JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

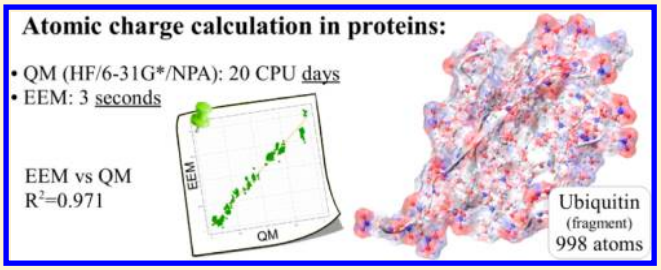
Rapid Calculation of Accurate Atomic Charges for Proteins via the Electronegativity Equalization Method

Crina-Maria Ionescu, Stanislav Geidl, Radka Svobodová Vařeková,* and Jaroslav Koča*

CEITEC—Central European Institute of Technology, and National Centre for Biomolecular Research, Faculty of Science, Masaryk University Brno, Kamenice 5, 625 00, Brno-Bohunice, Czech Republic

Supporting Information

ABSTRACT: We focused on the parametrization and evaluation of empirical models for fast and accurate calculation of conformationally dependent atomic charges in proteins. The models were based on the electronegativity equalization method (EEM), and the parametrization procedure was tailored to proteins. We used large protein fragments as reference structures and fitted the EEM model parameters using atomic charges computed by three population analyses (Mulliken, Natural, iterative Hirshfeld), at the Hartree−Fock level with two basis sets (6-31G*, 6-31G**) and in two environments (gas phase, implicit solvation). We parametrized and successfully validated 24 EEM models. When tested on insulin and ubiquitin, all models reproduced quantum mechanics level charges well and were consistent with respect to population analysis and basis set. Specifically, the models showed on average a correlation of 0.961, RMSD 0.097 e, and average absolute error per atom 0.072 e. The EEM models can be used with the freely available EEM implementation EEM_SOLVER.



INTRODUCTION

The concept of atomic point charges is well established in theoretical chemistry. Atomic point charges have played an important role in understanding and modeling chemical behavior by allowing the extraction and quantification of information stored in the molecular electron distribution of chemical compounds. Thus, atomic point charges have been used to estimate reactivity indices, dissociation constants, partition coefficients, the electrostatic contribution in molecular dynamics or docking studies, etc.[1−4] Virtual screening, the process by which potential drug candidates or targets are identified by screening huge libraries of compounds, employs atomic point charges, for instance, as parameters in the docking stage. Absorbtion, distribution, metabolism, and excretion profiling studies, involved in drug discovery programs, employ atomic point charges as descriptors useful in predicting various pharmacological properties. It is therefore desirable to have knowledge of the values of atomic charges.

While atomic charges are very intuitive, they are not physical observables. Nevertheless, atomic partial charges can be derived from physical observables, the values of which are most commonly obtained by quantum mechanical (QM) calculations. A remarkable fact is that a unique definition of atomic partial charges has yet to be accepted. As such, a score of methods have been developed to estimate the values of atomic charges, which is another indication that atomic charges are of great interest. Roughly, these methods can be classified as QM-based or empirical.

QM-based approaches first compute the wave function and subsequently the molecular electron density. In further, various population analyses can be performed in order to partition the molecular electron density into its atomic contributions. Some partitioning approaches are based on the wave function itself;[5−8] others on a wave function-dependent physical observable.[9−13] QM atomic charges can also be mapped to reproduce charge-dependent observables.[14] Along with physical soundness, the advantage of QM based approaches for charge calculation is their general applicability.

Empirical approaches, on the other hand, do not need the molecular wave function and instead employ a wide array of principles and intuitive derivations, coupled with cost-effective algorithms. Empirical approaches for atomic charge calculation can be divided into those which depend only on the 2D structure of the molecule,[15−18] and those which depend on the 3D structure of the molecule.[19−24] The majority of empirical approaches are based on partial or complete equalization of the electronegativity. Atomic charge calculation via empirical approaches is an extremely cost-effective alternative to QM-based methods, though appropriate parametrizations need to be performed before any empirical method can be used with confidence. Unlike QM based approaches, the applicability of empirical methods is many times limited to certain kinds of systems, either because of the intrinsic approximations that are made or by the parametrization procedure.

Although these atomic charge definitions differ in both the principles and algorithms they employ, many such definitions have already proven suitable for the prediction of relevant

Received: May 1, 2013
Published: August 22, 2013

chemical phenomena. Following a statistical analysis of 25 charge definition schemes, Meister and Schwarz[25] concluded that the same physical principles govern all the schemes under study, and therefore, all charge definition schemes are relevant, despite the differences in scale.

In the present study, we focus on proteins, due to their capital importance on all aspects of cellular life. Charge transfer was found to be significant in protein folding and many biomolecular interactions,[4,26−28] and functionally linked to protein structural dynamics.[29] Cho et al.[4] have shown that using conformationally dependent, QM-level atomic charges is critical to identifying the correct binding partners in docking studies on various proteins. Later, Cho and Rinaldo[30] found that, especially in the case of metalloproteins, it is necessary to compute QM charges for as many atoms as possible close to the metal ion in order to allow the correct prediction of ligand binding modes. Wallin et al.[31] emphasized the need to generalize and automatize the charge assignment procedure and include this procedure in the work flow of ligand preparation for binding free energy calculations.

In the light of the above-mentioned findings, it is clear that there is a need for a fast, accurate, and accessible procedure to calculate atomic charges in proteins. Nonetheless, an affordable, accurate, and highly available solution to handle large biomolecular systems at the QM level has not yet been found. Clearly, direct QM calculations are not yet a feasible option for charge calculation in proteins, and one must rely on some empirical method. In the present study, we focus on the electronegativity equalization method (EEM) as a method for rapid and accurate estimation of atomic partial charges.[20,32−45] EEM is fast, has a straightforward implementation, and is generally applicable and sensitive to a molecule's 3D structure. EEM has been successfully applied to zeolites, small organic molecules, and polypeptides.[46−50] A recent EEM investigation proved useful in mapping the allosteric activation mechanism of two apoptotic proteins.[44]

Before an empirical model can be used with confidence, appropriate parameters must be determined, i.e., the model must be parametrized. EEM models are parametrized by fitting the model parameters to a set of reference data usually made up of atomic charges obtained by various QM-based approaches. A QM-based charge calculation approach is characterized by the setup of the wave function calculation (theory level, basis set, environment), as well as by the population analysis (PA) used to partition the molecular electron density. We will refer to the sum of these characteristics as the "type" of charge produced by the QM approach. The maximum accuracy and potential application of any EEM model is given by the type of charge used during its parametrization. Previous works have parametrized EEM mainly for inorganic or small organic molecules, and especially for drug-like molecules,[20,33−45] but very few with specific focus on biomolecules. Some EEM models have been successfully tested on peptides of various size,[20,41] but these models cover only two types of charges and contain no parameters for sulfur, an element commonly found in proteins. Some EEM models were successfully tested on large protein fragments,[44] but these models only cover one type of charge.

Thus, the goal of the present study was to provide wider coverage and evaluation of the EEM approach for proteins. In this study, we have developed and validated 24 EEM models specifically tailored to cover 12 types of charges in proteins. Our reference data came from QM calculations at the HF level with two basis sets (6-31G*, 6-31G**), in the gas-phase and implicit solvent. Three atomic charge definitions were considered (Mulliken, Natural, iterative Hirshfeld). The accuracy of the proposed models was evaluated on insulin and ubiquitin.

THEORY AND METHODS

Electronegativity Equalization Method. The electronegativity equalization method (EEM)[20] enables the determination of atomic charges that are sensitive to the molecule's topology and three-dimensional structure. EEM is based on the electronegativity equalization principle,[51] which has received theoretical grounding within the density functional theory,[52,53] and which states that the electronegativity of all atoms is equalized throughout a molecule:

$$X_1 = X_2 = \ldots X_i = \ldots = \bar{X} \tag{1}$$

Within EEM, the electronegativity X_i of each atom i in a molecule can be approximated as a linear function of several terms:

$$X_1 = (X_i^0 + \Delta X_i) + 2(\eta_i^0 + \Delta\eta_i)q_i + k\sum_{i\neq j}\frac{q_j}{r_{ij}} \tag{2}$$

The first term is the electronegativity of the isolated atom X_i^0, empirically corrected for the presence of the molecular environment (ΔX_i). The second term is the product between the charge of the atom q_i and the hardness of the isolated atom η_i^0, empirically corrected for the presence of the molecular environment ($_\Delta\eta_i$). The last term $k\sum_{i\neq j}(q_j/r_{ij})$ accounts for the electrostatic interaction with every other charged atom j in the molecule. k is an adjusting factor first introduced by Yang and Shen.[54] Setting $A_i = X_i^0 + \Delta X_i$ and $B_i = 2(\eta_i^0 + \Delta\eta_i)$, the molecular electronegativity can be formally expressed as

$$\bar{X} = A_i + B_i q_i + k\sum_{i\neq j}\frac{q_j}{r_{ij}} \tag{3}$$

Additionally, the total molecular charge Q is the sum of all partial atomic charges q_i:

$$\sum q_i = Q \tag{4}$$

Taken all together, eqs 1, 3, and 4 can be expressed as a system of equations from which the partial atomic charges q_i and the molecular electronegativity $\bar{X}$ can be calculated, provided that the rest of the terms (Q, r_{ij}, k, A_i, B_i) are known:

$$\begin{pmatrix} B_1 & \frac{k}{r_{1,2}} & \ldots & \frac{k}{r_{1,N}} & -1 \\ \frac{k}{r_{2,3}} & B_2 & \ldots & \frac{k}{r_{2,N}} & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{k}{r_{N,1}} & \frac{k}{r_{N,2}} & \ldots & B_N & -1 \\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix}\begin{pmatrix} q_1 \\ q_2 \\ \ldots \\ q_N \\ \bar{X} \end{pmatrix} = \begin{pmatrix} -A_1 \\ -A_2 \\ \ldots \\ A_N \\ Q \end{pmatrix} \tag{5}$$

This formalism estimates atomic charges via a set of coupled linear equations which can be efficiently solved by a Gaussian elimination procedure.[55] The computational complexity is $\theta(N^3)$, where N is the number of atoms in the molecule.

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

Table 1. Atomic Composition of the Data Sets Used in This Study

	Nr molecules	Nr atoms	H	C	N	O	S	Ca
training set	41	40142	19879	11912	3188	4954	148	61
insulin	1	802	397	248	63	88	6	
ubiquitin	1	998	505	306	88	99		

EEM Model Parameterization Theory. Given the three-dimensional structure of the molecule and its total charge, partial charges for all atoms in the molecule can be calculated only if the values of the EEM parameters (k, $\langle A_i, B_i \rangle$) are known. Otherwise, the EEM model needs to be parametrized by fitting against reference data (training data set). We employed QM atomic charges to fit the EEM parameters given by the electronegativity and hardness contributions $A_i = X_i^0 + \Delta X_i$ and $B_i = 2(\eta_i^0 + \Delta\eta_i)$, respectively, as well as the value of the parameter k present in this formalism. The EEM parametrization methodology employed here was described by Svobodová Vařeková et al.,[37] and its principles are given below.

For the purpose of EEM model parametrization, eq 3 can be rearranged as a linear equation in A_i and B_i for each atom i in the system:

$$A_i + B_i q_i = \bar{X} - k \sum_{i \neq j} \frac{q_j}{r_{ij}} \tag{6}$$

For a given value of the parameter k, sets of eqs 6 can be grouped together according to the kind of atom they refer to. The values of atomic charges q_i and interatomic distances r_{ij} are taken from the training data set, and the value of the molecular electronegativity $\bar{X}$ can be calculated as an average of the isolated atom electronegativities X_i^0. Under these circumstances, each group of linear equations becomes an overdetermined system of equations, enabling the determination of a set of parameters $(A, B)_m$ for each atom type m by least-squares minimization. The classification of atoms into types can be done according to various criteria, such as chemical element, hybridization, binding partners, etc. Once a set of parameters $(A, B)_m$ has been obtained for all M atom types, for all given values of k, it is possible to determine the optimal EEM parameter set $k[(A, B)_1, (A, B)_2, ..., (A, B)_M]$, in further denoted simply as $k[(A, B)_m]_1^M$, as the parameter set which produces the best EEM model in the internal validation step (see below).

EEM Model Validation Theory. The accuracy of an EEM model in reproducing the reference data (here, QM atomic charges) can be evaluated by internal and external validation. In the internal validation step, the EEM model is evaluated for its ability to reproduce the reference data which was used during its parametrization. In other words, EEM is first validated on the training data set. Next, in the external validation step, the EEM model is evaluated for its ability to reproduce reference data which was not used during its parametrization. In other words, EEM is also validated on test molecules. The correlation between the reference QM charges and predicted EEM charges can be assessed by various indicators.

The first indicator used in the present study was the average correlation coefficient (squared Pearson's correlation coefficient), computed for each molecule, and averaged over all molecules in the data set:

$$R_{\text{avg}} = \frac{\sum_{I=1}^{N} \left(\frac{\sum_{i=1}^{n_I} (q_i^{\text{QM}} - \overline{q_I^{\text{QM}}})(q_i^{\text{EEM}} - \overline{q_I^{\text{EEM}}})}{(n_I - 1)\sigma_I^{\text{QM}} \sigma_I^{\text{EEM}}} \right)}{N} \tag{7}$$

where the index $i = 1...n_I$ described all atoms in molecule I, $\overline{q_I^{\text{QM}}}$ and $\overline{q_I^{\text{EEM}}}$ represented average atomic charges in molecule I, σ_I^{QM} and σ_I^{EEM} were standard deviations of the atomic charges in molecule I, n_I was the number of atoms in molecule I, and N was the number of molecules in a given set.

The second indicator was the root-mean-square deviation, computed for each molecule, and averaged over all molecules in the data set:

$$\text{RMSD}_{\text{avg}} = \frac{\sum_{I=1}^{N} \sqrt{\frac{\sum_{i=1}^{n_I} (q_i^{\text{QM}} - q_i^{\text{EEM}})^2}{n_I}}}{N} \tag{8}$$

The third indicator was the average absolute difference, computed for each molecule and averaged over all molecules in the data set:

$$D_{\text{avg}} = \frac{\sum_{I=1}^{N} \sqrt{\frac{\sum_{i=1}^{n_I} (q_i^{\text{QM}} - q_i^{\text{EEM}})^2}{n_I}}}{N} \tag{9}$$

Reference Structures—Training Data Set. The reference molecules used in the model parametrization and internal validation steps were 41 fragments of proteins from the Protein Data Bank (PDB), whose structures had been determined by X-ray crystallography or solution state NMR experiments. Protein fragments, rather than small molecules, were chosen as reference structures since they reflect the complex nature of proteins as long, non-neutral molecular chains with complex 3D assembly.

The 41 reference structures consisted of amino acid chains, water molecules, and calcium ions, and were obtained from the 3D structure of their parent protein using the program Triton.[56] Hydrogen atoms were added for all crystal structures to satisfy the missing valences. The protonation states of the amino acid residues were +1 for Arg, Lys, and His and charged amino ends of the polypeptide chains, −1 for Glu, Asp, and charged carboxyl ends of the polypeptidic chains, 0 for the rest. Only the first structural model was used in the case of NMR structures.

Reference Structures—Test Molecules. Two small proteins were used in the external validation step, namely insulin (PDB ID 3E7Y[57]) and ubiquitin (PDB ID 1UBQ[58]). The nature of the calculations imposed some limitations to the size of the systems used for testing. Thus, only one of the two insulin monomers was kept, containing chains C and D, along with a few water molecules cocrystallized with the protein. None of the Cl and Zn ions found cocrystallized with insulin were kept, as no EEM parameters were obtained for these ions. In the case of ubiquitin, the first 14 residues had to be removed in order to reduce the system to a manageable size.

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

Table 2. Overview of the 12 Types of QM Calculations Performed, and 24 EEM Models Obtained in This Study

EEM model characteristics						
	QM scheme			training data sets		
atom type classification	PA	basis set	environment	Nr molecules	Nr atoms	EEM model
E	MPA	6-31G*	gas phase	41	40142	E-MPA/6-31G*/gas
			PCM	41	40142	E-MPA/6-31G*/PCM
		6-31G**	gas phase	41	40142	E-MPA/6-31G**/gas
			PCM	41	40142	E-MPA/6-31G**/PCM
	NPA	6-31G*	gas phase	41	40142	E-NPA/6-31G*/gas
			PCM	40	39061	E-NPA/6-31G*/PCM
		6-31G**	gas phase	41	40142	E-NPA/6-31G**/gas
			PCM	40	39061	E-NPA/6-31G**/PCM
	HiI	6-31G*	gas phase	41	40142	E-HiI/6-31G*/gas
			PCM	38	36875	E-HiI/6-31G*/PCM
		6-31G**	gas phase	41	40142	E-HiI/6-31G**/gas
			PCM	40	39061	E-HiI/6-31G**/PCM
EX	MPA	6-31G*	gas phase	41	40142	EX-MPA/6-31G*/gas
			PCM	41	40142	EX-MPA/6-31G*/PCM
		6-31G**	gas phase	41	40142	EX-MPA/6-31G**/gas
			PCM	41	40142	EX-MPA/6-31G**/PCM
	NPA	6-31G*	gas phase	41	40142	EX-NPA/6-31G*/gas
			PCM	40	39061	EX-NPA/6-31G*/PCM
		6-31G**	gas phase	41	40142	EX-NPA/6-31G**/gas
			PCM	40	39061	EX-NPA/6-31G**/PCM
	HiI	6-31G*	gas phase	41	40142	EX-HiI/6-31G*/gas
			PCM	38	36875	EX-HiI/6-31G*/PCM
		6-31G**	gas phase	41	40142	EX-HiI/6-31G**/gas
			PCM	40	39061	EX-HiI/6-31G**/PCM

An overview of the composition of all molecules used for EEM model parametrization and validation is given in Table 1. The 3D structures of these fragments are available as Supporting Information in PDB format.

Reference QM Atomic Charges. For each reference structure, QM atomic charges were obtained from the Mulliken Population Analysis (MPA),[5,6] Natural Population Analysis (NPA),[8] and iterative Hirshfeld analysis (HiI)[11] performed at the Hartree−Fock (HF) theory level with 6-31G* and 6-31G** basis sets in the gas phase and polarizable continuum model (PCM).[59] These three population analyses were chosen because they are known to work well with EEM.[34,50] Moreover, these population analyses can be performed by currently available software tools for very large protein fragments that we used as training and test molecules.

A total of 12 QM reference data sets were thus obtained: MPA/6-31G*/gas, MPA/6-31G*/PCM, MPA/6-31G**/gas, MPA/6-31G**/PCM, NPA/6-31G*/gas, NPA/6-31G*/PCM, NPA/6-31G**/gas, NPA/6-31G**/PCM, HiI/6-31G*/gas, HiI/6-31G*/PCM, HiI/6-31G**/gas, HiI/6-31G**/PCM. All QM single point energy calculations and MPA were performed using Gaussian 09,[60] while NPA was performed using the NBO program,[61] and HiI was performed using the HiPart program.[62] The values of all QM atomic charges obtained in this study are included as Supporting Information in csv format.

EEM Models. Two classifications of atoms into atom types were employed here. One classification was based on chemical elements (denoted "E"), and the other on chemical elements and maximum bond multiplicity for each atom (denoted "EX", so that, for example, "O1" indicates sp^3 hybridized oxygen and "O2", sp^2 hybridized oxygen). Both atom classification approaches were applied to all 12 QM reference data sets.

Within a given atom classification, the parameter k was sampled as discrete values on several intervals. For each discrete value of k, the set of parameters $(A, B)_m$ were obtained for all atom types in the given classification. For each atom type m, the parameters $(A, B)_m$ were determined by least-squares minimization, as the values of all other variables were known: the interatomic distances were calculated from the 3D atomic coordinates of the reference structures, the atomic charges were calculated by the 12 QM schemes described above, and the value of the average electronegativity for each molecule was calculated as the harmonic average of the electronegativities of the constituent atoms[63] $\bar{X} = n(\sum_{i=1}^{n}(1/X_i^0))^{-1}$, where n is the number of atoms in the molecule, and the values of X_i^0 correspond to Pauling electronegativities.[64,65] Thus, within each of the two atom type classifications, for each discrete value of k, an EEM model described by the parameter set $k[(A, B)_m]_1^M$ was obtained. Internal validation was then performed for all such EEM models. Namely, each model was used to predict the EEM charges of the reference molecules in the training data set. The atomic charges predicted by the EEM model were compared against the reference QM atomic charges used during the model parametrization. The EEM model which gave the highest R_{avg} between the EEM predicted charges and the QM reference charges for the whole training set was designated as the final EEM model within the given atom type classification.

Finally, 12 EEM models were obtained for the first atom classification E, based on chemical elements: E-MPA/6-31G*/gas, E-MPA/6-31G*/PCM, E-MPA/6-31G**/gas, E-MPA/6-31G**/PCM, E-NPA/6-31G*/gas, E-NPA/6-31G*/PCM, E-NPA/6-31G**/gas, E-NPA/6-31G**/PCM, E-HiI/6-31G*/gas, E-HiI/6-31G*/PCM, E-HiI/6-31G**/gas, E-HiI/

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

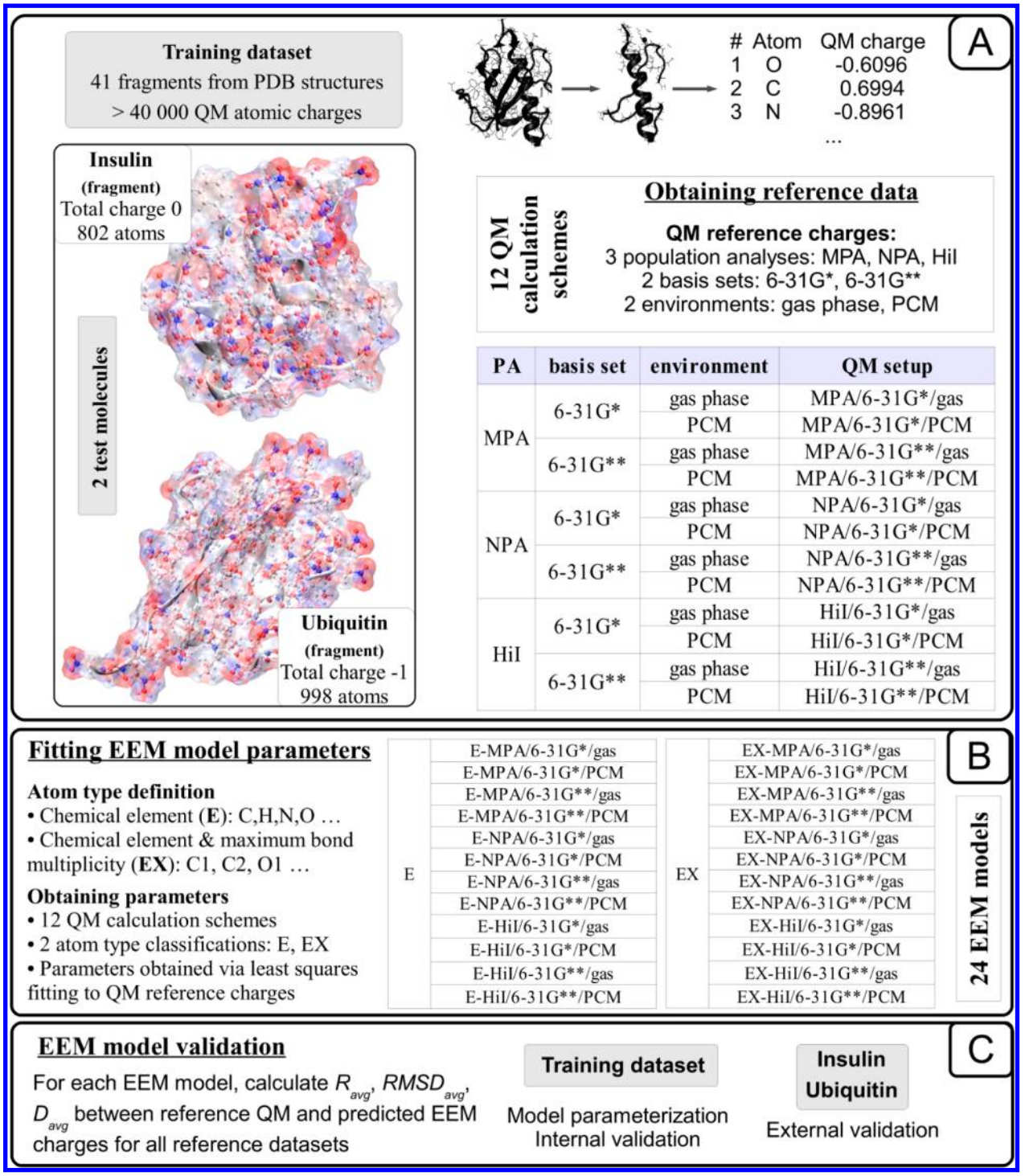


Figure 1. Flowchart of parametrization of EEM models for calculating partial atomic charges in proteins. (A) The reference data used in this study consisted of QM atomic charges for one training data set of protein fragments, and two test molecules (insulin and ubiquitin). Due to the system size limitations imposed by the QM calculation, only one insulin monomer was used, and the first 14 residues were removed from the ubiquitin structure. In total, 12 different QM atomic charge calculation schemes were used. (B) For each of the 12 QM charge schemes, two atom type classification approaches were employed for EEM model parametrization. Parameters for 24 EEM models were fitted onto the QM charges of the training data set. (C) Each EEM model was subjected to internal and external validation by comparing the EEM charges against reference QM charges of the training and test molecules, respectively. Performance was evaluated by 3 indicators, namely the average correlation coefficient (R_{avg}), the average root-mean-square deviation ($RMSD_{avg}$), and the average absolute difference (D_{avg}).

6-31G**/PCM. Additionally, 12 EEM models were obtained for the second atom classification EX, based on chemical elements and maximum bond multiplicity: EX-MPA/6-31G*/gas, EX-MPA/6-31G*/PCM, EX-MPA/6-31G**/gas, EX-MPA/6-31G**/PCM, EX-NPA/6-31G*/gas, EX-NPA/6-31G*/PCM, EX-NPA/6-31G**/gas, EX-NPA/6-31G**/PCM, EX-HIl/6-31G*/gas, EX-HIl/6-31G*/PCM, EX-HIl/6-31G**/gas, EX-HIl/6-31G**/PCM. All 24 EEM models

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

Table 3. Validation of the 24 EEM Models Obtained in This Study[a]

EEM model	internal validation (training data set)			external validation (insulin)			external validation (ubiquitin)		
	R_{avg}	$RMSD_{avg}$ [e]	D_{avg} [e]	R_{avg}	$RMSD_{avg}$ [e]	D_{avg} [e]	R_{avg}	$RMSD_{avg}$ [e]	D_{avg} [e]
E-MPA/6-31G*/gas	0.975	0.079	0.062	0.972	0.080	0.064	0.969	0.083	0.065
E-MPA/6-31G*/PCM	0.976	0.080	0.065	0.973	0.081	0.066	0.972	0.081	0.066
E-MPA/6-31G**/gas	0.977	0.067	0.053	0.976	0.065	0.052	0.973	0.069	0.054
E-MPA/6-31G**/PCM	0.977	0.068	0.053	0.976	0.066	0.052	0.973	0.071	0.054
E-NPA/6-31G*/gas	0.975	0.083	0.064	0.971	0.086	0.066	0.971	0.086	0.066
E-NPA/6-31G*/PCM	0.975	0.092	0.067	0.971	0.091	0.065	0.973	0.087	0.063
E-NPA/6-31G**/gas	0.976	0.084	0.065	0.971	0.086	0.067	0.971	0.086	0.066
E-NPA/6-31G**/PCM	0.975	0.093	0.068	0.971	0.092	0.066	0.973	0.088	0.063
E-HiI/6-31G*/gas	0.965	0.125	0.092	0.965	0.114	0.083	0.969	0.113	0.080
E-HiI/6-31G*/PCM	0.948	0.123	0.087	0.948	0.116	0.081	0.958	0.108	0.075
E-HiI/6-31G**/gas	0.967	0.121	0.088	0.965	0.111	0.080	0.970	0.109	0.077
E-HiI/6-31G**/PCM	0.962	0.128	0.094	0.961	0.119	0.087	0.965	0.116	0.084
EX-MPA/6-31G*/gas	0.976	0.076	0.056	0.970	0.082	0.059	0.974	0.076	0.058
EX-MPA/6-31G*/PCM	0.979	0.073	0.056	0.976	0.075	0.057	0.975	0.076	0.058
EX-MPA/6-31G**/gas	0.975	0.070	0.053	0.969	0.073	0.056	0.971	0.072	0.055
EX-MPA/6-31G**/PCM	0.961	0.090	0.069	0.958	0.087	0.067	0.959	0.088	0.065
EX-NPA/6-31G*/gas	0.979	0.077	0.059	0.974	0.081	0.062	0.975	0.080	0.061
EX-NPA/6-31G*/PCM	0.978	0.079	0.063	0.975	0.080	0.063	0.974	0.081	0.063
EX-NPA/6-31G**/gas	0.979	0.077	0.062	0.975	0.081	0.062	0.975	0.080	0.061
EX-NPA/6-31G**/PCM	0.980	0.076	0.062	0.977	0.078	0.062	0.976	0.079	0.062
EX-HiI/6-31G*/gas	0.919	0.148	0.109	0.913	0.146	0.107	0.934	0.130	0.095
EX-HiI/6-31G*/PCM	0.910	0.166	0.118	0.911	0.156	0.110	0.930	0.142	0.099
EX-HiI/6-31G**/gas	0.910	0.156	0.118	0.903	0.154	0.117	0.927	0.137	0.103
EX-HiI/6-31G**/PCM	0.917	0.161	0.117	0.912	0.156	0.112	0.931	0.141	0.101

[a]Internal validation denotes validation for the molecules in the training data set, while external validation denotes validation on the test molecules. Statistical descriptors comprising the average correlation coefficient (R_{avg}), the average root mean square deviation ($RMSD_{avg}$), and the average absolute difference (D_{avg}) are given. All quantities are given in elementary charges (1 e $\sim$ 1.602×10^{-19} coulombs).

were further subjected to external validation, whereby they were evaluated for their ability to predict the atomic charges of the test molecules insulin and ubiquitin.

An overview of all QM calculations performed, and all EEM models obtained is given in Table 2. An overview of the entire EEM model parametrization procedure employed in this study is given in Figure 1. The parametrization and validation of the EEM models were done using an in-house program. The visualization of structural models in various representations was performed using VMD.[66] The values of all EEM atomic charges obtained in this study are included as Supporting Information in csv format. The parameters of all 24 EEM models developed in this study are given in Table S1 of the Supporting Information.

RESULTS AND DISCUSSION

Performance of the EEM Models. In this study, we obtained 24 EEM models (Figure 1, Table 2). Each EEM model was validated with respect to R_{avg}, $RMSD_{avg}$, and D_{avg} between reference QM charges and predicted EEM charges. The results of the internal and external validation procedures are given in Table 3 and Figure 2 (see also Supporting Information Table S2).

Overall, there is good agreement between QM and EEM charges, suggesting that these EEM models can be used reliably for rapid atomic charge calculation in proteins. With respect to internal validation (i.e., validation against the training data sets), all 24 models have R_{avg} over 0.91, while 19 models have R_{avg} over 0.95, and 15 models have R_{avg} over 0.97. All models have $RMSD_{avg}$ less than 0.17 e and D_{avg} less than 0.12 e, while 16 models have $RMSD_{avg}$ less than 0.10 e and D_{avg} less than 0.07 e. With respect to external validation on insulin and ubiquitin, all EEM models perform comparably as in the internal validation step, suggesting high transferability. Specifically, all 24 models have R_{avg} over 0.9, while 19 models have R_{avg} over 0.95, and 14 models have R_{avg} over 0.97. All models have $RMSD_{avg}$ less than 0.16 e and D_{avg} less than 0.12 e, while 16 models have $RMSD_{avg}$ less than 0.10 e and D_{avg} less than 0.07 e. The values obtained for the majority of models are comparable to previous EEM parametrizations published in the literature.

There are diverging trends in literature regarding the level of detail that should be included in the atom type classification used for EEM model parametrization. One direction supports the use of general atom types,[33] while the other direction supports the use of very detailed atom types.[40,41] Our results in Table 3 (see also Supporting Information Table S2) suggested that the finer grained atom classification EX only modestly improved the accuracy compared to the classification E based on chemical elements alone, and not for all models. In any case, both atom classifications can provide satisfactory results for proteins, but we rather support the use of general atom types to avoid the extra step and possible errors associated with assigning hybridizations. Thus, in further we discuss only the results for the models based on general atom types E, while the complete results can be found in Supporting Information Table S2.

Accuracy of the EEM Models. A more detailed view of the accuracy that can be expected for the EEM models developed in this study is given in Figure 3. Approximately 50% of the predicted values are expected to contain an error of less than 0.05 e, while approximately 30% of the predicted values are expected to contain an error between 0.05 and 0.1 e. Finally, approximately 5% of the values are expected to contain an error larger than 0.2 e.

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

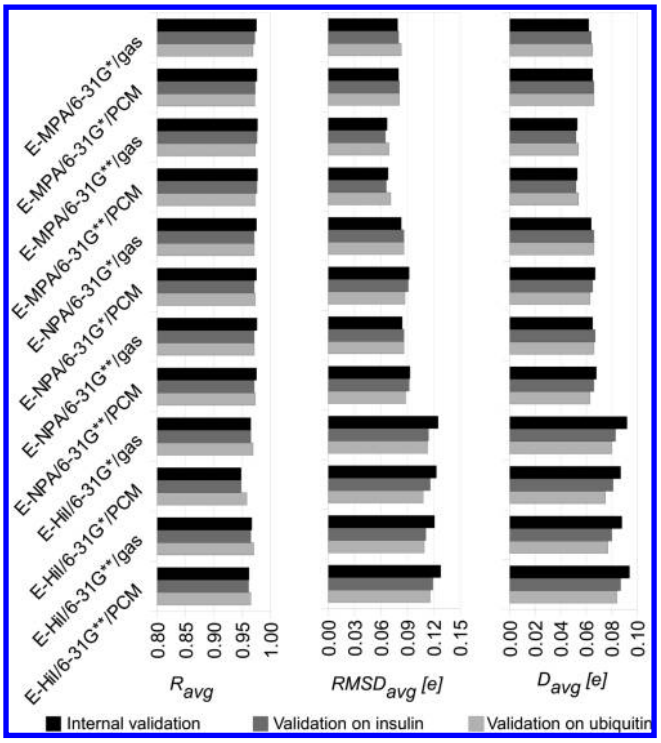


Figure 2. Validation of EEM models by comparing the predicted EEM atomic charges against the reference QM atomic charges. Internal validation denotes validation for the molecules in the training data set. Statistical descriptors comprising the average correlation coefficient (R_{avg}), the average root-mean-square deviation ($RMSD_{avg}$), and the average absolute difference (D_{avg}) are given. All quantities are given in elementary charges (1 e ~ 1.602×10^{-19} coulombs). The names of the EEM models encode the QM scheme and atom classification used in their parametrization. Only the models based on general atom types (E) are given here, while the complete results can be found in Table 3 (see also Supporting Information Table S2). Good agreement between QM and EEM charges was found for all data sets, as R_{avg} is close to 1, and $RMSD_{avg}$ and D_{avg} are minimal.

With respect to the practical implications of employing EEM models for calculating atomic partial charges, it is worth discussing the expected accuracy of the EEM models in reproducing not only QM atomic charges, but also various quantities derived from these charges. Since it is not possible to give a general evaluation of the expected accuracy of the EEM models in all possible practical applications, we focus here on two common uses of atomic charges, namely electrostatic potential (ESP) and docking calculations.

Figure 4 (see also Supporting Information Table S3) shows that the ESP computed from EEM charges deviates from the ESP computed from QM charges on average by 0.0071 au for insulin and 0.0058 au for ubiquitin, respectively; whereas, the average correlation coefficient is 0.9 for insulin and 0.87 for ubiquitin.

To evaluate the accuracy of EEM charges in a typical docking calculation, we performed the blind docking of glycerol onto the ubiquitin fragment previously used in our study as a test molecule for the validation of EEM models. Glycerol was chosen as a potential ligand because it has been found to stabilize the native state of ubiquitin.[68] The ubiquitin fragment was used as a receptor in favor of the native ubiquitin dimer because of the size restrictions imposed by the reference QM calculations. Figure 5 (see also Supporting Information Table S4) shows that the docking results obtained using QM charges are well reproduced using the EEM charges given by the corresponding EEM model. The EEM binding pose differs by 0.07 kcal/mol and an RMSD of 0.131 Å from the binding pose given using QM charges. By comparison, using Gasteiger−Marsili or AMBER ff94[69] charges on ubiquitin produces different binding poses. Specifically, the binding pose given by Gasteiger−Marsili charges differs from that given by QM charges by 0.99 kcal/mol and an RMSD of 3.244 Å. The binding pose given by AMBER ff94 charges differs by 1.57 kcal/mol and an RMSD of 3.235 Å.

Sensitivity of the EEM Models. We have evaluated the performance and predicted the accuracy of the EEM models developed in this study. It is also necessary to translate the meaning of this accuracy with respect to the ability of the EEM models to differentiate between types of charges, since many times different modeling applications rely on different types of charges. For example, in the particular case of pK_a prediction via

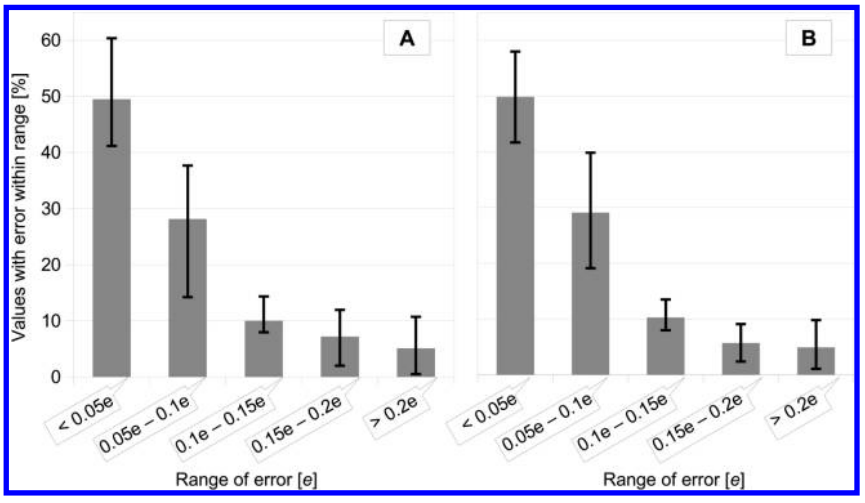


Figure 3. Expected accuracy of the EEM models developed in this study, measured as the percentage of atoms for which the error of the EEM vs QM prediction lies in a certain interval. Only the models based on general atom types E are included in this analysis. The error is computed as the absolute difference between the predicted EEM value and reference QM value of the atomic charge. The gray blocks represent the error distribution averaged over all 12 EEM models obtained in this study, while the black lines indicate the minimum and maximum values for these distributions. (A) Values obtained for insulin. (B) Values obtained for ubiquitin.

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

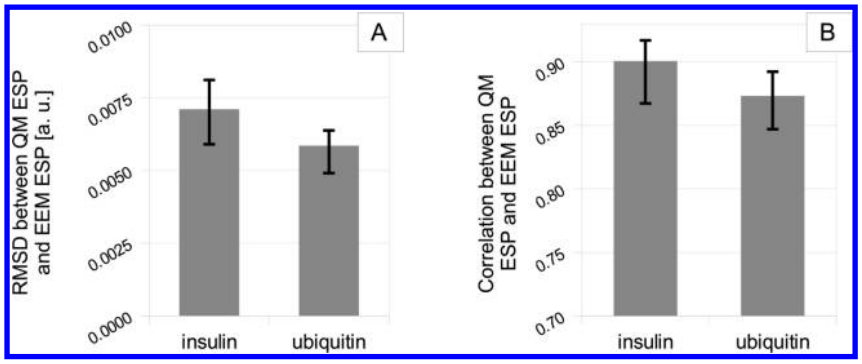


Figure 4. Comparison between the QM and EEM based ESP for insulin and ubiquitin. Only the models based on general atom types E are included in this analysis. The gray blocks represent averages over all 12 EEM models, while the black lines indicate the minimum and maximum values. The ESP was calculated using the biomolecular electrostatics software APBS.[67] (A) Deviation between QM and EEM based ESP measured as RMSD. (B) Correlation between QM and EEM based ESP measured as Pearson's squared correlation coefficient.

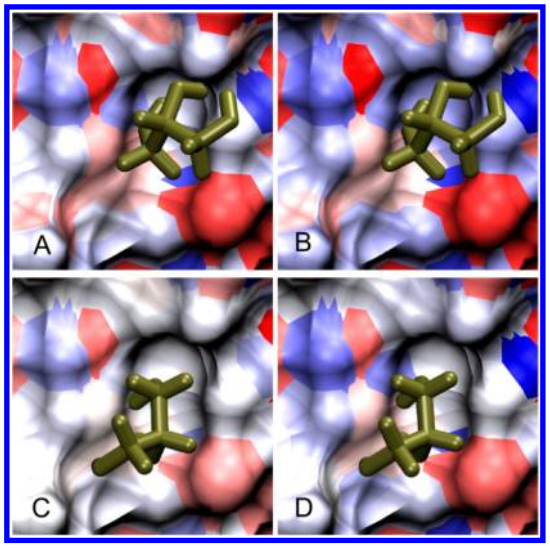


Figure 5. Blind docking of glycerol (ochre, licorice representation) onto the ubiquitin fragment (surface representation colored according to atomic charges, where red signifies more negative charge, while blue signifies more positive charge). The fixed receptor was placed in a 100 × 90 × 90 sized grid, with 0.35 Å spacing between the grid points. The ligand's initial conformation was taken from the coordinates of the ideal glycerol model available in Ligand Expo[70] and contained five rotatable bonds. All calculations were set up using Triton[56] and performed with AutoDock4.2.[71] Each docking run employed 200 iterations of the Genetic Algorithm, with up to 1 000 000 energy evaluations for each iteration (the rest of the parameters were kept in their default values). The calculations differ in the type of atomic charges used for the receptor, while the ligand always carried Gasteiger−Marsili charges. In each case, the binding pose which gives the best binding energy in five independent docking runs is chosen for comparison. (A) Using MPA/6-31G*/gas QM charges—estimated binding energy −9.64 kcal/mol. (B) Using E-MPA/6-31G*/gas EEM charges—estimated binding energy −9.71 kcal/mol, RMSD 0.132 Å compared to the QM pose. (C) Using Gasteiger−Marsili charges—estimated binding energy −8.65 kcal/mol, RMSD 3.244 Å compared to the QM pose; (D) Using AMBER ff94 charges—estimated binding energy −8.07 kcal/mol, RMSD 3.235 Å compared to the QM pose.

QSPR modeling, it was shown that QSPR models which use QM MPA charges as descriptors perform better than QSPR models which use QM charges derived from electrostatic potentials (ESP charges).[72,73] On the other hand, QSPR models which use EEM ESP charges perform comparably well as the QSPR models which use EEM MPA charges,[50] suggesting that the error of the EEM vs QM prediction could very well overwhelm the distinction between different atomic charge definitions.

Therefore, one last aspect to be studied is the sensitivity of the EEM models produced in this study with respect to population analysis, basis set, and environment. For this purpose, we inspected the correlation between the QM charges produced by each of the 12 QM schemes, and the EEM charges produced by all 24 EEM models (see Supporting Information Table S2). Table 4 contains the conclusions for the 12 EEM models based on general atom types E, for insulin and ubiquitin.

It is clear from Table 4 (see also Supporting Information Table S2) that the models obtained in this study easily distinguish between population analyses. In all cases, EEM models parametrized onto MPA charges provide the best prediction for QM MPA data. Similarly, EEM models parametrized onto NPA charges always provide the best prediction for QM NPA data, and EEM models parametrized onto HiI charges always provide the best prediction for QM HiI data. Further, in the case of MPA charges, the models easily distinguish between basis sets, which is expected, since MPA charges are known to be basis set dependent. As such, EEM models parametrized onto MPA/6-31G* charges provide the best prediction for QM MPA/6-31G* data. The equivalent is true for 6-31G**. In the case of NPA and HiI charges, the models do not distinguish between basis sets, which correlates with the fact that NPA and HiI charges have low basis set dependence.[74] As such, EEM models parametrized onto NPA/6-31G* charges provide the best prediction for both QM NPA/6-31G* and NPA/6-31G** data. Similarly, EEM models parametrized onto HiI/6-31G* charges provide the best prediction for both QM HiI/6-31G* and HiI/6-31G** data. On the other hand, the EEM models are not able to distinguish between the gas phase and implicit solvation conditions. In many cases, EEM models parametrized onto gas phase data provide the best prediction for QM PCM data, and vice versa.

Availability of Implementation. The models developed in the present study are fully compatible with any implementation of the EEM formalism given by eq 3, allowing the computation of accurate, conformationally dependent atomic charges in proteins with hundreds of residues in only a few minutes. In particular, these models can be used with EEM_SOLVER, our previously published EEM implementation, which has already been implemented in the Parallel Virtual Machine (PVM) environment and scales very favorably for large systems.[55]

dx.doi.org/10.1021/ci400448n | J. Chem. Inf. Model. 2013, 53, 2548−2558

Table 4. Verification of EEM Model Consistency with QM Calculation Scheme with Respect to Population Analysis, Basis Set, and Environment[a]

R_{avg}	best E-EEM model	
EEM model	insulin	ubiquitin
MPA/6-31G*/gas	MPA/6-31G*/gas	MPA/6-31G*/PCM
MPA/6-31G*/PCM	MPA/6-31G*/PCM	MPA/6-31G*/PCM
MPA/6-31G**/gas	MPA/6-31G**/gas	MPA/6-31G**/gas
MPA/6-31G**/PCM	MPA/6-31G**/gas	MPA/6-31G**/gas
NPA/6-31G*/gas	NPA/6-31G*/PCM	NPA/6-31G*/PCM
NPA/6-31G*/PCM	NPA/6-31G*/gas	NPA/6-31G*/PCM
NPA/6-31G**/gas	NPA/6-31G*/gas	NPA/6-31G**/PCM
NPA/6-31G**/PCM	NPA/6-31G*/gas	NPA/6-31G*/PCM
HiI/6-31G*/gas	HiI/6-31G**/gas	HiI/6-31G**/gas
HiI/6-31G*/PCM	HiI/6-31G**/gas	HiI/6-31G**/gas
HiI/6-31G**/gas	HiI/6-31G**/gas	HiI/6-31G**/gas
HiI/6-31G**/PCM	HiI/6-31G**/gas	HiI/6-31G**/gas
D_{avg}	best E-EEM model	
EEM model	insulin	ubiquitin
MPA/6-31G*/gas	MPA/6-31G*/gas	MPA/6-31G*/gas
MPA/6-31G*/PCM	MPA/6-31G*/gas	MPA/6-31G*/gas
MPA/6-31G**/gas	MPA/6-31G**/gas	MPA/6-31G**/gas
MPA/6-31G**/PCM	MPA/6-31G**/gas	MPA/6-31G**/gas
NPA/6-31G*/gas	NPA/6-31G**/PCM	NPA/6-31G**/PCM
NPA/6-31G*/PCM	NPA/6-31G*/gas	NPA/6-31G**/PCM
NPA/6-31G**/gas	NPA/6-31G*/gas	NPA/6-31G**/PCM
NPA/6-31G**/PCM	NPA/6-31G*/gas	NPA/6-31G**/PCM
HiI/6-31G*/gas	HiI/6-31G*/gas	HiI/6-31G*/PCM
HiI/6-31G*/PCM	HiI/6-31G*/PCM	HiI/6-31G*/PCM
HiI/6-31G**/gas	HiI/6-31G*/PCM	HiI/6-31G*/PCM
HiI/6-31G**/PCM	HiI/6-31G*/PCM	HiI/6-31G*/PCM
$RMSD_{avg}$	best E-EEM model	
EEM model	insulin	ubiquitin
MPA/6-31G*/gas	MPA/6-31G*/gas	MPA/6-31G*/PCM
MPA/6-31G*/PCM	MPA/6-31G*/gas	MPA/6-31G*/PCM
MPA/6-31G**/gas	MPA/6-31G**/gas	MPA/6-31G**/gas
MPA/6-31G**/PCM	MPA/6-31G**/gas	MPA/6-31G**/gas
NPA/6-31G*/gas	NPA/6-31G*/gas	NPA/6-31G**/PCM
NPA/6-31G*/PCM	NPA/6-31G*/gas	NPA/6-31G**/PCM
NPA/6-31G**/gas	NPA/6-31G*/gas	NPA/6-31G*/gas
NPA/6-31G**/PCM	NPA/6-31G*/gas	NPA/6-31G*/gas
HiI/6-31G*/gas	HiI/6-31G**/gas	HiI/6-31G*/PCM
HiI/6-31G*/PCM	HiI/6-31G*/PCM	HiI/6-31G*/PCM
HiI/6-31G**/gas	HiI/6-31G**/gas	HiI/6-31G*/PCM
HiI/6-31G**/PCM	HiI/6-31G*/PCM	HiI/6-31G*/PCM

[a]Only the analysis for models based on E atom type classification is given here, while the complete results can be found in Supporting Information Table S2.

CONCLUSION

The goal of the study was to extend the coverage of the Electronegativity Equalization Method (EEM) for atomic charge calculation in proteins. For the purpose of EEM model parametrization, fragments of experimentally determined protein structures were used as reference systems. Reference QM atomic charges given by three population analyses were calculated at the HF level, with two basis sets and in two environments. A total of 24 EEM models were parametrized and evaluated in the present study.

Upon validation on insulin and ubiquitin, the models exhibited high correlation and low deviations between the QM reference charges and EEM predicted charges. Very good accuracy is expected for 80% of the predictions, while poor accuracy is expected for 5% of the predictions. Last, the models were found to be consistent with respect to basis set and population analysis.

Overall, the results of the evaluation suggest that the EEM models obtained in this study can be used reliably for the rapid calculation of conformationally dependent atomic charges in proteins and protein complexes. The models can be used with our previously published EEM implementation EEM_SOLVER, which allows for an efficient estimation of atomic charges with application in biomolecular modeling investigations.

ASSOCIATED CONTENT

Supporting Information

Structures of all reference molecules (.pdb format), values of all QM and EEM atomic charges (.csv format), values of parameters for all EEM models (Table S1, .pdf format), results of the validation of all EEM models against all QM schemes (Table S2, .pdf format), reproduction of ESP for insulin and ubiquitin (Table S3, .pdf format), and comparison of docking runs using different types of charges for the receptor (Table S4, .pdf format). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jkoca@ceitec.muni.cz.

*E-mail: svobodova@chemi.muni.cz.

Author Contributions

The study was conceived by J.K., while the procedure was designed by C.-M.I. and R.S.V. The calculations were performed by C.-M.I. and S.G., and the data was analyzed by all authors. The manuscript was drafted by C.-M.I. and critically revised by all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

C.-M.I. is very grateful to Tomáš Bouchal for the graphic design ideas and to Dr. Sushil Kumar Mishra for the useful discussions. This work was funded by the Ministry of Education, Youth and Sports of the Czech Republic (contract number LH13055) and the European Community's Seventh Framework Programme (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund and from the "Capacities" specific program (Contract No. 286154). The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is highly appreciated. C.-M.I. would like to thank Brno City Municipality for the financial support provided to her through the program Brno Ph.D. Talent.

ABBREVIATIONS

E, atom type classification approach based on chemical elements; EEM, Electronegativity Equalization Method; ESP, electrostatic potential; EX, atom type classification approach based on chemical elements and maximum bond multiplicity; HiI, iterative Hirshfeld population analysis; MPA, Mulliken Population Analysis; NMR, nuclear magnetic resonance; NPA, natural

population analysis; PA, population analysis; PCM, polarizable continuum model; PDB, Protein Data Bank; PVM, parallel virtual machine; QM, quantum mechanics/mechanical; QSPR, quantitative structure−property relationships; RMSD, root-mean-square deviation

REFERENCES

(1) Wang, R.; Gao, Y.; Lai, L. Calculating partition coefficient by atom-additive method. *Persp. Drug Discov. Design* **2000**, *19*, 47−66.

(2) Czodrowski, P.; Dramburg, I.; Sotriffer, C. A.; Klebe, G. Development, validation, and application of adapted PEOE charges to estimate pKa values of functional groups in protein-ligand complexes. *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 424−437.

(3) Cherkasov, A.; Shi, Z.; Li, Y.; Jones, S. J. M.; Fallahi, M.; Hammond, G. L. 'Inductive' charges on atoms in proteins: comparative docking with the extended steroid benchmark set and discovery of a novel SHBG ligand. *J. Chem. Inf. Model.* **2005**, *45*, 1842−1853.

(4) Cho, A. E.; Guallar, V.; Berne, B. J.; Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **2005**, *26*, 915−931.

(5) Mulliken, R. S. Electronic Structures of Molecules XI. Electro-affinity, Molecular Orbitals and Dipole Moments. *J. Chem. Phys.* **1935**, *3*, 573−585.

(6) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.

(7) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365−375.

(8) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.

(9) Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; Macdougall, P. J.; Wiberg, K. B. Properties of atoms in molecules: Dipole moments and transferability of properties. *J. Chem. Phys.* **1987**, *87*, 1142−1152.

(10) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.

(11) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, 144111.

(12) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(13) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(14) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Accurate partial atomic charges for high-energy molecules using class IV charge models with the MIDI! basis set. *Theor. Chem. Acc.* **2005**, *113*, 133−151.

(15) Abraham, R. J.; Griffiths, L.; Loftus, P. Approaches to charge calculations in molecular mechanics. *J. Comput. Chem.* **1982**, *3*, 407−416.

(16) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(17) Oliferenko, A. A.; Pisarev, S. A.; Palyulin, V. A.; Zefirov, N. S. Atomic Charges via Electronegativity Equalization: Generalizations and Perspectives. *Adv. Quantum Chem.* **2006**, *51*, 139−156.

(18) Shulga, D. A.; Oliferenko, A. A.; Pisarev, S. A.; Palyulin, V. A.; Zefirov, N. S. Parameterization of empirical schemes of partial atomic charge calculation for reproducing the molecular electrostatic potential. *Dokl. Chem.* **2008**, *419*, 57−61.

(19) Cho, K.-H.; Kang, Y. K.; No, K. T.; Scheraga, H. A. A Fast Method for Calculating Geometry-Dependent Net Atomic Charges for Polypeptides. *J. Phys. Chem. B* **2001**, *105*, 3624−3634.

(20) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **1986**, *108*, 4315−4320.

(21) Rappé, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.

(22) Nistor, R. A.; Polihronov, J. G.; Muser, M. H.; Mosey, N. J. A generalization of the charge equilibration method for nonmetallic materials. *J. Chem. Phys.* **2006**, *125*, 094108−094118.

(23) Yang, Z.-Z.; Wang, C.-S. Atom−Bond Electronegativity Equalization Method. 1. Calculation of the Charge Distribution in Large Molecules. *J. Phys. Chem. A* **1997**, *101*, 6315−6321.

(24) Chaves, J.; Barroso, J. M.; Bultinck, P.; Carbó-Dorca, R. Toward an alternative hardness kernel matrix structure in the Electronegativity Equalization Method (EEM). *J. Chem. Inf. Model.* **2006**, *46*, 1657−1665.

(25) Meister, J.; Schwarz, W. H. E. Principal Components of Ionicity. *J. Phys. Chem.* **1994**, *98*, 8245−8252.

(26) Van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L.; Merz, K. M. Are many body effects important in protein folding? *J. Phys. Chem. B* **2000**, *104*, 9554−9563.

(27) Bucher, D.; Raugei, D.; Guidoni, L.; Dal Peraro, M.; Rothlisberger, U.; Carloni, P.; Klein, M. L. Polarization effects and charge transfer in the KcsA potassium channel. *Biophys. Chem.* **2006**, *124*, 292−301.

(28) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.

(29) Anisimov, V. M.; Cavasotto, C. N. In *Challenges and Advances in Computational Chemistry and Physics*; Paneth, P., Dybala-Defratyka, A., Eds.; Springer: Netherlands, 2010; Vol. *12*, Chapter 9, pp 247−266.

(30) Cho, A. E.; Rinaldo, D. Extension of QM/MM docking and its applications to metalloproteins. *J. Comput. Chem.* **2009**, *30*, 2609−2616.

(31) Wallin, G.; Nervall, M.; Carlsson, J.; Åqvist, J. Charges for Large Scale Binding Free Energy Calculations with the Linear Interaction Energy Method. *J. Chem. Theory. Comput.* **2009**, *5*, 380−395.

(32) Yang, Z.; Shen, E.; Wang, L. A scheme for calculating atomic charge distribution in large molecules based on density functional theory and electronegativity equalization. *J. Mol. Str.: Theochem* **1994**, *312*, 167−173.

(33) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The electronegativity equalization method I: Parametrization and validation for atomic charge calculations. *J. Phys. Chem. A* **2002**, *106*, 7887−7894.

(34) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Van Alsenoy, C.; Tollenaere, J. P. The Electronegativity Equalization Method II: Applicability of Different Atomic Charge Schemes. *J. Phys. Chem. A* **2002**, *106*, 7895−7901.

(35) Bultinck, P.; Vanholme, R.; Popelier, P. L. A.; De Proft, F.; Geerlings, P. High-Speed Calculation of AIM Charges through the Electronegativity Equalization Method. *J. Phys. Chem. A* **2004**, *108*, 10359−10366.

(36) Berente, I.; Czinki, E.; Naray-Szabo, G. A Combined Electronegativity Equalization and Electrostatic Potential Fit Method for the Determination of Atomic Point Charges. *J. Comput. Chem.* **2007**, *28*, 1936−1942.

(37) Svobodová Vařeková, R.; Jiroušková, Z.; Vaněk, J.; Suchomel, Š; Koča, J. Electronegativity Equalization Method: Parameterization and Validation for Large Sets of Organic, Organohalogene and Organometal Molecule. *Int. J. Mol. Sci.* **2007**, *8*, 572−582.

(38) Jiroušková, Z.; Svobodová Vařeková, R.; Vaněk, J.; Koča, J. Electronegativity Equalization Method: Parameterization and Validation for Organic Molecules using the Merz-Kollman-Singh Charge Distribution Scheme. *J. Comput. Chem.* **2009**, *30*, 1174−1178.

(39) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. The electronegativity equalization method and the split charge equilibration applied to organic systems: Parametrization, validation, and comparison. *J. Chem. Phys.* **2009**, *131*, 044127−19.

(40) Puranen, J. S.; Vainio, M. J.; Johnson, M. S. Accurate Conformation-Dependent Molecular Electrostatic Potentials for High-

Throughput *In Silico* Drug Discovery. *J. Comput. Chem.* **2010**, *31*, 1722−1732.
(41) Ouyang, Y.; Ye, F.; Liang, Y. A modified electronegativity equalization method for fast and accurate calculation of atomic charges in large biological molecules. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6082−6089.
(42) Verstraelen, T.; Bultinck, P.; Van Speybroeck, V.; Ayers, P. W.; Van Neck, D.; Waroquier, M. The Significance of Parameters in Charge Equilibration Models. *J. Chem. Theory Comput.* **2011**, *7*, 1750−1764.
(43) Verstraelen, T.; Sukhomlinov, S. V.; Van Speybroeck, V.; Waroquier, M.; Smirnov, K. S. Computation of Charge Distribution and Electrostatic Potential in Silicates with the Use of Chemical Potential Equalization Models. *J. Phys. Chem. C* **2012**, *116*, 490−504.
(44) Ionescu, C.-M.; Svobodová Vařeková, R.; Prehn, J. H. M.; Huber, H. J.; Koča, J. Charge Profile Analysis Reveals That Activation of Pro-apoptotic Regulators Bax and Bak Relies on Charge Transfer Mediated Allosteric Regulation. *PloS Comput. Biol.* **2012**, *8*, e1002565−11.
(45) Cedillo, A.; Van Neck, D.; Bultinck, P. Self-consistent methods constrained to a fixed number of particles in a given fragment and its relation to the electronegativity equalization method. *Theor. Chem. Acc.* **2012**, *131*, 1227−1233.
(46) Heidler, R.; Janssens, G. O. A.; Mortier, W. J.; Schoonheydt, R. A. Charge Sensitivity Analysis of Intrinsic Basicity of Faujasite-Type Zeolites Using the Electronegativity Equalization Method (EEM). *J. Phys. Chem.* **1996**, *100*, 19728−19734.
(47) Bultinck, P.; Langenaeker, W.; Carbó-Dorca, R.; Tollenaere, J. P. Fast Calculation of Quantum Chemical Molecular Descriptors from the Electronegativity Equalization Method. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 422−428.
(48) Shimizu, K.; Chaimovich, H.; Farah, J. P. S.; Dias, L. G.; Bostick, D. L. Calculation of the Dipole Moment for Polypeptides Using the Generalized Born-Electronegativity Equalization Method: Results in Vacuum and Continuum-Dielectric Solvent. *J. Phys. Chem. B* **2004**, *108*, 4171−4177.
(49) Chen, S.; Yang, Z. Molecular Dynamics Simulations of a β-Hairpin Fragment of Protein G by Means of Atom-Bond Electronegativity Equalization Method Fused into Molecular Mechanics (ABEEM$\sigma\pi$/MM). *Chin. J. Chem.* **2010**, *28*, 2109−2118.
(50) Svobodová Vařeková, R.; Geidl, S.; Ionescu, C.-M.; Skřehota, O.; Bouchal, T.; Sehnal, D.; Abagyan, R.; Koča, J. Predicting pKa values from EEM atomic charges. *J. Cheminf.* **2013**, *5*, 18.
(51) Sanderson, R. T. An Interpretation of Bond Lengths and a Classification of Bonds. *Science* **1951**, *114*, 670−672.
(52) Parr, R. G.; Donnelly, R. A.; Levy, M.; Palke, W. E. Electronegativity- the density functional viewpoint. *J. Chem. Phys.* **1978**, *68*, 3801−3807.
(53) Parr, R. G.; Pearson, R. G. Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512−7516.
(54) Yang, Z.; Shen, E. Molecular electronegativity in density functional theory (I). *Ser. B Sci. China* **1995**, *38*, 521−528.
(55) Svobodová Vařeková, R.; Koča, J. Optimized and Parallelized Implementation of the Electronegativity Equalization Method and the Atom-Bond Electronegativity Equalization Method. *J. Comput. Chem.* **2005**, *27*, 396−405.
(56) Prokop, M.; Adam, J.; Križ, Z.; Wimmerová, M.; Koča, J. TRITON: a graphical tool for ligand-binding protein engineering. *Bioinformatics* **2008**, *24*, 1955−1956.
(57) Timofeev, V. I.; Baidus, A. N.; Kislitsyn, Y. A.; Kuranova, I. P. 2009, DOI: 10.2210/pdb3e7y/pdb. http://www.rcsb.org/pdb/explore.do?structureId=3E7Y (accessed Oct 31, 2012).
(58) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(59) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117−29.
(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision C.01; Gaussian, Inc.: Wallingford, CT, 2009.
(61) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO Version 3.1*; Gaussian, Inc.: Wallingford, CT, 2009.
(62) Verstraelen T. HiPart Program. http://molmod.ugent.be/software (accessed May 23, 2013).
(63) Wilson, M. S.; Ichikawa, S. Comparison between the Geometric and Harmonic Mean Electronegativity Equilibration Techniques. *J. Phys. Chem.* **1989**, *93*, 3087−3089.
(64) Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.
(65) Allred, A. L. Electronegativity values from thermochemical data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215−221.
(66) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(67) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.
(68) Page, R. C.; Pruneda, J. N.; Amick, J.; Klevit, R. E.; Misra, S. Structural insights into the conformation and oligomerization of E2∼ubiquitin conjugates. *Biochemistry* **2012**, *51*, 4175−4187.
(69) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(70) Feng, Z.; Chen, L.; Maddula, H.; Akcan, O.; Oughtred, R.; Berman, H. M.; Westbrook, J. Ligand Depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics* **2004**, *13*, 2153−2155.
(71) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J. Comput. Chem.* **2009**, *16*, 2785−91.
(72) Gross, K. C.; Seybold, P. G.; Hadad, C. M. Comparison of different atomic charge schemes for predicting pKa variations in substituted anilines and phenols. *Int. J. Quantum Chem.* **2002**, *90*, 445−458.
(73) Svobodová Vařeková, R.; Geidl, S.; Ionescu, C.-M.; Skřehota, O.; Kudera, M.; Sehnal, D.; Bouchal, T.; Abagyan, R.; Huber, H. J.; Koča, J. Predicting pK_a values of substituted phenols from atomic charges. *J. Chem. Inf. Model.* **2011**, *51*, 1795−1806.
(74) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007**, *444*, 205−208.