

# Out-of-the-box deep learning prediction of quantum-mechanical partial charges by graph representation and transfer learning

Dejun Jiang<sup>†</sup>, Huiyong Sun<sup>†</sup>, Jike Wang<sup>†</sup>, Chang-Yu Hsieh, Yuquan Li, Zhenxing Wu, Dongsheng Cao<sup>†</sup>, Jian Wu and Tingjun Hou<sup>†</sup>

Corresponding authors: Tingjun Hou, Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, P. R. China. Tel: 0571-88208412; E-mail: [tingjunhou@zju.edu.cn](mailto:tingjunhou@zju.edu.cn); Jian Wu, College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, Zhejiang, P. R. China. Tel: 0571-86713379; E-mail: [wujian2000@zju.edu.cn](mailto:wujian2000@zju.edu.cn); Dongsheng Cao, Xiangya School of Pharmaceutical Sciences, Central South University, Changsha 410013, Hunan, P. R. China. Tel: 0731-89824761; E-mail: [oriental-cds@163.com](mailto:oriental-cds@163.com)

<sup>†</sup>Equivalent authors.

## Abstract

Accurate prediction of atomic partial charges with high-level quantum mechanics (QM) methods suffers from high computational cost. Numerous feature-engineered machine learning (ML)-based predictors with favorable computability and reliability have been developed as alternatives. However, extensive expertise effort was needed for feature engineering of atom chemical environment, which may consequently introduce domain bias. In this study, SuperAtomicCharge, a data-driven deep graph learning framework, was proposed to predict three important types of partial charges (i.e. RESP, DDEC4 and DDEC78) derived from high-level QM calculations based on the structures of molecules. SuperAtomicCharge was designed to simultaneously exploit the 2D and 3D structural information of molecules, which was proved to be an effective way to improve the prediction accuracy of the model. Moreover, a simple transfer learning strategy and a multitask learning strategy based on self-supervised descriptors were also employed to further improve the prediction accuracy of the proposed model. Compared with the latest baselines, including one GNN-based predictor and two ML-based predictors, SuperAtomicCharge showed better performance on all the three external test sets and had better usability and portability. Furthermore, the QM partial charges of new molecules predicted by SuperAtomicCharge can be efficiently used in drug design applications such as structure-based virtual screening, where the predicted RESP and DDEC4 charges of new molecules showed more robust scoring and screening power than the commonly used partial charges. Finally, two tools including an online server (<http://cadd.zju.edu.cn/deepchargepredictor>) and the source code command lines (<https://github.com/zjujdz/SuperAtomicCharge>) were developed for the easy access of the SuperAtomicCharge services.

**Dongsheng Cao** received his PhD degree in 2013 from Central South University, China. He is currently a professor in the Xiangya School of Pharmaceutical Sciences, Central South University, China. More information can be found at the website of his group: <http://www.scbdd.com>.

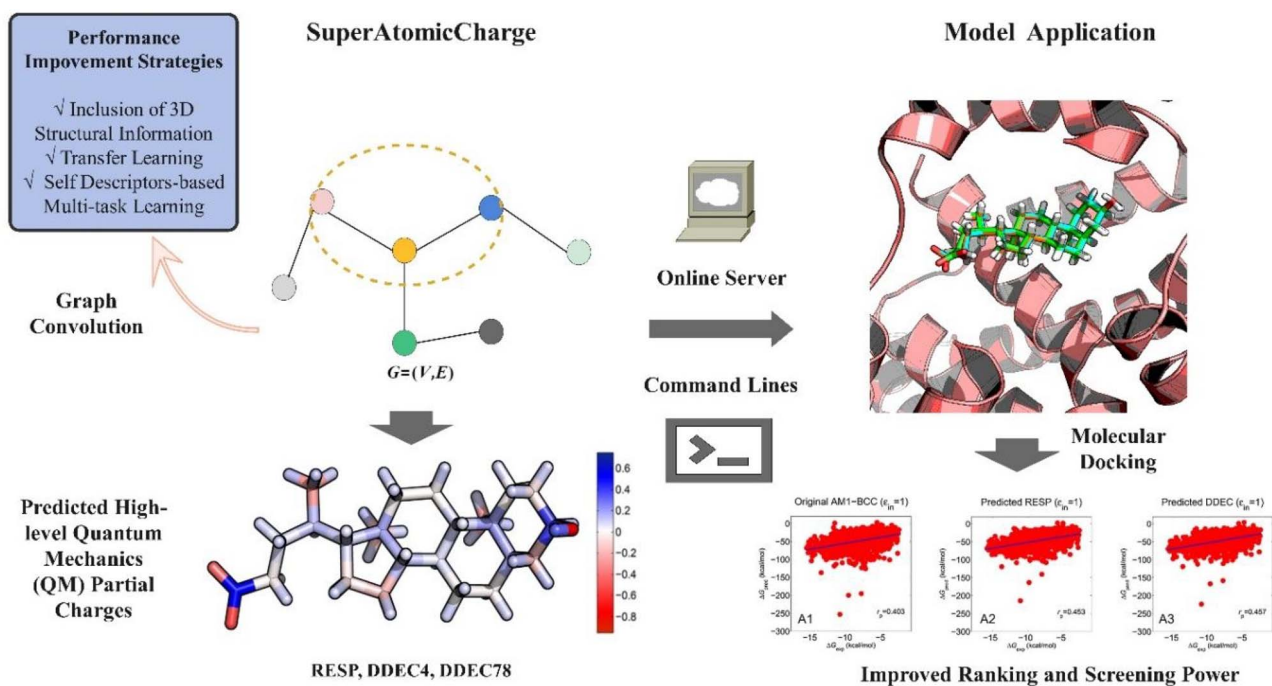
**Jian Wu** is the vice president of the National Academy of Health and Medical Big Data of ZJU, the director of RealDoctor AI Research Center of ZJU. More information can be found at the website of his group: <http://realdoctor.zju.edu.cn>.

**Tingjun Hou** received his PhD degree in 2002 from Peking University, China. He is currently a professor at the College of Pharmaceutical Sciences, Zhejiang University, China. More information can be found at the website of his group: <http://cadd.zju.edu.cn>.

**Received:** November 23, 2021. **Revised:** December 14, 2021. **Accepted:** December 23, 2021

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Graphical Abstract



**Keywords:** atomic partial charges, graph representation learning, transfer learning, multitask learning, drug design, structure-based virtual screening

## Introduction

Accurate prediction of atomic partial charges is an essential step for biochemical and biophysical computations, such as in conformational searches, molecular docking and molecular dynamics (MD) simulations [1]. Atomic partial charges can describe the electrostatic potentials of molecules and intermolecular electrostatic interactions [2]. Normally, high-quality quantum mechanics (QM) calculations provide the most reliable way to estimate the partial charges of small molecules, such as the restrained electrostatic potential (RESP) [3] algorithm by fitting the electrostatic potentials computed by high-quality QM calculations [4], but they are too time-consuming to process a large number of molecules for virtual screening [5]. With the aim of minimizing computational cost, some compromised methods represented by non-QM empirical methods such as Gasteiger charge [6] or semi-empirical QM methods such as AM1-bond charge correction (BCC) charge [7] were developed [8]. However, such compromised methods with high computability may correspondingly suffer from discounted prediction accuracy. Therefore, it is desirable to develop methods that can well handle the tradeoff between accuracy and computability in the estimation of atomic partial charges.

To this end, a myriad of machine learning (ML)-based charge predictors has been developed for fast and accurate predictions of partial charges [1, 5, 8–14]. In

a simplistic view, they can be divided into two major categories: traditional ML-based predictors and deep learning-based predictors. In the first category, each atom was described by a fixed-length vector and an arbitrary subsequent learner can be employed to extract charge assignment rules from the training vectors. The contents of the vector can vary from one to another. In 2013, Rai *et al.* adopted symmetry functions consisting of 77 radial components and 49 angular components to describe the surrounding chemical and geometric environments of atoms [12], and random forest (RF) was employed to construct the partial charges assignment model. The prediction on the test set with 5000 molecules reached a mean unsigned error of 0.03 e. Similarly, in 2018, Bleiziffer *et al.* employed RF to predict the density derived electrostatic and chemical (DDEC) charges of different elements including C, H, N, O, S, P, F, Cl, Br and I [5], but differently, only 2D-based topological descriptors were used in modeling. Finally, the RF model achieved R-square ( $R^2$ ) of 0.983 and 0.997 for the two external test sets, respectively. Based on the atom type symmetry function (ATSF), Wang *et al.* also developed an RF model to predict the conformational adaptive (CA) charges and dipole moments of furanoses [14]. The results indicated that the CA charges predicted by the RF model based on ATSF are superior to the ensemble-averaged atomic charge sets commonly employed in molecular mechanics force fields. For many years, traditional

ML-based partial charges predictors were caught in the feature engineering paradigm where accurate representations of atoms are the most important component of predictors. To better represent the geometric and chemical environments of atoms, Wang *et al.* developed the atom-path-descriptor (APD) descriptors that can represent the 2D and 3D structural information of the atoms in a molecule simultaneously [9], and the application of the APD descriptors using two popular algorithms including RF and eXtreme Gradient Boosting demonstrated that it can give better predictions for all types of atoms than those based on traditional molecular fingerprints.

There is no denying that traditional ML-based partial charge predictors have achieved nontrivial performance improvement in terms of both accuracy and efficacy. However, some points for this kind of methods are naturally inferior. First, the predefined atomic representations or descriptors may introduce bias of domain expertise and cannot consider all possibilities. Second, each atom was treated as an individual training sample and such treatment could cause tremendous inefficiency in computation where tens of thousands of molecules could entail millions of atoms. Third, different elements generally share different predictors even under the same algorithm and that is not a practical way in real-life applications. Therefore, there is a great demand for the development of improved computational models with affordable computability and high accuracy.

Molecular graphs consisting of a series of vertexes (atoms) and edges (chemical bonds) are the natural and effective representations for chemical molecules [15], and the graph neural network (GNN) can be employed to extract any level of representations including node-level, edge-level and graph-level for molecular graphs [16]. For atomic partial charges predictions, GNN can be one of the best ML architectures based on the following facts: first, the basic principles of GNN fall into neighboring information aggregation for central atoms through graph convolution, and this process can be achieved by automatically learning the surroundings for central atoms. In this manner, an atomic partial charges prediction task can be naturally converted into a node-level task on molecular graphs. Second, GNN is an end-to-end architecture where it is capable of automatically extracting task-specific features without any predefined rule or domain expertise. Third, the graph representations of molecules can well evade the individual atom treatment style in traditional ML-based predictors and acquire unrivalled computational efficiency. To this end, Wang *et al.* recently developed the first GNN-based model named DeepAtomicCharge for predicting the DDEC charges of small molecules with different dielectric constants [8]. Although DeepAtomicCharge exhibits better accuracy and computability in comparison with traditional ML-based models, it may still need further improvement. First, the construction of DeepAtomicCharge mainly considered the basic

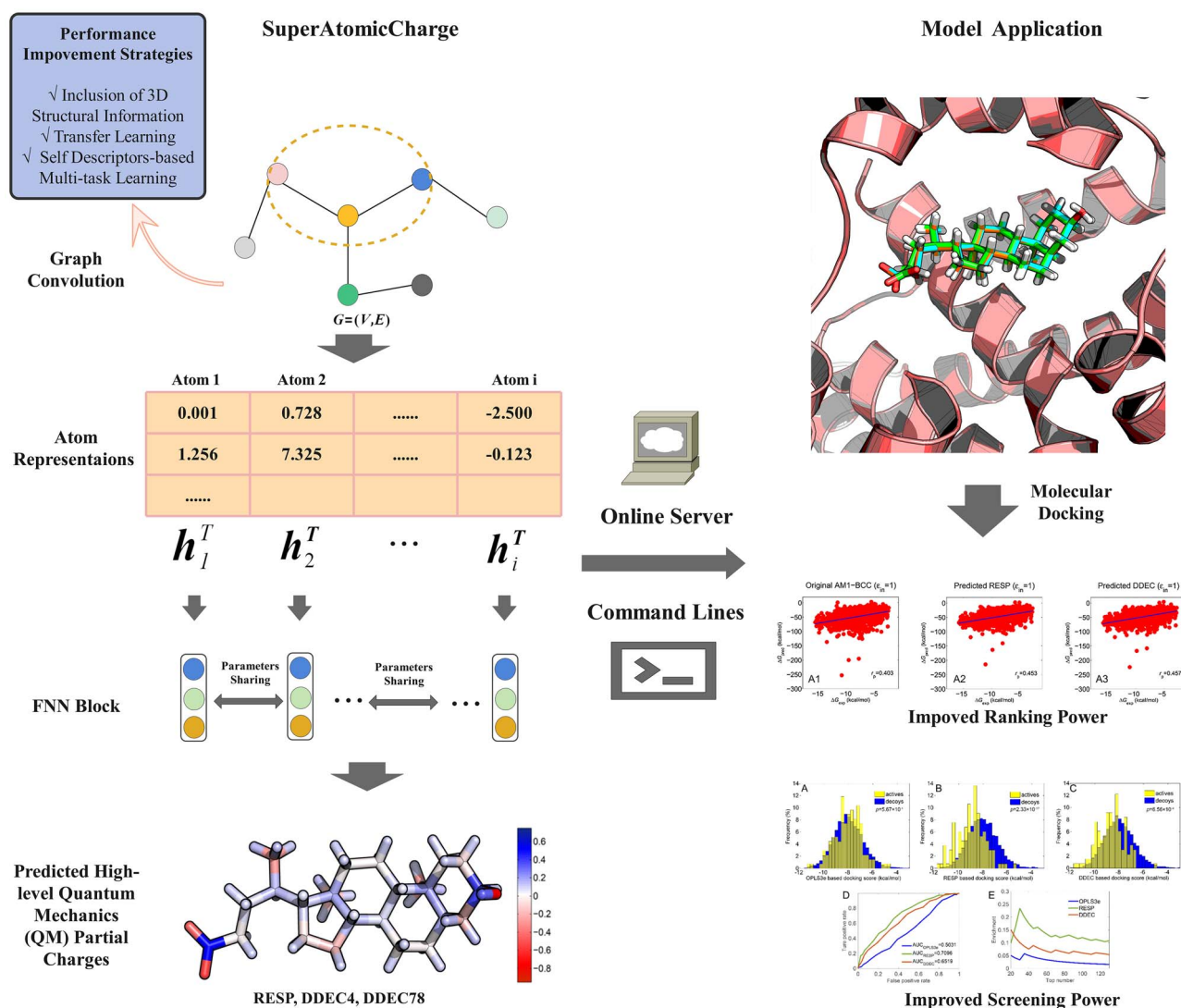
chemical information. However, atomic partial charges are conformationally dependent [2] and the properties of atoms can be critically governed by the spatial distances and directions between atoms in 3D space [17]. Second, only fixed-length molecules (with the same number of atoms) can be handled and the maximum number of atoms was limited to 65 in DeepAtomicCharge. When facing with different length of molecules, the extra fake atom padding technique was needed, which is not an efficient way in computation.

In this study, we proposed a novel and efficient deep graph learning framework named SuperAtomicCharge to predict the atomic partial charges directly from the 3D structures of molecules (Figure 1). In SuperAtomicCharge, both the 2D chemical information and 3D structural information of molecules were covered in model development, which was demonstrated as an effective way in improving the prediction accuracy of model. Then, a simple transfer learning strategy and the multitask learning strategy based on self-supervised descriptors were used to further improve the model's performance. The testing on the three kinds of partial charges including the DDEC4 ( $\epsilon = 4$ ), DDEC78 ( $\epsilon = 78$ ) and RESP charges demonstrated that SuperAtomicCharge outperformed the latest GNN-based predictor and ML-based predictors. Moreover, the predicted QM partial charges of new molecules by SuperAtomicCharge gained superior scoring power and screening power than the commonly used partial charges in the applications of structure-based virtual screening. Finally, two different tools including the source code command lines (<https://github.com/zjujdj/SuperAtomicCharge>) and the online server (<http://cadd.zju.edu.cn/deepchargepredictor>) were developed for the easy use of the SuperAtomicCharge services.

## Materials and methods

### Collection of datasets

A total of ~130 000 organic small molecules with three kinds of partial charges including DDEC4 ( $\epsilon = 4$ ), DDEC78 ( $\epsilon = 78$ ) and RESP were used to construct the SuperAtomicCharge models. Those molecules were collected from the study of Bleiziffer *et al.* [5] The first two kinds of partial charges [i.e., DDEC4 ( $\epsilon = 4$ ) and DDEC78 ( $\epsilon = 78$ )] calculated at high-level QM computation reported by Bleiziffer *et al.* were directly used in this study. Here, the dielectric constant  $\epsilon = 4$  was employed to model partial charges in protein environment and  $\epsilon = 78$  was aimed to model partial charges in solvent environment. Because no RESP charges were available for those organic small molecules, all the molecules in the dataset were optimized using the semi-empirical method (AM1) first [7], and then the RESP fitting algorithm based on the electrostatic potentials estimated at the Hartree Fock self-consistent field (SCF)/6 G-31G\* level [18] in Gaussian16 program was utilized to calculate the corresponding RESP charges for each molecule.



**Figure 1.** The detailed descriptions of the SuperAtomicCharge architecture and corresponding applications.

## SuperAtomicCharge formalism

In this section, we will elaborate on three key components in the SuperAtomicCharge model including input representation, node representation learning module and feed-forward neural network (FNN).

**Input Representation with Chemical and 3D Structural Information:** In this section, each small molecule was represented as a bidirected graph ( $G = (V, E)$ ) with the explicit existence of hydrogen atoms, and  $(i, j) \in E$  indicates that there exists a bond between atoms  $i$  and  $j$  [in  $(i, j)$ , atom  $i$  is the source atom and atom  $j$  is the destination atom, and vice versa]. Due to the bidirectional essence of the graph in this study,  $(j, i) \in E$  is another opposite edge between atoms  $i$  and  $j$ . All the small molecules were processed into the corresponding bidirected graphs with node features and edge features (Table 1). As shown in Table 1, different from the complicated atom descriptors reported in previous studies [5, 9, 12], only several basic atom features were considered in the graph representation. In terms of edge features,

the basic bond features were also considered. As for the spatial distances and directions between atoms in 3D space may govern the properties of atoms, another set of 3D-related edge features were considered. Those edge features including distance, angle and area statistics were employed to describe the geometric environments of atoms [19]. The detailed calculation algorithm of this set of edge features is formulized in Table S1, and another detailed example is shown in Figure 2A.

**Node Representation Learning with Graph Convolution:** This module was designed based on the studies reported by Xiong et al. [20] and Wang et al. [8], and it was intended to automatically learn the node representations of atoms under the message passing framework. The learned node representations were used as the inputs of the subsequent FNN module for partial charge predictions. At each time step  $t$ , the inputs of this module are a series of node states from the last time step  $t - 1$ . Here, the node states in time step 0 were acquired by the following steps that both consider the initial atom



**Table 1.** The detailed descriptions of the node/edge features used in graph representation

Features	Type	Attributes name	Descriptions	Length
Node Features	2D	atom_type_one_hot	One hot encoding for atom type ([‘C’, ‘N’, ‘O’, ‘S’, ‘F’, ‘P’, ‘Cl’, ‘Br’, ‘I’, ‘B’, ‘H’, ‘Si’, ‘Other’])	13
		atom_degree_one_hot	One hot encoding for the degree of an atom (Atom degrees to consider: 0–5)	6
		atom_formal_charge	The formal charge for an atom	1
		atom_num_radical_electrons	The number of radical electrons for an atom	1
		atom_hybridization_one_hot	One hot encoding for the hybridization of an atom (‘SP’, ‘SP2’, ‘SP3’, ‘SP3D’, ‘SP3D2’, ‘Other’)	6
		atom_is_aromatic	Whether the atom is aromatic	1
		atom_total_num_H_one_hot	One hot encoding for the total number of Hs of an atom (Total number of Hs to consider: 0–4)	5
		atom_chirality_one_hot	One hot encoding for chirality of an atom ([‘R’, ‘S’, ‘Other’])	3
Edge features	2D	bond_type_one_hot	One hot encoding for the type of a bond ([‘SINGLE’, ‘DOUBLE’, ‘TRIPLE’, ‘AROMATIC’])	4
		bond_is_conjugated	Whether the bond is conjugated	1
		bond_is_in_ring	Whether the bond is in a ring of any size	1
		bond_stereo_one_hot	One hot encoding for the stereo configuration of a bond ([‘STEREONONE’, ‘STEREOANY’, ‘STEREOZ’, ‘STEREOE’, ‘Other’])	5
Edge features	3D	Distance	the Euclidean distance between to connected atoms in 3D space	1
		Angle statistics	The scaled max, sum, and mean values of angle between atoms i, j, k in 3D space	3
		Area statistics	The max, sum, and mean values of areas between atoms i, j, k in 3D space	3
		Distance statistics	The scaled max, sum, and mean values of distances between atoms i, k in 3D space	3

and edge features simultaneously. For each atom  $i$ , the unnormalized attention score  $s_{ij}$  from certain bonded atom  $j$  was computed based on the initial edge features  $e_{ij}$  and initial atom features  $x_i$  and  $x_j$  according to:

$$s_{ij} = \sigma_1 (w_3 [\sigma_1 (w_1 x_j) \parallel \sigma_1 (w_2 [x_i \parallel e_{ij}])]) \quad (1)$$

where  $w_1, w_2, w_3$  are different learnable matrix, and  $\sigma_1$  is the *LeakyReLU* activation function. To make the attention score easily comparable across the bonded atoms, unnormalized attention score was normalized across all the bonded atoms  $N_{(i)}$  using the *softmax* function to get the normalized attention score  $\alpha_{ij}$ :

$$\alpha_{ij} = \text{softmax}(s_{ij}) = \frac{\exp(s_{ij})}{\sum_{k \in N_{(i)}} \exp(s_{ik})} \quad (2)$$

Then, atom  $i$  aggregates the information across its bonded neighbors weighted by  $\alpha_{ij}$ :

$$m_i = \sigma_2 \left( \sum_{k \in N_{(i)}} \alpha_{ik} w_4 \sigma_1 (w_2 [x_i \parallel e_{ik}]) \right) \quad (3)$$

where the calculation of term  $\sigma_1 (w_2 [x_i \parallel e_{ik}])$  is kept as the same with Equation (1);  $w_4$  is the learnable matrix, and  $\sigma_2$  is the *exponential linear unit* activation function. Finally, the initial node state  $h_i^0$  in time step 0 for atom  $i$  was computed as:

$$h_i^0 = \sigma_3 (\text{GRU} (m_i, \sigma_1 (w_1 x_i))) \quad (4)$$

where  $\sigma_3$  is the *rectified linear unit* activation function, and a gated recurrent unit (GRU) was used to fuse  $m_i$  and the transformed term  $\sigma_1 (w_1 x_i)$  of the initial atom features  $x_i$  into the initial node state for each atom. Afterwards, for each time step  $t$ , the node states were calculated according to previous node states under the message passing framework (Figure 2B), and the message function [Equations (5) and (6)] and vertex update function [Equation (7)] were defined as follows:

For  $t$  in  $\{1, \dots, T-1\}$ :

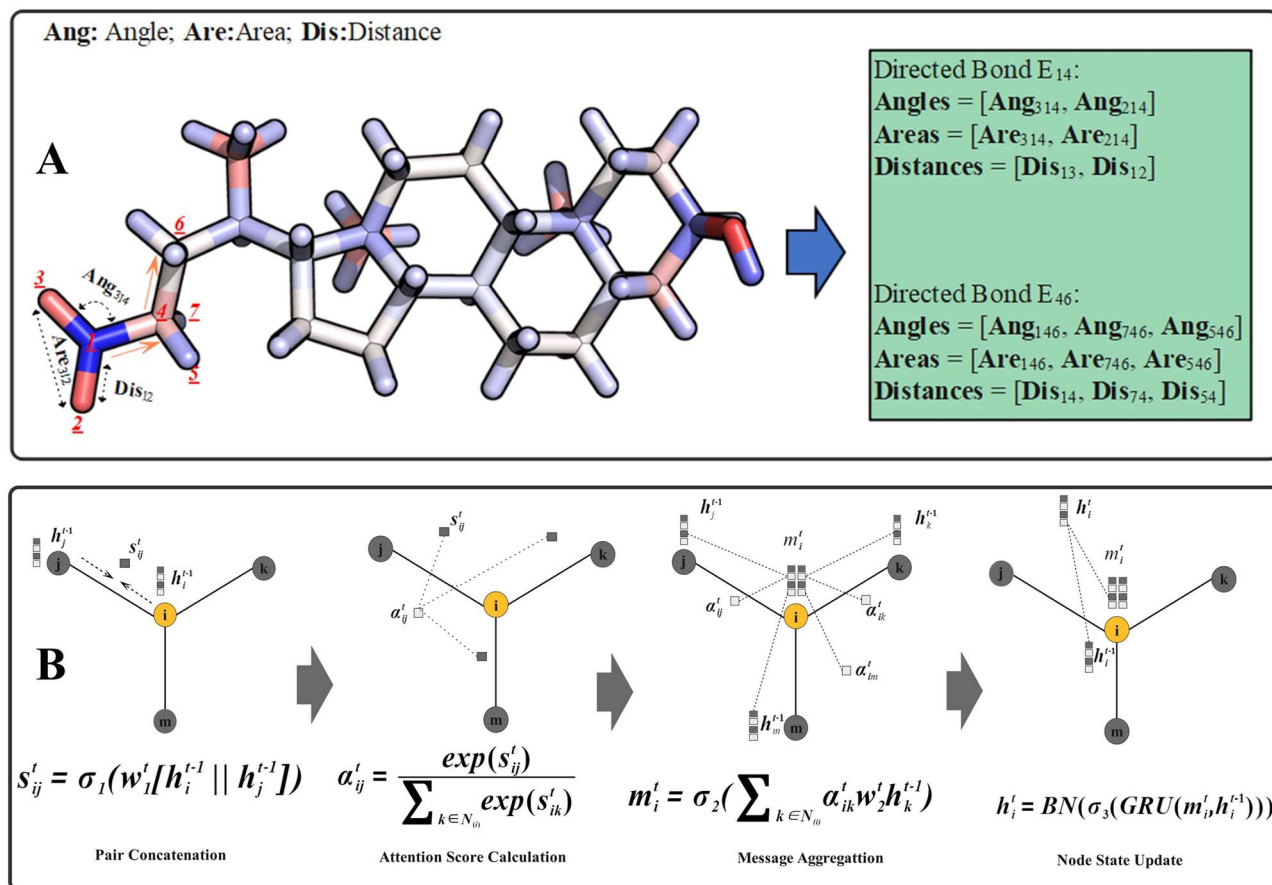
$$s_{ij}^t = \sigma_1 \left( w_1^t [h_i^{t-1} \parallel h_j^{t-1}] \right) \quad (5)$$

$$m_i^t = \sigma_2 \left( \sum_{k \in N_{(i)}} \alpha_{ik}^t w_2^t h_k^{t-1} \right) \quad (6)$$

$$h_i^t = \text{BN} (\sigma_3 (\text{GRU} (m_i^t, h_i^{t-1}))) \quad (7)$$

where the calculation of normalized attention score  $\alpha_{ik}^t$  is similar to Equation (2) based on  $s_{ij}^t$ . BN is a batch normalization operation with the aim of accelerating model training, and  $w_1^t$  and  $w_2^t$  are the learnable matrix in each time step. Finally, we summed the node states in each time step to get the final node representation  $h_i^T$  for each atom:

$$h_i^T = \sum_{t=1}^{T-1} h_i^t \quad (8)$$



**Figure 2.** (A) The visualization example of the calculation of the 3D edge features including distance, angle, and area statistics. (B) The detailed descriptions of message passing including pair concatenation, attention score calculation, message aggregation and node state update.

**Feed-forward Neural Network (FNN):** Normally, traditional ML-based methods would develop different predictors for different elements. For simplicity, all the elements share the same FNN predictor to reduce model's complexity in this study. The inputs of this module are the node representation  $h_i^T$  of each atom. Here a two hidden layer FNN was used and the training objective is to minimize the loss function (mean squared error (MSE) loss) among all the training atoms. Given a dataset with  $N$  atoms, the predicted and true values of a certain atom are  $\hat{y}_i$  and  $y_i$ , respectively. The loss function is defined as:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (9)$$

### Details of model training and evaluation

The SuperAtomicCharge model was implemented with the open-source DGL-CUDA90 (version: 0.4.3.post2) with PyTorch as the backend and RDKit (version: 2019.09.1) python package. For the three datasets, each dataset was divided into the training, validation and test sets at the ratio of 8:1:1. The validation set was used for model section and the test set was used to validate the generalization ability of the model. All the experiments in this study share the same hyperparameters as

shown in Table S2. In addition to the commonly used dropout and L2 regularization trick, early stopping was further considered for avoiding overfitting if no improvement of the validation performance was achieved in successive 50 epochs. For comparison, three latest methods including the GNN-based DeepAtomicCharge [8], atom-path-descriptor-based random forest (APD-RF) [9] and atom-pair fingerprint-based RF model (AP-RF) [5] were used as the baselines. For the sake of fairness, all the compared models were evaluated on the same data folds used in this study. Each experiment was repeated three times with different random seeds and the average performance with standard deviation was reported.

As for the evaluation metric, root mean square error (RMSE) is the main indicator. In a more diverse evaluation, Pearson correlation coefficient ( $R_p$ ), mean absolute error (MAE) and  $R^2$  were also reported.

### Transfer learning strategy

Transfer learning falls into the conception 'learn to learning' and it can be considered as a collection of ML methods that can gain generalizable knowledge from related tasks (source tasks) to enable learning of target tasks with data scarcity property [21–23]. Different from other transfer learning scenarios where the target tasks are usually lack of data volume and correspondingly the

pretraining strategy is commonly used [23–25], here the endpoint properties of our three datasets were similar. Therefore, a simple but efficient ‘model fine tuning’ strategy was used. In a more specific way, for each available model pair (source task model/target task model), we used the parameters from the well-trained source model to initialize the target task model and subsequently fine-tune the target task model, which can be regarded as the process that the target task model was optimized starting from the knowledge extracted from the source task, and therefore it is quite possible for the target task model to find more generalized answers in the tremendous solution space compared with random initialization. In the fine-tuning of the target task model, two kinds of strategies were used: the first is global tuning where all the parameters of the target model share the same learning rate ( $1.0 \times 10^{-03}$ ), and the second is local tuning where the parameters from the graph convolution layers share small learning rate ( $1.0 \times 10^{-05}$ ) but the parameters from the FNN layers share normal learning rate ( $1.0 \times 10^{-03}$ ). In the transfer learning process, the transfer learning model was trained three times. Finally, the average performance with standard deviation was reported.

### Validation of the proposed model in actual drug design issues

To test the performance of the SuperAtomicCharge model in practical drug design applications, we assessed the performance of the RESP and DDEC4 charges predicted by SuperAtomicCharge on two important metrics (scoring power and screening power) in structure-based virtual screening [26]. Here the DDEC78 charges are usually used to characterize the electrostatic effects of a molecule in bulk solvent, and therefore we did not evaluate the performance of this type of atomic charges in drug design applications (usually are protein systems). The scoring power was evaluated by the PDBbind dataset [27] containing ~1300 protein–ligand complexes and the screening power was evaluated using two drug targets i.e. caspase 8 and Farnesoid X Receptor (FXR). The more descriptions about scoring power and screening power evaluations are available in the Supporting Information.

## Results and discussion

### Inclusion of 3D structural information and transfer learning did improve the performance of SuperAtomicCharge

It is a common sense that the spatial distances and directions between atoms in 3D space can determine the nature of molecules [17]. However, many proposed methods for partial charge predictions only rely on 2D topological information [5, 8]. In the light of this, we checked whether the inclusion of 3D structural information can improve the performance of SuperAtomicCharge using a straightforward comparative study (Table 2). All the 3D edge features were removed and then the

model was trained correspondingly with the same procedure. As shown in Table 2, it is obvious that the inclusion of the 3D structural information is beneficial to the prediction accuracy of SuperAtomicCharge. For the DDEC4 charges, the inclusion of the 3D structural information can improve the performance by ~7% points, and the average RMSE value is dropped from 0.01147 e to 0.01061 e. For the DDEC78 charges, ~5% points improvement is gained after the inclusion of the 3D structural information, and the decrease of the average RMSE value is 0.00075 e. Similarly, ~5% points improvement is also gained for the RESP charges, and the corresponding decrease of the average RMSE value is 0.0028 e. In addition, it can be observed that the corresponding RMSE values for the training and validation sets also decrease with the inclusion of the 3D structural information, and overall, the standard deviations are also getting smaller, demonstrating that the SuperAtomicCharge model built on both the 2D and 3D structural information is more robust and generalized than that only built on the 2D structural information. To be more persuasive, the paired t-test was employed to check whether the performance difference between the model built on only the 2D structural information and that built on both the 2D and 3D structural information is significant (significance level=0.05), and totally nine RMSE values from three independent runs for the three types of partial charges were used to perform the statistical test. The corresponding P-value is 0.003, which is much lower than the 0.05, indicating that the performance difference between the model built on only the 2D structural information and that built on both the 2D and 3D structural information is significant. Finally, other metrics including  $R^2$ , MAE and  $R_p$  are also presented in Table S3, and those statistics also support the conclusion that the 3D structural information is essential to the reliable prediction of atomic partial charges.

Further, the effectiveness of the transfer learning strategy was also checked as described in ‘Transfer Learning Strategy’. It should be mentioned that all the models tested here were developed based on both the 2D and 3D structural information. As shown in Table 3, for strategy 1, the performance of two models (DDEC4 and DDEC78) is obviously improved by transferring from the RESP model. For the DDEC4 charge model, the average RMSE value is improved by ~5% points and decreases from 0.01061 e to 0.01008 e, and that for the DDEC78 model is ~8% points and decreases from 0.01312 e to 0.01204 e. In addition, the mutual transfer between the DDEC4 and DDEC78 models can also slightly improve the performance of either model compared with the corresponding original model. Here, the RESP model transferred from the DDEC4 and DDEC78 models is also slightly improved, indicating the effectiveness of this simple transfer learning strategy. As for the transfer strategy 2 where the learning rate of the graph convolution layer is much smaller than strategy 1, it is obvious that the DDEC4 and DDEC78 models transferred from the RESP model benefit much

**Table 2.** The performance comparison (average RMSE values from 3 independent runs) of SuperAtomicCharge built on only the 2D information and both the 2D and 3D information for the three kinds of partial charges

Model	Type of charge	Features	RMSE		
			Training	Validation	Testing
SuperAtomicCharge	DDEC4	2D	0.01129 ± 2.310 × 10 <sup>-04</sup>	0.01139 ± 2.311 × 10 <sup>-04</sup>	0.01147 ± 2.300 × 10 <sup>-04</sup>
	DDEC4	2D + 3D	0.01043 ± 1.198 × 10 <sup>-04</sup>	0.01053 ± 1.198 × 10 <sup>-04</sup>	<b>0.01061 ± 1.070 × 10<sup>-04</sup></b> (+7.50%)
	DDEC78	2D	0.01390 ± 9.516 × 10 <sup>-04</sup>	0.01404 ± 9.099 × 10 <sup>-04</sup>	0.01387 ± 9.482 × 10 <sup>-04</sup>
	DDEC78	2D + 3D	0.01292 ± 7.282 × 10 <sup>-04</sup>	0.01328 ± 6.794 × 10 <sup>-04</sup>	<b>0.01312 ± 6.632 × 10<sup>-04</sup></b> (+5.41%)
	RESP	2D	0.04903 ± 7.175 × 10 <sup>-04</sup>	0.05321 ± 2.116 × 10 <sup>-04</sup>	0.05283 ± 1.486 × 10 <sup>-04</sup>
	RESP	2D + 3D	0.04602 ± 5.409 × 10 <sup>-04</sup>	0.05026 ± 3.211 × 10 <sup>-04</sup>	<b>0.04999 ± 3.164 × 10<sup>-04</sup></b> (+5.38%)

Statistic with the performance improvement or decrease is larger than 3% points was represented as bold.

**Table 3.** The performance comparison (average RMSE values from 3 independent runs) between transferred SuperAtomicCharge and original SuperAtomicCharge on different atomic partial charges. Statistic with the performance improvement or decrease is larger than 3% points was represented as bold<sup>a</sup>

Strategy 1 (global fine-tuning, lr = 1.0 × 10 <sup>-03</sup> )			
Type of charges	DDEC4	DDEC78	RESP
DDEC4	0.01061 ± 1.070 × 10 <sup>-04</sup>	<b>0.01267 ± 2.826 × 10<sup>-04</sup></b> (+3.43%)	0.04958 ± 3.402 × 10 <sup>-04</sup>
DDEC78	0.01059 ± 4.003 × 10 <sup>-04</sup>	0.01312 ± 6.632 × 10 <sup>-04</sup>	0.04932 ± 3.750 × 10 <sup>-05</sup>
RESP	<b>0.01008 ± 1.919 × 10<sup>-04</sup></b> (+4.90%)	<b>0.01204 ± 2.356 × 10<sup>-04</sup></b> (+8.23%)	0.04999 ± 3.164 × 10 <sup>-04</sup>
Strategy 2 (local fine-tuning, GNN lr = 1.0 × 10 <sup>-05</sup> , FNN lr = 1.0 × 10 <sup>-03</sup> )			
Type of charges	DDEC4	DDEC78	RESP
DDEC4	0.01061 ± 1.070 × 10 <sup>-04</sup>	<b>0.01224 ± 3.376 × 10<sup>-04</sup></b> (+6.71%)	<b>0.05249 ± 7.291 × 10<sup>-04</sup></b> (-5.00%)
DDEC78	<b>0.01022 ± 7.927 × 10<sup>-04</sup></b> (+3.68%)	0.01312 ± 6.632 × 10 <sup>-04</sup>	<b>0.05287 ± 8.408 × 10<sup>-04</sup></b> (-5.76%)
RESP	<b>0.00942 ± 7.505 × 10<sup>-05</sup></b> (+11.21%)	<b>0.01179 ± 1.365 × 10<sup>-04</sup></b> (+10.14%)	0.04999 ± 3.164 × 10 <sup>-04</sup>

<sup>a</sup>The diagonal number in each table represents the performance of the original SuperAtomicCharge model. Vertical axis is the source model and horizontal axis is the target model. Taking 0.01059 ± 4.003 × 10<sup>-04</sup> as an example, it represents the performance of the target SuperAtomicCharge (DDEC4) model transferred from the source model [SuperAtomicCharge (DDEC78)].

more compared with strategy 1. For the DDEC4 model, the average RMSE value is improved by ~11% points and decreases from 0.01061 e to 0.00942 e, ~0.0012 e precision improvement gained using this transfer strategy (~5% points for strategy 1). As for the DDEC78 model, the average RMSE value is also improved by ~10% points and decreases from 0.01312 e to 0.01179 e (~8% points for strategy 1). In parallel, the mutual transfer between the DDEC4 and DDEC78 models using strategy 2 yields slightly better performance than strategy 1, as shown in Table 3.

Out of the exception, both the RESP models transferred from the DDEC4 and DDEC78 models using strategy 2 are getting much worse than the corresponding original models (0.05249 e versus 0.04999 e and 0.05287 e versus 0.04999 e). Our investigation revealed that this phenomenon is potentially caused by the very small learning rate used in the graph convolution layers (1.0 × 10<sup>-5</sup>). For illustration of this point, we checked the training logs of the RESP models transferred from the two DDEC models for both strategy 1 (graph convolution layers learning rate is 1.0 × 10<sup>-3</sup>) and strategy 2 (graph convolution layers learning rate is 1.0 × 10<sup>-5</sup>). We observed that an average training step of 250 epochs is enough for the RESP models transferred from the two DDEC

models under strategy 1 to be converged to better results (compared with its original untransferred models), but the average training step for the RESP models transferred from the two DDEC models under strategy 2 is about 1500 epochs, implying that the RESP models under strategy 2 were updated slowly and the models were potentially trapped into local minima. Therefore, it can be concluded that the RESP models with a very small learning rate in graph convolution layers may not be converged to better results. Similarly, to be more persuasive, the paired t-test was employed to check whether the performance difference between the transferred models and original untransferred models is significant, and a total of 18 RSME values (6 transferred models from three different runs versus 6 originally untransferred models from three different runs) were employed as the samples. For the transfer strategy 1, the P-value is 0.0008, indicating the effectiveness of this transfer strategy. However, the P-value for the transfer strategy 2 is 0.55, which was obviously contributed from the dramatical performance decrease of the RESP models transferred from the two DDEC models. After removal of those negative results, the P-value for this transfer strategy can reach a better value of 0.0006. All in all, the inclusion of the 3D structural information in the graph representation of molecules



is quite critical for accurate and reliable predictions of atomic partial charges, and overall, the simple transfer learning technique can also improve the prediction accuracy of models.

### Multitask learning using self-supervised descriptors

It has been reported many times that multitask learning can be more superior against single-task learning in chemical property predictions [15, 28, 29]. Therefore, it is of interest to explore whether the prediction of partial charges can be beneficial from multitask learning. Here the multitask learning model for each type of partial charges was trained based on the atom-centered symmetry functions (ACSFs) [30–32] and then compared with the corresponding single-task counterpart. Different from previous multitask learning paradigm where each target label was generally acquired based on the bioassays [33, 34], we took this convenient way where no expensive label acquisition was needed to explore whether multitask learning by self-supervised descriptors can improve the performance compared with the single-task learning. The detailed descriptions and calculation of ACSF descriptors are available in the Supporting Information.

As shown in Table 4, it is obvious that the multitask learning supervised by the ACSF descriptors is superior to the single-task learning. For all types of partial charges, the RMSE values for the training, validation and testing sets given by the multitask learning are smaller than those given by the single-task learning. As to the DDEC4 and DDEC78 models, the average RMSE improvements for the test set in comparison with those of the corresponding single-task learning are 0.0006 e and 0.0002 e, respectively. With regard to the RESP model, the average RMSE improvement is much better (0.0019 e) than those for the two DDEC models. According to the performance statistics, it can be concluded that the predictions of the RESP charges are much harder than those of the two DDEC charges. Satisfactorily, the multitask learning RESP model achieves the best performance with an average RMSE value of 0.04805 e for the test set, which is much better than the aforementioned best average RMSE value of 0.04932 e (Table 3, given by the RESP model transferred from the DDEC78 model). However, it seems that the performance improvement given by the multitask learning based on the ACSF descriptors is inferior to the simple transfer learning shown above for the two DDEC models, suggesting that different training strategies may have a different degree of benefits to different models. Similarly, the *P*-value of 0.015 given by the paired *t*-test demonstrated that the performance difference between the multitask learning model and single-task learning model is statistically significant (significant level = 0.05). Finally, other three metrics including  $R^2$ , MAE and  $R_p$  metrics given by the single-tasking learning and multitask learning for the three types of partial charges are also presented in Table S4, and the analysis of those

metrics also supports the conclusion that the multitask learning model based on the ACSF descriptors can give more reliable predictions than the single-task learning model.

### Performance comparison with other state-of-the-arts

To better check the superiority of SuperAtomicCharge, the SuperAtomicCharge model with the minimum validation error for each type of partial charges was compared with the three latest baselines, including the AP-RF model proposed by Bleiziffer *et al.* [5], the APD-RF model and the graph convolution-based DeepAtomicCharge model recently proposed by Wang *et al.* [8, 9] For the sake of fairness, all the baselines shared the same data splitting used in this study and their hyperparameter configurations directly followed the original reports. For the two classical ML models including APD-RF and AP-RF, individual models were trained for different elements to follow the original operations. Each experiment was repeated three times and the average metric with standard deviation was reported.

As can be seen from Table 5, the AP-RF model shows extremely unsatisfactory predictions to all the three types of atomic partial charges compared with the other models. It only gives the average RMSE values of 0.01659 e, 0.01886 e and 0.06353 e to the DDEC4, DDEC78 and RESP partial charges, respectively. The AP-RF model was developed only based on the 2D atom-pair fingerprints, and potentially more complicated descriptors comprehensively covering the 2D and 3D profiles of molecules are needed to improve the performance of the AP-RF model. Overall speaking, the recently proposed graph convolution-based DeepAtomicCharge model ranks second last among the compared models, and it showed the average RMSE values of 0.01148 e, 0.01404 e and 0.05469 e to the three types of atomic partial charges, respectively. Careful investigation of the source codes of DeepAtomicCharge indicates that only one 3D-related feature (bond length) was utilized by the model, which may also limit the comprehensive representations of molecules. Among the three baselines, the ADP-RF model is the top-performing estimator, and it gives satisfactory average RMSE values of 0.01043 e, 0.01208 e and 0.05472 e to the three types of atomic charges, respectively. The investigation of APD illustrated that it is a comprehensive descriptor to describe the local 2D and 3D structural information for each atom using iterative algorithms. As expected, the ADP-RF model performs the best among the compared baselines.

As shown in Table 5, our SuperAtomicCharge model gained ~43%, ~37% and 24% average RMSE improvements against the commonly used AP-RF model baseline [8, 9] for the three atomic partial charges, respectively, indicating the great superiority of our SuperAtomicCharge model. Stepping back, our SuperAtomicCharge model is also superior to the latest top-performing ADP-RF baseline. SuperAtomicCharge yielded the average

**Table 4.** The performance comparison (average RMSE values from 3 independent runs) between single-task learning and multitask learning of the SuperAtomicCharge model

Type of charges	Model	RMSE		
		Training	Validation	Testing
DDEC4	Single-task	$0.01043 \pm 1.198 \times 10^{-04}$	$0.01053 \pm 1.198 \times 10^{-04}$	$0.01061 \pm 1.070 \times 10^{-04}$
	Multitask	$0.00971 \pm 2.619 \times 10^{-04}$	$0.00991 \pm 2.587 \times 10^{-04}$	<b><math>0.01000 \pm 2.521 \times 10^{-04}</math></b> (+5.75%)
DDEC78	Single-task	$0.01292 \pm 7.282 \times 10^{-04}$	$0.01328 \pm 6.794 \times 10^{-04}$	$0.01312 \pm 6.632 \times 10^{-04}$
	Multitask	$0.01263 \pm 2.726 \times 10^{-04}$	$0.01303 \pm 2.494 \times 10^{-04}$	$0.01293 \pm 2.209 \times 10^{-04}$
RESP	Single-task	$0.04602 \pm 5.410 \times 10^{-04}$	$0.05026 \pm 3.211 \times 10^{-04}$	$0.04999 \pm 3.164 \times 10^{-04}$
	Multitask	$0.04287 \pm 3.315 \times 10^{-04}$	$0.04829 \pm 1.320 \times 10^{-04}$	<b><math>0.04805 \pm 2.852 \times 10^{-05}</math></b> (+3.90%)

Statistic with the performance improvement or decrease is larger than 3% points was represented as bold.

**Table 5.** The performance comparison (average RMSE values from 3 independent runs) between the optimal SuperAtomicCharge determined by the minimum validation error criterion in this study and other three baselines including AP-RF, ADP-RF and DeepAtomicCharge on the three types of partial charges

Type of charges	Model	RMSE		
		Training	Validation	Testing
DDEC4	AP-RF	$0.01433 \pm 4.131 \times 10^{-06}$	$0.01671 \pm 6.436 \times 10^{-06}$	$0.01659 \pm 5.868 \times 10^{-06}$
	ADP-RF	$0.00688 \pm 1.545 \times 10^{-06}$	$0.01038 \pm 5.941 \times 10^{-06}$	$0.01043 \pm 3.744 \times 10^{-06}$
	DeepAtomicCharge	$0.01127 \pm 3.972 \times 10^{-04}$	$0.01142 \pm 3.486 \times 10^{-04}$	$0.01148 \pm 3.411 \times 10^{-04}$
	SuperAtomicCharge	$0.00912 \pm 7.154 \times 10^{-05}$	$0.00935 \pm 6.528 \times 10^{-04}$	<b><math>0.00942 \pm 7.505 \times 10^{-05}</math></b> (+43.22%)
DDEC78	AP-RF	$0.01640 \pm 2.773 \times 10^{-06}$	$0.01914 \pm 5.999 \times 10^{-06}$	$0.01886 \pm 1.620 \times 10^{-05}$
	ADP-RF	$0.00819 \pm 5.486 \times 10^{-06}$	$0.01223 \pm 2.016 \times 10^{-05}$	$0.01208 \pm 1.308 \times 10^{-05}$
	DeepAtomicCharge	$0.01382 \pm 12.84 \times 10^{-04}$	$0.01413 \pm 11.72 \times 10^{-04}$	$0.01404 \pm 11.77 \times 10^{-04}$
	SuperAtomicCharge	$0.01134 \pm 1.466 \times 10^{-04}$	$0.01197 \pm 1.400 \times 10^{-04}$	<b><math>0.01179 \pm 1.365 \times 10^{-04}</math></b> (+37.49%)
RESP	AP-RF	$0.05541 \pm 1.954 \times 10^{-06}$	$0.06349 \pm 1.079 \times 10^{-05}$	$0.06353 \pm 5.074 \times 10^{-06}$
	ADP-RF	$0.03500 \pm 6.645 \times 10^{-06}$	$0.05509 \pm 2.178 \times 10^{-05}$	$0.05472 \pm 7.348 \times 10^{-06}$
	DeepAtomicCharge	$0.05041 \pm 5.854 \times 10^{-04}$	$0.05481 \pm 1.686 \times 10^{-04}$	$0.05469 \pm 2.148 \times 10^{-04}$
	SuperAtomicCharge	$0.04286 \pm 3.315 \times 10^{-04}$	$0.04829 \pm 1.320 \times 10^{-05}$	<b><math>0.04805 \pm 2.852 \times 10^{-05}</math></b> (+24.37%)

The percentage in bracket refers to the performance improvement over the commonly used AP-RF baseline.

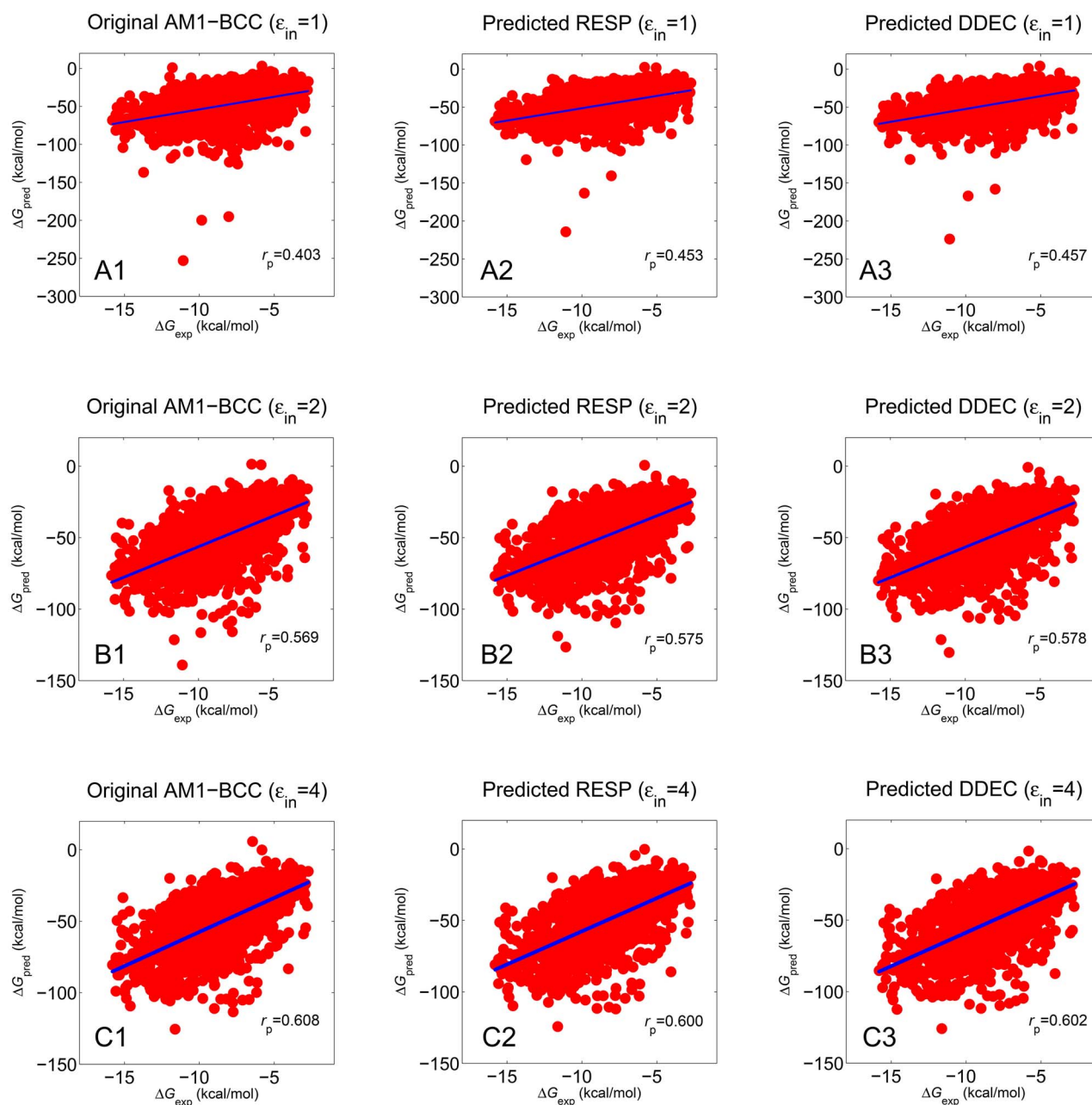
RMSE values of 0.00942, 0.01179 and 0.04805 e to the three types of partial charges respectively, which is approximately 9.68%, 2.40% and 12.19% lower than the top-performing ADP-RF baseline. As can be observed from Table 5, the APD-RF model generally shows much better performances to the training set than the SuperAtomicCharge model for all the three types of partial charges. However, the SuperAtomicCharge model always give the lowest RMSE values to the test sets of all the three types of partial charges, indicating that the APD-RF model is more likely to be overfitted and less generalized in comparison with the SuperAtomicCharge model. Based on the statistics of the paired t-test, the P-values of 0.001, 0.030 and  $9.645 \times 10^{-5}$  for the SuperAtomicCharge/DeepAtmoicCharge, SuperAtomicCharge/ADP-RF and SuperAtomicCharge/AP-RF model pairs indicated that the performance difference between the SuperAtomicCharge model and the other three baselines is statistically significant under the significant level of 0.05. The comparison of the three metrics including  $R^2$ , MAE and  $R_p$  (Table S5) also demonstrated that on average SuperAtomicCharge is the top-performing model for the predictions of atomic partial charges.

It should be noted that, compared with the most satisfactory baseline (the ADP-RF model), the SuperAtomicCharge model is also equipped with higher

convenience and lightweight essence. In the original ADP-RF model, one should train individual models for different elements, and more critically, different models may share different lengths of ADP descriptors, which shows a degree of impracticality in real-life applications. Moreover, the model size of ADP-RF for the considered elements can accumulate to above 4 GB of storage in this study (Table S6). On the contrary, the SuperAtomicCharge model naturally takes a molecule as a graph and therefore training different models for different elements is not necessary. The model size of SuperAtomicCharge is tremendously light weighting where only 6.7 MB of storage is needed to allocate, which is approximately one-thousandth of the ADP-RF model's size. Finally, we also analyzed the prediction capacity to different elements in three datasets given by the best SuperAtomicCharge model in the Supporting Information. In brief, the SuperAtomicCharge model developed in this study has the advantages of higher accuracy, usability and portability compared with the other reported models.

### Performance of the predicted atomic charges on drug design applications

To test the proposed algorithms in practical drug design applications, we investigated the performance of the

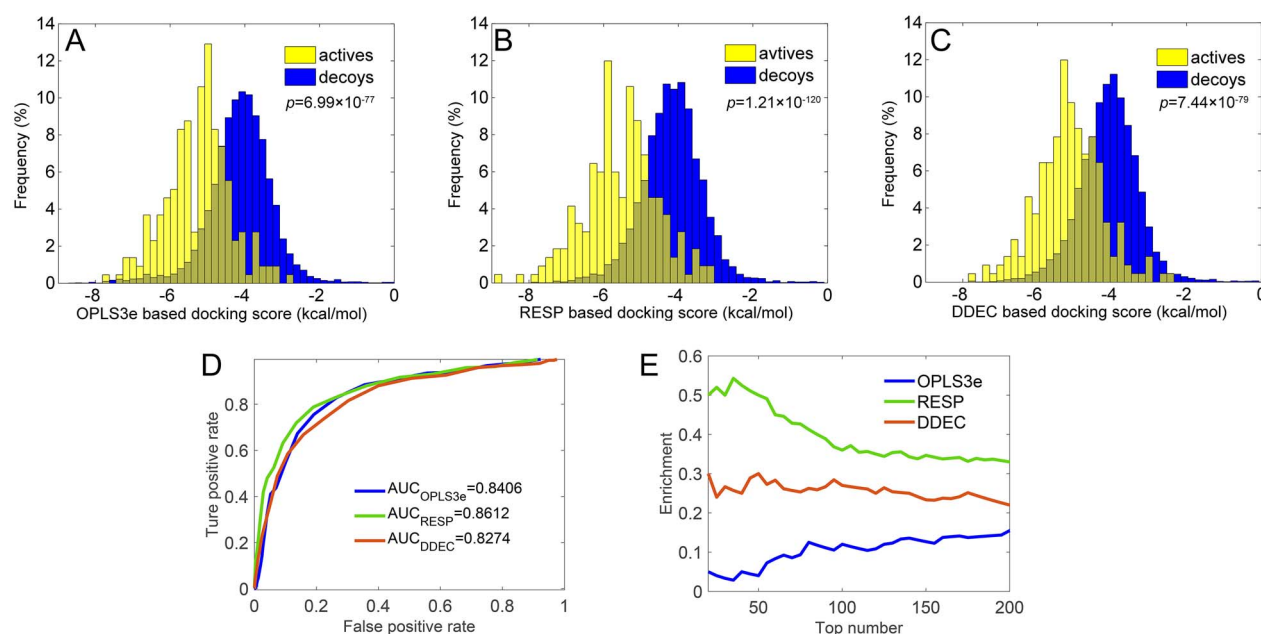


**Figure 3.** Scoring power (the ability of the algorithm to produce binding scores in a linear correlation with experimental binding data) for the predicted RESP (panels **A2**, **B2** and **C2**) and DDEC4 (panels **A3**, **B3** and **C3**) charges and the original AM1-BCC charges (panels **A1**, **B1** and **C1**), where the interior dielectric constant of 1 (panel **A**), 2 (panel **B**) and 4 (panel **C**) was used for the comparison.

RESP and DDEC4 atomic partial charges predicted by SuperAtomicCharge on the PDBbind dataset (1299 protein–ligand systems) for the estimation of the scoring power and two drug targets (caspase 8 and FXR) for the estimation of the screening power.

Assessment of the Algorithm Based on the Scoring Power. As shown in Figure 3, both the predicted RESP and DDEC4 atomic charges overall exhibit a higher accuracy compared with the predicted AM1-BCC charges. Especially for  $\epsilon_{in} = 1$  (panels **A1–A3**), the predicted RESP and DDEC4 atomic charges improve the  $R_p$  value of  $\sim 0.400$  for the AM1-BCC charges to  $\sim 0.450$ , suggesting that the proposed algorithms are useful for practical drug design

applications. For  $\epsilon_{in} = 2$  (panels **B1–B3**), the  $R_p$  values given by predicted RESP and DDEC4 atomic charges were also slightly improved in comparison with the results of the AM1-BCC charges. However, the predicted RESP and DDEC4 atomic charges gave almost the same  $R_p$  values of  $\sim 0.600$  for  $\epsilon_{in} = 4$  in comparison with the results of the AM1-BCC charges. The reason why there is no improvement based on the predicted charges for  $\epsilon_{in} = 4$  is potentially because the electrostatic interactions are dramatically weakened in the binding free energy calculations at such high dielectric constant. All in all, the high-level QM atomic charges predicted by SuperAtomicCharge can achieve a more reasonable accuracy for the



**Figure 4.** Screening power [the ability of algorithm to identify the known actives among the chemical background (decoys) toward a given target] for the caspase 8 system, where the P-values of the t-test for the original OPLS3e atomic charges, and the predicted RESP and DDEC4 atomic charges are shown in panels A–C, respectively. The corresponding AUC values of the ROC curves and the enrichment factors are illustrated in panels D and E, respectively, for the three types of atomic charges (blue for OPLS3e, green for RESP and red for DDEC4).

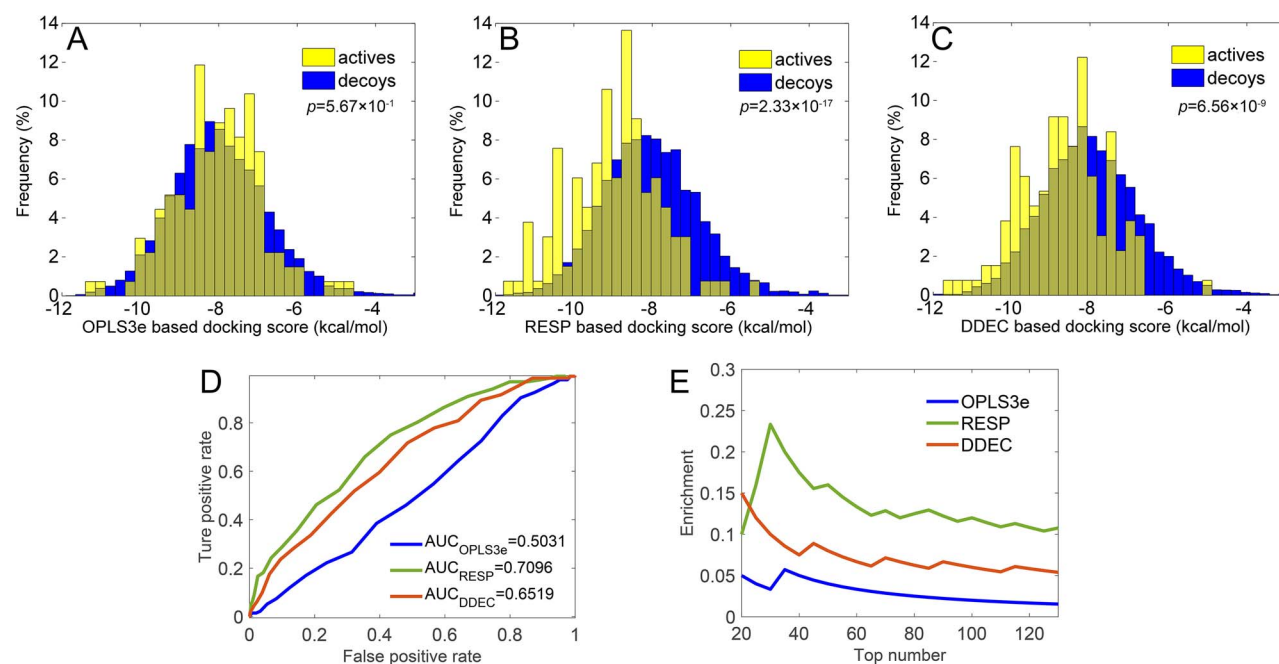
end-point binding free energy calculations according to the above analysis.

Assessment of the Algorithm Based on the Screening Power. To give a more comprehensive validation of the proposed algorithm, virtual screening for two drug targets (caspase 8 and FXR) was conducted to investigate the discrimination capability of the known actives and the chemical background. As shown in Figures 4 (caspase 8) and 5 (FXR), the predicted RESP charges exhibit the best screening power according to all the three metrics (P-value, area under the curve of roc (receiver operating characteristic curve) and enrichment factor) for both the two targets. It shows a P-value of  $1.21 \times 10^{-120}$  and an AUC of 0.861 to caspase 8, and a P-value of  $2.33 \times 10^{-17}$  and an AUC of 0.710 to FXR, which are much better than the results given by the original OPLS3e charges, where it only yielded a P-value of  $6.99 \times 10^{-77}$  and an AUC of 0.841 to caspase 8, and a P-value of  $5.67 \times 10^{-01}$  and an AUC of 0.503 to FXR. In addition, the predicted DDEC4 charges also give overall better results compared with the original OPLS3e charges, except for an AUC value of 0.827 to the caspase 8 system, which (red line in Figure 4D) is a bit lower than that of the original OPLS3e charges with an AUC value of 0.841 (blue line in Figure 4D). Here it should be mentioned that, for FXR, the OPLS3e charges do not show any capability to discriminate the actives from the decoys indicated by the completely overlapped distributions of the docking scores (P-value > 0.5, Figure 5A) and an AUC value of  $\sim 0.5$  (a random manner, blue line in Figure 5D). Moreover, both the predicted RESP and DDEC4 charges can reasonably discriminate the actives from the decoys, and additionally, both the predicted RESP

and DDEC charges showed much better results in terms of the intuitive enrichment factor in comparison with the original OPLS3e charges (Figure 5E), suggesting that the proposed methods are useful in large-scale virtual screening.

Further observation shows that the top-scored actives based on the predicted RESP and DDEC4 atomic charges for both caspase 8 (Figure 4B and C) and FXR (Figure 5B and C) prefer to gain stronger binding affinities compared with those derived from the OPLS3e atomic charge (Figures 4A and 5A), whereas leaving the distribution of the decoys nearly unchanged, which thereafter results in better enrichment factors for the two targets (Figures 4E and 5E). To better understand why the predicted RESP and DDEC4 atomic charges prefer to give better docking scores for the actives, the mean absolute atomic charges (MAACs) were calculated for the three types of atomic charges of the 135 FXR actives. As shown in Table 6, both the RESP and DDEC4 atomic charges exhibit larger absolute values (increased by 20% and 13% for the RESP and DDEC4 atomic charges, respectively) compared with the OPLS3e atomic charges, and the corresponding mean docking scores (MDSs) of the 135 FXR actives based on the RESP and DDEC4 atomic charges also increase by 11% (MDS =  $-8.91$  kcal/mol) and 8% (MDS =  $-8.66$  kcal/mol), respectively, in comparison with that based on the OPLS3e atomic charges (MDS =  $-8.02$  kcal/mol), indicating that larger absolute atomic charges may result in more favorable electrostatic interactions with the target for active compounds, whereas, on the contrary, may lead to a larger unfavorable electrostatic interaction with the target for decoy





**Figure 5.** Screening power for the FXR system, where the P-values of t-test for the original OPLS3e atomic charges, and the predicted RESP and DDEC4 atomic charges are shown in panels A–C, respectively. The AUC values of the ROC curves and the enrichment factors are illustrated in panels D and E, respectively, for the three types of atomic charges (blue for OPLS3e, green for RESP and red for DDEC4).

molecules. Furthermore, a case study was conducted on an example molecule (lithocholic acid) to give a more detailed comparison of the electrostatic effects for the three types of atomic charges. As shown in Figure 6A, the binding modes of lithocholic acid in the binding site of FXR predicted by Glide based on the three types of atomic charges (green, blue and orange for the OPLS3e, RESP and DDEC4 atomic charges, respectively) are highly overlapped, implying that the use of different types of atomic charges may not seriously affect the selection of the top-1 docking pose, and therefore any change of the binding affinity may probably contributed from the difference of the electrostatic effects arising from different types of atomic charges. As observed, much lower binding affinity was shown for the top-1 docking pose based on the OPLS3e atomic charges [docking score =  $-7.51$  kcal/mol, which is  $\sim 20\%$  higher than those based on the RESP and DDEC4 atomic charges ( $-9.11$  and  $-8.96$  kcal/mol)]. Further charge-spectrum analysis reveals that the OPLS3e atomic charges (green line in Figure 6B) usually exhibit much lower amplitude (with the MAAC of  $0.1124$  e) compared with the corresponding RESP (blue line in Figure 7B, MAAC =  $0.1561$  e) and DDEC4 (orange line in Figure 7B, MAAC =  $0.1606$  e) atomic charges. A further structural observation shows that the main difference between the OPLS3e and RESP/DDEC4 atomic charges lie in the carbon atoms, whose charges were usually parameterized to zero by empirical atomic charge models, whereas these electrically neutral atoms may exhibit much larger electrostatic effect within the more realistic high-level QM calculations.

Taken together, the above results clearly show that the high-level QM atomic charges predicted by the

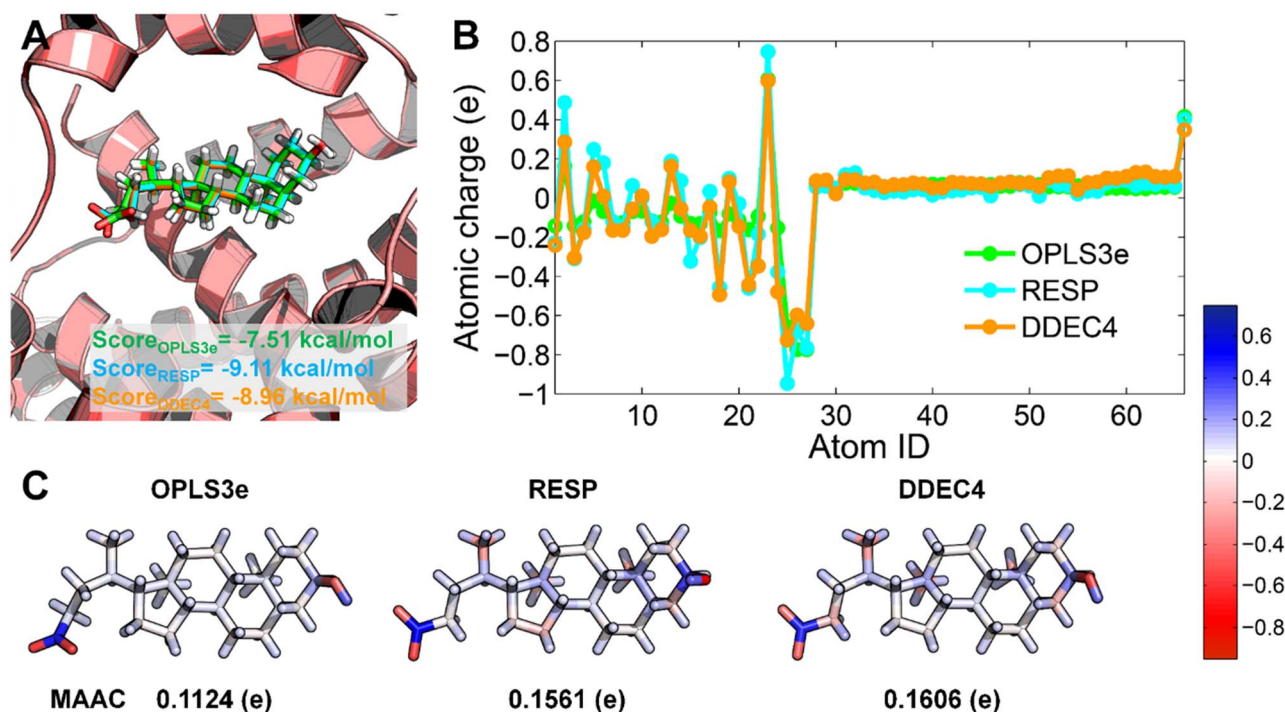
SuperAtomicCharge model can benefit the drug design campaign, and the SuperAtomicCharge models are useful tools in large-scale drug design applications.

#### Access of SuperAtomicCharge services

For those who wants to employ our proposed model to calculate the atomic partial charges in some drug design pipelines, we developed two easy-to-use tools including an online server and the source code command lines to satisfy such demands. As to the online server, we directly integrated the proposed model into the DeepChargePredictor server developed in our group (<http://cadd.zju.edu.cn/deepchargepredictor/>). Totally, three types of partial charges including DDEC4, DDEC78 and RESP can be predicted by SuperAtomicCharge. The inputs of the SuperAtomicCharge services are the same to those supported by DeepChargePredictor (SDF and MOL2 files) [35]. The outputs also include two types of files (SDF and MOL2). Among them, the MOL2 file that stores the predicted partial charges can be directly used as the inputs of some practical applications such as molecular docking and MD simulations. In addition, the outputted SDF file also stores the predicted partial charges and that can be easily processed by the commonly used cheminformatics tools such as RDKit (see the script: [https://github.com/zjujdj/SuperAtomicCharge/blob/main/scripts/get\\_sdf\\_charge.py](https://github.com/zjujdj/SuperAtomicCharge/blob/main/scripts/get_sdf_charge.py)). It should be noted that SuperAtomicCharge was developed based on the 3D molecular structures, and therefore the inputted molecules should contain the coordinates of atoms. It is recommended that the submitted molecule can be optimized by the merck molecular force field (MMFF) force field or PM7 method, and so on. In addition, the maximum number of the atoms in a molecule is

**Table 6.** Differences of MAACs and MDSs between the three types of atomic charges of the 135 FXR actives

	OPLS3e	RESP	DDEC4
MAAC (e)	0.1606	0.1921	0.1809
Ratio of change from the OPLS3e charge	–	20%	13%
MDS (kcal/mol)	–8.02	–8.91	–8.66
Ratio of change from the OPLS3e charge	–	11%	8%

**Figure 6.** Distributions of the three types of atomic charges for lithocholic acid. The binding modes and docking scores of lithocholic acid targeting FXR are shown in panel A, in which the small molecule with the OPLS3e, RESP and DDEC4 charges are colored in green, blue and orange stick models, respectively. The corresponding distributions of the atomic charges are illustrated with both spectra (panel B) and structures in panels B and C, respectively.

unlimited. More detailed information is available at the website of <http://cadd.zju.edu.cn/deepchargepredictor/>.

Due to the limited computing resources on the server and other irresistible factors, we also provide the source code command lines for the partial charge predictions using the well-trained models in this study. To use the source code command lines, two important steps including conda environment reproduce and repository clone should be accomplished first. As to troublesome environment reproduce, we provide the simplest way that uses the file packaged by conda-pack to reproduce the environment, and alternatively, the commonly used method that utilizes the package information files to reproduce the environment is also supported. After the two important steps, the ‘charge prediction’ and ‘model training’ usage is available in our GitHub page (<https://github.com/zjujdj/SuperAtomicCharge>).

## Conclusion

In this study, a deep graph learning-based model was developed to predict three kinds of important atomic

partial charges including DDEC4, DDEC78 and RESP based on the structures of molecules. A simple transfer learning strategy and an ACSF descriptors-based multitask learning strategy can be used to improve the prediction reliability of models. Compared with the latest GNN-based baseline and ML-based baseline, our proposed model achieved the best performance on the three external test sets. Moreover, the atomic partial charges of new molecules predicted by our model can be effectively used in some practical drug design applications such as structure-based virtual screening. Finally, the free services of our models were provided, and users can directly use the well-trained charge models to predict the partial charges of new molecules or trained their customized charge models using the new datasets.

### Key points

- A novel data-driven deep graph learning framework named SuperAtomicCharge was proposed to predict important quantum-mechanical (QM) partial charges only based on the structures of small molecules.

- SuperAtomicCharge showed better performance on all the external test sets and had better usability and portability compared with latest baselines.
- The predicted RESP and DDEC4 charges of new molecules given by SuperAtomicCharge showed more robust scoring and screening power than the commonly used partial charge in molecular docking.
- Two tools including an online server (<http://cadd.zju.edu.cn/deepchargepredictor>) and the source code command lines (<https://github.com/zjujdj/SuperAtomicCharge>) were developed for the easy access of the SuperAtomicCharge services, the results can be directly used in the biochemical and biophysical computations such as molecular docking.

## Supplementary data

Supplementary data are available online at <https://academic.oup.com/bib>.

## Funding

National Natural Science Foundation of China (81773632), Natural Science Foundation of China of Zhejiang Province (LZ19H300001), Key R&D Program of Zhejiang Province (2020C03010) and Fundamental Research Funds for the Central Universities (2020QNA7003).

## References

- Martin R, Heider D. ContraDRG: automatic partial charge prediction by machine learning. *Front Genet* 2019;**10**:00990.
- Racek T, Schindler O, Tousek D, et al. Atomic charge calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic Acids Res* 2020;**48**:W591–6.
- Bayly CI, Cieplak P, Cornell W, et al. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 1993;**97**:10269–80.
- Xu L, Sun HY, Li YY, et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *J Phys Chem B* 2013;**117**:8408–21.
- Bleiziffer P, Schaller K, Riniker S. Machine learning of partial charges derived from high-quality quantum-mechanical calculations. *J Chem Inf Model* 2018;**58**:579–90.
- Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 1980;**36**:3219–28.
- Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 2002;**23**:1623–41.
- Wang J, Cao D, Tang C, et al. DeepAtomicCharge: a new graph convolutional network-based architecture for accurate prediction of atomic charges. *Brief Bioinform* 2021;**22**(3):bbaa183.
- Wang J, Cao D, Tang C, et al. Fast and accurate prediction of partial charges using atom-path-descriptor-based machine learning. *Bioinformatics* 2020;**36**(18):4721–8.
- Cho MS, Sylvetsky N, Eshafi S, et al. The atomic partial charges arboretum: trying to see the Forest for the trees. *ChemPhysChem* 2020;**21**:688–96.
- Kato K, Masuda T, Watanabe C, et al. High-precision atomic charge prediction for protein systems using fragment molecular orbital calculation and machine learning. *J Chem Inf Model* 2020;**60**:3361–8.
- Rai BK, Bakken GA. Fast and accurate generation of ab initio quality atomic charges using nonparametric statistical regression. *J Comput Chem* 2013;**34**:1661–71.
- Sifain AE, Lubbers N, Nebgen BT, et al. Discovering a transferable charge assignment model using machine learning. *J Phys Chem Lett* 2018;**9**:4495–501.
- Wang XC, Gao J. Atomic partial charge predictions for furanoses by random forest regression with atom type symmetry function. *RSC Adv* 2020;**10**:666–73.
- Jiang D, Wu Z, Hsieh C-Y, et al. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *J Cheminform* 2021;**13**:12.
- Duvenaud DK, Maclaurin D, Iparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints. *Adv Neural Inform Process Syst*, Montreal, Canada, December 7–12. 2015;2224–32.
- Cho H, Lee EK, Choi IS. InteractionNet: Modeling and explaining of noncovalent protein-ligand interactions with noncovalent graph neural network and layer-wise relevance propagation. *arXiv preprint arXiv:2005.2020*;2020;**13438**.
- Kwiatkowski JS, Leszczyński J. Ab initio Hartree–Fock and post-Hartree–Fock studies on molecular structures and vibrational spectra of formamide and formamidic acid. *J Mol Struct* 1992;**270**:67–86.
- Li P, Li Y, Hsieh C-Y, et al. TrimNet: learning molecular representation from triplet messages for biomedicine. *Brief Bioinform* 2021;**22**:bbab266.
- Xiong Z, Wang D, Liu X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem* 2019;**63**:8749–60.
- Cai C, Wang S, Xu Y, et al. Transfer learning for drug discovery. *J Med Chem* 2020;**63**:8683–94.
- Playe B, Stoven V. Evaluation of deep and shallow learning methods in chemogenomics for the prediction of drugs specificity. *J Chem* 2020;**12**:11.
- Hu W, Liu B, Gomes J, et al. Strategies for pre-training graph neural networks. *arXiv preprint arXiv* 2019;**1905**:12265.
- Zhang X-C, Wu C-K, Yang Z-J, et al. MG-BERT: leveraging unsupervised atomic representation learning for molecular property prediction. *Brief Bioinform* 2021;**22**(6):bbab152.
- Liu X, Luo Y, Song S, et al. Pre-training of graph neural network for Modeling effects of mutations on protein-protein binding affinity. *arXiv preprint arXiv* 2020;**2008**:12473.
- Liu Z, Su M, Han L, et al. Forging the basis for developing protein–ligand interaction scoring functions. *Acc Chem Res* 2017;**50**:302–9.
- Wang R, Fang X, Lu Y, et al. The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J Med Chem* 2004;**47**:2977–80.
- Li X, Xu Y, Lai L, et al. Prediction of human cytochrome P450 inhibition using a multitask deep autoencoder neural network. *Mol Pharm* 2018;**15**:4336–45.
- Jiang D, Lei T, Wang Z, et al. ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning. *Journal of Chem* 2020;**12**:1–26.

30. Liu Z, Lin L, Jia Q, et al. Transferable multilevel attention neural network for accurate prediction of quantum chemistry properties via multitask learning. *J Chem Inf Model* 2021;**61**:1066–82.
31. Karlov DS, Sosnin S, Fedorov MV, et al. graphDelta: MPNN scoring function for the affinity prediction of protein–ligand complexes. *ACS omega* 2020;**5**:5150–9.
32. Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem Sci* 2017;**8**:3192–203.
33. Wu Z, Jiang D, Wang J, et al. Mining toxicity information from large amounts of toxicity data. *J Med Chem* 2021;**64**:6924–36.
34. Feinberg EN, Joshi E, Pande VS, et al. Improvement in ADMET prediction with multitask deep Featurization. *J Med Chem* 2020;**63**:8835–48.
35. Wang J, Sun H, Chen J, et al. DeepChargePredictor: a web server for predicting QM-based atomic charges via state-of-the-art machine-learning algorithms. *Bioinformatics* 2021;**37**(22):4255–7.