A Fast Empirical GAFF Compatible Partial Atomic Charge Assignment Scheme for Modeling Interactions of Small Molecules with Biomolecular Targets

Goutam Mukherjee and B. Jayaram*

Department of Chemistry and Supercomputing Facility for Bioinformatics and Computational Biology, Indian Institute of Technology Hauz Khas, New Delhi-110016, India.

Email: goutam@scfbio-iitd.res.in; bjayaram@chemistry.iitd.ac.in; Website: www.scfbio-iitd.res.in

ABSTRACT

We report here a new and fast approach (TPACM4: Transferable Partial Atomic Charge Model – up to 4 bonds) for deriving the partial atomic charges of small molecules for use in protein/DNA-ligand docking and scoring. In our method, we have used 5363 atom types to cover the chemical space of C, H, O, N, S, P, F, Cl and Br atoms in small molecules. Starting with a set of Cartesian coordinates, partial atomic charges are developed by considering diverse plausible chemical environments of each atom in a molecule. The model gives an average error ±1.16 kcal/mol and a correlation coefficient of 0.90 vis-à-vis an average error of ±1.02 kcal/mol and a correlation coefficient of 0.92 obtained with RESP fit charges in calculations of binding free energies of 161 protein-ligand complexes in comparison to experiment. This new method of charge derivation can also accurately predict hydrogen bond energetics, solvation free energies of small molecules. For a molecule containing 50-55 atoms, the method takes approximately 0.4 seconds to assign partial atomic charges. The TPACM4 programme has been web-enabled at http://scfbio-iitd.res.in/software/drug design/charge.jsp.

INTRODUCTION

Partial atomic charge is very crucial for computing physical, chemical and biological properties, and reactivity of molecules. Through the information of the atomic charge in a given species it is possible to predict the stability, solvation energetics of various molecules, course of a particular reaction, determine its interaction with biological molecules and so on. The usefulness, notwithstanding, there is no direct method to determine the partial atomic charges from experiment. Also, there is no universally agreed upon best procedure for computing partial atomic charges. During the last few decades various methods have been developed to determine the partial atomic charges, but all these methods have their limitations.

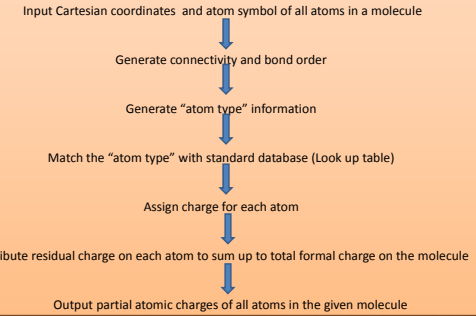
MATERIALS & METHODS

Dataset

For our study we have chosen six nucleic acids base pairs, and 24 of the 400 (20 × 20) possible interacting side chain pairs of amino acids and 24 hydrogen bonded small organic molecules to test the ability of the TPACM4 charge models to reproduce quantum mechanical/experimental secondary bonding interactions between these dimers. The energies of these base pairs were obtained from literature. For calculation of solvation free energies we have chosen 24 small molecules whose experimental solvation free energies are known in literature. For protein-ligand binding free energy calculation a dataset of 161 protein-ligand complexes were chosen from BAPPL data set. The experimental binding free energies for these complexes were also available from the same data base.

Methodology

The main idea of TPACM4 is based on a look up table of template fragments consisting of 4-bond paths around the atom being assigned charges. A computational flowchart for the assignment of partial atomic charge on an atom in a given molecule is shown below.



RESULTS AND DISCUSSION

In this study, a new model of partial charge derivation is developed on the basis of the bonded topology of the molecules. The model is tested on a dataset of 161 protein-ligand complexes whose experimental binding free energies is reported in literature. The model is found to perform fairly well, giving a correlation coefficient of 0.90 vis-à-vis a correlation coefficient of 0.92 obtained with RESP fit charges between the experimental and the predicted binding free energies. The standard error of estimate is found to be ±1.16 kcal/mol for TPACM4 model and ± 1.02 kcal/mol for RESP model. This model also performs fairly well for prediction of solvation free energies and dimer stabilization energies.

Comparison of base pair stabilization energies (kcal) computed by various charge models

S. No.	Base pairs	EXPT.	RESP	AM1BCC	TPACM4
1	AT	-13	-13.25	-14.72	-12.74
2	GC	-21	-28.28	-27.45	-22.30
3	CC	-16	-18.90	-21.23	-12.65
4	TT	-9	-12.09	-12.61	-9.70
5	AU	-14.5	-13.20	-14.62	-11.92
6	UU	-9.5	-10.35	-12.36	-8.67
	Average Error		2.61	3.33	1.50

A comparison of amino acid side chain interaction energies (kcal) computed via QM and various charge models

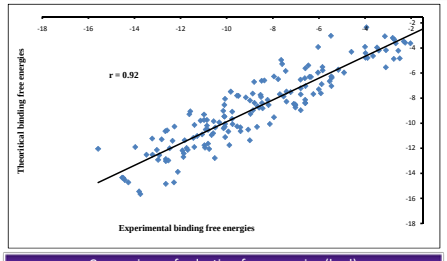
S. No.	Dimer	QM	RESP	AM1BCC	TPACM4
1	AL	-1.07	-0.96	-0.96	-0.91
2	DH(N)	-17.97	-11.62	-10.95	-10.14
3	DH	-30.64	-21.68	-18.91	-20.94
4	FF	-2.33	-2.02	-2.02	-1.24
5	II	-1.24	-1.21	-1.20	-1.29
6	IL	-1.39	-1.43	-1.43	-1.42
7	KE(N)	-10.76	-9.49	-8.12	-7.81
8	KE	-108.4	-98.52	-99.7	-98.54
9	LG	-0.77	-0.73	-0.73	-0.72
10	LL	-1.62	-1.67	-1.67	-1.50
11	LT	-1.09	-1.09	-1.10	-1.21
12	LW	-4.04	-3.41	-3.36	-3.63
13	LY	-1.72	-1.57	-1.56	-1.65
14	MC	-1.46	-1.36	-1.38	-1.13
15	MM	-2.03	-2.20	-2.17	-2.03
16	QN	-7.37	-6.72	-5.75	-6.56
17	RD(N)	-16.32	-15.33	-11.62	-8.32
18	RD	-110.8	-98.62	-98.5	-97.3
19	TS	-4.50	-4.46	-3.79	-4.37
20	TT	-6.50	-7.24	-6.25	-6.85
21	VL	-1.08	-1.10	-1.10	-1.11
22	VV	-1.39	-1.48	-1.49	-1.48
23	YP	-3.79	-3.35	-3.26	-2.8
24	YY	-4.66	-4.04	-3.85	-3.80
	Average Error		1.83	2.20	2.40

A comparison of hydrogen bonded dimer stabilization energies (kcal) computed via QM and various charge models

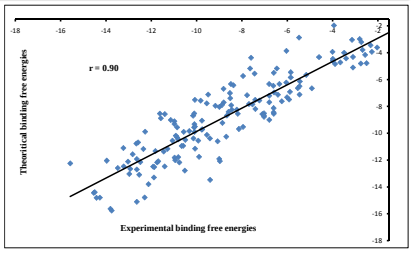
S. No.	Dimer	QM	RESP	AM1BCC	TPACM4
1	CH3OH---OH2	-6.20	-7.25	-6.70	-5.63
2	CH3HO---HOH	-6.74	-6.42	-5.65	-7.83
3	PhHO---HOH	-5.57	-5.00	-4.99	-5.31
4	PhOH---OH2	-8.60	-7.07	-8.12	-8.94
5	T-NMA---HOH	-7.97	-9.28	-9.12	-9.14
6	T-NMA---OH2	-6.66	-6.86	-6.03	-6.99
7	Formamide dimer	-14.36	-13.88	-11.5	-14.64
8	H3N---HOH	-7.65	-8.36	-6.96	-8.32
9	CH3COOH---OH2	-10.45	-10.95	-9.90	-10.19
10	Oxalic acid---OH2	-11.27	-12.37	-12.52	-11.86
11	Im---OH2	-7.86	-7.10	-6.29	-8.31
12	ImN---HOH	-7.84	-6.75	-7.33	-6.10
13	Indole---OH2	-7.34	-6.85	-6.04	-6.50
14	Pyrrole---OH2	-6.99	-6.40	-5.64	-6.44
15	Pyridine---HOH	-7.25	-5.73	-6.29	-5.15
16	Acetone---HOH	-6.57	-6.84	-6.58	-6.43
17	CH2O---HOH	-5.47	-5.29	-5.58	-5.27
18	PhNH2---OH2	-5.15	-5.08	-5.33	-5.45
19	Me2S---HOH	-3.04	-3.79	-3.41	-3.73
20	MeSSMe---HOH	-4.10	-3.54	-3.60	-3.40
21	H2S---SH2	-1.57	-1.57	-0.08	-1.66
22	PyNO---HOH	-9.04	-10.41	-8.49	-10.30
23	CH3COOH---NH3	-12.11	-11.97	-10.20	-11.73
24	MeNH3---OH2	-19.80	-18.41	-19.14	-19.32
	Average Error		0.71	0.89	0.65

Correlation between predicted and experimental binding free energies (kcal) for 161 protein-ligand complexes

RESP Charge model



TPACM4 Charge model



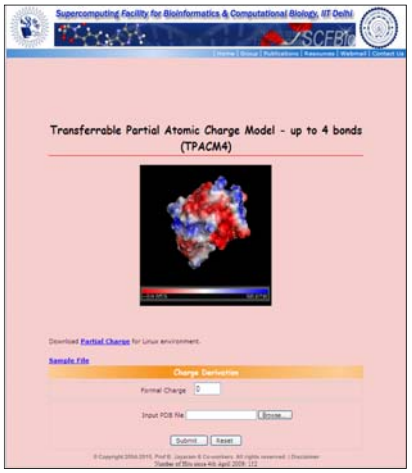
Comparison of solvation free energies (kcal) computed by various charge models

S. No.	Molecule	EXPT.	RESP	AM1BCC	TPACM4
1	Methanol	-5.08	-3.90	-3.29	-5.02
2	Ethanol	-4.9	-3.05	-2.6	-3.33
3	Ammonia	-4.31	-7.94	-5.89	-7.19
4	Methylamine	-4.57	-3.23	-2.84	-4.09
5	Ethylamine	-4.50	-2.16	-2.09	-3.27
6	Methylthiol	-1.24	-2.02	-1.36	-1.88
7	Acetone	-3.85	-6.29	-3.69	-3.56
8	2-Butanone	-3.64	-5.30	-3.08	-2.46
9	Acetaldehyde	-3.50	-4.53	-4.40	-4.5
10	Proponal	-3.44	-3.32	-3.76	-3.55
11	Acetic acid	-6.70	-7.92	-6.52	-5.13
12	Propionic acid	-6.47	-6.51	-5.91	-4.27
13	Acetamide	-9.72	-10.80	-9.28	-10.26
14	propionamide	-9.42	-9.44	-8.47	-9.00
15	Benzene	-0.87	-0.32	-0.21	-1.18
16	Toulene	-0.76	0.35	0.33	-0.28
17	Pyridine	-4.70	-2.82	-2.49	-2.07
18	Phenol	-6.62	-3.92	-3.92	-4.47
19	Meimidazole	-10.25	-7.06	-2.86	-5.56
20	Ammonium	-81.53	-95.56	-88.75	-89.01
21	n-Butylammonium	-69.24	-66.97	-65.58	-66.16
22	Acetate ion	-80.65	-80.75	-78.69	-86.74
23	Propionate ion	-79.12	-77.47	-75.92	-74.29
24	Methylimidazolium	-64.13	-58.78	-61.94	-57.60
	Average Error		2.15	1.93	2.18

Computational efficiency of TPACM4 method

Average computing time required for a molecule containing 50 to 55 atoms is ~0.4 seconds whereas RESP or AM1BCC take several minutes. In spite of the enormous speed with minimum computational efforts, TPACM4 does not loose it's accuracy in calculating the partial atomic charges.

http://scfbio-iitd.res.in/software/drugdesign/charge.jsp



CONCLUSION

The core of the model is the creation of a look up table of partial atomic charges of all possible atom types with different chemical environments which are generally found in organic chemistry. A given atom in any molecule is then matched against the look-up table and its charge assigned. This method thus overcomes the limitations of time complexity of deriving the partial atomic charges of a given molecule. The low errors against experimental/QM values of diverse physico-chemical properties indicates the reliability comparable to RESP/AM1BCC models.

REFERENCES:

Jayaram, B.; Sprous, D.; Beveridge, D. L. J Phys Chem B 1998, 102, 9571.
Jardine, W. K; Langler, R. F.; Macgregor, J. A. Can J Chem 1982, 60, 2069.
Faerman, C. H.; Price, S. L. J Am Chem Soc 1990, 112, 4915.
Mulliken, R. S. J Chem Phys 1955, 23, 1833, 1841, 2338, 2343.
Löwdin, P. O. J Chem Phys 1955, 21, 374.
Gasteiger, J.; Marsili, M. Tetrahedron Letters 1978, 34, 3181.
Gasteiger, J.; Marsili, M. Tetrahedron 1980, 36, 3219.
Cioslowski, J.; Hay, P. J.; Ritchie, J. P. J Phys Chem 1990, 94, 148.
Bader, R. F. W.; Beddall, P. M.; Cade, P. E. J Am Chem Soc 1971, 93, 3095.
Merz, K. M.; Kollman, P. A. J Comput Chem 1990, 4, 431.
Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. J Comput Chem 2000, 21, 132.
Jain, T.; Jayaram, B. FEBS Letters 2005, 579, 6659.
Shaikh, S., Jain. T.; Sandhu, G.; Latha, N.; Jayaram, B. Current Pharmaceutical Design 2007, 13, 3454.
Gupta, A.; Gandhimathi, A.; Sharma, P.; Jayaram, B. Protein and Peptide Letters 2007, 14, 632.
www.begdb.com
www.scfbio- iitd.res.in/software/drugdesign/proteinliganddataset.html