Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation

ARAZ JAKALIAN,[1] DAVID B. JACK,[2] CHRISTOPHER I. BAYLY[3]

[1]Boehringer Ingelheim (Canada) Ltd Research and Development, 2100 Rue Cunard, Laval, Quebec, Canada, H7S 2G5

[2]Department of Chemistry and Biochemistry, Concordia University, 1455 de Maisonneuve Blvd. Ouest, Montréal, Québec, Canada, H3G 1M8

[3]Merck Frosst Canada & Co., 16711 Trans-Canada Hwy, Kirkland, Québec, Canada, H9H 3L1

Received 1 February 2002; Accepted 26 April 2002

Published online 18 October 2002 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jcc.10128

Abstract: We present the first global parameterization and validation of a novel charge model, called AM1-BCC, which quickly and efficiently generates high-quality atomic charges for computer simulations of organic molecules in polar media. The goal of the charge model is to produce atomic charges that emulate the HF/6-31G* electrostatic potential (ESP) of a molecule. Underlying electronic structure features, including formal charge and electron delocalization, are first captured by AM1 population charges; simple additive bond charge corrections (BCCs) are then applied to these AM1 atomic charges to produce the AM1-BCC charges. The parameterization of BCCs was carried out by fitting to the HF/6-31G* ESP of a training set of >2700 molecules. Most organic functional groups and their combinations were sampled, as well as an extensive variety of cyclic and fused bicyclic heteroaryl systems. The resulting BCC parameters allow the AM1-BCC charging scheme to handle virtually all types of organic compounds listed in The Merck Index and the NCI Database. Validation of the model was done through comparisons of hydrogen-bonded dimer energies and relative free energies of solvation using AM1-BCC charges in conjunction with the 1994 Cornell et al. forcefield for AMBER.[13] Homo- and hetero-dimer hydrogen-bond energies of a diverse set of organic molecules were reproduced to within 0.95 kcal/mol RMS deviation from the *ab initio* values, and for DNA dimers the energies were within 0.9 kcal/mol RMS deviation from *ab initio* values. The calculated relative free energies of solvation for a diverse set of monofunctional isosteres were reproduced to within 0.69 kcal/mol of experiment. In all these validation tests, AMBER with the AM1-BCC charge model maintained a correlation coefficient above 0.96. Thus, the parameters presented here for use with the AM1-BCC method present a fast, accurate, and robust alternative to HF/6-31G* ESP-fit charges for general use with the AMBER force field in computer simulations involving organic small molecules.

 J Comput Chem 23: 1623–1641, 2002

Key words: organic small molecules; atomic charges; force-field simulations; AMBER; electrostatic potential fit charges

Introduction

The two-body additive force fields (FFs) commonly used in biomolecular simulations calculate electrostatics using atom-centered point charges (atomic charges) within a simple Coulombic model. Quantitatively reproducing electrostatic interactions, which play a key role in nonbonded interactions such as ligand binding to a receptor, thus becomes the task of generating a suitable set of atomic charges. An atomic charge must represent not only the influence of the atom itself and its immediate bonded neighbors, but it must also account for the local effects of any other nearby functional groups, extended π systems (with electron donating and withdrawing groups), and formal charges that may be present. Small molecule ligands such as those made and tested in pharmaceutical research can have all of these effects present simultaneously, and a satisfactory charge model must accommodate them all. Good atomic charges for two-body additive FFs can be ob-

Correspondence to: C. I. Bayly

tained by fitting to the electrostatic potential (ESP) of a molecule calculated at the HF/6-31G* level of theory; refs. 1 and 2 offer good background reviews on the subject. The HF/6-31G* level of theory (and the ESP-fit charges derived from it) overestimates the polarity of molecules by approximately 10 to 15% in addition to accounting for the electronic effects mentioned above. This overestimation offers a fortuitous "implicit" polarization to compensate for the fact that a two-body additive FF, by construction, does not include polarization; thus it balances the exaggerated polarity of widely used water models, for example, SPC, TIP3P, and TIP4P.[1,2] This ESP-fit charge model has proven itself through being able to reproduce relative free energies between organic systems of interest within a FF formalism.[3] While ESP-fit methods produce suitable charges for condensed phase simulations, there are inherent problems in current ESP-fit methods: charges are conformer dependent; charges on and near buried atoms (e.g., methyl carbon atoms) are numerically unstable, that is, their magnitude can vary widely while barely perturbing the quality of the fit; large molecules must be treated as a superposition of fragments due to computer resource limitations; and charges are not easily transferable between common functional groups in related molecules. In addition, the costs of the CPU intensive calculation of the ESP at the *ab initio* level pose a significant barrier to the routine use of ESP charges, especially in pharmaceutical applications where the computational evaluation of potential ligands requires ever-higher throughput.

The need for a faster and more efficient method to generate high-quality atomic charges has led to the recently introduced AM1-BCC methodology.[4] AM1 population charges[5] (hereafter referred to as AM1 atomic charges), although incapable of satisfactorily reproducing the QM ESP, capture underlying electronic features such as formal charge and electron delocalization. Simple additive bond charge corrections (BCCs) are then applied based on the general atom and bond types in the molecule so that the resulting charges emulate the HF/6-31G* ESP that would be generated for that molecule.

This approach differs fundamentally from previous efforts to approximate HF/6-31G* ESP-fit charges based on scaling charges derived from a semiempirical wave function[6a,b] in that no scaling is involved. Scaling produces erroneous and nonintegral atomic charges when applied to formally charged molecules, and is built on the questionable assumption that for all the chemistry considered, a single global scaling factor can quantitatively transform the features of a semiempirical ESP into those of a HF/6-31G* ESP. Also, scaling does nothing to rectify the inherent numerical instability of the charge-fitting procedure.[2] The AM1-BCC method is more similar in spirit to the CM*x* methods of Cramer and Truhlar,[6c,d] but the AM1-BCC method parameterizes additive corrections by fitting directly to the HF/6-31G* ESP of a training set of molecules, whereas the CM*x* formalism involves a bond-order calculation from the wave function and is parameterized to fit molecular dipole moments. A more detailed comparison between these methods is given in the first article in this series (ref. 4), which describes the AM1-BCC methodology.

The current work presents the global parameterization of the BCCs and the validation of the resulting AM1-BCC charge model. The goal of this work was to parameterize the BCCs for all organic small molecules that could potentially be ligands for macromolecules. This requires a general atom- and bond-typing scheme that spans the required chemistry and a large training set that samples the atom- and bond-types. Once parameterized, a validation procedure consisting of FF calculations of hydrogen-bonded dimers and free energies of solvation was used to verify the quality of the charge model.

Methods

Electrostatic Potentials

The *ab initio* ESPs used for both the parameterization and for generating RESP (restrained electrostatic potential) charges[7] were evaluated at the HF/6-31G* level of theory using GAUSSIAN92.[8] A relativistic effective core potential was used for iodine with a valence space consisting of 4d, 5s, and 5p shells.[9] The ESP was sampled using a face-centered cubic grid of points, with 0.5 Å spacing, located outside the van der Waals surface at 1.4 to 2.0 times the van der Waals radius of each atom. The standard two stage fitting protocol[7] was used to fit RESP charges, except that because a more dense face-centered cubic grid of points was used instead of the Connolly grid,[10] the charge restraints were recalibrated to 0.001 and 0.01 for the first and second stages, respectively.

AM1 Calculations

The AM1[5a] calculations were performed using MOPAC 6.0[5b] with the following input keywords: **AM1 GEO-OK MMOK EF CHARGE=***n*, where *n* is the net charge on the molecule. AM1 atomic charges used in this work are the atomic population charges produced by default by MOPAC 6.0 and are derived from the Coulson density matrix. The same charges can be obtained from GAUSSIAN94[11] using the **Iop(4/24=3)** keyword, except for sulfur- or phosphorus-containing molecules, which will have incompatible charges because GAUSSIAN94 uses MNDO parameters for these atoms. For the training set used in the parameterization, geometry optimizations were carried out. When performing the AM1 calculation to obtain the population charges as precharges for the AM1-BCC method, it is important to avoid short-range through-space polar interactions such as intramolecular hydrogen bonds. Such interactions locally polarize the AM1 wave function, perturbing the population charges significantly away from the typical range for which the BCCs are parameterized. Adding the BCCs to such charges will yield a charge set with exaggerated polarization in that area. In such cases, an extended conformation without intramolecular hydrogen bonds should be used.

AM1-BCC Charges

The AM1-BCC charge $q_j^{\mathrm{AM1-BCC}}$ for atom *j* is obtained by adding the net charge correction q_j^{BCC} to the AM1 population atomic charge q_j^{AM1}:

$$q_j^{\mathrm{AM1}} + q_j^{\mathrm{BCC}} = q_j^{\mathrm{AM1-BCC}} \quad (1)$$

The net charge correction q_j^{BCC} depends on only the immediate bonded neighbors of atom j and is given by

$$q_j^{BCC} = \sum_{\alpha=1}^{nb} T_{j\alpha} B_\alpha \quad (2)$$

where $T_{j\alpha}$ is the bond connectivity template matrix[4] (hereafter referred to as the **T** matrix), B_α is the BCC for bond type α, and the summation runs over all immediate bonded neighbors of atom j. A bond type α is composed of atom type I, the bond order between atoms i and j, and atom type J. The atom typing and bond typing definitions are given later.

Dimer Energies

The binding energy E_{bind} between two molecules was evaluated as

$$E_{bind} = E_{AB} - (E_A + E_B) \quad (3)$$

where E_{AB} is the energy of the hydrogen-bonded complex formed between molecules A and B, and E_A and E_B are the individual energies of molecules A and B, respectively. A validation set of 47 organic dimers and 27 DNA dimers was chosen to test the ability of the charge models to reproduce *ab initio* noncovalent interactions between homo- and hetero-dimers of molecules containing common organic functionalities. Monomers and dimers were energy minimized using the steepest descent method[12] for 10 cycles and then switched to the conjugate gradient method[12] until a DRMS of 0.1 kcal/mol Å was achieved. All FF dimer calculations were performed with the Cornell et al. force field[13] using the Sander module of AMBER-5.01.[14] The MP2/6-31G*(.25) energies for the DNA dimers were taken from ref. 15, and the MP2/6-31+G** energies for the organic dimers were calculated using GAMESS[16] at the HF/6-31G* geometries. All molecules were charged with the RESP, AM1, MMFF, and AM1-BCC charge models with the exception of the TIP3P water. Initial geometries for the DNA dimers were reproduced from ref. 15, while the organic molecule dimers were obtained from the MMFF validation suite available from the Computational Chemistry List.[17]

Free Energy Calculations

Free energy calculations were carried out using the Gibbs module of AMBER-5.01. Molecules were solvated in TIP3P water with 12 Å separating the solute from the edges of the box. The solvated systems were then minimized using the steepest descent method for 10 cycles and then switched to the conjugate gradient method until a DRMS of 0.1 kcal/mol Å was achieved. The systems were then subjected to 20 ps of molecular dynamics at 1 K in order to diffuse any buildup of potential energy that may have occurred during the placement of the waters. The next stage consisted of 20 ps of heating to 300 K followed by 40 ps of equilibration at the target temperature. The vdW + electrostatic perturbations were performed using slow growth over 200 ps. Free energy differences between charge models were evaluated using free energy perturbation (FEP) using 21 windows, with 400 steps of equilibration and 600 steps of data collection per window. The reported perturbation energies are the mean values of the forward and reverse directions. All simulations were run at constant temperature (300 K) and pressure (1 atm) using a constant dielectric of 1 with a 10 Å nonbonded cutoff, 2 fs. time step, and SHAKE applied to all bonds containing hydrogen atoms except those involved in the perturbations. The potential of mean force (PMF) correction was calculated for perturbed bonds. Separate scaling factors were used for solvent and solute atoms with heat bath coupling constants of 0.2 ps while in the gas phase a coupling constant of 0.4 ps was used. The pressure relaxation time was set to 0.2 ps. The 1-4 nonbonded vdW and electrostatic interactions were scaled by 1/2 and 1/1.2, respectively.

Global Parameterization of BCCs

The objective of the global parameterization was to generate a complete set of BCCs that would enable this charge model to be used on any organic small molecule that might serve as a ligand in a biological context. To accomplish this, several considerations had to be taken into account. First, the breadth of organic chemical structures to be parameterized had to be characterized. With no analytic way to do this, we selected The Merck Index[18] (an encyclopedia of >10,000 chemicals, drugs, and biological molecules) and the National Cancer Institute (NCI) structure database of about 250,000 molecules,[19] considering them to comprehensively sample the required chemistry. The bond types generated from the training set were compared to the bond types generated from The Merck Index and the NCI database. If bond types were present in either the Index or the database and not in the training set, then a minimum of four diverse molecules were created to sample the missing bond type. Second, an atom- and bond-typing scheme had to be chosen that would (a) encompass the required chemistry, (b) generate the smallest total number of parameters (BCCs) in order to make the computations tractable, and (c) give the best fit in terms of emulating the RHF/6-31G(d) ESP. Third, a training set of molecules had to be selected, against which the fitting was to be carried out. The fitting was carried out as described previously,[4] based on the AM1 population charges and the RHF/6-31G(d) ESP of each molecule in the training set. The overall process was iterative, applying a trial atom- and bond-typing scheme to the small molecule databases to see what BCCs would require fitting, expanding the training set as needed for any new BCC parameter to be fit, simultaneously fitting the entire set of BCCs, and testing the resulting parameters. If not satisfactory, the atom- and bond-typing scheme was modified and the process repeated. Throughout the development of the atom- and bond-types, the atom- and bond-typing schemes were kept simple and parsimonious so as to simplify the fitting process, decrease the degrees of freedom, and provide a robust method applicable to any organic molecule. Eleven elements were parameterized in this work: *C*, *N*, *O*, *F*, *S*, *P*, *Cl*, *Br*, *I*, *Si*, and *H*. Initially, atoms were typed according to element and valency. Seven bond orders were defined: single, double, triple, aromatic single, aromatic double, dative, and delocalized; however, the parameterization was based upon only six *unique* bond orders because aromatic single and

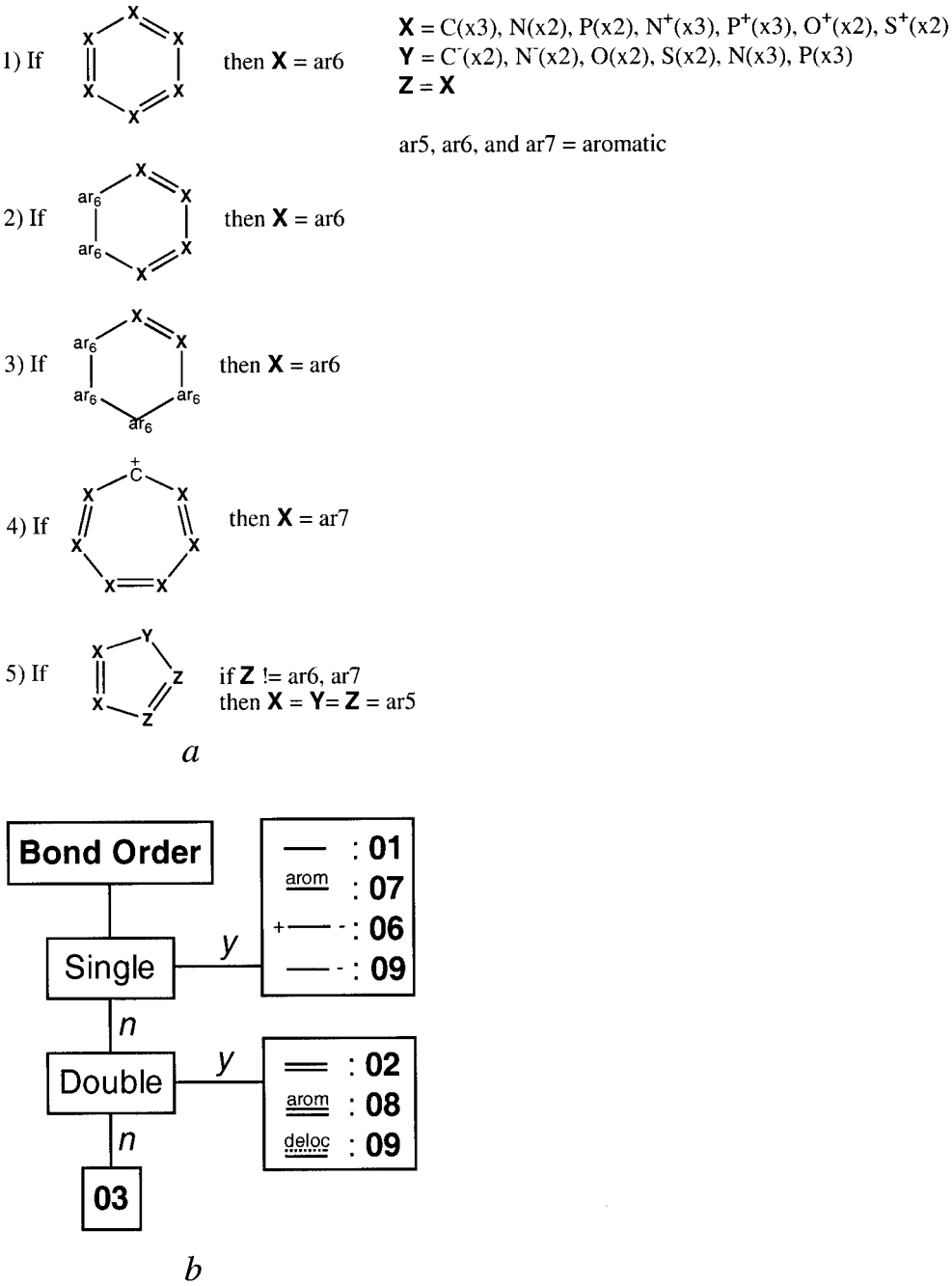


Figure 1. (a) Rules to define aromaticity for purposes of atom- and bond-typing. Five rules are given, beginning with a chemical substructure pattern and definitions of allowable atoms for X, Y, and Z. In the atom definitions, in parentheses after each element, x2 and x3 denote bivalent and trivalent atoms, respectively. If the chemical substructure for the rule has a match in the structure being typed, a property ar5, ar6, or ar7 may be attached to one or more atoms, identifying them as aromatic atoms in a 5, 6, or 7-membered ring, respectively. This property may then be used as part of the pattern in subsequent rules. The property of an atom may be modified by subsequent rules, so all five rules must be progressively applied. A bond between aromatic atoms is considered aromatic if and only if it is within the ring. In rule 5, the expression "**Z**!=ar6,ar7" means "atoms **Z** are not part of a six- or seven-membered aromatic ring." (b) Definition of the bond order codes: 01 = single bond, 02 = double bond, 03 = triple bond, 06 = dative bond (e.g., N-oxide), 07 = aromatic single bond, 08 = aromatic double bond, 09 = single bond with charge (e.g., methoxide, sulfoxide, etc.) or delocalized bond (e.g., nitro or carboxy).

Table 1. Summary of Atom-Type and Bond Order Definitions.

Atom-Type[a]	Description	Code
C_4	Tetravalent carbon	11
$C_3^{=C}$	Trivalent carbon, double bonded to carbon	12
$C_3^{=\mathrm{N,P}}$	Trivalent carbon, double bonded to nitrogen or phosphorus	13
$C_3^{=\mathrm{O,S}}$	Trivalent carbon, double bonded to oxygen or sulfur	14
$C_{1,2}$	Univalent or divalent carbon	15
C_{ar}	Aromatic carbon	16
$C_{\mathrm{ar}}^{\mathrm{lp}}$	Aromatic carbon bonded to an aromatic oxygen or nitrogen with a lone pair	17
$N_{2^-,3,4^+}$	Amine nitrogen	21
N_3^{deloc}	Trivalent nitrogen with a delocalized lone pair	22
N_3^{hdeloc}	Trivalent nitrogen with a highly delocalized lone pair	23
N_2	Neutral divalent nitrogen	24
$N_{1,2^+}$	Univalent or cationic divalent nitrogen	25
$O_{1,2}$	Univalent or divalent oxygen	31
$O_1^{\mathrm{ester,acid}}$	Double-bonded oxygen in an ester or acid	32
O_1^{lact}	Double-bonded oxygen in a lactone or lactam	33
$P_{2,3}$	Divalent or trivalent phosphorus	41
$P_{3,4}^{=}$	Trivalent or tetravalent double-bonded phosphorus	42
$S_{1,2}$	Univalent or divalent sulfur	51
S_3	Trivalent sulfur	52
S_4	Tetravalent sulfur	53
Si_4	Tetravalent silicon	61
F_1	Fluorine	71
Cl_1	Chlorine	72
Br_1	Bromine	73
I_1	Iodine	74
H_1	Hydrogen	91
	Single bond	01
	Double bond	02
	Triple bond	03
	Dative bond	06
	Aromatic single bond	07
	Aromatic double bond	08
	Single bond with charge or delocalized bond	09

[a]The first letter in the atom-type represents the element, the subscripts represent the valency, and the superscripts represent bonded environments (see text).

aromatic double bond orders were made equivalent. Reported here is only the final atom- and bond-typing scheme and the final training set of molecules, together with the derived final BCC parameter set.

Atom- and Bond-Typing

Atom- and bond-typing a molecule begins with assigning aromaticity to atoms using Figure 1a, starting with case 1 and progressing through case 5; bond types are then assigned using Figure 1b. An unusual feature of these rules concerns atoms in five-membered rings that are fused to six-membered aromatic rings. The uniquely five-membered ring bonds in these types of systems (e.g., the three atoms in the five-membered ring of indole) are not assigned aromatic bond types, although non-fused 5-membered heteroaromatic rings (e.g., pyrrole) are bond-typed as aromatic. This was necessary to achieve a good global fit. Once bond types are assigned, atom typing may then be performed.

Having started with elements subclassified only according to valency, further subclassification within the same valency was necessary for C, O, and N. In other cases, atom types of differing valencies could be combined. Carbon atom-types are thus C_4, $C_3^{=C}$, $C_3^{=N,P}$, $C_3^{=O,S}$, $C_{1,2}$, C_{ar}, and $C_{\mathrm{ar}}^{\mathrm{lp}}$, where the subscripts refer to the valency (e.g., a carbon atom in a molecule of methanol would be classified as C_4). The superscript, in general, refers to the bonded environment: $=C$ signifies a double bond to carbon, $=N,P$ signifies a double bond to nitrogen or phosphorus, and $=O,S$ signifies a double bond to oxygen or sulfur. $C_{1,2}$ refers to a mono- or divalent carbon (e.g., isonitriles and ketenes, respectively), and C_{ar} signifies an aromatic carbon atom. The superscript *lp* signifies that the atom is adjacent to an aromatic nitrogen or oxygen atom bearing a lone pair of electrons. Two aromatic carbon types were used because C_{ar} — X bond-types differed in their BCC value if they were adjacent to an aromatic atom containing a lone pair (i.e., N, O). For example, the BCC values for the C_{ar} — H bond-types in benzene and pyridine were 0.0128 and 0.1366, respectively. The

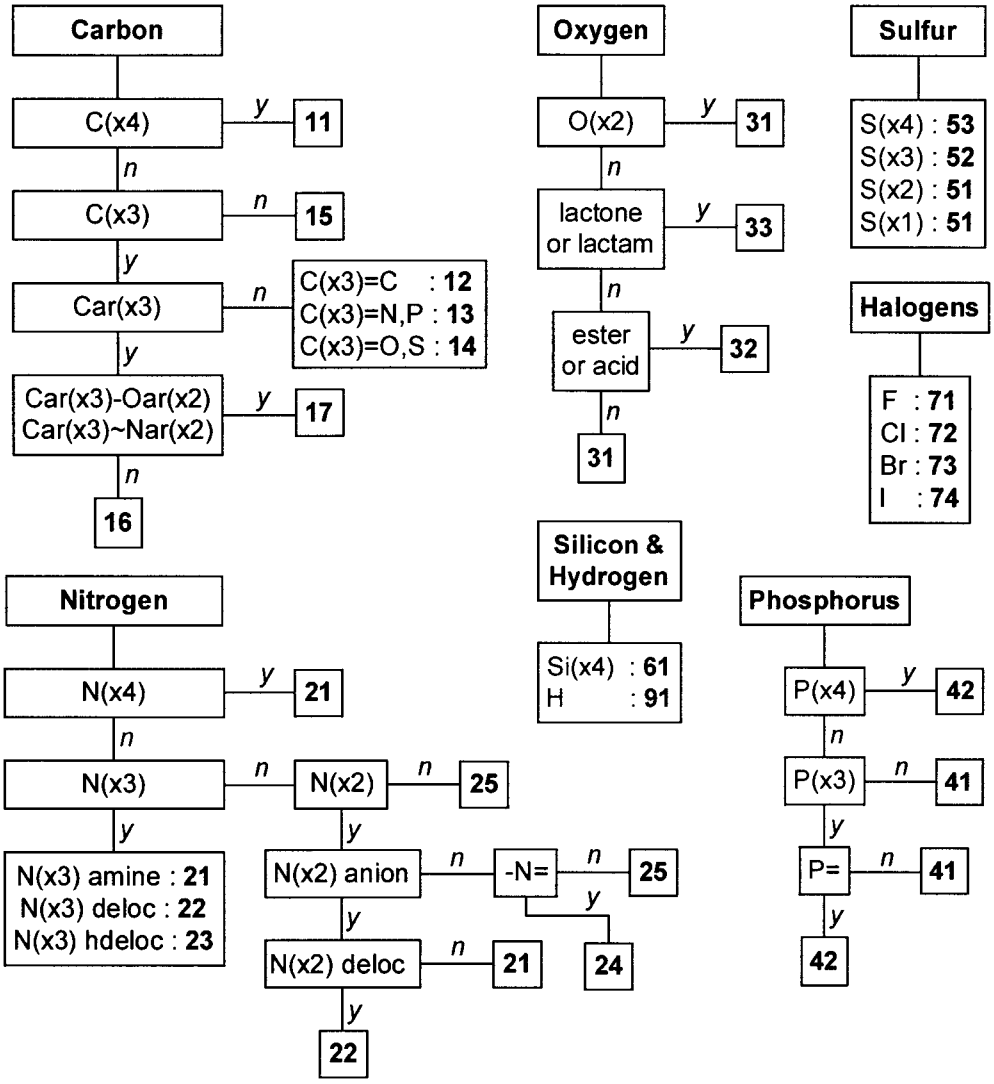
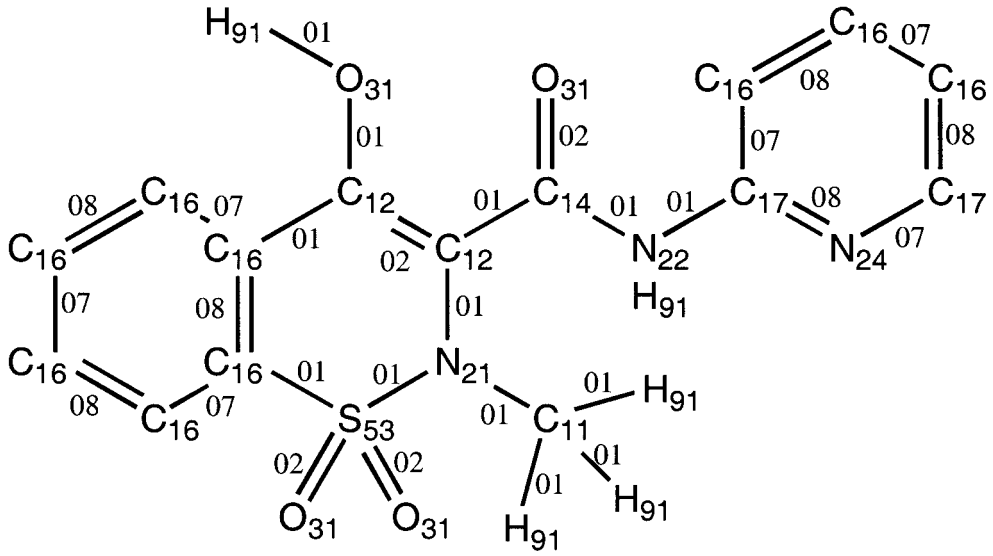


Figure 2. The atom-typing definitions. The atom-type of an atom in a molecule is determined by starting at the top of the tree for a given element and answering the questions in the boxes until the atom code is reached. Definitions: ar = aromatic; Oar = oxygen atom bearing lone pairs and in an aromatic ring; Nar = nitrogen atom bearing lone pairs and in an aromatic ring; deloc= delocalized lone pair, for example, amide nitrogen atoms; hdeloc= highly delocalized lone pair, for example, the nitrogen atom in indole, pyrrole, or nitro group.

adjacent hydrogen atom in pyridine is compensating for the lack of a point charge (or BCC-type in the present charge model) needed to describe the nitrogen lone pair in the plane of the molecule.[20] We term this phenomenon "atomic charge compensation" (ACC). In general, ACC occurs when a limited number of point charges are used to describe an ESP that varies quickly around a lone pair of electrons. In this work, we have allowed for ACC by differentiating two aromatic carbon atom-types: C_{ar} and C_{ar}^{lp}. Although aromatic carbons adjacent to aromatic sulfur and phosphorus should also suffer from ACC, it was found to be negligible for these atoms.

Nitrogen atom-types are $N_{2^-,3,4^+}$, N_3^{deloc}, N_3^{hdeloc}, N_2, and $N_{1,2^+}$, where the subscripts represent valency and charge (e.g., $N_{2^-,3,4^+}$ represents divalent anionic nitrogen, trivalent aliphatic nitrogen, or tetravalent cationic nitrogen). The superscripts *deloc* and *hdeloc* signify delocalized and highly delocalized, respectively. The subclassification of the N_3 into N_3^{deloc} and N_3^{hdeloc} was needed to distinguish between different levels of lone pair delocalization on conjugated nitrogen systems (e.g., to differentiate a delocalized amide nitrogen from a highly delocalized pyrrole nitrogen).

Oxygen atom-types are $O_{1,2}$, $O_1^{ester,acid}$, and O_1^{lact} where the superscript *ester,acid* denotes ester and acid functional groups, and *lact* signifies either a lactone or lactam. To get a good fit it was found necessary to separate carbonyl oxygen atom-types for the more stable trans form of esters and amides, found in most acyclic contexts, from the less-stable cis form usually encountered in



a

11 01 21	12 01 21	16 01 53
11 01 91	12 01 31	16 01 91
12 01 14	12 01 91	17 01 22
12 01 16	14 01 22	17 01 91
21 01 53	31 02 53	16 08 16
31 01 91	16 07 16	16 08 17
12 02 12	16 07 17	17 08 24
14 02 31	17 07 24	

b

Figure 3. (a) Atom typing of piroxicam, a nonsteroidal anti-inflammatory drug. The labels adjacent to the atoms represent the atom-types and the labels adjacent to the bonds represent the bond orders (see Figures 1 and 2 for atom and bond order definitions). (b) The BCC bond types needed to charge piroxicam. The first pair of digits represents atom-type *I*, the second pair represents the bond order, and the third pair represents atom-type *J*. The BCC values for each bond-type can be found in Table 2.

lactones and lactams. The difference might be caused by ACC or the different electronic characters of the delocalized lone pairs of electrons between the cis and trans conformations.

Phosphorus atom-types were classified simply as $P_{2,3}$ and $P^{=}_{3,4}$ where the superscript "=" signifies a double bond. Sulfur atom-types were classified as $S_{1,2}$, S_3, and S_4. It was not necessary to further divide the monovalent atoms (i.e., the halogens and hydrogen) because the subclassifications of the C, N, O, S, and P atom-types provided enough degrees of freedom to fit the BCCs to the data.

The final atom- and bond-type definitions are summarized in Table 1, with the last column giving the numeric codes used for the BCCs; there are a total of 26 atom-types and six unique bond orders. The atom-typing flowchart for each element is depicted in Figure 2. Once these are assigned to a molecule, the BCC-type for bonded atoms *i* and *j*, having atom-types *I* and *J*, respectively, is then constructed by combining the numeric codes for atom-type *I*, the bond order *O* between atoms *i* and *j*, and atom-type *J*, giving BCC bond-type *I-O-J*. This atom- and bond-typing scheme is robust and general and can be used to unambiguously assign atom-types in an automated fashion, for example, using PATTY,[21] a programmable atom-typer.

An example of this bond-typing process is shown for piroxicam (Fig. 3). The atoms in the two outer phenyl rings are assigned to be

Table 2. Table of Bond Codes, Number of Occurrences in the Training Set, and the BCC Values.

	Type	Occur.	BCC		Type	Occur.	BCC		Type	Occur.	BCC
1	110111	1876	0.0000	46	120174	4	0.2728	91	150121	10	0.0558
2	110112	262	0.0042	47	120191	1093	0.0000	92	150122	4	0.0062
3	110113	132	−0.0753	48	130113	26	0.0000	93	150123	5	−0.0543
4	110114	372	−0.0500	49	130114	26	0.0693	94	150124	4	0.0520
5	110115	103	−0.0269	50	130115	5	0.0302	95	150125	4	0.0798
6	110116	238	0.0073	51	130116	81	0.0839	96	150131	5	0.0286
7	110117	85	−0.0943	52	130117	15	−0.0297	97	150141	5	0.3987
8	110121	860	0.1582	53	130121	114	0.1522	98	150142	5	0.3545
9	110122	164	0.0374	54	130122	13	0.0069	99	150151	4	0.2251
10	110123	244	−0.0193	55	130123	68	0.0388	100	150152	5	0.4586
11	110124	117	0.1252	56	130124	27	0.2193	101	150153	5	0.4018
12	110125	12	0.1825	57	130125	4	0.2406	102	150161	7	0.2343
13	110131	750	0.0718	58	130131	48	0.1336	103	150171	4	0.0021
14	110141	96	0.2579	59	130141	6	0.3529	104	150172	4	0.0899
15	110142	131	0.4078	60	130142	5	0.5230	105	150173	4	0.2338
16	110151	173	0.1821	61	130151	41	0.2242	106	150174	4	0.2886
17	110152	82	0.4263	62	130152	5	0.4296	107	150191	32	0.0575
18	110153	110	0.4608	63	130153	7	0.5077	108	160116	35	0.0000
19	110161	117	0.1540	64	130171	7	0.1055	109	160117	44	−0.1075
20	110171	123	0.0713	65	130172	14	0.1123	110	160121	160	0.0816
21	110172	147	0.0734	66	130173	6	0.1253	111	160122	17	−0.0074
22	110173	36	0.1274	67	130174	4	0.2752	112	160123	140	−0.0452
23	110174	4	0.3009	68	130191	211	0.1292	113	160124	73	0.1394
24	110191	10885	0.0393	69	140114	21	0.0000	114	160125	5	0.1877
25	120112	116	0.0000	70	140115	4	−0.0308	115	160131	105	0.0451
26	120113	142	−0.0866	71	140116	86	0.0206	116	160141	23	0.3270
27	120114	187	−0.0192	72	140117	15	−0.0895	117	160142	20	0.3895
28	120115	51	0.0333	73	140122	352	0.0670	118	160151	48	0.2269
29	120116	62	0.0118	74	140123	12	−0.0432	119	160152	7	0.4285
30	120117	27	−0.1098	75	140124	29	0.1394	120	160153	51	0.4445
31	120121	77	0.0820	76	140125	5	0.0070	121	160161	5	0.1938
32	120122	24	0.0026	77	140131	208	0.0901	122	160171	7	0.0349
33	120123	42	−0.0271	78	140141	8	0.2828	123	160172	24	0.0784
34	120124	49	0.1406	79	140142	11	0.4820	124	160173	4	0.1401
35	120125	5	0.4897	80	140151	39	0.1826	125	160174	4	0.2859
36	120131	60	0.0441	81	140152	5	0.4319	126	160191	4215	0.0000
37	120141	4	0.2986	82	140153	8	0.4840	127	170117	4	0.0000
38	120142	17	0.4154	83	140171	6	0.1115	128	170121	160	0.1929
39	120151	48	0.2406	84	140172	8	0.0486	129	170122	7	0.0799
40	120152	11	0.4307	85	140173	6	0.1106	130	170123	35	0.0635
41	120153	27	0.4509	86	140174	4	0.2601	131	170124	25	0.2190
42	120161	5	0.2128	87	140191	60	0.0928	132	170125	5	0.2100
43	120171	9	0.0625	88	150115	10	0.0000	133	170131	19	0.1703
44	120172	38	0.0844	89	150116	16	0.0040	134	170141	5	0.4597
45	120173	10	0.1257	90	150117	6	−0.1070	135	170142	4	0.4543
136	170151	19	0.3273	181	230153	5	0.3418	226	420171	6	−0.1976
137	170152	5	0.5416	182	230161	5	0.1766	227	420172	15	−0.3077
138	170153	4	0.5169	183	230171	4	0.0561	228	420173	4	−0.2425
139	170171	9	0.1402	184	230172	7	0.0642	229	420191	25	−0.3756
140	170172	8	0.1660	185	230173	4	0.1802	230	510151	15	0.0000
141	170173	16	0.2312	186	230191	362	−0.0497	231	510152	5	0.2576
142	170174	4	0.3528	187	240124	13	0.0000	232	510153	4	0.2971
143	170191	991	0.1369	188	240125	4	0.1043	233	510161	4	−0.0936
144	210121	31	0.0000	189	240131	34	−0.0985	234	510171	4	−0.0547
145	210122	16	−0.0302	190	240141	6	0.1287	235	510172	4	−0.0260
146	210123	52	−0.1185	191	240142	17	0.1498	236	510173	4	0.1105
147	210124	16	0.0867	192	240151	19	0.0315	237	510191	36	−0.1718
148	210131	48	−0.0517	193	240152	8	0.1731	238	520152	1	0.0000

(*continued*)

Table 2 *(Continued)*

	Type	Occur.	BCC		Type	Occur.	BCC		Type	Occur.	BCC
149	210141	13	0.2527	194	240153	14	0.2714	239	520172	5	−0.2350
150	210142	24	0.2947	195	240161	4	0.1150	240	520191	10	−0.4040
151	210151	7	0.0757	196	240171	4	−0.0554	241	530171	7	−0.2041
152	210152	11	0.3531	197	240172	9	−0.0603	242	530172	9	−0.2761
153	210153	82	0.3359	198	240173	5	0.0192	243	530173	4	−0.1970
154	210161	10	0.1037	199	240174	4	0.1171	244	530174	4	−0.1490
155	210171	5	−0.0204	200	240191	72	−0.2444	245	530191	6	−0.4588
156	210172	4	−0.0369	201	250191	4	−0.1634	246	610172	20	−0.0555
157	210173	6	0.0876	202	310131	6	0.0000	247	610191	45	−0.0258
158	210174	4	0.1898	203	310141	18	0.2229	248	120212	613	0.0000
159	210191	878	−0.2048	204	310142	196	0.2336	249	120215	23	−0.0631
160	220122	5	0.0000	205	310151	5	0.0442	250	130223	39	0.0547
161	220123	4	−0.0327	206	310152	9	0.2152	251	130224	432	0.2877
162	220124	11	0.1326	207	310153	41	0.2093	252	130225	13	0.1624
163	220125	4	0.0235	208	310161	32	0.0083	253	130241	27	0.2496
164	220131	10	−0.0059	209	310171	4	0.0375	254	130242	10	0.6836
165	220141	9	0.1682	210	310172	4	0.0014	255	140231	301	0.2391
166	220142	10	0.3046	211	310173	4	0.0699	256	140232	115	0.1890
167	220151	4	0.1593	212	310174	4	0.1712	257	140233	217	0.2755
168	220152	4	0.3657	213	310191	386	−0.2010	258	140251	88	0.2834
169	220153	12	0.3322	214	410141	4	0.0000	259	140252	4	0.4447
170	220171	4	−0.0155	215	410142	6	0.2762	260	140253	4	0.5617
171	220172	4	−0.0015	216	410151	5	−0.1072	261	150215	6	0.0000
172	220173	6	0.0838	217	410161	6	0.0592	262	150224	13	0.2204
173	220174	4	0.2067	218	410171	6	−0.2769	263	150231	14	0.2335
174	220191	213	−0.0865	219	410172	8	−0.2065	264	150242	4	0.7150
175	230123	6	0.0000	220	410173	4	−0.1309	265	150251	6	0.3429
176	230124	31	0.1794	221	410174	4	0.0181	266	150252	4	0.3052
177	230131	11	0.0202	222	410191	49	−0.1593	267	230224	5	0.1731
178	230141	6	0.2538	223	420142	1	0.0000	268	240224	47	0.0000
179	230142	5	0.4175	224	420151	10	−0.3626	269	240225	12	0.0040
180	230151	5	0.2024	225	420152	4	0.0675	270	240231	12	−0.0467
271	240241	7	0.0535	299	160751	178	0.2699	327	170824	1003	0.2630
272	240242	22	0.4287	300	170717	97	0.0000	328	170831	4	0.1294
273	240251	6	0.1684	301	170723	244	0.0885	329	170841	6	0.2464
274	240252	4	0.3764	302	170724	813	0.2630	330	170851	4	0.3144
275	240253	5	0.4265	303	170731	265	0.1294	331	230824	4	0.1698
276	250225	25	0.0000	304	170741	4	0.2464	332	240824	226	0.0000
277	310241	4	0.1315	305	170751	126	0.3144	333	240841	8	0.0456
278	310242	95	0.2707	306	230723	4	0.0000	334	110931	5	0.1615
279	310251	5	0.1583	307	230724	227	0.1698	335	110951	4	0.0833
280	310252	74	0.2718	308	230731	4	0.0677	336	120931	7	0.1844
281	310253	385	0.2792	309	230741	8	0.1806	337	120951	5	0.1879
282	410241	1	0.0000	310	230751	4	0.1565	338	130931	5	0.3046
283	420251	14	−0.5630	311	240724	87	0.0000	339	130951	4	0.1969
284	510252	7	0.3295	312	240731	121	−0.0723	340	140931	56	0.2653
285	150315	88	0.0000	313	240741	5	0.0456	341	140951	23	0.1713
286	150325	160	0.3258	314	240751	131	0.0419	342	150931	4	0.2549
287	250325	17	0.0000	315	310741	4	0.0710	343	150951	4	0.3281
288	210631	6	−0.1168	316	310751	4	0.0983	344	160931	4	0.2797
289	220631	4	0.0465	317	410751	4	0.0587	345	160951	4	0.2541
290	230631	7	0.1317	318	510751	3	0.0000	346	170931	4	0.3370
291	240631	4	−0.0817	319	160816	2551	0.0000	347	170951	4	0.3476
292	250631	4	−0.0543	320	160817	504	−0.0653	348	230931	245	−0.1500
293	250651	4	0.1473	321	160823	47	−0.0243	349	310941	7	0.2354
294	160716	2351	0.0000	322	160831	4	−0.1518	350	310942	108	0.3195
295	160717	413	−0.0653	323	160841	14	0.2240	351	310951	4	0.1530
296	160723	535	−0.0243	324	160851	7	0.2699	352	310952	8	0.3916
297	160731	4	−0.1518	325	170817	129	0.0000	353	310953	30	0.3228
298	160741	11	0.2240	326	170823	11	0.0885	354	510953	4	0.5218

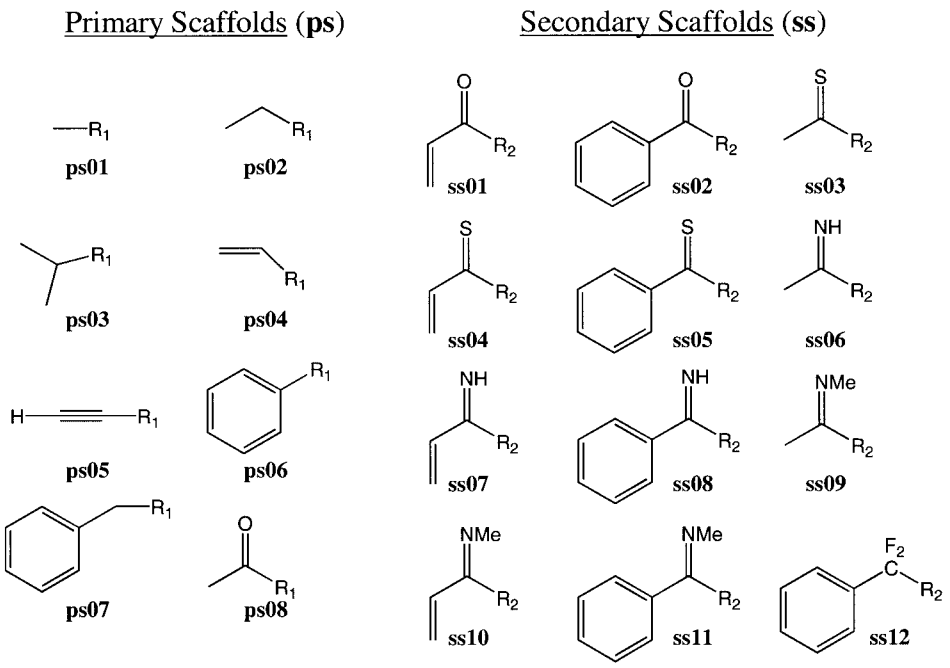


Figure 4. Primary and secondary scaffolds used in assembling the training set.

aromatic, while the remaining four atoms in the middle 6-membered ring and all other atoms in the molecule are assigned to be nonaromatic according to Figure 1. The aromatic bond-order codes (07 and 08 for aromatic single and double, respectively) are therefore assigned to the two outer aromatic rings while nonaromatic bond-order codes (01 and 02 for single and double, respectively) are assigned to the nonaromatic bonds. The atom-types are then determined using Figure 2. Using the amide carbon atom in piroxicam (Fig. 3) as an example, we begin at the top of the "carbon" tree in Figure 2 and descend by answering the questions in the boxes. Is the carbon tetravalent? No. Is it trivalent? Yes. Is it aromatic and trivalent? No. Is its double-bonded partner O or S? Yes, so the amide carbon atom-type is *C(x3)═O,S*, coded **14**. This process is repeated for all the atoms in the molecule. The atom-types and bond order codes are shown for piroxicam in Figure 3a and the resulting BCC-types are shown in Figure 3b. The BCC values for each bond-type in Figure 3b can be found in Table 2.

Training Set

The training set was composed of 2,755 organic molecules that sample organic bond-types composed of H, C, N, O, F, Si, P, S, Cl, Br, and I atoms. Only limited sampling of Si bonds was made because they were not of primary interest; no effort was made to ensure that all Si bond-types were present in the training set. In order to improve the robustness of the BCCs, each bond-type was sampled across a minimum of four different chemical contexts.

To ensure a diverse and systematic sampling of common organic bond-types, a scaffold/functional group scheme was developed. A set of molecular scaffolds (Fig. 4) was created onto which were attached a variety of functional groups (Fig. 5), creating an array of chemical functionalities. Primary (pf) and secondary (sf) functional groups were added to primary scaffolds (ps). For example, functional group pf01 is added to scaffold ps08 to build a molecule of acetone, functional group pf07 is added to scaffold ps06 to build a molecule of phenol, and so on. Similarly, a set of secondary scaffolds (ss) (Fig. 4) was used with the primary functional groups (Fig. 5) to extended the range of functionalities. This scaffold/functional group scheme generated a total of 390 molecules after elimination of a few redundancies (e.g., ps06_pf01 and ps07_pf02).

Although the above scheme sampled many bond-types, additional molecules were needed to sample cyclic and fused bicyclic heteroaryl bond-types. The building schemes for these molecules are shown in Figure 6, which generated a set of 723 molecules. To further improve sampling and extend the variety of bond-types, the MMFF94 validation suite of 761 molecules was obtained from the Computational Chemistry List[17] (CCL) and incorporated into the training set. Additionally, to complete the training set during the iterative modifications of the atom-typing scheme, approximately 900 extra molecules were created manually. In retrospect, a much more complete scaffold/functional-group approach would have reduced this large latter category.

BCCs

With the above atom- and bond-typing schemes (i.e., Figs. 1 and 2), The Merck Index and the NCI database showed that only 354 BCCs would cover the desired chemical variability. Of these, 29 were zero by atom-type symmetry (e.g., C_4— C_4, bond code 11 01

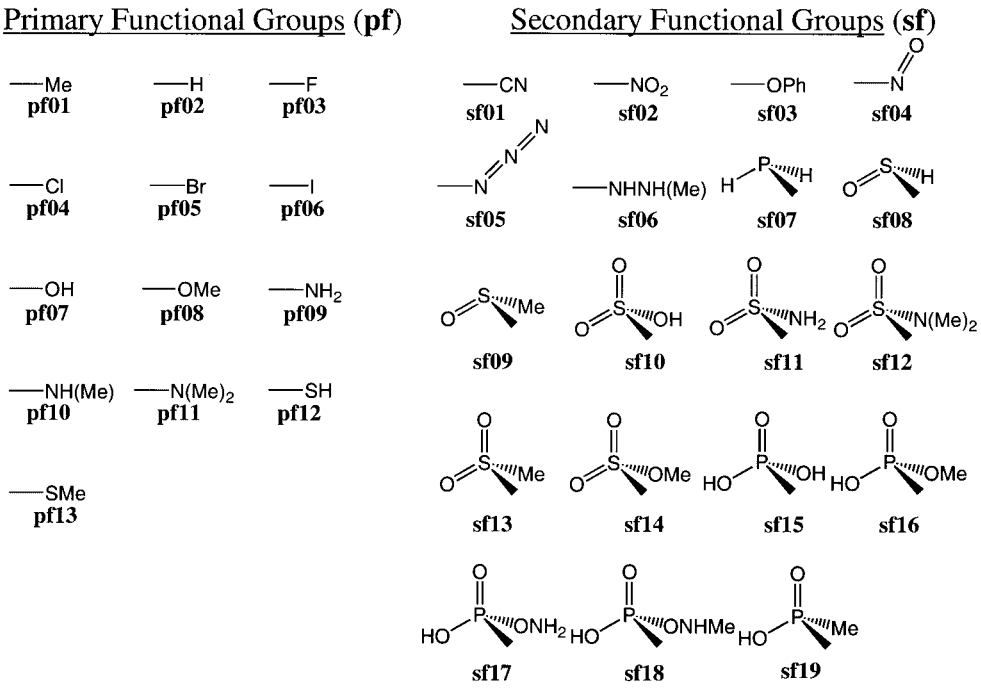


Figure 5. Primary and secondary functional groups used in assembling the training set.

11), and 16 of them were made equivalent to ensure symmetry in aromatic molecules (e.g., C_{ar}^{lp} *arom* N_2 and C_{ar}^{lp} *arom* N_2, bond codes 17 07 24 and 17 08 24, respectively). This left 309 unique BCCs to be fit to the 44,194 bonds in the training set of 2755 molecules. These BCCs were fit to the ESP of the training set using the methodology given in ref. 4, yielding a relative RMS (c.f. ref. 4) of 0.50. The Pearson correlation matrix for the 354 BCCs was calculated to evaluate interdependencies between them. Of the 62,500 nonequivalent BCC pairs all but 15 pairs showed a correlation of less than 0.6, indicating little correlation between most parameters. Of the 15 remaining BCC pairs, 12 were between 0.6 and 0.7, and the remaining three fell in the range of 0.7 to 0.8. These correlated BCCs were between bonds from H, C, O, and N to a common central atom, where that central atom was tetravalent silicon, hypervalent sulfur, or hypervalent phosphorus. The latter two cases were represented in the training set by sulfones, sulfonamides, sulfonyl esters, and the corresponding phosphorus species. These tetrahedral species tend to have four bulky and often similar substituents attached, which may lead to the correlation. The correlation matrix confirmed that, in general, the BCCs were independent from each other.

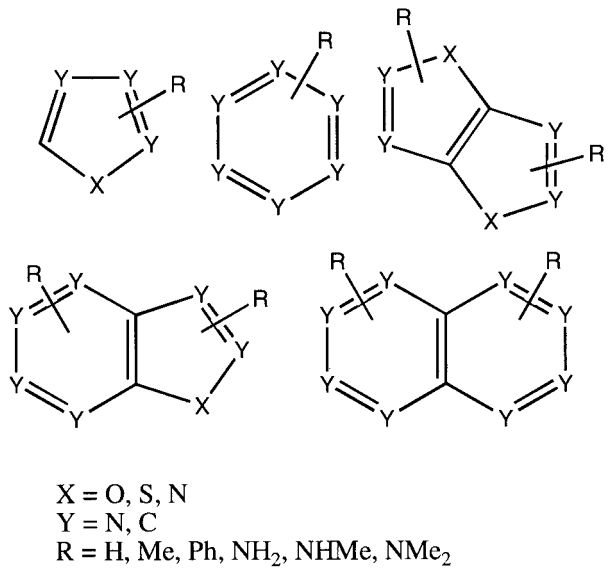


Figure 6. Construction scheme of cyclic and fused bicyclic heteroaryl systems.

Subsequent to the fit, five BCCs were adjusted *a posteriori* to improve the relative free energies of solvation (RFES) of amines, nitros, and unsaturated hydrocarbons. These adjustments will be discussed in the next section. The BCC-types parameterized in this work are listed in Table 2, which presents the bond-types, the BCC values, and the number of times they occur in the training set. These BCCs make it possible to charge any molecule in either The Merck Index or the NCI small-molecule database, except for molecules containing boron or molecules containing covalently bound metal atoms.

Validation of the AM1-BCC Charge Model

The validation examined the ability of AM1-BCC charges, in the context of a FF, to reproduce experimental or high-level *ab initio*

data of nonbonded interactions between organic molecules. AM1-BCC charges were used in the Cornell et al.[13] FF to test the ability to reproduce homo- and hetero-dimer energies of a set of diverse small organic molecules and experimental RFES of small organic molecules. Conformational energies were not examined here because the 1,4 electrostatic term is only a component of the torsional energy in a FF, and not necessarily a dominant one; conformational energies reflect the combined overall parameterization of the FF, in particular the explicit torsional term.

Dimer Energies

The hydrogen-bonded energies of diverse but simple organic homo- and hetero-dimers (see Fig. 7 for representative structures) were compared between *ab initio* calculated values (HF/6-31G* and MP2/6-31+G**, referred to as HF and MP2, respectively, hereafter) and RESP, AM1, MMFF, and AM1-BCC charge models (see Table 3). For comparison purposes, in Table 3 after the dimer energy for each charge model is given, the difference with the two *ab initio* values, for example, after each RESP value, is given $\Delta HF = E(HF) - E(RESP)$ and $\Delta MP2 = E(MP2) - E(RESP)$. As expected, the AM1 population charges alone showed poor performance here and throughout the validation, but they served as a means of comparison for the improvement introduced with the BCCs. The MMFF charge model performed best overall with respect to MP2 energies, with a mean unsigned error of 0.74 kcal/mol. Interestingly, the MMFF charge model was parameterized for a "buffered" vdW and electrostatic functional form,[22] not the AMBER electrostatic functional form[13] in which it was used for this work. However, the good performance of MMFF must also be interpreted in light of the fact that the dimer energies of dimers **1** through **40** were part of the dimer training set[23] of MMFF.

The AM1-BCC model has the same mean unsigned error (i.e., 0.93 kcal/mol) with respect to both HF and MP2 (see Table 3). However, examination of the mean error reveals that AM1-BCC tends to overestimate the binding energy with respect to HF (0.53 kcal/mol) and underestimate it with respect to MP2 (−0.36 kcal/mol), also observed with RESP (0.61 and −0.28 kcal/mol, respectively). Conversely, the AM1 charge model underestimates the binding energy (−1.98 and −2.87 kcal/mol) and the MMFF charge model overestimates it (1.24 and 0.35 kcal/mol) with respect to both HF and MP2, respectively. When compared to the HF dimer energies, the AM1-BCC charge model performs best overall with a mean unsigned error of 0.93 kcal/mol, while RESP has a mean unsigned error of 1.01 kcal/mol. However, compared to MP2, RESP slightly outperforms AM1-BCC (0.88 and 0.93 kcal/mol). Overall, the RESP and AM1-BCC charge models perform similarly, producing dimer energies that tend to fall in between HF and MP2 values. This behavior is consistent with the results found in earlier work.[4]

The DNA hydrogen-bonding dimer calculations were based on the extensive work of Šponer et al.[15] As well as the biological importance of these dimers, the DNA bases are also highly delocalized polar molecules and thus they served as a challenging arena to compare the performance of the AM1-BCC charge model against other models. Representative structures for the DNA dimers taken from ref. 15 are shown in Figure 8, while Table 4 compares the dimer energies of the RESP, AM1, MMFF, and AM1-BCC charge models to those taken from ref. 15, calculated at *ab initio* HF/6-31G** and MP2/6-31G*(.25) levels.

Again, the AM1 charge model performed very poorly, with large mean unsigned errors of 4.8 and 6.2 kcal/mol for HF and MP2 energies, respectively; this model will not be considered further. For the MP2 energies, the mean error shows that, on average, RESP and MMFF slightly underestimated the hydrogen-bond strengths while AM1-BCC slightly overestimated them. Compared to HF, MMFF performed the best; all three charge models overestimated the dimer strengths. The MP2 mean unsigned error for both RESP and AM1-BCC was <1 kcal/mol, consistent with their behavior in the organic homo- and hetero-dimers (Table 3). This demonstrates the ability of these two charge models to account for the delocalization and dense functionalization in the DNA bases. In contrast, while MMFF had the smallest mean unsigned error in the organic homo- and hetero-dimers (0.74 kcal/mol, see Table 3), this error more than doubled in the DNA dimer energies (1.7 kcal/mol), which were not included in the MMFF parameterization set. All charge models reproduced the correct ordering of the AT base pairs, that is, the AT Hoogsteen base pair is stronger than the Watson-Crick AT base pair.[24]

Densely Functionalized Molecules

A crucial requirement for a charge model in the pharmaceutical industry is that it properly accounts for the electrostatics of densely functionalized molecules. A large proportion of the molecules in The Merck Index, a wide representation of drug and druglike molecules, contain multiple functional groups crowded close together. In this section, we compare the ability of RESP, AM1, MMFF, and AM1-BCC to correctly capture the electrostatics (*vis-a-vis* the QM ESP) of four neutral molecules, shown in Figure 9, having varying degrees of functionalization: ethanol, a simple monofunctionalized molecule; homarine, a zwitterion; *p*-methoxy benzenesulfonate, having both electron-donating and -withdrawing groups; and the multi- and densely functionalized cation N-methyl-N′-cyano-N-nitrosoguanidinium (a variant of the carcinogen N-methyl-N′-nitro-N-nitrosoguanine).

Table 5 shows the RMS deviation of the four charge models with respect to the QM ESP for the molecules in Figure 9. The RESP, AM1, and AM1-BCC charge models perform as expected: poorly performing AM1 atomic charges are much improved by the addition of the BCCs, while RESP performs best overall. For the ethanol molecule, with a single functional group, MMFF performs moderately with an RMS of 0.0408 a.u. However, for the three molecules with more than one functional group, the quality of the MMFF fit greatly deteriorates, yielding RMS deviations of >0.5, significantly higher than even AM1. Even though MMFF was parameterized within the context of a FF having a different functional form[23] than the one in which it was applied here, these large deviations are nevertheless worrisome, possibly resulting from the neglect of explicit delocalized electron-donating and -withdrawing in that model.

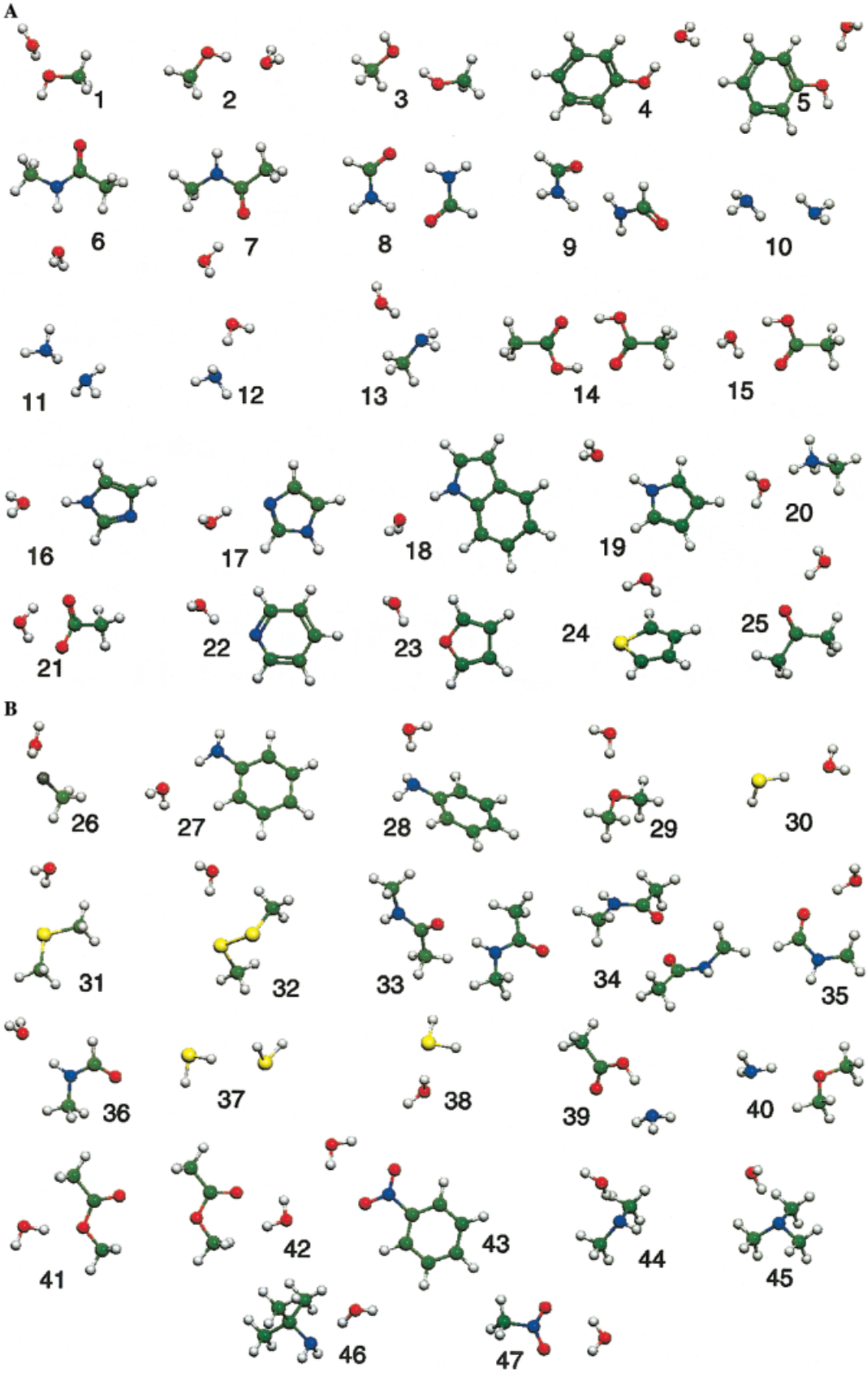


Figure 7. Representative structures for the hydrogen-bonded organic homo- and hetero-dimers. A description of the dimers can be found in the first column of Table 3.

Table 3. Comparison of Hydrogen-Bonded Organic Homo- and Hetero-Dimer Energies[a] between Various Charge Models and HF/6-31G* and MP2/6-31+G**.

Dimer	HF/6-31G*	MP2/6-31+G**	RESP	ΔHF[b]	ΔMP2	AM1	ΔHF	ΔMP2	MMFF	ΔHF	ΔMP2	AM1-BCC	ΔHF	ΔMP2
HOH . . . $OHCH_3$ **1**	−5.56	−6.80	−6.33	0.77	−0.47	−3.39	−2.17	−3.41	−6.44	0.88	−0.36	−5.70	0.14	−1.10
CH_3OH . . . OH_2 **2**	−5.59	−6.21	−6.90	1.31	0.69	−2.41	−3.18	−3.80	−6.36	0.77	0.15	−6.57	0.98	0.36
CH_3OH . . . $OHCH_3$ **3**	−5.54	−6.89	−6.67	1.13	−0.22	−1.90	−3.64	−4.99	−6.32	0.78	−0.57	−5.82	0.28	−1.07
C_6H_5OH . . . OH_2 **4**	−7.37	−8.61	−8.41	1.04	−0.20	−3.95	−3.42	−4.66	−8.99	1.62	0.38	−8.19	0.82	−0.42
HOH . . . OHC_6H_5 **5**	−4.71	−5.69	−5.54	0.83	−0.15	−4.59	−0.12	−1.10	−5.49	0.78	−0.20	−4.98	0.27	−0.71
T-NMA . . . OH_2 **6**	−5.43	−6.65	−6.60	1.17	−0.05	−3.90	−1.53	−2.75	−7.09	1.66	0.44	−6.25	0.82	−0.40
HOH . . . T-NMA **7**	−7.30	−7.97	−8.59	1.29	0.62	−6.43	−0.87	−1.54	−8.99	1.69	1.02	−9.29	1.99	1.32
Formamide, cyclic **8**	−13.44	−14.36	−14.10	0.66	−0.26	−5.80	−7.64	−8.56	−12.73	−0.71	−1.63	−11.27	−2.17	−3.09
Formamide, 1-HB **9**	−7.38	−7.86	−7.80	0.42	−0.06	−4.12	−3.26	−3.74	−8.1	0.72	0.24	−7.21	−0.17	−0.65
H_2NH . . . NH_3, bifurcated **10**	−3.19	−3.65	−4.43	1.24	0.78	−0.57	−2.62	−3.08	−3.9	0.71	0.25	−3.30	0.11	−0.35
H_2NH . . . NH_3, linear **11**	−3.07	−3.68	−4.47	1.40	0.79	−0.58	−2.49	−3.10	−3.95	0.88	0.27	−3.35	0.28	−0.33
HOH . . . NH_3 **12**	−6.56	−7.65	−8.01	1.45	0.36	−2.08	−4.48	−5.57	−7.48	0.92	−0.17	−6.81	0.25	−0.84
HOH . . . NH_2CH_3 **13**	−6.53	−8.00	−7.89	1.36	−0.11	−2.92	−3.61	−5.08	−7.76	1.23	−0.24	−7.25	0.72	−0.75
H_3CCOOH . . . $HOOCCH_3$ **14**	−15.55	−16.10	−15.96	0.41	−0.14	−5.65	−9.90	−10.45	−17.7	2.15	1.60	−14.70	−0.85	−1.40
H_2CCOOH . . . OHH **15**	−10.66	−10.75	−10.55	−0.11	−0.20	−5.93	−4.73	−4.82	−11.73	1.07	0.98	−10.23	−0.43	−0.52
Imidazole . . . OH_2 **16**	−6.37	−7.88	−7.13	0.76	−0.75	−4.65	−1.72	−3.23	−8.05	1.68	0.17	−6.47	0.10	−1.41
HOH . . . imidazole **17**	−7.06	−7.82	−6.52	−0.54	−1.30	−5.38	−1.68	−2.44	−8.13	1.07	0.31	−7.54	0.48	−0.28
Indole . . . OH_2 **18**	−5.75	−7.42	−6.90	1.15	−0.52	−4.53	−1.22	−2.89	−6.68	0.93	−0.74	−6.09	0.34	−1.33
Pyrrole . . . OH_2 **19**	−5.36	−7.11	−6.41	1.05	−0.70	−4.29	−1.07	−2.82	−6.42	1.06	−0.69	−5.67	0.31	−1.44
$CH_3NH_3^+$ . . . OH_2 **20**	−19.30	−19.80	−18.31	−0.99	−1.49	−17.48	−1.82	−2.32	−19.49	0.19	−0.31	−19.53	0.23	−0.27
OHH . . . $(-)O_2CCH_3$ **21**	−21.85	−20.74	−24.34	2.49	3.60	−21.32	−0.53	0.58	−26.15	4.30	5.41	−24.74	2.89	4.00
HOH . . . pyridine **22**	−6.03	−7.28	−6.13	0.10	−1.15	−4.72	−1.31	−2.56	−7.32	1.29	0.04	−6.33	0.30	−0.95
HOH . . . furan **23**	−3.64	−4.10	−4.62	0.98	0.52	−4.50	0.86	0.40	−4.45	0.81	0.35	−4.04	0.40	−0.06
HOH . . . thiophene **24**	−2.42	−3.54	−4.48	2.06	0.94	−6.48	4.06	2.94	−3.91	1.49	0.37	−4.13	1.71	0.59
HOH . . . acetone **25**	−6.25	−6.57	−6.76	0.51	0.19	−4.44	−1.81	−2.13	−7.56	1.31	0.99	−6.67	0.42	0.10
HOH . . . FCH_3 **26**	−4.42	−4.30	−3.32	−1.10	−0.98	−2.55	−1.87	−1.75	−3.94	−0.48	−0.36	−3.03	−1.39	−1.27
Aniline . . . OHH **27**	−4.20	−4.04	−4.76	0.56	0.72	−2.81	−1.39	−1.23	−4.78	0.58	0.74	−5.20	1.00	1.16
HOH . . . aniline **28**	−5.17	−5.02	−5.62	0.45	0.60	−4.82	−0.35	−0.20	−6.97	1.80	1.95	−5.97	0.80	0.95
HOH . . . $O(CH3)_2$ **29**	−5.31	−6.72	−4.68	−0.63	−2.04	−4.17	−1.14	−2.55	−6.42	1.11	−0.30	−5.14	−0.17	−1.58
SHH . . . OHH **30**	−2.66	−3.66	−3.07	0.41	−0.59	−0.98	−1.68	−2.68	−3.57	0.91	−0.09	−4.47	1.81	0.81
HOH . . . $S(CH_3)_2$ **31**	−3.24	−3.05	−3.51	0.27	0.46	−1.39	−1.85	−1.66	−4.36	1.12	1.31	−3.50	0.26	0.45
OHH . . . CH_3SSCH_3 cyclic **32**	−3.44	−4.11	−3.31	−0.13	−0.80	−1.87	−1.57	−2.24	−3.97	0.53	−0.14	−3.57	0.13	−0.54
T-NMA dimer, parallel **33**	−7.00	−8.95	−9.65	2.65	0.70	−5.57	−1.43	−3.38	−10.43	3.43	1.48	−8.74	1.74	−0.21
T-NMA antiparallel stacked **34**	−4.92	−7.46	−7.75	2.83	0.29	−5.57	0.65	−1.89	−9.54	4.62	2.08	−7.69	2.77	0.23
HOH . . . N-Me formamide **35**	−7.18	−7.65	−8.30	1.12	0.65	−6.13	−1.05	−1.52	−8.89	1.71	1.24	−8.97	1.79	1.32
N-Me formamide . . . OH_2 **36**	−5.50	−6.80	−6.25	0.75	−0.55	−4.00	−1.50	−2.80	−6.93	1.43	0.13	−6.31	0.81	−0.49
HSH . . . SH_2 **37**	−0.88	−1.57	−1.32	0.44	−0.25	−0.41	−0.47	−1.16	−1.66	0.78	0.09	−2.43	1.55	0.86
HOH . . . SH_2 **38**	−2.06	−2.88	−2.16	0.10	−0.72	−0.71	−1.35	−2.17	−2.51	0.45	−0.37	−3.15	1.09	0.27
CH_3COOH . . . NH_3, bidentate **39**	−10.85	−12.11	−11.73	0.88	−0.38	−2.02	−8.83	−10.09	−12.55	1.70	0.44	−9.65	−1.20	−2.46
H_2NH . . . $O(CH_3)_2$ **40**	−2.84	−4.10	−3.07	0.23	−1.03	−1.25	−1.59	−2.85	−3.79	0.95	−0.31	−2.97	0.13	−1.13
HOH . . . Me acetate (—O—) **41**	−3.29	−4.82	−3.60	0.31	−1.22	−3.72	0.43	−1.10	−3.03	−0.26	−1.79	−4.13	0.84	−0.69
HOH . . . Me acetate (O═) **42**	−6.02	−6.59	−7.89	1.87	1.30	−5.67	−0.35	−0.92	−7.6	1.58	1.01	−7.62	1.60	1.03
OHH . . . nitrobenzene **43**	−5.04	−5.28	−5.83	0.79	0.55	−5.21	0.17	−0.07	−5.95	0.91	0.67	−3.57	−1.47	−1.71
OHH . . . nitromethane **44**	−4.86	−4.95	−5.39	0.53	0.44	−4.96	0.10	0.01	−6.29	1.43	1.34	−3.43	−1.43	−1.52
HOH . . . dimethyl amine **45**	−6.19	−8.05	−6.63	0.44	−1.42	−3.72	−2.47	−4.33	−8.01	1.82	−0.04	−7.80	1.61	−0.25
HOH . . . trimethylamine **46**	−5.90	−8.46	−3.74	−2.16	−4.72	−4.16	−1.74	−4.30	−7.99	2.09	−0.47	−8.07	2.17	−0.39
HOH . . . t-butylamine **47**	−7.08	−8.32	−3.42	−3.66	−4.90	−3.42	−3.66	−4.90	−7.96	0.88	−0.36	−7.54	0.46	−0.78
Mean error				0.61	−0.28		−1.98	−2.87		1.24	0.35		0.53	−0.36
Mean unsigned error				1.01	0.88		2.24	3.04		1.30	0.74		0.93	0.93
Maximum				3.66	4.90		9.9	10.45		4.62	5.41		2.89	4.00
Standard deviation				1.12	1.33		2.36	2.47		0.97	1.10		1.07	1.15

[a]All energies are in units of kcal/mol.

[b]ΔHF is associated with the dimer energy value for the preceding charge model and evaluates the difference compared to the HF value, for example, ΔHF after the RESP value means $E(\mathrm{HF}) - E(\mathrm{RESP})$ similar to ΔMP2 compared to the MP2 dimer energy value.

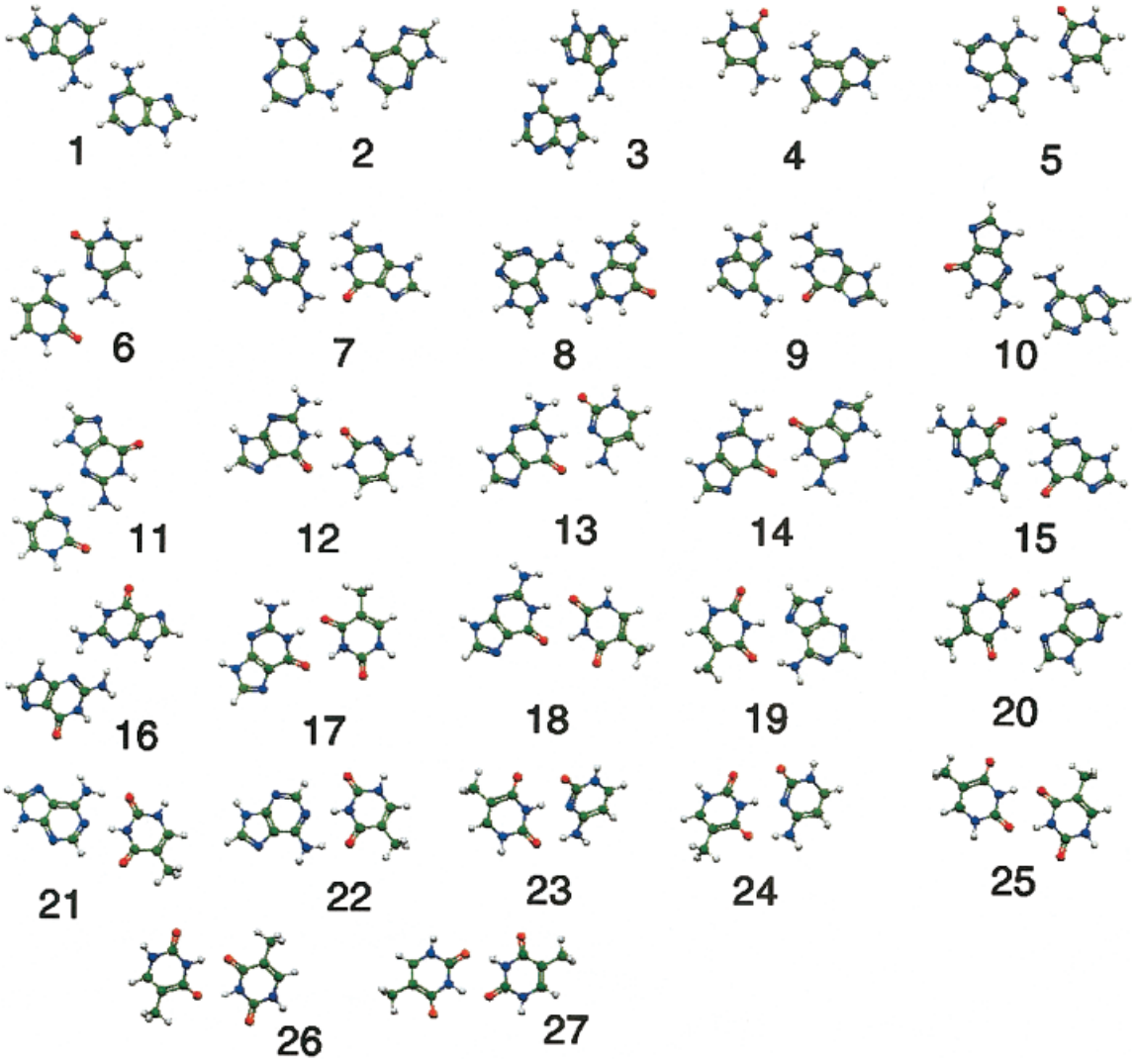


Figure 8. Representative structures for the DNA dimers as taken from ref. 15. (1) AA1, (2) AA2, (3) AA3, (4) AC1, (5) AC2, (6) CC, (7) GA1, (8) GA2, (9) GA3, (10) GA4, (11) GC1, (12) GCNEW, (13) GCWC, (14) GG1, (15) GG3, (16) GG4, (17) GT1, (18) GT2, (19) TAH, (20) TARH, (21) TARWC, (22) TAWC, (23) TC1, (24) TC2, (25) TT1, (26) TT2, (27) TT3.

Relative Free Energies of Solvation

The calculation of the relative free energies of solvation (RFES) is a critical test of the ability of a FF to correctly capture the dynamic noncovalent interactions between solute and solvent. Unlike the dimer energies, where comparisons were made to quantum mechanical data, the RFES allow a direct comparison to experimental data and therefore provide the most important comparisons. Also, the RFES reflect an ensemble average sampling of many configurations of solvent around the entire solute, as opposed to a single energy-minimized configuration as with the dimers. The RFES in this work were chosen to be between systems as isosteric as possible, but of very different polarity. This places the burden on the electrostatic term for the RFES and thus focuses on the charge model, minimizing the vdW contributions. Because a two-body additive FF has difficulty reproducing the free energy of charging for solutes bearing a formal charge,[25] ions were not included in these calculations. An explicit polarization term appears to be required to accurately reproduce the free energy of solvation of ions.[26] Table 6 shows a comparison of RFES for organic molecules using the same four charge models; the experimental values used came from refs. 27 and 28.

In general, RESP, AM1, and AM1-BCC underestimate the RFES while MMFF overestimates them, as shown by the mean errors. The mean unsigned errors show that AM1-BCC reproduced the experimental data better than RESP (0.69 vs. 1.36, respectively). This is at least in part due to five BCCs that, *a posteriori*, were

Table 4. Comparison of Hydrogen-Bonded DNA Dimer Energies[a] with HF/6-31G**[b] and MP2/6-31G*(.25)[b,c] Energies.

Dimers[d]	HF/6-31G**	MP2/6-31G*(.25)	RESP	ΔHF[d]	ΔMP2	AM1	ΔHF	ΔMP2	MMFF	ΔHF	ΔMP2	AM1-BCC	ΔHF	ΔMP2
AA1	−8.7	−11.5	−10.19	1.5	−1.3	−5.65	−3.1	−5.9	−7.87	−0.8	−3.6	−11.60	2.9	0.1
AA2	−8.0	−11.0	−10.50	2.5	−0.5	−7.06	−0.9	−3.9	−8.41	0.4	−2.6	−11.55	3.6	0.6
AA3	−6.9	−9.8	−10.79	3.9	1.0	−8.16	1.3	−1.6	−8.49	1.6	−1.3	−11.91	5.0	2.1
AC1	−11.9	−14.3	−12.26	0.4	−2.0	−6.28	−5.6	−8.0	−9.05	−2.9	−5.3	−13.58	1.7	−0.7
AC2	−11.4	−14.1	−12.98	1.6	−1.1	−6.16	−5.2	−7.9	−10.08	−1.3	−4.0	−13.67	2.3	−0.4
CC	−17.3	−18.8	−16.57	−0.7	−2.2	−9.31	−8.0	−9.5	−14.74	−2.6	−4.1	−19.07	1.8	0.3
GA1	−12.6	−15.2	−14.16	1.6	−1.0	−10.23	−2.4	−5.0	−12.83	0.2	−2.4	−13.96	1.4	−1.2
GA2	−7.5	−10.3	−10.35	2.9	0.0	−8.41	0.9	−1.9	−10.58	3.1	0.3	−12.51	5.0	2.2
GA3	−11.0	−13.8	−15.13	4.1	1.3	−11.34	0.3	−2.5	−13.84	2.8	0.0	−14.83	3.8	1.0
GA4	−8.8	−11.4	−9.77	1.0	−1.6	−11.34	2.5	−0.1	−9.57	0.8	−1.8	−12.21	3.4	0.8
GC1	−12.7	−14.3	−12.28	−0.4	−2.0	−8.95	−3.8	−5.4	−12.13	−0.6	−2.2	−15.00	2.3	0.7
GCNEW	−22.8	−22.2	−21.73	−1.1	−0.5	−9.64	−13.2	−12.6	−18.88	−3.9	−3.3	−19.92	−2.9	−2.3
GCWC	−25.5	−25.8	−25.35	−0.1	−0.4	−11.48	−14.0	−14.3	−23.11	−2.4	−2.7	−26.00	0.5	0.2
GG1	−25.0	−24.7	−24.41	−0.6	−0.3	−10.64	−14.4	−14.1	−21.65	−3.4	−3.1	−22.98	−2.0	−1.7
GG3	−16.8	−17.8	−17.41	0.6	−0.4	−11.28	−5.5	−6.5	−17.61	0.8	−0.2	−17.07	0.3	−0.7
GG4	−7.3	−10.0	−8.61	1.3	−1.4	−10.84	3.5	0.8	−10.26	3.0	0.3	−12.37	5.1	2.4
GT1	−14.0	−15.1	−15.62	1.6	0.5	−6.97	−7.0	−8.1	−15.65	1.7	0.6	−15.33	1.3	0.2
GT2	−13.7	−14.7	−15.36	1.7	0.7	−6.96	−6.7	−7.7	−15.36	1.7	0.7	−14.50	0.8	−0.2
TAH	−10.9	−13.3	−13.66	2.8	0.4	−7.98	−2.9	−5.3	−13.13	2.2	−0.2	−13.77	2.9	0.5
TARH	−10.9	−13.2	−13.70	2.8	0.5	−6.28	−4.6	−6.9	−12.75	1.9	−0.4	−13.66	2.8	0.5
TARWC	−10.2	−12.4	−12.51	2.3	0.1	−6.40	−3.8	−6.0	−12.15	2.0	−0.3	−13.27	3.1	0.9
TAWC	−10.3	−12.4	−12.50	2.2	0.1	−6.47	−3.8	−5.9	−12.52	2.2	0.1	−13.47	3.2	1.1
TC1	−9.1	−11.4	−11.31	2.2	−0.1	−5.76	−3.3	−5.6	−11.52	2.4	0.1	−11.78	2.7	0.4
TC2	−9.2	−11.6	−11.43	2.2	−0.2	−5.83	−3.4	−5.8	−11.85	2.7	0.3	−12.35	3.2	0.8
TT1	−9.1	−10.6	−12.06	3.0	1.5	−5.66	−3.4	−4.9	−12.56	3.5	2.0	−11.79	2.7	1.2
TT2	−9.0	−10.6	−11.93	2.9	1.3	−5.61	−3.4	−5.0	−12.84	3.8	2.2	−11.91	2.9	1.3
TT3	−9.1	−10.6	−12.23	3.1	1.6	−5.72	−3.4	−4.9	−12.85	3.8	2.3	−11.72	2.6	1.1
Mean error				1.7	−0.2		−4.2	−6.1		0.8	−1.1		2.3	0.4
Mean unsigned error				1.9	0.9		4.8	6.2		2.2	1.7		2.7	0.9
Max error				4.1	2.2		14.4	14.3		3.9	5.3		5.1	2.4
Standard deviation				1.4	1.1		4.5	3.6		2.3	2.0		1.8	1.1

[a]All energies are in units of kcal/mol.
[b]HF and MP2 energies are for HF/6-31G** optimized geometries and are taken from ref. 15.
[c]The 6-31G*(.25) basis set represents a standard split valence 6-31G* basis set augmented with d-polarization functions on the nonhydrogen atoms with an exponent of 0.25 added to the nonhydrogens (see ref. 15).
[d]ΔHF is associated with the dimer energy value for the preceding charge model and evaluates the difference compared to the HF value, for example, ΔHF after the RESP value means E(HF) − E(RESP); similar to ΔMP2 compared to the MP2 dimer energy value.

adjusted in order to improve agreement with the experimental data. Adjustments were made to the $N_{2^-,3,4^+}$—H (from −0.1898 to −0.2048) and C_4—$N_{2^-,3,4^+}$ (from 0.0864 to 0.1582) BCC values in order to reproduce the experimentally observed free energy of solvation trends of the amine series:[29] ammonia (−4.3 kcal/mol), methylamine (−4.6 kcal/mol), dimethylamine (−4.3 kcal/mol), and trimethylamine (−3.2 kcal/mol). Testing the validity of these changes, these adjustments also improved the RFES of pyrrolidine to cyclopentane from 3.96 to 5.73 kcal/mol (experimental value of 6.68 kcal/mol); piperidine to cyclohexane from 4.13 to 5.77 kcal/mol (experimental value of 6.34 kcal/mol); and methylpiperidine to methylcyclohexane from 2.96 to 5.48 kcal/mol (experimental value of 5.60 kcal/mol). Adjustments were also made to the N_3^{hdeloc} $\underline{\text{deloc}}$ $O_{1,2}$ BCC, used for the nitro-containing functionalities that appeared to be problematic.[30] The BCC value for bond-type N_3^{hdeloc} $\underline{\text{deloc}}$ $O_{1,2}$ required a large adjustment from 0.1203 to −0.1500, changing the RFES of nitroethane to isopentane from 10.54 to 5.57 kcal/mol, in close agreement with the experimental value of 6.08 kcal/mol. Similarly, the RFES of nitrobenzene to isopropylbenzene changed from 7.44 to 4.48 kcal/mol, approaching the experimental value of 3.80 kcal/mol. We noticed that the AM1 charge model better reproduced the RFES for perturbations to unsaturated hydrocarbons (e.g., cyclopentene to cyclopentane) than did AM1-BCC before adjustment, suggesting that a BCC was not needed for C—H bonds in unsaturated hydrocarbons. Consequently, unsaturated hydrocarbon

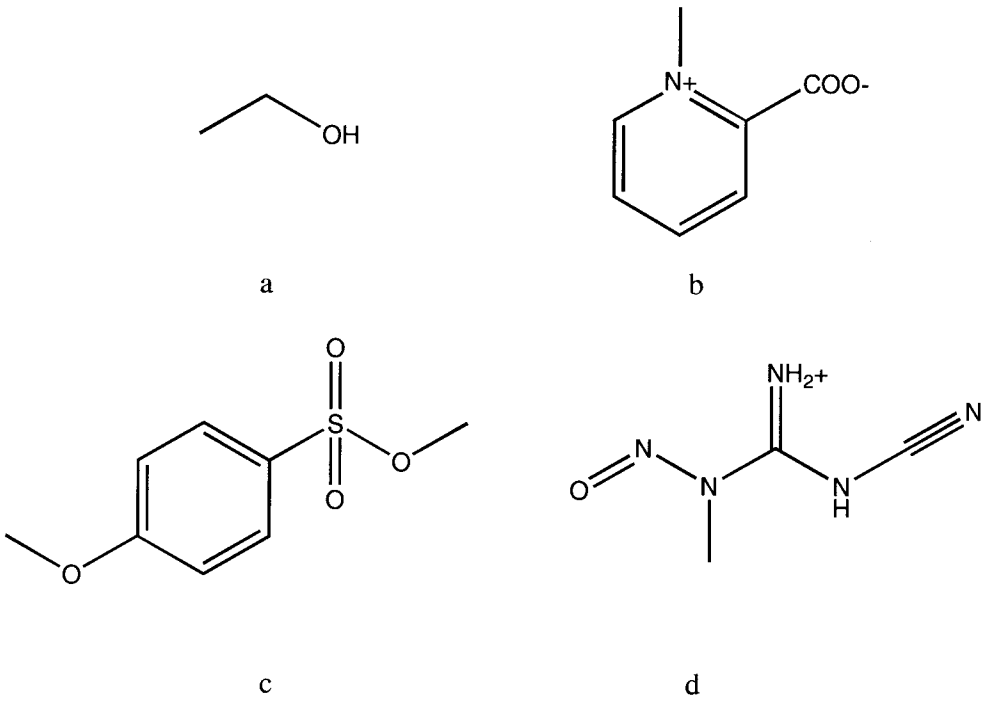


Figure 9. Representation of (a) ethanol, a simple mono-functionalized molecule; (b) homarine, a zwitterion; (c) *p*-methoxy benzenesulfonate, a delocalized multifunctionalized molecule; and (d) N-methyl-N′-cyano-N-nitrosoguanine cation (MCNG), a multi- and densely functionalized cation.

BCCs (i.e., $C_3^{=C}$—H_1 and C_{ar}—H_1) were set to zero, which gave rise to RFES closer to experiment. The change from 0.0116 to 0.0000 for $C_3^{=C}$—H_1 made only modest improvements, while the adjustment from 0.0128 to 0.000 for C_{ar}—H_1 improved the RFES of benzene to cyclohexane from 0.87 to 1.50 kcal/mol, approaching the experimental value of 2.09 kcal/mol. We found that setting the BCC for C_4—H_1 to zero had an overall unfavorable effect on the RFES of methylated molecules, so it was left untouched.

Overall Performance

The correlations between RESP, AM1, MMFF, and AM1-BCC for the *ab initio* dimer energies and the experimental RFES are shown in Table 7. For the organic dimer energies all models except for AM1 correlate well with both *ab initio* methods with correlation coefficients above 0.95. For the DNA dimers, both RESP and AM1-BCC had very good correlation coefficients with the *ab initio* dimer energies, although those for MMFF deteriorate. The AM1-BCC model achieves the best correlation with experiment for the RFES, partially attributable to the five adjusted BCCs (see above). In general, AM1 population charges perform poorly in every context. MMFF does well with molecules bearing simple organic functionality around which it was parameterized, but does less well in densely functionalized or charged molecules. RESP does well in all contexts except for known liabilities in the RFES of amines and nitro groups. AM1-BCC consistently maintained a high correlation coefficient (>0.95) in all contexts.

Table 5. RMS Deviations from the QM ESP[a] for RESP, MMFF, and AM1-BCC Charge Models for the Molecules in Figure 9.

	RMS deviation from QM ESP (a.u.)			
Molecule	RESP	AM1	MMFF	AM1-BCC
Ethanol	0.0254	0.0816	0.0408	0.0259
Homarine	0.0214	0.1079	1.0277	0.0424
p-Methoxy benzenesulfonate	0.0190	0.1159	0.7537	0.0230
MCNG	0.0294	0.1593	0.5029	0.0547

[a]The ESP was calculated at the HF/6-31G* level of theory.

Conclusions

The AM1-BCC atomic charge method adjusts an initial set of AM1 atomic charges, which do not emulate the HF/6-31G* ESP around a molecule, into a new set of charges that do, using simple additive correction terms applied to the AM1 charges. This charge model thus quickly generates high-quality atomic charges suitable for simulations in polar solution-phase systems: the user of the AM1-BCC model simply needs to calculate AM1 atomic charges and apply the BCCs; there are no *ab initio* calculations to perform. A training set of 2755 organic molecules was created to parameterize the 309 unique nonzero BCCs of the 354 sample bond-types

Table 6. Relative Free Energies[a] of Solvation of Organic Molecules Using Various Charge Models.

		RESP		AM1		MMFF		AM1BCC	
Perturbation	Exp.[b]	Calc.	Error[c]	Calc.	Error	Calc.	Error	Calc.	Error
Ammonia → methane	6.31	8.27	−1.96	1.21	5.11	7.36	−1.05	6.19	0.12
CH_3NH_2 → ethane	6.39	7.63	−1.24	1.61	4.79	7.08	−0.69	6.32	0.07
$(CH_3)_2NH$ → propane	6.25	5.15	1.10	1.86	4.39	6.40	−0.15	5.98	0.27
$(CH_3)_3N$ → isobutane	5.56	1.68	3.89	2.02	3.55	5.63	−0.07	5.64	−0.08
Phenol → aniline	−0.10	2.40	−2.50	−1.49	1.39	1.58	−1.68	0.15	−0.25
Pyridine → benzene	3.83	2.71	1.12	0.65	3.18	4.79	−0.96	2.64	1.19
THF → cyclopentane	4.67	4.61	0.06	2.66	2.02	5.06	−0.39	3.85	0.83
Cyclopentene → cyclopentane	0.64	0.13	0.51	0.73	−0.09	0.60	0.04	0.33	0.31
Pyrrolidine → cyclopentane	6.68	5.93	0.76	1.21	5.48	6.42	0.26	5.73	0.96
Benzene → cyclohexane	2.09	1.62	0.47	1.48	0.61	2.65	−0.56	1.50	0.59
Piperidine → cyclohexane	6.34	4.74	1.61	1.84	4.51	6.03	0.32	5.77	0.58
Acetaldehyde → ethane	5.33	5.80	−0.47	2.62	2.71	8.15	−2.82	5.84	−0.51
Bromomethane → ethane	2.63	2.09	0.54	2.99	−0.36	2.09	0.54	1.82	0.81
Chloromethane → ethane	2.43	0.08	2.35	−0.67	3.10	0.78	1.65	−0.13	2.56
Ethene → ethane	0.56	0.42	0.15	−0.13	0.69	0.22	0.35	−0.11	0.67
Fluoromethane → ethane	2.03	1.95	0.09	1.11	0.93	2.68	−0.65	1.65	0.39
Methanol → ethane	6.95	8.47	−1.52	3.01	3.94	8.11	−1.16	7.28	−0.33
Methylthiol → ethane	3.07	2.98	0.09	0.52	2.55	2.94	0.13	2.84	0.23
Anisole → ethylbenzene	0.24	0.87	−0.63	0.71	−0.47	0.75	−0.51	1.47	−1.23
Acetamide → isobutane	12.03	12.54	−0.51	6.89	5.15	12.16	−0.13	12.26	−0.23
Acetic acid → isobutane	9.02	10.97	−1.95	5.30	3.73	11.11	−2.09	10.00	−0.98
Nitrobenzene → *i*-Pr benzene	3.80	6.91	−3.11	5.67	−1.87	5.77	−1.97	4.48	−0.68
Me 2Pr ether → isopentane	4.39	2.76	1.63	2.53	1.86	4.64	−0.25	3.48	0.91
Me acetate → isopentane	5.70	6.59	−0.89	4.57	1.13	6.17	−0.47	6.66	−0.96
Me butene → isopentane	1.07	−0.27	1.34	0.76	0.31	0.80	0.27	0.22	0.85
Me Et ketone → isopentane	6.02	6.35	−0.33	3.21	2.81	8.17	−2.15	6.45	−0.43
Nitroethane → isopentane	6.08	9.39	−3.31	8.60	−2.52	11.13	−5.05	5.57	0.51
NMA → isopentane	12.50	10.98	1.53	6.67	5.84	11.77	0.73	12.19	0.32
Mepiperidine → mecyclohexane	5.60	1.25	4.36	1.80	3.81	5.25	0.36	5.48	0.12
Toluene → Me cyclohexane	2.59	1.77	0.83	2.29	0.31	3.32	−0.73	2.05	0.55
Me imidazole → Me pyrrole	5.52	3.64	1.89	1.02	4.51	7.05	−1.53	4.19	1.34
Me acetate → NMA	−6.76	−3.59	−3.17	−1.30	−5.46	−4.80	−1.96	−4.73	−2.03
Phenol → toluene	5.73	6.69	−0.96	0.51	5.23	6.41	−0.68	5.81	−0.07
$(CH_3)_2O$ → propane	3.85	3.18	0.67	2.39	1.46	4.93	−1.08	3.43	0.42
$(CH_3)_2S$ → propane	3.49	2.52	0.97	1.15	2.34	4.18	−0.69	2.82	0.67
Imidazole → pyrrole	4.85	3.57	1.28	1.12	3.73	8.63	−3.78	4.50	0.35
Benzaldehyde → toluene	3.14	4.47	−1.33	1.77	1.38	6.42	−3.28	4.51	−1.37
Bromobenzene → toluene	0.61	0.08	0.53	0.57	0.04	−0.19	0.80	0.06	0.55
Chlorobenzene → toluene	0.21	−1.88	2.09	−1.95	2.16	−2.09	2.30	−1.98	2.19
Thiobenzene → toluene	1.66	2.28	−0.62	1.60	0.06	1.20	0.47	1.33	0.34
Mean error			0.13		2.10		−0.71		0.24
Mean unsigned error			1.36		2.64		1.12		0.69
Maximum			4.36		5.84		5.05		2.56
Standard deviation			1.73		2.42		1.42		0.88

[a]All energies are in kcal/mol.
[b]Experimental values obtained from refs. 27 and 28.
[c]Error = Exp. − Calc.

composed of H, C, N, O, F, P, S, Si, Cl, Br, and I atoms. Five BCCs were adjusted *a posteriori* to better reproduce the RFES of the amine series, aromatic and olefinic hydrocarbons, and nitro-containing functionalities. The resulting set of BCC parameters allows the model to charge any molecule in The Merck Index and the NCI database of small molecules, except molecules that contain boron or covalently bound metal atoms. Although silicon BCCs are present, a complete set of silicon BCCs has not been parameterized in this work. Replacing RESP charges with AM1-BCC charges within the Cornell et al.[13] FF gave a model that was

Table 7. Correlation Coefficients between the Various Charge Models for the *Ab Initio* Dimer Energies and the Experimental Relative Free Energies of Solvation.

	Organic dimers		DNA dimers		
	HF	MP2	HF	MP2	RFES Exp.
RESP	0.96	0.95	0.97	0.97	0.88
AM1	0.82	0.79	0.50	0.54	0.70
MMFF	0.98	0.97	0.91	0.88	0.93
AM1-BCC	0.97	0.96	0.97	0.98	0.97

able to reproduce correlated *ab initio* hydrogen-bonded organic homo- and hetero-dimer energies within, on average, 0.95 kcal/mol, and hydrogen-bonded DNA dimer energies to within 0.9 kcal/mol, on average. These energies tend to fall in between HF and MP2 values. This model was also validated against RFES of isosteric small organic molecules, reproducing experimental relative free energies to within 0.69 kcal/mol, on average. The AM1-BCC charge model consistently maintained a correlation coefficient above 0.96 for all the validation tests with respect to *ab initio* and experiment, and therefore we propose it as a general method to charge organic molecules for use in solution phase simulations.

References

1. Williams, D. E. In Computational Chemistry Reviews, Vol. 2; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1991.
2. Francl, M. M.; Chirlain, L. E. In Computational Chemistry Reviews, Vol. 14; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2000.
3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. J Am Chem Soc 1993, 115, 9620.
4. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. J Comput Chem 2000, 21, 132.
5. (a) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902; (b) Stewart, J. J. P. MOPAC 6.0 (SGI version), Quantum Program Chemistry Exchange (QCPE) #455. AM1 atomic charges are the standard population charges produced by default by MOPAC 6.0 and are derived from the Coulson density matrix (see pp. 4–8, notes 18 and 20 of the MOPAC 6.0 manual).
6. (a) Besler, B. H.; Merz Jr, K. M.; Kollman, P. A. J Comput Chem 1990, 11, 431; (b) Alemán, C.; Luque, F. J.; Orozco, M. J Comput Chem 1993, 14, 799; (c) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. J Comput-Aided Mol Des 1995, 9, 87; (d) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 1998, 102, 1820.
7. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
8. Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian 92 Revision A; Gaussian, Inc.: Pittsburgh, PA, 1992.
9. LaJohn, L. A.; Christiansen, P. A.; Ross, R. B.; Atashroo, T.; Ermler, W. C. J Chem Phys 1987, 87, 2812.
10. Connolly, M. L. J Appl Cryst 1983, 16, 548.
11. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94; Gaussian, Inc.: Pittsburgh, PA, 1995.
12. McCammon, J. A.; Harvey, S. C. Dynamics of Proteins and Nucleic Acids; Cambridge University Press: Cambridge, 1989; Chapter 4.
13. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
14. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham III, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Siebel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 5, University of California: San Francisco, 1997.
15. Šponer, J.; Leszczynski, J.; Hobza, P. J Phys Chem 1996, 100, 1965.
16. Schmidt, M. W.; Baldridge, K. K.; Boatz, J .A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. J Comput Chem 1993, 14, 1347.
17. http://www.ccl.net/cca/data/ff_evaluation_suite (accessed 18 November 1998).
18. The CD-ROM of the Merck Index, 12th Edition, is commercially available.
19. The NCI database of ca. 250,000 molecules was downloaded from http://cactus.cit.nih.gov/ncidb/download.html
20. Dixon, R. W.; Kollman, P. A. J Comput Chem 1997, 18, 1632.
21. Bush, B. L.; Sheridan, R. P. J Chem Inf Comput Sci 1993, 33, 756.
22. Halgren, T. A. J Comput Chem 1996, 17, 490.
23. Halgren, T. A. J Comput Chem 1996, 17, 520.
24. Gould, I. R.; Kollman, P. A. J Am Chem Soc 1994, 116, 2493.
25. Meng, E. C.; Cieplak, P.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1994, 116, 12061.
26. Chipot, C.; Maigret, B.; Pearlman, D. A.; Kollman, P. A. J Am Chem Soc 1996, 118, 2998.
27. Gerber, P. R. J Comput-Aided Mol Des 1998, 12, 37.
28. Cramer, C. J.; Truhlar, D. G. J Comput-Aided Mol Des 1992, 6, 629.
29. Rizzo, R. C.; Jorgensen, W. L. J Am Chem Soc 1999, 121, 4827.
30. Duffy, E. M.; Jorgensen, W. L. J Am Chem Soc 2000, 122, 2878.