

'Inductive' Charges on Atoms in Proteins: Comparative Docking with the Extended Steroid Benchmark Set and Discovery of a Novel SHBG Ligand

Artem Cherkasov,^{*,†} Zheng Shi,[†] Yvonne Li,[‡] Steven J. M. Jones,[§] Magid Fallahi,^{||} and Geoffrey L. Hammond^{||}

Division of Infectious Diseases, Faculty of Medicine, University of British Columbia, 2733 Heather Street, Vancouver, British Columbia V5Z 3J5, Canada, Strategic CIHR/MSFHR Training Program in Bioinformatics and BC Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada, and Department of Obstetrics and Gynecology, University of British Columbia, BC Research Institute for Children's and Women's Health, British Columbia, Canada

Received June 4, 2004

We have developed a novel iterative approach for calculation of partial charges in proteins within the framework of the 'molecular capacitance' model. The method operates by an effective 'inductive' electronegativity scale derived from a number of the conventional charge systems including CHARMM, AMBER, MMFF, OPLS, and PEOE among others. Our novel 'inductive' electronegativity equalization procedure allows rapid and conformation sensitive computation of adequate partial charges in proteins. Accuracy of the 'inductive' values was confirmed by their correlation with DFT-computed partial charges in common amino acids. A comparative docking study with an extended steroid data set not only illustrated the adequacy of 'inductive' protein charges but also demonstrated their superior performance compared to several conventional protein charging systems. Subsequent docking with 'inductive' charges resulted in identification of five potential leads as human Sex Hormone Binding Globulin (SHBG) ligands from a commercial library of natural compounds. When the selected substances were evaluated for their ability to bind SHBG *in vitro*, three of them displaced testosterone from the SHBG steroid-binding site, and with one compound this was achieved at micromolar concentrations.

INTRODUCTION

Net atomic charges represent one of the most important classes of reactivity indices as they play an essential role in molecular electrostatic calculations and have a broad application in all areas of chemical and biochemical research. Although the actual distribution of electrons within a molecule is continuous, it is commonly accepted that discrete electrostatic charges can be assigned to (and centered on) bound atoms approximated as spheres. Partial charges on protein atoms play an essential role in protein folding simulations, structure-based drug design, conformational analysis, molecular dynamics, Monte Carlo simulations, etc.^{1–5} and, therefore, it is very important to develop reliable and easy-to-use approaches for their calculation.^{6,7}

Conventional methods of approximating atomic charges in proteins can be conditionally divided into two main classes. One group of approaches is based on the transferability principle and uses various properties of isolated amino acids or small peptides for benchmarking protein atomic charges. Thus, force field methods derive charges on atoms in amino acids from their solvation energies, dipole moments,

properties of crystals,^{8–19} etc. Several approaches utilize *ab initio* calculations allowing indirect quantification of deficit/excess of electronic density in amino acids either from the Mulliken population analysis (often seen as unsatisfactory²⁰) or by fitting the calculated electrostatic potential (method suffering from conformational sensitivity²¹).

The second group of methods for protein charge calculation operates by atomic and/or bond reactivity indices derived from the electronegativity equalization approximation.^{22–27} These methods often adopt an iterative scheme of calculation of partial charges and consider the connectivity of atoms in a protein without utilizing actual three-dimensional atomic coordinates.²⁶

Currently there is no 'golden standard' for calculating partial charges in macromolecules. The existing charge systems may reflect different aspects of unevenly distributed electronic density in proteins and may have good applicability for specific tasks, but they all suffer from significant drawbacks. Thus, the 'transferral' methods assign static partial charges according to atomic types, and this is an obvious oversimplification. Furthermore, the electronegativity equalizing approaches do not take three-dimensional molecular structures into account. In both cases, the resulting partial charges may adequately represent the overall distribution of intramolecular electronic density, but they do not usually reflect the finer aspects of molecular electrostatics.

Overview of Molecular Capacitance Model. In our previous works^{27–29} we developed an iterative scheme for

* Corresponding author phone: 604-875-4588; fax: 604-875-4013; e-mail: artc@interchange.ubc.ca.

[†] Faculty of Medicine, University of British Columbia.

[‡] Strategic CIHR/MSFHR Training Program in Bioinformatics, BC Cancer Agency.

[§] BC Genome Sciences Centre, BC Cancer Agency.

^{||} Department of Obstetrics and Gynecology, University of British Columbia.

calculating atomic partial charges $\Delta\Delta N_i$ in small molecules which is based on the ‘inductive’ electronegativity scale:

$$\Delta\Delta N_{i,k+1} = 0.172 \sum_{j \neq i}^{N-1} \frac{\Delta\chi'_{ji} (R_j^2 + R_i^2)}{r_{j-i}^2} \quad (1)$$

Here N is the number of atoms in the molecule, r_{i-j} is the distance between atoms i and j , R is a covalent radii, and $\Delta\chi_{ji}$ is the difference between the ‘inductive’ electronegativities of atoms i and j , and $\Delta\Delta N$ is the change in inductive charge ΔN_i between iterations k .

The physical interpretation of the inductive charges (1) has been introduced by considering a molecule as an electrical capacitor formed by charged atomic spheres. The capacitance of such a system has been related to the chemical hardness η and softness s of bound atoms (reflecting their ability to delocalize electrons):

$$s_i = 2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2} \quad \text{and} \quad \eta_i = \frac{1}{2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \quad (2)$$

Within this approximation, the repetitive applications of eq 1 and the formula for in situ (or equalized) electronegativity χ' (χ^0 at the initial step $i = 0$):

$$\chi'_{i+1} = \chi'_i + \frac{\Delta\Delta N_i}{2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \quad (3)$$

yield the cumulative atomic inductive charges $\Delta N_i = Q_i + \sum_k \Delta\Delta N_{i,k}$ (where k is the number of iterations, Q_i is the formal charge of atom i , and $\Delta\Delta N_{i,0}$ is set to zero). When calculated for low-molecular weight organic compounds, the cumulative inductive charges ΔN_i reproduce the corresponding experimental dipole moments and C 1s core electron binding energies with very high accuracy and thereby support the utility of the elaborated routine for those systems.²⁹

Several conventional descriptors have been developed within the capacitance model, including ‘inductive’ analogues of chemical hardness–softness; the Fukui index; and electronegativity as well as inductive and steric parameters for individual atoms, groups, and molecules. These QSAR descriptors have already been successfully applied for modeling the antibacterial activity of organic chemicals and cationic peptides³⁰ and in discovery of novel nonsteroidal analogues of sex steroids that compete for the steroid-binding site of plasma Sex Hormone Binding Globulin (SHBG).³¹ More details on ‘inductive’ parameters and their theoretical $h =$ justification can be found in a recent review.³⁰

On the other hand, the broader use of eqs 1–3 has been previously limited by unavailability of absolute electronegativity values for some atomic types.^{2,32} In particular, application of the developed approaches could not be readily

extended to proteins because they contain many atomic types that are not ‘covered’ by Taft’s substituent constants.

RESULTS AND DISCUSSION

Introduction of Effective ‘Inductive’ Electronegativity.

The variability of resonance forms and ionization, which are attributed to amino acids, represents a recognized challenge for χ -operating approaches.³³ Similarly, to make the inductive charge calculating scheme applicable for proteins, we had to develop an ‘inductive’ electronegativity scale covering specific atomic types present in free and bound amino acids. In the current work, we derived such electronegativities from the values of partial charges assigned to protein atoms by common conventional force field and electronegativity equalization approaches and conducted a comparative evaluation of ‘inductive’ partial charges in proteins.

To derive ‘inductive’ electronegativity for specific atomic types present in proteins, we first introduced a novel parameter of **effective** ‘inductive’ electronegativity $\chi_i^0(\text{eff})$. It is postulated that the **effective** ‘inductive’ electronegativity corresponds to an absolute (initial) atomic electronegativity χ^0 that allows equalization of electronegativities of all atoms in a protein/peptide/amino acid in a single run (in contrast to the previously utilized k -steps iterative procedure)

$$\chi'_i = \chi'_{\text{MOL}} = \chi_i^0 + \frac{\sum_k \Delta\Delta N_{i,k}}{2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} = \chi_i^0(\text{eff}) + \frac{Q_{\text{formal}} - q_i}{2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \quad (4)$$

(here Q_{formal} is the formal charge on i th atom).

This assumption could only be valid, however, if we could solve a system of equations (4) for all atomic charges in all common amino acids, and if such a solution would result in a unique set of effective electronegativities $\chi^0(\text{eff})$.

It should also be noted that eq 4 contains two types of unknowns: an effective atomic electronegativity $\chi_i^0(\text{eff})$ and equalized molecular electronegativity χ'_{MOL} . Thus, for each N -atomic molecule we could write N equations with $K+1$ variables where K represents the total number of distinct atomic types under consideration:

$$\left[\chi_i^0(\text{eff}) - \chi'_{\text{MOL}} = \frac{Q_{\text{formal}_i} - q_i}{2 \sum_{j \neq i}^{N-1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \right]_N \quad (5)$$

In the series of **A** compounds containing **K** distinct atomic types, all **K** effective absolute electronegativities $\chi_i^0(\text{eff})$ should retain their values throughout the system; therefore, we presented a system of **R** equations (5) with **(A+K)** unknowns where **R** is the total number of atoms contained in **A** compounds under study:

$$\left[\begin{aligned} \chi_{i,1}^0(\text{eff}) - \chi'_{\text{MOI},1} &= \frac{Q_{\text{formal},i,1} - q_{i,1}}{2 \sum_{j \neq i, j \in \text{MOI},1, j \in \text{MOI}_1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \\ \chi_{N,1}^0(\text{eff}) - \chi'_{\text{MOI},1} &= \frac{Q_{\text{formal},i} - q_i}{2 \sum_{j \neq i, j \in \text{MOI},1, j \in \text{MOI}_1} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \\ \chi_{i,A}^0(\text{eff}) - \chi'_{\text{MOI},A} &= \frac{Q_{\text{formal},i,A} - q_{i,A}}{2 \sum_{j \neq i, j \in \text{MOI},A, j \in \text{MOI}_A} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \\ \chi_{N,A}^0(\text{eff}) - \chi'_{\text{MOI},A} &= \frac{Q_{\text{formal},i,A} - q_{i,A}}{2 \sum_{j \neq i} \frac{R_j^2 + R_i^2}{r_{j-i}^2}} \end{aligned} \right]$$

We approached the solution of the system of linear equations (6) for $(\mathbf{A}+\mathbf{K})$ unknown variables through the integer matrix representation:

$$\left[\begin{array}{cccccc} \alpha_{1,1} & & \alpha_{1,K} & \beta_{1,K+1} & & \beta_{1,K+A} \\ & \alpha_{i,j} & & & \beta_{k,j} & \\ \alpha_{M,1} & & \alpha_{M,K} & \beta_{M,K+1} & & \beta_{M,K+A} \end{array} \right] \otimes \left[\begin{array}{c} \chi_1(\text{eff}) \\ \chi_K(\text{eff}) \\ \chi'_{K+1} \\ \chi'_{K+A} \end{array} \right] = \left[\begin{array}{c} \left(\frac{Q-q}{s} \right)_1 \\ \left(\frac{Q-q}{s} \right)_M \end{array} \right] \quad (7)$$

Here \mathbf{A} is the number of molecules in the data set under consideration, \mathbf{M} is the total number of times the distinct atomic types can be found in all \mathbf{A} molecules (when counted in each molecule independently), \mathbf{K} is the number of distinct atomic types present in the entire data set, and s substitutes the softness representation (2).

Effective ‘Inductive’ Electronegativity of Atoms in Amino Acids. We decided to solve the system of linear eq 7 for the molecular data set consisting of five configurations of common amino acids: neutral (_NEU), washed (_WSH), N-terminal (_Nt), C-terminal (_Ct), and central (_CTR) corresponding to the neutral form, fully ionized form, CH_3NH – substituted on the C terminus, $\text{CH}_3\text{C}(\text{O})$ substituted on the N – terminus and bisubstituted on both ends of the amino acids. The examples of five configurations of alanine are featured in Figure 1.

The data set was created to account for the maximal range of atomic types found in proteins. As for the dependent parameters of equations (7), we considered partial charges assigned to atoms in the studied amino acids by MMFF,^{8–10} CHARMM,^{11–13} Engh–Huber,¹⁵ AMBER,^{16,17} OPLS,¹⁸ PEF,^{23,24} and PEOE²⁶ as well as by TAFF (Tripos such as molecular field) and EE methods³⁴ (the total of 11 protein charge schemes). The reasoning for using the third party charges to establish ‘inductive’ electronegativity parameters can be formulated as the following. In our opinion, atomic electronegativity represents such a fundamental property, that even not necessarily agreeing ‘protein charging’ systems can be ‘reduced’ to a more–less generic set of atomic electronegativities. In other words, different charge calculation approaches can be based on the same, universal set of atomic electronegativity values—they just utilize/parametrize them differently.

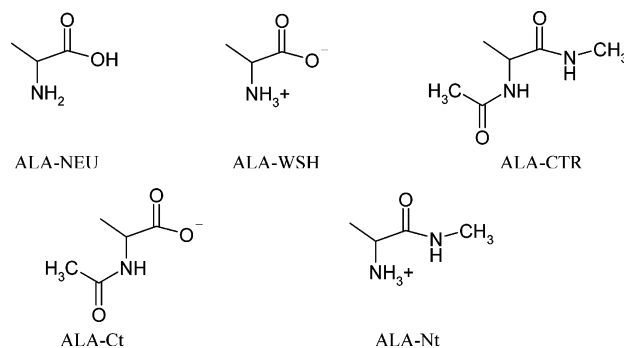


Figure 1. Five forms of alanine amino acid considered within the developed approach. The atomic charges have been assigned on the basis of the averaged effective electronegativity scale.

Thus, a total of 115 amino acid derivatives have been investigated (corresponding to 20 common amino acids plus three tautomeric forms of histidine). The geometry of each molecule was optimized by the MMFF force field, and the atomic types have been assigned according to the nature, hybridization state, and formal charge of each atom. The resonance structures were considered as mixtures of ‘pure’ atomic states: the carboxylic group being considered as comprising an sp^2 carbon ($\text{C}=\text{O}$), an sp^2 oxygen ($\text{O}=\text{O}$), and an sp^3 oxygen anion (O^-), and the benzene rings being treated as an assembly of sp^2 carbons, etc. This simplification was introduced in order to reduce the total number of atomic types.

For each of 115 molecules under study, we populated the matrix rows in (7) (corresponding to distinct atomic types present in a given structure) with two nonzero values α and β . Consequently, it was postulated that when the current atomic type i (row) matches a general type j (column) the corresponding parameter $\alpha_{i,j}$ acquires an integer value reflecting the number of atoms of the j th type present in the current molecule. The parameter $\beta_{k,j}$ can be either 0 or -1 depending on whether the atomic type i belongs to the molecule assigned to the column k . After all the rows were created for each molecule, they were appended into the general α/β matrix in a form of (7).

The force field and electronegativity-derived partial charges were then assigned to the molecules of the training set according to default values and procedures implemented within the MOE molecular modeling package.³⁴ Thus, for each of all 11 sets of partial charges $[q_i]_N$ we derived the corresponding $[Q_i + q_i/s]_M$ vectors, where s corresponds to the local inductive softness (2).

At the next step, for all cases when a molecule contains several atoms of the same type (such as hydrogens), the corresponding one-atom equations in a system (7) were combined, and the corresponding linear terms have been summed up. The resulting system of equations has then been solved using the Multiple Linear Regression approach as it is implemented within the MOE.³⁴ The fragment of the final equation system is presented in Table 1, and this illustrates the combined matrix coefficients for ALA_CTR, ALA_Ct, and VAL_WSH structures.

Finally, the resulting system of linear 2254 equations corresponding to 115 substances studied was solved independently against 11 sets of conventional atomic charges, and this resulted in 11 vectors of effective atomic and in situ molecular electronegativities $[\chi_1(\text{eff}), \dots, \chi_K(\text{eff}), \chi'_{K+1}$

Table 1. Representation of a Binary Matrix Used To Establish the Effective Inductive Electronegativities of Atoms and Molecules

name	atom type	H	C sp3	C sp2	N sp2	N sp3 +	O sp2	O sp3 -	Osp2-	ALA_CTR	ALA_Ct	VAL_WSH
ALA_Ct	O sp3 -	0	0	0	0	0	0	1	0	0	-1	0
	O sp2	0	0	0	0	0	2	0	0	0	-2	0
	N sp2	0	0	0	1	0	0	0	0	0	-1	0
	H	8	0	0	0	0	0	0	0	0	-8	0
	C sp3	0	3	0	0	0	0	0	0	0	-3	0
	C sp2	0	0	2	0	0	0	0	0	0	-2	0
ALA_CTR	O sp2	0	0	0	0	0	2	0	0	-2	0	0
	N sp2	0	0	0	2	0	0	0	0	-2	0	0
	H	12	0	0	0	0	0	0	0	-12	0	...
	C sp3	0	4	0	0	0	0	0	0	-4	0	0
	Csp2	0	0	2	0	0	0	0	0	-2	0	0
						...						
VAL_WSH	O sp3 -	0	0	0	0	0	0	1	0	0	0	-1
	O sp2	0	0	0	0	0	1	0	0	0	0	-1
	N sp3 +	0	0	0	0	1	0	0	0	0	0	-1
	H	11	0	0	0	0	0	0	0	0	0	-11
	C sp3	0	4	0	0	0	0	0	0	0	0	-4
	C sp2	0	0	1	0	0	0	0	0	0	0	-1

... χ'_{K+N}]. These electronegativity values, established as the linear coefficient of MLR solutions, were collected into Table 2 along with the resulting correlation coefficients r^2 reflecting the accuracy of the MLR fittings. It is worth mentioning, that the produced values of partial charges for model amino acids and their derivatives can be viewed as transferable parameters — as such methodology is broadly used by the conventional ‘protein charging’ approaches (they compute partial charges for isolated main amino acids and to transfer charge values to proteins). At the same time, the approach we developed allows differentiating partial charges in proteins depending on the actual local environment as it will be demonstrated in greater detail in the following sections.

As can be seen from the data in Table 2, the estimated r^2 parameters range from 0.93 to 0.98 and demonstrate that the solution of a system of 2254 linear equations accurately predicts 11 sets of unknowns $\chi_i^{0(\text{eff})}$. The derived effective electronegativity scales [$\chi_1(\text{eff})$, ... $\chi_K(\text{eff})$, χ'_{K+1} ... χ'_{K+N}] also agree among themselves, and Table 3 features the coefficients of their cross-correlations. Notably, all electronegativity scales derived from the force field charging schemes appear to be very similar (correlate with $r^2 > 0.93$), while only the results of the MOE electronegativity equalization method³⁴ do not correlate well with any other scales.

Considering the fact that the estimated sets of effective electronegativities agree with one another, we assumed that we could use the 11 derived effective electronegativity solutions to build a consensus electronegativity scale to reflect an averaged ability of protein atomic types to attract/donate electrons. Hence, using the arithmetic mean, we produced the average scale $\chi(\text{consensus})$ presented in Table 2, that correlates well with all the ‘parental’ electronegativity scales. Thus, the derived consensus inductive χ scale supports the idea that ‘charges do not correlate, but electronegativities — do’.

To evaluate the derived consensus effective electronegativity, we utilized the $\chi(\text{consensus})$ values for the calculation of partial charges in 23 selected amino acids according to (4). We also computed similar values using the DFT approach with a 6-31G* basis set and the B3LYP function (the molecular geometries have been preoptimized by the PM3 method). The DFT calculations were carried out within the *HyperChem* 7.5 package,³⁵ and the resulting values are collected into supplement file 1 in the Supporting Informa-

tion. When the atomic charges produced in the DFT calculation (requiring several hours of CPU time for each amino acid molecule) were compared with the corresponding inductive parameters (computed in a matter of few seconds), the resulting linear correlation (Figure 2) appeared to be of rather high quality with the corresponding correlation coefficient $r^2 = 0.89$. It should be noted, however, that the dependence featured in Figure 2 seemed to encompass two separate trends: one for positively and another — for negatively charged atoms. The exact reasoning for such separation is not obvious at the moment and can reflect the specifics of either or both DFT and ‘inductive’ charge calculations.

It is also feasible to stipulate that electronegativity represents such fundamental atomic property, that even generally not well-agreeing ‘protein charging’ systems can be ‘reduced’ to a single generic set of atomic electronegativities. In other words, different charge calculation approaches are directly or indirectly based on the same, ‘universal’ set of atomic electronegativity values — they just utilize/parametrize them differently. Thus, the derived consensus scale of inductive electronegativities illustrates the idea that “charges do not correlate, electronegativities — do”.

In addition, we calculated the global softness parameters for the selected amino acids and their derivatives (computed as a sum of the constituent atomic softness (2)), which are presented in Table 2. As mentioned previously, these parameters should reflect the ability of compounds to delocalize electron density. To explore that assumption we compared the calculated molecular softness with the polarizability of amino acids. As seen in Figure 3, the correspondence between the molecular polarizability and softness parameters is very significant, and this justifies the original physical interpretation of the molecular capacitance.

Thus these results demonstrate that the developed method allows accurate estimations of effective atomic and molecular electronegativities from predefined partial atomic charges. (The source code for the developed method can be obtained from the author upon request.) However, it has been demonstrated that the consensus inductive electronegativity derived from a number of conventional protein charge systems allows the reproduction of quantum-chemical charges in common amino acids with good accuracy.

Table 2. Scales of Effective Atomic and Molecular Electronegativities Derived from Conventional Sets of Amino Acids Partial Charges

	χ	softness		polariz- ability	PEOE ³²	TAAF (Tripos- like)	Amber 22	Amber 23	CHARMM 16–19	PEF ³⁰ interact.	PEF ³⁰	MMFF 14–16	Engh– Huber ²¹	Emp. MOE 34	OPLS 24
		mol.	av												
r^2					0.9719	0.941	0.9404	0.9436	0.9691	0.9758	0.9762	0.9230	0.9609	0.978	0.9526
hydrogen	2.10				2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.10	2.09
carbon sp ³	2.39				2.36	2.20	2.34	2.40	2.44	2.52	2.54	2.23	2.42	2.44	2.39
carbon sp ²	2.30				2.35	2.32	2.24	2.31	2.32	2.30	2.33	2.31	2.28	2.29	2.29
nitrogen sp ³	3.01				2.96	3.39	2.96	2.90	3.02	2.98	2.99	3.28	2.73	3.06	3.02
nitrogen sp ²	2.79				2.89	2.92	2.74	2.74	2.84	2.77	2.77	2.84	2.67	2.81	2.75
nitrogen sp ² +	3.62				4.17	4.01	3.42	3.40	3.71	3.51	3.53	3.75	3.44	3.73	3.63
nitrogen sp ³ +	3.52				4.23	3.51	3.44	3.21	3.39	3.60	3.39	3.68	3.43	3.82	3.45
oxygen sp ³	3.20				3.19	3.20	3.20	3.19	3.20	3.20	3.20	3.20	3.20	3.20	3.19
oxygen sp ²	3.12				3.08	3.21	3.14	3.13	3.16	3.18	3.19	3.16	3.10	2.83	3.17
oxygen sp ³ -	2.23				2.16	2.52	2.18	2.43	2.35	2.16	2.22	2.57	2.11	1.30	2.41
oxygen sp ² -	2.18				2.10	2.40	2.14	2.40	2.29	2.19	2.21	2.43	2.04	1.31	2.33
sulfur sp ³	2.63				2.76	2.41	1.98	2.79	2.74	2.68	2.94	2.89	2.70	2.63	2.93
ALA_CTR	2.25	89.76	4.08	14.66	2.28	2.18	2.20	2.24	2.29	2.32	2.33	2.17	2.26	2.28	2.23
ALA_Ct	2.23	66.28	3.90	11.65	2.25	2.15	2.17	2.22	2.26	2.33	2.35	2.14	2.24	2.27	2.20
ALA_NEU	2.29	46.37	3.57	8.35	2.36	2.32	2.21	2.21	2.34	2.35	2.35	2.31	2.32	2.35	2.28
ALA_Nt	2.32	69.61	3.87	10.33	2.50	2.28	2.30	2.23	2.35	2.33	2.38	2.29	2.35	2.32	2.31
ALA_WSH	2.34	46.49	3.58	7.32	2.60	2.29	2.27	2.25	2.36	2.35	2.41	2.32	2.38	2.33	2.32
ARG_CTR	2.33	167.41	4.65	24.35	2.38	2.28	2.31	2.31	2.38	2.31	2.40	2.29	2.31	2.31	2.35
ARG_Ct	2.38	143.67	4.49	19.91	2.51	2.34	2.37	2.33	2.43	2.32	2.43	2.37	2.39	2.33	2.40
ARG_NEU	2.39	117.02	4.33	17.94	2.46	2.38	2.37	2.34	2.45	2.33	2.44	2.39	2.39	2.36	2.42
ARG_Nt	2.38	143.67	4.49	19.91	2.51	2.34	2.37	2.33	2.43	2.32	2.43	2.37	2.39	2.33	2.40
ARG_WSH	2.41	116.19	4.30	16.91	2.57	2.38	2.39	2.36	2.47	2.33	2.47	2.41	2.41	2.35	2.45
ASN_CTR	2.29	113.59	4.37	17.93	2.33	2.26	2.27	2.30	2.33	2.33	2.36	2.24	2.19	2.30	2.28
ASN_Ct	2.29	88.68	4.22	14.93	2.33	2.26	2.27	2.30	2.32	2.35	2.38	2.25	2.16	2.30	2.28
ASN_NEU	2.36	66.25	3.90	11.63	2.41	2.43	2.33	2.32	2.41	2.37	2.40	2.41	2.18	2.39	2.36
ASN_Nt	2.36	91.23	4.15	13.6	2.53	2.35	2.36	2.31	2.39	2.34	2.40	2.35	2.24	2.33	2.35
ASN_WSH	2.38	66.70	3.92	10.59	2.59	2.40	2.37	2.35	2.41	2.35	2.43	2.41	2.20	2.35	2.38
ASP_CTR	2.25	104.25	4.34	16.76	2.28	2.19	2.21	2.23	2.28	2.33	2.35	2.18	2.26	2.28	2.23
ASP_Ct	2.24	78.08	4.11	13.76	2.28	2.17	2.19	2.21	2.26	2.36	2.38	2.16	2.25	2.28	2.21
ASP_NEU	2.32	61.55	3.85	10.46	2.44	2.39	2.17	2.17	2.40	2.38	2.38	2.39	2.37	2.38	2.33
ASP_Nt	2.32	83.62	4.18	12.43	2.50	2.28	2.29	2.23	2.34	2.33	2.38	2.29	2.34	2.31	2.30
ASP_WSH	2.32	57.34	3.82	9.42	2.59	2.28	2.25	2.22	2.33	2.35	2.42	2.30	2.35	2.32	2.29
CYS_CTR	2.25	99.76	4.34	17.66	2.29	2.19	2.21	2.26	2.29	2.30	2.33	2.19	2.26	2.27	2.25
CYS_Ct	2.24	75.33	4.19	14.65	2.28	2.17	2.18	2.25	2.27	2.31	2.35	2.17	2.24	2.26	2.23
CYS_NEU	2.31	54.82	3.92	11.35	2.39	2.34	2.23	2.25	2.35	2.33	2.35	2.34	2.30	2.34	2.31
CYS_Nt	2.33	78.18	4.11	13.33	2.51	2.29	2.31	2.27	2.36	2.31	2.38	2.31	2.35	2.31	2.33
CYS_WSH	2.35	55.18	3.94	10.32	2.62	2.31	2.28	2.29	2.36	2.32	2.41	2.35	2.37	2.32	2.35
GLN_CTR	2.27	130.13	4.49	19.77	2.30	2.24	2.25	2.27	2.31	2.32	2.35	2.22	2.19	2.29	2.27
GLN_Ct	2.27	104.14	4.34	16.76	2.30	2.23	2.25	2.27	2.31	2.33	2.36	2.22	2.15	2.29	2.26
GLN_NEU	2.32	81.91	4.10	13.46	2.36	2.37	2.30	2.28	2.37	2.35	2.37	2.35	2.18	2.36	2.32
GLN_Nt	2.33	108.04	4.32	15.43	2.47	2.31	2.33	2.28	2.36	2.32	2.38	2.32	2.22	2.32	2.32
GLN_WSH	2.35	81.97	4.10	12.43	2.52	2.35	2.33	2.30	2.38	2.34	2.40	2.35	2.21	2.33	2.35
GLU_CTR	2.23	123.20	4.56	18.6	2.25	2.16	2.20	2.19	2.26	2.31	2.32	2.15	2.24	2.26	2.21
GLU_Ct	2.21	95.42	4.34	15.59	2.24	2.14	2.18	2.17	2.24	2.32	2.33	2.13	2.23	2.25	2.19
GLU_NEU	2.29	77.91	4.10	12.75	2.37	2.32	2.20	2.14	2.35	2.35	2.34	2.32	2.34	2.36	2.28
GLU_Nt	2.29	102.19	4.44	14.26	2.43	2.24	2.27	2.20	2.31	2.31	2.35	2.25	2.32	2.30	2.27
GLU_WSH	2.29	73.41	4.08	11.26	2.50	2.24	2.25	2.19	2.31	2.33	2.38	2.26	2.33	2.30	2.27
GLY_CTR	2.26	72.34	3.81	12.82	2.31	2.20	2.21	2.25	2.30	2.33	2.35	2.19	2.28	2.28	2.25
GLY_Ct	2.24	50.22	3.59	9.82	2.29	2.17	2.17	2.23	2.27	2.35	2.37	2.16	2.25	2.28	2.21
GLY_NEU	2.35	31.32	3.13	6.52	2.46	2.43	2.23	2.23	2.41	2.39	2.39	2.42	2.37	2.40	2.33
GLY_Nt	2.37	53.43	3.56	8.49	2.64	2.34	2.34	2.26	2.40	2.35	2.42	2.36	2.40	2.34	2.36
GLY_WSH	2.42	31.71	3.17	5.49	2.85	2.39	2.32	2.30	2.42	2.38	2.48	2.43	2.47	2.37	2.39
HID_CTR	2.26	134.64	4.64	21.48	2.31	2.24	2.21	2.26	2.30	2.30	2.32	2.22	2.27	2.28	2.25
HID_Ct	2.26	105.57	4.40	18.48	2.32	2.24	2.20	2.25	2.29	2.31	2.32	2.21	2.25	2.28	2.24
HID_NEU	2.30	83.59	4.18	15.18	2.40	2.36	2.22	2.25	2.35	2.31	2.32	2.33	2.29	2.34	2.30
HID_Nt	2.32	111.73	4.47	17.15	2.48	2.32	2.28	2.25	2.34	2.29	2.34	2.31	2.34	2.31	2.31
HID_WSH	2.33	83.92	4.20	14.15	2.56	2.35	2.26	2.28	2.36	2.30	2.35	2.34	2.34	2.32	2.32
HIE_CTR	2.27	132.72	4.58	21.48	2.32	2.24	2.22	2.27	2.30	2.31	2.32	2.22	2.27	2.29	2.26
HIE_Ct	2.27	104.50	4.35	18.48	2.33	2.25	2.21	2.27	2.30	2.31	2.33	2.22	2.27	2.29	2.25
HIE_NEU	2.31	83.53	4.18	15.18	2.40	2.37	2.24	2.27	2.35	2.31	2.32	2.33	2.30	2.34	2.30
HIE_Nt	2.32	110.13	4.41	17.15	2.49	2.32	2.29	2.27	2.35	2.30	2.35	2.31	2.34	2.31	2.31
HIE_WSH	2.34	83.42	4.17	14.15	2.57	2.36	2.28	2.30	2.37	2.31	2.36	2.34	2.35	2.32	2.33
HIP_CTR	2.30	137.03	4.57	21.93	2.39	2.26	2.27	2.30	2.34	2.30	2.33	2.26	2.29	2.29	2.27
HIP_Ct	2.31	110.35	4.41	18.93	2.42	2.27	2.27	2.31	2.35	2.31	2.34	2.28	2.29	2.29	2.28
HIP_NEU	2.36	89.50	4.26	15.62	2.51	2.40	2.32	2.32	2.42	2.32	2.34	2.40	2.36	2.35	2.33
HIP_Nt	2.36	113.89	4.38	17.6	2.57	2.34	2.35	2.30	2.40	2.31	2.36	2.36	2.36	2.32	2.34
HIP_WSH	2.39	89.14	4.24	14.59	2.67	2.39	2.35	2.35	2.43	2.31	2.38	2.42	2.39	2.33	2.36
HIS_CTR	2.30	138.88	4.63	21.93	2.39	2.26	2.27	2.30	2.34	2.30	2.33	2.26	2.29	2.28	2.27
HIS_Ct	2.31	110.66	4.43	18.93	2.42	2.27	2.27	2.31	2.35	2.31	2.34	2.28	2.30	2.28	2.27

Table 20. (Table 2 contd)

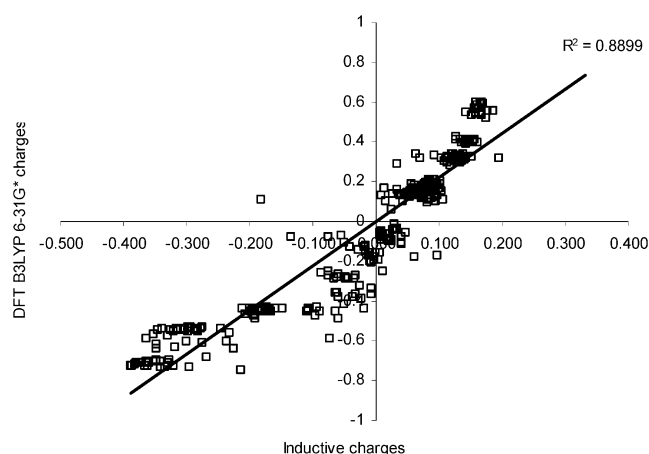
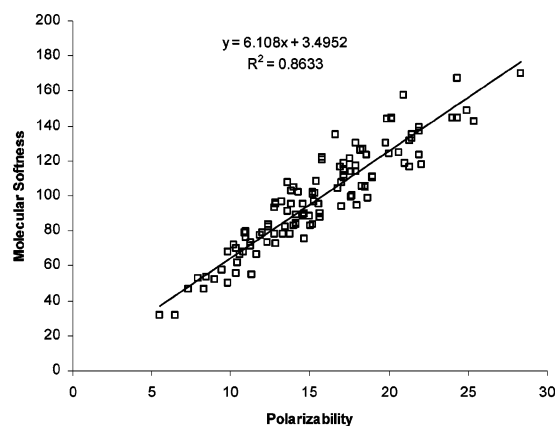
	χ	softness		polariz- ability	PEOE ³²	TAFF (Tripos- like)	Amber 22	Amber 23	CHARMM 16–19	PEF ³⁰ interact.	PEF ³⁰	MMFF 14–16	Engh– Huber ²¹	Emp. MOE 34	OPLS 24
		mol.	av												
HIS_NEU	2.37	87.75	4.18	15.62	2.53	2.41	2.32	2.33	2.43	2.33	2.36	2.42	2.36	2.35	2.34
HIS_Nt	2.36	113.89	4.38	17.6	2.57	2.34	2.35	2.30	2.40	2.31	2.36	2.36	2.36	2.32	2.34
HIS_WSH	2.39	88.56	4.22	14.59	2.68	2.39	2.36	2.35	2.44	2.31	2.38	2.42	2.39	2.33	2.37
HYP_CTR	2.28	126.10	4.67	18.19	2.29	2.19	2.29	2.30	2.29	2.32	2.32	2.18	2.27	2.29	2.24
HYP_Ct	2.27	100.89	4.59	15.19	2.28	2.17	2.29	2.30	2.28	2.33	2.33	2.16	2.27	2.29	2.23
HYP_NEU	2.32	77.60	4.31	11.89	2.36	2.28	2.29	2.30	2.35	2.35	2.35	2.27	2.34	2.35	2.30
HYP_Nt	2.31	102.70	4.47	13.86	2.43	2.25	2.29	2.30	2.32	2.32	2.34	2.25	2.29	2.31	2.29
HYP_WSH	2.32	78.53	4.36	10.85	2.48	2.25	2.29	2.30	2.31	2.34	2.36	2.26	2.29	2.33	2.30
ILE_CTR	2.22	144.58	4.66	20.16	2.23	2.15	2.18	2.20	2.26	2.29	2.30	2.15	2.24	2.26	2.21
ILE_Ct	2.21	118.71	4.57	17.16	2.22	2.13	2.16	2.19	2.24	2.30	2.31	2.13	2.23	2.25	2.19
ILE_NEU	2.24	95.15	4.32	13.86	2.26	2.21	2.18	2.19	2.28	2.31	2.30	2.20	2.26	2.30	2.23
ILE_Nt	2.26	121.82	4.51	15.83	2.35	2.20	2.23	2.21	2.29	2.30	2.32	2.21	2.29	2.28	2.25
ILE_WSH	2.26	95.64	4.35	12.83	2.37	2.19	2.20	2.21	2.29	2.31	2.34	2.21	2.28	2.28	2.25
LEU_CTR	2.22	143.74	4.64	20.16	2.24	2.15	2.18	2.22	2.26	2.30	2.30	2.15	2.24	2.26	2.21
LEU_Ct	2.21	118.26	4.55	17.16	2.22	2.13	2.16	2.21	2.24	2.30	2.31	2.13	2.23	2.25	2.19
LEU_NEU	2.24	94.92	4.31	13.86	2.26	2.20	2.18	2.21	2.28	2.30	2.30	2.20	2.27	2.30	2.23
LEU_Nt	2.26	120.69	4.47	15.83	2.35	2.20	2.23	2.22	2.29	2.30	2.32	2.21	2.29	2.28	2.25
LEU_WSH	2.26	95.22	4.33	12.83	2.37	2.19	2.21	2.23	2.29	2.31	2.33	2.21	2.30	2.28	2.25
LYS_CTR	2.28	157.78	4.64	20.94	2.38	2.23	2.26	2.26	2.31	2.31	2.34	2.23	2.30	2.29	2.27
LYS_Ct	2.29	130.09	4.49	17.93	2.40	2.23	2.27	2.26	2.31	2.32	2.35	2.24	2.31	2.29	2.27
LYS_NEU	2.30	101.95	4.25	15.21	2.31	2.29	2.29	2.30	2.32	2.33	2.32	2.28	2.29	2.33	2.27
LYS_Nt	2.33	135.06	4.50	16.6	2.51	2.29	2.32	2.27	2.35	2.31	2.36	2.31	2.36	2.31	2.32
LYS_WSH	2.35	107.29	4.29	13.6	2.57	2.31	2.33	2.29	2.36	2.33	2.39	2.33	2.38	2.32	2.33
MET_CTR	2.23	131.19	4.52	21.33	2.26	2.15	2.20	2.22	2.27	2.30	2.31	2.16	2.24	2.26	2.22
MET_Ct	2.22	105.79	4.41	18.32	2.24	2.13	2.18	2.20	2.25	2.31	2.32	2.14	2.23	2.25	2.20
MET_NEU	2.25	82.81	4.14	15.02	2.29	2.22	2.21	2.19	2.29	2.31	2.31	2.23	2.28	2.31	2.24
MET_Nt	2.28	107.59	4.30	17	2.40	2.21	2.26	2.21	2.31	2.30	2.33	2.23	2.30	2.29	2.27
MET_WSH	2.28	83.09	4.15	13.99	2.43	2.20	2.23	2.21	2.30	2.31	2.35	2.24	2.31	2.29	2.27
PHE_CTR	2.25	144.69	4.52	24.32	2.27	2.22	2.23	2.23	2.27	2.28	2.30	2.21	2.24	2.25	2.23
PHE_Ct	2.23	116.74	4.32	21.31	2.26	2.21	2.17	2.23	2.25	2.28	2.30	2.20	2.22	2.24	2.22
PHE_NEU	2.26	94.61	4.11	18.01	2.31	2.30	2.20	2.23	2.29	2.28	2.30	2.29	2.26	2.28	2.25
PHE_Nt	2.28	123.92	4.43	19.99	2.41	2.28	2.24	2.23	2.30	2.28	2.32	2.28	2.28	2.28	2.27
PHE_WSH	2.29	94.03	4.09	16.98	2.44	2.30	2.22	2.25	2.30	2.28	2.33	2.30	2.28	2.27	2.28
PRO_CTR	2.22	120.99	4.65	17.55	2.25	2.15	2.19	2.20	2.26	2.30	2.30	2.14	2.23	2.26	2.21
PRO_Ct	2.20	95.14	4.53	14.55	2.23	2.12	2.17	2.18	2.24	2.31	2.31	2.11	2.21	2.25	2.18
PRO_NEU	2.23	70.91	4.17	11.25	2.29	2.21	2.15	2.12	2.30	2.31	2.31	2.21	2.28	2.30	2.23
PRO_Nt	2.27	96.63	4.39	13.22	2.39	2.20	2.28	2.25	2.28	2.30	2.32	2.21	2.26	2.28	2.25
PRO_WSH	2.23	72.10	4.24	10.22	2.42	2.19	2.15	2.12	2.27	2.31	2.34	2.20	2.23	2.28	2.24
SER_CTR	2.28	96.86	4.21	15.3	2.33	2.23	2.25	2.26	2.32	2.33	2.35	2.21	2.30	2.31	2.27
SER_Ct	2.28	73.38	4.08	12.29	2.32	2.21	2.24	2.27	2.31	2.36	2.37	2.20	2.30	2.31	2.26
SER_NEU	2.37	52.05	3.72	8.99	2.45	2.40	2.31	2.29	2.41	2.39	2.38	2.38	2.40	2.41	2.36
SER_Nt	2.36	75.73	3.99	10.96	2.56	2.33	2.36	2.28	2.40	2.35	2.39	2.34	2.39	2.35	2.36
SER_WSH	2.40	52.56	3.75	7.96	2.68	2.37	2.35	2.32	2.42	2.38	2.44	2.39	2.44	2.38	2.39
THR_CTR	2.27	114.68	4.41	17.13	2.30	2.21	2.23	2.25	2.31	2.32	2.34	2.20	2.28	2.29	2.26
THR_Ct	2.26	89.02	4.24	14.13	2.29	2.20	2.22	2.24	2.30	2.35	2.35	2.19	2.27	2.30	2.25
THR_NEU	2.33	67.64	3.98	10.83	2.38	2.33	2.28	2.25	2.37	2.36	2.36	2.32	2.35	2.38	2.32
THR_Nt	2.33	93.01	4.23	12.8	2.48	2.29	2.31	2.26	2.36	2.33	2.37	2.29	2.36	2.33	2.32
THR_WSH	2.35	67.87	3.99	9.79	2.55	2.31	2.31	2.27	2.38	2.36	2.40	2.33	2.39	2.35	2.35
TRP_CTR	2.26	169.99	4.72	28.37	2.30	2.24	2.21	2.26	2.29	2.28	2.31	2.22	2.25	2.26	2.25
TRP_Ct	2.25	142.79	4.61	25.37	2.29	2.25	2.20	2.25	2.28	2.28	2.30	2.22	2.24	2.26	2.24
TRP_NEU	2.28	117.95	4.37	22.07	2.34	2.32	2.22	2.25	2.31	2.28	2.30	2.28	2.27	2.29	2.27
TRP_Nt	2.29	144.53	4.52	24.04	2.41	2.30	2.26	2.26	2.32	2.28	2.32	2.28	2.29	2.28	2.28
TRP_WSH	2.29	118.56	4.39	21.04	2.43	2.31	2.24	2.26	2.31	2.27	2.32	2.29	2.29	2.28	2.28
TYR_CTR	2.27	148.67	4.51	24.95	2.31	2.24	2.23	2.26	2.30	2.30	2.32	2.23	2.27	2.28	2.26
TYR_Ct	2.27	123.07	4.40	21.95	2.30	2.24	2.22	2.26	2.29	2.31	2.33	2.23	2.26	2.28	2.26
TYR_NEU	2.31	98.71	4.11	18.65	2.37	2.34	2.25	2.27	2.34	2.32	2.34	2.33	2.31	2.33	2.30
TYR_Nt	2.32	124.36	4.29	20.62	2.44	2.31	2.29	2.27	2.34	2.31	2.35	2.31	2.33	2.31	2.31
TYR_WSH	2.32	99.08	4.13	17.62	2.47	2.33	2.27	2.29	2.34	2.31	2.36	2.33	2.33	2.31	2.32
VAL_CTR	2.23	126.77	4.53	18.33	2.24	2.16	2.19	2.22	2.26	2.30	2.31	2.15	2.24	2.26	2.22
VAL_Ct	2.21	101.32	4.41	15.32	2.22	2.13	2.16	2.21	2.25	2.31	2.31	2.13	2.23	2.25	2.20
VAL_NEU	2.25	79.05	4.16	12.02	2.27	2.22	2.19	2.20	2.29	2.31	2.31	2.22	2.28	2.31	2.23
VAL_Nt	2.27	104.65	4.36	14	2.38	2.22	2.24	2.23	2.31	2.30	2.33	2.23	2.31	2.29	2.26
VAL_WSH	2.27	79.13	4.17	10.99	2.41	2.21	2.22	2.23	2.30	2.31	2.35	2.23	2.31	2.29	2.26

Calculation of Partial Charges in Proteins. To illustrate the performance of the consensus scale for proteins we have computed ‘inductive’ charges for human Sex Hormone Binding Globulin (SHBG) consisting of 170 amino acids (the Protein Databank³⁶ code 1D2S). To speed up the calculation we introduced the distance cutoff 6 Å for interatomic

distances taken into calculations by (4). The computation of all atomic charges in the entire protein (including H atoms) took ~15 s on the standard 2GHz single processor. We also used the 0.05 convergence factor in the calculation of equalized electronegativities (i.e. iterative procedure stops when the averaged difference in the atomic electronegativities

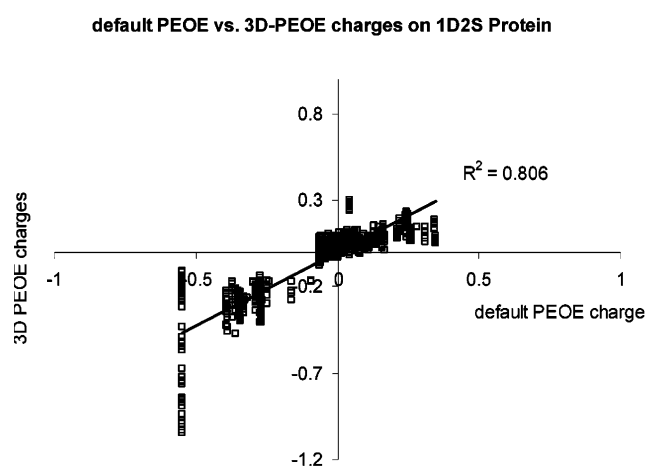
Table 3. Coefficients r^2 of Linear Regressions between Sets of Electronegativities Derived from Conventional Sets of Amino Acids Partial Charges

	PEOE ²⁶	TAFB (Tripos-like) ³⁴	AMBER		CHARMM refs 11–13	PEF interact. ²⁴	PEF ²³	MMFF ^{8–10}	Engh–Huber ¹⁵	empirical MOE ³⁴	OPLS ¹⁸	χ (consensus)
			ref 16	ref 17								
PEOE ²⁶	1											
TAFB (Tripos-like) ³⁴	0.8924	1										
AMBER ¹⁶	0.8961	0.9335	1									
AMBER ¹⁷	0.8525	0.9340	0.9071	1								
CHARMM ^{11–13}	0.9238	0.9696	0.9393	0.9686	1							
PEFinteract ²⁴	0.8869	0.9132	0.9241	0.9392	0.9630	1						
PEF ²³	0.9040	0.9137	0.9058	0.9577	0.9773	0.9804	1					
MMFF ^{8–10}	0.9126	0.9777	0.9014	0.9531	0.9758	0.9223	0.9395	1				
Engh–Huber ¹⁵	0.9303	0.8918	0.9075	0.9097	0.9563	0.9503	0.9575	0.9144	1			
empirical MOE ³⁴	0.8587	0.7811	0.8348	0.7671	0.8490	0.8960	0.8763	0.7783	0.8850	1		
OPLS 18	0.9130	0.9559	0.9114	0.9772	0.9909	0.9547	0.9792	0.9835	0.9470	0.8158	1	
χ (consensus)	0.9334	0.9655	0.9561	0.9694	0.9950	0.9753	0.9807	0.9722	0.9630	0.8701	0.9875	1

**Figure 2.** Dependence between the inductive partial charges in common amino acids and the corresponding charge values computed by the DFT B3LYP 6-31G* method.**Figure 3.** Dependence between polarizability values and calculated molecular softness parameters.

reaches this value). It should also be mentioned, that appropriate distance cutoff parameters can be established by a user empirically to ensure optimal performance by the approach.

The atomic charges were assigned to all atoms of the SHBG by PEOE and CHARMM approaches, as they are implemented within the MOE environment. The resulting dependences between the 3D-sensitive 'inductive' and PEOE charges for 2640 atoms of 1D2S were plotted in Figure 4.

**Figure 4.** Dependence between partial atomic charges in 1D2S protein calculated by the PEOE method and by the developed approach with the PEOE derived effective atomic electronegativities.

The estimated linear trend shown in Figure 4 corresponds to the correlation coefficient $r^2 = 0.7922$. Such satisfactory accuracy demonstrates that the calculated partial charges on atoms of a protein still generally reproduce the PEOE values. All noticeable differences between the two sets of charges could be seen in Figure 4 as vertical bars. Apparently, the PEOE procedure, implemented within the MOE package, equalizes atomic electronegativities only within the individual amino acids and transfers the estimated charges to all amino acids of the same type. Thus, within this approximation, the partial charges of bound atoms do not reflect their actual molecular surrounding and only depend on amino acids they belong to (as in the cases of force field approaches). The most noticeable differences in Figure 4 were observed for the NZ atomic type corresponding to the nitrogen cation in the side chain of lysine. Surprisingly, the PEOE method assigns negative charges for this atomic type, while it has a very high electronegativity value in the χ consensus scale. Similar deviations (and clustering of PEOE charges according to amino acid types) could be observed in Figure 5, reflecting the dependence between calculated inductive charges in the 1D2S protein and the corresponding numbers assigned by the static CHARMM force fields.

The correlation coefficient of this dependence was estimated as $r^2 = 0.7464$. As can be seen from the graphs, the

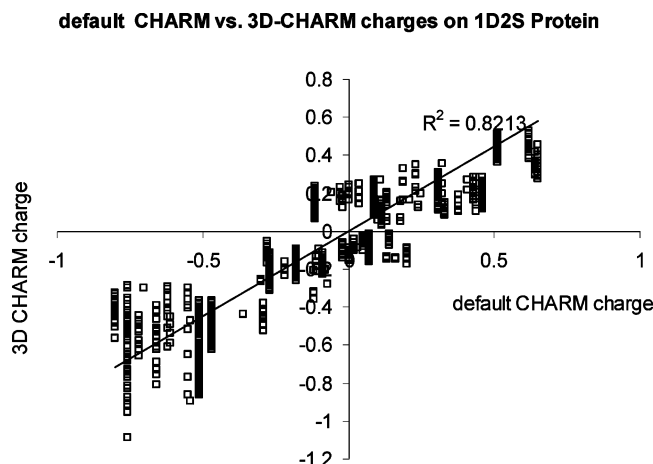


Figure 5. Dependence between partial atomic charges in 1D2S protein assigned by the CHARMM method and calculated with effective atomic electronegativities derived from the CHARMM data.

most noticeable differences between the ‘inductive’ and CHARMM charges also occur for the atomic types corresponding to cationic nitrogens NE and ND1, for which the CHARMM method also assigns uncharacteristic negative q values. Nonetheless, as the result of such qualitative comparisons it could be postulated that there is an overall agreement between the PEOE, CHARMM, and ‘inductive’ partial charges in the studied protein structure.

Use of Protein Inductive Charges in Docking. Human SHBG used in the previous example is a plasma protein that binds biologically active hormones and plays an important role in regulating sex steroid hormone transport and action.^{37–39} Numerous pathological conditions including endometrial cancer, osteoporosis, and some cardiovascular diseases are associated with abnormal levels of SHBG.^{40–44} Preventing the formation of active SHBG-hormone complexes can be expected to increase free sex steroid hormone levels, and this could be exploited therapeutically for the treatment of diseases associated with low endogenous sex hormone bioavailability.³¹ Thus, novel nonsteroidal SHBG ligands represent potentially useful alternatives or adjuncts to current hormone-replacement therapies.

In our previous work,³¹ we utilized a number of conventional ‘in silico’ drug design solutions including pharmacophore search, QSAR modeling with ‘inductive’ descriptors, and protein–ligand docking to identify novel nonsteroidal SHBG ligands. As a result, we identified 29 potential SHBG ligands, eight of which bound to SHBG in vitro (three ligands demonstrated 50% testosterone replacement at low micromolar level). For the discovery of potential drug leads in that study, we utilized several alternative crystal structures of the SHBG that have been used in the docking with the *InterBioScreen*⁴⁵ collection of natural substances. In that study we used MMFF and OPLS partial charges for protein atoms that allowed effective separation of SHBG ligands from the negative control substances. By contrast, we have now compared the accuracy of docking with MMFF, OPLS, and several other conventional charging systems such as CHARMM, AMBER, and PEOE with docking using ‘inductive’ partial charges. The comparative docking analysis involved three alternative structures of SHBG – 1D2S, 1KDM and 1LHV, which are presented in superimposed form in Figure 6.

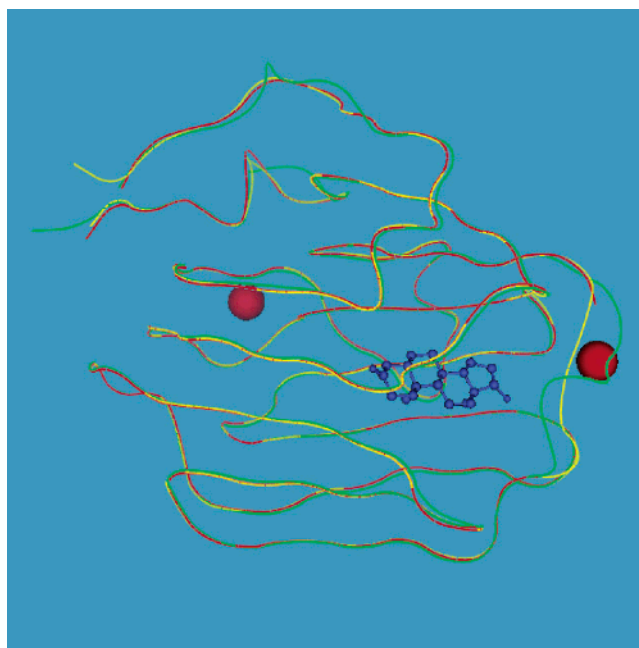


Figure 6. Superimposed structures of SHBG corresponding to 1D2S (red), 1KDM (green), and 1LHV (yellow).

As can be clearly seen from the chart, three SHBG structures reflect possible configurations of a flexible protein loop ‘gating’ the steroid binding pocket.⁴⁷ This flexible ‘gating’ loop represents one of the major challenges for docking studies with SHBG structures, along with the presence of a Zn^{2+} ion in close proximity to the loop, which may cause its distortion and alter SHBG ligand binding specificity.⁴⁸ Taking these factors into consideration, and to capture all possible implications of SHBG structural diversity, we conducted a docking study on three PDB entries that differ in the configuration of the ‘gating’ loop (see Figure 6).

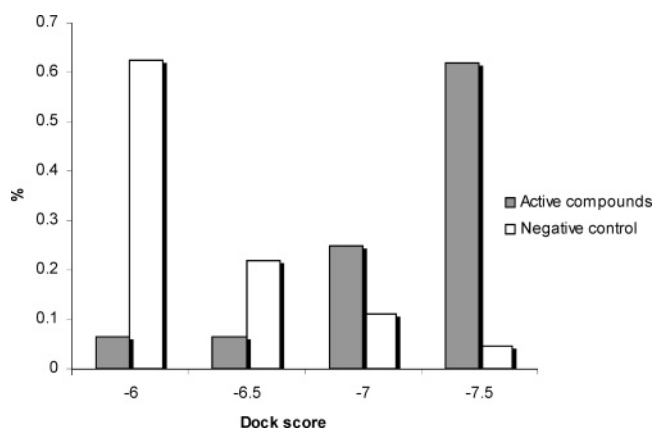
Thus, we used the Glide 2.7 package⁴⁹ to conduct 18 independent docking studies corresponding to three alternative configurations of the SHBG protein and six different partial charge systems. For the purpose of evaluating the docking accuracy, we compiled a set of 239 compounds that included numerous substances with known affinity toward the SHBG as well as a number of randomly selected organic molecules with no known SHBG binding. The data set covered the steroid benchmark used in various molecular modeling studies, and the role of polar effects in defining steroid-SHBG affinity is broadly acknowledged (for the latest examples see refs 50–59). Numerous nonsteroidal natural derivatives that demonstrated micromolar binding to SHBG have also been included into the data set as well as three nonsteroidal SHBG ligands identified in the previous study. Initially, the ‘negative control’ substances have been randomly derived from the ACD library of available chemicals. As the ‘negative control’ (compounds that do not exhibit investigated activity) we selected those 110 ACD molecules that passed two consecutive rounds of steroid similarity search with the use of 0.8 Tanimoto coefficients calculated on bit-packed MACCS Structural Key⁶¹ fingerprints as it is implemented within the MOE package.

We calculated PM3 atomic partial charges for all 239 substances under investigation and used them for docking with 18 protein structures (three protein configurations with six modifications of atomic partial charges) using Glide 2.7.⁴⁹

Table 4. Accuracy of Docking of 239 Studied Compounds into the SHBG Protein Using “Inductive” and Conventional Protein Charging Systems^a

parameters/system	CHARMM	AMBER	OPLS	PEOE	MMFF	IND
total true positives	77	72	85	61	79	57
total false positives	80	54	102	30	81	24
total true negatives	538	564	516	588	537	594
total false negatives	16	21	8	32	14	36
sensitivity	0.828	0.774	0.914	0.656	0.849	0.613
specificity	0.871	0.913	0.835	0.951	0.869	0.961
accuracy	0.865	0.895	0.845	0.913	0.866	0.916
PPV	0.490	0.571	0.455	0.670	0.494	0.704

^a The results have been averaged using 1D2S, 1KDM, and 1LHV PDB structures.

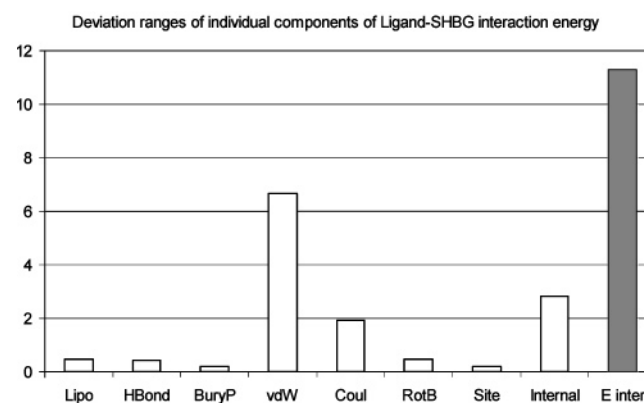
**Figure 7.** Separation of active and nonactive compounds by docking with the Glide package and using inductive charges for protein atoms.

‘Inductive’ charges calculated from the consensus ‘inductive’ electronegativity scale can be found in supplement 1 file in the Supporting Information. The docking results for all protein structures and charging systems are collated in supplement file 2 in the Supporting Information.

The docking outputs for all 18 studied SHBG structures were classified using the GlideScore = -7.5 and $\log K_a = -8.5$ thresholds (all association constants can be found in supplement 2 file in the Supporting Information). Based on these threshold parameters all real/predicted binders and nonbinders were then classified, and the corresponding confusion matrices were created. The estimated parameters of true- and false-positives and negatives as well as the resulting specificity, sensitivity, accuracy, and positive predictive values (PPV) are presented in Table 4 for all 18 docking experiments.

As can be seen from the data, all six charging systems resulted in generally accurate identification of effective SHBG binders for all three protein structures studied. Notable, the PEOE and ‘inductive’ partial charges produced the most results (with the slight advantage of the ‘inductive’ system). Moreover, the use of 3D sensitive ‘inductive’ partial charges resulted in the best positive predicting powers (PPV) for all three protein structures. This is an important development because high PPV represents one of the most important requirements for large scale ‘in silico’ lead discovery. The separation of sufficient from nonsufficient SHBG binders ($\log K_a = -8.5$ threshold) by the docking with ‘inductive’ charges is illustrated by Figure 7 for the 1D2S structure.

Thus, the docking study with 1D2S, 1KDM, and 1LHV proteins and the extended set of steroid substances (with negative control) demonstrated superior performance by

**Figure 8.** Deviation ranges of individual components of ligand-SHBG interaction energy.


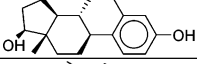
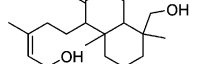
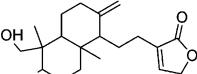
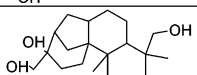
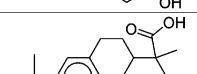

‘inductive’ partial charges. Docking with ‘inductive’ q also turned out to be slightly more accurate than docking with PEOE charges, as the total number of false positive predictions in the case of ‘inductive’ charges is significantly lower than for all other systems.

Analysis of individual components of protein–ligand docking energy demonstrated that variability in SHBG binding affinities of investigated steroids and steroid-like compounds may be attributed mainly to the van der Waals and Coulomb interactions. Thus, Figure 8 features the established deviation ranges for individual components to the docking energy in the ligand-SHBG system.

As can be seen from the graph, within the series of 239 studied compounds the standard deviation of overall docking energy approaches 12 kcal/mol, while the variability of the electrostatic component is close to 2 kcal/mol (the corresponding column is labeled as ‘Coul’). This may indicate that electrostatic interactions play an important role in defining such a broad range of SHBG-steroid binding affinities. It is feasible to assume that ‘inductive’ charges allow more accurate quantification of protein–ligand electrostatic interactions and therefore result in better differentiation of overall binding energies. Taking all of this into account, we decided to use ‘inductive’ charges on the SHBG protein to conduct the docking-based search for novel nonsteroidal SHBG ligands.

Identification of a Novel SHBG Inhibitor. We utilized the all-atom representation of the 1KDM protein structure with the assigned ‘inductive’ charges for a large scale docking study. We conducted flexible docking of $>20\,000$ natural molecules into the 1KDM active site using Glide 2.7. All atomic charges in the natural compounds investigated were calculated using the PM3 system implemented in the

Table 5. Bioactivity Data for Top 5 Lead Compounds

Compound	Structure	Binding	IC ₅₀ (μM)
Testosterone		Yes	0.016
Estradiol		Yes	0.062
NP-000680		Yes	70.54
NP-011179 (AD04130-4)		No	N/A
NP-009930 (AD04130-3)		Some	N/A
NP-002802 (AD04130-2)		Some	N/A
NP-010148 (AD04130-5)		No	N/A

MOE package. The docking was carried out with the batch mode on Glide and took approximately 6800 min on the single CPU time on a 2GHz Linux station. As a result, 5 top structures were identified for experimental testing (the formula of the selected compounds can be seen in Table 5).

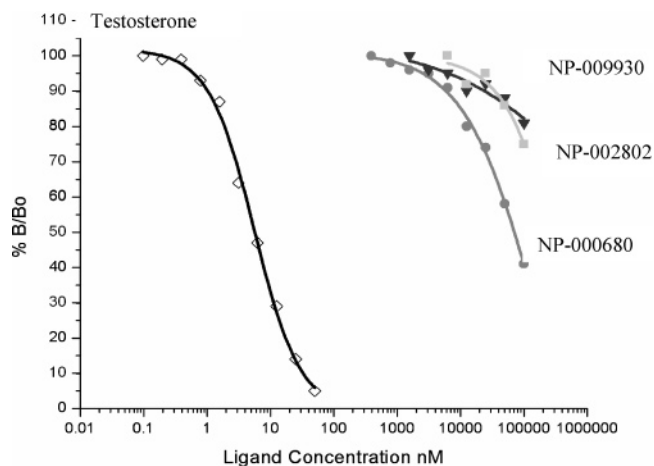
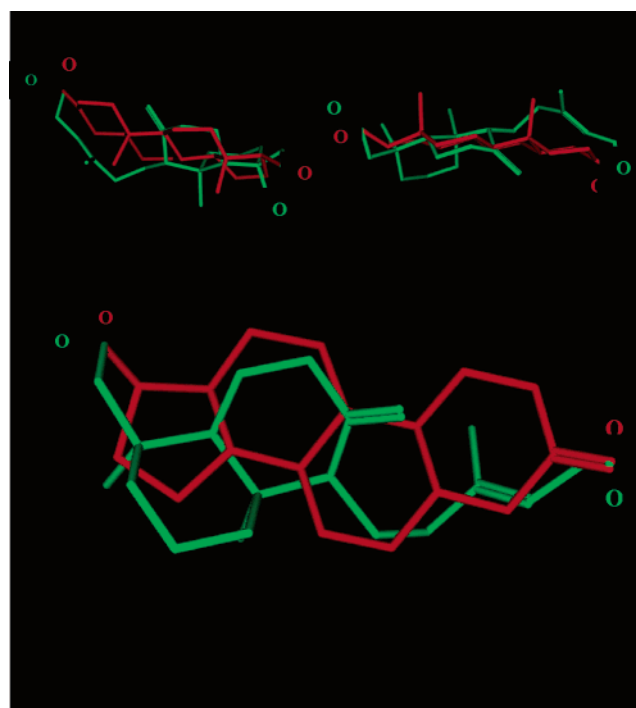
The experimental testing of compounds as potential SHBG binders was carried out using diluted human serum, as described previously,³¹ and 3 out of 5 substances tested demonstrated binding affinity toward human SHBG. Details of the experimental protocol can be found in the "Materials and Methods" section. The competition curve for three active substances is presented in Figure 8 along with testosterone, and this represents their relative affinities for the SHBG steroid-binding site, as defined by their abilities to displace tritium-labeled 5 α -dihydrotestosterone (DHT) from the binding site.

As Figure 8 illustrates, three test substances demonstrated binding to the SHBG protein, but only one compound AD001 displaced >50% of the [³H]DHT from the binding site. This allowed us to calculate the IC₅₀ value for AD001 as 70 μM, which is only approximately 4 orders of magnitude weaker SHBG binder as compared to natural sex steroids such as testosterone and estradiol (see more data in Table 5). The structure of AD001 superimposed with the best known SHBG binder dihydrotestosterone (DHT) is illustrated in Figure 9.

As the chart illustrates, the identified substance demonstrates good steric and polar complementarity to the steroid structure.

Thus, the conducted HTS study involving SHBG protein and its substrates demonstrated superior performance of docking with 'inductive' partial charges compared to such conventional protein charging systems as CHARMM, AMBER, MMMFF, OPLS, and PEOE. Moreover, the docking with 'inductive' parameters allowed identification of novel non-steroidal SHBG ligand that may have therapeutic potential.

In summary, the electronegativity equalization procedure we have developed together with the consensus 'inductive' electronegativity scale have enabled rapid, 3D sensitive

**Figure 9.** Substitution curve for the DHT replacement from SHBG in human serum.**Figure 10.** Configuration of NP-000680 compound (green) superimposed with DHT structure (red). Views from the front and from the sides.

calculation of atomic partial charges within formula (1), and we have demonstrated their usefulness in docking. To further improve our methodology, we are currently refining the charge calculating approach by introducing additional atomic types reflecting the finer details of protein intramolecular interactions.

CONCLUSIONS

On the basis of the previously developed models for inductive and steric substituent constants and inductive electronegativity scale, we have developed a new procedure for calculating atomic partial charges in proteins. The method utilizes effective electronegativity scales calibrated by partial charges assigned to protein atoms by several common conventional methods. The iterative method we have developed allows calculation of partial charges in proteins, within formats of conventional force field and electronegativity

equalization methods, and offers a consensus effective electronegativity scale reflecting the average ability of protein atomic types to attract/donate electrons. The use of 3D-sensitive 'inductive' charges in the docking study involving the 1D2S protein, and the extended set of steroid structures, not only demonstrated their superior performance but also resulted in discovery of a novel nonsteroidal ligand for human SHBG that may have therapeutic potential.

As a cautionary note it should also be mentioned that the developed approach has only been tested on docking results so far, and therefore the presented protein charge calculating approach and 'protein electronegativity scale' should generally be recommended for docking simulations at this point.

MATERIALS AND METHODS

Docking. The Maestro suite⁶⁰ was used to prepare protein structures for docking, which was conducted using the Glide 2.7 program⁴⁹ with default settings. The protein charges were assigned to the MMFF94 molecular mechanics force field, CHARMM22 partial charges, AMBER89, OPLS-AA, and PEOE (Gasteiger) charge system as they are implemented within the MOE package. The 'inductive' charges have been calculated by formula (1) with effective electronegativity values from Table 2 and using the SVL script for the MOE package.

All protein preparation and grid generation procedures have been carried out with the Maestro package with default settings. The SHBG active site has been defined as 10 Å surrounding of the bound ligand. The docking has been conducted within the Glide 2.7 program using the flexible mode and using the Glide batch with the default setting, except for the partial charges that have been set to the "user-defined".

SHBG Ligand-binding Assay. An established competitive ligand binding assay was used to determine the binding affinities of test compounds to human SHBG in relation to those of testosterone.⁶² In brief, the assay involved mixing 100 μ L aliquots of diluted (1:200) human pregnancy serum containing approximately 1 nM SHBG, which was pretreated with dextran-coated charcoal (DCC) to remove the endogenous steroid ligand, with 100 μ L of tritium labeled DHT (³H] DHT) at 10 nM as the labeled ligand. For the screening assay, triplicate aliquots (100 μ L) of a fixed amount (200 μ M) of a test compound was added to this SHBG/³H] DHT mixture and incubated overnight at room temperature, followed by a 10 min incubation with 500 μ L of a DCC slurry at 0 °C and centrifugation to separate SHBG-bound from free ³H] DHT. Compounds that displaced more than 35% of the ³H] DHT from the SHBG in this assay were then diluted serially, and triplicate aliquots (100 μ L) of known concentrations of test compounds were used to generate complete competition curves by incubation with the SHBG/³H]DHT mixture, and separation of SHBG-bound from free ³H]DHT, as in the screening assay. The amounts of ³H] DHT bound to SHBG at each concentration of the competitor ligand were determined by scintillation spectrophotometry and plotted in relation to the amount of ³H] DHT bound to SHBG at zero concentration of the competitor. From the resulting competition curves, IC₅₀ concentrations could be calculated if displacement of more than 50% of ³H] DHT from SHBG was achieved.

ACKNOWLEDGMENT

G.L.H. is a Canada Research Chair in Reproductive Health and is supported by operating grants from the Canadian Institutes of Health Research. Y.L. acknowledges the support of the CIHR/MSFHR Strategic Training Program in Bioinformatics (<http://bioinformatics.bcgsc.ca>).

Supporting Information Available: (a) Inductive and DFT charges for the studied 23 amino acids, (b) CHARMM and 'inductive' charges for 1D2S protein, and (c) PEOE and 'inductive' charges for 1D2S protein (supplement file 1) and docking results for 1D2S, 1KDM, and 1LNV structures with 'inductive', CHARMM, PEOE, MMFF, AMBER, and OPLS charges (supplement file 2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- (1) Antosiewicz, A.; McCammon, J. A. The determination of pK_as in proteins. *Biochemistry* **1996**, *35*, 78197833.
- (2) Berioza, P.; Fredkin, D. R. Calculation of amino acid pK_as in a protein from a continuum electrostatic model: method and sensitivity analysis. *J. Comput. Chem.* **1996**, *17*, 1229–1244.
- (3) Matthew, J. B.; Gurd, F. R. N. Calculation of electrostatic interactions in proteins. *Methods Enzymol.* **1986**, *130*, 413–436.
- (4) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.
- (5) *Structural Bioinformatics*; Bourne, P. E., Weissig, H., Eds.; Wiley & Sons: 2003; p 572.
- (6) Vereschagin, A. N. *Induktivnyi efekt (Inductive Effect)*; Nauka: Moscow, 1987; p 325.
- (7) Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A. Inductive effect of substituents in the correlation analysis. The problem of quantitative estimation. *Uspekhi Khimii (Russ. Chem. Rev.)* **1996**, *65*, 641–711.
- (8) Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (9) Halgren, T. A. Merck molecular force field .2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520–552.
- (10) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
- (11) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenker, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of protein. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (12) Feller et al. *Biophys. J.* **1997**, *73*, 2269.
- (13) MacKerell, A. D., Jr. Molecular dynamics simulation analysis of a sodium dodecyl sulfate micelle in aqueous solution: Decreased fluidity of the micelle hydrophobic interior. *J. Phys. Chem.* **1995**, *99*, 1846–1855.
- (14) Roland H.; Karplus, S.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 12.
- (15) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr.* **1991**, *A47*, 392–400.
- (16) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (17) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force-field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
- (18) Jorgensen, W. L.; Tirado-Rives, J. Development of the OPLS-AA force field for organic and biomolecular systems. *Abstr. PAP Am. Chem. Soc.* **1998**, S216.
- (19) Williams, D. E.; Yan, J. M. *Adv. Atom. Mol. Phys.* **1998**, *23*, 87.
- (20) Williams, D. E. Alanyl dipeptide potential-derived net atomic charges and bond dipoles and their variation with molecular conformation. *Biopolymers* **1990**, *29*, 1367–1386.
- (21) Exner, T. E.; Mezey, P. G. Comparison of nonlinear transformation methods for electron density approximation. *J. Phys. Chem. A* **2002**, *106*, 5504–5509.
- (22) Baeten, A.; Geerlings, P. The use of electronegativity equalization principle to study charge distributions in enzymes: application to dipeptides. *J. Mol. Struct. (THEOCHEM)* **1999**, *465*, 203–207.

- (23) Wang, C.-S.; Li, S.-M.; Yang, Z.-Z. Calculation of molecular energies by atom-bond electronegativity equalization method. *J. Mol. Struct. (THEOCHEM)* **1998**, *430*, 191–199.
- (24) Fabricius, J.; Engelsen, S. B.; Rasmussen, K. *J. Carbohydr. Chem.* **1997**, *16*.
- (25) Shankar, S.; Mortier, W. J.; Ghosh, S. K. Calculation of atomic charges in large molecules. *Ann. N. Y. Acad. Sci.* **1986**, *482*, 82–84.
- (26) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (27) Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A. "Inductive" Electronegativity Scale. *J. Mol. Struct. (THEOCHEM)* **1999**, *489*, 43–46.
- (28) Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A. "Inductive" Electronegativity Scale. 2. "Inductive" Analogue of Chemical Hardness. *J. Mol. Struct. (THEOCHEM)* **2000**, *497*, 115–123.
- (29) Cherkasov, A. Inductive Electronegativity Scale. Iterative Calculation of Inductive Partial Charges. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2039–2047.
- (30) Cherkasov, A. 'Inductive' Descriptors. 10 Successful Years in QSAR. *Curr. Comput.-Aided Drug Des.* **2005**, *1*, 21–42.
- (31) Cherkasov, A.; Shi, Z.; Fallahi, M.; Hammond G. L. Successful in Silico Discovery of Novel Non-Steroidal Ligands for Human Sex Hormone Binding Globulin (SHBG). *J. Med. Chem.* **2005**, *48*, 3203–3213.
- (32) Cherkasov, A. R.; Galkin, V. I.; Cherkasov, R. A. A New Approach to the Theoretical Estimation of Inductive Constants. *J. Phys. Org. Chem.* **1998**, *11*, 437–447.
- (33) Gilson, M. K.; Gilson, H. S. R.; Potter M. J. Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method that Accounts for Alternate Resonance Forms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1982–1997.
- (34) *Molecular Operational Environment*; Chemical Computation Group Inc., Montreal, Canada, 2003.
- (35) *HyperChem 7.5 for Windows*; by HyperCube Inc., 2002.
- (36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein databank. *Nucl. Acids Res.* **2000**, *28*, 235–242.
- (37) Hammond, G. L. Potential functions of plasma steroid-binding protein. *Trends Endocrinol. Metab.* **1995**, *6*, 298–304.
- (38) Joseph, D. R. Structure, function, and regulation of androgen-binding protein/sex hormone-binding globulin. *Vitam. Horm.* **1994**, *49*, 197–280.
- (39) Siiteri, P. K.; Murai, J. T.; Hammond, G. L.; Niskier, J. A.; Raymoure, W. J.; Kuhn, R. W. The serum transport of steroid hormones. *Recent Prog. Horm. Res.* **1982**, *38*, 457–510.
- (40) Niskier, J. A.; Hammond, G. L.; Davidson, B. J.; Frumar, A. M.; Takaki, N. K.; Judd, H. L.; Siiteri, P. K. Serum sex hormone-binding globulin capacity and the percentage of free estradiol in postmenopausal women with and without endometrial carcinoma. A new biochemical basis for the association between obesity and endometrial carcinoma. *Am. J. Obstet. Gynecol.* **1980**, *138*, 637–642.
- (41) Hogeveen, K. N.; Cousin, P.; Pugeat, M.; Dewailey, D.; Soudan, B.; Hammond, G. L. Human sex hormone-binding globulin variants associated with hyperandrogenism and ovarian dysfunction. *J. Clin. Invest.* **2002**, *109*, 973–981.
- (42) Anderson, D. C. Sex-hormone-binding globulin. *Clin. Endocrinol.* **1974**, *3*, 69–96.
- (43) Van Pttelgergh, I.; Goemaere, S.; Zmierczak, H.; Kaufman, J. M. Perturbed sex steroid status in men with idiopathic osteoporosis and their sons. *J. Clin. Endocrinol. Metab.* **2004**, *89*, 4949–4953.
- (44) Rapuri, P. B.; Gallagher, J. C.; Haynatzki, G. Endogenous levels of serum estradiol and sex hormone binding globulin determine bone mineral density, bone remodeling, the rate of bone loss, and response to treatment with estrogen in elderly women. *J. Clin. Endocrinol. Metab.* **2004**, *89*, 4954–4962.
- (45) Lindstedt, G.; Lundberg, P.-A.; Lapidus, L.; Lundgren, L.; Björntrop, P. Low sex-hormone-binding globulin concentration as independent risk factor for development of NIDDM. 12-yr follow-up of population study of women in Gothenburg, Sweden. *Diabetes* **1991**, *40*, 123–128.
- (46) *InterBioScreen* collection: <http://www.ibscreen.com/>
- (47) Avvakumov, G. V.; Muller, Y. A.; Hammond, G. L. Steroid-binding specificity of human sex hormone-binding globulin is influenced by occupancy of a zinc-binding site. *J. Biol. Chem.* **2000**, *275*, 25920–25925.
- (48) Hammond, G. L.; Avvakumov, G. V.; Muller, Y. A. Structure/function analyses of human sex hormone-binding globulin: effects of zinc on steroid-binding specificity. *J. Steroid Biochem. Mol. Biol.* **2003**, *85*, 195–200.
- (49) *Glide*; Version 2.7, Schrödinger Inc., San Diego, 2004.
- (50) Tuppurainen, K.; Viisas, M.; Perakyla, M. et al. Ligand intramolecular motions in ligand-protein interaction: ALPHA, a novel dynamic descriptor and a QSAR study with extended steroid benchmark dataset. *J. Comput.-Aided Drug Des.* **2004**, *18*, 175–187.
- (51) Asikainen, A. H.; Ruuskanen, J.; Tuppurainen, K. A. Performance of (consensus) kNN QSAR for predicting estrogenic activity in a large diverse set of organic compounds. *SAR QSAR Environ. Res.* **2004**, *15*, 19–32.
- (52) Korhonen, S. P.; Tuppurainen, K.; Laatikainen, R.; Perakyla, M. FLUFF-BALL, a template-based grid-independent superposition and QSAR technique: validation using a benchmark steroid data set. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1780–1793.
- (53) Liu, S. S.; Yin, C. S.; Wang, L. S. Combined MEDV-GA-MLR method for QSAR of three panels of steroids, dipeptides, and COX-2 inhibitors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 749–756.
- (54) Evaluation of a novel electronic eigenvalue (EEVA) molecular descriptor for QSAR/QSPR studies: validation using a benchmark steroid data set. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 607–613.
- (55) Liu, S. S.; Yin, C. S.; Li, Z. L.; Cai, S. X. QSAR study of steroid benchmark and dipeptides based on MEDV-13. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 321–329.
- (56) Polanski, J.; Walczak, B. The comparative molecular surface analysis (COMSA): a novel tool for molecular design. *Comput. Chem.* **2000**, *24*, 615–625.
- (57) Turner, D. B.; Willett, P.; Ferguson, A. M.; Heritage, T. W. Evaluation of a novel molecular vibration-based descriptor (EVA) for QSAR studies: 2. Model validation using a benchmark steroid dataset. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 271–296.
- (58) Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G. Self-organizing molecular field analysis: a tool for structure–activity studies. *J. Med. Chem.* **1999**, *42*, 573–583.
- (59) Jain, A. N.; Koile, K.; Chapman, D. Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark. *J. Med. Chem.* **1994**, *37*, 2315–2327.
- (60) *Maestero*; Schrödinger Inc., San Diego, 2004.
- (61) Brown, R. D.; Martin, Y. C. Use of Structure–Activity Data to Compare Structure-Based Clustering. Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572–584.
- (62) Hammond, G. L.; Lahteenmaki, P. L. A versatile method for the determination of serum cortisol binding globulin and sex hormone binding globulin binding capacities. *Clin. Chim. Acta* **1983**, *132*, 101–110.

CI0498158