Is Quantum Mechanics Necessary for Predicting Binding Free Energy?

Ting Zhou, Danzhi Huang,* and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received March 6, 2008

To take into account polarization effects, the linear interaction energy model with continuum electrostatic solvation (LIECE) is supplemented by the linear-scaling semiempirical quantum mechanical calculation of the intermolecular electrostatic energy (QMLIECE). QMLIECE and LIECE are compared on three enzymes belonging to different classes: the West Nile virus NS3 serine protease (WNV PR), the aspartic protease of the human immunodeficiency virus (HIV-1 PR), and the human cyclin-dependent kinase 2 (CDK2). QMLIECE is superior for 44 peptidic inhibitors of WNV PR because of the different amount of polarization due to the broad range of formal charges of the inhibitors (from 0 to 3). On the other hand, QMLIECE and LIECE show similar accuracy for 24 peptidic inhibitors of HIV-1 PR (20 neutral and 4 with one formal charge) and for 73 CDK2 inhibitors (all neutral). These results indicate that quantum mechanics is essential when the inhibitor/protein complexes have highly variable charge−charge interactions.

1. Introduction

Accurate methods for computing the binding affinity between small molecules and proteins are needed for drug discovery and design.[1] Approaches based on ab initio quantum mechanics (QM) are rigorous but slow for the studies of macromolecules of biological interest. In order to accelerate QM calculations the hybrid QM/molecular mechanics method[2,3] has been developed for the study of enzyme catalysis.[4–7] In addition linear-scaling QM approaches[8–11] have also been applied extensively for the evaluation of binding affinity between small molecules and proteins.[12,13] Because of their first principle nature, both the time-consuming ab initio methods[14] and fast semiempirical methods[15,16] do not suffer of the approximation inherent to the ball-spring description and the fix-charge approach used in the force field method. Raha and Merz[17] developed a semiempirical/linearly scaling QM-based scoring function and studied the ion-mediated ligand binding processes. They pointed out that QM is needed for metal-containing system because the ill-defined atom types of metal atoms in most of the force field parameters cannot describe the nature of the interactions between a small molecule and a metal ion in the active site. Nevertheless, even with the fast semiempirical/linearly scaling methods, QM approaches are still time-consuming compared with force field based methods especially for high throughput docking.[18,19] Moreover, most of the QM methods significantly underestimate the weak London dispersion forces which require highly correlated theoretical levels and large basis sets.[20] These weak forces, however, play a major role in the hydrophobic effect, molecular recognition, and ligand binding.[21–23] Therefore, it is important to find an optimal compromise between accuracy and efficiency in binding affinity calculation for high-throughput docking campaigns of multimillion library of compounds.

Recently, the linear interaction energy model with continuum electrostatic solvation (LIECE[a])[24] has been successfully applied in high-throughput docking resulting in the discovery of inhibitors of proteases[25,26] and kinases.[27] LIECE is about 2 orders of magnitude faster than the original LIE[28–30] method because molecular dynamics (MD) sampling is replaced by a simple energy minimization. In this paper, LIECE is further improved by using a linearly scaling semiempirical QM method[9] to calculate the electrostatic interaction energy between the ligand and the protein, and the new approach is termed

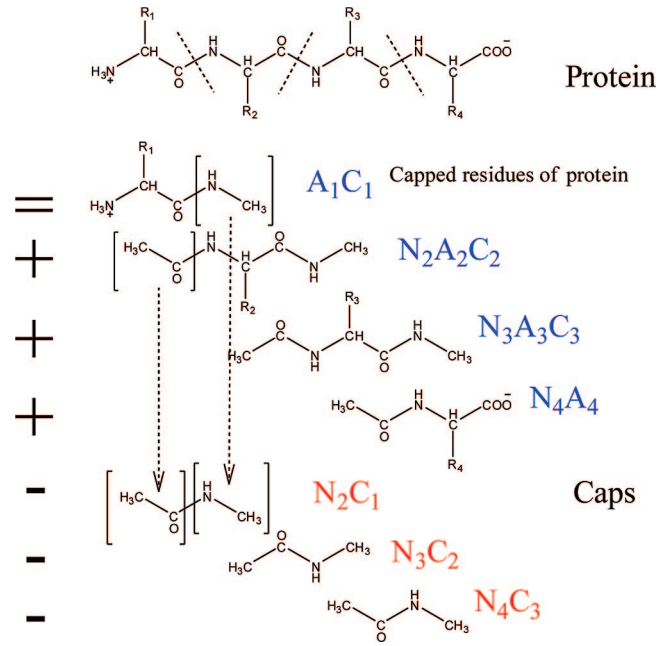

Figure 1. Divide and conquer protocol[9] for calculation of quantum mechanical interaction energy between a protein and a small molecule (ligand). The interaction energy between a protein with *m* residues and the ligand is decomposed into

$$E_{\text{ligand-protein}} = E_{\text{ligand-A}_1\text{C}_1} + E_{\text{ligand-N}_2\text{A}_2\text{C}_2} + \ldots + E_{\text{ligand-N}_{m-1}\text{A}_{m-1}\text{C}_{m-1}} + E_{\text{ligand-N}_m\text{A}_m} - E_{\text{ligand-N}_2\text{C}_1} - E_{\text{ligand-N}_3\text{C}_2} - \ldots - E_{\text{ligand-N}_m\text{C}_{m-1}}$$

where N_i and C_i are N-terminal and C-terminal cap, respectively, of residue A_i. The fragments with blue names are protein residues with conjugate caps,[10] while the ones with red names are pure "caps" that have to be subtracted to remove the duplication in energy calculation.

* To whom correspondence should be addressed. (D.H.) Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62. E-mail: huang@bioc.uzh.ch. (A.C.) Phone: (+41 44) 635 55 21. Fax: (+41 44) 635 68 62. E-mail: caflisch@bioc.uzh.ch.

[a] Abbreviations: LIECE, linear interaction energy model with continuum electrostatic solvation; QMLIECE, quantum mechanical linear interaction energy model with continuum electrostatic solvation; CDK2, cyclin-dependent kinase 2; HIV-1 PR, human immunodeficiency virus protease; WNV PR, West Nile virus NS3 serine protease.

10.1021/jm800242q CCC: $40.75
Published on Web 06/25/2008

Table 1. 44 Peptidic Inhibitors of WNV PR Tested at the Novartis Institute for Tropical Diseases[34–36,a]

ID	structure	no. of formal charges	IC_{50} (μm)	ΔG (kcal/mol)
1	Bz-Nle-Lys-Arg-Arg-H	3	4.1	−7.39
2	Bz-Nle-Lys-Lys(Z)-Arg-H	2	99.5	−5.49
3	Bz-Nle-Lys-Gln-Arg-H	2	1.7	−7.90
4	Bz-Nle-Lys-Lys-Arg-H	3	1.9	−7.86
5	Ac-Lys-Lys-Arg-H	3	**0.4**	−8.84
6	Bz-Lys-Arg-Arg-H	3	1.5	−7.99
7	Bz-Lys-Lys(Tos)-Arg-H	2	117.9	−5.39
8	Ac-Lys-Lys(Tos)-Arg-H	2	**463.4**	−4.58
9	Ac-Lys-Lys(Bz)-Arg-H	2	116.5	−5.40
10	Bz-Lys-(p-Me)Phe-Arg-H	2	194.3	−5.09
11	Bz-Lys-Lys(Bz)-Arg-H	2	68.1	−5.72
12	indole-Lys-Arg-Arg-H	3	2.4	−7.72
13	Bz-Lys-Asn-Arg-H	2	71.8	−5.69
14	Bz-Nle-Ala-Arg-Arg-H	2	3.8	−7.44
15	Bz-Ala-Lys-Arg-Arg-H	3	0.7	−8.45
16	Bz-Nle-Lys-Arg-Phe-H	2	109.8	−5.43
17	Bz-Nle-Lys-Phe-Arg-H	2	108.0	−5.44
18	Bz-Nle-Phe-Arg-Arg-H	2	4.2	−7.38
19	Bz-Phe-Lys-Arg-Arg-H	3	1.2	−8.14
20	Bz-Lys-Arg-Tyr-H	2	14.6	−6.64
21	Ac-KRR-H	3	0.5	−8.60
22	pyridine-KRR-H	3	0.8	−8.40
23	isoquinoline-KRR-H	3	0.6	−8.56
24	pyrazine-Lys-Arg-Arg-H	3	1.1	−8.18
25	3-pyridyl-KRR-H	3	1.0	−8.24
26	Bzl-Nle-Lys-Arg-(4-CN)-Phe-H	2	62.0	−5.77
27	Bzl-Nle-Lys-Arg-Trp-H	2	10.0	−6.86
28	Bz-Nle-Lys-Arg-Lys-H	3	57.7	−5.82
29	BZ-Nle-Lys-Arg-(4-guanidinyl)-Phe-H	3	11.8	−6.76
30	Bz-Nle-Lys-Arg-His-H	2	43.1	−5.99
31	Bz-Nle-Lys-Arg-Phg-H	2	90.9	−5.55
32	Bz-Arg-Arg-H	2	3.9	−7.42
33	Bz-Lys(Z)-Arg-H	1	436.3	−4.61
34	Bz-Lys-Arg-H	2	1.4	−8.03
35	Bz-Arg-Lys-H	2	57.5	−5.82
36	Bz-Lys(Z)-(2-napthyl)Ala-H	0	15.9	−6.59
37	Bz-Lys(Z)-Tyr(Bn)-H	0	17.2	−6.54
38	Bz-Tyr(Bn)-(p-Me)Ph-H	0	12.7	−6.72
39	Bz-Lys(Z)-(p-NH-1-isoquinoline)Phe-H	0	18.0	−6.51
40	Bz-Lys(Z)-(p-NH-1-indole)Phe-H	0	38.2	−6.06
41	Bz-Lys-Lys(Z)-(p-Me)Phe-H	1	43.3	−5.99
42	Bz-Lys-Arg-Phe-H	2	71.1	−5.69
43	Bz-Lys-Arg-(P-Me)Phe-H	2	17.7	−6.52
44	Bz-Lys-Arg-Tyr(Bn)-H	2	11.8	−6.76

[a] Weakest and strongest affinities are in bold.

Table 2. Energy Components for WNV PR Peptidic Inhibitors[34–36,a]

ID	ΔE_{vdW}	$\Delta E_{elec,coul}$	ΔE_{QM}	$\Delta G^{solvation}$	no. of formal charges
1	−45.2	−898.0	−856.8	846.9	3
2	−53.2	−603.1	−511.1	608.9	2
3	−41.1	−601.1	−516.6	566.1	2
4	−40.9	−880.5	−874.2	840.1	3
5	−38.4	−926.0	−950.3	883.2	3
6	−41.2	−899.4	−860.7	842.0	3
7	−46.3	−621.8	−530.3	616.5	2
8	−45.8	−617.6	−530.2	616.0	2
9	−48.3	−593.8	−513.1	601.3	2
10	−43.7	−583.8	−511.5	558.0	2
11	−47.2	−596.7	−503.7	595.3	2
12	−36.9	−881.8	−874.0	839.0	3
13	−35.7	−589.4	−506.0	556.0	2
14	−46.3	−620.5	−530.6	565.0	2
15	−44.9	−898.5	−862.2	845.9	3
16	−54.2	−654.5	−565.1	621.2	2
17	−43.5	−580.6	−510.4	556.6	2
18	−53.9	−623.6	−531.1	568.3	2
19	−45.9	−898.1	−860.2	846.0	3
20	−52.8	−657.1	−565.0	616.9	2
21	−39.3	−898.7	−869.7	847.2	3
22	−40.9	−905.7	−870.4	847.6	3
23	−41.0	−904.3	−868.3	847.8	3
24	−40.4	−900.4	−862.9	842.5	3
25	−40.9	−901.8	−866.5	845.4	3
26	−52.3	−661.8	−573.2	623.1	2
27	−56.4	−650.8	−559.7	619.0	2
28	−45.8	−950.8	−870.4	906.6	3
29	−47.7	−903.9	−851.7	867.8	3
30	−50.3	−656.0	−554.7	619.3	2
31	−52.6	−660.0	−556.5	629.5	2
32	−41.7	−628.0	−539.2	565.0	2
33	−49.7	−362.3	−269.2	357.1	1
34	−36.1	−649.8	−566.5	595.6	2
35	−37.3	−672.9	−566.3	615.7	2
36	−52.9	−111.5	−34.2	116.3	0
37	−59.6	−113.0	−35.4	121.3	0
38	−54.0	−70.9	−11.6	81.1	0
39	−58.0	−124.5	−44.2	136.4	0
40	−58.8	−121.4	−43.4	124.4	0
41	−60.1	−350.5	−260.1	347.3	1
42	−48.4	−654.1	−553.8	621.3	2
43	−47.1	−651.3	−552.4	616.2	2
44	−49.5	−653.5	−547.7	616.3	2

[a] All energy values are in kcal·mol^{-1}.

QMLIECE. The LIECE and QMLIECE models are tested on three enzyme/inhibitor systems: the West Nile virus NS3 protease (WNV PR, a serine protease), the HIV-1 protease (HIV-1 PR, an aspartic protease), and the human cyclin-dependent kinase 2 (CDK2). Because of the large variability of charge−charge interactions in the complexes of WNV PR with the 44 peptidic inhibitors (having between 0 and 3 positively charged side chains), the use of QM is necessary to capture polarization effects,[31,32] which are neglected in fixed-charge approximation of force field based methods. On the other hand, QM and force field methods show similar accuracy for the binding affinity evaluation of mainly neutral inhibitors of HIV-1 PR and CDK2.

2. Method

Preparation of WNV NS3-NS2B Protease. The coordinates of WNV PR in complex with the substrate-based inhibitor benzoyl-norleucine-lysine-arginine-arginine-aldehyde (Bz-Nle-Lys-Arg-Arg-H) were downloaded from the PDB database (PDB entry 2FP7).[33] All water molecules were removed. The spurious termini at the segments missing in the X-ray structure (residues 28−32 in chain B) were neutralized by a $-COCH_3$ and a $-NHCH_3$ group at the N-terminus and C-terminus, respectively. The 44 peptidic inhibitors of WNV PR used in this study include Bz-Nle-Lys-Arg-Arg-H (IC_{50} = 4.1 μM) and a series of 43 related inhibitors with an aldehyde warhead (IC_{50} values ranging from 0.4 to 463 μM) synthesized in the same laboratory and tested all with the same enzymatic assay (Table 1).[34–36] The initial binding conformations were modeled manually according to the binding mode of Bz-Nle-Lys-Arg-Arg-H because all inhibitors have similar backbone structure and are covalently bound to the Ser135 side chain by an ester linkage.

For interaction energy calculation, the ester bond between protein and inhibitor and the adjacent −OH group of Ser135 and −C(H)OH group of inhibitor were removed to avoid the artificial crash. The resulting empty valencies on both protein and inhibitors were filled with hydrogen atoms.

Preparation of HIV-1 PR and CDK2. The coordinates of the 24 complexes of HIV-1 PR (PDB code 1AAQ) with the inhibitors tested by Dreyer and co-workers[37] were available from a previous study.[24] The coordinates of the 73 complexes of

Table 3. Energy Components of HIV-1 PR Peptidic Inhibitors[37,a]

ID	ΔE_{vdW}	$\Delta E_{elec,coul}$	ΔE_{QM}	$\Delta G^{solvation}$	no. of formal charges[b]
1	−58.4	−22.3	−14.6	58.4	0
2	−61.2	−21.2	−14.9	61.3	0
3	−66.6	−18.1	−12.3	62.2	0
4	−64.9	−19.2	−8.7	62.4	0
5	−71.0	−21.2	−15.5	66.0	0
6	−64.6	−31.4	−21.3	74.5	0
7	−67.4	−29.4	−21.5	76.8	0
8	−72.4	−26.1	−20.3	79.2	0
9	−72.1	−24.6	−17.5	81.8	0
10	−77.5	−26.9	−23.4	86.1	0
11	−73.2	−27.5	−25.1	99.2	0
12	−76.2	−24.3	−25.8	102.7	0
13	−80.7	−22.5	−23.8	103.4	0
14	−80.4	−18.8	−16.4	105.6	0
15	−81.2	−22.9	−17.1	100.3	0
16	−75.4	−14.4	−29.3	129.2	0
17	−78.5	−12.2	−30.3	133.0	0
18	−82.6	−10.4	−27.5	133.7	0
19	−82.4	−7.5	−20.2	135.5	0
20	−83.2	−11.9	−21.0	132.5	0
21	−69.1	−56.2	−38.2	69.3	1
22	−71.7	−50.5	−34.3	71.8	1
23	−76.0	−46.2	−29.0	73.7	1
24	−76.2	−60.4	−32.1	64.2	1

[a] All energy values are in kcal·mol^{-1}. [b] Neutral blocking group or positive charge on unblocked N-terminal amino group of inhibitors 21−24. The C-terminal group is neutral; it is $-NH_2$ or −O−Me for inhibitors 1−10 or 11−24, respectively.

CDK2 (PDB code 1KE5) with the inhibitors published by Bramson[38] and Gibson[39] were also available from a previous study.[27]

Minimization. Standard protonation states at neutral pH were used for all ionizable side chains (i.e., neutral His, positively charged Arg and Lys, and negatively charged Asp and Glu) except for one of the two carboxy groups in the Asp catalytic dyad of HIV-1 PR.[24] The net charge of WNV PR, HIV-1 PR, and CDK2 is −10, +7, and +5, respectively. Hydrogen atoms were added to all structures and minimized with the program CHARMM[40] and the CHARMm22 force field[41] (Accelrys Inc.). Partial charges were assigned using the MPEOE method.[42,43] The WNV PR protein/inhibitor complexes were minimized with a two-step protocol. First, the inhibitors were minimized by 200 iterations of the steepest descent algorithm with rigid protein. The second step consisted of 10 000 iterations of the adopted basis Newton−Raphson algorithm to an rms of the gradient of 0.001 kcal mol^{-1} Å^{-1} with flexible protein but using harmonic restraints on all carbon atoms of protein and inhibitor. The value of the force constants was gradually decreased from 20 to 1 kcal mol^{-1} Å^2. In the first minimization the electrostatic energy term was screened by a distance-dependent dielectric function ($\varepsilon(r) = 4r$), and the default nonbonding cutoff of 14 Å was used. In the second minimization the Coulombic energy (constant dielectric of 1.0) was evaluated without truncation. The distance-dependent dielectric in the first minimization and the harmonic restraints on carbon atoms in the second minimization were applied to prevent artificial deviations due to vacuum effects. The second step of the minimization protocol with vacuum dielectric yields optimal QMLIECE model. Several optimization protocols were tested, and it was found that QMLIECE always outperforms LIECE model for WNV PR, irrespective of the protocol (see Supporting Information). Ideally, one should minimize the sum of van der Waals and QM energy contributions. However, optimization using QM, even with linear-scaling techniques, is still computationally too costly for enzyme/ligand complexes.[44] For HIV-1 PR and CDK2 complexes, the inhibitors were optimized by 200 iterations of the

Table 4. Energy Components of the CDK2 Inhibitors[a]

ID	ΔE_{vdW}	$\Delta E_{elec,coul}$	ΔE_{QM}	$\Delta G^{solvation}$
1	−24.5	−10.1	−8.3	36.0
2	−26.5	−10.2	−9.2	30.1
3	−28.2	−9.4	−9.2	32.1
4	−24.2	−11.0	−8.7	35.0
5	−26.7	−10.9	−10.0	34.7
6	−28.5	−8.7	−7.2	34.1
7	−28.9	−9.7	−4.5	28.6
8	−28.8	−8.1	−4.5	35.1
9	−22.3	−9.2	−7.9	29.5
10	−28.3	−10.4	−10.4	39.5
11	−23.2	−10.1	−8.5	35.7
12	−26.6	−11.1	−9.0	36.1
13	−28.5	−11.3	−6.2	33.6
14	−30.6	−8.8	−8.2	39.3
15	−31.2	−13.7	−15.9	46.2
16	−29.7	−10.5	−9.2	37.0
17	−29.3	−10.3	−12.0	40.4
18	−31.2	−9.9	−7.8	37.9
19	−29.9	−10.1	−10.4	39.0
20	−31.7	−1.5	2.1	35.3
21	−34.6	−11.3	−11.1	52.3
22	−31.3	−5.8	−4.2	31.2
23	−32.9	−14.0	−15.9	59.3
24	−39.3	−27.0	−21.8	54.3
25	−43.8	−26.6	−21.0	55.3
26	−43.6	−27.1	−22.9	56.2
27	−46.2	−26.4	−22.6	59.7
28	−47.3	−25.7	−21.0	63.1
29	−46.4	−27.3	−24.1	64.4
30	−44.3	−28.0	−24.4	64.7
31	−46.5	−29.1	−25.2	69.0
32	−48.9	−18.2	−19.3	66.3
33	−51.3	−32.7	−28.9	71.8
34	−51.8	−28.6	−27.4	69.7
35	−46.0	−27.2	−19.6	60.6
36	−42.7	−26.1	−21.9	53.8
37	−38.9	−26.0	−20.3	58.2
38	−40.0	−20.5	−19.2	59.1
39	−40.7	−20.4	−18.8	61.8
40	−42.2	−20.4	−17.8	59.3
41	−40.1	−21.8	−20.7	56.9
42	−39.4	−37.7	−27.4	63.1
43	−38.8	−17.4	−12.1	53.2
44	−39.9	−33.7	−25.8	61.5
45	−38.6	−30.8	−24.5	57.8
46	−41.4	−37.4	−26.7	79.1
47	−40.9	−41.9	−33.4	81.1
48	−41.9	−28.9	−28.7	67.5
49	−45.2	−33.3	−30.8	79.3
50	−50.9	−35.3	−22.4	80.8
51	−50.9	−40.9	−26.7	85.3
52	−44.3	−20.2	−23.3	66.6
53	−39.5	−25.8	−20.7	59.4
54	−42.9	−22.7	−16.1	52.8
55	−36.8	−17.6	−22.0	64.9
56	−42.0	−18.0	−19.3	58.0
57	−41.3	−15.5	−12.0	55.9
58	−41.5	−32.4	−25.6	61.1
59	−39.5	−27.1	−20.1	53.0
60	−39.6	−22.8	−15.5	50.6
61	−36.5	−23.3	−13.5	52.5
62	−40.3	−26.4	−20.1	53.2
63	−37.8	−23.3	−13.1	61.9
64	−41.6	−46.0	−31.8	71.0
65	−40.2	−24.7	−19.4	56.8
66	−40.5	−27.9	−22.2	64.2
67	−42.0	−30.2	−23.4	72.6
68	−42.4	−27.4	−26.2	65.9
69	−44.0	−29.4	−23.4	75.0
70	−47.1	−26.6	−24.8	71.7
71	−46.9	−32.9	−28.0	78.3
72	−47.8	−30.5	−28.2	83.4
73	−43.1	−31.1	−23.9	72.4

[a] All 73 inhibitors are nonpeptidic and devoid of formal charges.[38,39] All energy values are in kcal·mol^{-1}.

Table 5. QMLIECE and LIECE Models[a]

				leave-one-out cross-valid		
	α	β	$\Delta G_{tr,rot,bond}$ (kcal·mol^{-1})	rms error (kcal·mol^{-1})	q^2	rms error on test set[b] (kcal·mol^{-1})
			WNV PR (44 Peptidic Inhibitors with $0 \le Q \le 3$)			
$\beta\Delta G_{QM\_elesol} + \Delta G_{tr,rot,bond}$		0.022	−7.6	0.67	0.65	
standard deviation		±0.003	±0.2			
$\beta\Delta G_{MM\_elesol} + \Delta G_{tr,rot,bond}$		0.032	−5.7	0.91	0.35	
standard deviation		±0.006	±0.3			
			WNV PR (37 Peptidic Inhibitors with $2 \le Q \le 3$)			
$\beta\Delta G_{QM\_elesol} + \Delta G_{tr,rot,bond}$		0.024	−7.6	0.64	0.70	
standard deviation		±0.004	±0.2			
$\beta\Delta G_{MM\_elesol} + \Delta G_{tr,rot,bond}$		0.048	−5.0	0.84	0.49	
standard deviation		±0.009	±0.4			
			HIV-1 PR (24 Peptidic Inhibitors)			
$\alpha\Delta E_{vdW} + \beta\Delta G_{QM\_elesol} + \Delta G_{tr,rot}$	0.350	0.067	8.3	0.64	0.80	1.15
standard deviation	±0.063	±0.025	±2.8			
$\alpha\Delta E_{vdW} + \beta\Delta G_{MM\_elesol} + \Delta G_{tr,rot}$	0.299	0.032	7.9	0.67	0.78	1.30
standard deviation	±0.048	±0.013	±2.8			
			CDK2 (Nonpeptidic Inhibitors)			
$\alpha\Delta E_{vdW} + \beta\Delta G_{QM\_elesol}$	0.241	*0.002*[c]		0.99	0.79	
standard deviation	±0.022	*±0.022*[c]				
$\alpha\Delta E_{vdW} + \beta\Delta G_{MM\_elesol}$	0.265	0.029		0.98	0.79	
standard deviation	±0.018	±0.020				

[a] For each set of enzyme/inhibitor complexes the QMLIECE and LIECE models differ only in ΔG_{QM_elesol} and ΔG_{MM_elesol}, respectively (see Methods). [b] The test set was not used to derive the model. It contains five HIV-1 PR inhibitors with K_i values of 0.05, 0.38, 3.2, 437, and 1100 nM. [c] Parameters with leave-one-out standard deviation larger than the average value are statistically not significant and are given in italics.

steepest descent algorithm followed by 10 000 iterations of the adopted basis Newton−Raphson algorithm to an rms of the gradient of 0.001 kcal mol^{-1} Å^{-1} with rigid protein. Because of the predominance of the van der Waals term, which is identical in LIECE and QMLIECE, similar fitting results for these two complexes are obtained by minimizing with rigid protein or harmonically restrained protein (not shown).

Energy Calculation. All QM energy values were calculated on CHARMM-minimized structures. The vacuum interaction energies in QMLIECE were calculated with a divide and conquer approach[9,10] (Figure 1) using MOPAC[45] and the recently developed semiempirical Hamiltonian RM1.[16] The QM energy characterizes the nonclassical charge transfer effect, which is omitted in the fixed-charge model but can be strong if a cation group and an anion group are closed to each other, e.g., the positively charged side chains of Bz-Nle-Lys-Arg-Arg-H and negatively charged sub pockets of WNV PR.[34] van der Waals interactions are fundamentally charge−charge interactions consisting of attractive and repulsive interactions originating from dispersive forces and exchange forces, respectively. The interaction energies from semiempirical QM calculations include the repulsive part of van der Waals interaction energy but ignore the attractive part,[17] which needs highly correlated treatments and large basis sets.[46] Ideally, "pure" electrostatic part of QM interaction energy is needed for linear combination with van der Waals part from molecular mechanics (MM) calculation. However, the QM interaction energy cannot be decomposed as in classical force fields. A linear combination of the QM and MM energy contributions is used in QMLIECE to partially remove the double counting of the repulsive part of van der Waals interaction. In any case, minimized complexes have negligible repulsive interactions.

The van der Waals energy and the MM vacuum Coulombic energy ($\varepsilon(r) = 1$, infinite cutoff) were calculated using CHARMM[40] and the CHARMm[41] force field with the same protocol as in a previous publication.[24]

The electrostatic solvation energy was calculated by the finite-difference Poisson approach using the PBEQ[47] module in CHARMM[40] and a focusing procedure with a final grid spacing of 0.25 Å. The size of the initial grid was determined by considering a layer of at least 22.5 Å around the solute. The dielectric discontinuity interface was delimited by the molecular surface which is spanned by the surface of a rolling probe of 1.4 Å. The ionic strength was set to zero, and the temperature was set to 300 K. Two finite-difference Poisson calculations were performed for each of the three systems (inhibitor, protein, inhibitor/protein complex). The exterior dielectric constant was set to 78.5 and 1.0 for the first and second calculation, respectively, while the solute dielectric constant was set to 1.0, which is consistent with QM energy and parameters of the CHARMm22 force field.

Binding Free Energy. The equations used for the fitting are two-parameter models

$$\Delta G_{bind} = \beta\Delta G_{elesol} + \Delta G_{tr,rot,bond} \quad \text{for WNV PR} \tag{1}$$

$$\Delta G_{bind} = \alpha\Delta E_{vdW} + \beta\Delta G_{elesol} \quad \text{for CDK2}^{48} \tag{2}$$

and a three-parameter model

$$\Delta G_{bind} = \alpha\Delta E_{vdW} + \beta\Delta G_{elesol} + \Delta G_{tr,rot} \quad \text{for HIV-1 PR}^{24} \tag{3}$$

The electrostatic contribution to the binding energy ΔG_{elesol} is the sum of the ligand/protein electrostatic interaction energy in solvent ($\Delta G^{sol}_{pro/lig}$) and the change in solvation energy of ligand and protein upon binding:[49,50]

$$\begin{aligned}\Delta G_{elesol} &= \Delta G^{sol}_{prot/lig} - \Delta G^{solvation}_{prot} - \Delta G^{solvation}_{lig} \\ &= \Delta G^{vacuo}_{prot/lig} + \Delta G^{solvation}_{prot/lig} - \Delta G^{solvation}_{prot} - \Delta G^{solvation}_{lig} \\ &= \Delta G^{vacuo}_{prot/lig} + \Delta G^{solvation}\end{aligned} \tag{4}$$

For the vacuum electrostatic interaction energy $\Delta G^{vacuo}_{prot/lig}$, QM ($\Delta E_{QM}$ in Table 2–4) and MM ($\Delta E_{elec,coul}$) calculations were used in QMLIECE and LIECE, respectively. Note that ΔE_{QM} could be further decomposed into electrostatic and explicit polarization energy terms,[31,32] but such decomposition would require additional fitting parameters. The finite-difference Poisson approach was used to calculate the solvation energy changes upon binding ($\Delta G^{solvation}$). Details are given in the Supporting Information.

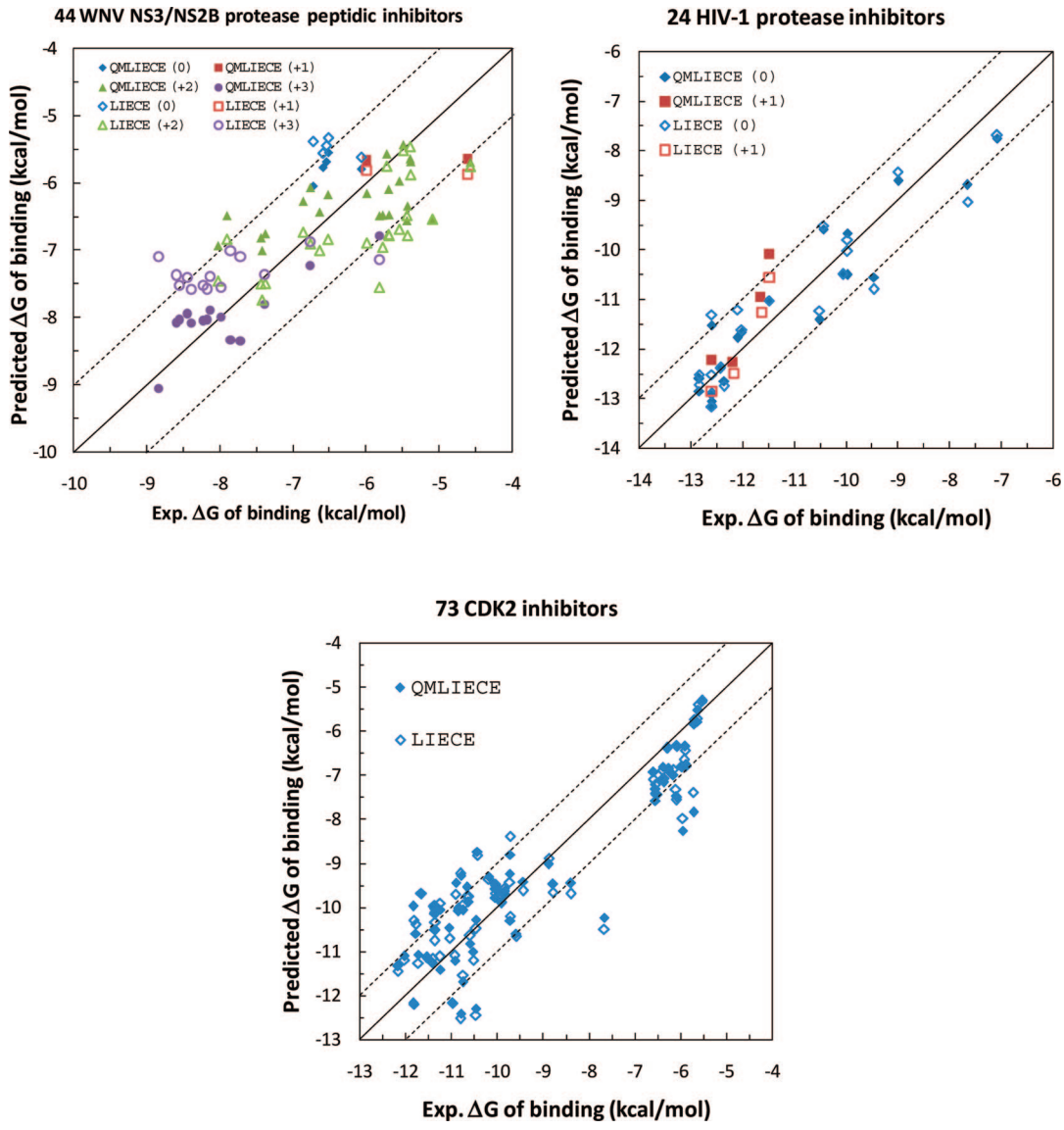

Figure 2. Comparison of the calculated (QMLIECE filled symbols, LIECE empty symbols) versus experimental binding free energies for 44 WNV PR[34–36] (top left), 24 HIV-1 PR[37] (top right), and 73 CDK2[38,39] (bottom) inhibitors. The experimental data are fitted with two-parameter models for WNV PR (eq 1), three-parameter models for HIV-1 PR (eq 3), and two-parameter models for CDK2 (eq 2). Digit in parentheses is the total charge of the inhibitor.

The term ΔE_{vdW} is the ligand/protein van der Waals interaction energy. Since the semiempirical QM calculation does not take into account the attractive part of the van der Waals energy, the van der Waals interaction energy of the force field is still used in QMLIECE.

The constant term $\Delta G_{tr,rot,bond}$ accounts for the loss of translational and rotational degrees of freedom upon binding and the energy of formation of the covalent bond for the 44 aldehyde inhibitors of WNV PR. The entropic penalty due to loss of translational and rotational degrees of freedom ($\Delta G_{tr,rot}$) is unfavorable and therefore positive, but its sum with the covalent bond energy can also be negative.

For WNV PR (eq 1), ΔE_{vdW} was neglected because the statistical significance of the fitting deteriorates (see Supporting Information). The same is observed upon addition of $\Delta G_{tr,rot}$ in the CDK2 models, which is consistent with the significantly smaller flexibility of the nonpeptidic inhibitors of CDK2 than the peptidic inhibitors of WNV PR and HIV-1 PR.

3. Results and Discussion

The energy values and the parameters obtained by least-squares fitting are given in Tables 2–4 and Table 5, respectively, while the correlation between LIECE/QMLIECE binding energies and experimental values is shown in Figure 2.

WNV PR. The two-parameter QMLIECE model yields a leave-one-out rms of the error of 0.67 kcal·mol^{-1} and cross-validated q^2 of 0.65. These results are significantly better than those obtained by LIECE (rms error of 0.91 kcal·mol^{-1} and cross-validated q^2 of 0.35). As an additional test, the LIECE and QMLIECE models were applied to a nonpeptidic inhibitor, discovered recently in our group (Ekonomiuk et al., unpublished results), which was not used to derive the model. The LIECE and QMLIECE binding affinity are −5.5 and −8.6 kcal·mol^{-1}, respectively, while the experimentally measured binding affinity is −7.2 kcal·mol^{-1}. Since this inhibitor does not bind covalently to the protein, the calculated binding free energy should be more favorable than the measured value because the LIECE and

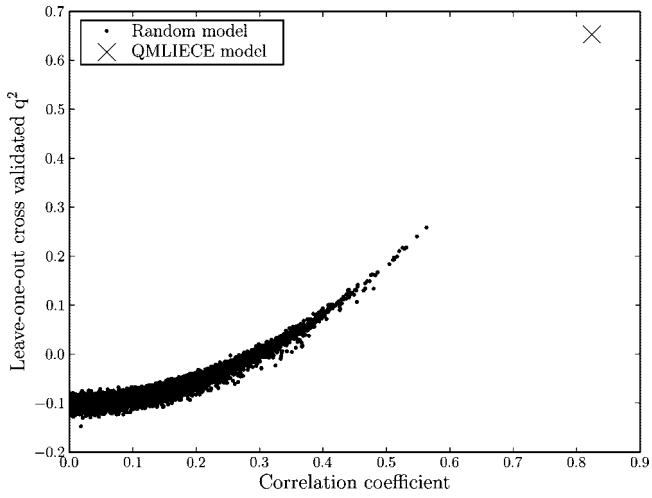

Figure 3. Statistical test to assess the predictive power of the QMLIECE two-parameter model for WNV PR (cross) by comparison with 10 000 models obtained by randomizing the activity values (dots). The QMLIECE model is clearly separated, which indicates that QMLIECE not only better fits the data than the random models but also has a better predictive ability. In other words, the plot shows that the QMLIECE model does not suffer from chance correlation.

QMLIECE models for WNV PR were derived from 44 peptidic inhibitors covalently bound to the protein. Therefore, the QMLIECE value is more reliable than the LIECE one.

A statistical test based on the randomization of the data points was used to analyze an eventual chance correlation between the QMLIECE model and the data points.[25,51] The binding free energies of the 44 peptidic inhibitors[34–36] were randomly chosen from uniformly distributed values in the same range as the experimental values (i.e., from −8.84 to −4.58 kcal·mol^{-1}), and the multiplicative parameters of ΔG_{elesol} and $\Delta G_{tr,rot,bond}$ constant term were determined by fitting to random "data points". The rationale behind this test is that the statistical significance of the real model is poor if there is a significant correlation between the descriptors and the randomized data points. The randomization and fitting were repeated 10 000 times, and Figure 3 shows the cross-validated q^2 (obtained by the leave-one-out procedure) plotted versus the correlation coefficient. The QMLIECE model with the two parameters fitted to the real data points is located at the top right corner and is significantly separated from the models generated by the randomization of the binding free energies. This separation provides further evidence that the QMLIECE two-parameter model not only fits the experimental data but also has very good predictive ability, i.e., chance correlation is not present. To further assess the significance of the QMLIECE model, the same statistical test was performed on two naive models suggested by an anonymous reviewer: a combination of LIECE and a binary descriptor for distinguishing inhibitors with charge +2 from all others and a simple five-parameter model based only on binary descriptors for the number of positive charges in the inhibitors (Supporting Information). Interestingly, the risk of chance correlation increases with increasing number of fitting parameters and decreasing physical soundness. In other words, for the QMLICE model only, there is a genuine correlation between descriptors and data points.

HIV-1 PR. The three-parameter QMLIECE model yields an rms of the error of 0.64 kcal·mol^{-1} and a cross-validated q^2 of 0.80. These results are similar to those obtained by LIECE (rms error of 0.67 kcal·mol^{-1} and a cross-validated q^2 of 0.78). (Note that the rms error of 0.67 kcal·mol^{-1} is slightly smaller than in ref 24, where it was 0.77 kcal·mol^{-1}, because of the different minimization protocol.) The predictive ability of the LIECE and QMLIECE approach was further tested on a set of five inhibitors available from a previous work[52] and not used to derive the models. Their PDB identifiers and K_i values are the following: 1HVR, K_i = 0.05 nM, K_i = 0.38 nM; 1HTG, K_i = 3.2 nM;

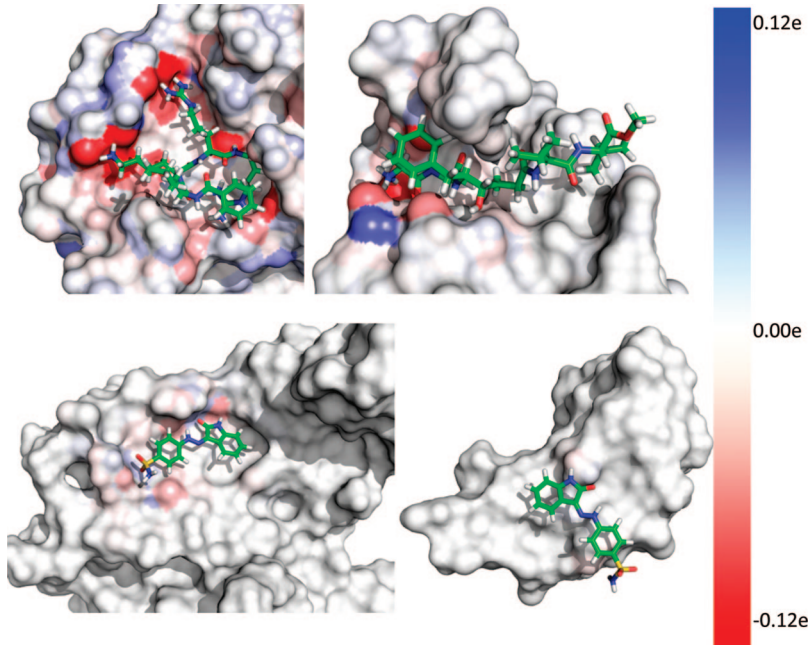

Figure 4. Polarization of protein atoms due to inhibitor binding: WNV PR in complex with its inhibitor 1 (top left), HIV-1 PR in complex with its inhibitor 21 (top right, only one monomer of the C_2-symmetric structure of the HIV-1 PR homodimer is shown), and CDK2 in complex with its inhibitor 24 (α-helical domain bottom left and β-sheet domain bottom right). The polarized charges were calculated by subtracting SCF atomic charges before binding from that after binding, using the divide and conquer protocol.[9] The protein surfaces were rendered with the blue-white-red spectrum according to polarized charges of atoms. The blue on the surface denotes atomic partial charges that become more positive upon binding, while red means more negative atomic charges upon binding, and white indicates atomic charges that do not change upon binding.

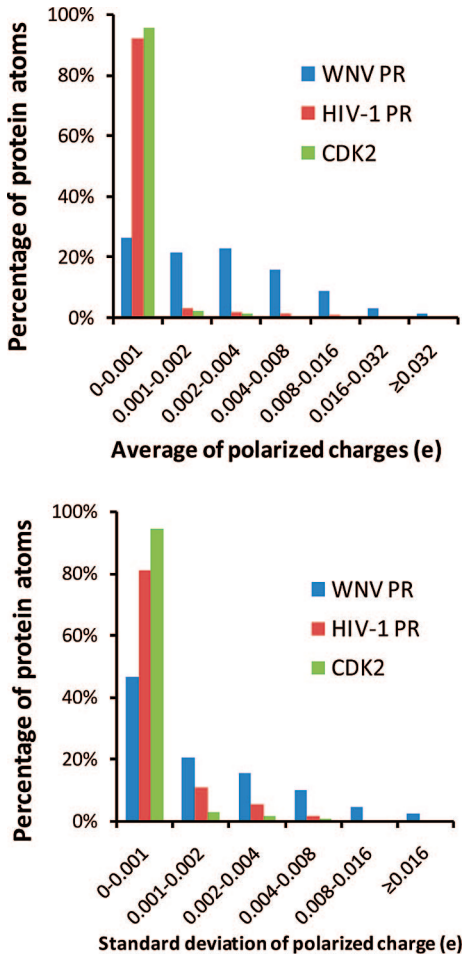

Figure 5. Distribution of average (top) and standard deviation (bottom) of individual polarized charges of proteins upon binding, calculated over all inhibitors (44, 24, and 73 inhibitors for WNV PR, HIV-1 PR, and CDK2, respectively). The polarized charges were calculated by subtracting SCF atomic charges before binding from that after binding, using the divide and conquer protocol.[9]

1HBV, K_i = 437 nM; 1HVS, 5HVP, K_i = 1100 nM.[53–57] The five inhibitors were minimized in the HIV-1 PR conformation from the 1HVR complex because of its highest resolution (1.8 Å). The error rms for the five inhibitors of the test set is 1.30 and 1.15 kcal·mol^{-1} for the LIECE and QMLIECE models, respectively. This comparison indicates that for the 24 mainly neutral inhibitors of HIV-1 PR the QMLIECE model is only slightly more predictive than the LIECE model.

CDK2. The two-parameter model of QMLIECE yields an rms of the error of 0.99 kcal·mol^{-1} and a cross-validated q^2 of 0.79. This accuracy is essentially identical to the one of the LIECE model. Moreover, the electrostatic parameter of the QMLIECE model is smaller than the standard deviation obtained by the leave-one-out procedure, which indicates that the QMLIECE model of CDK2 is not robust.

Applicability of QM. It is important to clarify under which circumstance it is necessary to use QM for calculating electrostatic energy contribution in linear interaction energy models. The advantage of QM compared with MM is that QM allows the evaluation of charge-transfer effects by self-consistent field (SCF) calculation. Upon inhibitor binding, the amount of polarization of WNV PR is larger than HIV-1 PR and much larger than CDK2 (Figure 4). As a matter of fact, for the complexes of HIV-1 PR and CDK2 the charge−charge interactions are relatively similar and small (Figure 4). Furthermore, more than 90% of their atoms are not significantly polarized (less than 0.001e) upon inhibitor binding (Figure 5). Therefore, the absolute errors originating from polarization are small for these two enzyme/inhibitor complexes, and can be rectified by the regression parameters without leading to poor fitting. On the other hand, the 44 peptidic inhibitors of WNV have between zero and three positively charged side chains resulting in a large variability of polarized charges; as a consequence, the energies calculated by MM are significantly different from QM energies because only the latter takes charge polarization effects into account. This explains the better predictive ability of QMLIECE than LIECE for the 44 inhibitors of WNV PR.

An additional test was performed to separate the effect of total charge from charge variability. The inhibitors of WNV PR with zero or one formal charge (7 of the 44 inhibitors) were removed from the fitting data of two-parameter model of LIECE. The variability of polarized charges, therefore, becomes smaller for this subset, while the average value of polarized charges becomes even larger. By application of leave-one-out cross-validation to the reduced set of data (37 inhibitors), it is found that QMLIECE does not change significantly whereas q^2 of LIECE improves from 0.35 to 0.49, Moreover, in the LIECE model generated using only 37 inhibitors with $2 \le Q \le 3$ the parameter of ΔG_{MM_elesol} increases from 0.032 to 0.048, and the constant term $\Delta G_{tr,rot,bond}$ changes from −5.7 to −5.0 kcal·mol^{-1} (Table 5). These results indicate that the weight of the electrostatic contribution in the LIECE regression model increases by reducing the formal charge variability of inhibitors despite the larger average total charge. In other words, the neglect of polarization in LIECE results in acceptable predictive ability for binding affinities of inhibitors with two or three formal charges.

Therefore, if the charge−charge interactions between inhibitors and protein are similar, even though the absolute values of them are large, the fixed charge model in the force field method can attain reasonable results for the evaluation of electrostatic energies. Otherwise, QM is needed to more accurately evaluate the variable influence of polarization effects on electrostatic interactions.

Computational Requirements. The time for the QM calculation is linearly related to the number of residues. For WNV PR (187 residues), the QM calculation needs about 40 min on an Opteron 252 (2.6 GHz). The total time required by QMLIECE is about 1 h for each inhibitor. The finite-difference Poisson and QM calculations require about 900 and 50 MB memory, respectively.

4. Conclusions

Previously the computationally expensive sampling by MD in the linear interaction energy model had been substituted with a simple energy minimization and continuum electrostatics calculation (LIECE).[24] QMLIECE is a further development of LIECE, in which the force field based electrostatic part of the inhibitor/protein interaction energy is replaced by the corresponding contribution evaluated by a QM calculation at the semiempirical RM1[16] level with a linear-scaling method. LIECE and QMLIECE models are assessed on three classes of inhibitor/enzyme complexes: 44 inhibitors of a flaviviral serine protease, 24 inhibitors of a retroviral aspartic protease, and 73 inhibitors of the CDK2 serine/threonine protein kinase. Only for the 44 inhibitors of the serine protease, which have between zero and three positive charges, did QMLIECE show a significant improvement compared to LIECE. However, for the subset of 37 of the 44 inhibitors with two or three positive charges LIECE

was more predictive than for the full set of 44 inhibitors but still not as robust as QMLIECE. Therefore, the comparison of LIECE and QMLIECE indicates that the use of QM is necessary when complexes with different inhibitors have significantly diverse charge−charge interactions, i.e., a large variability of polarized charges of protein atoms upon binding different inhibitors.

Acknowledgment. This work was supported in part by the Hartmann-Müller Foundation. Calculations were performed on the Matterhorn Beowulf cluster.

Supporting Information Available: Tables of different optimization protocols for QMLIECE and LIECE, including CHARMM commands, their cross-validated q^2, three-parameter LIECE and QMLIECE models, and statistical tests on simple models. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.
(2) Warshel, A.; Karplus, M. Calculation of ground and excited-state potential surfaces of conjugated molecules. 1. Formulation and parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.
(3) Field, M. J.; Bash, P. A.; Karplus, M. A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.
(4) Masgrau, L.; Roujeinikova, A.; Johannissen, L. O.; Hothi, P.; Basran, J.; Ranaghan, K. E.; Mulholland, A. J.; Sutcliffe, M. J.; Scrutton, N. S.; Leys, D. Atomic description of an enzyme reaction dominated by proton tunneling. *Science* **2006**, *312*, 237–241.
(5) Pu, J. Z.; Gao, J. L.; Truhlar, D. G. Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3140–3169.
(6) Gao, J. L.; Ma, S. H.; Major, D. T.; Nam, K.; Pu, J. Z.; Truhlar, D. G. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.* **2006**, *106*, 3188–3209.
(7) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11*, 182–187.
(8) Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular tailoring approach for simulation of electrostatic properties. *J. Phys. Chem.* **1994**, *98*, 9165–9169.
(9) Dixon, S. L.; Merz, K. M. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
(10) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein−molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
(11) Raha, K.; Merz, K. M. Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein−ligand complexes. *J. Med. Chem.* **2005**, *48*, 4558–4575.
(12) Cavalli, A.; Carloni, P.; Recanatini, M. Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* **2006**, *106*, 3497–3519.
(13) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; WollaCott, A. M.; Westerhoff, L. M.; Merz, K. M. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today* **2007**, *12*, 725–731.
(14) Vondrasek, J.; Bendova, L.; Klusak, V.; Hobza, P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.* **2005**, *127*, 2615–2619.
(15) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
(16) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RMI: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.
(17) Raha, K.; Merz, K. M. A quantum mechanics-based scoring function: study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020–1021.
(18) Alvarez, J. C. High-throughput docking as a source of novel drug leads. *Curr. Opin. Chem. Biol.* **2004**, *8*, 365–370.
(19) Joseph-McCarthy, D.; Baber, J. C.; Feyfant, E.; Thompson, D. C.; Humblet, C. Lead optimization via high-throughput molecular docking. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 264–274.
(20) Giese, T. J.; York, D. M. High-level ab initio methods for calculation of potential energy surfaces of van der Waals complexes. *Int. J. Quantum Chem.* **2004**, *98*, 388–408.
(21) Barratt, E.; Bingham, R. J.; Warner, D. J.; Laughton, C. A.; Phillips, S. E. V.; Homans, S. W. van der Waals interactions dominate ligand−protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* **2005**, *127*, 11827–11834.
(22) Johnson, E. R.; Becke, A. D. van der Waals interactions from the exchange hole dipole moment: application to bio-organic benchmark systems. *Chem. Phys. Lett.* **2006**, *432*, 600–603.
(23) Gonzalez-Diaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. *J. Proteome Res.* **2007**, *6*, 904–908.
(24) Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
(25) Huang, D. Z.; Luthi, U.; Kolb, P.; Cecchini, M.; Barberis, A.; Caflisch, A. In silico discovery of beta-secretase inhibitors. *J. Am. Chem. Soc.* **2006**, *128*, 5436–5443.
(26) Huang, D. Z.; Luthi, U.; Kolb, P.; Edler, K.; Cecchini, M.; Audetat, S.; Barberis, A.; Caflisch, A. Discovery of cell-permeable non-peptide inhibitors of beta-secretase by high-throughput docking and continuum electrostatics calculations. *J. Med. Chem.* **2005**, *48*, 5108–5111.
(27) Kolb, P.; Huang, D.; Dey, F.; Caflisch, A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* **2008**, *51*, 1179–1188.
(28) Hansson, T.; Aqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, *8*, 1137–1144.
(29) Aqvist, J.; Medina, C.; Samuelsson, J. E. New method for predicting binding-affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
(30) Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.
(31) Garcia-Viloca, M.; Truhlar, D. G.; Gao, J. L. Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J. Mol. Biol.* **2003**, *327*, 549–560.
(32) Hensen, C.; Hermann, J. C.; Nam, K. H.; Ma, S. H.; Gao, J. L.; Holtje, H. D. A combined QM/MM approach to protein−ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* **2004**, *47*, 6673–6680.
(33) Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat. Struct. Mol. Biol.* **2006**, *13*, 372–373.
(34) Knox, J. E.; Ma, N. L.; Yin, Z.; Patel, S. J.; Wang, W. L.; Chan, W. L.; Rao, K. R. R.; Wang, G.; Ngew, X.; Patel, V.; Beer, D.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors. *J. Med. Chem.* **2006**, *49*, 6585–6590.
(35) Yin, Z.; Patel, S. J.; Wang, W. L.; Chan, W. L.; Rao, K. R. R.; Wang, G.; Ngew, X.; Patel, V.; Beer, D.; Knox, J. E.; Ma, N. L.; Ehrhardt, C.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 40–43.
(36) Yin, Z.; Patel, S. J.; Wang, W. L.; Wang, G.; Chan, W. L.; Rao, K. R. R.; Alam, J.; Jeyaraj, D. A.; Ngew, X.; Patel, V.; Beer, D.; Lim, S. P.; Vasudevan, S. G.; Keller, T. H. Peptide inhibitors of dengue virus NS3 protease. Part 1: Warhead. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 36–39.
(37) Dreyer, G. B.; Lambert, D. M.; Meek, T. D.; Carr, T. J.; Tomaszek, T. A.; Fernandez, A. V.; Bartus, H.; Cacciavillani, E.; Hassell, A. M.; Minnich, M.; Petteway, S. R.; Metcalf, B. W.; Lewis, M. Hydroxyethylene isostere inhibitors of human immunodeficiency virus-1 protease - structure activity analysis using enzyme-kinetics, X-ray crystallography, and infected T-cell assays. *Biochemistry* **1992**, *31*, 6646–6659.
(38) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.
(39) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O6 substituted guanine derivatives. *J. Med. Chem.* **2002**, *45*, 3381–3393.
(40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm, a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*

1983, *4*, 187–217.
(41) Momany, F. A.; Rone, R. Validation of the general-purpose Quanta 3.2/CHARMm force-field. *J. Comput. Chem.* **1992**, *13*, 888–900.
(42) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic-molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4740–4746.
(43) No, K. T.; Grant, J. A.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4732–4739.
(44) Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Fully quantum mechanical energy optimization for protein−ligand structure. *J. Comput. Chem.* **2004**, *25*, 1431–1437.
(45) Stewart, J. J. P. Optimization of parameters for semiempirical methods. 1. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.
(46) Chalasinski, G.; Szczesniak, M. M. State of the art and challenges of the ab initio theory of intermolecular interactions. *Chem. Rev.* **2000**, *100*, 4227–4252.
(47) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson−Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59–75.
(48) Kolb, P.; Huang, D.; Dey, F.; Caflisch, A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J. Med. Chem.* **2008**, *51*, 1179–1188.
(49) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 88–105.
(50) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins: Struct., Funct., Genet.* **2001**, *42*, 256–268.
(51) So, S. S.; Karplus, M. Genetic neural networks for quantitative structure−activity relationships: improvements and application of benzodiazepine affinity for benzodiazepine/GABA(A) receptors. *J. Med. Chem.* **1996**, *39*, 5246–5256.
(52) Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J. Comput. Chem.* **2004**, *25*, 412–422.
(53) Fitzgerald, P. M. D.; Mckeever, B. M.; Vanmiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. Crystallographic analysis of a complex between human-immunodeficiency-virus type-1 protease and acetyl-pepstatin at 2.0-Å resolution. *J. Biol. Chem.* **1990**, *265*, 14209–14219.
(54) Jhoti, H.; Singh, O. M. P.; Weir, M. P.; Cooke, R.; Murrayrust, P.; Wonacott, A. X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease. *Biochemistry* **1994**, *33*, 8417–8427.
(55) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Ericksonviitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263*, 380–384.
(56) Baldwin, E. T.; Bhat, T. N.; Liu, B. S.; Pattabiraman, N.; Erickson, J. W. Structural basis of drug-resistance for the V82a mutant of HIV-1 proteinase. *Nat. Struct. Biol.* **1995**, *2*, 244–249.
(57) Hoog, S. S.; Zhao, B. G.; Winborne, E.; Fisher, S.; Green, D. W.; Desjarlais, R. L.; Newlander, K. A.; Callahan, J. F.; Moore, M. L.; Huffman, W. F.; Abdelmeguid, S. S. A check on rational drug design. Crystal-structure of a complex of human-immunodeficiency-virus type-1 protease with a novel gamma-turn mimetic inhibitor. *J. Med. Chem.* **1995**, *38*, 3246–3252.

JM800242Q