Hydrogen Bonding and π-Stacking: How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions

Robert S. Paton and Jonathan M. Goodman*

Department of Chemistry, Unilever Centre for Molecular Science Informatics, Lensfield Road, Cambridge CB2 1EW, U.K.

Received January 12, 2009

We have evaluated the performance of a set of widely used force fields by calculating the geometries and stabilization energies for a large collection of intermolecular complexes. These complexes are representative of a range of chemical and biological systems for which hydrogen bonding, electrostatic, and van der Waals interactions play important roles. Benchmark energies are taken from the high-level ab initio values in the JSCH-2005 and S22 data sets. All of the force fields underestimate stabilization resulting from hydrogen bonding, but the energetics of electrostatic and van der Waals interactions are described more accurately. OPLSAA gave a mean unsigned error of 2 kcal mol^{-1} for all 165 complexes studied, and outperforms DFT calculations employing very large basis sets for the S22 complexes. The magnitude of hydrogen bonding interactions are severely underestimated by all of the force fields tested, which contributes significantly to the overall mean error; if complexes which are predominantly bound by hydrogen bonding interactions are discounted, the mean unsigned error of OPLSAA is reduced to 1 kcal mol^{-1}. For added clarity, web-based interactive displays of the results have been developed which allow comparisons of force field and ab initio geometries to be performed and the structures viewed and rotated in three dimensions.

INTRODUCTION

The structure of biologically important macromolecules such as DNA, RNA, and proteins is largely determined by the noncovalent interactions between constituent base pairs and amino acids.[1] Molecular self-assembly, folding, molecular recognition, host−guest chemistry, mechanically interlocked molecular architectures, and dynamic covalent chemistry all rely on noncovalent interactions. Noncovalent intermolecular forces also govern the reversible binding of drug molecules to biological targets. The term *noncovalent interactions* is used here to encompass contributions from electrostatic interactions between charged species, hydrogen bonds between electronegative atoms, and hydrogen atoms bound to another electronegative atom and van der Waals forces.

The van der Waals term includes dipole−dipole interactions, dipole−induced dipole interactions and London or dispersion interactions (interactions between instantaneously induced multipoles). Electrostatic interactions are especially important for proteins, and ion pairs or salt bridges play an important role in stabilizing their structures. Hydrogen bonding is highly specific and directional. It is responsible for much of the stabilization of biomacromolecules, such as DNA, via the formation of complementary nucleic acid base pairs. Dispersion forces are the major attractive interactions between nonpolar molecules. Indeed, van der Waals complexes are formed by all neutral, closed-shell molecules.[2] Dispersion forces are also important between unsaturated organic groups and may be termed $\pi-\pi$ stacking interactions. Recent high-level ab initio calculations have demonstrated that dispersion forces between of stacked DNA base pairs and stacked amino acids can be large, contributing almost as much stabilization energy as hydrogen bonding interactions.[3,4] Recent DFT studies have suggested that, for small molecules, stacking interactions arise due to stabilizing van der Waals interactions rather than p-orbital overlap[5] and van der Waals interactions make the major contribution in the Hunter-Sanders model for $\pi-\pi$ interactions.[6] DFT calculations have also recently suggested that the interaction between polar substituents and apolar benzene rings significantly contribute to the stability of substituted benzene dimers.[7]

A computational description of extended macromolecules for which intermolecular forces are important is currently impossible using ab initio calculations at the post Hartree−Fock or DFT level because of the computational expense involved. Molecular mechanics methods, which use empirically parametrized force fields, provide an invaluable means to study the conformations, dynamics and energetics of these important systems. To obtain meaningful results, the accurate parametrization of the force fields to reproduce experimental or ab initio quantities is essential. As progress is made in high-level quantum chemical studies of nucleic acid bases and amino acid building blocks of large biomacromolecules, it becomes increasingly important that force field performance is evaluated against these benchmarks. Here we present a systematic comparison of the ability of widely used force fields to accurately reproduce high-level interaction energies and geometries, using a series of complexes for which noncovalent interactions play an important role. Calculations need to provide the correct balance between hydrogen bonding and stacking interactions to model accurately the structures of oligomers and biopolymers. Non-

* To whom correspondence should be addressed. E-mail: jmg11@cam.ac.uk.

10.1021/ci900009f CCC: $40.75
Published on Web 03/23/2009

Table 1. Nonbonded Interactions of the Force Fields[b]

Force field	van der Waals potential	H-bonding potential	Treatment of 1,4-interactions
MM2*	MM2 Hill equation $\varepsilon\left(2.9\times10^5 e^{-12.5\left(\frac{R_{AB}}{R^*_{AB}}\right)} - 2.25\left(\frac{R^*_{AB}}{R_{AB}}\right)^6\right)$	Angle-independent 10,12 Lennard-Jones potential	100%
MM3*	MM3 Hill equation $\varepsilon\left(1.84\times10^5 e^{-12.0\left(\frac{R_{AB}}{R^*_{AB}}\right)} - 2.25\left(\frac{R^*_{AB}}{R_{AB}}\right)^6\right)$	Angle-independent 10,12 Lennard-Jones potential	100%
AMBER*	6,12 Lennard-Jones potential $\varepsilon\left(\left(\frac{R^*_{AB}}{R_{AB}}\right)^{12} - 2\left(\frac{R^*_{AB}}{R_{AB}}\right)^6\right)$	No explicit hydrogen bonding[a]	van der Waals and electrostatics scaled by 50%
OPLSA*		No explicit hydrogen bonding	van der Waals scaled by 12.5% electrostatics scaled by 50%
OPLSAA		No explicit hydrogen bonding	van der Waals and electrostatics scaled by 50%
MMFF	Buffered 7,14 potential $\varepsilon\left(\frac{1.07R^*_{AB}}{R_{AB}+0.07R^*_{AB}}\right)^7\left(\frac{1.12R^{*\,7}_{AB}}{R_{AB}^{\,7}+0.12R^{*\,7}_{AB}}-2\right)$	No explicit hydrogen bonding	electrostatics scaled by 75%

[a] AMBER* has the nondefault option of explicit 10,12-Lennard-Jones hydrogen bonding. [b] MMFF and MMFFs have the same functional form. R_{AB} is the non-bonded interatomic distance, and R^*_{AB} is the sum of the predefined van der Waals radii. All force fields use a Coulombic potential between point charges with a constant dielectric for electrostatic interactions with the exception of MMFF which uses a buffered Coulombic potential: $E_{el} \propto (q_A q_B)/[\rho(R_{AB} + \delta)]$.

bonded interactions are also important for studies of reaction mechanisms and stereoselectivity. The controlling interactions in synthetic processes are usually nonbonded interactions between chiral auxiliaries or catalysts and the reacting center.[8] Simulations in solution are dominated by the nonbonded interactions between substrates, solvents, and ions.[9] As interest continues to grow into the use of hybrid quantum mechanics/molecular mechanics (QM/MM) methods to describe large molecules,[10] it is necessary to know the benefits and pitfalls of particular force fields. For this study ab initio data have been taken from the JSCH-2005 and S22 data sets from the Prague laboratory of Hobza and co-workers, which have been released to the community for the purpose of testing the performance of computational descriptions of intermolecular interactions.[11]

FORCE FIELDS STUDIED

The performance of a range of widely used all-atom force fields (those for which hydrogen atoms are explicitly included) has been evaluated by calculating the geometries and energies of noncovalent complexes in the gas-phase. The force fields evaluated are the *MacroModel* implementations of MM2*,[12] MM3*,[13] AMBER*,[14,15] and OPLS*,[16] and native versions of OPLSAA[17] and MMFF94/MMFF94s.[18,19] *MacroModel* atom types were assigned to each complex and remained the same for each of the force fields tested. All force fields were accessed through version 7.5 of *BatchMin*.[20] Where applicable, the effects of including explicit lone-pairs have been examined and different nonbonded cutoff values have been investigated. The functional form of nonbonded potentials and any special treatments of 1,4-interactions (atoms separated by three chemical bonds) are tabulated for each of the force fields tested (Table 1). Nonbonded interactions are only counted for atoms three or more bonds apart. The force fields are described in more detail in the Appendix.

BENCHMARK DATA SETS

The benchmark geometries and interaction energies of the complexes have been taken from the S22 and the JSCH-2005 data sets.[11] The S22 data set consists of small to medium-sized (30 atoms) complexes of common molecules containing only C, N, O, and H, with single, double, and triple bonds (Figure 1). Most typical noncovalent interactions, such as hydrogen bonds, dispersion interactions (stacked parallel, T-shaped), and mixed electrostatic-dispersion interactions are represented in this data set. A total of 22 complexes are divided into three subgroups: seven predominantly hydrogen bonded complexes, eight complexes bound predominantly by dispersion interactions, and seven complexes for which the contribution of electrostatic and dispersion interactions are similar in magnitude. The complexes were originally chosen to represent all the characteristic types of noncovalent interactions as well as possible. The set also spans a wide range of interaction strengths in order to represent the diversity of interactions in macromolecules. In each of the above-mentioned subgroups, the stability of the complexes ranges between 3 and 20, 0.5 and 15, and 1.2 and 8 kcal mol^{-1}, respectively. Some of the complexes appear twice as they are present in different geometries (e.g., the uracil dimer is present with a planar and a stacked geometry, and the benzene dimer is present with a parallel stacked and a T-shape stacked geometry).

While the S22 data set includes a wide-range of molecules, the JSCH-2005 data set is more focused on biological systems and contains almost exclusively nucleic acid bases

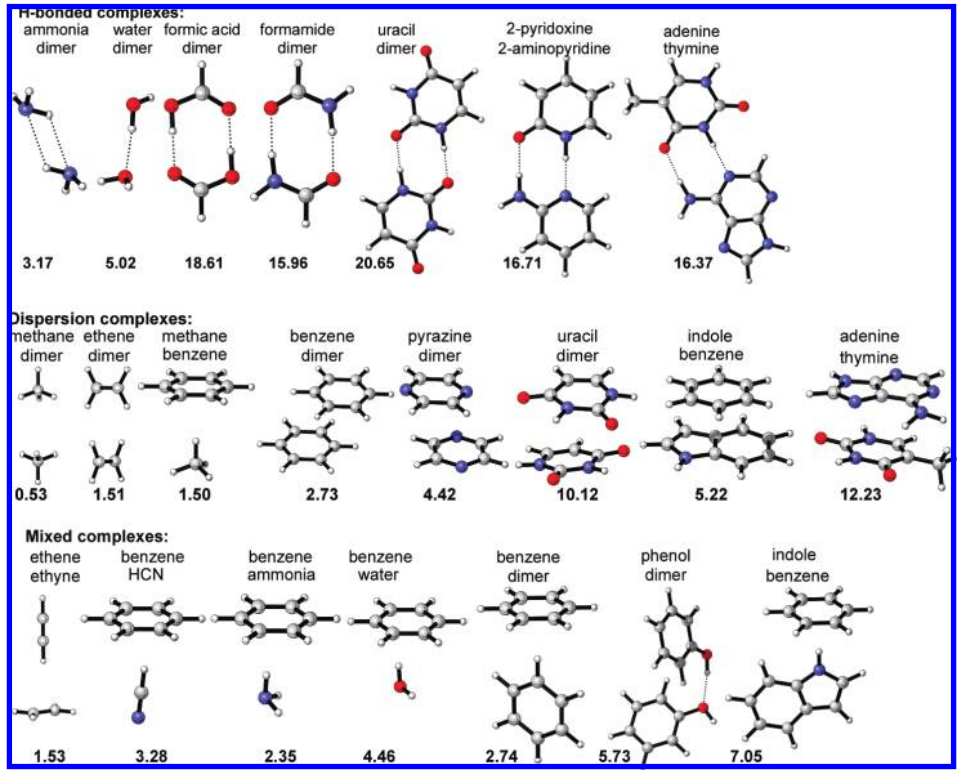


Figure 1. Molecular complexes contained in the S22 data set with benchmark ab initio interaction energies in kcal mol^{-1}.[21]

and amino acids (Figure 2). The set is divided into four subgroups. Three subgroups contain nucleic acid bases, divided according to the type of predominant noncovalent interaction: hydrogen-bonded DNA base pairs (38 structures),[22−24] interstrand base pairs that are characterized by smaller interaction energies (32 structures),[24,25] and stacked base pairs (54 structures).[22,24−27] The other subgroup consists of pairs of amino acid pairs (19 structures). Geometries are taken from a mixture of ab initio optimized complexes and experimental (X-ray and NMR) structures. Eleven neutral amino acid pairs (of which one component to each of the pairs is phenylalanine) were taken from the hydrophobic core of the protein rubredoxin, along with two amino acid/peptide bond pairs. Six charged amino acid pairs were taken from PDB crystal structures of rubredoxin (1IU5, 1BQ9, 1SMM, 1BRF).[28]

COMPUTATIONAL METHODS

The S22 and JSCH-2005 data sets contain only C, H, N, O, F and S atoms. *MacroModel* C2, N2 and O2 atom types were assigned to atoms formally sp^2 hybridized and C3, N3 and O3 atom types were assigned to sp^3 atoms. Fluorine atoms were assigned with the neutral F0 type, and thiocarbonyl sulfur atoms were assigned with the neutral S0 atom type, rather than the S2 atom type for sp^2 sulfurs that lacks nonbonded parameters for some of the force fields tested. For the MM3* force field only, there are insufficient parameters for sulfur atom types, so complexes containing sulfur were not tested with this force field. Hydrogen atoms were assigned automatically according to the bonded heavy atom: H1 if bonded to carbon, H2 if oxygen, and H3 if nitrogen. If the bonded atom was formally positively charged, the H4 atom type was used. For amino acids, the protonation states of carboxylate and amino groups were taken directly from that used in the ab initio investigation to allow for a direct comparison. Primary amino groups were protonated (the positively charged N5 atom type was used for nitrogen and the H4 atom type for the adjoining hydrogens) and carboxylate groups deprotonated, and so the negatively charged OM atom type was used for one of the oxygen atoms. *MacroModel* formatted input files containing Cartesian coordinates, atom types, and bond orders are available for download from the Supporting Information. All tautomeric forms used are identical to those contained in the ab initio data sets to allow for direct comparison with the force field results.

By default, *MacroModel* energy calculations are performed with cutoff distances in place for nonbonded interactions. If the distance between any nonbonded pair of atoms is greater than the cutoff distance, then the nonbonded term (electrostatic or van der Waals energies) for that pair will be ignored. Since these terms decrease in magnitude with increasing interatomic distance, nonbonded atom pairs separated by a large distance do not contribute greatly to the overall energy. The introduction of nonbonded cut-offs greatly reduces the time it takes to compute the energy for a large molecular system. We have used the default cutoff distance of 7 Å for van der Waals interactions and 12 Å for electrostatics and also tested nonbonded cut-offs that are effectively infinite (i.e., larger than the largest interatomic distance in the complexes taken from the data set, which is 15.9 Å). The default implementation of the "bond-dipole cutoff" (BDCO) method of truncating nonbonded interactions has also been tested, which uses a 12 Å cutoff for charge-dipole interactions and no cutoff for charge−charge interactions. A dielectric constant of 1.0 has been used during the evaluation of electrostatic interactions between atoms.

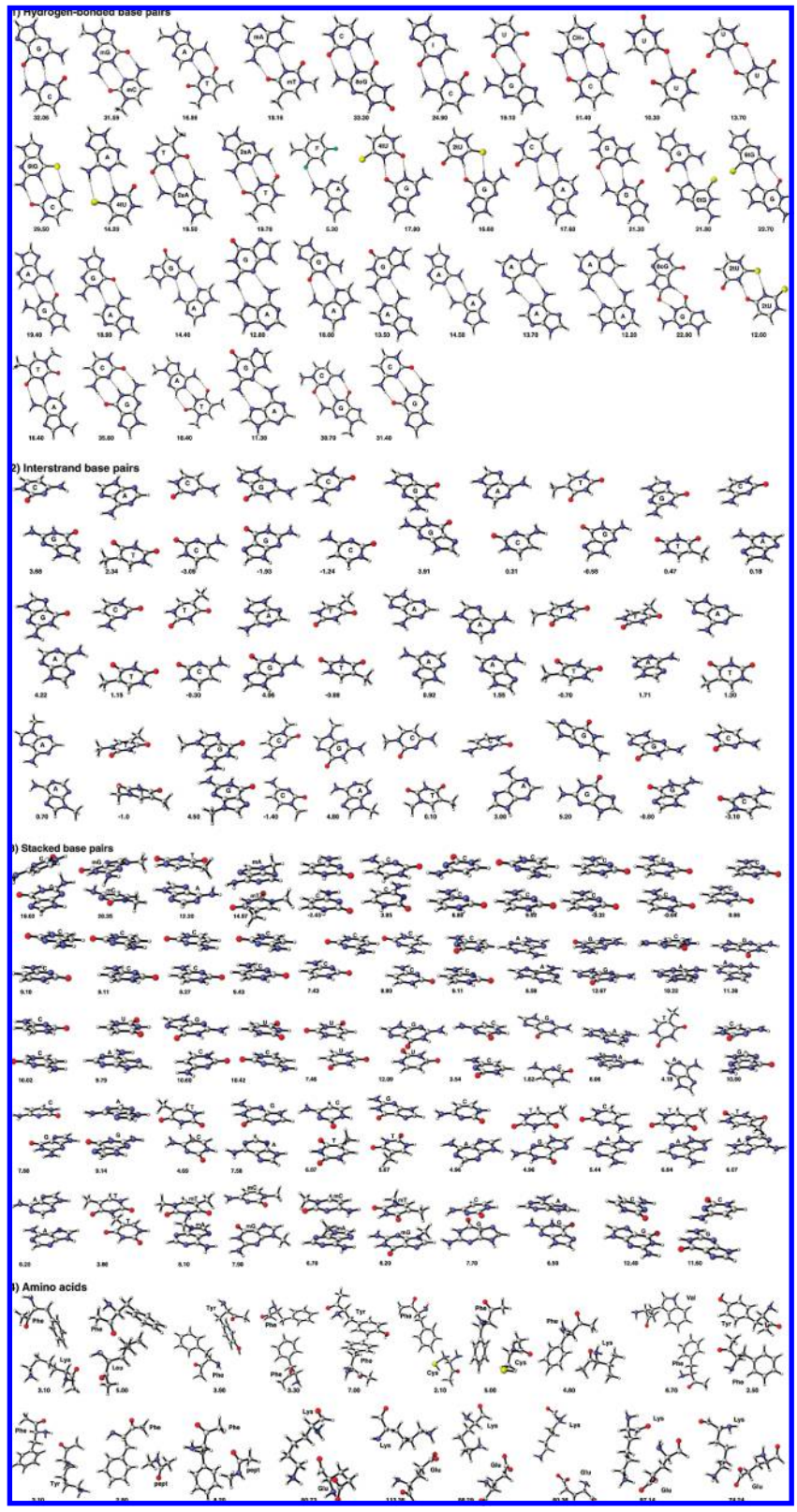


Figure 2. Molecules contained in the JSCH-2005 data set with benchmark interaction energies in kcal mol^{-1}. Abbreviations used: A = adenine, C = cytosine, F = difluorotoluene, G = guanine, I = inosine, T = thymine, U = uracil, a = amino, m = methyl, o = oxy, t = thio, pept = peptide bond, with standard three letter amino acid codes.

The complex geometries in the JSCH-2005 data set were taken either from ab initio optimization or from macromolecular structures experimentally determined from crystallographic and NMR studies. The high-level stabilization energies of Hobza and co-workers were corrected for the effects of BSSE. For complexes with experimentally determined geometries (such as the position of two interstrand DNA base pairs), geometries were not optimized, and the energies were obtained with single-point calculations. All of the complex geometries in the S22 data set were optimized and the interaction energies are corrected for the effects of BSSE. When making a comparison with ab initio optimized

Table 2. Interaction Energies (in kcal mol^{-1}) for Complexes in the S22 Data Set[a]

S22 test set	CCSD(T)/CBS	CH_4	CH_4	Amber*	OPLSA*	OPLSAA	MMFF94	MMFF94s	BHandH
				H-bonded complexes					
ammonia dimer	3.17	1.64	1.16	4.64	4.07	2.47	3.02	3.02	4.85
water dimer	5.02	4.87	7.11	7.09	6.84	6.72	6.61	6.61	7.6
formic acid dimer	18.61	14.21	3.95	10.74	13.76	12.61	16.99	16.99	26.68
fomamide dimer	15.96	9.63	8.49	15.04	16.01	13.80	12.26	12.26	21.64
uracil dimer	20.65	10.80	4.20	10.34	12.50	12.52	13.85	13.85	26.34
pyridoxine/aminopyridine	16.71	3.35	3.66	11.92	9.39	9.96	10.25	10.60	21.69
adenine/thymine (WC)	16.37	4.55	3.68	11.30	10.64	10.65	11.41	11.60	21.24
MSE		*−6.89*	*−9.18*	*−3.63*	*−3.33*	*−3.97*	*−3.16*	***−3.08***	*4.79*
MUE		*6.89*	*9.78*	*4.64*	*4.12*	*4.45*	*3.61*	***3.53***	*4.79*
				complexes with predominant dispersion contribution					
methane dimer	0.53	0.91	0.41	0.31	0.60	0.52	0.40	0.40	0.82
ethene dimer	1.51	0.98	0.77	0.98	1.02	1.07	0.74	0.74	2.53
benzene/methane	1.50	1.20	0.86	0.85	1.12	1.06	0.62	0.62	1.89
benzene dimer (PD)	2.73	3.69	2.42	2.86	2.12	2.11	1.88	1.88	1.91
pyrazine dimer	4.42	5.22	4.80	6.60	5.45	5.44	4.86	4.86	3.87
uracil dimer	10.12	9.88	3.69	8.69	7.30	7.18	12.20	12.20	10.63
indole/benzene (stack)	5.22	4.58	3.71	6.40	3.50	5.96	5.03	5.03	3.69
adenine/thymine (stack)	12.23	5.72	7.66	10.68	9.79	9.86	12.71	12.68	12.31
MSE		*−0.77*	*−1.74*	*−0.11*	*−0.92*	*−0.63*	*0.02*	*0.02*	***−0.13***
MUE		*1.30*	*1.84*	*0.98*	*1.20*	*1.07*	*0.73*	*0.72*	***0.70***
				mixed complexes					
ethene/ethyne	1.53	0.48	0.65	1.03	0.96	0.85	0.61	0.61	2.29
benzene/water	3.28	1.35	2.64	5.54	3.86	3.76	3.81	3.81	4.5
benzene/ammonia	2.35	1.51	1.88	3.44	2.57	2.16	2.48	2.48	3.07
benzene/HCN	4.46	1.25	3.07	5.91	3.78	3.60	3.66	3.66	6.34
benzene dimer (t-shape)	2.74	1.57	1.87	2.64	2.15	2.15	1.78	1.78	3.12
indole/benzene (t-shape)	5.73	2.09	2.26	6.34	6.16	5.91	5.00	5.00	6.12
phenol dimer	7.05	4.67	3.10	6.84	9.08	7.99	7.21	7.21	9.17
MSE		*−2.03*	*−1.67*	*0.66*	*0.20*	***−0.10***	*−0.37*	*−0.37*	*1.07*
MUE		*2.03*	*1.67*	*0.89*	*0.73*	***0.56***	*0.60*	*0.60*	*1.07*
overall MUE		*3.31*	*4.31*	*2.12*	*1.98*	*1.98*	*1.61*	***1.58***	*2.10*

[a] BHandHLYP/aug- ccpVTZ values taken from ref 34. WC = Watson-Crick, PD = parallel displaced, MSE = mean signed error, MUE = mean unsigned error. Boldface text indicates the best performing method (lowest MUE) for each subset and overall.

complexes, full geometry optimization has been performed at the molecular mechanics level. Polak-Ribiere conjugate gradient minimization[29] was employed, with an energy-derivative convergence criterion of 0.005 kJ mol^{-1} Å^{-1}. For unoptimized geometries for which the ab initio interaction energy is calculated from a single point calculation, the position of the hydrogen atoms were optimized at the molecular mechanics level while heavy atoms were constrained with a harmonic restraining potential (force constant 500 kcal mol^{-1} Å^{-2}). The interaction energy was obtained for each complex from the difference in energies between the dimer and the two isolated monomer geometries. For optimized complexes the monomer energies were obtained following optimization, while for constrained complexes the monomers were subject to the same constrained optimization in which heavy atoms are fixed by steep harmonic potentials.

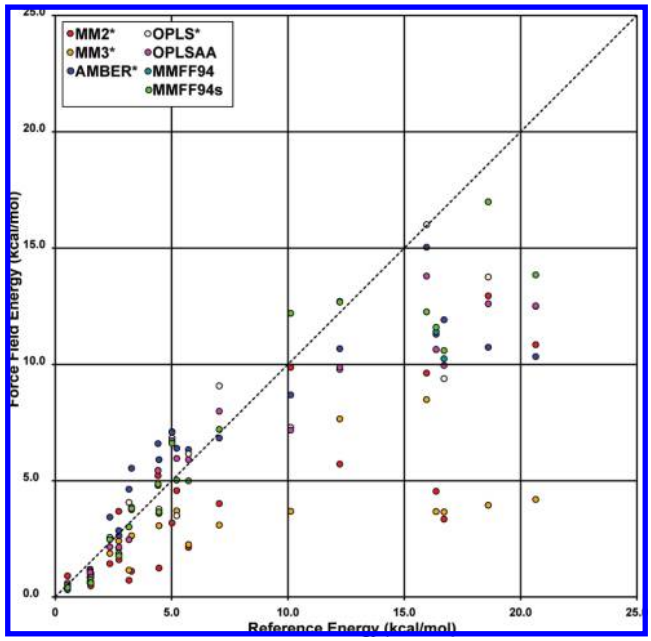
The MM2* force field was examined with and without explicit lone pairs on tetrahedral oxygen and nitrogen atoms, and the AMBER* force field was examined with and without explicit lone pairs on sulfur atoms were appropriate. The MMFF and MMFFs force fields were examined separately, but they gave identical results for complexes that do not contain sp^2 nitrogen atoms. AMBER* was also investigated using the option of an explicit 10,12 Lennard-Jones hydrogen bonding potential. Molecular mechanics complexation energies for the data sets were evaluated relative to the ab initio results using the mean signed error, MSE = $(1/n)\sum(E^{QM}_{complex} - E^{MM}_{complex})$ and the mean unsigned error, MUE = $(1/n)\sum|E^{QM}_{complex} - E^{MM}_{complex}|$.

RESULTS FOR THE S22 DATA SET

The 22 weakly bound molecular complexes of the S22 database are divided into three subgroups: hydrogen bonded, complexes with predominant dispersion interactions, and complexes for which a mixture of hydrogen bonding and dispersion are important. The set spans a wide range of interaction strengths in order to represent the diversity of interactions in macromolecules. In each subgroup the complexation energies range between 3 and 20, 0.5 and 15, and 1.2 and 8 kcal mol^{-1}, respectively. The benchmark complex geometries were obtained from full gradient optimization at the CCSD(T)/cc-pVQZ level (smaller complexes) or counterpoise-corrected gradient optimization at the MP2/cc-pVTZ level (larger complexes). The interaction energy is calculated at the MP2 level, extrapolating to the complete basis set (CBS) limit, while higher-order correlation energy contributions are accounted for with a CCSD(T) correction. High-quality interaction energies are available for the smaller complexes in the work of Truhlar[30−32] and Grimme.[33] However, the values for the larger complexes in the S22 database represent the most reliable estimates up to date.

Table 2 shows the calculated interaction energies along with the S22 benchmark values. This data set has been widely used in the benchmarking of ab initio and DFT methods,[34] and for comparison we tabulate recent DFT values obtained by Lovas and co-workers using the BHandHLYP functional,[35] which has been recommended as a computationally efficient means to study stacked systems, giving good agreement with MP2 and CCSD(T) energetics.[36] This DFT approach gives a MUE of 2.10 kcal mol^{-1}. Grimme has also

Chart 1. Comparison of Force Field Interaction Energy with Ab Initio Values for the S22 Data Set[a]



[a] Dashed line is $y = x$.

used Quantum Monte Carlo (QMC) calculations to obtain a MUE of 0.68 kcal mol^{-1} across the S22 data set in a large scale distributed computing project.[37] Interaction energies obtained with the OPLS and MMFF families of force fields actually outperform the DFT results overall and are comparable with the QMC results for the dispersion and mixed subsets. The overall performance is let down for these and all other force fields because of the underestimation of interaction energies largely caused by the hydrogen bonding.

The largest discrepancies between the ab initio and molecular mechanics complexation energies are observed for complexes for which the predominant interaction is hydrogen bonding. All force fields were found to systematically underestimate the stabilization of the hydrogen bonded complexes. The performance for complexes in which the predominant interaction is dispersion, and complexes stabilized by a mixture of hydrogen bonding and dispersion is much more satisfactory, with all of the force fields bar MM2* and MM3* giving a MUE less than 1.2 kcal mol^{-1} for these categories. A comparison of the calculated force field interaction energies against the ab initio values is shown in Chart 1 (an interactive version of this chart which allows the superimposition of benchmark and molecular mechanics geometries can be accessed at http://www-jmg.ch.cam.ac.uk/data/forcefield/). Evidently the underestimation of complexation energies is most prominent for the more strongly bound complexes, which are those bound by hydrogen bonding interactions. These discrepancies are particularly noticeable for the MM2* and MM3* force fields. Results for AMBER* are shown without explicit 10,12 Lennard-Jones hydrogen bonding, which fared marginally better in terms of MUE and RMSE by 0.06 and 0.09 kcal mol^{-1} respectively (for full results with explicit hydrogen bonding see Supporting Information).

Relative to the ab initio optimized structures, the geometries of hydrogen bonded complexes are described fairly well, except for the MM3* and AMBER* geometries of the 2-pyridoxine: 2-aminopyridine dimer which pucker from the planar arrangement, and the MM2* (stacked) and MM3* (puckered) optimized geometries of the planar adenine: thymine dimer. For predominantly dispersion-bonded complexes, deviations from the ab initio geometry are observed for the parallel-displaced benzene dimer (MM2* shifts from a displaced geometry) and for the pyrazine dimer (all force fields shift from the displaced geometry). Significant differences to the reference data set were also seen for the stacked uracil dimer (MM2* and MMFF(s) give hydrogen bonded structures), the indole: benzene dimer (all bar MM2* and MM3* give a T-shaped structure) and the adenine-thymine dimer (MMFF(s) give hydrogen bonded structures). For the mixed complexes, the positions of the heavy atoms in the complexes of benzene with water and ammonia were reproduced well, although the orientation of the O−H and N−H bonds differ between the force fields. The orientation of HCN in the benzene: HCN dimer also shows some variation. Geometries of the T-shaped benzene dimer and benzene: indole dimer are in good agreement bar the MM3* results. Structures of the phenol dimer are in broad agreement except for MM2* (stacked) and MM3* (hydrogen bonded). The average and maximum rms error between superimposed force field and ab initio geometry is shown for each force field in Table 3.

Molecular mechanics optimized geometries are shown alongside the benchmark structures in Figure 3. Structures are color coded according to the force field used in the optimization.

RESULTS FOR THE JSCH-2005 DATA SET

Force field energies are compared against benchmark interaction energies from accurate MP2 and CCSD(T) complete basis set calculations. In the JSCH-2005 data set, the geometries of thirty two hydrogen bonded DNA base pairs and four stacked base pairs were obtained by full ab initio geometry optimization. For our comparison with the ab initio geometry in these cases the complex was also fully optimized at the molecular mechanics level. The remaining nucleic acid complexes in the data set are not fully optimized: they were either taken from crystal structures and only the positions of H atoms were optimized, or the interbase geometries were fixed and intrabase geometries were optimized at the DFT level. Neutral amino acid pairs containing phenylalanine were taken from the hydrophobic core of Rubredoxin. Positions of heavy atoms were obtained from the PDB crystal structure (1RB9) while the positions of H atoms were optimized. The charged amino acid pairs were taken from PDB crystal structures of rubredoxin (1IU5, 1BQ9, 1SMM, 1BRF). For our force field calculations on complexes which were not fully optimized at the ab initio level, only the positions of H atoms were fully optimized and heavy atoms were fixed as explained in the Computational Methods section.

The coordinates of JSCH2005 data set were taken from the Supporting Information of the original publication.[38] For one of the hydrogen bonded entries (cytosine, protonated cytosine dimer) two resonance structures of the protonated molecule were investigated in which a different N atom bears the formal +1 charge (Figure 4). However, both of these resonance descriptions led to a calculated

Table 3. Mean and Maximum RMS Errors after Superimposition of Benchmark Geometries with Those Obtained from Full Force Field Optimization

S22 data set	MM2*	MM3*	Amber*	OPLSA*	OPLSAA	MMFF94	MMFF94s
H-bonded complexes (7 structures)							
mean rms error (Å)	0.42	0.35	0.19	0.13	0.11	0.13	0.11
max rms error (Å)	1.61	0.92	0.48	0.25	0.25	0.24	0.21
complexes with predominant dispersion (8 structures)							
mean rms error (Å)	0.59	0.30	0.43	0.32	0.40	0.99	1.01
max rms error (Å)	2.39	0.61	1.56	1.03	1.56	2.34	2.45
mixed complexes (7 structures)							
mean rms error (Å)	0.53	0.46	0.19	0.20	0.15	0.28	0.28
max rms error (Å)	2.17	1.08	0.45	0.61	0.38	0.61	0.61

complexation energy very different to the benchmark value of 46.5 kcal mol^{-1} even though the geometries were good. The MM2* complexation energy is an underestimate by 37.5 kcal mol^{-1}. Although the origin of this difference is not clear, this result was significantly worse than for any other complex, so we decided not to include this complex in the statistical analysis of the results. This was the only formally charged base pair dimer in the data set. It therefore seems likely that the problem arises due to the confinement of a positive charge on either of the two ring N atoms, whereas in reality it is delocalized around the pyrimidine ring.

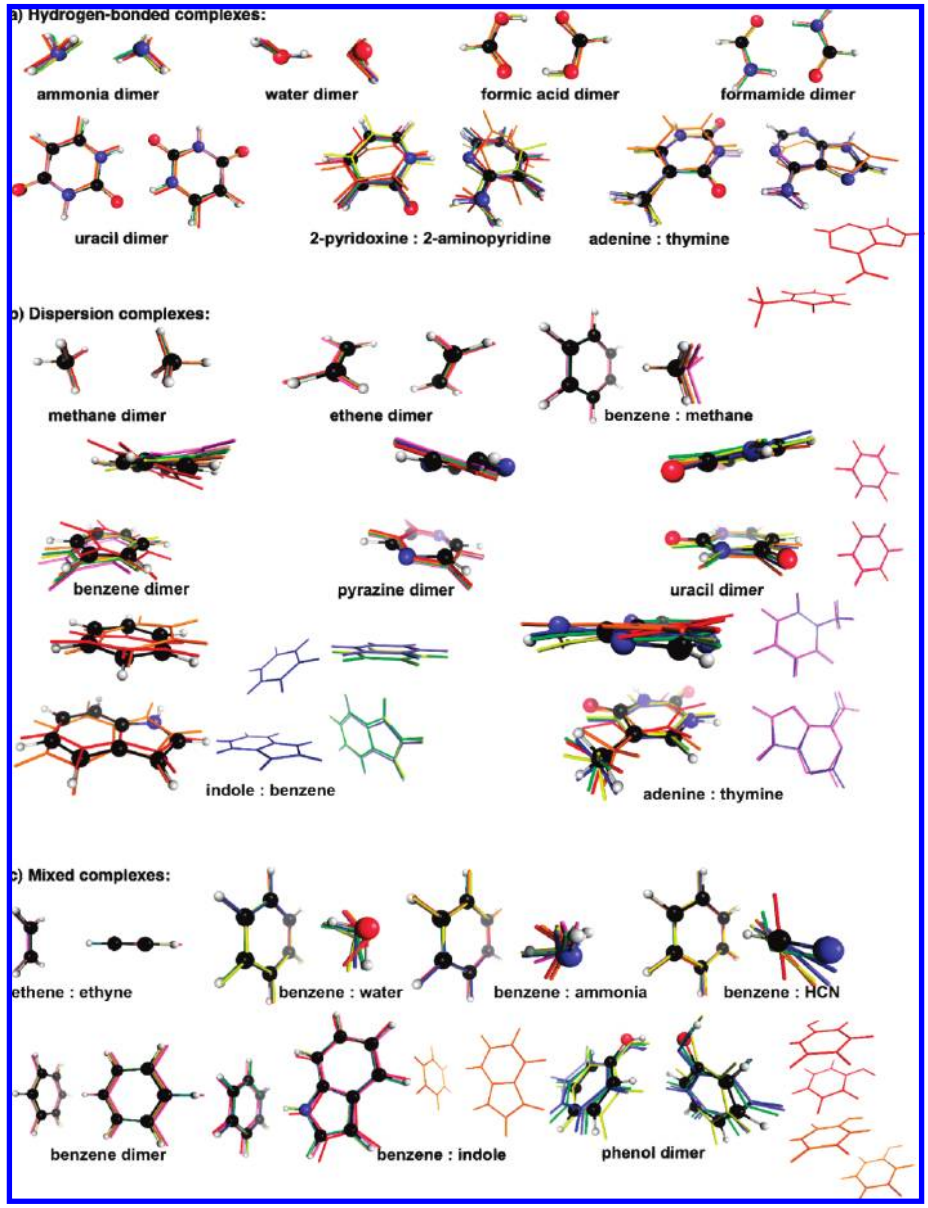


Figure 3. Optimized complexes from the S22 data set, colored by force field: MM2* (red), MM3* (orange), AMBER* (yellow), OPLSA* (green), OPLSAA (blue), MMFF (indigo), and MMFFs (violet). The ab initio optimized geometry is shown in ball and stick representation.

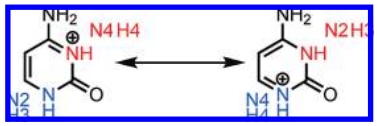


Figure 4. Two resonance structures of protonated cytosine use different atom type descriptions.

The JSCH-2005 ab initio interaction energies for hydrogen-bonded complexes all exceed 10 kcal mol^{-1} and for complexes with three hydrogen bonds are almost 30 kcal mol^{-1}. Although stacked complexes are less favorable, their interaction energies are usually larger than 10 kcal mol^{-1} and reach 18 kcal mol^{-1}. Interaction energies of neutral amino acid pairs reach 8 kcal mol^{-1} without hydrogen bonding or π-stacking contributions, largely due to the phenylalanine-peptide interaction. The Interaction energy between charged amino acid pairs is very large and in one case exceeds 100 kcal mol^{-1}.

Interaction energies were obtained for all complexes in the JSCH-2005 data set. Force field performance was quantified in terms of the mean signed error (MSE) and mean unsigned error (MUE). A negative MSE means that the force field underestimates the favorable complexation energy (i.e., the complex is destabilized relative to the ab initio results) and vice-versa. Using the MUE, the best results were obtained using infinite cut-offs, although the difference between this and default cutoff values was smaller than 0.3 kcal mol^{-1}. This is quite a small effect, but would become significant in systems, such as enzymes, where there are many more pairs of atoms affected by the cut-offs.

The use of explicit 10,12 Lennard-Jones hydrogen bonding in AMBER* increased the MUE by 0.14 kcal mol^{-1} for the hydrogen bonded base pairs. The performance of each force field for each of the sub groups is shown in Table 4 (full results showing the individual complexes are detailed in the Supporting Information). The overall MUE is shown with infinite and default nonbonded cut-offs. The best performing force field in terms of both MSE and MUE was found to be OPLSAA.

One of the key findings from the ab initio calculations on the JSCH-2005 data set was that the complexation energies of hydrogen bonded dimers were larger than previous estimates.[10] The complexation energies of the hydrogen bonded base pairs obtained from out molecular mechanics optimizations are almost always underestimated relative to the ab initio benchmarks. MM2* and MM3* show large mean errors of around 13 kcal mol^{-1}, while the other force fields do better, although all still produce average errors greater than 4 kcal mol^{-1} for the hydrogen bonded subset. This is a significant underestimate of the stabilization energy, which represents a 25% error in relation to the average interaction energy of 17.7 kcal mol^{-1} even for the best performing force field (MMFF94). By comparison, complexation energies for the interstrand base pairs show much smaller mean errors, although that the magnitude of these interactions are also much smaller. Stacking interactions and amino acid interactions are described with much greater precision. A comparison of the calculated molecular mechanics interaction energy against the ab initio energy is shown for each of the JSCH-2005 subgroups below (Chart 2). The underestimation of hydrogen bonding interactions is clear since nearly all points lie below $y = x$, while much better performance is seen for the interstrand bases, stacked bases and amino acid subgroups. This result follows on from Hobza's findings that high level interaction energies of hydrogen bonded complexes are much larger than previous calculated results.[11] The existing hydrogen bonding force field parameters will have to be reparameterized to reflect the revised ab initio values. An interactive version of these data can also be accessed at http://www-jmg.ch.cam.ac.uk/data/forcefield/ where the geometries can be visualized using *Jmol*.[39] The resulting force field geometries can be superimposed on benchmark geometries for a visual comparison.

Resulting force field complex geometries were superimposed with benchmark (*ab initio* and experimental) structures using a Java translation of the quaternion fit program, qtrfit, of David J. Heisterberg (Ohio Supercomputer Center).[40] The rms errors of the geometries obtained by constrained optimization (experimental benchmark geometry) are small, as expected. For fully optimized hydrogen bonded and stacked structures the rms errors are larger, and there are particular complexes for which the force fields fail noticeably: *ab initio* planar hydrogen bonded complexes optimize to stacked geometries and vice-versa at the molecular mechanics level. The average and maximum rms errors are shown for each force field in Table 5.

In some cases, molecular mechanics optimization leads to a geometry that is very different to the ab initio optimized structure. In the hydrogen-bonded subset, optimized geometries of the adenine/4-thiouracil, adenine/difluorotoluene, and guanine/2-thiouracil complexes are no longer in a planar hydrogen bonded conformation for all force fields. These complexes all optimize to a stacked geometry. For the four optimized stacked geometries most force fields give geometries which are similar to the ab initio structures, with a few exceptions (Figure 5).

The MM2* and MM3* force fields both perform well in optimizing the stacked complexes, giving geometries similar to the ab initio optimized structures (Figure 4). The hydrogen bonded complex should be a more stable global minimum, but some of the other force fields tend to miss the stacked minimum altogether. The success of MM2* and MM3* is due to the smaller magnitude of the calculated hydrogen bonding interactions, but also due to the use of explicit hydrogen bonding potentials which use default 4 Å cutoffs. This means that the longer-range electrostatic attractions of polar-hydrogens will be reduced in the parametrization process, thus stabilizing stacked local minima.

CONCLUSIONS

We have assessed the ability of various widely used force fields to describe molecular complexes that include a range of nucleic acid bases and amino acids. Geometries and complexation energies have been compared with ab initio benchmark values from the S22 and JSCH-2005 data sets.[41] We conclude the following:

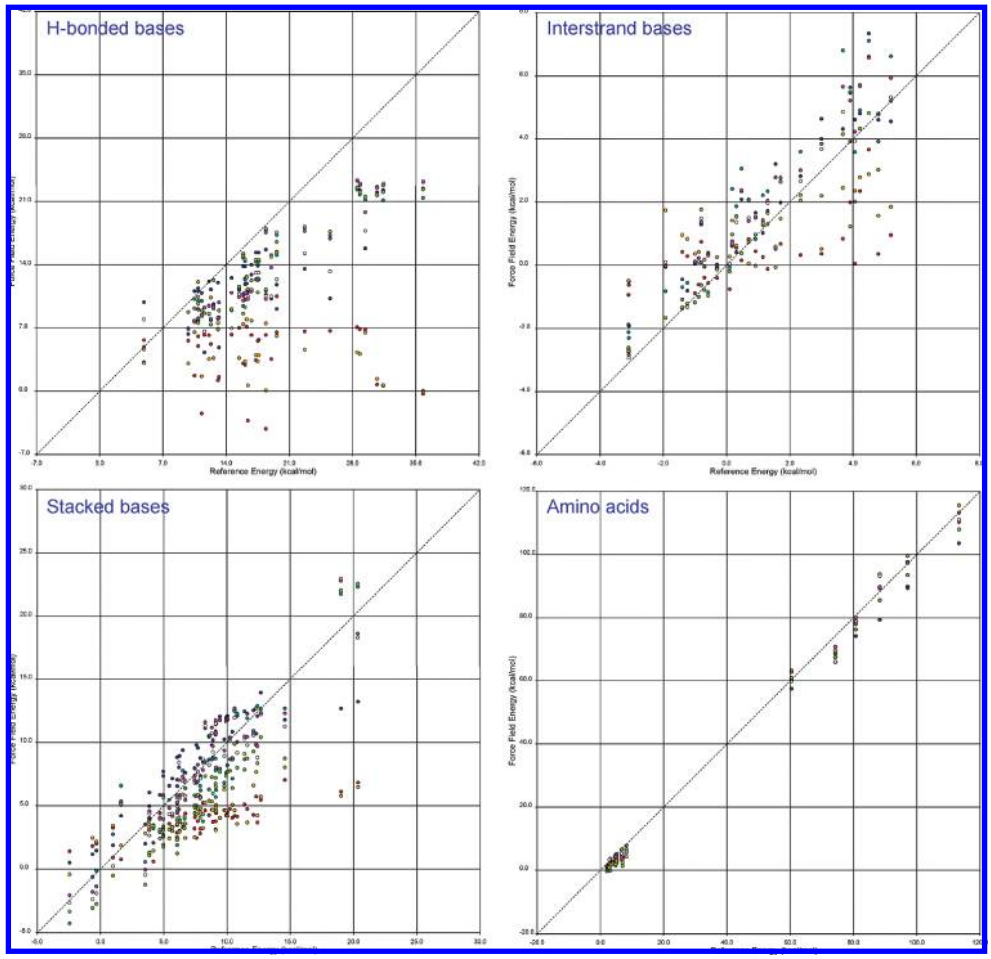
(1) In describing the geometries and complexation energies of intermolecular complexes we recommend the use of the OPLS and MMFF families of force fields, which outperform the other force fields (MM2, MM3, AMBER) we tested. In particular, the OPLSAA and MMFF94s force fields give the best reproduction of ab initio energies and geometries for the S22 and JSCH2005 data sets.

Table 4. Interaction Energies (in kcal mol^{-1}) for Complexes in the JSCH-2005 Data Set Calculated with Infinite Nonbonded Cut-Offs[a]

JSCH2005 data set	MM2*	MM2* (LP)	MM3*	Amber*	Amber* (LP)	OPLSA*	OPLSAA	MMFF94	MMFF94s
H-bonded DNA base pairs (38 structures, 32 optimized geometries) mean complexation energy = 17.75									
MSE	−13.04	−13.04	−13.09	−4.84	−4.47	−5.18	**−4.80**	−4.38	−4.55
MUE	12.74	12.74	13.13	5.04	4.68	5.21	**4.68**	4.26	4.43
interstrand base pairs (32 structures, 4 optimized geometries) mean complexation energy = 1.13									
MSE	−0.60	−0.61	0.14	0.86	0.86	0.80	0.89	0.82	**−0.06**
MUE	1.37	1.37	1.43	0.87	0.87	0.7	0.85	0.97	**0.41**
stacked base pairs (54 structures, experimental geometries) mean complexation energy = 7.84									
MSE	−3.96	−3.96	−3.46	0.92	0.92	−0.33	**−0.11**	−1.08	−2.44
MUE	4.37	4.37	3.91	1.75	1.75	1.45	**1.32**	1.74	2.77
amino acid pairs (19 structures, experimental geometries) mean complexation energy = 30.81									
MSE	−2.08	−2.41	−1.30	−3.21	−3.23	−1.02	**−0.94**	−2.71	−2.73
MUE	2.18	2.42	2.70	3.21	3.23	1.72	**1.39**	2.77	2.78
overall MUE	4.76	4.80	4.36	2.41	2.31	2.10	**1.87**	2.13	2.47
(default cut-offs)	4.88	4.55	4.41	2.59	2.51	2.44	**2.16**	2.37	2.68

[a] Bolface text indicates best performing method (lowest MUE) for each subset and overall.

Chart 2. Comparison of Force Field Interaction Energy with Ab Initio Values for the Four Sub-Groups of the JSCH-2005 Data Set[a]



[a] Dashed line is $y = x$.

(2) For the S22 data set, the OPLSAA and MMFF interaction energies are as good as those obtained with much more expensive DFT calculations, giving a MUE of 1.98 and 1.58 kcal mol^{-1}, respectively. The S22 geometries are mostly well reproduced by these two force fields, with the exception of the indole/benzene and adenine/thymine stacked complexes, which makes them an attractive alternative to DFT calculations.

(3) For the JSCH2005 data set, the OPLSAA and MMFF94 interaction energies give a MUE of 2.24 and 2.42 kcal mol^{-1} respectively. These figures are adversely affected by the severe underestimation of hydrogen bonding inter-

Table 5. Mean and Maximum RMS Errors after Superimposition of Benchmark Geometries with Those Obtained Following Full and Constrained Optimization with the Force Fields Shown[a]

JSCH2005 data set	MM2*	MM2* (LP)	MM3*	Amber*	Amber* (LP)	OPLSA*	OPLSAA	MMFF94	MMFF94s
full optimization (36 structures)									
mean rms error	1.314	1.314	1.414	**0.478**	0.638	0.550	0.629	0.735	0.540
max rms error	3.888	3.888	3.388	**3.887**	4.176	3.887	3.887	3.887	3.887
constrained optimization (107 structures)									
mean rms error	0.131	0.133	0.274	0.163	0.163	0.148	0.154	0.188	0.137
max rms error	0.339	0.344	0.559	0.435	0.435	0.431	0.200	0.437	0.403

[a] The best performance for full optimizations is shown in bold.

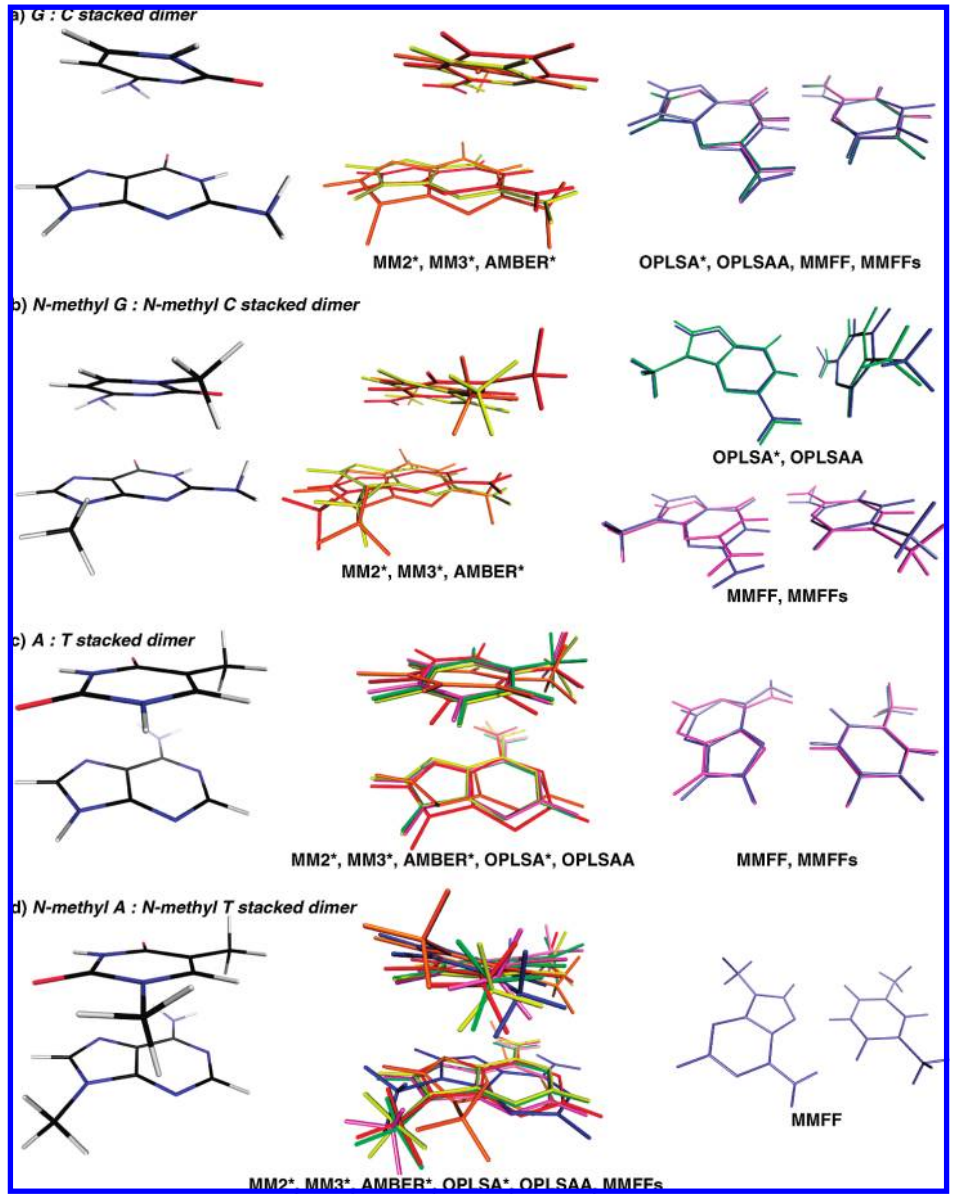


Figure 5. Optimized geometries of complexes for which stacking interactions are important. The ab initio optimized geometry is shown on the left, next to MM2* (red), MM3* (orange), AMBER* (yellow), OPLSA* (green), OPLSAA (blue), MMFF (indigo), and MMFFs (violet) geometries.

actions, which was seen for all force fields tested. Some of the optimized geometries were found to differ from the ab initio structures, such that hydrogen bonded complexes optimized to stacked structures and vice versa.

(4) High level ab initio hydrogen bonding interactions calculated at the CCSD level are stronger than previous estimates from calculations at lower levels of theory. In the light of this, existing force field hydrogen bonding parameters were found to underestimate the benchmark interaction energies. In the development or reparameterization of force fields, this issue needs to be addressed.

ACKNOWLEDGMENT

We thank Unilever and St Catharine's College, Cambridge for a Research Fellowship for R.S.P. Dr Joe Townsend is

thanked for assistance with SVG code used in the Supporting Information.

APPENDIX I

A description of the force fields evaluated in this study.

MM2*. Allinger's 1987 MM2 parameter set for simple organic molecules with some changes.[12] The electrostatic treatment in MM2 consists of dipole interactions, while MM2* instead uses a Coulomb's law treatment of charge−charge interactions with a distance-dependent dielectric constant as default, with atom point charges calculated from the MM2 bond dipoles. The difference is most important when selection rules in *MacroModel* exclude part of the electrostatic interaction present in MM2. The treatment of π-systems also differs, with MM2* using a torsional barrier description of conjugation. The substructure feature is used to recognize some conjugated systems (e.g., benzene and a range of heterocycles) and apply special parameters. No long-range effects from, for example, π-active benzene substituents are recognized by MM2*. Hydrogen bonding in MM2* has been modeled after AMBER, not MM2. The inclusion of explicit lone-pairs is optional for tetrahedral (sp^3) oxygen and nitrogen atoms.

MM3*. Derived from Allinger's 1990 MM3 parameter set with additions in a manner analogous to the creation of the MM2* force field from MM2 and differs from the authentic field by use of a Coulomb's law treatment of electrostatics and torsional barrier treatment of conjugation.[13]

AMBER*. Kollman's united atom and all atom fields with additional parameters for organic functionality.[14] AMBER* can be set to mimic different versions of AMBER.[15] By default a constant dielectric treatment is used, with charges assigned from substructure recognition and bond dipoles in a way that exactly reproduces the united atom AMBER charges and Kollman's 6,12 Lennard-Jones treatment of hydrogen bonding. AMBER* is a purely diagonal, harmonic force field, containing no cross terms (e.g., stretch−bend interactions). Many of the parameters for some of the more complex amino acid side chains (phenol, indole, imidazole, etc) are given separately from the actual residue itself in the substructure section. The inclusion of explicit lone-pairs is optional for sulfur atoms.

OPLS*. Jorgensen's nonbonded parameter set with AMBER bonded functions for liquid simulations.[16] Potential functions have the Coulomb form (electrostatics) plus Lennard-Jones form (van der Waals) and parameters were obtained and tested primarily in conjunction with Monte Carlo statistical mechanics simulations of pure organic liquids and numerous aqueous solutions of organic ions representative of subunits in the side chains and backbones of proteins.

OPLSAA. All-atom OPLS force field for organic molecules and peptides.[17] Parameters for both torsional and nonbonded energetics were derived from simulations of liquids and solutions, while the bond stretching and angle bending parameters have been adopted mostly from AMBER. The functional form of nonbonded interactions is the same as OPLS*.

MMFF94/MMFF94s. MMFF94 has been parametrized for a wide variety of chemical systems, using high-quality ab initio data at the MP2/6−31G level, although in many cases employing larger basis sets and/or higher correlation methods.[18] Where appropriate, MMFF94 has been supplemented and extended by parametrizing against a large number of crystallographically determined structures. MMFF94s was introduced as a static variant of MMFF94.[19] Appropriate out of plane and torsional parameters were modified to provide (nearly) planar structures at amide and unsaturated amine trigonal nitrogens. The van der Waals interactions are described by a buffered 7,14-Lennard-Jones potential, electrostatic interactions by a buffered constant dielectric constant and hydrogen bonding interactions by a 6,12-Lennard-Jones potential.

Supporting Information Available: Energies of complexation, monomer and dimer energies for each complex. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Černý, J.; Hobza, P. Non-covalent Interactions in Biomacromolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291–5303.

(2) Pyykkö, P. Strong Closed-Shell Interactions in Inorganic Chemistry. *Chem. Rev.* **1997**, *97*, 597–636.

(3) Šponer, J.; Riley, K. E.; Hobza, P. Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610.

(4) (a) For a review of ab initio studies on the structures, energetics and dynamics of nucleic acids see: (b) Hobza, P.; Šponer, J. Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab Initio Calculations. *Chem. Rev.* **1999**, *99*, 3247–3276.

(5) Grimme, S. Do Special Noncovalent $\pi-\pi$ Stacking Interactions Really Exist? *Angew. Chem., Int. Ed.* **2008**, *47*, 3430–3434.

(6) Hunter, C. A.; Sanders, J. K. M. The Nature of $\pi-\pi$ Interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

(7) Wheeler, S. E.; Houk, K. N. Substituent Effects in the Benzene Dimer are Due to Direct Interactions of the Substituents with the Unsubstituted Benzene Ring. *J. Am. Chem. Soc.* **2008**, *130*, 10854–10855.

(8) (a) Simón, L.; Goodman, J. M. Theoretical Study of the Mechanism of Hantzsch Ester Hydrogenation of Imines Catalyzed by Chiral BINOL-Phosphoric Acids. *J. Am. Chem. Soc.* **2008**, *130*, 8741–8747. (b) Paton, R. S.; Goodman, J. M.; Pellegrinet, S. C. Theoretical Study of the Asymmetric Conjugate Alkenylation of Enones Catalyzed by Binaphthols. *J. Org. Chem.* **2008**, *73*, 5078–5089. (c) Paton, R. S.; Goodman, J. M. 1,5-Anti Stereocontrol in the Boron-Mediated Aldol Reactions of β-Alkoxy Methyl Ketones: The Role of the Formyl Hydrogen Bond. *J. Org. Chem.* **2008**, *73*, 1253–1263. (d) Paton, R. S.; Goodman, J. M.; Pellegrinet, S. C. Mechanistic Insights into the Catalytic Asymmetric Allylboration of Ketones: Brønsted or Lewis Acid Activation. *Org. Lett.* **2008**, *11*, 37–40.

(9) (a) Fedorov, M. V.; Goodman, J. M.; Schumm, S. Solvent Effects and Hydration of a Tripeptide in Sodium Halide Aqueous Solutions: An *In Silico* Study. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5423–5435. (b) Fedorov, M. V.; Goodman, J. M.; Schumm, S. The effect of sodium chloride on poly-L-glutamate conformation. *Chem. Commun.* **2009**, 896−898.

(10) Gao, J. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. *Rev. Comput. Chem.* **1996**, *7*, 119–185.

(11) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(12) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.

(13) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(14) McDonald, D. Q.; Still, W. C. AMBER Torsional Parameters for the Peptide Backbone. *Tetrahedron Lett.* **1992**, *33*, 7743–7746.

(15) (a) 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. (b) 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.

(16) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(18) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, And Performance of Mmff94. *J. Comput. Chem.* **1996**, *17*, 490–519, 520−552, 553−586, 587−615, 616−641 and MMFF VII. Characterization of MMFF94, MMFF94s, and Other Widely Available Force Fields for Conformational Energies and for Intermolecular-Interaction Energies and Geometries. *J. Comput Chem.* **1999**, 20, 730−748.
(19) Halgren, T. A. MMFF VI. MMFF94s Option for Energy Minimization Studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
(20) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Acromodel−An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.
(21) Figures prepared with CYLview: Legault, C. Y. CYL*view*, version 1.0b. http://www.cylview.org (accessed Dec 02 2008).
(22) Jurečka, P.; Hobza, P. True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine···Cytosine, Adenine···Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.
(23) Šponer, J.; Jurečka, P.; Hobza, P. Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
(24) Dbkowska, I.; Gonzalez, H. V.; Jurečka, P.; Hobza, P. Stabilization Energies of the Hydrogen-Bonded and Stacked Structures of Nucleic Acid Base Pairs in the Crystal Geometries of CG, AT, and AC DNA Steps and in the NMR Geometry of the 5′-d(GCGAAGC)-3′ Hairpin: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels. *J. Phys. Chem. A* **2005**, *109*, 1131–1136.
(25) Perez, A.; Šponer, J.; Jurečka, P.; Hobza, P.; Luque, F. J.; Orozco, M. Are the Hydrogen Bonds of RNA (A−U) Stronger than Those of DNA (A−T)? A Quantum Mechanics Study. *Chem.−Eur. J.* **2005**, *11*, 5062–5066.
(26) Jurečka, P.; Šponer, J.; Hobza, P. Potential Energy Surface of the Cytosine Dimer: MP2 Complete Basis Set Limit Interaction Energies, CCSD(T) Correction Term, And Comparison with the AMBER Force Field. *J. Phys. Chem. B* **2004**, *108*, 5466–5471.
(27) Šponer, J.; Jurečka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. Nature of Base Stacking: Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps. *Chem.−Eur. J.* **2006**, *12*, 2854–2865.
(28) Vondrásek, J.; Bendová, L.; Klusák, 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.
(29) Polak, E.; Ribiere, G. Note sur la Convergence des Méthodes de Directions Conjuguées (Note on the Convergence of Conjugate Directions Methods). *Revue Française Inf. Rech. Oper.* **1969**, *16-R1*, 35–43.
(30) Zhao, Y.; Truhlar, D. G. How Well Can New-Generation Density Functional Methods Describe Stacking Interactions in Biological Systems. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2701–2705.
(31) Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.
(32) Zhao, Y.; Truhlar, D. G. Infinite-Basis Calculations of Binding Energies for the Hydrogen Bonded and Stacked Tetramers of Formic Acid and Formamide and Their Use for Validation of Hybrid DFT and Ab Initio Methods. *J. Phys. Chem. A* **2005**, *109*, 6624–6627.
(33) Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
(34) For recent examples, see: (a) DFT studies Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 1432–2234. (b) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406. (c) ab initio Hill, J. G.; Platts, J. A. Spin-Component Scaling Methods for Weak and Stacking Interactions. *J. Chem. Theory Comput.* **2007**, *3*, 80–85. (d) semi-empirical McNamara, J. P.; Hillier, I. H. Semi-empirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Prediction of the Full Range of Intermolecular Interactions in Biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362–2370.
(35) Csontos, J.; Palermo, N. Y.; Murphy, R. F.; Lovas, S. Calculation of Weakly Polar Interaction Energies in Polypeptides Using Density Functional and Local Møller−Plesset Perturbation Theory. *J. Comput. Chem.* **2008**, *29*, 1344–1352.
(36) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, J. A.; Williams, P. A. Hybrid Density Functional Theory for -Stacking Interactions: Application to Benzenes, Pyridines, and DNA Bases. *J. Comput. Chem.* **2006**, *27*, 491–504.
(37) Korth, M.; Lüchow, A.; Grimme, S. Toward the Exact Solution of the Electronic Schrödinger Equation for Noncovalent Molecular Interactions: Worldwide Distributed Quantum Monte Carlo Calculations. *J. Phys. Chem. A* **2008**, *112*, 2104–2109.
(38) We omitted two interstrand base pair complexes (originally labelled GG0/3.36 GCis036 and AA20/3.05 TAis 2005) from our analysis, as the deposited structures contained different base pairs from those described in the tabulated benchmark data. We also swapped round two neighbouring structures (originally labelled CG0/3.19 G//Gis and CG0/3.19 C//Cis), since it was apparent their names had been swapped.
(39) *Jmol*: An open-source Java viewer for chemical structures in 3D. http://www.jmol.org (accessed Dec 02 2008).
(40) Silva, M. A.; Goodman, J. M. QRC: A Rapid Method for Connecting Transition Structures to Reactants in the Computational Analysis of Organic Reactivity. *Tetrahedron Lett.* **2003**, *44*, 8233–8236.
(41) Additional information: Interactive versions of Charts 1 and 2 are available on the web at http://www-jmg.ch.cam.ac.uk/data/forcefield/ which allow for *Jmol* visualisation of the complexes and comparison with ab initio geometries. All atom-typed geometries used in this investigation are available for download in *MacroModel* (*.dat) format at the same web address.

CI900009F