Ab Initio Investigation of the Methylimidazole−Indole Complexes as Models of the Histidine−Tryptophan Pair

Giuliano Alagona* and Caterina Ghio

CNR-ICQEM, Institute of Quantum Chemistry and Molecular Energetics, Via Risorgimento 35, I-56126 Pisa, Italy

Susanna Monti

Menarini Ricerche, Via Sette Santi 3, I-50131 Florence, Italy

Received: November 25, 1997; In Final Form: March 11, 1998

Several histidine−tryptophan complexes, derived from the crystal structures available in the Brookhaven Protein Data Bank, have been examined with ab initio theoretical methods (using as model systems 5-methylimidazole and indole, respectively), in order to identify the most favorable arrangements of the two side chains, elucidating also the strength and the nature of the intermolecular interaction established between them. The equilibrium geometries of the isolated partners were optimized at the HF/6-31G* level and the interaction energy of the adducts was computed, employing the 6-31G* basis set with the *d* exponents reduced to 0.25, thus named 6-31G*(0.25), at the HF and MP2 (frozen-core approximation) levels. For a few typical orientations, the dependence of the interaction energy upon the intermolecular distance, as measured from the ring centroids, was then examined while keeping fixed reciprocal orientations and internal geometries of the partners. There is a fair linear correlation between the equilibrium distances (R_{eq}) at the MP2 level and the experimental (R_{exp}) ones and between the MP2 interaction energies at R_{eq} and those computed at R_{exp}. For three arrangements with a shallow or even repulsive HF interaction energy, the counterpoise correction to the basis set superposition error (BSSE) was introduced both at the HF and MP2 levels, using Pople's 6-31G* standard, 6-31G*(0.25), and Dunning's DZP basis sets, to test the reliability of the results obtained along the whole approaching path. This is made necessary by the noticeable displacement in the equilibrium distances usually found at the various levels. The DZP HF interaction energies turn out to be less affected by BSSEs than the 6-31G* and the 6-31G*(0.25) ones and are located in an intermediate position between them. As a general rule for these complexes, the counterpoise correction is larger at the correlated level; therefore the addition of the correlation effect to the counterpoise-corrected SCF energy produces a curve fairly close to the MP2 one that seems to represent a lower bound to the true interaction energy. Kitaura and Morokuma's decomposition analysis of the interaction energies was also carried out on these typical complexes.

Introduction

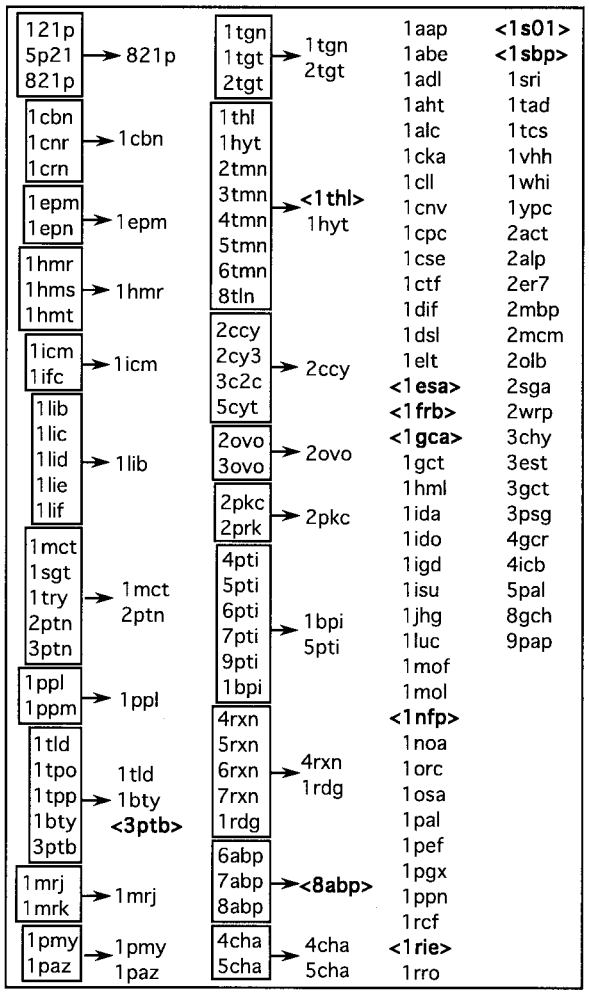
The theoretical study of molecular recognition intended as the molecular interaction between a system of limited size and a much more extended one (such as for instance the couples ligand/drug−receptor, substrate/inhibitor−enzyme, molecule−surface, etc.) represents a real challenge. In the first two examples mentioned, a noticeable complication is brought about by the conformational flexibility of both systems. Flexibility is almost always neglected in the available computer codes which consider rigid partners approaching each other and interacting without any possibility of mutual deformations and conformational changes as a consequence of the incipient or established interactions. It is, however, an extremely difficult task to overcome this inconvenience starting from scratch, because the problem closely resembles the protein folding one. Recently, we devised a tentative solution to it in an empirical way, though limiting ourselves to only one of the partners (the ligand), when we docked a good number of different conformers of a few NKA antagonists inside the NK2 receptor site,[1] model built using bacteriorhodopsin[2] as a template and site-directed mutagenesis studies.[3] From these studies, the presence of two residues of histidine (namely His 198 and His 267, belonging to the receptor site) turns out to play a major role in the interaction with both peptide and nonpeptide antagonists of tachykinines. A common feature shared by those antagonists is the presence of an indole group (as tryptophan (Trp) for the peptide ones).[4] Therefore, their activity might be explained with the preferential interaction of indole with one of the histidines.

These aromatic side chains, however, might interact in several reciprocal orientations: either stacked (and namely parallel, antiparallel, displaced, undisplaced, etc.), or T-shaped, H-bonding interactions. In this study, we follow a strategy opposite to that employed in the aforementioned paper,[1] trying to figure out which arrangements are more favorable and thus more likely to occur for that given pair of side chains with the aim of reproducing them eventually inside the receptor site.

This investigation, however, directly pertains to the vast literature, which we are not going to review, concerning the electronic and geometrical properties of bimolecular adducts and to the debate about stacked and T-shaped positions, in the gas phase[5,6] or in solution,[7,8] even inside a single molecule.[9−11] In addition, it constitutes a natural continuation of our previous studies on H-bonded dimers in the gas phase[12] and in the presence of an external field.[13] For these reasons, the highest

S1089-5639(98)00327-2 CCC: $15.00
Published on Web 07/01/1998

SCHEME 1



level ab initio calculations compatible with the size of the complexes under scrutiny were used, and the counterpoise correction to the basis set superposition error (BSSE) was introduced for test cases in both the HF and MP2 correlated wave functions.

Methodology

The Protein Data Bank (PDB) was searched to determine all the X-ray structures solved with a resolution of 1.7 Å or better, thus identifying a large number of proteins, which were grouped into families considering only one structure per family, where possible, to reduce their number (Scheme 1).

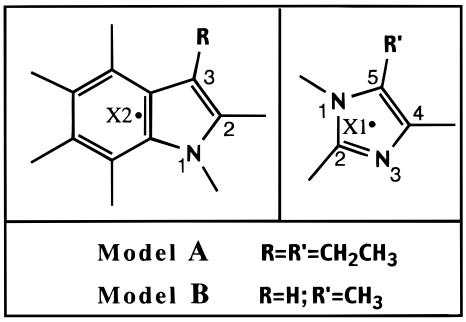
Several different structures were obtained, among which only those containing in their sequences Trp and His were kept. Then the final choice consisted of the structures presenting Trp and His close enough to interact, whose names are reported in Scheme 1 in boldface and brackets. 1rie and 1frb, however, resulted from a sample of about 40 additional structures obtained from an alternative search not limited to the "protein" class. To find more orientations in the region of the T-shaped adducts, we resorted to the structures solved with worse resolution (up to 2.8 Å), but presenting Trp and His close in their sequences, thus finding four more systems, reported at the bottom of Table 1. Nevertheless, we do not claim to have detected and examined all the possible arrangements of the Trp/His adducts contained in the PDB files. Our aim in those searches was to find a satisfactory variety of orientations among which to choose a few typical arrangements apt to be investigated in the absence

TABLE 1: Angles between the Ring Planes of the Trp and His Residues, Separations of Their Centroids, and HF and MP2 Interaction Energies at the 6-31G*(0.25) Level in the Chosen Structures

	structure	resolution (Å)	θ^a (deg)	centroid distance (Å)	ΔE^{HFb} (kcal/mol)	ΔE^{MP2c} (kcal/mol)
1	8abp	1.49	160.91	4.470	3.150	−6.486
2	3ptb	1.70	55.53	4.635	−0.069	−4.238
3	1rie	1.50	153.76	4.355	1.601	−6.237
4	1frbA	1.70	168.13	5.665	−1.977	−5.697
5	1frbB	1.70	126.99	6.346	−1.302	−2.765
6	1nfp	1.60	131.44	6.199	0.623	−0.821
7	1esaA	1.65	11.87	4.435	2.154	−5.365
8	1esaB	1.65	39.82	4.510	−3.732	−8.806
9	1s01	1.70	3.13	3.935	2.274	−8.304
10	1spbA	1.70	129.56	7.066	−0.037	−0.625
11	1spbB	1.70	115.12	4.595	−1.816	−6.099
12	1thl	1.70	31.30	6.920	−0.593	−2.926
13	1aoz	1.90	63.77	6.562	−1.371	−3.772
14	1bpl	2.20	31.07	8.926	−0.029	−0.072
15	1lla	2.20	81.31	5.464	−0.049	−3.751
16	1snv	2.80	133.45	4.967	−0.151	−5.142

[a] Larger than 90° when the imidazole ring is located below the indole plane. [b] $E_\infty^{HF} = -625.142\,360$ hartrees at the HF/6-31G*(0.25) level on the HF/6-31G* optimized internal geometries (HF/6-31G*(0.25)//HF/6-31G*); the HF/6-31G* optimized energies of the isolated compounds are −263.855 344 and −361.467 696 hartrees for 5-methylimidazole and indole, respectively. [c] $E_\infty^{MP2} = -626.834\,928$ hartrees at the MP2/6-31G*(0.25)//HF/6-31G* level.

SCHEME 2



of external fields. We realize of course that these orientations are determined not only by the interacting partners but also by the presence of the surrounding residues. We are fairly confident, however, that such close contacts (4.8 Å or less) should presumably be favorable, as indicated by the MP2/6-31G*(0.25) results in Table 1.

The Trp/His adducts were then extracted from the files and examined after adding the hydrogens to the side chains and optimizing their positions, while keeping fixed all the other atoms, with molecular mechanics (MM), using the SYBYL force field[14] and the Gasteiger−Hückel charges,[15] which produced δ-protonated imidazole rings (H on N_1) for all the adducts. Then two models were considered at first: (A) 3-ethylindole···5-ethylimidazole and (B) indole···5-methylimidazole, reported in Scheme 2. Since the trend of the results obtained was fairly similar in both cases, the ab initio calculations were carried out using the smaller model (i.e., B) throughout, though a plain imidazole could be suited as well. The 5-substitution, however, was requested to give at least a sense of the peptide backbone location.

The calculations on a few typical orientations of the adducts (with the internal geometry of the partners optimized at the HF/6-31G* level[16]) were carried out at the SCF and MP2[17] levels employing the 6-31G* basis set with the *d* exponents reduced

to 0.25[18] (6-31G*(0.25)[6]), because it is advisable for stacking interactions to use basis sets containing diffuse polarization functions and take into account also electron correlation effects.[6] These effects (very important because they account for the dispersion attraction) can be computed at the MP2 level, which is also named MP2 correction. In order to elucidate the quality of the description and the nature of the interaction, and to compare the trend of the interaction energies along the approaching path with standard and probably more accurate results, additional descriptions (namely, 6-31G* and Dunning's double-ζ plus polarization, DZP[19]) of a subset of the adducts were considered at the SCF and MP2 levels with and without counterpoise (CP) corrections[20] to the basis set superposition error (BSSE).

The BSSE is the artifact introduced by the use of a limited basis set; passing from the individual molecules to the adduct, the electrons obtain a beneficial effect from the availability of the virtual space of the partner, thus producing in general a deeper potential hole at shorter separations than in the CP-corrected calculations. The CP correction consists of a change in the reference energy that is now obtained from calculations of the individual partner energies using the adduct geometries and basis functions in the absence of the other's nuclei and electrons. Wide agreement has been reached about the convenience of using both the occupied and vacant functional space of the partners[12a,21] instead of just the virtual space as previously suggested.[22−24]

To clarify the notations used, we report hereafter a few definitions. The total interaction energy along the approaching path (R) may be defined as

$$\Delta E(R) = \Delta E^{HF}(R) + \mathrm{BSSE}(R) + \Delta E^{COR}(R) \qquad (1)$$

where $\Delta E^{HF}(R)$ is the interaction energy at the HF level, BSSE(R) is the basis set superposition error, and $\Delta E^{COR}(R)$ is the correlation effect.

The CP-corrected interaction energies at the HF and MP2 levels are given by

$$\Delta E^{CPHF}(R) = \Delta E^{HF}(R) + \Delta^{CPHF}(R) \qquad (2)$$

$$\Delta E^{CPMP}(R) = \Delta E^{MP2}(R) + \Delta^{CPMP}(R) \qquad (3)$$

with $\Delta E^{MP2}(R) = \Delta E^{HF}(R) + \Delta^{MP}(R)$, the MP2 approximation to $\Delta E^{COR}(R)$, while $\Delta^{CPHF}(R)$ and $\Delta^{CPMP}(R)$ are the HF and MP2 CP corrections, respectively, that turn out to be considerably different from each other. All these values are however different approximations to the true interaction energy of eq 1. It is not at all demonstrated thus far that the CP corrections or the MP2 level account for the whole BSSE or ΔE^{COR}, respectively. On the contrary, the CP correction has often been charged of overcorrecting the error, while MP2 probably underestimates the true electron correlation effect. As far as the CP correction is concerned, we showed that the overcorrection arises from the use of a minimal basis set, at least in H-bonded dimers.[12]

Since all the aforementioned terms depend on the partner separation, the calculations were performed at a few values of R, as already stated. It is improper in fact to evaluate the CP or MP2 corrections only at the HF or MP2 equilibrium distances which can be greatly affected by the corrections.[12a] For this reason, it is advisable to analyze the trend of the corrected and uncorrected interaction energies along the ring centroid separation. For the sake of comparison it is useful to consider separately the CP correction and the correlation effect.

For all the aforementioned calculations we made use of the Gaussian92−94 systems of programs[25] running on the IBM/RS6000-590 and SGI Indigo2 workstations at ICQEM. The interaction energy decomposition was performed with GAMESS,[26] in Morokuma's framework (KM)[27] using Deisz's scaled MINI-1 basis set[28] that for these systems compares fairly well with DZP. Additional energy decompositions were carried out using the 4-31G basis set[29] (the largest basis set consistent with our filesystem size) with the Pisa version of Monstergauss,[30] MGPIPC,[31] which includes the CP corrections for each monomer to the energy decomposition (details about the method used can be found in the source paper[12b]). Geometries were visualized with SYBYL or MidasPlus[32] on the IRIS/4D-420-GTXB workstation at ICQEM. Additional details are given in the proper place in the next section.

Results and Discussion

The adduct arrangements considered, named after the PDB file containing them, are reported in Table 1 together with the structure resolution, the indication of the angle between the ring planes, θ, and the ring centroid separation as measured from the distance d(X1−X2) of the imidazole and indole ring centers, also indicated in Scheme 2.

HF and MP2 Results. The interaction energy at the HF and MP2/6-31G*(0.25) levels for the adducts in the experimental arrangement (with the internal geometries of the partners optimized at the HF/6-31G* level, but at infinite separation) is reported to give a preliminary indication of their relative stability. The stereo pictures of the 5-methylimidazole structures with respect to indole (whose ring was exactly superimposed for all the adducts) are displayed in Figures 1−3. The calculations at the 6-31G*(0.25)/HF and MP2 levels were carried out on the adducts of Figures 1 and 2 for various values of the intermolecular distance. The use of a polarization function exponent more diffused (0.25) than the usual one was first proposed[18] in connection with the DZP basis set in order to use a single set of polarization functions instead of two, as in DZPP. Subsequently, this suggestion was widely exploited using the 6-31G*(0.25) basis set[6] mainly in the study of stacking interactions, to obtain more realistic values of the correlation interaction energy, ΔE^{COR}, which is considered to be of paramount importance in this kind of complexes. The structures reported in Figure 1 correspond to a T-shaped arrangement with some of the indole hydrogens pointing toward the imidazole ring (1lla) and to two stacked structures, one almost parallel (1s01) and one antiparallel but slightly distorted (θ = 39.8, 1esaB).

The structures reported in Figure 2 belong to two T-shaped (1spbB and 1aoz) and two parallel (1frbA and 1esaA, though slightly displaced with respect to a perfectly stacked position) arrangements, one differing from the other in the imidazole position—either below ($\theta > 90°$) or above ($\theta < 90°$) the indole plane—and in the ring centroid separation, while the angles between the ring planes are fairly similar. Both T-shaped adducts belong to a different class with respect to 1lla, because they show an imidazole H pointing toward the indole ring π density.

Their interaction energies were then evaluated at various separations, while keeping fixed mutual orientations and internal geometries of the partners at the HF and MP2 (in the frozen-core approximation) levels, because the incidence of correlation effects may heavily vary with the distance between the partners. The trend of the HF and MP2 interaction energies at the 6-31G*-(0.25) level along the ring centroid separation is shown in the mid part of Figures 4−6 for 1esaB, 1s01, and 1lla, respectively,

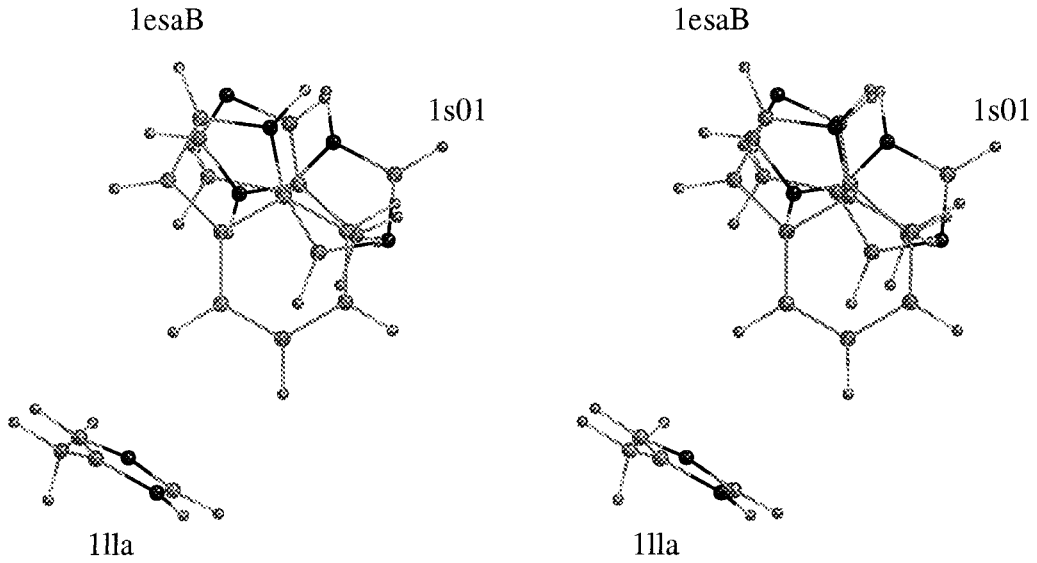


Figure 1. Stereo picture of three adducts with the indole ring superimposed, corresponding to the T-shaped arrangement, 1lla (with some of the indole hydrogens pointing toward the imidazole ring) and to two stacked structures, 1s01 (almost parallel) and 1esaB, skew antiparallel ($\theta = 39.8$, see Table 1).

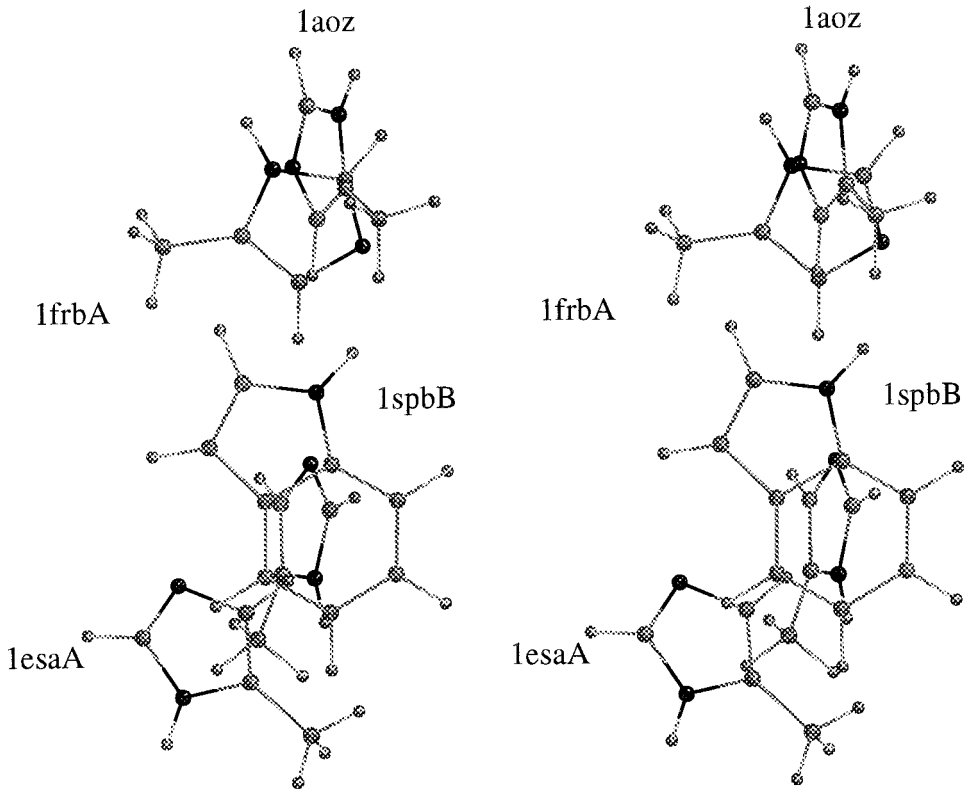


Figure 2. Stereo picture of four adducts with the indole ring superimposed, corresponding to two T-shaped arrangements, 1spbB and 1aoz (below and above the indole plane, see Table 1) both with an imidazole H pointing toward the indole ring, and to two parallel structures, 1frbA and 1esaA, displaced with respect to a stacked position.

while that of the structures in Figure 2, namely 1esaA, 1frbA, 1aoz, and 1spbB, is displayed in Figure 7. The corresponding equilibrium values are reported in Tables 2 and 3. The skew antiparallel structure (1esaB) is the most favorable in the set, with an interaction energy of about 4 and 13 kcal/mol at the HF and MP2 levels, respectively, in agreement with what was already found for the experimental separation. For all these adducts, however, the dispersion attraction plays a considerably important role in stabilizing the interaction energies and in shortening the equilibrium distances (up to 9 kcal/mol and 0.7 Å, respectively), as can be seen examining Tables 2 and 3.

The three T-shaped structures are attractive at the HF level (1lla only slightly, $\Delta E^{HF} = -0.4$ kcal/mol at 5.9 Å, not reported in Table 2), as the displaced stacked structure 1frbA, while 1esaA (the other shifted parallel structure) and 1s01 are repulsive. However, the dispersion estimated at the MP2 level stabilizes these repulsive interactions, making them favorable by 6 and 9 kcal/mol, respectively. The electron correlation effect is somewhat less effective for the T-shaped structures, attractive at the HF level, and for 1frbA, which gain only 3−5 kcal/mol.

CP Correction to BSSE. To evaluate the quality of the description and the reliability of the results obtained examining

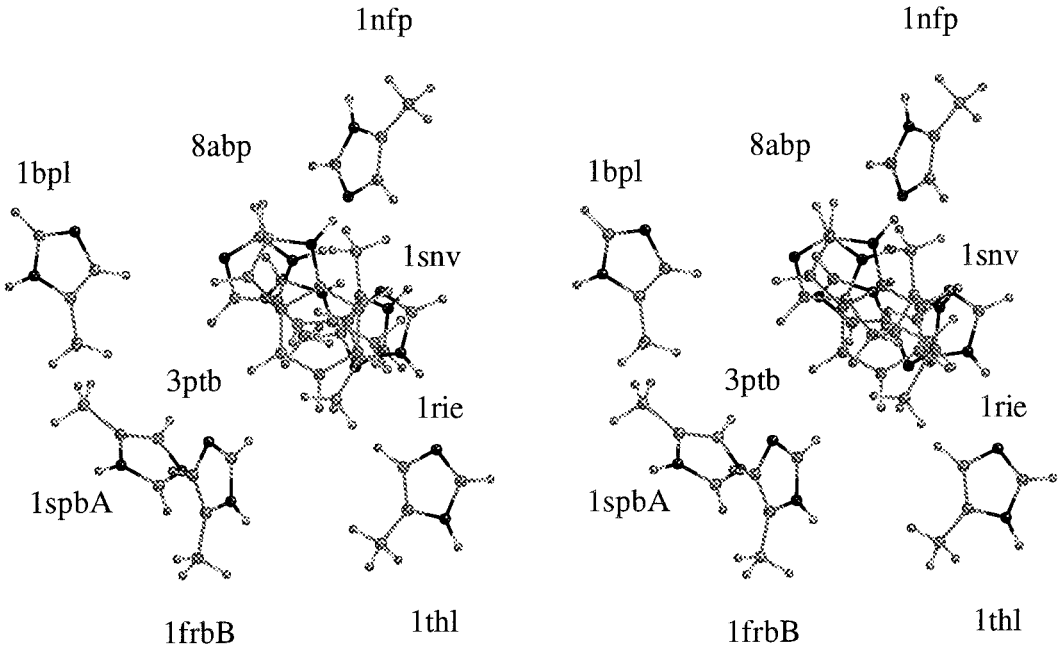


Figure 3. Stereo picture of nine adducts with the indole ring superimposed, corresponding to various arrangements. The angle between the ring planes and the centroid separation are reported in Table 1.

the basis set effect and related sources of errors, such as the BSSE, the calculations were repeated for the three adducts of Figure 1, with the 6-31G* (found to have a small BSSE in H-bonding interactions[12]) and the DZP basis sets. The trends of the interaction energies at the various levels along the centroid separation are displayed in the upper (6-31G*) and lower (DZP) parts of Figures 4−6. The counterpoise correction to the BSSE was also introduced for all the basis sets considered both at the HF and MP2 levels, to compare the relevant results, because the applicability of the CP method at the correlated level is still under debate.[5,21]

What is apparent at first sight is the analogous behavior of the three basis sets for the different adducts: the 6-31G* basis set (upper parts of the figures) produces corrected interaction energies less favorable than DZP (lower parts), which in turn gives interaction energies less favorable than 6-31G*(0.25), whereas the HF values are generally less sensitive to the basis set effect than the corrections (CP or MP2). Moreover, when CP corrected, the HF interaction energies turn out to be practically independent of the basis set, even when they are repulsive. Another feature observed almost everywhere is the extent of the CP correction to the HF interaction energy, Δ^{CPHF}, that is very low for DZP and somewhat higher for both the 6-31G* basis sets, with slightly higher values for the 6-31G*-(0.25) one. The CP correction to the MP2 interaction energy, Δ^{CPMP}, lower for 6-31G* than for DZP followed by 6-31G*-(0.25), is always much larger than Δ^{CPHF} and roughly proportional to ΔE^{MP2}. The ΔE^{CPMP} curve is thus located almost halfway between ΔE^{HF} and ΔE^{MP2}, suggesting the viability to limit further analyses of this kind of adducts to the trends of the HF and MP2 interaction energies avoiding the computational burden represented by the CP correction. The structures which are repulsive (or only feebly attractive) at the HF level are even more repulsive, of course, when CP corrected, whereas the correlation effects computed at the MP2 level stabilize these interactions that remain still favorable after CP correction, though only slightly.

Equilibrium Distances. In Table 2 the equilibrium distances at the various levels and the corresponding interaction energies

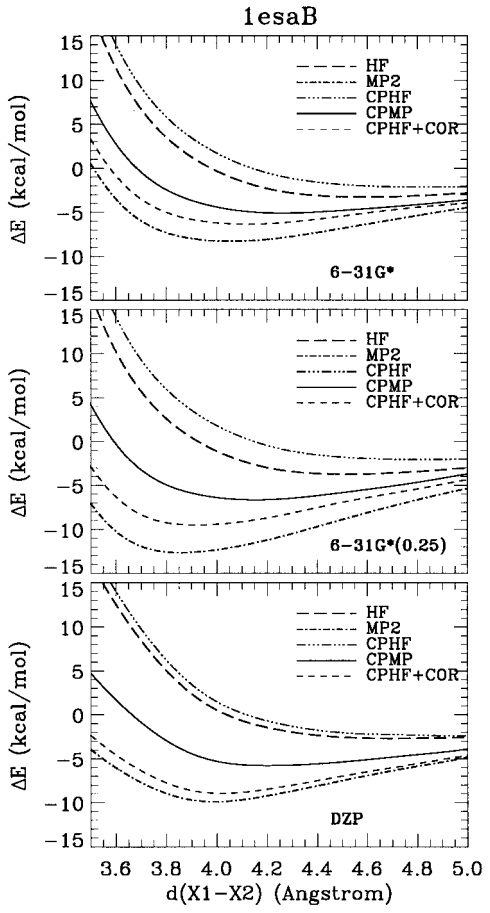


Figure 4. Interaction energy along the approaching path for 1esaB, at the HF (long dash) and MP2 (dot-dash) levels, with the inclusion of counterpoise corrections at both levels (HF (three dots-dash; CPHF) and MP2 (solid line; CPMP)); the curve corresponding to the inclusion of correlation effects at the MP2 level to the CPHF interaction energy is also displayed (short dash; CPHF + COR), as described by the three basis sets used: 6-31G* (upper part), 6-31G*(0.25) (mid part), and DZP (lower part).

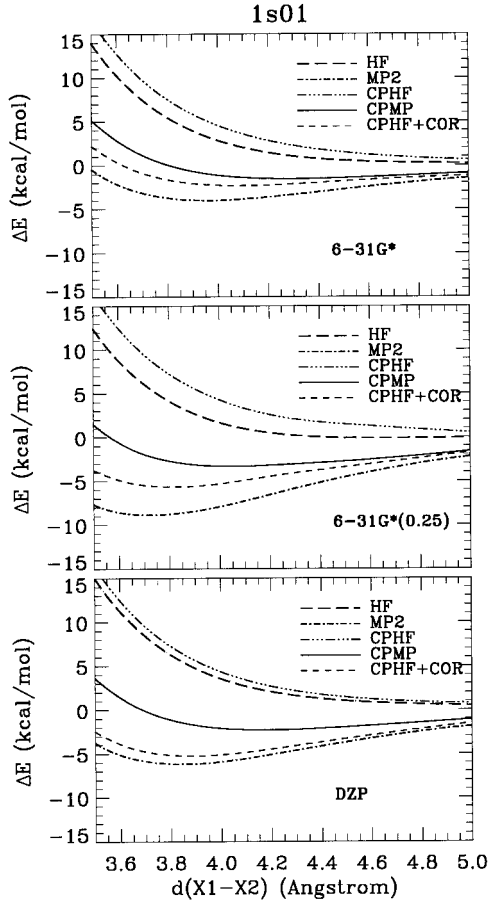


Figure 5. Interaction energy along the approaching path for 1s01, at the HF (long dash) and MP2 (dot-dash) levels, with the inclusion of counterpoise corrections at both levels (CPHF (three dots-dash) and CPMP (solid line)); the curve corresponding to the inclusion of correlation effects at the MP2 level to the CPHF interaction energy is also displayed (short dash; CPHF + COR), as described by the three basis sets used: 6-31G* (upper part), 6-31G*(0.25) (mid part), and DZP (lower part).

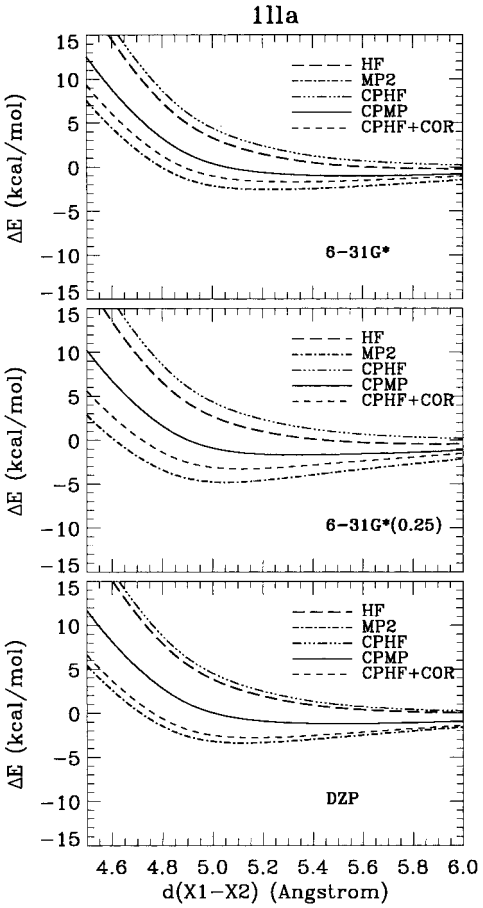


Figure 6. Interaction energy along the approaching path for 1lla, at the HF (long dash) and MP2 (dot-dash) levels, with the inclusion of counterpoise corrections at both levels (CPHF (three dots-dash) and CPMP (solid line)); the curve corresponding to the inclusion of correlation effects at the MP2 level to the CPHF interaction energy is also displayed (short dash; CPHF + COR), as described by the three basis sets used: 6-31G* (upper part), 6-31G*(0.25) (middle part), and DZP (lower part).

are reported to allow a comparison of values (not just of trends), even though this could be somewhat misleading because the curves are very smooth. The 6-31G*(0.25) description favors slightly shorter equilibrium distances than the other two basis sets, at least for the stacked adducts, while for the T-shaped adduct (1lla) the three basis sets show analogous equilibrium separations at the MP2 level. The HF and CPHF entries are not displayed in the bottom part of the table, even though the HF interaction energy turns out to be feebly attractive for 1lla at large separations with the 6-31G* and 6-31G*(0.25) basis sets (by 0.2 and 0.4 kcal/mol at 6.0 and 5.9 Å, respectively).

By comparing the equilibrium separations (Tables 2 and 3) with the experimental ones (reported in Table 1), it can be seen that for 1s01 the ring centroid distance derived from the crystal structure is intermediate between those obtained at the MP2 and CPMP levels using both the DZP and 6-31G*(0.25) basis sets, and slightly lower than those computed at the 6-31G* level. For 1esaB, 1spbB, 1aoz, and 1frbA on the contrary the separation is fairly similar to the values obtained for all the basis sets at the HF level (when CP uncorrected). For 1lla the experimental value is closer to the CPMP ones than to the results including only electron correlation. Interestingly enough, the MP2 equilibrium distances show a fair linear correlation with respect to the experimental separations, plotted in Figure 8a, while the MP2 interaction energies at the equilibrium distance are only slightly worse correlated with the MP2 interaction

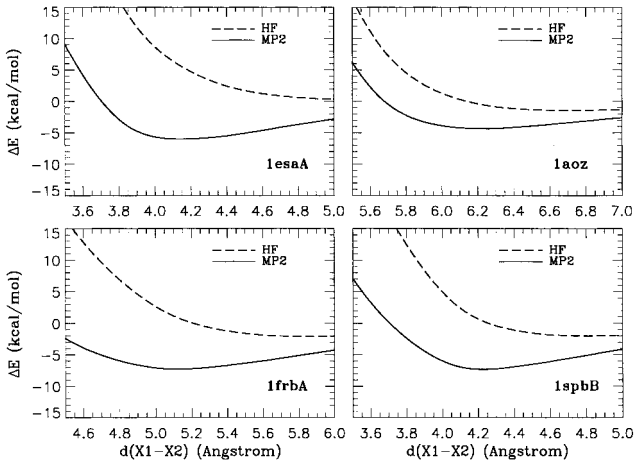


Figure 7. Interaction energy along the approaching path at the HF (long dash) and MP2 (solid line) levels, as described by the 6-31G*-(0.25) basis set, for 1esaA, 1aoz, 1frbA, and 1spbB.

energies at the experimental distance (Figure 8b). There is an overall fair agreement between the computed and experimental results if we consider that the calculations were carried out in the gas phase, in the absence of the field produced at least by the nearby residues, and that in the imidazole model the H bound to the N atom was placed on N_1 throughout.

TABLE 2: Basis Set Effect on Equilibrium Distances (in Å) and Corresponding Interaction Energies (in kcal/mol) along the Approaching Paths of the Adducts, Contained in 1esaB, 1s01, and 1lla, Kept in Their Crystallographic Mutual Orientation, at the HF and MP2 Levels with and without Counterpoise Corrections

1esaB	R_{eq}^{HF}	ΔE_{eq}^{HF}	R_{eq}^{MP2}	ΔE_{eq}^{MP2}	R_{eq}^{CPHF}	ΔE_{eq}^{CPHF}	R_{eq}^{CPMP}	ΔE_{eq}^{CPMP}
6-31G*	4.6	−3.2	4.1	−8.2	4.9	−2.14	4.3	−5.1
6-31G*(0.25)	4.5	−3.7	3.8	−12.7	4.7	−2.06	4.1	−6.6
DZP	4.6	−2.7	4.0	−9.9	4.8	−2.31	4.2	−5.8

1s01	R_{eq}^{MP2}	ΔE_{eq}^{MP2}	R_{eq}^{CPMP}	ΔE_{eq}^{CPMP}	1lla	R_{eq}^{MP2}	ΔE_{eq}^{MP2}	R_{eq}^{CPMP}	ΔE_{eq}^{CPMP}
6-31G*	4.0	−4.0	4.2	−1.5	6-31G*	5.2	−2.6	5.6	−1.0
6-31G*(0.25)	3.7	−8.9	4.1	−3.3	6-31G*(0.25)	5.1	−4.7	5.3	−1.7
DZP	3.8	−6.2	4.1	−2.3	DZP	5.1	−3.4	5.5	−1.2

TABLE 3: 6-31G*(0.25) Equilibrium Distances (Å) and Corresponding Interaction Energies (in kcal/mol) along the Approaching Paths of the Adducts, Kept in Their Crystallographic Mutual Orientation, at the HF and MP2 Levels

structures	R_{eq}^{HF}	ΔE_{eq}^{HF}	R_{eq}^{MP2}	ΔE_{eq}^{MP2}
1spbB	4.8	−2.0	4.2	−7.3
1aoz	6.7	−1.4	6.2	−4.3
1frbA	5.8	−2.1	5.1	−7.3
1esaA			4.2	−6.0

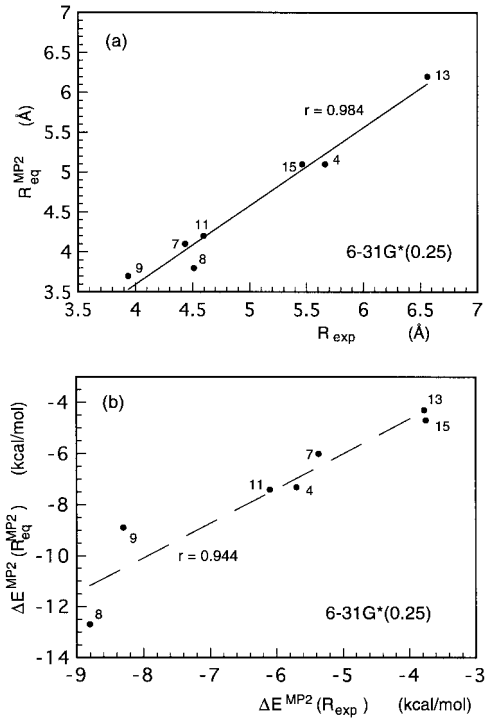


Figure 8. Correlation between (a) the MP2/6-31G*(0.25) equilibrium distances R_{eq}^{MP2} and the experimental separations (R_{exp}), and (b) the MP2/6-31G*(0.25) interaction energies at R_{eq}^{MP2} and the MP2 interaction energies computed at the same level at R_{exp} for the seven adducts in Figures 1 and 2. The adduct numbering corresponds to that reported in Table 1.

As a general rule at the HF and MP2 levels, the 6-31G*(0.25) basis set produces the most favorable interaction energies and equilibrium distances intermediate in general between 6-31G* and DZP. This should ensure that the interaction energy computed at the MP2 level is a lower bound to any other value obtainable with an alternative choice. Nonetheless, the inclusion of CP corrections is likely to almost halve the stabilization. The most favorable arrangements turn out to be stacked (1esaB > 1s01), followed by a couple of structures (the most stable T-shaped structure, 1spbB, and the displaced stacked adduct, 1frbA) that are almost as stable. Finally, there are the other shifted parallel adduct (1esaA) and two differently T-shaped adducts (1lla, 1aoz).

Energy Decomposition Analysis. In order to analyze as far as possible the nature of the interaction in the three adducts examined in detail, the Morokuma decomposition analysis was carried out employing the MINI-1 basis set, the only basis set affordable for these systems due to disk space limitations, using GAMESS which does not implement direct methods for this kind of calculations. The MINI-1 basis set was found to present a very limited BSSE and to offer results comparable to the extended basis sets of better quality (6-31G**),[12a−e] even though for H-bonded systems.

The HF interaction energy decomposition along the approaching path is reported in Figure 9. Let us focus first of all on the total interaction energy (dotted line), which is favorable only for one of the systems considered (1esaB). Despite the use of a minimal basis set, the trend and the value of ΔE are fairly similar to those obtained thus far with the extended basis sets: as already noticed the HF interaction energy for these adducts is only little dependent on the basis set. It is worth stressing that the behavior of minimal basis sets, such as MINI-1, obtained via a direct optimization, is much better than that of Gaussian minimal basis sets obtained by fitting to Slater orbitals.[33,34]

Because of the concern raised by one of the reviewers by the use of the MINI-1 basis set, additional calculations were carried out with MGPIPC, after providing our machine with a new disk, at the 4-31G level (the largest basis set compatible with her filesystem size). The relevant results we are going to discuss in parallel to the MINI-1 ones are reported in Figure 10. For all the systems considered, the MP2 correlation contribution to the interaction energy (thin solid line with cross markers) is also reported, in order to give at least a sense of the incidence of the dispersion attraction, together with the interaction energy at the MP2 level (solid line). An accurate calculation of the dispersion term would require computation of the intramolecular correlation contribution of each partner in the presence of the other using the localized MO description of the adduct;[37] we have on the contrary evaluated the intramolecular correlation contribution at infinite separation of the partners. Alternatively, the intermolecular Møller−Plesset perturbation theory might be used, which is a double perturbation formalism with two perturbations representing, respectively, the intermolecular interaction and the intramonomer correlation. Interested readers are referred to a recent review and to the references quoted therein.[38]

The interpretation of binding in terms of the ΔE components can be decidedly affected by the basis set. Therefore, it is almost impossible to find strong polarization effects without using polarization functions. The 4-31G basis set, which gives a trend for E_{PL} better (but still hardly satisfactory) than MINI-1 due to its split valence shell, is more heavily affected by BSSE than

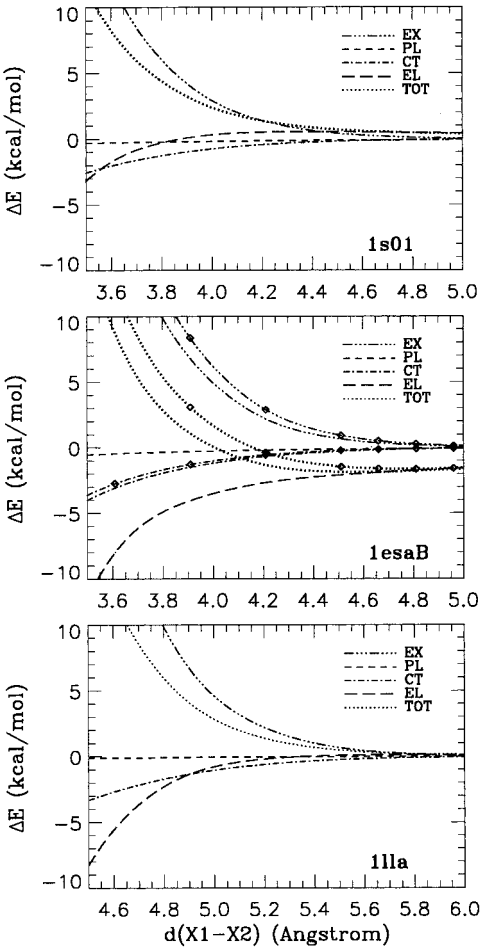


Figure 9. HF interaction energy decomposition in the KM scheme along the approaching path, as described by the MINI-1 basis set, for 1s01 (upper part), 1esaB (middle part), and 1lla (lower part). TOT (dotted line) corresponding to the total interaction energy, EL (long dash) to the electrostatic contribution, CT (dot-dash) to the charge transfer term, PL (short dash) to the polarization term, and EX (three dots-dash) to the exchange term. For 1esaB also the CP-corrected (marked with rhombs) TOT, EX, and CT contributions are reported.

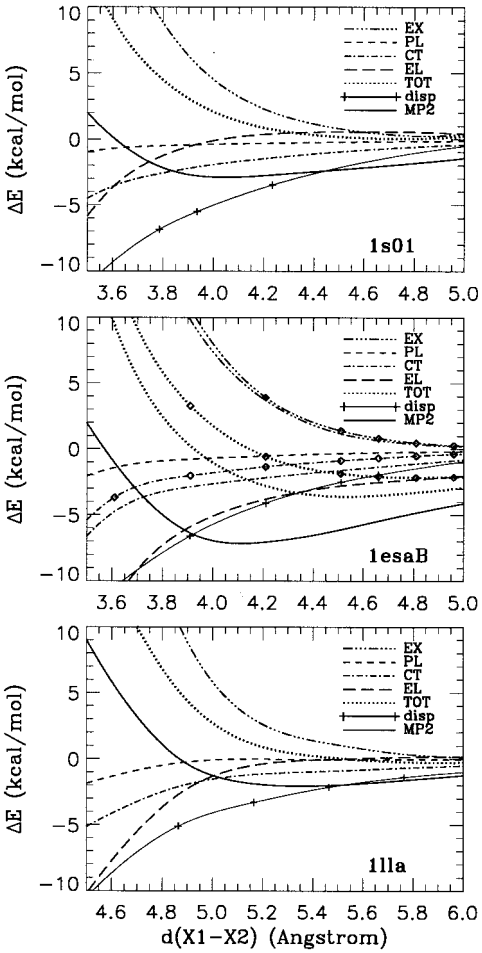


Figure 10. HF interaction energy decomposition in the KM scheme and dispersion contribution (see text, thin solid line marked with crosses) along the approaching path, as described by the 4-31G basis set, for 1s01 (upper part), 1esaB (middle part), and 1lla (lower part). TOT (dotted line) corresponding to the total HF interaction energy, EL (long dash) to the electrostatic contribution, CT (dot-dash) to the charge transfer term, PL (short dash) to the polarization term and EX (three dots-dash) to the exchange contribution. MP2 is the interaction energy at the MP2 level (solid line). For 1esaB also the CP-corrected (marked with rhombs) TOT, EX, and CT contributions are reported.

the MINI-1 one, as we observed in the past[12e] and as can be seen looking at the mid part (1esaB) of Figures 9 and 10, where the CP-corrected contributions[12b] are marked with rhombs.[39] The MINI-1 basis set shows a small BSSE, which decreases as the separation increases, especially for the exchange contribution, whereas the 4-31G basis presents an almost constant BSSE effective also at large separations, mainly related to the CT term. The electrostatic contribution is responsible for the favorable interaction already at the HF level in 1esaB, while for both systems (1s01 and 1lla) which are unfavorable at the HF level the EL term is smaller than the charge transfer contribution. For sufficiently large separations of the partners (∼4.5 Å for the stacked adducts) the attractive polarization and charge transfer terms almost counterbalance the exchange repulsion, thus making the electrostatic contribution coincide with the total interaction energy. At the 4-31G level this is true only for the CP-corrected values of ΔE.

The 4-31G dispersion term,[40] displayed in Figure 10 (thin solid line marked with crosses), is the largest attractive contribution for 1s01 and 1lla, whereas it is analogous in magnitude to the largest HF favorable contribution, EL, for 1esaB. The total interaction energy including correlation effects, MP2, also presented in Figure 10, shows a trend very similar to the 6-31G* one displayed in Figures 4−6, thus suggesting a possible use of the CP corrected 4-31G basis set at the MP2 level to study this kind of systems.

Conclusions

Several arrangements of the histidine and tryptophan residues within interacting distance have been extracted from the crystal structures contained in the Brookhaven Protein Data Bank and a few typical stacked, whether displaced or not, and T-shaped structures involving 5-methylimidazole and indole as models of their heteroaromatic side chains have been studied, making use of the 6-31G*(0.25) basis set. The HF interaction energy for these adducts is almost independent of the basis set, allowing to carry out the interaction energy decomposition at the MINI-1 and 4-31G levels, but the latter should be CP corrected. The interpretation of binding as derived from the decomposition analysis puts forward the importance of the electrostatic contribution to obtain favorable adducts already at the HF level. However, the dispersion attraction plays an important role also in adducts where the charge transfer term is larger than the electrostatic one.

The picture of the interaction does not heavily depend on the basis set, provided counterpoise corrections and electron

correlation effects are taken into account as shown by comparing results employing also the 4-31G, 6-31G*, and DZP basis sets. In summary, these basis sets may be used with confidence, even though the inclusion of polarization functions more diffused than the usual ones in our opinion ameliorates the description of the attraction, which, however, could be still underestimated.

The lowest incidence of BSSE is found for the standard 6-31G* basis set followed by DZP at the MP2 level, while the reverse holds at the HF level. The 6-31G*(0.25) results turn out to be affected by a BSSE larger than expected. The CP-uncorrected values obtained using the 6-31G*(0.25) basis set therefore are a lower bound to the true interaction energy when the correlation effects at the MP2 level are introduced. The CP-corrected interaction energy is in fact likely to be located halfway between uncorrected MP2 and HF values, and thus this basis set should not be used without taking into account CP corrections.

The inclusion of correlation effects allows to obtain sensible estimates for the interaction energy also at the experimental separation and even using standard basis sets, thus avoiding the computationally expensive procedures to locate the equilibrium distances and to evaluate the counterpoise corrections, when the experimental arrangements are available. In this way, it is possible to obtain rather reasonable results, at least as the first estimate, in all cases when the calculation of BSSE corrections is unaffordable or unavailable.

The importance of correlation effects not only for the stacked structures but also for the T-shaped ones emerges enhanced, if possible, by this study.

Acknowledgment. S.M. is grateful to Menarini Ricerche for a fellowship allowing her to carry out research activity at ICQEM.

References and Notes

(1) Alagona, G.; Ghio, C.; Monti, S. *J. Mol. Graphics Mod.* submitted.

(2) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. *J. Mol. Biol.* **1990**, *213*, 899.

(3) Huang, R.-R. C.; Vicario, P. P.; Strader, C. D.; Fong, T. M. *Biochemistry* **1995**, *34*, 10048.

(4) Alagona, G.; Ghio, C.; Monti, S. *J. Mol. Struct. (THEOCHEM)* **1998**, in press.

(5) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

(6) (a) Hobza, P.; Mehlhorn, A.; Čársky, P.; Zahradník, R. *J. Mol. Struct. (THEOCHEM)* **1986**, *138*, 387. (b) Šponer, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1994**, *12*, 671. (c) Hobza, P.; Šponer, J.; Polášek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792. (d) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965. (e) Šponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695. (f) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263.

(7) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.

(8) Smithrud, D. B.; Diederich, F. *J. Am. Chem. Soc.* **1990**, *112*, 339.

(9) Declercq, D.; Delbeke, P.; De Schryver, F. C.; Van Meervelt, L.; Miller, R. D. *J. Am. Chem. Soc.* **1993**, *115*, 5702.

(10) Alagona, G.; Ghio, C.; Nagy, P. I.; Durant, G. J. *J. Phys. Chem.* **1994**, *98*, 5422.

(11) (a) Cozzi, F.; Cinquini, M.; Annunziata, R.; Dwyer, T.; Siegel, J. S. *J. Am. Chem. Soc.* **1992**, *114*, 5729. (b) Cozzi, F.; Cinquini, M.; Annunziata, R.; Siegel, J. S. *J. Am. Chem. Soc.* **1993**, *115*, 5330.

(12) (a) Alagona, G.; Ghio, C.; Cammi, R.; Tomasi, J. *Int. J. Quantum Chem.* **1987**, *32*, 207. (b) Alagona, G.; Ghio, C.; Cammi, R.; Tomasi, J. *Int. J. Quantum Chem.* **1987**, *32*, 227. (c) Alagona, G.; Ghio, C.; Cammi, R.; Tomasi, J. In *Molecules in Physics, Chemistry, and Biology*, Maruani, J., Ed.; Kluwer: Dordrecht, 1988; Vol. II, pp 507−559. (d) Alagona, G.; Cammi, R.; Ghio, C.; Tomasi, J. *Int. J. Quantum Chem.* **1989**, *35*, 223. (e) Alagona, G.; Ghio, C.; Tomasi, J. *J. Phys. Chem.* **1989**, *93*, 5401. (f) Alagona, G.; Ghio, C.; Latajka, Z.; Tomasi, J. *J. Phys. Chem.* **1990**, *94*, 2267. (g) Alagona, G.; Ghio, C. *J. Comput. Chem.* **1990**, *11*, 930. (h) Alagona, G.; Biagi, A.; Ghio, C. *Mol. Eng.* **1992**, *2*, 137.

(13) Alagona, G.; Cammi, R.; Ghio, C.; Tomasi, J. *Theor. Chim. Acta* **1993**, *85*, 167.

(14) SYBYL, Tripos Associates, St. Louis, MO 63144-2913.

(15) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219; Gasteiger, J.; Saller, H. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687. Hückel, E. *Z. Phys.* **1932**, *76*, 628.

(16) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257; Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(17) (a) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618. (b) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem.* **1976**, *10*s, 1. (c) Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 4244.

(18) Van Duijneveldt-Van De Rijdt, J. G. C. M.; Van Duijneveldt, F. B. *J. Mol. Struct. (THEOCHEM)* **1982**, *89*, 185.

(19) Dunning, T. H.; Hay, P. J. *Modern Theoretical Chemistry*; Plenum: New York, 1976; Chapter 1, pp 1−28.

(20) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(21) Gutowski, M.; Van Lenthe, J. H.; Verbeek, J.; Van Duijneveldt, F. B.; Chałasiński, G. *Chem. Phys. Lett.* **1986**, *124*, 370.

(22) Daudey, J. P.; Malrieu, J. P.; Rojas, O. *Int. J. Quantum Chem.* **1974**, *8*, 17.

(23) Morokuma, K.; Kitaura, K. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981; p 215.

(24) Magnasco, V.; Musso, G. F.; Costa, C.; Figari, G. *Mol. Phys.* **1985**, *56*, 1249.

(25) (a) *Gaussian 92/DFT, Revision G.2*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Wong, M. W.; Foresman, J. B.; Robb, M. A.; Head-Gordon, M.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1993. (b) *Gaussian 94, Revision D.4*; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1995.

(26) *GAMESS*; Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(27) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325.

(28) (a) Andzelm, J.; Huzinaga, S.; Klobukowski, M.; Radzio-Andzelm, E.; Sakai, Y.; Tatewaki, H. *Gaussian Basis Sets for Molecular Calculations*; Huzinaga, S., Ed.; Elsevier: Amsterdam, 1984. (b) Tatewaki, H.; Huzinaga, S. *J. Comput. Chem.* **1980**, *1*, 205.

(29) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(30) *MONSTERGAUSS*; Peterson, M. R.; Poirier, R. A. Department of Chemistry, University of Toronto, Toronto, Ontario, Canada.

(31) MGPIPC, Bonaccorsi, R.; Cammi, R., Ghio, C. ICQEM, Via Risorgimento 35, 56126 Pisa, Italy.

(32) MidasPlus, Computer Graphics Lab., School of Pharmacy, University of California, San Francisco, CA 94143, USA.; Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13.

(33) Kołos, W. *Theor. Chim. Acta* **1979**, *51*, 219.

(34) A relevant consideration under this respect may be represented by the results we obtained using Slater type functions both minimal (BAZ) and double-ζ with polarization basis sets to compute interaction energies[35] for dimeric H-bonded systems (either neutral or anionic), previously evaluated using Gaussian functions of different quality.[12f] The BAZ wave function shows the same behavior as the STO-3G one whether CP corrected or not for the neutral systems, while for the anionic ones it compares well to MINI-1 when uncorrected, and gives interaction energies slightly less favorable than MINI-1 when CP corrected. Therefore the STO-3G wave function seems to derive its defects from its birth. It is, however, not an intrinsic fault of Slater basis sets, because on the contrary the double-ζ plus polarization basis set gives better results than the 6-31G** basis set when uncorrected and can hold the comparison with a basis set especially designed to be practically free from BSSE.[36]

(35) Alagona, G.; Ghio, C. *J. Mol. Structure (THEOCHEM)* **1995**, *330*, 77.

(36) Latajka, Z.; Scheiner, S. *J. Comput. Chem.* **1987**, *8*, 663.

(37) Vos, R. J.; Hendricks, R.; van Duijneveldt, F. B. *J. Comput. Chem.* **1990**, *11*, 1.

(38) Chałasiński, G.; Szczęśniak, M. *Chem. Rev.* **1994**, *94*, 1723.

(39) The EL and PL terms are unaffected by BSSE in the KM framework by definition; the MIX term is not displayed.[12b]

(40) It is not advisable to compute such a contribution with a minimal basis set.