Journal of Computer-Aided Molecular Design, 8 (1994) 653–668
ESCOM

J-CAMD 270

Extended electron distributions applied to the molecular mechanics of some intermolecular interactions

J.G. Vinter

Cambridge Centre for Molecular Recognition, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Received 16 May 1994
Accepted 12 July 1994

Key words: Molecular mechanics; Charge; Coulombic energy; Intermolecular interaction

SUMMARY

Extended electron distributions (XEDs) have been added to the molecular mechanics Coulombic term and applied to a selection of intermolecular interactions. The results from this approach have been compared with the commonly used atom-centred charges and more rigorous AM1-derived natural atom orbital point densities. The use of XEDs generally improves the simulation of experimental and ab initio results over the other two charge allocations and corrects geometries in those cases for which the others yield wrong results.

INTRODUCTION

One major drawback in the energy calculation phase of most methods based on molecular mechanics has been the use of atom-centred charges (ACCs) in the Coulombic term. Many empirical and quantum mechanical procedures exist for their determination, each one resulting in a different set of charges. Since there are as yet no experimental criteria on which to judge these values, except the molecular ground-state property of dipole moment and possibly high-quality quantum mechanics (QM), they are not easy to assess. The use of computer modelling tools has become standard practice for many disciplines and users rely on computer methods which are assumed to be reliable. This is not necessarily so if those methods use atom-centred charges.

Any method which collapses orbital charge distribution must be an approximation. How much of an approximation has been a source of extended debate [1,2], particularly since the introduction of Distributed Multipole Analysis (DMA) [3] which has highlighted the dangers of ignoring electron anisotropy. However, DMA and other QM procedures which retain electron distribution require the calculation of a full wave function. This restricts any study to a few conformations of relatively small molecules. The problems with ACCs led Hunter and Sanders [4] to propose

a simple model of π-electron distribution which considerably improved the π-stacking behaviour of porphyrins.

This paper describes the incorporation of Hunter and Sanders' seminal ideas [4] into an integrated molecular modelling framework [5,6] and its extension to include lone-pair electrons. Collectively, the new methodology implements extended electron distributions (XEDs). Its development has been a necessary stage in a larger project involving the use of electrostatic potential fields in molecular recognition. If XEDs are to be useful, they must necessarily be computationally fast, capable of handling large molecules and demonstrate a consistent improvement over other

TABLE 1
DATA FOR REAL AND XED ATOM TYPES

Atom type	Atom hybrid	Atom geometry	Connectivity
1	C sp^3	tetrahedral	4
2	C sp^2	trigonal	3
3	C aromatic	trigonal	3
4	C sp	linear	2
5	N sp^3 +	tetrahedral	4
6	N sp^3	tetrahedral	3
7	N azide	trigonal	2
8	N sp^2	trigonal	2
9	N trig	trigonal pl	3
10	O sp^3	single	2
11	O sp^2	double	1
12	S sp^3	tetrahedral	6 & >
13	S sp^2	double	1
14	P	–	6 & >
15	H	–	1
16	F	–	1
17	Cl	–	1
18	Br	–	1
19	I	–	1
20	User-defined	Metal	–
21	User-defined	Non-metal	–
22	O	(NO_2, NO_3^-, etc.)	1
23	N sp	(CN)	1
24	O	(PO and SO)	2 & >
25	N sp^2 +	trigonal	3
26	O $sp^3\frac{1}{2}-$	single	1
27	*spare*	–	–
28	*spare*	–	–
29	unassigned	–	–
XED type	XED description	XED geometry	XED connectivity
30	*lp (nat = –2)*	(*'rabbit ear'*)	*1*
31	pz-pi (nat = –1)	(pi points)	1
32	py-lp (nat = –2)	(90° to pi – single valence)	1
33	*px1-lp (nat = –3)*	*(linear to sigma/pi bond inner)*	*1*
34	px2-lp (nat = –3)	(linear to sigma/pi bond outer)	1
35	px3-lp (nat = –4)	(bond point)	1

The table contains extensions and modifications to the data in Refs. 5 and 6. The XED points have been allocated type numbers above 30 and a negative 'atomic number' in order to identify and distinguish them from real atoms. Italic entries are currently not used.

comparable methods. Intermolecular interactions for which there are experimental data or calculations based on ab initio wave functions have been used to validate the method against ACCs and AM1-derived natural atomic orbitals.

The philosophy of XEDs falls between ACCs and DMAs and may loosely be regarded as a form of Distributed Monopole Analysis. However, as molecules approach each other, their electronic arrangements change specifically to meet the incoming species. Molecular mechanics are parameterised from ground-state data. Molecular approach is accompanied by changes from the ground state, so XEDs have been parameterised to simulate intuitively the way molecules are expected to polarise on approach. This requires allocating integer positive (rather than partial) charges to nuclei and distributing the negative charge amongst the XEDs. At this stage, each atom is neutral. To account for atomic electronegativity and electron drift through bonds, partial charge from CHARGE3 [7], a simple empirical generator, is added to each XED pattern leaving the nuclear charge unchanged (except for nonpolar atoms with a Pauling electronegativity equal to carbon or less, which are allocated an unaltered ACC). Despite the criticism of ACCs, they are singularly successful, particularly when applied to intramolecular interactions. The application of XEDs to intramolecular minimisation procedures will be reported elsewhere.

Derivation of XEDs

The development of XEDs started with the incorporation of Hunter's aromatic points into COSMIC90 [5,6]. This was found to be adequate in reproducing many aromatic geometries without changing the original parameters. Further parameterisation was always accompanied by the question: 'How will the electron cloud of a target and bullet molecule respond as the molecules approach each other?' Adding variations of 'rabbit ears' to simulate lone pairs soon revealed that directionality could become too dominant and that different hybrid states (for example, O_{sp3} and O_{sp2}) could not be described by a common orbital arrangement. A full complement of orbital descriptors was needed which could be tailored to fit the hybridisation state from a parameter table. To this end, a 'minimal valence basis set' arrangement of a maximum of six 'orbital points' (or XEDs) was used as a starting pattern for every atom, the orthogonal components of which could be extended, retracted or eliminated according to the atom hybridisation

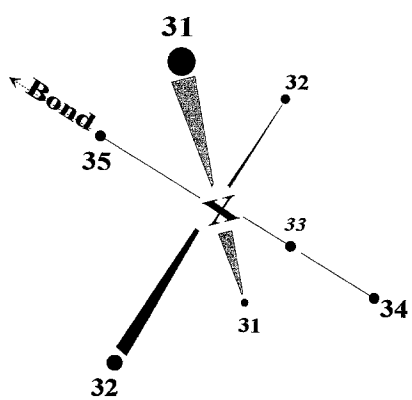

Fig. 1. General XED constructions (see Tables 1 and 2). Each XED is allocated a *type*, which defines its spatial arrangement and charge size. These data are taken from Table 2 according to the real atom type (hybrid) which bears the XED.

type. This afforded a general construct which could be varied in both charge (before electron distribution by CHARGE3) and length.

Table 1 lists the number codes of real atom types and XED types. The XED points are numbered from 30 through to 35. Following the description above, a pair of π-points is allocated type 31 and a pair, orthogonal to these first two, is labelled type 32. Type 34 is a single point extending away from the formal covalent bond, whilst type 35 is at 180° to 34 and extends backwards into the covalent bond. Type 33 was originally introduced to balance the large polarisation associated with 34, but was subsequently found to be unnecessary. Figure 1 illustrates the general scheme for XED construction. Negative atomic numbers (nat) are associated with the XED points for recognition purposes.

Individual entries for each XED type were fine-tuned for both XED lengths and XED charges, using one experimentally well-characterised complex. For example, the carbonyl O_{sp^2} XED distances and charges were finalised using the acetone–water complex for which good-quality data are available [12]. Application to other examples confirmed consistency. It was considered that, if major changes were found to be necessary for each new example, the concepts of extended electron distribution would have been wrong. This was not found; once set up to simulate a reliable experimental result, the XEDs worked well across all the other systems tested. The current set of XED parameters are reported in Table 2.

METHODS

The COSMIC molecular modelling package [5,6] was rewritten to include points (XEDs) which were designed to simulate orbital polarisation patterns. Each point is defined according to its data table entry (Table 2) which contains its XED type (Table 1), the associated atom type, the distance from the nucleus and the gross negative charge.

TABLE 2
XED LENGTHS (Å) AND CHARGES (e) IN ELECTRONS

Atom type	XED type							
	31d length	32d length	34d length	35d length	31e charge	32e charge	34e charge	35e charge
2	0.470	0.000	0.000	0.000	0.500	0.000	0.000	0.000
3	0.470	0.000	0.000	0.000	0.500	0.000	0.000	0.000
4	0.470	0.470	0.000	0.000	0.500	0.500	0.000	0.000
5	0.470	0.470	0.470	0.470	0.100	0.100	0.300	0.300
6	0.000	0.000	0.300	0.300	0.000	0.000	1.250	0.750
8	0.470	0.000	0.300	0.000	0.500	0.000	1.000	0.000
9	0.470	0.000	0.000	0.300	0.750	0.000	0.000	0.500
10	0.300	0.300	0.300	0.300	1.000	0.500	1.000	1.000
11	0.300	0.300	0.300	0.000	0.500	1.750	0.500	0.000
12	0.350	0.350	0.350	0.350	0.500	1.000	0.750	1.250
13	0.350	0.350	0.350	0.000	0.500	1.750	0.500	0.000
25	0.300	0.000	0.100	0.000	0.500	0.000	1.000	0.000
26	0.300	0.300	0.300	0.000	0.500	1.000	2.000	0.000

The table defines XED lengths and charges according to their types and the atom types they occupy as defined in Table 1. XED sets are constructed according to the atom types. d is the distance from the nucleus; e is the negative charge. Not all combinations are yet parameterised. Generally, type 33 (inner lone-pair point) is not necessary as part of the XED construct but is included for compatibility with VAMP NAOs (see Methods) which do generate these points.

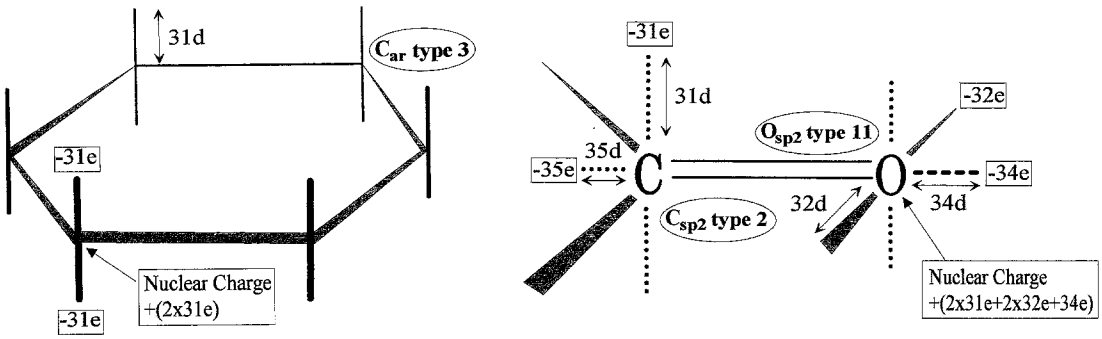

Fig. 2. XED constructions for aromatic and carbonyl arrangements (see Tables 1 and 2). Allocation from Table 2 would set −31e = −0.5e and 31d = 0.47 Å. The nuclear charge on the C_{ar} atom = (−0.5) + (−0.5) = 1.0e. The partial charge on C_{ar}, as determined from CHARGE3, is halved and added to −0.5e on each type 31 charge.

The nuclear charge is the absolute sum of the XEDs for any given ensemble. Figure 2 shows two examples of typical arrangements – the aromatic arrangement and the two ensembles making up the carbonyl group. Partial charges are calculated from the program CHARGE3 [7]. For each atom, the partial charge is divided equally among the XEDs by addition. Atoms with no XEDs are allocated their appropriate ACC. This procedure, whilst not removing the ACCs altogether, at least attenuates their effect, supplies an electronegativity contribution and polarises the molecule in an appropriate way.

From early results [9] with XEDs, Clark [8] was stimulated to add to his implementation of MOPAC (VAMP4.4) the ability to produce natural atomic orbitals (NAOs) from the AM1 Hamiltonian for ground-state molecules. These bore a remarkable resemblance to XEDs, but defined *every* atom in terms of extended orbital points (Fig. 3). It is interesting to note that an equivalent of type 33 is produced by these AM1-generated NAOs. Type 33 has not yet been found to be necessary for the systems so far examined.

XEDs are designed crudely to simulate interaction configurations on polar groups and are not expected to reproduce ground-state properties such as dipole moment or low-energy 'interstellar' conformations. With the new approach from Clark, XEDs could be compared with NAOs as well as ACCs.

The computational methodology was based on earlier docking concepts [5] and comprised a hard-sphere docking routine which places a 'mobile' molecule on evenly spaced points around

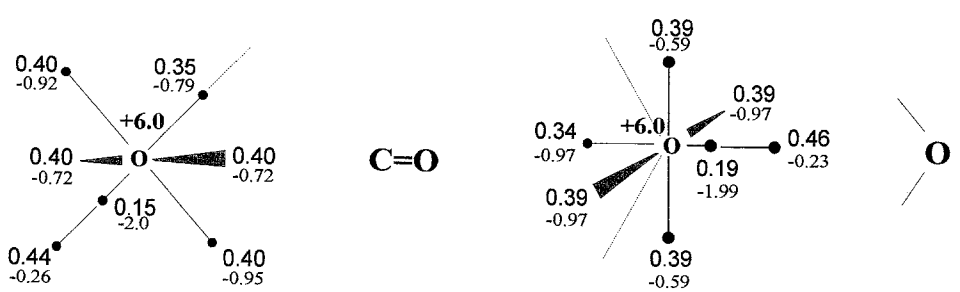

Fig. 3. Generalised AM1 NAOs averaged from 12 different environments, generated from VAMP4.4 for each oxygen motif. Signed numbers are charges; unsigned numbers are distances.

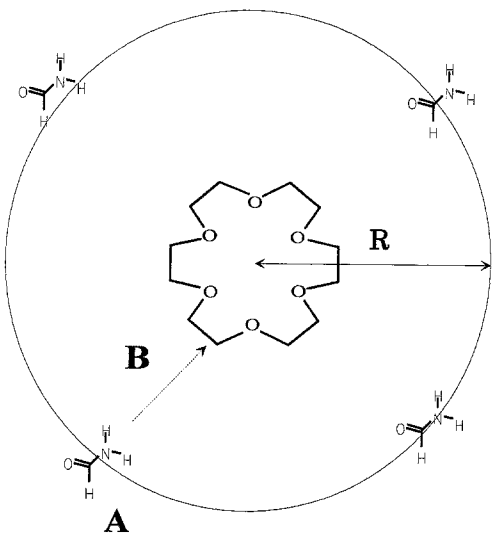

Fig. 4. Docking regime used in XEDOCK for the interaction of a mobile molecule (e.g. formamide) with a fixed molecule (e.g. 18-crown-6). R is chosen between 10 and 15 Å or automatically allocated. The first stage orients the molecule by rotation and the second stage moves the 'mobile' in towards the 'fixed', using nonbonding energy terms (vdW + Coulombic) in a 6D simplex minimiser.

a sphere of chosen radius with a 'fixed' molecule at its centre. Only nonbonding potentials are evaluated with the dielectric generally held at unity. Rotation of the mobile species on its spot, to set the most attractive start orientation, is followed by a six-dimensional simplex (rotation and translation) minimisation to an interaction energy difference of 0.0001 kcal/mol (Fig. 4) or a maximum of 1000 iterations. The number of start points was set to approximately 250 (dependent on the algorithm) and the resulting pictures represented the local minima associated with the docking of the two molecules from every available direction.

The software for molecule building (XEDRAW), for docking (XEDOCK) and for visualising and analysing the results (XEDAST) has been developed over five years on PC (486, DOS vs. 5), Silicon Graphics Personal Iris, Indigo (IRIX vs. 4) and Indy (IRIX vs. 5). Many changes to the original force field parameters have been made, the most important of which are those associated with the Morse van der Waals terms [6] reproduced in Table 3.

Both ACC and NAO calculations were incorporated into the building software to allow fast interaction. ACCs were added using the CHARGE3 [7] protocol from Abraham and VAMP NAOs were directly taken from a modified version of Clark's program [8].

RESULTS

A broad spectrum of examples was used to finalise the parameterisation of each atom type; Table 4 summarises the results from 12 representative binary complexes. The general usefulness of this approach is yet to be fully established. So far, no example has yielded poor results and further studies on larger systems will be published in due course.

Three runs were computed on each reaction with identical protocols, differing only in the type

of charge allocated to the reactants. The '_m' labels the molecules as having ACCs (unmodified partial charges from CHARGE3), the '_mx' label specifies that XEDs have been constructed and a molecule having a complete set of NAOs from an AM1 VAMP run is suffixed with an '_x'. Neutral porphyrin was constructed in two ways: with alternate type 8 (N_{sp2}) and type 9 (N_{trig}) nitrogen atoms (asymmetric) and again with all type 9 (N_{trig}) nitrogens (symmetric). This latter set was labelled with 't' added to the suffix.

Each docking run was analysed in terms of the following parameters: the molar partition function, '$Q' = \Sigma e^{-\Delta E_i/RT}$, where ΔE_i is the difference between the nonbonding energy of interaction of each dock and that of the lowest energy dock (the sum is over all docks, i.e., ca. 250) at 298 K. Q gives an idea of the number of states available to the system. For example, Q=415 for the 'XED' benzene dimer indicates a large number of complexes having similar energies. Q for valerolactam is very low because there are few states at low energy and the remaining states are separated from this small collection by a relatively large energy difference (ca. 3 kcal/mol).

Some of the column headings in Table 4 need expanding: 'Mean', 'Highest' and 'Lowest' record the mean energy over all docks, the energy of the highest (least stable) complex and the global minimum-energy complex. Again, the range, spread and species dominance of the run are reflected in these values. The breakdown into Coulombic ('Coul') and van der Waals ('vdW') contributions is noted for the global minimum-energy complex only. These values can be used to gauge the extent of electrostatic attraction versus the 'stickiness' or repulsion of contact.

The experimental ('Exp') or calculated ('Cal') energies are recorded in the next two columns. Literature references and summaries for these values are given as notes with the table and the note label is recorded in the 'Note' column. The 'Motif Fig' column contains the figure number defining the measurements for comparison with literature data (which use the same measurements) and the 'Geometries' column summarises the results of these measurements from each docking run. A figure number for the lowest energy picture in each case is held in the 'Min Fig' column. For further analysis, a screen dump ('Rng Fig') of the first few conformations from a selected docking run over the energy range ('En Rng') above the global minimum can be obtained

TABLE 3
VAN DER WAALS DATA E_{ij} (kcal/mol)

H	C	N	O	S	F	Cl	Br	I	H^+
0.022	0.081	0.076	0.083	0.082	0.139	0.139	0.164	0.196	0.051
	0.129	0.121	0.134	0.131	0.221	0.221	0.260	0.312	0.081
		0.115	0.127	0.123	0.208	0.208	0.246	0.294	*0.051*
			0.140	0.135	0.230	0.230	0.271	0.324	*0.055*
				0.132	0.224	0.224	0.262	0.313	*0.052*
					0.380	0.380	0.446	0.535	0.139
						0.380	0.446	0.535	*0.190*
							0.525	0.631	*0.200*
								0.753	0.196
									0.051

The table contains extensions and modifications to the data in Refs. 5 and 6. These values are appropriate for the Morse function in the 'XED' suite of programs originally introduced in the COSMIC90 [6] molecular mechanics package. Changes to the original table in Ref. 6 are italicised. $E_{nb} = E_{ij}\{z^2 - 2z\}$, where $z = e^{-b[1 - R/R_{ij}]}$, E_{nb}=nonbonding energy, b=constant, R=distance (in Å) between i and j, R_{ij}=sum of vdW radii (in Å), E_{ij}=energy cross term (in kcal/mol).

TABLE 4
INTERACTION ENERGIES BETWEEN SELECTED SIMPLE BINARY COMPLEXES USING ACC, XED AND

Fixed	Mobile	Complex	Q	Mean En	Highest En	Lowest En	Coul eng	vdW eng	Exp En
benzene_m	benzene_m	bzbzm	**64**	−2.32	−1.10	**−3.07**	−1.40	−1.67	
benzene_mx	benzene_mx	bzbzmx	**415**	−4.40	−3.09	**−4.64**	−1.79	−2.85	
benzene_x	benzene_x	bzbzx	**108**	−3.48	−1.06	**−4.22**	0.88	−5.10	
benzene_m	water_m	bzwm	**161**	−3.73	−1.69	**−4.22**	0.91	−5.13	−2.2
benzene_mx	water_mx	bzwmx	**98**	−3.52	−1.72	**−4.66**	−3.19	−1.47	−2.2
benzene_x	water_x	bzwx	**90**	−2.46	0.00	**−3.18**	−1.85	−1.33	−2.2
acetone_m	water_m	acwm	**271**	−3.20	−1.40	**−3.75**	−3.05	−0.70	
acetone_mx	water_mx	acwmx	**89**	−4.72	−1.56	**−5.60**	−6.38	−0.78	
acetone_x	water_x	acwx	**114**	−4.82	0.00	**−5.46**	−5.50	0.05	
aminoeth_m	water_m	amwm	**356**	−2.36	−0.59	**−2.63**	−2.17	−0.46	
aminoeth_mx	water_mx	amwmx	**88**	−4.38	−0.81	**−5.45**	−7.38	1.93	
aminoeth_x	water_x	amwx	**0**	0.00	0.00	**0.00**	0.00	0.00	
pyridine_m	water_m	pywm	**113**	−2.46	−1.55	**−3.34**	−2.04	−1.31	
pyridine_mx	water_mx	pywmx	**54**	−4.16	−2.62	**−5.05**	−5.42	0.38	
pyridine_x	water_x	pywx	**27**	−13.11	−7.57	**−14.16**	−15.22	1.06	
valerolactam_m	valerolactam_m	valvalm	**9**	−6.17	−2.45	**−8.16**	−2.21	−5.93	−10.3
valerolactam_mx	valerolactam_mx	valvalmx	**30**	−13.91	−2.38	**−14.60**	−17.36	2.76	−10.3
pyridone_m	pyridone_m	popom	**138**	−11.12	−3.26	**−11.91**	−9.69	−2.22	−5.9
pyridone_mx	pyridone_mx	popomx	**60**	−23.11	−4.80	**−23.92**	−29.84	5.92	−5.9
pyridone_x	pyridone_x	popox	**22**	−6.53	0.00	**−7.46**	−7.24	−0.22	−5.9
formamid_m	formamid_m	forform	**94**	−4.40	−3.43	**−5.73**	−5.51	−0.21	
formamid_mx	formamid_mx	forformx	**29**	−13.42	−6.15	**−14.93**	−18.48	3.55	
formamid_x	formamid_x	forforx	**0**	0.00	0.00	**0.00**	0.00	0.00	
cis-nmacet_m	formamid_m	nmbform	**137**	−4.85	−1.85	**−5.84**	−5.39	−0.45	
cis-nmacet_mx	formamid_mx	nmbformx	**20**	−13.09	−2.85	**−14.68**	−17.88	3.20	
cis-nmacet_m	acetamid_m	nmbacem	**64**	−5.42	−2.98	**−6.59**	−2.87	−3.72	
cis-nmacet_mx	acetamid_mx	nmbacemx	**15**	−14.39	−2.96	**−15.18**	−18.03	2.85	
acetic_m	acetic_m	acacm	**81**	−5.47	−1.47	**−6.63**	−6.75	0.11	−14.0
acetic_mx	acetic_mx	acacmx	**11**	−11.65	−2.10	**−12.16**	−15.87	3.41	−14.0
ala_m	ala_m	alalam	**64**	−5.56	−0.87	**−6.39**	−4.03	−2.36	
ala_mx	ala_mx	alalamx	**15**	−9.39	−1.54	**−11.87**	−15.02	3.15	
ala_x	ala_x	alalax	**51**	−46.55	−17.52	**−47.68**	−48.05	0.38	
porphyr_m (8/9)	porphyr_m (8/9)	prprm	**86**	−25.23	−7.71	**−25.64**	−4.52	−21.12	−11.4
porphyr_mx (8/9)	porphyr_mx (8/9)	prprmx	**50**	−25.13	−15.17	**−26.07**	−13.22	−12.85	−11.4
porphyr_x (8/9)	porphyr_x (8/9)	prprx	**17**	−43.64	−8.50	**−45.00**	−20.56	−24.44	−11.4
porphyr_mt (9)	porphyr_mt (9)	prprmt	**160**	−26.10	−7.48	**−26.54**	−2.13	−24.42	−11.4
porphyr_mxt (9)	porphyr_mxt (9)	prprmxt	**215**	−27.56	−6.49	**−27.97**	−4.38	−23.59	−11.4
porphyr_xt (9)	porphyr_xt (9)	prprxt	**30**	−27.59	−1.18	**−28.30**	−4.63	−23.67	−11.4

aminoeth = ethylamine, formamid = formamide, cis-nmacet = *cis-N*-methylacetamide, ala = neutral alanine, porphy = neutral porphyrin. 'm' indicates atom-centred charges, 'mx' indicates extended electron distributions, 'x' indicates AM1-derived natural atomic orbitals. Energies are in kcal/mol; distances are in Å; angles are in degrees.

[a] Ref. 10. Calculated results from DMA on a double zeta basis set wave function. The best motif is a T shape, R = 4.7. There are six other motifs. None was parallel with overlap. Electrostatic energy range: −1.12 to −0.65 kcal/mol.

[b] Ref. 11. Experimental results from vibration-rotation tunnelling spectra. r = 3.35, a = 20 ± 15. Calculation using 6-31G**-MP2, r = 3.195, a = 24. Free z axis rotation. When no MP2 CI was employed, a = 0. Calculated interaction energy is ±0.6 kcal/mol.

[c] Ref. 12. Combined 6-31G** and FT-IR. Energies: CO-HO = −4.3 kcal/mol; CH-OH = −1.9 kcal/mol. r1 = 1.90, r2 = 2.24, a1 = 113.9, a2 = 143.0, a3 = 73.6. Note that the 3-21G basis set gave a poorer fit to spectral results.

[d] Ref. 13. Calculation using the 6-31G** basis set, r = 2.17, a1 = 117.1, a2 = 185.3. Near-linear H-bonding.

[e] Ref. 13. Calculation using the 6-31G** basis set, r = 2.12, a1 = 141.0, a2 = 155.6. Nonlinear H-bonding.

[f] Ref. 14. Experimental energy from near-IR-derived association constants in carbon tetrachloride. Geometry not specified.

[g] Ref. 15. Experimental results from gas-phase high-resolution laser spectroscopy; r = 2.75 ± 0.03, R = 5.3 ± 0.03, q1 =

NAO CHARGE DISTRIBUTIONS

Cal En	Foot-note	Motif Fig.	Geometries	Min Fig.	Rng Fig.	En Rng
−1.1	a	5a	Parallel to tilted. No agreement with Ref. 1	5b	5aa	0.2
−1.1			T shape, R = 4.76 + six motifs over 0.45 range as in Ref. 1	5c	5bb	0.1
−1.1			Parallel to tilted. No agreement with Ref. 1	5d		0.2
	b	6a	r = 3.21, a = 2.4	6b		
			r = 3.20, a = 2.8	6c		
			r = 3.52, a = 110.1	6d		
−6.2	c	7a	r1 = 2.36, r2 = 2.58, a1 = 117.5, a2 = 137.6, a3 = 76.4	7b		
−6.2			r1 = 1.94, r2 = 2.48, a1 = 118.1, a2 = 142.9, a3 = 61.1	7c		
−6.2			r1 = 2.16, r2 = 2.60, a1 = 126.6, a2 = 119.5, a3 = 37.2	7d		
−6.3	d	8a	r = 2.69 (2.97), a1 = 83.3(109.8), a2 = 77.9(94.6)	8b	8aa	0.2
−6.3			r = 2.09, a1 = 106.5, a2 = 174.8	8c	8bb	0.5
−6.3			No docked species.			
−6.0	e	9a	r = 2.76(2.90), a1 = 93.8(79.3), a2 = 114.2(104.6)	9b	9aa	0.4
−6.0			r = 2.27, a1 = 126.1, a2 = 128.5	9c	9bb	0.2
−6.0			r = 2.34, a1 = 115.6, a2 = 82.9	9d	9cc	3.0
	f	10a	r = 3.63(4.30), R = 3.68, q1 = 74.4(69.5), q2 = 76.6(60.6)		10aa	1.0
			r = 2.83, R = 5.45, q1 = 40.6, q2 = 123.5, linear.		10bb	1.0
−21.0	g	10a	r = 2.77, R = 3.95, q1 = 66.9, q2 = 92.4, nonlinear.		11aa	0.5
−17.1			r = 2.72, R = 5.19, q1 = 41.7, q2 = 123.7, linear.	11b	11bb	0.5
−27.3			r = 2.80, R = 5.80, q1 = 26.3, q2 = 137.6, one H-bond.	11c	11cc	0.5
−14.2	h	11	r1 = 2.08(2.19), r2 = 3.03(3.19), a1 = 176(157), a2 = 32 (nonplanar)		12aa	0.7
−14.2			r1 = 1.83, r2 = 2.83, a1 = 166.1, a2 = 26 (planar)		12bb	1.0
−14.2			No docked species.			
−14.5	h	11	r1 = 2.12, r2 = 3.08(3.12), a1 = 149(160), a2 = 40 (nonplanar)			
−14.5			r1 = 1.83(1.86), r2 = 2.82(2.85), a1 = 171.8(172.0), a2 = 29 (planar)			
−14.5			Stacked – no linear H-bonding.		13cc	0.5
−14.5			r1 = 1.85(1.86), r2 = 2.85, a1 = 170.8(175), a2 = 26(28) (planar)		13dd	2.0
	i	12	r1 = 3.00, r2 = 0.17, a = 140.0			
			r1 = 2.76, r2 = 0.04, a = 127.6			
−14.8	j	13a	r1 = 2.13, r2 = 2.82, a = 113.7, one H-bond.	13b	15aa	0.1
−14.8			r1 = 1.84, r2 = 2.77, a = 165.0	13c	15bb	1.3
−14.8			Two H-bonds: C-O-HOC and C-O-HN	13d	15cc	0.5
	k	14a	r1 = 3.01, r2 = 3.72, no Zn used. No definite motif.	14b	16aa	2.0
			r1 = 3.03, r2 = 4.69, no Zn used. No parallel stack, 'fanned'.	14c	16bb	1.5
			r1 = 3.29, r2 = 0.00, no Zn used. Central axis of symmetry.	14d	16cc	1.0
−30.3			r1 = 3.07, r2 = 3.06, no Zn used. No definite motif.	14e	16dd	1.0
−22.5			r1 = 3.01, r2 = 3.41, no Zn used. Specific motif.	14f	16ee	2.0
			r1 = 3.00, r2 = 3.30, no Zn used. 'Vibration' on specific motif.	14g	16ff	2.0

43 ± 2, q2 = 122 ± 2. Ref. 16. Calculated energies from three approaches: (1) a van der Waals term plus ESP electrostatic charges derived from a 3-21G basis set, energy = −21.0 kcal/mol, R = 5.19, q1 = 39.2; (2) ab initio (GAMESS) 3-21G basis set, energy = −27.3 kcal/mol, R = 5.22, q1 = 39.6(av); (3) DMA using 3-21G basis set, energy = −17.1 kcal/mol, R = 5.48, q1 = 38.1.Ref. 17. Experimental association energy from ultrasonic attenuation measurements in chloroform.

[h] Refs. 18 and 19. Experimental geometries from X-ray data: r1 = 1.93 (0.12), r2 = 2.89 (0.11), a1 = 161.2 (15), a2 = 0°–90°, centred around 37° (standard deviations are given in brackets). Ref. 20. DMA on a 3-21G basis set wave function. a2 = 38.5 ± 6; a1 = 175.6 (3.3).

[i] Ref. 21. Experimental association energy not accurately determined, but known to be approximately 0.6 eV in the gas phase. Ref. 22. X-ray: r1 = 2.62–2.70, r2 = 0.05–0.38, a = 108.5–131.9.

[j] Ref. 23. GEOTRIPLEX** basis set with BSSI correction (Clementi): energy = −14.8 kcal/mol, r1 = 1.85, r2 = 2.79, a = 168.

[k] Refs. 4 and 24. Experimental results gave a specific motif with no Z rotation, r1 = 3.4–3.6, r2 = 3–4 (zinc included). Ref. 16. Calculations using: (1) a van der Waals term plus ESP electrostatic charges derived from a 3-21G basis set, energy = −30.3 kcal/mol, r1 = 3.0, r2 = 3.6 (zinc included); (2) DMA using a 6-31G basis set wave function, energy = −22.5 kcal/mol, r1 = 3.2, r2 = 2.3, rotation = 45° (zinc included).

as supplementary material from the author. These indicate the preponderance and stability of the best docked species.

DISCUSSION

The project set out to assess the value of using XEDs to describe Coulombic intermolecular interactions as compared with ACCs and AM1-derived NAOs. The general conclusion from the examination of Table 4 and its accompanying figures is that XEDs give rise to better overall geometries and energies of interaction than do ACCs or NAOs. Table 4 summarises the experimental and molecular mechanics results. At this stage of development, only qualitative agreement was sought. No attempt has been made to account for entropy or solvation, as can be seen by comparing interaction energies from experiment and calculations based on ab initio wave functions.

Aromatics

The geometry and parameter motifs are displayed in Figs. 5 and 6. The XED-derived 'T'-shaped benzene dimer (Fig. 5c) corresponds to the M1 structure of Price and Stone [10]. All other motifs described by these authors are present in the full DOCK picture. Both ACCs and NAOs produce stacked dimers with little differentiation (Figs. 5b,d). Although the interaction energy from XEDs is the highest of the three runs, its associated Q-value reflects the closeness of all dimer energies. The benzene–water complex is reasonably well simulated by both ACCs (Fig. 6b) and XEDs (Fig. 6c), although the 'tipping' of the water, only revealed after MP2 correlation in Ref. 11, is not reproduced. The free rotation of water about the benzene axis of symmetry is probably better described by ACCs (see the Q-value).

More water complexes

Both XEDs (Fig. 7c) and ACCs (Fig. 7b) reproduce the preferred lone-pair directionality (parameter definitions in Fig. 7a) of the acetone–water complex. This is not reproduced by NAOs

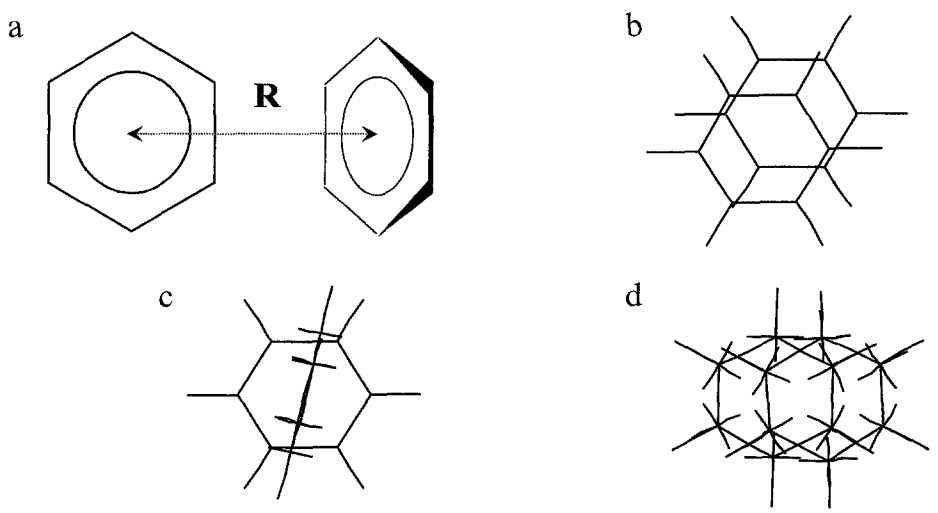

Fig. 5. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for benzene.

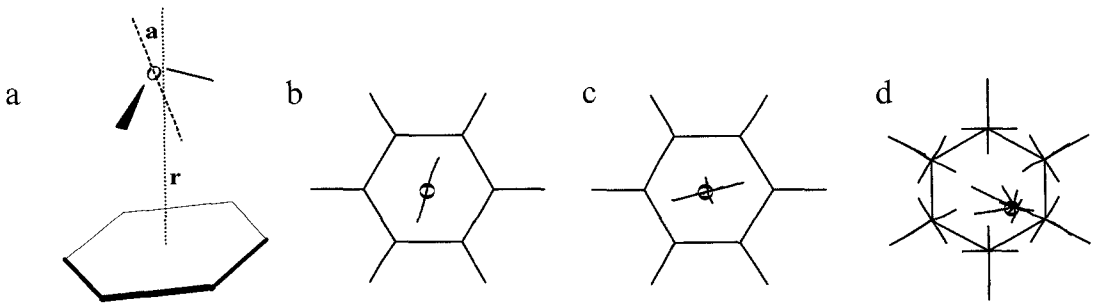

Fig. 6. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for a benzene–water complex.

(Fig. 7d). Both the interaction energy and the geometry of this complex are better described using XEDs. Further investigations of the lone-pair 'bond' using ethylamine (Fig. 8) and pyridine (Fig. 9) with water show more obvious deviations for both ACCs (Figs. 8b and 9b) and NAOs (Fig. 9d). Neither base positioned any of its docked waters in line with the lone pair when ACCs were allocated; a broad halo of solvent was produced around ethylamine and pyridine preferred to stack its solvent above and below the aromatic plane. With NAOs, only a dozen waters docked onto ethylamine, none of which were chemically acceptable. The NAO pyridine complex produced a concentrated water group roughly in the right place but with the wrong orientation. Again, XEDs simulated the experimental results [13] closest, with the best energy and geometry correspondence (Figs. 8c and 9c).

Cyclic amides

The next five examples were designed to probe the effectiveness of XEDs on amide complexes, so important in biological modelling. The best experimental geometry for this group exists for the pyridone complex [15] using the parameters shown in Fig. 10a. This is reproduced by XEDs (Fig. 10c), but not by either of the other techniques (Figs. 10b,d). ACCs stack the amide and the use of NAOs results in the formation of only one N-H--O=C bond. The experimental enthalpy was determined in chloroform (–5.9 kcal/mol) and differs from both the XED energy (–23.92 kcal/mol) and those derived from ESPs, 3-21G and DMA calculations [16]. By setting the dielectric (D) at

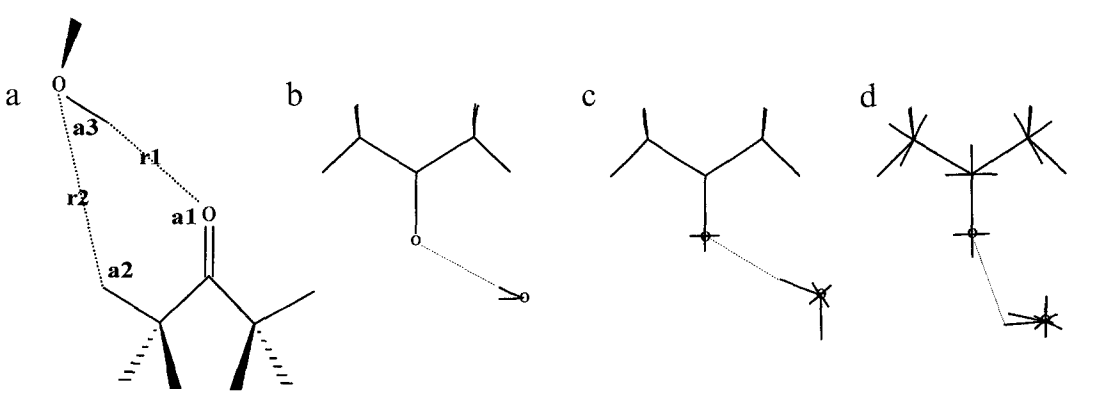

Fig. 7. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for an acetone–water complex.

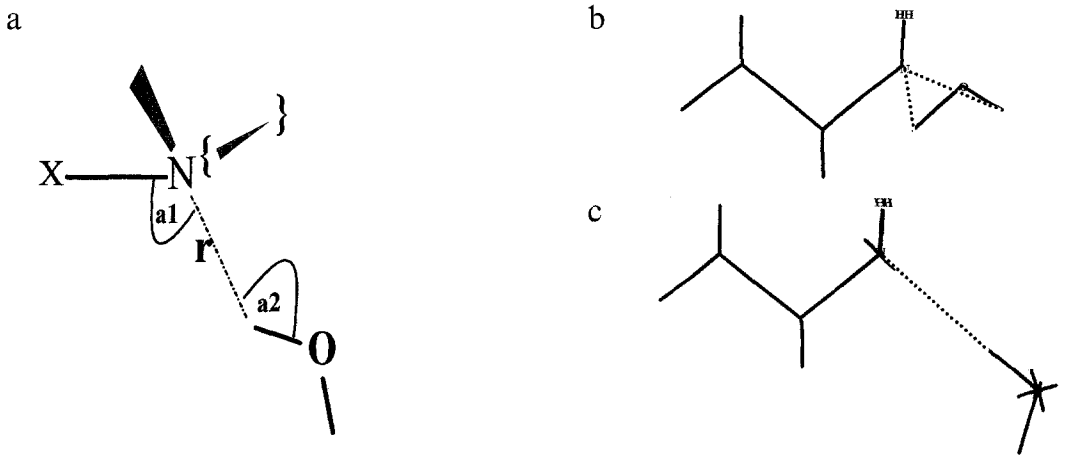

Fig. 8. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for an ethylamine–water complex.

4.8 (chloroform constant), a better interaction energy of –6.72 kcal/mol resulted from a dock run, but the linear double hydrogen bond was found to have reverted to a stacked arrangement.

Valerolactam behaved in a similar way, preferring to stack with ACCs but keeping linear with XEDs. NAOs gave few docked species with inconsistent geometries. An experimental geometry was not found for valerolactam and a linear arrangement similar to that of the pyridone dimer was assumed (Fig. 10a). The experimental association enthalpy was determined in carbon tetrachloride (D = 2.2) and found to be –10.3 kcal/mol. Again, by setting the dielectric at 2.2, a decrease to –6.71 kcal/mol was achieved. However, conversion from the linear to the stacked geometry occurred when the dielectric was between 1.5 and 2.0, with a marked change of energy and a difference of only 0.4 kcal/mol existing between the linear and stacked forms.

Table 5 records a series of docks over a dielectric range for both pyridone and valerolactam. The turnover range from linear to stacked geometry is highlighted.

Acyclic amides

The interaction energies for two acyclic amides were all taken from ab initio calculations [20]

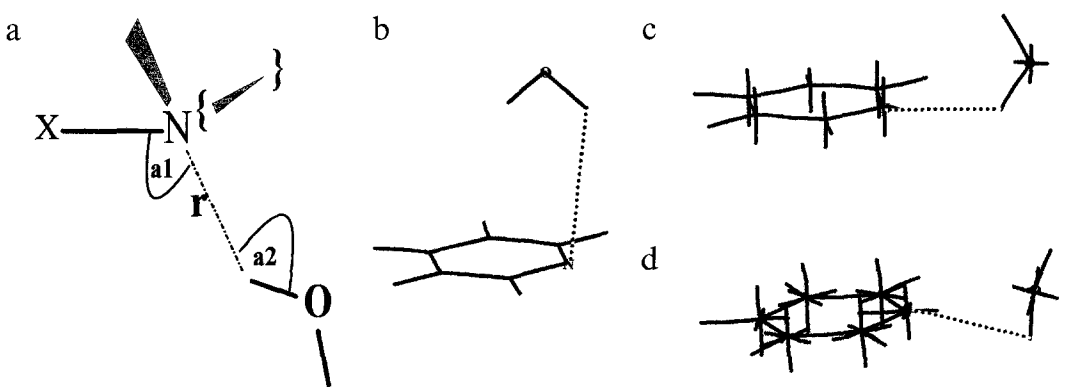

Fig. 9. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for a pyridine–water complex.

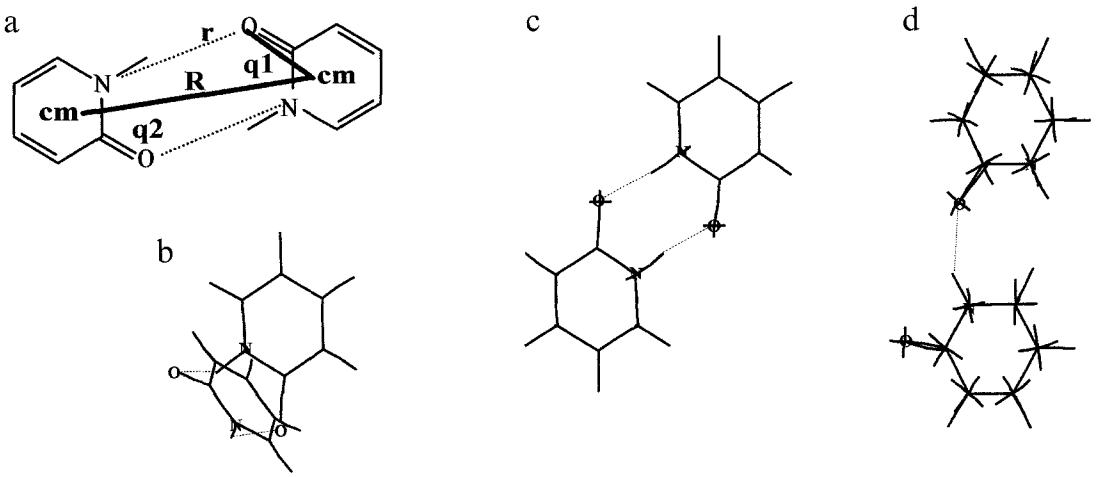

Fig. 10. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for a pyridone complex.

of the electrostatic energy only. The geometry motif and parameters recorded in Table 4 for both examples are defined in Fig. 11. Comparison with the 'Coul' energy column in Table 4 would be a fairer estimate. Nevertheless, the literature values are in reasonable agreement with the XED energies. Some NAOs were attempted for this group but did not dock consistently.

TABLE 5
PYRIDONE AND VALEROLACTAM DIMERISATION WITH XEDOCK AT DIFFERENT DIELECTRIC VALUES

Fixed and mobile	Complex	Dielectric	Q	Mean energy	Lowest energy
pyridone_mx	popomx_1	1	60	−23.11	−23.92
pyridone_mx	popomx_1h	1.5	62	−13.92	−14.75
pyridone_mx	popomx_2	2	79	−10.04	−10.76
pyridone_mx	**popomx_2h**	**2.5**	**67**	**−7.38**	**−8.56**
pyridone_mx	**popomx_3**	**3**	**68**	**−6.39**	**−7.54**
pyridone_mx	popomx_3h	3.5	28	−5.76	−7.18
pyridone_mx	popomx_4	4	19	−5.53	−7.01
pyridone_mx	popomx_4h	4.5	84	−5.46	−6.83
pyridone_mx	popomx_chcl3	4.8	62	−5.24	−6.72
pyridone_mx	popomx_5	5	63	−5.30	−6.68
valerolactam_mx	valvalmx_1	1	30	−13.90	−14.60
valerolactam_mx	**valvalmx_1h**	**1.5**	**28**	**−8.01**	**−9.27**
valerolactam_mx	**valvalmx_2**	**2**	**25**	**−5.68**	**−7.18**
valerolactam_mx	valvalmx_ccl4	2.2	75	−5.39	−6.71
valerolactam_mx	valvalmx_2h	2.5	42	−4.99	−6.55
valerolactam_mx	valvalmx_3	3	35	−4.79	−6.64
valerolactam_mx	valvalmx_3h	3.5	31	−4.74	−6.48
valerolactam_mx	valvalmx_4	4	28	−4.71	−6.51

All energies are in kcal/mol. For pyridone, the experimental association enthalpy [17] is −5.9 kcal/mol in chloroform (D = 4.8). Turnover from linear to stacked dimer occurs between D = 2.5 and 3.0 (bold). For valerolactam, the experimental association enthalpy [14] is −10.3 kcal/mol in carbon tetrachloride (D = 2.2). Turnover from linear to stacked dimer occurs between D = 1.5 and 2.0 (bold).

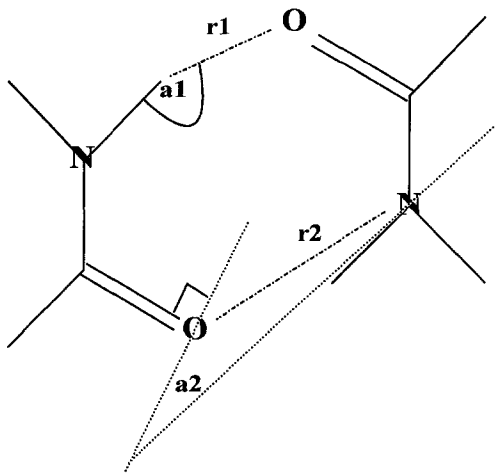

Fig. 11. Literature comparison motif, as defined in Table 4, for an acyclic amide.

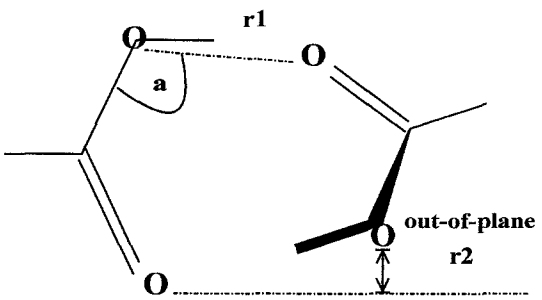

Fig. 12. Literature comparison motif, as defined in Table 4, for acetic acid.

Carboxyl dimers

Acetic acid (Fig. 12) and alanine (Fig. 13) were used to investigate carboxylic acid dimerisation. Both examples were well reproduced by XEDs. The interaction energy for acetic acid was drawn from Ref. 21 as an approximate value of −0.6 eV in the gas phase. There is an excellent correspondence between the geometry of the XED-generated dimer of alanine (Fig. 13c) and that derived from Clementi's best basis set [23].

Porphyrin

Finally, following Hunter and Sanders [4], the porphyrin complex was re-investigated. Addition of the central zinc atom was omitted based on the fact that accepted parameters for zinc in the molecular mechanics formalism are not available and a contrived arrangement was not felt to be

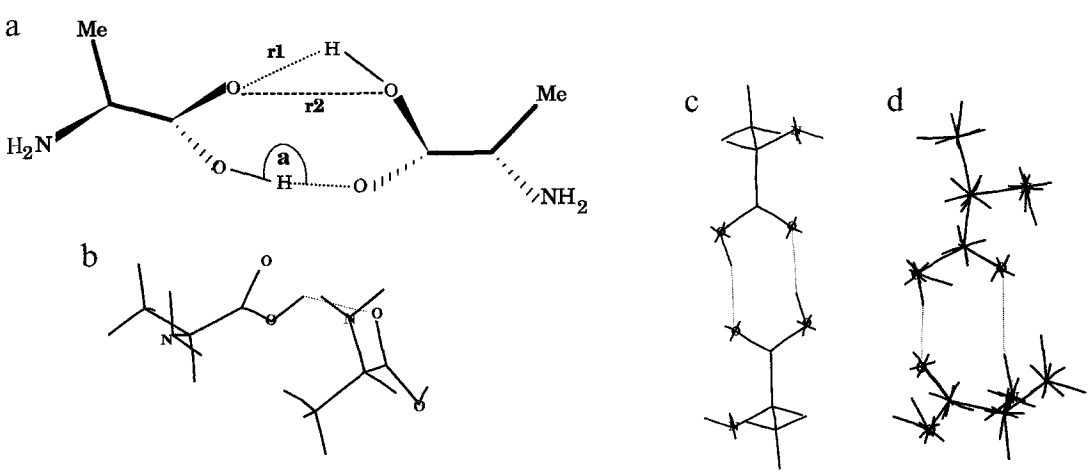

Fig. 13. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for alanine.

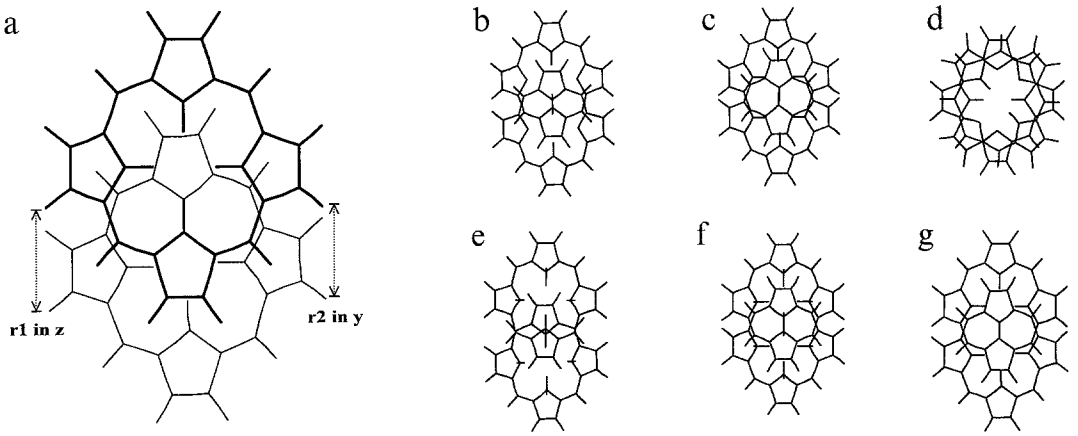

Fig. 14. Literature comparison motif (a) and lowest energy patterns, as defined in Table 4, for porphyrin.

justified. Neutral porphyrin could be built using either two type 8 and two type 9 nitrogens or four type 9 (N_{trig}) nitrogens. Both were investigated and the success of the methods can be judged from Figs. 14b–d for the mixed nitrogen type and Figs. 14e–g for the 'all N_{trig} type 9' structure. In general, the 'all type 9' symmetric porphyrin docking was scientifically preferred. The remarkable variation in organisation of the complexes was not expected and highlights the sensitivity of the protocols. The complex derived from XED allocations is again the best, but it will be noted that all the examples have an aromatic separation of about 3 Å. The experimental value is nearer 3.4 Å and it is assumed that the exclusion of the central zinc may have caused this discrepancy. Yet again, the interaction energies from ab initio [16] and XED calculations are somewhat higher than the experimental enthalpy of −11.4 kcal/mol [4].

Docking patterns

The ultimate column in Table 4 specifies the energy ranges from the global minimum (as defined by the docking procedure) depicted in Figs. 5aa to 16ff (supplementary material). The pictures show the patterns of the most significant docked binaries likely to contribute to the overall docked population, observable for example by NMR at room temperature.

CONCLUSIONS

The comparison of simple binary complex association energies and geometries from experimental and ab initio data with those calculated from molecules built with (i) atom-centred charges (ACCs) and (ii) AM1-derived (without polarisability or CI) natural atomic orbitals (NAOs) indicated little correspondence. This is to be expected if it is understood that both (i) and (ii) are derived from a 'resting' state. At large distances compared with the NAO lengths, an NAO array will be regarded as similar to an ACC by an incoming molecule. Also, at close quarters, the full NAO description on every atom leaves little geometric choice for the incoming molecule to specify a particular geometry. Extended electron distributions (XEDs) have been devised to simulate the electronic responses accompanying the approach and interaction of two nonbonding molecules and this method is more successful in its simulations of experimental and ab initio data.

Although still to be completed and honed, XEDs seem to be more reliable in all examples of intermolecular contact so far examined. Although the general usefulness of atom-centred charges is not denied, their use in some instances of intermolecular interactions gives wrong answers, as does the use of AM1-derived NAOs. No cases using XEDs are grossly in error.

It is accepted that the internal energy changes of moderately rigid molecules will be governed by bond, angle and close vdW energies and that Coulombic terms can often be ignored. It would be expected, however, that electron distribution should be important in intramolecular processes for those larger flexible molecules which can bring their groups together and interact through space. The fact that ACCs work well at close quarters (intramolecular distances) may be because the amount of distortion to XEDs might have to be much greater than the XEDs presented here for intermolecular association and the ACC is the 'lesser of two evils'. The XEDs in Table 3 have been parameterised for intermolecular communication and ultimately for the generation of more reliable electrostatic fields, both of which are dominated by nonbonding interactions.

ACKNOWLEDGEMENTS

The SERC is thanked for its financial support. Special thanks are due to Mark Gardner and Keith Trollope for their scientific input and support.

REFERENCES

1 Stone, A.J. and Price, S.L., J. Phys. Chem., 92 (1988) 3325.
2 a. Brobjer, J.T. and Murrell, J.N., J. Chem. Soc., Faraday Trans. 2, 78 (1982) 1853.
b. Buckingham, A.D. and Fowler, P.W., Can. J. Chem., 63 (1985) 2018.
3 Stone, A.J. and Alderton, M., Mol. Phys., 56 (1985) 1047.
4 Hunter, C.A. and Sanders, J.K.M., J. Am. Chem. Soc., 112 (1990) 5525.
5 Vinter, J.G., Davis, A. and Saunders, M.R., J. Comput.-Aided Mol. Design, 1 (1987) 31.
6 Morley, S.D., Abraham, R.J., Haworth, I.S., Jackson, D.E., Saunders, M.R. and Vinter, J.G., J. Comput.-Aided Mol. Design, 5 (1991) 475.
7 Abraham, R.J. and Smith, P.E., J. Comput.-Aided Mol. Design, 3 (1989) 175 and references cited therein.
8 Rauhut, G. and Clark, T., J. Comput. Chem., 14 (1993) 503.
9 Vinter, J.G., presented at the Molecular Graphics Society One Day Meeting, Harlow, 1991 and at the 1992 European Advanced Computational Chemistry Workshop, Oxford, 1992.
10 Price, S.L. and Stone, A.J., J. Chem. Phys., 86 (1987) 2859.
11 Susuki, S., Green, P.G., Bumgarner, S.D., Goddard III, W.A. and Blake, G.A., Science, 257 (1992) 942.
12 Zhang, X.K., Lewars, E.G., March, R.E. and Parnis, J.M., J. Phys. Chem., 97 (1993) 4320.
13 Gould, I.R. and Hillier, I.H., J. Mol. Struct. (THEOCHEM), 314 (1994) 128.
14 Affsprung, H.E., Christian, S.D. and Worley, J.D., Spectrochim. Acta, 20 (1964) 1415.
15 Held, A. and Pratt, D.W., J. Am. Chem. Soc., 112 (1990) 8629.
16 Trollope, K.I., Ph.D. Thesis, University of Manchester, Manchester, 1992, p. 99.
17 Hammes, G.G. and Park, A.C., J. Am. Chem. Soc., 94 (1969) 956.
18 Taylor, R., Kennard, O. and Versichel, W., Acta Crystallogr., B40 (1984) 280.
19 Taylor, R., Kennard, O. and Versichel, W., J. Am. Chem. Soc., 105 (1983) 5761.
20 Mitchell, J.O.B. and Price, S.L., Chem. Phys. Lett., 154 (1989) 267.
21 Faubel, M. and Kisters, Th., Nature, 339 (1989) 527.
22 Donohue, J., Acta Crystallogr., B24 (1968) 1558.
23 Sordo, J.A., Sordo, T.L., Fernandez, G.M., Gomberts, R., Chin, S. and Clementi, E., J. Chem. Phys., 90 (1989) 6361.
24 Leighton, P., Cowan, J.A., Abraham, R.J. and Sanders, J.K.L., J. Org. Chem., 53 (1988) 733.