Ab initio Models for Receptor–Ligand Interactions in Proteins. Part 1. Models for Asparagine, Glutamine, Serine, Threonine and Tyrosine

Jarmo Lindroos, Mikael Peräkylä, Jussi-Pekka Björkroth and Tapani A. Pakkanen*
University of Joensuu, Department of Chemistry, PO Box 111, 80101, Joensuu, Finland.

Model compounds to be used in quantum mechanical receptor–ligand model calculations have been generated for the amino acids asparagine, glutamine, serine threonine and tyrosine using 3-21G and 6-31G basis sets. Interaction energy surfaces of the amino acids and candidate model compounds have been studied using methanol as a probe molecule. Acetamide is found to model asparagine and glutamine, ethanol serine and threonine, and phenol tyrosine. In the case of phenol, the splicing of basis set technique is found to be useful in further reducing the computational demands of the model compound.

For the studies of receptor–ligand interactions in proteins by means of computational chemistry there are in the main two complementary methods available, classical mechanics and quantum mechanics. The advantages of each method are well recognized. Classical mechanics can be used for large systems and, by the use of molecular dynamics, can give insight into the dynamical processes in macromolecules.[1] Also free energy calculations, which apply classical mechanics, are useful in the studies of receptor–ligand interactions.[2] When accurate knowledge of interactions at the electronic level is needed, *ab initio* quantum mechanics is the method to be used. Because of the computational expense of *ab initio* quantum mechanical calculations, the number of atoms which can be included is severely limited and so the use of *ab initio* methods is restricted to relatively small receptor–ligand model systems. This is a serious limitation for a realistic model because the surrounding protein matrix has an important bearing on the receptor–ligand interactions and for example on enzyme catalysis.[3]

Several attempts have been made to take environmental effects into account. In one approach, a set of partial charges are used in the place of surrounding atoms to reproduce the electrostatic potential of the protein matrix.[3–5] More recently, reaction field methods, in which the ligand is placed within either a homogenous[6] or heterogenous reaction field[7] have been used. Although these methods may satisfactorily describe the electrostatic interactions, they do not appropriately describe the whole system because of the neglect of geometric changes introduced by the two interacting parts.

In the combined quantum mechanical–molecular mechanical methodology the most important residues of the receptor and the ligand are modelled using quantum mechanics within a large molecular mechanical system.[4,8] Such integrated calculations have previously been done for some systems. In these calculations only a limited part has been calculated using the *ab initio* method[4,5,9] or semiempirical quantum mechanics have been used.[10] Calculations using the empirical valence bond method have also been performed.[11] The use of semiempirical quantum mechanics in the calculations of intermolecular interactions and especially hydrogen bonding has well known limitations,[12] although lately the results of hydrogen bonding complexes using the semiempirical PM3 model have been shown to generally agree with *ab initio* calculations.[13] In contrast to the semiempirical methods, *ab initio* methods do not have the limitations of parametrization, but the accuracy of the calculations is limited by the ability of the selected basis set to describe the system. Since the basis sets in receptor–ligand calculations must often be of modest size, the reliability of such calculations needs a case-by-case validation with more extended basis sets.[14] The inclusion of electron correlation, which is especially important in bond-breaking and -making processes, is limited in *ab initio* calculations, owing to the high computational costs of correlation calculations.

The usefulness of *ab initio* calculations of receptor–ligand systems as well as *ab initio* quantum mechanical–molecular mechanical methodology would increase, if larger systems could be studied. One way to achieve this is to use model compounds instead of whole ligand-binding amino acids. The use of model compounds in the *ab initio* calculations to reduce the size of the system is a well known and widely used procedure.

We have started using this procedure by developing the smallest possible model compounds which can reproduce the key steric and electrostatic properties of the amino acids. These model compounds will be used in the quantum mechanical receptor–ligand models and the model assemblies will be incorporated into the integrated quantum mechanical–molecular mechanical calculations. Thus, in an *ab initio* calculation the essential amino acid residues can be replaced by these amino acid model compounds and a model assembly can be constructed, which reproduces the essential features of the receptor–ligand system. Using this approach, for example in the study of enzyme catalysis, the changes in electronic structure occurring during the reaction can be studied using an *ab initio* model assembly which does not need to suffer from major simplications.

In the present paper we report the results from the *ab initio* molecular orbital calculations aimed to develop simple model compounds for the amino acids asparagine, glutamine, serine, threonine and tyrosine. The study will demonstrate the usefulness of the model compound approach, and model compounds are generated for a representative set of amino acids with both hydrogen bond donor and acceptor nature and different steric properties.

Method of Calculation

The starting structures of the amino acids were taken from the fragment library of the molecular modelling program ChemX.[15] The amino acids were first minimized as *N*-acetyl- and *C*-methyl-amide derivatives using the AMBER[16] force-field. These derivatives were selected as starting structures because, unlike crystal structures of amino acids, they were found to be representative models for the amino acid residues in the peptide chains. For the *ab initio* calculations the acetyl and methyl groups of the *N*-acetyl- and *C*-methyl-amide derivatives were removed, the CO_2H-groups were replaced by CHO-groups and the amino acids were treated as neutral species (Figs. 1, 2 and 3).

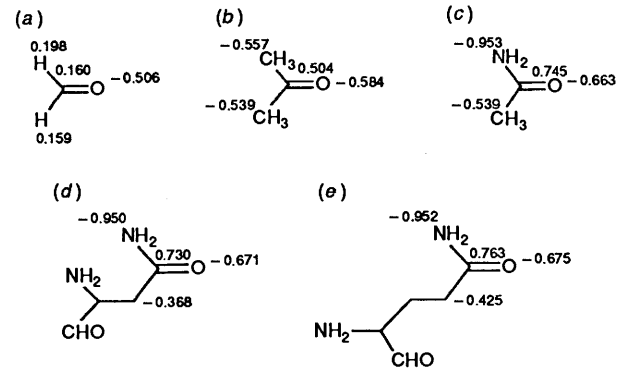

Fig. 1 Mulliken partial charges for selected atoms of (*a*) formaldehyde, (*b*) acetone, (*c*) acetamide, (*d*) asparagine model and (*e*) glutamine model in the optimized methanol complexes using 6-31//3-21G

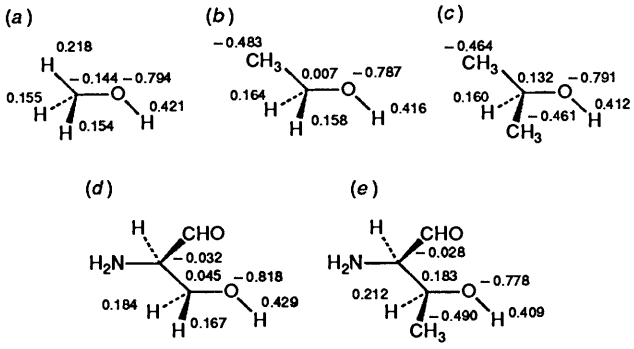

Fig. 2 Mulliken partial charges for selected atoms of (*a*) methanol, (*b*) ethanol, (*c*) isopropyl alcohol, (*d*) serine model and (*e*) threonine model in the optimized methanol complexes using 6-31//3-21G

Before the interaction calculations the amino acids were optimized using the 3-21G basis set.

In the interaction calculations methanol was selected as a probe, because it has the ability to act as both hydrogen bond donor and acceptor. It is also simple enough for the *ab initio* calculations and gives a more realistic view of the ligand binding, than smaller probes, for example H_2O. This selection is supported by the successful comparison of the calculated interaction energy surfaces of the present study and the observed hydrogen bonding patterns in the crystals.

The interaction energy surfaces were mapped by calculating the amino acid–methanol and model compound–methanol interactions at selected points around the binding groups (–OH, –C=O) of the amino acids and the model compounds. The definition of the interaction parameters is shown in Fig. 4. The amino acids and model compounds were kept partly frozen during the optimizations: only the binding functional OH and C=O groups were allowed to move freely. This was done because the structures of the amino acids change only slightly during the optimization of the complexes and the fixed amino acids model the real situation in proteins better than a completely free system. In all the interaction calculations complexes were optimized without any constraints on methanol. The points around the binding groups were defined by fixing the oxygen of the methanol at selected locations around it. The fixing was done by defining only the angle (A_0) and the torsion (T_0) of the oxygen of the methanol with respect to the binding group. The interaction parameter R_0 was optimized at each A_0/T_0 placement.

The ability of the model compounds to model the amino acids were determined in terms of the geometry of the binding, interaction energies, atomic charges obtained from the Mulliken population analysis,[17] and the shape of the interaction energy surfaces. The interaction energies were calculated

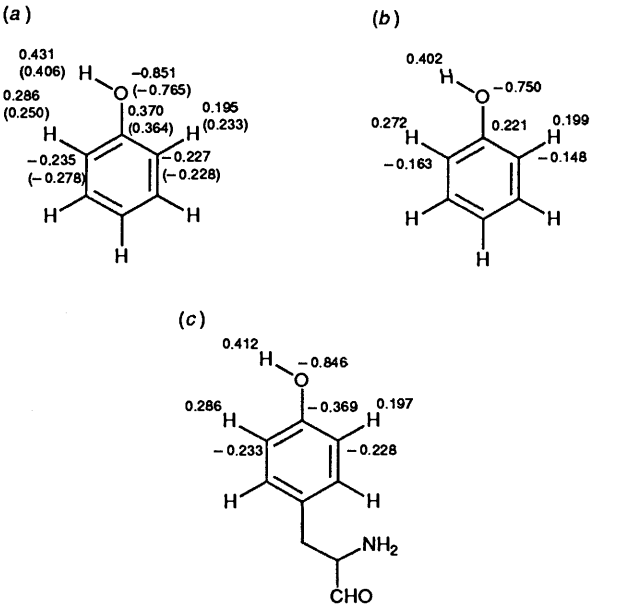

Fig. 3 Mulliken partial charges for selected atoms of (*a*) phenol (6-31//3-21G, 3-21G in parentheses), (*b*) phenol(spl) (see text for the basis set) and (*c*) tyrosine model (6-31//3-21G) in the optimized methanol complexes

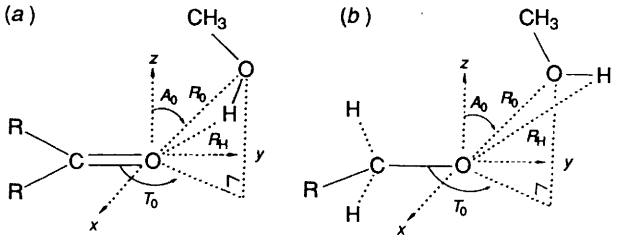

Fig. 4 Definition of the interaction parameters for the complexes of methanol with (*a*) C=O and (*b*) OH group. The *z*-axis is perpendicular to the RCO plane.

as the difference between the total energies of the isolated molecules and the energy of the supermolecule.

The basis set superposition error (BSSE) was not taken into account because we were primarily comparing similar complexes to each other and correction for the BSSE would not have changed the conclusions.

The *ab initio* molecular orbital calculations were performed using Gaussian 88[18] and Gaussian 90[19] programs implemented in VAX 785, VAX 6420 and Cray X-MP EA/432 computers. Geometry optimizations were mostly done using the 3-21G basis set. Interaction energies and atomic charges were also calculated with the 6-31G basis set using the 3-21G geometries. Additional calculations with standard basis sets with polarization and diffuse functions including the second-order Møller–Plesset (MP2) correlation corrections were done for the ethanol–methanol dimer in order to evaluate the performance of the basis sets.

Results and Discussion

The Role of the Basis Set.—The influence of the basis set on the description of the hydrogen bonding between two hydroxy groups was first studied. In Table 1 are the results of nine different basis sets for the ethanol–methanol model system. Although small hydrogen bonding systems have been extensively studied previously, we found it necessary to make these calculations because results for ethanol–methanol complexes have not been reported. In Table 2 the parameters describing the optimized binding geometries of ethanol–methanol com-

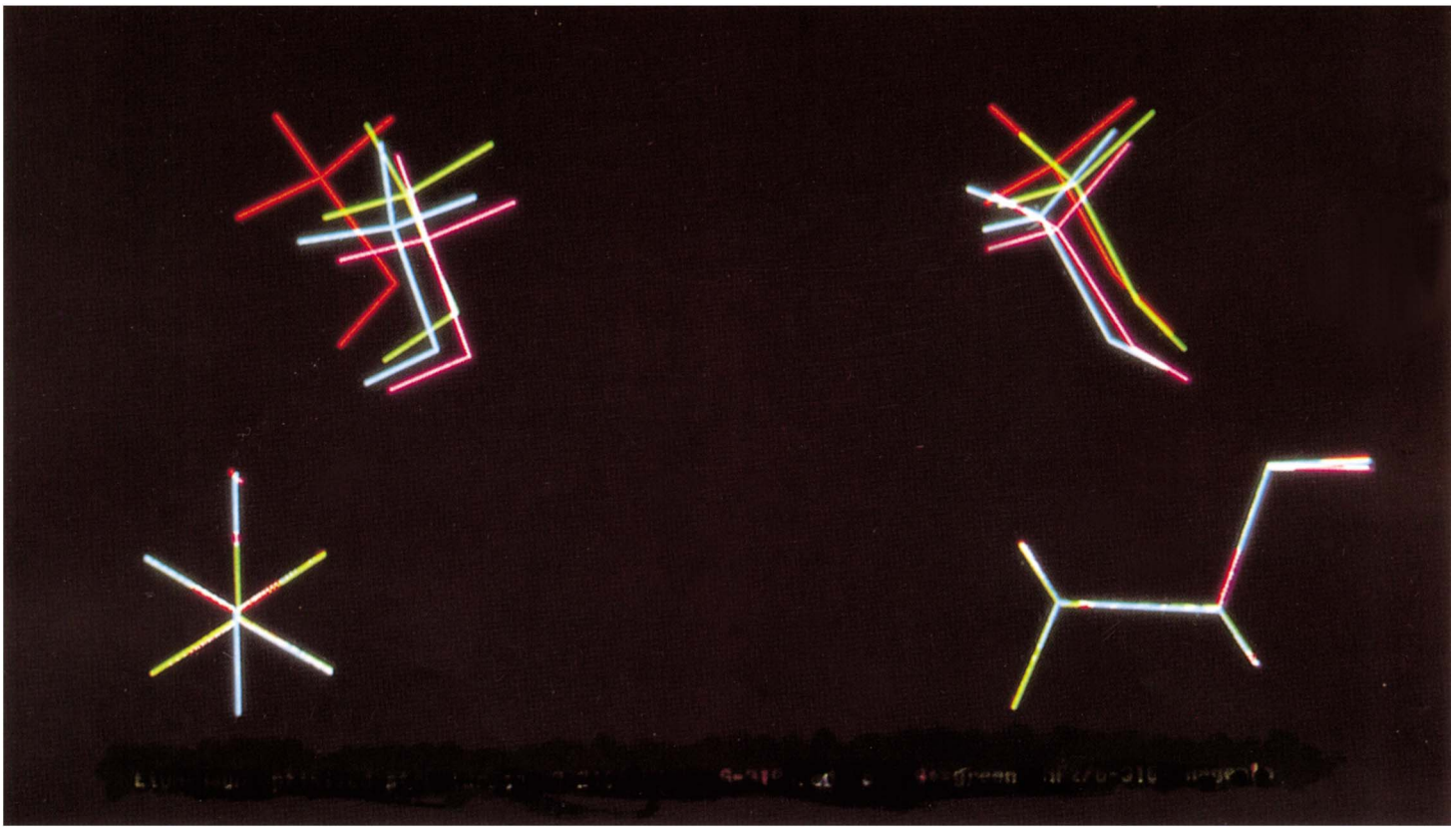

Fig. 5 Orthogonal representation of the superimposed optimized ethanol–methanol complexes using 3-21G (blue), 6-31G (red), 6-31G* (green) and MP2/6-31G* (magenta) basis sets

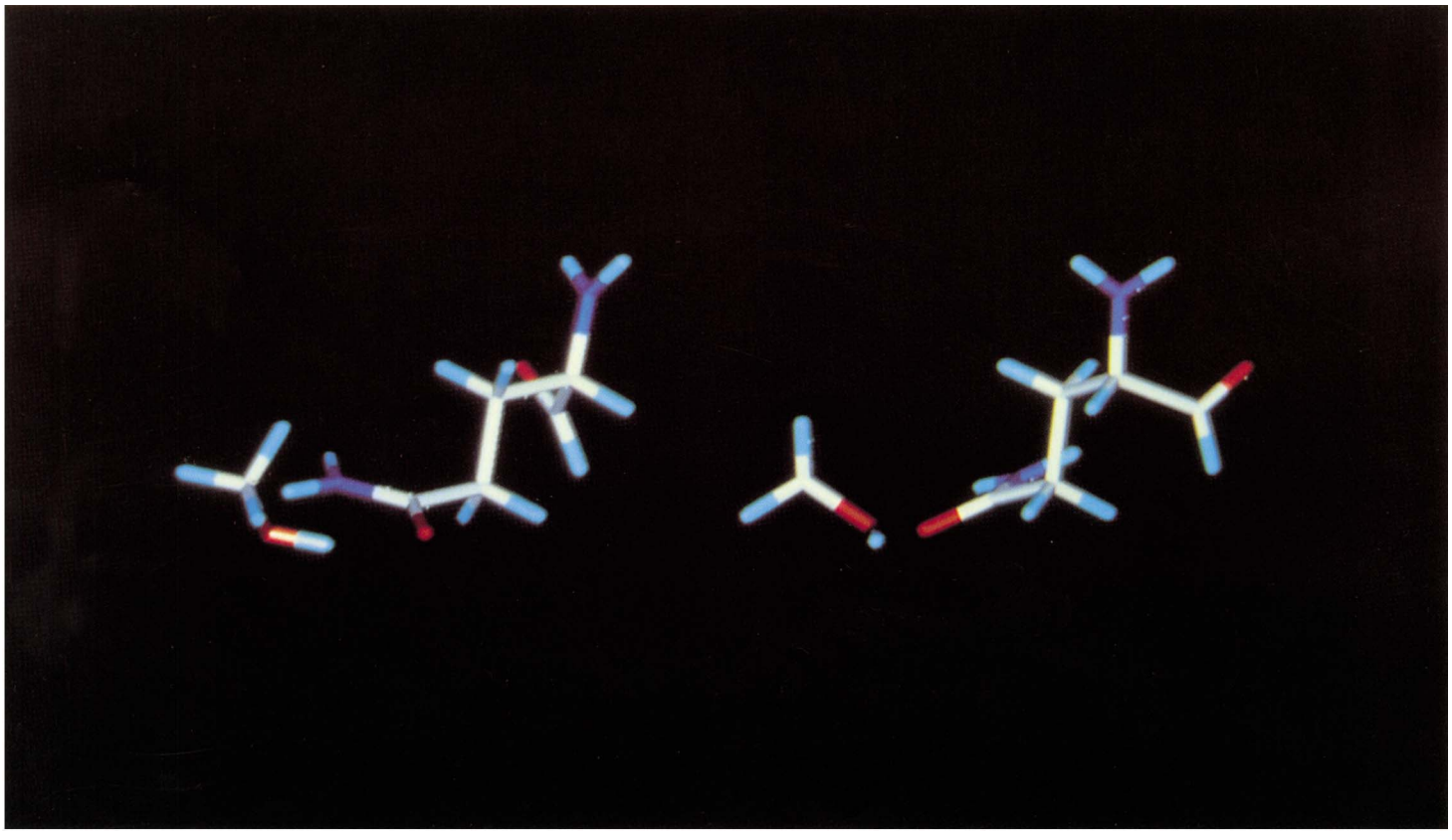

Fig. 6 Orthogonal representation of the optimized (3-21G) glutamine–methanol complex.

plexes are presented. The optimized complexes with selected basis sets are shown in Fig. 5. In the calculated ethanol–methanol complexes methanol is a hydrogen bond donor throughout. These results show that, excluding the STO-3G basis set, the interaction energies at the Hartree–Fock level decrease systematically as more flexible basis sets are used. The O–H and O–O distances between the interacting molecules show a reverse trend as compared to the interaction energies. One reason for this is the obvious lowering in the BSSE. When correlation correction is included the interaction energies increase slightly. Although the range in the interaction energies is as much as 30 kJ mol^{-1} and in the bond lengths 0.2 Å, the optimized interaction geometries in terms of A_0 and T_0 are satisfactorily described by all the basis sets (Table 2 and Fig. 5). The small torsional angles T_0 obtained with the 3-21G and 6-21G basis sets are probably due to the BSSE which stabilises the geometry in which the oxygen of the methanol is close to the methyl group of the ethanol. The differences in A_0 and T_0 cannot be regarded as significant, because on the flat energy surface the optimized geometries are sensitive to the differences in the basis sets. However, it must be pointed out that different basis sets may change even the qualitative features of the interaction energy surfaces.[20] For further discussion on the effects of basis sets see for example ref. 21. The 3-21G and 6-31G basis sets were selected to be used later in the interaction calculations because they give reasonable interaction geometries and are suitable for the comparison of the amino acids and their model compounds.

Table 1 Interaction energies for ethanol–methanol complexes

Basis set	E_{MeOH} [a]	E_{EtOH} [a]	$E_{Complex}$ [a]	E_{Int} [b]
STO-3G	113.549 193	152.132 078	265.690 654	24.6
3-21G	114.398 019	153.222 681	267.640 511	52.0
6-21G	114.872 515	153.868 692	268.760 181	49.8
4-31G	114.571 522	153.855 520	268.740 467	35.2
6-31G	114.988 165	154.013 112	269.013 821	32.9
6-31G*	115.035 418	154.075 609	269.120 352	24.5
6-31 + G*	115.040 965	154.081 497	269.130 870	22.1
MP2/6-31G*	115.353 295	154.528 461	269.896 422	38.5
MP2/6-31 + G*	115.365 289	154.542 000	269.920 596	34.9

[a] Absolute energies in −au. [b] Interaction energies in −kJ mol^{-1}.

Table 2 Interaction parameters for ethanol–methanol complexes[a]

Basis set	R_H/Å	R_O/Å	A_0/°	T_0/°
STO-3G	1.726	2.719	41.5	135.0
3-21G	1.793	2.753	43.3	108.9
6-21G	1.819	2.777	43.2	108.7
4-31G	1.865	2.820	51.9	120.2
6-31G	1.881	2.836	55.3	122.0
6-31G*	2.006	2.951	41.7	122.2
6-31 + G*	2.002	2.952	49.5	137.6
MP2/6-31G*	1.895	2.864	36.0	111.7
MP2/6-31 + G*	1.876	2.853	37.8	123.6

[a] Methanol as a hydrogen bond donor.

Model Compounds for Asparagine and Glutamine.—Amide-group-containing amino acids asparagine (**1d**) and glutamine (**1e**) and the corresponding candidate model compounds formaldehyde (**1a**), acetone (**1b**) and acetamide (**1c**) are presented in Fig. 1. Table 3 lists the total energies of the studied compounds and the interaction energies at the minimum and at the point A, were $A_0 = 90°$ and $T_0 = 180°$. The latter point was selected, because in this geometry the amide substituents do not interfere sterically with the binding of methanol. In Table 4 the interaction parameters of the optimized geometries are presented. The optimized glutamine–methanol complex is shown in Fig. 6.

Studies of the interaction energy surfaces show that the energy surfaces of acetamide and asparagine [Figs. 7(*c*) and (*d*)] resemble each other closely. The interaction energies are within 4 kJ mol^{-1} over the surface and they have two minima which are located in the plane of the C=O group ($A_0 = 90°$) near the direction of the lone-pairs of the carbonyl oxygen ($T_0 = 100°$ and 260°). The T_0 values of the minima deviate from 120° and 240°, which are the directions of the lone-pairs, because of the attraction between the oxygen of the methanol and the hydrogen of the NH_2 and CH_3 groups. The use of bigger basis sets would increase T_0 values as can be deduced from the calculations of the ethanol–methanol dimers.

One possible model compound for asparagine and glutamine not included in the present study is formamide. Calculations for the methanol–formamide complex have been reported by Jasien and Stevens.[22] Their most stable complex was similar to the minimum points reported here. The interaction energy for that complex using the double zeta basis (3,1 contracted valence shell) with compact effective potentials was 45.2 kJ mol^{-1}, which

Table 3 Interaction energies for amide-containing amino acids and the candidate model compounds

		Minimum point		Point A	
Compound	$E_{Compound}$ [a]	$E_{Complex}$ [a]	E_{Int} [b]	$E_{Complex}$ [a]	E_{Int} [b]
Formaldehyde					
3-21G	113.221 820	227.634 129	37.5	227.630 253	27.3
6-31G//3-21G	113.808 367	228.805 042	22.3	228.803 673	18.8
Acetone					
3-21G	190.887 221	305.303 876	48.9	305.297 408	32.0
6-31G//3-21G	191.875 162	306.873 289	26.2	306.871 833	22.3
Acetamide					
3-21G	206.815 803	321.241 014	71.4	321.228 396	38.3
6-31G//3-21G	207.886 440	322.891 364	44.0	322.884 545	26.1
Asparagine					
3-21G	412.419 566	526.846 105	74.9	526.832 517	39.2
6-31G//3-21G	414.554 109	529.560 252	47.2	529.550 954	22.8
Glutamine					
3-21G	451.238 493	565.664 766	74.2	565.651 830	40.2
6-31G//3-21G	453.572 776	568.578 531	46.2	568.571 502	27.7

[a] Absolute energies in −au. [b] Interaction energies in −kJ mol^{-1}.

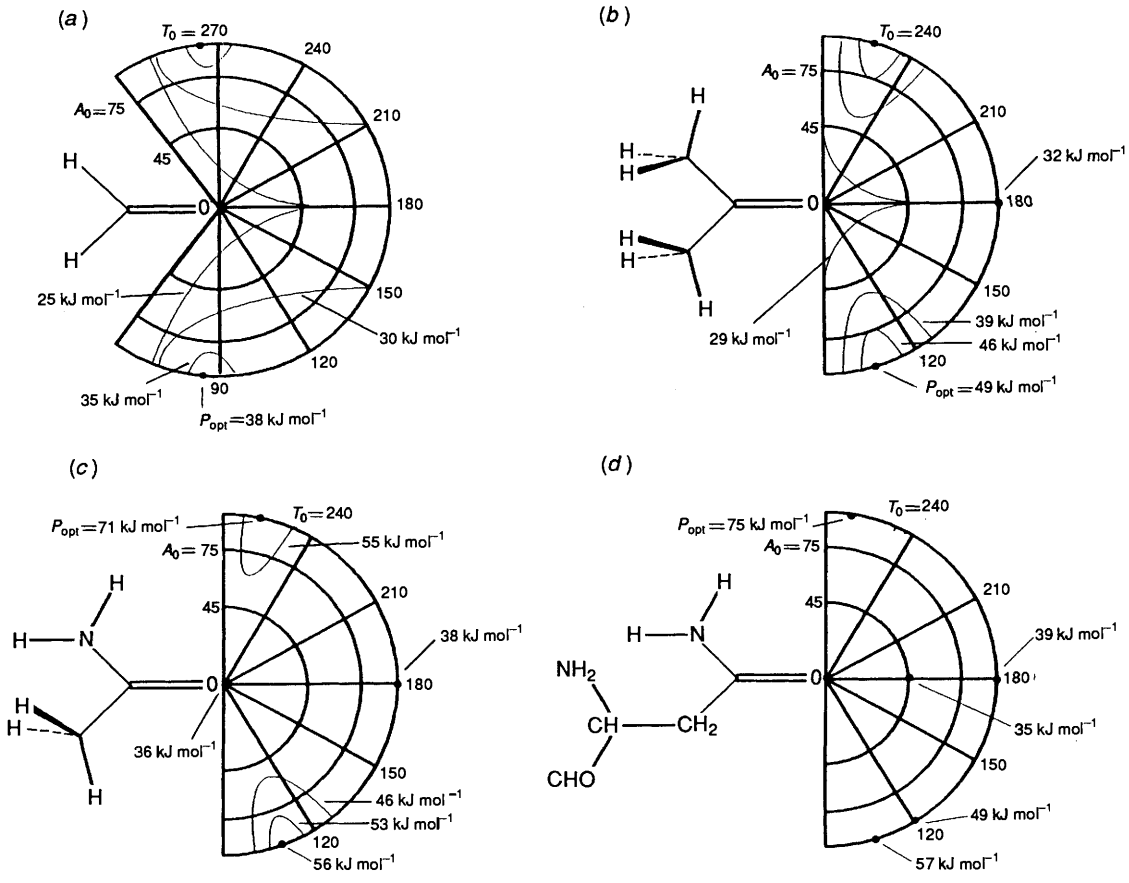

Fig. 7 Computed interaction energy surfaces for the complexes of methanol with (*a*) formaldehyde, (*b*) acetone, (*c*) acetamide and (*d*) asparagine model using 3-21G

Table 4 Interaction parameters for the optimized geometries[a]

Compound	R_H/Å	R_O/Å	A_0/°	T_0/°
Formaldehyde	1.957	2.779	89.6	86.0
Acetone	1.888	2.809	89.7	107.1
Acetamide	1.846	2.736	89.6	97.7
Asparagine	1.845	2.727	89.1	276.6
Glutamine	1.839	2.725	89.6	276.9

[a] Computed with the 3-21G basis set.

is less than 3 kJ mol^{-1} smaller than the energies reported here for acetamide, asparagine and glutamine at the 6-31G//3-21G level. This result indicates that formamide can be used as a model compound near the potential energy minimum, but because of the lack of steric substituent in the C=O group the shape of the potential energy surface around the other minimum would not be satisfactory as can be noted from the interaction surface of formaldehyde [Fig. 7(*a*)]. The minimum of the interaction surface of formaldehyde has shifted away (ΔT_0 = 15°) from the minima of asparagine, glutamine and acetamide. The interaction energy surfaces of formaldehyde and acetone [Figs. 7(*a*) and (*b*)] also lack the deep minimum of amides (71 and 75 kJ mol^{-1}) which is due to the second hydrogen bond between the oxygen of the methanol and the hydrogen of the NH_2 group. The interaction energies at the minimum point and point A, as well as the optimized geometries and atomic charges of asparagine, glutamine and acetamide, are almost identical. Again formaldehyde and acetone are different from the other molecules.

It can also be noted, that in this series the atomic charges manifest nicely the differences between the molecules. The atomic charges of carbonyl oxygen (Fig. 1) can be seen to change in parallel with the interaction energies. For glutamine, points other than the minimum point and point A were not studied. This was not thought to be necessary because asparagine and glutamine behave almost identically at the calculated points. The small difference in the atomic charges of the carbon adjacent to the C=O group is not likely to cause much change to their interaction surfaces.

Hydrogen bonding patterns of carbonyl and amide groups in crystals have been reported in several articles.[23–25] The hydrogen bonding group has been found to be located primarily in the direction of the lone-pairs of carbonyl oxygen (T_0 = 110°–140°) in the plane of the C=O group (A_0 = 70–90°). The hydrogen bonding distances are 2.7–3.0 Å depending on the donor and the surroundings of the carbonyl. All these findings in the crystals are qualitatively reproduced in the present calculations of asparagine, glutamine and acetamide.

On the basis of these results it seems that even the 3-21G basis set can reproduce at least the qualitative features of the hydrogen bonding to the carbonyl group of the amide. Also the suitability of acetamide as a model compound for asparagine and glutamine is obvious.

Model Compounds for Serine, Threonine and Tyrosine.—Hydroxy-group-containing amino acids serine (**2d**), threonine (**2e**) and tyrosine (**3c**) as well as corresponding model compounds methanol (**2a**), ethanol (**2b**), isopropyl alcohol (**2c**), phenol (**3a**) and phenol(spl) (**3b**) are presented in Figs. 2 and 3. In the case of phenol the usefulness of the mixing (splicing) of an extended and minimal basis set in different parts of the molecule to reduce further the size of the model compound was investigated. In this approach the STO-3G basis set was used

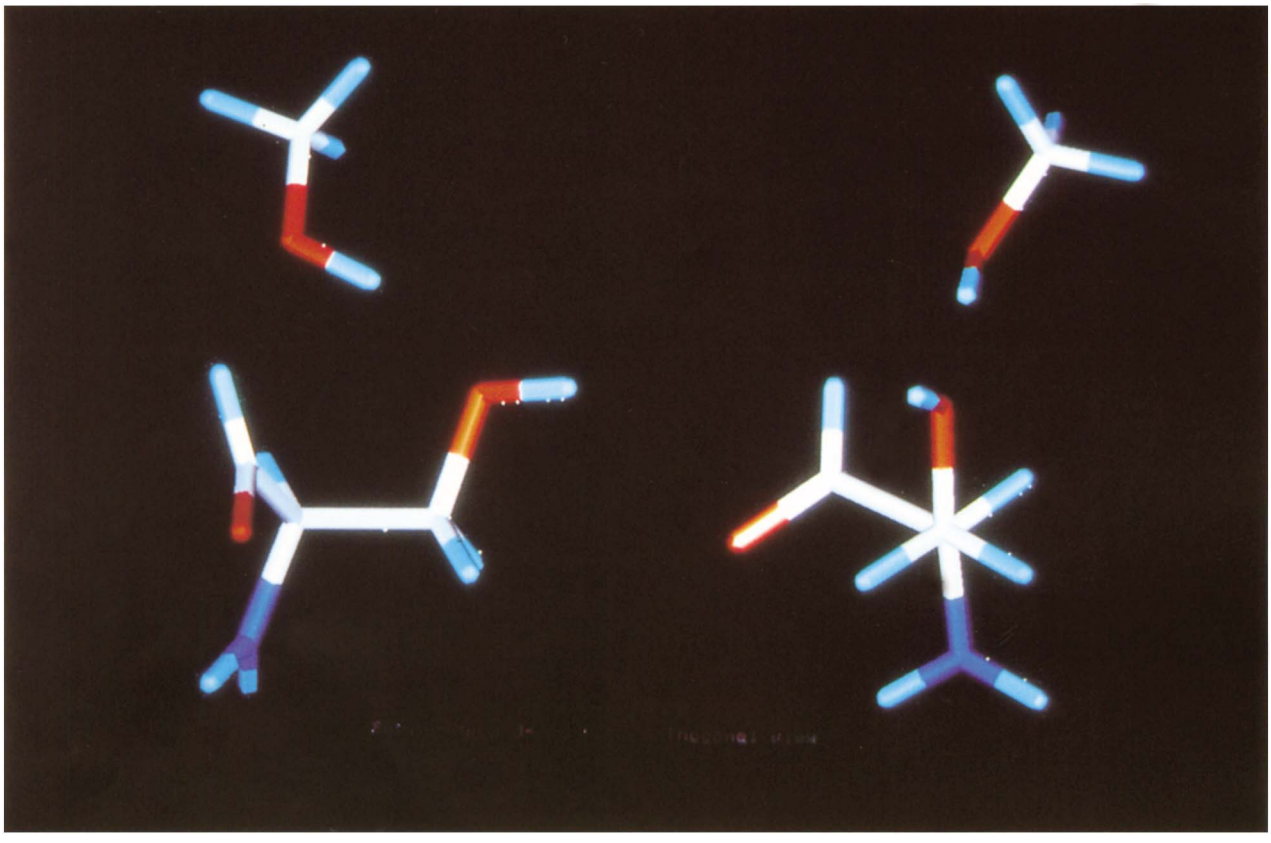

Fig. 8 Orthogonal representation of optimized (3-21G) serine–methanol complex

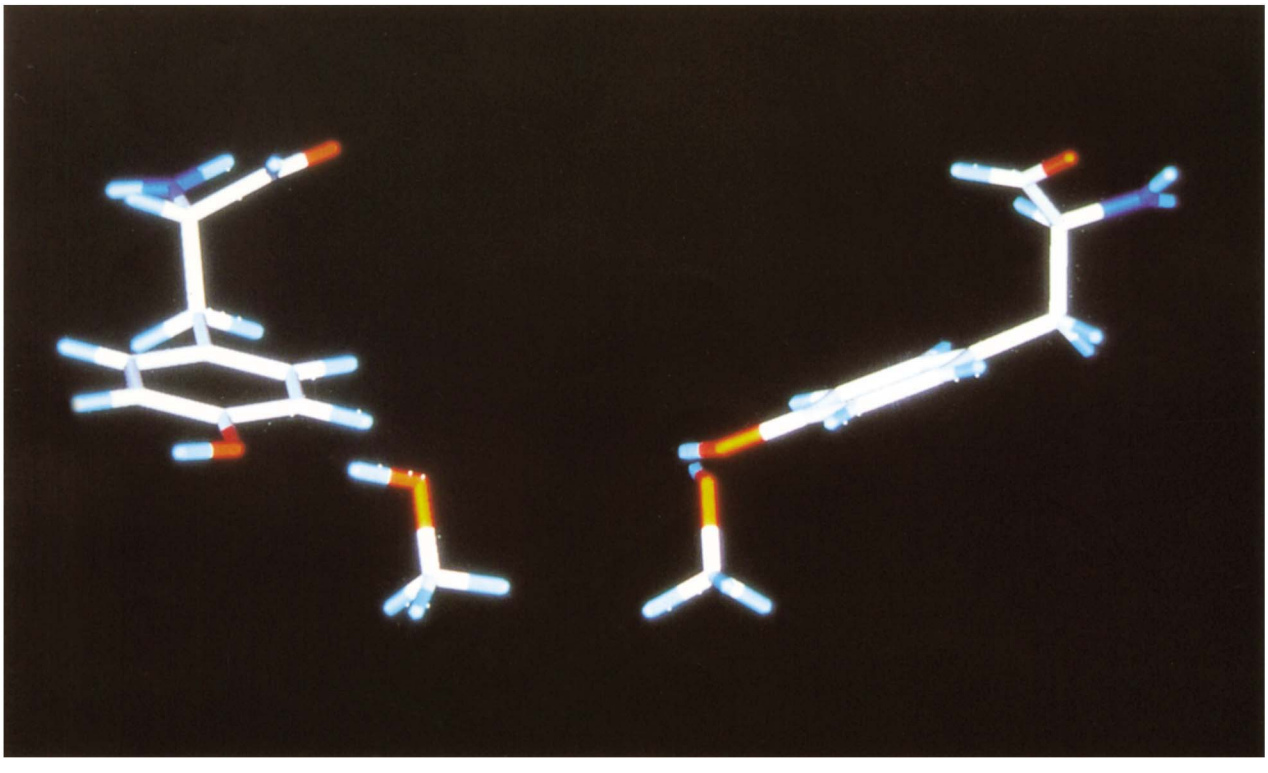

Fig. 9 Orthogonal representation of the optimized (3-21G) tyrosine–methanol complex.

Table 5 Interaction energies for hydroxy-containing amino acids and the candidate model compounds

Compound	$E_{Compound}$ [a]	Minimum point		Point A	
		$E_{Complex}$ [a]	E_{Int} [b]	$E_{Complex}$ [a]	E_{Int} [b]
Methanol					
3-21G	114.398 019	228.813 847	46.8	228.810 306	37.5
6-31G//3-21G	114.988 159	229.986 774	27.4	229.985 306	23.6
Ethanol					
3-21G	143.222 681	267.640 511	52.0	267.634 827	36.7
6-31G//3-21G	154.013 112	269.011 404	26.6	269.010 205	23.4
Serine					
3-21G	320.012 613	434.433 092	59.0	434.423 474	33.7
6-31G//3-21G	321.670 558	436.671 338	33.1	436.665 985	19.1
Isopropyl alcohol					
3-21G	192.048 394	306.465 062	49.3	306.460 473	36.9
6-31G//3-21G	193.036 937	308.034 778	24.2	308.034 322	22.4
Threonine					
3-21G	358.845 033	473.262 042	49.9	473.255 306	32.2
6-31G//3-21G	475.694 572	475.691 643	23.4	375.690 834	21.3
Phenol					
3-21G	303.860 103	418.275 715	46.2	418.270 598	32.8
6-31G//3-21G	305.446 408	420.442 940	22.0	420.442 406	20.5
Phenol(spl)					
3-21G	303.117 854	417.533 701	46.8	417.528 988	34.4
Tyrosine					
3-21G	548.297 463	662.712 970	45.9	662.707 837	32.4
6-31G//3-21G	551.146 164	666.143 020	22.8	666.142 466	21.4

[a] Absolute energies in −au. [b] Interaction energies in −kJ mol^{-1}.

Table 6 Interaction parameters for the optimized geometries[a]

Compound	R_H/Å	R_O/Å	A_0/°	T_0/°
Methanol	1.778	2.691	83.4	92.9
Ethanol	1.793	2.753	43.4	108.9
Serine	1.774	2.714	98.3	110.1
Isopropyl alcohol	1.800	2.745	46.0	111.1
Threonine	1.793	2.708	66.9	127.4
Phenol	1.880	2.802	87.2	106.1
Phenol(spl)	1.851	2.786	87.3	107.7
Tyrosine	1.888	2.807	87.6	105.6

[a] Computed with the 3-21G basis set.

Table 7 Interaction energies and binding geometries for the methanol complexes of phenol and phenol(spl)[a]

		Phenol		Phenol(spl)	
A_0/°	T_0/°	R_O/Å	−E/kJ mol^{-1}	R_O/Å	−E/kJ mol^{-1}
90	90	3.561	26.8	3.595	24.5
90	120	2.775	39.7	2.765	42.3
50	120	2.819	34.2	2.804	37.6
90	150	2.799	33.0	2.800	36.6
50	150	2.833	32.7	2.811	36.4
90	180	2.651	32.7	2.652	34.4

[a] 3-21G was used for phenol, see text for phenol(spl).

for the three carbon and hydrogen atoms furthest away from the OH group and the 3-21G basis set was used for the other atoms. This system is denoted as phenol(spl). Mixing of basis sets has been used in earlier studies on cyclohexa-1,2-diene,[26] proton transfer reactions[27] and on the molecular properties of several test molecules.[28] Table 5 lists the total energies of the OH-group-containing molecules and the interaction energies at the minimum point and point A. In Table 6 the interaction parameters of the optimized geometries are presented. The optimized serine–methanol complex is shown in Fig. 8 and the optimized tyrosine–methanol complex in Fig. 9.

Studies of the interaction energy surfaces of ethanol, isopropyl alcohol and serine show [Figs. 10(*b*), (*c*) and (*d*)] that the energy differences between the molecules over the studied surfaces are about 5 kJ mol^{-1}. Methanol [Fig. 10(*a*)] has the minimum point (T_0 = 92.9°, A_0 = 83.4°) in the area where ethanol has a less favourable binding due to the steric effects of the methyl group. Although the locations of the minimum binding sites differ between the aliphatic alcohols, apart from methanol they are all in the direction of the lone-pairs of the sp^3-hybridized oxygen (T_0 = 109–130°). The differences between the minimum binding sites of the studied aliphatic alcohols are partly due to the flatness of the interaction energy surfaces. Instead the differences in the energies of the optimized geometries and at the point A and differences in atomic charges (Fig. 2) are small. On the basis of these results it can be concluded that ethanol is a suitable model compound for serine and threonine.

From the comparison of the optimized minimum point geometries, interaction energies and atomic charges (Fig. 3) of phenol, phenol(spl) and tyrosine it can be noted, that both phenol and phenol(spl) model tyrosine very well. In contrast to the flat interaction energy surfaces of aliphatic alcohols, the aromatic ones have well defined minimum binding sites in the direction of the lone-pairs of oxygen in the plane of the aromatic ring (indicative of sp^2-hybridization) [Fig. 10(*e*)]. The effect of the basis set splicing was further studied by calculating more points on the interaction energy surfaces of phenol and phenol(spl). The energy values of these points are presented in

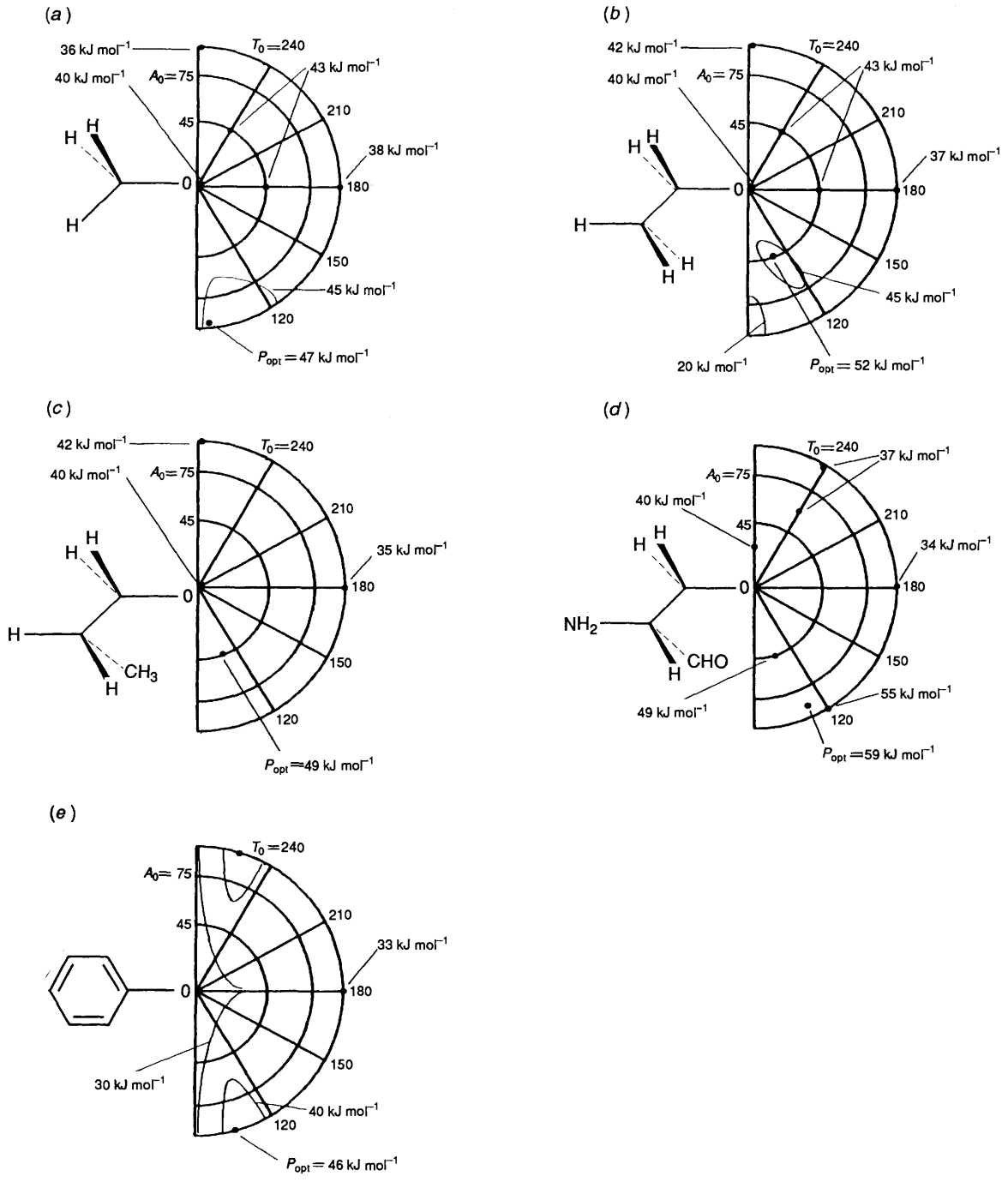

Fig. 10 Computed interaction energy surfaces for the complexes of methanol with (*a*) methanol, (*b*) ethanol, (*c*) isopropyl alcohol, (*d*) serine model and (*e*) phenol using 3-21G

Table 7. The values of phenol are 2–4 kJ mol^{-1} lower than those of phenol(spl). The exception is the point $A_0 = 90°$, $T_0 = 90°$, at which the oxygen of methanol is in close contact with the aromatic ring, where the effect of the STO-3G basis set is appreciable. In this position the distance between O_{Phe} and O_{Met} is 3.6 Å. This comparison shows that, in the interaction studies, basis set splicing is a technique which can be used with only minor effects on the interaction energies, geometries and on the shape of the interaction energy surfaces. Other studies have also shown that basis set splicing can be done when more extended basis sets are used.[28]

Crystal structure studies of hydrogen bonding to hydroxy group have shown that there is a wide distribution of bonding around the oxygen of serine and threonine.[24,25] The site of the most favourable binding is in the direction of the lone-pairs of the oxygen. The O–O distances in hydrogen bonding are 2.8–2.9 Å. Hydrogen bonding to tyrosine has preferred regions in the direction of the lone-pairs ($T_0 = 120°$) of oxygen near the plane of the aromatic ring ($A_0 = 90°$). These findings in crystal structures are in accordance with the calculations presented here.

Conclusions

The small model compounds for the amino acids to be used in the quantum mechanical receptor–ligand interaction calculations were generated in the present study using mainly 3-21G and 6-31G basis sets and methanol as a probe molecule. The investigated amino acids were asparagine, glutamine, serine, threonine and tyrosine.

Acetamide reproduced well the interaction energy surfaces of asparagine and glutamine. The other candidate model compounds were formaldehyde and acetone. Ethanol was found to model satisfactorily serine and threonine. Ethanol reproduced well the magnitude of the interaction and the general features of the interaction energy surfaces. The minima of the interaction energy surfaces differ between the aliphatic alcohols partly due to the flatness of the surfaces. Phenol was found to be a good model compound for tyrosine. Also the suitability of the basis set splicing technique was investigated in the case of phenol. The spliced system gave a similar interaction energy surface to that of phenol, although the interaction energies of the spliced system were uniformly slightly higher. It seems obvious, that the basis set splicing technique can also be used to model other amino acids with aromatic ring structures. All the calculated interaction energy surfaces of the amino acids and their model compounds reproduced qualitatively the general features of the hydrogen bonding found in the crystal structures.

In conclusion, the present results indicate that it is possible to reduce notably the size of the receptor–ligand system and at the same time to preserve its properties by using well selected model compounds instead of entire amino acids. However, before the receptor–ligand systems can be studied using this model assembly approach, the reliability of the approach for the larger systems should be investigated.

References

1 C. L. Brooks III, M. Karplus and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, John Wiley and Sons, New York, 1988.

2 C. A. Reynolds and P. M. King, in *Computer-Aided Molecular Design*, ed. W. G. Richards, IBC Technical Services Ltd, Eynsham, 1989, p. 51.

3 (*a*) D. M. Hayes and P. A. Kollman, *J. Am. Chem. Soc.*, 1976, **98**, 3335; (*b*) D. M. Hayes and P. A. Kollman, *J. Am. Chem. Soc.*, 1976, **98**, 7811.

4 B. Waszkowycz, I. H. Hillier, N. Gensmantel and D. W. Payling, *J. Chem. Soc., Perkin Trans. 2*, 1991, 225.

5 B. Waszkowycz, I. H. Hillier, N. Gensmantel and D. W. Payling, *J. Chem. Soc., Perkin Trans. 2*, 1991, 2025.

6 (*a*) M. Sola, A. Lledos, M. Duran, J. Bertran and J.-L. M. Abboud, *J. Am. Chem. Soc.*, 1991, **113**, 2873; (*b*) M. Sola, A. Lledos, M. Duran and J. Bertran, *J. Am. Chem. Soc.*, 1992, **114**, 869.

7 (*a*) O. Tapia and G. Johannin, *J. Chem. Phys.*, 1981, **75**, 3624; (*b*) J. A. C. Rullman, M. N. Bellido and P. T. van Duijnen, *J. Mol. Biol.*, 1989, **206**, 101.

8 (*a*) S. J. Weiner, G. L. Seibel and P. A. Kollman, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 649; (*b*) U. C. Sighn and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718.

9 (*a*) G. Alagona, P. Desmeules, C. Ghio and P. A. Kollman, *J. Am. Chem. Soc.*, 1984, **106**, 3623; (*b*) G. Bolis, M. Ragazzi, D. R. Ferro and E. Clementi, *Gazz. Chim. Ital.*, 1978, **108**, 425; (*c*) B. Waszkowycz, I. H. Hillier, N. Gensmantel and D. W. Payling, *J. Chem. Soc., Perkin Trans. 2*, 1989, 1795; (*d*) B. Waszkowycz, I. H. Hillier, N. Gensmantel and D. W. Payling, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1259; (*e*) B. Waszkowycz, I. H. Hillier, N. Gensmantel and D. W. Payling, *J. Chem. Soc., Perkin Trans. 2*, 1991, 1819.

10 (*a*) V. Daggett, S. Schröder and P. Kollman, *J. Mol. Biol.*, 1976, **103**, 227; (*b*) S. Schröder, V. Daggett and P. Kollman, *J. Am. Chem. Soc.*, 1991, **113**, 8926.

11 (*a*) A. Warshel and R. M. Weiss, *J. Am. Chem. Soc.*, 1980, **102**, 6218; (*b*) A. Warshel, S. Russell and F. Sussman, *Isr. J. Chem.*, 1986, **27**, 217; (*c*) J. Åqvist and A. Warshel, *Biochemistry*, 1989, **28**, 4680.

12 T. Clark, *A Handbook of Computational Chemistry*, John Wiley and Sons, New York, 1985, pp. 145–151.

13 (*a*) S. Schröder, V. Daggett and P. Kollman, *J. Am. Chem. Soc.*, 1991, **113**, 8922; (*b*) H. S. Rzepa and M. Y. Yi, *J. Chem. Soc., Perkin Trans. 2*, 1991, 531.

14 J. A. Sordo, S. Chin and T. L. Sordo, *Theor. Chim. Acta*, 1988, **74**, 101.

15 Chem-X, Chemical Design Ltd, Oxford, England.

16 P. K. Weiner and P. Kollman, *J. Comput. Chem.*, 1981, **2**, 287.

17 R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833.

18 Gaussian 88, M. J. Frisch, M. Head-Gordon, H. R. Schleger, K. Raghavachari, J. S. Binkley, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, R. L. Kahn, J. J. P. Stewart, E. M. Fluder, S. Topiol and J. A. Pople, Gaussian Inc., Pittsburg PA.

19 Gaussian 90, M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. R. Schleger, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzales, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, R. L. Kahn, J. J. P. Stewart, S. Topiol and J. A. Pople, Gaussian Inc., Pittsburg PA.

20 (*a*) R. A. Kumpf and J. R. Damewood, Jr., *J. Phys. Chem.*, 1989, **93**, 4478; (*b*) J. R. Damewood, Jr. and R. A. Kumpf, *J. Phys. Chem.*, 1987, **91**, 3449.

21 (*a*) M. J. Frisch, J. E. Del Bene, J. S. Binkley and H. F. Schaefer III, *J. Chem. Phys.*, 1986, **84**, 2279; (*b*) J. A. Sordo, T. L. Sordo, G. M. Fernandes, R. Gomperts, S. Chin and E. Clementi, *J. Chem. Phys.*, 1989, **90**, 6361.

22 P. G. Jasien and W. J. Stevens, *J. Chem. Phys.*, 1985, **84**, 3271.

23 (*a*) R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1983, **105**, 5761; (*b*) R. Taylor, O. Kennard and W. Versichel, *J. Am. Chem. Soc.*, 1984, **106**, 244; (*c*) P. Murray-Rust and J. P. Glusker, *J. Am. Chem. Soc.*, 1984, **106**, 1018; (*d*) R. Taylor, O. Kennard and W. Versichel, *Acta Crystallogr., Sect. B*, 1984, **40**, 280; (*e*) R. Taylor, O. Kennard and W. Versichel, *Acta Crystallogr., Sect. B*, 1984, **40**, 280; (*f*) A. Vedani and J. D. Dunitz, *J. Am. Chem. Soc.*, 1985, **107**, 7653.

24 (*a*) C. Ramakrishnan and N. Prasad, *Int. J. Pept. Protein Res.*, 1971, **3**, 209; (*b*) J. Mitra and C. Ramakrishnan, *Int. J. Pept. Protein Res.*, 1977, **9**, 27; (*c*) S. N. Vinogradov, *Int. J. Pept. Protein Res.*, 1979, **14**, 281; (*d*) N. Thanki, J. M. Thornton and J. M. Goodfellow, *J. Mol. Biol.*, 1988, **202**, 637.

25 M. Tintelnot and P. Andrews, *J. Comp. Aided Mol. Design*, 1989, **3**, 67.

26 M. W. Schmidt, R. O. Angus, Jr. and R. P. Johnson, *J. Am. Chem. Soc.*, 1982, **104**, 6838.

27 L. Pardo, A. P. Mazurek and R. Osman, *Int. J. Quantum Chem.*, 1990, **37**, 701.

28 J. H. Jensen and M. S. Gordon, *J. Comput. Chem.*, 1991, **12**, 421.

Paper 2/02783F
Received 28*th May* 1992
Accepted 20*th August* 1992