Journal of Computer-Aided Molecular Design **18:** 421–431, 2004.

Tuning of hydrogen bond strength using substituents on phenol and aniline: A possible ligand design strategy

Jóhannes Reynisson* & Edward McDonald
Cancer Research UK Centre For Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

Received 13 May 2004; accepted in revised form 12 September 2004

Key words: density functional theory (DFT), drug design, hydrogen bonding energy, ligand–protein interactions

Summary

Using Density Functional Theory, the hydrogen bonding energy is calculated for the interaction of phenol and aniline with four model compounds representing the protein backbone and various amino acid side chain residues. The models are methanol, protonated methylamine, formaldehyde and acetate anion. The H-bond energies for the uncharged species are ~2.5 kcal mol^{-1}, whereas the charged model compounds bind with much higher energies of ~20 kcal mol^{-1}. The effect of *para*-substitution on the hydrogen bond energies is determined. Substitution has little effect on the H-bond energy of the neutral complexes (< 2 kcal mol^{-1}), but for the positively and negatively charged systems substitution drastically alters the binding energies, e.g., 14.3 kcal mol^{-1} for *para*-NO_2. In the context of protein–ligand binding, relatively small changes in binding energy can cause large changes in affinity due to their exponential relationship. This means that for $-NO_2$ an enormous change of 10 orders of magnitude for the affinity constant is predicted. These calculations allow prediction of H-bonds, using different substituents, in order to fine-tune and optimize ligand–protein interactions in the search for drug candidates.

Introduction

Hydrogen bonding in chemical biology is of paramount importance. An important early example was the proposal by Pauling, that the secondary structure of proteins is determined by H-bonding between backbone amino acid residues. This concept led him to predict that α-helixes and β-sheets would be prominent structures in proteins [1, 2]. H-bonding is also a significant factor when ligands bind to proteins and is therefore an important aspect of structure-based design of drugs [3–6].

One way to increase affinity is to modify the structure of a ligand by adding extra functional groups that can form *additional* interaction sites, e.g., by H-bonding [8]. However, it has been found, using the isothermal titration calorimetry method (ICT), that when modifying the structures of an interacting ligand–protein system, increased bonding (favourable enthalpy change, ΔH) tends to be offset by increased order (negative entropy change, ΔS), leading to a relatively small change in Gibbs free energy (ΔG), and therefore in the affinity equilibrium dissociation constant (K_d) [9, 10]. The relation between these properties is given by Equations 1 and 2:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

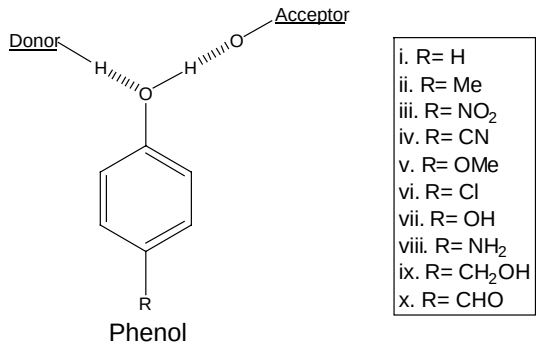
$$K_d = e^{-\Delta G/RT} \quad (2)$$

An alternative approach to increase affinity would be to strengthen an existing H-bond without significantly altering the degrees of freedom of

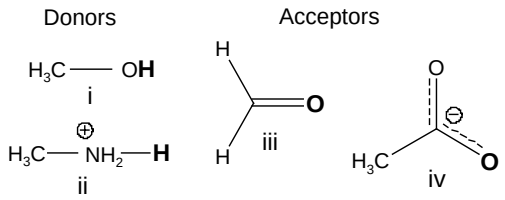
*To whom correspondence should be addressed.
E-mail: Johannes.Reynisson@icr.ac.uk

the system. This would result in an increased enthalpy with minimal change in entropy. Adding electron donating or withdrawing substituents to a ligand in the vicinity of the pertinent hydrogen bond, thereby altering the electron density of the hydrogen-binding group, can lead to changes in H-bond energy. Abraham and Platt investigated this phenomenon in a systematic way [11] and small systems have been calculated with *ab initio* methods [12, 13]. The typical ΔG for a ligand–protein complex lies between −7 and −11 kcal mol^{-1} [9] and relatively small modifications to it can lead to large changes in affinity (K_d) due to the exponential relationship between the two (see Equation 2). A change of only 1.4 kcal mol^{-1} leads to a tenfold increase or decrease in the affinity of a ligand at physiological temperature. It is therefore pertinent to investigate the effect of substituents on hydrogen bonding energy to find out whether this could be an attractive design strategy for molecular modelers. In the current work the H-bond energies were calculated using the density functional theory (DFT). This method has previously been proven successful in calculating various properties of organic molecules including hydrogen bond energies (see for example Refs. 14–20 and references therein). Limitations to the DFT method exist [21] but many research groups are working to refine the method (e.g., Refs. 22 and 23).

As a simple model for investigating this effect, phenol and aniline were used; hydrogen bonded either with formaldehyde and acetate anion acting as hydrogen bond acceptors. In addition, methanol and protonated methylamine ($CH_3N^+H_3$) were used as hydrogen bond donors. Phenol and aniline are substituted on the *para*-positions as shown in Scheme 1 for the former. The model donor/acceptor systems are depicted in Scheme 2.



Scheme 1. Phenol and its derivatives used in this work as hydrogen bond donors and acceptors with small model molecules. Aniline is used in an analogous way.



Scheme 2. The small model donor/acceptor molecules. They are methanol (i), protonated methylamine (ii), formaldehyde (iii) and acetate anion (iv). The H-bonding sites are depicted in bold.

Computational methods

All the energy calculations and geometry optimisations were performed with the GAUSSIAN 03 program package [24] utilising restricted DFT. The Lee, Yang and Parr functional was used for the correlation part [25] and the combined Becke's parameter functional for the exchange part (B3LYP) [26, 27]. The basic split valence standard 6-31 + G(d,p) basis set, with a diffuse function [28], was used for the geometry optimisation and frequency analysis. In all cases the normal modes of the molecular motions revealed no imaginary frequencies for the calculated structures, which is in support of them representing a minimum on the potential energy surface. The zero-point vibrational energies (ZPE) were scaled according to Wong (0.9804) [29]. Subsequent single-point electronic energy calculations were performed again at the B3LYP level with Pople's triply split valence and polarised 6-311 + G(2df,2pd) basis set. The basis set superposition error (BSSE) was calculated and taken into account by using the Counterpoise Theory [30, 31].

Results and discussion

The results of the single-point electronic energy calculations and ZPEs of the calculated structures of the H-bonded pairs and their subunits are listed in Tables S1–S8 (see supplementary information at the end of this paper). Also, the BSSE corrections are given in Tables S1–S8.

Formaldehyde was chosen as a model for a hydrogen bond acceptor because it approximates the carbonyl groups in proteins, e.g. the amide carbonyl in the backbone and in the side chains of asparagine and glutamine amino acids. The acetate anion was used as a model for the aspartic and glutamic amino acids. Methanol was utilized as a H-bond donor, simulating serine and threonine amino acids and, to some extent, tyrosine. Finally protonated methylamine was used for lysine [32]. Phenol and aniline were chosen because they often are components of drug molecules [5]. The model compounds presented here are small but should be adequate to theoretically reflect the situation for ligand–protein interactions.

In this work we have chosen larger basis sets than the popular 6-31*G*(d,p)/6-311(2df,p) combination, i.e. the 6-31 + *G*(d,p)/6-311 + *G*(2df,2pd) combination. Adding a diffuse function makes a more thorough representation of the hydrogen bonds. Furthermore, for the energy calculations, more orbitals are calculated for the hydrogen atoms for the same purpose.

Hydrogen bonding energies

Hydrogen bonding energy is defined as the difference between the energy of the fully optimized H-bonded pair and the sum of the energies of the two individually optimized components making up the pair, with ZPE corrections taken into account (e.g. for the phenol-formaldehyde pair: Hydrogen bond energy = E(phenol-formaldehyde) − E(phenol) − E(formaldehyde)). This value is therefore an estimate of the interaction energy of the components. The calculated hydrogen bond energies of *para*-substituted phenol and aniline with the small model compounds as derived from the energies shown in Tables S1–S8 and corrected for their BSSE, are reported in Tables 1–4. The tables also depict the change (Δ) that is caused by the substituent relative to their pertinent unsubstituted (-H) systems. Furthermore, the Δ-values are used in Equation 3 to generate the effect on K_d. This is justified by the fact that any change in the H-bonding energy is directly transmitted to the ΔH, resulting in a changed enthalpy ($\Delta\Delta H$), which again changes the Gibbs free energy ($\Delta\Delta G$) according to Equations 1 and 2, resulting in Equation 3 [33]:

$$\Delta K_d = e^{(\Delta\Delta G/RT)} \quad (3)$$

The ΔK_d gives the ratio of equilibrium constants, i.e., as mentioned earlier 1.4 kcal mol^{-1} results in a 10-fold increase in K_d and 2.7 kcal mol^{-1} in a 100-fold increase, etc.

In Table 1 for the phenol–formaldehyde system the H-bond energies lie between 3.5 and 5.5 kcal mol^{-1} and for aniline–formaldehyde they lie between 1.1 and 3.4 kcal mol^{-1}. Aniline is a poorer H-bond donor than phenol, which is consistent with nitrogen having a lower electron affinity than oxygen.

In Table 2 for the phenol–methanol system the H-bond energies lie between 1.7 and 3.2 kcal mol^{-1}

Table 1. The H-bonding energies and their change (Δ) from the un-substituted system in kcal mol^{-1} for the *para*-substituted phenol/aniline-formaldehyde systems. Also, the effect on the ratio of the equilibrium constants (K_d) is given according to Equation 3.

Para-substituents	Phenol			Aniline		
	H-bond	Δ	Δ K_d	H-bond	Δ	Δ K_d
–H	3.9	0.0	1.0	2.3	0.0	1.0
–Me	3.7	−0.1	0.8	1.8	−0.5	0.4
–NO_2	5.5	1.7	16.5	3.4	1.1	6.2
–CN	5.1	1.2	7.4	2.8	0.6	2.6
–OMe	3.8	0.0	0.9	1.6	−0.7	0.3
–Cl	4.3	0.4	2.0	2.0	−0.3	0.6
–OH	3.6	−0.3	0.6	1.1	−1.2	0.1
–NH_2	3.5	−0.4	0.5	1.6	−0.6	0.3
–CH_2OH	3.9	0.0	1.0	2.0	−0.3	0.6
–CHO	4.8	1.0	5.1	2.7	0.4	2.1

Table 2. The H-bonding energies and their change from the un-substituted system (Δ) in kcal mol^{-1} for the *para*-substituted phenol/aniline-methanol systems. Also, the effect on the ΔK_d is given according to Equation 3.

Para-substituents	Phenol			Aniline		
	H-bond	Δ	Δ K_d	H-bond	Δ	Δ K_d
–H	2.2	0.0	1.0	3.4	0.0	1.0
–Me	2.3	0.1	1.2	3.3	−0.1	0.8
–NO_2	1.9	−0.3	0.6	1.6	−1.8	4.8×10^{-2}
–CN	1.7	−0.4	0.5	2.0	−1.5	7.9×10^{-2}
–OMe	2.5	0.3	1.7	3.4	0.0	1.0
–Cl	2.0	−0.1	0.8	2.8	−0.7	0.3
–OH	3.2	1.0	5.8	2.7	−0.7	0.3
–NH_2	2.7	0.5	2.3	4.0	0.6	2.8
–CH_2OH	2.2	0.0	1.0	3.1	−0.4	0.5
–CHO	1.7	−0.5	0.4	2.0	−1.5	0.1

Table 3. The H-bonding energies and their change from the un-substituted system (Δ) in kcal mol^{-1} for the *para*-substituted phenol/aniline-acetate anion systems. Also, the effect on the ratio of the equilibrium constants (K_d) is given according to Equation 3.

Para-substituents	Phenol			Aniline		
	H-bond	Δ	Δ K_d	H-bond	Δ	Δ K_d
–H	25.4	0.0	1.0	17.9	0.0	1.0
–Me	24.7	−0.7	0.3	17.3	−0.6	0.4
–NO_2	39.6	14.3	2.9×10^{10}	30.4	12.5	1.5×10^{9}
–CN	36.6	11.2	1.6×10^{8}	27.5	9.6	1.1×10^{7}
–OMe	24.3	−1.1	0.2	16.7	−1.3	0.1
–Cl	29.6	4.2	117.4	21.4	3.5	365.7
–OH	24.7	−0.7	0.3	16.5	−1.4	0.1
–NH_2	22.9	−2.5	1.5×10^{-2}	15.3	−2.6	1.2×10^{-2}
–CH_2OH	26.2	0.8	4.0	18.7	0.8	3.6
–CHO	35.6	10.2	2.9×10^{7}	26.7	8.8	2.9×10^{6}

Table 4. The H-bonding energies and their change from the un-substituted system (Δ) in kcal mol^{-1} for the *para*-substituted phenol/aniline-protonated methylamine systems. Also, the effect on the ΔK_d is given according to Equation 3.

Para-substituents	Phenol			Aniline		
	H-bond	Δ	Δ K_d	H-bond	Δ	Δ K_d
–H	16.3	0.0	1.0	20.5	0.0	1.0
–Me	17.7	1.4	11.1	22.4	1.9	24.5
–NO_2	8.1	−8.2	9.8×10^{-7}	10.7	−9.9	5.9×10^{-8}
–CN	9.0	−7.2	4.9×10^{-6}	12.2	−8.4	7.3×10^{-7}
–OMe	18.7	2.4	62.5	23.5	2.9	137.5
–Cl	13.8	−2.5	0.1	18.0	−2.6	1.3×10^{-2}
–OH	17.6	1.3	9.0	21.6	1.1	6.2
–NH_2	20.7	4.5	1.8×10^{3}	25.8	5.3	7.7×10^{3}
–CH_2OH	18.2	1.9	24.5	22.6	2.1	33.2
–CHO	10.9	−5.4	1.1×10^{-4}	14.2	−6.4	2.1×10^{-5}

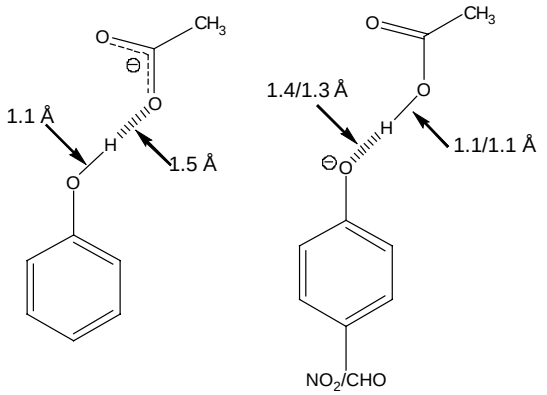
and for aniline–methanol they lie between 1.6 and 4.0 kcal mol^{-1}. According to Table 2 phenol and aniline are similar H-bond acceptors, in contrast to their H-bond donor abilities.

In Table 3 for the phenol–acetate anion system the H-bond energies lie between 22.9 and 39.6 kcal mol^{-1} and for aniline-acetate they lie between 15.3 and 30.4 kcal mol^{-1}. These are significantly larger H-bond energies than found for the uncharged systems. As before (Table 1) phenol has better H-bond donating abilities than aniline.

In Table 4 for the phenol–protonated methylamine system the H-bond energies lie between 8.1 and 20.7 kcal mol^{-1} and for the aniline they lie between 10.7 and 25.8 kcal mol^{-1}. Aniline binds better than phenol to protonated methylamine, consistent with oxygen being more electrophilic than nitrogen. In the case of methanol (Table 2), the H-binding energy is low both for phenol and aniline, which masks this phenomenon. Usually, charged systems have higher H-bonding energies than their neutral counterparts [34].

The H-bond lengths are ~2.0 Å for the phenol/aniline–methanol complexes. The phenol–formaldehyde system has a H-bond length of ~1.9 Å and its aniline counterpart ~2.2 Å. In all cases the H-bond lengths are slightly increased and decreased by ~0.05 Å depending on the H-bond strength. The covalent bonds in O–H and N–H involved in H-bonding were calculated to be ~1.0 Å and did not significantly change when new substituents were added to the rings. The H-bond lengths for the complexes containing CH_3–N^+H_3 lie mostly around 1.7 Å, depending on the substituents. The covalent bonds ($CH_3N^+H_2$ –H) are in the region of 1.1 Å. In the case of the aniline-acetate anion the H-bond lengths lie between 1.7 and 1.8 Å and the pertinent covalent bonds are again ~1.1 Å. For the phenol–acetate anion system the H-bonds lie between 1.3 and 1.5 Å except for the $–NO_2$ and –CHO derivatives [35]. For these the proton shifts from the substituted phenol towards the acetate anion, as seen in Scheme 3.

In a similar case, a proton shift was observed upon oxidation of H-bonded complexes of violuric acid, 1-hydroxybenzotriazole, N-hydroxyacetanilide and 2-aminopurine with imidazole [36]. A further example is the oxidation of the guanine nucleotide, which leads to a drastic lowering of the pK_a value compared to its parent compound [37].



Scheme 3. The phenol–acetate anion system. The proton shifts from the phenolic moiety towards the acidic anion for the $–NO_2$ and –CHO substituted phenol systems compared to the unsubstituted complex.

When oxidised, whilst paired with cytosine, the N1 proton of guanine is shifted to the cytosine [38]. It can be argued that the substitution of strongly electron withdrawing groups simulates, to some extent, the removal of an electron from the pair shown in Scheme 3.

Correlation with the Hammett constants

The classical way of evaluating the effects of substituents is to use the well known Hammett constants (see for example Ref. 39). In all the cases discussed here a linear regression is observed when σ is plotted against the Δ-values. This is depicted for the *para*-substituted phenol–formaldehyde system in Figure 1.

The slopes of the Hammett plots, i.e. the reaction constant ρ, are listed in Table 5 with the statistical errors for the slopes. The constant ρ is a measure of the sensitivity of the system to the effects of electronic perturbation. The good linearity of these correlations serves to validate the DFT calculations, since the Hammett constants are derived from experimental data [40].

From Table 5 it is apparent that H-bond energy increases with electron donating groups for the species complexed with formaldehyde and the acetate anion. Furthermore, a decrease is observed in the case of methanol and protonated methylamine complexes. The changes are much more pronounced with the charged species. In general, charged chemical species form stronger hydrogen bonds than their neutral counterparts [34].

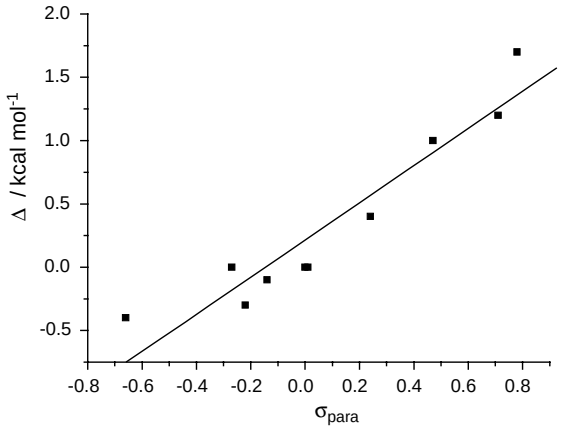


Figure 1. The change in H-bond strength (Δ) relative to the original phenol–formaldehyde system plotted against the Hammett substituent constants (σ_{para}).

Table 5. The ρ-constants for the calculated H-bonded systems.

Complexed systems	ρ (kcal mol^{-1})	±
para-phenol – formaldehyde	1.5	0.2
para-aniline – formaldehyde	1.3	0.3
para-phenol – methanol	−0.8	0.2
para-aniline – methanol	−1.6	0.2
para-phenol – acetate[a]	12.7	1.4
para-aniline – acetate[a]	11.3	1.2
para-phenol – methylamine[b]	−9.4	0.8
para-aniline – methylamine[b]	−11.0	0.9

[a]Acetate anion is deprotonated.
[b]The methylamine is protonated.

The trifluoromethyl- and methylsulphone-substituents

The nitro-group in drugs is often readily metabolised in the body, sometimes with toxic consequences. Medicinal chemists therefore seek alternative electron withdrawing species such as trifluoromethyl- and methylsulphone-groups with similar Hammett constants, which are usually more stable *in vivo* [41]. The H-bonding energy was checked for two systems, i.e. the *para*-phenol with methanol and *para*-aniline with protonated methylamine. In the first case both substituents yielded 1.9 kcal mol^{-1}, which is the same number as for $-NO_2$ (see Table 2). For the second case, *para*-aniline with protonated methylamine, the nitro derivative has a H-bond strength of 10.7 kcal mol^{-1}, whereas the trifluoromethyl and methylsulphone both have energies of 14.6 kcal mol^{-1}. Both substituents should therefore enhance H-bonding based affinity to a similar extent, as does $-NO_2$.

Relevance to ligand design

High affinity ligands generally need to have excellent shape complementarity and strong specific interactions via charge–charge interactions, H-bonding and hydrophobic effects. Overall electrostatic complementarity is also likely to be important. In this study we have shown that H-bond energies to phenol and aniline ligands may be strongly affected by substituents with potential changes in K_d as large as ten orders of magnitude. However, an added substitute will undoubtedly have further effects than just increasing H-bonding strength. First, a steric clash is possible, i.e., there is not enough room for the substituent within the binding pocket of a protein target. Second, electron perturbation effects rendered by the groups can change the locations of the minima on the potential energy surface. This is important because it has been established that high affinity ligands correspond to low free energy conformations [42, 43]. Third, the hydrophobicity can be changed and this may result in significant changes in entropy (ΔS) [44]. For hydrophobic interactions, energetic gain is mainly due to replacements and the release of ordered water molecules with a smaller enthalpic contribution from dispersion interactions favourable contacts between lipophilic groups [3, 4].

For structure-based design the challenge is to use all of these factors to improve the affinity and selectivity of a pertinent ligand. The present study suggests that in H-bonding interaction, there is considerable potential for increasing the energy of H-bonds especially when the slope of the Hammett plot (ρ) may be much greater when charged protein residues are targeted by the ligand.

Conclusions

DFT calculations have revealed that H-bonding between phenol and aniline with methanol and formaldehyde results in relatively low H-bonding energies, and *para*-substitution has little effect. In the case of protonated methylamine cation and

acetate anion, H-bond energies are much larger, and substitution at the *para*-position leads to pronounced changes ($\leq$14 kcal mol^{-1}). These methods may therefore be used as ligand design tools in the framework of drug discovery projects.

Acknowledgements

We gratefully acknowledge financial support from Cancer Research UK Development Committee (J. Reynisson) and the Institute of Cancer Research.

References

1. Pauling, L., Corey, R.B. and Branson, H.R., Proc. Natl. Acad. Sci. USA, 37 (1951) 205.
2. Pauling, L. and Corey, R.B., Proc. Natl. Acad. Sci. USA, 37 (1951) 729.
3. Böhm, H.J. and Klebe, G., Angew. Chem. Int. Ed. Engl., 35 (1996) 2588.
4. Klebe, G. and Böhm, H.J., Recept. Signal Transduction Res., 17 (1997) 459.
5. Höltje, H.D., Sippl, W., Rognan, D. and Folkers, G., Molecular Modeling, Second Edition. Wiley-VCH, Weinheim, Germany, 2003.
6. An analog of the highly successful Glivec kinase inhibitor interacts with the protein through hydrogen bonds, some of which confer specificity [7].
7. Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B. and Kuriyan, J., Science, 289 (2000) 1938.
8. Very weak correlation has been found between the number of contributing H-bonds to binding affinity [3].
9. Ward, W.H.J. and Holdgate, G.A., In King, F.D. and Oxford, A.W. (Eds.), Progress in Medicinal Chemistry, Vol. 38. Elsevier Science, 2001, pp. 309–376.
10. Roberts, D.A. and Ward, W.H.J., In King, F.D. (Ed.), Medicinal Chemistry: Principles and Practice, Royal Society of Chemistry, Cambridge, UK, 2002, pp. 64–90.
11. Abraham, M.H. and Platts, J.A., J. Org. Chem., 66 (2001) 3484.
12. Ahn, D., Park, S., Lee, S. and Kim, B., J. Phys. Chem. A, 107 (2003) 131.
13. Pejov, L., Chem. Phys., 285 (2002) 177.
14. Koch, W. and Holthausen, M.C., A Chemist's Guide to Density Functional Theory. Wiley-VCH, Weinheim, Germany, 1999.
15. Reynisson, J. and Steenken, S., J. Mol. Struct. (Theochem), 635 (2003) 133.
16. Reynisson, J. and Steenken, S., Phys. Chem. Chem. Phys., 4 (2002) 5353.
17. Abdali, S., Jalkanen, K.J., Cao, X., Nafie, L.A. and Bohr, H., Phys. Chem. Chem. Phys., 6 (2004) 2434.
18. Tuma, C. and Sauer, J., Chem. Phys. Lett., 387 (2004) 388.
19. Shishkin, O.V., Elstner, M., Frauenheim, T. and Suhai, S., Int. J. Mol. Sci., 4 (2003) 537.
20. Guerra, C.F., Bickelhaupt, F.M., Snijders, J.G. and Baerends, E.J., J. Am. Chem. Soc., 122 (2000) 4117.
21. Braida, B., Hiberty, P.C. and Savin, A., J. Phys. Chem. A, 102 (1998) 7872.
22. Pollet, R., Savin, A., Leininger, T. and Stoll, H., J. Chem. Phys., 116 (2002) 1250.
23. Kamiya, M., Tsuneda, T. and Hirao, K., J. Chem. Phys., 117 (2002) 6010.
24. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., J.A. Montgomery, J., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., N. Rega, G. A. Petersson, Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. and Pople, J.A. GAUSSIAN 03 (B.3). Gaussian, Inc., Pittsburgh, PA, 2003.
25. Lee, C., Yang, W. and Parr, R.G., Phys. Rev. B., 37 (1988) 785.
26. Becke, A.D., Phys. Rev. A., 38 (1988) 3098.
27. Becke, A.D., J. Chem. Phys., 98 (1993) 5648.
28. Frisch, M.J., Pople, J.A. and Binkley, J.S., J. Chem. Phys., 80 (1984) 3265.
29. Wong, M.W., Chem. Phys. Lett., 256 (1996) 391.
30. Boys, S.F. and Bernardi, F., Mol. Phys., 19 (1970) 553.
31. Duijneveldt, F.B.v., Rijdt, J.G.C.M.v.D.-v.d. and Lenthe, J.H.v., Chem. Rev., 94 (1994) 1873.
32. It also applies to some extent to arginine. In general, we do not try to cover all of the amino acids but focus on the most interesting ones in our opinion.
33. Here the consequences on the entropy term are ignored in order to keep the model simple. The qualitative effects on entropy are discussed later in the paper.
34. Desiraju, G.R. and Steiner, T., The Weak Hydrogen Bond. University Press, Oxford, UK, 1999, p. 12.
35. Additionally, the aniline–acetate anion system, substituted with –CH_2OH, forms bidentate bonding, i.e., each respective hydrogen on aniline forms a bond with each oxygen atom on the acetate molecule with H-bond lengths of 1.8 and 2.6 Å, respectively. This is the only complex, which forms a bidentate bonding observed in this work.
36. Reynisson, J. and Steenken, S., Org. Biomol. Chem., 2 (2004) 578.
37. Steenken, S., Chem. Rev., 89 (1989) 503.
38. Reynisson, J. and Steenken, S., Phys. Chem. Chem. Phys., 4 (2002) 5346.
39. Isaacs, N., Physical Organic Chemistry. Longman Scientific and Technical, Essex, UK, 1987, pp. 146–192.
40. The R values for the slopes lie between 0.97509 and 0.87004.

41. Nitro ($\sigma_{para} = 0.78$), trifluoromethyl ($\sigma_{para} = 0.53$) and methylsulphonite ($\sigma_{para} = 0.73$).
42. Boström, J., Norrby, P. and Liljefors, T., J. Comput.-aided Mol. Des., 12 (1998) 383.
43. Vieth, M., Hirst, J.D. and Brooks III, C.L., J. Comput.-aided Mol. Des., 12 (1998) 563.
44. For intercalation of ligands into the DNA stack hydrophobicity provides the overwhelming driving force for complex formation [45, 46].
45. Chaires, J.B., Curr. Opin. Struct. Biol., 8 (1998) 314.
46. Chaires, J.B., Biopolymers, 44 (1997) 201.

Supplementary information

Table S1. The single point and zero point vibrational energies (ZPE) of the optimised *para*-phenol-formaldehyde derivatives, single components and hydrogen-bonded pairs in hartrees (a.u.). The molecular structures are shown in Schemes 1 and 2.

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Formaldehyde	−114.5481455	0.026673	x
Phenol	−307.5799554	0.104473	x
Phenol + formaldehyde	−422.1362995	0.133197	0.30
Methyl-phenol	−346.9094458	0.131655	x
Methyl-phenol + formaldehyde	−461.4656109	0.160429	0.30
Nitro-phenol	−512.1577109	0.107165	x
Nitro-phenol + formaldehyde	−626.7168311	0.136021	0.16
Cyano-phenol	−399.852864	0.103174	x
Cyano-phenol + formaldehyde	−514.4112412	0.132041	0.33
Methoxy-phenol	−422.1414786	0.136724	x
Methoxy-phenol + formaldehyde	−536.6978603	0.165557	0.12
Chloro-phenol	−881.7634888	0.123679	x
Chloro-phenol + formaldehyde	−881.7629873	0.123679	0.31
Hydroxy-phenol	−382.8292732	0.108244	x
Hydroxy-phenol + formaldehyde	−497.3865577	0.137014	0.42
Amine-phenol	−362.9567875	0.120753	x
Amine-phenol + formaldehyde	−477.5125924	0.14961	0.30
Hydroxymethyl-phenol	−422.1495804	0.137071	x
Hydroxymethyl-phenol + formaldehyde	−536.7059168	0.165753	0.31
Aldehyde-phenol	−420.9460751	0.113918	x
Aldehyde-phenol + formaldehyde	−535.5040449	0.142731	0.33

Table S2. The single point and zero point vibrational energies (ZPE) of the optimised *para*-aniline-formaldehyde derivatives, single components and hydrogen-bonded pairs in hartrees (a.u.). The molecular structures are shown in Schemes 1 and 2.

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Formaldehyde	−114.5481455	0.026673	x
Aniline	−287.7078514	0.116926	x
Aniline + formaldehyde	−402.2613792	0.145395	0.13
Methyl-aniline	−327.03711	0.144107	x
Methyl-aniline + formaldehyde	−441.5899449	0.172659	0.23
Nitro-aniline	−492.28871	0.119451	x
Nitro-aniline + formaldehyde	−606.8434037	0.147353	0.06
Cyano-aniline	−379.9830355	0.115555	x
Cyano-aniline + formaldehyde	−494.5369472	0.143503	0.24
Methoxy-aniline	−402.2691194	0.149269	x

Table S2. Continued.

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Methoxy-aniline + formaldehyde	−516.8215578	0.177751	0.35
Chloro-aniline	−747.3345596	0.107395	x
Chloro-aniline + formaldehyde	−861.8876159	0.135808	0.24
Hydroxy-aniline	−362.9578536	0.120753	x
Hydroxy-aniline + formaldehyde	−477.5095092	0.149227	0.24
Amine-aniline	−343.084158	0.133306	x
Amine-aniline + formaldehyde	−457.6367204	0.161823	0.23
Hydroxymethyl-aniline	−402.2778439	0.149521	x
Hydroxymethyl-aniline + formaldehyde	−516.8305754	0.177648	0.23
Aldehyde-aniline	−401.0762952	0.126252	x
Aldehyde-aniline + formaldehyde	−515.6299422	0.154148	0.24

Table S3. The single point and zero point vibrational energies (ZPE) of the optimised *para*-phenol-methanol derivatives, single components and hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Methanol	−115.7728612	0.051235	x
Phenol + methanol	−423.3579047	0.157377	0.37
Methyl-phenol + methanol	−462.6876037	0.184623	0.38
Nitro-phenol + methanol	−627.9349489	0.159826	0.10
Cyano-phenol + methanol	−515.6300077	0.15594	0.32
Methoxy-phenol + methanol	−537.9199508	0.189632	0.38
Chloro-phenol + methanol	−882.983537	0.147713	0.35
Hydroxy-phenol + methanol	−498.6088945	0.16117	−0.45
Amine-phenol + methanol	−478.7355515	0.173667	0.15
Hydroxymethyl-phenol + methanol	−537.9274779	0.189882	0.37
Aldehyde-phenol + methanol	−536.7229998	0.166582	0.33
Para-trifluoromethyl-phenol	−644.7471976	0.109051	x
Trifluoromethyl-phenol + methanol	−760.5245997	0.161779	0.35
Methylsulphone-phenol	−895.6014841	0.14234	x
Methylsulphone-phenol + methanol	−1011.378778	0.195088	0.34

Table S4. The single point and zero point vibrational energies (ZPE) of the optimised *para*-aniline-methanol derivatives, hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Aniline + methanol	−403.4883229	0.170359	−0.18
Methyl-aniline + methanol	−442.8175072	0.197633	−0.02
Nitro-aniline + methanol	−608.0661522	0.172707	0.06
Cyano-aniline + methanol	−495.7609951	0.168814	0.08
Methoxy-aniline + methanol	−518.0496967	0.202773	0.12
Chloro-aniline + methanol	−863.1139042	0.160739	0.11
Hydroxy-aniline + methanol	−478.7373258	0.174291	0.12
Amine-aniline + methanol	−458.8656163	0.186762	−0.19
Hydroxymethyl-aniline + methanol	−518.0576446	0.202863	0.10
Aldehyde-aniline + methanol	−516.8543598	0.179608	0.08

Table S5. The single point and zero point vibrational energies (ZPE) of the optimised *para*-phenol-acetate anion derivatives, single components and hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Acetate anion	−228.614475	0.047931	x
Phenol + acetic acid	−536.2356969	0.153216	0.61
Methyl-phenol + acetic acid	−575.5640625	0.18045	0.59
Nitro-phenol + acetic acid	−740.8356819	0.155417	0.72
Cyano-phenol + acetic acid	−628.5252538	0.150728	0.67
Methoxy-phenol + acetic acid	−650.7956304	0.185561	0.59
Chloro-phenol + acetic acid	−995.8682033	0.143519	0.63
Hydroxy-phenol + acetic acid	−611.4839271	0.157045	0.58
Amine-phenol + acetic acid	−591.6088894	0.169806	0.57
Hydroxymethyl-phenol + acetic acid	−650.8064065	0.18559	0.62
Aldehyde-phenol + acetic acid	−649.6162778	0.160862	0.76

Table S6. The single point and zero point vibrational energies (ZPE) of the optimised *para*-aniline-acetate anion derivatives, hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Aniline + acetic acid	−516.3519707	0.165963	0.36
Methyl-aniline + acetic acid	−555.6803097	0.193171	0.21
Nitro-aniline + acetic acid	−720.9531209	0.168878	0.46
Cyano-aniline + acetic acid	−608.6427189	0.164875	0.43
Methoxy-aniline + acetic acid	−630.9111866	0.198275	0.38
Chloro-aniline + acetic acid	−975.9841478	0.156331	0.40
Hydroxy-aniline + acetic acid	−591.5996728	0.169723	0.37
Amine-aniline + acetic acid	−571.724132	0.182378	0.35
Hydroxymethyl-aniline + acetic acid	−630.9231282	0.198526	0.37
Aldehyde-aniline + acetic acid	−629.7347891	0.175636	0.43

Table S7. The single point and zero point vibrational energies (ZPE) of the optimised *para*-phenol-protonated methylamine derivatives, hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Protonated methylamine	−96.25525249	0.079429	x
Phenol + methylamine	−403.8621797	0.184946	0.35
Methyl-phenol + methylamine	−443.193867	0.212049	0.35
Nitro-phenol + methylamine	−608.426725	0.187482	0.32
Cyano-phenol + methylamine	−496.123399	0.183487	0.31
Methoxy-phenol + methylamine	−518.4274257	0.217015	0.35
Chloro-phenol + methylamine	−863.4840043	0.17516	0.33
Hydroxy-phenol + methylamine	−479.1134769	0.188622	0.35
Amine-phenol + methylamine	−459.2457703	0.200881	0.39
Hydroxymethyl-phenol + methylamine	−518.4346452	0.217364	0.34
Aldehyde-phenol + methylamine	−517.2196217	0.19428	0.32

Table S8. The single point and zero point vibrational energies (ZPE) of the optimised *para*-aniline-protonated methylamine derivatives, hydrogen-bonded pairs in hartrees (a.u.).

Para-substituted	Energy (a.u.)	ZPE (a.u.)	BSSE (kcal mol^{-1})
Aniline + methylamine	−383.9973196	0.197859	0.56
Methyl-aniline + methylamine	−423.329605	0.22505	0.21
Nitro-aniline + methylamine	−588.562548	0.200465	0.32
Cyano-aniline + methylamine	−476.2592175	0.196535	0.31
Methoxy-aniline + methylamine	−498.5629809	0.229945	0.34
Chloro-aniline + methylamine	−843.6198237	0.188213	0.34
Hydroxy-aniline + methylamine	−459.2489731	0.201608	0.35
Amine-aniline + methylamine	−439.3810511	0.213207	−0.04
Hydroxymethyl-aniline + methylamine	−498.5704703	0.230306	0.26
Aldehyde-aniline + methylamine	−497.355616	0.20718	0.30
Trifluoromethyl-aniline	−624.876829	0.121369	x
Trifluoromethyl-aniline + methylamine	−721.1569225	0.202485	0.46
Methylsulphone-aniline	−875.7313293	0.154665	x
Methylsulphone-aniline + methylamine	−972.0115123	0.235767	0.33