Aromatic Interactions in Proteins, DNA and Synthetic Receptors

C. A. Hunter

Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Aromatic interactions in proteins, DNA and synthetic receptors

By C. A. Hunter

Department of Chemistry, University of Sheffield, Sheffield S3 7HF, U.K.

Non-covalent interactions between aromatic molecules (π–π interactions) play a major role in biological molecular recognition. A simple theoretical model which accounts for the geometric properties of π–π interactions is described. The key feature of this model is that it specifically allows for out-of-plane π-electron density in the calculation of electrostatic interactions. Experimental evidence for the validity of the model comes from studies of the geometric distribution of phenylalanine-phenylalanine interactions in protein X-ray crystal structures. The model has also been used to design a synthetic molecular receptor which recognizes *p*-benzoquinone using H-bonds and edge-to-face π–π interactions.

Aromatic stacking interactions provide the crucial link between sequence and three-dimensional structure in double-helical DNA. The π–π interaction model has been used to calculate the conformational preferences of all ten DNA base-pair steps and the results provide new insight into the molecular basis of sequence-dependent DNA structure.

---

1. Introduction

Interactions between aromatic molecules represent one of the most important intermolecular forces in chemical and biological systems. We will use the term π–π interaction to describe all non-covalent interactions between delocalized π–systems, including interactions between aromatic molecules (Hunter & Sanders 1990). These interactions control, or have an important influence upon, a range of molecular recognition and self-assembly phenomena such as:

(*a*) the packing of aromatic molecules in the crystalline state and hence the materials properties of these compounds (Desiraju & Gavezzotti 1989);

(*b*) the base-stacking interactions which determine the sequence-dependent structure and properties of DNA as well as recognition of DNA by drugs and regulatory proteins (Hunter 1993);

(*c*) the three-dimensional structures of proteins (Burley & Petsko 1985);

(*d*) molecular recognition of drugs by biological receptors or enzymes and of guests by synthetic hosts (Hunter 1991).

Molecular recognition in these systems involves the interplay of many different effects which can be divided into three categories.

1. van der Waals (vdW) interactions. These define the size and shape specificity of non-covalent interactions.

2. Solvation. Two molecules which form a complex in solution must be desolvated before complexation can occur. The solvent may compete for recognition sites thereby destabilizing the complex. Alternatively, in polar solvents, solvophobic effects can stabilize the complex.

3. Electrostatic interactions. These are particularly important in conferring specificity on molecular recognition events.

Aromatic molecules are normally planar so that vdW interactions are maximized when the molecules adopt a perfectly stacked arrangement. The flat π-electron surfaces of the molecules are non-polar so that solvophobic effects also favour stacking. Despite this, interactions between aromatic molecules are frequently observed in quite different orientations. The reasons are to be found in the electrostatic interactions. As we shall see, it is the detailed molecular charge distributions which determine the properties of these interactions.

2. A model for π–π interactions

Aromatic molecules have a characteristic charge distribution. A positively charged σ-framework is sandwiched between two regions of negatively charged π-electron density (figure 1). Figure 2 illustrates how the electrostatic interaction between two such charge distributions varies as a function of orientation (Hunter & Sanders 1990). The face-to-face stacked orientation (the origin in figure 2) maximizes π-electron repulsion. For edge-to-face and offset stacked orientations, attractive interactions between the positively charged σ-frameworks and the negatively charged π-electrons dominate. Thus π-stacking is associated with large repulsive electrostatic interactions and is therefore much less favourable than one might expect on the basis of VdW interactions and solvophobic effects.

Most commercial molecular mechanics force-fields use partial atomic charges to calculate non-covalent electrostatic interactions. However, figure 2 shows that if we are to understand how aromatic molecules interact, it will be important to take into account the out-of-plane π-electron density. All of the calculations discussed in this paper use molecular charge distributions which contain π-electron charges as illustrated in figure 1.

The ideas summarized in figure 2 hold for non-polarized π-systems, i.e. hydrocarbons which contain no heteroatoms. Polarization of the π-system by heteroatoms alters this picture and will be discussed later.

3. π–π Interactions in proteins

Experimental evidence for the picture presented in figure 2 comes from an analysis of the orientations of aromatic interactions in proteins. Singh & Thornton (1990) have compiled a database of the geometry of all side-chain–side-chain interactions in high resolution X-ray crystal structures of proteins. We will concentrate on the interactions observed between the aromatic rings of phenylalanine (Phe) side-chains (Hunter *et al.* 1991). Figure 3*a* shows how the calculated electrostatic interaction between two benzene molecules varies as a function of orientation. This can be compared with the experimental distribution of Phe–Phe interaction geometries observed in the Singh–Thornton database (figure 3*b*). The experimental scatter maps out the region of conformational space which corresponds to an attractive electrostatic interaction. Orientations which would result in repulsive electrostatic interactions, such as face-to-face π-stacking, are not observed.

A strong clustering of experimental observations which would indicate a deep well-defined energy minimum is not observed. This is consistent with our calculations which indicate that the maximum electrostatic interaction is 6 kJ mol^{-1}. This is a rather weak interaction. However, aromatic side-chains often occur in clusters in
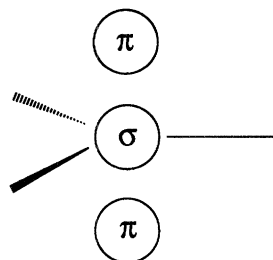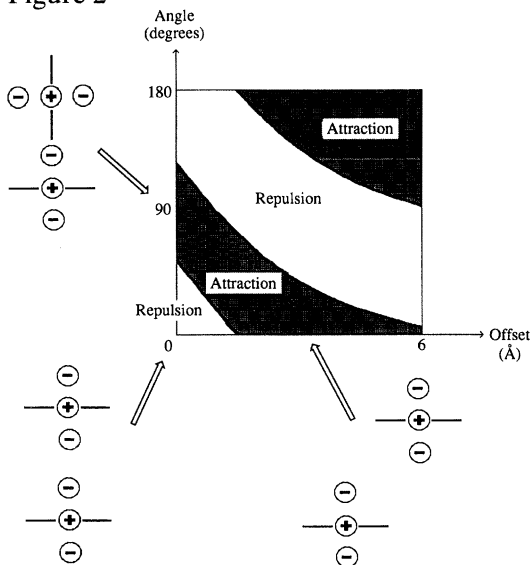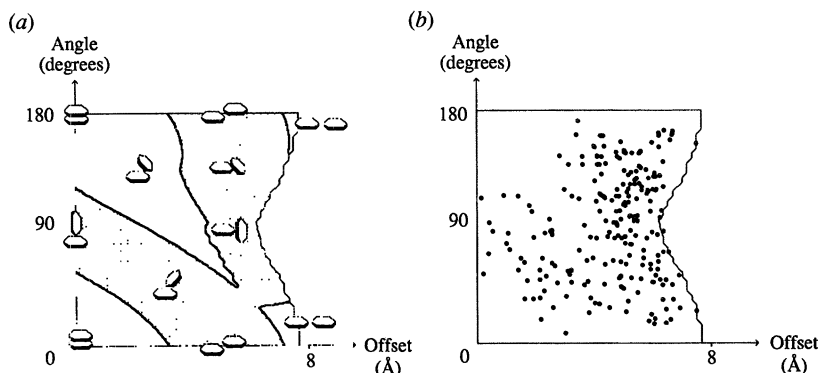
Figure 1

Figure 2

Figure 1. An sp^2 hybridized atom in a π-system.

Figure 2. Electrostatic interaction between two idealized π-atoms as a function of orientation: two attractive geometries and the repulsive face-to-face geometry are illustrated. Reproduced with permission from *J. Am. chem. Soc.*

Figure 3. (*a*) Electrostatic interaction between two benzene rings as a function of orientation. The geometries of interaction are illustrated. The shaded area is attractive. The unshaded area is repulsive. (*b*) A scatter plot of the experimentally observed geometries of interaction between phenylalanine aromatic rings in proteins. Reproduced with permission from *J. mol. Biol.*

proteins (Burley & Petsko 1985). These clusters contain several π–π interactions and so make a significant contribution to the stability of that particular structural motif.

4. Design of a synthetic receptor

To test these ideas further, we decided to design a synthetic host which would complex a guest using the edge-to-face π–π interactions which are predicted to be attractive (Hunter 1991). We chose *p*-benzoquinone as the guest since we were interested in mimicking some of the properties of bacterial photosynthetic reaction centres where quinone cofactors play an essential role in the primary photoinduced charge separation process (Michel *et al.* 1986; Allen *et al.* 1988).

p-Benzoquinone has a range of possible recognition sites which could be exploited in host design (figure 4*a*). In order to select the optimum π–π interactions, we must consider the effects of the polarizing heteroatoms. The quinone carbonyl oxygens withdraw π-electron density from the ring. This will weaken an edge-to-face interaction with the face of the quinone π-system. In contrast, the polarization will enhance an interaction between the edge of the quinone and the face of another π-system. The host which uses these edge-to-face interactions and maximizes the hydrogen-bonding interactions is illustrated in figure 4*b*.

This host has been synthesized. It complexes *p*-benzoquinone with an association constant (K_a) of 10^3 M^{-1} in chloroform. Complexation-induced changes in the ^{1}H-NMR chemical shifts of both the host and the guest indicate that the structure of the complex is as shown in figure 4*b*. These observations do not prove that there is an attractive π–π interaction in this system, but they are consistent with such an interaction. Moreover, the host is selective for *p*-benzoquinone: substitution of methyl groups, chlorines, fluorines or fused aromatic rings for the quinone hydrogens completely inhibits complexation ($K_a < 5$ M^{-1}). These changes would block the edge-to-face π–π interactions.

5. The effect of polarization by heteroatoms

In the example above, we have touched on the role of polarizing heteroatoms in modulating π–π interactions. The systems discussed in the first half of the paper were aromatic hydrocarbons which lacked heteroatoms. In these systems, face-to-face stacking is inhibited by π-electron repulsion. However, the introduction of polarizing substituents which perturb the molecular charge distribution changes this picture. Heteroatoms cause large partial atomic charges and lead to additional electrostatic interactions. We divided the electrostatic interactions in these systems into three categories (Hunter 1993).

1. πσ–πσ interactions. These are the interactions associated with the out-of-plane π-electron density if no polarization were present, i.e. the kind of interaction we have already discussed.

2. Atom–atom. These are the interactions between the partial atomic charges.

3. Atom–πσ. This is the cross-term of the other two interactions, i.e. the interaction between the partial atomic charges on one molecule and the out-of-plane π-electron density on the other.

For molecules that are highly polarized such as the DNA bases, πσ–πσ interactions are not very important. It is usually the atom–atom term which is the largest electrostatic interaction, but the atom–π term can also be significant. Moreover, the atom–π term is very sensitive to changes in geometry and can play an important role in determining the orientation of π–π interactions between polarized π-systems.

6. Sequence-dependent DNA structure

The three-dimensional structures and properties of double helical DNA depend critically on the sequence of the aromatic bases. Evidence comes from a wide variety of sources including, X-ray fibre diffraction studies of polymers (Leslie *et al.* 1980), X-ray crystal structures of oligomers (Drew *et al.* 1990; Kennard & Hunter 1991), gel running experiments (Calladine *et al.* 1988) and recognition of DNA by small organic molecules and proteins (Kennard & Hunter 1991; Travers & Klug 1990). An examination of the chemical structure of DNA shows that the only difference
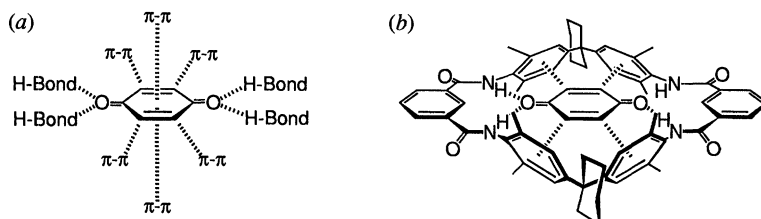
Figure 4. (*a*) Interaction sites available for molecular recognition of *p*-benzoquinone. (*b*) Complexation of *p*-benzoquinone by a synthetic macrocyclic host. Reproduced with permission from *J. chem. Soc. chem. Commun.*
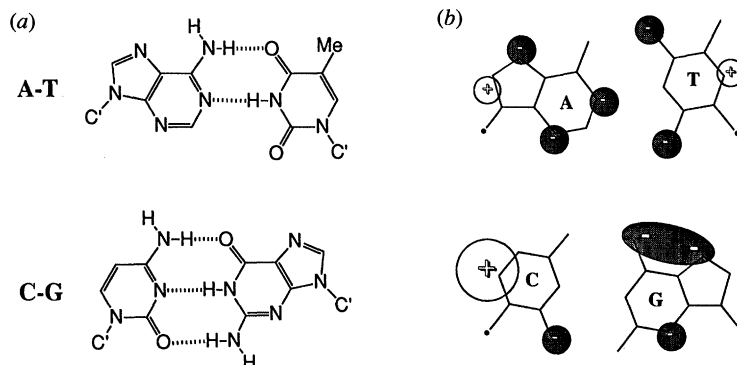

Figure 5. (*a*) The chemical structures of the DNA base-pairs. C′ is the atom which connects the bases to the sugar. (*b*) The molecular charge distributions of the base-pairs. Reproduced with permission from *J. mol. Biol.*

between different sequences lies in the aromatic bases (figure 5*a*). The sugar–phosphate backbone is identical regardless of sequence. It therefore follows that the cause of sequence-dependent variations in DNA structure is the π–π interactions between the aromatic bases which are stacked up the centre of the double helix.

We have applied the π–π interaction model described above to elucidate the molecular basis for these structural variations (Hunter 1993). The conceptual basis for tackling this problem was established by Calladine & Drew (1984). They showed how the conformation of a single base-pair step can be related to the overall three-dimensional structure of the double helix. The conformation of a base-pair step can be defined in terms of six degrees of freedom (figure 6) (Diekmann 1989). (There are another six internal degrees of freedom for each base-pair. One of these parameters, propeller twist (a rotation about the long axis of the base-pair), plays an important role in defining sequence-dependent changes in DNA structure but it will not be discussed here.) An analysis of DNA oligomer X-ray crystal structures shows that the parameters which are most important for defining sequence-dependent variations in structure are twist, slide and roll (Calladine & Drew 1984; C. R. Calladine & M. A. Elhassan, person communication). We have therefore analysed the π–π interactions for all ten possible base-pair steps as a function of these parameters. The results correlate rather well with experiment and throw new light on many properties of nucleic acids including the different conformational preferences of RNA and DNA, the formation of left-handed Z-DNA and the role of TATA in originating replication.

Inspection of the chemical structures of the bases (figure 5) reveals that all sequence-dependent effects must be caused by:

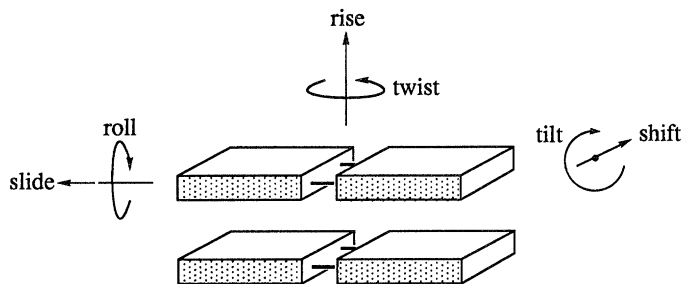(*a*) steric interactions with the guanine amino group in the minor groove;

Figure 6. The conformation of a base-pair step in double-helical DNA can be defined by three rotations (twist, roll and tilt) and three translations (rise, slide and shift). The block edges which correspond to the minor groove are shaded.
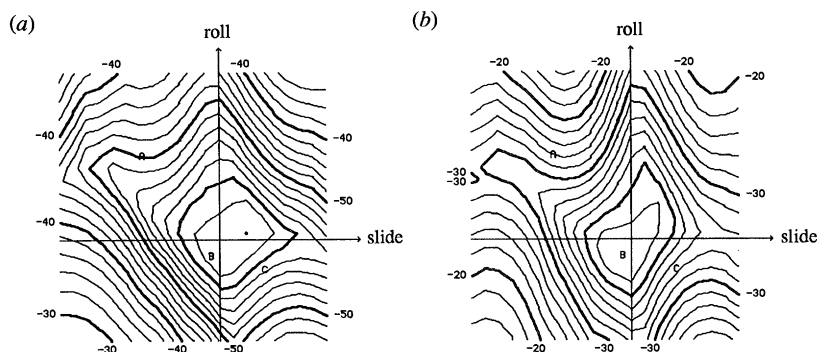

Figure 7. Contour plots showing the interaction energy for the AA/TT step (in kJ mol^{-1}) as a function of slide and roll. The contour spacing is 2 kJ mol^{-1}. The slide axis runs for -3 Å to 2 Å and the roll axis runs from $-15°$ to $25°$. Helical twist $= 36°$, propeller twist $= 15°$ and all other parameters are set to zero. Energy minima and conformations corresponding to A, B and C-DNA are labelled. (*a*) vdW interaction energy; (*b*) total π–π interaction energy (vdW + electrostatic). Reproduced with permission from *J. mol. Biol.*

(*b*) steric interactions associated with the configuration of the step, whether the sequence of bases in purine–purine, purine–pyrimidine or pyrimidine–purine;

(*c*) steric interactions with the thymine methyl group in the major groove;

(*d*) electrostatic interactions associated with the different molecular charge distribution across the A-T and C-G base-pairs (figures 5*b*).

The first two effects have been discussed in detail by Calladine (1982) and Calladine & Drew (1984). Their observations are reproduced by our calculations, but in this paper, we will concentrate on the last two effects. We will consider just two base-pair steps to illustrate the approach. A detailed analysis of the conformational properties all ten base-pair steps can be found in Hunter (1993).

Two base-pair steps which have been thoroughly characterized by experiment are AA/TT and CC/GG. These steps exemplify the two most common conformational families which were first observed by fibre diffraction methods, the A and B polymorphs. AA/TT has a very strong preference for the B-form (roll $\approx 0°$ and slide ≈ 0 Å) (Leslie *et al.* 1980; Nelson *et al.* 1987; Calladine *et al.* 1988). CC/GG has a very strong preference for the A-form (roll $\approx 12°$ and slide ≈ -1.5 Å) (Leslie *et al.* 1980; McCall *et al.* 1985). Why is this?

Contour plots of the magnitude of the π–π interaction between the base-pairs as a function of slide and roll are shown in figures 7 and 9. The calculation for AA/TT shows a deep well-defined energy minimum at the origin (slide $= 0$ Å and roll $= 0°$),
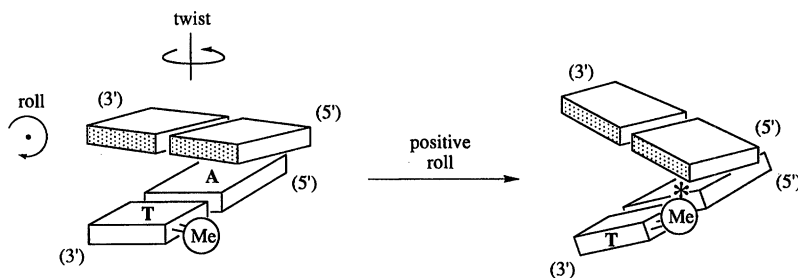
Figure 8. An AX/XT step viewed along the slide/roll axis (X is any base). The direction of positive slide is towards the reader. The block edges which correspond to the minor groove are shaded. The thymine methyl group prevents positive roll. Increasing roll leads to a steric clash (asterisk) between the thymine methyl and the other base on the same strand. Reproduced with permission from *J. mol. Biol.*
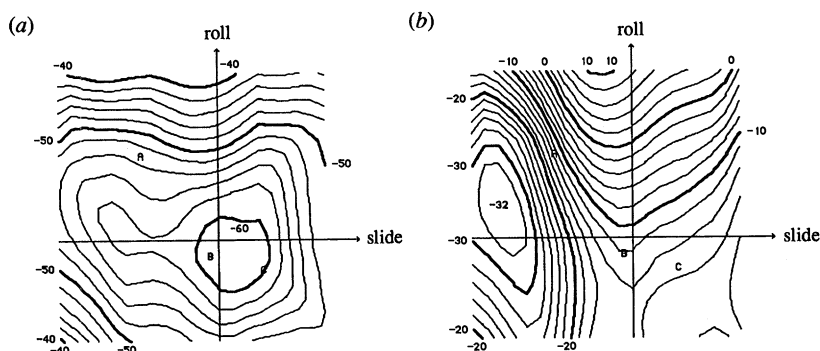

Figure 9. Contour plots showing the interaction energy for the CC/GG step (in kJ mol^{-1}) as a function of slide and roll. The contour spacing is 2 kJ mol^{-1}. The slide axis runs from -3 Å to 2 Å and the roll axis runs from $-15°$ to $25°$. Helical twist $= 36°$, propeller twist $= 15°$ and all other parameters are set to zero. Energy minima and conformations corresponding to A, B and C-DNA are labelled. (a) vdW interaction energy; (b) total π–π interaction energy (vdW + electrostatic). Reproduced with permission from *J. mol. Biol.*

the region of conformational space which corresponds to the *B*-form that is observed experimentally (figure 7). Comparison of the VdW interaction energy with the total π–π interaction energy shows that the conformational properties of this step are dominated by steric effects. Electrostatics are relatively unimportant because the AT molecular charge distribution has no large regions of high charge density.

The major steric interaction which locks AA/TT into the B-form is a clash between the thymine methyl group and the 5′-neighbouring base which occurs in positive roll A-type conformations (figure 8). This interaction also provides an explanation for the different conformational properties of double-helical DNA and RNA, RNA has uracil in place of thymine and so lacks the major groove methyl groups. Thus double-helical RNA has only been observed in the A-form (Arnott *et al.* 1973; Dock-Bregon *et al.* 1988), whereas the A-form of DNA is destabilized by the thymine methyl groups so that DNA is polymorphic (Leslie *et al.* 1980).

Electrostatic interactions play a more important role in the CC/GG step. C–G base-pairs have two regions of very high charge density, a positive charge over cytosine and a negative charge over guanine (figure 5*b*). The vdW interaction for CC/GG shows a broad flat energy minimum close to the origin, the B-form (figure 9*a*). Addition of the electrostatic interaction moves the energy minimum to large
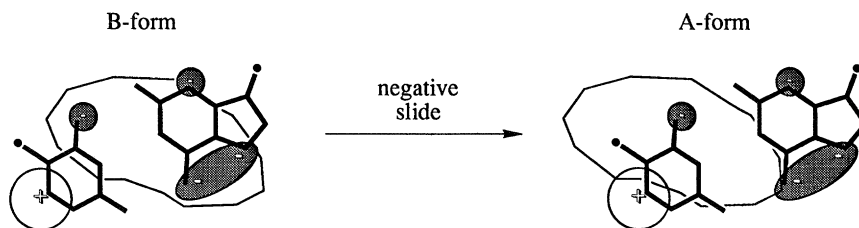
C. A. Hunter

Figure 10. Atom–πσ interactions in a CX/XG step (X is any base). The thin lines map out the area covered by the π-electron density of the X-X base-pair. The molecular charge distribution of the C-G base-pair is shown. · indicates the position of the C$_1'$ atom where the bases are attached to the sugar. Reproduced with permission from *J. mol. Biol.*

negative slide, close to the A-type conformation, in agreement with the experimental results for this sequence.

When the calculations were repeated omitting the out-of-plane π-electron density and using only partial atomic charges to calculate the electrostatic interaction, a very broad flat energy minimum with no clearly defined conformational preference was obtained. Clearly, it is essential to allow for out-of-plane π-electron density to accurately model the properties of π–π interactions even for these very highly polarized π-systems. The important interaction which causes the negative slide A-type conformation for CC/GG is the atom–πσ interaction. Figure 10 shows that the A-form is associated with a movement of the π-electron density of the bottom base-pair away from the guanine negative charge and towards the cytosine positive charge.

Thus calculations using this model for π–π interactions are capable of reproducing the experimental conformational preferences of different DNA sequences and allow us to probe the molecular basis for these properties (Hunter 1993).

7. Conclusion

These studies show that to understand or model non-covalent interactions between aromatic molecules, it is essential to consider electrostatic interactions involving the out-of-plane π-electron density. The pictures presented in this paper show how the properties of π–π interactions can be understood simply on the basis of the shapes and charge distributions of the individual molecules. Our calculations show that π–π interactions have quite strong geometrical requirements: these interactions are in fact directional to a much greater extent than was previously thought. They therefore play a very important role in controlling specificity in biological molecular recognition.

References

Allen, J. P., Feher, G., Yeates, T. O., Komiya, H. & Rees, D. C. 1988 Structure of the reaction center from *Rhodobacter sphaeroides* R-26: protein-cofactor (quinones and Fe²⁺) interactions. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8487–8491.

Arnott, S., Hukins, D. W. L., Dover, S. D., Fuller, W. & Hodgson, A. R. 1973 Structures of synthetic polynucleotides in the A-RNA and A′-RNA conformations: X-ray diffraction analyses of the molecular conformations of polyadenylic acid·polyuridylic acid and polyinosinic acid·polycytidylic acid. *J. mol. Biol.* **81**, 107–122.

Burley, S. K. & Petsko, G. A. 1985 Aromatic-aromatic interaction: a mechanism of protein structure stabilisation. *Science, Wash.* **229**, 23–28.

Calladine, C. R. 1982 Mechanics of sequence-dependent stacking of bases in B-DNA. *J. mol. Biol.* **161**, 343–352.

Calladine, C. R. & Drew, H. R. 1984 A base-centred explanation of the B-to-A transition in DNA. *J. mol. Biol.* **178**, 773–782.

Calladine, C. R., Drew, H. R. & McCall, M. J. 1988 The intrinsic curvature of DNA in solution. *J. mol. Biol.* **201**, 127–137.

Desiraju, G. R. & Gavezzotti, A. 1989 From molecular to crystal structure; polynuclear aromatic hydrocarbons. *J. chem. Soc. chem. Commun.*, 621–623.

Diekmann, S. 1989 Definitions and nomenclature of nucleic acid structure parameters. *J. mol. Biol.* **205**, 787–791.

Dock-Bregon, A. C., Chevrier, B., Podjarny, A., Moras, D., de Bear, J. S., Gough, G. R. & Gilham, P. T. 1988 High-resolution structure of the RNA duplex [U(UA)$_6$A]$_2$. *Nature, Lond.* **335**, 375–378.

Drew, H. R., McCall, M. J. & Calladine, C. R. 1990 New approaches to DNA in the crystal and in solution. In *DNA topology and its biological effects*, pp. 1–56. Cold Spring Harbor.

Hunter, C. A. 1991 Molecular recognition of *p*-benzoquinone by a macrocyclic host. *J. chem. Soc. chem. Commun.*, 749–751.

Hunter, C. A. 1993 Sequence-dependent DNA structure. The role of base stacking interactions. *J. mol. Biol.*, 1025–1054.

Hunter, C. A. & Sanders, J. K. M. 1990 The nature of π–π interactions. *J. Am. chem. Soc.* **112**, 5525–5537.

Hunter, C. A., Singh, J. & Thornton, J. M. 1991 π–π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. mol. Biol.* **218**, 837–846.

Kennard, O. & Hunter, W. N. 1991 Single-crystal X-ray diffraction studies of oligonucleotides and oligonucleotide-drug complexes. *Angew. Chem. Int. End Engl.* **30**, 1254–1277.

Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. 1980 Polymorphism of DNA double helices. *J. mol. Biol.* **143**, 49–72.

Michel, H., Epp, O. & Diesenhofer, J. 1986 Pigment-protein interactions in the photosynthetic reaction centre from *Rhodopseudomonas viridis. EMBO J.* **5**, 2445–2451.

McCall, M. J., Brown, T. & Kennard, O. 1985 The crystal structure of d(GGGGCCCC). A model for poly(dG)·poly(dC). *J. mol. Biol.* **183**, 385–396.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. 1987 The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature, Lond.* **330**, 221–226.

Singh, J. & Thornton, J. M. 1990 SIRIUS – an automated method for the analysis of the preferred packing arrangements between protein groups. *J. mol. Biol.* **211**, 595–615.

Travers, A. A. & Klug, A. 1990 Bending of DNA in nucleoprotein complexes. In *DNA Topology and Its Biological Effects*, pp. 57–106. Cold Spring Harbor.