ACCOUNTS
of chemical research

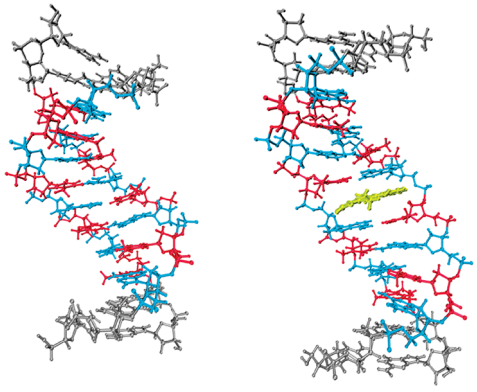
On the Importance and Origin of Aromatic Interactions in Chemistry and Biodisciplines

KEVIN E. RILEY[†] AND PAVEL HOBZA[*,†,‡]

[†]*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague, Czech Republic, and* [‡]*Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, 771 46 Olomouc, Czech Republic*

RECEIVED ON MARCH 14, 2012

CONSPECTUS



Aromatic systems contain both σ- and π-electrons, which in turn constitute σ- and π-molecular orbitals (MOs). In discussing the properties of these systems, researchers typically refer to the highest occupied and lowest unoccupied MOs, which are π MOs. The characteristic properties of aromatic systems, such as their low ionization potentials and electron affinities, high polarizabilities and stabilities, and small band gaps (in spectroscopy called the $N \rightarrow V_1$ space), can easily be explained based on their electronic structure. These one-electron properties point to characteristic features of how aromatic systems interact with each other.

Unlike hydrogen bonding systems, which primarily interact through electrostatic forces, complexes containing aromatic systems, especially aromatic stacked pairs, are predominantly stabilized by dispersion attraction. The stabilization energy in the benzene dimer is rather small (~2.5 kcal/mol) but strengthens with heteroatom substitution. The stacked interaction of aromatic nucleic acid bases is greater than 10 kcal/mol, and for the most stable stacked pair, guanine and cytosine, it reaches approximately 17 kcal/mol. Although these values do not equal the planar H-bonded interactions of these bases (~29 kcal/mol), stacking in DNA is more frequent than H-bonding and, unlike H-bonding, is not significantly weakened when passing from the gas phase to a water environment.

Consequently, the stacking of aromatic systems represents the leading stabilization energy contribution in biomacromolecules and in related nanosystems. Therefore stacking (dispersion) interactions predominantly determine the double helical structure of DNA, which underlies its storage and transfer of genetic information. Similarly, dispersion is the dominant contributor to attractive interactions involving aromatic amino acids within the hydrophobic core of a protein, which is critical for folding.

Therefore, understanding the nature of aromatic interactions, which depend greatly on quantum mechanical (QM) calculations, is of key importance in biomolecular science. This Account shows that accurate binding energies for aromatic complexes should be based on computations made at the (estimated) CCSD(T)/complete basis set limit (CBS) level of theory. This method is the least computationally intensive one that can give accurate stabilization energies for all common classes of noncovalent interactions (aromatic–aromatic, H-bonding, ionic, halogen bonding, charge-transfer, etc.). These results allow for direct comparison of binding energies between different interaction types. Conclusions based on lower-level QM calculations should be considered with care.

Introduction

What makes interactions of aromatic systems unique? The answer is simple, it is the character of aromatic systems. The electronic structure of aromatic molecules (as well as conjugated π-systems) is determined by the fact that these systems are composed mainly of carbon atoms in their sp^2 hybridization. There are two important consequences of this. First, the aromatic systems are either linear or planar. Second, their valence electrons are of s and p character and by linear combination of these atomic orbitals σ- and π-molecular orbitals (MO's) are formed. Both the highest occupied and lowest unoccupied MO's in aromatic systems are of π-character, which is responsible for the fact that ionization potentials and electronic affinities are low, much

Published on the Web 08/08/2012 www.pubs.acs.org/accounts
10.1021/ar300083h

lower than those in nonaromatic systems. Finally, systems with delocalized π electrons exhibit high polarizabilities, as demonstrated in a recent study showing that the aromatic amino acid side chains have higher polarizabilities than aliphatic ones.[1]

The planarity (or linearity) of aromatic systems together with their high polarizability and multipole moments (quadrupole moment (Q) in hydrocarbons and potential dipole moment(s) (μ) in heteroatomic systems) is of key importance for the 3D architecture of their complexes. The planarity of aromatic systems is especially significant in stacked (or parallel displaced) complexes, because it allows for the largest amount of dispersion contact between two species. Results by Wheeler and co-workers indicate that it is the planarity of aromatic systems that leads to strong interactions in stacked and parallel displaced aromatic complexes.[2]

Let us mention here the fascinating world of nucleic acids. The basic function of DNA, the storage and transfer of genetic information, is due to its double helical structure. It is now evident that this double helical structure is mainly due to interactions of planar nucleic acid bases that are of H-bonded and stacked types. Xenobiotic systems interact with DNA in different ways, a very important one of which is intercalation, where a planar aromatic system intercalates between stacked DNA base pairs.

Characteristics of Aromatic Dimers

There are many computational tools that have been used to describe complexes of aromatic systems in the past several years. All accurate binding energies described in this work are based on computations made at the (estimated) CCSD(T)/complete basis set limit (CBS) level of theory.[3] This is the least computationally intensive method giving accurate stabilization energies for all common classes of noncovalent interactions (aromatic−aromatic, H-bonding, ionic, halogen bonding, etc.), allowing for direct comparison of binding energies between interaction types. To understand the 3D structure of aromatic systems, from their dimers up to polymers, we must understand the nature of their stabilization. Insight into the character of noncovalent interactions can be gained using symmetry-adapted perturbation theory (SAPT),[4] with which the total stabilization energy is computed as the sum of various energy contributions of the first and second order: electrostatic, exchange repulsion, induction, and dispersion. The second term is systematically repulsive. The third and fourth terms, describing interactions between permanent multipoles of one subsystem and induced multipoles of the second subsystem (induction) and interactions between instantaneous multipoles of one subsystem and induced multipoles of the second subsystem (dispersion), are systematically attractive. The first term, which describes an interaction of electric multipoles in both subsystems is either attractive or repulsive, depending on the orientation of the subsystems. When SAPT calculations are performed with suitably large subsystem wave functions the resulting stabilization energies generally agree very closely with those of CCSD(T). The results of SAPT calculations are often used to complement CCSD(T) results, because they offer more insight into the character of an interaction. Multipole analysis, usually considering the dipoles and quadrupoles contained within aromatic and heterocyclic aromatic systems, can be used as a qualitative tool that aids in understanding the electrostatic forces that contribute to stability in aromatic complexes and offers a quick indication of the structure and stability of a given complex.

The characteristics and specific features of aromatic interactions will be demonstrated in the following text for several typical complexes of aromatic systems.

Benzene Dimer

The benzene dimer represents, without doubt, the most widely studied aromatic dimer, and the literature devoted to it is enormous. However, its most significant features were elucidated only in the last 20 years, and we have had the privilege of participating in this exciting research.[5−7] Experiments at the end of the 1980s clearly excluded the existence of the parallel stacked (PS) structure (Figure 1a) of the dimer, which was expected on the basis of the importance of dispersion energy.[8] Maximal overlap in this dimer corresponds to the largest contribution of the dispersion energy. It was further shown that the subsystems in the dimer are not identical, which lead to prediction of the T-shaped structure (Figure 1c). Other experiments at the beginning of the 1990s indicated the existence of more than one dimer structure. On the basis of relatively primitive correlated QM calculations, we were able to show that the parallel-displaced (PD) structure (Figure 1b) is either comparably stable or even more stable than the T-shaped one, which has been confirmed using the highly accurate QM calculations available at the present time. The interpretation of these results shed more light on the character and specificity of aromatic interactions in general. The PS structure is strongly stabilized by the dispersion energy but is even more strongly destabilized by electrostatic $Q-Q$ and exchange-repulsion interactions. It should be kept in mind that the attraction of the

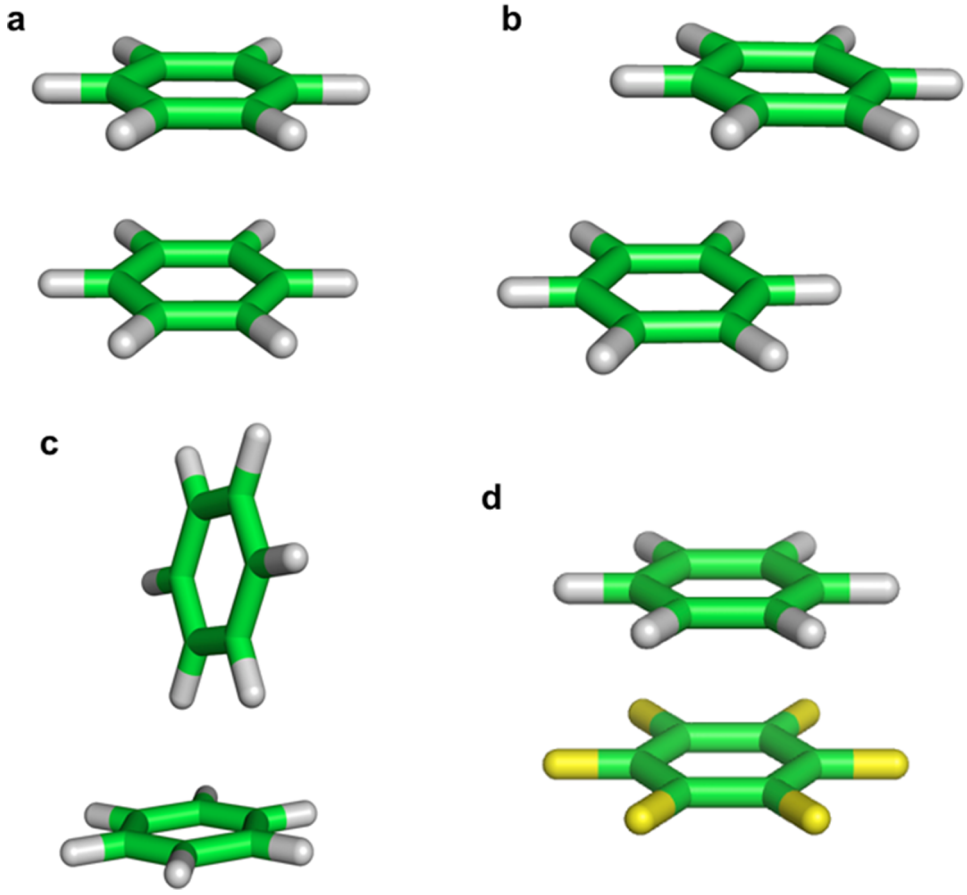


FIGURE 1. Equilibrium structures of (a) the parallel stacked benzene dimer, (b) the parallel displaced benzene dimer, (c) the T-shaped benzene dimer, and (d) the parallel stacked benzene–hexafluorobenzene complex.

dispersion energy is proportional to R^{-6}, while the repulsion of the electrostatic $Q-Q$ energy is proportional to R^{-5}, where R is the distance between centers of mass. The quadrupole repulsion in this structure changes to attraction when passing to the T-shaped structure, and because the dispersion energy is also attractive here (though less so than in the PS structure), the T-shaped structure becomes stable. For a long time, it was believed that stabilization of the PD structure is partially due to a favorable $Q-Q$ interaction; however, accurate perturbation analysis has shown that this interaction is actually slightly repulsive.[9] The overall electrostatic attraction is attributable to a favorable interaction of hexadecapoles. The fact that the stabilities of the PD and T-shaped structures are comparable is thus due to attractive electrostatic and dispersion energies.

Experiments on the benzene dimer, the first (rather primitive) calculations, and recent highly accurate calculations have shown the stabilization of the benzene dimer (both PD and T-shaped) to be about 2.7 kcal/mol.[10−13] Evidently the interactions in the benzene dimer are too weak to be regarded as being critical interaction types in nature. There are several ways to make these interactions more attractive, many of which are utilized by nature. These will be demonstrated in the following paragraphs.

Substituted Benzene Dimers and Heterocyclic Aromatic Rings

The main reason that attraction in the benzene dimer is rather weak is the fact that the PS structure, which exhibits a great deal of attraction from dispersion, is destabilized by electrostatic interactions. This feature can be removed by passing from the interaction of identical monomers, as in the benzene dimer, to an interaction of different monomers, such as in benzene···hexahalogenbenzene complexes.[14] In the hexafluorobenzene···benzene complex, the quadrupole moment of each monomer is practically identical but opposite in sign. This has one very important consequence: the $Q-Q$ electrostatic interaction will be attractive for the PS structure, while it will be repulsive for the T-shaped and PD structures. The PS structure (Figure 1d) is now stabilized by dispersion and electrostatic energies, which leads to an important stabilization increase. In the cases

of benzene···hexafluorobenzene and benzene···hexachlorobenzene, the CCSD(T)/CBS binding energies for the PS complexes are about 6.3 and 8.8 kcal/mol, respectively.[14] The most stable configuration of the benzene···hexafluorobenzene dimer is a PD structure whose parallel displacement (∼1.0 Å)[15] is smaller than that of the benzene dimer (∼1.6 Å).[16] It should also be noted that, since halogens are large and have large polarizabilities (especially chlorine, bromine, and iodine), it is likely that direct dispersion interactions involving the halogens also play a role in stabilizing these complexes. These stabilization energies are substantial, suggesting that this motif may be considered as a powerful tool in supramolecular construction.

The introduction of heteroatoms into either of the rings in a PS or PD benzene dimer also has a significant effect on the strength and character of the aromatic−aromatic interaction. If a dipole moment is created by the presence of the heteroatom (most commonly nitrogen), it will interact attractively with the benzene quadrupole moment, resulting in a stronger interaction that is more electrostatic in nature.[17−19] Considering a complex in which both benzene dimers contain heterocyclic substituents (e.g., the pyridine dimer), there are now possibilities for dipole−dipole, dipole−quadrupole, and quadrupole−quadrupole interactions. In such a complex, the interaction will tend to be yet stronger and more electrostatic in character. The presence of a dipole−dipole interaction makes the relative (rotational) orientation of the two monomers relevant, because there is now a possibility for the dipoles to be aligned or antialigned. Further substitutions of heterocyclic atoms will tend to produce stronger, more electrostatic interactions whose orientational dependencies become more complex.

Substitution of benzene hydrogens with other atoms or chemical groups can also have a strong impact on aromatic−aromatic binding strengths.[20−22] Interestingly, for the stacked benzene dimer near its equilibrium separation, it is found that the substitution of either an electron-withdrawing or electron-donating atom/group tends to increase the strength of the interaction.[20,22] This behavior contradicts the π-resonance model, which is the most common and long-held hypothesis used to describe substitutional effects in aromatic dimers. This model is based on the polarization of the π electron cloud by the aromatic substituent, with an electron-withdrawing substituent yielding a less negatively charged π system and, thus, a stronger interaction. Substitution of an electron-donating group would be expected to produce the opposite behavior. The disagreement of the π-resonance model with observed computational results has been the focus of intense investigation, and several newer models describing aromatic substitution effects have been developed in the last 2 years. These are based on direct interactions involving the substituent and on the increased size of the electron cloud brought about by substitution, which results in increased interpenetration effects. We will not discuss these new theoretical models at length here but will refer the reader to the excellent review by Wheeler and co-workers[23] and to articles by Sherrill and co-workers,[24] Lewis and co-workers,[25] and Wheeler.[26]

Figure 2 shows the six PS and PD aromatic complexes contained in the S66 data set of noncovalent interactions, which is a data set containing accurate CCSD(T)/CBS binding energies for a wide variety of noncovalent interaction types represented by 66 complexes (further information on S66 is available in the original papers[19,27] and in a recent review[28]). These six complexes are composed of different combinations of benzene, pyridine, and uracil, which represent an unsubstituted aromatic system, a simple heterocyclic aromatic system, and a complex heterocyclic and aromatically substituted system, respectively. An aromatic ring containing both heterocyclic atoms and aromatic substituents will tend to have a very complex charge distribution and interact with similar aromatic moieties in a geometrically specific manner largely governed by the alignment of dipole and quadrupole moments between the two monomers. Such interactions are, among the aromatic−aromatic complexes, the most electrostatic in character (although still dominated by dispersion) and generally tend to favor geometries that are more stacked (PS), as opposed to PD, orientations. Only in the case of the benzene dimer (Figure 2a) is the binding energy of the T-shaped dimer (not shown) comparable to that of the PD (or PS) structure. The T-shaped dimers of the other complexes given in Figure 2 are all less stable than the stacked dimers. It can be seen in Figure 2 that the inclusion of heteroatoms and aromatic substituents generally results in structures approaching the stacked (face-to-face) arrangement.

Table 1 gives the relative DFT-SAPT (a DFT variant of the SAPT method)[4,29,30] contributions of dispersion and electrostatics to attraction for the six aromatic systems as well as their estimated CCSD(T)/CBS binding energies (please see refs 19 and 27 for calculation details). Here the binding energy contributions coming from induction and higher-order effects are very small compared with dispersion and electrostatics and are not given. It can be seen that the introduction of heteroatoms and aromatic substitutents generally results in stronger interactions that are more

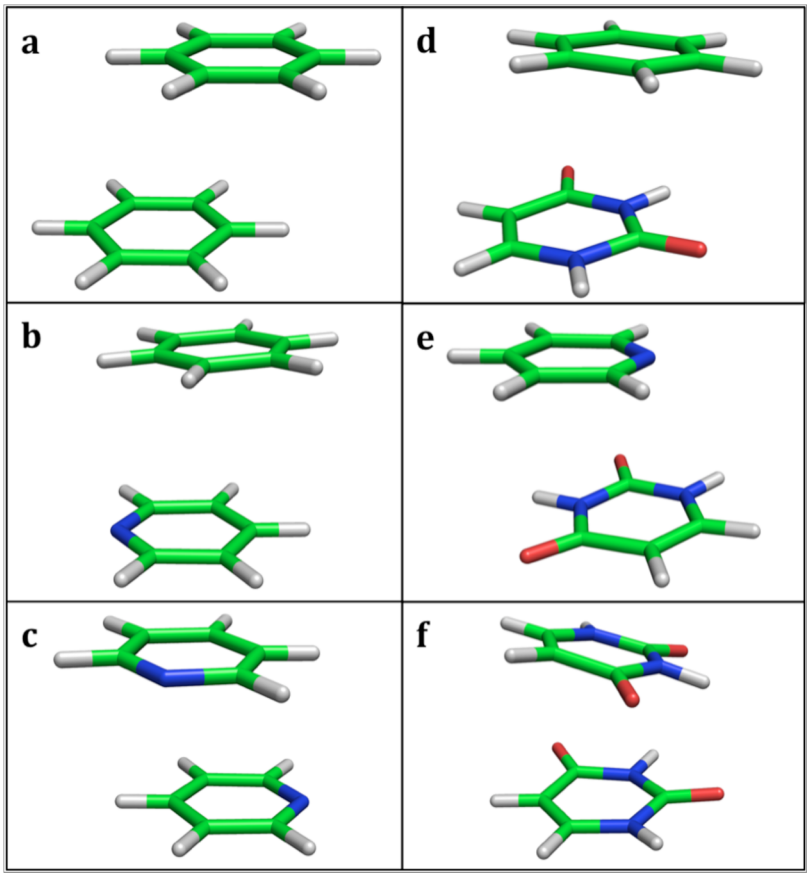


FIGURE 2. Parallel displaced and parallel stacked aromatic complexes contained within the S66 data set. The complexes are (a) benzene dimer, (b) pyridine···benzene, (c) pyridine dimer, (d) uracil···benzene, (e) uracil···pyridine, and (f) uracil dimer.

TABLE 1. DFT-SAPT Electrostatic (*E*(elec)) and Dispersion (*E*(disp)) Contributions and CCSD(T) Benchmark Interaction Energies for Selected Stacked Aromatic Complexes from the S66 Database

	E(elec), %	*E*(disp), %	Δ*E*(CCSD(T))
benzene···benzene	19.2	73.5	−2.72
benzene···pyridine	24.0	68.6	−3.34
pyridine···pyridine	27.1	65.4	−3.80
benzene···uracil	29.6	63.9	−5.59
pyridine···uracil	34.0	59.4	−6.70
uracil···uracil	39.8	53.7	−9.75

electrostatic in nature. The increase in interaction strength is quite dramatic, with the uracil dimer having a binding energy that is more than 3.5 times that of the benzene dimer. The change in the electrostatic character of these interactions is also substantial, ranging from about 19% of attraction (benzene dimer) to about 40% of attraction (uracil dimer). Despite the increased relative importance of electrostatic effects in the heterocyclic dimers, it should be noted that both the electrostatic and the dispersion components of the total interaction energies increase substantially, with dispersion still playing a dominant role in the uracil dimer. There are two reasons for this increase in the dispersion energy for heterocyclic aromatic complexes. First, these complexes become more PS-like (and less PD-like) with the addition of heteroatoms, meaning that the contact area between the two monomers increases, thus increasing the dispersion attraction. Second, the additional electrostatic attraction brought about by the substitution of heterocyclic atoms into an aromatic complex generally produces a contraction of the interplane distance between the monomers, which also results in an increase in the dispersion energy. It has been demonstrated that the interplanar separation of the PS uracil dimer (∼3.4 Å) is shorter than that of the PD benzene dimer (∼3.6 Å).

Dimers of Nucleic Acid Bases, DNA

Purines (guanine and adenine), as well as pyrimidines (cytosine, thymine, uracil), are heterocyclic aromatic systems in which aromatic ring(s) are planar while exocyclic amino groups are nonplanar (Figure 3 depicts a GC···GC nucleoside step). As noted above, the presence of heteroatoms in aromatic systems leads to stronger electrostatic interactions and more specific relative orientations. Because of the

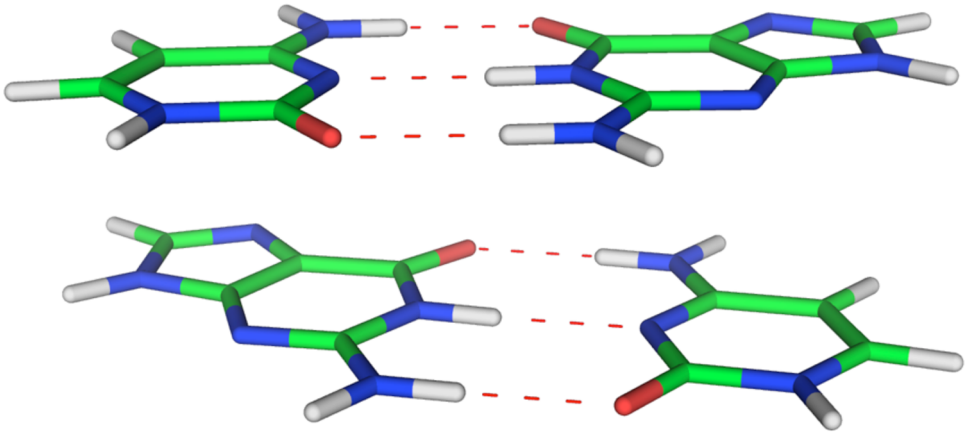


FIGURE 3. GC–GC base pair step from the structure of DNA.

presence of heteroatoms, the unsymmetrical bases possess dipole moments as well as rather high quadrupole moments. Each system contains several proton donor and proton acceptor groups, which is a prerequisite for the existence of strong H-bonds between bases. The electrostatic origin of stabilization in H-bonded base pairs has long been widely accepted. However, since planar nucleic bases have large polarizabilities, much larger than that of benzene, they can also interact via stacking interactions. Unlike H-bonding, base stacking is determined by an interplay of the three most commonly encountered molecular interactions: dispersion, electrostatic, and short-range exchange repulsion. Despite the fact that the existence of stacking has long been known, its strength and, most importantly, its role in stabilizing DNA double helices has not been understood until recently. This is partially because, in this case (as opposed to the case of the benzene dimer), no relevant experiments existed and elucidation of the role of stacking could come only from QM calculations. MP2 calculations performed at the complete basis set limit (CBS) overestimate the stacking attraction while providing accurate H-bonding stabilization energies. Thus, reliable stabilization energies for H-bonding and stacking of nucleic acids (as well as for other interaction motifs) could only be obtained at the CCSD(T)/CBS level.[3,31] The resulting stabilization energies were very large, much larger than published previously.[32] H-bonding energies for Watson–Crick G···C and A···T pairs (−28.8 and −15.4 kcal/mol, respectively) are still larger than the respective stacking energies (−16.9 and −11.6 kcal/mol, respectively), but the fact that a complex could have a stabilization of about 17 kcal/mol without any H-bonding was not readily accepted by many practitioners in the field. Nucleic acid bases possess large dipole moments and dipole–dipole interactions play a key role in H-bond

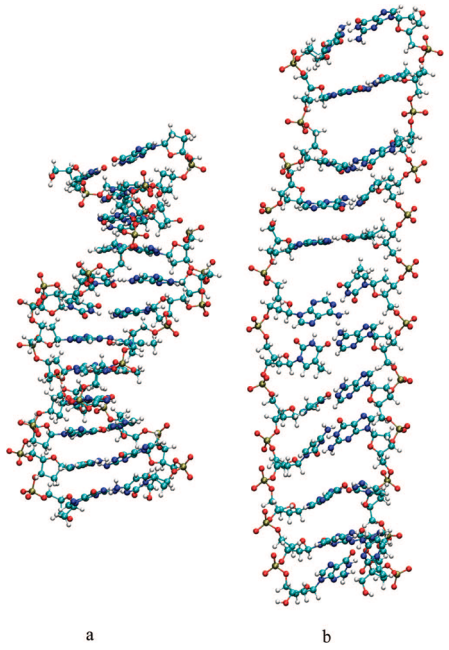


FIGURE 4. Snapshot figure of crystal double helical structure of DNA (a) and final ladder-like structure (b) after force field MD simulations where force-field dispersion energy was eliminated.

stabilization. The role of dipole–dipole electrostatic interactions in stacking is smaller, and here dispersion plays a key role. The electrostatic energy is, however, not negligible; it determines the relative orientation in the stacked pair. The importance of the dispersion energy in stacked complexes can be easily proven by considering the fact that no stacked minimum is obtained when optimization is performed using a method that does not describe the dispersion energy (e.g., Hartree–Fock or standard DFT methods). The consequences for the double-helical structure of DNA, which is essential for its basic function, that is, storing and transfer of genetic information, is enormous. Without dispersion energy, a double-helical structure will unwind (Figure 4).[33] This is also

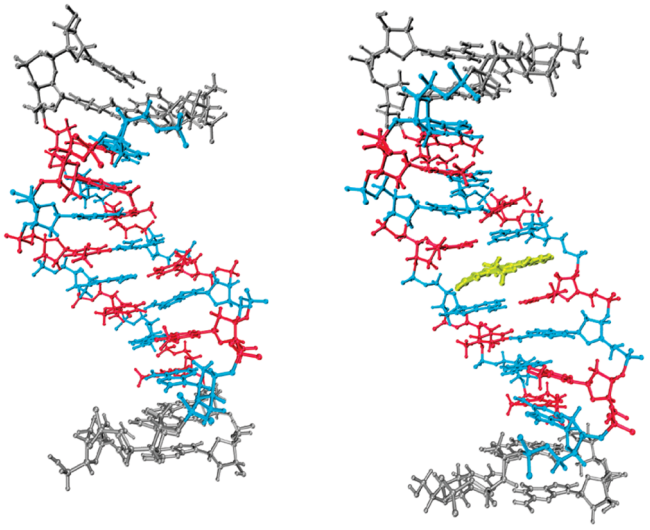


FIGURE 5. Snapshot figure of the double-helical structure of DNA without (left) and with (right) an ellipticine intercalator (yellow).

partially because dispersion energy, unlike the electrostatic energy, is not damped by passing from the gas phase to a water environment. This surprising finding is fully confirmed by biological experiments showing that base stacking is the main stabilizing factor in the DNA double helix.[34] There is another independent experiment showing the importance of stacking that concerns the surprising stability of DNA where natural purines or pyrimidines are replaced by unnatural nucleobases.[35] The melting point of DNA containing hydrophobic unnatural bases is comparable or even higher than that of the original DNA containing natural bases. Since unnatural hydrophobic bases either do not form H-bonds or form weaker H-bonds than in the case of complementary base pairs, the higher stability of modified DNA cannot be due to H-bonding but should be mainly assigned to stacking (the smaller desolvation energy of unnatural nucleobases plays a role as well). Evidently, aromatic stacking of nucleic acid bases is one of the key players in determining the structure and dynamics of nucleic acids. We believe this finding represents one of the most important demonstrations of the concept for aromaticity in science.

Stacking interactions not only are responsible for the stability, and thus function, of DNA but also play an important role in intercalation. Intercalation is the process by which a planar aromatic molecule strongly binds to DNA by inserting itself between adjacent base-pair steps of a nucleic-acid double helix (Figure 5). Recently there has been a great deal of interest in the behavior of intercalators because of their mutagenic, teratogenic, and carcinogenic effects as well as their antitumor and antiviral pharmacologic activity. The dominant energy contribution with all intercalators, neutral or charged, is represented by the dispersion energy.[24] As might be expected, the electrostatic energy is not negligible, and it also determines the relative orientation in a stacked pair.

Aromatic Interactions in Protein Structure

Interactions between the aromatic amino acids have been recognized as being particularly significant in the structure and function of proteins.[36] There are several characteristics of aromatic–aromatic interactions in proteins that distinguish them from other interaction types, such as H-bonds and aliphatic–aliphatic dispersion interactions. The three neutral amino acids, phenylalanine, tyrosine, and tryptophan, are generally quite hydrophobic, meaning that they tend to reside within a protein's interior regions. Unlike their aliphatic counterparts, aromatic amino acids usually interact in geometrically specific ways, forming pairs that are arranged in geometries approaching either the stacked (PD) or T-shaped motifs. This specificity of interaction geometry may be of great importance in overall protein structure, with the positions of specific aromatic pairs dictating secondary and tertiary structural elements of a protein.

Interactions between aromatic amino acids are involved in the stabilization of several protein structural elements, including the hydrophobic cores of globular proteins,[37,38] helical bundles,[39,40] and β-hairpins,[41−43] and have also been shown to be important in protein–ligand complexes.[44,45] The unique properties of aromatic residues contribute to the specificity of these interactions and, thus, can modify the overall structure of a protein in several different ways. Through protein data bank analysis, it has been observed that aromatic residues often occur in clusters with more than two interacting aromatic groups.[46] This is a significant finding and may point to the important role that aromatic amino acids play in protein structure, because conglomerations of aromatic groups have a much larger capacity to stabilize a protein than simple aromatic pairs.

There are several phenomena that contribute to the overall stability of a protein in its folded state, the most important of which are solvation effects (hydrophobic effect) and enthalpic effects (inter-residue interactions). The past several years have seen many arguments favoring either the hydrophobic effect or inter-residue interactions as being dominant in stabilizing protein structures. This question is far from resolved, and it seems that both phenomena are important, but modern computational chemistry techniques have been able to shed some light onto the issue. Much of

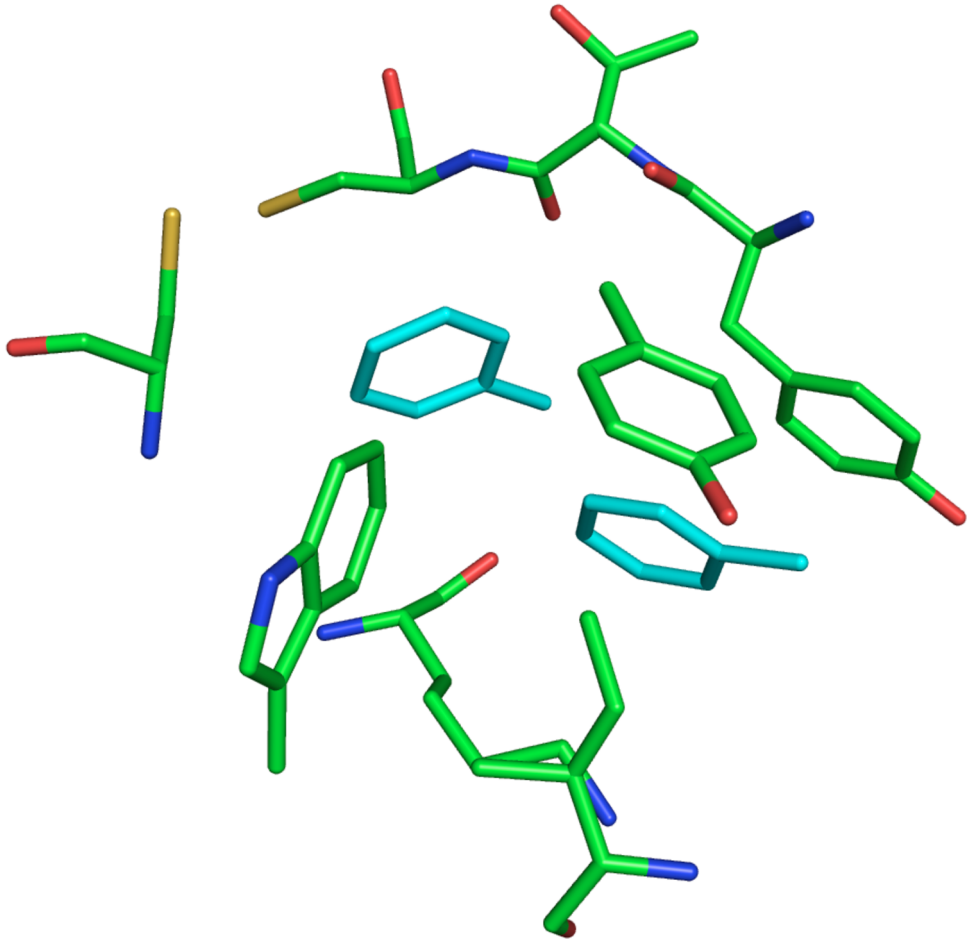


FIGURE 6. Hydrophobic core of the thermophilic protein rubredoxin. The central phenylalanines (in blue) interact with five and seven amino acids, respectively. Some backbone atoms removed for clarity.

the research done in our laboratory, and in others, points toward enthalpic inter-residue interactions as being critically important to the stability of a folded protein.[37,38,47−51]

Within the interior of a globular protein is found a conglomeration of nonpolar amino acids known as the hydrophobic core. The formation of the hydrophobic core was long believed to be a consequence of exterior hydrophobic forces of an entropic nature. We have shown that the stabilization inside the hydrophobic core of the thermophilic protein rubredoxin can in large part be attributed to the noncovalent interactions of the nonpolar amino acids within the protein core. This core is dominated by aromatic residues and contains several aromatic−aromatic interactions (Figure 6). The total stabilization within the core is estimated (at the MP2/CBS level) to be 51.2 kcal/mol in the gas phase, 28.7 kcal/mol in implicit ether solvent (which mimics a protein core environment), and 24.0 kcal/mol in implicit water solvent.[37,38]

The role of dispersion in the trypophan cage miniprotein, whose core is dominated by aromatic−aromatic and aromatic−aliphatic interactions, has been studied using MD simulations, based both on force-field (molecular modeling) potentials and on a more reliable QM/MM potential.[50] As in the case of double-helical DNA described above, suppression of the dispersion term led to protein denaturation (unfolding). When the dispersion term was restored, the protein refolded into its native structure very quickly. This study shows the importance of dispersion interactions (specifically those involving aromatic residues) in protein stability, as well as folding/unfolding equilibrium.

Noncovalent interactions between two (or more) amino acids located on two distinct secondary structural elements of a protein contribute to the establishment of a protein's tertiary structure. A specific example of the importance of inter-residue interactions in stabilizing tertiary structures can be found in the transmembrane proteins, which contain a bundle of several (usually six) α-helices that all traverse the phospholipid membrane.[39,40] Because a membrane protein resides, in large part, within a lipid environment, the hydrophobic effect cannot be a dominant contributor to stability.

Instead, inter-residue interactions, including hydrogen bonds, aliphatic dispersion contacts, salt bridges, aromatic–aromatic interactions, and cation–π interactions, play the dominant role in establishing the final conformations of these proteins. Interactions involving aromatic residues have been identified as being particularly important, and the presence of aromatic residues in the α-helices of some transmembrane proteins has been determined to be critical to the establishment of tertiary structure.[40]

Aromatic Interactions in Protein–Nucleic Acid Complexes

Protein–DNA/RNA interactions are important in a variety of biological processes, such as translation, transcription, and DNA repair, and are also critical in the structure of the ribosome. Interactions between aromatic amino acids and the heterocyclic rings of nucleobases have been shown to be important in protein–DNA/RNA interactions and may contribute to the recognition of nucleic acid initiation sequences.[50,52,53]

Analysis of the Protein Data Bank for protein–nucleic acid interactions indicates that aromatic residues may play some part in the stabilizing protein–DNA/RNA complexes and, perhaps more importantly, in the recognition of a GC or AT (AU) pair.[52] Histidine was found to form the greatest number of contacts with G and T, while phenylalanine was observed to have a large number of contacts with T and A. This may be particularly important in the recognition of specific sequences, such as in the TATA box unit, which is a common recognition site for initiation of transcription.

Wetmore and co-workers have thoroughly investigated the interactions between the aromatic units contained within nucleic acids and proteins.[53,54] Potential energy surfaces for all possible amino acid–nucleoside pair combinations were produced for both stacked and T-shaped geometries. The strengths of these aromatic–aromatic interactions involving heterocyclic groups were found to be comparable to hydrogen bonds in biological systems. It was also found that the strength of the interaction in an amino acid–nucleoside complex is much more strongly dependent on the relative parallel displacement and relative rotational orientation of the monomers than on their vertical separation.

In terms of binding energies, this study found no binding selectivity among the amino acid–nucleoside pairs. That is, for these model complexes, there is no propensity for phenylalanine to interact with A or T or for histidine to interact with G or C. However, these aromatic interactions were found to be strong enough to contribute significantly to the strength of a protein–DNA complex. It should be noted that the chemical environment, including both the neighboring protein and nucleic acid frameworks and the solvent, may play some role in determining the propensity of a given amino acid to interact with a particular nucleic acid.

Conclusions

The structures and stabilization energies of aromatic systems are determined mainly by dispersion and, to a lesser extent, electrostatic energies. Consequently, the stacking of aromatic systems plays an important role in many aspects of biology. Specifically, stacking (and dispersion) is responsible for the double helical structure of DNA, for the structure of DNA with aromatic intercalators, for the structure of hydrophobic cores of proteins, and also for the structure of DNA–protein complexes. These findings have fundamental consequences, since the structure of biomacromolecules is responsible for their function (e.g., the double helical structure of DNA is essential for the storage and transfer of genetic information).

This work was a part of research Project No. RVO: 613 889 63 of the Institute of Organic Chemistry and Biochemistry, ASCR, and was supported by the Operational Program Research and Development for Innovations - European Regional Development Fund (Project CZ.1.05/2.1.00/03.0058 of the MEYS of the CR). The support of Praemium Academiae, ASCR, awarded to P.H. in 2007 is acknowledged. This work was also supported by the Czech Science Foundation (P208/12/G016) and by NRF (WCU: R32-2008-000-10180-0).

BIOGRAPHICAL INFORMATION

Kevin E. Riley was born in New Mexico in 1976. He is currently an Assistant Professor of Chemistry at Xavier University of Louisiana. His main research interests relate to noncovalent interactions and their role in biomolecular systems, such as proteins and protein–ligand complexes.

Pavel Hobza was born in Prerov, Czech Republic in 1946. He is currently Distinguished Chair at the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague. His research focuses on noncovalent interactions and their applications to biodisciplines.

FOOTNOTES

The authors declare no competing financial interest.

REFERENCES

1 Krishtal, A.; Senet, P.; Van Alsenoy, C. Local softness, softness dipole, and polarizabilities of functional groups: Application to the side chains of the 20 amino acids. *J. Chem. Phys.* **2009**, *131*, No. 044312.

2 Bloom, J. W. G.; Wheeler, S. E. Taking the aromaticity out of aromatic interactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 7847–7849.

3 Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.* **2010**, *110*, 5023–5063.

4 Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation-theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.

5 Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential-energy surface of the benzene dimer: Ab initio theoretical-study. *J. Am. Chem. Soc.* **1994**, *116*, 3500–3506.

6 Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential energy surface for the benzene dimer. Results of ab initio CCSD(T) calculations show two nearly isoenergetic structures: T-shaped and parallel-displaced. *J. Phys. Chem.* **1996**, *100*, 18790–18794.

7 Spirko, V.; Engvist, O.; Soldan, P.; Selzle, H. L.; Schlag, E. W.; Hobza, P. Structure and vibrational dynamics of the benzene dimer. *J. Chem. Phys.* **1999**, *111*, 572–582.

8 Hobza, P.; Selzle, H. L.; Schlag, E. W. Structure and properties of benzene-containing molecular clusters: Nonempirical ab-initio calculations and experiments. *Chem. Rev.* **1994**, *94*, 1767–1785.

9 Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential energy surface for the benzene dimer and perturbational analysis of $\pi-\pi$ interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

10 Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

11 Pitonak, M.; Neogrady, P.; Rezac, J.; Jurecka, P.; Urban, M.; Hobza, P. Benzene dimer: High-level wave function and density functional theory calculations. *J. Chem. Theory Comput.* **2008**, *4*, 1829–1834.

12 Sinnokrot, M. O.; Sherrill, C. D. Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations. *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

13 Sinnokrot, M. O.; Sherrill, C. D. High-accuracy quantum mechanical studies of $\pi-\pi$ interactions in benzene dimers. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.

14 Pluhackova, K.; Jurecka, P.; Hobza, P. Stabilisation energy of $C_6H_6\cdots C_6X_6$ (X = F, Cl, Br, I, CN) complexes: Complete basis set limit calculations at MP2 and CCSD(T) levels. *Phys. Chem. Chem. Phys.* **2007**, *9*, 755–760.

15 Tsuzuki, S.; Uchimaru, T.; Mikami, M. Intermolecular interaction between hexafluorobenzene and benzene: Ab initio calculations including CCSD(T) level electron correlation correction. *J. Phys. Chem. A* **2006**, *110*, 2027–2033.

16 Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Estimates of the ab initio limit for $\pi-\pi$ interactions: The benzene dimer. *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893.

17 Hohenstein, E. G.; Sherrill, C. D. Effects of heteroatoms on aromatic $\pi-\pi$ interactions: Benzene–pyridine and pyridine dimer. *J. Phys. Chem. A* **2009**, *113*, 878–886.

18 Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Highly accurate CCSD(T) and DFT-SAPT stabilization energies of H-bonded and stacked structures of the uracil dimer. *ChemPhysChem* **2008**, *9*, 1636–1644.

19 Rezac, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

20 Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of assembly phenomena by aromatic–aromatic interactions: Benzene dimer and the substituted systems. *J. Phys. Chem. A* **2007**, *111*, 3446–3457.

21 Riley, K. E.; Merz, K. M. Effects of fluorine substitution on the edge-to-face interaction of the benzene dimer. *J. Phys. Chem. B* **2005**, *109*, 17752–17756.

22 Ringer, A. L.; Sherrill, C. D. Substituent effects in sandwich configurations of multiply substituted benzene dimers are not solely governed by electrostatic control. *J. Am. Chem. Soc.* **2009**, *131*, 4574–4575.

23 Raju, R. K.; Bloom, J. W. G.; An, Y.; Wheeler, S. E. Substituent effects on non-covalent interactions with aromatic rings: Insights from computational chemistry. *ChemPhysChem* **2011**, *12*, 3116–3130.

24 Hohenstein, E. G.; Duan, J. N.; Sherrill, C. D. Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers. *J. Am. Chem. Soc.* **2011**, *133*, 13244–13247.

25 Watt, M.; Hardebeck, L. K. E.; Kirkpatrick, C. C.; Lewis, M. Face-to-face arene–arene binding energies: Dominated by dispersion but predicted by electrostatic and dispersion/polarizability substituent constants. *J. Am. Chem. Soc.* **2011**, *133*, 3854–3862.

26 Wheeler, S. E. Local nature of substituent effects in stacking interactions. *J. Am. Chem. Soc.* **2011**, *133*, 10262–10274.

27 Rezac, J.; Riley, K. E.; Hobza, J. Extensions of the S66 data set: More accurate interaction energies and angular-displaced nonequilibrium geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

28 Hobza, J. Calculations on noncovalent interactions and databases of benchmark interaction energies. *Acc. Chem. Res.* **2012**, *45*, 663–672.

29 Hesselmann, A.; Jansen, G. First-order intermolecular interaction energies from Kohn-Sham orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464–470.

30 Hesselmann, A.; Jansen, G.; Schutz, M. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, No. 014103.

31 Hobza, P.; Sponer, J. Structure, energetics, and dynamics of the nucleic acid base pairs: Nonempirical ab initio calculations. *Chem. Rev.* **1999**, *99*, 3247–3276.

32 Jurecka, P.; Hobza, P. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine···cytosine, adenine···thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.

33 Cerny, J.; Kabelac, P.; Hobza, P. Double-helical → ladder structural transition in the B-DNA is introduced by a loss of dispersion energy. *J. Am. Chem. Soc.* **2008**, *130*, 16055–16059.

34 Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564–574.

35 Kool, E. T. Replacing the nucleohases in DNA with designer molecules. *Acc. Chem. Res.* **2002**, *35*, 936–943.

36 Burley, S. K.; Petsko, G. A. Aromatic-aromatic interaction - a mechanism of protein-structure stabilization. *Science* **1985**, *229*, 23–28.

37 Riley, K. E.; Merz, K. M. Role of solvation in the energy stabilization inside the hydrophobic core of the protein rubredoxin. *J. Phys. Chem. B* **2006**, *110*, 15650–15653.

38 Vondrasek, J.; Bendova, L.; Klusak, V.; Hobza, P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated ab initio quantum chemical calculations. *J. Am. Chem. Soc.* **2005**, *127*, 2615–2619.

39 Fiedler, S.; Broecker, J.; Keller, S. Protein folding in membranes. *Cell. Mol. Life Sci.* **2010**, *67*, 1779–1798.

40 Haeger, S.; Kuzmin, D.; Detro-Dassen, S.; Lang, N.; Kilb, M.; Tsetlin, V.; Betz, H.; Laube, B.; Schmalzing, G. An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nat. Struct. Mol. Biol.* **2010**, *17*, 90–U116.

41 Hughes, R. M.; Waters, M. L. Model systems for beta-hairpins and beta-sheets. *Curr. Opin. Struct. Biol.* **2006**, *16*, 514–524.

42 Santiveri, C. M.; Jimenez, M. A. Tryptophan residues: Scarce in proteins but strong stabilizers of beta-hairpin peptides. *Biopolymers* **2010**, *94*, 779–790.

43 Tatko, C. D.; Waters, M. L. Selective aromatic interactions in β-hairpin peptides. *J. Am. Chem. Soc.* **2002**, *124*, 9372–9373.

44 Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **2010**, *53*, 5061–5084.

45 Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic rings in chemical and biological recognition: Energetics and structures. *Angew. Chem., Int. Ed.* **2011**, *50*, 4808–4842.

46 Chourasia, M.; Sastry, G. M.; Sastry, G. N. Aromatic-aromatic interactions database, A(2)ID: An analysis of aromatic pi-networks in proteins. *Int. J. Biol. Macromol.* **2011**, *48*, 540–552.

47 Bendova-Biedermannova, L.; Hobza, P.; Vondrasek, J. Identifying stabilizing key residues in proteins using interresidue interaction energy matrix. *Proteins* **2008**, *72*, 402–413.

48 Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrasek, J. Representative amino acid side chain interactions in proteins. A comparison of highly accurate correlated ab initio quantum chemical and empirical potential procedures. *J. Chem. Theory Comput.* **2009**, *5*, 982–992.

49 Berka, K.; Laskowski, R. A.; Hobza, P.; Vondrasek, J. Energy matrix of structurally important side-chain/side-chain interactions in proteins. *J. Chem. Theory Comput.* **2010**, *6*, 2191–2203.

50 Cerny, J.; Vondrášek, J.; Hobza, P. Loss of dispersion energy changes the stability and folding/unfolding equilibrium of the Trp-cage protein. *J. Phys. Chem. B* **2009**, *113*, 5657–5660.

51 Vondrasek, J.; Kubar, T.; Jenney, F. E.; Adams, M. W. W.; Kozisek, M.; Cerny, J.; Sklenar, V.; Hobza, P. Dispersion interactions govern the strong thermal stability of a protein. *Chem.—Eur. J.* **2007**, *13*, 9022–9027.

52 Baker, C. M.; Grant, G. H. Role of aromatic amino acids in protein-nucleic acid recognition. *Biopolymers* **2007**, *85*, 456–470.

53 Churchill, C. D. M.; Rutledge, L. R.; Wetmore, S. D. Effects of the biological backbone on stacking interactions at DNA-protein interfaces: The interplay between the backbone···π and $\pi\cdots\pi$ components. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14515–14526.

54 Rutledge, L. R.; Durst, H. F.; Wetmore, S. D. Evidence for stabilization of DNA/RNA–protein complexes arising from nucleobase-amino acid stacking and T-shaped interactions. *J. Chem. Theory Comput.* **2009**, *5*, 1400–1410.