PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

CH/π hydrogen bonds in biological macromolecules

Motohiro Nishio,[1] Yoji Umezawa,[2] Jacques Fantini,[3] Manfred. S. Weiss[4] and Pinak Chakrabarti[5]

This is a sequel to the previous *Perspective* "*The CH/π hydrogen bond in chemistry. Conformation, supramolecules, optical resolution and interactions involving carbohydrates*", which featured in a PCCP themed issue on "*Weak Hydrogen Bonds - Strong Effects?*": *Phys. Chem. Chem. Phys.*, 2011, **13**, 13873–13900 (Ref. 11 in the text).

Evidence that weak hydrogen bonds play an enormously important role in chemistry and biochemistry has now accumulated to an extent that the rigid classical concept of hydrogen bonds formulated by Pauling needs to be seriously revised and extended. The concept of a more generalized hydrogen bond definition is indispensable for understanding the folding-mechanisms of proteins. The CH/π hydrogen bond, a weak molecular force occurring between a soft acid CH and a soft base π-electron system, among all, is one of the most important and plays a functional role in defining the conformation and stability of 3D structures as well as in many molecular recognition events. This concept is also valuable in structure-based drug design efforts. Despite their frequent occurrence in organic molecules and bio-molecules the importance of CH/π hydrogen bonds is still largely unknown to many chemists and biochemists. Here we present a review that deals with the evidence, nature, characteristics and consequences of

[1] *The CHPI Institute, 705-6-338, Minamioya, Machida-shi, Tokyo 194-0031, Japan. E-mail: dionisio@tim.hi-ho.ne.jp*

[2] *Institute of Microbial Chemistry (BIKAKEN) Tokyo, 3-14-23 Kamiosaki, Shinagawa-ku, Tokyo 141-0021, Japan. E-mail: umezaway@bikaken.or.jp*

[3] *Université d'Aix-Marseille, EA-4674, Faculté des Sciences et Techniques St-Jerome, Marseille, France. E-mail: jacques.fantini@univ-amu.fr*

[4] *Helmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography (HZB-MX), Albert-Einstein-Str. 15, D-12489 Berlin, Germany. Email: msweiss@helmholtz-berlin.de*

[5] *Department of Biochemistry, Bose Institute, P1/12 CIT Scheme VIIM, Kolkata 700 054, India. E-mail: pinak@jcbose.ac.in*

the CH/π hydrogen bond in biological macromolecules (proteins, nucleic acids, lipids and polysaccharides). It is hoped that the present *Perspective* will show the importance of CH/π hydrogen bonds and stimulate interest in the interactions of biological macromolecules, one of the most fascinating fields in bioorganic chemistry. Implication of this concept is enormous and valuable in the scientific community.

Motohiro Nishio *obtained his Ph.D. in physical organic chemistry from the University of Tokyo in 1965. He worked at Meiji Seika Kaisha Ltd (1959-1996), after which he was a visiting professor for three years at Chiba University. During that period he also acted as a visiting scientist (1996-2000) at the Institute of Physical and Chemical Research (RIKEN). In 1999 he established the CHPI Institute in his home address. He has published 93 papers, 21 reviews, 4 book chapters, and 5 books (4 in Japanese). His research interests include stereochemistry, weak hydrogen bonds in chemistry and structural biology, and structure-based drug design.*

Yoji Umezawa *graduated in 1974 from the Faculty of Engineering, Keio University, Japan and subsequently obtained his Ph.D. in synthetic organic chemistry from Keio University in 1982. He has been working at Institute of Microbial Chemistry (BIKAKEN)*

from 1975. He has published 51 papers, 7 reviews and 1 book (with M. Nishio and M. Hirota). His research interests include weak hydrogen bonds and rational drug design.

Jacques Fantini *is a professor of Biochemistry at Aix-Marseille University. He is the leader of the research team "Molecular Interactions in Biological Membranes". He is particularly interested in aiding the elucidation of the molecular mechanisms controlling the interaction of proteins with membrane lipids, especially sphingolipids and cholesterol. With his long-time teammate Prof. Dr. Nouara Yahi, he has discovered a universal protein fold, referred to as the sphingolipid-binding domain (SBD), which mediates the selective association a wide range of cellular and pathogen-associated proteins with lipid rafts. He is the author of 160 articles (PubMed), has been cited 5500 times and has a current h-index of 40.*

Manfred S. Weiss *obtained his Ph.D. in chemistry from the Albert-Ludwigs-Universität Freiburg i. Br. (Germany) in 1992. He is currently a holding a permanent scientist position in macromolecular crystallography at the Helmholtz-Zentrum Berlin (HZB), which is one of the two German institutes operating a synchrotron storage ring with facilities for macromolecular crystallography. He has authored or co-authored more than 100 scientific publications in peer-reviewed journals (current H-index 30). Since 2002 he is also co-editor of Acta Crystallographica Section D and since 2008 he is Section Editor of Acta Crystallographica Section F, two journals, which are among the leading journals in the field of macromolecular crystallography.*

Pinak Chakrabarti *obtained his Ph.D in organic chemistry from the Indian Institute of Science in 1981. For his post-doctoral work he was associated with Profs. Jack Dunitz, Michael Rossmann and Doug Rees. Currently he is a professor of Biochemistry, and in-charge of the Bioinformatics Centre at Bose Institute. He is the author of 130 publications. His research interests include protein structure and conformation, protein-protein interaction, molecular recognition, protein-nanoparticle interaction, Vibrio cholera proteins, structural bioinformatics.*

1. Introduction

1.1 Three-dimensional structure

The conformation of both small and large organic compounds, biological macromolecules, as well as the specificity of molecular recognition are governed by the same kinds of intermolecular forces, attractive as well as repulsive. It should therefore be

possible to explain the three-dimensional structures of these compounds and their recognition specificities on the basis of common principle.

"The conformation of both small and large organic compounds, biological macromolecules, as well as the specificity of molecular recognition are governed ..."

1.2 Weak molecular interactions

Before moving towards the discussion of interactions involving biological macromolecules, a brief introduction to weak molecular forces shall be given. The most important of these forces are the van der Waals force, electrostatic interactions and the hydrogen bond.

1.2.1 Van der Waals force. Among weak intermolecular interactions, the van der Waals force is most general and ubiquitous. This concept has originated from the work of van der Waals, in the 19th century, who investigated the physical origin of the difference of the properties of real gasses such as Ar, N_2, CO_2, *etc*., from the ideal gas theory. Today, the van der Waals force is understood as a blend of various intermolecular interactions. The attractive part of the van der Waals force includes the dispersion force, electrostatic forces such as Coulombic, dipole/dipole interactions, *etc*. Typically, the largest contribution is from the dispersion force: 100% for neutral compounds such as noble gasses, methane, propane, *etc*., but somewhat smaller in the cases of compounds bearing partial charges or dipole moment (99, 96, 86, 68, 57 and 24 %, respectively, for HI, HBr, HCl, CH_3Cl, NH_3 and H_2O).[1]

1.2.2 Electrostatic interactions. Among the electrostatic interactions, the Coulombic, the dipole/dipole and the dipole/quadrupole interactions are the most important ones and need to be mentioned. The Coulombic energy is the strongest of these, and it exhibits a distance dependence of $1/r$. For further details, interested readers may be referred to the comprehensive review on electrostatic interactions contributed by Israelachvili.[1]

1.2.3 The hydrogen bond. According to Pauling, the hydrogen bond is formed between X–H and Y, where X and Y are electronegative atoms such as O and N.[2] In this so-called classical or conventional hydrogen bond, the largest energetic contribution comes from

the Coulomb energy.[3,4] On the other hand, Pimentel and McClellan argued that a hydrogen bond exists between a donor A–H and an acceptor B, when there is evidence of bond formation and that this new bond specifically involves the hydrogen atom already bonded to A (Fig. 1).[5] This definition does not restrict the chemical nature of the donors and acceptors, the energy or the geometry of the participants. This very general definition has proven to be very useful to the progress of modern chemistry and biology. While the hydrogen bond covers very wide energy range, the ones familiar to organic chemists and biochemists, range from *ca.* 1-7 kcal mol^{-1} per a one-unit interaction.[6] Nowadays, the hydrogen bond is recognized as a much broader phenomenon than envisaged earlier.[7] One of the aims of this review is to demonstrate that the concept of Pimentel and McClellan is particularly suited to explain hydrogen bond formation given all the knowledge that has been accumulated on this topic in the past three or four decades.

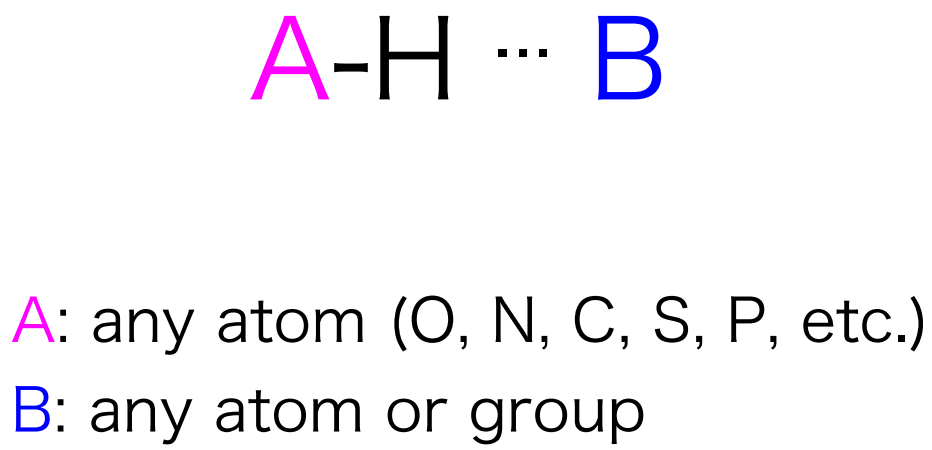


Fig. 1 Definition of the hydrogen bond according to Pimentel and McClellan.

In 1966, Pearson argued that the strength of an interaction depends on hardness and softness of the components, acids and bases (HSAB principle).[8] Nishio and Hirota classified the hydrogen bond into four categories: HA/HB, HA/SB, SA/HB and SA/SB, according to this HSAB concept.[9] Desiraju and Steiner wrote a comprehensive treatise on weak hydrogen bonds.[10] As for the XH/π and CH/*n* hydrogen bonds (a SA/HB combination), see Section 1.1 of our previous *Perspective*.[11] Examples have been reported in which the peptide NH/π hydrogen bonds (a HA/SB combination) are of

functional importance in the stabilization of protein α-helix termini, β-strand ends, β-strand edges, β-bulges and regular turns. Further, the occurrence of XH/π interactions involving the N-H, O-H and S-H donors in proteins was analysed and it has been found that about one such interaction occurs per 10.8 aromatic amino acid residues.[12]

1.3 The CH/π hydrogen bond

More recently, yet another molecular force, the CH/π hydrogen bond (a SA/SB combination), has been recognized to play significant roles in organic chemistry.[13] In 1977 and subsequent years, Nishio presented a hypothesis that CH/π hydrogen bonds, involving aliphatic and aromatic CHs as the hydrogen donor, bear implications in a variety of molecular scenarios, including chemical as well as biochemical backgrounds.[14, 15] In short, a notable feature of the CH/π hydrogen bond is that the dispersion energy contributes most in typical cases where an aliphatic CH or an aromatic CH is the hydrogen donor; this is *not* another term of the van der Waals force. The dual nature of the interaction, with dispersion and electrostatic terms, is the basis for the importance of this attractive molecular force in chemistry and biochemistry. Another noteworthy characteristic of the CH/π hydrogen bond is that it works in polar, protic solvents such as *water*. This is of paramount importance in the consideration of the effect in biochemistry, structural biology and molecular biology. Reviews are available for the topics of conformation,[16,17,18] supramolecular chemistry,[19,20,21] selectivity of organic reactions,[22,23] optical resolution[24] and interactions involving carbohydrates.[25]

1.3.1 Different classes of the CH/π hydrogen bond. The energy of the CH/π hydrogen bond depends on the nature of the participating molecular fragments, the CH donor and the π-acceptor groups. For typical cases involving aliphatic and aromatic CH groups as the hydrogen donor and the C_6 aromatic ring as the CH-acceptor, the energy of a one unit CH/π hydrogen bond is ca. 1.5-2.5 kcal mol^{-1}, as obtained from high-level *ab initio* MO calculations.[26]

1.3.1.1 CH/π hydrogen bonds involving activated C-H groups. For CH/π hydrogen bonds involving activated or acidic C–H groups, the energy of interaction increases. For instance, the energy of a CH/π hydrogen bond involving acetylene or chloroform has

been estimated to be *ca.* 3-5 kcal mol^{-1}. These values are very similar to the classical or conventional hydrogen bonds.

1.3.1.2 CH/π hydrogen bonds involving typical C-H groups. The energy of CH/π hydrogen bonds involving typical CH donor groups in alkyl and aromatic groups is generally weaker than when activated donor groups are involved. An aromatic or sp^2-CH group is stronger than an aliphatic or sp^3-CH. Nevertheless, though nonconventional, the CH/π hydrogen bond is a genuine hydrogen bond. It should not be confused with the so-called 'anti-hydrogen bond'[27] nor the so-called 'improper-hydrogen bond'.[28] Barnes published critical comments on the above discussions.[29] Nakagawa suggested that interaction between the C–H dipole and the quadrupole of an aromatic group is also important.[30] In today's view, this would also qualify as an account supporting the CH/π hydrogen bond although it was not explicitly named in that way.

1.3.1.3 Aromatic CH/π hydrogen bond. A somewhat special case of the CH/π hydrogen bond is the CH/π hydrogen bond involving aromatic CH donor atoms and aromatic π–systems as acceptors. Such interactions are frequently referred to as 'edge-to-face', 'T-shape', 'π/π-' or 'arene/arene'-interactions.[31] However, if looked at in detail with respect to the position of all atoms involved and the direction of the bonds, it is absolutely clear that this interaction falls into the category CH/π hydrogen bond. Therefore, we refer to this as the 'aromatic CH/π hydrogen bond'.

As an example, the antigen-binding site of immunoglobulins (Fab fragment) is composed of a number of nonpolar residues. Novotny and Haber retrieved the crystal structures of Fab fragments NEW, KOL and MCPC603, VL-VL dimers of RHE and REI and Bence-Jones protein MCG from the Protein Data Bank (PDB).[32] There they found that invariant aromatic residues close-pack at the bottom of the binding-site with their ring planes perpendicularly oriented in the characteristic "herringbone" packing mode. In our view, this is indication of aromatic CH/π hydrogen bonds. To analyze the invariant features of the binding region (VL-VH domain interface), they compared the known immunoglobulin three-dimensional structures. The geometry of the interface is preserved *via* invariance of some aromatic side-chains on their surface of the domains.

As a second example, a representative set of X-ray crystal structures of nonhomologous proteins was examined to determine the preferred positions and orientations of

noncovalent interactions between the aromatic side-chains of Phe, Tyr, His and Trp.[33] It has been found that pairs of aromatic side-chain amino acids preferentially align their respective aromatic rings in an off-centered parallel orientation; this parallel-displaced structure is 0.5–0.75 kcal mol^{-1} more stable than a T-shaped structure for Phe interactions and 1 kcal mol^{-1} more stable than a T-shaped structure for the full set of aromatic side-chain amino acids. This result, however, is not in agreement with previously published analyses of aromatic amino acids in proteins.

Importance of attractive interactions involving π groups in biochemistry should be emphasized; examples illustrated were CH/π hydrogen bonds between Phe80 and Trp108 in myohemerythrin, those between Tyr94 and flavodoxin, Phe139 and Tyr125, Trp62 of lysozyme with its inhibitor N-acetylchitotriose, Phe179 and Tyr33, Phe134 and Trp113 in dihydrofolate reductase.[34] The nature of the π stacking was argued in relation to interactions including amyloid proteins.[35]

Sastry and coworkers recently examined networks formed by aromatic residues (Phe, Tyr, Trp and His) in proteins;[36,37] they retrieved 17911 crystal structures from the PDB and systematically examined the connectivity, the plane-plane angle (ϕ) of aromatic residues, and the plane distances *d*. There they found that the preference for right angle orientation is distinctly higher, between cutoff distance of 5.0 and 7.0 Å. A sudden change occurred in the ϕ angle distribution at ca. *d* 4.5 Å. It is clear that when the aromatic-aromatic distance is large enough to allow for a CH/π hydrogen bond to form, it will form. At shorter distance, π/π parallel stacking interactions are the dominant interactions. Fig. 2 illustrates the maximum length of networks and the interconnectivity presenting the proteins observed with different cutoff radii, 4.5-6.0 Å. This type interaction works cooperatively; the interplay of the aromatic CH/π hydrogen bonds undoubtedly contributes to stabilizing the protein 3D structure.

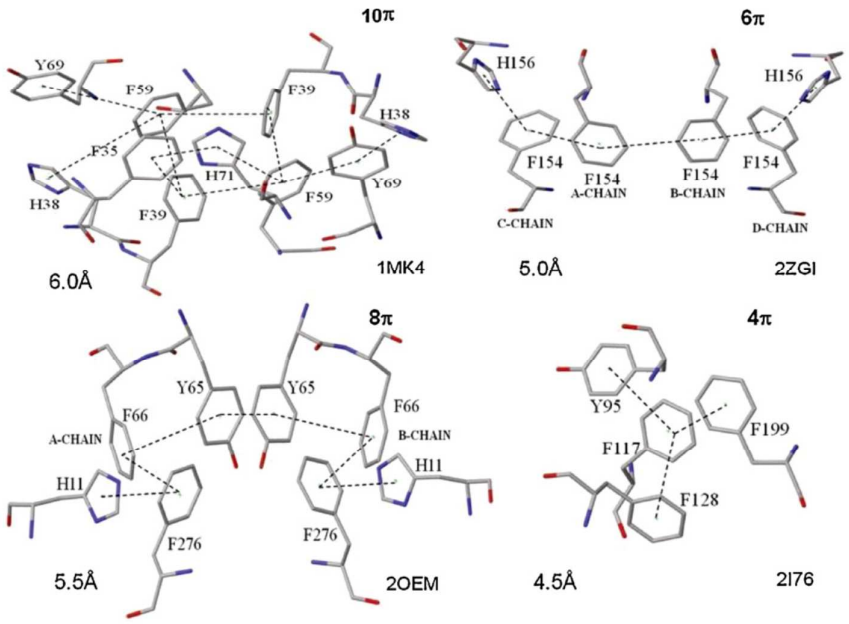


Fig. 2 Maximum length of networks and the interconnectivity presenting the proteins with different cut-off radii, 4.5, 5.0, 5.5 and 6.0 Å. From M. Chourasia *et al.*, *Int. J. Biol. Macromol.*, 2011, **48**, 540, with permission.

1.3.2 Geometric aspects of the CH/π hydrogen bond.

1.3.2.1 Program CHPI. To investigate CH/π hydrogen bonds in biological macromolecules, Umezawa and Nishio wrote a program termed "CHPI".[38] They retrieved a number of crystal structures from the PDB and examined, by CHPI, the occurrence of short distances between CH-groups and π-systems in protein structures. The protein 3D structures included RNase, 2-Zn insulin, BPTI, sperm whale myoglobin, human growth hormone and ligand/protein complexes such as *Ras* p21 G-proteins, transducin α, EF-Tu and *Src* homology 2 (SH2) domains of various sources. There, they used certain geometric criteria, which are illustrated in Fig. 3

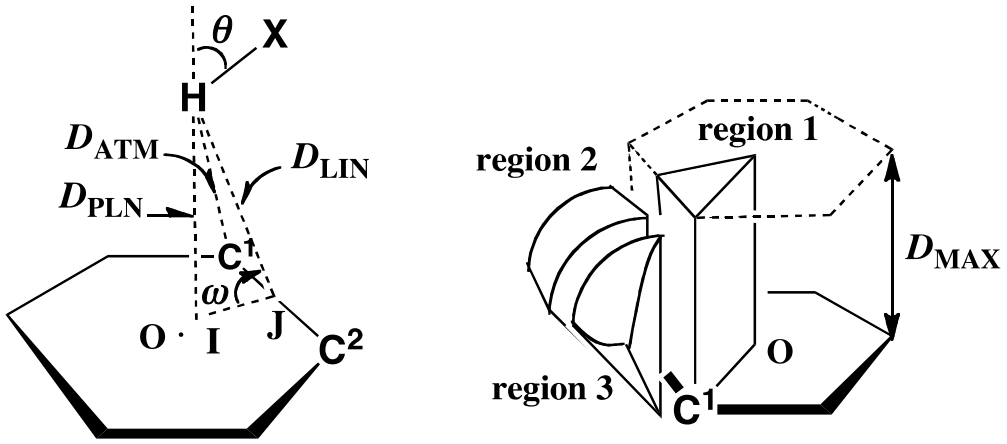


Fig. 3 Method of exploring CH/π contacts: (A) O: centre of the plane. C^1 and C^2: nearest and second nearest sp^2-carbons, respectively, to H. ω: dihedral angle defined by C^1OC^2 and HC^1C^2 planes. θ: 180°-∠X-H-I. D_{PLN}: H/π-plane distance (H/I). D_{ATM}: interatomic distance (H/C^1). D_{LIN}: distance between H and line C^1-C^2 (H/J). (B) 1: region where H is above the aromatic ring. 2 and 3: regions where H is out of region 1 but may interact with π-orbitals. The program was run to search for H/π distance shorter than a cut-off value D_{MAX} in every region: $D_{PLN} < D_{MAX}$, $\theta < 63°$, $|\omega| < 90°$ for region 1, $D_{LIN} < D_{MAX}$, $\theta < 63°$, $90° < |\omega| < 130°$ for region 2, and $D_{ATM} < D_{MAX}$, $\theta < 63°$, $50° < \phi < 90°$ for region 3 (ϕ: ∠H-C^1-I). This program (CHPI) has been implemented in ABINIT-MP software for free use (BioStation Viewer), by Centre for Research on Innovative Simulation Software: http://www.ciss.iis.u-tokyo.ac.jp/english/dl/index.php.

1.3.2.2 The Brandl Weiss system. In their large database study on CH/π hydrogen bonds in proteins published in 2001, Brandl, Weiss and coworkers introduced a slightly different geometric system for the identification of CH/π hydrogen bonds than the one presented in the previous paragraph.[39] Here three parameters are used: (1) the distance of the donor carbon atoms to the geometric centre of the π-acceptor system $d_{C\text{-}X}$, (2) the angle at the hydrogen atom of the donor carbon atom $\angle_{C\text{-}H\text{-}X}$ and the distance of the vertical projection of the donor H atom onto the acceptor π-system from the geometric centre of the acceptor π-system $d_{Hp\text{-}X}$. A schematic representation of these parameters is shown in Fig. 4

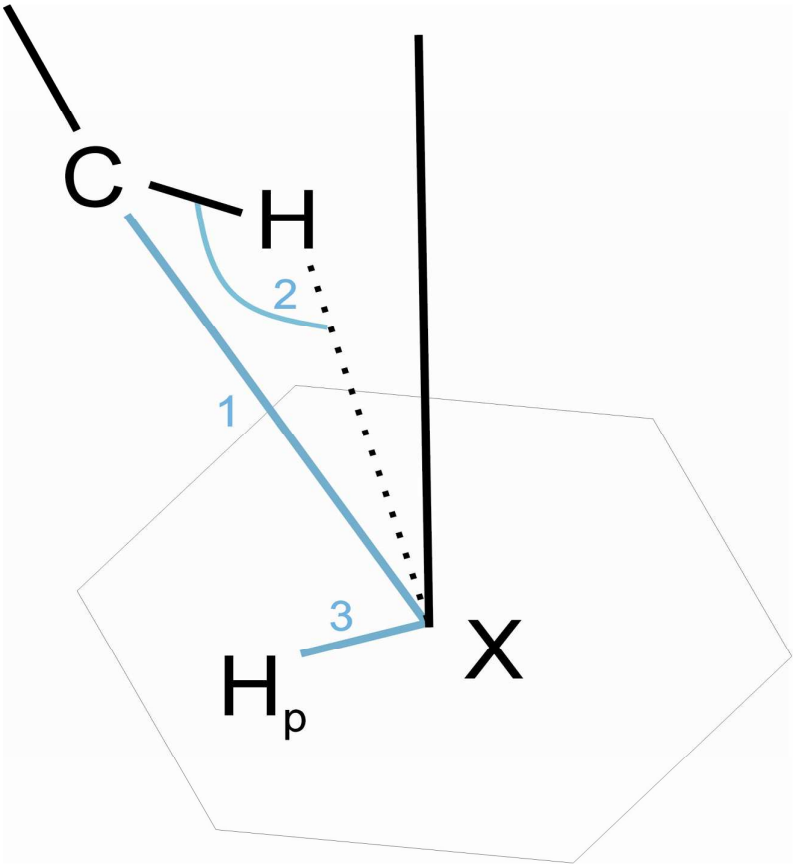


Fig. 4 The Brandl Weiss system for defining CH/π hydrogen bonds. Three criteria are employed: (1) the distance of the donor carbon atoms to the geometric centre of the π-acceptor system $d_{C\text{-}X}$, (2) the angle at the hydrogen atom of the donor carbon atom $\angle_{C\text{-}H\text{-}X}$ and (3) the distance of the vertical projection of the donor H atom onto the acceptor π-system from the geometric centre of the acceptor π-system $d_{Hp\text{-}X}$. From Brandl *et al.*, *J. Mol. Biol.,* 2001, **307**, 357, with permission.

Typically used cut-off values for these parameters are 4.5 Å for $d_{C\text{-}X}$, 120° for $\angle_{C\text{-}H\text{-}X}$ and 1.0 or 1.2 Å for different sized π-acceptor systems. While this system appears to be at first sight a bit more intuitive than the system implemented in the program CHPI discussed in the previous paragraph, because it is conceptually closer to the Baker and Hubbard system for defining a hydrogen bond,[40] it yields essentially the same results as the CHPI program. Although there is no obvious direct geometric relationship between the two systems, in practice it does not really matter which of the systems is employed to

identify a CH/π hydrogen bond. One aspect, which has to be taken into account, however, is that the most important parameter used in the program CHPI, namely the distance between the hydrogen donor atom and the acceptor π-system D_{MAX} is based on an "optimized" hydrogen position. Since hydrogen atoms are typically not seen in a macromolecular crystal structure determination, such a definition may be regarded as a potential weak point.

1.3.3 Direct experimental evidence for CH/π hydrogen bonds in proteins. While nearly all evidence for the occurrence and the importance of CH/π hydrogen bond in proteins until recently stems from geometric analysis of high-resolution structures of groups of structures or whole databases of protein structures, a few very important contributions appeared recently, which provide more direct (experimental) evidence for CH/π hydrogen bonds. The first such contribution was reported, in 2010, by nuclear magnetic resonance (NMR) in combination with isotope labeling (Fig. 5).[41] Based on the observation of *J*-coupling it has been shown that CH/π hydrogen bonds between methyl group and aromatic systems occur and are important.

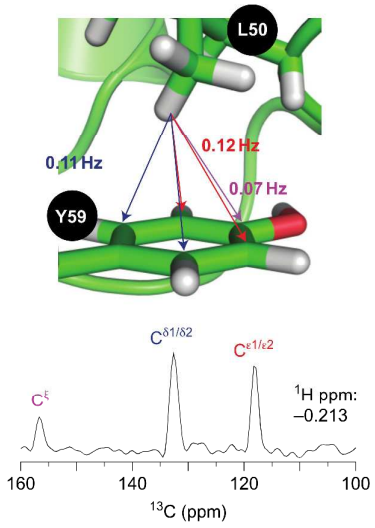


Fig. 5 CH/π hydrogen bond between a methyl group of Leu50 and the aromatic ring of Tyr59 in ubiquitin (top panel). The identification of weak scalar coupling signals in NMR experiments of isotope labeled proteins (bottom panel) unambiguously identifies this

interaction as a hydrogen bond. From Plevin *et al. Nat. Chem.,* 2010, **2**, 466, with permission.

Two further papers appeared in 2012. In the first contribution, a 35-residue model peptide, α_2D was studied.[42] The peptide α_2D has been shown to fold into a dimeric four-helix bundle containing two stacked Phe/Phe pairs in the core, where water interference does not exist. They introduced a Phe to cyclohexyl mutation and analysed the folding behaviour of the resulting protein and concluded that CH/π hydrogen bonds are energetically significant. Fig. 6 illustrates that the stacking residue pairs are sequestered in the core of the protein. Interaction energies between several sets of artificial peptide dimers were estimated and analyzed by using the van't Hoff analysis. The CH/π hydrogen bonds, occurring between Phe and cyclohexylalanine (Cha) in α_2D, have been found to significantly contribute to the stability of the dimers: they contribute as much as 2 kcal mol^{-1} to the overall protein stability.

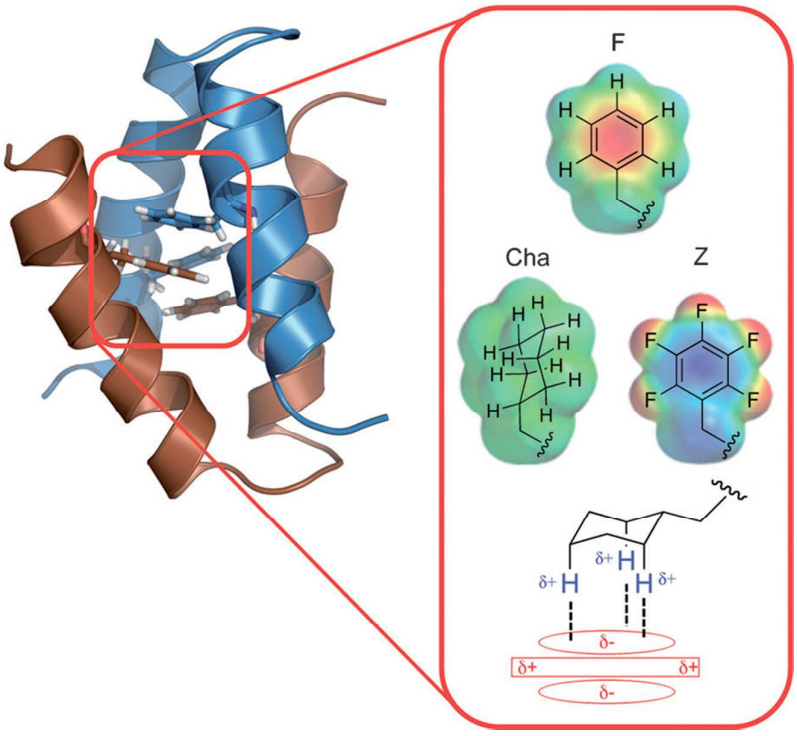


Fig. 6 Cartoon representation of the dimeric structure of α_2D with the aromatic residues

shown as sticks (left). The inset illustrates a Phe/Cha stacking pair in the core of α_2D, with the electron-rich π cloud of Phe interacting with Cha. In the bottom are illustrated CH/π hydrogen bonds between three hydrogens in Cha and benzene ring of Phe. Fig. 1 of Pace *et al., Chem.- Eur. J.,* 2012, **18**, 5832, with permission.

Finally, the chemical shifts of the C^{α}-hydrogen atoms of the first Pro in the *cis* and *trans* isomers of tripeptides of the type Pro-Pro-Xaa were examined, where Xaa stands for the aromatic amino acids His, Phe, Trp and Tyr (Fig. 7).[43] Upfield shifts clearly identify the occurrence of a CH/π hydrogen bond and from free energies associated with *trans* to *cis* transformation of the Pro-Pro moiety, it could be shown that this interaction contributes about 1.0 kcal mol^{-1} to the overall stability.

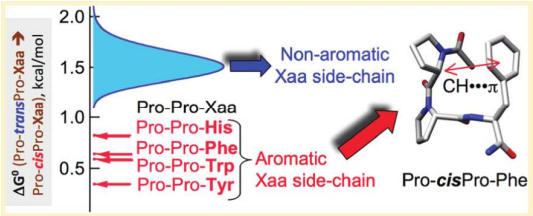


Fig. 7 Stabilization of the Pro-*cis*Pro peptide bond by a following aromatic residue. The *cis*-conformation is stabilized by about 0.35-0.82 kcal mol^{-1} when Xaa in the peptide Ac-Pro-Pro-Xaa is an aromatic residue compared to when Xaa is non-aromatic. This stabilization is attributed to the presence of an attractive CH/π hydrogen bond, between the C^{α}-H of the first Pro and the aromatic ring of Xaa. Additional indirect evidence for the presence of this CH/π hydrogen bond comes from a ring-induced upfield chemical shift of the Pro-C^{α}-H. From Ganguly *et al., J. Am. Chem. Soc.,* 2012, **134**, 4661, with permission.

2. CH/π hydrogen bonds in proteins

2.1 Overview

In 1978, Nishio investigated for the first time the occurrence and importance of the CH/π hydrogen bond in biological macromolecules.[44] The literature survey revealed indeed the

frequent occurrence of the CH/π hydrogen bond in protein structures such as lysozyme, erabutoxin, Bence-Jones proteins and immunoglobulin fragment Fab McPC603. Later Umezawa and Nishio examined, by program CHPI (see Section 1.3.2.1) CH/π hydrogen bonds in haemoglobin α-subunit, carp parvalbumin, FKBP/FK506 complex, GBP/glucose complex, lysozyme/$(GlcNAc)_3$ complex, xylose isomerase/D-sorbitol complex and human plasminogen kringle/ε-aminocaproic acid complex, *etc.*; there they found many CH/π short contacts in the protein/specific ligand complexes. With respect to the CH-donor groups, every CH is potentially relevant. As to the π-acceptor groups, typical π-systems are the side-chain aromatic groups of the amino acids Trp, Tyr, Phe and His. The carboxyl/carboxamide side chains (in Asp, Glu, Asn and Gln) may also participate. In principle the main chain peptide groups constitute π-systems with some degree of delocalization, which may also act as π-acceptors in some cases.

In 1985, Burley and Petsko,[45] and Singh and Thornton[46] carried out analyses of the interaction geometry of aromatic pairs (based on 34 high resolution protein structures) and Phe-Phe pairs (in 29 proteins), respectively, and found a preference for the perpendicular (T-shaped) arrangement. As to the origin of this effect, they argued for the importance of a polar-π interaction. As already demonstrated in a previous section, this so-called T-shaped interaction can nicely be explained using the concept of the aromatic CH/π hydrogen bond. Although an aromatic face-to-face π/π arrangement appeared to not be very common in the study by Burley and Petsko, a subsequent analysis involving a larger sample of 505 non-homologous proteins from the PDB,[47] suggests that the preferred orientation of aromatic pairs indeed corresponds to a displaced parallel π stacking.

In the early 1990s, several papers appeared, commenting on the possible importance of the CH/π hydrogen bond in protein chemistry. In 1995, Chakrabarti and Samanta showed that CH/π hydrogen bonds play also an important role in the packing of adenine ring of cofactors, such as NAD^+, NADP, FAD and ADP, *etc*. in enzymes (dehydrogenases, reductases and kinases, *etc*.).[48] The incorporation of nitrogen atoms makes the heterocyclic ring in the cofactor electron-rich, which can results in an enthalpically favourable interaction with CH donor groups.[49] The adenine ring is usually sandwiched between two amino acid residues, there being a preponderance of branched aliphatic side-chains, such that methyl or methylene carbon atoms sit on top of the ring nitrogen

atoms at distances of approximately 3.7 Å. In many instances, an amino acid residue can employ its main-chain atom to form conventional hydrogen bond along the ring periphery, while the side-chain atoms are positioned to interact with the adenine face, thereby enabling the protein engage an adenine moiety by employing only a very limited number of residues. The usefulness of the concept of stabilizing CH/π hydrogen bonds is due to the fact that they normally occur in protein interiors or binding sites involving rather conformationally rigid groups (main-chain CH, branched-aliphatic and aromatic side chains) so that the loss of conformational entropy from such interaction is less. Since they are numerous they can contribute significantly to the overall stability. Besides, the two groups involved in CH/π hydrogen bonds are nonpolar and are easily desolvated. A normal hydrogen bond may provide a stronger interaction in the complex; however, the participating groups are much more difficult to desolvate prior to hydrogen bond formation. Following this work there have been many other studies involving protein structures and binding, which will be dealt with in a separate section.

2.2 Analyses of crystallographic data in the PDB

2.2.1 Large-scale database studies. The initial evidence for the importance of CH/π hydrogen bonds in proteins arose from the examination of individual proteins or protein-ligand complexes, such as guanine-nucleotide binding proteins, SH2 domains, human growth hormone or MHC proteins.[50] Subsequently, it was recognized in two independent database studies on the occurrence of *cis* peptide bonds in protein structures[51,52] that the special geometry imposed by a *cis* peptide bond often leads to close contacts between the side chains of the two residues flanking the *cis* peptide bond. In both studies, it was proposed that this special arrangement is frequently stabilized by CH/π hydrogen bonds. In support of that is the higher than expected occurrence of aromatic residues as one of the *cis* peptide flanking residues. A striking example is the Trp-Pro *cis* peptide bond a copper amine oxidase (PDB entry 1OAC),[53] where the distance between the Pro-C^{α}-H to the indole ring of Trp is only d = 3.54 Å (Fig. 12A), just slightly larger than the limit that is frequently suggested for a conventional hydrogen bond and significantly shorter than what one would expect if it was a simple van der Waals interaction.

In one of the first large-scale database studies Samanta *et al.*[54] examined 719 Trp

residues occurring in 180 protein structures. They realized that these Trp side chains are involved in all kinds of interactions including CH/π hydrogen bonds. Almost at the same time Brandl *et al.* published a paper in which they described the examination of a non-redundant set of 1154 proteins for the occurrence of CH/π hydrogen bonds.[39] More than 30000 individual CH/π hydrogen bonds were identified in this study according to the Brandl Weiss system. These were then grouped into different donor and acceptor classes. Three C-H donor classes (C^{α}-H, $C_{aliphatic}$-H and $C_{aromatic}$-H) were defined and four π-acceptor classes (aromatic (His, Phe, Trp, Tyr), guanidinium (Arg), carboxylate (Asp, Glu), and carboxamide (Asn, Gln)). The only π-systems, which had been left out in this analysis, are the protein main chain peptide bond π-systems and the C-terminal carboxylate. By careful analysis, it could be observed that the combinations $C_{aromatic}$-H/$\pi_{aromatic}$ and $C_{aliphatic}$-H/$\pi_{aromatic}$ are much frequently observed than the other combinations. As a matter of fact, the majority of aromatic π-systems, which occur in the core of a protein, are involved in a CH/π hydrogen bond as a π-acceptor. Many of such interactions, in particular the ones involving $C_{aromatic}$-H donors had previously been named aromatic edge-to-face interaction but this new study suggested that these interactions are really mostly CH/π hydrogen bonds (see also the discussion in Section 1.3.2.3). Brandl *et al.* also reported that there was a dependence of the mean geometric parameters of CH/π hydrogen bonds dependent on the resolution of the structure. In essence this means that lower resolution structures are so much influenced by the lower observables to parameter ratio that some atom-atom distances are significantly different from the ones observed in higher resolution structures. Finally, it was observed that a large number of CH/π hydrogen bonds occur between residues, which are relatively close in sequence space. This means that such interactions are often involved in stabilizing secondary structures. Aspects of this are also supported by later studies, *e.g.* in Thomas *et al*.[55] and Bhattacharyya *et al.*[56] In summary, these large-scale database studies revealed for the first time that the number of CH/π hydrogen bonds in proteins is probably much higher and that such interactions may be much more widespread in macromolecular structures than was previously thought.

The concept of local stabilization of dipeptides involving one Pro residue by NH/π and CH/π hydrogen bonds was picked up by Tóth *et al*.[57] The authors in this study arrived at essentially the same conclusions as Jabs *et al*. and Pal and Chakrabarti earlier. A more extensive discussion on stabilizing CH/π hydrogen bonds around *cis* peptide bonds will

follow in the chapter on local structures.

2.2.2 Analysis of individual proteins or protein classes using CHPI

2.2.2.1 The G-protein/GTP-analogue complex p21/GppNp. With the help of the program CHPI, Umezawa and Nishio analysed the 3D structure of a G-protein/GTP-analogue complex (Fig. 8). This highlights a striking feature observed in the *Ras* p21/GppNp complex structure. The G-protein/GTP-analogue complex revealed that the ligand GppNp, sandwiched by side-chains of Lys117 and Lys147 of the protein, is interacting with the CH_2 groups in these residues. The charged N^+H_3–groups of the Lys residues are not involved. Lys117 is located in the unfavourable region of the Ramachandran diagram, indicating that this unusual main chain conformation may be stabilized by this arrangement. For all amino acid residues interacting with the cofactor, rather low atomic displacement parameters were observed after crystallographic refinement. This indicates that these residues certainly contribute significantly to the G-specificity of p21. Lys117 is well conserved. Lys147 is well conserved also. The inspection of the crystal structure showed that the CH/π hydrogen bond is possible for diverse residues; in other words, substitution may occur but the necessary interactions are conserved.

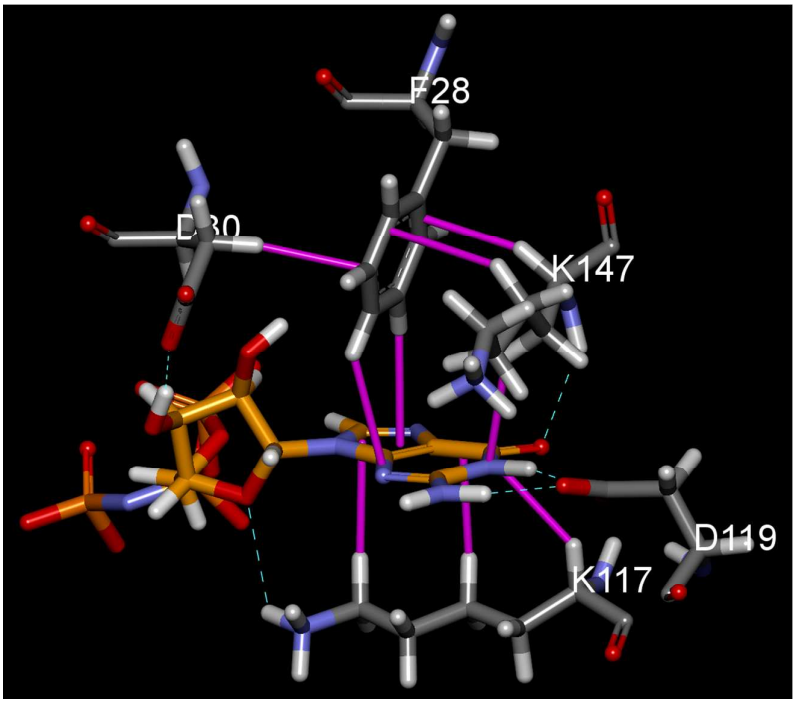


Fig. 8 G-protein (*Ras* p21) complexed with a GTP-analogue (PDB entry 5P21). Orange sticks: GTP-analogue, GppNp. Gray sticks: protein. Magenta sticks and blue dotted lines indicate CH/π hydrogen bonds and conventional hydrogen bonds, respectively, based on Fig. 3 of Umezawa and Nishio, *Bioorg. Med. Chem.*, 1998, **6**, 493, with permission.

2.2.2.2 Human growth hormone. A striking feature involving many CH/π hydrogen bonds is shown in Fig. 9. A whole ladder of CH/π hydrogen bonds occurs in the structure of human growth hormone. The CH_2-groups of Lys and Arg side-chains are sandwiched by aromatic residues and *vice versa*. A number of short contacts can be observed between Lys or Arg and the aromatic ring of Tyr, Phe or Trp. Abundance of this type interaction in protein 3D structures has been reported previously and termed as the "Lys-Arg CH/π hydrogen bond".[58,59]

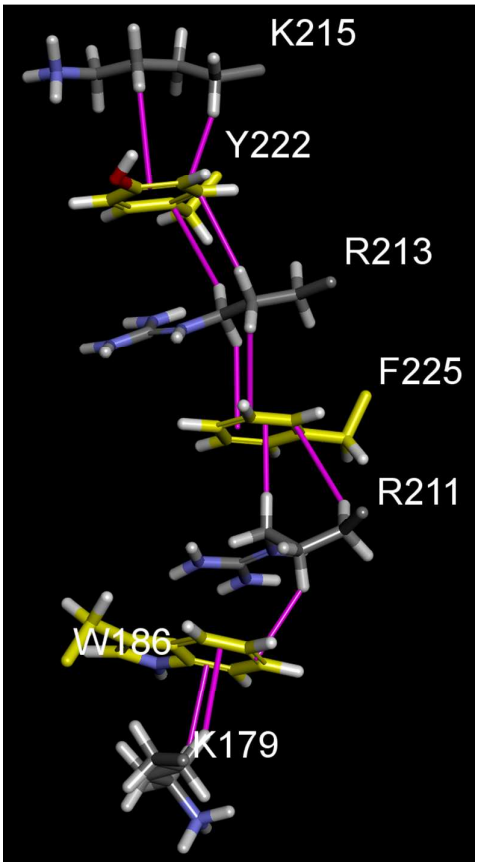


Fig. 9 A series of Lys-Arg CH/π hydrogen bonds observed in human growth hormone (PDB entry 3HHR). Lys215/Tyr222/Arg213/Phe225/Arg211/Trp186/Lys179. Aromatic residues are in yellow sticks. Fig. 10 of Umezawa and Nishio, *Bioorg. Med. Chem.,* 1998, **6**, 493, with permission.

2.2.2.3 CH/π hydrogen bonds vs. so-called cation/π interaction. In their publication dealing with the human growth hormone, Ma and Dougherty had argued that the so-called cation/π interaction is responsible for sustaining the stable 3D structure of human growth hormone.[60] A careful examination of the structure revealed, however, that many CH/π hydrogen bonds are formed between the residues they had implicated in cation/π interactions. Lys and Arg contain four and three CH_2-groups, respectively, between their Cα and their terminal ammonium or guanidinium group, which means that they harbor many potential CH donor groups. Fig. 9 shows indeed that multiple pairs of CH atoms in Lys and Arg are involved in CH/π hydrogen bonds. Therefore, it seems

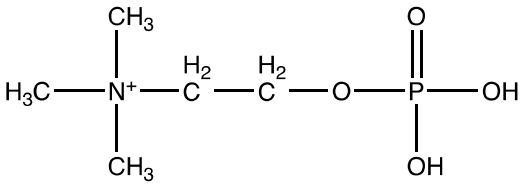
likely that the interactions cooperatively occur by use of the CH_2 groups, and not with the ammonium cation. Cohen and coworkers studied the kinetics of binding of acetylcholine (ACh) analogues such as 3,3-dimethylbutylacetate and 3,3-dimethylbutanol.[61] There they found that the above neutral compounds bind as effectively as ACh to the same subsite of acetylcholine esterase (AChE). Comparisons of quaternary compounds (Me_3N^+-R and Me_3C-R) as ligands with their lower analogues (Me_2N^+-R and Me_2C-R) showed the former to be much more effective than the latter with regard to its binding capabilities to AChE.

The binding constant *K* in a host calixresorcinalene/guest (Me_nN^+-R *vs*. Me_nC-R) complex system has been found to become much larger on increasing the number of methyl groups of the guest.[62] Recently the interaction between Lys and Trp was examined using a model β-hairpin dodecapeptides incorporating Trp at position 2.[63] The ratio of the folded conformer was 77% for a peptide bearing Lys at position 9, while norleucine and acylated Lys analogues gave almost identical results in conformational equilibria.[64] This demonstrates that the CH_2 groups are interacting with the aromatic side-chain of Trp *via* CH/π hydrogen bonds; it is clear that the positive charge does not play any significant role.[65]

2.2.2.4 Photoactive yellow protein. Concerning the interaction of an aromatic system with the Lys side-chain, the result for photoactive yellow protein (PYP) is worthy to note.[66] The authors reported that a single CH/π hydrogen bond occurring between the side-chain methylene groups of Lys123 and the aromatic ring of Phe6 governs the stability and the biological activity of PYP. They found, by site-specific mutagenesis, that Lys123Leu and Lys123Glu retain the 3D structure and biological activity of wild-type PYP. This indicates that the isobutyl group in Leu and two methylene groups in the Glu side-chain are almost equally effective as those in the aliphatic part of the Lys side-chain. On the contrary, the mutations Phe6Leu and Phe6Asp completely abolish the stability and the activity of the protein. This demonstrates nicely, that neither the so-called hydrophobic effect nor the classical hydrogen bond plays a role in stabilizing PYP at this site.

2.2.2.5. Immunoglobulin fragment in complex with phosphorylcholine. A CHPI analysis was also carried out on the crystal structure of an immunoglobulin fragment Fab

McPC603 in complex with its specific substrate, phosphorylcholine.[67] It is clear from Fig. 10 that methyl and methylene protons are interacting with the aromatic ring of Trp107(H) and Tyr100(L) in Fab'.



Phosphorylcholine

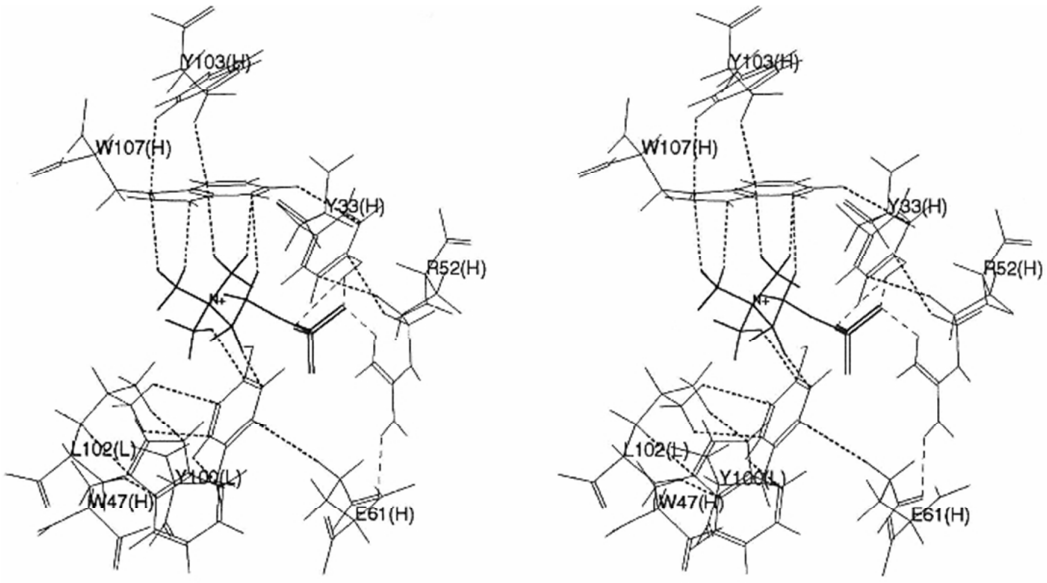


Fig. 10 Crystal structure of McPC603 in complex with phosphorylcholine (stereo view). Dashed lines indicate hydrogen bonds. A number of CH/π short distances (dotted lines) are disclosed between CHs of the ligand phosphorylcholine and aromatic side-chains of Trp107(H), Tyr103(H), Trp47(H), and Tyr100(L) in McPC603. H: heavy chain. L: light chain. Fig. 11.20 of Ref. 13, with permission.

Umezawa and Nishio analysed the crystal structure of class I major histocompatibility complex antigens (MHC) bound to their specific ligand peptides. With dissociation constants of around 10^{11} M^{-1} the stability of these complexes is very high. A number of

short CH/π distances have in fact been shown to occur right at the boundary of the heavy chain and β2 microglobulin and between CHs and aromatic side-chain groups in the MHC/peptide complexes. The CH/π hydrogen bond has been suggested to contribute to the specificity in the complex formation of class I MHC. [68]

Chakrabarti and Bhattacharyya wrote comprehensible reviews on the geometry of nonbonded interactions involving planar groups in proteins. [69] Inoue *et al.* wrote comprehensible reviews on weak interactions in protein structures.[70] Dasgupta *et al.*[71] also performed database analyses but these topics will be thoroughly dealt with in the next section. For other reviews, see Meyer *et al*.,[72] Chakkaravarthi *et al.*,[73] Terrón *et al.*,[74] Cohen *et al.*,[75] Suresh *et al.*,[76] ad Bissantz *et al.*[77]

2.3 Stabilization of protein local structures by CH/π hydrogen bonds

2.3.1 Turns. The amino acid Pro has a high propensity to have aromatic side chains in its immediate environment,[78] and since it is also a preferred residue in the turn regions in protein structures,[79] such local structures often have Pro and aromatic residues in contact. Though a constrained backbone torsion angle enables Pro to occur in turns,[80] having an adjacent aromatic residue in sequence enhances the stability of turn conformation through stacking interaction, many of which can be classified as a CH/π hydrogen bond.[81] There are many types of turns.[82,83] In a β-turn (defined by four residues at positions i to i + 4) of type I, a Pro at the i + 1 position is usually preceded by a His – the latter can use its ring N atom to accept a proton from the backbone NH group at position i + 2, simultaneously engaging a Pro CH group through a CH/π hydrogen bond (Fig. 11). Pro, occupying the usual i + 1 position, can also be favourably disposed to interact with an aromatic residue following it. Another category of β-turn, which can be mediated by aromatic/Pro interactions (at positions i + 1 and i + 2, respectively, linked by a *cis* peptide bond – discussed in the next section) is of type VI.

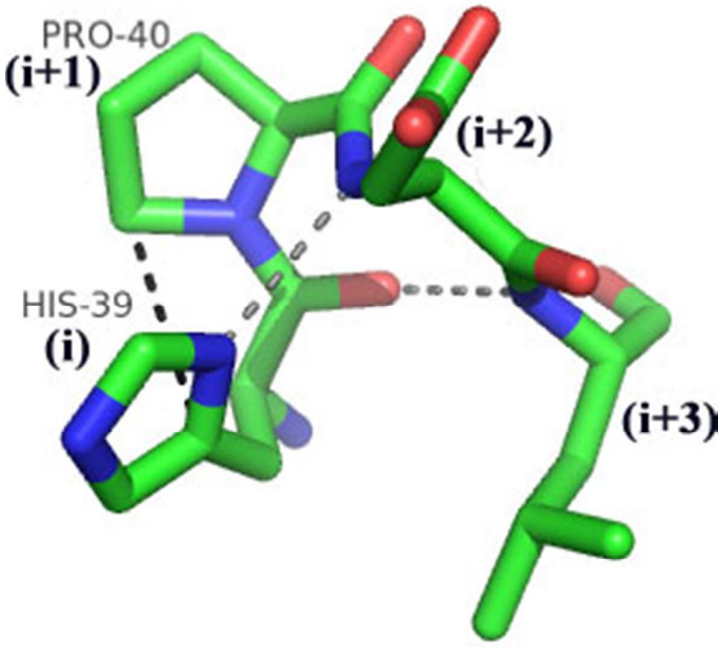


Fig. 11 Pro-aromatic interaction (in black dashed line) in β-turn; the normal hydrogen bonds are shown in grey. The main-chain atoms at positions i, (i+1), (i+2) and (i+3) of β-turn are shown along with the side-chain atoms of the interacting Pro and aromatic residues. Made following Fig. 6 of Bhattacharyya and Chakrabarti, *J. Mol. Biol.*, 2003, **331**, 925, using the PDB file 1GCI and the software PYMOL (http://www.pymol.org).

2.3.2 Main-chain conformation: cis peptide bond. Cis peptide bonds in the peptide unit X(i)-X(i+1) (X stands for any amino acid) are rare in proteins mainly due to the unfavourable steric interactions between C^{α}(i) and C^{α}(i+1). However, with a Pro at the i + 1 position there is also an unfavourable C^{α}(i) and C^{δ}(i+1) steric interaction even in the *trans* conformer, decreasing the energy difference between the *cis*-and *trans*-conformation. Thus, there is a clear over-representation of Pro at i+1 when the peptide unit occurs in the *cis* conformation.[84,85] An aromatic residue preceding the Pro has a higher chance of making the X-Pro bond occur in *cis*-conformation as has also been seen in NMR studies of the *cis-trans* equilibrium in short designed peptides.[86,87,88] It has been proposed that the CH/π hydrogen bond is responsible for the stability of such motifs

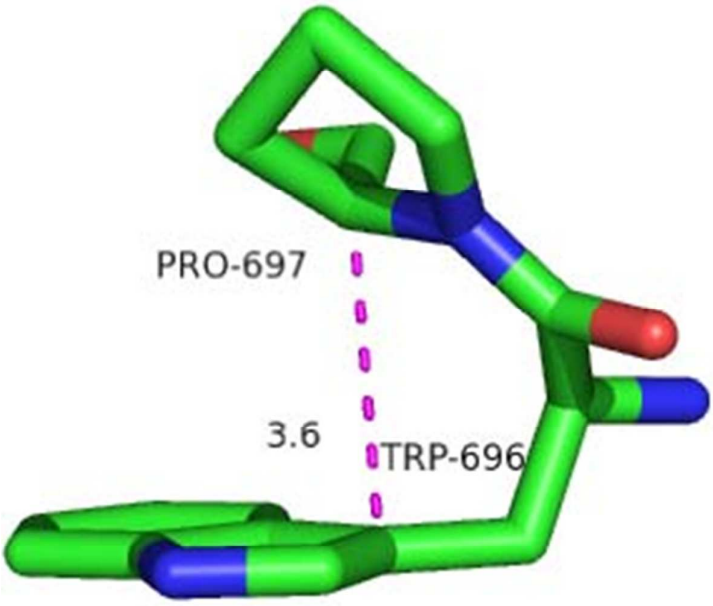
– the C^{α} atom of Pro is usually at a distance of 3.6(±1) Å from the C^{γ} atom and interacting with the face of the aromatic residue (Fig. 12A). Using model peptides, Ac-Thr-X-Pro-NH_2, with X varying from electron-rich to electron-poor aromatic amino acids, it was observed that the population of *cis* amide bond is tunable by aromatic electronics.[89] A CH/π hydrogen bond can form more easily when the CH donor group is made more acidic by an adjacent electron-withdrawing nitrogen atom, as in the C^{α} and C^{δ} positions of the Pro ring, or the peptide C^{α} atom. The CH/π hydrogen bond is ubiquitous not only between residues involved in *cis* peptide bond, but sometimes also across it, and both these atom types could be used. For example, quite frequently an aromatic residue (Aro) in the sequence X-Pro-Aro can interact with the C^{α}-H of a non-Pro residue at X (Fig. 12B) or the C^{δ}-H of Pro if X happens to be an aromatic residue (Fig. 12C). In the latter structure two CH groups of Pro interact with the π-face of aromatic residues on either side. The stability conferred by two CH/π hydrogen bonds to the type VIa turn displayed in the structure is demonstrated by the occurrence of a high population of the *cis* isomeric form in solution of the polypeptide, Ser-Tyr-Pro-Tyr-Asp-Val.[90]

It has also been observed that a *cis* peptide bond between two Pro residues is usually followed by an aromatic residue in such a way that the C^{α} of the first proline interacts with the π-face of the aromatic residue at a distance of about 3.8 Å (Fig. 12D). That the type of residue following it can modulate the *cis-trans* isomerization of Pro-Pro peptide has been demonstrated with two model peptides, Ac-Pro-Pro-Phe-NH_2 and Ac-Pro-Pro-Ala-NH_2, the former having 47% of the *cis* Pro-Pro conformer in water, and the latter only 17%. Van't Hoff analyses of temperature-dependent NMR spectra yielded entropic and enthalpic contributions for the *trans* to *cis* transition – both the peptides yielded unfavourable entropies while the associated enthalpy was favourable for the former and unfavourable for the latter.[71] Using all possible amino acids at X in the three-residue peptide sequence Pro-Pro-X it was shown that the CH/π hydrogen bond (between the C^{α}-H of the first Pro and the aromatic side chain) stabilizes the *cis* conformer by ca. 0.9(±0.2) kcal mol^{-1}. The interaction causes an upfield chemical shift (by 1-2 ppm) of the proton (due to the ring current effect of the π-system) and a restriction of rotation of the aromatic side chain as compared to any non-aromatic residue at X.[43]

It is not only in Pro-containing *cis* peptides, however, that the CH/π hydrogen bonds are noticeable. Though rare, *cis* peptide bonds have also been identified in proteins, and these interactions have been reported in Xnp-Aro and Aro-Xnp cases (Xnp being any

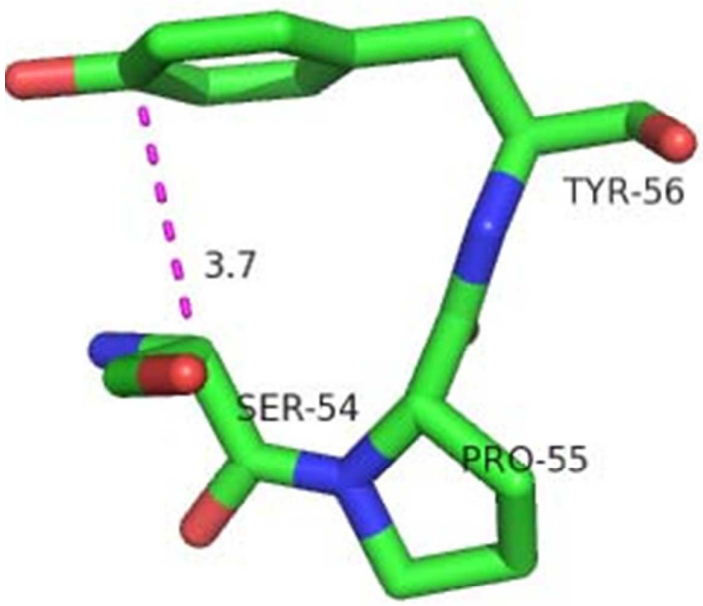
non-proline residue), where a C^{β}-H of Xnp usually points to the centre of the aromatic residue at distances ranging from 3.4 to 4.4 Å (Fig. 12E).[51] It was earlier observed that when there is an interaction involving the face of Trp residues, of all the ring atoms $N^{\varepsilon 1}$ has the maximum number of CH groups interacting with it.[54] Trp seems to be a favoured residue in Aro-Xnp examples, and even here C^{β} of Xnp has the shortest contact with $N^{\varepsilon 1}$ of Trp, indicating a relatively stronger interaction resulting from a CH (or NH) group being directed towards the π electrons of an N atom.[52]

(A)

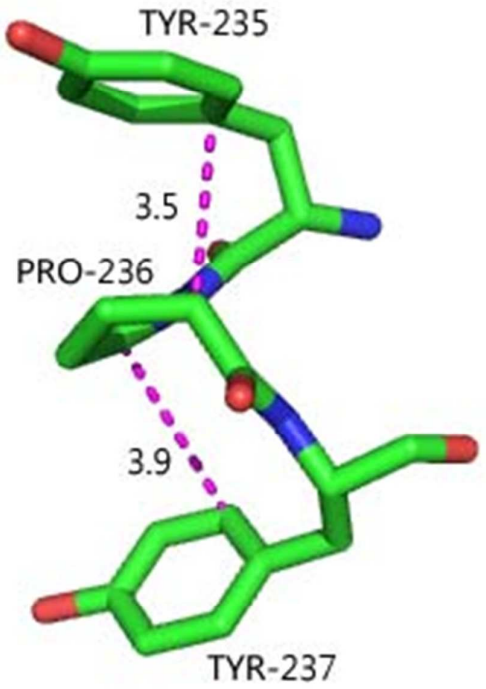


(B)

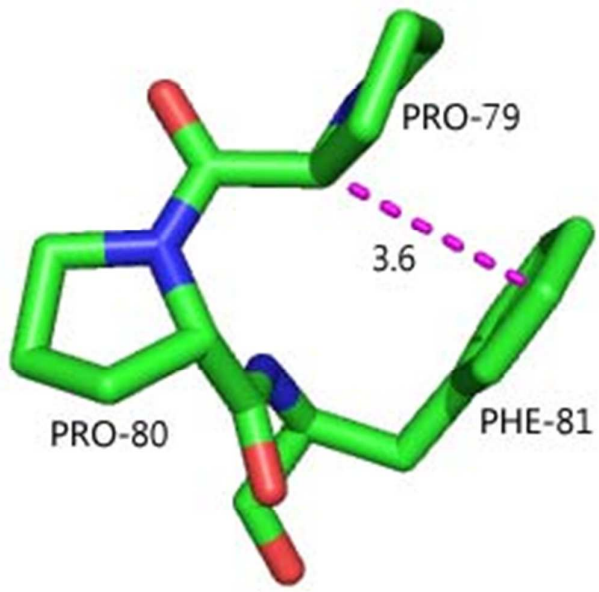
30



(C)



(D)



(E)

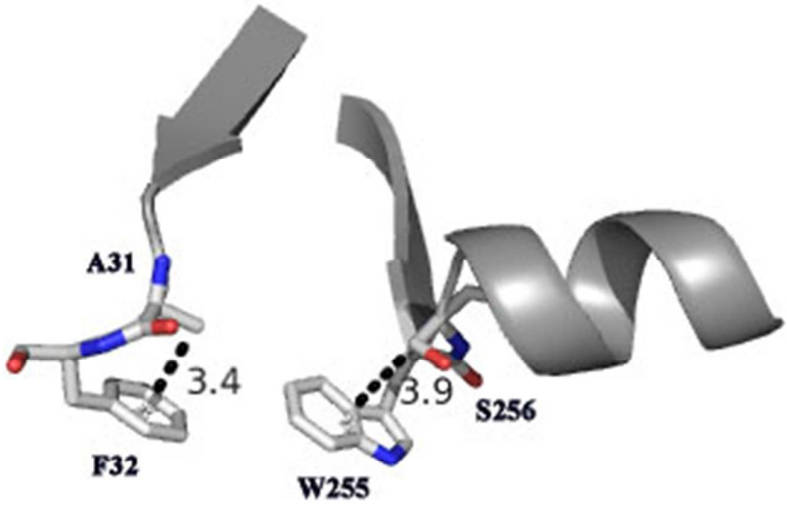


Fig. 12 CH/π hydrogen bonds (with distances, Å) in (A-D) X-Pro and (E) X-nonPro *cis* peptide bonds. The X-Pro *cis* peptides are: (A) Trp696-Pro697 (PDB, 1OAC), (B) Ser54-Pro55, with C^{α} atom of the former having CH/π hydrogen bond with Tyr56

(1RGA), (C) Tyr235-Pro236, with C^{α} and C^{δ} of Pro having CH/π bonds with Tyr rings on both sides (1ADE), and (D) Pro79-Pro80, with C^{α} of Pro79 having CH/π bond with Phe81 (1AOC). In (E) (PDB, 2HVM) there are two spatially close *cis* peptides, both at the end of β-strands (residues 27-30, and 249-255), and one of them is followed by an α-helix (residues 257-263); in both the cases the C^{β} atom is the donor. (A)-(D) are based on Fig. 12 of Pal and Chakrabarti, *J. Mol. Biol.*, 1999, **294**, 271; (E) is based on Table 2 of Jabs *et al*., *J. Mol. Biol.*, 1999, **286**, 291.

Another striking example for CH/π hydrogen bonds stabilizing a non-Pro *cis*-peptide bond is the two *cis*-peptide bonds found in human blood coagulation factor XIII (Fig. 13). Here both *cis*-peptide bonds Arg-Tyr and Gln-Phe are obviously stabilized by CH/π hydrogen bonds. It is not always necessary that an aromatic residue provides the π system – in the proteins of the lectin family there is an Ala-Asp *cis* peptide bond involving an interaction between the C^{β}-H and the π-electrons of the carboxylate group.[51] Interestingly, most of the non-proline *cis* peptide bonds occur in or close to the active site of proteins, in particular carbohydrate-binding or processing proteins. This alone may already be suggesting a functional role of some sort of these peptide bonds.

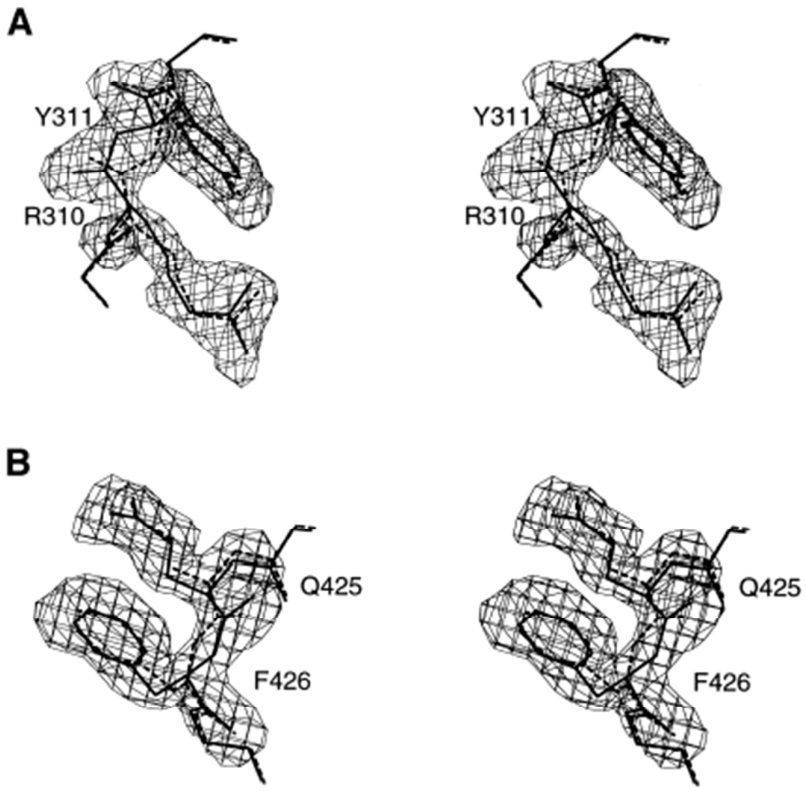


Fig. 13 Stereo diagrams of the peptide bonds (A) Arg-310-Tyr311 and (B) Gln425-Phe426 in blood coagulation factor XIII (PDB entry 1F13).[91] The two superimposed structures show the peptide bond modeled in *cis*-conformation (solid lines) and in *trans*-conformation (dotted lines). It is clearly evident from the figure that the fit of the atoms to the electron density is much better if the peptide bond is modeled in *cis*. The figure also nicely highlights the proximity of the aliphatic portions of the first amino acids Arg and Gln to the aromatic rings of the second amino acids Tyr and Phe, respectively. Obviously, the *cis*-conformation of the two peptide bonds in blood coagulation factor XIII is stabilized by CH/π hydrogen bonds.

2.3.3. Interaction in and around Helices. α-Helices are quite enriched in CH/π hydrogen bonds – in α-helical proteins 20 CH/π interactions can be identified per 100 residues, while there are only 9 in β-sheet proteins.[39] If the geometry of interaction between aromatic residues located in and around α-helices is analyzed, it is evident that the interaction depends on the sequence difference (Δ) between the two residues.[56] Many such pairs, especially when Δ is 1, 3 or 4, exhibit CH/π or NH/π (for the imidazole ring of

His it is not always possible to distinguish between the two possible proton donors, and both may exist simultaneously), causing a restriction on the distribution of the side chain torsion angle (χ1). The order of occurrence of aromatic pairs in α helix determines the stability conferred to the α-helix. For example, using Ala-based peptide systems it has been shown that a Phe (or Trp), His pair placed at i, i+4 positions, respectively, gives rise to the highest helical content (*i.e.*, most stabilizing effect) when the His is protonated.[92,93] A non-His aromatic residue preceding a His (by 4 residues) is found significantly more in protein structures than the two residues placed in the reverse order with the same spacing. The explanation is that in the preferred arrangement a C-H or N-H group of His can be directed towards the π-electron cloud of the other aromatic ring. In case the His ring is protonated, the X-H protons in the ring carry more partial positive charge, resulting in the occurrence (or the strengthening) of the XH/π interaction (Fig. 14). The interaction of His with another aromatic ring can be observed not only in α-helices, but also in the overall tertiary structure. For example, in barnase the edge of His18 interacts with the face of Trp94. The protonation of His increases the stability of the protein by about 1 kcal mol^{-1},[94] This may be attributed to the participating XH/π hydrogen bonds.

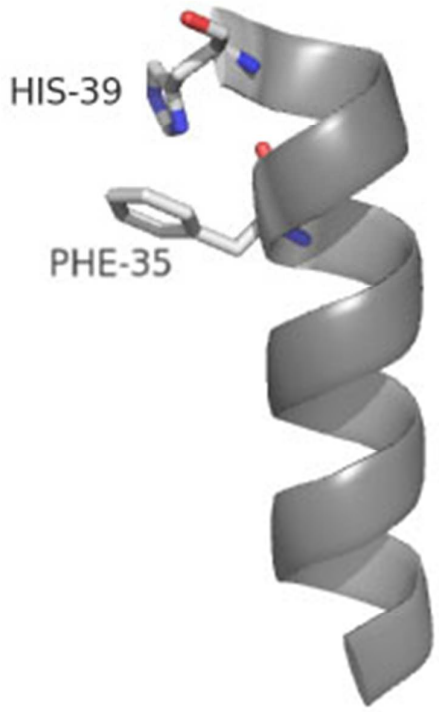
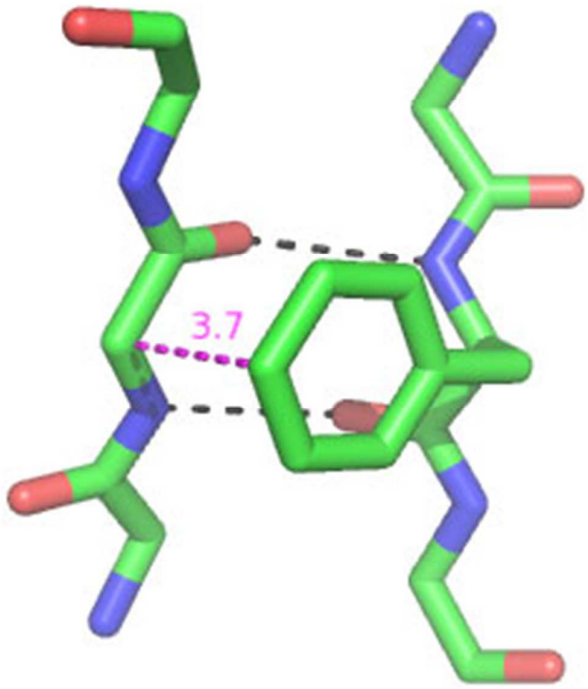


Fig. 14 Interaction between Phe and His with sequence difference of 4, occurring at the C-terminal end of a helix (residues 25-39) in the PDB entry 1MTY[95] (chain G). Made following Fig. 6 of Bhattacharyya *et al.*, *Protein Eng.*, 2002, **15**, 91.

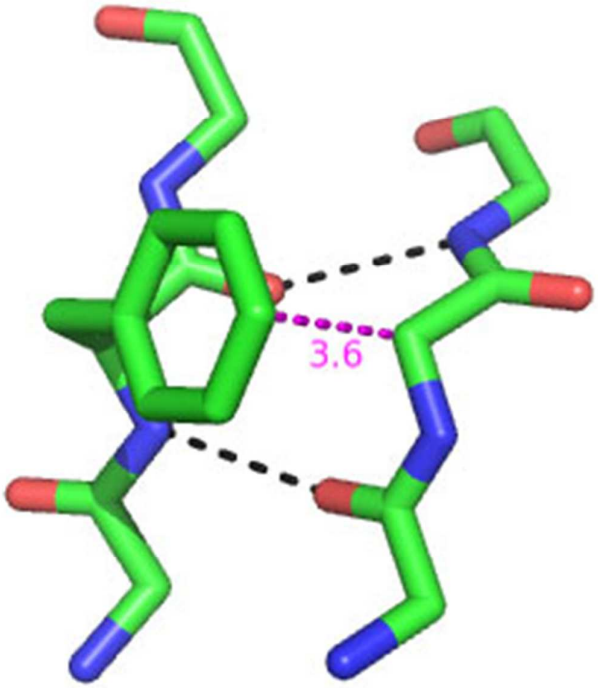
(A)

37

(B)



(C)



(D)

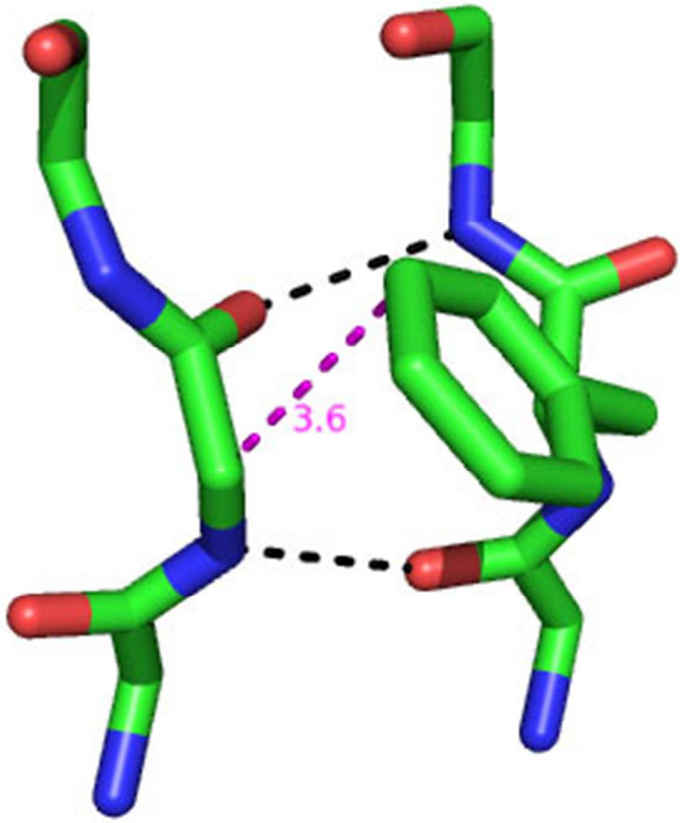


Fig. 15 CH/π hydrogen bond in β-sheets. (A) Interaction between Pro and Tyr, residues 4-10 and 17-23 constituting the two β-strands (PDB entry 1BXO,[96] chain A). Interaction between Gly and Phe are shown in (B)-(D). (B) Gly94 and Phe82 in antiparallel β-sheet in 1PLC.[97] (C) Gly56 (at non-hydrogen-bonding position) and Phe 91 in parallel β-sheet in 2FX2.[98] (D) Gly38 (at hydrogen-bonding position) and Phe134 in parallel β-sheet in 1DEA.[99] The closest contacts in residue pairs are shown, as well as hydrogen bonds in (B)-(D).

2.3.4 Interaction across strands in β-sheet. Interaction between a proline and an aromatic residue can occur between two strands of an antiparallel β-sheet, thereby accommodating a Pro in the sheet.[81] With no main-chain NH that can form hydrogen bond, a Pro can be located at the edge strand (quite frequently at the C-terminal position) with its CO group pointing outside. An aromatic residue (a preference exists for Trp and Tyr) placed at the non-hydrogen-bonded site can stack against the pyrrolidine ring of Pro (Fig. 15A) – the

stretches of the peptides centered at the two residues in different structures can be superimposed nicely. Stacking against Pro is the only way a large aromatic residue like Trp can have interaction within an antiparallel β-sheet.

Like Pro-aromatic pairs, Gly-aromatic pairs are also overrepresented in protein structures, although by themselves Gly and Pro residues are the most destabilizing residues in β-sheets.[100] The cross-strand pairing of Gly with aromatics in antiparallel strands has been shown to increase protein stability synergistically, while in parallel β-sheets the protein is stable only when Gly occurs at the non-hydrogen-bonding site,[101] paired with cross-strand Phe at the hydrogen-bonding site. Though the stability has been ascribed to the positioning of the aromatic ring over the Gly residue, which prevents competing water molecules to form hydrogen bonds with the amide and carbonyl groups of the residue, there is clearly a case for a CH/π hydrogen bond in such pairing. In both the cases the Gly C^{α} has a distance of less than 3.7 Å from a ring C atom of Phe (Figs. 15B and 15C). Interestingly, the reversal of the positions, *i.e.*, Gly at the hydrogen-bonding site and Phe at the non-hydrogen-bonding site, is not commonly observed. However, there is one such case, in which a CH/π hydrogen bond is also formed. However, in this case the regular β-strand structure is disrupted (Fig. 15D).

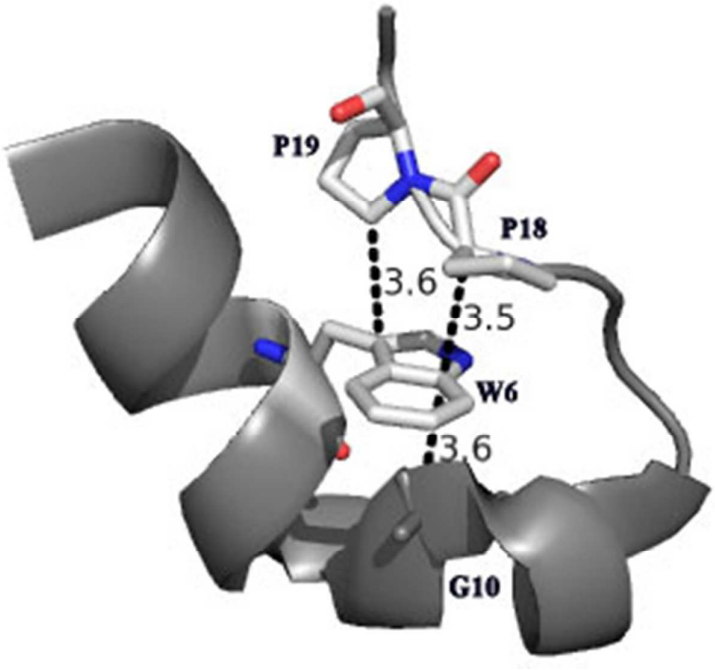
Being rich in Pro and Hyp (hydroxyproline) collagen provides another example where the interaction with aromatic residues can be important. Quantification of the conformational propensities of all 20 amino acids in collagen $(Pro\text{-}Hyp\text{-}Gly)_n$ repeats has indicated the destabilizing nature of aromatic residues.[102]

2.3.5 Interaction in tertiary structures. The hydrophobic effect cannot contribute significantly to stability in small proteins lacking a large core. Here, the CH/π hydrogen bond may be of utmost importance. This is revealed in the structure of Trp-cage,[103] a 20-residue polypeptide that folds rapidly, where a Trp residue in an α-helix is positioned at the centre of the molecule and packs against a polyproline helix such that C^{α} of Pro18 and C^{δ} of Pro19 are at a distance of ~3.6 Å from a ring C atom of Trp6 (Fig. 16A). Interestingly, C^{α} of Gly11 located in a short stretch of a helix is at a similar distance on the other face of the ring, and one of its protons is particularly upfield-shifted, consistent with an unusually tight interaction with Trp6.[104] Lovas and coworkers quantified the strength of the weakly polar interactions found in the miniprotein using all the available 38 NMR structures (in the PDB file 1L2Y) with DFT quantum chemical calculations, and reported

values in the range −2.32 ± 0.17 to −2.93 ± 0.12 kcal mol^{-1} for the CH/π hydrogen bonds.[105] The importance of weakly polar interactions in the structure and stability of miniproteins was also shown in the case of CLN025, a β-hairpin decapeptide, a variant of chignolin, in which the terminal Gly residues are replaced with Tyr residues, which leads to a 29.7 K increase in melting temperature.[106] The molecule was found to retain its β-hairpin conformation in presence of denaturants.[107] Predictions of small-protein structures such as Trp-cage, displaying all the CH/π hydrogen bonds, can now be efficiently carried out using a modal Monte Carlo approach.[108]

Kang and Byun studied the strength of CH/πhydrogen bonds in a series of tripeptides with the sequence Pro-Leu-X (X = Trp, Tyr, Phe, His, Ala), a motif found in the C-terminal subdomain of villin (a tissue-specific actin-binding protein associated with the actin core bundle) headpiece.[109] The folded structure having CH/π hydrogen bonds between Pro-CH and the aromatic ring was found to be more stable than the extended structure, in particular when a Trp is involved (Fig. 16B).

(A)



(B)

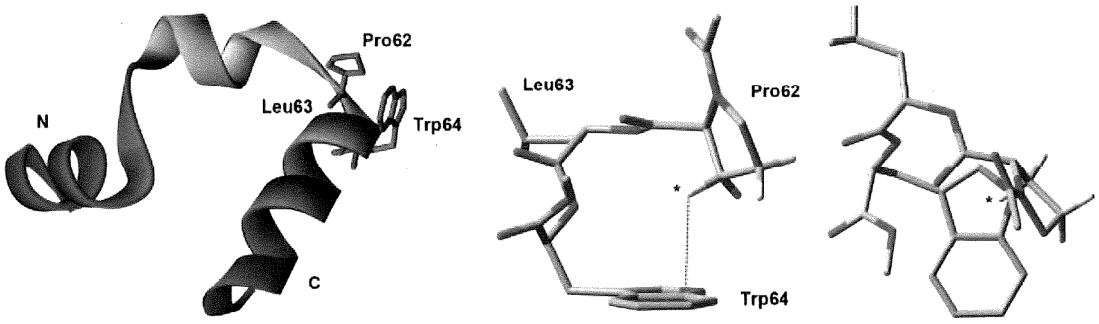


Fig. 16 (A) Structure of the Trp-cage (consisting of α-, 3_{10}- and PPII-helices) showing the CH/π interactions of the central Trp with Pro residues and Gly (PDB entry 1L2Y). (B) Conformation of the C-terminal subdomain of villin with Pro-Leu-Trp sequence (PDB entry 1YU5[110]). Fig. 1 of Kang and Byun, *Biopolymers*, 2012, **97**, 778, with permission.

2.3.6 Interaction in transmembrane domains. CH/π hydrogen bonds may also play a crucial role in the folding of integral membrane proteins and their association into the physiologically relevant oligomers. By sequence analysis, it has been observed that the amino acid sequence Phe-x-x-Gly-x-x-x-Gly is significantly over-represented in transmembrane domains of bitopic membrane proteins.[111] This then led to the suggestion that the first Gly of the membrane-spanning GxxxG motif could be involved in a C^{α}-H/π hydrogen bond with the Phe residue of the partner helix. The low dielectric environment of membranes may enhance the strength of this interaction. The proposed interaction is compatible with the experimental results indicating the first Gly of the motif to be more sensitive to mutation than the second, and the inability of Trp or Tyr to replace Phe without loss of affinity. The large side chain in Trp and the anchoring of Tyr side chain through a hydrogen bond interaction with the carbonyl at the -3 or -4 position on the same helix may hinder close association of the two-helix backbones.

2.3.7 Protein-protein interactions. CH/π hydrogen bonds are copiously observed in protein-protein interactions, as the interface, especially its core, formed during the complexation are enriched in aromatic residues.[112] Many of these constitute hot spots for binding (contributing more than 2 kcal mol^{-1} to the free energy of binding).[113] About 16% of the contacts involving Pro or an aromatic residue on one side and another aromatic group on the other are CH/π hydrogen bonds.[114] The interfaces formed in crystal lattices

of monomeric proteins are usually depleted in aromatic residues, and thus have fewer of such interactions. These interfaces are typically non-specific, weaker and not relevant under physiological conditions.

As in individual protein structures the Pro-aromatic interaction is important in protein-protein complexes as well. Nature seems to have segregated Pro and aromatic residues between the two partner molecules, so as to facilitate multiple intermolecular interactions and prevent competing intramolecular contacts. Many domains critical to the assembly of signaling complexes and pathways usually bind Pro-rich motifs that exist in the form of a polyproline type II (PPII) helix.[115] One such domain of about 60 amino acids and a β-barrel fold is the *Src* homology 3 (SH3) domain, which employs a number of aromatic residues to bind proteins at epitopes having consensus sequence Pro-x-x-Pro (x being any amino acid) (Fig. 17) – the planar aromatic rings are highly complementary to the ridges and grooves present on the PPII helix surface. CH/π hydrogen bonds between Pro and aromatic residues provide sequence-specific recognition that is not particularly high in affinity. Interactions that are specific and low affinity can be quite useful in intracellular signaling environments where rapidly reversible interactions may be required.

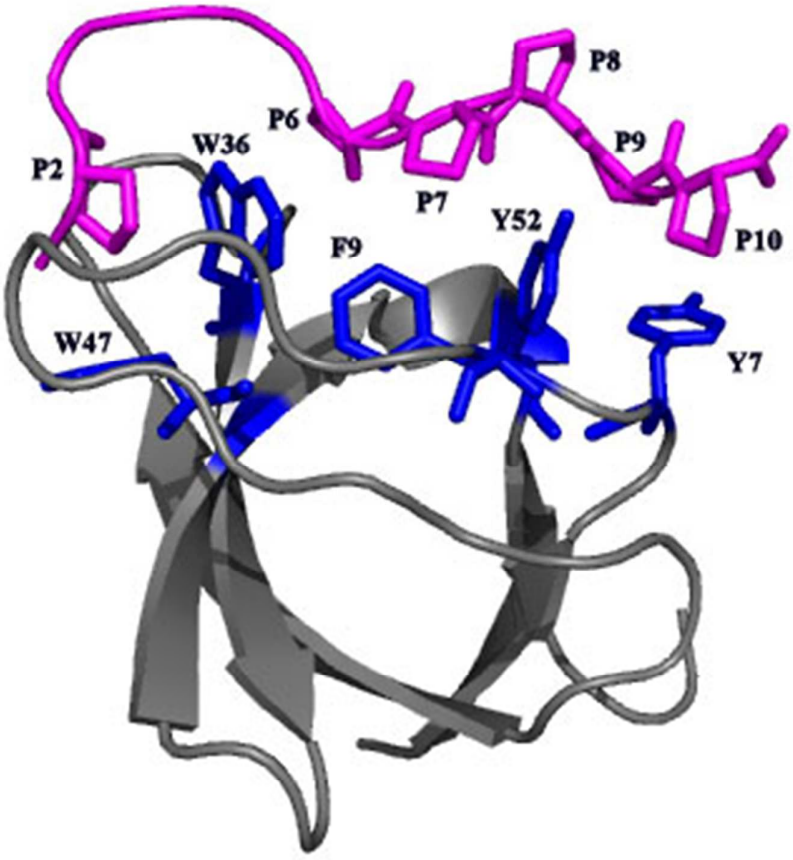


Fig. 17 Juxtaposition of Pro and aromatic residues at protein-protein binding sites. Interdigitation of Pro (magenta) and aromatic (blue) residues at the binding site of ABL-SH3 domain (grey ribbon) for a Pro-rich ligand peptide (magenta) (PDB entry 1BBZ[116]).

Dvir *et al.* studied the synaptic acetylcholinesterase (AChE) tetramer assembled around a polyproline II helix.[117] Functional localization of AChE in vertebrate muscle and brain depends on the interaction of the Trp amphiphilic tetramerization (WAT) sequence at the C-terminus of its major splice variant with a Pro-rich domain of the anchoring proteins (PRAD). The crystal structure of WAT/PRAD complex revealed a supercoil structure in which four parallel WAT chains form a left-handed superhelix around an antiparallel left-handed PRAD helix resembling polyproline II (Fig. 18). The Trp side chains form a spiral staircase, fitting into grooves formed by the left-handed PRAD helix, with the

apparent formation of CH/π hydrogen bonds.

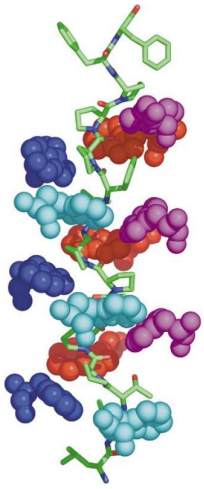


Fig. 18 View of the four parallel WAT helices wrapped around a single antiparallel PRAD helix. Colour coding is from blue at the N- to red at the C-termini for each chain. (C) Side view of the complex, with the superhelical axis running vertically, and PRAD depicted as a stick model. Only the Trps of WAT are shown, in space-filling format, with those of each WAT chain colored individually. The Trp side chains form a spiral staircase around the PRAD, fitting into grooves formed by the extended left-handed PRAD helix. Figs. 5 and 7 of Dvir *et al., EMBO J.*, 2004, **23**, 4394, with permission.

2.4 Interaction of proteins with carbohydrates and glycolipids

At neutral pH, natural monosaccharides with 5 and 6 carbon atoms are mostly observed as furane and/or pyrane-like cyclic structures. The rings are made up from four carbons plus one oxygen and five carbons plus one oxygen atom, respectively. According to the specific stereochemistry of each monosaccharide, several axial CH groups are located on the same face of the sugar ring, thereby forming a planar apolar surface. Such apolar faces form an ideal complementary surface for the planar aromatic rings of Phe, Tyr and Trp residues of proteins.[118]

In 1984, Quiocho and Vyas reported that, in a series of carbohydrate-binding proteins, aromatic residues sandwiched the substrate sugars.[119] The interaction between Trp183 of *E. coli* hexose-binding protein and glucose is a good example of this type of interaction.

When the complex is viewed with glucose in front of Trp183 (Fig. 19), the optimal geometric arrangement of the CH groups just above the carbon atoms of the aromatic ring is particularly obvious.

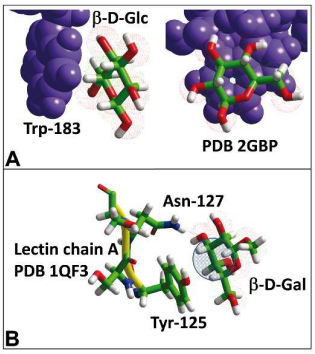


Fig. 19 CH/π hydrogen bonds between water-soluble proteins and natural monosaccharides. (A) β-D-Glc bound to the *E. coli* galactose chemoreceptor protein at a resolution of 1.90 Å (PDB entry 2GBP[120]). The CH/π hydrogen bond involves the CH groups 1, 3, and 5 of the sugar and the aromatic 6-ring of the tryptophan residue Trp183. The total energy of interaction is -5.7 kcal mol^{-1}. (B) Methyl-β-D-Gal bound peanut lectin (chain A) at a resolution of 2.8 Å (PDB entry 1QF3[121]).

The CH/π hydrogen bond involves the apolar cluster of CH groups 1, 3, 4 and 5 and the aromatic ring of Tyr125. The total energy of interaction is -5.0 kcal mol^{-1}. Note that the phenolic OH group of Tyr125 forms a weak non-axial H-bond with the amide group of Asn127. One should note that the oxygen atom of the sugar ring is often pointing away from the aromatic ring, due to the chair conformation of the pyranose ring. This is driven by the electrostatic repulsion between the π cloud of the aromatic ring and the electronegative oxygen atom of the sugar ring. As a consequence, the remaining carbon atoms of the sugar rings get closer to the aromatic structure, which favours the establishment of CH/π hydrogen bonds even more. Because of its two conjugated aromatic rings, Trp appears to have the highest stacking capability among the aromatic residues commonly found in proteins. However, CH/π stacking can involve any aromatic residue, including Tyr, as shown for methyl-β-D-galactose bound the peanut lectin. In this case, the apolar face of methyl-β-D-galactose establishes several CH/π contacts with the

aromatic ring of Tyr125. Interestingly, the sugar-binding site also contains an Asn residue whose amide group forms a hydrogen bond with the OH group of carbon 3 in methyl-β-D-galactose. This indicates that both conventional hydrogen bonds and CH/π hydrogen bonds can cooperate to optimize the binding of carbohydrates to proteins. Umezawa and Nishio explored, by CHPI, CH/π short contacts in various proteins in complex with specific carbohydrate ligands. In every case, the axial CHs in the carbohydrates were found to be involved in the interaction with the aromatic residues such as Trp.

In 1997, Harata and Muraki first pointed out, by crystal structure determinations, the importance of CH/π hydrogen bonds in the binding of lysozyme to its specific substrate tri-N-acetylchitotriose.[122] Mutagenesis and affinity labeling studies followed.[123] Several experimental studies have underscored the importance of carbohydrate-aromatic interactions in the molecular recognition of chitin (a β-1,4-linked polymer of N-acetylglucosamine) by proteins. For instance, Muraki *et al.* have demonstrated that the binding affinity of an antimicrobial peptide (Ac-AMP2) for chitin involves a triad of aromatic residues including Phe18, Tyr20 and Tyr27.[124] Replacing any of these three aromatic residues by alanine resulted in decreased binding affinity. Interestingly, mutations of Phe18 to non-natural residues with larger aromatic rings (*e.g.* β-(1-naphthyl)alanine) increased the affinity for chitin.

In a similar mutagenesis study of the Ac-AMP2 protein, it has been demonstrated that the binding affinity was clearly related to the area of the aromatic group.[125] Indeed, both the association constant and the binding enthalpy were increased with the size of the aromatic side chain. Overall these data strongly support the view that protein binding to carbohydrates relies largely on sugar-aromatic CH/π hydrogen bonds. A triad of aromatic residues is also involved in the binding of *Urtica dioica* agglutinin to chitin.[126] Yet in some cases the CH/π hydrogen bond can also involve the imidazole ring of histidine, as shown for wheat-germ agglutinin whose carbohydrate-binding site includes His66 together with Tyr64 and Tyr73. The studies concerning the mutagenesis of Phe18 of AcAMP2 into several types of amino acid residues including non-natural residues, and characterization by quantitative binding analysis reported by Muraki *et al.* suggest the possibility of strengthening the binding affinity toward carbohydrate ligands by artificial manipulation of CH/π hydrogen bonds *via* protein engineering.[127] In 2002, Muraki wrote a comprehensive review, focusing on the structural aspects of the complexes of these

proteins.[128]

By a combined-mutagenesis and affinity labeling study of human lysozyme (HL), it has been found that Tyr63 ➔ Leu mutant of HL substantially lost the affinity to its specific substrate 2', 3'-epoxypropyl β-glycoside of N, N'-diacetylchitobiose, NAG-NAG-EPO. The result was attributed to the loss of four CH/π hydrogen bonds brought about by the mutation. Fig. 20 illustrates this. Due to the lack of a π-electron system in the side-chain group of Leu63; the L63-HL/NAG-NAG-EPO complex lost four possible CH/π hydrogen bonds compared to the WT-HL/NAG-NAG-EPO complex. By the replacement of Tyr63 with Leu, the decrease in apolar interactions altered the position of NAG131 in the L63-HL/NAG-NAG-EPO complex from its original position in the corresponding WT-HL complex, resulting in a significant decrease of polar hydrogen-bonding interactions. Evidently, the stacking interaction with the aromatic side chain of Tyr63 is essential in positioning the NAG residue in the productive binding mode. Effect of the loss of the above stabilizing interactions in the actual protein-carbohydrate ligand complex seems clear.

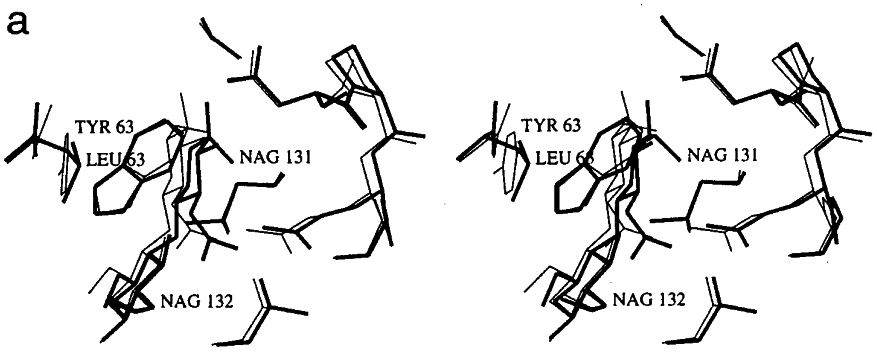


Fig. 20 Stereoview of the superimposed binding structures of the disaccharide part of L63-HL/NAG-NAG-EPO complex (thick line) and WT-HL/NAG-NAG-EPO complex (thin line). Fig. 4 (a) of Muraki *et al.*, *Biochemistry*, 2000, **39**, 292, with permission.

Support for the above finding has been provided by a number of computational studies. For instance, Spiwok *et al.* calculated, at the MP2/6-31+G(d) level approximation, the interaction energy between Trp-999 of *E. coli* β-galactosidase with glucose and galactose moieties of allolactose (-5.2 and -2.4 kcal mol^{-1}, respectively) and glucose moiety of

lactose (-5.0 kcal mol^{-1}). [129] Balaji *et al.* also carried out calculations, on a galactose-specific lectin at the MP2/6-311G++(d,p)//B3LYP/6-31G+(d) level. In every case agreement of the computed results with the experimental data was satisfactory, both in view of the non-bonded distances and energies.[130]

A specific case of protein-carbohydrate interactions concerns glycolipids.[131] These complex lipids are formed by the covalent association of a membrane-embedded apolar tail (which can be either a ceramide or a diacylglycerol) with a sugar head group (mono-, di-, or oligosaccharide). The sugar moiety (referred to as the glycone part) of these glycolipids is often involved in the recognition of proteins through sugar-aromatic CH/π hydrogen bonds.[132] Protein-glycolipid interactions have been characterized for bacterial adhesins[133] and toxins,[134] viral glycoproteins,[135] as well as amyloid-forming proteins such as α-synuclein, PrP and Aβ respectively involved in neurodegenerative Parkinson,[136] Creutzfeldt-Jakob[137] and Alzheimer diseases.[138] Very few structures of protein-glycolipid complexes are available in the PDB. Among these, the case of the *E. coli* Shiga-like toxins is particularly informative. The pentameric B subunit of these toxins interacts with Gb3 (Galα1-Galβ1-4Glcβ1-Cer), a glycolipid expressed by the plasma membrane of epithelial cells (Fig. 21A).

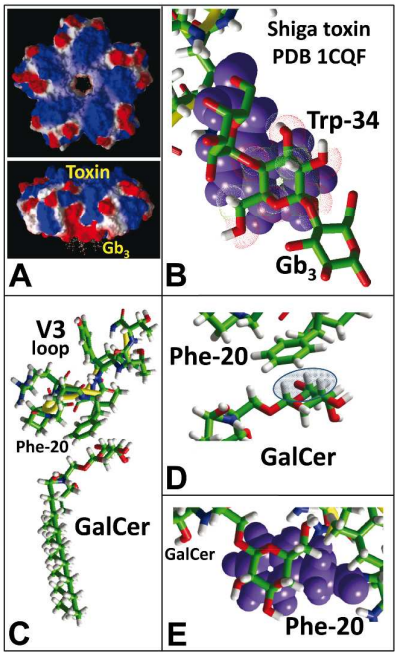


Fig. 21 CH/π hydrogen bonds between proteins and glycolipids. (A) Pentameric structure of the Shiga toxin B subunit complexed with the trisaccharide of its plasma membrane glycolipid receptor Gb_3 (Galα1-Galβ1-4Glcβ1-Cer) (PDB entry 1CQF). In these surface views, acidic regions are colored in red and basic regions in blue. The toxin is viewed above the plasma membrane with the trisaccharide behind the toxin (upper panel) and by the side with the trisaccharide under the toxin (lower panel). (B) Galα1-Galβ1-4Glc trisaccharide bound to chain A of Shiga toxin at a resolution of 2.20 Å. The CH/π hydrogen bonds involve the CH groups 1, 3, 4 and 5 of the central galactose residue of the trisaccharide and the aromatic rings of Trp34 (spacefill rendering in purple) for a total energy of interaction of -6.0 kcal mol^{-1}. (C-E) Molecular modeling simulations of GalCer bound to the V3 loop of the HIV-1 surface envelope glycoprotein gp120.[139] In this case the CH/π hydrogen bonds involve the planar apolar surface of galactose (CH groups 1, 3, 4 and 5) and the aromatic side chain of the phenylalanine residue Phe20 of the V3 loop for a predicted energy of interaction of -5.2 kcal mol^{-1}.

The complex of Shiga toxin B subunit with the trisaccharide of Gb_3 has been obtained at a resolution of 2.20 Å. The central galactose residue of the trisaccharide stacks onto the aromatic ring of the Trp34 (Fig. 21). The pyranose ring of the sugar is in a chair conformation, and its oxygen atom is moved away from the aromatic ring to relax the electrostatic repulsion (compare Figs. 21A and 21B). Although the central galactose of Gb_3 does not have the mobility of a free sugar, it can adapt its conformation within the oligosaccharide so that its apolar surface faces Trp34. Correspondingly, the energy of interaction between Trp34 and the central galactose unit of Gb_3 is high (-6.0 kcal mol^{-1}). The screening of phage display libraries allowed identifying high affinity GM1-binding peptides in which aromatic residues were also critical.[140] Molecular dynamics simulations of glycolipid-protein interactions fully confirmed and extended these structural and physicochemical data.[141] In silico studies suggested that the galactose part of GalCer, a glycolipid acting as a receptor for HIV-1 gp120 on both neural and intestinal cells, stacks onto the aromatic ring of the V3 loop residue Phe20. As for free galactose, the cluster of CH groups 1, 3, 4 and 5 form an apolar surface that is recognized by the aromatic ring of Phe20 (Fig. 21 (C-E)). In full agreement with structural data of both sugar-protein and glycolipid-protein binding, the calculated energy of interaction of the GalCer-V3 loop complex is as high as -5.2 kcal mol^{-1}. Overall these data suggested that

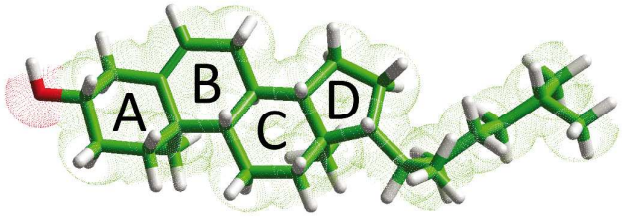
natural glycolipids have the intrinsic capability to bind to proteins through their glycone part just like free sugars bind to proteins

There are other papers dealing with the nature of carbohydrate/aromatic interactions in proteins. The titles and references are given in ESI Table (a).

2.5 Interaction of proteins with cholesterol

Cholesterol is a polycyclic molecule derived from the sterane backbone. It contains four carbon rings referred to as A, B, C and D. It is the principal lipid able to interact with proteins through CH/π hydrogen bonds.

Aromatic rings can stack onto any of these rings and form a CH/π hydrogen bond. For this reason, transmembrane domains of proteins containing such aromatic residues are particularly well adapted to accommodate membrane cholesterol.[142] Recent structural and molecular modelling data strongly support this view. In 2008, Hanson *et al.* described a specific cholesterol-binding site in the 3D structure of the human β_2-adrenergic receptor; a typical G-protein coupled receptor (GPCR) with seven transmembrane (TM) domains.[143] The crystal structure of this receptor in complex with cholesterol was determined to a resolution of 2.8 Å. The cholesterol-binding site was defined as a 3D motif involving four TM domains. Such TM domains consist of a α-helix enriched in apolar amino acid residues that crosses the lipid bilayer. The amino acid side chains are oriented towards the lipids, allowing the establishment of numerous molecular interactions between the TM-domain and surrounding lipids. Interestingly, an aromatic Trp residue that is conserved among GPCRs appeared to play a key role in cholesterol binding. It is located in the fourth TM domain (TM4) of the human β_2-adrenergic receptor at position 158. The complex between cholesterol and this TM domain is shown in Figs. 22A and 22B). This model has been retrieved from the PDB entry 3D4S and visualized with the Hyperchem software. The whole TM4 domain associated with cholesterol is shown in Fig. 22A.



Chemical structure of cholesterol with alphabetical numbering of the four rings.

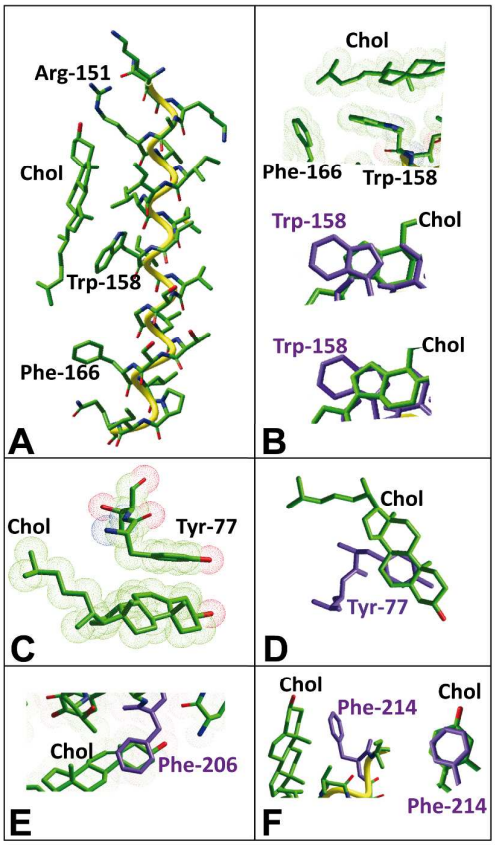


Fig. 22 CH/π hydrogen bonds between cholesterol and aromatic residues of membrane proteins: structural and modelling data. (A) Interaction between cholesterol and the fourth transmembrane domain (TM4) of the human β_2-adrenergic receptor (PDB entry 3D4S). TM4 (151-RVIILMV**W**IVSGLTSFLPIQMHWY-174) is an important part of the cholesterol-binding domain of the receptor. It contains the tryptophan residue Trp158 (bold and underlined in the sequence) whose aromatic side chain stacks onto cholesterol through a CH/π hydrogen bond with ring D of the sterol. (B) Distinct views of the

stacking between Trp158 of the human β_2-adrenergic receptor and cholesterol. The involvement of the D ring of cholesterol in CH/π hydrogen bonds is particularly well visible in the bottom model, with cholesterol carbons in green and Trp158 in purple. (C) Docking of cholesterol to a typical CRAC domain located near the second transmembrane domain (TM2) TM2 domain of the human delta-type opioid receptor. The aromatic side chain of Tyr77 plays a critical role in this interaction. (D) The aromatic side chain of Tyr77 (in purple) stacks onto ring B of cholesterol (carbon atoms in green) through a CH/π hydrogen bond. (E) Docking of cholesterol to a typical CARC domain located in the fifth transmembrane domain (TM5) of the human type-3 somatostatin receptor. In this case, the phenyl ring of Phe206 (in purple) stacks onto ring A of cholesterol. (F) Docking of cholesterol to the unique transmembrane domain (TM) of human CD40 showing a CH/π hydrogen bond between ring A of cholesterol and the phenyl side chain of Phe214.

Several residues including Trp158 are involved in cholesterol binding, for a total energy of interaction estimated to be -10.6 kcal mol^{-1}. By itself, Trp158 accounts for -5.4 kcal mol^{-1}, which represents 50% of the total energy of interaction. Although the aromatic indole group of Trp158 is not parallel to cholesterol, it faces ring D cholesterol at mean distance of 3.75 Å (from carbon to carbon). This is consistent with the establishment of a coordinated network of CH/π hydrogen bonds resulting in a stable TM4-cholesterol complex. Interestingly, the cholesterol-binding site of the Na^+/K^+ ATPase revealed a conserved Tyr40 residue on the β-subunit that, similar to Trp158 in the β_2-adrenergic receptor, is in a stacking interaction with the rings of cholesterol.[144] As a matter of fact, the efficient contribution of aromatic rings to cholesterol binding is a hallmark of protein-cholesterol interactions.[145] Indeed, the presence of at least one aromatic residue at a central position is an absolute requirement of sequence-based cholesterol consensus domains. The **C**holesterol **R**ecognition/interaction **A**mino acid **C**onsensus sequence, referred to as **CRAC** domain[146] is a short linear motif which fulfils the following algorithm: (Leu/Val)-X_{1-5}-(Tyr)-X_{1-5}-(Lys, Arg), where X can be any of the 20 amino acids. The CRAC motif has been found in various proteins known to bind cholesterol and in many cases the interaction between cholesterol and CRAC has been confirmed by physicochemical approaches.[147] Given the scarcity of 3D structures for most membrane

proteins, structural homology modelling coupled with docking studies can shed some light on the molecular mechanisms involved in protein-cholesterol interactions.[148]

An example of a potential CRAC-cholesterol complex studied by such in silico approaches is illustrated in Figs. 22C and 22D. This CRAC domain (74-I**V**R**Y**T**K**M**K**-81) is located near the second TM domain (TM2) of the human δ-type opioid receptor. It has been predicted to display a high affinity for cholesterol, with a total energy of interaction of -11.7 kcal mol^{-1}. The aromatic side chain of Tyr77 binds to cholesterol through a CH/π hydrogen bond with ring B of cholesterol. This aromatic residue contributes -6.2 kcal mol^{-1}, which represents about 50 % of the total energy of interaction of the CRAC/cholesterol complex. This is fully consistent with the structural data obtained for the β_2-adrenergic receptor. Recently, a second consensus cholesterol binding motif has been unveiled in the nicotinic acetylcholine receptor and referred to as an inverted CRAC motif defined by the following 'CARC' algorithm: (Lys/Arg)-X_{1-5}-(Tyr/Phe)-X_{1-5}-(Leu/Val).

The fifth TM domain (TM5) of the human type 3 somatostatin receptor contains such a CARC domain (Fig. 22E): 203-**R**AG**F**II**Y**TAA**L**-213. There are two aromatic residues in this motif, and both are predicted to contribute to cholesterol binding. The main contribution is provided by Phe206, whose phenyl group stacks onto ring A of cholesterol through near-perfect CH/π hydrogen bonds. Among the three aromatic residues, Phe is unique in that its side chain does not contain any polar group or atom. Thus, Phe is particularly suited to provide an apolar planar surface with which cholesterol rings can interact in the most apolar zone of the membrane bilayer. On the basis of this consideration, it is not totally surprising that optimal parallel fits could arise from cholesterol/Phe interactions. This is illustrated by the molecular docking of cholesterol to the TM domain of human CD40, a member of the tumor necrosis factor receptor family expressed by antigen-presenting cells and recruited in cholesterol-enriched plasma membrane domains.[149] This domain neither contains a CRAC nor a CARC motif, but its Phe214 residue could efficiently stack onto ring A of cholesterol, thereby allowing the establishment of a CH/π hydrogen bond (Fig. 22F). A similar case concerns NPC1, a protein that regulates intracellular cholesterol transport. The aromatic residue Phe108 of NPC1 is not part of a consensus cholesterol-binding motif, but structural data indicate that it interacts with ring D of cholesterol, consistent with the establishment of a CH/π

hydrogen bond at a mean distance of 3.60 Å from carbon to carbon.[150]

Overall these structural and modeling data underline the critical role of CH/π hydrogen bonds between cholesterol and aromatic residues of proteins. One important parameter that probably favours such interactions is that the aromatic structure is linked to the protein backbone by a CH_2 group that may act as a hinge. This hinge allows the rotation of the aromatic ring so that it can freely adjust its orientation with respect to cholesterol. This can be viewed as a fine-tuning of the aromatic structure leading to an optimal geometry for a CH/π hydrogen bond. In absence of a functional hinge, the energy of interaction between cholesterol and an inert aromatic surface would be significantly weaker, as recently calculated for carbon nanotubes and graphene.[151] Incidentally, the amino acid residues of the TM domain that surround the aromatic residue might affect the capability of the latter to optimize the stacking interaction with the sterol rings. Indeed, in addition to the critical aromatic residue, the cholesterol-binding site of several membrane proteins also contains Ser and/or Gly residues whose function is to secure the orientation of the aromatic side chain to optimize a stacking interaction with the sterol rings.[152] A challenge of future 3D studies of proteins will be to understand how secondary residues - that do not bind to cholesterol - contribute to cholesterol binding through fine tuning of aromatic side chains that are involved in CH/π hydrogen bond with cholesterol rings.

2.6 Other topics

In the 1980s, Sigel and his group studied, by potentiometric titrations and NMR spectroscopy, the solution conformation of a series of ternary complexes between a metal ion, an amino acid and a bipyridyl (bipy) or phenanthroline (phen) moiety.[153] Though not strictly related to biological macromolecules, such complexes are very useful to focus on specific interactions. Examples of investigated complexes are: M(bipy)(AA) and M(phen)(AA) [M = Cu^{2+}; bipy = 2,2'-bipyridyl; phen = 1,10-phenanthroline; AA = $H\text{-}(CH_2)_n\text{-}CH(NH_2)COO^-$ [Ala^- (n = 1), Aba^- (n = 2), Nva^- (n = 3), Nle^- (n = 4)] or $(CH_3)_2CH\text{-}(CH_2)_n\text{-}CH(NH_2)COO^-$ [Val^- (n = 0), Leu^- (n = 1)] or $CH_3CH_2(CH_3)CH\text{-}CH(NH_2)COO^-$ [Ile^-].It could be observed that in the complexes the alkyl groups of the amino acids were curled over the aromatic rings to form a folded conformation. Fig. 23 illustrates this for the Leu^- complex. The proportion of the folded species increased according to the size of the alkyl moiety and reached for

alkylcarboxylates[154] with 6-methylheptanoate a plateau. The result is in accord with the formation of a CH/π hydrogen bond.

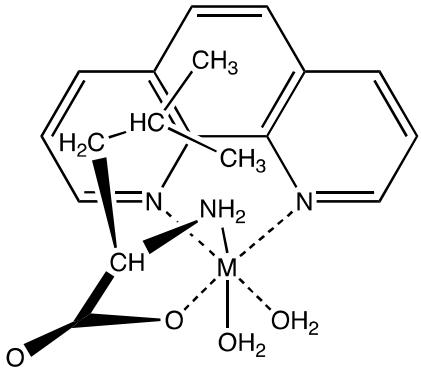


Fig. 23 Folded conformation of M(phen)[$(CH_3)_2CH\text{-}CH_2\text{-}CH(NH_2)COO^-$].

Zsila and coworkers investigated the induced CD activity of β-lactoglobulin/retinoic acid complexes correlated with the helically distorted, chiral conformation of the ligand molecule entrapped in the central cavity of the protein.[155] Among 10 aromatic residues surrounding the retinoic acid, Phe82 and Phe105 were found very close to a methyl group and axial CHs of the ligand cyclohexene ring. It was proposed that CH/π hydrogen bonds between retinoic acid and β-lactoglobulin residues are important in stabilizing the structure of the complex.

The nature of binding of a pheromone-binding protein was studied by *ab initio* MO calculations.[156] Analysis of the crystal structure of the protein in complex with a pheromone (bombykol) identified that a number of aromatic residues (Phe12, Trp35, Phe36, Phe76, Phe118) are involved in a variety of intermolecular interactions. Using simple model fragments as representatives of these residues, the interaction energies of their complexes were estimated. This enabled to describe the nature and origin of the binding forces in terms of the contribution of the individual amino acids and types of interaction such as, CH/O, CH/S, CH/π and aromatic CH/π hydrogen bonds to the overall stability. Interaction energy of Phe12 and Phe 118 with the pheromone indicated a total of -8.2 kcal mol^{-1} (Fig. 24).

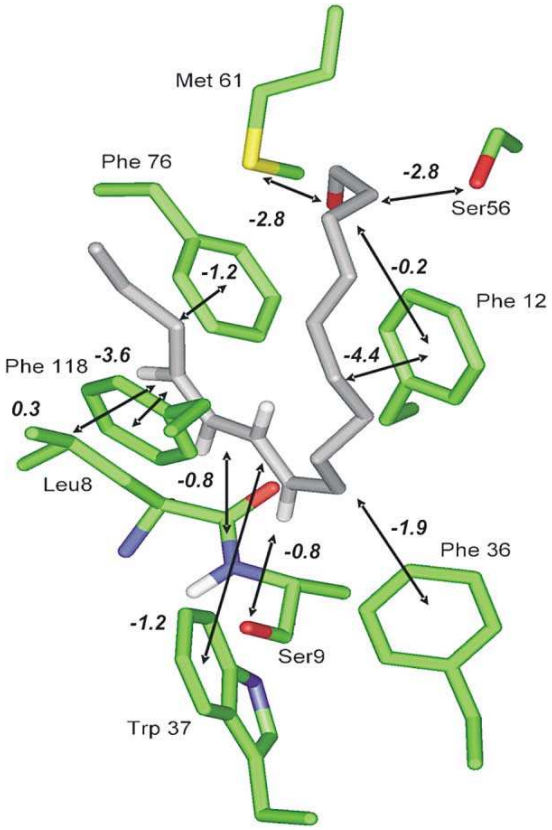


Fig. 24 Schematic representation of the binding cavity, showing bombykol as well as the side-chains of the neighbouring amino acids. The arrows show the main interactions of the pheromone in the binding pocket. Interaction energies are given in kcal mol^{-1}. Fig. 2 of Klusák *et al*., *Chem. Biol.*, 2003, **10**, 331, with permission.

The solution NMR structure of a pheromone-binding protein, a carrier transporting volatile pheromone molecule to the membrane-bound GPCR proteins was studied.[157] The cavity of the pheromone-binding protein has been found lined with side-chains from nine helices. The authors suggested that Trp37 plays an important role in the initial interaction with the ligand, while Phe12, Phe36, Trp37, Phe76 and Phe118 are responsible for the binding of pheromones.

The structural role of carotenoids in the photosystem I was studied by quantum chemical analysis.[158] It has been found that aromatic CH/π hydrogen bonds or π/π stacking, CH/π hydrogen bonds between aromatic side-chains and β-carotene and

chlorophyll are responsible in stabilizing the structure of the photosynthetic complexes.

The structure of photosystem I, an electron-transfer centre of plants, algae, *etc.*, was studied. It has been found that a number of CH/π hydrogen bonds play a role between aromatic residues.[159] A number of CH/π hydrogen bonds are formed by use of aromatic CHs of PheB691, PheB667, PheA600, TyrA735, *etc*. Fig. 25 illustrates this.

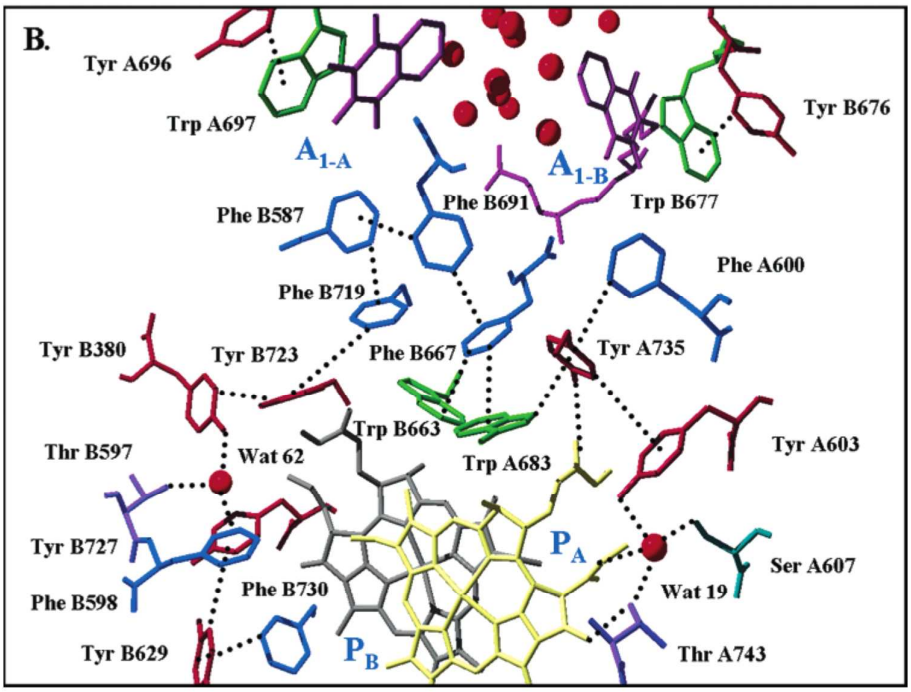


Fig. 25 Crystal structure of *S. elongatus* PSI showing aromatic CH/π hydrogen bonds with the electron-transfer cofactor. Fig. 1 of Sacksteder *et al., J. Am. Chem. Soc.,* 2005, **127**, 7879.

The structure of a TSG-6/hyaluronan complex was studied by modeling and NMR spectroscopy.[160] Thus involvement of aromatic residues such as Phe70, Tyr59 and Tyr58 has been found; implication to the CH/π hydrogen bonds was suggested.

Tóth and Borics studied the mechanism of closing the flaps of retroviral aspartic protease (HIV-1 PR) by molecular dynamics (MD) simulation.[161] An extensive conformational change of the flaps has been found to take place upon binding of substrate to the active site. The nature of the residues of HIV-1 PR has been shown important in the flap closing mechanism, which is conserved across known structures of retroviral aspartic

proteases family. As to the opening process, X-ray diffraction studies suggested that in the free enzyme the two flaps are loosely packed onto each other in a semi-open conformation, while MD studies have shown that the flaps can also separate into open conformations.[162] The flaps showed complex dynamic behaviour as two distinct mechanisms of flap opening and various stable flap conformations (semi-open, open, and curled) were observed during the simulations. A network of weak hydrogen bonds between the flaps was proposed to be responsible for stabilizing the semi-open flap conformation (Fig. 26).

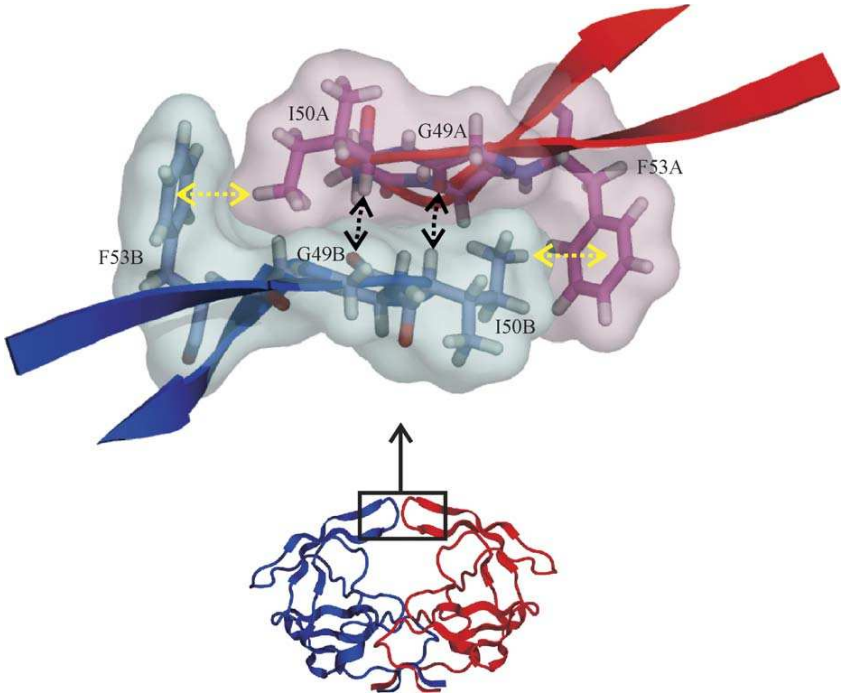


Fig. 26 Selected weak interactions between the two flaps of HIV-1 PR. Residues Gly49, Ile50 and Phe53 are shown in both surface and stick representations. Backbone is drawn in cartoon representations: monomer A is in red, while monomer B is in blue. Yellow arrows indicate CH/π hydrogen bonds, while black arrows indicate CH/O=C interactions. Fig. 7 of Tóth and Borics, *J. Mol. Graph. Model.*, 2006, **24**, 465, with permission.

CH/π hydrogen bonds were examined in 59 RNA-binding proteins. In the 59 crystal structures studied, it has been found that an average of 55 CH/π hydrogen bonds per protein. The donor atom contribution to CH/π hydrogen bonds was mainly from Phe, Tyr,

Trp, Pro, Gly, Lys, His and Ala.[163] The acceptor residues were Trp, Phe and Tyr. Significant proportion of CH/π interacting residues had one or more stabilization centres. The authors also explored the roles played by CH/π hydrogen bonds in twelve interleukin structures in the PDB.[164] There they found an average of 15 CH/π hydrogen bonds per protein and also there was an average of one significant CH/π hydrogen bond for every 14 residues, in the interleukins investigated. Trp contributed both donor and acceptor atoms in main-chain to side-chain, main-chain to side-chain 5 member aromatic ring and side-chain to side-chain CH/π hydrogen bonds.

The role played by Tyr92 in the extrinsic PsbU protein of red algal photosystem II was examined.[165] Mutagenesis experiments have revealed that Tyr92 was essential for maintaining its function. From the crystal structure analysis of the photosystem, Tyr92 was found located close to Pro340, suggesting that the aromatic ring of Tyr92 interacts with CH groups of Pro340; they attributed the result to the CH/π hydrogen bond.

The crystal structures were examined for profilin IIa in complex with two peptide ligands bearing a number of Pro residues: formin homology 1 of *Drosophia diaphanous* (mDia1) and vasodilator-stimulated phosphoprotein (VASP).[166] The ligands were found to bind aromatic residues, Phe, Tyr and Trp, which are well conserved in profilin. The crystallographic data allowed detection of conserved CH/π hydrogen bonds between the peptides and profilin in the complexes. Fig. 27 shows that the peptides from VASP and mDia1 bind to profilin IIa in a distinct mode.

Fig. 17

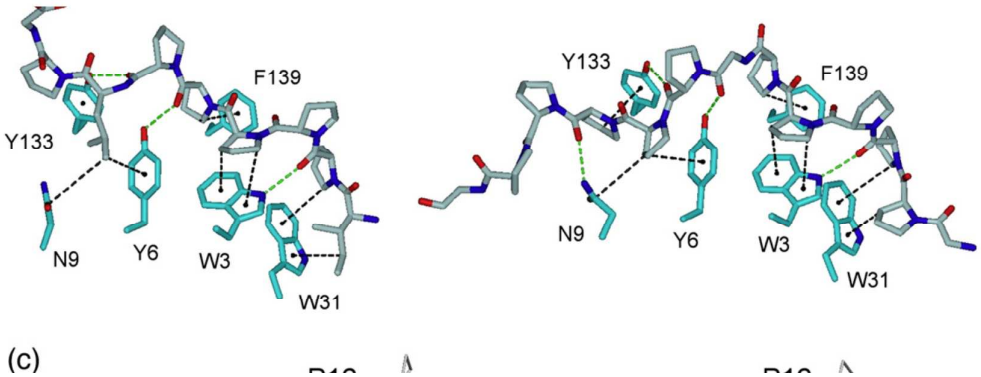


Fig. 27 (a) Details of peptide binding in the profilin/mDia1 complex. (b) Profilin/VASP complex. Broken green lines and black lines indicate hydrogen bonds and CH/π hydrogen

bonds, respectively. From Fig. 4 of Kursula *et al., J. Mol. Biol.*, 2008, **375**, 270, with permission.

The binding of Pro-rich segments of myelin basic protein to SH3 domains was investigated by spectroscopic and computational methods.[167] Molecular docking simulations of the interaction of a putative SH3 ligand of with the human *Fyn* SH3 domain has indicated that the molecular recognition and association is mediated by CH/π hydrogen bonds between Pro residues of the ligand and Trp119 and Tyr132 of the SH3 binding site.

Divergent patterns within a multiple alignment of *Ras*-like GTPase sequences were analyzed. A structural element termed the glycine brace has been identified, as the feature distinguishing *Rab, Rho/Rac, Ran* and *Ras* family GTPases from other *Ras*-like GTPases.[168] The glycine brace consists of an aromatic residue that forms a stabilizing CH/π hydrogen bond with a conserved Gly at the start of the guanine-binding loop; a second aromatic residue, Trp, that forms stabilizing CH/π and NH/π hydrogen bonds with Gly, and two other residues that form a network of interactions connecting the two aromatic residues. It was proposed that this component stabilizes binding of guanine nucleotide in the absence of associated exchange factors, but that exchange factor-induced disruption of the CH/π and NH/π hydrogen bonds leads to release of GDP by destabilizing nucleotide binding loops at their Gly residue hinge points. This illustrates how CH/π hydrogen bonds can contribute to biological function, namely the signaling pathway on-off switching mechanism associated with these GTPases.

The chemistry and diverse actions of neonicotinoids (Nic) on insect and mammalian nicotinic ACh receptors (nAChRs) was reviewed.[169] Electrophysiological studies on native nAChRs and recombinants have shown that basic residues particular to loop D of insect nAChRs are likely to interact with the nitro group of Nics. The crystal structure showed that 1) glutamine in loop D, corresponding to the basic residues of insect nAChRs, hydrogen bonds with the NO_2 group and 2) Nic-unique stacking and CH/π hydrogen bonds at the ligand-binding domain. The prospects of designing Nics, which are safe not only for mammals but also for beneficial insects such as honeybee are discussed in terms of interactions with non-α nAChR subunits.

An antimicrobial 21-residue cyclic peptide, arenicin-2, was studied by NMR and CD spectra.[170] It has been suggested that the structure of arenicin-2 in water represented a

well formed, but highly twisted β-hairpin, stabilized by noncovalent interactions. To investigate the spatial arrangement of the peptide side-chains and to get a clear view of its possible amphipathic properties, the authors performed MD simulations in explicit water. Arenicin-2 retained its β-hairpin structure during simulations, although the residues close to the strand ends were found to escape from the ideal hairpin conformation. Several nonbonded interactions, like nonpolar interactions between aliphatic side chains, aromatic CH/π hydrogen bonds, and water bridges, contributed to the hairpin stabilization.

CH/π, OH/π, and NH/π hydrogen bonds were studied for complexes of Phe in its cationic, anionic, neutral and zwitterionic forms with CH_4, H_2O, NH_3 and NH, by DFT calculations such as MPWB1K and M06-2X.[171] All noncovalent interactions were identified by the presence of bond critical points (bcp) of electron density ρ(*r*) and the values of ρ(*r*) showed a linear relationship to the binding energies. Solvation has no significant effect on the overall geometry of the complex though slight weakening of noncovalent interactions by 1-2 kcal mol^{-1} is observed. These workers also devised a prediction tool, HBPredicT, for detecting structural water molecules and CH/π hydrogen bonds in PDB files of protein-ligand complexes.[172] This program was tested using 19 HIV-1 proteases and 11 PTP1B-inhibitor complexes.

A distinct role of non-covalent interactions to the function and structural stability of glutaredoxins was investigated.[173] Among the proteins, an average of 1 CH/π hydrogen bond per 16 residues have been found. These interactions were influenced by long-range contacts, whereas short- and medium-range contacts were found insignificant.

Interactions involving CH/π hydrogen bonds on the structural stability of immunoglobulin proteins were examined.[174] A total of 128 CH/π hydrogen bonds were identified in 33 structures retrieved from the PDB. The most prominent representatives are the interactions between aromatic CH donor groups and aromatic π acceptor groups. The secondary structure preference, solvent accessibility and stabilization centres of CH/π interacting residues were estimated. Their study has shown that 46% of the donor residues and 52% of the acceptor residues are well conserved. They concluded that the CH/π hydrogen bond could be categorized as a true stabilizing force in immunoglobulins.

To improve the catalytic activity and enantioselectivity of *B. cepacia* lipase, the enzyme was redesigned, by manipulation of the transition state of the reaction.[175] A double mutant I1287F/I1290A showed a conversion rate and *E*-value (>200) toward a poor substrate,

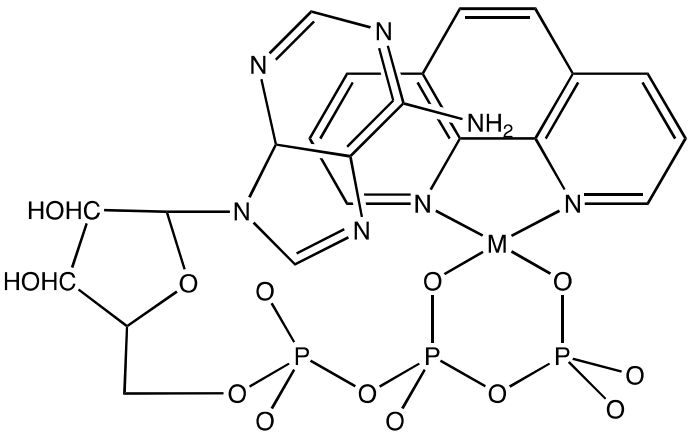
while the WT enzyme showed much less conversion rate and *E*-value (*ca*. 5). The result was interpreted as a consequence of CH/π hydrogen bonds between Phe287 and the substrate.

The pharmacological relevance of weak hydrogen bonds (CH/π, OH/π, NH/π, CH/O, CH/FC and OH/FC hydrogen bonds) and aromatic CH/π hydrogen bonds was reviewed.[176] Reviews dealing with the rational drug design are available.[177] Readers interested in structure-based drug design are referred to topics and discussions appeared in the above treatises.

There are other papers dealing with CH/π hydrogen bonds interactions in proteins. The titles and references are given in ESI Table (b).

3. CH/π hydrogen bonds involving nucleic acids

In 1975 and subsequent years, Sigel and his group studied, by NMR, a number of ternary metal complexes bearing aromatic moieties at both termini of the ligands such as Trp and ATP,[178] and ATP and 1,10-phenanthroline (phen).[179] Folded conformations generally occur in complexes with aromatic ligands, as illustrated in Fig. 28.[180] At the outset, the reason for this phenomenon was not necessarily apparent, however, it has become gradually clear that aromatic CH/π hydrogen bonds or π/π stacking interactions play an important role. In 2004, Sigel reviewed the aspect of coordination chemistry of a multitalented biological substrate, using adenosine 5'-triphosphate as an example.[181] Sigel and Griesser wrote a comprehensive review dealing with the self-association, acid-base and metal ion-binding properties of nucleoside 5'-triphosphates in solution.[182]



65

Fig. 28 Ternary metal ion complex with ATP^{4-} and 1,10-phenanthroline as ligands.

The conformational equilibria of ternary complexes were studied, by potentiometric pH titrations of a series of acyclic nucleoside phosphonates (ANP; R = H or NH_2 and R'= H, NH_2 or NMe_2) with antiviral properties. Detailed stability-constant comparisons revealed that aromatic CH/π hydrogen bonds or π/π stacking between the aromatic rings (phen or bipy) and the purine moiety within the Cu^{2+} complexes is important.[183] The stability enhancements were attributed to intramolecular-stack formation in the Cu(phen)(ANP) complexes (Fig. 29).[184] The quantitative analysis of the intramolecular equilibria showed that the formation degrees of the π/π stacked conformation was 88-97%. The result is consistent with NMR data.

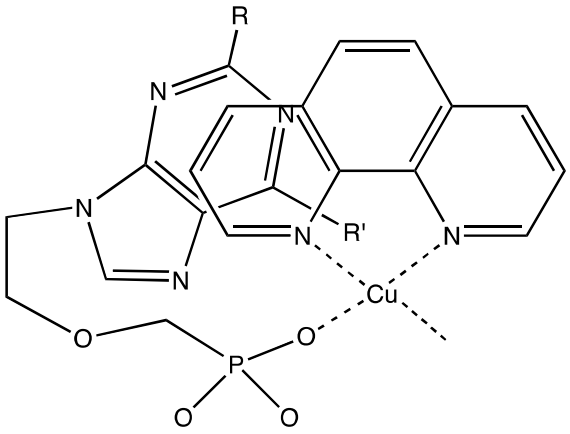


Fig. 29 ANP^{2-}/Cu^{2+}/1,10-phenanthroline complex.

In 1978, Viswamitra, Kennard and coworkers determined the crystal structure of a deoxytetranucleotide 5'-P-adenylyl-3',5'-thymidylyl-3',5'-adenylyl-3',5'-thymidine (pATAT).[185] An extraordinary feature of their finding was that the conformation of the sugar-phosphate backbone of pATAT is significantly different from the canonical structure of B-DNA. Thus the first two bases A and T pair with the complementary bases from another molecule, but then the phosphate backbone swings away from the helical orientation, to make another pairing of the subsequent A-T sequence with a third

molecule. T is stacked with the preceding adenine base, but there is no interaction of T with A next to it. Klug *et al*. noted that the methyl group of T in an A-T step places itself over the π-ring of the purine, whereas in the T-A step there was no such an interaction. They argued this as the primary cause of the alternating B-DNA structure and suggested importance of the CH_3 group in thymidine.[186]

Inspired from the above findings, Umezawa and Nishio analyzed all crystal structures of DNAs from the Nucleic Acid Database by program CHPI.[187] In every case the structure of pATAT has been shown stabilized by the 5-methyl group in T with the A π-ring preceding it. The structure of DNAs incorporating a long sequence of A (A-tracts: AAAA, *etc*.) was also analyzed. Since an A-tract is lined with an oligo-T sequence (TTTT, *etc*.) in the complementary strand, a successive N/T-CH_3 stacking may contribute in making the A-tracts robust and straight; this is a result consistent with biochemical findings. The role of methyl groups in modified DNA was discussed on a similar basis (Fig. 30). Notice that the CH_3 group of 5-methyl cytosine is interacting with a base ring preceding it.

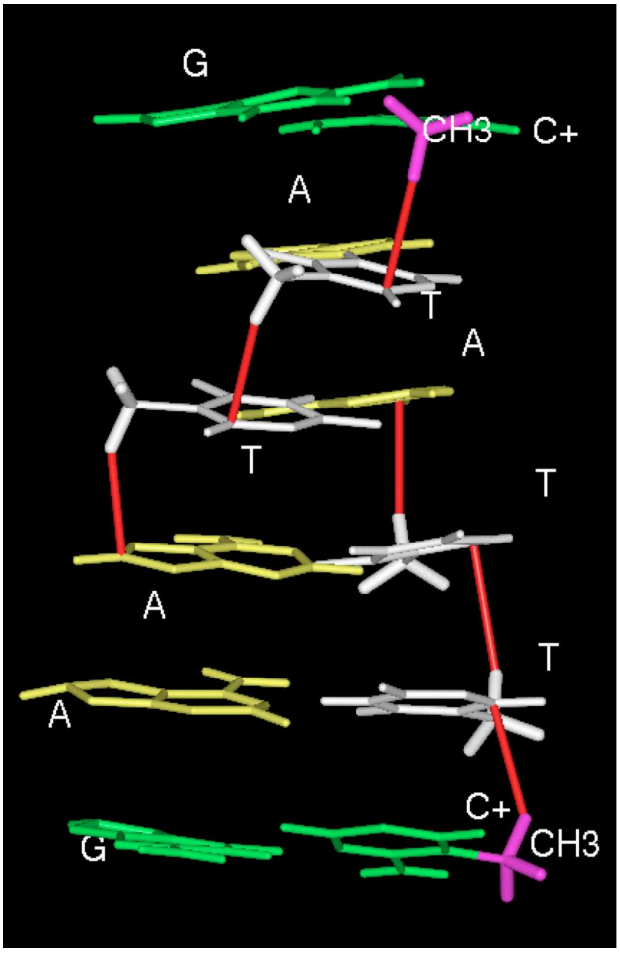


Fig. 30 A part of the crystal structure of CGCGAATTC^{+}GCG (BDLB73), C^{+}: 5-methyl cytosine). Fig. 7 of Umezawa and Nishio, *Nucleic Acids Res*., 2002, **30**, 2183, with permission.

The crystal structure of TATA-box binding proteins (TBP) of various sources in complex with the promoter was examined. A number of short CH/C^{sp2} contacts have been unveiled in these complexes at the boundary of TBP and the DNA minor groove. Thus, the nature of nonpolar forces, reported in the past at the interface of the two components, was attributed to the CH/π hydrogen bond.[188] Furthermore, many CH/π contacts have been disclosed within the same strand of the promoter DNA. The structure of the TATA element, partially unwound and severely bent on complexation, seems to be stabilized by CH/π hydrogen bonds; H2' of the deoxyribose moiety and the methyl group in T nucleotides also play important roles.

Waters and his group examined the recognition of nucleotides with model β-hairpin receptors and found that a relationship exists between critical contacts and nucleotide selectivity.[189] The order of binding affinity was shown to follow dTTP > GTP > ATP > CTP, with differences in binding energies spanning as much as 1.6 kcal mol^{-1}. NMR analyses demonstrated that both aromatic interactions with the Trp side-chains and CH/π hydrogen bonds between the ribose protons and the Trp residues might contribute significantly to the binding.

Cox and coworkers analyzed the π/π stacking interactions between an aminoglycoside antibiotic kinase and its nucleotide ligands.[190] A key contact in the active site of this enzyme is a π/π stacking interaction between Tyr42 and the adenine ring of the bound nucleotides. They investigated the prevalence of similar Tyr-adenine contacts and found that many different protein systems employ Tyr residues in the recognition of the adenine ring. The geometry of these stacking interactions suggests that electrostatics play a role in the attraction between these aromatic systems. Kinetic and calorimetric experiments on wild-type and mutant forms of the enzyme yielded further experimental evidence of the importance of electrostatics in the adenine-binding region and suggested that the stacking interaction contributes approximately 2 kcal mol^{-1} of the binding energy. This type of information concerning the forces that govern the nucleotide binding will facilitate inhibitor design strategies that target the nucleotide-binding site of APH-type enzymes.

Yahi *et al.* used molecular docking techniques to delineate the ATP binding pocket in various mutant reverse transcriptases from clinical HIV-1 isolates.[191] They showed that the binding of ATP to Thr215Tyr mutant reverse transcriptase involves i) π/π stacking between adenine and the aromatic ring of Tyr215, and ii) CH/π hydrogen bonds between Tyr215 and the furanose ring of ribose. This unique combination of π/π and CH/π hydrogen bonds ensures a very efficient binding of ATP in its active, phosphorolysis-competent conformation. Correspondingly, the suboptimal stacking capability of the Thr215Phe mutant reverse transcriptase is associated with a lower ATP-dependent phosphorolysis activity compared with the Thr215Tyr mutant.

In 2008, Morales *et al*. reported on the highly polar carbohydrates stack onto DNA duplexes via CH/π hydrogen bonds.[192] Carbohydrate-nucleic acid contacts are known to be a fundamental part of some drug-DNA recognition processes. Most of these interactions occur through the minor groove of DNA. Highly polar mono- and disaccharides are capable of aromatic CH/π hydrogen bond onto the terminal DNA base

pair of a duplex as shown by NMR spectroscopy. These results reveal carbohydrate-DNA base stacking as a potential recognition motif to be used in drug design, supramolecular chemistry or bio-based nanomaterials.

In 2012, Lucas *et al*. investigated carbohydrate/aromatic interactions using a dangling-end DNA model system.[193] Every conjugate containing a carbohydrate moiety showed higher stability than control conjugates. It has become clear that the interaction of axial CHs of the carbohydrate with the benzene π-system contributes in stabilizing the structure of the dangling-end DNA. Undoubtedly, the interaction is enthalpy driven. Contribution from CH/π hydrogen bonds between the carbohydrates and the benzene ring seems clear. Neither the conventional hydrogen bond nor the so-called hydrophobic effect plays any significant role. It has been demonstrated also that the carbohydrate/DNA interactions are possible through sugar capping of a DNA double helix. Highly polar mono- and disaccharides are also capable of establishing CH/π hydrogen bonds with the terminal DNA base pair of a duplex as has been demonstrated by NMR spectroscopy.

In 2007, Bertran and coworkers studied, by DFT calculations (MPWB1K/6-31+G(d,p)), the mutual relationship between stacking and hydrogen bonding and the possible influence of stacking in the different behaviour of DNA and RNA base pairs.[194] The different behaviour of DNA and RNA when replacing U by T has been interpreted through the formation of a stabilizing CH/π hydrogen bond between the methyl group of T and the five-member ring of A (Fig. 31). This is expected given that the increase of polarization produced by the methyl group enhances the stabilization of the system through CH/π hydrogen bonds.

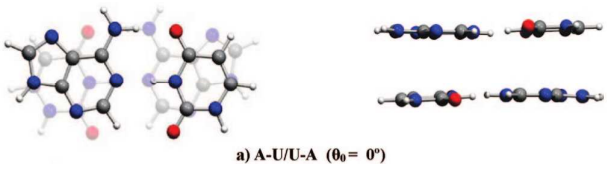


Fig. 31 Optimized structure of an AT/TA dimer with Watson-Crick base-pairing. Notice that methyl group of T is pointing to the five-member ring of A. Fig. 3 of Gil *et al., J. Phys. Chem. B*, 2009, **113**, 4907, with permission.

Molecular modeling, absorption studies and binding-constant measurements support

the different binding patterns. The interaction of amiloride with T was found to depend on the bases flanking the AP site and different binding modes were observed for different flanking bases. Molecular modeling, absorption studies and binding-constant measurements supported the different binding patterns. The flanking base dependent recognition of AP site phosphates was investigated by ^{31}P NMR experiments. The thermodynamics of the ligand-nucleotide interaction was demonstrated by isothermal titration calorimetry. The emission behaviour of amiloride was found to depend on the bases flanking the AP site.

In 2012, the structural basis for the transcriptional regulation of heme homeostasis in *Lactococcus lactis* was studied.[195] The crystal structure of heme-regulated transporter regulator (HrtR) in complex with its target DNA has indicated that two residues, Arg46 and Tyr50, of HrtR play a crucial role for sequence-specific DNA binding using a hydrogen bond and a CH/π hydrogen bond. The methyl group in T12 is oriented perpendicularly to the aromatic ring of Tyr50, the distance between the T-methyl carbon and Cγ of Tyr50 is in a reasonable range in view of the CH/π hydrogen bond (Fig. 32).

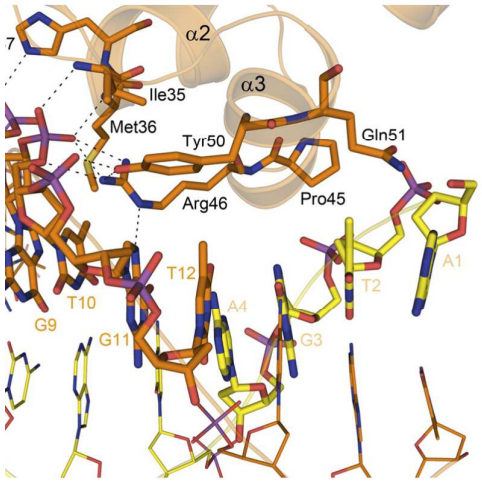


Fig. 32 Interactions of Arg46 and Tyr50 with the DNA bases G11 and T12. Numbers near the dotted lines represent the distances between the two atoms. From Fig. 5 of Sawai *et al., J. Biol. Chem.,* 2012, **287**, 30755, with permission.

The crystal structure was studied for an aminoglycoside antibiotic sisomicin (with an

unsaturated 6-membered ring) bound to the bacterial ribosomal decoding site and compared to a gentamicin C1a (with a saturated ring) complex.[196] A remarkable difference has been noted in the stacking interaction between a ring and G1491 in the bacterial A-site helix. While the typical saturated ring of gentamicin stacks on the aromatic G1491 ring through CH/π hydrogen bonds, the unsaturated ring of sisomicin shares its π-electron density with G1491 and fits well within the A-site helix.

There are other papers dealing with CH/π hydrogen bonds in nucleic acids. The titles and references are given in ESI Table (c).

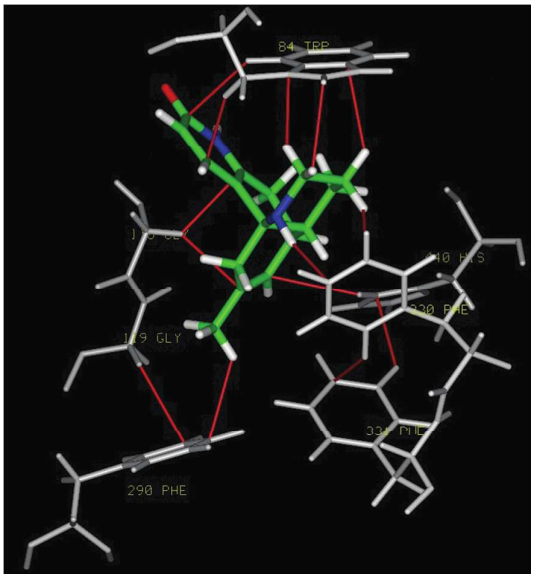
4. Rational drug design

4.1 Structure-based drug design (SBDD)

In cannabinoid receptor CB1, a member of GCPR family, transmembrane helix (TMH) region 3-4-5-6 includes a microdomain, comprised of a number of aromatic residues. McAllister *et al.* investigated interactions involved in the microdomain at the cannabinoid CB1-agonist and inverse agonist-binding region. NMR, MD simulations and modeling studies suggested that in the inactive state of CB1, the binding site of a CB1 inverse agonist/antagonist within the trans-membrane helix (TMH) region 3-4-5-6 involves aromatic CH/π hydrogen bonds with Phe194, Phe244, Phe381, Phe432 and Trp531. In the binding pocket, the CB agonist anandamide binds in the TMH region 2-3-6-7 in which CH/π hydrogen bonds between Phe3.25 and the C=C double bonds in the ligand seem to be important.[197]

Umezawa and Nishio retrieved the crystal structure of acetylcholine esterase (AChE) in complex with twelve effective inhibitors from the PDB and analyzed by CHPI.[198] In these compounds, N^+ is absent except for two cases, suggesting that the ammonium cation plays little role if any in the biding with AChE. Many CH/π hydrogen bonds have in fact been disclosed in the binding pocket of AChE (Fig. 33). For instance the number of short CH/π contacts is greater in the AChE/HupA-like dimer E10 complex than in the AChE/(-)Huperidine B complex. The inhibitory activity of the former (K_i 0.8 n*M*) is much greater than that of the latter (K_i 334 n*M*). This shows that CH/π hydrogen bonds paly an important role in stabilizing the structure of the complexes.

(A)



(B)

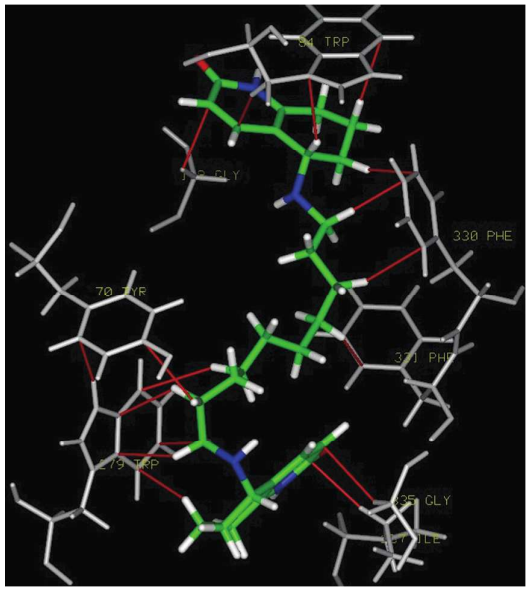


Fig. 33 (A) AChE/(-)Huperzine B complex. (B) AChE/HupA-like dimer E10 complex. Umezawa and Nishio, *Biopolymers,* 2005*,* **79**, 248, with permission.

In 2005, Irie and coworkers found that indolactam-V (IL-V) is involved in the CH/π interaction with PKCδ C1B domain (Fig. 34).[199] Compared to the wild-type PKCδ C1B

peptide, a lower binding affinity was noted to the mutant peptide with a fluorine-substituted proline, suggesting that the CH/π hydrogen bond plays a pivotal role in the binding of IL-V to the PKCδ C1B domain. On the other hand, the binding affinity of benzolactam-V8 (BL-V8) with a benzene ring instead of the indole ring of IL-V was weak. This might be ascribable to lack of the CH/π hydrogen bond, and the binding affinity was enhanced by effective formation of the CH/π hydrogen bond as exemplified by the synthesis of naphtholactam-V8 (NL-V8).

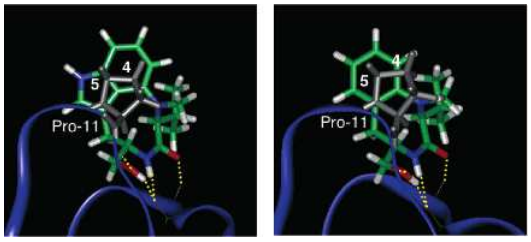


Fig. 34 Docking models of IL-V and BL-V8 with the PKCδ C1B domain. Yellow dotted lines represent hydrogen bonds. Fig. 2 of Nakagawa *et al. J. Am. Chem. Soc.* 2005, **127**, 5746, with permission.

In 2010, Boger and coworkers determined the crystal structures of several α-ketoheterocycle inhibitors bound to a variant of fatty acid amide hydrolase (FAAH).[200] They analyzed interactions involved in the active site and discussed the result in view of the structure-activity relationship. Fig. 35 is a part of their figures reporting the importance of CH/π hydrogen bonds in the enzyme co-crystalized with an inhibitor. Notice that a phenyl ring in the biphenylethyl group is aromatic-CH/π-bonded with Phe192 and Phe381 of FAAH. The other phenyl ring in the biphenylethyl group interacts with Val491 and Thr488.

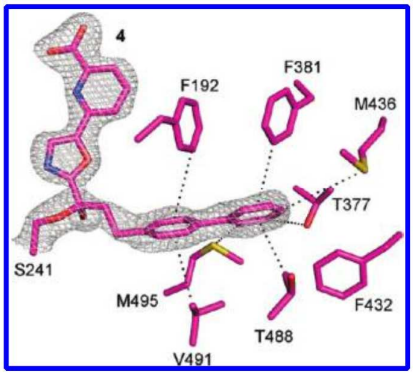


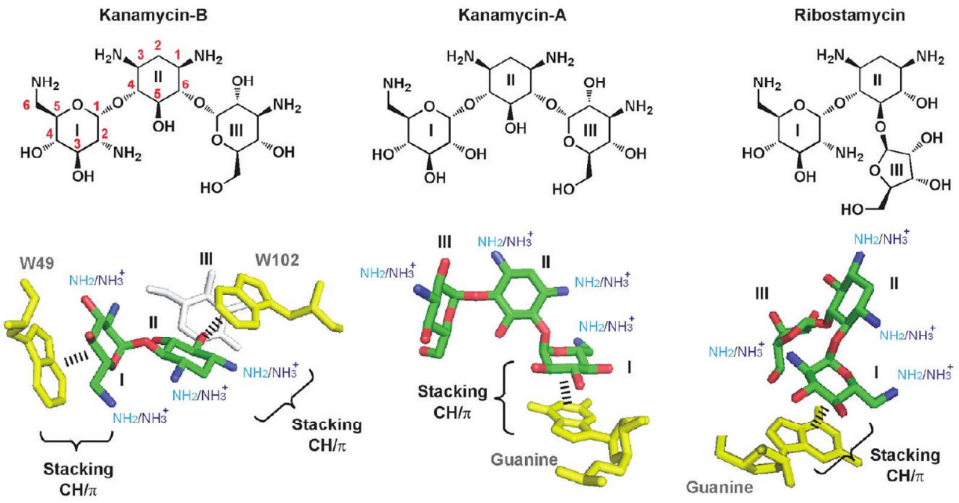
Fig. 35 FAAH active site with bound ligand. Aromatic CHs in F192 and F381, and methyl groups in V491 and T488 are CH/π hydrogen bonded with the biphenylethyl chain mimicking that of arachindonoyl substrates. Fig. 4 of Mileni *et al., J. Med. Chem*., 2010, **53**, 230, with permission.

The molecular mechanisms of inactivation of a serotonin receptor (5-HT6) by antagonists were investigated.[201] Thus, a series of benzimidazole inhibitors were prepared and examined, by site-directed mutagenesis and β-2-adrenoceptor-based homology modeling. Substitution of Trp6.48 or Phe6.52 by Ala impeded the inhibitor to block the 5-HT-induced activation. The aromatic rings of the inhibitors are placed between TMs 3 and 5 or 6 in these complexes. There, a number of CH/π and aromatic CH/π hydrogen bonds were detected.

Aminoglycoside antibiotics are known to be involved in stacking interactions with aromatic rings present in their protein or RNA receptors. Asensio and coworkers pursed this issue by analyzing the conformational properties of synthetic aminoglycoside derivatives equipped with aromatic systems.[202] The experimental data were complemented with an extensive theoretical analysis of simpler models. Their results have shown that the CH/π-hydrogen bonds contribute to the molecular recognition of aminoglycoside antibiotics. As shown in Fig. 36, the crystal structures of the resistance enzyme AAC(6'), the ribosomal A-site and the HIV kissing-loop in complex with aminoglycoside antibiotics suggest that CH/π bonds are involved in the molecular recognition of these drugs. Thus, in the former case, two Trp side-chains (W49 and W102) are in contacts with neamine fragment of the antibiotic. The later two RNA receptors also recognize the antibiotics by stacking of axial CHs of amino sugars to the

aromatic G-base. Recently, they provided an interesting review on this issue.[203] Waters and coworkers reported the importance of this type interaction.[204]

(A)



(B)

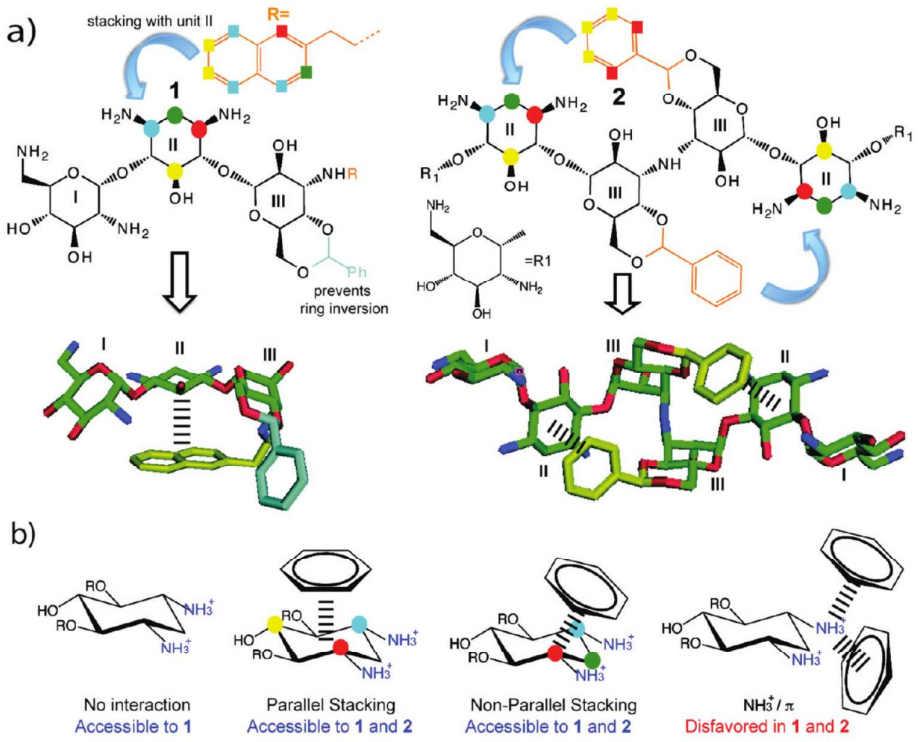


Fig. 36 (A) Top: Kanamycin B, kanamycin A and ribostamycin. Bottom: X-ray data

illustrating CH/π hydrogen bonds present in their respective complexes (from left to right) with the resistance enzyme, the ribosomal A-site RNA and the viral HIV DIS kissing-loop. (B) Top: Aminoglycosides **1** and **2** and corresponding molecular mechanics minimized models. Bottom: The different interacting modes between unit II and the corresponding aromatic ring accessible to **1** and **2**. Figs. 1 and 4 of Vacas *et al. J. Am. Chem. Soc.,* 2010, **132**, 12074, with permission.

A conformational-restriction approach was reported for β-secretase (BACE1) inhibitors.[205] The stability of a ligand/enzyme complex has been found enhanced by using cyclopropane derivatives; the authors applied this finding to seek more potent BACE1 inhibitors. In the structure of BACE1, the side-chain of Tyr71 flipped, making a space for the compound to be accommodated. In the complex, a cyclopropane ligand with the *cis*-(1*S*, 2*R*) configuration was located closer to the flap region of BACE1, compared with that of the ethylene linker compound, making it possible for the cyclopropane ring to interact with Tyr71. A tight conformational restriction by the small ring might have caused the induced-fit; there the CH/π hydrogen bond likely stabilized the structure of the complex.

There are other papers dealing with the effect of CH/π hydrogen bonds in SBDD. The titles and references are given in ESI Table (d).

4.2 Fragment molecular orbital calculation

Kitaura and his group reported on the *ab initio* fragment MO (FMO) method,[206] by which biological macromolecules can be treated with chemical accuracy. In the FMO method, a large molecule or a molecular cluster is divided into M fragment (monomers). *Ab initio* calculations are then carried out, repeatedly, on the fragments under the electrostatic potential from the surrounding (M-1) monomers, until all the monomer densities become self-consistent. Next, the equations for the dimers are solved under the influence of the electrostatic potential from the neighboring (M-2) monomers. Finally, the total energy of the system is calculated, using the total energies of the monomer and the dimer. Many commonly used wave functions and solvent models have been interfaced with FMO. The method has been efficiently parallelized suitable for petascale computing. Reviews[207] and books[208,209]are available. Nakano *et al.* implemented the four-body-corrected fragment

MO method (FMO4)[210] into APINIT-MP BioStation Viewer.[211] This new version provides better accuracy in total energies, in comparison with the reference values by regular FMO calculations.

4.3 Application of the FMO method in structure-based drug design

The FMO method can be used in biochemical applications involving protein/ligand binding and structure-based drug design (SBDD). A review of SBDD appeared, focusing on the application of the FMO method and the CH/π hydrogen bond.[212]

By using the FMO method, Ozawa and coworkers reported that CH/π hydrogen bonds determine the selectivity of SH2 domain to tyrosine peptides[213] and leukocyte-specific protein tyrosine kinase.[214] The calculations were carried out at the Hartree-Fock (HF) and Møller–Plesset (MP2) levels for three SH3 domains and five Pro-rich domains (PRDs) in complex with their ligand peptides. Thus they examined CH/π hydrogen bonds in protein/ligand complexes involving at least one Pro residue, using the FMO method and program CHPI.[215] It has been found that SH3 uses a conserved set of aromatic residues to recognize Pro-rich sequences of specific ligands (Fig. 38. See also Section 2.3.7 and Fig. 17). Their results suggest that CH/π hydrogen bonds are important in the recognition of SH3 by PRDs and plays, by implication, a vital role in the signal transduction system. They also examined CH/π hydrogen bonds in complexes of active and inactive β2 adrenergic receptor, a member of GPCR.[216] Comparison of the active and inactive states suggested that CH/π hydrogen bonds play a key role in ligand recognition and conversion between these two states. Combined use of the FMO method and CHPI analysis is becoming a valuable tool for the study of protein/ligand interactions and is certainly useful in SBDD.

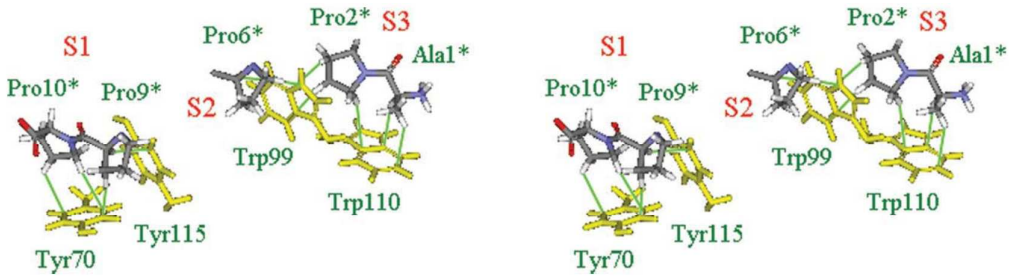


Fig. 37 CH/π hydrogen bonds unveiled in the Pro-recognition domains in a tyrosine kinase (stereo view). Fig. 3 of Ozawa *et al.*, *J. Comput. Chem.,* 2011, **32**, 2774, with permission.

Tsukamoto *et al.* performed a geometry optimization with the FMO scheme, using Trp-cage as an example (for Trp-cage, see Section 2.3.5 and Fig. 16).[217] The geometries of the central part of the peptide were partially optimized at the FMO-MP2 and FMO-HF levels (Fig. 38); the result by the FMO-MP2 calculations showed a more reasonable agreement with the experimental NMR structure reported by Lovas and coworkers.

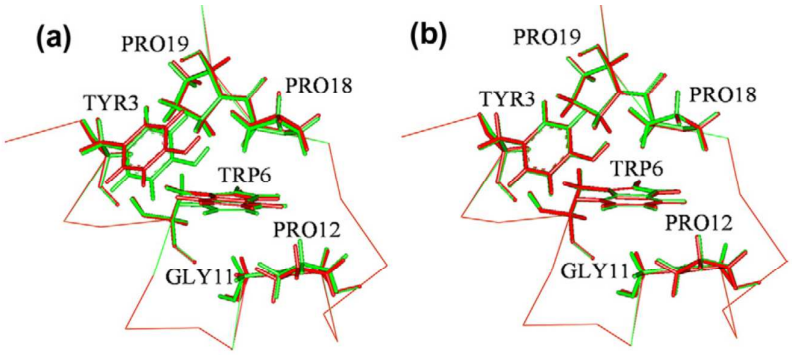
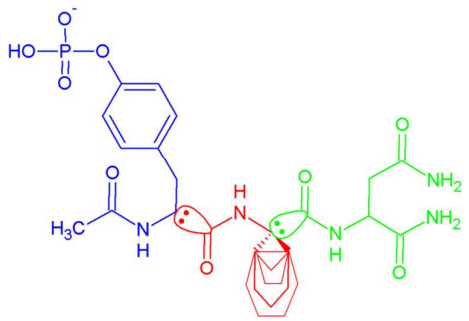


Fig. 38 Geometries of Trp-cage optimized by (a) FMO-HF and (b) FMO-MP2 calculations, respectively, superimposed on the NMR structure. From Tsukamoto *et al. Chem. Phys. Lett.,* 2012, **535**, 157, with permission.

The thermodynamic parameters were determined[218] for complex formation between the *Grb2* SH2 domain and Ac-pTyr-Xaa-Asn derived tripeptides, where the Xaa is a α,α–cycloaliphatic amino acid varying in ring size from three (n=1) to seven (n=5). The binding affinities of those having three- to six-membered rings increased incrementally with the ring size, a finding consistent with an enthalpy-driven effect. Their crystallographic analysis revealed that the only significant differences in structures of the complexes were shown in the number of van der Waals contacts between the domain and the methylene groups in the Xaa residues (Fig. 39).



n = 1-5

Fig. 39 Structure of tripeptide derivatives whose thermodynamic parameters for complex formation with the *Grb2* SH2 domain was investigated.

Watanabe *et al.* investigated the issue by calculations and found that CH/π hydrogen bonds are in fact responsible for the above effect.[219] Thus the energies of stabilization, estimated by the FMO4 method, increased from n=1 (cyclopropane derivative) to n=4 (cyclohexane derivative), a result being consistent with our hypothesis.

Umezawa and Nishio carried out a CSD (Cambridge Structural Database) study to investigate the potentiality of guanidinium group acting as an effective CH acceptor of the CH/π hydrogen bond; it has been demonstrated that this group is a good receptor of CH in crystals of arginine and its derivatives.[220] They also found that the crystal structure of influenza neuraminic acid neuraminidase (NA) in complex with oseltamivir (marketed as tamifulu, a specific inhibitor of NA) is involved in such a kind of CH/π hydrogen bond. Implication of the finding to protein biochemistry and SBDD has been suggested.

5. Summary and prospects

We presented, throughout this treatise, evidence that CH/π hydrogen bonds play an important role in every field of biochemistry. The frequent occurrence of aliphatic (normal CH/π) as well as aromatic CH/π hydrogen bonds (so-called π/π stacking) in biological macromolecules (proteins, nucleic acids, membrane lipids and polysaccharides) suggests a functional role for them in defining stability of 3D structures,

in many molecular recognition events as well as the folding-mechanisms of macromolecules. Clearly, a more generalized hydrogen bond definition is necessary. We conclude that the classical concept of hydrogen bonds needs to be revised and extended.

The topics dealt with in this *Perspective* are limited to those noticed by the authors; it is certain that many papers escaped their attention. More up-to-date information is available from the literature list in one of the authors' website (http://www.tim.hi-ho.ne.jp/dionisio). We hope that the present treatise will show the importance of CH/π hydrogen bonds and stimulate interest in the interactions of biological macromolecules, one of the most fascinating fields in chemistry.

Acknowledgments

We thank following scientists for reading the relevant parts of the manuscript: Helmut Sigel (Basel), Gergely Tóth (South San Francisco), Sandor Lovas (Omaha), Jyoti Chattopadhyaya (Uppsala), Dale L. Boger (La Jolla), Kazuhiro Irie (Kyoto), Kazuo Umezawa (Nagoya), Norio Matsushima (Sapporo), Juan L. Asensio (Barcelona), Jianmin Gao (Boston), G. Narahari Sastry (Hyderabad), Xiche Hu (Toledo), George Harauz (Guelph), Fenenc Zsila (Budapest), Tomonaga Ozawa (Azumino), Kazuhiko Matsuda (Nara), Isao Momose (Numazu), Eugenia Polverini (Parma), Shigetoshi Aono (Okazaki), Norio Teramae (Sendai), Isao Nakanishi (Higashi Osaka), Jiro Kondo (Tokyo) and Satoshi Shuto (Sapporo). MN thanks Jun Suzuki for his help in preparing the manuscript.

It is our pleasure to dedicate this review to the late Professor Max F. Perutz, the founder of structural biology.

References

[1] J. N. Israelachvili, *Intermolecular Forces and Surface Forces*, 2nd edition, Academic Press Ltd., London, 1992, Table 6.3.
[2] L. Pauling, *The Nature of the Chemical Bond*, 3rd edition, Cornell Univ. Press, Ithaca, New York, 1960, Chapter 12.
[3] G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford Univ. Press, Oxford, 1997.
[4] S. Scheiner, *Hydrogen Bonding – A Theoretical Perspective*, Oxford Univ. Press, Oxford, 1997.

[5] G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, W. H. Freeman, San Francisco, 1960.
[6] S. J. Grabowski, W. A. Sokalski, E. Dyguda and J. Leszczynski, *J. Phys. Chem. B,* 2006, **110**, 6444-6446.
[7] There is an updated IUPAC definition of the hydrogen bond (IUPAC Technical Report): E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.,* 2011, **83**, 1619-1636; 1637-1641. See also Desiraju, G. R. *Angew. Chem. Int. Ed.,* 2011, **50**, 52-59.
[8] R. G. Pearson, *Science,* 1966, **151**, 172-177.
[9] M. Nishio and M. Hirota, *Tetrahedron,* 1989, **45**, 7201-7245.
[10] G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford Univ. Press, Oxford, 1999.
[11] M. Nishio, *Phys. Chem. Chem. Phys.,* 2011, **13**, 13873-13900.
[12] T. Steiner and G. Koellner, *J. Mol. Biol.,* 2001, **305**, 535-557.
[13] M. Nishio, M. Hirota and Y. Umezawa, *The CH/π interaction. Evidence, Nature, and Consequences*, Wiley-VCH, New York, 1998.
[14] M. Nishio, (a) *Kagaku no Ryoiki,* 1977, **31**, 998-1006; (b) 1979, **33**, 422-432; (c) 1983, **37**, 243-251.
[15] For evidence, nature, characteristics and implications of the CH/π hydrogen bond, see Section 1.1 of Ref. 11.
[16] O. Takahashi, Y. Kohno and M. Nishio, *Chem. Rev.,* 2010, **110**, 6049-6076.
[17] Ref. 11, Section 3.1.
[18] M. Nishio and Y. Umezawa, *Top. Stereochem.,* 2006, **25**, 255-302.
[19] M. Nishio, *CrystEngComm,* 2004, **6**, 130-156.
[20] M. Nishio, Y. Umezawa, K. Honda, S. Tsuboyama and H. Suezawa, *CrystEngComm,* 2009, **11**, 1757-1788.
[21] M. Nishio, Y. Umezawa, H. Suezawa and S. Tsuboyama, *The CH/π hydrogen bond: Implication in crystal engineering*, Chapter 1 in *Frontiers in Crystal Engineering - Pi Interactions, III*, Eds. E. Tiekink and J. Zukerman-Schpector, Wiley, Chichester, 2012.
[22] M. Nishio, M. Hirota and Y. Takeuchi, *Tetrahedron,* 1995, **51**, 8665-8701.
[23] M. Nishio, *Tetrahedron,* 2005, **61**, 6923-6950.
[24] Ref. 11, Section 3.3.
[25] Ref. 11, Section 3.4.
[26] See Table 2 of Ref. 11.
[27] P. Hobza, V. Spirko, H. L. Selze and E. W. Schlag, *J. Phys. Chem. A,* 1998, **102**, 2501-2504.
[28] P. Hobza and Z. Havlas, *Chem. Rev.,* 2000, **100**, 4253-4264.
[29] A. J. Barnes, *J. Molec. Struct.,* 2004, **704**, 3-9.
[30] N. Nakagawa, *Nippon Kagaku Zasshi,* 1961, **82**, 141-147.

[31] (a) C. A. Hunter and J. K. M. Sanders, *J. Am. Chem. Soc.,* 1990, **112**, 5525-5534; (b) C. A. Hunter, *Chem. Soc. Rev.,* 1994, **23**, 101-109.
[32] J. Novotny and E. Haber, *Proc. Natl. Acad. Sci. U.S.A.,* 1985, **82**, 4592-4596.
[33] G. B. McGaughey, M. Gagné and A. K. Rappé, *J. Biol. Chem.*, 1998, **273**, 15458-15463.
[34] J. R. Cox, *J. Chem. Ed.,* 2000, **77**, 1424-1428.
[35] (a) E. Gazit, *Bioinformatics Discovery Note,* 2002, **18**, 880-883; (b) E. Gazit, *FASEB J.,* 2002, **16**, 77-83.
[36] M. Chourasia, G. M. Sastry and G. N. Sastry, *Int. J. Biol. Macromol.,* 2011, **48**, 540-552. For comparison of T-shape *vs.* stacking interactions in coordination compounds, see: C. Janiak, *J. Chem. Soc. Dalton Trans.,* 2000, 3885-3896.
[37] A. S. Mahadevi, A. P. Rahalkar, S. R. Gadre and G. N. Sastry, *J. Chem. Phys.,* 2010, **133**, 164308.
[38] Y. Umezawa and M. Nishio, *Bioorg. Med. Chem.,* 1998, **6**, 493-504. See also Section 11.2 of Ref. 13.
[39] M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel and R. Hilgenfeld, *J. Mol. Biol.*, 2001, **307**, 357-377.
[40] E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.,* 1984, **44**, 97-179.
[41] M. J. Plevin, D. L. Bryce and J. Boisbouvier, *Nat. Chem.,* 2010, **2**, 466-471.
[42] C. J. Pace, D. Kim and J. Gao, *Chem.- Eur. J.*, 2012, **18**, 5832-5836.
[43] H. K. Ganguly, B. Majumder, S. Chattopadhyay, P. Chakrabarti and G. Basu, *J. Am. Chem. Soc.,* 2012, **134**, 4661-4669.
[44] M. Nishio, *29th Symposium on Protein Structures, Osaka*, 1978, Abstract, 161-164.
[45] (a) S. K. Burley and G. A. Petsko, *Science*, 1985, **229**, 23-28; (b) S. K. Burley and G. A. Petsko, *Adv. Prot. Chem*., 1988, **39**, 125-189.
[46] (a) J. Singh and J. M. Thornton, *FEBS Lett.*, 1985, **191**, 1-6; (b) J. Singh and J. M. Thornton, *J. Mol. Biol.*, 1990, **211**, 595-615.
[47] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res*., 2000, **28**, 235-242.
[48] P. Chakrabarti and U. Samanta, *J. Mol. Biol*., 1995, **251**, 9-14.
[49] U. Samanta, P. Chakrabarti and J. Chandrasekhar, *J. Phys. Chem. A*, 1998, **102**, 8964-8969.

[50] Y. Umezawa and M. Nishio, *Bioorg. Med. Chem*., 1998, **6**, 2507-2515.

[51] A. Jabs, M. S. Weiss and R. Hilgenfeld, *J. Mol. Biol.*, 1999, **286**, 291-304.

[52] D. Pal and P. Chakrabarti, *J. Mol. Biol.*, 1999, **294**, 271-288.

[53] M. R. Parsons, M. A. Convey, C. M. Wilmot, K. D. Yadav, V. Blakeley, A. S. Corner, S. E. Phillips, M. J. McPherson and P. F. Knowles, *Structure*, 1995, **3**, 1171-1184.

[54] U. Samanta, D. Pal and P. Chakrabarti, *Proteins*, 2000, **38**, 288-300.

[55] A. Thomas, R. Meurisse, B. Chartloteaux and R. Brasseur, *Proteins*, 2002, **48**, 628-634.

[56] R. Bhattacharyya, U. Samanta and P. Chakrabarti, *Prot. Eng.*, 2002, **15**, 91-100.

[57] G. Tóth, R. F. Murphy and S. Lovas, *Protein Eng.,* 2001, **14**, 543-547.

[58] Ref. 13, Section 11.4.

[59] M. Nishio, Y. Umezawa and M. Hirota, *Yuki Gosei Kagaku Kyoukai-shi*, 1997, **55**, 17-28, Fig. 23.

[60] J. C. Ma and D. A. Dougherty, *Chem. Rev.,* 1997, **97**, 1303-1324.

[61] (a) F. B. Hasan, S. G. Cohen and F. B. Hasan, *J. Biol. Chem.,* 1980, **255**, 3898-3904; (b) F. B. Hasan, J. L. Elkind, S. G. Cohen and J. B. Cohen, *J. Biol. Chem.*, 1981, **256**, 7781-7785; (c) S. G. Cohen, D. L. Lieberman, F. B. Hasan and J. B. Cohen, *J. Biol. Chem.,* 1982, **257**, 14087-14092.

[62] K. Kobayashi, Y. Asakawa and Y. Aoyama, *Supramol. Chem.,* 1993, **2**, 133-135. See Table 9 in Ref. 11.

[63] C. D. Tatko and M. L. Waters, *J. Am. Chem. Soc.,* 2004, **126**, 2028-2034. See Fig. 43 in Ref. 11.

[64] R. M. Huges and M. L. Waters, *J. Am. Chem. Soc.,* 2006, **128**, 13586-13591.

[65] See Fig. 44 in Ref. 11.

[66] M. Harigai, M. Kataoka and Y. Imamoto, *J. Am. Chem. Soc.*, 2006, **128**, 10646-10647.

[67] Ref. 13, Fig. 11.20.

[68] Y. Umezawa and M. Nishio, *Bioorg. Med. Chem*., 1998, **6**, 2507-2515.

[69] P. Chakrabarti and P. Bhattacharyya, *Progr. Biophys. Mol. Biol.*, 2007, **95**, 83-137.

[70] (a) Y. Inoue, N. Nakamura and T. Inagami, *J. Hypertension,* 1997, **15**, 703-714; (b) Y. N. Imai, Y. Inoue and Y. Yamamoto, *J. Med. Chem.*, 2007, **50**, 1189-1196.

[71] B. Dasgupta, P. Chakrabarti and G. Basu, *FEBS Lett*., 2007, **581**, 4529-4532.

[72] E. A. Meyer, R. K. Castellano and F. Diederich, *Angew. Chem. Int. Ed.*, 2003, **42**, 1210-1250.

[73] S. Chakkaravarthi, M. M. Babu, M. M. Gromiha, G. Jayaraman and R. Sethumadhavan, *Proteins*, 2006, **65**, 75-86.

[74] A. Terrón, J. J. Fiol, A. García-Raso, M. Barceló-Oliver and V. Moreno, *Coord. Chem.*

Rev., 2007, **251**, 1973-1986.
[75] M. Cohen, D. Reichmann, H. Neuvirth and G. Schreiber, *Proteins*, 2008, **72**, 741-753.
[76] C. H. Suresh, N. Mohan, K. P. Vijayalakshmi, R. George and J. M. Martin, *J. Comput. Chem.*, 2009, **30**, 1392-1404.
[77] C. Bissantz, B. Kuhn and M. Stahl, *J. Med. Chem*., 2010, **53**, 5061-5084.
[78] R. P. Saha, R. P. Bahadur and P. Chakrabarti, *Proteome Res.,* 2005, **4**, 1600-1609.
[79] P. Chakrabarti and D. Pal, *Prog. Biophys. Mol. Biol.,* 2001, **76**, 1-102.
[80] E. G. Hutchinson and J. M. Thornton, *Protein Sci.,* 1994, 3, 2207-2216.
[81] R. Bhattacharyya and P. Chakrabarti, *J. Mol. Biol.,* 2003, 331, 925-940.
[82] J. S. Richardson, *Adv. Protein Chem.,* 1981, **34**, 167-339.
[83] G. D. Rose, L. M. Gierasch and J. A. Smith, *Adv. Protein Chem.,* 1985, **37**, 1-109.
[84] D. E. Stewart, A. Sarkar and J. E. Wampler, *J. Mol. Biol.,* 1990, **214**, 253-260.
[85] M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.,* 1991, **218**, 397-412.
[86] U. Reimer, G. Scherer, M. Drewello, S. Kruber, M. Schutkowski and G. Fischer, *J. Mol. Biol.,* 1998, **279**, 449-460.
[87] W.-J. Wu and D. P. Raleigh, *Biopolymers,* 1998, **45**, 381-394.
[88] H. Y. Meng, K. M. Thomas, A. E. Lee and N. J. Zondlo, *Biopolymers,* 2006, **84**, 192-204.
[89] N. J. Zondlo, *Acc. Chem. Res.,* 2012, **46**, 1039-1049.
[90] J. Yao, V. A. Feher, B. F. Espejo, M. T. Reymond, P. E. Wright and H. J. Dyson, *J. Mol. Biol.,* 1994, **243**, 736-753.
[91] M. S. Weiss, H. J. Metzner and R. Hilgenfeld, *FEBS Lett.,* 1998, **423**, 291-296.
[92] K. M. Armstrong, R. Fairman and R. L. Baldwin, *J. Mol. Biol.,* 1993, **230**, 284-291.
[93] J. Fernández-Recio, A. Vázquez, C. Civera, P. Sevilla and J. Sancho, *J. Mol. Biol.,* 1997, **267**, 184-197.
[94] R. Loewenthal, J. Sancho and A. R. Fersht, *J. Mol. Biol.,* 1992, **224**, 759-770.
[95] A. C. Rosenzweig, H. Brandstetter, D. A. Whittington, P. Nordlund, S. J. Lippard and C. A. Frederick, *Proteins,* 1997, 29, 141-152.
[96] A. R. Khan, J. C. Parrish, M. E. Fraser, W. W. Smith, P. A. Bartlett and M. N. James, *Biochemistry,* 1998, 37*,* 16839-16845.
[97] J. M. Guss, H. D. Bartunik and H. C. Freeman, *Acta Cryst.,* 1992, B48, 790-810.
[98] W. Watt, A. Tulinsky, R. P. Swenson and K. D. Watenpaugh, *J. Mol. Biol.,* 1991, 218, 195-208.
[99] G. Oliva, M. R. Fontes, R. C. Garratt, M. M. Altamirano, M. Calcagno and E. Horjales,

Structure, 1995, 3, 1323-1332.

[100] M. A. Wouters and P. M. G. Curmi, *Proteins,* 1995, **22**, 119-131.

[101] J. S. Merkel and L. Regan, *Fold. Des.,* 1998, **3**, 449-455.

[102] A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry*, 2000, **39**, 1496014967.

[103] J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat. Struct. Biol.,* 2002, **9**, 425-430.

[104] D. Naduthambi and N. J. Zondlo, *J. Am. Chem. Soc.,* 2006, **128**, 12430-12431.

[105] M. P. D. Hatfield, R. F. Murphy and S. Lovas, *J. Phys. Chem. B,* 2011, **115**, 4971-4981.

[106] M. P. D. Hatfield, N. Y. Palermo, J. Csontos, R. F. Murphy and S. Lovas, *J. Phys. Chem. B,* 2008, **112**, 3503-3508.

[107] M. P. D. Hatfield, R. F. Murphy and S. Lovas, *J. Phys. Chem. B,* 2010, **114**, 3028-3037.

[108] P. Carnevali, G. Tóth, G. Toubassi and S. N. Meshkat, *J. Am. Chem. Soc.,* 2003, **125**, 14244-14245.

[109] Y. K. Kang and B. J. Byun, *Biopolymers,* 2012, **97**, 778-788.

[110] M. Meng, D. Vardar, Y. Wang, H. C. Guo, J. F. Head and C. J. McKnight, *Biochemistry,* 2005, 44, 11963-11973.

[111] S. Unterreitmeier, A. Fuchs, T. Schäffler, R. G. Heym, D. Frishman and D. Langosch, *J. Mol. Biol.* 2007, **374**, 705-718.

[112] P. Chakrabarti and J. Janin, *Proteins,* 2002, **47**, 334-343.

[113] A. A. Bogan and K. S. Thorn, *J. Mol. Biol.,* 1998, **280**, 1-9.

[114] R. P. Saha, R. Bhattacharyya and P. Chakrabarti, *Proteins*, 2007, **67**, 84-97.

[115] A. Zarrinpar, R. P. Bhattacharyya and W. A. Lim, *Sci. STKE,* 2003, **179**, re8.

[116] M. T. Pisabarro, L. Serrano and M. Wilmanns, *J. Mol. Biol.,* 1998, **281**, 513-521.

[117] H. Dvir, M. Harel, S. Bon, W. Q. Liu, M. Vidal, C. Garbay, J. L. Sussman, J. Massoulie and I. Silman, *EMBO J.,* 2004, **23**, 4394-4405.

[118] M. Maresca, A. Derghal, C. Carravagna, S. Dudin and J. Fantini, *Phys. Chem. Chem. Phys.,* 2008, **10**, 2792-2800.

[119] F. A. Quiocho and N. K. Vyas, *Nature,* 1984, **310**, 381-386.

[120] N. K. Vyas, M. N. Vyas and F. A. Quiocho, *Science,* 1988, 242*,* 1290-1295.

[121] R. Ravishankar, K. Suguna, A. Surolia and M. Vijayan, *Acta Crystallogr., Sect. D,* 1999, **55**, 1375-1377.

[122] M. Muraki, K. Harata, N. Sugita and K. Sato, *Acta Crystallogr., Sect. D,* 1998, **54**, 834-843.

[123] (a) M. Muraki and K. Harata, *J. Molec. Recogn*., 2003, **16**, 72-82; (b) M. Muraki, K. Harata, N. Sugita and K. Sato, *Biochemistry,* 2000, **39**, 292-299.

[124] M. Muraki, H. Morii and K. Harata, *Protein Eng.,* 2000, **13**, 385-389.

[125] M. I. Chávez, C. Andreu, P. Vidal, N. Aboitiz, F. Freire, P. Groves, J. L. Asensio, G. Asensio, M. Muraki, F. J. Cañada and J. Jiménez-Barbero, *Chem.- Eur. J.,* 2005, **11**, 7060-7074.

[126] K. Harata and M. Muraki, *J. Mol. Biol.,* 2000, **297**, 673-681.

[127] K. Harata and M. Muraki, *Acta Crystallogr., Sect. D,* 1997, **53**, 650-657.

[128] M. Muraki, *Prot. Peptide Lett.,* 2002, **9**, 195-209.

[129] (a) V. Spiwok, P. Lipovová, T. Skálová, E. Buchtelová, J. Dohnálek, J. Hasek and B. Králová, *Carbohydr. Res.,* 2004, **339**, 2275-2280; (b) V. Spiwok, P. Lipovová, T. Skálová, E. Buchtelová, J. Dohnálek, J. Hasek and B. Králová, *J. Comput-Aided Drug Des.,* 2006, **19**, 887-901.

[130] (a) M. S. Sujatha and P. V. Balaji, *Proteins,* 2004, **55**, 44-65; (b) M. S. Sujatha, Y. Sasidhar and P. V. Balaji, *Protein Sci.,* 2004, **13**, 2502-2514 ; (c) M. S. Sujatha and P. V. Balaji, *Biochemistry,* 2005, **44**, 8554-8562 ; (d) M. S. Sujatha, Y. Sasidhar and P. V. Balaji, *J. Molec. Struct: THEOCHEM,* 2007, **814**, 11-24.

[131] J. Fantini, *Cell. Mol. Life Sci.,* 2003, **60**, 1027-1032.

[132] J. Fantini and N. Yahi, *Expert Rev. Mol. Med.,* 2011, **12**, e27.

[133] J. Fantini, N. Garmy and N. Yahi, *Biochemistry,* 2006, **45**, 10957-10962.

[134] H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, C. D. Armstrong, J. L. Brunton and R. J. Read, *Biochemistry,* 1988, **37**, 1777-1788.

[135] D. Hammache, N. Yahi, G. Piéroni, F. Ariasi, C. Tamalet and J. Fantini, *Biochem. Biophys. Res. Commun.,* 1998, **246**, 117-122.

[136] J. Fantini and N. Yahi, *J. Mol. Biol.,* 2011, **408**, 654-669.

[137] R. Mahfoud, N. Garmy, M. Maresca, N. Yahi, A. Puigserver and J. Fantini, *J. Biol. Chem.,* 2002, **277**, 11292-11296.

[138] K. Matsuzaki, K. Kato and K. Yanagisawa, *Biochim. Biophys. Acta*, 2010, **1801**, 868.

[139] N. Yahi, J. M. Sabatier, S. Baghdiguian, F. Gonzalez-Scarano and J. Fantini, *J. Virol.,* 1995, **69**, 320-325.

[140] T. Matsubara, D. Ishikawa, T. Taki, Y. Okahata and T. Sato, *FEBS Lett.,* 1999, **456**, 253-256.

[141] N. Yahi, A. Aulas and J. Fantini, *PLoS One,* 2010, **5**, e9079.

[142] J. Fantini and F. J. Barrantes, *Biochim. Biophys. Acta*, 2009, **1788**, 2345-2361.

[143] M. A. Hanson, V. Cherezov, M. T. Griffith, C. B. Roth, V. P. Jaakola, E. Y. Chien, J. Velasquez, P. Kuhn and R. C. Stevens, *Structure,* 2008, **16**, 897-905.

[144] L. Adamian, H. Naveed and J. Liang, *Biochim. Biophys. Acta,* 2011, **1808**, 1092-1102.

[145] C. J. Baier, J. Fantini and F. J. Barrantes, *Sci. Reports,* 2011, **1**, 69.

[146] H. Li and V. Papadopoulos, *Endocrinology,* 1998, **139**, 4991-4997.

[147] (a) R. M. Epand, *Prog. Lipid Res.,* 2006, **45**, 279-294; (b) R. F. Epand, A. Thomas, R. Brasseur, S. A. Vishwanathan, E. Hunter and R. M. Epand, *Biochemistry,* 2006, **45**, 6105-6114.

[148] R. M. Epand, A. Thomas, R. Brasseur and R. F. Epand, *Subcell. Biochem.*, 2010, **51**, 253-278.

[149] J. Bock and E. Gulbins, *FEBS Lett.,* 2003, **534**, 169-174.

[150] H. J. Kwon, L. Abi-Mosleh, M. L. Wang, J. Deisenhofer, J. L. Goldstein, M. S. Brown and R. E. Infante, *Cell*, 2009, **137**, 1213-1224.

[151] A. J. Ciani, B. C. Gupta and I. P. Patra, *Solid State Commun.,* 2008, **147**, 146-150.

[152] J. Fantini and F. J. Barrantes, *Front Physiol.,* 2013, **4**, 31.

[153] B. E. Fisher and H. Sigel, *J. Am. Chem. Soc.,* 1980, **102**, 2998.

[154] G. Liang, R. Tribolet and H. Sigel, *Inorg. Chem.,* 1988, **27**, 2877-2887.

[155] F. Zsila, Z. Bikadi and M. Simonyi, *Biochem. Pharmacol.,* 2002, **64**, 1651-1660.

[156] V. Klusák, Z. Havlas, L. Rulísek, J. Vondrásek and A. Svatos, *Chem. Biol*., 2003, **10**, 331-340.

[157] S. Mohanty, S. Zubkov and A. M. Gronenborn, *J. Mol. Biol.,* 2004, **337**, 443-451.
[158] Y. Wang, L. Mao and X. Hu, *Biophys. J.,* 2004, **86**, 3097-3111.
[159] C. A. Sacksteder, S. L. Bender and B. A. Barry, *J. Am. Chem. Soc.,* 2005, **127**, 7879-7890.
[160] C. D. Blundell, A. Almond, D. J. Mahoney, P. L. DeAngelis, I. D. Campbell and A. J. Day, *J. Biol. Chem.,* 2005, **280**, 18189-18210.
[161] G. Tóth and A. Borics, *Biochemistry,* 2006, **45**, 6606-6614.
[162] G. Tóth and A. Borics, *J. Mol. Graph. Model.,* 2006, **24**, 465-474.
[163] A. Anbarasu, S. Anand, M. M. Babu and R. Sethumadhavan, *Int. J. Biol. Macromol.,* 2007, **41**, 251-259.
[164] S. Anand, A. Anbarasu and R. Sethumadhavan, *In Silico Biology,* 2008, **8**, 261-273.
[165] A. Okumura, M. Sano, T. Suzuki, H. Tanaka, R. Nagao, K. Nakazato, M. Iwai, H. Adachi, J.-R. Shen and I. Enami, *FEBS Lett.,* 2007, **581**, 5255-5258.
[166] P. Kursula, L. Kursula, M. Massimi, Y. H. Song, J. Downer, W. A. Stanley, W. Witke and M. Wilmanns, *J. Mol. Biol.,* 2008, **375**, 270-290.
[167] E. Polverini, G. Rangaraj, D. S. Libich, J. M. Boggsand G. Harauz, *Biochemistry,* 2008, **47**, 267-282.
[168] A. F. Neuwald, *BMC Structural Biology,* 2009, **9**, 11.
[169] K. Matsuda, S. Kanaoka, M. Akamatsu and D. B. Sattelle, *Molec. Pharmcol.,* 2009, **76**, 1-10.
[170] A. Stavrakoudis, I. G. Tsoulos, Z. O. Shenkarev and T. V. Ovchinnikova, *Biopolymers,* 2009, **92**, 143-155.
[171] C. H. Suresh, N. Mohan, K. P. Vijayalakshmi, R. George and J. M. Mathew, *J. Comp. Chem.,* 2009, **30**, 1392-1404.
[172] J. P. Yesudas, F. B. Sayyed and C. H. Suresh, *J. Molec. Model.,* 2011, **17**, 401-413.
[173] V. R. Prasad, B. N. Tripathi and R. Sethumadhavan, *Int. J. Bioinform. Res. Appl.,* 2010, ***6***, 241-259.
[174] I. A. Tayubi and R. Sethumadhavan, *Int. J. Pharm. Pharmaceut. Sci.,* 2011, **3***(Suppl. 2)*, 212-218.
[175] T. Ema, Y. Nakano, D. Yoshida, S. Kamata and T. Sakai, *Org. Biomol. Chem.,* 2012, **10**, 6299-6308.

[176] G. Tóth, S. G. Bowers, A. P. Truong and G. Probst, *Curr. Pharmac. Des.,* 2007, **13**, 3476-3493: *The Role and Significance of Unconventional Hydrogen Bonds in Small Molecule Recognition by Biological Receptors of Pharmaceutical Relevance.*

[177] (a) K. Irie, Y. Nakagawa and H. Ohigashi, *Chem. Rec.*, 2005, **5**, 185-195: *Toward the development of new medicinal leads with selectivity for protein kinase C isozymes.* (b) K. Irie, *Farumashia*, 2006, **42**, 427-430: *Molecular design of agents based on the CH/π interaction*.

[178] (a) H. Sigel and C. F. Naumann, *J. Am. Chem. Soc.,* 1976, *98*, 730; (b) P. R. Mitchell, B. Prijs and H. Sigel, *Helv. Chim. Acta,* 1979, *62*, 1723.

[179] P. R. Mitchell and H. Sigel, *J. Am. Chem. Soc.,* 1978, **100**, 1564-1570.

[180] H. Sigel, *Angew. Chem. Int. Ed.,* 1975, **14**, 394-402.

[181] H. Sigel, *Pure Appl. Chem.,* 2004, **76**, 375-388.

[182] H. Sigel and R. Griesser, *Chem. Soc. Rev.,* 2005, **34**, 875-900.

[183] R. B. Gómez-Coca, C. A. Blindauer, A. Sigel, B. B. Operschall, A. Holý and H. Sigel, *Chemistry & Biodiversity, Special Issue: Bioinorganic Chemistry (ICBIC XV),* 2012, **9**, 2008-2034.

[184] H. Sigel, *Chem. Soc. Rev.,* 2004, **33**, 191-200.

[185] (a) M. A. Viswamitra, O. Kennard, P. G. Jones, G. M. Sheldrick, S. Salisbury, L. Falvello and Z. Shakked, *Nature,* 1978, **273**, 687-690; (b) M. A. Viswamitra, Z. Shakked, P. G. Jones, G. M. Sheldrick, S. A. Salisbury and O. Kennard, *Biopolymers,* 1982, **21**, 513-533.

[186] A. Klug, A. Jack, M. A. Viswamitra, O. Kennard, Z. Shakked and T. A. Steitz, *J. Mol. Biol.,* 1979, **131**, 669-680.

[187] Y. Umezawa and M. Nishio*, Nucleic Acids Res*., 2002, **30**, 2183-2192.

[188] Y. Umezawa and M. Nishio*, Bioorg. Med. Chem*., 2000, **8**, 2463-2650.

[189] S. M. Butterfield, M. M. Sweeney and M. L. Waters, *J. Org. Chem.,* 2005, **70**, 1105-1114.

[190] D. D. Boehr, A. R. Farley, G. D. Wright and J. R. Cox, *Chem. Biol.,* 2002, **9**, 1209-1217.

[191] N. Yahi, J. Fantini, M. Henry, C. Tourrès and C. Tamalet, *J. Biomed. Sci.,* 2005, **12**, 701-710.

[192] J. C. Morales, J. J. Reina, I. Diaz, A. Avino, P. M. Nieto and R. Eritja, *Chem.- Eur. J.,* 2008, **14**, 7828-7835.

[193] R. Lucas, I. Gomez-Pinto, A. Avinos, J. J. Reina, R. Eritja, C. Gonzarez and J. C. Morales, *J. Am. Chem. Soc*., 2011, **133**, 1909-1916.

[194] A. Gil, V. Branchadell, J. Bertran and A. Oliva, *J. Phys. Chem. B,* 2007, **111**, 9372-9379.

[195] H. Sawai, M. Yamanaka, H. Sugimoto, Y. Shiro and S. Aono, *J. Biol. Chem.,* 2012, **287**, 30755-30768.

[196] J. Kondo, M. Koganei and T. Kasahara, *ACS Med. Chem. Lett.*, 2012, **3**, 741-744.

[197] S. D. McAllister, G. Rizvi, S. Anavi-Goffer, D. P. Hurst, J. Barnett-Norris, D. L. Lynch, P. H. Reggio and M. E. Abood, *J. Med. Chem.,* 2003, **46**, 5139-5152.

[198] Y. Umezawa and M. Nishio, *Biopolymers,* 2005*,* **79**, 248-258.

[199] (a) Y. Nakagawa, K. Irie, R. C. Yanagita, H. Ohigashi and K. Tsuda, *J. Am. Chem. Soc.*, 2005, **127**, 5746-5747; (b) R. C. Yanagita, Y. Nakagawa, N. Yamanaka, K. Kashiwagi and K. Irie, *J. Med. Chem.,* 2008, **51**, 46-56; (c) T. Sugimoto, K. Itagaki and K. Irie, *Bioorg. Med. Chem.,* 2008, **16**, 650-657.

[200] M. Mileni, J. Garfunkle, C. Ezzili, F. S. Kimball, B. F. Cravatt, R. Stevens and D. B. Boger, *J. Med. Chem*., 2010, **53**, 230-240.

[201] T. de la Fuente, M. Martín-Fontecha, J. Sallander, B. Benhamú, M. Campillo, R. A. Medina, L. P. Pellissier, S. Claeysen, A. Dumuis, L. Pardo and M. L. López-Rodríguez, *J. Med. Chem*., 2010, **53**, 1357-1369.

[202] T. Vacas, F. Corzana, G. Jimenez-Oses, C. Gonzalez, A. M. Gomez, A. Bastida, J. Revuelta and J. L. Asensio, *J. Am. Chem. Soc.,* 2010, **132**, 12074-12690.

[203] J. L. Asensio, A. Arda, F. J. Canada and A. S. Jimenez-Barbero, *Acc. Chem. Res.*, 2013, **46**, 946–954.

[204] (a) S. E. Kiehna, Z. R. Laughrey and M. L. Waters, *Chem. Commun.*, 2007, 4026-4028; (b) Z. R. Laughrey, S. E. Kiehna and M. L. Waters, *J. Am. Chem. Soc.*, 2008, **130**, 14625-14633.

[205] S. Yonezawa, T. Yamamoto, H. Yamakawa, C. Muto, M. Hosono, K. Hattori, K. Higashino, T. Yutsudo, H. Iwamoto, Y. Kondo, M. Sakagami, H. Togame, Y. Tanaka, T.

Nakano, H. Takemoto, M. Arisawa and S. Shuto, *J. Med. Chem.,* 2012, **55**, 8838-8858.

[206] (a) K. Kitaura, I. Ikeo, T. Asada, T.; Nakano and M. Uebayashi, *Chem. Phys. Lett.,* 1999, **313**, 701-706; (b) T. Nakano, T. Kakinuma, T. Sato, Y. Akiyama, M. Uebayashi and K. Kitaura, *Chem. Phys. Lett.,* 2000, **318**, 684.

[207] (a) D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A,* 2007, **111**, 6904-6914; (b) D. G. Fedorov, T. Nagata and K. Kitaura, *Phys. Chem. Chem. Phys*., 2012, **14**, 7562-7577.

[208] *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, Eds. E. B. Starikov, J. P. Lewis and S. Tanaka, Elsevier: New Canaan, 2006.

[209] *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*. Eds. D. G. Fedorov and K. Kitaura, CRC Press: New York, 2009.

[210] T. Nakano, Y. Mochizuki, K. Yamashita, C. Watanabe, K. Fukuzawa, K. Segawa, Y. Okiyama, T. Tsukamoto and S. Tanaka, *Chem. Phys. Lett.,* 2012, **523**, 128-133.

[211] MIZUHO/ABINIT-MP http://www.mizuho-ir.co.jp/solution/research/life/macromolecule/biostation/index.html

[212] T. Ozawa, K. Okazaki and M. Nishio, *FMO as a Tool for Structure-Based Drug Design*: Chapter 10 in *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*. Eds. D. G. Fedorov and K. Kitaura, CRC Press: New York, 2009.

[213] T. Ozawa and K. Okazaki, *J. Comput. Chem.,* 2008, **29**, 2656-2666.

[214] T. Ozawa, E. Tsuji, M. Ozawa, C. Handa, H. Mukaiyama, T. Nishimura, S. Kobayashi and K. Okazaki, *Bioorg. Med. Chem.,* 2008, **16**, 10311-10318.

[215] T. Ozawa, K. Okazaki and K. Kitaura, *J. Comput. Chem.,* 2011, **32**, 2774-2782.

[216] T. Ozawa, K. Okazaki and K. Kitaura, *Bioorg. Med. Chem.,* 2011, **19**, 5231-5237.

[217] T. Tsukamoto, Y. Mochizuki, N. Watanabe, K. Fukuzawa and T. Nakano, *Chem. Phys. Lett.,* 2012, **535**, 157-162.

[218] J. M. Myslinski, J. E. DeLorbe, J. H. Clements and S. F. Martin, *J. Am. Chem. Soc.,* 2011, **133**, 18518-18521.

[219] C. Watanabe, K. Fukuzawa, H. Mochizuki, T. Nakano, Y. Umezawa and M. Nishio, #15 *Symposium on Theoretical Organic Chemistry*, Aug. 2012, Sendai.

[220] Y. Umezawa and M. Nishio, *Supramol. Chem.*, 2013, **25**, 581-585.