PAPER

Check for updates

Preferred microenvironments of halogen bonds and hydrogen bonds revealed using statistics and QM/MM calculation studies†

Liping Zhou,[ab] Jintian Li,[ab] Yulong Shi,[ab] Leyun Wu,[ab] Weiliang Zhu [iD] *[ab] and Zhijian Xu [iD] *[ab]

Hydrogen bonds (HBs) and halogen bonds (XBs) are two essential non-covalent interactions for molecular recognition and drug design. As proteins are heterogeneous in structure, the microenvironments of protein structures should have effects on the formation of HBs and XBs with ligands. However, there are no systematic studies reported on this effect to date. For quantitatively describing protein microenvironments, we defined the local hydrophobicities (LHs) and local dielectric constants (LDCs) in this study. With the defined parameters, we conducted an elaborate database survey on the basis of 22 011 ligand–protein structures to explore the microenvironmental preference of HBs (91 966 in total) and XBs (1436 in total). The statistics show that XBs prefer hydrophobic microenvironments compared to HBs. The polar residues like ASP are more likely to form HBs with ligands, while nonpolar residues such as PHE and MET prefer XBs. Both the LHs and LDCs (10.69 ± 4.36 for HBs; 8.86 ± 4.00 for XBs) demonstrate that XBs are prone to hydrophobic microenvironments compared with HBs with significant differences ($p < 0.001$), indicating that evaluating their strengths in the corresponding environments should be necessary. Quantum Mechanics-Molecular Mechanics (QM/MM) calculations reveal that in comparison with vacuum environments, the interaction energies of HBs and XBs are decreased to varying degrees given different microenvironments. In addition, the strengths of HBs are impaired more than those of XBs when the local dielectric constant's difference between the XB microenvironments and the HB microenvironments is large.

Introduction

Non-covalent interactions (NCIs), including hydrogen bonds (HBs), electrostatic interaction, hydrophobic interaction, cation-π, π–π interaction and halogen bonds (XBs), play crucial roles in molecular recognition and rational drug design.[1–3] Apart from the hydrophobic effect, HBs are the most frequent among these interactions in ligand–protein structures.[4] HBs stabilize the three-dimensional structures of drug targets, such as proteins, DNA and RNA, and is also an important determinant of binding affinity of drugs for their targets.[5,6] Similarly, XBs have drawn much attention in recent years, because of their structural and functional roles in molecular interactions in biological systems, especially for ligand–protein interactions.[7,8] Moreover, HBs and XBs can not only be used to optimize the binding affinity of

ligands, but also have effects on ADME/T properties.[9,10] So far, several reports have revealed the essential roles of XBs and HBs in nucleic acids.[11–14] As for protein–ligand complexes, a research study reported that the lengths of HBs between a protein and a halogenated ligand are shorter than those estimated for non-halogenated ligands.[15] In addition, there are some researches focusing on comprehensive explorations for various non-covalent interactions.[4,16]

Because of the vital roles of HBs and XBs in ligand–protein recognition, determining their strengths and contributions to ligand binding affinity accurately is of great importance. It is generally believed that HBs are slightly stronger than XBs,[17] but the difference in specific microenvironments is not taken into account. The microenvironment exerts an influence on the intensities of HBs and XBs.[18] It is reported that a solvent-exposed HB contributes significantly less to net interaction energy than the same HB in a buried hydrophobic pocket[19] and a HB in polar environments is much weaker.[20,21] Although the strengths of HBs and XBs depend on their microenvironments to a certain degree, there is no consensus about the environmental bias of HBs and XBs. Craig *et al.* demonstrated that the formation of hydrogen-bonded co-crystals is favored

a Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, 201203, China. E-mail: zjxu@simm.ac.cn, wlzhu@simm.ac.cn

b School of Pharmacy, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing, 100049, China

from less polar solvents, but halogen-bonded co-crystals from more polar solvents,[22] which suggested that XBs prefer polar environments compared with HBs. Nonetheless, other results described that the amount of solvent used, and not just its polarity, is critical for co-crystallization.[23]

Herein, we carried out an elaborate study on the hydrophobicity and dielectric constants of microenvironments for HBs and XBs and further explored their impact on the strengths to better understand the microenvironmental preference of HBs and XBs. Our results reveal that XBs are more likely to be formed in nonpolar environments than HBs. Furthermore, Quantum Mechanics–Molecular Mechanics (QM/MM) calculations show that the strengths of HBs and XBs are attenuated in different degrees in a solvent in comparison with that in a vacuum depending on the local dielectric constants (LDCs). These findings not only deepen the understanding of HBs and XBs, but also provide guidance for drug design, *e.g.* lead optimization.

Methods

Database preparation

The three subsets of Protein Data Bank (PDB)[24] (September 13, 2022 release, 195 093 biological macromolecular structures in total) were explored for HBs and XBs detection. The first set is composed of 3D structures with approved small molecule drugs and their protein targets from DrugBank[25] (July 18, 2022 release, 698 complexes), the second includes complex structures with halogenated compounds extracted from PDB (13 476 complexes), and the third subset is PDBbind[26] (v2020, 19 443 complexes). Only ligand–protein systems were considered. In order to remove redundancy, only the structure with the highest resolution was retained among complexes with the same ligand and protein. Identical XBs or HBs may exist in homomultimer, as well as NMR structures, which would bias the results and thus only the monomer with the maximum number of atoms within 10 Å of the ligand was considered. After the above processes, the numbers of ligand–protein complexes for three subsets are 545 (DrugBank with approved small molecule drugs), 11 919 (PDB with halogenated ligands) and 15 703 (PDBbind) respectively, and the union of three sets consists of 22 011 structures in total.

Detection of XBs and HBs

For the detection of XBs with X⋯Y form, the cutoffs were set as X⋯Y distances shorter than the sum of vdW radii[27] ($d(\text{Cl}\cdots\text{N}) < 3.30$ Å, $d(\text{Br}\cdots\text{N}) < 3.40$ Å, $d(\text{I}\cdots\text{N}) < 3.53$ Å, $d(\text{Cl}\cdots\text{O}) < 3.27$ Å, $d(\text{Br}\cdots\text{O}) < 3.37$ Å, $d(\text{I}\cdots\text{O}) < 3.50$ Å, $d(\text{Cl}\cdots\text{S}) < 3.55$ Å, $d(\text{Br}\cdots\text{S}) < 3.65$ Å, $d(\text{I}\cdots\text{S}) < 3.78$ Å), and the C–X⋯Y angle is larger than $140°$[28] (Fig. S1, ESI†). For detection of XBs with X⋯π form, π systems from aromatic residues (PHE, TYR, HIS, and TRP) were considered in this study with the following criteria: $d(\text{Cl}\cdots\pi) < 4.2$ Å, $d(\text{Br}\cdots\pi) < 4.3$ Å, $d(\text{I}\cdots\pi) < 4.5$ Å, $\alpha < 60°$ and $\theta > 120°$ (Fig. S1, ESI†).

The HBs were analyzed using find_pairs function in PyMOL2.5[29] with the following parameters, respectively, *i.e.*, distance < 3.2 Å for N/O⋯N/O, distance < 3.5 Å for S⋯N/O

and distance < 3.8 Å for S⋯S, angle $> 125°$. For HBs composed of halogens as acceptors, the standards were set as X⋯Y distances shorter than the sum of vdW radii plus 1 Å to accommodate a hydrogen, and the C–X⋯Y angle smaller than $120°$.

Local hydrophobicity calculations

In order to describe the local hydrophobicity (LH) of microenvironments, the popular and efficient formalism named Molecular Hydrophobicity Potential (MHP) approach was used.[30–33] The midpoint of XB/HB donor and acceptor was considered as the center of the sphere, and the residue heavy atoms within 4 Å of the center were selected to calculate the local hydrophobicity. The LH was calculated using eqn (1), where k represents the number of selected atoms, f_i is the empirical parameterization of atomic hydrophobicity constants as reported,[34] and r is the distance of each atom away from the center, and α has the dimension of Å^{-1} (here, $\alpha = 1$ Å^{-1}).[35]

$$\text{LH} = \sum_{i=1}^{k} f_i \times e^{-\alpha r} \qquad (1)$$

Local dielectric constant calculations

The local dielectric constant (LDC) was described as eqn (2). The dielectric constants (diele_cons) of each amino acid in every structure were predicted by protCAD[36] and the dielectric constant of each atom lies on the type of residue it belongs to. Similar to LH calculation, the midpoint of the XB/HB donor and acceptor was considered as the center of the sphere, and the residue atoms within 5 Å of the center were selected to calculate LDC. The LDC was calculated using eqn (2), where k represents the number of selected atoms, dielec_cons is the dielectric constant of each atom, ASA_{atm} is the accessible surface area of each atom, ASA_{res} is the accessible surface area of the residue that the atom belongs to, and Distance is the distance of each atom away from the center. ASA was obtained by get_area script in PyMOL2.5.[29]

$$\text{LDC} = \sum_{i=1}^{k} \text{diele\_cons}_i \times (\text{ASA}_{\text{atm}}/\text{ASA}_{\text{res}})_i \times \left(\frac{1}{\text{Distance}}\right)_i \qquad (2)$$

QM/MM optimization

There were 6 systems of ligand–protein complexes chosen for QM/MM calculations with the two-layer ONIOM (our own N-layered integrated molecular orbital and molecular mechanics) method.[37] Only the ligand involved in the HBs and XBs was kept, and the other small molecules, water, and ions were all removed. The ligands in the QM/MM systems were protonated with pH = 7.35 by the LigPrep module in Schrödinger software (LigPrep, Schrödinger, LLC, New York, NY, 2020). Missing loops and sidechains in protein were modelled by SWISS-MODEL.[38] The pKa values of ionizable residues in the proteins were calculated by the PDB2PQR[39] at pH = 7.35, and hydrogen atoms were added accordingly. The complex systems were optimized by the QM/MM approach, *i.e.*, ONIOM. The ligands and

residues involved in XBs or HBs along with two linked residues were placed in the QM region which were described at the b3lyp[40,41]/LanL2DZ[42] level for iodine atoms and b3lyp/6-31g(d)[43] level for other atoms, and the rest were contained in the MM layer depicted by the AMBER parm96 force field.[44] QM/MM geometry optimization was performed by Gaussian 16 suite[45] without any constraints.

Calculation of the interaction energy

After structure optimization, the ligands and protein residues involved in the interaction in the QM layer were then chosen for single-point energy calculation at the m062x[46]/6-311++g(d,p)[47] level, while the SDD basis set[48] was adapted for iodine atoms. The interaction energies between the ligands and the protein residues were then assessed from eqn (3).

$$\Delta E = E_{com} - E_{lig} - E_{res} + BSSE \qquad (3)$$

Where ΔE is the interaction energy (IE), E_{com} is the energy of the whole complex in the QM layer, E_{lig} and E_{res} are the energies of the ligand and the protein residue, separately, and BSSE stands for the basis set superposition error corrections.[49] The energies were calculated in vacuums or under the corresponding LDC, and BSSE was estimated in vacuums. All these calculations were conducted with the Gaussian 16 suite of programs.

Statistical analysis

The homogeneity of variance was tested using the Bartlett method and then the independent two-sample T test with or without homogeneity of variance was used to analyze the significance for LHs or LDCs between HBs and XBs.

Results and discussion

The amino acids preference for HBs and XBs

Based on 22 011 ligand-complex structures, there were up to 91 966 HBs and 1436 XBs detected (945 for X···Y and 491 for X···π). Firstly, we analyzed the amino acids preference for HBs and XBs. The amino acids preference for HBs and XBs in the three subsets, *viz.* DrugBank with approved small molecule drugs (DrugBank with approved sm), PDB with halogenated ligands (PDB with X) and PDBbind is shown in Fig. 1A and B. For better comparison, the propensity is defined as the relative frequency of each amino acid in HBs or XBs divided by its background frequency in the respective data set. Hence, if the propensity is greater than 1, the residue is prone to being involved in HBs/XBs interaction.

For HBs, in DrugBank with approved sm, there are 11 residues with a propensity greater than 1, among which the 4 prominent residues are ARG, HIS, SER and THR. In PDB with X, the most prominent residues are ARG, ASP, CYS, HIS, TRP and TYR, while ARG, ASP and HIS dominate the PDBbind database (Fig. 1A). Most of these amino acids are polar except for TPR, indicating HBs tend to be formed between ligands and polar residues. Among these three subsets, the propensity of ARG and HIS are
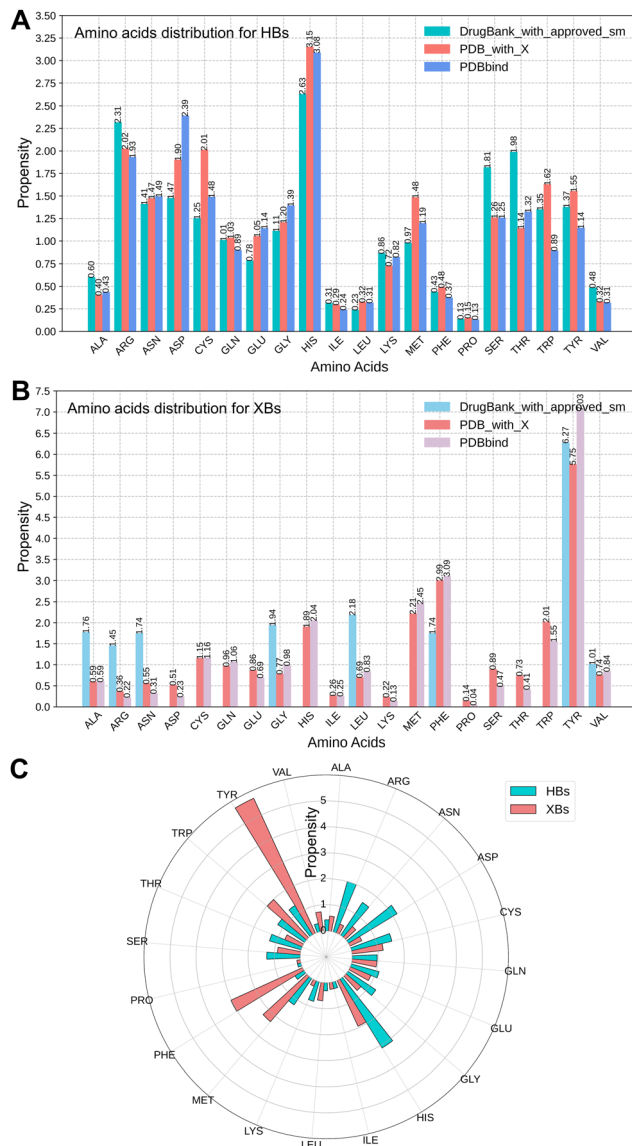

Fig. 1 The propensity of amino acids for HBs and XBs. Propensity is defined as the relative frequency of each residue in HBs/XBs divided by its relative background frequency in the respective data set. (A) Propensity of amino acids involved in HBs in three subsets (cyan: DrugBank with sm; salmon: PDB with X; blue: PDBbind). (B) Propensity of amino acids involved in XBs in three subsets (skyblue: DrugBank with sm; lightcoral: PDB with X; thistle: PDBbind). (C) Propensity of amino acids involved in HBs and XBs (cyan: HBs; salmon: XBs) in the union of three subsets. Propensity is calculated by the relative frequencies divided by the relative background frequency of each residue in the union of three subsets.

arresting, especially for HIS with the highest propensity (2.63 in DrugBank, 3.15 in PDB with X and 3.08 in PDBbind).

For XBs, there are in total 8 kinds of amino acids detected in the DrugBank with approved sm dataset and the propensities of residues GLY (1.94), LEU (2.18) and TYR (6.27) are the most significant. In PDB with X and PDBbind, there are 5 residues with propensity significantly greater than 1, *viz.* HIS, MET, PHE, TYR and TRP (Fig. 1B). TYR dominates in the three subsets (6.27 in DrugBank with approved sm, 5.75 in PDB with X and

17694 | *Phys. Chem. Chem. Phys.*, 2023, **25**, 17692–17699

This journal is © the Owner Societies 2023

7.03 in PDBbind) partly attributed to its role in the X···π interaction. The amino acids distribution in DrugBank with approved sm differs from that in the other two subsets (PDB with X and PDBbind) because only 14 XBs were detected, which is much less than that in the other two databases.

To compare the difference of amino acids preference between HBs and XBs, the results in the three subsets were combined and redundant systems were removed. All relative frequencies are divided by the background frequency of each residue in the union of three subsets. As indicated from Fig. 1C, TYR, PHE, MET, TRP and HIS are more prone to being involved in XBs while HIS, ASP and ARG prefer HBs. The propensity for TYR is so remarkable (1.31 in HBs $vs.$ 5.71 in XBs), because it accounts for 50.92% of the X···π halogen bonds and the same is true for PHE (data not shown). As for the difference between HBs and XBs, the propensities of TRP, PHE and MET in XBs are 0.74, 2.62, and 1.07 more than that in HBs respectively, while HBs are more prone to existing between ligands and amino acids HIS, ASP, ARG, ASN compared with XBs (Fig. 1C and Table S1 for details, ESI†). Obviously, the former three are nonpolar amino acids and the latter four are polar, indicating that XBs prefer nonpolar microenvironments compared with HBs. Without considering the X···π XBs, the propensities of three nonpolar residues (VAL, MET and LEU) in XBs are significantly greater than that in HBs (Fig. S2, ESI†). These findings suggest that XBs are more likely to be generated between ligands and nonpolar residues, while HBs are not.

The elements preference for HBs and XBs

Furthermore, we analyzed the elements preference for HBs and X···Y XBs. As indicated in Fig. 2, the N in sidechain ($N_{sidechain}$) with propensity 2.03 and O in sidechain ($O_{sidechain}$) with propensity 1.60 are more inclined to interact with ligands through HBs than N in mainchain ($N_{mainchain}$) (0.79) and O in

mainchain ($O_{mainchain}$) (0.47). For XBs, the propensity of $N_{sidechain}$ is a little greater than that of $N_{mainchain}$ (0.44 $vs.$ 0.15), while the tendency of $O_{sidechain}$ and $O_{mainchain}$ is reversed (1.33 $vs.$ 1.68). In other words, $N_{sidechain}$ prefers XBs than $N_{mainchain}$ and $O_{mainchain}$ is more apt to interact with ligands through XBs than $O_{sidechain}$. In addition, the propensity of $O_{mainchain}$ (1.68) is remarkably greater than that of $N_{mainchain}$ (0.15) since $N_{mainchain}$ has a hydrogen, making it hard to be the acceptors of XBs owing to the steric hindrance and weak electronegativity.

When it comes to the difference between HBs and XBs, both $N_{mainchain}$ and $N_{sidechain}$ tend to form HBs with ligands instead of XBs ($N_{mainchain}$: 0.79 in HBs $vs.$ 0.15 in XBs; $N_{sidechain}$: 2.03 in HBs $vs.$ 0.44 in XBs). In contrast, the propensity of $O_{mainchain}$ in XBs is significantly greater than that in HBs (1.68 $vs.$ 0.47), while there is no conspicuous difference for $O_{sidechain}$ between HBs and XBs (1.60 $vs.$ 1.33). Impressively, the propensity of S in HBs is strikingly less than that in XBs (1.30 $vs.$ 6.36), which may be attributed to the fact that the sidechain of methionine can't serve as an HB donor, but can be a strong XB receptor. In general, the preferences of elements involved in HBs and XBs for these two kinds of non-covalent interactions (NCIs) are diverse.

The distribution of microenvironmental hydrophobicities for HBs and XBs

To quantify the hydrophobicity value of microenvironments, Molecular Hydrophobicity Potential (MHP) was used to calculate the local hydrophobicity (LH) for each HB or XB, which was proved to be popular and efficient.[30–33] There are several functions to model the MHP-decay from analogy with electrostatic potentials including exponential,[35] hyperbolic[50] and Fermi-like potential,[51] but it's reported that in most MHP applications to proteins, a simple exponent is sufficient to obtain reasonable results.[30] Moreover, only atoms closer than 4 Å are effectively contributing to the MHP.[51] Hence, we used the exponential formula as eqn (1) with a distance of 4 Å to describe LHs for HBs and XBs. Since there is no specific experimental data for comparison, the accuracy of the absolute value of LHs calculated by this formula cannot be determined. Perhaps, there will be relevant researches on this in the future. Nonetheless, the current results can reflect the relative differences of microenvironmental hydrophobicity between HBs and XBs to some extent.

As shown in Fig. 3, the LHs for HBs are dispersively distributed between −0.38 and −0.07, while that of XBs are concentrated between −0.25 and −0.12. In addition, the mean value of LHs for HBs (−0.23 ± 0.13) is smaller than that of XBs (−0.19 ± 0.10) (Table S2, ESI†). Statistically, the results of an independent two-sample T-test also demonstrate that the LHs between HBs and XBs have a significant difference ($p < 0.001$), indicating that XBs are more prone to hydrophobic environments than HBs.

In order to explore whether different halogens affect the results, we analyzed the difference of LH distribution between HBs and XBs with Cl, Br and I, separately (Fig. S3, ESI†). There are 1436 XBs and the ratios of Cl, Br, I are 69%, 21% and 10%, respectively. As indicated in Fig. S3 (ESI†), the trends of LHs for
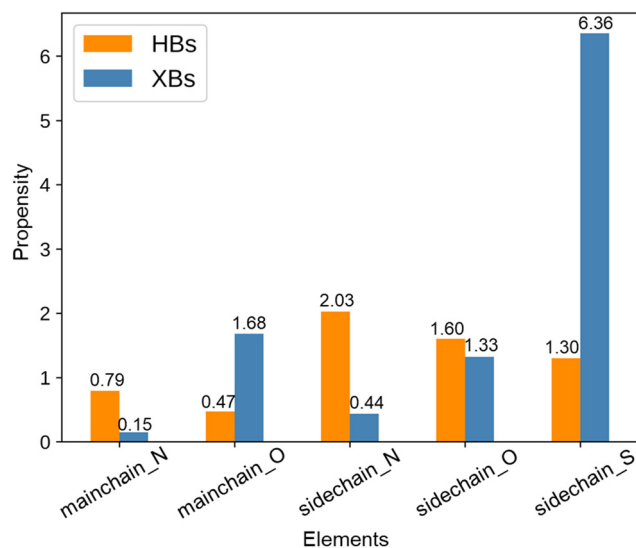

Fig. 2 The elements propensity for HBs (orange) and X···Y XBs (blue). The definition of propensity is that the relative frequency of each element is divided by its relative background frequency in the whole database.
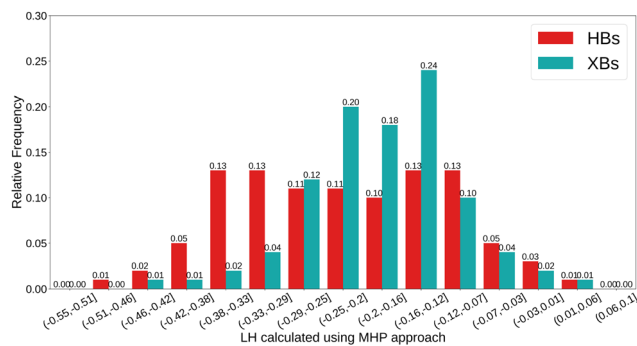
This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 17692–17699 | **17695**

Fig. 3 The frequency distribution of LHs for HBs (red) and XBs (cyan) based on MHP. The *x*-coordinate represents the range of LH, and the *y*-coordinate represents the relative frequency. The higher the LH value, the greater the hydrophobicity. The unit of LH is log *P* (octanol–water).
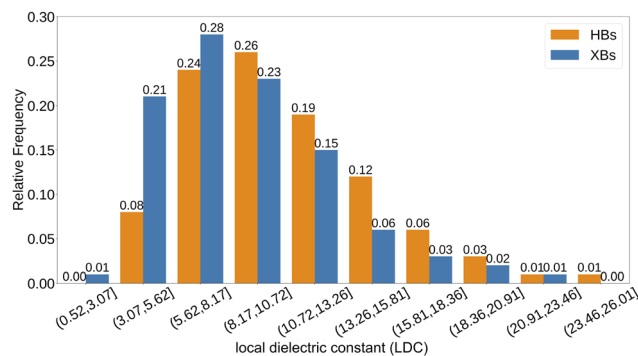

Fig. 4 The frequency distribution of LDCs for HBs (orange) and XBs (navy blue). The *x*-coordinate represents the range of LDC, and the *y*-coordinate represents the relative frequency.

HBs and XBs with three halogens are all in accord with the overall tendency in Fig. 3. Moreover, the specific values of LHs for XBs with Cl, Br and I are -0.20 ± 0.10, -0.19 ± 0.10 and -0.18 ± 0.07, which are all greater than that of HBs (-0.23 ± 0.13) with significant differences ($p < 0.001$) (Table S2, ESI†). Therefore, the kind of halogens has no influence on the conclusion that HBs prefer hydrophilic microenvironments than XBs. In conclusion, the microenvironments of XBs are strikingly more hydrophobic than that of HBs, which may affect the strength of each HB/XB. Therefore, it may be better to predict the strength of XB/HB under their specific environments. Also, for structure-based drug design, halogen modification for the sake of introducing a potential XB in a hydrophobic protein subpocket may be a more rational choice than considering a potential HB. Besides, although halogens are considered hydrophobic in nature, the polarization and enthalpy stabilization make them favorable for the formation of non-hydrophobic interactions like XBs,[52] which needs further investigation.

The distribution of microenvironmental dielectric constants for HBs and XBs

The dielectric constant is associated with the polarity of the environment and the average dielectric constant inside the protein is relatively low, about 7, and reaches a value of about 30 at the protein's surface.[53] However, in terms of previous research, there are no relevant experiments to determine the dielectric constant at the atomic level or residue level like hydrophobicity determination. Secondly, the dielectric constant is a more macroscopic concept affected by more factors. Certainly, it does not rule out that there will be follow-up studies on this crucial topic. Hence, a method as described in eqn (2) was applied to predict the local dielectric constants (LDCs) for HBs and XBs in this study.

As indicated in Fig. 4, the peak of LDC of HBs is slightly higher than that of XBs. Besides, the relative frequencies of less than 8.17 for HBs and XBs LDCs are 0.32 and 0.50 separately. In terms of the specific values, the average value of LDCs for HBs is 10.69 ± 4.36, while that of XBs is 8.86 ± 4.00 with a difference of 1.83 as for the mean value. Statistically, the LDC of HBs is larger than that of XBs with a significant difference ($p < 0.001$),

which is in agreement with the trend of LH. Both LDCs and LHs demonstrate that HBs prefer hydrophilic environments than XBs. In terms of XBs with specific halogen elements, the tendencies of Cl, Br and I are consistent with the overall trend (Fig. S4 (ESI†) and Fig. 4). The mean values of LDCs for XBs with Cl, Br, and I are 8.97 ± 3.97, 8.53 ± 3.92, and 8.87 ± 4.26 respectively, which are all smaller than that of HBs (10.69 ± 4.36) with a significant difference ($p < 0.001$). Both the global tendency for HBs and XBs and statistical results of XBs with different halogens demonstrate that XBs show a preference toward hydrophobic environments. Therefore, given the micro-environmental preference of HBs and XBs, taking specific LDCs into account is advisable when estimating the binding capability of ligands or comparing the strengths of different NCIs.

Interaction energy for HBs and XBs

HBs and XBs both make great contributions to the binding affinity of ligands, which are influenced by the dielectric constant. Based on the calculated LDCs, 6 systems as described in Fig. 5A–F were selected as examples to determine the interaction energies of HBs and XBs in a vacuum (IE$_{vacuum}$) and environment with the corresponding LDC (IE$_{solvent}$). After geometry optimization, the distances and angles for HBs and XBs are shown in Fig. 5. As demonstrated in Table 1, the strength of HBs is stronger than that of XBs either in a vacuum or under the corresponding LDCs in each system, except for 3sw8 (PDB ID). In the complex structure 3sw8, the XB is formed by a chlorine and carboxyl group with a negative charge of residue E (Fig. 5F), so the IE is extremely strong (IE$_{vacuum}$: -16.68 kcal mol^{-1}). Both in the two different environments, the strengths of these two HBs (IE$_{vacuum}$: -0.58 kcal mol^{-1}, IE$_{solvent}$: -0.36 kcal mol^{-1}; IE$_{vacuum}$: -3.67 kcal mol^{-1}, IE$_{solvent}$: -2.55 kcal mol^{-1}) are weaker than the XB (IE$_{vacuum}$: -16.68 kcal mol^{-1}, IE$_{solvent}$: -3.57 kcal mol^{-1}) in 3sw8.

Apparently, the strengths of all interactions are attenuated in solvent to varying degrees. When it comes to the extent of IE$_{solvent}$ reduction compared to IE$_{vacuum}$, the IE change ratios for XBs are smaller than that of HBs in 2 systems (-0.20 for 5xdv, -0.07 and -0.03 for 6rlw). In 5xdv, the difference between LDCs for HBs and XBs is 7.83 and those in 6rlw are 11.37 and
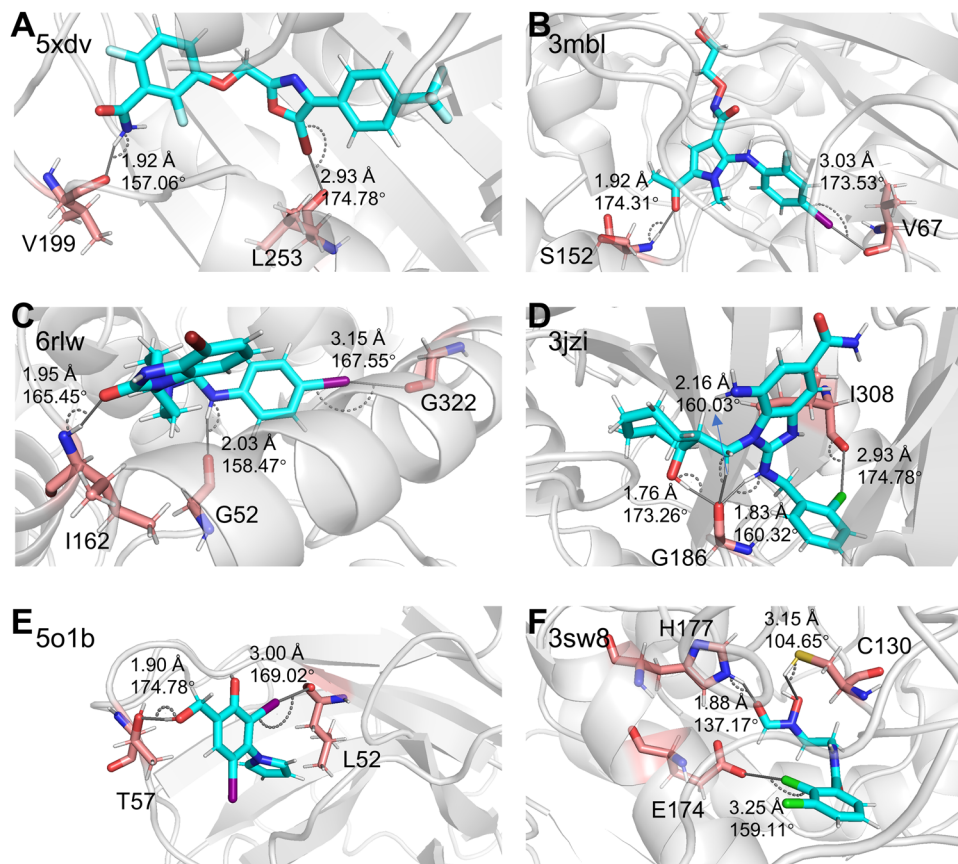
Fig. 5 The binding patterns of 6 systems after QM/MM optimization. The distance of XBs is measured between the XB acceptor and halogen atom and that of HBs is between the HB acceptor and hydrogen. The angle of XBs is measured on C–X···Y and that of HBs is measured on D–H···A where D is the donor of HB and A is the acceptor. PDB IDs are labeled on the upper left. Proteins are shown in cartoon with key residues in salmon sticks and ligands in cyan sticks.

Table 1 The interaction energies in 6 systems (kcal mol^{-1})

PDB ID	residue	NCI	X···Y	LDC	IE$_{vacuum}$	IE$_{solvent}$	ΔLDC[a]	IE change ratio[b]
5xdv	L253	XB	Br···O	5.66	−4.19	−2.98	—	−0.29
5xdv	V199	HB	N···O	13.49	−9.03	−4.60	7.83	−0.49
3mbl	V67	XB	I···O	10.3	−4.23	−2.61	—	−0.38
3mbl	S152	HB	N···O	14.12	−5.96	−3.89	3.82	−0.35
6rlw	G322	XB	I···O	3.83	−2.83	−1.95	—	−0.31
6rlw	G52	HB	N···O	15.2	−10.68	−6.57	11.37	−0.38
6rlw	I162	HB	N···O	13.84	−6.25	−4.15	10.01	−0.34
3jzi	I308	XB	Cl···O	5.69	−4.78	−0.64	—	−0.87
3jzi	G186	HB	N···O\O···O\C···O	28.82	−27.28	−15.21	23.13	−0.44
5o1b	L52	XB	I···O	6.5	−4.85	−3.52	—	−0.27
5o1b	T57	HB	O···O	12.01	−7.89	−6.59	5.51	−0.17
3sw8	E174	XB	Cl···O	7.29	−16.68	−3.57	—	−0.79
3sw8	C130	HB	S···O	13.02	−0.58	−0.36	5.73	−0.38
3sw8	H177	HB	N···O	5.59	−3.67	−2.55	−1.7	−0.31

[a] For each system: ΔLDC = LDC (HB) − LDC (XB). [b] IE change ratio = (IE$_{solvent}$ − IE$_{vacuum}$)/IE$_{vacuum}$.

10.01. In another three systems, namely 3mbl, 5o1b and 3sw8, the differences between LDCs are smaller than 6, and the change ratios of IEs for XBs are greater than that of HBs (0.03 for 3mbl, 0.10 for 5o1b and 0.41 for 3sw8). That is to say, if the LDC difference between HBs and XBs is relatively large (more than 7 here), the strength of HBs may be weakened more than that of XBs. The system with PDB ID 3jzi is an exception, but it's

hard to explain the variation ratio on account of three HBs here (Fig. 5D).

Generally speaking, HBs are considered to be stronger or equal to XBs under the same environment. However, XBs prefer nonpolar environments according to our statistics. Moreover, HBs in polar environments tend to be in the solvent regions and are susceptible to solvent competition, which may be more

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 17692–17699 | **17697**

unstable kinetically than XBs. Hence, it is possible for the strength of XB to be greater than that of HB in practical situations.

Conclusion

In this study, we carried out a statistical survey based on 22 011 structures extracted from PDB to determine the microenvironmental preference for HBs and XBs and studied its influence on the strengths with QM/MM calculations. First of all, polar amino acids like HIS, ASP, ARG, and ASN are more likely to interact with ligands through HBs, while nonpolar amino acids including TRP, PHE and MET are more likely to interact with ligands through XBs. Furthermore, $N_{mainchain}$ and $N_{sidechain}$ are both more prone to being involved in HBs than XBs, while the tendencies of $O_{mainchain}$ and S are different. Additionally, even though previous research revealed that XBs may prefer polar solvent by the means of competitive co-crystal formation,[22] the calculated LHs and LDCs based on 91 966 HBs and 1436 XBs in ligand–protein systems both show that XBs show preference for hydrophobic microenvironments compared with HBs. Our results also demonstrate that the strengths of XBs and HBs are both reduced to different extents hinging on their local environments, and the strengths of HBs are impaired more than those of XBs when the difference between their LDCs is greater than 7. Our findings provide a useful hint for structure-based drug design.

Data and software availability

All data used for analyses are extracted from the public database named PDB. The PDB IDs lists, the python scripts used for HBs/XBs detection and the input files of 6 systems for QM/MM calculation are available at the following github repository: https://github.com/Zhijian-Xu/hb_xb_microenv.

Author contributions

Conceptualization: Z. X., W. Z. Methodology: L. Z., J. L., Y. S. Investigation: L. Z. Visualization: L. Z., Y. S., L. W. Supervision: Z. X., W. Z. Writing – original draft: L. Z. Writing – review & editing: Z. X., W. Z., L. Z. All authors read and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 C. Bissantz, B. Kuhn and M. Stahl, *J. Med. Chem.*, 2010, **53**, 5061–5084.

2 E. Persch, O. Dumele and F. Diederich, *Angew. Chem., Int. Ed.*, 2015, **54**, 3290–3327.

3 D. K. Smith, *J. Chem. Educ.*, 2005, **82**, 393.

4 R. Ferreira de Freitas and M. Schapira, *MedChemComm*, 2017, **8**, 1970–1981.

5 P. W. Kenny, *J. Cheminf.*, 2019, **11**, 8.

6 D. Chen, N. Oezguen, P. Urvil, C. Ferguson, S. M. Dann and T. C. Savidge, *Sci. Adv.*, 2016, **2**, e1501240.

7 M. R. Scholfield, C. M. V. Zanden, M. Carter and P. S. Ho, *Protein Sci.*, 2013, **22**, 139–152.

8 P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16789–16794.

9 Z. Xu, Z. Yang, Y. Liu, Y. Lu, K. Chen and W. Zhu, *J. Chem. Inf. Model.*, 2014, **54**, 69–78.

10 P. W. Kenny, *J. Med. Chem.*, 2022, **65**, 14261–14275.

11 K. Mu, Z. Zhu, A. Abula, C. Peng, W. Zhu and Z. Xu, *J. Med. Chem.*, 2022, **65**, 4424–4435.

12 D. Mitra, N. Bankoti, D. Michael, K. Sekar and T. N. G. Row, *J. Chem. Sci.*, 2020, **132**, 93.

13 A. Frontera and A. Bauzá, *J. Chem. Theory Comput.*, 2020, **16**, 4744–4752.

14 S. K. Panigrahi and G. R. Desiraju, *J. Biosci.*, 2007, **32**, 677–691.

15 J. Poznański, A. Poznańska and D. Shugar, *PLoS One*, 2014, **9**, e99984.

16 Z. Xu, Q. Zhang, J. Shi and W. Zhu, *J. Chem. Inf. Model.*, 2019, **59**, 3389–3399.

17 Y. Q. Wang, R. J. Wang, Q. Z. Li and Z. W. Yu, *Molecules*, 2022, **27**, 8523.

18 M. L. Bogado, R. N. Villafañe, J. L. Gómez Chavez, E. L. Angelina, G. L. Sosa and N. M. Peruchena, *J. Chem. Inf. Model.*, 2022, **62**, 6494–6507.

19 B. K. Shoichet, *Nat. Biotechnol.*, 2007, **25**, 1109–1110.

20 J. Gao, D. A. Bosco, E. T. Powers and J. W. Kelly, *Nat. Struct. Mol. Biol.*, 2009, **16**, 684–690.

21 W. Cai, D. Xu, F. Zhang, J. Wei, S. Lu, L. Qian, Z. Lu and S. Cui, *Nano Res.*, 2022, **15**, 1517–1523.

22 C. C. Robertson, J. S. Wright, E. J. Carrington, R. N. Perutz, C. A. Hunter and L. Brammer, *Chem. Sci.*, 2017, **8**, 5392–5398.

23 J. Lombard, T. le Roex and D. A. Haynes, *Cryst. Growth Des.*, 2020, **20**, 7384–7391.

24 S. K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, L. Chen, G. V. Crichlow, C. H. Christie, K. Dalenberg, L. Di Costanzo, J. M. Duarte, S. Dutta, Z. Feng, S. Ganesan, D. S. Goodsell, S. Ghosh, R. K. Green, V. Guranović, D. Guzenko, B. P. Hudson, C. L. Lawson, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Persikova, C. Randle, A. Rose, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, Y.-P. Tao, M. Voigt, J. D. Westbrook, J. Y. Young, C. Zardecki and M. Zhuravleva, *Nucleic Acids Res.*, 2021, **49**, D437–D451.

25 D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, *Nucleic Acids Res.*, 2018, **46**, D1074–d1082.

26 Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu and R. Wang, *Bioinformatics*, 2015, **31**, 405–412.

17698 | *Phys. Chem. Chem. Phys.*, 2023, **25**, 17692–17699

This journal is © the Owner Societies 2023

27 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.

28 Y. Lu, Y. Wang and W. Zhu, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4543–4551.

29 Schrodinger, LLC, unpublished work.

30 R. G. Efremov, A. O. Chugunov, T. V. Pyrkov, J. P. Priestle, A. S. Arseniev and E. Jacoby, *Curr. Med. Chem.*, 2007, **14**, 393–415.

31 T. V. Pyrkov, A. O. Chugunov, N. A. Krylov, D. E. Nolde and R. G. Efremov, *Bioinformatics*, 2009, **25**, 1201–1202.

32 A. A. Polyansky, M. Hlevnjak and B. Zagrovic, *Nat. Commun.*, 2013, **4**, 2784.

33 A. A. Polyansky and B. Zagrovic, *J. Phys. Chem. Lett.*, 2012, **3**, 973–976.

34 A. K. Ghose, V. N. Viswanadhan and J. J. Wendoloski, *J. Phys. Chem. A*, 1998, **102**, 3762–3772.

35 J.-L. Fauchére, P. Quarendon and L. Kaetterer, *J. Mol. Graphics*, 1988, **6**, 203–206.

36 D. H. Pike and V. Nanda, *Biopolymers*, 2015, **104**, 360–370.

37 M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, A. Sieber and K. Morokuma, *J. Phys. Chem.*, 1996, **100**, 19357–19363.

38 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res.*, 2018, **46**, W296–w303.

39 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, *Nucleic Acids Res.*, 2004, **32**, W665–W667.

40 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

41 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

42 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270–283.

43 G. A. Petersson and M. A. Al-Laham, *J. Chem. Phys.*, 1991, **94**, 6081–6090.

44 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

45 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. MontgomeryJr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

46 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **119**, 525.

47 K. Raghavachari and G. W. Trucks, *J. Chem. Phys.*, 1989, **91**, 1062–1065.

48 M. Dolg, U. Wedig, H. Stoll and H. Preuss, *J. Chem. Phys.*, 1987, **86**, 866–872.

49 S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

50 E. Audry, J. C. Colleter and P. Dallet, *Eur. J. Med. Chem.*, 1986, **21**, 71–72.

51 W. Heiden, G. Moeckel and J. Brickmann, *J. Comput.-Aided Mol. Des.*, 1993, **7**, 503–514.

52 P. S. Ho, *Top. Curr. Chem.*, 2015, **358**, 241–276.

53 L. Li, C. Li, Z. Zhang and E. Alexov, *J. Chem. Theory Comput.*, 2013, **9**, 2126–2136.

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 17692–17699 | **17699**