J. Chem. Sci. (2020)132:93
https://doi.org/10.1007/s12039-020-01794-1

REGULAR ARTICLE

C-halogen…pi interactions in nucleic acids: a database study

DEBARGHYA MITRA[a], NAMRATA BANKOTI[a], DALIAH MICHAEL[a], K SEKAR[a,*] and T N GURU ROW[b,*]

[a]Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560 012, India
[b]Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560 012, India
E-mail: sekar@iisc.ac.in; gururow@iisc.ac.in

MS received 22 August 2019; revised 10 April 2020; accepted 16 April 2020

Abstract. A Protein Data Bank study was conducted to check the role of C-halogen(X)…pi interactions (X = F, Cl, Br, I) in nucleic acids. The presence of halogens can be attributed to not only the modified residues of the nucleic acid but also the ones associated with the ligand. The study reports the presence of bromine amongst modified residues of the nucleic acid being the maximum than any other halogens. It is important to consider these interactions as they seem to be responsible for increasing the affinity of the ligand to the biomolecule concerned. The role of halogen-pi interactions becomes important to be assessed in order to design halogenated ligands that will certainly be more effective in increasing the binding of the corresponding ligand to the target nucleic acid. The study attempts at listing the halogens that are prominent in the existing nucleic acid and ligand interactions which are deposited in the PDB. Inter-atomic interactions based on C-X…pi were analyzed to be found in several PDB IDs and similar ligand halogenation can be carried out to effectively increase binding affinity of halogenated ligands to nucleic acid based targets in case of several diseases.

Keywords. C-halogen…pi interactions; nucleic acids; database study; halogenated ligands; weak macromolecular interaction.

1. Introduction

Nucleic acids play a key role and serve as the data repository for genetic information in the living system. We find that nucleic acids are double-stranded DNA or single-stranded RNA and vary in terms of nitrogenous bases (pyrimidines or purines depicted in Figure 1) which along with the pentose sugar and the phosphate group give rise to a nucleotide, serving as the basic functional unit of nucleic acids. Exceptions include the absence of the phosphate group in nucleosides. Based on base-stacking which involves hydrogen bonds and other non-covalent interactions, like van der Waals and dipole–dipole interactions, the major bonding patterns play an important role in conferring the stability to these biomolecules. Although very weak in nature, these non-covalent interactions are the most numerous, amounting to their predominance in playing an important role in stability. Complementary base pairing also plays a pivotal role in conferring the defined three dimensional structures to the nucleic acid.

Several articles have established the presence of halogen bonds present amongst biomolecules with halogen-pi interactions being broadly studied among proteins, involving the residues that possess an aromatic ring. These interactions have been described in protein–ligand interactions and how these halogen bonds play an important role in binding and recognition of the subsequent ligand to the biomolecule. The surge in use of halogenated ligands also serves as a reminder and points to the need of unravelling the intricacies of this non-covalent interaction. Synonymous with hydrogen bonding interactions, the halogen-pi interaction essentially involves a halogen atom present either in the ligand or the modified residue and its corresponding interaction (pi–pi interaction) with the neighboring aromatic residue. A typical halogen-pi bonding involves the aromatic benzene rings which act as donors and the C-X (where X = F, Cl, Br, I) act as the acceptors. The Lewis bases prefer to approach the ligands (C-X) from head-on orientation or a face-on approach. Halogen bonds are generally characterized

*For correspondence

Published online: 04 August 2020

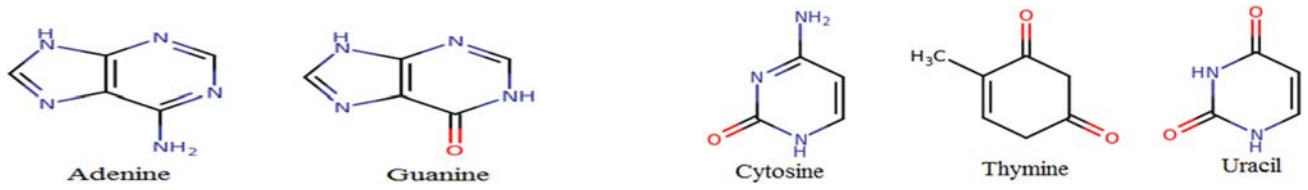

Figure 1. Nitrogenous bases of nucleic acids with aromatic rings that were used in the study.

by the attractive nature and lead to shorter lengths than the corresponding van der Waals radii. The bond strength increases with increase in molecular weight (F < Cl < Br < I) leading to the shorter contacts. As one may have already assumed as to the prevalence of these bonds predominantly around hexane derivatives, we bring to fore the structure of the nitrogenous bases, purines and pyrimidines and their subsequent attachment with the sugar residues and phosphate group which leads to the formation of nucleotides or nucleosides (without phosphate group), the building blocks of DNA or RNA.

The role of hydrogen bonds in any biomolecule can never be over-emphasized, with energies varying from 30 kcal/mol for the strongest to less than 0.5 kcal/mol for the weakest. These C-H…O interactions were reported by Sutor[1] amongst nucleic acids. Halogen interactions although weak are found to be of profound importance in supramolecular assembly and in protein–nucleic acid interactions. Exhaustive work has been carried out on halogen-pi interactions in protein ligand complexes.[2–4] Several articles have conducted screening of halogen-pi interactions using PDB as the source. In case of prevalent halogen-pi interactions in nucleic acids, we find that Auffinger *et al.* and Vallejos *et al.*[5,6] have conducted studies on a few such nucleic acid structures. The study of C-X…pi interactions in nucleic acids by incorporating and refining all the available nucleic acids in PDB has never been sufficiently elucidated.

The cutoff value (D_{max} depicted in Figure 2) is defined for each halogen based on the sum of the specific halogen's van der Waals radius and includes half the thickness of the benzene ring along with an added tolerance to incorporate the electrostatic interaction.[7]

2. Preparation of structure

The Protein Data Bank[8] (issue:June,2018) was utilized in this database survey and the corresponding nucleic acids structures (DNA,RNA,DNA/RNA–protein, including even hybrid) were considered and were analyzed in search of existing halogen-pi interactions.

An approach similar to that of Saraogi *et al.*[3,9] was used. Table 2 depicts the total number of PDB files with the appropriate interactions as listed. Based on R-factor and resolution refinements, the number of files were reduced. PDB structures solved by X-Ray crystallography with a resolution equal to 2.5 Å (Angstroms) and R factor refinements amounting to or less than 0.25 were considered only in this study. Table 3 shows the results based on the additional filters through values of θ (theta) and ϕ (phi) to include the directionality of these interactions.

The analysis is mainly based upon the contacts between the halogen and the six-membered benzene ring prevalent among all the nucleic acid nitrogenous bases. The filters applied ensured that only possible halogen bond forming interactions were considered which are defined to be less than the van der Waals radii of the respective halogens (Table 1). The C-X..pi contacts which fell below the defined Dmax were listed (Table 2) as hits. Additional filters were used in order to determine directionality through defined angles θ and ϕ for each of the selected structures.

3. Results and Discussion

The need for phasing of crystallographic data through halogenation and with the increase in halogenated ligands acting as either inhibitors or promoters, we find that the role of halogen bonding and specifically halogen-pi interactions must be adequately elucidated to ensure that the subsequent studies involving recognition and binding to halogenated ligands or biomolecules are not misinterpreted.

The dataset that was assembled and used in this work to study the halogen-pi interactions in nucleic acids are specifically tabulated and results of the same are shown as in Table 2. Table 3 indicates the θ and ϕ constraints which were used for determining the directionality based on the allowed cut-offs for θ and ϕ values. The total number of files analyzed to narrow

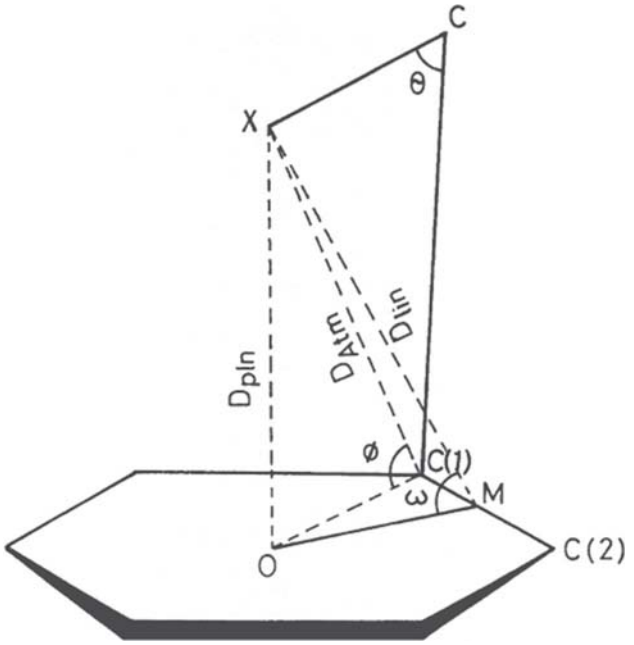

Figure 2. The parameters taken to study the C-X..pi interactions include: Angle OCX = θ; Angle OMX = omega; Angle OC(1)X = ϕ; Distance: OX = $D_{planar;}$ MX = Dlin; C(1)X = Datm; Cutoffs:Region 1: θ < 60°; − 90° < omega < 90°; Dplanar < Dmax; Region 2: θ < 60°; −130° < omega < 130°; Dlinear < Dmax; Region 3: θ < 60°; 50° < ϕ 90°; Datm < Dmax.

Table 1. Dmax (cutoff values) used.

Interaction	vdW radius of the halogen (Å)	D_{max} (Å)
C-F…pi	1.47	4
C-Cl….pi	1.75	4.2
C-Br….pi	1.85	4.5
C-I…pi	1.98	4.8

Table 2. Occurrence of C-X…pi interactions in nucleic acids.

Interaction	Number of files	Number of selected files	Number of hits	Mean distance (Å)
C-F…pi	63	1	1	NA
C-Cl…pi	299	2	10	3.93
C-Br…pi	260	30	85	4.11
C-I…pi	28	1	2	4.35

down to the number of files is listed based on the type of interaction. In Table 2, the number of selected files points to the structures that met the Resolution and R-factor criteria and in Table 3, the number of selected files indicates the files filtered from Table 2 which

Table 3. Occurrence of C-X…pi interactions in nucleic acids after putting in θ and ϕ constraint.

Interaction	Number of files	Number of selected files	Number of hits	Mean distance (Å)
C-F…pi	63	1	1	NA
C-Cl…pi	299	2	5	3.92
C-Br…pi	260	15	19	3.95
C-I…pi	28	1	1	4.30

meet the directionality criteria of specific θ and ϕ as specified in Figure 1. In each table, hits refer to number of non-bonded contacts which are identified as C-halogen…pi interactions.

It is observed that the number of contacts involving C-Br….pi is more than any other halogen. This can be generally attributed to the presence of brominated residues being introduced amongst nucleic acids as the primary stabilizing influence of the concerned biomolecule. Even C-Cl…pi interactions are significant and their corresponding R-factor versus resolution are shown in Figure 3.

The heavier halogen, iodine, is rarely used for modifying residues in nucleic acids and the presence of fluorine is less, owing to the instituted resolution and refinements filters while selecting the files from PDB. Therefore, the number of hits based on these two C-I and C-F interactions with pi interactions is too few to draw any significant conclusion. The scattergrams that have been constructed for each of these C-Br..pi and C-Cl..pi interactions are present in Figures 4 and 5 respectively based on the θ and ϕ angles exhibited by the interactions.

Even now, on removing the refinements as in Auffinger 2004 the results in Table 4 indicate that maximum number of the C-X..pi interactions are prevalent in case of interactions associated with brominated ligands. The number of hits for chlorinated and iodinated halogens remains considerably reduced based on their application in nucleic acids. A considerable number of hits are seen in case of fluorinated ligands which show over 800 hits spread over 19 files from the dataset assembled from PDB. Interactions with fluorinated ligands are generally neglected owing to the smaller size of the halogen and the resultant strength of C-X…pi interaction. The results indicate that the interactions fall well within the defined cutoffs (D_{max}).

In the same analysis, we came across several interatomic C-halogen…pi interactions prevalent for several DNA-based targets and its corresponding

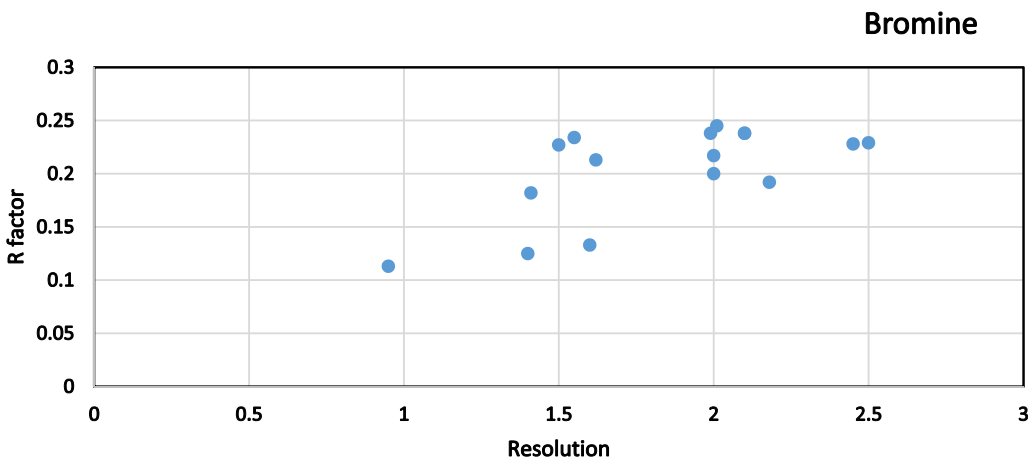

Figure 3. Plots of the crystallographic R-factor versus the Resolution of the nucleic acid structures used in the study for the different halogens.

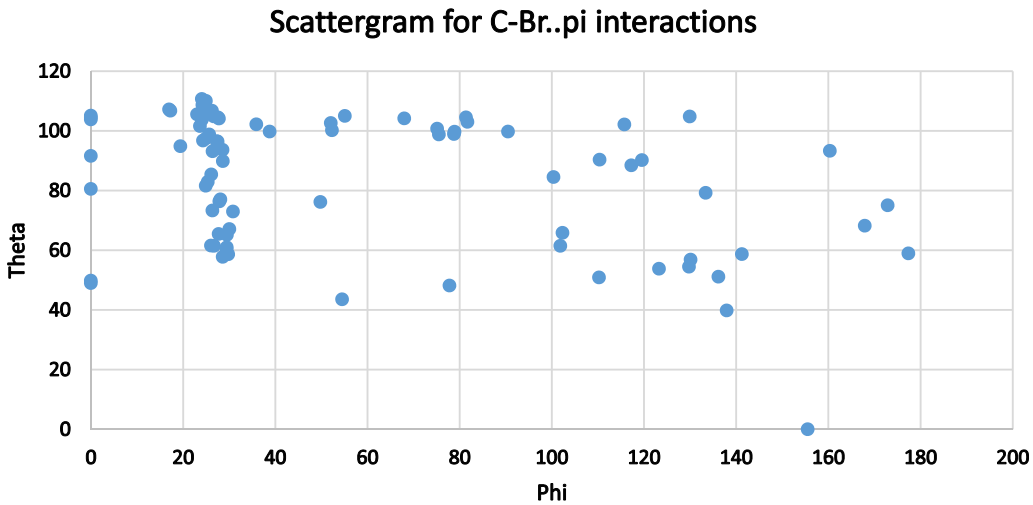

Figure 4. Scattergram for C-Br…pi interactions (θ vs ϕ).

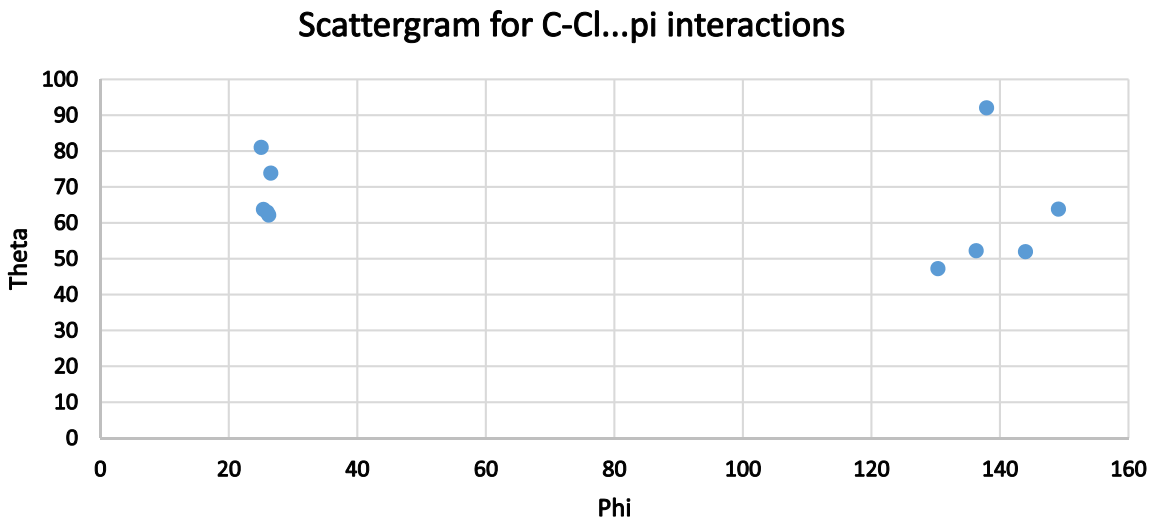

Figure 5. Scattergram for C-Cl…pi interactions (θ vs ϕ).

Table 4. Occurrence of C-X…pi interactions without refinements in nucleic acids.

Interaction	Number of selected files	Number of hits	Mean distance (Å)
C-F…pi	19	842	3.72
C-Cl…pi	6	9	3.95
C-Br…pi	118	244	3.91
C-I…pi	4	4	4.18

Table 5. PDB IDs exhibiting inter-atomic C-halogen…pi interactions.

Interaction	PDB IDs
C-F…pi	5EIX,5BJP,5BJO,4Z3O,4Z2D,4Z2C,3RAD,3RAE,3K9F,3FOE
C-Cl…pi	4ZTJ,4ZTF
C-Br…pi	NA
C-I…pi	NA

Table 6. Representative PDB IDs with inter-atomic C-halogen…pi interaction.

PDB ID	Type of interaction[a]	Number of hits	Nucleic acid based target	Halogenated ligand
5EIX	C-F…pi	4	Topoisomerase IV	Quinoline
5BJP	C-F…pi	1	Corn RNA aptamer	DFHO
5BJO	C-F…pi	1	Corn RNA aptamer	DFHO
4Z3O	C-F…pi	1	Topoisomerase IV	Moxifloxacin
4Z2D	C-F…pi	3	Topoisomerase IV	Levofloxacin
4Z2C	C-F…pi	1	Gyrase	Moxifloxacin
3RAD	C-F…pi	1	Topoisomerase IV	Clinafloxacin
3RAE	C-F…pi	2	Topoisomerase IV	Levofloxacin
3K9F	C-F…pi	1	Topoisomerase IIA	Quinoline
3FOE	C-F…pi	1	Topoisomerase IIA	Quinoline
4ZTJ	C-Cl…pi	1	HIV-1 Integrase	2-Pyridinone
4ZTF	C-Cl…pi	1	HIV-1 Integrase	2-Pyridinone

[a]The pi region is present on the nucleic acids and the C-X is on the ligand.

interaction with halogenated ligands. This points to the relevance of C-X…pi interactions in increasing binding affinity of the ligand to the target and helps us infer that a similar approach with halogenated ligands for nucleic acid based targets can be utilized in case of malaria, leishmanial, filariasis, ESKAPE pathogens and even in HIV. 12 PDB-IDs representing nucleic acid targets with attached halogenated ligands can be studied, which include fluorinated and chlorinated ligands. Table 6 represents the presence of inter-atomic C-halogen…pi interactions (in Table 5) across several PDB IDs and points to their use in design of halogenated ligands for efficient binding with the nucleic acid associated target of choice be it in malarial, leishmanial parasites, resistant microbes and also in Human Immunodeficiency Virus (HIV). The halogenated ligands under study are mainly chlorine and fluorine indicating their use and incorporation in ligand design (Table 6).

4. Conclusions

Halogenation has been used considerably in the past decade which effectively acts as inhibitors or agonists against several identified drug targets and also in protein engineering to increase functional capacities of the enzyme concerned.[2,10] As more and more DNA–protein, DNA-DNA/RNA and other hybrid interactions are studied and corresponding sites against which inhibitors

or drugs are designed, it becomes necessary to ensure that these sites are properly analyzed and ligands are designed effectively to ensure that the binding to the target site occurs spontaneously and also involves stability.[11] Therefore, the analysis of these interactions becomes predominant as they will indicate the form of halogenation to be introduced in order to correspondingly bind to the identified target or the site to be exploited in DNA/RNA-protein interactions.

This database survey summarizes the presence of C-X…pi interactions present amongst the halogenated residues in nucleic acids. The number of hits although insufficient for most of the halogens does necessarily imply as to the prevalence of these non-conventional interactions and their role in ensuring ligand binding and subsequent recognition amongst nucleic acids. The presence of such interactions will influence molecular modeling and dynamics studies to include such conformational changes introduced by the presence of halogens and halogenated biomolecules or compounds in nucleic acid–protein or ligand interactions. The utilization of halogenated ligands to conclusively improve binding to the nucleic acid based target is implied in this database study and points to the role of non-covalent interaction like C-halogen…pi in conferring increased affinity for the target site by the ligand and a stable interaction.

The analysis of PDB IDs leads us to at least some of the listed C-halogen…pi interactions between fluorinated and chlorinated ligands and nucleic acids (DNA) in the case of malaria and other neglected tropical diseases wherein, existing targets associated with DNA with bound ligands show the presence of interatomic C-halogen…pi interactions.

References

1. Sutor D J 1963 Evidence for the existence of C–H··· O hydrogen bonds in crystals *J. Chem. Soc. (Resumed)* **204** 1105
2. Lu Y, Wang Y and Zhu W 2010 Nonbonding interactions of organic halogens in biological systems: implications for drug discovery and biomolecular design *Phys. Chem. Chem. Phys.* **12** 4543
3. Prasanna M and Row T G 2000 C–halogen··· π interactions and their influence on molecular conformation and crystal packing: a database study *Cryst. Eng.* **3** 135
4. Prasanna M and Row T G 2001 Weak interactions involving organic fluorine: analysis of structural motifs in Flunazirine and Haloperidol *J. Mol. Struct.* **562** 55
5. Auffinger P, Hays F A, Westhof E and Ho P S 2004 Halogen bonds in biological molecules *Proc. Natl Acad. Sci.* **101** 16789
6. Vallejos M, Auffinger P and Ho P 2006 Halogen interactions in biomolecular crystal structures *Int. Tables Crystallogr.* Vol. F, Ch. 23.6, pp. 821–826
7. Desiraju, G R 1991 The CH… O hydrogen-bond in crystals-what is it *Acc. Chem. Res.* **24** 290
8. Berman H, Henrick K, Nakamura H and Markley J L 2006 The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data *Nucl. Acids Res.* **35(suppl_1)** D301
9. Saraogi I, Vijay V, Das S, Sekar K and Row T G 2003 C–halogen… π interactions in proteins: a database study *Cryst. Eng.* **6** 69
10. Robinson-Rechavi M, Alibés A and Godzik A 2006 Contribution of electrostatic interactions, compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritima *J. Mol. Biol.* **356** 547
11. Umezawa Y, Tsuboyama S, Honda K, Uzawa J and Nishio M 1998 CH/π interaction in the crystal structure of organic compounds. A database study *Bull. Chem. Soc. Jpn.* **71** 1207