

Sandhya Kortagere and William J. Welsh, Dept. of Pharmacology, UMDNJ-RWJMS, 675, Hoes lane, Piscataway, NJ 08904. kortagsa@umdnj.edu

Sandhya Kortagere and William J. Welsh, Dept. of Pharmacology, UMDNJ-RWJMS, 675, Hoes lane, Piscataway, NJ 08904. kortagsa@umdnj.edu

A **toxiphore** is a feature or group within a chemical structure that is thought to be responsible for the toxic properties. A toxic substance exerts its toxicity through interaction with a biomolecule, such as a protein or DNA. This causes changes in the normal physiological state eliciting toxic effects. Sometimes the toxiphore requires bioactivation (modification by an enzyme), to produce a more reactive chemical species, that is able to covalently bind to cellular macromolecules. Generally, different chemical compounds that contain the same toxiphore elicit similar toxic effects within the same organ system or area of the body. Since the toxic properties of a compound are related to its chemical structures, a thorough molecular knowledge of toxiphore interactions with biomolecules can help predict the toxicity of a given compound. This study is designed to understand the structural and molecular basis of toxiphore interactions with biomolecules. A database of all these biomolecular interactions with various toxiphores will serve as a guide to efficiently design future drugs and pesticides and hence limit the release of harmful pollutants. A case study of the interactions of alkyl and aryl halides that are extensively used as pesticides, disinfectants and components of most drugs is presented. Although many halogenated compounds are very beneficial, they can also be harmful causing carcinogenic and mutagenic toxicity in the body. They are also known to cause genotoxicity based on the nature, number and position of the halogen substituent. Figure:1 describes a few examples of halogenated interactions with proteins that are known to be toxic in nature.

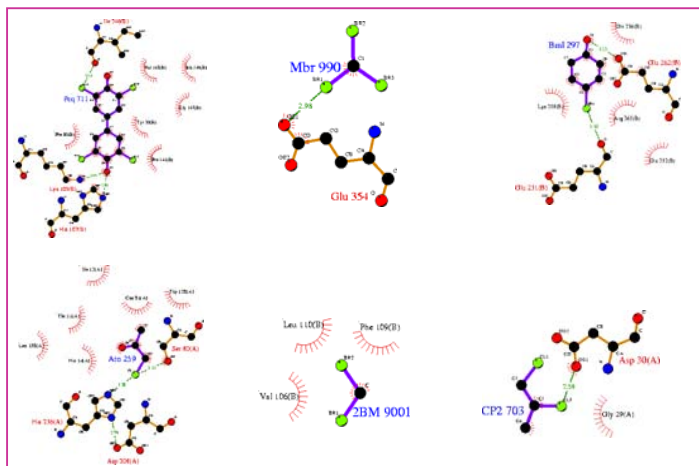


Figure-1: Schematic representation of the interaction of the toxicophores Alkyl halides and Aryl halides in biological systems. a) Polychlorinated biphenyl is bound to the estrogen binding site of human estrogen sulfotransferase b) interaction of bromoform with firefly luciferase enzyme c) 4-bromophenol bound to toluene/o-xylene monooxygenase hydroxylase d) interaction of chloroacetone with hydroxynitrile lyase that has implications in cyanogenesis e) dibromomethane interaction with methane monooxygenase hydroxylase f) interaction of dichloro propane with LINB – a haloalkane dehalogenase enzyme.

- All halogenated ligands co-crystallized with proteins were chosen from PDB.
- Only those structures that had a resolution of 3 Å or better resolution were considered, resulting in a set of 680 structures.
- All interactions between the halogenated atom of the ligand and any of the amino acid side-chain atoms less than 5 Å in distance were included. Only the closest distance to an atom of an amino acid side chain was taken into account as an interaction.
- Since the interactions are likely to be influenced by the moiety to which the halogen is attached (e.g. alkyl or aryl groups), we classified the halogenated ligands into Alkyl-X and Aryl-X groups, where X is any halogen F, Cl, Br or I (see Figure-2).
- Ligands in which the halogen atom was covalently attached to any atoms other than C, O, N, S were not considered in this dataset, excluding interactions with molecules such as Beryllium Trifluoride or other metallic halides from the data set.
- The propensity of interaction of a halogen X with 19 amino acids in the entire data set was recorded as the observed frequency (OBS-Freq). Also, the expected frequency (EXP-Freq) for each such interacting pair was determined from a set of N(X,X) interactions and the results were normalized to get the Nor.Freq as described below.

$$P(A, X) = \frac{N(A, X)}{N(X, X)} \quad P(B, X) = \frac{N(B, X)}{N(X, X)}$$

$$Exp.Freq = P(A, X) * P(B, X) * N(X, X)$$

$$Nor.Freq = \frac{X_i}{\sum_{i=19} X_i} * 100 \quad C = 1.96 * \frac{\sigma}{\sqrt{n}}$$

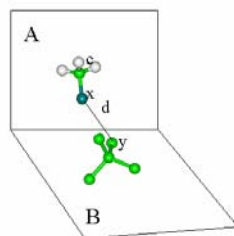


Figure-2: Schematic representation of halogenated ligand interaction. x represents the halogen atom, where x = F, Cl, Br or I. d is the distance of interaction, θ is the angle formed by atoms c, x and y and ϕ is the interplanar angle between planes A and B.

- A total of 6854 interactions between the 19 amino acids and various halogenated ligands were found, of which 3565 interactions were with fluorinated, 899 with chlorinated, 1259 with brominated and 1131 with iodinated ligands.
- We found 1926 interactions involving Alkyl-F, and 1446 interactions with Aryl-F groups. Similarly, 400 interactions were found for Alkyl-Cl, 383 for Aryl-Cl, 690 for Alkyl-Br, 476 for Aryl-Br, 490 for Alkyl-I and 577 for Aryl-I groups.
- Fluorine and chlorine substitutions were found to have the highest propensities for interactions with Leu, while bromine interacts most often with Phe, and iodine with Ser. Fluorine has the lowest interaction preference for Cys, chlorine for Gln, bromine for Asp and iodine for Pro. Among the Alkyl-X groups interacting with alkyl amino acids, the highest propensity was observed for Ser and Leu, while the lowest preference was observed for the sulphur containing amino acids Cys or Met and the nitrogen containing Asp or Lys. For the Aryl-X ligands interacting with aromatic amino acids the highest preference of interaction was observed for Phe.

AA	F			C			CL			CL			CL			BR			BR			BR			I			I			I		
	ALK	ALK	ALK	ARK	ARK	ARK	ALK	ALK	ALK	ARK	ARK	ARK	ALK	ALK	ALK	ARK	ARK	ARK	ALK	ALK	ALK	ARK	ARK	ARK	ALK	ALK	ALK	ARK	ARK	ARK			
OB1	NOE	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP			
ALA	121	506	1346	36	307	6612	18	1653	963	62	852	410	19	512	2030	36	405	5081	1	121	929	92	108	108	108	108	108	108	108	108	108		
ASN	871	1757	3505	309	503	6629	14	819	966	25	343	411	23	620	2085	32	360	5014	2	121	932	92	108	108	108	108	108	108	108	108	108		
ASN	871	1757	3505	309	503	6629	14	819	966	25	343	411	23	620	2085	32	360	5014	2	121	932	92	108	108	108	108	108	108	108	108	108	108	
ASP	203	840	1172	53	451	3584	10	585	813	44	647	3463	3	1081	1765	7	079	4242	0	545	7385	7	072	166	166	166	166	166	166	166	166	166	
CYS	51	213	431	61	1094	2141	0	1058	3312	12	1673	238	4	108	677	11	124	1620	19	1152	301	7	108	108	108	108	108	108	108	108	108	108	
GLN	304	856	3710	17	145	308	0	858	444	6	087	1891	17	458	963	8	060	7506	0	485	439	17	176	176	176	176	176	176	176	176	176	176	
GLU	57	238	7814	100	452	3837	7	409	559	9	124	2379	16	431	1231	16	108	2902	2	121	939	17	176	176	176	176	176	176	176	176	176	176	
GLY	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	
ILE	142	408	10396	52	1435	1194	0	858	745	41	531	3165	15	108	1813	29	225	3861	2	121	937	14	108	108	108	108	108	108	108	108	108	108	108
ILE	1224	397	28011	176	4493	13904	26	1120	2038	98	13466	878	53	14294	4242	44	1055	85	0	485	1967	18	108	108	108	108	108	108	108	108	108	108	108
LYS	172	319	110603	33	281	5447	3	175	793	14	165	3378	10	270	1251	15	169	4120	0	100	766	7	108	108	108	108	108	108	108	108	108	108	108
MEU	28	117	6384	49	417	3135	4	234	457	31	426	1944	6	162	991	19	214	2371	4	242	441	17	108	108	108	108	108	108	108	108	108	108	108
PHI	165	600	30768	79	673	15017	16	936	22801	102	1401	9358	73	1640	4754	339	181	11427	3	182	2135	10	108	108	108	108	108	108	108	108	108	108	108
PRO	59	247	548	24	204	2762	6	334	3389	13	453	1657	2	054	844	25	282	2021	2	121	9376	7	108	108	108	108	108	108	108	108	108	108	108
SER	266	1103	17407	93	7592	8547	15	877	12425	29	396	5305	12	061	2701	21	236	6465	36	2103	1021	36	373	373	373	373	373	373	373	373	373	373	
THR	212	887	1442	445	232	898	15	173	1033	26	337	4397	17	458	2421	5	574	5344	5	301	997	7	108	108	108	108	108	108	108	108	108	108	108
THR	247	9768	73	622	456	4	234	669	9	257	2974	12	323	1536	7	800	3268	18	1059	674	17	176	176	176	176	176	176	176	176	176	176	176	
TRP	90	42	19536	36	733	7837	19	458	1163	106	1420	4020	33	849	557	628	620	38	127	1577	9	108	108	108	108	108	108	108	108	108	108	108	

Table-1: Propensity of interaction between the various classes of ligands and the amino acids is shown. OBS-Freq represents the observed propensity, NOR-freq represents the normalized propensity and EXP-freq represents the expected propensity.

The highest propensity is colored red and lowest blue.

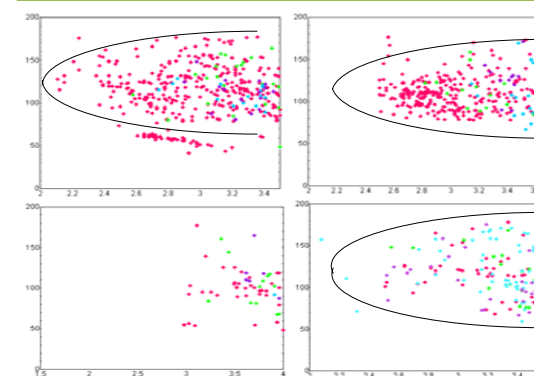


Figure-2: Illustration of the interactions of halogens with hydrogen bonding partners, X-axis represents distance and the Y-axis represents the corresponding angle A) with Oxygen B) with Nitrogen C) with Sulphur D) with Oxygen of water.

Fluorine is represented as red diamonds, Chlorine as green diamonds, Bromine as blue diamonds and Iodine as Purple diamonds.

The results illustrate the fact that bromine and iodine are better leaving groups.

- Although the introduction of halogen atoms to potential drug compounds improves their affinity and lipophilicity, they can also be potentially toxic due to the number of halogens, nature of substitution and nature of interaction with other functional groups. Thus by understanding the nature and propensity of interactions of halogens with various amino acids, it is feasible to design drugs, pesticides and disinfectants so as to minimize these toxic effects.
- The present study shows the various structural parameters such as distance, angle and nature of interaction of two classes of toxicophores namely alkyl halides and aryl halides. These structural details serve as a guidelines for future design of molecules with reduced toxicity using halogens.
- Future goals for the project is to provide a complete database of molecular interactions of all known toxicophores with biomolecules.

Support for this work has been provided primarily by the USEPA-funded Environmental Bioinformatics and Computational Toxicology Center (ebCTC), under STAR Grant number GAD R 832721-010.