Computational Biology and Chemistry 47 (2013) 231–239

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

Halogen bonding in complexes of proteins and non-natural amino acids

Sunčica Z. Borozan [a], Srđan Đ. Stojanović [b,*]

[a] Department of Chemistry, Faculty of Veterinary Medicine, University of Belgrade, Belgrade, Serbia
[b] ICTM – Department of Chemistry, University of Belgrade, Belgrade, Serbia

ARTICLE INFO

Article history:
Received 3 September 2013
Received in revised form 10 October 2013
Accepted 10 October 2013

Keywords:
Halogen bonding
Proteins
Non-natural amino acids
Complexes
Stabilization centers

ABSTRACT

In this work, we have analyzed the influence of halogen bonding to the stability of 44 complexes of proteins and non-natural amino acids. Fluorine- and chlorine-containing non-natural amino acids are more prevalent in the dataset, and an even larger number of contacts made by iodine-containing ligands are found. Only few halogen bonds with the hydroxyl oxygens and carboxylate side chains are found in the dataset. Halogen bonds with the nitrogen-containing side chains have higher occurrence than other acceptors. Backbone carbonyl oxygens and nitrogens are to a substantial extent involved in our dataset. We have observed a small percentage of interactions involving water as hydrogen bond donors. Additionally, most of the interacting residues comprising the interfaces also show a great degree of conservation. There is a clear interaction hot spot at distances of 3.5–3.7 Å and Θ_1 angles of 100–120°. There is also a cluster of contacts featuring short distances (2.6–2.9 Å) but only nearly optimal Θ_1 angles (140–160°). 51.3% of stabilizing residues are involved in building halogen bonds with the non-natural amino acids. We discovered three types of structural motifs significantly over-represented: beta-turn-ir, beta-turn-il and niche-4r. The halogen-bonding statistics of the dataset do not show any preference for α-helices (36%), β-sheets (36%), or turns/coils (28%) structures. Most of the amino acid residues that were involved in halogen bonds prefer to be in the solvent excluded environment (buried). Furthermore, we have shown that in amino acid–protein complexes halogen atoms can sometimes be involved in hydrogen bonding interactions with hydrogen bonding-donors. The results from this study might be used for the rational design of halogenated ligands as inhibitors and drugs, and in biomolecular engineering.

1. Introduction

Supramolecular assemblies, in both the solid state and in solution, rely on a range of weak, noncovalent interactions to hold the constituent molecules together (Anbarasu et al., 2009; Arzhanik et al., 2010; Bissantz et al., 2010; Panigrahi and Desiraju, 2007; Petkau-Milroy and Brunsveld, 2013; Reddy et al., 2008). The most common of these are hydrogen bonding interactions (Steiner, 2002), which are relatively strong and highly directional, and thus they have highly predictable geometries and fine control of the assembly process can be achieved. An alternative interaction that has more recently been exploited as a construction tool in the field of supramolecular chemistry and crystal engineering is the halogen bond (Beale et al., 2013; Erdelyi, 2012; Riley and Hobza, 2011; Wilcken et al., 2013). Halogen atoms usually form a single, covalent bond with one other atom. As a result of this, halogen atoms are typically found on the periphery of a molecule, making them ideally placed for taking part in intermolecular interactions. Despite the fact that halogens have higher electronegativity than carbon, which creates a negative partial charge on halogens in organic molecules, halogens favorably interact with a Lewis base atom, such as oxygen or nitrogen with a lone electron pair (Fig. 1). This counterintuitive attraction is commonly explained by the existence of a region of positive electrostatic potential (ESP), located on top of the halogen atom (Clark et al., 2007). This region, usually referred to as the σ-hole, is an inherent feature of compounds containing covalently bound halogens; i.e., it is not induced by the interacting partner in a complex. The interaction is typically linear at the (organic) halogen, consistent with maximizing the two main directional attractive contributions to the interaction energy, electrostatics and charge transfer, and minimizing the exchange repulsion, which is also directional (Huber et al., 2013). Its bond strength is similar to that of the hydrogen bond, in the range 1–40 kcal/mol. Its strength increases in the order fluorine < chlorine < bromine < iodine (Clark et al., 2007).

Numerous computational and experimental studies have been performed to better understand halogen bonding, and in recent years, there has been an increasing interest in studying halogen

* Corresponding author. Tel.: +381 11636061.
E-mail address: srdjanst@chem.bg.ac.rs (S.Đ. Stojanović).

1476-9271/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiolchem.2013.10.002

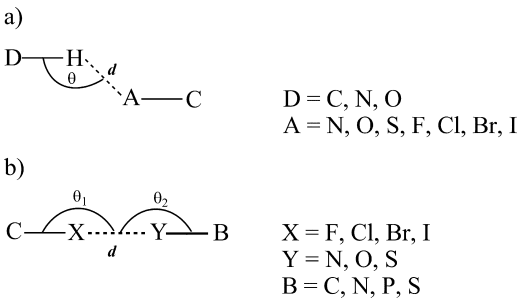

Fig. 1. Parameters for (a) hydrogen bonds and (b) halogen bonds.

bonding in biological systems and drug discovery (Wilcken et al., 2013 and references therein). A large number of pharmaceutical ligands have halogens incorporated into their structures. There are several reasons for this, the most important of which are that halogens tend to increase the membrane permeability of small molecules while also increasing their catabolic stability (Parisini et al., 2011). Protein data bank survey studies done in the past several years have identified many halogen bonds that occur within protein ligand complexes (Auffinger et al., 2004; Lu et al., 2010a). One of the principle findings in these surveys is that most of the halogen bonds in these complexes involved halogens bound to aromatic groups. Halogen bonds involving aromatically bound halogens are generally stronger than those of aliphatic halogens because aromatic moieties have electron withdrawing properties that lead to larger σ-holes. Halogen bonding has found applications in a number of fields including liquid crystals (Prasang et al., 2008), conducting materials (Yamamoto et al., 2008), and structural biology (Auffinger et al., 2004) as well as via supramolecular chemistry in anion recognition (Kilah et al., 2010) and polymerization (Sun et al., 2006).

Amino acids form the building blocks of all proteins. Naturally occurring amino acids are restricted to a few tens of sidechains, even when considering post-translational modifications and rare amino acids such as selenocysteine and pyrrolysine. However, the potential chemical diversity of amino acid sidechains is nearly infinite. Exploiting this diversity by using non-natural sidechains to expand the building blocks of proteins and peptides has recently found widespread applications in biochemistry, protein engineering and drug design. For instance, in a recent study, it was shown that non-natural sidechains dramatically increase the affinity of amyloid fiber inhibitors (Sievers et al., 2011). Similarly, several cyclic and other kinds of modified peptides with non-natural amino acid sidechains have been developed for therapeutics use, such as cilengitide (Burke et al., 2002) or crafilzomib (Parlati et al., 2009). Non-natural sidechains have also been used as independent ligands. For instance, L-3,4-dihydroxyphenylalanine is a non-natural amino acid used in Parkinson disease treatment (Smith et al., 2012), and 5-hydroxytryptophan (oxitriptan) has been used as an antidepressant (Turner and Blackwell, 2005). In addition to therapeutics use, non-natural sidechains have found many other applications in biochemistry and protein studies (Wang et al., 2009). These include photo-crosslinking amino acids to probe in vivo protein interactions (Ai et al., 2011; Kessler et al., 1999), fluorescent amino acids used as markers of specific proteins (Wang et al., 2006) or phosphorylated amino acid mimetics to probe the effect of post-translational modifications (Lemke et al., 2007).

In this work, we have analyzed the influence of halogen bonds in complexes of proteins and halogen-containing non-natural amino acids. We have focused our study at the interface between proteins and halogen-containing non-natural amino acids and hence the halogen bonds within a protein are not considered. The quantification of such bonds in binary complexes, as well as its in silico prediction, can provide not only insights into the mechanism by which the ligand achieves specificity but also a greater understanding of the binding pocket that will enable more selective drug candidates to be discovered.

2. Methodology

2.1. Dataset

For this study we used the molecular and structural database of non-natural sidechains (SwissSidechain) (Gfeller et al., 2013). The SwissSidechain database contains molecular and structural data for 210 non-natural alpha amino acid sidechains, both in L- and D-configurations, in addition to the 20 natural ones. These amino acids have been selected based on two main criteria: first, the presence of non-natural sidechains in publicly available protein structures in the PDB (Protein Data Bank) (Rose et al., 2011). We used the following procedure to select the desired complex structures for further analysis: (1) no theoretical model structures and no NMR structures were accepted; these structures were not included since it was difficult to define the accuracy of the ensemble of structures in terms of displacement that was directly comparable to the X-ray diffraction studies. (2) Only crystal structures with the resolution of 3.0 Å or better and a crystallographic R-factor of 25.0% or lower were accepted, and (3) halogen-containing amino acid–protein complexes. To avoid redundancy in our data set, a standalone PISCES package was used to select a single structure with the best resolution in cases where proteins in different complexes had >30% sequence identity (with all other options set to their default) (Wang and Dunbrack, 2005). Using these criteria, we created a dataset of 44 protein–halogen containing amino acid complexes. The PDB IDs are as follows: 1C0L, 1CF0, 1CZI, 1GA1, 1NLU, 1OKW, 1OL2, 1ORW, 1PFV, 1PN3, 1RRV, 1TF9, 1TZM, 1WQ3, 2AG6, 2AKW, 2AR8, 2AXI, 2C5V, 2GV2, 2NW9, 2UUE, 2V7L, 2WDX, 2WHB, 2X1N, 2X68, 2XAD, 2ZP1, 3D39, 3D3V, 3F3C, 3FEA, 3GFD, 3GH8, 3KTJ, 3MG9, 3Q4K, 3Q9I, 3RUL, 3TNZ, 3VFJ, 4EAR, and 4EEC.

2.2. Hydrogen and halogen bond analysis

If not already present, all hydrogen atoms were added using the program Discovery Studio Visualizer 3.5 (Accelrys Software Inc., 2012). The H-atom positions were then refined, keeping the position of the non-H atoms fixed, using the MMFF94 force field (Halgren, 1996). When multiple alternative conformations of certain residues were present, as indicated by the altLoc field in the PDB file, only the first conformation was considered. All the optimized structures were exported to the hydrogen bond analysis tool (HBAT) for calculation of halogen and hydrogen bonds and their properties (Tiwari and Panigrahi, 2007) with default settings. The positions and geometry of donor and acceptor atoms are shown in Fig. 1. The used criteria for hydrogen bonds are d (H···A) ≤ 3.0 Å and Θ (D—H···A) ≥ 90°. Parameters are d, distance between the H atom and the A (acceptor, A = N, O, S) atom; Θ, defined as the angle between the D—H bond and the center of the acceptor atom. The geometry of the halogen bond is defined by the d (X···Y) distance (≤3.7 Å), the Θ_1 angle (≥90°) of the Y atom relative to the C—X bond, and the Θ_2 angle (≥90°) of the halogen relative to the Y—B bond.

The program HBAT was used for statistical analysis, providing distance–angle distributions as well as furcation and halogen–water interactions.

2.3. Computation of conservation of amino acid residues

We computed the conservation of amino acid residues in each protein using the ConSurf server (DeLano, 2002). This server computes the conservation based on the comparison of the sequence of