

Structural bioinformatics

Census of halide-binding sites in protein structures

Rostislav K. Skitchenko ¹, Dmitrii Usoltsev¹, Mayya Uspenskaya¹,
Andrey V. Kajava^{1,2,*} and Albert Guskov ^{3,*}

¹Institute BioEngineering, ITMO University, Saint-Petersburg 197101, Russia, ²Centre de Recherche en Biologie cellulaire de Montpellier (CRBM), UMR 5237 CNRS, Université Montpellier, Montpellier 34293, France and ³Groningen Biomolecular Sciences & Biotechnology Institute, University of Groningen, Groningen 9747 AG, the Netherlands

*To whom correspondence should be addressed.

Associate Editor: Arne Elofsson

Received on October 24, 2019; revised on January 24, 2020; editorial decision on January 28, 2020; accepted on January 29, 2020

Abstract

Motivation: Halides are negatively charged ions of halogens, forming fluorides (F⁻), chlorides (Cl⁻), bromides (Br⁻) and iodides (I⁻). These anions are quite reactive and interact both specifically and non-specifically with proteins. Despite their ubiquitous presence and important roles in protein function, little is known about the preferences of halides binding to proteins. To address this problem, we performed the analysis of halide–protein interactions, based on the entries in the Protein Data Bank.

Results: We have compiled a pipeline for the quick analysis of halide-binding sites in proteins using the available software. Our analysis revealed that all of halides are strongly attracted by the guanidinium moiety of arginine side chains, however, there are also certain preferences among halides for other partners. Furthermore, there is a certain preference for coordination numbers in the binding sites, with a correlation between coordination numbers and amino acid composition. This pipeline can be used as a tool for the analysis of specific halide–protein interactions and assist phasing experiments relying on halides as anomalous scatters.

Availability and implementation: All data described in this article can be reproduced via compiled pipeline published at https://github.com/rostkick/Halide_sites/blob/master/README.md.

Contact: andrey.kajava@crbm.cnrs.fr or a.guskov@rug.nl

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Halides are the common anionic forms of halogens, in which the latter interact with the less electronegative atoms, thus forming fluorides (F⁻), chlorides (Cl⁻), bromides (Br⁻) and iodides (I⁻). In this work, we are not considering astatides (At⁻), due to its unstable (radioactive) nature and not known biological functions. In the biological context, halides are often considered merely as the components of buffers to maintain certain ionic strength of a solution, this is especially true for Cl⁻ as it is the most common counterion. Iodides owing to its electron-rich configuration [Kr] 4d¹⁰5s²5p⁶ have found an interesting niche in the macromolecular X-ray crystallography as a phasing agent to solve the phase problem (*vide infra*). Fluoride seems to be a rather exotic anion, for example currently only 67 entries contain fluoride as a ligand in the Protein Data Bank (PDB) (i.e. <0.05% entries), and fluorides are typically rather incorporated into the scaffold of other ligands, i.e. drugs or inhibitors to fine-tune their properties. Bromide seems to have a few highly specialized functions (discussed below) and similarly to iodide can be used for phasing.

For the purpose of this work on characterization of biologically relevant halide-binding sites, we focus only on ionic forms of halides

and not on covalent ones. The former are characterized by –1 charge and large radii: 1.19, 1.67, 1.82 and 2.06 Å for F⁻, Cl⁻, Br⁻ and I⁻, respectively (Shannon, 1976).

1.1 Fluoride

The biological role of fluoride is typically reviewed in the context of dentistry, as it has been shown as an excellent anti-caries agent (Aoun *et al.*, 2018; Wong *et al.*, 2011). It has a dual role, as it reinforces enamel via slowdown of demineralization and promotion of remineralization via formation of fluorohydroxyapatite (Fincham *et al.*, 1999; Kanduti *et al.*, 2016), but also via suppression of mouth bacteria by inhibiting certain intracellular enzymes, leading to the decreased production of lactic acid and hence lowering the risk of caries formation (Buzalaf *et al.*, 2011). Clearly, taking into account such an inhibitory function, the valid question is the safety of fluoride for humans, especially since in many countries tap water and table salt come fluorinated. According to the current research and guidelines, there is no actual risk for humans under normal circumstances, apart for those who live in the areas, where water is naturally enriched in fluoride—some parts of India, China and Africa continent. The prolonged exposure to the excess of fluoride might

lead to the disorders, such as skeletal fluorosis, dental fluorosis and kidney failure (Aoba and Fejerskov, 2002; Perumal *et al.*, 2013; Weatherell, 1969).

However, the environmental levels of F^- of 10–100 μM found in soil and water are toxic for many organisms, hence several transport systems to expel fluoride from cells have evolved. One of them is Fluc proteins, which can be found in prokaryotes and lower eukaryotes (Stockbridge *et al.*, 2013). Flucs are anion channels, which are highly specific to F^- over Cl^- (Stockbridge *et al.*, 2015). The other major class of F^- transporting proteins is a subfamily of CLC chloride transporters, strictly confined to bacterial species only, so-called CLCF F^-/H^+ antiporters, which couple F^- efflux to a proton gradient (Brammer *et al.*, 2014). The selectivity of these transporters for F^- over Cl^- is also extremely high (Stockbridge *et al.*, 2012).

Clearly the exporters cannot immediately remove the excess of incoming F^- flux, hence it readily interacts with intracellular proteins. Among the best characterized targets to which F^- binds are F^- -specific riboswitch, where F^- is complexed with Mg^{2+} and RNA phosphate groups (Ren *et al.*, 2012); enolase (Qin *et al.*, 2006), which is an essential enzyme for glycolysis; heme containing proteins and numerous phosphatases (Marquis *et al.*, 2003). Fluorides readily react with aluminum and beryllium, and the formed complexes are highly cytotoxic as they mimic phosphate group hence inhibiting numerous enzymes, which exert ATPase or GTPase activity (Li, 2016). In eukaryotic cells, the deleterious effect of exposure to fluoride is even more dramatic, as fluoride interferes with the cell cycle, respiration, gene expression, oxidative stress and G protein activation (Mendoza-Schulz *et al.*, 2009; Murthy and Makhlof, 1994; Strunecka *et al.*, 2007).

It is important to note that in many cases the inhibition effect might be caused not via binding of fluorides to proteins, but also as an effect of acidification of cytoplasm, as F^- might play a role of transmembrane proton carrier (Barbier *et al.*, 2010; Eisenberg and Marquis, 1981).

The inhibitory effect of fluoride complexes with aluminum and beryllium was quickly recognized as an invaluable tool for structural studies as it allows stabilization of intermediate states. Using such complexes, numerous enzymes as well as several membrane transporters were trapped in specific conformations, e.g. maltose uptake transporter MalFGK2 (Oldham and Chen, 2011) and calcium ATPase (Toyoshima *et al.*, 2004).

1.2 Chloride

Chloride is a ubiquitous anion both in the environment and in cells, and it plays important roles in all kingdoms of life. Its most common role is as a counterion for sodium and potassium, in combination with which they form an electrolyte mix essential to maintain the concentration and charge differences across cell membranes. The second most important role of chloride (especially in certain bacteria and archaea) is as an osmolyte during osmoadaptation (Saum and Müller, 2008). In a high saline milieu, halophilic and halotolerant microorganisms accumulate up to molar concentrations of chloride (Lanyi, 1974; Oren, 1986). Interestingly some of these organisms (e.g. *Halobacillus halophilus*) are strictly dependent on chloride for their growth, as Cl^- is directly involved in regulation of transcription and translation of several essential proteins (Hänelt and Müller, 2013; Roessler and Müller, 2001; Roessler *et al.*, 2000). Interestingly, some bacteria which are not normally halotolerant, are capable of growth in the high salt medium, but only if the counterion is a chloride, implying the involvement of chloride in the osmoadaptation also in these species and/or regulation of sodium export (Galinski, 1995). In higher organisms, Cl^- also plays important roles. For example, in plants chloride is required for turgor generation and the regulation of cell volume, as well as for generation of Cl^- currents (Wege *et al.*, 2017). In photosynthetic organisms (hence also including cyanobacteria and algae), chloride ion plays an essential role in photosystem II function, namely it facilitates the proton flux from the oxygen evolving complex to the thylakoid lumen (Brahmachari *et al.*, 2017; Guskov *et al.*, 2009; Kawakami *et al.*, 2009). In mammals, chloride in the form of hydrochloric acid maintains the very acidic pH of gastric juice (pH 1.5–3.5) necessary

to unfold consumed proteins, to activate digestive enzymes and to kill microorganisms susceptible to such acidic environments (Freeman and Kim, 1978; Lu *et al.*, 2010). Furthermore, there are numerous human proteins, where specific binding sites for chloride were revealed and/or which are shown to be affected upon interaction with chloride. Among these α -amylase (Aghajari *et al.*, 2002), angiotensin-converting enzyme I (Yates *et al.*, 2014), hemoglobin (Prange *et al.*, 2001), kinases (Chen *et al.*, 2019), acute myeloid leukemia-1 transcription factor (Wolf-Watz *et al.*, 2001) and many others.

To regulate the flux of Cl^- , numerous chloride channels and transporters have evolved. The most studied chloride transporting proteins belong to the chloride channel (CLC) family (Jentsch and Pusch, 2018) and chloride intracellular ion channel (CLIC) family (Argenzio and Moolenaar, 2016).

CLC proteins are integral membrane proteins, residing either in plasma or intracellular membranes and encompassing both channels and transporters (Accardi, 2015; Accardi and Picollo, 2010). They are involved in the control of excitability during muscle contraction, acidification of endosomes and lysosomes, and epithelial transport (Jentsch, 2007; Pedersen *et al.*, 2016; Piwon *et al.*, 2000; Smith and Lippiat, 2010). Intriguingly all members form dimers with a separate translocation pathway.

Malfuncions of CLC proteins cause severe diseases, such as myotonia congenita (Imbrici *et al.*, 2015a, b), neuronal ceroid lipofuscinosis (Poët *et al.*, 2006; Yoshikawa *et al.*, 2002), Dent's disease (Devuyst and Thakker, 2010) and Bartter syndrome (Andrini *et al.*, 2015; Cunha and Heilberg, 2018).

CLIC proteins are quite unique, since they can exist both in soluble and membrane-embedded forms. They are not located in the plasma membrane but abundant in intracellular organelles (Gururaja Rao *et al.*, 2018). They are involved in signaling (Argenzio and Moolenaar, 2016), endosomal trafficking (Argenzio *et al.*, 2014), phagosomal acidification (Jiang *et al.*, 2012), angiogenesis (Ulmasov *et al.*, 2009), actin-dependent membrane remodeling (Berryman and Bretscher, 2000) and other intracellular processes (Littler *et al.*, 2010).

Another well-studied chloride channel is cystic fibrosis transmembrane conductance regulator (Hwang *et al.*, 2018; Liu *et al.*, 2017, 2019). It is an ATP-gated chloride channel evolved from ABC transporter scaffold. Mutations rendering this protein defunct lead to the increased viscosity of mucus on membranes (e.g. in the lungs) which can be lethal. There are several more families of chloride channels, such as calcium-activated chloride channels (Berg *et al.*, 2012), maxi Cl^- channels (Sabirov and Okada, 2009) and volume-regulated chloride channels (Osei-Owusu *et al.*, 2018), which are beyond the scope of this work.

1.3 Bromide

There is no solid evidence for a certain role of bromide in prokaryotes, although there is a large class of marine and soil microorganisms capable of oxidizing methyl bromides via transmethylation or monooxygenase pathway (Hoeft *et al.*, 2000). For the majority of microorganisms though, bromide is toxic at high concentrations, and in fact it is used as a disinfectant agent, typically in the form of hypobromous acid.

In eukaryotes, the role of bromide was for a long time rather elusive and only recently it has been established that it is essential for the assembly of collagen IV scaffolds during tissue development (McCall *et al.*, 2014). Furthermore, bromide is a preferred substrate for eosinophil peroxidases (Henderson *et al.*, 2001; Thomas *et al.*, 1995), which catalyze the conversion of bromide to hypobromous acid for the host defense.

Bromide is localized mainly extracellularly, and its concentration seems to be tightly regulated (Barratt and Walser, 1969). Bromide deficiency leads to diminished tissue growth and causes failures in tissue development and remodeling (McCall *et al.*, 2014). However, at excess, bromide can cause bromism—the collective name of several neurological disorders caused by the neurotoxic effect of prolonged consumption of bromide (Frances *et al.*, 2003; Lugassy and Nelson, 2009; van Leeuwen and Sangster, 1987).

Interestingly, some marine algae accumulate large amounts of bromide (and iodide) but in the cell wall and usually not in the cytosol (Küpper et al., 2018; Saenko et al., 1978). Their genomes encode a specific set of proteins to deal with halides, such as haloacid and haloalkane dehalogenases as well as vanadium haloperoxidase (Almeida et al., 2001; Kunka et al., 2018). The latter enzyme is responsible for production of methyl halides. The accumulation of bromide is probably a consequence of the relative abundance of this anion in seawater: the average concentration of Br⁻ is ~65 mg l⁻¹, whereas F⁻ and I⁻ are ~1 mg l⁻¹ (however, Cl⁻ is predominant with concentration of 19 000–23 000 mg l⁻¹). Surprisingly despite the concentration of Br⁻ is 300 times lower than that of Cl⁻, in many cases it is rather bromide than chloride (or their combination) which is used for organohalogen production (Cabrita et al., 2010).

In the world of structural biology, bromide has caught an eye due to its phasing potential—with 36 electrons and easily accessible X-ray absorption edge (K edge ~0.92 Å), it is a good choice for single or multi-wavelength anomalous diffraction (SAD/MAD) phasing (Dauter and Dauter, 2007). In the easiest application, a crystal should be shortly soaked in a cryoprotectant containing bromide just before the flash freezing in liquid nitrogen (Dauter et al., 2000). Bromide ions will quickly diffuse via the solvent channels and settle within the ordered solvent shell around the protein surface (Dauter and Dauter, 1999, 2001).

1.4 Iodide

Iodide is one of the largest monoatomic anions and one of the heaviest elements utilized by living organisms. In vertebrates, it is utilized for the production of growth-regulating thyroid hormones (thyroxine and triiodothyronine), which are essential regulators of virtually nearly all processes during different life phases (Bernal et al., 2015; Eales, 1997; Senese et al., 2014).

Uptake of iodide into thyroid occurs via the sodium iodide transporter (SLC5A5) (Darrouzet et al., 2014) residing in the basolateral membrane of thyroid follicular cells. The transport is active and the inward translocation of sodium down its electrochemical gradient is coupled to inward translocation of iodine against its electrochemical gradient.

Other organisms such as algae, zooplankton and plants are capable to accumulate iodine/iodide as thyroid hormone precursors, which can be used as developmental regulators. Many bacteria are capable of extracting necessary iodine/iodide from the host environment (DiStefano et al., 1993). Marine microorganisms are especially agile in iodide accumulation as they are capable to reduce inorganic iodate (the most thermodynamically stable form) to iodide and produce numerous iodinated organic compounds (Küpper et al., 2018). However, many other bacteria, both aerobic and anaerobic, are also able to convert iodate to iodide. It seems that even in the absence of highly specialized enzymes of thyroid gland, iodotyrosines can form spontaneously and due to its reactivity play a crucial role in cell–cell signaling (Cahnmann and Funakoshi, 1970; Crockford, 2009).

Therefore, a plausible scenario is that during evolution iodine/iodide reacting with tyrosines might had been recruited as a potent signaling molecule somewhat after last universal common ancestor (Crockford, 2009).

Some macroalgae (kelp) developed an extreme concentrating capacity for iodine, e.g. *Laminaria digitata* can concentrate up to 30 000 more of iodide in its apoplasts compared to iodide concentration in seawater (Küpper et al., 2011). Such accumulation leads to a buildup of antioxidant reservoir that is mobilized during oxidative stress. Iodide can scavenge not only H₂O₂ and ozone, but also hydroxyl radicals and superoxides (Küpper et al., 2008, 2018).

Iodine deficiency in humans is well documented and leads to the numerous mental and physical developmental delays (Zimmermann, 2011) and currently up to two billion people are affected worldwide according to World Health Organization. The iodized table salt turned out to be an excellent tool to compensate for insufficient uptake of iodine in affected populations. The large excess of iodine, however, can be toxic, especially in the case of selenium deficiency,

when the function of Se-containing antioxidative enzymes is impaired.

One can exploit anomalous signal of iodides bound to proteins the same way as it is done with bromide.

The X-ray absorption edge for iodide is not readily accessible (L–I edge ~2.39 Å), however, even far from it (at wavelengths of ~1.8 Å, which are accessible at modern synchrotrons), the anomalous signal is roughly three times higher than for bromide. The fast iodide soaking before cryo-freezing turned out to be successful phasing technique for numerous soluble proteins (Abendroth et al., 2011). Recently it has been proposed that iodide SAD phasing might be universally applied to membrane proteins (Melnikov et al., 2017) as their positively charged residues found at the hydrophobic–hydrophilic interface need a compensatory negative charge, hence increasing the odds of (ordered) binding of iodides at these areas.

Additionally, all four halides were shown to restore the oxidative redding (green to red photoconversion) in fluorescent proteins (Bogdanov et al., 2016), adding another possible application role in the fine tuning of properties in this important class of proteins.

Taking into account the importance of halides in the biochemistry of all life forms and possible applications we became intrigued whether there are certain patterns of halide binding to proteins—preferred amino acids involved in the binding sites and its geometry. We have analyzed all the PDB entries, containing halides, available on August 23, 2019, in the PDB and revealed the following patterns.

2 Materials and methods

2.1 Data acquisition and filtering

X-ray data analysis of protein structures containing fluoride, chloride, bromide and iodide, which are coordinated only by protein without ligands, were obtained from the PDB using Biopython (module Bio.PDB). List of entries was obtained with an advanced search (Search parameters: Chemical name—chloride/bromide/iodide/fluoride, Name—Equals, Polymeric type—Any). 67, 12 686, 455 and 864 entries were obtained for F⁻, Cl⁻, Br⁻ and I⁻, respectively. PDB structures obtained by NMR, powder diffraction, cryo-electron microscopy and neutron diffraction were excluded. For entries with the same name of proteins, those with the highest resolution were selected. Entries with a resolution of lower than 2 Å were excluded. The final non-redundant dataset includes 25, 3229, 206 and 246 structures with F⁻, Cl⁻, Br⁻ and I⁻, respectively. Analysis was performed with and without water molecules. If a PDB entry contained several identical halide sites (i.e. the case of homologous sites), only one site was taken and the rest were excluded from further study (the similarity threshold was arbitrary set at 0.5 Å rmsd). Sites containing non-protein atoms such as small ligands from HETATM or DNA/RNA from ATOM field were excluded. Sites consisting of several chains (~10% of all entries) have not been taken into account.

2.2 Calculations of distances, angles and accessible surface area

Distances and angles were calculated for each atom within a sphere with a radius of 5 Å around the halide using NumPy. Interactions of halides with carbon and hydrogen atoms were not considered. Interactions were examined both in the absence and in the presence of water molecules. In cases when the halide had more than one coordinating atom, the angles were calculated between two vectors: halide—the nearest atom and halide—atom. Fractional accessible surface area (fASA) of each halide atom was calculated with FreeSASA (Mitternacht, 2016) as a ratio between ASA of the halide within the protein structure and ASA of the sphere with the radius including halide radius (1.19, 1.67, 1.82 and 2.06 Å for F⁻, Cl⁻, Br⁻ and I⁻, respectively) and water molecule radius (1.4 Å). fASA allows to distinguish buried versus surface-bound halides.

2.3 Workflow reproducibility

Since the PDB is continuously expanding, the addition of new entries will make our analysis more robust. Snakemake pipeline was

constructed by using Python programming language to provide the ability to reproduce the results of this study. The pipeline covers all stages of the current study from downloading of halide-bound protein structures from the PDB to generation of the output files and graphical output in the course of several minutes. The output is tab-separated file with information about each halide atom in PDB structures. Additionally, there is an anaconda-environment file, which provides instructions for the installation of the dependencies required for the workflow. The pipeline is available at https://github.com/rostkick/Halide_sites/blob/master/README.md.

3 Results

The distribution of entries in PDB containing different halides is not equal, there are only 67 entries with fluoride and over 12 000 for chloride; for bromide and iodide, there are 455 and 864 entries. After applying strict selection criteria (see Section 2), the resulting working dataset contains 25, 3229, 206 and 246 entries for F^- , Cl^- , Br^- and I^- , respectively.

3.1 Bimodal distributions of distances between halides and atoms of amino acid residues

For all four halides, the median distance between an anion and docking residue is about 4.16 Å (Fig. 1A), with the smallest value of 4.09 Å for fluoride followed by 4.11 Å for chloride, 4.17 Å for bromide and the largest value of 4.28 Å for iodide reflecting difference in their ionic radii (*vide supra*). Such distances are indicative that the bound anions are either partially or completely dehydrated, as their hydrated radii are within 3.3–3.5 Å (Ayala *et al.*, 2003; Paula *et al.*, 1998).

However, if we look at the distribution of distances (Fig. 1A), the largest deviation is observed for chloride and iodide, with the distances going closer than 1.5 Å—these must be clear outliers as at such a short distance the repulsion of atoms is inevitable. The possible explanation of such a spread is that with the higher number of entries the odds of getting erroneous assignment is also higher; additionally, in the case of chloride since its anomalous signal is very low, it is hard to verify its assignment and it can be easily confused with water molecules. Fortunately, the number of such cases is low.

Surprisingly there is a large number of all four halides with the distances around 5 Å, those represent situations where the anions interact with atoms of residues via a water molecule. With that notion, we decided also to include water molecules in the analysis of interactions. It turns out that iodide tends to be less solvated (Supplementary Fig. S1A), with the preference for two water molecules in its network, whereas Br^- and Cl^- prefer three water molecules. Furthermore, the distribution of distances and angles in the presence of water molecules has not changed considerably, apart from fluoride (Supplementary Fig. S1B and C). These observations agree well with the Hoffmeister series on $\Delta G_{\text{hydration}}$ energies, where F^- is the most and I^- is the least strongly solvated anion (Fox *et al.*, 2015).

3.2 Angular distribution of coordinating atoms

The distribution of angles between the coordinating bonds is quite wide (Fig. 1B and Supplementary Fig. S1C), reflecting the possible positional errors (which can be of various origins and influenced by data quality and resolution). Nevertheless, Cl^- , Br^- and I^- have a noticeable maximum at 45 degrees and also less visible blurred maximum at around 90 degrees, indicative of bipyramidal or octahedral arrangement of binding sites. However, in the presence of water, the maximum is rather shifted toward larger angular values.

3.3 Typical compositions of halide-binding sites

The analysis of accessible surface area of bound halides revealed that in the majority of cases, the binding sites form pockets that surround the halides. At the same time, on average, 35% of the halide surface is accessible to the solvent. It means that the binding sites are

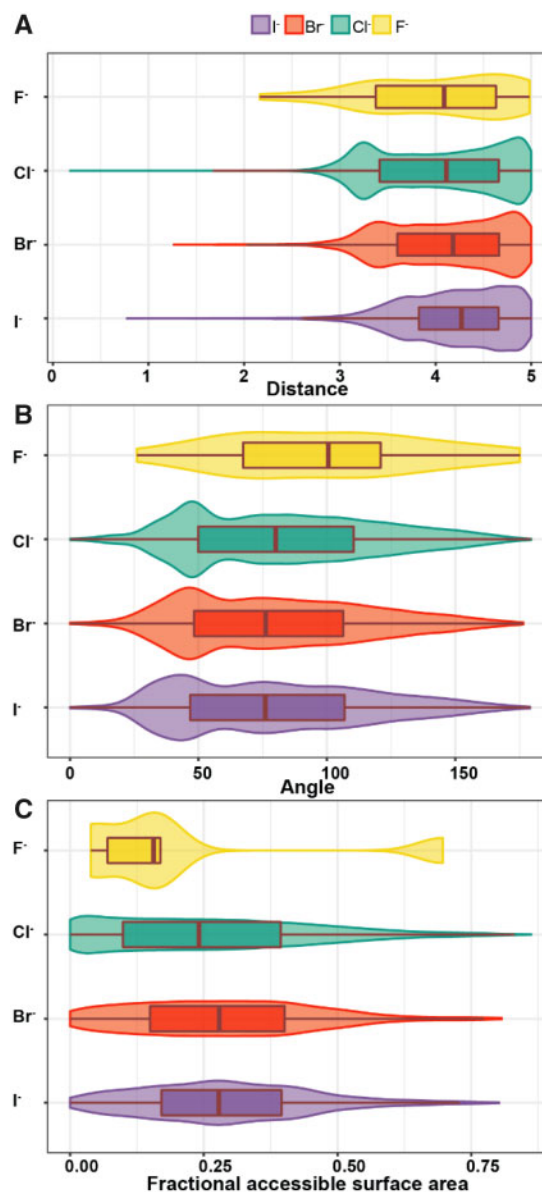


Fig. 1. Halide-binding sites in proteins. Distribution of (A) distances between halide and coordinating atoms, (B) angles between two vectors (halide-nearest-coordinating atom, halide-coordinating atom) and (C) fractional ASA values

located on the surface of proteins, apart from fluoride, which seem to be buried deeper inside (Fig. 1C).

We have also analyzed preferences of halides binding to different secondary structure elements (Fig. 2A) and it shows that the binding sites are predominantly located in the loop regions between α -helices and β -strands. At the same time, the α -helices are more frequently found to provide the interacting residues than β -strands.

In terms of amino acid residues which compose the binding sites, there is a strong preference for positively charged amino acid residues for all four halides (Fig. 2B). Arg side chains are the most universal anchor for halide binding via its guanidinium moiety. Interestingly for halide-containing small molecules, this is in general also true (Supplementary Fig. S2), however, in this case, polar Ser, Thr, Asn, Gln and also negatively charged Asp and Glu (especially in case of fluorine- and bromine-containing compounds) also play a significant role. This might be dictated by the exact composition of a small molecule with an incorporated halogen atom.

Polar Ser and Thr (with hydroxyl side chain) and Asn (with carboxamide group) also participate frequently in the coordination of

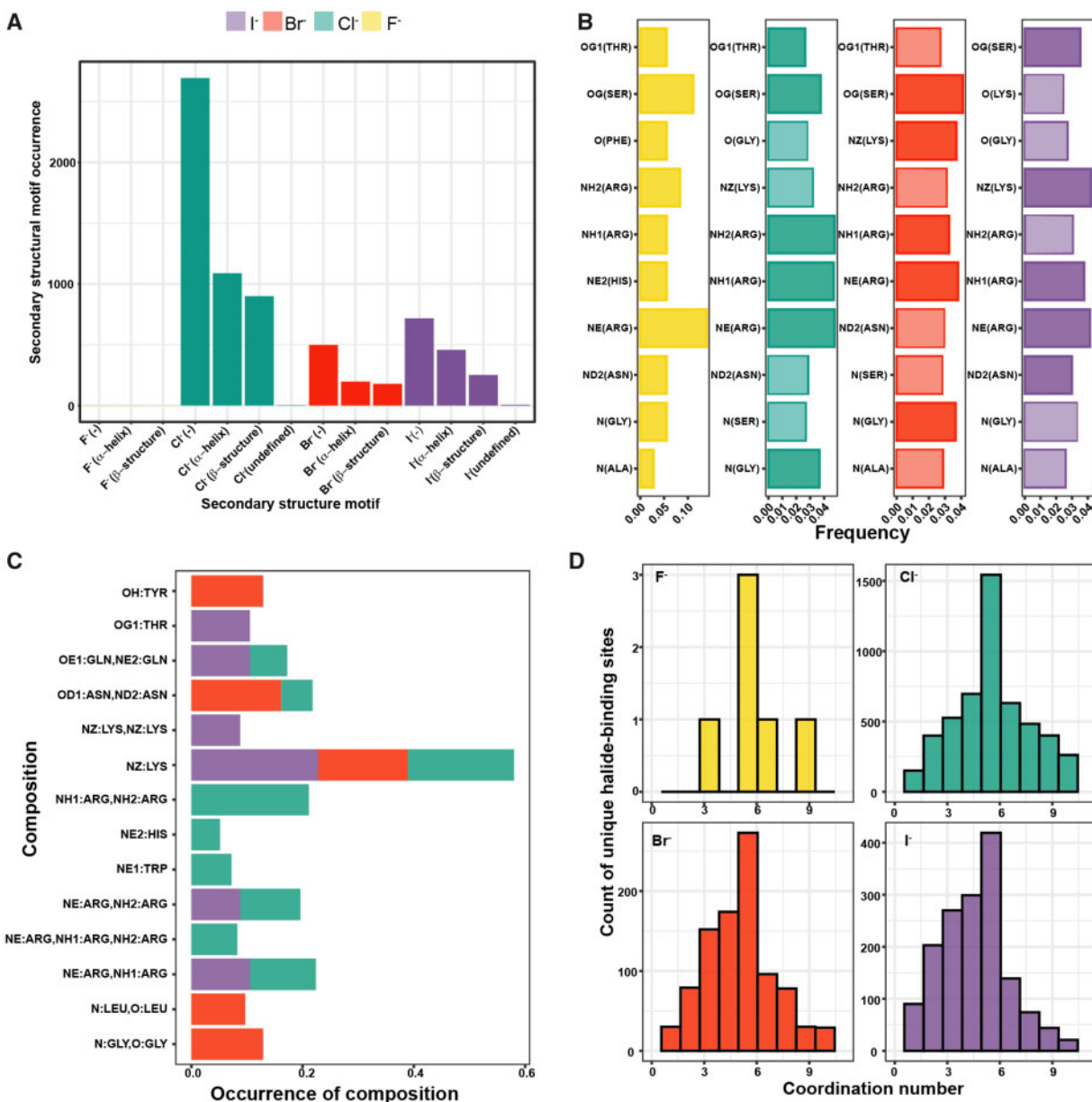


Fig. 2. Composition of halide-binding sites in proteins. (A) Preference to bind to certain secondary structure elements, namely α -helix, β -strand and – none. (B) The amino acid composition of binding sites, (C) certain combinations of amino acids in binding sites and (D) coordination numbers of halides

halides (Fig. 2B). However, to our big surprise, there are numerous cases, where negatively or partially negatively charged atoms from Asp and Glu side chains as well as mainchain carbonyl are involved (~30% of cases). At the physiological pH values, these atoms, in principle, should have repulsion with halides. However, it seems that at least in some structures these interactions are mediated by water molecules. For Br⁻, I⁻ and Cl⁻ sites there is an additional positive charge provided by side chains of Lys via its ϵ -amino group and in case of F⁻, it is provided by imidazole group of histidine. Interestingly for all halides, the main chain (namely its NH group) of small residues (Ala, Ser and especially Gly), is frequently involved in the interactions (Fig. 2B).

When we checked for the common combinations in the binding sites for various halides, we revealed that some combinations are specific to a certain halide (Fig. 2C). For example, the combination of two Lys side chains occurs only for I⁻, whereas single Lys residue might be a part of Cl⁻, Br⁻ or I⁻ binding sites. The Tyr side chain is unique to Br⁻ binding sites, as well as the combination of two Gly

residues, whereas Trp and His side chains are found in Cl⁻ binding sites only.

3.4 Coordination numbers of halides in binding sites

The coordination numbers for halides vary a lot (Fig. 2D)—from two, indicating the simplest linear configuration, up to nine, corresponding to either tricapped trigonal prismatic or capped square antiprismatic configuration (Supplementary Fig. S3). However, it turns out that the most common coordination number for all halides is 5 (Fig. 2D), corresponding to trigonal bipyramidal or square pyramidal geometry of the binding sites.

Interestingly, there is a certain correlation between the coordination numbers of halides and the amino acid residues involved in the respective binding sites (Supplementary Fig. S4). For example, Lys residues are way more common in the binding sites of halides with the low coordination numbers, whereas Cys residues are exclusively found in the sites of halides with the high coordination numbers.

4 Discussion

Halides are ubiquitous in the environment and impact all living organisms on our planet. Whereas some of them, such as chloride became a universal counterion for positively charged potassium and sodium, and contribute to buildup of electrochemical gradients, others, such as bromide and iodide play rather very defined roles, for example as strong antioxidants. Fluoride is mostly considered as a toxic compound, and many organisms developed fluoride expelling channels, and the only well-documented case where it is actively used as an essential compound is the class of adenosyl-fluoride synthases, catalyzing the formation of a carbon-fluorine bond to S-adenosyl-L-methionine, with the concomitant release of L-methionine (O'Hagan *et al.*, 2002). Furthermore, there are many other enzymes which catalyze the formation of numerous halogenated products, some of which are actively being investigated for their potent pharmacological properties (Jesus *et al.*, 2019; Kasanah and Triyanto, 2019; Rocha *et al.*, 2018).

Considering all the importance of halides in biology, it is surprising that rather little is known about their binding to proteins, apart from the logical suggestion that negatively charged anions will be recruited to the positively charged side chains via the formation of ion pair as dictated by the so-called 'law of matching water affinities' (Collins, 2006). Although our extensive analysis confirms this general observation, it additionally reveals certain patterns and preferences among four halides for its binding partners. For example, the occurrence of Arg side chain might be a general flag for the halide-binding site, however, the presence of other positively charged residues such as Lys and His, or polar residues can hint to the certain halide.

Based on amino acid composition of the binding sites, interaction lengths and angles, fASA and coordination numbers, it should be feasible to distinguish halides from other ions, however, the predictor is certainly not powerful enough to unambiguously assign the exact identity of a halide in question (Supplementary Fig. S5). In cases when the anomalous data are present, such an identification can be done more properly.

Furthermore, our analysis is of value only for the proteins with the structural models obtained; it is not possible to highlight the putative halide-binding sites based on the amino acid sequence only. At the same time since in many cases the formation of binding sites involves conformational changes of interacting side chains, the apo structures are not very likely to yield proper predictions. Nevertheless, the revealed patterns might be used either for fast evaluation of available PDB models in general to verify the assignments of ions in the absence of any other source of information and/or engineering of halide-binding sites in proteins.

The assembled pipeline can also be used just for analysis purposes. Our present analysis of the dataset of halide-binding sites obtained by using this pipeline demonstrates its usefulness and opens an avenue for more detailed future studies on halides binding to proteins.

Funding

This work was supported by the Government of the Russian Federation through the ITMO Fellowship and Professorship Program and by the grant 08-08.

Conflict of Interest: none declared.

References

Abendroth, J. *et al.* (2011) SAD phasing using iodide ions in a high-throughput structural genomics environment. *J. Struct. Funct. Genomics*, **12**, 83–95.
 Accardi, A. (2015) Structure and gating of CLC channels and exchangers. *J. Physiol.*, **593**, 4129–4138.
 Accardi, A. and Picollo, A. (2010) CLC channels and transporters: proteins with borderline personalities. *Biochim. Biophys. Acta*, **1798**, 1457–1464.
 Aghajari, N. *et al.* (2002) Structural basis of alpha-amylase activation by chloride. *Protein Sci.*, **11**, 1435–1441.

Almeida, M. *et al.* (2001) Vanadium haloperoxidases from brown algae of the Laminariaceae family. *Phytochemistry*, **57**, 633–642.
 Andrini, O. *et al.* (2015) CLC-K chloride channels: emerging pathophysiology of Bartter syndrome type 3. *Am. J. Physiol. Renal Physiol.*, **308**, F1324–F1334.
 Aoba, T. and Fejerskov, O. (2002) Dental fluorosis: chemistry and biology. *Crit. Rev. Oral Biol. Med.*, **13**, 155–170.
 Aoun, A. *et al.* (2018) The fluoride debate: the pros and cons of fluoridation. *Prev. Nutr. Food Sci.*, **23**, 171–180.
 Argenzio, E. and Moolenaar, W.H. (2016) Emerging biological roles of Cl⁻ intracellular channel proteins. *J. Cell Sci.*, **129**, 4165–4174.
 Argenzio, E. *et al.* (2014) CLIC4 regulates cell adhesion and β 1 integrin trafficking. *J. Cell Sci.*, **127**, 5189–5203.
 Ayala, R. *et al.* (2003) On the halide hydration study: development of first-principles halide ion-water interaction potential based on a polarizable model. *J. Chem. Phys.*, **119**, 9538–9548.
 Barbier, O. *et al.* (2010) Molecular mechanisms of fluoride toxicity. *Chem. Biol. Interact.*, **188**, 319–333.
 Barratt, T.M. and Walsler, M. (1969) Extracellular fluid in individual tissues and in whole animals: the distribution of radi sulfate and radiobromide. *J. Clin. Invest.*, **48**, 56–66.
 Berg, J. *et al.* (2012) Ca²⁺-activated Cl⁻ channels at a glance. *J. Cell Sci.*, **125**, 1367–1371.
 Bernal, J. *et al.* (2015) Thyroid hormone transporters—functions and clinical implications. *Nat. Rev. Endocrinol.*, **11**, 690.
 Berryman, M. and Bretscher, A. (2000) Identification of a novel member of the chloride intracellular channel gene family (CLIC5) that associates with the actin cytoskeleton of placental microvilli. *Mol. Biol. Cell*, **11**, 1509–1521.
 Bogdanov, A.M. *et al.* (2016) Turning on and off photoinduced electron transfer in fluorescent proteins by π -stacking, halide binding, and Tyr145 mutations. *J. Am. Chem. Soc.*, **138**, 4807–4817.
 Brahmachari, U. *et al.* (2017) Chloride maintains a protonated internal water network in the photosynthetic oxygen evolving complex. *J. Phys. Chem. B*, **121**, 10327–10337.
 Brammer, A.E. *et al.* (2014) F⁻/Cl⁻ selectivity in CLCF-type F⁻/H⁺ antiporters. *J. Gen. Physiol.*, **144**, 129–136.
 Buzalaf, M.A.R. *et al.* (2011) Mechanisms of action of fluoride for caries control. *Monogr. Oral Sci.*, **22**, 97–114.
 Cabrita, M.T. *et al.* (2010) Halogenated compounds from marine algae. *Mar. Drugs*, **8**, 2301–2317.
 Cahnmann, H.J. and Funakoshi, K. (1970) Model reactions for the biosynthesis of thyroxine. Nonenzymic formation of 3,5,3'-triiodothyronine from 4-hydroxy-3-iodophenylpyruvic acid, 3,5-diiodotyrosine, and oxygen. *Biochemistry*, **9**, 90–98.
 Chen, J.-C. *et al.* (2019) WNK4 kinase is a physiological intracellular chloride sensor. *Proc. Natl. Acad. Sci. USA*, **116**, 4502–4507.
 Collins, K.D. (2006) Ion hydration: implications for cellular function, polyelectrolytes, and protein crystallization. *Biophys. Chem.*, **119**, 271–281.
 Crockford, S.J. (2009) Evolutionary roots of iodine and thyroid hormones in cell-cell signaling. *Integr. Comp. Biol.*, **49**, 155–166.
 Cunha, T.D.S. and Heilberg, I.P. (2018) Bartter syndrome: causes, diagnosis, and treatment. *Int. J. Nephrol. Renovasc. Dis.*, **11**, 291–301.
 Darrouzet, E. *et al.* (2014) The sodium/iodide symporter: state of the art of its molecular characterization. *Biochim. Biophys. Acta*, **1838**, 244–253.
 Dauter, M. and Dauter, Z. (2007) Phase determination using halide ions. *Methods Mol. Biol.*, **364**, 149–158.
 Dauter, Z. and Dauter, M. (1999) Anomalous signal of solvent bromides used for phasing of lysozyme. *J. Mol. Biol.*, **289**, 93–101.
 Dauter, Z. *et al.* (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 232–237.
 Dauter, Z. *et al.* (2001) Entering a new phase: using solvent halide ions in protein structure determination. *Structure*, **9**, R21–R26.
 Devuyst, O. and Thakker, R.V. (2010) Dent's disease. *Orphanet J. Rare Dis.*, **5**, 28.
 DiStefano, J.J. *et al.* (1993) Binding and degradation of 3,5,3'-triiodothyronine and thyroxine by rat intestinal bacteria. *Am. J. Physiol.*, **264**, E966–E972.
 Eales, J.G. (1997) Iodine metabolism and thyroid-related functions in organisms lacking thyroid follicles: are thyroid hormones also vitamins? *Proc. Soc. Exp. Biol. Med.*, **214**, 302–317.
 Eisenberg, A.D. and Marquis, R.E. (1981) Enhanced transmembrane proton conductance in *Streptococcus mutans* GS-5 due to ionophores and fluoride. *Antimicrob. Agents Chemother.*, **19**, 807–812.
 Fincham, A.G. *et al.* (1999) The structural biology of the developing dental enamel matrix. *J. Struct. Biol.*, **126**, 270–299.

- Fox, J.M. et al. (2015) Interactions between Hofmeister anions and the binding pocket of a protein. *J. Am. Chem. Soc.*, **137**, 3859–3866.
- Frances, C. et al. (2003) Bromism from daily over intake of bromide salt. *J. Toxicol. Clin. Toxicol.*, **41**, 181–183.
- Freeman, H.J. and Kim, Y.S. (1978) Digestion and absorption of protein. *Annu. Rev. Med.*, **29**, 99–116.
- Galinski, E.A. (1995) Osmoadaptation in bacteria. *Adv. Microb. Physiol.*, **37**, 272–328.
- Gururaja Rao, S. et al. (2018) Three decades of chloride intracellular channel proteins: from organelle to organ physiology. *Curr. Protoc. Pharmacol.*, **80**, 11.21.1–11.21.17.
- Guskov, A. et al. (2009) Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.*, **16**, 334–342.
- Hänelt, I. and Müller, V. (2013) Molecular mechanisms of adaptation of the moderately halophilic bacterium *Halobacillus halophilus* to its environment. *Life*, **3**, 234–243.
- Henderson, J.P. et al. (2001) The eosinophil peroxidase-hydrogen peroxide–bromide system of human eosinophils generates 5-bromouracil, a mutagenic thymine analogue. *Biochemistry*, **40**, 2052–2059.
- Hoefl, S.E. et al. (2000) Metabolism of methyl bromide and dimethyl sulfide by marine bacteria isolated from coastal and open waters. *Aquat. Microb. Ecol.*, **21**, 221–230.
- Hwang, T.-C. et al. (2018) Structural mechanisms of CFTR function and dysfunction. *J. Gen. Physiol.*, **150**, 539–570.
- Imbrici, P. et al. (2015a) ClC-1 chloride channels: state-of-the-art research and future challenges. *Front. Cell. Neurosci.*, **9**, 156.
- Imbrici, P. et al. (2015b) ClC-1 mutations in myotonia congenita patients: insights into molecular gating mechanisms and genotype-phenotype correlation. *J. Physiol.*, **593**, 4181–4199.
- Jentsch, T.J. (2007) Chloride and the endosomal-lysosomal pathway: emerging roles of CLC chloride transporters. *J. Physiol.*, **578**, 633–640.
- Jentsch, T.J. and Pusch, M. (2018) CLC chloride channels and transporters: structure, function, physiology, and disease. *Physiol. Rev.*, **98**, 1493–1590.
- Jesus, A. et al. (2019) Isolation and potential biological applications of haloaryl secondary metabolites from macroalgae. *Mar. Drugs*, **17**, 73.
- Jiang, L. et al. (2012) Intracellular chloride channel protein CLIC1 regulates macrophage function through modulation of phagosomal acidification. *J. Cell Sci.*, **125**, 5479–5488.
- Kanduti, D. et al. (2016) Fluoride: a review of use and effects on health. *Mater. Soc. Med.*, **28**, 133–137.
- Kasanah, N. and Triyanto, T. (2019) Bioactivities of halometabolites from marine actinobacteria. *Biomolecules*, **9**, 225.
- Kawakami, K. et al. (2009) Location of chloride and its possible functions in oxygen-evolving photosystem II revealed by X-ray crystallography. *Proc. Natl. Acad. Sci. USA*, **106**, 8567–8572.
- Kunka, A. et al. (2018) Haloalkane dehalogenases from marine organisms. *Methods Enzymol.*, **605**, 203–251.
- Küpper, F.C. et al. (2008) Iodide accumulation provides kelp with an inorganic antioxidant impacting atmospheric chemistry. *Proc. Natl. Acad. Sci. USA*, **105**, 6954–6958.
- Küpper, F.C. et al. (2011) Commemorating two centuries of iodine research: an interdisciplinary overview of current research. *Angew. Chem. Int. Ed. Engl.*, **50**, 11598–11620.
- Küpper, F.C. et al. (2018) Emission of volatile halogenated compounds, speciation and localization of bromine and iodine in the brown algal genome model *Ectocarpus siliculosus*. *J. Biol. Inorg. Chem.*, **23**, 1119–1128.
- Lanyi, J.K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.*, **38**, 272–290.
- Li, L. (2016) The biochemistry and physiology of metallic fluoride: action, mechanism, and implications. *Crit. Rev. Oral Biol. Med.*, **14**, 100–114.
- Littler, D.R. et al. (2010) The enigma of the CLIC proteins: ion channels, redox proteins, enzymes, scaffolding proteins? *FEBS Lett.*, **584**, 2093–2101.
- Liu, F. et al. (2017) Molecular structure of the human CFTR ion channel. *Cell*, **169**, 85–95.e8.
- Liu, F. et al. (2019) Structural identification of a hotspot on CFTR for potentiation. *Science*, **364**, 1184–1188.
- Lu, P.-J. et al. (2010) Gastric juice acidity in upper gastrointestinal diseases. *World J. Gastroenterol.*, **16**, 5496–5501.
- Lugassy, D. and Nelson, L. (2009) Case files of the medical toxicology fellowship at the New York City poison control: bromism: forgotten, but not gone. *J. Med. Toxicol.*, **5**, 151–157.
- Marquis, R.E. et al. (2003) Fluoride and organic weak acids as modulators of microbial physiology. *FEMS Microbiol. Rev.*, **26**, 493–510.
- McCall, A.S. et al. (2014) Bromine is an essential trace element for assembly of collagen IV scaffolds in tissue development and architecture. *Cell*, **157**, 1380–1392.
- Melnikov, I. et al. (2017) Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci. Adv.*, **3**, e1602952.
- Mendoza-Schulz, A. et al. (2009) The effects of fluoride on cell migration, cell proliferation, and cell metabolism in GH4C1 pituitary tumour cells. *Toxicol. Lett.*, **190**, 179–186.
- Mitternacht, S. (2016) FreeSASA: an open source C library for solvent accessible surface area calculations. *F1000 Res.*, **5**, 189.
- Murthy, K.S. and Makhlof, G.M. (1994) Fluoride activates G protein-dependent and -independent pathways in dispersed intestinal smooth muscle cells. *Biochem. Biophys. Res. Commun.*, **202**, 1681–1687.
- O'Hagan, D. et al. (2002) Biochemistry: biosynthesis of an organofluorine molecule. *Nature*, **416**, 279.
- Oldham, M.L. and Chen, J. (2011) Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. USA*, **108**, 15152–15156.
- Oren, A. (1986) Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobius*. *Can. J. Microbiol.*, **32**, 4–9.
- Osei-Owusu, J. et al. (2018) Molecular biology and physiology of volume-regulated anion channel (VRAC). *Curr. Top. Membr.*, **81**, 177–203.
- Paula, S. et al. (1998) Permeation of halide anions through phospholipid bilayers occurs by the solubility-diffusion mechanism. *Biophys. J.*, **74**, 319–327.
- Pedersen, T.H. et al. (2016) Role of physiological ClC-1 Cl⁻ ion channel regulation for the excitability and function of working skeletal muscle. *J. Gen. Physiol.*, **147**, 291–308.
- Perumal, E. et al. (2013) A brief review on experimental fluorosis. *Toxicol. Lett.*, **223**, 236–251.
- Piwon, N. et al. (2000) ClC-5 Cl⁻-channel disruption impairs endocytosis in a mouse model for Dent's disease. *Nature*, **408**, 369–373.
- Poët, M. et al. (2006) Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6. *Proc. Natl. Acad. Sci. USA*, **103**, 13854–13859.
- Prange, H.D. et al. (2001) Physiological consequences of oxygen-dependent chloride binding to hemoglobin. *J. Appl. Physiol.*, **91**, 33–38.
- Qin, J. et al. (2006) Fluoride inhibition of enolase: crystal structure and thermodynamics. *Biochemistry*, **45**, 793–800.
- Ren, A. et al. (2012) Fluoride ion encapsulation by Mg²⁺ ions and phosphates in a fluoride riboswitch. *Nature*, **486**, 85–89.
- Rocha, D.H.A. et al. (2018) Seaweed secondary metabolites in vitro and in vivo anticancer activity. *Mar. Drugs*, **16**, 410.
- Roefler, M. and Müller, V. (2001) Chloride dependence of glycine betaine transport in *Halobacillus halophilus*. *FEBS Lett.*, **489**, 125–128.
- Roefler, M. et al. (2000) Motility and flagellum synthesis in *Halobacillus halophilus* are chloride dependent. *J. Bacteriol.*, **182**, 532–535.
- Sabirov, R.Z. and Okada, Y. (2009) The maxi-anion channel: a classical channel playing novel roles through an unidentified molecular entity. *J. Physiol. Sci.*, **59**, 3–21.
- Saenko, G.N. et al. (1978) Concentration of iodine and bromine by plants in the seas of Japan and Okhotsk. *Mar. Biol.*, **47**, 243–250.
- Saum, S.H. and Müller, V. (2008) Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies. *Saline Syst.*, **4**, 4–15.
- Senese, R. et al. (2014) Thyroid: biological actions of 'nonclassical' thyroid hormones. *J. Endocrinol.*, **221**, R1–R12.
- Shannon, R.D.I. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A*, **32**, 751–767.
- Smith, A.J. and Lippiat, J.D. (2010) Direct endosomal acidification by the outwardly rectifying ClC-5 Cl⁻/H⁺ exchanger. *J. Physiol.*, **588**, 2033–2045.
- Stockbridge, R.B. et al. (2012) Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *Proc. Natl. Acad. Sci. USA*, **109**, 15289–15294.
- Stockbridge, R.B. et al. (2013) A family of fluoride-specific ion channels with dual-topology architecture. *Elife*, **2**, e01084.
- Stockbridge, R.B. et al. (2015) Crystal structures of a double-barrelled fluoride ion channel. *Nature*, **525**, 548–551.
- Strunecka, A. et al. (2007) Fluoride interactions: from molecules to disease. *Curr. Signal Transduct. Ther.*, **2**, 190–213.
- Thomas, E.L. et al. (1995) Oxidation of bromide by the human leukocyte enzymes myeloperoxidase and eosinophil peroxidase. Formation of bromamines. *J. Biol. Chem.*, **270**, 2906–2913.
- Toyoshima, C. et al. (2004) Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature*, **432**, 361–368.

- Ulmasov, B. *et al.* (2009) Chloride intracellular channel protein-4 functions in angiogenesis by supporting acidification of vacuoles along the intracellular tubulogenic pathway. *Am. J. Pathol.*, **174**, 1084–1096.
- van Leeuwen, F.X. and Sangster, B. (1987) The toxicology of bromide ion. *Crit. Rev. Toxicol.*, **18**, 189–213.
- Weatherell, J.A. (1969) Uptake and distribution of fluoride in bones and teeth and the development of fluorosis. In: Barltrop, W. and Burland, W.L. (eds.) *Mineral Metabolism in Paediatrics*. Blackwell Scientific Publications, Oxford-Edinburgh, pp. 53–70.
- Wege, S. *et al.* (2017) Chloride: not simply a ‘cheap osmoticum’, but a beneficial plant macronutrient. *J. Exp. Bot.*, **68**, 3057–3069.
- Wolf-Watz, M. *et al.* (2001) Chloride binding by the AML1/Runx1 transcription factor studied by NMR. *FEBS Lett.*, **488**, 81–84.
- Wong, M.C.M. *et al.* (2011) Cochrane reviews on the benefits/risks of fluoride toothpastes. *J. Dent. Res.*, **90**, 573–579.
- Yates, C.J. *et al.* (2014) Molecular and thermodynamic mechanisms of the chloride-dependent human angiotensin-I-converting enzyme (ACE). *J. Biol. Chem.*, **289**, 1798–1814.
- Yoshikawa, M. *et al.* (2002) CLC-3 deficiency leads to phenotypes similar to human neuronal ceroid lipofuscinosis. *Genes Cells*, **7**, 597–605.
- Zimmermann, M.B. (2011) The role of iodine in human growth and development. *Semin. Cell Dev. Biol.*, **22**, 645–652.