



## **Census of halide-binding sites in protein structures**

Rostislav Skitchenko, Dmitrii Usoltsev, Mayya Uspenskaya, Andrey Kajava, Albert Guskov

### **► To cite this version:**

Rostislav Skitchenko, Dmitrii Usoltsev, Mayya Uspenskaya, Andrey Kajava, Albert Guskov. Census of halide-binding sites in protein structures. *Bioinformatics*, 2020, 36 (10), pp.3064-3071. <10.1093/bioinformatics/btaa079>. <hal-03089275>

**HAL Id: hal-03089275**

**<https://hal.science/hal-03089275v1>**

Submitted on 29 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Census of halide binding sites in protein structures

Rostislav K Skitchenko<sup>1</sup>, Dmitrii Usoltsev<sup>1</sup>, Maya Uspenskaya<sup>1</sup>, Andrey V. Kajava<sup>1,2</sup>  
and Albert Guskov<sup>3,\*</sup>

<sup>1</sup> ITMO University, Saint-Petersburg 197101, Russia

<sup>2</sup> Centre de Recherche en Biologie cellulaire de Montpellier (CRBM), UMR 5237  
CNRS, Universite Montpellier, CEDEX 5, 34293 Montpellier, France

<sup>3</sup> Groningen Biomolecular & Biotechnology Institute, University of Groningen,  
Nijenborgh 7, 9747 AG, Groningen, the Netherlands

\*correspondence to a.guskov@rug.nl

**Abstract**

Halides are negatively charged ions of halogens, forming fluoride (F<sup>-</sup>), chloride (Cl<sup>-</sup>), bromide (Br<sup>-</sup>) and iodide (I<sup>-</sup>). These ions are quite reactive and interact both specifically and non-specifically with proteins. In this study we have developed a protocol and a pipeline for the analysis of halide binding sites in proteins. 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. Taken together now it is possible to assign the identity of bound ion based on the geometry and composition of binding sites and it should be possible to predict halide binding sites in future. This is of use for characterization of specific halide-protein interactions and phasing techniques relying on halides as anomalous scatters.

## 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 ( $At^-$ ), as due to its unstable (radioactive) nature they have no biological implication. 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  $[Ar]3d^{10}4s^24p^6$  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 ion as a ligand in the pdb data bank (i.e. less than 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 a large radii: 1.19 Å, 1.67 Å, 1.82 Å and 2.06 Å for  $F^-$ ,  $Cl^-$ ,  $Br^-$  and  $I^-$ , respectively<sup>1</sup>.

### 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<sup>2,3</sup>. It has a dual role, as it reinforces enamel via slowdown of demineralization and promotion of remineralization via formation of fluorohydroxyapatite<sup>4,5</sup>, but also via suppression of mouth bacteria by inhibiting certain intracellular enzymes, leading to the decreased production of lactic

51 acid and hence lowering the risk of caries formation<sup>6</sup>. Clearly, taking into account such  
52 an inhibitory function, the valid question is the safety of fluoride for humans, especially  
53 since in many countries tap water and table salt come fluorinated. According to the  
54 current research and guidelines there is no actual risk for humans under normal  
55 circumstances, apart for those who live in the areas, where water is naturally enriched  
56 in fluoride – some parts of India, China and Africa continent. The prolonged exposure  
57 to the excess of fluoride might lead to the disorders such skeletal fluorosis, dental  
58 fluorosis and kidney failure<sup>7-9</sup>.

59 However, the environmental levels of  $F^-$  of 10-100 $\mu$ M found in soil and water are toxic  
60 for many organisms, hence several transport systems to expel fluoride from cells, have  
61 evolved. One of them is Fluc proteins, which can be found in prokaryotes and lower  
62 eukaryotes<sup>10</sup>. Flucs are anion channels, which are highly specific to  $F^-$  over  $Cl^-$  (ref<sup>11</sup>).  
63 The other major class of  $F^-$  transporting proteins is a clade of CLC transporters, strictly  
64 confined to bacterial species only, so-called CLCF  $F^-/H^+$  antiporters, which couple  $F^-$   
65 efflux to a proton gradient<sup>12</sup>. The selectivity of these transporters for  $F^-$  over  $Cl^-$  is also  
66 extremely high<sup>13</sup>.

67 Clearly the exporters cannot immediately remove the excess of incoming  $F^-$  flux, hence  
68 it readily interacts with intracellular proteins. Among the best characterized targets to  
69 which  $F^-$  binds are  $F^-$ -specific riboswitch, where  $F^-$  is complexed with  $Mg^{2+}$  and RNA  
70 phosphate groups<sup>14</sup>; enolase<sup>15</sup>, which is an essential enzyme for glycolysis; heme  
71 containing proteins and numerous phosphatases<sup>16</sup>. Fluoride ions readily react with  
72 aluminum and beryllium, and the formed complexes are highly cytotoxic as they mimic  
73 phosphate group hence inhibiting numerous enzymes, which exert ATPase or  
74 GTPase activity<sup>17</sup>. 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<sup>18-20</sup>.

It is important to note that in many cases the inhibition effect might be caused not via binding of fluoride ions to proteins, but also as an effect of acidification of cytoplasm, as F<sup>-</sup> might play a role of transmembrane proton carrier<sup>21,22</sup>.

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 (ref<sup>23</sup>) and calcium ATP-ase<sup>24</sup>.

### *Chloride*

Chloride is a ubiquitous ion 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 ions (especially in certain bacteria and archaea) is as an osmolyte during osmoadaptation<sup>25</sup>. In a high saline milieu halophilic and halotolerant microorganisms accumulate up to molar concentrations of chloride<sup>26,27</sup>. Interestingly some of these organisms (e.g. *Halobacillus halophilus*) are strictly dependent on chloride ions for their growth, as Cl<sup>-</sup> is directly involved in regulation of transcription and translation of several essential proteins<sup>28-30</sup>. 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<sup>31</sup>. In higher organisms chloride ions also play important roles. For example, in plants chloride is

required for turgor generation and the regulation of cell volume, as well as for generation of  $\text{Cl}^-$  currents<sup>32</sup>. 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<sup>33-35</sup>. In mammals, chloride ions 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<sup>36,37</sup>. 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  $\alpha$ -amylase<sup>38</sup>, angiotensin-converting enzyme I<sup>39</sup>, hemoglobin<sup>40</sup>, kinases<sup>41</sup>, acute myeloid leukemia-1 transcription factor<sup>42</sup>, and many others.

To regulate the flux of chloride ions numerous chloride channels and transporters evolved. The most studied chloride transporting proteins belong to the chloride channel (CLC) family<sup>43</sup> and Chloride Intracellular Ion Channel (CLIC) family<sup>44</sup>.

CLC proteins are integral membrane proteins, residing either in plasma or intracellular membranes and encompassing both channels and transporters<sup>45,46</sup>. They are involved in the control of excitability during muscle contraction, acidification of endosomes and lysosomes, and epithelial transport<sup>47-50</sup>. Intriguingly all members form dimers with a separate translocation pathway.

Malfunctions of CLC proteins cause severe diseases such as myotonia congenita<sup>51,52</sup>, Neuronal ceroid lipofuscinosis<sup>53,54</sup>, Dent's disease<sup>55</sup>, Bartter syndrome<sup>56,57</sup>.

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<sup>58</sup>. They are involved in signaling<sup>44</sup>, endosomal trafficking<sup>59</sup>,

phagosomal acidification<sup>60</sup>, angiogenesis<sup>61</sup>, actin-dependent membrane remodeling<sup>62</sup> and other intracellular processes<sup>63</sup>.

Another well-studied chloride channel is cystic fibrosis transmembrane conductance regulator (CFTR)<sup>64-66</sup>. 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<sup>67</sup>, maxi Cl<sup>-</sup> channels<sup>68</sup>, volume-regulated chloride channels<sup>69</sup>, which are beyond the scope of this work.

### *Bromide*

There is no solid evidence for a certain role of bromide ions in prokaryotes, although there is a large class of marine and soil microorganisms capable of oxidizing methyl bromides via transmethylation or monooxygenase pathway<sup>70-72</sup>. 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<sup>71</sup>. Furthermore, bromide is a preferred substrate for eosinophil peroxidases<sup>72,73</sup>, which catalyze the conversion of bromide to hypobromous acid for the host defense.

Bromide ions are localized mainly extracellularly, and its concentration seems to be tightly regulated<sup>74</sup>. Bromide deficiency leads to diminished tissue growth and causes failures in tissue development and remodelling<sup>71</sup>. However, at excess, bromide can cause bromism – the collective name of several neurological disorders caused by the neurotoxic effect of prolonged consumption of bromide<sup>75-77</sup>.

150 Interestingly, some marine algae accumulate large amounts of bromide (and iodide)  
151 but in the cell wall and usually not in the cytosol<sup>78,79</sup>. Their genomes encode a specific  
152 set of proteins to deal with halides, such as haloacid and haloalkane dehalogenases  
153 as well as vanadium haloperoxidase<sup>80,81</sup>. The latter enzyme is responsible for  
154 production of methyl halides. The accumulation of bromide is probably a consequence  
155 of the relative abundance of this ion in seawater: the average concentration of Br<sup>-</sup> is  
156 ~65 mg L<sup>-1</sup>, whereas F<sup>-</sup> and I<sup>-</sup> are ~1 mg L<sup>-1</sup> (however Cl<sup>-</sup> is predominant with  
157 concentration of 19000-23000 mg L<sup>-1</sup>). Surprisingly despite the concentration of Br<sup>-</sup> is  
158 300 times lower than that of Cl<sup>-</sup>, in many cases it is rather bromide than chloride (or  
159 their combination) which is used for organohalogen production<sup>82</sup>.

160 In the world of structural biology, bromide has caught an eye due to its phasing  
161 potential – with 36 electrons and easily accessible X-ray absorption edge (K edge  
162 ~0.92 Å) it is a good choice for single or multiwavelength anomalous diffraction (SAD  
163 / MAD) phasing<sup>83</sup>. In the easiest application a crystal should be shortly soaked in a  
164 cryoprotectant containing bromide just before the flash freezing in liquid nitrogen<sup>84</sup>.  
165 Bromide ions will quickly diffuse via the solvent channels and settle within the ordered  
166 solvent shell around the protein surface<sup>85,86</sup>.

### 167 *Iodide*

168 Iodide is one of the largest monoatomic anions and one of the heaviest elements  
169 utilized by living organisms. In vertebrates it is utilized for the production of growth-  
170 regulating thyroid hormones (thyroxine and triiodothyronine), which are essential  
171 regulators of virtually nearly all processes during different life phases<sup>87-89</sup>.

172 Uptake of iodide into thyroid occurs via the sodium iodide transporter (SLC5A5)<sup>90</sup>  
173 residing in the basolateral membrane of thyroid follicular cells. The transport is active



174 and the inward translocation of sodium down its electrochemical gradient is coupled  
175 to inward translocation of iodine against its electrochemical gradient.

176 Other organisms such as algae, zooplankton and plants are capable to accumulate  
177 iodine / iodide as thyroid hormone precursors, which can be used as developmental  
178 regulators. Many bacteria are capable of extracting necessary iodine /iodide from the  
179 host environment<sup>91</sup>. Marine microorganisms are especially agile in iodide  
180 accumulation as they are capable to reduce inorganic iodate (the most  
181 thermodynamically stable form) to iodide and produce numerous iodinated organic  
182 compounds<sup>79</sup>. However, many other bacteria, both aerobic and anaerobic, are also  
183 able to convert iodate to iodide. It seems that even in the absence of highly specialized  
184 enzymes of thyroid gland, iodothyrosines can form spontaneously and due to its  
185 reactivity play a crucial role in cell-cell signaling<sup>92,93</sup>.

186 Therefore, a plausible scenario is that during evolution iodine /iodide reacting with  
187 tyrosines might had been recruited as a potent signaling molecule somewhat after  
188 LUCA<sup>93</sup>.

189 Some macroalgae (kelp) developed an extreme concentrating capacity for iodine –  
190 e.g. *Laminaria digitata* can concentrate up to 30.000 more of iodide in its apoplasts  
191 compared to iodide concentration in seawater<sup>94</sup>. Such accumulation leads to a buildup  
192 of antioxidant reservoir that is mobilized during oxidative stress. Iodide can scavenge  
193 not only H<sub>2</sub>O<sub>2</sub> and ozone, but also hydroxyl radicals and superoxides<sup>79,95</sup>.

194 Iodide deficiency in humans is well documented and leads to the numerous mental  
195 and physical developmental delays<sup>96</sup> and currently up to two billion people are affected  
196 worldwide. Excess of iodide can also be toxic, especially in the case of selenium  
197 deficiency, when the function of Se-containing antioxidative enzymes is impaired.

198 Similarly to bromide one can exploit anomalous signal of iodide ions bound to proteins.

The X-ray absorption edge for iodide is not readily accessible (L-I edge  $\sim 2.39\text{\AA}$ ), however even far from it (at wavelengths of  $\sim 1.8\text{\AA}$ , 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<sup>97</sup>. Recently it has been proposed that iodide SAD phasing might be universally applied to membrane proteins<sup>98</sup> as their positively charged residues found at the hydrophobic-hydrophilic interface need a compensatory negative charge, hence increasing the odds of (ordered) binding of iodide ions at these areas.

Taking into account the importance of halides in the biochemistry of all life forms 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 the 23th of August 2019 in the PDB bank and revealed the following patterns.

## **Materials and methods**

### *Data acquisition and filtering*

X-ray data analysis of protein structures containing fluoride, chlorine, bromine and iodine atoms, which are coordinated only by protein without ligands, were obtained from the Protein Data Bank (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). 66, 12686, 455, and 864 entries were obtained for  $\text{F}^-$ ,  $\text{Cl}^-$ ,  $\text{Br}^-$  and  $\text{I}^-$  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<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup> and I<sup>-</sup> respectively. Water atoms were excluded from PDB files before the analysis. 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 have not been taken into account.

### *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. Water atoms were excluded from the analysis. 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<sup>99</sup> 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 Å, 1.67 Å, 1.82 Å, 2.06 Å for F<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup> and I<sup>-</sup> respectively) and water molecule radius (1.4 Å). fASA allows to distinguish buried vs surface-bound halides.

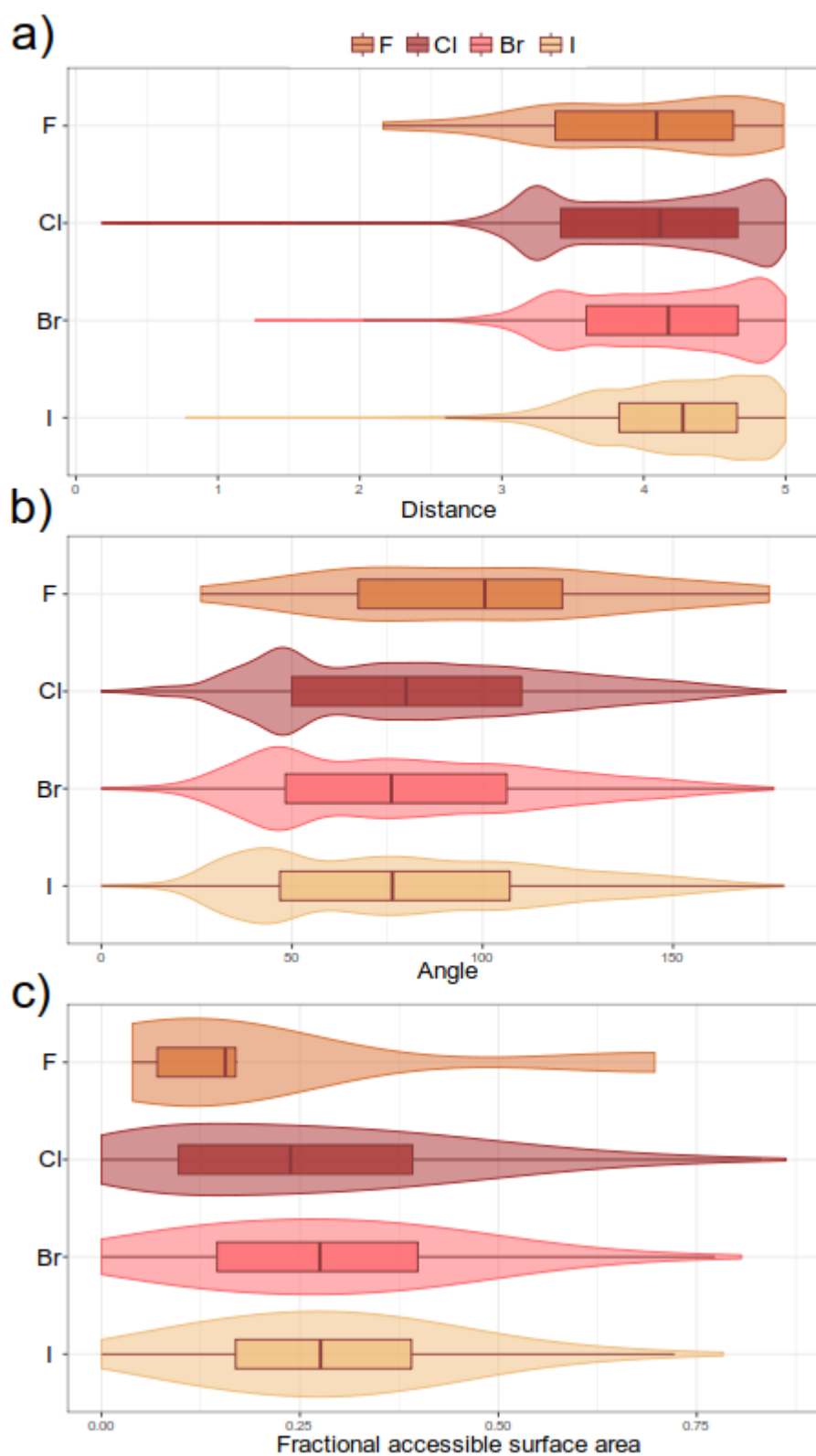
## **Results**

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 Materials and

249 methods), the resulting working dataset contains 25, 3229, 206 and 246 entries for F<sup>-</sup>  
250 , Cl<sup>-</sup>, Br<sup>-</sup> and I<sup>-</sup> respectively.

251 *Bimodal distributions of distances between halides and atoms of amino acid residues*

252 For all four halides the median distances between an anion and docking residue is  
253 about 4.16 Å (Fig. 1a), with the smallest value of 4.09 Å for fluoride followed by 4.11  
254 Å for chloride, 4.17 Å for bromide and the largest value of 4.28 Å for iodide reflecting  
255 difference in their ionic radii (*vide supra*). Such distances are indicative that the bound  
256 anions are either partially or completely dehydrated, as their hydrated radii are within  
257 3.3-3.5Å<sup>100,101</sup> .



258

259 Figure 1. Halide binding sites in proteins. Distribution of (a) distances between halide

260 and coordinating atoms, (b) angles between two vectors (halide-nearest coordinating

261 atom, halide-coordinating atom), (c) fractional ASA values. Halides are color-coded

262

263 However, if we look at the distribution of distances (Fig. 1a), the largest deviation is  
264 observed for chloride and iodide, with the distances going closer than 1.5 Å (those  
265 must be clear outliers though, as at such a short distance the repulsion of atoms is  
266 inevitable). The possible explanation of such a spread is that with the higher number  
267 of entries the odds of getting erroneous assignment is also higher; additionally, in the  
268 case of chloride since its anomalous signal is very low, it is hard to verify its assignment  
269 and it can be easily confused with water molecules. Surprisingly there is a large  
270 number of all four halides with the distances around 5Å, those most probably represent  
271 situations where the anions interact with atoms of residues via a water molecule.

272

#### 273 *Angular distribution of coordinating atoms*

274 The distribution of angles between the coordinating bonds is quite wide (Fig.1b),  
275 reflecting the possible positional errors (which can be of various origins and influenced  
276 by data quality and resolution). Nevertheless, Cl<sup>-</sup>, Br<sup>-</sup> and I<sup>-</sup> have a noticeable  
277 maximum at 45 degrees and also less visible blurred maximum at around 90 degrees,  
278 indicative of trigonal-bipyramidal or octahedral arrangement of binding sites.

279

#### 280 *Typical compositions of halide binding sites*

281 The analysis of accessible surface area of bound halides revealed that in the majority  
282 of cases, the binding sites form pockets that surround the halides. At the same time,  
283 on average, 35% of the halide surface is accessible to the solvent. It means that the  
284 binding sites are located on the surface of proteins, apart from fluoride, which seem to  
285 be buried deeper inside (Figure 1c).

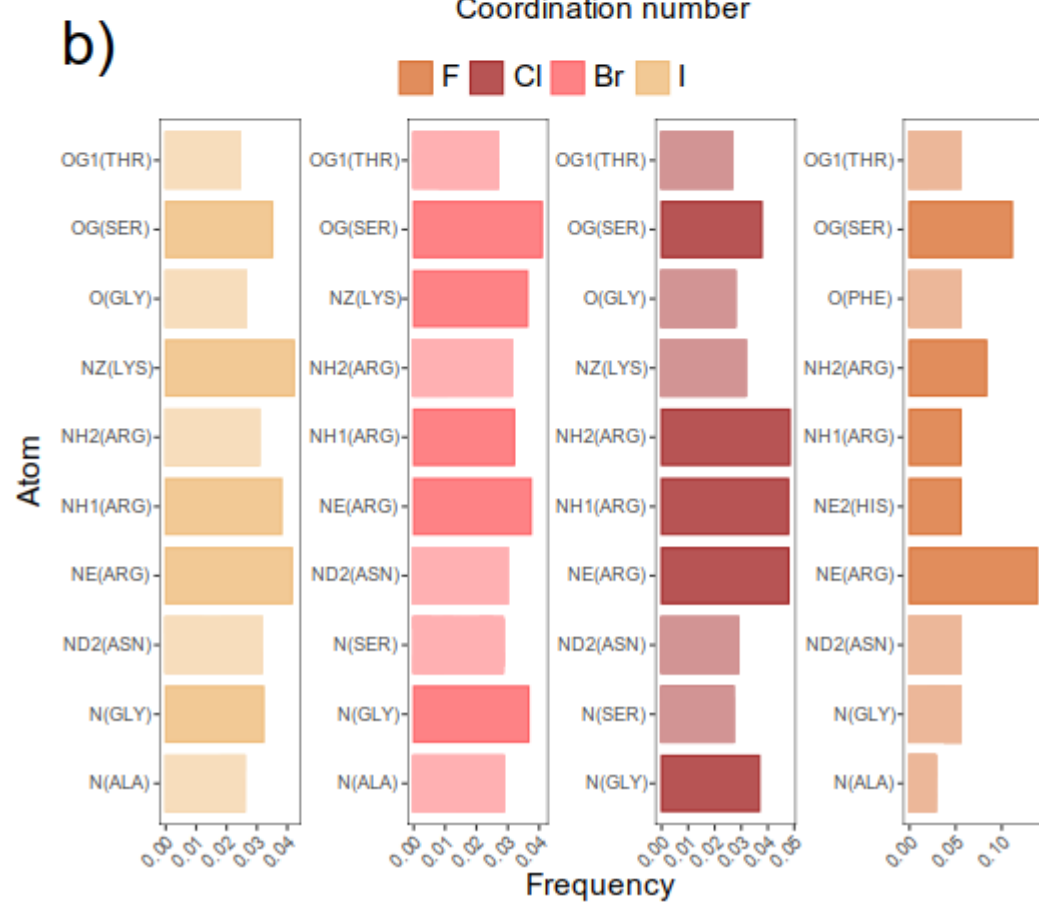
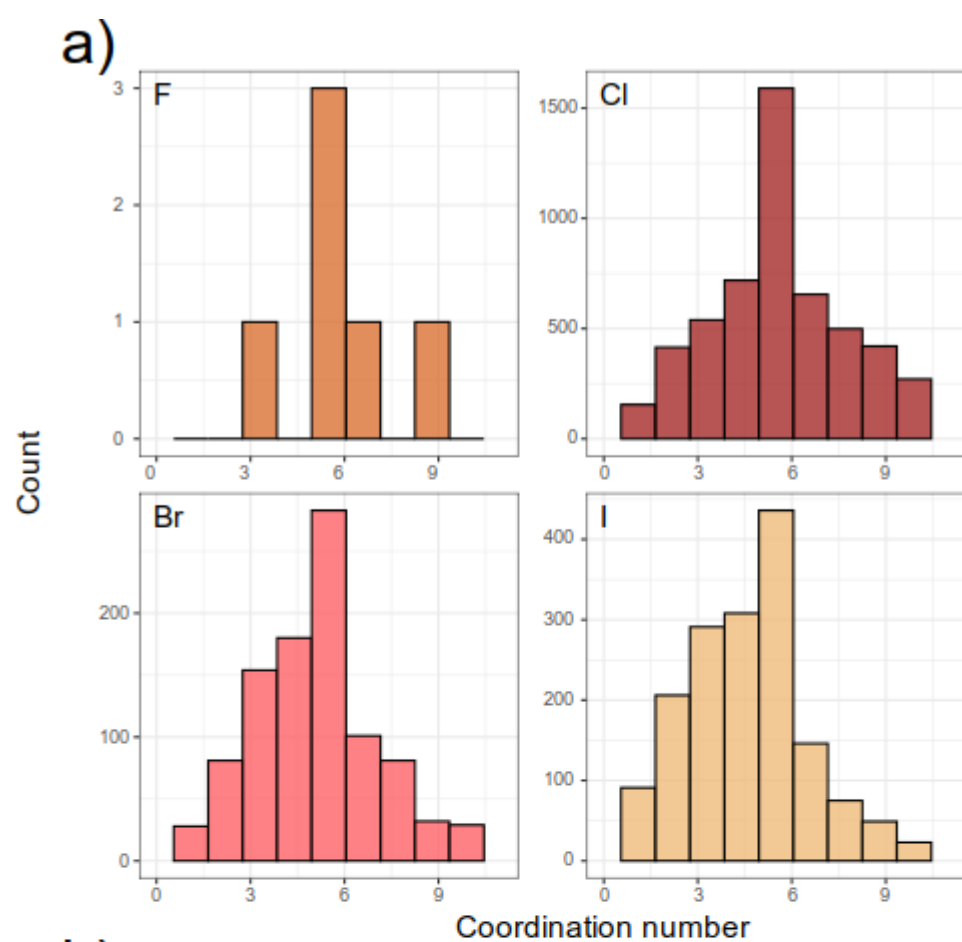


Figure 2. Coordination and residues forming halide binding sites. (a) Distribution of coordination numbers. (b) The top-10 most frequent coordinating atoms of the halide binding sites.

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. Polar Ser and Thr (with hydroxyl side chain) and Asn (with carboxamide group) also participate frequently in the coordination of halides. However, to our big surprise, there are some cases, where negatively or partially negatively charged atoms from Asp and Glu sidechains as well as mainchain carbonyl are involved. At the physiological pH values, these atoms, in principle, should have repulsion with halides. The only possibility which can be considered that in these cases interactions occur via water molecules of hydration shells. For Br<sup>-</sup>, I<sup>-</sup> and Cl<sup>-</sup> sites there is an additional positive charge provided by side chains of Lys via its ε-amino group and in case of F<sup>-</sup> 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.

When we checked for the common combinations in the binding sites for various halides we revealed some combinations are specific to a certain halide (Fig. 3.)



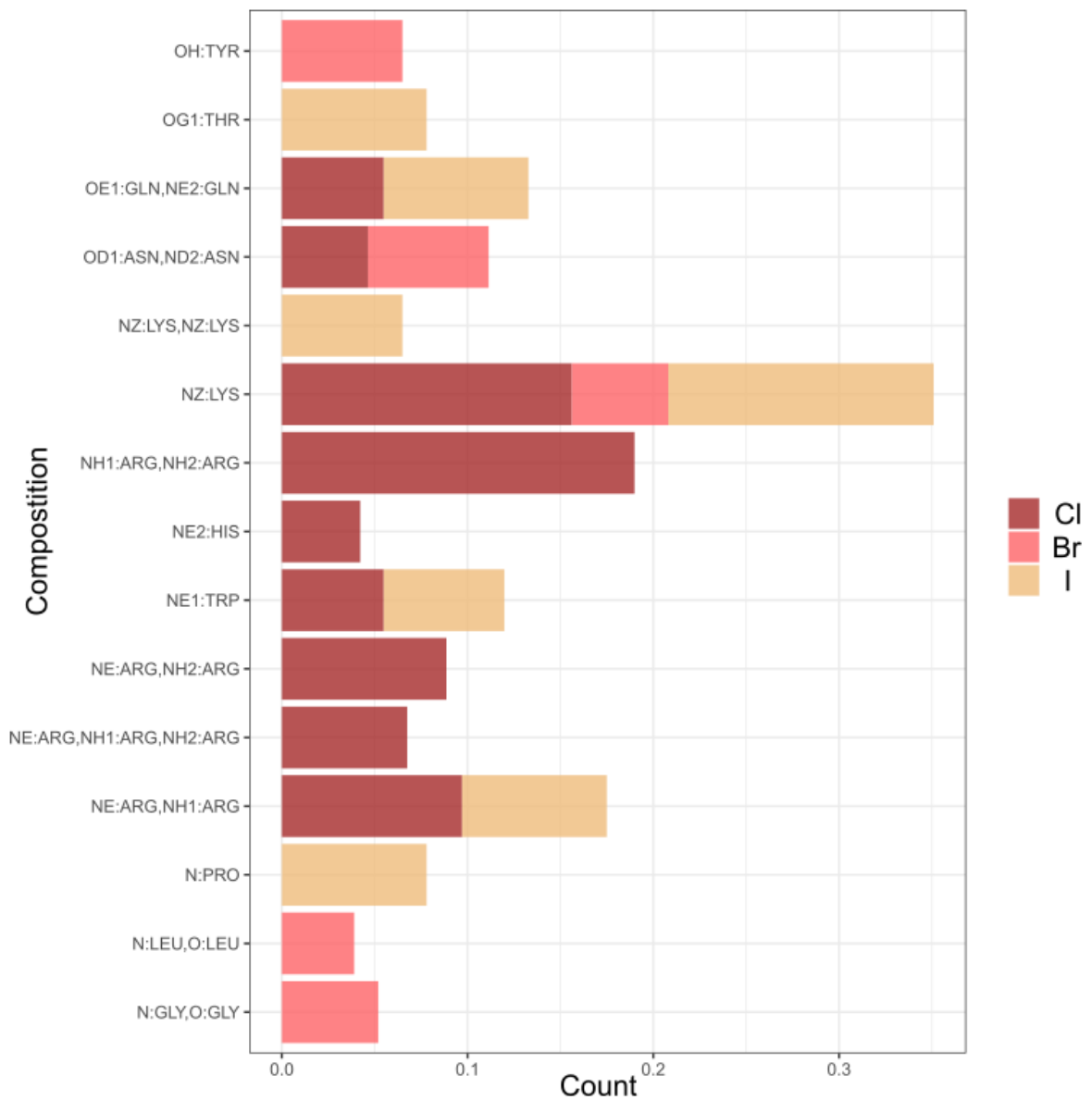


Figure 3. The revealed patterns of halide binding sites composition (color coded). Note that F<sup>-</sup> is not present due to the limited number of entries in the database.

For example, the combination of two Lys side chains occurs only for I<sup>-</sup>, whereas single Lys residue might be a part of Cl<sup>-</sup>, Br<sup>-</sup> or I<sup>-</sup> binding sites. The Tyr side chain is unique to Br<sup>-</sup> binding sites, as well as the combination of two Gly residues.

The coordination numbers for halides vary a lot (Fig.2a) – from two, indicating the simplest linear configuration, up to nine, corresponding to either tricapped trigonal prismatic or capped square antiprismatic configuration (Fig.4). However, it turns out that the most common coordination number for all halides is 5, corresponding to trigonal bipyramidal or square pyramidal geometry of the binding sites (Fig 4).

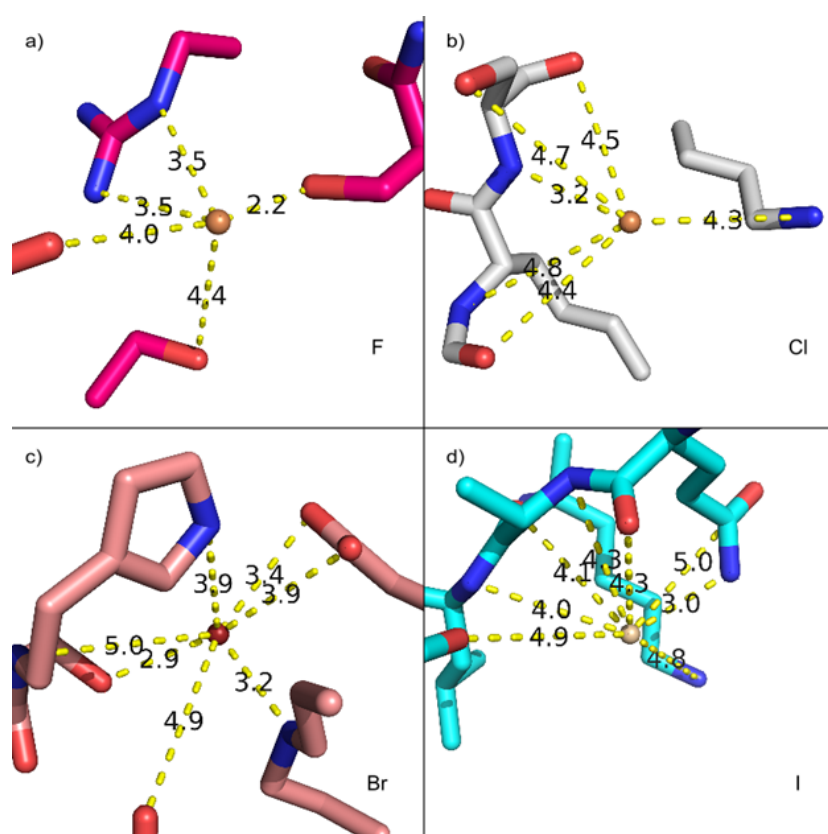


Figure 4. Examples of different arrangements in binding sites. Protein is shown in sticks, halides is a sphere. Distances are in Å. (a) trigonal bipyramidal, coordination number 5 (PDB id 2WSL) (b) trigonal prismatic, coordination number 6 (PDB id 1HZJ) (c) capped trigonal prismatic, coordination number 7 (PDB id 1MGY) (d) bicapped trigonal prismatic, coordination number 8 (PDB id 2D8W).

330 Since the PDB databank is continuously expanding, the addition of new entries will  
331 make our analysis more robust. Snakemake pipeline was constructed by using Python  
332 programming language to provide the ability to reproduce the results of this study. The  
333 pipeline covers all stages of the current study from downloading of halide-bound  
334 protein structures from the PDB to generation of the output files and graphical output.  
335 The output is tab-separated file with information about each halide atom in PDB  
336 structures. Additionally, there is an anaconda-environment file, which provides  
337 instructions for the installation of the dependencies required for the workflow. The  
338 pipeline is available at  
339 [https://github.com/rostick/Halide\\_sites/blob/master/README.md](https://github.com/rostick/Halide_sites/blob/master/README.md)

340

## 341 **Discussion**

342 Halides are ubiquitous in the environment and impact all living organisms on our  
343 planet. Whereas some of them, such as chloride became a universal counterion for  
344 positively charged potassium and sodium, and contribute to build-up of  
345 electrochemical gradients, others, such as bromide and iodide play rather very defined  
346 roles, for example as strong antioxidants. Fluoride is mostly considered as a toxic  
347 compound, and many organisms developed fluoride expelling channels, and the only  
348 well documented case where it is actively used as an essential compound is the class  
349 of adenosyl-fluoride synthases, catalysing the formation of a carbon-fluorine bond, via  
350 connecting fluoride ion to S-adenosyl-L-methionine, with the concomitant release of  
351 L-methionine<sup>102</sup>. Furthermore, there are many other enzymes which catalyse the  
352 formation of numerous halogenated products, some of which are actively being  
353 investigated for their potent pharmacological properties<sup>103-105</sup>.

354 Considering all the importance of halides in biology, it is surprising that rather little is  
355 known about their binding to proteins, apart from the logical suggestion that negatively  
356 charged anions will be recruited to the positively charged side chains. Although our  
357 extensive analysis confirms this general observation, it additionally reveals certain  
358 patterns and preferences among four halides for its binding partners. For example, the  
359 occurrence of Arg side chain might be a general flag for the halide binding site,  
360 however the presence of other positively charged residues such as Lys and His, or  
361 polar residues can hint to the certain halide.

362 Based on amino acid composition of the binding sites, interaction lengths and angles,  
363 and coordination numbers it is possible to easily distinguish halides from anything else,  
364 however the exact identification of halide identity might still be problematic  
365 (Supplementary Figure 1). In cases when the anomalous data are present, such an  
366 identification can be done more properly. Nevertheless, the revealed patterns might  
367 be used to develop a prediction algorithm, which can be useful for an engineering of  
368 halide binding sites (e.g. for phasing purposes) and to improve automatic ligand  
369 assignment.

370 The constructed pipeline is the first step for the development of such an algorithm. Our  
371 present analysis of the dataset of halide-binding sites obtained by using this pipeline  
372 demonstrates its usefulness and opens an avenue for their more detailed future  
373 studies.

## 374 **Acknowledgements**

376 This work was financially supported by the Government of the Russian Federation  
377 through the ITMO Fellowship and Professorship Program and by the Grant 08-08

## Author contributions

R.S. and D.U. performed all calculations. All authors analyzed the data and wrote the manuscript.

## Competing interests

Authors declare no competing interests

## Data availability

All data reported in this research are available from the corresponding author on reasonable request.

## References

1. Shannon, R. D. IUCr. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr Sect A Cryst Phys Diffr Theor Gen Crystallogr* **32**, 751–767 (1976).
2. Wong, M. C. M. *et al.* Cochrane reviews on the benefits/risks of fluoride toothpastes. *J. Dent. Res.* **90**, 573–579 (2011).
3. Aoun, A., Darwiche, F., Hayek, Al, S. & Doumit, J. The Fluoride Debate: The Pros and Cons of Fluoridation. *Prev Nutr Food Sci* **23**, 171–180 (2018).
4. Kanduti, D., Sterbenk, P. & Artnik, B. FLUORIDE: A REVIEW OF USE AND EFFECTS ON HEALTH. *Materia Socio-Medica* **28**, 133–137 (2016).
5. Fincham, A. G., Moradian-Oldak, J. & Simmer, J. P. The Structural Biology of the Developing Dental Enamel Matrix. *Journal of Structural Biology* **126**, 270–299 (1999).
6. Buzalaf, M. A. R., Pessan, J. P., Honório, H. M. & Cate, ten, J. M. Mechanisms of Action of Fluoride for Caries Control. *Fluoride and the Oral Environment* **22**, 97–114 (2011).
7. Perumal, E., Paul, V., Govindarajan, V. & Panneerselvam, L. A brief review on experimental fluorosis. *Toxicology Letters* **223**, 236–251 (2013).
8. Weatherell, J. A. *Uptake and distribution of fluoride in bones and teeth and the development of fluorosis. Barltrop W, Burland WL: Mineral Metabolism in Paediatrics 1 ....* (1969).
9. Aoba, T. & Fejerskov, O. Dental fluorosis: chemistry and biology. *Crit. Rev. Oral Biol. Med.* **13**, 155–170 (2002).
10. Stockbridge, R. B., Robertson, J. L., Elife, L. K.-P. 2013. A family of fluoride-specific ion channels with dual-topology architecture. *cdn.elifesciences.org*
11. Stockbridge, R. B. *et al.* Crystal structures of a double-barrelled fluoride ion channel. *Nature* **525**, 548–551 (2015).

12. Brammer, A. E., Stockbridge, R. B. & Miller, C. F<sup>-</sup>/Cl<sup>-</sup> selectivity in CLCF-type F<sup>-</sup>/H<sup>+</sup> antiporters. *The Journal of General Physiology* **144**, 129–136 (2014).
13. Stockbridge, R. B. *et al.* Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *PNAS* **109**, 15289–15294 (2012).
14. Ren, A., Rajashankar, K. R. & Patel, D. J. Fluoride ion encapsulation by Mg<sup>2+</sup> ions and phosphates in a fluoride riboswitch. *Nature* **486**, 85–89 (2012).
15. Jie Qin, Geqing Chai, John M Brewer, Leslie L Lovelace, A.Lukasz Lebioda. Fluoride Inhibition of Enolase: Crystal Structure and Thermodynamics†,‡. *Biochemistry* **45**, 793–800 (2005).
16. Marquis, R. E., Clock, S. A. & Mota-Meira, M. Fluoride and organic weak acids as modulators of microbial physiology. *FEMS Microbiology Reviews* **26**, 493–510 (2003).
17. Li, L. The Biochemistry and Physiology of Metallic Fluoride: Action, Mechanism, and Implications:. *Critical Reviews in Oral Biology & Medicine* **14**, 100–114 (2016).
18. Murthy, K. S. & Makhlof, G. M. Fluoride activates G protein-dependent and -independent pathways in dispersed intestinal smooth muscle cells. *Biochem. Biophys. Res. Commun.* **202**, 1681–1687 (1994).
19. Mendoza-Schulz, A. *et al.* The effects of fluoride on cell migration, cell proliferation, and cell metabolism in GH4C1 pituitary tumour cells. *Toxicology Letters* **190**, 179–186 (2009).
20. Strunecka, A., Patocka, J., Blaylock, R. L. & Chinoy, N. J. Fluoride Interactions: From Molecules to Disease. *CST* **2**, 190–213 (2007).
21. Barbier, O., Arreola-Mendoza, L. & Del Razo, L. M. Molecular mechanisms of fluoride toxicity. *Chemico-Biological Interactions* **188**, 319–333 (2010).
22. Eisenberg, A. D. & Marquis, R. E. Enhanced transmembrane proton conductance in *Streptococcus mutans* GS-5 due to ionophores and fluoride. *Antimicrob. Agents Chemother.* **19**, 807–812 (1981).
23. Oldham, M. L. & Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15152–15156 (2011).
24. Toyoshima, C., Nomura, H. & Tsuda, T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* **432**, 361–368 (2004).
25. Saum, S. H. & Müller, V. Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies. *Saline Syst.* **4**, 4–15 (2008).
26. Lanyi, J. K. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriological Reviews* **38**, 272–290 (1974).
27. Oren, A. Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobius*. *Canadian Journal of Microbiology* **32**, 4–9 (2011).
28. Roeßler, M., Wanner, G. & Müller, V. Motility and Flagellum Synthesis in *Halobacillus halophilus* Are Chloride Dependent. *J. Bacteriol.* **182**, 532–535 (2000).
29. Roeßler, M. & Müller, V. Chloride dependence of glycine betaine transport in *Halobacillus halophilus*. *FEBS Lett.* **489**, 125–128 (2001).

30. Hänelt, I. & Müller, V. Molecular Mechanisms of Adaptation of the Moderately Halophilic Bacterium *Halobacillus halophilus* to Its Environment. *Life (Basel)* **3**, 234–243 (2013).
31. Galinski, E. A. Osmoadaptation in bacteria. *Adv. Microb. Physiol.* **37**, 272–328 (1995).
32. Wege, S., Gilliam, M. & Henderson, S. W. Chloride: not simply a 'cheap osmoticum', but a beneficial plant macronutrient. *J. Exp. Bot.* **68**, 3057–3069 (2017).
33. Guskov, A. *et al.* Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.* **16**, 334–342 (2009).
34. Kawakami, K., Umena, Y., Kamiya, N. & Shen, J.-R. Location of chloride and its possible functions in oxygen-evolving photosystem II revealed by X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8567–8572 (2009).
35. Brahmachari, U., Gonthier, J. F., Sherrill, C. D. & Barry, B. A. Chloride Maintains a Protonated Internal Water Network in the Photosynthetic Oxygen Evolving Complex. *J. Phys. Chem. B* **121**, 10327–10337 (2017).
36. Freeman, H. J. & Kim, Y. S. Digestion and absorption of protein. *Annu. Rev. Med.* **29**, 99–116 (1978).
37. Lu, P.-J. *et al.* Gastric juice acidity in upper gastrointestinal diseases. *World J. Gastroenterol.* **16**, 5496–5501 (2010).
38. Aghajari, N., Feller, G., Gerday, C. & Haser, R. Structural basis of alpha-amylase activation by chloride. *Protein Sci.* **11**, 1435–1441 (2002).
39. Yates, C. J. *et al.* Molecular and thermodynamic mechanisms of the chloride-dependent human angiotensin-I-converting enzyme (ACE). *J. Biol. Chem.* **289**, 1798–1814 (2014).
40. Prange, H. D., Shoemaker, J. L., Westen, E. A., Horstkotte, D. G. & Pinshow, B. Physiological consequences of oxygen-dependent chloride binding to hemoglobin. *J. Appl. Physiol.* **91**, 33–38 (2001).
41. Chen, J.-C. *et al.* WNK4 kinase is a physiological intracellular chloride sensor. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4502–4507 (2019).
42. Wolf-Watz, M., Bäckström, S., Grundström, T., Sauer, U. & Härd, T. Chloride binding by the AML1/Runx1 transcription factor studied by NMR. *FEBS Lett.* **488**, 81–84 (2001).
43. Jentsch, T. J. & Pusch, M. CLC Chloride Channels and Transporters: Structure, Function, Physiology, and Disease. *Physiol. Rev.* **98**, 1493–1590 (2018).
44. Argenzio, E. & Moolenaar, W. H. Emerging biological roles of Cl<sup>-</sup> intracellular channel proteins. *J. Cell. Sci.* **129**, 4165–4174 (2016).
45. Accardi, A. Structure and gating of CLC channels and exchangers. *J. Physiol. (Lond.)* **593**, 4129–4138 (2015).
46. Accardi, A. & Picollo, A. CLC channels and transporters: proteins with borderline personalities. *Biochim. Biophys. Acta* **1798**, 1457–1464 (2010).
47. Jentsch, T. J. Chloride and the endosomal-lysosomal pathway: emerging roles of CLC chloride transporters. *The Journal of physiology* **578**, 633–640 (2007).
48. Piwon, N., Günther, W., Schwake, M., Bösl, M. R. & Jentsch, T. J. CLC-5 Cl<sup>-</sup> channel disruption impairs endocytosis in a mouse model for Dent's disease. *Nature* **408**, 369–373 (2000).

- 514 49. Smith, A. J. & Lippiat, J. D. Direct endosomal acidification by the outwardly  
515 rectifying CLC-5 Cl<sup>-</sup>/H<sup>+</sup> exchanger. *J. Physiol. (Lond.)* **588**, 2033–2045  
516 (2010).
- 517 50. Pedersen, T. H., Riisager, A., de Paoli, F. V., Chen, T.-Y. & Nielsen, O. B.  
518 Role of physiological CLC-1 Cl<sup>-</sup> ion channel regulation for the excitability and  
519 function of working skeletal muscle. *The Journal of General Physiology* **147**,  
520 291–308 (2016).
- 521 51. Imbrici, P. *et al.* CLC-1 chloride channels: state-of-the-art research and future  
522 challenges. *Front Cell Neurosci* **9**, 156 (2015).
- 523 52. Imbrici, P. *et al.* CLC-1 mutations in myotonia congenita patients: insights into  
524 molecular gating mechanisms and genotype-phenotype correlation. *J.*  
525 *Physiol. (Lond.)* **593**, 4181–4199 (2015).
- 526 53. Yoshikawa, M. *et al.* CLC-3 deficiency leads to phenotypes similar to human  
527 neuronal ceroid lipofuscinosis. *Genes Cells* **7**, 597–605 (2002).
- 528 54. Poët, M. *et al.* Lysosomal storage disease upon disruption of the neuronal  
529 chloride transport protein CLC-6. *PNAS* **103**, 13854–13859 (2006).
- 530 55. Devuyst, O. & Thakker, R. V. Dent's disease. *Orphanet J Rare Dis* **5**, 28–8  
531 (2010).
- 532 56. Andrini, O. *et al.* CLC-K chloride channels: emerging pathophysiology of  
533 Bartter syndrome type 3. *Am. J. Physiol. Renal Physiol.* **308**, F1324–34  
534 (2015).
- 535 57. Cunha, T. D. S. & Heilberg, I. P. Bartter syndrome: causes, diagnosis, and  
536 treatment. *Int J Nephrol Renovasc Dis* **11**, 291–301 (2018).
- 537 58. Gururaja Rao, S., Ponnalagu, D., Patel, N. J. & Singh, H. Three Decades of  
538 Chloride Intracellular Channel Proteins: From Organelle to Organ Physiology.  
539 *Curr Protoc Pharmacol* **80**, 11.21.1–11.21.17 (2018).
- 540 59. Argenzio, E. *et al.* CLIC4 regulates cell adhesion and  $\beta$ 1 integrin trafficking.  
541 *J. Cell. Sci.* **127**, 5189–5203 (2014).
- 542 60. Jiang, L. *et al.* Intracellular chloride channel protein CLIC1 regulates  
543 macrophage function through modulation of phagosomal acidification. *J. Cell.*  
544 *Sci.* **125**, 5479–5488 (2012).
- 545 61. Ulmasov, B., Bruno, J., Gordon, N., Hartnett, M. E. & Edwards, J. C. Chloride  
546 intracellular channel protein-4 functions in angiogenesis by supporting  
547 acidification of vacuoles along the intracellular tubulogenic pathway. *Am. J.*  
548 *Pathol.* **174**, 1084–1096 (2009).
- 549 62. Berryman, M. & Bretscher, A. Identification of a novel member of the chloride  
550 intracellular channel gene family (CLIC5) that associates with the actin  
551 cytoskeleton of placental microvilli. *Mol. Biol. Cell* **11**, 1509–1521 (2000).
- 552 63. Littler, D. R. *et al.* The enigma of the CLIC proteins: Ion channels, redox  
553 proteins, enzymes, scaffolding proteins? *FEBS Lett.* **584**, 2093–2101 (2010).
- 554 64. Liu, F., Zhang, Z., Csanády, L., Gadsby, D. C. & Chen, J. Molecular Structure  
555 of the Human CFTR Ion Channel. *Cell* **169**, 85–95.e8 (2017).
- 556 65. Liu, F. *et al.* Structural identification of a hotspot on CFTR for potentiation.  
557 *Science* **364**, 1184–1188 (2019).
- 558 66. Hwang, T.-C. *et al.* Structural mechanisms of CFTR function and dysfunction.  
559 *The Journal of General Physiology* **150**, 539–570 (2018).
- 560 67. Berg, J., Yang, H. & Jan, L. Y. Ca<sup>2+</sup>-activated Cl<sup>-</sup> channels at a glance. *J.*  
561 *Cell. Sci.* **125**, 1367–1371 (2012).



- 562 68. Sabirov, R. Z. & Okada, Y. The maxi-anion channel: a classical channel  
563 playing novel roles through an unidentified molecular entity. *J Physiol Sci* **59**,  
564 3–21 (2009).
- 565 69. Osei-Owusu, J., Yang, J., Vitery, M. D. C. & Qiu, Z. Molecular Biology and  
566 Physiology of Volume-Regulated Anion Channel (VRAC). *Curr Top Membr*  
567 **81**, 177–203 (2018).
- 568 70. Hoeft, S. E., Rogers, D. R. & Visscher, P. T. Metabolism of methyl bromide  
569 and dimethyl sulfide by marine bacteria isolated from coastal and open  
570 waters†. *Aquatic Microbial Ecology* **21**, 221–230 (2000).
- 571 71. McCall, A. S. *et al.* Bromine is an essential trace element for assembly of  
572 collagen IV scaffolds in tissue development and architecture. *Cell* **157**, 1380–  
573 1392 (2014).
- 574 72. Thomas, E. L., Bozeman, P. M., Jefferson, M. M. & King, C. C. Oxidation of  
575 bromide by the human leukocyte enzymes myeloperoxidase and eosinophil  
576 peroxidase. Formation of bromamines. *J. Biol. Chem.* **270**, 2906–2913  
577 (1995).
- 578 73. Henderson, J. P., Byun, J., Mueller, D. M. & Heinecke, J. W. The eosinophil  
579 peroxidase-hydrogen peroxide-bromide system of human eosinophils  
580 generates 5-bromouracil, a mutagenic thymine analogue. *Biochemistry* **40**,  
581 2052–2059 (2001).
- 582 74. Barratt, T. M. & Walser, M. Extracellular fluid in individual tissues and in  
583 whole animals: the distribution of radiosulfate and radiobromide. *J. Clin.*  
584 *Invest.* **48**, 56–66 (1969).
- 585 75. van Leeuwen, F. X. & Sangster, B. The toxicology of bromide ion. *Crit. Rev.*  
586 *Toxicol.* **18**, 189–213 (1987).
- 587 76. Frances, C., Hoizey, G., Lamiabie, D., Millart, H. & Trenque, T. Bromism from  
588 daily over intake of bromide salt. *J. Toxicol. Clin. Toxicol.* **41**, 181–183  
589 (2003).
- 590 77. Lugassy, D. & Nelson, L. Case files of the medical toxicology fellowship at  
591 the New York City poison control: bromism: forgotten, but not gone. *J Med*  
592 *Toxicol* **5**, 151–157 (2009).
- 593 78. Saenko, G. N., Kravtsova, Y. Y., Ivanenko, V. V. & Sheludko, S. I.  
594 Concentration of Iodine and bromine by plants in the seas of Japan and  
595 Okhotsk. *Mar. Biol.* **47**, 243–250 (1978).
- 596 79. Küpper, F. C. *et al.* Emission of volatile halogenated compounds, speciation  
597 and localization of bromine and iodine in the brown algal genome model  
598 *Ectocarpus siliculosus*. *J. Biol. Inorg. Chem.* **23**, 1119–1128 (2018).
- 599 80. Kunka, A., Damborsky, J. & Prokop, Z. Haloalkane Dehalogenases From  
600 Marine Organisms. *Meth. Enzymol.* **605**, 203–251 (2018).
- 601 81. Almeida, M. *et al.* Vanadium haloperoxidases from brown algae of the  
602 Laminariaceae family. *Phytochemistry* **57**, 633–642 (2001).
- 603 82. Cabrita, M. T., Vale, C. & Rauter, A. P. Halogenated compounds from marine  
604 algae. *Mar Drugs* **8**, 2301–2317 (2010).
- 605 83. Dauter, M. & Dauter, Z. Phase determination using halide ions. *Methods Mol.*  
606 *Biol.* **364**, 149–158 (2007).
- 607 84. Dauter, Z., Dauter, M. & Rajashankar, K. R. Novel approach to phasing  
608 proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr. D*  
609 *Biol. Crystallogr.* **56**, 232–237 (2000).
- 610 85. Dauter, Z. & Dauter, M. Entering a new phase: using solvent halide ions in  
611 protein structure determination. *Structure* **9**, R21–6 (2001).

86. Dauter, Z. & Dauter, M. Anomalous signal of solvent bromides used for phasing of lysozyme. *J. Mol. Biol.* **289**, 93–101 (1999).
87. Eales, J. G. Iodine metabolism and thyroid-related functions in organisms lacking thyroid follicles: are thyroid hormones also vitamins? *Proc. Soc. Exp. Biol. Med.* **214**, 302–317 (1997).
88. Senese, R., Cioffi, F., de Lange, P., Goglia, F. & Lanni, A. Thyroid: biological actions of 'nonclassical' thyroid hormones. *J. Endocrinol.* **221**, R1–12 (2014).
89. Bernal, J., Guadaño-Ferraz, A. & Morte, B. Thyroid hormone transporters-functions and clinical implications. *Nat Rev Endocrinol* **11**, 690–690 (2015).
90. Darrouzet, E., Lindenthal, S., Marcellin, D., Pellequer, J.-L. & Pourcher, T. The sodium/iodide symporter: state of the art of its molecular characterization. *Biochim. Biophys. Acta* **1838**, 244–253 (2014).
91. DiStefano, J. J., de Luze, A. & Nguyen, T. T. Binding and degradation of 3,5,3'-triiodothyronine and thyroxine by rat intestinal bacteria. *Am. J. Physiol.* **264**, E966–72 (1993).
92. Cahnmann, H. J. & Funakoshi, K. 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 (1970).
93. Crockford, S. J. Evolutionary roots of iodine and thyroid hormones in cell-cell signaling. *Integr. Comp. Biol.* **49**, 155–166 (2009).
94. Küpper, F. C. *et al.* Commemorating two centuries of iodine research: an interdisciplinary overview of current research. *Angew. Chem. Int. Ed. Engl.* **50**, 11598–11620 (2011).
95. Küpper, F. C. *et al.* Iodide accumulation provides kelp with an inorganic antioxidant impacting atmospheric chemistry. *PNAS* **105**, 6954–6958 (2008).
96. Zimmermann, M. B. The role of iodine in human growth and development. *Seminars in Cell & Developmental Biology* **22**, 645–652 (2011).
97. Abendroth, J. *et al.* SAD phasing using iodide ions in a high-throughput structural genomics environment. *J. Struct. Funct. Genomics* **12**, 83–95 (2011).
98. Melnikov, I. *et al.* Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci Adv* **3**, e1602952 (2017).
99. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **5**, 189 (2016).
100. Ayala, R., Martínez, J. M., Pappalardo, R. R. & Marcos, E. S. On the halide hydration study: Development of first-principles halide ion-water interaction potential based on a polarizable model. *Journal of Chemical Physics* **119**, 9538–9548 (2003).
101. Paula, S., Volkov, A. G. & Deamer, D. W. Permeation of halide anions through phospholipid bilayers occurs by the solubility-diffusion mechanism. *Biophys. J.* **74**, 319–327 (1998).
102. O'Hagan, D., Schaffrath, C., Cobb, S. L., Hamilton, J. T. G. & Murphy, C. D. Biochemistry: biosynthesis of an organofluorine molecule. *Nature* **416**, 279–279 (2002).
103. Rocha, D. H. A., Seca, A. M. L. & Pinto, D. C. G. A. Seaweed secondary metabolites in vitro and in vivo anticancer activity. *Mar Drugs* **16**, 410 (2018).
104. Jesus, A., Correia-da-Silva, M., Afonso, C., Pinto, M. & Cidade, H. Isolation and potential biological applications of haloaryl secondary metabolites from macroalgae. *Mar Drugs* **17**, 73 (2019).

662 105. Kasanah, N. & Triyanto, T. Bioactivities of Halometabolites from Marine  
663 Actinobacteria. *Biomolecules* **9**, 225 (2019).  
664