Structural introspection of a putative fluoride transporter in plants

Aditya Banerjee¹ · Aryadeep Roychoudhury¹

Received: 3 December 2018 / Accepted: 10 February 2019 / Published online: 22 February 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

The xenobiotic fluoride is a detrimental toxin which is exported by a double-barrelled fluoride channel (Fluc) in prokaryotes and a single-barrelled fluoride exporter (FEX) in lower eukaryotes. The presence of *FEX* gene in higher plant species has not been reported till date. In this communication, we have identified FEXs across nine plant species belonging to five different families. Homology modelling predicted the sequential and structural conservation of crucial residues in plant FEXs. It was predicted that the N-terminal segments were the main target sites for phosphorylation. Other post-translational modifications included Arg methylation, Lys acetylation and Cys *S*-nitrosylation, all of which contribute to protein stability and function. FEXs were predicted to contain lipid moieties which anchor the transporters to the plasma membrane. The crystallizability along with associated properties of plant FEXs was also analyzed to facilitate future experimental analyses. The predicted RNA structure of *OsFEX* formed several pseudoknots, though no riboswitches could be detected from our prediction. The analysis of upstream regions of *FEX* in representative plants represented that this gene could be responsive to phytohormone-mediated signaling. This is the first novel report highlighting that *FEX* genes in plants are probably expressed in a tissue-specific pattern and possibly via a phytohormone-dependent pathway to encode functional FEX proteins during fluoride stress.

Keyword Fluoride transporter · Bioinformatics · Phylogram · Post-translational modifications · Gene ontology function · Intrinsic disorder · Homology modelling · Riboswitch · Promoter analysis

Introduction

Fluorine is the 13th most abundant element found on earth's crust. It acts as a xenobiotic even at sparingly low concentrations and deteriorates human health if ingested beyond the safe limit of 1.5 ppm as prescribed by the World Health Organization (WHO) (Choubisa 2013; Banerjee and Roychoudhury 2019). Groundwater and soil stretches across 20 developed and developing countries of Asia, Africa, Australia and South America have recorded alarming concentrations of fluoride (Susheela 1999; Farooqi et al. 2007; Nanyaro et al. 1984), severely deteriorating the quality of the ecosystem. Plants and crops cultivated in fluoride-infested soil or with fluoride-contaminated water exhibit gradual physiological damages such as necrotic lesions, burning,

chlorosis, leaf damage, and developmental and reproductive retardation (Hong et al. 2016). Crop plants also bioaccumulate fluoride in the edible parts. As a result, fluoride biomagnification is facilitated at higher trophic levels and the ecosystem biosafety is being challenged (Mondal and Gupta 2015; Yadu et al. 2018).

Unicellular organisms such as *Escherichia coli*, *Bordetella pertussis*, and *Saccharomyces cerevisiae* possess a functional transmembrane fluoride exporter in their cell membrane to ameliorate the toxic accumulation of the anion (Baker et al. 2012). These fluoride exporters belong to camphor-resistance-like B (CRCB) family. The fluoride exporter is phylogenetically distinguished as 'Fluc' in prokaryotes and 'FEX' in eukaryotes. Li et al. (2013) demonstrated that possibly due to constant contact with environmental fluoride, the *Fluc* gene is constitutively expressed in *E. coli*, enabling the bacteria to behave as fluoride-tolerant species. Double-knockout (KO) *E. coli* mutants of *Fluc* yielded a fluoride-hypersensitive strain (Li et al. 2013). The Fluc channels are homodimers with each monomer forming four transmembrane helices (TMs) and an active fluoride pore,

✉ Aryadeep Roychoudhury
aryadeep.rc@gmail.com

¹ Department of Biotechnology, St. Xavier's College (Autonomous), 30, Mother Teresa Sarani, Kolkata, West Bengal 700016, India

and the structure corresponds to a two-stave barrel (Berbasova et al. 2017). The monomeric sequence contains short N- and C-termini and a conserved sequence of GxGGxT-TFSTFxxE in TM3 (Stockbridge et al. 2012). The structure of Fluc is precisely unconventional as most ion channels form a barrel-stave plane where three to seven homologous subunits are built around a central axial pore in a direction perpendicular to the membrane plane (Stockbridge et al. 2013). The dual topology found in Flucs has been usually observed in the small multidrug resistance (SMR) family of proton-coupled multidrug antiporters (Morrison et al. 2012). Interestingly, even the oppositely oriented repeats found in Flucs are unprecedented among ion channels. Such unique molecular architecture can be attributed to gene-pair duplications with the intermittent capture of a TM acting as an inversion linker. Thus, the antiparallel topology among the subunits remains conserved (Stockbridge et al. 2013).

Functional FEX channels have been reported in eukaryotic organisms such as *S. cerevisiae*, *Neurospora crassa* and *Candida albicans* (Li et al. 2013). These are monomers formed due to the covalent fusion of two subunits/domains (Berbasova et al. 2017). The FEX double-KO strains exhibited 200-folds higher sensitivity to fluoride, but not to other halides, and the mutants could not even tolerate the fluoride concentration in normal tap water (Li et al. 2013). Smith et al. (2015) reported that the yeast FEX proteins, FEX1 (NM_001183810) and FEX2 (NM_001184093), are constitutively produced in the plasma membrane. The FEX proteins contain four conserved motifs: ooFooX, ooXooFxxE, PxGTxxxN and YxxxS/T ('X/x' can be any amino acid; 'o' is either Ser or Thr) in the hydrophobic helices TM3, TM8, TM7 and TM9 respectively. A thorough structural and mutation analysis confirmed that unlike Flucs, FEX proteins have only one functional fluoride export pore (Pore II) formed by TM3 (TM3a and TM3b associated via a small linker), TM8 (TM8a and TM8b associated via a small linker), TM7 and TM9. The other pore (Pore I formed by TM1, TM2, TM4, TM5 and TM6) is a non-functional vestigial remnant of homologous gene fusion. The evolutionarily conserved residues in yeast FEX are all associated with the proper functioning of Pore II. Apart from these conserved sequences, there can be a GxxxR sequence in the first helix and one N in the middle of the second helix of each FEX domain (Berbasova et al. 2017).

FEX transporters can be promising molecular targets in plants to overcome fluoride susceptibility as well as to reduce alarming levels of fluoride accumulation in edible parts. However, this protein is totally uncharacterized in higher plants. In this communication, we have computationally analyzed the FEX proteins in nine species across five plant families. The FEX from rice (*Oryza sativa* var. japonica) was selected as a representative member for modelling and RNA structure analysis, since rice is a model plant for

both genomic and epigenomic studies in monocots (Banerjee and Roychoudhury 2018a). Identification of *cis*-acting elements via analysis of promoter region of FEX genes in representative plant species provided an insight to the inducible nature and tissue-specific expression pattern of FEX in plants. We expect our observations to play a trailblazing role in the field of fluoride transporter research that will develop into a scaffolding platform for future scientific endeavors regarding plant FEXs.

Materials and methods

Finding plant FEXs

The BLASTp search of yeast FEX1 (NM_001183810) against "Genes in Rice TIGR pseudomolecules-Proteins" in The Institute of Genomic Research (TIGR) Rice Genome Annotation Project-Web BLAST Server (Kawahara et al. 2013) identified *O. sativa* var. japonica FEX (OsFEX) as an expressed, hypothetical protein. The NCBI-BLASTp against "Non-redundant protein sequences" in plant families such as Brassicaceae, Fabaceae, Cucurbitaceae and Vitaceae were used to identify the FEX sequences which exhibited > 80% sequence coverage. The information regarding FEX gene in the monocot model, rice and dicot model, *Arabidopsis*, was extracted from TIGR and The *Arabidopsis* Information Resource (TAIR) databases, respectively (Kawahara et al. 2013; Berardini et al. 2015).

Multiple sequence alignment

The FEX protein sequences from the plants and *S. cerevisiae* were aligned, using Clustal Omega in Clustal W2 suite (McWilliam et al. 2013) and the alignment file was colored by sequence similarity using ESPrpt 3.0 (Robert and Gouet 2014).

Phylogram

The phylogenetic tree was prepared using Phylogeny.fr (Dereeper et al. 2008). Stringent selection (not allowing many contiguous non-conserved positions), approximate likelihood ratio test using chi-square-based parametric and tree rendering using TreeDyn were utilized to build the phylogram (Dereeper et al. 2008).

Prediction of post-translational modifications in plant FEXs

Ser, Thr and Tyr phosphorylation were predicted using NetPhos 3.1 Server (Blom et al. 1999) and DISPHOS 1.3 (Iakoucheva et al. 2004). The possible Arg methylations,

Lys acetylations, Cys *S*-nitrosylation and addition of fatty acid residues in selected plant FEXs were predicted using PRmePRed (Kumar et al. 2017), PAIL (<http://bdmpail.biocuckoo.org/>), iSNO-PseAAC (Xu et al. 2013) and GPS-Lipid (<http://lipid.biocuckoo.org/>), respectively.

Intrinsic disorder profiling and functional characterization of OsFEX protein

The intrinsic disorder of the rice FEX protein was deciphered using DISOPRED3 in the PSIPRED Protein Sequence Analysis Workbench (Jones and Cozzetto 2015) and DisEMBL (Linding et al. 2003). The functional and evolutionary significance of rice FEX was extracted from QuickGO, EMBL-EBI (<https://www.ebi.ac.uk/>) by using GO:1903425 (UniProtKB) as the input. The domain nomenclature of the sequence was retrieved from Pfam, EMBL-EBI (<http://pfam.xfam.org/>).

Homology modelling of OsFEX

Homology modelling was performed using SWISSMODEL BETA Server (Biasini et al. 2014) using N- and C-terminal domain sequences of rice (*O. sativa* var. japonica) FEX as separate targets, and the dual topology fluoride ion channels, PDB: 5A43 and PDB: 5A40 (Stockbridge et al. 2015) as the respective templates (based on sequence coverage and identity) for the N- and C-terminal domains. The N- and C-terminal domain sequences of rice FEX used as targets were selected from the target–template alignments in Clustal Omega (McWilliam et al. 2013). The selected N-terminal domain of rice FEX target ranged between Leu184 and Gly294. The selected C-terminal domain of rice FEX target ranged between His363 and Val493. The root mean square deviation (RMSD) values and TM-scores were calculated using TM-align to develop confidence regarding the accuracy of the constructed models (Zhang and Skolnick 2005). Each rice FEX model was aligned with one separate monomer in the crystal structure of Fluc from *Bordetella pertussis* (PDB: 5NKQ) (Stockbridge et al. 2015) in PyMOL (DeLano 2002). The coordinates of the fluoride and Na⁺ ions from PDB: 5NKQ was used to estimate their location in rice FEX model. The possible occurrence of disulphide bonds was

predicted using Dipro 2.0 (Cheng et al. 2006). To anticipate the interaction between the chains in the predicted model, protein–protein interface statistics in the templates were viewed in PDBsum, EMBL-EBI (Laskowski et al. 2018). The protein contact potential of the predicted models was identified by using vacuum electrostatics in PyMOL. The solvent accessible surface area was calculated using PyMOL (DeLano 2002). The solvent accessibility of OsFEX was predicted using ACCpro (Magnan and Baldi 2014).

Prediction of crystallizability of OsFEX

The solubility of OsFEX upon overexpression in *E. coli* was predicted using SOLpro (Magnan et al. 2009). The possibility of OsFEX to be crystallized for future structural studies was determined using XtalPred Server (Slabinski et al. 2007). The crystallizable probability based on length, isoelectric point, gravity index, longest disorder region, instability index, percentage of coil structures, surface entropy, surface hydrophobicity, surface ruggedness, coiled coils and insertion score of OsFEX was calculated. The final cumulative predictions regarding crystal formation was depicted using “expert pool crystallizability class” and the advanced “random forest crystallizability class”.

Predicting RNA structure and involvement of riboswitch and small RNA-mediated regulation in OsFEX

The RNA structure of *OsFEX* was predicted using RNAs-structure (Reuter and Mathews 2010). The Riboswitch finder (Bengert and Dandekar 2004) was used to identify and understand the interplay of any riboswitch during *OsFEX* RNA processing and translation. For further analysis, the genomic DNA sequence of OsFEX (downloaded from TIGR) was used as a query in Denison Riboswitch Detector (DRD) (Havill et al. 2014). The DRD detects riboswitches belonging to cobalamin, lysine, SAM-II, SAM-IV, PreQ1, FMN, glmS, purine, TPP, glycine, SAM-I, ykkC-yxkD and yybP-ykoY classes from genomic DNA sequences. The KineFold (Xayaphoummine et al. 2005) was used to predict the formation of pseudoknots for every 400 nucleotides/bases in the *OsFEX* RNA. The structures with least free

Table 1 Characteristics and structure of *FEX* gene in two model species (TIGR, TAIR)

Species	Chromosome Number	Chromosomal position	Gene ID/locus	Number of exons	Predicted protein length	Predicted pI	Predicted protein molecular weight (kDa)
<i>O. sativa</i>	10	22,466,999–22,473,150	LOC_Os10g41749	6	499	9.0433	54.0853
<i>A. thaliana</i>	2	17,397,270–17,400,081	LOC_At2g41705	7	461	7.53	50.4724

Fig. 1 Multiple sequence alignment using FEX protein sequences from *Saccharomyces cerevisiae* (yeast), *Oryza sativa* var. japonica, *Arabidopsis thaliana*, *Brassica oleracea*, *Brassica rapa*, *Medicago truncatula*, *Lupinus angustifolius*, *Phaseolus vulgaris*, *Cucumis sativus* and *Vitis vinifera*. The nomenclature of domains 1 and 2 has been made following the *S. cerevisiae* FEX sequence according to Berbasova et al. (2017). The residues marked with ‘asterisk’ are evolutionarily conserved among which the ones within the ooFoGxxxK (domain 1); PxGTxxxN, ooX-ooFxxE and YxxxS/T (domain 2) are crucial for the fluoride transport function of the channel protein

Fig. 1 (continued)

energy were represented. The involvement of any small RNA (sRNA) in regulating *OsFEX* post-transcriptional functioning was tested using psRNATarget (<http://plantgrn.noble.org/psRNATarget/>).

Promoter analysis of *FEX* gene in representative plant species

The ~1 Kbp upstream region of *FEX* gene from *O. sativa* var. japonica was downloaded from the Genome Browser in TIGR Rice Genome Annotation Project-Web BLAST Server (Kawahara et al. 2013). The ~1 Kbp upstream sequences of

Arabidopsis thaliana, *Medicago truncatula* and *Vitis vinifera* were downloaded from the Genome Browsers in PlantGDB (Dong et al. 2004). The promoter sequence of *Cucumis sativus*, the representative member of Cucurbitaceae was unavailable due to the presence of draft genome sequence in PlantGDB. The presence of the conserved *cis*-acting motifs (considering both + and – strands) was tabulated by analyzing the promoter sequences in PLACE-Plant *Cis*-Acting Regulatory DNA Elements (Higo et al. 1999).

Table 2 Inter-kingdom/family conservation of crucial FEX residues

Families	Saccharo- myceta- ceae	Poaceae	Brassicaceae/ Cruciferae	Fabaceae/ Leguminosae			Cucurbitaceae	Vitaceae	Corresponding mutations in yeast FEX lowered IC ₅₀ by (Berbasova et al. 2017)		
Regions	Yeast	<i>O. sativa</i>	<i>A. thaliana</i>	<i>B. oleracea</i>	<i>B. rapa</i>	<i>M. truncatula</i>	<i>L. angustifolius</i>	<i>P. vulgaris</i>		<i>C. sativus</i>	<i>V. vinifera</i>
PxGTxxxN	Pro270	Pro404	Pro366	Pro353	Pro347	Arg352	Pro365	Pro340	Pro373	Pro378	1.1X (P270A)
	Gly272	Gly406	Gly368	Gly355	Gly349	Gly354	Gly367	Gly342	Gly375	Gly380	N/S
	Thr273	Thr407	Thr369	Thr356	Thr350	Thr355	Thr368	Thr343	Thr376	Thr381	120X (T273V)
	Asn277	Asn411	Asn373	Asn360	Asn354	Asn359	Asn372	Asn347	Asn380	Asn385	2.7X (N277Q)
YxxxS/T	Tyr345	Tyr473	Tyr435	Tyr422	Tyr416	Tyr421	Tyr434	Tyr409	Tyr442	Tyr447	910X (Y345A)
	Ser349	Thr477	Thr439	Thr426	Thr420	Thr425	Thr438	Thr413	Thr446	Thr451	1.2X (S349A)
ooFooXXXXE	Ser94	Thr255	Thr224	Thr204	Thr204	Thr206	Thr221	Thr194	Thr238	Thr232	1000X (S94V:S95V)
	Ser95	Thr256	Thr225	Thr205	Thr205	Thr207	Thr222	Thr195	Thr239	Thr233	
	Phe96	Phe257	Phe226	Phe206	Phe206	Phe208	Phe223	Phe196	Phe240	Phe234	1.1X (F96A)
	Ser97	Ser258	Ser227	Ser207	Ser207	Ser209	Ser224	Ser197	Ser241	Ser235	7.9X (S97V:S98V)
ooXooFFXXE	Ser98	Gly259	Gly228	Gly208	Gly208	Gly210	Gly225	Gly198	Gly242	Gly236	IC₅₀ increased by: 1.1X (E102A)
	Glu102	Lys263	Lys232	Lys212	Lys212	Lys214	Lys229	Lys202	Lys246	Lys240	
	Ser325	Ser449	Ser411	Ser398	Ser392	Ser397	Ser410	Ser385	Ser418	Ser423	1000X (S325V:T326V)
	Thr326	Thr450	Thr412	Thr399	Thr393	Thr397	Thr411	Thr386	Thr419	Thr424	
	Ser328	Ser452	Ser414	Ser401	Ser395	Ser399	Ser413	Ser388	Ser421	Ser426	72X (S328V:T329V)
	Thr329	Thr453	Thr415	Thr402	Thr396	Thr400	Thr414	Thr389	Thr422	Thr427	
	Phe330	Phe454	Phe416	Phe403	Phe397	Phe402	Phe415	Phe390	Phe423	Phe428	1.4X (F330A)
	Glu333	Glu457	Glu419	Glu406	Glu400	Glu405	Glu418	Glu393	Glu426	Glu431	88X (E333A)

F96A:F330A double mutant in yeast lowered IC₅₀ by > 1000XE102A:E333A double mutant in yeast lowered IC₅₀ by > 1000X

N/S not studied

Fig. 2 The phylogram represents the evolutionary origin and relatedness among the FEX proteins in selected plant species. The horizontal lines are branches showing evolutionary lineages changing over time. The scale bar at the bottom of the figure represents the length of the branch that corresponds to an amount of genetic change of ‘1.0’. The unit of branch length is the number of nucleotide substitutions

in a particular FEX sequence divided by the length of the sequence. The length of the vertical dimension is irrelevant. The numbers next to each node is a statistical measure of support (ranging between 0 and 1) for evolutionary divergence. It has been calculated by approximate likelihood ratio test (aLRT: using chi-square-based parametrics) and tree rendering

Table 3 NetPhos 3.1 Server: Ser, Thr and Tyr phosphorylation sites with confidence score above 0.5 (threshold)

Species	Serine	Threonine	Tyrosine
<i>O. sativa</i>	43	16	4
<i>A. thaliana</i>	42	15	2
<i>B. oleracea</i>	28	16	2
<i>B. rapa</i>	24	16	2
<i>M. truncatula</i>	16	12	3
<i>L. angustifolius</i>	35	11	2
<i>P. vulgaris</i>	30	13	2
<i>C. sativus</i>	31	15	5
<i>V. vinifera</i>	36	12	1

Results

Identified plant FEXs

The BLASTp searches with yeast FEX1 identified the selected FEX protein sequences of *O. sativa* var. japonica (Locus: LOC_Os10g41749 and UniProtKB accession: Q108X7) [Poaceae], *Arabidopsis thaliana* (Gene: At2g41705 and UniProtKB accession: Q8RYE2), *Brassica oleracea* (NCBI Reference Sequence: XP_013631222.1), *Brassica rapa* (NCBI Reference Sequence: XP_009142769.1) [Brassicaceae], *Medicago truncatula* (NCBI Reference Sequence: XP_003601083.1), *Lupinus angustifolius* (NCBI Reference Sequence: XP_019414963.1), *Phaseolus vulgaris* (NCBI Reference Sequence: XP_007153147.1) [Fabaceae], *Cucumis sativus* (NCBI Reference Sequence: XP_011655170.1) [Cucurbitaceae] and *Vitis vinifera* (NCBI

Reference Sequence: XP_010664242.1) [Vitaceae]. Table 1 highlights the information regarding gene structure and gene products in the FEXs of model plants, *O. sativa* and *Arabidopsis*.

Multiple sequence alignment revealed the presence of conserved sequence stretches among plant FEXs

Figure 1 represents the multiple sequence alignment (MSA) of FEX sequences from yeast and selected plants. The motifs, GxxTR, ooFooX, ooXooFxxE, PxGTxxxN and YxxxS/T, associated with FEX were found to be conserved in all plants with minor exceptions. According to Berbasova et al. (2017), a conserved Asn (corresponding to N56 in yeast FEX) must be present in the middle of TM2. The plant FEX proteins also contain a conserved Asn (marked with ‘*’ in Fig. 1) which lies close to N56 of yeast FEX1. Actually, TM2 forms the non-functional Pore I (Berbasova et al. 2017). So, in spite of the sequence shift of this particular Asn in plant FEXs (with respect to yeast FEX1), fluoride transport activity should not be hampered. Following the sequence similarity detected in MSA, Table 2 has been constructed to compare similarly localized residues among plant and yeast FEXs, and the effect of mutation of these residues in yeast, according to Berbasova et al. (2017).

Phylogram

The MSA analysis showed that though the plant and yeast FEXs contain some crucial conserved motifs, they possess significant differences in sequences. The development of a phylogenetic tree revealed the evolutionary relatedness among different plant FEXs (Fig. 2). The phylogram

Statistics

Number of phosphorylated serines: 13 out of 57 (22.807%)
 Number of phosphorylated threonines: 0 out of 24 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 14 (0.000%)

Statistics

Number of phosphorylated serines: 11 out of 56 (19.643%)
 Number of phosphorylated threonines: 0 out of 28 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 11 (0.000%)

Statistics

Number of phosphorylated serines: 1 out of 45 (2.222%)
 Number of phosphorylated threonines: 0 out of 30 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 11 (0.000%)

Statistics

Number of phosphorylated serines: 1 out of 42 (2.381%)
 Number of phosphorylated threonines: 0 out of 29 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 11 (0.000%)

Statistics

Number of phosphorylated serines: 1 out of 35 (2.857%)
 Number of phosphorylated threonines: 0 out of 23 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 12 (0.000%)

Statistics

Number of phosphorylated serines: 6 out of 53 (11.321%)
 Number of phosphorylated threonines: 0 out of 21 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 9 (0.000%)

Statistics

Number of phosphorylated serines: 6 out of 44 (13.636%)
 Number of phosphorylated threonines: 0 out of 25 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 8 (0.000%)

Statistics

Number of phosphorylated serines: 2 out of 49 (4.082%)
 Number of phosphorylated threonines: 0 out of 24 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 14 (0.000%)

Statistics

Number of phosphorylated serines: 8 out of 53 (15.094%)
 Number of phosphorylated threonines: 0 out of 21 (0.000%)
 Number of phosphorylated tyrosines: 0 out of 11 (0.000%)

Fig. 3 DISPHOS scores of plant FEX proteins show that phosphorylation occurs only at Ser residues located towards the N-terminus. In the closely related *B. oleracea* and *B. rapa* FEXs, a single Ser phosphorylation was detected towards the middle of the sequence. Since DISPHOS 1.3 predicts phosphorylation in disordered regions of the protein, it is proposed that the single Ser phosphorylation in the two FEX sequences are probably located within an unstructured loop or coil situated in between successive helices

showed that among the selected sequences, *M. truncatula* FEX has the earliest origin. FEX among the members of Brassicaceae are closely related with *B. oleracea* and *B. rapa* FEXs having the most recent origin. *O. sativa* FEX exhibited the maximum amount of genetic change among the other plant FEXs.

Post-translational modifications in plant FEXs

NetPhos 3.1 Server detected a large number of Ser and Thr, and comparatively lesser number of Tyr phosphorylations in plant FEXs. The probable phosphorylations with confidence score above the threshold of 0.5 were represented (Table 3). DISPHOS 1.3 on the other hand predicted only Ser phosphorylations especially towards the N-terminal region of plant FEXs except in phylogenetically closely related *B. oleracea* and *B. rapa* FEX sequences (Fig. 3). Table 4 shows the large number of Arg methylations (confidence scores higher than 0.9 except in *M. truncatula* due to unavailability), Lys acetylations (confidence scores higher than 1.0) and Cys S-nitrosylation among plant FEX sequences. Being membrane proteins, each plant FEX revealed potential fatty acid modifications at target peptide sequences (Table 5).

Intrinsic disorder and ancestor chart of OsFEX

We next analyzed the intrinsic disorder profile of the OsFEX (as a representative example) and found that mostly the residues from position 1 to 170 (N-terminus) formed disordered secondary structures (coils) (Fig. 4a). The DisEMBL predicted the periodic presence of loops or coils (lacking regular secondary structure) within OsFEX. The disorder probability calculated based on Remark465 (regions that could lack electron density in crystal structure) and hot loops (highly mobile loops) also corroborated with the disordered regions formed by the loops or coils. The probability of occurrence of hot loops was much lower than the cutoff threshold of 0.4 (Fig. 4b). OsFEX was tagged with the Gene Ontology accession, GO:1903424 in UniProtKB. The gene ontology ancestor chart from QuickGO, EMBL-EBI predicted that OsFEX is an anion transmembrane transporter localized in the plasma membrane (Fig. 4c). Analysis in Pfam, EMBL-EBI revealed the presence of two CRCB domains ranging

between amino acid numbers 176–295 and 368–489 in the OsFEX sequence (Fig. 4d).

Homology model of FEX protein from *O. sativa* var. japonica

Figure 5a shows the double-barrelled Fluc in *B. pertussis* with two functional fluoride export pores and a centrally placed Na^+ . On the contrary, the rice FEX model is depicted in Fig. 5b with the conserved Phe257 and Phe454 residues (from Table 2) present in the functional Pore II and participating in fluoride transport. Pore I in eukaryotic FEXs is non-functional and evolutionarily vestigial. TM-align between rice FEX N-terminal domain model and PDB: 5A43 revealed a RMSD value of 0.74 and TM-score of 0.88714 (after normalization by the length of the N-terminal domain). Similarly, TM-align between rice FEX C-terminal domain model and PDB: 5A40 revealed a RMSD value of 0.33 and TM-score of 0.91066 (after normalization by the length of the C-terminal domain). Small RMSD values represent minute deviations between the backbone $\text{C}\alpha$ atoms in the model and the template. TM-scores lying between 0.5 and 1.0 ascertain that the structural coordinates are in the same fold in SCOP/CATH (Zhang and Skolnick 2005). Figures 6 and 7 depict the interactions of OsFEX amino acid residues with fluoride and sodium ions, respectively.

Two disulfide bonds were predicted between Cys210–Cys221 and Cys273–Cys349. The interface statistics of template 5A43 showed an interface area of 1638:1644 $\text{\AA}^2$ between the CRCB domain containing chains A and B. 14 hydrogen bonds and 165 non-bonded contacts could be identified. The interface statistics of template 5A40 showed an interface area of 1667:1668 $\text{\AA}^2$ between the CRCB domain containing chains A and B. In this template structure, 17 hydrogen bonds and 178 non-bonded contacts could be identified. Both the template structures did not record any salt bridges or disulphide bonds at least between the chains A and B. These data help to anticipate the presence of significant interactive area stabilized by hydrogen and disulphide bonds, and non-bonded contacts in the predicted OsFEX model as well. The solvent accessible surface areas of N-terminal and C-terminal models of OsFEX were calculated as 9284.502 $\text{\AA}^2$ and 10446.502 $\text{\AA}^2$. Both N- and C-terminal models of OsFEX recorded positively (blue) and negatively (red) charged residues at the ends of the α -helices where the loops or coils existed (Fig. 5c, d). ACCpro also predicted the residues within these loops and coils to be exposed and relatively accessible to solvents at 25% exposed threshold.

Crystallizability of *O. sativa* FEX

The OsFEX was predicted to be insoluble upon heterologous overexpression in *E. coli* with a probability of 0.737985. Multiple parameters were used to assess crystallizability of OsFEX. Length, isoelectric point, gravity index, predicted structural disorder and surface hydrophobicity of OsFEX were not in the optimum crystallizable range (Fig. 8a–d, h). As a result, according to “expert pool” analysis, OsFEX occupied class 5 (least crystallizable). However, “random forest” analysis is an improved version, which uses protein surface entropy, surface ruggedness, instability index, percentage of coil structures, coiled coils and insertion score (Fig. 8e–g, i–k), for predicting the crystallizability class. Since these parameters of OsFEX were in the optimum crystallizable range, “Random forest” analysis revealed that OsFEX protein occupied class 1 (highest probability of being crystallizable) (Fig. 8m).

Structure and regulation of *O. sativa* FEX RNA

The *OsFEX* RNA structure with least free energy ($-492.3 \text{ Kcal mol}^{-1}$) is represented in Fig. 9a. The riboswitch scanning softwares and psRNATarget could not identify any riboswitch and sRNA interaction, respectively. Interestingly, a large number of pseudoknots were predicted for *OsFEX* RNA as represented in Fig. 9b–e. For every 400 bases as input, an average of 11,000,000 ms was utilized to estimate RNA structural folding and formation of pseudoknots.

Identification of *cis*-acting elements in the promoters of plant FEXs

Though bioinformatic characterization of DNA sequence motifs might not justify a functionally active site, the *in silico* analysis of the probable *cis*-acting elements in ~ 1 Kbp upstream region of representative plant FEXs provides an idea regarding the nature of stress signals that might regulate their expressions (Table 6). A number of motifs such as TATA, GATA, CAAT, ROOTMOTIFTAPOX1, POLLEN-1LELAT52, G-box, EBOXBNNAPA, WRKY71OS, W-box, PYRIMIDINEBOX, GT1CONSENSUS, OSE2ROOTNODULE and CIRCADIANLELHC were detected in all representative FEX promoters. These *cis*-acting elements provide an insight regarding the inducible nature and tissue-specific expression of FEX gene in plants.

Discussion

Plant FEXs possess conserved functional stretches

Fluoride is an ecological pollutant found abundantly in the earth's crust. Alarming levels of fluoride have been detected

in water sources around the world. As a result, water-intensive crops and plants cultivated on fluoride-infested soils experience systemic physiological abnormalities and damages. In this regard, identification and computational characterization of a potential fluoride transporter, FEX, is crucial. We identified FEX sequences having > 80% mutual sequence coverage from nine plant species belonging to five different major plant families. Due to ubiquitous fluoride contamination, it might be possible that FEX channels exist in all organisms. However, due to very low sequence identity (< 20%), current bioinformatic searches might be unsuccessful in detecting them in higher organisms (Berbasova et al. 2017). The MSA analysis detected crucial conserved stretches among yeast and plant FEXs which are quintessential for fluoride transport activity. Two further interesting observations were documented. Berbasova et al. (2017) noted a conserved ooFooxxxE motif among bacterial Flucs and yeast FEX1. However, instead of being a Ser or Thr, the last 'o' has been changed to a Gly in plant FEXs. Even the end Glu has been converted to an oppositely charged amino acid, viz., Lys (Fig. 1). Thus, the motif in plant FEXs has been altered to ooFoGxxxK. Table 2 shows that simultaneous mutations at both Ser97 and Ser98 (corresponding to ooFooxxxE) lowered the lethal dose (IC_{50}) by 7.9-fold in yeast. The change of the last Ser to Gly in plant FEXs indicates that single mutation of Ser98 in yeast FEX1 possibly would not harm fluoride transport activity. Interestingly, single mutation of Glu102 (corresponding to ooFooxxxE) with a normal Glu333 (corresponding to ooXooFxxE) in yeast FEX1 increased the IC_{50} by about 1.1-fold (Table 2). Thus, this residue subtly suppresses the activity of wild-type FEX1. Glu102 was also found to be located in the vestigial Pore I (Berbasova et al. 2017). Hence, substitution of Glu in the ooFooxxxE motif with Lys might actually promote the fluoride transport activity of the channel. However, the importance of a conserved Glu in the ooXooFxxE motif in plants is highly valid as it was observed that simultaneous mutations at both Glu102 and Glu333 in yeast FEX1 lowered the IC_{50} by > 1000-fold (Table 2).

The PxGTxxxN motif is conserved among all FEXs except in that of *M. truncatula*. However, mutation of Pro270 (corresponding to PxGTxxxN) in yeast FEX1 lowered the IC_{50} by only 1.1-fold and hence it can be inferred that the change of Pro to Arg might not interfere with FEX activity (Table 2). In the YxxxS/T motif, the end Ser in yeast FEX was substituted by Thr in all the plant FEXs, thus keeping the motif conserved. Each Phe residue in ooFooX and ooXooFxxE is very crucial, as each faces the inner lining of Pore II and indirectly interacts with a fluoride ion. These Phe residues correspond to those of the GxGGxTTESTExxE motif in prokaryotic Flucs. Simultaneous mutations at these Phe residues (Phe96 and Phe330) in yeast FEX1 lowered the

Table 4 Prediction of post-translational modifications in FEX proteins across representative species

Species	Post-translational modifications							
	Methylation			Acetylation			S-nitrosylation	
	Sequence	Arg no	Score	Sequence	Lys no	Score	Sequence	Cys no
<i>O. sativa</i>	DSAGHSLGAR- SASILSRRS	34	0.948	APVSAMKPKDTNV	144	1.05	GERLRGWILKCIIEKN- SIGSK	308
	GARSASILSRRSSRQG- SRG	41	0.959	PASSSIKVEPYKL	159	1.66		
	ARSASILSRRSSRQG- SRGS	42	0.974	KNSIGSKCGWEDW	318	1.24		
	ASILSRRSSRQGSRGISL	45	0.979	GGIGIGKRRHLKW	396	1.13		
	SRRSSRQGSRG- SISLSREM	49	0.988					
	GSRGSISLSREMGSILSS	56	0.964					
<i>A. thaliana</i>	VSEAGDIGDRSLRRRH- SAG	54	0.938	ASPLPTKSLLSPE	116	1.49	VLPKSLEYISCLIH LAVF- GIF	143
	DIGDRSLRRRHSA- GRSSRL	59	0.946	RASSEDKFHCLKV	283	1.20	RRASSEDKFHCLKVNT- FQSHI	286
	SLRRRHSAGRSSRL- SADDF	64	0.947	LVKEFDKGTSEAQ	323	1.00		
	RRHSAGRSSRLSADD- FIEQ	67	0.936	AALATLKKS VNTR	387	2.28		
	RWFLARLNGRGLGKD- SQNL	353	0.972					
<i>B. oleracea</i>	RWFLARLNGRGLG- KNGQYL	340	0.967	XXMDSIKSHEEPI	5	5.12	VLPKSLEYISCLAHL- AVFGIL	123
				GSDSVSKAGDTGD	49	1.29	RASSEEKKQSCLKAN- TYKRNI	273
				RASSEEKKQSCLK	269	1.67	GTSKAQLWLGCLVAAT- GVWIR	321
				ASSEEKKQSCLKA	270	1.88		
				KKQSCLKANTYKR	275	2.72		
				LVKEFDKGTSKAQ	310	1.63		
<i>B. rapa</i>				AALATLKKSVDTP	374	1.92		
	RWFLARLNGRGLG- KNGQYL	334	0.968	XXMDSIKSHEEPI	5	5.12	RASSEEKKQSCLKA- NAYKRNI	267
				RASSEEKKQSCLK	263	1.67	GTSKAQLWLG- CLVAAPGVWLR	315
				ASSEEKKQSCLKA	264	1.88		
				KKQSCLKANAYKR	269	2.70		
				LVKEFEKGTSKAQ	304	1.58		
<i>M. truncatula</i>				AALATLKKSVDTP	368	1.92		
	TSSWGS AVSRLSVCQN- VGV	13	0.843	AFQGRRKSES GGF	46	1.62	VGDDEAEDHECFSA- GDIAFQ	32
				ISKVSWKVD SFPR	275	1.10	KIPKLLEYVSC LTHLAVF- GIL	125
				AALATLKKS VNNTT	373	2.14	FGFVFKRDISCVS- EHLAIGLT	188
							RWFLATRLNE- CEIGGGFLKWL	341
							TLKKS VNNTTNC DTVIT- GLQFG	381

Table 4 (continued)

Species	Post-translational modifications							
	Methylation			Acetylation			S-nitrosylation	
	Sequence	Arg no	Score	Sequence	Lys no	Score	Sequence	Cys no
<i>L. angustifolius</i>	TTSSLGSS M RRYSLSISSH	13	0.923	PNQEP H KGLLVLL	128	2.12	RADAYAIITICVSFSL-	440
	TSSLGSS M RRYSLSISSHE	14	0.918	SGNGN F KIKSKLD	284	1.72	GILY	
	AGDIGDRAL R SRRFSENS	51	0.914	NFKIK S KLDSCSC	288	1.13		
	RWFLARLN G RGLGK-AGLFK	353	0.970	VSGALV K AEFKHG	318	1.42		
				LV K AE F KHGGSAA	322	2.09		
<i>P. vulgaris</i>				AALAS L K EAVNTR	386	2.07		
	RMSSTGSS I RRRSTSISSH	12	0.951	VPFAEE K ILHPNS	77	1.72	SVKKAVHTRDCDVTVT-	369
							GTQFG	
	MSSTGSS I RRRSTSISSHA	13	0.911	TRRGLP K LLDYAS	106	1.08	VTGTQFGLLGCLSTVS-	383
							TFAAE	
<i>C. sativus</i>	IGDRALPSTRFSQSGSIRF	50	0.939	ASRDG S KINCKVD	259	1.33	VSFVLGILIYCVPVWTK-	426
	RWFLARLN G RGLGK-AGFLK	328	0.972	GSKIN C KVDSNSH	263	1.30	GVXX	
				AALASV K KAVHTR	361	2.18		
				ALASV K KAVHTRD	362	1.23		
	DYGSSNAE A RQGIPLN-RAR	11	0.974	GSLLD K DFSRGK	331	1.23	KLSKLLEYVT-CLVHLAVFGIL	157
<i>V. vinifera</i>	EARQGIPLN- R ARSLGSSFR	18	0.967	AALATV K KAVKTE	394	2.73	LRRKKISYNW C CK-VDSYKRHI	298
	RQGIPLNR A RLGSS-FRRN	20	0.965	TVKKAV K TERVET	398	1.94	RRKKISYNW C CK-VDSYKRHI	299
	RARSLGSS F RRNSFNSSGS	27	0.940					
	ARSLGSS F RRNSFNSSGSG	28	0.965					
	GSLLD K D F S R GKGAEL-WVG	335	0.915					
<i>V. vinifera</i>	RWFLARLN G RGVGQWK-WVP	364	0.935					
	HKTNDLEET R SGSF-GRSSS	12	0.910	XXXMD H KTNDLEE	4	11.13	TLKKAVDSKN C DTIVT-GIEFG	407
	EETRSGSF G RSS-SAGSSLR	18	0.994	MVYSAD K EDDNKK	134	1.02		
	RSSSAGSS L RRR-SCGLPGP	27	0.990	RLNSGG K RGILRS	290	2.25		
	SSSAGSS L RRR-SCGLPGPV	28	0.965	LCGILE K REFQSG	331	1.26		
<i>V. vinifera</i>	SSAGSS L RRRSCGLPG-PVP	29	0.962	AGLST L KKAVDSK	399	1.87		
	LKRLNSGG K R-GILRSNGNW	291	0.989	GLST L KKAVDSKN	400	1.56		
	RWWLARF N G R GLGK-AGLFK	366	0.957					

Residues predicted to be modified have been highlighted in bold

PRmePRed sequences with confidence score higher than 0.9 have been documented (except for *M. truncatula* due to unavailability), *PAIL* sequences with confidence score higher than 1.0 have been documented, *S-nitrosylation* iSNO-PseAAC

Table 5 Prediction of lipid anchors bound to FEXs

Organism	Position	Peptide	Score	Cutoff	Type
<i>O. sativa</i>	469	IQLGFLG C LSTVSTF	1.557	1.079	S-palmitoylation
<i>A. thaliana</i>	4	XXXXXMDT G QSSIEPY	5.892	5.686	N-myristoylation
	378	IANVVAAC V MAALAT	2.481	2.293	S-palmitoylation
	409	IQFGLLG C LSTVSTF	1.925	1.079	S-palmitoylation
<i>B. oleracea</i>	396	VQFGLLG C LSTVSTF	1.599	1.079	S-palmitoylation
<i>B. rapa</i>	390	VQFGLLG C LSTVSTF	1.599	1.079	S-palmitoylation
<i>M. truncatula</i>	17	AVSRLSV C QNVGVGD	4.335	2.293	S-palmitoylation
	395	LQFGLLG C LSTVSTF	1.136	1.079	S-palmitoylation
<i>L. angustifolius</i>	9	ERTTSSL G SSMRRYS	7.1	5.686	N-myristoylation
	408	IQFGLMG C LSTVSTF	1.552	1.079	S-palmitoylation
<i>P. vulgaris</i>	8	MERMSST G SSIRRS	6.265	5.686	N-myristoylation
	383	TQFGLLG C LSTVSTF	1.25	1.079	S-palmitoylation
	426	VLGILIY C VPVWTKG	1.112	1.072	S-palmitoylation
<i>C. sativus</i>	299	KISYNW C CKVDSYKR	1.403	1.072	S-palmitoylation
	416	LQLGLLG C LSTVSTF	2.059	1.079	S-palmitoylation
<i>V. vinifera</i>	31	SSLRRRS C GLPGVP	2.714	2.293	S-palmitoylation
	326	LGLLYSL C GILEKRE	1.772	1.072	S-palmitoylation

Predicted amino acid residues to be modified have been highlighted in bold

IC₅₀ by > 1000-fold. Both these Phe residues are conserved in the plant FEXs (Fig. 1; Table 2).

Post-translational modifications and intrinsic disorder

Post-translational modifications increase the proteome and signalosome complexities. These generally occur on amino acid side chains within a specific signature peptide sequence and regulate protein signaling (Khoury et al. 2011). The NetPhos 3.1 Server predicted large number of Ser/Thr phosphorylations throughout the plant FEX sequences (Table 3). The usage of FEX disorder information by DISPHOS 1.3 on the contrary localized Ser phosphorylations towards the N-terminal domain in plant FEXs (Fig. 3). Taking OsFEX as the representative member, we analyzed its intrinsic disorder pattern and found the N-terminal region to be disordered (Fig. 4a). This probably indicates towards the possibility that kinase-dependent phosphorylations in FEXs occur predominantly in the intrinsically disordered N-terminus. The concerned N-terminal residues possibly do not participate in the core structural assembly of the transporter. The residues beyond position 171 formed proper secondary structures (TMs) (Fig. 4a). Hence, the statement by Smith et al. (2015) that the mutations in the N-terminal domain of yeast FEX are much less detrimental than similar ones in the C-terminal domain might also appear justified for plant FEXs.

However, N-terminal phosphorylation was not observed in case of *B. oleracea* and *B. rapa* FEXs, exhibiting close evolutionary relatedness (Fig. 2). The FEXs in these two

species exhibited a single Ser phosphorylation located towards the middle of the sequence (Fig. 3). DisEMBL predicted significant disorder probability for the loops and coils lying between successive helices in OsFEX (Fig. 4b). These unstructured portions lack electron density in crystal structures and form hot loops. Thus, it is possible that the single Ser phosphorylation site in *B. oleracea* and *B. rapa* FEXs detected by DISPHOS 1.3 lie within these unstructured coils or loops accessible to kinases.

Non-histone Arg methylations are catalyzed by protein Arg methyltransferases which utilize S-adenosylmethionine as the methyl group donor. Such methylations can be either a dynamic process or serve regulatory purposes or act as static modifications that extend the functional repertoire of amino acids (Alban et al. 2014). N^ε-lysine acetylation in internal Lys residues orchestrates several cellular processes such as protein stability, function and protein–protein interactions (Glozak et al. 2005). S-nitrosylation triggered by the signaling molecule, nitric oxide (NO), is targeted on specific Cys thiols. This modification triggered by Cu/Zn superoxide dismutase and thioredoxin can modulate protein stability, activity and trafficking (Hess et al. 2005). All the above modifications were detected with high confidence scores in plant FEXs (Table 4), thereby indicating at the cellular complexity dedicated towards FEX activity and stability. FEX is a transmembrane channel protein. Hence, fatty acid anchors comprising of S-palmitoylation (at Cys residues) and N-myristoylation (at Gly residues) were predicted to occur in the proteins (Table 5). These lipid moieties probably anchor and orient FEXs to the plasma membrane in plants.

Fig. 4 Intrinsic disorder prediction and evolutionary functional significance of OsFEX. **a** The N-terminus ranging between 1 and 170 amino acids in OsFEX is intrinsically disordered. Possibly, this region is not crucial for FEX transport function and hence mutations in this portion are less detrimental for the transport activity. The residues beyond about 170 synchronize to form proper secondary structure, i.e., transmembrane helices. It is to be noted that all the conserved motifs quintessential for the FEX activity are located in this stretch. An intermittent small stretch of disorder was visible between residues 300–340, though the confidence score was low. This region might form an unstructured loop between two adjacent TMs. **b** DisEMBL predicted the periodic existence of seven disordered regions (seven peaks) [dotted line (purple colored) loops or coil] between 200

and 450 residues. These are probably due to the loops or coils formed between the successive nine TMs formed by FEX proteins. These loops and coils usually lack electron densities in crystal structures and so the remark 465 peaks [dotted line (yellow colored)] also corroborated with those of loops or coil. Similar observation was noted for the peaks for unstructured hot loops [dotted line (red colored)]. However, the probability of occurrence of hot loops was much lower than the threshold value of 0.4; **C**: The GO Ancestor Chart of OsFEX validates that it is a transmembrane transporter capable of transporting the inorganic anion, fluoride; **D**: Pfam search of OsFEX determines that the protein contains two CRCB domains ranging from 176–295 to 368–489 residues. These domains co-operate to form the functional fluoride transporting pore II in rice FEX

Analyzing the OsFEX model

Gene ontology ancestor chart (Fig. 4c) predicted the biological process of OsFEX (GO:1903424) as an inorganic anion (fluoride) transporter localized in the plasma membrane. The unstructured coils and loops located at either ends of the helices are composed of charged amino acids (Fig. 5c, d) which are solvent accessible. The rice FEX model revealed some striking structural similarities with yeast FEX1. The highlighted conserved residues in $PxGTxxxN$, $YxxxS/T$,

ooFooX and ooXooFxxE have been found to line the yeast FEX1 Pore II (Berbasova et al. 2017). It was predicted that these residues also decorate the inner face of Pore II in rice FEX (Fig. 6). This points out that plant FEXs have maintained both the sequential (Fig. 1) and structural conservation of all the necessary amino acids that are crucial for transporter functioning. Hence, FEXs should be probable fluoride transporters in plants due to overall conservation of the Pore II topology. The two Pfam CRCB domains detected in OsFEX (Fig. 4d) co-operate to form the functional Pore

Fig. 5 Schematic differences in fluoride transport between Fluc and OsFEX; and the protein contact potential of OsFEX. **a** Prokaryotic Fluc is a double-barrelled exporter with a centrally placed Na^+ (blue) and two functional fluoride export pores. Each pore exports two fluoride ions (red); **b** similar to other eukaryotic FEXs, OsFEX is a single-barrelled transporter with a centrally placed Na^+ (yellow). This channel is a fusion of two subunits/domains (backbones marked in gray and light blue). Pore I (evolutionarily vestigial) cannot transport fluoride and is located behind the functional Pore II. The conserved Phe residues (F257 from domain 1 and F454 from domain 2)

in Pore II weakly interact with the fluoride ions (red) via aromatic C–H and aid in their transport; **c** protein contact potential of predicted N-terminal model of OsFEX; **d** protein contact potential of predicted C-terminal model of OsFEX. Charge smoothing has shown that the charged amino acids are basically located towards the two tips of both the predicted models where the unstructured loops or coils between successive helices exist. These charged amino acids located within unstructured local coils are highly solvent accessible as predicted by ACCpro at 25% exposed threshold

II. In the rice FEX model, domain 1 contributes Phe257, whereas the other conserved residues within Pore II are contributed by domain 2. The two Phe residues weakly interact with the fluoride ions via aromatic C–H. This probably dictates the fluoride specificity of this channel (Last et al. 2016). The conserved amino acids lining the inner side of Pore II usually have high hydrogen bonding potential and hence can associate with the fluoride ions which are good hydrogen bond acceptors. The chemically favored hydrated state of fluoride ions might be mimicked by the residues

possibly to facilitate fluoride desolvation from the aqueous solution (Hille 2001).

The Ser/Thr residues of **ooFooX** and **ooXooF** motifs do not lie within Pore II, but are evolutionarily conserved. Table 2 shows that mutation of such Ser/Thr residues in particular combinations can significantly lower the IC_{50} values in yeast. In the rice FEX model, the main chain carbonyls of Ser449 and Thr255 along with those of Gly446 and Gly252 stabilize the Na^+ ion (Fig. 7). The side chains of the other Ser/Thr residues make hydrogen bonds with each other. We have previously mentioned that the end ‘o’ in **ooFooX**

Fig. 6 Structural analysis of Pore II in the FEX model of *O. sativa* var. japonica. The sequentially conserved residues in the Pore II (labelled in black) of OsFEX are located at similar structural positions with respect to the residues in Pore II (labelled in red) of yeast FEX1. The helices with green backbone belong to domain 1 and contribute Phe257. The other residues are contributed by domain 2 (helices with light blue backbone). The fluoride ions in Pore II are highlighted as orange spheres

motif has been altered to Gly in all the plant FEX sequences that were analyzed. We observed that in rice FEX, this Gly (Gly259) was located far away from the Na^+ and was not associated with its stabilization, though further experimental

studies are required to endorse this view. The substitution possibly might not have much effect on the structural conservation among FEX proteins. Berbasova et al. (2017) reported that eukaryotic FEXs contain flexible linkers which split TM3a/b and TM8a/b, so that these TMs can crossover each other. The interaction of Ser/Thr residues with the Na^+ and between each other probably aids in the formation and stabilization of two such helix crossovers, so that Phe257 of domain 1 can face the inner lining of Pore II. The entire structure is stabilized by hydrogen bonding, two disulphide bonds and several non-bonded interactions. Though predicted to be insoluble upon heterologous overexpression in *E. coli*, we assessed the crystallizable probability of OsFEX. The protein occupied class 1 (most promising) of “Random forest” crystallizable class (Fig. 8m) after comparison with the standard features of crystallized proteins submitted in TargetDB (Slabinski et al. 2007).

Characterizing OsFEX RNA

The *OsFEX* RNA formed multiple pseudoknots. Though no riboswitch or sRNA regulation could be detected from the current database, further experimental investigations are required to confirm these observations. This is because a riboswitch associated with a pseudoknot and specifically sensitive to fluoride ion has been identified in the *CrcB* RNA of *Pseudomonas syringae* (Baker et al. 2012). The bacterium, *Methylobacterium extorquens* DM4, encodes ten fluoride riboswitches in its genome, indicating that a robust fluoride sensor is essential for mitigating fluoride-induced toxicity (Baker et al. 2012).

Fig. 7 Structural and functional significance of S/T fragments in the FEX model of *O. sativa* var. japonica. Thr255, Thr256, Ser258, Gly259 (ooFoG) and Gly252 belong to domain 1 (helices with red backbone). Ser449, Thr450, Ser452, Thr453 (ooXoo) and Gly446 belong to domain 2 (helices with green backbone). The red dotted lines indicate polar interactions

Fig. 8 Comparison of features of the OsFEX with distributions of crystallization probability calculated from TargetDB and prediction of its crystallizability class. Predicted parameters of the queried OsFEX protein have been compared with multiple factors such as length (a), isoelectric point (b), gravy index (c), disordered region (d), instability index (e), percentage of coil structures (f), surface entropy (g), surface hydrophobicity (h), surface ruggedness (i), coiled coils (j) and insertion score (k) which determine the crystallizable probability. The X-axis represents the measures for the respective parameters. The extreme left Y-axis represents the number of crystallized proteins, whereas the extreme right Y-axis illustrates the probability of crystallization. The blue and gray bars show the observed

Promoter analysis

We next analyzed the presence of common *cis*-acting elements in the promoters of representative *FEX* genes from plants (Table 6). The *cis*-acting elements such as TATA boxes and enhancer elements such as GATA and CAAT boxes were detected in the *FEX* promoters of *O. sativa*, *A. thaliana*, *M. sativa* and *V. vinifera*. A root-specific *cis*-acting element, ROOTMOTIFTAPOX1, represented that the *FEX* genes might be expressed in plant roots along with shoots.

successes and failures at the particular measures of the respective parameters. The black line gives a crystallization trendline for the tested parameters, whereas the red line determines the predicted value of OsFEX for that parameter. OsFEX was predicted to be least crystallizable as it qualified class 5 according to the expert pool crystallizability class analysis (l). The “random forest” crystallizability class analysis presents a more developed and authentic prediction, as unlike the expert pool analysis, it considers crucial parameters such as surface entropy, surface ruggedness, instability index, percentage of coil structures, coiled coils and insertion score. OsFEX qualified class 1 (highest probability of being crystallized) according to “random forest” analysis (m)

Interestingly, repeats of a pollen-specific element, POLLEN1LELAT52, were also identified (Table 6). Fluoride translocates to the shoots and often accumulates in reproductive parts such as fruits and grains (Mondal 2017). The pollen-specific element probably ensures *FEX* expression in the pollens to exclude toxic fluoride ions and mediate proper reproductive development.

Phytohormones such as abscisic acid (ABA), gibberellic acid (GA) and salicylic acid (SA) are important regulators of plant growth physiology. ABA is regarded as the universal

Fig. 9 Structural predictions of *OsFEX* RNA. **a** The overall *OsFEX* RNA secondary structure was predicted with free energy of $-492.3 \text{ kcal mol}^{-1}$; proposed pseudoknots (colored lines) predicted within base numbers 1–400 (**b**), 401–800 (**c**), 801–1200 (**d**) and 1201

to end (**e**) after an average of 11,000,000 ms was utilized to estimate RNA structural folding. The structures with the least free energy as assigned beneath the respective figures (**b–e**) have been selected and represented

stress phytohormone (Roychoudhury and Banerjee 2017; Banerjee and Roychoudhury 2018a, b, c, 2019). Thus, the presence of ABA-responsive elements (ABREs) in multiple repeats is common in ABA-inducible genes. The ABREs have an ACGT core consensus sequence (G-box) to which the basic leucine zipper (bZIP) transcription factors (TFs) bind (Banerjee and Roychoudhury 2017). Another type of ABRE, EBOXBNNAPA, was also identified in multiple copies. EBOXBNNAPA has been regarded as the most well-distributed ABRE in the promoters of calcium-dependent protein kinases in rice (Roychoudhury and Banerjee 2017). We identified multiple ABREs consisting of G-box and EBOXBNNAPA in the upstream region of *FEX* genes across different plant species (Table 6). ABA and GA are known to

work antagonistically during stress. However, GA-responsive elements (GAREs) could be observed in *FEX* promoters (Table 6). Such cross-regulation of *FEX* via two antagonistically interacting phytohormones might be interesting. Possibly under control conditions, the GA-responsive silencing factors bind to the GAREs and down regulate unnecessary *FEX* expression to channelize the metabolic equivalents for plant growth. During fluoride stress, GA levels are reduced (Gadi et al. 2016) probably due to ABA accumulation. The *FEX* repression is removed in the absence of GA-responsive inhibitors and because of the binding of ABA-dependent bZIP TFs to the ABREs, the gene is up regulated. However, detailed investigations at the molecular level are required for establishing such hypotheses. The reverse regulation, i.e.,

Table 6 Analyses of upstream region (~ 1 Kbp) of *FEX* gene from representative plant species

Type	<i>Cis</i> elements	Sequence	<i>O. sativa</i>	<i>A. thaliana</i>	<i>M. truncatula</i>	<i>V. vinifera</i>
TATA box	TATABOX2	TATAAAT	–	1	2	–
	TATABOX3	TATTAAT	–	1	2	–
	TATABOX4	TATATAA	–	1	1	–
	TATABOX5	TTATTT	1	6	3	–
	TATABOXOSPAL	TATTTAA	–	–	1	1
GATA box	GATABOX	GATA	7	9	6	8
CAAT box	CAATBOX1	CAAT	11	11	7	16
Root-specific	ROOTMOTIFTAPOX1	ATATT	3	9	9	4
Pollen-specific	POLLEN1LELAT52	AGAAA	9	5	3	8
ABRE	ACGTATERD1 (GBOX)	ACGT	2	2	3	1
	EBOXBNNAPA	CANNTG	6	3	5	6
WRKY binding element	WRKY71OS	TGAC	13	5	11	10
	WBOX	TTGAC	8	3	4	5
GARE	PYRIMIDINEBOX	CCTTTT/TTTTTTTCC	1	2	1	2
SARE	GT1CONSENSUS	GRWAAW	8	9	6	5
PERE	OSE2ROOTNODULE	CTCTT	5	1	2	5
LRE	CIACADIANLELHC	CAANNNNATC	1	1	1	1

ABRE abscisic acid-responsive element, *W-box* WRKY protein-binding element, *GARE* gibberellic acid-responsive element, *SARE* salicylic acid responsive element, *PERE* pathogen elicitor-responsive element, *LRE* light-responsive element

gene induction by GA and silencing in the presence of ABA has been reported in case of α -amylase gene of *Hordeum vulgare* (Gubler and Jacobsen 1992). The upstream region of this gene contained both GAREs and ABREs (Gubler and Jacobsen 1992).

WRKY proteins are crucial TFs which regulate several abiotic and biotic stresses. They bind to conserved motifs called W-boxes (Banerjee and Roychoudhury 2015). A large number of WRKY71-binding sites (WRKY71OS) were detected in all the *FEX* promoters in the selected plant species (Table 6). *WRKY71* is an ABA-inducible gene in rice and the encoded TF has been found to block GA-induced α -amylase gene expression in *H. vulgare* (Zou et al. 2008).

The *cis*-acting elements such as the ABREs and WRKY71OSs detected in the plant *FEX* promoters can be present in those of any ABA-responsive gene. *FEX*, being fluoride-specific transporter, should not be produced abundantly under any sub-optimal condition which induces endogenous ABA levels. This would not be metabolically beneficial for the system. Thus, the presence of a fluoride-specific riboswitch in *FEX* RNAs is quite evident because it acts as a post-transcriptional regulatory checkpoint. It determines the optimum tertiary RNA structure in presence of fluoride ions to ultimately promote *FEX* translation. So, in spite of the fact that bioinformatics could not detect a riboswitch in *FEX* RNA via sequence similarity searches within a limited database, we propose the existence of a robust fluoride sensor within the plant cell.

FEX promoter was also found to contain SA-responsive elements (SAREs) (Table 6). SAREs are recognized by bZIP TFs and though SA is primarily involved in systemic acquired resistance, it also plays a role in developing abiotic stress responses (Roychoudhury et al. 2016). The presence of an element associated with pathogen infection indicates that *FEX* expression might represent crosstalk between abiotic and biotic stresses.

Interestingly, a light-responsive element (LRE) associated with the maintenance of circadian rhythm has been detected in single copy in all the *FEX* promoters. Recent studies have correlated the existence and importance of melatonin as a circadian rhythm associated-phytohormone, actively participating in abiotic stress responses (Nawaz et al. 2016). The LRE might actually act as a melatonin-responsive element and promote the binding of melatonin-responsive TFs to regulate *FEX* expression. A single LRE was identified in the *FEX* promoters of all selected plant species (Table 6).

Conclusion

Our findings showed that similar to yeast, plant *FEX*s are single-barrelled channels with nine TMs which co-operate to form one functional fluoride transport pore and one vestigial pore with no activity. The C-terminal domain of the protein is more important than the unstructured N-terminal domain. All the crucial residues quintessential for the fluoride

transport function of FEX are sequentially and structurally conserved in plant FEXs. This indicates that FEXs might also work as functional fluoride transporter in the analyzed plant species. The proteins might be post-translationally modified and anchored to the plasma membrane via lipid anchors. The crystallizability of OsFEX was also predicted from our analyses. The *OsFEX* RNA has been observed to form multiple pseudoknots, which increase the possibility of finding a novel riboswitch essential for robust fluoride sensing. The *FEX* promoter analyses revealed the uniform presence of *cis*-acting elements in all the plant species. This prediction supports the possible phytohormone-mediated regulation of *FEX* expression across different plant species.

Acknowledgements Financial assistance from Council of Scientific and Industrial Research (CSIR), Government of India, through the research grant [38(1387)/14/EMR-II], Science and Engineering Research Board, Government of India through the grant [EMR/2016/004799] and Department of Higher Education, Science and Technology and Biotechnology, Government of West Bengal, through the grant [264(Sanc.)ST/P/S&T/1G-80/2017] to Dr. Aryadeep Roychoudhury is gratefully acknowledged. The authors are thankful to the University Grants Commission (UGC), Government of India, for providing Junior Research Fellowship to Aditya Banerjee.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest in publishing the manuscript.

References

- Alban C, Tardif M, Mininno M, Brugire S, Gilgen A, Ma S et al (2014) Uncovering the protein lysine and arginine methylation network in *Arabidopsis* chloroplasts. *PLoS One* 9:e95512
- Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* 335:233–235
- Banerjee A, Roychoudhury A (2015) WRKY proteins: Signaling and regulation of expression during abiotic stress responses. *Sci World J* 2015:807560
- Banerjee A, Roychoudhury A (2017) Absciscic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress. *Protoplasma* 254:3–16
- Banerjee A, Roychoudhury A (2018a) The gymnastics of epigenomics in rice. *Plant Cell Rep* 37:25–49
- Banerjee A, Roychoudhury A (2018b) Strigolactones: multi-level regulation of biosynthesis and diverse responses in plant abiotic stresses. *Acta Physiol Plant* 40:86
- Banerjee A, Roychoudhury A (2018c) Interactions of brassinosteroids with major phytohormones: antagonistic effects. *J Plant Growth Regul* 37:1025–1032
- Banerjee A, Roychoudhury A (2019) Fluorine: a biohazardous agent for plants and phytoremediation strategies for its removal from the environment. *Biol Plant* 63:104–112
- Bergert P, Dandekar T (2004) Riboswitch finder—a tool for identification of riboswitch RNAs. *Nucleic Acids Res* 32:W154–W159
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E (2015) The Arabidopsis Information Resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis* 53:474–485
- Berbasova T, Nallur S, Sells T, Smith KD, Gordon PB et al (2017) Fluoride export (FEX) proteins from fungi, plants and animals are ‘single barrelled’ channels containing one functional and one vestigial ion pore. *PLoS One* 12:e0177096
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:W252–W258
- Blom N, Gammeltoft S, Brunak S (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294:1351–1362
- Cheng J, Saigo H, Baldi P (2006) Large-scale prediction of disulphide bridges using Kernel methods, two-dimensional recursive neural networks, and weighted graph matching. *Proteins* 62:617–629
- Choubisa SL (2013) Fluoride toxicosis in immature herbivorous domestic animals living in low fluoride water endemic areas of Rajasthan, India: an observational survey. *Fluoride* 46:19–24
- DeLano WL (2002) The PyMOL molecular graphics system. Delano Scientific, San Carlos
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36:W465–W469
- Dong Q, Schlueter SD, Brendel V (2004) PlantGDB, plant genome database and analysis tools. *Nucleic Acids Res* 32:D354–D359
- Farooqi A, Masuda H, Kusakabe M, Naseem M, Firdous N (2007) A geochemical model for the abnormal fluoride concentrations in waters in parts of northern Tanzania. *Geochem J* 41:213–234
- Gadi BR, Bhati K, Goswami B, Rankawat R, Kumar S et al (2016) Sources and phytotoxicity of fluorides in the environment. *Environ Impact Biodivers* 2016, 251–266
- Glozak MA, Sengupta N, Zhang X, Seto E (2005) Acetylation and deacetylation of non-histone proteins. *Gene* 363:15–23
- Gubler F, Jacobsen JV (1992) Gibberellin-responsive elements in the promoter of a barley high-pI α -amylase gene. *Plant Cell* 4:1435–1441
- Havill JT, Bhatiya C, Johnson SM, Sheets JD, Thompson JS (2014) A new approach for detecting riboswitches in DNA sequences. *Bioinformatics* 30:3012–3019
- Hess DT, Matsumoto A, Kim SO, Marshall HE, Stamler JS (2005) Protein S-nitrosylation: purview and parameters. *Nat Rev Mol Cell Biol* 6:150–166
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297–300
- Hille B (2001) Ion channels of excitable membranes, 3 Sinauer Associates, Sunderland, 814
- Hong BD, Joo RN, Lee KS, Lee DS et al. (2016) Fluoride in soil and plant. *Korean J Agric Sci* 43:522–536
- Iakoucheva LM, Radivojac P, Brown CJ, O’Connor TR et al (2004) Intrinsic disorder and protein phosphorylation. *Nucleic Acids Res* 32:1037–1049
- Jones DT, Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31:857–863
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4
- Khoury GA, Baliban RC, Floudas CA (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* 1:90
- Kumar P, Joy J, Pandey A, Gupta D (2017) PRmePRed: A protein arginine methylation prediction tool. *PLoS One* 12:e0183318

- Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM (2018) PDBsum: Structural summaries of PDB entries. *Prot Sci* 27:129–134
- Last NB, Kolmakova-Partensky L, Shane T, Miller C (2016) Mechanistic signs of double-barrelled structure in a fluoride ion channel. *eLife* 5:e18767
- Li S, Smith KD, Davis JH, Gordon PB, Breaker RR, Strobel SA (2013) Eukaryotic resistance to fluoride toxicity mediated by a widespread family of fluoride export proteins. *Proc Natl Acad Sci USA* 110:19018–19023
- Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB (2003) Protein disorder prediction: implications for structural proteomics. *Structure* 11:1453–1459
- Magnan CN, Baldi P (2014) SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* 30:2592–2597
- Magnan CN, Randall A, Baldi P (2009) SOLpro: accurate sequence-based prediction of protein solubility. *Bioinformatics* 25:17
- McWilliam H, Li WZ, Uludag M, Squizzato S, Park YM, Buso N et al (2013) Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res* 41:W597–W600
- Mondal NK (2017) Effect of fluoride on photosynthesis, growth and accumulation of four widely cultivated rice (*Oryza sativa* L.) varieties in India. *Ecotoxicol Environ Saf* 144:36–44
- Mondal D, Gupta S (2015) Influence of fluoride contaminated irrigation water on biochemical constituents of different crops and vegetables with an implication to human risk through diet. *J Mater Environ Sci* 6:3134–3142
- Morrison EA, DeKoster GT, Dutta S, Vafabakhsh R, Clarkson MW, Bahl A et al (2012) Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* 481:45–50
- Nanyaro JT, Aswathanarayana U, Mungure JS (1984) A geochemical model for the abnormal fluoride concentrations in waters in parts of northern Tanzania. *J Afr Earth Sci* 2:129–140
- Nawaz MA, Huang Y, Bie Z, Ahmed W, Reiter RJ, Niu M, Hameed S (2016) Melatonin: current status and future perspectives in plant science. *Front Plant Sci* 6:1230
- Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinform* 11:129
- Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324
- Roychoudhury A, Banerjee A (2017) Absciscic acid signaling and involvement of mitogen activated protein kinases and calcium-dependent protein kinases during plant abiotic stress. In: Pandey G (Ed.) *Mechanism of Plant Hormone Signaling under Stress*. Wiley, Hoboken, vol. 1, 197–241
- Roychoudhury A, Ghosh S, Paul S, Mazumdar S, Das G, Das S (2016) Pre-treatment of seeds with salicylic acid attenuates cadmium chloride-induced oxidative damages in the seedlings of mungbean (*Vigna radiata* L. Wilczek). *Acta Physiol Plant* 38:11
- Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* 23:3403–3405
- Smith KD, Gordon PB, Rivetta A, Allen KE, Berbasova T et al. (2015) Yeast Fex1p is a constitutively expressed fluoride channel with functional asymmetry of its two homologous domains. *J Biol Chem* 290:19874–19887
- Stockbridge RB, Lim HH, Otten R, Williams C, Shane T, Weinberg Z et al. (2012) Fluoride resistance and transport by riboswitch-controlled CLC antiporters. *Proc Natl Acad Sci USA* 109:15289–15294
- Stockbridge RB, Robertson JL, Kolmakova-Partensky L, Miller C (2013) A family of fluoride-specific ion channels with dual topology architecture. *eLife* 2:e01084
- Stockbridge RB, Kolmakova-Partensky L, Shane T, Koide A, Koide S et al (2015) Crystal structures of a double-barrelled fluoride ion channel. *Nature* 525:548
- Susheela AK (1999) Fluorosis management programme in India. *Curr Sci* 77:10
- Xayaphoummine A, Bucher T, Isambert H (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acid Res.* 33:605–610
- Xu Y, Ding J, Wu LY, Chou KC (2013) iSNO-PseAAC: Predict cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition. *PLoS One* 8:e55844
- Yadu B, Chandrakar V, Meena RK, Poddar A, Keshavkant S (2018) Spermidine and melatonin attenuate fluoride toxicity by regulating gene expression of antioxidants in *Cajanus cajan* L. *J Plant Growth Regul* 37:1113–1126
- Zhang Y, Skolnick J (2005) TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Res* 33:2302–2309
- Zou X, Neuman D, Shen QJ (2008) Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellin signaling in aleurone cells. *Plant Physiol* 148:176–186