



# Restriction of access to the central cavity is a major contributor to substrate selectivity in plant ABCG transporters

Konrad Pakuła<sup>1,2</sup> · Carlos Sequeiros-Borja<sup>3,4</sup> · Wanda Biała-Leonhard<sup>1</sup> · Aleksandra Paweła<sup>1</sup> · Joanna Banasiak<sup>1</sup> · Aurélien Bailly<sup>5</sup> · Marcin Radom<sup>6,7</sup> · Markus Geisler<sup>8</sup> · Jan Brezovsky<sup>3,4</sup> · Michał Jasiński<sup>1,9</sup>

Received: 25 October 2022 / Revised: 22 February 2023 / Accepted: 6 March 2023 / Published online: 23 March 2023  
© The Author(s) 2023

## Abstract

ABCG46 of the legume *Medicago truncatula* is an ABC-type transporter responsible for highly selective translocation of the phenylpropanoids, 4-coumarate, and liquiritigenin, over the plasma membrane. To investigate molecular determinants of the observed substrate selectivity, we applied a combination of phylogenetic and biochemical analyses, AlphaFold2 structure prediction, molecular dynamics simulations, and mutagenesis. We discovered an unusually narrow transient access path to the central cavity of MtABCG46 that constitutes an initial filter responsible for the selective translocation of phenylpropanoids through a lipid bilayer. Furthermore, we identified remote residue F562 as pivotal for maintaining the stability of this filter. The determination of individual amino acids that impact the selective transport of specialized metabolites may provide new opportunities associated with ABCGs being of interest, in many biological scenarios.

**Keywords** ABC transporters · AlphaFold2 · Selectivity · Phenylpropanoids · Access path

## Introduction

ATP-binding cassette (ABC) transporters are prominent proteins that translocate molecules through biological membranes using ATP as a source of energy [1]. Members of the ABC family are common in all domains of life, but plant genomes are exceptionally rich in genes encoding

them [2, 3]. Taking into account their structure and phylogenetic relationships, most ABC proteins have been classified into eight subfamilies, designated ABCA-ABCH [4]. The particularly numerous plant ABC transporters translocate diverse molecules, such as lipids, phytohormones, carboxylates, heavy metals, chlorophyll catabolites, and xenobiotic conjugates, across various biological membranes [5]. In this manner, they participate in diverse processes, including organ growth, nutrition, development,

Konrad Pakuła, Carlos Sequeiros-Borja and Wanda Biała-Leonhard contributed equally to the work.

✉ Jan Brezovsky  
janbre@amu.edu.pl; jbrezovsky@iimcb.gov.pl

✉ Michał Jasiński  
jasinski@ibch.poznan.pl

<sup>1</sup> Department of Plant Molecular Physiology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznan, Poland

<sup>2</sup> NanoBioMedical Centre, Adam Mickiewicz University, Wszechnicy Piastowskiej 3, 61-614 Poznan, Poland

<sup>3</sup> Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznan, Poland

<sup>4</sup> International Institute of Molecular and Cell Biology in Warsaw, Ks. Trojdena 4, 02-109 Warsaw, Poland

<sup>5</sup> Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

<sup>6</sup> Department of Structural Bioinformatics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z.Noskowskiego12/14, 61-704 Poznan, Poland

<sup>7</sup> Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland

<sup>8</sup> Department of Biology, University of Fribourg, Chem. du Musée 10, 1700 Fribourg, Switzerland

<sup>9</sup> Department of Biochemistry and Biotechnology, Poznan University of Life Sciences, Dojazd 11, 60-632 Poznan, Poland

responses to abiotic stresses, as well as both symbiotic and antagonistic relationships [2, 6, 7].

Especially abundant in plants are genes encoding so-called full-size ABCG transporters. A full-size ABCG transporter is a single polypeptide forming two transmembrane domains (TMDs), which constitute a membrane-spanning region, and two cytosolic domains called nucleotide-binding domains (NBDs). Depending on the transporter subfamilies, single TMD has five to ten transmembrane  $\alpha$ -helices. Full-size members of the G family are characteristic for plants and fungi and distinct in their reverse organization of domains in the subunits namely NBD1-TMD1-NBD2-TMD2 [4, 8]. Moreover, in comparison to thoroughly studied ABCB proteins, where  $\alpha$ -helices-forming TMDs represent a so-called domain swap arrangement, ABCG transporters revealed a different TMD organization, in which not individual helices but the entire TMD rotates as a solid body during the transport of molecules [9]. Full-size ABCGs were initially identified in *Saccharomyces cerevisiae* and the clinically relevant fungus, *Candida albicans*. They were described as pleiotropic drug resistance (PDR) proteins because they can act as efflux pumps, removing diverse molecules from these unicellular organisms, including exogenously applied drugs used in medical treatments, thus conferring resistance against large sets of chemicals [10]. Consequently, PDR proteins have attracted interest of biotechnologists and medical scientists. However, despite strenuous efforts, we still have limited understanding of molecular bases of their action.

Characterized plant full-size ABCGs include proteins that can transport several molecules that are not necessarily related but are usually endogenous metabolites. For instance, ABCG37 of *Arabidopsis thaliana* is involved in translocation of the auxin precursor, indolyl-3-butyric acid (IBA) [11], and the phenolic compound scopoletin [12, 13]. *Arabidopsis* ABCG36/PEN3/PDR8 transports, among others, IBA, the Brassicales-specific phytoalexin camalexin, heavy metals, and possibly monolignols [14–21]. However, despite initially considered as functional homologs of yeast multidrug pumps, at least certain plant full-size ABCG proteins appear to be selective toward translocated molecules. The specialization is proposed to be a consequence of a sophisticated chemodiversity exemplified by specialized metabolism that requires tightly controlled distribution of metabolites via dedicated transporters [3, 5].

Phenylpropanoids are a large class of specialized plant metabolites with many important roles in plant biology, medical applications, and industrial uses [22]. ABCG46 (formerly known as ABCG10) of the legume *Medicago truncatula*, which is required for efficient de novo production of the phenylpropanoid-derived phytoalexin medicarpin, selectively translocates 4-coumarate and

liquiritigenin. Notably, structurally similar phenylpropanoids like naringenin, isoliquiritigenin, and 7,4'-dihydroxyflavone are not transported by MtABCG46 [23, 24].

Despite the recent progress in structural research regarding ABCG proteins [25–29], progressing from static atomic structures to an understanding of molecular mechanisms behind the substrate recognition and transport has been challenging [30]. This is partly because of the difficulty in obtaining experimental data at atomic detail, but also because of the lack of efficient sampling of the intricate process of a complete transport cycle. Fortunately, AlphaFold2 [31] has proven utility for addressing the first of these problems by providing accurate 3D structures of proteins from the amino acid sequence, even for large transmembrane hydrophobic proteins that are difficult to crystalize [32]. Structurally, although the complete transport process is still elusive, it is accepted that ABCG exporters start this process with an inward-facing (IF) conformation that allows substrate migration to the central cavity. The protein then undergoes large structural rearrangements to an outward-facing (OF) conformation, enabling release of the substrate to the extracellular environment. Binding, hydrolysis, and release of ATP as well as substrate recognition followed by its migration contribute to the intricacy of the overall transport process [28, 29].

So far, due to its clinical importance—as connected to the multidrug resistance—a non-selective transporter HsABCG2 is one of the most studied proteins from the G subfamily. Obtained HsABCG2 cryo-EM data provide some structure–function insights. For instance, it enabled for the identification of the short loop after TMD helix 5 called the valve/plug. The latter is proposed as an important structural element regulating conformation-dependent substrate release from the central (binding) cavity and preventing substrate reflux [29, 33]. Furthermore, previous simulation work has highlighted the valve region as crucial for substrate binding, structurally dividing the protein interior into two cavities, which are accessible at different stages of the transport process and permeable for at least some substrates with energy costs not greater than hydrolysis of an ATP molecule [33–35]. Also, highly conserved residues located in TMD helix 2 were found to be vital for trapping the substrate within the central cavity [36]. However, the molecular details behind the substrate recognition still remain unclear, as well as whether those are analogous within the structurally and functionally diverse group of ABCG transporters.

Using a combination of phylogenetic and biochemical analyses, AlphaFold2 structure prediction, molecular dynamics simulations, and mutagenesis, we have identified a transient access path in MtABCG46 that is directly involved in the recognition and passage of 4-coumarate and liquiritigenin through the plasma membrane. Moreover, we

have identified F562 as a critical residue for the architecture of this access path responsible for selective transport.

## Materials and methods

### Plant material

*Nicotiana tabacum* Bright Yellow 2 (BY2) suspension cell cultures [37] were grown in Murashige and Skoog medium supplemented with 2.72 mM  $\text{KH}_2\text{PO}_4$ , 0.56 mM myoinositol, 3  $\mu\text{M}$  thiamine, 0.9  $\mu\text{M}$  2,4-dichlorophenoxyacetic acid, and 87.64 mM sucrose, in the dark at 26 °C on an orbital shaker (130 rpm), and diluted 1:5 every week.

### Genetic constructs

All genetic constructs were based on the pMDC43 vector, carrying a GFP tag sequence [38]. p35S::GFP-MtABCG46 construct as well as particular mutants of p35S::GFP-MtABCG46 were generated by GenScript. All plasmid constructs were confirmed by DNA sequencing.

### Plant transformation

Stably transformed BY2 cells were generated by co-cultivation of 5-day-old BY2 suspension cells with *Agrobacterium tumefaciens* strain AGL1 [39] carrying the pMDC43 vector containing a particular variant of the MtABCG46 sequence or the empty pMDC43 vector, as previously described [24].

### Confocal microscopy

Five-day-old suspension cell cultures overexpressing GFP-fused MtABCG46 variants were observed by laser scanning confocal microscopy (with a Leica TCS SP5 AX v.2.7 instrument). Plasma membranes of sampled cells were stained with FM4-64 (ThermoFisher Scientific) according to the manufacturers' protocol, no fixation was applied. Obtained pictures were analyzed using Leica LAS AF software. Fluorescent signals from GFP and FM4-64 were pseudo-colored in green and magenta, respectively. The excitation wavelength for GFP and FM4-64 was 488 nm. Fluorescence signals were collected at 485–547 nm (GFP) and 570–650 nm (FM4-64).

### Preparation of plasma membrane vesicles

Microsomal fractions were isolated from 12 g portions of BY2 suspension cell cultures as previously described [40]. Plasma membrane fractions of the microsomal isolates were enriched by partitioning in an aqueous two-phase system, also as previously described [41]. The quality of obtained

microsomes was tested with 9-amino-6-chloro-2-methoxy-acridine (ACMA; Invitrogen A1324) fluorescence quenching assays.

### Transport analysis with plasma membrane vesicles

The transport of phenolic compounds uptake was studied by the rapid filtration technique with 4-coumarate, liquiritigenin, isoliquiritigenin, 7,4'-dihydroxyflavone, and the plasma membrane microsomes using nitrocellulose filters (0.45  $\mu\text{m}$  pore-size; Millipore). The transport assays were performed with microsomes corresponding to 520  $\text{ng } \mu\text{L}^{-1}$  protein concentration mixed with transport buffer (10 mM Tris-HCl, 10 mM EDTA, 10% sucrose, pH 5.0), the selected phenolic (750  $\mu\text{M}$ ), 100  $\mu\text{g mL}^{-1}$  creatine kinase, 10 mM creatine-phosphate, and 1 mM of  $\text{MgCl}_2$ , in the presence and absence of 4 mM ATP. After 3 min incubation at 24 °C, 0.3 mL of each reaction mixture was immediately loaded on a prewetted filter and rapidly washed with 10 mL of ice-cold transport buffer. The filters were air dried for an hour, and then incubated in 80% MeOH with 0.1% formic acid. Phenolic compounds were extracted by adding chloroform:water mixture (1:0.25 sample volume) to the sample and centrifugation for 30 min in 13,200 rpm. The dried samples were dissolved in 80% methanol and subjected to HPLC/MS analyses, as described below. For competition assays, tested molecules were added, each at 750  $\mu\text{M}$ , together to the transport buffer. Experiments were repeated three times with independent vesicle preparations unless stated otherwise.

### HPLC/MS analysis

Samples were analyzed by liquid chromatography–electrospray ionization–tandem mass spectrometry (LC/ESI/MS) using a Waters UPLC Acquity system, equipped with a C18 RP column, connected to a Bruker micrOTOF-Q II mass spectrometer. The mobile phase consisted of a gradient of 0.5% formic acid (v/v) in water (A) and 0.5% formic acid (v/v) in acetonitrile (B). The  $m/z$  range of the recorded MS spectra was 50–1000. The MS was operated in positive and negative ion modes for phenolics and carboxylic acids, respectively.

### Multiple sequence alignment and data filtering

To select full-size ABCG from multiple groups of plants, we subjected One Thousand Plant Transcriptomes (1KP) transcript data (<https://db.cngb.org/onekp/>) [42, 43] to tBLASTn searches using MtABCG46 as a query sequence with an E-value cutoff of  $1\text{e}-5$ . Next, 26,889 1KP samples longer than 1000 bp were translated into six frames using software written in Visual C#. The longest ORFs starting with the ATG codon, after excluding the duplicates, were selected

and further verified, by a BLASTp search against known ABC transporters belonging to different subfamilies was conducted. One thousand eight hundred and thirty-nine 1KP samples were assigned to a full-size ABCG subfamily with over 70% coverage and an E-value of 0.0, and used for the subsequent analysis. Multiple sequence alignment (MSA) of the full-size ABCG amino acid sequences was performed using the MUSCLE algorithm [44] in MEGA X [45]. For conservation analysis, complete alignment of predicted amino acid sequences of 1576 plant full-size ABCG transporters was submitted to the ConSurf server with default settings [46, 47]. We also submitted extracted taxa-specific alignments (66, 205, 91, 388, and 725—green alga, bryophyte, pteridophyte, monocots, and core eudicot sequences, respectively) to the ConSurf server, and subjected the complete alignment (1576 sequences) to co-evolution analysis using the Gremlin server [48, 49] with default settings.

### Modeling of the 3D structure

The amino acid sequence of the ABCG46 transporter of *Medicago truncatula* was obtained from the UniProt database (accession no. A0A396JDZ5), and submitted to a local installation of AlphaFold2 v2.1.0 using the default settings [31]. The resulting models were evaluated using the pLDDT score [31] and PROCHECK software [50].

To define positions of the  $Mg^{2+}$  ions, experimental structures of ABCG2 from *Homo sapiens* (PDB ID: 6hbu) and Pdr5 from *Saccharomyces cerevisiae* (PDB ID: 7p06) were employed as templates. The NBD region of the model and experimental structures were superimposed using TM-align software [51], and positions of the  $Mg^{2+}$  ions were copied to the model. For the ATP molecules, the NBD regions were also superimposed, but coordinates of the ATP atoms were not used directly. Instead, a docking box enclosing the molecules was built, then AutoDock Vina v1.1.2 [52] was used to determine their most suitable positions. Mutations at residue F562 to alanine, leucine, and tyrosine were performed with the tleap module of the Amber20 package [53], and one system of the wild-type protein was built without ATP and  $Mg^{2+}$  ions as an apo variant.

### Molecular dynamics (MD) simulations

The protein model was protonated with the H++ server [54] at pH 7.0. The protonated protein was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer using the CHARMM-GUI server with integrated PPM 2.0 method to obtain the system's coordinates in PDB format [55–59]. For the system, the size of the simulation box was set to 130 Å, with water thickness of 15 Å, and KCl salt concentration of 0.1 M. Further ATP molecules and  $Mg^{2+}$  ions were inserted as described above and the PDB file

was converted to AMBER parameters and topology using tleap module of Amber20 package [53]. The water model employed was OPC [60], and the HMR [61] method was used to enable 4 fs simulation timesteps. The MD engine employed was Amber20 [53] with the pmemd GPU implementation [62], the ff19SB force field [63] was used for the protein, lipid17 for the POPC membrane, and previously presented parameters for the ATP molecules [64]. The system was minimized with 2500 steps of steepest descent, followed by 2500 steps of conjugated gradient, applying only positional restraints to the protein, ATP, and  $Mg^{2+}$  ions, but positional and dihedral restraints to the membrane (Supplementary Table 1). After the minimization, a series of 1 ns NVT and NPT equilibration simulations with Langevin thermostat and Monte Carlo barostat used as appropriate was employed to release restraints applied on the system (Supplementary Table 1). After the equilibration stage, an extra 100 ns NPT equilibration simulation was applied in which restraints were only maintained for the  $Mg^{2+}$  and ATP molecules (Supplementary Table 1). Finally, 100 ns unrestrained NPT simulation to fully equilibrate the system was performed, followed by a production phase of 400 ns unrestrained NPT simulation (Supplementary Table 1). For every variant considered, five replicas of the production phase were simulated.

To simulate conversion to more open IF conformation, umbrella sampling (US) [65] was employed on all protein variants with two aims: to obtain the structures in their IF-open conformations and study the dynamics in that conformation, and to evaluate the energetic cost of the opening process. Both processes employed the same procedure and input structures, varying in the biasing potential and simulation time for each umbrella. The open ScPDR5 structure with bound rhodamine 6G substrate was used as the target (PDB ID: 7p05). The US simulations were performed with the Amber20 package, using the RMSD to backbone atoms of the transmembrane helices of the target as the collective variable (Supplementary Fig. 1). For each variant, the US simulations were initiated from the structures exhibiting the lowest RMSD to the backbone atoms of the transmembrane helices of the target. Following, the ATP molecules of each system were replaced by ADP molecules present in the target structure, and the parameters were modified accordingly [64]. The systems were minimized with 1000 steps of steepest descent, followed by 1000 steps of conjugated gradient. Next, 250 ps of equilibration with NPT simulation were performed at 303.15 K and 1 bar using Langevin thermostat and Monte Carlo barostat. The US protocol to obtain IF-open structures consisted of 22, 21, 22, and 24 simulation windows for WT, F562L, F562Y, and F562A, respectively, equidistantly positioned (step 0.1 Å RMSD) along the collective variable until the target RMSD of 0 Å. For each window, a 1 ns NPT simulation was performed at 303.15 K and 1 bar

using the Langevin thermostat and Monte Carlo barostat and the force constant of  $1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . At the end of the US runs, the final structures for each variant were used in MD simulations that followed the same simulation protocol described earlier. For each variant, 100 ns of MD without restraints was performed as the equilibration stage, followed by five replicates with 400 ns of unrestrained MD as the production stage. To evaluate the energetic cost of transition from IF-closed to IF-open states in each variant, three opening simulations were performed with 1, 2, and 4 ns of simulation time per window. In these US simulations, the RMSD step for each window was reduced to  $0.02 \text{ \AA}$  to guarantee sufficiently large overlaps between windows for later analysis, using the following force constants:  $500 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ , and increased to  $1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  from 1.7 to  $1.40 \text{ \AA}$  RMSD, resulting in 36, 31, 36, and 46 simulation windows for WT, F562L, F562Y, and F562A, respectively. The potential of mean force was calculated with the weighted histogram analysis method [66] implemented by Grossfield, in version 2.0.11 [67] ([http://membrane.urmc.rochester.edu/?page\\_id=126](http://membrane.urmc.rochester.edu/?page_id=126)).

### Access path detection, classification, and selection

For path computation and detection, CAVER v3.0 software [68] was employed with a  $0.9 \text{ \AA}$  probe radius,  $6 \text{ \AA}$  shell radius, and  $4 \text{ \AA}$  shell depth. For clustering, the average-link hierarchical algorithm was used, the maximum number of clusters was set to 50, and clustering threshold to 3.5. The starting position of the CAVER calculation was set to employ the center of mass of residues 562, 566, 1213, and 1217. Subsequently, TransportTools software v0.9.0 [69] was used to obtain a comprehensive and comparative view of the path network across all variants sourced from all converged CAVER calculations. In TransportTools, the clustering method was set to complete with a clustering cutoff of  $0.5 \text{ \AA}$  for the analysis of the wild-type protein, then switched to the average method for the comparison of all variants in IF-closed and IF-open states simultaneously using clustering cutoffs of 2.0 and  $0.5 \text{ \AA}$ , respectively, while all other parameters were left as default. The candidate selected as the correct access path was further sorted by the bottleneck radius and the 100 widest paths of each variant were used in the analysis and further ligand migration experiments.

### Ligand migration and energy barrier calculation

To study the capability of each variant to transport the four tested phenolic compounds, molecular docking across selected paths was performed with CaverDock software v1.1 [70]. 3D structures of the ligands of the four compounds were obtained from the PubChem database: liquiritigenin (CID 114829), isoliquiritigenin (CID 638278), 4-coumarate

(CID 637542), and 7,4'-dihydroxyflavone (CID 5282073), then the MGLTools v1.5.6 [71] was used to prepare the files for CaverDock with the `prepare_ligand4.py` script with the default settings. In the same way, all the snapshots of the MDs from the variants where a path was selected were processed with the `prepare_receptor4.py` script with default settings. The exhaustiveness for CaverDock runs was set to one, using a single CPU core per task whereas eight parallel workers unique to CaverDock algorithm were used to enable extensive sampling of each path segment [70]. Finally, the upper-bound trajectories describing continuous ligand migration were analyzed with in-house Python scripts.

### Auxiliary analyses of data from MD simulations

Using the `cpptraj` [72] module of the Amber20 package, initially the heavy atoms' root mean square deviation (RMSD) and the root mean square fluctuation (RMSF) of the whole system were calculated. However, since the main focus was the TMD region, separate RMSDs for specific regions were also calculated (Supplementary Fig. 2b and Supplementary Table 2). For the membrane, the lipid order parameters of lipid tails were evaluated together with the mass density with `cpptraj` to ensure the proper equilibrium behavior of the POPC membrane. To evaluate the behavior of the TMD helices forming the candidate access path,  $\alpha$  helices 2, 5, 8, and 11 were analyzed with the HELANAL module of MDAnalysis v2.0.0 [73, 74].

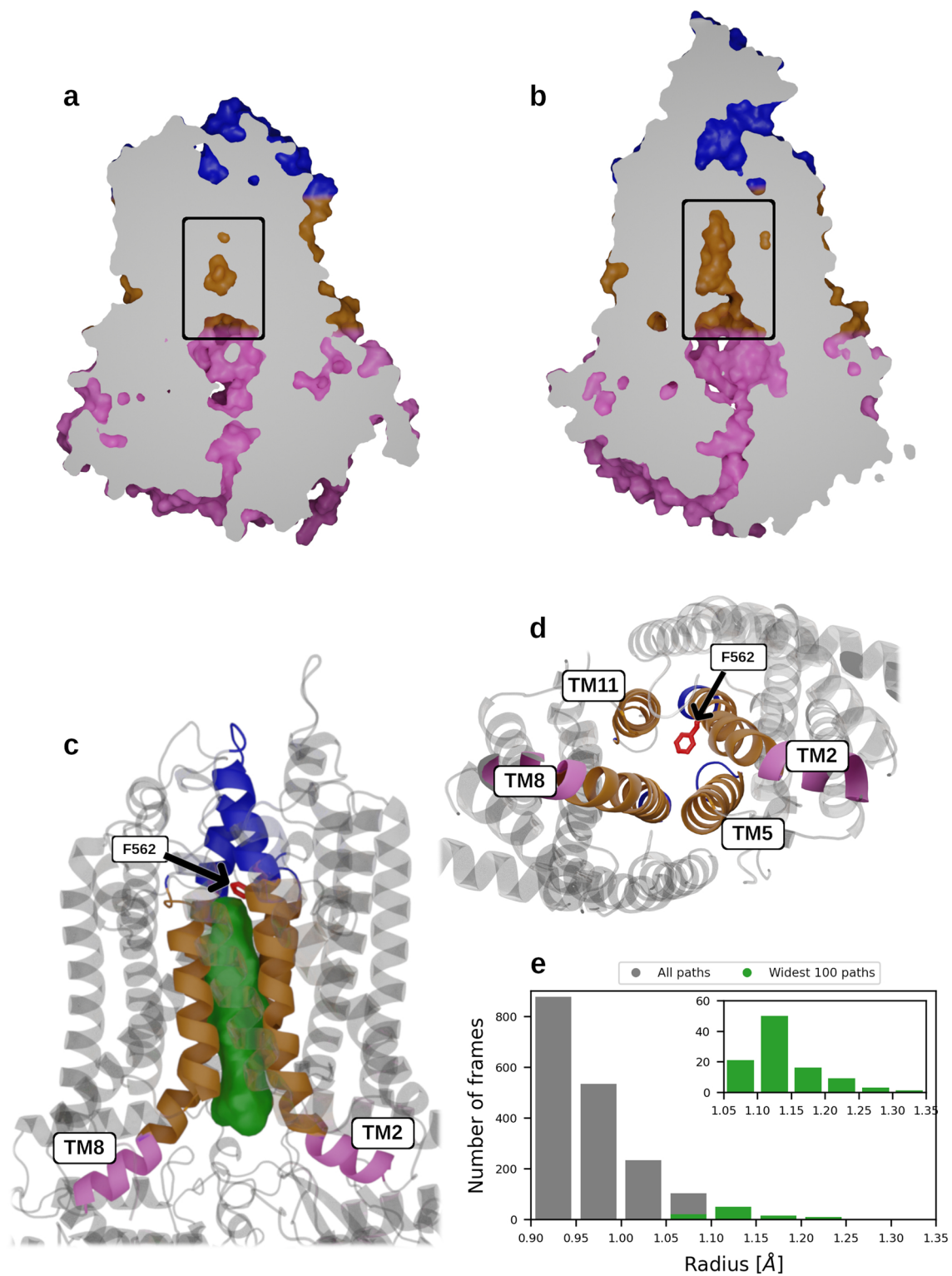
### Statistical analyses

Statistical analyses were performed using GraphPad Prism software v9.0. The normality of distribution assumption was assessed for particular groups of values by Anderson–Darling, Shapiro–Wilk, and Kolmogorov–Smirnov normality tests. If data met normal distribution criteria, an unpaired  $t$  test with Welch's correction was applied. If data did not meet normal distribution criteria, non-parametric tests were applied (the two-tailed Mann–Whitney test or Kruskal–Wallis test with post hoc Dunn's multiple comparison test).  $P$  values obtained can be found in the Supplementary Statistical Data.

## Results

### The MtABCG46 model features an unusually occluded central cavity connected with the intracellular environment by a transient access path

The MtABCG46 model obtained from AlphaFold2 has good overall quality according to the pLDDT score [31],



with limited confidence for only a few regions (Supplementary Fig. 2a). Moreover, these regions with the low pLDDT scores correspond to highly dynamic motifs in related ABCG proteins [26, 28, 29], for which single structures are hard to define. According to the template

modeling score (TM-score) [51], the model has a very similar fold to the recent full-size PDR5 structure of *Saccharomyces cerevisiae* (ScPDR5) [28], human ABCG1 (HsABCG1) [26], and half-size human ABCG2 (HsABCG2) [29] with TM-scores of 0.82, 0.72, and 0.66,

**Fig. 1** Structure and dynamics of the central cavity and its access path in MtABCG46. **a** Sliced surface side view of the MtABCG46 model obtained from AlphaFold2 featuring an occluded central cavity (indicated by a black rectangle). **b** Sliced surface side view of a cryo-EM structure of ScPDR5 (PDB ID: 7p04) with an open cavity (indicated by a black rectangle). **c** Structural representation of part of MtABCG46, showing the mutated residue F562 (in red sticks), the overall volume of the access path ensemble (green), and TMD helices forming it (intracellular region in pink, transmembrane region in gold, and extracellular region in blue). **d** The bottom view on transmembrane region of MtABCG46 from intracellular side, showing mutated residues F562 (in red sticks), and four TMD helices forming the path and central cavity (intracellular region in pink and transmembrane region in gold). **e** Bottleneck radius distribution of access path ensemble connecting the cavity in MD simulations of MtABCG46, the widest 100 paths are colored green and shown in the inset

respectively (Supplementary Fig. 3). In the structure, the typical arrangement of domains in a full-size ABCG transporter is clearly visible (Supplementary Fig. 2b), with TMD regions, each composed of 6  $\alpha$ -helices, and the two half-size domains joined by a ~55-residue linker region. A comparison of central cavities of MtABCG46 and ScPDR5 revealed a small disconnected cavity in the TMD region of MtABCG46 (Fig. 1a). In contrast, a wide-open cavity in the TMD region that directly connects with the cytosolic environment was observed in ScPDR5 (Fig. 1b).

To explore possible paths enabling access of cognate ligands into the central cavity, we analyzed molecular dynamics (MD) simulations for temporarily opened continuous internal voids formed within the MtABCG46 structure using CAVER 3.0 [68] and TransportTools [69]. This revealed a network of putative transport paths from the bottom of the cavity to the intracellular, extracellular, and membrane regions (Supplementary Fig. 4a). To help identification of the likeliest localization of a functionally relevant entrance to such a path along the membrane's Z-axis, we ran six independent MD simulations of systems composed of the endogenous MtABCG46 substrate, liquiritigenin, and a membrane. These simulations showed that liquiritigenin preferentially stayed between the heads and tails of the membrane phospholipids (Supplementary Fig. 4b, c). Consequently, among all paths that opened in this region, the most prevalent (open for ~9% of the total simulation time) was the third-ranked one (Supplementary Table 3) with an average bottleneck radius of  $0.97 \pm 0.06$  Å (Fig. 1e). This path provided the most straightforward access to the central cavity from the intracellular region (Fig. 1c and Supplementary Fig. 4d). Interestingly, an equivalent path was practically undetectable in simulations of MtABCG46 without bound ATP molecules (Supplementary Fig. 5). The remaining paths leading to the intracellular region were not considered due to their much lower frequency and much longer, curved geometry (Supplementary Table 4).

## Phylogenetic analyses of residues of the MtABCG46 central cavity

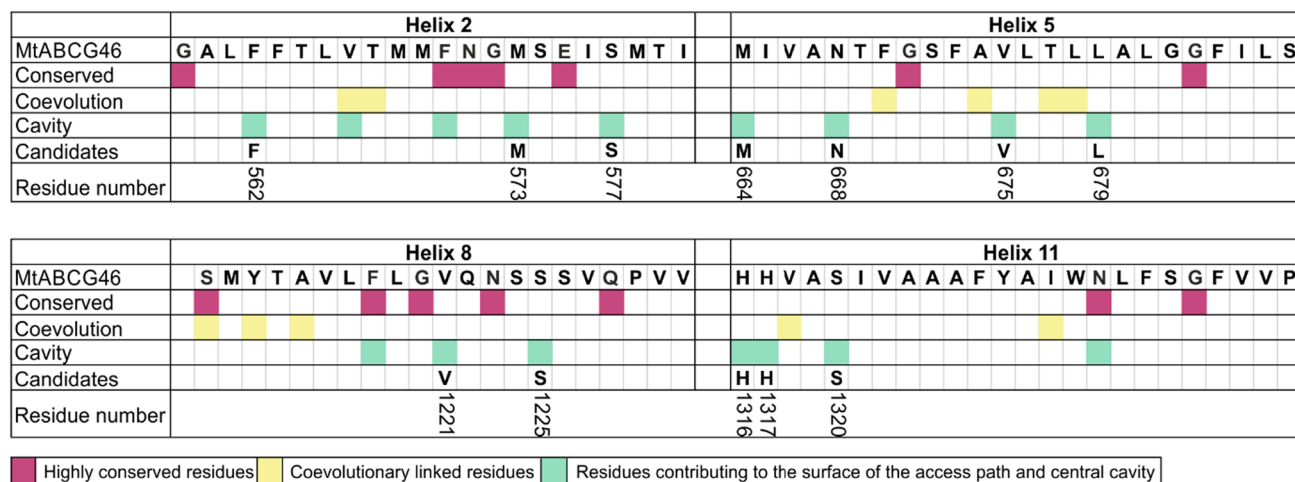
To investigate the importance of the architecture of internal voids in MtABCG46 for the passage of specific phenylpropanoids, we mapped residues contributing to the surface of the access path and central cavity during the MD simulations (Supplementary Fig. 6). We hypothesized that variations in the recognition and transport of diverse molecules by various ABCG proteins may arise from differences in the amino acid sequences that form the cavity. Based on this assumption, we generated multiple sequence alignment (MSA) of predicted amino acid sequences of 1839 plant full-size ABCG transporters extracted from the 1KP project [42, 43]. The latter collected transcriptomic data from more than 1000 species spanning a diversity of plant kingdom. The sequences were analyzed using the ConSurf [46, 47] and Gremlin [48, 49] servers. We then selected residues that contribute to the central cavity, are not fully conserved (ConSurf grade  $\leq 8$ ), and display variability that could not be readily explained by co-evolutional links with other residues (Fig. 2a).

One of the selected amino acids, residue F562 in TMD helix 2, particularly drew our attention due to its correspondence to F431 in HsABCG2 (Fig. 2 and Supplementary Fig. 7), a putatively important residue for ligand recognition and binding [75]. Residue F431 is highly conserved in human ABCG transporters and fully conserved among ABCG2 homologs in several animal species (Fig. 2b, c). However, in plant full-size ABCG transporters, variations at the F562 position include amino acids such as tyrosine, leucine, and isoleucine (Fig. 2d). Intriguingly phylogenetic analyses revealed that the prevalence of amino acids other than phenylalanine at this position is significantly higher in seed plants than in non-seed plants (Fig. 2e). Notably, such expansion of variability in seed plants was not observed for other residues that directly contribute to the cavity surface (Supplementary Fig. 8). Since seed plants have a higher degree of chemodiversity, emphasized in associated specialized metabolism, this led to the conclusion that variability at this position might be meaningful for plant ABCG transporters and/or possibly MtABCG46 selectivity.

## F562 substitutions profoundly affect the selectivity of MtABCG46-mediated transport

To address the importance of F562 for MtABCG46-mediated transport of phenylpropanoids, we substituted it for the other two most frequent amino acids at this position in plant full-size ABCG transporters, tyrosine and leucine (Fig. 2e), as well as alanine. Alanine is often used in such analyses as it eliminates the sidechain beyond the  $\beta$  carbon, but does not alter the main-chain conformation [76].

a



b

**Human ABCG paralog**

HsABCG2	G	V	L	F	F
HsABCG1	G	F	L	F	F
HsABCG4 isoform a	G	C	L	F	F
HsABCG4 isoform b	G	C	L	F	F
HsABCG5	G	L	L	Y	Q
HsABCG8	A	L	L	F	M

c

**HsABCG2 homologs in different animal species**

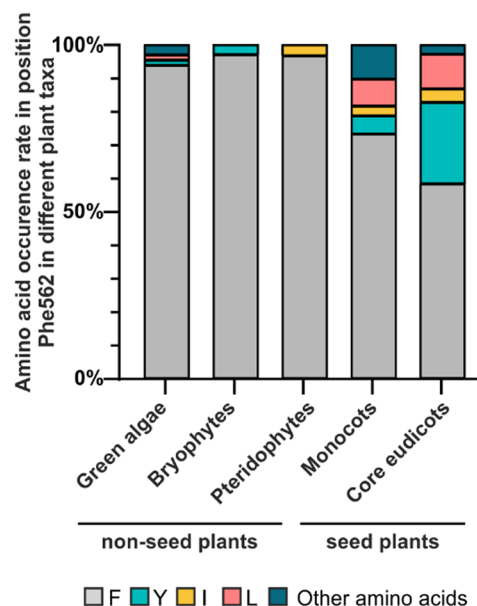
Homo sapiens (HsABCG2)	G	V	L	F	F
Pan troglodytes	G	V	L	F	F
Macaca mulatta	G	V	L	F	F
Canis lupus familiaris	G	V	L	F	F
Bos taurus	G	V	L	F	F
Mus musculus	G	V	L	F	F
Rattus norvegicus	G	V	F	F	F
Gallus gallus	G	S	L	F	F
Danio rerio	G	A	L	F	F
Drosophila melanogaster	G	A	I	F	L
Xenopus tropicalis	G	S	L	F	F

d

**Selected plant full-size ABCGs**

MtABCG46	G	A	L	F	F
CrTPT2	G	A	L	Y	F
HvABCG31	G	A	L	Y	F
OsABCG31	G	A	L	Y	F
AtABCG32	G	S	L	Y	F
AtABCG35	G	S	L	L	F
AtABCG36	G	A	L	L	F
VmABCG1/VmTPT2	G	A	L	Y	F
NtPDR3	G	A	L	F	Y
NbABCG1	G	A	L	F	F
NbABCG2	G	A	L	F	F
NtPDR1	G	A	L	F	F
MtABCG52	G	A	I	L	F
MtABCG54	G	A	I	L	F
MtABCG56	G	A	L	F	F

e



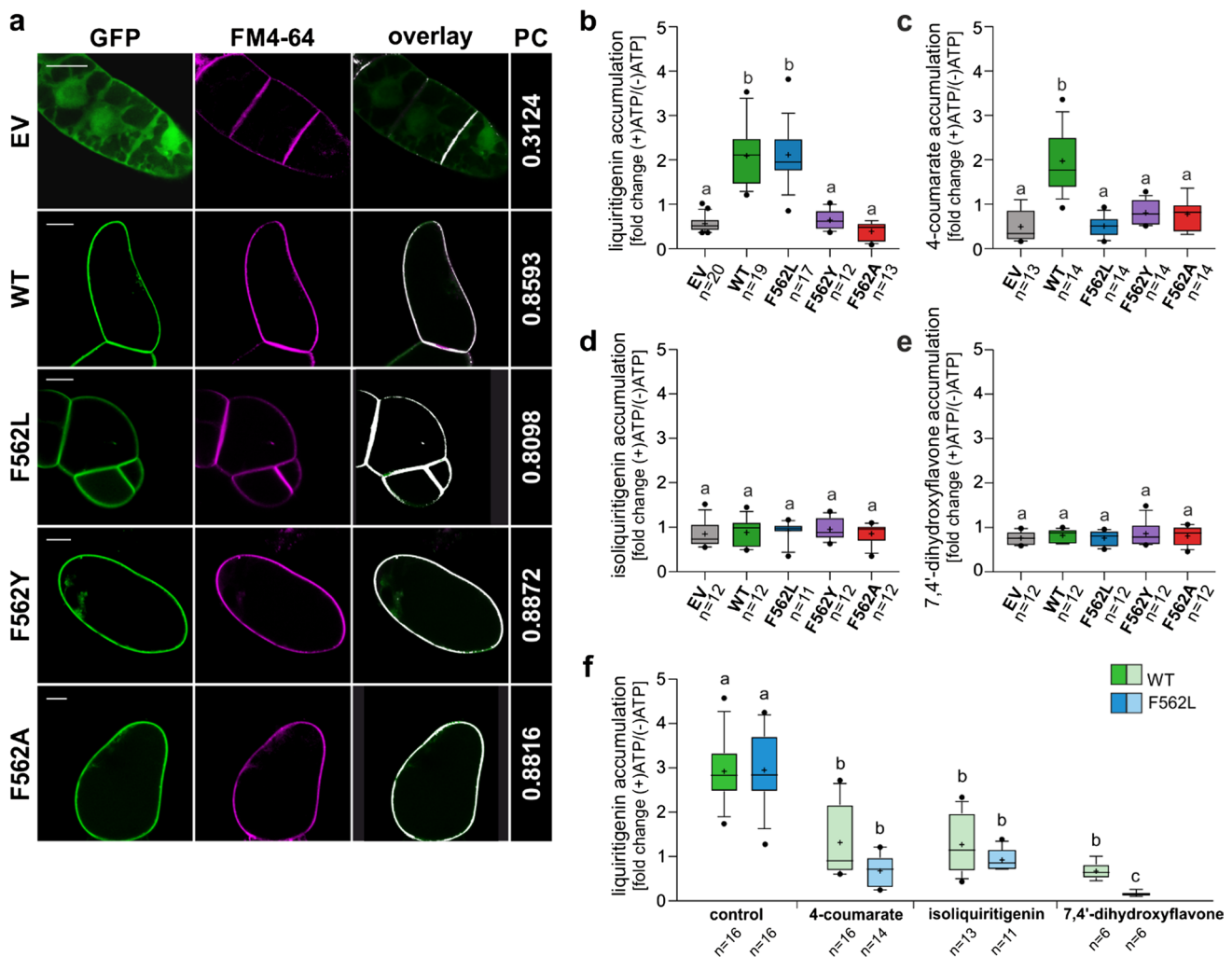
**Fig. 2** Residue selection for the site-directed MtABCG46 mutagenesis. **a** Schematic representation of candidate residues in sequences of MtABCG46 transmembrane helices 2, 5, 8, and 11 selected using conservation degrees (with Consurf), co-evolutionary parameters (with Gremlin), and the structural data. Candidate residues numbered in accordance with the MtABCG46 sequence. Alignment of

the region corresponding to surroundings of F431 in HsABCG2 for human ABCGs (**b**) and animal homologs (**c**). **d** Alignment of the region corresponding to surroundings of F562 in MtABCG46 for selected plant full-size ABCG transporters. **e** Frequencies of occurrence of indicated amino acids in the residue corresponding to MtABCG46 F562 in full-size ABCG sequences of indicated taxa

All our MtABCG46 variants, including the native form (hereafter: wild type, WT), tagged with GFP at the N terminus, were introduced into BY2 tobacco suspension cell cultures, a well-established heterologous expression systems for biochemical studies of ABCG proteins [8, 77]. To confirm the presence and correct localization in the plasma membrane (PM) of the transporters in BY2 lines, the colocalization of GFP-tagged MtABCG46 variants with a PM marker, FM4-64, was checked with confocal microscopy (Fig. 3a).

Effects of mutations were investigated by ATP-dependent, MtABCG46-mediated, transport assays of liquiritigenin, 4-coumarate, isoliquiritigenin, and 7,4'-dihydroxyflavone into PM inside-out vesicles derived from BY2 lines. Consistent with previous observations [24], liquiritigenin and 4-coumarate accumulated in vesicles from lines expressing WT MtABCG46 (Fig. 3b, c) but not isoliquiritigenin or 7,4'-dihydroxyflavone (Fig. 3d, e). Further experiments with lines expressing variants of MtABCG46



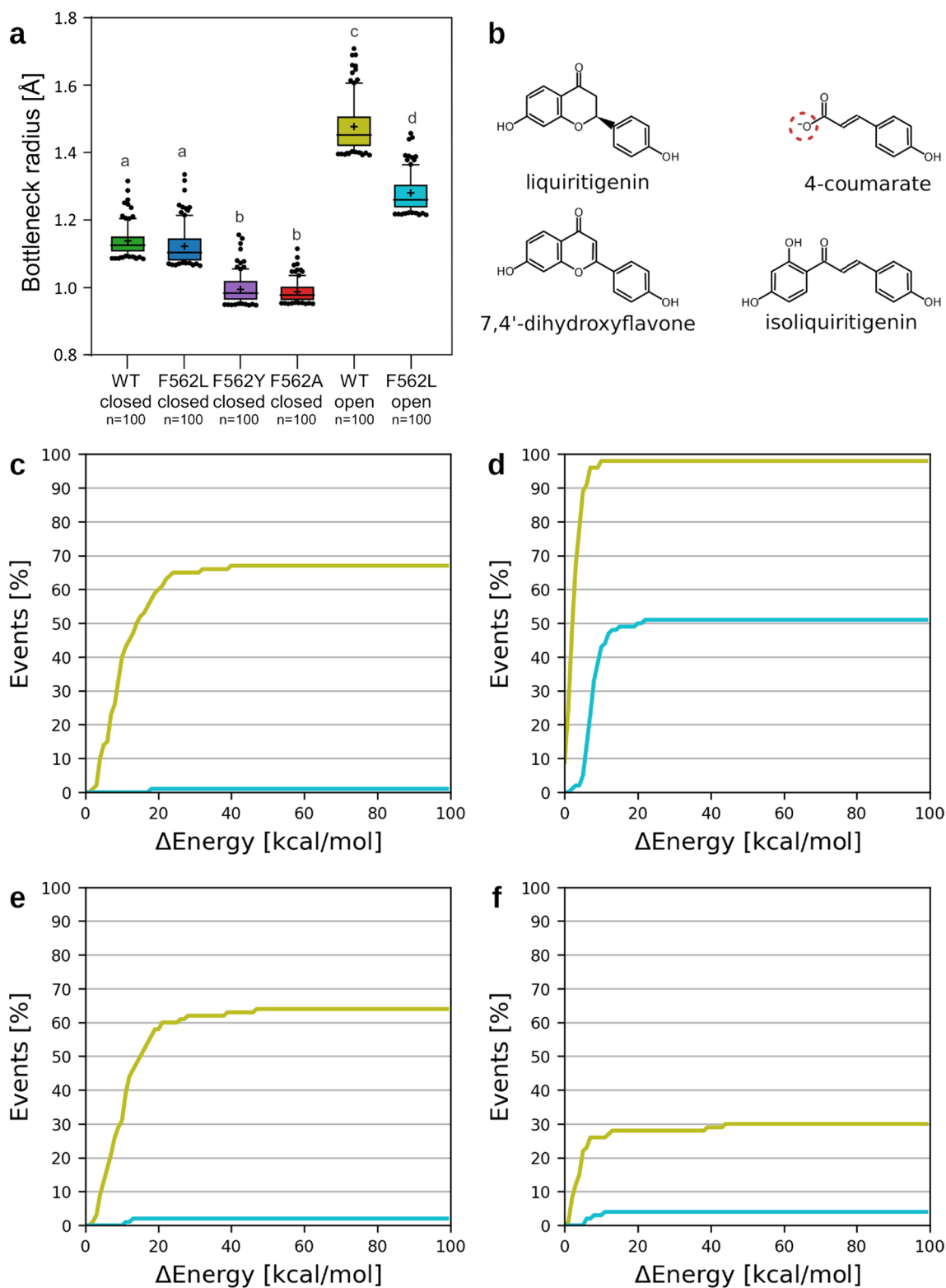


**Fig. 3** Transport assays with microsomes derived from BY2 suspension cell cultures. **a** Plasma membrane localization of the MtABCG46 variants tagged with GFP: WT, F562L, F562A, and F562Y in BY2 suspension cell cultures. Cell cultures expressing empty vector (EV) were used as a control. Images with GFP and FM4-64 fluorescence pseudo-colored in green and magenta, respectively. PC—Pearson's correlation coefficients for the colocalization of GFP and FM4-64 in the plasma membrane, visualized as an overlay. Scale bars, 20  $\mu$ m. Transport of liquiritigenin (**b**), 4-coumarate (**c**), isoliquiritigenin (**d**), and 7,4'-dihydroxyflavone (**e**) in microsomes derived from BY2 suspension cell cultures expressing empty vector (gray) or indicated variants of MtABCG46: WT (green), F562L (blue), F562Y (purple), and F562A (red). **f** Competitive transport of liquiritigenin versus 4-coumarate, isoliquiritigenin or 7,4'-dihydroxy-

flavone in microsomes expressing WT and F562L MtABCG46,  $n \geq 6$  means  $\pm$ SD from two to three biological replications, each with at least three technical replications. **b–f** Values are presented as fold change between (+)ATP and (–)ATP as a control. In each box-and-whiskers plot: the central black line indicates the median; '+' indicates the mean; the box extends from the 25th to the 75th percentile; the whiskers extend from the 10th to the 90th percentile, and points below and above the whiskers are marked by individual dots. Different lowercase letters indicate significant differences:  $P < 0.001$  (**b**);  $P < 0.005$  (**f**);  $P < 0.01$  (**c**);  $P < 0.05$  (**d**, **e**).  $P$  values, determined by the Kruskal–Wallis test with a post hoc Dunn's multiple comparison test (**b–f**) and unpaired  $t$  test with Welch's correction or Mann–Whitney test (**f**) can be found in Supplementary Statistical Data

revealed that F562Y and F562A substitutions abolished transport of liquiritigenin and 4-coumarate (Fig. 3b, c). None of the variants were able to translocate isoliquiritigenin nor 7,4'-dihydroxyflavone (Fig. 3d, e). Interestingly, the F562L variant, accumulated liquiritigenin similarly to the WT MtABCG46, however, it was not able to transport 4-coumarate (Fig. 3b, c).

Previous research has shown that liquiritigenin is a competitor of 4-coumarate in MtABCG46-mediated transport [24]. Analyses of liquiritigenin transport in the presence of 4-coumarate confirmed that this effect is mutual, i.e., 4-coumarate also reduces the rate of MtABCG46-dependent liquiritigenin transport (Fig. 3f). Experiments also revealed that although F562L cannot transport



4-coumarate, the latter is still a competitor for liquiritigenin (Fig. 3f). Interestingly, presence of non-transported compounds isoliquiritigenin and 7,4'-dihydroxyflavone also negatively affected the MtABCG46-mediated

transport of liquiritigenin. Moreover, this susceptibility was increased in F562L variant compared to the WT, what was significantly visible in case of 7,4'-dihydroxyflavone (Fig. 3f).

**Fig. 4** Effects of F562 mutations on the accessibility of the path to the central cavity of MtABCG46. **a** Radius of the probe fitting the bottleneck region of the 100 most open access paths in MtABCG46. In each box-and-whiskers plot: the center black line indicates the median; '+' indicates the mean; the box extends from the 25th to 75th percentile; the whiskers extend from the 10th to the 90th percentile, points below and above the whiskers are marked as individual dots. Different lowercase letters indicate significant differences,  $P < 0.0005$ .  $P$  values, determined by the Kruskal–Wallis test with a post hoc Dunn's multiple comparison test, can be found in Supplementary Statistical Data. **b** Structures of ligands evaluated with CaverDock. The atom of 4-coumarate in red circle could be also be present in protonated form. Comparison of the migration for the two protonation forms is available in Supplementary Fig. 22. Cumulative distributions of energetic barriers for migration to the internal cavity for the widest 100 paths for liquiritigenin (**c**), 4-coumarate (**d**), 7,4'-dihydroxyflavone (**e**), and isoliquiritigenin (**f**). In **c–f** curves obtained from experiments with WT, and F562L in IF-open states are marked in pale green and cyan, respectively

### F562 substitutions affect the viability of the transient access path through rearranged TMD helices

In MD simulations of our four variants (WT, F562L, F562Y, and F562A), there were no significant differences in overall stability or flexibility between the mutant and WT proteins (Supplementary Figs. 9–12). Structurally, the introduced modifications of residue 562 disrupted non-covalent contacts (Supplementary Fig. 13), changing the bending and twisting angles of the TMD helices forming the access path (Supplementary Figs. 14, 15). In WT, residue F562 maintained a parallel displaced  $\pi$ -stacking interaction with residue F684 during the simulations (Supplementary Fig. 13a), a feature shared with the F562Y mutant (Supplementary Fig. 13c). Loss of this non-covalent contact in F562A resulted in a considerable displacement of TMD helix 5 (Supplementary Fig. 13d). F562L also lost the  $\pi$ -stacking interaction but maintained stability of this helix (Supplementary Fig. 13b) with interhelical space maintained at the level of WT (Supplementary Fig. 16a, b), likely due to remaining Van der Waals interactions of its sidechain. Interestingly, the displacement of TMD helix 5 in F562A allowed marked bending and to some extent twisting of TMD helix 8 (Supplementary Figs. 14c, 15c). Furthermore, the bending of TMD helices markedly reduced the available space between the helices and resulting in a more packed structure. This reduction of available space was observed in both F562A and F562Y mutants (Supplementary Fig. 16c, d), although no significant bending of TMD helix 8 was observed in the F562Y mutant. In this case, the additionally introduced hydroxyl group in F562Y formed an H-bond with Y1213 in TMD helix 8 during the whole simulation time. This pulled the modified residue and Y1213 closer to each other, thereby rearranging the orientation of the neighboring residues in TMD helix 8 and changing its twist helical angle

(Supplementary Fig. 15c). Moreover, residue Y1213 formed an H-bond with N1331, a highly conserved residue (Fig. 2a), in all performed MD simulations. Consequently, disturbance of the normal behavior of these residues could influence ligands' binding in the central cavity. To explore this possibility, we investigated the availability of polar interactions provided by residues of the central cavity. We found only three polar residues around the deepest part of the central cavity: Y1213, T1214, and N1331. Of these, T1214 has limited accessibility as it is oriented outwards of the cavity, and Y1213 maintains constant H-bond contact with the ketone group of N1331 in all MD simulations. Hence, the amine group of N1331 is the only one to act as a hydrogen donor. Notably, a high conservation score was obtained for residue N1331 in our MSA analysis (ConSurf grade 9), indicating that it has functional importance. In our MD simulations, all investigated mutations of residue F562 resulted in a considerable change in the conformation of N1331 (Supplementary Fig. 17) perturbing the putative interaction with bound ligands.

Nonetheless, there were considerable differences in the availability of access paths leading to the central cavity, which were lower in all mutants than in the WT (Supplementary Table 3 and Supplementary Fig. 18a). The 100 widest paths were almost as wide in the F562L mutant as in WT, but they were markedly narrower in F562Y and F562A mutants (Supplementary Fig. 18b), mainly because they had more constricted entrances (Fig. 4a, Supplementary Figs. 16, 19), presumably hampering access of bulky molecules to the central cavity. Accordingly, we hypothesized that F562Y and F562A mutations affect the structural arrangement of residues forming the access path, making it less permissive for migration of ligands and thereby significantly reducing its overall transport capability.

To address this possibility, we assessed the viability of the 100 widest paths in each variant for the access of liquiritigenin and 4-coumarate, as well as isoliquiritigenin and 7,4'-dihydroxyflavone, which are chemically similar but not effectively transported by MtABCG46 (Fig. 4b), from the intracellular environment into the central cavity using CaverDock [70]. Each run of CaverDock yields an energetic profile of the binding energy between a ligand and protein along the path, which can be translated as an energetic barrier that must be overcome for a ligand to reach the central cavity (Supplementary Fig. 20). Such an approach has proven utility in detecting hotspot residues for protein engineering [78], and correlating energetic barriers with biochemical rates [79, 80]. The cumulative distribution of the energy barriers for all ligand migration events resulting in favorable binding of ligands into the central cavity showed that none of the ligands except for 4-coumarate could reach the central cavity effectively (Supplementary Fig. 21). Even for WT and F562L, the

fraction of successful migration events was surprisingly low (about 20%, see Supplementary Fig. 21), clearly indicating that a more open state of IF conformation is required to efficiently initiate the transport process by ligand migration into the central cavity. Interestingly, the migration efficiency was similar also when the neutral form of this ligand was considered (Supplementary Fig. 22a).

To investigate the putative open state, we have employed umbrella sampling simulations to achieve the partial opening of the transmembrane helices by driving the system toward the conformation observed in the most open state of ScPDR5 as the only viable template for full-size ABCG transporters. Here, we would like to point out that the available ScPDR5 structures feature only limited openings of their central cavity in different IF conformations despite their pleiotropic nature [28] unlike widely open states observed in half-size human ABCGs, which are among others lacking the covalent linker present in full-size ABCGs. Notably, the opening processes in the slowest-driven US simulations were found to have 1.2, 9.4, and 11.6 kcal mol<sup>-1</sup> higher energetic costs in F562L, F562Y, and F562A, respectively, when compared to WT (Supplementary Fig. 23) at the conformation ensemble most similar to the arrangements observed in the target ScPDR5 (RMSD of ~1.5 Å) (Supplementary Fig. 24). This observation suggested that the compact interhelical space formed due to F562A and F562Y mutations (Supplementary Fig. 16c, d) rendered these variants unlikely to adopt such an open state in IF conformation and perform their function in contrast to WT and F562L, which were much more prone to undergo this transition.

The subsequent unbiased simulations of WT and F562L starting from their induced IF-open states, showed a partial reversal toward their closed states, reaching an RMSD of ~2.2 and 2.4 Å compared to the target ScPDR5, respectively (Supplementary Fig. 25). Despite such relaxations, these simulations still featured a more frequent opening of the continuous access paths for at least 31 and 12% of the total simulation time in WT and F562L, respectively (Supplementary Table 5). The bottlenecks of these paths could be fitted with spherical probes of up to 1.7 and 1.4 Å radii in WT and F562L, respectively (Fig. 4a and Supplementary Table 5). Notably, the access paths ensemble in the open states of MtABCG46 also included branches that opened laterally to the lipid bilayer (Supplementary Fig. 26), allowing ligands to enter either from the membrane or through the cytosolic region, in contrast to the access paths found in the closed states. Additionally, the interhelical region of WT in the open state underwent significant enlargement along the majority of its length (Supplementary Fig. 16e), expanding not only the entry filter but also the volume of the central cavity. In contrast, conformations from F562L simulations

mostly showed only minor enlargements in this region (Supplementary Fig. 16f).

When considering the explicit calculation of ligand migration, performed with CaverDock, the revealed expansions enabled favorable access of all investigated ligands to the central cavity of MtABCG46 WT, exhibiting distinct preference for 4-coumarate. It was illustrated by almost 100% rate of successful migration events (Fig. 4d), irrespectively of the protonation form (Supplementary Fig. 22b). The much bulkier liquiritigenin and 7,4'-dihydroxyflavone reached the cavity in about 60% of migration events, with a slight preference for the former (Fig. 4c, e). In contrast, the least favorable migration was calculated for isoliquiritigenin, with only 30% rate of successful migration events (Fig. 4f). In the F562L mutant, the preference concerned only the 4-coumarate, which retained about half of its efficiency, compared to WT (Fig. 4d). All bulkier compounds reached similar efficiencies when it comes to migrating to the central cavity of F562L, with much lower success rates and considerably higher energy costs (Fig. 4c, e, f). The partial discrepancy, visible in the overall transport assays, corresponds well with the disruption of the helices and the central cavity, caused, to different degrees, by all three mutations investigated in F562 (Supplementary Fig. 16). This suggests that these mutations may also affect subsequent stages of the transport cycle.

## Discussion

AlphaFold2 has recently proposed suitability for modeling ABCG proteins, with the ability to provide similar levels of accuracy as for soluble proteins [32]. The MtABCG46 structure we obtained using AlphaFold2 has the typical architecture of full-size ABCG transporters described in the literature [81]. Moreover, our MD simulations identified transiently formed access paths to the central cavity from the intracellular region that were narrower than the ones observed in IF-ScPRD5 structures, irrespective of the adopted states (Fig. 1, Supplementary Fig. 27 and Supplementary Tables 4 and 5). Such open cavities were observed in HsABCG1 [26], HsABCG2 [29], and a recently obtained cryo-EM structure, ScPDR5 [29] (Supplementary Fig. 28). Also several substrates such as rhodamine 6G, cholesterol, and mitoxantrone have been found to occupy equivalent regions of ScPDR5, HsABCG1, and HsABCG2, respectively [26, 28, 29]. However, context of the access path to the cavity, as well as its role in substrate recognition and transport, have not been thoroughly investigated.

Tested substitutions of the selected residue F562 in the central cavity (F562L, F562Y, and F562A) severely affected MtABCG46-mediated transport of 4-coumarate and liquiritigenin, highlighting this residue's importance

for phenylpropanoid transport in *M. truncatula*. Transport assays of four phenylpropanoids with MtABCG46 WT showed lack of isoliquiritigenin or 7,4'-dihydroxyflavone transport (Fig. 3d, e), in accordance with indications from CaverDock analyses in IF-open like conformation predicted very good efficiency of migration to the central cavity for both 4-coumarate and liquiritigenin (Fig. 4c, d). In contrast, likelihood of other two compounds to access the central cavity is lower, in particular for metabolic precursor isoliquiritigenin (Fig. 4e, f). We have previously shown that the MtABCG46 promotes response of *Medicago truncatula* to pathogen infection by efficient transport of liquiritigenin but not isoliquiritigenin resulting in the de novo biosynthesis of the pterocarpan phytoalexin medicarpin, derived from the 5-deoxyisoflavonoid branch. Our observation regarding the efficiency of migration to the central cavity supports the potential role of it in this process. This is because, both scenario and pathway wise, such preference toward liquiritigenin but not isoliquiritigenin could be beneficial since the latter is a direct precursor that has to be converted and not transported [23, 24].

Moreover, the observed migration priority is in agreement with the ability of isoliquiritigenin or 7,4'-dihydroxyflavone to interfere with liquiritigenin transport in MtABCG46 WT (Fig. 3f), while lack of clear preference for liquiritigenin entry into the central cavities in F562L mutant (Fig. 4c, e, f) agrees well with the increased susceptibility of this protein to the interference from the same compounds (Fig. 3f). In contrast, the lack of transport capabilities in F562A or F562Y (Fig. 3) is well in line with the observed collapse of the interhelical space lining the access path and consequent overstabilization of IF-closed state caused by their mutations, disallowing the necessary opening for efficient ligand transport (Fig. 4a and Supplementary Fig. 16, 23). Thus, our data indicate that restriction of access to the central cavity is a contributor to substrate selectivity in early stages of the transport process, probably highly relevant for overall activity of the transporter in competitive environment inside the cell, where structurally analogous metabolites are present. Also, the fact that all four phenylpropanoids can still access the central cavity of MtABCG46 WT and F562L clearly reveals the presence of additional molecular mechanisms behind the observed selectivity in subsequent stages of the transport cycle, e.g., recognition of the substrate bound in the central cavity or its stabilization during the subsequent, likely rate-determining, conformational change (from IF to OF) of the transporter.

Similarly to the corresponding residue F431 of HsABCG2, in ABCG46 F562 is located in the direct proximity of the short loop after TMD helix 5 (and TMD helix 11). This loop is called a valve and has been proposed to work as a molecular gate crucial for regulating the conformation-dependent substrate release [29, 33]. A

residue in this loop in another full-size plant ABCG transporter from *A. thaliana*, AtABCG36 was also highlighted as potentially important for substrate specificity [18], also supporting our results. Our simulations showed that F562 interacts with at least one of the residues within the valve, namely F684, behavior of which was substantially altered in F562A and F562L variants (Supplementary Fig. 13). This enabled us to speculate, that introduced substitutions of F562 could significantly alter this structural element important for the transport mechanism. The substitutions were also considerably affecting the conformation of highly conserved N1331 (Supplementary Fig. 17), being the only hydrogen donor in the deep part of the central cavity, which could in the consequence affect the substrate binding.

In general, the molecular determinants of sequence changes resulting from protein adaptation to various biochemical needs remain obscure. Our experimental data, together with phylogenetic observations, suggest that adaptation of ABCG proteins associated with evolutionary pressures in plants has resulted in a variability of some key residues, such as F562 (Fig. 2e). Such variability is essential for transporters to fulfill their functional roles in diverse, complex chemical and biological scenarios. We foresee sequence–structure–dynamics exploration fueled by AlphaFold2 presented here as an alternative mean to overcome limitations in structural studies of membrane transporters, which can help to identify residues that define functional properties of that important subfamily of ABC transporters.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00018-023-04751-6>.

**Acknowledgements** The computations were performed at the Poznan Supercomputing and Networking Center. We thank Dr. Mariusz Czarnocki-Cieciura (IIMCB Warsaw) for obtaining structural predictions from the local installation of AlphaFold2, enabling predictions using the full-length sequence of MtABCG46.

**Author contributions** MJ devised and supervised the project. JBr designed, interpreted, and supervised the computational experiments. KP and MG selected residues for the site-directed mutagenesis (conservation and co-evolution analyses). CSB performed, analyzed, and interpreted results of computational experiments. WBL and KP designed, performed, and interpreted results of the transport experiments (BY2 transformation, isolation of microsomes, transport and competition assays). AP contributed to BY2 transformation, isolation of microsomes, transport and competition assays. WBL was responsible for HPLC/MS analysis and statistical analysis. JBa, AB, MR. ABCG selected sequences acquired in the 1KP project. JBa and WBL performed microscopic observation. CSB, WBL, and KP prepared figures. KP, CSB, WBL, JBr, and MJ wrote the manuscript with the help of the co-authors. All authors discussed and approved the manuscript.

**Funding** This work was supported by the National Science Centre, Poland (Grant nos. 2017/27/B/NZ1/01090 to MJ and 2017/25/B/NZ1/01307 to JBr). KP and CSB are recipients of scholarships associated with the POWER project (ref. nos. POWR.03.02.00-00-1032/16 and POWR.03.02.00-00-1022/16, respectively).

**Data availability** Sequence data reported in this article can be found in the NCBI database under the following accession numbers: MtABCG46 (Medtr\_2g102670), MtABCG52 (Medtr\_8g014360), MtABCG54 (Medtr\_5g070320), MtABCG56 (Medtr\_2g101090), AtABCG32 (AT2G26910), AtABCG35 (AT1G15210), AtABCG36 (AT1G59870), NbABCG1a (BAR94041), NbABCG2b (BAR94044), NtPDR3 (Q5W274), NtPDR1 (NP\_001312599), HvABCG31/EIBI1 (BAK52288), OsABCG31 (Os01g0177900), CrTPT2 (KC511771), VmABCG1/VmTPT2 (KC511773), HsABCG1 isoform X1 XP\_011528108, HsABCG2 NP\_004818.2, HsABCG4 isoform a NP\_001335120.1, HsABCG4 isoform b NP\_001335121.1, HsABCG5 NP\_071881.1, HsABCG8 NP\_071882.1; HsABCG2 homologs: *Homo sapiens* NP\_004818.2, *Pan troglodytes* XP\_526633.3, *Macaca mulatta* NP\_001028091.1, *Canis lupus familiaris* NP\_001041486.1, *Bos taurus* NP\_001032555.2, *Mus musculus* NP\_036050.1, *Rattus norvegicus* NP\_852046.1, *Gallus* XP\_421638.4, *Danio rerio* NP\_001036240.1, *Drosophila melanogaster* NP\_001039227.1, *Xenopus tropicalis* NP\_476787.1. The sequences from 1KP analysis, predicted AlphaFold2 models, parameters and input files for MD simulations of all MtABCG46 variants, as well as key restart files, results of tunnel geometry and ligand migration analyses are available at <https://doi.org/10.5281/zenodo.7002738>.

## Declarations

**Conflict of interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. *Nat Rev Mol Cell Biol* 10(3):218–227. <https://doi.org/10.1038/nrm2646>
- Hwang JU, Song WY, Hong D, Ko D, Yamaoka Y, Jang S, Yim S, Lee E, Khare D, Kim K, Palmgren M, Yoon HS, Martinoia E, Lee Y (2016) Plant ABC transporters enable many unique aspects of a terrestrial plant's lifestyle. *Mol Plant* 9(3):338–355. <https://doi.org/10.1016/j.molp.2016.02.003>
- Banasiak J, Jasiński M (2022) ATP-binding cassette transporters in nonmodel plants. *New Phytol* 233(4):1597–1612. <https://doi.org/10.1111/nph.17779>
- Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, Geisler M, Klein M, Kolukisaoglu Ü, Lee Y, Martinoia E, Murphy A, Rea PA, Samuels L, Schulz B, Spalding EJ, Yazaki K, Theodoulou FL (2008) Plant ABC proteins—a unified nomenclature and updated inventory. *Trends Plant Sci* 13(4):151–159. <https://doi.org/10.1016/j.tplants.2008.02.001>
- Lefèvre F, Boutry M (2018) Towards identification of the substrates of ATP-binding cassette transporters. *Plant Physiol* 178(1):18–39. <https://doi.org/10.1104/pp.18.00325>
- Do THT, Martinoia E, Lee Y, Hwang JU (2021) 2021 update on ATP-binding cassette (ABC) transporters: how they meet the needs of plants. *Plant Physiol* 187(4):1876–1892. <https://doi.org/10.1093/plphys/kiab193>
- Do THT, Martinoia E, Lee Y (2018) Functions of ABC transporters in plant growth and development. *Curr Opin Plant Biol* 41:32–38. <https://doi.org/10.1016/j.pbi.2017.08.003>
- Kang J, Hwang JU, Lee M, Kim YY, Assmann SM, Martinoia E, Lee Y (2010) PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc Natl Acad Sci USA* 107(5):2355–2360. <https://doi.org/10.1073/pnas.0909222107>
- Ford RC, Beis K (2019) Learning the ABCs one at a time: structure and mechanism of ABC transporters. *Biochem Soc Trans* 47(1):23–36. <https://doi.org/10.1042/BST20180147>
- Prasad R, Goffeau A (2012) Yeast ATP-binding cassette transporters conferring multidrug resistance. *Annu Rev Microbiol* 66(1):39–63. <https://doi.org/10.1146/annurev-micro-092611-150111>
- Ito H, Gray WM (2006) A gain-of-function mutation in the Arabidopsis pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides. *Plant Physiol* 142(1):63–74. <https://doi.org/10.1104/pp.106.084533>
- Fourcroy P, Sisó-Terraza P, Sudre D, Savirón M, Rey G, Gaymard F, Abadía A, Abadía J, Álvarez-Fernández A, Briat JF (2014) Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. *New Phytol* 201(1):155–167. <https://doi.org/10.1111/nph.12471>
- Ziegler J, Schmidt S, Strehmel N, Scheel D, Abel S (2017) Arabidopsis transporter ABCG37/PDR9 contributes primarily highly oxygenated coumarins to root exudation. *Sci Rep* 7(1):3704. <https://doi.org/10.1038/s41598-017-03250-6>
- Aryal B, Huynh J, Schneuwly J, Siffert A, Liu J, Alejandro S, Ludwig-Müller J, Martinoia E, Geisler M (2019) ABCG36/PEN3/PDR8 is an exporter of the auxin precursor, indole-3-butyric acid, and involved in auxin-controlled development. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2019.00899>
- Matern A, Böttcher C, Eschen-Lippold L, Westermann B, Smolka U, Döll S, Trempel F, Aryal B, Scheel D, Geisler M, Rosahl S (2019) A substrate of the ABC transporter PEN3 stimulates bacterial flagellin (flg22)-induced callose deposition in *Arabidopsis thaliana*. *J Biol Chem* 294(17):6857–6870. <https://doi.org/10.1074/jbc.RA119.007676>
- Strader LC, Bartel B (2009) The Arabidopsis PLEIOTROPIC DRUG RESISTANCE8/ABCG36 ATP binding cassette transporter modulates sensitivity to the auxin precursor Indole-3-butyric acid. *Plant Cell* 21(7):1992–2007. <https://doi.org/10.1105/tpc.109.065821>
- Stein M, Dittgen J, Sánchez-Rodríguez C, Hou BH, Molina A, Schulze-Lefert P, Lipka V, Somerville S (2006) Arabidopsis PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell* 18(3):731–746. <https://doi.org/10.1105/tpc.105.038372>
- Lu X, Dittgen J, Pislewska-Bednarek M, Molina A, Schneider B, Svatos A, Doudsky J, Schneeberger K, Weigel D, Bednarek P, Schulze-Lefert P (2015) Mutant allele-specific uncoupling of penetration3 functions reveals engagement of the ATP-binding cassette transporter in distinct tryptophan metabolic pathways. *Plant Physiol* 168(3):814–827. <https://doi.org/10.1104/pp.15.00182>
- Kim DY, Bovet L, Maeshima M, Martinoia E, Lee Y (2007) The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant J* 50(2):207–218. <https://doi.org/10.1111/j.1365-3113X.2007.03044.x>

20. Takeuchi M, Watanabe A, Tamura M, Tsutsumi Y (2018) The gene expression analysis of *Arabidopsis thaliana* ABC transporters by real-time PCR for screening monoglucosyl-transferase candidates. *J Wood Sci* 64(5):477–484. <https://doi.org/10.1007/s10086-018-1733-9>
21. He Y, Xu J, Wang X, He X, Wang Y, Zhou J, Zhang S, Meng X (2019) The *Arabidopsis* pleiotropic drug resistance transporters PEN3 and PDR12 mediate camalexin secretion for resistance to *Botrytis cinerea*. *Plant Cell* 31(9):2206–2222. <https://doi.org/10.1105/TPC.19.00239>
22. Neelam KA, Sharma KK (2020) Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. *Crit Rev Food Sci Nutr* 60(16):2655–2675. <https://doi.org/10.1080/10408398.2019.1653822>
23. Banasiak J, Biała W, Staszuk A, Swarczewicz B, Kępczyńska E, Figlerowicz M, Jasiński M (2013) A *Medicago truncatula* ABC transporter belonging to subfamily G modulates the level of isoflavonoids. *J Exp Bot* 64(4):1005–1015. <https://doi.org/10.1093/jxb/ers380>
24. Biała W, Banasiak J, Jarzyski K, Pawela A, Jasiński M (2017) *Medicago truncatula* ABCG10 is a transporter of 4-coumarate and liquiritigenin in the medicarpin biosynthetic pathway. *J Exp Bot* 68(12):3231–3241. <https://doi.org/10.1093/jxb/erx059>
25. Kowal J, Ni D, Jackson SM, Manolaridis I, Stahlberg H, Locher KP (2021) Structural basis of drug recognition by the multidrug transporter ABCG2. *J Mol Biol*. <https://doi.org/10.1016/j.jmb.2021.166980>
26. Sun Y, Wang J, Long T, Qi X, Donnelly L, Elghobashi-Meinhardt N, Esparza L, Cohen JC, Xie XS, Hobbs HH, Li X (2021) Molecular basis of cholesterol efflux via ABCG subfamily transporters. *Proc Natl Acad Sci USA*. <https://doi.org/10.1073/PNAS.2110483118>
27. Skarda L, Kowal J, Locher KP (2021) Structure of the human cholesterol transporter ABCG1. *J Mol Biol* 433(21):167218. <https://doi.org/10.1016/J.JMB.2021.167218>
28. Harris A, Wagner M, Du D, Raschka S, Nentwig LM, Gohlke H, Smits SHJ, Luisi BF, Schmitt L (2021) Structure and efflux mechanism of the yeast pleiotropic drug resistance transporter Pdr5. *Nat Commun*. <https://doi.org/10.1038/s41467-021-25574-8>
29. Orlando BJ, Liao M (2020) ABCG2 transports anticancer drugs via a closed-to-open switch. *Nat Commun*. <https://doi.org/10.1038/s41467-020-16155-2>
30. Khunweeraphong N, Kuchler K (2021) Multidrug resistance in mammals and fungi—from mdr to pdr: a rocky road from atomic structures to transport mechanisms. *Int J Mol Sci* 22(9):4806. <https://doi.org/10.3390/IJMS22094806>
31. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2>
32. Hegedűs T, Geisler M, Lukács GL, Farkas B (2022) Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cell MolLife Sci* 79(1):73. <https://doi.org/10.1007/s00018-021-04112-1>
33. Khunweeraphong N, Szöllösi D, Stockner T, Kuchler K (2019) The ABCG2 multidrug transporter is a pump gated by a valve and an extracellular lid. *Nat Commun* 10(1):5433. <https://doi.org/10.1038/s41467-019-13302-2>
34. Nagy T, Tóth Á, Telbisz Á, Sarkadi B, Tordai H, Tordai A, Hegedűs T (2021) The transport pathway in the ABCG2 protein and its regulation revealed by molecular dynamics simulations. *Cell Mol Life Sci* 78(5):2329–2339. <https://doi.org/10.1007/S00018-020-03651-3>
35. Ferreira RJ, Bonito CA, Cordeiro MNDS, Ferreira MJU, dos Santos DJVA (2017) Structure–function relationships in ABCG2: insights from molecular dynamics simulations and molecular docking studies. *Sci Rep* 7(1):15534. <https://doi.org/10.1038/s41598-017-15452-z>
36. Gose T, Shafi T, Fukuda Y, Das S, Wang Y, Allcock A, Gavan MA, Lynch J, Chen T, Tamai I, Shelat A, Ford RC, Schuetz JD (2020) ABCG2 requires a single aromatic amino acid to “clamp” substrates and inhibitors into the binding pocket. *FASEB J* 34(4):4890–4903. <https://doi.org/10.1096/fj.201902338RR>
37. Nagata T, Nemoto Y, Hasezawa S (1992) Tobacco BY-2 cell line as the “HeLa” cell in the cell biology of higher plants. *Int Rev Cytol* 132(C):1–30. [https://doi.org/10.1016/S0074-7696\(08\)62452-3](https://doi.org/10.1016/S0074-7696(08)62452-3)
38. Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* 133(2):462–469. <https://doi.org/10.1104/pp.103.027979>
39. Hellens R, Mullineaux P, Klee H (2000) A guide to agrobacterium binary Ti vectors. *Trends Plant Sci* 5(10):446–451. [https://doi.org/10.1016/S1360-1385\(00\)01740-4](https://doi.org/10.1016/S1360-1385(00)01740-4)
40. Jasiński M, Stukkens Y, Degand H, Purnelle B, Marchand-Brynaert J, Boutry M (2001) A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell* 13(5):1095–1107
41. Larsson C, Widell S, Kjellbom P (1987) Preparation of high-purity plasma membranes. *Methods Enzymol* 148:558–568. [https://doi.org/10.1016/0076-6879\(87\)48054-3](https://doi.org/10.1016/0076-6879(87)48054-3)
42. Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA, Li L, Du W, Ullrich K, Wickett NJ, Barkmann TJ, Barker MS, Leebens-Mack JH, Wong GKS (2019) Access to RNA-sequencing data from 1173 plant species: the 1000 Plant transcriptomes initiative (1KP). *Gigascience*. <https://doi.org/10.1093/gigascience/giz126>
43. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, Porsch M, Quint M, Rensing SA, Soltis DE, Soltis PS, Stevenson DW, Ullrich KK, Wickett NJ, DeGironimo L, Edger PP, Jordon-Thaden IE, Joya S, Liu T, Melkonian B, Miles NW, Pokorny L, Quigley C, Thomas P, Villarreal JC, Augustin MM, Barrett MD, Baucom RS, Beerling DJ, Benstein RM, Biffin E, Brockington SF, Burge DO, Burris JN, Burris KP, Burtet-Saramagna V, Caicedo AL, Cannon SB, Çebi Z, Chang Y, Chater C, Cheeseman JM, Chen T, Clarke ND, Clayton H, Covshoff S, Crandall-Stotler BJ, Cross H, dePamphilis CW, Der JP, Determann R, Dickson RC, di Stilio VS, Ellis S, Fast E, Feja N, Field KJ, Filatov DA, Finnegan PM, Floyd SK, Fogliani B, García N, Gâteblé G, Godden GT, Goh F, Qi Y, Greiner S, Harkess A, Heaney JM, Helliwell KE, Heyduk K, Hibberd JM, Hodel RGJ, Hollingsworth PM, Johnson MTJ, Jost R, Joyce B, Kapralov MV, Kazamia E, Kellogg EA, Koch MA, von Konrat M, Könyves K, Kuchan TM, Lam V, Larsson A, Leitch AR, Lentz R, Li FW, Lowe AJ, Ludwig M, Manos PS, Mavrodiev E, McCormick MK, McKain M et al (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574(7780):679–685. <https://doi.org/10.1038/s41586-019-1693-2>
44. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797. <https://doi.org/10.1093/nar/gkh340>
45. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35(6):1547–1549. <https://doi.org/10.1093/molbev/msy096>

46. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44(1):344–350. <https://doi.org/10.1093/NAR/GKW408>
47. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem* 53(3–4):199–206. <https://doi.org/10.1002/ijch.201200096>
48. Kamisetty H, Ovchinnikov S, Baker D (2013) Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci USA* 110(39):15674–15679. <https://doi.org/10.1073/pnas.1314045110>
49. Ovchinnikov S, Kamisetty H, Baker D (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife*. <https://doi.org/10.7554/eLife.02030>
50. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26(2):283–291. <https://doi.org/10.1107/S0021889892009944>
51. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins Struct Funct Bioinform* 57(4):702–710. <https://doi.org/10.1002/PROT.20264>
52. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461. <https://doi.org/10.1002/JCC.21334>
53. Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Berryman JT, Brozell SR, Cerutti DS, Cheatham III TE, Cisneros GA, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R, Izadi S, Izmailov SA, Kasavajhala K, Kaymak MC, King E, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Machado M, Man V, Manathunga M, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, O’Hearn KA, Onufriev A, Pan F, Pantano S, Qi R, Rahnamoun A, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shajan A, Shen J, Simmerling CL, Skrynnikov NR, Smith J, Swails J, Walker RC, Wang J, Wang J, Wei H, Wolf RM, Wu X, Xiong Y, Xue Y, York DM, Zhao S, Kollman PA (2020) AMBER 20. <http://ambermd.org/>
54. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40(W1):W537–W541. <https://doi.org/10.1093/nar/gks375>
55. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 29(11):1859–1865. <https://doi.org/10.1002/JCC.20945>
56. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614. <https://doi.org/10.1002/JCC.21287>
57. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, Jo S, Pande VS, Case DA, Brooks CL, MacKerell AD, Klauda JB, Im W (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J Chem Theory Comput* 12(1):405–413. <https://doi.org/10.1021/ACS.JCTC.5B00935>
58. Lee J, Hitzenberger M, Rieger M, Kern NR, Zacharias M, Im W (2020) CHARMM-GUI supports the Amber force fields. *J Chem Phys* 153(3):035103. <https://doi.org/10.1063/5.0012280>
59. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, Klauda JB, Im W (2014) CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J Comput Chem* 35(27):1997–2004. <https://doi.org/10.1002/JCC.23702>
60. Izadi S, Anandakrishnan R, Onufriev AV (2014) Building water models: a different approach. *J Phys Chem Lett* 5(21):3863–3871. <https://doi.org/10.1021/JZ501780A>
61. Hopkins CW, Grand S, Le WRC, Roitberg AE (2015) Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* 11(4):1864–1874. <https://doi.org/10.1021/CT5010406>
62. Le GS, Götz AW, Walker RC (2013) SPFP: speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun* 184(2):374–380. <https://doi.org/10.1016/J.CPC.2012.09.022>
63. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Miguez AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C (2020) Ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* 16(1):528–552. <https://doi.org/10.1021/ACS.JCTC.9B00591>
64. Meagher KL, Redman LT, Carlson HA (2003) Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* 24(9):1016–1025. <https://doi.org/10.1002/jcc.10262>
65. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys* 23(2):187–199. [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8)
66. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13(8):1011–1021. <https://doi.org/10.1002/JCC.540130812>
67. Grossfield A WHAM: the weighted histogram analysis method. [http://membrane.urmc.rochester.edu/?page\\_id=126](http://membrane.urmc.rochester.edu/?page_id=126). Accessed 26 Jan 2023
68. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, Biedermannova L, Sochor J, Damborsky J (2012) CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* 8(10):e1002708. <https://doi.org/10.1371/JOURNAL.PCBI.1002708>
69. Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Fomthum CJ, Agrawal N (2022) TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them. *Bioinformatics* 38(6):1752–1753. <https://doi.org/10.1093/BIOINFORMATICS/BTAB872>
70. Vavra O, Filipovic J, Plhak J, Bednar D, Marques SM, Brezovsky J, Stourac J, Matyska L, Damborsky J (2019) CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels. *Bioinformatics* 35(23):4986–4993. <https://doi.org/10.1093/bioinformatics/btz386>
71. Morris GM, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785–2791. <https://doi.org/10.1002/JCC.21256>
72. Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9(7):3084–3095. <https://doi.org/10.1021/CT400341P>



73. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAanalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* 32(10):2319–2327. <https://doi.org/10.1002/JCC.21787>
74. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, Domański J, Dotson DL, Buchoux S, Kenney IM, Beckstein O (2016) MDAanalysis: a python package for the rapid analysis of molecular dynamics simulations. Proceedings of the 15th Python in Science Conference. pp 98–105. <https://doi.org/10.25080/MAJORA-629E541A-00E>
75. Jackson SM, Manolaridis I, Kowal J, Zechner M, Taylor NMI, Bause M, Bauer S, Bartholomaeus R, Bernhardt G, Koenig B, Buschauer A, Stahlberg H, Altmann KH, Locher KP (2018) Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol* 25(4):333–340. <https://doi.org/10.1038/s41594-018-0049-1>
76. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* (1979) 244(4908):1081–1085. <https://doi.org/10.1126/science.2471267>
77. Toussaint F, Pierman B, Bertin A, Lévy D, Boutry M (2017) Purification and biochemical characterization of NpABCG5/NpPDR5, a plant pleiotropic drug resistance transporter expressed in *Nicotiana tabacum* BY-2 suspension cells. *Biochem J* 474(10):1689–1703. <https://doi.org/10.1042/BCJ20170108>
78. Papadopoulou A, Meierhofer J, Meyer F, Hayashi T, Schneider S, Sager E, Buller R (2021) Re-programming and optimization of a L-Proline *cis*-4-Hydroxylase for the *cis*-3-halogenation of its native substrate. *Chem Cat Chem* 13(18):3914–3919. <https://doi.org/10.1002/CCTC.202100591>
79. Marques SM, Bednar D, Damborsky J (2019) Computational study of protein-ligand unbinding for enzyme engineering. *Front Chem* 7(JAN):650. <https://doi.org/10.3389/FCHEM.2018.00650>
80. Brodsky K, Kutý M, Pelantová H, Cvačka J, Rebroš M, Kotik M, Smatanová IK, Křen V, Bojarová P (2020) Dual substrate specificity of the Rutinosidase from *Aspergillus niger* and the role of its substrate tunnel. *Int J Mol Sci* 21(16):5671. <https://doi.org/10.3390/IJMS21165671>
81. Thomas C, Tampé R (2020) Structural and mechanistic principles of ABC transporters. *Annu Rev Biochem* 89:605–636. <https://doi.org/10.1146/ANNUREV-BIOCHEM-011520-105201>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.