Role of a transient access path within ABCG protein involved in a selective transport of legume phenylpropanoids

Konrad Pakuła[1,2*], Carlos Sequeiros-Borja[3,4*], Wanda Biała-Leonhard[1*], Aleksandra Pawela[1], Joanna Banasiak[1], Aurélien Bailly[5], Marcin Radom[6,7], Markus Geisler[8], Jan Brezovsky[3,4]✉, Michał Jasiński[1,9] ✉

1. Department of Plant Molecular Physiology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
2. NanoBioMedical Centre, Adam Mickiewicz University, Poznan, Poland
3. Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland
4. International Institute of Molecular and Cell Biology, Warsaw, Poland
5. Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland
6. Department of Structural Bioinformatics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
7. Institute of Computing Science, Poznan University of Technology, Poznan, Poland
8. Department of Biology, University of Fribourg, Fribourg, Switzerland
9. Department of Biochemistry and Biotechnology, Poznan University of Life Sciences, Poznan, Poland

* These authors contributed equally to the work

✉ For correspondence: **Michał Jasiński**, Department of Plant Molecular Physiology, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznan, Poland (+48) 61 852 85 03 (e-mail jasinski@ibch.poznan.pl); **Jan Brezovsky**, Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland, and International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland (+48) 61 829 5839 (e-mail janbre@amu.edu.pl, jbrezovsky@iimcb.gov.pl).

Running title: ABCG protein and selective transport of phenylpropanoids

ABCG46 of the legume *Medicago truncatula* is an ABC-type plasma membrane transporter that selectively translocates endogenous phenylpropanoids from the biosynthetic pathway of the phytoalexin, medicarpin, namely 4-coumarate and liquiritigenin, and no other structurally very similar metabolites. 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 these 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' impact on the selective transport of specialized metabolites may provide new opportunities for the exploitation of transporters' activities in novel applications concerning agriculture and biotechnology.

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 (Rees et al., 2009). Members of the ABC family are common in all domains of life, but plant genomes are exceptionally rich in genes encoding them (Hwang et al., 2016; Banasiak and Jasiński, 2022). Taking into account their structure and phylogenetic relationships, most ABC proteins have been classified into eight subfamilies, designated ABCA-ABCH (Verrier et al., 2008). 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 (Lefèvre and Boutry, 2018). In this manner, they participate in diverse processes, including organ growth, nutrition, development, responses to abiotic stresses, as well as both symbiotic and antagonistic relationships (Hwang et al., 2016; Do et al., 2018; Do et al., 2021).

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 5 to 10 transmembrane α-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 (Verrier et al., 2008; Kang et al., 2010). Moreover, in comparison to thoroughly studied ABCB proteins, where α-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 (Ford and Beis, 2019). 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 (Prasad and Goffeau, 2012).

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) (Ito and Gray, 2006) and the phenolic compound scopoletin (Fourcroy et al., 2014; Ziegler et al., 2017). Arabidopsis ABCG36/PEN3/PDR8 transports, *inter alia*, IBA, the Brassicales-specific phytoalexin camalexin, heavy metals, and possibly monolignols (Stein et al., 2006; Kim et al., 2007; Strader and Bartel, 2009; Lu et al., 2015; Takeuchi et al., 2018; Matern et al., 2019; Aryal et al., 2019; He et al., 2019). However, despite initially considered as functional homologs of yeast's multidrug pumps, plant full-size ABCG proteins appear to be more selective towards translocated molecules. The specialization is proposed to be a consequence of sophisticated chemodiversity exemplified by specialized metabolism that requires tightly controlled distribution of metabolites via dedicated transporters and notably ABCG proteins (Lefèvre and Boutry, 2018; Banasiak and Jasiński, 2022). Currently the molecular determinants behind plant ABCG's selectivity are not known. A *Medicago truncatula* ABCG46 (formerly known as ABCG10), 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 (Banasiak et al., 2013; Biała et al., 2017). Thus, we decided to use this model to better understand the discrimination phenomenon.

Despite the recent improvement in structural research regarding ABCG proteins (Orlando and Liao, 2020; Kowal et al., 2021; Sun et al., 2021; Skarda et al., 2021; Harris et al., 2021), progressing from static atomic structures to an understanding of molecular mechanisms of

ligand recognition and transport has been challenging (Khunweeraphong and Kuchler, 2021), partly because of the difficulty in obtaining experimental data at atomic detail, and partly because of the lack of efficient sampling of the intricate process of a complete transport cycle. Of note, AlphaFold2 (Jumper et al., 2021) 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 trans-membrane hydrophobic proteins that are difficult to crystalize (Hegedűs et al., 2022). 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 ligand migration from the cytosolic environment to the central cavity. The protein then undergoes large structural rearrangements to an outward-facing (OF) conformation, enabling release of the ligand to the extracellular environment. Binding, hydrolysis and release of ATP as well as ligand migration and recognition contribute to the intricacy of the overall transport process (Orlando and Liao, 2020; Harris et al., 2021).

As described here, using a combination of phylogenetic and biochemical analyses, AlphaFold2 structure prediction, molecular dynamics simulations, and mutagenesis we have identified a region in MtABCG46, specifically its transient access path, that is directly involved in the recognition and passage of 4-coumarate and liquiritigenin through the cellular membrane. Moreover, we have identified F562 as a critical residue for the architecture of the central cavity and in particular the access path, mutations of which affect this selective transport.

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 per-residue lDDT-Cα metric (Mariani et al., 2013), with limited confidence for only a few regions (Supplemental Figure 1). Moreover, these regions correspond to highly dynamic motifs in related ABCG proteins (Orlando and Liao, 2020; Sun et al., 2021; Harris et al., 2021), for which single structures are hard to define, attesting to high model quality. According to the template modeling score (TM-score) (Zhang and Skolnick, 2004), the model has a very similar fold to the recent full-size PDR5 structure of *Saccharomyces cerevisiae* (ScPDR5) (Harris et al., 2021), human ABCG1 (HsABCG1) (Sun et al., 2021), and half-size human ABCG2 (HsABCG2) (Orlando and Liao, 2020) with TM-scores of 0.82, 0.72, and 0.66, respectively (Supplemental Figure 2). In the structure, the typical arrangement of domains in a full-size ABCG transporter is clearly visible (Supplemental Figure 3), with TMD regions, each

composed of six α-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 (Figure 1A). In contrast, a wide-open cavity in the TMD region that directly connects with the cytosolic environment was observed in ScPDR5 (Figure 1B).

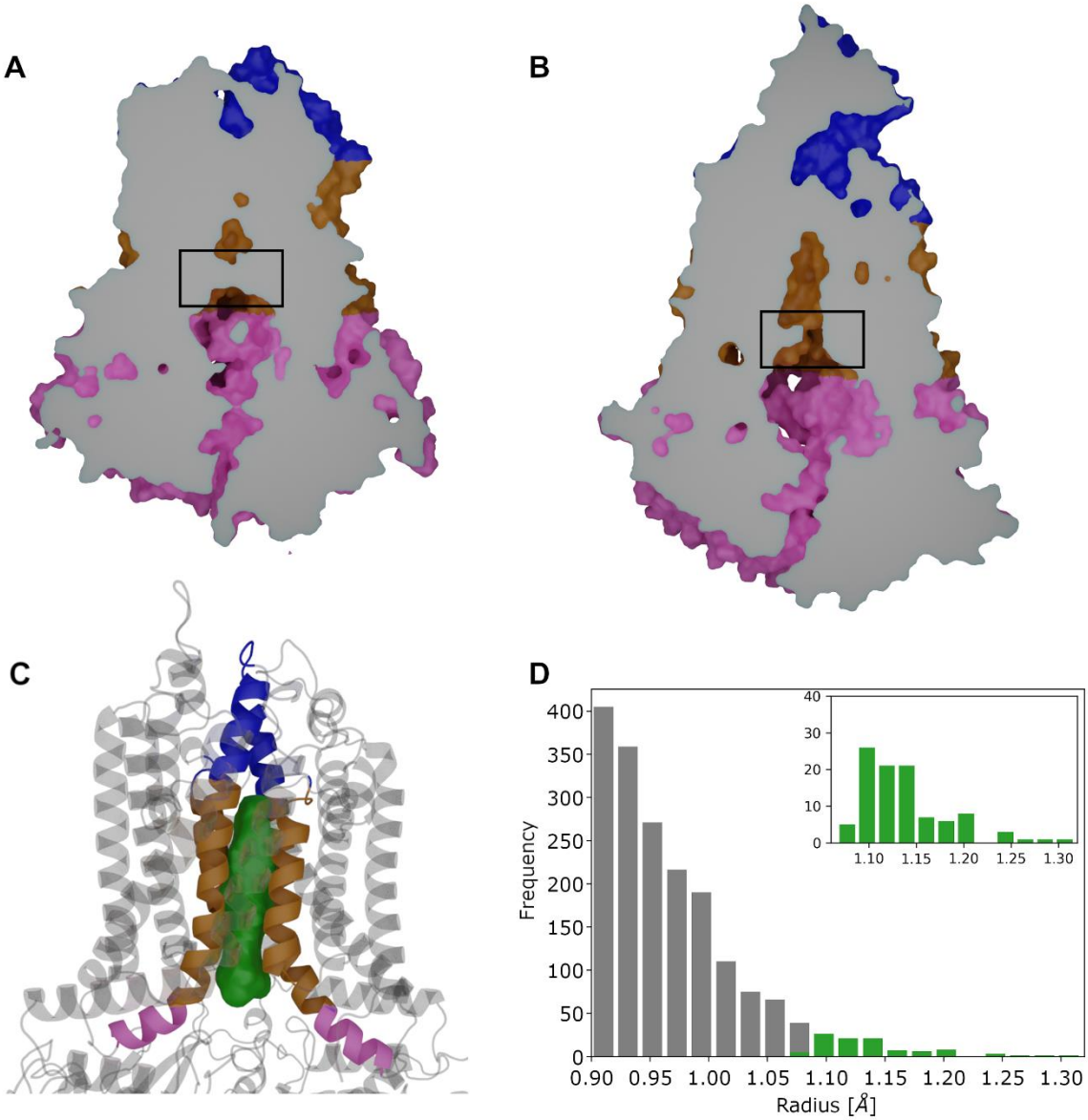

Figure 1. Structure and dynamics of the central cavity and its access path in MtABCG46. (A) Sliced surface side view of the MtABCG46 model obtained from AlphaFold2 featuring an occluded central cavity (the closure indicated by a black rectangle). **(B)** Sliced surface side view of a cryo-EM structure of ScPDR5 (PDB ID: 7p04) with an open cavity (indicated by a black rectangle). **(C)** Structural representation of part of MtABCG46, showing the overall volume of the access path ensemble (green) and TMD helices forming it (intra-cellular region in pink, trans-membrane region in gold, and extra-

cellular region in blue). **(D)** 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.

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 (Chovancova et al., 2012) and TransportTools (Brezovsky et al., 2022). This revealed a network of putative transport paths from the bottom of the cavity to the intracellular, extracellular and membrane regions (Supplemental Figure 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 the endogenous substrate of MtABCG46, liquiritigenin, with the membrane. These simulations showed that liquiritigenin preferentially stayed between the heads and tails of the membrane phospholipids (Supplemental Figure 4B and 4C). 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 (Supplemental Table 1) with an average bottleneck radius of 0.97±0.06 Å (Figure 1D). This path provided the most straightforward access to the central cavity from the intracellular region (Figure 1D and Supplemental Figure 4D). Interestingly, an equivalent path was practically undetectable in simulations of MtABCG46 without bound ATP molecules (Supplemental Figure 5). The remaining paths leading to the intracellular region were not considered due to their much lower frequency and much longer, curved geometry (Supplemental Table 1).

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 (Supplemental Figure 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 1 KP project (Carpenter et al., 2019; Leebens-Mack et al., 2019), and analyzed it using the ConSurf (Celniker et al., 2013; Ashkenazy et al., 2016) and Gremlin (Kamisetty et al., 2013; Ovchinnikov et al., 2014) servers. We then selected residues that contribute to the central cavity, are not fully

conserved (ConSurf grade ≤ 8), and display variability that could not be readily explained by co-evolutional links with other residues (Figure 2A).

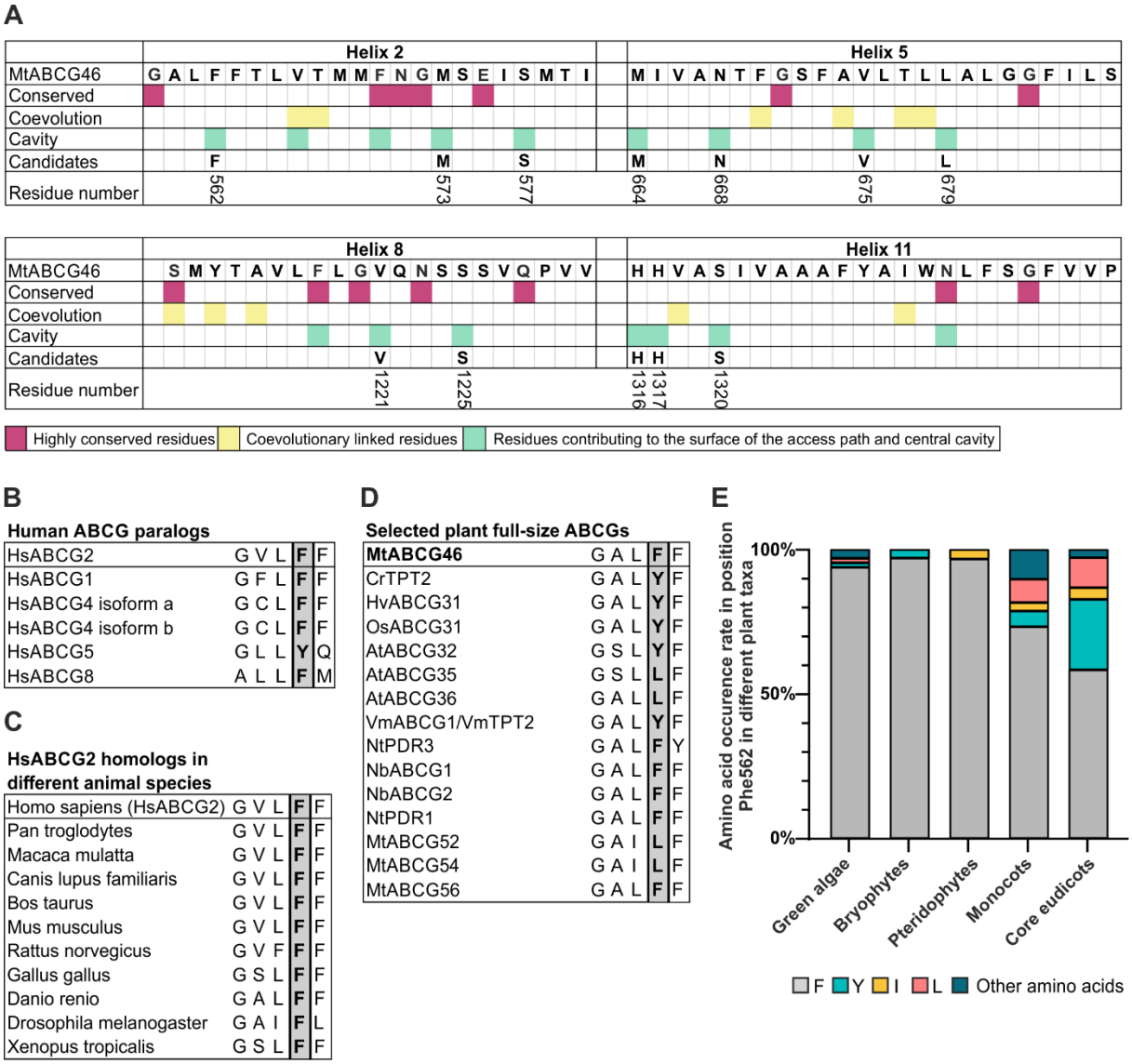

Figure 2. Residue selection for the site-directed MtABCG46 mutagenesis. (A) Schematic representation of candidate residues in sequences of MtABCG46 transmembrane helices 2, 5, 8 and 11 selected using conservation degrees (with Consurf), coevolutionary parameters (with Gremlin) and the structural data.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

One of the selected amino acids, residue F562 in TMD helix 2, particularly drew our attention due to its correspondence to F431 in HsABCG2 (Figure 2 and Supplemental Figure 7), a putatively crucial residue for ligand recognition and binding (Jackson et al., 2018). Residue F431 is highly conserved in human ABCG transporters and fully conserved among ABCG2 homologs in several animal species (Figure 2B and 2C). However, in plant full-size ABCG transporters, variations at the F562 position include amino acids such as tyrosine, leucine, and isoleucine (Figure 2D). Phylogenetic analyses revealed that the prevalence of amino acids other than phenylalanine at this position is significantly higher in monocots and eudicots than in non-seed plants (Figure 2E). Notably, such expansion of variability in angiosperms was not observed for other residues that directly contribute to the cavity surface (Supplemental Figure 8). Since angiosperms have a higher degree of chemodiversity (and 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 of phenylpropanoids

To address F562's importance for the MtABCG46-mediated transport of endogenous *M. truncatula* phenylpropanoids, we substituted it for the other two most frequent amino acids at this position in plant full-size ABCG transporters, tyrosine and leucine (Figure 2E), as well as alanine. Alanine is often used in such analyses as it eliminates the side-chain beyond the β carbon, but does not alter the main-chain conformation (Cunningham and Wells, 1989).

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, which are commonly used as heterologous expression systems for biochemical studies of ABCG proteins (Kang et al., 2010; Toussaint et al., 2017). 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 (Figure 3A).

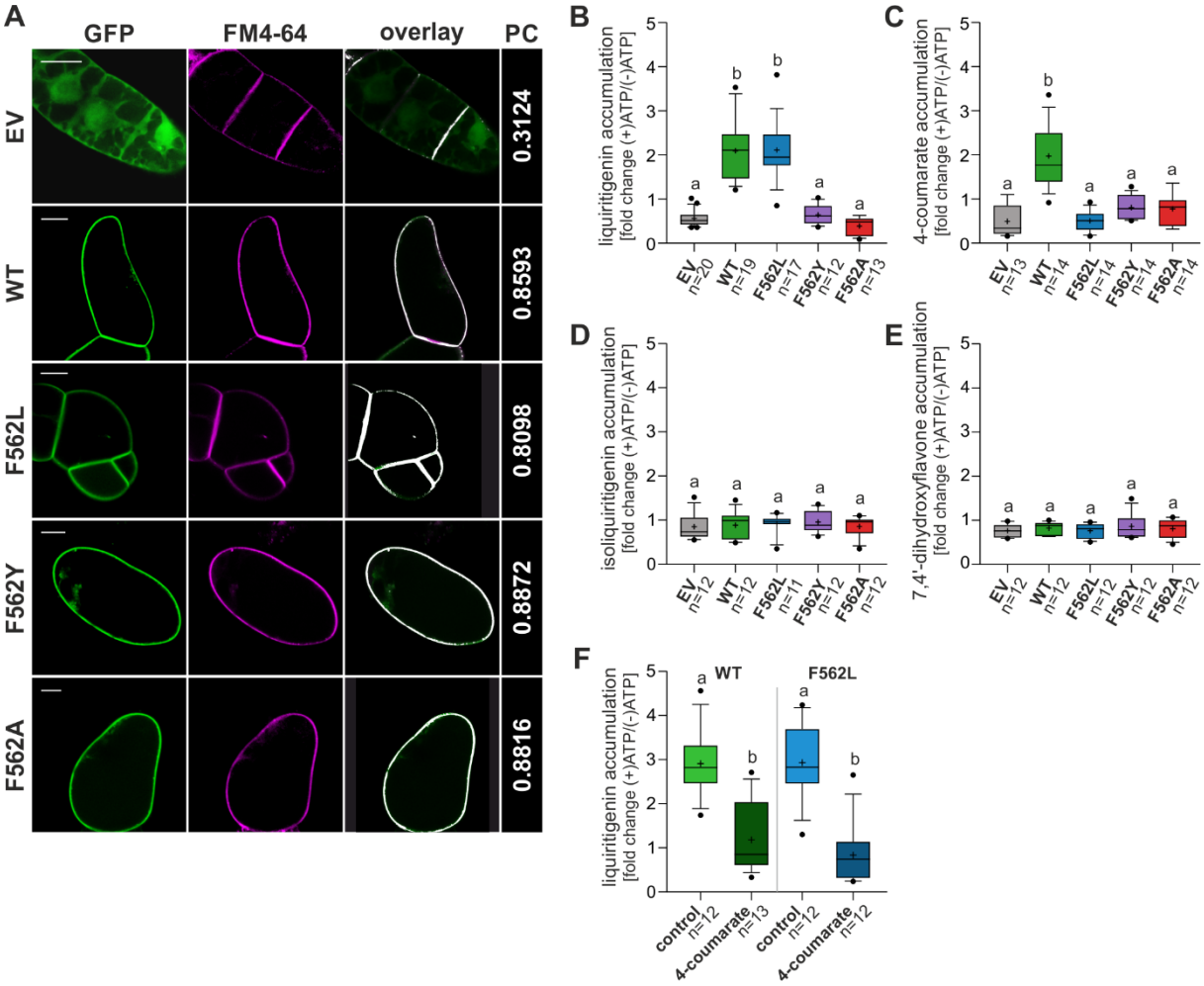

Figure 3. Transport assays with microsomes derived from BY2 suspension cell cultures. (A) Plasma membrane localization of the MtABCG46 variants tagged with GFP: WT, F562L, F562A, and F562Y in BY2 suspension cell cultures. Images with GFP and FM4-64 fluorescence pseudo-coloured in green and magenta, respectively. PC - Pearson's correlation coefficients for the colocalization of GFP and FM4-64 in the plasma membrane, visualized as an overlay. Scale bars, 20 µm. **(B-E)** 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 and 4-coumarate in microsomes expressing WT and F562L MtABCG46: means ±SD from 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 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-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-E)** or Mann-Whitney test **(F)** can be found in Supplemental Statistical Data.

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. Liquiritigenin and 4-coumarate accumulated in vesicles from lines possessing WT MtABCG46 (Figure 3B and 3C), but not isoliquiritigenin or 7,4'-dihydroxyflavone (Figure 3D and 3E). These results are consistent with previous observations (Biała et al., 2017). Further experiments with lines expressing variants of MtABCG46 revealed that F562Y and F562A substitutions abolished transport of liquiritigenin and 4-coumarate (Figure 3B and 3C). None of the variants were able to translocate isoliquiritigenin nor 7,4'dihydroxyflavone (Figure 3D and 3E). Interestingly, the F562L variant, accumulated liquiritigenin similarly to the WT MtABCG46, however, it was not able to transport 4-coumarate (Figure 3B and 3C).

Previous research has shown that liquiritigenin is a competitor of 4-coumarate in MtABCG46-mediated transport (Biała et al., 2017). 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 (Figure 3F). Experiments also revealed that although F562L cannot transport 4-coumarate, the latter is still a competitor for liquiritigenin (Figure 3F).

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 stability or flexibility between the mutant and WT proteins (Supplemental Figures 9-12). However, there were considerable differences in the availability of open access paths leading to the central cavity, which were lower in all mutants than in the WT (Supplemental Table 2 and Supplemental Figure 13A). 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 (Supplemental Figure 13B), mainly because they had more constricted entrances (Figure 4A and Supplemental Figure 14), presumably hampering access of bulky molecules to the central cavity. Accordingly, we hypothesized that mutation of residue F562 affects the structural arrangement of residues forming the access path, thereby significantly reducing its overall transport capability.

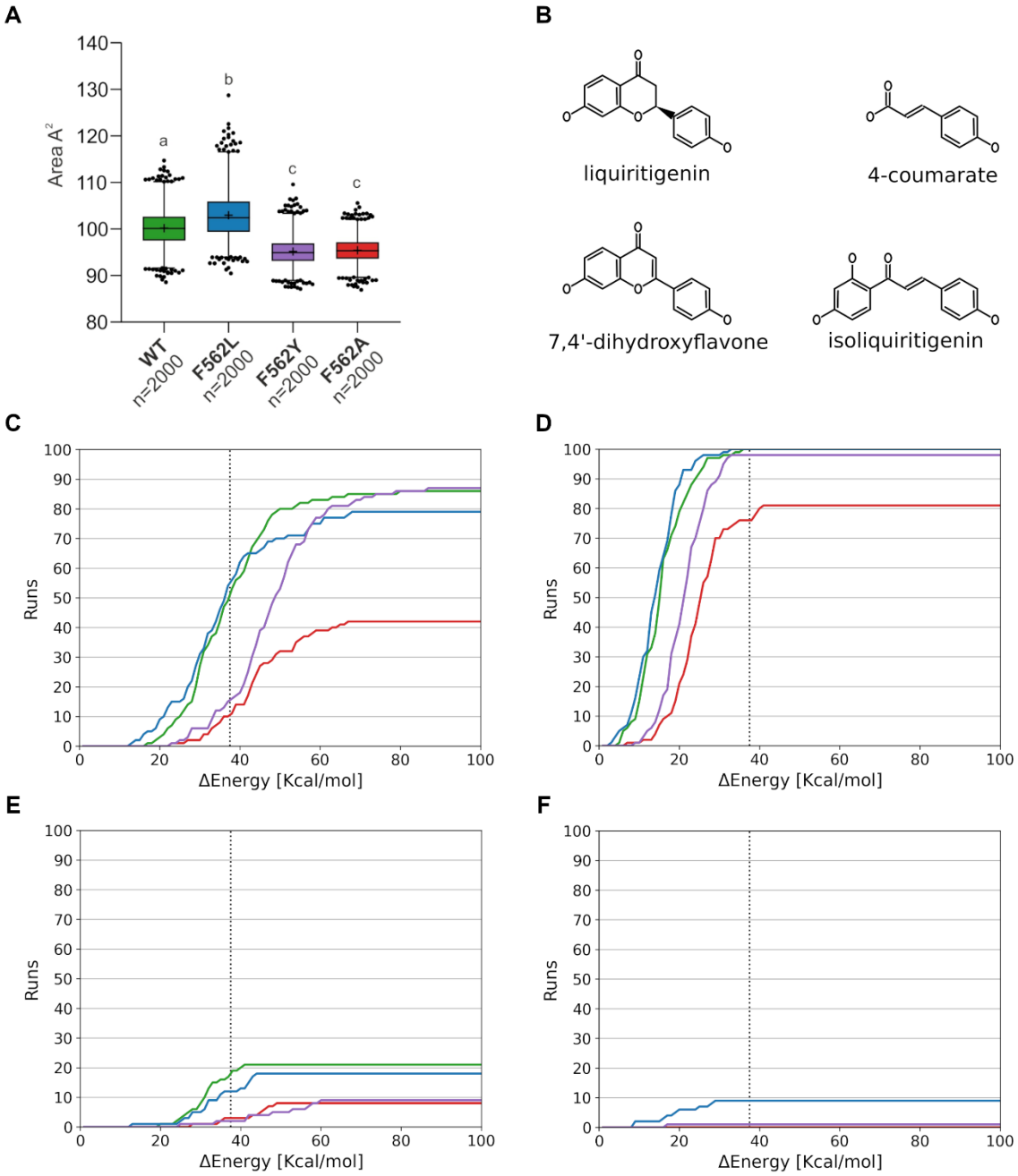

Figure 4. Effects of F562 mutations on the accessibility of the path to the central cavity of MtABCG46. (A) Area of the entrance to the access path in MtABCG46. For the definition of the area, see Supplemental Figure 14. In each box-and-whiskers plot: the center black line indicates the median; '+' indicates the mean; the box extends from the 25th to 75th percentile; the whiskers extend from the 1st to the 99th percentile, points below and above the whiskers are marked as individual dots. Different lowercase letters indicate significant differences, $P < 0.0001$. P values, determined by the Kruskal–Wallis test with a post hoc Dunn's multiple comparison test can be found in Supplemental Statistical Data. **(B)** Structures of ligands evaluated with CaverDock. **(C-F)** Cumulative distributions of energetic barriers for transport 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, F562L, F562Y, and F562A variants are marked in green, blue, purple, and red, respectively, and the median value for WT-liquiritigenin transport is indicated by a black dashed vertical line.

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 transported by MtABCG46 (Figure 4B), from the intracellular environment into the central cavity using CaverDock (Vavra et al., 2019). 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 (Supplemental Figure 15). Such an approach has proven utility in detecting hotspot residues for protein engineering (Papadopoulou et al., 2021), and correlating energetic barriers with biochemical rates (Marques et al., 2019; Brodsky et al., 2020). The cumulative distribution of the energy barriers for all variants showed that liquiritigenin could reach the central cavity equally well in WT and F562L mutant (with a median energy barrier of 38.5 $kcal \cdot mol^{-1}$) (Figure 4C). Frequencies of successful migration events in F562Y and F562A were much lower: 16% and 11% at the median WT energy barrier, respectively (Figure 4C). Moreover, to reach the same frequency of successful access to the central cavity, liquiritigenin had to overcome energy barriers of up to 50 $kcal \cdot mol^{-1}$ in F562Y and this frequency could not be reached at all in F562A (Figure 4C). In contrast, 4-coumarate reached the central cavity in almost all simulated migration events within this threshold in all variants (Figure 4D). Such surprising discrepancy with overall transport assays of this substrate indicates that mutations at F562 also affect transport of the substrates after they have reached the central cavity. Conversely, isoliquiritigenin and 7,4'dihydroxyflavone rarely reached the cavity in any variant (Figure 4E and 4F).

Structurally, the introduced modifications of residue 562 disrupted non-covalent contacts (Supplemental Figure 16), changing the bending and twisting angles of the TMD helices forming the access path (Supplemental Figures 17 and 18). In WT, residue F562 maintained a parallel displaced π-stacking interaction with residue F684 during the simulations (Supplemental Figure 16A), a feature shared with the F562Y mutant (Supplemental Figure 16C). Loss of this non-covalent contact in F562A resulted in a considerable displacement of TMD helix 5 (Supplemental Figure 16D). F562L also lost the π-stacking interaction, but maintained stability of this helix (Supplemental Figure 16B), 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 (Supplemental Figure17C and 18C). Furthermore, the bending of TMD helices affected not only the structure of the central cavity but also the path entrance, narrowing the available space for molecules to enter (Figure 4A). This reduction of available space at the entrance to the access path was observed in F562A and F562Y mutants, although no TMD helix deformations were observed in the F562Y mutant. 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 (Supplemental Figure 18C). Moreover, residue Y1213 formed an H-bond with N1331, a highly conserved residue (Figure 2A), in all performed MD simulations. Hence, disturbance of the normal behavior of these residues could influence ligands' binding in the central cavity.

DISCUSSION

AlphaFold2 has recently proposed suitability for modelling ABC proteins, with the ability to provide similar levels of accuracy as for soluble proteins (Hegedűs et al., 2022). The MtABCG46 structure we obtained using AlphaFold2 has the typical architecture of full-size ABCG transporters described in the literature (Thomas and Tampé, 2020). Moreover, our MD simulations identified transiently formed access paths to the central cavity from the intracellular region (Figure 1 and Supplemental Table 1). This cavity and its access paths were localized in a similar region to the open cavities observed in HsABCG1 (Sun et al., 2021), HsABCG2 (Orlando and Liao, 2020), and a recently obtained cryo-EM structure, ScPDR5 (Harris et al., 2021) (Supplemental Figure 19). Several substrates (such as rhodamine 6G, cholesterol, and mitoxantrone) have been found to occupy equivalent regions of ScPDR5, HsABCG1, and HsABCG2, respectively (Orlando and Liao, 2020; Sun et al., 2021; Harris et al., 2021), indicating the potential importance of this access path and central cavity of MtABCG46 for ligand transport.

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 the four MtABCG46 variants showed that none of them can transport isoliquiritiginen or 7,4'dihydroxyflavone (Figure 3D and 3E), in accordance with indications from CaverDock analyses that likelihoods of these compounds accessing the central cavity are extremely low (Figure 4E and 4F). We found no significant

difference in the energy barrier for liquiritigenin's entry into the central cavities of WT and F562L MtABCG46, which is sufficiently low for successful transport of this ligand by both variants. In contrast, liquiritigenin could not enter the cavity of F562A or F562Y, probably due to a collapse of the access path caused by their mutations (Figure 4A), which consequently lack ability to transport it (Figure 4C and 3B). Thus, our data indicate that restriction of access to the central cavity is a major contributor to substrate selectivity in early stages of the transport process.

Curiously, we found that only MtABCG46 WT could transport 4-coumarate (Figure 3C). CaverDock analysis detected no difference in accessibility of the central cavities of MtABCG46 WT and F562L for this metabolite, and only limited disadvantage in the other two mutants (Figure 4D). Nevertheless, 4-coumarate showed similar competitive behavior towards liquiritigenin in F562L to that observed in WT (Figure 3F). This suggests that 4-coumarate may still be able to reach the central cavity, and the transport failure may be linked to perturbation in subsequent stages of the transport cycle, *e.g.*, recognition of the ligand bound in the central cavity or its stabilization during the subsequent conformational change (from IF to OF) of the transporter. To explore this possibility, we investigated the availability of polar interactions provided by residues of the central cavity to 4-coumarate. 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 (Supplemental Figure 20) perturbing the putative interaction between N1331 and 4-coumarate, and possibly hampering transport of this ligand.

The phylogenetic analyses of various plant full-size ABCGs revealed family-specific ABCG expansion, which could be associated with chemodiversity reflected in various types of specialized metabolites in modern angiosperms. This hypothesis is supported not only by the phylogenetic analysis, but also by transport experiments and functional analyses of ABCGs, as previously reviewed (Do et al., 2021; Banasiak and Jasiński, 2022). Nevertheless, the molecular determinants of this specialization and sequence changes resulting from protein adaptation to various biochemical contexts remain obscure. Our experimental data, together with phylogenetic observations, suggest that adaptation of ABCG proteins associated with

evolutionary pressures in plants has resulted, *inter alia*, in variability of some key residues, such as F562 (Figure 2E). Such variability is essential for transporters to fulfill their functional roles in diverse, complex chemical scenarios. MtABCG46 is responsible for the membrane translocation of 4-coumarate and liquiritigenin, molecules at key branching points in phenylpropanoid biosynthesis (Biała et al., 2017). Thus, such functions may include selective precursor distribution (in addition to ABCG proteins' widely recognized role in secretion of end-products of specialized metabolism), which may be valuable for agriculture and biotechnological applications. Designing ABC-targeted research strategies for particular lineages, metabolite profiles or processes is of interest (Gani et al., 2021; Li et al., 2022). It allows e.g. plant breeding and genetic engineering to generate cultivars or transgenic plants accumulating higher levels of the desired compounds, but also the development of scalable plant cell or root cultures to biosynthesize and/or secrete desired product. Phenylpropanoids are a large class of specialized plant metabolites with many important roles in plant biology, medical applications and industrial uses (Neelam et al., 2020). Elucidation of the transport mechanisms involved in targeted metabolite distribution will facilitate metabolic engineering of these high-value natural products. We foresee sequence-structure-dynamics exploration fueled by AplhaFold2 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.

METHODS

Plant material

Nicotiana tabacum Bright Yellow 2 (BY2) suspension cell cultures (Nagata et al., 1992) were grown in Murashige and Skoog medium supplemented with 2.72 mM KH_2PO_4, 0.56 mM myoinositol, 3 μM thiamine, 0.9 μM 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 (Curtis and Grossniklaus, 2003). Particular mutants were prepared from a p35S::GFP-MtABCG46 construct 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 (Hellens et al., 2000) carrying the pMDC43 vector containing a particular variant of the MtABCG46 sequence or the empty pMDC43 vector, as previously described (Biała et al., 2017).

Confocal microscopy

5-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 FM4-46 (ThermoFisher Scientific). Obtained pictures were analyzed using Leica LAS AF software. Fluorescent signals from GFP and FM4-64 were pseudo-coloured in green and magenta, respectively.

Preparation of plasma membrane vesicles

Microsomal fractions were isolated from 12 g portions of BY2 suspension cell cultures as previously described (Jasiński et al., 2001). Plasma membrane fractions of the microsomal isolates were enriched by partitioning in an aqueous two-phase system, also as previously described (Larsson et al., 1987). The quality of obtained microsomes was tested with 9-amino-6-chloro-2-methoxyacridine (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 mm pore-size; Millipore). The transport assays were performed with microsomes corresponding to 520 $ng \cdot \mu l^{-1}$ protein concentration mixed with transport buffer (10 mM Tris-HCl, 10 mM EDTA, 10% sucrose, pH 5.0), the selected phenolic (750 μM), 100 $\mu g \cdot mL^{-1}$ creatine kinase, 10 mM creatine-phosphate and 1 mM of $MgCl_2$, in the presence and absence of 4 mM ATP. After 3 minutes 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 then incubated in 80% MetOH 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 minutes in 13200 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 μM, 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 1KP transcript data to tBLASTn searches using MtABCG46 as a query sequence with an E-value cutoff of 1e-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. 1839 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 (Edgar, 2004) in MEGA X (Kumar et al., 2018). 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 (Celniker et al., 2013; Ashkenazy et al., 2016). 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 (Kamisetty et al., 2013; Ovchinnikov et al., 2014) 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 (Jumper et al., 2021). The resulting models were evaluated using the per-

residue lDDT-Cα metric (Mariani et al., 2013) and results from PROCHECK software (Laskowski et al., 1993).

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 (Zhang and Skolnick, 2004), 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 (Trott and Olson, 2010) 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 (Case et al., 2020)**,** 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 (Anandakrishnan et al., 2012) 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 (Jo et al., 2008; Brooks et al., 2009; Wu et al., 2014; Lee et al., 2016; Lee et al., 2020). 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. The water model employed was OPC (Izadi et al., 2014), and the HMR (Hopkins et al., 2015) method was used to enable 4 fs simulation timesteps. The MD engine employed was Amber20 (Case et al., 2020) with the pmemd GPU implementation (le Grand et al., 2013), the ff19SB force field (Tian et al., 2020) was used for the protein, lipid17 for the POPC membrane, and previously presented parameters for the ATP molecules (Meagher et al., 2003). 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 (Supplemental Table 3). After the minimization, a series of short equilibration steps was applied to increase temperature and release restraints in the system, each consisting of 1 ns of simulation time. After the equilibration stage, an extra 100 ns equilibration step was applied in which restraints were only maintained for the Mg^{2+} and ATP molecules. Finally, 100 ns unrestrained MD to fully equilibrate the system was followed by a production phase of 400 ns unrestrained MD. For every variant considered, five replicas of the production phase were simulated.

Access path detection, classification, and selection

For path computation and detection, CAVER v3.0 software (Chovancova et al., 2012) was employed with a 0.9 Å probe radius, 6 Å shell radius, and 4 Å 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 (Brezovsky et al., 2022) was used to obtain a comprehensive and comparative view of the path network across all variants. In TransportTools, the clustering method was set to complete with a clustering cutoff of 0.5 Å for the analysis of the wild-type protein, then switched to the average method with a clustering cutoff of 2.0 Å for the comparison of all variants simultaneously, with all other parameters 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 (Vavra et al., 2019). 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 (Morris et al., 2009) 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 and eight parallel workers per combination of variant-path-ligand. Finally, the results were analyzed with in-house Python scripts.

Auxiliary analyses of data from MD simulations

Using the cpptraj (Roe and Cheatham, 2013) 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 (Supplemental Figure 3 and Supplemental Table 4). 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, α helices 2, 5, 8, and 11 were analyzed with the HELANAL module of MDAnalysis v2.0.0 (Michaud-Agrawal et al., 2011; Gowers et al., 2016).

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 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 Supplemental Statistical Data.

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.

FUNDING

This work was supported by the National Science Centre, Poland (grant nos. 2017/27/B/NZ1/01090 and 2017/25/B/NZ1/01307). K.P. and C.S.B. are recipients of

scholarships associated with the POWER project (ref. nos. POWR.03.02.00-00-1032/16 and POWR.03.02.00-00-I022/16, respectively).

CONTRIBUTIONS

M.J. devised and supervised the project. J.Br. designed, interpreted and supervised the computational experiments. K.P. and M.G. selected residues for the site-directed mutagenesis (conservation and coevolution analyses). C.S.B. performed, analyzed, and interpreted results of computational experiments. W.B.L. and K.P. designed, performed and interpreted results of the transport experiments (BY2 transformation, isolation of microsomes, transport and competition assays). A.P. contributed to BY2 transformation, isolation of microsomes, transport and competition assays. W.B.L was responsible for HPLC/MS analysis and statistical analysis. J.Ba., A.B., M.R. ABCG selected sequences acquired in the 1KP project. J.Ba. and W.B.L., performed microscopic observation. C.S.B., W.B.L. and K.P. prepared figures. K.P., C.S.B, W.B.L., J.Br., and M.J. wrote the manuscript with the help of the co-authors. All authors discussed and approved the manuscript.

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.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

Anandakrishnan, R., Aguilar, B., and Onufriev, A. v. (2012). H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**:W537–W541.

Aryal, B., Huynh, J., Schneuwly, J., Siffert, A., Liu, J., Alejandro, S., Ludwig-Müller, J., Martinoia, E., and 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* **10**.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**:W344–W350.

Banasiak, J., and Jasiński, M. (2022). ATP-binding cassette transporters in nonmodel plants. *New Phytologist* **233**:1597–1612.

Banasiak, J., Biała, W., Staszków, A., Swarcewicz, B., Kępczyńska, E., Figlerowicz, M., and Jasiński, M. (2013). A Medicago truncatula ABC transporter belonging to subfamily G modulates the level of isoflavonoids. *J Exp Bot* **64**:1005–1015.

Biała, W., Banasiak, J., Jarzyniak, K., Pawela, A., and Jasiński, M. (2017). Medicago truncatula ABCG10 is a transporter of 4-coumarate and liquiritigenin in the medicarpin biosynthetic pathway. *J Exp Bot* **68**:3231–3241.

Brezovsky, J., Thirunavukarasu, A. S., Surpeta, B., Sequeiros-Borja, C. E., Mandal, N., Sarkar, D. K., Dongmo Foumthuim, C. J., and Agrawal, N. (2022). TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them. *Bioinformatics* **38**:1752–1753.

Brodsky, K., Kutý, M., Pelantová, H., Cvačka, J., Rebroš, M., Kotik, M., Smatanová, I. K., Křen, V., and Bojarová, P. (2020). Dual substrate specificity of the rutinosidase from aspergillus niger and the role of its substrate tunnel. *Int J Mol Sci* **21**:1–21.

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: The biomolecular simulation program. *J Comput Chem* **30**:1545–1614.

Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Jimenez Vieira, F. R., Bowler, C., Dorrell, R. G., Gitzendanner, M. A., et al. (2019). Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *Gigascience* **8**.

Case, D. A., Aktulga, H. M., Belfon, K., Ben-Shalom, I. Y., Berryman, J. T., Brozell, S. R., Cerutti, D. S., Cheatham III, T. E., Cisneros, G. A., Cruzeiro, V. W. D., et al. (2020). AMBER 20 Advance Access published 2020.

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2013). ConSurf: Using evolutionary data to raise testable hypotheses about protein function. *Isr J Chem* **53**:199–206.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., et al. (2012). CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol* **8**:e1002708.

Cunningham, B. C., and Wells, J. A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science (1979)* **244**:1081–1085.

Curtis, M. D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**:462–469.

Do, T. H. T., Martinoia, E., and Lee, Y. (2018). Functions of ABC transporters in plant growth and development. *Curr Opin Plant Biol* **41**:32–38.

Do, T. H. T., Martinoia, E., Lee, Y., and Hwang, J.-U. (2021). 2021 update on ATP-binding cassette (ABC) transporters: how they meet the needs of plants. *Plant Physiol* **187**:1876–1892.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797.

Ford, R. C., and Beis, K. (2019). Learning the ABCs one at a time: Structure and mechanism of ABC transporters. *Biochem Soc Trans* **47**:23–36.

Fourcroy, P., Sisó-Terraza, P., Sudre, D., Savirón, M., Reyt, G., Gaymard, F., Abadía, A., Abadia, J., Álvarez-Fernández, A., and Briat, J. F. (2014). Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. *New Phytologist* **201**:155–167.

Gani, U., Vishwakarma, R. A., and Misra, P. (2021). Membrane transporters: the key drivers of transport of secondary metabolites in plants. *Plant Cell Rep* **40**.

Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., Seyler, S., Domański, J., Dotson, D., Buchoux, S., Kenney, I., et al. (2016). MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. In *Proceedings of the 15th Python in Science Conference*, pp. 98–105. SciPy.

Harris, A., Wagner, M., Du, D., Raschka, S., Nentwig, L. M., Gohlke, H., Smits, S. H. J., Luisi, B. F., and Schmitt, L. (2021). Structure and efflux mechanism of the yeast pleiotropic drug resistance transporter Pdr5. *Nat Commun* **12**:5254.

He, Y., Xu, J., Wang, X., He, X., Wang, Y., Zhou, J., Zhang, S., and Meng, X. (2019). The arabidopsis pleiotropic drug resistance transporters PEN3 and PDR12 mediate camalexin secretion for resistance to botrytis cinerea. *Plant Cell* **31**:2206–2222.

Hegedűs, T., Geisler, M., Lukács, G. L., and Farkas, B. (2022). Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cellular and Molecular Life Sciences* **79**:73.

Hellens, R., Mullineaux, P., and Klee, H. (2000). A guide to Agrobacterium binary Ti vectors. *Trends Plant Sci* **5**:446–451.

Hopkins, C. W., le Grand, S., Walker, R. C., and Roitberg, A. E. (2015). Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* **11**:1864–1874.

Hwang, J. U., Song, W. Y., Hong, D., Ko, D., Yamaoka, Y., Jang, S., Yim, S., Lee, E., Khare, D., Kim, K., et al. (2016). Plant ABC transporters enable many unique aspects of a terrestrial plant's lifestyle. *Mol Plant* **9**:338–355.

Ito, H., and Gray, W. M. (2006). A gain-of-function mutation in the Arabidopsis pleiotropic drug resistance transporter PDR9 confers resistance to auxinic herbicides. *Plant Physiol* **142**:63–74.

Izadi, S., Anandakrishnan, R., and Onufriev, A. v. (2014). Building Water Models: A Different Approach. *J Phys Chem Lett* **5**:3863–3871.

Jackson, S. M., Manolaridis, I., Kowal, J., Zechner, M., Taylor, N. M. I., Bause, M., Bauer, S., Bartholomaeus, R., Bernhardt, G., Koenig, B., et al. (2018). Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol* **25**:333–340.

Jasiński, M., Stukkens, Y., Degand, H., Purnelle, B., Marchand-Brynaert, J., and Boutry, M. (2001). A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion. *Plant Cell* **13**:1095–107.

Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **29**:1859–1865.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* **596**:583–589.

Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-based residue–residue contact predictions in a sequence- and structure-rich era. *Proc Natl Acad Sci U S A* **110**:15674–15679.

Kang, J., Hwang, J. U., Lee, M., Kim, Y. Y., Assmann, S. M., Martinoia, E., and Lee, Y. (2010). PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc Natl Acad Sci U S A* **107**:2355–2360.

Khunweeraphong, N., and 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**:4806.

Kim, D. Y., Bovet, L., Maeshima, M., Martinoia, E., and Lee, Y. (2007). The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant Journal* **50**:207–218.

Kowal, J., Ni, D., Jackson, S. M., Manolaridis, I., Stahlberg, H., and Locher, K. P. (2021). Structural basis of drug recognition by the multidrug transporter ABCG2. *J Mol Biol* **433**:166980.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**:1547–1549.

Larsson, C., Widell, S., and Kjellbom, P. (1987). [52] Preparation of high-purity plasma membranes. In *Methods in enzymology*, pp. 558–568.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., and IUCr (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *urn:issn:0021-8898* **26**:283–291.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., et al. (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**:405–413.

Lee, J., Hitzenberger, M., Rieger, M., Kern, N. R., Zacharias, M., and Im, W. (2020). CHARMM-GUI supports the Amber force fields. *Journal of Chemical Physics* **153**:035103.

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**:679–685.

Lefèvre, F., and Boutry, M. (2018). Towards identification of the substrates of ATP-binding cassette transporters. *Plant Physiol* **178**:18–39.

le Grand, S., Götz, A. W., and Walker, R. C. (2013). SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun* **184**:374–380.

Li, H., Matsuda, H., Tsuboyama, A., Munakata, R., Sugiyama, A., and Yazaki, K. (2022). Inventory of ATP-binding cassette proteins in Lithospermum erythrorhizon as a model plant producing divergent secondary metabolites. *DNA Research* **29**:1–12.

Lu, X., Dittgen, J., Pislewska-Bednarek, M., Molina, A., Schneider, B., Svatos, A., Doubsky, J., Schneeberger, K., Weigel, D., Bednarek, P., et al. (2015). Mutant allele-specific uncoupling of penetration3 functions reveals engagement of the ATP-binding cassette transporter in distinct tryptophan metabolic pathways. *Plant Physiol* **168**:814–827.

Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**:2722–2728.

Marques, S. M., Bednar, D., and Damborsky, J. (2019). Computational study of protein-ligand unbinding for enzyme engineering. *Front Chem* **7**:650.

Matern, A., Böttcher, C., Eschen-Lippold, L., Westermann, B., Smolka, U., Döll, S., Trempel, F., Aryal, B., Scheel, D., Geisler, M., et al. (2019). A substrate of the ABC transporter PEN3 stimulates bacterial flagellin (flg22)-induced callose deposition in Arabidopsis thaliana. *Journal of Biological Chemistry* **294**:6857–6870.

Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003). Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **24**:1016–1025.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem* **32**:2319–2327.

Morris, G. M., Ruth, H., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**:2785–2791.

Nagata, T., Nemoto, Y., and Hasezawa, S. (1992). Tobacco BY-2 cell line as the "HeLa" cell in the cell biology of higher plants. In *International Review of Cytology*, pp. 1–30.

Neelam, Khatkar, A., and Sharma, K. K. (2020). Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. *Crit Rev Food Sci Nutr* **60**:2655–2675.

Orlando, B. J., and Liao, M. (2020). ABCG2 transports anticancer drugs via a closed-to-open switch. *Nature Communications 2020 11:1* **11**:1–11.

Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **2014**.

Papadopoulou, A., Meierhofer, J., Meyer, F., Hayashi, T., Schneider, S., Sager, E., and Buller, R. (2021). Re-programming and optimization of a L-proline cis-4-hydroxylase for the cis-3-halogenation of its native substrate. *ChemCatChem* **13**:3914–3919.

Prasad, R., and Goffeau, A. (2012). Yeast ATP-binding cassette transporters conferring multidrug resistance. *Annu Rev Microbiol* **66**:39–63.

Rees, D. C., Johnson, E., and Lewinson, O. (2009). ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**:218–227.

Roe, D. R., and Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9**:3084–3095.

Skarda, L., Kowal, J., and Locher, K. P. (2021). Structure of the human cholesterol transporter ABCG1. *J Mol Biol* **433**:167218.

Stein, M., Dittgen, J., Sánchez-Rodríguez, C., Hou, B.-H., Molina, A., Schulze-Lefert, P., Lipka, V., and 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**:731–746.

Strader, L. C., and Bartel, B. (2009). The arabidopsis PLEIOTROPIC DRUG RESISTANCE8/ABCG36 ATP binding cassette transporter modulates sensitivity to the auxin precursor lndole-3-butyric acid. *Plant Cell* **21**:1992–2007.

Sun, Y., Wang, J., Long, T., Qi, X., Donnelly, L., Elghobashi-Meinhardt, N., Esparza, L., Cohen, J. C., Xie, X. S., Hobbs, H. H., et al. (2021). Molecular basis of cholesterol efflux via ABCG subfamily transporters. *Proc Natl Acad Sci U S A* **118**.

Takeuchi, M., Watanabe, A., Tamura, M., and Tsutsumi, Y. (2018). The gene expression analysis of Arabidopsis thaliana ABC transporters by real-time PCR for screening monolignol-transporter candidates. *Journal of Wood Science* **64**:477–484.

Thomas, C., and Tampé, R. (2020). Structural and mechanistic principles of ABC transporters. *Annu Rev Biochem* **89**:605–636.

Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., et al. (2020). Ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput* **16**:528–552.

Toussaint, F., Pierman, B., Bertin, A., Lévy, D., and Boutry, M. (2017). Purification and biochemical characterization of NpABCG5/NpPDR5, a plant pleiotropic drug resistance

transporter expressed in Nicotiana tabacum BY-2 suspension cells. *Biochemical Journal* **474**:1689–1703.

Trott, O., and Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**:455–461.

Vavra, O., Filipovic, J., Plhak, J., Bednar, D., Marques, S. M., Brezovsky, J., Stourac, J., Matyska, L., and Damborsky, J. (2019). CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels. *Bioinformatics* **35**:4986–4993.

Verrier, P. J., Bird, D., Burla, B., Dassa, E., Forestier, C., Geisler, M., Klein, M., Kolukisaoglu, Ü., Lee, Y., Martinoia, E., et al. (2008). Plant ABC proteins - a unified nomenclature and updated inventory. *Trends Plant Sci* **13**:151–159.

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., et al. (2014). CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J Comput Chem* **35**:1997–2004.

Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics* **57**:702–710.

Ziegler, J., Schmidt, S., Strehmel, N., Scheel, D., and Abel, S. (2017). Arabidopsis transporter ABCG37/PDR9 contributes primarily highly oxygenated coumarins to root exudation. *Sci Rep* **7**:3704.

SUPPLEMENTAL INFORMATION

Supplemental Figures and Tables (PDF) include details on the MtABCG46 model; the localization of access pathways leading into central cavity and descrition of their characteristics; structural and sequential alignment of MtABCG46 and HsABCG2; conservation score of the inner cavity forming residues; RMSD and RMSF of MtABCG46 variants observed in simulations; structural coupling between access path and transmebrane helices; interactions between key residues of transmebrane helices; the description of perturbation of transmebrane helices by mutagenesis; comparison of internal cavities among related ABC transporters; the description of confromational space adopted by N1331 in MtABCG46 variants observed in simulations; duration and restraints applied in individual simulation stages. Additionally, the results of statistical analyses are summarized in the Supplemental Statistical Data (XLSX).