Title: Mutational landscape and molecular bases of echinocandin resistance

Authors: Romain Durand[1,2,3,4,5], Alexandre G. Torbey[3,6], Mathieu Giguere[1,2,3,4], Alicia Pageau[1,2,3,4,5], Alexandre K. Dubé[1,2,3,4,5], Patrick Lagüe[1,2,3*] and Christian R. Landry[1,2,3,4,5*]

Affiliations:

[1]Département de Biochimie, de Microbiologie et de Bio-informatique, Faculté des Sciences et de Génie, Université Laval; Québec, G1V 0A6, Canada.

[2]Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval; Québec, G1V 0A6, Canada.

[3]PROTEO, Le regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, Université Laval; Québec, G1V 0A6, Canada.

[4]Centre de Recherche sur les Données Massives (CRDM), Université Laval; Québec, G1V 0A6, Canada.

[5]Département de Biologie, Faculté des Sciences et de Génie, Université Laval; Québec, G1V 0A6, Canada.

[6]Institut National de la Recherche Scientifique, Centre Armand-Frappier; Laval, H7V 1B7, Canada.

*Corresponding authors. Emails: patrick.lague@bcm.ulaval.ca or christian.landry@bio.ulaval.ca

Abstract

One of the front-line drug classes used to treat invasive fungal infections is echinocandins, which target the fungal-specific beta-glucan synthase (Fks). Treatment failure due to resistance often coincides with mutations in three protein regions defined as hotspots. The biophysical bases by which such mutations confer resistance and cross-resistance among echinocandins are largely unknown. Here, we combine molecular docking and molecular dynamics simulations to construct and refine a model of echinocandin binding to Fks1. This structural framework is integrated with deep-mutational scanning to comprehensively assess the impact of mutations across the three hotspots in the model yeast *Saccharomyces cerevisiae*. We detail the positioning and binding constraints of the three most widely used echinocandins; anidulafungin, caspofungin and micafungin and elucidate several key molecular bases of resistance. Our findings will enable DNA sequence-based predictions of resistance to this important drug family and the improvement of future molecules that could overcome current resistance mutations.

One sentence summary: Disruption of specific interactions between echinocandins and key residues from their target lead to drug-specific resistance.

Introduction

Echinocandins are among the most important compounds to combat human fungal infections. They are routinely used to treat invasive candidiasis and are often the only recourse against the emerging pathogen *Candidozyma auris* [1,2]. Unfortunately, their efficacy is hindered by the acquisition of resistance, which occurs mostly through mutations in the drug target: the β-1,3-glucan synthase Fks1 [3]. Mutations reported to correlate with treatment failure are concentrated in three regions of the coding sequence, which have hence been referred to as “hotspots” [4–6]. Despite the importance of echinocandins and the growing threat of resistant pathogens, little is known regarding the exact molecular mechanisms at play.

Echinocandins act as non-competitive inhibitors of the enzyme, although their mechanism of action is still not entirely clear [7,8]. The correlation between treatment failure and amino acid changes in the hotspots has initially led to the hypothesis that echinocandins bind these regions [9]. Since then, many studies have shown that the presence of amino acid substitutions in the hotspots correlates with echinocandin resistance [10]. In a pivotal study, Johnson et al. [11] significantly improved our understanding of the mechanisms underlying echinocandin resistance. They proposed that all three hotspots are adjacent to each other and are located close to, or embedded in the outer membrane. Additionally, they suggested that echinocandins are likely to bind directly to hotspot 1 and 2, through their peptidomimetic macrocycle, and to hotspot 3 through their lipophilic tail [11]. Until very recently, this model had not been supported with biophysical data, including from computational experiments. In the past two years, the structures of Fks1 from the budding yeast *Saccharomyces cerevisiae*, alone and in complex with the regulatory GTPase Rho1, have been solved by three independent groups by cryo-electron microscopy [12–14]. The structures enabled confirmation, through molecular docking, that binding of echinocandins in the spatially close hotspots is favorable [15]. If it is indeed the case, the secretion of the β-1,3-glucan polymer, as speculated from the structure of the Fks1-Rho1 complex, could potentially be inhibited by steric hindrance caused by echinocandin binding [7,14]. Although these recent advances have been insufficient to characterize the exact

mechanisms leading to resistance, there is little doubt that the missing piece of the puzzle lies in the hotspots.

One major concern with the apparition of resistance mutations in fungi is the extent of cross-resistance, especially given the limited number of antifungal drug classes and the small number of molecules available within each class [16]. If a mutation were to confer resistance to only a single echinocandin, diagnostics testing could guide clinicians to an alternative molecule within the same class, rather than abandoning echinocandins altogether. Notably, the W695C mutation (in hotspot 3) was shown to confer caspofungin-specific resistance [6]. This observation supports a model in which echinocandins bind directly to the hotspot region, with each molecule adopting a slightly different fit in the binding pocket.

In this study, we leverage the structure of Fks1 to improve our understanding of echinocandin resistance at the biophysical level. We first performed molecular docking using 3D free energy maps generated by Site Identification by Ligand Competitive Saturation (SILCS) [17] combined with molecular dynamics (MD) simulations to predict the binding modes of three clinically relevant echinocandins to Fks1: anidulafungin, caspofungin and micafungin. In parallel, we used deep-mutational scanning (DMS) to generate all possible single amino acid substitutions across the three known hotspots of Fks1. All mutants were screened against the three echinocandins. Our experimental data corroborate the binding predictions, together providing molecular insights into how specific interactions govern drug-specific binding, thereby explaining why certain mutations confer drug-specific resistance while others confer cross-resistance. Finally, we present an example of machine-learning assisted prediction of intrinsic resistance that makes use of the mutational data, based solely on hotspot sequences.

Results

Echinocandins bind to solvent-exposed Fks1 hotspots

Echinocandins are large, semi-synthetic lipopeptides composed of a hydrophilic macrocycle core derived from six amino acid residues, specifically modified forms of tyrosine, ornithine, proline, serine and threonine, linked to a lipophilic tail. Both the composition of the hexapeptide and the nature of the tail vary among different echinocandin molecules (S1 Fig). Anidulafungin lacks strong hydrogen bond donors or acceptors but features a rigid alkoxytriphenyl tail, which is critical for its insertion into the membrane [18]. Caspofungin contains two ornithine residues capable of participating in π-cation interactions and is characterized by a small, flexible tail (S1 Fig). Micafungin is distinct in having a sulfate group attached to its dihydroxy-homotyrosine residue and a tail that includes an isoxazole ring.

First, we performed 100-ns MD simulations to equilibrate the structure of Fks1. These simulations were followed by an additional 300 ns of simulation for each of three independently constructed systems, to allow sufficient sampling of water distribution within the hotspot region. In all cases, Fks1 was embedded in a lipid bilayer and solvated with explicit water molecules and neutralizing ions. Water density maps were generated from the final 100 ns of each trajectory. When combined, these maps revealed a water-filled pocket between the helices of hotspot 1 and hotspot 2, extending downward toward the base of the helix at hotspot 3 (residue D691) (Fig. 1A).

The observed membrane deformation and water pocket formation are driven by several polar residues within the hotspot cavity, including K632, E635, S636, D684 and D691 (HS3), as well as R1357 and S1361 (within and adjacent to HS2, respectively). These results indicate that the upper portion of the hotspot cavity is accessible to extracellular solvent, whereas the lower portion remains embedded within the lipid bilayer. This configuration suggests that drug binding likely occurs in a partially solvent-exposed environment, influencing drug positioning and interactions profiles. These findings are consistent with what was proposed by Johnson and Edlind [11]. It also underscores the importance of the echinocandins' hydrophobic tail in ensuring stable binding and proper orientation within the cavity. This tail is critical for inhibition, as first demonstrated with cilofungin; the first semi-synthetic echinocandin drug developed [19].

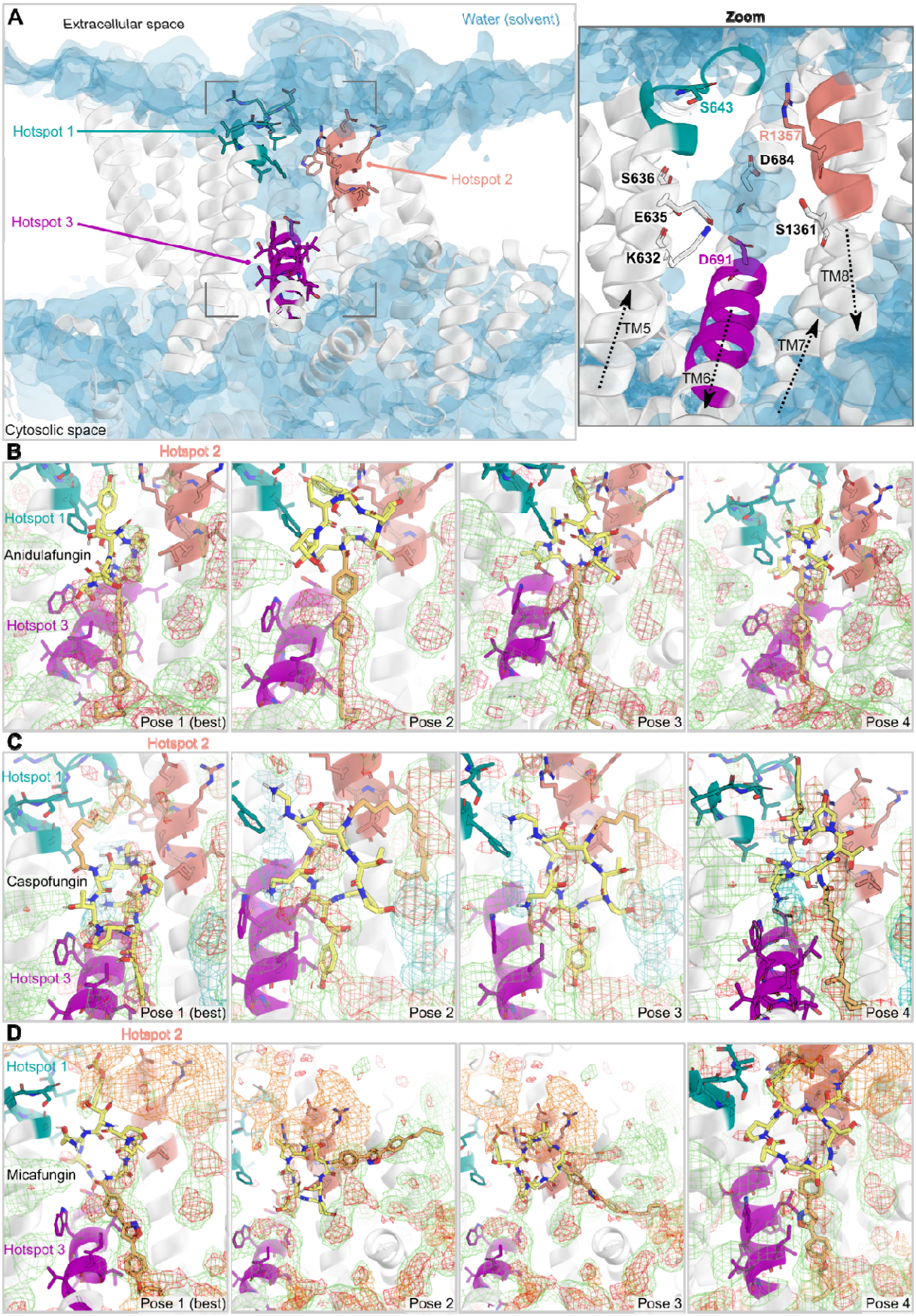

A
Extracellular space
Water (solvent)
Hotspot 1
Hotspot 2
Hotspot 3
Cytosolic space
Zoom
S643
R1357
D684
S636
E635
S1361
K632
D691
TM8
TM5
TM7
TM6
B
Hotspot 2
Hotspot 1
Anidulafungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4
C
Hotspot 2
Hotspot 1
Caspofungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4
D
Hotspot 2
Hotspot 1
Micafungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4

Fig. 1. Molecular bases of echinocandin binding. (**A**) Water occupancy map (blue surface, isocontour 0.2) derived from 300-ns MD simulations of Fks1 (white ribbon) in the absence of echinocandin reveals a water-filled pocket in the transmembrane region. This feature is located primarily between hotspot 1 (dark cyan) and hotspot 2 (salmon), extending to hotspot 3 (purple). The dark gray corners outline the region shown in the magnified view on the right, where only polar residues contributing to the formation of the water pocket are highlighted. Transmembrane helices TM5, TM6, TM7 and TM8 are shown, with arrow directions indicating the N- to C-terminal orientation of the protein sequence. (**B-D**) Docking poses, guided by SILCS FragMaps, are shown on the MD-equilibrated structure of Fks1 for anidulafungin (B, LGFE scores: -35.462 kcal/mol, -32.079 kcal/mol, -31.364 kcal/mol and -30.482 kcal/mol for poses 1-4, respectively), caspofungin (C, -24.295 kcal/mol, -23.239 kcal/mol, -22.512 kcal/mol and -16.021 kcal/mol) and micafungin (D, -31.821 kcal/mol, -31.464 kcal/mol, -29.445 kcal/mol and -26.669 kcal/mol). Atoms are colored as follows: polar hydrogen, white; carbon, yellow (drug macrocycles only) or orange (lipophilic tail of drugs only); nitrogen, blue; oxygen, red. In all panels, the following FragMaps are represented by colored meshes: generic apolar (green) and H-bond donor (red). The methylamine nitrogen FragMap (cyan) is shown in panel C, and the acetate carboxylate carbon FragMap (orange) is shown in panel D. Correspondences between FragMaps and drug chemical groups are provided in S1 Table. Figures were generated using PyMOL v2.5.5 [20].

Docking of the echinocandins in the hotspots

Due to their large size and intrinsic flexibility, combined with the even greater size and complexity of their receptor, inference of echinocandin binding modes using traditional molecular docking remains challenging. To address this and gain mechanistic insight into potential resistance pathways, molecular docking was instead performed using 3D free energy maps generated by SILCS [17].

In docking, multiple binding poses, i.e. predicted orientations and conformations of the ligand within the binding site, are generated and scored based on estimated binding affinity. To identify the most plausible poses, we first conducted SILCS-MD simulations to map potential binding pockets (S2 Fig), using a set of standard pharmacophoric fragments that dynamically competed for binding during the simulation. Each fragment features a distinct electrostatic profile, enabling the detection of atom-specific interaction sites. These sites are represented as volumetric probability maps, referred to here as FragMaps, which reflect the likelihood of favorable interactions at specific positions. The FragMaps correspond to key chemical features present in the echinocandins; for instance, the hydrogen bond acceptor FragMap captures interactions involving hydroxyl groups common to all three compounds, as well as the sulfate moiety unique to micafungin. In total, 14 distinct FragMaps (listed in S1 Table) were used to guide flexible molecular docking of the echinocandins, resulting in binding poses optimized for relative free energy.

The top SILCS-docking poses for each of the echinocandins are presented in Fig. 1B-D. For anidulafungin and micafungin, the four poses with the most favorable (i.e., lowest) ligand grid free energy (LGFE) scores were selected (LGFE scores estimate binding affinity based on FragMap overlap; lower values indicating stronger binding). For caspofungin, the three top-scoring poses were chosen, along with a fourth pose that, despite a slightly higher LGFE score, exhibited a distinct orientation of the hydrophobic tail directed toward HS3. All selected poses were subsequently refined through MD simulations. The refined poses were then used to assess

how specific amino acid substitutions might alter drug binding or interaction networks, providing structural insight into potential resistance mechanisms.

In all four poses generated by SILCS docking, anidulafungin adopted a consistent orientation, with its hydrophobic tail positioned near HS3 and the peptidomimetic macrocycle embedded in the solvent-exposed cavity between HS1 and HS2 (Fig. 1B). Due to the limited presence of strong hydrogen bond donors or acceptors (S1 Fig), this binding mode was primarily driven by the insertion of the hydrophobic tail into apolar FragMaps adjacent to HS3, and by the placement of the homotyrosine moiety into one of the two apolar subpockets located at the opposite end of the binding site, near the extracellular space. Additional stabilization was provided by hydroxyl groups aligning with the H-bond acceptor pockets near HS1 and HS2. Poses 1 and 4 differed from poses 2 and 3 in the orientation of the macrocycle: in poses 2 and 3, the methylhydroxyproline group faced the hotspot cavity, whereas in poses 1 and 4, it was oriented toward the lipid membrane (Fig. 1B).

The tail positioning near HS3 in our model is consistent with the proposal by Johnson and Edlind [11]. However, compared to the docking model of Jospe-Kaufman et al. [15], our poses positioned anidulafungin deeper within the HS1-HS2 cavity, with the homotyrosine moiety partially shielded from solvent exposure. In contrast, in the Jospe-Kaufman model, the homotyrosine remained exposed to the extracellular space, and the tail extended only shallowly into the membrane. Their poses reflect a cavity occupied by lipids, which is incompatible with the polar FragMaps generated in our simulations and with the presence of nearby polar residues E635, D684, and D691.

Unlike anidulafungin, the top-ranked caspofungin poses adopted an opposite orientation, with the hydrophobic tail positioned between HS1 and TM6, and the homotyrosine moiety oriented toward the distal end of HS3 (Fig. 1C). This orientation was primarily driven by the placement of the two ornithine residues (S1 Fig) into methylammonium nitrogen FragMaps near D684 (between HS1 and HS2) and D691 (at the junction of all three hotspots), as well as the insertion of the homotyrosine group into apolar and H-bond acceptor FragMaps at the extremity of HS3. The smaller size of caspofungin's hydrophobic tail, compared to that of the other two echinocandins, facilitated its accommodation within the apolar cavity between HS1 and TM6. As with anidulafungin, additional stabilization was provided by hydroxyl groups aligning with H-bond acceptor FragMaps.

A lower-ranked caspofungin pose (fourth panel in Fig. 1C) adopted an orientation similar to that of anidulafungin, with the hydrophobic tail inserted into apolar FragMaps along HS3 and the homotyrosine placed into one of the two apolar subpockets between HS1 and HS2. This alternative pose also maintained favorable interactions via the two ornithine residues occupying the methylammonium nitrogen FragMaps near D684 and D691.

The first and fourth poses of micafungin adopted an orientation similar to that of anidulafungin, with the sulfated dihydroxy-homotyrosine (S1 Fig) positioned within the acetate carboxylate carbon FragMap near D1353, at the top of HS2 (Fig. 1D). In contrast, the second and third poses featured the hydrophobic tail aligned along apolar and H-bond donor FragMaps that follow the HS2 helix, on the side opposite the main hotspot pocket. Notably, in the third pose, the sulfate moiety was also located within an acetate carboxylate carbon FragMap, near R1357 (HS2).

Drug binding is mediated through interactions with key residues both in and out of the hotspots

The twelve selected docking poses were each subjected to a 100-ns MD simulation, during which the interaction energy between Fks1 and the corresponding echinocandin was monitored. For each compound, the trajectory exhibiting the most favorable (i.e., lowest) interaction energy between Fks1 and the ligand at 100 ns was selected for an additional 100 ns of MD simulation. In the case of caspofungin, two trajectories exhibited similar energy values at 100 ns, both were extended, and the trajectory with the most favorable interaction energy at the end of the simulation was retained for further analysis. The final frames from these trajectories are shown in Fig. 2 and correspond to pose 1 of anidulafungin (first panel in Fig. 1B), pose 1 of caspofungin (first panel in Fig. 1C), and pose 4 of micafungin (last panel in Fig. 1D).

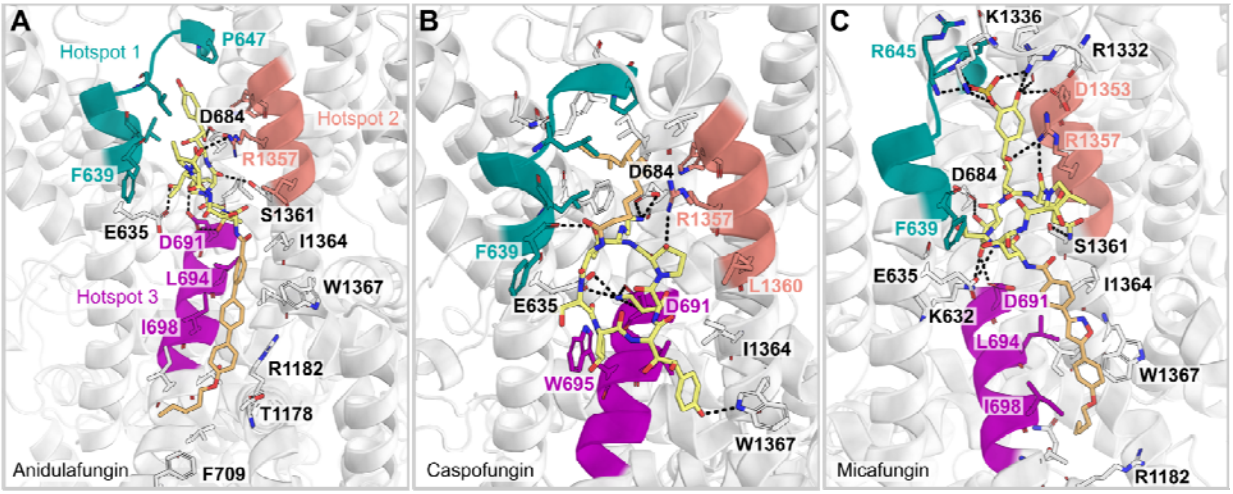


Fig. 2. Docking poses after 200 ns of MD simulations. Shown here are the selected poses for (**A**) anidulafungin, (**B**) caspofungin, and (**C**) micafungin. The twelve initial SILCS-guided poses (Fig. 1B-D) were refined by MD simulations, and the structures shown correspond to those with the most favorable interaction energies with Fks1. Hotpots and the labels of their associated residues are colored as follows: HS1, dark cyan; HS2, salmon; HS3, purple. Labels for residues outside the hotspots are colored black. Atoms are colored as follows: carbon, yellow (for drug macrocycles), orange (for drug lipophilic tails) or white; nitrogen, blue; oxygen, red. Predicted hydrogen bonds are depicted as black dashed lines.

For anidulafungin (Fig. 2A), the homotyrosine moiety remained stably positioned within a hydrophobic pocket formed by residues F463, F467 and N470 from TM4; L640, V641, L644, P647 (from hotspot 1) and I648; V680 and I681 from TM6; and W1354 and R1357 from hotspot 2 (S3 Fig). At the opposite end, the hydrophobic tail inserted into a groove composed of L694 and I698 from hotspot 3, along with T701, I702, V705, F709, and extended further through contacts with residues along TM8, including I1364, W1367, and I1368, as well as T1178 and R1182 in the cytoplasmic domain (Fig. 2A and S3 Fig). These hydrophobic interactions were complemented by hydrogen bonds involving hydroxyl groups and residues E635, D684, D691 (HS3) (twice), R1357 (HS2) and S1361, along with interactions with structured water molecules

within the binding pocket.. The peptidomimetic macrocycle also established nonpolar contacts with F639 (HS1), L687 and L1360 (HS2).

For caspofungin (Fig. 2B), the two ornithine residues (S1 Fig) remained stably coordinated with E635, D684 and D691 (HS3) throughout the trajectory, along with water molecules within the binding pocket. Additionally, hydroxyl groups from the macrocycle formed hydrogen bonds with the backbone carbonyl of F639 (HS1), and with the side chain of R1357 (HS2), while the hydroxyl of the homotyrosine engaged in an H-bond with W1367 (TM8). The hydrophobic tail occupied the same pocket as the one occupied by the homotyrosine in the anidulafungin-bound structure (S3 Fig). The peptidomimetic moiety also established hydrophobic contacts with F639 (HS1), L694 and W695 (HS3), as well as with L1360 (HS2) and I1364.

In the case of micafungin (Fig. 2C), the distinct sulfated dihydroxy-homotyrosine moiety (S1 Fig) engaged in multiple stabilizing interactions, which may contribute to its higher inhibitory potency in the absence of human serum compared to the other two echinocandins [21]. It formed hydrogen bonds with the backbone of R645 and D646 from hotspot 1, and with the side chains of R1332 and K1336 from EL4 (flexible loop linking TM7 to TM8). These latter interactions occur within a positively charged cluster of residues that defines an oxyanion pocket, an electropositive environment at the membrane-solvent interface capable of stabilizing negatively charged groups, such as the sulfate moiety of micafungin. The hydroxyl of the homotyrosine made H-bonds with R1332 (EL4), D1353 (HS2) and the backbone of P1350, and also established a π–cation interaction with R1357 (HS2). Additional hydrogen bonds were observed between hydroxyl groups of the macrocycle and residues K632, D684, D691 (HS3), and R1357 (HS2). E635 contributed a hydrogen bond to an amide hydrogen of the macrocycle, and S1361 interacted with the nitrogen of the asparagine side chain within the macrocycle. F639 (HS1) also established contacts with hydroxyl groups. The hydrophobic tail of micafungin occupied the same groove as observed for anidulafungin, forming similar interactions with Fks1 residues, including L694, I698 (both from HS3), T701, R1182 and Q1186 in the cytoplasmic domain, and S1361, I1364, W1367, and I1368 from TM8 (S3 Fig).

For all three echinocandins, multiple hydrogen bonds were also formed with structured water molecules within the binding pocket, further stabilizing the complexes. In addition, residues E635, D684 and D691 (HS3) were consistently involved with all three compounds, which means substituting them would likely confer resistance. However, these residues are highly conserved, suggesting that substitutions at these positions may incur a substantial fitness cost. This could partly explain why relatively few resistance-associated mutations have been reported at these sites [10], although it is worth noting that routine sequencing often omits regions outside the defined hotspots.

To test this hypothesis and validate our binding model, we recreated 13 Fks1 single mutants targeting residues predicted by our molecular dynamics simulation to contribute to drug binding and resistance. As expected, the hotspot 3 mutant D691V conferred resistance to all three echinocandins (S4 Fig). In our model, D691 forms key electrostatic and hydrogen-bonding interactions with positively charged or polar groups from the echinocandin macrocycle (notably the ornithine moiety of caspofungin), anchoring the drug within the binding pocket. Substitution to a valine disrupts these polar interactions by introducing a nonpolar side chain, thereby destabilizing ligand binding across all compounds and resulting in pan-echinocandin resistance.

Similarly, substitution of E635 to either a serine or an arginine also led to pan-resistance (S4 Fig), likely due to the loss or inversion of charge that disrupts both the water-mediated network

and ligand-specific hydrogen bonding. Interestingly, replacing E635 with a valine conferred resistance only to anidulafungin and micafungin, but not caspofungin. This substitution eliminates hydrogen bonding capacity and introduces a hydrophobic side chain, which likely perturbs interactions with the threonine-derived groups of anidulafungin and micafungin more significantly than with the distal ornithine group of caspofungin.

Mutating D684 to arginine conferred pan-echinocandin resistance, whereas substitution with valine or serine did not, underscoring the importance of both charge and interaction geometry at this position. Altogether, these findings provide functional support for the structural specificity of echinocandin binding and highlight the critical role of residues shaping the membrane-solvent interface.

Mutations at R1332, K1334 or K1336 to either an apolar or a negatively charged residue did not confer resistance to any echinocandin (S4 Fig). However, nearly all substitutions at these positions resulted in a significant fitness cost in the absence of drug (S4 Fig), suggesting that they play an important role in protein function. The EL4 loop is defined by two highly conserved cysteines (C1328 and C1345) that form a disulfide bridge. This structural motif is remarkably conserved across Fks homologs, 98% of sequences retain this cysteine pair, and 176 out of these 200 unique loops contain at least one lysine or one arginine. EL4 was also recently shown to be essential for enzymatic activity [13]. Given the flexibility of the loop and the apparent functional redundancy among its positively charged residues, this region is likely more tolerant to mutations than other critical residues such as E635 or E691.

Deep-mutational scanning of the putative binding sites of echinocandins

In order to complement our binding model predictions and enhance our understanding of the mechanisms of resistance, we performed deep-mutational scanning of all three hotspots of Fks1 by engineering the corresponding positions in the genome of *S. cerevisiae*. In this organism and many pathogenic fungi such as the closely related yeast *Nakaseomyces glabratus*, Fks1 homologs and particularly the three hotspots are very conserved (Fig. 3A and S5 Fig). *S. cerevisiae* is not only the most appropriate model to generate extensive mutagenesis data directly on the protein that was used for solving the structure, it also does not preclude extrapolating molecular mechanisms to pathogenic species.

Mutant oligo libraries (NNK codon at each position) were amplified and integrated at the endogenous hotspot loci using CRISPR-*Cas9* editing (Fig. 3B). We also amplified libraries containing unique hotspot sequences found in homologous *FKS* genes (used to validate machine-learning prediction of intrinsic resistance, as discussed below). The mutant strains were cultured overnight without selective pressure. They were then exposed to the three echinocandins (anidulafungin, caspofungin, micafungin). The concentrations used were estimated to inhibit around 80% of wild-type growth in most conditions. Cells were harvested before and after two consecutive rounds of selection and the genomic loci corresponding to the three hotspots were sequenced to estimate the extent of selection acting on each variant using changes in relative frequencies.

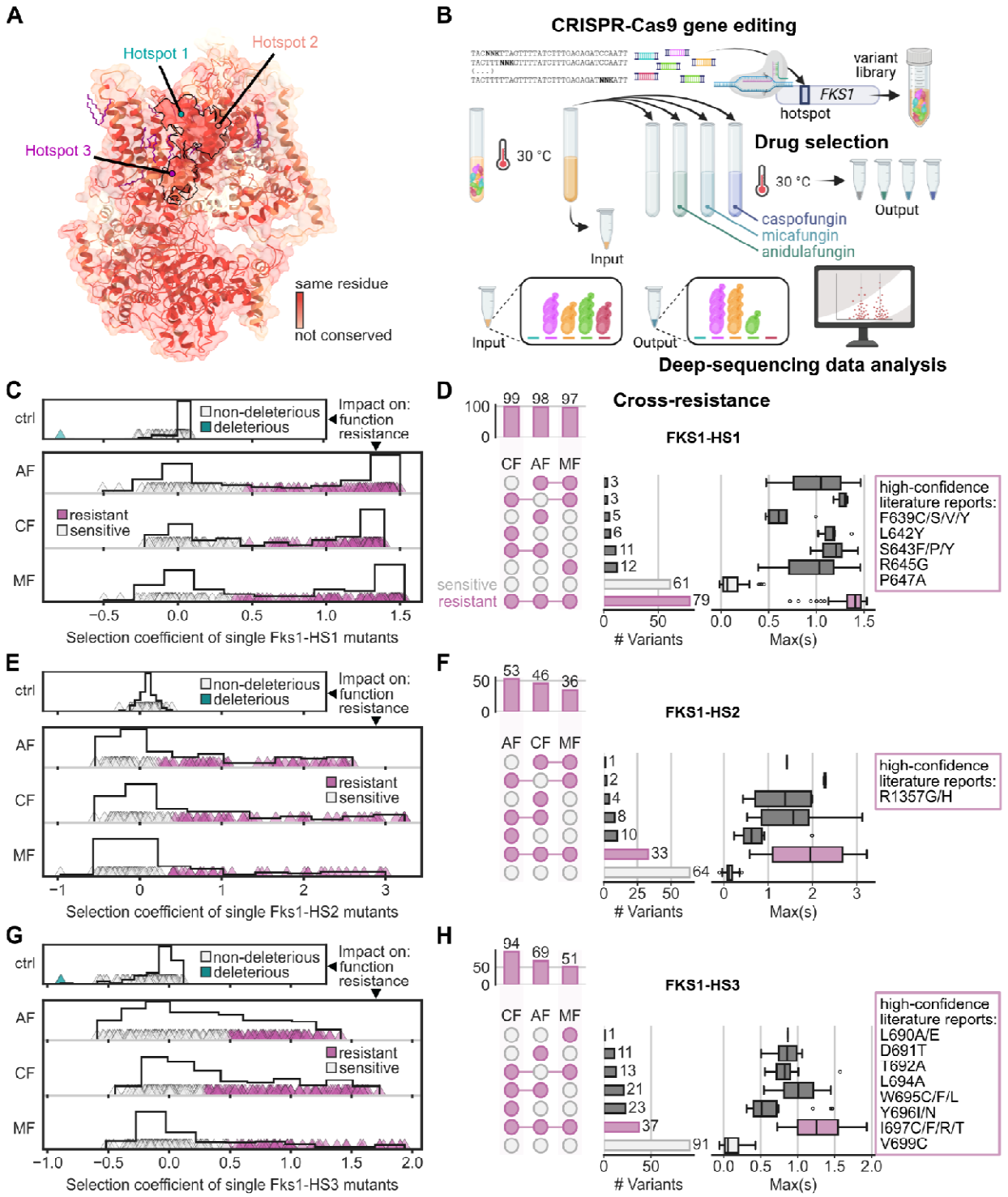


Fig. 3. Deep-mutational scanning of β-glucan synthase (Fks) hotspots identifies specific and cross-resistance mutations to echinocandins. (**A**) Fks1 structure from *S. cerevisiae* (PDB: 7XE4) [12], visualized using UCSF ChimeraX [22], colored according to Jalview's conservation score [23] from the structural alignment of Fks1 from *N. glabratus*, *Candida albicans* and *C. auris* (AlphaFold predictions for these three homologs, as shown in S5B Fig). (**B**) Simplified experimental design for the DMS assay, created with BioRender, showing how a library of single

mutants can be obtained by integrating oligonucleotides with NNK codons at the endogenous locus of any *FKS* hotspot. The library is cultured and subjected to bulk competition. Following genomic DNA extraction and amplicon sequencing, selection coefficients are calculated from the change in allele frequencies before/after screening. (**C, E, G**) Distribution of effects for single mutants of the (**C**) hotspot 1 (**E**) hotspot 2 and (**G**) hotspot 3 of Fks1 (including wild-type) in each condition. In the control condition (ctrl), the effect relates to protein function, whereas in the presence of drug (AF, anidulafungin; CF, caspofungin; MF, micafungin), the effect informs on resistance. Effects were quantified using the selection coefficient (the wild-type is expected to have a selection coefficient around 0). Variants were classified based on their effect using a Gaussian mixture model (see methods). The black line indicates the distribution regardless of the classification. (**D, F, H**) UpSet plots showing the number of unique (**D**) Fks1-HS1, (**F**) Fks1-HS2 and (**H**) Fks1-HS3 amino acid sequences conferring resistance (pink circles) or being sensitive (white circles) to each drug, and their overlap. For example, in panel D, 79 single mutants are resistant to all three echinocandins. Barplots at the top left of each panel indicate how many variants are resistant to the indicated drug. On the right, a boxplot shows the corresponding maximum selection coefficient across drugs for each set. The mutations listed are high-confidence reports from previously published studies for which there was at least one observation of resistance that matched the results presented here (S1-S2 Data). These reports include observations made from several fungal pathogens featuring on the WHO priority list [24].

For all mutated hotspots, at least one nucleotide variant could be successfully sequenced in the initial pools for every expected amino acid variant for both types of libraries (NNK and homologous hotspots). However, sequencing the diversity in the DMS input revealed that culturing overnight in rich medium was sufficient for several variants to be depleted, because of their impact on fitness in the absence of drug (or the impact of surviving variants, S6 Fig) and/or due to experimental bottleneck(s) (S7 Fig). Depletion was mainly observed for NNK and homologous variants of hotspot 2 (22% and 33%, respectively, S7 Fig), ruling out experimental bottleneck(s) as possible cause. The average mutant coverage (assuming even attribution of reads) for the corresponding sequencing run exceeded 600X, which leads us to infer that the depleted variants had a deleterious effect on fitness. We hypothesize that deleteriousness associated with hotspot 2 is most likely due to its proximity to the putative glucan translocation channel, as we discuss more in detail in later sections. In contrast, NNK and homolog libraries of hotspot 1 and hotspot 3 showed little to no depletion (at least 92% of variants remaining, S7 Fig).

Selection coefficients were computed for each nucleotide variant and then averaged within each amino acid variant. Replicates among cultures (at the nucleotide level) correlated well, with a Pearson's correlation coefficient above 0.88 for all drug conditions (S8 Fig).

Overall, we obtained a bimodal distribution of effects when selecting for echinocandin-resistant mutants, regardless of the drug used (Fig. 3C). This bimodality is consistent with strong selective pressure allowing only resistant mutants to grow significantly. In the case of hotspot 2 and hotspot 3 mutants, the two modes were less clearly defined, mostly because there were more sensitive mutants than for hotspot 1 (Fig. 3). We also observed an average selection coefficient for sensitive mutants in hotspots 2 and 3 below 0 (Figs. 3E,G and S9 Fig), again indicative of fitness cost. In the case of hotspot 2, harsh selection manifested from the generation of libraries (see methods) up until the end of the first screening round (t1), with cultures growing systematically slower than other conditions, reaching fewer mitotic generations (S6B Fig). In

summary, selection coefficients for hotspot 2 mutants were not as low as expected because of survivorship bias. In contrast, fitness costs associated with mutating hotspot 3 did not manifest until the screening round.

Validation and classification of mutational effects

In order to validate the effects of these mutations on resistance, 28 individual Fks1-HS1 mutants spanning the full fitness range were recreated by genome editing. The estimated fitness values strongly correlated with the selection coefficients from the bulk-competition assay (Spearman's r = 0.92-0.96 across drugs, p-value <= 4.4e-11, S10 Fig).

499 Fks1 single mutants were confidently classified as either sensitive or resistant to caspofungin, micafungin or anidulafungin across hotspots 1, 2 and 3, using a Gaussian mixture model (Fig. 3 and S9 Fig).

We find that 72 mutants across all three hotspots (14% of classified mutants) were specifically resistant to a single echinocandin molecule (Figs. 3D,F,H). 57% were resistant to at least one echinocandin drug, while 30% were resistant to all three tested molecules. More than half of these pan-echinocandin resistant mutants were in hotspot 1 (n=79, Fig. 3D), which may partly be explained by the lower fitness cost compared to the two other hotspots.

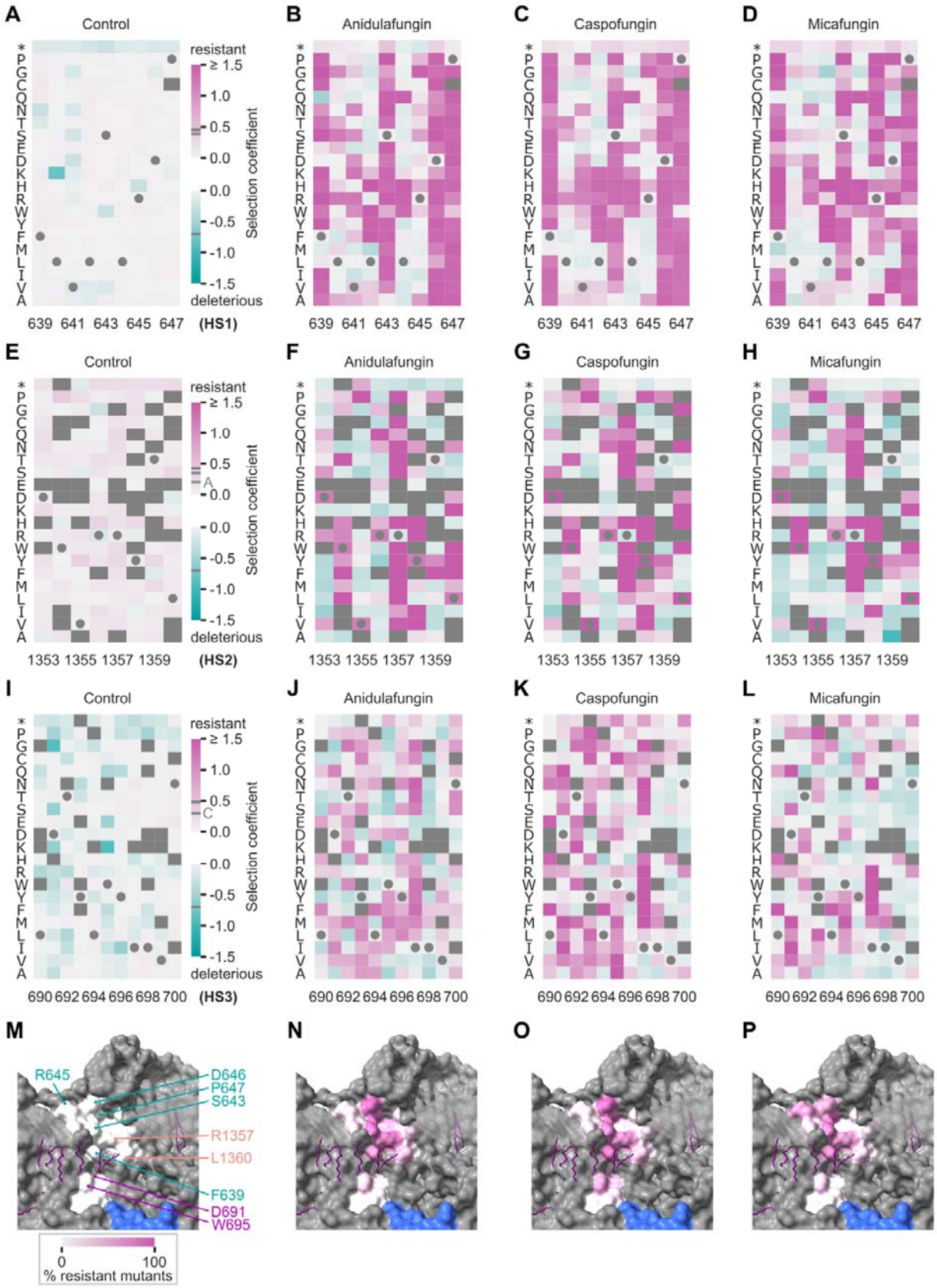
Consistent with our binding model, the same number of mutations in hotspot 1 conferred resistance to each drug, although we did observe more mutants conferring micafungin-specific resistance compared to the two other hotspots (12 in hotspot 1, none in hotspot 2, 1 in hotspot 3, Figs. 3D,F,H). This was most likely caused by the loss of hydrogen bonds normally formed with R645/D646 and micafungin's distinct sulfated dihydroxy-homotyrosine moiety (Fig. 2C).

In contrast, a greater number of hotspot 3 mutants conferred resistance to caspofungin (n=94) than to anidulafungin (n=69) or micafungin (n=51, Fig. 3H), confirming the crucial role played by three hotspot 3 residues (D691, L694 and W695) in caspofungin binding (Fig. 2B).

Instead, hotspot 2 mutations were more associated with anidulafungin resistance (Fig. 3F). This pattern is consistent with the homotyrosine moiety of anidulafungin fitting in the hydrophobic pocket involving W1354 and R1357, as well as the hydrogen bond and nonpolar contact with R1357 and L1360, respectively (Fig. 2A).

Several of these mutations have previously been engineered or identified in lab-evolved and clinical strains of a number of fungal species, including seven that feature on the WHO fungal priority pathogens list [24]: *C. auris*, *Aspergillus fumigatus*, *C. albicans*, *N. glabratus*, *Candida tropicalis*, *Candida parapsilosis* and *Pichia kudriavzevii* (S1 Data obtained, from the FungAMR database [10]). Our results validate the reported resistance (or sensitivity) phenotype for 27 of these mutants (S2 Data). As pointed out in the FungAMR study, which aggregates most reports of antifungal resistance to date, resistance among species is likely to be acquired through parallel evolution to the same variants [10]. Our results support this hypothesis, suggesting that observations made in one species are informative for predicting resistance phenotypes in other species.

Echinocandin-specific resistance depends on the position and electrostatic profile of substituted residues


A
Control
B
Anidulafungin
C
Caspofungin
D
Micafungin
resistant
≥ 1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
deleterious
Selection coefficient
639 641 643 645 647
(HS1)
E
F
G
H
1353 1355 1357 1359
(HS2)
I
J
K
L
690 692 694 696 698 700
(HS3)
M
N
O
P
R645
D646
P647
S643
R1357
L1360
F639
D691
W695
0
100
% resistant mutants

419

Fig. 4. Crucial positions in Fks1 for echinocandin resistance. (**A-L**) Selection coefficients (median of synonymous variants) of individual mutants in hotspot 1 (A-D), hotspot 2 (E-H) and hotspot 3 (I-L) of Fks1 in control (A, E, I), anidulafungin (B, F, J), caspofungin (C, G, K) and micafungin (D, H, L). The wild-type sequence is indicated by gray circles. We computed the selection coefficients of five mutants absent from the initial mutant libraries: F639C, L640D, L642G, P647N and P647Q in validation assays. In all panels, some variants have a selection coefficient above 1.5 but were set to 1.5 to ease the data representation (see the distribution of values in Fig. 3). (**M-P**) For each condition (same order as previous panels), all three hotspots in the structure of Fks1 are colored according to the proportion of mutants classified as resistant at each position (relative to the number of mutants for which we were able to obtain a selection coefficient). The labels of key residues are colored according to the hotspot, residues in blue at the bottom of each panel belong to the catalytic domain, while residues in gray correspond to other non-mutated residues in the transmembrane domains. The approximate location of membrane bilayers is shown by two semi-opaque circular planes (spheres overlapping with the hotspots are hidden for clarity). Lipids present in the structure [12] are represented in purple sticks.

Next, we used the full deep-mutational scanning dataset generated and compared it to our binding model to even more precisely interpret the molecular bases of resistance.

Our data revealed a strong position-dependent pattern of echinocandin resistance (Fig. 4B-D). Nearly all mutations at positions 639, 643, 646 and 647 conferred resistance to anidulafungin and caspofungin, while resistance to micafungin was primarily associated with mutations at positions 639, 643, 645 and 647 (Fig. 4D). Notably, these resistance-conferring mutations imparted little to no fitness cost (Fig. 4A), which may partly explain why they are so frequently associated with resistance (S1 Data).

Hotspot 1 is located at the water-membrane interface (Fig. 1A). F639 is embedded in the hydrophobic core of the membrane, S643 lies near the lipid-water interface, and R645 and D646 are fully solvent-exposed. In this region, the alpha-helix undergoes a transition in orientation: residues upstream F639 adopt a transmembrane orientation, whereas those following P647 align more parallel to the membrane surface and are positioned toward the extracellular side of the protein. This shift is favored by a combination of structural features, including the helix-breaking effect of P647, the propensity of S643 to induce local distortion in the helix backbone [25], the membrane anchoring of F639, and the strong solvation of R645 and D646. Mutations that disrupt this delicate balance may promote helix extension, altering access to the hydrophobic pocket where anidulafungin and caspofungin bind, and modifying exposure of the D646-P647 backbone region, which is critical for micafungin binding.

Although the model predicts interactions with L640 and L644, relatively few substitutions at these positions conferred resistance (Fig. 4B-D). This may be due to their location within the hydrophobic pocket, which means mutating them would not affect binding much, compared to mutating residues directly lining the cavity. A notable exception is the L644W mutation, which retained hydrophobic character but introduced substantial steric bulk, effectively filling, and potentially occluding the pocket, leading to cross-resistance.

A key distinction across the three drugs is the mutational sensitivity of positions 645 and 646. Substitutions at D646 more frequently conferred resistance to caspofungin and anidulafungin, whereas mutations at R645 were more commonly associated with micafungin-specific resistance.

This pattern is consistent with the micafungin docking pose, in which its sulfate group formed hydrogen bonds with the backbone of R645 and D646 (Fig. 2C), rather than occupying the hydrophobic pocket. Although micafungin also interacted with the side chain of D646, it tolerated steric variation in this region, as illustrated by the D646E mutation, which did not affect micafungin sensitivity (Fig. 4D).

In the case of hotspot 2, many variants exhibited a deleterious effect on fitness, both in the presence and absence of echinocandin (Fig. 4E–H). In particular, mutations introducing negatively charged residues, such as aspartate or glutamate, consistently resulted in a strong fitness cost (Fig. 4E). This sensitivity is likely due to the proximity of hotspot 2 to residues F1311 (TM7b) and F1475 (TM12), which have been proposed to form part of the exit of the putative glucan translocation channel (S11 Fig) [12,14]. Disruption of the delicate balance between TM8, which contains hotspot 2, and the neighboring transmembrane helices TM6, TM7b and TM12 could interfere with gating of the translocation channel at this site. This interpretation is further supported by the observation that mutations at positions 1353, 1356, and 1357 incurred a lower fitness cost (Fig. 4E), likely because these residues are exposed to the solvent or lipid environment rather than being involved in helix-helix packing.

Many substitutions at position 1357 conferred strong pan-echinocandin resistance (Fig. 4F-H). This is consistent with the refined docking poses, which revealed key polar and non-polar interactions between this residue and the macrocyclic core of each echinocandin (Fig. 2A-C). Overall, the DMS of hotspot 2 also showed a higher number of mutations conferring resistance to anidulafungin, while none conferred micafungin-specific resistance (Fig. 3F). Notably, the series of mutants W1354Y/F/M/L did not affect micafungin sensitivity, in contrast to their impact on anidulafungin and caspofungin resistance. This observation is consistent with our proposed binding mode of micafungin, where the sulfate moiety of the sulfated dihydroxy-homotyrosine engaged the backbone of R645 and D646 from hotspot 1, as well as the side chains of R1332 and K1336 from EL4. As a result, micafungin adopted a more superficial binding mode, in contrast to anidulafungin and caspofungin, which were positioned deeper within a hydrophobic pocket formed by residues L640, L644, and P647 from hotspot 1, in close proximity to W1354 (S3 Fig).

Consistent with our binding model and confirming early experimental validations, 6 missense mutations at position 691 in hotspot 3 (including the recreated mutant D691V) conferred pan-echinocandin resistance (Fig. 4J-L), highlighting the central role of the aspartate in drug binding. Also in hotspot 3, 12 missense mutations at position 695 conferred caspofungin resistance, compared to only 6 mutations conferring resistance to anidulafungin or micafungin (Fig. 4J–L). We also corroborated the earlier observation that the W695C mutation confers caspofungin-specific resistance [6]. With limited biophysical data available, this phenotype was proposed to result from differential binding of the echinocandin lipid tails. Our results now detail a structural mechanism supporting distinct binding modes: the lipid tails of anidulafungin and micafungin bind near this region, whereas for caspofungin, it is the macrocycle itself that directly interacts with D691 via an ornithine moiety (Fig. 2B), positioning the macrocycle in close proximity to W695. Given the demanding nature and computational cost of these simulations, we were limited to 200-ns trajectories in the present study. While these are sufficient to observe key binding interactions, we cannot exclude the possibility of a direct interaction between W695 and the macrocycle of caspofungin.

Similarly, mutations at positions 694 or 698 had a greater impact on caspofungin sensitivity than on anidulafungin or micafungin, further highlighting the distinct role that hotspot 3 plays

depending on the drug. Substitutions in this region most likely have minimal effect on the groove occupied by the tails of anidulafungin and micafungin, resulting in limited impact on their binding. Consistently, we observed mutants with differential cross-resistance profiles; for example, the I697S/T/C/H mutations conferred resistance to anidulafungin and caspofungin, but not to micafungin (Fig. 4J–L). This observation may be explained by the positioning of micafungin's tail, which appears slightly more distant from hotspot 3 compared to anidulafungin (Fig. 2).

Mutational landscapes of paralog hotspots overlap

In several clinically relevant fungi, as well as *S. cerevisiae*, there are at least two duplicated *FKS* genes [26]. In our model yeast, *FKS2* is expressed at a very low level during vegetative growth, unless *FKS1* is repressed or inactivated [27]. As a result, both paralogs are synthetically lethal, which allows editing them directly in the genome. To assess the impact of amino acid substitutions in Fks2, we built libraries for hotspot 1 and hotspot 2 as we did for Fks1 and measured fitness cost with and without repression of *FKS1*, as well as fitness effects in the presence of drug with repression of *FKS1*.

As expected for synthetic lethal gene duplicates, introducing premature stop codons in *FKS2* hotspots had little effect in the WT background but resulted in a strongly deleterious effect when *FKS1* was repressed (S12 Fig). Despite lower resolution for *FKS2* compared to *FKS1*, the pattern observed for other substitutions largely corroborated those obtained with Fks1 (S12 Fig), which we attribute to the high degree of protein sequence identity (87.7%) and functional redundancy between the two paralogs. More specifically, substitutions at positions 662, 1376 and 1379 (homologous to S643, R1357 and L1360 in Fks1, respectively) frequently conferred pan-echinocandin resistance, while substitutions at R664 (homologous to R645 in Fks1) led to micafungin-specific resistance (S12 Fig). Although we did not mutate hotspot 3 in Fks2, previous studies have reported overlaps in resistance profiles for equivalent mutations across both paralogs. For example, the Fks2 W715L mutation (homologous to W695 in Fks1) was shown to confer resistance to anidulafungin and micafungin in *N. glabratus* [28].

The simple sequence-resistance maps allow prediction of resistance from sequences alone

The full sequence-resistance relationship we obtained from the DMS may allow the prediction of resistance from DNA sequences alone. To test this hypothesis, we present a proof-of-concept random forest classifier trained on the DMS labels (sensitive/resistant) of all single Fks1-HS1 mutants classified with high confidence (n=180), which predicted echinocandin resistance for 46 unique hotspot 1 sequences found in homologous *FKS* genes, regardless of how many residues differ from the wild-type Fks1-HS1 sequence (2 to 7).

First, on the experimental side, we made sure to spike the initial libraries of single mutants with the homologous hotspot sequences. That way, when we would get the predictions from the machine-learning model, we would also have the experimental data to test their accuracy. On the computational side, we only trained and ran the model on hotspot 1 mutants, rather than on all hotspots, because it was the only hotspot for which the dataset was balanced in terms of resistant vs sensitive mutants, with good consistency across all three echinocandins. Finally, we chose a random forest classifier, which is relatively simple to put into place, while still generating interpretable data. The classifier used a combination of position (in the hotspot) and amino acid

properties to build decision trees. The training and testing sets were obtained by splitting all single mutants 70-30 randomly. After optimization (see methods, included precautions taken to minimize overtraining), we used the model to predict resistance for all homologous hotspots.

Our model accurately predicted resistance to any of the three tested echinocandins for more than 36 homologous hotspots, and up to 43 of them in the case of caspofungin (Fig. 5). Notably, the model predicted insensitivity to all echinocandins for the hotspot found in Fks1 from the fungal pathogen *C. parapsilosis*. This supports the hypothesis that the hotspot 1 sequence, which ends with an alanine instead of a proline because of a naturally-occuring polymorphism, is the main cause for the decreased sensitivity exhibited by several strains of *C. parapsilosis*, *Candida orthopsilosis* and *Candida metapsilosis* [29]. Several species of *Mucorales* encode at least two Fks paralogs, each with a different hotspot 1 sequence (S3 Data). Interestingly, the model accurately predicted resistance from one and sensitivity for the other. The former differs from *S. cerevisiae* by substitutions V641A, L644F and P647A, while the latter differs from *S. cerevisiae* by substitutions V641A, L644F and R645K. Although proper validation in the appropriate background is needed, it would seem that again, the presence of an alanine instead of a proline at position 647 is responsible for the intrinsic resistance observed in *Mucorales*. The model also accurately predicted intrinsic resistance to caspofungin and anidulafungin, but not micafungin, for the hotspot sequence found in Fks1 of several species of *Fusarium*. The hotspot, with its F639Y mutation (among others) compared to *S. cerevisiae*, was previously shown to play a critical role in intrinsic resistance of these pathogenic fungal species [30]. The reason why the model failed to accurately classify this hotspot (and eight others) as micafungin-resistant might lie in one of the model's known weaknesses, which is that many amino acid properties are assumed to correlate well with each other, impeding the decision process.

For all three drugs, the model predicted that the hotspot sequences would confer resistance (or not) mainly based on the residues at position 639 and 647 (S13 Fig). More specifically, if the residues favored extension of the helix, the model would predict resistance. This observation further confirms the role of the loop conformation of the first half of hotspot 1 as seen in the MD-relaxed structure of Fks1, and shows that a simple but interpretable machine-learning model can help gain insight into the molecular mechanisms of resistance.

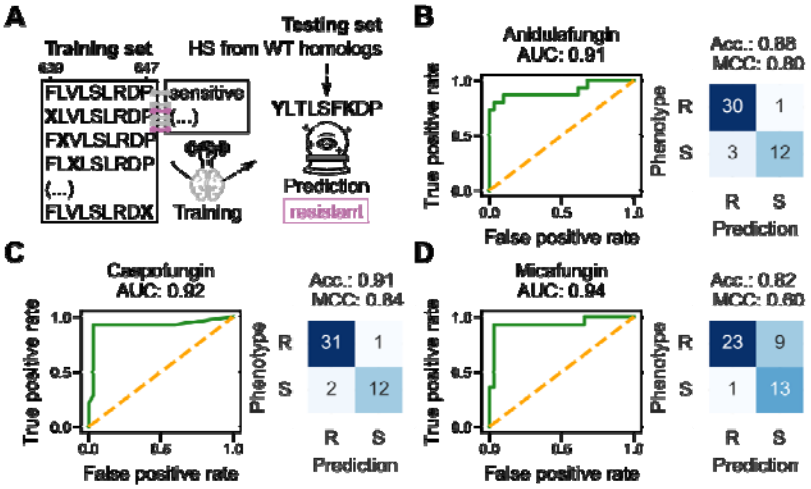


Fig. 5. Machine-learning assisted prediction of intrinsic resistance. (**A**) Resistance was predicted for all Fks1-HS1 homologous hotspots with two or more amino acid changes compared to *S. cerevisiae* WT. A random forest classifier was trained on single mutants of the DMS dataset that were classified with high confidence as sensitive or resistant to the drugs, using a combination of hotspot position and amino acid properties as features. "Training" icon by Creative Mahira and "Crystal Ball" icon by Agus Raharjo from Noun Project (CC BY 3.0). (**B-D**) For each drug, the AUC-ROC curve is displayed alongside the confusion matrix. Acc: balanced accuracy score, MCC: Matthew's correlation coefficient, R: resistant, S: sensitive.

Discussion

Interpreting the impact of mutations on drug susceptibility or improving drugs to overcome resistance requires both a comprehensive understanding of resistance-associated mutations and a detailed molecular model of drug binding. Using a combination of molecular modeling and screening of a large number of single amino acid mutants, we show that (1) the three known hotspots of Fks1 and their composition shape a single binding pocket partially exposed to the solvent, (2) echinocandins bind differently depending on their molecular structure, (3) combinations of factors including the loss of crucial interactions between the drugs and their target prevent binding-mediated inhibition, leading to specific and cross resistance.

Previous studies have proposed binding modes for echinocandins using molecular docking approaches [15,31]. Notably, the anidulafungin poses predicted by Jospe-Kaufman et al. [15] assumed the presence of lipids within the binding cavity. This arrangement is incompatible with the polar FragMaps observed in our simulations, as well as with the nearby polar residues E635, D684, and D691 (Fig. 1). The presence of lipids in this pocket is further contradicted by mutational data: substitutions at positions D691V, E635S, E635R, and D684R conferred resistance to all three echinocandins (Fig. S4), supporting a direct role for these residues in drug recognition.

Our simulations incorporate more physiologically relevant features of the system, including the flexibility of both the drugs and their target enzyme, as well as the surrounding membrane environment in which Fks1 is embedded. Simulating the dynamics of echinocandin-bound Fks1 confirms some but not all hypotheses previously emitted. For instance, an early topological model previously suggested that the lipophilic tail of echinocandins inserts into the membrane and interacts with hotspot 3 [11]. The structural models presented here support this for both anidulafungin and micafungin, with anidulafungin binding even deeper in the cavity than previously anticipated (Fig. 2). In contrast, we bring evidence that caspofungin's shorter and more flexible tail adopts an opposite orientation, favoring interactions between the macrocycle and hotspots 1 and 2, with the tail directed toward TM6 (Fig. 2B).

Our mutational data help refine and validate many of the interpretations drawn from the binding model. Unlike standard structure-function approaches, we estimate the mutational effects of several hundreds of mutants, with endogenous expression levels, and compare the landscapes obtained with the three most widely used echinocandins. Most resistance-associated mutations that have been identified in Fks1, whether it be from lab-evolved or clinical strains of fungi or here, are localized in hotspot 1 (Fig. 3 and S1 Data). Our results suggest that a lower fitness cost associated with mutations in this region may confer higher mutational tolerance in the absence of drug (Fig. 3).

In contrast, identifying the mechanisms underlying resistance in the absence of structural data has remained challenging. Although recent advances have provided insights into the overall architecture and mechanism of action of the enzyme [12,14], our model helps further elucidate the complex molecular mechanisms by which substitutions in the hotspots can confer resistance.

Located at the membrane-solvent interface in hotspot 1, F639 likely anchors the hotspot 1 helix, contributing to its proper positioning within the binding cavity. Substitutions at this position are expected to disrupt this structural balance, potentially reducing access to the binding site or altering drug affinity. These effects may explain why mutations at F639 are frequently associated with pan-echinocandin resistance (S1 and S2 Data). Our model also provides structural evidence

for the mechanisms underlying echinocandin-specific resistance. For instance, substitutions at position 645 in hotspot 1 led to two distinct mutational effects: one associated with resistance to anidulafungin and caspofungin, and another specific to micafungin (Fig. 4). Our modeling suggests that R645, located at the water–membrane interface, plays a key role in the transition of helix orientation in this region, a shift stabilized by membrane anchoring, solvation, and helix-disrupting residues such as P647 and S643. R645 contributes to the solvation-driven positioning of the helix, and mutations at this site may alter the accessibility and conformation of the binding pocket, thereby affecting echinocandin binding. For anidulafungin and caspofungin, such mutations may restrict access to the hydrophobic pocket, while for micafungin, R645 is directly involved in binding through electrostatic interactions (Fig. 2).

Although future work is needed to precisely characterize the binding modes of molecules beyond the three echinocandins studied here, our model can be used to infer the binding of closely related analogs, such as rezafungin. Rezafungin is a recently USDA-approved, long-acting echinocandin that enables once-weekly dosing (in contrast to the daily administration required for standard echinocandins) [32]. The molecule is an analog of anidulafungin that differs by a single substitution at the C5 ornithine side chain, where a hemiaminal is replaced with a choline-derived quaternary ammonium group [18]. This positively charged moiety is expected to localize within the methylammonium nitrogen FragMaps near D684 and D691, thereby favoring an orientation in which the methylhydroxyproline group faces the hotspot cavity, similar to poses 2 and 3 of anidulafungin (Fig. 1B).

While the resistance landscapes of anidulafungin and rezafungin are likely to overlap, differences in pharmacokinetics and therefore in effective drug concentrations in the host, may impact the number and nature of resistance mutations that are selected. Other echinocandin analogs currently under investigation include dehydroxylated versions of anidulafungin, rezafungin, caspofungin and micafungin. Interestingly, removal of both the hemiaminal and benzylic alcohol groups from micafungin abolishes antifungal activity, whereas selective removal of the benzylic alcohol from the dihydro-homotyrosine moeity of anidulafungin, rezafungin and caspofungin results in increased efficacy [15,33]. Our binding model provides structural insight into this observation. The removal of the benzylic alcohol may reduce steric hindrance or shift the position of the dihydroxy-homotyrosine group, allowing improved accommodation within the hotspot cavity and enhanced interaction with conserved polar residues.This structural rearrangement could contribute to the observed reduction in minimum inhibitory concentration (MIC) for some echinocandin-resistant strains of *C. albicans* and *N. glabratus* [15,33].

While modification of the existing compounds has generated a diversity of available molecules for this essential class of antifungals [18], there is a knowledge gap when it comes to their exact mechanism of action, particularly how each drug-specific binding mode is affected by amino acid changes in the hotspot sequences. Predictions of the echinocandin susceptibility profile from sequence alone could help prescribe the most relevant echinocandin, increasing the chances of treatment success while helping to slow down the development of resistance. This approach has successfully been applied to the etiologic agents of malaria or tuberculosis [34,35], but has yet to be implemented in fungi. Our data and recent work on other antifungal drug targets [36,37] contribute to this collective effort. Using a resistance-predicting model such as the one presented here does however come with several limitations. Most importantly, the training data was generated by an experiment in which only the hotspot 1 sequence was inserted in the *FKS1* gene of *S. cerevisiae*, in a *S. cerevisiae* background. However, if the resistance phenotype observed in

the DMS matches the phenotype of the original species, then the hotspot sequence may at the very minimum be sufficient to confer intrinsic resistance (or decreased sensitivity). On the other hand, such an experimental design cannot show that the hotspot sequence is necessary to confer intrinsic resistance. In other words, the model can provide insight into why certain species are intrinsically resistant to echinocandins, if there is a significant chance that resistance or reduced susceptibility is indeed caused by residues in hotspot 1 altering the binding pocket. The model is not meant to be used to understand any other cause of resistance either, for example interactions between alleles in diploid species such as *C. albicans*. As with any predictive tool, results should be interpreted with care, given that inhibiting Fks1 might not be enough to translate to organismal sensitivity. For example, our model accurately predicts that Fks1 of *Cryptococcus neoformans* is sensitive to all echinocandins, which has been shown before in vitro [38]. Despite this, *C. neoformans* displays decreased sensitivity [26]. Similar patterns in other species might be linked to the level of essentiality of Fks1, for example growth could depend more on a different cell wall component than β-1,3-glucan. In the case of *C. neoformans*, data from several studies point to a combination of factors, rather than a single cause of intrinsic resistance [39–41]. At this time, there is simply not enough data to quantify the extent of epistasis, i.e. mutations behaving differently in different species, a limitation that our model could help overcome. Indeed, to our knowledge, there is only a single mutation that was previously reported (with high confidence) to lead to different phenotypes in different fungal species: R647G in *C. albicans* (equivalent to R645G in *S. cerevisiae*) was reported sensitive to anidulafungin and caspofungin [42], whereas the equivalent mutation in *K. marxianus* (R657G) was reported to confer resistance to both drugs [43]. The same studies reported that the mutation conferred micafungin resistance in both species. If epistasis lies only within hotspot 1, then our model, which bases its classification on biophysical properties of each residue that make up the hotspot, can help pinpoint which molecular mechanisms are likely causes of intrinsic resistance exhibited by other species of fungi. One of the key contributing features of our model to classify a homologous hotspot as intrinsically resistant is the frequency at which the first hotspot 1 residue (at position 639) is generally found in alpha helices (of any protein): if it is seldom found, the model classifies the mutation as resistant, particularly to anidulafungin and micafungin (S13 Fig).

Beyond improving drug stewardship, elucidating the molecular mechanisms of echinocandin resistance and binding specificity has highlighted key elements in the structures of both the drugs and their target, which provide valuable information for the rational modification of existing drugs and design of novel compounds targeting Fks1.

For instance, our results have highlighted residues outside of the hotspots that contribute to drug binding. These residues may have been underreported in earlier resistance studies due to diagnostics biases, such as sequencing limited to the “traditional” hotspot regions. Alternatively, they may be less mutationally tolerant than the residues in hotspot 1, possibly due to essential structural or functional roles within the enzyme. This insight could be leveraged to guide the rational re-design of existing echinocandins to stabilize interactions with these mutationally fragile residues. Strengthening these interactions could slow the emergence of resistance by imposing a greater fitness cost on potential mutations.

In contrast, certain regions exhibit a different mutational profile. For example, the EL4 loop, which connects TM7 to TM8, contains conserved arginine and lysine residues that appear more tolerant to mutation. Our model reveals that EL4 forms an oxyanion cavity at the top of the binding site, primarily involving residues R1332 and K1336, and potentially K1334. These positively charged residues lie within interaction distance of the sulfated dihydroxy-

homotyrosine moiety of micafungin, suggesting a role in electrostatic stabilization. Consistent with this, our DMS results show that single substitutions in EL4 typically incur a strong fitness cost, underscoring the functional importance of this region. However, none of these individual mutations conferred resistance (S4 Fig), which aligns with our structural model: R1332 and K1336 can contribute redundantly to electrostatic interactions with micafungin, limiting the likelihood of resistance through single-site substitution. These interactions likely enhance micafungin's binding affinity and may help explain its distinct resistance profile compared to other echinocandins.

Recent findings in *N. glabratus* shed new light on this model: gene conversion events between the *FKS1* and *FKS2* paralogs were shown to produce a chimeric Fks1 that differs at multiple positions, including in EL4 [44]. More specifically, this structural rearrangement was identified in clinical isolates from a patient treated with micafungin, and likely an adaptive response to micafungin selection. This resulting chimeric protein included eight amino acid changes, four in or adjacent to EL4 (at positions equivalent to 1327, 1332 and 1336 in *S. cerevisiae*), one in hotspot 2 (V1355I in *S. cerevisiae*) and three close to the active site of Fks1 (positions 1411, 1417 and 1447 in *S. cerevisiae*). These combined changes correlated with an increase in MIC not only for micafungin, but also for fluconazole and voriconazole. The authors speculate that such gene conversion events may offer an evolutionary advantage by distributing the resistance burden across multiple positions, avoiding the high fitness cost of single critical mutations. To use the popular analogy of the adaptive fitness landscape, there might be a hiking trail around the fitness peak that is EL4.

Investigating the molecular bases of antifungal resistance can provide invaluable insights into both the likely evolutionary paths leading to resistance and opportunities for the rational design of improved therapeutics. By combining molecular simulations of target protein in a realistic environment with large-scale mutagenesis, our study advances the understanding of the mechanisms of action and resistance of echinocandins. Notably, our findings confirm that the key determinants of resistance are largely concentrated in the hotspot region.

Materials and Methods

Strains, plasmids, primers and culture media

Strains, plasmids, primers and oPools used in this study are listed in S4 Data. The following culture media were used: either YPD (1% yeast extract, 2% bio-tryptone, 2% glucose, with or without 2% agar) or SC (0.174% yeast nitrogen base without amino acids, 2% glucose, 0.5% ammonium sulfate, standard drop-out mix). Most chemical products were acquired from Fisher Scientific or BioShop Canada. When indicated, the following compounds were added to the medium: caspofungin (Cedarlane Labs), micafungin (Toronto Research Chemicals), anidulafungin (Millipore Sigma), hygromycin B (BioShop Canada), G418 (BioShop Canada).

Genetic constructions

Unless specified, yeast strain construction was done by standard transformation from competent cells and colony PCR validation by standard lithium acetate DNA extraction [45]. First, each hotspot (FKS1-HS1, F639-P647; FKS1-HS2, D1353-L1360; FKS2-HS1, L659-P666; FKS2-HS2, D1372-L1379) was replaced by a hygromycin (HYG) resistance marker in BY4741 (s_001 in S4 Data) and R1158 (s_002) using primers o_001-008 to generate strains s_003-010. In the resulting R1158 strains, the promoter of *FKS1* or *FKS2* was replaced by a doxycycline-repressible promoter using primers o_014-021 to generate strains s_011-014 [46]. Strains s_003-010 were used to introduce amplified mutated hotspot sequences: either synthetic libraries (for the deep-mutational scans) or amplicons obtained by fusion PCR (for individual reconstruction of single mutants). In order to reconstruct individual *FKS1*-HS1 mutants, we used primers mainly designed using NEBaseChanger (o_036-092) to amplify two non-overlapping fragments for each mutation. These fragments were individually amplified from genomic DNA of BY4741 using the KAPA HiFi HotStart DNA polymerase (Roche). For each mutation, we diluted and pooled both amplified fragments 1:20 and used 0.75 µL of the obtained pool as template for the fusion PCR using the Q5 High-Fidelity DNA polymerase (NEB). Strain s_003 was transformed as described elsewhere [47]. Each reaction was prepared with 20 µL competent cells, 20 µg carrier DNA, 250 ng pCas plasmid (with a guide RNA targeting the hygromycin resistance marker and carrying a G418 (KAN) selection marker), 8 µL PCR-amplified oPool and 200 µL PLATE buffer (PEG 3350 40%, 100 mM lithium acetate, 10 mM Tris-Cl, 1 mM EDTA pH 8.0). Briefly, transformants (selected on YPD + G418) were tested by colony PCR and cultured in YPD to cure the pCas plasmid. We used spot assays to confirm the antibiotic susceptibility profile, ensuring that the constructed strains had lost both selection markers (HYG (hotspot 1) + KAN (pCas)) and Sanger sequenced the mutated region with primer pair o_032-033 to confirm the introduced mutation. The wild-type hotspot was also reconstituted to ensure that the fitness of the obtained strain would not differ from the parental strain in subsequent growth assays.

In addition to missense mutations, we also attempted to introduce all possible codon deletions in the hotspot 1 of *FKS1*, as this type of mutation has previously been reported in *N. glabratus* *[5,48–50]*. However, all attempts yielded no colonies, hinting at a strong fitness cost for these mutations.

We followed essentially the same steps to reconstruct the other individual mutants in Fks1, including N470R, D691V in hotspot 3 (out of three mutations attempted at that position), three mutants at position E635, three mutants at position E684, two mutants at position R1332, two mutants at position K1334 and two mutants at position K1336. The notable difference with hotspot 1 mutants is that we first had to independently replace four more regions in the gene (corresponding to M465-F474, E635-P647, D684-N700 and D1331-L1340 in the protein sequence) with HYG, using primers o_093-099 to generate strains s_043-46. Fragments were amplified with primers o_101-126. The mutated regions were ultimately Sanger sequenced with primers o_032/034/100/. For each deletion strain (s_043-46), the wild-type was reconstituted and the resulting strains were included as controls in the growth assay.

Out of all missense mutations that we tried to engineer, only a few could not be obtained, including P647A, D691S and D691R.

Deep-mutational scanning (DMS)

Design of synthetic libraries

Two types of synthetic libraries for each of five Fks hotspots (FKS1-HS1, FKS1-HS2, FKS1-HS3, FKS2-HS1 and FKS2-HS2) were designed: the first one containing a single NNK codon for each position of the hotspot, the second containing unique codon-optimized sequences encoding for hotspots found in orthologs retrieved from Metaphors (S3 Data) [51]. To retrieve as many Fks1 and Fks2 orthologous sequences as possible, we used Fks1 from *Aspergillus niger* (M!42720474_ASPNG) and Fks2 from *N. glabratus* (M!115416204_CANGB) as queries. Because of the high percentage of sequence similarity between paralogs, hits for Fks1 sometimes included both paralogs of a given species (e.g. *S. cerevisiae* or *N. glabratus*). In contrast, using Fks2 from *N. glabratus* as the query did not return the paralog sequence, but returned two paralogs from *C. auris*. For this reason, we refer to the corresponding oPools and their associated variants throughout the text as homologous hotspots. Fks1 orthologous sequences were filtered based on their length using the following criteria: minimum of 1,500 residues (80% length of Fks1 from *S. cerevisiae*) and maximum 1,970 residues (105% length of Fks1 from *S. cerevisiae*). Fks2 orthologous sequences were filtered using the same criteria (80% = 1,517 residues, 105% = 1,991 residues). Three sequences were filtered out because they contained multiple stop codons throughout. The remaining sequences were all pooled and aligned with Muscle [52]. Aligned sequences were trimmed with trimal v1.2 using the automated heuristic approach [53]. Unique hotspot sequences were manually selected from the alignments and converted to nucleotide

sequences by choosing the most frequent codon in *S. cerevisiae* for each amino acid. For both types of synthetic libraries, the final sequences were generated using custom Python scripts by adding homology regions of 40 bp on either side of the hotspot (S4 Data).

Library preparation

Synthetic libraries were ordered as oPools from IDT. Each oPool was resuspended in biomolecular grade water to a final concentration of 0.5 µM for NNK pools and 5 µM for homolog pools, with an additional 10-fold or 100-fold dilution so that all pools were at a final concentration of 0.05 µM for the PCR amplification step. We estimated the total molar amount per tube to be equal to 50 pmol times the number of oligonucleotides (for example for the NNK pool: 9 oligonucleotides x 50 pmol = 450 pmol, resuspended in 900 µL). oPools were amplified with KAPA to add homology regions with sequences directly upstream and downstream of each hotspot, using primers o_022-031. CRISPR-*Cas9* transformations for genome editing were performed as described above, using strains s_003-006 and s_011-014. Transformants were scraped in 5 mL of YPD. The OD_{595} was measured, a glycerol stock was preserved at -70°C and two aliquots of 1 mL each were pelleted for downstream genomic DNA extraction. We estimated a single transformation per condition to yield enough colonies to saturate the expected diversity of variants. Despite a lower number of colonies for FKS1-HS2 (NNK and homolog pools), we managed to have sufficient coverage as we describe in the next section.

Pre-screening sequencing runs

For all hotspots except hotspot 3, all libraries (NNK or homologs, obtained from one transformation each) were sequenced to estimate the diversity of variants. For hotspot 3, libraries from four transformations (for each type, NNK and homologs) were sequenced separately. First, genomic DNA was extracted using a standard phenol-chloroform protocol [54]. For all hotspots except hotspot 3, PS sites were added by PCR using Q5 and primers o_127-134. Libraries were multiplexed by row-column indexing [55] using KAPA and primers o_135-148. Finally, they were amplified using KAPA and primer pair o_149-150 to add Illumina indexes, SPRI-purified and sequenced in paired-end 300 bp on an Illumina MiSeq system (three runs in total, Institut de Biologie Intégrative et des Systèmes sequencing platform, Université Laval). For hotspot 3, PBS sites were added using Q5 and a touch-down PCR (5 cycles with annealing at 55°C followed by 17 cycles of 2-step amplification, no annealing) with primers o_167-170. For the last step, we added i5/i7 dual barcodes by PCR using KAPA and primers o_187-194/207, SPRI-purified and sequenced in single-end 150 bp on an Aviti system (Institut de Biologie Intégrative et des Systèmes sequencing platform, Université Laval). For all libraries, we confirmed that all expected amino acid variants were present in the initial pools.

Inhibition curves

The parental strain BY4741 and strains in which either the hotspot 1 of *FKS1* or the hotspot 1 of *FKS2* was replaced with an hygromycin resistance marker (s_001, s_003 and s_005) were streaked on YPD agar medium supplemented (or not) with hygromycin B. Precultures were prepared from three isolated colonies in 5 mL of YPD medium with appropriate selection and incubated overnight at 30°C with shaking. Precultures were diluted to 1 OD_{600} in sterile water. 3 sterile Greiner 96-well plates were prepared with SC medium supplemented with two-fold dilutions of each of the three echinocandins (for caspofungin, from 0.0156 µg/mL to 0.25 µg/mL; for micafungin, from 0.0625 µg/mL to 1 µg/mL and for anidulafungin, from 0.125 µg/mL to 2 µg/mL) and cultures were inoculated at a final concentration of 0.1 OD_{600}. OD_{600} measurements were performed at 30°C every 15 min until a plateau was reached in a Tecan Infinite M Nano (Tecan Life Sciences). The maximum growth rate was transformed into the inhibition coefficient, with an inhibition coefficient of 0 corresponding to the maximum growth rate measured in the absence of drug.

Experimental design considerations

Preliminary growth assays using the parental strain BY4741 showed that drug concentrations previously estimated to inhibit growth at 50% or 75% in plates (S14 Fig) were close to completely inhibiting growth in 5 mL tubes, with a doubling time multiplied by more than 7.5 (compared to a culture without drug). Additional tests confirmed that concentrations corresponding to 10% inhibition in plates would be adequate for a screening in 5 mL tubes, with an actual growth inhibition estimated around 82% (corresponding to a doubling time multiplied by roughly 4.5). During our preliminary assays, we noticed that mutating Fks1 or Fks2 could have an effect on cell wall integrity. More specifically, we corroborated previous observations that a functional defect in glucan synthesis could lead to increased levels of chitin in the membrane [56]. Growth is usually monitored by simple OD_{600} measurements, but in this case, we would risk misestimating cell concentration, which we need to precisely calculate the number of mitotic generations (which is then used to calculate the selection coefficient of each variant). To solve this issue, we measured cell concentration by flow cytometry with a Guava easyCyte HT cytometer (Cytek). Ultimately, we did confirm that concentrations obtained by cytometry do not always correlate with OD_{600} measurements (S6A Fig).

Screening

Precultures were prepared in 5 mL of YPD medium supplemented with chloramphenicol to prevent any potential bacterial growth. They were inoculated in triplicates with 100 µL of glycerol stock and incubated overnight at 30°C with shaking. Homologous hotspot variants were spiked in NNK pools by mixing saturated precultures in ratios accounting for the total number of expected variants, based on the measured OD_{600}. Later on, read frequencies confirmed that the pools were mixed in the appropriate proportions and that the ratio of NNK variants to

homologous hotspot variants was not drastically impacted by selective pressure (S15 Fig). The mixed precultures were diluted to 1.5 OD_{600} in sterile water and used to inoculate 5 mL tubes containing SC medium supplemented with either sterile water, 0.022 µg/mL caspofungin, 0.15 µg/mL micafungin or 0.22 µg/mL anidulafungin, at a final concentration of 0.15 OD_{600}. Additionally, pellets of 1 mL of mixed precultures were prepared and stored at -70°C until genomic DNA extraction ("t0/input" samples). Cultures were incubated at 30°C with shaking until they reached approximately 1 OD_{600}, at which point they were adjusted to 1.5 OD_{600}. Pellets of 1 mL were prepared and stored at -70°C ("t1 samples") and fresh cultures were inoculated at a final concentration of 0.15 OD_{600} for a second round of selection. After diluting the t1 cultures at 0.6 OD_{600}, cell concentrations were measured by flow cytometry. After reaching approximately 1 OD_{600}, t2 cultures were processed similarly to obtain "t2/output" samples (except for hotspot 3 cultures supplemented with drugs, which were incubated overnight). The number of mitotic generations per selection round for each sample was calculated from cytometry measurements of cell concentrations by subtracting the logarithm of values before/after selection and dividing by the logarithm of 2. The total number of generations over both rounds of selection was obtained by summing up the numbers obtained for each round.

Genomic DNA extraction

Genomic DNA was extracted from t0 and t2 samples using the DNeasy Blood and Tissue kit (Qiagen), with the regular format for a couple of samples and all hotspot 3 samples, and with the DNeasy 96 protocol for the remaining samples. We followed the kit's protocol except for the modifications: initial lysis was performed with 6 units of zymolyase for 1h at 37°C, spheroplasts lysis was performed for 30 min at 56°C, RNase treatment was performed immediately after elution by adding 4 µL of 10 mg/mL RNase A and incubating 1h at 37°C. Samples were then SPRI-purified with a bead ratio of 0.4 and ultimately eluted in 50 µL of 10 mM Tris-HCl.

Post-screening sequencing runs

Two preliminary post-screening sequencing runs were performed from 10 and 4 samples, respectively. The first run was performed on an Illumina MiSeq system, whereas the second one was performed on an Illumina iSeq system. For the MiSeq run, PS sites were added using homemade Phusion polymerase and a touch-down PCR (5 cycles with annealing at 50°C followed by 15 cycles with annealing at 68°C), with the same primers as before (o_127-134). The rest of the protocol was similar to what was done previously (row-column indexing and final amplification using KAPA, SPRI purification and sequencing in paired-end 300 bp at the Institut de Biologie Intégrative et des Systèmes sequencing platform). For the iSeq preliminary run and both ultimate runs (NovaSeq for hotspot 1 and 2, Aviti for hotspot 3), PBS sites were added similarly to what was done with the pre-screening Aviti sequencing run of hotspot 3, i.e. using Q5 and a touch-down PCR (5 cycles with annealing at 55°C followed by 17 cycles of 2-step amplification, no annealing) with primers o_151-170. For the last step, we added i5/i7 dual

barcodes by PCR using KAPA and primers o_171-237. iSeq sequencing was performed in paired-end 150 bp at the CHUL sequencing platform, Université Laval. For both ultimate runs, all libraries were quantified in four technical replicates using the AccuClear Ultra High Sensitivity dsDNA Quantitation kit (Biotium). Quantified libraries for the NovaSeq run were manually pooled in four subpools of different concentrations (1X, 4X, 50X, 200X) depending on the concentration of individual libraries to optimize the pipetting volume. Overall, we aimed at an output of 800 million reads, with a coverage of 237X for every library. The four subpools were finally assembled into a single pool at a final concentration of 14 nM. Libraries were sequenced in paired-end 150 bp on an Illumina NovaSeq 6000 S4 system at the Centre d'expertise et de services Génome Québec. Hotspot 3 libraries were sequenced in single-end 150 bp on an Aviti system (Institut de Biologie Intégrative et des Systèmes sequencing platform, Université Laval).

Computational analyses of DMS data

The following sections describe the main steps of each computational analysis. Technical details and all scripts are available online on the dedicated GitHub repository. Demultiplexing of the MiSeq data and the DiMSum pipeline (see below) were run on the IBIS servers.

Quality control

Initial assessment of sequencing data showed that the number of obtained reads was almost consistently larger than the number of expected reads (2.9 times on average), with very little variability between samples. Around 10% of reads on average were lost upon merging, resulting in an overall coverage of 617X (across all conditions, based on the number of expected nucleotide variants in the DMS input). Very few reads were lost from that point on (due for example to indels or unexpected variants).

Variant counts

Variant counts were obtained using DiMSum v.1.3 [57]. DiMSum requires an experimental design file for every "experiment". Here, we considered an experiment to correspond to a single set of conditions (locus, compound, etc), for example BY4741 in which the hotspot 1 of *FKS1* was modified using NNK/homolog oPools and subjected to caspofungin selection (all input and output replicates). A custom Python script was used to generate all experimental design files. The DiMSum pipeline was run directly on demultiplexed samples for each "experiment". It includes the following steps: checking sequencing quality of raw reads using FastQC, trimming constant regions using cutadapt v2.4, align paired-end reads using VSEARCH v2.26.1, tally unique sequences using starcode v1.4.

Selection coefficients

A custom Jupyter notebook was used to process variant counts for each "experiment". The following steps were performed: discard variants with indels, keep only variants with at least 3 reads in at least 1 input replicate, label all variants with a confidence score based on the read count in input replicates (1 = 10+ reads in all input replicates, 2 = at least 4 reads in at least 1 input replicate, 3 = up to 3 reads in any input replicate), keep only expected variants based on the matching oPool design, add 1 to all read counts, calculate read frequencies, calculate log2 fold-change (frequency in the output replicate divided by frequency in the matching input replicate), normalize it with the number of mitotic generations (g; previously obtained from cytometry measurements of cell concentrations before and after each round of selection), calculate the selection coefficient by subtracting the median of normalized log2 fold-change values obtained for variants encoding the wild-type amino acid sequence, except for the actual wild-type nucleotide sequence. The reason why the wild-type nucleotide sequence is omitted from this normalization is because the oPool design overrepresents the wild-type in pools of variants. A Pearson correlation coefficient was calculated for pairwise comparisons of replicates. Replicates were well correlated ($\rho >= 0.88$, S8 Fig). For each nucleotide variant, we calculated the median selection coefficient (across replicates) and for each amino acid variant, the median selection coefficient (across synonymous variants). In downstream analyses, we only kept "high confidence variants", i.e. variants with a confidence score of 1 or 2 (at least 4 reads in at least 1 input replicate).

Classification of mutational effects

The multimodal distribution of selection coefficients for each condition was used to classify variants, using a Gaussian mixture model (implemented in Python with the scikit-learn package). For each combination of mutated hotspot and compound (drug or control condition), we first optimized the model by calculating the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC) for up to five components (S9 Fig). We then ran the model using the number of components that minimized the BIC: 3. Based on the mean of each distribution, we converted the labels predicted by the model to biologically relevant effects on fitness (either deleterious, slightly deleterious or WT-like in the absence of drug i.e. effect on protein function, or WT-like, intermediate or resistant in the presence of drug i.e. effect on resistance). To resolve overlaps between distributions, we set thresholds 3 standard deviations above and below the mean of the narrowest distributions (more details in S9 Fig). Finally, we collapsed bins to simplify the classification: variants were considered either sensitive or resistant (non-sensitive) to a drug and either deleterious or non-deleterious (initially classified as slightly deleterious or WT-like) in the absence of drug.

Random forest classifier

The random forest classifier was implemented in Python with the scikit-learn package. The model uses the 57 Expasy Protscale amino acid profiles [58] and the propensity scale from [59] as properties, for each position in the hotspots. Properties were imported in January 2019 and concatenated into a single csv file. The combinations of properties and position were used as features. Classification of each amino acid hotspot sequence as either sensitive or resistant provided the labels. To train the model, we first split all single FKS1-HS1 mutants with the proportions 70%-30%, and optimized the model until performances were adequate, meaning we tried to maximize the balanced accuracy score, minimize the interval between the score on the training set and the score on the testing set, while trying as much as possible to not overtrain the model (which would typically be indicated by a balanced accuracy score on training of 100%). In order to optimize the model, each of the following hyper-parameters were tuned for each echinocandin separately by grid search cross-validation: 'bootstrap', 'class_weight', 'max_depth', 'max_features', 'min_samples_leaf', 'min_samples_split' and 'n_estimators'. Balanced accuracy was used as the grid search's scoring function to account for slightly imbalanced datasets being fed to the model. SHAP values (computed in Python with the shap package) were plotted to visualize the most contributing features (S13 Fig). One of the identified limitations of the model is that many amino acid properties correlate with each other. In future studies, we recommend performing a dimension reduction analysis first, then appropriately name each group to ensure explainability.

Validations of mutational effects

Two validation assays were performed similarly: the first one with 28 single mutants of Fks1-HS1 to validate the DMS and the second one to validate the docking, using mutants outside the hotspots (and D691V in hotspot 3), as well as F639S, F639D and R645D in hotspot 1 from the first assay to be used as controls. We also included the parental strain BY4741 and strains in which the wild-type sequence was reinserted in the appropriate deletion background. Strains were grown overnight in 1 mL of YPD in a 96-deep-well plate at 30°C with shaking. Precultures were diluted to 1 OD_{600} in sterile water and inoculated at a final concentration of 0.1 OD_{600} in sterile Greiner 96-well plates with SC medium supplemented with either sterile water or echinocandin. For the assay designed to validate the DMS assay, we used 0.058 µg/mL caspofungin or 0.44 µg/mL micafungin or 0.63 µg/mL anidulafungin. The concentrations were chosen because they correspond to 80% inhibition based on our inhibition curves (S14 Fig), which is approximately what we estimated was the actual inhibition coefficient for the DMS screening (82 %). For the assay designed to validate the docking, we used concentrations that completely inhibit WT growth to ensure discrimination of resistant vs sensitive mutants: 0.232 µg/mL caspofungin or 1.76 µg/mL micafungin or 2.52 µg/mL anidulafungin. OD_{600} measurements were performed at 30°C every 15 min until a plateau was reached in a BioTek

Epoch 2 Microplate Spectrophotometer (Agilent). For each mutant, we calculated the area under the curve (AUC) using the composite trapezoidal rule (implemented with the trapezoid function from the Python package numpy) and converted it to a log2 fold-change, normalizing with the value obtained with the parental strain. The recreated wild-type control had a fitness effect similar to that of the parental strain, confirming that the observed fitness effect is largely a consequence of the introduced mutation, rather than that of one or several putative off-target mutations.

Correlation between the obtained relative fitness values and the selection coefficients from the DMS assay was evaluated by calculating Spearman's rank correlation coefficient (S10 Fig). The level of micafungin resistance of only two mutants (V641W and L642K) was underestimated in the DMS experiment. For both, a corrected DMS score was obtained by linear regression, which allowed us to re-classify the variants appropriately.

Molecular modeling

Fks1 structure preparation

The 3.4 Å cryo-EM structure of the Fks1 protein (PDB ID: 7XE4 [12]), was used. The ten missing loop segments (residues 244-278, 475-487, 799-805, 897-931, 1159-1167, 1247-1266, 1419-1435, 1516-1554, 1627-1637, and 1698-1723) were reconstructed using the GalaxyFill tool [60] within the CHARMM-GUI framework [61]. The reconstructed Fks1 was validated using the Molecular Operating Environment's (MOE) structure preparation tool [62], using the CHARMM27 force-field. The protonation states of residues were determined at pH 7.0, employing the APBS and PDB2PQR tools [63] under default CHARMM force-field parameters.

Fks1-membrane system construction

The position of the resulting Fks1 in the membrane was predicted using the PPM 3.0 (N-ter: in; fungi plasma membrane) server [64]. The Membrane Builder tool from the CHARMM-GUI server [65] was used to construct the protein-membrane system. This preparation incorporated post-translational modifications as specified by the cryo-EM structure article [12], including termini truncation with acetylation (ACE) and amidation (CT3) at the N- and C-termini, respectively, the disulfide bonds (658-669, 1328-1345), and the addition of a di-N-acetylglucosamine moiety (β-D-GlcNAc (1→4) β-D-GlcNAc) at N1849. A lipid bilayer mimicking the asymmetrical fungal plasma membrane's composition was constructed [66], consisting in the following lipid compositions : DYPC, YOPC, POPE, PYPE, YOPE, POPI, POPS, YOPA, ERG, and PI-Ceramide (18:1,16) with (->2) linked D-mannose, and the ratio of 8:3:3:1:3:5:7:3:49:54 for the outer leaflet, and 10:6:10:5:6:20:31:8:4:0 for the inner leaflet, respectively, for a total of 907 lipids. The system was neutralized and the ionic concentration

was 0.15 mM NaCl. The final systems measured approximately 173 x 173 x 182 Å on the XYZ dimensions, and water molecules were added to create a 30 Å layer on either side of the membranes along the Z axis, resulting in systems of 507,096 atoms.

Simulation details and parameters

Three independent 100-ns Molecular Dynamics (MD) trajectories were conducted to allow protein relaxation. These simulations were performed with the software NAMD [67], using the CHARMM36m force field parameters [68] and TIP3P waters [69]. Simulations were carried out at 303.15 K under isothermal-isobaric (NPT) ensemble conditions with a 4-fs time step with hydrogen mass repartitioning [70] and periodic boundary conditions. Langevin damping with a coefficient of 1 ps^{-1} was used to maintain constant temperature, while pressure was controlled by a Nosé-Hoover Langevin piston at 1 atm. The length of the bonds between hydrogen and heavy atoms was constrained using SETTLE [71] for water molecules and SHAKE [72] for all other molecules. Cut-off distance for the short-range electrostatics and the Lennard-Jones interactions was 12 Å and a switching function was used from 10 to 12 Å. Nonbonded pair lists were updated every 10 steps. The particle mesh Ewald (PME) method [73] was applied for the calculation of long-range electrostatic interactions at each integration step using a sixth-order interpolation and a grid spacing of ~1 Å. The non-bonded pair lists were updated every ten steps. The CHARMM-GUI Membrane Builder protocol for system equilibration was applied [65]. The last frame of the trajectory with the lowest transmembrane-domain RMSD was selected for SILCS and docking simulations [74].

In addition to the initial simulations, three independent 300-ns MD trajectories were conducted for the Fks1-membrane system. These extended simulations were carried out using a 1-fs time step.

SILCS simulation with pharmacophoric fragments

The selected MD-equilibrated Fks1 structure was inserted into a POPC:cholesterol membrane, at the ratio 9:1, available in the SILCSBio suite pipeline [74], for a combined 2 µs MD simulation in the absence of echinocandins. SILCS-MD was run with the standard set of pharmacophoric fragments to identify cavities and site-specific interactions. To accelerate the diffusion of the fragments through the hotspot cavity, the intracellular domains were excluded from the SILCS trajectories. The protein-membrane-fragments system was built using the standard SILCS command-line protocol for SILCS-setup, GCMC/MD simulation, and the volumetric maps of the various pharmacophores (FragMaps) generation was applied. Residues 437, 712, and 1165 were capped and constraints were applied. The default set of fragments was used, excluding halogenated fragments.

Molecular docking

The structures of the three echinocandins (caspofungin, micafungin, and anidulafungin) were built using their SMILES strings, then protonated and optimized for the CHARMM27 force field before being minimized in MOE. The standard SILCS FragMap-directed docking parameters [74] were used and a 15 Å radius docking grid was set around the geometric center of the hotspot cavity (hotspot residues Cβ atoms). Minimization of the input structures prior to docking was enabled. After 50 replicate runs of the docking simulation, generating each around 300 poses, the top scoring docking poses with the lowest ligand grid free energy (LGFE) values were extracted.

CHARMM force field topologies and parameters for the echinocandins

CHARMM-compatible topology and parameter files for the three echinocandins were obtained from the CHARMM-GUI *Ligand Reader and Modeller* tool [75], by analogy with CHARMM's CGenFF parameters [76]. Parameters associated with a penalty score above 10 were further optimized using the Force Field Toolkit (FFTK) plugin [77] in VMD [78]. For this, only molecular subsets, including high-penalty atoms and their neighboring functional groups, were considered. A single representative molecule per echinocandin was sufficient. The corresponding molecules are provided in S16 Fig.

Following the general CHARMM parameterization protocol [76], partial atomic charges were optimized first, followed by bond and angle parameters, and finally dihedral terms. All quantum mechanical (QM) calculations were performed using Gaussian09 [79], with geometry optimization conducted via energy minimization at the MP2(6-31G*) level of theory.

The PDB, Protein Structure File (PSF), and Residue topology File (RTF) files of each molecular subset, generated with CHARMM-GUI [61], were imported into FFTK. Geometry optimization was performed as described above. Atomic charges for atoms with high penalty scores were refined by fitting to water interaction energies; the resulting partial charges are shown in S2 Table. Bond and angle equilibrium values were adjusted to match the QM-minimized geometries, and force constants were refined using the Hessian matrix. Dihedral potential energy surfaces (PES) were generated from relaxed QM dihedral scans, and dihedral parameters were subsequently optimized using the simulated annealing protocol developed by Guvench and MacKerell [80], as implemented in FFTK [77], with a PES fitting cutoff of 10 kcal/mol. Final bond, angle and dihedral parameters are presented in S3 Table.

Molecular Dynamics simulations of Echinocandin-Fks1-Membrane Complexes

The Membrane Builder tool from the CHARMM-GUI server [65] was used to construct the protein–echinocandin–membrane systems based on the selected docking poses, incorporating the post-translational modifications of Fks1 and a lipid bilayer composition that mimics the asymmetric fungal plasma membrane. For each echinocandin, four independent 100-ns molecular dynamics trajectories were initiated from distinct top-scoring docking poses. The

system exhibiting the most favorable Fks1 hotspot–echinocandin interactions was then selected for an additional 100-ns production run, resulting in extended trajectories of 200 ns. All simulations were performed using the parameters described above, with the exception of the extended runs, which used a 1-fs time step to enhance integration accuracy.

Sequence and structure comparisons of Fks homologs

Sequence comparisons were conducted in MEGA X [81] and inferred by using the maximum likelihood method based on the LG matrix-based model [82]. Fks homologous sequences were aligned and trimmed as described above. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with the superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.7192)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 2.57% sites).

The Fks1 structure from *C. albicans* was retrieved from the AlphaFold database (AFDB accession AF-A0A1D8PCT0-F1). Fks1 structures from *N. glabratus*, and *C. auris* were predicted using AlphaFold v2.3.2. Structural alignment was obtained with US-align [83]. Conservation scores computed with Jalview were used to visualize conservation across the aligned structures [23], using ChimeraX [22].

Softwares, packages and versions

The following Python (v3.11.6) packages were used: matplotlib v3.8.1 [84], numpy v1.26.1 [85], pandas v2.1.2 [86], papermill v2.5.0, scikit-learn v1.3.2 [87], scipy v1.11.4 [88], seaborn v0.13.0 [89], shap v0.45.1 [90] and upsetplot v0.9.0. Variant counts were obtained with DiMSum v1.3 [57]. Molecular models were manipulated in PyMOL v2.5.5 [20] or ChimeraX v1.8 [22].

References

1. Pappas PG, Kauffman CA, Andes DR, Clancy CJ, Marr KA, Ostrosky-Zeichner L, et al. Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America. Clin Infect Dis. 2016;62: e1–50.
2. Chowdhary A, Sharma C, Meis JF. Candida auris: A rapidly emerging cause of hospital-acquired multidrug-resistant fungal infections globally. PLoS Pathog. 2017;13: e1006290.
3. Lee Y, Robbins N, Cowen LE. Molecular mechanisms governing antifungal drug resistance. npj Antimicrobials and Resistance. 2023;1: 1–9.
4. Walker LA, Gow NAR, Munro CA. Fungal echinocandin resistance. Fungal Genet Biol. 2010;47: 117–126.
5. Pfaller MA, Diekema DJ, Turnidge JD, Castanheira M, Jones RN. Twenty years of the SENTRY antifungal surveillance program: Results for Candida species from 1997-2016. Open Forum Infect Dis. 2019;6: S79–S94.
6. Johnson ME, Katiyar SK, Edlind TD. New Fks hot spot for acquired echinocandin resistance in Saccharomyces cerevisiae and its contribution to intrinsic resistance of Scedosporium species. Antimicrob Agents Chemother. 2011;55: 3774–3781.
7. Garcia-Effron G. Rezafungin-Mechanisms of Action, Susceptibility and Resistance: Similarities and Differences with the Other Echinocandins. J Fungi (Basel). 2020;6. doi:10.3390/jof6040262
8. Garcia-Effron G, Park S, Perlin DS. Correlating echinocandin MIC and kinetic inhibition of fks1 mutant glucan synthases for Candida albicans: implications for interpretive breakpoints. Antimicrob Agents Chemother. 2009;53: 112–122.
9. Douglas CM. Fungal β(1,3)-d-glucan synthesis. Med Mycol. 2001;39 Suppl 1: 55–66.
10. Bédard C, Pageau A, Fijarczyk A, Mendoza-Salido D, Alcañiz AJ, Després PC, et al. FungAMR: A comprehensive portrait of antimicrobial resistance mutations in fungi. bioRxiv. 2024. p. 2024.10.07.617009. doi:10.1101/2024.10.07.617009
11. Johnson ME, Edlind TD. Topological and mutational analysis of Saccharomyces cerevisiae Fks1. Eukaryot Cell. 2012;11: 952–960.
12. Hu X, Yang P, Chai C, Liu J, Sun H, Wu Y, et al. Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1. Nature. 2023. doi:10.1038/s41586-023-05856-5
13. Zhao C-R, You Z-L, Chen D-D, Hang J, Wang Z-B, Ji M, et al. Structure of a fungal 1,3-β-glucan synthase. Sci Adv. 2023;9: eadh7820.
14. Li J, Liu H, Li J, Liu J, Dai X, Zhu A, et al. Cryo-EM structure of the β-1,3-glucan synthase FKS1-Rho1 complex. Nat Commun. 2025;16: 2054.
15. Jospe-Kaufman M, Ben-Zeev E, Mottola A, Dukhovny A, Berman J, Carmeli S, et al. Reshaping echinocandin antifungal drugs to circumvent glucan synthase point□mutation□mediated resistance. Angew Chem Weinheim Bergstr Ger. 2024. doi:10.1002/ange.202314728
16. Puumala E, Fallah S, Robbins N, Cowen LE. Advancements and challenges in antifungal therapeutic development. Clin Microbiol Rev. 2024;37: e0014223.
17. Faller CE, Raman EP, MacKerell AD Jr, Guvench O. Site Identification by Ligand Competitive Saturation (SILCS) simulations for fragment-based drug design. Methods Mol Biol. 2015;1289: 75–87.
18. Szymański M, Chmielewska S, Czyżewska U, Malinowska M, Tylicki A. Echinocandins - structure, mechanism of action and use in antifungal therapy. J Enzyme Inhib Med Chem. 2022;37: 876–894.

19. Taft CS, Selitrennikoff CP. Cilofungin inhibition of (1-3)-beta-glucan synthase: the lipophilic side chain is essential for inhibition of enzyme activity. J Antibiot (Tokyo). 1990;43: 433–437.
20. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.5.5. 2015.
21. Spreghini E, Orlando F, Sanguinetti M, Posteraro B, Giannini D, Manso E, et al. Comparative effects of micafungin, caspofungin, and anidulafungin against a difficult-to-treat fungal opportunistic pathogen, Candida glabrata. Antimicrob Agents Chemother. 2012;56: 1215–1222.
22. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30: 70–82.
23. Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci. 1993;9: 745–756.
24. WHO fungal priority pathogens list to guide research, development and public health action. World Health Organization; 25 Oct 2022 [cited 8 Apr 2024]. Available: https://www.who.int/publications/i/item/9789240060241
25. Deupi X, Olivella M, Sanz A, Dölker N, Campillo M, Pardo L. Influence of the g-conformation of Ser and Thr on the structure of transmembrane helices. J Struct Biol. 2010;169: 116–123.
26. Denning DW. Echinocandin antifungal drugs. Lancet. 2003;362: 1142–1151.
27. Mazur P, Morin N, Baginsky W, el-Sherbeini M, Clemas JA, Nielsen JB, et al. Differential expression and function of two homologous subunits of yeast 1,3-beta-D-glucan synthase. Mol Cell Biol. 1995;15: 5671–5681.
28. Bordallo-Cardona MÁ, Escribano P, Marcos-Zambrano LJ, Díaz-García J, de la Pedrosa EG, Cantón R, et al. Low and constant micafungin concentrations may be sufficient to lead to resistance mutations in FKS2 gene of Candida glabrata. Med Mycol. 2018;56: 903–906.
29. Garcia-Effron G, Katiyar SK, Park S, Edlind TD, Perlin DS. A naturally occurring proline-to-alanine amino acid change in Fks1p in Candida parapsilosis, Candida orthopsilosis, and Candida metapsilosis accounts for reduced echinocandin susceptibility. Antimicrob Agents Chemother. 2008;52: 2305–2312.
30. Katiyar SK, Edlind TD. Role for Fks1 in the intrinsic echinocandin resistance of Fusarium solani as evidenced by hybrid expression in Saccharomyces cerevisiae. Antimicrob Agents Chemother. 2009;53: 1772–1778.
31. Kumar V, Huang J, Dong Y, Hao G-F. Targeting Fks1 proteins for novel antifungal drug discovery. Trends Pharmacol Sci. 2024;45: 366–384.
32. Thompson GR 3rd, Soriano A, Honore PM, Bassetti M, Cornely OA, Kollef M, et al. Efficacy and safety of rezafungin and caspofungin in candidaemia and invasive candidiasis: pooled data from two prospective randomised controlled trials. Lancet Infect Dis. 2024;24: 319–328.
33. Logviniuk D, Jaber QZ, Dobrovetsky R, Kozer N, Ksiezopolska E, Gabaldón T, et al. Benzylic Dehydroxylation of Echinocandin Antifungal Drugs Restores Efficacy against Resistance Conferred by Mutated Glucan Synthase. J Am Chem Soc. 2022;144: 5965–5975.
34. Murphy SG, Smith C, Lapierre P, Shea J, Patel K, Halse TA, et al. Direct detection of drug-resistant Mycobacterium tuberculosis using targeted next generation sequencing. Front Public Health. 2023;11: 1206056.
35. Pines G, Fankhauser RG, Eckert CA. Predicting Drug Resistance Using Deep Mutational Scanning. Molecules. 2020;25. doi:10.3390/molecules25092265
36. Bédard C, Gagnon-Arsenault I, Boisvert J, Plante S, Dubé AK, Pageau A, et al. Most azole

antifungal resistance mutations in the drug target provide cross-resistance and carry no intrinsic fitness cost. bioRxiv (In press, Nature Microbiology). 2023. p. 2023.12.13.571438. doi:10.1101/2023.12.13.571438

37. Després PC, Cisneros AF, Alexander EMM, Sonigara R, Gagné-Thivierge C, Dubé AK, et al. Asymmetrical dose responses shape the evolutionary trade-off between antifungal resistance and nutrient use. Nat Ecol Evol. 2022;6: 1501–1515.
38. Maligie MA, Selitrennikoff CP. Cryptococcus neoformans resistance to echinocandins: (1,3)beta-glucan synthase activity is sensitive to echinocandins. Antimicrob Agents Chemother. 2005;49: 2851–2856.
39. Hori Y, Shibuya K. Role of FKS gene in the susceptibility of pathogenic fungi to echinocandins. Med Mycol J. 2018;59: E31–E40.
40. Cao C, Wang Y, Husain S, Soteropoulos P, Xue C. A mechanosensitive channel governs lipid flippase-mediated echinocandin resistance in Cryptococcus neoformans. MBio. 2019;10. doi:10.1128/mbio.01952-19
41. Kalem MC, Subbiah H, Leipheimer J, Glazier VE, Panepinto JC. Puf4 mediates post-transcriptional regulation of cell wall biosynthesis and caspofungin resistance in Cryptococcus neoformans. MBio. 2021;12. doi:10.1128/mBio.03225-20
42. Lackner M, Tscherner M, Schaller M, Kuchler K, Mair C, Sartori B, et al. Positions and numbers of FKS mutations in Candida albicans selectively influence in vitro and in vivo susceptibilities to echinocandin treatment. Antimicrob Agents Chemother. 2014;58: 3626–3635.
43. Staab JF, Neofytos D, Rhee P, Jiménez-Ortigosa C, Zhang SX, Perlin DS, et al. Target enzyme mutations confer differential echinocandin susceptibilities in Candida kefyr. Antimicrob Agents Chemother. 2014;58: 5421–5427.
44. Zajac C, Scott NE, Kline S, Erayil SE, Selmecki A. Hotspot gene conversion between FKS1 and FKS2 in echinocandin resistant Candida glabrata serial isolates. NPJ Antimicrob Resist. 2025;3: 31.
45. Lõoke M, Kristjuhan K, Kristjuhan A. Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques. 2011;50: 325–328.
46. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng W-T, Zhang W, et al. Exploration of essential gene functions via titratable promoter alleles. Cell. 2004;118: 31–44.
47. Ryan OW, Poddar S, Cate JHD. CRISPR-Cas9 Genome Engineering in Saccharomyces cerevisiae Cells. Cold Spring Harb Protoc. 2016;2016. doi:10.1101/pdb.prot086827
48. Alexander BD, Johnson MD, Pfeiffer CD, Jiménez-Ortigosa C, Catania J, Booker R, et al. Increasing echinocandin resistance in Candida glabrata: clinical failure correlates with presence of FKS mutations and elevated minimum inhibitory concentrations. Clin Infect Dis. 2013;56: 1724–1732.
49. Pham CD, Iqbal N, Bolden CB, Kuykendall RJ, Harrison LH, Farley MM, et al. Role of FKS Mutations in Candida glabrata: MIC values, echinocandin resistance, and multidrug resistance. Antimicrob Agents Chemother. 2014;58: 4690–4696.
50. Garcia-Effron Guillermo, Lee Samuel, Park Steven, Cleary John D., Perlin David S. Effect of Candida glabrata FKS1 and FKS2 Mutations on Echinocandin Sensitivity and Kinetics of 1,3-β-d-Glucan Synthase: Implication for the Existing Susceptibility Breakpoint. Antimicrob Agents Chemother. 2009;53: 3690–3699.
51. Pryszcz LP, Huerta-Cepas J, Gabaldón T. MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. Nucleic Acids Res. 2011;39: e32.
52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high

throughput. Nucleic Acids Res. 2004;32: 1792–1797.
53. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973.
54. Amberg DC, Burke DJ, Strathern JN. Yeast DNA Isolations, Techniques and Protocols 3. Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press; 2005.
55. Yachie N, Petsalaki E, Mellor JC, Weile J, Jacob Y, Verby M, et al. Pooled-matrix protein interaction screens using Barcode Fusion Genetics. Mol Syst Biol. 2016;12: 863.
56. Cota JM, Grabinski JL, Talbert RL, Burgess DS, Rogers PD, Edlind TD, et al. Increases in SLT2 expression and chitin content are associated with incomplete killing of Candida glabrata by caspofungin. Antimicrob Agents Chemother. 2008;52: 1144–1146.
57. Faure AJ, Schmiedel JM, Baeza-Centurion P, Lehner B. DiMSum: an error model and pipeline for analyzing deep mutational scanning data and diagnosing common experimental pathologies. Genome Biol. 2020;21: 207.
58. Gasteiger E, Hoogland C, Gattiker A, Duvaud S ’everine, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. pp. 571–607.
59. Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci U S A. 2012;109: 20461–20466.
60. Coutsias EA, Seok C, Jacobson MP, Dill KA. A kinematic view of loop closure. J Comput Chem. 2004;25: 510–528.
61. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29: 1859–1865.
62. 2022. 02 Chemical Computing Group ULC. Molecular Operating Environment (MOE). 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7; 2024.
63. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, et al. Improvements to the APBS biomolecular solvation software suite. Protein Sci. 2018;27: 112–128.
64. Lomize AL, Todd SC, Pogozheva ID. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. Protein Sci. 2022;31: 209–220.
65. Wu EL, Cheng X, Jo S, Rui H, Song KC, Dávila-Contreras EM, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem. 2014;35: 1997–2004.
66. Pogozheva ID, Armstrong GA, Kong L, Hartnagel TJ, Carpino CA, Gee SE, et al. Comparative Molecular Dynamics Simulation Studies of Realistic Eukaryotic, Prokaryotic, and Archaeal Membranes. J Chem Inf Model. 2022;62: 1036–1051.
67. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys. 2020;153: 044130.
68. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14: 71–73.
69. Durell SR, Brooks BR, Ben-Naim A. Solvent-Induced Forces between Two Hydrophilic Groups. J Phys Chem. 1994;98: 2198–2202.
70. Gao Y, Lee J, Smith IPS, Lee H, Kim S, Qi Y, et al. CHARMM-GUI Supports Hydrogen Mass Repartitioning and Different Protonation States of Phosphates in Lipopolysaccharides. J Chem Inf Model. 2021;61: 831–839.
71. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem. 1992;13: 952–962.

72. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977;23: 327–341.
73. Darden T, York D, Pedersen L. Particle mesh Ewald: An N□log(N) method for Ewald sums in large systems. J Chem Phys. 1993;98: 10089–10092.
74. MacKerell AD Jr, Jo S, Lakkaraju SK, Lind C, Yu W. Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). Biochim Biophys Acta Gen Subj. 2020;1864: 129519.
75. Kim S, Lee J, Jo S, Brooks CL 3rd, Lee HS, Im W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. J Comput Chem. 2017;38: 1879–1886.
76. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem. 2010;31: 671–690.
77. Mayne CG, Saam J, Schulten K, Tajkhorshid E, Gumbart JC. Rapid parameterization of small molecules using the Force Field Toolkit. J Comput Chem. 2013;34: 2757–2770.
78. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14: 33–8, 27–8.
79. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT. Gaussian 09, Revision A.02. 2016.
80. Guvench O, MacKerell AD Jr. Automated conformational energy fitting for force-field development. J Mol Model. 2008;14: 667–679.
81. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35: 1547–1549.
82. Le SQ, Gascuel O. An Improved General Amino Acid Replacement Matrix. Mol Biol Evol. 2008;25: 1307–1320.
83. Zhang C, Shine M, Pyle AM, Zhang Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. Nat Methods. 2022;19: 1109–1115.
84. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9: 90–95.
85. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585: 357–362.
86. McKinney W. Data Structures for Statistical Computing in Python. In: van der Walt S, Millman J, editors. Proceedings of the 9th Python in Science Conference. SciPy; 2010. doi:10.25080/majora-92bf1922-00a
87. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;abs/1201.0490. doi:10.5555/1953048.2078195

88. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17: 261–272.
89. Waskom M. seaborn: statistical data visualization. J Open Source Softw. 2021;6: 3021.
90. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.; 2017. pp. 4765–4774.

Acknowledgments: We thank Anna Fijarczyk, François D. Rouleau and Stéphane Larose for bioinformatic assistance, Jordan Jalbert-Ross with technical assistance on preliminary assays, Philippe C. Després, François D. Rouleau and Soham Dibyachintan for critical reading of the manuscript and the whole team for feedback and emotional support on a regular basis throughout the project.

Funding:

Canadian Institutes of Health Research (CIHR) (Foundation grant 387697, CRL)

Genome Canada and Genome Quebec (grant 6569, CRL)

Canada Research Chair in Cellular Systems and Synthetic Biology (CRL)

Natural Sciences and Engineering Research Council of Canada (NSERC) (Grant RGPIN-2022-04721, PL)

Fonds de Recherche du Québec - Santé (FRQS) postdoctoral fellowship (RD)

NSERC postdoctoral fellowship (RD)

Author contributions:

Conceptualization: RD, AGT, PL, CRL

Methodology: RD, AGT, MG, AP, AKD, PL, CRL

Software: RD, AGT, MG, AP, PL

Validation: RD, MG, AP, AKD, PL

Formal analysis: RD, AGT, MG, AP, PL

Investigation: RD, AGT, MG, AP, AKD, PL

Resources: PL, CRL

Data curation: RD, MG, AP

Writing - Original draft: RD, AGT, MG

Writing - Review and editing: RD, AGT, PL, CRL

Visualization: RD, AGT, MG, AP, PL

Supervision: RD, AGT, AP, AKD, PL, CRL

Project administration: PL, CRL

Funding acquisition: PL, CRL

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials. Sequencing data are available at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1126009: https://www.ncbi.nlm.nih.gov/sra/PRJNA1126009. All data and code to reproduce results and figures for the DMS and the machine learning sections are available at https://github.com/Landrylab/Durand_et_al_2024. All strains listed in S4 Data are available and can be requested by email at christian.landry@bcm.ulaval.ca.

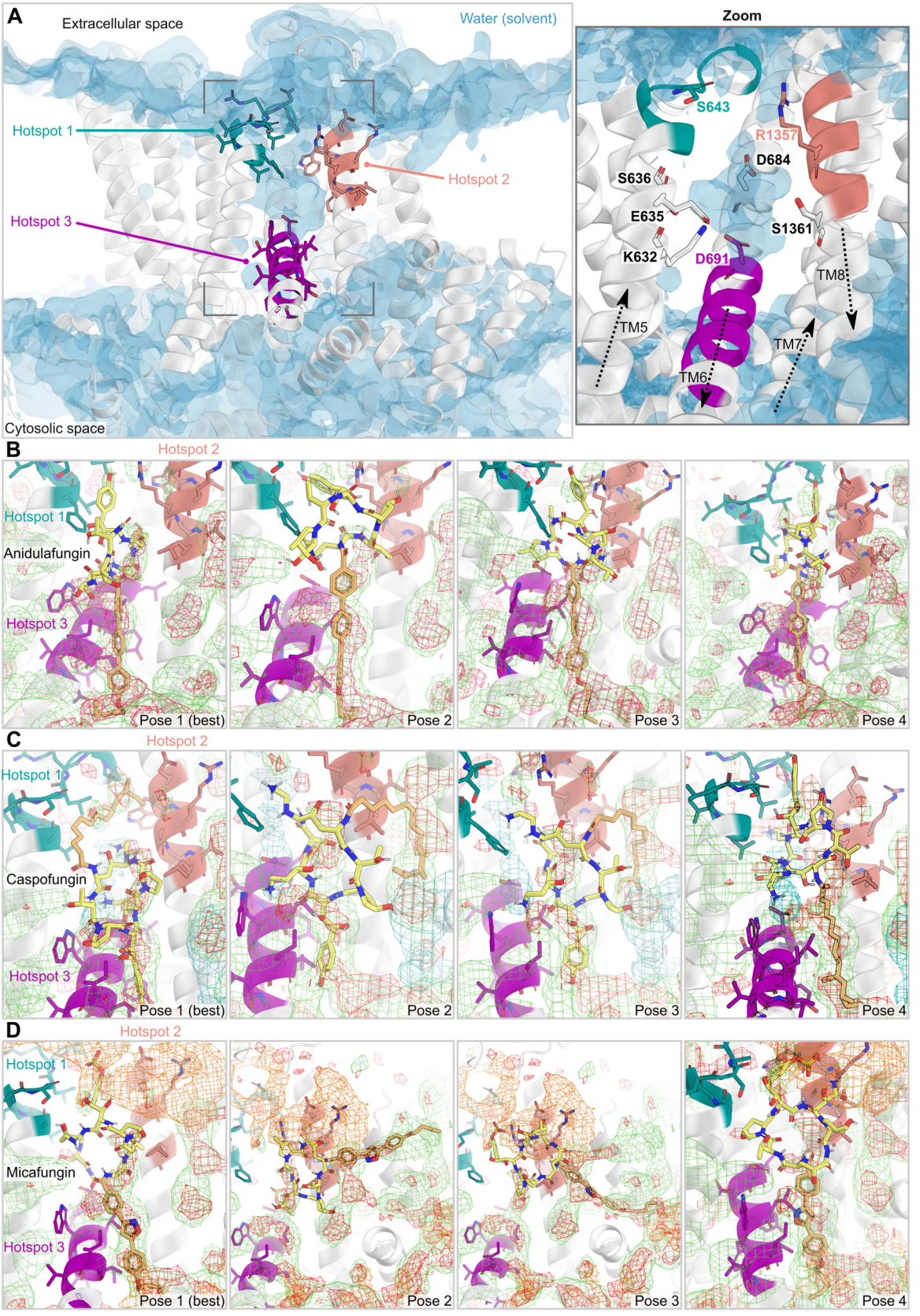
Supplementary Materials

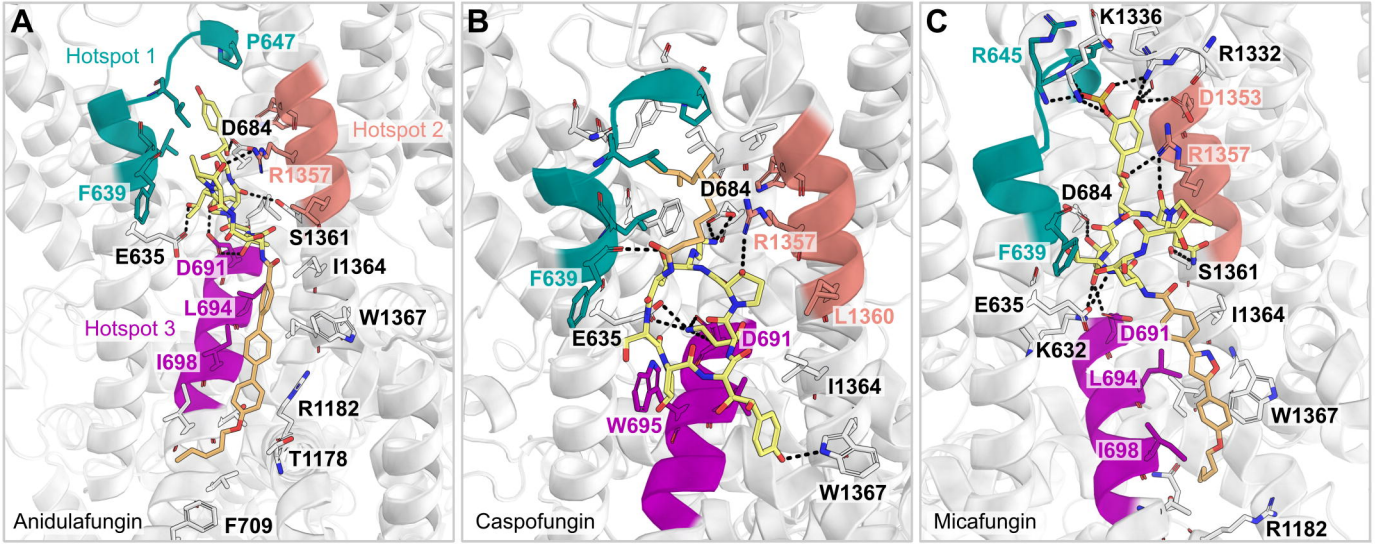
S1 to S16 Figures

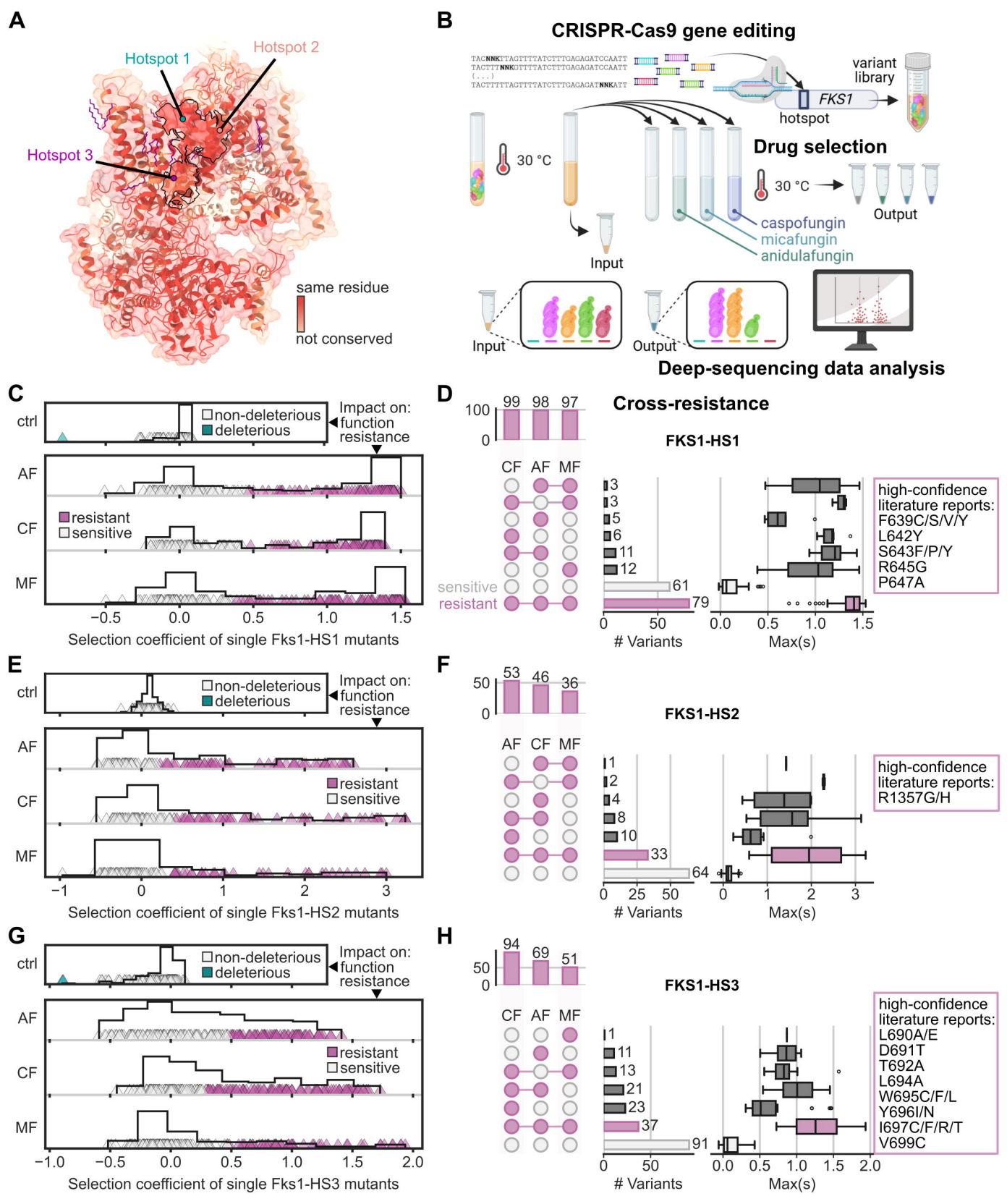
S1 to S4 Tables

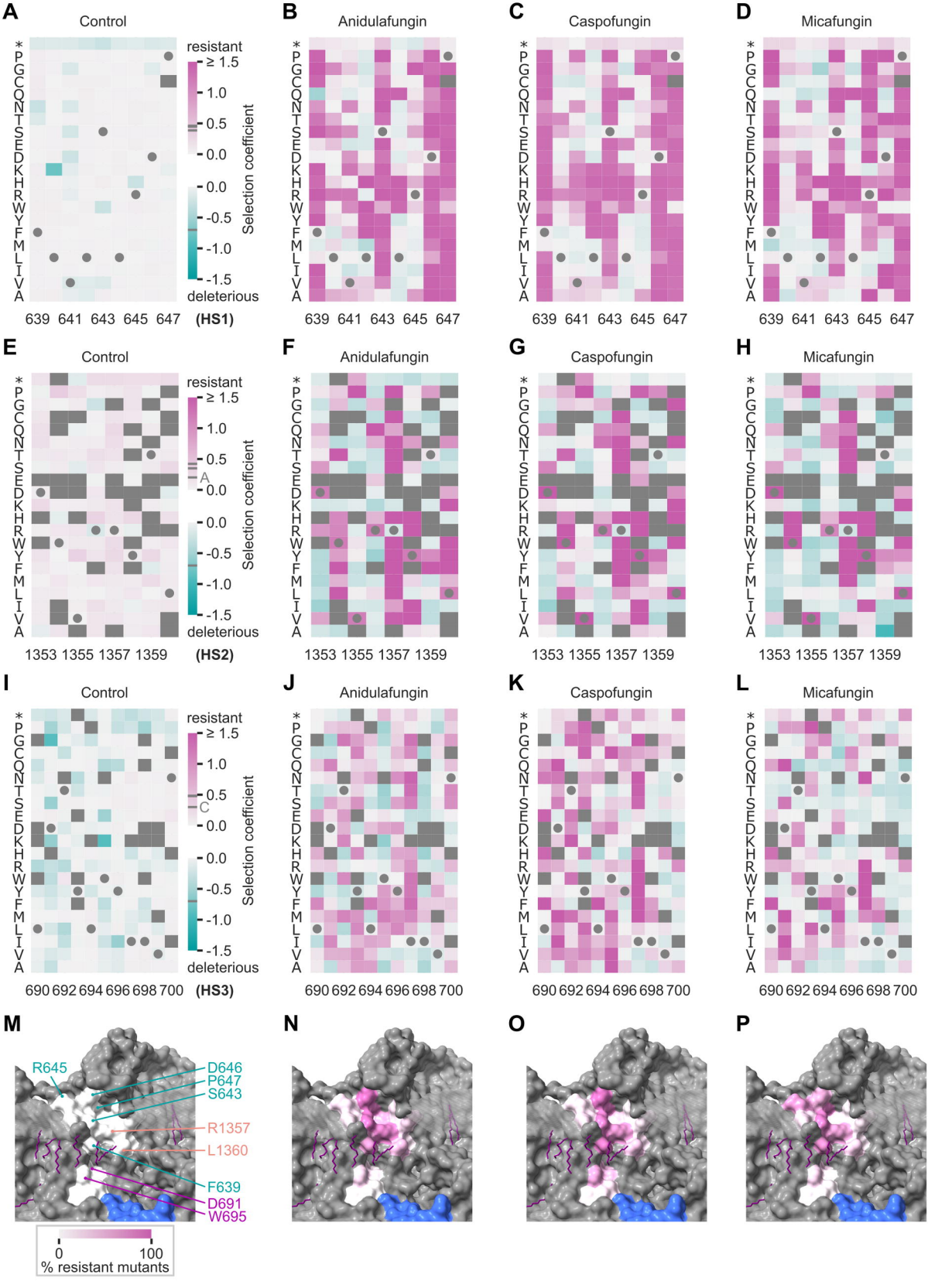
S1 to S4 Data

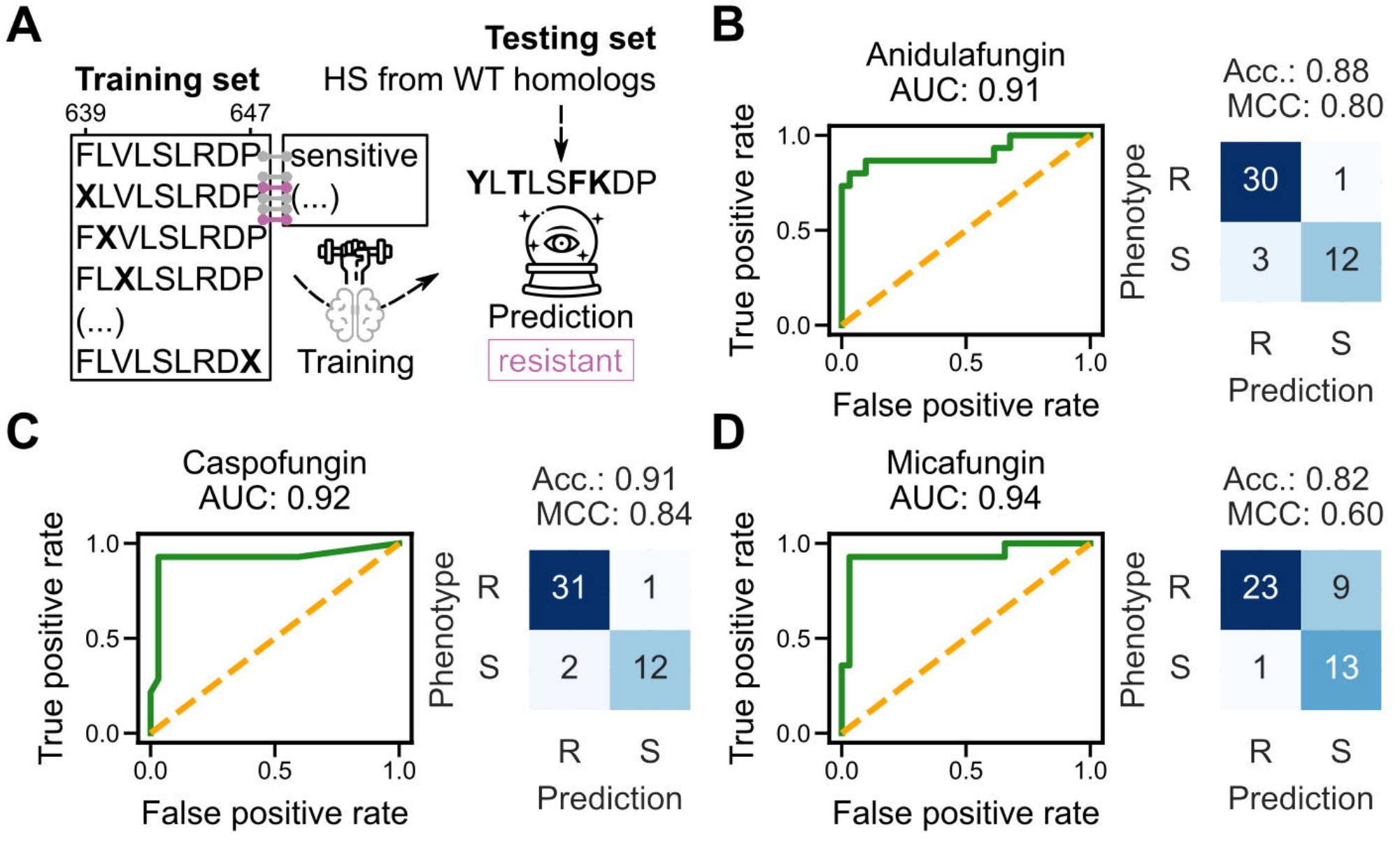
S1 to S3 Compressed files


A
Extracellular space
Water (solvent)
Hotspot 1
Hotspot 2
Hotspot 3
Cytosolic space
Zoom
S643
R1357
D684
S636
E635
S1361
K632
D691
TM8
TM5
TM7
TM6
B
Hotspot 2
Hotspot 1
Anidulafungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4
C
Hotspot 2
Hotspot 1
Caspofungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4
D
Hotspot 2
Hotspot 1
Micafungin
Hotspot 3
Pose 1 (best)
Pose 2
Pose 3
Pose 4


A
Hotspot 1
P647
D684
Hotspot 2
R1357
F639
S1361
E635
D691
I1364
L694
W1367
Hotspot 3
I698
R1182
T1178
Anidulafungin
F709
B
D684
R1357
F639
L1360
E635
D691
I1364
W695
W1367
Caspofungin
C
K1336
R645
R1332
D1353
R1357
D684
F639
S1361
E635
I1364
K632
D691
L694
W1367
I698
Micafungin
R1182


A
Hotspot 1
Hotspot 2
Hotspot 3
same residue
not conserved
B
CRISPR-Cas9 gene editing
variant library
FKS1
hotspot
30 °C
Drug selection
30 °C
Output
Input
caspofungin
micafungin
anidulafungin
Input
Output
Deep-sequencing data analysis
C
ctrl
non-deleterious
deleterious
Impact on: function resistance
AF
CF
MF
resistant
sensitive
Selection coefficient of single Fks1-HS1 mutants
D
Cross-resistance
FKS1-HS1
99 98 97
CF AF MF
sensitive
resistant
# Variants
Max(s)
high-confidence literature reports:
F639C/S/V/Y
L642Y
S643F/P/Y
R645G
P647A
E
Selection coefficient of single Fks1-HS2 mutants
F
FKS1-HS2
53 46 36
AF CF MF
high-confidence literature reports:
R1357G/H
G
Selection coefficient of single Fks1-HS3 mutants
H
FKS1-HS3
94 69 51
CF AF MF
high-confidence literature reports:
L690A/E
D691T
T692A
L694A
W695C/F/L
Y696I/N
I697C/F/R/T
V699C


A
Control
B
Anidulafungin
C
Caspofungin
D
Micafungin
resistant
≥ 1.5
1.0
0.5
0.0
-0.0
-0.5
-1.0
-1.5
deleterious
Selection coefficient
639 641 643 645 647
(HS1)
E
Control
F
Anidulafungin
G
Caspofungin
H
Micafungin
1353 1355 1357 1359
(HS2)
I
Control
J
Anidulafungin
K
Caspofungin
L
Micafungin
690 692 694 696 698 700
(HS3)
M
N
O
P
R645
D646
P647
S643
R1357
L1360
F639
D691
W695
0
100
% resistant mutants


A
Training set
639
647
FLVLSLRDP
XLVLSLRDP
FXVLSLRDP
FLXLSLRDP
(...)
FLVLSLRDX
sensitive
(...)
Training
Testing set
HS from WT homologs
YLTLSFKDP
Prediction
resistant
B
Anidulafungin
AUC: 0.91
True positive rate
False positive rate
Acc.: 0.88
MCC: 0.80
Phenotype
R
S
30
1
3
12
Prediction
C
Caspofungin
AUC: 0.92
Acc.: 0.91
MCC: 0.84
31
1
2
12
D
Micafungin
AUC: 0.94
Acc.: 0.82
MCC: 0.60
23
9
1
13