



**UNIVERSIDADE DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO**

***In silico* cross-species approach for the discovery of new  
antifungal drug targets and new antifungal drugs**

**Romeu Filipe Nobre Viana**

**Supervisor:** Doctor Miguel Nobre Parreira Cacho Teixeira

**Thesis approved in public session to obtain the PhD Degree  
in Biotechnology and Biosciences**

**Jury final classification: Pass with Distinction and Honour**





**UNIVERSIDADE DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO**

***In silico* cross-species approach for the discovery of new  
antifungal drug targets and new antifungal drugs**

**Romeu Filipe Nobre Viana**

**Supervisor:** Doctor Miguel Nobre Parreira Cacho Teixeira

**Thesis approved in public session to obtain the PhD Degree in  
Biotechnology and Biosciences**

**Jury final classification: Pass with Distinction and Honour**

**Jury**

**Chairperson:** Doctor Arsénio do Carmo Sales Mendes Fialho, Instituto Superior Técnico, Universidade de Lisboa

**Members of the Committee:**

**Doctor** Cláudio Manuel Simões Loureiro Nunes Soares, ITQB NOVA - Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa

**Doctor** Arsénio do Carmo Sales Mendes Fialho, Instituto Superior Técnico, Universidade de Lisboa

**Doctor** Miguel Nobre Parreira Cacho Teixeira, Instituto Superior Técnico, Universidade de Lisboa

**Doctor** Teresa Maria Fonseca de Oliveira Gonçalves, Instituto de Microbiologia, Faculdade de Medicina, Universidade de Coimbra

**Doctor** Nuno Gonçalo Pereira Mira, Instituto Superior Técnico, Universidade de Lisboa

**Doctor** Oscar Manuel Lima Dias, Centre of Biological Engineering, Universidade do Minho

**Funding Institution:** FCT – Fundação para a Ciência e a Tecnologia

**2024**



# Abstract

There is an urgent need for new antifungal therapeutic approaches due to limited efficacy, increase in multidrug resistance, and high mortality rates associated with invasive fungal infections, particularly involving *Candida* and *Cryptococcus* species. However, the process of developing new drugs remains a difficult and costly task, as evidenced by the reduced number of new classes of antifungals that have emerged in recent decades. Genome-scale metabolic models (GSMMs) and virtual drug screening are emerging as attractive *in silico* approaches to accelerate and innovate drug discovery. In this thesis, GSMMs were reconstructed for four pathogenic invasive yeast species, namely *Candida albicans*, *Candida parapsilosis*, *Candida auris*, and *Cryptococcus neoformans var grubii*. These GSMMs were used to identify new potential drug targets through enzyme/gene essentiality prediction and to elucidate the complex metabolic mechanisms that enable these yeasts to thrive in the human host. The reconstructed models were experimentally validated, being able to accurately predict the compounds that can be used as sole carbon and nitrogen sources, as well as growth parameters such as specific growth rate. The reconstructed models enabled the identification of unique metabolic features for each pathogen, mainly related to alternative assimilation pathways, host adaptation, virulence mechanisms, or drug/stress resistance. Moreover, 47 fungal drug targets common to all analyzed yeast species were identified, including those already targeted by clinically used antifungals and others that represent potentially new drug targets that have been vaguely explored in fungi. *In silico* virtual drug screening from large compound databases was performed for two of the most promising targets, namely chitin synthase and dihydropteroate synthase, in order to identify potential inhibitors among compounds exhibiting affinity to the active site of those targets. A vast list of compounds with predicted high affinity

was identified, and some were tested for antifungal activity, revealing 5 compounds with moderate activity against *Candida* or *Cryptococcus* species and a potential synergistic effect with fluconazole for one of them. However, further studies are needed to understand their mechanisms of action, interactions with host cells and other drugs, and to optimize their efficacy and safety. Altogether, this thesis provides valuable insights into fungal pathogenesis and drug discovery with the potential to have an impact on public health by addressing the challenges posed by fungal infections and drug resistance.

**Keywords:** GSMM's, Virtual drug screening, Pathogenic Fungi, Drug targets, Drug discovery

# Resumo

Existe uma necessidade urgente de novas abordagens terapêuticas antifúngicas devido à eficácia limitada das utilizadas correntemente, ao aumento da resistência a múltiplos fármacos e às elevadas taxas de mortalidade associadas a infecções fúngicas invasivas, especialmente envolvendo espécies de *Candida* e *Cryptococcus*. Contudo, o processo de desenvolvimento de novos fármacos permanece uma tarefa difícil e dispendiosa, como é evidenciado pelo reduzido número de novas classes de antifúngicos que emergiram nas últimas décadas. Os modelos metabólicos à escala genómica (GSMMs) e a triagem virtual de fármacos estão a emergir como abordagens *in silico* atrativas para acelerar e inovar a descoberta de novos fármacos. Nesta tese, foram reconstruídos GSMMs para quatro espécies patogénicas invasivas de leveduras, nomeadamente *Candida albicans*, *Candida parapsilosis*, *Candida auris* e *Cryptococcus neoformans var grubii*. Estes GSMMs foram usados para identificar novos potenciais alvos terapêuticos através da previsão da essencialidade de enzimas/genes e para elucidar os complexos mecanismos metabólicos que permitem a estas leveduras proliferar no hospedeiro humano. Os modelos reconstruídos foram validados experimentalmente, sendo capazes de prever com precisão compostos que podem ser usados como fontes únicas de carbono e de azoto, bem como parâmetros de crescimento, como a taxa de crescimento específica. Os modelos permitiram ainda a identificação de características metabólicas únicas para cada patogénio, principalmente relacionadas com vias de assimilação alternativas, adaptação ao hospedeiro, mecanismos de virulência ou resistência a fármacos/stress. Além disso, foram identificados 47 alvos terapêuticos fúngicos comuns a todas as espécies de leveduras analisadas, incluindo aqueles já visados pelos antifúngicos clinicamente utilizados, e outros que representam potencialmente novos alvos que foram vagamente explorados em fungos. A

triagem virtual *in silico* a partir de grandes bases de dados de compostos foi realizada para dois dos alvos mais promissores, nomeadamente a quitina sintase e dihidropteroato sintase, com o objetivo de identificar potenciais inibidores destas enzimas entre compostos que apresentam elevada afinidade para o seu centro ativo. Foi identificada uma vasta lista de compostos com elevada afinidade prevista, e alguns foram testados quanto à atividade antifúngica, levando à identificação de 5 compostos com atividade moderada contra espécies de *Candida* ou *Cryptococcus* e em um deles, um potencial efeito sinérgico com fluconazol. No entanto, são necessários mais estudos para compreender os seus mecanismos de ação, interações com as células hospedeiras e outros fármacos, e para otimizar a sua eficácia e segurança. Globalmente, esta tese fornece perspectivas valiosas sobre a patogénese fúngica e a descoberta de fármacos, com potencial impacto na saúde pública ao abordar os desafios colocados pelas infeções fúngicas e pela resistência a fármacos.

**Palavras-Chave:** GSMMs, Triagem virtual de fármacos, Fungos patogénicos, Alvos de fármacos, Descoberta de fármacos

# Acknowledgments

I would like to acknowledge my supervisor Professor Miguel Teixeira for giving me the opportunity to carry out my PhD work under his guidance. His guidance, motivation, expertise, creativity, and support were fundamental for my personal and professional growth and for the development of this thesis.

Secondly, I must acknowledge Professor Isabel Sá-Correia and Professor Arsénio Fialho as the former and current head of the Biological Sciences Research Group (BSRG), respectively, for the opportunity to work with such an incredible team. I also acknowledge Professor Margarida Casal as the coordinator of the Doctoral Program in Applied and Environmental Microbiology (DP\_AEM) for the opportunity to join such a dynamic program and collaborative community.

I acknowledge the collaboration of all co-authors of the papers and manuscripts presented in this thesis, particularly Dr. Oscar Dias and his team at CEB from Universidade do Minho and Dr. Claudio Soares and all his team from ITQB NOVA who have received me so well, and for the opportunity to learn from their expertise and expand my scientific knowledge. A special thanks also to Dr. Isabel Rocha for serving as a connection bridge between the two research teams and for his guidance. I would also like to thank Dr. Carolina Coelho for the collaboration with the University of Exeter in the reconstruction of the GSMM for *C. neoformans*; to Dr Luis Schang from the University of Cornell for the collaboration with compound experimental testing, and for kindly providing some of the compounds from their library; and to Dr. Nuno Bernardes for the assistance provided in the cytotoxicity assays.

I acknowledge “Fundação para a Ciência e a Tecnologia” (FCT) PhD grant PD/BD/150586/2020 attributed to the candidate under the framework of the DP\_AEM PhD program in Biotechnology and Biosciences. I further acknowledge

funding from FCT [Contract PTDC/BII-BIO/28216/2017] and Programa Operacional Regional de Lisboa 2020 [LISBOA-01-0145-FEDER-022231], through the Biodata.pt Research Infrastructure. National funds from FCT in the scope of the project UIDB/04565/2020 and UIDP/04565/2020 of the Research Unit Institute for Bioengineering and Biosciences—iBB, project UIDB/04469/2020 for the Centre of Biological Engineering—CEB, and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB, are also acknowledged.

The following personal acknowledgments will be addressed in Portuguese:

Gostaria de deixar um obrigado especial a todos os membros do grupo MCTeixeira'sLab em especial à Mafalda e ao Pedro por toda a ajuda e paciência nos meus primeiros anos no grupo, mas também à Mónica e Inês pela sua ótima companhia, ideias criativas e toda a ajuda prestada e paciência durante toda esta jornada. Um grande agradecimento também a todos os restantes elementos do grupo BSRG que de alguma forma contribuíram para o meu trabalho, em especial para a Mónica Rato, que foi crucial para a realização de todo este trabalho.

Por fim, gostaria de agradecer a todos os que me apoiaram fora do BSRG e que sempre acreditaram no sucesso desta etapa. Um obrigado à Mafalda por todo o apoio incansável e por acreditar sempre em mim. Um obrigado aos meus pais, à Cláudia e ao Duarte por todo o amor incondicional e apoio durante estes anos. E finalmente um obrigado aos meus avós por todos os valores passados ao longo da minha vida e por serem a minha inspiração. Um obrigado também à Carolina, à Sylvie, ao Zé e à Júlia e a todos os restantes amigos por todos os momentos partilhados ao longo de todo o meu percurso académico, um grande agradecimento a todas as pessoas da RUCA e RURS.

# Contents

<b>ABSTRACT</b> .....	<b>V</b>
<b>RESUMO</b> .....	<b>VII</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>IX</b>
<b>CONTENTS</b> .....	<b>XI</b>
<b>LIST OF FIGURES</b> .....	<b>XIV</b>
<b>LIST OF TABLES</b> .....	<b>XIX</b>
<b>LIST OF ACRONYMS</b> .....	<b>XXII</b>
<b>I. INTRODUCTION</b> .....	<b>1</b>
I.1. THESIS OUTLINE .....	2
I.2. INVASIVE FUNGAL INFECTIONS .....	5
<i>I.2.1. Antifungal Frontline: Current treatment options</i> .....	6
<i>I.2.2. Antifungal Frontline: Why are we losing the fight</i> .....	9
<i>I.2.3. Antifungal development - current drug targets</i> .....	11
I.2.3.1. Ergosterol .....	11
I.2.3.2. $\beta$ -1,3-glucan .....	13
I.2.3.3. Other currently targeted pathways in fungi .....	15
<i>I.2.4. The antifungal discovery strategy</i> .....	22
I.3. GENOME-SCALE METABOLIC MODELS .....	24
<i>I.3.1. Current challenges in GSMM reconstruction and exploitation</i> .....	29
I.4. MOLECULAR DOCKING .....	30
<i>I.4.1. Docking methodologies</i> .....	32
<i>I.4.2. Compound and receptor preparation for docking</i> .....	34
<i>I.4.3. Sampling algorithms</i> .....	35
<i>I.4.4. Scoring functions</i> .....	37
<i>I.4.5. Autodock Vina</i> .....	39
<i>I.4.6. Homology Modeling</i> .....	40
<i>I.4.7. AlphaFold</i> .....	43
<i>I.4.8. Application of molecular docking to virtual drug screening</i> .....	45
<i>I.4.9. Current challenges in molecular docking</i> .....	46
<b>II. RECONSTRUCTION OF GENOME-SCALE METABOLIC MODELS FOR PATHOGENIC FUNGI</b> .....	<b>49</b>
II.1. GENOME-SCALE METABOLIC MODEL OF THE HUMAN PATHOGEN <i>C. ALBICANS</i> : A PLATFORM FOR DRUG TARGET PREDICTION .....	50
<i>II.1.1. Abstract</i> .....	50
<i>II.1.2. Introduction</i> .....	51
<i>II.1.3. Materials and Methods</i> .....	52
II.1.3.1. Model development .....	52
II.1.3.2. Genome Annotation and Assembling the Metabolic Network .....	53
II.1.3.3. Reversibility and Balancing .....	54
II.1.3.4. Compartmentalization .....	55
II.1.3.5. Transport reactions .....	55
II.1.3.6. Biomass Equation .....	55
II.1.3.7. Curation of the Model .....	57
II.1.3.8. Strains and growth media .....	57
II.1.3.9. Carbon and nitrogen source utilization assessment .....	57
II.1.3.10. Network simulation and analysis .....	58

<b>II.1.4. Results and Discussion</b> .....	58
II.1.4.1. Model characteristics .....	58
II.1.4.1.1 Gap filling and Model Curation .....	62
II.1.4.1.2 Biomass Equation.....	62
II.1.4.2. Validation of the iRV781 model.....	64
II.1.4.2.1 Carbon and nitrogen source utilization.....	64
II.1.4.2.1 Growth parameters in batch culture .....	67
II.1.4.3. Gene essentiality assessment: a tool for drug target discovery? .....	69
<b>II.1.5. Conclusions</b> .....	73
<b>II.1.6. Acknowledgments</b> .....	75
<b>II.2. A GENOME-SCALE METABOLIC MODEL FOR THE HUMAN PATHOGEN <i>C. PARAPSILOSIS</i> AND EARLY IDENTIFICATION OF PUTATIVE NOVEL ANTIFUNGAL DRUG TARGETS</b> .....	76
<b>II.2.1. Abstract</b> .....	76
<b>II.2.2. Introduction</b> .....	77
<b>II.2.3 Materials and Methods</b> .....	78
II.2.3.1. Model reconstruction .....	78
II.2.3.1.1. Enzyme and Reaction Annotation.....	78
II.2.3.1.2. Correcting Reaction Reversibility, Directionality and Balance .....	79
II.2.3.1.3. Compartmentalization .....	79
II.2.3.1.4. Defining the Biomass Equation .....	80
II.2.3.2. Model Simulations and Enzyme Essentiality Prediction .....	81
II.2.3.3. Model Validation .....	82
II.2.3.3.1. Strains and Growth Media.....	82
II.2.3.3.2. Aerobic Batch Cultivation .....	82
II.2.3.3.3. Cell Density, Dry weight and Metabolite Concentration Assessment.....	83
<b>II.2.4. Results and Discussion</b> .....	83
II.2.4.1. Model Characteristics, highlighting <i>C. parapsilosis</i> unique metabolic features.....	83
II.2.4.2. Model Validation .....	88
II.2.4.2.1. Assessing the Model's Ability to Predict Carbon and Nitrogen Source Usage .....	88
II.2.4.2.2. Assessing the Model's Ability to Quantitatively Predict Growth Parameters .....	90
II.2.4.3. Enzyme essentiality assessment: looking for new drug targets.....	91
<b>II.2.5. Conclusions</b> .....	96
<b>II.2.6. Acknowledgments</b> .....	97
<b>II.3. METABOLIC RECONSTRUCTION OF THE HUMAN PATHOGEN <i>C. AURIS</i>: USING A CROSS-SPECIES APPROACH FOR DRUG TARGET PREDICTION</b> .....	98
<b>II.3.1. Abstract</b> .....	98
<b>II.3.2. Introduction</b> .....	99
<b>II.3.3. Materials and Methods</b> .....	100
II.3.3.1. Model Construction .....	100
II.3.3.1.1. Enzyme and Reaction Annotation.....	101
II.3.3.1.2. Correcting Reaction Reversibility, Directionality, and Balance .....	101
II.3.3.1.3. Compartmentalization .....	102
II.3.3.1.4. Defining the Biomass Equation .....	102
II.3.3.2. Model Simulations and Enzyme Essentiality Prediction .....	103
II.3.3.3. Model Validation .....	104
II.3.3.3.1. Strains and Growth Media.....	104
II.3.3.3.2. Aerobic Batch Cultivation .....	104
II.3.3.3.3. Cell Density, Dry Weight, and Metabolite Concentration Assessment .....	104
II.3.3.3.4. Carbon and Nitrogen Source Utilization Assessment.....	105
<b>II.3.4. Results and Discussion</b> .....	106
II.3.4.1. Model characteristics .....	106
II.3.4.2. Model validation.....	108
II.3.4.2.1 Carbon and nitrogen source utilization.....	108
II.3.4.2.1 Growth parameters in batch culture .....	110
II.3.4.3. <i>C. auris</i> unique metabolic features .....	111
II.3.4.4. Drug target forecast based on gene essentiality prediction: a cross-species evaluation .....	114
<b>II.3.5. Conclusions</b> .....	115

II.3.6. Acknowledgments.....	117
II.4. UNVEILING NEW FEATURES OF THE HUMAN PATHOGEN <i>C. NEOFORMANS</i> THROUGH THE CONSTRUCTION AND EXPLOITATION OF A DEDICATED GENOME-SCALE METABOLIC MODEL.....	118
II.4.1. Abstract.....	118
II.4.2. Introduction .....	119
II.4.3. Materials and Methods.....	121
II.4.3.1. Model Development.....	121
II.4.3.2. Genome Annotation and Assembling of the Metabolic Network.....	122
II.4.3.3. Reversibility, Directionality and Balancing.....	122
II.4.3.4. Compartmentalization and Transport reactions.....	123
II.4.3.5. Biomass Equation .....	123
II.4.3.7 Network simulation and model curation .....	124
II.4.3.8 Model Validation .....	125
II.4.3.8.1. Strains and Growth Media .....	125
II.4.3.8.2. Aerobic Batch Cultivation .....	125
II.4.3.8.3. Cell Density, Dry Weight, and Metabolite Concentration Assessment .....	125
II.4.3.8.4. Network simulation and analysis .....	126
II.4.4. Results and Discussion .....	127
II.4.4.1. Model characteristics .....	127
II.4.4.2. Model validation.....	129
II.4.4.2.1 Carbon and nitrogen source utilization.....	129
II.4.4.2.2 Growth parameters in batch culture .....	130
II.4.4.3. <i>C. neoformans</i> unique metabolic features.....	131
II.4.4.4. Drug target analysis based on gene essentiality prediction.....	138
II.4.5. Conclusions .....	143
<b>III: IDENTIFICATION OF POTENTIAL NEW ANTIFUNGAL DRUGS: FROM VIRTUAL DRUG SCREENING TO EXPERIMENTAL VALIDATION. ....</b>	<b>145</b>
III.1. ABSTRACT.....	146
III.2. INTRODUCTION .....	147
III.3. MATERIALS AND METHODS.....	149
III.3.1. Protein structure selection and modeling .....	149
III.3.3. Compounds .....	150
III.3.4. Strains and growth media.....	152
III.3.5. Minimal inhibitory concentration and Checkerboard assays.....	152
III.3.6. Cytotoxicity assays.....	153
III.3.7. Fluorescence microscopy .....	154
III.4. RESULTS AND DISCUSSION .....	154
III.4.1 Virtual screening for potential inhibitors of <i>Fol1</i> .....	154
III.4.1.1. Modeling <i>Fol1</i> protein structure.....	154
III.4.1.2. <i>Fol1</i> virtual drug screening .....	157
III.4.1.3. Antifungal activity assessment .....	160
III.4.1.4. C5 cytotoxicity assessment.....	162
III.4.2 Virtual screening for potential chitin synthase inhibitors .....	164
III.4.2.1. Virtual drug screening.....	164
III.4.2.2. Antifungal activity assessment and cellular localization .....	167
III.5. CONCLUSIONS .....	170
<b>IV. FINAL DISCUSSION .....</b>	<b>172</b>
<b>V. REFERENCES: .....</b>	<b>180</b>
<b>APPENDIX – SUPPLEMENTARY DATA .....</b>	<b>209</b>

# List of Figures

Figure I.1. Cellular target of the main treatment options currently used in the treatment of invasive fungal infections. Adapted from [5].	8
Figure I.2 - Methodology for the reconstruction of the <i>C. albicans</i> iRV781 metabolic model using <i>merlin</i> . Adapted from [111].	26
Figure I.3 – Virtual Drug screening methodology using molecular docking. 1. Receptor structure selection and preparation for molecular docking, which can be retrieved from a 3D structure available in PDB, or from <i>in silico</i> predicted structures using homology modeling or from the AlphaFold database. 2. Ligand preparation for molecular docking, which can be a group of compounds from a compound database, or specific selected ligand(s) of interest. 3. Searching box design, defining the special restraints within the target structure where the molecular docking calculation will be performed, using one of the different docking programs available. 4. Analysis and selection of the best-hit results based on criteria's such as binding energy calculated by scoring functions. 5. <i>In vivo</i> validation of the predicted activity (e.g. antimicrobial activity) and potential cytotoxic effects.	33
Figure II.1 - Methodology for the reconstruction of the <i>C. albicans</i> iRV781 metabolic model. Adapted from [111].	53
Figure II.2 - Comparison between <i>C. albicans</i> , <i>S. cerevisiae</i> , and <i>C. glabrata</i> proteins with associated EC Numbers present in the genome-scale metabolic models iRV781, iIN800 and iNX804, respectively. Diagram obtained using VENNY2.1 tool [247].	61

Figure II.3 - Utilization of glucose (control), cellobiose, D-Ribose, and mannitol by *C. albicans* reference strain SC5314 as carbon source in solid YNB medium. Initial OD600nm = 0.5 ± 0.05. Growth was assessed after incubation at 37°C for 24h..... 65

Figure II.4 - Comparison between *C. parapsilosis*, *C. albicans*, *S. cerevisiae*, and *C. glabrata* proteins with associated EC Numbers present in the iDC1003, iRV781, iIN800 and iNX804 genome-scale metabolic models, respectively. A: Venn Diagram. B: Pairwise Intersections. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)). ..... 87

Figure II.5 - Intersection of *C. parapsilosis*, *C. albicans*, and *C. glabrata* essential EC Numbers in RPMI medium environmental conditions in the genome-scale metabolic models iDC1003, iRV781, and iNX804, respectively. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)). ..... 93

Figure II.6 - Folate biosynthetic pathway. Red boxes highlight the enzymes considered essential in the three analysed models..... 95

Figure II.7 - Utilization of glucose (positive control), no carbon source (negative control), glycerol, ethanol, and dulcitol as carbon source by *C. auris* strains I: B8441 II: B11220 and III: B11221 in solid YNB medium. Initial OD600nm = 0.5 ± 0.05. Growth was assessed after incubation at 37 °C for 48 h. .... 109

Figure II.8: Comparison between *C. auris*, *C. parapsilosis*, *C. albicans*, *S. cerevisiae*, and *C. glabrata* proteins with associated EC Numbers present in the iRV973, iDC1003, iRV781, iIN800, and iNX804 genome-scale metabolic models, respectively. A: Venn Diagram. B: Pairwise Intersections. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)). ..... 112

Figure II.9: Multi-species comparison in terms of proteins with an associated EC number present in the *C. neoformans* iRV890, *C. albicans* iRV781, *C. auris* iRV973,

*S. cerevisiae* iIN800 and *C. glabrata* iNX804 GSMMs. The multiple intersection was performed using jvenn [340]. ..... 132

Figure II.10 – *C. neoformans* pathway for ascorbate biosynthesis, with the respective *C. neoformans var. grubii* EC numbers present in the iRV890 model. The 1.1.3.8 and 3.1.1.17 enzymes, which are unique to *C. neoformans* among other pathogenic yeasts, are highlighted in purple. .... 134

Figure II.11 - Metabolic pathways for inositol assimilation as carbon source, A - based on the proposed fungal inositol assimilation pathway reported in Kuivanen et al. 2016 [348]. B – based on the animal inositol assimilation pathway, and C- based on the bacterial inositol assimilation pathway. The respective *C. neoformans var. grubii* genes present in iRV890 are highlighted in purple. The currently unknown genes are highlighted in red and the proposed reactions with an unknown EC number are represented as a question mark in red. .... 136

Figure III.1 – Molecular 2D structure of all the compounds acquired and experimentally tested in this work. 2a - 5-(perylene-2-ylethynyl)-arabino-uridine. 3a - 5-(perylene-3-ylbutadiynyl)-arabino-uridine. 4a - 5-(perylene-3-ylethynyl)-2'-O-methyl-uridine. 5a - 5-(4-phenyl-1,2,3-triazol-1-yl)uridine. 6a - 5-[4-(perylene-3-yl)-1,2,3-triazol-1-yl]uridine. 7a - 5-[4-(perylene-2-yl)-1,2,3-triazol-1-yl]uridine. y7 - 5-[3-(4-benzoylphenoxy)prop-1-ynyl]-1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl]pyrimidine-2,4-dione. y11 – 2'-Deoxy-5-[(perylene-3-yl)ethynyl]uridine. C1 - N10-(Trifluoroacetyl)pteroic acid. C5 - Homopteroic acid. .... 151

Figure III.2 - Multiple Sequence Alignments of Fol1 protein of *C. albicans* (C5\_00770C\_A) and respective orthologs of *C. parapsilosis* (CPAR2\_303390), *C. auris* (B9J08\_001330), *C. glabrata* (CAGL0J07920g) and *C. neoformans* (CNAG\_02786). The residues involved in the active site are highlighted in blue,

with 73% identity and 93% similarity among all the proteins. Alignment performed with CLUSTAL 2.1..... 155

Figure III.3 - Sequence alignment of *S. cerevisiae* (PDB code 2BMB) and *C. albicans* Fol1. Residues from the active site are highlighted in orange. Similar residues are represented in green, and identical in yellow. The two sequences share globally 62% similarity and 42% identity, and 100% similarity and 93% identity if considering the active site alone..... 156

Figure III.5 - Overlap of 7,8-dihydropteroate (white) and 5 of the compounds with higher affinity to *C. albicans* Fol1 modeled active site 3D structure obtained from virtual drug screening through molecular docking studies. The respective binding energy is indicated for each compound. C1: 4-[(2-amino-4-oxo-1H-pteridin-6-yl)methoxy]benzoic acid; C5: Pteronic acid; C8: ethyl 4-[acetyl-[2-(2-amino-4-oxo-7,8-dihydro-1H-pteridin-6-yl)ethyl]amino]benzoate; C6: Homopteroic acid; C13: 4-((2-(2-Amino-4-oxo-3,4,7,8-tetrahydropteridin-6-yl)ethyl)amino)benzoic acid..... 158

Figure III.6 - A – PAB (yellow) and (7,8-dihydropterin-6-yl)methyl diphosphate (pink) docked in *C. albicans* Fol1 3D structure. B – overlap of compound C5 (dark blue) and 7,8-dihydropteroate (light blue) docked in *C. albicans* Fol1 3D structure. C – mesh representation of compound C5. D - Polar interactions between C5 and protein active site residues. .... 160

Figure III.7 - Cytotoxicity assessment of compound C5 with a range of concentrations from 2 to 64 mg/L against VK, L929, and HeLa cell lines, upon 48h drug exposure. Cell viability was assessed by the tetrazolium (MTT) reduction assay, reading 570 nm absorbance. C5 showed cytotoxicity at all concentration ranges especially against, VK (murine) and HeLa (human) cell line..... 163

Figure III.8 - Surface representation of the CaChs2, highlighting the central tunnel and active site.....	165
Figure III.9 - Multiple Sequence Alignments of residues 307-664 of Chs2 protein of <i>C. albicans</i> (CR_09020C_A) and respective orthologs of <i>C. parapsilosis</i> (CPAR2_701490), <i>C. auris</i> (B9J08_005077), <i>C. glabrata</i> (CAGL0J11506g) and <i>C. neoformans</i> (CNAG_03326), performed with CLUSTAL 2.1. The residues involved in the active site are highlighted in blue, with 100% identity among all the proteins.....	166
Figure III.10 – A – Compound 7a (Best affinity in virtual drug screening -11.8 Kcal/mol) docked in the active site of CaChs2 with respective observed polar interactions. B – overlap of compound 7a (light blue) and Chitin (pink). C and D – mesh and stick representation, respectively, of compound 7a in surface view of CaChs2 active site.....	167
Figure III.11 – Fluorescence of <i>C. neoformans</i> cells after exposure to Chs2 predicted inhibitors 2a, 4a, and y11, which exhibit intrinsic fluorescence. The higher preponderance of intracellular fluorescence observed for compounds 2a and 4a is possibly connected to their stronger antifungal activity when compared to y11, that mostly accumulates in the cell periphery. Phase contrast images of the same cells are used as controls.....	169

# List of Tables

Table I.1 - Statistics of the most significant invasive fungal infections, highlighting in purple those caused by <i>Candida</i> and <i>Cryptococcus</i> . Data collected from [3].....	6
Table I.2 – Main different antifungal agents under development targeting intracellular fungal metabolic pathways, and their respective in vitro activity, clinical trials, and state of development.....	12
Table II.1 - Number of reactions in the main pathways of the <i>C. albicans</i> iRV781 model in comparison to <i>C. glabrata</i> iNX804 model and <i>S. cerevisiae</i> iMM904 model. ....	59
Table II.2. Biomass Composition used in the model iRV781.....	63
Table II.3. Comparison between <i>in vivo</i> and <i>in silico</i> phenotypic behavior of <i>C. albicans</i> under different carbon and nitrogen sources. ....	65
Table II.4 - Growth parameters of iRV781 and comparison with <i>in vivo</i> values for <i>C. albicans</i> and <i>S. cerevisiae</i> . ....	68
Table II.5 - Anaerobic growth assessment of iRV781 model in defined media with or without anaerobic supplements. DMM [263] (defined minimal medium); DMMsup. [263] (defined minimal medium supplemented with ergosterol and Tween 80); GPP [262] (glucose-phosphate-proline); GPPsup. [262] (glucose-phosphate-proline supplemented with oleic acid and nicotinate).....	69
Table II.6 - Drug targets evaluated for gene essentiality prediction in RPMI medium, as identified by the iRV781. Data retrieved from DrugBank database; only drugs with known pharmacological action were selected .....	74

Table II.7 - Biomass composition used in the model iDC1003. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.2.1.....	85
Table II.8 - <i>In silico</i> predictions versus <i>in vivo</i> described data regarding <i>C. parapsilosis</i> ability to grow in the presence of sole carbon and nitrogen sources. From the 62 different tested compounds, iDC1003 correctly predicted positive or null biomass production on 95%. A plus represents biomass production (+), a minus (-) no biomass production, and prediction disparities highlighted in bold. Referenced data from Westerdijk fungal collection refer to strains CBS 1954 and CBS 604.....	88
Table II.9 - Growth parameters of iDC1003 and comparison with experimentally determined values. ....	91
Table II.10 - Enzymes predicted to be essential in RPMI medium, based on the screening of the genome-scale metabolic models of <i>C. parapsilosis</i> , iDC1003, <i>C. albicans</i> , iRV781, and <i>C. glabrata</i> , iNX804.....	93
Table II.11 - Biomass composition used in the model iRV973. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.3.1.....	107
Table II.12 - Comparison between <i>in vivo</i> and <i>in silico</i> phenotypic behavior of <i>C. auris</i> under different carbon and nitrogen sources.....	110
Table II.13 - Growth parameters of iRV973 and comparison with experimentally determined values for <i>C. auris</i> . ....	111
Table II.14: Biomass composition used in the model iRV890. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.4.1. ....	128

Table II.15 - Growth parameter values predicted by the iRV890 model, in comparison with those determined experimentally.....	131
Table II.16 - Enzymes predicted to be essential in RPMI medium in 5 pathogenic fungal species, based on the screening of the genome-scale metabolic models of <i>C. neoformans</i> iRV890, <i>C. auris</i> IRV973, <i>C. parapsilosis</i> iDC1003, <i>C. albicans</i> iRV781, and <i>C. glabrata</i> iNX804. Grey rows highlight enzymes which are not encoded in the human genome. Data regarding the drug association retrieved from DrugBank database; only drugs with known pharmacological action against pathogens were selected.....	140
Table III.1: MIC50 of compound C5 and Fluconazole alone and in combination against <i>C. albicans</i> , <i>C. glabrata</i> , <i>C. parapsilosis</i> , <i>C. auris</i> , and <i>C. neoformans</i> . .....	161
Table III.2: MIC80 of compound C5 and Fluconazole against <i>C. albicans</i> , <i>C. glabrata</i> , <i>C. parapsilosis</i> , and <i>C. neoformans</i> . .....	162
Table III.3 – MIC50 values in $\mu\text{M}$ of different compounds targeting the Chs2 protein against the 5 indicated pathogenic fungal species. MIC50 for compounds 2a, 4a, 5a, and y7 with a higher inhibition activity against <i>C. neoformans</i> are highlighted in bold. ....	168

# List of Acronyms

AIDS - Acquired immunodeficiency syndrome

*A. fumigatus* - *Aspergillus fumigatus*

ATP - Adenosine triphosphate

BioISO - Biological networks constraint-based *In silico* Optimisation

BLAST - Basic Local Alignment Search Tool

BSRG - Biological Sciences Research Group

*C. albicans* - *Candida albicans*

*C. auris* - *Candida auris*

*C. dubliniensis* - *Candida dubliniensis*

*C. glabrata* - *Candida glabrata*

*C. krusei* - *Candida krusei*

*C. neoformans* - *Cryptococcus neoformans*

*C. parapsilosis* - *Candida parapsilosis*

DCW - Dry Cell Weight

CLSI - Clinical and Laboratory Standards Institute

DMEM - Dulbecco's Modified Eagle Medium

Cryo-EM - Cryogenic electron microscopy

CoA - Coenzyme A

COBRA - Constraints-Based Reconstruction and Analysis

CTP - Cytidine triphosphate

DOPE - Discrete Optimized Protein Energy

CPU - Central Processing Unit

dATP - Deoxyadenosine triphosphate

dCTP - Deoxycytidine triphosphate

dGTP - Deoxyguanosine triphosphate

dTTP - Deoxythymidine triphosphate

DMSO - Dimethyl sulfoxide  
DNA - Deoxyribonucleic Acid  
E.C. - Enzyme Commission  
FAD - Flavin adenine dinucleotide  
FBA - Flux balance analysis  
FCT - Fundação para a Ciência e a Tecnologia  
FDA - Food and Drug Administration  
FIC - Fractional Inhibitory Concentration  
FMN - Flavin mononucleotide  
GAM - Growth Associated Maintenance  
GlcNAc - N-Acetylglucosamine  
GPI - Glycosylphosphatidylinositol  
GPP - Glucose-phosphate-proline  
GPR - Gene-protein-reaction  
GSMMs - Genome-scale metabolic models  
GTP - Guanosine triphosphate  
HIV - Human Immunodeficiency Virus  
HPLC - High-performance liquid chromatography  
HMMs - Hidden Markov Models  
HSP90 - Heat shock protein 90  
iBB - Institute for Bioengineering and Biosciences  
IBM - International Business Machines Corporation  
IC50 - Half-maximal inhibitory concentration  
ID - Identification  
KEGG - Kyoto Encyclopedia of Genes and Genomes  
L-DOPA - L-3,4-dihydroxyphenylalanine  
MCSS - Multiple Copy Simultaneous Search  
MEGA - Molecular Evolutionary Genetics Analysis  
MOPS - Morpholinepropanesulfonic acid

mRNA - Messenger RNA

MIC - Minimum inhibitory concentration

MTT - 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide

NAD - Nicotinamide adenine dinucleotide

NADP - Nicotinamide adenine dinucleotide phosphate

NCAC - Non-*C. albicans Candida* species

NCBI - National Center for Biotechnology Information

NGAM - Non-Growth Associated Maintenance

NMR - Nuclear magnetic resonance

OD - Optical Density

PDB - Protein Data Bank

PDBQT - Protein Data Bank, Partial Charge (Q), & Atom Type (T)

*P. aeruginosa* - *Pseudomonas aeruginosa*

QIDP - Qualified Infectious Disease Product

RNA - Ribonucleic acid

rRNA – Ribosomal RNA

RPMI - Roswell Park Memorial Institute

*S. cerevisiae* - *Saccharomyces cerevisiae*

SMILES - Simplified Molecular Input Line Entry System

SBML - Systems biology markup language

SMM – Synthetic minimal defined medium

STAT - Stationary phase yeasts

TCDB - Transporter Classification Database

TranSyT - Transport Systems Tracker

tRNA - Transfer RNA

UTP - Uridine-5'-triphosphate

VVC - Vulvovaginal candidiasis

YNB - Yeast Nitrogen Base

YPD - Yeast Extract-Peptone-Dextrose

WHO - World Health Organization

# **I. Introduction**

## I.1. Thesis outline

Fungal infections pose a significant public health threat, particularly for immunocompromised patients. While current antifungal treatments are reasonably effective, concerns about emerging drug resistance and the limitations of existing therapies persist. Mortality rates, particularly in cases involving *Candida* and *Cryptococcus* infections, remain alarmingly high. The proliferation of resistant clinical isolates, particularly within *Candida* and *Aspergillus* species, further worsens this issue. Some isolates of non-*albicans* pathogenic *Candida* species, such as *C. auris*, have even demonstrated resistance to the three antifungal classes currently used for its treatment. Given the increasing rates of therapeutic failure and mortality, there is an urgent need for new drug targets and novel drugs. However, the process of developing new drugs remains challenging and currently must go through not only the improvement or repurposing of existing drugs, but also through the discovery of other strategies of addressing fungal infections, searching for new distinct targetable cellular processes, and entirely novel drugs. With recent technological advances, *in silico* approaches and artificial intelligence offer a whole new perspective that could be the solution to this problem. In this scope, the use of GSMs and molecular docking could play an important role in the search for the perfect antifungal agent and recent examples showcase their potential in identifying new drug targets for combating pathogenic organisms. Considering these challenges, with this work we propose an innovative *in silico* pipeline to assist the discovery of new antifungal drugs. First by reconstructing Genome-scale Metabolic Models for relevant pathogenic fungal species a comprehensive understanding of cell metabolism and its unique metabolic features is provided. These models can then be used as a platform for the identification of potential drug targets through enzyme essentiality prediction in conditions similar to the

ones faced by pathogens inside the human host. Then by using virtual drug screening through molecular docking, compounds with the potential to inhibit the identified drug targets are identified and tested for antifungal activity.

Altogether, this thesis provides comprehensive insights into discovering new drugs to combat fungal infections, providing new lists of enzymes and pathways suitable as future drug targets, along with sets of compounds with antifungal potential worthy of experimental testing. The integration of GSMs and virtual drug screening is an innovative methodology with the potential to accelerate the development of novel antifungal agents, ultimately improving patient outcomes and addressing the growing challenges posed by fungal infections.

Chapter I provides a literature review of current treatment options for invasive fungal infections while also focusing on the drug targets that have been under scrutiny for the development of novel antifungal agents. Additionally, it provides an overview of the evolutionary trajectory, and development stages, as well as the challenges and limitations inherent in the development of new antifungal drugs. The chapter introduces Genome-scale Metabolic Models as a promising avenue in drug discovery, offering insights into their methodology, applications, and the significance of tools like the *merlin* software. Furthermore, it delves into an extensive review of existing methodologies and applications pertaining to molecular docking, exploring its associated techniques and variations for a comprehensive understanding.

Chapter II presents the reconstruction of the genome-scale metabolic models for the human pathogen *Candida albicans* (*C. albicans*) iRV781, *Candida parapsilosis* (*C. parapsilosis*) iDC1003, *Candida auris* (*C. auris*) iRV973, and *Cryptococcus neoformans* (*C. neoformans*) iRV890. The reconstruction steps are described and the validation strategy demonstrating the model's accuracy in predicting the utilization of different carbon and nitrogen sources and specific growth rate is highlighted. Some unique metabolic features of each pathogen that the models allowed to

predict are also addressed. This chapter also explores the models' potential as platforms for identifying drug targets, by comparing known targets with predictions of gene essentiality under conditions mimicking the human host. This chapter sets the stage for further exploration and validation of this innovative approach, with the potential to contribute to the field of antifungal drug discovery.

In Chapter III some of the drug targets identified in Chapter II (Fol1 and Chs1) are explored through *in silico* virtual drug screening of large databases of compounds homologous to compounds that naturally participate in the enzyme reaction and bind to the active site. Selected compounds were tested experimentally for antifungal activity, alone or in combination with other antifungal drugs, and for human cytotoxicity, leading to the identification of potential new antifungal principles.

Chapter IV discusses all the results presented in this work, highlighting the importance of *in silico* approaches like GSMMs and virtual drug screening towards an increased comprehension of the mechanisms of virulence of fungi, metabolic characterization, and modern drug discovery, and the potential impact of this work in public health.

## I.2. Invasive fungal infections

More than one billion people worldwide are estimated to be afflicted by fungal diseases per year [1]. While many of these infections are superficial and easily treatable, mainly dermatomycoses, diseases that do not invade deep tissues and rarely lead to life-threatening complications [2], the toll rises significantly due to invasive fungal infections. According to the last systematic review published by The Lancet Infectious Diseases using data from more than 120 countries, it is estimated that 6.55 million people annually are affected by invasive fungal infection, resulting in 2.55 million deaths annually directly attributable to fungal diseases [3]. Species from the *Aspergillus*, *Candida*, *Cryptococcus*, and *Pneumocystis* genera are the most significant pathogenic fungi, *Candida* and *Cryptococcus* infections being particularly concerning since they have particularly high mortality rates (Table I.1). Over the years efficient antifungals have been developed for the treatment of infections caused by these pathogens. However, the current challenge lies in combating antimicrobial resistance. Recent years have witnessed a shift in therapeutic strategies within intensive care units, emphasizing timely and effective treatment that targets specific pathogens while minimizing the risk of inducing microbial resistance [4]. This requires not only swift and accurate identification of the causative agent but also the exploration and development of alternative drugs with unique mechanisms of action. Understanding the mechanisms and unique features of these antifungal agents is crucial for the effective management of invasive fungal infections and for mitigating the challenges posed by drug resistance.

**Table I.1** - Statistics of the most significant invasive fungal infections, highlighting in purple those caused by *Candida* and *Cryptococcus*. Data collected from [3]

Disease	People affected worldwide/year	Deaths worldwide/year	Mortality Rate
Chronic pulmonary aspergillosis	1.84 million	340 000	18 %
<b>Invasive candidiasis</b>	<b>1.57 million</b>	<b>995 000</b>	<b>64 %</b>
Pneumocystis pneumonia	0.51 million	214 000	42 %
<b>Cryptococcal meningitis</b>	<b>0.19 million</b>	<b>147 000</b>	<b>76 %</b>
Other invasive fungal infections	0.30 million	161 000	54 %

### I.2.1. Antifungal Frontline: Current treatment options

When dealing with invasive fungal infections, azoles and echinocandins are the first-line treatment options, though polyenes and fluoropyrimidines can also be used. These four families of antifungal agents are Food and Drug Administration (FDA)-approved for clinical use and represent the main treatment options currently available in the battle against fungal infections (Figure 1). The first antifungal drug to be approved was amphotericin B, in the 1950s, a polyene known for its potency as a fungicidal, but also for significant toxic effects [5]. This drug acts by binding to sterols, particularly ergosterol, creating pores in the fungal membrane causing leakage of cellular components, and ultimately leading to cell death [4]. While amphotericin B exhibits broad-spectrum fungicidal activity, its use is often limited due to toxicity concerns, leading to a preference for other antifungals available. Amphotericin is currently used as a second-line treatment option for invasive fungal infections [6,7] as prophylactic

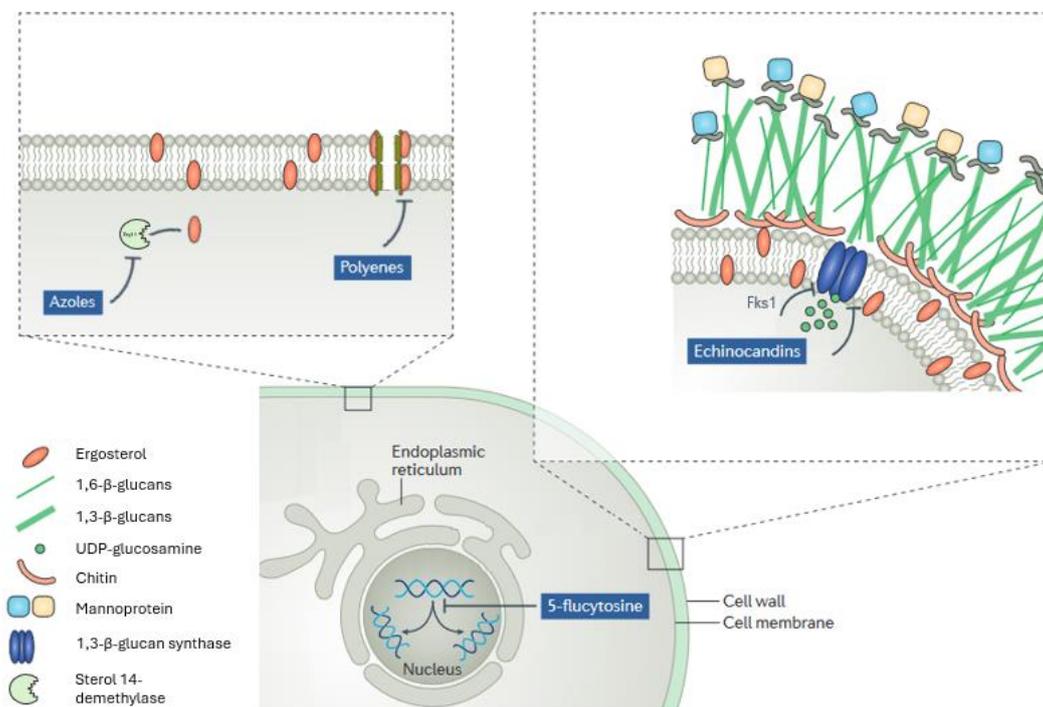
therapy in immunosuppressed patients [4] and as first option in Cryptococcal meningitis in combination with flucytosine [8].

Flucytosine is a pyrimidine analogue approved in the 1960s, that acts by inhibiting DNA and RNA synthesis in fungi. Flucytosine is metabolized inside fungal cells into 5-fluorouracil, which is then incorporated into RNA, while also inhibiting thymidylate synthase activity which affects DNA synthesis [9]. Since this drug is associated with high rates of drug resistance development its use is restricted, being always administrated in combination with other drugs. Currently, flucytosine is mostly used to treat Cryptococcal meningitis, as referred before, in combination with amphotericin B.

Although discovered much earlier, azoles, one of the most commonly used families of antifungals in clinical practice, only began to be used for the treatment of fungal infections in the 1980s and were the only available oral agent for the treatment of systemic fungal infections for almost a decade [10]. Azoles inhibit 14- $\alpha$ -lanosterol demethylase (Erg11), a cytochrome P-450 enzyme involved in the synthesis of ergosterol, resulting in decreased ergosterol synthesis and accumulation of toxic sterols that disturb membrane functions. While first-generation azoles like miconazole and ketoconazole were followed by second-generation extended-spectrum azoles such as fluconazole and itraconazole, third-generation options like voriconazole, posaconazole, and isavuconazole offer improved efficacy, safety, and pharmacokinetics [5]. Azoles are fungistatic and are commonly employed as first-line treatment strategy for invasive aspergillosis, as an alternative option for invasive candidiasis and cryptococcal meningitis, as step-down therapy for most fungal infections, and in prophylactic treatments [4,6–8]. Among azoles, fluconazole has been used preferentially, voriconazole being selected in the case of fluconazole-resistant strains and *Candida krusei* (*C. krusei*) (intrinsically resistant). Itraconazole, posaconazole, or

isavuconazole, have also been used in specific cases as second-line therapeutics [4,6].

Echinocandins, the most recent class of antifungal drugs whose development began in the early 2000s, target the catalytic subunit of  $\beta$ -1,3-glucan synthase, an enzyme essential for fungal cell wall integrity. Echinocandins unlike azoles are fungicidal so they go well beyond preventing growth and are currently recommended to treat invasive candidiasis. Since  $\beta$ -1,3-glucan synthases are fungal-specific enzymes, echinocandins exhibit low host toxicity [5]. Within the echinocandin drug family, caspofungin is typically used to treat *C. krusei* or *Candida glabrata* (*C. glabrata*) infections, for which azole therapy has a higher propensity to fail. Micafungin and anidulafungin are additional options within the echinocandin family [4,6].



**Figure I.1.** Cellular target of the main treatment options currently used in the treatment of invasive fungal infections. Adapted from [5].

### **I.2.2. Antifungal Frontline: Why are we losing the fight**

As presented before, fungal infections pose a significant threat to public health, especially among immunocompromised patients and those with underlying medical conditions. While antifungal agents play a crucial role in combating these infections, the emergence of drug resistance and the limitations of existing therapies raise concerns on our ability to effectively manage fungal diseases. As observed in Table I.1, despite the availability of antifungals, mortality rates remain extremely high, especially for *Candida* and *Cryptococcus*. The current antifungal treatment is often prolonged, raising concerns about poor immediate-term fungicidal activity, the emergence of direct antifungal drug resistance, or even poor patient compliance. Additionally, invasive fungal infections are more prevalent in critically ill patients who may already be burdened with multiple therapies and are more vulnerable to organ toxicity and drug interactions [5].

In recent decades, there has been a concerning increase in the prevalence of drug-resistant and multidrug-resistant clinical isolates, characterized by both acquired and intrinsic resistance, observed especially in *Candida* and *Aspergillus* species. This is one of the main factors contributing to the high mortality rates. This trend is particularly alarming in non-*albicans* pathogenic *Candida* species. Indeed, around 10% of *C. glabrata* clinical isolates are resistant to fluconazole [11,12] and nearly 40% of *C. glabrata* isolates resistant to at least one echinocandin were also resistant to fluconazole [13]. Also, some *C. auris* isolates display resistance to all three classes of available antifungal drugs. In a study involving 41 clinical isolates from different continents, 93% of isolates were resistant to fluconazole, 35% to amphotericin B, and 7% to echinocandins; 41% were resistant to 2 antifungal classes and 4% were resistant to the 3 classes [14].

The same scenario is observed in *Aspergillus* species. *Aspergillus terreus* and *Aspergillus flavus* are known for their intrinsic resistance to amphotericin B. Some species pose an even greater challenge by exhibiting intrinsic resistance to both

amphotericin B and azoles, complicating treatment strategies [15]. Acquired resistance to azoles has also seen a significant surge, with some recent studies reporting a prevalence of 19% in some medical centers [16]. Additionally, clinical strains containing mutations that confer resistance to multiple azole drugs, even without prior exposure to azoles, are being reported in several hospitals around the world including Europe, Asia, the Middle East, Africa, Australia, and the USA [17–25].

*Cryptococcus* species exhibit drug resistance less frequently, as they are typically treated with a combination therapy approach. Initial treatment often involves a combination of amphotericin B and flucytosine, followed by a consolidation regimen with fluconazole and long-term maintenance [26]. Despite its rarity, an increase in fluconazole resistance has been noted, rising from 7.3% to 11.7% between 1997 and 2007 [27], fluconazole resistance being particularly observed in relapse cases [28]. Notably, *Cryptococcus* can transform into large polyploid titan cells during pulmonary infection, crucial for virulence and dissemination to the central nervous system. Interestingly, these titan cells also generate daughter cells that exhibit increased resistance to fluconazole [29].

Given the increasing therapeutic failure, the high mortality rates, and the increase of multidrug resistance, the pursuit of new drug targets and novel drugs is essential, ideally with innovative mechanisms of action to mitigate cross-resistance. An ideal antifungal agent should possess attributes such as reduced toxicity and side effects; a broad spectrum of activity against a range of fungal species; preferably fungicidal activity; act in fungal-specific primary pathways and targets; bioavailability; and a cost-effective synthesis [30]. Developing the ideal antifungal remains a big challenge. However, there have been notable strides in recent years, including the FDA approval and integration of new antifungal drugs into clinical practice (Table I.2).

### I.2.3. Antifungal development - current drug targets

#### I.2.3.1. Ergosterol

Ergosterol is an important target for antifungal drugs due to its crucial role in the structure and function of fungal cell membranes. This sterol, analogous to cholesterol in animal cells, maintains membrane integrity, regulates fluidity, and is vital for fungal survival. Targeting ergosterol synthesis disrupts membrane integrity, making it an attractive strategy for antifungal therapy. As mentioned in a previous section, azoles, and polyenes target ergosterol biosynthesis or ergosterol itself, respectively. More recently, new progress has been made in developing more effective drugs targeting sterols, the most relevant cases being oteseconazole, isavuconazole, and Encochleated amphotericin B.

**Oteseconazole:** this new oral drug is a tetrazole, a synthetic doubly unsaturated five-membered ring aromatic heterocycles, consisting of one carbon and four nitrogen atoms. Some tetrazole-based drugs have been under development in recent years, all developed by Mycovia Pharmaceuticals Inc. [31–33]. The most successful case is Oteseconazole (VT-1161) which was recently FDA approved. Oteseconazole was rationally designed as a potent inhibitor of fungal lanosterol 14- $\alpha$  demethylase (Erg11) with improved selectivity over human enzymes [31,34]. This increased selectivity is due to the use of a tetrazole as the heme iron-binding moiety instead of the common imidazole or triazole metal-binding group found in previously approved azole drugs. Studies demonstrate its efficacy against several fungal species, including *Candida* (also fluconazole-resistant ones), *Cryptococcus*, *Coccidioides*, and dermatophyte species [31,35,36]. Clinical trials phase I, II, and III have been successfully completed, showcasing Oteseconazole's effectiveness and safety in treating recurrent vulvovaginal candidiasis (VVC), leading to its FDA approval as the first and only medication authorized for chronic yeast infection therapy [37].

**Table I.2** – Main different antifungal agents under development targeting intracellular fungal metabolic pathways, and their respective in vitro activity, clinical trials, and state of development.

Drug	In vitro activity			Clinical trials			Approval		State of development
	Candida spp	Aspergillus spp	Cryptococcus spp	Phase I	Phase II	Phase III	FDA	Clinical	
Oteseconazole	✓	✓	✓	✓	✓	✓	✓	✓	Complete
Rezafungin	✓	✓	✗	✓	⌚	⌚	✓	✓	Complete
Encochleated AMP	✓	✓	✓	⌚	⌚	✗	✓	✗	In phase III clinical trials
Ibrexafungerp	✓	✓	?	✓	✓	✓	✓	✓	Complete
Olorofim	✗	✓	?	✓	⌚	⌚	✗	✗	In phase III clinical trials
Fosmanogepix	✓	✓	✓	✓	⌚	⌚	✗	✗	In phase III clinical trials
Nikkomycin Z	✓	✗	✗	✗	✗	✗	✗	✗	Stopped, financial problems
ATI-2307	✓	✓	✓	✗	✗	✗	✗	✗	In preclinical evaluations
Illicolin H	✓	✓	✓	✗	✗	✗	✗	✗	Under study
AR-12	✓	✓	✓	✗	✗	✗	✗	✗	Stopped, financial problems
Aureobasidin A	✓	✓	✓	✗	✗	✗	✗	✗	Under study
VL-2397	✗	✓	✓	✗	✗	✗	✗	✗	Stopped, financial problems
Mohangamides	✓	✓	?	✗	✗	✗	✗	✗	Under study
FK506 analogs	✓	✓	✓	✗	✗	✗	✗	✗	Under study
MGCD290	✓	✓	?	✗	✗	✗	✗	✗	Clinical trials failed
Efungumab	✓	?	?	✗	✗	✗	✗	✗	Stopped, safety concerns
HONV	✓	?	?	✗	✗	✗	✗	✗	Under study

**Encochleated amphotericin B:** Encochleated amphotericin B (MAT2203) represents a novel formulation of amphotericin B, designed to overcome challenges associated with its low solubility and poor oral bioavailability. Traditional amphotericin B formulations require intravenous administration due to its tendency to self-aggregate in aqueous media. However, MAT2203 uses cochleate's, multilayered phospholipid structures, to encapsulate amphotericin B, facilitating oral delivery while protecting the drug from gastrointestinal degradation. Preclinical studies in murine models have demonstrated the efficacy of MAT2203 against several fungal infections, including *Candida*, *Aspergillus*, and cryptococcal meningitis [38–40]. In these studies, MAT2203 exhibited dose-dependent reductions in fungal load and mortality rates comparable to conventional amphotericin B treatments. Currently undergoing phase II clinical trials, MAT2203 has shown promise in the treatment of moderate to severe vulvovaginal candidiasis (VVC). Although fluconazole demonstrated superior clinical and mycological outcomes compared to MAT2203, it exhibited a favorable safety profile with no associated mortality or serious adverse events (same as fluconazole). MAT2203 was designated by the FDA as a Qualified Infectious Disease Product (QIDP) with Fast Track status, with the potential for Orphan Drug designation.

#### **1.2.3.2. $\beta$ -1,3-glucan**

$\beta$ -1,3-glucan is a polysaccharide found in fungal cell walls. Targeting  $\beta$ -1,3-glucan biosynthesis represents an effective strategy for antifungal drug development since  $\beta$ -1,3-glucan synthase is fungal specific and its inhibition by echinocandins compromises cell wall integrity, leading to cell lysis and death. Recently, Rezafungin was added as the newest member of the echinocandin family, and Ibrexafungerp also a glucan synthase inhibitor appeared as the first approved drug in a novel antifungal class in more than 20 years.

**Rezafungin:** this drug developed by Cidara Therapeutics, represents the first member of the second generation of echinocandins and is a promising antifungal agent. It is a structural analog of anidulafungin, offering improved stability against host degradation pathways, enhanced solubility, and an extended half-life while maintaining the typical efficacy and safety profile of echinocandins [41]. Rezafungin demonstrates potent antifungal activity against *Candida* and *Aspergillus* species, including strains resistant to fluconazole [42–44]. However, *C. neoformans* is resistant to Rezafungin, just like for the remaining echinocandins, and strains with mutations in FKS genes exhibit higher MIC values against Rezafungin as well [42]. Clinical trials, including phase I, II, and III studies, confirm Rezafungin's safety and efficacy and showed comparable outcomes to caspofungin in patients with candidemia and/or invasive candidiasis [45]. Rezafungin was designated by the FDA as a QIDP with Fast Track status for intravenous and topical use in the treatment of candidemia and invasive candidiasis in 2015 and 2016 respectively [41]. In March 2023, rezafungin received approval for the treatment of candidemia and invasive candidiasis in adults with limited or no alternative treatment options [46].

**Ibrexafungerp:** Ibrexafungerp, formerly known as SCY-078 and developed by Scynexis, is the first oral  $\beta$ -1,3-glucan synthase inhibitor. Unlike other glucan synthase inhibitors, Ibrexafungerp can be administered intravenously or orally, due to its excellent oral bioavailability, and permeability with pH-dependent solubility reaching the highest concentrations in acidic media like gastric and intestinal fluids [41]. This versatility expands treatment options for fungal infections. In vitro studies demonstrate Ibrexafungerp's potent fungicidal activity against *Candida* species [47], including biofilms, surpassing fluconazole in efficacy [48]. It also exhibits fungistatic activity against *Aspergillus* species [47]. This new drug is effective against azole and echinocandin-resistant strains since it has an independent target site with only a partial overlap with  $\beta$ -1,3-glucan

synthase active site, which confers different resistance mechanisms from those observed for echinocandins, resulting in very limited cross-resistance between the two drugs [49]. Phase I, II, and III clinical trials showed minimal side effects, notable responses in severe fungal infections, and successful treatment of candidemia caused by *C. auris*, resulting in FDA approval in 2021 for the treatment of vulvovaginal candidiasis. This was a significant milestone as it was the first approved drug in a new antifungal class in over two decades and the first and only non-azole treatment for VVC [50].

### **I.2.3.3. Other currently targeted pathways in fungi**

**Chitin:** Chitin, a polysaccharide consisting of  $\beta$ -(1,4)-linked N-acetylglucosamine units, is a fundamental component of the fungal cell wall, providing rigidity and integrity, so it is expected that the disruption of chitin biosynthesis results in a weakening of the cell wall, ultimately compromising fungal cell integrity and viability. Additionally, this enzyme is not present in humans representing an optimal drug target. Despite not being so much explored, there is a drug, Nikkomycin Z, a pyrimidine nucleoside produced as a secondary metabolite by *Streptomyces tendaeis* discovered in the 1970s [51], which is a competitive inhibitor of chitin-synthase. Nikkomycin demonstrates efficacy against *C. albicans* and *C. parapsilosis*, but its effectiveness is lower for other *Candida* species and fungal pathogens. Despite that, when used in combination with fluconazole or itraconazole, a synergistic effect is observed for most strains [52]. Preclinical animal studies have indicated no adverse effects or detectable toxicity of Nikkomycin [53]. Despite these promising findings, two phase II clinical trials initiated in 2014 were terminated due to recruitment challenges and lack of funding, and its development was abandoned.

**GPI-anchored proteins:** Glycosylphosphatidylinositol (GPI)-anchored proteins are found in eukaryotic organisms. These proteins are functionally diverse and play important roles including in fungal adhesion to the host cells, cell wall integrity, cell signaling, and virulence [54]. Targeting GPI anchor biosynthesis may represent a promising antifungal strategy.

Currently, there is one drug targeting GPI-anchors in development, Fosmanogepix, developed by Amplyx Pharmaceuticals, which targets the enzyme inositol acyltransferase (Gwt1) involved in the glycosylphosphatidylinositol (GPI)-anchored biosynthesis pathway [55]. By inhibiting Gwt1, Fosmanogepix disrupts the proper trafficking and anchoring of mannoproteins to the fungal cell membrane and outer cell wall, crucial for cell wall integrity, adhesion, and evasion of the host immune system. Currently, in Phase II clinical trials for the treatment of invasive fungal infections, Fosmanogepix has shown promising results. In completed studies, a significant proportion of patients treated with Fosmanogepix achieved clearance of *Candida* from blood cultures without requiring additional antifungal treatment. Ongoing trials aim to further evaluate the safety and efficacy of both intravenous and oral administration of Fosmanogepix. Fosmanogepix exhibits broad-spectrum activity against clinically important yeasts and molds, including *Candida* [56–58], *Aspergillus* [59]), *Scedosporium* [59,60] and *Fusarium* [59,60]. Studies in various animal models, including those with multi-drug resistant infections, have also demonstrated its effectiveness [61]. Although resistance development has been evaluated, Fosmanogepix shows comparable spontaneous mutation frequencies to echinocandins like Anidulafungin and Caspofungin in *Candida* species. Serial passage experiments have shown a modest increase in MIC values for some *Candida* species [62].

**Mitochondria:** The novel arylamidine ATI-2307, exhibits promising efficacy against *Candida*, *Aspergillus*, and *Cryptococcus* species, demonstrating in vitro

fungicidal activity [63]. Acting through a mechanism involving disruption of yeast mitochondrial membrane potential. ATI-2307 is efficiently internalized into *C. albicans* cells by a specific polyamine transporter, likely Agp2, related to the uptake of spermine and spermidine [64,65]. Once inside the cell, ATI-2307 reaches mitochondria and disrupts the mitochondrial membrane potential, resulting in mitochondrial dysfunction. This inhibition seems to be fungal-specific [66]. It was suggested by Yamashita et al. that mitochondrial dysfunction may be related to the inhibition of respiratory chain complexes III and IV in *Saccharomyces cerevisiae* and *C. albicans*, resulting in decreased intracellular ATP levels [63].

Another mitochondrial inhibitor Ilicolin H, a polyketide produced by *Cylindrocladium ilicicola* and *Gliocadium roseum*, displays broad-spectrum antifungal activity against *Candida* spp., *Cryptococcus* spp., and *Aspergillus fumigatus* (*A. fumigatus*) [67]. This drug acts as a mitochondrial respiratory-chain inhibitor, it targets the yeast mitochondrial cytochrome bc1 reductase. Despite its unique inhibition mechanism, Ilicolin H's clinical development has stagnated due to challenges associated with plasma protein binding, limiting its *in vivo* potency, and efforts to enhance its efficacy have yielded limited success.

**Pyrimidines:** Olorofim a member of the orotomide class of antifungal agents, targets the fungal dihydroorotate dehydrogenase enzyme responsible for pyrimidine biosynthesis required for deoxyribonucleic acids, cell wall, phospholipid synthesis, cell regulation, and protein production [41,68]. This drug shows low toxicity to mammalian cells despite the enzyme being also found in humans [68]. With good bioavailability and distribution into tissues, including the brain, it exhibits activity against various pathogenic fungi, including *Aspergillus* spp., but not against *Candida* spp. and zygomycetes [68]. Olorofim's resistance is not easily induced, as demonstrated by microevolution studies [68].

Currently in phase III trials to evaluate its effectiveness compared to Amphotericin B.

**Protein kinase:** AR-12 is a celecoxib derivative initially it was explored as an anticancer agent, however demonstrated antifungal activity against *Candida*, *Cryptococcus*, and *Aspergillus* species. AR-12 displays two mechanisms of action: first, it targets the acetyl coenzyme A synthetase competing with ATP and inhibiting the production of acetyl-CoA, with impact on the regulation of histone acetylation and carbon metabolism [69]; additionally, it downregulates host chaperone expression, particularly GRP78, HSP90 and HSP27, reducing the host immune response [70]. Fungal cells treated with AR-12 show induction of autophagy, decreased histone acetylation, and loss of cellular integrity. AR-12 is a weak inhibitor of human acetyl CoA synthetase furthermore acetyl CoA is primarily synthesized by an alternate enzyme, ATP-citrate lyase, in mammalian cells, making this enzyme an excellent drug target [69]. Despite its potential, no ongoing studies exist due to the bankruptcy of the developing company, and developments are not expected in the near future [71].

**Sphingolipids:** Aureobasidin A is a cyclic depsipeptide isolated from *Aureobasidium pullulans*, targets inositol phosphorylceramide synthase, an enzyme responsible for sphingolipid synthesis exclusively found in the fungal cell membrane [72], offering low toxicity due to its fungal-specific nature. In vitro studies demonstrate antifungal activity against *Candida* and *Aspergillus* spp. [73]. Although mutations in the AUR1 gene confer resistance to this drug [74–76], which is an obstacle to its development, however, there are ongoing studies that aim to develop novel derivatives with enhanced antifungal activity [77].

**Siderophores:** VL-2397 is a cyclic hexapeptide from *Acremonium persicinum* and exhibits potent antifungal activity against various species, including *Aspergillus* and *Candida* [78,79]. This new drug is structurally similar to the siderophore ferrichrome and acts through the chelation of aluminum ions instead of iron. Iron

is pivotal for the survival of microorganisms and siderophores are used to acquire iron from the environment [80]. While its mechanism of action is not fully understood, studies suggest dependency on the SIT1 gene for uptake in *A. fumigatus* and susceptibility influenced by iron availability [78]. Although showing promise in phase I trials with manageable adverse effects, a phase II trial for invasive aspergillosis was discontinued due to financial constraints, with no further development [71].

**Histone deacetylase:** MGCD290 is a fungal histone deacetylase inhibitor (Hos2). Histone deacetylases are enzymes responsible for deacetylation of lysines on core histones and other cellular proteins and play an important role in gene regulation and regulation of cellular proliferation and motility [81]. While this drug shows modest antifungal activity alone, it demonstrates synergistic effects when combined with fluconazole, posaconazole, voriconazole, and echinocandins against several fungal species, including *Candida*, *Aspergillus*, *Fusarium*, and *Zygomycetes* [81]. Despite safety demonstrated in initial clinical trials, a phase II clinical trial for moderate to severe VVC did not show significant improvement when MGCD290 was added to fluconazole compared to fluconazole alone [82]. No further clinical trials have been initiated.

**Hsp90:** Efungumab is a monoclonal antibody targeting heat shock protein 90 (HSP90), HSP90 is a conserved chaperone responsible for the regulation of function and stability of several proteins in pathogenic fungi under stressful environmental conditions such as exposure to antifungals or high temperatures [71,83]. This drug binds to the middle region of HSP90, preventing a conformational change needed for fungal viability, resulting in decreased resistance to antifungal agents and increased antifungal activity [84]. Efungumab shows synergy with amphotericin B and caspofungin against *Candida* spp. Clinical trials demonstrate improved outcomes in invasive candidiasis when combined with amphotericin B. Despite efficacy, marketing authorization was

denied due to quality and safety concerns related to the manufacturing process [85].

**Glyoxylate cycle:** The glyoxylate cycle is a pathway that allows pathogens to thrive on alternative carbon sources in host environments, it was shown to be required for *C. albicans* virulence [86]. Since this metabolic cycle is not present in humans it is considered a promising antimicrobial drug target [87]. Currently, there are 2 promising drugs targeting this pathway.

Mohangamides A and B are novel dilactone-tethered pseudodimeric peptides isolated from *Streptomyces* sp. that target the isocitrate lyase enzyme encoded by the *ICL1* gene in *C. albicans*. Bae et al. showed that when grown on glucose Mohangamides is unable to inhibit several fungal species, including *C. albicans* and *A. fumigatus*, however, when grown on acetate it inhibited *C. albicans* growth. Additionally, Mohangamide A showed potent inhibition against *C. albicans* isocitrate lyase with an IC<sub>50</sub> value of 4.4  $\mu$ M while Mohangamide B exhibited moderate inhibition [87]. Mohangamides did not display cytotoxicity against human carcinoma cell lines [87]. This drug has only recently been reported as an antifungal and there are no ongoing clinical trials so far.

Monoterpenoid perillyl alcohol was also recently reported [88] as a possible new drug to treat *Candida* infections, however only preliminary studies were presented, and more studies will be needed. Nevertheless, this molecule appears as another inhibitor of the glyoxylate cycle. Enzyme kinetics showed that monoterpenoid perillyl alcohol inhibits *C. albicans* isocitrate lyase in a competitive manner and *C. albicans* malate synthase in a non-competitive manner, both key enzymes of the glyoxylate cycle. The antifungal efficacy of this new drug was also shown by enhanced survival of *Caenorhabditis elegans* model infected with *C. albicans* [88].

**Calcineurin pathway:** Calcineurin pathway is crucial for activating virulence genes and proteins essential for fungal growth, hyphal development, and survival. This signaling pathway is highly conserved among pathogenic fungi like *Cryptococcus* spp, *Aspergillus* spp, and *Candida* spp [89]. However, developing drugs targeting this pathway is challenging due to Calcineurin's conservation from fungi to mammals and the immunosuppressive nature of compounds targeting it [90]. Despite that some attempts have been made, Juvvadi et al. reported a new compound, APX879, analog to FK506, a known calcineurin-targeting molecule, with an acetohydrazine substitution of the C22-carbonyl, which exhibits reduced immunosuppressive activity and shows efficacy against some pathogenic fungi tested, including *A. fumigatus*, *C. albicans*, and *C. neoformans* [91]. Further research led to the discovery of JH-FK-05, a novel compound from a panel of C-22-modified compounds. JH-FK-05 exhibits broad-spectrum antifungal activity *in vitro* without causing immunosuppression *in vivo*. Treatment with JH-FK-05 in murine models infected with *C. neoformans* significantly decreased fungal burden and increased survival rates [89].

**Homoserine dehydrogenase:** the homoserine dehydrogenase enzyme, encoded by the HOM6 gene, plays a vital role in fungal cells. Disruption of this gene leads to growth inhibition due to the accumulation of toxic aspartate  $\beta$ -semialdehyde [92]. In *C. albicans*, disruption of the HOM6 gene results in growth defects under amino acid starvation conditions, reduced cell adhesion, and subsequently diminished virulence [93]. The natural compound 5-hydroxy-4-oxo-l-norvaline (HONV) acts as an inhibitor of homoserine dehydrogenase and exhibits antifungal activity against *Candida* spp. However, HONV faces limitations in penetrating the yeast cell cytosol, where homoserine dehydrogenase is located, thus limiting its antifungal potential [94]. Despite this challenge, ongoing studies are exploring solutions to overcome this barrier. In a recent study, Skwarecki et al reported a novel strategy for the internalization using antifungal dipeptides

incorporating HONV, that once internalized are cleaved by intracellular peptidases to release the inhibitor [94]. Using this strategy, they were able to produce a dipeptide incorporating HONV that showed a slightly higher in vitro *anticandidal* activity in RPMI-1640 medium compared to HONV alone.

#### **I.2.4. The antifungal discovery strategy**

The journey of drug discovery spans across millennia, reaching back to the early times of human civilization. Treatments were often stumbled upon by chance or derived from keen observations of nature. Ingredients sourced from plants or animals played a pivotal role [95]. In the early 1900s, scientists began to explore drugs in a more systematic manner, marking a shift toward modern pharmaceutical research. Despite all the advances in medicine nowadays and the steady increase in research and development, the discovery of new drugs seems to be drying-up or to remain essentially stable [95]. The discovery of penicillin by Alexander Fleming in 1928 [96] is considered one of the most important breakthroughs in medical history. The drug discovery landscape was forever changed after the arrival of penicillin [97]. The following period, until the 1960s, witnessed the flourishing of antibiotic development, famously dubbed the golden age of antibiotics. However, the extensive search through soil microbes eventually led to the rediscovery of known compounds, prompting a shift in focus within the pharmaceutical industry during the 1960s and 1970s to the chemical tailoring of existing antibiotics to create successive generations of antibiotics in order to improve the efficacy or pharmacological properties [97]. Presently, traditional drug discovery strategies aim to identify the next breakthrough chemical or molecular entity with a novel mechanism of action to combat antimicrobial resistance. However, the journey from discovery to approval by the FDA, and similar regulatory entities, is fraught with challenges, with only a mere 1-2 drugs out of an initial pool of 10,000 compounds succeeding

[97]. Following target identification and validation, researchers develop high-throughput screening assays from compound libraries, aiming to identify promising hits. Each hit then undergoes further screening and potential chemical modifications to enhance its effectiveness before progressing to *in vitro* and *in vivo* testing in animal models [97]. Through these rigorous processes, only a handful of drug candidates from the initial pool of 10,000 compounds typically advance to clinical trials. Furthermore, the success rates for drugs entering Phase I clinical trials hover around a mere 10% [98].

There is currently a demand for alternative and innovative approaches in order to accelerate the complicated process of discovery of new drugs including high-throughput screening [99,100], omics technologies [101,102], structural biology [103,104], computational modeling [105–107] and more recently artificial intelligence [108–110].

Genome-scale metabolic models (GSMM) offer a comprehensive insight into cell metabolism. Historically, these global mathematical descriptions of cell metabolism have mostly been linked to the metabolic engineering of microbial cell factories given their potential to simulate global metabolic behavior and provide hints to guide experimental optimization of such organisms for the production of added-value compounds [111]. However, several recent examples have demonstrated their potential as an approach to the identification of new targets for drugs against pathogenic organisms [112–117].

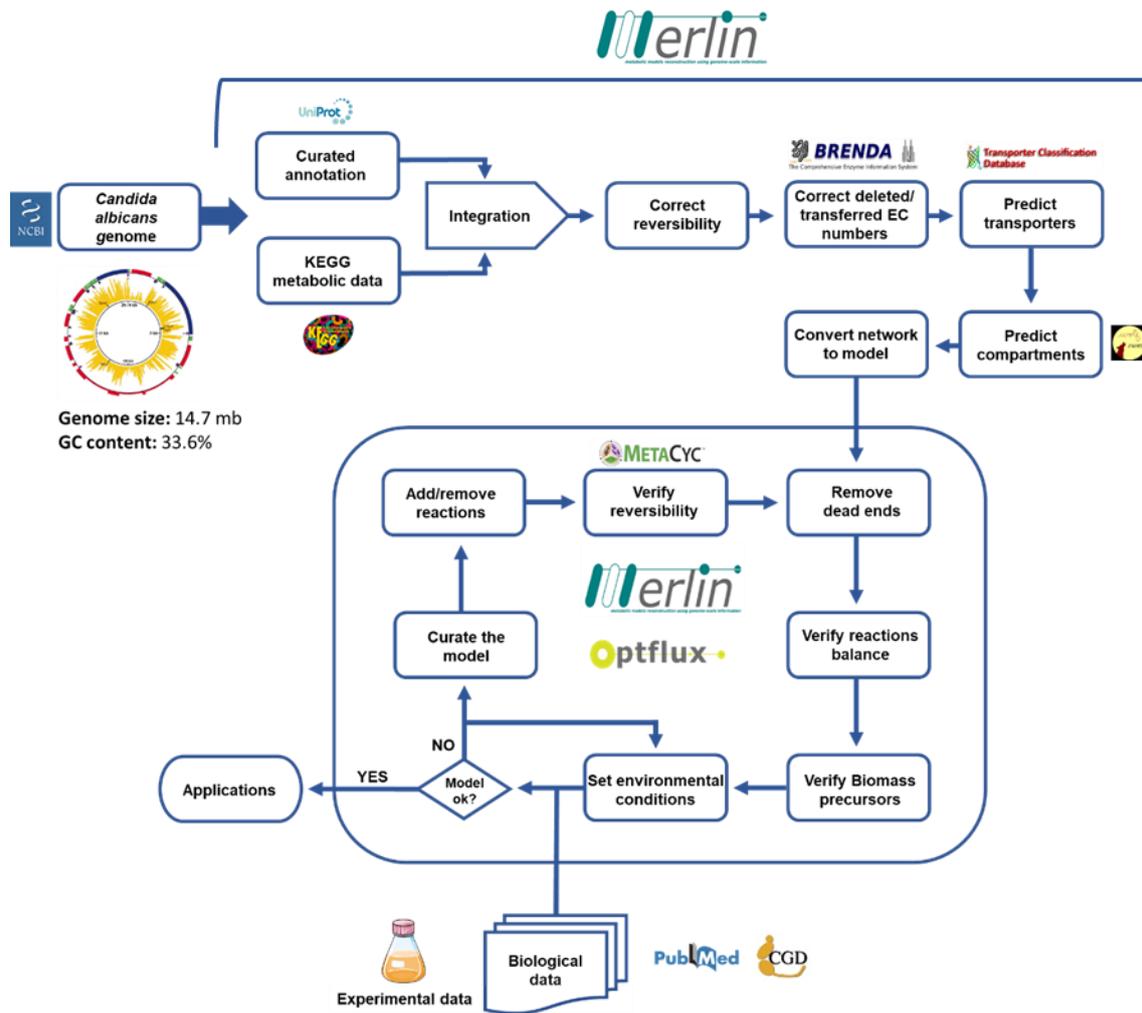
### I.3. Genome-Scale Metabolic Models

The first genome-scale metabolic model was reconstructed in 1999 for *Haemophilus influenzae* [118]. Since then, at an increasing pace more than 6000 distinct models have been reconstructed, for different species of bacteria, archaea, and eukaryotes, including *Homo sapiens* [119]. A GSMM represents a reconstruction of the complete metabolic network and may be applied to different types of cells, including, microorganisms, plants, and mammals. It may even represent entire tissues or body of a multicellular organism instead of a single cell [120]. These computational models encompass a whole set of stoichiometry-based, mass-balanced metabolic reactions within an organism using gene-protein-reaction (GPR) relationships that are formulated based on genome annotation data and experimentally obtained information [119]. Using optimization techniques like flux balance analysis (FBA), GSMMs can predict metabolic flux values for an entire set of reactions, allowing the prediction of microbial responses to different genetic or environmental stressors through *in silico* simulations, not offered by conventional metabolic pathway databases like KEGG [120,121] and circumventing the need for costly and time-consuming wet-lab experiments [122].

GSMM's offer a comprehensive insight into cell metabolism. Historically, these global mathematical descriptions of cell metabolism have mostly been linked to the metabolic engineering of microbial cell factories given their potential to simulate global metabolic behavior. This approach provides hints to guide the experimental optimization of such organisms for the production of added-value compounds, by predicting which genes may be manipulated (knocked-out, up- or down-regulated) to yield the desired effects [111]. Through analysis of simulation outcomes, it's possible to identify new metabolic features, uncovering previously unknown reactions or enzyme functions [119]. Additionally, several

recent examples have demonstrated the potential of GSMMs as an approach for the identification of new drug targets to be used against pathogenic organisms [112–117]. Using GSMMs, it is possible to predict an organism's essential genes, reactions, or those that, not being strictly essential, are required for growth under different environmental conditions, including for example the ones faced inside the human host. These predicted essential genes constitute promising new drug targets.

The process of reconstructing a GSMM involves more than 100 steps and can take from approximately six months to two years or more depending on the complexity of the organism, size of its genome, and degree of study [123]. Very often, this reconstruction process is progressive and can have continuous expansion and refinement over the years [124]. Recent software, such as *merlin*, seeks to solve this problem, offering options that allow optimizing reconstruction time [122]. The extensive process of GSMM reconstruction can be divided into 4 stages: draft reconstruction; refinement of reconstruction; network assembly; and model validation. The last 3 stages are cyclical and must be repeated successively until the model is refined and completed (Figure I.2)



**Figure I.2** - Methodology for the reconstruction of the *C. albicans* iRV781 metabolic model using *merlin*. Adapted from [111].

Thiele and Palsson provided a detailed methodology for the reconstruction of GSMMs, [123], which is followed by the *merlin* platform [122], including the 4 stages summarized below:

**Draft reconstruction:** initially, a draft reconstruction is generated based on the genome annotation of the target organism and biochemical databases. This automated process collects genome-encoded metabolic functions but may include false positives or miss crucial functions due to incomplete annotation. In this step manual curation is essential. The quality of genome annotation directly

impacts the subsequent reconstruction quality. In this stage, metabolic reactions catalyzed by the identified gene products are connected with the draft reconstruction by using the enzyme commission (E.C.) numbers and biochemical reaction databases, e.g., KEGG [125] or Brenda [126]. In *merlin*, this stage is optimized, as the platform provides tools to automatically retrieve the genome and perform a genome functional annotation [122]. Similarly, in *merlin*, the process of integrating transporter annotation into the draft network is optimized and can be performed using the state-of-the-art tool Transport Systems Tracker (TranSyT), developed by the same team [127]. Additionally, *merlin* enables performing and integrating compartment annotation into a draft genome-scale metabolic network using results provided by Psortb3, WoLFPSORT, and LocTree3, automatically loading all information regarding compounds, enzymes, reactions, and pathways into *merlin's* internal database.

**Refinement of reconstruction:** in this stage, the draft reconstruction undergoes thorough evaluation and refinement. Each gene and reaction entry is scrutinized to determine its relevance and completeness within the network. This manual assessment involves comparing metabolic functions and reactions with organism-specific literature, as not all annotations are highly reliable, and biochemical databases may list organism-unspecific reactions. Additionally, data on biomass composition, maintenance parameters, and growth conditions are gathered to support subsequent simulations. This manual evaluation is crucial to enhance the accuracy of the reconstruction. *merlin* provides a tool (e-Biomass equation tool) to automatically generate the biomass equation which, based on genome content, is able to estimate the protein, DNA, and RNA content of the cell. Additional detailed information must, then, be provided based on literature or experimental data. *merlin* also provides several tools to help the model refinement process, including correct reversibility, find unbalanced reactions, and find blocked reactions.

**Network assembly:** In the third stage, the reconstruction undergoes conversion into a mathematical format, and condition-specific models are established. This stage is largely automated, with defined system boundaries to transform the general reconstruction into a condition-specific model. The initial model may differ from the final model due to multiple iterations of validation and refinement. This final model is then used for simulating phenotypic behavior. Simulation constraints are established using functions in the Constraints-Based Reconstruction and Analysis (COBRA) Toolbox, facilitating easy adjustment of reaction constraints. In this process, it is important to meticulously track these changes to prevent simulation errors. *merlin* provides the Biological networks constraint-based *In silico* Optimisation (BioISO) tool, based on the COBRA and FBA frameworks, which can be used to evaluate biomass formulation and genome-scale metabolic network and find potential errors in the biomass formulation or gaps in the metabolic network.

**Model validation:** in the fourth stage of the reconstruction process, the network undergoes verification, evaluation, and validation. Common errors such as wrong reaction constraints, missing transport reactions, missing exchange reactions, metabolites not consumed or produced, and lack of transport for metabolites between compartments should be identified in this stage and corrected. The metabolic model created in the third stage is tested for its capacity to synthesize all the biomass precursors allowing the identification of missing metabolic functions, known as network gaps. This iterative process highlights the ongoing refinement of the reconstruction. Determining when to conclude the process and consider the reconstruction "finished" is crucial and typically based on the defined scope and purpose of the reconstruction. *merlin* facilitates this process with the BioISO tool, however, it's common to utilize other external software's such as Optflux [128] because it is more practical and offers more testing features.

### **I.3.1. Current challenges in GSMM reconstruction and exploitation**

While the methodology for reconstructing genome-scale metabolic models has become standardized, eukaryotic organisms, such as fungi present ongoing challenges due to their large genomes and complexity [129]. Despite efforts to approximate reality as closely as possible, the complexity inherent to these large models inevitably introduces errors, leading to minor discrepancies in predictions. These errors may include incorrect assignment of GPR associations, inaccuracies in reaction directionality or reversibility, inconsistencies in stoichiometric parameters, the absence of crucial reactions, and inaccuracies in biomass composition [129]. Moreover, current modeling tools are limited in addressing network properties beyond metabolism at a global scale, such as the regulatory network which has a big impact on the regulation of cell metabolism. Another limitation of GSMMs is not taking into account the time or the mechanisms required for cells to adapt to genetic perturbations or environmental variability [129]. Furthermore, interactions between yeast and other microorganisms, as well as the secretion of compounds influencing the surrounding environment, are often disregarded [130]. Despite these limitations, genome-scale metabolic reconstructions have proven highly effective in discovering new drug targets [112–117]. Once a model is constructed, predicting drug targets becomes relatively straightforward. However, the experimental validation of these targets and the identification of effective drugs pose an even more difficult challenge.

## I.4. Molecular Docking

The exponential growth in available genome sequences witnessed over recent decades has substantially expanded our access to an increasing number of new potential therapeutic targets for drug discovery. At the same time, advances in spectroscopy techniques have played a pivotal role in elucidating the structures of numerous proteins, including information on molecular details on protein interactions with solutes, such as substrates, co-factors, or inhibitors [131]. These advances have coincided with the evolution of computational strategies aimed at guiding drug discovery endeavors. Since the 1980s, computer technologies have revolutionized the drug discovery landscape, giving rise to innovative approaches such as computer-guided drug design and virtual drug screening approaches [131]. The ability to predict ligand binding modes holds immense value, facilitating the optimization and suggestion of novel ligands, in fact, with the rise in frequency of multidrug-resistant clinical isolates, the identification of new drug targets and new drugs is crucial to overcome the increase in therapeutic failure.

Currently, the process of identifying a drug target, synthesizing an active compound with reduced toxicity, bioavailability, and cost-effective synthesis is a time-consuming, complex, and expensive process [132]. Molecular docking is an attractive computational technique that can be used to understand drug biomolecular interactions for rational drug design and discovery, and currently, plays an important role in the discovery and development of new drugs [133]. The aim of this technique is to predict ligand-receptor complex structure interaction using computational methods, being one of the most applied virtual screening methods.

Molecular docking applied to protein-ligand interactions consists of two interrelated steps: 1) sampling conformations of the ligand in the active site of

the protein; and 2) ranking these conformations via a scoring function. In an ideal scenario, sampling algorithms should be able to reproduce the experimental binding mode and the scoring function should be able to distinguish correct from incorrect poses or the binders from the non-binders in the pool of poses generated by the sampling process in a reasonable computation time [131,134]. A molecular docking protocol needs some general requirements, common to all the different docking programs:

- The availability of the three-dimensional structure of the molecular target. This structure can be obtained experimentally (by X-ray crystallography, NMR, or cryo-EM) or predicted based on computational techniques (homology modeling or AlphaFold) [131].
- A database containing existing or virtual compounds for the docking process [135].
- Sampling and scoring methods, which require a computational framework for its efficient exploitation [135].

Another very important aspect of molecular docking involves determining the precise location of the binding site. Preselecting the precise binding site location prior to docking processes is crucial and significantly enhances docking efficiency. This selection procedure often relies on available information regarding substrate or co-factor binding sites within the target protein. Ideally, such information is acquired experimentally, for example through co-crystallization of the protein with its ligand(s). Alternatively, binding sites can be predicted, such as by comparing the target protein with a family of proteins sharing similar functions or with proteins co-crystallized with other ligands. In cases where knowledge about binding sites is entirely lacking, docking can serve as a tool to gain insights into their location by conducting blind docking experiments [134]. However, there are some cavity detection tools, such as GRID

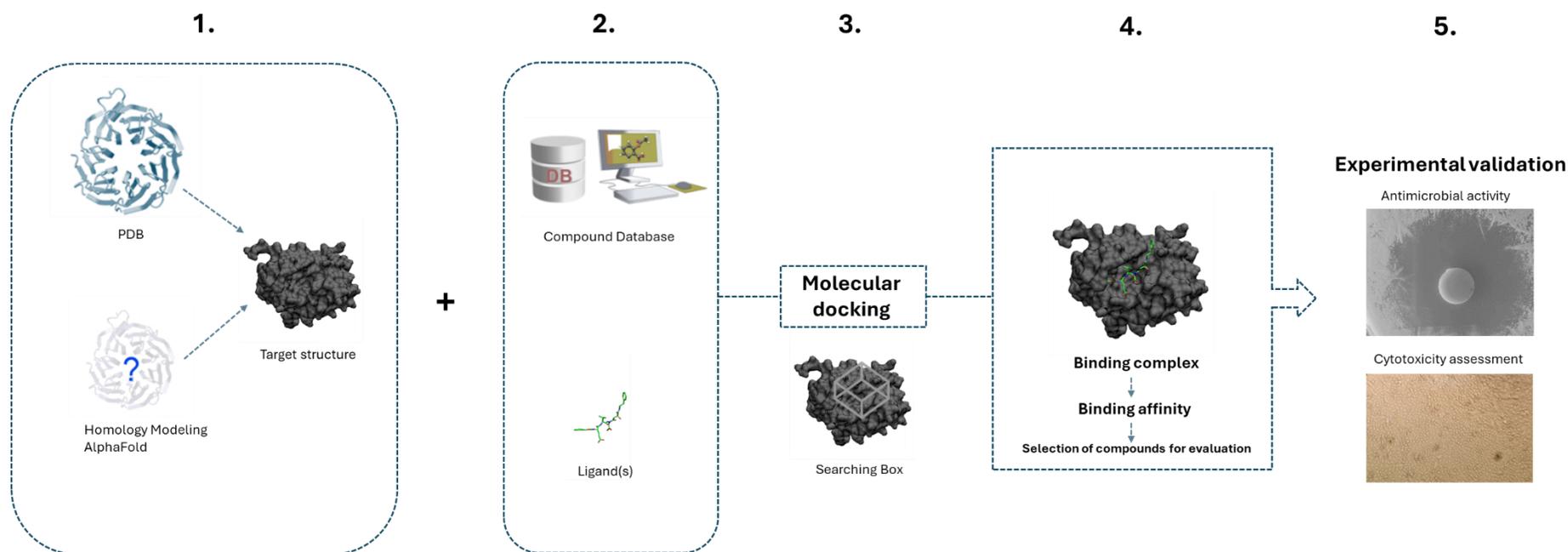
[136], POCKET [137], SurfNet [138], PASS [139], and MMC [140] that can be used to identify putative active sites in an attempt to overcome this problem.

Molecular docking pipelines for virtual drug screening consist of 4 steps: preparing the ligand; preparing the receptor; defining the search box and running the docking calculation. An additional step involving the analysis and interpretation of the obtained results is usually also considered (Figure I.3). In the following sections, the methodologies and approaches available to pursue each step are presented and discussed. Protein structure prediction methods such as homology modeling and AlphaFold are examined for their importance in achieving reliable docking predictions. Finally, applications, current challenges, and future perspectives of molecular docking, namely in virtual drug screening, are discussed.

#### **I.4.1. Docking methodologies**

Docking methods can be divided into 3 categories according to the degrees of flexibility of the molecules involved in the calculation:

- **Rigid docking:** in this methodology, the ligand and the receptor are both treated as rigid bodies resulting in very limited search space considering only three translational and three rotational degrees of freedom [134]. Due to its limitations, this methodology is no longer widely used, it can still be found in programs such as early versions of DOCK [141], FLOG [142], FTDOCK [143].



**Figure I.3** – Virtual Drug screening methodology using molecular docking. 1. Receptor structure selection and preparation for molecular docking, which can be retrieved from a 3D structure available in PDB, or from *in silico* predicted structures using homology modeling or from the AlphaFold database. 2. Ligand preparation for molecular docking, which can be a group of compounds from a compound database, or specific selected ligand(s) of interest. 3. Searching box design, defining the special restraints within the target structure where the molecular docking calculation will be performed, using one of the different docking programs available. 4. Analysis and selection of the best-hit results based on criteria such as binding energy calculated by scoring functions. 5. *In vivo* validation of the predicted activity (e.g. antimicrobial activity) and potential cytotoxic effects.

- **Semi-flexible docking:** in this methodology, the ligand is treated as flexible while the receptor is kept rigid during docking. This is the most common methodology, because of the trade-off between accuracy and computation time, since considering both ligand and receptor as flexible has very high computational costs [134]. Almost all the docking programs use this methodology, including AutoDock Vina [144].
- **Flexible docking:** it considers the flexibility of both the ligand and receptor, being the closest scenario to reality since both the ligand and receptor change their conformations to form a minimum energy perfect-fit complex. However, incorporating the receptor flexibility in docking is significantly challenging due to its high computational expense, which prevents this method from being used in the screening of large chemical databases [134]. Some programs have tried to implement this methodology, and they have options for the user to consider the flexibility of the receptor, including GOLD [145], AutoDock Vina [144], ICM [146], AutoDock 4 [147], DOCK [148] and FlexE [149].

#### **I.4.2. Compound and receptor preparation for docking**

When starting a structure-based docking project, the optimal scenario is starting from a high-resolution ligand-bound structure. Ligand-bound structures are favored due to the more clearly defined geometries of the binding pocket in the bound state compared to the unbound state [150]. Furthermore, it is advantageous to utilize small, enclosed binding pockets that closely complement a ligand rather than large, flat, and solvent-exposed binding sites [150]. Once a suitable structure has been identified as a starting point, it is important to

meticulously assess for any structural anomalies and rectify them as needed. For instance, certain protein structures may have been determined in a mutant form. In such cases, it is advisable to revert mutated structures back to their predicted wild-type configuration, especially if the mutations occur within the targeted ligand site. Additionally, if any atoms are found to be inadequately solved and missing from the structure, particularly in proximity to the binding site, their predicted position should be added to the structure [150]. Furthermore, it is essential to account for any water molecules enclosed within the binding pocket, as they may partake in interactions that influence side-chain conformations or provide additional hydrogen-bonding sites. Similarly, cofactors such as heme groups or metal ions should be considered, as they can also play crucial roles in the binding environment. [150]. Buffer components, specific to the crystallization conditions should be removed. Additionally, molecular docking requires to have both the protein binding site and the putative ligands correctly protonated. If this is not the case, it may not be possible to even dock compounds, due to steric or electrostatic clashes [151]. Usually, ligand and protein are processed and prepared separately. In virtual screening, ligands are retrieved from sources other than the Protein Data Bank (PDB), such as compound databases like PubChem, Zinc15, Sigma Aldrich, and others. In these instances, the procedure may vary but sometimes involves reconstructing the molecule from its Simplified Molecular Input Line Entry System (SMILES) format or sketching the molecule and saving it in an SDF/MOL file format [152].

#### **I.4.3. Sampling algorithms**

Docking a ligand into a receptor pocket involves six degrees of translational and rotational freedom per atom from both the ligand and protein. Consequently, this yields a huge number of possible binding modes between two molecules [135]. Thus, a large computational power is needed to sample the conformational space.

The sampling algorithms seek to find a balance between performance and required computation. The following sampling algorithms are among the most common:

- **Matching algorithms:** in these algorithms, the protein and the ligand atoms are superimposed into predefined binding-site points. Shape complementarity is measured by Fast Fourier Transform or Geometric Hashing functions [153] and ligand conformations are then defined by the distance matrix between the protein and the corresponding ligand atoms, taking into consideration chemical properties, like hydrogen-bond donors and acceptors [134]. These algorithms allow a fast generation of results, with no high computational power requirement and are used in programs like DOCK [154], FLOG [142], LibDock [155] and SANDOCK [156].
- **Fragment-based algorithms:** incremental construction algorithms are more common in this category. In this methodology the ligand is divided into several fragments by breaking its rotatable bonds and one of these fragments is selected to dock into the active site first, working as an anchor, usually the largest fragment or a part with a known important role in protein interaction. The remaining fragments are then added incrementally resulting in the generation of different orientations in order to fit in the active site according to the ligand flexibility [134]. These algorithms are used in the programs DOCK 4.0 [157], FlexX [158], Hammerhead [159], SLIDE [160] and eHiTS [161]. There are other available fragment-based algorithms using different methodologies, such as Multiple Copy Simultaneous Search (MCSS) [162] or LUDI [163]. MCSS makes 1,000 to 5,000 copies of a functional group which are then randomly placed in the binding site of interest; while LUDI focuses on the hydrogen bonds and hydrophobic contacts which could be formed between the ligand and protein.

**Stochastic methods:** These algorithms explore the conformational space by randomly modifying a ligand or a population of ligand conformations. This class includes Monte Carlo and genetic algorithms. In the Monte Carlo approach, ligand poses are generated through bond rotation, rigid-body translation, or rotation. Subsequently, the resulting conformation is tested with an energy-based selection criterion. If it passes the criterion, it is saved and further modified to generate the subsequent conformation [134]. The iterations will proceed until the predefined quantity of conformations is collected. This algorithm is used in programs like AutoDock [164], AutoDock Vina [144], ICM [146] and QXP [165]. Genetic algorithms utilize binary strings, referred to as genes, that encode the degrees of freedom of the ligand. In turn, these genes constitute the so-called chromosomes, which represent the poses of the ligand. Mutations represent random changes to the genes while crossover entails exchanging genes between two chromosomes. These genetic operations result in the generation of new ligand structures. Subsequently, these structures are evaluated using a scoring function, and those surpassing a predetermined threshold are retained for the next generation [134]. These algorithms can be found in programs like AutoDock [166], GOLD [145], DIVALI [167] and DARWIN [168].

#### **I.4.4. Scoring functions**

During the docking process, a sampling algorithm generates a vast number of different poses for each ligand. The scoring functions play a critical role in assessing the quality of these docking poses, helping us identify the most relevant ligand conformations [169]. A reliable scoring function must be able to distinguish the experimentally observed binding modes from all the other poses

discovered by the search algorithm. Those poses should be associated with the lowest binding energy and accurately discriminate between ligands that bind and those that do not, and they should predict the absolute binding affinity, effectively ranking compounds based on their potential [132,169]. Additionally, scoring functions must be fast enough to assess millions of compounds within a set computational timeframe. Popular scoring functions have an adequate balance between accurately estimating binding energy and computational efficiency [132]. Scoring functions can be divided into 3 categories:

- **Force-field based scoring functions:** binding energy is calculated by the sum of the physical atomic interactions: electrostatic potential, described by the Coulomb function where a distance-dependent dielectric may be introduced to mimic the solvent [131]; and van der Waals interactions described by the Lennard-Jones potential. Additional terms have been added to the force-field scoring functions, such as solvation terms, hydrogen bonds, and entropy contributions [134]. Programs, such as DOCK [154], GOLD [145], AutoDock [166] use this type of scoring functions.
- **Empirical scoring functions:** in these types of functions binding energy is calculated by the sum of several energy components including van der Waals energy, electrostatics, hydrogen bond, desolvation, entropy, hydrophobicity, etc [170] and uses known binding affinity data of experimentally determined structures to determine the coefficients associated with the functional terms. Each energy component is then multiplied by these coefficients and then summed up to give a final score [134]. **Empirical scoring functions** are used by LUDI [163], PLP [171], ChemScore [172], ID-Score [173], GlideScore [174] and Autodock Vina [144].

**Knowledge-based scoring functions:** this type of functions, uses statistical analysis of ligand-protein complex crystal structures from a database. From this database, the frequencies of ligand-protein atom pair contacts are computed and converted into an energy component, that is then used to evaluate a pose and assign a score to it [131]. The score is calculated by favoring preferred contacts and penalizing repulsive interactions between each atom in the ligand and the protein within a given cutoff [134]. DrugScore [175], GOLD/ASP [176], PMF [177], SMOG [178] and Bleep [179] use this type of scoring functions.

#### **I.4.5. Autodock Vina**

There are several docking programs as mentioned in the previous sections, each resorting to different methodologies, algorithms, and purposes. Autodock Vina [144] is one of the fastest and most widely used open-source docking engines, partly due to its ease of use, but also due to its speed, up to 100x faster than other docking engines. It is based on a simple scoring function and rapid gradient-optimization conformational search. Vina uses Monte-Carlo as sampling algorithm treating docking as a stochastic global optimization of the scoring function, and a new specific hybrid empirical and knowledge-based scoring function considered by the creators more like a machine learning than directly physics-based in its nature. This scoring function extracts empirical information from both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements [180]. For its input and output, Vina uses PDBQT molecular structure file format and provides a tool, MGLTools where the PDBQT files can be generated, edited, and viewed. It is an open-source free software and allows the user to take advantage of multiple CPUs or CPU cores to significantly shorten its running time.

#### **I.4.6. Homology Modeling**

Homology Modeling is a computational structure prediction method that goes hand in hand with molecular docking. Since molecular docking requires a protein structure to be able to perform the docking, when the structure is not available, resorting to prediction techniques is an attractive alternative to its crystallographic determination. Despite notable advancements in the quality of 3D structures derived from crystallography methods, determining the structure of a novel protein remains an arduous, time-consuming, and costly endeavor. In contrast, homology modeling offers a fast method to predict protein structures within a matter of days [181].

Moreover, high-resolution 3D structures are lacking for certain protein classes, particularly membrane proteins, due to the challenges associated with their purification and crystallization [181]. Given the role membrane proteins play in drug internalization by pathogens, as well as their importance in cell structure and survival, they represent promising therapeutic drug targets. Therefore, progress in elucidating their protein structures is vital for advancing the drug discovery pipeline.

Homology modeling is based on two foundational principles: first, that protein 3D structures are determined by their amino acid sequences, and second, that there exists conservation in the 3D structure relative to a primary sequence [182]. For this second point, it's known that evolutionary-related proteins share a similar structure, and consequently, sequences with high similarity fold into identical structures, facilitating the prediction of a protein's structure solely based on its amino acid sequence, as long as homologous proteins with established structures are available [181]. A homology modeling methodology typically can be explained in six steps:

- **Template selection:** Every homology modeling project starts with the search for a suitable template to be used for structure prediction. During this phase, the target sequence is used to perform sequence–sequence alignment to identify the most closely related available proteins. The Basic Local Alignment Search Tool (BLAST) is commonly utilized for this purpose, and this search is narrowed down by selecting only crystal structures available in the Protein Data Bank (PDB) (<http://www.wwpdb.org/>) database [182]. Once template candidates are identified, a selection process ensues to determine the most optimal structures. Sequence similarity between the template and the target sequence stands as the primary criterion. Additionally, factors such as phylogenetic proximity, ligand binding, and experimental structure resolution are also taken into consideration. The predictive accuracy of the homology modeling depends largely on the degree of similarity between the model and template sequences, with a similarity threshold of 25% considered the limit [181]. In cases where homology is low, alternative alignment methods may be employed to minimize shifts and gaps. These methods include profile-profile alignments, Hidden Markov Models (HMMs), and position-specific iterated BLAST (psi-BLAST) [182]. In some cases, more than one structure may be selected as template, for example, different regions of the target sequence may have an ideal template from different proteins or structures.
- **Sequence alignment and correction:** After selecting the most appropriate templates, the alignments undergo correction and optimization to ensure accurate alignment of protein sequences and construct the entire backbone. Multiple alignment tools, such as ClustalW [183], are commonly employed for this purpose. In regions where sequence identity is particularly low, aligning template and model sequences can pose challenges. One approach to address this is by increasing the identity

percentage through additional sequences from other homologous proteins [184]. Accurate alignment of sequences is crucial in homology modeling and can significantly enhance the quality of the results. Therefore, meticulous checks and corrections during the alignment process are essential for ensuring high-quality 3D protein structure construction [182].

- **Model building:** several tools can be used to generate 3D models in homology modeling. However, Modeller [185] is the most commonly used. Modeller uses a spatial restraint approach to generate three-dimensional models by optimally satisfying spatial restraints derived from the alignment and expressed as probability density functions for the restrained features [185]. These restraints include homology-derived restraints on the distances and torsional angles, stereochemical restraints such as bond length and bond angle parameters, and optional restraints derived from experimental data [182].
- **Loop modeling:** most loops have no function other than to connect secondary structural elements and typically exhibit low sequence conservation, leading to significant variation in conformation among related proteins. As a result, predicting changes in loop conformation is a difficult task. Two main approaches are commonly employed for loop modeling: knowledge-based and energy-based methods. Knowledge-based loop modeling assesses the orientation and separation of backbone segments flanking the target region and searches the PDB repository for loops of similar length with comparable end-point geometry [186]. Energy-based loop modeling utilizes an energy function to evaluate loop quality and identify the optimal loop conformation based on statistical preferences of atoms for several geometries, derived from a database of known protein structures [187].

**Side-chain modeling:** side-chain modeling plays an important role in protein structure prediction since protein side chains make a dominant

contribution to molecular recognition. Typically, protein side chains occupy a limited number of energetically favorable conformations known as rotamers [181]. The general strategy of modeling programs is to align the target side chains as closely as possible to their counterparts in the template structure. However, this approach is not always possible due to variations in amino acid composition between the target and template sequences. To address this problem, side-chain modeling adopts a knowledge-based approach, utilizing libraries of rotamers to position side chains correctly [181,186].

- **Model evaluation and validation:** after generating a model, it is necessary to evaluate the success of the process. Different scores can be used to evaluate and validate the models. Distance-based and contact-based similarity scores are used in experimental evaluation of homology models. Other scores can be used for quality check such as physics-based, knowledge-based, and combined scores [182]. For example, the Discrete Optimized Protein Energy (DOPE) score is implemented in MODELLER and this knowledge-based quality score evaluates the model based on atomic distance-dependent statistical potential calculated from a set of native crystallographic structures and can be used to estimate the obtained model quality [188]. Besides those scores, it is also important to perform a manual inspection and evaluate the residue placement and the interaction of neighboring residues. After taking into account all these aspects, subsequent analyses are carried out on the best model.

#### **I.4.7 AlphaFold**

AlphaFold, developed by Google DeepMind, surged in 2018 as an alternative and revolutionary approach for protein structure prediction. Unlike conventional homology modeling, AlphaFold uses the power of machine learning to predict a

protein's three-dimensional structure from its amino acid sequence. While homology modeling relies heavily on sequence similarity with known proteins, AlphaFold utilizes machine learning that incorporates physical and biological knowledge about protein structure to predict protein structures with remarkable accuracy, even in cases where no similar structures are known [189].

The AlphaFold database is currently open access, containing over 200 million protein structure predictions available for research purposes. It surges as an important alternative approach for accessing the protein 3D structure, especially in cases where conventional prediction methods fall short of atomic accuracy, for example when no homologous structure is available [189]. Another compelling aspect of AlphaFold is its speed and scalability. While conventional homology modeling techniques can be time-consuming, AlphaFold can rapidly generate accurate structural predictions, accelerating research in drug discovery, protein engineering, and molecular biology and contributing to a better understanding of human health, disease, and the environment.

The AlphaFold algorithm combines physical and geometric biases, learning from PDB data with high efficiency while accommodating the complexity and diversity of structural data, being trained to produce structures that align with constraints implicit in the sequence, such as specific stoichiometry, ligands, or ions. In a simple definition, the AlphaFold algorithm is trained to produce the protein structure most likely to appear as part of a PDB structure [189].

AlphaFold usage has increased within the scientific community over the past five years, with more than 1,000 entries in the PubMed database between 2021 and 2024, being applied in numerous successful cases, the most notable being SARS-CoV-2 protein structure prediction during the pandemic [190].

Google DeepMind is currently developing a new service for the scientific community, the AlphaFold Server, still in beta test and launched on May 8, 2024.

This server is powered by the newest AlphaFold 3 model and serves as a platform to modulate complex binding structure, including options for the user to modulate the structure of complexes, including proteins, nucleic acids, biologically common ligands, and ions, and modified residues [191]. AlphaFold 3 demonstrates greater accuracy in protein-ligand and protein-nucleic acid interactions than state-of-the-art docking tools/predictors [191]. Despite being in the beta phase only, it will certainly revolutionize the future of virtual drug screening and the drug discovery process.

#### **I.4.8. Application of molecular docking to virtual drug screening**

As mentioned before, molecular docking enables the modeling of interactions between small molecules and proteins at the atomic level, allowing for the characterization of small molecule behavior within the binding site of their predicted target and the elucidation of fundamental biochemical processes. It is generally employed for three different purposes: binding mode prediction, virtual drug screening, and free energy of binding prediction [134]. Virtual drug screening, in particular, is an important application of molecular docking, as this procedure saves time and money, and being widely used by many pharmaceutical industries [132].

In virtual drug screening, basically, a group of compounds from a database are docked against a target structure, allowing for large-scale prediction of which compounds bind with greater affinity to the target protein, thereby identifying which are the best candidates as potential inhibitors. There are many successful cases of application of this technique for drug discovery. For example, Boehm et al. used de novo design for a DNA gyrase and successfully obtained several new inhibitors for this antibacterial target [192]. Schames et al., using molecular docking, discovered a new binding site for drugs in the HIV 1 Integrase, which

resulted in the development of drugs with a new mode of inhibition for AIDS therapy [193]. Becker et al. screened 40,000 compounds from a database, using molecular docking, against the serotonin receptor (5HT1A). Based on this approach, 78 virtual hits were discovered and experimentally tested. The most potent molecule was selected as a lead molecule for further optimization reaching Phase I clinical trial very quickly [194]. Moreover, molecular docking also plays a prominent role in the initial prediction of drugs' nucleic acid binding properties, which can be useful in predicting their cytotoxicity. This is particularly valuable not only for developing new antimicrobial drugs but also for understanding the molecular mechanisms underlying anticancer effects [133].

#### **I.4.9. Current challenges in molecular docking**

One of the primary challenges encountered in the field of docking is receptor flexibility [132,152]. Proteins can assume various conformations depending on the ligand they bind to. Arguably, considering the receptor as flexible is the most appropriate approach to studying protein-ligand complex behavior [195]. However, due to its high computational demand, this methodology is not commonly utilized. As a result, docking performed with a rigid receptor represents only a single receptor conformation, often resulting in false negatives where the ligand later proves to be active [132]. There have been some advances in addressing this limitation, and, in fact, some programs already have options for performing docking with a flexible receptor.

Another significant challenge in docking is the lack of confidence in the accuracy of scoring functions to provide precise binding energies. Scoring functions often struggle to accurately predict certain intermolecular interactions, such as solvation effects and entropy changes. Additionally, some intermolecular

interactions are rarely considered [132]. Currently, machine learning is being used to improve docking and scoring functions [196].

Ligand preparation also represents a big challenge in this technique, having a big impact on docking results. Most databases contain molecules in their neutral forms, whereas, under physiological conditions, they are predominantly ionized. During the docking process, it's common to ionize ligands before calculating docking scores. This task involves removing or adding hydrogens, which is relatively straightforward. However, when dealing with tautomeric and/or protomeric states of the molecules being docked, significant discrepancies may persist [133]. Addressing water molecules or metal ions in the binding pocket can also be a challenge. The high importance of metal cofactors on many protein families is well known, this is particularly important when modeling interactions with zinc, calcium, and magnesium, which are prominent metal ions in drug discovery [152]. Autodock tool developers have put some effort into this issue and yielded a more accurate solution to pose and scoring in ligand-zinc interaction, for example [197,198].

Sometimes, x-ray crystal structures lack coordinate information for hydrogen atoms. The absence of precise hydrogen positions can lead to inaccuracies in identifying water molecules, potentially impacting receptor-ligand binding. Furthermore, predicting the number of water molecules in the binding pocket that might be displaced by potential ligands, and understanding how ligand binding disturbs the hydrogen bonding network, presents challenges [132]. Most docking programs can now consider the presence of water molecules during calculations and some methods have made advances in accurately representing water molecules in proteins, including Monte Carlo [199], molecular dynamics of water on the binding site implemented by Schrödinger [200], "Attachment" of water molecules to ligands as additional torsions [201], free energy perturbation methods, water displacement as implemented by PLANTS [171], QM/MM

hybrid methods [202], COSMO solvation and semi-empirical charges for ligands [203], “hydrated docking” scripts used by Autodock [204], protein-centric and ligand centric hydration implemented by Rossetta [205], Water docking using Autodock Vina [206], WScore [207] and grid inhomogeneous solvation theory applied by Autodock [208].

In recent decades, molecular docking has seen significant advances. Nowadays, docking millions of compounds has become a routine procedure, based on the existence of extensive compound databases. The development of fast docking programs and software platforms enabling the automation of virtual screening processes made its use more widespread. While there is still much room for improvement, the ongoing advancements in physics-based scoring methodologies and machine-learning techniques hold great promise for the future [209].

## II. Reconstruction of Genome-scale metabolic models for pathogenic fungi

Journal papers:

Chapter II.1 - Viana R, Dias O, Lagoa D, Galocha M, Rocha I, Teixeira MC. Genome-Scale Metabolic Model of the Human Pathogen *C. albicans*: A Promising Platform for Drug Target Prediction. *J Fungi (Basel)*. 2020 Sep 11;6(3):171. doi: 10.3390/jof6030171. PMID: 32932905; PMCID: PMC7559133.

Chapter II.2 - Viana R, Couceiro D, Carreiro T, Dias O, Rocha I, Teixeira MC. A Genome-Scale Metabolic Model for the Human Pathogen *C. parapsilosis* and Early Identification of Putative Novel Antifungal Drug Targets. *Genes (Basel)*. 2022 Feb 5;13(2):303. doi: 10.3390/genes13020303. PMID: 35205348; PMCID: PMC8871546.

Chapter II.3 - Viana R, Carreiro T, Couceiro D, Dias O, Rocha I, Teixeira MC. Metabolic reconstruction of the human pathogen *C. auris*: using a cross-species approach for drug target prediction. *FEMS Yeast Res*. 2023 Jan 4;23:foad045. doi: 10.1093/femsyr/foad045. PMID: 37852663.

Chapter II.4 - Viana R, Newton W, Dias O, Coelho C, Teixeira MC, “Unveiling new features of the human pathogen *C. neoformans* through the construction and exploitation of a dedicated Genome-Scale Metabolic Model”, manuscript in preparation.

## II.1. Genome-scale metabolic model of the human pathogen *C. albicans*: a platform for drug target prediction

### II.1.1. Abstract

*C. albicans* is one of the most impactful fungal pathogens and the most common cause of invasive candidiasis, which is associated with very high mortality rates. With the rise in frequency of multidrug-resistant clinical isolates, the identification of new drug targets and new drugs is crucial to overcome the increase in therapeutic failure. In this study, the first validated genome-scale metabolic model for *C. albicans*, iRV781, is presented. The model consists of 1221 reactions, 926 metabolites, 781 genes, and four compartments. This model was reconstructed using the open-source software tool *merlin* 4.0.2. It is provided in the well-established systems biology markup language (SBML) format, thus, being usable in most metabolic engineering platforms, such as OptFlux or Cobra. The model was validated, proving accurate when predicting the capability of utilizing different carbon and nitrogen sources when compared to experimental data. Finally, this genome-scale metabolic reconstruction was tested as a platform for the identification of drug targets, through the comparison between known drug targets and the prediction of gene essentiality in conditions mimicking the human host. Altogether, this model provides a promising platform for global elucidation of the metabolic potential of *C. albicans*, with an expected impact in guiding the identification of new drug targets to tackle human candidiasis.

**Keywords:** *C. albicans*, global stoichiometric model, drug targets, metabolic reconstruction, gene essentiality

## II.1.2. Introduction

In the last few decades, a significant increase in nosocomial fungal infections has been observed, and *Candida* species are by far the most common cause of invasive fungemia in humans [210,211]. Among *Candida* species, *C. albicans* is the main etiological agent of invasive candidiasis [212,213], being associated with high mortality rates [212,213]. Together with its virulence traits [214,215], its ability to acquire drug resistance [216–218] makes this opportunistic pathogen a severe threat.

Only three classes of antifungal drugs are licensed to treat *Candida* infections (azoles, echinocandins, and amphotericin B), and only fluconazole and echinocandins are recommended as first-line agents [219]. Currently, there has been a rise in the frequency of multidrug-resistant clinical isolates, and therapeutic options are running low. This is true for *C. albicans*, but even more so for other emerging non-*albicans* *Candida* species, such as *C. glabrata*, *C. krusei*, and *C. auris*. For example, in recent studies, almost 40% of the *C. glabrata* isolates shown to be resistant to at least one echinocandin were also resistant to fluconazole [12]. In non-*albicans* pathogenic *Candida* species, the scenario is even more frightening, as several of them display either intrinsic or easily acquired resistance to several of the available antifungal agents. For example, in a recent case, *C. auris* isolates were identified as resistant to the three classes of available antifungal drugs, further raising public concern on the future efficacy of current antifungal therapeutic options [14]. The identification of new drug targets and new drugs is crucial to overcome the increase in therapeutic failure.

Genome-scale metabolic models have the potential to provide a holistic view of cell metabolism. Historically, these global mathematical descriptions of cell metabolism have mostly been linked to metabolic engineering of microbial cell factories, given their potential to simulate global metabolic behavior and provide hints to guide experimental optimization of such organisms for the production

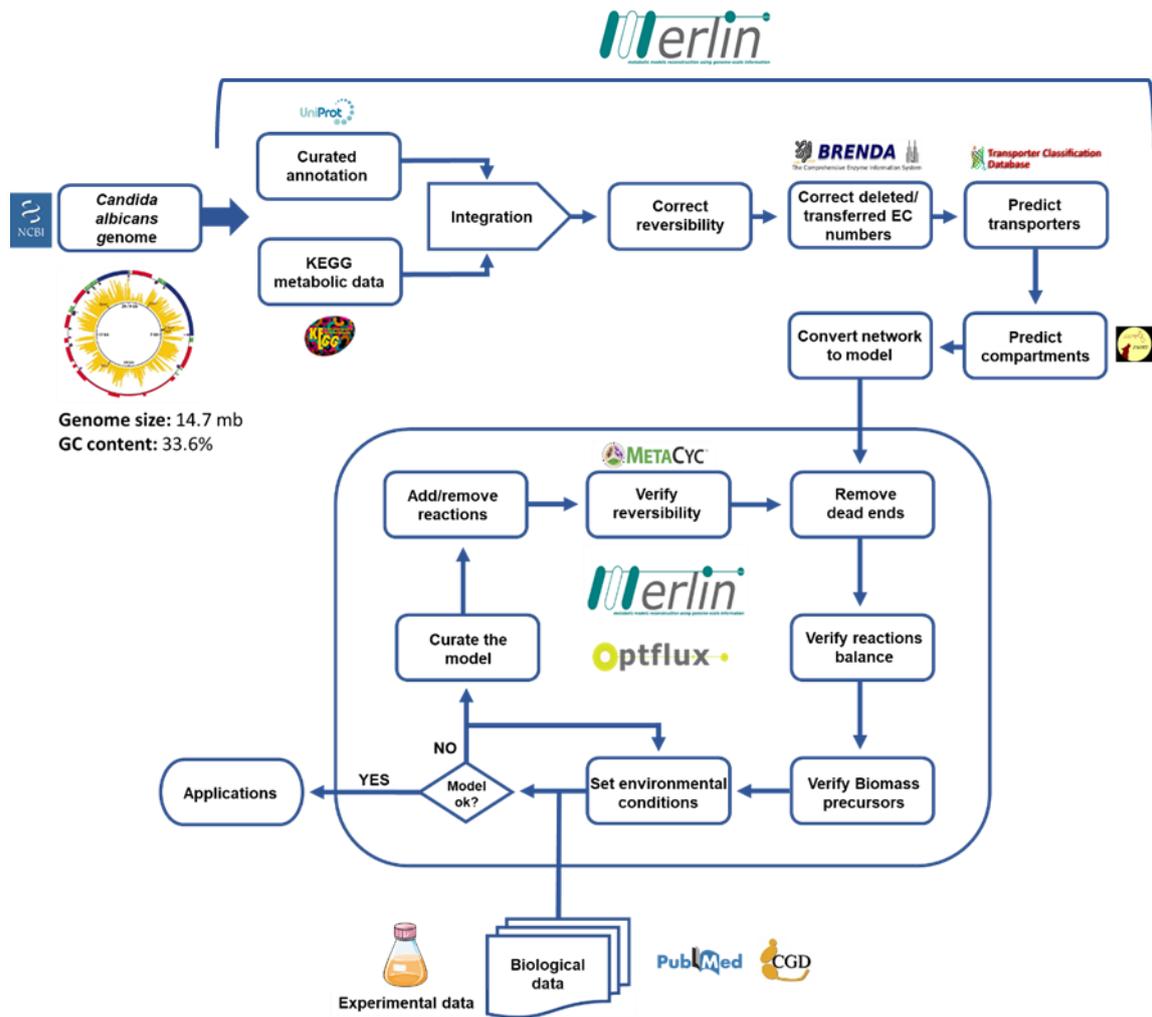
of added-value compounds [111]. However, recent examples have shown the potential of these models in the quest for novel drug targets in pathogenic organisms [112–116]. For example, Abdel-Haleem et al., 2018, described the reconstruction of genome-scale metabolic models for five life cycle stages of *Plasmodium falciparum*, enabling the identification of potential drug targets that could be used as both anti-malarial drugs and transmission-blocking agents [117].

Here, we present the first validated *in silico* genome-scale metabolic reconstruction of *C. albicans*, the iRV781. This model is provided in the well-established SBML format and can easily be read in most metabolic engineering platforms such as OptFlux [128] and COBRA [220]. The model validation procedure is detailed, and an evaluation of the potential of this model for the search for new drug targets in this fungal pathogen is put forward.

### **II.1.3. Materials and Methods**

#### **II.1.3.1. Model development**

The *C. albicans* iRV781 genome-scale metabolic model was developed following the methodology represented in Figure II.1, using *merlin* 4.0.2 [122] for the reconstruction process, as described elsewhere [221], and OptFlux 3.0 [128], for the curation and validation of the model. All predictions were performed using the IBM CPLEX solver. *merlin* is a platform that allows performing the reconstruction process of metabolic models semi-automatically, providing a user-friendly interface that assists the user in the manual curation process [122].



**Figure II.1** - Methodology for the reconstruction of the *C. albicans* iRV781 metabolic model. Adapted from [111].

### II.1.3.2. Genome Annotation and Assembling the Metabolic Network

The genome sequence of the reference strain *C. albicans* SC5314 was obtained from NCBI's Assembly database, accession number ASM18296v3 ([www.ncbi.nlm.nih.gov/assembly](http://www.ncbi.nlm.nih.gov/assembly)) [222] and the Taxonomy ID from NCBI ([www.ncbi.nlm.nih.gov/taxonomy](http://www.ncbi.nlm.nih.gov/taxonomy)) [223], which is required by *merlin* to univocally identify the organism under study throughout the reconstruction process. In order to establish a proximity between species, the 16S rRNA gene of several known closely related species was used to construct a Phylogenetic tree,

the sequences being retrieved from NCBI's database and aligned using MEGA X 10.0.5 [224] the evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [225] (Supplementary Figure II.1.1). The genome-wide functional annotation was processed by *merlin* based on taxonomy and frequency of similar sequences through remote BLAST [226] similarity searches to the UniProtKB/Swiss-Prot database [227] (<http://www.UniProt.org/>) and HMMER [228]. Protein-reaction associations available in the KEGG BRITE database were used to assemble the draft network. All reactions classified as spontaneous or non-enzymatic were also included in the first draft of the model. The assembly of the metabolic network is performed by *merlin*, using genome annotation to determine which reactions will be included in the model based on an algorithm described in detail elsewhere [122].

### II.1.3.3. Reversibility and Balancing

In order to ensure that all reactions in the network are balanced is important to verify the stoichiometry. *merlin* includes a tool to identify unbalanced reactions, that were manually verified and corrected. The reaction reversibility should also be confirmed to avoid gaps and mispredictions of the model, *merlin* also provides a tool to correct reversibility using BRENDA [126] as reference and the data provided elsewhere [229]. Since there are no guarantees that the EC numbers available in the different databases are updated, a manual inspection was also performed to correct a few cases of enzymes with deleted/transferred EC numbers, using public databases (BRENDA [230], UniProt, MetaCyc [231] and KEGG [125]) and literature search.

#### II.1.3.4. Compartmentalization

This model includes four compartments: extracellular, cytoplasm, mitochondrion, and cytoplasmic membrane. The prediction of compartments for each enzyme and carrier was performed using the WoLF PSORT [232].

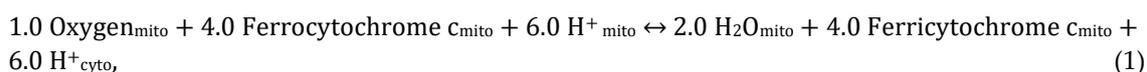
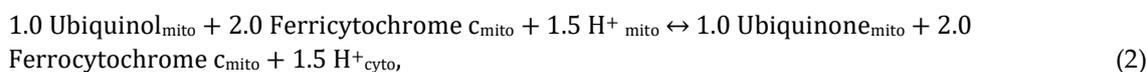
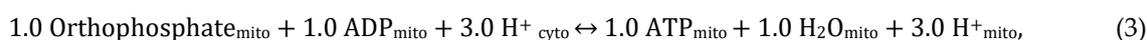
#### II.1.3.5. Transport reactions

Given the existence of compartments in the model, it is necessary to create transport reactions for the metabolites. Transport reactions were generated using genomic information together with the public database TCDB [233] by *merlin's* TranSyT [127]. Transport reactions across internal and external membranes for currency metabolites, such as H<sub>2</sub>O, CO<sub>2</sub>, and NH<sub>3</sub>, which are often carried by facilitated diffusion, were added to the model with no gene association.

#### II.1.3.6. Biomass Equation

The biomass formation was represented by an equation that includes proteins, DNA, RNA, lipids, carbohydrates, and cofactors, and detailed information for the composition of each one of these macromolecules. The content of each component was determined based on literature or using experimental data. All the calculations were performed as described previously [234].

For the phosphorus to oxygen ratio, the same theoretical ratio used in the *S. cerevisiae* iMM904 metabolic model was applied, 1.5. This ratio represents the relationship between ATP synthesis and oxygen consumption, indicating the number of orthophosphate molecules used for ATP synthesis per atom of oxygen consumed during oxidative phosphorylation [111]. Three generic reactions contributing to this ratio were automatically generated by *merlin*, and were updated to replicate the same ratio as in the iMM904 model:

**Reaction R00081\_C4:****Reaction T02161\_C4:****Reaction T00485\_C4:****The final balance reaction:**

This model also includes ATP requirements for biomass formation and maintenance (non-growth). The growth ATP requirements, 23.346 mmoles ATP/gDCW, were introduced directly into the biomass equation; this value was calculated based on ATP requirements for biosynthesis of cell polymers for *S. cerevisiae*, adjusted for the composition in macromolecules of the biomass equation [235].

Non-growth associated ATP maintenance, the amount of ATP required by the cell even when it is not growing, was represented in the model by an equation that forces ATP consumption via a specific flux. The boundaries of this flux were inferred from *Candida tropicalis* [236]. For more detailed information on the computation of the biomass equation, see Supplementary Data II.1.1.

### **II.1.3.7. Curation of the Model**

Throughout the curation process, reactions were edited, manually added to, or removed from the model to correct some gaps in the network, using KEGG pathways, MetaCyc Database, and literature data as standards.

### **II.1.3.8. Strains and growth media**

*C. albicans* reference strain SC5314 was batch-cultured at 37°C, with orbital agitation (250 rpm) in Yeast Nitrogen Base (YNB) medium without amino acids: 5g/L glucose (Merck), 6.8 g/L YNB (Difco). Solid media contained, besides the above-indicated ingredients, 20 g/L agar (Iberagar).

### **II.1.3.9. Carbon and nitrogen source utilization assessment**

The capability of utilizing different carbon and nitrogen sources for cell growth was assessed by comparing *in silico* predictions to literature data for *C. albicans*. For the few carbon or nitrogen sources for which the model predictions were not consistent with literature data, wet-lab experiments were conducted. Specifically, the utilization of cellobiose, D-Ribose, and mannitol as carbon sources, by the *C. albicans* reference strain SC5314, was evaluated in solid YNB medium containing either 5g/L glucose as control, or 5g/L of either one of the mentioned carbon sources. *C. albicans* cell suspensions used to inoculate the agar plates, were mid-exponential cells grown in YNB medium with 5g/L glucose, until culture OD<sub>600nm</sub> = 0.5 ± 0.05 was reached and then diluted in sterile water to obtain suspensions with OD<sub>600nm</sub> = 0.05 ± 0.005. These cell suspensions and subsequent dilutions (10<sup>-1</sup>; 10<sup>-2</sup>; 10<sup>-3</sup>) were applied as 4 µL spots onto the surface of solid YNB media, with the indicated carbon sources. Growth was assessed after incubation at 37°C for 24h.

### II.1.3.10. Network simulation and analysis

All the phenotype simulations were performed with FBA in OptFlux 3.0 [128] using the IBM CPLEX solver. Gene essentiality was also determined by OptFlux 3.0 which provides a tool that allows to determine critical genes automatically by performing individual gene knockouts and stimulating growth in a given environmental condition. In these simulations, environmental conditions that simulated the Roswell Park Memorial Institute (RPMI) medium were used, in order to replicate the human serum conditions.

## II.1.4. Results and Discussion

### II.1.4.1. Model characteristics

The final version of the iRV781 (Supplementary Data II.1.4) model includes 781 genes associated with 1221 reactions, among which, 174 are transport reactions, and 196 are external drain reactions (exchange constraints set to mimic the environmental conditions), involving 927 metabolites and four different compartments. Analyzing the distribution of genes by the compartments, 205 are plasma membrane genes, 521 cytoplasmatic genes, and 139 mitochondrial genes.

In order to elucidate the characteristics of our model we selected well-characterized genome-scale metabolic models of *C. glabrata* [237] and *S. cerevisiae* [238] as a comparison. Table II.1 shows the distribution of those reactions by the main pathways in the three models. In general, the number of reactions by pathway is quite similar to *C. glabrata*, *S. cerevisiae*, or both.

Although our model has common standard identifiers for reactions (KEGG ID), it is not possible to assess how the reactions differ among the three models, since

the remaining two models do not possess the same identifiers. However, considering only the proteins associated with an EC number, it is possible to make a comparison across the existing models. More than 80% of the proteins with an associated EC number in our model are also present at least in one of the other 2 models (*S. cerevisiae* or *C. glabrata*). Furthermore, about 65% of the proteins are shared by the three models while about 20% are unique in iRV781 (Figure II.2). The complete list of unique EC numbers can be found in Supplementary Data II.1.3.

**Table II.1** - Number of reactions in the main pathways of the *C. albicans* iRV781 model in comparison to *C. glabrata* iNX804 model and *S. cerevisiae* iMM904 model.

	<i>C. albicans</i> iRV781	<i>C. glabrata</i> iNX804	<i>S. cerevisiae</i> iMM904
Amino acid metabolism	218	223	217
NAD biosynthesis	20	20	24
Cofactors and vitamins	122	120	127
Nucleotide metabolism	120	138	135
Alternate carbon metabolism	27	31	27
Glycolysis/gluconeogenesis	26	18	22
Citrate cycle	24	20	13
Pentose phosphate pathway	18	16	13
Pyruvate metabolism	31	28	18
Oxidative phosphorylation	10	13	19
Sterol metabolism	29	30	49
Fatty acid metabolism	87	81	108
Glycerolipid metabolism	13	9	12
Phospholipid metabolism	34	44	52

In most cases, the observable differences in EC numbers were related to outdated EC numbers or were compensated by other enzymes that are responsible for the same reactions in the model. However, some cases stand out as potential unique features of *C. albicans*:

The enzyme 1.13.99.1, inositol oxygenase, responsible for the conversion of myo-inositol into D-glucuronate. This enzyme seems to be involved in resistance to

toxic ergosterol analogs [239], is also present in other *Candida* species, including some important pathogens (*C. parapsilosis*, *Candida dubliniensis* (*C. dubliniensis*), *C. auris*), however, is absent in *C. glabrata*.

The enzyme 1.1.1.289, sorbose reductase, responsible for the interconversion of L-Sorbose into D-Sorbitol. In fact, the presence of this enzyme allows *C. albicans* to use L-Sorbose as carbon source. Contrary to *S. cerevisiae*, the utilization of L-Sorbose is confirmed experimentally [240].

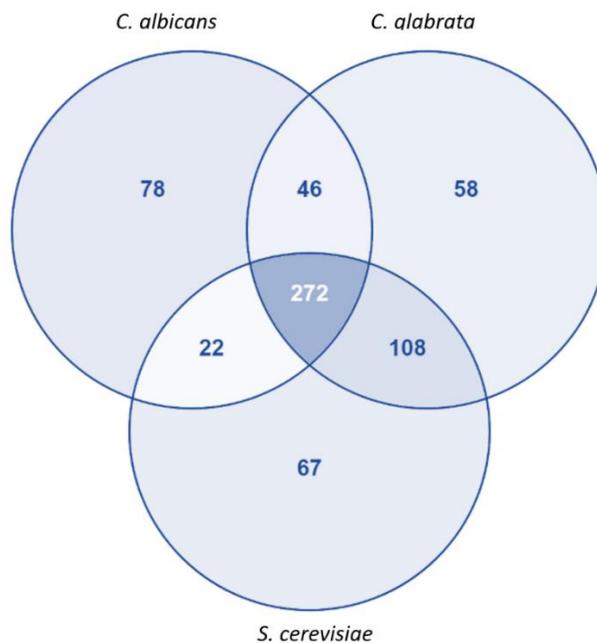
The enzyme 1.14.19.17, sphingolipid 4-desaturase, responsible for the conversion of dihydroceramide into N-Acylsphingosine. This protein is involved in sphingolipid metabolism, and it has been shown that altering sphingolipid composition is a possible mechanism of azole resistance in *C. albicans* [241]. The presence of this enzyme may represent a specific resistance feature of some *Candida* species, which is present in some pathogenic species, *C. parapsilosis*, *C. dubliniensis*, and *C. auris*, but not in *C. glabrata* for example.

The enzyme 1.1.99.2, L-2-hydroxyglutarate dehydrogenase, is a metabolite repair enzyme responsible for the conversion of (S)-2-hydroxyglutarate into 2-oxoglutarate, in other organisms such as plants [242] or humans [243], the inactivation of this enzyme leads to the accumulation of the toxic (S)-2-hydroxyglutarate.

The enzyme 2.7.1.59, N-acetylglucosamine kinase, responsible for the conversion of N-acetyl-D-glucosamine into N-acetyl-D-glucosamine 6-phosphate. Many yeast species including *S. cerevisiae* have lost their ability to utilize N-acetyl-D-glucosamine as carbon source, however, genetically altered yeasts are able to use it, based on expression of *C. albicans* genes [244]. In fact, this enzyme allows *C. albicans* to utilize this carbon source, feature that is particularly important for Its survival inside the phagosomes [245].

The enzyme 3.5.1.25, N-acetylglucosamine-6-phosphate deacetylase, responsible for the conversion of N-acetyl-D-glucosamine 6-phosphate into D-glucosamine 6-phosphate. Like 2.7.1.59, this enzyme is also involved in N-acetyl-D-glucosamine metabolism.

The enzyme 1.4.3.3, D-amino-acid oxidase, responsible for the conversion of a D-amino acid into a 2-oxo carboxylate and ammonia, is the first enzyme involved in the catabolism of D-amino acids and may allow the utilization D-amino acids as a source of carbon or nitrogen in some yeasts [246], may be an interesting feature to be explored in *C. albicans*.



**Figure II.2** - Comparison between *C. albicans*, *S. cerevisiae*, and *C. glabrata* proteins with associated EC Numbers present in the genome-scale metabolic models iRV781, iIN800, and iNX804, respectively. Diagram obtained using VENNY2.1 tool [247].

#### **II.1.4.1.1 Gap filling and Model Curation**

During the process of manual curation described in the methods section, a total of 66 reactions were manually added to the initial model obtained from the results of re-annotation to fill gaps. For additions, evidence from the literature was always considered, or of the well-studied *S. cerevisiae*. On the other hand, 336 reactions were removed from the initial model, have been removed for being unconnected reactions, general reactions, reactions using metabolites that are not included in the model, or reactions for which it was manually verified that the model does not have the enzyme coding gene. Additionally, the compartment of 79 reactions was changed, and 94 reactions were altered in order to become balanced. The complete list of alterations can be found in Supplementary Data II.1.2.

#### **II.1.4.1.2 Biomass Equation**

The biomass equation (Table II.2) includes the composition of proteins, DNA, RNA, lipids, carbohydrates, and cofactors. For the composition of DNA, the whole genome sequence was used to estimate the amount of each deoxyribonucleotide as described in [123], while mRNA, rRNA, and tRNA were used to estimate the total RNA in the cell as described in [111]. For the amino acid composition, the percentage of each codon usage was calculated from the translated genome sequence [123] using the e-BiomassX tool [248].

Carbohydrate [249], Lipid [250], Sterol [250], Phospholipid [251], and Fatty acid [252] compositions were inferred from literature data. Essential metabolites were included in the biomass composition to qualitatively account for the essentiality

of their synthesis pathways [237,253]. The growth and non-growth ATP requirements were adopted from *S. cerevisiae* [254].

**Table II.2.** Biomass Composition used in the model iRV781

Metabolite	g/gDCW	Metabolite	g/gDCW
<b>Proteins</b>		<b>Lipids</b>	
L-Valine	0,02001	Lanosterol	0,00166
L-Tyrosine	0,02153	Squalene	0,00088
L-Tryptophan	0,00671	Ergosterol	0,00247
L-Threonine	0,02311	Phosphatidylserine	0,00299
L-Serine	0,02908	Phosphatidylinositol	0,00417
L-Proline	0,01616	Phosphatidylcholine	0,00681
L-Phenylalanine	0,02407	Phosphatidylethanolamine	0,00542
L-Methionine	0,00869	Cardiolipin	0,00201
L-Lysine	0,03535	Phosphatidic acid	0,00271
L-Leucine	0,03874	Phosphatidylglycerol	0,00174
L-Isoleucine	0,02992	Tetradecanoic acid	0,00003
L-Histidine	0,01067	Hexadecanoic acid	0,00073
L-Glutamate	0,03084	Palmitoleic acid	0,00022
L-Cysteine	0,00410	Octadecanoic acid	0,00035
L-Aspartate	0,02508	Oleic acid	0,00163
L-Asparagine	0,02841	Linoleate	0,00054
L-Arginine	0,02203	Linolenate	0,00008
L-Alanine	0,01334	Triacylglycerol	0,00573
Glycine	0,01077	Monoacylglycerol	0,00620
L-Glutamine	0,02158	Diacylglycerol	0,00087
		Sterol esters	0,01177
<b>Carbohydrates</b>		<b>Ribonucleotides</b>	
Chitin	0,01368	UTP	0,00603
Mannan	0,14669	GTP	0,00714
$\beta$ (1,3)-Glucan	0,23962		

		CTP	0,00561
<b>Deoxyribonucleotides</b>		ATP	0,00714
dTTP	0,02072		
dGTP	0,01266	<b>Soluble Pool</b>	
dCTP	0,01118	Thiamine	0,00290
dATP	0,02114	Ubiquinone-6	0,00290
		NADP+	0,00290
<b>More detailed information in Supplementary Data II.1.1.</b>		NAD+	0,00290
		FMN	0,00290
		FAD	0,00290
		CoA	0,00290
		Biotin	0,00290
		Pyridoxal phosphate	0,00290
		5-Methyltetrahydrofolate	0,00290

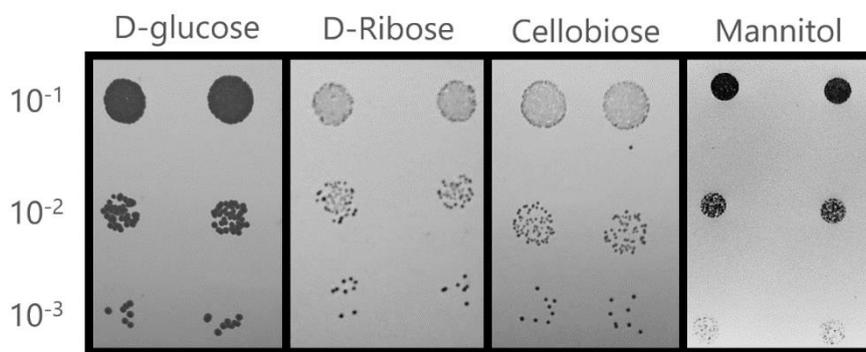
## II.1.4.2. Validation of the iRV781 model

### II.1.4.2.1 Carbon and nitrogen source utilization

Based on the literature, phenotypic growth data were collected from different sources. Data related to *C. albicans* strains, other than the reference SC5314 strain, was also considered in the analysis to increase the number of carbon and nitrogen sources tested.

In a first simulation, this model correctly predicted the usability of 92% of the 39 tested carbon sources. According to data available on Royal Netherlands Academy of Arts and Sciences (CBS-KNAW) Fungal Biodiversity Centre webpage [255], the *C. albicans* CBS562 strain seems not to be able to use cellobiose and D-Ribose, contrary to the model's prediction. The prediction also failed for mannitol that, according to the model, could not be used as carbon source by *C. albicans*. However, *C. albicans* SC5314 strain has been shown to be able to grow on mannitol as single carbon source [256],[257]. Thus, the utilization of cellobiose, D-Ribose, and mannitol as carbon source by *C. albicans* SC5314 was evaluated

experimentally to assess if the prediction failure could result from a different metabolic capacity exhibited by the reference strain. The results (shown in figure II.3) confirmed the model's prediction regarding the utilization of cellobiose and D-Ribose, but not for the mannitol, suggesting that the reference *C. albicans* strain has higher metabolic capabilities, when compared to other strains, while showing that the model is able to correctly predict the usability of 97% of the tested substrates.



**Figure II.3** - Utilization of glucose (control), cellobiose, D-Ribose, and mannitol by *C. albicans* reference strain SC5314 as carbon source in solid YNB medium. Initial OD600nm =  $0.5 \pm 0.05$ . Growth was assessed after incubation at 37°C for 24h.

Altogether, the constructed model proved accurate when predicting the utilization of different carbon and nitrogen sources, when compared to experimental data (Table II.3). It correctly predicts the usability of 97% of the tested carbon sources, and 80% of the 15 tested nitrogen sources. It should be noted that in the nitrogen tests, none of the data belongs to the reference strain; therefore, the prediction accuracy of the model might be even closer to reality.

**Table II.3.** Comparison between *in vivo* and *in silico* phenotypic behavior of *C. albicans* under different carbon and nitrogen sources.

Carbon Source	Biomass		Reference
	<i>In vivo</i>	<i>In silico</i>	
N-acetylglucosamine	+	+	[256,258]
Glucose	+	+	[256–258]
Maltose	+	+	[258]
Galactose	+	+	[256–258]
Sucrose	+	+	[258]
Fructose	+	+	[256–258]
Mannitol	+	-	This study
Acetate	+	+	[258]
Ethanol	+	+	[258]
Glycerol	+	+	[256–258]
Mannose	+	+	[256,257]
Citrate	+	+	[255]
Lactate	+	+	[256]
Sorbitol	+	+	[256]
L-sorbose	+	+	[255]
D-xylose	+	+	[255]
L-rhamnose	-	-	[255]
$\alpha,\alpha$ -trehalose	+	+	[255]
Cellobiose	+	+	This study
Salicin	-	-	[255]
Myo-inositol	-	-	[255]
D-ribose	+	+	This study
Ribitol	-	-	[255]
D-glucuronate	-	-	[255]
D-galacturonate	-	-	[255]
Succinate	+	+	[255]
D-gluconate	+	+	[255]
Arbutin	-	-	[255]
D-arabinose	-	-	[255]
Galactitol	-	-	[255]
Starch	+	+	[255]
D-glucosamine	+	+	[255]
Inulin	-	-	[255]
Melibiose	-	-	[255]
Lactose	-	-	[255]
Raffinose	-	-	[255]

Erythritol	-	-	[255]
Xylitol	+	+	[255]
L-arabinitol	-	-	[255]
<b>Nitrogen Source</b>			
Nitrate	-	-	[255,259]
Nitrite	-	-	[255,259]
Ethylamine	+	-	[255]
L-Lysine	+	+	[255]
Ammonia	+	+	[255,259]
Cadaverine	+	-	[255]
Glucosamine	-	+	[255]
Creatine	-	-	[255]
Creatinine	-	-	[255]
Imidazole	-	-	[255]
L-asparagine	+	+	[255,259]
Urea	+	+	[255,259]
Hydroxylamine	-	-	[255,259]
Hydrazine	-	-	[255,259]
D-Tryptophan	-	-	[255]

Growth (+); lack of growth (-).

#### II.1.4.2.1 Growth parameters in batch culture

Experimental data obtained elsewhere [260] from synthetic minimal media batch cultures with glucose as carbon source were used to validate the model quantitatively. The model was simulated in environmental conditions that simulate the medium used in [260]. The glucose uptake flux was fixed to  $q_{\text{Glucose}}=7.56 \text{ mmol.g}^{-1} \text{ dry weight.h}^{-1}$  as per such work, and the remaining nutrients flux were left unconstrained, as the model in this condition is glucose-limited. Once again, the model proved to be robust as the experimentally observed growth rate is similar to that predicted by the model (table II.4). Additionally, the formation of glycerol, acetic acid, and ethanol as by-products

was not predicted to occur, which is in agreement with the experimental data, except for ethanol, that appears to be produced in trace amounts. *C. albicans*, as a Crabtree-negative yeast [261], under aerobic conditions does not produce significant concentrations of ethanol. Nonetheless, the model predicts ethanol production under low-oxygen conditions ( $q_{\text{Oxygen}} < 5.5 \text{ mmol.g}^{-1} \text{ dry weight.h}^{-1}$ ).

**Table II.4** - Growth parameters of iRV781 and comparison with *in vivo* values for *C. albicans* and *S. cerevisiae*.

	Specific growth rate ( $\text{h}^{-1}$ )	q ( $\text{mmol g}^{-1} \text{ dry weight h}^{-1}$ )			
		Glucose	Ethanol	Glycerol	Acetic acid
<i>In silico C. albicans</i>	0.53	7.56	0	0	0
<i>In vivo C. albicans</i> [260]	0.51	7.56	0.38	0	0
<i>In vivo S. cerevisiae</i> [260]	0.38	13.26	21.87	1.98	<0.1

*C. albicans* is unable to grow in anaerobic conditions in minimal media. However, *C. albicans* colonization is known to spread into anaerobic niches of the gastrointestinal tract or in the inner sections of biofilms where the oxygen availability is scarce or null. Dumitru et al., 2004 reported a defined anaerobic growth medium for studying *C. albicans*. In this medium (GPP) oleic acid and nicotinic acid were added as required growth factors for anaerobic growth [262]. Interestingly, *S. cerevisiae* under anaerobiosis also requires growth factors, such as ergosterol and Tween 80, a source of oleic acid, if growing in a defined medium (SMM), it should be noticed that SMM medium also contains nicotinic acid in its composition despite not being a required growth factor in anaerobic conditions [263]. Cultivation was simulated in the absence of oxygen in GPP medium and GPP medium supplemented with oleic acid and nicotinic acid, to assess if this model is capable of predicting growth in anaerobic conditions. Our model

predicts the growth only in media supplemented with specific anaerobic growth factors. For the simulation, the glucose uptake flux was set to  $q_{\text{Glucose}}=6.58$  mmol.g<sup>-1</sup> dry weight.h<sup>-1</sup>, to compare the growth parameters with the reported values for *S. cerevisiae*. Indeed, for the same anaerobic conditions, the specific growth rate of the model and the ethanol production are similar to data reported for *S. cerevisiae* [263], though the model does not predict the production of glycerol in such conditions (Table II.5).

**Table II.5** - Anaerobic growth assessment of iRV781 model in defined media with or without anaerobic supplements. DMM [263] (defined minimal medium); DMMsup. [263] (defined minimal medium supplemented with ergosterol and Tween 80); GPP [262] (glucose-phosphate-proline); GPPsup. [262] (glucose-phosphate-proline supplemented with oleic acid and nicotinate).

Condition	Specific growth rate (h <sup>-1</sup> )	q (mmol g <sup>-1</sup> dry weight h <sup>-1</sup> )		
		Glucose	Ethanol	Glycerol
<i>In silico</i> GPP	0	0	0	0
<i>In silico</i> GPP <sub>sup.</sub>	0.08	6.58	10.80	0
<i>In silico</i> DMM	0	0	0	0
<i>In silico</i> DMM <sub>sup.</sub>	0.08	6.58	10.80	0
<i>S. cerevisiae</i> DMM	0.10	6.58	9.47	1.11

#### II.1.4.3. Gene essentiality assessment: a tool for drug target discovery?

A set of *C. albicans* essential genes collected from other work [264], were used to evaluate the model's ability to predict essentiality. For each gene, a simulation was performed, on the same environmental conditions described in the reference [264] (YNB medium), eliminating the corresponding reactions for that gene. Only protein-coding genes present in the model were considered. The model was able to correctly predict 78% (84 out of 108) of the identified essential genes (Supplementary Data II.1.5). It is important to highlight that in this type of

models, the regulatory network is not considered, so it will always be expected that some predictions are not close to reality.

To evaluate if gene essentiality assessment could be a promising tool in drug target discovery, each one of the identified essential enzymes in the RPMI medium, was searched in the DrugBank database [265] as a possible drug target of known antimicrobial agents. RPMI medium simulates human serum, thus allowing to simulate the natural environment faced by *C. albicans* in systemic infections.

Interestingly, 11 ERG genes, including the well-known azole drug target ERG11, were predicted by the model to be essential in RPMI medium. Although most ERG genes are not essential, the inhibition of the activity of this pathway has a fungistatic effect indeed. They encode the enzymes that guide the last steps of ergosterol biosynthesis. This pathway is the main target of azole drugs, one of the most common antifungal agents to treat Candida infections [266]. These drugs act by blocking ergosterol biosynthesis inhibiting the Erg11 encoded by the ERG11 gene. When an azole drug binds to this enzyme, ergosterol synthesis is inhibited, leading to lower concentrations of this metabolite in the plasma membrane [267]. Given that ergosterol is part of the *C. albicans* biomass, it is acceptable to consider that enzymes that participate in its synthesis pathway can be essential, making most ERG genes attractive alternatives as new drug targets [268].

Many additional proteins stand out as promising new drug targets, including some for which there are already predicted inhibitory drugs, based on results for homologous proteins in other organisms. For example, Atovaquone is a drug used as a fixed-dose combination with Malarone for treating uncomplicated malaria cases or as chemoprophylaxis in travelers. This drug is an analogue of ubiquinone and targets enzyme 1.3.5.2 encoded by URA9 in *Plasmodium falciparum*. Atovaquone acts as a competitive inhibitor of ubiquinol inhibiting

the mitochondrial electron transport chain at the bc1 complex, resulting in a loss of mitochondrial function [269]. It would be interesting to check if these drugs are also active against *C. albicans* by targeting CaUra9.

Another promising example of a predicted *C. albicans* drug target is Fol1, which corresponds to enzyme 2.5.1.15. Fol1 is the target of the sulfa drugs (sulfonamides and sulfones), a very well-known class of drugs, used to treat infectious diseases [270]. The effect of sulfa drugs on *C. albicans* has not been sufficiently investigated; however, it seems that sulfa-fluconazole combination results in increased antifungal activity against *C. albicans*, leading to the reversal of azole resistance in previously resistant strains [271].

Since a reaction can be catalyzed by a protein encoded by more than one gene, and genes may encode more than one protein, we decided to analyze the model's critical reactions. This analysis allowed to increase the number of confirmed drug targets predicted as essential enzymes.

As example, the FKS genes stand out. They are not considered essential as the enzyme beta-1,3-glucan synthase can be encoded by more than one FKS/GSC/GSL genes [272]. However, the model predicts the reaction in which this enzyme participates as essential. The beta-1,3-glucan synthase is the target of the echinocandin class of antifungal drugs. Via noncompetitive inhibition, these drugs block the enzyme and stop beta-1,3-glucan synthesis, compromising the integrity of the cell wall [273].

Other drugs, such as ethionamide, sulfacetamide, azelaic acid, cerulenin, or trimethoprim, were identified as targeting proteins from various organisms. These proteins are homologous to the ones encoded by genes identified as essential in RPMI, in the *C. albicans* model (Table II.6). Altogether, these results indicate a high capability of the iRV781 to predict drug targets, offering predicted essential genes that have great potential as targets for new antifungal drugs. The

predictions of this model may even be applied to other pathogenic *Candida* species given the proximity of most species. In fact, if we search the 12 genes present in Table 6 in emerging non-*albicans* *Candida* species, 12 have ortholog genes in *C. parapsilosis* and *C. dubliniensis*, 11 in *C. auris*, and 8 in *C. glabrata*.

Despite the methodology for the reconstruction of genome-scale metabolic models being standardized, eukaryotic models remain a challenge, due to their larger genomes and complexity [129]. These models always seek to get as close as possible to reality; however, given the complexity of the networks, they are always subject to some errors, which may cause small deviations in the predictions. Some errors may include incorrect assignment of GPR associations, reaction directionality or reversibility, incongruous stoichiometric parameters, missing reactions, and inaccurate biomass composition [129].

Genome-scale metabolic reconstructions have been effective in drug target prediction and they are expected to continue to expand in the future [274]. Gene essentiality assessment is the most common method to identify potential drug targets, and for a better prediction, it is necessary to consider the medium in which the organism is exposed. In this work, we simulate gene essentiality in RPMI medium in order to simulate the natural environment faced by *C. albicans* in systemic infections. However, it is important to highlight that in these reconstructions, it is not considered that cells may need time to adapt to genetic perturbations or environmental variability [129]. Additionally, yeast interactions with other microorganisms and the secretion of compounds that can influence their surrounding environment are not taken into account [130]. Despite these inaccuracies, genome-scale metabolic reconstructions have proved to be very efficient in discovering new drug targets, and once a model is built, drug targets can be predicted relatively easily. In fact, the experimental validation of the targets and the identification of the effective drugs represent a more demanding challenge [119].

### II.1.5. Conclusions

The first validated global metabolic model for the human pathogen *C. albicans* is presented in this study. The model was manually curated and validated thoroughly, constituting a powerful platform for the study of *C. albicans* metabolic potential and weaknesses. The iRV781 model includes 781 genes associated with 1221 reactions, the number of reactions in the main pathways being similar to those in *C. glabrata* and *S. cerevisiae* models. However, about 20% of the proteins associated with EC numbers in iRV781 are unique in relation to these models. The model proved accurate when predicting the utilization of different carbon and nitrogen sources, and in anaerobic growth in defined anaerobic media. *In silico* growth parameters are also in agreement with the experimental data. We were able to identify as essential genes in the RPMI medium, a medium that simulates the human body environment, previously known targets of antifungal agents and other antimicrobial agents used in clinical practice. This observation demonstrates that the *C. albicans* global stoichiometric model, presented herein, constitutes a promising platform for the identification of targets for new antifungal drugs, that may circumvent the current tendency of growing therapeutic failure.

**Table II.6** - Drug targets evaluated for gene essentiality prediction in RPMI medium, as identified by the iRV781. Data retrieved from DrugBank database; only drugs with known pharmacological action were selected.

Systematic Name	Standard Name	EC number	Organism	Drug	PDB Entry	Similarity	Coverage
C1_08590C_A	ERG1	1.14.14.17	<i>Candida albicans</i>	Terbinafine	-	-	-
			<i>Candida albicans</i>	Tolnaftate	-	-	-
C1_09720W_A	URA1	1.3.5.2	<i>Plasmodium falciparum</i>	Atovaquone	5DEL	37%	81%
C2_02460W_A	ERG7	5.4.99.7	<i>Candida albicans</i>	Oxiconazole	-	-	-
Ø5_00190C_A	FAS1	1.3.1.9	<i>Mycobacterium tuberculosis</i>	Ethionamide	4V8W	30%	45%
			<i>Mycobacterium tuberculosis</i>	Isoniazid			
C5_00770C_A	FOL1	4.1.2.25	<i>Saccharomyces cerevisiae</i>	Sulfacetamide	2BMB	42%	65%
C5_02710W_A	TRR1	1.8.1.9	<i>Staphylococcus aureus</i>	Azelaic acid	4GCM	42%	98%
C7_03130C_A	DFR1	1.5.1.3	<i>Escherichia coli</i>	Trimethoprim	4GH8	35%	77%
C5_00770C_A	FOL1	2.5.1.15	<i>Escherichia coli</i>	Sulfonamides and sulfones	1AJ2	36%	40%
			<i>P. falciparum</i>	Sulfonamides and sulfones	6KCM	26%	65%
C1_02420C_A	GSC1		<i>Candida albicans</i>	Anidulafungin	-	-	-
C1_05600W_A	GSL1	2.4.1.34	<i>Candida albicans</i>	Caspofungin	-	-	-
CR_00850C_A	GSL2		<i>Candida albicans</i>	Micafungin	-	-	-
C3_04830C_A	FAS2	2.3.1.41	<i>Escherichia coli</i>	Cerulenin	2BYX	31%	8%
CR_00850C_A	ERG11	1.14.14.154	<i>Candida albicans</i>	Azoles	-	-	-

### **II.1.6. Acknowledgments**

This research was funded by “Fundação para a Ciência e a Tecnologia” (FCT) [Contract PTDC/BII-BIO/28216/2017] and by Programa Operacional Regional de Lisboa 2020 [LISBOA-01-0145-FEDER-022231], through the Biodata.pt Research Infrastructure. Funding received by iBB-Institute for Bioengineering and Biosciences from FCT [Contract UIDB/04565/2020] is also acknowledged.

## II.2. A Genome-scale metabolic model for the human pathogen *C. parapsilosis* and Early Identification of Putative Novel Antifungal Drug Targets

### II.2.1. Abstract

*C. parapsilosis* is an emerging human pathogen whose incidence is rising worldwide, while an increasing number of clinical isolates display resistance to first-line antifungals, demanding alternative therapeutics. Genome-Scale Metabolic Models (GSMMs) have emerged as a powerful *in silico* tool for understanding pathogenesis due to their systems view of metabolism but also to their drug target predictive capacity. This study constructed the first validated GSMM for *C. parapsilosis* – iDC1003 – comprising 1003 genes, 1804 reactions, and 1278 metabolites across four compartments and an intercompartment. *In silico* growth parameters as well as predicted utilization of several metabolites as sole carbon or nitrogen sources were experimentally validated. Finally, iDC1003 was exploited as a platform for predicting 147 essential enzymes in mimicked host conditions, in which 56 are also predicted to be essential in *C. albicans* and *C. glabrata*. These promising drug targets include, besides those already used as targets for clinical antifungals, several others that seem to be entirely new and worth further scrutiny. The obtained results strengthen the notion that GSMMs are promising platforms for drug target discovery and guide the design of novel antifungal therapies.

**Keywords:** *C. parapsilosis*; Genome-scale metabolic model; Drug target; Drug discovery.

## II.2.2. Introduction

In a world of climate and social change, human susceptibility to microbial disease is increased. In particular, fungal infections have seen a significant rise in incidence worldwide since the 1980's, with *Candida* spp. accounting for the majority of cases [275]. Although *C. albicans* is the most commonly isolated species from candidiasis patients, the 1990s saw a shift in incidence within the genus towards Non-*C. albicans Candida* species (NCAC) [276]. From these, *C. parapsilosis* has seen one of the most significant increases, often surging as the second most common etiological agent of *Candida* spp. infections, subverting historical trends in species incidence and even outranking *C. albicans* in some European countries [277]. Non-geographically restricted and with a broad range of virulence factors, adding to *C. parapsilosis*' already complex pathogenicity, is both the rise in resistance to first-line antifungals and intrinsically lower susceptibility to alternative therapies – such as azoles [278] and echinocandins [275], respectively. Thus, there is a strong need to develop new antifungal therapies and develop new research tools to understand the metabolism of pathogenesis and, if possible, to use metabolic impairment as an antifungal strategy.

Genome-Scale Metabolic Models (GSMMs) have emerged as a systems biology approach to tackle this issue [119]. GSMMs correspond to the *in silico* reconstructed metabolic network of a given organism [120] and thus enable a systems perspective of metabolism. In the little more than 20 years since the publication of the first model [118], GSMMs have proven their applicability and versatility from guiding strain design in metabolic engineering to elucidate novel drug target discovery in molecular medicine [112]. In the past, GSMMs have been mostly associated with the metabolic engineering of microbial cell factories due to their potential to simulate global metabolic behavior and provide clues for the optimisation of added-value compound production [111]. However, recent

examples have demonstrated the potential of these models in the search for new drug targets in pathogenic organisms [112–117].

This work presents the first validated *in silico* genome-scale metabolic reconstruction of the human pathogen *C. parapsilosis*, iDC1003. This model is provided in the well-established SBML format and can easily be read in most metabolic engineering platforms, such as OptFlux [128] and COBRA [220]. A set of predicted essential genes and reactions common to other pathogenic *Candida* spp. was obtained from the validated model, and their targetability as putative novel antifungal drug targets is discussed.

## **II.2.3 Materials and Methods**

### **II.2.3.1. Model reconstruction**

The herein described metabolic model reports to the yeast *C. parapsilosis* with the taxonomic ID 5480. Model reconstruction was performed using *merlin* 4.0.5 [122], and further curation and validation were performed on OptFlux 3.0 [128] using the IBM CPLEX solver. Throughout the curation process, reactions were edited, manually added to, or removed from the model to correct gaps in the network, using KEGG pathways, MetaCyc Database, and literature data as standards.

#### **II.2.3.1.1. Enzyme and Reaction Annotation**

The initial draft model construction comprised enzyme and subsequent reaction annotation. The genome sequence of the reference strain *C. parapsilosis* CDC317 was obtained from NCBI's Assembly database, accession number ASM18276v2 ([www.ncbi.nlm.nih.gov/assembly](http://www.ncbi.nlm.nih.gov/assembly)) [222] and the Taxonomy ID from NCBI

([www.ncbi.nlm.nih.gov/taxonomy](http://www.ncbi.nlm.nih.gov/taxonomy)) [223], which is required by *merlin* to identify the organism under study throughout the reconstruction process univocally. The genome-wide functional annotation was processed by *merlin* based on taxonomy and frequency of similar sequences through remote BLAST [226] similarity searches to the UniProtKB/Swiss-Prot database [227] (<http://www.uniprot.org/>). Hit selection was performed as described elsewhere [221], and phylogenetic proximity was implemented as described in Tsui *et al.* 2008 [279]. Protein-reaction associations available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [280] database were used to assemble the draft network.

#### **II.2.3.1.2. Correcting Reaction Reversibility, Directionality and Balance**

The initial reversibility curation was automatically performed by *merlin*, which implements information from remote tools such as eQuilibrator [281] as described by Dias *et al.* [221]. Further curation was entirely manual and justified, resorting to information from MetaCyc [231] and existing literature. Unbalanced reactions were identified automatically, and balancing was performed manually and explained with MetaCyc [231], ChEBI [282], Brenda [283] and existing literature. All the reactions manually edited during the curation process can be found in Supplementary Data II.2.1.

#### **II.2.3.1.3. Compartmentalization**

Compartmentalization was implemented using WoLF PSORT [232], a protein localization predictor, on the already connected non-compartmentalized model to simplify issue-solving. The constructed model includes four compartments:

extracellular, cytoplasm, mitochondrion, and peroxisome, and one intercompartment, the cytoplasmic membrane.

A compartmentalised model calls for the implementation of transport reactions to connect inter-compartment pathways. Transport reactions were generated using genomic information together with the public database TCDB [233] using *merlin*'s integrated tool TranSyT [127]. Transport reactions across internal and external membranes for currency metabolites, such as H<sub>2</sub>O, CO<sub>2</sub>, and NH<sub>3</sub>, often carried by facilitated diffusion, were added to the model with no gene association.

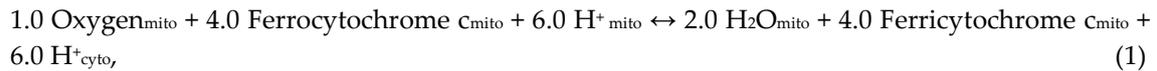
#### **II.2.3.1.4. Defining the Biomass Equation**

The Biomass equation encompasses the cells' major components and their relative numerical contributions – DNA, RNA, Carbohydrates, Lipids and Proteins – and acted as the objective function in the presented essentiality analysis. The content of each component was determined based on the literature. All the calculations were performed as described previously [234].

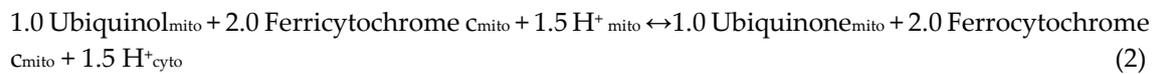
The reconstructed model also includes ATP requirements for both biomass production and cell maintenance – Growth Associated Maintenance (GAM) and Non-Growth Associated Maintenance (NGAM), respectively. A GAM value of 25.65 mmol ATP/gDCW was considered for the biomass equation, calculated based on the ATP requirements for the biosynthesis of cell polymers as shown in Mishra *et al.*[236], then adjusted for the considered biomass macromolecule composition. Non-growth associated ATP maintenance, the amount of ATP required by the cell repair and similar processes was implemented in this model as an autonomous equation, thus forcing a basal ATP consumption – flux bounds inferred from *Candida tropicalis* [236]. The biomass equation's components and

their relative content are shown in Supplementary Data II.2.1. The theoretical ratio used in the *S. cerevisiae* iMM904 metabolic model for the phosphorus to oxygen ratio was applied. Three generic reactions contributing to this ratio were automatically generated by *merlin* and were updated to replicate the same ratio as in the iMM904 model:

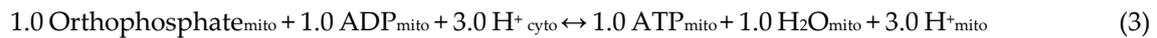
**Reaction R00081:**



**Reaction T02161:**



**Reaction T00485:**



**The final balance reaction:**



**II.2.3.2. Model Simulations and Enzyme Essentiality Prediction**

The model simulations were performed using the FBA [284] formulation on OptFlux 3.0 [128] using the IBM CPLEX solver. The determination of critical essential genes or reactions was performed with the following rationale: a gene/reaction is considered essential if, when removed from the model, this leads to a value of biomass flux of less than 5% of the reference value calculated for the wild-type strain. The essentiality of a gene/reaction was assessed by setting the flux of the reactions corresponding to a particular gene to zero and simulating the optimal growth rate with FBA. If deletion of one gene/reaction leads to non-

growth, that gene/reaction is defined as essential. The simulations for gene/reaction essentiality were performed in environmental conditions simulating the RPMI medium, which mimics human serum.

### **II.2.3.3. Model Validation**

#### **II.2.3.3.1. Strains and Growth Media**

*C. parapsilosis* type strain *CDC317* was batch cultured at 30 °C in orbital agitation (250 rpm) in Yeast Extract-Peptone-Dextrose (YPD) for inoculating cultivation and in Synthetic Minimal Media (SMM) for growth parameter determination. Media composition are as follows: YPD: 20 g/L glucose (Merck), 20 g/L peptone (Merck) and 10 g/L yeast extract (Merck); SMM: 20g/L glucose (Merck), 2.7 g/L ammonium sulphate (Merck), 0.05 g/L magnesium sulphate (Riddle-de-Haen), 2 g/L potassium dihydrogen phosphate (Panreac), 0.5 g/L calcium chloride (Panreac) and 100 µg/L biotin (Sigma).

#### **II.2.3.3.2. Aerobic Batch Cultivation**

Precultures (100 mL) for aerobic batch experiments were grown in SMM in 500mL flasks at 30 °C in orbital agitation (250 rpm). Cells were grown until the exponential phase and used to inoculate fresh media at an initial Optical Density at 600nm ( $OD_{600nm}$ ) of 0.3. Aerobic batch cultivations were incubated in SMM at 30 °C with orbital agitation (250 rpm) for 10 hours.

### **II.2.3.3.3. Cell Density, Dry weight and Metabolite Concentration Assessment**

During *C. parapsilosis* cell cultivation in SMM medium, 4 mL samples were harvested from the cell culture every two hours, aiming for the quantification of biomass and extracellular metabolites. Cell density was monitored by measuring the OD<sub>600nm</sub>. For dry weight determination, culture samples were centrifuged at 13000 rpm for 3 min, and the pellets were lyophilised for 72 h at -80 °C and weighted. The supernatants were centrifuged once more for clarification, and the concentrations of glucose, ethanol, glycerol and acetic acid in the supernatants were determined by HPLC on an Aminex HPX-87 H Ion Exchange chromatography column, eluted with 0.0005 M H<sub>2</sub>SO<sub>4</sub> at a flow rate of 0.6 mL/min, at room temperature. Concentrations were determined through the corresponding calibration curves. Samples from any batch cultivation were analysed in triplicate. The specific growth rate, specific glucose consumption rate, and the specific production rates of ethanol, glycerol, and acetic acid were calculated during the exponential growth phase as indicated elsewhere [285].

## **II.2.4. Results and Discussion**

### **II.2.4.1. Model Characteristics, highlighting *C. parapsilosis* unique metabolic features**

The herein described *C. parapsilosis* metabolic model, iDC1003, comprises 1003 genes associated with 1804 reactions – of which 358 are drains (exchange constraints set to mimic the environmental conditions) and 536 are transport reactions - and 1278 metabolites across four compartments (extracellular, cytoplasm, mitochondria, and peroxisome) and within an intercompartment, the

plasma membrane. The model can be found in SBML format in Supplementary Data II.2.2.

Manual curation assessed a total of 847 reactions from which 83 were mass balanced, 162 were corrected regarding reversibility or directionality, and 625 were added or removed from the model or had their annotation corrected or completed, as detailed in Supplementary Data II.2.1.

The Biomass equation (Table II.7) encompasses the cell's major components along with their respective and relative contributions – DNA, RNA, Carbohydrate, Protein, Lipid, and Cofactor content. The equation's composition in Carbohydrate [286], Lipid [236,287] – Sterol [236], Phospholipid [236], and Fatty acid [252] – and Protein [236] was inferred from literature data. On the other hand, for the composition of DNA, the whole genome sequence was used to estimate the amount of each deoxyribonucleotide as described in [123] while mRNA, rRNA, and tRNA were used to estimate the total RNA in the cell as described in [111]. Essential metabolites were included in the biomass composition to account for the essentiality of their synthesis pathways 39 qualitatively. The growth and non-growth ATP requirements were inferred from *Candida tropicalis* [236].

The iDC1003 model was compared to the well-characterised genome-scale metabolic models of *C. glabrata* [237] *S. cerevisiae* [238] and *C. albicans* [288] to highlight unique metabolic features of *C. parapsilosis*. Although iDC1003 uses standard identifiers for reactions (KEGG ID), it is not possible to assess how the reactions differ among the four models, as the remaining models do not use the same identifiers (except for *C. albicans*). Thus, a comparison across the existing models was carried out based on the proteins associated with an EC number.

**Table II.7** - Biomass composition used in the model iDC1003. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.2.1.

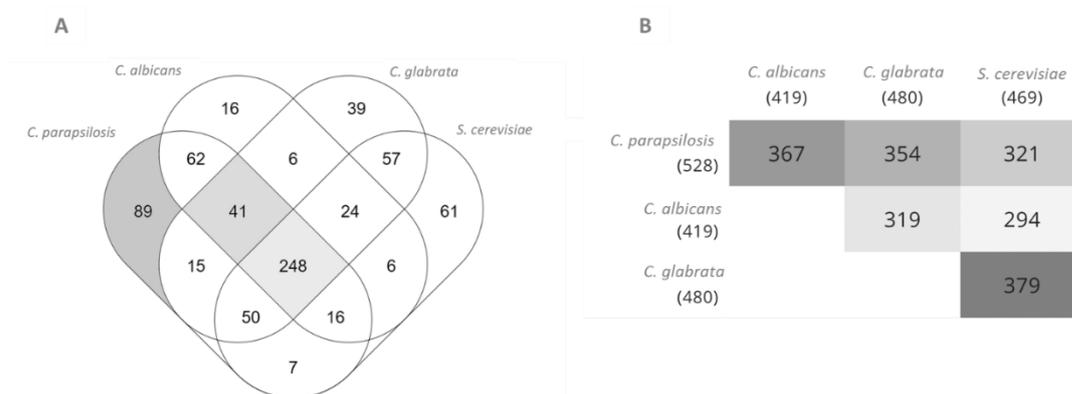
Metabolite	g/gDCW	Metabolite	g/gDCW
<b>Lipids</b>		<b>Proteins</b>	
Lanosterol	0.00063	L-Valine	0.03536
Squalene	0.00017	L-Tyrosine	0.02771
Ergosterol	0.00455	L-Tryptophan	0.01356
Phosphatidylserine	0.00237	L-Threonine	0.02230
1-Phosphatidyl-D-myo-inositol	0.00173	L-Serine	0.02478
Phosphatidylcholine	0.00288	L-Proline	0.01902
Phosphatidylethanolamine	0.00194	L-Phenylalanine	0.02845
Phosphatidic acid	0.00052	L-Methionine	0.04275
Phosphatidylglycerol	0.00186	L-Lysine	0.06440
Tetradecanoic acid	0.00001	L-Leucine	0.03933
Hexadecanoic acid	0.00074	L-Isoleucine	0.02115
(9Z)-Hexadecenoic acid	0.00010	L-Histidine	0.01887
Octadecanoic acid	0.00032	L-Glutamate	0.03987
(9Z)-Octadecenoic acid	0.00278	L-Cysteine	0.00487
(9Z,12Z)-Octadecadienoic acid	0.00071	L-Aspartate	0.00346
(9Z,12Z,15Z)-Octadecatrienoic acid	0.00016	L-Asparagine	0.00362
Triacylglycerol	0.00467	L-Arginine	0.00008
Monoacylglycerol	0.00401	L-Alanine	0.03551
Diacylglycerol	0.00316	Glycine	0.02150
Sterol esters	0.00445	L-Glutamine	0.03987
<b>Soluble Pool</b>		<b>Carbohydrates</b>	
Thiamine	0.00096	Chitin	0.01170

Ubiquinone-6	0.00096	Mannan	0.23437
NADP+	0.00096	$\beta$ (1,3)-Glucan	0.13621
NAD+	0.00096	<b>Deoxyribonucleotides</b>	
FMN	0.00096	UTP	0.01599
FAD	0.00096	GTP	0.01378
CoA	0.00096	CTP	0.01313
Biotin	0.00096	ATP	0.01730
Pyridoxal phosphate	0.00096	<b>Ribonucleotides</b>	
Tetrahydrofolate	0.00096	dTTP	0.00111
		dGTP	0.00074
		dCTP	0.00086
		dATP	0.00111

After intersecting the EC numbers present in each of the three models, 85% of the proteins with an associated EC number in the *C. parapsilosis* model were found to also be present in at least one of the other three models (Figure II.4). However, the remaining 15% (89/528) are exclusive to this model and may represent unique metabolic features of *C. parapsilosis*. It's also interesting to observe that 41 EC numbers are shared exclusively by Candida species and not by *S. cerevisiae*. These 41 enzymatic activities may be related to unique features of this genus, eventually linked to its pathogenicity. The complete list of unique EC numbers can be found in Supplementary Data II.2.1.

Occasionally, EC numbers might be related to outdated EC numbers or associated with other enzymes responsible for the same reactions in the different models. Nevertheless, specific cases stand out as potentially unique features of *C. parapsilosis*:

- 1.1.1.138: mannitol 2-dehydrogenase enables *C. parapsilosis* to use mannitol as a carbon source.
- 1.3.1.38: trans-2-enoyl-CoA reductase involved in fatty acid elongation, likely affecting membrane properties.
- 3.1.4.12: sphingomyelin phosphodiesterase participates in sphingolipid metabolism, responsible for sphingomyelin hydrolysis.
- 3.5.1.75: urethane amidohydrolase enables *C. parapsilosis* to use urethane as a nitrogen source.
- 1.16.1.7: ferric-chelate reductase, which is involved in iron starvation, catalyses the reduction of bound ferric iron in a variety of iron chelators (siderophores), resulting in the release of ferrous iron.
- 1.16.3.1: ferroxidase, involved in iron homeostasis, oxidises Fe(II) to Fe(III), which allows the subsequent incorporation of the latter into proteins such as apotransferrin and lactoferrin.



**Figure II.4** - Comparison between *C. parapsilosis*, *C. albicans*, *S. cerevisiae* and *C. glabrata* proteins with associated EC Numbers present in the iDC1003, iRV781, iIN800, and iNX804 genome-scale metabolic models, respectively. A: Venn

Diagram. B: Pairwise Intersections. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)).

## II.2.4.2. Model Validation

### II.2.4.2.1. Assessing the Model's Ability to Predict Carbon and Nitrogen Source Usage

Simulations were performed in SMM and compared to phenotypic growth data from existing literature to assess iDC1003's reliability in predicting biomass production from different sole carbon or nitrogen sources. Data related to *C. parapsilosis* strains, other than the reference CDC317 strain, were also considered in the analysis to increase the number of carbon and nitrogen sources accounted for. A total of 47 sole carbon sources and 17 sole nitrogen sources were evaluated. The *C. parapsilosis* model correctly predicted growth in 94% of the carbon sources tested and in 100% of the nitrogen sources (Table II.8). The model only failed for three carbon sources, 2-Keto-D-gluconic acid, L-arabinose and ribitol, which the literature indicates that *C. parapsilosis* can use for growth. Interestingly, as far as it could be assessed, there is no experimental evidence of any enzymes behind these pathways existing in yeasts. It would be interesting to look deeper into these organisms' eventually unique 2-Keto-D-gluconic acid, L-arabinose and ribitol assimilation pathways.

**Table II.8** - *In silico* predictions versus *in vivo* described data regarding *C. parapsilosis* ability to grow in the presence of sole carbon and nitrogen sources. From the 62 different tested compounds, iDC1003 correctly predicted positive or null biomass production on 95%. A plus represents biomass production (+), a

minus (-) no biomass production, and prediction disparities are highlighted in bold. Referenced data from Westerdijk fungal collection refer to strains CBS 1954 and CBS 604.

	<i>In vivo</i>	<i>In silico</i>	Reference		<i>In vivo</i>	<i>In silico</i>	Reference
<b>Carbon Source</b>							
Glucose	+	+	[255,289–293]	L-Sorbose	+	+	[255]
Maltose	+	+	[255,290–293]	D-arabinose	-	-	[255]
Sucrose	+	+	[255,289,291–293]	<b>L-arabinose</b>	+	-	[255]
Lactose	-	-	[255,290,292,293]	i-Erythritol	-	-	[255]
Galactose	+	+	[255,290–293]	Fucose	-	-	[290]
Melibiose	-	-	[255,290,291,293]	Salicin	-	-	[255]
Cellobiose	-	-	[255,290–293]	Arbutin	-	-	[255]
Inusitol	-	-	[255,290,293]	D-ribose	+	+	[255]
Xylose	+	+	[255,290,292,293]	D-Gluconate	+	+	[255]
Raffinose	-	-	[255,290–293]	<b>2-Keto-D-gluconic acid</b>	+	-	[255]
Trehalose	+	+	[255,290–293]	Inulin	-	-	[255]
Galactitol	-	-	[255,290,292,293]	D-Glucosamine	-	-	[255]
Rahmnose	-	-	[255]	D-Galacturonate	-	-	[255]
Glycerol	+	+	[255,290]	Quinate	-	-	[255]
<b>Ribitol</b>	+	-	[255,290]	D-Glucono-1,5-lactone	+	+	[255]
Mannitol	+	+	[255,290]	Propane-1,2-diol	-	-	[255]
Sorbitol	+	+	[255,290]	D-Glucarate	-	-	[255]
Ethanol	+	+	[255,290]	L-Arabinitol	-	-	[255]
Methanol	-	-	[255]	D-Glucuronate	-	-	[255]
Succinate	+	+	[255]	Butane 2,3 diol	-	-	[255]
Lactate	-	-	[255]	D-Galactonate	-	-	[255]
Citrate	+	+	[255]	D-Tagaturonate	-	-	[255]
Starch	-	-	[255]	Fructose	+	+	[289]
Xylitol	+	+	[255]				
<b>Nitrogen Source</b>							
Ammonium	+	+	[255,294]	Urethane	+	+	[295]
Citrate	-	-	[255]	Creatine	-	-	[255]
L-Lysine	+	+	[255]	Imidazole	-	-	[255]

Creatinine	-	-	[255]	L-Glutamate	+	+	[294]
D-Tryptophan	-	-	[255]	L-Proline	+	+	[294]
Nitrite	-	-	[255]	L-Isoleucine	+	+	[294]
Cadaverine	+	+	[255]	Allantoin	+	+	[294]
Glucosamine	-	-	[255]	4-Aminobutanoate	+	+	[294]
Ethylamine	+	+	[255]				

#### II.2.4.2.2. Assessing the Model's Ability to Quantitatively Predict Growth Parameters

Specific growth rate, glucose consumption rate, and metabolite production rates were experimentally determined to validate the model quantitatively due to the lack of literature data for *C. parapsilosis*. The *C. parapsilosis* CDC 317 cells were cultivated in SMM medium, and growth was followed by regular measurements of the OD<sub>600nm</sub> and the cell dry weight. Samples were harvested to assess glucose concentration as a function of time during the exponential growth phase. In these conditions, a glucose consumption rate of 2.098 +/- 0.404 mmol.gDCW<sup>-1</sup>.h<sup>-1</sup> was determined, while no ethanol, glycerol or acetate production could be detected. A simulation of the system's behaviour in environmental conditions of SMM medium, with a fixed glucose uptake flux of 2.098 mmol.g<sup>-1</sup> dry weight.h<sup>-1</sup> was performed to evaluate the model's ability to predict the specific growth rate. The remaining nutrient fluxes were left unconstrained, as the system in these conditions is glucose-limited. Considering the experimentally determined glucose consumption rate of 2.098 mmol.gDCW<sup>-1</sup>.h<sup>-1</sup>, the model predicted a specific growth rate of 0.172 h<sup>-1</sup>. Compared to the experimentally determined rate of 0.159 +/- 0.027 h<sup>-1</sup>, the predicted growth rate is within the uncertainty interval of the experimentally determined parameter. Thus there is no significant difference between both, which suggests iDC1003 can predict *C. parapsilosis*

growth parameters (Table II.9) quantitatively. Additionally, the formation of glycerol, acetic acid, and ethanol as by-products was not predicted to occur, which agrees with the experimental data and the notion that *C. parapsilosis* is a crabtree-negative yeast. Altogether, iDC1003 proved to predict the main metabolic features of *C. parapsilosis* quantitatively.

**Table II.9** - Growth parameters of iDC1003 and comparison with experimentally determined values.

	Specific growth rate (h <sup>-1</sup> )	q (mmol g <sup>-1</sup> dry weight h <sup>-1</sup> )			
		Glucose	Ethanol	Glycerol	Acetic acid
<i>In silico</i>	0.172	2.098	0	0	0
<i>In vivo</i>	0.159 ± 0.027	2.098 ± 0.404	0	0	0

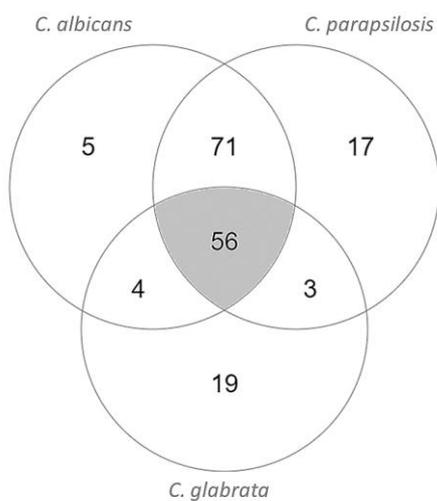
#### II.2.4.3. Enzyme essentiality assessment: looking for new drug targets

Identification of essential enzymes of a given pathogen should, in principle, lead to the identification of good drug targets, since if one enzyme is essential for its growth or survival, a compound capable of inhibiting it could turn out to be used as a drug with pharmacological activity against that pathogen. The drug target will be ideal if it is essential, or at least essential under conditions that mimic the human host environment, while having no human homolog. Based on these principles, iDC1003 was used to identify potential new drug targets by determining enzyme essentiality. For that, a list of essential reactions was obtained through simulation of the system's behaviour in RPMI medium, which mimics the environmental conditions of human serum. A total of 147 enzymes were predicted as essential in iDC1003, excluding drains, transport reactions, and

those without an associated or incomplete EC number. The complete list of predicted essential enzymes can be found in Supplementary Data II.2.1.

Finally, we decided to intersect the potential drug targets obtained through the *C. parapsilosis* model with those resulting from the two existing models for pathogenic *Candida* species, *C. albicans* [288] and *C. glabrata* [237] to focus on essential enzymes that can be used as targets in the treatment of infections caused by all *Candida* species. After the intersection, 56 essential enzymes common to the three models were found (Table II.10, Figure II.5). Consistently, well established antifungal drug targets were identified, including the well-known targets of azoles and echinocandins, Erg11 and Fks1, respectively.

Additionally, some of the identified predicted drug targets are homologs of enzymes used as drug targets against other pathogenic organisms, including Imh3, which is targeted by inosinic acid in *Streptococcus pyogenes*, Fol1, a target of sulfonamides, and Fas1, a target of ethionamide, used in the treatment against *Mycobacterium tuberculosis*. These confirmatory results illustrate the potential of the used approach in the quest for new drug targets. More interesting, however, is the identification of numerous potential targets, as identified herein, that are not currently targeted by any drug used in clinical practice. There are some new targets with tremendous potential as these do not have an orthologous enzyme in humans. Although not an excluding factor, the absence of a human orthologous is a preferable attribute as this translates into lower chances of host drug toxicity and may allow for greater freedom of drug design.



**Figure II.5** - Intersection of *C. parapsilosis*, *C. albicans*, and *C. glabrata* essential EC Numbers in RPMI medium environmental conditions in the genome-scale metabolic models iDC1003, iRV781, and iNX804, respectively. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)).

**Table II.10** - Enzymes predicted to be essential in RPMI medium, based on the screening of the genome-scale metabolic models of *C. parapsilosis*, iDC1003, *C. albicans*, iRV781, and *C. glabrata*, iNX804.

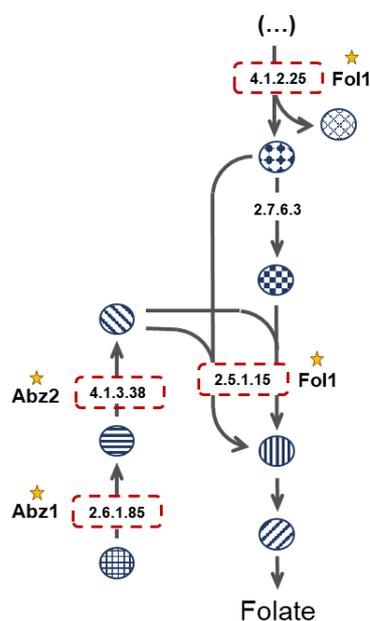
Gene name				Gene name			
<i>C. parapsilosis</i>	<i>S. cerevisiae</i>	Human	EC Number	<i>C. parapsilosis</i>	<i>S. cerevisiae</i>	Human	EC Number
	Homologue	Homologue			Homologue	Homologue	
CPAR2_302110	ERG26	NSDHL	1.1.1.170	CPAR2_805350	PEL1	PGS1	2.7.8.5
CPAR2_104580	IMD4	IMPDH	1.1.1.205	CPAR2_804250	PHO8	ALPL	3.1.3.1
CPAR2_801560	ERG27	DHRS11	1.1.1.270	CPAR2_602700	GEP4	PTPMT1	3.1.3.27
CPAR2_110330	HMG1	HMGCR	1.1.1.34	CPAR2_100500	URA4	CAD	3.5.2.3
CPAR2_405900	ERG24	TM7SF2	1.3.1.70	CPAR2_806200	IPP1	PPA2	3.6.1.1
ERG4	ERG4	-	1.3.1.71	CPAR2_805940	ADE2	PAICS	4.1.1.21
CPAR2_206550	TMP1	TYMS	2.1.1.45	URA3	URA3	UMPS	4.1.1.23

CPAR2_202250	ADE17	ATIC	2.1.2.3	CPAR2_109530	MVD1	MVD	4.1.1.33
CPAR2_203160	URA2	CAD	2.1.3.2	CPAR2_800750	CAB3	PPCDC	4.1.1.36
CPAR2_203160	URA2	CAD	6.3.5.5	CPAR2_303390	FOL1	-	4.1.2.25
CPAR2_106400	FKS1	-	2.4.1.34	CPAR2_303390	FOL1	-	2.5.1.15
CPAR2_802790	URA5	UMPS	2.4.2.10	CPAR2_212310	ABZ2	-	4.1.3.38
CPAR2_208260	ADE4	PPAT	2.4.2.14	CPAR2_204960	ADE13	ADSL	4.3.2.2
CPAR2_302840	BTS1	GGPS1	2.5.1.1	CPAR2_401630	IDI1	IDI1	5.3.3.2
CPAR2_103950	ERG20	FDPS	2.5.1.10	CPAR2_301800	ERG7	LSS	5.4.99.7
CPAR2_403110	ABZ1	-	2.6.1.85	CPAR2_212740	MET7	FPGS	6.3.2.17
CPAR2_502760	CAB5	COASY	2.7.1.24	CPAR2_500190	ADE1	PAICS	6.3.2.6
CPAR2_202590	FMN1	RFK	2.7.1.26	CPAR2_208400	ADE5,7	GART	6.3.3.1
CPAR2_602050	CAB1	PANK	2.7.1.33	CPAR2_208400	ADE5,7	GART	6.3.4.13
CPAR2_105320	URA6	CMPK2	2.7.4.14	CPAR2_803640	ADE12	ADSS	6.3.4.4
CPAR2_400710	ERG8	PMVK	2.7.4.2	CPAR2_204070	ADE6	PFAS	6.3.5.3
CPAR2_304260	PRS1	PRPS1	2.7.6.1	CPAR2_804060	ACC1	ACACA	6.4.1.2
CPAR2_500260	PIS1	CDIPT	2.7.8.11	CPAR2_803530	ERG12	MVK	2.7.1.36
CPAR2_211620	ADE8	GART	2.1.2.2	CPAR2_701400	ERG13	HMGCS	2.3.3.10
CPAR2_602300	COQ3	COQ3	2.1.1.114	FAS1	FAS1	-	2.3.1.86
CPAR2_209250	COQ5	COQ5	2.1.1.201	CPAR2_803560	GUA1	GMPS	6.3.5.2
ERG11	ERG11	CYP51A1	1.14.14.154	CPAR2_100620	URA7	CTPS1	6.3.4.2
CPAR2_303080	GUK1	GUK1	2.7.4.8	CPAR2_804900	URA1	-	1.3.98.1

Blue: enzymes without any human homolog or drug association. Red: enzymes targeted by drugs currently used to treat *Candida* infections. Green: enzymes with homologs that are currently targeted in the treatment of infections caused by other pathogens.

Among the identified potential new drug targets, Abz1/2, Erg4, and Ura1 emerge as enzymes without any human homolog or drug association. Fungi rely on folate de novo biosynthesis given their inability to uptake folate from the environment [296]. *FOL1*, *ABZ1* and *ABZ2* encode key enzymes in the folate

biosynthesis pathway (Figure II.6). Furthermore, these enzymes have no human orthologs, as humans rely on a diet derived folate [296]. The dihydropteroate synthase encoded by *FOL1* has been shown to be successfully inhibited by antifolates such as sulfonamides in a series of microorganisms [296,297]. However, antifolate therapy for *Candida* infections is not particularly effective considering current antifolate compounds [298]. In fact, for *C. albicans* only sulfanilamide is used clinically, and it is restricted to topical use [271]. Given the efficacy of antifolates in treating infections by other etiologic agents, this might present the opportunity to design new effective antifungal compounds against *Candida* Fol1. On the other hand, no inhibitors of the para-aminobenzoate synthetase encoded by *ABZ1*, or of the 4-amino-4-deoxychorismate lyase encoded by *ABZ2*, are currently known, making these two enzymes fully novel putative drug targets worth for further exploitation.



**Figure II.6** - Folate biosynthetic pathway. Red boxes highlight the enzymes considered essential in the three analysed models.

The Ura1 and Erg4 proteins may also be interesting drug targets. Ura1 has no human ortholog and, although this protein is not the target of any known drug, Aro9, from the same pathway, is a known target of atovaquone used to treat *Plasmodium falciparum* infections. In turn, Erg4, also with no human ortholog, is involved in ergosterol biosynthesis, a pathway currently targeted by azole drugs.

## II.2.5. Conclusions

The first validated global metabolic model for the human pathogen *C. parapsilosis* is presented in this study. The model was manually curated and validated experimentally and proved to be able to predict the main metabolic features of *C. parapsilosis* quantitatively. Furthermore, iDC1003 is robust in predicting the use of several carbon and nitrogen sources. The model shares 85% of the proteins with an associated EC number in other already published yeast models. However, 15% of them are exclusive to this model and may represent some unique metabolic features of *C. parapsilosis*. Using iDC1003, several enzymes were predicted to be essential in RPMI medium, including some already known targets of antifungal agents and other antimicrobial agents used in clinical practice, illustrating the potential of the used approach in the quest for new drug targets. Several of the identified potential drug targets are not currently targeted by any drug used in clinical practice, deserving further study. Among the identified potential new drug targets, Abz1/2, Erg4, and Ura1 stand out as enzymes without any human homolog or drug association. However, all other targets are worthy of scrutiny as, in fact, many of the drug targets currently targeted by drugs used in clinical practice have human orthologs. In these cases, the strategy may involve taking advantage of the structural differences between the proteins in the two organisms to design efficient compounds.

Despite the clear usefulness of these types of models, it is important to highlight that these reconstructions have limitations. Firstly, the basis of GSMMs is the genome's functional annotation. Depending on the stringency of the criteria imposed on hit selection, such a procedure may lead either to compromising rates of false positive or negative annotations. Furthermore, such models do not encompass regulatory processes due to the high complexity of such an integration. Note how enzymatic activity can be regulated at different levels – from gene expression to post-translational modifications - and may result in given pathways being preferential in certain environmental conditions. The exclusion of such mechanisms, particularly regarding essentiality prediction, may result in predicted essential ECs that would otherwise not be metabolically relevant in the conditions of interest. Lastly, these simulations do not consider supra-metabolic factors, such as stress factors. Even so, and considering all these limitations, GSMMs allow for increasingly guided and reliable drug target discovery – as illustrated in this paper.

#### **II.2.6. Acknowledgments**

This work was supported by “Fundação para a Ciência e a Tecnologia” (FCT) (Contract PTDC/BII-BIO/28216/2017 and AEM PhD grant to RV). Funding received from project LISBOA-01-0145-FEDER-022231-the BioData.pt Research Infrastructure, is acknowledged. This work was further financed by national funds from FCT in the scope of the project UIDB/04565/2020 and UIDP/04565/2020 of the Research Unit Institute for Bioengineering and Biosciences – iBB, project UIDB/04469/2020 for the Centre of Biological Engineering - CEB, and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy - i4HB.

## II.3. Metabolic reconstruction of the human pathogen *C. auris*: using a cross-species approach for drug target prediction

### II.3.1. Abstract

*C. auris* is an emerging human pathogen, associated to antifungal drug resistance and hospital candidiasis outbreaks. Genome-scale metabolic models (GSMMs) have recently been used as a platform for drug target identification, a feature that may prove particularly valuable in the search for new *C. auris* drug targets.

In this work, we present iRV973, the first reconstructed GSMM for *C. auris*. The model, provided in the SBML format, was manually curated and experimentally validated, being able to accurately predict the specific growth rate of *C. auris* and the utilization of several sole carbon and nitrogen sources. The model was compared to GSMMs available for other pathogenic *Candida* species and exploited as a platform for cross-species comparison, aiming the analysis of their metabolic features and the identification of potential new antifungal targets common to the most prevalent pathogenic *Candida* species. From a metabolic point of view, we were able to identify some unique enzymes in *C. auris* in comparison with other existing *Candida* models, which may represent unique metabolic features of *C. auris*. Additionally, 50 enzymes were identified as potential drug targets, given their essentiality in conditions mimicking human serum, common to all the 4 different *Candida* models analyzed. These enzymes represent interesting drug targets for antifungal therapy, some are already known targets of antifungal agents used in clinical practice, but other new potential drug targets also stand out without any human homolog or drug association in *Candida* species.

**Keywords:** *C. auris*; Global stoichiometric model; Gene essentiality; Drug target; Metabolic features

### II.3.2. Introduction

*C. auris* is an emerging human pathogen. First isolated in 2009 [299], *C. auris* is geographically non-restricted [300] and thrives in nosocomial environments [301]. While *C. albicans* is still the most frequent etiological agent of candidiasis, a shift towards non-*C. albicans* *Candida* species (NCAC) [302] has been registered since the 1990s [302]. Among the NCAC, *C. glabrata*, *C. parapsilosis*, and *Candida tropicalis* account for the majority of candidiasis cases. Although the prevalence of *C. auris* among candidiasis patients is relatively low, concern is raising on its impact as a serious threat to public health due to three factors: 1) its diagnosis is difficult using standard methods, which may lead to non-optimal therapeutic choices; 2) resistance against multiple antifungal drug families has been often identified in clinical isolates; and 3) it causes outbreaks in healthcare settings [299], possibly due to its ability to resist for long periods on hospital surfaces. The isolation of *C. auris* strains displaying resistance to the three main classes of available clinically used antifungal drugs (azoles, echinocandins, and polyenes), highlights the need to identify novel antifungal drug targets and design entirely new effective antifungal therapies.

Genome-Scale Metabolic Models (GSMMs) are *in silico* reconstructions of global metabolic networks. They have been used to model the full metabolism of individual organisms [120] constituting a Systems Biology tool to analyze metabolic features [119]. GSMMs have been used for decades for guiding strain design, in a metabolic engineering perspective. More recently, however, they have also been used in the elucidation of pathogenesis and drug target

identification [112]. Interestingly, such models have been built for the most clinically relevant *Candida* species, including *C. albicans* [288], *C. glabrata* [237], and *C. parapsilosis* [303], but not yet for *C. auris*. The exploitation of these GSMMs provide recent examples of the reliability of this approach in the identification of putative novel drug targets in pathogenic organisms.

Herein, the construction and validation of the first GSMM for the human pathogen *C. auris* is described. This model is provided in the well-established SBML format and can easily be read in most metabolic engineering platforms, such as OptFlux [128] and COBRA [304]. A set of predicted essential genes was obtained from the validated model, and their targetability as putative novel antifungal drug targets is discussed. Additionally, the existent *Candida* GSMMs were exploited as a platform for cross-species comparison, aiming the analysis of their metabolic features, including the study of reactions and pathways which are common to all *Candida* species and the identification of reactions and pathways that are unique to each pathogen.

### **II.3.3. Materials and Methods**

#### **II.3.3.1. Model Construction**

The metabolic model described here refers to the yeast *C. auris* with the taxonomic ID 498019. Network reconstruction was performed using *merlin* 4.05 [122] and subsequent curation and validation were performed on OptFlux 3.0 [128] using the IBM CPLEX 12.10 solver. Curation consisted of the editing, addition, or removal of reactions so to correct previous gaps in the network using KEGG pathways, MetaCyc Database, and literature data as reference.

### **II.3.3.1.1. Enzyme and Reaction Annotation**

Draft model construction consisted of an initial enzyme and reaction annotation from the reference genome sequence of the reference strain *C. auris* B11221\_V1, retrieved from NCBI's Assembly Database, accession ID "Cand\_auris\_B11221\_V1" along with the respective Taxonomic ID [223]. Genome-wide functional annotation was performed by *merlin* by homology alignment assessment with resource to BLAST [226] against the remote databases UniProtKB/Swiss-Prot [227]. Hit selection was performed as described in Dias *et al.* 2018 [221] and phylogenetic proximity was implemented based on phylogenetic tree from literature [305]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [280] database served as reference for protein-reaction associations and consequent draft network construction.

### **II.3.3.1.2. Correcting Reaction Reversibility, Directionality, and Balance**

Reaction reversibility curation comprised an initial automatic curation step followed by extensive manual curation. Automatic curation was performed by *merlin* referencing remote databases such as eQuilibrator [281], as described by Dias *et al.* [221]. Manual curation referenced other databases such as Meta-Cyc [231] as well as existing literature. Unbalanced reactions were identified automatically and manually balanced with resort to information from databases such as MetaCyc, ChEBI [282], Brenda [283] and existing literature. All manually edited reactions can be found in Supplementary Data II.3.1.

#### **II.3.3.1.3. Compartmentalization**

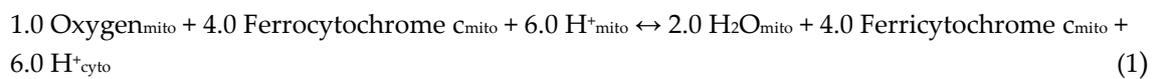
The model described herein comprises four compartments: extracellular, cytoplasm, mitochondrion, peroxisome and one intercompartment: the cytoplasmic membrane. The online protein localization predictor tool WoLF PSORT [232] was used in model compartmentalization and was implemented on the already connected non-compartmentalized model through direct import of the report, as *merlin* has an integrated import tool for WoLF PSORT and other similar tools. Transport reactions – which connect the networks compartments – were generated from the genome sequence with resort to the public database TCDB [233] and using *merlin's* integrated tool TranSyT [127]. Transport reactions across internal and external membranes for currency metabolites, such as H<sub>2</sub>O, CO<sub>2</sub>, and NH<sub>3</sub>, often carried by facilitated diffusion, were added to the model with no gene association.

#### **II.3.3.1.4. Defining the Biomass Equation**

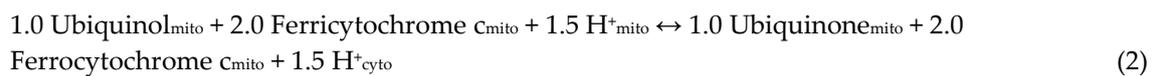
The biomass equation comprises all the cell's major components along with their respective relative numerical contributions - DNA, RNA, carbohydrates, lipids, and proteins – and was used as the objective function in all performed simulations, including essentiality prediction. The reconstructed model also includes ATP requirements for both biomass production and cell maintenance – growth-associated maintenance (GAM) and non-growth-associated maintenance (NGAM), respectively. A GAM value of 25.65 mmol ATP/gDCW was considered for the biomass equation, calculated based on the ATP requirements for the biosynthesis of cell polymers, as shown in Mishra et al. [236], then adjusted for the considered biomass macromolecule composition. Non-growth-associated ATP maintenance was implemented in this model as an autonomous equation,

thus forcing a basal ATP consumption —flux bounds inferred from *Candida tropicalis* [236]. The Supplementary Data II.3.1 details the biomass equation's components and their relative content. The theoretical ratio used in the *Saccharomyces cerevisiae* iMM904 metabolic model for the phosphorus-to-oxygen ratio was applied. Three generic reactions contributing to this ratio were automatically generated by *merlin* and were updated to replicate the same ratio as in the iRV973 model:

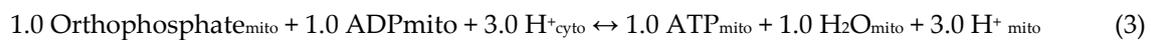
**Reaction R00081:**



**Reaction T02161:**



**Reaction T00485:**



**The final balance reaction:**



**II.3.3.2. Model Simulations and Enzyme Essentiality Prediction**

Flux balance analysis (FBA) [284] on OptFlux 3.0 [128] using the IBM CPLEX solver was used to conduct model simulations. Genes and reactions were considered essential if when removed from the model the resulting simulation's predicted biomass flux is less than 5% of the wild-type strain. Gene and reaction removal was simulated by restraining its corresponding flux bounds to zero. Gene/reaction essentiality was searched for in environmental conditions mimicking the human serum, namely the composition of the RPMI medium.

### **II.3.3.3. Model Validation**

#### **II.3.3.3.1. Strains and Growth Media**

*C. auris* type strain B11221 was routinely maintained in Yeast extract–Peptone–Dextrose (YPD) and synthetic minimal media (SMM) was used for batch cultured experiments. YPD contained: 20 g/L glucose (Merck, Darmstadt, Germany), 20 g/L peptone (Merck, Darmstadt, Germany), and 10 g/L yeast extract (Merck, Darmstadt, Germany), while SMM included: 20g/L glucose (Merck, Darmstadt, Germany), 2.7 g/L ammonium sulphate (Merck, Darmstadt, Germany), 0.05 g/L magnesium sulphate (Riddle-de-Haen), 2 g/L potassium dihydrogen phosphate (Panreac, Barcelona, Spain), 0.5 g/L calcium chloride (Merck, Darmstadt, Germany), and 100 µg/L biotin (Sigma).

#### **II.3.3.3.2. Aerobic Batch Cultivation**

Aiming the validation of the constructed model, *C. auris* B11221\_V1 cells were batch cultivated in Erlenmeyer flasks containing 250ml of SMM medium, at 30 °C (250 rpm). Exponential phase inocula (Optical Density at 600 nm (OD<sub>600nm</sub>) of 0.3) were prepared and cells were transferred Erlenmeyer flasks containing 250ml of fresh SMM medium and cultivated at 30 °C with orbital agitation (250 rpm) for 10 h.

#### **II.3.3.3.3. Cell Density, Dry Weight, and Metabolite Concentration Assessment**

Throughout cell cultivation in SMM 4 mL samples were retrieved every two hours for subsequent biomass and extracellular metabolite quantification, and cell density was monitored through OD<sub>600nm</sub> measuring. For dry weight determination, culture samples were centrifuged at 13,000 rpm for 3 min and the pellets were lyophilized for 72h at -80 °C and weighted. For extracellular

metabolite identification and quantification and the concentrations of glucose, ethanol, glycerol, and acetic acid in the supernatants were determined by HPLC on an Aminex HPX-87 H Ion Exchange chromatography column, eluted with 0.0005 M H<sub>2</sub>SO<sub>4</sub> at a flow rate of 0.6 mL/min at room temperature. Samples were analysed in triplicate and concentrations determined via adequate calibration curves. The specific growth rate, specific glucose consumption rate, and the specific production rates of ethanol, glycerol and acetic acid were calculated during the exponential growth phase as indicated elsewhere [285].

#### **II.3.3.3.4. Carbon and Nitrogen Source Utilization Assessment**

The capability of utilizing different carbon and nitrogen sources for cell growth was assessed by comparing *in silico* predictions to literature data for *C. auris*. For the carbon sources for which the model predictions were not consistent with literature data, experimental testing was conducted with the B11221 *C. auris* reference strain used in the model reconstruction, and also with representative strains from clades I and II, *C. auris* B8441 and *C. auris* B11220, respectively. The utilization of ethanol, glycerol, and galactitol as sole carbon source, by the *C. auris* strains was evaluated in solid YNB medium containing: either 5 g/L glucose as control, or 5 g/L of either one of the mentioned carbon sources; 1.7 g/L Yeast nitrogen base without amino-acids without ammonium sulfate (Sigma); 5 g/L Ammonium sulfate (Merck, Darmstadt, Germany) and 20g/L Agar (Iberagar, Barreiro, Portugal). *C. auris* cell suspensions used to inoculate the agar plates, were mid-exponential cells grown in YNB medium with 5 g/L glucose, until culture OD<sub>600nm</sub> = 0.5 ± 0.05 was reached and then diluted in sterile water to obtain suspensions with OD<sub>600nm</sub> = 0.05 ± 0.005. These cell suspensions and subsequent dilutions (10<sup>-1</sup> ; 10<sup>-2</sup> ; 10<sup>-3</sup> ) were applied as 4 µL spots onto the surface

of solid YNB media, with the indicated carbon sources. Growth was assessed after incubation at 37 °C for 48 hours.

## **II.3.4. Results and Discussion**

### **II.3.4.1. Model characteristics**

The *C. auris* metabolic model, iRV973, comprises 973 genes associated with 2863 reactions – of which 352 are drains (exchange constraints for external metabolites used to simulate the environmental conditions allow metabolites to be removed from the system) and 1035 are transport reactions - and 2150 metabolites across four compartments (extracellular, cytoplasm, mitochondria, and peroxisome) and within an intercompartment, the plasma membrane. The model can be found in SBML format in Supplementary Data II.3.2.

Manual curation assessed a total of 776 reactions requiring alterations, from which 75 were mass balanced and 701 were corrected regarding reversibility, directionality, annotation, or added or removed from the model as detailed in Supplementary Data II.3.1.

The Biomass equation (Table II.11) encompasses the cell's major components along with their respective and relative contributions, DNA, RNA, lipids, carbohydrates, and cofactors. The equation's composition in Carbohydrate (M. Pfaller & Riley, 1992), Lipid [236,287], Sterol [236], Phospholipid [236], Fatty acid [252] and Protein [236] was inferred from literature data. On the other hand, for the composition of DNA, the whole genome sequence was used to estimate the amount of each deoxyribonucleotide as described in [123], this was automatically performed using the e-BiomassX *merlin* tool, while mRNA, rRNA, and tRNA

were used to estimate the total RNA in the cell as described in [111] automatically performed using the e-BiomassX *merlin* tool. The growth and non-growth ATP requirements were inferred from *Candida tropicalis* [236].

**Table II.11** - Biomass composition used in the model iRV973. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.3.1.

Metabolite	g/gDCW	Metabolite	g/gDCW
<b>Lipids<sup>1</sup></b>		<b>Proteins<sup>2</sup></b>	
Lanosterol	0,000054	L-Valine	0,021208
Zymosterol	0,000189	L-Tyrosine	0,025060
5,7,22,24(28)-Ergostatetraenol	0,000269	L-Tryptophan	0,015297
Ergosterol	0,002047	L-Threonine	0,011456
Phosphatidylserine	0,018028	L-Serine	0,004758
Phosphatidylinositol	0,005408	L-Proline	0,015616
Phosphatidylcholine	0,008089	L-Phenylalanine	0,021862
Phosphatidylethanolamine	0,006425	L-Methionine	0,007562
Cardiolipin	0,003883	L-Lysine	0,020142
Tetradecanoic acid	0,000026	L-Leucine	0,017778
Hexadecanoic acid	0,001329	L-Isoleucine	0,002905
Octadecanoic acid	0,000489	L-Histidine	0,016385
Oleic acid	0,002276	L-Glutamate	0,015260
alpha-Linolenic acid	0,000109	L-Cysteine	0,027463
Triacylglycerol	0,013575	L-Aspartate	0,014186
Diacylglycerol	0,001001	L-Asparagine	0,007636
Sterol esters	0,014687	L-Arginine	0,017066
		L-Alanine	0,006159
		Glycine	0,020616
		L-Glutamine	0,009458
<b>Carbohydrates<sup>3</sup></b>		<b>Soluble Pool<sup>4</sup></b>	
Chitin	0,026552	Thiamine	0,000174
Mannan	0,181732	Ubiquinone-9	0,000174
$\beta$ (1,3)-Glucan	0,381756	NADP+	0,000174
		NAD+	0,000174
<b>Ribonucleotides</b>		FMN	0,000174
UTP	0,008262	FAD	0,000174
GTP	0,006585	CoA	0,000174
CTP	0,008267	Biotin	0,000174
ATP	0,006984	Pyridoxal phosphate	0,000174
<b>Deoxyribonucleotides</b>			
dTTP	0,000647		

dGTP	0,000537	5-Methyltetrahydrofolate	0,000174
dCTP	0,000534		
dATP	0,000646		

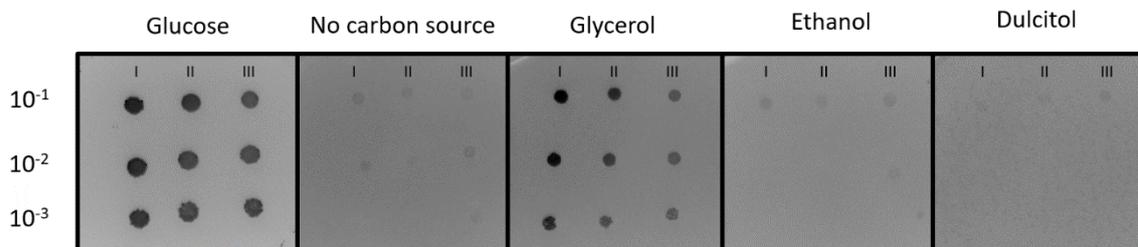
---

## II.3.4.2. Model validation

### II.3.4.2.1 Carbon and nitrogen source utilization

Simulations were performed *in silico* SMM medium with different molecules as sole carbon and nitrogen sources and compared to phenotypic growth data from existing literature to assess iRV973's reliability in predicting biomass production from different sole carbon or nitrogen sources. Data related to *C. auris* strains, other than the reference B11221 strain, were also considered in the analysis to increase the number of carbon and nitrogen sources accounted for. A total of 36 sole carbon sources and 8 sole nitrogen sources were evaluated. In an initial assessment our model correctly predicted growth in 92% (33/36) of the carbon sources tested and in 100% (8/8) of the nitrogen sources. The model failed to predict growth in galactitol and wrongly predicted growth in ethanol and glycerol, however, since the literature data is based on strains different from the one used to reconstruct our model, CBS10913 [299] and several other clinical isolates [306,307], we decided to test experimentally these 3 cases with one representative strain of *C. auris* clades I, II and III (B8441, B11220 and B11221). The results confirmed the model's prediction regarding the utilization of glycerol as carbon source and regarding the non-utilization of dulcitol (Figure II.7). Interestingly, none of the *C. auris* strains tested was able to grow in ethanol as the sole carbon source, contradicting the model's prediction and suggesting that, although having all the required enzymes for ethanol utilization, *C. auris* cannot

use it, which may be due to regulatory reasons not accounted for in the model. Altogether our model only failed to predict growth in ethanol and was able to correctly predict 97% of the carbon sources tested (Table II.12).



**Figure II.7** - Utilization of glucose (positive control), no carbon source (negative control), glycerol, ethanol, and dulcitol as carbon source by *C. auris* strains I: B8441 II: B11220 and III: B11221 in solid YNB medium. Initial OD<sub>600nm</sub> =  $0.5 \pm 0.05$ . Growth was assessed after incubation at 37 °C for 48 h.

A particular case worth mentioning is the use of melezitose as a sole carbon source. According to the literature, it is known that *C. auris* is able to use this compound as a carbon source [299,306,308], this being a metabolic feature that distinguishes *C. auris* from other *Candida* species. However, as far as it could be assessed, there is no experimental evidence of any enzymes behind melezitose degradation. As far as we know, the melezitose degrading enzymes remain unknown for any species, making it impossible to add the corresponding reactions to the model.

**Table II.12** - Comparison between *in vivo* and *in silico* phenotypic behavior of *C. auris* under different carbon and nitrogen sources.

	<i>In vivo</i>	<i>In silico</i>	Reference	<i>In vivo</i>	<i>In silico</i>	Reference
<b>Carbon Source</b>						
D-Glucose	+	+	1,4,5	D-Xylose	-	- 4
D-Glucosamine	-	-	4	L-Rhamnose	-	- 3,4
L-Arabinose	-	-	3,4	a,a-Trehalose	+	+ 4,5
Sucrose	+	+	1,4,5	Salicin	-	- 4
β-D-galactoside	-	-	4	Raffinose	+	+ 4
Melibiose	-	-	4	Starch	+	+ 4
Glycerol	+	+	3,4	Ribitol	-	- 4
Xylitol	-	-	3,4	D-Mannitol	+	+ 4
Galactitol	-	-	4	2-Keto-D-Gluconate	-	- 4
D-Gluconate	-	-	3,4	Succinate	-	- 3,4
Citrate	+	+	4	Ethanol	-	+ 1,3,4
D-Galactose	-	-	3,4	Methanol	-	- 3,4
D-Ribose	-	-	4	N-acetylglucosamine	-	- 2,4
D-Arabinose	-	-	3,4			
Maltose	+	+	4,5	<b>Nitrogen Source</b>		
Cellobiose	-	-	4	Nitrite	-	- 4
Lactose	-	-	1,4	Cadaverine	+	+ 4
Inulin	W	-	3,4	Nitrate	-	- 1,2,4
Erythritol	-	-	4	L-Lysine	+	+ 4
D-Sorbitol	+	+	4	Ethylamine	-	- 4
myo-Inositol	-	-	4	Ammonia	+	+ 4
DL-Lactate	-	-	4	Urea	-	- 2,4
L-Sorbose	-	-	3,4	Glycine	-	- 1

Growth (+); lack of growth (-); 1:[307]; 2:[306]; 3:[308]; 4:[299]; 5:[309]

#### II.3.4.2.1 Growth parameters in batch culture

Specific growth rate, glucose consumption rate, and metabolite production rates were experimentally determined to validate the model quantitatively. A specific growth rate of 0.110 h<sup>-1</sup> was observed for a glucose consumption rate of 2.19 mmol.gDCW<sup>-1</sup>.h<sup>-1</sup>, while no ethanol, glycerol or acetate production could be detected. In order to compare with the *in silico* results, a simulation of the

system's behavior in SMM medium, with a fixed glucose uptake flux of 2.19 mmol.g<sup>-1</sup> dry weight.h<sup>-1</sup> was performed to evaluate the model's ability to predict the specific growth rate. The remaining nutrient fluxes were left unconstrained, as the system in these conditions is glucose-limited. In this simulation the model predicted a specific growth rate of 0.155 h<sup>-1</sup>, similar to the experimentally determined value (Table II.13). Additionally, the formation of glycerol, acetic acid, and ethanol as by-products was not predicted to occur, which agrees with the experimental data and the notion that *C. auris* is a Crabtree-negative yeast.

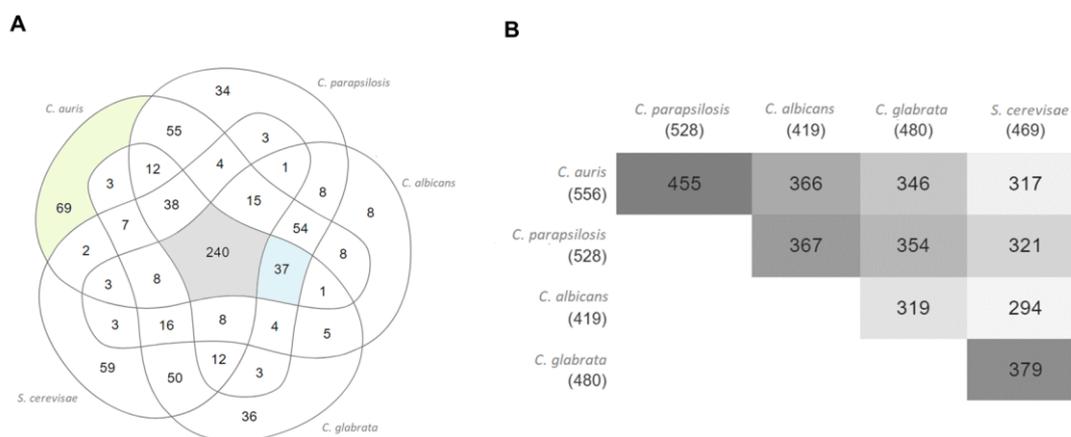
**Table II.13** - Growth parameters of iRV973 and comparison with experimentally determined values for *C. auris*.

	Specific growth rate (h <sup>-1</sup> )	q (mmol g <sup>-1</sup> dry weight h <sup>-1</sup> )			
		Glucose	Ethanol	Glycerol	Acetic acid
<i>In silico</i>	0.153	2.19	0	0	0
<i>In vivo</i>	0.110	2.19	0	0	0

#### II.3.4.3. *C. auris* unique metabolic features

To identify potentially unique metabolic features of *C. auris*, the iRV973 model was compared to the *C. glabrata* [237], *S. cerevisiae* [238], *C. albicans* [288] and *C. parapsilosis* [303] GSMMs. A comparison across the existing models was carried out based on shared EC numbers. After intersecting the EC numbers present in each of the five models, 88% (467/556) of the proteins with an associated EC number in the *C. auris* model were found to also be present in at least one of the other four models (Figure II.8). However, the remaining 12% (69/556) are exclusive to this model and may represent unique metabolic features of *C. auris*. It is also interesting to point out that 37 EC numbers are shared by all *Candida* species and absent in *S. cerevisiae*, which could represent unique features of pathogenic yeasts. However, after a literature search and a manual inspection of

the *S. cerevisiae* model we were not able to identify any *Candida* specific enzyme, in all the cases the enzymes were also present in *S. cerevisiae*, and not in the GSMM, probably due to changes in annotation, or update of EC numbers. The complete list of “unique” EC numbers can be found in Supplementary Data II.3.1.



**Figure II.8:** Comparison between *C. auris*, *C. parapsilosis*, *C. albicans*, *S. cerevisiae*, and *C. glabrata* proteins with associated EC Numbers present in the iRV973, iDC1003, iRV781, iIN800 and iNX804 genome-scale metabolic models, respectively. A: Venn Diagram. B: Pairwise Intersections. Diagrams were obtained using Multiple List Comparator ([www.molbiotools.com](http://www.molbiotools.com)).

From the 69 *C. auris* unique EC numbers, some are related to outdated/incomplete EC numbers or associated with other enzymes responsible for the same reactions in the different models. However, some EC numbers stand out as potentially real unique features of *C. auris*:

- 1.1.1.24: this EC number represents a quinate dehydrogenase that catalyzes the oxidation of quinate to 3-dehydroquinate and is involved in quinate degradation, this enzyme seems to be a unique feature of *C. auris* among *Candida* species and

may allow this yeast to utilize quinate as source for the biosynthesis of folates, quinones, and aromatic amino acids [310].

- 1.1.1.10: this EC number represents a chloride peroxidase, which is able to catalyze the production of hypochlorous acid by transferring one oxygen atom from H<sub>2</sub>O<sub>2</sub> to chloride. This is a very interesting feature since hypochlorous acid is used for microbial killing by phagocytic cells and in nature this enzyme is often associated with the production of halogenated antibiotics [311]. In *C. auris* this enzyme may represent a competitive advantage or even a potential mechanism of defense to the host.

- 1.13.11.24: this EC number represents a quercetinase which catalyzes the degradation of quercetin, a molecule that exhibits antioxidant and antibacterial properties and is present in edible fruits and vegetables, being part of the human diet and also of soil composition due to degradation of those fruits. Some microorganisms exposed to these compounds developed the ability to degrade it [312]. This may be the case for *C. auris*.

- 3.1.1.65: this EC number represents a L-rhamnono-1,4-lactonase, which participates in the L-rhamnose degradation pathway. Although having some enzymes of this pathway, *C. auris* seems to be unable to use L-rhamnose as sole carbon source.

- 3.5.1.128: this EC number represents a deaminated glutathione amidase which may be involved in the clearance of toxic deaminated glutathione, working as a repair enzyme [313].

#### II.3.4.4. Drug target forecast based on gene essentiality prediction: a cross-species evaluation

IRV973 was used to identify potential new drug targets by determining enzyme essentiality. A list of essential reactions was obtained through simulation of the system's behavior in RPMI medium, which mimics the environmental conditions of human serum, in order to narrow down the results to enzymes that should be essential in the conditions that a pathogen faces inside the host. A total of 158 reactions and 82 genes were predicted as essential in iRV973. The complete list of predicted essential genes and reactions in *C. auris* can be found in Supplementary Data II.3.1.

As seen before using a similar approach for *C. albicans* and *C. parapsilosis*, proteins which are targeted by currently used antifungal drugs were identified, including the targets of azoles, Erg11, and echinocandins, Fks1. This observation confirms the potential of using this approach in the quest for new drug targets. The most interesting outcome of this analysis, however, is the identification of numerous potential targets that are not currently used in clinical practice.

In the search for potential new antifungal drug targets suitable to be used against all forms of candidiasis, the enzymes predicted to be essential in *C. auris* were compared with those predicted to be essential in *C. albicans*, *C. parapsilosis* and *C. glabrata* [237,288,303]. A total of 50 enzymes were predicted to be essential in the 4 species, according to the simulations run in the 4 corresponding GSMs Supplementary Table II.3.1. From this list we excluded drains, transport reactions, and those without an associated or incomplete EC number. The 50 essential predicted enzymes can be grouped in 13 different pathways: steroid biosynthesis, purine metabolism, pyrimidine metabolism, terpenoid backbone biosynthesis, 1,3-beta-glucan biosynthesis, CoA biosynthesis, riboflavin

metabolism, glycerophospholipid metabolism, ubiquinone biosynthesis, folate biosynthesis, oxidative phosphorylation, fatty acid biosynthesis, and chitin biosynthesis. Some of these pathways are already targets of drugs used in clinical practice (Supplementary Table II.3.1), however, some could be explored as potential targets, in this work we not only propose these pathways as potential targets but also identify possible enzymes that should receive special focus in each of them.

For example, Erg4 (ergosterol biosynthesis) Chs2 (chitin biosynthesis), and Fol1 (folate biosynthesis) are targets without an orthologous enzyme in humans that are currently not targeted by any drug in *Candida* species. The absence of a human orthologous is a preferable attribute as this translates into lower chances of host drug toxicity and may allow for greater freedom of drug design, making these targets good candidates for drug targeting. However, the potential of the remaining candidates should not be discarded. Indeed, there are several reported cases of drug targets for which there are human orthologs. In those cases, the drug specificity relies on its design favoring affinity to the pathogen's target, when compared to the eventual human target.

### **II.3.5. Conclusions**

This study represents a significant milestone in the field of fungal pathogen research, as it presents the first validated global metabolic model for the human pathogen *C. auris*. The manual curation and experimental validation of the model have demonstrated its capability to quantitatively predict the main metabolic features of *C. auris* and its utilization of various metabolites as sole carbon or nitrogen sources. The overlap of 88% of the proteins with associated EC numbers in other already published yeast models establishes a foundation of shared

metabolic pathways among yeast species. Notably, the remaining 12% of exclusive proteins in this model may unveil unique metabolic characteristics specific to *C. auris*.

The utilization of iRV973 to predict essential genes in RPMI medium has unveiled a list of potential drug targets. Remarkably, 50 of these essential genes have been predicted as vital in all *Candida* species for which GSMMs are available. Among these identified potential drug targets, some encode proteins already known to be targets of antifungal agents and other antimicrobial agents used in clinical practice, demonstrating the potential of this approach in identifying new drug targets. Of particular interest are the enzymes lacking any human homolog or drug association, making them highly promising candidates for new drug targets. Targeting such enzymes could lead to novel therapeutic strategies with reduced risks of off-target effects on human cells.

Looking to the future, this validated global metabolic model for *C. auris* opens exciting prospects for further research and applications. It can be used as a platform to investigate the metabolic network of *C. auris* and with advancements in computational biology and systems biology, this model could be further refined and expanded to encompass additional aspects of the pathogen's biology.

In summary, the publication of this validated global metabolic model for *C. auris* provides valuable insights into its metabolism and offering potential new clues for therapeutic interventions against *C. auris* and potentially other *Candida* species, given the alarming increase in drug-resistant fungal infections worldwide, the identification of new drug targets and the understanding of unique metabolic features are of utmost importance in combating these pathogens effectively.

### **II.3.6. Acknowledgments**

This work was supported by “Fundação para a Ciência e a Tecnologia” (FCT) (Contract PTDC/BII-BIO/28216/2017 and AEM PhD grant to RV). Funding received from project LISBOA-01-0145-FEDER-022231-the BioData.pt Research Infrastructure is acknowledged. This work was further financed by national funds from FCT in the scope of the project UIDB/04565/2020 and UIDP/04565/2020 of the Research Unit Institute for Bioengineering and Biosciences—iBB, project UIDB/04469/2020 for the Centre of Biological Engineering—CEB, and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB.

## **II.4. Unveiling new features of the human pathogen *C. neoformans* through the construction and exploitation of a dedicated Genome-Scale Metabolic Model**

### **II.4.1. Abstract**

*C. neoformans* is notorious for causing severe pulmonary and central nervous system infections, particularly in immunocompromised patients. High mortality rates even with access to the best medical care associated with its tropism and adaptation to the brain microenvironment and its intrinsic resistance to echinocandins and an observed increase in fluconazole resistance in the past decades, makes this pathogen a public health threat and a World Health Organization (WHO) priority.

In this study, the GSMM iRV890 was constructed for *C. neoformans var. grubii*, offering a promising platform for comprehensively understanding the unique metabolic features of *C. neoformans*, shedding light on its complex tropism for the microenvironment within the human and mammalian brain and potentially guiding the discovery of new drug targets. The model, available in the SBML format, underwent validation using experimental data for nitrogen and carbon assimilation, as well as specific growth and glucose consumption rates. Based on the comparison with GSMMs available for other pathogenic yeasts, unique metabolic features were predicted for *C. neoformans*, including key pathways shaping the dynamics between *C. neoformans* and the human host, and underlying its adaptation to the brain environment. Finally, predicted essential genes from the validated model are explored herein as potential antifungal drug targets.

**Keywords:** *C. neoformans*, Global stoichiometric model, Drug targets, Metabolic features, neurotropism.

#### II.4.2. Introduction

Cryptococcal meningitis is a disease caused by some pathogenic species of basidiomycetous yeasts, namely *Cryptococcus neoformans* (*C. neoformans*) and *Cryptococcus gatii*. These yeasts are notorious for their potential to induce severe pulmonary and central nervous system infections [314]. While this pathogen is harmless in healthy individuals, it poses a serious threat to immunocompromised patients, especially with acquired immunodeficiency syndrome (AIDS) or those undergoing immunosuppressive therapies, causing severe meningoencephalitis and other grave neurological complications [315–317]. According to a recent systematic review using data from more than 120 countries, it is estimated that Cryptococcal meningitis affects 190 000 people worldwide annually, however, this situation is particularly concerning due to the extremely high mortality rates of those infections, with 147 000 deaths annually and a mortality rate of 76% [3]. *Cryptococcus* species show relatively infrequent drug resistance since they are commonly treated with combination therapy, usually flucytosine in combination with amphotericin B in a first induction stage, followed by consolidation and long-term maintenance with high dose fluconazole [26]. Fluconazole, due to its fungistatic mode of action, is not considered optimal for anti-cryptococcal monotherapy, with the risk of developing drug resistance [318]. In fact, an increase in fluconazole resistance among *C. neoformans* isolates was noted in the past decades, rising from 7.3% in 1997 to 11.7% in 2007 [27]. Fluconazole resistance is particularly notorious in relapse cases [28]. One entire class of antifungal drugs is not effective against

Cryptococcus, the echinocandins. This class of drugs is not used clinically to treat cryptococcosis due to intrinsic resistance, despite in vitro susceptibility [319], which is attributed to an inability of the drug to penetrate the host brain. Another possible cause of echinocandin resistance in *Cryptococcus* species is the fungal cell wall melanization that occurs in the brain environment, through the action of a fungal laccase, that uses L-DOPA and dopamine found in the human brain as precursors [320]. Melanin is an important virulence factor in *C. neoformans* since it can neutralize oxidative stress radicals protecting cells against host oxidative attack [321]. Melanin can also neutralize some toxic compounds, including some antifungal drugs, such as caspofungin and amphotericin B [322,323].

*C. neoformans* is widely spread in the environment, with worldwide distribution, mainly attributed to the particles in bird guano found in soil and trees that are then inhaled by humans and other mammals [324]. This pathogen is characterized by their high resistance to harsh environments in nature and in mammalian hosts [325], and after entering the host it can stay in a dormant latent granulomatous form for long periods of time. However, their preference for causing central nervous system infections in response to immunosuppression is not yet fully understood [314,324]. Despite being a public health threat and a WHO priority pathogen [326], *C. neoformans* still has many aspects of its peculiar metabolism associated with the central nervous system and interactions with the host that remain poorly understood [327].

Cryptococcosis is caused by three *Cryptococcus* species/variants, *C. neoformans* var. *grubii* (serotype A), responsible for 95% of *Cryptococcus* infections worldwide [328]; *C. neoformans* var. *neoformans* (serotype D) and *Cryptococcus gattii* (serotypes B and C) geographically restricted to tropical and/or subtropical regions [324]. In this work, iRV890 the first reconstructed GSMM for the human pathogen *C. neoformans* var. *grubii*, is presented. The model is provided in the widely used

SBML format. Model validation was conducted using experimental data for nitrogen and carbon assimilation from phenotypic arrays covering 222 different sources [329].

Specific growth and glucose consumption rates were experimentally determined in order to quantitatively validate the model's predictive power. A set of essential genes derived from the validated model is predicted and discussed in terms of their potential as novel antifungal drug targets. An additional comparison, with GSMM's for other pathogenic yeast species and *S. cerevisiae* was performed regarding the gene essentiality prediction and unique metabolic features of *C. neoformans*. Some peculiar characteristics and pathways of this fungus relevant to its pathogenicity are also discussed based on our findings. The iRV890 model provides a promising platform for global elucidation of the metabolic features of *C. neoformans var. grubii*, with an expected impact in guiding the identification of new drug targets and understanding the complex metabolism of this pathogen in the context of the human brain.

### **II.4.3. Materials and Methods**

#### **II.4.3.1. Model Development**

The genome-scale metabolic model of *C. neoformans var. grubii* H99, designated as iRV890, was reconstructed using *merlin* 4.0.5 [122] following the methodology described elsewhere [221] and OptFlux 3.0 [128], for curation and subsequent validation stages. All computational analyses were executed utilizing the IBM CPLEX 12.10 solver.

### II.4.3.2. Genome Annotation and Assembling of the Metabolic Network

The genome sequence of the *C. neoformans var. grubii* was retrieved from NCBI's Assembly database, with the accession number ASM1180120v1 [222] and the Taxonomy ID 235443 from NCBI's Taxonomy database [223]. The genome-wide functional annotation was based on taxonomy and frequency of similar sequences through remote DIAMOND alignment [330] and similarity searches using the UniProtKB/Swiss-Prot database. Draft network assembly relied on protein-reaction associations available in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database [280], with all reactions categorized as spontaneous or non-enzymatic also incorporated in the initial draft model. Hit selection was performed as described elsewhere [221] and phylogenetic proximity was implemented based on phylogenetic tree from the literature [122], this process is automated via the "Automatic workflow" *merlin* tool and then integrated into the draft model [122].

### II.4.3.3. Reversibility, Directionality and Balancing

Reaction reversibility and stoichiometry curation involved a multi-step process combining both automated and manual efforts. Initially, automatic curation was conducted by *merlin*, utilizing references from remote databases like eQuilibrator [281] to predict reaction directionality, as described by [221]. This was followed by extensive manual curation, exploiting databases such as MetaCyc [231], Brenda [230], UniProt [331], FungiDB [332], RHEA [333], KEGG [231] and existing literature, in order to ensure that all reactions in the network are balanced, and with the correct directionality. All manually edited reactions can be found in Supplementary Data II.4.1.

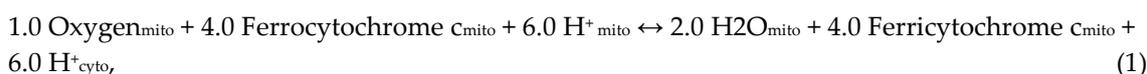
#### II.4.3.4. Compartmentalization and Transport reactions

This model includes four compartments: extracellular, cytoplasm, mitochondrion, and peroxisome, and one intercompartment, the cytoplasmic membrane. The prediction of compartments for each enzyme was performed using the DeepLoc - 2.0 [334] and directly imported to *merlin*. The transport reactions were automatically generated by the *merlin* integrated tool TranSyT [127] based on the public database TCDB [233]. Additional transport reactions across internal and external membranes for common metabolites, such as H<sub>2</sub>O, CO<sub>2</sub>, and NH<sub>3</sub>, often carried out without a transporter, were added to the model with no gene association.

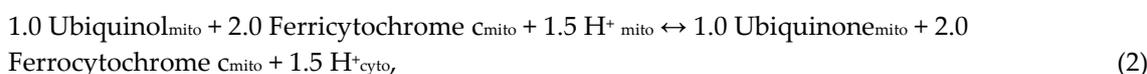
#### II.4.3.5. Biomass Equation

The biomass formation was depicted through an equation including proteins, DNA, RNA, lipids, carbohydrates, and cofactors, with detailed composition information for each macromolecule sourced from literature or experimental data. All calculations were performed as in previously described methodology [39] and are detailed in Supplementary Data II.4.1. ATP requirements for biomass production and growth-associated maintenance (GAM) were added to the biomass equation with a value of 25.65 mmol ATP/gDCW, based on the ATP requirements for the biosynthesis of cell polymers as reported in [236], and ATP requirements for non-growth-associated maintenance (NGAM) was inserted in the model by an equation with specific fixed flux boundaries inferred from *Candida tropicalis* [236]. The theoretical phosphorus to-oxygen ratio used in the *Saccharomyces cerevisiae* iMM904 metabolic model was applied to our model adding three generic reactions contributing to this ratio:

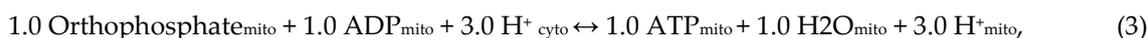
### Reaction R00081:



### Reaction R\_Ubiquinol\_Cytochrome\_Reductase:



### Reaction T\_ATP\_Synthase:



### The final balance reaction:



## II.4.3.7 Network simulation and model curation

During the model reconstruction process, extensive manual curation was needed in order to correct gaps in some pathways, due to incorrect reversibility; incomplete reactions; annotation errors; and blocked metabolites. Each case was meticulously inspected and studied, and reactions were edited, manually added to, or removed from the model based on literature evidence or databases such as KEGG pathways, MetaCyc, FungiDB, etc. The detailed list of all the alterations performed can be found in Supplementary Data II.4.1.

During this process, the *merlin* tool “*Find blocked reactions*” was used to assist and accelerate the process. Additionally, the *BioISO* tool based on COBRA and FBA [284] frameworks, also integrated in *merlin*, assisted in the process of identifying potential errors in the network and accelerated the process of correcting the gaps.

## **II.4.3.8 Model Validation**

### **II.4.3.8.1. Strains and Growth Media**

*C. neoformans var. grubii* H99 strain, obtained via the Fungal Genomic Stock Center, was routinely maintained in Yeast extract–Peptone–Dextrose (YPD) and synthetic minimal media (SMM) was used for batch cultured experiments. YPD contained: 20 g/L glucose (Merck, Darmstadt, Germany), 20 g/L peptone (Merck, Darmstadt, Germany), and 10 g/L yeast extract (Merck, Darmstadt, Germany), while SMM included: 20g/L glucose (Merck, Darmstadt, Germany), 2.7 g/L ammonium sulphate (Merck, Darmstadt, Germany), 0.05 g/L magnesium sulphate (Riddle-de-Haen), 2 g/L potassium dihydrogen phosphate (Panreac, Barcelona, Spain), 0.5 g/L calcium chloride (Merck, Darmstadt, Germany), and 100 µg/L biotin (Sigma).

### **II.4.3.8.2. Aerobic Batch Cultivation**

*C. neoformans var. grubii* H99 cells were batch cultivated in Erlenmeyer flasks containing 250ml of SMM medium, at 30 °C (250 rpm). Exponential phase inocula, with an Optical Density (OD) (Hitachi u2001) at 600nm of 0.3, were prepared and cells were transferred to Erlenmeyer flasks containing 250ml of fresh SMM medium and cultivated at 30 °C with orbital agitation (250 rpm) for 10 h.

### **II.4.3.8.3. Cell Density, Dry Weight, and Metabolite Concentration Assessment**

Throughout cell cultivation in SMM, 4 mL samples were collected every two hours for subsequent quantification of biomass and extracellular metabolites.

Cell density was monitored by measuring OD<sub>600nm</sub>. For dry weight determination, culture samples were centrifuged at 13,000 rpm for 3 minutes, and the resulting pellets were freeze-dried for 72 hours at -80 °C before being weighed. Extracellular metabolites, including glucose, ethanol, glycerol, and acetic acid, were identified and quantified by High-performance liquid chromatography (HPLC) on an Aminex HPX-87 H Ion Exchange chromatography column, eluted with 0.0005 M H<sub>2</sub>SO<sub>4</sub> at a flow rate of 0.6 mL/min at room temperature. Samples were analyzed in triplicate, and concentrations were determined using appropriate calibration curves. During the exponential growth phase, the specific growth rate, specific glucose consumption rate, and specific production rates of ethanol, glycerol, and acetic acid were calculated as described elsewhere [285].

#### **II.4.3.8.4. Network simulation and analysis**

All the phenotype simulations were performed with FBA in OptFlux 3.0 using the IBM®CPLEX solver, including: gene and reaction essentiality; growth assessment; metabolite production and consumption; and carbon and nitrogen source utilization. For gene and reaction essentiality, *in silico* growth was simulated in environmental conditions mimicking RPMI medium (described above), and a biomass flux lower than 5% of the wild-type strain, after the respective gene/reaction knockout, was considered the threshold for essentiality. Gene and reaction knockout was simulated by restraining its corresponding flux bounds to zero.

## II.4.4. Results and Discussion

### II.4.4.1. Model characteristics

The *C. neoformans var. grubii* genome-scale metabolic model constructed herein, and denominated iRV890, comprises 890 genes associated with 2598 reactions, of which 683 are transport reactions, and 2047 metabolites across four compartments (extracellular, cytoplasm, mitochondria, and peroxisome). The model can be found in SBML format in Supplementary Data II.4.2. Among the 2598 reactions, 1747 are cytoplasmic, 351 mitochondrial, 60 peroxisomal, and 440 are drains (exchange constraints for external metabolites used to simulate the environmental conditions allow metabolites to be removed from the system).

During the manual curation process, a total of 639 reactions/genes required alterations, including 80 that were mass balanced, 518 that were corrected for reversibility, directionality, or added or removed from the model, and 41 whose annotation was corrected, as detailed in Supplementary Data II.4.1.

The Biomass equation (Table II.14) encompasses the cell's major components along with their respective and relative contributions, including DNA, RNA, lipids, carbohydrates, and cofactors. The equation's composition in carbohydrates [335], and lipids [336–338] was inferred from literature data for *C. neoformans*. The composition of Proteins, DNA, and RNA was determined by the e-BiomassX where the whole genome sequence was used to estimate the amount of each deoxyribonucleotide as described in [123] mRNA, rRNA, and tRNA being used to estimate the total RNA in the cell as described in [111,123].

**Table II.14:** Biomass composition used in the model iRV890. The full individual validated contributions of each of these metabolites are shown in Supplementary Data II.4.1.

<b>Metabolite</b>	<b>g/gDCW</b>	<b>Metabolite</b>	<b>g/gDCW</b>
<b>Lipids</b>		<b>Proteins</b>	
Lanosterol	0,000122	L-Valine	0,021577
Zymosterol	0,000254	L-Tyrosine	0,015049
Squalene	0,000209	L-Tryptophan	0,009327
Ergosterol	0,000724	L-Threonine	0,021232
Phosphatidylserine	0,005024	L-Serine	0,028926
Phosphatidylinositol	0,004638	L-Proline	0,023588
Phosphatidylcholine	0,031241	L-Phenylalanine	0,018311
Phosphatidylethanolamine	0,017714	L-Methionine	0,010414
Cardiolipin	0,002254	L-Lysine	0,024010
Phosphatidic acid	0,000644	L-Leucine	0,036460
Phosphatidylglycerol	0,000644	L-Isoleucine	0,019954
Tetradecanoic acid	0,000020	L-Histidine	0,011797
Hexadecanoic acid	0,000097	L-Glutamate	0,030725
Octadecanoic acid	0,000038	L-Cysteine	0,003906
Dodecanoic acid	0,000021	L-Aspartate	0,022303
Decanoic acid	0,000011	L-Asparagine	0,014317
Octanoic acid	0,000038	L-Arginine	0,033954
Octadecanoic acid	0,000038	L-Alanine	0,021391
(9Z)-Octadecenoic acid	0,000093	Glycine	0,014693
(9Z,12Z)-Octadecadienoic acid	0,000116	L-Glutamine	0,018063
(9Z,12Z,15Z)-Octadecatrienoic acid	0,000002		
Triacylglycerol	0,032969	<b>Soluble Pool</b>	
Sterol esters	0,001127	Pyridoxine 5'-phosphate	0,000833
		FAD	0,000833
		Thiamine(1+) diphosphate	0,000833
		NAD	0,000833
		Glutathione	0,000833
		Riboflavin	0,000833
		Eumelanin	0,000833
		Ubiquinone-6	0,000833
		NADP	0,000833
		COA	0,000833
		FMN	0,000833
		5-Methyltetrahydrofolate	0,000833
<b>Carbohydrates</b>			
Chitin	0,005645		
Mannan	0,033956		
$\beta$ (1,3)-Glucan	0,360399		
<b>Ribonucleotides</b>			
UTP	0,006713		
GTP	0,006806		
CTP	0,005381		
ATP	0,007101		
<b>Deoxyribonucleotides</b>			
dTTP	0,016718		
dGTP	0,017029		
dCTP	0,015059		
dATP	0,017193		

The translated genome sequence was used to calculate the amino acid composition using the percentage of each codon usage as described in [123]. Essential metabolites were included in the biomass composition to qualitatively account for the essentiality of their synthesis pathways [237,253]. The growth and non-growth ATP requirements were adopted from *S. cerevisiae* [254].

#### **II.4.4.2. Model validation**

##### **II.4.4.2.1 Carbon and nitrogen source utilization**

*In silico* simulations were conducted using 222 different compounds as the exclusive carbon or nitrogen sources, under environmental conditions mimicking those of the minimal medium reported in [329]. The *in silico* growth was then compared to public available data on experimental growth data from commercially available phenotypic microarrays (Biolog) for *C. neoformans var. grubii* performed in [329]. For the analyses we used the data from STAT (stationary phase yeasts) condition after calculating the difference from the respective negative control group. A total of 155 sole carbon sources and 67 sole nitrogen sources were evaluated. iRV890 model correctly predicted growth in 85% (133/155) of the carbon sources tested and in 85% (57/67) of the nitrogen sources Supplementary Table II.4.1. We tried to analyze and understand the reason behind the cases that failed the prediction. In some cases, mainly associated with amino acids, L-ornithine and glycerol (carbon source) and amino acids and D-Glucosamine (nitrogen source) the model seems to have all the necessary tools to assimilate the compounds as carbon/nitrogen sources, however, the experimental evidence states no growth. In those cases, the failed prediction may be related to regulatory information that is not taken into account

in this model simulations, or misinformation regarding the transporter annotation which is still a big challenge in the current model development process. In many cases, however, the prediction failed because specific enzymes are not yet characterized for *C. neoformans*. The comparison between the model's prediction and experimental evidence suggests that the following enzyme activities are likely to be present in *C. neoformans*, although the underlying genes and proteins were not yet identified: 1.2.1.3 (aldehyde dehydrogenase), 1.1.1.21 (aldose reductase), 3.1.1.65 (L-rhamnono-1,4-lactonase), 1.1.1.56 (ribitol 2-dehydrogenase), 5.1.3.30 (D-psicose 3-epimerase), 2.7.1.55 (allose kinase), 4.1.2.10 ((R)-mandelonitrile lyase), 5.3.1.3 (D-arabinose isomerase), 3.2.1.86 (6-phospho-beta-glucosidase), 4.1.2.4 (deoxyribose-phosphate aldolase), 3.2.1.86 (6-phospho-beta-glucosidase) and 1.1.1.16 (galactitol 2-dehydrogenase). The identification and characterization of these predicted functions and their underlying gene(s) will shed light on the specific pathways of carbon or nitrogen assimilation in this pathogen, potentially revealing new mechanisms of virulence related to adaptation to the host environment and contributing to understanding its neurotropism. Altogether, the model achieved 85% predictability which is a high value, especially considering that the extensive list of carbon and nitrogen sources tested includes many that are not commonly used in traditional metabolic and phenotypic experiments and thus lack biochemical characterization. The specific example of the inositol assimilation pathway is paradigmatic in this sense, as explored later on.

#### **II.4.4.2.2 Growth parameters in batch culture**

To quantitatively validate the model, the specific growth rate, glucose consumption rate, and metabolite production rates were experimentally

determined, and compared with *in silico* predicted values. For a glucose consumption rate of 1.72 mmol.gDCW<sup>-1</sup>.h<sup>-1</sup>, a specific growth rate of 0.188 h<sup>-1</sup> was experimentally determined, leading to no detectable production of ethanol, glycerol, or acetate. For comparison with *in silico* results, we simulated the system's behavior in SMM medium with a fixed glucose uptake flux of 1.72 mmol.g<sup>-1</sup> dry weight.h<sup>-1</sup>. Other nutrient fluxes were left unconstrained, as the system was glucose-limited under these conditions. The simulation predicted a specific growth rate of 0.124 h<sup>-1</sup>, a difference of 0.06 h<sup>-1</sup> to the experimentally determined value (Table II.15). In these conditions, the model did not predict the formation of glycerol, acetic acid, or ethanol as by-products, consistent with the experimental data. Moreover, the model does not predict the growth of *C. neoformans* under anaerobic conditions which is expected since this pathogen is an obligate aerobic fungus.

**Table II.15** - Growth parameter values predicted by the iRV890 model, in comparison with those determined experimentally.

	Specific growth rate (h <sup>-1</sup> )	q (mmol g <sup>-1</sup> dry weight h <sup>-1</sup> )			
		Glucose	Ethanol	Glycerol	Acetic acid
<i>In silico</i>	0.124	1.72	0	0	0
<i>In vivo</i>	0.188	1.72	0	0	0

#### II.4.4.3. *C. neoformans* unique metabolic features

To uncover unique metabolic features of this pathogen, a comparison was made between the *C. neoformans* GSMM with those previously built for *C. glabrata* [237], *C. albicans* [288], *C. auris* [339] and *S. cerevisiae* [238]. A comparison across the existing models was carried out based on shared EC numbers. After intersecting the EC numbers present in each of the five models, 83% (470/566) of the proteins

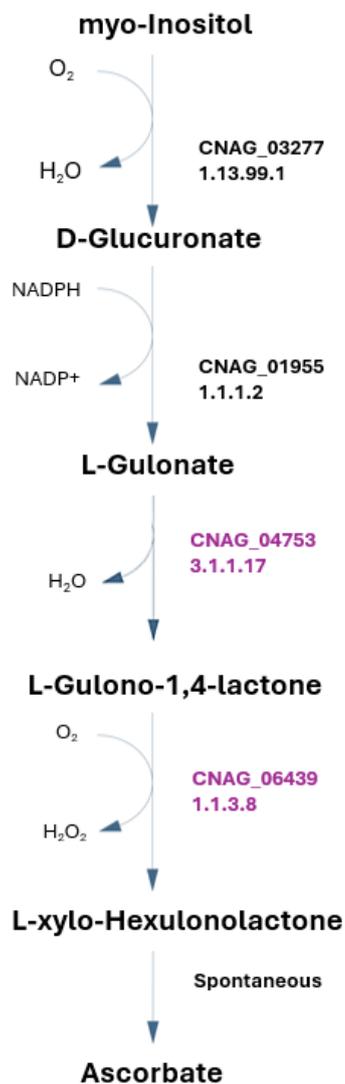


From the list of 96 *C. neoformans* unique EC numbers (Supplementary Data II.4.1), we searched for metabolic features or pathways relevant in the context of fungal infection in the host brain, either by constituting defense mechanisms or enabling host adaptation, through degradation or biosynthesis of specific metabolites. A few of these unique EC numbers with higher potential of impacting *C. neoformans* pathogenesis are discussed below:

**1.1.1.12 and 1.1.1.287-** L-arabinitol 4-dehydrogenase and D-arabinitol dehydrogenase are two enzymes that are required for L-arabinitol assimilation as carbon source. Neither *Candida* species [288,303] nor *S. cerevisiae* [341], unless genetically engineered [342], can assimilate L-arabitol, which is a particular metabolic feature of *C. neoformans* when compared to other yeast species (Supplementary Table II.4.1). In fact, it was shown that the deletion of the transporter for arabitol and other polyols, Ptp1 (polyol transporter protein 1) was associated with patient survival [343], while PTP1 expression is highly induced in macrophage and amoeba infection [343]. Since, arabitol is present in the cerebrospinal fluid, or is a known product of other human pathogens, for example, *Candida* species [344] this underscores the metabolic flexibility of *C. neoformans*.

**1.1.3.8 and 3.1.1.17 -** L-gulonolactone oxidase and gluconolactonase are two enzymes that participate in ascorbate metabolism, allowing the utilization of Inositol and D-glucuronate as source for L-ascorbate biosynthesis (Figure II.10). Interestingly, it was reported by two independent studies that the presence of ascorbate, an antioxidant, lowers the susceptibility towards fluconazole in *C. neoformans* [345,346]. However, this effect seems to not be related to its antioxidant potential but to ascorbate-induced up-regulation of Upc2, a transcriptional regulator of genes involved in ergosterol biosynthesis, as shown in *C. albicans* [347]. The ability of *C. neoformans* to synthesize ascorbate from

inositol is particularly noteworthy, given the abundance of inositol in the human brain [327] and the widespread use of fluconazole in treating infections. Having a mechanism to produce a compound that mitigates the toxicity of fluconazole represents a significant adaptive advantage for this species.

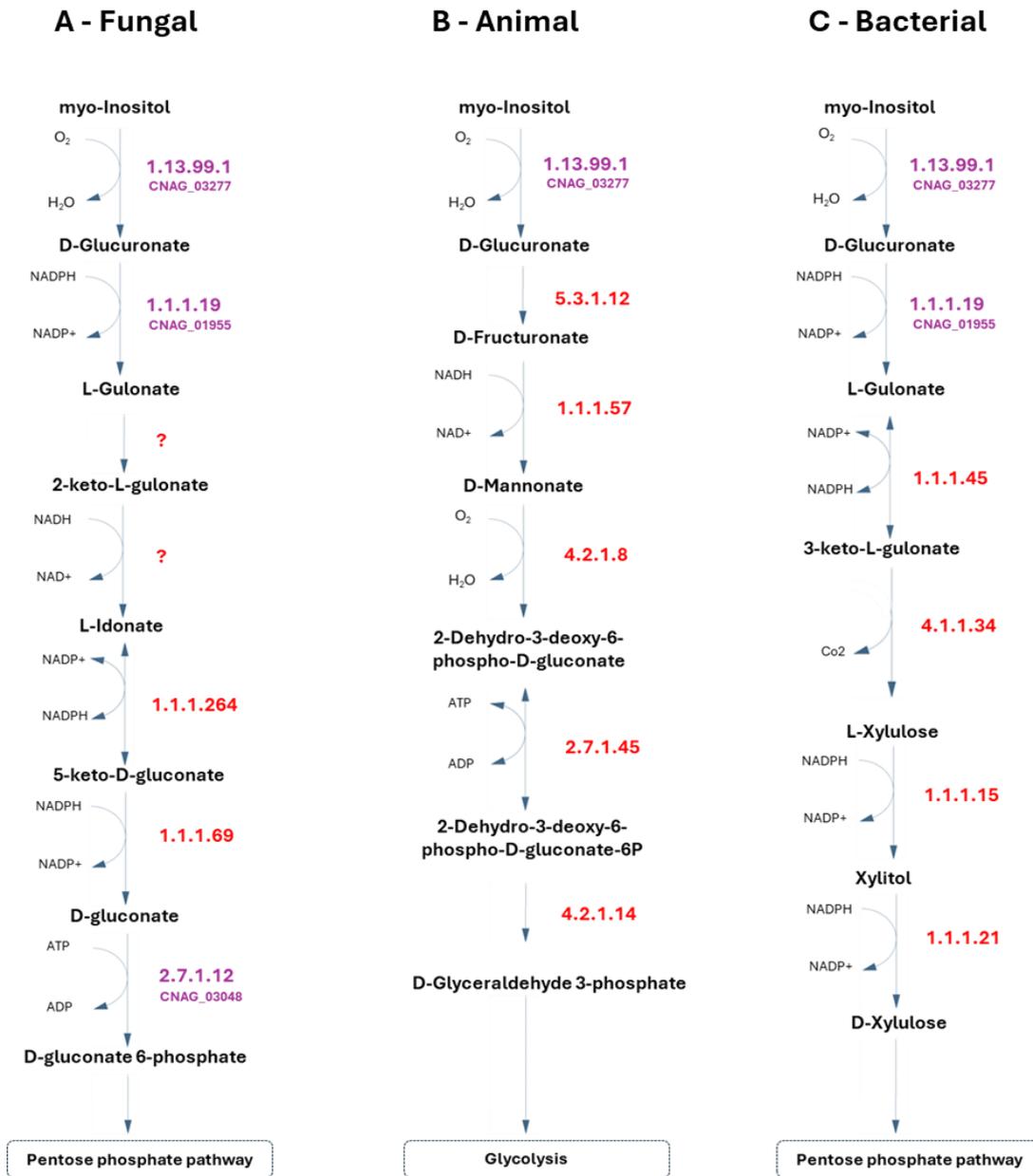


**Figure II.10** – *C. neoformans* pathway for ascorbate biosynthesis, with the respective *C. neoformans* var. *grubii* EC numbers present in the iRV890 model. The 1.1.3.8 and 3.1.1.17 enzymes, which are unique to *C. neoformans* among other pathogenic yeasts, are highlighted in purple.

Additionally, the 1.1.3.8 and 3.1.1.17 enzymes are also important for inositol assimilation as carbon source through a variation of the previous pathway (Figure II.11). This pathway was suggested recently as an alternative pathway in fungi for inositol assimilation, and since inositol is highly abundant in the human brain, this may represent a very important metabolic feature for *C. neoformans*, as

an alternative carbon source in that environment, where some nutrients may be scarce. In fact, in order to implement that pathway in the model, two of the reactions reported were recreated and attributed with the names R2\_Inositol\_Pathway and R1\_Inositol\_Pathway in the model, although the corresponding EC numbers and genes have not been identified in the annotated *C. neoformans* genome [348]. This pathway was recreated exclusively from literature, and while it lacks more study, we suggest the gene CNAG\_02553 is a probable candidate for encoding the 1.1.1.69 enzyme, predicted by OrthoMCL [349] as an ortholog of other genes with the same function [349]. Additional pathways for inositol assimilation are reported for animals (Figure III.11.B) and bacteria (Figure III.11.C), however, since *C. neoformans* lacks almost all the enzymes in those pathways, we considered that the new pathway reported in fungi [348] was the most probable to occur in this pathogen.

**1.1.1.377** - L-rhamnose 1-dehydrogenase is an enzyme required for L-rhamnose assimilation as carbon source. *Candida* species [288,303,339] and *S. cerevisiae* (unless engineered) [350] cannot assimilate L-rhamnose, which is a particular metabolic feature of *C. neoformans* when compared to these yeast species (Supplementary Table II.4.1). Rhamnose is used by some pathogens, for example *Pseudomonas aeruginosa*, to produce rhamnolipids, and constitutes an important virulence factor in those bacteria, with roles in biofilm formation, hydrophobic nutrient uptake, and host immunity evasion, characterized for increasing lung epithelial permeability [351,352] and inhibition of macrophage phagocytosis [353]. Rhamnose utilization may represent an important adaptative advantage if *C. neoformans*, whose primary route of infection is, like *P. aeruginosa*, the respiratory system.



**Figure II.11** - Metabolic pathways for inositol assimilation as carbon source, A - based on the proposed fungal inositol assimilation pathway reported in Kuivanen et al. 2016 [348]. B – based on the animal inositol assimilation pathway, and C- based on the bacterial inositol assimilation pathway. The respective *C. neoformans var. grubii* genes present in iRV890 are highlighted in purple. The

currently unknown genes are highlighted in red and the proposed reactions with an unknown EC number are represented as a question mark in red.

**1.14.13.231** - tetracycline 11a-monooxygenase is an enzyme that allows the direct conversion of tetracycline into 11a-hydroxytetracycline, and it was reported to confer resistance to all clinically relevant tetracyclines, by efficient degradation of a broad range of tetracycline analogues but also conferring resistance to these antibiotics *in vivo*. The hydroxylated product, 11a-hydroxytetracycline, is very unstable and leads to intramolecular cyclization and non-enzymic breakdown to undefined products, completely neutralizing the tetracycline effects [354,355]. Although tetracyclines are generally used as antibacterial antibiotics and have poor antifungal activity, the presence of this enzyme in *C. neoformans* should be taken into consideration when designing tetracyclines against fungi.

**3.1.3.8** - 3-phytase is an enzyme involved in inositol metabolism that may be involved in the production of phytic acid from inositol, a primary storage molecule of phosphorus and inositol. It was previously reported that mutation of this enzyme reduces the utilization of alternative carbon sources and confers attenuated virulence of *C. neoformans* associated with failed dissemination into the brain [356].

**3.5.2.17** – hydroxyisourate hydrolase is an enzyme essential for the assimilation of uric acid as a nitrogen source. Uric acid is a normal component of urine, which is an interesting feature since this yeast is frequently isolated from bird guano, which is rich in uric acid. Additionally, uric acid enhances the production of key cryptococcal virulence factors, including capsule and urease, an enzyme required for full fitness at mammalian pH and dissemination to the brain [357]. *C. neoformans* capsule is induced in the presence of uric acid, which is a metabolite also found at high concentrations in the human blood [358,359].

**4.1.1.105** – L-tryptophan decarboxylase catalyzes the conversion of L-tryptophan into Tryptamine, which can then be converted into serotonin, and shares structure with several aminergic neuromodulators. However, the reaction is bidirectional, and Tryptamine can also be converted into L-tryptophan. While it is unclear which may be the role of this enzyme in *C. neoformans*, it may be related to the brain environment, specifically in the utilization of serotonin as nitrogen source through its conversion into L-tryptophan.

**4.1.1.28, 1.14.18.1, and 1.10.3.2** – DOPA decarboxylase, tyrosinase, and laccase are particularly important in *C. neoformans*, as they are involved in the biosynthesis of melanin, an important virulence factor in *C. neoformans*. Melanin is able to neutralize oxidative stress radicals as well as protecting the pathogen against the host immune system and antifungal drugs, such as caspofungin and amphotericin B. L-DOPA and Dopamine are present in the human brain and serve as precursors for dopamine biosynthesis in this pathogen.

#### **II.4.4.4. Drug target analysis based on gene essentiality prediction**

Pathogen's GSMM are particularly useful to identify potential new drug targets, among predicted essential genes. For that purpose, a list of all predicted essential genes and enzymes in *C. neoformans* was obtained through simulation of the system's behaviour in RPMI medium, which mimics the environmental conditions of human serum. A total of 157 enzymes and 101 genes were identified as essential in RPMI medium. Among these targets, some have been previously identified as essential genes in other pathogenic yeasts (see Table 4), indicating potential drug targets common to all *Candida* species and *C. neoformans*. Notably, Erg11 and Fks1 are already targets of currently used antifungals, fluconazole, and echinocandins, respectively. Additionally, Erg26, Erg27, Erg24, Erg4, Erg7,

Erg12, and Erg13 have all been identified herein as potential drug targets within the ergosterol biosynthetic pathway. Particularly interesting is Erg4, as it lacks a human ortholog, suggesting an opportunity for designing compounds with enhanced selectivity and lower toxicity. Similarly to *Candida*, which lacks a Folate transporter [360] and relies on its *de novo* biosynthesis, *C. neoformans* seems to also not have a cytoplasmic folate transporter, leading to the identification of Fol1 as a promising multi-yeast drug target. Furthermore, Fas1, a fatty acid synthase enzyme, and Chs1, a chitin synthase, also lack human orthologs and constitute promising alternative antifungal drug targets due to their important role in membrane and cell wall structure and integrity. Other noteworthy targets span various pathways, including purine metabolism, terpenoid backbone biosynthesis, pyrimidine metabolism, CoA biosynthesis, glycerophospholipid biosynthesis, and ubiquinone biosynthesis (Table II.16). However, exploring these targets requires leveraging potential structural differences in the enzyme's active site compared to their human counterparts.

Since *C. neoformans* colonizes a different host environment and is phylogenetically distant from *Candida* spp. we extended our evaluation to include potential new drug targets that are unique to this species, and not shared by any of the *Candida* spp.. We identified only two such targets: the 1.14.18.1 tyrosinase, encoded by the gene CNAG\_03009, and the 2.5.1.83 hexaprenyl diphosphate synthase, encoded by the gene CNAG\_04375. While tyrosinase, responsible for melanin production, has a human ortholog (since humans also synthesize melanin), hexaprenyl diphosphate synthase (2.5.1.83) is fungal-specific and may represent an interesting target. This enzyme plays a crucial role in terpenoid backbone biosynthesis, serving as a key contributor to the synthesis of precursors for ubiquinone biosynthesis.

**Table II.16** - Enzymes predicted to be essential in RPMI medium in 5 pathogenic fungal species, based on the screening of the genome-scale metabolic models of *C. neoformans* iRV890, *C. auris* IRV973, *C. parapsilosis* iDC1003, *C. albicans* iRV781, and *C. glabrata* iNX804. Grey rows highlight enzymes that are not encoded in the human genome. Data regarding the drug association was retrieved from the DrugBank database; only drugs with known pharmacological action against pathogens were selected.

<i>C. neoformans</i>	<i>C. albicans</i>	<i>C. glabrata</i>	<i>C. parapsilosis</i>	<i>C. auris</i>	<i>S. cerevisiae</i>	Human	Pharmacological action	EC Number	Pathway/Target
CNAG_04605	ERG26	CAGLOG00594g	CPAR2_302110	CJI97_000938	ERG26	NSDHL		1.1.1.170	Steroid
CNAG_00441	IMH3	CAGL0K10780g	CPAR2_104580	CJI97_000080	IMD4	IMPDPH	-	1.1.1.205	Purine
CNAG_07437	ERG27	CAGL0M11506g	CPAR2_801560	CJI97_004310	ERG27	DHRS11	-	1.1.1.270	Steroid
CNAG_06534	HMG1	CAGL0L11506g	CPAR2_110330	CJI97_003299	HMG1	HMGCR	-	1.1.1.34	Terpenoid backbone
CNAG_00117	ERG24	CAGL0I02970g	CPAR2_405900	CJI97_003097	ERG24	TM7SF2	-	1.3.1.70	Steroid
CNAG_02830	ERG4	ERG4	ERG4	CJI97_002908	ERG4	-	-	1.3.1.71	Steroid
CNAG_04692	CDC21	CDC21	CPAR2_206550	CJI97_005101	TMP1	TYMS	-	2.1.1.45	Pyrimidine
CNAG_00700	ADE17	CAGL0A03366g	CPAR2_202250	CJI97_002511	ADE17	ATIC	-	2.1.2.3	Purine
CNAG_07373	URA2	CAGL0L05676g	CPAR2_203160	CJI97_002269	URA2	CAD	-	2.1.3.2	Pyrimidine
CNAG_06508	GSC1	FKS1	CPAR2_106400	FKS1	FKS1	-	Echinocandins	2.4.1.34	1,3-beta-glucan
CNAG_03196	URA5	URA5	CPAR2_802790	CJI97_002422	URA5	UMPS	-	2.4.2.10	Pyrimidine
CNAG_02853	ADE4	CAGL0M13717g	CPAR2_208260	CJI97_001833	ADE4	PPAT	-	2.4.2.14	Purine
CNAG_02084	BTS1	CAGL0H05269g	CPAR2_302840	CJI97_003197	BTS1	GGPS1	-	2.5.1.1	Terpenoid backbone
CNAG_07780	ERG20	ERG20	CPAR2_103950	CJI97_001757	ERG20	FDPS	-	2.5.1.10	Terpenoid backbone
CNAG_02787	C5_05130C	CAGL0F05555g	CPAR2_502760	CJI97_003836	CAB5	COASY	-	2.7.1.24	CoA
CNAG_02976	CR_03740C	CAGL0K11022g	CPAR2_202590	CJI97_005311	FMN1	RFK	-	2.7.1.26	Riboflavin
CNAG_02866	C6_02980C	CAGL0H01551g	CPAR2_602050	CJI97_004586	CAB1	PANK	-	2.7.1.33	CoA

CNAG_05935	URA6	CAGL0L09867g	CPAR2_105320	CJI97_000033	URA6	CMPK2	-	2.7.4.14	Pyrimidine
CNAG_06001	ERG8	ERG8	CPAR2_400710	CJI97_001215	ERG8	PMVK	-	2.7.4.2	Terpenoid backbone
CNAG_03335	C5_00260W	CAGL0D00550g	CPAR2_304260	CJI97_000019	PRS1	PRPS1	-	2.7.6.1	Purine
CNAG_05384	C4_05210W	CAGL0G03157g	CPAR2_500260	CJI97_005306	PIS1	CDIPT	-	2.7.8.11	Glycerophospholipid
CNAG_02795	ADE8	CAGL0F02761g	CPAR2_211620	CJI97_002826	ADE8	GART	-	2.1.2.2	Purine
CNAG_02609	COQ3	CAGL0I07601g	CPAR2_602300	CJI97_005452	COQ3	COQ3	-	2.1.1.114	Ubiquinone
CNAG_00138	COQ5	CAGL0J06710g	CPAR2_209250	CJI97_003704	COQ5	COQ5	-	2.1.1.201	Ubiquinone
CNAG_00040	ERG11	ERG11	ERG11	ERG11	ERG11	CYP51A1	Azoles	1.14.14.154	Steroid
CNAG_02844	PEL1	PGS1	CPAR2_805350	CJI97_000224	PEL1	PGS1	-	2.7.8.5	Glycerophospholipid
CNAG_02878	C6_01340C	CAGL0H04389g	CPAR2_602700	CJI97_005490	GEP4	PTPMT1	-	3.1.3.27	Glycerophospholipid
CNAG_00734	URA4	CAGL0J04598g	CPAR2_100500	CJI97_002941	URA4	CAD	-	3.5.2.3	Pyrimidine
CNAG_02294	ADE2	ADE2	CPAR2_805940	CJI97_004071	ADE2	PAICS	-	4.1.1.21	Purine
CNAG_04961	URA3	URA3	URA3	CJI97_003384	URA3	UMPS	-	4.1.1.23	Pyrimidine
CNAG_02786	FOL1	CAGL0J07920g	CPAR2_303390	CJI971_001274	FOL1	-	Sulfacetamide	4.1.2.25	Folate biosynthesis
CNAG_02786	FOL1	CAGL0J07920g	CPAR2_303390	CJI971_001274	FOL1	-	Sulfonamides	2.5.1.15	Folate biosynthesis
CNAG_05125	MVD	CAGL0C03630g	CPAR2_109530	CJI97_001340	MVD1	MVD	-	4.1.1.33	Terpenoid backbone
CNAG_00909	CAB3	CAGL0L05302g	CPAR2_800750	CJI97_003563	CAB3	PPCDC	-	4.1.1.36	CoA
CNAG_03270	ADE13	CAGL0B02794g	CPAR2_204960	CJI97_000801	ADE13	ADSL	-	4.3.2.2	Purine
CNAG_00265	IDI1	CAGL0J06952g	CPAR2_401630	CJI97_001183	IDI1	IDI1	-	5.3.3.2	Terpenoid backbone
CNAG_01129	ERG7	CAGL0J10824g	CPAR2_301800	CJI97_005090	ERG7	LSS	Oxiconazole	5.4.99.7	Steroid
CNAG_00143	ADE1	CAGL0I04444g	CPAR2_500190	CJI97_003065	ADE1	PAICS	-	6.3.2.6	Purine
CNAG_06314	ADE5,7	CAGL0H07887g	CPAR2_208400	CJI97_001704	ADE5,7	GART	-	6.3.3.1	Purine
CNAG_06314	ADE5,7	CAGL0H07887g	CPAR2_208400	CJI97_001704	ADE5,7	GART	-	6.3.4.13	Purine
CNAG_04192	ADE6	CAGL0K04499g	CPAR2_204070	CJI97_002160	ADE6	PFAS	-	6.3.5.3	Purine
CNAG_05759	ACC1	CAGL0L10780g	CPAR2_804060	CJI97_001038	ACC1	ACACA	-	6.4.1.2	Fatty acid
CNAG_02686	ERG12	CAGL0F03861g	CPAR2_803530	CJI97_005606	ERG12	MVK	-	2.7.1.36	Terpenoid backbone

CNAG_03311	ERG13	ERG13	CPAR2_701400	CJI97_004952	ERG13	HMGCS	-	2.3.3.10	Terpenoid backbone
CNAG_02099	FAS1	CAGL0D00528g	FAS1	CJI97_001309	FAS1	-	-	2.3.1.86	Fatty acid
CNAG_01877	GUA1	CAGL0F03927g	CPAR2_803560	CJI97_005609	GUA1	GMPS	-	6.3.5.2	Pyrimidine
CNAG_03099	CHS1	CAGL0I04818g	CPAR2_805640	CHS2	CHS2	-	-	2.4.1.16	Chitin

---

## II.4.5. Conclusions

The construction and validation of iRV890, the first genome-scale metabolic model for *C. neoformans var. grubii* is presented herein. iRV890 constitutes a robust platform for exploring and elucidating the metabolic features of this poorly understood pathogen, particularly concerning its interaction within the central nervous system and the human host. By encompassing 890 genes associated with 1466 reactions, this model offers a comprehensive view of the metabolic landscape of the pathogen. Through *in silico* simulations, we predicted the use of more than 200 compounds as sole carbon or nitrogen sources, and after comparison to experimental data from phenotypic microarrays, we gained valuable insights into the metabolic capabilities of *C. neoformans*. The model correctly predicts 85% of the carbon and nitrogen sources tested. The model was able to predict accurately the organism's specific growth rate and confirmed its inability to grow under anaerobic conditions or to accumulate glycerol, acetic acid, or ethanol as metabolic by-products during growth in synthetic minimal medium, with glucose as carbon source. Additionally, we propose a list of yet unidentified enzymes expected to be present in *C. neoformans*, based on the carbon and nitrogen utilization and with potential to represent new host adaptation or virulence mechanisms.

Our investigation into the unique metabolic features of *C. neoformans* has unveiled several pathways and enzymatic activities that are proposed to play pivotal roles in fungal infection within the host brain. Some enzymes constitute important virulence factors, such as Tyrosinase and laccase, enzymes responsible for the production of melanin which has an important role in host immune evasion [321], infection proliferation, and drug resistance [322,323]. Other enzymes are related to drug and stress resistance, such as tetracycline 11a-monooxygenase, L-gulonolactone oxidase, and gluconolactonase. The remaining

enzymes are directly related to alternative carbon/nitrogen source utilization and are important for environmental adaptation. For example, hydroxyisourate hydrolase is essential for the assimilation of uric acid as a nitrogen source, an important virulence factor mechanism, and 3-phytase is involved in inositol metabolism and storage, important for brain dissemination.

In this work, we also propose several potential drug targets in *C. neoformans*. Notably, enzymes such as Erg4, Chs1, Fol1, and Fas1 present promising opportunities for targeted drug development, due to their absence in human cells, offering opportunities for the development of selective and low-toxicity compounds. The CNAG\_03009 and CNAG\_04375 genes, encoding a tyrosinase and a hexaprenyl diphosphate synthase, are presented as potential antifungal drug targets specific of *C. neoformans*.

Our model contributes to a better understanding of *C. neoformans* metabolism, especially within the host environment. With this work, we not only propose new metabolic enzymes awaiting characterization but also offer insights into key pathways and interactions shaping the dynamics between host and pathogen and its adaptive strategies. We also propose some potential antifungal targets, for *C. neoformans* and confirmed the coverage of already identified targets also to that species. These results hold promise for the discovery of novel drug targets and for the full comprehension of this pathogen's metabolic network with an expected impact in combating cryptococcosis.

### **III: Identification of Potential New Antifungal Drugs: from virtual drug screening to experimental validation.**

Journal Paper: Viana R, Souza CS, Bernardes N, Schang LM, Soares CM, Teixeira MC, "Identification of Potential New Antifungal Drugs: from virtual drug screening to experimental validation", manuscript in preparation.

### III.1. Abstract

Fungal infections pose a significant threat to human health, driven by increasing drug resistance and limited treatment options. The emergence of multidrug resistance further increases the urgency of discovering novel antifungal agents with enhanced efficacy and safety profiles.

In this work, a multidisciplinary approach integrating computational modeling, molecular docking, and experimental validation was used to identify and evaluate potential antifungal agents targeting essential enzymes in pathogenic fungi. Specifically, focus was given to finding inhibitors of chitin synthase (Chs2) or dihydropteroate synthase (Fol1), previously identified as potential drug targets. Through high-throughput *in silico* molecular docking, compound libraries were screened and among 317 and 20,000 tested molecules, 37 and 76 were identified as promising candidates as inhibitors of Fol1 and Chs2, respectively. Upon selection of some of the compounds exhibiting higher predicted affinity to the active site of Fol1 or Chs2, Homopteroic acid (C5) and 5-(perylene-2-ylethynyl)-arabino-uridine (2a), were identified as the most promising compounds, exhibiting moderate antifungal effects. C5 was found to be fungicidal against the tested *Candida* species with an MIC of 256 mg/L and exhibiting synergy with fluconazole against *C. albicans*. However, cytotoxicity against mammalian cell lines was observed for this antifolate compound. 2a, with no detected cytotoxicity, was found to exhibit fungistatic activity against *C. neoformans* at 400  $\mu$ M.

Altogether, the compounds identified in this study show potential as active molecules against fungal pathogens, although further optimization to increase their antifungal activity and decrease cytotoxicity is still required. Additionally, a vast list of new alternative compounds for antifungal therapy is provided,

laying the foundation for future research and development in the field of antifungal drug design.

**Key-words:** Molecular docking; Virtual drug screening; Antifungal agents; Drug discovery; Pathogenic fungi

### **III.2. Introduction**

Fungal infections pose a significant threat to human health, with morbidity and mortality rates on the rise due to increasing drug resistance and limited treatment options [3,4]. Addressing this urgent clinical need requires the discovery and development of novel antifungal agents with improved efficacy and safety profiles. Particularly concerning are infections caused by pathogenic fungi belonging to the *Aspergillus*, *Candida*, *Cryptococcus*, and *Pneumocystis* genera, which exhibit high mortality rates, especially in immunocompromised patients [3]. The emergence of multidrug resistance in these pathogens further exacerbates the urgency of the situation [11–15].

There is currently a demand for alternative and innovative approaches to accelerate the process of discovery of new drugs and deal with the rise of multidrug resistance in pathogenic fungi. Indeed, the traditional drug development process is hindered by its time-consuming, complex, and costly nature [132].

Molecular docking emerged as a promising computational technique in rational drug design and discovery, and currently, plays an important role in the discovery and development of new drugs [133]. Molecular docking enables the modeling of interactions between small molecules and proteins at the atomic level, allowing the prediction of ligand-receptor complex structures within the

binding site of their predicted target and the elucidation of fundamental biochemical processes [131,134]. This approach can be used to perform a virtual drug screening of a large database of compounds against a certain target, offering large-scale prediction of compounds with superior binding affinities as potential inhibitors. Successful applications of molecular docking in drug discovery underscore its potential and widespread utilization in modern drug development [192–194].

In previous work from our lab [288,303,339] we identified chitin synthase (EC number 2.4.1.16) and dihydropteroate synthase enzyme (EC: 2.5.1.15) as potential drug targets in pathogenic fungal species using a cross-species approach integrating enzyme essentiality within host prediction from 5 different genome-scale metabolic models (GSMM's) for *Candida* and *Cryptococcus* species.

In this work we used a multidisciplinary approach, integrating computational modeling, molecular docking, and experimental validation, to identify and evaluate potential new drugs targeting dihydropteroate synthase and chitin synthase in different *Candida* and *Cryptococcus* species. The integration of GSMMs and virtual drug screening is an innovative methodology with the potential to accelerate the development of novel antifungal agents. Our findings are anticipated to offer new alternative compounds for drug discovery, contributing to new advancements in addressing the challenges posed by fungal infections.

### III.3. Materials and Methods

#### III.3.1. Protein structure selection and modeling

For the chitin synthase enzyme Chs1, the structure of *Candida albicans* (*C. albicans*) co-crystalized with UDP-N-acetylglucosamine and deposited in Protein Data Bank (PDB) with the code 7STM was used. For the Dihydropteroate synthase enzyme, Fol1, there was no available structure for any *Candida* or *Cryptococcus* species in PDB, so we decided to model its structure by homology modeling using MODELLER version 9.6 [185]. The structure of *Saccharomyces cerevisiae* (*S. cerevisiae*) Fol1 (PDB code: 2BMB) was used as a template to model the *C. albicans* Fol1 protein, except for some small regions where the *S. cerevisiae* Fol1 structure exhibited low confidence. In those cases, the *Plasmodium vivax* Fol1 structure (PDB code: 5Z79) was used as a complementary template. Multiple amino acid sequence alignments were performed using Clustal Omega [361]. The modeling was performed with 2 ligands fixed in the active site (4-aminobenzoate, and (7,8-dihydropterin-6-yl)methyl diphosphate), and a magnesium cofactor. Thirty independent models were generated and ranked using the normalized Discrete Optimized Protein Energy (DOPE) score. Subsequent analyses were carried out on the model with the lowest score and with no visually detected clashes between non-bound atoms. The structures were visualized and analyzed using PyMOL 2.3 [362].

#### III.3.2. Molecular Docking simulations and Virtual drug screening

The docking calculations for the *C. albicans* Fol1 and Chs1 structures were performed using AutoDock Vina [144]. The receptor and ligands for each docking experiment were pre-treated and converted to PDBQT file format using

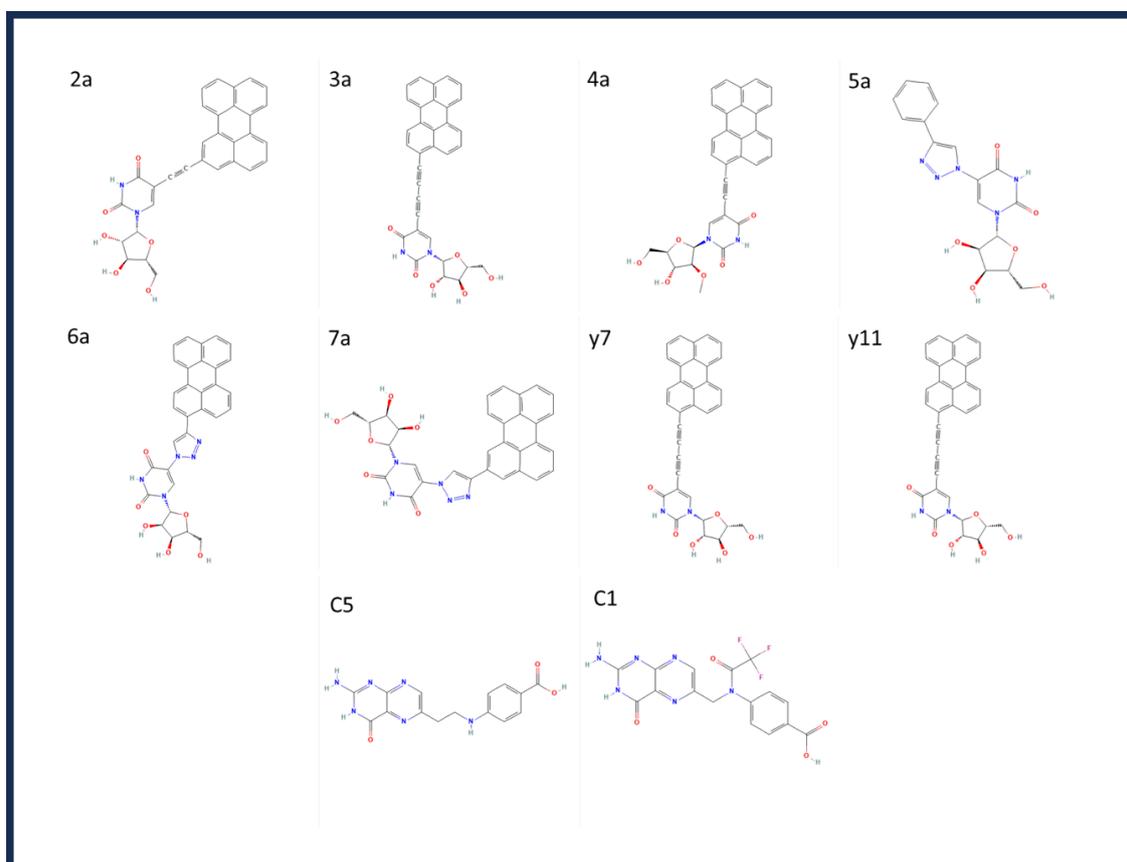
the AutoDockTools and used in AutoDock Vina to perform the simulations with an exhaustiveness value of 50 and a grid box centralized at the ligands covering the active site for each protein, with sizes 18,20,18 Å for Fol1 and 40, 40, 35 Å for Chs1. For each docking calculation, 30 solutions were generated and ordered by the binding energy. The docking results with the lowest binding energies were selected for visual inspection and further analysis.

For the virtual drug screening on the Fol1 protein active site, analogous compounds of the product of the enzymatic reaction, 7,8-dihydropteroate, were searched in the Zinc15 database using two different search approaches, SMARTS search and Tanimoto 30 search, resulting in 317 analogs selected. The selection was filtered by the availability of the compounds to be purchased. For the Virtual drug screening on Chs1 active site, N-Acetyl-D-Glucosamine (chitin monomer) analogous compounds were searched in the PubChem database also using two different search approaches, SMARTS search and Tanimoto 30 search, leading to the selection of 22,000 compounds. This search was not restricted by compound purchase availability.

### III.3.3. Compounds

The compounds 5-(perylene-2-ylethynyl)-arabino-uridine (2a), 5-(perylene-3-ylbutadiynyl)-arabino-uridine (3a), 5-(perylene-3-ylethynyl)-2'-O-methyl-uridine (4a), 5-(4-phenyl-1,2,3-triazol-1-yl)uridine (5a), 5-[4-(perylene-3-yl)-1,2,3-triazol-1-yl]uridine (6a), 5-[4-(perylene-2-yl)-1,2,3-triazol-1-yl]uridine (7a), 5-[3-(4-benzoylphenoxy)prop-1-ynyl]-1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl]pyrimidine-2,4-dione (y7) and 2'-Deoxy-5-[(perylene-3-yl)ethynyl]uridine (y11) were kindly provided by Luis M. Schang, Baker Institute for Animal Health, Cornell University, USA [363]. Homopteroic acid (compound C5) was purchased from Fluorochem, and N10-

(Trifluoroacetyl)pteroic acid (compound CH1) was purchased from Musechem (Figure III.1).



**Figure III.1** – Molecular 2D structure of all the compounds acquired and experimentally tested in this work. 2a - 5-(perylene-2-ylethynyl)-arabino-uridine. 3a - 5-(perylene-3-ylbutadiynyl)-arabino-uridine. 4a - 5-(perylene-3-ylethynyl)-2'-O-methyl-uridine. 5a - 5-(4-phenyl-1,2,3-triazol-1-yl)uridine. 6a - 5-[4-(perylene-3-yl)-1,2,3-triazol-1-yl]uridine. 7a - 5-[4-(perylene-2-yl)-1,2,3-triazol-1-yl]uridine. y7 - 5-[3-(4-benzoylphenoxy)prop-1-ynyl]-1-[(2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl]pyrimidine-2,4-dione. y11 - 2'-Deoxy-5-[(perylene-3-yl)ethynyl]uridine. C1 - N10-(Trifluoroacetyl)pteroic acid. C5 - Homopteroic acid.

### III.3.4. Strains and growth media

*C. albicans* SC5315, *C. glabrata* CBS138, *C. parapsilosis* CDC317, *C. auris* B8441 and *C. neoformans* var. *grubii* H99 strains were used in this study. Cells were pre-cultured at 30°C, with orbital agitation (250 rpm) in yeast extract-peptone-dextrose (YPD) liquid medium containing per liter: 20 g glucose (Merck), 20 g yeast extract (Difco) and 10 g bacterial-peptone (LioChem). For minimal inhibitory concentration assays and checkerboard assays the Roswell Park Memorial Institute (RPMI) 1640 medium was used, containing per liter: 18 g glucose (Merck), 10.4 g RPMI-1640 (Sigma), 34.53 g morpholinepropanesulfonic acid (MOPS; Sigma).

### III.3.5. Minimal inhibitory concentration and Checkerboard assays

The susceptibility of the fungal cells towards the selected compounds and determination of the minimal inhibitory concentration (MIC) was performed by broth microdilution method according to Clinical and Laboratory Standards Institute (CLSI) M27-S4 [364], using a range of concentrations for the new compounds from 4µM to 1000µM. In order to evaluate the eventual interaction of the compounds with known effective drugs, fluconazole, amphotericin B, and caspofungin, checkerboard assays were also performed. Several combinations of each pair of drugs were used in each well, using several dilutions for each one. All plates were incubated at 37°C for 24 h to 48h and the growth was assessed by measuring the value of optical density (OD) at 600 nm in a microplate reader (SPECTROstar Nano, BMG Labtech). In all experiments, a negative control without any drug concentration and a control with only DMSO at the same volume of the higher drug concentration were carried out in order to exclude a potential cytotoxic effect of DMSO.

The procedure for testing the compounds Y7, Y11, 2a, 3a, 4a, 5a, 6a and 7a was slightly different. Those compounds are amphipathic and have the peculiar characteristic of adhering strongly to plastic surfaces, for this reason, the traditional 96-well polystyrene microplates could not be used. Instead, the experiments were performed in individual glass vials, one vial for each different condition. Additionally, the cells were added to the vials before the compounds, and the medium with the cells was previously acclimatized at 37°C. By the end of the growth inhibition experiments, the optical density was measured using a spectrophotometer (Hitachi u-2001) at 600 nm.

### **III.3.6. Cytotoxicity assays**

The cytotoxicity of compounds Y7, Y11, 2a, 3a, 4a, 5a, 6a and 7a was evaluated elsewhere [365]. The cytotoxicity of compound CH1 was not assessed given that it did not present inhibitory activity against the tested fungal pathogens. The cytotoxicity of compound C5 was evaluated in three cellular models purchased from European Collection of Authenticated Cell Cultures, HeLa human cervical cancer cell line, L929 mouse Fibroblast cell line, and VK Murine myeloma cells. The cells were cultured and routinely maintained at 37°C in a humidified chamber containing 5% of CO<sub>2</sub> (Binder CO<sub>2</sub> incubator C150) in Dulbecco's Modified Eagle Medium (DMEM; Gibco® by Life Technologies), supplemented with 10% of heat-inactivated Fetal Bovine Serum (FBS; Gibco® by Life Technologies), penicillin/streptomycin 100 U/ml (Gibco, ThermoFisher). Cell viability was assessed by the tetrazolium (MTT) reduction assay. Cells were seeded on 96-well plates (Orange Scientific) at a density of  $1 \times 10^4$ . After 24 h, medium was exchanged with 100ul of medium with a range of concentrations of C5 compound from 64 to 2 mg/L and incubated for 24 hours, DMSO was also

used at the highest concentration used for the drug dilution. Then MTT was added to each well and incubated at 37°C for 2 h. Absorbance was measured at 570 nm in a microplate reader (SPECTROstar Nano, BMG Labtech). Untreated cells were used as control, in order to determine the relative cell proliferation of treated cells.

### **III.3.7. Fluorescence microscopy**

Since compounds Y11, 2a, 3a, 4a, 6a, and 7a exhibit intrinsic fluorescence, it was possible to measure their intracellular localization in exposed fungal cells, with a Zeiss Axioplan fluorescence microscope (Carl Zeiss Microimaging) using excitation and emission wavelengths of 395 and 509 nm, respectively. Fluorescence images were captured using a cooled Zeiss AxioCam 503 color camera (Carl Zeiss Microscopy).

## **III.4. Results and discussion**

### **III.4.1 Virtual screening for potential inhibitors of Fol1**

#### **III.4.1.1. Modeling Fol1 protein structure**

In previous work from our lab, we identified Fol1, dihydropteroate synthase enzyme (EC: 2.5.1.15) as a potential drug target in pathogenic fungal species using a cross-species approach integrating enzyme essentiality within host prediction from 5 different genome-scale metabolic models (GSMM's) for *Candida* and *Cryptococcus* species. Information regarding the active site residues of this protein is available, and the residues are conserved among the 5 species, with a 73% identity and 93% similarity (Figure III.2). This conservation opens the

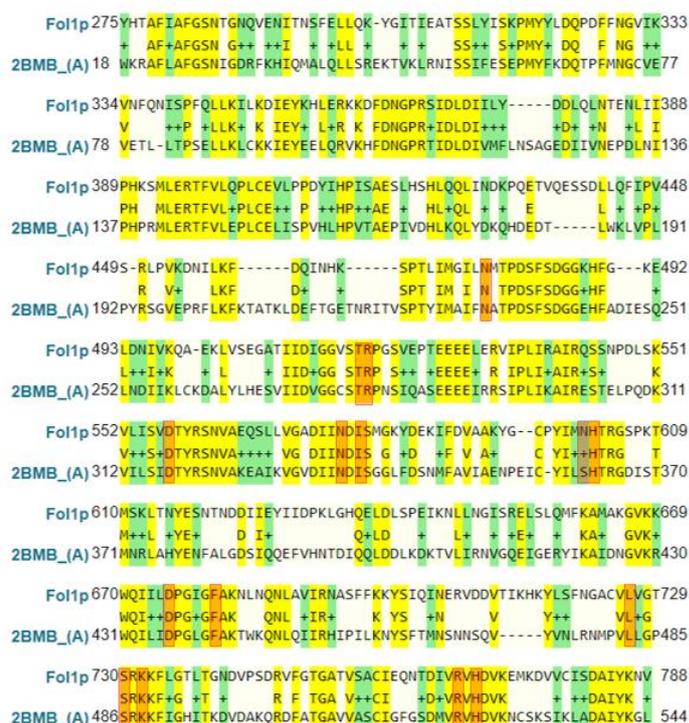
possibility of developing novel and effective inhibitory compounds targeting all five species.



**Figure III.2** - Multiple Sequence Alignments of Fol1 protein of *C. albicans* (C5\_00770C\_A) and respective orthologs of *C. parapsilosis* (CPAR2\_303390), *C. auris* (B9J08\_001330), *C. glabrata* (CAGL0J07920g) and *C. neoformans* (CNAG\_02786). The residues involved in the active site are highlighted in blue, with 73% identity and 93% similarity among all the proteins. Alignment performed with CLUSTAL 2.1.

Fol1 protein has no available structures for *Candida* or *Cryptococcus* species, the closest related structure being for *S. cerevisiae* (PDB code: 2BMB). In order to perform molecular docking studies and screening for compounds capable of inhibiting this target, the structure of Fol1 from *C. albicans* was predicted by

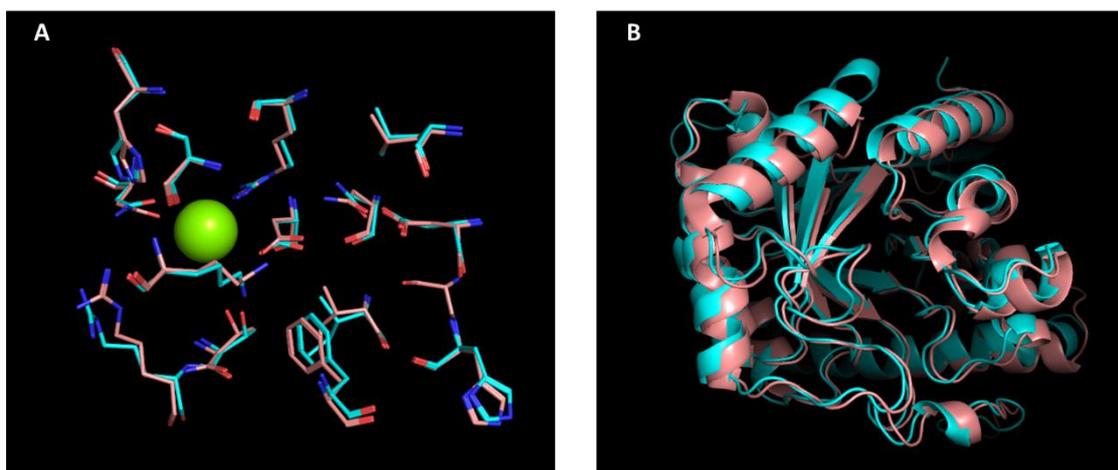
homology modeling using as a template the 3D structure of *S. cerevisiae* with a resolution of 2.30 Å and PMM an analog of (7,8-dihydropterin-6-yl)methyl diphosphate, one of the reagents of the enzyme. The *P. vivax* Fol1 structure (PDB code: 5Z79) with 2.90 Å resolution and PAB and HH2, also an analog of (7,8-dihydropterin-6-yl)methyl diphosphate, was also used as a template for some regions with no coverage in the Fol1 *S. cerevisiae* structure. *C. albicans* and *S. cerevisiae* 2BMB Fol1 sequences share 62% similarity and 42% identity, however, this identity is even higher if we consider only the residues reported to be involved in the active site [296], with 100% similarity and 93% identity (Figure III.3), making it possible to perform the modeling of *C. albicans* 3D protein structure with a high confidence, especially in the active site region, the most critical for the docking studies.



**Figure III.3** - Sequence alignment of *S. cerevisiae* (PDB code 2BMB) and *C. albicans* Fol1. Residues from the active site are highlighted in orange. Similar residues are

represented in green, and identical in yellow. The two sequences share globally 62% similarity and 42% identity, and 100% similarity and 93% identity if considering the active site alone.

The modeling was performed using MODELLER 9.23 and the best solution without atomic clashes and a DOPE score of -1.87 was used in the subsequent docking studies (Supplementary Data III.1). Comparison of the 2BMB structure with the modeled Fol1 *C. albicans* structure revealed a consistent overlap (Figure III.4B), with no significant differences observed in the positioning of side chains within the active site (Figure III.4A).

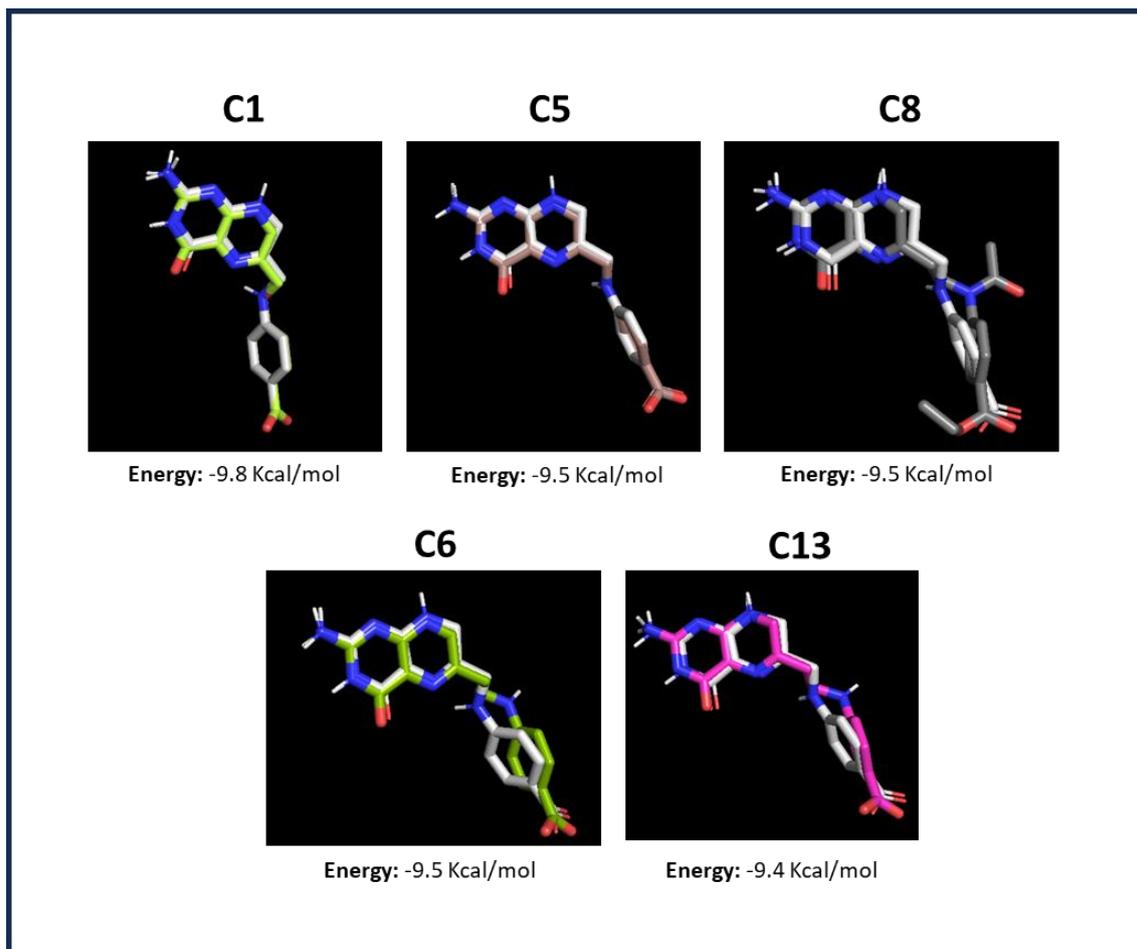


**Figure III.4** - Overlap between *S. cerevisiae* 3D structure (PDB code: 2BMB) and modeled structure of *C. albicans* using MODELLER 9.23. A – overlap of active side residues. B – overlap of the full structures.

#### III.4.1.2. Fol1 virtual drug screening

In order to find new compounds with the potential to bind to the active site of Fol1 and inhibit its activity, the modeled 3D structure of *C. albicans* Fol1 was used as the receptor for molecular docking studies against libraries of compounds

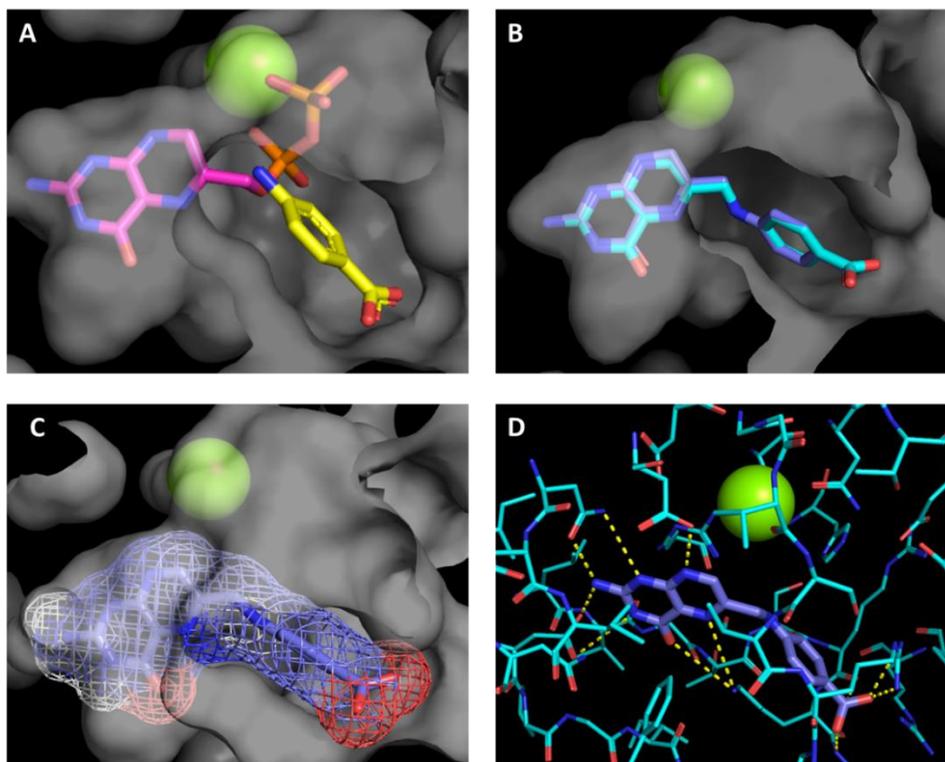
analogous to 7,8-dihydropteroate, the product of the Fol1 enzymatic reaction. Out of a total of 317 compounds tested, 37 exhibited a binding energy lower than -9.0 kcal/mol, with an additional 91 displaying energies lower than -8 kcal/mol. Some of the compounds with the lowest binding energies are represented in Figure III.5, and the full list of the compounds with all the binding energies can be found in Supplementary Data III.2.



**Figure III.5** - Overlap of 7,8-dihydropteroate (white) and 5 of the compounds with higher affinity to *C. albicans* Fol1 modeled active site 3D structure obtained from virtual drug screening through molecular docking studies. The respective binding energy is indicated for each compound. C1: 4-[(2-amino-4-oxo-1H-pteridin-6-yl)methoxy]benzoic acid; C5: Pterotic acid; C8: ethyl 4-[acetyl-2-(2-

amino-4-oxo-7,8-dihydro-1H-pteridin-6-yl)ethyl]amino]benzoate; C6:  
Homopteroic acid; C13: 4-((2-(2-Amino-4-oxo-3,4,7,8-tetrahydropteridin-6-  
yl)ethyl)amino)benzoic acid.

Among the most promising compounds, experimental testing was focused on compound C5, used as a case-study. Although not further pursued herein, all other identified potential Fol1 inhibitors constitute promising avenues for future exploitation. The molecular docking simulations of compound C5 in the active site of *C. albicans* Fol1 revealed that the compound seems to occupy the exact same position of 7,8-dihydropteroate, as expected (Figure III.6B). Similarly and also expected, the compound shares the same binding mode to the active site with the same polar interactions between the ligand and protein (R519, D557, N576, D675, R731, and K732) as observed for PAB in the carboxylate group and for (7,8-dihydropteridin-6-yl)methyl diphosphate in the amide groups [296], products of the enzymatic reaction catalyzed by Fol1 (Figures III.6A, C and D). Given this similar binding mode and the occupation of the same space as the ligands within the active site and its high affinity predicted by molecular docking studies, it is expected that the compound C5 will be an efficient inhibitor of this protein, competing with the substrate for the binding to the active site.



**Figure III.6** - A – PAB (yellow) and (7,8-dihydropterin-6-yl)methyl diphosphate (pink) docked in *C. albicans* Fol1 3D structure. B – overlap of compound C5 (dark blue) and 7,8-dihydropteroate (light blue) docked in *C. albicans* Fol1 3D structure. C – mesh representation of compound C5. D - Polar interactions between C5 and protein active site residues.

#### III.4.1.3. Antifungal activity assessment

In order to assess compound C5 antifungal activity, broth dilution antifungal susceptibility testing was performed against several pathogenic fungal species. C5 exhibited moderate inhibition against all tested species, including *C. albicans*, *C. glabrata*, *C. parapsilosis*, *C. auris*, and *C. neoformans*, with MIC50 values of 200mg/L (820  $\mu$ M). Interestingly the MIC80 against all tested *Candida* strains was also of 200mg/L, the compound displaying fungicidal activity at this concentration (Table III.1 and III.2). The MIC80 against *C. neoformans* was higher

than 512 mg/L. This lower capacity of C5 to inhibit *C. neoformans* may possibly be related to the presence of a capsule in this pathogen, which constitutes an important mechanism of resistance against toxic compounds.

Given the moderate fungicidal activity of C5, we explored the potential for synergistic effects when combined with other known effective antifungal drugs. Checkerboard assays were performed to evaluate a possible synergistic effect of C5 when in combination with Fluconazole, Amphotericin B, and Caspofungin, used as representatives of each drug class. These drugs are known to impact fungal cell wall or plasma membrane integrity, potentially facilitating C5's access to the interior of fungal cells.

C5 when in combination with Fluconazole was found to display a synergistic effect against *C. albicans* and *C. neoformans* with an Fractional Inhibitory Concentration (FIC) Index of 0.31 and 0.38, respectively, and an additive effect against *C. glabrata* and *C. parapsilosis* with FIC indexes between 0.75 and 1 (Table III.1). Synergistic assessment against *C. auris* was not feasible due to the strain's resistance to fluconazole. In contrast, no synergistic or additive effects were observed when C5 was combined with Amphotericin B or Caspofungin for any of the species or drug concentrations.

**Table III.1:** MIC50 of compound C5 and Fluconazole alone and in combination against *C. albicans*, *C. glabrata*, *C. parapsilosis*, *C. auris*, and *C. neoformans*.

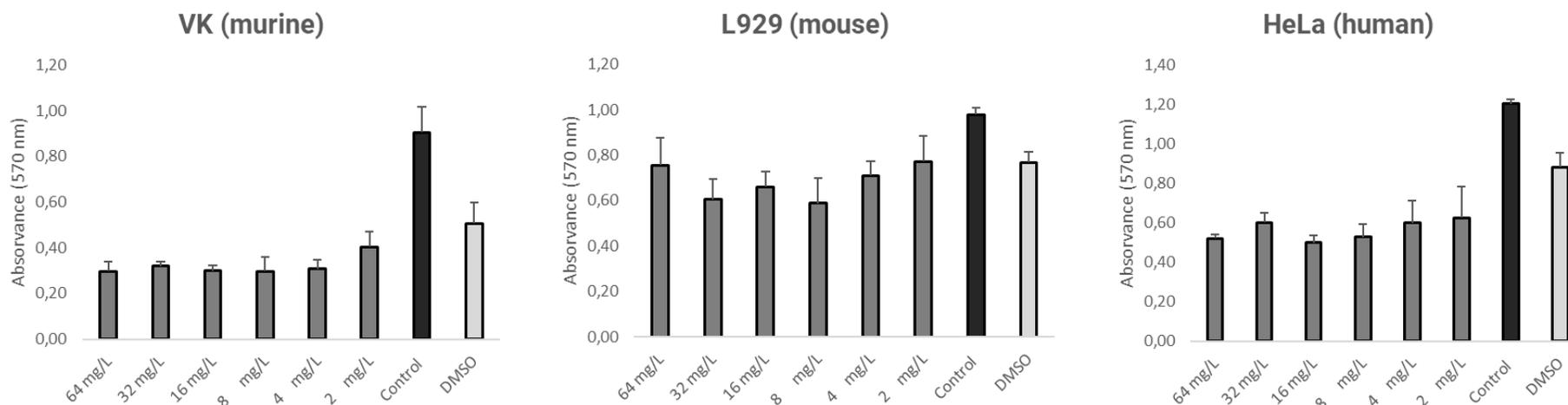
Strain	MIC50 Drug alone		MIC50 Drug Synergy		FC Index	Effect
	C5 (mg/L)	Fluco (mg/L)	C5 (mg/L)	Fluco (mg/L)		
<i>C. albicans</i>	256	1	16	0,25	0,31	synergistic
<i>C. glabrata</i>	256	16	2	16	1,01	additive
<i>C. parapsilosis</i>	256	2	64	1	0,75	additive
<i>C. neoformans</i>	256	0,5	32	0,125	0,38	synergistic
<i>C. auris</i>	256	>64	>128	>64	nd	indifferent

**Table III.2:** MIC80 of compound C5 and Fluconazole against *C. albicans*, *C. glabrata*, *C. parapsilosis* and *C. neoformans*.

Strain	MIC80 Drugalone	
	C5 (mg/L)	Fluco (mg/L)
<i>C. albicans</i>	256	128
<i>C. glabrata</i>	256	128
<i>C. parapsilosis</i>	256	4
<i>C. neoformans</i>	>512	0,5
<i>C. auris</i>	256	>64

#### III.4.1.4. C5 cytotoxicity assessment

Compound C5 showed moderate antifungal activity against *Candida* species, with a synergistic effect with fluconazole against *C. albicans*. In order to assess the possibility of this compound as a human antifungal drug we decided to evaluate its cytotoxicity against three different mammalian cell lines, HeLa (human), L929 (mouse), and VK (Murine). The cytotoxicity was evaluated by MTT re-duction assay upon exposure to different concentrations of C5, using absorbance measurements to quantitatively evaluate cell proliferation indicated by the reducing power of living cells. Compound C5 seems to have a toxic effect even at lower concentrations in VK and HeLa cell lines. For L929 cells, no cytotoxic effect was detected, when compared to control conditions (Figure III.7).



**Figure III.7** - Cytotoxicity assessment of compound C5 with a range of concentrations from 2 to 64 mg/L against VK, L929, and HeLa cell lines, upon 48h drug exposure. Cell viability was assessed by the tetrazolium (MTT) reduction assay, reading 570 nm absorbance. C5 showed cytotoxicity at all concentration ranges especially against, VK (murine) and HeLa (human) cell line

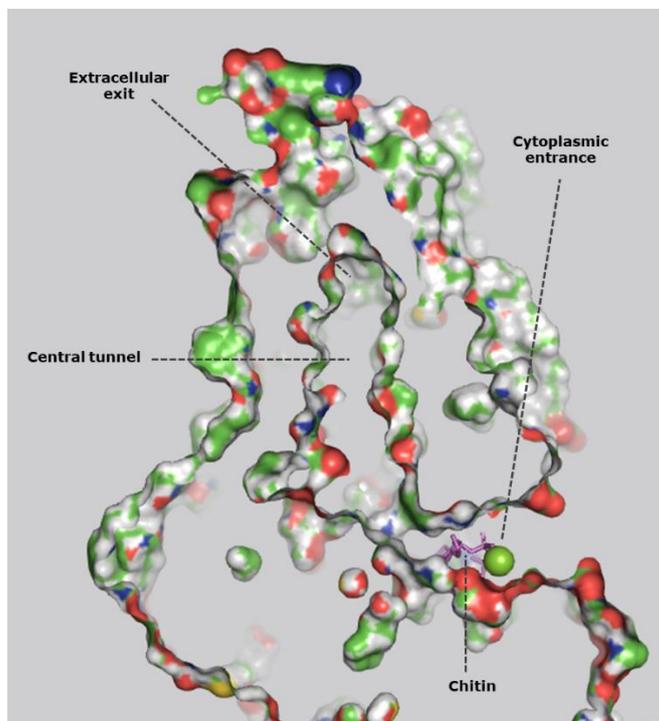
## III.4.2 Virtual screening for potential chitin synthase inhibitors

### III.4.2.1. Virtual drug screening

Adopting a similar strategy to our approach with Foll, molecular docking was used to conduct virtual drug screening of a large compound database targeting the active site of the *C. albicans* chitin synthase (EC number 2.4.1.16), previously identified as a promising new antifungal drug target lab [288,303,339]. *C. albicans*, similar to most fungi, has multiple chitin synthase isoforms, including Chs1, Chs2, and Chs3. Although these isoforms share the same function and catalyze the same enzyme reaction, they are expressed under different conditions throughout the fungal cell life cycle. Chs1 is primarily involved in primary septum formation, while Chs2 participates in repairing damaged chitin during cell separation and depositing chitin in hyphae. Chs3 serves as an additional chitin synthase in both yeast and hyphal forms. Despite the challenges posed in protein crystallization by its transmembrane nature, there is an available 3D structure for the *C. albicans* Chs2, deposited in the PDB with the code 7STM, at a resolution of 3.02 Å. This structure, co-crystallized with an analog of chitin and a magnesium cofactor in the active site, is ideal for molecular docking studies, given its resolution and information on active site location and substrate-binding residues.

The virtual drug screening was performed using molecular docking to access from a large compound database which ones are predicted to bind with greater affinity to the reported active site. From the PubChem database, we searched for compounds analogous to N-Acetyl-D-Glucosamine (chitin monomer) and selected all identified compounds, regardless of their availability for purchase. This screening encompassed a total of 20,000 compounds, leading to the

identification of 76 compounds exhibiting a high predicted affinity to the *C. albicans* Chs2, with a free energy of binding prediction lower than -10 kcal/mol (Supplementary Data III.3). Table III.3 displays 10 compounds with higher predicted affinity that were selected for further experimental validation.



**Figure III.8** - Surface representation of the CaChs2, highlighting the central tunnel and active site.

In the structure of *C. albicans* Chs2, it is possible to identify a central tunnel within the transmembrane domain, approximately 30 Å long. This tunnel, with the capacity to accommodate a chitin chain comprising approximately five or six GlcNAc units, is where chitin polymerization is expected to occur (Figure III.8). The binding site of the UDP-GlcNAc, is positioned close to the cytoplasmic entrance of this tunnel, where the residues Y319, E321, K441, D465, E603, D604, Q643, R644, R646 and W647 take an important role in the binding process and catalysis [366], which are highly conserved among fungi with 100% identity in *Candida* and *C. neoformans* (Figure III.9). Among these residues, Q643 and W647,

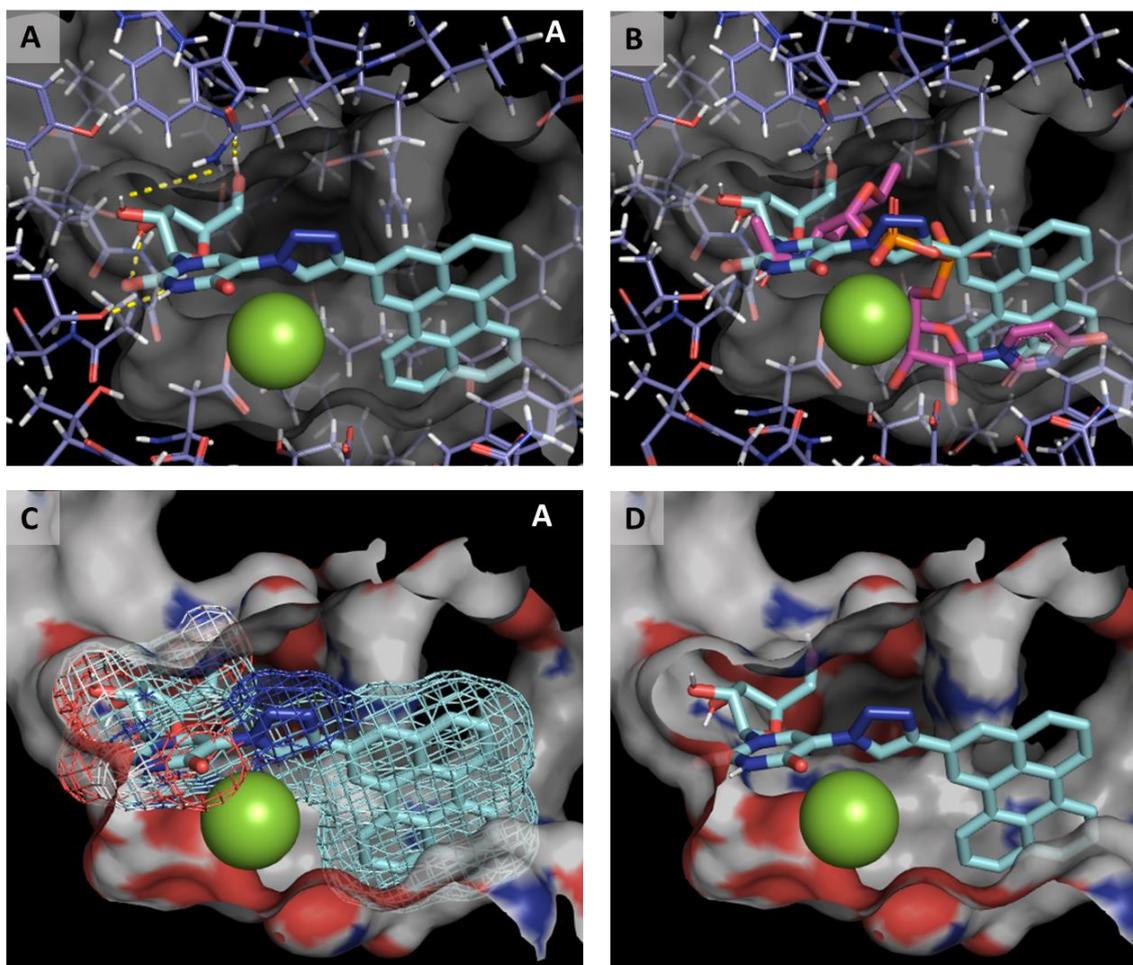
located in the polar region terminal, appear to be involved in the binding of the analogous compounds tested in this study (Figure III.10).

```

CR_09020C_A|CHS2      PRQTELMIVITMYNEDDILLGRTLKGVFKNIKYLESKARSSTWGKDSWKKIVVCIVSDGR
CPAR2_701490          PRQTEMMIVITMYNEDDVLGRTLKGVFKNIKYLESKRSSTWGKDSWKKIVVCIVSDGR
B9J08_005077         RRQTEMMVVITMYNEDDILLARTLKGVFKNIKHLESRSRSPVWGRDSWKKVVVCVSDGR
CAGL0J11506g         PRSTELLIVITMYNEDHILLGRTLKGVMDNIKHMVKKTRSSTWGPDAWKIVVCIVSDGR
CNAG_03326           -RKTELFIVMTMYNENSELLRLTNAVIAKNIHLTTRTRSKTWGPDWKKVVVCIVADGR
                      *.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:
CR_09020C_A|CHS2      TKINERAQALLAGLVYQEGLAQRSVDDKKVQAHMFEYTRRVGISKVTDVVKLTTEK-V
CPAR2_701490          TKINERAQALLAGLVYQEGLAQRSVDDKKVQAHMFEYTRRVGISSVSDVVKLTTEK-I
B9J08_005077         SKINERAQALLAALGVYQEGLAQRSRIDDKKVQGHIEYTRRVGIS-SVEDTVKLTTEK-V
CAGL0J11506g         AKINEKSLALLSSLGCYQDGFADKDEINGKKVTVHVEHTTMMNVVDVTDNVELRCDQST
CNAG_03326           KVVDPRLVKVLQMGVYAEVMDKDHVVDKETQAHIFEYTSQVVVSETGEVGFGS-----
                      :: :  :*  :*  :*  *..*..*..*  *:::*:  ::  :  :
CR_09020C_A|CHS2      VPVQMLFCLKETNAKKINSHRWCFQAIGQVLDPKIVVLLDCGTQPSGRSLYELWKEFDRD
CPAR2_701490          VPVQMLFCLKETNAKKINSHRWCFQAIGQVLDPKIVVLLDCGTQPTGKSLYKLWKEFDND
B9J08_005077         VPVQMLFCLKENNAKKINSHRWCFEALSQVLDPNIVVLLDCGTQPSGKSLYHLWKEFDKD
CAGL0J11506g         VPIQMLFCLKEQNKKKINSHRWAFEGFAELLRPNIIVTLLDAGTMPGKDSIYELWREFR-N
CNAG_03326           TPIQLLFCLKEQNKKKLNSHRWFFNAFGPLIKPNVVCVLLDVGTKPSGHSIYELYKCFEKH
                      .*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:
CR_09020C_A|CHS2      HRVAGACGEITTSCLKK-RQMITNPLVYGQNFYKISNILDKPTESSFGFISVLPGAFSAY
CPAR2_701490          HRVAGACGEIIASLKK-RQIVTNPLVYGQNFYKISNILDKPTESSFGFISVLPGAFSAY
B9J08_005077         PRVAGACGEIKASLKK-RQMLTNPIVYGQNFYKISNILDKPTESVFGFISVLPGAFSAY
CAGL0J11506g         PQVGGACGEIRTDLGNYSNLINPLIASQNFYKMSNILDKTTESNFGFITVLPGAFSAY
CNAG_03326           PTVGGACGEIFADTGKWKYLWNLPLVAGQNFYKMSNILDKPFESVFLGISVLPGAFSAY
                      *.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:.*:
CR_09020C_A|CHS2      RFIALQNDINGVGPLEKYFKGEFLHSSGELDPNDDEFQMKHMLKKEEAGIFTSNMYLAED
CPAR2_701490          RNIALQNDINGRGPLEKYFKGEFLHSSGELDPQDEYEMKRTLLKKEEAGIFTSNMYLAED
B9J08_005077         RYVALQNDINGKGPLEKYFKGEFLHSGELDPNDDEYDLKVKMLKKEEAGIFTSNMYLAED
CAGL0J11506g         RLEAVQG-----LPLQKYFYGESMEH-----EKFHFFSSNMYLAED
CNAG_03326           RYDAPANHADGTGLAAYFRGELMNQPG-----ATATIFDRNKFLAED
                      *  *:  .      **  **  **  :.  :  :  :  :  :  :  :  :  :  :  :  :  :
CR_09020C_A|CHS2      RILCFELVAKRGCNWLRLRYCKSARAETDVPEGLAEFILQRRRWLNGSFFAAIYSLVHFYK
CPAR2_701490          RILCFELVAKQGCNWLRLRYCKSARAETDVPERLAEFILQRRRWLNGSFFAAIYSLVHFFK
B9J08_005077         RILCYELVAKPGCSWVLRYSKASAEADVPEGLAEFILQRRRWLNGSFFAAIYSLAHFYK
CAGL0J11506g         RILCFEVMKKNANWILKYSRSSHATDVPDRVPEFILQRRRWLNGSFFASVYSFCHFYR
CNAG_03326           RILAFEIVVKKNARWRLQYVKAAGDVPATVPEFISQRRRWLNGSIFAATYAMVCFWR
                      ***.:** * .. * :.*:  ::  *  ***  ..***  *****:***:  *::  *::

```

**Figure III.9** - Multiple Sequence Alignments of residues 307-664 of Chs2 protein of *C. albicans* (CR\_09020C\_A) and respective orthologs of *C. parapsilosis* (CPAR2\_701490), *C. auris* (B9J08\_005077), *C. glabrata* (CAGL0J11506g) and *C. neoformans* (CNAG\_03326), performed with CLUSTAL 2.1. The residues involved in the active site are highlighted in blue, with 100% identity among all the proteins.



**Figure III.10** – A – Compound 7a (Best affinity in virtual drug screening -11.8 Kcal/mol) docked in the active site of CaChs2 with respective observed polar interactions. B – overlap of compound 7a (light blue) and Chitin (pink). C and D – mesh and stick representation, respectively, of compound 7a in surface view of CaChs2 active site.

#### III.4.2.2. Antifungal activity assessment and cellular localization

In order to assess the antifungal activity of the 10 selected potential Chs2 inhibitors, broth dilution antifungal susceptibility testing was performed against several pathogenic fungal species, including *C. albicans*, *C. glabrata*, *C. parapsilosis*,

*C. auris*, and *C. neoformans*. Compounds 2a, 4a, 5a, and y7 were found to exhibit moderate fungistatic activity against *C. neoformans*, with a MIC values of 400  $\mu$ M (Table III.3). Among these, compound 2a displayed the strongest inhibition, with an MIC80 value of 400  $\mu$ M against *C. neoformans*.

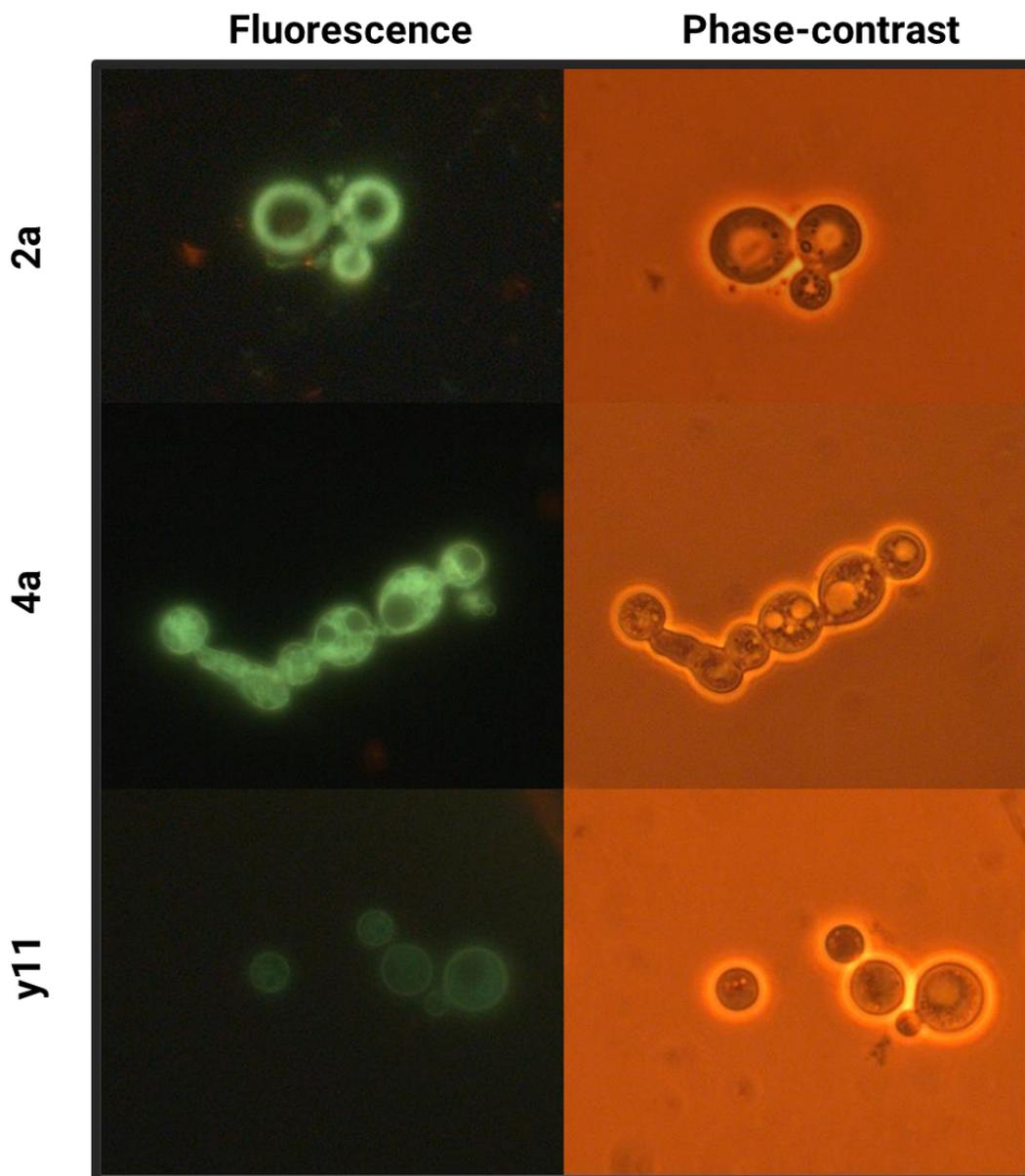
**Table III.3** – MIC50 values in  $\mu$ M of different compounds targeting the Chs2 protein against the 5 indicated pathogenic fungal species. MIC50 for compounds 2a, 4a, 5a, and y7 with a higher inhibition activity against *C. neoformans* are highlighted in bold.

Drug ID	MIC50 ( $\mu$ M)				
	<i>C. albicans</i>	<i>C. glabrata</i>	<i>C. neoformans</i>	<i>C. auris</i>	<i>C. parapsilosis</i>
c1	>820	>820	>820	>820	>820
<b>2a</b>	>400	>400	<b>400</b>	>400	nt
<b>3a</b>	>400	>400	>400	>400	nt
<b>4a</b>	>400	>400	<b>400</b>	>400	nt
<b>5a</b>	>400	>400	<b>400</b>	>400	nt
<b>6a</b>	>400	>400	>400	>400	nt
<b>7a</b>	>400	>400	>400	>400	nt
<b>y7</b>	>400	>400	<b>400</b>	>400	nt
<b>y11</b>	>400	>400	>400	>400	nt

nt: not tested

Notably, compounds 2a, 4a, and y11 emit fluorescence, allowing us to assess their subcellular localization following exposure. The localization studies revealed that all three compounds were localized in the membrane and cytoplasm of *C. neoformans*. This effect was more pronounced in compounds 2a and 4a compared to y11, potentially contributing to explain the higher inhibitory effect of these compounds (Figure III.11). However, these observations allow us to infer that the compounds are being internalized into the cells, especially compounds 2a and 4a. To verify if any of the compounds could be used in combination therapy, checkerboard assays were performed with each compound in combination with

Amphotericin B. However, none of the compounds showed any synergistic effect with FIC values between 0.75 and 1, indicating only an additive effect between the 2 drugs, with the best value being achieved for compound 5a (0.75).



**Figure III.11** – Fluorescence of *C. neoformans* cells after exposure to Chs2 predicted inhibitors 2a, 4a, and y11, which exhibit intrinsic fluorescence. The higher preponderance of intracellular fluorescence observed for compounds 2a and 4a is possibly connected to their stronger antifungal activity when compared

to y11, which mostly accumulates in the cell periphery. Phase contrast images of the same cells are used as controls.

Moreover, compounds 2a, 3a, 4a, 5a, 6a, 7a, y7, and y11 demonstrated low cytotoxic effects against mammalian cells (Vero cells), with MIC<sub>50</sub> values higher than 50 μM (maximum tested concentration), as determined in [363].

### III.5. Conclusions

This study highlights the importance of molecular docking as a valuable tool for virtual drug screening, enabling the identification of promising compounds targeting specific fungal enzymes. Through the used approach, successful identification of new compounds with potential inhibitory activity against key enzymes involved in important cellular processes and fungal cell wall biosynthesis, Fol1 and chitin synthase respectively, was achieved.

The obtained results led to the identification of homopteroic acid (C5) as having fungicidal activity against a range of pathogenic *Candida* species. The observed synergistic effects with fluconazole against *C. albicans* suggest the possibility of combination therapy to enhance the efficacy of existing antifungal drugs. However, the cytotoxicity of C5 raises concerns regarding its safety profile for clinical use. Further studies are warranted to elucidate the mechanisms underlying the interaction between compound C5 and Fol1 and to optimize its pharmacokinetic properties for improved safety and efficacy. High-throughput *in silico* molecular docking also uncovered 76 potential Chs2 inhibitors. Among others, 5-(perylene-2-ylethynyl)-arabino-uridine (2a) was identified as the most promising compound, with no detected cytotoxicity [365] and exhibiting fungistatic activity against *C. neoformans* at 400 μM.

In conclusion, our study highlights the importance of a multidisciplinary approach that combines systems biology, computational modeling, experimental validation the discovery and development of novel antifungal agents. The compounds identified in this study show potential for further exploration and optimization, with the ultimate goal of addressing the urgent clinical need for effective antifungal therapies. Additionally, it is important to note that the compounds tested represent only a fraction of those identified in the virtual drug screening experiments. Their potential effectiveness, if tested experimentally, could yield valuable insights and potentially lead to the discovery of even more promising candidates.

## IV. Final Discussion

The currently available antifungal treatments are limited and only reasonably effective, especially against invasive fungal infections in immunocompromised patients. This is evidenced by the alarmingly high mortality rates, particularly in cases involving *Candida* and *Cryptococcus* infections [3]. The proliferation of resistant clinical isolates, with some isolates demonstrating multi-resistance to the three main antifungal classes [14], further reinforces the urgent need for the development of new antifungal strategies, new drug targets, and new antifungal drugs. However, the process of developing new drugs remains a difficult and costly task, as evidenced by the reduced number of new classes of antifungals that have emerged in recent decades. Building upon this issue, this thesis focused on the exploitation of an *in silico* drug discovery pipeline to disclose new pathways/enzymes essential for the pathogen survival or adaptation to the host environment, followed by *in silico* and *in vivo* screening for inhibitors of such essential enzymes to be used as potential new antifungal drugs. Focusing on clinically relevant pathogenic invasive fungi, including *C. albicans*, *C. auris*, *C. glabrata*, *C. parapsilosis*, and *C. neoformans*, an innovative multidisciplinary approach was used, integrating GSMM's, virtual drug screening using molecular docking and experimental testing, contributing for the comprehension of fungal metabolic intricacies, the discovery of new drug targets and new potential drug candidates (Figure IV.1).

Chapter II describes the reconstructions of previously unavailable GSMM's for *C. albicans* SC5314, *C. parapsilosis* CDC317, *C. auris* B11221\_V1, and *C. neoformans* var. *grubii* H99 strains. The four models were validated using experimental data and proved accurate when predicting compounds that can be used as sole carbon

or nitrogen sources and specific growth rates. Globally, all the built models are able to predict the assimilation of sole carbon or nitrogen sources with at least 85% certainty. As important as correct predictions are, the cases where predictions have failed are also relevant, as they often revealed the lack of characterization of certain enzymes or metabolic pathways in the pathogen, allowing the identification of gaps in the understanding of pathogen metabolism that may play a crucial role in its environmental/host adaptation and virulence mechanisms. Indeed, we provide an extensive list of enzymes lacking characterization in those pathogens involved in assimilation pathways, especially in *C. neoformans*.

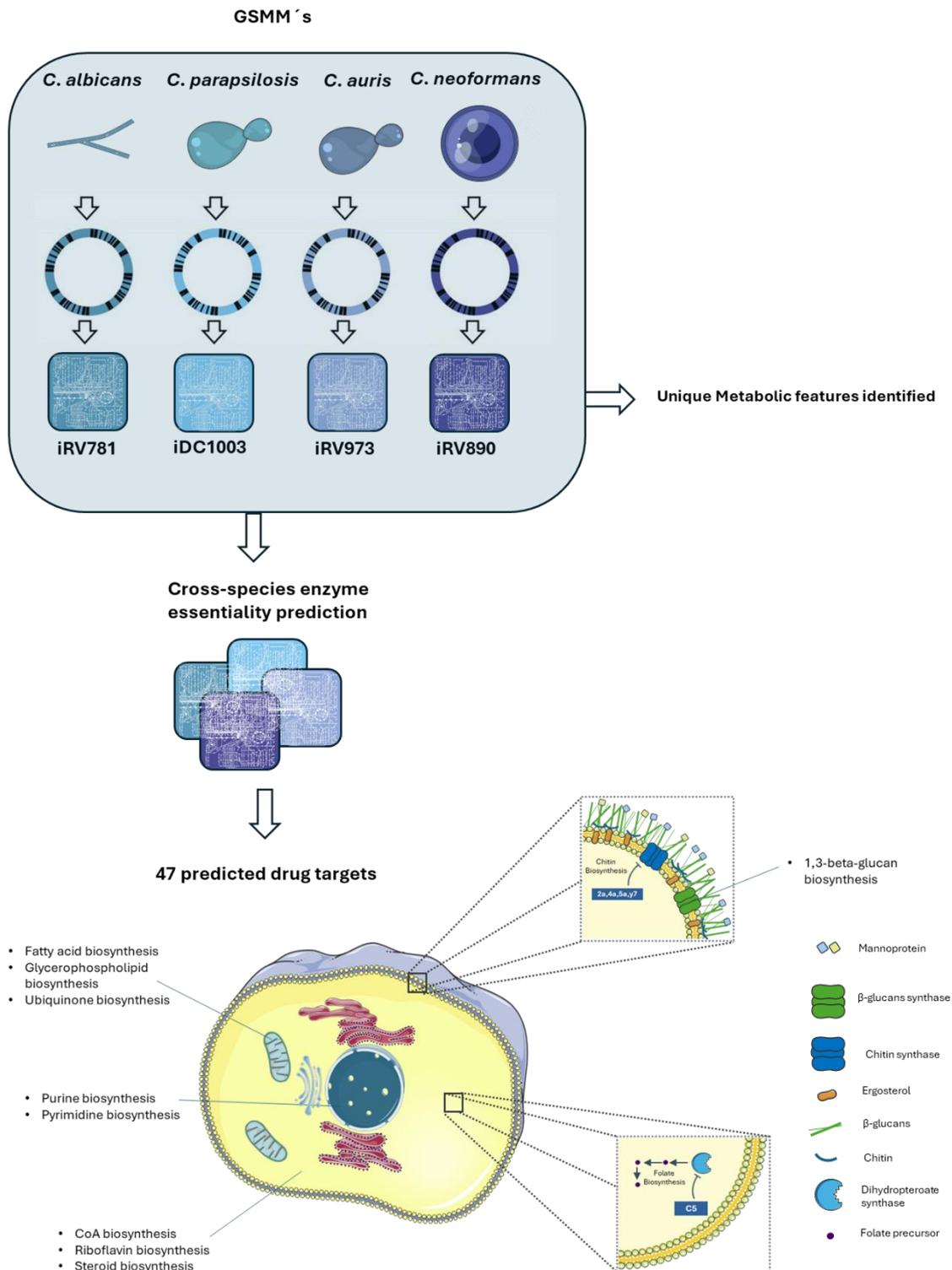
*Candida* species are commensal microorganisms commonly found as part of the normal microbial flora in various mucosal surfaces in the human body or even in the skin [367]. The GSMM's for the *Candida* species led to the identification of some unique metabolic characteristics in these species, mostly related to the utilization of alternative carbon and nitrogen sources or the acquisition of other nutrients. Additionally, there are potential mechanisms directly related to competitive advantages and interplay with the rest of the microflora, such as in *C. auris*, the enzyme chloride peroxidase associated with the production of halogenated antibiotics. Potential defense mechanisms against the host or toxic compounds were also identified, such as the enzyme inositol oxygenase in *C. albicans* and *C. parapsilosis*, involved in the resistance to toxic ergosterol analogs, or the enzyme sphingolipid 4-desaturase possibly related to resistance to azoles through alterations in membrane composition. Additionally, the enzyme N-acetylglucosamine kinase may represent a survival mechanism for *Candida* species within phagosomes. Another example is the enzyme quercetinase involved in resistance to quercetin in *C. auris*, possibly due to exposure to this compound in the environment or as a result of the human diet.

*C. neoformans*, on the other hand, is characterized for colonizing the central nervous system [368], thriving in an environment that is scarcer in certain nutrients and requiring unique adaptation capabilities compared to *Candida* species [369]. *C. neoformans* model allowed the identification of some interesting and unique metabolic features related to this particular environment. For example, L-gulonolactone oxidase and gluconolactonase are involved in ascorbate biosynthesis from inositol, a metabolite that is abundant in the human brain [327], ascorbate conferring resistance to fluconazole in this pathogen. Additional enzymes related to inositol assimilation as carbon source were also proposed, as well as enzymes related to inositol storage such as 3-phytase, which is important for *C. neoformans* dissemination into the brain [356]. Additionally, enzymes from melanin biosynthesis pathway were also identified, an important mechanism of defense in this pathogen against human defenses and toxic compounds such as caspofungin and amphotericin B [322,323].

GSMM's enabled the identification of several unique metabolic features, which may open new doors for exploring new potential mechanisms of inhibition or drug resistance in these pathogens. However, our approach was also focused on the identification of essential pathways and enzymes for the survival of these fungi inside the human host environment, preferably common to more than one fungal species in order to identify potential new fungal-specific drug targets with reduced potential of cross-resistance. For that purpose, a cross-species comparison was performed utilizing GSMM from five different *Candida* species and *C. neoformans* to identify essential enzymes for the survival of the pathogens when facing conditions similar to the ones of the human serum. With this approach, we identified 47 enzymes predicted as essential in all *Candida* species included in the study and *C. neoformans*. Notably, targets of azoles and echinocandins were identified in this list, revealing the potential of this approach

for the identification of new drug targets. Most of the remaining targets have a human homolog, which may make it difficult to design effective compounds with reduced toxicity. However, if the host and pathogen protein structures differ significantly from each other it may be possible to take advantage of those structural differences. The most interesting compounds are those that are fungal specific, capable of inhibiting all fungal targets, and able to enter or be transported into the cell, as these should exhibit high antifungal activity with reduced cytotoxicity. However, in the design of a compound, it is necessary to take into account factors beyond its specificity and intracellular access. For example, the pathogen may metabolically inactivate the compound through metabolism or degradation, or it may possess efflux pumps that actively remove the compound from inside the cell. Additionally, it is challenging to be able to predict, for instance, the compound's access to cells growing within biofilms, which are characteristic of *Candida* species, or ability to cross the blood-brain barrier, as required for the treatment of cryptococcosis, or even the interaction of the compound with other compounds in the environment.

Aware of these limitations, in Chapter III, we also decided to expand our strategy for discovering new compounds with antifungal activity against some of the previously identified targets by using an *in silico* pipeline, integrating protein 3D determination by homology modeling (if required), virtual drug screening from large compound databases using molecular docking and experimental testing in order to assess antifungal activity and human cytotoxicity.



**Figure IV.1** - Schematic representation of the work performed in this thesis, with a focus on GSMM reconstruction and the compounds found from virtual drug screening that went through experimental testing and showed promising results (C5, 2a, 4a, 5a, y7).

Chitin synthase and dihydropteroate synthase were the two enzymes selected for virtual drug screening and experimental testing, given their potential as possible drug targets. Both enzymes are fungal-specific and have key roles in fungal cellular structure or function.

Dihydropteroate synthase, an enzyme encoded by the *FOL1* gene and with the EC number 2.5.1.15 is essential in most fungal cells, including *Candida* species which do not have a folate transporter [296], and rely on folate *de novo* biosynthesis. Folate has a crucial role for fungal cells, being required for example as a precursor for S-adenosyl-L-methionine biosynthesis, which in turn is required in the ergosterol biosynthesis pathway. Dihydropteroate synthase 3D structure was determined by homology modeling, and virtual drug screening was performed in the active site of the protein against 317 homologous compounds to 7,8-dihydropteroate, the product of the enzymatic reaction. C5, one of the identified compounds with a higher predicted affinity to the protein was found to have fungicidal activity against all *Candida* species tested with a MIC value of 200mg/L. More interesting was the synergistic effect of this compound against *C. albicans* when in combination with fluconazole with an FIC value of 0,31. This synergistic effect suggests that compound C5 might not be able to access the cell effectively unless there is membrane permeabilization (such as that caused by fluconazole). Despite the initial promising results, compound C5 showed cytotoxicity against human cell lines, even at lower concentrations. Nevertheless, we highlight the potential of this compound for the possibility of combination therapy to enhance the efficacy of existing antifungal drugs, keeping

in mind the concerns regarding its safety profile and the need for optimization of its pharmacokinetic properties for improved safety and efficacy.

Chitin synthase was the second enzyme included in the virtual drug screening. This enzyme is encoded by several genes in fungal species and plays a critical role in the integrity of the cell wall, a structure that does not exist in human cells, making this enzyme a great candidate for drug targeting. High-throughput virtual drug screening was performed in the active site of the 3D structure of the Chs2 protein of *C. albicans* available in PDB against 20,000 chitin homologous compounds, uncovering 76 potential Chs2 high-affinity inhibitors. Among the 9 compounds selected for experimental testing, none showed antifungal activity against *Candida* species, however, compounds 2a, 4a, 5a, and y7 showed promising fungistatic activity against *C. neoformans*, with MIC<sub>50</sub> values of 400  $\mu$ M, compounds 2a, 4a and y11 being effectively internalized into *C. neoformans* cells. While the antifungal activity of these compounds is relatively low even against *C. neoformans*, the active principal may be explored in the future in the design of an inhibitor with higher efficacy. Why the compounds are ineffective against *Candida* species remains to be explained, but it is possible to hypothesize that *Candida* species may have efflux pumps that actively export these compounds, structural differences in the cell membranes that prevent drug diffusion, or even the presence of multiple chitin synthases with redundant activities.

Looking to the future, the recent launch of AlphaFold3 [191] and its new beta-phase server opens exciting new possibilities in the field of fungal pathogen research and drug discovery. By integrating AlphaFold3's advanced capabilities in complex structure prediction into our *in silico* results, we can expect to be able to perform the virtual drug screening process for the remaining drug targets predicted by GSMMs much more efficiently and accurately. Additionally, this

new tool could also be very useful in the design of new compounds based on those already identified in this thesis, with improved efficiency and safety.

Overall, this thesis highlights the complexity of fungal metabolism and the importance of innovative approaches in drug discovery to address the growing threat of fungal infections and drug resistance. With an innovative *in silico* approach, integrating GSMMs, a cross-species genes essentiality prediction, and virtual drug screening against selected potential drug targets, we not only were able to identify new drug targets and potential antifungal compounds, but also enlighten some specific and possibly important metabolic features of the fungal pathogens, related to its host adaptation, infection dissemination or drug/stress resistance. Ultimately, we expect that this work will have an impact on the discovery of new therapeutic strategies to target fungal pathogens, which requires our utmost attention due to its impact on public health, resulting from high mortality rates, the emerging increase in drug resistance, and the limited therapeutic options currently available.

## V. References:

- [1] Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 2012;380. [https://doi.org/10.1016/S0140-6736\(12\)61729-2](https://doi.org/10.1016/S0140-6736(12)61729-2).
- [2] Casadevall A. Immunity to Invasive Fungal Diseases. *Annu Rev Immunol* 2022;40. <https://doi.org/10.1146/annurev-immunol-101220-034306>.
- [3] Denning DW. Global incidence and mortality of severe fungal disease. *Lancet Infect Dis* 2024. [https://doi.org/10.1016/S1473-3099\(23\)00692-8](https://doi.org/10.1016/S1473-3099(23)00692-8).
- [4] Nami S, Aghebati-Maleki A, Morovati H, Aghebati-Maleki L. Current antifungal drugs and immunotherapeutic approaches as promising strategies to treatment of fungal diseases. *Biomedicine and Pharmacotherapy* 2019;110. <https://doi.org/10.1016/j.biopha.2018.12.009>.
- [5] Perfect JR. The antifungal pipeline: A reality check. *Nat Rev Drug Discov* 2017;16. <https://doi.org/10.1038/nrd.2017.46>.
- [6] 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. *Clinical Infectious Diseases* 2015;62. <https://doi.org/10.1093/cid/civ933>.
- [7] Walsh TJ, Anaissie EJ, Denning DW, Herbrecht R, Kontoyiannis DP, Marr KA, et al. Treatment of aspergillosis: Clinical practice guidelines of the infectious diseases society of America. *Clinical Infectious Diseases* 2008;46. <https://doi.org/10.1086/525258>.
- [8] Perfect JR, Dismukes WE, Dromer F, Goldman DL, Graybill JR, Hamill RJ, et al. Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the infectious diseases society of America. *Clinical Infectious Diseases* 2010;50. <https://doi.org/10.1086/649858>.
- [9] Odds FC, Brown AJP, Gow NAR. Antifungal agents: Mechanisms of action. *Trends Microbiol* 2003;11. [https://doi.org/10.1016/S0966-842X\(03\)00117-3](https://doi.org/10.1016/S0966-842X(03)00117-3).
- [10] Maertens JA. History of the development of azole derivatives. *Clinical Microbiology and Infection* 2004;10. <https://doi.org/10.1111/j.1470-9465.2004.00841.x>.
- [11] Pfaller MA, Castanheira M, Lockhart SR, Ahlquist AM, Messer SA, Jones RN. Frequency of decreased susceptibility and resistance to echinocandins among fluconazole-resistant bloodstream isolates of *Candida glabrata*. *J Clin Microbiol* 2012;50. <https://doi.org/10.1128/JCM.06112-11>.
- [12] Pfaller MA, Messer SA, Woosley LN, Jones RN, Castanheira M. Echinocandin and triazole antifungal susceptibility profiles for clinical opportunistic yeast and mold isolates

- collected from 2010 to 2011: Application of new CLSI clinical breakpoints and epidemiological cutoff values for characterization of geographic and temporal trends of antifungal resistance. *J Clin Microbiol* 2013;51. <https://doi.org/10.1128/JCM.00308-13>.
- [13] 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. <https://doi.org/10.1128/AAC.03255-14>.
- [14] Lockhart SR, Etienne KA, Vallabhaneni S, Farooqi J, Chowdhary A, Govender NP, et al. Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. *Clinical Infectious Diseases* 2017;64. <https://doi.org/10.1093/cid/ciw691>.
- [15] Arendrup MC. Update on antifungal resistance in *Aspergillus* and *Candida*. *Clinical Microbiology and Infection* 2014;20. <https://doi.org/10.1111/1469-0691.12513>.
- [16] Lestrade PP, Bentvelsen RG, Schauwvlieghe AFAD, Schalekamp S, Van Der Velden WJFM, Kuiper EJ, et al. Voriconazole resistance and mortality in invasive aspergillosis: A multicenter retrospective cohort study. *Clinical Infectious Diseases* 2019;68. <https://doi.org/10.1093/cid/ciy859>.
- [17] Beer KD, Farnon EC, Jain S, Jamerson C, Lineberger S, Miller J, et al. Multidrug-Resistant *Aspergillus fumigatus* Carrying Mutations Linked to Environmental Fungicide Exposure — Three States, 2010–2017 . *MMWR Morb Mortal Wkly Rep* 2018;67. <https://doi.org/10.15585/mmwr.mm6738a5>.
- [18] Snelders E, Van Der Lee HAL, Kuijpers J, Rijs AJMM, Varga J, Samson RA, et al. Emergence of azole resistance in *Aspergillus fumigatus* and spread of a single resistance mechanism. *PLoS Med* 2008;5. <https://doi.org/10.1371/journal.pmed.0050219>.
- [19] Kidd SE, Goeman E, Meis JF, Slavin MA, Verweij PE. Multi-triazole-resistant *Aspergillus fumigatus* infections in Australia. *Mycoses* 2015;58. <https://doi.org/10.1111/myc.12324>.
- [20] Chowdhary A, Kathuria S, Xu J, Sharma C, Sundar G, Singh PK, et al. Clonal Expansion and Emergence of Environmental Multiple-Triazole-Resistant *Aspergillus fumigatus* Strains Carrying the TR34/L98H Mutations in the *cyp51A* Gene in India. *PLoS One* 2012;7. <https://doi.org/10.1371/journal.pone.0052871>.
- [21] Chowdhary A, Sharma C, van den Boom M, Yntema JB, Hagen F, Verweij PE, et al. Multi-azole-resistant *Aspergillus fumigatus* in the environment in Tanzania. *Journal of Antimicrobial Chemotherapy* 2014;69. <https://doi.org/10.1093/jac/dku259>.
- [22] Verweij PE, Mellado E, Melchers WJG. Multiple-Triazole-Resistant Aspergillosis. *New England Journal of Medicine* 2007;356. <https://doi.org/10.1056/nejmc061720>.
- [23] Chowdhary A, Sharma C, Kathuria S, Hagen F, Meis JF. Azole-resistant *aspergillus fumigatus* with the environmental TR46/Y121F/T289A mutation in India. *Journal of Antimicrobial Chemotherapy* 2014;69. <https://doi.org/10.1093/jac/dkt397>.

- [24] Lockhart SR, Frade JP, Etienne KA, Pfaller MA, Diekema DJ, Balajee SA. Azole resistance in *Aspergillus fumigatus* isolates from the ARTEMIS global surveillance study is primarily due to the TR/L98H mutation in the *cyp51A* gene. *Antimicrob Agents Chemother* 2011;55. <https://doi.org/10.1128/AAC.00185-11>.
- [25] Wiederhold NP, Gil VG, Gutierrez F, Lindner JR, Albatineh MT, McCarthy DI, et al. First detection of TR34 L98H and TR46 Y121F T289A Cyp51 mutations in *aspergillus fumigatus* isolates in the United States. *J Clin Microbiol* 2016;54. <https://doi.org/10.1128/JCM.02478-15>.
- [26] Bermas A, Geddes-McAlister J. Combatting the evolution of antifungal resistance in *Cryptococcus neoformans*. *Mol Microbiol* 2020;114. <https://doi.org/10.1111/mmi.14565>.
- [27] Pfaller MA, Diekema DJ, Gibbs DL, Newell VA, Bijie H, Dzierzanowska D, et al. Results from the ARTEMIS DISK global antifungal surveillance study, 1997 to 2007: 10.5-year analysis of susceptibilities of noncandidal yeast species to fluconazole and voriconazole determined by CLSI standardized disk diffusion testing. *J Clin Microbiol* 2009;47. <https://doi.org/10.1128/JCM.01747-08>.
- [28] Zafar H, Altamirano S, Ballou ER, Nielsen K. A titanic drug resistance threat in *Cryptococcus neoformans*. *Curr Opin Microbiol* 2019;52. <https://doi.org/10.1016/j.mib.2019.11.001>.
- [29] Gerstein AC, Fu MS, Mukaremera L, Li Z, Ormerod KL, Fraser JA, et al. Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *MBio* 2015;6. <https://doi.org/10.1128/mBio.01340-15>.
- [30] Mota Fernandes C, Dasilva D, Haranahalli K, McCarthy JB, Mallamo J, Ojima I, et al. The Future of Antifungal Drug Therapy: Novel Compounds and Targets. *Antimicrob Agents Chemother* 2021;65. <https://doi.org/10.1128/AAC.01719-20>.
- [31] Hoekstra WJ, Garvey EP, Moore WR, Rafferty SW, Yates CM, Schotzinger RJ. Design and optimization of highly-selective fungal CYP51 inhibitors. *Bioorg Med Chem Lett* 2014;24. <https://doi.org/10.1016/j.bmcl.2014.05.068>.
- [32] Lockhart SR, Fothergill AW, Iqbal N, Bolden CB, Grossman NT, Garvey EP, et al. The investigational fungal Cyp51 inhibitor VT-1129 demonstrates potent in vitro activity against *cryptococcus neoformans* and *cryptococcus gattii*. *Antimicrob Agents Chemother* 2016;60. <https://doi.org/10.1128/AAC.02770-15>.
- [33] Yates CM, Garvey EP, Shaver SR, Schotzinger RJ, Hoekstra WJ. Design and optimization of highly-selective, broad spectrum fungal CYP51 inhibitors. *Bioorg Med Chem Lett* 2017;27. <https://doi.org/10.1016/j.bmcl.2017.06.037>.
- [34] Warrilow AGS, Hull CM, Parker JE, Garvey EP, Hoekstra WJ, Moore WR, et al. The clinical candidate VT-1161 is a highly potent inhibitor of *candida albicans* CYP51 but fails to bind the human enzyme. *Antimicrob Agents Chemother* 2014;58. <https://doi.org/10.1128/AAC.03707-14>.

- [35] Shubitz LF, Trinh HT, Galgiani JN, Lewis ML, Fothergill AW, Wiederhold NP, et al. Evaluation of VT-1161 for treatment of coccidioidomycosis in murine infection models. *Antimicrob Agents Chemother* 2015;59. <https://doi.org/10.1128/AAC.00593-15>.
- [36] Sobel JD, Nyirjesy P. Oteseconazole: An advance in treatment of recurrent vulvovaginal candidiasis. *Future Microbiol* 2021;16. <https://doi.org/10.2217/fmb-2021-0173>.
- [37] Hoy SM. Oteseconazole: First Approval. *Drugs* 2022;82. <https://doi.org/10.1007/s40265-022-01734-y>.
- [38] Lu R, Hollingsworth C, Qiu J, Wang A, Hughes E, Xin X, et al. Efficacy of oral encochleated amphotericin B in a mouse model of cryptococcal meningoencephalitis. *MBio* 2019;10. <https://doi.org/10.1128/mBio.00724-19>.
- [39] Delmas G, Park S, Chen ZW, Tan F, Kashiwazaki R, Zarif L, et al. Efficacy of orally delivered cochleates containing amphotericin B in a murine model of aspergillosis. *Antimicrob Agents Chemother* 2002;46. <https://doi.org/10.1128/AAC.46.8.2704-2707.2002>.
- [40] Santangelo R, Paderu P, Delmas G, Chen ZW, Mannino R, Zarif L, et al. Efficacy of oral cochleate-amphotericin B in a mouse model of systemic candidiasis. *Antimicrob Agents Chemother* 2000;44. <https://doi.org/10.1128/AAC.44.9.2356-2360.2000>.
- [41] Van Daele R, Spriet I, Wauters J, Maertens J, Mercier T, Hecke S Van, et al. Antifungal drugs: What brings the future? *Med Mycol* 2019;57:328–43. <https://doi.org/10.1093/mmy/myz012>.
- [42] Pfaller MA, Messer SA, Rhomberg PR, Jones RN, Castanheira M. Activity of a long-acting echinocandin, CD101, determined using CLSI and EUCAST reference methods, against *Candida* and *Aspergillus* spp., including echinocandin- and azole-resistant isolates. *Journal of Antimicrobial Chemotherapy* 2016;71. <https://doi.org/10.1093/jac/dkw214>.
- [43] Tóth Z, Forgács L, Locke JB, Kardos G, Nagy F, Kovács R, et al. In vitro activity of rezafungin against common and rare *Candida* species and *Saccharomyces cerevisiae*. *Journal of Antimicrobial Chemotherapy* 2019;74. <https://doi.org/10.1093/jac/dkz390>.
- [44] Pfaller MA, Messer SA, Rhomberg PR, Castanheira M. CD101, a long-acting echinocandin, and comparator antifungal agents tested against a global collection of invasive fungal isolates in the SENTRY 2015 Antifungal Surveillance Program. *Int J Antimicrob Agents* 2017;50. <https://doi.org/10.1016/j.ijantimicag.2017.03.028>.
- [45] Smith HL, Bensman TJ, Mishra S, Li X, Dixon CA, Sheikh J, et al. Regulatory Considerations in the Approval of Rezafungin (Rezzayo) for the Treatment of Candidemia and Invasive Candidiasis in Adults. *J Infect Dis* 2024. <https://doi.org/10.1093/INFDIS/JIAE146>.
- [46] Syed YY. Rezafungin: First Approval. *Drugs* 2023;83. <https://doi.org/10.1007/s40265-023-01891-8>.
- [47] Davis MR, Donnelley MA, Thompson GR. Ibrexafungerp: A novel oral glucan synthase inhibitor. *Med Mycol* 2019;58. <https://doi.org/10.1093/mmy/myz083>.

- [48] Marcos-Zambrano LJ, Gómez-Perosanz M, Escribano P, Bouza E, Guinea J. The novel oral glucan synthase inhibitor SCY-078 shows in vitro activity against sessile and planktonic *Candida* spp. *Journal of Antimicrobial Chemotherapy* 2017;72. <https://doi.org/10.1093/jac/dkx010>.
- [49] Jiménez-Ortigosa C, Perez WB, Angulo D, Borroto-Esoda K, Perlin DS. De novo acquisition of resistance to SCY-078 in *Candida glabrata* involves FKS mutations that both overlap and are distinct from those conferring echinocandin resistance. *Antimicrob Agents Chemother* 2017;61. <https://doi.org/10.1128/AAC.00833-17>.
- [50] Sobel R, Nyirjesy P, Ghannoum MA, Delchev DA, Azie NE, Angulo D, et al. Efficacy and safety of oral ibrexafungerp for the treatment of acute vulvovaginal candidiasis: a global phase 3, randomised, placebo-controlled superiority study (VANISH 306). *BJOG* 2022;129. <https://doi.org/10.1111/1471-0528.16972>.
- [51] Dähn U, Hagenmaier H, Höhne H, König WA, Wolf G, Zähner H. Stoffwechselprodukte von Mikroorganismen - 154. Mitteilung. Nikkomycin, ein neuer Hemmstoff der Chitinsynthese bei Pilzen. *Arch Microbiol* 1976;107. <https://doi.org/10.1007/BF00446834>.
- [52] Li RK, Rinaldi MG. In vitro antifungal activity of nikkomycin Z in combination with fluconazole or itraconazole. *Antimicrob Agents Chemother* 1999;43. <https://doi.org/10.1128/aac.43.6.1401>.
- [53] Nix DE, Swezey RR, Hector R, Galgiani JN. Pharmacokinetics of nikkomycin Z after single rising oral doses. *Antimicrob Agents Chemother* 2009;53. <https://doi.org/10.1128/AAC.01609-08>.
- [54] Paulick MG, Bertozzi CR. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. *Biochemistry* 2008;47:6991–7000.
- [55] Miyazaki M, Horii T, Hata K, Watanabe NA, Nakamoto K, Tanaka K, et al. In vitro activity of E1210, a novel antifungal, against clinically important yeasts and molds. *Antimicrob Agents Chemother* 2011;55. <https://doi.org/10.1128/AAC.00291-11>.
- [56] Pfaller MA, Hata K, Jones RN, Messer SA, Moet GJ, Castanheira M. In vitro activity of a novel broad-spectrum antifungal, e1210, tested against *Candida* spp. as determined by clsi broth microdilution method. *Diagn Microbiol Infect Dis* 2011;71. <https://doi.org/10.1016/j.diagmicrobio.2011.05.001>.
- [57] Pfaller MA, Huband MD, Flamm RK, Bien PA, Castanheira M. In vitro activity of APX001A (Manogepix) and comparator agents against 1,706 fungal isolates collected during an international surveillance program in 2017. *Antimicrob Agents Chemother* 2019;63. <https://doi.org/10.1128/AAC.00840-19>.
- [58] Wiederhold NP, Najvar LK, Fothergill AW, McCarthy DI, Bocanegra R, Olivo M, et al. The investigational agent E1210 is effective in treatment of experimental invasive candidiasis caused by resistant *Candida albicans*. *Antimicrob Agents Chemother* 2015;59. <https://doi.org/10.1128/AAC.03944-14>.

- [59] Rivero-Menendez O, Cuenca-Estrella M, Alastruey-Izquierdo A. In vitro activity of APX001A against rare moulds using EUCAST and CLSI methodologies. *Journal of Antimicrobial Chemotherapy* 2019;74. <https://doi.org/10.1093/jac/dkz022>.
- [60] Castanheira M, Duncanson FP, Diekema DJ, Guarro J, Jones RN, Pfaller MA. Activities of E1210 and comparator agents tested by CLSI and EUCAST broth microdilution methods against *Fusarium* and *Scedosporium* species identified using molecular methods. *Antimicrob Agents Chemother* 2012;56. <https://doi.org/10.1128/AAC.05414-11>.
- [61] Shaw KJ, Ibrahim AS. Fosmanogepix: A review of the first-in-class broad spectrum agent for the treatment of invasive fungal infections. *Journal of Fungi* 2020;6. <https://doi.org/10.3390/jof6040239>.
- [62] Kapoor M, Moloney M, Soltow QA, Pillar CM, Shaw KJ. Evaluation of resistance development to the GWT1 inhibitor manogepix (APX001A) in *Candida* species. *Antimicrob Agents Chemother* 2020;64. <https://doi.org/10.1128/AAC.01387-19>.
- [63] Yamashita K, Miyazaki T, Fukuda Y, Mitsuyama J, Saijo T, Shimamura S, et al. The novel arylamidine T-2307 selectively disrupts yeast mitochondrial function by inhibiting respiratory chain complexes. *Antimicrob Agents Chemother* 2019;63. <https://doi.org/10.1128/AAC.00374-19>.
- [64] Nishikawa H, Yamada E, Shibata T, Uchihashi S, Fan H, Hayakawa H, et al. Uptake of T-2307, a novel arylamidine, in *Candida albicans*. *Journal of Antimicrobial Chemotherapy* 2010;65. <https://doi.org/10.1093/jac/dkq177>.
- [65] Nishikawa H, Sakagami T, Yamada E, Fukuda Y, Hayakawa H, Nomura N, et al. T-2307, a novel arylamidine, is transported into *Candida albicans* by a high-affinity spermine and spermidine carrier regulated by Agp2. *Journal of Antimicrobial Chemotherapy* 2016;71. <https://doi.org/10.1093/jac/dkw095>.
- [66] Shibata T, Takahashi T, Yamada E, Kimura A, Nishikawa H, Hayakawa H, et al. T-2307 causes collapse of mitochondrial membrane potential in yeast. *Antimicrob Agents Chemother* 2012;56. <https://doi.org/10.1128/AAC.05954-11>.
- [67] Singh SB, Liu W, Li X, Chen T, Shafiee A, Card D, et al. Antifungal spectrum, *in vivo* efficacy, and structure-activity relationship of ilicicolin H. *ACS Med Chem Lett* 2012;3. <https://doi.org/10.1021/ml300173e>.
- [68] Oliver JD, Sibley GEM, Beckmann N, Dobb KS, Slater MJ, McEntee L, et al. F901318 represents a novel class of antifungal drug that inhibits dihydroorotate dehydrogenase. *Proc Natl Acad Sci U S A* 2016;113. <https://doi.org/10.1073/pnas.1608304113>.
- [69] Koselny K, Green J, Favazzo L, Glazier VE, Didone L, Ransford S, et al. Antitumor/Antifungal Celecoxib Derivative AR-12 is a Non-Nucleoside Inhibitor of the ANL-Family Adenylating Enzyme Acetyl CoA Synthetase. *ACS Infect Dis* 2016;2. <https://doi.org/10.1021/acsinfecdis.5b00134>.

- [70] Booth L, Roberts JL, Ecroyd H, Tritsch SR, Bavari S, Reid SP, et al. AR-12 Inhibits Multiple Chaperones Concomitant With Stimulating Autophagosome Formation Collectively Preventing Virus Replication. *J Cell Physiol* 2016;231. <https://doi.org/10.1002/jcp.25431>.
- [71] Fernandes CM, Dasilva D, Haranahalli K, Mccarthy JB, Mallamo J, Ojima I, et al. The Future of Antifungal Drug Therapy: Novel Compounds and Targets 2021. <https://doi.org/10.1128/AAC.01719-20>.
- [72] Rauseo AM, Coler-Reilly A, Larson L, Spec A. Hope on the horizon: Novel fungal treatments in development. *Open Forum Infect Dis* 2020;7. <https://doi.org/10.1093/ofid/ofaa016>.
- [73] Kastuhisa U, Tamio H, Hideyo Y, Fumiyo H, Yoshie Y, Ikunoshin K. Biological Properties Of Aureobasidin A, A Cyclic Depsipeptide Antifungal Antibiotic. *J Antibiot (Tokyo)* 1993;46. <https://doi.org/10.7164/antibiotics.46.1414>.
- [74] Wang XH, Guo XJ, Li HY, Gou P. Characteristics of Inositol phosphorylceramide synthase and effects of Aureobasidin a on growth and pathogenicity of botrytis Cinerea. *Journal of General and Applied Microbiology* 2015;61. <https://doi.org/10.2323/jgam.61.108>.
- [75] Heidler SA, Radding JA. The AUR1 gene in *Saccharomyces cerevisiae* encodes dominant resistance to the antifungal agent aureobasidin A (LY295337). *Antimicrob Agents Chemother* 1995;39. <https://doi.org/10.1128/AAC.39.12.2765>.
- [76] Hashida-Okado T, Ogawa A, Endo M, Yasumoto R, Takesako K, Kato I. AUR1, a novel gene conferring aureobasidin resistance on *Saccharomyces cersvisiae*: A study of defective morphologies in Aur1p-depleted cells. *Molecular and General Genetics* 1996;251. <https://doi.org/10.1007/BF02172923>.
- [77] Du L, Haldar S, King JB, Mattes AO, Srivastava S, Wendt KL, et al. Persephacin Is a Broad-Spectrum Antifungal Aureobasidin Metabolite That Overcomes Intrinsic Resistance in *Aspergillus fumigatus*. *J Nat Prod* 2023;86. <https://doi.org/10.1021/acs.jnatprod.3c00382>.
- [78] Nakamura I, Ohsumi K, Takeda S, Katsumata K, Matsumoto S, Akamatsu S, et al. ASP2397 is a novel natural compound that exhibits rapid and potent fungicidal activity against *Aspergillus* species through a specific transporter. *Antimicrob Agents Chemother* 2019;63. <https://doi.org/10.1128/AAC.02689-18>.
- [79] Arendrup MC, Jensen RH, Cuenca-Estrella M. In vitro activity of ASP2397 against *Aspergillus* isolates with or without acquired azole resistance mechanisms. *Antimicrob Agents Chemother* 2016;60. <https://doi.org/10.1128/AAC.02336-15>.
- [80] Nakamura I, Kanasaki R, Yoshikawa K, Furukawa S, Fujie A, Hamamoto H, et al. Discovery of a new antifungal agent ASP2397 using a silkworm model of *Aspergillus fumigatus* infection. *Journal of Antibiotics* 2017;70. <https://doi.org/10.1038/ja.2016.106>.
- [81] Pfaller MA, Messer SA, Georgopapadakou N, Martell LA, Besterman JM, Diekema DJ. Activity of MGCD290, a Hos2 histone deacetylase inhibitor, in combination with azole

- antifungals against opportunistic fungal pathogens. *J Clin Microbiol* 2009;47. <https://doi.org/10.1128/JCM.00618-09>.
- [82] Augenbraun M, Livingston J, Parker R, Lederman S, Chavoustie S, Morgan F, et al. Fluconazole and MGCD290 in vulvo vaginal candidiasis (VVC): Results from a randomized phase II study. IDWeek 2013 Meeting of the Infectious Diseases Society of America, 2013.
- [83] Cowen LE, Lindquist S. Cell biology: Hsp90 potentiates the rapid evolution of new traits: Drug resistance in diverse fungi. *Science* (1979) 2005;309. <https://doi.org/10.1126/science.1118370>.
- [84] Karwa R, Wargo KA. Efungumab: A novel agent in the treatment of invasive candidiasis. *Annals of Pharmacotherapy* 2009;43. <https://doi.org/10.1345/aph.1M218>.
- [85] EMEA. REFUSAL CHMP ASSESSMENT REPORT FOR MYCOGRAB. Evaluation of Medicines for Human Use 2007.
- [86] Lorenz MC, Fink GR. The glyoxylate cycle is required for fungal virulence. *Nature* 2001. <https://doi.org/10.1038/35083594>.
- [87] Bae M, Kim H, Moon K, Nam SJ, Shin J, Oh KB, et al. Mohangamides A and B, new dilactone-tethered pseudo-dimeric peptides inhibiting candida albicans isocitrate lyase. *Org Lett* 2015;17. <https://doi.org/10.1021/ol5037248>.
- [88] Ansari MA, Fatima Z, Ahmad K, Hameed S. Monoterpenoid perillyl alcohol impairs metabolic flexibility of *Candida albicans* by inhibiting glyoxylate cycle. *Biochem Biophys Res Commun* 2018;495. <https://doi.org/10.1016/j.bbrc.2017.11.064>.
- [89] Hoy MJ, Park E, Lee H, Lim WY, Cole DC, DeBouver ND, et al. Structure-Guided Synthesis of FK506 and FK520 Analogs with Increased Selectivity Exhibit *In vivo* Therapeutic Efficacy against *Cryptococcus*. *MBio* 2022:e0104922. <https://doi.org/10.1128/mbio.01049-22>.
- [90] Juvvadi PR, Lamoth F, Steinbach WJ. Calcineurin as a multifunctional regulator: Unraveling novel functions in fungal stress responses, hyphal growth, drug resistance, and pathogenesis. *Fungal Biol Rev* 2014;28. <https://doi.org/10.1016/j.fbr.2014.02.004>.
- [91] Juvvadi PR, Fox D, Bobay BG, Hoy MJ, Gobeil SMC, Venters RA, et al. Harnessing calcineurin-FK506-FKBP12 crystal structures from invasive fungal pathogens to develop antifungal agents. *Nat Commun* 2019;10. <https://doi.org/10.1038/s41467-019-12199-1>.
- [92] Arévalo-Rodríguez M, Pan X, Boeke JD, Heitman J. FKBP12 controls aspartate pathway flux in *Saccharomyces cerevisiae* to prevent toxic intermediate accumulation. *Eukaryot Cell* 2004;3. <https://doi.org/10.1128/EC.3.5.1287-1296.2004>.
- [93] Tsai PW, Chien CY, Yeh YC, Tung L, Chen HF, Chang TH, et al. *Candida albicans* Hom6 is a homoserine dehydrogenase involved in protein synthesis and cell adhesion. *Journal of Microbiology, Immunology and Infection* 2017;50. <https://doi.org/10.1016/j.jmii.2016.03.001>.

- [94] Skwarecki AS, Schielmann M, Martynow D, Kawczyński M, Wiśniewska A, Milewska MJ, et al. Antifungal dipeptides incorporating an inhibitor of homoserine dehydrogenase. *Journal of Peptide Science* 2018;24. <https://doi.org/10.1002/psc.3060>.
- [95] Singh N, Vayer P, Tanwar S, Poyet J-L, Tsaïoun K, Villoutreix BO. Drug discovery and development: introduction to the general public and patient groups. *Frontiers in Drug Discovery* 2023;3. <https://doi.org/10.3389/fddsv.2023.1201419>.
- [96] Fleming A. *Classics in infectious diseases: on the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae* by Alexander Fleming, Reprinted from the *British Journal of Experimental Pathology* 10:226-236, 1929. *Rev Infect Dis* 1980;2.
- [97] Boyd NK, Teng C, Frei CR. Brief Overview of Approaches and Challenges in New Antibiotic Development: A Focus On Drug Repurposing. *Front Cell Infect Microbiol* 2021;11. <https://doi.org/10.3389/fcimb.2021.684515>.
- [98] Mullard A. Parsing clinical success rates. *Nat Rev Drug Discov* 2016;15. <https://doi.org/10.1038/nrd.2016.136>.
- [99] Yang, Wang, Zang, Tang, Li. Cell-Based Assays in High-Throughput Screening for Drug Discovery. *Int J Biotechnol Wellness Ind* 2012. <https://doi.org/10.6000/1927-3037.2012.01.01.02>.
- [100] Carnero A. High throughput screening in drug discovery. *Clinical and Translational Oncology* 2006;8. <https://doi.org/10.1007/s12094-006-0048-2>.
- [101] Matthews H, Hanison J, Nirmalan N. "Omics"-informed drug and biomarker discovery: Opportunities, challenges and future perspectives. *Proteomes* 2016;4. <https://doi.org/10.3390/proteomes4030028>.
- [102] Paananen J, Fortino V. An omics perspective on drug target discovery platforms. *Brief Bioinform* 2020;21. <https://doi.org/10.1093/bib/bbz122>.
- [103] Scapin G. Structural Biology and Drug Discovery. *Curr Pharm Des* 2006;12. <https://doi.org/10.2174/138161206777585201>.
- [104] Congreve M, Murray CW, Blundell TL. Keynote review: Structural biology and drug discovery. *Drug Discov Today* 2005;10. [https://doi.org/10.1016/S1359-6446\(05\)03484-7](https://doi.org/10.1016/S1359-6446(05)03484-7).
- [105] Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. *Molecules* 2020;25. <https://doi.org/10.3390/molecules25061375>.
- [106] Adelusi TI, Oyedele AQK, Boyenle ID, Ogunlana AT, Adeyemi RO, Ukachi CD, et al. Molecular modeling in drug discovery. *Inform Med Unlocked* 2022;29. <https://doi.org/10.1016/j.imu.2022.100880>.
- [107] Anh Vu L, Thi Cam Quyen P, Thuy Huong N. *In silico* Drug Design: Prospective for Drug Lead Discovery. *International Journal of Engineering Science Invention* 2015;4.

- [108] Liu Z, Roberts RA, Lal-Nag M, Chen X, Huang R, Tong W. AI-based language models powering drug discovery and development. *Drug Discov Today* 2021;26. <https://doi.org/10.1016/j.drudis.2021.06.009>.
- [109] Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, Piñeiro Á, et al. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals* 2023;16. <https://doi.org/10.3390/ph16060891>.
- [110] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. *Drug Discov Today* 2021;26. <https://doi.org/10.1016/j.drudis.2020.10.010>.
- [111] Dias O, Pereira R, Gombert AK, Ferreira EC, Rocha I. iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnol J* 2014;9. <https://doi.org/10.1002/biot.201300242>.
- [112] Kim TY, Kim HU, Lee SY. Metabolite-centric approaches for the discovery of antibacterials using genome-scale metabolic networks. *Metab Eng* 2010. <https://doi.org/10.1016/j.ymben.2009.05.004>.
- [113] Raškevičius V, Mikalayeva V, Antanavičiūtė I, Ceslevičienė I, Skeberdis VA, Kairys V, et al. Genome scale metabolic models as tools for drug design and personalized medicine. *PLoS One* 2018. <https://doi.org/10.1371/journal.pone.0190636>.
- [114] Robinson JL, Nielsen J. Anticancer drug discovery through genome-scale metabolic modeling. *Curr Opin Syst Biol* 2017. <https://doi.org/10.1016/j.coisb.2017.05.007>.
- [115] Mienda BS, Salihu R, Adamu A, Idris S. Genome-scale metabolic models as platforms for identification of novel genes as antimicrobial drug targets. *Future Microbiol* 2018. <https://doi.org/10.2217/fmb-2017-0195>.
- [116] Dias O, Saraiva J, Faria C, Ramirez M, Pinto F, Rocha I. iDS372, a Phenotypically Reconciled Model for the Metabolism of *Streptococcus pneumoniae* Strain R6. *Front Microbiol* 2019;10:1283. <https://doi.org/10.3389/fmicb.2019.01283>.
- [117] Abdel-Haleem AM, Hefzi H, Mineta K, Gao X, Gojobori T, Palsson BO, et al. Functional interrogation of *Plasmodium* genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Comput Biol* 2018;14. <https://doi.org/10.1371/journal.pcbi.1005895>.
- [118] Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry* 1999;274. <https://doi.org/10.1074/jbc.274.25.17410>.
- [119] Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol* 2019;20. <https://doi.org/10.1186/s13059-019-1730-3>.
- [120] Zhang C, Hua Q. Applications of genome-scale metabolic models in biotechnology and systems medicine. *Front Physiol* 2016;6. <https://doi.org/10.3389/fphys.2015.00413>.

- [121] Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol* 2010;28. <https://doi.org/10.1038/nbt.1614>.
- [122] Dias O, Rocha M, Ferreira EC, Rocha I. Reconstructing genome-scale metabolic models with *merlin*. *Nucleic Acids Res* 2015;43. <https://doi.org/10.1093/nar/gkv294>.
- [123] Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 2010;5. <https://doi.org/10.1038/nprot.2009.203>.
- [124] Feist AM, Palsson B. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol* 2008;26. <https://doi.org/10.1038/nbt1401>.
- [125] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 1999. <https://doi.org/10.1093/nar/27.1.29>.
- [126] Placzek S, Schomburg I, Chang A, Jeske L, Ulbrich M, Tillack J, et al. BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Res* 2017;45:D380–8. <https://doi.org/10.1093/nar/gkw952>.
- [127] Lagoa D, Faria JP, Liu F, Cunha E, Henry CS, Dias O. TranSyT, the Transport Systems Tracker. *BioRxiv* 2021.
- [128] Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, et al. OptFlux: An open-source software platform for *in silico* metabolic engineering. *BMC Syst Biol* 2010;4. <https://doi.org/10.1186/1752-0509-4-45>.
- [129] Lopes H, Rocha I. Genome-scale modeling of yeast: chronology, applications and critical perspectives. *FEMS Yeast Res* 2017;17. <https://doi.org/10.1093/femsyr/fox050>.
- [130] Jouhten P, Ponomarova O, Gonzalez R, Patil KR. *Saccharomyces cerevisiae* metabolism in ecological context. *FEMS Yeast Res* 2016;16. <https://doi.org/10.1093/femsyr/fow080>.
- [131] Salmaso V, Moro S. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Front Pharmacol* 2018. <https://doi.org/10.3389/fphar.2018.00923>.
- [132] Sethi A, Joshi K, Sasikala K, Alvala M. Molecular Docking in Modern Drug Discovery: Principles and Recent Applications. *Drug Discovery and Development - New Advances*, 2020. <https://doi.org/10.5772/intechopen.85991>.
- [133] Dar AM, Mir S. Molecular Docking: Approaches, Types, Applications and Basic Challenges. *J Anal Bioanal Tech* 2017. <https://doi.org/10.4172/2155-9872.1000356>.
- [134] Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided-Drug Design* 2012. <https://doi.org/10.2174/157340911795677602>.
- [135] Kumar A, Goyal R, Jain S. Docking methodologies and recent advances. *Oncology: Breakthroughs in Research and Practice*, 2016. <https://doi.org/10.4018/978-1-5225-0549-5.ch031>.

- [136] Goodford PJ. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J Med Chem* 1985. <https://doi.org/10.1021/jm00145a002>.
- [137] Levitt DG, Banaszak LJ. POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J Mol Graph* 1992. [https://doi.org/10.1016/0263-7855\(92\)80074-N](https://doi.org/10.1016/0263-7855(92)80074-N).
- [138] Laskowski RA. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 1995. [https://doi.org/10.1016/0263-7855\(95\)00073-9](https://doi.org/10.1016/0263-7855(95)00073-9).
- [139] Brady GP, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 2000. <https://doi.org/10.1023/A:1008124202956>.
- [140] Mezei M. A new method for mapping macromolecular topography. *J Mol Graph Model* 2003. [https://doi.org/10.1016/S1093-3263\(02\)00203-6](https://doi.org/10.1016/S1093-3263(02)00203-6).
- [141] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982. [https://doi.org/10.1016/0022-2836\(82\)90153-X](https://doi.org/10.1016/0022-2836(82)90153-X).
- [142] Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: A system to select “quasi-flexible” ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* 1994. <https://doi.org/10.1007/BF00119865>.
- [143] Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 1997. <https://doi.org/10.1006/jmbi.1997.1203>.
- [144] Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* 2021;61. <https://doi.org/10.1021/acs.jcim.1c00203>.
- [145] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics* 2003. <https://doi.org/10.1002/prot.10465>.
- [146] Abagyan R, Totrov M, Kuznetsov D. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994. <https://doi.org/10.1002/jcc.540150503>.
- [147] Morris GM, Huey R, Weng F, Lindstrom W, Lindstrom W, Sanner MF, Sanner MF, Belew RK, Belew RK, Goodsell DS, Goodsell DS, Olson AJ, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009.
- [148] Knegtel RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J Mol Biol* 1997. <https://doi.org/10.1006/jmbi.1996.0776>.

- [149] Claußen H, Buning C, Rarey M, Lengauer T. FLEXE: Efficient molecular docking considering protein structure variations. *J Mol Biol* 2001. <https://doi.org/10.1006/jmbi.2001.4551>.
- [150] Bender BJ, Gahbauer S, Luttens A, Lyu J, Webb CM, Stein RM, et al. A practical guide to large-scale docking. *Nat Protoc* 2021. <https://doi.org/10.1038/s41596-021-00597-z>.
- [151] Bortolato A, Fanton M, Mason JS, Moro S. Molecular docking methodologies. *Methods in Molecular Biology* 2013. [https://doi.org/10.1007/978-1-62703-17-5\\_13](https://doi.org/10.1007/978-1-62703-17-5_13).
- [152] Prieto-Martínez FD, Arciniega M, Medina-Franco JL, Prieto-Martínez FD, Arciniega M, Medina-Franco JL. Molecular docking: current advances and challenges. *TIP Revista Especializada En Ciencias Químico-Biológicas* 2018;21.
- [153] Choi V, Goyal N. A combinatorial shape matching algorithm for rigid protein docking. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 2004. [https://doi.org/10.1007/978-3-540-27801-6\\_21](https://doi.org/10.1007/978-3-540-27801-6_21).
- [154] Allen WJ, Balias TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, et al. DOCK 6: Impact of new features and current docking performance. *J Comput Chem* 2015. <https://doi.org/10.1002/jcc.23905>.
- [155] Diller DJ, Merz KM. High throughput docking for library design and library prioritization. *Proteins: Structure, Function and Genetics* 2001. [https://doi.org/10.1002/1097-0134\(20010501\)43:2<113::AID-PROT1023>3.0.CO;2-T](https://doi.org/10.1002/1097-0134(20010501)43:2<113::AID-PROT1023>3.0.CO;2-T).
- [156] Burkhard P, Taylor P, Walkinshaw MD. An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex. *J Mol Biol* 1998. <https://doi.org/10.1006/jmbi.1997.1608>.
- [157] Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 2001. <https://doi.org/10.1023/A:1011115820450>.
- [158] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996. <https://doi.org/10.1006/jmbi.1996.0477>.
- [159] Welch W, Ruppert J, Jain AN. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol* 1996. [https://doi.org/10.1016/S1074-5521\(96\)90093-9](https://doi.org/10.1016/S1074-5521(96)90093-9).
- [160] Schneck V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design*, 2000. <https://doi.org/10.1023/A:1008737207775>.
- [161] Bentham Science Publisher BSP. eHiTS: An Innovative Approach to the Docking and Scoring Function Problems. *Curr Protein Pept Sci* 2006. <https://doi.org/10.2174/138920306778559412>.

- [162] Miranker A, Karplus M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins: Structure, Function, and Bioinformatics* 1991. <https://doi.org/10.1002/prot.340110104>.
- [163] Böhm HJ. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 1992. <https://doi.org/10.1007/BF00126217>.
- [164] Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics* 1990. <https://doi.org/10.1002/prot.340080302>.
- [165] McMartin C, Bohacek RS. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* 1997. <https://doi.org/10.1023/A:1007907728892>.
- [166] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998. [https://doi.org/10.1002/\(SICI\)1096-987X\(19981115\)19:14<1639::AID-JCC10>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B).
- [167] Clark KP, Ajay. Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *J Comput Chem* 1995. <https://doi.org/10.1002/jcc.540161004>.
- [168] Taylor JS, Burnett RM. DARWIN: A program for docking flexible molecules. *Proteins: Structure, Function and Genetics* 2000. [https://doi.org/10.1002/1097-0134\(20001101\)41:2<173::AID-PROT30>3.0.CO;2-3](https://doi.org/10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3).
- [169] Guedes IA, Pereira FSS, Dardenne LE. Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Front Pharmacol* 2018;9. <https://doi.org/10.3389/fphar.2018.01089>.
- [170] Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics* 2010. <https://doi.org/10.1039/c0cp00151a>.
- [171] Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *J Chem Inf Model* 2009. <https://doi.org/10.1021/ci800298z>.
- [172] Eldridge MD, Murray CW, Auton TR, Paolini G V., Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997. <https://doi.org/10.1023/A:1007996124545>.
- [173] Li GB, Yang LL, Wang WJ, Li LL, Yang SY. ID-score: A new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J Chem Inf Model* 2013. <https://doi.org/10.1021/ci300493w>.
- [174] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006. <https://doi.org/10.1021/jm051256o>.

- [175] Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 2000. <https://doi.org/10.1006/jmbi.1999.3371>.
- [176] Mooij WTM, Verdonk ML. General and targeted statistical potentials for protein-ligand interactions. *Proteins: Structure, Function and Genetics* 2005. <https://doi.org/10.1002/prot.20588>.
- [177] Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J Med Chem* 1999. <https://doi.org/10.1021/jm980536j>.
- [178] DeWitte RS, Shakhnovich EI. SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J Am Chem Soc* 1996. <https://doi.org/10.1021/ja960751u>.
- [179] Mitchell JBO, Laskowski RA, Alex A, Thornton JM. BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential. *J Comput Chem* 1999. [https://doi.org/10.1002/\(SICI\)1096-987X\(199908\)20:11<1165::AID-JCC7>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-987X(199908)20:11<1165::AID-JCC7>3.0.CO;2-A).
- [180] Oleg T, Arthur J. O. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* 2010.
- [181] Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des* 2019. <https://doi.org/10.1111/cbdd.13388>.
- [182] Hameduh T, Haddad Y, Adam V, Heger Z. Homology modeling in the time of collective and artificial intelligence. *Comput Struct Biotechnol J* 2020;18. <https://doi.org/10.1016/j.csbj.2020.11.007>.
- [183] Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;23. <https://doi.org/10.1093/bioinformatics/btm404>.
- [184] Larsson P, Wallner B, Lindahl E, Elofsson A. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Science* 2008;17. <https://doi.org/10.1110/ps.073344908>.
- [185] Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234. <https://doi.org/10.1006/jmbi.1993.1626>.
- [186] Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology modeling a fast tool for drug discovery: Current perspectives. *Indian J Pharm Sci* 2012;74. <https://doi.org/10.4103/0250-474X.102537>.
- [187] Fiser A, Do RKG, Šali A. Modeling of loops in protein structures. *Protein Science* 2000;9. <https://doi.org/10.1110/ps.9.9.1753>.
- [188] Shen M, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Science* 2006;15. <https://doi.org/10.1110/ps.062416606>.

- [189] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596. <https://doi.org/10.1038/s41586-021-03819-2>.
- [190] Verba K, Gupta M, Azumaya C, Moritz M, Pourmal S, Diallo A, et al. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. *Res Sq* 2021. <https://doi.org/10.21203/rs.3.rs-515215/v1>.
- [191] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024 2024:1–3. <https://doi.org/10.1038/s41586-024-07487-w>.
- [192] Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, et al. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem* 2000;43. <https://doi.org/10.1021/jm000017s>.
- [193] Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a Novel Binding Trench in HIV Integrase. *J Med Chem* 2004;47. <https://doi.org/10.1021/jm0341913>.
- [194] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, Cheruku S, et al. An integrated *in silico* 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. *J Med Chem* 2006;49. <https://doi.org/10.1021/jm0508641>.
- [195] Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem* 2004;47. <https://doi.org/10.1021/jm049756p>.
- [196] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. *Sci Rep* 2017;7. <https://doi.org/10.1038/srep46710>.
- [197] Bai F, Liao S, Gu J, Jiang H, Wang X, Li H. An accurate metalloprotein-specific scoring function and molecular docking program devised by a dynamic sampling and iteration optimization strategy. *J Chem Inf Model* 2015. <https://doi.org/10.1021/ci500647f>.
- [198] Chen D, Menche G, Power TD, Sower L, Peterson JW, Schein CH. Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. *Proteins: Structure, Function and Genetics* 2007. <https://doi.org/10.1002/prot.21249>.
- [199] Parikh HI, Kellogg GE. Intuitive, but not simple: Including explicit water molecules in protein-protein docking simulations improves model quality. *Proteins: Structure, Function and Bioinformatics* 2014. <https://doi.org/10.1002/prot.24466>.
- [200] Kumar A, Zhang KYJ. Investigation on the effect of key water molecules on docking performance in CSARdock exercise. *J Chem Inf Model* 2013. <https://doi.org/10.1021/ci400052w>.

- [201] Lie MA, Thomsen R, Pedersen CNS, Schiøtt B, Christensen MH. Molecular docking with ligand attached water molecules. *J Chem Inf Model* 2011. <https://doi.org/10.1021/ci100510m>.
- [202] Xu M, Lill MA. Induced fit docking, and the use of QM/MM methods in docking. *Drug Discov Today Technol* 2013. <https://doi.org/10.1016/j.ddtec.2013.02.003>.
- [203] Oferkin I V., Katkova E V., Sulimov A V., Kutov DC, Sobolev SI, Voevodin V V., et al. Evaluation of Docking Target Functions by the Comprehensive Investigation of Protein-Ligand Energy Minima. *Adv Bioinformatics* 2015. <https://doi.org/10.1155/2015/126858>.
- [204] Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 2016. <https://doi.org/10.1038/nprot.2016.051>.
- [205] Lemmon G, Meiler J. Towards Ligand Docking Including Explicit Interface Water Molecules. *PLoS One* 2013. <https://doi.org/10.1371/journal.pone.0067536>.
- [206] Sridhar A, Ross GA, Biggin PC. Waterdock 2.0: Water placement prediction for Holostructures with a pymol plugin. *PLoS One* 2017. <https://doi.org/10.1371/journal.pone.0172743>.
- [207] Murphy RB, Repasky MP, Greenwood JR, Tubert-Brohman I, Jerome S, Annabhimoju R, et al. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J Med Chem* 2016. <https://doi.org/10.1021/acs.jmedchem.6b00131>.
- [208] Uehara S, Tanaka S. AutoDock-GIST: Incorporating thermodynamics of active-site water into scoring function for accurate protein-ligand docking. *Molecules* 2016. <https://doi.org/10.3390/molecules21111604>.
- [209] Waszkowycz B, Clark DE, Gancia E. Outstanding challenges in protein-ligand docking and structure-based virtual screening. *Wiley Interdiscip Rev Comput Mol Sci* 2011. <https://doi.org/10.1002/wcms.18>.
- [210] Wisplinghoff H, Bischoff T, Tallent SM, Seifert H, Wenzel RP, Edmond MB. Nosocomial Bloodstream Infections in US Hospitals: Analysis of 24,179 Cases from a Prospective Nationwide Surveillance Study. *Clinical Infectious Diseases* 2004. <https://doi.org/10.1086/421946>.
- [211] Perlroth J, Choi B, Spellberg B. Nosocomial fungal infections: Epidemiology, diagnosis, and treatment. *Med Mycol* 2007. <https://doi.org/10.1080/13693780701218689>.
- [212] Yapar N. Epidemiology and risk factors for invasive candidiasis. *Ther Clin Risk Manag* 2014. <https://doi.org/10.2147/TCRM.S40160>.
- [213] Horn DL, Neofytos D, Anaissie EJ, Fishman JA, Steinbach WJ, Olyaei AJ, et al. Epidemiology and Outcomes of Candidemia in 2019 Patients: Data from the Prospective Antifungal Therapy Alliance Registry. *Clinical Infectious Diseases* 2009. <https://doi.org/10.1086/599039>.

- [214] Galocha M, Pais P, Cavalheiro M, Pereira D, Viana R, Teixeira MC. Divergent approaches to virulence in *C. Albicans* and *C. Glabrata*: Two sides of the same coin. *Int J Mol Sci* 2019. <https://doi.org/10.3390/ijms20092345>.
- [215] Cavalheiro M, Teixeira MC. *Candida* Biofilms: Threats, challenges, and promising strategies. *Front Med (Lausanne)* 2018. <https://doi.org/10.3389/fmed.2018.00028>.
- [216] Hampe IAI, Friedman J, Edgerton M, Morschhäuser J. An acquired mechanism of antifungal drug resistance simultaneously enables *Candida albicans* to escape from intrinsic host defenses. *PLoS Pathog* 2017. <https://doi.org/10.1371/journal.ppat.1006655>.
- [217] Pfaller MA, Rhomberg PR, Messer SA, Jones RN, Castanheira M. Isavuconazole, micafungin, and 8 comparator antifungal agents' susceptibility profiles for common and uncommon opportunistic fungi collected in 2013: Temporal analysis of antifungal drug resistance using CLSI species-specific clinical breakpoints and prop. *Diagn Microbiol Infect Dis* 2015. <https://doi.org/10.1016/j.diagmicrobio.2015.04.008>.
- [218] Cleveland AA, Farley MM, Harrison LH, Stein B, Hollick R, Lockhart SR, et al. Changes in incidence and antifungal drug resistance in candidemia: Results from population-based laboratory surveillance in Atlanta and Baltimore, 2008-2011. *Clinical Infectious Diseases* 2012. <https://doi.org/10.1093/cid/cis697>.
- [219] Arendrup MC, Patterson TF. Multidrug-resistant *Candida*: Epidemiology, molecular mechanisms, and treatment. *Journal of Infectious Diseases* 2017. <https://doi.org/10.1093/infdis/jix131>.
- [220] Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 2019. <https://doi.org/10.1038/s41596-018-0098-2>.
- [221] Dias O, Rocha M, Ferreira EC, Rocha I. Reconstructing high-quality large-scale metabolic models with *merlin*. *Methods in Molecular Biology*, vol. 1716, 2018. [https://doi.org/10.1007/978-1-4939-7528-0\\_1](https://doi.org/10.1007/978-1-4939-7528-0_1).
- [222] Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, et al. Assembly: A resource for assembled genomes at NCBI. *Nucleic Acids Res* 2016. <https://doi.org/10.1093/nar/gkv1226>.
- [223] Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res* 2012. <https://doi.org/10.1093/nar/gkr1178>.
- [224] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018. <https://doi.org/10.1093/molbev/msy096>.
- [225] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.

- [226] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- [227] Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot: The manually annotated section of the UniProt KnowledgeBase. *Methods in Molecular Biology* 2007;406.
- [228] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011;39:W29-37. <https://doi.org/10.1093/nar/gkr367>.
- [229] Stelzer M, Sun J, Kamphans T, Fekete SP, Zeng AP. An extended bioreaction database that significantly improves reconstruction and analysis of genome-scale metabolic networks. *Integrative Biology* 2011. <https://doi.org/10.1039/c1ib00008j>.
- [230] Schomburg I. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* 2002. <https://doi.org/10.1093/nar/30.1.47>.
- [231] Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2014. <https://doi.org/10.1093/nar/gkt1103>.
- [232] Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res* 2007. <https://doi.org/10.1093/nar/gkm259>.
- [233] Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res* 2021;49. <https://doi.org/10.1093/nar/gkaa1004>.
- [234] Santos S, Rocha I. Estimation of biomass composition from genomic and transcriptomic information. *J Integr Bioinform* 2016;13:161–9. <https://doi.org/10.2390/biecoll-jib-2016-285>.
- [235] Verduyn C. Physiology of yeasts in relation to biomass yields. *Antonie Van Leeuwenhoek* 1991. <https://doi.org/10.1007/BF00430373>.
- [236] Mishra P, Park GY, Lakshmanan M, Lee HS, Lee H, Chang MW, et al. Genome-scale metabolic modeling and *in silico* analysis of lipid accumulating yeast *Candida tropicalis* for dicarboxylic acid production. *Biotechnol Bioeng* 2016. <https://doi.org/10.1002/bit.25955>.
- [237] Xu N, Liu L, Zou W, Liu J, Hua Q, Chen J. Reconstruction and analysis of the genome-scale metabolic network of *Candida glabrata*. *Mol Biosyst* 2013;9. <https://doi.org/10.1039/c2mb25311a>.
- [238] Mo ML, Palsson B, Herrgård MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 2009. <https://doi.org/10.1186/1752-0509-3-37>.

- [239] Xu D, Jiang B, Ketela T, Lemieux S, Veillette K, Martel N, et al. Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog* 2007. <https://doi.org/10.1371/journal.ppat.0030092>.
- [240] Greenberg JR, Price NP, Oliver RP, Sherman F, Rustchenko E. *Candida albicans* SOU1 encodes a sorbose reductase required for L-sorbose utilization. *Yeast* 2005. <https://doi.org/10.1002/yea.1282>.
- [241] Gao J, Wang H, Li Z, Wong AH-H, Wang Y-Z, Guo Y, et al. *Candida albicans* gains azole resistance by altering sphingolipid composition. *Nat Commun* 2018;9:4495. <https://doi.org/10.1038/s41467-018-06944-1>.
- [242] Hüdig M, Maier A, Scherrers I, Seidel L, Jansen EEW, Mettler-Altmann T, et al. Plants Possess a Cyclic Mitochondrial Metabolic Pathway similar to the Mammalian Metabolic Repair Mechanism Involving Malate Dehydrogenase and L-2-Hydroxyglutarate Dehydrogenase. *Plant Cell Physiol* 2014. <https://doi.org/10.1093/pcp/pcv108>.
- [243] Rzem R, Veiga-Da-Cunha M, Noël G, Goffette S, Nassogne MC, Tabarki B, et al. A gene encoding a putative FAD-dependent L-2-hydroxyglutarate dehydrogenase is mutated in L-2-hydroxyglutaric aciduria. *Proc Natl Acad Sci U S A* 2004. <https://doi.org/10.1073/pnas.0404840101>.
- [244] Wendland J, Schaub Y, Walther A. N-acetylglucosamine utilization by *Saccharomyces cerevisiae* based on expression of *Candida albicans* NAG genes. *Appl Environ Microbiol* 2009. <https://doi.org/10.1128/AEM.00053-09>.
- [245] Vesely EM, Williams RB, Konopka JB, Lorenz MC. N -Acetylglucosamine Metabolism Promotes Survival of *Candida albicans* in the Phagosome . *MSphere* 2017. <https://doi.org/10.1128/msphere.00357-17>.
- [246] Molla G, Motteran L, Piubelli L, Pilone MS, Pollegioni L. Regulation of D-amino acid oxidase expression in the yeast *Rhodotorula gracilis*. *Yeast* 2003. <https://doi.org/10.1002/yea.1023>.
- [247] Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>. *BioinfogpCnbCsicEs/Tools/Venny/IndexHtml* 2007.
- [248] Santos ST. Development of computational methods for the determination of biomass composition and evaluation of its impact in genome-scale models predictions. Universidade do Minho, 2013.
- [249] Brondz I, Olsen I. Multivariate analyses of cellular carbohydrates and fatty acids of *Candida albicans*, *Torulopsis glabrata*, and *Saccharomyces cerevisiae*. *J Clin Microbiol* 1990. <https://doi.org/10.1128/jcm.28.8.1854-1857.1990>.
- [250] Ghannoum MA, Swairjo I, Soll DR. Variation in lipid and sterol contents in *Candida albicans* white and opaque phenotypes. *Med Mycol* 1990;28. <https://doi.org/10.1080/02681219080000151>.

- [251] Abu-Elteen KH, Whittaker PA. Effect of sub-inhibitory concentration of chlorhexidine on lipid and sterol composition of *Candida albicans*. *Mycopathologia* 1997. <https://doi.org/10.1023/A:1006852207366>.
- [252] Mayatepek E, Herz A, Leichsenring M, Kappe R. Fatty acid analysis of different *Candida*, species by capillary column gas-liquid chromatography. *Mycoses* 2009;34. <https://doi.org/10.1111/j.1439-0507.1991.tb00619.x>.
- [253] Xavier JC, Patil KR, Rocha I. Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metab Eng* 2017;39. <https://doi.org/10.1016/j.ymben.2016.12.002>.
- [254] Verduyn C, Postma E, Scheffers WA, Van Dijken JP. Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 1990. <https://doi.org/10.1099/00221287-136-3-395>.
- [255] CBS-KNAW Collections n.d. <http://www.cbs.knaw.nl/Collections> (accessed July 24, 2020).
- [256] Huang X, Chen X, He Y, Yu X, Li S, Gao N, et al. Mitochondrial complex I bridges a connection between regulation of carbon flexibility and gastrointestinal commensalism in the human fungal pathogen *Candida albicans*. *PLoS Pathog* 2017. <https://doi.org/10.1371/journal.ppat.1006414>.
- [257] Brown V, Sexton JA, Johnston M. A glucose sensor in *Candida albicans*. *Eukaryot Cell* 2006. <https://doi.org/10.1128/EC.00186-06>.
- [258] Guan G, Wang H, Liang W, Cao C, Tao L, Naseem S, et al. The mitochondrial protein Mcu1 plays important roles in carbon source utilization, filamentation, and virulence in *Candida albicans*. *Fungal Genetics and Biology* 2015. <https://doi.org/10.1016/j.fgb.2015.01.006>.
- [259] Choudary P V., Ramananda Rao G. Molecular analysis of inorganic nitrogen assimilation in yeasts. *Arch Microbiol* 1984. <https://doi.org/10.1007/BF00402116>.
- [260] Rozpędowska E, Galafassi S, Johansson L, Hagman A, Piškur J, Compagno C. *Candida albicans*- a pre-whole genome duplication yeast - is predominantly aerobic and a poor ethanol producer. *FEMS Yeast Res* 2011. <https://doi.org/10.1111/j.1567-1364.2010.00715.x>.
- [261] Askew C, Sellam A, Epp E, Hogues H, Mullick A, Nantel A, et al. Transcriptional regulation of carbohydrate metabolism in the human pathogen *Candida albicans*. *PLoS Pathog* 2009. <https://doi.org/10.1371/journal.ppat.1000612>.
- [262] Dumitru R, Hornby JM, Nickerson KW. Defined anaerobic growth medium for studying *Candida albicans* basic biology and resistance to eight antifungal drugs. *Antimicrob Agents Chemother* 2004;48. <https://doi.org/10.1128/AAC.48.7.2350-2354.2004>.
- [263] Jouhten P, Rintala E, Huuskonen A, Tamminen A, Toivari M, Wiebe M, et al. Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Syst Biol* 2008. <https://doi.org/10.1186/1752-0509-2-60>.

- [264] Omeara TR, Veri AO, Ketela T, Jiang B, Roemer T, Cowen LE. Global analysis of fungal morphology exposes mechanisms of host cell escape. *Nat Commun* 2015. <https://doi.org/10.1038/ncomms7741>.
- [265] Wishart DS. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res* 2006. <https://doi.org/10.1093/nar/gkj067>.
- [266] Robbins, N., Wright, G.D., and Cowen LE. Antifungal Drugs: The Current Armamentarium and Development of New Agents. *Microbiology Spectrum* 4 2016;Volume 4. <https://doi.org/10.1128/microbiolspec.FUNK-0002-2016>.
- [267] Lupetti A, Danesi R, Campa M, Tacca M Del, Kelly S. Molecular basis of resistance to azole antifungals. *Trends Mol Med* 2002. [https://doi.org/10.1016/S1471-4914\(02\)02280-3](https://doi.org/10.1016/S1471-4914(02)02280-3).
- [268] Rodrigues ML. The multifunctional fungal ergosterol. *MBio* 2018. <https://doi.org/10.1128/mBio.01755-18>.
- [269] Nixon GL, Moss DM, Shone AE, Lalloo DG, Fisher N, O'neill PM, et al. Antimalarial pharmacology and therapeutics of atovaquone. *Journal of Antimicrobial Chemotherapy* 2013. <https://doi.org/10.1093/jac/dks504>.
- [270] Iliades P, Berglez J, Meshnick S, Macreadie I. Promoter strength of folic acid synthesis genes affects sulfa drug resistance in *Saccharomyces cerevisiae*. *Microbial Drug Resistance* 2003. <https://doi.org/10.1089/107662903322286454>.
- [271] Eldesouky HE, Mayhoub A, Hazbun TR, Seleema MN. Reversal of azole resistance in *Candida albicans* by sulfa antibacterial drugs. *Antimicrob Agents Chemother* 2018. <https://doi.org/10.1128/AAC.00701-17>.
- [272] Garcia-Effron G, Lee S, Park S, Cleary JD, Perlin DS. Effect of *Candida glabrata* FKS1 and FKS2 mutations on echinocandin sensitivity and kinetics of 1,3- $\beta$ -D-glucan synthase: Implication for the existing susceptibility breakpoint. *Antimicrob Agents Chemother* 2009. <https://doi.org/10.1128/AAC.00443-09>.
- [273] Walker LA, Gow NAR, Munro CA. Fungal echinocandin resistance. *Fungal Genetics and Biology* 2010. <https://doi.org/10.1016/j.fgb.2009.09.003>.
- [274] Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin JA. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol* 2012. <https://doi.org/10.1016/j.tim.2011.12.004>.
- [275] Silva S, Negri M, Henriques M, Oliveira R, Williams DW, Azeredo J. *Candida glabrata*, *Candida parapsilosis* and *Candida tropicalis*: Biology, epidemiology, pathogenicity and antifungal resistance. *FEMS Microbiol Rev* 2012;36:288–305. <https://doi.org/10.1111/j.1574-6976.2011.00278.x>.
- [276] Fridkin SK. The changing face of fungal infections in health care settings. *Clinical Infectious Diseases* 2005;41:1455–60. <https://doi.org/10.1086/497138>.
- [277] Trofa D, Gácsér A, Nosanchuk JD. *Candida parapsilosis*, an emerging fungal pathogen. *Clin Microbiol Rev* 2008;21:606–25. <https://doi.org/10.1128/CMR.00013-08>.

- [278] Castanheira M, Deshpande LM, Messer SA, Rhomberg PR, Pfaller MA. Analysis of global antifungal surveillance results reveals predominance of Erg11 Y132F alteration among azole-resistant *Candida parapsilosis* and *Candida tropicalis* and country-specific isolate dissemination. *Int J Antimicrob Agents* 2020;55:105799. <https://doi.org/10.1016/j.ijantimicag.2019.09.003>.
- [279] Tsui CKM, Daniel HM, Robert V, Meyer W. Re-examining the phylogeny of clinically relevant *Candida* species and allied genera based on multigene analyses. *FEMS Yeast Res* 2008;8:651–9. <https://doi.org/10.1111/j.1567-1364.2007.00342.x>.
- [280] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28. <https://doi.org/10.1093/nar/28.1.27>.
- [281] Flamholz A, Noor E, Bar-Even A, Milo R. EQuilibrator - The biochemical thermodynamics calculator. *Nucleic Acids Res* 2012;40. <https://doi.org/10.1093/nar/gkr874>.
- [282] Degtyarenko K, De matos P, Ennis M, Hastings J, Zbinden M, Mcnaught A, et al. ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008;36. <https://doi.org/10.1093/nar/gkm791>.
- [283] Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* 2002;30:47–9. <https://doi.org/10.1093/nar/30.1.47>.
- [284] Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 1994. <https://doi.org/10.1128/aem.60.10.3724-3731.1994>.
- [285] Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, et al. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J Bacteriol* 1999;181:6679–88. <https://doi.org/10.1128/jb.181.21.6679-6688.1999>.
- [286] Pfaller M, Riley J. Effects of fluconazole on the sterol and carbohydrate composition of four species of *Candida*. *European Journal of Clinical Microbiology & Infectious Diseases* 1992;11:152–6. <https://doi.org/10.1007/BF01967067>.
- [287] Ghannoum MA, Abu-Elteen KH. Pathogenicity determinants of *Candida*. *Mycoses* 1990;33:265–82. <https://doi.org/10.1111/myc.1990.33.6.265>.
- [288] Viana R, Dias O, Lagoa D, Galocha M, Rocha I, Teixeira MC. Genome-Scale Metabolic Model of the Human Pathogen *Candida albicans*: A Promising Platform for Drug Target Prediction. *J Fungi* 2020;6. <https://doi.org/10.3390/jof6030171>.
- [289] Endoh R, Horiyama M, Ohkuma M. D-fructose assimilation and fermentation by yeasts belonging to saccharomycetes: Rediscovery of universal phenotypes and elucidation of fructophilic behaviors in *Ambrosiozyma platypodis* and *Cyberlindnera americana*. *Microorganisms* 2021. <https://doi.org/10.3390/microorganisms9040758>.
- [290] Fell JW, Meyer SA. Systematics of yeast species in the *Candida parapsilosis* group. *Mycopathol Mycol Appl* 1967. <https://doi.org/10.1007/BF02049795>.

- [291] Tambosis E, Atkins BL, Capizzi T, Gottlieb T. Rapid and cost-effective identification of *Candida* species using multipoint inoculation of CHROMagar *Candida* media, cycloheximide sensitivity and carbohydrate assimilation tests. *Pathology* 2003. <https://doi.org/10.1097/01268031-200335020-00010>.
- [292] Devadas SM, Ballal M, Prakash PY, Hande MH, Bhat G V., Mohandas V. Auxanographic carbohydrate assimilation method for large scale yeast identification. *Journal of Clinical and Diagnostic Research* 2017;11:DC01–3. <https://doi.org/10.7860/JCDR/2017/25967.9653>.
- [293] Deorukhkar, Roushani S. Identification of *Candida* species: conventional methods in the era of molecular diagnosis. *Annals of Microbiology and Immunology* 2018;1:1002.
- [294] Turner SA, Ma Q, Ola M, Martinez de San Vicente K, Butler G. Dal81 Regulates Expression of Arginine Metabolism Genes in *Candida parapsilosis* . *MSphere* 2018;3:1–13. <https://doi.org/10.1128/msphere.00028-18>.
- [295] Masaki K, Mizukure T, Kakizono D, Fujihara K, Fujii T, Mukai N. New urethanase from the yeast *Candida parapsilosis*. *J Biosci Bioeng* 2020;130:115–20. <https://doi.org/10.1016/j.jbiosc.2020.03.005>.
- [296] Lawrence MC, Iliades P, Fernley RT, Berglez J, Pilling PA, Macreadie IG. The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8- dihydropterin pyrophosphokinase/dihydropteroate synthase of *Saccharomyces cerevisiae*. *J Mol Biol* 2005;348. <https://doi.org/10.1016/j.jmb.2005.03.021>.
- [297] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46. <https://doi.org/10.1093/nar/gkx1037>.
- [298] Navarro-Martínez MD, Cabezas-Herrera J, Rodríguez-López JN. Antifolates as antimycotics?. Connection between the folic acid cycle and the ergosterol biosynthesis pathway in *Candida albicans*. *Int J Antimicrob Agents* 2006. <https://doi.org/10.1016/j.ijantimicag.2006.07.012>.
- [299] Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. *Candida auris* sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. *Microbiol Immunol* 2009;53:41–4. <https://doi.org/10.1111/j.1348-0421.2008.00083.x>.
- [300] Ademe M, Girma F. *Candida auris*: From multidrug resistance to pan-resistant strains. *Infect Drug Resist* 2020;13:1287–94. <https://doi.org/10.2147/IDR.S249864>.
- [301] Kim MN, Shin JH, Sung H, Lee K, Kim EC, Ryoo N, et al. *Candida haemulonii* and closely related species at 5 university hospitals in Korea: Identification, antifungal susceptibility, and clinical features. *Clinical Infectious Diseases* 2009;48:57–61. <https://doi.org/10.1086/597108>.
- [302] Lass-Flörl C. The changing face of epidemiology of invasive fungal disease in Europe. *Mycoses* 2009;52:197–205. <https://doi.org/10.1111/j.1439-0507.2009.01691.x>.

- [303] Viana R, Couceiro D, Carreiro T, Dias O, Rocha I, Teixeira MC. A Genome-Scale Metabolic Model for the Human Pathogen *Candida Parapsilosis* and Early Identification of Putative Novel Antifungal Drug Targets. *Genes (Basel)* 2022;13(2). <https://doi.org/DOI:10.3390/genes13020303>.
- [304] Freeman, Starr DA, O'connor. 乳鼠心肌提取 HHS Public Access. *Physiol Behav* 2016;176:139–48. <https://doi.org/10.1038/s41596-018-0098-2>.Creation.
- [305] Muñoz JF, Gade L, Chow NA, Loparev VN, Juieng P, Berkow EL, et al. Genomic insights into multidrug-resistance, mating and virulence in *Candida auris* and related emerging species. *Nat Commun* 2018;9. <https://doi.org/10.1038/s41467-018-07779-6>.
- [306] Lee WG, Shin JH, Uh Y, Kang MG, Kim SH, Park KH, et al. First three reported cases of nosocomial fungemia caused by *Candida auris*. *J Clin Microbiol* 2011;49:3139–42. <https://doi.org/10.1128/JCM.00319-11>.
- [307] Tamanna P, Saifali A, Jayati C, Rishika S, Pankaj C, Ravi Ranjan KN, et al. Comprehensive characterization of *Candida* isolates in a given geographical area for the determination of prevalence and drug sensitivity. *MedRxiv* 2022.
- [308] Cendejas-Bueno E, Kolecka A, Alastruey-Izquierdo A, Theelen B, Groenewald M, Kostrzewa M, et al. Reclassification of the *Candida haemulonii* complex as *Candida haemulonii* (*C. haemulonii* group I), *C. duobushaemulonii* sp. nov. (*C. haemulonii* group II), and *C. haemulonii* var. *vulnera* var. nov.: three multiresistant human pathogenic yeasts. *J Clin Microbiol* 2012;50:3641–51. <https://doi.org/10.1128/JCM.02248-12>.
- [309] Ninan MM, Sahni RD, Chacko B, Balaji V, Michael JS. *Candida auris*: Clinical profile, diagnostic challenge and susceptibility pattern: Experience from a tertiary-care centre in South India. *J Glob Antimicrob Resist* 2020;21. <https://doi.org/10.1016/j.jgar.2019.10.018>.
- [310] Kubota T, Tanaka Y, Hiraga K, Inui M, Yukawa H. Characterization of shikimate dehydrogenase homologues of *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* 2013;97:8139–49. <https://doi.org/10.1007/S00253-012-4659-Y>.
- [311] van Schijndel JWPM, Vollenbroek EGM, Wever R. The chloroperoxidase from the fungus *Curvularia inaequalis*; a novel vanadium enzyme. *Biochim Biophys Acta* 1993;1161:249–56. [https://doi.org/10.1016/0167-4838\(93\)90221-C](https://doi.org/10.1016/0167-4838(93)90221-C).
- [312] Merkens H, Kappl R, Jakob RP, Schmid FX, Fetzner S. Quercetinase QueD of *Streptomyces* sp. FLA, a monocupin dioxygenase with a preference for nickel and cobalt. *Biochemistry* 2008;47:12185–96. [https://doi.org/10.1021/BI801398X/ASSET/IMAGES/MEDIUM/BI-2008-01398X\\_0007.GIF](https://doi.org/10.1021/BI801398X/ASSET/IMAGES/MEDIUM/BI-2008-01398X_0007.GIF).
- [313] Peracchi A, Veiga-Da-Cunha M, Kuhara T, Ellens KW, Paczia N, Stroobant V, et al. Nit1 is a metabolite repair enzyme that hydrolyzes deaminated glutathione. *Proc Natl Acad Sci U S A* 2017;114:E3233–42. [https://doi.org/10.1073/PNAS.1613736114/SUPPL\\_FILE/PNAS.1613736114.SAPP.PDF](https://doi.org/10.1073/PNAS.1613736114/SUPPL_FILE/PNAS.1613736114.SAPP.PDF).

- [314] Lin X, Heitman J. The biology of the *Cryptococcus neoformans* species complex. *Annu Rev Microbiol* 2006;60. <https://doi.org/10.1146/annurev.micro.60.080805.142102>.
- [315] Del Valle L, Piña-Oviedo S. HIV disorders of the brain; pathology and pathogenesis. *Frontiers in Bioscience* 2006;11. <https://doi.org/10.2741/1830>.
- [316] Park BJ, Wannemuehler KA, Marston BJ, Govender N, Pappas PG, Chiller TM. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS* 2009;23. <https://doi.org/10.1097/QAD.0b013e328322ffac>.
- [317] Ventura Aguiar P, Lopes V, Martins LS, Santos J, Almeida M, Pedroso S, et al. Cryptococcal infection in non-HIV immunosuppressed patients - Three case reports in a nephrology setting. *Med Mycol Case Rep* 2014;3. <https://doi.org/10.1016/j.mmcr.2013.11.003>.
- [318] Hope W, Stone NRH, Johnson A, McEntee L, Farrington N, Santoro-Castelazo A, et al. Fluconazole monotherapy is a suboptimal option for initial treatment of cryptococcal meningitis because of emergence of resistance. *MBio* 2019;10. <https://doi.org/10.1128/mBio.02575-19>.
- [319] Coelho C, Casadevall A. Cryptococcal therapies and drug targets: the old, the new and the promising. *Cell Microbiol* 2016;18. <https://doi.org/10.1111/cmi.12590>.
- [320] Williamson PR. Laccase and melanin in the pathogenesis of *Cryptococcus neoformans*. *Front Biosci* 1997;2. <https://doi.org/10.2741/A231>.
- [321] Wang Y, Aisen P, Casadevall A. *Cryptococcus neoformans* melanin and virulence: Mechanism of action. *Infect Immun* 1995;63. <https://doi.org/10.1128/iai.63.8.3131-3136.1995>.
- [322] Nosanchuk JD, Casadevall A. Impact of melanin on microbial virulence and clinical resistance to antimicrobial compounds. *Antimicrob Agents Chemother* 2006;50. <https://doi.org/10.1128/AAC.00545-06>.
- [323] Van Duin D, Casadevall A, Nosanchuk JD. Melanization of *Cryptococcus neoformans* and *Histoplasma capsulatum* reduces their susceptibilities to amphotericin B and caspofungin. *Antimicrob Agents Chemother* 2002;46. <https://doi.org/10.1128/AAC.46.11.3394-3400.2002>.
- [324] Velagapudi R, Hsueh YP, Geunes-Boyer S, Wright JR, Heitman J. Spores as infectious propagules of *Cryptococcus neoformans*. *Infect Immun* 2009;77. <https://doi.org/10.1128/IAI.00542-09>.
- [325] Botts MR, Hull CM. Dueling in the lung: How *Cryptococcus* spores race the host for survival. *Curr Opin Microbiol* 2010;13. <https://doi.org/10.1016/j.mib.2010.05.003>.
- [326] WHO. WHO fungal priority pathogens list to guide research, development and public health action. vol. 1. 2022.
- [327] Wang Y, Wear M, Kohli G, Vij R, Giamberardino C, Shah A, et al. Inositol Metabolism Regulates Capsule Structure and Virulence in the Human Pathogen *Cryptococcus neoformans*. *MBio* 2021;12. <https://doi.org/10.1128/mBio.02790-21>.

- [328] Levitz SM. *Cryptococcus neoformans* by Casadevall, Arturo & Perfect, John R. (1998) ASM Press, Washington, DC. Hardcover. 542 pp. \$89.95. (ASM Member price: \$79.95). *Med Mycol* 2008;37. <https://doi.org/10.1111/j.1365-280x.1999.00238.x>.
- [329] Hommel B, Sturny-Leclère A, Volant S, Veluppillai N, Duchateau M, Yu CH, et al. *Cryptococcus neoformans* resists to drastic conditions by switching to viable but non-culturable cell phenotype. *PLoS Pathog* 2019;15. <https://doi.org/10.1371/journal.ppat.1007945>.
- [330] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2014;12. <https://doi.org/10.1038/nmeth.3176>.
- [331] Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2023;51. <https://doi.org/10.1093/nar/gkac1052>.
- [332] Amos B, Aurrecochea C, Barba M, Barreto A, Basenko EY, Bazant W, et al. VEuPathDB: The eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res* 2022;50. <https://doi.org/10.1093/nar/gkab929>.
- [333] Bansal P, Morgat A, Axelsen KB, Muthukrishnan V, Coudert E, Aimo L, et al. Rhea, the reaction knowledgebase in 2022. *Nucleic Acids Res* 2022;50. <https://doi.org/10.1093/nar/gkab1016>.
- [334] Thumhuri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res* 2022;50. <https://doi.org/10.1093/nar/gkac278>.
- [335] Mukaremera L, Lee KK, Wagener J, Wiesner DL, Gow NAR, Nielsen K. Titan cell production in *Cryptococcus neoformans* reshapes the cell wall and capsule composition during infection. *Cell Surface* 2018;1. <https://doi.org/10.1016/j.tcs.2017.12.001>.
- [336] Ghannoum MA, Spellberg BJ, Ibrahim AS, Ritchie JA, Currie B, Spitzer ED, et al. Sterol composition of *Cryptococcus neoformans* in the presence and absence of fluconazole. *Antimicrob Agents Chemother* 1994;38. <https://doi.org/10.1128/AAC.38.9.2029>.
- [337] Kaneko H, Hosohara M, Tanaka M, Itoh T. Lipid composition of 30 species of yeast. *Lipids* 1976;11. <https://doi.org/10.1007/BF02532989>.
- [338] Rawat DS, Upreti HB, Das SK. Lipid composition of *Cryptococcus neoformans*. *Microbiologica* 1984;7.
- [339] Viana R, Carreiro T, Couceiro D, Dias O, Rocha I, Teixeira MC. Metabolic reconstruction of the human pathogen *Candida auris*: using a cross-species approach for drug target prediction. *FEMS Yeast Res* 2023;23. <https://doi.org/10.1093/femsyr/foad045>.
- [340] Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C. Jvarkit: An interactive Venn diagram viewer. *BMC Bioinformatics* 2014;15. <https://doi.org/10.1186/1471-2105-15-293>.

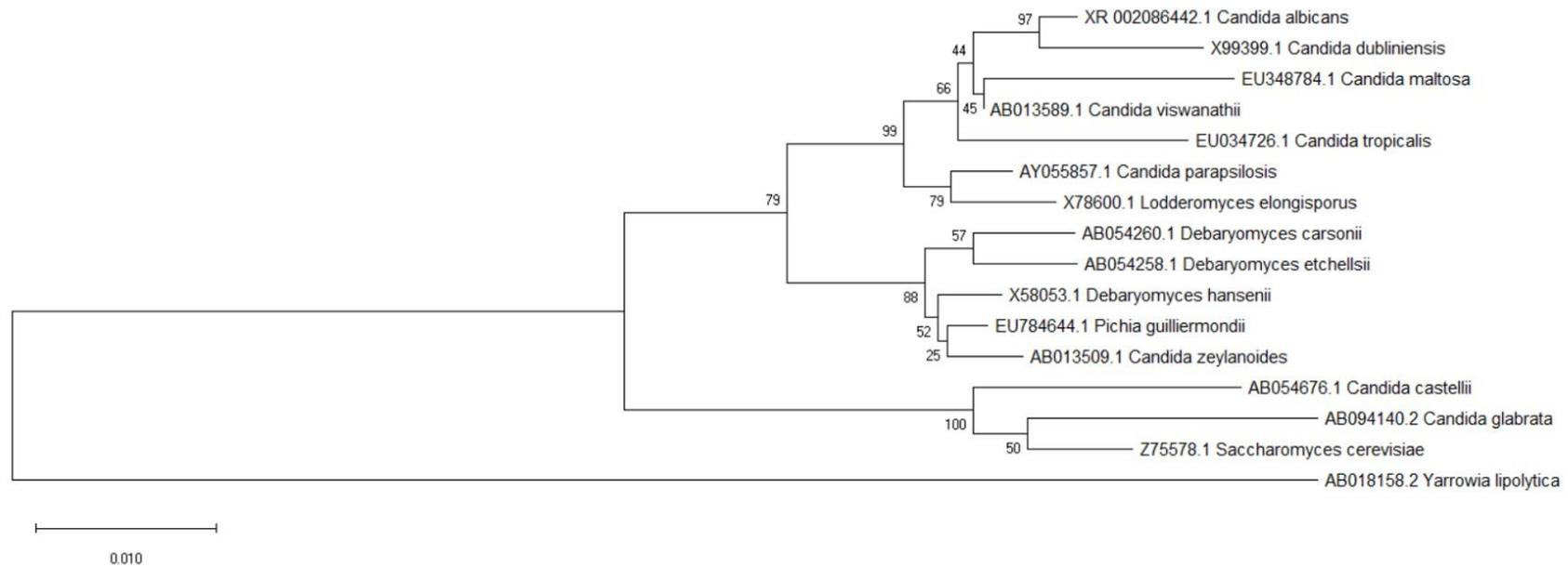
- [341] Fonseca C, Romão R, Rodrigues De Sousa H, Hahn-Hägerdal B, Spencer-Martins I. L-Arabinose transport and catabolism in yeast. *FEBS Journal* 2007;274. <https://doi.org/10.1111/j.1742-4658.2007.05892.x>.
- [342] Bettiga M, Bengtsson O, Hahn-Hägerdal B, Gorwa-Grauslund MF. Arabinose and xylose fermentation by recombinant *Saccharomyces cerevisiae* expressing a fungal pentose utilization pathway. *Microb Cell Fact* 2009;8. <https://doi.org/10.1186/1475-2859-8-40>.
- [343] Gerstein AC, Jackson KM, McDonald TR, Wang Y, Lueck BD, Bohjanen S, et al. Identification of pathogen genomic differences that impact human immune response and disease during *Cryptococcus neoformans* infection. *MBio* 2019;10. <https://doi.org/10.1128/mBio.01440-19>.
- [344] Christensson B, Roboz J. Arabinitol enantiomers in cerebrospinal fluid. *J Neurol Sci* 1991;105. [https://doi.org/10.1016/0022-510X\(91\)90150-6](https://doi.org/10.1016/0022-510X(91)90150-6).
- [345] Dbouk NH, Covington MB, Nguyen K, Chandrasekaran S. Increase of reactive oxygen species contributes to growth inhibition by fluconazole in *Cryptococcus neoformans*. *BMC Microbiol* 2019;19. <https://doi.org/10.1186/s12866-019-1606-4>.
- [346] Carlson T, Lupinacci E, Moseley K, Chandrasekaran S. Effects of environmental factors on sensitivity of *Cryptococcus neoformans* to fluconazole and amphotericin B. *FEMS Microbiol Lett* 2021;368. <https://doi.org/10.1093/femsle/fnab040>.
- [347] Van Hauwenhuyse F, Fiori A, Van Dijck P, Van Hauwenhuyse F, Fiori A, Van Dijck P. Ascorbic acid inhibition of *Candida albicans* Hsp90-mediated morphogenesis occurs via the transcriptional regulator Upc2. *Eukaryot Cell* 2014;13. <https://doi.org/10.1128/EC.00096-14>.
- [348] Kuivanen J, Sugai-Guérios MH, Arvas M, Richard P. A novel pathway for fungal D-glucuronate catabolism contains an L-idonate forming 2-keto-L-gulonate reductase. *Sci Rep* 2016;6. <https://doi.org/10.1038/srep26329>.
- [349] Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13. <https://doi.org/10.1101/gr.1224503>.
- [350] Bahia FM, De Almeida GC, De Andrade LP, Campos CG, Queiroz LR, Da Silva RLV, et al. Rhamnolipids production from sucrose by engineered *Saccharomyces cerevisiae*. *Sci Rep* 2018;8. <https://doi.org/10.1038/s41598-018-21230-2>.
- [351] Azghani AO, Miller EJ, Peterson BT. Virulence factors from *Pseudomonas aeruginosa* increase lung epithelial permeability. *Lung* 2000;178. <https://doi.org/10.1007/s004080000031>.
- [352] Zulianello L, Canard C, Köhler T, Caille D, Lacroix JS, Meda P. Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. *Infect Immun* 2006;74. <https://doi.org/10.1128/IAI.01772-05>.
- [353] McClure CD, Schiller NL. Inhibition of macrophage phagocytosis by *Pseudomonas aeruginosa* rhamnolipids *in vitro* and *in vivo*. *Curr Microbiol* 1996;33. <https://doi.org/10.1007/s002849900084>.

- [354] Volkens G, Palm GJ, Weiss MS, Wright GD, Hinrichs W. Structural basis for a new tetracycline resistance mechanism relying on the TetX monooxygenase. *FEBS Lett* 2011;585. <https://doi.org/10.1016/j.febslet.2011.03.012>.
- [355] Yang W, Moore IF, Koteva KP, Bareich DC, Hughes DW, Wright GD. TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *Journal of Biological Chemistry* 2004;279. <https://doi.org/10.1074/jbc.M409573200>.
- [356] Li C, Lev S, Saiardi A, Desmarini D, Sorrell TC, Djordjevic JT. Identification of a major IP5 kinase in *Cryptococcus neoformans* confirms that PP-IP5/IP7, not IP6, is essential for virulence. *Sci Rep* 2016;6. <https://doi.org/10.1038/srep23927>.
- [357] Fu MS, Coelho C, De Leon-Rodriguez CM, Rossi DCP, Camacho E, Jung EH, et al. *Cryptococcus neoformans* urease affects the outcome of intracellular pathogenesis by modulating phagolysosomal pH. *PLoS Pathog* 2018;14. <https://doi.org/10.1371/journal.ppat.1007144>.
- [358] Lee IR, Yang L, Sebetso G, Allen R, Doan THN, Blundell R, et al. Characterization of the Complete Uric Acid Degradation Pathway in the Fungal Pathogen *Cryptococcus neoformans*. *PLoS One* 2013;8. <https://doi.org/10.1371/journal.pone.0064292>.
- [359] Russel Lee I, Chow EWL, Morrow CA, Djordjevic JT, Fraser JA. Nitrogen metabolite repression of metabolism and virulence in the human fungal pathogen *Cryptococcus neoformans*. *Genetics* 2011;188. <https://doi.org/10.1534/genetics.111.128538>.
- [360] DeJarnette C, Luna-Tapia A, Estredge LR, Palmer GE. Dihydrofolate Reductase Is a Valid Target for Antifungal Development in the Human Pathogen *Candida albicans*. *MSphere* 2020;5. <https://doi.org/10.1128/msphere.00374-20>.
- [361] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7. <https://doi.org/10.1038/msb.2011.75>.
- [362] DeLano WL. The PyMOL Molecular Graphics System, Version 2.3. Schrödinger LLC 2020.
- [363] St. Vincent MR, Colpitts CC, Ustinov A V., Muqadas M, Joyce MA, Barsby NL, et al. Rigid amphipathic fusion inhibitors, small molecule antiviral compounds against enveloped viruses. *Proc Natl Acad Sci U S A* 2010;107. <https://doi.org/10.1073/pnas.1010026107>.
- [364] CLSI. M27-S4 Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts; Fourth Informational Supplement. vol. 32. 2012.
- [365] Speerstra S, Chistov AA, Proskurin G V., Aralov A V., Ulashchik EA, Streshnev PP, et al. Antivirals acting on viral envelopes via biophysical mechanisms of action. *Antiviral Res* 2018;149. <https://doi.org/10.1016/j.antiviral.2017.11.018>.
- [366] Ren Z, Chhetri A, Guan Z, Suo Y, Yokoyama K, Lee SY. Structural basis for inhibition and regulation of a chitin synthase from *Candida albicans*. *Nat Struct Mol Biol* 2022;29. <https://doi.org/10.1038/s41594-022-00791-x>.

- [367] Talapko J, Juzbašić M, Matijević T, Pustijanac E, Bekić S, Kotris I, et al. *Candida albicans*—the virulence factors and clinical manifestations of infection. *Journal of Fungi* 2021;7. <https://doi.org/10.3390/jof7020079>.
- [368] Chen Y, Shi ZW, Strickland AB, Shi M. *Cryptococcus neoformans* Infection in the Central Nervous System: The Battle between Host and Pathogen. *Journal of Fungi* 2022;8. <https://doi.org/10.3390/jof8101069>.
- [369] Czarniak N, Kamińska J, Matowicka-Karna J, Koper-Lenkiewicz OM. Cerebrospinal Fluid—Basic Concepts Review. *Biomedicines* 2023;11. <https://doi.org/10.3390/biomedicines11051461>.

## Appendix – Supplementary Data





**Supplementary Figure II.1.1** – 16S rRNA phylogenetic tree of several known *C. albicans* closely related species. Sequences retrieved from NCBI’s database and aligned using MEGA X 10.0.5, the evolutionary history was inferred by using Maximum Likelihood method and Tamura-Nei model. The percentage of trees in which associated taxa clustered together is shown next to the branches. The Tree is drawn to scale, with lengths measured in the number of substitutions per site.

**Supplementary Data II.1.1** – Detailed information regarding biomass composition of *Candida albicans* used in iRV781.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qciNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.1.2** – Detailed information regarding model iRV781 curation process.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qciNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.1.3** – Unique EC numbers of iRV781 in comparison to the GSMMS iIN800 and iNX804.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qciNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.1.4** – Model iRV781 in sbml format.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.1.5** – Comparison of gene essentiality in *C. albicans* iRV781 with literature data.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.2.1** – 1 and 2: Detailed information regarding biomass composition of *Candida parapsilosis* used in iDC1003. 3,4 and 6: Detailed information regarding model iRV781 curation process. 6: List of essential enzymes in RPMI medium environmental conditions in *C. parapsilosis* iDC1003, *C. albicans* iRV781, and *C. glabrata* iNX804. 7: List of all EC numbers present in *C. parapsilosis* iDC1003, *C. albicans* iRV781, *C. glabrata* iNX804, and *S. cerevisiae* iIN800, and respective *C. parapsilosis* iDC1003 unique EC numbers.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.2.2** - Model iDC1003 in sbml format.  
<https://drive.google.com/drive/folders/1oZUtNrh9nMAFv7qciNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.3.1** - Detailed information regarding *Candida auris* iRV973 GSMM. Including: Detailed biomass composition; Detailed information regarding model curation process; List of essential enzymes and List of Unique enzymes.  
<https://drive.google.com/drive/folders/1oZUtNrh9nMAFv7qciNYZrw3L6ZU5ty?usp=sharing>



Supplementary Data II.3.2 - Model iRV973 in sbml format.  
<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qjiNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Table II.3.1** - Enzymes predicted to be essential in RPMI medium, based on the screening of the genome-scale metabolic models of *C. auris* IRV973, *C. parapsilosis* iDC1003, *C. albicans* iRV781, and *C. glabrata* iNX804. Bold: fungal-specific enzymes. Data regarding the drug association was retrieved from DrugBank database; only drugs with known pharmacological action against pathogens were selected.

Gene name						Pharmacological action	EC Number	Pathway
<i>C. auris</i>	<i>C. albicans</i>	<i>C. glabrata</i>	<i>C. parapsilosis</i>	<i>S. cerevisiae</i>	Human			
CJI97_000938	ERG26	CAGLOG00594g	CPAR2_302110	ERG26	NSDHL	-	1.1.1.170	Steroid biosynthesis
CJI97_000080	IMH3	CAGL0K10780g	CPAR2_104580	IMD4	IMPDH	-	1.1.1.205	Purine metabolism
CJI97_004310	ERG27	CAGL0M11506g	CPAR2_801560	ERG27	DHRS11	-	1.1.1.270	Steroid biosynthesis
CJI97_003299	HMG1	CAGL0L11506g	CPAR2_110330	HMG1	HMGCR	-	1.1.1.34	Terpenoid backbone biosynthesis
CJI97_003097	ERG24	CAGL0I02970g	CPAR2_405900	ERG24	TM7SF2	-	1.3.1.70	Steroid biosynthesis
<b>CJI97_002908</b>	<b>ERG4</b>	<b>ERG4</b>	<b>ERG4</b>	<b>ERG4</b>	-	-	<b>1.3.1.71</b>	Steroid biosynthesis
CJI97_005101	CDC21	CDC21	CPAR2_206550	TMP1	TYMS	-	2.1.1.45	Pyrimidine metabolism
CJI97_002511	ADE17	CAGL0A03366g	CPAR2_202250	ADE17	ATIC	-	2.1.2.3	Purine metabolism
CJI97_002269	URA2	CAGL0L05676g	CPAR2_203160	URA2	CAD	-	2.1.3.2	Pyrimidine metabolism
<b>FKS1</b>	<b>GSC1</b>	<b>FKS1</b>	<b>CPAR2_106400</b>	<b>FKS1</b>	-	<b>Echinocandins</b>	<b>2.4.1.34</b>	1,3-beta-glucan biosynthesis
CJI97_002422	URA5	URA5	CPAR2_802790	URA5	UMPS	-	2.4.2.10	Pyrimidine metabolism
CJI97_001833	ADE4	CAGL0M13717g	CPAR2_208260	ADE4	PPAT	-	2.4.2.14	Purine metabolism
CJI97_003197	BTS1	CAGL0H05269g	CPAR2_302840	BTS1	GGPS1	-	2.5.1.1	Terpenoid backbone biosynthesis
CJI97_001757	ERG20	ERG20	CPAR2_103950	ERG20	FDPS	-	2.5.1.10	Terpenoid backbone biosynthesis
CJI97_003836	C5_05130C	CAGL0F05555g	CPAR2_502760	CAB5	COASY	-	2.7.1.24	CoA biosynthesis
CJI97_005311	CR_03740C	CAGL0K11022g	CPAR2_202590	FMN1	RFK	-	2.7.1.26	Riboflavin metabolism

<i>CJI97_004586</i>	<i>C6_02980C</i>	<i>CAGL0H01551g</i>	<i>CPAR2_602050</i>	<i>CAB1</i>	<i>PANK</i>	-	2.7.1.33	CoA biosynthesis
<i>CJI97_000033</i>	<i>URA6</i>	<i>CAGL0L09867g</i>	<i>CPAR2_105320</i>	<i>URA6</i>	<i>CMPK2</i>	-	2.7.4.14	Pyrimidine metabolism
<i>CJI97_001215</i>	<i>ERG8</i>	<i>ERG8</i>	<i>CPAR2_400710</i>	<i>ERG8</i>	<i>PMVK</i>	-	2.7.4.2	Terpenoid backbone biosynthesis
<i>CJI97_000019</i>	<i>C5_00260W</i>	<i>CAGL0D00550g</i>	<i>CPAR2_304260</i>	<i>PRS1</i>	<i>PRPS1</i>	-	2.7.6.1	Purine metabolism
<i>CJI97_005306</i>	<i>C4_05210W</i>	<i>CAGL0G03157g</i>	<i>CPAR2_500260</i>	<i>PIS1</i>	<i>CDIPT</i>	-	2.7.8.11	Glycerophospholipid metabolism
<i>CJI97_002826</i>	<i>ADE8</i>	<i>CAGL0F02761g</i>	<i>CPAR2_211620</i>	<i>ADE8</i>	<i>GART</i>	-	2.1.2.2	Purine metabolism
<i>CJI97_005452</i>	<i>COQ3</i>	<i>CAGL0I07601g</i>	<i>CPAR2_602300</i>	<i>COQ3</i>	<i>COQ3</i>	-	2.1.1.114	Ubiquinone biosynthesis
<i>CJI97_003704</i>	<i>COQ5</i>	<i>CAGL0J06710g</i>	<i>CPAR2_209250</i>	<i>COQ5</i>	<i>COQ5</i>	-	2.1.1.201	Ubiquinone biosynthesis
<i>ERG11</i>	<i>ERG11</i>	<i>ERG11</i>	<i>ERG11</i>	<i>ERG11</i>	<i>CYP51A1</i>	Azoles	1.14.14.154	Steroid biosynthesis
<i>CJI97_000224</i>	<i>PEL1</i>	<i>PGS1</i>	<i>CPAR2_805350</i>	<i>PEL1</i>	<i>PGS1</i>	-	2.7.8.5	Glycerophospholipid metabolism
<i>CJI97_004676</i>	<i>C1_08780W</i>	<i>PHO8</i>	<i>CPAR2_804250</i>	<i>PHO8</i>	<i>ALPL</i>	-	3.1.3.1	Folate biosynthesis
<i>CJI97_005490</i>	<i>C6_01340C</i>	<i>CAGL0H04389g</i>	<i>CPAR2_602700</i>	<i>GEP4</i>	<i>PTPMT1</i>	-	3.1.3.27	Glycerophospholipid metabolism
<i>CJI97_002941</i>	<i>URA4</i>	<i>CAGL0J04598g</i>	<i>CPAR2_100500</i>	<i>URA4</i>	<i>CAD</i>	-	3.5.2.3	Pyrimidine metabolism
<i>CJI97_000499</i>	<i>IPP1</i>	<i>IPP1</i>	<i>CPAR2_806200</i>	<i>IPP1</i>	<i>PPA2</i>	-	3.6.1.1	Oxidative phosphorylation
<i>CJI97_004071</i>	<i>ADE2</i>	<i>ADE2</i>	<i>CPAR2_805940</i>	<i>ADE2</i>	<i>PAICS</i>	-	4.1.1.21	Purine metabolism
<i>CJI97_003384</i>	<i>URA3</i>	<i>URA3</i>	<i>URA3</i>	<i>URA3</i>	<i>UMPS</i>	-	4.1.1.23	Pyrimidine metabolism
<i>CJI97_001340</i>	<i>MVD</i>	<i>CAGL0C03630g</i>	<i>CPAR2_109530</i>	<i>MVD1</i>	<i>MVD</i>	-	4.1.1.33	Terpenoid backbone biosynthesis
<i>CJI97_003563</i>	<i>CAB3</i>	<i>CAGL0L05302g</i>	<i>CPAR2_800750</i>	<i>CAB3</i>	<i>PPCDC</i>	-	4.1.1.36	CoA biosynthesis
<b><i>CJI971_001274</i></b>	<b><i>FOL1</i></b>	<b><i>CAGL0J07920g</i></b>	<b><i>CPAR2_303390</i></b>	<b><i>FOL1</i></b>	-	<b>Sulfacetamide</b>	<b>4.1.2.25</b>	Folate biosynthesis
<b><i>CJI971_001274</i></b>	<b><i>FOL1</i></b>	<b><i>CAGL0J07920g</i></b>	<b><i>CPAR2_303390</i></b>	<b><i>FOL1</i></b>	-	<b>Sulfonamides</b>	<b>2.5.1.15</b>	Folate biosynthesis
<i>CJI97_000801</i>	<i>ADE13</i>	<i>CAGL0B02794g</i>	<i>CPAR2_204960</i>	<i>ADE13</i>	<i>ADSL</i>	-	4.3.2.2	Purine metabolism
<i>CJI97_001183</i>	<i>IDI1</i>	<i>CAGL0J06952g</i>	<i>CPAR2_401630</i>	<i>IDI1</i>	<i>IDI1</i>	-	5.3.3.2	Terpenoid backbone biosynthesis
<i>CJI97_005090</i>	<i>ERG7</i>	<i>CAGL0J10824g</i>	<i>CPAR2_301800</i>	<i>ERG7</i>	<i>LSS</i>	Oxiconazole	5.4.99.7	Steroid biosynthesis
<i>CJI97_003065</i>	<i>ADE1</i>	<i>CAGL0I04444g</i>	<i>CPAR2_500190</i>	<i>ADE1</i>	<i>PAICS</i>	-	6.3.2.6	Purine metabolism
<i>CJI97_001704</i>	<i>ADE5,7</i>	<i>CAGL0H07887g</i>	<i>CPAR2_208400</i>	<i>ADE5,7</i>	<i>GART</i>	-	6.3.3.1	Purine metabolism
<i>CJI97_001704</i>	<i>ADE5,7</i>	<i>CAGL0H07887g</i>	<i>CPAR2_208400</i>	<i>ADE5,7</i>	<i>GART</i>	-	6.3.4.13	Purine metabolism

<i>CJI97_002160</i>	<i>ADE6</i>	<i>CAGL0K04499g</i>	<i>CPAR2_204070</i>	<i>ADE6</i>	<i>PFAS</i>	-	6.3.5.3	Purine metabolism
<i>CJI97_001038</i>	<i>ACC1</i>	<i>CAGL0L10780g</i>	<i>CPAR2_804060</i>	<i>ACC1</i>	<i>ACACA</i>	-	6.4.1.2	Fatty acid biosynthesis
<i>CJI97_005606</i>	<i>ERG12</i>	<i>CAGL0F03861g</i>	<i>CPAR2_803530</i>	<i>ERG12</i>	<i>MVK</i>	-	2.7.1.36	Terpenoid backbone biosynthesis
<i>CJI97_004952</i>	<i>ERG13</i>	<i>ERG13</i>	<i>CPAR2_701400</i>	<i>ERG13</i>	<i>HMGCS</i>	-	2.3.3.10	Terpenoid backbone biosynthesis
<b><i>CJI97_001309</i></b>	<b><i>FAS1</i></b>	<b><i>CAGL0D00528g</i></b>	<b><i>FAS1</i></b>	<b><i>FAS1</i></b>	-	-	<b>2.3.1.86</b>	Fatty acid biosynthesis
<i>CJI97_005609</i>	<i>GUA1</i>	<i>CAGL0F03927g</i>	<i>CPAR2_803560</i>	<i>GUA1</i>	<i>GMPS</i>	-	6.3.5.2	Pyrimidine metabolism
<i>CJI97_004456</i>	<i>URA7</i>	<i>URA7</i>	<i>CPAR2_100620</i>	<i>URA7</i>	<i>CTPS1</i>	-	6.3.4.2	Pyrimidine metabolism
	<b><i>CHS1</i></b>	<b><i>CAGL0I04818g</i></b>	<b><i>CPAR2_805640</i></b>	<b><i>CHS2</i></b>	-	-	<b>2.4.1.16</b>	Chitin biosynthesis

---

**Supplementary Data II.4.1** – 1, 2, 3, and 4: Detailed information regarding biomass composition of *Cryptococcus neoformans* used in iRV890. 3,4 and 6: Detailed information regarding model iRV890 curation process. 5: List of essential enzymes in RPMI medium environmental conditions in *C. neoformans* iRV890. 6: List of all EC numbers present in *C. neoformans* iRV890, *C. parapsilosis* iDC1003, *C. albicans* iRV781, *C. glabrata* iNX804, and *S. cerevisiae* iIN800, and respective *C. neoformans* iRV890 unique EC numbers, with a detailed description.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjiNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data II.4.2** - Model iRV890 in sbml format.  
<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjiNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Table II.4.1** - Comparison between experimental and *in silico* phenotypic behavior of *C. neoformans* under different carbon and nitrogen sources. Highlighted in grey are the cases that are not in accordance with both

evidence. Carbon source utilization was predicted correctly in 86% (133/155) of the cases and nitrogen source 85% (57/67). Growth (+); lack of growth (-);

<b>Carbon source</b>	<b><i>In silico</i></b>	<b>Experimental</b>
Acetamide	-	-
Acetic acid	+	+
Acetoacetate	-	-
Adenosine	-	-
a-D-Glucose	+	+
a-D-Lactose	-	-
a-Hydroxyglutaric acid-g-Lactone	-	-
a-Keto-Valeric acid	-	-
Amygdalin	-	+
Arbutin	-	+
b-Hydroxybutyric acid	-	-
b-Methyl-D-Galactoside	-	-
b-Methyl-D-Xyloside	-	-
Bromosuccinic acid	-	-
Butylamine	-	-
Butyric acid	-	-
Capric acid	-	-
Caproic acid	-	-
Citric acid	-	-
D,L-Carnitine	-	-
D,L-Malic acid	-	-
D-Alanine	-	-
D-Allose	-	+
D-Arabinose	-	+
D-Arabitol	+	+
D-Aspartic acid	-	-
D-Cellobiose	+	+
Deoxyadenosine	-	-
Deoxyribose	-	+
Dextrin	+	+
D-Fructose	+	+
D-Fructose-6-Phosphate	-	-
D-Galactarate	-	+
D-Galactose	+	+
D-Galacturonate	-	+
D-Glucarate	-	-
D-Gluconic acid	+	+
D-Glucosamine	+	+
D-Glucose-1-Phosphate	-	-
D-Glucose-6-Phosphate	-	-
D-Glucuronate	+	+
D-Lactic acid Methyl Ester	-	-
D-Lactitol	-	-
D-Malic acid	-	-

D-Mannitol	+	+
D-Mannose	+	+
D-Melibiose	-	-
D- Psicose	-	+
D-Raffinose	+	+
D-Ribose	+	+
D-Serine	+	-
D-Sorbitol	+	+
D-Tagatose	-	+
D-Tartaric acid	-	-
D-Threonine	-	-
D-Trehalose	+	+
D-Xylose	+	+
Ethanolamine	-	-
Formate	-	-
Fumaric acid	-	-
Galactitol	-	+
Gelatin	-	-
g-Hydroxybutyric acid	-	-
Glucuronamide	-	-
Glycerol	+	-
Glycerone phosphate	+	+
Glycine	-	-
Glycolate	-	-
Glyoxylate	-	-
Hydroxyproline	-	-
Inosine	-	-
Inulin	-	-
Itaconic acid	-	-
Lactulose	-	-
L-Alanine	+	+
L-Arabinose	+	+
L-Arabitol	+	+
L-Arginine	-	-
L-Asparagine	-	-
L-Aspartate	+	-
L-Fucose	-	-
L-glutamate	+	+
L-Glutamine	+	-
L-Gulono-1,4-lactone	+	+
L-Histidine	-	-
L-Homoserine	+	-
L-Isoleucine	-	-
L-Leucine	-	-
L-Lysine	-	-
L-Lyxose	+	+
L-Malic acid	-	-
L-Methionine	-	-
L-Ornithine	+	-

L-Phenylalanine	-	-
L-Proline	+	+
L-Rhamnose	-	+
L-Serine	+	-
L-Sorbose	+	+
L-Threonine	+	-
L-Valine	-	-
Malonate	-	-
Maltose	+	+
Mannan	-	-
Methyl beta-D-galactoside	-	-
Methyl ethyl ketone	-	-
Methylpyruvate	-	-
Mono-Methylsuccinate	-	-
myo-Inositol;	+	+
N-Acetyl-D-Galactosamine	-	-
N-Acetyl-D-Glucosamine	+	+
N-Acetyl-D-mannosamine 6- phosphate	-	-
N-Acetyl-L-glutamate	-	-
N-Acetylneuraminic acid	-	-
Octopamine	-	-
Oxalate	-	-
Pectin	-	-
Phenylethylamine	-	-
Propane-1,2,3-tricarboxylate	-	-
Propane-1,2-diol	-	-
Propanoate	-	-
Putrescine	-	-
Pyruvate	-	-
Quinate	-	-
Ribitol	-	+
Salicin	-	+
Sebacic acid	-	-
Sedoheptulose	-	-
sn-Glycerol 3-phosphate	-	-
Stachyose	+	+
Starch	-	-
Succinate	-	-
Sucrose	+	+
Thymidine	-	-
Tyramine	-	-
Uridine	-	-
Xylitol	-	-
(R)-2-Methylmalate	-	-
(R)-Acetoin	-	-
(R,R)-Butane-2,3-diol	-	-
(R,R)-Tartaric acid	-	-
(S)-Lactate	-	-

2-Amino-2-deoxy-D-gluconate	-	-
2-Hydroxybenzoic acid	-	-
2-Hydroxybutanoic acid	-	-
2-Methylmaleate	-	-
2-Oxobutanoate	-	-
2-Oxoglutarate	-	-
3-Hydroxyphenylacetate	-	-
3-Oxalomalate	-	-
4-Aminobutanoate	+	+
4-Hydroxybenzoic acid	-	-
4-Hydroxyphenylacetate	-	-
5-Aminopentanoate	-	-
5-Oxoproline	-	-
6-Deoxy-D-galactose	-	-
<b>Nitrogen source</b>	<b><i>In silico</i></b>	<b>Experimental</b>
Ammonia	+	+
Nitrite	-	-
Nitrate	-	-
Urea	+	+
Biuret	-	-
L-Alanine	+	+
L-Arginine	+	+
L-Asparagine	-	+
L-Aspartic acid	+	+
L-Cysteine	-	-
L-glutamate	+	+
L-Glutamine	+	+
Glycine	+	+
L-Histidine	-	-
L-Isoleucine	-	+
L-Leucine	-	+
L-Lysine	-	+
L-Methionine	-	+
L-Phenylalanine	-	+
L-Proline	+	+
L-Serine	+	+
L-Threonine	+	+
L-Tryptophan	+	+
L-Tyrosine	-	+
L-Valine	-	+
D-Alanine	-	-
D-Aspartic acid	+	+
D-Glutamic acid	-	-
D-Lysine	-	-
D-Serine	+	+

D-Valine	-	-
L-Citrulline	+	+
L-Homoserine	+	+
L-Ornithine	+	+
N-Acetyl-L-Glutamic acid	+	+
N-Acetyl-L-glutamate	-	-
Hydroxylamine	-	-
Methylamine	-	-
Ethylamine	-	-
Ethanolamine	+	+
Putrescine	-	-
Agmatine	+	+
Histamine	-	-
Phenethylamine	-	-
Tyramine	-	-
Acetamide	-	-
Formamide	-	-
D-Glucosamine	+	-
N-Acetyl-D-Glucosamine	+	+
N-Acetyl-D-Galactosamine	-	-
N-Acetyl-D-Mannosamine	-	+
Adenine	-	-
Adenosine	-	-
Cytidine	-	-
Cytosine	-	-
Guanine	+	+
Guanosine	-	-
Thymine	-	-
Thymidine	-	-
Uracil	-	-
Uridine	-	-
Inosine	-	-
Xanthine	-	-
Xanthosine	-	-
Uric acid	+	+
Allantoin	+	+
4-Aminobutanoate	+	+

**Supplementary Data III.1** – 3D structure of Fol1 protein of *C. albicans* modeled structure using MODELLER 9.23 with a DOPE score of -1.87 used in the docking studies.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data III.2** - List of the compounds used in virtual drug screening of *C. albicans* Fol1 modeled structure with all the respective binding energies.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>



**Supplementary Data III.3** - List of the compounds with binding energies lower than -10 kcal/mol calculated in virtual drug screening of *C. albicans* Chs2 3D structure.

<https://drive.google.com/drive/folders/1oZUtNrh9nMAtFv7qcjNYZrw3L6ZU5ty?usp=sharing>

