

Exploração de ferramentas bioinformáticas para a análise do proteoma na descoberta de medicamentos para o controlo da candidíase

Diana Silva¹, Daniela Araújo^{2,3}, Bruna Gonçalves²

¹Departamento de Informática, Universidade do Minho

²Centro de Engenharia Biológica, Universidade do Minho

³INIAV – Instituto Nacional de Investigação Agrária e Veterinária,
Pólo de Vairão

Abstract.

1. Introdução

1.1. Infecções causadas por espécies de *Candida*

1.1.1 A candidíase e a problemática associada à resistência antifúngica

A candidíase, uma das infecções fúngicas mais comuns em todo o mundo, afeta indivíduos de todas as idades, tendo uma prevalência significativa e um grande impacto na saúde pública (Brown et al., 2012). Estas infecções são causadas em grande maioria por fungos do género *Candida* que são microrganismos oportunistas que colonizam diversos nichos do corpo humano como a boca, os genitais, a pele, a garganta e outras áreas mucosas (Calderone & Clancy, 2012). O género *Candida* é composto por mais de 150 espécies, apresentando uma taxa de mortalidade de cerca de 30%-50% em pacientes hospitalizados (Quindós, 2018). Esta infeção representa um desafio significativo para a saúde pública, tendo também um impacto substancial nos recursos médicos e financeiros que são necessários para o tratamento e hospitalização de pacientes. Dentro das espécies de *Candida*, a *Candida albicans* continua a ser a mais prevalente (Pfaller, 2012).

No entanto, o aumento do uso de agentes antifúngicos para o tratamento destas infeções, incluindo o recurso à automedicação, prescrição inadequada e terapias prolongadas, tem conduzido a um aumento da resistência antifúngica (Araújo et al., 2019). A resistência microbiana acarreta uma preocupação clínica, mas também socioeconómica (Sobel, 2007). Assim, com o aumento da resistência antifúngica, novas abordagens inovadoras para o tratamento eficaz de infeções por *Candida* são emergentes. Desta forma, é necessário implementar estratégias de prevenção, investir em pesquisas para o desenvolvimento de novas terapias, aumentar a conscientização pública sobre a importância do uso de novas terapias, mas também sobre a importância do uso indiscriminado de antimicrobianos (Pfaller & Diekema, 2007).

1.1.2 Virulência da espécie *Candida albicans*

Os mecanismos de virulência desempenham um papel crucial na patogenicidade de *C. albicans* os quais contribuem para o desenvolvimento de infeções oportunistas em diferentes áreas do corpo, e que resultam em elevadas taxas de morbilidade e mortalidade (Araújo et al., 2017).

A patogenicidade de *C. albicans* é suportada por uma série de fatores de virulência, sendo um dos mais alarmantes a sua capacidade de mudar da forma de levedura para uma forma filamentosa, a qual é designada de hifa (Figura 1).



Figura 1. Morfologias da espécie de *Candida albicans*. (a) Levedura; (b) Pseudo-hifa; (c) Hifa.

A capacidade de alternar entre estas formas morfológicas é designada de dimorfismo. Esta capacidade de alteração de morfologia permite a adaptação das espécies de *Candida* a diferentes ambientes no hospedeiro. Em condições normais, as células de *C. albicans* assumem a forma de levedura, que é geralmente não invasiva, sendo predominantemente encontrada na microbiota normal da pele, mucosas e trato gastrointestinal humano. No entanto, em resposta a estímulos ambientais específicos, como mudanças de temperatura, pH e a presença de certos produtos químicos ou células do sistema imunológico, as células de *C. albicans* podem alternar para forma

filamentosa. As hifas são estruturas alongadas e multicelulares que podem penetrar os tecidos e as membranas, facilitando a invasão do hospedeiro (Jacobsen et al., 2012).

O dimorfismo é regulado por uma rede complexa genética, sendo que o gene *EFG1* desempenha um papel central na regulação da morfologia de *C. albicans*. Foi identificado pela primeira vez como indutor do crescimento das pseudo-hifas em *Saccharomyces cerevisiae*, sendo posteriormente identificado como essencial para o crescimento de hifas em *C. albicans* (Stoldt et al., 1997). As estirpes de *C. albicans* manipuladas geneticamente com eliminação do gene *EFG1*, apresentam uma incapacidade de produzir hifas em resposta à maioria dos estímulos. O gene *EFG1* também está associado a capacidade de formar biofilme, outro fator de virulência importante para as espécies de *Candida*. O biofilme é uma estrutura multicelular aderida a uma superfície, caracterizado pela presença de microrganismos devidamente organizados, composto por uma matriz extracelular que permite a manutenção de nutrientes essenciais (Lo et al., 1997; Stoldt et al., 1997; Sudbery, 2011).

1.2. Terapia Antisense

O aumento da resistência aos antimicrobianos e a falta de identificação de novas classes de agentes antimicrobianos, têm levado ao desenvolvimento de novas estratégias terapêuticas. A terapia antisense tem sido reconhecida como uma terapia promissora para o tratamento de doenças crônicas humanas não infecciosas (Araújo et al., 2019). No entanto, tem sido também aplicada nos últimos tempos em doenças infecciosas humanas. Esta terapia baseia-se na aplicação de moléculas de antisense, designadas de oligonucleotídeos de antisense que são complementares a um mRNA específico, permitindo a inibição da expressão genética, induzindo um bloqueio na transferência de informação genética do DNA para a proteína. Os ácidos nucleicos são potenciais candidatos a novos fármacos para tratar geneticamente infecções, silenciando um gene básico de crescimento ou combatendo a resistência aos antimicrobianos ou fatores de virulência de um agente patogénico (Costerton et al., 2009).

1.2.1 Mecanismo de ação

Os oligonucleotídeos de antisense são sequências curtas de ácidos nucleicos complementares a sequências específicas de RNA mensageiro que codificam proteínas de interesse. Estes oligonucleotídeos são geralmente modificados quimicamente de forma a aumentar a sua estabilidade no hospedeiro e a sua resistência às nucleases (Vickers & Crooke, 2015).

Existem diferentes mecanismos de ação associados à terapia antisense, sendo divididos em dois principais: os que ativam a enzima RNase H e os que não a ativam. No primeiro caso, os oligonucleotídeos ligam-se ao RNA mensageiro alvo por meio de hibridização, ativam a RNase H permitindo a degradação do mRNA, e inibindo assim a síntese da proteína correspondente. Por outro lado, os oligonucleotídeos de antisense podem ligar-se à região codificadora do mRNA, bloqueando a ligação das duas unidades ribossomais, e impedindo consequentemente a tradução em proteínas (Figura 2) (Kole et al., 2012).

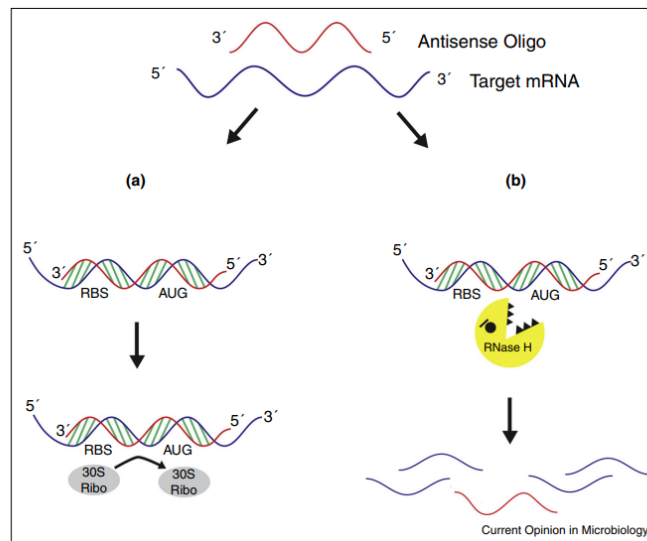


Figura 2. Mecanismos de ação de oligonucleotídeos antisense na redução da expressão genética. (a) Bloqueio da subunidade ribossômica 30S e o início da tradução. (b) Ativação da RNase H e consequente degradação do mRNA do gene alvo.

1.2.2 A aplicação da Terapia Antisense no controlo da virulência de *Candida albicans*

Apesar da terapia antisense estar largamente aplicada em doenças não infecciosas humanas, a sua aplicação no controlo de infeções fúngicas ainda foi pouco explorada (Bennett et al., 2017). Recentemente, um estudo pioneiro permitiu o controlo de um fator de virulência de *C. albicans* a partir do bloqueio de um gene que tem um papel fundamental na filamentação (Costerton et al., 2009). Para tal, foi desenhado um oligonucleotídeo modificado quimicamente, anti-*EFG1* 2'OMe, para bloquear o gene *EFG1*. Este oligonucleotídeo permite ativar a RNase H e degradar o mRNA do gene alvo, bloqueando a sua tradução em proteína e, consequentemente, controlar a formação de hifas (Georgiadou et al., 2021).

Este oligonucleotídeo foi quimicamente modificado por uma modificação de açúcar, nomeadamente 2'OMethyl e foi desenhado tendo em conta diversas características cruciais para um bom desempenho de hibridação, nomeadamente, ter temperaturas de fusão em torno de 39°C-42°C, composição de guanina-citosina (GC) de aproximadamente 50% a 60%, e ter um tamanho entre 12 e 20 nt (nucleótidos) (Araújo et al., 2019).

Este estudo demonstrou que o anti-*EFG1* 2'OMe inibiu a expressão genética do *EFG1*, assim como a tradução da proteína e consequentemente reduziu a filamentação em *C. albicans* (Araújo et al., 2019). Assim, este estudo permitiu comprovar pela primeira vez a aplicação da terapia antisense para controlar fatores de virulência das espécies de *Candida*.

1.3 Bioinformática

1.3.1 Análise de dados ómicos

A área da biotecnologia requer um amplo conjunto de informações para alcançar os seus objetivos. Nesse sentido, compreender a biologia dos microrganismos por meio de estudos ómicos oferece não só apoio, como também abre novas perspectivas para a expansão desse conhecimento, através de uma análise mais aprofundada das moléculas e das funções celulares (Oliveira, 2018).

A genómica consiste na análise de todos os elementos do genoma, tais como, genes, reguladores, genes de RNA não codificantes e outros elementos móveis. A transcriptómica é o estudo do transcriptoma, o conjunto completo de transcrições de RNA originadas pelo genoma. Através da análise dos transcritos de RNA é possível inferir a função dos genes e o estado atual da célula, permitindo analisar informações

importantes. Como a transcriptômica nem sempre reflete a expressão proteica numa dada célula/população (Oliveira, 2018), o campo da proteômica surgiu como uma nova abordagem para estudar todo o conjunto de proteínas numa célula/organismo, as suas características e interações, em diferentes situações (Schmidt et al., 2014). Apesar da valiosa informação obtida através dos estudos ômicos, a manipulação e análise dos numerosos dados que estes fornecem constitui um desafio. Consequentemente, a utilização de tecnologia computacional facilita este processo (Wiwanitkit, 2013).

A bioinformática tem vindo a ganhar extrema importância em estudos ômicos devido à sua capacidade de armazenar, recuperar e analisar dados biológicos através de ferramentas desenvolvidas e algoritmos adequados. Devido à sua importância, várias tecnologias no domínio biotecnológico usadas em estudos ômicos, como os microarrays, RNAseq (transcriptoma) e LC-MS/MS (proteoma), já estão interligadas com a bioinformática e necessitam da sua intervenção direta para funcionarem de forma ótima (Yadav, 2015).

Os estudos ômicos, aliados às ferramentas bioinformáticas, têm proporcionado uma compreensão mais aprofundada da progressão, tratamento e diagnóstico de doenças, assim como da patogenicidade dos microrganismos e das interações entre hospedeiro e agentes patogénicos.

1.3.2 Bases de dados bioinformáticas: descrição e ferramentas

As bases de dados permitem o armazenamento de informação de forma organizada e de fácil acesso. As bases de dados biológicas podem, portanto, armazenar informação sobre uma ou mais espécies, nomeadamente, o seu genoma, o seu proteoma e potenciais vias, permitindo uma visão geral e organizada da informação atual sobre o organismo/ gene/ composto de interesse. Entre as diferentes bases de dados usadas para o estudo de dados ômicos de espécies de *Candida* as mais usadas são: Pathoyeasttract, *Candida* genome database, FungiFun, Uniprot e String.

PathoYeasttract

A PathoYeasttract (Pathogenic Yeast Search for Transcriptional Regulators And Consensus Tracking) é uma base de dados que reúne todas as associações regulatórias conhecidas entre fatores de transcrição e genes-alvo em espécies patogénicas de *Candida*, incluindo *Candida albicans*. Esta plataforma permite a identificação de reguladores de transcrição documentados, ou potenciais, para um gene específico, assim como identificação dos genes alvo documentados ou potenciais para cada fator de transcrição conhecido. Além disso, possibilita a comparação entre os motifs de DNA e os locais de ligação dos fatores de transcrição descritos na literatura, oferecendo uma ferramenta eficaz para agrupar genes com base nas suas associações regulatórias.

***Candida* genome database**

A *Candida* genome database (CGD) é uma ferramenta online estabelecida em 2004 que fornece informações genómicas e de biologia molecular sobre as espécies de *Candida*. O seu objetivo principal é vincular dados de sequência com dados experimentais derivados da literatura científica, permitindo uma fácil rastreabilidade do processo de investigação da espécie. A CGD oferece uma visão abrangente e atualizada das informações disponíveis sobre as espécies de *Candida*. A CGD tem diversas ferramentas com utilização para múltiplos objetivos, incluindo análise de genes/proteínas associados direta ou indiretamente (por ortologia) a fenótipos de interesse, como por exemplo a formação de biofilme e a produção de hifas.

FungiFun

A FungiFun é uma ferramenta poderosa que atribui anotações funcionais a genes ou proteínas fúngicas, possibilitando análises de distribuição funcional de conjuntos de

genes/proteínas. Utilizando três métodos de classificação distintos - FunCat, GO e KEGG - a FungiFun categoriza genes e proteínas de uma variedade de espécies de fungos em diferentes níveis de detalhe de anotação. Essa abordagem permite uma compreensão mais abrangente das funções biológicas desses genes/proteínas e simplifica a interpretação dos dados ômicos.

UniProt

A UniProt (Universal Protein) é uma base de dados altamente qualificada e abrangente que disponibiliza informações detalhadas sobre sequências de proteínas e as respectivas funções, estruturas, interações e anotações. A maioria das entradas da UniProt é proveniente de projetos de sequenciamento de genomas, contendo uma vasta quantidade de dados sobre as funções biológicas das proteínas, extraídos da literatura científica. Além disso, esta base de dados oferece uma variedade de recursos e ferramentas para análise e pesquisa, incluindo pesquisas de sequências de proteínas, identificação de homólogos, acesso a dados de expressão genética, análise de interações proteína-proteína, bem como a capacidade de prever estruturas de proteínas. Com a sua vasta gama de funcionalidades, a UniProt é uma ferramenta essencial numa ampla variedade de aplicações, desde a descoberta de novos alvos terapêuticos até a compreensão de processos biológicos fundamentais.

String

A String (Search Tool for the Retrieval of Interacting Genes/Proteins) é uma base de dados altamente sofisticada de interações proteína-proteína que reúne uma variedade de dados experimentais e computacionais. O principal objetivo é prever e visualizar redes de interações moleculares, proporcionando uma compreensão abrangente das interações entre diferentes proteínas. Com a integração de dados experimentais, a String oferece uma plataforma robusta para explorar e analisar complexas redes de interação molecular. Essa base de dados desempenha um papel crucial ajudando na compreensão da função e regulação das proteínas, bem como a identificar potenciais alvos terapêuticos.

1.4 Objetivos do trabalho

Desta forma, o principal objetivo deste trabalho é analisar os dados proteômicos do fungo patogênico *C. albicans*, utilizando 5 bases de dados diferentes: Pathoyeast, *Candida* genome database, FungiFun, Uniprot e String, com o intuito de clarificar o efeito antifúngico de um novo fármaco inovador para a candidíase. O foco será a compreensão do modo de ação e aplicabilidade deste novo medicamento, denominado anti-*EFG1* 2'OMe, contra a espécie causadora da candidíase. As tarefas incluirão:

- A comparação dos dados proteômicos de células de *C. albicans* expostas/não expostas ao anti-*EFG1* 2'OMe;
- A identificação dos alvos proteicos do anti-*EFG1* 2'OMe;
- A determinação da sobreposição entre as proteínas alvo do anti-*EFG1* 2'OMe e as proteínas codificadas pelos genes regulados pelo fator de transcrição *Efg1*
- A análise dos alvos mais relevantes em termos de função biológica, distribuição funcional, fenótipos putativos, nível de tradução proteica e interação molecular.

Referências

- Araújo, D., Henriques, M., and Silva, S. (2017). Portrait of *Candida* species biofilm regulatory network genes. *Trends Microbiol.* 25, 62–75.
- Casadevall, A., and Pirofski, L. (2001). Host-pathogen interactions: the attributes of virulence. *J. Infect. Dis.* 184, 337–344
- Connolly, L.A., Riccombeni, A., Grózer, Z., Holland, L.M., Lynch, D.B., Andes, D.R., Gácsér, A., and Butler, G. (2013). The APSES transcription factor *Efg1* is a global regulator that controls morphogenesis and biofilm formation in *Candida parapsilosis*. *Mol. Microbiol.* 90, 36–53.
- Eggimann, P., Garbino, J., and Pittet, D. (2003). Epidemiology of *Candida* species infections in critically ill non-immunosuppressed patients. *Lancet Infect. Dis.* 3, 685–702
- Ernst, J.F. (2000). Transcription factors in *Candida albicans* – environmental control of morphogenesis. *Microbiology* 146, 1763–1774.
- Gonçalves, B., Ferreira, C., Alves, C.T., Henriques, M., Azeredo, J., and Silva, S. (2016). Vulvovaginal candidiasis: epidemiology, microbiology and risk factors. *Crit. Rev. Microbiol.* 42, 905–927.
- Koehler, P., Stecher, M., Cornely, O.A., Koehler, D., Vehreschild, M.J.G.T., Bohlius, J., Wisplinhoff, H., and Vehreschild, J.J. (2019). Morbidity and mortality of candidaemia in Europe: an epidemiologic meta-analysis. *Clin. Microbiol. Infect.* 25, 1200–1212.
- Mayer, F.L., Wilson, D., and Hube, B. (2013). *Candida albicans* pathogenicity mechanisms. *Virulence* 4, 119
- Negri, M., Henriques, M., Svidzinski, T.I., Paula, C.R., and Oliveira, R. (2009). Correlation between Etest, disk diffusion, and microdilution methods for antifungal susceptibility testing of *Candida* species from infection and colonization. *J. Clin. Lab. Anal.* 23, 324–330.
- Negri, M., Silva, S., Henriques, M., and Oliveira, R. (2012). Insights into *Candida tropicalis* nosocomial infections and virulence factors. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 1399–1412
- Nobile, C.J., and Mitchell, A.P. (2005). Regulation of cell-surface genes and biofilm formation by the *C. albicans* transcription factor Bcr1p. *Curr. Biol.* 15, 1150–1155.
- Nobile, C.J., Fox, E.P., Nett, J.E., Sorrells, T.R., Mitrovich, Q.M., Hernday, A.D., Tuch, B.B., Andes, D.R., and Johnson, A.D. (2012). A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell* 148, 126–138.
- Pfaller M. A. (2012). Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. *The American journal of medicine*, 125(1 Suppl), S3–S13.
- Pfaller, M. A., & Diekema, D. J. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. *Clinical Microbiology Reviews*, 20(1), 133–163.
- Quindós, G. (2018). Epidemiology of invasive mycoses: a landscape in continuous change. *Rev. Iberoam. Micol.* 35, 171–178.
- Ramage, G., VandeWalle, K., López-Ribot, J.L., and Wickes, B.L. (2002). The filamentation pathway controlled by the *Efg1* regulator protein is required for normal biofilm formation and development in *Candida albicans*. *FEMS Microbiol. Lett.* 214, 95–100.

Saville, S.P., Lazzell, A.L., Monteagudo, C., and Lopez-Ribot, J.L. (2003). Engineered control of cell morphology in vivo reveals distinct roles for yeast and filamentous forms of *Candida albicans* during infection. *Eukaryot. Cell* 2, 1053–1060

Stoldt, V.R., Sonneborn, A., Leuker, C.E., and Ernst, J.F. (1997). *Efg1p*, an essential regulator of morphogenesis of the human pathogen *Candida albicans*, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. *EMBO J.* 16, 1982–1991.

Yapar, N. (2014). Epidemiology and risk factors for invasive candidiasis. *Ther. Clin. Risk Manag.* 10, 95–105.