Prioritizing and modelling of putative drug target proteins of *Candida albicans* by systems biology approach

Tariq Ismail[1✉], Nighat Fatima[1], Syed Aun Muhammad[2], Syed Saoud Zaidi[3], Nisar Rehman[1], Izhar Hussain[1], Najam us Sahr Tariq[4], Imran Amirzada[1] and Abdul Mannan[1]

[1]Department of Pharmacy, COMSATS Institute of Information Technology, Abbottabad, 22060, Pakistan; [2]Institute of Molecular Biology and BioTechnology, Bahauddin Zakariya University, Multan, 46000, Pakistan; [3]Department of Pharmaceutics, University of Florida, Gainesville, 36090, USA; [4]Department of Pharmaceutics Margella Institute of Health Science, Rawalpindi, 44000 Pakistan

Candida albicans (*Candida albicans*) is one of the major sources of nosocomial infections in humans which may prove fatal in 30% of cases. The hospital acquired infection is very difficult to treat affectively due to the presence of drug resistant pathogenic strains, therefore there is a need to find alternative drug targets to cure this infection. *In silico* and computational level frame work was used to prioritize and establish antifungal drug targets of *Candida albicans*. The identification of putative drug targets was based on acquiring 5090 completely annotated genes of *Candida albicans* from available databases which were categorized into essential and non-essential genes. The result indicated that 9% of proteins were essential and could become potential candidates for intervention which might result in pathogen eradication. We studied cluster of orthologs and the subtractive genomic analysis of these essential proteins against human genome was made as a reference to minimize the side effects. It was seen that 14% of *Candida albicans* proteins were evolutionary related to the human proteins while 86% are non-human homologs. In the next step of compatible drug target selections, the non-human homologs were sequentially compared to the human microbiome data to minimize the potential effects against gut flora which accumulated to 38% of the essential genome. The sub-cellular localization of these candidate proteins in fungal cellular systems indicated that 80% of them are cytoplasmic, 10% are mitochondrial and the remaining 10% are associated with the cell wall. The role of these non-human and non-gut flora putative target proteins in *Candida albicans* biological pathways was studied. Due to their integrated and critical role in *Candida albicans* replication cycle, four proteins were selected for molecular modeling. For drug designing and development, four high quality and reliable protein models with more than 70% sequence identity were constructed. These proteins are used for the docking studies of the known and new ligands (unpublished data). Our study will be an effective framework for drug target identifications of pathogenic microbial strains and development of new therapies against the infections they cause.

✉e-mail: tariqismail@ciit.net.pk
Abbreviations: DEG, database of essential gene; NHHG, non human homolog gene; NHNGH, non human non gut homolog gene; CELLO, subcellular localization predictor; STRING, search tool for the retrieval of interacting genes/protein; PTMs, post translational modification; BLASTp, basic local alignment search tool-protein; KEGG, Kyoto encyclopedia of gene and genome; NCBI, National Centre for Biotechnology Information; UNIPROT, universal protein resource; ATM, ataxia-telangiectasia mutated; CKI, cyclin dependent kinase inhibitor I; CKII, cyclin dependent kinase inhibitor II; CAMII, cell adhesion molecules II; DNA-PK, DNA dependent protein kinase; EGFR, epidermal growth factor receptor; GSK3, glycogen synthase kinase 3; INSR, insulin receptor; PKA, protein kinase A; PKB, protein kinase B; PKC, protein kinase C; rsk, ribosomal S6 kinase; src, sarcoma; cdc2, cyclin dependent kinase Inhibitor II; cdk5, cyclin dependent kinase inhibitor V; p38MAPK, p38 mitogen activated protein kinase

INTRODUCTION

Drug discovery process starts with identification of the drug target (Chan *et al.*, 2010), and provision of a complete fungal/pathogen genome sequence has sped up this procedure. Due to the availability of both, pathogen and host genome sequences, through international institutions hosting systems such as Entrez which contain databases like NCBI, identification of drug targets at the genomic level has become easier for any given pathogen (Marton *et al.*, 1998). Historically, targets were identified based on the knowledge related to the function of each individual protein molecule involved in various steps of the reaction. This procedure not only incurredhuge costs in terms of finances, but also in terms of the time which increased the cost of research and grew steeply and steadily in the last few decades (Volkswirtschaftliche, 2005). The advantages of using *in silico* methods are speed, low cost and provision of viewing whole microbes as a complete system. These methods have advantage of asking questions which were otherwise difficult to address especially experimentally. Now instead of using ligand based approaches to target individual molecules, *in-silico* approach can target molecular pathways of the disease (Mdluli & Spigelman, 2006).

Inthe last decades there was paradigm shift in discovering novel targets by drug designing based on genomics, proteomics and metabolomics approaches, especially for drug resistant pathogens (Lin & Qian, 2007; Fischbach & Walsh, 2010). The most common approach is based on genomics and specifically, *in silico* comparative, subtractive and functional genomics are used to identify new drug targets for effective antifungal agents against the fungal pathogens which are becoming resistant to existing regimens in certain setting (Ji, 2002; Pucci, 2006). It is usually proposed that analysis at the level of a system involving genes, proteins and interactions between them provide a deep insight into the route required for drug targets and resistance profile of pathogen. The pre-

requisite of any such analysis required the presence of a protein interactome of a given pathogen (Raman & Chandra, 2008).

Candida albicans is a diploid fungal pathogen, harmless to nearly 70% of the human population, while nearly 75% of women suffer from its infection at least once in their lifetime (Meiller *et al.*, 2009; Schulze & Sonnenborn, 2009). *Candida albicans* turn into an opportunistic organism for immunocompromised patients and causes a disease known as candidiasis. Candidiasis is categorized as a spotty white infected membrane known as "thrush" in GIT epithelial cells, vaginal or oropharyngeal mucosa, or a life threatening infection in severely ill patients, with mortality rate of 30% (Scully, EI-Kabir, & Samaranayake, 1994; Soll *et al.*, 1988). *Candida species*are the fourth largest cause of cardiovascular and blood stream infections with mortality rate as high as 50% among infected patient in the US (Kabir *et al.*, 2012). It causes infection by changing different phenotypic states playing a particular role in infection, by displaying certain external proteins and by secreting digestive enzymes (Braun *et al.*, 2005). *Candida albicans* has different strategies for its survival against treatment by having a thick cell wall and by releasing aspartic proteases which prevent opsonization by cleaving C3b (a component of complementary protein formed by cleavage of Component C3 used in cascade reaction for generation of killing protein) (Riesbeck, 2013; Garcı, 2004).

Here, we report a comprehensive framework for analysis of the entire *Candida albicans* genome to identify and prioritize alternative and novel drug target proteins (Fig. 1). This systems level approach involves (i) the use of genome databases and biological network tools (ii) analysis of network-based relationships between the essential and non-essential proteins (iii) using subtractive and comparative genomic analysis to minimize the risks of disturbing human gastrointestinal microbial flora (iv) predict subcellular localization of drug targets, and (v) integrated pathways analysis and modeling of targeted proteins. This study emphasized on the importance of computational analysis and its potential as generalized strategy for identification of putative drug targets of pathogenic strains.

MATERIAL AND METHODS

***Candida albicans* genome coverage**. The categorized annotated gene and protein sequences of *Candida*
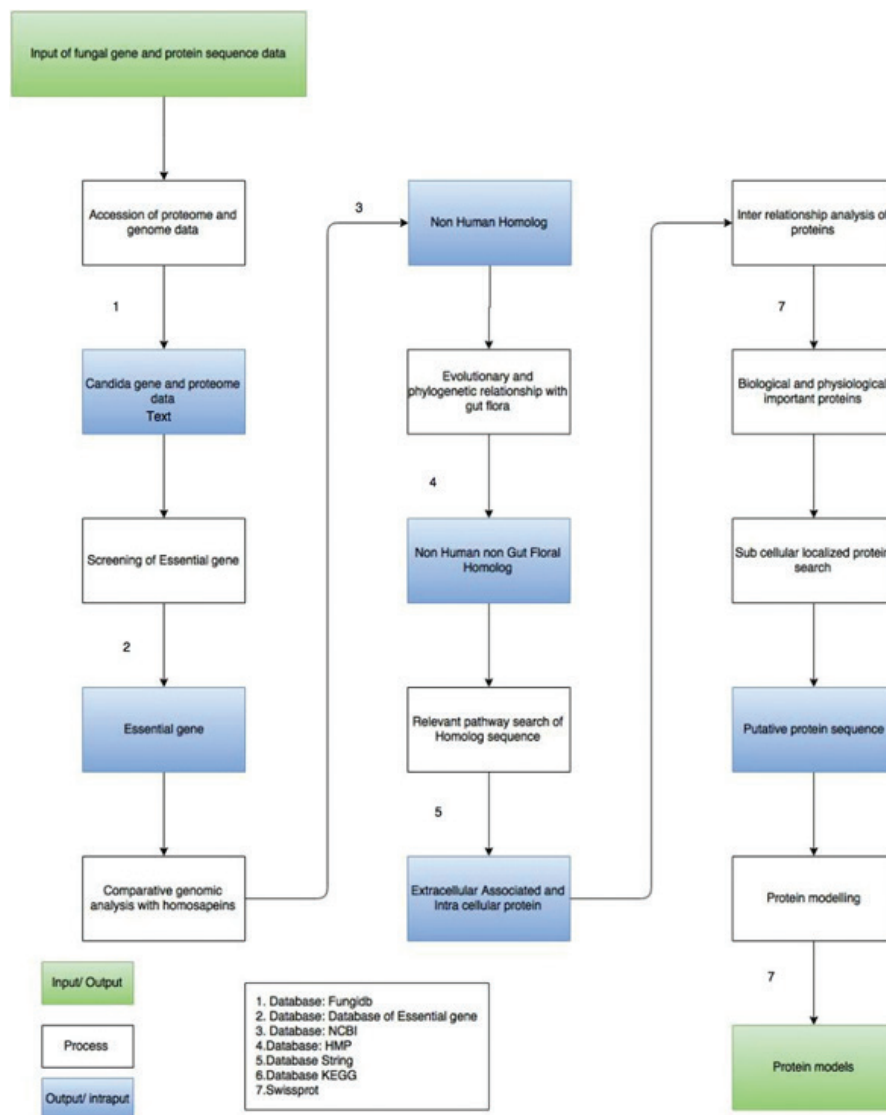

Figure 1. Steps involved in identification and prioritization lof putative protein targets of *Candida albicans* using systems biology approach

Table 1. Data bases, Tools and Software used in this study for prediction of potential fungal drug targets

Tool/Software/Database	Accessibility	Utility
Database of Essential Genes (DEG)	http://tubic.tju.edu.cn/deg/	For screening of essential genes
STRING database version 8.3	http://string-db.org/	For known and predicted protein/COGs interaction
Cytoscape version 2.8.1	http://www.cytoscape.org/	For network analysis and visualization
Human Microbiome Project (HMP)	http://www.hmpdacc.org/	Comparative human microbiome genetic analysis
JGI Genome Portal database	http://genome.jgi-psf.org/	Integrated Genomic Information
MEGA 4.0.2 package	http://www.megasoftware.net/	Molecular evolutionary genetic analysis software
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/	Pathways analysis and comparison
CELLO v.2.5	http://cello.life.nctu.edu.tw/	Subcellular localization prediction for bacteria and Eukaryotes
Uniprot	http://www.uniprot.org/	Resource of protein sequence and functional information
National Centre for Biotechnology information	http://blast.ncbi.nlm.nih.gov/	Biomedical and genomic information source
David Bioinformatic tool	http://david.abcc.ncifcrf.gov	Functional Annotational tool

albicans strain SC3514 in FASTA format were downloaded and retrieved from the Fungidb database (http://fungidb.org/common/downloads/release-2.0/Calbicans_SC5314/fasta/data/, n.d.). All of the tools and softwares used in this study are listed in Table 1. The *Candida albicans* genome delivers a lot of information regarding overall conservation, role categories and therapeutic potential of antifungal targets. These genes coded for essential and non-essential proteins, as well as other regulatory factors. A complete genome of this species comprises 15.7mb, with the total of 5090 genes. Out of 5090 coding genes, only 1205 have role category. The general feature of the *Candida albicans* genome is shown in Table 2. The methodology has been illustrated in Fig. 1.

Screening of essential genes. The Database of Essential Genes (DEG) was used for screening of essential genes (Luo *et al.*, 2014; Zhang & Lin, 2009; Zhang *et al.*, 2004) and sequence alignments were done by using the BLASTp function of the database of essential genes against each gene and protein sequence of the pathogen. The cut off parameter was set at a minimum 100 bit score (>80% identity), along with 0.0001 expectation value-for short listing the essential genes (Barh *et al.*, 2011).

Comparative analysis against human genome. Essential protein sequences of *Candida albicans* were queried to NCBI for finding human homolog and non-human homolog proteins. The BLASTp algorithm of NCBI was used for human-homologs based on E=1 cutoff and sequence similarity of less than 50% at the NCBI server (Muhammad *et al.*, 2014).

Comparative analysis against human microbiota. Many different kinds of organisms are present to in the gut of normal healthy individuals and they perform a very important function in the health and disease of an individual (Garcı, 2004). They not only participate in a commensal relationship between the gut flora and humans, but their role is more than mutualism (Arias &

Murray, 2009). The microorganisms perform many useful functions such as fermentation, enhancing the immune system, preventing harmful species growth (Guarner & Malagelada, 2003), gut developmental regulation, vitamins production and defending host against various diseases (Arias & Murray, 2009). If proteins of these microorganisms were inhibited, then there was a number of side effects (Riesbeck, 2013). To overrule this possibility and to screen out the matched proteins the short listed NHHG (non-human homolog genes) of *Candida albicans* were compared to proteins of the gut flora by sequence analysis with cutoff value (E=0.01) and bit score (>100) at the Human Microbiome Project Database server (Lewis, 2013).

Protein-protein interaction (PPI) studies. The functional annotation and topology analysis was done using protein-protein interactions in the *Candida albicans* system to examine the role of each proteins in a cellular system. For this analysis, the interacting proteins of selected non-human and non-gut flora homologs (NHNGH) were retrieved from the STRING database version 10 with confidence score of >7.0 (Szklarczyk *et*

Table 2. General features of the genome of *Candida albicans*.

Features	Specificity
Total Size	15.6Mb
Total gene	5090
Protein coding gene	4068
Annotated gene	4068
Conserved hypothetical genes	42
True hypothetical genes	2253
tRNA	282
rRNA	10

al., 2011). The PPI-network was constructed by using the Cytoscape software (Bauer-Mehren, 2013).

Prediction of sub-cellular localization. NHNGH-werefound to be the potential fungal targets and their sub-cellular localization (Cytoplasmic membrane, secreted inside cell or secreted extracellularly) were important due to access of drug ligand to the putative target. This localization of NHNGHwere predicted for biological significance and distribution of these target genes by using CELLO v 2.5 sub-cellular localization prediction tool (Lu *et al.*, 2004).

Pathway analysis. The interaction between integrated metabolic and protein network of *Candida albicans* was analyzed and it was found that a correlation existed between these pathways. The identified sequences, which were essential, non-human and non-gut flora homolog genes, were mapped to metabolic pathways. Metabolic pathways were prepared using KEGG database (Altschul *et al.*, 1997). For individual genesanalysis, gene ontology (GO) network was created for cataloging the homologous pathways with in fungus. This analysis also showed those pathways which were absent in the host but were present in the *Candida albicans.*

Protein structure modeling and posttranslational modifications (PTMs). For structure based drug design, there is a need ofprotein structure with accurate conformation (Rausch *et al.*, 2014). The selected essential non-human and non-gut flora homolog protein sequences whose structures were not available in the existing protein databases were comparatively modelled with maximum homology (Arnold *et al.*, 2006) using online SWISS-MODEL server (Bordoli *et al.*, 2009). The stereo-chemical quality of these protein models were based on the Q-means and Z-score (Benkert *et al.*, 2009; Benkert *et al.*, 2011). Further, we predicted the posttranslational modifications (PTMs) of these protein models by sequencebased predictors. We used different statistical tools of the ExPASY resource server (https://www.expasy.org/, n.d.). The protein sequences of these models were analyzed for N-glycosylation, acetylation, O-glycosylation with N-Acetylglucosamine (GlcNAc) or NAcetylgalactosamine (GalNAc) and phosphorylation with NetAcet (Kiemer *et al.*, 2005), NetNGlyc (Gupta, *et al.*, 2002) NetOGlyc (Steentoft *et al.*, 2013), YinOYang (Gupta *et al.*, 2002), and NetPhos (Blom *et al.*, 1999; Jensen *et al.*, 2002), respectively. In this study the probability of palmitoyl chains bound to free cysteine side chains was also examined CSS-Palm (Ren *et al.*, 2008).

RESULTS

Finding of essential genes

The database of essential gene was used for homologous sequence search of individual fungal proteins using BLASTp options. Based on the cutoff parameters, we shortlisted 538 essential proteins which are associated with various structural and physiological functions.

Orthology analysis with *Homo sapiens*

Each essential protein sequence of *Candida albicans* was analyzed for sequence homology with human genome using standard human BLASTp at the NCBI server with an E-value cutoff of 1. The comparative analysis was performed to exclude the human homolog proteins and to retain non-human homolog protein sequences. 464 genes ofto tal 538 protein sequences were
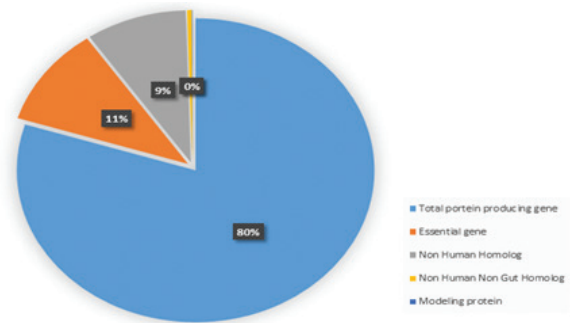

Figure 2. Screening of potential drug targets in *Candida albicans* by sequential systems biology analysis

found to be non-human homologous sequences, while 160 genes of these 464 were found to be conserved.

Comparative analysis with human microbiota

The most commonly used antibiotics kill pathogenic microbiota and healthy microbial flora non-selectively. This had a profound and prolonged harmful impact on normal gut flora (Jernberg, Löfmark, Edlund, & Jansson, 2010), so here efforts to design novel drugs that specifically targetthe fungus proteins. The short listed 464 essential proteins were compared with the gut flora by sequence analysis. The outcomes of sequence homolog analysis revealed that 20 (4.4%) were non-homologous to the gut flora (Fig. 2), and these may be further used for interactomic analysis and designing of drug ligand to cure the disease. The complete list of putative targets is provided as a supporting document (Supplementary Table 1 at www.actabp.pl).

Interactomic network

To understand and analyze the topology and functional annotation of protein-protein interaction in *Candida albicans*a protein networks was constructed (20-proteins) that contained 111-nodes and 122-edges. The entire protein network consisted of high confidence scoring partners (STRING relevant confidence score >0.8). The main component of this network connected nodes and edges (nodes represent protein and edges represent interaction). In Fig. 3, the interaction network was largely segregated into three neighborhoods: one enriched for modelled proteins shown in dark green among light green nodes representing potential drug targets and blue nodes indicates the interacting proteins with source proteins. The network was analyzed by Cytoscape network statistics with undirected score based on the centrality relationship among nodes.

The topological analysis of this network revealed the direct interaction of these critical and essential proteins that can be probable drug targets. 20 proteins were mapped to the PPI network. The visual analysis of this network showed that not only these essential proteins interact with each other and but were involved also in biosynthesis of amino acids signal transduction systems, nucleic acid synthesis, ribosomal proteins-functioning and other physiological transportation of chemicals and neurotransmitter.

Prediction of sub-Cellular localization

We predicted sub-cellular localization of the putative 20 target proteins. Among these targets, 80% of the
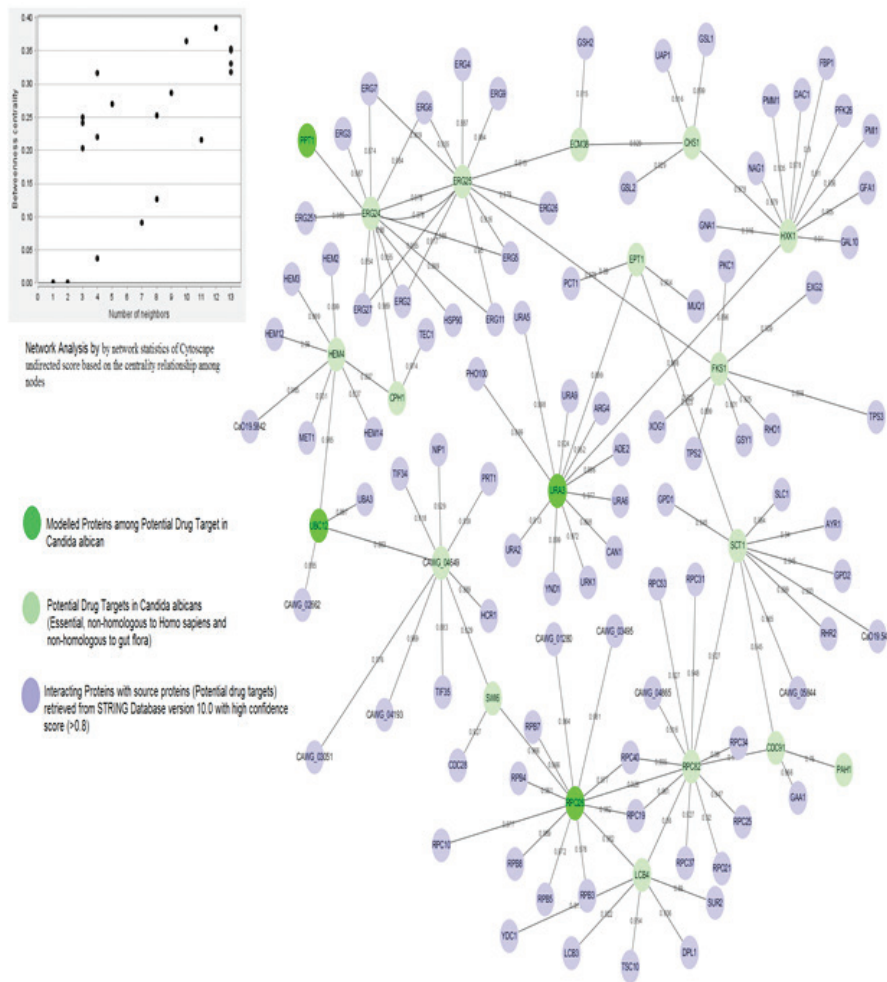
Figure 3. Protein interaction network. Dark nodes represent the modelled proteins, light green shows the potential drug targets of *Candida albicans*, and blue nodes indicates the interactors retrieved from STRING database version 10.0 with high confidence score (>0.8).

proteins belonged to the cytoplasmic compartment, followed by the membrane associated protein (10%) and extracellular proteins (10%) (Fig. 4).

Identifying pathways of potential drug targets

The role of the 20 sorted proteins in the metabolic pathways of *Candida albicans* was studied and correlated by KEGG database (Fig. 5). An illustrated map of the metabolic network was created showing system level in-

Figure 4. Prediction of sub-cellular localization of potential drug targets of *Candida albicans*.

vestigations of this fungus'cell machinery. The 20 inputs genes encode enzymes important in various biochemical reactions related to 40 pathways (KEGG database). These reactions include carbohydrate metabolism, protein synthesis, and its degradation, two component systems, phosphotransferase system, cofactor biosynthesis, DNA synthesis and others.

Protein modeling and analysis of posttranslational modification

Among the potential drug targets of *Candida albicans*, four proteins including RNA polymerase I, orotidine 5 phosphate decarboxylase, ubiquitin-like protein-conjugating E2 enzyme and serine-theonine phosphatase5 were selected based on the maximum homology of protein models and critical analysis of the pathways. The protein models of these potential targets were validated by the Q-Mean and Z-score which shows that the geometry and conformation of these protein models are stable (Fig. 6).

The post translation modifications showed that according to the NetAcet server, for Sequence 5 (Orotidine 5-phosphate decarboxylase), sequence 17 (ubiquitin-like protein-conjugating E2 enzyme) and sequence 20 (delta14-sterol reductase), acetylationis not possible as there is no alanine, glycine threonine and serine at this position. The sequence 12 (serine/threonine-protein
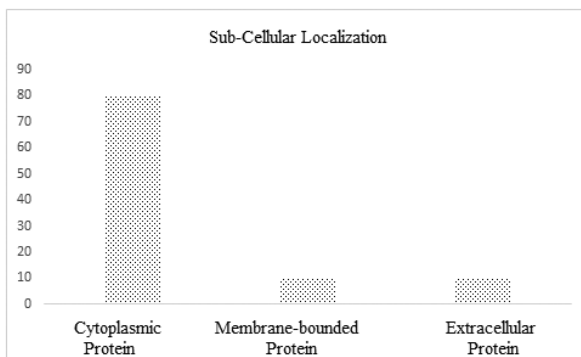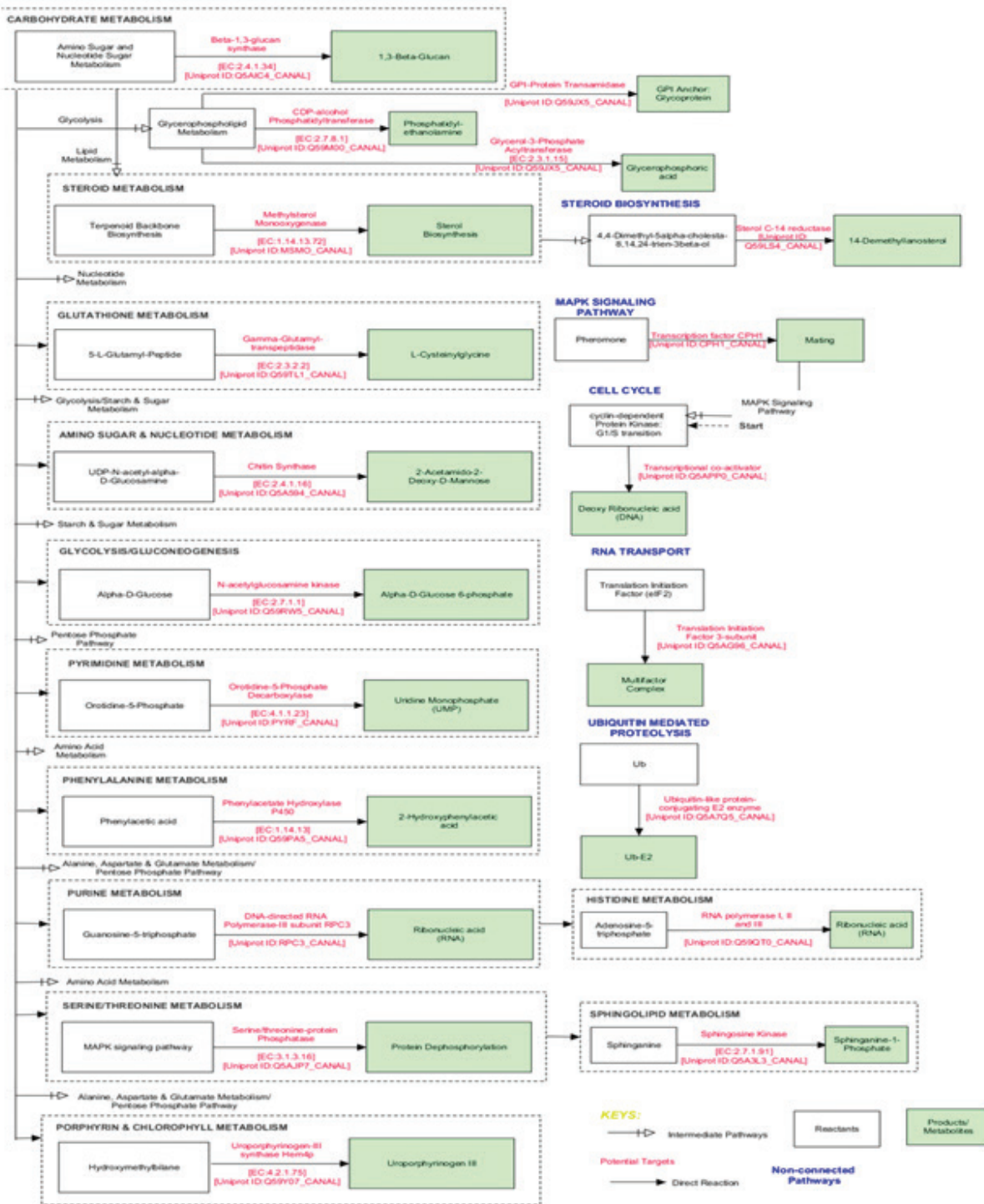
Figure 5. Integrated pathways analysis in *Candida albicans*.

phosphatase 5) and sequence 14 (RNA polymerase I) show positive results as acetylation is possible in both proteins. Similarly, According to NetAcet server, Sequence 5 (hexokinase), sequence 17 (ubiquitin-like protein-conjugating E2 enzyme) and sequence 20 (delta14-sterol reductase), the acetylation of this protein is not possible as there of no alaline, glycine threonine and serine at this position. The sequence 12 (serine/threonine-protein phosphatase 5) and sequence 14 (RNA polymerase I) show positive results as acetylation is possible in both proteins.For each sequence, potential glycosylation

sites or location are identified by using prediction confidence of predicted score higher than 0.5. The Net-Phos prediction indicated about the presence of serine threonine and tyrosine phosphorylation sites in our protein for specific and general kinases. This site predicts for following kinases: ATM CKI CKII CAMII DNAPK EGFR GSK3 INSR PKA PKB PKC RSK SRC cdc2 cdk5 p38MAPK. We predicted that serine/ threonine residues of protien are used to predict it O-GlcNAcylated as well as phosphorylated in YinOYang prediction, these sites are acylated reversibly at different time for
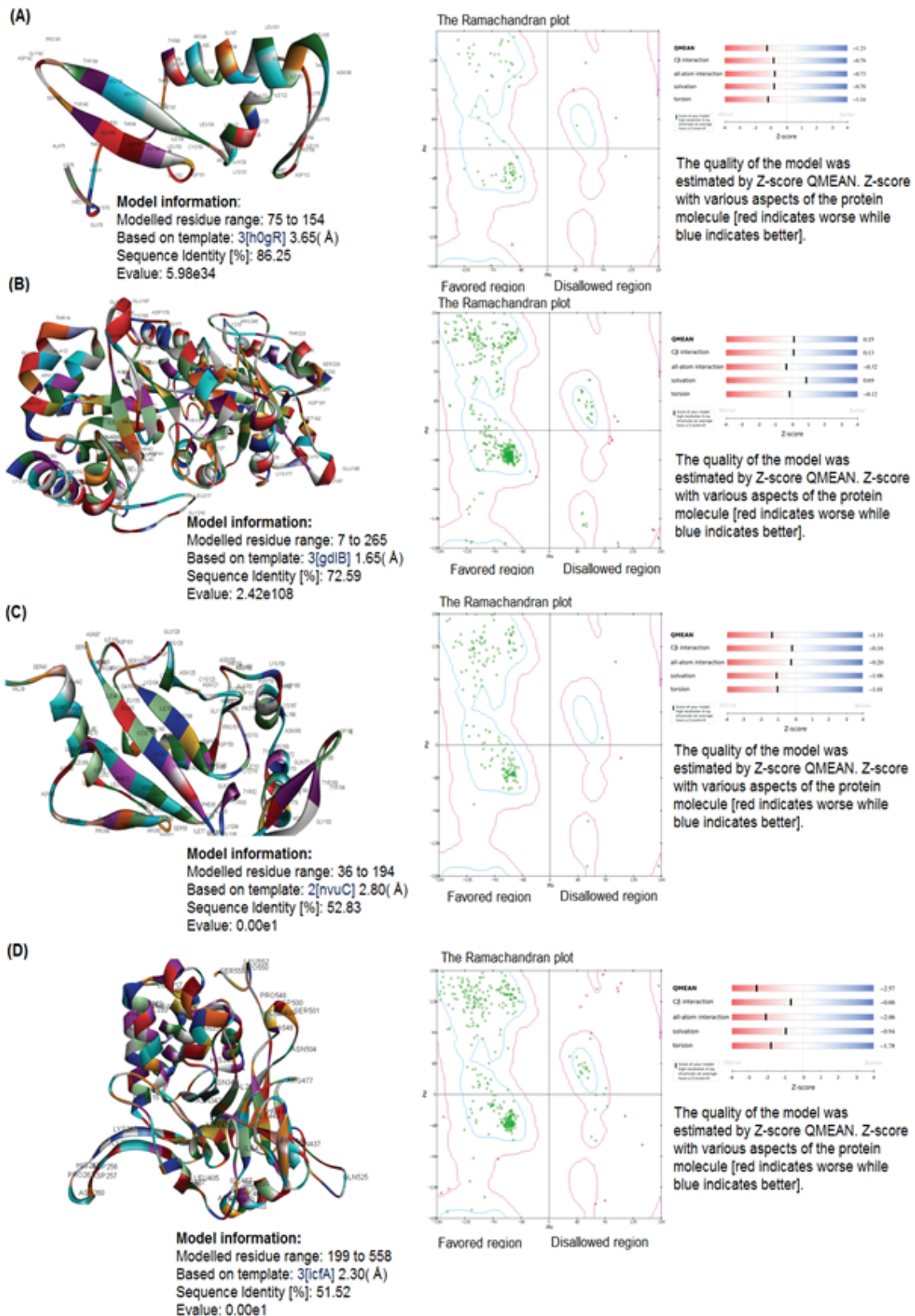
(A)

Model information:
Modelled residue range: 75 to 154
Based on template: 3[h0gR] 3.65(Å)
Sequence Identity [%]: 86.25
Evalue: 5.98e34

The Ramachandran plot

Favored region　　Disallowed region

The quality of the model was estimated by Z-score QMEAN. Z-score with various aspects of the protein molecule [red indicates worse while blue indicates better].

(B)

Model information:
Modelled residue range: 7 to 265
Based on template: 3[gdIB] 1.65(Å)
Sequence Identity [%]: 72.59
Evalue: 2.42e108

The Ramachandran plot

Favored region　　Disallowed region

The quality of the model was estimated by Z-score QMEAN. Z-score with various aspects of the protein molecule [red indicates worse while blue indicates better].

(C)

Model information:
Modelled residue range: 36 to 194
Based on template: 2[nvuC] 2.80(Å)
Sequence Identity [%]: 52.83
Evalue: 0.00e1

The Ramachandran plot

Favored region　　Disallowed region

The quality of the model was estimated by Z-score QMEAN. Z-score with various aspects of the protein molecule [red indicates worse while blue indicates better].

(D)

Model information:
Modelled residue range: 199 to 558
Based on template: 3[icfA] 2.30(Å)
Sequence Identity [%]: 51.52
Evalue: 0.00e1

The Ramachandran plot

Favored region　　Disallowed region

The quality of the model was estimated by Z-score QMEAN. Z-score with various aspects of the protein molecule [red indicates worse while blue indicates better].

Figure 6. Top four predicted protein models by Swiss-Model server from A–D.
Maximum sequence identity percentage shows the strength of the predicted model. The quality of the model was estimated by Z-score and Q-MEAN. These Ramachandran plots of top four predicted protein models wereconstructed by Drug Discovery Studio version 3.0 and indicated that the amino acid residues occur in the "favored region" (saturated green nodes) and few in "disallowed region" of the plot. For a good protein model there must be ≥90% amino acid residues in the most favored region or <2% in the disallowed region of the plot. (**A**) RNA polymerase I (**B**) Orotidine 5-phosphate decarboxylase (**C**) Ubiquitin-like based protein-conjugating E2 enzyme (**D**) Serine/threonine phosphatase 5.

Table 3: Analysis of post translational modification (PTM) of modelled proteins of *Candida albicans*.

Serial No.	Name of sequence	NetAcet	NetGlyC	NetOGlyc	NetPhos	YinYong
1	Orotidine 5 phosphate	No	3 positive locations	4 positive locations	Total 67 kinase active sites: 29 Unspecified kinases sites, 9 CKI sites, 11 PKC sites, 7 cdc sites.	Nil
2	Serine/threonine-protein phosphatase 5	3 S -MSSPKE 0.512 YES	5 positive locations	17 positive locations	Total 69 kinase sites: 20 unspecified kinase, 4 egfr sites, 10 ck sites 13 PKC sites	Nil
3	RNA polymerase I, II and III	2 S --MSDYE 0.531 YES	No sites	9 Positive locations	Total 29 sites: 12 unspecified kinase	Nil
4	Ubiquitin-like protein-conjugating E2 enzyme	No	1 positive location	9 positive locations	Total 26 sites 10 unspecified kinase,	Nil

the different function performed in the cell. None of the proteins in our scheme showed YinOYong positive results (Table 3).

DISCUSSION/CONCLUSION

Alternative drug targets are needed for nosocomial *Candida albicans* infections, as this fungus has become resistant to all therapies currently available (Arias & Murray, 2009; Boyanova *et al.*, 2007). One innovative way is to target genes essential for fungal survival. The product of such essential gene could become a new antifungal drug target (Materi & Wishart, 2007). In this study, we found that in *Candida albicans* 20 proteins were potential targets which were non-homologous to human genes as well as the gut flora. These proteins are recommended as drug targets having minimum side effects. By having a deep knowledge on a biological system by using genomic databases and system biological studies we can formulate better drugs. These systems help us to explore various pathways to find the most appropriate drugs by exploration of pathways and whole cellular systems which are involved in given disease. Computational systems help us in assessing the criticalness of any individual protein by studying the involvement of that proteins in alternative pathways or to the alternative pathway mechanismsthat compensate for that protein. System level studies are making it clearer day by day that a whole new plethora of techniques may be used for simulation and analysis of biochemical systems (von Mering *et al.*, 2007; Szklarczyk *et al.*, 2011).

The protein network map shows the behavior of essential and non-essential proteins. Protein-protein interaction are comprehensive due to availability of experimentally mapped interactionsin the integrated databases (Perumal *et al.*, 2007). Along with the essentiality of a selected protein target for *Candida albicans*. The other most common requirement is non-similarity with human proteins. Inhibition of the human proteins could lead to adverse effects. Querying the selected proteins against human proteome data at NCBI and UNIPROT databases was simplest and efficient way to achieve this. Sequence information for hundreds of eukaryotes/prokaryotes is available and this kind of analysis has been reported for *Mycobacterium tuberculosis*, *Pseudomonas* and *Helicobacter* species (Singh *et al.*, 2006; Beaugerie & Petit, 2004).

The role of normal human flora is ecological balance of the body system along with controlling the human mood and immunity (O'Hara & Shanahan, 2006; Hugot, 2004; Payne *et al.*, 2003; Carman *et al.*, 2004). The normal flora usually resides in digestive tract, upper respiratory tract and otic region of the human body and perform immunity control, mood regulation, synthesis of vitamins and prevention of superinfections. The metabolic activities performed by microflora of any species is now considered equal to a virtual organ which might be named as a forgotten organ (Guarner & Malagelada, 2003). Broad spectrum antibiotics affect the host's health by changing composition of the gut flora (Carman *et al.*, 2004). Due to this reason, non-human homologous sequences could not be our targets as many microbes are needed for balancing of ecology of the human body. The short listed 238 protein sequences which were non-human homologs were then compared with the gut flora and it was found that 20 of them were non-human and non-gut floral sequence homologs, leaving 87% of the non-homologous sequences to gut flora as our confident drug targets. The products of these selected genes were mapped to biological pathwaysand it was found these putative target proteins were involved in processes. The involvement of putative proteins in biological make them an essentiality for structure and survival ability of Candida. The gene ontology of twenty selected proteins indicated thatthese proteins weremostly cytoplasmic in origin, followed by mitochondrial and cell wall proteins (Muhammad *et al.*, 2014). Out of 20 proteins, 4were further modelled for the docking studies using a Swiss model template. The template with the highest quality has been identified and selected for the model building (Remmert *et al.*, 2011; Guex & Peitsch, 1997; Bordoli *et al.*, 2009; Sali & Blundell, 1993; Benkert *et al.*, 2011; Benkert *et al.*, 2009; Biasini *et al.*, 2014). The QMEAN scoring function was used for the assessment of the global and per-residue model. These homology models may be used for used for computational analysis of new ligand, ligand designing and ligand modification. The present antifungal drug can also be studied for their bonding affinities with these putative target proteins. These homology model are already being used for some antifungal docking studies (unpublished data). These protein models provide useful template for designing of novel drugs, as well as studies of old ligands against these putative targets. The homology models provide a good template to identify the selective targets in any species and selective toxicity of drug/ligand toward that species.

The selective toxicity of compound is very important as the natural human flora act as an organ of human body, providing nutrients like folic acid, are responsible for mood stabilization and have a commensal function to not allowing pathogenic species to grow. Further studies and a refined modeling algorithm will help us to design a tailor made drug towards a selected target of a given species

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Arias CA, Murray BE (2009) Antibiotic-Resistant Bugs in the 21st Century — A Clinical Super-Challenge. *New England Journal of Medicine* **360**: 439–443. https://doi.org/10.1056/NEJMp0804651

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195–201. https://doi.org/10.1093/bioinformatics/bti770

Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Kumar A (2011) *In silico* subtractive genomics for target identification in human bacterial pathogens. *Drug Develop Res* **72**: 162–177. https://doi.org/10.1002/ddr.20413

Bauer-Mehren A (2013) Integration of genomic information with biological networks using Cytoscape. *Methods in Mol Biol* **1021**: 37–61. https://doi.org/10.1007/978-1-62703-450-0-3

Beaugerie L, Petit J-C (2004) Microbial-gut interactions in health and diseaseAntibiotic-associated diarrhoea. *Best Practice Res Clin Gastroenterol* **18**: 337–52. https://doi.org/10.1016/j.bpg.2003.10.002

Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**: 343–350. https://doi.org/10.1093/bioinformatics/btq662

Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res* **37** (Suppl 2). https://doi.org/10.1093/nar/gkp322

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Schwede T (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**: 252–258. https://doi.org/10.1093/nar/gku340

Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**: 1351–62. https://doi.org/10.1006/jmbi.1999.3310

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protocols* **4**: 1–13. https://doi.org/10.1038/nprot.2008.197

Boyanova L, Kolarov R, Mitov I (2007) Antimicrobial resistance and the management of anaerobic infections. *Expert Rev Anti-Infective Ther* **5**: 685–701. https://doi.org/10.1586/14787210.5.4.685

Braun BR, van het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, Nantel A (2005) A human-curated annotation of the *Candida albicans* genome. *PLoS Genetics* **1**: 0036–0057. https://doi.org/10.1371/journal.pgen.0010001

Carman RJ, Simon MA, Fernández H, Miller MA, Bartholomew MJ (2004) Ciprofloxacin at low levels disrupts colonization resistance of human fecal microflora growing in chemostats. *Regul Toxicol Pharmacol* **40**: 319–326. https://doi.org/10.1016/j.yrtph.2004.08.005

Chan JNY, Nislow C, Emili A (2010) Recent advances and method development for drug target identification. *Trends Pharmacol Sci* **31**: 82–88. https://doi.org/10.1016/j.tips.2009.11.002

Fischbach MA, Walsh CT (2010) NIH Public Access **325** (5944) 1089–1093. https://doi.org/10.1126/science.1176667.Antibiotics

Garcı S (2004) *Candida albicans*. *Society* **3**: 536–545. https://doi.org/10.1128/EC.3.2.536

Guarner F, Malagelada J-R (2003) Gut flora in health and disease. *Lancet* **360**: 512–519. https://doi.org/10.1016/S0140-6736 (03)12489-0

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723. https://doi.org/10.1002/elps.1150181505

Gupta R, Jung E, Brunak S (2002) Prediction of N-glycosylation sites in human proteins. *Pacific Symposium on Biocomputing* **7**: 310–322

http://fungidb.org/common/downloads/release-2.0/Calbicans_SC5314/fasta/data/ (n.d.) No Title.

https://www.expasy.org/ (n.d.) No Title. Retrieved from https://www.expasy.org/

Hugot J-P (2004) Inflammatory bowel disease: a complex group of genetic disorders. *Best Practice Res Clin Gastroenterol* **18**: 451–462. https://doi.org/10.1016/j.bpg.2004.01.001

Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Brunak S (2002) Prediction of human protein function from posttranslational modifications and localization features. *J Mol Biol* **319**: 1257–1265. https://doi.org/10.1016/S0022-2836 (02)00379-0

Jernberg C, Löfmark S, Edlund C, Jansson JK (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* **156**: 3216–3223. https://doi.org/10.1099/mic.0.040618-0

Ji YD (2002) The role of genomics in the discovery of novel targets for antibiotic therapy. *Pharmacogenomics* **3**: 315–323. https://doi.org/doi 10.1517/14622416.3.3.315

Kabir MA, Hussain MA, Ahmad Z (2012) *Candida albicans*: a model organism for studying fungal pathogens. *Int Scholar Res Network Microbiol* **2012**: 1–15. https://doi.org/10.5402/2012/538694

Kiemer L, Bendtsen JD, Blom N (2005) NetAcet: Prediction of N-terminal acetylation sites. *Bioinformatics* **21**: 1269–1270. https://doi.org/10.1093/bioinformatics/bti130

Lewis GDWand JD (2013) Analysis of the human gut microbiome and association with diseaseblic access. *Clin Gastroenterol Hepatol* **11**: 774–777. https://doi.org/10.1016/j.cgh.2013.03.038.Analysis

Lin J, Qian J (2007) Systems biology approach to integrative comparative genomics. *Expert Rev Proteomics* **4**: 107–19. https://doi.org/10.1586/14789450.4.1.107

Lu Z, Szafron D, Greiner R, Lu P, Wishart DS, Poulin B, Eisner R (2004) Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* **20**: 547–556. https://doi.org/10.1093/bioinformatics/btg447

Luo H, Lin Y, Gao F, Zhang CT, Zhang R (2014) DEG 10 an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res* **42**: 574–580. https://doi.org/10.1093/nar/gkt1131

Marton MJ, DeRisi JL, Bennett Ha Iyer VR, Meyer MR, Roberts CJ, Friend SH (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* **4**: 1293–1301. https://doi.org/10.1038/3282

Materi W, Wishart DS (2007) Computational systems biology in drug discovery and development: methods and applications. *Drug Discovery Today* **12**: 295–303. https://doi.org/10.1016/j.drudis.2007.02.013

Materi W, Wishart DS (2007) Computational systems biology in drug discovery and development: Methods and applications. *Drug Discov Today* **12**: 295–303. https://doi.org/10.1016/j.drudis.2007.02.013

Mdluli K, Spigelman M (2006) Novel targets for tuberculosis drug discovery. *Curr Opin Pharmacol* **6**: 459–467. https://doi.org/10.1016/j.coph.2006.06.004

Meiller TF, Hube B, Schild L, Shirtliff ME, Scheper MA, Winkler R, Jabra-Rizk MA (2009) A novel immune evasion strategy of *Candida albicans*: Proteolytic cleavage of a salivary antimicrobial peptide. *PLoS One* **4**. https://doi.org/10.1371/journal.pone.0005039

Muhammad SA, Ahmed S, Ali A, Huang H, Wu X, Yang XF, Chen J (2014) Muhammad_14 Prioritizing drug targets in *Clostridium botulinum* with a computational systems biology approach. *Genomics*. https://doi.org/10.1016/j.ygeno.2014.05.002

O'Hara AM, Shanahan F (2006) The gut flora as a forgotten organ. *EMBO Rep* **7**: 688–693. https://doi.org/10.7400731. [pii]\r10.1038/sj.embor.7400731

Payne S, Gibson G, Wynne A, Hudspith B, Brostoff J, Tuohy K (2003) *In vitro* studies on colonization resistance of the human gut microbiota to *Candida albicans* and the effects of tetracycline and *Lactobacillus plantarum* LPK. *Curr Issues Intest Microbiol* **4**: 1–8

Perumal D, Lim CS, Sakharkar KR, Sakharkar MK (2007) Differential genome analyses of metabolic enzymes in *Pseudomonas aeruginosa* for drug target identification. *In Silico Biol* **7**: 453–465

Pucci MJ (2006) Use of genomics to select antibacterial targets. *Biochem Pharmacol*. https://doi.org/10.1016/j.bcp.2005.12.004

Raman K, Chandra N (2008) Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol* **8**. https://doi.org/10.1186/1471-2180-8-234

Rausch F, Schicht M, Br??uer L, Paulsen F, Brandt W (2014) Protein modeling and molecular dynamics simulation of the two novel surfactant proteins SP-G and SP-H. *J Mol Model* **20**: 2513. https://doi.org/10.1007/s00894-014-2513-0

Remmert M, Biegert A, Hauser A, Söding J (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**: 173–175. https://doi.org/10.1038/nmeth.1818

Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X (2008) CSS-Palm 2.0: An updated software for palmitoylation sites prediction. *Protein Engin Design Select* **21**: 639–644. https://doi.org/10.1093/protein/gzn039

Riesbeck K (2013) *Candida albicans* is a crafty microbe that deceives its host by using complement regulators and proteases. *J Infect Dis* **207**: 550–552. https://doi.org/10.1093/infdis/jis722

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779–815. https://doi.org/10.1006/jmbi.1993.1626

Schulze J, Sonnenborn U (2009) Pilze im Darm – von kommensalen Untermietern zu Infektionserregern. *Dtsch Ärztebl* **106**: 837–842. https://doi.org/10.3238/arztebl.2009.0837 (in German)

Scully C, EI-Kabir M, Samaranayake LP (1994) Candida and Oral Candidosis: A Review. *Crit Rev Oral Biol Med* **5**: 125–157. https://doi.org/10.1177/10454411940050020101

Singh NK, Selvam SM, Chakravarthy P (2006) T-iDT: tool for identification of drug target in bacteria and validation by Mycobacterium tuberculosis. *In Silico Biol* **6**: 485–493. https://doi.org/2006060045

Soll DR, Staebell M, Langtimm C, Pfaller M, Hicks J, Rao TVG (1988) Multiple Candida strains in the course of a single systemic infection. *J Clin Microbiol* **26**: 1448–1459

Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT.-BG, Clausen H (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**: 1478–88. https://doi.org/10.1038/emboj.2013.79

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Von Mering C (2011) The STRING database in 2011: Functional interaction networks of proteins globally integrated and scored. *Nucleic Acids Res* **39** (Suppl 1). https://doi.org/10.1093/nar/gkq973

Volkswirtschaftliche Z (2005) Innovation in the Pharmaceutical Industry — Future Prospects. *Roche: Prevention* 1–8

von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krüger B, Bork P (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35** (Database issue) D358-D362. https://doi.org/10.1093/nar/gkl825

Zhang R, Lin Y (2009) DEG 5.0 a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res* **37** (Suppl 1) 455–458. https://doi.org/10.1093/nar/gkn858

Zhang R, Ou H.-Y, Zhang C.-T (2004) DEG: a database of essential genes. *Nucleic Acids Res* **32** (Database issue) D271-D272. https://doi.org/10.1093/nar/gkh024