ISSN Online: 2349-7785

Functional Elucidation and Structure Prediction of Certain Hypothetical Proteins in *Candida glabrata* CBS 138: an *In silico* Approach

Shilpee Pal [1], Sanghati Bhattacharya [2], Arnab Sen [2], Bikash Ranjan Pati [1], Keshab Chandra Mondal [1], Pradeep Kumar DasMohapatra [1]*

[1] Bioinformatics Infrastructure Facility Centre, Department of Microbiology, Vidyasagar University, Midnapore 721 102, India
[2] Department of Botany, School of Life Sciences, University of North Bengal, Darjeeling 734 013, India

Received 27 October 2015; accepted in revised form 20 November 2015

Abstract: Transition from an opportunistic pathogen to a deadly disease forming one, depends on a number of aspects, which activate causative protein production and their functions. In recent world, most of the proteins are still functionally unknown i.e. hypothetical proteins (HPs) and are not negligible. In the present study, functional categorization and structure prediction of 20 HPs of *Candida glabrata* was performed. Their physicochemical properties, functional domains, as well as interacted proteins were predicted and analyzed. Signal peptide, subcellular localizers, transmembrane regions were determined to explore physical characteristics of selected HPs. They were involved in various important functions such as nuclear-vacuolar junction, metabolic pathways, ATPase execution, DNA binding, mitochondrial transportation, amino acid biosynthesis and catabolism, vesicular transportation, extracellular, DNA repair and cell cycle control, nuclear functions, mitochondrial RNA synthesis and translation, RNA synthesis, glucose transportation and Ca^{2+} ion exchanger. 3D structure of each HP was determined and energy minimization was done by using GROMOS96 force field implicated in Swiss-Pdb Viewer. Ligand binding sites of the HPs showed the active regions, involved in functional modulation. Thus *in silico* analysis of HPs is more easy to reveal the structures and functions, which was experimentally very expensive and tedious. It would also be helpful in recognizing the mechanism of pathogenesis and in developing therapeutic drug molecules and their docking studies.

Key words: *Candida glabrata***,** hypothetical proteins, domains, subcellular localizer, functional categorization, structure prediction, active site.

Introduction

Candida glabrata, earlier known as *Cryptococcus glabrata*, identified by Anderson in 1917 and renamed as *Torulopsis glabrata* in 1938 [1] is becoming second highest frequent cause of 'candidiasis' as *Candida albicans* [2] and showed equal mortality rate [3]. The genus *Candida* consist of more than 150 species [4] among which *C. glabrata* is a normal commensal of human gastro-intestinal tract, oral cavity, alimentary tract, respiratory tract [5]. Phylogenetically it is closely related with *Saccharomyces cerevisiae* [6] and comparatively less pathogenic agent with respect to *C. albicans*. Naturally, *Cryptococcus glabrata* does not infect its host but takes advantages of impaired host immune system and causes fungal infections (candidiasis) of skin, oropharyngeal, esophageal, bloodstream infections, etc.[7]. Transition from normal commensal to pathogenic one is an outcome of various aspects such as

*Corresponding author (Pradeep Kumar DasMohapatra)
E-mail: < pkdmvu@gmail.com >

© 2015, Har Krishan Bhalla & Sons

environmental conditions, host immune system, host-pathogen interactions, pathogen's adaptation efficiency, etc.[8]. These situations actually modulate the pathogenesis causing proteins. From this point of view we cannot neglect hypothetical proteins (HPs), as their functions are still unknown [9]. Revelation of these proteins by analyzing their structure and functions may improve the knowledge about the pathogen.

An interesting fact is, in living world, almost half of the proteins belong to HP [10]. According to IMG-JGI database (https://img.jgi.doe.gov/), complete genome of *C. glabrata* has already been sequenced but very little amount of proteins have been characterized. During study of pathogenesis, characterization of HPs can explore their roles in a number of functional pathways. Structure determination will help to reveal the active sites where ligands bind and accelerate proteins function. It will also help to predict drug molecules against pathogen as well as in docking studies.

Materials and methods

Sequence retrieval

Total 20 hypothetical protein sequences of *C. glabrata* CBS 138 were randomly selected and retrieved from IMG-JGI database (https://img.jgi.doe.gov). Structural as well as functional properties of the selected HPs were analyzed by using several bioinformatics tools and databases.

Physicochemical properties analysis

Physicochemical properties such as amino acid length, molecular weight, iso-electric point (pI), instability index, aliphatic index and hydrophobicity were calculated by ProtParam tool of ExPASy server (web.expasy.org/protparam/) and showed in Table 1.

Motifs and domains prediction

The conserved domains and motifs of HPs were predicted by using several databases like Conserved Domain Database(CDD-BLAST),

Table 1. Physiochemical properties of selected HPs

Accession No.	Amino acid number	Molecular weight	pI	Positively charged residues	Negatively charged residues	Instability ndex	Aliphatic ndex	Grand average of hydropathy (GRAVY)
XP_444880	1049	119916.6	9.30	163	138	53.57	73.85	-0.878
XP_444843	568	62918.1	8.09	48	45	36.85	92.31	0.357
XP_444844	563	62315.2	7.81	43	41	33.35	89.15	0.318
XP_444845	533	59869.7	8.47	45	39	35.96	93.21	0.322
XP_444846	934	104760.8	6.15	91	101	37.96	86.81	-0.321
XP_444847	794	85613.7	4.28	40	80	29.30	80.52	-0.096
XP_445514	868	96648.3	6.27	92	100	41.25	91.54	-0.452
XP_444860	552	61546.4	8.09	46	43	38.58	90.04	0.321
XP_444851	675	75493.7	8.17	92	89	34.73	101.01	-0.334
XP_444829	707	76529.7	5.59	70	91	30.53	85.64	-0.190
XP_444861	549	61215.1	8.27	46	42	37.65	89.65	0.337
XP_444859	605	69296.2	5.88	78	87	33.22	83.17	-0.396
XP_444820	801	88290.7	8.90	92	83	40.51	96.99	-0.100
XP_444856	548	61736.6	5.57	71	82	46.59	83.30	-0.385
XP_444788	784	86523.0	5.74	91	108	32.67	81.21	-0.368
XP_444794	902	98376.1	4.86	77	116	34.67	103.22	0.179
XP_444780	756	83646.4	8.67	91	85	57.28	59.30	-0.894
XP_444808	851	98674.3	9.19	108	82	33.72	103.49	-0.185
XP_444842	605	69153.8	9.05	82	71	34.98	94.38	-0.307
XP_444795	1122	122964.7	6.42	128	133	33.55	98.94	-0.033

Pfam, ScanProsite and TIGRFAMs.

CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) identifies domain in query protein sequence by searching CDD, which is linked with other databases such as PubMed, Entrez Protein and NCBI BioSystems, etc. The search algorithm is based on Reverse Position-Speciûc BLAST (RPS-BLAST), a variation of Position Specific Iterated Blast (PSI-BLAST) method to scan query sequence for conserved domains [11].

Pfam (http://pfam.sanger.ac.uk/) is an assembly of protein families, superfamilies, domains, Hidden Markov Models (HMMs), repeats, etc.[12].

Scan Prosite (http://www.expasy.org/tools/scanprosite/) identifies motif of query sequence by scanning remote homologues from PROSITE database. It provides signature sequences built by manually derived alignments and also delivers intra-domain topologies [13].

TIGRFAMs (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi) is a collection of manually curated protein families as well as superfamilies defined by HMMs and provides structurally as well as functionally conserved domains of full-length query protein sequence [14].

Domains with 100 % confidence level i.e. delivered by all above four tools were tabulated in Table 2.

Signal peptide identification

SignalP 4.1 Server [15] was used to detect signal peptides among the selected HPs.

Subcellular localization site determination

PSORT II (http://psort.hgc.jp/form2.html) was used to predict the subcellular localization of selected HPs. It is a new version of PSORT, considers the eukaryotic amino acid sequence as input [16]. Reinhardt's method was used for cytoplasmic/nuclear discrimination.

Transmembrane region prediction

Transmembrane Hidden Markov Models [TMHMM (at http://www.cbs.dtu.dk/services/TMHMM)] was used to predict transmembrane protein topologies in selected HPs. It is constructed by HMM and can predict 97-98 % correct transmembrane helices [17].

Protein-protein interaction prediction

Search Tool for the Retrieval of Interacting Genes/Proteins; [STRING (http://string-db.org/)] was used to evaluate protein-protein interaction of nominated HPs. In this database the co-occurrences and association between proteins are derived from statistical analyses [18].

Protein structure determination

Iterative Threading ASSEmbly Refinement [I-tasser (http://zhanglab.ccmb.med.umich.edu/I-tasser)] server was considered to predict 3D structures of selected HPs. It is an automated protein structure prediction tool which depends on threading analysis [19]. Total energy of each predicted structure was calculated using GROMOS96 (GROningen molecular dynamics simulation) force field associated in Swiss-Pdb Viewer [20] and energy minimization was performed to get optimum structure.

Active site determination

DogSiteScorer (http://dogsite.zbh.uni-hamburg.de/) was accessed to determine pockets in predicted structures. Pockets with highest scores were considered as active sites of selected HPs.

Results

Physicochemical properties of selected 20 HPs of *C. glabrata* CBS 138 revealed the length of proteins varied from 530bp to 1122bp (Table 1). pI of protein XP_444880, XP_444808, XP_444842, XP_444820, XP_444780, XP_444845, XP_444861, XP_444843, XP_444860 and XP_444844 was higher. The surface of protein XP_444880, XP_444808, XP_444851, XP_444820, XP_444780, XP_444842, XP_444843, XP_444861, XP_444860, XP_444845 and XP_444844 were more positively charged than other HPs (Table 1). *In vitro* stability of proteins were determined by calculating instability index, which was less than 40 (<40) for stable proteins. In present study, protein XP_444880, XP_444849, XP_444856 and XP_444780 showed instability index above 40. The volume occupied by aliphatic side chains (Ala, Val, Ile and Leu) were measured by calculating aliphatic index. Majority of the HPs

Table 2. Domain description of selected HPs

Functional domains	Descriptions	Hypothetical proteins
ZF_FYVE	The FYVE zinc finger was first found in Fab1, YOTB/ZK632.12, Vac1 and EEA1. The FYVE finger has eight potential zinc coordinating cysteine positions.	XP_444780
Aconitase	Aconitase (aconitate hydratase; EC: 4.2.1.3) is an iron-sulphur protein contains a [4Fe-4S]-cluster and catalyses the interconversion of isocitrate and citrate via a *cis*-aconitate intermediate.	XP_444788
E1-E2_ATPase	Transmembrane ATPases are membrane-bound enzyme complexes/ion transporters that use ATP hydrolysis to drive the transport of protons across a membrane.	XP_444794; XP_444795
RVT_1	Reverse transcriptase gene is usually indicative of a mobile element such as a retrotransposon or retrovirus.	XP_444808
ABC_membrane	Water-soluble domain of transmembrane ABC transporters	XP_444820
Trp_syntA	Tryptophan synthase catalyses the last step in the biosynthesis of tryptophan	XP_444829
AA_TRNA_LIGASE	Aminoacyl-tRNA synthetase (also known as aminoacyl-tRNA ligase) catalyses the attachment of an amino acid to its cognate transfer RNA molecule	XP_444842; XP_444856; XP_444859
Sugar_tr	ATP-binding cassette (ABC) superfamily and major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family.	XP_444845; XP_444844; XP_444843; XP_444860; XP_444861
GATA_ZN_FINGER	GATA-type zinc fingers (Znf), a transcription factors (including erythroid-specific transcription factor and nitrogen regulatory proteins), specifically bind the DNA sequence (A/T) GATA (A/G) in the regulatory regions of genes.	XP_444846
Collagen	Generally extracellular structural proteins involved in formation of connective tissue structure. The sequence is predominantly repeats of the G-X-Y and the polypeptide chains form a triple helix.	XP_444847
LRR_1	Composed of repeating 20–30 amino acid stretches that are unusually rich in the hydrophobic amino acid leucine.	XP_445514
IF2	Translation initiation factor 2 (IF2) promotes 30S initiation complex (IC) formation and 50S subunit joining, which produces the 70S IC.	XP_444851
SANT	SANT domain is a motif of ~50 amino acids present in proteins involved in chromatin-remodelling and transcription regulation.	XP_444880

displayed aliphatic index from 70 to 103 without protein XP_444780 (Table 1). GRAVY revealed the hydrophobicity of a protein. Grand Average hydropathy (GRAVY) of the studied HPs were varied between -0.878 to 0.357. A significant positive correlation was observed between aliphatic index and GRAVY ($r = 0.583$, $p<0.01$). For functional analysis of selected HPs, a number of parameters such as, protein domains, cleavage sites, transmembrane regions, subcellular localizations and protein-protein interactions were predicted and analyzed. Functional domains, determined by all four databases were ZF_FYVE, Aconitase, E1-E2_ATPase, RVT_1, ABC_membrane, Trp_syntA, AA_TRNA_LIGASE, Sugar_tr, GATA_ZN_FINGER, Collagen, LRR_1, IF2 and SANT (Table 2). Protein XP_444847 revealed signal peptide and the cleavage site was between 17th (Ala) and 18th (Phe) amino acids (Fig. 1). According to PSORTII, protein XP_444880, XP_444846, XP_445514, XP_444780 and XP_444808 were nuclear; XP_444829, XP_444859, XP_444856 and XP_444788 were cytoplasmic; XP_444843, XP_444844, XP_444845, XP_444860, XP_444861, XP_444794 and XP_444795 were membrane proteins; XP_444851 and XP_444842 were mitochondrial and protein XP_444820 was localizing in endoplasmic reticulum (Table 3). Protein XP_444843, XP_444844, XP_444845, XP_444860 and XP_444861 showed similar number of transmembrane regions (Table 4) localized in plasma membrane (Table 3) and possessed Sugar_tr domain (Table 2); protein XP_444794 and XP_444795 also localized in plasma membrane (Table 3) showed same number of transmembrane regions (Table 4) and comprised of E1-E2_ATPase domain (Table 2). Another transmembrane protein XP_444820 (Table 4) contained ABC_membrane domain (Table 2).

Protein-protein interactions of selected 20 HPs were shown in Fig. 2. Associating the above results, functional categorization was done (Table 5). HPs were involved in nuclear–vacuolar junction, metabolic pathways, ATPase execution, DNA binding, mitochondrial transport, amino acid

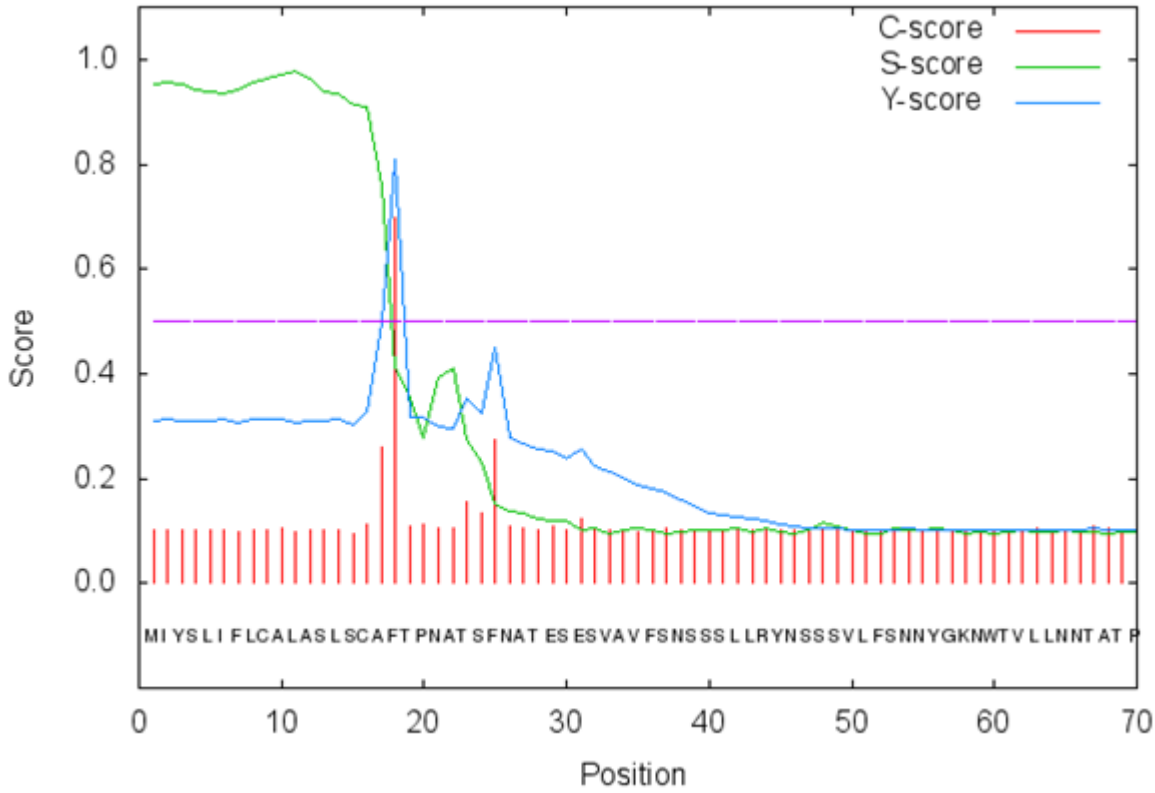

Figure 1. Protein XP_444847 showed max. C at position 18, max. Y at position 18 and mean S value between position 1-17. Cleavage site between positions 17 (Ala) and 18 (Phe)

Table 3. Subcellular localization of the HPs

Protein id	Localization
XP_444880	Nuclear
XP_444843	Plasma membrane
XP_444844	Plasma membrane
XP_444845	Plasma membrane
XP_444846	Nuclear
XP_444847	Extracellular
XP_445514	Nuclear
XP_444860	Plasma membrane
XP_444851	Mitochondrial
XP_444829	Cytoplasmic
XP_444861	Plasma membrane
XP_444859	Cytoplasmic
XP_444820	Endoplasmic reticulum
XP_444856	Cytoplasmic
XP_444788	Cytoplasmic
XP_444794	Plasma membrane
XP_444780	Nuclear
XP_444808	Nuclear
XP_444842	Mitochondrial
XP_444795	Plasma membrane

Table 4. Transmembrane regions of selected HPs

Sequence Id	N-terminal	C-terminal	Transmembrane region	Length
XP_444843	60	82	ASAYVTVSIFCLFIAFGGFVFGW	23
	116	135	NGTHYLSKVRTGLVVSIFNIGCAIGGVILS	20
	145	164	PGLIIVVVIYVVGIIIQIAT	20
	171	193	YFIGRIISGLGVGGIAVLSPMLI	23
	203	225	ATLVACYQLMITLGIFLGYCTNF	23
	238	257	VPLGLCFAWAIFMISGMTFV	20
	361	383	FETSIVIGVVNFFSTFVGIFLVG	23
	390	410	CLLWGAATMTACMVVFASVGV	21
	430	452	MIVFTCFYIFCFATTWAPLAFVI	23
	465	487	CMALAQASNWIWGFLISFFTPFI	23
	491	513	INFNYGYVFMGCLCFSYFYVFFF	23
XP_444844	54	73	AFVGVIISCFMVAFGGFVFG	20
	110	127	LIVSIFNIGCAIGGIILS	18
	137	156	MGLVVVVVIYIVGIIIQIAS	20
	163	185	YFIGRIISGLGVGGISVLSPMLI	23
	195	217	GSLVSCYQLMITLGIFLGYCTNF	23
	230	249	VPLGLCFAWALFMIGGMTFV	20
	352	374	SFETSIVFGVVNFFSTCCSLLTV	23
	381	403	NCLLYGAIGMVCCYVVYASVGVT	23

table 4. (continued).

Sequence Id	N-terminal	C-terminal	Transmembrane region	Length
	423	445	IVFSCFYIFCFATTWAPIAYVII	23
	457	479	AMSIATAANWMWGFLIAFFTPFI	23
	483	505	INFYYGYVFMGCMVFAYFYVFFF	23
XP_444845	25	44	LIFVSLCCIMVAFGGFVFGW	20
	78	100	TGLIVAIFNIGCAIGGITLSKLG	23
	107	124	LGLVTVVVVYTIGIVIQI	17
	134	156	FIGRIISGLGVGGIAVLSPMLIS	23
	163	185	LRGTLVSCYQLMITCGIFLGYCT	23
	200	219	VPLGLCFAWALFMIFGVMCV	20
	300	319	LTGANYFFYYGTTIFRAVGL	20
	323	342	FQTAIVLGVVNFVSTFYALY	20
	355	377	WGCVGMVCCYVVYASVGVTRLWP	23
	392	414	MIVFACFFIFCFATTWAPIAYVI	23
	427	449	AMSIAIAANWIWGFLIAFFTPFI	23
	453	475	INFYYGYVFMGCMVFAYFYVFFF	23
XP_444860	44	66	ASAYVSISIFCLFIAFGGFVFGW	23
	100	119	TGLVVSIFNIGCAIGGIVLS	20
	129	148	IGLISVVVIYIVGIVIQIAT	20
	155	177	YFIGRIISGLGVGGIAVLSPMLI	23
	187	209	GSLVSCYQLMITCGIFLGYCTNY	23
	222	241	VPLGLCFAWALFMIGGMTFV	20
	344	366	SFETSIVIGIVNFASTFVALYVV	23
	375	394	LLWGAAAMTACMVVFASVGV	20
	414	436	MIVFTCFYIFCFATTWAPIPFVV	23
	449	471	CMAIAQASNWIWGFLIAFFTPFI	23
	475	497	INFYYGYVFMGCLCFSYFYVFFF	23
XP_444861	41	63	ASAYVAISIFCLFIAFGGFVFGW	23
	97	116	TGLIVSIFNIGCAIGGVVLS	20
	126	145	IGLISVVVIYIVGIVIQIAT	20
	152	174	YFIGRIISGLGVGGIAVLSPMLI	23
	184	206	GSLVSCYQLMITCGIFLGYCTNY	23
	219	238	VPLGLCFAWALFMIGGMTFV	20
	341	363	SFETSIVIGIVNFASTFVALYVV	23
	372	391	LLWGAAAMTACMVVFASVGV	20
	411	433	MIVFTCFYIFCFATTWAPIPFVV	23
	446	468	CMAIAQASNWIWGFLIAFFTPFI	23
	472	494	INFYYGYVFMGCLCFSYFYVFFF	23
XP_444820	133	155	LLLTAVGLLTISCSIGMTIPKVI	23
	193	215	FLGGFALALLVGIAANYGRIILL	23
	272	294	GFKALICGSVGIGMMLALSPQLS	23
XP_444794	98	120	VKFVMFFVGPIQFVMEAAAILAA	23
	125	144	WVDFGVICGLLMLNACVGFI	20
	273	295	VLNGIGILLLVLVIVTLLGVWAA	23
	315	337	IIGVPVGLPAVVTTTMAVGAAYL	23

table 4. (continued).

Sequence Id	N-terminal	C-terminal	Transmembrane region	Length
	675	697	YVVYRIALSLHLELFLGLWIIIL	23
	703	722	IELIVFIAIFADVATLAIAY	20
	742	764	MSIILGIVLAIGTWICLTTMFLP	23
	807	829	WQLAGAVFAVDIIATMFTLFGWF	23
	842	861	IYIWSIGVFCVLGGFYYIMS	20
XP_444795	83	102	LLLTGAAVVSFTLGIYEVLT	20
	117	139	VDWIEGLAIMMAVLVVVLVSAAN	23
	307	329	ISVYGCVAAITLFVVLFARYLSY	23
	352	374	IFITAITVIVVAVPEGLPLAVTL	23
	854	876	FIQFQLIVNVTAVLLTFVTSVISS	23
	886	905	VQLLWVNLIMDTLAALALAT	20
	1022	1044	YFIFIMSLIAVLQVLIMFFGGAP	23
	1054	1076	MWLVSVSSGILAIPVGALIRICP	23

Table 5. Putative function of selected HPs

Functional categories	Hypothetical proteins
Nuclear–vacuolar junction	XP_444780
Metabolic pathways	XP_444788
ATPase execution	XP_444794
DNA binding	XP_444808
Mitochondrial transporter	XP_444820
Amino acid biosynthesis and catabolism	XP_444829
Vesicular transport	XP_444846
Extracellular	XP_444847
DNA repair and cell cycle control	XP_444880
Nuclear functions	XP_445514
Mitochondrial RNA synthesis and translation	XP_444851; XP_444842
Cellular RNA synthesis	XP_444856; XP_444859
Glucose transporter	XP_444845; XP_444844; XP_444843; XP_444860; XP_444861
Ca^{2+} ion exchanger	XP_444795

biosynthesis and catabolism, vesicular transport, extracellular, DNA repair and cell cycle control, nuclear functions, mitochondrial RNA synthesis and translation, cellular RNA synthesis, glucose transport and Ca^{2+} ion exchange. Template structures of predicted protein models were arranged in Table 6. Optimized 3D structures were predicted and shown in Fig. 3. Among them transmembrane protein XP_444843, XP_444844, XP_444845, XP_444860, XP_444861, XP_444820, XP_444794 and XP_444795 showed large amount of helical structure (Fig. 3). Signal peptide containing protein XP_444847 revealed large quantity of â-barrel (Fig. 3). Active sites of each HP were shown in Table 7.

Discussion

Physicochemical properties analysis uncovers the basic knowledge about the nature of proteins. A very important property i.e. pI revealed the pH

Table 6. Template structures of the HPs

Accession No	Template
XP_445514	4mn8
XP_444780	3gav
XP_444794	1mhs
XP_444795	3ba6
XP_444808	4c0o
XP_444842	1f7u
XP_444843	4gc0
XP_444844	4pyp
XP_444845	4gby
XP_444846	3efo
XP_444847	2pff
XP_444860	4pyp
XP_444880	3gaw
XP_444859	1f7u
XP_444861	4gby
XP_444820	4f4c
XP_444856	2xti
XP_444829	1k8z
XP_444788	1c96
XP_444851	3j4j

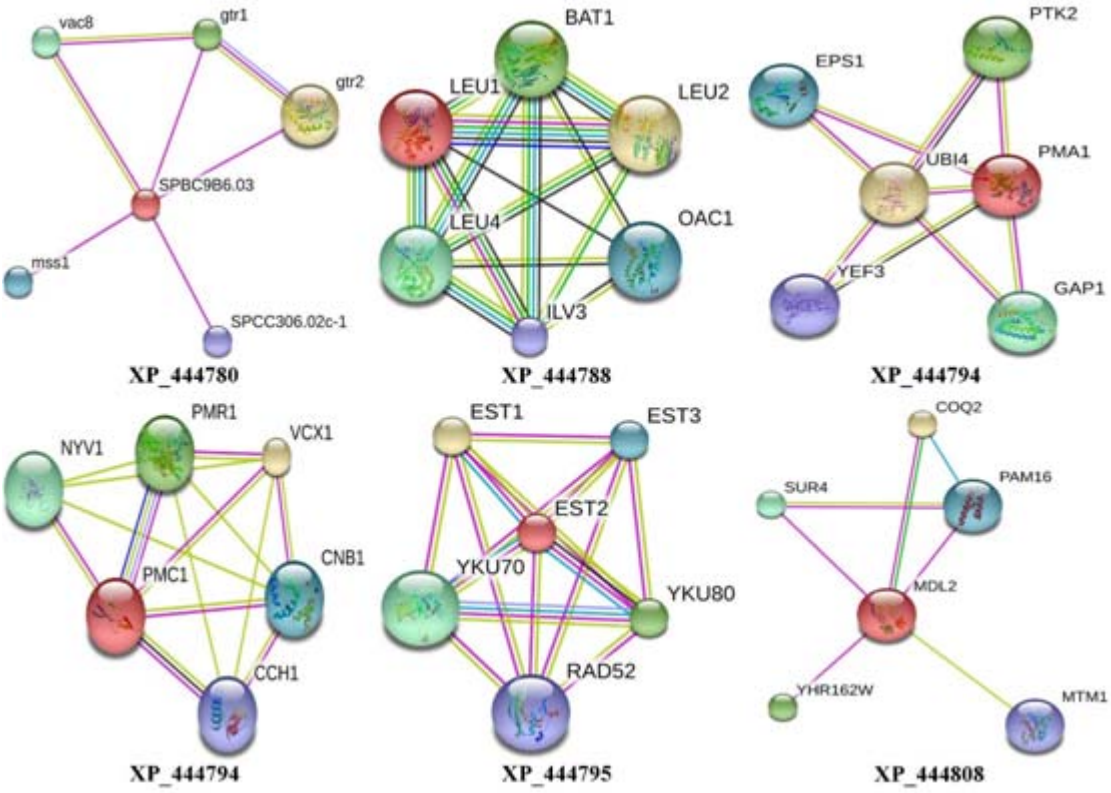

Figure 2. (continued).

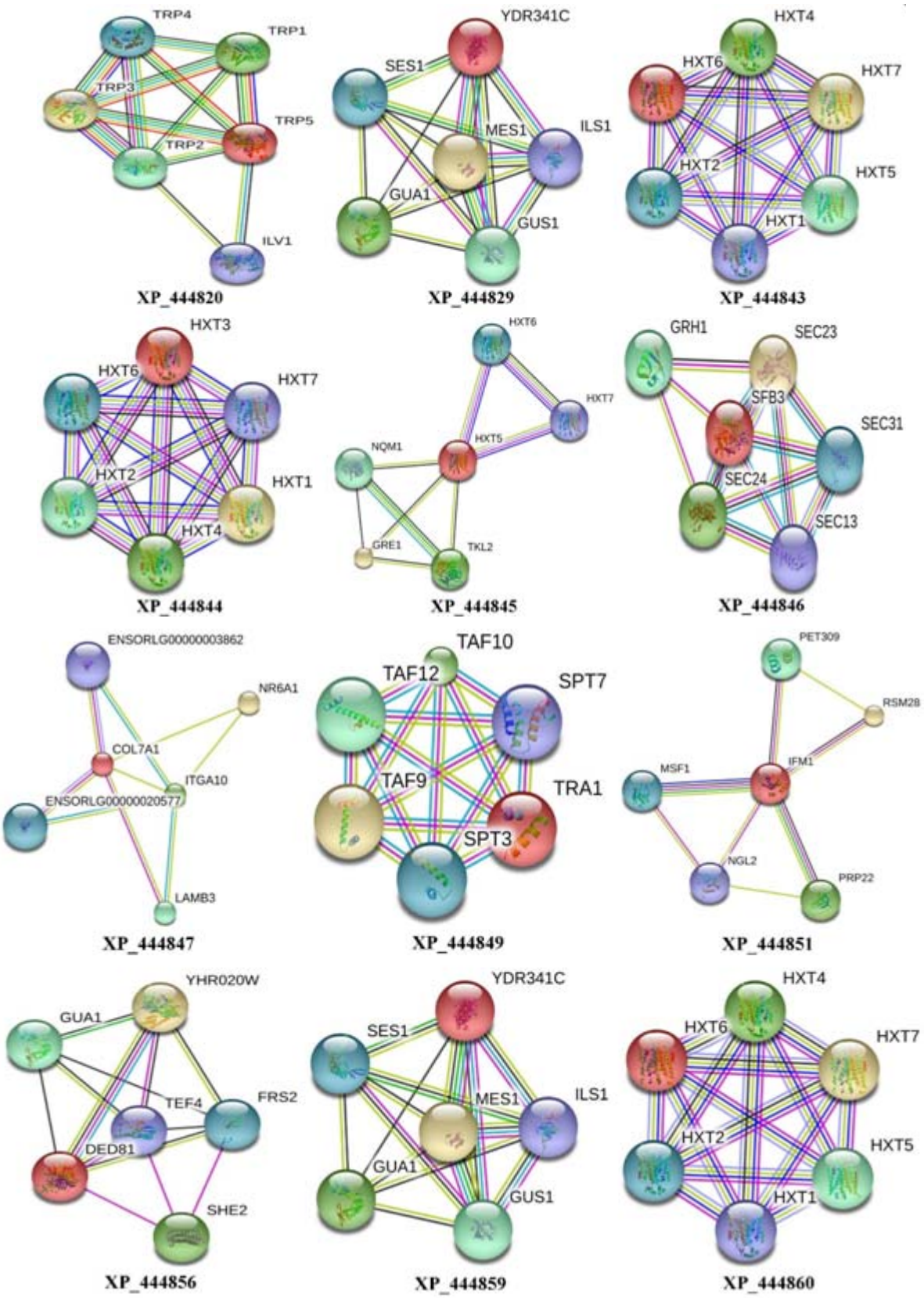

Figure 2. (continued).

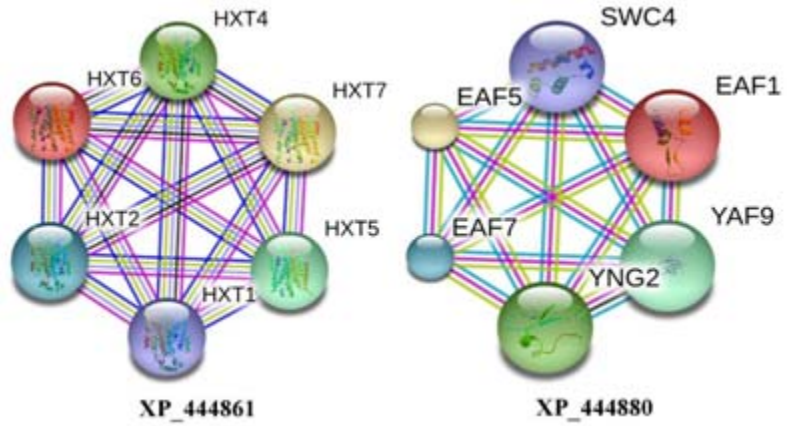

Figure 2. Protein-protein interaction of selected hypothetical proteins

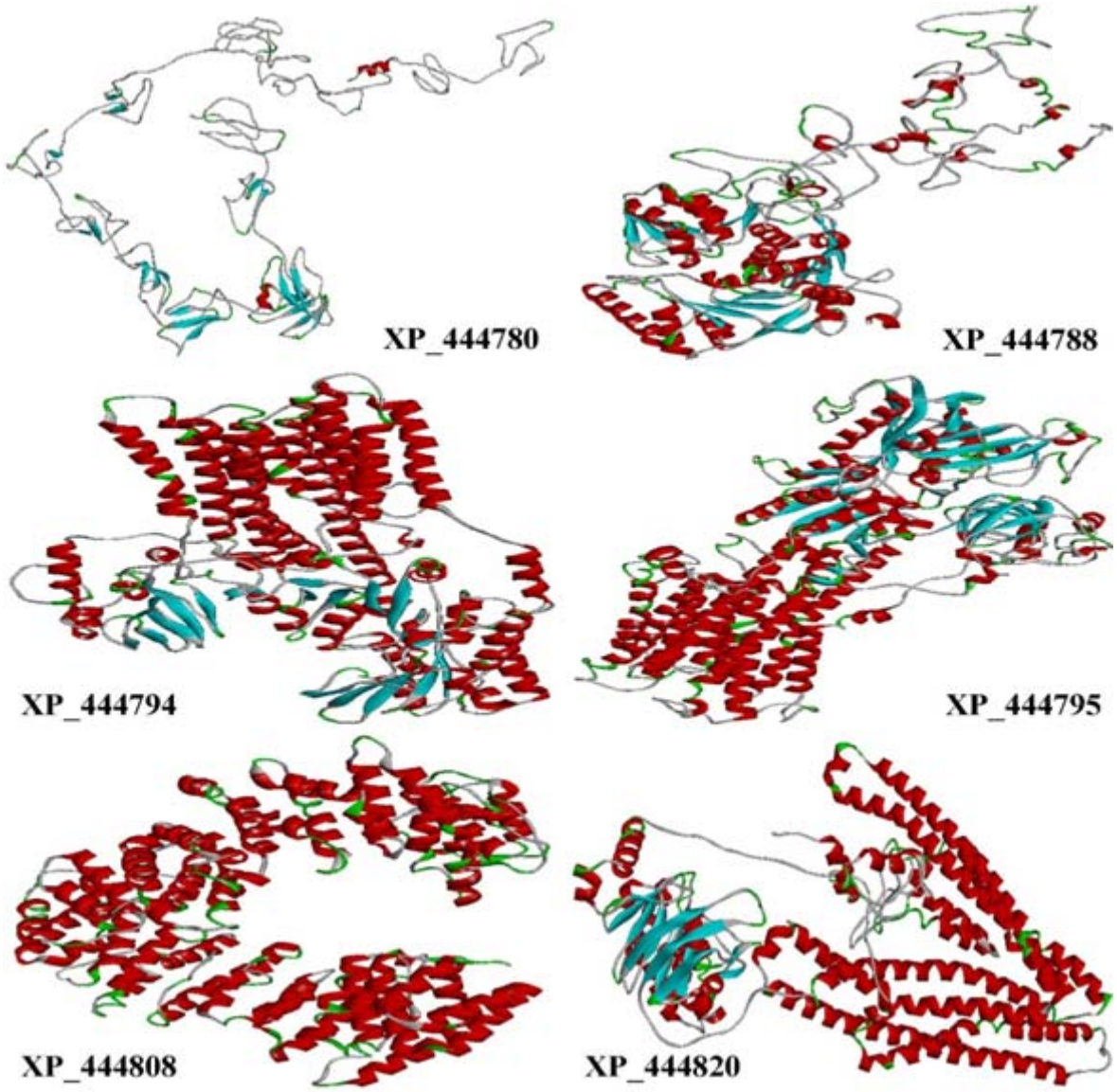

Figure 3. (continued).

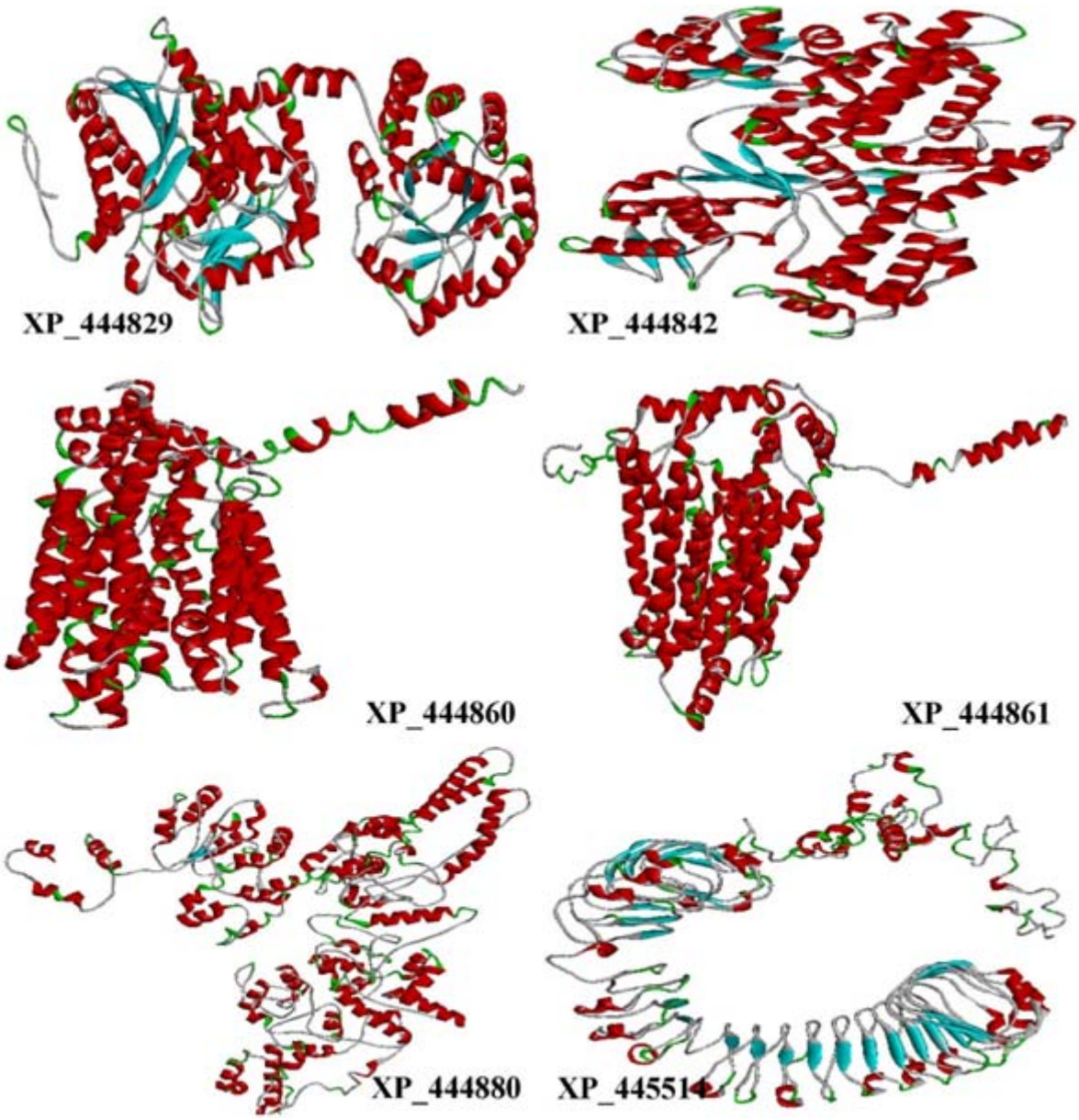

Figure 3. Predicted structures of selected hypothetical proteins

at which they are stable, least soluble and immobile in an electro focusing system i.e. contain no net charge [21]. In present study, protein XP_444880, XP_444808, XP_444842, XP_444820, XP_444780, XP_444845, XP_444861, XP_444851, XP_444843 and XP_444860 showed higher pI, so, during protein purification by isoelectric focusing method, require basic buffer. *In vitro* protein stability is measured by calculating instability index (II) ranges from 1 to 40 and depends on the primary sequence of protein[22]. In the current study, protein XP_444780, XP_444880, XP_444856, XP_444849 and XP_444820 displayed instability index (II) greater than 40, thus unstable and rests were seemed to be stable. Ligand binding residues on protein surface require flexible side chains to undergo conformational changes. Thus positive charged residues (Arg, Lys and His) as well as negatively charged (Glu and Asp) amino acids reside in active sites of proteins [23]. Surface of selected protein XP_444880, XP_444808, XP_444851, XP_444820, XP_444780, XP_444842, XP_444843, XP_444861, XP_444860,

Table 7. Predicted active sites of selected HPs

Sequence Id	Active site volume	Residues at active site
XP_444780	208	ARG 673; ASN 684, 692, 699 and 701; ASP 628; CYS 664; GLU 667 and 703; GLY 650, 651, 652, 653, 688 and 693; HIS 665; ILE 694; LEU 671 and 700; LYS 633, 685, 686 and 697; MET 704; PHE 690; PRO 683 and 689; SER 675, 679, 698 and 705; THR 677; TYR 668 and 687; VAL 648, 649, 654, 680, 682,695 and 702.
XP_444794	153	ALA 94 and 95; ASN 96; GLU 97 and 98; GLY 100 and 101; ILE 102, 104, 105 and 108; LEU 146, 149, 271, 274 and 275; MET 278, 279 and 281; PHE 282 and 285; PRO 311 and 314; SER 315; THR 318, 319 and 320; VAL 321, 322, 323, 324, 326, 327 and 330.
XP_444795	189	ALA 65, 76, 137 and 377; ARG 64 and 80; ASP 79, 140 and 293; GLN 148 and 297; GLU 75 and 144; LEU 72, 145, 300, 368, 370, 376 and 393; LYS 67 and 143; MET 82; PHE 69 and 147; PRO 66 and 369; SER 68, 136 and 301; THR 81, 304, 373 and 380; TYR 141; VAL 73, 372 and 390.
XP_444808	397	ALA 70,75 and 92; ASN 77,,88,,90, 107 and 126; ASP 37, 52 and 152; CYS 86; GLN 60, 132 and 155; GLU 87, 102, 128 and 156; GLY 71, 109 and 176; HIS 105 and 120; ILE 34, 51, 55, 83, 108, 117, 148, 149, 179 and 183; LEU 68, 74, 84, 96, 103, 104, 116, 118 and 138; LYS 38, 73, 76, 94, 97, 140 and 180; PHE 41, 113, 123, 129 and 177; PRO 141, 144 and 145; SER 91, 98, 99, 100, 110, 121 and 137; THR 79, 95, 106 and 124; TRP 101 and 147; TYR 72 and 153; VAL 67, 69, 93, 114 and 133.
XP_444842	192	ALA 353, 354 and 371; ARG 357; ASN 153 and 187; ASP 350; GLN 295 and 374; GLU 148, 356 and 397; GLY 161, 190 and 292; HIS 159, 162, 377 and 398; ILE 351; LEU 146, 189, 381, 385 and 396; MET 186, 366 and 378; PHE 149 and 401; PRO 152; SER 150, 151, 165 and 291; TYR 188, 346, 368 and 382; VAL 296, 369 and 370.
XP_444788	266	ALA 220, 228, 462, 502 and 503; ARG 364 and 477; ASN 223, 234, 362 and 370; ASP 184, 195, 367 and 549; CYS 360; GLU 219, 227, 366, 500 and 501; GLY 185, 190, 294 and 363; HIS 122 and 199; ILE 191, 292, 297 and 298; LEU 187, 196, 224 and 499; LYS 186, 194, 293 and 373; MET 230 and 548; PHE 551; PRO 189, 461 and 550; SER 188, 193, 225, 231, 295 and 460; THR 123, 146, 192, 300 and 361; TYR 504; VAL 183, 251, 374 and 553.
XP_444829	400	ALA 407, 438, 525, 553, 565, 598, 622 and 644; ARG 436, 443, 462 and 674; ASN 440 and 532; ASP 433, 592, 601, 625, 676 and 678; CYS 526; GLN 409 and 437; GLU 404, 552 and 646; GLY 379, 406, 408, 484, 488, 528, 529, 530, 550, 554, 555, 599, 604, 606 and 673; HIS 381, 410, 563, 576, 594 and 649; ILE 487 and 621; LEU 439, 461, 567, 600 and 645; LYS 382 and 677; PHE 469, 575 and 620; PRO 489, 603 and 607; SER 380, 485, 531, 564, 595, 597, 672 and 681; THR 378, 405, 460, 465, 566, 568, 580, 593 and 624; TYR 481 and 602; VAL 412, 483, 527, 551, 578, 582, 596 and 605.

table 7. (continued).

Sequence Id	Active site volume	Residues at active site
XP_445514	519	ALA 394, 402, 429, 446, 494 and 509; ARG 413, 473, 481 and 505; ASN 428, 448, 449, 471 and 545; ASP 407, 458, 492 and 512; CYS 410 and 421; GLN 397, 436 and 454; GLU 405, 406, 425, 456, 501, 502 and 514; GLY 478; HIS 442, 460 and 507; ILE 412, 433, 453, 463, 493 and 524; LEU 378, 395, 396, 409, 419, 439, 440, 441, 479, 485, 498, 543 and 544; LYS 398, 401, 443, 444, 452, 467, 482, 503 and 526; MET 414, 417 and 426; PHE 400, 403, 404, 411, 415, 416 and 432; PRO 377, 425, 483, 491 and 525; SER 418, 422, 431, 445, 447, 450, 457, 466, 468, 480, 511, 515, 516, 522 and 542; THR 424, 434 and 438; TRP 430; TYR 423, 427 and 506; VAL 420, 477, 488, 496, 510 and 517.
XP_444880	294	ALA 737; ARG 510, 715, 733 and 744; ASN 528 and 714; ASP 505 and 515; GLN 509, 534, 710 and 711; GLU 502, 506, 507, 527 and 560; HIS 707; ILE 503 and 520; LEU 500, 523, 530, 538 and 568; LYS 516, 517, 519 and 708; MET 511 and 743; PHE 521, 541 and 740; PRO 501, 542, 563 and 571; SER 504, 514, 518, 531, 535, 539 and 567; THR 524 and 526; TRP 736; TYR 508 and 544; VAL 543 and 564.
XP_444861	203	ALA 167, 359, 423, 427 and 450; ASN 105, 323, 352, 434 and 454; ASP 322 and 364; CYS 108; GLN 191, 314, 317, 318 and 451; GLY 62 and 184; ILE 194, 198 and 461; LEU 360; MET 447; PHE 61, 104, 326, 422 and 431; PRO 171, 428 and 430; SER 188, 355 and 435; THR 65, 195 and 356; TRP 455; TYR 190 and 327; VAL 163, 168, 187, 363 and 432.
XP_444860	629	ALA 76, 170, 430 and 465; ARG 82, 83, 159 and 396; ASN 180, 208, 216, 326, 355, 412, 457 and 473; ASP 67, 79, 325, 343 and 403; CYS 199, 206, 227 and 419; GLN 77, 145, 194, 219, 317, 320, 321 and 454; GLY 65, 69, 73, 101, 141, 200, 204, 332, 340, 352, 411 and 461; HIS 93; ILE 71, 142, 197, 201, 349, 351, 353, 415 and 422; LEU 84 and 224; LYS 97, 212, 337 and 408; MET 414; PHE 64, 74, 80, 107, 202, 328, 329, 336, 345, 356, 389, 425, 434, 466, 470, 477 and 483; PRO 223 and 469; SER 72, 96, 105, 162, 217 and 348; THR 68, 70, 100, 198, 211, 333, 334, 347 and 418; TRP 66, 220, 398, 429 and 458; TYR 94, 138, 155, 205, 209, 327, 330, 331, 421 and 479; VAL 75, 98, 104, 335 and 339.
XP_444859	240	ALA 146, 147 and 148; ARG 149 and 151; ASN 152; ASP 153, 154 and 159; GLN 160, 163 and 186; GLU 187, 188 and 189; GLY 190, 191, 192 and 193; HIS 194 and 197; ILE 222 and 250; LEU 251, 254, 292 and 338; LYS 341, 342, 343 and 344; MET 345; PHE 348, 349, 356, 367, 368 and 369; SER 370, 373 and 374; THR 377, 378 and 380; TRP 381; TYR 395, 396, 397 and 400; VAL 402.
XP_444851	557	ALA 142, 143, 144, 145 and 146; ARG 147, 150 and 151; ASN 152, 153, 154 and 155; ASP 156, 159, 160, 163, 164, 169, 173 and 175; GLN 178, 179 and 180; GLU 181, 182, 192 and 193; GLY 194, 195, 196, 197,

table 7. (continued).

Sequence Id	Active site volume	Residues at active site
		198, 199 and 201; HIS 202, 203, 204, 205, 206 and 207; ILE 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219 and 220; LEU 223, 224, 225, 226, 227, 228, 229 and 230; LYS 231, 232, 233, 234, 235, 236, 237, 238 and 239; MET 241, 243, 244 and 245; PHE 248, 249 and 251; PRO 252, 266 and 269; SER 270, 273, 274, 275, 284, 290 and 291; THR 292, 326, 327, 328, 329, 330, 331, 372 and 373; THR 329, 330, 331, 372 and 373; TRP 394; VAL 395, 396, 397, 398, 399, 475, 477, 478, 479, 506, 507, 508 and 510.
XP_444847	159	ALA 22, 662 and 733; ASN 21, 26 and 731; ASP 706 and 713; GLN 718; GLU 29 and 31; GLU 724; GLY 57, 583, 664, 716 and 723; ILE 708 and 720; LEU 663; LYS 586; PHE 719; PRO 20 and 712; SER 24, 30, 32, 585, 709, 714 and 715; THR 23, 28, 710, 711 and 722; TRP 60; TYR 56 and 584; VAL 707 and 735.
XP_444846	315	ALA 59, 60 and 61; ARG 62, 63 and 64; ASN 65, 66, 67, 68, 69 and 70; ASP 71, 72 and 73; CYS 75; GLN 76 and 80; GLU 81 and 82; GLY 83, 85, 86, 87, 96 and 97; HIS 98 and 99; ILE 102; LEU 103, 105, 106, 107 and 108; LYS 109, 118, 119 and 120; MET 121; PHE 122, 123, 124 and 685; PRO 686, 687, 689, 694, 695, 697 and 698; SER 699, 701, 702, 740, 741, 742, 779, 780 and 781; THR 782, 783 and 784; TYR 785, 883 and 884; TYR 885; VAL 886, 893, 894, 895 and 898.
XP_444845	211	ALA 36; ASP 24; CYS 32, 170 and 177; GLN 10; GLU 12, 14 and 16; GLY 38 and 39; ILE 2, 33 and 155; LEU 9, 17, 25, 163, 167, 173, 181 and 341; LYS 20, 21 and 22; MET 1 and 174; PHE 37, 40, 180, 334 and 338; PRO 5 and 152; SER 3, 29, 151 and 169; THR 166; TYR 171; VAL 13, 28, 35, 331 and 335.
XP_444844	462	ALA 72, 73 and 74; ASN 76, 77, 78, 79, 80 and 81; ASP 83; CYS 84, 85 and 102; GLN 105, 106, 109 and 110; GLY 112, 113, 115, 116, 119, 174, 198, 199 and 201; ILE 202, 205, 206, 207, 208 and 209; LEU 210, 211, 212 and 213; LYS 214, 215 and 216; PHE 217, 220, 231, 234, 235, 328, 329, 333, 334, 336, 337, 338, 340 and 341; PRO 342, 344 and 345; SER 352, 353, 355, 356 and 357; THR 359, 360, 361, 363, 364 and 429; TRP 430, 433 and 437; TYR 438, 439, 442, 443, 465 and 466; VAL 469, 470, 473, 474, 477, 478 and 481.
XP_444843	553	ALA 92, 186, 406, 446 and 490; ARG 175 and 412; ASN 124, 224, 342, 371, 473 and 492; ASP 83, 95, 341 and 489; CYS 222, 243 and 435; GLN 93, 161, 210, 235, 333, 336, 337 and 470; GLY 81, 85, 89, 117, 131, 179, 183, 216, 220, 348, 368, 409, 419, 477 and 496; ILE 87, 122, 158, 213, 217, 350, 365, 367, 438, 480 and 491; LEU 215, 240 and 338; LYS 228 and 353; MET 430; PHE 80, 90, 123, 218, 225, 344, 345, 352, 361, 372, 405, 441, 450, 478, 482, 486, 493 and 499; PRO 239 and 485; SER 88, 121, 178, 360, 364 and 481; THR 84, 86, 214, 227, 349, 434, 484 and 488; TRP 82, 236, 445 and 474; TYR 154, 221, 343, 346, 347, 437 and 495; VAL 119, 120, 187, 351, 369, 408 and 498.

table 7. (continued).

Sequence Id	Active site volume	Residues at active site
XP_444856	304	ARG 245, 248, 269, 270, 271, 277, 278 and 296; ASP 297; CYS 298 and 299; GLN 300, 301 and 304; GLU 315, 317, 318, 319, 321, 323, 327, 328, 329 and 330; GLY 331, 332, 333, 334 and 335; HIS 336 and 337; ILE 338, 340 and 421; LEU 445, 446, 447, 448 and 449; LYS 461; MET 463 and 471; PHE 472 and 473; PRO 474, 475 and 476; SER 478, 479, 485, 502, 506, 515, 516 and 517; THR 518, 519, 520, 521 and 522; TYR 533, 537, 538 and 539; VAL 540 and 541.
XP_444820	269	ALA 7, 8 and 11; ARG 12 and 13; ASN 14, 15, 16, 17 and 18; GLN 20, 49 and 50; GLU 51, 52, 53, 56 and 57; GLY 63 and 67; HIS 70 and 71; LEU 72 and 73; LYS 74, 75, 76, 77, 79, 80, 81 and 82; PRO 83, 84 and 85; SER 86, 87, 89, 90, 92, 93, 98 and 99; THR 101, 102, 103, 105 and 106; TRP 107; TYR 108 and 109; VAL 111, 316, 319, 320 and 323.

XP_444845 and XP_444844 contained more number of positively charged amino acids which might be involved in ligand binding. On the other hand, protein XP_444795, XP_444794, XP_444788, XP_444846, XP_445514, XP_444829, XP_444859, XP_444856 and XP_444847 comprised more amounts of negatively charged amino acids at their surface area which might display binding sites of ligand. Hydrophobicity of a protein depends on hydrophobic amino acids [24]. Protein XP_444843, XP_444861, XP_444845, XP_444860, XP_444844 and XP_444794 was more hydrophobic than others (Table 1). Moreover, a significant positive correlation was found between aliphatic index of selected HPs and their GRAVY ($r = 0.583$, $p<0.01$), i.e. hydrophobic proteins contained long chain aliphatic side chains [24].

Functional categorization of proteins was performed by determining protein domains, signal peptide, subcellular localizations, transmembrane regions, protein network analysis, etc. [21]. According to evolutionary conservation theory, domains are the most conserved parts as well as the functional regions of proteins [25]. They can modulate the function of a protein by changing its arrangement (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/). So, only sequence homology based methods may not predict the actual function of proteins [26]. Furthermore, hypothetical proteins do not have any known homologous protein sequences. So, a number of domain/superfamily prediction tools were used to predict accurate domains of selected 20 hypothetical proteins of *C. glabrata* CBS 138.

In the present study, protein XP_444780 showed Zn+ ion binding domain ZF_FYVE, involved in TGFâ signalling[27]. XP_444780 showed interaction (Fig.2) with GTP-binding protein (gtr1), PRA1-like (Prenylated Rab acceptor 1) protein (SPCC306.02c-1) involved in cell trafficking [28]; vacuolar protein 8 (vac8), important for vacuolar sorting [29]. These functions were generally occurred by NVJ (nuclear–vacuolar junction) proteins, which resided in nuclear–vacuolar junctions of yeast cell and helped in engulfment of nucleus into vacuole during carbon and nitrogen depletion [30]. Very naturally, protein XP_444780 was localized

in nucleus (Table 3), thus it might be concluded that protein XP_444780 was functioning as nuclear–vacuolar junction protein. Protein XP_444788 showed Aconitase domain, involved in catalysis of interconversion between isocitrate and citrate via cis-aconitate intermediate [31] and interacted with (Fig.2) 3-isopropylmalate dehydrogenase (LEU2); branched-chain-amino-acid aminotransferase (BAT1); 2-isopropylmalate synthase (LEU4); mitochondrial oxaloacetate transport protein (OAC1); dihydroxy-acid dehydratase (ILV3), which were involved in cellular metabolic pathways [32]. These functions used to occur in cytoplasm[33], which was the subcellular localizer of protein XP_444788. Protein XP_444808 showed RVT_1 domain, a reverse transcriptase gene usually present in mobile element like retrotransposon [34]. The protein interacted with (Fig. 2) telomere elongation protein (EST1; EST3); high affinity DNA-binding factor subunit 2 (YKU70, YKU80); DNA repair and recombination protein (RAD52). Thus protein XP_444808 might function as a DNA binding protein in nucleus. ABC_membrane domain acts as ABC transporter [35] existed in protein XP_444820. It was connected with (Fig.2) para-hydroxybenzoate-polyprenyl transferase (COQ2), an integral membrane protein involved in ubiquinone biosynthesis [36], fatty acid elongation protein 3 (SUR4); mitochondrial inner membrane translocase subunit TIM16 (PAM16); mitochondrial carrier (MTM1). Above functions generally happened in mitochondria and endoplasmic reticulum, which was the subcellular localization of protein XP_444820 (Table 3). Moreover the protein revealed transmembrane regions (Table 4), thus might be involved in mitochondrial transportation. Protein XP_444829 comprised Trp_syntA domain, responsible for tryptophan biosynthesis [37]. XP_444829 interacted with N-(5'-phosphoribosyl) anthranilate isomerase (TRP1); anthranilate synthase components (TRP2; TRP3); anthranilate phosphoribosyltransferase (TRP4); threonine dehydratase (ILV1), all involved in amino acid biosynthesis and catabolism [38-41], used to occur in cytoplasm [33], which was subcellular localizer of protein XP_444829 (Table 3). Thus it might be concluded that protein XP_444829 was associated with amino acid biosynthesis and catabolism. Protein XP_444846 showed GATA_ZN_FINGER domain, a DNA binding domain also performed in vesicular trafficking [42]. It was associated with (Fig.2) transport proteins (SEC23, SEC24, SEC31, SEC13); GRASP65 homolog protein 1 (GRH1), a coat protein complex II (COPII) that promote formation of transport vesicles from the endoplasmic reticulum [43] and localized above nucleus which was also a localizer of XP_444846 (Table 3). Thus protein XP_444846 might be involved in vesicular transportation.

Protein XP_444847 revealed collagen domain, involved in formation of connective tissue [44]. XP_444847 showed interaction with nuclear receptor subfamily 6- group A- member 1 (NR6A1), involved in integrin-mediated cell-matrix interaction[45]; integrin (ITGA10); laminin (LAMB3); fibronectin, component of extracellular matrix [46], which was the subcellular localizer of the studied protein. Cleavage site was present in this protein between 17th (Ala) and 18th (Phe) amino acids (Fig. 1). So it could be concluded that XP_444847 was an extracellular protein and might be involved in host-pathogen associations [47].

Protein XP_445514 contained LRR_1 domain, a structural basis of various purposes such as formation of protein-protein interactions [48], tyrosine kinase receptors, cell-adhesion molecules, virulence factors and extracellular matrix-binding glycoproteins [49]. It interacted with (Fig. 2) protein HYM1, helped in cell cycle regulation [50]; serine/threonine-protein kinase (KIC1), required for cell integrity, cellular polarity and morphogenesis [51]; nicotinamide riboside kinase 1(NRK1), coenzyme of oxidoreductase and performed as a source of ADP-ribosyl groups used in various reactions [52]; Serine/threonine-protein kinase (CBK1), seemed to play role in regulation of cell morphogenesis and proliferation [53]; autophagy-related protein 17 (ATG17), responsible for pexophagy and nucleo-phagy [54]. All the above functions typically occurred in nucleus, which was the subcellular localizer of protein XP_445514 (Table 3). Thus it might be involved in nuclear functions. Translation initiation factor 2 (IF2) and 30S initiation complex (IC) forming domain IF2 [55] was found in protein

XP_444851. Moreover, phenylalanyl-tRNA synthetase (MSF1); RNA exonuclease (NGL2); 37S ribosomal protein (RSM28), involved in mitochondrial protein translation [56]; ATP-dependent RNA helicase (PRP22) a pre-mRNA-splicing factor; mitochondrial precursor required for initiation of translation of the COX1 coding region (PET309) were interacted with the protein (Fig. 2). The above functions were involved in gene translation of mitochondria, which was the subcellular localizer of protein XP_444851 (Table 3). So, the protein might be involved in mitochondrial gene translation. Protein XP_444880 showed SANT domain, involved in chromatin-remodelling and transcription regulation [57]. It was interacted with (Fig. 2) chromatin modification-related protein (EAF5, YNG2, EAF7); SWR1 complex mediated ATP- dependent exchange of histone H2A (YAF9, SWC4), which were involved in centromere functions, DNA damage control, cell cycle control [58] in nucleus the subcellular localizer of protein XP_444880 (Table 3). Protein XP_444880 might be involved in DNA repair and cell cycle control of *C. glabrata* CBS 138.

Protein XP_444794 and XP_444795 both contained E1-E2_ATPase domain, a transmembrane ATPase that transport membrane-bound enzyme complexes or ions [34]. Protein XP_444794 was associated with (Fig.2) ubiquitin (UBI4); serine/threonine-protein kinase PTK2/STK2 (PTK2); general amino-acid permease (GAP1); yeast elongation factor 3 (YEF3), requiredATPase for functioning [59-61]. XP_444794 and XP_444795 showed transmembrane regions (Table 4) and plasma membrane as subcellular localizer (Table 3). They might be involved in ATPase execution. On the other hand proteinXP_444795 shown interaction with vacuolar Ca^{2+}/H^{+} exchanger (VCX1); golgi Ca^{2+}-ATPase (PMR1); calcium-channel protein (CCH1); vacuolar v-SNARE (NYV1), which were involved in voltage-gated Ca^{2+} channels[62]. Subcellular localizer of the HP was plasma membrane (Table 3) and transmembrane regions were present (Table 4). Thus protein XP_444795 might be involved in Ca^{2+} ion exchange. AA_TRNA_LIGASE domain, catalysed aminoacyl-tRNA ligase [63], was observed in three proteins XP_444842, XP_444856 and XP_444859. Protein XP_444842 shown interaction with (Fig.2) methionine-tRNA ligase (MES1); GMP synthase (GUA1); glutamyl-tRNA synthetase (SES1); isoleucyl-tRNA synthetase (ILS1). Above proteins were involved in RNA synthesis andsubcellular localizer of protein XP_444842 was mitochondria (Table 3). So, it might be concluded that protein XP_444842 was involved in mitochondrial RNA synthesis. On another hand protein XP_444856 and XP_444859 interacted with cytoplasmic tRNA synthetase (Fig. 2) and PSORT II shown cytoplasm as subcellular localizer (Table 3) of above HPs. Thus protein XP_444856 and XP_444859 might be involved in RNA synthesis. Protein XP_444845, XP_444844, XP_444843, XP_444860 and XP_444861 showed Sugar_tr domain, played role in uptake of sugar [64]. They interacted with glucose transporters (HXT 1-7) (Fig. 2) and subcellular localizer was plasma membrane (Table 3). They also revealed transmembrane regions (Table 4). Thus the above HPs might be involved in glucose transportation.

Among two types of transmembrane proteins, helical proteins are more abundant than β-barrel [65]. In the present study, protein XP_444843, XP_444844, XP_444845, XP_444860, XP_444861, XP_444820, XP_444794 and XP_444795 were membrane proteins; contained large quantity of helical structure (Fig. 3) and their hydrophobicity as well as aliphatic index was higher (Table 1). On the other side, outer membrane proteins generally contain more amount of β-barrel [65] and formerly established outer membrane protein XP_444847 was containing large amount of â-barrel (Fig. 3). Active sites were also identified and shown in Table 7. Determination of tertiary structure of the hypothetical proteins (Fig. 3), their templates (Table 6) and active sites might be helpful to study the conformations and to identify molecular docking sites that would help in in silico drug designing.

Conclusion

In silico approach in revelation of protein structure as well as function is less time consuming and cost effective than experimental investigation. Functional annotation may distinguish required

proteins from others. Screening of significant protein from a number of hypothetical proteins by experimental analysis is tedious and very expensive job. Since *C. glabrata* CBS 138 is a new emerging pathogen, it is very important to become acquainted about each protein. In the present study, randomly selected HPs have been categorized in different important intracellular as well as extracellular functions. Protein structures and their active site prediction would help in drug designing and docking studies.

List of abbreviations

CDD, Conserved Domain Database; RPS-BLAST, Reverse Position-Speciûc BLAST; PSI-BLAST, Position Specific Iterated Blast; HMM, Hidden Markov Models; TMHMM, Trans-membrane Hidden Markov Models; STRING, Search Tool for the Retrieval of Interacting Genes/ Proteins; I-tasser, Iterative Threading ASSEmbly Refinement; GROMOS, GROningen molecular dynamics simulation; GRAVY, Grand Average hydropathy.

Acknowledgement

The Department of Biotechnology (DBT), Government of India, New Delhi is acknowledged gratefully for creation of Bioinformatics Infrastructure Facility Centre at Vidyasagar University, Midnapore, West Bengal, India.

References

1. **Bethea, E.K., Carver, B.J., Montedonico, A.E., Reynolds, T.B. (2009).** The inositol regulon controls viability in *Candida glabrata*. Microbiology156: 452-462.
2. **Paul, L.F., Jose, A.V., Jack, D.S. (1999).** *Candida glabrata*: review of epidemiology, pathogenesis, and clinical disease with comparison to *C. albicans*. Clinical Microbiology Reviews 12: 80-96.
3. **Wingard, J.R. (1995).** Importance of *Candida* species other than *C. albicans* as pathogens in oncology patients. Clinical Infectious Diseases 20: 115-125.
4. **Mandell G.L., Bennett J.E., Dolin R. (2010).** Mandell, Douglas, and Bennett's principles and practice of infectious diseases. 7th ed. Philadelphia.
5. **Corno, F., Caldart, M., Toppino, M., Tapparo, A., Capozzi, M.P., Goitre, M., Cervetti, O., Forte, M., Forcheri, V. (1989).** Ano-rectal candidiasis. Minerva Chirurgica 44: 2251-2253.
6. **Roetzer, A., Gabaldon, T., Schuller, C. (2011).** From *Saccharomyces cerevisiae* to *Candida glabrata* in a few easy steps: important adaptations for an opportunistic pathogen. FEMS Microbiology Letters 314: 1-9.
7. **Papon, N., Courdavault, V., Clastre, M., Bennett, R.J. (2013).** Emerging and emerged pathogenic *Candida* species: beyond the *Candida albicans* paradigm. PLOS Pathogens 9: e1003550.
8. **Seider, K., Heyken, A., Luttich, A., Miramon, P., Hube, B. (2010).** Interaction of pathogenic yeasts with phagocytes: survival, persistence and escape. Current Opinion in Microbiology 13: 392-400.
9. **Zarembinski, T.I., Hung, L.W., Mueller-Dieckmann, H.J., Kim, K.K., Yokota, H., Kim, R., Kim, S.H. (1998).** Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. Proceedings of the National Academy of Sciences of the United States of America 95: 15189-15193.
10. **Nimrod, G., Schushan, M., Steinberg, D.M., Ben-Tal, N. (2008).** Detection of functionally important regions in ''hypothetical proteins'' of known structure. Structure 16: 1755-1763.
11. **Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., Bryant, S.H. (2011).** CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Research 39: D225-D229.
12. **Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A.,**

Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M. (2014). Pfam: the protein families database. Nucleic Acids Research 42: D222-D230.

13. **Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N. (2006).** Scan Prosite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Research 34:W362-W365.
14. **Haft, D.H., Selengut, J.D., White, O. (2003).** The TIGRFAMs database of protein families. Nucleic Acids Research 31: 371-373.
15. **Petersen, T.N., Brunak, S., Heijne, G., Nielsen, H. (2011).** SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8: 785-786.
16. **Nakai, K., Kanehisa, M. (1992).** A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14: 897-911.
17. **Krogh, A., Larsson, B., Heijne, G., Sonnhammer, E.L. (2001).** Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of Molecular Biology 305: 567-580.
18. **Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering. C., Jensen. L.J. (2013).** STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research 41:D808-D815.
19. **Roy, A., Kucukural, A., Zhang, Y. (2010).** I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.
20. **Guex, N., Peitsch, M.C. (1997).** SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modelling. Electrophoresis 18: 2714-2723.
21. **Mohan, R., Venugopal, S. (2012).** Computational structural and functional analysis of hypothetical proteins of *Staphylococcus aureus*. Bioinformation 8: 722-728.
22. **Guruprasad, K., Reddy, B.V., Pandit, M.W. (1990).** Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering 4: 155-161.
23. **Najmanovich, R., Kuttner, J., Sobolev, V., Edelman, M. (2000).** Side-chain flexibility in proteins upon ligand binding. Proteins 39: 261-268.
24. **Kyte, J., Doolittle, R.F. (1982).** A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157: 105-132.
25. **Pettit, F.K., Bare, E., Tsai, A., Bowie, J.U. (2007).** HotPatch: a statistical approach to finding biologically relevant features on protein surfaces. Journal of Molecular Biology, 369: 863-879.
26. **Schnoes, A.M., Brown, S.D., Dodevski, I., Babbitt, P.C. (2009).** Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLOS Computational Biology 5: e1000605.
27. **Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., Wrana, J.L. (1998).** SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. Cell 95: 779-791.
28. **Kamei, C.L.A., Boruc, J., Vandepoele, K., Van den Daele, H. (2008).** The PRA1 gene family in Arabidopsis. Plant Physiology 147: 1735-1749.
29. **Chen, Y.J., Stevens, T.H. (1996).** The VPS8 gene is required for localization and trafficking of the CPY sorting receptor in *Saccharomyces cerevisiae*. European Journal of Cell Biology 70: 289-297.
30. **Honscher, C., Ungermann, C. (2014).** A close-up view of membrane contact sites between the endoplasmic reticulum and the endolysosomal system: From yeast to man. Critical Reviews in Biochemistry and Molecular Biology 49: 262-268.
31. **Juang, H.H. (2004).** Modulation of iron on mitochondrial aconitase expression in human prostatic

carcinoma cells. Molecular and Cellular Biochemistry 265: 185-194.
32. **Smith, E., Morowitz, H. (2004).**Universality in intermediary metabolism. Proceedings of the National Academy of Sciences of the United States of America 101: 13168-13173.
33. **Goodsell, D.S. (1991).** Inside a living cell. Trends in Biochemical Sciences 16:203-206.
34. **Auer, M., Scarborough, G.A., Kühlbrandt, W. (1998).** Three-dimensional map of the plasma membrane H^+-ATPase in the open conformation. Nature 392: 840-843.
35. **Hung, L.W., Wang, I.X., Nikaido, K., Liu, P.Q., Ames, G.F., Kim, S.H. (1998).** Crystal structure of the ATP-binding subunit of an ABC transporter. Nature 396: 703-707.
36. **Uchida, N., Suzuki, K., Saiki, R., Kainou, T., Tanaka, K., Matsuda, H., Kawamukai, M. (2000).** Phenotypes of fission yeast defective in ubiquinone production due to disruption of the gene for p-hydroxybenzoate polyprenyl diphosphate transferase. Journal of Bacteriology 182: 6933-6939.
37. **Crawford, I.P. (1989).** Evolution of a biosynthetic pathway: the tryptophan paradigm. Annual Review of Microbiology 43: 567-600.
38. **Baker, T.I., Crawford, I.P. (1966).** Anthranilate synthetase: Partial purification and some kinetic studies on the enzyme from *Escherichia coli*. The Journal of Biological Chemistry 241: 5577-5584.
39. **He, U., Brown, B. (1957).** Threonine deamination in *Escherichia coli*. II. Evidence for two L-threonine deaminases. Journal of Bacteriology 73: 105-112.
40. **Creighton, T.E., Yanofsky, C. (1970).** Chorismate to tryptophan (*Escherichia coli*) - Anthranilate synthetase, PR transferase, PRA isomerase, InGP synthetase, tryptophan synthetase. Methods in Enzymology 17A: 365-380.
41. **Mayans, O., Ivens, A., Nissen, L.J., Kirschner, K., Wilmanns, M. (2002).** Structural analysis of two enzymes catalysing reverse metabolic reactions implies common ancestry. The EMBO Journal 21: 3245-3254.
42. **Matthews, J., Sunde, M. (2002).** Zinc fingers - folds for many occasions. IUBMB Life 54: 351-355.
43. **Jensen, D., Schekman, R. (2011).** COPII-mediated vesicle formation at a glance. Journal of Cell Science 124: 1-4.
44. **Sundaramoorthy, M., Meiyappan, M., Todd, P., Hudson, B.G. (2002).** Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. The Journal of Biological Chemistry277: 31142-31153.
45. **Barreto, G., Reintsch, W., Kaufmann, C., Dreyer, C. (2003).** The function of Xenopus germ cell nuclear factor (xGCNF) in morphogenetic movements during neurulation. Developmental Biology 257: 329-342.
46. **Xiong, Y., Eickbush, T.H. (1990).** Origin and evolution of retro elements based upon their reverse transcriptase sequences. The EMBO Journal 9: 3353-3362.
47. **Sandini, S., La Valle, R., De Bernardis, F., Macri, C., Cassone, A. (2007).** The 65 kDa manno-protein gene of *Candida albicans* encodes a putative b-glucanase adhesin required for hyphal morphogenesis and experimental pathogenicity. Cellular Microbiology 9: 1223-1238.
48. **Kobe, B., Kajava, A.V. (2001).** The leucine-rich repeat as a protein recognition motif. Current Opinion in Structural Biology 11: 725-732.
49. **Rothberg, J.M., Jacobs, J.R., Goodman, C.S., Artavanis-Tsakonas, S. (1990).** Slit: an extra-cellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. Genes & Development 4: 2169-2187.
50. **Hsu, J., Weiss, E.L. (2013).** Cell cycle regulated interaction of a yeast hippo kinase and its activator MO25/Hym1. PLoS ONE 8: e78334.
51. **Sullivan, D.S., Biggins, S., Rose, M.D. (1998).** The yeast centrin, cdc31p, and the interacting

protein kinase, Kic1p, are required for cell integrity. The Journal of Cell Biology 143:751-765.

52. **Bieganowski, P., Brenner, C. (2004).** Discoveries of nicotinamide riboside as a nutrient and conserved NRK genes establish a Preiss-Handler independent route to NAD^+ in fungi and humans. Cell 117:495-502.
53. **Du, L.L., Novick, P. (2002).** Pag1p, a novel protein associated with protein kinase Cbk1p, is required for cell morphogenesis and proliferation in *Saccharomyces cerevisiae*. Molecular Biology of the Cell 13: 503-514.
54. **Kamada, Y., Funakoshi, T., Shintani, T., Nagano, K., Ohsumi, M., Ohsumi, Y. (2000).** Tor-mediated induction of autophagy via an Apg1 protein kinase complex. The Journal of Cell Biology 150:1507-1513.
55. **Simonetti, A., Marzi, S., Billas, I.M., Tsai, A., Fabbretti, A., Myasnikov, A.G., Robline,P., Vaianag,A.C., Hazemanna,I., Eilerh,D., Steitzi,T.A., Puglisic, J.D., Gualerzik, C.O., Klaholz, B.P. (2013).**Involvement of protein IF2 N domain in ribosomal subunit joining revealed from architecture and function of the full-length initiation factor. Proceedings of the National Academy of Sciences of the United States of America 110: 15656-15661.
56. **Williams, E.H., Bsat, N., Bonnefoy, N., Butler, C.A., Fox, T.D. (2005).** Alteration of a novel dispensable mitochondrial ribosomal small-subunit protein, Rsm28p, allows translation of defective COX2 mRNAs. Eukaryotic Cell 4:337-345.
57. **Horton, J.R., Elgar, S.J., Khan, S.I., Zhang, X., Wade, P.A., Cheng, X. (2007).** Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. Proteins: Structure, Function, and Bioinformatics 67: 1198-1202.
58. **Krogan, N.J., Baetz, K., Keogh, M.C., Datta, N., Sawa, C., Kwok, T.C., Thompson, N.J., Davey, M.G., Pootoolal, J., Hughes, T.R., Emili, A., Buratowski, S., Hieter, P., Greenblatt, J.F. (2004).** Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4. Proceedings of the National Academy of Sciences of the United States of America 101: 13513-13518.
59. **Peters, J.M., Franke, W.W., Kleinschmidt, J.A. (1994).** Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. The Journal of Biological Chemistry 269: 7709-7718.
60. **Goossens, A., Fuente, N., Forment, J., Serrano, R., Portillo, F. (2000).** Regulation of Yeast H^+-ATPase by protein kinases belonging to a family dedicated to activation of plasma membrane transporters. Molecular and Cellular Biology 20: 7654-7661.
61. **Dasmahapatra, B., Chakraburtty, K. (1981).** Protein synthesis in yeast. I. Purification and properties of elongation factor 3 from *Saccharomyces cerevisiae*. The Journal of Biological Chemistry 256: 9999-10004.
62. **Ramakrishnan, N., Drescher, M., Drescher, D. (2012).** The SNARE complex in neuronal and sensory cells. Molecular and Cellular Neuroscience 50: 58-69.
63. **Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D. (1990).** Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347: 203-206.
64. **Pao, S.S., Paulsen, I.T., Saier, M.H. Jr. (1998).** Major facilitator superfamily. Microbiology and Molecular Biology Reviews 62: 1-34.
65. **Schulz, G.E. (2002).**The structure of bacterial outer membrane proteins. Biochimica et Biophysica Acta 1565: 308-317.