Assignment for subcellular protein localization

1. Name of chosen protein: HWP1
   Organism: *Candida albicans*
   Locus: CAU64206
   NCBI Nucleotide accession #: U64206
   NCBI protein accession #: AAC96368

2. Predicted function: HWP1 possibly functions as an adhesin in *Candida albicans.* It has also been demonstrated to play a role in the morphological change (from yeast form to hyphae form) of *Candida albicans.*

3. The adhesin function of HWP1 implies that it localizes to the cell wall. I did not find any data directly proving this though. In addition, I think it would be interesting to check the change of HWP1 localization during the morphological change of *Candida albicans.*

4. Results from different protein localization:

I. TargetP 1.1 Server - prediction results (Technical University of Denmark)

TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP).

targetp v1.1 prediction results
Number of query sequences: 1
Cleavage site predictions not included.
Using NON-PLANT networks.

Name	Len	mTP	SP	other	Loc	RC
Sequence	634	0.031	0.859	0.098	S	2
cutoff		0.000	0.000	0.000		

DESCRIPTION

Name	Sequence name truncated to 20 characters
Len	Sequence length
cTP, mTP, SP, other	Final NN scores on which the final prediction is based (Loc, see below). Note that the scores are not really probabilities, and they do not necessarily add to one. However, the location with the highest score is the most likely according to TargetP, and the relationship between the scores (the reliability class, see below) may be an indication of how certain the prediction is.

Loc — Prediction of localization, based on the scores above; the possible values are:

- C Chloroplast, i.e. the sequence contains cTP, a chloroplast transit peptide;
- M Mitochondrion, i.e. the sequence contains mTP, a mitochondrial targeting peptide;
- S Secretory pathway, i.e. the sequence contains SP, a signal peptide;
- _ Any other location;
- * "don't know"; indicates that cutoff restrictions were set (see instructions) and the winning network output score was below the requested cutoff for that category.

RC — Reliability class, from 1 to 5, where 1 indicates the strongest prediction. RC is a measure of the size of the difference ('diff') between the highest (winning) and the second highest output scores. There are 5 reliability classes, defined as follows:

1 : diff > 0.800
2 : 0.800 > diff > 0.600
3 : 0.600 > diff > 0.400
4 : 0.400 > diff > 0.200
5 : 0.200 > diff

Thus, the lower the value of RC the safer the prediction.

In a short word, TargetP 1.1 predicts HWP1 will go through the secretory pathway, and it contains a secretory pathway signal peptide (SP).

II. Results of PSORTII

PSG: a new signal peptide prediction method
N-region: length 2; pos.chg 1; neg.chg 0
H-region: length 23; peak value 8.53
PSG score: 4.13
GvH: von Heijne's method for signal seq. recognition
GvH score (threshold: -2.1): -4.50
possible cleavage site: between 27 and 28
>>> **Seems to have no N-terminal signal peptide**
ALOM: Klein et al's method for TM region allocation
Init position for calculation: 1
Tentative number of TMS(s) for the threshold 0.5: 2
Number of TMS(s) for threshold 0.5: 1
INTEGRAL Likelihood = -3.93 Transmembrane 618 - 634
PERIPHERAL Likelihood = 5.46 (at 398)
ALOM score: -3.93 (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
Center position for calculation: 625
Charge difference: 0.0 C(0.0) - N(0.0)
N >= C: **N-terminal side will be inside**
>>> **Single TMS is located near the C-terminus**
>>> **membrane topology: type Nt (cytoplasmic tail 1 to 617)**
MITDISC: discrimination of mitochondrial targeting seq
R content: 1 Hyd Moment(75): 7.98
Hyd Moment(95): 7.26 G content: 1
D/E content: 1 S/T content: 4
Score: -2.96
Gavel: prediction of cleavage sites for mitochondrial preseq
R-2 motif at 12 MRL|ST
NUCDISC: discrimination of nuclear localization signals
pat4: none
pat7: none
bipartite: none
content of basic residues: 2.1%
NLS Score: -0.47
KDEL: ER retention motif in the C-terminus: none
ER Membrane Retention Signals:
XXRR-like motif in the N-terminus: RLST
none
SKL: peroxisomal targeting signal in the C-terminus: none
SKL2: 2nd peroxisomal targeting signal: none
VAC: possible vacuolar targeting motif: none
RNA-binding motif: none
Actinin-type actin-binding motif:
type 1: none
type 2: none
NMYR: N-myristoylation pattern : none
Prenylation motif: none
memYQRL: transport motif from cell surface to Golgi: none
Tyrosines in the tail: too long tail
Dileucine motif in the tail: none
checking 63 PROSITE DNA binding motifs: none
checking 71 PROSITE ribosomal protein motifs: none
checking 33 PROSITE prokaryotic DNA binding motifs: none
NNCN: Reinhardt's method for Cytplasmic/Nuclear discrimination
Prediction: nuclear
Reliability: 94.1
COIL: Lupas's algorithm to detect coiled-coil regions
total: 0 residues

Results of the k-NN Prediction

k = 9/23

21.7 %: nuclear

13.0 %: vesicles of secretory system

13.0 %: cytoplasmic

13.0 %: Golgi

13.0 %: mitochondrial

8.7 %: endoplasmic reticulum

8.7 %: plasma membrane

4.3 %: extracellular, including cell wall

4.3 %: cytoskeletal

>> **prediction for QUERY is nuc (k=23)**

III. Results of WOLFPSORT

queryProtein (634 aa)

PSG: a new signal peptide prediction method

N-region: length 2; pos.chg 1; neg.chg 0

H-region: length 23; peak value 3.51

PSG score: -0.89

GvH: von Heijne's method for signal seq. recognition

GvH score (threshold: -2.1): -4.50

possible cleavage site: between 27 and 28

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation

Init position for calculation: 1

Tentative number of TMS(s) for the threshold 0.5: 2

Number of TMS(s) for threshold 0.5: 1

INTEGRAL Likelihood = -3.93 Transmembrane 618 - 634

PERIPHERAL Likelihood = 5.46 (at 398)

ALOM score: -3.93 (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)

Center position for calculation: 625

Charge difference: 0.0 C(0.0) - N(0.0)

N >= C: N-terminal side will be inside

>>> Single TMS is located near the C-terminus

>>> membrane topology: type Nt (cytoplasmic tail 1 to 617)

MITDISC: discrimination of mitochondrial targeting seq

R content: 1 Hyd Moment(75): 7.98

Hyd Moment(95): 7.26 G content: 1

D/E content: 1 S/T content: 4

Score: -2.96

Gavel: prediction of cleavage sites for mitochondrial preseq

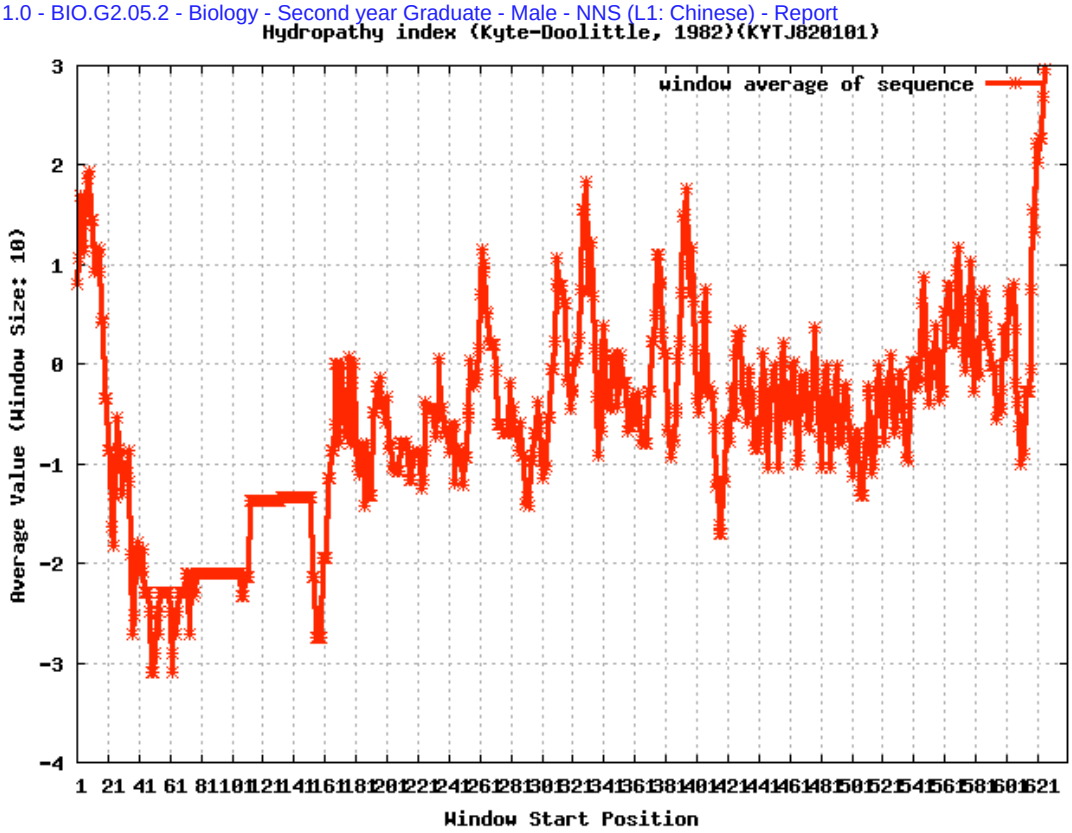
R-2 motif at 12 MRL|ST

NUCDISC: discrimination of nuclear localization signals

pat4: none
pat7: none
bipartite: none
content of basic residues: 2.1%
NLS Score: -0.47
KDEL: ER retention motif in the C-terminus: none
ER Membrane Retention Signals:
XXRR-like motif in the N-terminus: RLST none
SKL: peroxisomal targeting signal in the C-terminus: none
PTS2: 2nd peroxisomal targeting signal: none
VAC: possible vacuolar targeting motif: none
RNA-binding motif: none
Actinin-type actin-binding motif:
type 1: none
type 2: none
NMYR: N-myristoylation pattern : none
Farnesylation/Geranylgeranylation motif: none
memYQRL: transport motif from cell surface to Golgi: none
Tyrosines in the tail: too long tail
Dileucine motif in the tail: none
checking 63 PROSITE DNA binding motifs: none
checking 71 PROSITE ribosomal protein motifs: none
checking 33 PROSITE prokaryotic dna binding motifs: none
Final Results:
48.0 %: extracellular, including cell wall 12.0 %: cytoplasmic 12.0 %: nuclear 12.0 %: endoplasmic reticulum 8.0 %: vesicles of secretory system 4.0 %: plasma membrane 4.0 %: mitochondrial >> **prediction for queryProtein is exc**

WOLFPSORT is an updated version of PSORTII. Comparing the results from two programs, we can see that they gave same/similar results in most aspects. Both programs said that HWP1 doesn't seem to have an N-terminal signal peptide. HWP1 looks more like a peripheral protein rather than an integral one, with single trans-membrane sequence near its C-terminus and its N-terminus facing inside the cell. Nonetheless, those two programs gave controversial results of HWP1 localization. PSORTII suggests that HWP1 localizes to nucleus, while WOLFPSORT predicts HWP1 localizes to extracellular space or cell wall. Given adhesion function of HWP1, the result of WOLFPSORT makes more sense. And it is consistent with the result of TargetP (Because WOLFPSORT has been updated?) However, we can not rule out the possibility that HWP1 may localize to nucleus.

5. HWP1 were overall predicted to be a peripheral membrane protein by all the programs. I further used iPSORT to do an AAindex Analysis, and got



Based on this Hydropathy Index of HWP1, I would predict a trans-membrane sequence at its N-terminus. (If I have understood the hydropathy index correctly) I don't know why both PSORTII and WOLFPSORT predict a trans-membrane sequence at the C-terminus of HWP1.

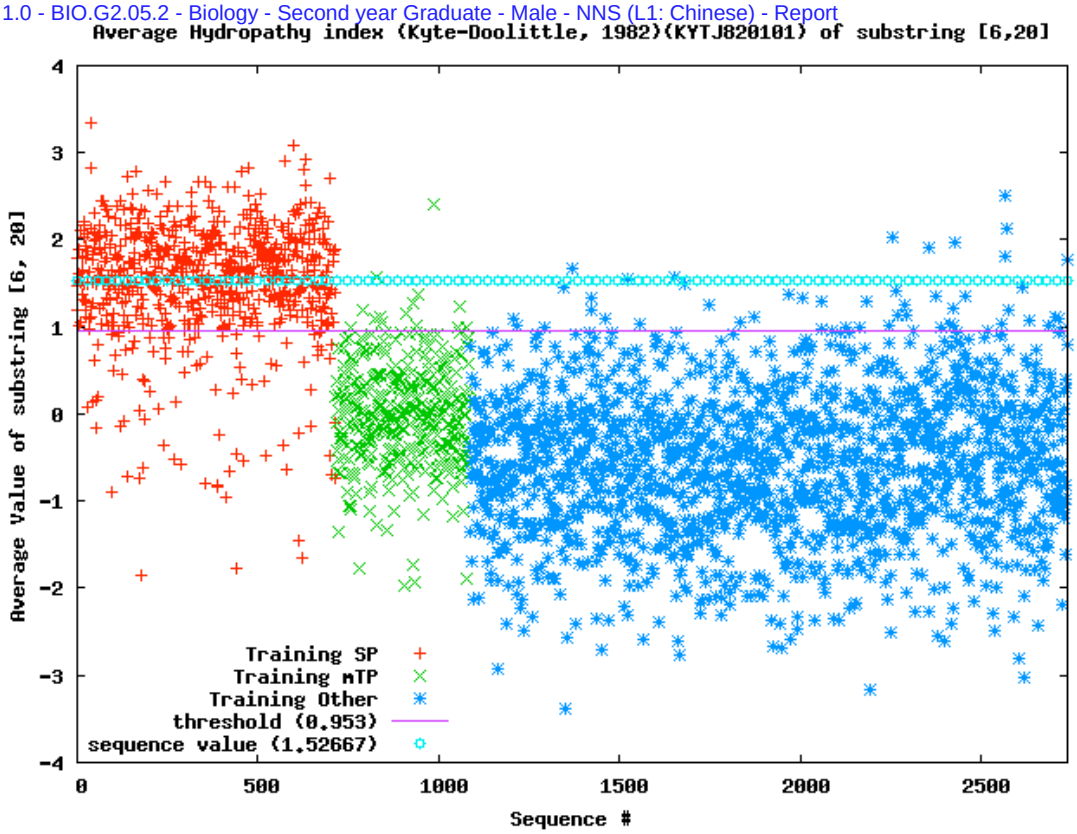
6. Although either PSORTII or WOLFPSORT did not find any N-terminal signal peptide, iPSORT does find a signal peptide in the N-terminus of HWP1. This is consistent with and further consolidates the result of TargetP, which says HWP1 has a signal sequence and is likely a secretory protein.

iPSORT Prediction

Predicted as: having a signal peptide

Values used for reasoning

Node	Answer	View	Substring	Value(s)	Plot
1. Signal peptide?	Yes	Average Hydropathy (KYTJ820101)	[6,20]	1.52667 (>= 0.953? Yes)	show



7. Base on the results of prediction programs, as well as previous data in literature. I would propose that HWP1 has a low expression level in the yeast-form cells of *Candida.* The expression of HWP1 will be induced during *Candida* cells change from yeast-form to hyphae-form, and HWP1 will localize to the cell wall.

8. Considering the expression level may affect a protein's subcellular localization, I would take the advantage of homologous recombination in *Candida*, and replace the endogenous gene of HWP1 with a fusion gene in which HWP1 being tagged with a fluorophore. In this case, the fusion gene will be expressed under endogenous promoter of HWP1, and hopefully has a similar expression level as HWP1. Some better-known adhesion (e.g. ALS3) will be used as cell wall marker. I expected to see dim or no fluorescence from fusion protein in yeast-form cells. After shifting conditions and cells becoming hyphae, I expect to see strong signal and the fluorescence localizes to the cells' surface. If I also tagged ALS3 with another fluorophore, I expect to see co-localization of HWP1 and ALS3. A possible problem is the fluorophore may affect the localization of HWP1. An alternative way to get around is to use HWP1 antibody (which we already have) to do immuno-staining. The down-side of this approach is non-specific binding and staining.

Both approaches may require certain amount of HWP1 to exist inside cells. If HWP1 keeps being expressed at a low level, we can make *Candida* cultures, do cell fractionation. Use HWP1 antibody to detect the amount of HWP1 in each fraction.

Appendix

Candida albicans hyphal wall protein 1 (HWP1) gene, complete cds

FEATURES Location/Qualifiers
source 1..2682
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="SC5314"
/db_xref="taxon:5476"
gene 445..2682
/gene="HWP1"
mRNA 445..2682
/gene="HWP1"
CDS 503..2407
/gene="HWP1"
/note="hyphal surface protein"
/codon_start=1
/transl_table=12
/product="hyphal wall protein 1"
/protein_id="AAC96368.1"
/db_xref="GI:1915979"

/translation="MRLSTAQLIAIAYYMLSIGATVPQVDGQGETEEALIQKRSYDYY
QEPCDDYPQQQQQQEPCDYPQQQQQEEPCDYPQQQPQEPCDYPQQPQEPCDYPQQPQE
PCDYPQQPQEPCDNPPQPDVPCDNPPQPDVPCDNPPQPDIPCDNPPQPDIPCDNPPQP
DQPDDNPPIPNIPTDWIPNIPTDWIPDIPEKPTTPATTPNIPATTTTSESSSSSSSSS
SSTTPKTSASTTPESSVPATTPNTSVPTTSSESTTPATSPESSVPVTSGSSILATTSE
SSSAPATTPNTSVPTTTTETKSSSTPLTTTTEHDTTVVTVTSCSNSVCTESEVTTGVI
VITSKDTIYTTYCPLTETTPVSTAPATETPTGTVSTSTEQSTTVITVTSCSESSCTES
EVTTGVVVVTSEETVYTTFCPLTENTPGTDSTPEASIPPMETIPAGSESSMPAGETSP
AVPKSDVPATESAPVPEMTPAGSQPSIPAGETSPAVPKSDVPATESAPAPEMTPAGTE
TKPAAPKSSAPATEPSPVAPGTESAPAGPGASSSPKSSVLASETSPIAPGAETAPAGS
SGAITIPESSAVVSTTEGAIPTTLESVPLMQPSANYSSVAPISTFEGAGNNMRLTFGA
AIIGIAAFLI"
ORIGIN
1 ggatccaaaa acaaggaatt cggaaattct gacgataaat gtcgactcac aattcattgt
61 aaaaagggag agttttggta ggctcataat cgcttataat gtacctctaa agtaatctaa
121 aacaaacaca acctttctaa aacctataat aataacccta atggctcaca accgggataa
181 gttagttagc ccagctgttt ttttttgcc ttatttttat gactacattt tgtttcactt
241 tttgttgcga ctttaatacc gtttttgcaa cttctctttg tatcacctgt atccgccttt
301 tttaacatag caactcttgt aaagtccctt tcttttccca ctattttatc attcttgaaa
361 tatgtaatca gaatagtttt tcaaaaacta taaataacgg tcaaaataac cggctatttt
421 caatttccat tcaacttgtt ttctcaacaa tatcaaacac aacaggaatc tcctatagtc

481 actcgctttt agtttcgtca atatgagatt atcaactgct caacttattg ctatcgctta
541 ttacatgtta tcaattgggg ccactgtccc acaggtagac ggtcaaggtg aaacagagga
601 agctcttatt caaaagagat cttatgatta ctatcaagaa ccatgtgatg attacccaca
661 acaacaacaa caacaagagc cttgtgatta cccacaacaa caacagcagg aagaaccttg
721 tgattaccca caacaacaac cacaagagcc atgtgactat ccacaacagc cacaagaacc
781 ttgtgactac ccacaacaac cacaagaacc ttgtgactac ccacaacaac cacaagaacc
841 ttgcgacaat ccacctcaac ctgatgttcc ttgtgacaat cctcctcaac ctgatgttcc
901 ttgtgacaat cctcctcaac ctgatattcc ttgtgacaat cctcctcaac ctgatattcc
961 ttgtgacaat cctcctcaac ctgatcagcc tgatgacaat cctcctattc caaacattcc
1021 aaccgattgg attccaaata ttccaactga ttggatccca gatattccag aaaagccaac
1081 aactccagct actactccaa acattcctgc tacaactact acttctgaat catcatcttc
1141 ttcttcttct tcatcatcat ctactactcc aaaaacttct gcttcaacta cacctgaatc
1201 ttctgttcca gctaccactc caaacacttc tgttccaaca acttcttcag aatcaactac
1261 tccagctact agcccagaaa gttctgttcc agttacttct ggatcatcta ttttagctac
1321 cacttcagaa tcatcatctg ctccagctac tactccaaat acatctgttc caaccactac
1381 tactgaaacc aaatcatcaa gtactccatt aactactact actgaacatg atacaactgt
1441 tgtcactgtt acttcatgtt ctaacagtgt ttgtaccgaa agtgaagtta ctactggtgt
1501 tattgtcatc acatctaaag atactattta caccacttac tgtccattga ctgaaactac
1561 tccagtttct actgctccag ccactgaaac accaactggt acagtatcca cttctactga
1621 acaatcaact actgttatta ctgttacttc atgttctgaa agctcttgta ccgaatctga
1681 agttactact ggtgttgttg ttgttacttc tgaggaaact gtctacacta cattctgtcc
1741 attgactgaa aacactccag gtactgattc aactccagaa gcttccattc cacctatgga
1801 aacaattcct gctggttcag aatcatccat gcctgccggt gaaacctctc cagctgttcc
1861 aaaatcagat gttccagcta ctgaatcagc tccagttcct gaaatgactc cagctggttc
1921 acaaccatct attcctgccg gtgaaacctc tccagctgtt ccaaaatcag atgttccagc
1981 tactgaatct gctcctgctc ctgaaatgac tccagctggt actgaaacta aaccagctgc
2041 tccaaaatca tcagctcctg ccactgaacc ttccccagtt gctccaggta ctgaatccgc
2101 accagctggt ccaggtgctt cttcttctcc aaaatcttct gttttggcta gtgaaacctc
2161 accaattgct ccaggtgctg aaaccgctcc agctggctca agtggtgcta ttactattcc
2221 ggaatctagt gctgtcgtct ctacgactga aggtgctatt ccaactacat tagaatcagt
2281 tccactcatg caaccatctg ccaattactc aagtgtcgct cctatttcta catttgaagg
2341 tgctggtaac aacatgagat tgactttcgg tgctgctatt attggtattg ctgcattctt
2401 gatctaattc taataactga tactaagttt tgttcttttt ttgggatttc tttttttct
2461 aattttgatt gtttttcaat tttgggtttt caatattatt gacaagagtc attttattga
2521 atatttgttt tgtttactac attaaaggtg ataggtactt ttagttttta aaaattgttt
2581 tgttcaaatt gtttatcttt ttcttcttct tctacttgct ttgttttctg ttttcggttc
2641 atagttgata gcttttaata aataccccctt ttttttttaca at