WoLF PSORT、蛋白質細胞内局在予測ツール

産業技術総合研究所・ＣＢＲＣ　　藤田直也、原田肇、今井賢一郎、○Paul Horton

東京大学・医科学研究所・ＨＧＣ　中井謙太

概要

WoLF PSORTはアミノ酸配列から蛋白質の細胞内局在部位を予測する。WoLF PSORTの予測法は、PSORTIIプログラムの大幅な改良で、局在化モチーフのほかに、アミノ酸組成にように局在化との因果関係は弱いが局在部位と相関がみられる特徴に基づいて予測を行なう。WoLF PSORTはPSORTとPSORTIIと同じく、予測結果以外にも局在化シグナルについて有効な情報を表示するようになっている。

我々の計算実験(APBC06)ではWoLF PSORTの予測率は80%を越えている。特に、数の多い局在部位では、配列上の類似度があまりなくても、WoLF PSORT から有意義な予測結果が得れ、BLASTのような配列類似度検索ツールを補うことができる。

データセット

WoLF PSORTのデータセットは主にUniprotの記述に基づいて作られている。現在使われているデータセットは動物で12,000個以上、植物と真菌類それぞれ 2,000個以上の蛋白質を含んでいる。データセットは主にUniprotを元にして集めたが、Gene Ontologyから取った、数百個のナズナ(Arabidopsis thaliana) 蛋白質も含まれている。

特徴徴量

PSORT(中井、金久)、とiPSORT(坂内ら)、の特徴の一部分に、アミノ酸組成や 配列の長さを加えた集合を特徴量の候補として採用している。

分類法

我々が開発したWoLFプログラムで特徴量の選択と 重み付けを行なった後、重み付きkNN法で分類を行なう。つまり、クエリ配列を予測する方法は、局在予測用の特徴量から計算した類似度で、データ セットの中からクエリ配列にもっとも類似したk個の蛋白の局在部位を 参考にして予測を行なう。

詳しくは wolfpsort.org

独立比較で高い評価！

Computational classification of classically secreted proteins

Eric W. Klee[1] and Carlos P. Sosa[2,3]

[1] Stabile 3-15, Department of Laboratory Medicine, Mayo Clinic, Rochester, MN 55905, USA
[2] IBM, 650 3rd Ave. South, Minneapolis, MN 53402 USA
[3] University of Minnesota Supercomputing Institute, 599 Walter Library, 117 Pleasant St, Minneapolis, MN 55455m USA

The ability to identify classically secreted proteins is an important component of targeted therapeutic

予測ツール	感度	特異度	相関係数
WoLF PSORT	0.85(0.97)	0.94(0.95)	0.79(0.80)
PredSL	0.69(0.94)	0.94(0.95)	0.67(0.81)
Protein Prowler	0.58(0.91)	0.99(0.99)	0.66(0.90)
MultiLoc	0.78(0.88)	0.87(0.89)	0.66(0.63)
BaCello	0.60(0.91)	0.98(0.98)	0.65(0.88)
LocTree	0.66(0.94)	0.95(0.95)	0.65(0.78)
HSL pred	0.67(0.21)	0.94(0.95)	0.65(0.21)
SignalP 3.0	0.60(0.91)	0.96(0.98)	0.63(0.87)
Phobius	0.56(0.94)	0.96(0.97)	0.60(0.84)
pTarget	0.65(0.79)	0.92(0.92)	0.60(0.63)

参考文献

"WoLF PSORT: Protein Localization Predictor",
Paul Horton, Keun-Joon Park, Takeshi Obayashi, Naoya Fujita, Hajime Harada, C.J. Adams-Collier, & Kenta Nakai,
Nucleic Acids Research, doi:10.1093/nar/gkm259, 2007.

"Protein Subcellular Localization Prediction with WoLF PSORT",
Paul Horton, Keun-Joon Park, Takeshi Obayashi & Kenta Nakai,
Proc. 4th Annual Asia Pacific Bioinformatics Conference APBC06, Taipei, Taiwan. pp. 39-48, 2006.

Under Construction

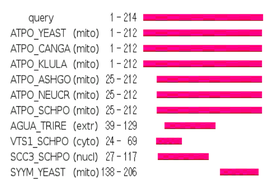
配列類似度とWoLF PSORT特徴量類似度を組み合わせた局在予測とデータベース検索

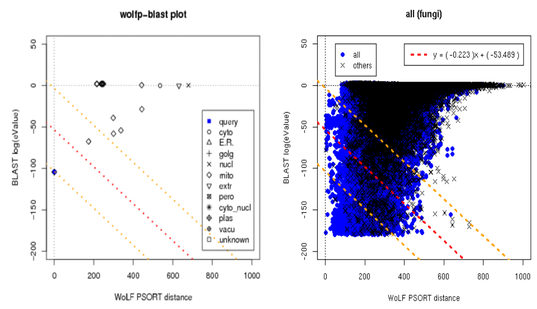
WoLF PSORT DB

result id cf0972b20cd812c9be1df3da33c66c99
database fungi
query query1_mito
seq MFNRVFTRSFASSLRAAASKAAAPPPVRLFGVEGTYATALYQAAAKNSSIDAAFQSLQKVESTVKKNPKLGHLLLNPALS LKDRNSVIDAIVETHKNLDGYVVNLLKVLSENNRLGCFEKIASDFGVLNDAHNGLLKGTVTSAEPLDPKSFKRIEKALSA SKLVGQGKSLKLENVVKPEIKGGLIVELGDKTVDLSISTKIQKLNKVLEDSI
self (blast, wolfp) = (6e-105, 0), eval = 0.0

BLAST							WoLF PSORT						
subject	loc	cluster	c_loc	BLAST E value	WOLFP distance	eval	neighbor	loc	cluster	c_loc	BLAST E value	WOLFP distance	eval
ATPO_YEAST	mito	c0061 (4)		5e-105	0.0 (100.0%)	-0.4	ATPO_YEAST	mito	c0061 (4)		5e-105	0.0 (100.0%)	-0.4
ATPO_CANGA	mito	c0061 (4)		3e-68	173.5 (40.9%)	338.1	SODM_CANAL	mito	c8047 (1)			214.5 (31.0%)	690.8
ATPO_KLULA	mito	c0061 (4)		9e-59	299.8 (19.6%)	506.9	QCR1_YEAST	mito	c4907 (1)			216.5 (30.7%)	692.8
ATPO_ASHGO	mito	c0061 (4)		8e-55	337.3 (16.9%)	562.1	YOS9_SCHPO	cyto	c8250 (1)			235.4 (23.9%)	711.7
ATPO_NEUCR	mito	c8253 (1)		9e-40	298.6 (19.8%)	590.9	PYRD_SACKL	mito	c5073 (1)			238.0 (27.1%)	714.3
ATPO_SCHPO	mito	c8327 (1)		3e-29	443.2 (12.0%)	782.7	RM44_YEAST	mito	c9861 (1)			241.3 (26.6%)	717.6
AGUA_TRIRE	extr	c1803 (1)		0.65	631.2 (0.3%)	1097.7	ARG56_SCHPO	mito	c1666 (1)			242.1 (26.5%)	718.4
VTS1_SCHPO	cyto	c2390 (1)		1.1	536.0 (12.0%)	1003.5	SYWM_YEAST	mito	c5980 (1)			242.8 (26.3%)	719.1
SCC3_SCHPO	nucl	c1429 (1)		1.9	679.5 (11.2%)	1148.1	CH10_YEAST	mito	c9773 (1)			245.7 (25.9%)	722.0
SYYM_YEAST	mito	c4464 (1)		3.2	442.4 (12.0%)	912.0	FMP52_CANGA	mito	c8059 (1)			250.0 (25.2%)	726.3

Count (=#obs): mito (16), cyto (2), extr (1), nucl (1)
U Count (=#obs*wolfu): mito (16), cyto_mito (7.2), mito_nucl (6.8), mito_pero (6.4), golg_mito (6.4), E.R._mito (6.4)
r_freq (=#obs/#exp): mito (5.8), extr (1.8), cyto (1.0), nucl (0.3)
Rank : mito (5.9), cyto (5.0), extr (4.0), nucl (2.0)
Eval : mito (587.1), cyto (857.6), extr (1097.7), nucl (1148.1)
Eval2 (=Eval/#exp): extr (614.7), mito (1620.3), cyto (1801.0), nucl (3857.6)
Eval3 (=Eval/r_freq): mito (101.3), extr (614.7), cyto (900.5), nucl (3857.6)
Eval4 (=Eval3 w/o singleton): mito (101.3), extr (614.7), cyto (900.5), nucl (3857.6)
Eval5 (=100 - Eval4/1500*100) mito (93.2), extr (59.0), cyto (40.0), nucl (-157.2)





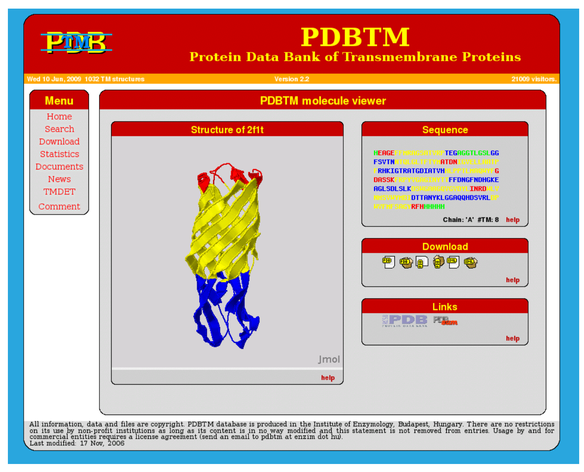
WoLF PSORT次期版では、配列の上に特徴量の一致個所や、WoLFPSORT類似度と配列類似度の判別面、を表示する画面を加える予定。

バクテリアβバレル膜蛋白質予測

P0A915

Uniprot ID OMPW_ECOLI
Uniprot AC P0A915
Uniprot Loc. Cell outer membrane

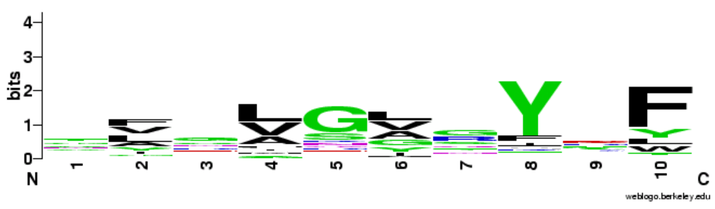
[feature]	score	(positive/negative)
Amph N term	2.27	(2.52/2.12)
Amph C term	2.42	(2.83/2.13)
Amph global	2.27	(2.29/1.84)
N term sig	-123	(-130/-142)
TM motif	-19.3	(-19.9/-25.0)
C term sig	-20.5	(-24.5/-33.7)
Asn	4.25	(6.41/3.74)
Tyr	3.77	(4.65/2.77)
Met	3.30	(1.62/2.70)
Ser	6.13	(7.78/5.86)
Asn-Tyr	0.95	(0.34/0.11)
Gly-Tyr	0.47	(0.65/0.26)
Tyr-Asn	0.00	(0.30/0.11)



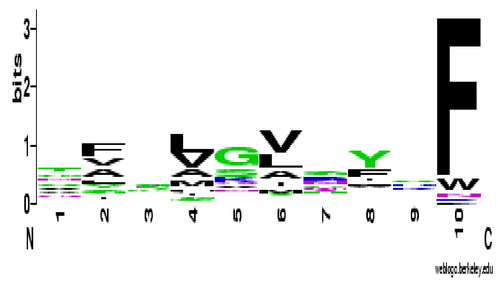

```
seq   MKKLTVAALAVTTLLSGSAFAHEAGEFFMRAGSATVRPTEGAGGTLGSLGGFSVTNNTQLGLTFTYMATDNIGVELLAAT 80
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Amph N
      ------------------------------------------------------------------------------------ Amph C
      NNNHHCCCCCCCCCCCCCCCC------------------------------------------------------------- N  sig
      ------------------------------------------------------------------------------------ TM mtf
      ------------------------------------------------------------------------------------ C  sig

      PFRHKIGTRATGDIATVHHLPPTLMAQWYFGDASSKFRPYVGAGINYTTFFDNGFNDHGKEAGLSDLSLKDSWGAAGQVG 160
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Amph N
      ---------------------------<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<< Amph C
      ------------------------------------------------------------------------------------ N  sig
      ---------------------------------------YVGAGINYTT-------------------------------- TM mtf
      ------------------------------------------------------------------------------------ C  sig

      VDYLINRDWLVNMSVWYMDIDTTANYKLGGAQQHDSVRLDPWVFMFSAGYRF 212
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>> Amph N
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<< Amph C
      ---------------------------------------------------- N  sig
      ---------------------------------------------------- TM mtf
      ------------------------------------------VFMFSAGYRF C  sig
```



内部膜組み込みシグナル



C末端膜組み込みシグナル