

: BLAST (profile)

1. BLAST

(bioinformatics) 가
 BLAST(Basic Local Alignment Search Tool) [1].
 (DNA RNA)
 ,
 (sequence
 database)
 .
 가 BLAST .
 BLAST 가 가 .
 NCBI (<http://www.ncbi.nlm.nih.gov/blast/>) BLAST
 가
 BLAST ()
 . 가 BLAST (executables)
 NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/>)
 . BLAST ()
 NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>)
 . "nr.tar.gz"(nr non-redundant
)
 . 가 ()
 "pdb.tar.gz" ("pataa.tar.gz") .

1: blastall 가

blastn		
blastp		
blastx		
tblastn		
tblastx		

BLAST 가 blastall .
blastall 가 “blastn”, “blastp”, blastx”,
“tblastn”, “tblastx” . BLAST
[2] . blastall .
blastall -p blastp -d nr -i query
“-p” (가)
“-d” (nr
non-redundant). “-i”
FASTA (가
)

BLAST 가 (scoring
matrix) BLOSUM62
(2). BLOSUM BLOcks SUBstitution Matrix
가 ()
[3]. BLOSUM
. 0

가 (homology)가
. BLOSUM62 (2) 가 11 (가
: (W)), 가
4 (8 가 가). (W) (N),
(D), (P) , (G)
(I) (L) , (D) (L)
, (E) (C) , (P)
(F) 가 4 .

2. PSI-BLAST (profile)

BLAST(PSI Position-Specific Iterated) PSI-
[4]. PSI-BLAST BLAST
BLAST .

BLAST , PSI-BLAST
 가 (A) (S) PSI-BLAST
 가 BLOSUM62
 (position specific) 가
 가

2: BLOSUM62 (scoring matrix)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

BLOSUM62

가 PSI-BLAST , PSI-BLAST
 “ ()
 ”(position-specific scoring matrix; PSSM
) . PSSM (profile) ,
 n×20 . n
 , 20 20 가 . PSI-BLAST
 “ ” BLOSUM62

BLAST . PSI-BLAST
 PSSM . PSSM PSI-
 BLAST PSSM .
 (iterated) 가 PSSM .

PSI-BLAST BLAST
 “blastpgp” . blastpgp
 blastpgp -d nr -i query -j 3 -h 0.001 -C query.chk
 “-d” “-i” blastall . “-j” PSI-
 BLAST (iteration) . “-h” PSSM
 (threshold) . “-h” 0.001
 (default) . “-C” PSSM (query.chk)
 . PSSM
 (ASCII) BLAST
 “makemat” .

3.

(secondary structure) 가
 PSIPRED [5], (domain boundary)
 가 PPRODO [6], (solvent
 accessibility) SVMpsi [7]
 PSSM .
 (tertiary structure) 가
 . ab initio 가
 (fragment assembly) [8,9]

threading 가 (fold recognition) threading
 alignment) threading (profile - profile [10,11].
 PSI-BLAST (comparative homology modeling)

가

(<http://www.cheric.org>).

[]

- [1] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* **215** (1990) 403-410.
- [2] I. Korf, M. Yandell, and J. Bedell, *BLAST* (2003) O'Reilly.
- [3] S. Henikoff and J. G. Henikoff, Amino acid substitution matrices from protein blocks, *Proc. Natl. Acad. Sci. U.S.A.* **89** (1992) 10915-10919.
- [4] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25** (1997) 3389-3402.
- [5] D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol.* **292** (1999) 195-202.
- [6] J. Sim, S. -Y. Kim, and J. Lee, PPRODO: prediction of protein domain boundaries using neural network, *Proteins: Structure, Function, and*

Bioinformatics **59** (2005) 627-632.

- [7] H. Kim and H. Park, Prediction of protein relative solvent accessibility with support vector machines and long-range interaction 3D local descriptor, *Proteins: Structure, Function, and Bioinformatics* **54** (2004) 557-562.
- [8] K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, *J. Mol. Biol.* **268** (1997) 209-225.
- [9] J. Lee, S.-Y. Kim, K. Joo, I. Kim, and J. Lee, Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing, *Proteins: Structure, Function, and Bioinformatics* **56** (2004) 704-714.
- [10] K. Ginalski *et al.*, ORFeus: detection of distant homology using sequence profiles and predicted secondary structure, *Nucleic Acids Res.* **31** (2003) 3804-3807.
- [11] J. Sim, S.-Y. Kim, J. Lee, and A. Yoo, Predicting the three-dimensional structures of proteins: combined alignment approach, *J. Korean Phys. Soc.* **44** (2004) 611-616.