

ESERCIZIO 1

Comparazione di FASTA e BLAST

>1O5X:B | PDBID | CHAIN | SEQUENCE

```
MARKYFVAANWKCNGTLESIKSLTNSFNLLDFPSKLDVVVFPVSVHYDHRKLLQSKFSTGIQNVSKFGNGSYTGEVSA  
EIAKDLNIEYVIIGHFERRKYFHETDEDVREKLOASLKNNLKAVVCFGESLEQREONKTIEVITKQVKAFVDLIDNFDNV  
ILVYEPLWAIGTGKTATPEQAQLVHKEIRKIVKDTTCGEKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF  
VDIIKSAM
```

è stata utilizzata per confrontare i risultati ottenibili con FASTA e BLAST

Esaminate le tabelle riassuntive della ricerca effettuata in DB con i due programmi e traete le vostre conclusioni sui seguenti punti:

Quali sono i parametri di ricerca usati rispettivamente dai due programmi?

In particolare

Quale database?

BLAST
FASTA

In quale delle due ricerche sono state confrontate il maggior numero di sequenze in DB?

BLAST
FASTA

Quale finestra?

BLAST
FASTA

Quale score o threshold?

BLAST
FASTA

Quale matrice di sostituzione?

BLAST
FASTA

Quale penalità ai Gap?

BLAST
FASTA

Quale parametro è stato utilizzato per valutare la significatività statistica?

BLAST
FASTA

Le sequenze allineate in FASTA si ritrovano nella tabella dei risultati di BLAST?

Se non si ritrovano quale è il motivo di tale risultato?

Da quale specie sono state tratte le sequenze allineate che vi sono mostrate e qual è il loro nome comune?

/ebi/extserv/bin/fasta-35.4.10/fasta35_t -l /ebi/services/idata/v2738/fastacfg/fasta3db
-Q -H -p -b 50 -d 50 -s BL50 -E 10.0 -f -10 -g -2 -z 1 @:1- +uniprot+ 2
FASTA searches a protein or DNA sequence data bank
version 35.04 Jan. 24, 2010
Please cite:
W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query: @
1>>>Sequence - 248 aa
Library: UniProt 3623243416 residues in 11163426 sequences

3623243416 residues in 11163426 sequences
Statistics: Expectation_n fit: rho(ln(x))= 5.3130+/-0.000183; mu= 6.0417+/- 0.010
mean_var=74.3799+/-15.170, 0's: 14 Z-trim: 31 B-trim: 0 in 0/67
Lambda= 0.148712
statistics sampled from 60000 to 11159454 sequences
Algorithm: FASTA (3.5 Sept 2006) [optimized]
Parameters: BL50 matrix (15:-5) ktup: 2
join: 42, opt: 30, open/ext: -10/-2, width: 16
Scan time: 498.000

The best scores are:

	opt	bits	E(11163426)
UNIPROT:TPIS_PLAF7 Q7KQM0 Triosephosphate isomeras (248)	1595	351.4	1.2e-94
UNIPROT:TPIS_PLAFA Q07412 Triosephosphate isomeras (248)	1595	351.4	1.2e-94
UNIPROT:Q4YHF9_PLABE Q4YHF9 Triosephosphate isomer (248)	1390	307.4	2e-81
UNIPROT:Q4XU30_PLACH Q4XU30 Triosephosphate isomer (248)	1377	304.6	1.4e-80
UNIPROT:A5K411_PLAVI A5K411 Triosephosphate isomer (248)	1363	301.6	1.1e-79
UNIPROT:B3LA18_PLAKH B3LA18 Triosephosphate isomer (248)	1345	297.7	1.6e-78
UNIPROT:Q7RGP0_PLAYO Q7RGP0 Triosephosphate isomer (210)	1166	259.3	5.2e-67
UNIPROT:Q1KSE2_TOXGO Q1KSE2 Triosephosphate isomer (252)	834	188.1	1.7e-45
UNIPROT:B9Q7W9_TOXGO B9Q7W9 Triosephosphate isomer (253)	834	188.1	1.7e-45
UNIPROT:B9PMJ9_TOXGO B9PMJ9 Triosephosphate isomer (253)	834	188.1	1.7e-45
UNIPROT:B6KEI3_TOXGO B6KEI3 Triosephosphate isomer (253)	834	188.1	1.7e-45
UNIPROT:Q4P9B4_USTMA Q4P9B4 Triosephosphate isomer (248)	788	178.2	1.5e-42
UNIPROT:Q4N293_THEPA Q4N293 Triosephosphate isomer (250)	787	178.0	1.8e-42
UNIPROT:D1H4M5_VITVI D1H4M5 Triosephosphate isomer (254)	782	176.9	3.8e-42
UNIPROT:B9VRG4_9ROSI B9VRG4 Triosephosphate isomer (254)	781	176.7	4.4e-42
UNIPROT:D0NKB9_PHYIN D0NKB9 Triosephosphate isomer (250)	780	176.5	5.1e-42
UNIPROT:Q38IW8_SOYBN Q38IW8 Triosephosphate isomer (253)	777	175.9	8e-42
UNIPROT:Q5CMF9_CRYHO Q5CMF9 Triosephosphate isomer (250)	775	175.4	1.1e-41
UNIPROT:Q5JZZ3_PHAVU Q5JZZ3 Triosephosphate isomer (254)	775	175.4	1.1e-41
UNIPROT:Q6GW08_SOYBN Q6GW08 Triosephosphate isomer (253)	774	175.2	1.3e-41
UNIPROT:B9GJN0_POPTR B9GJN0 Triosephosphate isomer (255)	770	174.4	2.3e-41
UNIPROT:A4RLV9_MAGGR A4RLV9 Triosephosphate isomer (250)	768	173.9	3e-41
UNIPROT:Q5CSE7_CRYPV Q5CSE7 Triosephosphate isomer (250)	764	173.1	5.5e-41
UNIPROT:A5AZX9_VITVI A5AZX9 Triosephosphate isomer (250)	764	173.1	5.5e-41
UNIPROT:Q0H294_PTEVI Q0H294 Triosephosphate isomer (253)	760	172.2	1e-40
UNIPROT:A7F1G3_SCLS1 A7F1G3 Triosephosphate isomer (247)	757	171.6	1.5e-40
UNIPROT:Q3HRV9_SOLTU Q3HRV9 Triosephosphate isomer (257)	757	171.6	1.6e-40
UNIPROT:TPIS_COPJA P21820 Triosephosphate isomeras (253)	756	171.4	1.8e-40
UNIPROT:A6S6P3_BOTFB A6S6P3 Triosephosphate isomer (247)	755	171.1	2.1e-40
UNIPROT:C6TKJ5_SOYBN C6TKJ5 Triosephosphate isomer (252)	754	170.9	2.4e-40
UNIPROT:TPIS_ARATH P48491 Triosephosphate isomeras (254)	754	170.9	2.5e-40
UNIPROT:Q4U9J9_THEAN Q4U9J9 Triosephosphate isomer (329)	755	171.2	2.6e-40
UNIPROT:D1IDL4_VITVI D1IDL4 Triosephosphate isomer (254)	753	170.7	2.9e-40
UNIPROT:A5BV65_VITVI A5BV65 Triosephosphate isomer (254)	753	170.7	2.9e-40
UNIPROT:D1LWT8_9ROSI D1LWT8 Triosephosphate isomer (254)	753	170.7	2.9e-40
UNIPROT:A9PE68_POPTR A9PE68 Triosephosphate isomer (254)	750	170.1	4.5e-40
UNIPROT:B2W7G7_PYRTR B2W7G7 Triosephosphate isomer (248)	748	169.6	5.9e-40
UNIPROT:A9NRN6_PICSI A9NRN6 Triosephosphate isomer (253)	748	169.6	6e-40
UNIPROT:B6AI68_9CRYPT B6AI68 Triosephosphate isomer (249)	747	169.4	6.9e-40
UNIPROT:Q0TWR8_PHANO Q0TWR8 Triosephosphate isomer (248)	744	168.8	1.1e-39
UNIPROT:TPIS_SCHPO P07669 Triosephosphate isomeras (249)	744	168.8	1.1e-39
UNIPROT:B3TLL4_ELAGV B3TLL4 Triosephosphate isomer (254)	741	168.1	1.7e-39
UNIPROT:A2I3Y4_MACHI A2I3Y4 Triosephosphate isomer (247)	739	167.7	2.2e-39
UNIPROT:C1BV50_9MAXI C1BV50 Triosephosphate isomer (249)	739	167.7	2.3e-39
UNIPROT:Q33BU3_CHLPY Q33BU3 Triosephosphate isomer (251)	739	167.7	2.3e-39
UNIPROT:Q22T89_TETTH Q22T89 Triosephosphate isomer (256)	739	167.7	2.3e-39
UNIPROT:A1CBS9_ASPLC A1CBS9 Triosephosphate isomer (249)	738	167.5	2.6e-39
UNIPROT:Q69K00_ORYSJ Q69K00 Triosephosphate isomer (304)	739	167.8	2.6e-39

UNIPROT:A2Z3G7_ORYSI A2Z3G7 Triosephosphate isomer (304) 739 167.8 2.6e-39
UNIPROT:TPIS_PETHY P48495 Triosephosphate isomeras (254) 738 167.5 2.7e-39

>>UNIPROT:TPIS_PLAF7 Q7KQM0 Triosephosphate isomerase OS (248 aa)
initn: 1595 initl: 1595 opt: 1595 Z-score: 1858.4 bits: 351.4 E(): 1.2e-94
Smith-Waterman score: 1595; 99.6% identity (100.0% similar) in 248 aa overlap (1-248:1-248)

```
          10      20      30      40      50      60
Sequen  MARKYFVAANWKCNGTLESIKSLTNSFNLDLDFPSKLDVVVFPVSVHYDHTRKLQSKFS
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  MARKYFVAANWKCNGTLESIKSLTNSFNLDLDFPSKLDVVVFPVSVHYDHTRKLQSKFS
          10      20      30      40      50      60

          70      80      90     100     110     120
Sequen  TGIQNVSKFNGSYTGEVSAEIAKDLNIEYVVIIGHFERRKYFHETDEDVREKLQASLKNN
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  TGIQNVSKFNGSYTGEVSAEIAKDLNIEYVVIIGHFERRKYFHETDEDVREKLQASLKNN
          70      80      90     100     110     120

          130     140     150     160     170     180
Sequen  LKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILVYEPLWAIAGTGKTATPEQ
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  LKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILAYEPLWAIAGTGKTATPEQ
          130     140     150     160     170     180

          190     200     210     220     230     240
Sequen  AQLVHKEIRKIVKDTCEGKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  AQLVHKEIRKIVKDTCEGKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF
          190     200     210     220     230     240
```

```
Sequen  VDIIKSAM
        : : : : : :
UNIPRO  VDIIKSAM
```

>>UNIPROT:Q4U9J9_THEAN Q4U9J9 Triosephosphate isomerase (329 aa)
initn: 606 initl: 430 opt: 755 Z-score: 882.7 bits: 171.2 E(): 2.6e-40
Smith-Waterman score: 755; 46.1% identity (77.1% similar) in 245 aa overlap (3-246:84-327)

```
          10      20      30
Sequen  MARKYFVAANWKCNGTLESIKSLTNSFNLDLDF
        . . . . . : : : : : : : : : : : : : : : : : : :
UNIPRO  GSSNNGVLTQRKLLRCLIKRFIIFYLKYIKRKWLGWNWKCNGTKQISIDLLLFNRHQ
          60      70      80      90      100     110

          40      50      60      70      80      90
Sequen  DPSKLDVVVFPVSVHYDHTRKLQSK-FSTGIQNVSKFNGSYTGEVSAEIAKDLNIEYV
        . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  KNNNLDVVLFPPLSYVEHTRNELKSEVFELGVQNVQSQSKGAFTGELSLTMFTDFGLKWS
          120     130     140     150     160     170

          100     110     120     130     140     150
Sequen  IIGHFERRKYFHETDEDVREKLQASLKNNLKAVVCFGESLEQREQNKTIEVITKQVKAFV
        . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  LVGHSERRQLFNEDDSYVCEKVMMLQENGVNAVVCFGETLSEREQQTENVLKRQLDAFV
          180     190     200     210     220     230

          160     170     180     190     200     210
Sequen  DLIDNFDNVILVYEPLWAIAGTGKTATPEQAQLVHKEIRKIVKDTCEGKQANQIRILYGG
        . . . . . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
UNIPRO  KHVKDWDKVVLAYEPVWAIAGTGKVAIVDQVKEAHKFVRDYVRGLVGDV-ADKVRVLVYGG
          240     250     260     270     280     290

          220     230     240
Sequen  VNTENCSSLIQQEDIDGFLVGNASLKESFVDIIKSAM
```


BLASTP 2.2.23+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Reference for compositional score matrix adjustment: Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", *FEBS J.* 272:5101-5109.

RID: WT0811JU01N

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

10,871,739 sequences; 3,704,264,808 total letters

Query= 105X:B|PDBID|CHAIN|SEQUENCE

Length=248

Sequences producing significant alignments:	Score (Bits)	E Value
pdb 1YDV A Chain A, Triosephosphate Isomerase (Tim) >pdb 1YDV...	511	2e-143
ref XP_001348552.1 triosephosphate isomerase [Plasmodium fal...	510	6e-143
pdb 2VFH A Chain A, Crystal Structure Of The F96w Mutant Of P...	509	1e-142
pdb 2VFF A Chain A, Crystal Structure Of The F96h Mutant Of P...	509	1e-142
pdb 1VGA A Chain A, Structures Of Unligated And Inhibitor Com...	509	1e-142
pdb 2VFD A Chain A, Crystal Structure Of The F96s Mutant Of P...	508	2e-142
ref XP_674608.1 triose-phosphate isomerase [Plasmodium bergh...	442	1e-122
ref XP_744298.1 triose-phosphate isomerase [Plasmodium chaba...	438	2e-121
ref XP_001615992.1 triosephosphate isomerase [Plasmodium viv...	431	3e-119
ref XP_002260474.1 triose-phosphate isomerase [Plasmodium kn...	426	1e-117
ref XP_724583.1 triosephosphate isomerase [Plasmodium yoelii...	373	1e-101
gb ABE76515.1 triose-phosphate isomerase 1 [Toxoplasma gondii]	262	2e-68
ref XP_002366256.1 triosephosphate isomerase, putative [Toxo...	262	2e-68
ref XP_759446.1 hypothetical protein UM03299.1 [Ustilago may...	252	3e-65
ref XP_002283671.1 PREDICTED: hypothetical protein [Vitis vi...	249	3e-64
gb EEY59956.1 triosephosphate isomerase [Phytophthora infest...	248	4e-64
gb ABA86966.1 triosephosphate isomerase [Glycine max] >gb AC...	248	6e-64
emb CAI43251.1 triose-phosphate isomerase [Phaseolus vulgaris]	247	7e-64
gb AAT46998.1 triosephosphate isomerase [Glycine max]	247	1e-63
ref XP_002299871.1 predicted protein [Populus trichocarpa] >...	246	2e-63
emb CAN67342.1 hypothetical protein [Vitis vinifera]	244	7e-63
ref XP_364060.2 triosephosphate isomerase [Magnaporthe grise...	243	1e-62
gb ABC59698.1 triosephosphate isomerase [Pteris vittata]	243	1e-62
gb ABA46792.1 triosephosphate isomerase-like protein [Solanum...	243	2e-62
ref XP_668007.1 triose-phosphate isomerase [Cryptosporidium ...	241	5e-62
ref XP_002283693.1 PREDICTED: hypothetical protein [Vitis vi...	241	5e-62
ref NP_191104.1 TPI (TRIOSEPHOSPHATE ISOMERASE); triose-phos...	241	5e-62
gb AAA03449.1 cytosolic triose phosphate isomerase [Arabidop...	241	6e-62
ref XP_764099.1 triosephosphate isomerase [Theileria parva s...	241	7e-62
emb CAN70587.1 hypothetical protein [Vitis vinifera]	240	9e-62
gb ACU23435.1 unknown [Glycine max]	239	1e-61
pdb 3KRS A Chain A, Structure Of Triosephosphate Isomerase Fr...	238	4e-61
ref XP_001587441.1 triosephosphate isomerase [Sclerotinia sc...	238	4e-61
ref XP_628120.1 triosephosphate isomerase [Cryptosporidium p...	238	5e-61
ref XP_001552990.1 hypothetical protein BC1G_08882 [Botryoti...	237	1e-60
sp P48495.1 TPIS_PETHY RecName: Full=Triosephosphate isomeras...	236	1e-60
ref XP_001806125.1 hypothetical protein SNOG_15994 [Phaeosph...	236	1e-60
gb ACF06445.1 triose phosphate isomerase cytosolic isoform [...	236	2e-60
gb ACJ11723.1 triosephosphate isomerase [Gossypium hirsutum]	236	2e-60
ref XP_001936088.1 triosephosphate isomerase [Pyrenophora tr...	236	2e-60

gb ACM67308.2	triosphosphate isomerase-like protein type I [...	235	3e-60
gb ACO12903.1	Triosephosphate isomerase [Lepeophtheirus salm...	234	5e-60
ref XP_002533147.1	triosephosphate isomerase, putative [Rici...	234	5e-60
ref XP_002455200.1	hypothetical protein SORBIDRAFT_03g006130...	234	5e-60
gb AAR04017.2	chloroplast trisophosphate isomerase [Euglena ...	234	6e-60
gb ACG24648.1	triosephosphate isomerase, cytosolic [Zea mays...	234	7e-60
ref NP_588024.1	triosephosphate isomerase [Schizosaccharomyc...	234	8e-60
gb AAU93945.1	triose phosphate isomerase [Helicosporidium sp...	233	1e-59
ref NP_001063777.1	Os09g0535000 [Oryza sativa (japonica cult...	233	1e-59
sp P46226.3	TPIS_SECCE RecName: Full=Triosephosphate isomeras...	233	1e-59
gb ADD38004.1	Triosephosphate isomerase B [Lepeophtheirus sa...	233	1e-59
gb EAZ09878.1	hypothetical protein OsI_32171 [Oryza sativa I...	233	1e-59
ref XP_001274623.1	triosephosphate isomerase [Aspergillus cl...	233	1e-59
ref XP_386878.1	hypothetical protein FG06702.1 [Gibberella z...	233	2e-59
gb ABM55553.1	putative triosephosphate isomerase [Maconellic...	233	2e-59
ref NP_001140424.1	triosephosphate isomerase, cytosolic [Zea...	233	2e-59
ref XP_002314179.1	predicted protein [Populus trichocarpa] >...	232	3e-59
ref XP_002175311.1	triosephosphate isomerase [Schizosaccharo...	232	3e-59
ref XP_001744473.1	hypothetical protein [Monosiga brevicolli...	232	3e-59
gb ABE11555.1	cytosolic triosephosphate isomerase [Euglena i...	232	3e-59
ref XP_002142258.1	triosephosphate isomerase [Cryptosporidiu...	231	4e-59
gb AAR11379.1	triose phosphate isomerase cytosolic isoform [...	231	4e-59
ref XP_001008794.2	triosephosphate isomerase [Tetrahymena th...	231	5e-59
gb AAV65492.1	plastid triosephosphate isomerase [Euglena longa]	231	5e-59
gb EEU46622.1	triose-phosphate isomerase [Nectria haematococ...	231	6e-59
ref NP_001132639.1	hypothetical protein LOC100194114 [Zea ma...	231	6e-59
emb CAC14917.1	triosephosphat-isomerase [Triticum aestivum]	230	9e-59
gb ABB02628.1	triose phosphate isomerase cytosolic isoform-l...	230	9e-59
gb ACN33440.1	unknown [Zea mays]	230	1e-58
ref XP_953129.1	triosephosphate isomerase [Theileria annulat...	230	1e-58
ref XP_002426302.1	triosephosphate isomerase, putative [Pedi...	230	1e-58
gb EEY64624.1	triosephosphate isomerase [Phytophthora infest...	230	1e-58
ref XP_002462733.1	hypothetical protein SORBIDRAFT_02g031030...	229	2e-58
ref XP_001633516.1	predicted protein [Nematostella vectensis...	229	2e-58
gb ACF85433.1	unknown [Zea mays] >gb ACR37699.1 unknown [Ze...	229	2e-58
ref NP_001130128.1	hypothetical protein LOC100191222 [Zea ma...	229	2e-58
ref XP_002621266.1	triosephosphate isomerase [Ajellomyces de...	229	3e-58
ref NP_001152578.1	triosephosphate isomerase [Zea mays] >gb ...	229	3e-58
ref XP_002340271.1	triosephosphate isomerase [Talaromyces st...	228	3e-58
sp Q9C401.1	TPIS_ZYGBA RecName: Full=Triosephosphate isomeras...	228	3e-58
ref XP_002117956.1	hypothetical protein TRIADDRAFT_38509 [Tr...	228	4e-58
ref XP_001425288.1	hypothetical protein [Paramecium tetraure...	228	4e-58
emb CAF90849.1	unnamed protein product [Tetraodon nigroviridis]	228	5e-58
sp P46225.1	TPIC_SECCE RecName: Full=Triosephosphate isomeras...	228	6e-58
gb ACY66804.1	triosphosphate isomerase-like protein type II ...	228	7e-58
gb EFE30002.1	hypothetical protein ARB_03343 [Arthroderma be...	227	7e-58
sp P21820.1	TPIS_COPJA RecName: Full=Triosephosphate isomeras...	227	7e-58
ref XP_002145107.1	triosephosphate isomerase [Penicillium ma...	227	7e-58
ref XP_001600119.1	PREDICTED: similar to triosephosphpate iso...	227	8e-58
gb ABD92704.1	triosephosphate isomerase 1 [Sus scrofa]	227	9e-58
ref NP_001133174.1	triosephosphate isomerase 1b [Salmo salar...	227	9e-58
ref NP_001131642.1	hypothetical protein LOC100193000 [Zea ma...	227	1e-57
ref XP_001419463.1	predicted protein [Ostreococcus lucimarin...	227	1e-57
gb AAV65489.1	chloroplast triosephosphate isomerase [Porphyr...	227	1e-57
emb CAA45835.1	triosephosphate isomerase + glyceraldehyde-3-...	227	1e-57
ref XP_975493.2	PREDICTED: similar to triosephosphate isomer...	227	1e-57
ref XP_001422959.1	predicted protein [Ostreococcus lucimarin...	227	1e-57
gb EER00893.1	triose-phosphate isomerase 1, putative [Perkin...	226	1e-57
ref XP_002311168.1	predicted protein [Populus trichocarpa] >...	226	1e-57
gb EER13902.1	triose-phosphate isomerase 1, putative [Perkin...	226	1e-57

ALIGNMENTS

```

>pdb|1YDV|A Chain A, Triosephosphate Isomerase (Tim)
  pdb|1YDV|B Chain B, Triosephosphate Isomerase (Tim)
  pdb|1M7O|A Chain A, Plasmodium Falciparum Triosephosphate Isomerase (Pftim)
  Compled To Substrate Analog 3-Phosphoglycerate (3pg)
  pdb|1M7O|B Chain B, Plasmodium Falciparum Triosephosphate Isomerase (Pftim)
  Compled To Substrate Analog 3-Phosphoglycerate (3pg)
  pdb|1M7P|A Chain A, Plasmodium Falciparum Triosephosphate Isomerase (Pftim)
  Compled To Substrate Analog Glycerol-3-Phosphate (G3p).
  pdb|1M7P|B Chain B, Plasmodium Falciparum Triosephosphate Isomerase (Pftim)

```

Compled To Substrate Analog Glycerol-3-Phosphate (G3p).
 pdb|1LYX|A Chain A, Plasmodium Falciparum Triosephosphate Isomerase (Pftim)-
 Phosphoglycolate Complex
 pdb|1LZO|A Chain A, Plasmodium Falciparum Triosephosphate Isomerase- Phosphoglycolate
 Complex
 pdb|1LZO|B Chain B, Plasmodium Falciparum Triosephosphate Isomerase- Phosphoglycolate
 Complex
 pdb|1LZO|C Chain C, Plasmodium Falciparum Triosephosphate Isomerase- Phosphoglycolate
 Complex
 pdb|1LZO|D Chain D, Plasmodium Falciparum Triosephosphate Isomerase- Phosphoglycolate
 Complex
 pdb|105X|A Chain A, Plasmodium Falciparum Tim Complexed To 2-Phosphoglycerate
 pdb|105X|B Chain B, Plasmodium Falciparum Tim Complexed To 2-Phosphoglycerate
 pdb|2VFI|A Chain A, Crystal Structure Of The Plasmodium Falciparum Triosephosphate
 Isomerase In The Loop Closed State With 3- Phosphoglycerate
 Bound At The Active Site And Interface
 pdb|2VFI|B Chain B, Crystal Structure Of The Plasmodium Falciparum Triosephosphate
 Isomerase In The Loop Closed State With 3- Phosphoglycerate
 Bound At The Active Site And Interface
 Length=248

Score = 511 bits (1317), Expect = 2e-143, Method: Compositional matrix adjust.
 Identities = 248/248 (100%), Positives = 248/248 (100%), Gaps = 0/248 (0%)

Query	1	MARKYFVAANWKCNGTLESIKSLTNSFNLLDFDPSKLDVVVFPVSVHYDHTRKLQSKFS	60
		MARKYFVAANWKCNGTLESIKSLTNSFNLLDFDPSKLDVVVFPVSVHYDHTRKLQSKFS	
Sbjct	1	MARKYFVAANWKCNGTLESIKSLTNSFNLLDFDPSKLDVVVFPVSVHYDHTRKLQSKFS	60
Query	61	TGIQNVSKFGNGSYTGEVSAEIAKDLNIEYVIIGHFERRKYFHETDEDVREKLQASLKNN	120
		TGIQNVSKFGNGSYTGEVSAEIAKDLNIEYVIIGHFERRKYFHETDEDVREKLQASLKNN	
Sbjct	61	TGIQNVSKFGNGSYTGEVSAEIAKDLNIEYVIIGHFERRKYFHETDEDVREKLQASLKNN	120
Query	121	LKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILVYEPLWAIGTGKTATPEQ	180
		LKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILVYEPLWAIGTGKTATPEQ	
Sbjct	121	LKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILVYEPLWAIGTGKTATPEQ	180
Query	181	AQLVHKEIRKIVKDTTCGEKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF	240
		AQLVHKEIRKIVKDTTCGEKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF	
Sbjct	181	AQLVHKEIRKIVKDTTCGEKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKESF	240
Query	241	VDIIKSAM 248	
		VDIIKSAM	
Sbjct	241	VDIIKSAM 248	

 >ref|XP_002311168.1| predicted protein [Populus trichocarpa]
 gb|ABK94671.1| unknown [Populus trichocarpa]
 gb|EEE88535.1| predicted protein [Populus trichocarpa]
 Length=254

Score = 226 bits (577), Expect = 1e-57, Method: Compositional matrix adjust.
 Identities = 115/248 (46%), Positives = 155/248 (62%), Gaps = 1/248 (0%)

Query	1	MARKYFVAANWKCNGTLESIKSLTNSFNLLDFDPSKLDVVVFPVSVHY-DHTRKLQSKF	59
		MARK+V NWKCNGT E +K + ++ NN S + VV + + L+ F	
Sbjct	1	MARKFFVGGNWKCNGTSEEVKIVSALNNSQVPSSDVVEVVVSPFFVFLPLVKSTLRPDF	60
Query	60	STGIQNVSKFGNGSYTGEVSAEIAKDLNIEYVIIGHFERRKYFHETDEDVREKLQASLKN	119
		QN G++TGEVSAE+ +L+I +VI+GH ERR +E++E V +K+ +L	
Sbjct	61	HVAAQNCWVKKGGAFTGEVSAEMLVNLDIPWVILGHSERRSLLNESNEFVGDKVAYALSQ	120
Query	120	NLKAVVCFGESLEQREQNKTIEVITKQVKAFVDLIDNFDNVILVYEPLWAIGTGKTATPE	179
		LK + C GE+LEQRE T+EV+ Q KA + N+ +V+L YEP+WAIGTGK ATP	
Sbjct	121	GLKVIACVGETLEQREAGSTVEVVAAQTKAIAARVSNWADVVLAYEPVWAIGTGKVATPA	180
Query	180	QAQLVHKEIRKIVKDTTCGEKQANQIRILYGGSVNTENCSSLIQQEDIDGFLVGNASLKES	239

Sbjct 181 QAQ VH E+RK +++ + A RI+YGGSV+ NC L + D+DGFLVG ASLK
QAQEVHYELRKWLQENTSP EVAATTRIIYGGSVSGANCKELAAKPDVDGFLVGGASLKPE 240

Query 240 FVDIIKSA 247
F DIIKSA

Sbjct 241 FNDIIKSA 248

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF
excluding environmental samples from WGS projects

Posted date: Apr 18, 2010 5:42 PM

Number of letters in database: -590,702,484

Number of sequences in database: 10,871,739

Lambda K H
0.317 0.133 0.377

Gapped

Lambda K H
0.267 0.0410 0.140

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Number of Sequences: 10871739

Number of Hits to DB: 168392806

Number of extensions: 6672291

Number of successful extensions: 20502

Number of sequences better than 100: 377

Number of HSP's better than 100 without gapping: 0

Number of HSP's gapped: 19883

Number of HSP's successfully gapped: 378

Length of query: 248

Length of database: 3704264808

Length adjustment: 134

Effective length of query: 114

Effective length of database: 2247451782

Effective search space: 256209503148

Effective search space used: 256209503148

T: 11

A: 40

X1: 16 (7.3 bits)

X2: 38 (14.6 bits)

X3: 64 (24.7 bits)

S1: 41 (20.4 bits)

S2: 70 (31.6 bits)

ESERCIZIO 2

Identificazione di ortologhi e paraloghi

The vascular endothelial growth receptor VEGFR-1 (Swissprot Accession Number P17948) has a 332 AA protein kinase domain. Extract the sequence of this domain and try to answer the following questions.

Use the BLAST tools available at the European Bioinformatics Institute at <http://www.ebi.ac.uk/blastall/>. We propose to use the NCBI-BLAST2 program

- a) Identify the closest ortholog stored in UniProt.
- b) Identify the closest paralog stored in UniProt.
- c) The top-hit in the previous BLAST searches is obviously the query sequence itself; a perfect match, 100% identical residues, no gaps. Still, the associated E-Value is not 0. Why?
- d) Repeat the search with a reduced query length (e.g. take the 30 first AAs only) and watch the associated E-Values. What happens?

ESERCIZIO 3

Comparazione di FASTA e BLAST

Utilizzando la query

```
>pdb|4bc1A00
TTTAHSDYEIILEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLDGVVRFPTTKIESVVDSVKNTLNVEVDIANE
TKDRRIAVGEGSLSVGDFSHSFSFEGSVVNMYRSDAVRRNIPNPIYMQGRQFHDILMKVPLDNNDLVDTWEGFQQSIG
ANFGDWIREFWFIGPAFAAINEGGQRISPIVVNSSNVEGGPVGVTRWKF SHAGSGVVDSISRWTELPVEQLNKPASIEG
GFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGMVKGFNDFTVDTQLKIVLPKGYKIRYAAPQFRSQNLE
EYRWSGGAYARWVEHVCKGGTGQFEVLYA
```

effettua una ricerca con FASTA contro SwissProt (<http://www.ebi.ac.uk/Tools/fasta33/>) e con la stessa query con BLAST contro SwissProt (<http://www.ebi.ac.uk/Tools/blastall/>) e confronta gli elenchi ottenuti.

Quali sono i parametri di ricerca usati rispettivamente dai due programmi?

In particolare

Quale database?

BLAST

FASTA

In quale delle due ricerche sono state confrontate il maggior numero di sequenze in DB?

BLAST

FASTA

Quale finestra?

BLAST

FASTA

Quale score o threshold?

BLAST

FASTA

Quale parametro è stato utilizzato per valutare la significatività statistica?

BLAST

FASTA

Le sequenze allineate in FASTA si ritrovano nella tabella dei risultati di BLAST?

Se non si ritrovano quale è il motivo di tale risultato?

Da quale specie sono state tratte le sequenze allineate che vi sono mostrate e qual è il loro nome comune?

BLASTP 2.2.23 [Feb-03-2010]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= Sequence
(349 letters)

Database: swissprot
516,081 sequences; 181,677,051 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
SW:BCPA_PROAE P11741 Bacteriochlorophyll a protein OS=Prostheco...	697	0.0
SW:BCPA_CHLTE Q46393 Bacteriochlorophyll a protein OS=Chlorobium...	582	e-165
SW:BCPA_CHLLT Q46135 Bacteriochlorophyll a protein (Fragment) OS...	575	e-163
SW:B3GL2_MOUSE Q8BG28 UDP-GalNAc:beta-1,3-N-acetylgalactosaminyl...	34	1.8
SW:B3GL2_HUMAN Q8NCR0 UDP-GalNAc:beta-1,3-N-acetylgalactosaminyl...	33	3.9
SW:PDDC1_DANRE Q3B7H1 Parkinson disease 7 domain-containing prot...	32	5.1

>SW:BCPA_PROAE P11741 Bacteriochlorophyll a protein OS=Prosthecochloris aestuarii
PE=1 SV=1
Length = 366

Score = 697 bits (1800), Expect = 0.0
Identities = 348/358 (97%), Positives = 348/358 (97%), Gaps = 9/358 (2%)

Query: 1 TTTAHSYDIIELEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLD---GVVRF 57
TTTAHSYDIIELEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLD GVVRF
Sbjct: 8 TTTAHSYDIIELEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLDAQKGVVRF 67

Query: 58 TKIESVVDSVKNTLNVEVDIANETKDRRIAVGEGSLSVGDFSHSFSFEGSVVNMYYRSD 117
TKIESVVDSVKNTLNVEVDIANETKDRRIAVGEGSLSVGDFSHSFSFEG VVNMYYRSD
Sbjct: 68 TKIESVVDSVKNTLNVEVDIANETKDRRIAVGEGSLSVGDFSHSFSFEGQVVNMYYRSD 127

Query: 118 AVRRNIPNPIYMQGRQFHDILMKVPLDNNDLVDTWEGFQOSI---GANFGDWIREFWFIG 174
AVRRNIPNPIYMQGRQFHDILMKVPLDNNDLVDTWEGFQOSI GANFGDWIREFWFIG
Sbjct: 128 AVRRNIPNPIYMQGRQFHDILMKVPLDNNDLVDTWEGFQOSISGGGANFGDWIREFWFIG 187

Query: 175 PAFAAINEGGQRISPIVNSSNVEGG---PVGVTRWKFSSHAGSGVVDSISRWTELFVPEQ 231
PAFAAINEGGQRISPIVNSSNVEGG PVGVTRWKFSSHAGSGVVDSISRWTELFVPEQ
Sbjct: 188 PAFAAINEGGQRISPIVNSSNVEGGKGPVGVTRWKFSSHAGSGVVDSISRWTELFVPEQ 247

Query: 232 LNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGLRRLNHPLIPLVHHGMVGFNFDF 291
LNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGLRRLNHPLIPLVHHGMVGFNFDF
Sbjct: 248 LNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGLRRLNHPLIPLVHHGMVGFNFDF 307

Query: 292 TVDTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQFEVLYA 349
TVDTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQFEVLYA
Sbjct: 308 TVDTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQFEVLYA 365

>SW:BCPA_CHLTE Q46393 Bacteriochlorophyll a protein OS=Chlorobium tepidum GN=fmoA
PE=1 SV=3
Length = 366

Score = 582 bits (1499), Expect = e-165
Identities = 276/356 (77%), Positives = 310/356 (87%), Gaps = 8/356 (2%)

Query: 2 TTAHSYDIIELEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLD---GVVRF 58
TTAHSYDIIELEGGSSSWG+VK RAKVN P A PLLP DC++++ KPLD G VR +
Sbjct: 10 TTAHSYDIIELEGGSSSWGKVKARAKVNAPPASPLLPADCDVKNVVKPLDPKGFVRISA 69

Query: 59 KIESVDSVKNTLNVEVDIANETKDRRIAVGEGSLSVGDFSHSFSFEGSVVNMYYYRSDA 118
ES+VDS KN L +E DIANETK+RRI+VGEG +SVGDFSH+FSFEGSVVN++YYRSDA
Sbjct: 70 VFESIVDSTKNKLTIEADIANETKERRISVGEVMVSVGDFSHSFSFEGSVVNLFYYRSDA 129

Query: 119 VRRNIPNPIYMQGRQFHDILMKVPLDNNLDVDTWEGFQOSIGAN--FGDWIREFWFIGPA 176
VRRN+PNPIYMQGRQFHDILMKVPLDNNDL+DTWEG ++IG+ F DWIR+FWFIGPA
Sbjct: 130 VRRNVPNPIYMQGRQFHDILMKVPLDNNLDLDTWEGTVKAIGSTGAFNDWIRDFWFIGPA 189

Query: 177 FAAINEGGQRISPIVNSSNVEGGP---VGVTRWKFSHAGSGVVDISISRWTELPVEQLN 233
F A+NEGGQRIS I VN N E GP VGV+RW+FSH GSG+VDSISRW ELFP ++LN
Sbjct: 190 FTALNEGGQRISRIEVNGLNTEGSPKGPVGVSRWRF SHGGSGMVVDISISRWAELEFPSPDKLN 249

Query: 234 KPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGMVGKFNDFTV 293
+PA +E GFRSDSQGIEVKVDG PGVS DAGGGLRRILNHPLIPLVHHGMVGKFN+F V
Sbjct: 250 RPAQVEAGFRSDSQGIEVKVDGEFPGVSVVDAGGGLRRILNHPLIPLVHHGMVGKFNDFNV 309

Query: 294 DTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQFEVLYA 349
D QLK+VLPKGYKIRYAAPQ+RSQNLEEYRWSGGAYARWVEHVCKGG QQFE+LYA
Sbjct: 310 DAQLKVLPKGYKIRYAAPQYRSQNLEEYRWSGGAYARWVEHVCKGGVQFEVLYA 365

>SW:BCPA_CHLLT_Q46135 Bacteriochlorophyll a protein (Fragment) OS=Chlorobium
limicola f.sp. thiosulfatophilum GN=fmoA PE=4 SV=1
Length = 354

Score = 575 bits (1483), Expect = e-163
Identities = 274/353 (77%), Positives = 305/353 (86%), Gaps = 8/353 (2%)

Query: 5 HSDYEIILEGGSSSWGQVKGRAKVNVPAAIPLLPDCNIRIDAKPLD---GVVRFTTKIE 61
H DYEI+LEGGSSSWG+VK RAKVNVP A PLLP DCN++++ KPLD G VR + E
Sbjct: 1 HRDYEIVLEGGSSSWGKVKARAKVNVPASP LLPADCNVKNVPLDPAKGFVRISAVFE 60

Query: 62 SVVDSVKNTLNVEVDIANETKDRRIAVGEGSLSVGDFSHSFSFEGSVVNMYYYRSDAVRR 121
S+VDS KN L +E DIANETK+RRI+VGEG +SVG FSHSFSFEGSVVNMYYYRSDAVRR
Sbjct: 61 SIVDSTKNKLTIEADIANETKERRISVGEVMVSVGGFSHSFSFEGSVVNMFYYRSDAVRR 120

Query: 122 NIPNPIYMQGRQFHDILMKVPLDNNLDVDTWEGFQOSIGAN--FGDWIREFWFIGPAAFA 179
N+PNPIY QGRQFHDILMKVPLDNNDL+DTWEG ++IG+ F DWIR+FWFIGPAF A
Sbjct: 121 NVPNPIYRQGRQFHDILMKVPLDNNLDLDTWEGTVRAIGSTGTFNWIRDFWFIGPAFTA 180

Query: 180 INEGGQRISPIVNSSNVEGGP---VGVTRWKFSHAGSGVVDISISRWTELPVEQLNKPA 236
+NEGGQRIS I VN N E GP VGV+RW+FSH GSG+VDSISRW ELFP ++LN+PA
Sbjct: 181 LNEGGQRISRIEVNGLNTEGSPKGPVGVSRWRF SHGGSGMVVDISISRWAELEFPFDKLN RPA 240

Query: 237 SIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGMVGKFNDFTVDTQ 296
+E GFRSDSQGIEVKVDG PGVS DAGGGLRRILNHPLIPLVHHGMVGKFN+F VD Q
Sbjct: 241 QVEAGFRSDSQGIEVKVDGEFPGVSVVDAGGGLRRILNHPLIPLVHHGMVGKFNDFNVDAQ 300

Query: 297 LKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQFEVLYA 349
LK+VLPKGYK+RYAAPQ+RSQNLEEYRWSGGAYARWVEHVCKGG QQFEVLYA
Sbjct: 301 LKVLPKGYKIRYAAPQYRSQNLEEYRWSGGAYARWVEHVCKGGVQFEVLYA 353

>SW:B3GL2_MOUSE_Q8BG28 UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 2
OS=Mus musculus GN=B3galnt2 PE=1 SV=1
Length = 504

Score = 33.9 bits (76), Expect = 1.8
Identities = 23/59 (38%), Positives = 29/59 (49%)

Query: 228 PVEQLNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGMVG 286
PVEQ P S EG +SQ + V NL V+ + GGG+ R+L L H M G
Sbjct: 211 PVEQFILPESFEGTIVWESQDLHGLVSRNLHRVTVNDGGGVLRLVLAAGEGALPHEFMEG 269

>SW:B3GL2_HUMAN_Q8NCR0 UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 2
OS=Homo sapiens GN=B3GALNT2 PE=1 SV=1
Length = 500

Score = 32.7 bits (73), Expect = 3.9
Identities = 21/59 (35%), Positives = 29/59 (49%)

Query: 228 PVEQLNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGMVG 286
PVEQ P S EG +SQ + V NL V+ + GGG+ R++ L H + G
Sbjct: 209 PVEQFILPESFEGTIVWESQDLHGLVSRNLHKVTVNDGGGVLRVITAGEGALPHEFLEG 267

>SW:PDDC1_DANRE_Q3B7H1 Parkinson disease 7 domain-containing protein 1 OS=Danio
rerio GN=pddc1 PE=2 SV=1
Length = 213

Score = 32.3 bits (72), Expect = 5.1
Identities = 20/63 (31%), Positives = 29/63 (46%), Gaps = 5/63 (7%)

Query: 214 SGVVDSISRWTELPVEQLNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDA--GGGLRRI 271
+GV DS RW + F ++ PA +E D + + + PG D G L RI
Sbjct: 49 TGVDDSTGRWFQEFSEIKPYANPAKLES---IDGARYQALLIPDCPGAMNDLAHSGSLARI 105

Query: 272 LNH 274
L+H
Sbjct: 106 LSH 108

Database: swissprot
Posted date: Mar 23, 2010 12:24 AM
Number of letters in database: 181,677,051
Number of sequences in database: 516,081

Lambda	K	H
0.318	0.138	0.422

Gapped

Lambda	K	H
0.267	0.0410	0.140

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 516081
Number of Hits to DB: 137,149,996
Number of extensions: 6258484
Number of successful extensions: 12339
Number of sequences better than 10.0: 6
Number of HSP's gapped: 12617
Number of HSP's successfully gapped: 6
Length of query: 349
Length of database: 181,677,051
Length adjustment: 118
Effective length of query: 231
Effective length of database: 120,779,493
Effective search space: 27900062883
Effective search space used: 27900062883
Neighboring words threshold: 11
Window for multiple hits: 40
X1: 16 (7.3 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.7 bits)
S2: 70 (31.6 bits)

/ebi/extserv/bin/fasta-35.4.10/fasta35_t -l /ebi/services/idata/v2738/fastacfg/fasta3db
-Q -H -p -b 50 -d 50 -s BL50 -E 10.0 -f -10 -g -2 -z 1 @:1- +swissprot+ 2
FASTA searches a protein or DNA sequence data bank
version 35.04 Jan. 24, 2010
Please cite:
W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query: @
1>>>Sequence - 349 aa
Library: SWISS-PROT Protein Sequence Database 181677051 residues in 516081 sequences

181677051 residues in 516081 sequences
Statistics: Expectation_n fit: rho(ln(x))= 5.8184+/-0.000193; mu= 5.6551+/- 0.011
mean_var=75.5972+/-15.827, 0's: 63 Z-trim: 63 B-trim: 0 in 0/65
Lambda= 0.147510
statistics sampled from 60000 to 516040 sequences
Algorithm: FASTA (3.5 Sept 2006) [optimized]
Parameters: BL50 matrix (15:-5) ktup: 2
join: 43, opt: 31, open/ext: -10/-2, width: 16
Scan time: 22.190

The best scores are: opt bits E(516081)
SW:BCPA_PROAE P11741 Bacteriochlorophyll a protein (366) 1989 432.6 4e-120
SW:BCPA_CHLTE Q46393 Bacteriochlorophyll a protein (366) 1948 423.8 1.7e-117
SW:BCPA_CHLLT Q46135 Bacteriochlorophyll a protein (354) 1926 419.2 4.2e-116
SW:BGAL_XANMN P48982 Beta-galactosidase OS=Xanthom (598) 120 34.9 3.3
SW:UREF_BURCM Q0BHN7 Urease accessory protein ureF (226) 113 33.2 4
SW:B3GL2_MOUSE Q8BG28 UDP-GalNAc:beta-1,3-N-acetyl (504) 116 34.0 5.2
SW:HUTU_ALISL B6ERX7 Urocanate hydratase OS=Aliivi (564) 115 33.8 6.6
SW:B3GL2_XENTR Q5M900 UDP-GalNAc:beta-1,3-N-acetyl (488) 114 33.6 6.7
SW:B3GL2_XENLA Q6NRQ1 UDP-GalNAc:beta-1,3-N-acetyl (486) 112 33.2 9

>>SW:BCPA_PROAE P11741 Bacteriochlorophyll a protein OS= (366 aa)
initn: 2055 init1: 1041 opt: 1989 Z-score: 2291.6 bits: 432.6 E(): 4e-120
Smith-Waterman score: 2311; 97.2% identity (97.5% similar) in 358 aa overlap (1-349:8-365)

```

                10         20         30         40         50
Sequen  TTTAHS DY EI ILEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLD--
          :
SW:BCP  ALFGTKDTTTAHS DY EI ILEGGSSSWGQVKGRAKVNVPAAIPLLPTDCNIRIDAKPLDAQ
          10         20         30         40         50         60

          60         70         80         90        100        110
Sequen  -GVVRF TTKIESVVD SVKNTLNVEVDIANETKDRRIAVGEGSLSVGF SHSFSFEGSVVN
          :
SW:BCP  KGVVRF TTKIESVVD SVKNTLNVEVDIANETKDRRIAVGEGSLSVGF SHSFSFEGQVVN
          70         80         90        100        110        120

          120        130        140        150        160
Sequen  MYYRSDAVRRNIPNPIYMQGRQFHDILMKVPLDNDLVDTWEGFQQSI---GANFGDWI
          :
SW:BCP  MYYRSDAVRRNIPNPIYMQGRQFHDILMKVPLDNDLVDTWEGFQQSISSGGGANFGDWI
          130        140        150        160        170        180

          170        180        190        200        210        220
Sequen  REFWFIFGPAFAAINEGGQRISPIVNSSNVEGG---PVGVTRWKFSHAGSGVVD SISRWT
          :
SW:BCP  REFWFIFGPAFAAINEGGQRISPIVNSSNVEGGKGPVGVTRWKFSHAGSGVVD SISRWT
          190        200        210        220        230        240

          230        240        250        260        270        280
Sequen  ELFPVEQLNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGM
          :
SW:BCP  ELFPVEQLNKPASIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRILNHPLIPLVHHGM
          250        260        270        280        290        300

          290        300        310        320        330        340
Sequen  VGKFNDF TVDTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSGGAYARWVEHVCKGGTGQF
          :

```



```
180      190      200      210      220      230
Sequen AAINEGGQRISPIVNVSSNVEGGPVGVTWRKFSHAGSGV-VDSISRWTELPVEQLNKPA
      .:: :.::: :.::: :. : :.::: :
SW:B3G VFQHDSAAGFHRNITVKLFQTEHEEALFSARFSPASSGVQVNGI--WYK--PVEQFILPE
      150      160      170      180      190      200
```

```
240      250      260      270      280      290
Sequen SIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRI--LNHPLIPL-VHHGMVGKFNDFTV
      .::: :. : :.::: :. :.::: :. :.::: :. :. :. :. :. :. :
SW:B3G GFEGTVVWESHDPGLLSGNVHRVIVNDGGGIFRITTVKEGLLPYEFTEGVEGIAGGFTY
      210      220      230      240      250      260
```

```
300      310      320      330      340
Sequen DTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSSGAYARWVEHVCKGGTGQFEVLYA
SW:B3G TIHEGEALLNTLETRPERIQIHLAALEKEDALLQEESTTFODIVFVHVVDTYRNVPSKLL
      270      280      290      300      310      320
```

>>SW:B3GL2_XENLA Q6NRQ1 UDP-GalNAc:beta-1,3-N-acetylgala (486 aa)
initn: 83 init1: 83 opt: 112 Z-score: 130.9 bits: 33.2 E(): 9
Smith-Waterman score: 112; 34.8% identity (60.7% similar) in 89 aa overlap (208-292:178-262)

```
180      190      200      210      220      230
Sequen AAINEGGQRISPIVNVSSNVEGGPVGVTWRKFSHAGSGV-VDSISRWTELPVEQLNKPA
      .:: :.::: :.::: :. : :.::: :
SW:B3G VFQHDSAAGFQRNITVKLFQTEHEEALFSARFSPASSGVQVNGI--WYK--PVEQFILPE
      150      160      170      180      190      200
```

```
240      250      260      270      280      290
Sequen SIEGGFRSDSQGIEVKVDGNLPGVSRDAGGGLRRI--LNHPLIPL-VHHGMVGKFNDFTV
      .::: :. : :.::: :. :.::: :. :.::: :. :. :. :. :. :. :
SW:B3G GFEGTVVWESHDPGLLSGNVHHVIVNDGGGIFRLTTVKEGLLPYEFTEGVEGIAGGFTY
      210      220      230      240      250      260
```

```
300      310      320      330      340
Sequen DTQLKIVLPKGYKIRYAAPQFRSQNLEEYRWSSGAYARWVEHVCKGGTGQFEVLYA
SW:B3G TIHEGETLLNTLETRPERIQNHLAALEKEDALLQEESTTFODIVFVNVVDTYRNVPSKLL
      270      280      290      300      310      320
```

349 residues in 1 query sequences
181677051 residues in 516081 library sequences
Tcomplib [35.04] (16 proc)
start: Mon Apr 19 19:07:22 2010 done: Mon Apr 19 19:07:24 2010
Total Scan time: 22.190 Total Display time: 0.040

Function used was FASTA [version 35.04 Jan. 24, 2010]