# Supplementary Material A

# for "Improving taxonomy-based protein fold recognition by using global and local features"

Jian-Yi Yang and Xin Chen

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore, 637371

## The revised DD dataset (RDD)

The training and testing sequences of the original DD dataset can be downloaded from http://ranger.uta.edu/~chqding/protein/. We found that quite a number of sequences in this dataset were already updated in the latest release of SCOP database (release 1.75, June 2009, http://scop.mrc-lmb.cam.ac.uk/scop/). Therefore, we revised the DD dataset based on the latest release of SCOP database. Consequently, 11 training domain sequences and 76 testing sequences were updated, and one domain (`1BUCA1`) was deleted. We use the following few examples to explain why we made these updates.

The sequence of the domain "1CGT1" in the original DD dataset is given as follows:

```
>1GOF1
ASAPIGSAISRNNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
TTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRP
ARYVRLVAITEANGQPWTSIAEINVFQASSYTA
```

This domain was classified to the fold "`IMMUNOGLOBULIN-LIKE BELTA-SANDWICH`" according to the original DD dataset.

However, we found a different sequence for the above domain in the PDB database, as can be seen from the following **Figures SA1** and **SA2**. the sequence for the above domain was not correct. This extracted from the PDB database (http://www.pdb.org/pdb/home/home.do).

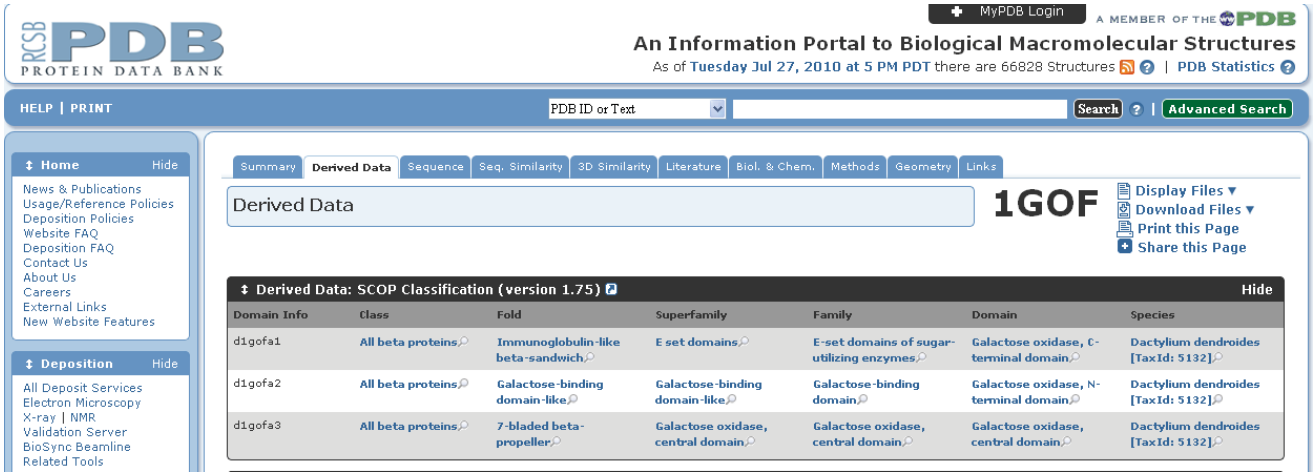



**Figure SA1.** The domain information and SCOP classification of the protein 1GOF. This webpage was retrieved from http://www.pdb.org/pdb/explore/derivedData.do?structureId=1GOF#SCOP (as of 10 September 2010), which shows that the protein 1GOF has three domains *d1gofa1*, *d1gofa2*, and *d1gofa3*. These three domains are classified into three different folds. The first domain *d1gofa1* belongs to the fold "`IMMUNOGLOBULIN-LIKE BELTA-SANDWICH`", whose sequence can also be found in the PDB database, as shown in **Figure SA2**.

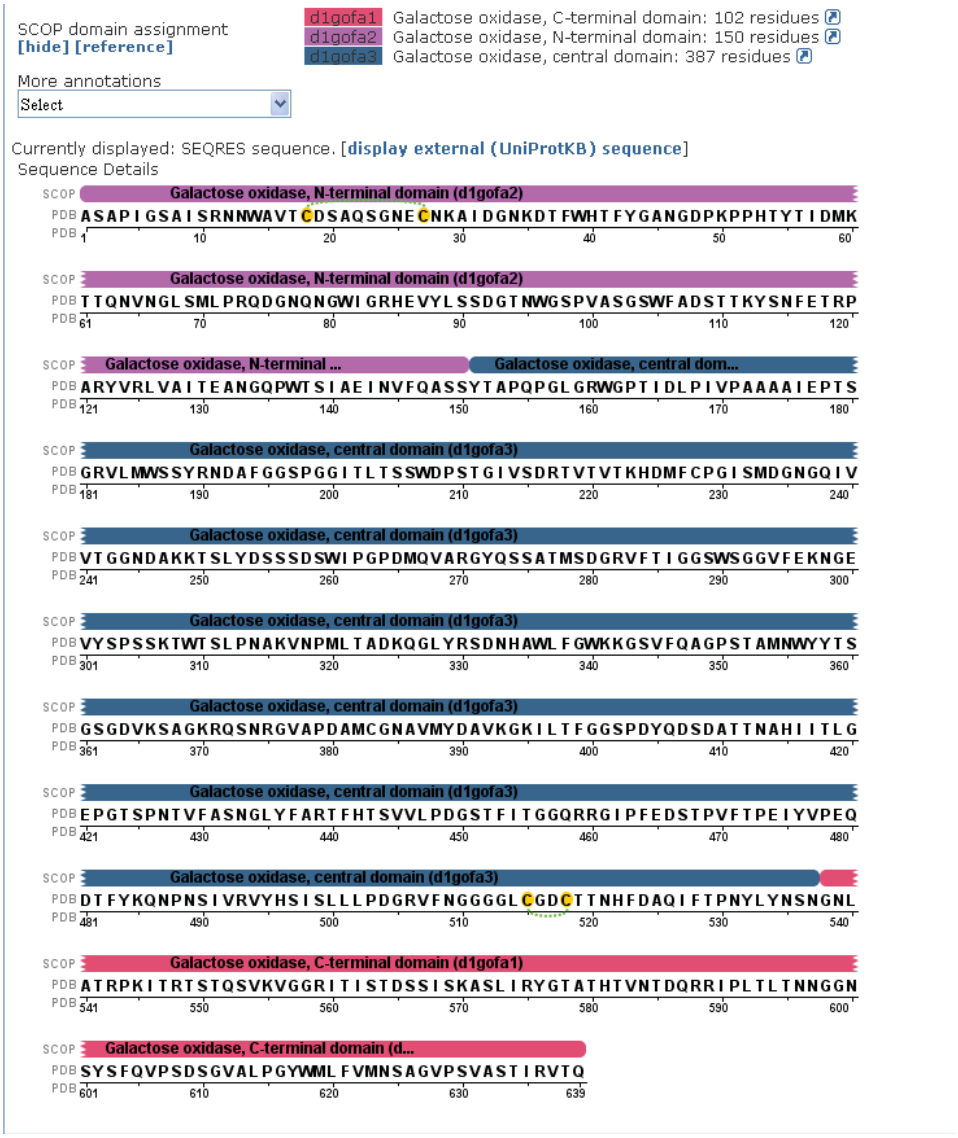



**Figure SA2.** The domain sequences of the protein 1GOF. This page was retrieved from http://www.pdb.org/pdb/explore/remediatedSequence.do?structureId=1GOF (as of 10 September 2010), where we can see that the three domains are marked by different colors. The sequence of the domain *d1gofa1* appears at the bottom and is marked by red color; that is,

GNLATRPKITRTSTQSVKVGGRITISTDSSISKASLIRYGTATHTVNTDQRRIPLTLTNNGGNSYSFQVPSDSGVALPGYWMLF
VMNSAGVPSVASTIRVTQ.

It is easy to see that this sequence is different from the one given in the original DD training dataset; that is,

ASAPIGSAISRNNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMKTTQNVNGLSMLPRQDGNQNGWIGR
HEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRPARYVRLVAITEANGQPWTSIAEINVFQASSYTA

In fact, this is the sequence of the domain *d1gofa2*, which is classified into the fold "Galactose-binding domain-like", instead of "IMMUNOGLOBULIN-LIKE BELTA-SANDWICH" as the original DD dataset indicates. In such a case, we will replace the domain sequence with the correct one in our revised DD dataset. Note that the three SCOP domains *d1gofa1*, *d1gofa2*, and *d1gofa3* have their sequences appeared **NOT** in the same order as their domain names suggest. So, we suspect that the discontinuous labeling of domain sequences had caused the sequence inconsistencies between the original DD dataset and the PDB database.

In addition, the domain 1BUCA1 was classified as the "4-HELICAL UP-AND-DOWN BUNDLE" in the original DD dataset. However, according to the SCOP classification, there is no domain of this protein classified into this fold, as can be seen from **Figure SA3**. Therefore, this protein is removed from the DD dataset.

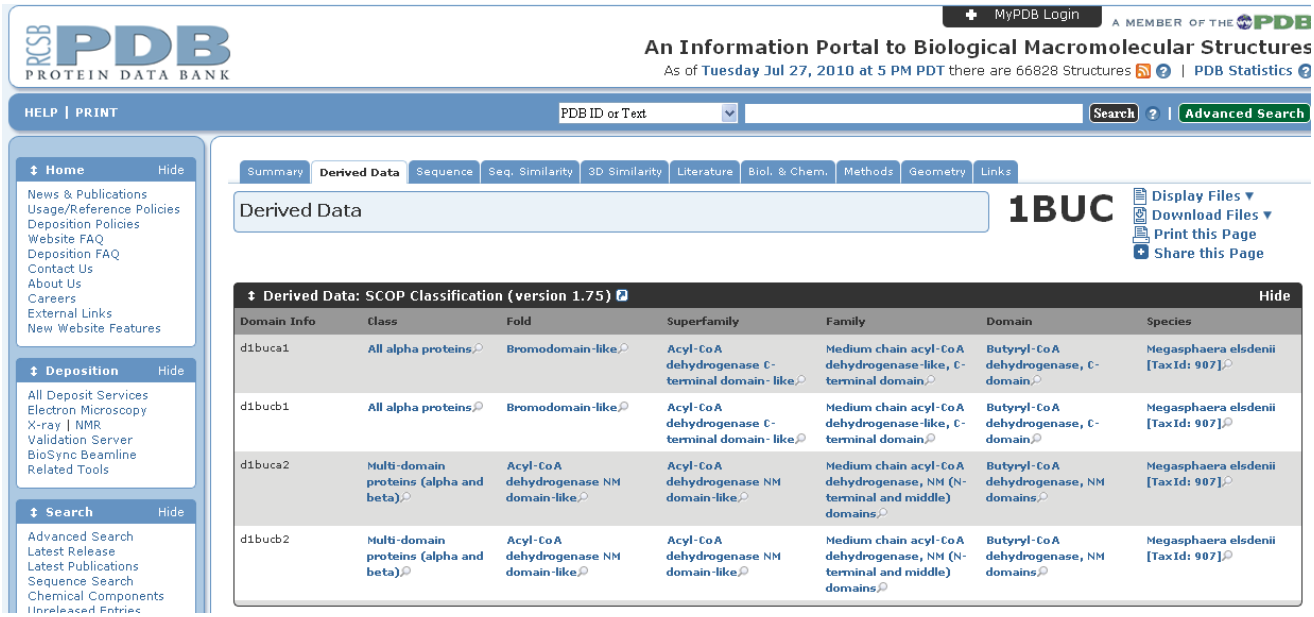



**Figure SA3.** The domain information and SCOP classification of the protein 1BUC. We can see that there are four domains for this protein, but none of them belong to the fold "4-HELICAL UP-AND-DOWN BUNDLE".

In summary, there are 88 domain sequences in the original DD dataset updated, which account for 88/(311+383)=12.7% proportion. The detailed comparisons are presented blow.

## Appendix:

## Comparison of sequences in SCOP release 1.50, 1.75 and the original DD dataset

For each domain below, the *first* sequence given is taken from the release 1.50 of SCOP and the *second* sequence from the release 1.75. The sequence differences are highlighted in yellow color. The *third* sequence (in capital letter) is taken from the original DD dataset. Note that, except for two domain sequences of 1SCUB2 and 3RUBL1, the first and the second sequence are identical. However, the third sequence is completely different from the first and second one.

### Domains used in the original DD dataset for training

>1GOF1
>d1gof_1 2.1.1.5.1 (538-639) Galactose oxidase, C-terminal domain {Dactylium dendroides}
gnlatrpkitrtstqsvkvggritistdssiskaslirygtathtvntdqrripltltnn

ggnsysfqvpsdsgvalpgywmlfvmnsagvpsvastirvtq

>d1gofa1 b.1.18.2 (A:538-639) Galactose oxidase, C-terminal domain {Dactylium dendroides [TaxId: 5132]}
gnlatrpkitrtstqsvkvggritistdssiskaslirygtathtvntdqrripltltnn
ggnsysfqvpsdsgvalpgywmlfvmnsagvpsvastirvtq

ASAPIGSAISRNNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
TTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRP
ARYVRLVAITEANGQPWTSIAEINVFQASSYTA

>1CGT1
>d1cgt_1 2.1.1.5.4 (495-579) Cyclodextrin glycosyltransferase, domain E {Bacillus circulans, different strains}
ettptighvgpvmgkpgnvvtidgrgfgstkgtvyfgttavtgaaitswedtqikvtips
vaagnyavkvaasgvnsnaynnfti

>d1cgta1 b.1.18.2 (A:495-579) Cyclomaltodextrin glycanotransferase, domain D {Bacillus circulans, different strains [TaxId: 1397]}
ettptighvgpvmgkpgnvvtidgrgfgstkgtvyfgttavtgaaitswedtqikvtips
vaagnyavkvaasgvnsnaynnfti

DPDTAVTNKQSFSTDVIYQVFTDRFLDGNPSNNPTGAAYDATCSNLKLYCGGDWQGLINK
INDNYFSDLGVTALWISQPVENIFATINYSGVTNTAYHGYWARDFKKTNPYFGTMADFQN
LITTAHAKGIKIVIDFAPNHTSPAMETDTSFAENGRLYDNGTLVGGYTNDTNGYFHHNGG
SDFSSLENGIYKNLYDLADFNHNNATIDKYFKDAIKLWLDMGVDGIRVDAVKHMPLGWQK
SWMSSIYAHKPVFTFGEWFLGSAASDADNTDFANKSGMSLLDFRFNSAVRNVFRDNTSNM
YALDSMINSTATDYNQVNDQVTFIDNHDMDRFKTSAVNNRRLEQALAFTLTSRGVPAIYY
GTEQYLTGNGDPDNRAKMPSFSKSTTAFNVISKLAPLRKS

>1OXY1
>d1oxy_1 1.81.1.1.1 (1-109) Hemocyanin, N-terminal domain {Limulus polyphemus}
tlhdkqirichlfeqlssatvigdgdkhkhsdrlknvgklqpgaifscfhpdhleearhl
yevfweagdfndfieiakeartfvneglfafaaevavlhrddckglyvp

>d1oxya1 a.85.1.1 (A:1-109) Hemocyanin, N-terminal domain {Horseshoe crab (Limulus polyphemus) [TaxId: 6850]}
tlhdkqirichlfeqlssatvigdgdkhkhsdrlknvgklqpgaifscfhpdhleearhl
yevfweagdfndfieiakeartfvneglfafaaevavlhrddckglyvp

TLHDKQIRICHLFEQLSSATVIGDGDKHKHSDRLKNVGKLQPGAIFSCFHPDHLEEARHL
YEVFWEAGDFNDFIEIAKEARTFVNEGLFAFAAEVAVLHRDDCKGLYVP

>1CGT3
>d1cgt_4 3.1.7.1.4 (1-406) Cyclodextrin glycosyltransferase {Bacillus circulans}
dpdtavtnkqsfstdviyqvftdrfldgnpsnnptgaaydatcsnlklycggdwqglink
indnyfsdlgvtalwisqpvenifatinysgvtntayhgywardfkktnpyfgtmadfqn
littahakgikividfapnhtspametdtsfaengrlydngtlvggytndtngyfhhngg
sdfsslengiyknlydladfnhnnatidkyfkdaiklwldmgvdgirvdavkhmplgwqk

swmssiyahkpvftfgewflgsaasdadntdfanksgmslldfrfnsavrnvfrdntsnm
yaldsminstatdynqvndqvtfidnhdmdrfktsavnnrrleqalaftltsrgvpaiyy
gteqyltgngdpdnrakmpsfsksttafnvisklaplrksnpaiay

>d1cgta4 c.1.8.1 (A:1-406) Cyclodextrin glycosyltransferase {Bacillus circulans, different strains [TaxId: 1397]}
dpdtavtnkqsfstdviyqvftdrfldgnpsnnptgaaydatcsnlklycggdwqglink
indnyfsdlgvtalwisqpvenifatinysgvtntayhgywardfkktnpyfgtmadfqn
littahakgikividfapnhtspametdtsfaengrlydngtlvggytndtngyfhhngg
sdfsslengiyknlydladfnhhnnatidkyfkdaiklwldmgvdgirvdavkhmplgwqk
swmssiyahkpvftfgewflgsaasdadntdfanksgmslldfrfnsavrnvfrdntsnm
yaldsminstatdynqvndqvtfidnhdmdrfktsavnnrrleqalaftltsrgvpaiyy
gteqyltgngdpdnrakmpsfsksttafnvisklaplrksnpaiay

ETTPTIGHVGPVMGKPGNVVTIDGRGFGSTKGTVYFGTTAVTGAAITSWEDTQIKVTIPS
VAAGNYAVKVAASGVNSNAYNNFTIL

>4ENL1
>d4enl_1 3.1.10.1.1 (142-436) Enolase {Baker's yeast (Saccharomyces cerevisiae)}
spyvlpvpflnvlnggshaggalalqefmiaptgaktfaealrigsevyhnlksltkkry
gasagnvgdeggvapniqtaeealdlivdaikaaghdgkvkigldcasseffkdgkydld
fknpnsdkskwltgpqladlyhslmkrypivsiedpfaeddweawshffktagiqivadd
ltvtnpkriataiekkaadalllkvnqigtlsesikaaqdsfaagwgvmvshrsgetedt
fiadlvvglrtgqiktgaparserlaklnqllrieeelgdnavfagenfhhgdkl

>d4enla1 c.1.11.1 (A:142-436) Enolase {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}
spyvlpvpflnvlnggshaggalalqefmiaptgaktfaealrigsevyhnlksltkkry
gasagnvgdeggvapniqtaeealdlivdaikaaghdgkvkigldcasseffkdgkydld
fknpnsdkskwltgpqladlyhslmkrypivsiedpfaeddweawshffktagiqivadd
ltvtnpkriataiekkaadalllkvnqigtlsesikaaqdsfaagwgvmvshrsgetedt
fiadlvvglrtgqiktgaparserlaklnqllrieeelgdnavfagenfhhgdkl

AVSKVYARSVYDSRGNPTVEVELTTEKGVFRSIVPSGASTGVHEALEMRDGDKSKWMGKG
VLHAVKNVNDVIAPAFVKANIDVSDQKAVDDFLISLDGTANKSKLGANAILGVSLAASRA
AAAEKN

>2TMDA2
>d2tmda2 3.3.1.1.1 (490-645) Trimethylamine dehydrogenase, C-terminal domain {Escherichia coli}
rwntdgtnclthdpipgadaslpdqltpeqvmdgkkkigkrvvilnadtyfmapslaekl
ataghevtivsgvhlanymhftleypnmmrrlhelhveelgdhfcsriepgrmeiyniwg
dgskrtyrgpgvsprdantshrwiefdslvlvtgrh

>d2tmda2 c.3.1.1 (A:490-645) Trimethylamine dehydrogenase, C-terminal domain {Methylophilus methylotrophus, w3a1 [TaxId: 17]}
rwntdgtnclthdpipgadaslpdqltpeqvmdgkkkigkrvvilnadtyfmapslaekl
ataghevtivsgvhlanymhftleypnmmrrlhelhveelgdhfcsriepgrmeiyniwg
dgskrtyrgpgvsprdantshrwiefdslvlvtgrh

QTKNKDSVLIVGAGPSGSEAARVLMESGYTVHLTDTAEKIGGHLNQVAALPGLGEWSYHR
DYRETQITKLLKKNKESQLALGQKPMTADDVLQYGADKVIIATGARWRHSECTLWNELKA
RESEWAENDIKGIYLIGDAEAPRLIADATFTGHRVAREIEEANPQIAI

>1SCUB2
>d1scub1 3.16.3.1.2 (245-388) Succinyl-CoA synthetase, beta-chain, C-terminal domain {Escherichia coli}
aaqwelnyvaldgnigcmvngaglamgtmdivklhggepanfldvgggatkervteafki
ilsddkvkavlvnifggivrcdliadgiigavaevgvnvpvvvrlegnnaelgakklads
glniiaakgltdaaqqvvaavegk

>d1scub1 c.23.4.1 (B:239-388) Succinyl-CoA synthetase, beta-chain, C-terminal domain {Escherichia coli [TaxId: 562]}
dpreaqaaqwelnyvaldgnigcmvngaglamgtmdivklhggepanfldvgggatkerv
teafkiilsddkvkavlvnifggivrcdliadgiigavaevgvnvpvvvrlegnnaelga
kkladsglniiaakgltdaaqqvvaavegk

VGYACTTPREAEEAASKIGAGPWVVKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKR
LVTYQTDANGQPVNQILVEAATDI

>2NADA2
>d2nada2 3.16.11.1.1 (1-147,336-391) Formate dehydrogenase {Pseudomonas sp. 101}
akvlcvlyddpvdgypktyarddlpkidhypggqtlptpkaidftpgqllgsvsgelglr
kylesnghtlvvtsdkdgpdsvferelvdadvvisqpfwpayltperiakaknlklalta
gigsdhvdlqsaidrnvtvaevtycnsXttltaqaryaagtreilecffegrpirdeyli
vqggalagtgahsyskgnatggse

>d2nada2 c.23.12.1 (A:1-147,A:336-391) Formate dehydrogenase {Pseudomonas sp., strain 101 [TaxId: 306]}
akvlcvlyddpvdgypktyarddlpkidhypggqtlptpkaidftpgqllgsvsgelglr
kylesnghtlvvtsdkdgpdsvferelvdadvvisqpfwpayltperiakaknlklalta
gigsdhvdlqsaidrnvtvaevtycnsXttltaqaryaagtreilecffegrpirdeyli
vqggalagtgahsyskgnatggse

NSISVAEHVVMMILSLVRNYLPSHEWARKGGWNIADCVSHAYDLEAMHVGTVAAGRIGLA
VLRRLAPFDVHLHYTDRHRLPESVEKELNLTWHATREDMYPVCDVVTLNCPLHPETEHMI
NDETLKLFKRGAYIVNTARGKLCDRDAVARALESGRLAGYAGDVWFPQPAPKDHPWRTMP
YNGMTPHISG

>2NADA1
>d2nada1 3.2.1.4.1 (148-335) Formate dehydrogenase {Pseudomonas sp. 101}
isvaehvvmmilslvrnylpshewarkggwniadcvshaydleamhvgtvaagriglavl
rrlapfdvhlhytdrhrlpesvekelnltwhatredmypvcdvvtlncplhpetehmind
etlklfkrgayivntargklcdrdavaralesgrlagyagdvwfpqpapkdhpwrtmpyn
gmtphisg

>d2nada1 c.2.1.4 (A:148-335) Formate dehydrogenase {Pseudomonas sp., strain 101 [TaxId: 306]}
isvaehvvmmilslvrnylpshewarkggwniadcvshaydleamhvgtvaagriglavl

rrlapfdvhlhytdrhrlpesvekelnltwhatredmypvcdvvtlncplhpetehmind
etlklfkrgayivntargklcdrdavaralesgrlagyagdvwfpqpapkdhpwrtmpyn
gmtphisg

AKVLCVLYDDPVDGYPKTYARDDLPKIDHYPGGQTLPTPKAIDFTPGQLLGSVSGELGLR
KYLESNGHTLVVTSDKDGPDSVFERELVDADVVISQPFWPAYLTPERIAKAKNLKLALTA
GIGSDHVDLQSAIDRNVTVAEVTYCTTLTAQARYAAGTREILECFFEGRPIRDEYLIVQG
GALA

>1TADC2
>d1tadc2 3.31.1.7.12 (27-56,178-344) Transducin (alpha subunit) {Bovine (Bos taurus)}
artvkllllgagesgkstivkqmkiihqdgXtgiietqfsfkdlnfrmfdvggqrserkk
wihcfegvtciifiaalsaydmvlveddevnrmheslhlfnsicnhryfattsivlflnk
kdvfsekikkahlsicfpdyngpntyedagnyikvqflelnmrrdvkeiyshmtcatdtq
nvkfvfdavtdiiikenl

>d1tadc2 c.37.1.8 (C:27-56,C:178-344) Transducin (alpha subunit) {Cow (Bos taurus) [TaxId: 9913]}
artvkllllgagesgkstivkqmkiihqdgXtgiietqfsfkdlnfrmfdvggqrserkk
wihcfegvtciifiaalsaydmvlveddevnrmheslhlfnsicnhryfattsivlflnk
kdvfsekikkahlsicfpdyngpntyedagnyikvqflelnmrrdvkeiyshmtcatdtq
nvkfvfdavtdiiikenl

SLEECLEFIAIIYGNTLQSILAIVRAMTTLNIQYGDSARQDDARKLMHMADTIEEGTMPK
EMSDIIQRLWKDSGIQACFDRASEYQLNDSAGYYLSDLERLVTPGYVPTEQDVLRSRVK

>1TSSA2 (the PDB ID in release 1.75 is updated as 2TSSA2)
>d2tssa2 4.13.7.1.3 (94-194) Toxic shock syndrome toxin-1 (TSST-1) {Staphylococcus aureus}
lptpielplkvkvhgkdsplkywpkfdkkqlaistldfeirhqltqihglyrssdktggy
wkitmndgstyqsdlskkfeyntekppinideiktieaein

>d2tssa2 d.15.6.1 (A:94-194) Toxic shock syndrome toxin-1 (TSST-1) {Staphylococcus aureus [TaxId: 1280]}
lptpielplkvkvhgkdsplkywpkfdkkqlaistldfeirhqltqihglyrssdktggy
wkitmndgstyqsdlskkfeyntekppinideiktieaein

DTFTNSEVLDNSLGSMRIKNTDGSISLIIFPSPYYSPAFTKGEKVDLNTKRTKKSQHTSE
GTYIHFQISGV

**Domains used in the original DD dataset for testing**

>1OCTC1
>d1octc1 1.4.1.1.5 (102-161) Oct-1 POU Homeodomain {Human (Homo sapiens)}
rkkrtsietnirvaleksflenqkptseeitmiadqlnmekevirvwfcnrrqkekrinp

>d1octc1 a.4.1.1 (C:102-161) Oct-1 POU Homeodomain {Human (Homo sapiens) [TaxId: 9606]}
rkkrtsietnirvaleksflenqkptseeitmiadqlnmekevirvwfcnrrqkekrinp

DLEELEQFAKTFKQRRIKLGFTQGDVGLAMGKLYGNDFSQTTISRFEALNLSFKNMCKLK

PLLEKWLNDAE

>1SFE_1
>d1sfe_1 1.4.2.1.1 (93-176) Ada DNA repair protein {Escherichia coli}
gtafqqqvwqalrtipcgetvsyqqlanaigkpkavravasacaanklaivipchrvvrg
dgslsgyrwgvsrkaqllrreaen

>d1sfea1 a.4.2.1 (A:93-176) Ada DNA repair protein {Escherichia coli [TaxId: 562]}
gtafqqqvwqalrtipcgetvsyqqlanaigkpkavravasacaanklaivipchrvvrg
dgslsgyrwgvsrkaqllrreaen

LAVRYALADCELGRCLVAESERGICAILLGDDDATLISELQQMFPAADNAPADLMFQQHV
REVIASLNQRDTPLTLPLDIR

>1CGPA1
>d1cgpa1 1.4.4.4.1 (138-205) Catabolite gene activator protein (CAP), C-terminal domain {Escherichia coli}
dvtgriaqtllnlakqpdamthpdgmqikitrqeigqivgcsretvgrilkmledqnlis
ahgktivv

>d1cgpa1 a.4.5.4 (A:138-205) Catabolite gene activator protein (CAP), C-terminal domain {Escherichia coli [TaxId: 562]}
dvtgriaqtllnlakqpdamthpdgmqikitrqeigqivgcsretvgrilkmledqnlis
ahgktivv

PTLEWFLSHCHIHKYPSKSTLIHQGEKAETLYYIVKGSVAVLIKDEEGKEMILSYLNQGD
FIGELGLFEEGQERSAWVRAKTACEVAEISYKKFRQLIQVNPDILMRLSAQMARRLQVTS
EKVGNLAFL

>1XGSA1
>d1xgsa1 1.4.4.21.1 (195-271) Methionine aminopeptidase, insert domain {Pyrococcus furiosus}
gqvievpptliymyvrdvpvrvaqarfllakikreygtlpfayrwlqndmpegqlklalk
tlekagaiygypvlkei

>d1xgsa1 a.4.5.25 (A:195-271) Methionine aminopeptidase, insert domain {Archaeon Pyrococcus furiosus [TaxId: 2261]}
gqvievpptliymyvrdvpvrvaqarfllakikreygtlpfayrwlqndmpegqlklalk
tlekagaiygypvlkei

MDTEKLMKAGEIAKKVREKAIKLARPGMLLLELAESIEKMIMELGGKPAFPVNLSINEIA
AHYTPYKGDTTVLKEGDYLKIDVGVHIDGFIADTAVTVRVGMEEDELMEAAKEALNAAIS
VARAGVEIKELGKAIENEIRKRGFKPIVNLSGHKIERYKLHAGISIPNIYRPHDNYVLKE
GDVFAIEPFATIGARNGIVAQFEHTIIVEKDSVIVTTE

>1CD1A1
>d1cd1a1 2.1.1.2.154 (186-279) CD1, beta2-microglobulin and alpha-3 domain {Mouse (Mus musculus)}
qekpvawlssvpssahghrqlvchvsgfypkpvwvmwmrgdqeqqgthrgdflpnadetw

ylqatldveageeaglacrvkhsslggqdiilyw

>d1cd1a1 b.1.1.2 (A:186-279) CD1, alpha-3 domain {Mouse (Mus musculus) [TaxId: 10090]}
qekpvawlssvpssahghrqlvchvsgfypkpvwvmwmrgdqeqqgthrgdflpnadetw
ylqatldveageeaglacrvkhsslggqdiilyw

NYTFRCLQMSSFANRSWSRTDSVVWLGDLQTHRWSNDSATISFTKPWSQGKLSNQQWEKL
QHMFQVYRVSFTRDIQELVKMMSPKEDYPIEIQLSAGCEMYPGNASESFLHVAFQGKYVV
RFWGTSWQTVPGAPSWLDLPIKVLNADQGTSATVQMLLNDTCPLFVRGLLEAGKSDLE

>1DLHA1
>d1dlha1 2.1.1.2.156 (82-182) Class II MHC, C-terminal domains of alpha and beta chains {Human HLA-dr1}
itnvppevtvltnspvelrepnvlicfidkftppvvnvtwlrngkpvttgvsetvflpre
dhlfrkfhylpflpstedvydcrvehwgldepllkhwefda

>d1dlha1 b.1.1.2 (A:82-182) Class II MHC alpha chain, C-terminal domain {Human (Homo sapiens), HLA-DR group [TaxId: 9606]}
itnvppevtvltnspvelrepnvlicfidkftppvvnvtwlrngkpvttgvsetvflpre
dhlfrkfhylpflpstedvydcrvehwgldepllkhwefda

EEHVIIQAEFYLNPDQSGEFMFDFDGDEIFHVDMAKKETVWRLEEFGRFASFEAQGALAN
IAVDKANLEIMTKRSNYTP

>1GOF_1
>d1gof_1 2.1.1.5.1 (538-639) Galactose oxidase, C-terminal domain {Dactylium dendroides}
gnlatrpkitrtstqsvkvggritistdssiskaslirygtathtvntdqrripltltnn
ggnsysfqvpsdsgvalpgywmlfvmnsagvpsvastirvtq
>d1gofa1 b.1.18.2 (A:538-639) Galactose oxidase, C-terminal domain {Dactylium dendroides [TaxId: 5132]}
gnlatrpkitrtstqsvkvggritistdssiskaslirygtathtvntdqrripltltnn
ggnsysfqvpsdsgvalpgywmlfvmnsagvpsvastirvtq

ASAPIGSAISRNNWAVTCDSAQSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYTIDMK
TTQNVNGLSMLPRQDGNQNGWIGRHEVYLSSDGTNWGSPVASGSWFADSTTKYSNFETRP
ARYVRLVAITEANGQPWTSIAEINVFQASSYTA

>1QBA_1
>d1qba_1 2.1.1.5.2 (781-885) Bacterial chitobiase, c-terminal domain {Serratia marcescens}
gethfvdtqalekdwlrfanilgqrelakldkggvayrlpvpgarvaggkleanialpgl
gieystdggkqwqrydakakpavsgevqvrsvspdgkrysraekv

>d1qbaa1 b.1.18.2 (A:781-885) Bacterial chitobiase (N-acetyl-beta-glucoseaminidase), C-terminal domain {Serratia marcescens [TaxId: 615]}
gethfvdtqalekdwlrfanilgqrelakldkggvayrlpvpgarvaggkleanialpgl
gieystdggkqwqrydakakpavsgevqvrsvspdgkrysraekv

DQQLVDQLSQLKLNVKMLDNRAGENGVDCAALGADWASCNRVLFTLSNDGQAIDGKDWVI
YFHSPRQTLRVDNDQFKIAHLTGDLYKLEPTAKFSGFPAGKAVEIPVVAEYWQLFRNDFL
PRWYATSGDAKPKMLANTDTEN

>1SVB_1
>d1svb_1 2.1.1.5.3 (303-395) Envelope glycoprotein, domain III (C-terminal) {Tick-borne encephalitis virus, TBE}
tytmcdktkftwkraptdsghdtvvmevtfsgtkpcripvravahgspdvnvamlitpnp
tiennggfiemqlppgdniiyvgelshqwfqk

>d1svba1 b.1.18.4 (A:303-395) Envelope glycoprotein {Tick-borne encephalitis virus [TaxId: 11084]}
tytmcdktkftwkraptdsghdtvvmevtfsgtkpcripvravahgspdvnvamlitpnp
tiennggfiemqlppgdniiyvgelshqwfqk

SRCTHLENRDFVTGTQGTTRVTLVLELGGCVTITAEGKPSMDVWLDAIYQENKIVYTVKV
EPHTGDYVAANETHSGRKTASFTISSEKTILTMGEYGDVSLLCRVASGPVAHIEGTKYHL
KSGHVTCEVGLEKLKM

>1CDG_1
>d1cdg_1 2.1.1.5.4 (496-581) Cyclodextrin glycosyltransferase, domain E {Bacillus circulans, different strains}
tatptighvgpmmakpgvtitidgrgfgsskgtvyfgttavsgaditswedtqikvkipa
vaggnynikvanaagtasnvydnfev

>d1cdga1 b.1.18.2 (A:496-581) Cyclomaltodextrin glycanotransferase, domain D {Bacillus circulans, different strains [TaxId: 1397]}
tatptighvgpmmakpgvtitidgrgfgsskgtvyfgttavsgaditswedtqikvkipa
vaggnynikvanaagtasnvydnfev

APDTSVSNKQNFSTDVIYQIFTDRFSDGNPANNPTGAAFDGTCTNLRLYCGGDWQGIINK
INDGYLTGMGVTAIWISQPVENIYSIINYSGVNNTAYHGYWARDFKKTNPAYGTIADFQN
LIAAAHAKNIKVIIDFAPNHTSPASSDQPSFAENGRLYDNGTLLGGYTNDTQNLFHHNGG
TDFSTTENGIYKNLYDLADLNHNNSTVDVYLKDAIKMWLDLGIDGIRMDAVKHMPFGWQK
SFMAAVNNYKPVFTFGEWFLGVNEVSPENHKFANESGMSLLDFRFAQKVRQVFRDNTDNM
YGLKAMLEGSAADYAQVDDQVTFIDNHDMERFHASNANRRKLEQALAFTLTSRGVPAIYY
GTEQYMSGGTDPDNRARIPSFSTSTTAYQVIQKLAPLRKC

>1LLA_2
>d1lla_3 2.1.1.5.10 (380-628) Hemocyanin, C-terminal domain {Horseshoe crab (Limulus polyphemus), hemolymph}
pydhdvlnfpdiqvqdvtlharvdnvvhtfmreqelelkhginpgnarsikaryyhldhe
pfsyavnvqnnsasdkhatvriflapkydelgneikadelrrtaieldkfktdlhpgknt
vvrhsldssvtlshqptfedllhgvglnehkseycscgwpshllvpkgnikgmeyhlfvm
ltdwdkdkvdgsesvacvdavsycgardhkypdkkpmgfpfdrpihtehisdfltnnmfi
kdikikfhe

>d1llaa3 b.1.18.3 (A:380-628) Arthropod hemocyanin, C-terminal domain {Horseshoe crab (Limulus polyphemus) [TaxId: 6850]}
pydhdvlnfpdiqvqdvtlharvdnvvhtfmreqelelkhginpgnarsikaryyhldhe
pfsyavnvqnnsasdkhatvriflapkydelgneikadelrrtaieldkfktdlhpgknt
vvrhsldssvtlshqptfedllhgvglnehkseycscgwpshllvpkgnikgmeyhlfvm
ltdwdkdkvdgsesvacvdavsycgardhkypdkkpmgfpfdrpihtehisdfltnnmfi
kdikikfhe

PVQEIFPDKFIPSAAINEAFKKAHVRPEFDESPILVDVQDTGNILDPEYRLAYYREDVGI
NAHHWHWHLVYPSTWNPKYFGKKKDRKGELFYYMHQQMCARYDCERLSNGMHRMLPFNNF
DEPLAGYAPHLTHVASGKYYSPRPDGLKLRDLGDIEISEMVRMRERILDSIHLGYVISED
GSHKTLDELHGTDILGALVESSYESVNHEYYGNLHNWGHVTMARIHDPDGRFHEEPGVMS
DTSTSLRDPIFYNWHRFIDNIFHEYKNTLK

>1HC2_2 (the PDB ID in release 1.75 is updated as 1HC1)
>d1hc1_3 2.1.1.5.11 (399-653) Hemocyanin, C-terminal domain {Spiny lobster (Panulirus interruptus)}
ppythdnlefsgmvvngvaidgelitffdefqyslinavdsgeniedveinarvhrlnhn
eftykitmsnnndgerlatfriflcpiednngitltldearwfcieldkffqkvpsgpet
iersskdssvtvpdmpsfqslkeqadnavngghdldlsayerscgipdrmllpkskpegm
efnlyvavtdgdkdteghngghdyggthaqcgvhgeaypdnrplgyplerripdervidg
vsnikhvvvkivhhl

>d1hc1a3 b.1.18.3 (A:399-653) Arthropod hemocyanin, C-terminal domain {Spiny lobster (Panulirus interruptus) [TaxId: 6735]}
ppythdnlefsgmvvngvaidgelitffdefqyslinavdsgeniedveinarvhrlnhn
eftykitmsnnndgerlatfriflcpiednngitltldearwfcieldkffqkvpsgpet
iersskdssvtvpdmpsfqslkeqadnavngghdldlsayerscgipdrmllpkskpegm
efnlyvavtdgdkdteghngghdyggthaqcgvhgeaypdnrplgyplerripdervidg
vsnikhvvvkivhhl

PLYQITPHMFTNSEVIDKAYSAKMTQKPGTFNVSFTGTKKNREQRVAYFGEDIGMNIHHV
TWHMDFPFWWEDSYGYHLDRKGELFFWVHHQLTARFDFERLSNWLDPVDELHWDRIIREG
FAPLTSYKYGGEFPVRPDNIHFEDVDGVAHVHDLEITESRIHEAIDHGYITDSDGHTIDI
RQPKGIELLGDIIESSKYSSNVQYYGSLHNTAHVMLGRQGDPHGKFNLPPGVMEHFETAT
RDPSFFRLHKYMDNIFKKHTDSF

>1CLC_2
>d1clc_2 2.1.1.5.12 (35-134) CelD cellulase, N-terminal domain {Clostridium thermocellum}
ietkvsaakitenyqfdsrirlnsigfipnhskkatiaancstfyvvkedgtivytgtat
smfdndtketvyiadfssvneegtyylavpgvgksvnfki

>d1clca2 b.1.18.2 (A:35-134) CelD cellulase, N-terminal domain {Clostridium thermocellum [TaxId: 1515]}
ietkvsaakitenyqfdsrirlnsigfipnhskkatiaancstfyvvkedgtivytgtat
smfdndtketvyiadfssvneegtyylavpgvgksvnfki

NVYEDAFKTAMLGMYLLRCGTSVSATYNGIHYSHGPCHTNDAYLDYINGQHTKKDSTKGW
HDAGDYNKYVVNAGITVGSMFLAWEHFKDQLEPVALEIPEKNNSIPDFLDELKYEIDWIL

TMQYPDGSGRVAHKVSTRNFGGFIMPENEHDERFFVPWSSAATADFVAMTAMAARIFRPY
DPQYAEKCINAAKVSYEFLKNNPANVFANQSGFSTGEYATVSDADDRLWAAAEMWETLGD
EEYLRDFENRAAQFSKKIEADFDWDNVANLGMFTYLLSERPGKNPALVQSIKDSLLSTAD
SIVRTSQNHGYGRTLGTTYYWGCNGTVVRQTMILQVANKISPNNDYVNAALDAISHVFGR
NYYNRSYVTGLGINPPMNPHDRRSGADGIWEPWPGYLVGGGWPGPKDWVDIQDSYQTNEI
AINWNAALIYALAGFVNYN

>1BGLA1
>d1bgla1 2.1.4.1.1 (220-333) beta-Galactosidase, domains 2 and 4 {Escherichia coli}
tqisdfhvatrfnddfsravleaevqmcgelrdylrvtvslwqgetqvasgtapfggeii
derggyadrvtlrlnvenpklwsaeipnlyravvelhtadgtlieaeacdvgfr

>d1bgla1 b.1.4.1 (A:220-333) beta-Galactosidase, domains 2 and 4 {Escherichia coli [TaxId: 562]}
tqisdfhvatrfnddfsravleaevqmcgelrdylrvtvslwqgetqvasgtapfggeii
derggyadrvtlrlnvenpklwsaeipnlyravvelhtadgtlieaeacdvgfr

ITDSLAVVLQRRDWENPGVTQLNRLAAHPPFASWRNSEEARTDRPSQQLRSLNGEWRFAW
FPAPEAVPESWLECDLPEADTVVVPSNWQMHGYDAPIYTNVTYPITVNPPFVPTENPTGC
YSLTFNVDESWLQEGQTRIIFDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRAGEN
RLAVMVLRWSDGSYLEDQDMWRMSGIFRDVSLLHKP

>1BGLA2
>d1bgla2 2.1.4.1.1 (626-730) beta-Galactosidase, domains 2 and 4 {Escherichia coli}
ffqfrlsgqtievtseylfrhsdnellhwmvaldgkplasgevpldvapqgkqlielpel
pqpesagqlwltvrvvqpnatawseaghisawqqwrlaenlsvtl

>d1bgla2 b.1.4.1 (A:626-730) beta-Galactosidase, domains 2 and 4 {Escherichia coli [TaxId: 562]}
ffqfrlsgqtievtseylfrhsdnellhwmvaldgkplasgevpldvapqgkqlielpel
pqpesagqlwltvrvvqpnatawseaghisawqqwrlaenlsvtl

TTQISDFHVATRFNDDFSRAVLEAEVQMCGELRDYLRVTVSLWQGETQVASGTAPFGGEI
IDERGGYADRVTLRLNVENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFR

>1BHGA1
>d1bhga1 2.1.4.1.2 (226-328) beta-Glucuronidase {Human (Homo sapiens)}
tyidditvttsveqdsglvnyqisvkgsnlfklevrlldaenkvvangtgtqgqlkvpgv
slwwpylmherpaylyslevqltaqtslgpvsdfytlpvgirt

>d1bhga1 b.1.4.1 (A:226-328) beta-Glucuronidase {Human (Homo sapiens) [TaxId: 9606]}
tyidditvttsveqdsglvnyqisvkgsnlfklevrlldaenkvvangtgtqgqlkvpgv
slwwpylmherpaylyslevqltaqtslgpvsdfytlpvgirt

GLQGGMLYPQESPSRECKELDGLWSFRADFSDNRRRGFEEQWYRRPLWESGPTVDMPVPS
SFNDISQDWRLRHFVGWVWYEREVILPERWTQDLRTRVVLRIGSAHSYAIVWVNGVDTLE
HEGGYLPFEADISNLVQVGPLPSRLRITIAINNTLTPTTLPPGTIQYLTDTSKYPKGYFV
QNTYFDFFNYAGLQRSVLLYTTP

>1GGTA2

>d1ggta2 2.1.5.1.1 (516-627) Coagulation factor XIII, two C-terminal domains {Human (Homo sapiens), blood}
snvdmdfevenavlgkdfklsitfrnnshnrytitaylsanitfytgvpkaefkketfdv
tleplsfkkeavliqageymgqlleqaslhffvtarinetrdvlakqkstvl

>d1ggta2 b.1.5.1 (A:516-627) Transglutaminase, two C-terminal domains {Human (Homo sapiens), blood isozyme [TaxId: 9606]}
snvdmdfevenavlgkdfklsitfrnnshnrytitaylsanitfytgvpkaefkketfdv
tleplsfkkeavliqageymgqlleqaslhffvtarinetrdvlakqkstvl

VYLDNEKEREEYVLNDIGVIFYGEVNDIKTRSWSYGQFEDGILDTCLYVMDRAQMDLSGR
GNPIKVSRVGSAMVNAKDDEGVLVGSWDNIYAYGVPPSAWTGSVDILLEYRSSENPVRYG
QCWVFAGVFNTFLRCLGIPARIVTNYFSAHDNDANLQMDIFLEEDGNVNSKLTKDSVWNY
HCWNEAWMTRPDLPVGFGGWQAVDSTPQENSDGMYRCGPASVQAIKHGHVCFQFDAPFVF
AEVNSDLIYITAKKDGTHVVENVDATHIGKLIVTKQIGGDGMMDITDTYKFQEGQEEERL
ALETALMYGAKKPLNTEGVMKSR

>1GGTA3
>d1ggta3 2.1.5.1.1 (628-729) Coagulation factor XIII, two C-terminal domains {Human (Homo sapiens), blood}
tipeiiikvrgtqvvgsdmtvtveftnplketlrnvwvhldgpgvtrpmkkmfreirpns
tvqweevcrpwvsghrkliasmssdslrhvygeldvqiqrrp

>d1ggta3 b.1.5.1 (A:628-729) Transglutaminase, two C-terminal domains {Human (Homo sapiens), blood isozyme [TaxId: 9606]}
tipeiiikvrgtqvvgsdmtvtveftnplketlrnvwvhldgpgvtrpmkkmfreirpns
tvqweevcrpwvsghrkliasmssdslrhvygeldvqiqrrp

SNVDMDFEVENAVLGKDFKLSITFRNNSHNRYTITAYLSANITFYTGVPKAEFKKETFDV
TLEPLSFKKEAVLIQAGEYMGQLLEQASLHFFVTARINETRDVLAKQKSTVLTIP

>4KBPA1
>d4kbpa1 2.1.11.1.1 (9-120) Purple acid phosphatase, N-terminal domain {Kidney bean (Phaseolus vulgaris)}
rdmpldsdvfrvppgynapqqvhitqgdlvgramiiswvtmdepgssavrywsekngrkr
iakgkmstyrffnyssgfihhttirklkyntkyyyevglrnttrrfsfitpp

>d4kbpa1 b.1.12.1 (A:9-120) Purple acid phosphatase, N-terminal domain {Kidney bean (Phaseolus vulgaris) [TaxId: 3885]}
rdmpldsdvfrvppgynapqqvhitqgdlvgramiiswvtmdepgssavrywsekngrkr
iakgkmstyrffnyssgfihhttirklkyntkyyyevglrnttrrfsfitpp

APQQVHITQGDLVGRAMIISWVTMDEPGSSAVRYWSEKNGRKRIAKGKMSTYRFFNYSSG
FIHHTTIRKLKYNTKYYYEVGLRNTTRRFSFITPPQT

>1OCCB1
>d1occb1 2.5.1.2.2 (91-227) Cytochrome c oxidase {Bovine (Bos taurus)}
nnpsltvktmghqwywsyeytdyedlsfdsymiptselkpgelrllevdnrvvlpmemti
rmlvssedvlhswavpslglktdaipgrlnqttlmssrpglyygqcseicgsnhsfmpiv

lelvplkyfekwsasml

>d1occb1 b.6.1.2 (B:91-227) Cytochrome c oxidase {Cow (Bos taurus) [TaxId: 9913]}
nnpsltvktmghqwywsyeytdyedlsfdsymiptselkpgelrllevdnrvvlpmemti
rmlvssedvlhswavpslglktdaipgrlnqttlmssrpglyygqcseicgsnhsfmpiv
lelvplkyfekwsasml

MAYPMQLGFQDATSPIMEELLHFHDHTLMIVFLISSLVLYIISLMLTTKLTHTSTMDAQE
VETIWTILPAIILILIALPSLRILYMMDEIN

>1KIT_2
>d1kit_2 2.28.1.8.1 (347-543) Vibrio cholerae sialidase, N-terminal and insertion domains {Vibrio cholerae}
dvtdqvkersfqiagwggselyrrntslnsqqdwqsnakirivdgaanqiqvadgsrkyv
vtlsidesgglvanlngvsapiilqsehakvhsfhdyelqysalnhtttlfvdgqqittw
agevsqenniqfgnadaqidgrlhvqkivltqqghnlvefdafylaqqtpevekdleklg
wtkiktgntmslygnas

>d1kita2 b.29.1.8 (A:347-543) Vibrio cholerae sialidase, N-terminal and insertion domains {Vibrio cholerae [TaxId: 666]}
dvtdqvkersfqiagwggselyrrntslnsqqdwqsnakirivdgaanqiqvadgsrkyv
vtlsidesgglvanlngvsapiilqsehakvhsfhdyelqysalnhtttlfvdgqqittw
agevsqenniqfgnadaqidgrlhvqkivltqqghnlvefdafylaqqtpevekdleklg
wtkiktgntmslygnas

QGDVIFRGPDRIPSIVASSVTPGVVTAFAEKRVGGGDPGALSNTNDIITRTSRDGGITWD
TELNLTEQINVSDEFDFSDPRPIYDPSSNTVLVSYARWPTDAAQNGDRIKPWMPNGIFYS
VYDVASGNWQAPIDVTDQVKNASVNPGPGHGITLTRQQNISGSQNGRLIYPAIVLDRFFL
NVMSIYSDDGGSNWQTGSTLPIPFRWKSSSILETLEPSEADMVELQNGDLLLTARLDFNQ
IVNGVNYSPRQQFLSKDGGITWSLLEANNANVFSNISTGTVDASITRFEQSDGSHFLLFT
NPQGNPAGTNGRQNLGLWFSFDEGVTWKGPIQLVNGASAYSDIYQLDSENAIVIVETDNS
NMRILRMPITLLKQKLTLSQN

>1BIA_2
>d1bia_2 2.32.1.1.1 (271-317) Biotin repressor/biotin holoenzyme synthetase, C-terminal domain {Escherichia coli}
finrpvkliigdkeifgisrgidkqgalleqdgiikpwmggeislr

>d1biaa2 b.34.1.1 (A:271-317) Biotin repressor/biotin holoenzyme synthetase, C-terminal domain {Escherichia coli [TaxId: 562]}
finrpvkliigdkeifgisrgidkqgalleqdgiikpwmggeislr

QLLNAKQILGQLDGGSVAVLPVIDSTNQYLLDRIGELKSGDACIAEYQQAGSPFGANLYL
SMFWRLEQPAAAIGLSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAGILVELTGA
AQIVIGAGINMAMWITLQEAGINLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLD
N

>1PRTB1

>d1prtb1 2.38.2.1.6 (90-199) Pertussis toxin S2/S3 subunits, C-terminal domain {Bordetella pertussis}
ttrntgqpatdhyysnvtatrllsstnsrlcavfvrsgqpvigactspydgkywsmysrl
rkmlyliyvagisvrvhvskeeqyydyedatfetyaltgisicnpgsslc

>d1prtb1 b.40.2.1 (B:90-199) Pertussis toxin S2/S3 subunits, C-terminal domain {Bordetella pertussis [TaxId: 520]}
ttrntgqpatdhyysnvtatrllsstnsrlcavfvrsgqpvigactspydgkywsmysrl
rkmlyliyvagisvrvhvskeeqyydyedatfetyaltgisicnpgsslc

GIVIPPQEQITQHGSPYGRCANKTRALTVAELRGSGDLQEYLRHVTRGWSIFALYDGTYL
GGEYGGVIKDGTPGGAFDLKTTFCIM

>1ESFA1
>d1esfa1 2.38.2.2.1 (1-120) Staphylococcal enterotoxin A, SEA {Staphylococcus aureus}
sekseeinekdlrkkselqgtalgnlkqiyyynekaktenkeshdqflqhtilfkgfftd
hswyndllvdfdskdivdkykgkkvdlygayygyqcaggtpnktacmyggvtlhdnnrlt

>d1esfa1 b.40.2.2 (A:1-120) Staphylococcal enterotoxin A, SEA {Staphylococcus aureus [TaxId: 1280]}
sekseeinekdlrkkselqgtalgnlkqiyyynekaktenkeshdqflqhtilfkgfftd
hswyndllvdfdskdivdkykgkkvdlygayygyqcaggtpnktacmyggvtlhdnnrlt

YNEKAKTENKESHDQFLQHTILFKGFFTYNDLLVDFDSKDIVDKYKGKKVDLYGAYYGYQ
CAGGTPNKTACMYGGVTLH

>1SE4_1
>d1se4_1 2.38.2.2.4 (1-121) Staphylococcal enterotoxin B, SEB {Staphylococcus aureus}
esqpdpkpdelhksskftglmenmkvlyddnhvsainvksidqflyfdliysikdtklgn
ydnvrvefknkdladkykdkyvdvfganyyyqcyfskktndinshqtdkrktcmyggvte
h

>d1se4a1 b.40.2.2 (A:1-121) Staphylococcal enterotoxin B, SEB {Staphylococcus aureus [TaxId: 1280]}
esqpdpkpdelhksskftglmenmkvlyddnhvsainvksidqflyfdliysikdtklgn
ydnvrvefknkdladkykdkyvdvfganyyyqcyfskktndinshqtdkrktcmyggvte
h

HVSAINVKSIDQFLYFDLIYSIKDTKLGNYDNVRVEFKNKDLADKYKDKYVDVFGANYYY
QCYFSKKTNDINSHQTDKRKTCMYGGVTEH

>1CUK_3
>d1cuk_3 2.38.4.2.1 (1-64) DNA helicase RuvA subunit, N-terminal domain {Escherichia coli}
migrlrgiiiekqpplvlievggvgyevhmpmtcfyelpeagqeaivfthfvvredaqll
ygfn

>d1cuka3 b.40.4.2 (A:1-64) DNA helicase RuvA subunit, N-terminal domain {Escherichia coli [TaxId: 562]}

migrlrgiiiekqpplvlievggvgyevhmpmtcfyelpeagqeaivfthfvvredaqll
ygfn

TDDAEQEAVARLVALGYKPQEASRMVSKIARPDASSETLIREALRAAL

>1CKMA1
>d1ckma1 2.38.4.6.1 (239-327) RNA guanylyltransferase (mRNA capping enzyme) {Chlorella virus, PBCV-1}
thhtidfiimsedgtigifdpnlrknvpvgkldgyynkgsivecgfadgtwkyiqgrsdk
nqandrltyektllnieenitidelldlf

>d1ckma1 b.40.4.6 (A:239-327) RNA guanylyltransferase (mRNA capping enzyme) {Chlorella virus PBCV-1 [TaxId: 10506]}
thhtidfiimsedgtigifdpnlrknvpvgkldgyynkgsivecgfadgtwkyiqgrsdk
nqandrltyektllnieenitidelldlf

NITTERAVLTLNGLQIKLHKVVGESRDDIVAKMKDLAMDDHKFPRLPGPDGIRFMMFFTR
VFGFKVCTIIDRAMTVYLLPFKNIPRVLFQGSIFDGELCVDIVEKKFAFVLFDAVVVSGV
TVSQMDLASRFFAMKRSLKEFKNVPEDPAILRYKE

>1CDG_4
>d1cdg_4 3.1.7.1.4 (1-406) Cyclodextrin glycosyltransferase {Bacillus circulans}
apdtsvsnkqnfstdviyqiftdrfsdgnpannptgaafdgtctnlrlycggdwqgiink
indgyltgmgvtaiwisqpveniysiinysgvnntayhgywardfkktnpaygtiadfqn
liaaahaknikviidfapnhtspassdqpsfaengrlydngtllggytndtqnlfhhngg
tdfsttengiyknlydladlnhnnstvdvylkdaikmwldlgidgirmdavkhmpfgwqk
sfmaavnnykpvftfgewflgvnevspenhkfanesgmslldfrfaqkvrqvfrdntdnm
yglkamlegsaadyaqvddqvtfidnhdmerfhasnanrrkleqalaftltsrgvpaiyy
gteqymsggtdpdnraripsfststtayqviqklaplrkcnpaiay

>d1cdga4 c.1.8.1 (A:1-406) Cyclodextrin glycosyltransferase {Bacillus circulans, different strains [TaxId: 1397]}
apdtsvsnkqnfstdviyqiftdrfsdgnpannptgaafdgtctnlrlycggdwqgiink
indgyltgmgvtaiwisqpveniysiinysgvnntayhgywardfkktnpaygtiadfqn
liaaahaknikviidfapnhtspassdqpsfaengrlydngtllggytndtqnlfhhngg
tdfsttengiyknlydladlnhnnstvdvylkdaikmwldlgidgirmdavkhmpfgwqk
sfmaavnnykpvftfgewflgvnevspenhkfanesgmslldfrfaqkvrqvfrdntdnm
yglkamlegsaadyaqvddqvtfidnhdmerfhasnanrrkleqalaftltsrgvpaiyy
gteqymsggtdpdnraripsfststtayqviqklaplrkcnpaiay

SGDQVSVRFVVNNATTALGQNVYLTGSVSELGNWDPAKAIGPMYNQVVYQYPNWYYDVSV
PAGKTIEFKFLKKQGSTVTWEGGSNHTFTAPSSGTATINVNWQ

>1PPI_2
>d1ppi_2 3.1.7.1.8 (1-403) Animal alpha-amylase {Porcine (Sus scrofa)}
qyapqtqsgrtsivhlfewrwvdialecerylgpkgfggvqvsppnenvvvtnpsrpwwe
ryqpvsyklctrsgnenefrdmvtrcnnvgvriyvdavinhmcgsgaaagtgttcgsycn

pgsrefpavpysawdfndgkcktasggiesyndpyqvrdcqlvglldlalekdyvrsmia
dylnklidigvagfridaskhmwpgdikavldklhnlntnwfpagsrpfifqevidlgge
aiksseyfgngrvtefkygaklgtvvrkwsgekmsylknwgegwgfmpsdralvfvdnhd
nqrghgaggssiltfwdarlykvavgfmlahpygftrvmssyrwarnfvngedvndwigp
pnnngvikevtinadttcgndwvcehrwreirnmvwfrnvvdg

>d1ppia2 c.1.8.1 (A:1-403) Animal alpha-amylase {Pig (Sus scrofa) [TaxId: 9823]}
qyapqtqsgrtsivhlfewrwvdialecerylgpkgfggvqvsppnenvvvtnpsrpwwe
ryqpvsyklctrsgnenefrdmvtrcnnvgvriyvdavinhmcgsgaaagtgttcgsycn
pgsrefpavpysawdfndgkcktasggiesyndpyqvrdcqlvglldlalekdyvrsmia
dylnklidigvagfridaskhmwpgdikavldklhnlntnwfpagsrpfifqevidlgge
aiksseyfgngrvtefkygaklgtvvrkwsgekmsylknwgegwgfmpsdralvfvdnhd
nqrghgaggssiltfwdarlykvavgfmlahpygftrvmssyrwarnfvngedvndwigp
pnnngvikevtinadttcgndwvcehrwreirnmvwfrnvvdg

EPFANWWDNGSNQVAFGRGNRGFIVFNNDDWQLSSTLQTGLPGGTYCDVISGDKVGNSCT
GIKVYVSSDGTAQFSISNSAEDPFIAIHAESKL

>2AAA_2
>d2aaa_2 3.1.7.1.11 (1-381) Fungal alpha-amylases {Aspergillus niger, acid amylase}
lsaaswrtqsiyflltdrfgrtdnsttatcntgneiycggswqgiidhldyiegmgftai
wispiteqlpqdtadgeayhgywqqkiydvnsnfgtadnlkslsdalhargmylmvdvvp
dhmgyagngndvdysvfdpfdsssyfhpyclitdwdnltmvedcwegdtivslpdldtte
tavrtiwydwvadlvsnysvdglridsvlevqpdffpgynkasgvycvgeidngnpasdc
pyqkvldgvlnypiywqllyafesssgsisnlynmiksvasdcsdptllgnfienhdnpr
fakytsdysqaknvlsyiflsdgipivyageeqhyaggkvpynreatwlsgydtsaelyt
wiattnairklaiaadsayit

>d2aaaa2 c.1.8.1 (A:1-381) Fungal alpha-amylases {Aspergillus niger, acid amylase [TaxId: 5061]}
lsaaswrtqsiyflltdrfgrtdnsttatcntgneiycggswqgiidhldyiegmgftai
wispiteqlpqdtadgeayhgywqqkiydvnsnfgtadnlkslsdalhargmylmvdvvp
dhmgyagngndvdysvfdpfdsssyfhpyclitdwdnltmvedcwegdtivslpdldtte
tavrtiwydwvadlvsnysvdglridsvlevqpdffpgynkasgvycvgeidngnpasdc
pyqkvldgvlnypiywqllyafesssgsisnlynmiksvasdcsdptllgnfienhdnpr
fakytsdysqaknvlsyiflsdgipivyageeqhyaggkvpynreatwlsgydtsaelyt
wiattnairklaiaadsayit

ADSAYITYANDAFYTDSNTIAMAKGTSGSQVITVLSNKGSSGSSYTLTLSGSGYTSGTKL
IEAYTCTSVTVDSSGDIPVPMASGLPRVLLPASVVDSSSLCG

>1JDC_2
>d1jdc_2 3.1.7.1.15 (1-357) G4-amylase (1,4-alpha-D-glucan maltotetrahydrolase) {Pseudomonas stutzeri}
dqagkspnavryhggdeiilqgfhwnvvreapndwynilrqqaatiaadgfsaiwmpvpw
rdfsswsdgsksgggegyfwhdfnkngrygsdaqlrqaasalggagvkvlydvvpnhmnr
gypdkeinlpagqgfwrndcadpgnypndcddgdrfiggdadlntghpqvygmfrdeftn
lrsqygaggfrfdfvrgyapervnswmtdsadnsfcvgqlwkgpseypnwdwrntaswqq
iikdwsdrakcpvfdfalkermqngsiadwkhglngnpdprwrevavtfvdnhdtgyspg

qnggqhhwalqdglirqayayiltspgtpvvywdhmydwgygdfirqliqvrraagv

>d1jdca2 c.1.8.1 (A:1-357) G4-amylase (1,4-alpha-D-glucan maltotetrahydrolase) {Pseudomonas stutzeri [TaxId: 316]}
dqagkspnavryhggdeiilqgfhwnvvreapndwynilrqqaatiaadgfsaiwmpvpw
rdfsswsdgsksgggegyfwhdfnkngrygsdaqlrqaasalggagvkvlydvvpnhmnr
gypdkeinlpagqgfwrndcadpgnypndcddgdrfiggdadlntghpqvygmfrdeftn
lrsqygaggfrfdfvrgyapervnswmtdsadnsfcvgqlwkgpseypnwdwrntaswqq
iikdwsdrakcpvfdfalkermqngsiadwkhglngnpdprwrevavtfvdnhdtgyspg
qnggqhhwalqdglirqayayiltspgtpvvywdhmydwgygdfirqliqvrraagv

RADSAISFHSGYSGLVATVSGSQQTLVVALNSDLGNPGQVASGSFSEAVNASNGQVRVWR

>1AMY_2
>d1amy_2 3.1.7.1.16 (1-346) Plant alpha-amylase {Barley (Hordeum vulgare), seeds, AMY2 isozyme}
qvlfqgfnweswkhnggwynflmgkvddiaaagithvwlppasqsvaeqgympgrlydld
askygnkaqlksligalhgkgvkaiadivinhrtaehkdgrgiycifeggtpdarldwgp
hmicrddrpyadgtgnpdtgadfgaapdidhlnlrvqkelvewlnwlkadigfdgwrfdf
akgysadvakiyidrsepsfavaeiwtslayggdgkpnlnqdqhrqelvnwvdkvggkgp
attfdfttkgilnvavegelwrlrgtdgkapgmigwwpakavtfvdnhdtgstqhmwpfp
sdrvmqgyayilthpgtpcifydhffdwglkeeidrlvsvrtrhgi

>d1amya2 c.1.8.1 (A:1-346) Plant alpha-amylase {Barley (Hordeum vulgare), seeds, AMY2 isozyme [TaxId: 4513]}
qvlfqgfnweswkhnggwynflmgkvddiaaagithvwlppasqsvaeqgympgrlydld
askygnkaqlksligalhgkgvkaiadivinhrtaehkdgrgiycifeggtpdarldwgp
hmicrddrpyadgtgnpdtgadfgaapdidhlnlrvqkelvewlnwlkadigfdgwrfdf
akgysadvakiyidrsepsfavaeiwtslayggdgkpnlnqdqhrqelvnwvdkvggkgp
attfdfttkgilnvavegelwrlrgtdgkapgmigwwpakavtfvdnhdtgstqhmwpfp
sdrvmqgyayilthpgtpcifydhffdwglkeeidrlvsvrtrhgi

IHNESKLQIIEADADLYLAEIDGKVIVKLGPRYDVGNLIPGGFKVAAHGNDYAVWEKI

>1BGLA5
>d1bgla5 3.1.7.3.13 (334-625) beta-Galactosidase, domain 3 {Escherichia coli}
evriengllllngkpllirgvnrhehhplhgqvmdeqtmvqdillmkqnnfnavrcshyp
nhplwytlcdryglyvvdeaniethgmvpmnrltddprwlpamservtrmvqrdrnhpsv
iiwslgnesghganhdalyrwiksvdpsrpvqyegggadttatdiicpmyarvdedqpfp
avpkwsikkwlslpgetrplilceyahamgnslggfakywqafrqyprlqggfvwdwvdq
slikydengnpwsayggdfgdtpndrqfcmnglvfadrtphpalteakhqqq

>d1bgla5 c.1.8.3 (A:334-625) beta-Galactosidase, domain 3 {Escherichia coli [TaxId: 562]}
evriengllllngkpllirgvnrhehhplhgqvmdeqtmvqdillmkqnnfnavrcshyp
nhplwytlcdryglyvvdeaniethgmvpmnrltddprwlpamservtrmvqrdrnhpsv
iiwslgnesghganhdalyrwiksvdpsrpvqyegggadttatdiicpmyarvdedqpfp
avpkwsikkwlslpgetrplilceyahamgnslggfakywqafrqyprlqggfvwdwvdq
slikydengnpwsayggdfgdtpndrqfcmnglvfadrtphpalteakhqqq

PAASHAIPHLTTSEMDFCIELGNKRWQFNRQSGFLSQMWIGDKKQLLTPLRDQFTRAPLD
NDIGVSEATRIDPNAWVERWKAAGHYQAEAALLQCTADTLADAVLITTAHAWQHQGKTLF
ISRKTYRIDGSGQMAITVDVEVASDTPHPARIGLNCQLAQVAERVNWLGLGPQENYPDRL
TAACFDRWDLPLSDMYTPYVFPSENGLRCGTRELNYGPHQWRGDFQFNISRYSQQQLMET
SHRHLLHAEEGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVWCQK

>1EBHA1
>d1ebha1 3.1.10.1.1 (142-436) Enolase {Baker's yeast (Saccharomyces cerevisiae)}
spyvlpvpflnvlnggshaggalalqefmiaptgaktfaealrigsevyhnlksltkkry
gasagnvgdeggvapniqtaeealdlivdaikaaghdgkvkigldcasseffkdgkydld
fknpnsdkskwltgpqladlyhslmkrypivsiedpfaeddweawshffktagiqivadd
ltvtnpkriataiekkaadalllkvnqigtlsesikaaqdsfaagwgvmvshrsgetedt
fiadlvvglrtgqiktgaparserlaklnqllrieeelgdnavfagenfhhgdkl

>d1ebha1 c.1.11.1 (A:142-436) Enolase {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}
spyvlpvpflnvlnggshaggalalqefmiaptgaktfaealrigsevyhnlksltkkry
gasagnvgdeggvapniqtaeealdlivdaikaaghdgkvkigldcasseffkdgkydld
fknpnsdkskwltgpqladlyhslmkrypivsiedpfaeddweawshffktagiqivadd
ltvtnpkriataiekkaadalllkvnqigtlsesikaaqdsfaagwgvmvshrsgetedt
fiadlvvglrtgqiktgaparserlaklnqllrieeelgdnavfagenfhhgdkl

AVSKVYARSVYDSRGNPTVEVELTTEKGVFRSIVPSGASTGVHEALEMRDGDKSKWMGKG
VLHAVKNVNDVIAPAFVKANIDVKDQKAVDDFLISLDGTANKSKLGANAILGVSLAASRA
AAAEKN

>2MNR_1
>d2mnr_1 3.1.10.2.3 (133-359) Mandelate racemase {Pseudomonas putida}
pvqaydshsldgvklateravtaaelgfravktkigypaldqdlavvrsirqavgddfgi
mvdynqsldvpaaikrsqalqqegvtwieeptlqhdyeghqriqsklnvpvqmgenwlgp
eemfkalsigacrlampdamkiggvtgwirasalaqqfgipmsshlfqeisahllaatpt
ahwlerldlagsvieptltfeggnavipdlpgvgiiwrekeigkylv

>d2mnra1 c.1.11.2 (A:133-359) Mandelate racemase {Pseudomonas putida [TaxId: 303]}
pvqaydshsldgvklateravtaaelgfravktkigypaldqdlavvrsirqavgddfgi
mvdynqsldvpaaikrsqalqqegvtwieeptlqhdyeghqriqsklnvpvqmgenwlgp
eemfkalsigacrlampdamkiggvtgwirasalaqqfgipmsshlfqeisahllaatpt
ahwlerldlagsvieptltfeggnavipdlpgvgiiwrekeigkylv

EVLITGLRTRAVNVPLAYPVHTAVGTVGTAPLVLIDLATSAGVVGHSYLFAYTPVALKSL
KQLLDDMAAMIVNEPLAPVSLEAMLAKRFCLAGYTGLIRMAAAGIDMAAWDALGKVHEEK
EIGKYL

>2CHR_1
>d2chr_1 3.1.10.2.4 (127-370) Chlormuconate cycloisomerase {Alcaligenes eutrophus}
plrsaipiawtlasgdtkrdldsavemierrrhnrfkvklgfrspqddlihmealsnslg
skaylrvdvnqawdeqvasvyipelealgvelieqpvgrentqalrrlsdnnrvaimade
slstlasafdlardrsvdvfslklcnmggvsatqkiaavaeasgiasyggtmldstigts
valqlystvpslpfgceligpfvladtlshepleirdyelqvptgvghgmtldedkvrqy

arvs

>d2chra1 c.1.11.2 (A:127-370) Chlormuconate cycloisomerase {Alcaligenes eutrophus [TaxId: 106590]}
plrsaipiawtlasgdtkrdldsavemierrrhnrfkvklgfrspqddlihmealsnslg
skaylrvdvnqawdeqvasvyipelealgvelieqpvgrentqalrrlsdnnrvaimade
slstlasafdlardrsvdvfslklcnmggvsatqkiaavaeasgiasyggtmldstigts
valqlystvpslpfgceligpfvladtlshepleirdyelqvptgvghgmtldedkvrqy
arvs

MKIDAIEAVIVDVPTKRPIQMSITTVHQQSYVIVRVYSEGLVGVGEGGSVGGPVWSAECA
ETIKIIVERYLAPHLLGTDAFNVSGALQTMARAVTGNASAKAAVEMALLDLKARALGVSI
AELLGGP

>1PII_1
>d1pii_2 3.1.2.2.3 (255-452) Indole-3-glycerophosphate synthase, IPGS {Escherichia coli}
genkvcgltrgqdakaaydagaiygglifvatsprcvnveqaqevmaaaplqyvgvfrnh
diadvvdkakvlslaavqlhgneeqlyidtlrealpahvaiwkalsvgetlparefqhvd
kyvldngqggsgqrfdwsllngqslgnvllagglgadncveaaqtgcagldfnsavesqp
gikdarllasvfqtlray

>d1piia1 c.1.2.4 (A:255-452) N-(5'phosphoribosyl)antranilate isomerase, PRAI {Escherichia coli [TaxId: 562]}
genkvcgltrgqdakaaydagaiygglifvatsprcvnveqaqevmaaaplqyvgvfrnh
diadvvdkakvlslaavqlhgneeqlyidtlrealpahvaiwkalsvgetlparefqhvd
kyvldngqggsgqrfdwsllngqslgnvllagglgadncveaaqtgcagldfnsavesqp
gikdarllasvfqtlray

MQTVLAKIVADKAIWVEARKQQQPLASFQNEVQPSTRHFYDALQGARTAFILECKKASPS
KGVIRDDFDPARIAAIYKHYASAISVLTDEKYFQGSFNFLPIVSQIAPQPILCKDFIIDP
YQIYLARYYQADACLLMLSVLDDDQYRQLAAVAHSLEMGVLTEVSNEEEQERAIALGAKV
VGINNRDLRDLSIDLNRTRELAPKLGHNVTVISESGINTYAQVRELSHFANGFLIGSALM
AHDDLHAAVRRVLLGQTLRAY

>1PII_2
>d1pii_1 3.1.2.2.1 (1-254) N-(5'phosphoribosyl)antranilate isomerase, PRAI {Escherichia coli}
mqtvlakivadkaiwvearkqqqplasfqnevqpstrhfydalqgartafileckkasps
kgvirddfdpariaaiykhyasaisvltdekyfqgsfnflpivsqiapqpilckdfiidp
yqiylaryyqadacllmlsvldddqyrqlaavahslemgvltevsneeeqeraialgakv
vginnrdlrdlsidlnrtrelapklghnvtvisesgintyaqvrelshfangfligsalm
ahddlhaavrrvll

>d1piia2 c.1.2.4 (A:1-254) Indole-3-glycerophosphate synthase, IPGS {Escherichia coli [TaxId: 562]}
mqtvlakivadkaiwvearkqqqplasfqnevqpstrhfydalqgartafileckkasps
kgvirddfdpariaaiykhyasaisvltdekyfqgsfnflpivsqiapqpilckdfiidp
yqiylaryyqadacllmlsvldddqyrqlaavahslemgvltevsneeeqeraialgakv
vginnrdlrdlsidlnrtrelapklghnvtvisesgintyaqvrelshfangfligsalm

ahddlhaavrrvll

ENKVCGLTRGQDAKAAYDAGAIYGGLIFVATSPRCVNVEQAQEVMAAAPLQYVGVFRNHD
IADVVDKAKVLSLAAVQLHGNEEQLYIDTLREALPAHVAIWKALSVGETLPAREFQHVDK
YVLDNGQGGSGQRFDWSLLNGQSLGNVLLAGGLGADNCVEAAQTGCAGLDFNSAVESQPG
IKDARLLASVF

>1DIK_1
>d1dik_1 3.1.11.2.1 (510-874) Pyruvate phosphate dikinase, C-terminal domain {Escherichia coli}
ietqeasvsgsferimvwadkfrtlkvrtnadtpedtlnavklgaegiglcrtehmffea
drimkirkmilsdsveareealnelipfqkgdfkamykalegrpmtvryldpplhefvph
teeeqaelaknmgltlaevkakvdelhefnpmmghrgcrlavtypeiakmqtravmeaai
evkeetgidivpeimiplvgekkelkfvkdvvvevaeqvkkekgsdmqyhigtmieipra
altadaiaeeaeffsfgtndltqmtfgfsrddagkfldsyykakiyesdpfarldqtgvg
qlvemavkkgrqtrpglkcgicgehggdpssvefchkvglnyvscspfrvpiarlaaaqa
alnnk

>d1dika1 c.1.12.2 (A:510-874) Pyruvate phosphate dikinase, C-terminal domain {Clostridium symbiosum [TaxId: 1512]}
ietqeasvsgsferimvwadkfrtlkvrtnadtpedtlnavklgaegiglcrtehmffea
drimkirkmilsdsveareealnelipfqkgdfkamykalegrpmtvryldpplhefvph
teeeqaelaknmgltlaevkakvdelhefnpmmghrgcrlavtypeiakmqtravmeaai
evkeetgidivpeimiplvgekkelkfvkdvvvevaeqvkkekgsdmqyhigtmieipra
altadaiaeeaeffsfgtndltqmtfgfsrddagkfldsyykakiyesdpfarldqtgvg
qlvemavkkgrqtrpglkcgicgehggdpssvefchkvglnyvscspfrvpiarlaaaqa
alnnk

AKWVYKFEEGNASMRNLLGGKGCNLAEMTILGMPIPQGFTVTTEACTEYYNSGKQITQEI
QDQIFEAITWLEELNGKKFGDTEDPLLVSVRSAARASMPGMMDTILNLEPKDQLMGAVKA
VFRSWDNPRAIVYRRMNDIPGDWGTAVNVQTMV

>3RUBL1
>d3rubl1 3.1.13.1.1 (148-468) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Tobacco (Nicotiana tabacum), variant turkish samsun}
fqgpphgiqverdklnkygrpllgctikpklglsaknygravyeclrggldftkddenvn
sqpfmrwrdrflfcaealykaqaetgeikghylnatagtceemikravfarelgvpivmh
dyltggftantslahycrdnglllhihramhavidrqknhgihfrvlakalrmsggdhih
sgtvvgklegerditlgfvdllrddfveqdrsrgiyftqdwvslpgvlpvasggihvwhm
palteifgddsvlqfggmtlghpwgnapgavanrvaleacvkarnegrdlaqegneiire
ackwspelaaacevwkeivfn

>d3rubl1 c.1.14.1 (L:148-467) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Tobacco (Nicotiana tabacum), variant turkish samsun [TaxId: 4097]}
fqgpphgiqverdklnkygrpllgctikpklglsaknygravyeclrggldftkddenvn
sqpfmrwrdrflfcaealykaqaetgeikghylnatagtceemikravfarelgvpivmh
dyltggftantslahycrdnglllhihramhavidrqknhgihfrvlakalrmsggdhih
sgtvvgklegerditlgfvdllrddfveqdrsrgiyftqdwvslpgvlpvasggihvwhm

palteifgddsvlqfggmtlghpwgnapgavanrvaleacvkarnegrdlaqegneiire
ackwspelaaacevwkeivf

LTYYTPEYQTKDTDILAAFRVTPQPGVPPEEAGAAVAAESSTVWTDGLTSLDRYKGRCYR
IERVVGEKDQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYVK
TF

>5RUBA1
>d5ruba1 3.1.13.1.6 (138-457) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Rhodospirillum rubrum}
gpsvnisalwkvlgrpevdgglvvgtiikpklglrpkpfaeachafwlggdfikndepqg
nqpfaplrdtialvadamrraqdetgeaklfsanitaddpfeiiargeyvletfgenash
vallvdgyvagaaaittarrrfpdnflhyhraghgavtspqskrgytafvhckmarlqga
sgihtgtmgfgkmegessdraiaymltqdeaqgpfyrqswggmkactpiisggmnalrmp
gffenlgnanviltagggafghidgpvagarslrqawqawrdgvpvldyarehkelaraf
esfpgdadqiypgwrkalgv

>d5ruba1 c.1.14.1 (A:138-457) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Rhodospirillum rubrum [TaxId: 1085]}
gpsvnisalwkvlgrpevdgglvvgtiikpklglrpkpfaeachafwlggdfikndepqg
nqpfaplrdtialvadamrraqdetgeaklfsanitaddpfeiiargeyvletfgenash
vallvdgyvagaaaittarrrfpdnflhyhraghgavtspqskrgytafvhckmarlqga
sgihtgtmgfgkmegessdraiaymltqdeaqgpfyrqswggmkactpiisggmnalrmp
gffenlgnanviltagggafghidgpvagarslrqawqawrdgvpvldyarehkelaraf
esfpgdadqiypgwrkalgv

SRYVNLALKEEDLIAGGEHVLCAYIMKPKAGYGYVATAAHFAAESSTGTRGVDALVYEVD
EARELTKIAYPVALFDRNITDGKAMIASFLTLTMGNNQGMGDVEYAKMHDFYVPEAYRAL
F

>1QAPA1
>d1qapa1 3.1.16.1.1 (130-296) Quinolinic acid phosphoribosyltransferase, C-terminal domain {Salmonella typhimurium}
vasevrryvgllagtqtqlldtrktlpglrtalkyavlcggganhrlgltdaflikenhi
iasgsvrqavekafwlhpdvpvevevenldelddalkagadiimldnfntdqmreavkrv
ngqarlevsgnvtaetlrefaetgvdfisvgaltkhvraldlsmrfc

>d1qapa1 c.1.17.1 (A:130-296) Quinolinic acid phosphoribosyltransferase (Nicotinate-nucleotide pyrophosphorylase, NadC), C-terminal domain {Salmonella typhimurium [TaxId: 90371]}
vasevrryvgllagtqtqlldtrktlpglrtalkyavlcggganhrlgltdaflikenhi
iasgsvrqavekafwlhpdvpvevevenldelddalkagadiimldnfntdqmreavkrv
ngqarlevsgnvtaetlrefaetgvdfisvgaltkhvraldlsmrfc

IPAAVAQALREDLGGEVDAGNDITAQLLPADTQAHATVITREDGVFCGKRWVEEVFIQLA
GDDVRLTWHVDDGDAIHANQTVFELQGPARVLLTGERTALNFVQTLSGVASVRALDLSMR
FC

>1DJXA3

>d1djxa3 3.1.17.1.1 (299-625) Phospholipase C isozyme D1 (PLC-D1) {Rat (Rattus norvegicus)}
dqplshylvssshntylledqltgpssteayiralckgcrcleldcwdgpnqepiiyhgy
tftskilfcdvlrairdyafkaspypvilslenhcsleqqrvmarhlrailgpilldqpl
dgvttslpspeqlkgkillkgkklggllpaggengseatdvsdeveaaemedeavrsqvq
hkpkedklklvpelsdmiiycksvhfggfsspgtsgqafyemasfsesralrllqesgng
fvrhnvsclsriypagwrtdssnyspvemwnggcqivalnfqtpgpemdvylgcfqdngg
cgyvlkpaflrdpnttfnsraltqgpw

>d1djxa3 c.1.18.1 (A:299-625) Phospholipase C isozyme D1 (PLC-D1) {Rat (Rattus norvegicus) [TaxId: 10116]}
dqplshylvssshntylledqltgpssteayiralckgcrcleldcwdgpnqepiiyhgy
tftskilfcdvlrairdyafkaspypvilslenhcsleqqrvmarhlrailgpilldqpl
dgvttslpspeqlkgkillkgkklggllpaggengseatdvsdeveaaemedeavrsqvq
hkpkedklklvpelsdmiiycksvhfggfsspgtsgqafyemasfsesralrllqesgng
fvrhnvsclsriypagwrtdssnyspvemwnggcqivalnfqtpgpemdvylgcfqdngg
cgyvlkpaflrdpnttfnsraltqgpw

WRPERLRVRIISGQQLPKVNKNKNSIVDPKVIVEIHGVGRDTGSRQTAVITNNGFNPRWD
MEFEFEVTVPDLALVRFMVEDYDSSSKNDFIGQSTIPWNSLKQGYRHVHLLSKNGDQHPS
ATLFVKISIQD

>1SFTA2
>d1sfta2 3.1.5.1.1 (12-244) Alanine racemase {Bacillus stearothermophilus}
vdldaiydnvenlrrllpddthimavvkanayghgdvqvartaleagasrlavafldeal
alrekgieapilvlgasrpadaalaaqqrialtvfrsdwleeasalysgpfpihfhlkmd
tgmgrlgvkdeeetkrivalierhphfvleglythfatadevntdyfsyqytrflhmlew
lpsrpplvhcansaaslrfpdrtfnmvrfgiamyglapspgikpllpyplkea

>d1sfta2 c.1.6.1 (A:12-244) Alanine racemase {Bacillus stearothermophilus [TaxId: 1422]}
vdldaiydnvenlrrllpddthimavvkanayghgdvqvartaleagasrlavafldeal
alrekgieapilvlgasrpadaalaaqqrialtvfrsdwleeasalysgpfpihfhlkmd
tgmgrlgvkdeeetkrivalierhphfvleglythfatadevntdyfsyqytrflhmlew
lpsrpplvhcansaaslrfpdrtfnmvrfgiamyglapspgikpllpyplkea

NDFHRDTWAEVDAFSLHSRLVHVKKLQPGEKVSYGATYTAQTEEWIGTIPIGYADGWLRR
LQHFHVLVDGQKAPIVGRICMDQCMIRLPGPLPVGTKVTLIGRQGDEVISIDDVARHLET
INYEVPCTISYRVPRIFFRHKRIMEVRNAIGA

>1GND_1
>d1gnd_1 3.3.1.3.1 (1-291,389-430) Guanine nucleotide dissosiation inhibitor, GDI {Bovine (Bos taurus)}
mdeeydvivlgtgltecilsgimsvngkkvlhmdrnpyyggesssitpleelykrfqlle
gppetmgrgrdwnvdlipkflmangqlvkmllytevtryldfkvvegsfvykggkiykvp
stetealasnlmgmfekrrfrkflvfvanfdendpktfegvdpqntsmrdvyrkfdlgqd
vidftghalalyrtddyldqpcletinriklyseslarygkspylyplyglgelpqgfar
lsaiyggtymlnkpvddiimengkvvgvksegevarckqlicdpsyvpdrvXpiddgses
qvfcscsydatthfettcndikdiykrmagsafd

>d1gnda1 c.3.1.3 (A:1-291,A:389-430) Guanine nucleotide dissociation inhibitor, GDI {Cow (Bos taurus) [TaxId: 9913]}
mdeeydvivlgtgltecilsgimsvngkkvlhmdrnpyyggesssitpleelykrfqlle
gppetmgrgrdwnvdlipkflmangqlvkmllytevtryldfkvvegsfvykggkiykvp
stetealasnlmgmfekrrfrkflvfvanfdendpktfegvdpqntsmrdvyrkfdlgqd
vidftghalalyrtddyldqpcletinriklyseslarygkspylyplyglgelpqgfar
lsaiyggtymlnkpvddiimengkvvgvksegevarckqlicdpsyvpdrvXpiddgses
qvfcscsydatthfettcndikdiykrmagsafd

VPSTETEALASNLMGMFEKRRFRKFLVFVANFDENDPKTFEGVDPQNTSMRDVYRKFDLG
QDVIDFTGHALALYRTDDYLDQPCLETINRIKLYSESLARYGKSP

>1RNL_2
>d1rnl_2 3.16.2.1.4 (5-142) Nitrate/nitrite response regulator (NARL), receiver domain {Escherichia coli}
epatilliddhpmlrtgvkqlismapditvvgeasngeqgielaesldpdlilldlnmpg
mngletldklrekslsgrivvfsvsnheedvvtalkrgadgyllkdmepedllkalhqaa
agemvlsealtpvlaasl

>d1rnla2 c.23.1.1 (A:5-142) Nitrate/nitrite response regulator (NarL), receiver domain {Escherichia coli [TaxId: 562]}
epatilliddhpmlrtgvkqlismapditvvgeasngeqgielaesldpdlilldlnmpg
mngletldklrekslsgrivvfsvsnheedvvtalkrgadgyllkdmepedllkalhqaa
agemvlsealtpvlaasl

VLSEALTPVLAASLQLTPRERDILKLIAQGLPNKMIARRLDITESTVKVHVKHMLKKMKL
KSRVEAAVWVHQERIF

>1SCUB1
>d1scub1 3.16.3.1.2 (245-388) Succinyl-CoA synthetase, beta-chain, C-terminal domain {Escherichia coli}
aaqwelnyvaldgnigcmvngaglamgtmdivklhggepanfldvgggatkervteafki
ilsddkvkavlvnifggivrcdliadgiigavaevgvnvpvvvrlegnnaelgakklads
glniiaakgltdaaqqvvaavegk

>d1scub1 c.23.4.1 (B:239-388) Succinyl-CoA synthetase, beta-chain, C-terminal domain {Escherichia coli [TaxId: 562]}
dpreaqaaqwelnyvaldgnigcmvngaglamgtmdivklhggepanfldvgggatkerv
teafkiilsddkvkavlvnifggivrcdliadgiigavaevgvnvpvvvrlegnnaelga
kkladsglniiaakgltdaaqqvvaavegk

MNLHEYQAKQLFARYGLPAPAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLI
HKVALDPLTGPMPYQGRELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVIT
KQGDLICLDGKLGADGNALFRQPDLREMRDQSQED

>1HLPA1
>d1hlpa1 3.2.1.5.5 (21-162) Malate dehydrogenase {Haloarcula marismortui}

tkvsvvgaagtvgaaagynialrdiadevvfvdipdkeddtvgqaadtnhgiaydsntrv
rqggyedtagsdvvvitagiprqpgqtridlagdnapimediqssldehnddyislttsn
pvdllnrhlyeagdrsreqvigfg

>d1hlpa1 c.2.1.5 (A:21-162) Malate dehydrogenase {Archaeon Haloarcula marismortui [TaxId: 2238]}
tkvsvvgaagtvgaaagynialrdiadevvfvdipdkeddtvgqaadtnhgiaydsntrv
rqggyedtagsdvvvitagiprqpgqtridlagdnapimediqssldehnddyislttsn
pvdllnrhlyeagdrsreqvigfg

FGGRLDSARFRYVLSEEFDAPVQNVEGTILGEHGDAQVPVFSKVRVDGTDPEFSGDEKEQ
LLGDLQESAMDVIERKGATEWGPARGVAHMVEAILHDTGEVLPASVKLEGEFGHEDTAFG
VPVRLGSNGVEEIVEWDLDDYEQDLMADAAEKLSDQYDKIS

>1YVEI2
>d1yvei2 3.2.1.6.1 (83-307) Acetohydroxy acid isomeroreductase, ketoacid reductoisomerase (KARI) {Spinach (Spinacia oleracea)}
attfdfdssvfkkekvtlsghdeyivrggrnlfpllpdafkgikqigvigwgsqapaqaq
nlkdslteaksdvvvkiglrkgsnsfaearaagfseengtlgdmwetisgsdlvlllisd
saqadnyekvfshmkpnsilglshgfllghlqslgqdfpknisviavcpkgmgpsvrrly
vqgkevngaginssfavhqdvdgratdvalgwsialgspftfatt

>d1yvei2 c.2.1.6 (I:83-307) Class II ketol-acid reductoisomerase (KARI) {Spinach (Spinacia oleracea) [TaxId: 3562]}
attfdfdssvfkkekvtlsghdeyivrggrnlfpllpdafkgikqigvigwgsqapaqaq
nlkdslteaksdvvvkiglrkgsnsfaearaagfseengtlgdmwetisgsdlvlllisd
saqadnyekvfshmkpnsilglshgfllghlqslgqdfpknisviavcpkgmgpsvrrly
vqgkevngaginssfavhqdvdgratdvalgwsialgspftfatt

FTFATTLEQEYKSDIFGERGILLGAVHGIVECLFRRYTESGMSEDLAYKNTVECITGVIS
KTISTKGMLALYNSLSEEGKKDFQAAYSASYYPSMDILYECYEDVASGSEIRSVVLAGRR
FYEKEGLPAFPMGKIDQTRMWKVGEKVRSVRPAGDLGPLYPFTAGVYVALMMAQIEILRK
KGHSYSEIINESVIEAVDSLNPFMHARGVSFMVDNCSTTARLGSRKWAPRFDYILSQQAL
VAVDNGAPINQDLISNFLSDPVHEAIGVCAQLRPSVDISVTADADFVRPELRQ

>2PGD_2
>d2pgd_2 3.2.1.6.2 (1-176) 6-phosphogluconate dehydrogenase {Sheep (Ovis orientalis aries)}
aqadialiglavmgqnlilnmndhgfvvcafnrtvskvddflaneakgtkvlgahsleem
vsklkkprriillvkagqavdnfieklvplldigdiiidggnseyrdtmrrcrdlkdkgi
lfvgsgvsggedgarygpslmpggnkeawphikaifqgiaakvgtgepccdwvgdd

>d2pgda2 c.2.1.6 (A:1-176) 6-phosphogluconate dehydrogenase {Sheep (Ovis orientalis aries) [TaxId: 9940]}
aqadialiglavmgqnlilnmndhgfvvcafnrtvskvddflaneakgtkvlgahsleem
vsklkkprriillvkagqavdnfieklvplldigdiiidggnseyrdtmrrcrdlkdkgi
lfvgsgvsggedgarygpslmpggnkeawphikaifqgiaakvgtgepccdwvgdd

FVKMVHNGIEYGDMQLICEAYHLMKDVLGLGHKEMAKAFEEWNKTELDSFLIEITASILK
FQDADGKHLLPKIRDSAGQKGTGKWTAISALEYGVPVTLIGEAVFARCLSSLKDERIQAS

KKLKGPQNIPFEGDKKSFLEDIRKALYASKIISYAQGFMLLRQAATEFGWTLNYGGIALM
WRGGCIIRSVFLGKIKDAFDRNPGLQNLLLDDFFKSAVENCQDSWRRAISTGVQAGIPMP
CFTTALSFYDGYRHA

>1HRDA1
>d1hrda1 3.2.1.7.1 (195-449) Glutamate dehydrogenase {Clostridium symbiosum}
karsfggslvrpeatgygsvyyveavmkhendtlvgktvalagfgnvawgaakklaelga
kavtlsgpdgyiydpegitteekinymlemrasgrnkvqdyadkfgvqffpgekpwgqkv
diimpcatqndvdleqakkivannvkyyievanmpttnealrflmqqpnmvvapskavna
ggvlvsgfemsqnserlswtaeevdsklhqvmtdihdgsaaaaeryglgynlvaganivg
fqkiadammaqgiaw

>d1hrda1 c.2.1.7 (A:195-449) Glutamate dehydrogenase {Clostridium symbiosum [TaxId: 1512]}
karsfggslvrpeatgygsvyyveavmkhendtlvgktvalagfgnvawgaakklaelga
kavtlsgpdgyiydpegitteekinymlemrasgrnkvqdyadkfgvqffpgekpwgqkv
diimpcatqndvdleqakkivannvkyyievanmpttnealrflmqqpnmvvapskavna
ggvlvsgfemsqnserlswtaeevdsklhqvmtdihdgsaaaaeryglgynlvaganivg
fqkiadammaqgiaw

EATGYGSVYYVEAVMKHENDTLVGKTVALAGFGNVAWGAAKKLAELGAKAVTLSGPDGYI
YDPEGITTEEKINYMLEMRASGRNKVQDYADKFGVQFFPGEKPWGQKVDIIMPCATQNDV
DLEQAKKIVANNVKYYIEVANMPTTNEALRFLMQQPNMVVAPSKAVN

>1LEHA1
>d1leha1 3.2.1.7.6 (135-364) Leucine dehydrogenase {Bacillus sphaericus}
gispafgssgnpspvtaygvyrgmkaaakeafgsdsleglavsvqglgnvakalckklnt
egaklvvtdvnkaavsaavaeegadavapnaiygvtcdifapcalgavlndftipqlkak
viagsadnqlkdprhgkylhelgivyapdyvinaggvinvadelygynrtramkrvdgiy
dsiekifaiskrdgvpsyvaadrmaeeriakvakarsqflqdqrnilngr

>d1leha1 c.2.1.7 (A:135-364) Leucine dehydrogenase {Bacillus sphaericus [TaxId: 1421]}
gispafgssgnpspvtaygvyrgmkaaakeafgsdsleglavsvqglgnvakalckklnt
egaklvvtdvnkaavsaavaeegadavapnaiygvtcdifapcalgavlndftipqlkak
viagsadnqlkdprhgkylhelgivyapdyvinaggvinvadelygynrtramkrvdgiy
dsiekifaiskrdgvpsyvaadrmaeeriakvakarsqflqdqrnilngr

MEIFKYMEKYDYEQLVFCQDEASGLKAVIAIHDTTLGPALGGARMWTYNAEEEAIEDALR
LARGMTGGGKTVIIGDPFADKNEDMFRALGRFIQGLNGRYITAEDVGTTVDDMDLIHQET
DYVRMAEERIAKVAKARSQFLQDQRNILNGR

>1TADA2
>d1tada2 3.31.1.7.12 (27-56,178-342) Transducin (alpha subunit) {Bovine (Bos taurus)}
artvkllllgagesgkstivkqmkiihqdgXtgiietqfsfkdlnfrmfdvggqrserkk
wihcfegvtciifiaalsaydmvlveddevnrmheslhlfnsicnhryfattsivlflnk
kdvfsekikkahlsicfpdyngpntyedagnyikvqflelnmrrdvkeiyshmtcatdtq
nvkfvfdavtdiiike

>d1tada2 c.37.1.8 (A:27-56,A:178-342) Transducin (alpha subunit) {Cow (Bos taurus) [TaxId: 9913]}
artvkllllgagesgkstivkqmkiihqdgXtgiietqfsfkdlnfrmfdvggqrserkk
wihcfegvtciifiaalsaydmvlveddevnrmheslhlfnsicnhryfattsivlflnk
kdvfsekikkahlsicfpdyngpntyedagnyikvqflelnmrrdvkeiyshmtcatdtq
nvkfvfdavtdiiike

SLEECLEFIAIIYGNTLQSILAIVRAMTTLNIQYGDSARQDDARKLMHMADTIEEGTMPK
EMSDIIQRLWKDSGIQACFDRASEYQLNDSAGYYLSDLERLVTPGYVPTEQDVLRSRVK

>1EFT_3
>d1eft_3 3.31.1.7.15 (1-212) Elongation factor Tu (EF-Tu), N-terminal (G) domain {Thermus aquaticus}
akgefirtkphvnvgtighvdhgkttltaaltfvtaaenpnvevkdygdidkapeerarg
itintahveyetakrhyshvdcpghadyiknmitgaaqmdgailvvsaadgpmpqtrehi
llarqvgvpyivvfmnkvdmvddpelldlvemevrdllnqyefpgdevpvirgsallale
emhknpktkrgenewvdkiwelldaideyipt

>d1efta3 c.37.1.8 (A:1-212) Elongation factor Tu (EF-Tu), N-terminal (G) domain {Thermus aquaticus [TaxId: 271]}
akgefirtkphvnvgtighvdhgkttltaaltfvtaaenpnvevkdygdidkapeerarg
itintahveyetakrhyshvdcpghadyiknmitgaaqmdgailvvsaadgpmpqtrehi
llarqvgvpyivvfmnkvdmvddpelldlvemevrdllnqyefpgdevpvirgsallale
emhknpktkrgenewvdkiwelldaideyipt

TPHTKFEASVYVLKKEEGGRHTGFFSGYRPQFYFRTTDVTGVVRLPQGVEMVMPGDNVTF
TVELIKPVALEEGLRFAIREGGRTVGAGVVTKIL

>1DAR_2
>d1dar_2 3.31.1.7.18 (1-282) Elongation factor G (EF-G), N-terminal (G) domain {Thermus thermophilus}
mavkveydlkrlrnigiaahidagkttttterilyytgrihkigevhegaatmdfmeqere
rgititaavttcfwkdhriniidtpghvdftieversmrvldgaivvfdssqgvepqset
vwrqaekykvpriafankmdktgadlwlvirtmqerlgarpvvmqlpigredtfsgiidv
lrmkaytygndlgtdireipipeeyldqareyheklvevaadfdenimlkylegeeptee
elvaairkgtidlkitpvflgsalknkgvqllldavvdylps

>d1dara2 c.37.1.8 (A:1-282) Elongation factor G (EF-G), N-terminal (G) domain {Thermus thermophilus [TaxId: 274]}
mavkveydlkrlrnigiaahidagkttttterilyytgrihkigevhegaatmdfmeqere
rgititaavttcfwkdhriniidtpghvdftieversmrvldgaivvfdssqgvepqset
vwrqaekykvpriafankmdktgadlwlvirtmqerlgarpvvmqlpigredtfsgiidv
lrmkaytygndlgtdireipipeeyldqareyheklvevaadfdenimlkylegeeptee
elvaairkgtidlkitpvflgsalknkgvqllldavvdylps

PDPNGPLAALAFKIMADPYVGRLTFIRVYSGTLTSGSYVYNTTKGRKERVARLLRMHANH
REEVEELKAGDLGAVVGLKETITGDTLVGEDAPRVILE

>1BMFA3
>d1bmfa3 3.31.1.10.3 (95-379) Central domain of alpha and beta subunits of F1 ATP synthase {Bovine (Bos taurus)}
vdvpvgeellgrvvdalgnaidgkgpigskarrrvglkapgiiprisvrepmqtgikavd
slvpigrgqreliigdrqtgktsiaidtiinqkrfndgtdekkklyciyvaigqkrstva
qlvkrltdadamkytivvsatasdaaplqylapysgcsmgeyfrdngkhaliiyddlskq
avayrqmslllrrppgreaypgdvfylhsrlleraakmndafgggsltalpvietqagdv
sayiptnvisitdgqifletelfykgirpainvglsvsrvgsaaq

>d1bmfa3 c.37.1.11 (A:95-379) Central domain of alpha subunit of F1 ATP synthase {Cow (Bos taurus) [TaxId: 9913]}
vdvpvgeellgrvvdalgnaidgkgpigskarrrvglkapgiiprisvrepmqtgikavd
slvpigrgqreliigdrqtgktsiaidtiinqkrfndgtdekkklyciyvaigqkrstva
qlvkrltdadamkytivvsatasdaaplqylapysgcsmgeyfrdngkhaliiyddlskq
avayrqmslllrrppgreaypgdvfylhsrlleraakmndafgggsltalpvietqagdv
sayiptnvisitdgqifletelfykgirpainvglsvsrvgsaaq

TRAMKQVAGTMKLELAQYREVAAFAQFGSDLDAATQQLLSRGVRLTELLKQGQYSPMAIE
EQVAVIYAGVRGYLDKLEPSKITKFENAFLSHVISQHQALLGKIRTDGKISEESDAKLKE
IVTNFLAGFEA

>1BMFD3
>d1bmfd3 3.31.1.10.3 (82-357) Central domain of alpha and beta subunits of F1 ATP synthase {Bovine (Bos taurus)}
iripvgpetlgrimnvigepidergpiktkqfaaihaeapefvemsveqeilvtgikvvd
llapyakggkiglfggagvgktvlimelinnvakahggysvfagvgertregndlyhemi
esgvinlkdatskvalvygqmneppgararvaltgltvaeyfrdqegqdvllfidnifrf
tqagsevsallgripsavgyqptlatdmgtmqeritttkkgsitsvqaiyvpaddltdpa
pattfahldattvlsraiaelgiypavdpldstsri

>d1bmfd3 c.37.1.11 (D:82-357) Central domain of beta subunit of F1 ATP synthase {Cow (Bos taurus) [TaxId: 9913]}
iripvgpetlgrimnvigepidergpiktkqfaaihaeapefvemsveqeilvtgikvvd
llapyakggkiglfggagvgktvlimelinnvakahggysvfagvgertregndlyhemi
esgvinlkdatskvalvygqmneppgararvaltgltvaeyfrdqegqdvllfidnifrf
tqagsevsallgripsavgyqptlatdmgtmqeritttkkgsitsvqaiyvpaddltdpa
pattfahldattvlsraiaelgiypavdpldstsri

MDPNIVGSEHYDVARGVQKILQDYKSLQDIIAILGMDELSEEDKLTVSRARKIQRFLSQP
FQVAEVFTGHLGKLVPLKETIKGFQQILAGEYDHLPEQAFYMVGPIEEAVAKADKLAE

>2GSTA2
>d2gsta2 3.42.1.5.1 (1-84) Glutathione S-transferase {Rat (Rattus rattus)}
pmilgywnvrglthpirllleytdssyeekryamgdapdydrsqwlnekfklgldfpnlp
ylidgsrkitqsnaimrylarkhh

>d2gsta2 c.47.1.5 (A:1-84) Class mu GST {Rat (Rattus norvegicus) [TaxId: 10116]}

pmilgywnvrglthpirllleytdssyeekryamgdapdydrsqwlnekfklgldfpnlp
ylidgsrkitqsnaimrylarkhh

CGETEEERIRADIVENQVMDNRMQLIMLCYNPDFEKQKPEFLKTIPEKMKLYSEFLGKRP
WFAGDKVTYVDFLAYDILDQYHIFEPKCLDAFPNLKDFLARFEGL

>1GSEA2
>d1gsea2 3.42.1.5.9 (2-80) Glutathione S-transferase {Human (Homo sapiens), class alpha}
aekpklhyfnargkmestrwllaaagvefeekfiksaedldklrndgylmfqqvpmveid
gmklvqtrailnyiaskyn

>d1gsea2 c.47.1.5 (A:2-80) Class alpha GST {Human (Homo sapiens), (a1-1) [TaxId: 9606]}
aekpklhyfnargkmestrwllaaagvefeekfiksaedldklrndgylmfqqvpmveid
gmklvqtrailnyiaskyn

YGKDIKERALIDMYIEGIADLGEMILLLPVCPPEEKDAKLALIKEKIKNRYFPAFEKVLK
SHGQDYLVGNKLSRADIHLVELLYYVEELDSSLISSFPLLKALKTRISN

>1GNWA2
>d1gnwa2 3.42.1.5.16 (2-85) Glutathione S-transferase {Mouse-ear cress (Arabidopsis thaliana)}
gikvfghpasiatrrvlialheknldfelvhvelkdgehkkepflsrnpfgqvpafedgd
lklfesraitqyiahryenqgtnl

>d1gnwa2 c.47.1.5 (A:2-85) Class phi GST {Mouse-ear cress (Arabidopsis thaliana) [TaxId: 3702]}
gikvfghpasiatrrvlialheknldfelvhvelkdgehkkepflsrnpfgqvpafedgd
lklfesraitqyiahryenqgtnl

QTDSKNISQYAIMAIGMQVEDHQFDPVASKLAFEQIFKSIYGLTTDEAVVAEEEAKLAKV
LDVYEARLKEFKYLAGETFTLTDLHHIPAIQYLLGTPTKKLFTERPRVNEWVAEITKRPA

>1HPM_1
>d1hpm_1 3.50.1.1.1 (4-188) Heat shock protein 70kDa, ATPase fragment {Bovine (Bos taurus)}
gpavgidlgttyscvgvfqhgkveiiandqgnrttpsyvaftdterligdaaknqvamnp
tntvfdakrligrrfddavvqsdmkhwpfmvvndagrpkvqveykgetksfypeevssmv
ltkmkeiaeaylgktvtnavvtvpayfndsqrqatkdagtiaglnvlriineptaaaiay
gldkk

>d1hpma1 c.55.1.1 (A:4-188) Heat shock protein 70kDa, ATPase fragment {Cow (Bos taurus) [TaxId: 9913]}
gpavgidlgttyscvgvfqhgkveiiandqgnrttpsyvaftdterligdaaknqvamnp
tntvfdakrligrrfddavvqsdmkhwpfmvvndagrpkvqveykgetksfypeevssmv
ltkmkeiaeaylgktvtnavvtvpayfndsqrqatkdagtiaglnvlriineptaaaiay
gldkk

GPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNP
TNTVVLTKMKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKDAGTIAGLNVLRIINEPTA
AAIAYGLDKKSINPDEAVAYGAAVQAAILS

>1HPM_2
>d1hpm_2 3.50.1.1.1 (189-381) Heat shock protein 70kDa, ATPase fragment {Bovine (Bos taurus)}
vgaernvlifdlgggtfdvsiltiedgifevkstagdthlggedfdnrmvnhfiaefkrk
hkkdisenkravrrlrtacerakrtlssstqasieidslyegidfytsitrarfeelnad
lfrgtldpvekalrdakldksqihdivlvggstripkiqkllqdffngkelnksinpdea
vaygaavqaails

>d1hpma2 c.55.1.1 (A:189-381) Heat shock protein 70kDa, ATPase fragment {Cow (Bos taurus) [TaxId: 9913]}
vgaernvlifdlgggtfdvsiltiedgifevkstagdthlggedfdnrmvnhfiaefkrk
hkkdisenkravrrlrtacerakrtlssstqasieidslyegidfytsitrarfeelnad
lfrgtldpvekalrdakldksqihdivlvggstripkiqkllqdffngkelnksinpdea
vaygaavqaails

FDAKRLIGRRFDDAVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSM

>2BTFA1
>d2btfa1 3.50.1.1.4 (2-146) Actin {Bovine (Bos taurus), pancreas}
dddiaalvvdngsgmckagfagddapravfpsivgrprhqgvmvgmgqkdsyvgdeaqsk
rgiltlkypiehgivtnwddmekiwhhtfynelrvapeehpvllteaplnpkanrekmtq
imfetfntpamyvaiqavlslyasg

>d2btfa1 c.55.1.1 (A:2-146) Actin {Cow (Bos taurus) [TaxId: 9913]}
dddiaalvvdngsgmckagfagddapravfpsivgrprhqgvmvgmgqkdsyvgdeaqsk
rgiltlkypiehgivtnwddmekiwhhtfynelrvapeehpvllteaplnpkanrekmtq
imfetfntpamyvaiqavlslyasg

DDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQSKR
GILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQI
MFETFNTPAMYVAIQVWIGGSILASLSTFQQMWISKQEYDESGPSIVHRK

>2BTFA2
>d2btfa2 3.50.1.1.4 (147-375) Actin {Bovine (Bos taurus), pancreas}
rttgivmdsgdgvthtvpiyegyalphailrldlagrdltdylmkiltergysfttaer
eivrdikeklcyvaldfeqemataasssleksyelpdgqvitignerfrcpealfqpsf
lgmescgihettfnsimkcdvdirkdlyantvlsggttmypgiadrmqkeitalapstmk
ikiiapperkysvwiggsilaslstfqqmwiskqeydesgpsivhrkcf

>d2btfa2 c.55.1.1 (A:147-375) Actin {Cow (Bos taurus) [TaxId: 9913]}
rttgivmdsgdgvthtvpiyegyalphailrldlagrdltdylmkiltergysfttaer
eivrdikeklcyvaldfeqemataasssleksyelpdgqvitignerfrcpealfqpsf
lgmescgihettfnsimkcdvdirkdlyantvlsggttmypgiadrmqkeitalapstmk
ikiiapperkysvwiggsilaslstfqqmwiskqeydesgpsivhrkcf

AVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDCGIHETTFNSIMKCDVDI
RKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYS

>1BCO_2
>d1bco_2 3.50.3.3.1 (258-480) mu transposase, core domain {Bacteriophage mu}
ehldamqwingdgylhnvfvrwfngdvirpktwfwqdvktrkilgwrcdvsenidsirls
fmdvvtrygipedfhitidntrgaankwltggapnryrfkvkeddpkglfllmgakmhwt
svvagkgwgqakpverafgvggleeyvdkhpalagaytgpnpqakpdnygdravdaelfl
ktlaegvamfnartgretemcggklsfddvfereyartivrkp

>d1bcoa2 c.55.3.3 (A:258-480) mu transposase, core domain {Bacteriophage mu [TaxId: 10677]}
ehldamqwingdgylhnvfvrwfngdvirpktwfwqdvktrkilgwrcdvsenidsirls
fmdvvtrygipedfhitidntrgaankwltggapnryrfkvkeddpkglfllmgakmhwt
svvagkgwgqakpverafgvggleeyvdkhpalagaytgpnpqakpdnygdravdaelfl
ktlaegvamfnartgretemcggklsfddvfereyartivrkp

AEAVNVSRKGEFTLKVGGSLKGAKNVYYNMALMNAGVKKVVVRFDPQQLHSTVYCYTLDG
RFICEAEC

>1SFE_2
>d1sfe_2 3.50.4.1.1 (12-92) Ada DNA repair protein {Escherichia coli}
lavryaladcelgrclvaesergicaillgdddatliselqqmfpaadnapadlmfqqhv
reviaslnqrdtpltlpldir

>d1sfea2 c.55.7.1 (A:12-92) Ada DNA repair protein {Escherichia coli [TaxId: 562]}
lavryaladcelgrclvaesergicaillgdddatliselqqmfpaadnapadlmfqqhv
reviaslnqrdtpltlpldir

GTAFQQQVWQALRTIPCGETVSYQQLANAIGKPKAVRAVASACAANKLAIVIPCHRVVRG
DGSLSGYRWGVSRKAQLLRREAEN

>1HPLA2
>d1hpla2 3.64.1.16.1 (1-336) Pancreatic lipase, N-terminal domain {Horse (Equus caballus)}
nevcyerlgcfsddspwagiverplkilpwspekvntrfllytnenpdnfqeivadpsti
qssnfntgrktrfiihgfidkgeeswlstmcqnmfkvesvncicvdwksgsrtaysqasq
nvrivgaevaylvgvlqssfdyspsnvhiighslgshaageagrrtngavgritgldpae
pcfqgtpelvrldpsdaqfvdvihtdiapfipnlgfgmsqtaghldffpnggkempgcqk
nvlsqivdidgiwqgtrdfaacnhlrsykyytdsilnpdgfagfscasysdftankcfpc
ssegcpqmghyadrfpgrtkgvgqlfylntgdasnfa

>d1hpla2 c.69.1.19 (A:1-336) Pancreatic lipase, N-terminal domain {Horse (Equus caballus) [TaxId: 9796]}
nevcyerlgcfsddspwagiverplkilpwspekvntrfllytnenpdnfqeivadpsti
qssnfntgrktrfiihgfidkgeeswlstmcqnmfkvesvncicvdwksgsrtaysqasq
nvrivgaevaylvgvlqssfdyspsnvhiighslgshaageagrrtngavgritgldpae
pcfqgtpelvrldpsdaqfvdvihtdiapfipnlgfgmsqtaghldffpnggkempgcqk
nvlsqivdidgiwqgtrdfaacnhlrsykyytdsilnpdgfagfscasysdftankcfpc
ssegcpqmghyadrfpgrtkgvgqlfylntgdasnfa

WRYRVDVTLSGKKVTGHVLVSLFGNKGNSRQYEIFQGTLKPDNTYSNEFDSDVEVGDLEK
VKFIWYNNVINLTLPKVGASKITVERNDGSVFNFCSEETVREDVLLTLTAC

>1ALO_2 (the PDB ID in release 1.75 is updated as 1VLBa2)
>d1alo_2 4.13.6.2.2 (1-80) Aldehyde oxidoreductase, N-terminal domain {Desulfovibrio gigas}
miqkvitvngieqnlfvdaeallsdvlrqqlgltgvkvgceqgqcgacsvildgkvvrac
vtkmkrvadgaqittiegvg

>d1vlba2 d.15.4.2 (A:1-80) Aldehyde oxidoreductase, N-terminal domain {Desulfovibrio gigas [TaxId: 879]}
miqkvitvngieqnlfvdaeallsdvlrqqlgltgvkvgceqgqcgacsvildgkvvrac
vtkmkrvadgaqittiegvg

TIEGVGQPENLHPLQKAWVLHGGAQCGFCSPGFIVSAKGLLDTNADPSREDVRDWFQKHR
NACRCTGYKPLVDAVMDAAAVINGKKPETDLEFKMPADGRIWGSKYPRPTAVAKVTGTL

>3RUBL2
>d3rubl2 4.48.9.1.1 (22-147) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Tobacco (Nicotiana tabacum), variant turkish samsun}
ltyytpeyqtkdtdilaafrvtpqpgvppeeagaavaaesstgtwttvwtdgltsldryk
grcyriervvgekdqyiayvaypldlfeegsvtnmftsivgnvfgfkalralrledlrip
payvkt

>d3rubl2 d.58.9.1 (L:22-147) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Tobacco (Nicotiana tabacum), variant turkish samsun [TaxId: 4097]}
ltyytpeyqtkdtdilaafrvtpqpgvppeeagaavaaesstgtwttvwtdgltsldryk
grcyriervvgekdqyiayvaypldlfeegsvtnmftsivgnvfgfkalralrledlrip
payvkt

QGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNS
QPFMRWRDRFLFCAEALYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHD
YLTGGFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGIHFRVLAKALRMSGGDHIHS
GTVVGKLEGERDITLGFVDLLRDDFVEQDRSRGIYFTQDWVSLPGVLPVASGGIHVWHMP
ALTEIFGDDSVLQFGGMTLGHPWGNAPGAVANRVALEACVKARNEGRDLAQEGNEIIREA
CKWSPELAAACEVWKEIV

>5RUBA2
>d5ruba2 4.48.9.1.6 (2-137) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Rhodospirillum rubrum}
dqssryvnlalkeedliaggehvlcayimkpkagygyvataahfaaesstgtnvevcttd
dftrgvdalvyevdeareltkiaypvalfdrnitdgkamiasfltltmgnnqgmgdveya
kmhdfyvpeayralfd

>d5ruba2 d.58.9.1 (A:2-137) Ribulose 1,5-bisphosphate carboxylase-oxygenase {Rhodospirillum rubrum [TaxId: 1085]}
dqssryvnlalkeedliaggehvlcayimkpkagygyvataahfaaesstgtnvevcttd

dftrgvdalvyevdeareltkiaypvalfdrnitdgkamiasfltltmgnnqgmgdveya
kmhdfyvpeayralfd

DGPSVNISALWKVLGRPEVDGGLVVGTIIKPKLGLRPKPFAEACHAFWLGGDFIKNDEPQ
GNQPFAPLRDTIALVADAMRRAQDETGEAKLFSANITADDPFEIIARGEYVLETFGENAS
HVALLVDGYVAGAAAITTARRRFPDNFLHYHRAGHGAVTSPQSKRGYTAFVHCKMARLQG
ASGIHTGTSSDRAIAYMLTQDEAQGPFYRQSWGGMKACTPIISGGMNALRMPGFFENLGN
ANVILTAGGGAFGHIDGPVAGARSLRQAWQAWRDGVPVLDYAREHKELARAFES

>1DAR_4
>d1dar_4 4.48.11.1.1 (600-689) Elongation factor G (EF-G), domain V {Thermus thermophilus}
vilepimrvevttpeeymgdvigdlnarrgqilgmeprgnaqvirafvplaemfgyatdl
rsktqgrgsfvmffdhyqevpkqvqeklik

>d1dara4 d.58.11.1 (A:600-689) Elongation factor G (EF-G) {Thermus thermophilus [TaxId: 274]}
vilepimrvevttpeeymgdvigdlnarrgqilgmeprgnaqvirafvplaemfgyatdl
rsktqgrgsfvmffdhyqevpkqvqeklik

RETITKPVDVEGKFIRQTGGRGQYGHVKIKVEPLPRGSGFEFVNAIVGGVIPKEYIPAVQ
KGIEEAMQSGPLIGFPVVDIKVTLYDGSYHEVDSSEMAFKIAGSMAIKEAVQKGDPVIL

>1MLA_2
>d1mla_2 4.48.20.1.1 (128-197) Probable ACP-binding domain of malonyl-CoA ACP transacylase {Escherichia coli}
gtgamaaiigldasiakaceeaaegqvvspvnfnspgqvviaghkeaveragaackaag
akralplpvs

>d1mlaa2 d.58.23.1 (A:128-197) Probable ACP-binding domain of malonyl-CoA ACP transacylase {Escherichia coli [TaxId: 562]}
gtgamaaiigldasiakaceeaaegqvvspvnfnspgqvviaghkeaveragaackaag
akralplpvs

QFAFVFPGQGSQTVGMLADMAASYPIVEETFAEASAALGYDLWALTQQGPAEELNKTWQT
QPALLTASVALYRVWQQQGGKAPAMMAGHSLGEYSALVCAGVIDFADAVRLVEMRGKFMQ
EAVPSHCALMKPAADKLAVELAKITFNAPTVPVVNNVDVKCETNGDAIRDALVRQLYNPV
QWTKSVEYMAAQGVEHLYEVGPGKVLTGLTKRIVDTLTASALNEPSAMAAAL

>1VAOA1
>d1vaoa1 4.48.27.1.1 (274-560) Vanillyl-alcohol oxidase {Fungus (Penicillium simplicissimum)}
rgyqsylitlpkdgdlkqavdiirplrlgmalqnvptirhilldaavlgdkrsyssrtep
lsdeeldkiakqlnlgrwnfygalygpepirrvlwetikdafsaipgvkfyfpedtpens
vlrvrdktmqgiptydelkwidwlpngahlffspiakvsgedammqyavtkkrcqeagld
figtftvgmremhhivcivfnkkdliqkrkvqwlmrtliddcaangwgeyrthlafmdqi
metynwnnssflrfnevlknavdpngiiapgksgvwpsqyshvtwkl

>d1vaoa1 d.58.32.1 (A:274-560) Vanillyl-alcohol oxidase {Fungus (Penicillium simplicissimum) [TaxId: 69488]}
rgyqsylitlpkdgdlkqavdiirplrlgmalqnvptirhilldaavlgdkrsyssrtep

lsdeeldkiakqlnlgrwnfygalygpepirrvlwetikdafsaipgvkfyfpedtpens
vlrvrdktmqgiptydelkwidwlpngahlffspiakvsgedammqyavtkkrcqeagld
figtftvgmremhhivcivfnkkdliqkrkvqwlmrtliddcaangwgeyrthlafmdqi
metynwnnssflrfnevlknavdpngiiapgksgvwpsqyshvtwkl

EFRPLTLPPKLSLSDFNEFIQDIIRIVGSENVEVISVDGSYMKPTHTHDPHHVMDQDYFL
ASAIVAPRNVADVQSIVGLANKFSFPLWPISIGRNSGYGGAAPRVSGSVVLDMGKNMN

>1GEO_1 (the PDB ID in release 1.75 is updated as 1AOPa1)
>d1aop_1 4.48.29.1.1 (81-145) Sulfite reductase, domains 1 and 3 {Escherichia coli}
llrcrlpggvittkqwqaidkfagentiygsirltnrqtfqfhgilkknvkpvhqmlhsv
gldal

>d1aopa1 d.58.36.1 (A:81-145) Sulfite reductase, domains 1 and 3 {Escherichia coli [TaxId: 562]}
llrcrlpggvittkqwqaidkfagentiygsirltnrqtfqfhgilkknvkpvhqmlhsv
gldal

NDMNRNVLCTSNPYESQLHAEAYEWAKKISEHLLPTYLPRKFKTTVVIPPQNDIDLHAND
MNFVAIAENGKLVGFNLLVGGGLSIEHGNKKTYARTASEFGYLPLEHTLAVAEAVVTTQR
DWGNRTDRKNAKTKYTLERVGVETFKAEVERRAGIKFEPIRPYEFT

>1GEO_2 (the PDB ID in release 1.75 is updated as 1AOPa2)
>d1aop_2 4.48.29.1.1 (346-425) Sulfite reductase, domains 1 and 3 {Escherichia coli}
igwvkgiddnwhltlfiengrildyparplktglleiakihkgdfritanqnliiagvpe
sekakiekiakesglmnavt

>d1aopa2 d.58.36.1 (A:346-425) Sulfite reductase, domains 1 and 3 {Escherichia coli [TaxId: 562]}
igwvkgiddnwhltlfiengrildyparplktglleiakihkgdfritanqnliiagvpe
sekakiekiakesglmnavt

LLRCRLPGGVITTKQWQAIDKFAGENTIYGSIRLTNRQTFQFHGILPVHQMLHSVGLDAL
GRGDRIGWVKGIDDQWHLTLFIENGRILDYPARPLKTGLLEIAKIHKGDFRITANQNLII
AGVPESEKAKIEKIAKESGLMNAV

>1PRHA2
>d1prha2 7.3.11.1.9 (33-73) Prostaglandin H2 synthase-1, EGF-like module {Sheep (Ovis aries)}
vnpccyypcqhqgicvrfgldryqcdctrtgysgpnctipe

>d1prha2 g.3.11.1 (A:33-73) Prostaglandin H2 synthase-1, EGF-like module {Sheep (Ovis aries)
[TaxId: 9940]}
vnpccyypcqhqgicvrfgldryqcdctrtgysgpnctipe

IWTWLRTTLRPSPSFIHFLLTHGRWLWDFVNATFIRDTLMRLVLTVRSNLIPSPPTYNIA
HDYISWESFSNVSYYTRILPSVPRDCPTPMGTKGKKQLPDAEFLSRRFLLRRKFIPDPQG
TNLMFAFFAQHFTHQFFKTSGKMGPGFTKALGHGVDLGHIYGDNLERQYQLRLFKDGKLK
YQMLNGEVYPPSVEEAPVLMHYPRGIPPQSQMAVGQEVFGLLPGLMLYATIWLREHQRVC
DLLKAEHPTWGDEQLFQTAKLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGAQFQY
RNRIAMEFNQLYHWHPLMPDSFRVGPQDYSYEQFLFNTSMLVDYGVEALVDAFSRQPAGR

IGGGRNIDHHILHVAVDVIKESRVLRLQPFNEYRKRFGMKPYTSFQELTGEKEMAAELEE
LYGDIDALEFYPGLLLEKCHPNSIFGESMIEMGAPFSLKGLLGNPICSPEYWKASTFGGE
VGFNLVKTATLKKLVCLNTKTCPYVSFHVP

# Supplementary Material B for "Improving taxonomy-based protein fold recognition by using global and local features"

Jian-Yi Yang and Xin Chen
Division of Mathematical Sciences, School of Physical and Mathematical Sciences,
Nanyang Technological University, 21 Nanyang Link, Singapore, 637371.

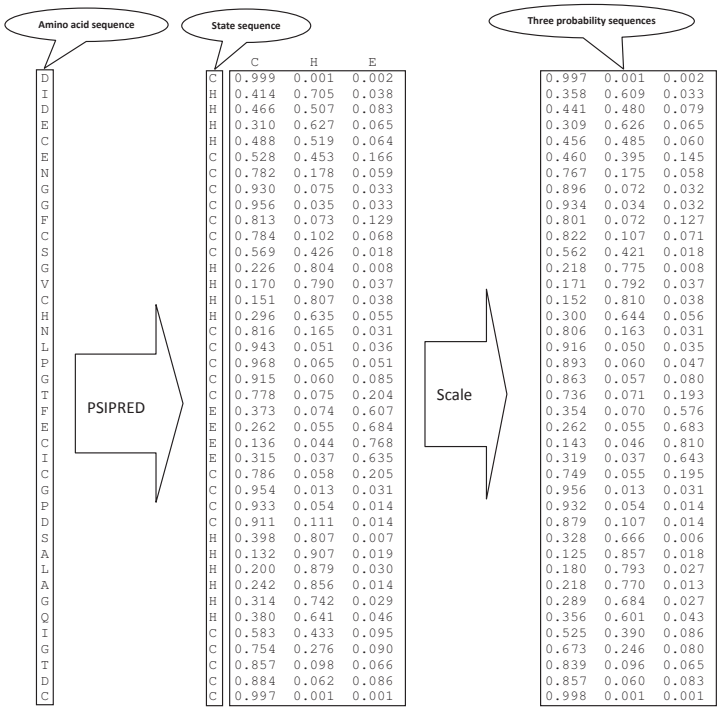
## 1 An example of PSIPRED profiles

Figure 1 gives an example of the PSIPRED profile. The amino acid sequence of a protein domain is submitted to PSIPRED to predict its secondary structure. Besides the predicted secondary structure sequence (we call it a *state sequence*), there are the other three sequences returned by PSIPRED, which are used to measure the confidence levels of the corresponding predictions. We scale these sequences so that the sum of the three confidence values at each position along the amino acid sequence is one. We call the scaled sequences as the *probability sequences*. Therefore, a PSIPRED profile is made of a state sequence and three probability sequences.

## 2 Chaos game representation of the state sequence

We propose here a way to build a new set of features based on the *chaos game representation* (CGR) of a state sequence. The CGR was initially developed to visualize DNA sequences [1], and later applied to protein amino acid sequences as well [2, 3, 4, 5]. Given a state sequence, we start with a equilateral triangle with the unit length of sides and each vertex associated with a distinct letter of H, E and C. For each letter of the given state sequence, we then plot a point inside the triangle in the following way. The first point is placed half way between the center of the triangle and the vertex corresponding to the first letter of the state sequence, and the $i$-th point is then placed half way between the $(i-1)$-th point and the vertex corresponding to the $i$-th letter. The obtained plot is then called the CGR of the state sequence. Figure 2 depicts the CGR for a protein.

Observe that each state sequence gives rise to a distinct $(x, y)$-coordinate sequence of the plotted points. Hence we can faithfully model a CGR plot as a combination of two time series, one composed of the $x$-coordinates and the other of the $y$-coordinates. For simplicity, we call them $x$-time series and $y$-time series, respectively. Because a CGR plot can be fully reconstructed from the corresponding $x$ and $y$ time series, no information present in the CGR plot would be




| | | C | H | E |
|---|---|---|---|---|
| D | C | 0.999 | 0.001 | 0.002 |
| I | H | 0.414 | 0.705 | 0.038 |
| D | H | 0.466 | 0.507 | 0.083 |
| E | H | 0.310 | 0.627 | 0.065 |
| C | H | 0.488 | 0.519 | 0.064 |
| E | C | 0.528 | 0.453 | 0.166 |
| N | C | 0.782 | 0.178 | 0.059 |
| G | C | 0.930 | 0.075 | 0.033 |
| G | C | 0.956 | 0.035 | 0.033 |
| F | C | 0.813 | 0.073 | 0.129 |
| C | C | 0.784 | 0.102 | 0.068 |
| S | C | 0.569 | 0.426 | 0.018 |
| G | H | 0.226 | 0.804 | 0.008 |
| V | H | 0.170 | 0.790 | 0.037 |
| C | H | 0.151 | 0.807 | 0.038 |
| H | H | 0.296 | 0.635 | 0.055 |
| N | C | 0.816 | 0.165 | 0.031 |
| L | C | 0.943 | 0.051 | 0.036 |
| P | C | 0.968 | 0.065 | 0.051 |
| G | C | 0.915 | 0.060 | 0.085 |
| T | C | 0.778 | 0.075 | 0.204 |
| F | E | 0.373 | 0.074 | 0.607 |
| E | E | 0.262 | 0.055 | 0.684 |
| C | E | 0.136 | 0.044 | 0.768 |
| I | E | 0.315 | 0.037 | 0.635 |
| C | C | 0.786 | 0.058 | 0.205 |
| G | C | 0.954 | 0.013 | 0.031 |
| P | C | 0.933 | 0.054 | 0.014 |
| D | C | 0.911 | 0.111 | 0.014 |
| S | H | 0.398 | 0.807 | 0.007 |
| A | H | 0.132 | 0.907 | 0.019 |
| L | H | 0.200 | 0.879 | 0.030 |
| A | H | 0.242 | 0.856 | 0.014 |
| G | H | 0.314 | 0.742 | 0.029 |
| Q | H | 0.380 | 0.641 | 0.046 |
| I | C | 0.583 | 0.433 | 0.095 |
| G | C | 0.754 | 0.276 | 0.090 |
| T | C | 0.857 | 0.098 | 0.066 |
| D | C | 0.884 | 0.062 | 0.086 |
| C | C | 0.997 | 0.001 | 0.001 |

| | | | |
|---|---|---|---|
| 0.997 | 0.001 | 0.002 |
| 0.358 | 0.609 | 0.033 |
| 0.441 | 0.480 | 0.079 |
| 0.309 | 0.626 | 0.065 |
| 0.456 | 0.485 | 0.060 |
| 0.460 | 0.395 | 0.145 |
| 0.767 | 0.175 | 0.058 |
| 0.896 | 0.072 | 0.032 |
| 0.934 | 0.034 | 0.032 |
| 0.801 | 0.072 | 0.127 |
| 0.822 | 0.107 | 0.071 |
| 0.562 | 0.421 | 0.018 |
| 0.218 | 0.775 | 0.008 |
| 0.171 | 0.792 | 0.037 |
| 0.152 | 0.810 | 0.038 |
| 0.300 | 0.644 | 0.056 |
| 0.806 | 0.163 | 0.031 |
| 0.916 | 0.050 | 0.035 |
| 0.893 | 0.060 | 0.047 |
| 0.863 | 0.057 | 0.080 |
| 0.736 | 0.071 | 0.193 |
| 0.354 | 0.070 | 0.576 |
| 0.262 | 0.055 | 0.683 |
| 0.143 | 0.046 | 0.810 |
| 0.319 | 0.037 | 0.643 |
| 0.749 | 0.055 | 0.195 |
| 0.956 | 0.013 | 0.031 |
| 0.932 | 0.054 | 0.014 |
| 0.879 | 0.107 | 0.014 |
| 0.328 | 0.666 | 0.006 |
| 0.125 | 0.857 | 0.018 |
| 0.180 | 0.793 | 0.027 |
| 0.218 | 0.770 | 0.013 |
| 0.289 | 0.684 | 0.027 |
| 0.356 | 0.601 | 0.043 |
| 0.525 | 0.390 | 0.086 |
| 0.673 | 0.246 | 0.080 |
| 0.839 | 0.096 | 0.065 |
| 0.857 | 0.060 | 0.083 |
| 0.998 | 0.001 | 0.001 |

Figure 1: An example of PSIPRED profile.

lost in the combination of two time series. For example, the two time series corresponding to the CGR plots of Figure 2 are depicted in Figure 3.

The average values of $x$- and $y$-time series points are calculated respectively as

$$\bar{x} = \frac{1}{L}\sum_{i=1}^{L} x_i \quad \text{and} \quad \bar{y} = \frac{1}{L}\sum_{i=1}^{L} y_i, \tag{1}$$

where $L$ denotes the length of the time series and $x_i$ and $y_i$ are the coordinate values of the $i$-th point in CGR. These two variables are included into our feature sets.

## 3 Recurrence plot

*Recurrence plot* (RP) is a purely graphical tool originally proposed by Eckmann *et al.* [6] to detect patterns of recurrence in the data. Here, it is used to describe natural time correlation information in a time series. Given a time series $z_1 z_2 \cdots z_L$ of length $L$, we first embed it into the space $R^m$ of dimension $m$ using a time delay $\tau$. Let us define

$$Z_i = (z_i, z_{i+\tau}, z_{i+2\tau}, \cdots, z_{i+(m-1)\tau}), \quad i = 1, 2, \cdots, N_m, \tag{2}$$

where $N_m = L - (m-1)\tau$. Hence, we obtain $N_m$ vectors (*i.e.*, points) in the embedding space $R^m$. While the values of $m$ and $\tau$ have to be chosen appropriately based on nonlinear dynamical theory [7], $\tau$ is often set to be 1 in practical. Because an $\alpha$-helix segment generally comprises

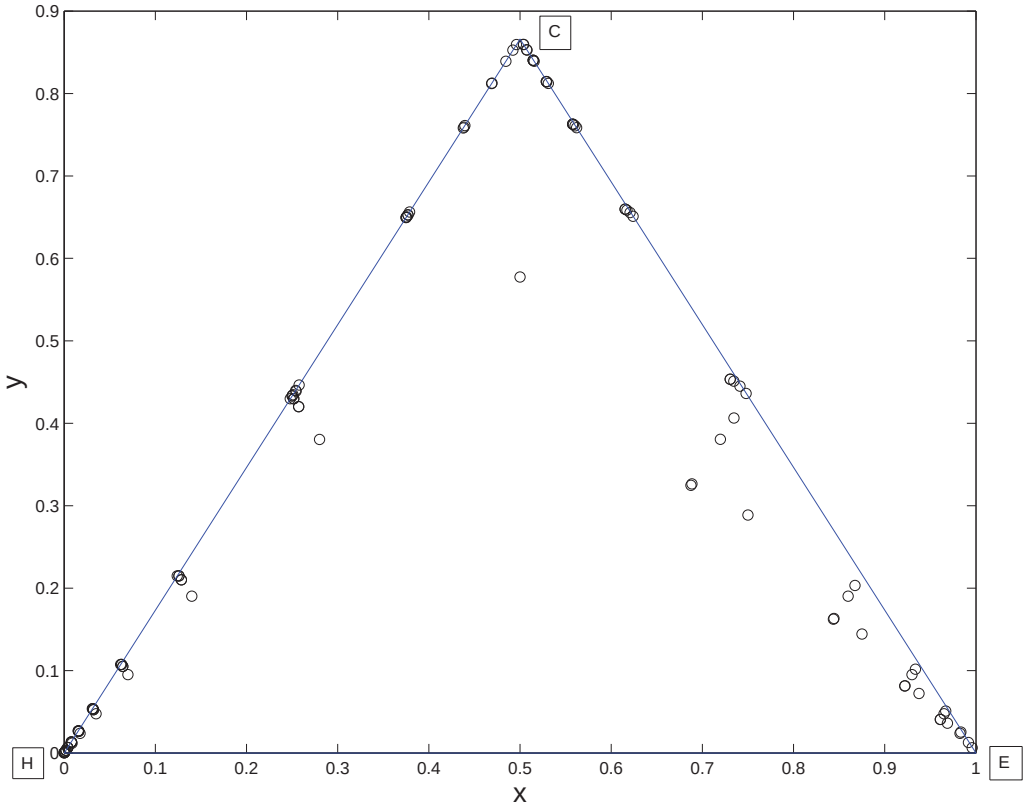



Figure 2: CGR plot of a state sequence.

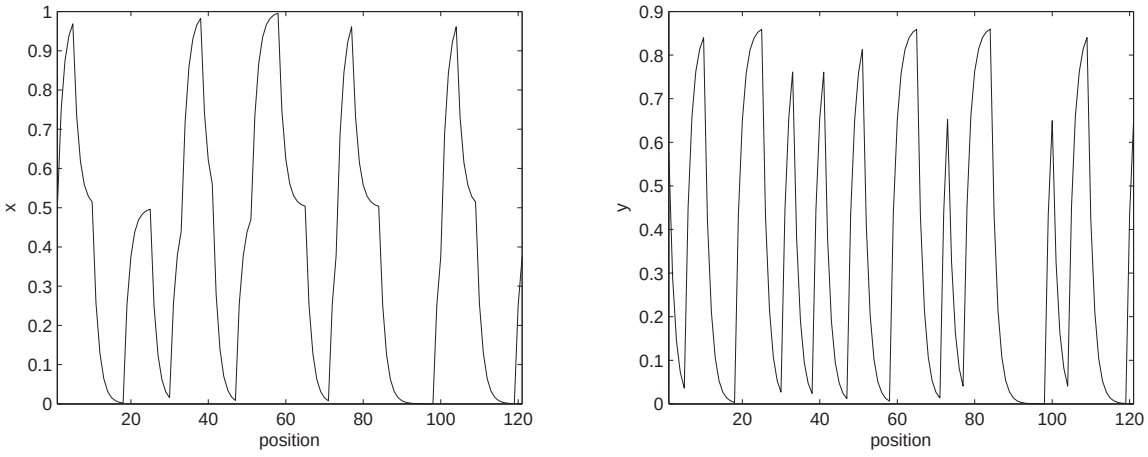



Figure 3: The two time series corresponding to Figure 2.

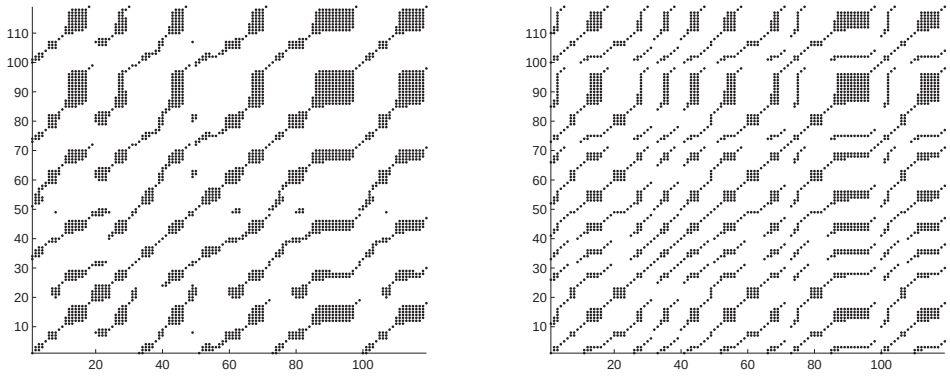



Figure 4: The RP plots of the two time series in Figure 3.

at least three residues, we set $m$ to be 3 in this study. We further construct a *distance matrix* (DM) of size $N_m \times N_m$ from the $N_m$ points. Its elements are the (Euclidean) distances between all pairs of points after scaled down by the maximum distance. As a result, all the element values of DM are located in the interval between 0 and 1, one advantage of which is to allow the recurrence plots in different scales to be statistically compared [7]. Finally, we define a *recurrence matrix* (RM) by applying a *threshold* $\varepsilon$ (namely radius) on the element values of DM. Formally, let RM$=(R_{i,j}(\varepsilon))_{N_m \times N_m}$ and

$$R_{i,j}(\varepsilon) = H(\varepsilon - D_{i,j}), \quad i,j = 1,2,\cdots,N_m \tag{3}$$

where $H$ is the *Heaviside function*; that is,

$$H(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0. \end{cases} \tag{4}$$

RP is simply a visualization of RM by plotting points on $i$-$j$ plane for those elements in RM with values equal to 1. If $R_{i,j}(\varepsilon) = 1$, we say the $j$-th point recurs with reference to the $i$-th point. For any $\varepsilon > 0$, the RP has always a black line along main diagonal since $R_{i,i}(\varepsilon) \equiv 1$. Furthermore, the RP is symmetric with respect to the main diagonal as $R_{i,j}(\varepsilon) = R_{j,i}(\varepsilon)$. For example, the RPs for the two time series in Figure 3 are shown in Figure 4. It can be seen that $\varepsilon$ is a crucial parameter in the construction of a RP. If $\varepsilon$ is chosen too small, then there might leave only a few of recurrence points so that we can not learn any recurrence structure of the underlying time series. But if $\varepsilon$ is too large, almost all the points will be enclosed in the neighbor of a point, thereby introducing a lot of structure artifacts. In this study, $\varepsilon$ is set to be 39% according to [3].

## 4 Recurrence quantification analysis

*Recurrence quantification analysis* (RQA) is a nonlinear technique used to quantify the information supplied by a recurrence plot [8, 9]. In the following we briefly introduce the RQA

techniques, where eight recurrence variables are defined to quantify a RP. These variables will be included into our set of features. Because the RP is symmetric with respect to the main diagonal, the recurrence points considered in the following definitions will only comprise those in the upper triangle of a RP (excluding the main diagonal line as well).

The first recurrence variable is called *recurrence* ($REC$). It is a measure of the density of recurrence points in a RP, taking a value ranging from 0 (when there is no recurrent point) to 1 (when all points are recurrent). That is,

$$REC = \frac{\#\text{ recurrent points in upper triangle}}{N_m(N_m-1)/2}, \tag{5}$$

where # stands for counting the number of points.

The second recurrence variable is called *determinism* ($DET$). It measures the proportion of recurrent points that form diagonal line structures. Before evaluating this variable, we need to set the minimum number of recurrent points that a diagonal line segment requires. The commonly used number is 2, which is used in this study as well. Formally, we define determinism as

$$DET = \frac{\#\text{ recurrent points in diagonal lines}}{\#\text{ recurrent points}}. \tag{6}$$

The third recurrence variable is called *linemax* and denoted by $L_{max}$. It simply represents the length of the longest diagonal line segment in RP, and essentially inversely scales with the largest positive Lyapunov exponent [6]. Note that in general, the longer a time series, the longer diagonal line segments as well. In order to cancel the length influence of the time series (equal to the length of the corresponding amino acid sequence), we normalize the length of the longest diagonal line segment by dividing $N_m$. That is,

$$L_{max} = \frac{\text{length of longest diagonal line in RP}}{N_m}. \tag{7}$$

The fourth recurrence variable is *entropy* ($ENT$), which is the Shannon information entropy of the distribution probability of the length of the diagonal lines. That is,

$$ENT = -\sum_{k=L_{min},\ p(k)\neq 0}^{L_{max}} p(k)\log_2(p(k)), \tag{8}$$

where $L_{min}$ is the minimum length of diagonal lines in RP and

$$p(k) = \frac{\#\text{ diagonal lines of length } k \text{ in RP}}{\#\text{ diagonal lines in RP}}. \tag{9}$$

The fifth recurrence variable is called *trend* ($TND$), which quantifies the stationarity degree of time series. It is calculated as the level that the *local recurrences* of diagonal lines fits their displacements from the main diagonal by least squares regression, where the *local recurrence* of a diagonal line refers to the proportion of points on the diagonal line that are the recurrence points. We would like to emphasize that the variable *recurrence* is defined on the whole upper triangle of RP while the *local recurrence* is instead defined only on a certain diagonal line of RP.

The remaining three variables are defined based on the vertical line structure. The sixth recurrence variable is called *laminarity* ($LAM$). It is analogous to $DET$, but calculated using recurrence points forming vertical line structures. That is,

$$LAM = \frac{\#\text{recurrent points in vertical lines}}{\#\text{recurrent points}}. \tag{10}$$

The seventh variable, called *trapping time* ($TT$), is the *normalized* average length of vertical line structures (i.e., average length of vertical line structures divided by $N_m$). The eighth recurrence variable is the *maximum normalized length of the vertical lines* in RP, which is analogous to the definition of $L_{max}$ and denoted by $V_{max}$.

## References


[1] H J Jeffrey. Chaos game representation of gene structure. *Nucleic Acids Research*, 18:2163–2170, 1990.

[2] J Y Yang, Z L Peng, Z G Yu, R J Zhang, V Anh, and D Wang. Prediction of protein structural classes by recurrence quantification analysis based on chaos game representation. *Journal of Theoretical Biology*, 257:618–626, 2009.

[3] J Y Yang, Z L Peng, and X Chen. Prediction of protein structural classes for low-homology sequences based on predicted secondary structure. *BMC Bioinformatics*, 11:S9, 2010.

[4] A Fiser, G E Tusnády, and I Simon. Chaos game representation of protein structures. *Journal of Molecular Graphics and Modelling*, 12:302–304, 1994.

[5] Z G Yu, V Anh, and K S Lau. Chaos game representation of protein sequences based on the detailed hp model and their multifractal and correlation analyses. *Journal of Theoretical Biology*, 226:341–348, 2004.

[6] J P Eckmann, S O Kamphorst, and D Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 4:973–977, 1987.

[7] M A Riley and G C Van Orden. Tutorials in contemporary nonlinear methods for the behavioral sciences, retrieved march 1, 2005, from `http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp`. 2005.

[8] J P Zbilut and C L Jr Webber. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 171:199–203, 1992.

[9] C L Jr Webber and J P Zbilut. Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76:965–973, 1994.

# Supplementary Material C
# for "Improving taxonomy-based protein fold recognition by using global and local features"

Jian-Yi Yang and Xin Chen
Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore, 637371

## I. SVM parameters optimization

In our experiments, grid search is used to select the optimal $C$ and $\gamma$. The grid is set to be $C = [2^0, 2^1, \ldots, 2^{10}]$ and $\gamma = [2^{-1}, 2^{-2}, \ldots, 2^{-10}]$. For the original DD dataset and the RDD datasets, the $C$ and $\gamma$ in SVM are optimized based on ***standard*** 5-fold cross-validation on the training datasets. Then the independent testing datasets are used to evaluate the accuracies of TAXFOLD. While for the EDD, F95 and F194 datasets, *n*-fold cross-validation is adopted to assess TAXFOLD. In order to select the optimal parameters $C$ and $\gamma$, we have slightly modified the procedure of standard *n*-fold cross-validation as follows. First, a dataset (eg., EDD) is randomly partitioned into *n* subsets of equal size. Second, 80% of (n-1) subsets are used to train SVMs, the remaining 20% of the (*n*-1) subsets are used to find the optimal $C$ and $\gamma$, and the remaining subset (called *validation set*) is used to evaluate the prediction accuracies. The second step is repeated *n* times with each of the *n* subsets used exactly once as the validation set. This *n*-fold cross-validation is called ***adjusted*** *n*-fold cross-validation. In this study, the optimal parameters $C$ and $\gamma$ for the ***EDD, F95, F194*** and ***F710*** datasets are obtained using the adjusted *n*-fold cross-validation.

## II. Optimal values of λ and $l_{max}$

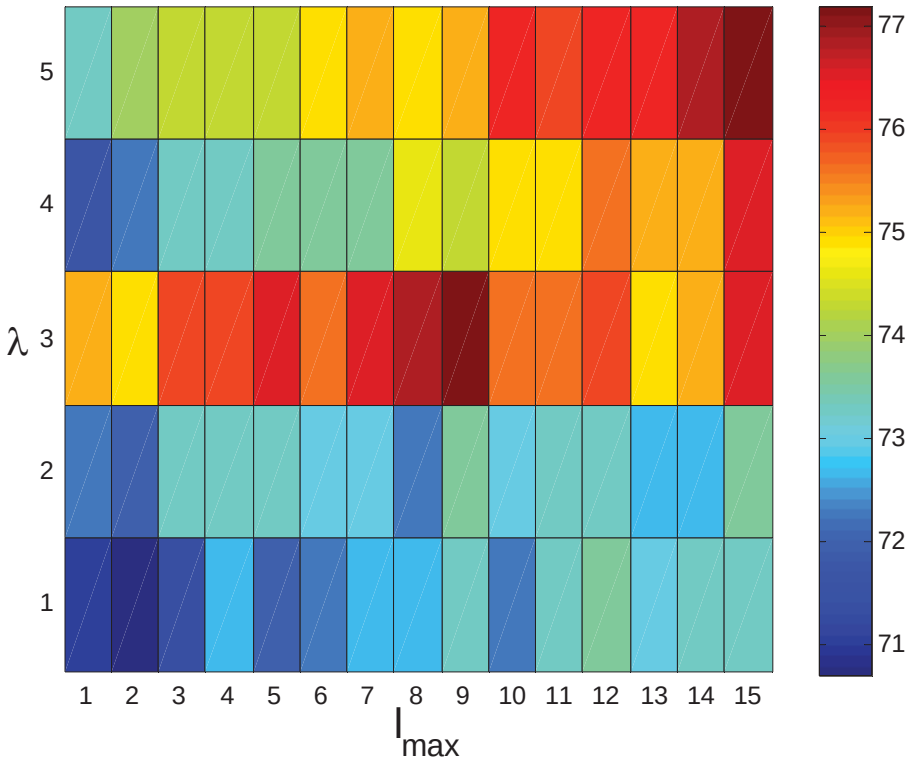



**Figure SC1** The overall accuracies for the RDD training dataset obtained by varying the values of λ and $l_{max}$ from 1 to 5 and 1 to 15, respectively. There are multiple values of λ and $l_{max}$ leading to the same highest accuracy 77.2%. λ =3 and $l_{max}$=9 are selected because they give rise to the smallest number of features while achieving the highest accuracy.

# III. Detailed accuracies of TAXFOLD and other methods

**Table SC1** Performances of TAXFOLD on the **original DD** and **RDD** and **EDD** datasets. W.A. represents weighted average. The accuracies for the original DD (resp. RDD) are evaluated on the independent testing sequences of the DD (resp. RDD) dataset; the accuracies for the EDD dataset are evaluated via *10-fold cross-validation*.

| Fold | Original DD (%) ($C$=4, $\gamma$=0.25) | | | RDD (%) ($C$=4, $\gamma$=0.25) | | | EDD (%) ($C$=4, $\gamma$=0.5) | | |
|---|---|---|---|---|---|---|---|---|---|
| | *recall* | *prec.* | *F-m.* | *recall* | *prec.* | *F-m.* | *recall* | *prec.* | *F-m.* |
| 1 | 100 | 66.7 | 80.0 | 100 | 100 | 100 | 97.6 | 100 | 98.8 |
| 2 | 100 | 100 | 100 | 100 | 100 | 100 | 94.1 | 94.1 | 94.1 |
| 3 | 60.0 | 75.0 | 66.7 | 85.0 | 94.4 | 89.5 | 95.7 | 90.6 | 93.1 |
| 4 | 100 | 57.1 | 72.7 | 100 | 87.5 | 93.3 | 88.4 | 93.8 | 91.0 |
| 5 | 100 | 75.0 | 85.7 | 100 | 90.0 | 94.7 | 86.7 | 83.9 | 85.2 |
| 6 | 66.7 | 100 | 80.0 | 66.7 | 100 | 80.0 | 88.1 | 96.3 | 92.0 |
| 7 | 79.5 | 77.8 | 78.7 | 95.5 | 95.5 | 95.5 | 96.2 | 90.4 | 93.2 |
| 8 | 83.3 | 83.3 | 83.3 | 75.0 | 90.0 | 81.8 | 85.1 | 100 | 92.0 |
| 9 | 92.3 | 85.7 | 88.9 | 92.3 | 80.0 | 85.7 | 90.0 | 84.4 | 87.1 |
| 10 | 83.3 | 100 | 90.9 | 66.7 | 100 | 80.0 | 78.9 | 86.5 | 82.6 |
| 11 | 50.0 | 57.1 | 53.3 | 50.0 | 100 | 66.7 | 79.1 | 81.6 | 80.3 |
| 12 | 73.7 | 43.8 | 54.9 | 89.5 | 58.6 | 70.8 | 72.4 | 71.5 | 72.0 |
| 13 | 100 | 66.7 | 80.0 | 100 | 80.0 | 88.9 | 88.9 | 95.2 | 92.0 |
| 14 | 50.0 | 66.7 | 57.1 | 50.0 | 66.7 | 57.1 | 84.4 | 97.4 | 90.5 |
| 15 | 100 | 87.5 | 93.3 | 100 | 100 | 100 | 75.7 | 100 | 86.2 |
| 16 | 68.8 | 57.9 | 62.9 | 93.8 | 69.2 | 79.6 | 97.3 | 94.0 | 95.6 |
| 17 | 91.7 | 78.6 | 84.6 | 91.7 | 78.6 | 84.6 | 94.8 | 97.9 | 96.3 |
| 18 | 38.5 | 62.5 | 47.6 | 69.2 | 81.8 | 75.0 | 91.8 | 95.7 | 93.7 |
| 19 | 74.1 | 90.9 | 81.6 | 77.8 | 95.5 | 85.7 | 79.2 | 88.8 | 83.7 |
| 20 | 33.3 | 50.0 | 40.0 | 50.0 | 66.7 | 57.1 | 93.7 | 90.7 | 92.2 |
| 21 | 37.5 | 60.0 | 46.2 | 75.0 | 75.0 | 75.0 | 84.7 | 94.0 | 89.1 |
| 22 | 58.3 | 53.8 | 56.0 | 83.3 | 71.4 | 76.9 | 81.3 | 86.7 | 83.9 |
| 23 | 71.4 | 55.6 | 62.5 | 85.7 | 66.7 | 75.0 | 94.0 | 94.0 | 94.0 |
| 24 | 50.0 | 100 | 66.7 | 50.0 | 100 | 66.7 | 100 | 100 | 100 |
| 25 | 25.0 | 100 | 40.0 | 37.5 | 100 | 54.5 | 86.0 | 90.4 | 88.1 |
| 26 | 48.1 | 68.4 | 56.5 | 63.0 | 85.0 | 72.3 | 87.9 | 83.2 | 85.5 |
| 27 | 96.3 | 100 | 98.1 | 100 | 100 | 100 | 99.0 | 96.3 | 97.7 |
| W.A. | 71.5 | 73.8 | 71.0 | 83.2 | 85.5 | 82.9 | 90.0 | 90.1 | 90.0 |

**Table SC2** The performance of TAXFOLD on **F95** dataset evaluated via *10-fold cross-validation*. ($C$=16, $\gamma$=0.5)

| **Fold** | **Recall** (%) | **Precision** (%) | **F-measure** (%) |
|---|---|---|---|
| 1 | 92.7 | 100 | 96.2 |
| 2 | 56.1 | 57.5 | 56.8 |
| 3 | 94.1 | 97.0 | 95.5 |
| 4 | 92.2 | 76.3 | 83.5 |
| 5 | 71.1 | 91.4 | 80.0 |
| 6 | 61.9 | 74.3 | 67.5 |
| 7 | 76.8 | 70.7 | 73.6 |
| 8 | 75.5 | 78.7 | 77.1 |
| 9 | 83.3 | 83.3 | 83.3 |
| 10 | 62.9 | 78.6 | 69.8 |
| 11 | 51.7 | 83.3 | 63.8 |
| 12 | 88.1 | 94.5 | 91.2 |
| 13 | 82.8 | 92.3 | 87.3 |
| 14 | 72.4 | 71.4 | 71.9 |
| 15 | 92.0 | 95.8 | 93.9 |
| 16 | 97.8 | 95.7 | 96.7 |
| 17 | 86.7 | 82.7 | 84.7 |
| 18 | 85.3 | 90.6 | 87.9 |
| 19 | 93.4 | 83.7 | 88.3 |
| 20 | 52.2 | 61.5 | 56.5 |
| 21 | 85.1 | 95.2 | 89.9 |
| 22 | 69.2 | 100 | 81.8 |
| 23 | 88.3 | 75.7 | 81.5 |
| 24 | 73.2 | 69.5 | 71.3 |
| 25 | 68.4 | 83.0 | 75.0 |
| 26 | 79.3 | 76.7 | 78.0 |
| 27 | 78.3 | 71.1 | 74.5 |
| 28 | 96.5 | 100 | 98.2 |
| 29 | 65.4 | 59.6 | 62.4 |
| 30 | 84.4 | 88.4 | 86.4 |
| 31 | 62.8 | 81.8 | 71.1 |
| 32 | 77.4 | 92.3 | 84.2 |
| 33 | 86.7 | 97.5 | 91.8 |
| 34 | 82.1 | 79.0 | 80.5 |
| 35 | 75.7 | 100 | 86.2 |
| 36 | 82.1 | 71.9 | 76.7 |
| 37 | 81.1 | 88.2 | 84.5 |
| 38 | 44.7 | 56.7 | 50.0 |
| 39 | 69.0 | 95.2 | 80.0 |
| 40 | 84.2 | 78.4 | 81.2 |
| 41 | 71.4 | 87.0 | 78.4 |
| 42 | 50.0 | 72.7 | 59.3 |
| 43 | 94.3 | 87.3 | 90.7 |
| 44 | 94.3 | 94.3 | 94.3 |
| 45 | 89.0 | 89.0 | 89.0 |
| 46 | 100 | 100 | 100 |
| 47 | 82.1 | 92.0 | 86.8 |
| 48 | 76.2 | 83.2 | 79.5 |

| | | | |
|---|---|---|---|
| 49 | 72.6 | 81.8 | 76.9 |
| 50 | 89.3 | 92.6 | 90.9 |
| 51 | 82.1 | 95.8 | 88.5 |
| 52 | 90.0 | 100 | 94.7 |
| 53 | 90.8 | 81.9 | 86.1 |
| 54 | 84.7 | 91.3 | 87.9 |
| 55 | 65.4 | 89.5 | 75.6 |
| 56 | 48.9 | 47.8 | 48.4 |
| 57 | 72.7 | 68.9 | 70.7 |
| 58 | 81.8 | 88.2 | 84.9 |
| 59 | 92.6 | 96.2 | 94.3 |
| 60 | 89.3 | 82.1 | 85.6 |
| 61 | 92.3 | 93.8 | 93.0 |
| 62 | 65.6 | 87.5 | 75.0 |
| 63 | 94.0 | 92.9 | 93.4 |
| 64 | 86.7 | 96.3 | 91.2 |
| 65 | 91.5 | 94.7 | 93.1 |
| 66 | 72.7 | 69.6 | 71.1 |
| 67 | 53.7 | 78.6 | 63.8 |
| 68 | 81.8 | 79.2 | 80.5 |
| 69 | 80.5 | 86.4 | 83.3 |
| 70 | 84.0 | 100 | 91.3 |
| 71 | 90.6 | 100 | 95.1 |
| 72 | 90.2 | 100 | 94.8 |
| 73 | 46.2 | 85.7 | 60.0 |
| 74 | 85.5 | 71.3 | 77.7 |
| 75 | 70.3 | 86.7 | 77.6 |
| 76 | 68.6 | 77.4 | 72.7 |
| 77 | 63.9 | 85.2 | 73.0 |
| 78 | 82.1 | 92.0 | 86.8 |
| 79 | 85.7 | 87.0 | 86.3 |
| 80 | 77.1 | 82.2 | 79.6 |
| 81 | 86.7 | 92.9 | 89.7 |
| 82 | 75.9 | 89.1 | 82.0 |
| 83 | 86.7 | 92.9 | 89.7 |
| 84 | 91.7 | 91.7 | 91.7 |
| 85 | 84.2 | 91.4 | 87.7 |
| 86 | 96.4 | 96.4 | 96.4 |
| 87 | 91.7 | 97.1 | 94.3 |
| 88 | 93.3 | 96.6 | 94.9 |
| 89 | 88.6 | 95.1 | 91.8 |
| 90 | 96.2 | 78.3 | 86.3 |
| 91 | 89.3 | 96.2 | 92.6 |
| 92 | 50.0 | 76.9 | 60.6 |
| 93 | 61.4 | 72.9 | 66.7 |
| 94 | 56.6 | 76.9 | 65.2 |
| 95 | 66.7 | 90.0 | 76.6 |
| W.A. | 82.4 | 82.9 | 82.3 |

**Table SC3** The performance of TAXFOLD on **F194** dataset evaluated via *10-fold cross-validation*. ($C$=16, $\gamma$=0.5)

| **Fold** | **Recall** (%) | **Precision** (%) | **F-measure** (%) |
|---|---|---|---|
| 1 | 92.7 | 97.4 | 95.0 |
| 2 | 51.2 | 44.7 | 47.7 |
| 3 | 20.0 | 50.0 | 28.6 |
| 4 | 91.2 | 93.9 | 92.5 |
| 5 | 89.8 | 66.3 | 76.3 |
| 6 | 68.9 | 77.5 | 72.9 |
| 7 | 42.9 | 100 | 60.0 |
| 8 | 66.7 | 68.3 | 67.5 |
| 9 | 30.4 | 63.6 | 41.2 |
| 10 | 52.4 | 100 | 68.8 |
| 11 | 63.8 | 55.7 | 59.5 |
| 12 | 81.6 | 78.4 | 80.0 |
| 13 | 83.3 | 78.1 | 80.6 |
| 14 | 53.8 | 87.5 | 66.7 |
| 15 | 60.0 | 70.0 | 64.6 |
| 16 | 44.8 | 72.2 | 55.3 |
| 17 | 83.1 | 89.1 | 86.0 |
| 18 | 66.7 | 88.9 | 76.2 |
| 19 | 25.0 | 50.0 | 33.3 |
| 20 | 82.8 | 88.9 | 85.7 |
| 21 | 36.4 | 80.0 | 50.0 |
| 22 | 72.4 | 63.2 | 67.5 |
| 23 | 78.3 | 81.8 | 80.0 |
| 24 | 61.5 | 88.9 | 72.7 |
| 25 | 52.9 | 100 | 69.2 |
| 26 | 73.7 | 93.3 | 82.4 |
| 27 | 80.0 | 90.9 | 85.1 |
| 28 | 95.6 | 89.6 | 92.5 |
| 29 | 100 | 100 | 100 |
| 30 | 86.7 | 76.5 | 81.3 |
| 31 | 85.3 | 85.3 | 85.3 |
| 32 | 100 | 100 | 100 |
| 33 | 100 | 78.6 | 88.0 |
| 34 | 84.6 | 100 | 91.7 |
| 35 | 14.3 | 100 | 25.0 |
| 36 | 86.4 | 79.2 | 82.6 |
| 37 | 92.8 | 81.0 | 86.5 |
| 38 | 52.2 | 57.1 | 54.5 |
| 39 | 61.9 | 100 | 76.5 |
| 40 | 85.1 | 93.0 | 88.9 |
| 41 | 73.1 | 100 | 84.4 |
| 42 | 88.3 | 74.6 | 80.9 |
| 43 | 58.3 | 100 | 73.7 |
| 44 | 78.6 | 66.7 | 72.1 |
| 45 | 92.9 | 100 | 96.3 |
| 46 | 62.5 | 100 | 76.9 |
| 47 | 71.9 | 73.2 | 72.6 |
| 48 | 72.4 | 72.4 | 72.4 |

| | | | |
|---|---|---|---|
| 49 | 100 | 100 | 100 |
| 50 | 79.1 | 68.0 | 73.1 |
| 51 | 85.0 | 100 | 91.9 |
| 52 | 96.5 | 100 | 98.2 |
| 53 | 63.6 | 93.3 | 75.7 |
| 54 | 67.3 | 53.0 | 59.3 |
| 55 | 86.7 | 79.6 | 83.0 |
| 56 | 62.8 | 77.1 | 69.2 |
| 57 | 36.4 | 100 | 53.3 |
| 58 | 74.2 | 85.2 | 79.3 |
| 59 | 53.8 | 77.8 | 63.6 |
| 60 | 84.4 | 90.5 | 87.4 |
| 61 | 66.7 | 100 | 80.0 |
| 62 | 94.1 | 100 | 97.0 |
| 63 | 69.6 | 76.2 | 72.7 |
| 64 | 79.5 | 72.9 | 76.1 |
| 65 | 75.7 | 93.3 | 83.6 |
| 66 | 52.9 | 100 | 69.2 |
| 67 | 82.1 | 69.7 | 75.4 |
| 68 | 78.4 | 87.9 | 82.9 |
| 69 | 50.0 | 65.5 | 56.7 |
| 70 | 69.0 | 90.9 | 78.4 |
| 71 | 66.7 | 92.3 | 77.4 |
| 72 | 85.3 | 82.7 | 83.9 |
| 73 | 56.5 | 92.9 | 70.3 |
| 74 | 75.0 | 85.7 | 80.0 |
| 75 | 67.9 | 86.4 | 76.0 |
| 76 | 56.3 | 72.0 | 63.2 |
| 77 | 95.8 | 84.1 | 89.6 |
| 78 | 94.3 | 91.5 | 92.9 |
| 79 | 89.0 | 89.0 | 89.0 |
| 80 | 70.0 | 93.3 | 80.0 |
| 81 | 55.0 | 57.9 | 56.4 |
| 82 | 100 | 100 | 100 |
| 83 | 85.7 | 88.9 | 87.3 |
| 84 | 64.3 | 100 | 78.3 |
| 85 | 78.5 | 72.3 | 75.3 |
| 86 | 84.6 | 100 | 91.7 |
| 87 | 69.4 | 79.6 | 74.1 |
| 88 | 81.8 | 90.0 | 85.7 |
| 89 | 77.8 | 100 | 87.5 |
| 90 | 25.0 | 57.1 | 34.8 |
| 91 | 78.9 | 100 | 88.2 |
| 92 | 89.3 | 92.6 | 90.9 |
| 93 | 75.0 | 84.0 | 79.2 |
| 94 | 86.7 | 100 | 92.9 |
| 95 | 89.1 | 77.7 | 83.0 |
| 96 | 72.7 | 100 | 84.2 |
| 97 | 75.0 | 88.2 | 81.1 |
| 98 | 95.2 | 100 | 97.6 |
| 99 | 83.8 | 86.9 | 85.3 |

| | | | |
|---|---|---|---|
| 100 | 69.2 | 90.0 | 78.3 |
| 101 | 55.6 | 49.0 | 52.1 |
| 102 | 73.4 | 57.3 | 64.4 |
| 103 | 87.3 | 88.9 | 88.1 |
| 104 | 88.9 | 96.0 | 92.3 |
| 105 | 84.6 | 100 | 91.7 |
| 106 | 86.4 | 74.2 | 79.8 |
| 107 | 92.3 | 93.8 | 93.0 |
| 108 | 59.4 | 82.6 | 69.1 |
| 109 | 94.0 | 89.7 | 91.8 |
| 110 | 100 | 100 | 100 |
| 111 | 93.3 | 100 | 96.6 |
| 112 | 58.3 | 87.5 | 70.0 |
| 113 | 100 | 100 | 100 |
| 114 | 41.2 | 87.5 | 56.0 |
| 115 | 83.3 | 100 | 90.9 |
| 116 | 93.8 | 100 | 96.8 |
| 117 | 91.5 | 94.7 | 93.1 |
| 118 | 87.5 | 87.5 | 87.5 |
| 119 | 70.6 | 100 | 82.8 |
| 120 | 78.9 | 88.2 | 83.3 |
| 121 | 65.9 | 65.9 | 65.9 |
| 122 | 80.0 | 100 | 88.9 |
| 123 | 72.7 | 100 | 84.2 |
| 124 | 60.0 | 85.7 | 70.6 |
| 125 | 53.7 | 73.3 | 62.0 |
| 126 | 84.3 | 77.9 | 81.0 |
| 127 | 54.2 | 86.7 | 66.7 |
| 128 | 79.3 | 76.7 | 78.0 |
| 129 | 66.7 | 100 | 80.0 |
| 130 | 90.5 | 100 | 95.0 |
| 131 | 75.0 | 100 | 85.7 |
| 132 | 66.7 | 100 | 80.0 |
| 133 | 80.0 | 100 | 88.9 |
| 134 | 93.8 | 100 | 96.8 |
| 135 | 88.9 | 100 | 94.1 |
| 136 | 95.8 | 100 | 97.9 |
| 137 | 90.2 | 100 | 94.8 |
| 138 | 81.0 | 100 | 89.5 |
| 139 | 54.5 | 92.3 | 68.6 |
| 140 | 79.2 | 100 | 88.4 |
| 141 | 46.2 | 85.7 | 60.0 |
| 142 | 69.6 | 88.9 | 78.0 |
| 143 | 81.3 | 100 | 89.7 |
| 144 | 84.1 | 65.7 | 73.7 |
| 145 | 91.7 | 100 | 95.7 |
| 146 | 95.2 | 100 | 97.6 |
| 147 | 55.0 | 91.7 | 68.7 |
| 148 | 46.2 | 100 | 63.2 |
| 149 | 70.3 | 86.7 | 77.6 |
| 150 | 63.6 | 100 | 77.8 |

| | | | |
|---|---|---|---|
| 151 | 68.6 | 77.4 | 72.7 |
| 152 | 66.7 | 100 | 80.0 |
| 153 | 80.6 | 80.6 | 80.6 |
| 154 | 87.0 | 95.2 | 90.9 |
| 155 | 46.2 | 85.7 | 60.0 |
| 156 | 71.4 | 83.3 | 76.9 |
| 157 | 83.3 | 100 | 90.9 |
| 158 | 82.1 | 85.2 | 83.6 |
| 159 | 88.6 | 86.1 | 87.3 |
| 160 | 76.5 | 100 | 86.7 |
| 161 | 75.0 | 76.6 | 75.8 |
| 162 | 86.7 | 92.9 | 89.7 |
| 163 | 81.3 | 100 | 89.7 |
| 164 | 83.3 | 90.9 | 87.0 |
| 165 | 75.9 | 73.2 | 74.5 |
| 166 | 77.3 | 94.4 | 85.0 |
| 167 | 54.5 | 85.7 | 66.7 |
| 168 | 83.3 | 89.3 | 86.2 |
| 169 | 91.7 | 91.7 | 91.7 |
| 170 | 85.7 | 85.7 | 85.7 |
| 171 | 81.6 | 100 | 89.9 |
| 172 | 85.7 | 96.0 | 90.6 |
| 173 | 91.3 | 91.3 | 91.3 |
| 174 | 93.8 | 100 | 96.8 |
| 175 | 37.5 | 66.7 | 48.0 |
| 176 | 86.1 | 93.9 | 89.9 |
| 177 | 100 | 100 | 100 |
| 178 | 82.6 | 95.0 | 88.4 |
| 179 | 100 | 100 | 100 |
| 180 | 81.3 | 92.9 | 86.7 |
| 181 | 88.6 | 90.7 | 89.7 |
| 182 | 86.4 | 95.0 | 90.5 |
| 183 | 94.3 | 63.5 | 75.9 |
| 184 | 58.8 | 83.3 | 69.0 |
| 185 | 21.4 | 100 | 35.3 |
| 186 | 63.6 | 100 | 77.8 |
| 187 | 50.0 | 85.7 | 63.2 |
| 188 | 85.7 | 96.0 | 90.6 |
| 189 | 13.3 | 66.7 | 22.2 |
| 190 | 57.5 | 74.2 | 64.8 |
| 191 | 64.9 | 68.5 | 66.7 |
| 192 | 54.7 | 64.4 | 59.2 |
| 193 | 66.7 | 75.0 | 70.6 |
| 194 | 35.7 | 71.4 | 47.6 |
| W.A. | 79.6 | 81.1 | 79.4 |

**Table SC4** Comparison with the major existing methods on the testing sequences of the **RDD** dataset. For each of the *Recall*, *Precision*, and *F-measure* columns, the five sub-columns from left to right are the accuracies for the Shamim, ACCFold_AC, ACCFold_ACC, PFRES, and TAXFOLD, respectively.

| **Fold** | **Recall** (%) | | | | | **Precision** (%) | | | | | **F-measure** (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 100 | 100 | 100 | 100 | 100 | 54.5 | 100 | 100 | 100 | 100 | 70.6 | 100 | 100 | 100 | 100 |
| 2 | 66.7 | 77.8 | 100 | 100 | 100 | 100 | 100 | 100 | 90.0 | 100 | 80.0 | 87.5 | 100 | 94.7 | 100 |
| 3 | 60.0 | 60.0 | 60.0 | 75.0 | 85.0 | 80.0 | 75.0 | 85.7 | 88.2 | 94.4 | 68.6 | 66.7 | 70.6 | 81.1 | 89.5 |
| 4 | 42.9 | 100 | 100 | 100 | 100 | 75.0 | 87.5 | 100 | 87.5 | 87.5 | 54.5 | 93.3 | 100 | 93.3 | 93.3 |
| 5 | 88.9 | 88.9 | 100 | 88.9 | 100 | 88.9 | 80.0 | 100 | 80.0 | 90.0 | 88.9 | 84.2 | 100 | 84.2 | 94.7 |
| 6 | 44.4 | 55.6 | 55.6 | 66.7 | 66.7 | 80.0 | 100 | 100 | 100 | 100 | 57.1 | 71.4 | 71.4 | 80.0 | 80.0 |
| 7 | 93.2 | 97.7 | 97.7 | 93.2 | 95.5 | 77.4 | 86.0 | 91.5 | 87.2 | 95.5 | 84.5 | 91.5 | 94.5 | 90.1 | 95.5 |
| 8 | 50.0 | 66.7 | 66.7 | 25.0 | 75.0 | 66.7 | 100 | 100 | 100 | 90.0 | 57.1 | 80.0 | 80.0 | 40.0 | 81.8 |
| 9 | 76.9 | 92.3 | 100 | 84.6 | 92.3 | 83.3 | 52.2 | 33.3 | 100 | 80.0 | 80.0 | 66.7 | 50.0 | 91.7 | 85.7 |
| 10 | 33.3 | 66.7 | 66.7 | 83.3 | 66.7 | 66.7 | 80.0 | 100 | 83.3 | 100 | 44.4 | 72.7 | 80.0 | 83.3 | 80.0 |
| 11 | 37.5 | 37.5 | 37.5 | 50.0 | 50.0 | 100 | 100 | 100 | 100 | 100 | 54.5 | 54.5 | 54.5 | 66.7 | 66.7 |
| 12 | 36.8 | 36.8 | 52.6 | 73.7 | 89.5 | 41.2 | 41.2 | 52.6 | 63.6 | 58.6 | 38.9 | 38.9 | 52.6 | 68.3 | 70.8 |
| 13 | 75.0 | 100 | 100 | 100 | 100 | 37.5 | 66.7 | 66.7 | 66.7 | 80.0 | 50.0 | 80.0 | 80.0 | 80.0 | 88.9 |
| 14 | 50.0 | 50.0 | 50.0 | 50.0 | 50.0 | 66.7 | 100 | 100 | 66.7 | 66.7 | 57.1 | 66.7 | 66.7 | 57.1 | 57.1 |
| 15 | 85.7 | 85.7 | 100 | 100 | 100 | 54.5 | 100 | 100 | 63.6 | 100 | 66.7 | 92.3 | 100 | 77.8 | 100 |
| 16 | 89.6 | 79.2 | 89.6 | 85.4 | 93.8 | 61.4 | 61.3 | 58.9 | 85.4 | 69.2 | 72.9 | 69.1 | 71.1 | 85.4 | 79.6 |
| 17 | 83.3 | 91.7 | 91.7 | 83.3 | 91.7 | 71.4 | 68.8 | 91.7 | 71.4 | 78.6 | 76.9 | 78.6 | 91.7 | 76.9 | 84.6 |
| 18 | 38.5 | 76.9 | 69.2 | 61.5 | 69.2 | 35.7 | 66.7 | 75.0 | 72.7 | 81.8 | 37.0 | 71.4 | 72.0 | 66.7 | 75.0 |
| 19 | 51.9 | 66.7 | 70.4 | 77.8 | 77.8 | 63.6 | 94.7 | 95.0 | 80.8 | 95.5 | 57.1 | 78.3 | 80.9 | 79.2 | 85.7 |
| 20 | 41.7 | 41.7 | 41.7 | 66.7 | 50.0 | 45.5 | 71.4 | 71.4 | 61.5 | 66.7 | 43.5 | 52.6 | 52.6 | 64.0 | 57.1 |
| 21 | 37.5 | 75.0 | 62.5 | 75.0 | 75.0 | 50.0 | 85.7 | 100 | 60.0 | 75.0 | 42.9 | 80.0 | 76.9 | 66.7 | 75.0 |
| 22 | 75.0 | 66.7 | 58.3 | 100 | 83.3 | 60.0 | 72.7 | 63.6 | 66.7 | 71.4 | 66.7 | 69.6 | 60.9 | 80.0 | 76.9 |
| 23 | 71.4 | 100 | 100 | 85.7 | 85.7 | 55.6 | 53.8 | 58.3 | 54.5 | 66.7 | 62.5 | 70.0 | 73.7 | 66.7 | 75.0 |
| 24 | 25.0 | 75.0 | 50.0 | 50.0 | 50.0 | 33.3 | 60.0 | 66.7 | 50.0 | 100 | 28.6 | 66.7 | 57.1 | 50.0 | 66.7 |
| 25 | 25.0 | 25.0 | 37.5 | 37.5 | 37.5 | 66.7 | 50.0 | 50.0 | 60.0 | 100 | 36.4 | 33.3 | 42.9 | 46.2 | 54.5 |
| 26 | 37.0 | 66.7 | 44.4 | 74.1 | 63.0 | 55.6 | 62.1 | 63.2 | 80.0 | 85.0 | 44.4 | 64.3 | 52.2 | 76.9 | 72.3 |
| 27 | 100 | 77.8 | 63.0 | 100 | 100 | 96.4 | 95.5 | 100 | 100 | 100 | 98.2 | 85.7 | 77.3 | 100 | 100 |
| W.A. | 66.2 | 73.6 | 73.8 | 80.1 | 83.2 | 67.4 | 76.4 | 79.9 | 81.8 | 85.5 | 64.9 | 73.2 | 74.1 | 79.5 | 82.9 |

**Table SC5** Comparison with the major existing methods via *2-fold cross-validation* on the **EDD** dataset. For each of the *Recall*, *Precision*, and *F-measure* columns, the five sub-columns from left to right are the accuracies for the Shamim, ACCFold_AC, ACCFold_ACC, PFRES, and TAXFOLD, respectively.

| **Fold** | **Recall** (%) | | | | | **Precision** (%) | | | | | **F-measure** (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 56.1 | 95.1 | 92.7 | 92.7 | 97.6 | 71.9 | 97.5 | 100 | 100 | 100 | 63.0 | 96.3 | 96.2 | 96.2 | 98.8 |
| 2 | 73.5 | 58.8 | 55.9 | 91.2 | 91.2 | 83.3 | 100 | 100 | 96.9 | 96.9 | 78.1 | 74.1 | 71.7 | 93.9 | 93.9 |
| 3 | 82.9 | 78.9 | 88.8 | 93.5 | 94.1 | 79.2 | 71.3 | 80.1 | 86.5 | 89.4 | 81.0 | 74.9 | 84.2 | 89.9 | 91.7 |
| 4 | 62.3 | 78.3 | 81.2 | 82.6 | 84.1 | 64.2 | 80.6 | 88.9 | 90.5 | 90.6 | 63.2 | 79.4 | 84.8 | 86.4 | 87.2 |
| 5 | 56.7 | 66.7 | 70.0 | 83.3 | 86.7 | 73.9 | 95.2 | 100 | 89.3 | 81.3 | 64.2 | 78.4 | 82.4 | 86.2 | 83.9 |
| 6 | 69.5 | 71.2 | 66.1 | 86.4 | 84.7 | 82.0 | 100 | 100 | 87.9 | 98.0 | 75.2 | 83.2 | 79.6 | 87.2 | 90.9 |
| 7 | 86.2 | 90.8 | 95.7 | 93.4 | 93.9 | 72.5 | 78.4 | 77.1 | 87.1 | 90.2 | 78.7 | 84.1 | 85.4 | 90.1 | 92.0 |
| 8 | 23.4 | 57.4 | 46.8 | 63.8 | 76.6 | 44.0 | 96.4 | 100 | 78.9 | 94.7 | 30.6 | 72.0 | 63.8 | 70.6 | 84.7 |
| 9 | 50.0 | 60.0 | 73.3 | 68.3 | 91.7 | 53.6 | 48.0 | 52.4 | 75.9 | 82.1 | 51.7 | 53.3 | 61.1 | 71.9 | 86.6 |
| 10 | 49.1 | 59.6 | 52.6 | 78.9 | 75.4 | 50.9 | 87.2 | 93.8 | 76.3 | 79.6 | 50.0 | 70.8 | 67.4 | 77.6 | 77.5 |
| 11 | 41.1 | 51.9 | 45.7 | 67.4 | 69.8 | 52.0 | 64.4 | 89.4 | 72.5 | 77.6 | 45.9 | 57.5 | 60.5 | 69.9 | 73.5 |
| 12 | 44.2 | 48.7 | 55.1 | 59.0 | 71.8 | 44.8 | 37.1 | 59.3 | 63.9 | 69.6 | 44.5 | 42.1 | 57.1 | 61.3 | 70.7 |
| 13 | 33.3 | 66.7 | 73.3 | 64.4 | 80.0 | 55.6 | 83.3 | 100 | 82.9 | 94.7 | 41.7 | 74.1 | 84.6 | 72.5 | 86.7 |
| 14 | 73.3 | 77.8 | 75.6 | 80.0 | 82.2 | 86.8 | 100 | 100 | 90.0 | 97.4 | 79.5 | 87.5 | 86.1 | 84.7 | 89.2 |
| 15 | 43.2 | 64.9 | 73.0 | 75.7 | 75.7 | 69.6 | 96.0 | 100 | 90.3 | 100 | 53.3 | 77.4 | 84.4 | 82.4 | 86.2 |
| 16 | 80.4 | 93.2 | 97.0 | 91.1 | 95.5 | 60.8 | 73.8 | 86.5 | 81.4 | 88.4 | 69.2 | 82.4 | 91.4 | 86.0 | 91.8 |
| 17 | 56.7 | 85.1 | 89.7 | 83.0 | 89.7 | 53.1 | 86.4 | 92.6 | 82.6 | 96.7 | 54.9 | 85.7 | 91.1 | 82.8 | 93.0 |
| 18 | 41.1 | 80.8 | 83.6 | 75.3 | 94.5 | 52.6 | 93.7 | 96.8 | 83.3 | 94.5 | 46.2 | 86.8 | 89.7 | 79.1 | 94.5 |
| 19 | 16.9 | 63.8 | 66.2 | 59.2 | 74.6 | 29.3 | 80.6 | 96.6 | 74.8 | 88.2 | 21.5 | 71.2 | 78.5 | 66.1 | 80.8 |
| 20 | 66.1 | 77.8 | 87.9 | 82.4 | 91.2 | 49.5 | 68.1 | 70.9 | 80.7 | 88.3 | 56.6 | 72.7 | 78.5 | 81.6 | 89.7 |
| 21 | 38.7 | 52.3 | 52.3 | 77.5 | 82.0 | 59.7 | 87.9 | 98.3 | 79.6 | 90.1 | 47.0 | 65.5 | 68.2 | 78.5 | 85.8 |
| 22 | 19.5 | 35.9 | 46.9 | 52.3 | 65.6 | 35.2 | 74.2 | 90.9 | 68.4 | 83.2 | 25.1 | 48.4 | 61.9 | 59.3 | 73.4 |
| 23 | 41.0 | 80.7 | 84.3 | 84.3 | 90.4 | 54.0 | 91.8 | 97.2 | 83.3 | 90.4 | 46.6 | 85.9 | 90.3 | 83.8 | 90.4 |
| 24 | 12.5 | 62.5 | 68.8 | 43.8 | 81.3 | 100 | 100 | 100 | 87.5 | 100 | 22.2 | 76.9 | 81.5 | 58.3 | 89.7 |
| 25 | 41.3 | 54.5 | 46.3 | 71.9 | 81.8 | 54.3 | 66.0 | 90.3 | 74.4 | 88.4 | 46.9 | 59.7 | 61.2 | 73.1 | 85.0 |
| 26 | 64.6 | 76.4 | 88.8 | 82.9 | 86.7 | 54.3 | 66.4 | 49.9 | 73.2 | 73.9 | 59.0 | 71.1 | 63.9 | 77.7 | 79.8 |
| 27 | 97.1 | 82.9 | 43.8 | 100 | 100 | 94.4 | 90.6 | 100 | 98.1 | 94.6 | 95.8 | 86.6 | 60.9 | 99.1 | 97.2 |
| W.A. | 61.0 | 73.9 | 77.3 | 81.1 | 86.9 | 60.2 | 75.7 | 82.2 | 80.9 | 87.2 | 59.6 | 73.7 | 77.0 | 80.8 | 86.8 |

**Table SC6** Comparison with the major existing methods via *2-fold cross-validation* on the **F95** dataset. For each of the *Recall*, *Precision*, and *F-measure* columns, the five sub-columns from left to right are the accuracies for the Shamim, ACCFold_AC, ACCFold_ACC, PFRES, and TAXFOLD, respectively.

| **Fold** | **Recall** (%) | | | | | **Precision** (%) | | | | | **F-measure** (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 43.9 | 85.4 | 90.2 | 75.6 | 92.7 | 56.3 | 97.2 | 100 | 86.1 | 100 | 49.3 | 90.9 | 94.9 | 80.5 | 96.2 |
| 2 | 24.4 | 36.6 | 36.6 | 61.0 | 56.1 | 52.6 | 75.0 | 88.2 | 64.1 | 57.5 | 33.3 | 49.2 | 51.7 | 62.5 | 56.8 |
| 3 | 76.5 | 52.9 | 70.6 | 91.2 | 94.1 | 74.3 | 100 | 100 | 88.6 | 97.0 | 75.4 | 69.2 | 82.8 | 89.9 | 95.5 |
| 4 | 74.2 | 75.2 | 85.4 | 89.4 | 92.2 | 51.1 | 52.4 | 63.2 | 66.8 | 76.3 | 60.5 | 61.7 | 72.7 | 76.5 | 83.5 |
| 5 | 28.9 | 33.3 | 37.8 | 44.4 | 71.1 | 44.8 | 78.9 | 100 | 80.0 | 91.4 | 35.1 | 46.9 | 54.8 | 57.1 | 80.0 |
| 6 | 19.0 | 40.5 | 45.2 | 47.6 | 61.9 | 30.8 | 65.4 | 65.5 | 74.1 | 74.3 | 23.5 | 50.0 | 53.5 | 58.0 | 67.5 |
| 7 | 26.1 | 68.1 | 78.3 | 66.7 | 76.8 | 26.9 | 54.0 | 58.1 | 58.2 | 70.7 | 26.5 | 60.3 | 66.7 | 62.2 | 73.6 |
| 8 | 42.9 | 63.3 | 67.3 | 71.4 | 75.5 | 46.7 | 86.1 | 94.3 | 72.9 | 78.7 | 44.7 | 72.9 | 78.6 | 72.2 | 77.1 |
| 9 | 46.7 | 63.3 | 73.3 | 86.7 | 83.3 | 51.9 | 57.6 | 84.6 | 72.2 | 83.3 | 49.1 | 60.3 | 78.6 | 78.8 | 83.3 |
| 10 | 11.4 | 60.0 | 60.0 | 48.6 | 62.9 | 17.4 | 95.5 | 91.3 | 60.7 | 78.6 | 13.8 | 73.7 | 72.4 | 54.0 | 69.8 |
| 11 | 0.0 | 34.5 | 34.5 | 13.8 | 51.7 | 0.0 | 83.3 | 100 | 50.0 | 83.3 | 0.0 | 48.8 | 51.3 | 21.6 | 63.8 |
| 12 | 52.5 | 62.7 | 67.8 | 76.3 | 88.1 | 64.6 | 94.9 | 100 | 81.8 | 94.5 | 57.9 | 75.5 | 80.8 | 78.9 | 91.2 |
| 13 | 24.1 | 79.3 | 75.9 | 75.9 | 82.8 | 36.8 | 100 | 100 | 75.9 | 92.3 | 29.2 | 88.5 | 86.3 | 75.9 | 87.3 |
| 14 | 23.7 | 53.9 | 67.1 | 47.4 | 72.4 | 38.3 | 65.1 | 89.5 | 53.7 | 71.4 | 29.3 | 59.0 | 76.7 | 50.3 | 71.9 |
| 15 | 4.0 | 56.0 | 52.0 | 64.0 | 92.0 | 12.5 | 82.4 | 92.9 | 80.0 | 95.8 | 6.1 | 66.7 | 66.7 | 71.1 | 93.9 |
| 16 | 75.6 | 84.4 | 91.1 | 97.8 | 97.8 | 79.1 | 76.0 | 95.3 | 78.6 | 95.7 | 77.3 | 80.0 | 93.2 | 87.1 | 96.7 |
| 17 | 66.7 | 64.8 | 74.3 | 76.2 | 86.7 | 43.8 | 54.4 | 66.7 | 85.1 | 82.7 | 52.8 | 59.1 | 70.3 | 80.4 | 84.7 |
| 18 | 50.0 | 64.7 | 64.7 | 85.3 | 85.3 | 45.9 | 73.3 | 81.5 | 67.4 | 90.6 | 47.9 | 68.8 | 72.1 | 75.3 | 87.9 |
| 19 | 83.1 | 87.7 | 92.3 | 89.8 | 93.4 | 55.6 | 65.0 | 69.2 | 80.0 | 83.7 | 66.6 | 74.6 | 79.1 | 84.6 | 88.3 |
| 20 | 15.2 | 32.6 | 37.0 | 32.6 | 52.2 | 21.9 | 50.0 | 53.1 | 57.7 | 61.5 | 17.9 | 39.5 | 43.6 | 41.7 | 56.5 |
| 21 | 31.9 | 66.0 | 83.0 | 70.2 | 85.1 | 40.5 | 86.1 | 100 | 73.3 | 95.2 | 35.7 | 74.7 | 90.7 | 71.7 | 89.9 |
| 22 | 3.8 | 50.0 | 57.7 | 50.0 | 69.2 | 20.0 | 100 | 100 | 59.1 | 100 | 6.5 | 66.7 | 73.2 | 54.2 | 81.8 |
| 23 | 48.3 | 75.0 | 81.7 | 63.3 | 88.3 | 52.7 | 43.3 | 46.7 | 76.0 | 75.7 | 50.4 | 54.9 | 59.4 | 69.1 | 81.5 |
| 24 | 10.7 | 37.5 | 53.6 | 60.7 | 73.2 | 22.2 | 46.7 | 73.2 | 63.0 | 69.5 | 14.5 | 41.6 | 61.9 | 61.8 | 71.3 |
| 25 | 26.3 | 54.4 | 64.9 | 75.4 | 68.4 | 23.8 | 68.9 | 66.1 | 64.2 | 83.0 | 25.0 | 60.8 | 65.5 | 69.4 | 75.0 |
| 26 | 17.2 | 72.4 | 75.9 | 75.9 | 79.3 | 45.5 | 87.5 | 88.0 | 62.9 | 76.7 | 25.0 | 79.2 | 81.5 | 68.8 | 78.0 |
| 27 | 30.2 | 51.2 | 57.4 | 61.2 | 78.3 | 40.2 | 55.9 | 61.2 | 62.2 | 71.1 | 34.5 | 53.4 | 59.2 | 61.7 | 74.5 |
| 28 | 61.4 | 93.0 | 96.5 | 96.5 | 96.5 | 66.0 | 100 | 100 | 94.8 | 100 | 63.6 | 96.4 | 98.2 | 95.7 | 98.2 |
| 29 | 26.9 | 46.2 | 53.2 | 47.4 | 65.4 | 25.5 | 36.0 | 35.5 | 43.0 | 59.6 | 26.2 | 40.4 | 42.6 | 45.1 | 62.4 |
| 30 | 26.7 | 57.8 | 64.4 | 57.8 | 84.4 | 42.9 | 76.5 | 93.5 | 76.5 | 88.4 | 32.9 | 65.8 | 76.3 | 65.8 | 86.4 |
| 31 | 7.0 | 30.2 | 41.9 | 34.9 | 62.8 | 14.3 | 52.0 | 100 | 50.0 | 81.8 | 9.4 | 38.2 | 59.0 | 41.1 | 71.1 |
| 32 | 22.6 | 51.6 | 54.8 | 58.1 | 77.4 | 46.7 | 84.2 | 100 | 72.0 | 92.3 | 30.4 | 64.0 | 70.8 | 64.3 | 84.2 |
| 33 | 71.1 | 71.1 | 73.3 | 86.7 | 86.7 | 64.0 | 100 | 100 | 84.8 | 97.5 | 67.4 | 83.1 | 84.6 | 85.7 | 91.8 |
| 34 | 41.0 | 65.4 | 79.5 | 78.2 | 82.1 | 50.8 | 70.8 | 87.3 | 63.5 | 79.0 | 45.4 | 68.0 | 83.2 | 70.1 | 80.5 |
| 35 | 32.4 | 67.6 | 67.6 | 73.0 | 75.7 | 48.0 | 100 | 100 | 87.1 | 100 | 38.7 | 80.6 | 80.6 | 79.4 | 86.2 |
| 36 | 17.9 | 50.0 | 64.3 | 78.6 | 82.1 | 22.7 | 100 | 72.0 | 71.0 | 71.9 | 20.0 | 66.7 | 67.9 | 74.6 | 76.7 |
| 37 | 45.9 | 75.7 | 86.5 | 64.9 | 81.1 | 58.6 | 80.0 | 80.0 | 77.4 | 88.2 | 51.5 | 77.8 | 83.1 | 70.6 | 84.5 |
| 38 | 13.2 | 34.2 | 34.2 | 28.9 | 44.7 | 45.5 | 48.1 | 72.2 | 57.9 | 56.7 | 20.4 | 40.0 | 46.4 | 38.6 | 50.0 |
| 39 | 24.1 | 55.2 | 58.6 | 72.4 | 69.0 | 31.8 | 80.0 | 85.0 | 80.8 | 95.2 | 27.5 | 65.3 | 69.4 | 76.4 | 80.0 |
| 40 | 32.6 | 68.4 | 78.9 | 65.3 | 84.2 | 24.6 | 66.3 | 85.2 | 61.4 | 78.4 | 28.1 | 67.4 | 82.0 | 63.3 | 81.2 |
| 41 | 7.1 | 60.7 | 64.3 | 46.4 | 71.4 | 9.5 | 60.7 | 94.7 | 59.1 | 87.0 | 8.2 | 60.7 | 76.6 | 52.0 | 78.4 |
| 42 | 3.1 | 18.8 | 25.0 | 21.9 | 50.0 | 16.7 | 33.3 | 88.9 | 43.8 | 72.7 | 5.3 | 24.0 | 39.0 | 29.2 | 59.3 |
| 43 | 77.1 | 86.3 | 97.3 | 86.0 | 94.3 | 41.7 | 63.5 | 70.9 | 70.5 | 87.3 | 54.1 | 73.1 | 82.1 | 77.5 | 90.7 |
| 44 | 53.6 | 78.9 | 88.7 | 82.5 | 94.3 | 37.3 | 75.0 | 85.6 | 69.0 | 94.3 | 44.0 | 76.9 | 87.1 | 75.1 | 94.3 |
| 45 | 39.7 | 80.8 | 89.0 | 71.2 | 89.0 | 31.5 | 74.7 | 92.9 | 77.6 | 89.0 | 35.2 | 77.6 | 90.9 | 74.3 | 89.0 |
| 46 | 61.5 | 92.3 | 88.5 | 96.2 | 100 | 84.2 | 100 | 100 | 100 | 100 | 71.1 | 96.0 | 93.9 | 98.0 | 100 |
| 47 | 0.0 | 57.1 | 71.4 | 57.1 | 82.1 | 0.0 | 88.9 | 100 | 88.9 | 92.0 | 0.0 | 69.6 | 83.3 | 69.6 | 86.8 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 48 | 15.4 | 57.7 | 73.8 | 53.1 | 76.2 | 19.6 | 56.0 | 82.1 | 55.2 | 83.2 | 17.2 | 56.8 | 77.7 | 54.1 | 79.5 |
| 49 | 3.2 | 32.3 | 45.2 | 21.0 | 72.6 | 16.7 | 69.0 | 90.3 | 65.0 | 81.8 | 5.4 | 44.0 | 60.2 | 31.7 | 76.9 |
| 50 | 10.7 | 66.1 | 73.2 | 53.6 | 89.3 | 24.0 | 97.4 | 100 | 69.8 | 92.6 | 14.8 | 78.7 | 84.5 | 60.6 | 90.9 |
| 51 | 0.0 | 42.9 | 67.9 | 14.3 | 82.1 | 0.0 | 85.7 | 100 | 57.1 | 95.8 | 0.0 | 57.1 | 80.9 | 22.9 | 88.5 |
| 52 | 16.7 | 66.7 | 76.7 | 66.7 | 90.0 | 62.5 | 90.9 | 100 | 95.2 | 100 | 26.3 | 76.9 | 86.8 | 78.4 | 94.7 |
| 53 | 66.1 | 68.6 | 85.4 | 72.8 | 90.8 | 32.2 | 45.8 | 52.7 | 60.6 | 81.9 | 43.3 | 54.9 | 65.2 | 66.2 | 86.1 |
| 54 | 31.5 | 56.8 | 75.7 | 78.4 | 84.7 | 47.9 | 75.0 | 91.3 | 72.5 | 91.3 | 38.0 | 64.6 | 82.8 | 75.3 | 87.9 |
| 55 | 0.0 | 26.9 | 30.8 | 15.4 | 65.4 | 0.0 | 63.6 | 100 | 33.3 | 89.5 | 0.0 | 37.8 | 47.1 | 21.1 | 75.6 |
| 56 | 0.0 | 17.8 | 22.2 | 22.2 | 48.9 | 0.0 | 14.0 | 15.9 | 34.5 | 47.8 | 0.0 | 15.7 | 18.5 | 27.0 | 48.4 |
| 57 | 19.5 | 45.3 | 61.7 | 43.0 | 72.7 | 21.2 | 39.5 | 53.7 | 44.4 | 68.9 | 20.3 | 42.2 | 57.5 | 43.7 | 70.7 |
| 58 | 14.5 | 58.2 | 72.7 | 56.4 | 81.8 | 25.8 | 74.4 | 90.9 | 60.8 | 88.2 | 18.6 | 65.3 | 80.8 | 58.5 | 84.9 |
| 59 | 25.9 | 63.0 | 66.7 | 63.0 | 92.6 | 46.7 | 100 | 100 | 77.3 | 96.2 | 33.3 | 77.3 | 80.0 | 69.4 | 94.3 |
| 60 | 29.1 | 60.2 | 77.7 | 68.0 | 89.3 | 25.0 | 68.1 | 86.0 | 63.6 | 82.1 | 26.9 | 63.9 | 81.6 | 65.7 | 85.6 |
| 61 | 23.1 | 73.8 | 84.6 | 86.2 | 92.3 | 65.2 | 87.3 | 98.2 | 83.6 | 93.8 | 34.1 | 80.0 | 90.9 | 84.8 | 93.0 |
| 62 | 0.0 | 21.9 | 31.3 | 25.0 | 65.6 | 0.0 | 63.6 | 76.9 | 53.3 | 87.5 | 0.0 | 32.6 | 44.4 | 34.0 | 75.0 |
| 63 | 34.9 | 75.9 | 85.5 | 78.3 | 94.0 | 39.2 | 81.8 | 93.4 | 78.3 | 92.9 | 36.9 | 78.8 | 89.3 | 78.3 | 93.4 |
| 64 | 6.7 | 50.0 | 66.7 | 60.0 | 86.7 | 40.0 | 83.3 | 100 | 78.3 | 96.3 | 11.4 | 62.5 | 80.0 | 67.9 | 91.2 |
| 65 | 39.0 | 69.5 | 81.4 | 66.1 | 91.5 | 37.7 | 65.1 | 100 | 83.0 | 94.7 | 38.3 | 67.2 | 89.7 | 73.6 | 93.1 |
| 66 | 22.7 | 22.7 | 36.4 | 52.3 | 72.7 | 29.4 | 29.4 | 48.5 | 43.4 | 69.6 | 25.6 | 25.6 | 41.6 | 47.4 | 71.1 |
| 67 | 9.8 | 22.0 | 29.3 | 29.3 | 53.7 | 28.6 | 69.2 | 92.3 | 40.0 | 78.6 | 14.5 | 33.3 | 44.4 | 33.8 | 63.8 |
| 68 | 38.8 | 60.3 | 65.3 | 66.1 | 81.8 | 39.2 | 44.2 | 67.5 | 61.5 | 79.2 | 39.0 | 51.0 | 66.4 | 63.7 | 80.5 |
| 69 | 28.7 | 58.6 | 59.8 | 66.7 | 80.5 | 32.1 | 51.5 | 61.9 | 61.7 | 86.4 | 30.3 | 54.8 | 60.8 | 64.1 | 83.3 |
| 70 | 8.0 | 24.0 | 40.0 | 40.0 | 84.0 | 22.2 | 85.7 | 100 | 76.9 | 100 | 11.8 | 37.5 | 57.1 | 52.6 | 91.3 |
| 71 | 18.8 | 68.8 | 81.3 | 78.1 | 90.6 | 33.3 | 95.7 | 100 | 92.6 | 100 | 24.0 | 80.0 | 89.7 | 84.7 | 95.1 |
| 72 | 37.3 | 78.4 | 84.3 | 82.4 | 90.2 | 50.0 | 97.6 | 100 | 91.3 | 100 | 42.7 | 87.0 | 91.5 | 86.6 | 94.8 |
| 73 | 15.4 | 26.9 | 38.5 | 42.3 | 46.2 | 80.0 | 100 | 100 | 61.1 | 85.7 | 25.8 | 42.4 | 55.6 | 50.0 | 60.0 |
| 74 | 57.5 | 69.0 | 80.8 | 75.5 | 85.5 | 33.1 | 43.7 | 47.7 | 58.6 | 71.3 | 42.0 | 53.5 | 60.0 | 66.0 | 77.7 |
| 75 | 8.1 | 16.2 | 18.9 | 16.2 | 70.3 | 33.3 | 85.7 | 100 | 50.0 | 86.7 | 13.0 | 27.3 | 31.8 | 24.5 | 77.6 |
| 76 | 0.0 | 28.6 | 45.7 | 34.3 | 68.6 | 0.0 | 62.5 | 100 | 40.0 | 77.4 | 0.0 | 39.2 | 62.7 | 36.9 | 72.7 |
| 77 | 0.0 | 33.3 | 47.2 | 27.8 | 63.9 | 0.0 | 70.6 | 77.3 | 50.0 | 85.2 | 0.0 | 45.3 | 58.6 | 35.7 | 73.0 |
| 78 | 7.1 | 39.3 | 64.3 | 46.4 | 82.1 | 66.7 | 61.1 | 100 | 72.2 | 92.0 | 12.9 | 47.8 | 78.3 | 56.5 | 86.8 |
| 79 | 45.7 | 78.6 | 78.6 | 71.4 | 85.7 | 45.7 | 88.7 | 96.5 | 72.5 | 87.0 | 45.7 | 83.3 | 86.6 | 71.9 | 86.3 |
| 80 | 6.3 | 39.6 | 54.2 | 52.1 | 77.1 | 14.3 | 79.2 | 83.9 | 64.1 | 82.2 | 8.7 | 52.8 | 65.8 | 57.5 | 79.6 |
| 81 | 30.0 | 70.0 | 86.7 | 80.0 | 86.7 | 34.6 | 95.5 | 100 | 92.3 | 92.9 | 32.1 | 80.8 | 92.9 | 85.7 | 89.7 |
| 82 | 16.7 | 59.3 | 68.5 | 50.0 | 75.9 | 42.9 | 71.1 | 97.4 | 81.8 | 89.1 | 24.0 | 64.6 | 80.4 | 62.1 | 82.0 |
| 83 | 3.3 | 26.7 | 43.3 | 43.3 | 86.7 | 12.5 | 72.7 | 100 | 59.1 | 92.9 | 5.3 | 39.0 | 60.5 | 50.0 | 89.7 |
| 84 | 50.0 | 85.4 | 85.4 | 81.3 | 91.7 | 80.0 | 89.1 | 100 | 68.4 | 91.7 | 61.5 | 87.2 | 92.1 | 74.3 | 91.7 |
| 85 | 47.4 | 57.9 | 78.9 | 73.7 | 84.2 | 72.0 | 81.5 | 93.8 | 80.0 | 91.4 | 57.1 | 67.7 | 85.7 | 76.7 | 87.7 |
| 86 | 17.9 | 82.1 | 89.3 | 75.0 | 96.4 | 50.0 | 100 | 100 | 72.4 | 96.4 | 26.3 | 90.2 | 94.3 | 73.7 | 96.4 |
| 87 | 61.1 | 66.7 | 75.0 | 80.6 | 91.7 | 81.5 | 92.3 | 100 | 76.3 | 97.1 | 69.8 | 77.4 | 85.7 | 78.4 | 94.3 |
| 88 | 16.7 | 73.3 | 73.3 | 86.7 | 93.3 | 100 | 95.7 | 100 | 83.9 | 96.6 | 28.6 | 83.0 | 84.6 | 85.2 | 94.9 |
| 89 | 52.3 | 31.8 | 43.2 | 86.4 | 88.6 | 88.5 | 73.7 | 90.5 | 90.5 | 95.1 | 65.7 | 44.4 | 58.5 | 88.4 | 91.8 |
| 90 | 89.5 | 82.9 | 77.1 | 92.4 | 96.2 | 71.8 | 81.3 | 73.6 | 78.9 | 78.3 | 79.7 | 82.1 | 75.3 | 85.1 | 86.3 |
| 91 | 46.4 | 53.6 | 67.9 | 82.1 | 89.3 | 72.2 | 88.2 | 100 | 92.0 | 96.2 | 56.5 | 66.7 | 80.9 | 86.8 | 92.6 |
| 92 | 47.5 | 27.5 | 30.0 | 70.0 | 50.0 | 76.0 | 57.9 | 85.7 | 66.7 | 76.9 | 58.5 | 37.3 | 44.4 | 68.3 | 60.6 |
| 93 | 38.6 | 49.1 | 63.2 | 56.1 | 61.4 | 48.9 | 50.0 | 59.0 | 71.1 | 72.9 | 43.1 | 49.6 | 61.0 | 62.7 | 66.7 |
| 94 | 30.2 | 39.6 | 52.8 | 49.1 | 56.6 | 42.1 | 58.3 | 60.9 | 60.5 | 76.9 | 35.2 | 47.2 | 56.6 | 54.2 | 65.2 |
| 95 | 29.6 | 22.2 | 40.7 | 55.6 | 66.7 | 47.1 | 100 | 100 | 75.0 | 90.0 | 36.4 | 36.4 | 57.9 | 63.8 | 76.6 |
| W.A. | 41.6 | 62.5 | 71.8 | 68.0 | 82.4 | 40.3 | 66.5 | 77.2 | 67.9 | 82.9 | 38.5 | 62.2 | 71.8 | 67.0 | 82.3 |

**Table SC7** Comparison with the major existing methods via *2-fold cross-validation* on **F194** dataset. For each of the *Recall*, *Precision*, and *F-measure* columns, the five sub-columns from left to right are the accuracies for the Shamim, ACCFold_AC, ACCFold_ACC, PFRES, and TAXFOLD, respectively.

| **Fold** | **Recall** (%) | | | | | **Precision** (%) | | | | | **F-measure** (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 43.9 | 82.9 | 87.8 | 80.5 | 78.0 | 39.1 | 97.1 | 100 | 89.2 | 86.5 | 41.4 | 89.5 | 93.5 | 84.6 | 82.1 |
| 2 | 29.3 | 29.3 | 39.0 | 43.9 | 46.3 | 42.9 | 60.0 | 61.5 | 48.6 | 48.7 | 34.8 | 39.3 | 47.8 | 46.2 | 47.5 |
| 3 | 0.0 | 0.0 | 0.0 | 13.3 | 0.0 | 0.0 | 0.0 | 0.0 | 25.0 | 0.0 | 0.0 | 0.0 | 0.0 | 17.4 | 0.0 |
| 4 | 64.7 | 76.5 | 76.5 | 88.2 | 85.3 | 71.0 | 96.3 | 100 | 83.3 | 93.5 | 67.7 | 85.2 | 86.7 | 85.7 | 89.2 |
| 5 | 74.5 | 72.4 | 82.0 | 92.2 | 90.4 | 41.0 | 40.4 | 54.0 | 47.9 | 58.1 | 52.9 | 51.8 | 65.1 | 63.1 | 70.7 |
| 6 | 33.3 | 22.2 | 37.8 | 33.3 | 46.7 | 50.0 | 76.9 | 94.4 | 75.0 | 75.0 | 40.0 | 34.5 | 54.0 | 46.2 | 57.5 |
| 7 | 0.0 | 21.4 | 21.4 | 0.0 | 14.3 | 0.0 | 100 | 100 | 0.0 | 100 | 0.0 | 35.3 | 35.3 | 0.0 | 25.0 |
| 8 | 26.2 | 54.8 | 59.5 | 52.4 | 54.8 | 28.2 | 60.5 | 78.1 | 68.8 | 56.1 | 27.2 | 57.5 | 67.6 | 59.5 | 55.4 |
| 9 | 0.0 | 8.7 | 13.0 | 13.0 | 26.1 | 0.0 | 50.0 | 60.0 | 50.0 | 66.7 | 0.0 | 14.8 | 21.4 | 20.7 | 37.5 |
| 10 | 4.8 | 47.6 | 57.1 | 42.9 | 47.6 | 20.0 | 83.3 | 100 | 100 | 76.9 | 7.7 | 60.6 | 72.7 | 60.0 | 58.8 |
| 11 | 37.7 | 60.9 | 73.9 | 56.5 | 55.1 | 26.8 | 47.7 | 50.0 | 46.4 | 55.1 | 31.3 | 53.5 | 59.6 | 51.0 | 55.1 |
| 12 | 51.0 | 61.2 | 69.4 | 61.2 | 75.5 | 41.7 | 76.9 | 73.9 | 73.2 | 57.8 | 45.9 | 68.2 | 71.6 | 66.7 | 65.5 |
| 13 | 16.7 | 73.3 | 76.7 | 63.3 | 76.7 | 35.7 | 57.9 | 79.3 | 70.4 | 71.9 | 22.7 | 64.7 | 78.0 | 66.7 | 74.2 |
| 14 | 0.0 | 30.8 | 38.5 | 23.1 | 53.8 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 47.1 | 55.6 | 37.5 | 70.0 |
| 15 | 5.7 | 45.7 | 45.7 | 48.6 | 51.4 | 12.5 | 76.2 | 88.9 | 77.3 | 69.2 | 7.8 | 57.1 | 60.4 | 59.6 | 59.0 |
| 16 | 0.0 | 34.5 | 44.8 | 10.3 | 44.8 | 0.0 | 83.3 | 100 | 100 | 72.2 | 0.0 | 48.8 | 61.9 | 18.8 | 55.3 |
| 17 | 52.5 | 62.7 | 66.1 | 78.0 | 76.3 | 63.3 | 97.4 | 95.1 | 80.7 | 91.8 | 57.4 | 76.3 | 78.0 | 79.3 | 83.3 |
| 18 | 8.3 | 0.0 | 8.3 | 8.3 | 33.3 | 33.3 | 0.0 | 100 | 25.0 | 80.0 | 13.3 | 0.0 | 15.4 | 12.5 | 47.1 |
| 19 | 0.0 | 25.0 | 12.5 | 18.8 | 18.8 | 0.0 | 66.7 | 100 | 75.0 | 60.0 | 0.0 | 36.4 | 22.2 | 30.0 | 28.6 |
| 20 | 37.9 | 86.2 | 79.3 | 86.2 | 82.8 | 32.4 | 100 | 100 | 89.3 | 92.3 | 34.9 | 92.6 | 88.5 | 87.7 | 87.3 |
| 21 | 0.0 | 0.0 | 0.0 | 9.1 | 9.1 | 0.0 | 0.0 | 0.0 | 100 | 100 | 0.0 | 0.0 | 0.0 | 16.7 | 16.7 |
| 22 | 23.7 | 60.5 | 61.8 | 31.6 | 61.8 | 28.6 | 60.5 | 68.1 | 49.0 | 53.4 | 25.9 | 60.5 | 64.8 | 38.4 | 57.3 |
| 23 | 0.0 | 21.7 | 39.1 | 43.5 | 52.2 | 0.0 | 100 | 100 | 66.7 | 92.3 | 0.0 | 35.7 | 56.3 | 52.6 | 66.7 |
| 24 | 0.0 | 38.5 | 38.5 | 46.2 | 46.2 | 0.0 | 100 | 100 | 60.0 | 100 | 0.0 | 55.6 | 55.6 | 52.2 | 63.2 |
| 25 | 0.0 | 35.3 | 58.8 | 47.1 | 58.8 | 0.0 | 100 | 100 | 80.0 | 100 | 0.0 | 52.2 | 74.1 | 59.3 | 74.1 |
| 26 | 15.8 | 42.1 | 57.9 | 36.8 | 52.6 | 23.1 | 88.9 | 91.7 | 77.8 | 76.9 | 18.7 | 57.1 | 71.0 | 50.0 | 62.5 |
| 27 | 12.0 | 40.0 | 56.0 | 36.0 | 64.0 | 27.3 | 100 | 100 | 90.0 | 94.1 | 16.7 | 57.1 | 71.8 | 51.4 | 76.2 |
| 28 | 77.8 | 71.1 | 77.8 | 86.7 | 88.9 | 62.5 | 69.6 | 92.1 | 73.6 | 90.9 | 69.3 | 70.3 | 84.3 | 79.6 | 89.9 |
| 29 | 23.5 | 82.4 | 88.2 | 82.4 | 88.2 | 50.0 | 93.3 | 100 | 100 | 100 | 32.0 | 87.5 | 93.8 | 90.3 | 93.8 |
| 30 | 59.0 | 62.9 | 79.0 | 80.0 | 81.0 | 33.5 | 45.5 | 51.2 | 61.8 | 64.4 | 42.8 | 52.8 | 62.2 | 69.7 | 71.7 |
| 31 | 32.4 | 61.8 | 70.6 | 76.5 | 76.5 | 44.0 | 56.8 | 66.7 | 70.3 | 81.3 | 37.3 | 59.2 | 68.6 | 73.2 | 78.8 |
| 32 | 41.2 | 94.1 | 100 | 100 | 100 | 77.8 | 100 | 100 | 100 | 100 | 53.8 | 97.0 | 100 | 100 | 100 |
| 33 | 0.0 | 81.8 | 81.8 | 90.9 | 81.8 | 0.0 | 90.0 | 100 | 71.4 | 90.0 | 0.0 | 85.7 | 90.0 | 80.0 | 85.7 |
| 34 | 23.1 | 61.5 | 61.5 | 69.2 | 69.2 | 37.5 | 100 | 100 | 69.2 | 90.0 | 28.6 | 76.2 | 76.2 | 69.2 | 78.3 |
| 35 | 0.0 | 21.4 | 14.3 | 14.3 | 14.3 | 0.0 | 75.0 | 66.7 | 50.0 | 100 | 0.0 | 33.3 | 23.5 | 22.2 | 25.0 |
| 36 | 72.7 | 68.2 | 81.8 | 72.7 | 72.7 | 88.9 | 100 | 100 | 80.0 | 88.9 | 80.0 | 81.1 | 90.0 | 76.2 | 80.0 |
| 37 | 84.9 | 85.9 | 89.8 | 91.8 | 91.3 | 50.1 | 58.3 | 66.5 | 61.5 | 73.0 | 63.0 | 69.5 | 76.4 | 73.6 | 81.1 |
| 38 | 8.7 | 19.6 | 30.4 | 21.7 | 37.0 | 18.2 | 24.3 | 28.6 | 58.8 | 39.5 | 11.8 | 21.7 | 29.5 | 31.7 | 38.2 |
| 39 | 4.8 | 23.8 | 52.4 | 0.0 | 38.1 | 16.7 | 100 | 100 | 0.0 | 88.9 | 7.4 | 38.5 | 68.8 | 0.0 | 53.3 |
| 40 | 17.0 | 59.6 | 68.1 | 57.4 | 76.6 | 29.6 | 82.4 | 88.9 | 69.2 | 87.8 | 21.6 | 69.1 | 77.1 | 62.8 | 81.8 |
| 41 | 3.8 | 53.8 | 65.4 | 46.2 | 69.2 | 16.7 | 100 | 100 | 92.3 | 100 | 6.3 | 70.0 | 79.1 | 61.5 | 81.8 |
| 42 | 50.0 | 65.0 | 80.0 | 63.3 | 83.3 | 41.1 | 27.5 | 32.2 | 69.1 | 70.4 | 45.1 | 38.6 | 45.9 | 66.1 | 76.3 |
| 43 | 8.3 | 33.3 | 50.0 | 16.7 | 58.3 | 33.3 | 100 | 100 | 100 | 100 | 13.3 | 50.0 | 66.7 | 28.6 | 73.7 |
| 44 | 8.9 | 48.2 | 60.7 | 60.7 | 76.8 | 17.9 | 56.3 | 79.1 | 53.1 | 66.2 | 11.9 | 51.9 | 68.7 | 56.7 | 71.1 |
| 45 | 28.6 | 85.7 | 85.7 | 78.6 | 85.7 | 80.0 | 100 | 92.3 | 100 | 100 | 42.1 | 92.3 | 88.9 | 88.0 | 92.3 |
| 46 | 6.3 | 50.0 | 62.5 | 18.8 | 62.5 | 25.0 | 80.0 | 90.9 | 100 | 90.9 | 10.0 | 61.5 | 74.1 | 31.6 | 74.1 |
| 47 | 42.1 | 50.9 | 71.9 | 77.2 | 63.2 | 29.3 | 58.0 | 74.5 | 60.3 | 72.0 | 34.5 | 54.2 | 73.2 | 67.7 | 67.3 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 48 | 13.8 | 58.6 | 69.0 | 79.3 | 75.9 | 36.4 | 68.0 | 90.9 | 74.2 | 75.9 | 20.0 | 63.0 | 78.4 | 76.7 | 75.9 |
| 49 | 7.1 | 78.6 | 92.9 | 57.1 | 100 | 50.0 | 100 | 100 | 100 | 100 | 12.5 | 88.0 | 96.3 | 72.7 | 100 |
| 50 | 35.7 | 50.4 | 57.4 | 62.8 | 72.9 | 38.3 | 52.0 | 61.2 | 52.6 | 62.3 | 36.9 | 51.2 | 59.2 | 57.2 | 67.1 |
| 51 | 20.0 | 55.0 | 70.0 | 70.0 | 85.0 | 50.0 | 100 | 100 | 82.4 | 100 | 28.6 | 71.0 | 82.4 | 75.7 | 91.9 |
| 52 | 59.6 | 93.0 | 98.2 | 93.0 | 94.7 | 73.9 | 100 | 100 | 84.1 | 100 | 66.0 | 96.4 | 99.1 | 88.3 | 97.3 |
| 53 | 13.6 | 40.9 | 54.5 | 27.3 | 50.0 | 33.3 | 90.0 | 85.7 | 66.7 | 91.7 | 19.4 | 56.3 | 66.7 | 38.7 | 64.7 |
| 54 | 31.4 | 46.2 | 56.4 | 43.6 | 67.9 | 22.2 | 25.4 | 31.3 | 35.8 | 43.8 | 26.0 | 32.7 | 40.3 | 39.3 | 53.3 |
| 55 | 31.1 | 73.3 | 82.2 | 55.6 | 80.0 | 45.2 | 78.6 | 80.4 | 71.4 | 83.7 | 36.8 | 75.9 | 81.3 | 62.5 | 81.8 |
| 56 | 11.6 | 39.5 | 58.1 | 25.6 | 46.5 | 16.7 | 60.7 | 86.2 | 73.3 | 83.3 | 13.7 | 47.9 | 69.4 | 37.9 | 59.7 |
| 57 | 0.0 | 27.3 | 36.4 | 18.2 | 27.3 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 42.9 | 53.3 | 30.8 | 42.9 |
| 58 | 25.8 | 54.8 | 64.5 | 32.3 | 64.5 | 44.4 | 89.5 | 100 | 83.3 | 71.4 | 32.7 | 68.0 | 78.4 | 46.5 | 67.8 |
| 59 | 0.0 | 23.1 | 23.1 | 23.1 | 23.1 | 0.0 | 100 | 100 | 75.0 | 60.0 | 0.0 | 37.5 | 37.5 | 35.3 | 33.3 |
| 60 | 73.3 | 75.6 | 75.6 | 84.4 | 86.7 | 62.3 | 94.4 | 100 | 76.0 | 92.9 | 67.3 | 84.0 | 86.1 | 80.0 | 89.7 |
| 61 | 6.7 | 46.7 | 73.3 | 13.3 | 60.0 | 100 | 100 | 100 | 66.7 | 100 | 12.5 | 63.6 | 84.6 | 22.2 | 75.0 |
| 62 | 41.2 | 76.5 | 88.2 | 88.2 | 94.1 | 87.5 | 100 | 100 | 100 | 100 | 56.0 | 86.7 | 93.8 | 93.8 | 97.0 |
| 63 | 8.7 | 43.5 | 56.5 | 39.1 | 52.2 | 33.3 | 100 | 100 | 81.8 | 70.6 | 13.8 | 60.6 | 72.2 | 52.9 | 60.0 |
| 64 | 48.7 | 59.0 | 75.6 | 74.4 | 70.5 | 45.8 | 63.9 | 89.4 | 54.7 | 68.8 | 47.2 | 61.3 | 81.9 | 63.0 | 69.6 |
| 65 | 27.0 | 70.3 | 70.3 | 70.3 | 70.3 | 47.6 | 92.9 | 100 | 92.9 | 96.3 | 34.5 | 80.0 | 82.5 | 80.0 | 81.3 |
| 66 | 11.8 | 47.1 | 52.9 | 11.8 | 47.1 | 100 | 80.0 | 100 | 100 | 100 | 21.1 | 59.3 | 69.2 | 21.1 | 64.0 |
| 67 | 17.9 | 42.9 | 67.9 | 82.1 | 78.6 | 29.4 | 85.7 | 55.9 | 76.7 | 71.0 | 22.2 | 57.1 | 61.3 | 79.3 | 74.6 |
| 68 | 37.8 | 59.5 | 67.6 | 75.7 | 70.3 | 58.3 | 73.3 | 62.5 | 87.5 | 92.9 | 45.9 | 65.7 | 64.9 | 81.2 | 80.0 |
| 69 | 7.9 | 34.2 | 42.1 | 0.0 | 36.8 | 20.0 | 48.1 | 76.2 | 0.0 | 56.0 | 11.3 | 40.0 | 54.2 | 0.0 | 44.4 |
| 70 | 41.4 | 51.7 | 65.5 | 58.6 | 65.5 | 41.4 | 75.0 | 65.5 | 89.5 | 79.2 | 41.4 | 61.2 | 65.5 | 70.8 | 71.7 |
| 71 | 38.9 | 66.7 | 72.2 | 44.4 | 55.6 | 46.7 | 100 | 100 | 100 | 100 | 42.4 | 80.0 | 83.9 | 61.5 | 71.4 |
| 72 | 29.5 | 54.7 | 70.5 | 53.7 | 74.7 | 14.9 | 50.5 | 65.0 | 53.7 | 68.9 | 19.8 | 52.5 | 67.7 | 53.7 | 71.7 |
| 73 | 4.3 | 47.8 | 56.5 | 26.1 | 43.5 | 20.0 | 64.7 | 92.9 | 85.7 | 90.9 | 7.1 | 55.0 | 70.3 | 40.0 | 58.8 |
| 74 | 4.2 | 50.0 | 54.2 | 25.0 | 50.0 | 20.0 | 85.7 | 100 | 60.0 | 92.3 | 6.9 | 63.2 | 70.3 | 35.3 | 64.9 |
| 75 | 10.7 | 53.6 | 60.7 | 21.4 | 64.3 | 23.1 | 71.4 | 94.4 | 75.0 | 85.7 | 14.6 | 61.2 | 73.9 | 33.3 | 73.5 |
| 76 | 3.1 | 21.9 | 31.3 | 12.5 | 37.5 | 5.9 | 50.0 | 71.4 | 57.1 | 66.7 | 4.1 | 30.4 | 43.5 | 20.5 | 48.0 |
| 77 | 79.8 | 90.5 | 99.1 | 91.7 | 92.6 | 31.4 | 52.1 | 62.1 | 50.8 | 74.9 | 45.1 | 66.1 | 76.4 | 65.4 | 82.8 |
| 78 | 50.5 | 81.4 | 89.7 | 78.9 | 91.2 | 31.6 | 69.9 | 81.7 | 61.0 | 83.9 | 38.9 | 75.2 | 85.5 | 68.8 | 87.4 |
| 79 | 49.3 | 74.0 | 90.4 | 57.5 | 86.3 | 31.0 | 63.5 | 82.5 | 60.0 | 85.1 | 38.1 | 68.4 | 86.3 | 58.7 | 85.7 |
| 80 | 0.0 | 30.0 | 55.0 | 35.0 | 55.0 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 46.2 | 71.0 | 51.9 | 71.0 |
| 81 | 5.0 | 15.0 | 25.0 | 25.0 | 40.0 | 14.3 | 75.0 | 100 | 100 | 53.3 | 7.4 | 25.0 | 40.0 | 40.0 | 45.7 |
| 82 | 69.2 | 76.9 | 100 | 92.3 | 100 | 85.7 | 100 | 100 | 96.0 | 100 | 76.6 | 87.0 | 100 | 94.1 | 100 |
| 83 | 3.6 | 60.7 | 71.4 | 60.7 | 78.6 | 33.3 | 89.5 | 100 | 100 | 91.7 | 6.5 | 72.3 | 83.3 | 75.6 | 84.6 |
| 84 | 0.0 | 7.1 | 28.6 | 0.0 | 14.3 | 0.0 | 100 | 100 | 0.0 | 100 | 0.0 | 13.3 | 44.4 | 0.0 | 25.0 |
| 85 | 17.7 | 63.1 | 71.5 | 51.5 | 72.3 | 14.5 | 50.0 | 63.7 | 43.8 | 67.1 | 15.9 | 55.8 | 67.4 | 47.3 | 69.6 |
| 86 | 7.7 | 69.2 | 76.9 | 61.5 | 76.9 | 33.3 | 100 | 100 | 100 | 100 | 12.5 | 81.8 | 87.0 | 76.2 | 87.0 |
| 87 | 4.8 | 45.2 | 64.5 | 21.0 | 64.5 | 13.0 | 45.2 | 66.7 | 86.7 | 67.8 | 7.1 | 45.2 | 65.6 | 33.8 | 66.1 |
| 88 | 0.0 | 0.0 | 18.2 | 0.0 | 36.4 | 0.0 | 0.0 | 100 | 0.0 | 80.0 | 0.0 | 0.0 | 30.8 | 0.0 | 50.0 |
| 89 | 0.0 | 38.9 | 55.6 | 38.9 | 55.6 | 0.0 | 100 | 100 | 87.5 | 90.9 | 0.0 | 56.0 | 71.4 | 53.8 | 69.0 |
| 90 | 0.0 | 12.5 | 12.5 | 12.5 | 18.8 | 0.0 | 50.0 | 100 | 50.0 | 100 | 0.0 | 20.0 | 22.2 | 20.0 | 31.6 |
| 91 | 5.3 | 68.4 | 73.7 | 57.9 | 73.7 | 100 | 100 | 100 | 91.7 | 93.3 | 10.0 | 81.3 | 84.8 | 71.0 | 82.4 |
| 92 | 8.9 | 66.1 | 78.6 | 42.9 | 78.6 | 14.3 | 88.1 | 97.8 | 96.0 | 81.5 | 11.0 | 75.5 | 87.1 | 59.3 | 80.0 |
| 93 | 0.0 | 46.4 | 60.7 | 21.4 | 60.7 | 0.0 | 100 | 100 | 100 | 77.3 | 0.0 | 63.4 | 75.6 | 35.3 | 68.0 |
| 94 | 33.3 | 63.3 | 73.3 | 40.0 | 80.0 | 76.9 | 95.0 | 100 | 100 | 100 | 46.5 | 76.0 | 84.6 | 57.1 | 88.9 |
| 95 | 62.8 | 70.3 | 82.8 | 78.7 | 88.3 | 25.9 | 40.9 | 46.0 | 41.0 | 69.0 | 36.7 | 51.7 | 59.2 | 53.9 | 77.4 |
| 96 | 0.0 | 18.2 | 18.2 | 18.2 | 18.2 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 30.8 | 30.8 | 30.8 | 30.8 |
| 97 | 0.0 | 50.0 | 75.0 | 60.0 | 75.0 | 0.0 | 100 | 100 | 92.3 | 88.2 | 0.0 | 66.7 | 85.7 | 72.7 | 81.1 |
| 98 | 23.8 | 81.0 | 90.5 | 61.9 | 81.0 | 71.4 | 100 | 100 | 86.7 | 100 | 35.7 | 89.5 | 95.0 | 72.2 | 89.5 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 99 | 24.3 | 51.4 | 70.3 | 53.2 | 76.6 | 29.3 | 60.0 | 85.7 | 53.2 | 85.9 | 26.6 | 55.3 | 77.2 | 53.2 | 81.0 |
| 100 | 0.0 | 38.5 | 38.5 | 30.8 | 34.6 | 0.0 | 66.7 | 83.3 | 88.9 | 75.0 | 0.0 | 48.8 | 52.6 | 45.7 | 47.4 |
| 101 | 2.2 | 6.7 | 20.0 | 8.9 | 40.0 | 7.1 | 6.0 | 16.7 | 23.5 | 31.6 | 3.4 | 6.3 | 18.2 | 12.9 | 35.3 |
| 102 | 18.8 | 39.8 | 60.9 | 39.8 | 64.1 | 15.7 | 32.3 | 46.2 | 45.9 | 48.2 | 17.1 | 35.7 | 52.5 | 42.7 | 55.0 |
| 103 | 3.6 | 49.1 | 69.1 | 52.7 | 83.6 | 7.7 | 57.4 | 84.4 | 67.4 | 76.7 | 4.9 | 52.9 | 76.0 | 59.2 | 80.0 |
| 104 | 11.1 | 77.8 | 88.9 | 55.6 | 92.6 | 50.0 | 91.3 | 100 | 83.3 | 89.3 | 18.2 | 84.0 | 94.1 | 66.7 | 90.9 |
| 105 | 0.0 | 61.5 | 69.2 | 7.7 | 69.2 | 0.0 | 100 | 100 | 50.0 | 100 | 0.0 | 76.2 | 81.8 | 13.3 | 81.8 |
| 106 | 34.0 | 67.0 | 85.4 | 65.0 | 76.7 | 25.0 | 69.7 | 74.6 | 53.6 | 66.9 | 28.8 | 68.3 | 79.6 | 58.8 | 71.5 |
| 107 | 10.8 | 66.2 | 80.0 | 81.5 | 93.8 | 58.3 | 84.3 | 94.5 | 72.6 | 87.1 | 18.2 | 74.1 | 86.7 | 76.8 | 90.4 |
| 108 | 0.0 | 28.1 | 46.9 | 28.1 | 53.1 | 0.0 | 60.0 | 93.8 | 90.0 | 70.8 | 0.0 | 38.3 | 62.5 | 42.9 | 60.7 |
| 109 | 34.9 | 74.7 | 84.3 | 68.7 | 89.2 | 35.8 | 73.8 | 92.1 | 65.5 | 92.5 | 35.4 | 74.3 | 88.1 | 67.1 | 90.8 |
| 110 | 7.1 | 85.7 | 100 | 92.9 | 100 | 100 | 100 | 100 | 86.7 | 100 | 13.3 | 92.3 | 100 | 89.7 | 100 |
| 111 | 6.7 | 50.0 | 80.0 | 66.7 | 83.3 | 40.0 | 88.2 | 100 | 87.0 | 96.2 | 11.4 | 63.8 | 88.9 | 75.5 | 89.3 |
| 112 | 0.0 | 16.7 | 58.3 | 25.0 | 50.0 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 28.6 | 73.7 | 40.0 | 66.7 |
| 113 | 0.0 | 57.1 | 78.6 | 7.1 | 57.1 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 72.7 | 88.0 | 13.3 | 72.7 |
| 114 | 0.0 | 29.4 | 29.4 | 11.8 | 29.4 | 0.0 | 100 | 100 | 66.7 | 83.3 | 0.0 | 45.5 | 45.5 | 20.0 | 43.5 |
| 115 | 0.0 | 55.6 | 77.8 | 16.7 | 55.6 | 0.0 | 90.9 | 93.3 | 100 | 83.3 | 0.0 | 69.0 | 84.8 | 28.6 | 66.7 |
| 116 | 0.0 | 50.0 | 62.5 | 50.0 | 81.3 | 0.0 | 80.0 | 100 | 100 | 100 | 0.0 | 61.5 | 76.9 | 66.7 | 89.7 |
| 117 | 37.3 | 71.2 | 84.7 | 62.7 | 89.8 | 44.0 | 65.6 | 92.6 | 78.7 | 89.8 | 40.4 | 68.3 | 88.5 | 69.8 | 89.8 |
| 118 | 0.0 | 54.2 | 83.3 | 41.7 | 91.7 | 0.0 | 100 | 100 | 76.9 | 84.6 | 0.0 | 70.3 | 90.9 | 54.1 | 88.0 |
| 119 | 0.0 | 35.3 | 47.1 | 29.4 | 58.8 | 0.0 | 100 | 100 | 83.3 | 100 | 0.0 | 52.2 | 64.0 | 43.5 | 74.1 |
| 120 | 0.0 | 42.1 | 52.6 | 52.6 | 84.2 | 0.0 | 100 | 100 | 83.3 | 69.6 | 0.0 | 59.3 | 69.0 | 64.5 | 76.2 |
| 121 | 13.6 | 13.6 | 20.5 | 38.6 | 43.2 | 18.2 | 27.3 | 45.0 | 51.5 | 61.3 | 15.6 | 18.2 | 28.1 | 44.2 | 50.7 |
| 122 | 33.3 | 73.3 | 86.7 | 86.7 | 80.0 | 100 | 100 | 100 | 92.9 | 100 | 50.0 | 84.6 | 92.9 | 89.7 | 88.9 |
| 123 | 0.0 | 9.1 | 54.5 | 27.3 | 54.5 | 0.0 | 100 | 100 | 75.0 | 100 | 0.0 | 16.7 | 70.6 | 40.0 | 70.6 |
| 124 | 5.0 | 35.0 | 35.0 | 30.0 | 50.0 | 25.0 | 77.8 | 70.0 | 54.5 | 71.4 | 8.3 | 48.3 | 46.7 | 38.7 | 58.8 |
| 125 | 7.3 | 26.8 | 26.8 | 31.7 | 41.5 | 20.0 | 61.1 | 100 | 65.0 | 60.7 | 10.7 | 37.3 | 42.3 | 42.6 | 49.3 |
| 126 | 42.1 | 60.3 | 71.9 | 71.9 | 81.8 | 39.5 | 42.9 | 63.0 | 56.5 | 73.3 | 40.8 | 50.2 | 67.2 | 63.3 | 77.3 |
| 127 | 12.5 | 29.2 | 33.3 | 41.7 | 54.2 | 60.0 | 87.5 | 100 | 71.4 | 61.9 | 20.7 | 43.8 | 50.0 | 52.6 | 57.8 |
| 128 | 24.1 | 55.2 | 60.9 | 69.0 | 75.9 | 28.0 | 57.8 | 60.2 | 68.2 | 70.2 | 25.9 | 56.5 | 60.6 | 68.6 | 72.9 |
| 129 | 40.0 | 40.0 | 60.0 | 66.7 | 66.7 | 85.7 | 100 | 100 | 100 | 100 | 54.5 | 57.1 | 75.0 | 80.0 | 80.0 |
| 130 | 33.3 | 66.7 | 66.7 | 61.9 | 76.2 | 77.8 | 100 | 100 | 65.0 | 100 | 46.7 | 80.0 | 80.0 | 63.4 | 86.5 |
| 131 | 0.0 | 50.0 | 58.3 | 16.7 | 75.0 | 0.0 | 100 | 100 | 66.7 | 100 | 0.0 | 66.7 | 73.7 | 26.7 | 85.7 |
| 132 | 8.3 | 16.7 | 25.0 | 25.0 | 41.7 | 100 | 100 | 100 | 100 | 100 | 15.4 | 28.6 | 40.0 | 40.0 | 58.8 |
| 133 | 4.0 | 44.0 | 52.0 | 44.0 | 68.0 | 11.1 | 100 | 100 | 91.7 | 100 | 5.9 | 61.1 | 68.4 | 59.5 | 81.0 |
| 134 | 25.0 | 78.1 | 87.5 | 87.5 | 84.4 | 50.0 | 100 | 100 | 62.2 | 100 | 33.3 | 87.7 | 93.3 | 72.7 | 91.5 |
| 135 | 72.2 | 83.3 | 83.3 | 88.9 | 88.9 | 100 | 100 | 100 | 94.1 | 100 | 83.9 | 90.9 | 90.9 | 91.4 | 94.1 |
| 136 | 54.2 | 91.7 | 91.7 | 91.7 | 87.5 | 92.9 | 100 | 100 | 91.7 | 100 | 68.4 | 95.7 | 95.7 | 91.7 | 93.3 |
| 137 | 49.0 | 76.5 | 78.4 | 80.4 | 80.4 | 38.5 | 95.1 | 97.6 | 68.3 | 91.1 | 43.1 | 84.8 | 87.0 | 73.9 | 85.4 |
| 138 | 9.5 | 42.9 | 47.6 | 42.9 | 52.4 | 28.6 | 100 | 100 | 90.0 | 100 | 14.3 | 60.0 | 64.5 | 58.1 | 68.8 |
| 139 | 22.7 | 31.8 | 50.0 | 50.0 | 54.5 | 50.0 | 87.5 | 100 | 100 | 100 | 31.3 | 46.7 | 66.7 | 66.7 | 70.6 |
| 140 | 20.8 | 54.2 | 62.5 | 54.2 | 70.8 | 50.0 | 100 | 100 | 81.3 | 100 | 29.4 | 70.3 | 76.9 | 65.0 | 82.9 |
| 141 | 3.8 | 30.8 | 42.3 | 15.4 | 34.6 | 50.0 | 88.9 | 100 | 100 | 90.0 | 7.1 | 45.7 | 59.5 | 26.7 | 50.0 |
| 142 | 0.0 | 30.4 | 39.1 | 39.1 | 52.2 | 0.0 | 87.5 | 100 | 81.8 | 85.7 | 0.0 | 45.2 | 56.3 | 52.9 | 64.9 |
| 143 | 0.0 | 75.0 | 75.0 | 75.0 | 81.3 | 0.0 | 100 | 100 | 80.0 | 100 | 0.0 | 85.7 | 85.7 | 77.4 | 89.7 |
| 144 | 57.8 | 64.0 | 74.9 | 78.2 | 81.1 | 25.9 | 34.6 | 43.7 | 43.1 | 54.8 | 35.7 | 44.9 | 55.2 | 55.6 | 65.4 |
| 145 | 41.7 | 83.3 | 83.3 | 83.3 | 83.3 | 71.4 | 100 | 100 | 90.9 | 100 | 52.6 | 90.9 | 90.9 | 87.0 | 90.9 |
| 146 | 14.3 | 52.4 | 66.7 | 57.1 | 95.2 | 75.0 | 100 | 93.3 | 92.3 | 90.9 | 24.0 | 68.8 | 77.8 | 70.6 | 93.0 |
| 147 | 5.0 | 30.0 | 30.0 | 25.0 | 35.0 | 33.3 | 100 | 100 | 100 | 63.6 | 8.7 | 46.2 | 46.2 | 40.0 | 45.2 |
| 148 | 0.0 | 15.4 | 23.1 | 7.7 | 15.4 | 0.0 | 66.7 | 100 | 100 | 100 | 0.0 | 25.0 | 37.5 | 14.3 | 26.7 |
| 149 | 0.0 | 37.8 | 37.8 | 18.9 | 37.8 | 0.0 | 66.7 | 87.5 | 77.8 | 73.7 | 0.0 | 48.3 | 52.8 | 30.4 | 50.0 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 150 | 0.0 | 63.6 | 54.5 | 27.3 | 45.5 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 77.8 | 70.6 | 42.9 | 62.5 |
| 151 | 5.7 | 45.7 | 51.4 | 22.9 | 54.3 | 28.6 | 72.7 | 78.3 | 72.7 | 59.4 | 9.5 | 56.1 | 62.1 | 34.8 | 56.7 |
| 152 | 0.0 | 27.8 | 33.3 | 11.1 | 33.3 | 0.0 | 83.3 | 100 | 66.7 | 100 | 0.0 | 41.7 | 50.0 | 19.0 | 50.0 |
| 153 | 2.8 | 38.9 | 44.4 | 38.9 | 69.4 | 9.1 | 73.7 | 59.3 | 73.7 | 80.6 | 4.3 | 50.9 | 50.8 | 50.9 | 74.6 |
| 154 | 43.5 | 87.0 | 87.0 | 78.3 | 87.0 | 71.4 | 100 | 100 | 90.0 | 95.2 | 54.1 | 93.0 | 93.0 | 83.7 | 90.9 |
| 155 | 0.0 | 0.0 | 0.0 | 0.0 | 15.4 | 0.0 | 0.0 | 0.0 | 0.0 | 50.0 | 0.0 | 0.0 | 0.0 | 0.0 | 23.5 |
| 156 | 0.0 | 21.4 | 35.7 | 21.4 | 35.7 | 0.0 | 100 | 100 | 75.0 | 100 | 0.0 | 35.3 | 52.6 | 33.3 | 52.6 |
| 157 | 0.0 | 58.3 | 66.7 | 33.3 | 83.3 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 73.7 | 80.0 | 50.0 | 90.9 |
| 158 | 0.0 | 46.4 | 57.1 | 25.0 | 64.3 | 0.0 | 76.5 | 84.2 | 63.6 | 85.7 | 0.0 | 57.8 | 68.1 | 35.9 | 73.5 |
| 159 | 38.6 | 68.6 | 80.0 | 72.9 | 81.4 | 38.0 | 82.8 | 98.2 | 65.4 | 80.3 | 38.3 | 75.0 | 88.2 | 68.9 | 80.9 |
| 160 | 0.0 | 64.7 | 70.6 | 58.8 | 82.4 | 0.0 | 100 | 100 | 100 | 100 | 0.0 | 78.6 | 82.8 | 74.1 | 90.3 |
| 161 | 2.1 | 31.3 | 45.8 | 35.4 | 50.0 | 6.3 | 48.4 | 68.8 | 58.6 | 60.0 | 3.1 | 38.0 | 55.0 | 44.2 | 54.5 |
| 162 | 20.0 | 60.0 | 73.3 | 66.7 | 80.0 | 35.3 | 100 | 100 | 64.5 | 92.3 | 25.5 | 75.0 | 84.6 | 65.6 | 85.7 |
| 163 | 6.3 | 75.0 | 87.5 | 6.3 | 75.0 | 50.0 | 100 | 100 | 100 | 100 | 11.1 | 85.7 | 93.3 | 11.8 | 85.7 |
| 164 | 0.0 | 83.3 | 83.3 | 66.7 | 83.3 | 0.0 | 90.9 | 100 | 100 | 100 | 0.0 | 87.0 | 90.9 | 80.0 | 90.9 |
| 165 | 20.4 | 57.4 | 57.4 | 53.7 | 64.8 | 35.5 | 70.5 | 91.2 | 76.3 | 61.4 | 25.9 | 63.3 | 70.5 | 63.0 | 63.1 |
| 166 | 27.3 | 68.2 | 68.2 | 77.3 | 72.7 | 66.7 | 100 | 78.9 | 89.5 | 100 | 38.7 | 81.1 | 73.2 | 82.9 | 84.2 |
| 167 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 168 | 3.3 | 43.3 | 60.0 | 40.0 | 70.0 | 33.3 | 76.5 | 81.8 | 60.0 | 61.8 | 6.1 | 55.3 | 69.2 | 48.0 | 65.6 |
| 169 | 62.5 | 81.3 | 85.4 | 81.3 | 91.7 | 75.0 | 88.6 | 100 | 62.9 | 81.5 | 68.2 | 84.8 | 92.1 | 70.9 | 86.3 |
| 170 | 9.5 | 76.2 | 81.0 | 66.7 | 76.2 | 66.7 | 100 | 100 | 87.5 | 88.9 | 16.7 | 86.5 | 89.5 | 75.7 | 82.1 |
| 171 | 44.7 | 63.2 | 86.8 | 71.1 | 81.6 | 58.6 | 70.6 | 97.1 | 73.0 | 83.8 | 50.7 | 66.7 | 91.7 | 72.0 | 82.7 |
| 172 | 25.0 | 75.0 | 78.6 | 60.7 | 85.7 | 58.3 | 100 | 100 | 73.9 | 88.9 | 35.0 | 85.7 | 88.0 | 66.7 | 87.3 |
| 173 | 0.0 | 73.9 | 73.9 | 47.8 | 78.3 | 0.0 | 85.0 | 94.4 | 68.8 | 78.3 | 0.0 | 79.1 | 82.9 | 56.4 | 78.3 |
| 174 | 31.3 | 87.5 | 87.5 | 87.5 | 87.5 | 83.3 | 100 | 100 | 100 | 100 | 45.5 | 93.3 | 93.3 | 93.3 | 93.3 |
| 175 | 6.3 | 25.0 | 31.3 | 31.3 | 31.3 | 50.0 | 57.1 | 71.4 | 83.3 | 100 | 11.1 | 34.8 | 43.5 | 45.5 | 47.6 |
| 176 | 63.9 | 72.2 | 75.0 | 83.3 | 86.1 | 60.5 | 100 | 100 | 76.9 | 93.9 | 62.2 | 83.9 | 85.7 | 80.0 | 89.9 |
| 177 | 18.2 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 91.7 | 100 | 30.8 | 100 | 100 | 95.7 | 100 |
| 178 | 0.0 | 17.4 | 69.6 | 30.4 | 82.6 | 0.0 | 57.1 | 29.1 | 87.5 | 82.6 | 0.0 | 26.7 | 41.0 | 45.2 | 82.6 |
| 179 | 10.0 | 86.7 | 86.7 | 86.7 | 86.7 | 100 | 89.7 | 100 | 86.7 | 96.3 | 18.2 | 88.1 | 92.9 | 86.7 | 91.2 |
| 180 | 18.8 | 68.8 | 68.8 | 43.8 | 75.0 | 75.0 | 91.7 | 84.6 | 100 | 85.7 | 30.0 | 78.6 | 75.9 | 60.9 | 80.0 |
| 181 | 50.0 | 25.0 | 54.5 | 72.7 | 81.8 | 91.7 | 52.4 | 88.9 | 88.9 | 92.3 | 64.7 | 33.8 | 67.6 | 80.0 | 86.7 |
| 182 | 68.2 | 59.1 | 72.7 | 72.7 | 81.8 | 88.2 | 76.5 | 84.2 | 94.1 | 100 | 76.9 | 66.7 | 78.0 | 82.1 | 90.0 |
| 183 | 81.0 | 77.1 | 73.3 | 93.3 | 97.1 | 58.2 | 61.4 | 65.3 | 57.0 | 55.4 | 67.7 | 68.4 | 69.1 | 70.8 | 70.6 |
| 184 | 29.4 | 47.1 | 47.1 | 35.3 | 64.7 | 62.5 | 80.0 | 100 | 66.7 | 100 | 40.0 | 59.3 | 64.0 | 46.2 | 78.6 |
| 185 | 0.0 | 14.3 | 14.3 | 0.0 | 14.3 | 0.0 | 66.7 | 100 | 0.0 | 100 | 0.0 | 23.5 | 25.0 | 0.0 | 25.0 |
| 186 | 9.1 | 45.5 | 63.6 | 9.1 | 54.5 | 33.3 | 100 | 100 | 100 | 100 | 14.3 | 62.5 | 77.8 | 16.7 | 70.6 |
| 187 | 8.3 | 33.3 | 33.3 | 16.7 | 33.3 | 100 | 100 | 100 | 100 | 100 | 15.4 | 50.0 | 50.0 | 28.6 | 50.0 |
| 188 | 42.9 | 64.3 | 78.6 | 85.7 | 82.1 | 60.0 | 100 | 100 | 80.0 | 92.0 | 50.0 | 78.3 | 88.0 | 82.8 | 86.8 |
| 189 | 0.0 | 26.7 | 13.3 | 6.7 | 6.7 | 0.0 | 80.0 | 100 | 100 | 100 | 0.0 | 40.0 | 23.5 | 12.5 | 12.5 |
| 190 | 37.5 | 12.5 | 35.0 | 50.0 | 47.5 | 51.7 | 41.7 | 82.4 | 44.4 | 76.0 | 43.5 | 19.2 | 49.1 | 47.1 | 58.5 |
| 191 | 29.8 | 52.6 | 68.4 | 35.1 | 52.6 | 34.7 | 42.3 | 54.9 | 43.5 | 54.5 | 32.1 | 46.9 | 60.9 | 38.8 | 53.6 |
| 192 | 30.2 | 35.8 | 52.8 | 26.4 | 37.7 | 34.0 | 51.4 | 73.7 | 45.2 | 62.5 | 32.0 | 42.2 | 61.5 | 33.3 | 47.1 |
| 193 | 25.9 | 25.9 | 51.9 | 33.3 | 63.0 | 41.2 | 100 | 93.3 | 45.0 | 77.3 | 31.8 | 41.2 | 66.7 | 38.3 | 69.4 |
| 194 | 14.3 | 14.3 | 42.9 | 21.4 | 28.6 | 33.3 | 100 | 100 | 60.0 | 80.0 | 20.0 | 25.0 | 60.0 | 31.6 | 42.1 |
| W.A. | 35.4 | 58.6 | 68.8 | 60.0 | 72.6 | 34.4 | 65.9 | 75.8 | 64.8 | 75.5 | 31.1 | 58.3 | 68.8 | 57.6 | 71.8 |

## IV. Fold recognition for 710 folds

The current version (1.75) of SCOP database has 1195 folds, which is significantly more than those (27, 94, and 194 folds) we discussed. This fact makes it difficult for real-world use of TAXFOLD. In order to make prediction for as more folds as possible, we decrease the threshold when constructing dataset from the 10493 sequences in the Datasets Section. Those folds containing at least 2 sequences are used to construct classifiers. As a result, we obtain a large dataset consisting of **710** folds, which contain **10011** sequences. For convenience, this dataset is called **F710**. By using this large dataset, the likelihood of a newly sequenced protein to be the prediction target of TAXFOLD should be about 60% (710/1195).

With the dataset F710, we are able to perform fold recognition for more 710 folds, which is an approximation of real-world situation. We assess the prediction accuracy of TAXFOLD on this dataset based on 2-fold cross-validation. The overall accuracy 68.1% is achieved and the distribution of F-measures for individual folds is shown in Figure SC2. There are 417 folds that can be predicted with F-measures higher than 60%. On the other hand, the F-measures for 185 folds are lower than 10%. Among these folds, there are 183 folds containing less than 10 sequences, which is a major reason for the low accuracies. In general, it is hard for taxonomic methods to make accurate predictions for folds that contain only small number of sequences. It is anticipated that when more sequences are accumulated in the SCOP database, we are able to improve the prediction and at the same time make predictions for more folds.

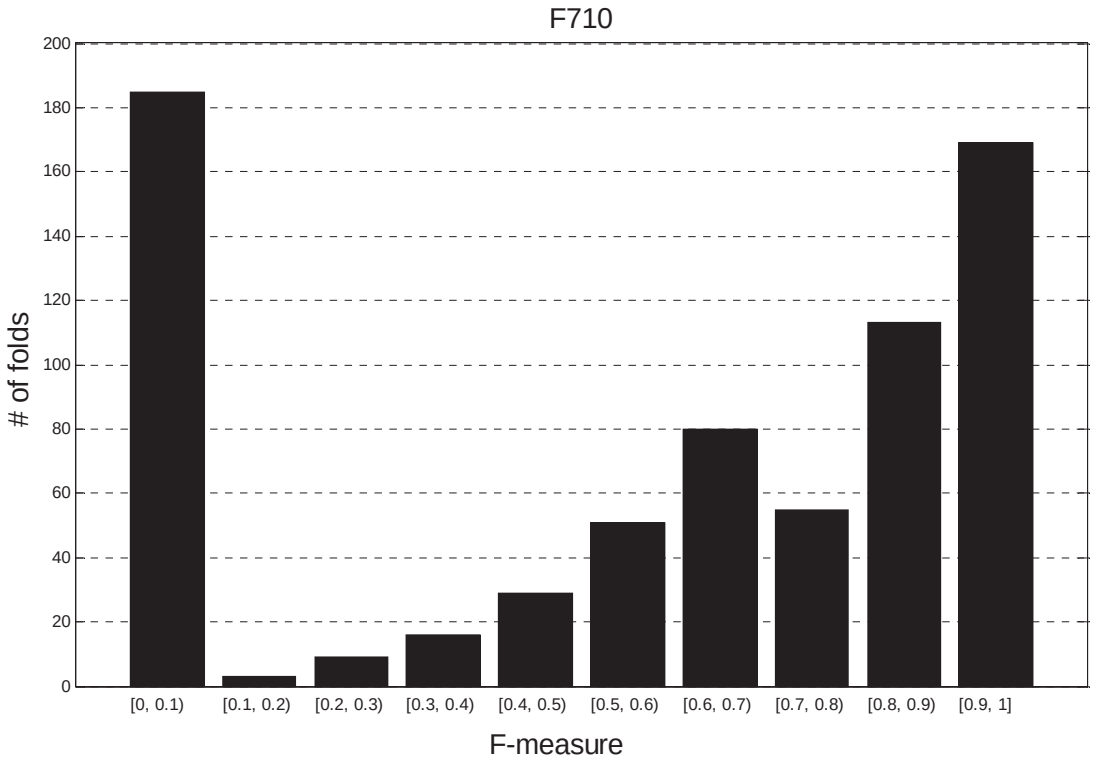



**Figure SC2** Histogram of the F-measure values for the dataset F710. The parameters $C$ and $\gamma$ in SVM are 32 and 0.25, respectively.