Table S4. DOS gene reference sequences used in HMM construction.

BetC (HMM cutoff score = 410)

```
>WP_011046830.1 choline-sulfatase SPO1083 [Ruegeria pomeroyi]
MTRPNILILMVDQLNGTLFPDGPADWLHAPNLKRLAARSTRFANAYTASPLCAPGRASFMSGQLPSRTGVYDNAAEFRSDIPTYAH
HLRRAGYYTCLSGKMHFVGPDQLHGFEERLTTDIYPADFGWTPDYRKPGERIDWWYHNMGSVTGAGVAEITNQMEYDDEVAYHAKA
KLYDLARGKDDRPWCVTVSFTHPHDPYVARRKYWDLYEDCEHLLPQVGDLGYDNHDPHAKRIFDANDWRNFDITEEDIRKSRRAYF
ANISYLDDKLGEILEVLETTRQEAIILFVSDHGDMLGERGLWFKMNFYEGSARVPLMVAAPGMEPGRIDTPVSTIDVTPTLGELAG
VDMAEIAPWTDGTSLVSLANGTERTAPVLMEYAAEASYAPLVSMRYGKWKYNRCALDPDQLFDLDADPHEMTNLADHPDHQGTVQS
LRAKSQARWDLDRFDAEVRASQARRWVVYEALRQGGYYPWDYQPLQKASERYMRNHMNLDNLEESQRFPRGE
```

```
>WP_009828548.1 choline-sulfatase [Rhodobacteraceae bacterium HTCC2083]
MSKLNILIFMVDQLNGTFFPDGPADWLHVPNIKKLAARSTRFKNAYTASPLCAPGRASYMSGQLPSTTGVYDNAAEFPSSIPTYAH
HLRRAGYQTCLSGKMHFVGPDQLHGFEERMTTDIYPPDFGWTPDYRKPGERIEWWYHNMGSVTGSGIAEITNQMEYDDEVAYHATR
KIYDYARGKDERPWCMTVSFTHPHDPYVARRKYWDLYEDCAHLMPEVGPIPYEDQDAHSQRILDANDRDNFNISDKDIRRSRRAYF
ANISYLDDKLGEVMEALESTRQEAIIMFVADHGDMLGERGLWFKMSFFEGSARVPMMISSPDMEPGLIETPVSNIDVCPTLCDLAG
VDMEEIMPWTEGMSLKPMGQGVERTAPVAMEYAAEASYSPLVCLRYGKWKHIRCALDPDQLFDLDADPHELTNLADAPEHQGTLQT
LRAKSEQRWDLDRFDAEVRRSQARRWVVYEALRNGDYFPWDYQPLQKASERYMRNHMDLNVLEESQRFPRGE
```

```
>WP_008035825.1 choline-sulfatase [Rhodobacterales bacterium HTCC2255]
MTKPNLLIFMVDQLNGTFFPDGPADWLHAPNLKKLAGKSTRFKNCYTASPLCAPGRASFMSGLLPSKSRVYDNAAEFSSDIPTYAH
HLRRAGYQTCLSGKMHFVGPDQMHGFEERLTTDIYPADFGWTPDYRKPGERIDWWYHNMGSVTGSGIAEISNQMEYDDEVAYNAKR
KIYDLSKGKDNRPWSLTVSFTHPHDPYVARKKYWDLYEDCAHLMPTVSAFEYEKHDTHSKRIFDANNWRDFDITDDDIKRSRRAYF
ANISYLDDKIGEILEALETTKQEATILFLSDHGDMLGERGLWFKMSFYEGSARVPLMVCSPKMKPGLVTYPVSNIDLAPTLCDLAD
VSIEEIAPWCEGESLVHLGQGGERTSPVAMEYAAEASYSPMVSLRYGPWKYNRCLLDEDQLFNLDEDPNELINLVNTHKHQGTLNQ
LRVKSEKRWNLNKFDTDVRESQAKRWIVYEALRNGNYFPWDYQPLQDASERYMRNHMDLNEVENDNRFPRGE
```

```
>2524484855 choline-sulfatase [Donghicola xiamenensis DSM 18339 :
G455DRAFT_scaffold00007.7]
MSQPNILIVMVDQLNGTLFPDGPVDWLHVPNLKKLASRSSRFANAYTASPLCAPGRASFMSGQLPSRTGVYDNAAEFRSDIPTYAH
HLRRAGYYTCLSGKMHFVGPDQLHGFEDRLTTDIYPADFGWTPDYRKPGERIDWWYHNMGSVTGAGVAEISNQLEYDDEVAYHARA
KLYDLARGRDGRPWCLTVSFTHPHDPYVARRKYWDLYADCDHLQPDIPAIPYEDQDAHSKRIFDANDWRNFDITPRDIERSRRAYF
ANISYLDDKIGEILQVLEETQQEAIVVFLSDHGDMLGERGLWFKMNFFEGSARVPLMISAPGLPVGRVETPVSTLDVVPTLGALAG
VDLGEIAPWTDGEDLCPLARGARRDTPVALEYAAEASESPLVCLRQGQWKYTRCLLDPEQLFDLEADPHELTNLAADPAHADTLER
FREMAGAHWQLDRFDAEVRASQARRWVVYEALRQGGYYPWDYQPLQKASERYMRNHMDLNVLEDSQRFPRGD
```

```
>WP_093152573.1 choline-sulfatase [Thalassobaculum litoreum]
MARPNILILMVDQLNGTLFPDGPAPWLHTPNLRRLAERSTRFANAYTASPLCAPGRASFMSGQLPSLTGVYDNAAEFSSAIPTYAH
HLRRAGYQTCLSGKMHFVGPDQLHGFEERLTTDIYPADFGWTPDYRKPGERIDWWYHNLGSVTGAGVAEISNQMEYDDEVAYHAVR
KLYDLSRGRDGRPWCLTVSFTHPHDPYVARRKYWDLYEDCEHLQPTVPAMDYEDHDPHSQRIFDANDWRSFDITEENIRRSRQAYF
ANISYLDDKIGEILDTLERTRQEAIILFVSDHGDMLGERGLWFKMSFFDGSSRVPLMISAPGMAPGLIAAPASTLDVTPTLAELAG
ISMEEVTPWTEGKSLVSVPGGTPRPGSVAMEYAAEGSYAPLVGLRRGRWKYTRCVLDPEQLFDLEADPHELTNLAADPAHADTLAE
FRAEADGRWDLARFDAEVRSSQARRWVVYEALRNGAYYPWDHQPLLKASERFMRNHMDLNVVEENQRFPRGE
```

CuyA2 (HMM cutoff score = 271)

```
>WP_084790992.1 D-cysteine desulfhydrase SPO2657 [Ruegeria pomeroyi]
MPIQGGIVKTFVDAPVSGAVDDQALLRRALERFARIRLGHLPSPLEPLDRLSEALGGPRIWVKRDDCTGLSSGGNKTRKLEFLMAS
ACEAGADTIITQGAIQSNHARQTAAAAARLGLHCHLLLEDRTGAADLAYTLNGNVLLDRLHGATISRRAGGSDMAAEMERLAATLR
GKGKRPYVIPGGGSNRVGALGYVNCALELSEQALSMGLDIDALIHATGSCGTQAGLVAGLAGLSWNTHLLGIGVRAPQERQEAMVF
DLACETMDYLGMGIPIVRSGVRANCDHVGPGYGLPTEAMCEAVRMLARLEGLLLDPVYTGKAMAGLIDLTRQGYFDGMANIVFLHT
GGSAALFGYPEIFGLPGYQP
```

```
>WP_008033848.1 D-cysteine desulfhydrase [Rhodobacterales bacterium HTCC2255]
MTNATQTNEVEMPFSPEMEQSLCKDLSLTLAKFPKVRLGHLPTPLEPMDRLSEILGGPRLWVKRDDCTGLSSGGNKTRKLEFLMAD
AQSKGADTIITQGATQSNHARQTTAAAAKLGMECHILLEDRTGSNDHSYILNGNVLLDRLHGASVSKRSGGTDMNAEMQDFAESLI
EKGKNPYIIPGGGSNAIGALGYVNCARELTEQASEIGLKVDALVHATGSAGTQAGLVAGLAAIQSDIHLLGIGVRAPKDKQEQMVF
DLAQKTADYLDTGIKIERDKVRAICDYVGAGYGLPTDGMIKAVKLLAQSEGLLFDPVYSGKGLDGLIDQIKKGYFAGMDNVVFLHT
GGSAALFGYPETFELPGYAN
```

```
>CUI00049.1 L-cysteate sulfo-lyase [Leisingera aquaemixtae]
MNTAVTEDTMAGGGNLALVKRALENFPRVRLGHLPTPLEPLDRLSEELGGPRIWVKRDDCTGLASGGNKTRKLEFLMAEALKLGAD
TVITQGATQSNHARQTAAAAAKLGLDCQILLEDRTGSNDVAYNFNGNVLLDQLHGATISKRAGGADMNAEMETLAGALQDQGRKTY
IIPGGGSNRVGALGYVNCAAELEEQAGALGLKVDALVHATGSSGTQAGLVAGLAGLASDIHLLGIGVRAPQEKQESMVFDLACKTM
```

```
DYLGMGAEINRASVRANCDYVGPGYGLPTEGMRAALLKLARLEGLLFDPVYSGKGLDGLIDQVRKGYFDGMDNIVFLHTGGSAALF
GYPGIFGLPGYQD

>2236438413 alpha proteobacterium sp. SCGC AAA076-
C03MGRLRIKHYFFVSVFVIVSVNHIKGINLVTNATQTNEVEMPFSPEMEQSLCKDLSLTLAKFPKVRLGHLPTPLEPMDRLSEIL
GGPRLWVKRDDCTGLSSGGNKTRKLEFLMADAQSKGADTIITQGATQSNHARQTTAAAAKLGMECHILLEDRTGSNDHSYILNGNV
LLDRLHGASVSKRSGGTDMNAEMQDFADSLIEKGKNPYIIPGGGSNAIGALGYVNCARELTEQASEIGLKIDALVHATGSAGTQAG
LVAGLAAIQSDIHLLGIGVRAPKDKQEQMVFDLAQKTADHLDTGIKIERDKVRAICDYVGAGYGLPTDGMIKAVKLLAQSEGLLFD
PVYSGKGLDGLIDQIKKGYFAGMDNVVFLHTGGSAALFGYPETFELPGYAN

>2624356654 Rhizobiales sp. genome_bin_27
MNLARFPRVRLGHMPTPLEPMDNLSKHLGGPRIWVKRDDCTGLSSGGNKTRKLEFLMAEALAQKADVVITQGATQSNHARQTAAAA
AKLGMACMILLEDRTGYTDEAYTLNGNVLLDQLHGASISKRKSGADMQAEMESVAADLRKTGGRPYVIPGGGSTAVGALGYVNCAI
ELLNQANEMGLKIDLLVHATGSAGTQAGLVAGLVALNSGIPVLGIGVRAPKDRQEANVLALAEKTAAHLGLAGIMKREHVVANCDY
VGQGYGIPTEGMVEALKLVARMEGILLDPVYSGKGMAGLIDLIKKGEMREHENIVFLHTGGSVGLFGYPTAFGLQGYQA
```

DddK (HMM cutoff score = 188)

```
>gi|499600944|ref|WP_011281678.1| cupin [Candidatus Pelagibacter ubique HTCC1062]
/dddK/
MIFVKNLASVLSQEWSSTEKYPGVRWKFLIDADFDGSSGLSLGFAEIAPGGDLTLHYHSPAEIYVVTNGKGILNKSGKLETIKKGD
VVYIAGNAEHALKNNGKETLEFYWIFPTDRFSEVEYFPAKQKSG

>gi|654569998|ref|WP_028037226.1| cupin [Candidatus Pelagibacter ubique HTCC9022]
/dddK/
MIFVNNLKSVSDQEWSSTEKYPGVRWKFLIDADYTKSSGLSLGFAEIAPGGDLTLHYHSPAEIYVVTNGTGILNKSGQLEEIKKGD
VVYIAGNAKHALKNNGKETLEFYWIFPTDRFSEVKYLS

>WP_014953073.1 cupin domain-containing protein [alpha proteobacterium HIMB5]
MIFIKNMNSVSDQDWTTSEKYPGVRWKFLIDEDYNGSKGLSCGFAEIEPGGNLTLHHHAPDEIYVVTNGS
GTLNKSGELEEIKKGDVVYIAGNAKHALQNNGKEVLGFYWVFPTNKFKDVEYISDE

>WP_027306832.1 cupin domain-containing protein [Candidatus Pelagibacter ubique
HTCC1013]
MIFVKNLASVLSQEWSSTEKYPGVRWKFLIDADFDGSSGLSLGFAEIAPDGDLTLHYHSPAEIYVVTNGK
GILNKSGKLETIKKGDVVYIAGNAEHALKNNGKETLEFYWIFPTDRFSEVEYFPAKQKSG

>WP_018413735.1 cupin domain-containing protein [Candidatus Pelagibacter ubique
HTCC1016]
MIFVKNLASVLSQEWSSTEKYPGVRWKFLVDADFDGSSGLSLGFAEIVPGGDLTLHYHSPAEIYVVTNGT
GILNKSGKLETIKKGDVVYIAGNAEHALKNNGKETLEFYWIFPTDRFSEVEYFPAKQKSV

>WP_006997514.1 cupin domain-containing protein [Candidatus Pelagibacter ubique
HTCC1002]
MIFVKNLASVLSQEWSSTEKYPGVRWKFLVDADFDGSSGLSLGFAEIVPGGNLTLHYHSPAEIYVVTNGT
GILNKSGKLETIKKGDVVYIAGNAEHALKNNGKETLEFYWIFPTDRFSEVEYFPAKQKSV

>2503364858 IMG accession Cupin domain-containing protein [Candidatus Pelagibacter
ubique SAR11 HTCC9565 : HTCC9565_HTCC9565_contig2]
MIFVNNLKSVSDQEWSTTEKYPGVRWKFLIDADFAGSSGLSLGFAEIAPG
GDLTLHYHSPAEIYVVTGGSGILNKSGELEKIKEGDVVYIAGNAKHALKN
NGKDTLKFYWIFPTDRFSEVEYFN
```

DddP (HMM cutoff score = 382)

```
>gi|83950505|ref|ZP_00959238.1| metallopeptidase, family M24 [Roseovarius nubinhibens
ISM] /dddP/
MNQHYSETRKIDPSRGATLGDNTPNDNNRIEIGPTQLAFGEWATAGLALPDLQRMREFRWNRLTQAVVDRDYGGVLMFDPLNIRYA
TDSTNMQLWNAHNPFRALLVCADGYMVIWDYKNSPFLSTFNPLVREQRFGADLFYFDRGDKVDVAADAFSNEVRTLIAEHGGGNMR
LAVDKIMLHGLRALEAQGFEIMEGEELTEKTRAIKGPDEILAMRCAVHACETSVAAMEHFAREAVPQGNTSEDDVWAVLHAENIKR
GGEWIETRLLASGPRTNPWFQECGPRIIQNNEIISFDTDLIGSYGICVDISRSWWVGDAAPPADMVYAMQHAHEHIMTNMEMLKPG
VTIPELSERSHRLDEQFQAQKYGCLMHGVGLCDEWPLVAYPDQAVPGSYDYPLEPGMVLCVEAAVGAVGGNFTIKLEDQVLITETG
YENLTSYPFDPALMGR
```

```
>gi|163736748|ref|ZP_02144167.1| peptidase M24 [Phaeobacter gallaeciensis BS107]
/dddP/
MSSDTFETMSNTEPEMNEHYRDTRKIDPTRGATLGDNTPNDQDRVEIGPTQLAFGEWAAAGLQLPDLQAMRRYRWERLTRFINDRD
YAGLLVFDPMNIRYATDSTNMQLWNTHNPFRALLICADGYMVMWDYKQAPFLSEFNPLVREQRAGADLFYFDRGDKVDVAADAFAN
EVRTLLAEHSGGNTRLAVDKIMLHGLRALEAQGLEVFPGEELTEKCRAVKGPDEILAMRCANHACETTVAEMERYARSAIPGGQIS
EDDVWAVLHAENIRRGGEWIETRLLTSGPRTNPWFQECGGRIIQNNEIISFDTDLVGSYGICIDISRSWWIGDRAPPADMVYAMQH
GVEHIQSNMEMLKPGVNLQELSRNCHLLDAQFQKQKYGCMMHGVGLCDEWPLVAYPDAMVEGAFDYDLEPGMVLCVEALVSPEGGD
FSIKLEDQVLITETGYENLTTYPFDPALMGTTR

>gi|294084994|ref|YP_003551754.1| peptidase M24 [Candidatus Puniceispirillum marinum
IMCC1322] /dddP/
MNQLIVGSNRKIDPTRRLHLKPDNTPDDNDRVEIGPTALAFEEWKQLGLTAPDMPALRAYRLERLQQQIRIHDCAGLLLFDPLNIR
YATDATNMQLWTSHNMARACFVPPEGKMILWDFHNCEHLSAHLPLVGELRGGASFFYFETGDRTAEAAKAFADQMLDIMHHYAPGN
KRLAVDKMENLGYAALVGLGVEVLEGQVLTEHGRSIKNENELNAMRCAIATCELAVEEMRDEMRAGISENELWATLHAGNIKRGGE
WIETRILSSGPRTNPWFQECGPRIMQDGDLMAFDTDLVGTYGYCCDISRTWLVGDGSPSDAQKHLYQVAYDHVMTNIGLIEPGMRF
ADMTRIAHRLPEEYRALRYGVLAHGVGLCDEYPSVRYPEDVEHHGYGGCFEVGMTLCVEAYVGAVGGRDGVKLEEQVVVTDQGAIP
LSTYRYEDAFLS
```

DddW (HMM cutoff score = 91)

```
>gi|56695368|ref|YP_165716.1| ortholog hypothetical protein SPO0453 [Ruegeria pomeroyi
DSS-3] /dddW/
MTAMLDSFATDLTAATSLHQPGNLPHPPHAIDAENVPLSGGTDPTYGEVRWRTLINGTEAAPRDMVLGIAEFGPGHQLRPHRHTPP
EFYLGLEGSGIVTIDGVPHEIRAGVALYIPGDAEHGTVAGPEGLRFAYGFASASFEAIEYRFTASA

>gi|86139257|ref|ZP_01057827.1| ortholog hypothetical protein MED193_09710
[Roseobacter sp. MED193] /dddW/
MQQETTLDFAPDMTTLATDLGETEQFHQASNLPHAPQDTLLVVDAAQVPLSGGTDAAYGEVRWRTLINSCSEKSREMVLGIAEFEP
HGRLLPHRHDPPEFYLGLEGSGVVTIDGTPHEIRPGVAIYVPANAEHDTQAGPDGLRFTYGFAEASFEQITYRFSATTEARG

>gi|260426459|ref|ZP_05780438.1| conserved hypothetical protein [Citreicella sp. SE45]
/dddW/
MEDPMAEPGIDVANLPQIAALMTLAVAANARGAPLEGGGDPAFGTVRWRSLFSDDRTPTRGMVPGIAEFAPFGTLLPHRHGHEKIC
FGLSGGGVVTIDGVPHPLGPEVALYVPEDAEHGTVAGPGGLRFL
```

DmdA (HMM cutoff score = 411)

```
>gi|294083644|ref|YP_003550401.1| aminomethyl transferase [Candidatus Puniceispirillum
marinum IMCC1322] /dmdA/
MAQSGLNMSRRIRRSPFTDKVEEYGVRGFSVVNHMLLPKAFETSVEDDYWHLREHVQIWDVGVQRQVQITGLDAARLVQMMTPRDV
RQAKIGQCLYVPMIDEDAGMLNDPVLIKLADDKFWLSIADSDILLWVKGLALGLKLNVDVEEPDVSPLAIQGPKAIALMADLFGEA
IRDLGYFQYGIFDVLGTRQLIARSGYSKQGGFEIYLHGGHLGSDLWDMIYQAGKQYNIMPGCPNLIERIEGGLMSYGNEFTRDNNP
LECGFEELCYFGDDIDYIGKIALRRIAEEGPQKLIRGIKFGGGKAPPCGKPFLVTTRDNIHIGQITSGIYSPRLKCNVGMSMMAKG
HWDFGTVVFVHTPDGIVREGTVSPLPF

>gi|114771227|ref|ZP_01448647.1| aminomethyl transferase family protein [alpha
proteobacterium HTCC2255] /dmdA/
MASAMIFPSRRLRATPFTSRVSKLGVSGFTVYNHMLLPTVFESLQEDYKHLKEYVQMWDVSVERQVQLLGKDAHKLACMISARDLT
NAQTGRCYYAPICDQSGAIINDPIALRLADDKYWFSIADSDLLLWVQGIALGLDLNVEICEPDVSPLAIQGPMAEDLMVDVFGAEI
RNIKFFHFKEFPFNGRMLNIARSGWSKQGGFEIYLNDSQLGPELWDTIWEKGEKYNIRPGCPNLIERIEAGLLSYGNDMNREDSPL
EIGLEKYISLDSNVDFIGKKALLKQRKDGIKKRLLGIEIDGSEMPPLSMPEEVFKDGKKIGIVTSAVFSPDYNGNIGFAMIEASNA
TAGTEVSVDSKAGIRKGKLCEIGDFWSQVQSKN

>gi|254456019|ref|ZP_05069448.1| aminomethyltransferase [Candidatus Pelagibacter sp.
HTCC7211] /dmdA/
MKKFPIAKSRRLRSTPYTDRIERQGVSSYTVYNHMLLPASFVSVEADYHHLKEFVQVWDVAAERQVEISGKDSAQLVQLMTCRDLS
KSKVGKCYYAPIIDGQGNLVNDPIINKLAENRWWLSIADSDVIFFAKGLASGNKFDVDIKEPDVNILAVQGPLSDKLMSKVFGEKI
SQLKFFNFDYFEFKGMKHFIARSGWSKQGGFEIYVENAEAGKELYDYLFEAGLEFNVKPGCPNLIERIEGALLSYGNDFDNRDNPL
EANFEKYTNLDSEVEFLGKDRLKKIRDKGVKRKLMGVKIDHDQIDMYCEKTLFDDNNNIIGFVRSATYSPTFKKVIGIAMINKPYW
NSKNSFKIEINEKIHLGNICDLPFI

>gi|254456670|ref|ZP_05070099.1| probable aminomethyltransferase, putative [Candidatus
Pelagibacter sp. HTCC7211] /dmdA/
MSKNIKLNMSRRIRRTPYTNRVEQHGVSDFTVVNHMLLPKGFKNTVEEDYLHLSKEVQMWDVSCQRQVQICGPDAAKLIQKLTPRS
IKDMTIGKCFYIPMLNENAGMINDPVLLKLDDDMFWISIADSDILLWAKGLALGLNLNVVIEEPDVYPLAIQGPKSEELMVSIFGD
```

```
EIKKIKFFNFRVIDFEGTKQIIARSGYSKQDGFEIYFKVHENYFDKVEMGEKLWDTIWEAGKKFNISPGCPNLIDRIEAGLMSYGN
DFTGENNPLECNLEKYCKADASHDFVGKQALTKIQSEGIIQKMRGIIFDGAPCAATGQPLKIFSKDNKRIGQITSGIFSPRIKKNI
GLSMILKDYWNVGNEVIIETLDGEKRNGTITSLPFPD

>gi|119503015|ref|ZP_01625100.1| aminomethyl transferase family protein [marine gamma
proteobacterium HTCC2080] /dmdA/
MNAPVISFSRRLRVSPFELRSLEGSKSASVYNHLVLPTCYESLEADYWHLREHVQLWDVACQVQVEVQGPDAAEFVEYLTPRDVSR
CQVGQCIYTPLIDEAAGIINDPLVLRLAEDRFWISLSDSDVLLWAKGLALGKGFDVRVFDPDVFPMSIQGPKSADLLSRVLGDSIR
ELKFFRFVETEIAGTPVVVARTGWSGQGGYEIYLQEPDAGVTLWDTLAAAGEDLQVRVGCPNLIERIESGLLSFGNDMTLANNPLE
AGLDRFFKLGKSADYLGRAALEAIAEEGVKNRLVKLVIEGEPIANPRTVYTVQGESGENIGTVTSAVYSPRLCCNIGLGYLPVSYC
DEGKAAIVLTPQGPRELRIANNDWSS

>gi|86139921|ref|ZP_01058486.1| aminomethyl transferase family protein [Roseobacter
sp. MED193] /dmdA/
MAFISPSRRLRRTPFSEGVEAAQVKAYTVYNHMLLPTVFESPEADYHHLKKHVQIWDVSCERQVELRGPDAGRLMQMLTPRDLRGM
LPGQCYYVPIVDETGGMLNDPVAVKLAEDRWWISIADSDLLYWVKGIANGWRLDVLVDEPDVSPLAIQGPKSEELLVRVFGESIRS
IRFFRFGTFQFQGRDLVIARSGYSKQGGFEIYVEDSEIGMPLWNKLFEAGQDLEVRAGCPNLIERIEGGLLSYGNDMTDDNTPHEC
GLGRFCDTHTAIGCIGRDALLRVAKEGPVQQIRAISIAGEPVPACTEFWPLYAGGKRVGRVSSAAWSPDFRTNVAIGMVRMTHWDA
GAKLEVETPDGMRQATVREKFWI

>gi|56696787|ref|YP_167148.1| dmdA gene product [Ruegeria pomeroyi DSS-3] /dmdA/
MASIFPSRRVRRTPFSAGVEAAGVKGYTVYNHMLLPTVFDSLQADCAHLKEHVQVWDVACERQVSIQGPDALRLMKLISPRDMDRM
ADDQCYYVPTVDHRGGMLNDPVAVKLAADHYWLSLADGDLLQFGLGIAIARGFDVEIVEPDVSPLAVQGPRADDLMARVFGEAVRD
IRFFRYKRLAFQGVELVVARSGWSKQGGFEIYVEGSELGMPLWNALFAAGADLNVRAGCPNNIERVESGLLSYGNDMTRENTPYEC
GLGKFCNSPEDYIGKAALAEQAKNGPARQIRALVIGGEIPPCQDAWPLLADGRQVGQVGSAIHSPEFGVNVAIGMVDRSHWAPGTG
MEVETPDGMRPVTVREGFWR

>gi|71082952|ref|YP_265671.1| dmdA gene product [Candidatus Pelagibacter ubique
HTCC1062] /dmdA/
MKNFSIAKSRRLRSTPYTSRIEKQGVTAYTIYNHMLLPAAFGSIEDSYKHLKEHVQIWDVAAERQVEISGKDSAELVQLMTCRDLS
KSKIGRCYYCPIIDENGNLVNDPVVLKLDENKWWISIADSDVIFFAKGLASGHKFDVKIVEPVVDIMAIQGPKSFALMEKVFGKKI
TELKFFGFDYFDFEGTKHLIARSGWSKQGGYEVYVENTQSGQKLYDHLFEVGKEFNVGPGCPNLIERIESALLSYGNDFDNNDNPF
ECGFDQYVSLDSDINFLGKEKLKEIKLKGPQKKLRGVKIDIKEISLTGSKNIYDENNNVIGELRSACYSPHFQKVIGIAMIKKSHW
EASQGFKIQINDNTINGNVCDLPFI

>gi|399156270|ref|ZP_10756337.1| dimethyl sulfoniopropionate demethylase [SAR324
cluster bacterium SCGC AAA001-C10] /dmdA/
MRPELLISSRTRSTPFSSRVEACGVKAYSVYNHMLLPLIFRSLEEDYWHLCESVQVWDVSCQRQVEITGPDTQKLVQLMTPRDLSQ
AELGQCFYAPLCDETGGMINDPILIKHSNNHWWLSIADSDVMLWAKGLATGFGLDALVTEPDIWPLAVQGPKAEELLSRVFGKEIS
KILFFRSSTFNYQGTKMLVARSGWSKQGGFEIYVNDAELGGQLWDELFAKGEDLNVGPGCPNLIERIESGLLSFGGDMGYDTTPFE
CGLEKYVSLDADIESLSLDELRTRKSKTKLIGIVIDRKVNLINKRVFIGSNVVGKITSDTWSPRYSAFLAFARCELKHLEDAQNXG
IDXGTXX
```

HpsN (HMM cutoff score = 682)

```
>RpomDSS3 [Ruegeria pomeroyi DSS-3]2,3-dihydroxypropane-1-sulfonate dehydrogenase
MTIEYLKKASLTSKSDASDVQETVRAILADIEAGGDQVALDYAAKFDRYEGSIILSPEEIEAACAKVPEK
LKADIRFAHDNVRRFAETQKATLTDVELEVVPGVITGQKAIPVDAAGCYVPGGRYSHIASAIMTVTTAKV
AGCKHIMACSPPRPGVGVAPAIVYAAHICGADTIMAIGGVQGVASMAFGLFGLPKAKILVGPGNQFVAEA
KRMLFGRVGIDMIAGPTDSLILADRTADPHIVTTDLVSQAEHGYNSPVWLVTDDRALAEKVIEMIPSYIA
DLPEVNRDNAAAAWRDYAEVILCADREEMAATSDRYAPEHLTVMAEDLDWWLDRLSCYGSLFLGEESTVS
YGDKAAGTNHVLPTSGAASYTGGLSVHKYMKIVTWQRGTREGYKPVAEATARIARLEGMEGHARAADVRL
AKYFPDETFDLTANG

>EDQ35720.1 histidinol dehydrogenase [Hoeflea phototrophica DFL-43]
MSRIYLKKAEKTASTDSGSVRDTVQSILDAIEAGGEASAREYAEKFDRYTGNIIVTREEIERAAASLPQK
LKDDIRFAHDNVRRFAEAQKGTLGDVELEIVPGLVAGQKTIPIESAGCYVPGGRYSHVASAIMTVTTAKV
AGCKHITVCSPPRPDVGVHPAILYTADICGADTILAMGGVQGVASMAFGLFDVPPAGILVGPGNQFVAEA
KRILFGRVGIDMFAGPTDSLILADKDADPEIVAADLVGQAEHGYNSPVWLVTDDKSLAETVMALVPRLIA
DLPQVNRDNAEAAWRDFAEVILCDSREEMARTSDDYAPEHLTVQAADLDWWLDRLSCYGSLFLGEETTVA
FGDKASGTNHVLPTSGAARYTGGLSVHKFMKIVTWQRATRDGSRPIAEATARISRLEGMEGHARTADIRL
RKYFPLDNFDLSGVDQDI

>ABG33272.1 histidinol dehydrogenase [Roseobacter denitrificans OCh 114]
MAREYLKKAEKTAMTDAGDVREIVQNILSDIEAGGDAAAMRYAEQFDKYAGNVILTDDEIEAASAKVSQK
```

```
LKDDIAFAHDNVRRFAEVQKSTMKDVELEIVPGLIAGQKAIPVRAAGCYVPGGRYSHIASAIMTVTTAKV
AGCQHITACSPPRPDVGINPAIVYAANLCGADKILSMGGVQGVAAMTFGLFGLPKADILVGPGNQFVAEA
KRILFGRVGIDMIAGPTDSLVLADATADPFMVATDLVSQAEHGYNSPVWLVTDSRVLAEKVMELVPGMID
DLPELNRDNAFAAWRDYAEVIVCANREEMAACSDDYAPEHLTVQAEDLPWWLDRLQCYGSLFLGEETTVS
FGDKASGTNHVLPTSRSAKYTGGLSVHKYMKIVTWQRSTRDGARPVAEATARISRLEGMEGHARAADARL
HKYFPNENFDLSAVD

>EAP77625.1 histidinol dehydrogenase [Roseovarius nubinhibens ISM]
MARDYLKKAHLTSRSDASDVHDTVRKILADIEAGGDQAALDYAAKFDKYDGNVMLTAEEIDAACALVPEK
LKADIRFAHDNVKRFAEQQKSTVANVEYEIVPGVIAGQKAIPVDAAGCYAPGGRYSHVASAIMTVTTAKV
AGCKHITACSPPRPGVGVNPAIVYAAHICGADKIMAMGGVQAIAAMTFGLFGLPKANILVGPGNQFVAEA
KRILFGRVGIDMIAGPTDSLILADKTADPHIVATDLVSQAEHGYNSPVWLVTDDRALAEDVMARVPALID
DLPEVNRDNAAAAWRDYAEVILCADREEMAATSDDYAPEHLTVQAEDLDWWLDRLTCYGSLFLGEETTVS
YGDKATGTNHVLPTSGAASYTGGLSVHKYMKIVTWQRATRTGSKPIAEATARISRLEGMEGHARAADVRL
AKYFPDETFDLTAHG

>EAQ46924.1 histidinol dehydrogenase [Roseobacter sp. MED193]
MAREYLKRATLTSKSDASDVHDTVVKILSDIETGGDAAALDYAAKFDKYNGSVLLSAEEIEAACALVPDK
LKADIQFAHDNVKRFAEKQKSTVADVEYEIVPGLVAGQKAIPVDAAGCYAPGGRYSHVASAIMTVTTAKV
AGCKHITCCSPPRPGIGIAPAIVYAAHICGADKIMAMGGVQGVAAMTFGLFGLPKANILVGPGNQFVAEA
KRILFGRVGIDMIAGPTDSLILADKTADPHIVATDLVSQAEHGYNSPVWLVTDDRDLAQDVMSRVPELID
DLPQVNRENATAAWRDYAEVIVCADREEMAATSDDYAPEHLTVQADDLDWWLNRLTCYGSLFLGEETTVS
YGDKAAGTNHVLPTSGAANYTGGLSVHKYMKIVTWQRATREGSKPVAEATARISRLEGMEGHARAADVRL
AKYFPDETFDLTANG

>EAQ24107.1 histidinol dehydrogenase [Roseovarius sp. 217]
MTREYLKKATLTSTSNAGDVHETVQGILADIEAGGDKVALEYAAKFDRYEGKIILTPEDIAAASALVPEK
LKRDIEFSHANVRRFAQAQRETLRDIEVEVVPGLIAGQKSIPVRAAGCYVPGGRYSHIASAIMTVTTAKV
AGCKHIVACSPPRPGVGVAPAIIYAAHVCGADVIMAMGGVQGVAAMTFGLFGLPPANILVGPGNQFVAEA
KRILFGRVGIDMIAGPTDSLVLADATADPMVVATDLVGQAEHGYNSPVWLVTDNRPLAEEVMRLVPDLIA
DLPEVNRDNAAAAWRDYAEVILCADREEMAATSDDYAPEHLTVQAEDLNWWLDRLTCYGSLFLGEETTVA
FGDKASGTNHVLPTSGAANYTGGLSVHKYMKIVTWQRATREGAKPVAEATARISRLEGMEGHARTADIRL
RKYFPDETFDLTANG

>AAZ65418.1 histidinol dehydrogenase (plasmid) [Ralstonia eutropha JMP134]
MISYLKKAEKTPQTETATAQKVVTEMLAEIQARGKDAVRQYAKQLDGWSGDIVLTPDQIREQTKDVPAGV
RADIDFAIRQVTDFALAQRESLKEFSVELHPGVTAGQRVLPVNVVGCYAPAGRYAHIASAYMGVATAKAA
GVKTVVACSSPFRGQGIHPHVLYAFQAAGADVIMALGGVQAIASMAYGLFTGKPADVVVGPGNKFVAEAK
RSLYGQVGIDVFAGPSEVAVIADETADPAIVASDLVGQAEHGHESPAWLFTTSRDLADRVMALVPELIAK
LPPTARDAATAAWRDYGEVILCGTREEVVEISDRYASEHLEVHTADLDWWLANLTCYGSLFLGEETTVAF
GDKTSGPNHVLPTKGAARYSGGLSVHKFMKTLTWQQMTREATRQIGQVTARISRLEGMEAHARTADDRMA
KYFPNASFEMGTPVEV

>2504637969 sulfopropanediol 3-dehydrogenase [Gammaproteobacteria sp. OM252 HIMB30 :
HIMB30_contig00039]
MTIEYLKKAKKTSTSDASDVTKIVQGILNDIESGGEDACLKFRDQFDKYDGNVVLTADEIQAACDQVPEK
LKRDIQFAHGNVKRFAEAQKSTIVDTEVEVIPGLIAGQKAIPCNAAGCYAPGGRYAHIASALMTVTTAKV
AGCQNIVACSPPRPGVGIHPAIVYTAHICGADHIVALGGVQGIAAMAFGLFGLPKADILVGPGNQFVAEA
KRILFGRVGIDMFAGPTDSLILADASADPELVAADLVGQAEHGYNSPVWLVTDHEPLARRVMDLVPEYIS
TLPELNRDNATKAWNDYAEVVVCADREEMAATSDRYAPEHLTVQADDLDWWLDRLTCYGSLFLGNETTVA
FGDKASGPNHVLPTSGAAKYTGGLSVHKYMKLVTWQRSNTDAMKSVAEATARISRLEGMEAHARSADIRL
QKYFPEEVFDLTGTDNG

>ADE40381.1 histidinol dehydrogenase [Candidatus Puniceispirillum marinum IMCC1322]
MTITYLKKASKTPQTGTDETQKIVADMLATIEAGGEEAALEYGRKLDGYSGNAVVSAEQIAAAADAVSPQ
LKDDIQFSLDRVRTFAEAQRDSIKSFETELSPGLWAGQKLIPIETAGCYVPGGRYAHVASAIMSIATAKV
AGVKHIVACSAPNGEDGPNPAILYAMNLCGADTILSLGGVQGIAAMAYGLFTSKPARILVGPGNRFVAEA
KRSLYGRVGIDMFAGPTEIAVIADESADPAVVAEDLVSQAEHGPDSPAWLFTTSQKLADDVMAQMQTHIE
ALPEIARNAATAAWRDYGEVILCDTDEEVAAISDEYAAEHLEVHTSKDDWFMDRLQNYGSLFVGEETTVT
YGDKCSGTNHILPTKGAAHYTGGLSVHKFMKVVTTQRMTREANREVGAAAARISRLEGMEGHARAADVRL
RKYFPGENLA
```

IseK (HMM cutoff score = 202)

```
>637289836 SPO2358 iseK TRAP dicarboxylate transporter, DctP subunit, putative
[Silicibacter pomeroyi DSS-3: NC_003911]
MKVSDKLSGVSRRDFFRVAGRFGLSSTVLGLGGLTGAVTLPSLAAAAESTYDKRFKKEAKHTLKFGAAGFNARNLLIERAGCLEFA
RDLEERTDGEIRVEFIGDNQICGQLSCVEKTQNGIVDIYAASTQNSAGGAPYLNVLDYAYVFPGRASQYHFLYSPESNAILRDPLE
KRHGLKFLFSHCELRGLQMGKSFADKPTVTKLEELFGTKNRVTGTQLGRIAMQLLNLNPVPVAWEETLDGLKTGLIDGAETWASAV
AYANMSPVVSQSVDLRFFCGTEHTSMSAKVFDSLDGYLQDAVMESAYLAQVHVQAANEAALVKTVGFSDPVLPGTLFDEHDVRPAF
LAEDQIKMAEEMCSPEYNPEPWAQWRERLNNWSGGIDTYAELHRVAREVPADMKPENVEPRRWWKG
```

```
>2551559243 A142DRAFT_01565  tripartite ATP-independent transporter solute receptor,
DctP family [Vibrio splendidus 12E03 : A142DRAFT_AJZD01000156_1.156]
MMMTRKTLISAIVGATFAFGATASAYAATTLKLSHNHPRDHAVHEAMTFMAKEVREMTDGEVRIRIYPDAQLGTQRESMELMQNGA
LDMVKSNAAELEAFSPAYSSFNMPYLFRDQAHYYSVTDGEVGRDILNSSRDSGFIGVTYYDAGARSFYTSKPINTPEDLKGLKVRV
QPSPSAIAMVKALGGNPTPLAYGELYTALQQGVVDAAENNIPSFSLSRHSEVSKYFSLDEHTMVPDVLVISTKTYDKLTTEQQDAL
MKAASDSSELMKKLWAESEAKERAKAEKMGVTFVEPDKAAFALAVQPMYDALETSNPELSALIDRIQSVK
```

```
>2654246990 Ga0114105_107642  TRAP-type C4-dicarboxylate transport system, substrate-
binding protein [SAR324 cluster deltaproteobacterium REDSEA-S21_B5 Ga0114105 :
Ga0114105_1076]
MFTHKISRRDLIKVAGRYGLSSTLLGAATLGGALTLPELAAAANSTYKKRFKNTPKVKLKFGAAGLNERNLLIERAGCLQFVNDLE
ERTDGAIRVEFIGSNQLCGQLNCVKKTMQGITDIYASSTQNAAGGAPYLNVLDYAYVFPGSSAIYYFLYHPDSEKILREPMRRRHN
IQFLFSHAELRGIMLGLKWKDKPTVTSVEQLAGSKNRVTGTQLGRIAMNLMKLNPVPIAWEETLDGLKQGLIDGAETWMSAVAYAN
MSPVVSQCVDLRFFCGTEHTAMSSKVFDKLSGDLQDAVMESAYTTQIHVQGVHEASLVHTCGATNPQMPDTIFAKHGTRVANISTS
ERRKAEEMCSPEFQPQHWEEWRERINKWSGGIDTYSSIYKIAREIPADTRAENVEPRRWWK
```

```
>2592588220 MOLA814_00455  TRAP-type C4-dicarboxylate transport system, substrate-
binding protein [Betaproteobacteria bacterium MOLA814 : AYMW01000001]
MWFAQELEKRTNGALKIEFIGSNAICNQLDCVKKTQQGIVDLFTASTQNSAGSAPYYNVLDFAYMFPSRAAQHHFLYSKKSEALLR
EPMRRLHNIQFLYSHCELRGLMMGKKFEDKPLVTSIDDLAGTKNRVTGTQLGRIAMQLMNLNPVPIAWEETLDGLKQGLIDGAETW
MGAAAYANMAPVLSQAVDLKFFCGTEHTAMNFKTFQKLPPKIQDIVMETSFDAQQYTQQRQEKALIDVVGAVANPGKDTVFGKAGV
RVATLSDAEILKAEKMCSPEFNPKPWEEWRERLNKMSGGRDVYKEIYDIAREIPKNMNAVDVTPNRWWV
```

```
>2523988608 G511DRAFT_01008  TRAP-type C4-dicarboxylate transport system, substrate-
binding protein [Neptunomonas japonica DSM 18939 : G511DRAFT_scaffold00003.3]
MAKIITSSKPEQGSVSRRDFFRVAATYGMSSTMLAATALGSFTLPQLAQAAEGNAKRRYKKEAKHVLKFGASGFNENNLLIERAGC
IDFLNDIEERTDGEIRVEFIGDNQICGQLNCVKKTQQGIVDMFTASTQNSAGGAPYLNVLDYAYMFPTRASQYHFLYHPDSQKLLR
DPLREKHGLQFLFSHCELRGLQMGLGWKDKPLITSIEQLRGTKNRVTGTQLGRIAMQLLELNPVPIAWSETLDGLKQGLIDGAETW
AGAVGYANMSPVVSQSVDLRFFCGTEATMMSASMFDSFSGDLQDAVMESSYLTQVKIQAAQEAALVNTVGATSPSLPGTLFHKHGV
RMANLSDEERAKAERICAPEFNPEPWTQWRERLDKWSGGQDTYGSIHKIAREISRDTLAENVAPRRWWKS
```

```
>647613112 RB2083_572  twin-arginine translocation pathway signal [Rhodobacterales
bacterium HTCC2083 scf_1106200183017 genomic scaffold: NZ_DS995276]
MGTSDKINHLTRRDLFKVAGQYGMSSTLMAAGAFGGAMSLANLASAAESTYEKRFAKEPKHTLKFGAAGFNARNLLIERAGALEFA
RDLESRSDGEIRVEFIGDNQICGQVSCVEKCQQGIVDIYAASTQNSAGGAPYLNVLDYAYQFRNRADQYHFLYHPDSQRLLRDPYE
KRHGLKFLFSHAELRGIQLGLKWEDKATVTSVDQLAGTKNRVTGTQLGRIAMTAMNLNPVPVAWEETLDALKTGLIDGAETWSSAV
AYANMSPVVSQSVYLDFFCGTEHTAMNAGVFDSLSGELQDVVMESSYWAQTHVQAANEAALVKTVGVSDPILPDTIFAKNNVRNAF
LSKEEKRKAEEMCSPEFQPQLWEEWRERLNGWSGGHDTYQDIYNISREIDENTLPENVEPRRWWKA
```

```
>647670959 RKLH11_3142  twin-arginine translocation pathway signal [Rhodobacteraceae
bacterium KLH11 scf_1108856220443 genomic scaffold: NZ_DS999532]
MEIGKKLSGISRRDFFRVAGRYGMSSTLLAAGGFGGAMSLANLAQAAESNYEKRFSKEPKHTLKFGASGFNTRNLLIERAGAIEFA
RDLEERTDGEIRIEFIGDNQICGQLSCVEKTQQGIVDIYAASTQNSAGGAPYLNVLDYAYMFPSRAAQYHFFYSPESQRILRDPLE
KRHGLKFLFTHCELRGLQMGSTFADKPTVTKLEELFGTKNRVTGTQLGRIAMQLLNLNPVPVAWEETLDALKTGLIDGAETWASAV
AYANMSPVVSQSVNLKFFCGTEHTSMSAKVFDSLGGELQDAVMESSYFTQALIQGANEAALLNTVGFSEPQLPNTIFAENGVRPAF
LPDDQIKLAEEMCAPNYNPEPWAQWRERLNKWAGGIDTYQEIYDIAREIPADTLPENVEPRRWWMGQA
```

NaaA (HMM cutoff score = 266)

```
>637288147 SPO0660 naaA peptide ABC transporter, periplasmic peptide-binding protein
[Silicibacter pomeroyi DSS-3: NC_003911]
MKHLFKGLIAATAVASVLVGTTATAEEPRRGGTLKTQLGNNLRTLNSAVQSGIVTGFPGAQLFASPLRYDDSWTPQPYLAESWEVS
ADGLKVTLNLVKNAVFHDGHPITSSDIAFSVKVIQEAHPFKSMFAPVTSVDTPDDHTAVLNLSHPHPGLMLAMSGQLMSIIPEHIY
GDGQDFKTHPRNIENVVGSGPFKLAEYKNGEHVILERFDDFFIEGRPYLDKIVMRIITDSSARAIAYENGELDFGAFESEPLIINR
LKKVDHLTVTPTGYGAIGALDWLAINTTKGPLADKRVRQAIAYAIDKNFIQNALMQGTVQDSNTGIHPDSPFYNADVNAYALDLDK
```

```
ANALLDEAGYPMKGDERFAVTIDYGWPSAKPHVEYVRAALKKVGINVEVRASADFPTWAKRMGTMDFDLSWDAVFNWGDPSIGVHR
TYLTSNIAKGVWSNTQGYSNARVDELLDMAGKENDPEKRKALYAEFQSIVNDEVPIYFVQTLPYHTVYSNKVMNAPVNGIWSTSSP
LDQVWLKE

>640640659 RB2150_09984  peptide ABC transporter, periplasmic peptide-binding protein
[Rhodobacterales bacterium HTCC2150, unfinished sequence: NZ_AAXZ01000008]
MKQFLKGTFTAAAALSVLATGVIAEDTPKQGGTLVTVLSTNVRNLNSAVQSGIVTGYPGAQLFASPLRYDEDWTPQPYLAESWNVS
DDGLTVTLNLAKNAKFHDGVALTSKDVAFSVDVIKANHPFKSMFAPVESVETPDDHTAVLKLSKPHPALMLAMSGQLMSIIPQHIY
GDVPIDEVKTHPRNNENTVGSGPFKLVEYKSGEHVILERFDDFFIEGRPYLDKVVMRIISDPAARAIAYENGEVHMGAFESLPRII
NRLKKVDGLTVTSEGYGGIGPLDWLAMNTTKGPLADVRVRKAIAHAVDKNFIRNALMQGTTDESRTGIHPDSPFYNADVEAYELDL
DKANALLDEAGFPMNGGTRFALTIDFGWPGVKPQVEYVKAALKKVGIDVTVRASADFPTWAKRMGTMDFDMSWDTVFNWGDPVIGV
HRTYNSENIAKGVWSNTQGYSNSRVDELMAMAAVENDMAKRTALYAEFQQIIADELPVYHVNTLPYHTVYSDKVGNPPLGIWGTST
PIDMTYLKD

>2551215194 RCA23_c03720  peptide/nickel transport system substrate-binding protein
[Planktomarina temperata RCA23, DSM 22400 (RCA23) : RCA23c]
MKHILKSMLTATAAVAILASGAIAEDAPRKGGTLITVMATNVRNLNPAVQSGIVTGYPGAQLFAAPLRYDEDWTPQPYLAKSWDVS
EDGLTVTLNLVDNAVFHDGTPITSEDVAFSVDTIKAHHPFKSMFAPVVSVDTPDAHTAVLNLSKPHPALMLAMSGQLMAIIPKHIY
GDGQDPKTHPRNTENVVGSGPFKLVEYKSGEHVILERFDDFFIEGRPYLDKVVMRIITDPAARAIAYENGEVHMGAFESLPRIINR
LKKVDSLTVTDEGYGGIGPLDWLAMNTTKGPLADVNVRKAIAYAVDKNFIHKALMQGTASDSKTGIHPDSPFYNANVEGYDLDLDK
SRALLDAAGYPMQGDSRFSLTIDFGWPGVKPQVEYVKAALKKVGIDVEVRASADFPTWAARMGEMDFDMSWDTVFNWGDPVIGVHR
TYSSDNIAKGVWSNTQGYSNARVDELIAMAAVETDPAKRTALYAEFQEIIADEVPVYHTNTLPYHTVYNDNVGNPPLGIWGTSTPI
DMTYLKK

>2654239695 Ga0114103_10743  peptide/nickel transport system substrate-binding protein
[SAR324 cluster deltaproteobacterium REDSEA-S14_B10 Ga0114103 : Ga0114103_1074]
MKNLLKSLILSIFIILPGVLSVQAGSHKKGGTLIMSVGATPRHLNPAVQSGIDTGAPGAQLFATPLRFDENWNPQPYLAESWSISD
DGLSVTLNLVKGATFHDGKPITSKDVAFSIKTIKAHHPFKTMFGAVKRVDTPDSHTAVIRLSKPHPALLLALSSQLGAVIPEHIYG
DGQDPKTHPRNSENVVGSGPFKLVEFVRDQHIILERNENFFLEGKPYLDKIVMRIIKDPSARSLGRENGEIHLSKPTSDVRVRQAI
AYAVDRNFLINAVMRGTSKPAYTGIHPDSIFHEPNVERYDLDINKANQILDDAGYSKNSDGIRFPLNIDYGWPSAKPMAEYMKPQL
KKIGIDVTVRTSAGFPAWAKTVSNWEFDMTVDVVFNWGDPVIGVHRTYLCENAKKGVIWSNTQQYCNPEVDQILEKAGQENDRNQR
IAFYSKAQKLISKDVPIYFTDTVPYHTIYNHKKVGNPPSTIWGTCSPLDEVFLK
```

TauA (HMM cutoff score = 374)

```
>637288161 SPO0674 tauA taurine ABC transporter, periplasmic taurine-binding protein
[Silicibacter pomeroyi DSS-3: NC_003911]
MMTAKTKLLGAVAGLALAIGAMPAAVAAEGEITVAYFLEWPMPFEYAKQKGTYDEVLGVKVNWVSFESGVKMSAAMASGDVHLSVS
QGVPPFVVATSAGQDLQILDVAVSYADNDNCVVRSELEIDKDNAGELAGKKVAVPLGTAAHYGFLKQMNHFGVDVASMDVVDMDPP
DGAAAIAQGAVDMFCGWGGSLRRALEHGNVLLTGDEKTELGILVFDVTSGPAGWVAENSDLVAKFLKVTADANAMWADEANRAEML
PLIAKDAGMDEEATASTIATFKFPTIEQQLSGGWLGGNAQTFMKGVADVFVEAGSIDSALDTYENAVNTGPLSAAGGM

>2517702672   taurine transport system substrate-binding protein [Rhodobacterales sp.
HTCC2255 (original sequence, contaminants removed) : AATR01000003]
MKITKTLTAMAVALAATTGTSAVALDELVVGYFMEWPTANQYAQHNKLYDEALGIPVKWVSFDAGTAMSAAMASGDVHISYSQGVT
PFLVATAAGQDLQVIDIAETYSDNNNCVVRSSLEIDADNTDELKGKQVALPIGTAAHSDFLAQIGHFGLAESDFKIVDMTPVDSAA
AFASGEFDMVCGWGGPLRRMKEHGNVLLSGPAKEAVSGVIYDVTSTPSSFAEANADVVTTFLKVTSEMNTMYAENPEPMYEAISIE
AGMDMEGTKAILAVMAFMSAETQLSDQWMGKWLAGDLKRQALALEAAGKITALDDYDALVNTTYLAAAAK

>2528297236 AB747F18903_00590  taurine transport system substrate-binding protein
[SUP05 cluster bacterium AB-747_F18AB-903 : AB747F18903_contig_11_0.12]
MKSIYKKIGATVVSAIASVNVMALDELNVAYFEEWPMPFEYAKQIGAYDEALGMTVNWKAFGTGTAMSAAMASGDIHISVSQGVPP
FIVAASTGQDIQIVDVAVSYADNDNCVVASGLEIDASNASELAGKKVSVPIGTAAHYGFLKQMSHFGVSVDSMEVVDMAPPEGAAA
FAQGSTDMFCGWGGSLRRALEHGNVLLTGAEKTALGILVWDVTSTPASFADENAEVLQTFLGVTAASNAMWNSGGFTSLMLPHIAR
DAGMDEGDTADTMATFVFPSVSNQLGSNWLGGSGAAFLKGVADVFVESGNIPSARGSYANAINTDGLEGLAAQ

>2592588209 MOLA814_00444  taurine transport system substrate-binding protein
[Betaproteobacteria bacterium MOLA814 : AYMW01000001]
MTNKTIKFAIAAALGAVATVSFAQPKEITVAYFLEWPMPYLAAKAAGTYDKEMGVKVNWRSFETGTAMSAAMASGDVQIAVSQGVP
PFVVAASGGQSIQVVDIAVSYSDNDNCVVSKKLEIDKSSAKELAGKTVAVPLGTAAHYGFLRQMDHFGVAVSSLKVINMAPPEGAA
ALAQGSVDFACGYGGGLTRMKEYGNVLLTGKEKEALGILVFDVTSSPTNFIAENQQLLSKFLKVTANANTQWNAKKDPAMLKVIAK
ESGMDEAAAAKSIATMSFPSAKDQLSKQWLGGNAQTFMKGVADVFMTAGNIKKVLPSYDKSVNTAPLMQAVK

>WP_011281961.1 taurine ABC transporter substrate-binding protein [Candidatus
Pelagibacter ubique]
```

```
MKKIISFILGTVVALNLSMSVANAAAKEVRVAFFLEWPTPNQEDKVKKMFDKALGVPVKWTNFSNGGAMTDAMLAGDIDISYSQGL
VPFINAVKSKAPLKLVDIAMEYGMGGTTCVTSNASGITKANASELEGKKVAVPLGTMAEYVFDESMKVVGADKSKMTVIQMDPEEG
AAALVSGDVVMACLFGGNSIKAATAVGSRLLTVQEARDAGILGIDITSVTDKFMKENPGMLRTFIEVTHEANARYAAGKSDLNVIA
KDAEMKLGDMKETLSGFKFLTPEETEKSMTSGNLSGFLKGMGTPNGAVDTSFLPL

>WP_085114116.1 taurine ABC transporter substrate-binding protein [Candidatus
Pelagibacter sp. HIMB1321]
MKKILSFILSSIVALNLSISVANSAANEVRVAYFLEWPTPNLEDMNKKMYEKALGVPVKWTNFTNGGAMTDAMLAGDIDISYSQGL
VPFINAVKSKAPIKLVDIAMEYGMGGTTCVTSNASGITKANGSELEGKKVAVPLGTMAEYVFDESMKVVGADRNKMDIIQMDPEEG
AAALVSGDVVMACLFGGNSIKAATAVGSRLLTVQEARDAGILGIDITSVTTKFMKENPGMLRTFIEVTHEANARYAAGKSDLNAMA
KESEMSVGDMKDTLSGFKFLTPEETAQSMTKGNLHGFLEGMGTPKGNVDTSFLPL

>WP_014952993.1 taurine ABC transporter substrate-binding protein [alpha
proteobacterium HIMB5]
MKKIISFILSCFVALNLSISVANSAANEVRVAYFLEWPTPNLEDMHKKLFAKALGVPVKWTNFTNGGAMTDAMLAGDIDISYSQGL
VPFINAVKSKAPIKLVDIAMEYGMGGTTCVTSNASGITKANGSELEGKKVAVPLGTMAEYVFDESMKVVGADRNKMDIIQMDPEEG
AAALVSGDVVMACLFGGNSIKAATAVGSRLLTVQEARDAGILGIDITSVTDKFMKENPGMVRTFIEVTHEANARYAAGKSDLNAMA
KESEMKVGDMKDTLSGFKFLNAEETAQSMTKGNLHKFLKGMGTPRGNVDTSFLPL

>WP_009604762.1 taurine ABC transporter substrate-binding protein [SAR116 cluster
alpha proteobacterium HIMB100]
MKSFITRTAAAAVAALTLSGAAQAADSVNVAFFLEWATPNQIAKVDKAYDEAMGVDVNWTDFSTGVAMTEAMLAGDIDIAYSQGLA
PFVTAIQQGAPLKMVGVAVVYEANDCFVANGLGIDSSNASELEGKTVAVPLNTMADFSFRKQMAALNVDTSTMTIVDQAPPDGAVS
LADGSVDMACIFGGASAKAAGEVGTPIMSTQQKTDAGIGSFDVISVTEKFANENPELVSSFLEVTEEANAAWMATPEQIRKVASDA
GMDFTTTQNQMAGFLFPTVAEQAADYFGNDGIAAAAAESLGAVFVKTNIKNYKGSVGAAIDGSFLQ

>2698166757 Ga0131154_107133 taurine transport system substrate-binding protein
[Sulfitobacter pseudonitzschiae DSM 26824 : Ga0131154_107]
MTFKSKLMSATAAIAMLTGAQAATAAGHGEITVAYFLEWPMPFQYAKETGMYDEAMGTKINWVSFDTGTA
MSAAMASGDVQLSVSQGVPPFVVATSAGQDLQIVDVAVSYSDNDNCVVASGLEIDKDSAGELAGKKVAVP
LGTAAHYGFLSQMNHFGVDISTMDIVDMAPAEGAAALAQGSLDMACGWGGALRRMKEHGNVLLTGAEKEE
LGILVFDATTAPATYIAENGEMVAKFLAVTAKANAMWADEANHATMLPVIAKDAGMSEEDTMSTLSTFKF
PTVDEQLSDKWLNGRAQTFMKGVAGVFVDAGSIPSALDSYAGTVNTGPLEAAKDM

>2551214976 tauA taurine transport system substrate-binding protein [Planktomarina
temperata RCA23, DSM 22400 (RCA23) : RCA23c]
MKMKTALKSAVAGLTLVSGSAAFAGGHGELNVAYFLEWPMPFQAAKVSGAYDAALGMKVNWVSFDTGTAM
SAAMASGDIDISVSQGVPPFVVATSAGQDLQIVDVAVSYADNDNCVVASGLEIDKNSAGELAGKKVAVPL
GTAAHYGFLKQMEHFGVSLDSLTVVDMAPAEGQAALAQGAVDMACGWGGALRRMKESGNVLLTGAEKTEL
GILVFDVTTAPSGFVAENADIVAKFLAVTADANTAWNTSQPEFMANMIAQDAGMSVEDAKSSMSTFVFPD
IDTQLSQSWLGGNAQTFMKGVADVFVNAGSIDSAKGSYADNVNTGPLASAKGM
```

Xsc (HMM cutoff score = 258)

```
>WP_044028790.1 sulfoacetaldehyde acetyltransferase SPO3561 [Ruegeria pomeroyi]
MKMTTEEAFVKTLQMHGIQHAFGIIGSAMMPISDIFPAAGITFWDCAHEGSGGFMADGYTRATGKMSMMI
AQNGPGITNFVTAVKTAYWNHTPLLLVTPQAANKTIGQGGFQEMEQMRMFADCVAYQEEVRDPSRVAEVL
NRVIMNAKRASAPAQINIPRDMWTQVIDIALPAIVEFERPSGGEEAVAAAAEMLSNAKNPVILNGAGVVL
SKGGIAASKALAERLDAPVCVGYQHNDAFPGSHPLFAGPLGYNGSKAGMELIKTADVVLCLGTRLNPFST
LPGYGMEYWPADAKIIQVDINPDRIGLTKKVSVGIIGDAAKVARGILAQLSDTAGDEGRAARKDRIAQTK
SAWAQQLSSMTHEDDDPGTTWNERARADKPDWMSPRMAWRAIQAALPREAIISSDIGNNCAIGNAYPDFD
EGRKYLAPGLFGPCGYGLPAVVGAKIGCPDVPVVGFSGDGAFGIAVNELTAIGRGEWPAVTQIVFRNYQW
GAEKRNSTLWFDDNFVGTELDQRVSYAGIAKACGLQGVVARTMDELTAALAKAIEDQKAGKTTLIEAMIN
QELGEPFRRDAMKKPVRVAGIDKADMREQQV

>WP_008033816.1 sulfoacetaldehyde acetyltransferase [Rhodobacterales bacterium
HTCC2255]
MKMTTEEAFVKVLQMHGIEHAFGIIGSAMMPISDLFPAAGINFWDAAHECNAGMMADGYTRASGKMSMMV
AQNGPGITNFVTPVKTAYWNHTPLLLVTPQAANKTIGQGGFQEMEQMNLFADCVAYQEEVRDPSRIAETL
NRVILQAKRASGPAQINIPRDFWTQVIDIELPKVVEFERPSGGSDALNQAAELLSNAKFPVILNGAGVVL
SSGIPASAALAERLDAPVCCGYQHNDAFPGNHPLFAGPLGYNGSKAGMELISKADVVLALGTRLNPFSTL
PGYGIDYWPKDAKIIQVDINADRIGLTKDVTVGIVGDAKKVAEELLSSLSDHAGETNRADRKATIAKTKS
AWAQELTSMDHEDDDPGTTWNERARSDKPDHMSPRMAWRAIQAAMPNNAIISSDIGNNCAIGNAYPSFQE
GRKYLAPGLFGPCGYGFPAIAGAKIAQPNVPVIGFAGDGAFGISMNEMTAVGRGEWPAITMVIFRNYQWG
AEKRNTTLWFDDNFVGTELDEDVSYAAIANACGLVGVQATSMDDLTDKLDAAIKAQMNEGKTTFIEIILN
```

```
KELGEPFRRDAMKAPVSVAGISKNDMIPQKGA

>WP_088919161.1 sulfoacetaldehyde acetyltransferase [Granulosicoccus antarcticus]
MKMTTEEAFVKVLQRHGIEHAFGIIGSAMMPISDLFPAAGIKFWDCAHECNAGMSADGYTRATGKMSMAI
AQNGPGITNFVTPVKTAYWNHTPMLLVTPQAANKTIGQGGFQEIKQMALFEDMVAYQEEVRDPSRIAEVL
NRVILQAHRLAGPAQINIPRDYWTQVIDIELPQVFRLERPRGGEQAIQDAAALLSEAKFPVILNGAGVVL
AGGIDASAKLAEKLDAPVCCNYQHNDAFPGSHPLSMGPLGYNGSKAAMEIISKADVVLALGTRLNPFSTL
PGYGIDYWPKDAKLIQVDINADRIGLTKAVAVPIQGDAKRVAEQLLEQLSASAGNADRAERKSLIDQTKK
RWSQELVSMEHEDDDPGTNWNERAREREPKKMSPRNAWLAIRDAMPADAIISSDIGNNCAIGNAYPTFEQ
GRKYLSPGLFGPCGYGFPAILGAKIGCPDVPVIGFAGDGAFGISMNEMTACGRNDWPPITMVIFRNYQWG
AEKRNTTLWFDDNFVGTELDLQVNYAKIAEACGLKGVQVDSTEALTEALNTAVKEQMEDGVTTFIEILLN
QELGEPFRRDAMKQPVSVAGIDLADMIPQQTV

>2779882284 sulfoacetaldehyde acetyltransferase  [SAR116 cluster alphaproteobacterium
MED880 : Ga0267689_113]
MKMTTEEAFVKVLQRHGIEHAFGIIGSAMMPISDLFPQAGITFWDAAHENAGMMADGYTRASGKVCMMVAQNGPGITNFVTPVKTA
YWNHTPLLLVTPQAANKTIGQGGFQEVEQMALFKDMVCYQEEVRDPTRVAEVLNRVITQAVRLSAPAQVNIPRDFWTQVIDIEIPE
ILSFERPSGGDDALSKAAQMLSEAEFPVILNGAGVVIGGAIPASMALAERLDAPVCCGYQHNDAFPGSHPLFAGPLGYNGSKAGME
LIAKADVVLALGTRLNPFSTLPGYGIDYWPEDAKIIQVDINPDRIGLTKPVTIGIIGDARKVAEGLLEKLGPQAGEKGRSSRTAVI
AKTKSAWAQELTSLDHEDDDPGTTWNERARKREPGKMSPRMAWRAITSVLPKDAIISSDIGNNCAIGNAYPTFEEGRKYLAPGLFG
PCGYGFPAICGAKIAQPDVPVVGFAGDGAFGISMNEMVSVGRDEWPPITMIIFRNYQWGAEKRNTTLWFDDNFVGTELDTDVTYAG
IAQQCGVKGVQVDGMDALAEALDKAIEAQMKDKVTTFIEVVLNQELGEPFRRDAMKKPVSVAGIDPADFRPQQPVS

>2654250678  sulfoacetaldehyde acetyltransferase [SAR324 cluster deltaproteobacterium
REDSEA-S27_B3
MTTEEAFVKVLQMHGIEHTFGIIGSAFMPISDYFPKAGITFWDCAHECNGGYMADGFTRSTGKMAMIIAQNGPGITNFVTCVKTAY
WNHTPMLLVTPQAANKTIGQGGFQEVEQMAAFKDMVAYQEEIRDPSRVAEVLNRLITKAKRMSAPTQVNVPRDLWTQVIDIELPPI
VNFERPSGGVGAIAAAAELLSNAKFPVILNGAGVVLGDAISASQALAERLDAPVCCNYQHNDAFPGSHPLFAGPLGYNGSKAGMEL
IKQADVVLALGTRLNPFSTLPGYGMDYWPNDAKLIQVDINPDRIGLTKPVSVGIVGDAKQVAEQILAQLSGSAGDAGREERRNRIA
TTKSAWAETLASMNHEDDDDGTSWNERTRKRQPDHMSPRMAWRAIMEALPKEAIISSDIGNNCAIGNAYPTFEQGRKYLAPGLFGP
CGYGFPTIIGAKIGNPDTPVVGFAGDGAFGISMNEMTAVGRNDWPAITMVIFRNYQWGAEKRNTTLWYDDNFVGTELDLQVSFAEI
AVACGVKGVKAKSMNELTEQLETAIKAQMEQGETTFIEAVLNQELGEPFRRDAMKTPVPVAGINKDDMCVQQG

>2731912897 Gammaproteobacterium bacterium SCGC AG-487_F09
MKMTTEEAFIKTLQMHGIEHAFGIIGSAMMPISDLFGRAGISFWDCAHEGSGGMMADGYTRATGKMSMMIAQNGPGITNFVTAVKT
AYWNHTPLLLVTPQAANKTIGQGGFQEVEQMALFQDMVCYQEEVRDPSRIAEVLDRVIANAHRASAPAQINVPRDMFTQIVDIELP
TMIDFELPSGGESAISAAAELLSNAKAPVILNGAGVVLSAGGIASSMALAERLDAPVCCGYQHNDAFPGSHPLFAGPLGYNGSKAG
MELIREADVVLALGTRLNPFSTLPGYGIDYWPKGAKIIQVDLKPERIGLTKAVSVGIVGDAAKVANGIMAQLSATAGDAGREARKA
HIAQTKSNWVAELAAMTHEDDDPGTTWNQRARADKPEWLSPRMAWRAIEAVLPKEAIISSDIGNNCAIGNAYPTFEKGRKYLAPGL
FGPCGYGLPAVVGAKIGCPDTPVVGFSGDGAFGIAVTELTAIGREEWPAITQIVLRNYQWGAEKRNSTLWFEDNFVGTELDENVSY
AGIAQACGLQGVQARTMDEVSAALSQALEDQASGKTTLIEILINQELGEPFRRDAMKNPVEVAGVSIEDMRPQKGS
```

dddQ (HMM cutoff score = 223)

```
>gi|83951699|ref|ZP_00960431.1| hypothetical protein ISM_14090 [Roseovarius
nubinhibens ISM] /dddQ/
MDPRWTTLLQEARAAHGATPELRDFCAFPDALRDQPGDPRPDPLAVTLQDAPGDTSARWQGFRDAACAVGPIARWRDTYRHTAIGA
DLHRHFGCYELLGQDAPYGTEQMRGFLVYQRPGYHYPPHHHPAEEIYLVVAGEAEFHLDGHAPRRLGPGGTVFHPSGVAHALTTHD
SPVLAWVLWRGEMDTRPVFSDPALQGEAQR

>gi|83951698|ref|ZP_00960430.1| hypothetical protein ISM_14085 [Roseovarius
nubinhibens ISM] /dddQ/
MSGRAALTHLLEAARDWHEALPEFRAFATWPEDLRWADRPAHALPVIDHLTRDPGHASDQSQPLRDALVAAAPHVEWRHSYTEAEV
GRDFLNRFGWFELAGPSGHFLTQSLRVTVGYWGPGLDYGWHEHLPEELYSVVSGRALFHLRNAPDLMLEPGQTRFHPANAPHAMTT
LTDPILTLVLWRGAGLGDDPRMSQ

>gi|56696480|ref|YP_166837.1| hypothetical protein SPO1596 [Ruegeria pomeroyi DSS-3]
/dddQ/
MTQTDPAFQNLLAEFQALHAREPALAGFVALPDSLTPQPVTPVRIPPAALMESDPDLTTTAYAAIRDAFIAAGAVAQWRLTYQGSR
LGADFMDRFACYCLIGEGGPFASDSLAAYVVYMPAGLYYPFHQHPAEEIYFILAGEAEFLMEGHPPRRLGPGDHVFHPSGHPHATR
TYDRPFMALVLWRGDLETAPVLTYPEGEI

>gi|254465895|ref|ZP_05079306.1| conserved hypothetical protein [Rhodobacterales
bacterium Y4I] /dddQ/
MSLSQALQAARDLHAKLPALREFAPWPGDLEPAALAPAAMAAQDQVAGFNLPGTADTQPLIDALRTAAGHAHWTQTYTEAEVGAHF
LQNYGYFELFGPAGHFRSTRLRGYIAYWGAGLKYDWHSHEAEELYVVLGGGAQFLTEENESWLQAGDTRLHGGWQPHAMNTGAEPV
LTYVLWRGQGLAGLPLMGGKRTAAA
```

dddD (HMM cutoff score = 150)

```
>sp|A6W2K8.1|DDDD_MARMS RecName: Full=CoA-transferase/lyase DddD MWYL1
MNKQNQLPLVGVRVADFGQQIAGPAVAMVLADLGATVVHIDPPGGPSWKHPANAILNRNKASLCIDLKTQ
AGLDQALELIENVDIVIESFRPGVMKRLGIDFVALRESRPELITLSMPGFASNDELHRDWKATEAIVAAT
SGTFTDMGFNRVLMGLNPSFSPLPLGSSYAISLAASSIALALFEREKTGRGDNIEVPIAAALMEGLSYNS
YVVDQLPERYKTMRELEIEHRKSNNIKMDVSYAQLQEYLDPFYRTYVCADGRQFYCVCPSHRNHAERALK
VLGIYDELVAEGLPEVKDLHVPISEWDGETSIGVYPLPKKWADLISEKMKKAFLQKTSDEWGVIFGEGQI
PGAPHRSTEEWVNSEHCNASGLIVEVEGTEFGTMKQPGPIVWFENESEAMLKPKPQEHVSFEQALARLQS
VAKIEKISRPTGQDIQPASGKGWLDGVKILDLTNVIAGPHSTAFMSRFGAEITKLDPVTPLYDPLIGILF
TFQTGVGKQSALVNIMTKEGREVFERLVRSVDIVVINAPDRQMKPLGLDQDSLSAINPDVLFCRLDCFGG
PRTGSKTNYIGYDDIIQANSGIMSRFGKPETPEEHAHLGTLDVNCGFAAALGMVIALYQKRKTGKVCRVR
TSLSAVTNIAQIPFAFDYEGRAPFNEASGREAMGNHALSHFYRTNSGWVFLDSHQGELAKLDAIKGLNGI
QQSQDMGQFLRDQLVKESSAYWLKEFAAADIACAEPFSIEYLREHNSRVADQKVGTDLGSYAFSIFPDHP
SGHCITQVDPYSIRPREAKIRAVTPTEKFGCSTIKVLQGLGYSESDINDMLEKKIAATGWGREFLPS

>EBA08656.1 CaiB/BaiF family protein [Sagittula stellata E-37]
MRKSFSNLPLTGVKVVDFGQYIAGPAVAMLLGDLGATVVHIDPPDGPMWDSPANAALMRNKLIVNLDLKS
EDGLAKARALCSEADIIVENFRPGKLAKLGIDFAAMREERPELITISIPGFASNDELRRELRAFESVVAA
SSGVFTDMGLNRVLMGVNPSFSPLPLSSSYASQIAASATVLALQSRQLTGLGDQIEVPLAAAVMEGLCYN
SIHVSDVPKRYLTQREIEIERRRIEGLPMNVSYEDLQELLDPFFRSYMCKDGRMFYVVCPSHKNHARRCL
EVLGLYEEMVEEGLQEEDDTYKPQAEWETDTSLGVYPMPKFWADKIAARMKEVFMTRTSHEWKRMFGRAG
IPGAPQRWLQEWINDEYAEASGLMIDVWDPEYGEMTQPGPVVWMEESGEEALTPEPRRWVEFDEALKVLR
KLRTDIPDPEDGIEQEGWLSGVRVLDMCNVIAGPHSASYLARFGAEVIKLDPAVPLYDSWNTVMYGMSQG
RGKRSILADVKSEHGRKVFEDLVKTVDVIFWNAPDSQIKRMGLDEAGLRKLNPEAIFCKLDCFSGVRRGV
RSDYIGYDDLVQATTGIMLRFGGAMDRPEEHAHVGTIDVMCGFGGALAVAAALYQKNRFGRIGRGRTSLS
ANSGLLQIPFAYDYRGRGLFDEPTGPEVNGYDALSRFYSTASGIYILLSAYEADLPRFRNVEGLEEFPDV
PEEDRAAFLAGAFQSLPATEWIDRLRAADLACAICQNIEALRAENLQEADGTPGTDRGSYSFSLYTDHPS
GHDVTQLDPYAVRPARGRVFALSPTEKFGTSTRDILNELGYSVSAVERMLKSGQVSESWSAEYLPS

>Ruegeria pomeroyi dddD MQNRPLSHIRVLDFGHYLAGPLVGMMLADLGAEVVRIDPPAGPRWKDPAFDMLSRGKRAL
TLDLKTADGRDTALDLVRRADVVIENFRPGVMERLGLGPDALRQANADIVSLSLPGFAST
DPEFAGWAAWEAVIAARTGQFTDMGLNRRLMGINPSFTPLGLASAYGAAFGTMSVLFALG
ARSRMGGDHIEVPLASALLEGLIYNCEQIEDYPDRYKSPRELELERRAGEGLPNNLSFAE
LSEFLDPFYRTYTCADGRGFYIVSCSIVNHPQRVLEVLGLGELLKELPDFDVYVDQADWP
GEWALRSYPVGADDRKRLSDAMKAAFLTRPAHEWEELFGVAKAPATAQRSTAEWLVDPHA
LASGLVVALDDPRHGQMRQMGNVAWLTDDPGAMKKVAGPEQDDFRDALSGVLAEPPRRPT
GGDSKGVWLDGLKVLDLTNVIAGPTIGSTLARFGAQVTLVQPVRPSVDPWNAVVFGLHAQ
RGKESVLLDLRSEQGQEALWRLVAEADVITMNGTDQQRDALGLTEARLNEVNPRLILVQL
DAWGGPRRGPKSDHLGYDDLAQAATGVMTRFGGGPETPEEHAHFGTIDALTGHCACVALG
AALERLRVTGKGGVARASLAAAGEMIQAQFMYDFDGRPAFDEPSGREVRGWGSFYRCYAA
ADGWMFFAAPTERDAALQRVPDLSDLVGKDDADLEDLLAERFAQKRVADWMRAFAGGSVG
ITPLGSLHGTRDAGLQRESEGEIDISKATFRAVRHDRHPMGRWVDLVAPNAVRPEKTRIT
IPGPAPKYGQHTREVLAGVGYADDEIGRMIESGAAAQSWSDKYLPE

>643822352 YP_002822700 dimethylsulfoniopropionate cleavage enzyme DddD [Rhizobium sp.
NGR234 plasmid pNGR234b: NC_012586]
MTSISRSALPLSGVRVVDFGQYIAGPAVAMILGDLGATVVHVDPPGGPMW
DNPANAILNRNKLIVTIDLKTAHGLEEARALIAGADVVIENFRPGVMARL
GLDFEALRTVRPELITLSIPGFGSNDQLRRDWRAFETVIAASSGVFTDMG
LNRVLMGINPSFSPLPLASAYGTMLAASATVLALQARERTGHGDHIEVPL
ASAVMEGLSYNSIKIDDYPLRYQTQREREIERRQREGLPMDMSYEDLQQF
LDPFYRSYLCSDGRMFYVVCPSHRNHAKRCLQALGLYDELAAEGLREEED
TYLPVSQWSSDVSLGVYPLPKHWADKIAARMKEVFLTKTAAEWERIFGEG
LFPGAPQRWLKEWISDDHANAAGLMIEVDDPVYGRMTQPGPLTWLQESAE
AMLTPQPRRWGSVDEAVELLSRTGSLKIPAATSSNSGGWLDGVKVLDLCN
VIAGPHSVAYLARFGAEVIKIDPAKPLYDCWNTVIFGMSHMCGKQSVLLN
IGSPDGRVVFEHLVKSVDVVVWNATDRQVKWMGLDAESLEALNPNAIFCQ
LDCFGGVRKGPRTDYLGYDDLVQSATGIMLRFGGSMQTPEEHAHVGTIDV
MCGFGAALGVAAALYQKSKTGIVGRPRTSLSALTGLAQIPFCYEYEGRGP
FDEPSGRDIKGYDALSRLYETGSGDYILLCATEADLPRLVTVEGLERLAS
VAVPDREGWLARAFMSAPAETWQHRLLQADVGVSLCENIETIRARSARVS
DGTPGTDRGSYSFSIFPDHPSGHTVTQLDPFAIRPAVGKVIAVTPAEKYG
ASTRSVLKALNYTDAEIDRMIASGTASETWSAEYLPS
```

dddL (HMM cutoff score = 130)

```
>dimethylsulfoniopropionate lyase DddL (EC 4.4.1.3) [Sulfitobacter sp. EE-36]
MTFDTAAPTTLRDVPDWRYMMQECDELYRYLPAGGSDRIKSHQRKAREAI
ARLLRGNAELRLQPRAEKPVTAHLRRALDEGKQGALAPAVRALDAVAHDL
SWQYGYEKVPKGLTNSYAYAELAGAQGPVVSHDIILGVVLFAPGCTYPAH
AHKGITESYVCLSGAVSENHQGVYVPGSMIFNPPEHLHRITVGDREPALL
AYAWMGDPADLHQQKMVFTRARK

>dimethylsulfoniopropionate lyase DddL [Dinoroseobacter shibae DFL-12]
MPDLPNPETPAEASALRLMDCPNWVYLLREFDALYRYGSAGGSPAIRSHR
KRVRDRLSAVLAANPAMVDRPRETKPVVAHFARALDLGERAAMQGMVRAL
REVKDDLTWEYGYEKVPKALAQKYAYCEVLGPRGPVQGTTLTLGFVLFAP
NTTYPQHSHHDIEESYTSVSGAWSENNAAVFAPGSLILNTSGHEHRITTG
DRDPCLLAYAWAGPAERLATPDMKLTAPRRARGAGV

> dimethylsulfoniopropionate lyase DddL [Rhodobacter sphaeroides 2.4.1]
MHSLSERVEQLRLNDCPDWLYLLHEFDALYRQGSDGGSRPIRTHRKRVRD
SLALIVEANPAVNDRPPEVKPVTAHLGRALDLGERGAVQGMSRALARVAG
RLTWEYGYEKVPKALARKYAYCEILGPRGPICAERLILGFVLFAPSTTYP
QHSHKDIEESYISVAGAWSENDAAVHAPGSLILNRPGLEHRITTGDLSPC
LLAYAWTGSEERLNQPGMKLSSPRKARIEKGI
```