Supplementary Data 1

Table 1S Specificity of PKS domains

Domain	Activitity	Choices
AT starter	acyl transferase	**selection of substrate** malonyl propionyl methylbutyryl
Starter no AT	-	**selection of substrate** 3,4-DHCHC 3,5-AHBA p-nitrobenzoate trans-1,2-CPDA hydroxymalonyl phenylacetyl p-aminobenzoate 3-methylbutyryl cyclohexanecarboxylic acid benzoyl 3-amino-2-methylpropionate
ACP	acyl carrier protein	generic, no specificity
KS	ketosynthase	generic, no specificity
AT extender	acyl transferase	**selection of substrate** malonyl methoxymalonyl methylmalonyl ethylmalonyl
KR	ketoreductase reduces C=O to OH	active vs inactive stereochemistry of OH (R or S) stereochemistry of β-carbon atom (R or S)
DH	dehydratase reduces OH to C=C	active vs inactive *cis*/*trans* orientation of double bond
ER	enoyl reductase reduces C=C	active vs inactive stereochemistry of OH (R or S) stereochemistry of β-carbon atom (R or S)
TE	thioesterase release from enzyme cyclization	generic, no specificity

Table 2S Fingerprints of KR domains

Reference sequence = 'Erythromycin_KR_module01'

Key residues:
Activity = 113, 137, 150, 154
Specificity = 93, 94, 95, 142, 147, 150, 152, 154
Hmmalign is used with the profile constructed for all KR's and the amino-acid alignment from which the profile was built

Type	**Regular expressions describing activity of this KR type**
active	K[S,A,G]YN or E[S,A,G]HH or K[S,A,G]Y[N,G]

Type	**Regular expressions describing specificity of this KR type**
A1	LDD and W[anything except H]YAN
A2	LDD and WHYAN
B1	LDDXXY[anything except P]N
B2	LDDXXYPN
C1	XXXXX[anything except Y]XX
C2	XXXXXYX[anything except N]

X – Stands for any amino acid including gaps (-)

Table 3S Fingerprints of AT domains

Reference sequence = 'Erythromycin_AT_module01_C3'
Key residues = 7, 70, 96, 97, 98, 99, 100, 123, 197, 198, 229, 248, 253

Hmmalign is used with the profile constructed for all AT's and the amino-acid alignment from which the profile was built

LEGEND:

Building block	**Regular expressions describing specificity of this AT type**
malonyl	QQGHS[L,V,I,F,M]GR[F,P]H[A,N,T,G,E,D,S,-][N,H,Q]V
methylmalonyl	QQGHS[Q,M,I]GRSHT[N,S]V
ethylmalonyl	Q[Q,H]G[H,S]S[Q,L]GR[G,T,A]HTNV
propionyl	QQGHS[Q,M,I]GWAH[S,G]SV
methylbutyryl	QQGHS[Q,M,I]GWAH[S,G]NV

Methoxymalonyl-specific AT's were first assigned as an unknown group. These sequences were then checked for positions specific for methoxymalonyl by looking at:

Reference sequence 1 = 'Concanamycin A_AT006_1_hidro'
Reference sequence 2 = 'Ascomycin_AT007_1_hidro'

Key residues 1 = 182, 184
Key residues 2 = 221, 222

For the methoxymalonyl a new profile with only methoxymalonyl AT's was built and used for the alignment and position extraction.

LEGEND:

Building block	**Regular expressions describing specificity of this AT type**
methoxymalonyl	XW from key residues 1 or PX from key residues 2 is sufficient

With this fingerprints all AT containing loading modules and most of extender modules were covered. Unknown/non-predictable building blocks were named internally as CX and were assigned with "C(*)~C(*)" SMARTS structure.

X – Stands for any amino acid including gaps (-)

Extenders:
Malonyl, methoxymalonyl, methylmalonyl, ethylmalonyl

Starters: malonyl, propionyl, methylbutyryl

Table 4S SMILES of starter and extender building blocks

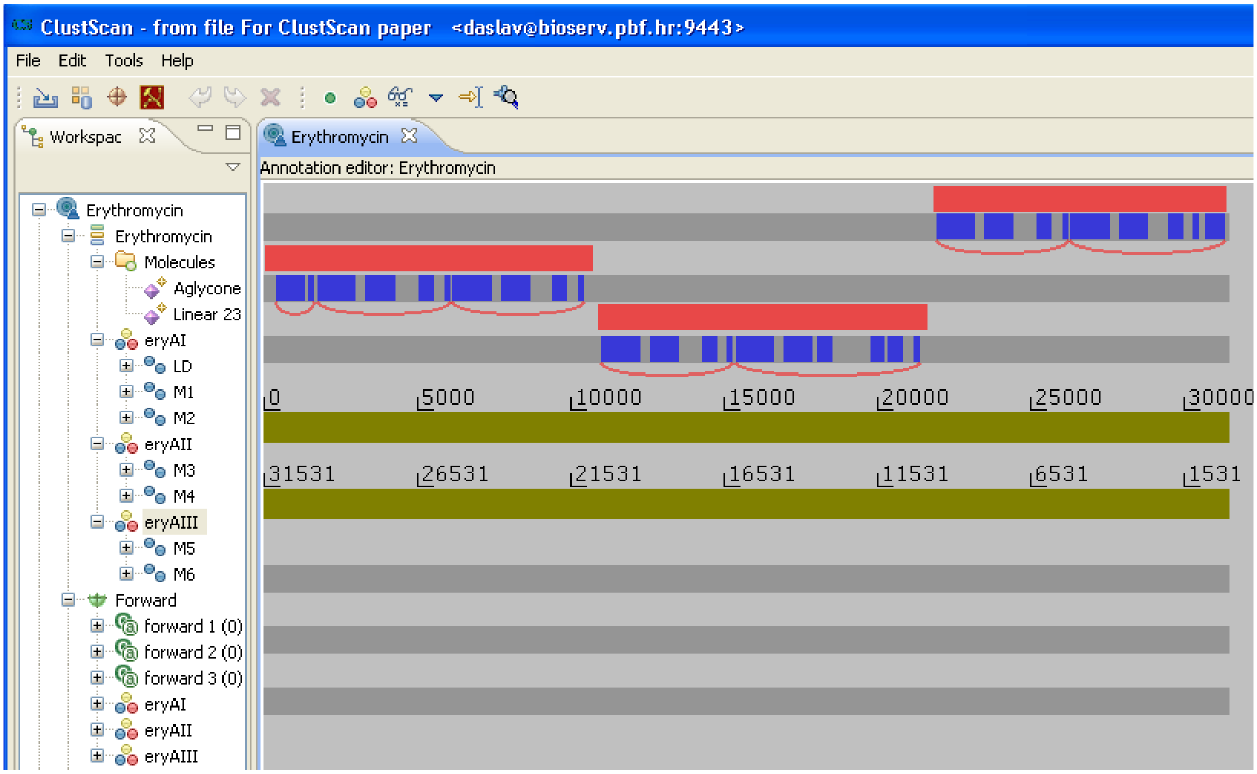
C2_1 C2_2 C2_3 C2_4	CC(=O); CC(O); C[C@H](O); C[C@@H](O); C=C; CC
C3_1 C3_2 C3_3 C3_4	C(C)C(=O); [C@H](C)C(=O); [C@@H](C)C(=O); C(C)C(O); C(C)[C@H](O); C(C)[C@@H](O); [C@H](C)C(O); [C@@H](C)C(O); [C@H](C)[C@H](O); [C@H](C)[C@@H](O); [C@@H](C)[C@H](O); [C@@H](C)[C@@H](O); C(C)=C; C(C)C [C@H](C)C [C@@H](C)C); C(C)C; [C@H](C)C [C@@H](C)C
MM_1 MM_2 MM_3 MM_4	C(CC)C(=O); [C@H](CC)C(=O); [C@@H](CC)C(=O); C(CC)C(O); C(CC)[C@H](O); C(CC)[C@@H](O); [C@H](CC)C(O); [C@@H](CC)C(O); [C@H](CC)[C@H](O); [C@H](CC)[C@@H](O); [C@@H](CC)[C@H](O); [C@@H](CC)[C@@H](O); C(CC)=C; C(CC)C; [C@H](CC)C; [C@@H](CC)C;
EM_1 EM_2 EM_3 EM_4	C(CC)C(=O); [C@H](CC)C(=O); [C@@H](CC)C(=O); C(CC)C(O) C(CC)[C@H](O); C(CC)[C@@H](O); [C@H](CC)C(O); [C@@H](CC)C(O); [C@H](CC)[C@H](O); [C@H](CC)[C@@H](O); [C@@H](CC)[C@H](O); [C@@H](CC)[C@@H](O); C(CC)=C); C(CC)C; [C@H](CC)C [C@@H](CC)C
MB_1 MB_2 MB_3 MB_4	CCC(C)C(=O); CC[C@H](C)C(=O); CC[C@@H](C)C(=O); CCC(C)C(O) CCC(C)[C@H](O); CCC(C)[C@@H](O); CC[C@H](C)C(O); CC[C@@H](C)C(O); CC[C@H](C)[C@H](O); CC[C@H](C)[C@@H](O); CC[C@@H](C)[C@H](O); CC[C@@H](C)[C@@H](O); CCC(C)=C; CCC(C)C; CC[C@H](C)C; CC[C@@H](C)C
CX_1	C(*)C(*)

Extenders: malonyl (C2), methylmalonyl (C3), methoxymalonyl (MM), ethylmalonyl (EM), unknown (CX)

Starters: malonyl (C2), propionyl (C3), methylbutyryl (MB)

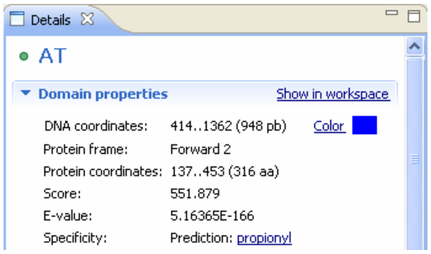
Fig. 1S Functional prediction of the DNA sequence (AY771999) of the erythromycin gene-cluster. The three genes (in red), six modules (underlined in red) and 29 catalytically active domains (in blue) of the erythromycin gene-cluster are shown. In particular, the inactive KR in module 3 responsible for the hydroxyl-stereochemistry: S stereochemistry was predicted. In module 4, the inability of the program to predict the stereochemical outcome from the KR module is unimportant, because of the presence of the DH domain that destroys the chirality. However, the inability to predict the β-carbon stereochemistry of the ER domain prevents prediction of the stereochemistry at this position in the final linear chain. (A). The isomeric SMILES as well as the 3-D structure of the predicted linear chain is also shown (B top and right). The cyclization function predicts a ring structure that can also be displayed (B left).

A



Gene: *ery*AI

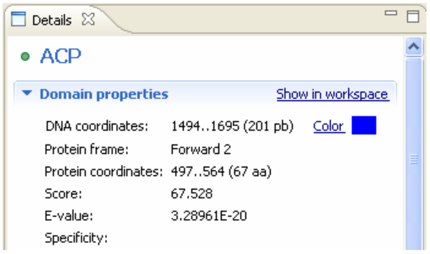
LD



AT

Domain properties

DNA coordinates:	414..1362 (948 pb)
Protein frame:	Forward 2
Protein coordinates:	137..453 (316 aa)
Score:	551.879
E-value:	5.16365E-166
Specificity:	Prediction: propionyl

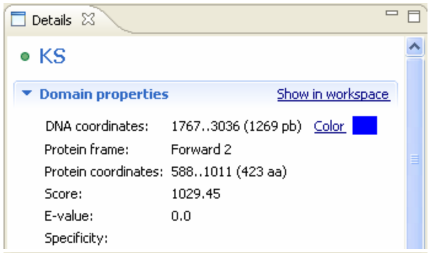


ACP

Domain properties

DNA coordinates:	1494..1695 (201 pb)
Protein frame:	Forward 2
Protein coordinates:	497..564 (67 aa)
Score:	67.528
E-value:	3.28961E-20
Specificity:	

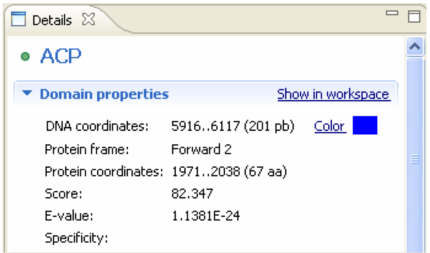
M1



KS

Domain properties

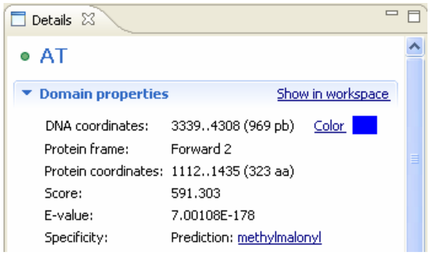
DNA coordinates:	1767..3036 (1269 pb)
Protein frame:	Forward 2
Protein coordinates:	588..1011 (423 aa)
Score:	1029.45
E-value:	0.0
Specificity:	



ACP

Domain properties

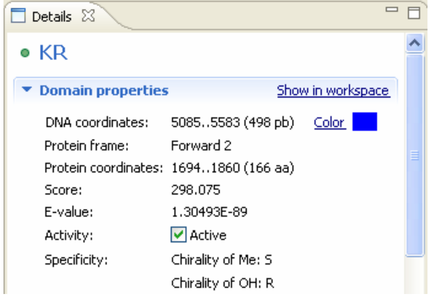
DNA coordinates:	5916..6117 (201 pb)
Protein frame:	Forward 2
Protein coordinates:	1971..2038 (67 aa)
Score:	82.347
E-value:	1.1381E-24
Specificity:	



AT

Domain properties

DNA coordinates:	3339..4308 (969 pb)
Protein frame:	Forward 2
Protein coordinates:	1112..1435 (323 aa)
Score:	591.303
E-value:	7.00108E-178
Specificity:	Prediction: methylmalonyl

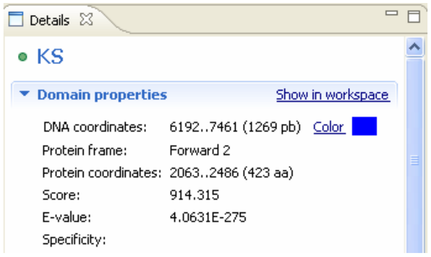


KR

Domain properties

DNA coordinates:	5085..5583 (498 pb)
Protein frame:	Forward 2
Protein coordinates:	1694..1860 (166 aa)
Score:	298.075
E-value:	1.30493E-89
Activity:	☑ Active
Specificity:	Chirality of Me: S Chirality of OH: R

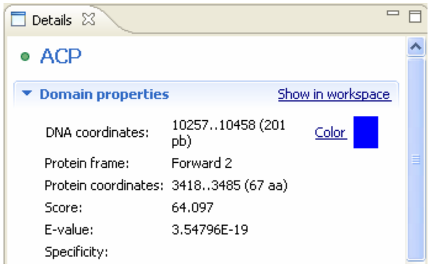
M2



KS

Domain properties

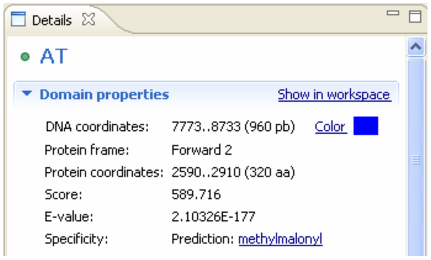
DNA coordinates:	6192..7461 (1269 pb)
Protein frame:	Forward 2
Protein coordinates:	2063..2486 (423 aa)
Score:	914.315
E-value:	4.0631E-275
Specificity:	



ACP

Domain properties

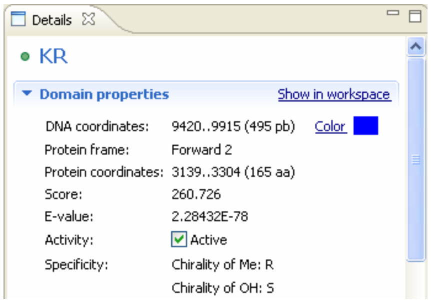
DNA coordinates:	10257..10458 (201 pb)
Protein frame:	Forward 2
Protein coordinates:	3418..3485 (67 aa)
Score:	64.097
E-value:	3.54796E-19
Specificity:	



AT

Domain properties

DNA coordinates:	7773..8733 (960 pb)
Protein frame:	Forward 2
Protein coordinates:	2590..2910 (320 aa)
Score:	589.716
E-value:	2.10326E-177
Specificity:	Prediction: methylmalonyl



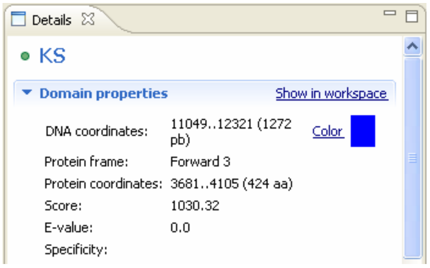
KR

Domain properties

DNA coordinates:	9420..9915 (495 pb)
Protein frame:	Forward 2
Protein coordinates:	3139..3304 (165 aa)
Score:	260.726
E-value:	2.28432E-78
Activity:	☑ Active
Specificity:	Chirality of Me: R Chirality of OH: S

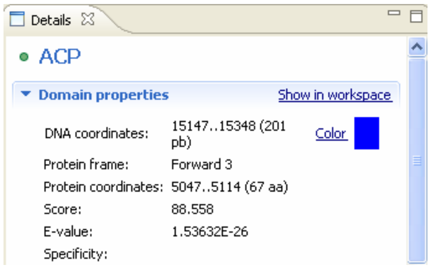
Gene: *ery*AII

M3



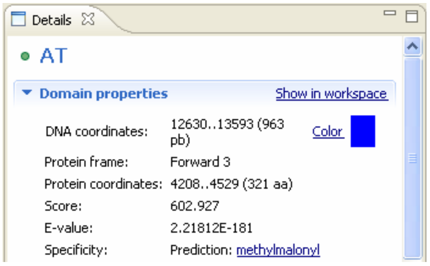
KS

DNA coordinates: 11049..12321 (1272 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 3681..4105 (424 aa)
Score: 1030.32
E-value: 0.0
Specificity:



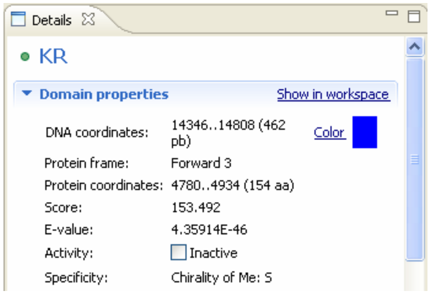
ACP

DNA coordinates: 15147..15348 (201 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 5047..5114 (67 aa)
Score: 88.558
E-value: 1.53632E-26
Specificity:



AT

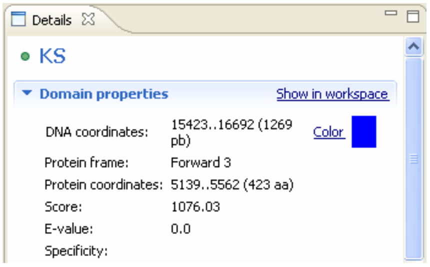
DNA coordinates: 12630..13593 (963 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 4208..4529 (321 aa)
Score: 602.927
E-value: 2.21812E-181
Specificity: Prediction: methylmalonyl



KR

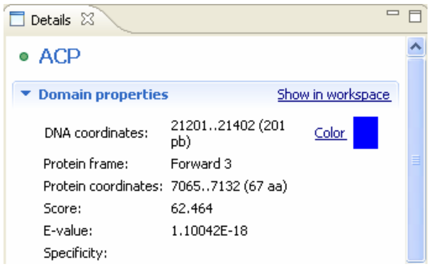
DNA coordinates: 14346..14808 (462 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 4780..4934 (154 aa)
Score: 153.492
E-value: 4.35914E-46
Activity: ☐ Inactive
Specificity: Chirality of Me: S

M4



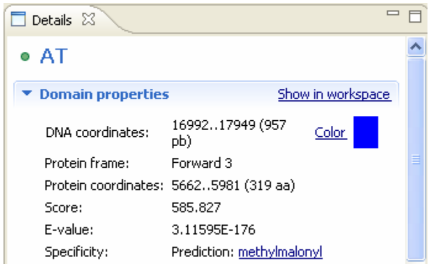
KS

DNA coordinates: 15423..16692 (1269 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 5139..5562 (423 aa)
Score: 1076.03
E-value: 0.0
Specificity:



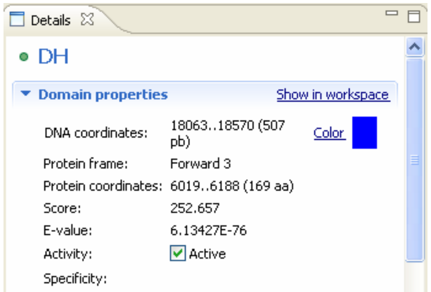
ACP

DNA coordinates: 21201..21402 (201 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 7065..7132 (67 aa)
Score: 62.464
E-value: 1.10042E-18
Specificity:



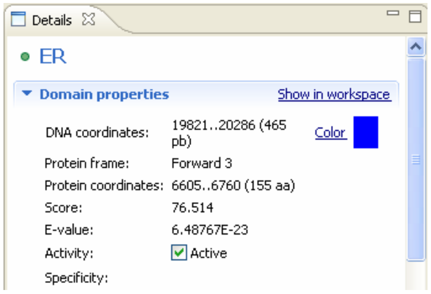
AT

DNA coordinates: 16992..17949 (957 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 5662..5981 (319 aa)
Score: 585.827
E-value: 3.11595E-176
Specificity: Prediction: methylmalonyl



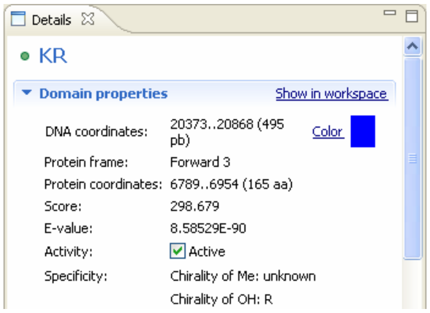
DH

DNA coordinates: 18063..18570 (507 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 6019..6188 (169 aa)
Score: 252.657
E-value: 6.13427E-76
Activity: ☑ Active
Specificity:



ER

DNA coordinates: 19821..20286 (465 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 6605..6760 (155 aa)
Score: 76.514
E-value: 6.48767E-23
Activity: ☑ Active
Specificity:

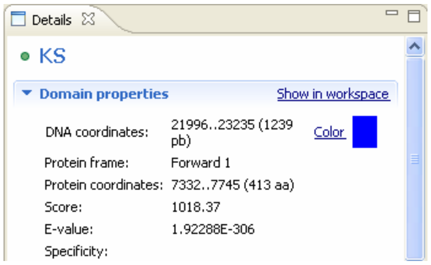


KR

DNA coordinates: 20373..20868 (495 pb) 🟦
Protein frame: Forward 3
Protein coordinates: 6789..6954 (165 aa)
Score: 298.679
E-value: 8.58529E-90
Activity: ☑ Active
Specificity: Chirality of Me: unknown
Chirality of OH: R

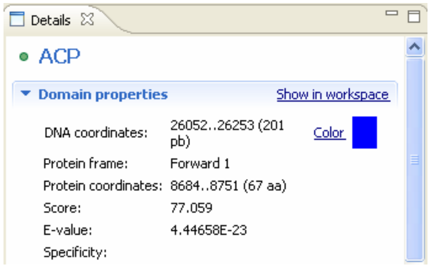
Gene: *ery*AIII

M5



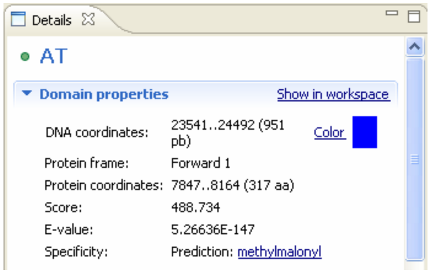
KS

DNA coordinates: 21996..23235 (1239 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 7332..7745 (413 aa)
Score: 1018.37
E-value: 1.92288E-306
Specificity:



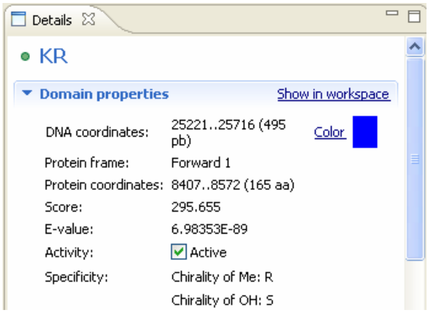
ACP

DNA coordinates: 26052..26253 (201 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 8684..8751 (67 aa)
Score: 77.059
E-value: 4.44658E-23
Specificity:



AT

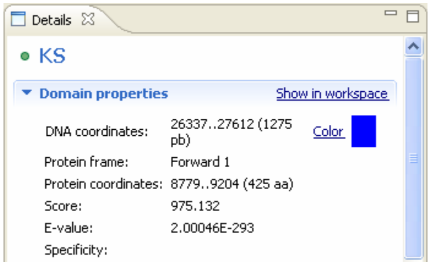
DNA coordinates: 23541..24492 (951 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 7847..8164 (317 aa)
Score: 488.734
E-value: 5.26636E-147
Specificity: Prediction: methylmalonyl



KR

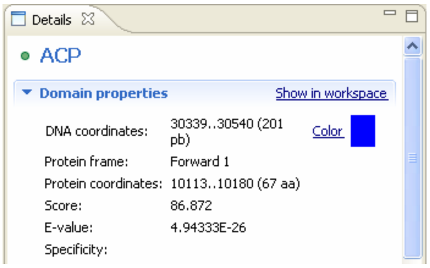
DNA coordinates: 25221..25716 (495 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 8407..8572 (165 aa)
Score: 295.655
E-value: 6.98353E-89
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: S

M6



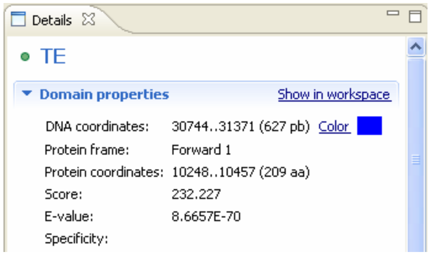
KS

DNA coordinates: 26337..27612 (1275 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 8779..9204 (425 aa)
Score: 975.132
E-value: 2.00046E-293
Specificity:



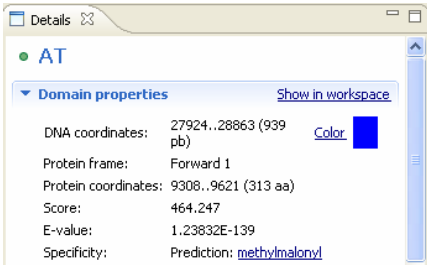
ACP

DNA coordinates: 30339..30540 (201 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 10113..10180 (67 aa)
Score: 86.872
E-value: 4.94333E-26
Specificity:



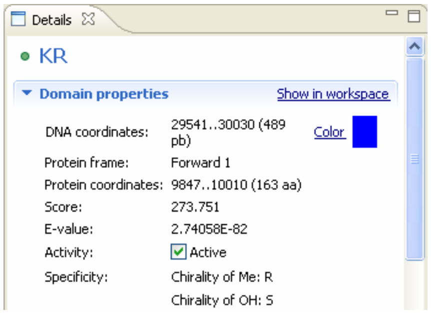
TE

DNA coordinates: 30744..31371 (627 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 10248..10457 (209 aa)
Score: 232.227
E-value: 8.6657E-70
Specificity:



AT

DNA coordinates: 27924..28863 (939 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 9308..9621 (313 aa)
Score: 464.247
E-value: 1.23832E-139
Specificity: Prediction: methylmalonyl



KR

DNA coordinates: 29541..30030 (489 pb) 🟦
Protein frame: Forward 1
Protein coordinates: 9847..10010 (163 aa)
Score: 273.751
E-value: 2.74058E-82
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: S

B

SMILES:
[C@H](C)[C@@H](O1)[C@@H](C)[C@H](O)[C@H](C)C(=O)C(C)
C[C@@H](C)[C@H](O)[C@@H](C)[C@H](O)C(C)C1(=O)

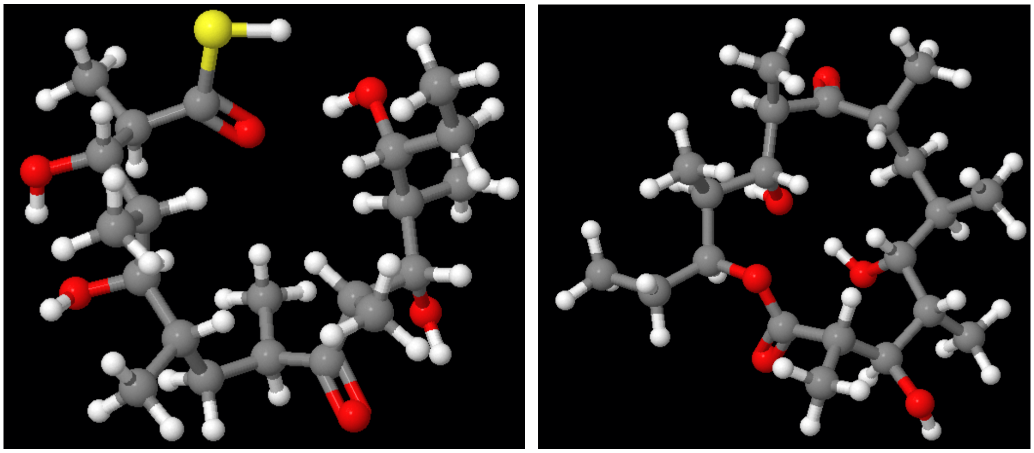
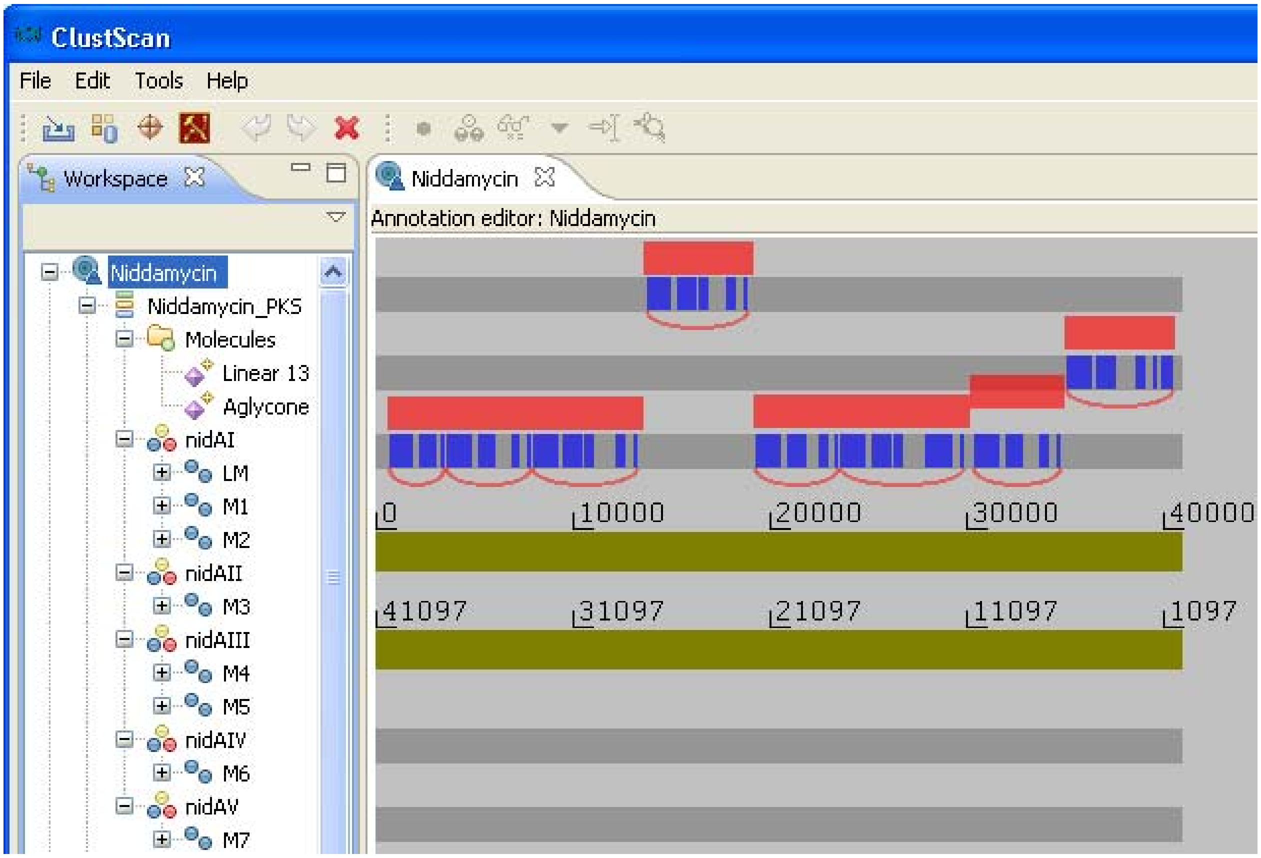


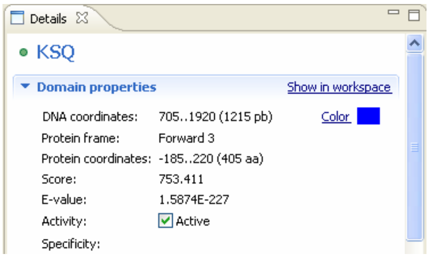
Fig. 1S Functional prediction of the DNA sequence (AF016585) of the niddamycin gene-cluster. The five genes (in red), seven modules (underlined in red) and 36 catalytically active domains (in blue) of the niddamycin gene-cluster are shown. In particular, the inactive KR in module 4 responsible for the hydroxyl-stereochemistry: S stereochemistry was predicted (A). The isomeric SMILES as well as the 3-D structure of the predicted linear chain is also shown (B top and right). The cyclization function predicts a ring structure that can also be displayed (B left).

A



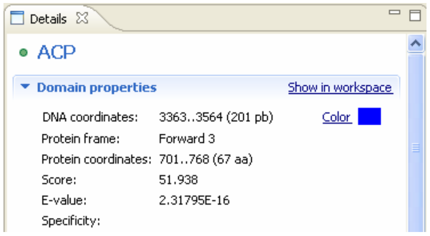
Gene: *nid*AI

LM



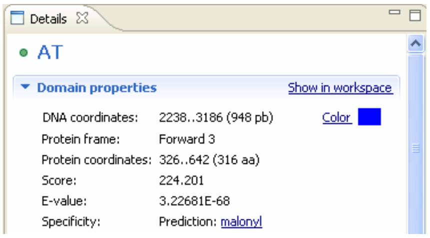
KSQ

DNA coordinates: 705..1920 (1215 pb)
Protein frame: Forward 3
Protein coordinates: -185..220 (405 aa)
Score: 753.411
E-value: 1.5874E-227
Activity: ☑ Active
Specificity:



ACP

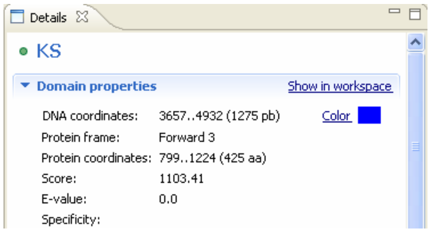
DNA coordinates: 3363..3564 (201 pb)
Protein frame: Forward 3
Protein coordinates: 701..768 (67 aa)
Score: 51.938
E-value: 2.31795E-16
Specificity:



AT

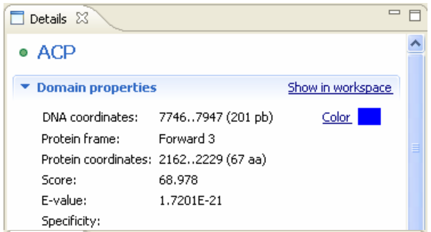
DNA coordinates: 2238..3186 (948 pb)
Protein frame: Forward 3
Protein coordinates: 326..642 (316 aa)
Score: 224.201
E-value: 3.22681E-68
Specificity: Prediction: malonyl

M1



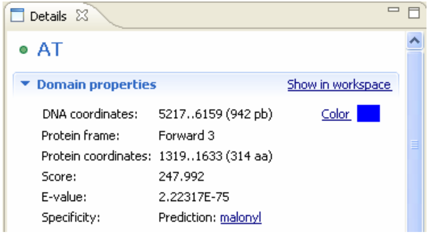
KS

DNA coordinates: 3657..4932 (1275 pb)
Protein frame: Forward 3
Protein coordinates: 799..1224 (425 aa)
Score: 1103.41
E-value: 0.0
Specificity:



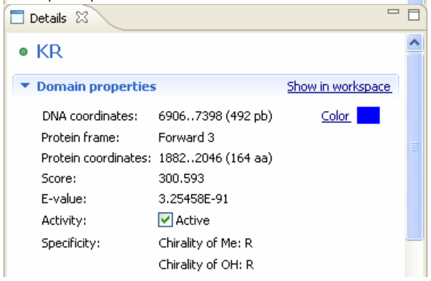
ACP

DNA coordinates: 7746..7947 (201 pb)
Protein frame: Forward 3
Protein coordinates: 2162..2229 (67 aa)
Score: 68.978
E-value: 1.7201E-21
Specificity:



AT

DNA coordinates: 5217..6159 (942 pb)
Protein frame: Forward 3
Protein coordinates: 1319..1633 (314 aa)
Score: 247.992
E-value: 2.22317E-75
Specificity: Prediction: malonyl



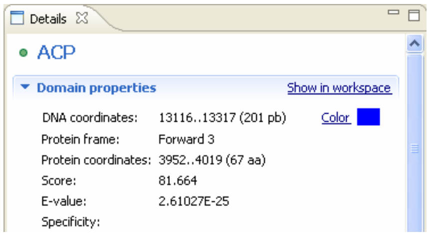
KR

DNA coordinates: 6906..7398 (492 pb)
Protein frame: Forward 3
Protein coordinates: 1882..2046 (164 aa)
Score: 300.593
E-value: 3.25458E-91
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: R

M2

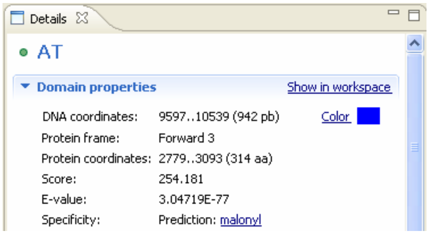
KS

DNA coordinates:	8037..9312 (1275 pb) Color 🟦
Protein frame:	Forward 3
Protein coordinates:	2259..2684 (425 aa)
Score:	1109.98
E-value:	0.0
Specificity:	



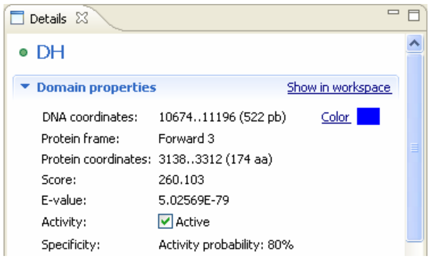
ACP

DNA coordinates:	13116..13317 (201 pb) Color 🟦
Protein frame:	Forward 3
Protein coordinates:	3952..4019 (67 aa)
Score:	81.664
E-value:	2.61027E-25
Specificity:	



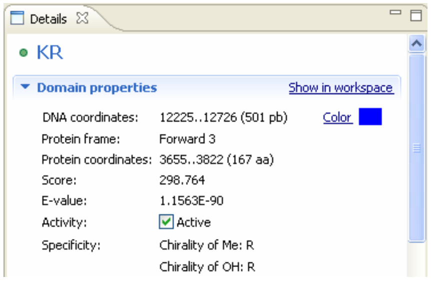
AT

DNA coordinates:	9597..10539 (942 pb) Color 🟦
Protein frame:	Forward 3
Protein coordinates:	2779..3093 (314 aa)
Score:	254.181
E-value:	3.04719E-77
Specificity:	Prediction: malonyl



DH

DNA coordinates:	10674..11196 (522 pb) Color 🟦
Protein frame:	Forward 3
Protein coordinates:	3138..3312 (174 aa)
Score:	260.103
E-value:	5.02569E-79
Activity:	☑ Active
Specificity:	Activity probability: 80%

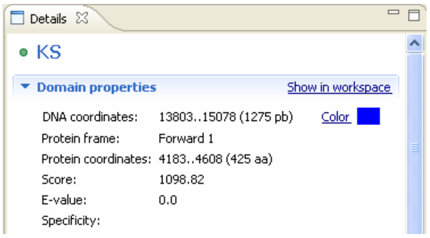


KR

DNA coordinates:	12225..12726 (501 pb) Color 🟦
Protein frame:	Forward 3
Protein coordinates:	3655..3822 (167 aa)
Score:	298.764
E-value:	1.1563E-90
Activity:	☑ Active
Specificity:	Chirality of Me: R Chirality of OH: R

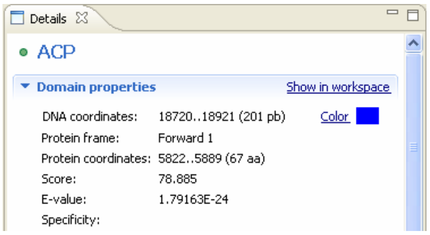
Gene: *nid*AII

M3



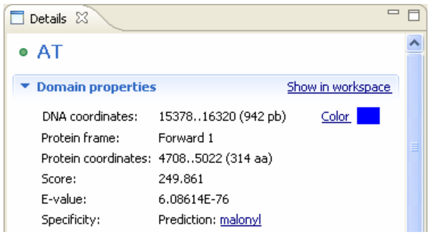
KS

DNA coordinates:	13803..15078 (1275 pb) Color 🟦
Protein frame:	Forward 1
Protein coordinates:	4183..4608 (425 aa)
Score:	1098.82
E-value:	0.0
Specificity:	



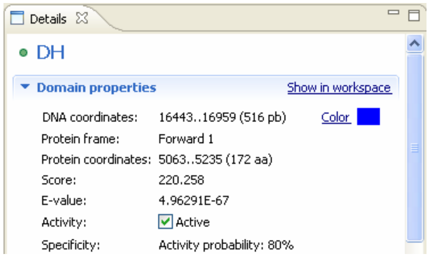
ACP

DNA coordinates:	18720..18921 (201 pb) Color 🟦
Protein frame:	Forward 1
Protein coordinates:	5822..5889 (67 aa)
Score:	78.885
E-value:	1.79163E-24
Specificity:	



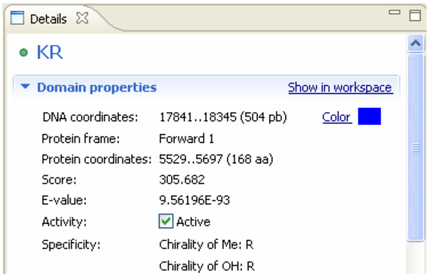
AT

DNA coordinates:	15378..16320 (942 pb) Color 🟦
Protein frame:	Forward 1
Protein coordinates:	4708..5022 (314 aa)
Score:	249.861
E-value:	6.08614E-76
Specificity:	Prediction: malonyl



DH

DNA coordinates:	16443..16959 (516 pb) Color 🟦
Protein frame:	Forward 1
Protein coordinates:	5063..5235 (172 aa)
Score:	220.258
E-value:	4.96291E-67
Activity:	☑ Active
Specificity:	Activity probability: 80%

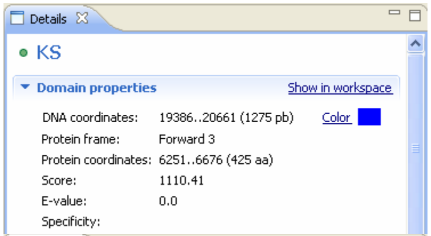


KR

DNA coordinates:	17841..18345 (504 pb) Color 🟦
Protein frame:	Forward 1
Protein coordinates:	5529..5697 (168 aa)
Score:	305.682
E-value:	9.56196E-93
Activity:	☑ Active
Specificity:	Chirality of Me: R Chirality of OH: R

Gene: *nid*AIII

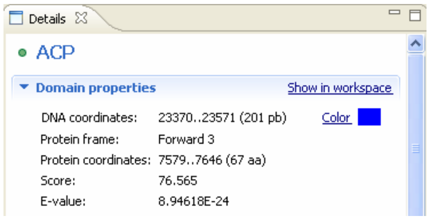
M4



KS

Domain properties

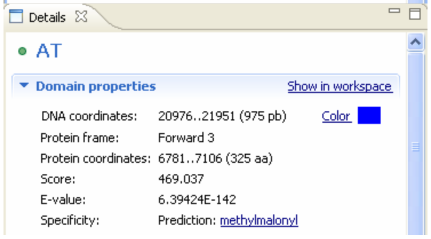
DNA coordinates: 19386..20661 (1275 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 6251..6676 (425 aa)
Score: 1110.41
E-value: 0.0
Specificity:



ACP

Domain properties

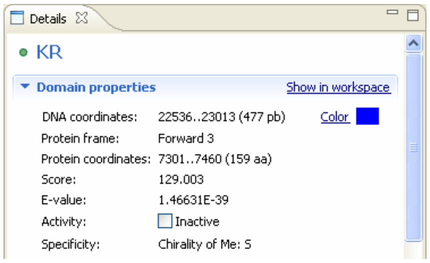
DNA coordinates: 23370..23571 (201 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 7579..7646 (67 aa)
Score: 76.565
E-value: 8.94618E-24



AT

Domain properties

DNA coordinates: 20976..21951 (975 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 6781..7106 (325 aa)
Score: 469.037
E-value: 6.39424E-142
Specificity: Prediction: methylmalonyl

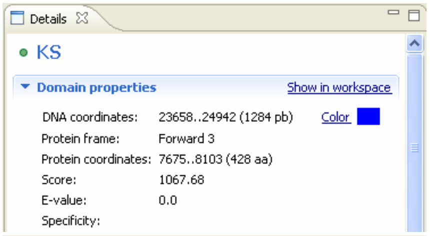


KR

Domain properties

DNA coordinates: 22536..23013 (477 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 7301..7460 (159 aa)
Score: 129.003
E-value: 1.46631E-39
Activity: ☐ Inactive
Specificity: Chirality of Me: S

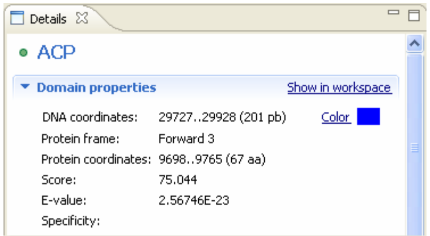
M5



KS

Domain properties

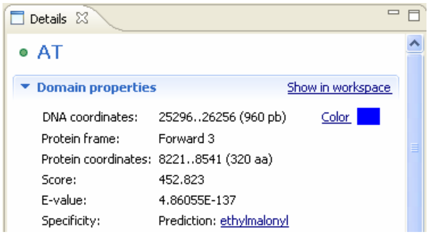
DNA coordinates: 23658..24942 (1284 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 7675..8103 (428 aa)
Score: 1067.68
E-value: 0.0
Specificity:



ACP

Domain properties

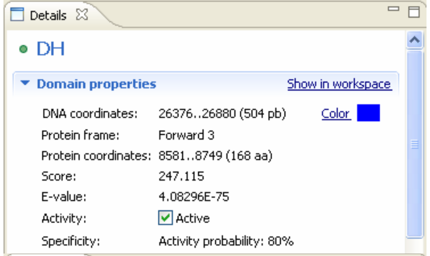
DNA coordinates: 29727..29928 (201 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 9698..9765 (67 aa)
Score: 75.044
E-value: 2.56746E-23
Specificity:



AT

Domain properties

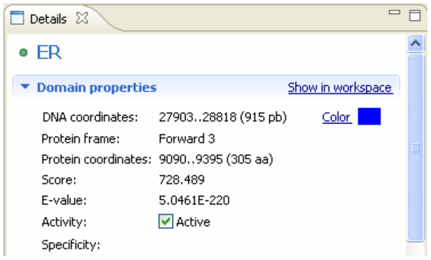
DNA coordinates: 25296..26256 (960 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 8221..8541 (320 aa)
Score: 452.823
E-value: 4.86055E-137
Specificity: Prediction: ethylmalonyl



DH

Domain properties

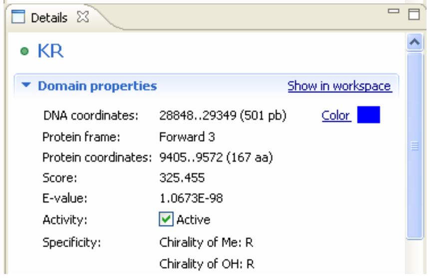
DNA coordinates: 26376..26880 (504 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 8581..8749 (168 aa)
Score: 247.115
E-value: 4.08296E-75
Activity: ☑ Active
Specificity: Activity probability: 80%



ER

Domain properties

DNA coordinates: 27903..28818 (915 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 9090..9395 (305 aa)
Score: 728.489
E-value: 5.0461E-220
Activity: ☑ Active
Specificity:



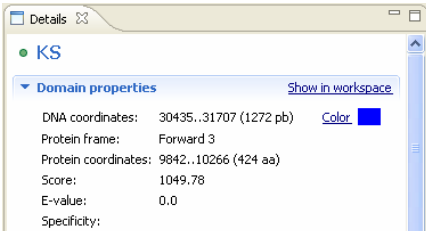
KR

Domain properties

DNA coordinates: 28848..29349 (501 pb) Color 🟦
Protein frame: Forward 3
Protein coordinates: 9405..9572 (167 aa)
Score: 325.455
E-value: 1.0673E-98
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: R

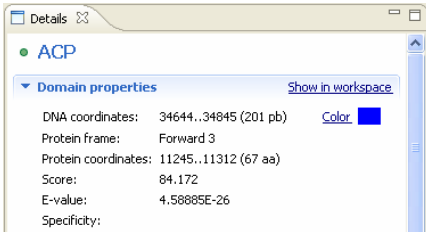
Gene: *nid*AIV

M6



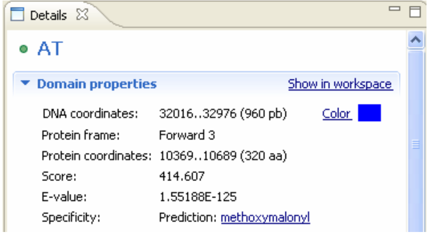
KS

DNA coordinates: 30435..31707 (1272 pb)
Protein frame: Forward 3
Protein coordinates: 9842..10266 (424 aa)
Score: 1049.78
E-value: 0.0
Specificity:



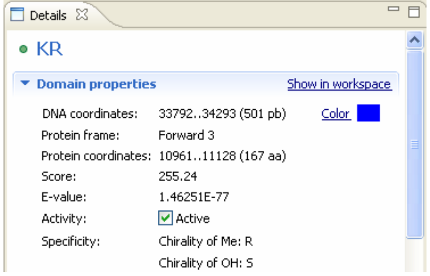
ACP

DNA coordinates: 34644..34845 (201 pb)
Protein frame: Forward 3
Protein coordinates: 11245..11312 (67 aa)
Score: 84.172
E-value: 4.58885E-26
Specificity:



AT

DNA coordinates: 32016..32976 (960 pb)
Protein frame: Forward 3
Protein coordinates: 10369..10689 (320 aa)
Score: 414.607
E-value: 1.55188E-125
Specificity: Prediction: methoxymalonyl

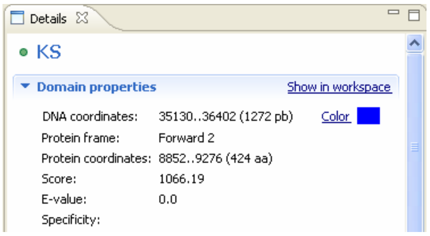


KR

DNA coordinates: 33792..34293 (501 pb)
Protein frame: Forward 3
Protein coordinates: 10961..11128 (167 aa)
Score: 255.24
E-value: 1.46251E-77
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: S

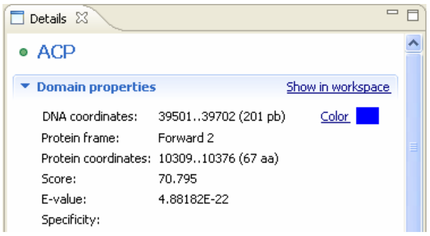
Gene: *nid*AV

M7



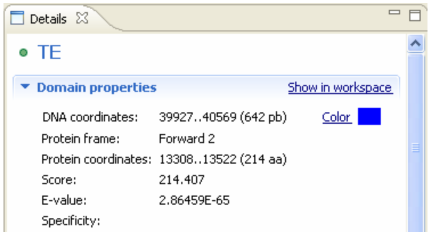
KS

DNA coordinates: 35130..36402 (1272 pb)
Protein frame: Forward 2
Protein coordinates: 8852..9276 (424 aa)
Score: 1066.19
E-value: 0.0
Specificity:



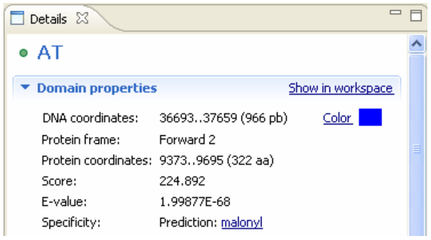
ACP

DNA coordinates: 39501..39702 (201 pb)
Protein frame: Forward 2
Protein coordinates: 10309..10376 (67 aa)
Score: 70.795
E-value: 4.88182E-22
Specificity:



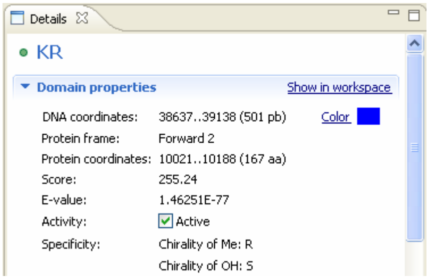
TE

DNA coordinates: 39927..40569 (642 pb)
Protein frame: Forward 2
Protein coordinates: 13308..13522 (214 aa)
Score: 214.407
E-value: 2.86459E-65
Specificity:



AT

DNA coordinates: 36693..37659 (966 pb)
Protein frame: Forward 2
Protein coordinates: 9373..9695 (322 aa)
Score: 224.892
E-value: 1.99877E-68
Specificity: Prediction: malonyl



KR

DNA coordinates: 38637..39138 (501 pb)
Protein frame: Forward 2
Protein coordinates: 10021..10188 (167 aa)
Score: 255.24
E-value: 1.46251E-77
Activity: ☑ Active
Specificity: Chirality of Me: R
Chirality of OH: S

B

SMILES:
C[C@@H](O)C=CC=CCC(=O)[C@@H](C)C[C@@H](CC)
[C@H](O)[C@@H](CC)[C@H](O)CC(=O)S