

Electronic Supplementary Material for:

Applied Microbiology and Biotechnology

**Identification of catalytically important residues of the carotenoid  
1,2-hydratases from *Rubrivivax gelatinosus* and *Thiocapsa roseopersicina***

**Aida Hiseni, Linda G. Otten and Isabel W.C.E. Arends\***

Biocatalysis Group, Department of Biotechnology, Delft University of Technology, Delft, NL

\* Corresponding author: Prof. Dr. Isabel W.C.E. Arends, Biocatalysis Group, Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands. Tel.: +31 15 2786659; Fax: +31 15 2781415; E-mail: I.W.C.E.Arends@tudelft.nl; URL: [www.bt.tudelft.nl/boc](http://www.bt.tudelft.nl/boc).

Contents

<b>1. Importance of the N-terminal part of CrtC</b>	<b>p.24</b>
<b>2. Homology model of CrtC</b>	<b>p.26</b>
<b>Table S1</b>	<b>p.24</b>
<b>Figure S1</b>	<b>p.2</b>
<b>Figure S2</b>	<b>p.25</b>
<b>Figure S3</b>	<b>p.26</b>
<b>Figure S4</b>	<b>p.27</b>
<b>Figure S5</b>	<b>p.28</b>

**Figure S1.** Multiple sequence alignment of all CrtC protein sequences showing conserved amino acids. Identical amino acids are marked with an asterisk. *RgCrtC* sequence is highlighted in yellow (next page).







```

gi|495340016|ref|WP_008064750.1|-----FS 2
gi|517223140|ref|WP_018411958.1|-----ARRFD 5
gi|889919304|ref|WP_048861895.1|-----FS 2
gi|496441596|ref|WP_009150441.1|DGVFQSTGTHQAAGPVSGGRPPPTGTAEGDPIRPSDSADRARCPAFD 77
gi|146409846|gb|ABQ38352.1|_10|DGVVPAPGGADETAGTLHSRCGRASGARADGGAVGFDGSRAGDSGPRFD 91
gi|908659074|ref|WP_049795525.1|-----
gi|365289544|emb|CCD93979.1|_1|DGVVPAAGCADEAGRALRGGGRASGTGRADGSAVGPDGSRTGDSGPRFD 91
gi|41018902|gb|AAR98493.1|_10|DGIIPAAGRADEAARALRGGGRPSGTGRPDGGLVGPDGRRQDLRPRFD 91
gi|653027709|ref|WP_027279528.1|-----FN 2
gi|499791021|ref|WP_011471755.1|-----FN 2
gi|499823412|ref|WP_011504146.1|-----FN 2
gi|499762149|ref|WP_011442883.1|-----FN 2
gi|829076993|ref|WP_047310044.1|-----FN 2
gi|503269006|ref|WP_013503667.1|-----FN 2
gi|499470438|ref|WP_011157078.1|-----FN 2
gi|501486651|ref|WP_012495141.1|-----FN 2
gi|550975051|ref|WP_022723265.1|-----FN 2
gi|764538658|ref|WP_044412067.1|-----FN 2
gi|754113006|ref|WP_041748751.1|-----MGGSPARGPDFS 13
gi|737680904|ref|WP_035649991.1|-----DRGPDFS 7
gi|503405597|ref|WP_013640258.1|-----RFD 3
gi|657196689|ref|WP_029312388.1|-----RFD 3
gi|703385326|ref|WP_033417843.1|-----
gi|653032131|ref|WP_027283892.1|-----FD 2
gi|378403832|emb|CCG08948.1|_1|ERLLFSDGERDGAAGPVRGGRPSGAGSDGGALGPAGTGGPAGDLM 82
gi|754176764|ref|WP_041795395.1|VGDLRPAGGENGDARPLSGGQRPSGTGGADGGLIGTVGGGPGGSGGPDFA 89
gi|25989732|gb|AAN75036.1|_12|-----
gi|860504149|ref|WP_048445338.1|-----
gi|860474721|ref|WP_048434413.1|-----
gi|501291381|ref|WP_012333472.1|-----
gi|860453591|ref|WP_048427078.1|-----
gi|860560381|ref|WP_048464709.1|-----
gi|860527577|ref|WP_048454269.1|-----
gi|760871815|dbj|BAQ48812.1|_1|-----PRFD 4
gi|373567336|gb|EHP93308.1|_6|-----
gi|24009448|gb|ACS40674.1|_4|-----
gi|218523206|gb|ACK83791.1|_4|-----
gi|254269168|emb|CAX25134.1|_4|-----
gi|163663749|gb|ABY31116.1|_2|-----
gi|494837541|ref|WP_007563641.1|-----
gi|914802769|ref|WP_050733395.1|-----
gi|657246165|ref|WP_029357447.1|-----
gi|501276758|ref|WP_012319776.1|-----
gi|640229117|ref|WP_024829694.1|-----
gi|518936109|ref|WP_020091984.1|-----
gi|760075276|ref|WP_043757989.1|-----

```

```

gi | 759662308 | ref | WP_043379914. | -----
gi | 498368472 | ref | WP_010682628. | -----
gi | 497224411 | ref | WP_009538673. | -----
gi | 589291345 | gb | AHK79488.1 | _12
gi | 644125233 | ref | WP_025281955. | -----
gi | 496218797 | ref | WP_008932834. | -----
gi | 495659402 | ref | WP_008383981. | -----
gi | 687960265 | gb | KGE05197.1 | _16
gi | 496312569 | ref | WP_009021747. | -----
gi | 495571757 | ref | WP_008296336. | -----
gi | 908380581 | ref | WP_049751705. | -----
gi | 658524263 | ref | WP_029718812. | -----
gi | 332109922 | gb | FGJ10550.1 | _1-
gi | 760103377 | ref | WP_043785587. | -----
gi | 760129436 | ref | WP_043810804. | -----
gi | 7416798 | gb | BAA94047.1 | _1-4
gi | 75345570 | sp | p95619.1 | CRTC_R
gi | 913634290 | gb | KNZ32951.1 | _1-
gi | 809277751 | ref | WP_046331236. | -----
gi | 518258896 | ref | WP_019429104. | -----
gi | 518255752 | ref | WP_019425960. | -----
gi | 918152310 | gb | KOF08324.1 | _1-
gi | 61653243 | gb | AAX48196.1 | _1-2

VSVPRGYAWWYVDALSDDGKHGLTLIAFVGSVFSPIYFRGRRGL--- 146
--VGPFGYAWWYVDALSDDGKHGLTLIAFVGSVFSPIYFRGRRGL--- 44
-----MDALSDDGKHGLTLIAFVGSVFSPIYFKGRRRNK--- 34
-----MDALSDDGKHGLTLIAFVGSVFSPIYFKGRRRGR--- 34
IDVPPRGYAWWYVDALSDDGKHGLTLIAFVGSVFSPIYFKGHRGK--- 51
IDVPPRGYAWWYVDALSDDGKHGLTLIAFVGSVFSPIYFKGHRGK--- 51
RTIPSGYSWWYIDALSDDGKHGLTLIAFVGSVFSPIYAWARRKGG--T 92
--MPGSGYSWWYIDALSDDGKHGLTLIAFVGSVFSPIYAWARRKGG--T 45
RQIPRSGYSWWYIDALSDDGKHGLTLIAFVGSVFSPIYAWARRRGGG--A 129
-----MSDDGKHGLTLIAFVGSVFSPIYAWARRRGR--- 31
RKVPRDGYAWWYVDALSDDGKHGLTLIAFVGSVFSPIYAWARRRGE--- 48
-----MSDDGKHGLTLIAFVGSVFSPIYAWARRRGR--- 31
VHVAANGYAWWYIDATSDAGKHGLTLIAFVGSVFSPIYKLSGR--- 47
VAVAPNGYAWWYIDATSDAGKHGLTLIAFVGSVFSPIYKLSGR--- 47
-----MSADGKHGLTLIAFVGSVFSPIYAWMSGR--- 28
ETVEPNGYAWWYVDASFADHKYGLTLIAFVGSVFSPIYAKERRRAI--- 119
-----MSDDGKHGLTLIAFVGSVFSPIYAWARARET--- 31
-----MSDDGKHGLTLIAFVGSVFSPIYAWMSGW--- 28
LTVNDGYVWVYVDALSDDEQYGLTLIAMLGSVFSPIYAWMSRQ---GN 119
-----MSDDEQYGLTLIAMLGSVFSPIYAWMSRQ---GN 31
-----MSDDGKHGLTLIALIGSVFSPIYAWMSRQ---GQ 31

```







gi   739596279	ref   WP_037453842.	GDPLNHCTVNVVLTGAGGA	---	RWCMTBTRGRGKTRRDASSLTI	GFSSLN	77
gi   764363842	ref   WP_044336870.	GDPEHCAINVSLNGPRGS	---	AWAMTBRPSKRMRESADHF	FAVGFSGMH	93
gi   644239143	ref   WP_025293539.	QRPEHCAINVSLNGPRGS	---	AWAMTBRPSARVHF	SADHFSVGFSGVR	93
gi   503035479	ref   WP_013270455.	GAPENHCALNVAFYGPDCQ	---	RWTMTBRPAGQVEREARHI	QIGFSSLS	74
gi   617472745	gb KAI93218.1 _2-	VDPAQHCSINVAVYGRGAA	---	RWAMTECRHALGATPEQLA	IGFSSLS	165
gi   737603811	ref   WP_035574467.	VPAENHVALNVALYGPRG	---	AWAMTBRGAGALARERADRF	VVVPSSLT	76
gi   648530966	ref   WP_026222717.	ADPFNHCANVVALYGPRGA	---	RWAMTBRGRKSLKFRDVL	LLIGFSSLE	74
gi   288897641	gb ADC63477.1 _28	ADPLDHCALNVALYGRRGK	---	RWAMTBRGRAALQRDR	TNLVIGPSSHVH	165
gi   760114352	ref   WP_043796307.	ADPLDHCALNVALYGRRGK	---	RWAMTBRGRAALQRDR	TNLVIGPSSHVH	77
gi   494098690	ref   WP_007039487.	ADPLDHCALNVALYGRNGK	---	RWAMTBRGRTSLHRTGDS	LTIIGFSSALA	77
gi   75388571	sp Q7X3G5.1 CRTC_T	PDPLNHCAINVALYQKAGK	---	RWTMTBRGRKALRQAPGR	LLDIGPSSHLT	166
gi   390426908	gb AFL73973.1 _35	SDPLNHNSLNVALYQ-ACK	---	RWSMTBRGRSALRRDAS	RRLEIGFSSALS	126
gi   752807498	ref   WP_041446922.	SDPLNHNSLNVALYQ-ACK	---	RWSMTBRGRSALRRDAS	RRLEIGFSSALS	87
gi   760074757	ref   WP_043757486.	SDPLNHNSLNVALYQ-EVK	---	RWTMTBRGRSALRRDAS	SLEIGFSSALS	95
gi   918211006	ref   WP_052348303.	SDPLNHNSLNVALYQ-EVK	---	RWTMTBRGRSALRRDAS	SLEIGFSSALS	87
gi   499982040	ref   WP_011662758.	GDPLHHVSLNVALYQ-EGK	---	RWSMTBRGRSLODKT	WLQIGFSSLE	96
gi   493840365	ref   WP_006787509.	LDPLNHCAINVALYGPNGK	---	RWALDTRGRSSVERGR	DFLRIGFSQLC	80
gi   653037601	ref   WP_027289289.	ADPLNHCSINVALYGEQPN	---	RWAMTBRGRTRIDR	GARHFSVGFSGLH	103
gi   519009821	ref   WP_020165696.	ADPHAHCANVVALYGDK-R	---	RWTMTBRGASALQR	DTHSLRIGFSRLH	99
gi   518759539	ref   WP_019916969.	ADPHAHCANVVALYGDK-R	---	RWTMTBRGASALQR	DAHSLRIGFSRLH	95
gi   495340016	ref   WP_008064750.	ADPHAHCANVVALYGDK-R	---	RWAMTBRGASALQR	DARNLRIGFSRLH	95
gi   517223140	ref   WP_018411958.	ADPHAHCANVVALYQDN-R	---	RWTMTBRGPQALQR	DAQHLRIGFSLH	97
gi   889919304	ref   WP_048861895.	VDPERHCAVNVLYGERGR	---	RWTMTBRGASHLHRT	ATSFALGPSTLA	94
gi   496441596	ref   WP_009150441.	ADPEHFCAINVAIVYASGK	---	RWAMTBRGRQLSR	TPTELRIGFSQLH	169
gi   146409846	gb ABQ38352.1 _10	ADPEDHCALNVALYGARAR	---	RWSMTBRGRGAI	VRDAESFKIGFSRLA	183
gi   908659074	ref   WP_049795525.	ADPEDHCALNVALYGARAR	---	RWSMTBRGRGAI	VRDAESFKIGFSRLA	90
gi   365289544	emb CCD93979.1 _1	AEPENHCALNVALYKRRAR	---	RWSMTBRGRGAI	ARDEARFAVGFSSLS	183
gi   41018902	gb AAR98493.1 _10-	ADAENHCALNVALYGERAR	---	RWAMTBRGRGAI	ARDEARFAVGFSSLA	183
gi   653027709	ref   WP_027279528.	VDPNMHCAINVALYRENGN	---	RWTMTBRGRGHVER	SATQLSIGFSRVF	94
gi   499791021	ref   WP_011471755.	TDPLNHCAINVALYRNTGN	---	RWTMTBRGRRTVAF	SAPSLVIGFSDLS	94
gi   499823412	ref   WP_011504146.	ADPLNHCAINVLGYHSAGK	---	RWTMTBRARGRVAR	NLSLLIIGFSDLA	94
gi   499762149	ref   WP_011442883.	ADPLNHCAINVALYHSAGK	---	RWTMTBRGRVARNL	HSLLIIGFSDLA	94
gi   829076993	ref   WP_047310044.	ADPLDHCALNVALYHQAGK	---	RWSMTBRGRGRVAR	NVNAFLIGPDDL	94
gi   503269006	ref   WP_013503667.	ADPLDHCALNVALYHQAGK	---	RWSMTBRGRGRVAR	NVNAFLIGPDDL	94
gi   499470438	ref   WP_011157078.	ADPLDHCALNVALYHEAGK	---	RWTMTBRGRGRVAR	NANSFLIGPDDL	94
gi   501486651	ref   WP_012495141.	ADPLDHCALNVALYHEAGK	---	RWTMTBRGRGRVAR	NANSFLIGPDDL	94
gi   509975051	ref   WP_022732365.	ADPLDHCALNVALYRPGGK	---	TWTMTBRGRGRVAR	NVHSFLIGPDDL	94
gi   764538658	ref   WP_044412067.	ADPMNHCAINVALYRPGGN	---	RWTMTBRARGRVAR	TVNALLVGPDDL	94
gi   754113006	ref   WP_041748751.	ADPLNHCAINVAIVYRSGGN	---	RWAMTBRPRGAVG	FTDTFTVGFPSRLC	105
gi   737680904	ref   WP_035649991.	ADPINHCAINVAIVYRSGGN	---	RWAMTBRPEAVSR	IDTFAVGFPSNLS	99
gi   503405597	ref   WP_013640258.	AAPEDFVAINALYRPGGN	---	RWTMTERRRHS	LRAEADSLAIGFSRLA	97
gi   657196689	ref   WP_029312388.	ADPEYVAINALYRPRRT	---	LWAMTERRRRSL	IREPGTLIGFSRLD	97
gi   703385326	ref   WP_033417843.	SEPLDHSAINLALYKSSH	---	RWAMTERRRRGS	VLFAPDVLAIIGFSELR	87
gi   653032131	ref   WP_027283892.	GDPMHCCLNVALYQ-RPR	---	RWAMTERRRRGAL	ARGADFLQIGFSRLD	93
gi   378403832	emb CCG08948.1 _1	-DPHDHVAINVALYQAPQG	---	WTMTBRGRRAL	ARDTRLTIIGFSSALS	170
gi   754176764	ref   WP_041795395.	-DPHDHVAINVALYQAPQG	---	WTMTBRGRRAL	ARDTRLTIIGFSSALS	86

```

gi | 25989732 | gb | AAN75036.1 | _1_2- | DPENHVAINVALYNRGG | ---RWAMTEGREALVERDDSTLAIGPSALR | 178
gi | 860504149 | ref | WP_048445338. | DPIDHSANVVLYGKPA | ---AWAMTERRRHSHVARDAASLAIGPSLSD | 86
gi | 860474721 | ref | WP_048434413. | DPIDHSANVVLYGRPK | ---AWAMTERRRARSKTATSLAIGPSALD | 86
gi | 501291381 | ref | WP_012333472. | DPLDHSANVVLYGAPA | ---RWAMTERRRSVARDAASLSIGPSSLA | 86
gi | 860453591 | ref | WP_048427078. | DPINHSAINVLYGRPK | ---AWAMTERRRHVNRATSLAIGPSLSD | 73
gi | 860560381 | ref | WP_048464709. | DPINHSANVVLYGRPK | ---AWAMTERRRHVSRTATSLAIGPSLSD | 73
gi | 860527577 | ref | WP_048454269. | DPINHSAINVLYGRPK | ---AWAMTERRRHVSRTATSLAIGPSLSD | 86
gi | 760871815 | db | BAQ48812.1 | _1_1 | DPINHSANVVLYGRPK | ---AWAMTERRRHVDRTATSLAIGPSLSD | 73
gi | 373567336 | gb | EHP93308.1 | _6- | DPFAHCAMNVLYGERGN | ---RFAMTEARNARSLRDATSFSLGPSLLE | 93
gi | 240009448 | gb | ACS40674.1 | _4- | DPFAHCAMNVLYGERGN | ---RFAMTEARNARSLRDATSFSLGPSLLE | 84
gi | 218523206 | gb | ACK83791.1 | _4- | DPFAHCAMNVLYGERGN | ---RFAMTEARNARSLRDATSFSLGPSLLE | 84
gi | 254269168 | emb | CAX25134.1 | _4 | DPFAHCAMNVLYGERGN | ---RFAMTEARNARSLRDATSFSLGPSLLE | 84
gi | 163663749 | gb | ABY31116.1 | _2- | DPFAHCAMNVLYGERGN | ---RFAMTEARNARSLRDATSFSLGPSLLE | 87
gi | 494837541 | ref | WP_007563641. | DPFAHCAMNVLYGDRT | ---RFAMTEGAAASLSQSRHSIAIGPSAMA | 73
gi | 914802769 | ref | WP_050733395. | DPFAHCAMNVLYGDRT | ---RFAMTEGAAASLSRSHSVAIGRSAMA | 73
gi | 657246165 | ref | WP_029357447. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSADHIAVGPSAMH | 73
gi | 501276758 | ref | WP_012319776. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSADHIAVGPSAMH | 73
gi | 640229117 | ref | WP_024829694. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSADHIAVGPSAMH | 73
gi | 518936109 | ref | WP_020091984. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSADHIAVGPSAMH | 73
gi | 760075276 | ref | WP_043757989. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSGPDHIAIGPSAMH | 73
gi | 759662308 | ref | WP_043379914. | DPFAHCAMNVLYGDRG | ---RFAMTEARRAASLSRSGPDHIAIGPSAMH | 73
gi | 498368472 | ref | WP_010682628. | DPFAHCAMNVLYGDRS | ---RFAMTEARRAASLSRSADHIAVGPSVMH | 73
gi | 497224411 | ref | WP_009538673. | TDPARHCCMNVLYGASAN | ---RWAMTEGAAASLARDAATTIRIGPSGLR | 86
gi | 589291345 | gb | AHK79488.1 | _1_2 | ---PNHHCALNVALYGDVGR | ---WAMTEGRGLARDSHNLIIGPSAMS | 131
gi | 644125233 | ref | WP_025281955. | ---PNHHCALNVALYGDVGR | ---WAMTEGRGLARDSHNLIIGPSAMS | 76
gi | 496218797 | ref | WP_008932834. | ---PDHHVAINVALYGNAPR | ---RWAMTEGRGLQDRDNLVIGPSALE | 90
gi | 495659402 | ref | WP_008383981. | VDPDFCSLNVALYGDGR | ---RWAMTEGRGALSRDASSLAIGPSALA | 80
gi | 687960265 | gb | KGE05197.1 | _1_16 | ---DPEQYCGFNAVLYNPKR | ---HWAMTEGRGDLIRDSEHLRARGSAIH | 91
gi | 496312569 | ref | WP_009021747. | QNAENHCAFNTILYGPSGRK | ---RWSMTEFSGARLEHSDSYQLGPSRID | 119
gi | 495571757 | ref | WP_008296336. | TAAENHCAFNTILYGPSGKK | ---RWSMTEFSGARLKRADSYALGPSRID | 95
gi | 908380581 | ref | WP_049751705. | PNPEDHCAFVALYGEK - RG | ---HWSMTEDRHALQDRDRLIVGPSQAR | 77
gi | 658524263 | ref | WP_029718812. | ADPENHCALNVALYGDAG | ---KRWMTTEGRRWMRRSRDEFVIGPSRLH | 71
gi | 332109922 | gb | EGJ10550.1 | _1- | --- | ---MTBRGRRWMRRSRDEFVIGPSRLH | 24
gi | 760103377 | ref | WP_043785587. | ADPENHCALNIALYGDAG | ---KRWMTTEGRRWMRRSRDEFVIGPSRLH | 71
gi | 760129436 | ref | WP_043810804. | ADPENHCALNIALYGDAG | ---KRWMTTEGRRWMRRSRDEFVIGPSRLH | 71
gi | 7416798 | db | BAA94047.1 | _1_1-4 | ADPENHCALNIALYGDAG | ---KRWMTTEGRRWMRRSRDEFVIGPSRLH | 194
gi | 75345570 | sp | P95619.1 | CRTCR | ADPENHCALNIALYGPQG | RGAKRWTMTEFSRASITREARRFVIGPSRLE | 194
gi | 913634290 | gb | KNZ32951.1 | _1- | ADPENHCALNIALYTPNS | ---KYWTMTEAQNAINFTEHTFTICPSHLH | 76
gi | 809277751 | ref | WP_046331236. | ANANDYCAINVALYKGG | ---GRWAMTEGQRHCQSESQFTIGPSQLS | 77
gi | 518258896 | ref | WP_019429104. | ANPDNYCALNVALYKGG | ---GRWAMTEGQRHCQSESQFTIGPSQLS | 77
gi | 518255752 | ref | WP_019425960. | ANPDNHCALNVALYKGG | ---GRWAMTEGQRHCQSESQFTIGPSQLS | 77
gi | 918152310 | gb | KOF08324.1 | _1- | ANPDHHCALNVALYKGG | ---GRWAMTEGQRHCQSESQFTIGPSQLS | 77
gi | 616553243 | gb | AAX48196.1 | _1-2 | MNPDNHCALNVALYKGG | ---GRWAMTEGQRHCQSESQFTIGPSQLS | 77
*
* : * : *
WNADGTLTIAIDBIGTP | ---LPQVRGIRVITPLFVNERAFTLDSNG | 235

```







gi   365289544   emb   CCD93979.1   _1	RHRWQFIAPCCRYSDVDFHPDLHWRGDGYFDLNHGDAPLERDFQSQWQR	276
gi   41018902   gb   AAR98493.1   _10-	RHRWQFIAPCCRYSDVDFHPDLHWRGDGYFDLNHGDAPLERDFQSQWQR	276
gi   653027709   ref   WP_027279528.	NHRWMP IAPCSRYEVELEKPKLRWSGEGYFDMNAGDVP IEHGFTVMEWSR	187
gi   499791021   ref   WP_011471755.	NHRWMP IAP SARVEVEMAAPDLAWQEGYFDMNAGDVPLEQGFVMEWSR	187
gi   499823412   ref   WP_011504146.	NHRWMP IAPRSRIEVLDPKLNWQGEAYIDT NAGDAP IETGFSHWEWAR	187
gi   499762149   ref   WP_011442883.	NHRWMP IAPRSRIEVLDPKLNWQGEAYIDT NAGDAP IETGFSHWEWAR	187
gi   829076993   ref   WP_047310044.	NHRWMP IAPRARVEVDLRRPNLSWQEGYLDNAGDEPIETGFRHWEWAR	187
gi   503269006   ref   WP_013503667.	NHRWMP IAPRARVEVDLRRPNLSWQEGYLDNAGDEPIETGFRHWEWAR	187
gi   499470438   ref   WP_011157078.	NHRWMP IAPRARVEVDLRRPNLSWQEGYLDNAGDEPIETGFRHWEWAR	187
gi   501486651   ref   WP_012495141.	NHRWMP IAPRARVEVDLRRPNLSWQEGYLDNAGDEPIETGFRHWEWAR	187
gi   550975051   ref   WP_022723265.	NHRWMP IAPRSRIEVDLRRPNLSWQEGYLDNAGDAP IETGFKHWEWAR	187
gi   764538658   ref   WP_044412067.	NHRWMP IAPCARVEVEMAQPSLRWQEGYLDNAGDAP IETGFRDWEWSR	187
gi   754113006   ref   WP_041748751.	NHQWMP IAPCARVQVRLDHPRLSWQGDGYFDMNHGDTPLQGFSDWQWQR	198
gi   737680904   ref   WP_035649991.	RHQWMP IAPFARVQVRLDHPRLSWQGDGYFDMNHGDTPLQGFSDWQWQR	192
gi   503405597   ref   WP_013640258.	RHRWQFIAPCARIRVAFTAPELAWEGNAYLDRNDGDEPIAHAFRRHWSR	190
gi   657196689   ref   WP_029312388.	AHRWQFIAPCARISVTFEKPALSWEGDAYFDRNEGDAPMHTAFSTWNSR	190
gi   703385326   ref   WP_033417843.	HHVWTP IAP SARVEVELENSPKLSWSGEAYLDGNAGDTP IEDAFRSHWWSR	180
gi   653032131   ref   WP_027283892.	RHLWHP IAAARVEVALENPALRWSGPA YFDWNAGAAPLEQDFREWDWCR	186
gi   378403832   emb   CCG08948.1   _1	RHRWFVWMPRARIEVDLDAPAARWSGSA YFDTNDGDEPMQDAFQRWDWSR	263
gi   754176764   ref   WP_041795395.	RHRWFVWMPRARIEVDLDAPAARWSGSA YFDTNDGDEPMQDAFQRWDWSR	179
gi   25989732   gb   AAN75036.1   _12-	RHVWMPFAPRSRYELRMDNPDIRWNGHG YFDFNCGSEPLQDAFSDWWSR	271
gi   860504149   ref   WP_048445338.	RHRWMP LSPAARLEVAFDRLSHWQGHAYLD TNEGDPLEHSHFRSWTWSR	179
gi   860474721   ref   WP_048434413.	RHRWMP LAPSSRIEVALDRPSLRWNHGHAYFD TNDGVALEDSFRSWTWSR	179
gi   501291381   ref   WP_012333472.	RHRWMP IAPASRIEVDLDRPALRWRGHAYFD TNDGDTPLEASFRSWTWSR	179
gi   860453591   ref   WP_048427078.	RHRWMP LAPSSRYEVAFDPSLRWSGHAYFD TNEGDPLESDSFRSWTWSR	166
gi   860560381   ref   WP_048464709.	HRWMP LAPTSRYEVAFDPSLRWSGHAYFD TNEGDPLEASFKSWTWSR	166
gi   860527577   ref   WP_048454269.	HRWMP LAPTSRYEVAFDPSLRWSGHAYFD TNEGDPLEASFKSWTWSR	179
gi   760871815   dbj   BAQ48812.1   _1	RHRWMP LAPSSRYEVAFDPSLRWSGHAYFD TNDGDPLESDSFRSWTWSR	166
gi   373567336   gb   EHP93308.1   _6-	LHRWMP MAPDCEVEAAFTAPALSWRSGSYHD TNDGGALEDAFDTWTWCR	186
gi   240009448   gb   ACS40674.1   _4-	LHRWMP MAPDCEVEAAFTAPALSWRSGSYHD TNDGGALEDAFDTWTWCR	177
gi   218523206   gb   ACK83791.1   _4-	LHRWMP MAPDCEVEAAFTAPALSWRSGSYHD TNDGGALEDAFDTWTWCR	177
gi   254269168   emb   CAX25134.1   _4	LHRWMP MAPDCEVEAAFTAPALSWRSGSYHD TNDGGALEDAFDTWTWCR	177
gi   163663749   gb   ABY31116.1   _2-	LHRWMP MAPDCEVEAAFTAPALSWRSGSYHD TNDGGALEDAFDTWTWCR	180
gi   494837541   ref   WP_007563641.	RHRWMP MAPSSRYEVELDQPNLRWTGSGYFD TNDGDEPLASAFIRWTWCR	166
gi   914802769   ref   WP_050733395.	RHRWMP MAPSSRYEVELDQPNLRWTGSGYFD TNDGDEPLASAFIRWTWCR	166
gi   657246165   ref   WP_029357447.	RHRWMP MSPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   501276758   ref   WP_012319776.	RHRWMP MSPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   640229117   ref   WP_024829694.	RHRWMP MSPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   518936109   ref   WP_020091984.	RHRWMP MSPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   760075276   ref   WP_043757989.	RHRWMP MAPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   759662308   ref   WP_043379914.	RHRWMP MAPSSRYEVDLDPDLRWSGSGYFD TNDGDEPLEAAFAFWTWSR	166
gi   498368472   ref   WP_010682628.	RHRWMP MAPSSRYEVAFDQGLSWSGSGYFD TNDGDEPLGAAFAFWTWSR	166
gi   497224411   ref   WP_009538673.	AHRWMP IAPSSRYEVLLEKPKALSWKGRAYFD TNDGSEPLEKGFKDWDCR	179
gi   589291345   gb   AHK79488.1   _12	LHQWRP IAP SARVEVDLEQGLQWEGRYLDSNRGEE PVEKGFYWDWSR	224
gi   644125233   ref   WP_025281955.	LHQWRP IAP SARVEVDLEQGLQWEGRYLDSNRGEE PVEKGFYWDWSR	169
gi   4962218797   ref   WP_008932834.	RHHWRP IAPCARVEVDLQQPGLRWLGHGYMDSN IGTGLEQGFYWDWSR	183



```

gi | 495659402 | ref | WP_008383981.1 | 16
gi | 687960265 | gb | KGE05197.1 | 16
gi | 496312569 | ref | WP_009021747.1
gi | 495571757 | ref | WP_008296336.1
gi | 908380581 | ref | WP_049751705.1
gi | 658524263 | ref | WP_029718812.1
gi | 332109922 | gb | EGJ10550.1 | 1-
gi | 760103377 | ref | WP_043785587.1
gi | 760129436 | ref | WP_043810804.1
gi | 7416798 | dbj | BAA94047.1 | 1-4
gi | 75345570 | sp | P95619.1 | CRTC_R
gi | 913634290 | gb | KNZ32951.1 | 1-
gi | 809277751 | ref | WP_046331236.1
gi | 518258896 | ref | WP_019429104.1
gi | 518255752 | ref | WP_019425960.1
gi | 918152310 | gb | KOF08324.1 | 1-
gi | 61653243 | gb | AAX48196.1 | 1-2

gi | 350284778 | gb | AEQ27763.1 | 1-1-
gi | 521992738 | ref | WP_020504009.1
gi | 557030915 | gb | ESQ08988.1 | 1-1-
gi | 494353717 | ref | WP_007191968.1
gi | 570733543 | gb | AHF04706.1 | 24
gi | 760081160 | ref | WP_043763735.1
gi | 528190751 | gb | EPY01080.1 | 43
gi | 748594904 | ref | WP_039853084.1
gi | 488815067 | ref | WP_002727473.1
gi | 505094729 | ref | WP_015281831.1
gi | 754165669 | ref | WP_041786299.1
gi | 739596279 | ref | WP_037453842.1
gi | 764363842 | ref | WP_044336870.1
gi | 644239143 | ref | WP_025293539.1
gi | 503035479 | ref | WP_013270455.1
gi | 617472745 | gb | KAI93218.1 | 2-
gi | 737603811 | ref | WP_0355774467.1
gi | 648530966 | ref | WP_026222717.1
gi | 288897641 | gb | ADC63477.1 | 28
gi | 760114352 | ref | WP_043796307.1
gi | 494098690 | ref | WP_007039487.1
gi | 75388571 | sp | Q7X3G5.1 | CRTC_T
gi | 390426908 | gb | AFL73973.1 | 35
gi | 752807498 | ref | WP_041446922.1
gi | 760074757 | ref | WP_043757486.1
gi | 918211006 | ref | WP_0522348303.1
gi | 499982040 | ref | WP_011662758.1

LHRWPLAPCSDEYVDLARPGLSWRCHGYLDSNI GDEPVE SAFVRWDSR 173
RHRWVPVAPSCRARVTMEKPGLSWEGHYFDN RGAEPLEAAFRWDSR 184
RHHWNP IAPIARIEVD IPSLDTQWSGDGYLDSNHGSEPLARFGSGDWTR 212
RHHWNP IAPIARVDVLP SLGTAMSGDGYLDSNRGTPLADGFGSGDWTR 188
RHRWCALAPCARIEVEMEKPRVWQEGYLD SNADEPLEDGFSGHWSR 170
RHRWGPLAPCSRIEVELDSPRVRSCHAYLDS NEGDEPIDRPFHEWDSR 164
RHRWGPLAPCSRIEVELDSPRVRSCHAYLDS NEGDEPIDRPFHEWDSR 117
RHRWGPLAPCSRIEVELDSPRNARWSGHAYV DSNEGDEPIDRPFREWDSR 164
RHRWGPLAPCSRIEVELDSPRNARWSGHAYV DSNEGDEPIDRPFREWDSR 287
RHRWGPLAPCSRIEVELDSPRVRSCHAYLDS NEGDEPIDRPFREWDSR 287
RHHWGPLAP SARIEVAFDSPTRWSGHAYLDS NEGTPIERPFTWDSR 169
KHRWGPLAP SARIEVFTKPNLHWKGSAYFDS NEGDEAVSQFSEWDSR 170
RHRWGPLAPASRIEVDINAPDLKQGHAYLDS NEGDEPDRGFHNWDSR 170
RHRWGPLAPSSRIEVLNAPDLKQGHAYLDS NEGDEPVEQGFHTWDSR 170
RHRWGPLAPASRIEVALDAPDLKQGHAYLDS NEGDEPVERGFHTWDSR 170
RHRWGPLAPASRIEVELSSPDLKQGHAYLDS NEGDEPIDQGFHTWDSR 170
* * * * *
ASMG---EESAILYHVD R-----SDGP--LSLALRFAADGTLSEFFEP 324
ASMG---EESAILYHVD R-----GDGP--LSLALRFAADGTLSEFFESP 222
AGMDG--AESAILYNI DR-----DGT P--LSLALHFANDGSLSEFFEP 213
ASLGG--DESLILYHAD RR-----EGGR--LSLGLHFAGDGLSHFDAP 213
ATLGE--DESAVLYNMD RR-----DGT P--FSLALNFAADGTLSEFFDP 235
ATLGE--DESAVLYNMD RR-----DGT P--FSLALNFAADGTLSEFFDP 235
APLQ---NGTAVLYDVT RR-----DGS D--LAVALRIGRDGKVEESDSP 269
APLQ---NGTAVLYDVT RR-----DGS D--LAVALRIGRDGKVEESDSP 222
APLK---DGTAILYDVT RR-----DGS D--LAVALRIGHDGKVEESDSP 307
TGLD---DGAAILYDVT RR-----DAER--MSLALR IAGNGEVEAFEP 209
APVGFQGRDGAAILYDVT RR-----DGS G--KVIAIRIDRAGQVESFAVP 230
ADFG---GDSAVLYNVD RR-----VGS H--LALALMIGQDGTVEFFEPL 209
AHLK---NDVAVLYEGRR R-----DGT P--FDLALKFDRQGRWHDVVQP 225
AHLK---NDVAVLYEGRR R-----DGT P--FDLALRFRQGRWHDVVQP 225
AHLK---SDSAVLYDLKR K-----DGS A--LSLALKFDRHGVPETVEPP 212
ACLS---DSAVMLYEGTR Q-----NGSK--FNMGIQIAADGSGVSPVELR 297
APVG---AGAAVLYDAIR R-----DGS A--VSLALRIRADGRVTPAEAP 208
ANIG---DDAAVLYDAER R-----REAP--LSLALRFSASGEMETLDP 206
AELKHG---AAILYDTRAR-----DDQG--AHLALRFDANA EVD EFALP 297
AELKHG---AAILYDTRAR-----DDQG--AHLALRFDANA EVD EFALP 209
AELKQG---AAILYDVT R-----EGRFGAELALRCDHRGQLDRFDPP 211
ANTPSG---TTWMLYDVTAR-----HGTG--ASLALRFNASGEVEFFPPP 298
AELKDGG---TAVLYDAARR-----DGH R--TELGLHFD PAGDVERIDVP 259
AELKDGG---TAVLYDAARR-----DGH R--TELGLHFD PAGDVERIDVP 220
AMLSDGG---AAILYDVTQR-----EGHR--PPLGLRFAANGDVEHFEAP 228
AMLSDGG---AAILYDVTQR-----EGHR--PPLGLRFAANGDVEHFEAP 220
TALKHG---AAVLYDVTGR-----EGPG--PVLGLRFRDPNGDVEHFEAP 228

```



gi   493840365	ref   WP_006787509.	AEIKDG	---	TAILYDITRR	---	NGEA	---	FALALKYHPD	SGLTQLPPP	212	
gi   653037601	ref   WP_027289289.	APLPSGG	---	TAITYDTFRR	---	DGSD	---	HALSLLVVRK	DGGIERFDNA	236	
gi   519009821	ref   WP_020165696.	SNLRGGD	---	TAVLYDVQRR	---	DGTP	---	HTLALHFAD	DGGCRSFAAP	232	
gi   518759539	ref   WP_019916969.	SNLRGGD	---	TAVLYDVERR	---	DGTP	---	HSLALHFAD	DGGCRSFAAP	228	
gi   495340016	ref   WP_008064750.	SSLRGGD	---	TAVLYDVQRR	---	DGTP	---	HSLALHFAD	DGGCRHFAAP	228	
gi   517223140	ref   WP_018411958.	CTLRGGD	---	TAVLYDVSGR	---	DGAR	---	TALALHFAD	GQRRFDAP	230	
gi   889919304	ref   WP_048861895.	APLRDG	---	AAILYDVRRR	---	DGSV	---	QNLSLRVGR	DGRVADYAAP	226	
gi   496441596	ref   WP_009150441.	AKRRQGG	---	AVVLYEANRR	---	DGSA	---	LSLALQVNR	DGSVEPFEPP	302	
gi   146409846	gb ABQ38352.1  _10	AAATRG	---	TVISYDITAR	---	TGAG	---	KSIALHIDP	DARLDRLAPL	315	
gi   908659074	ref   WP_049795525.	AAATRG	---	TVISYDITAR	---	TGAG	---	KSIALHIDP	DARLDRLAPL	222	
gi   365289544	emb   CCD93979.1  _1	AAATRG	---	TVISYDITVAR	---	AGTD	---	KSIALHVDP	DARLRRIAPL	315	
gi   41018902	gb AAR98493.1  _10	AAATSG	---	AVISYDAIAR	---	DHAI	---	SPLALHYDHT	GRLEPIAPL	315	
gi   653027709	ref   WP_027279528.	AALRNG	---	TAILYEAERR	---	DGSR	---	LDLALQFDK	SGAMQMFDFP	226	
gi   499791021	ref   WP_011471755.	AALRDG	---	TAILYEAARR	---	DGTR	---	FDLAIKFDG	NGEMEHFEAP	226	
gi   499823412	ref   WP_011504146.	GALRDK	---	TAILYEAERR	---	DGSR	---	LDLAVTFDA	QGNMERFAQP	226	
gi   499762149	ref   WP_011442883.	GALRDK	---	TAILYEAERR	---	DGSR	---	VDLAITFDA	QGNMERFTPP	226	
gi   829076993	ref   WP_047310044.	GAMRDK	---	TVILYEAERR	---	DGSR	---	VDLAVTFDADG	NKQEFTHPP	226	
gi   503269006	ref   WP_013503667.	GAMRDK	---	TVILYEAERR	---	DGSR	---	VDLAVTFDADG	NKQEFTHPP	226	
gi   499470438	ref   WP_011157078.	GAMRDK	---	TVILYEAERR	---	DGSR	---	VDLAVTFDAE	GNKQEFTHPP	226	
gi   501486651	ref   WP_012495141.	GAMRDK	---	TVILYEAERR	---	DGSR	---	VDLAVTFDAE	GNKQEFTHPP	226	
gi   550975051	ref   WP_022723265.	GALRDK	---	TAILYEAERR	---	DGSR	---	IDLAMTFDA	QGNMOPFTPP	226	
gi   764538658	ref   WP_044412067.	GALRDG	---	TAILYEAERR	---	DGSR	---	FDLAVTFDAK	GNMOPFERP	226	
gi   754113006	ref   WP_041748751.	GAMQDG	---	TAILYEAQRR	---	DGSR	---	IDLAMTFDP	NGRMQTFEPP	237	
gi   737680904	ref   WP_035649991.	GAMQDG	---	TAILYEAERR	---	DGSR	---	VDLAMTFDR	NGRMQAFEP	231	
gi   503405597	ref   WP_013640258.	AAATQG	---	AAILYHAERR	---	DGGQ	---	TSLALGFA	AGGGATAFTPP	229	
gi   657196689	ref   WP_029312388.	APSRQG	---	ARVYHANRR	---	DGGT	---	TALS LGFTA	AGGAAPLHAP	229	
gi   703395326	ref   WP_033417843.	NPLRDG	---	SAVLYDVAR	---	DGSA	---	INLALRF	DGHGEVSRFEPP	219	
gi   653032131	ref   WP_027283892.	APMADR	---	TAILYHAERR	---	DGSS	---	QSLALRF	DEDEGAVGAFPP	225	
gi   378403832	emb   CCG08948.1  _1	GDLADGG	---	CALLYDLTPRH	---	RGPA	---	RLLSLR	LTPGQWEEVAPP	305	
gi   754176764	ref   WP_041795395.	GDLADGG	---	CALLYDLTPRH	---	RGPA	---	RLLSLR	LTPGQWEEVAPP	221	
gi   25989732	gb AAN75036.1  _12	APLADGG	---	AAILYEVTPR	---	QGAD	---	RLLSLR	VDRD	GALSASFSP	311
gi   860504149	ref   WP_048445338.	STLRHG	---	AAILYDVLRR	---	DGSR	---	QDLSLR	FDPDGT	PRPIAPP	218
gi   860474721	ref   WP_048434413.	STLRHG	---	AAILYDVLRR	---	DGSR	---	QDLSLR	FDPDGT	PRPIAPP	218
gi   501291381	ref   WP_012333472.	ATLRHG	---	AAILYDVLRR	---	DGTR	---	QDLSLR	FDPDGT	PRPIAPP	218
gi   860453591	ref   WP_048427078.	ATLRHG	---	AAILYDVLRR	---	DGSR	---	QDLSLR	FDPDGT	PRPIAPP	205
gi   860560381	ref   WP_048464709.	ATLRHG	---	AAILYDVLRR	---	DGSR	---	QDLSLR	FDPDGT	PRPIAPP	205
gi   860527577	ref   WP_048454269.	ATLRHG	---	AAILYDVLRR	---	DGSR	---	QDLSLR	FDPDGT	PRPIAPP	218
gi   760871815	dbj BAQ48812.1  _1	ATLRHG	---	AAILYDVAR	---	DGSR	---	QELSLRF	FDPDGT	PRPIAPP	205
gi   373567336	gb EHP93308.1  _6	APLKRK	---	SAILYDVLRR	---	DGGG	---	QNLTLRF	FADGT	RREMRPP	225
gi   240009448	gb ACS40674.1  _4	APLKRK	---	SAILYDVLRR	---	DGGG	---	QNLTLRF	FADGT	RREMRPP	216
gi   218523206	gb ACK83791.1  _4	APLKRK	---	SAILYDVLRR	---	DGGG	---	QNLTLRF	FADGT	RREMRPP	216
gi   254269168	emb   CAX25134.1  _4	APLKRK	---	SAILYDVLRR	---	DGGG	---	QNLTLRF	FADGT	RREMRPP	216
gi   163663749	gb ABY31116.1  _2	APLKRK	---	SAILYDVLRR	---	DGGG	---	QNLTLRF	FADGT	RREMRPP	219
gi   494837541	ref   WP_007563641.	ADLPDG	---	AAILYDVRHR	---	DGGG	---	QNLSLRF	CHDGNRREIRPP	205	
gi   914802769	ref   WP_050733395.	ADLPDG	---	AAILYDVRHR	---	NGGG	---	QNLSLRF	SHDGSRRREIRPP	205	
gi   657246165	ref   WP_029357447.	ADLPDG	---	AAILYDVVHR	---	DGGG	---	QNLSLRF	SADGSRRREIRPP	205	

gi   501276758   ref   WP_012319776.	ADLPDG	---	AAIYDVVHR	---	DGGG	---	QNL	SLRFS	ADG	SRR	IRPP	205																																
gi   640229117   ref   WP_024829694.	ADLPDG	---	AAIYDVVHR	---	DGGG	---	QNL	SLRFS	ADG	SRR	IRPP	205																																
gi   518936109   ref   WP_020091984.	ADLPDG	---	AAIYDVVHR	---	DGGG	---	QNL	SLRFS	ADG	SRR	IRPP	205																																
gi   760075276   ref   WP_043757989.	ADLPDG	---	AAIYDVVHR	---	DGGG	---	QNL	SLRFS	ADG	SRR	IRPP	205																																
gi   759662308   ref   WP_043379914.	ADLPDG	---	AAIYDVVHR	---	DGGG	---	QNL	SLRFS	ADG	SRR	IRPP	205																																
gi   498368472   ref   WP_010682628.	ADLQDG	---	AAIYDVQHR	---	EGGG	---	QNL	SLRFS	SDG	SRR	IRPP	205																																
gi   497224411   ref   WP_009538673.	APLDGG	---	AAIYETRPL	---	DGPS	---	RCL	SLRIG	RDG	SSR	IEIPP	219																																
gi   589291345   gb   AHK79488.1  _1-12	AVMEDQE	---	TVVLYEADR	---	DGGH	---	SLLS	LRFD	ADG	KARE	VEAP	264																																
gi   644125233   ref   WP_025281955.	AVMEDQE	---	TVVLYEADR	---	DGGH	---	SLLS	LRFD	ADG	KARE	VEAP	209																																
gi   496218797   ref   WP_008932834.	AVMPNGD	---	TVVLYEATQR	---	DDQE	---	RML	TLRFD	PRG	VRD	VEPP	223																																
gi   495659402   ref   WP_008383981.	AALPDGS	---	AAVLYHAI	RRGAD	GGEAE	---	TGL	ALRFD	RS	GA	VF	FDAP	218																															
gi   687960265   gb   KGE05197.1  _1-16	APLADGG	---	SLVQYHARL	---	DGSE	---	RL	LALR	F	GAT	GA	VELDPL	224																															
gi   496312569   ref   WP_009021747.	ARLNNGD	---	CAVRYEARYP	---	GLEP	---	HL	ALR	FA	Q	DS	IAE	PSQ	252																														
gi   495571757   ref   WP_008296336.	ARLSNGD	---	CAVRYEARHG	---	ARKP	---	HL	ALR	FK	D	GS	IA	SE	TR	228																													
gi   908380581   ref   WP_049751705.	ASDTPG	---	TTVLYDVTPR	---	QGAP	---	YR	KAI	H	F	H	Q	T	V	ET	PAP	209																											
gi   658524263   ref   WP_029718812.	ATMADGS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	204																					
gi   332109922   gb   EGJ10550.1  _1-1	ATMADGS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	157																					
gi   760103377   ref   WP_043785587.	ATMADGS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	204																					
gi   760129436   ref   WP_043810804.	ATMADGS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	204																					
gi   7416798   gb   BAA94047.1  _1-4	ATMADGS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	327																					
gi   75345570   sp   P95619.1  CRTC_R	ATMADSS	---	TAVIYDVRQK	---	RDGD	---	R	VIA	E	R	F	L	L	D	G	S	T	S	F	E	A	P	327																					
gi   913634290   gb   KNZ32951.1  _1-1	CSLRDGS	---	TAVIYDVRQK	---	QGAD	---	R	V	L	A	L	R	F	T	Q	D	G	E	V	E	P	F	E	P	209																			
gi   809277751   ref   WP_046331236.	AHLKDG	---	TAVIYDVRQS	---	NGAE	---	R	I	A	S	K	E	N	V	D	G	T	V	E	P	F	A	A	P	210																			
gi   518258896   ref   WP_019429104.	ARTRDGS	---	TVVLYDMQAP	---	ATED	---	R	V	L	S	L	R	F	T	P	R	G	T	V	E	P	I	A	T	P	210																		
gi   518255752   ref   WP_019425960.	ARMRDGS	---	TVVLYDMQAP	---	GTED	---	R	V	L	S	L	R	F	S	P	K	G	T	V	E	P	M	A	T	P	210																		
gi   918152310   gb   KOF08324.1  _1-1	ARMRDGS	---	TVVLYDMQAP	---	GAED	---	R	V	L	S	L	L	F	T	P	Q	G	A	V	E	P	I	A	A	P	210																		
gi   61653243   gb   AAX48196.1  _1-2	ARMRDGS	---	TLVLYDMQAP	---	HTQD	---	R	V	L	S	L	R	F	T	P	K	G	V	E	I	D	T	P			210																		
gi   350284778   gb   AEQ27763.1  _1-1	PNHRLP	R	T	P	I	W	R	S	S	Q	A	D	A	D	H	---	P	P	K	I	V	K	T	L	E	D	T	P	F	Y	S	R	S	V	I	A	S	H	L	370				
gi   521992738   ref   WP_020504009.	PNHRLP	R	T	P	I	W	R	S	S	Q	A	D	A	D	H	---	P	P	R	I	V	K	T	L	E	D	T	P	F	Y	S	R	S	V	I	A	S	H	L	268				
gi   557030915   gb   ESQ08988.1  _1-1	PNHRLP	R	T	P	I	W	R	S	S	Q	A	D	A	D	H	---	P	P	R	V	V	K	T	L	E	D	T	P	F	Y	S	R	S	V	I	D	S	H	L	259				
gi   494353717   ref   WP_007191968.	PDHRLA	R	T	P	I	W	R	M	P	R	S	S	Q	A	D	A	K	H	---	P	P	R	V	V	K	T	L	E	D	T	P	F	Y	S	R	S	V	I	A	S	H	L	259	
gi   570733543   gb   AHF04706.1  _24	PNHKL	P	R	G	T	V	M	R	S	P	R	S	M	Q	A	D	E	---	P	P	R	I	V	E	T	L	E	D	T	P	F	Y	S	R	S	V	S	A	K	L	279			
gi   760081160   ref   WP_043763735.	PNHKL	P	R	G	T	V	M	R	S	P	R	S	M	Q	A	D	E	---	P	P	R	I	V	E	T	L	E	D	T	P	F	Y	S	R	S	V	S	A	K	L	279			
gi   528190751   gb   EPY01080.1  _43	PMSAL	P	R	G	-	A	W	G	V	E	R	V	S	R	A	D	S	G	H	---	T	A	R	L	A	E	V	L	E	D	T	P	F	Y	T	R	S	V	I	D	T	R	L	314
gi   748594904   ref   WP_039853084.	PMSAL	P	R	G	-	A	W	G	V	E	R	V	S	R	A	D	S	G	H	---	T	A	R	L	A	E	V	L	E	D	T	P	F	Y	T	R	S	V	I	D	T	R	L	267
gi   488815067   ref   WP_002727473.	PVADL	P	R	G	-	V	M	G	D	R	A	T	R	A	D	S	G	H	---	A	A	Q	L	A	E	I	L	E	D	T	P	F	Y	T	R	S	M	V	E	T	R	L	352	
gi   505094729   ref   WP_015281831.	PPAPL	A	T	S	V	M	R	I	A	R	A	T	Q	A	D	T	Q	---	P	A	R	V	E	T	L	E	D	T	P	F	Y	A	R	S	V	I	A	T	H	L	255			
gi   754165669   ref   WP_041786299.	PRVAL	P	G	S	F	W	R	I	N	G	V	R	C	E	S	G	---	T	A	A	V	R	E	L	V	D	A	P	F	A	R	S	V	A	T	R	M			276				
gi   739596279   ref   WP_037453842.	PRVTL	P	A	S	R	W	R	I	P	R	T	R	A	D	P	Q	---	P	V	T	V	E	T	L	V	D	A	P	F	A	R	S	V	L	A	T	H			255				
gi   764363842   ref   WP_044336870.	TAARL	P	R	T	-	G	L	V	D	R	A	T	R	V	D	A	G	H	---	R	P	R	V	T	K	T	W	I	D	A	P	F	A	R	S	A	L	S	T	H	L	270		
gi   644239143   ref   WP_025293539.	APARL	P	R	T	-	G	L	V	M	R	Q	T	R	A	D	S	G	A	---	A	P	R	I	L	K	T	W	I	D	A	P	F	A	R	S	S	I	A	T	R	L	270		
gi   503035479   ref   WP_013270455.	PHAA	L	P	A	T	-	L	W	E	M	P	R	H	T	R	A	D	A	R	---	P	V	S	V	E	R	T	W	E	D	T	P	F	Y	S	R	S	T	I	R	T	H	L	257
gi   617472745   gb   KAI93218.1  _2-	ERCD	L	P	L	T	R	Y	W	K	M	P	R	E	T	R	V	D	A	G	H	---	T	A	N	V	E	Q	L	E	D	T	P	F	Y	S	R	S	M	I	A	T	H	L	343
gi   737603811   ref   WP_035574467.	PEAG	L	P	G	T	-	L	W	R	I	P	R	R	T	R	A	D	A		---	T	P	G	V	L	A	T	L	E	D	A	P	F	A	R	S	T	L	A	T	R	L	253	

gi   648530966	ref   WP_026222717.1	PLAPLPI-T-KWRVQRRT-RADDG-----VAQAIRSRFEDAPFFYRSLSVAAKF	250
gi   288897641	gb   ADC63477.1   _28	PRAPLP-TTSVMRIERGIQCEAGDE---PRVVTILEDTPFYARSVVETRL	343
gi   760114352	ref   WP_043796307.1	PRAPLP-TTSVMRIERGIQCEAGDE---PRVVTILEDTPFYARSVVETRL	255
gi   494098690	ref   WP_007039487.1	PRVPIQTTTRVWICRGQTQSEPGHE---TRVVDTLEDTPFYARSVLESQI	257
gi   75388571	sp   Q7X3G5.1   CRTC_T	PRVPLPTTGIWRIKRGTCQCEAGHQ---ARVVEITLEDTPFYARSLVETRL	344
gi   390426908	gb   AFL73973.1   _35	PHATLP-T-AWLMQRGTRVDPQQT---ARVTEITLNTFFYTRSVLETHL	304
gi   752807498	ref   WP_041446922.1	PHATLP-T-AWLMQRGTRVDPQQT---ARVTEITLNTFFYTRSVLETHL	265
gi   760074757	ref   WP_043757486.1	PHVRLPNT-SWLMRRRGRTRVANKF---ARVAETLEDTPFYTRSVLESHL	273
gi   918211006	ref   WP_052348303.1	PHVRLPNT-SWLMRRRGRTRVANKF---ARVAETLEDTPFYTRSVLESHL	265
gi   499982040	ref   WP_011662758.1	PRIKLPNT-LWRVRETRVDHQSS---ARVATLEDTPFYARSVLETNL	273
gi   493840365	ref   WP_006787509.1	PAHLPSPGPIWIRAGRNQADEGHQ---PRIAKTFEDTPFYARSMVHSQI	258
gi   653037601	ref   WP_027289289.1	VLNDLPSTSIWIRRRKGRADAGR---ACVSKTLEDTPFYARSLVETHL	282
gi   519009821	ref   WP_020165696.1	AAHELPG-TRWGIAARTRSEGGTT---PQAR-TLEDTPFYARSLLDTCI	276
gi   518759539	ref   WP_019916969.1	AAHELPG-TRWGIAARTRSEGGTT---PQAR-TLEDTPFYARSLLDTCI	272
gi   495340016	ref   WP_008064750.1	AERTLPG-TRWGIAARTRSEADTS---PQVR-TLEDTPFYARSLLDTRL	272
gi   517223140	ref   WP_018411958.1	PECTLPG-TRWGIAARTRCERGA---PQVR-TLEDTPFYARSLLDTRL	274
gi   889919304	ref   WP_048861895.1	AACPDPG-TRWGIAARTRAEPTGG---TRLRITLTDAPFYARSLLDTRI	271
gi   496441596	ref   WP_009150441.1	ARVPLPTTWWIARARTRSEHPVGTETRAQVLELEDTPFYARSLIATRL	352
gi   146409846	gb   ABQ38352.1   _10	ADAPLRR-T-LWRIDRSARADSETP---VRVSTLEDAPFYARSHLAARI	360
gi   908659074	ref   WP_049795525.1	ADAPLRR-T-LWRIDRSARADSETP---VRVSTLEDAPFYARSHLAARI	267
gi   365289544	emb   CCD93979.1   _1	AESPLRKT-LWRVDRSARADAFS---VRVSTLEDAPFYARSHLAARI	360
gi   41018902	gb   AAR98493.1   _10	VDCRLPKT-LWRVRSARADAGAP---VRVSTLEDAPFYARSHLAARV	360
gi   653027709	ref   WP_027279528.1	TIHPLPRT-PWRVRSARSQIE---AAVVRTLEDSPFYARSMIATTL	269
gi   499791021	ref   WP_011471755.1	ALERLQRT-PWRVDRSARSEAG---ARVVKTLEDSPFYARSMIATTL	269
gi   499823412	ref   WP_011504146.1	PLMTLPRS-GWKVRSARSEID---AKVLRITLEDTPFYARSMVATTL	269
gi   499762149	ref   WP_011442883.1	PLLPPLRS-GWKVRSARSETD---ARVIRTLEDSPFYARSMVATTL	269
gi   829076993	ref   WP_047310044.1	QLLPPLRS-GWKVDRSARSETG---AKLVRTLEDTPFYARSMVETTL	269
gi   530269006	ref   WP_013503667.1	QLLPPLRS-GWKVDRSARSETG---AKLVRTLEDTPFYARSMVETTL	269
gi   499470438	ref   WP_011157078.1	QLLPPLRS-GWKVDRSARSEAS---VKLVRTLEDTPFYARSMVETTL	269
gi   501486651	ref   WP_012495141.1	QLLPPLRS-GWKVDRSARSEAS---VKLVRTLEDTPFYARSMVETTL	269
gi   550975051	ref   WP_022723265.1	PLLPPLRS-GWKVARSARSETD---ARLLRITLEDTPFYARSMIDTTL	269
gi   764538658	ref   WP_044412067.1	QLVPLARS-GWRVNRNARSNDG---ARVVKTLEDAPFYTRSVVATTL	269
gi   754113006	ref   WP_041748751.1	PTIDLRR-T-GWRVARSARSDSA---AKIVKTLEDAPFYARSVVSASL	280
gi   737680904	ref   WP_035649991.1	PTVKLRR-T-GWRVARSVRSDGA---AKIVKTLEDAPFYARSVLSAKL	274
gi   503405597	ref   WP_013640258.1	PAALPAT-RWRILARETRSDGAS---LIRNLEDTPFYARARIGASL	272
gi   657196689	ref   WP_029312388.1	PEAPLPI-T-RWRINRATRCADAE---PRLIRAFEDTPFYARARIAATI	274
gi   703385326	ref   WP_033417843.1	PAVLPKT-GWRIDRASRGGGAA---A-IVQITLEDTPFYARSLSTTL	263
gi   653032131	ref   WP_027283892.1	PPVSLPR-T-LWRIRRPTRSEDGHA---R-VLRTLEDTPFYARSVIETRL	269
gi   378403832	emb   CCG08948.1   _1	PPAALPAGPSWVGARGIRSTG---PARVETLEDTPFYTRSLVESVF	349
gi   754176764	ref   WP_041795395.1	PPAALPAGPSWVGARGIRSTG---PARVETLEDTPFYTRSLVESVF	265
gi   25989732	gb   AAN75036.1   _12	PKAPLGA-TGWRVARGTRCDVG---APTVLRTLEDTPFYARSLSTRL	356
gi   860504149	ref   WP_048445338.1	VPAPLPR-T-LWRILPRTRSDD---GTAAVLRRFEDAPFYRSLSLSARI	262
gi   860474721	ref   WP_048434413.1	LPAALPRT-LWRILPRATRSDD---GQARVLRFRFEDAPFYRSLSLAARI	262
gi   501291381	ref   WP_012333472.1	IPAAIPPT-LWRILKRATRSDD---GQARVLRFRFEDAPFYRSLSLAARI	262
gi   860453591	ref   WP_048427078.1	VPAALPRT-LWRILPRATRSDD---GQAAVLRFRFEDAPFYRSLSLSARI	249
gi   860560381	ref   WP_048464709.1	VPAALPRT-LWRILPRATRSDD---GRAAVLRRFEDAPFYRSLSLSARI	249

gi   860527577	ref   WP_048454269.	VPAALPRT-LWR	LPRATRSDD	----	GRAALLRRR	FE	DAPFFYRSLLAARI	262		
gi   760871815	dbj   BAQ48812.1   _1	VPAALPRT-LWR	LERATRSDD	----	GQASVLR	RRR	FE	DAPFFYRSLLSARI	249	
gi   373567336	gb   EHP93308.1   _6	LAAGLPSTGL	WRMPRTTRGDD	----	GRAQVIR	T	FE	DPFFYARSLAATL	270	
gi   240009448	gb   ACS40674.1   _4	LAAGLPSTGL	WRMPRTTRGDD	----	GRAQVIR	T	FE	DPFFYARSLAATL	261	
gi   218523206	gb   ACK83791.1   _4	LAAGLPSTGL	WRMPRTTRGDD	----	GRAQVIR	T	FE	DPFFYARSLAATL	261	
gi   254269168	emb   CAX25134.1   _4	LAAGLPPTDL	WRMPRTTRSD	----	GRAQVIR	T	FE	DPFFYARSLAATL	261	
gi   163663749	gb   ABY31116.1   _2	LAATLPPTGL	WRMPRRSRSD	----	GRAQVIR	T	FE	DPFFYARSLAATL	264	
gi   494837541	ref   WP_007563641.	LAAPLPSTSF	WRMPRETRSD	----	GRARVLT	T	FE	DPFFYARSLAATL	250	
gi   914802769	ref   WP_050733395.	LAAPLPSTSF	WRMPRETRSD	----	GRARVLT	T	FE	DPFFYARSLAATL	250	
gi   657246165	ref   WP_029357447.	PGATLPATRF	WRMPRETRSDG	----	GSRVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   501276758	ref   WP_012319776.	PGATLPATRF	WRMPRETRSDG	----	GSRVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   640229117	ref   WP_024829694.	PGATLPATRF	WRMPRETRSDG	----	GRARVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   518936109	ref   WP_020091984.	PGAALPATRF	WRMPRETRSDG	----	GSRVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   760075276	ref   WP_043757989.	PAASLPATGF	WRMPRETRSD	----	GSRVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   759662308	ref   WP_043379914.	PAASLPATGF	WRMPRETRSD	----	GSRVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   498368472	ref   WP_010682628.	TAASLPATSF	WRMGRETRSD	----	GRARVLIAT	Y	ED	DPFFYARSLAATL	250	
gi   497224411	ref   WP_009538673.	PPVMLPKTRL	WRIRRASRQAP	----	EAAARVAG	T	LE	DPFFYARSLSLRL	265	
gi   589291345	gb   AHK79488.1   _12	PAHGLPTARI	WRAPRATRADGS	----	ATIRHTL	E	DPFFYARSLVDARI	309		
gi   644125233	ref   WP_025281955.	PAHGLPTARI	WRAPRATRADGS	----	ATIRHTL	E	DPFFYARSLVDARI	254		
gi   496218797	ref   WP_008932834.	RRQGLPTS	AWRIDRATRADGR	----	AAVRQTL	I	DPFFYVRSLLDTRI	268		
gi   495659402	ref   WP_008383981.	PPVMLP	ASAWRPRETRADPGTA	----	VTIVDTL	E	DPFFYARSLITGTL	263		
gi   687960265	gb   KGE05197.1   _16	PRQQLP	RASIMWARSTQADRDTP	----	VTVQRTL	E	DPFFYRSLLQYRA	270		
gi   496312569	ref   WP_009021747.	PSTALP	STGIWRVARRMPGAE	----	GKPAKLE	T	LE	DPFFYARSLIETTF	297	
gi   495571757	ref   WP_008296336.	DAAAALPRT	GIVRHRMPGAE	----	ARPTRLQ	T	LE	DPFFYARSLIKTEF	273	
gi   908380581	ref   WP_049751705.	DPAPLP	TALWGRRRIPSDTP	----	EKARLIET	F	ED	DPFFYARSLVATEV	254	
gi   658524263	ref   WP_029718812.	PRQLP	TT-LWRIGRTMRTEPGV	----	PAVVEQT	L	ED	DPFFYARSMVRSGL	249	
gi   332109922	gb   EGT10550.1   _1	PRQLP	TT-LWRIGRTMRTEPGV	----	PAVVEQT	L	ED	DPFFYARSMVRSGL	202	
gi   760103377	ref   WP_043785587.	PRQLP	TT-LWRIDRTMRTEPGV	----	PAFVEQT	L	ED	DPFFYARSMVRSGL	249	
gi   760129436	ref   WP_043810804.	PRQLP	TT-LWRIDRTMRTEPGV	----	PALVEQT	L	ED	DPFFYARSMVRSGL	249	
gi   7416798	dbj   BAA94047.1   _1-4	PRQLP	TT-LWRIDRTMRTEPGV	----	PAFVEQT	L	ED	DPFFYARSMVRSGL	372	
gi   75345570	sp   P95619.1   CRIC_R	PRQLP	TT-LWRIGRTMRTEPGV	----	PALVEQT	L	ED	DPFFYARSMVRSGL	372	
gi   913634290	gb   KNZ32951.1   _1	PRQALP	RT-LWRIGRTMRTEPGV	----	PATVTRT	L	ED	DPFFYARSMVRSGL	254	
gi   809277751	ref   WP_046331236.	ERVSURKT	GWGIKRNMRSSEK	TG	----	IVELLNT	Y	ED	DPFFYARSLQSHL	255
gi   518258896	ref   WP_019429104.	PVQLPKT	AWRIDRRMRSTQ	----	AVRVHEQ	L	ED	DPFFYQALLQFDY	253	
gi   518255752	ref   WP_019425960.	PVQNL	SKT-GWRIDRRMRSTQ	----	AVRVHEQ	L	ED	DPFFYQALLQFDH	253	
gi   918152310	gb   KOF08324.1   _1	PVQNL	SKT-GWRIDRRMRSAN	----	AVRVQEQ	L	ED	DPFFYQALLQFNF	253	
gi   61653243	gb   AAX48196.1   _1-2	PVQRLPKT	GWRIDRRMRSEQ	----	PVQVQEQ	L	ED	DPFFYQALLQFDH	253	
* * * * *										
gi   350284778	gb   AEQ27763.1   _1	LDRPVT	AVHESLSDRFRS	RWVQHLL	LPYRMPRR	----	403			
gi   521992738	ref   WP_020504009.	LDRPVT	AVHESLSDRFRS	RWVQHLL	LPYRMPRR	----	301			
gi   557030915	gb   ESQ08988.1   _1	LDRPLQ	AVHESLSDRFRS	PVQHML	LPYRMPRH	----	292			
gi   494353717	ref   WP_007191968.	QGRPIE	AVHESLSDRFRS	AWVQHLL	LPYRMPRR	----	292			
gi   570733543	gb   AHF04706.1   _24	QGRQAE	AVHESLSDRFTS	PWVQPLL	LPYRMPKR	----	312			
gi   760081160	ref   WP_043763735.	QGRQAE	AVHESLSDRFTS	PWVQPLL	LPYRMPKR	----	312			
gi   528190751	gb   EYP01080.1   _43	FGEKY	TGFHESLCLDRFRS	GWVQALL	LPFRMPRR	----	347			

gi   748594904	ref   WP_039853084.	FGEKVTGFHESLCLDRFRSGWVQALLPFRMPRR--	300
gi   488815067	ref   WP_002727473.	FGEEVTGFHESLCLDRFRFTGWVQALLPFRMPRR--	385
gi   505094729	ref   WP_015281831.	LGPVTVGHESLCLDRFRQGWVQMLLPFRMPRLA	289
gi   754165669	ref   WP_041786299.	AGDVTAIHESLCLDRFDSTWVQVLLPFRMPRLA	310
gi   739596279	ref   WP_037453842.	GSRPATAIHESLCLDRFRAGWVQAMLFFRMPR--	287
gi   764363842	ref   WP_044336870.	FGEDAQAVHESLGLGRFRSPVQSMPLPRMPR--	302
gi   644239143	ref   WP_025293539.	FGEDYQAVHESLGLGRFRSPVQSMPLPRMPR--	302
gi   503035479	ref   WP_013270455.	LGEACGVHESLGLRRLRSLPLVRAMLFFRMPR--	289
gi   617472745	gb KAI93218.1 _2-	LGEPVHAMHESVLDLRFKRKRVQSLIGWEMPR--	375
gi   737603811	ref   WP_035574467.	YQDPTAVHESLSLVRFTRWVKLLLPFRMPRRA	287
gi   648530966	ref   WP_026222717.	RGETIHWNESLCLDRFANPLIRLMLPFRMPRRA	284
gi   288897641	gb ADC63477.1 _28	LGRATSVEHESLCLDRFSTQWVKVLLPFRMPR--	375
gi   760114352	ref   WP_043796307.	LQRATSVEHESLCLDRFSTQWVKVLLPFRMPR--	287
gi   494098690	ref   WP_007039487.	LGEHATSVEHESLCLDRFSSWVTRTLLPFRMPRIA	291
gi   75388571	sp Q7X3G5.1 CRTC_T	AGETATCVHESLCLDRFASPVVQMLPFRMPR--	376
gi   390426908	gb AFL73973.1 _35	LGEPYTAMHESVCLDRFDSRWVQVLLPFRIPR--	336
gi   752807498	ref   WP_041446922.	LGEPYTAMHESVCLDRFDSRWVQVLLPFRIPR--	297
gi   760074757	ref   WP_043757486.	LGEPYTAMHESVCLDRFNSRWVQTLPPFRIPR--	305
gi   918211006	ref   WP_052348303.	LGEPYTAMHESVCLDRFNSRWVQTLPPFRIPR--	297
gi   499982040	ref   WP_011662758.	LGETATAMHESLCLDRFNSRWVQTLPPFRIPR--	305
gi   493840365	ref   WP_006787509.	FGPEVTAMHESLCLDRYSKWKVRSMLPFRMPR--	290
gi   653037601	ref   WP_027289289.	AGEPYRAMHESLSLTRFDTRIVRLMLPFRMPR--	314
gi   519009821	ref   WP_020165696.	LGEVHGVEHESLLELRFSGWVQMLLPFRMPRRS	310
gi   518759539	ref   WP_019916969.	LGEVHGVEHESLLELRFSGWVQMLLPFRMPRRS	306
gi   495340016	ref   WP_008064750.	LGEVHGVEHESLLELRFSSGWVQMLLPFRMPRRS	306
gi   517223140	ref   WP_018411958.	LGEVYRGVHESLLELRFSSGWVQMLLPFRMPRRR	308
gi   889919304	ref   WP_048861895.	DGEVYIAMHEALSMRRFVAPWVQAMLFFRMPR--	303
gi   496441596	ref   WP_009150441.	LGEPTAVHESLCLDRFAQPWVYRMLLPFRMPR--	384
gi   146409846	gb ABQ38352.1 _10	GGEDVAVMHESLCLDRFKQPIVQAMLFFRMPR--	392
gi   908659074	ref   WP_049795525.	GGEDVAVMHESLCLDRFKQPIVQAMLFFRMPR--	299
gi   365289544	emb CCD93979.1 _1	QGEDVAVMHESLCLDRFQMPVQAMLFFRMPRRA	394
gi   41018902	gb AAR98493.1 _10-	HGEDVAVMHESLCLDRFQMPVQAMLFFRMPR--	392
gi   653027709	ref   WP_027279528.	LGEQTTLMHESLSLELRFMPVQAMLFFRMPRR--	302
gi   499791021	ref   WP_011471755.	FGEHATLMHESLCLDRFKMPVYQAMLFFRMPRR--	302
gi   499823412	ref   WP_011504146.	LGEPTLVHESLCLDRFRMPVYQAMLFFRMPRR--	302
gi   499762149	ref   WP_011442883.	LGERVTLMHESLCLDRFKMPVYQAMLFFRMPRR--	302
gi   829076993	ref   WP_047310044.	LGEKVTLMHESLCLDRFRMPVYQAMLFFRMPRR--	302
gi   503269006	ref   WP_013503667.	LGEKVTLMHESLCLDRFRMPVYQAMLFFRMPRR--	302
gi   499470438	ref   WP_011157078.	LGEKVTLVHESLCLDRFRMPVYQAMLFFRMPRR--	302
gi   501486651	ref   WP_012495141.	LGEKVTLVHESLCLDRFRMPVYQAMLFFRMPRR--	302
gi   550975051	ref   WP_022723265.	LGEKVTLMHESLCLDRFKLPVYQAMLFFRMPRR--	302
gi   764538658	ref   WP_044412067.	LGEKVTLMHESLCLDRFKLPVYQAMLFFRMPRR--	302
gi   754113006	ref   WP_041748751.	LGEKVTLMHESLCLDRFKLPVYQAMLFFRMPRR--	312
gi   737680904	ref   WP_035649991.	LGPVTLIHESLCLDRFKMPLVQAMLFFRMPR--	306
gi   503405597	ref   WP_013640258.	GGEPVAVHETLDCDRLVHPAVQFMLPFRMPRR--	305
gi   657196689	ref   WP_029312388.	DGERLDGVEHETLDDDRFTMPVYQAMLFFRMPRR--	307

gi   703385326	ref   WP_033417843.	CGERTTG	VHESL	DLDR	FRQP	VVQ	WML	PF	RMP	--	295														
gi   653032131	ref   WP_027283892.	LGERATA	VHESL	DLDR	FAHPS	WY	AML	PF	KV	VP	--	301													
gi   378403832	emb   CCG08948.1   _1	HGETVRS	VHESL	DLDR	LTSR	WVP	WLL	LP	FR	MP	RR	382													
gi   754176764	ref   WP_041795395.	HGETVRS	VHESL	DLDR	LTSR	WVP	WLL	LP	FR	MP	RR	298													
gi   25989732	gb   AAN75036.1   _12-	AGEETRA	VHESL	IDLDR	LKAG	WV	WPL	LP	FR	MP	RR	389													
gi   860504149	ref   WP_048445338.	CGEAYQP	VHEAL	CLDR	FS	HL	VR	LML	PF	FR	MP	--	294												
gi   860474721	ref   WP_048434413.	CGEAYEP	VHEAL	DLDR	FRN	PL	VR	AML	PF	FR	MP	--	294												
gi   501291381	ref   WP_012333472.	CGEAYEP	VHEAL	CLDR	FR	HP	PL	VR	LML	PF	FR	MP	--	294											
gi   860453591	ref   WP_048427078.	CGEAYQP	VHESL	DC	DR	LTH	PL	VR	FML	PF	FR	MP	--	281											
gi   860560381	ref   WP_048464709.	CGEAYQP	VHESL	DLDR	LTH	PL	VR	FML	PF	FR	MP	--	281												
gi   860527577	ref   WP_048454269.	CGEAVRPI	HESL	DLDR	LTH	PL	VR	FML	PF	FR	MP	--	294												
gi   760871815	dbj   BAQ48812.1   _1	CGEAYEP	VHESL	DLDR	LTH	PL	VR	FML	PF	FR	MP	--	281												
gi   373567336	gb   EHP93308.1   _6-	DGEAYQP	VHESL	TR	FR	N	PL	VR	LML	PF	FR	MP	RR	303											
gi   240009448	gb   ACS40674.1   _4-	DGEAYQP	VHESL	TR	FR	N	PL	VR	LML	PF	FR	MP	RR	294											
gi   218523206	gb   ACK83791.1   _4-	DGEAYQP	VHESL	TR	FR	N	PL	VR	LML	PF	FR	MP	RR	294											
gi   254269168	emb   CAX25134.1   _4	DGEAYQP	VHESL	TR	FR	N	PL	VR	LML	PF	FR	MP	RR	294											
gi   163663749	gb   ABY31116.1   _2-	DGEAYQP	VHESL	TR	FR	N	PL	VR	LML	PF	FR	MP	RR	297											
gi   494837541	ref   WP_007563641.	CGEPVRA	VHESL	DLDR	FRN	PL	VQ	LML	PF	FR	MP	--	282												
gi   914802769	ref   WP_050733395.	AGEPVRA	VHESL	DLDR	FRN	PL	VQ	LML	PF	FR	MP	--	282												
gi   657246165	ref   WP_029357447.	CGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   501276758	ref   WP_012319776.	CGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   640229117	ref   WP_024829694.	CGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   518936109	ref   WP_020091984.	GGEPVRA	VHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   760075276	ref   WP_043757989.	GGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   759662308	ref   WP_043379914.	GGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   498368472	ref   WP_010682628.	GGEPVRA	MHESL	DLDR	FRN	P	M	V	R	L	M	L	PF	FR	IP	--	282								
gi   497224411	ref   WP_009538673.	AGEDATA	VHESL	DM	RR	FT	N	P	L	V	Q	A	M	L	PF	FR	MP	RR	288						
gi   589291345	gb   AHK79488.1   _12	HGEKAVAF	HESL	S	L	N	R	F	E	S	P	V	Q	L	M	L	PF	FR	MP	RR	343				
gi   644125233	ref   WP_025281955.	HGEKAVAF	HESL	S	L	N	R	F	E	S	P	V	Q	L	M	L	PF	FR	MP	RR	288				
gi   496218797	ref   WP_008932834.	EGEPGLAF	HESL	S	L	E	R	F	E	T	T	V	Q	L	M	L	PF	FR	MP	RR	302				
gi   495659402	ref   WP_008383981.	GGEPVTA	VHESL	DLDR	FA	P	W	V	Q	A	M	L	PF	-----	291										
gi   687960265	gb   KGE05197.1   _16	LGEVRAV	AHESL	DLDR	F	R	R	W	V	Q	L	L	PF	FR	IP	RR	304								
gi   496312569	ref   WP_009021747.	GGEPGAGI	HESL	S	W	Q	R	F	E	Q	W	V	Q	L	L	PF	FR	MP	RR	--	329				
gi   495571757	ref   WP_008296336.	GGETGTGI	HESL	S	M	E	R	F	E	Q	W	V	Q	L	L	PF	FR	MP	RR	TT	307				
gi   908380581	ref   WP_049751705.	DGRPLHAV	HESL	DLDR	FAS	A	W	K	P	L	L	PF	FR	MP	RR	--	287								
gi   688524263	ref   WP_029718812.	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	235
gi   332109922	gb   FGJ10550.1   _1-	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	235
gi   760103377	ref   WP_043785587.	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	283
gi   760129436	ref   WP_043810804.	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	283
gi   7416798	dbj   BAA94047.1   _1-4	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	406
gi   75345570	sp   P95619.1   CRTC_R	LGEVTSV	HET	M	L	P	R	V	I	T	L	P	V	R	I	M	L	P	W	E	R	MP	RR	--	406
gi   913634290	gb   KNZ32951.1   _1-	LGEAVES	VHET	L	S	P	R	L	S	P	A	V	R	F	M	L	P	W	E	R	MP	RR	--	286	
gi   809277751	ref   WP_046331236.	LGEVISM	HET	L	N	R	L	K	S	I	V	Q	L	M	L	P	W	E	R	MP	RR	N	--	288	
gi   518258896	ref   WP_019429104.	AGESLLAF	HET	L	S	P	R	L	S	P	V	Q	A	M	L	P	W	E	R	MP	RR	RR	--	287	
gi   518255752	ref   WP_019425960.	AGEPLLA	FHET	L	S	P	R	L	S	P	V	Q	A	M	L	P	W	E	R	MP	RR	RR	--	287	
gi   918152310	gb   KOF08324.1   _1-	AGQLLA	FHET	L	S	P	R	L	S	P	V	Q	A	M	L	P	W	E	R	MP	RR	RR	--	287	

gi|61653243|gb|AAx48196.1|\_1-2      ACEPILAFHEITLSVPRIIVSPVYQAMLPEMPRRA    287  
..            .:\*:\*            :            :            :  
              .\*:\*            \*            :            :

PLEASE NOTE: Showing colors on large alignments is slow.

## 1. Importance of the N-terminal part of CrtC

In our previous study we have already shortly discussed the preliminary results on the importance of the N-terminal part of CrtC for the catalytic activity (Hiseni et al. 2011). In order to investigate this further, truncated (*Tr*- and *Rg*CrtC) and N-terminal point mutants (*Tr*CrtC) were constructed (Table S1). We have shown here that the activity of the truncated versions is fully retained (Figure 6, main text). It should be noted here that the truncated version of *Tr*CrtC shows a larger molecular weight (Figure S2, lane 2) compared to the native ‘cleaved’ form (Figure S2, lanes 3,4), which is due to the attached N-terminal His-Tag in all constructs that is cleaved off from the full length protein, resulting in a smaller protein.

**Table S1.** N-terminal mutants of *Rubrivivax gelatinosus* (*Rg*CrtC) and *Thiocapsa roseopersicina* (*Tr*CrtC) carotenoid 1,2-hydratase.

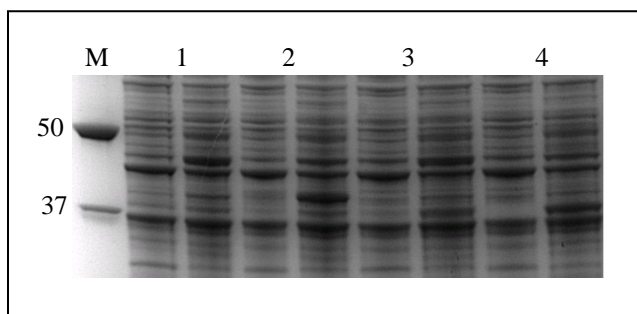
<b>Name</b>	<b>Mutation</b>
<i>Rg</i> CrtC truncated	<i>Rg</i> CrtC without first 45 AAs
<i>Tr</i> CrtC truncated	<i>Tr</i> CrtC without first 57 AAs
<i>Tr</i> CrtC S58V	<i>Tr</i> CrtC bearing mutation S58V
<i>Tr</i> CrtC S58Q	<i>Tr</i> CrtC bearing mutation S58Q

Apart from expression levels, the activities of crude cell extracts of the N-terminal constructs were measured with lycopene as substrate. Since crude cell extracts were used, the results cannot be quantitated. However, in combination with the expression levels as shown in Figure S2, indicative conclusions can be drawn. As stated in our previous study, the N-terminus is not important for catalytic activity (Hiseni et al. 2011). This again was proven here, as the truncated *Rg*- as well as *Tr*CrtC were active. Furthermore, although the introduction of N-terminal point mutations did affect the cleavage rate of the N-terminal part, it did not affect the activity of *Tr*CrtC.



**Figure S2.** SDS-PAGE (10%) analysis of expression of N-terminal mutants from *T. roseopersicina* CrtC (*TrCrtC*). M, Precision plus protein standard (BioRad). First lane of each sample shows cells before induction with 0.1 mM IPTG and the second lane shows cells after overnight expression at 25°C. 1, *TrCrtC* wild type; 2, *TrCrtC* truncated; 3, *TrCrtC* S58V; 4, *TrCrtC* S58Q.

Full length protein = 44 kDa, truncated protein w/o His-tag = 38 kDa, truncated protein with His-tag = 39 kDa.

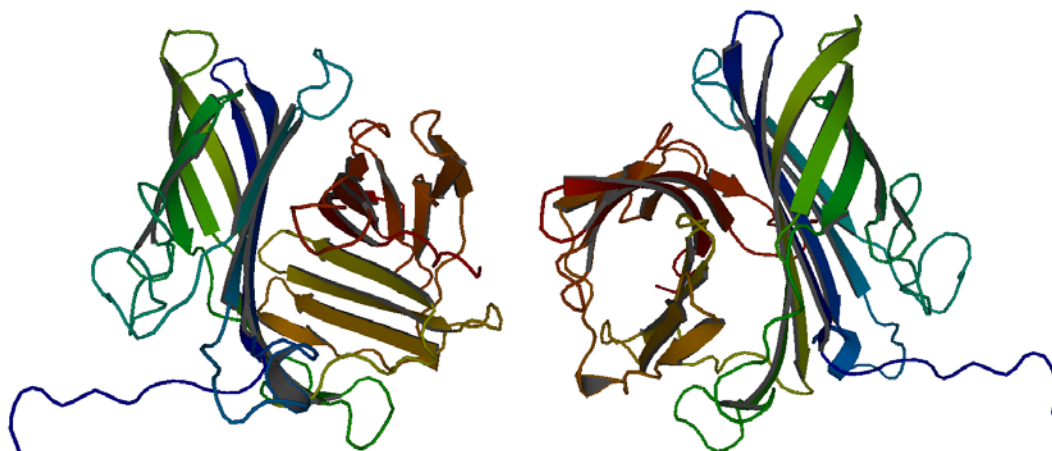


Although the reason for the truncation is still unknown, and the N-terminal region is not involved in the catalytic activity, we did explore this phenomenon a bit further. The cleavage site was determined by MS analysis, and appeared between S57 and S58. To determine whether the truncation of recombinant *TrCrtC* was processed by *E. coli* endogenous proteases or the enzyme itself, two mutants at the first amino acid after cleavage were produced. The S58 position was modified by substitutions with valine (same size, different chemical features) or glutamine (different size, same chemical features), which are no targets for known *E. coli* proteases. All resulting proteins were cleaved, but at a different rate. It seems that the cleavage rate increased in the order of wild type < S58V < S58Q. This conclusion is based on the fact that in the wild type sample in Figure S2 mainly the 44 kDa band is visible (lane 1). In the S58V mutant about the same amount of both, the 38- as well as the 44 kDa bands could be detected (lane 3), while in the S58Q mutant mainly the 38kDa band could be identified (lane 4). These differences in cleavage rate might be explained by the fact that the substitution of an amino acid by a larger or more hydrophobic amino acid could result in conformational changes which promote the processing activity, either by host proteases or through self-cleavage. Since this truncation was not the focus of this study, we did not pursue this any further. Nevertheless, additional experiments are needed to get more insight into this occurrence. For example, by expressing larger amounts of the corresponding proteins, and quickly purifying them, one could follow the change of the protein size in time. With the wild type *TrCrtC*, we have already seen that even though the purified enzyme consisted of both sizes proteins, after few days of storage, the 44 kDa could not be detected anymore, pointing towards self-cleavage (Hiseni et al. 2011), which is not disturbed by the mutations prepared in this study.

## 2. Homology model of CrtC

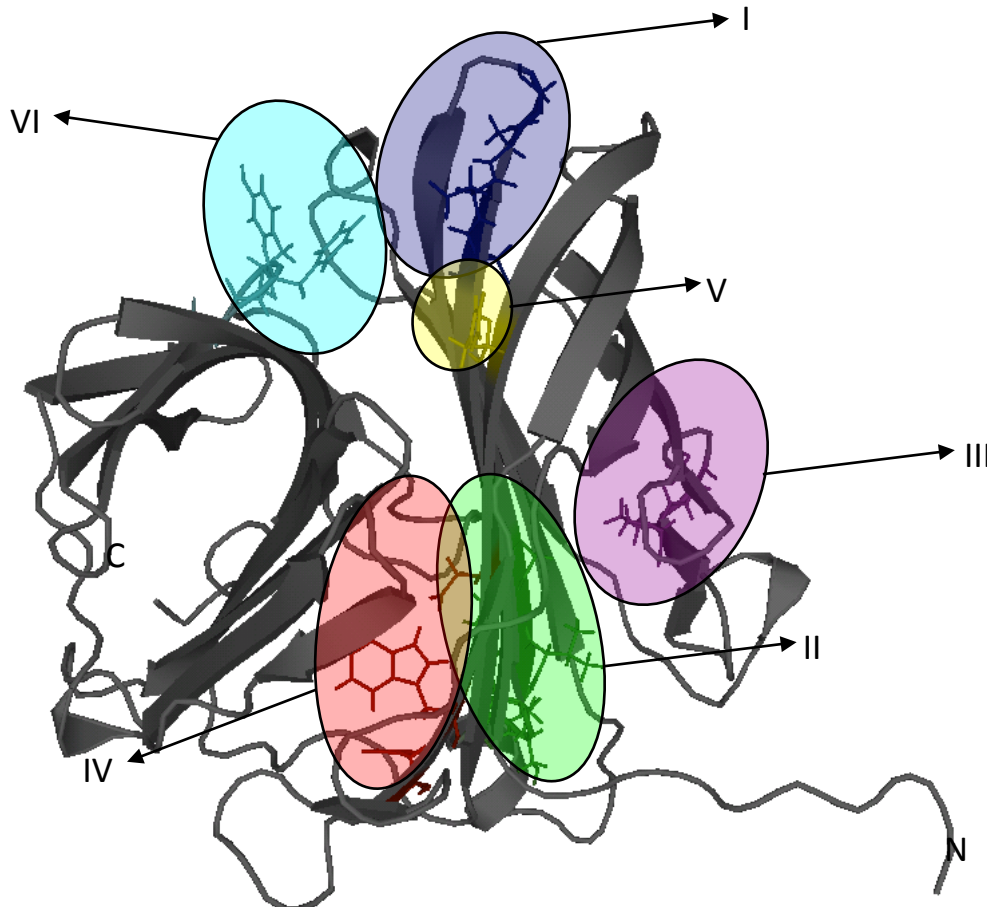
In order to investigate which one of the fully conserved residues could be involved in the catalytic hydration reaction, we tried to build a 3D structure model of *TrCrtC*. The CrtC secondary structure prediction was carried out with the program PolyView 2D (Porollo et al. 2004). Then the SWISS-MODEL program was used to model the structure of *TrCrtC* (Arnold et al. 2006). All programs were used from their respective website, using the default settings.

**Figure S3.** A homology model of CrtC from *T. roseopersicina* based on the crystal structure of a putative AttH (PDB id: 2ICH) from *Nitrosomonas europaea*. The ribbon diagrams depict front (left-hand side) and back (right-hand side) view of CrtC. The structure is color-coded from the N-terminus (blue) to the C-terminus (red).



The homology model was based on the only known 3D structure (PDB 2ICH) that showed some sequence identity (15.4%) to the CrtC (Figure S3), which is generally regarded as too low to build a good model. Modelled residue range was between 95 and 342 (QMEAN Z-Score: -6.054). The homologue is the putative AttH protein from *Nitrosomonas europaea* (Chiu et al. 2010). It belongs to a protein family of unknown function (DUF2006), which has remote similarity to the family PF07143 encompassing carotenoid 1,2-hydratases. The topology of the CrtC model structure shows similarity to lipocalins, proteins that bind and transport small hydrophobic molecules (Flower 1996). The lipocalin fold is typically formed by a large, twisted beta-sheet that closes in the back to form a central, internal, ligand-binding cavity. This folding motif is frequently found in porins, transmembrane proteins or in general in proteins that bind hydrophobic ligands/substrates (LaLonde et al. 1994).

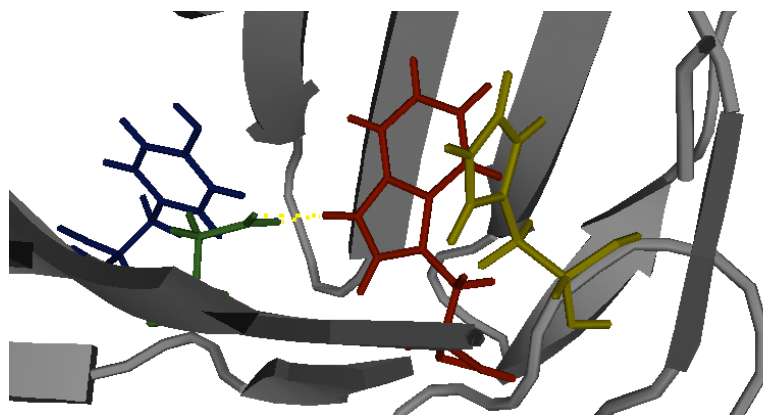
**Figure S4.** CrtC conserved residues. Ribbon diagram of *Tr*CrtC with marked regions that contain highly conserved amino acid residues. The sequence motifs, which correspond to those regions, are shown in boxes (I – VI). N and C indicate N- and C-terminus, respectively.



	I	II	III	IV	V	IV	VI
<i>Rg</i> CrtC .....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY
<i>Tr</i> CrtC .....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY
<i>Br</i> CrtC .....	SDDG.	GSVFSP.Y	GPS.	H.W.	D.Y.D.		PFY
<i>Rr</i> CrtC .....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY
<i>Rp</i> CrtC .....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY
Fos13CrtC.....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY
Fos06CrtC.....	SDDG.	GSVFSP.Y	GPS.	H.W.	H.Y.D.		PFY

We visualized the conserved residues onto the model to investigate if there is any relationship between the conserved regions and specific locations in the homology model of CrtC (Figure S4). Six striking sequence motifs were selected (I – VI) and the results show that, indeed, most of the highly conserved residues that were found across all (putative) CrtC's (Figure S1) are located either at the bottom or upper part of the cavity, with one motif (III) located outside the cavity. From a catalytic point of view, the amino acids located in regions IV and V, i.e. aspartic acid (D), tyrosine (Y) and histidine (H), are most probably involved in the catalytic hydration, as these amino acids are commonly involved as active residues in acid-base type catalyzed reactions in the active sites of enzymes (Bartlett et al. 2002). Furthermore, in this model they are all aligned and in close distance to each other (Figure S5), which is important for the contact with the substrate.

**Figure S5.** View of the potential active site of *Tr*CrtC. The conserved residues H237 (yellow), W239 (red), Y264 (blue) and D266 (green) are shown as sticks, with hydrogen bond as yellow dots (region IV).



Although the model is based on very low homology, this structure is in agreement with our final hypothesis, where Y266 could make a bond to D268 through a water molecule, and H239 is in H-bridging distance (Figure S5), so all amino acids are correctly placed to assume their hypothetical role (Figure 8, main text). Interestingly, according to the homology model, the residues H264 (*Rg*) and H262 (*Tr*), which mutation to Ala does not impair the activity of the enzyme so much, are located away (region V) from the potential active site residues. This observation further allows our hypothesis that region IV is the active site of CrtC. On the other hand, the three totally conserved hydrophobic residues proline, phenylalanine and tyrosine in region VI might play the role of attracting the hydrophobic substrate and placing it in the right position. The highly conserved mainly hydrophobic amino acids in region II, which are located close to potential active site residues in region IV, might play a role in stabilization of the substrate during catalytic activity.

## Conclusion

The homology model built for this study is merely a working model, given the limited level of sequence identity. However, the position of key residues correspond to the activity data of the mutants and allows our hypothesis on the hydration mechanism of CrtC. Due to this fact that the preserved active site residues do end up in close proximity to each other (see figure S5), we consider this model to provide a good basis for future studies, although a real crystal structure is of the utmost importance to confirm our hypothesis.

## Acknowledgements

We thank Jan van Leeuwen (Groningen University) for making the CrtC models.

## References

- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2):195-201. doi:10.1093/bioinformatics/bti770
- Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. *J Mol Biol* 324(1):105-121. doi:10.1016/S0022-2836(02)01036-7
- Chiu H-J, Bakolitsa C, Skerra A, Lomize A, Carlton D, Miller MD, Krishna SS, Abdubek P, Astakhova T, Axelrod HL, Clayton T, Deller MC, Duan L, Feuerhelm J, Grant JC, Grzechnik SK, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Kumar A, Marciano D, McMullan D, Morse AT, Nigoghossian E, Okach L, Paulsen J, Reyes R, Rife CL, van den Bedem H, Weekes D, Xu Q, Hodgson KO, Wooley J, Elsliger M-A, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) Structure of the first representative of Pfam family PF09410 (DUF2006) reveals a structural signature of the calycin superfamily that suggests a role in lipid metabolism. *Acta Crystallogr, Sect F: Struct Biol Cryst Commun* 66(10):1153-1159. doi:10.1107/S1744309109037749
- Flower DR (1996) The lipocalin protein family: Structure and function. *Biochem J* 318(1):1-14. doi:10.1042/bj3180001
- Hiseni A, Arends IWCE, Otten LG (2011) Biochemical characterization of the carotenoid 1,2-hydratases (CrtC) from *Rubrivivax gelatinosus* and *Thiocapsa roseopersicina*. *Appl Microbiol Biotechnol* 91(4):1029-1036. doi:10.1007/s00253-011-3324-1
- LaLonde JM, Bernlohr DA, Banaszak LJ (1994) The up-and-down beta-barrel proteins. *FASEB J* 8(15):1240-7.
- Porollo AA, Adamczak R, Meller J (2004) POLYVIEW: A flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics* 20(15):2460-2462. doi:10.1093/bioinformatics/bth248