

Acta Crystallographica Section F

Volume 70 (2014)

Supporting information for article:

The Structure of Rice Weevil Pectin Methylesterase

David C. Teller, Craig A. Behnke, Kirk Pappan, Zicheng Shen, John C. Reese, Gerald R. Reeck and Ronald E. Stenkamp

Table S1 Sequence alignments of CE8 proteins.

The PDB entries providing protein sequences aligned with clustalw (Sievers *et al.*, 2011) are:

4PMH	RW PME
3GRH	YbhC lipoprotein from <i>E. coli</i> (Eklöf <i>et al.</i> , 2009)
1GQ8	PME from Carrot, <i>Daucus carota</i> (Johansson <i>et al.</i> , 2002)
1XG2	PME from Tomato, <i>Solanum lycopersicum</i> (Di Matteo <i>et al.</i> , 2005)
1QJV	PME from <i>Dickeya dadantii</i> (Jenkins <i>et al.</i> , 2001)
3UW0	PME from <i>Yersinia enterocolitica</i> (Boraston & Abbott, 2012).

Residues highlighted in yellow are found in beta strands (E in dssp output) (Kabsch & Sander, 1983). Helices are denoted by green highlights.

Runs of consecutive residues with alpha carbon atoms within 3 Å of the RW PME alpha carbon atom for all five compared proteins are considered part of a core structure denoted in magenta.

Residue numbers are taken from the PDB files.

Symbols: *, conserved residues; #, binding site; ^, transition state stabilizing; =, active site

Note: The clustalw sequence alignment was taken as the master and the structure alignment was placed within it. Differences between the sequence and structure alignments result in the lack of registration especially seen beyond residue 276 in the 4PMH (RW PME) sequence.

Table S1A. Alignment of the N-terminal regions of 4PMH and 3GRH with the C-terminal peptides of 1GQ8, 1XG2, 1QJV, and 3UW0.

Symbols												
4PMH N-terminal		-----DQTAP	GTASRP	-----ILT	ASES	NYFT	TATYLQ	-----	GWSPPSIST			37
4PMH-core			P		ILTA ES		TTATYL					
3GRH N-terminal		YFQSMQRPSDQTAP	GTSSRP	-----ILS	AKEA	QNFD	DAQHYFASLTPGAAAWNPSPITL					48
3GRH-core			P		ILSA EA		DAQHYF					
1GQ8 C-terminal	273	-----AATSGRVTWKG	FKVITS	STEAG	GFT	PGSFI	IAAGGSWLKKATTFPFSLG					318
1GQ8-core					KVITS EA		TPGSFI					
1XG2 C-terminal	269	-----AGTSKRVKWPG	YHVITD	PAKAMPFT	VAKLI	IQG	GSWL	RSTGV	AYVDG			314
1XG2-core					HVITD KA		TVAKLI					
1QJV C-terminal	335	-----AAVSKDRR		QLT	DAQA	AEYT	TQSKVL	GD	--WTPTLP	-----		366
1QJV-core			R		QLTD QA		TQSKVL					
3UW0 C-terminal	332	-----AAINEGRR		QLS	AEQL	KAFT	LPMIFPD	--WAV	-----			360
3UW0-core			R		QLSA QL		TLPMIF					
seq. identities			*									

Table S1B. Alignment of the full sequences modeled in the PDB files.

Symbols												
4PMH		-----DQTAP	GTASRP	ILT	ASES	NYFT	TATYLQ	-----	GWSPPSIST	SKADYT		43
4PMH-core					PILTA ES		TTATYL					
3GRH		PAQPD	FVYFQSMQRPSDQTAP	GTSSRP	ILS	AKEA	QNFD	DAQHYFASLTPGAAAWNPSPITL				55
3GRH-core					PILSA EA		DAQHYF					
1GQ8 N-terminal		-----SSTVGPN	VVVAADGSGDYK	-----								20
1GQ8-core							YK					
1XG2 N-terminal		-----IIAN	AVVAQDGTGDYQ	-----								16
1XG2-core							YQ					
1QJV N-terminal		-----ATTYN	AVVSKSSSDGKTFK	-----								43
1QJV-core							FK					
3UW0 N-terminal		-----					FS	-----				42
3UW0-core							FS					
seq. identities												

Symbols

```

4PMH      VGNGL-----YNTIQAAVNAAINA GGTTRKYIKIINAGTYQEVVYI PNTKVP LTIYGGSS 97
4PMH-core YNTIQAAVNAA          RKYIKINAGTYQEVVYIP      PLTIYGG
3GRH      VGPAQTQGVTHHTTIQAAVDAA IIKRTNKRQYIAVMPGEYQGTVYV PAAPGGITLYGTGEK 115
3GRH-core HTTIQAAVDAA          RQYIAVMPGEYQGTVYVP      GITLYGT
1GQ8      -----TVSEAVAAA APEDSK-TRYVIRIKAGVYRENVDP PPKKKNIMFLGDGRT 67
1GQ8-core TVSEAVAAA          RYVIRIKAGVYRENVDP      NIMFLGD
1XG2      -----TLAEAVAAA APDKSK-TRYVIYVKRGTYKENVEV ASNKMLMIVGDGMY 63
1XG2-core TLAEAVAAA          RYVIYVKRGTYKENVEVA     NLMIVGD
1QJV      -----TIADAIASAPAGS--TPFVILIKNGVYNERLTI TRN--NLHLKGESRN 87
1QJV-core TIADAIASA          PFVILIKNGVYNERLTI     NLHLKGE
3UW0      -----SINAALKSAPKDD--TPFIIFLKNGVYTERLEVAR S--HVTLKGENRD 86
3UW0-core SINAALKSA          PFIIFLKNGVYTERLEVA     HVTLKGE
Seq. identities          *      *          *      *      *          *

```

```

Symbols
4PMH      PSDTLITLNMPAQTTPSAYKSLVG--SLFNSADPAYSMYNCSASK-SGTIGTSCSTVFWV 154
4PMH-core PSDTLITLNMPA          TIGTSCSTVFWV
3GRH      PIDVKIGLSLDGGMSPADWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAV FWS 175
3GRH-core PIDVKIGLSLDG          SIGVLCSAV FWS
1GQ8      STIITASKNV---QDGS-----TTFNSATVAA 91
1GQ8-core RTSTIITASKNV      GS          TTFNSATVAA
1XG2      ATTITGSLNV---VDGS-----TTFRSATLAA 87
1XG2-core MYATTITGSLNV      GS          TTFRSATLAA
1QJV      GAVIAAATAAGTLKSDGSK-----WGTAGSSTITI 117
1QJV-core RNGAVIAAATAA      K          WGTAGSSTITI
3UW0      GTVIGANTAAGMLNPQGEK-----WGTSGSSTVLV 116
3UW0-core RDGTVIGANTAA      K          WGTSGSSTVLV
Seq. identities          *

```

Symbols		#	#		^				
4PMH	KAPAVQIVNLSIEN	-----	SAKN	---	TGDQQAVALQTN	--	SDQIQIHNARLLGHQ	199	
4PMH-core	KAPAVQIVNLSIEN		SA		QQAVALQTN		SDQIQIHNARLLGHQ		
3GRH	QNNGLQLQNLT IEN	-----	TLGDSVDAGNH	PAVALRTD	--	GDQVQINNVN	ILGRQ	223	
3GRH-core	QNNGLQLQNLT IEN		TL		HPAVALRTD		GDQVQINNVN	ILGRQ	
1GQ8	VGAGFLARDITFQN	-----	TAGA	AKHQAV	ALRVG	--	SDL	SAFYRCDILAYQ	135
1GQ8-core	VGAGFLARDITFQN		TA		HQAV	ALRVG	SDL	SAFYRCDILAYQ	
1XG2	VGQGFILQDICIQN	-----	TAGPAK	DQAV	ALRVG	--	ADM	SVINRCRIDAYQ	131
1XG2-core	VGQGFILQDICIQN		TA		DQAV	ALRVG	ADM	SVINRCRIDAYQ	
1QJV	SAKDFSAQSLTIRNDF		FPANQ	AKSDSDSS	IKDTQAV	ALYVTKSG	DRAYFKDVSLVGYQ	177	
1QJV-core	SAKDFSAQSLTIRNDF				TQAV	ALYVT	DRAYFKDVSLVGYQ		
3UW0	NAPNFTAENLTIRNDF		FPANKK	KADTDP	TKLKD	TQAV	ALLLAENS	DKARFKAVKLEGYQ	176
3UW0-core	NAPNFTAENLTIRNDF				TQAV	ALLLA	DKARFKAVKLEGYQ		
Seq. identities		*			****		*	*	

Symbols	=		#	=	#		#						
4PMH	DTLYA	GSG	-----	SSSVE	RSYYT	NTYIEG	DIDFVFGG	SAIFE	SCTFYV	KADRRSDT	251		
4PMH-core	DTLYA				ERSYYT	NTYIEG	DIDFVFGG	SAIFE	SCTFYV	KAT			
3GRH	NTFFV	TNSGVQNR	LETNRQP	R	TLVT	NSYIEG	DVDIVS	GRGAV	VFDNTE	FRVVNS	RQTQE	282	
3GRH-core	NTFFV				PRTLVT	NSYIEG	DVDIVS	GRGAV	VFDNTE	FRVVNS	ETQE		
1GQ8	DSL	YV	VHSN	-----	RQFFI	NCFIAG	TVDF	IFGNA	AVVLQDC	DIHARR	PGSGQK	182	
1GQ8-core	DSL	YV	N		RQFFI	NCFIAG	TVDF	IFGNA	AVVLQDC	DIHARR	K		
1XG2	DTLY	A	HSQ	-----	RQFYR	DSYVTG	TVDF	IFGNA	AVVFQK	QLVARK	PGKYQQ	178	
1XG2-core	DTLY	A	Q		RQFYR	DSYVTG	TVDF	IFGNA	AVVFQK	QLVARK	Q		
1QJV	DTLY	V	SGG	-----	RSFFS	DCRISG	TVDF	IFGDG	TALFN	NCDLV	SRYRAD	VKSG	225
1QJV-core	DTLY	V	G		RSFFS	DCRISG	TVDF	IFGDG	TALFN	NCDLV	SRY	V	
3UW0	DTLYS	K	TGS	-----	RSYFS	DCEISG	HVDF	IFGSG	ITVFD	NCNIV	ARDRSD	IEP	224
3UW0-core	DTLYS	K	S		RSYFS	DCEISG	HVDF	IFGSG	ITVFD	NCNIV	ARD	P	
Seq. identities					*	*	*	*					

Symbols		#	#
4PMH	---AVVFAP-DTDPHKMYGYFVYKSTITGDS-----AWSSSKKAYLGRAW	292	
4PMH-core	AVVFAP DT PH GYFVYKSTITG KAYLGRAW		
3GRH	---AYVFAP-ATLSNIYYGFLAVNSRFN-----AFGDG-VAQLGRSL	319	
3GRH-core	AYVFAP AT SN GFLAVNSRFN A VAQLGRSL		
1GQ8	---NMVTAQGRTPNQNTGIVIQKSRIGATSDDLQPVQSSFP TYLGRPWKEYSRTVVMQSS	239	
1GQ8-core	NMVTAQGR PN GIVIQKSRIGA PTYLGRPWK YSRTVVMQSS		
1XG2	---NMVTAQGRTPNQATGTSIQFCNIIASSDLEPVLKEFP TYLGRPWKEYSRTVVMESY	235	
1XG2-core	NMVTAQGR PN GTSIQFCNIIA PTYLGRPWK YSRTVVMESY		
1QJV	NVSGYLTAP-STNINQKYGLVITNSRVIR ESDSVPAKSYG---LGRPWHPTTTFSDGRYA	281	
1QJV-core	GYLTAP ST IN GLVITNSRVIR SYG LGRPWH		
3UW0	-PYGYITAP-STLTTSPYGLIFINSRLTK EP-GVPANSFA---LGRPWHPTTTFADGRYA	278	
3UW0-core	GYITAP ST TT GLIFINSRLTK SFA LGRPWH		
Seq. identities	* * *		

Symbols			
4PMH	DSGVSSSSAYVPGT---SPNGQLIIKESTIDGIINTSGPWT TATSG-RTYSGNNANS---	345	
4PMH-core	D NGQLIIKESTIDGIIN PWTTA		
3GRH	DVDANT-----NGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVDND	366	
3GRH-core	D NGQVVIRDSAINEGFN PWADA		
1GQ8	ITNVINPAGWFPWDGNFALDTLYYGEYQNTGAGAATSGRVTWKG---FKVITS STEAQGF	296	
1GQ8-core	ITNVIN GWFPW LYYGEYQNTGAGA KVITS EA		
1XG2	LGGLINPAGWAEWDGDFALKTLYYGEFMNNGPGAGTSKRVKWPG---YHVITD PAKAMPF	292	
1XG2-core	LGGLIN GWAEW LYYGEFMNNGPGA HVITD KA		
1QJV	DPNAIGQTVFLNTS----MDNHIYGWDKMSGKDKNGNTIWFNPEDSRFFEYKSYGAGAAV	337	
1QJV-core	IGQTVFLNTS MDNHIYGWDKM SRFFEYKSYGAGA		
3UW0	DPAAIGQSVFINTT----MDDHIYGWDKMSGKDKQGEKIWFYPQDSRFFEANSQGPAAI	334	
3UW0-core	IGQSVFINTT MDDHIYGWDKM SRFFEANSQGPAAI		
Seq. identities			

Symbols

```

4PMH      ----RDLNNDNYNRFWEYNNSGNGA---- 366
4PMH-core          NRFWEYNNSGNGA
3GRH      DEIQRNLNDTNYNRMWEYNNRGVGSK--- 392
3GRH-core          NRMWEYNNRGVGSK
1GQ8      TPGSFIAGGSWLKATTFPFSLGL----- 319
1GQ8-core  TPGSFI
1XG2      TVAKLIQGGSWLRSTGVAYVDGLYD---- 317
1XG2-core  TVAKLI
1QJV      SKDRRQLTDAQA AEYTQSKVLGDWTP TLP 366
1QJV-core          RQLTD QA TQSKVL
3UW0      NEGRRQLSAEQLKAFTLPMIFPDWAV--- 360
3UW0-core          RQLSA QL TLP MIF
Seq. identities

```

References

- Boraston, A. B. & Abbott, D. W. (2012). *Acta Cryst.* F68, 129-133.
- Di Matteo, A., Giovane, A., Raiola, A., Camardella, L., Bonivento, D., De Lorenzo, G., Cervone, F., Bellincampi, D. & Tsernoglou, D. (2005). *Plant Cell.* 17, 849-858.
- Eklöf, J. M., Tan, T.-C., Divne, C. & Brumer, H. (2009). *Proteins - Structure Function and Bioinformatics* 76, 1029-1036.
- Jenkins, J., Mayans, O., Smith, D., Worboys, K. & Pickersgill, R. (2001). *J. Mol. Biol.* 305, 951-960.
- Johansson, K., El-Ahmad, M., Friemann, R., Jornvall, H., Markovic, O. & Eklund, H. (2002). *FEBS Lett.* 514, 243-249.
- Kabsch, W. & Sander, C. (1983) *Biopolymers* 22, 2577-2637.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). *Molecular Systems Biology* 7, 539-544.