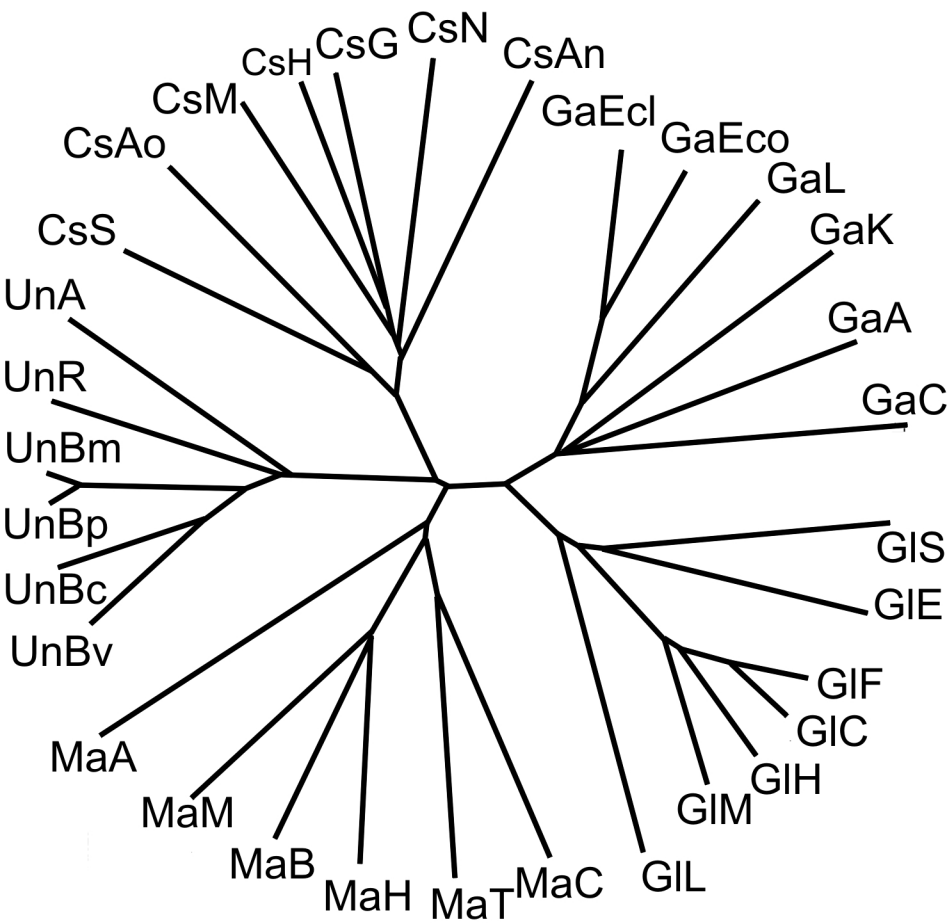


SUPPLEMENTARY FIGURES LEGEND

Suppl. Fig. 1. Phylogenetic analysis of the $(\alpha/\beta)_8$ -barrel sequences of GH2 family members. The name of the originating organism and the database entry are indicated. Cs: CsxA and related ORFs. CsS: *Streptomyces avermitilis* (NP_822398); CsAo: *Amycolatopsis orientalis* (this work; CsxA, AAX62629); CsN: *Neurospora crassa* (XP_331434.1); CsM: *Magnaporthe grisea* (MG05864.4); CsG: *Gibberella zeae* (FG02314.1); CsH: *Hypocrea jecorina* (BAD99604); CsAn: *Aspergillus nidulans* (AN2824.2); Ga: β -galactosidases from: GaK: *Kluyveromyces lactis* (P00723; BGAL_KLULA); GaL: *Lactococcus lactis* (Q48727; BGAL_LACLA); GaEcl: *Enterobacter cloacae* (Q47077, BGAL_ENTCL); GaEco: *Escherichia coli* (P00722, BGAL_ECOLI); GaA: *Arthrobacter* sp. (Q59140, GAL_ARTSB); GaC: *Clostridium acetobutylicum* (P24131, BGAL_CLOAB); Gl: β -glucuronidases from: GIC: *Canis familiaris* (O18835, BGLR_CANFA); GIF: *Felis silvestris catus* (O97524, BGLR_FELCA); GIM: *Mus musculus* (P12265, BGLR_MOUSE); GIH: *Homo sapiens* (P08236, BGLR_HUMAN); GIE: *Escherichia coli* (P05804, BGLR_ECOLI); GIS: *Staphylococcus* sp. (patent PCT/US98/19217); GIL: *Lactobacillus gasserii* (AAK07836, Q9AHJ8_9LACO); Ma: β -mannosidases from: MaA: *Aspergillus niger* (Q9UUZ3, MANBA_ASPNG); MaB: *Bos taurus* (Q29444, MANBA_BOVIN); MaH: *Homo sapiens* (O00462, MANBA_HUMAN); MaM: *Mus musculus* (Q8K2I4_MOUSE); MaT: *Thermobifida fusca* (CAD33708, Q8KLI9_THEFU); MaC: *Cellulomonas fimi* (AAD42775, Q9XCV4_CELFI); Un: proteins of unknown catalytic activity having a glutamate as a putative acid/base residue preceded by a serine: UnA: *Agrobacterium tumefaciens* (NP_533240, ATU2575); UnBm: *Burkholderia mallei* (YP_101889, BMA0039); UnBp: *Burkholderia pseudomallei* (YP_107240, BPSL0611); UnBc: *Burkholderia* sp. 383 cepacia (YP_372566_ZP_00215636); UnBv: *Burkholderia vietnamiensis* (ZP_00425692); UnR: *Ralstonia eutropha* (ZP_00166278).

Suppl. Fig. 2. Assignment of putative catalytic residues in GH2 family members through structure-guided alignment. Symbols as in Suppl. Fig. 1. Alignment was obtained by the 3DCoffee structure guided alignment method [20] using the domain 3 from human β -glucuronidase as template. Secondary structures are shaded in dark gray (α -helices) or light gray (β -sheets). Secondary structures formally identified in crystallized proteins are shown for human β -glucuronidase (PDB ID: 1bhg:A) and *E.coli* β -galactosidase LacZ (PDB ID: 1DPO:A). These two proteins are indicated by bold symbols **GIH** and **GaEco**, respectively. Secondary structures in other proteins were predicted by the PSIPRED algorithm [21]. The putative catalytic acid/base and nucleophile residues are indicated with asterisks. Catalytic residues formally identified by mechanistic studies [35-39] are underlined. Other strictly conserved residues are indicated with triangles.



Supplementary Figure 2

	340	350	360	370	380	390	400				
CsAo	D--VKA--TLNSSGG	QYSVNGKPLLR	IRGGYTPDLFL--RW-NE	TAAADKLKYV	LNGLNTVRLEGH--IE	PDEFFDIA					
CsS	D--VKA--PLNSD	GARQYSVNGRRL	LKGGWSPDEFL--RW-DS	TYVEDRLRYALD	LGLNTRIRLEGH--IE	PDEFFDLA					
CsH	K--VTS--FVNQF	NDTQYSVNGHP	FQVIGGGYGADMFL--RW-DS	ERFTRIVEYMLDM	HQNTIRLEGG--MEHPELYEIC						
CsN	T--VTS--QLNSF	NDITFLINSHPF	QVLGAGYAPDMFL--RF-S	PAKFESEARYVLD	LGFNTRIRLEGG--NEHPELYQIA						
GLH	VA-----	VTKSQFLINGK	PFYFHGVNKHEDADIR	KGKGFDP	WPLLKDFNLLR	WLGANAFRTSHY--PYAEEVMQMC					
G1M	VA-----	VTKSKFLINGK	PFYFQGVNKHEDSIR	KGKGFDP	WPLLKDFNLLR	WLGANSFRTSHY--PYSEEVLQLC					
G1E	VA-----	VKGEQFLINHK	PFYFTGFRHEDADLR	KGKGFDP	NVLMVHDHALMD	WIGANSYRTSHY--PYAEMLDWA					
G1L	VK-----	IVNDKILLN	NHPIYLKGF	GKHEDFNVLGKAVNESI	IKRDYECMKWIGANC	FRSSHY--PYAEEWYQYA					
GaC	FE-----	MKDKIMCLK	WKRIIFKGVNRHEFS	ARRGRSIT	KEDMLWDIKFLKQHN	INAVRTSHY--PNOSLWYRLC					
GaA	VK-----	IVGDQFLV	NGRKFVIFHGVNRHET	NADRGRV	FEASAREDLALMKR	FNVAIRTSHY--PHHPRFLDLA					
GaEc1	VE-----	ISNGLLKL	NGKPLLRGVNRHEH	SENGQVMD	EATMRRDIETMKQ	HSFNAVRTSHY--PYAEMLDWA					
GaEco	VR-----	IENGLLLL	NGKPLLRGVNRHEH	PHLHGQVMD	EQTMVQDILLMKQ	NNFNVAVRTSHY--PNHPLWYTL					
MaB	VELVEEPI	QNSPGLSFYFK	INGLPIFLKGSNWIP	ADSFQDRV-T	SAMRLLQLQSVVD	ANMNLRVWGGGYE	QDEFYELC				
MaH	VELIEEPI	KGSPGLSFYFK	INGLPIFLKGSNWIP	ADSFQDRV-T	SELLRLLQLQSVVD	ANMNLTRVWGGGYE	QDEFYELC				
MaM	VQLIEEPI	KGSPGLSFYFK	INGLPIFLKGSNWIP	ADSFQDRV-T	SRLQLLQSVVD	ANMNLTRVWGGGYE	QDEFYALC				
MaC	VEVD--	TPDE	DGTPFTRVNGRPV	FVKGANWIP	DDHLLTRI-T	RERLAHRLDQAVE	ANLNLRVWGGGYE	SEDFYDLC			
UnA	-----	GNGFGLK	VNGQPVFCRGAV	TTAIVRLPG-T	RPDYE	PWLKKAEEAGNM	MIRVGGTMA	YETPEFFALC			
UnBc	-----	GKGFGLR	INGVVPV	FARGACWSSA	APALHA-D	DATYARLLGL	LARDAGFN	MIRVGGTMTY	EADAFHAWC		
UnBv	-----	GNGFGLR	VNGTPV	FARGACWSSA	APALHA-D	DATYARLLGL	LARDAGFN	MIRVGGTMTY	EADAFHAWC		
UnBm	-----	GRGFALS	VNGTPLF	FARGACWTS	ADPVLGHA-D	APAYRRAL	LVLARDAGC	NMIRVGGTMI	YADAFYALC		
		::	▲			▲	▲				
	410	420	430	440	450	460	470				
CsAo	DDLGLV	LTMPGW	ECCDKWEGQ---	VNGEEK-G---	EPWVES	DYPIAKASMF	SEAERLRD	HPSVISFHIGSDF	FAPDRRIE		
CsS	DRYGIL	TLPGW	CCNKWEGN---	VNGSGS-G---	DEWTA	ADYPVAKAS	MAAEEARLRD	HPSVVSFLIGSDF	FAPDAKIE		
CsH	DEYGLM	VMPGW	ECCDKWEAW---	AYNDEL-A	I	FPFVW	ANDYETANY	SMIHEAAMQ	PHPSVLTFLVGSDF	FWNDEAV	
CsN	DRLGIM	L	AGWECCDKWEAW---	SYNQDL-TV	PTPV	WSEDDY	NIANASMF	HEAGMLQ	THPSILGYLIGSDY	WDP	
GLH	DRYGI	VVIDE	CPGVGLALPQFFN---	-----	-----	NVSLHHHMQ	MEVVRRD	KNHPAVVM	SVANEP	PASHLESA	
G1M	DRYGI	VVIDE	CPGVGLVLPQSF	-----	-----	NESLRHHLE	VMEELVRRD	KNHPAVVM	SVANEP	PSSALKPA	
G1E	DEHGI	VVIDE	TA	AVGFNLSL	GIGFEAGN	PKELYSEEAVN	GETQQAHL	QAIKELI	ARDKNHPSV	VMWSIANEP	DPTRPQGA
G1L	DKYGF	LIDEV	PAVGLNRSIT	NFLNVTNS	NQSHFFASK	TP	PELKKVHEQE	IKEMIDRD	QRHPSVIA	WSLFNEPE	STQES
GaC	DEYGI	YLIDE	TNLESHGS---	WQKMGQIE	PSWNV	PGSLP	QWQAAVLD	RASSM	VERDKNHPSV	LWSCGNE	SYAGE
GaA	DELGF	VVILE	CDLETHGF---	HALKWV---	GNP	SDDPA	WRDALVDR	MERTVERD	KNHASI	VMWSLGN	ESGTGANLA
GaEc1	DRYGL	VVDE	ANIETHGM---	VPM	SRL---	ADD	PRWLPAM	SERVTR	RMVQRDRN	HPSII	IWSLGN
GaEco	DRYGL	VVDE	ANIETHGM---	VPM	NRL---	TDD	PRWLPAM	SERVTR	RMVQRDRN	HPSVI	IWSLGN
MaB	DELGI	MVQD	FMFAC---	-----	ALY	PTD-K	DFMDSV	REEVTHQ	VRRLLK	SHPSII	IWSGNN
MaH	DELGI	MVQD	FMFAC---	-----	ALY	PTD-Q	GFLDSV	TAEVAYQ	IKRLK	SHPSII	IWSGNN
MaM	DELGI	MVQD	FMFAS---	-----	ALY	PTD-P	GFLASV	RKEVTV	QVRRLLK	SHPSII	IWSGNN
MaC	DERGL	LVVQD	FLLAC---	-----	AAY	PEE-Q	IWDELE	AEAREN	VARLTP	HASLV	WNGGN
UnA	DELGI	MVQD	DAMLAN---	-----	FDY	PAKDD	GLRQHI	VMEIE	QLLE	STALSP	SFAIF
UnBc	DRLGL	LVVQD	FMFAN---	-----	FDY	ALDD	PAFADN	VREARQ	FLSRHS	SASPS	SLAVLC
UnBv	DRLGL	LVVQD	FMFAN---	-----	FDY	ALDD	RAFADA	VDAEAD	QFLARR	RASPS	SLAVLC
UnBm	DELGL	LVVQD	FMLAN---	-----	FDY	PSND	PRFAES	LKREAE	QFLGRH	MARPSI	AVLC
	▲	▲									* :
	480	490	500	510							
CsAo	QGYLD	-----	AMKAA	DFLLP	VPIPAAS-AR	PSP-ITG	ASGMK-----				
CsS	KTYLD	-----	ALKAA	DWPTP	VVAAS-DK	SSP-VSG	SSGMK-----				
CsH	VLYAN	-----	ALKN	AGWQ	TFIIASAS-K	RGF	PALLG	PGGMK-----			
CsN	PMYIN	-----	TLRL	QDWQ	TPVLS	SSAS-K	RGF	PSMG	SPGLK-----		
GLH	GYI	-----	LKM	VI	AHTKSLDP-SR	PVTV	FVSN	SNYAA--DK	GAPYVD	VICL	NSYYSWY
G1M	AYY	-----	FKTL	I	THTKALDL-TR	PVTV	FVSN	AKYDA--DL	GAPYVD	VICV	NSYYSWY
G1E	REY	-----	FAP	LAE	ATRKLDP-TR	PIT	CVN	MFCDAHTDTIS	DLFDV	LCLN	RYYGWY
G1L	YDY	-----	FKDI	F	AFARKLDP	QNR	PYTGT	LV	MGSG	PKVD	KLHPL
GaC	Q	-----	MSKY	FR	KKDP-SR	L	VH	YEGV	TRC-----	REFM-----	TR-RHESRM
GaA	A	-----	MAA	W	THARDL-SR	PV	H	YEGDY	TG-A-----	YTDVYS-----	RMYSSTP-ETDSIG
GaEc1	A	-----	LYR	W	LKTDP-TR	PV	Q	YEGG	ANTAA-----	TDIVC-----	PMYARVD-RDQPPF
GaEco	A	-----	LYR	W	I	K	S	VDP-SR	PV	Q	YEGG
MaB	WYD	TK	PGYLQ	TYIK	DYV	TL	V	VKNIR	TIVLE	G	DQ-TR
MaH	WYH	S	TDR	PIYIK	DYV	TL	V	VKNIR	ELV	L	AGDK-SR
MaM	WFH	V	NR	DM	KTYID	DYV	TL	V	VKNIR	KIV	LSEDK-SR
MaC	WG	P	Q	E	L	E	G	R	T	W	G
UnA	GL	P	E	Q	-----	I	W	K	-----	T	L
UnBc	GL	G	P	K	-----	Q	R	F	-----	E	L
UnBv	GL	G	P	K	-----	Q	R	A	-----	E	L
UnBm	GL	A	P	D	-----	E	R	R	-----	P	A

	520	530	540	550	560	570	580
CsAo	NGPYDYV---PPVYWDK---SQKDRGGAW---SFNSETSAGVDIPTMDTLKR-MMSASELDTMWKNPSAKQYHRSSS						
CsS	TGPYDWI---PPNYWYA---KREGGAT---GFNSETSAGVDIPTLDTLRR-MMTPAELDTLWKNPGAKQYHRSPS						
CsH	DGPYDWW---PPNYWYDV---EPS ED RLGAAF---GFGSELGAGVGTPELSSLR-FLNQSDLDDLWKNPNKNL FHM ST-						
CsN	EGPYDWW---PPNYWYDT---EPSAGRFGAAF---GFGSELGAGVGTPELSSLR-FLSPDLTDLWKNPNKDLF HMS ST						
GlH	LQLATQF---ENWYK---KY-----QKPIIQSE-YGAETIAGFHQDPPL-----						
GlM	PQLNSQF---ENWYK---TH-----QKPIIQSE-YGADAI PGIHEDPPR-----						
GlE	KVLEKEL---LAWQE---KL-----HQPIITE-YGVDTLAGLHSMYTD-----						
GlL	KMLEDEL---DGWQNL-KL-----NKPFVFE-FGADTLSSSHRLPDE-----						
GaC	YAKAAEI---EEYLN-DNP-----KKPYSCE-YMHSM-----GNS-----						
GaA	RNDSHALLGCNAIESARQR-----TRPFILCE-YVHAM-----GNG-----						
GaEcl	AVPKWSI---KKWIGMPDE-----TRPLILCE-YAHAM-----GNS-----						
GaEco	AVPKWSI---KKWLSLPGE-----TRPLILCE-YAHAM-----GNS-----						
MaB	YDLNYGD---VHFYDY---VSDCWNWRTFPKARFVSE-YGYQSWPSFSTLEK-VSSEEDWSYRSS---FALHRQHL						
MaH	NSNYFGD---VHFYDY---ISDCWNWVFPKARFVSE-YGYQSWPSFSTLEK-VSSTEDWSFNK---FSLHRQHH						
MaM	YSIQYGD---IHFYNY---ADDCWNWVFPKARFVSE-YGYQSWPSFSTLEK-VSSQEDWAYNSR---FSLHRQHH						
MaC	NDPDHGT---HHEWEV---WNRVDYSA Y RDDV---PRFCE-FGFQGPPTWSTLTR-AVRADDGGPLTKDDPT FLLH -QKA						
UnA	FFPNAGV---GHYYGVGAYCRPLEDARRADL---RFAAECLAFANIPEQETLDRYLPGVAVHDP RWK ---ARTPRDRG						
UnBc	FAPRERV---SHYYGVGAYLRPLDDARRADV---RFASECLAFSNVPCDATLAE-LGWPGVHEPRWK---AAVPRDPG						
UnBv	FVPRERV---SHYYGVGAYLRPLDDARRADV---RFASECLAFANVPCDATLER-IGRPHPEPRWK---AAVPRDPG						
UnBm	FAPREGV---THYYGVGAYLRPLEDARRAGV---RFASECLAFANVPCDATLAS-IGSPAAHEPAWK---RAVPRDPG						

*

	590	600	610	620	630	640
CsAo	--DTFGLKLFGDALTKRYGASAN---LNDFVR-KAQLSQYENVRAEFESHRYTD-----STNPSTGLIYWMLNSP					
CsS	--SVFGLTKIYDAALAGRYGAPTG---LTDYVR-KAQLAQYENVRAQFEAYGRGATD-----ASKPATGVIYWMFNNG					
CsH	NASSFYNRKIYNQGLWKRYGAPTS---LDDYLL-KAQMDYEA TR AQYEGFGALWT---ASRPATGVIYWMNLNA					
CsN	ETSSFRNRRIYNTGLWNRWSAPTS---LEDYVQ-KSQITDYEA TR AQFEGYAANWGN-----SORPATGMIYWMLNGA					
GlH	-----MFT EEY QKSLLEQY---HLGLDQKRRKYVVGELIWNFADF					
GlM	-----MF SEY QKAVLENY---HSVLDQKRRKEYVVGELIWNFADF					
GlE	-----MW SEY QCAWLDMY---HRVFD--RVSAVVGEGV QV WNFADF					
GlL	-----MWS Q EYQNEY Y QMY---FDIFK--KYPFICGELVWNFADF					
GaC	-----TGG-----MMKY---TELED--KYL MY QGG F IWDYGDQ					
GaA	-----PGA---IDQY---EDLVD--KYPRLHGGFVWEWRDH					
GaEcl	-----FGG---FAKY---WQAFR--SHPR L QGGFVWDVVDQ					
GaEco	-----LGG---FAKY---WQAFR--QYPR L QGGFVWDVVDQ					
MaB	--INGNNEMLHQIELHFKLPNSTDQLRRFKDTLY--LTQVMQAQCVKTE TE EFYRRSRSEIVN--GKGHTMGALYWLQNDI					
MaH	--EGGNKQMLYQAGLHFKLPQSTDP LR TFKDTIY--LTQVMQAQCVKTE TE EFYRRSRSEIVD--QQGHTMGALYWLQNDI					
MaM	--EDGNHQMLHQVKMHFKLPQGTDP LR TFKDTIY--LTQVMQAQC IK TE TE EFYRLSRSEIVD--GKGHTMGALYWLQNDI					
MaC	--EDGNGLDRGLAPHLGV P AG---FVDWHW--ATQLNQARAVAF AE IHYRS-----WWPRTAGAI V WLQNDI					
UnA	--ASWDFEDV RD HYL K LLYEAPDVLRRREDGALYL DM SRA TA EVM EAT FAEWRR-----SGSSCGALVW Q LDL					
UnBc	--TSWDFED IR DHYL Q TLY D VV P DLR LR RED PS RY F EL S RS VI AD LM RET F SEWRR-----TGSR C AGALV W Q F QDV					
UnBv	--ASWDFED D VR D HYL R TLY D V A PER L R R RED P ARY F EL S RA VI AD VM RET F SEWRR-----TGSR C AGALV W Q F QDV					
UnBm	--APWDFED D VR D HYL R ALY G VE P AR L R S ID P ARY L T L SRA V AD L V ET LAEWRR-----VGSS C AGALV W Q F QDV					

▲ ▲

	650	660	670
CsAo	----W T SLHW Q LFDAYMD-----QNGAYYGAKKAN E PLHI		
CsS	----W T SLHW Q LLDRYLD-----QGGAYFGAKKAN E PLHV		
CsH	----W P SLHW N QFDYYLH-----PAGSYFGTKVGSRIEHV		
CsN	----F P SLHWSIWDYYMH-----PAGAYFGAKVGSRIEHV		
GlH	----MTEQSPTRVLGN-----KKGIFTRQRQPKSA A FL L RERY W KIAN		
GlM	----MTNQSPLRVIGN-----KKGIFTRQRQPKTSA F IL L RERY W RIAN		
GlE	----ATSQGI L RVGGN-----KKGIFTRDRKPKSA A FL L QKRW T GMN		
GlL	----KTSEGI M RVGGN-----DKGIFTRDREPKDIA F T L KKR W QQLN		
GaC	ALYRKL P DGKEV L AYGGDF T DRPTD Y NFSGNGLI Y ADRTI S PKAQEV K LY Q NVKL		
GaA	GIRTRTADGTEFFAYGGDF E VIH D GNF V MDGMIL S DSTPT F GLFEY K QIV S PIRL		
GaEcl	ALTKREDGNTFFWYGGDFGDKPNDRQ F CLNGLV F PDRT P HPAL E YAHG P Q F FFTF		
GaEco	SLIKYDENG N PWSAYGGDFG D TPNDRQ F CMNGLV F ADRT P HPAL E AKH Q Q F FFQF		
MaB	----WQAPSW S LE Y GGK-----WKMLHYFAR H FAPL--		
MaH	----WQAPSW A SLE Y GGK-----WKMLHYFA Q NFFAPL--		
MaM	----WQAPSW A SLE Y GGK-----WKMLHYFA R RFAPL--		
MaC	----WPVTSWAAIDGDER-----VKPLW H ALRRAYAPR--		
UnA	----VPGAGWGI I DAAGR-----PKSV W HALKRAFR--		
UnBc	----MPGAGW G MIDAHR-----PKSA W HALRQVLQ---		
UnBv	----MPGAGW L LDAHL-----PKSG W YALRQVLQ---		
UnBm	----MPGAGW G LVDAGR-----PKSA W HALRRV-----		