

## Supplementary Material

### **Unique structural fold of LonBA protease from *Bacillus subtilis*, a member of a newly identified subfamily of Lon proteases**

Alla Gustchina<sup>1</sup>, Mi Li<sup>1,2</sup>, Anna G. Andrianova<sup>3</sup>, Arsen M. Kudzhaev<sup>3</sup>, George Lountos<sup>1,2</sup>, Bartosz Sekula<sup>1,4</sup>, Scott Cherry<sup>1</sup>, Joseph E. Tropea<sup>1</sup>, Ivan V. Smirnov<sup>3</sup>, Alexander Wlodawer<sup>1\*</sup>, Tatyana V. Rotanova<sup>\*3</sup>

<sup>1</sup>Center for Structural Biology, National Cancer Institute, Frederick, MD 21702, USA

<sup>2</sup>Basic Science Program, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

<sup>3</sup>Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, 117997, Russia

<sup>4</sup>Institute of Molecular and Industrial Biotechnology, Lodz University of Technology, Lodz, 90-573, Poland

Chymotrypsin (M234-K235)		NTD <sup>5H</sup> subdomain										Linker-1										α/β subdomain									
		250		260		270		280		290		300		310		320		330		340		350									
EcLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	DAPDENEALKRKIDAAKMPKEAKEKAEELOQLKLMSPMSAEATVVRGYIDWMVQVPWNARSKVKDLRQAQEI <sup>LD</sup> TDHYGLERVKDRILEYLAVQSRVNKIKGPIL																						352							
BsLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKTGEVQTLTEKIEEAGMPDHVKETALKELNRYEKIPSSSAESSVIRNYIDWLVALPWTDETDDKL <sup>LD</sup> KEAGRL <sup>LD</sup> DEHHGLEKVKERILEYLAVQKLT <sup>LD</sup> KSLKGPIL																						350							
BcLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKGGEVEELREKIEQSGMPEETMKAALKELDRYEKLVPSSAESGVIHNYIDWLLALPWTATTEDMIDLTHSEEILNKDHYGLEKVKERVLEYLAVQKLTNSLKGPIL																						352							
BhLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKTGEVASLREQIDEANMPENVEEKALKELERYEKMPASSAESSVLRYLDWLIQLPWWKETEDILDVNRAERILDE <sup>LD</sup> DHYGLEKVKERVIEYLAVQQLTKELKGPIL																						351							
BmLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKGGEVEELREKIEQSGMPEETMKAALKELDRYEKLVPSSAESGVIHNYIDWLLALPWTATTEDMIDLTHSEEILNKDHYGLEKVKERVLEYLAVQKLTNSLKGPIL																						352							
BpLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKTGEVSSLSMKIESSMPSVRE <sup>LD</sup> TALKELNRYEKIPSSSAESSVIRNYIDWLI <sup>LD</sup> GLPGWYITED <sup>LD</sup> RLDKIA <sup>SE</sup> ILDEHHGLEKVKERVLEYLAVQKLTNSLKGPIL																						350							
BbLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GRQGEVDEL <sup>LD</sup> RAGLEKSDAPERIKAKIEKELERLEKMPATSAEGSVIRTYIDTLFALPWTKT <sup>LD</sup> TEDNLDIHHAQQVLDE <sup>LD</sup> DHYGLDKPKERVLEYLAVQKL <sup>LD</sup> VNSMRGPIL																						345							
GtLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GKTGEVEALKEKIEAAGMPEHVKETALKELDRYEKIPATSAESAVIRNYLDWLI <sup>LD</sup> ALPWSKETKD <sup>LD</sup> IHDIKRAETILNEEHYGLDKVKERVLEFLSVQQLTQSLKGPIL																						356							
PiLonA	◀ N and NTD <sup>5H</sup> subdomains ▶	GRAGEVEELRSQLEQLP <sup>LD</sup> ERVHEKVEKIDRLEKMPASSAEGGVIRNYVDWLLALPWT <sup>LD</sup> KN <sup>LD</sup> TED <sup>LD</sup> DLIAKAEQVLDE <sup>LD</sup> DHYGLEKPKERVLEYLAVQKL <sup>LD</sup> VKKMKGPIL																						352							
		10		20		30		40		50		60		70		80		90													
BsLonBA-1	-----	MSWTGIALFIQLFFGIIIGLYFWNLLKNQRTQKVTIDKESKKEMEQLRKMRAISLSEPLSEKVRPKSFKDIVGQEDGIK-----ALKAALCGPNPQH <sup>LD</sup> V																						93							
BsLonBA-2	-----	MSWTGIALFIQLFFGIIIGLYFWNLLKNQRTQKVTIDKESKKEMEQLRKMRAISLSEPLSEKVRPKSFKDIVGQEDGIK-----ALKAALCGPNPQH <sup>LD</sup> V																						93							
BcLonBA	-----	MMSWTNIFLLVQLVFGVIGLYFWHLLRNQRTQKVSIDRESKKEQLRKMREISLTEPLAEKVRPTSF <sup>LD</sup> LDIVGQEDGIK-----SLKAALCGPNPQH <sup>LD</sup> V																						94							
BhLonBA	-----	MNNWTGILFVQLFFGVVIGLYFWNLLKTQ <sup>LD</sup> RSQRVSIDRESKKEQLRKLQ <sup>LD</sup> ISLSEPLSEKVRPESFEDIVGQKEGIR-----ALKAALCGPNPQH <sup>LD</sup> V																						93							
BmLonBA	-----	MSWTNIFLLVQLVFGVIGLYFWHLLRNQRTQKVSIDRESKKEQLRKMREISLTEPLAEKVRPTSF <sup>LD</sup> LDIVGQEDGIR-----SLKAALCGPNPQH <sup>LD</sup> V																						93							
BpLonBA	-----	MSWTGIVLFIQLFFGIVIGLYFWNLLRNQRTQKVTIDKESKKEMEALRKMRSIKLSEPLSERVRPKGFQDIVGQEDGIR-----ALKAALCGPNPQH <sup>LD</sup> V																						93							
BbLonBA	-----	MDYTTLVIAVIEVVVGIVIGTYFWNLLRAQRNTKTSTEKSRKDLDSIRKMRMVALTEPLSERTRPATLEEIVGQEDGLR-----ALRAALCGPNPQH <sup>LD</sup> V																						94							
GtLonBA	-----	MNNWTNIVLVQLFFGVVIGLYFWNLLKGQRVKVSIDRESKKEMEQLRKLRSILTEPLAEKVRPKSFDDIVGQEDGIK-----ALKAALCGPNPQH <sup>LD</sup> V																						93							
PiLonBA	-----	MNLISIVIMLVQLFFAVIIGIYFWNLLRGQRTSKTAVDRESRKELDKRLKRSVSLTKPLAEKTRPQTMDDIVGQKDGLR-----ALKAALCSANPQH <sup>LD</sup> V																						93							

		α/β subdomain																														
		Walker A						Pore loop-1				Walker B		Pore loop-2				RKH loop														
		360	370	380	390	400	410	420	430	440	450																					
EcLonA	CLV	GPPGVGKT	SLGQSI	AKAT	-----	GRKYVR	MALG	-----	GVRDEAE	IRG	-----	HRRTY	IGSMP	-----	GKLIQ	MAKV	GVKNP	LFLLDE	IDKMS	SDMRGD	PASALLE	VLVDPE	QNVA	FSDH	455							
BsLonA	CLAG	PPGVGKT	SLAKSI	AKSL	-----	GRKFVR	ISLG	-----	GVRDESE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	KAGKLN	PVFL	LLDE	IDKMS	SDFRGD	PSSAMLE	VLVDPE	QNSFS	SDH	453						
BcLonA	CLV	GPPGVGKT	SLARSI	ATSL	-----	NRNFVR	VSLG	-----	GVRDESE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	KAKTVN	PVFL	LLDE	IDKMS	NDFRGD	PAAALLE	VLVDPE	QNHNF	SDH	455						
BhLonA	CLAG	PPGVGKT	SLARS	VARAL	-----	NREFVR	ISLG	-----	GVRDEAE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	KAGSVN	PVFL	LLDE	IDKMA	NDFRGD	PAAALLE	VLVDPE	QNHFT	SDH	454						
BmLonA	CLV	GPPGVGKT	SLARSI	ATSL	-----	NRNFVR	VSLG	-----	GVRDESE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	KAKTVN	PVFL	LLDE	IDKMS	NDFRGD	PAAALLE	VLVDPE	QNHNF	SDH	455						
BpLonA	CLAG	PPGVGKT	SLAKSI	AKSL	-----	DRKFIR	VSLG	-----	GVRDESE	IRG	-----	HRRTY	VGAMP	-----	GRIIR	GMSK	AGTMNP	PVFL	LLDE	IDKMS	SDFRGD	PSSAMLE	VLVDPE	QNHNF	SDH	453						
BbLonA	CLV	GPPGVGKT	SLARS	VARAI	-----	GREFVR	ISLG	-----	GVRDEAE	IRG	-----	HRRTY	VGALP	-----	GRIIQ	GMMQ	AGTVNP	PVFL	LLDE	IDKL	ASDFRGD	PASALLE	VLDPN	QNDKF	SDH	454						
GtLonA	CLAG	PPGVGKT	SLARSI	AKAL	-----	GRRFVR	MSLG	-----	GVRDESE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	AGTINP	PVFL	LLDE	IDKMS	SDFRGD	PSSAMLE	VLVDPE	QNHFT	SDH	458						
PlLonA	CLV	GPPGVGKT	SLARSI	ARSL	-----	ERKFVR	ISLG	-----	GVRDEAE	IRG	-----	HRRTY	VGAMP	-----	GRIIQ	GMMK	TAGSLNP	PVFL	LLDE	IDKMS	ASDFRGD	PSSALLE	VLVDPE	QNTFT	SDH	454						
		100	110	120	130	140	150	160	170	180	190	200	210																			
BsLonBA-1	IVY	GPPGVGKT	AAARLV	LEEAK	HKQS	PFKEQ	AVFVEL	DATTAR	FDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKQG	-AVTHA	-HGG	--	VL	FIDE	EIGE	--	LH	--	PIQMNKML	KVLEDRK	VFLD	SA	213				
BsLonBA-2	IVY	GPPGVGKT	AAARLV	LEEAK	HKQS	PFKEQ	AVFVEL	DATTAR	FDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKQG	-AVTHA	-HGG	--	VL	FIDE	EIGE	--	LH	--	PIQMNKML	KVLEDRK	VFLD	SA	213				
BcLonBA	IIY	GPPGVGKT	AAARLV	LEEAK	RNP	KSPFRT	NATFI	ELDTAT	ARFDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKKG	-AVTDA	-HGG	--	IL	FIDE	EIGE	--	LH	--	PIQMNKML	KVLEDRK	VFL	SA	214				
BhLonBA	IIY	GPPGVGKT	AAARLV	LVEEAK	KNP	SPFAH	DAVFVELD	GT	TARFDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKQG	-AVTKA	-HGG	--	ML	FIDE	EIGE	--	LH	--	SIQINKLL	KVLEDRK	VFL	SA	213				
BmLonBA	IIY	GPPGVGKT	AAARLV	LEEAK	KNP	SPFRM	NATFI	ELDTAT	ARFDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKKG	-AVTDA	-HGG	--	IL	FIDE	EIGE	--	LH	--	PIQMNKML	KVLEDRK	VFL	SA	213				
BpLonBA	IVY	GPPGVGKT	AAARLV	MEEAK	RHADS	SPFKK	DAVFVELD	TAT	ARFDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKQG	-AVTHA	-HGG	--	VL	FIDE	EIGE	--	LH	--	PIQMNKML	KVLEDRK	VFLD	SA	213				
BbLonBA	IIY	GPPGVGKT	AAARLV	LEEAK	KNP	QSPFSS	DAKFIE	IDATTAR	FDERGI	ADPLIGSV	HDPIY	QAGS	SLGQAGIPQ	PKPG	-AVTKA	-HGG	--	ML	FIDE	EIGE	--	LH	--	PIQMNKLL	KVLEDRK	VFL	SA	214				
GtLonBA	IIY	GPPGVGKT	AAARLV	LEEAK	KNP	SPFKK	NAVFVELD	TAT	ARFDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKQG	-AVTNA	-HGG	--	VL	FIDE	EIGE	--	LH	--	PIQMNKLL	KVLEDRK	VFL	SA	213				
PlLonBA	IIY	GPPGVGKT	AAARLV	MEEAK	KNP	SPFKY	DAKFTIE	IDATTAR	FDERGI	ADPLIGSV	HDPIY	QAGAMG	QAGIPQ	PKPG	-AVTKA	-HGG	--	IL	FIDE	EIGE	--	LH	--	PIQMNKLL	KVLEDRK	VFL	SA	213				

← α/β subdomain   α subdomain →																	
RKH loop			SRH region														
			s1	RF	▼Chymotrypsin			s2									
460	470	480	490	500	510	520	530	540	550	560							
EcLonA	YLEVDYD-----	LSDVMFVATSN-----	S-MNI-PAPLLDRMEVIRLSGYT	EDEKLNIAKRHLHPKQIERN	ALKKGE	LTVD	DS	AI	IGIIRYYTREAGV	RLEREISKL	CRKAVKQLLLDKS	564					
BsLonA	YIETTFD-----	LSKVLFIATAN-----	NLATI-PGPLDRMEIINI	IAGYTEIEKLEIVKDHL	LPKQIKEHGLKKS	NLQLRD	QA	ILDIIRYYTREAGV	RS	LERQLAAICRKA	AKAIVAEER	563					
BcLonA	YIEEPYD-----	LSKVMFVATAN-----	TLATANTLASV-PGPLDRMEIISI	AGYTELEKVHIAREHLL	PKQLKEHGLRKG	NLQVRDE	ALLEIIRYYTREAGV	RTL	ERQIAKVCR	KA	AKIIVTAER	571					
BhLonA	YIEEPYD-----	LSKVMFITAN-----	NIMTI-PGPLDRMEIISI	AGYTEVEKQRIALDYLL	PKQMEHGLTKGT	FQVKEE	AILKVIRYYTREAGV	RS	LERQMATLC	CRGAKMLV	SGKK	564					
BmLonA	YIEEPYD-----	LSKVMFVATAN-----	TLASV-PGPLDRMEIISI	AGYTELEKVHIAREHLL	PKQLKEHGLRKG	NLQVRDE	ALLEIIRYYTREAGV	RTL	ERQIAKVCR	KA	AKIIVTAER	565					
BpLonA	YIETTFD-----	LSQVLFIATAN-----	NLATI-PGPLDRMEIITI	AGYTEVEKAEIVKDHL	LPKQLKEHGLKKN	LQLEAAIYDTIRYYTREAGV	RL	ERQLAAICRKA	ARAIVAE	DR		563					
BbLonA	YIEETYD-----	LTNVMFITAN-----	SLDTI-PRPLDRMEVIS	ISSYTELEKLNILRDYLL	PKQM	QDHGLGDKLKMND	DAMKLVLRYTREAGV	RNL	NREANVCR	KA	AKLIVGGEK	564					
GtLonA	YIEEPYD-----	LSKVMFIATAN-----	NLATI-PQPLDRMEIITH	IPGYTEVEKLHIAKRHL	LPKQLAEHGLKKA	ALQVRDDAMLDIIRHYTREAGV	RE	LERQLAAICRKA	ARLIVS	GEK		569					
PILonA	FIEIPFD-----	LSQVMFVTAN-----	ALHNI-PRPLDRMEVLY	IPGYTEIEKLQIANRYLL	LPQKSEHGLEPE	QLQVDEGALLKVIREYTRESGVR	NLE	QQVAALCRKA	AKRIVT	NES		565					
▼Chymotrypsin																	
220	230	240	250	260	270	280	290	300	310	320							
BsLonBA-1	YYSEENTQIPNHIH	DIFQNGLPAD	FRILGATT	-----RMPNEIPP	AIRSR	CLEVF	FRELEKDELKT	VAKTAADKI	-----EKNIS	-----EEGLDL	LLTSY-TRN--GREAVNMIQIAAGMAVTE---NRK	322					
BsLonBA-2	YYSEENTQIPNHIH	DIFQNGLPAD	FRILGATT	-----RM	PNEIPP	AIRSR	CLEVF	FRELEKDELKT	VAKTAADKI	-----EKNIS	-----EEGLDL	LLTSY-TRN--GREAVNMIQIAAGMAVTE---NRK	322				
BcLonBA	YYSEENTMIPTIYI	HDIFQKGLPAD	FRILGATT	-----RSPEEIPPAIRS	CLEVF	FRLEDT	EEIQK	VAKNAADKV	-----EMQIG	-----ENGIE	MIGMY-ARN--GREAINLVQISAGMAINE---ERS	330					
BhLonBA	YYSEENQQIPRHIH	EIFQKGLPAD	FRILGATT	-----RMPEEIPPAIRS	CLEVF	FRSLK	KREEIVTI	AKAVNRL	-----GFKAE	-----EGVYE	KISAY-ATN--GRET	VNIVQIAAGMATTE---GRH	322				
BmLonBA	YYSEENS	MIPTIYIHDIFQKGLPAD	FRILGATT	-----RSPDEIPPAIRS	CLEVF	FRLEDT	EEIQK	VAKNAAEKV	-----KMEIG	-----KQG	VE	MIGMY-ARN--GREAINLVQISAGMAINE---GRS	322				
BpLonBA	YYSEENTQIPNHIH	DIFQNGLPAD	FRILGATT	-----RTPDEIPPAIRS	CLEIF	FKDL	DQHEL	I	VA	AKAVQKI	-----QKELS	-----DEGL	NLLTSY-VKN--GREAVNMMVQIAAGMAVTE---NRK	322			
BbLonBA	YYSEENNQIPSHIH	DVFKYGLPAD	FRILGATT	-----RLPEELPAALRS	CLEIF	FRPLK	AGEIG	CI	VRT	AVPKM	-----NMNIE	-----DS	AVSVIERY-ATN--GREAINLVQISAGMAINE---ERQ	323			
GtLonBA	YYSKENQIPSHIH	DIFQNGLPAD	FRILGATT	-----RTPNEIPPAIRS	CLEVF	FRLEDT	QDE	I	AL	IAKRAAEKI	-----RLNVS	-----ES	GIRLLAAY-ARN--GREAVNMMQIAAGLAITE---NRE	322			
PILonBA	YNSSEDANTPAYI	HDIFQNGLPAD	FRILGATT	-----RSPEEVAPALRS	CMEIF	FRPL	P	EEIG	KIA	EDAIQKI	-----GLGAA	-----PE	AVDVVKQY-CSN--GREAVNMIQIAAGLAITE---KRD	322			
▼Chymotrypsin																	
← α subdomain   Linker-2 →   P domain →																	
570	580	590	600	610	620	630	640	650	660	670							
EcLonA	LKHIEINGDN	LHDYLG---	VQRF	DYGRADNENRV	QGV	TGLAWTEVGG--	DLT	TIETACVP	KGK---	LYT	TGSL-----GEVMQESIQAA	LTVVRAAEKLG	INPDFYEKRD	IHVHVP	PEG	671	
BsLonA	-KRITVTEKN	LQDFIG---	KRIF	RYGQAE	TDQVG	VVTGLAYTTVGG--	DTLS	IEVSLSPGK	GK---	LILT	TGKL-----GDVMRESAQAA	FSYVRSKTEELGI	EPDFHEKYD	IHIHVP	PEG	669	
BcLonA	-KRIVVTEKN	IVDLLG---	KHIF	RYGQAE	KTQVG	MATGLAYTAAGG--	DTLA	IEVSVAPG	GK---	LILT	TGKL-----GDVMKESAQAA	FSYIRSRAEELH	IDPDFHEKND	IHIHVP	PEG	677	
BhLonA	-KRIVVVTQNT	ITETMLG---	KKRF	RYGLAE	EQDQV	GAAATGLAYTTAGG--	DTLS	IEVSVVP	GK---	LILT	TGKL-----GDVMKESAQAA	FSYIRSRSKELD	IDPSFHEKND	IHIHVP	PEG	670	
BmLonA	-KRIVVTEKN	IVDLLG---	KHIF	RYGQAE	KTQVG	MATGLAYTAAGG--	DTLA	IEVSVAPG	GK---	LILT	TGKL-----GDVMKESAQAA	FSYIRSRAEELH	IDPDFHEKND	IHIHVP	PEG	671	
BpLonA	-KRITVTEKN	LSEFLG---	KRLY	RYGQAE	TTDQV	GVVTGLAYTTVGG--	DTLS	IEVSLSPG	GK---	LILT	TGKL-----GDVMRESAQAA	FSYIRSKADELN	IDPNFNEKHD	IHIHVP	PEG	669	
BbLonA	-KRIVVTVKT	LETLLG---	KPRY	RYGLAE	IKDQV	GSVTGLAWTQAGG--	DTLN	VEVSILP	GK---	LILT	TGKL-----GDVMKESAQAA	FSYIRSRAEQW	IDPAFHEKND	IHIHVP	PEG	670	
GtLonA	-KRIVVVTQNN	IEEFLG---	KRKY	RYGQAE	EDQIG	VATGLAYTAFGG--	DTLA	IEVSLAK	NGK---	LILT	TGKL-----GDVMKESAQAA	FSYVRSRAEELG	IDPKFHETYD	IHIHVP	PEG	675	
PILonA	-DSIVIGADDI	KDYL---	IPKF	RYGMAE	LEDQV	GTVTGLAWTEVGG--	ETLT	IEVTVP	PGTGK---	LILT	TGKL-----GDVMKESAQAA	FSFTRSKAELG	IDPDFHEKLD	VHIHVP	PEG	671	
▼Chymotrypsin																	
340	350	360	370	380	390	400	410	420	430								
BsLonBA-1	DIT-----	IEDIEWIHSSQL	TPKH	EQKIG	VEPQVG	IVNGLAVY	GPNSGSLLEIEVSVTA	AQD--	KGSINITGIAE	EEESIGSQSKS	IRRKSMAGKS	SVENVLT	VLRTMG-----MKPSDYD	IHINFP-G	437		
BsLonBA-2	DIT-----	IEDIEWIHSSQL	TPKH	EQKIG	VEPQVG	IVNGLAVY	GPNSGSLLEIEVSVTA	AQD--	KGSINITGIAE	EEESIGSQSKS	IRRKSMAGKS	SVENVLT	VLRTMG-----MKPSDYD	IHINFP-G	437		
BsLonBA-3D					GVEPQVG	IVNGLAVY	GPNSGSLLEIEVSVTA	AQD--	KGSINITGIAE	EEESIGSQSKS	IRRKSMAGKS	SVENVLT	VLRTMG-----MKPSDYD	IHINFP-G	437		
BcLonBA	FIK-----	DEDIEWVHSSQL	TPKYE	KHIYPIPRIG	LVNGLAVY	GPNTGALLEIEVTA	IAKAKD--	IGSVNVTG	IV	EEESIGSQTKS	IRRKSMAGKS	VDNVL	RVLRSLD-----VLPEGYD	IHINFP-G	438		
BhLonBA	EVT-----	VADVEWVIHSSQL	SPRPERQIP	SEP	KVGLVNG	LAVFGPNTGALLEIEV	VDVIPSVA--	EGSVTITG	IAE	EEETGNQARS	IRRKSMAGKS	SVENVLT	TVLKKMD-----LPIDQYD	IHINFP-G	437		
BmLonBA	FIK-----	DEDIEWVHSSQL	TPKYE	KRIYPIPKIG	LVNGLAVY	GPNTGALLEIEVMA	IPAAE--	KGTINVTG	IV	EEESIGSQTKS	IRRKSMAGKS	VDNVL	TVLRSLD-----LLPEGYD	IHINFP-G	437		
BpLonBA	DIT-----	IEDVEWVIHSSQL	TPRYEQ	KVPKPGV	GVNGLAVY	GPNSGALLEIEVNI	CKALE--	KGSINITGIAE	EEESIGNQSKS	IRRKSMAGKS	SVENVLT	TVLRNLG-----IRTNEYD	IHINFP-G	437			
BbLonBA	DIQ-----	AADVEWVMHSSQ	KSPREKQ	VHDT	TPQVGLV	NGLAVYGP	NMGSM	LEV	TASPA	AVPGQGRMAMT	GMAEEEMG	SR	SRTIRRKSMAGKS	IENVLT	TVLSRMG-----VRPYD	YDLHINFP-G	440
GtLonBA	KIL-----	DKDIEWVIHSSQ	MAPRYE	KKIGSAP	AVGVNGLAVY	GPNTGALLEIEV	TALPAK--	KGSINVTG	IV	EEESIGSPEKS	VRRKSMARG	SAENVIT	TVLRAMG-----VPADRYD	IHINFP-G	434		
PILonBA	SIR-----	ASDVEWVASSQ	IQPRPDR	KVPEK	PQIGV	VNGLAVYGP	NMGALLEIEV	SAVPVEK--	QGQTYNIT	GV	VEEELGGGSR	KRLRRKSMAGKS	IENVLT	TVLEAMG-----FSPKDYD	YDLHINFP-G	435	

		*680	690	700	710	*	730	740	750	760	770	780	← P domain									
EcLonA		ATPKD	GPSAGIAMCTALV	SCLTGNPVRAD	VAMTGEITL	RGQVLPIGGLKEKLLAAHRG	GIKTVLIPFENKRDLEE	IPDNVIADLDIHPVKRIEEVLT	TALQNEPSGMQVVTAK	-----	-----	-----	784									
BsLonA		AVPKD	GPSAGITMATALV	SALTGRAVSR	EVGMTGEITL	RGRVLP	PIGGLKEKALGAHRAGL	TTIIPAPKDNEKDIED	IPESVREGLTFILASHLDEVLEHALVGEKK	-----	-----	-----	774									
BcLonA		AVPKD	GPSAGITMATALV	SALTGIPVSKEV	GMTGEITL	RGRVLP	PIGGLKEKTL	SAHRAGLTKIILPAENEKDLDD	IPESVKENLTFVLASHLDEVLEHALVGKQ	-----	-----	-----	782									
BhLonA		ATPKD	GPSAGITMATALV	SALTGRVVKKEV	GMTGEITL	RGRVLP	PIGGLKEKMS	SAHRAGLTTIIPKENEKDLDD	IPKSVREDLNLYILVDHLDQVLKHALGEKK	-----	-----	-----	774									
BmLonA		AVPKD	GPSAGITMATALV	SALTGIPVSKEV	GMTGEITL	RGRVLP	PIGGLKEKTL	SAHRAGLTKIILPAENEKDLDD	IPESVKENLTFVLASHLDEVLEHALVGKQ	-----	-----	-----	776									
BpLonA		AVPKD	GPSAGITMATALV	SALTGRPVSK	EVGMTGEITL	RGRVLP	PIGGLKEKAL	GAHRAGLTKIILPKDNEKDIDD	IPESVREGLTFIPVSHLDEVLEKALVGEGR	-----	-----	-----	774									
BbLonA		AIPKD	GPSAGITMATALV	SALTGKIPVKKEV	GMTGEITL	RGRVLP	PIGGLKEKMS	SAHRAGLTTIILPKDNEKDIED	IPESVREGLTFYPVDHLDVLRHALTKQFPVGDKA	-----	-----	-----	779									
GtLonA		AVPKD	GPSAGITMATALV	SALTGKPVSRF	VGMTGEITL	RGRVLP	PIGGLKEKTL	SAHRAGLTKVILPKDNEKDLAD	IPDTPVKRDLRFVLVSHLDEVLP	PHALVGWER	-----	-----	780									
PILonA		AIPKD	GPSAGITMATALV	SSLT	KRHVAR	DVAMTGEITL	RGRVLP	PIGGLKEKSLAAHRAGY	KKILLPKDNERDLRDI	IPDSVKNDVEFVPSVSHMDQVLEHALVEQAGVH	-----	-----	778									
		440	*	450	460	470	480	*	490	500	510	520	530	540	550							
BsLonBA-1		GIPID	GPSAGIAMAGIFSAI	HKIPIDNT	VAMTGEIS	LNGLV	KPIGGV	IPKIKAAKQ	SGAKKVIIPYENQQAI	---	LKQIDG	-IEII	AVKTFQEV	LDEILVNP	PTEQKPFHI-E	---	INKE	---	SV	552		
BsLonBA-2		GIPID	GPSAGIAMAGIFSAI	HKIPIDNT	VAMTGEIS	LNGLV	KPIGGV	IPKIKAAKQ	SGAKKVIIPYENQQAI	---	LKQIDG	-IEII	AVKTFQEV	LDEILVNP	PTEQKPFHI-E	---	INKE	---	SV	552		
BsLonBA-3D		GIPID	GPSAGIAMAGIFSAI	HKIPIDNT	VAMTGEIS	LNGLV	KPIGGV	IPKIKAAKQ	SGAKKVIIPYENQQAI	---	LKQIDG	-IEII	AVKTFQEV	LDEILVNP	PTEQKPFHI-E	---	INKE	---	SV	552		
BcLonBA		GIPID	GPSAGIAMATGV	YSAVHHHTY	VNNEVAMT	GEISIH	GEVVKPIGGV	YAKIKAAKQ	AGAKKVIIPYENQQAI	---	LYTIKG	-IEII	PVRKLKEV	FELTFM	QENMHRELDI	HTTI--	DETD	QAQ	---	SM	557	
BhLonBA		GAPVD	GPSAGVAIATGI	YSAIKEEK	VNDNTVAMT	GEISIH	GTVPVGGV	IAKVEAAKQ	AGATRV	LIPKENEQAI	---	LQIEG	-ITII	PMKHLDDV	LSYALIRE	EETEEETIP	ANQSFVSAP--	A	---	PV	556	
BmLonBA		GIPID	GPSAGIAMATGV	YSALHQVYV	NNEIAMT	GEISIH	GEVVKPIGGV	YAKIKAAKQ	AGAKKVIIPYENQQAI	---	LYTIKG	-IEII	PVRKLKEV	FALTFV	QEKHREIDV	HREL--	DEGD	VQ	---	SM	556	
BpLonBA		GAPID	GPSAGIAMATAI	FSAIHQIP	IDHLTAMT	GEIGLQ	QGVKPIGGV	IPKIKAAKQ	AGATRV	IIPYDNQ	QSI--	LKEIEG	-VTIV	PVKHFQEV	LDQCLV	NPPDEK	KGLEEE--	MKRK	---	SV	553	
BbLonBA		GIPVD	GPSAGITMATALV	SALTGIPVSKEV	GMTGEITL	RGRVLP	PIGGLKEKTL	SAHRAGLTKIILPAENEKDLDD	IPESVKENLTFVLASHLDEVLEHALVGKQ	-----	-----	-----	782									
GtLonBA		GVPVD	GPSAGVAIATGI	YSAIQVLP	VNDNTVAMT	GEISIH	GVKVPVGGV	FAKIKAAKQ	AGAKKVIIPYENQQAI	---	LREVS	G-IQII	AVRRLE	EALIHV	FGEETLRR	GTAFLP	AAAAD	RSGKK	---	LV	558	
PILonBA		GTPID	GPSAGIAMATAI	VSAIKGIE	VNDNKIAMT	GEMSIH	GKVPVGGV	IAKVEAAKQ	AGATRV	LIPKDNWQSL	---	FQDLSG	GLRVI	PVETV	DEVFRQV	FGGSLA	DIVELP	NTDIY	PASSASLL	14	GV	576

**Figure S1. Alignment of the primary and available or predicted secondary structures of *EcLonA* and bacillary *LonA* and *LonBA* proteases.**

*LonAs* and *LonBAs* are from: *Escherichia coli* (*EcLonA*, MER0000485, UniProt P0A9M0); *Bacillus subtilis* (*BsLonA*, MER0000487, UniProt P37945 and *BsLonBA*, MER0002228, UniProt G4ESU3); *B. cereus* (*BcLonA*, MER0029007, UniProt J7YLY7 and *BcLonBA*, MER0029009, UniProt C2NNG6); *B. halodurans* (*BhLonA*, MER0013686, UniProt Q9K8F6 and *BhLonBA*, MER0013716, UniProt Q9K8F5); *B. mycoides* (*BmLonA*, MER0231248, UniProt A0A084IV40 and *BmLonBA*, MER0399002, UniProt C3AR76); *B. pumilus* (*BpLonA*, MER0109288, UniProt A8FFV7 and *BpLonBA*, MER0164590, UniProt A0A8G1JA29); *Brevibacillus brevis* (*BbLonA*, MER0000486, UniProt C0ZAG5 and *BbLonBA*, MER0166780, UniProt C0ZAG4); *Geobacillus thermodenitrificans* (*GtLonA*, MER0090423, UniProt A4IRH0 and *GtLonBA*, MER0090424, UniProt A4IRH1); *Paenibacillus lactis* (*PILonA*, MER0299899, UniProt G4HHX1 and *PILonBA*, MER0361527, UniProt G4HHX2).

Domain organization is shown for *LonAs*. Primary structures were compared (by fragments) using the program <https://www.ebi.ac.uk/Tools/msa/clustalo/>, fragment boundaries were determined according to the secondary structures. Secondary structures of *BsLonBA-1* and *BsLonBA-2* were predicted using the programs <http://distilldeep.ucd.ie/porter/> and <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>, correspondingly. Secondary structure of *BsLonBA* P domain (*BsLonBA-3D*) was determined in this work. Red color indicates amino acids that form  $\alpha$ -helices, purple – 3/10 helices, blue –  $\beta$ -strands; amino acids marked in black are not included in the secondary structure elements. Enzyme consensus elements are highlighted in yellow: Walker motifs A and B, Pore loop-1 with the conserved Tyr residue, Pore loop-2, RKH loop, SRH region, the residues sensor-1 (s1), sensor-2 (s2), and Arg finger (RF), catalytic residues Ser and Lys and their environment. Amino acid numbering is shown for the sequences of *EcLonA* and *BsLonBA*. ▼ – sites of limited proteolysis by chymotrypsin of *EcLonA* and *BsLonBA*.



		380	390	400	410	420	430	440	450	460	470		
BsLonBA	374	SVTAAQD	---KGSINITGIAEESIGSQSK	SIRRKS	MAKGSVENVLTVLRT	MG	-----MKPSDYDIHINFP	GGIPID	GPSAGIAMAAGIFS	AIHKIPIDNT	VAMTG	471	
		470	480		490	500	510	520	530	540			
TonLonB	461	AVVAPAA	SKKEGKIIVT	GKLG	-----EIAKEAVQNVSAI	IKRYKGE	-----DISRYDIHVQFLQTYEGVE	GDSASISV	ATAVISALEGIP	IRQDVAMTG	549		
AfLonB	447	AEVTPSM	SKSEGRVIATGRL	Q	-----EIAREAVMNVS	AIKKYTGR	-----DISNMDVHIQFVGTYEGVE	GDSASISI	ATAVISALEGIP	VDQSVAMTG	535		
MjLonB	479	IIVQILE	SKNPGTHLLN	-ISG	-----DIAKHSITLASALS	SKKLVAEKKLPLPKKDIDLNNKEI	YIQFSQSYSKID	GDSATA	AAVCLAIISALLDIP	LKQDFAITG	576		
		520	530		540	550	560	570	580	590	600		
MtaLonC	517	LTARAAP	GRDHL-ISIDRE	AGLGG	-----QIFHKAVLTL	AGYLRSRYIE	-----HGSLPVTISLAF	EQNYVSIE	GDSAGLAELVA	ALSAIGNLPLRQD	LAVTG	608	
		480	490	500	510	520	530	540	550	← P domain			
BsLonBA	472	EISLNGLV	KPIGGV	VIPKIK	-----AAQSGAK	KVIIPYEN	QQ	-----AI	-----LKQIDGIE	IIIAVKTFQEV	LDEILVNP	PTEQKPFHIEINKESV	552
		550	560	570	580	590	600	610	620	630			
TonLonB	550	SLSVRGE	VLPIGGAT	PKIE	-----AAIEAGIK	MVVIIPKSNEK	-----DVFLSKDKAEKI	QIFPVETIDEV	LEIALEESEK	KRELLRRIRET	TLPLSL		635
AfLonB	536	SLSVKGE	VLPGGV	TQKIE	-----AAIQAGL	KKVVIIPKDNID	-----DVLLDAEHGKI	EVIPVSRINEV	LEHVLEDG	KKKNRLMSK	FKELELAAV		621
MjLonB	577	SIDL	SGNVLAIGGV	NEKIE	-----AAKRYGFK	RVIIPEANMI	-----DVIETEG	-----IEIIPVKT	LDEIVPLV	FDL	DNRRGGAERFN		649
		610	620	630	640	650	660	670	680	690	700	710	
MtaLonC	609	AVDQTGK	VLAVGAI	INAKVEG	FFRVCKALGLSGTQ	GVILPEANLANLT	LRAEVLEAVRAGQF	HIYAVETAEO	QALEILAGARME	GFRGLQEKIR	RAGLEAFARLEEGHDKEDRE		719

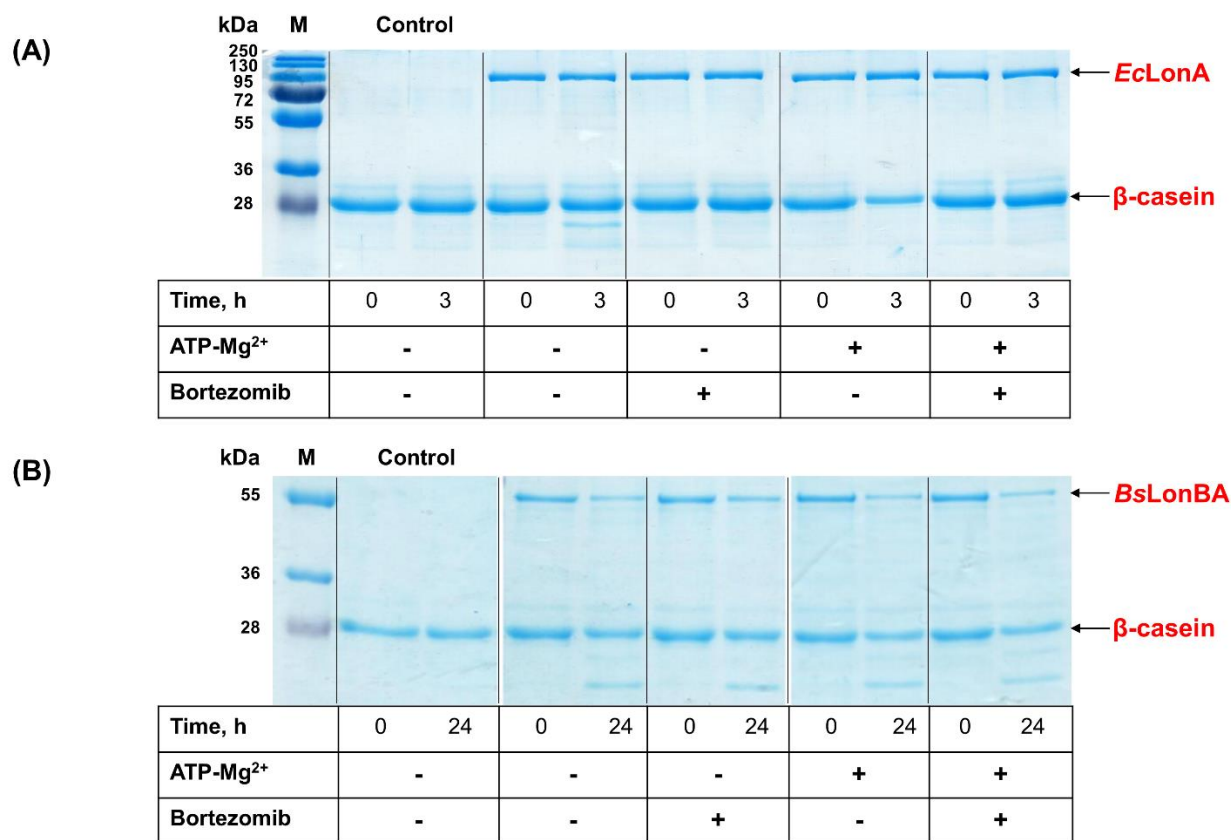
**Figure S2. Primary and secondary structure alignment of *BsLonBA*, LonBs and *MtaLonC* proteases based on experimental structures**

Enzyme sources: *Bacillus subtilis* – *BsLonBA* (MER0002228, UNIPROT P42425); *Thermococcus onnurineus* – *TonLonB* (MER0145010, UNIPROT B6YU74, PDB: 3K1J); *Archaeoglobus fulgidus* – *AfLonB* (MER0003868, UNIPROT O29883, PDB: 1Z0W/B,C,E,G,T,V); *Methanocaldococcus jannaschii* – *MjLonB* (MER0003359, UNIPROT Q58812, PDB: 1XHK); *Meiothermus taiwanensis* – *MtaLonC* (MER0423045, UNIPROT C9DRU9, PDB: 4FW9/D,G,H,V).

Predicted fragments of structures are shown in the standard font; experimentally determined structures and/or their fragments are highlighted in bold; sequence fragment of *MtaLonC* not seen in the crystal structure is italicized. Amino acid numbering is shown for the sequences of *BsLonBA*, *TonLonB*, and *MtaLonC*.

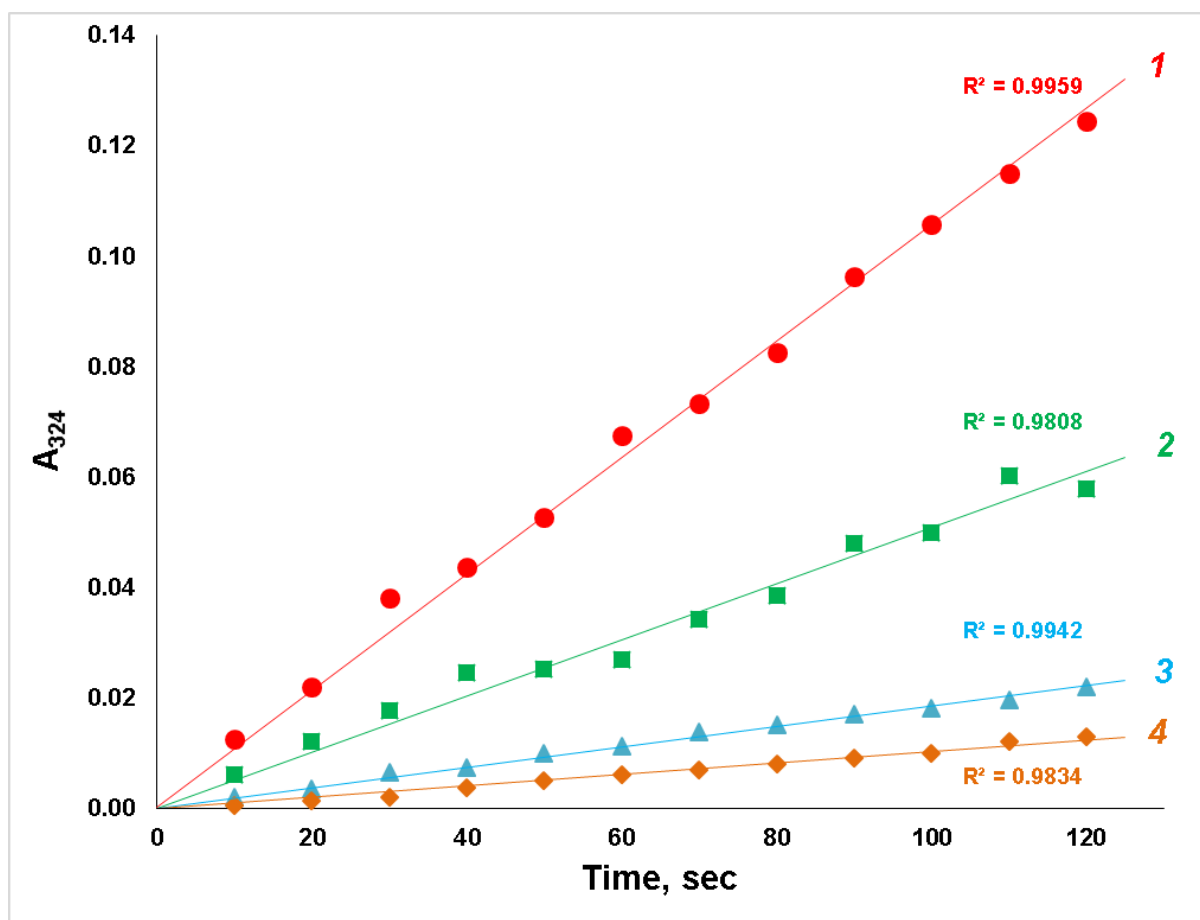
See all designations in the caption to Fig. S1. Substitutions of amino acid residues in Walker motifs are highlighted in gray.





**Figure S3. Hydrolysis of  $\beta$ -casein by *EcLonA*-H6 (A) and *BsLonBA*-H6 (B).**

*Experimental conditions:* 15% PAGE, 50 mM Tris pH 8.3, 200 mM NaCl, 8 mM CHAPS buffer, 1 mM ATP, 10 mM MgCl<sub>2</sub>, 7.0 (A) or 5.0 (B)  $\mu$ M  $\beta$ -casein, 1.5  $\mu$ M *EcLonA*-H6, 3.0  $\mu$ M *BsLonBA*-H6, 35 (A) or 70 (B)  $\mu$ M bortezomib; 30 °C. Preincubation of the enzymes with bortezomib was carried out during 1 hour at 25 °C. M, molecular weight markers.



**Figure S4. Hydrolysis of PepTBE by *EcLonA*-H6 (1, 2) and *BsLonBA*-H6 (3, 4).**  
*Experimental conditions:* 50 mM Tris pH 8.3, 200 mM NaCl, 10% DMSO, 8 mM CHAPS buffer, 0.1 mM PepTBE, 0.2 mM DTDP, 0.5  $\mu$ M enzymes, 0.05 (2) or 1.0 (4)  $\mu$ M bortezomib; 30°C. The reactions were initiated by the addition of the substrate.