

(12) Patent Application Publication
Chen et al.

(43) **Pub. Date:** **Aug. 14, 2025**

Publication Classification

(57) **ABSTRACT**

Specification includes a Sequence Listing.

(63) Continuation of application No. 17/936,744, filed on Sep. 29, 2022.

[illegible]

	20	40	60	80
Tgo	M			
Kod1	M			
9 ^N	M			
Pfu	M			
Vent	M			
Mma	M			
Mec	M			
hPOLD	MDGKRRPGG	GLWDDADPR	PSQF	
ScePOLD	MSEKRSI	PMVDVKI		
Pis	MELKWWP			
Seo	MTKQLTLFDI	PSSKPAKSEQ	NT	
Pae	MELL			
Eco	MAQA			
RB69	MKE			
T4	MKE			
Phi29	MK			
Consensus	M			
	100	120	140	160
Tgo				
Kod1				
9 ^N				
Pfu				
Vent				
Mma				
Mec				
hPOLD	QEQEEELQS	VLEGVADGQV	FPSAIDPRWL	RPTPALDPQ
ScePOLD	ESTFEQELSQ	MEHDMADQEE	HDLSSEFERKK	LPTD
Pis				
Sso				
Pae				
Eco				
RB69				
T4				
Phi29				
Consensus				
	100	120	140	160
Tgo				
Kod1				
9 ^N				
Pfu				
Vent				
Mma				
Mec				
hPOLD	QEQEEELQS	VLEGVADGQV	FPSAIDPRWL	RPTPALDPQ
ScePOLD	ESTFEQELSQ	MEHDMADQEE	HDLSSEFERKK	LPTD
Pis				
Sso				
Pae				
Eco				
RB69				
T4				
Phi29				
Consensus				

FIG.1

	180	200	220	240	
Tgo	KENGFEKIDY DRNFEPIYA LL	-SAIEDVKKI TAERH--G-	-TTVRVVRAE KVKKKF--LG	----RPIEVW 83	
Kod1	KENGFEKIEY DRTFEPYFYA LL	-SAIEEVKKI TAERH--G-	-TVVTYKRAE KVQKKF--LG	----RPVEVW 83	
9 ^N	KENGFEKIEY DRTFEPYFYA LL	-SAIEDVKKV TAKRH--G-	-TVVAVKRAE KVQKKF--LG	----RPIEVW 83	
Ptu	KENGFEKIEH DRTFRPIYA LL	-SKIEEVKKI TGERH--G-	-KIVRIVDVE KVEKKF--LG	----KPITVW 83	
Vent	KENGFEKIEL DPHFQPIYA LL	-SAIEEIKAI KGERH--G-	-KTVRVLDVAV KVRKKF--LG	----REVEVW 83	
Mma	S-----IIK EKDFKPYFYV NS	-EQILEFLKO YEKHKHKLDS-	-EISKMIENI ETVKKI--VF	DENYQEKELS 86	
Mac	GADGKSVCCE VPDFEPYFYL KA	-LHAVARLIKD TFEQ-----	-----VKKV EIVEKFEPVG	YQKTKK-EML 89	
hPOLD	TDEGFSVCCCH IHGFAPYFYT P	MGDLQRELNI AISRDSRQGR	ELTGPVAVLAV ELCSRESMFG	YHGHGSPFPL 210	
ScePOLD	TSEGHSVLCN VTGFKNYLYV PAP	AND-QEQINK FVHYLN--	ETFDHAIDS EVVSKQSIWG	YSGDTKLFPFW 214	
Pis	ISSGERVLI DRSFKPYFYV	CAVCEPAALK TA-----	SRVAPIDDDVQ IVERRF--LG	----RSKKFL 84	
Sao	ETQKIYALYD NTGCHKPYFLV	DLE-----PDK VGPPIKIVRD	PSFDH-----	ETVSKIDPYT W--NKFKLT 131	
Pae	TPAGT-----	---EVAFWL ATEQGGPROVR	LPPOPSVAFV LAEQGR--	---VESLLAG ETGAELRPLA	LRFQQRPLV 80
Eco	TPQGT-----	---EVSFWL ATONGPLOVT	LAPQESVAFI PADQVPR--	---ACHILQG EQGFRLTPLA	LKDFHRQPVY 79
RB69	DSNGR-----	ERTREVEYKPS SLFAHCPEEQ	ATKYFDIYGK PCIRK----	-LFANMRDAS QWIKRMEDIG	LEALGMDQFK 89
T4	DENSK-----	ERTREVEYLP TMFRHCKEE-	-SKYKDIYGK NCAPQ-----	-KFPSPMKDAR DWMKRMEDIG	LEALGMNDFK 87
Phi29	-----	-----	-----	-----	---HMPRKMV 9
Consensus	KENGE--I-- DRTFEPYFYA LL	-----KDD -SAIEEVKKI TAERH-----	---VRVVDAAE EVVKKF--LG	----RPIEVW	
	260	280	300	320	
Tgo	KLYFTHPODV PAIRDKIKEH PAVV	DI--YEYDIP FAKRYLIDKG LIP	-----M EGD-----	----- 132	
Kod1	KLYFTHPODV PAIRDKIREH PAVI	DI--YEYDIP FAKRYLIDKG LVP	-----M EGD-----	----- 132	
9 ^N	KLYFNHPQDV PAIRDRIKRAH PAVV	DI--YEYDIP FAKRYLIDKG LIP	-----M EGD-----	----- 132	
Ptu	KLYLEHPQDV PTIREKVRERH PAVV	DI--FEYDIP FAKRYLIDKG LIP	-----M EGE-----	----- 132	
Vent	KLYFEHPQDV PAMRGKIREH PAVV	DI--FEYDIP FAKRYLIDKG LIP	-----M EGD-----	----- 132	
Mma	KVTVKYPNNV KTVREILMEF ER	---L--YEYDIP FVRRYLIDNS	VIP-TSTWDF ENN-----	----- 137	
Mac	RVTTRLPKDV PEIRDEILKI RDVLR	AEQDW QV--YESDIL FRNRLIDRA	LGG-MVWVSA EGKPVDPVRY	LGAGSAWRSR 186	
hPOLD	RITVALPRLV APARRLEQG -IRV	AGLSTP SFAPYEANVD FEIRFMVDT	IVG-ONWLEL PAGKYAL-RL	KEKATQCQLE 287	
ScePOLD	KIIVTYPHMV NKLRTAFERG HLS	FNWSFN GTTTYD-NIA YTLRLMVDCG	IVG-MSWITL PKGKYSMIEP	NNRVSSCQLE 282	
Pis	KVIKAKIPEDV RKLREAAMS I	PRV-----	-SGVYEADIR FYMRYMIDMG	VVP-CSWNVA EVEEGORLO-	----- 144
Sao	KIVVROPLAV RRLRNDVPK -	-----	---AYEAHIK YFNNYMYDIO	LIPGMPYVVK NGKLESVYLS	LDEKDVVEEK 187
Pae	GLYCCQHRQL MNLEKRLRQA	-----	GVEFEADIR PPERYLMERF	ITAP---VSL EASVEADGSL	LAR----- 140
Eco	GLYCAHRLQL MNYEKRLREG	-----	GVTVEADVR PPERYLMERF	ITSP---VWV EGDH-NGTI	VNA----- 138
RB69	LAYLS-----	-----	-DTYNVEIK YDHTKI-----	-----	----- 108
T4	LAYS-----	-----	-DTYQSEIV YDRKFKV-----	-----	----- 108
Phi29	SODFETITKV EDCR-----	-----	-VWAYG-----	-----	----- 29
Consensus	KLYFXHPQDV PAIRDKIREH	--V----- -I--YEYDIP FAKRYLIDKG LIP	-----M EGD-----	-----	

FIG.1(continued)

	340	360	380	400
Tgo	EE	L	LA	LA
Kod1	EE	EE	EE	EE
9'N	EE	EE	EE	EE
Pfu	EE	EE	EE	EE
Vent	EE	EE	EE	EE
Mna	EE	EE	EE	EE
Mec	EE	EE	EE	EE
hPOLD	EE	EE	EE	EE
ScoPOLD	EE	EE	EE	EE
Pis	EE	EE	EE	EE
Sao	EE	EE	EE	EE
Pae	EE	EE	EE	EE
Eco	EE	EE	EE	EE
RB69	EE	EE	EE	EE
T4	EE	EE	EE	EE
Phi29	EE	EE	EE	EE
Consensus	EE	EE	EE	EE
	420	440	460	480
Tgo	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Kod1	DEEGAR	DEEGAR	DEEGAR	DEEGAR
9'N	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Pfu	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Vent	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Mna	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Mec	DEEGAR	DEEGAR	DEEGAR	DEEGAR
hPOLD	DEEGAR	DEEGAR	DEEGAR	DEEGAR
ScoPOLD	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Pis	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Sao	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Pae	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Eco	DEEGAR	DEEGAR	DEEGAR	DEEGAR
RB69	DEEGAR	DEEGAR	DEEGAR	DEEGAR
T4	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Phi29	DEEGAR	DEEGAR	DEEGAR	DEEGAR
Consensus	DEEGAR	DEEGAR	DEEGAR	DEEGAR

FIG.1(continued)

		500	520	540	560
Tgo	-GVKFI LGR-	-EGS	GDRFAVEV-	DLYPVIRRTI	PTYTLEAVYE
Kod1	-GINFALGR-	-DGS	GDRFAVEV-	DLYPVIRRTI	PTYTLEAVYE
9 th	-GIKFTLGR-	-DGS	GDRFAVEV-	DLYPVIRRTI	PTYTLEAVYE
Pfu	-GKLT LGR-	-DGS	GDMTAVEV-	DLYPVIRRTI	PTYTLEAVYE
Vent	-GVRLV LGR-	-DKE	GDSFAVEV-	DLFPVIRRTI	PTYTLEAVYE
Mna	-GLELKLGR-	-NDE	GMSKSYI-	DLPIARRLL	TKYRLENVTE
Mec	-ETINSVGR-	-DGS	GLITRTEM-	DALPLVRRAF	KQYTLRAVSK
hPOLD	-LGRV AGLCNIRDS	-DGS	SFQSKQIGR	DMLQVLLREY	RSYTLNAVGF
ScePOLD	-FGRL KTVKQEKES	-DGS	VFSKAYGT-	DLQFIQREY	RSYTLNAVGA
Pis	-GVPLKVDR L	-DGS	GHW-	DLQFIQREY	KLKTLDRVAE
Sao	-EIPIDVA-	-DGS	GKDEAKYL-	DLQFIQREY	KLKTLDRVAE
Pae	-VPLRLGR-	-DGS	MGWREHGR-	DLQFIQREY	KLKTLDRVAE
Eco	-LPLRLGR-	-DGS	LEWREHGR-	DLQFIQREY	KLKTLDRVAE
R869	-FGES TAKRLSPHRK	-DGS	TRVKVNIENMY	DLQFIQREY	KLKTLDRVAE
T4	-LGER SMKRFSPIGR	-DGS	VKSKLIONMY	DLQFIQREY	KLKTLDRVAE
Phi29	-RN GFKWSADGLP	-DGS	NTYNTIISRM	DLQFIQREY	KLKTLDRVAE
Consensus	-GI-L-LGR-	-DGS	GDRFAVEV-	DLQFIQREY	KLKTLDRVAE
		580	600	620	640
Tgo	-AIFGQPKK-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Kod1	-AVFGQPKK-	-VYAEI T	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
9 th	-AVFGQPKK-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Pfu	-AIFGQPKK-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Vent	-AVLGKTKSK-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Mna	-ALFDVKKVD-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Mec	-ELLSREKLD-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
hPOLD	-HFLGQEKED-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
ScePOLD	-HFLGQEKED-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Pis	-YFGVMKREE	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Sao	-ALLGTSKVKV	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Pae	-TLLEGK AID	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Eco	-ELLGEGKSID	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
R869	-FELNVGKLKY	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
T4	-HETKKGKLPY	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Phi29	-DF--	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL
Consensus	-ALGXKKEK-	-VYAEI A	ERVARYSMED	KEFFPME--A	OLSRVVGQSL

FIG.1(continued)

	660	680	700	720				
Tgo	NLVEWFLLRK	AYERNELAPN	KPDERELARR	-----RES-----	YAGGYV	GLWE-NIVYL	DFRSLYPSI	413
Kod1	NLVEWFLLRK	AYERNELAPN	KPDEKELARR	-----ROS-----	YEGGYV	GLWE-NIVYL	DFRSLYPSI	413
9 th	NLVEWFLLRK	AYERNELAPN	KPDERELARR	-----RGG-----	YAGGYV	GLWD-NIVYL	DFRSLYPSI	413
Pfu	NLVEWFLLRK	AYERNEVAPN	KPSEEEYQRR	-----LRES-----	YTGQGV	GLWE-NIVYL	DFRALYPSI	414
Vent	NLVEWFLLRK	AYERNELAPN	KPDEEEYKRR	-----LRIT-----	YLGQGV	GLWE-NIVYL	DFRSLYPSI	416
Mna	QMVYLLLRK	SYKMGVAPN	RPSGKEYQKR	-----IRSS-----	YEGGYV	GLHE-DIVSM	DFLSLYPSI	422
Mec	SMVETLLRRE	FGLKDRVILP	KPGDELSAER	YDMSSD-----	LKSGEV	GLLE-NVLII	DYKSLYPTIM	490
hPOLD	KVVSQ-LLRQ	AMHEGLLMPV	VKSEGG-----	-----ED-----	YTGATV	GLLE-NVLII	DYKSLYPTIM	490
SacPOLD	KVVSQ-LLRQ	CLEIDTVIPN	MOSQAS-----	-----DDQ-----	YEGATV	GLLE-NVLII	DYKSLYPTIM	490
Pis	NRVEWMLLRK	AYRLGEVAPN	R-EEREY-----	-----EP-----	YKGAIV	GLYE-SVLVL	DFSSLYPSIM	611
Sso	TWVKNLYYWE	HRKRNWLIPL	K-EELAKS	SNIRTSALIK	GKGYKGAVV	GLYE-SVLVL	DFSSLYPSIM	617
Pae	AAFTHLVMP	MHRAGFVAPN	LGEKRPEAS-	-----	PGGFV	GLYE-SVLVL	DFSSLYPSIM	617
Eco	AAFGHLYFPR	MHRAGFVAPN	LGEVFPFAS-	-----	PGGFV	GLYE-SVLVL	DFSSLYPSIM	617
RB69	DAI-IFNS	LKEONKVIPO	GRSHVQCP-	-----	YPGAFV	GLYE-SVLVL	DFSSLYPSIM	617
T4	DAI-IFNS	LKEONKVIPO	QGSHVQCS-	-----	FPGAFV	GLYE-SVLVL	DFSSLYPSIM	617
Phi29	KGFKDITTK	KFKKVFTLS	LGLDKEV----	-----	RYAYRGGFTW	GLYE-SVLVL	DFSSLYPSIM	617
Consensus	NLVEWLLLRK	AYERNEVAPN	KPSEXEXA--	---R-----	YPGGYV	GLYE-NIVXL	DFRSLYPSI	
Tgo	ITHNVSPDIL	-NREGCEEY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Kod1	ITHNVSPDIL	-NREGCEEY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
9 th	ITHNVSPDIL	-NREGCEEY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Pfu	ITHNVSPDIL	-NREGCEEY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Vent	ITHNVSPDIL	-NREGCEEY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Mna	MSHNLSPETI	-DGTCCSDE-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Mec	MAHNLCTYTV	VTRDRPDGK-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
hPOLD	MAHNLCTYTV	LRPGTAOKL-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
SacPOLD	MAHNLCTYTV	LRPGTAOKL-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Pis	RTWNLSPDIT	DIQCKKPY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Sso	RTWNLSPDIT	DIQCKKPY-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Pae	RTFLIDPVGL	-----VEGL-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Eco	RTFLIDPVGL	-----VEGL-	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
RB69	RQVNLSPETI	AGTFKVAPLH	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
T4	RQVNLSPETI	RQCFKVAPLH	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Phi29	YSRLLPYG--	-----EPIV	-----DVAPO-	---V-----	GHK	FC--KDFPGF	IPSL	800
Consensus	XTHNLSPDIL	-NREGCXPY-	---V-----	GHX	FC--KDFPGF	IPSL	800	

FIG.1(continued)

		820	840	860	880
Tgo	---	---	---	---	---
Kod1	---	---	---	---	---
9°N	---	---	---	---	---
Plu	---	---	---	---	---
Vent	---	---	---	---	---
Mma	---	---	---	---	---
Mac	---	---	---	---	---
hPOLD	---	---	---	---	---
SeePOLD	---	---	---	---	---
Pls	---	---	---	---	---
Sso	---	---	---	---	---
Pae	---	---	---	---	---
Eco	---	---	---	---	---
R869	---	---	---	---	---
T4	---	---	---	---	---
Phi29	---	---	---	---	---
Consensus	---	---	---	---	---
		900	920	940	960
Tgo	---	---	---	---	---
Kod1	---	---	---	---	---
9°N	---	---	---	---	---
Plu	---	---	---	---	---
Vent	---	---	---	---	---
Mma	---	---	---	---	---
Mac	---	---	---	---	---
hPOLD	---	---	---	---	---
SeePOLD	---	---	---	---	---
Pls	---	---	---	---	---
Sso	---	---	---	---	---
Pae	---	---	---	---	---
Eco	---	---	---	---	---
R869	---	---	---	---	---
T4	---	---	---	---	---
Phi29	---	---	---	---	---
Consensus	---	---	---	---	---

FIG.1(continued)

	980	1000	1020	1040					
Tgo	-----ADAE	TVKK-----K	AKEFLDYINA	-----KLPGLLELEY	EGFYKRGFF	-----V-TK	KKYAV-IDE	599	
Kod1	-----ADAE	TVKK-----K	AMEFLKYINA	-----KLPGALELEY	EGFYERGFF	-----V-TK	KKYAV-IDE	599	
9°N	-----ADAE	TVKK-----K	AKEFLKYINP	-----KLPGLLELEY	EGFYVRGFF	-----V-TK	KKYAV-IDE	599	
Phu	-----GESE	EIKK-----K	AKEFLKYINS	-----KLPGLLELEY	EGFYLRGFF	-----V-TK	KKYAV-IDE	600	
Vent	-----EKPE	EIKK-----K	AKEFLNYINS	-----KLPGLLELEY	EGFYLRGFF	-----V-TK	KKYAV-IDE	602	
Mma	-----DDKE	KLKSYELLEK	TREFLKNINN	-----TLPGEMELEY	EGFYKRGIF	-----V-TK	KKYAL-IDE	618	
Mac	-----KGDL	SLEEVSLVGN	RLS--EIVSA	-----SLDPMELEF	ESVAKRALL	-----I-AK	KRYALWLFE	710	
hPOLD	-----SSVA	EAMAL-----	GREAADWVSG	-----HFPSPIRLEF	EKVYFYPYL	-----I-SK	KRYAGLFFSS	816	
SoaPOLD	-----IDLK	EAMDL-----	GTEAAKYVST	-----LFKHPINLEF	EKAYFYPYL	-----I-NK	KRYAGL-FWT	821	
Pis	-----D-----	VDKLIKYYVEE	-----	-----KYG-IDIKV	DKDYAKVLF	-----TEAK	KRYAGLL--	822	
Sso	-----KNSL	E-----	--NIIKWVKT	-----TFN--LDLEV	DKTYKFVAF	-----SGLK	KNYFGVYQ--	707	
Pae	-----PRAE	EEAAK-----	GRALVARVND	-----WWEHLEKEEF	GLDLSALELQF	-----ETHYRRFLMP	TVRGAEEGSK	KRYAGLVRRA	627
Eco	-----AHSE	EEAAK-----	GRALVQHUNA	-----WFAETLQKO-	RLTSALELEY	-----ETHFCRFLMP	TIRGADITGSK	KRYAGLIOEG	624
RB69	-----VGESKFRDIN	HWVDFLDKFA	ERMEMPADIR	-----NKQHLMFMOR	EALACPPLOS	-----KGIGGFWTGK	KRYALNVWDM	715	
T4	-----VGLDRFKEQN	DLVEFMNQFG	KKKMEPMIDV	-----NREHLMHMDR	EALSCPPLOS	-----KVGGGFWAK	KRYALNVYDM	712	
Phi29	-----	GTEIPOVVKD	-----	-----	-----	-----IVDPK	K-----	479	
Consensus	-----ADAE	EXKK-----	XXEFLKYINA	-----KLPGLLELEY	EGFYKRGLF-	-----V-TK	KRYAG-IDE-	1120	
Tgo	-----EDKI	TTRGLEIVRR	DWSEIAKETQ	ARVLEAIIK-	-----HGDVEEA-	-----VRIVKEV	TEKLSKYEVP	P-----EKLIV	661
Kod1	-----EGKI	TTRGLEIVRR	DWSEIAKETQ	ARVLEAIIK-	-----DGDVEEA-	-----VRIVKEV	TEKLSKYEVP	P-----EKLIV	661
9°N	-----EGKI	TTRGLEIVRR	DWSEIAKETQ	ARVLEAIIK-	-----HGDVEEA-	-----VRIVKEV	TEKLSKYEVP	P-----EKLIV	661
Phu	-----EGKV	TTRGLEIVRR	DWSEIAKETQ	ARVLETILK-	-----HGDVEEA-	-----VRIVKEV	IQKLANYEIP	P-----ECLA	862
Vent	-----EGRI	TTRGLEIVRR	DWSEIAKETQ	AKVLEAIIK-	-----EGSVEKA-	-----VEVVRDV	VEKIAKYRVP	L-----EKLIV	864
Mma	-----NEKI	TVKGLEIVRR	DWSEIVSKNTQ	KNVLEALLK-	-----EGSVENKA-	-----KKVIQDT	IKELDKGKVN	N-----EDLL	880
Mac	-----RNSGW-ENKI	KVKGMETVRR	DWSELTSITL	KNVLEFVLI-	-----EGSDVKA-	-----VEHVVRV	VSDVRNLDPO	KDAGIEKLIV	782
hPOLD	-----RPDAH-D-RM	DKKGLEAVRR	DNCPILVANLV	TASLRRLI-	-----DRDPEGA-	-----VAHAODV	ISDILCNRID	-----ISQIV	882
SoaPOLD	-----NPDKF-D-KL	DKKGLASVRR	DSCLSIVSIM	NKVLKKILI-	-----ERNVDGA-	-----LAFVRET	INDILHNRVD	-----ISKL	883
Pis	-----R-DGRI	DIVGFEVVRG	DWSELAKDVO	LRVIEIILK-	-----SRDIVEAR	-----HGVIKYIREI	IERLKNKYFN	-----IDEL	678
Sso	-----DGKV	DIKQMLVKRR	DNTEFVKKVF	NEVKELMIS-	-----NSPDVKEIK	-----RKIVDVVKGS	VEKLKNKGYN	-----LDELA	778
Pae	-----DGEM	VFKQMLETVRT	DWSEPLAQRFQ	QELYLRFNR	QFYQD-----	-----VYRDY	VRRITLAGEID	-----DILLV	687
Eco	-----DK-----ORM	VFKQMLETVRT	DWTPLAQOFO	QELYLRFNR	EPYQE-----	-----YVRET	IDKLMAGED	-----ARLV	683
RB69	-----EGTRYAEPKL	KIMGMEIOQS	STPKAVQKAL	KECIRRMIL	-----QEGE	-----ESLQEYFKEF	EKEFRQLNY	-----IS	779
T4	-----EDKRFAEPHL	KIMGMEIOQS	STPKAVQKAL	EESIRRMIL	-----QEGE	-----ESLQEYFKEF	EKEYRQLDY	-----KV	776
Phi29	-----	--LGYWAHES	TFKRAKYLKQ	KTYIQOIMK	-----EVOCKLVEGS	-----PDDYTDIKFS	VK-----	-----CAGMTDK	536
Consensus	-----EGKI	TTKQLETVRR	DWSEIAKETQ	ARVLEXIILK	-----EGDVEEA-	-----VXYVKEV	IEKLXXYEVP	-----EKLIV	

	1140	1160	1180	1200
Tgo	!YEQITRDLK	DYKATG----	!GKIRPGTIV	ISYIVLKQSG
Kod1	!HEQITRDLK	DYKATG----	!GVKIRPGTIV	ISYIVLKQSG
9 ^N	!HEQITRDLK	DYKATG----	!GVKIRPGTIV	ISYIVLKQSG
Phi	!YEQITRDLK	EYKATG----	!GVKIRPGTIV	ISYIVLKQSG
Vent	!HEQITRDLK	EYKATG----	!GVKIRPGTIV	ISYIVLKQSG
Mna	!HTQLTKR!E	DYKTTA----	!GKRVNTGSD	ISYIVLKQSG
Mec	!TRTLTRKAD	SYKNKQ----	!GIMPSTGTR	ISYIVLKQSG
hPOLD	!TKELTRAAS	DYAGKO----	!GIMPSTGTR	ISYIVLKQSG
ScePOLD	!SKTL-----	NYTNPO----	!GIMPSTGTR	ISYIVLKQSG
Pis	!WKTLDKELD	EYKATG----	!GKRVNTGSD	ISYIVLKQSG
Sso	!FKVMSKPLD	AYKNT-----	!GKRVNTGSD	ISYIVLKQSG
Pae	!YRKRLRRRLD	DYGRNV----	!GKRVNTGSD	ISYIVLKQSG
Eco	!YRKRLRRRLS	EYGRNV----	!GKRVNTGSD	ISYIVLKQSG
RB69	!ASVSSANN:A	KYDVGGFPGP	!GKRVNTGSD	ISYIVLKQSG
T4	!AEVKTAND:A	KYDDKGWPGF	!GKRVNTGSD	ISYIVLKQSG
Phi29	!KKEVT-----	NFK-VGFSRK	!GKRVNTGSD	ISYIVLKQSG
Consensus	!HKQLTRDLK	DYKATG----	!GKRVNTGSD	ISYIVLKQSG
Tgo	YDAEYYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
Kod1	YDAEYYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
9 ^N	YDAEYYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
Phi	YDAEYYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
Vent	YDPDYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
Mna	YDTNYYIENQ	IL--PPVERI	!GKRVNTGSD	ISYIVLKQSG
Mec	IDVQYYVKKQ	IL--PPVERI	!GKRVNTGSD	ISYIVLKQSG
hPOLD	IDTQYYLEQQ	LA--KPLLR!	!GKRVNTGSD	ISYIVLKQSG
ScePOLD	VDSTRYLTNQ	LO--NPIS!	!GKRVNTGSD	ISYIVLKQSG
Pis	IDIDYYIERQ	IL--PAALR!	!GKRVNTGSD	ISYIVLKQSG
Sso	IDAEKYLE-A	LR--STFEQ!	!GKRVNTGSD	ISYIVLKQSG
Pae	IDYEHYVQKQ	LO--PVADA!	!GKRVNTGSD	ISYIVLKQSG
Eco	LDYEHYLTQ	LO--PVAEG!	!GKRVNTGSD	ISYIVLKQSG
RB69	DDVLHWMYD	YLTFTFKP	!GKRVNTGSD	ISYIVLKQSG
T4	SDVLSWIDHS	TLFQKSFVKP	!GKRVNTGSD	ISYIVLKQSG
Phi29	VDOTFTIK--	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG
Consensus	YDAEYYIENQ	VL--PAVERI	!GKRVNTGSD	ISYIVLKQSG

FIG.1(continued)

	1300	1320	1340	1360
Tgo	---RYQKTR QVGLGAW--L	---KPK-T---	---	---
Kod1	---RYQKTR QVGLSAW--L	---KPKGT---	---	---
g*N	---RYQKTK QVGLGAW--L	---KVKGKK---	---	---
Plu	---RYQKTR QVGLTSW--L	---NIK-KS---	---	---
Vent	---RYQSSK QTGLDAW--L	---K---R---	---	---
Mna	---KDSK-K QYTLHHF--L	---K---K---	---	---
Mac	---DFD-AK QKGLDFEVK	---KPEAKKOEKS	---SSOKGTNGKI	---LEKAPEEKAR
hPOLD	FCQPRESELY QKEVSHLNAL	EERFSRLWTO	CORCQGSLSHE	DVICTSRDOP
ScePOLD	NCLARSSELY IKALYDVRDL	EEKYSRLWTO	CORCAGNLHS	EVLCNKNCO
Pis	---KTK RMERSLLDFL	S---	---	---
Sso	---AA TMSIDSEFSY	PSKGN-	---	---
Pat	---DR QMAL---	---	---	---
Eco	---TG QLGL---	---	---	---
R889	---EK KASLFDM---	---	---	---
T4	---EE KASLDFL---	---	---	---
Phi29	---	---	---	---
Consensus	---R---TK QVGLXAW--L	K---	---	---

FIG.1(continued)

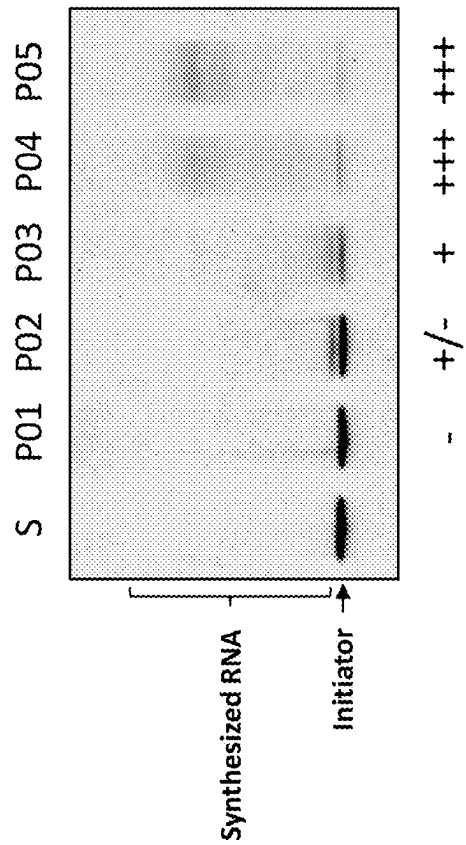


FIG.3A

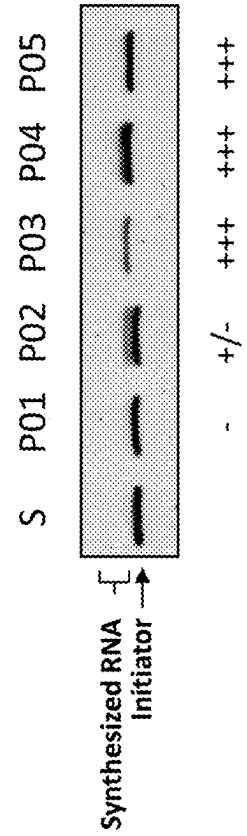


FIG.3B

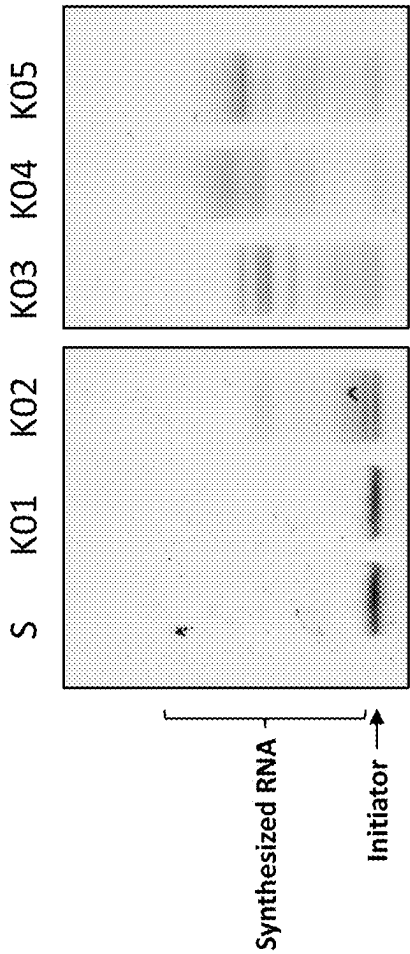


FIG.4A



FIG.4B

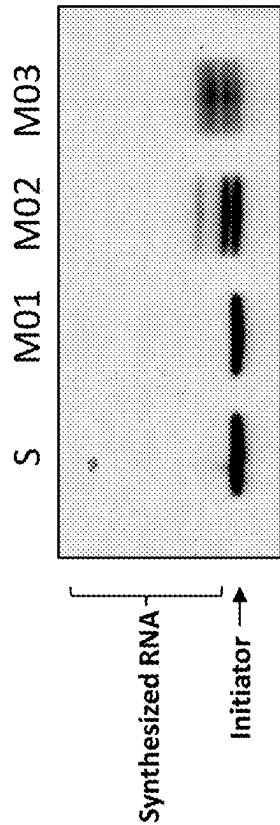


FIG.5A

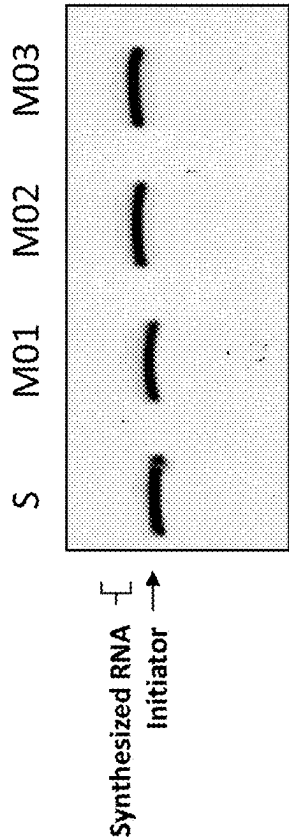


FIG.5B

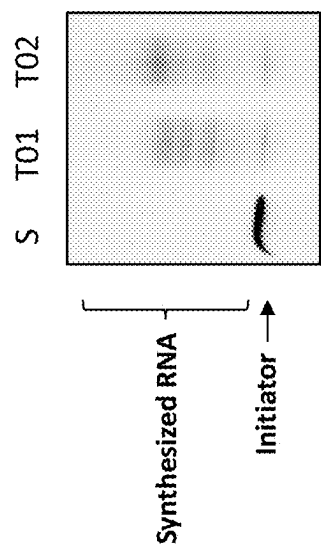


FIG.6A

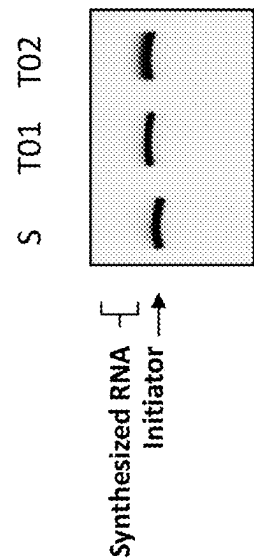


FIG.6B

NUCLEIC ACID POLYMERASE VARIANTS, KITS AND METHODS FOR TEMPLATE-INDEPENDENT RNA SYNTHESIS

CROSS REFERENCE

[0001] This application is a continuation application of U.S. application Ser. No. 17/936,744, filed on Sep. 29, 2022, which claims the benefit of U.S. Provisional Application No. 63/249,819, filed on Sep. 29, 2021. The contents of these applications are incorporated herein by reference.

TECHNICAL FIELD

[0002] The present disclosure relates to nucleic acid polymerase variants and kits comprising the same for use particularly in the context of de novo enzymatic nucleic acid synthesis.

SEQUENCE LISTING

[0003] The present application contains a computer readable Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML format file, created on May 2, 2025, is named YDBP0012USA3_A3_1Sequence Listing XML.xml and is 92.9 kilobytes in size.

BACKGROUND

[0004] Synthetic oligonucleotides are crucial to many aspects of biotechnological research in both the academic and industrial settings. However, there are many limitations of traditional chemical synthesis methods developed decades ago. This is especially true for de novo RNA synthesis, which remains largely inaccessible to those heavily invested in advancing genome engineering technologies, RNA-based diagnostics, RNA-based therapeutics, RNA-based sequencing technologies, nucleic acid-based information storage, and even biological computing. Chemical synthesis of RNA is troubled by lengthy reaction steps that require both harsh chemical reagents and biologically incompatible organic solvents. These reaction conditions often lead to the depurination of the nucleotide bases, unexpected insertions or deletions from the overall sequence, and the premature irreversible capping of the oligonucleotide resulting in unwanted truncated products. This substantially increases the overall error-rate of RNA synthesis, limits the synthesis length of RNA oligonucleotides (less than 120 nucleotides), and requires longer lead-times to obtain acceptable yields of a desired product. Moreover, the chemical synthesis of RNA oligonucleotide is toxic and labor-intensive. Overcome the current limitations of RNA oligonucleotide synthesis is therefore important.

[0005] Enzyme-based de novo nucleic acid synthesis is an emerging, non-toxic method to substitute for the decades-old, toxic, chemical phosphoramidite-based nucleic acid synthesis. All living organisms rely on nucleic acid polymerases to efficiently duplicate their nucleic acid. Owing to their nucleic acid duplication mechanism, most nucleic acid polymerases require a template to direct synthesis and incorporation of nucleotides into a growing nucleic acid strand. The template-dependent manner of nucleic acid synthesis requires the polymerase to associate with a primer-template nucleic acid before the nucleotide can be added to the 3'-terminus of primer by the polymerase.

[0006] Unlike most replicative nucleic acid polymerases, the X-family terminal deoxynucleotidyl transferase (Tdt) is a unique class of mesophilic enzyme, which doesn't rely on a template for adding nucleotides during nucleic acid synthesis. Tdt only requires a short initiator DNA or primer to direct synthesis and incorporation of nucleotides into a growing initiator DNA or primer. Previous studies revealed that Tdt can incorporate both deoxyribonucleotides (dNTPs) and ribonucleotides (rNTPs) with a minor discrimination (Bould, J B et al. *The Journal of biological chemistry* (2001)). However, Tdt fails to extend the newly synthesized RNA strand beyond around 4-5 ribonucleotides, suggesting that Tdt has impeded accommodation of ribo- or mixed ribo/deoxyribonucleic acid substrates for further synthesis. Therefore, Tdt enzyme is not suitable for de novo RNA synthesis.

SUMMARY OF THE INVENTION

[0007] Owing to the diverse structure-function relationships mentioned above, the naturally occurring, replicative nucleic acid polymerases cannot readily utilize ribonucleotides (rNTPs) as a substrate for de novo ribonucleic acid synthesis. Thus, the tailor-made, modified nucleic acid polymerase is a prerequisite for exerting the utilities of a variety of nucleic acid synthesis applications.

[0008] The inventor has discovered the novel positions/regions in the amino acid sequences of B-family DNA polymerase variants that play crucial parts in endowing the said polymerases with a template-independence and an enhancing nucleotide substrate binding affinity of said polymerases for ribonucleotides, thereby providing a new option for the template-independent RNA synthesis method.

[0009] Accordingly, in one aspect, the present disclosure provides a, so-called, RNA polymerase variant comprising: a motif A, and a motif B corresponding to positions 706 to 730, and 843 to 855, respectively, of a consensus sequence (SEQ ID NO:1); and at least one amino acid substitution at a position in the motif A, the motif B, or the combination thereof; wherein the RNA polymerase variant has a reduced or eliminated the intrinsic 3' to 5' exonuclease activity.

[0010] In one embodiment, the representative RNA polymerase variant is modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.

[0011] In one embodiment, the wild-type B-family DNA polymerase is *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanobrevibacterium acetivorans* DNA polymerase (Mac), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Methanococcus maripaludis* DNA polymerase (Mma), human DNA polymerase, delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), *Escherichia* phage RB69 DNA polymerase (RB69), *Escherichia* phage T4 DNA polymerase (T4), or *Bacillus* phage Phi29 DNA polymerase (Phi29).

[0012] In one embodiment, the representative RNA polymerase variant comprises a motif Exo I corresponding to

positions 349 to 364 of the consensus sequence (SEQ ID NO:1), and the RNA polymerase variant has at least one amino acid substitution at a position in the motif Exo I. Preferably, an amino acid L or M corresponding to position 715 of SEQ ID NO: 1 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y; an amino acid Y corresponding to position 716 of SEQ ID NO: 1 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P corresponding to position 717 of SEQ ID NO: 1 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0013] In one embodiment, the selected RNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein: an amino acid L at position 408 of SEQ ID NO: 2 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0014] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein: an amino acid L at position 408 of SEQ ID NO: 2 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 485 of SEQ ID NO: 2 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0015] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein: an amino acid L at position 408 of SEQ ID NO: 3 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0016] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein: an amino acid L at position 408 of SEQ ID NO: 3 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 485 of SEQ ID NO: 3 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0017] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein: an amino acid L at position 408 of SEQ ID NO: 4 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 4 is remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid

P at position 410 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0018] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein: an amino acid L at position 408 of SEQ ID NO: 4 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 409 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 410 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 485 of SEQ ID NO: 4 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0019] In one embodiment, the representative RNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein: an amino acid L at position 409 of SEQ ID NO: 5 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0020] In one embodiment, the representative RNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein: an amino acid L at position 409 of SEQ ID NO: 5 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 486 of SEQ ID NO: 5 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0021] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein: an amino acid L at position 411 of SEQ ID NO: 6 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0022] In one embodiment, the representative RNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein: an amino acid L at position 411 of SEQ ID NO: 6 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 488 of SEQ ID NO: 6 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0023] In one embodiment, the representative RNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein: an amino acid L at position 485 of SEQ ID NO: 7 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at

position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0024] In one embodiment, the representative RNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein: an amino acid L at position 485 of SEQ ID NO: 7 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 565 of SEQ ID NO: 7 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0025] In one embodiment, the representative RNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein: an amino acid M at position 426 of SEQ ID NO: 8 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y; an amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0026] In one embodiment, the representative RNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein: an amino acid M at position 426 of SEQ ID NO: 8 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y; an amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 508 of SEQ ID NO: 8 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0027] In one embodiment, the representative RNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein: an amino acid L at position 518 of SEQ ID NO: 9 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0028] In one embodiment, the representative RNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein: an amino acid L at position 518 of SEQ ID NO: 9 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; an amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V; an amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and an amino acid A at position 601 of SEQ ID NO: 9 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

[0029] In one embodiment, the representative RNA polymerase variant exhibits an activity of synthesizing nucleic acids in a template-independent manner by adding at least

one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator.

[0030] In one embodiment, the extendible initiator comprises a single-stranded oligonucleotide initiator, a blunt-ended double-stranded oligonucleotide initiator, or a mixture thereof.

[0031] In one embodiment, the extendible initiator is a free form nucleic acid to be reacted in a liquid phase.

[0032] In one embodiment, the extendible initiator is immobilized on a solid support, wherein the solid support comprises a particle, bead, slide, array surface, membrane, flow cell, well, microwell, nano-well, chamber, microfluidic chamber, channel, or microfluidic channel.

[0033] In one embodiment, the at least one nucleotide is linked with a detectable label.

[0034] In one embodiment, the at least one nucleotide comprises a ribose.

[0035] In one embodiment, the representative RNA polymerase variant exhibits the activity at reaction temperatures ranging from 10° C. to 100° C.

[0036] In one aspect, the present disclosure further provides a kit for performing de novo enzymatic nucleic acid synthesis, comprising an RNA polymerase variant derived from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17, wherein the RNA polymerase variant exhibits activity of synthesizing nucleic acids in a template-independent manner by adding at least one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator, thereby synthesizing a desired nucleic acid sequence.

[0037] In another aspect, the present disclosure further provides method for template-independent synthesis of an RNA oligonucleotide, comprising:

[0038] (a) providing an initiator oligonucleotide,

[0039] (b) providing an RNA polymerase variant;

[0040] (c) combining the initiator oligonucleotide, the RNA polymerase variant and one or more nucleotides under conditions sufficient for the addition of at least one nucleotide to the 3' end of the initiator oligonucleotide;

[0041] wherein the selected RNA polymerase variant comprising:

[0042] a motif A, and a motif B corresponding to positions 706 to 730, and 843 to 855, respectively, of a consensus sequence (SEQ ID NO:1); and at least one amino acid substitution at a position in the motif A, the motif B, or the combination thereof; wherein the RNA polymerase variant has a reduced or eliminated the intrinsic 3' to 5' exonuclease activity.

[0043] Accordingly, the present invention relates to the specific RNA polymerase variants that exhibit an improved performance on incorporating a variety of nucleotides for nucleic acid synthesis at various reaction temperatures in the absence of nucleic acid template. More particularly, the de novo RNA synthesis method can be efficiently performed by said RNA polymerase variants with a broad-spectrum of nucleotides and nucleotide analogues.

BRIEF DESCRIPTION OF THE DRAWINGS

[0044] The present disclosure will become more readily appreciated with reference to the following description in conjunction with the accompanying drawings.

[0045] FIG. 1 shows an amino acid sequence alignment of the wild-type B-family DNA polymerases (PolB) related to the present invention and their consensus sequence.

[0046] FIGS. 2A and 2B show the results of the RNA synthesis reactions described in Example 3.1.

[0047] FIGS. 3A and 3B show the results of the RNA synthesis reactions described in Example 3.2.

[0048] FIGS. 4A and 4B show the results of the RNA synthesis reactions described in Example 3.3.

[0049] FIGS. 5A and 5B show the results of the RNA synthesis reactions described in Example 3.4.

[0050] FIGS. 6A and 6B show the results of the RNA synthesis reactions described in Example 3.5.

DETAILED DESCRIPTION

Definition

[0051] All terms including descriptive or technical terms which are used herein should be construed as having meanings that are understandable to one of ordinary skill in the art. However, the terms may have different meanings according to an intention of the user, case precedents, or the appearance of new technologies. Also, some terms may be arbitrarily selected by the applicant, and in this case, the meaning of the selected terms will be described in detail in the descriptions of the present disclosure. Thus, the terms used herein are defined based on the meaning of the terms together with the descriptions throughout the specification. In addition, the titles and subtitles may be attached to the description for readability, but these titles do not affect the scope of the present invention.

[0052] As used herein, the term “a,” “an,” or “the” includes plural referents unless expressly and unequivocally limited to one referent. The term “or” is used interchangeably with the term “and/or” unless the context clearly indicates otherwise.

[0053] Also, when a part or a method “includes” or “comprises” a component or a step, respectively, unless there is a particular description contrary thereto, the part or the method can further include other components or other steps, not excluding the others.

[0054] As used herein, an “amino acid” refers to any monomer unit that can be incorporated into a peptide, polypeptide, or protein. As used herein, the term “amino acid” includes the following twenty natural or genetically encoded alpha-amino acids: alanine (Ala or A), arginine (Arg or R), asparagine (Asn or N), aspartic acid or aspartate (Asp or D), cysteine (Cys or C), glutamine (Gln or Q), glutamic acid or glutamate (Glu or E), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V). In cases where “X” residues are undefined, these should be defined as “any amino acid”.

[0055] The term “functionally equivalent” or “equivalent” is used to describe a specific B-family DNA polymerase (PolB) variant and any of the derivative RNA polymerase variant provided herein having the substitution or mutation

that is considered to occur at the amino acid position in the other PolB, PolB variant according to the sequence alignment, or a reference sequence, which has the same functional or structural role in the enzyme. The equivalent positions may be defined according to homologues, conserved motifs, user-defined, or derived, consensus sequence.

[0056] Generally, the homologous PolBs have similar or identical amino acid sequences and functional structure, and thereby the equivalent amino acid substitution mutations among different PolBs generally occur at homologous amino acid positions. The term “functionally equivalent” or “equivalent” used herein also encompass mutations that are “homologous” or “positionally equivalent” to a given mutation in view of protein sequence or structural alignment, regardless of the actual function of the mutated amino acid. Practically, the “functionally equivalent”, “homologous” and/or “positionally equivalent” amino acid residues of different polymerases can be identified according to the protein sequence or structural alignment. Accordingly, a cross-species alignment was made on multiple wild-type PolBs, as illustrated in FIG. 1, and the consensus sequence (SEQ ID NO: 1) is used as a positional reference sequence.

[0057] For example, the substitution of amino acid aspartic acid (D) with alanine (A) at position 141 of the wild-type *Thermococcus kodakarensis* (Kod1) (D141A) amino acid sequence would be functionally equivalent to the amino acid substitution mutation D114A at the conserved residue of wild-type *Escherichia coli* phage RB69 DNA polymerase (RB69) amino acid sequence. When the positional reference sequence is used to describe these equivalent amino acid substitutions, the functionally equivalent positions of both amino acid residue 141 of Kod1 and amino acid residue 114 of RB69 corresponds to position 354 of the consensus sequence (SEQ ID NO: 1).

[0058] The term “conserved” means the segment of polymerase having the same amino acid residue in the homologous or equivalent position of different PolBs from various sources. The term “semi-conserved” used herein refers to the segment of polymerase that has a similar property of amino acid residue or an identical amino acid residue in the homologous position of different PolBs from various sources.

[0059] The terms “nucleic acid”, “nucleic acid sequence”, “oligonucleotide”, “polynucleotide”, and “nucleic acid fragment” as used herein refer to a deoxyribonucleotide or ribonucleotide sequence in a single-stranded or a double-stranded form of which the sources and length are not limited herein; and generally, includes naturally occurring nucleotides or artificial chemical mimics. The term “nucleic acid” as used herein is interchangeable with the terms including natural or unnatural “oligonucleotide”, “polynucleotide”, “DNA”, “RNA”, “gene”, “complementary DNA” (cDNA), and “messenger RNA” (mRNA) in use.

[0060] The “nucleic acid”, “oligonucleotide”, or “polynucleotide” used herein refers to a polymer that can be corresponded to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or an analogue thereof. This includes polymers of nucleotides such as RNA and DNA, as well as synthetic forms, modified (e.g., chemically or biochemically modified) forms thereof, and mixed polymers (e.g., including both RNA and DNA subunits). Exemplary modifications include methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages

(e.g., methyl phosphonates, phosphotriesters, phosphoamides, carbamates, and the like), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, and the like), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids and the like). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Typically, the nucleotide monomers are linked via phosphodiester bonds, although synthetic forms of nucleic acids can comprise other linkages (e.g., peptide nucleic acids as described in Nielsen et al. *Science* 254:1497-1500, 1991). A nucleic acid can be or can include, e.g., a chromosome or chromosomal segment, a vector (e.g., an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, and a primer. A nucleic acid can be, e.g., single-stranded, double-stranded, or triple-stranded and is not limited to any particular length. Unless otherwise indicated, a particular nucleic acid sequence optionally comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

[0061] The nucleic acid as used herein also includes nucleic acid analogue. The term nucleic acid analogue is known to describe compounds or artificial nucleic acids which are functionally or structurally equivalent to naturally existing RNA and DNA. A nucleic acid analogue may have one or more parts of a nucleotide (the phosphate backbone, pentose sugar, and nucleobase) being modified. These modifications on the nucleotide change the structure and geometry of the nucleic acid and its interactions with nucleic acid polymerases. The nucleic acid analogue also encompasses the emerging category of artificial nucleic acids, such as XNA, which is designed to have new-to-nature forms of sugar backbone.

[0062] Examples of nucleic acid analogues include but are not limited to: the universal bases, such as inosine, 3-nitropyrrole, and 5-nitroindole, which can form a base-pair with all four canonical bases; the phosphate-sugar backbone analogues, such as peptide-nucleic acids (PNA), which affect the backbone properties of the nucleic acid; chemical linker or fluorophore-attached analogues, such as amine-reactive aminoallyl nucleotide, thiol-containing nucleotides, biotin-linked nucleotides, rhodamine-linked nucleotides, and cyanine-linked nucleotides; the fluorescent base analogues, such as 2-aminopurine (2-AP), 3-methylisoxanthopterin (3-MI), 3-methylisoxanthopterin (6-MI), 4-amino-6-methylisoxanthopterin (6-MAP), and 4-dimethylaminopyridine (DMAP); the nucleic acid probes for various genetic applications, such as the oligonucleotide-conjugated with a fluorescent reporter dye (ALEXA, FAM, TET, TAMRA, CY3, CY5, VIC, JOE, HEX, NED, PET, ROX, Texas Red and others) and/or a fluorescent quenchers (BHQs); the molecular beacons (MBs), which are single-stranded nucleic acid probes containing a stem-loop structure and a dual fluorophore-and-quencher label; and the nucleic acid aptamers.

[0063] Generally, as used herein, a “template” is a polynucleotide, or a polynucleotide mimic, that contains the desired or unknown target nucleotide sequence. In some instances, the terms “target sequence”, “template polynucleotide”, “target nucleic acid”, “target polynucleotide”, “nucleic acid template”, “template sequence, and variations thereof, are used interchangeably. Specifically, the term

“template” refers to a strand of nucleic acid from which a complimentary copy is synthesized from nucleotides or nucleotide analogues through the replication by a template-dependent or template-directed nucleic acid polymerase. Within a nucleic acid duplex, the template strand is, by the convention definition, depicted and described as the “bottom” strand. Similarly, the non-template strand is often depicted and described as the “top” strand. The “template” strand may also be referred to as the “sense”, or “plus”, strand and the non-template strand as the “antisense”, or “minus”, strand.

[0064] The term “initiator” refers to a mononucleoside, a mononucleotide, and oligonucleotide, a polynucleotide, or modified analogues thereof, from which a nucleic acid is to be synthesized by nucleic acid polymerase de novo. The term “initiator” may also refer to a xeno nucleic acids (XNA) or a peptide nucleic acid (PNA) having a 3'-hydroxyl group.

[0065] The terms “nucleotide incorporation”, “analogue incorporation”, “incorporating nucleotide” and “incorporating analogue” are known to those skilled in the art and used to describe a process or reaction for nucleic acid synthesis. Thus, as used herein, the term “incorporation” is known to flexibly refer to add one, or more nucleotides, or any specified nucleic acid precursors to the 3'-hydroxyl terminus of a nucleic acid initiator or a primer.

[0066] The term “nucleotide analogue” is known to those skilled in the art to describe the chemically modified nucleotides or artificial nucleotides, which are structural mimics of canonical nucleotides. These nucleotide analogues can serve as substrates for nucleic acid polymerases to synthesize nucleic acid. A nucleotide analogue may have one or more altered components of a nucleotide (e.g., the phosphate backbone, pentose sugar, and nucleobase), which changes the structure and configuration of a nucleotide and affects its interactions with other nucleobases and the nucleic acid polymerases. For example, a nucleotide analogue having altered nucleobase may confer alternative base-pairing and base-stacking properties in the DNA or RNA. Furthermore, by way of example, the modification at the base may generate various nucleosides such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, diamino-2,6-purine or bromo-5-deoxyuridine, and any other analogues which permits hybridization. In other exemplary aspects, modifications may take place at the level of sugar moiety (for example, replacement of a deoxyribose by an analogue), and/or at the level of the phosphate group (for example, boronate, alkylphosphonate, or phosphorothioate derivatives). A nucleotide analogue monomer may have a phosphate group selected from a monophosphate, a diphosphate, a triphosphate, a tetraphosphate, a pentaphosphate, and a hexaphosphate.

[0067] Other examples of nucleotide analogues also include nucleotides having a removable blocking moiety. Examples of the removable blocking moiety include, but are not limited to, a 3'-O-blocking moiety, a base blocking moiety, and a combination thereof. Examples of the 3'-O-blocking moiety include, but are not limited to, O-azido (O—N₃), O-azidomethyl, O-amino, O-allyl, O-phenoxyacetyl, O-methoxyacetyl, O-acetyl, O-(p-toluene)sulfonate, O-phosphate, O-nitrate, O-[4-methoxy]-tetrahydrothiopyran-yl, O-tetrahydrothiopyran-yl, O-[5-methyl]-tetrahydrofuran-yl, O-[2-methyl, 4-methoxy]-tetrahydropyran-yl, O-[5-methyl]-tetrahydropyran-yl, and O-tetrahydrothiofuran-yl,

O-2-nitrobenzyl, O-methyl, and O-acyl groups. Examples of the base blocking moiety may be a reversible dye-terminator. Examples of the reversible dye-terminator include, but are not limited to, a reversible dye-terminator of Illumina MiSeq, a reversible dye-terminator of Illumina HiSeq, a reversible dye-terminator of Illumina Genome Analyzer IIX, a reversible dye-terminator of Helicos Biosciences Heliscope, and a reversible dye-terminator of LaserGen's Lighting Terminators.

[0068] As used herein, "B-family DNA polymerases (PolBs)" refers to the most common template-dependent DNA polymerases or replicases in all domains of life and many DNA viruses. Like most nucleic acid polymerases, natural PolBs require a duplex primer-template DNA with a free 3'-hydroxyl (3'-OH) group at the primer terminus, all four nucleoside triphosphates (dATP, dTTP, dCTP, and dGTP), and catalytic divalent cations (Mg^{2+} or Mn^{2+} , etc.) for catalyzing the nucleotidyl transferase reaction of adding nucleotides to the 3'-OH terminus of a primer. The PolB enzymes, such as bacterial Pol II and archaeal B-family DNA polymerases, are replicative and repair polymerases that inherently contain a catalytic polymerase domain and a 3' to 5' exonucleolytic, or proofreading, domain for removing the mis-incorporated nucleotide from the growing primer strand during nucleic acid replication. The term "3' to 5' exonucleolytic domain" (Exo domain) refers to a region of the amino acid sequence of a polymerase, which exerts the nucleic acid degradation activity from the 3'-terminus of the primer or the polynucleotide chain. Coordinately, the term "catalytic polymerase domain" (Pol domain) refers to a region of the amino acid sequence of a polymerase, which exerts the catalytic DNA/RNA polymerase activity for adding nucleotides to the 3'-terminus of a primer or a polynucleotide chain.

[0069] All known structures of PolB catalytic polymerase domain resemble the shape of human right hand, where the key functional regions are characterized as fingers, palm, and thumb subdomains. The most conserved region is the palm subdomain, which contains the essential residues for catalysis. The protein sequence-alignment among various B-family DNA polymerases from different kingdoms of life and DNA viruses reveals that the PolB polymerases generally harbor six semi-conserved or conserved motifs (I-VI) for their essential exonuclease and polymerase functions. The first three sequence-motifs (Exo I, Exo II, Exo III) are in the Exo domain, while the other three motifs (designated as Motif A, B, and C, respectively) reside in the Pol domain (Hopfner et al, Proc. Natl. Acad. Sci. USA 96, 3600-3605, 1999). In some embodiments of the present invention, without being limited by any theory, it is discovered that through modifying some novel positions/regions in said motifs of B-family DNA polymerases, the polymerases can thereby effectively catalyze de novo RNA synthesis as a template-independent RNA polymerases. Based on the acquired activities of de novo RNA synthesis, these modified B-family DNA polymerases are referred to as template-independent RNA polymerase variants. In other words, the RNA polymerase variant is capable of incorporating ribonucleotides, such as rATP, rUTP, rCTP and rGTP, at the 3'-terminus of an initiator oligonucleotide under impeding reaction conditions.

[0070] As used herein, the term "mutant" in the context of the present invention, means a polypeptide, typically recom-

binant, that comprises one or more amino acid substitutions relative to a corresponding, functional DNA polymerase.

[0071] As used herein, in the context of the present invention, "corresponding to another sequence" (e.g., regions, fragments, nucleotide or amino acid positions, or the like) is based on the convention of numbering according to nucleotide or amino acid position number and then aligning the sequences in a manner that maximizes the percentage of sequence identity. An amino acid "corresponding to position X of specific sequence" refers to an amino acid in a polypeptide of interest that aligns with the equivalent amino acid of a specified sequence. Generally, as described herein, the amino acid corresponding to a position of a polymerase can be determined using an alignment algorithm such as BLAST and other currently available tools for conducting amino acid sequence alignment. Because not all positions within a given "corresponding region" need to be identical, non-matching positions within a corresponding region may be regarded or define as "corresponding positions". Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position X of a specified DNA polymerase" refers to equivalent positions, based on alignment, in other DNA polymerases and structural homologues and families.

[0072] As used herein, the term "consensus sequence of SEQ ID NO: 1" used herein refers to a reference sequence comprising the conserved or semi-conserved amino acids of cross-species B-family DNA polymerase. The consensus sequence of SEQ ID NO: 1 is a virtual sequence and is generated by aligned the following 16 wild-type B-family DNA polymerases to obtain the conserved amino acids: *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanococcus maripaludis* DNA polymerase (Mma), *Methanosarcina acetivorans* DNA polymerase (Mac), human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), *Escherichia coli* phage RB69 DNA polymerase (RB69), *Escherichia coli* phage T4 DNA polymerase (T4), or *Bacillus* phage Phi29 DNA polymerase (Phi29). These PolB sequences are aligned for obtaining the alignment sequence as a reference of functionally equivalent positions.

[0073] The positions of motifs Exo I, Exo II, Exo III, A, B, and C are defined by the inventor using the consensus sequence of SEQ ID NO: 1 of the present invention; therefore, it shall be noted that the positions of these motifs defined in the present invention are not totally the same as those described in the literature or prior art.

OBJECTIVES

[0074] The inventor has discovered PolB variants that have an improved function and activity for utilizing canonical nucleotides, nucleotide analogues, and initiators for synthesizing polynucleotides in a template-independent manner. These PolB variants can efficiently add said canonical nucleotides or nucleotide analogues to said initiator in

the absence of a replicative template to synthesize a polynucleotide with a random or defined sequence.

[0075] More specifically, the inventor has discovered PolB variants can efficiently catalyze the additions of natural ribonucleotides (rNTP), such as rATP, rUTP, rCTP and rGTP, to the 3'-OH ends of a single-stranded nucleic acid initiator or a blunt-end duplex nucleic acid initiator, in the absence of replicative template, to generate polynucleotides with desired nucleic acid sequences. Furthermore, the PolB variants provided herein generally have a broader substrate specificity, which means the PolB variants can utilize not only naturally occurring nucleotides, but also varieties of modified nucleotides and nucleic acid analogues for the de novo nucleic acid synthesis. Thus, the modified nucleotide can be further designed for being incorporated to the initiator to generate certain functional polynucleotides. Therefore, these PolB variants broaden the scope and utility of template-independent enzymatic nucleic acid synthesis applications for synthesizing polynucleotides with desired sequences and features.

Protein Sequence Alignment of B-Family DNA Polymerases

[0076] FIG. 1 shows the amino acid sequence alignment of 16 different wild-type B-family DNA polymerases (PolBs) utilized by the inventor, and the outcome of consensus sequence alignment is listed in the bottom (SEQ ID NO:1). The 16 wild-type PolBs used for alignment are *Thermococcus gorgonarius* DNA polymerase (Tgo, SEQ ID NO:2), *Thermococcus kodakarensis* DNA polymerase (Kod1, SEQ ID NO:3), *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N, SEQ ID NO:4), *Pyrococcus furiosus* DNA polymerase (Pfu, SEQ ID NO:5), *Thermococcus litoralis* DNA polymerase (Vent, SEQ ID NO:6), *Methanobacterium acetivorans* DNA polymerase (Mac, SEQ ID NO:7), *Pyrobaculum islandicum* DNA polymerase (Pis, SEQ ID NO:8), *Sulfolobus solfataricus* DNA polymerase (Sso, SEQ ID NO:9), *Methanococcus maripaludis* DNA polymerase (Mma, SEQ ID NO:10), human DNA polymerase delta catalytic p125 subunit (hPOLD, SEQ ID NO:11), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD, SEQ ID NO:12), *Pseudomonas aeruginosa* DNA polymerase II (Pae, SEQ ID NO:13), *Escherichia coli* DNA polymerase II (Eco, SEQ ID NO:14), *Escherichia coli* phage RB69 DNA polymerase (RB69, SEQ ID NO:15), *Escherichia coli* phage T4 DNA polymerase (T4, SEQ ID NO:16), and *Bacillus* phage Phi29 DNA polymerase (Phi29, SEQ ID NO:17).

[0077] As shown in FIG. 1, various sequence regions among these exemplary wild-type PolBs are highly conserved while other regions are more variable. Those of skill in the art will immediately recognize and understand that mutations in addition to those specifically identified and discussed herein may be also made in the variable regions of wild-type PolBs without altering, or without substantially altering, the polymerase activity of the mutated enzyme. Likewise, conservative mutations at conserved residues/positions of any of PolBs may be made without altering, or substantially altering, the polymerase activity of the mutated enzyme. Enzyme engineering based on comparative structure analysis with other functionally related enzymes or homologs is a useful technique in the molecular biology field that allows the inventor to reasonably predict the effect of a given mutation on the catalytic activity of the enzyme. Based on the present disclosure, using the sequence, struc-

tural data, and known physical properties of amino acids, those of skill in the art can mutate enzymes, such as the DNA polymerases encompassed by the present invention, without altering, or without substantially altering, the essential, intrinsic characteristics of the enzymes.

[0078] Besides, the motifs Exo I, Exo II, Exo III, A, B, and C corresponding to the positions 349 to 364, 450 to 476, 590 to 608, 706 to 730, 843 to 855, and 940 to 956 respectively, of the consensus sequence of SEQ ID NO:1 are focused in the present disclosure. More specifically, the polymerase variant in the present invention is based on substitution mutations at one or more residues correspondingly residing in said motifs.

RNA Polymerase Variant

[0079] In view of the above, provided herein are altered polymerase, which is described based on the amino acid sequence of the consensus sequence of SEQ ID NO: 1. An altered polymerase includes substitution mutations at one or more residues when compared to the consensus sequence of SEQ ID NO: 1. A substitution mutation can be at the same, a homologous, or a functionally equivalent position as compared to the consensus sequence of SEQ ID NO: 1. The skilled person can readily appreciate that an altered polymerase described herein is not naturally occurring. Therefore, an altered polymerase described herein is based on the consensus sequence of SEQ ID NO: 1 and further includes substitution mutations at one or more residues of the corresponding wild-type polymerase (individual parent polymerase). In one embodiment, at least one substitution mutation is at a position functionally equivalent to an amino acid of the consensus sequence of SEQ ID NO: 1. "Functionally equivalent" means that the altered polymerase has the amino acid substitution at the amino acid position according to the consensus sequence of SEQ ID NO: 1 that has the same functional or structural role in both the consensus sequence and the altered polymerase.

[0080] In general, functionally equivalent substitution mutations in two or more different polymerases occur at homologous amino acid positions in the amino acid sequences of the polymerases. Hence, "functionally equivalent" also encompasses mutations that are "positionally equivalent" or "homologous" to a given mutation, regardless of whether or not the particular function of the mutated amino acid is known. It is possible to identify the regions of functionally equivalent and positionally equivalent amino acid residues in the amino acid sequences of two or more different polymerases on the basis of sequence alignment and/or molecular modelling. For instance, the amino acid sequence alignment of exemplary 16 wild-type B-family DNA polymerases from different domains of life are used to identify positionally equivalent and/or functionally equivalent residues. The result of the protein sequence alignment among these PolBs is set forth in FIG. 1. Thus, for the exemplary residue 141 of the Tgo, Kod1, 9° N, Pfu, and Vent polymerases, residue 171 of the Pis, residue 231 of the Sso, and residue 198 of the Mac polymerase are functionally equivalent and positionally equivalent. Likewise, for the exemplary residue 143 of the Tgo, Kod1, 9° N, Pfu, and Vent polymerases, residue 173 of the Pis, residue 233 of the Sso, and residue 200 of the Mac polymerase are functionally equivalent and positionally equivalent. The skilled person can easily identify functionally equivalent residues in different polymerases.

[0081] In accordance with some embodiments, the provided RNA polymerase variant comprising: a motif A, and a motif B corresponding to positions 706 to 730, and 843 to 855, respectively, of a consensus sequence (SEQ ID NO:1); and at least one amino acid substitution (one or more amino acid substitutions, or a combination of amino acid substitutions) at a position in the motif A, the motif B, or the combination thereof; wherein the RNA polymerase variant has reduced or deficient in the 3' to 5' exonuclease activity. Said the 3' to 5' exonuclease activity deficiency can be reached by any means. For example, practically, the 3' to 5' exonuclease activity can be reduced, attenuated, removed or inactivated by modifying the 3' to 5' exonucleolytic domain of the polymerase to generate a polymerase that has reduced or is deficient in the 3' to 5' exonuclease activity. Preferably, the means of amino acid substitution is adapted to modify the 3' to 5' exonucleolytic domain. In other words, the provided RNA polymerase variant may comprise at least one amino acid substitution in Exo I, Exo II, Exo III or the combination. For example, the PolB variants may have functionally equivalent or positionally equivalent substitutions of the native D with A at position 354 (D354A) and the native E with A at position 356 (E356A) in the motif Exo I of SEQ ID NO:1, thereby causing a 3'-5' exonuclease deficiency.

[0082] In accordance with some embodiments, the RNA polymerase variant is modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17., which are respectively derived from the wild-type B-family DNA polymerase of *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanosarcina acetivorans* DNA polymerase (Mac), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Methanococcus maripaludis* DNA polymerase (Mma), human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (SecPOLD), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), *Escherichia* phage RB69 DNA polymerase (RB69), *Escherichia* phage T4 DNA polymerase (T4), and *Bacillus* phage Phi29 DNA polymerase (Phi29).

[0083] In accordance with certain embodiments, the representative RNA polymerase variant comprises a motif Exo I corresponding to positions 349 to 364 of the consensus sequence (SEQ ID NO:1), and the RNA polymerase variant has at least one amino acid substitution at a position in the motif Exo I. Preferably, an amino acid L or M corresponding to position 715 of SEQ ID NO: 1 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y; an amino acid Y corresponding to position 716 of SEQ ID NO: 1 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and an amino acid P corresponding to position 717 of SEQ ID NO: 1 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

[0084] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein:

an amino acid L at position 408 of SEQ ID NO: 2 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0085] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein: an amino acid L at position 408 of SEQ ID NO: 2 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 485 of SEQ ID NO: 2 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0086] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein: an amino acid L at position 408 of SEQ ID NO: 3 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0087] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein: an amino acid L at position 408 of SEQ ID NO: 3 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 485 of SEQ ID NO: 3 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0088] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein: an amino acid L at position 408 of SEQ ID NO: 4 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 410 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0089] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein: an amino acid L at position 408 of SEQ ID NO: 4 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 409 of SEQ ID

NO: 4 remains unchanged or is substituted with A, C, D, G, N, S, T or V preferably A; an amino acid P at position 410 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 485 of SEQ ID NO: 4 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0090] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein: an amino acid L at position 409 of SEQ ID NO: 5 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0091] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein: an amino acid L at position 409 of SEQ ID NO: 5 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 486 of SEQ ID NO: 5 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0092] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein: an amino acid L at position 411 of SEQ ID NO: 6 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0093] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein: an amino acid L at position 411 of SEQ ID NO: 6 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 488 of SEQ ID NO: 6 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0094] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Methanoscarchina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein: an amino acid L at position 485 of SEQ ID NO: 7 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G,

N, S, T or V, preferably A; and an amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0095] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Methanoscarchina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein: an amino acid L at position 485 of SEQ ID NO: 7 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 565 of SEQ ID NO: 7 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0096] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein: an amino acid M at position 426 of SEQ ID NO: 8 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0097] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein: an amino acid M at position 426 of SEQ ID NO: 8 is substituted with A, C, D, F, G, H, K, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at position 508 of SEQ ID NO: 8 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0098] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein: an amino acid L at position 518 of SEQ ID NO: 9 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; and an amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G.

[0099] In accordance with some embodiments, the representative RNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein: an amino acid L at position 518 of SEQ ID NO: 9 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, preferably Y; an amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V, preferably A; an amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V, preferably G; and an amino acid A at

position 601 of SEQ ID NO: 9 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y, preferably E or L.

[0100] In accordance with some embodiments, the representative RNA polymerase variant exhibits an activity of synthesizing nucleic acids in a template-independent manner by adding at least one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator.

[0101] In certain embodiments, the extendible initiator comprises a single-stranded oligonucleotide initiator, a blunt ended double-stranded oligonucleotide initiator, or a mixture thereof. In certain embodiments, the extendible initiator is a free form nucleic acid and can be reacted in a liquid phase.

[0102] In certain embodiments, the extendible initiator is immobilized on a solid support, wherein the solid support comprises a particle, bead, slide, array surface, membrane, flow cell, well, microwell, nano-well, chamber, microfluidic chamber, channel, microfluidic channel, or any other surfaces.

[0103] In certain embodiments, the at least one nucleotide comprises a ribose. Furthermore, the at least one nucleotide is linked with a detectable label, such as fluorophores, enzymes, radioactive phosphates, digoxigenin, or biotin.

[0104] In accordance with some embodiments, the representative RNA polymerase variant exhibits the template-independent nucleic acid synthesis activity at reaction temperatures ranging from 10° C. to 100° C. For example, the reaction temperature is between 10° C. and 20° C., between 20° C. and 30° C., between 30° C. and 40° C., between 40° C. and 50° C., between 50° C. and 60° C., between 60° C. and 70° C., between 70° C. and 80° C., between 80° C. and 90° C., between 90° C. and 95° C., between 95° C. and 100° C., or any reaction temperatures within a range defined by an upper limit of 15° C., 20° C., 25° C., 30° C., 35° C., 37° C., 40° C., 45° C., 50° C., 55° C., 60° C., 65° C., 70° C., 75° C., 80° C., 85° C., 90° C., 95° C., or 100° C. and an lower limit of 10° C., 15° C., 20° C., 25° C., 30° C., 35° C., 37° C., 40° C., 45° C., 50° C., 55° C., 60° C., 65° C., 70° C., 75° C., 80° C., 85° C., 90° C., or 95° C.

Creation of Polymerase Variants

[0105] Various types of mutagenesis techniques are optionally used in the present disclosure, e.g., to modify polymerases to create the variants of the subject application, or using random or semi-random mutational approaches. In general, any available mutagenesis procedure can be used for making polymerase mutants. Such mutagenesis procedures optionally include selection of altered nucleic acids and polypeptides for one or more activity of interest. Procedures that can be used include, but are not limited to: the site-directed point mutagenesis, random point mutagenesis, in vitro or in vivo homologous recombination (DNA shuffling and combinatorial overlap PCR), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to skilled person.

Kit for Performing Template-Independent Nucleic Acid Synthesis Reaction

[0106] The present invention also provides a kit that includes the RNA polymerase variant described herein, for performing de novo enzymatic nucleic acid synthesis reaction, comprising: an RNA polymerase variant as described above, wherein the RNA polymerase variant exhibits activity of synthesizing nucleic acids in a template-independent manner by adding at least one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator, thereby synthesizing a desired nucleic acid sequence.

[0107] Optionally, other reagents such as buffers and solutions required for the RNA polymerase variant and nucleotide solution are also included. Instructions for use of the assembled or packaged components are also typically but not necessarily included.

Methods and Uses of RNA Polymerase Variant

[0108] In some embodiments, the RNA polymerase variants described herein can be used to add natural ribonucleotides (rNTP), such as rATP, rUTP, rCTP and rGTP, to the 3'-hydroxyl (3'-OH) terminus of a single-stranded or a blunt-end, duplex nucleic acids initiator in a template-independent synthesis manner to produce polynucleotides with desired sequences.

[0109] In some embodiments, the RNA polymerase variants described herein can be used to add natural ribonucleotides (rNTP), such as rATP, rUTP, rCTP and rGTP, to the 3'-OH termini of arrays of clustered single-stranded or a blunt-end, duplex nucleic acids initiators, which are immobilized or physically confined, and separated on a solid support as described previously; and preferably, the solid support is made of glass and implemented in the form of silicon wafer. Thus, a multiplexing, parallel de novo nucleic acid synthesis can be performed to synthesize large numbers of various polynucleotides or nucleic acids with distinct sequences.

[0110] In certain embodiments, the RNA polymerase variants described herein can be used to incorporate the nucleotide conjugates (one of the types of nucleotide analogue defined previously) covalently linked with an enzyme, an antibody, a chemical group, such as a biotin, a desthiobiotin, or a fluorophore on the base, phosphate moiety, or pentose sugar of nucleotide, to the 3'-terminus of the nucleic acid initiator in a template-independent synthesis manner.

[0111] The incorporation of these nucleotide analogues into the nucleic acids by RNA polymerase variants during the nucleic acid synthesis concurrently add the desired component, such as an associated enzyme, an antibody, or a chemical group to the newly synthesized nucleic acids in a base-specific, site-specific, or sequence-specific manner. Common components used to label or generate nucleic acid probes and conjugates are known in the art, which include, but are not limited to, radiolabeled nucleotides and nucleotide analogues, modified linkers, such as a biotin, a thiol, an azido, or an amine group, fluorophores, enzymes, and antibodies.

[0112] Alternatively, in other embodiments, to label or generate nucleic acid probes, the post-synthetic modifications of nucleic acids can be achieved by covalently or non-covalently coupling with an enzyme, an antibody, a chemical group, or a fluorophore via a modified linker on the

base, the phosphate moiety, or the pentose sugar of synthesis nucleotide. As a result, the desired component can be covalently or non-covalently associated with the specific base or connected to the 5'- or 3'-terminus of newly synthesized nucleic acids.

[0113] In some embodiments, the RNA polymerase variant-dependent incorporation of linker-modified nucleotide analogues may be used to facilitate the newly synthesized polynucleotides or nucleic acids to be attached, immobilized or physically confined on various solid surfaces. Retrospectively, in other embodiments, the newly synthesized sequence-specific nucleic acids with unique labels, tags, or fluorophores can be used in various nucleic acid-based molecular detections, which include, but are not limited to, the fluorescence in situ hybridization (FISH), TaqMan real-time PCR (RT-PCR), real-time fluorescence ligase chain reaction (RT-LCR), real-time fluorescence recombinase-polymerase amplification (RPA) assay, and real-time fluorescence loop-mediated isothermal amplification assay.

[0114] The present disclosure further provides a method for template-independent synthesis of an RNA oligonucleotide, comprising:

- [0115]** (a) providing an initiator oligonucleotide,
- [0116]** (b) providing an RNA polymerase variant; and
- [0117]** (c) combining the initiator oligonucleotide, the RNA polymerase variant and one or more nucleotides under conditions sufficient for the addition of at least one nucleotide to the 3' end of the initiator oligonucleotide.

[0118] Once the one or more nucleotides are added to the initiator oligonucleotide, one or more additional nucleotides can be added subsequently in order to synthesize a desired RNA oligonucleotide. Therefore, in certain embodiments, the method further comprises adding one or more natural or modified nucleotides to the 3' end of the resulting RNA oligonucleotide (i.e., the RNA oligonucleotide formed in step (c)) until a desired RNA sequence is obtained. In certain embodiments, the method further comprises: (d) repeating steps (a)-(c) until a desired RNA sequence is obtained.

[0119] In certain embodiments, step (c) is carried out in the presence of one or more additional enzymes. In certain embodiments, step (c) is carried out in the presence of a mixture of two or more different enzymes. The mixture of enzymes may comprise more than one distinct RNA polymerase variants (e.g., 2 or 3 RNA polymerase variants).

[0120] In certain embodiments, step (c) is carried out in the presence of one or more additional enzyme (i.e., auxiliary enzyme), such as specific phosphatases in addition to the RNA polymerase variant. In certain embodiments, step (c) is carried out in the presence of a yeast inorganic pyrophosphatase (PPI-ase) in addition to the RNA polymerase variant.

[0121] In certain embodiments, the reaction in step (c) is carried out in the presence of one or more additional additives. In certain embodiments, step (c) is carried out in the presence of a crowding agent. In certain embodiments, the crowding agent is polyethylene glycol (PEG) or Ficoll. In certain embodiments, the crowding agent is polyethylene glycol (PEG). In certain embodiments, step (c) is carried out in the presence of an RNase inhibitor. In certain embodiments, step (c) is carried out in the presence of a non-hydrolyzable nucleotide.

EXAMPLES

Example 1: Preparation of RNA Polymerase Variants

[0122] The gene synthesis approach and mutagenesis technique are adapted to create exemplary RNA polymerase variants according to the properties of conserved/consensus amino acids in the conserved and semi-conserved regions of selective PolBs, which are disclosed herein. For instance, the well-known site-directed mutagenesis approach is conducted to change the amino acid residues in the motif Exo I, motif Exo II, motif Exo III, motif A, motif B, and motif C regions of an exemplary wild-type PolB listed herein.

[0123] In some embodiments, the procedure for obtaining exemplary RNA polymerase variants is generally divided into three steps, including Step 1: Gene synthesis of wild-type PolB and its 3' to 5' exonuclease-deficient (Exo⁻) mutant, Step 2: Construction of the specific exemplary RNA polymerase variant in the desired region, and Step 3: Expression and purification of wild-type PolB, Exo⁻ mutant, and RNA polymerase variant. As described in more detail below, the techniques used in said procedure are well-known to those skilled in the art.

[0124] In Step 1, the codon-optimized gene fragment encoding the wild-type, intein-free PolB polymerase is synthesized by Genomics BioSci & Tech Co. (New Taipei City, Taiwan). The 3' to 5' exonuclease-deficient (designated as Exo⁻) PolB polymerase is also provided by the same vendor. The superscript "Exo⁻" following an abbreviated name of any PolB listed herein means that the designated wild-type PolB has been modified to eliminate the intrinsic 3' to 5' exonuclease activity, which indicates said PolB is a 3' to 5' exonuclease-deficient PolB. Preferably, in the Examples of this disclosure, the Exo means a PolB mutant carrying combinatory mutations at the positions corresponding to D354 of SEQ ID NO: 1, which is substituted with an alanine residue (D354A), and E356 of SEQ ID NO: 1, which is also substituted with an alanine residue (E356A), respectively.

[0125] In Step 2, the synthetic wild-type and Exo PolB genes are respectively subcloned into the pET28b vector using the flanking NdeI and NotI restriction sites. The sequences of recombinant plasmids are confirmed by DNA sequencing. To create the RNA polymerase variant at the desired motif region of the PolB Exo protein backbone, the site-directed mutagenesis, is conducted. Briefly, the site-directed mutagenesis PCR is performed with the recombinant plasmids using the Q5 Site-directed Mutagenesis Kit from New England Biolabs (Ipswich, MA) to introduce the amino acid substitution. The products are first analyzed by 1% agarose gel to confirm the amplicon size and the rest of PCR reaction mixture is then treated with DpnI at 37° C. for an hour. The mixture is further incubated at 70° C. for 10 mins to inactivate the DpnI function. The DpnI-treated PCR reaction mixture is then purified by the Qiagen's QIAquick PCR Purification Kit (Whatman, MA). The purified DNA fragment is treated with the mixture of T4 PNK and T4 DNA ligase. The re-circularized PCR-amplified DNA is transformed back into the *E. coli* cells. The plasmid DNA was later extracted from the *E. coli* cells using the Qiagen Plasmid Mini Kit (Whatman, MA). The mutated sequences of the polymerase variants at the desired motif region, or regions, are confirmed by DNA sequencing.

[0126] In Step 3, *E. coli* Acella cells harboring the plasmid DNA carrying specific polymerase variant gene are grown in 2 L of LB medium supplemented with 0.5% glucose and 50 µg/ml carbenicillin at 37° C. When the cell density reaches an absorbance value at OD_{600 nm} around 0.6-0.8, an 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to induce protein expression. Cells are grown for additional 4 hours at 37° C. and then harvested by centrifugation at 4° C. for 10 min at 7,000×g. Cell pellets are resuspended with buffer A [50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol] containing 1 mM benzamidinium hydrochloride. Cell lysis is achieved by incubation with 50 mg of lysozyme on ice for 1 hour followed by sonication. The cell lysate is clarified by centrifugation at 18,000×g for 25 min at 4° C. The clarified crude cell extract is incubated at 70° C. for 30 minutes and then cooled down at 4° C. The heat-treated cell extract is further clarified by centrifugation at 18,000×g for 25 minutes at 4° C. After centrifugation, the supernatant is diluted with buffer A without NaCl and loaded onto a HiTrap Heparin column (Cytiva Life Sciences, Marlborough, MA, USA) pre-equili-

brated in buffer A in the AKTA pure chromatography system (Cytiva Life Sciences, Marlborough, MA, USA). The protein is eluted with the linear 100 mM to 1 M NaCl gradient using the buffer B [50 mM Tris-HCl (pH 8.0), 1 M NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol]. Column fractions are analyzed by 10% SDS-PAGE. Fractions containing desired protein are pooled and dialyzed against the storage buffer [50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol] at 4° C. overnight. The dialyzed protein fraction pool containing the target protein is concentrated using an Amicon filter unit (MW cut-off 50,000). The concentrated protein pool is aliquoted and stored at -20° C. Each mutant polymerase variant was purified with the same procedures as described above. The final protein concentration is determined by the Bradford reaction (Bradford, 1976) using the Bio-Rad Protein Assay (Hercules, CA) with bovine serum albumin as a standard. **[0127]** In this example, selected (exemplary) Exo RNA polymerase variants aimed for the following assays with functionally or positionally substitutions residing in the motif A and/or motif B are summarized and listed in Table 1

TABLE 1

List of amino acid substitutions in the exemplary RNA polymerase variants					
Type of PolB Enzymes	SEQ ID NO	Equivalent substitutions in the motif A and/or motif B corresponding to the consensus sequence (SEQ ID NO: 1)			
		Position 715 (motif A)	Position 716 (motif A)	Position 717 (motif A)	Position 854 (motif B)
Tgo	2	L408A, L408C, L408D, L408F, L408G, L408H, L408K, L408M, L408N, L408Q, L408S, L408W, L408Y	Y409A, Y409C, Y409D, Y409G, Y409N, Y409S, Y409T, Y409V	P410A, P410C, P410G, P410I, P410L, P410M, P410N, P410S, P410T, P410V	A485C, A485D, A485E, A485F, A485G, A485H, A485I, A485K, A485L, A485M, A485N, A485P, A485Q, A485R, A485T, A485V, A485W, A485Y
Kod1	3	L408A, L408C, L408D, L408F, L408G, L408H, L408K, L408M, L408N, L408Q, L408S, L408W, L408Y	Y409A, Y409C, Y409D, Y409G, Y409N, Y409S, Y409T, Y409V	P410A, P410C, P410G, P410I, P410L, P410M, P410N, P410S, P410T, P410V	A485C, A485D, A485E, A485F, A485G, A485H, A485I, A485K, A485L, A485M, A485N, A485P, A485Q, A485R, A485T, A485V, A485W, A485Y
9°N	4	L408A, L408C, L408D, L408F, L408G, L408H, L408K, L408M, L408N, L408Q, L408S, L408W, L408Y	Y409A, Y409C, Y409D, Y409G, Y409N, Y409S, Y409T, Y409V	P410A, P410C, P410G, P410I, P410L, P410M, P410N, P410S, P410T, P410V	A485C, A485D, A485E, A485F, A485G, A485H, A485I, A485K, A485L, A485M, A485N, A485P, A485Q, A485R, A485T, A485V, A485W, A485Y
Pfu	5	L409A, L409C, L409D, L409F, L409G, L409H, L409K, L409M, L409N, L409Q, L409S, L409W, L409Y	Y410A, Y410C, Y410D, Y410G, Y410N, Y410S, Y410T, Y410V	P411A, P411C, P411G, P411I, P411L, P411M, P411N, P411S, P411T, P411V	A486C, A486D, A486E, A486F, A486G, A486H, A486I, A486K, A486L, A486M, A486N, A486P, A486Q, A486R, A486T, A486V, A486W, A486Y
Vent	6	L411A, L411C, L411D, L411F, L411G, L411H, L411K, L411M, L411N, L411Q,	Y412A, Y412C, Y412D, Y412G, Y412N, Y412S, Y412T, Y412V	P413A, P413C, P413G, P413I, P413L, P413M, P413N, P413S, P413T, P413V	A488C, A488D, A488E, A488F, A488G, A488H, A488I, A488K, A488L, A488M,

TABLE 1-continued

List of amino acid substitutions in the exemplary RNA polymerase variants					
Type of PolB Enzymes	SEQ ID NO	Equivalent substitutions in the motif A and/or motif B corresponding to the consensus sequence (SEQ ID NO: 1)			
		Position 715 (motif A)	Position 716 (motif A)	Position 717 (motif A)	Position 854 (motif B)
Mac	7	L411S, L411W, L411Y			A488N, A488P, A488Q, A488R, A488T, A488V, A488W, A488Y
		L485A, L485C, L485D, L485F, L485G, L485H, L485K, L485M, L485N, L485Q, L485S, L485W, L485Y	Y486A, Y486C, Y486D, Y486G, Y486N, Y486S, Y486T, Y486V	P487A, P487C, P487G, P487I, P487L, P487M, P487N, P487S, P487T, P487V	A565C, A565D, A565E, A565F, A565G, A565H, A565I, A565K, A565L, A565M, A565N, A565P, A565Q, A565R, A565T, A565V, A565W, A565Y
Pis	8	M426A, M426C, M426D, M426F, M426G, M426H, M426K, M426M, M426N, M426Q, M426S, M426W, M426Y	Y427A, Y427C, Y427D, Y427G, Y427N, Y427S, Y427T, Y427V	P428A, P428C, P428G, P428I, P428L, P428M, P428N, P428S, P428T, P428V	A508C, A508D, A508E, A508F, A508G, A508H, A508I, A508K, A508L, A508M, A508N, A508P, A508Q, A508R, A508T, A508V, A508W, A508Y
Sso	9	L518A, L518C, L518D, L518F, L518G, L518H, L518K, L518M, L518N, L518Q, L518S, L518W, L518Y	Y519A, Y519C, Y519D, Y519G, Y519N, Y519S, Y519T, Y519V	P520A, P520C, P520G, P520I, P520L, P520M, P520N, P520S, P520T, P520V	A601C, A601D, A601E, A601F, A601G, A601H, A601I, A601K, A601L, A601M, A601N, A601P, A601Q, A601R, A601T, A601V, A601W, A601Y

Example 2: Template-Independent RNA Synthesis Assay

[0128] The RNA polymerase variants provided herein are evaluated for template-independent RNA synthesis approach. To further determine the activities (performance on incorporating naturally occurring ribonucleotides) of the RNA polymerase variants, normal rNTPs and a single-stranded DNA initiator or a blunt-end duplex DNA initiator are used herein. Besides, the reaction temperature of said approach is set up as 55° C. to evaluate the heat-resistance of the exemplary RNA polymerase variants.

[0129] In this example, the following synthetic oligonucleotides are used as the initiators to determine the template-independent RNA synthesis activity of RNA polymerase variants.

[0130] FAM-45-mer ssDNA initiator for Mode I assay: a single-stranded DNA having the sequence of 5'-CTCGGCCTGGCACAGGTCCGTTTCAGTGCTGC GGCGACCACCGAGG-3' (SEQ ID NO: 18). This single-stranded oligonucleotide is labeled with a fluorescent fluorescein amidite (FAM) dye at the 5'-end.

[0131] Blunt-end duplex DNA initiator for Mode II assay: a duplex DNA composed of the FAM-45-mer ssDNA initiator pre-annealed with its complementary 45-mer oligonucleotide.

[0132] The blunt-end duplex DNA initiator is formed by annealing the FAM-45-mer ssDNA initiator primer with the

complementary 45-mer DNA at a molar ratio of 1:1.5 in the 1×TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] containing 100 mM NaCl. The DNA annealing reaction is performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98° C. for 3 minute and then gradually cooling it down (5° C./30 seconds) to 4° C. The annealing product without overhang is used as the blunt-end duplex DNA initiator.

[0133] The template-independent RNA synthesis reaction is performed in the reaction mixtures (10 µl) containing 100 nM FAM-45-mer ssDNA initiator (Mode I assay) or the blunt-end duplex DNA initiator (Mode II assay), 0.25 mM manganese chloride (MnCl₂), and 200 nM exemplary RNA polymerase variant. The de novo enzymatic RNA synthesis reactions was initiated by the addition of 100 µM of rNTPs. The reactions were allowed to proceed for a defined period of time (e.g., 2 minutes for the Mode I assay and 10 minutes for the Mode II assay) and then terminated by adding 10 µl of 2× quench solution (95% de-ionized formamide and 25 mM EDTA) at a selected reaction temperature (e.g., 55° C. for either the Mode I or Mode II assay). After either the reaction of Mode I or Mode II assay, the sample mixtures were first denatured at 95° C. for 10 min and analyzed by 20% polyacrylamide gel electrophoresis containing 8M urea (Urea-PAGE). The de novo enzymatic RNA synthesis reaction products are then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

[0134] Based on the assays described above, the relative template-independent RNA synthesis activity of each variant is scored and represented by the number of symbol “+”. The overall activity score for each variant is divided into 4 distinct levels: 1) the “+++” indicates that the initiator is completely elongated to various lengths of newly synthesized RNA as compared to the band intensity and position of the substrate control. Hence, the variant is considered to possess an 100% of RNA synthesis activity; 2) the “++” indicates that the initiator is elongated around 50% to 100% to various lengths of newly synthesized RNA as compared to the band intensity and position of the substrate control. Hence, the variant is considered to possess a 50% to an 100% of RNA synthesis activity; 3) the “+” indicates that the initiator is elongated around 10% to 50% to various lengths of newly synthesized RNA as compared to the band intensity and position of the substrate control. Hence, the variant is considered to possess a 10% to 50% of RNA synthesis activity; and 4) the “+/-” indicates that the initiator is elongated less than 10% to various lengths of newly synthesized RNA as compared to the band intensity and position of the substrate control. Therefore, the variant is considered to possess <10% of RNA synthesis activity. This principle for activity scoring is applied throughout the present disclosure and resulting data are primarily listed in the corresponding tables.

Example 3: Catalytic Activity of RNA Polymerase
Variants on Incorporating rNTPs to the
FAM-45-Mer ssDNA Initiator and Blunt-End
Duplex DNA Initiator

[0135] In this section, the selected exonuclease-deficient RNA polymerase variants (e.g., Tgo^{exo-}, Kod1^{exo-}, 9^N^{exo-}, Pfu^{exo-}, Vent^{exo-}, Mac^{exo-} and Sso^{exo-}) were modified to include on or more amino acid substitutions with different amino acids in varied conserved regions or motifs of each protein.

[0136] Additionally, in the preliminary screening, the inventor has discovered that a conserved motif of those exonuclease-deficient RNA polymerase variants, which is functionally and positionally equivalent to the L715, Y716, and P717 residing in the motif A of the consensus sequence (SEQ ID NO: 1) as defined herein, may be predominantly

related with the function of de novo RNA synthesis. More specifically, said conserved motif may be an essential site for template-independent RNA synthesis activity. Furthermore, a conserved residue residing in the motif B, which is functionally and positionally equivalent to the A854 of the consensus sequence (SEQ ID NO: 1), serves as a reinforceable site for enhancing the template-independent RNA synthesis activity (data not shown).

Example 3.1: Template-Independent RNA Synthesis
Activity of Vent Variants

[0137] In this example, the RNA polymerase variants derived from Vent (SEQ ID NO: 6) is used exemplarily for evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B. Additionally, the United States Patent NO. U.S. Ser. No. 11/136,564B2 disclosed an AAI motif for substituting said conserved motif of some archaeal DNA polymerases to improve the incorporation of nucleotide analogues for template-dependent DNA synthesis reactions (i.e., DNA sequencing). The conserved motif AAI is functionally and positionally equivalent to the L715, Y716, and P717 of the consensus sequence (SEQ ID NO: 1), therefore, the conserved motif is also functionally and positionally equivalent to the L411, Y412 and P413 residing in the motif A of the wild-type Vent (SEQ ID NO: 6). Thus, in view of the effects of the AAI motif on the template-directed nucleotide incorporation, the AAI motif substitution is equivalently included in this example for comparison. Moreover, in this example, the combinatory effect of substitution of A854 residing in the motif B of the consensus sequence, which is functionally and positionally equivalent to the A488 residing in the motif B of the wild-type Vent, is also evaluated.

[0138] In this example, the variants modified from Vent Exo backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I assay are shown in Table 2.1 and FIG. 2A; and the results of Mode II assay are shown in Table 2.2 and FIG. 2B, where “S” denoted in the figures stands for the substrate (FAM-45-mer ssDNA initiator or blunt-end duplex DNA initiator) and serves as a blank DNA control. Besides, for the sake of brevity, only the exemplary results for the representative Vent variants are shown in FIGS. 2A and 2B.

TABLE 2.1

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Vent variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
V01	D141A + E143A	-
V02	D141A + E143A + A488L	-
V03	D141A + E143A + P413E	+/-
V04	D141A + E143A + L411Y	+/-
V05	D141A + E143A + L411A + A488L	+/-
V06	D141A + E143A + L411C + A488L	+/-
V07	D141A + E143A + L411D + A488L	+/-
V08	D141A + E143A + L411F + A488L	++
V09	D141A + E143A + L411G + A488L	+/-
V10	D141A + E143A + L411H + A488L	++
V11	D141A + E143A + L411K + A488L	+/-
V12	D141A + E143A + L411M + A488L	+/-
V13	D141A + E143A + L411Q + A488L	+/-
V14	D141A + E143A + L411Y + A488L	++
V15	D141A + E143A + Y412A + A488L	+
V16	D141A + E143A + Y412C + A488L	+/-

TABLE 2.1-continued

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Vent variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
V17	D141A + E143A + Y412G + A488L	+
V18	D141A + E143A + Y412N + A488L	+/-
V19	D141A + E143A + Y412S + A488L	+
V20	D141A + E143A + P413S + A488L	++
V21	D141A + E143A + L411A + Y412A + P413G + A488L	++
V22	D141A + E143A + L411C + Y412A + P413G + A488L	+++
V23	D141A + E143A + L411D + Y412A + P413G + A488L	+
V24	D141A + E143A + L411E + Y412A + P413G + A488L	+
V25	D141A + E143A + L411F + Y412A + P413G + A488L	+++
V26	D141A + E143A + L411G + Y412A + P413G + A488L	+
V27	D141A + E143A + L411H + Y412A + P413G + A488L	+++
V28	D141A + E143A + L411I + Y412A + P413G + A488L	+
V29	D141A + E143A + L411K + Y412A + P413G + A488L	+++
V30	D141A + E143A + Y412A + P413G + A488L	+
V31	D141A + E143A + L411M + Y412A + P413G + A488L	+++
V32	D141A + E143A + L411N + Y412A + P413G + A488L	++
V33	D141A + E143A + L411Q + Y412A + P413G + A488L	+++
V34	D141A + E143A + L411S + Y412A + P413G + A488L	++
V35	D141A + E143A + L411T + Y412A + P413G + A488L	+
V36	D141A + E143A + L411V + Y412A + P413G + A488L	+
V37	D141A + E143A + L411W + Y412A + P413G + A488L	++
V38	D141A + E143A + L411Y + Y412A + P413G + A488L	++
V39	D141A + E143A + L411Y + Y412C + P413G + A488L	+
V40	D141A + E143A + L411Y + Y412D + P413G + A488L	++
V41	D141A + E143A + L411Y + Y412F + P413G + A488L	+
V42	D141A + E143A + L411Y + Y412G + P413G + A488L	++
V43	D141A + E143A + L411Y + Y412H + P413G + A488L	+
V44	D141A + E143A + L411Y + Y412I + P413G + A488L	+
V45	D141A + E143A + L411Y + Y412L + P413G + A488L	+
V46	D141A + E143A + L411Y + Y412M + P413G + A488L	+
V47	D141A + E143A + L411Y + Y412N + P413G + A488L	++
V48	D141A + E143A + L411Y + Y412Q + P413G + A488L	+
V49	D141A + E143A + L411Y + Y412S + P413G + A488L	++
V50	D141A + E143A + L411Y + Y412T + P413G + A488L	++
V51	D141A + E143A + L411Y + Y412V + P413G + A488L	++
V52	D141A + E143A + L411Y + P413G + A488L	++
V53	D141A + E143A + L411Y + Y412A + P413A + A488L	++
V54	D141A + E143A + L411Y + Y412A + P413C + A488L	++
V55	D141A + E143A + L411Y + Y412A + P413D + A488L	+
V56	D141A + E143A + L411Y + Y412A + P413E + A488L	+
V57	D141A + E143A + L411Y + Y412A + P413F + A488L	+
V58	D141A + E143A + L411Y + Y412A + P413H + A488L	+
V59	D141A + E143A + L411Y + Y412A + P413I + A488L	++
V60	D141A + E143A + L411Y + Y412A + P413K + A488L	+
V61	D141A + E143A + L411Y + Y412A + P413L + A488L	++
V62	D141A + E143A + L411Y + Y412A + P413M + A488L	++
V63	D141A + E143A + L411Y + Y412A + P413N + A488L	++
V64	D141A + E143A + L411Y + Y412A + A488L	++
V65	D141A + E143A + L411Y + Y412A + P413Q + A488L	+
V66	D141A + E143A + L411Y + Y412A + P413S + A488L	++
V67	D141A + E143A + L411Y + Y412A + P413T + A488L	++
V68	D141A + E143A + L411Y + Y412A + P413V + A488L	++
V69	D141A + E143A + L411Y + Y412A + P413W + A488L	+
V70	D141A + E143A + L411Y + Y412A + P413Y + A488L	+
V71	D141A + E143A + L411Y + Y412A + P413G + A488A	+
V72	D141A + E143A + L411Y + Y412A + P413G + A488C	++
V73	D141A + E143A + L411Y + Y412A + P413G + A488D	++
V74	D141A + E143A + L411Y + Y412A + P413G + A488E	++
V75	D141A + E143A + L411Y + Y412A + P413G + A488F	++
V76	D141A + E143A + L411Y + Y412A + P413G + A488G	++
V77	D141A + E143A + L411Y + Y412A + P413G + A488H	++
V78	D141A + E143A + L411Y + Y412A + P413G + A488I	++
V79	D141A + E143A + L411Y + Y412A + P413G + A488K	++
V80	D141A + E143A + L411Y + Y412A + P413G + A488M	++
V81	D141A + E143A + L411Y + Y412A + P413G + A488N	++
V82	D141A + E143A + L411Y + Y412A + P413G + A488P	+
V83	D141A + E143A + L411Y + Y412A + P413G + A488Q	++
V84	D141A + E143A + L411Y + Y412A + P413G + A488R	++

TABLE 2.1-continued

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Vent variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
V85	D141A + E143A + L411Y + Y412A + P413G + A488T	++
V86	D141A + E143A + L411Y + Y412A + P413G + A488V	++
V87	D141A + E143A + L411Y + Y412A + P413G + A488W	++
V88	D141A + E143A + L411Y + Y412A + P413G + A488Y	++

TABLE 2.2

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Vent variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
V02	D141A + E143A + A488L	–
V05	D141A + E143A + L411A + A488L	+++
V06	D141A + E143A + L411C + A488L	+++
V07	D141A + E143A + L411D + A488L	+++
V08	D141A + E143A + L411F + A488L	+++
V09	D141A + E143A + L411G + A488L	–/+
V10	D141A + E143A + L411H + A488L	++
V11	D141A + E143A + L411K + A488L	–/+
V12	D141A + E143A + L411M + A488L	++
V13	D141A + E143A + L411Q + A488L	++
V14	D141A + E143A + L411Y + A488L	+++
V16	D141A + E143A + Y412C + A488L	+++
V17	D141A + E143A + Y412G + A488L	+++
V18	D141A + E143A + Y412N + A488L	+++
V19	D141A + E143A + Y412S + A488L	+++
V20	D141A + E143A + P413S + A488L	+++
V38	D141A + E143A + L411Y + Y412A + P413G + A488L	+++
V21	D141A + E143A + L411A + Y412A + P413G + A488L	+++
V22	D141A + E143A + L411C + Y412A + P413G + A488L	+++
V23	D141A + E143A + L411D + Y412A + P413G + A488L	+++
V24	D141A + E143A + L411E + Y412A + P413G + A488L	+++
V25	D141A + E143A + L411F + Y412A + P413G + A488L	+++
V26	D141A + E143A + L411G + Y412A + P413G + A488L	+++
V27	D141A + E143A + L411H + Y412A + P413G + A488L	+++
V28	D141A + E143A + L411I + Y412A + P413G + A488L	+++
V29	D141A + E143A + L411K + Y412A + P413G + A488L	+++
V30	D141A + E143A + Y412A + P413G + A488L	+++
V31	D141A + E143A + L411M + Y412A + P413G + A488L	+++
V32	D141A + E143A + L411N + Y412A + P413G + A488L	+++
V33	D141A + E143A + L411Q + Y412A + P413G + A488L	+++
V34	D141A + E143A + L411S + Y412A + P413G + A488L	+++
V35	D141A + E143A + L411T + Y412A + P413G + A488L	+++
V36	D141A + E143A + L411V + Y412A + P413G + A488L	+++
V37	D141A + E143A + L411W + Y412A + P413G + A488L	+++
V52	D141A + E143A + L411Y + P413G + A488L	+++
V39	D141A + E143A + L411Y + Y412C + P413G + A488L	+++
V40	D141A + E143A + L411Y + Y412D + P413G + A488L	+++
V41	D141A + E143A + L411Y + Y412F + P413G + A488L	+++
V43	D141A + E143A + L411Y + Y412H + P413G + A488L	+++
V44	D141A + E143A + L411Y + Y412I + P413G + A488L	+++
V45	D141A + E143A + L411Y + Y412L + P413G + A488L	+++
V46	D141A + E143A + L411Y + Y412M + P413G + A488L	+++
V47	D141A + E143A + L411Y + Y412N + P413G + A488L	+++
V48	D141A + E143A + L411Y + Y412Q + P413G + A488L	+++
V49	D141A + E143A + L411Y + Y412S + P413G + A488L	+++
V50	D141A + E143A + L411Y + Y412T + P413G + A488L	+++
V51	D141A + E143A + L411Y + Y412V + P413G + A488L	+++
V64	D141A + E143A + L411Y + Y412A + A488L	+++
V53	D141A + E143A + L411Y + Y412A + P413A + A488L	+++
V54	D141A + E143A + L411Y + Y412A + P413C + A488L	+++
V55	D141A + E143A + L411Y + Y412A + P413D + A488L	+++
V56	D141A + E143A + L411Y + Y412A + P413E + A488L	+++
V57	D141A + E143A + L411Y + Y412A + P413F + A488L	+++
V58	D141A + E143A + L411Y + Y412A + P413H + A488L	+++
V59	D141A + E143A + L411Y + Y412A + P413I + A488L	+++
V60	D141A + E143A + L411Y + Y412A + P413K + A488L	+++
V61	D141A + E143A + L411Y + Y412A + P413L + A488L	+++

TABLE 2.2-continued

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Vent variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
V62	D141A + E143A + L411Y + Y412A + P413M + A488L	+++
V63	D141A + E143A + L411Y + Y412A + P413N + A488L	+++
V65	D141A + E143A + L411Y + Y412A + P413Q + A488L	+++
V66	D141A + E143A + L411Y + Y412A + P413S + A488L	+++
V67	D141A + E143A + L411Y + Y412A + P413T + A488L	+++
V68	D141A + E143A + L411Y + Y412A + P413V + A488L	+++
V69	D141A + E143A + L411Y + Y412A + P413W + A488L	+++
V70	D141A + E143A + L411Y + Y412A + P413Y + A488L	+++
V72	D141A + E143A + L411Y + Y412A + P413G + A488C	+++
V73	D141A + E143A + L411Y + Y412A + P413G + A488D	+++
V74	D141A + E143A + L411Y + Y412A + P413G + A488E	+++
V75	D141A + E143A + L411Y + Y412A + P413G + A488F	+++
V77	D141A + E143A + L411Y + Y412A + P413G + A488H	+++
V78	D141A + E143A + L411Y + Y412A + P413G + A488I	+++
V79	D141A + E143A + L411Y + Y412A + P413G + A488K	+++
V80	D141A + E143A + L411Y + Y412A + P413G + A488M	+++
V81	D141A + E143A + L411Y + Y412A + P413G + A488N	+++
V82	D141A + E143A + L411Y + Y412A + P413G + A488P	+++
V83	D141A + E143A + L411Y + Y412A + P413G + A488Q	+++
V84	D141A + E143A + L411Y + Y412A + P413G + A488R	+++
V85	D141A + E143A + L411Y + Y412A + P413G + A488T	+++
V86	D141A + E143A + L411Y + Y412A + P413G + A488V	+++
V87	D141A + E143A + L411Y + Y412A + P413G + A488W	+++
V88	D141A + E143A + L411Y + Y412A + P413G + A488Y	+++

[0139] As shown in Table 2.1, Table 2.2, FIG. 2A, and FIG. 2B, the variants carrying amino acid substitutions in the motif Exo I, motif A and/or motif B (A488L) have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays. More specifically, the preferable substitution combinations occurs in the motif A, such as L411A+Y412A+P413G (AAG), L411C+Y412A+P413G (CAG), L411F+Y412A+P413G (FAG), L411H+Y412A+P413G (HAG), L411K+Y412A+P413G (KAG), L411M+Y412A+P413G (MAG), L411Q+Y412A+P413G (QAG), and L411Y+Y412A+P413G (YAG) for the catalytic activity of template-independent enzymatic RNA synthesis.

Example 3.2: Template-Independent RNA Synthesis Activity of Pfu Variants

[0140] In this example, the RNA polymerase variants derived from Pfu (SEQ ID NO: 5) is used exemplarily for evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0141] Specifically, the variants modified from Pfu Exo backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I assay are shown in Table 3.1 and FIG. 3A; and the results of Mode II assay are shown in Table 3.2 and FIG. 3B, where “S” denoted in the figures stands for the substrate (FAM-45-mer ssDNA initiator or blunt-end duplex DNA initiator) and serves as a blank DNA control. Besides, for the sake of brevity, only the exemplary results for the representative Pfu variants are shown in this example.

TABLE 3.1

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Pfu variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
P01	D141A + E143A	–
P02	D141A + E143A + A486L	+/-
P03	D141A + E143A + L409Y + Y410A + P411G	+
P04	D141A + E143A + L409Y + Y410A + P411G + A486L	+++
P05	D141A + E143A + L409A + Y410A + P411I + A486L	+++

TABLE 3.2

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Pfu variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
P01	D141A + E143A	–
P02	D141A + E143A + A486L	+/-
P03	D141A + E143A + L409Y + Y410A + P411G	+++
P04	D141A + E143A + L409Y + Y410A + P411G + A486L	+++
P05	D141A + E143A + L409A + Y410A + P411I + A486L	+++

[0142] As shown in FIG. 3A, FIG. 3B, Table 3.1 and Table 3.2, the variants carrying amino acid substitutions in the motif Exo I and motif A, such as variants P03, P04, and P05, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays. Moreover, the variants carrying combinatory amino acid substitutions in the motif Exo I, motif A, and motif B, such as variants P04 and P05, further enhanced the said catalytic activity.

Example 3.3: Template-Independent RNA Synthesis Activity of Kod1 Variants

[0143] In this example, the RNA polymerase variants derived from Kod1 (SEQ ID NO: 3) is used exemplarily for

evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0144] Specifically, the variants modified from Kod1 Exo⁻ backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I assay are shown in Table 4.1 and FIG. 4A; and the results of Mode II assay are shown in Table 4.2 and FIG. 4B, where “S” denoted in the figures stands for the substrate (FAM-45-mer ssDNA initiator or blunt-end duplex DNA initiator) and serves as a blank DNA control. Besides, for the sake of brevity, only the exemplary results for the representative Kod1 variants are shown in this example.

TABLE 4.1

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Kod1 variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
K01	D141A + E143A	–
K02	D141A + E143A + A485L	+
K03	D141A + E143A + L408Y + Y409A + P410G	++
K04	D141A + E143A + L408Y + Y409A + P410G + A485L	+++
K05	D141A + E143A + L408A + Y409A + P410I + A485L	++

TABLE 4.2

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Kod1 variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
K01	D141A + E143A	–
K02	D141A + E143A + A485L	+
K03	D141A + E143A + L408Y + Y409A + P410G	+++
K04	D141A + E143A + L408Y + Y409A + P410G + A485L	+++
K05	D141A + E143A + L408A + Y409A + P410I + A485L	+++

[0145] As shown in FIG. 4A, FIG. 4B, Table 4.1 and Table 4.2, the variants carrying amino acid substitutions in the motif Exo I, motif A and/or motif B, such as variants K02, K03, K04, and K05, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays. Moreover, the variants carrying combinatory amino acid substitutions in the motif Exo I, motif A, and motif B (A485L), such as K04 and K05, further enhanced the said catalytic activity.

Example 3.4: Template-Independent RNA Synthesis Activity of Mac Variants

[0146] In this example, the RNA polymerase variants derived from Mac (SEQ ID NO: 7) is used exemplarily for evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0147] Specifically, the variants modified from Mac Exo⁻ backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I assay are shown in Table 5.1 and FIG. 5A; and the results of Mode II assay are shown in Table 5.2 and FIG. 5B, where “S” denoted in the figures stands for the substrate (FAM-45-mer ssDNA initiator or blunt-end duplex DNA initiator) and serves as a blank DNA control. Besides, for the sake of brevity, only the exemplary results for the representative Mac variants are shown in this example.

[0148] As shown in FIG. 5A, FIG. 5B, Table 5.1 and Table 5.2, the variants carrying amino acid substitutions in the motif Exo I and motif A, such as variant M02, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays. Moreover, the variants carrying combinatory amino acid substitutions in the motif Exo I, motif A, and motif B, such as variant M03, further enhanced the said catalytic activity.

Example 3.5: Template-Independent RNA Synthesis Activity of Tgo Variants

[0149] In this example, the RNA polymerase variants derived from Tgo (SEQ ID NO: 2) is used exemplarily for evaluating the template-independent RNA synthesis activity the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0150] Specifically, the variants modified from Tgo Exo backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I assay are shown in Table 6.1 and FIG. 6A; and the results of Mode II assay are shown in Table 6.2 and FIG. 6B, where “S” denoted in the figures stands for the substrate (FAM-45-mer ssDNA initiator or blunt-end duplex DNA initiator) and serves as a blank DNA control. Besides, for the sake of brevity, only the exemplary results for the representative Tgo variants are shown in this example.

TABLE 5.1

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Mac variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
M01	D198A + E200A	–
M02	D198A + E200A + L485Y + Y486A + P487G	+/-
M03	D198A + E200A + L485Y + Y486A + P487G + A565L	+

TABLE 5.2

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Mac variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
M01	D198A + E200A	–
M02	D198A + E200A + L485Y + Y486A + P487G	+++
M03	D198A + E200A + L485Y + Y486A + P487G + A565L	+++

TABLE 6.1

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Tgo variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
T01	D141A + E143A + L408A + Y409A + P410A + A485L	+++
T02	D141A + E143A + L408Y + Y409A + P410G + A485L	+++

TABLE 6.2

List of amino acid substitutions and RNA synthesis activity scorings of the exemplary Tgo variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
T01	D141A + E143A + LA08A + Y409A + P410A + A485L	+++
T02	D141A + E143A + L408Y + Y409A + P410G + A485L	+++

[0151] As shown in FIG. 6A, FIG. 6B, Table 6.1 and Table 6.2, the variants carrying amino acid substitutions in motif Exo I, motif A, and motif B, such as variants T01 and T02, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays.

Example 3.6: Template-Independent RNA Synthesis
Activity of Sso Variants

[0152] In this example, the RNA polymerase variants derived from Sso (SEQ ID NO: 9) is used exemplarily for evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0153] Specifically, the variants modified from Sso Exo backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I and Mode II assays are shown in Tables 7.1 and 7.2 respectively. For the sake of brevity, the Urea-PAGE image results in this example is currently omitted.

[0154] As shown in the Tables 7.1 and 7.2, the variants carrying amino acid substitutions in motif Exo I, motif A, and/or motif B, such as variants S03 and S06, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays.

Example 3.7: Template-Independent RNA Synthesis
Activity of 9° N Variants

[0155] In this example, the RNA polymerase variants derived from 9N (SEQ ID NO: 4) is used exemplarily for evaluating the template-independent RNA synthesis activity of the variants carrying combinatory substitutions in the motif Exo I, the motif A, and the motif B.

[0156] Specifically, the variants modified from 9° N Exo⁻ backbone are evaluated using the Mode I and Mode II activity assays as described above. The results of Mode I and Mode II assays are shown in Tables 8.1 and 8.2 respectively. For the sake of brevity, the Urea-PAGE image results in this example is currently omitted.

TABLE 7.1

List of amino acid substitutions and activity scorings of the exemplary Sso variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
S01	D231A + E233A	–
S03	D231A + E233A + L518Y + Y519A + P520G	+++
S06	D231A + E233A + L518Y + Y519A + P520G + A601L	+++

TABLE 7.2

List of amino acid substitutions and activity scorings of the exemplary Sso variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
S01	D231A + E233A	–
S03	D231A + E233A + L518Y + Y519A + P520G	+++
S06	D231A + E233A + L518Y + Y519A + P520G + A601L	+++

TABLE 8.1

List of amino acid substitutions and activity scorings of the exemplary 9°N variants determined by Mode I activity assay		
Variant Name	Substitutions	Activity Scoring
N02	D141A + E143A + L408Y + Y409A + P410E + A485V	++
N03	D141A + E143A + L408Y + Y409A + P410F + A485V	++
N04	D141A + E143A + L408Y + Y409A + P410G + A485V	+++
N05	D141A + E143A + L408Y + Y409A + P410H + A485V	+++
N06	D141A + E143A + L408Y + Y409A + P410T + A485V	+++
N07	D141A + E143A + L408Y + Y409A + P410V + A485V	+++
N08	D141A + E143A + L408Y + Y409C + P410G + A485V	++
N09	D141A + E143A + L408Y + Y409G + P410G + A485V	++
N10	D141A + E143A + L408Y + Y409I + P410G + A485V	++
N11	D141A + E143A + L408Y + Y409K + P410G + A485V	+/-
N12	D141A + E143A + L408Y + Y409L + P410G + A485V	+/-
N13	D141A + E143A + L408Y + Y409Q + P410G + A485V	+/-
N14	D141A + E143A + L408Y + Y409Y + P410G + A485V	+++
N15	D141A + E143A + L408A + Y409A + P410G + A485V	++
N16	D141A + E143A + L408S + Y409A + P410G + A485V	++
N17	D141A + E143A + L408V + Y409A + P410G + A485V	+++

TABLE 8.2

List of amino acid substitutions and activity scorings of the exemplary 9°N variants determined by Mode II activity assay		
Variant Name	Substitutions	Activity Scoring
N02	D141A + E143A + L408Y + Y409A + P410E + A485V	+++
N03	D141A + E143A + L408Y + Y409A + P410F + A485V	+++
N04	D141A + E143A + L408Y + Y409A + P410G + A485V	+++
N05	D141A + E143A + L408Y + Y409A + P410H + A485V	+++
N06	D141A + E143A + L408Y + Y409A + P410T + A485V	+++
N07	D141A + E143A + L408Y + Y409A + P410V + A485V	+++
N08	D141A + E143A + L408Y + Y409C + P410G + A485V	+++
N09	D141A + E143A + L408Y + Y409G + P410G + A485V	+++
N10	D141A + E143A + L408Y + Y409I + P410G + A485V	+++
N11	D141A + E143A + L408Y + Y409K + P410G + A485V	+++
N12	D141A + E143A + L408Y + Y409L + P410G + A485V	+++
N13	D141A + E143A + L408Y + Y409Q + P410G + A485V	+++
N14	D141A + E143A + L408Y + Y409Y + P410G + A485V	+++
N15	D141A + E143A + L408A + Y409A + P410G + A485V	+++
N16	D141A + E143A + L408S + Y409A + P410G + A485V	+++
N17	D141A + E143A + L408V + Y409A + P410G + A485V	++

[0157] As shown in the Tables 8.1 and 8.2, the variants carrying amino acid substitutions in motif Exo I, motif A, and motif B, such as the variants listed in Tables 8.1 and 8.2, have exerted prominent catalytic activity of template-independent enzymatic RNA synthesis in both Mode I and Mode II assays.

[0158] In view of the results observed, the RNA polymerase variants and the kit provided herein have been further proven in various scenarios to use rNTPs effectively and efficiently for de novo enzymatic RNA synthesis. Furthermore, these RNA polymerase variants are also proven to successfully exert the conferred template-independent RNA synthesis function under broader reaction temperatures covering from atmospheric temperatures to the hyperthermal

conditions, demonstrating a higher thermotolerance. Therefore, the RNA polymerase variants and the kit within the scope of the present disclosure can broaden the scope of various applications of template-independent enzymatic nucleic acids synthesis in different reaction conditions.

[0159] The present disclosure has been described with embodiments thereof, and it is understood that various modifications, without departing from the scope of the present disclosure, are in accordance with the embodiments of the present disclosure. Hence, the embodiments described are intended to cover the modifications within the scope of the present disclosure, rather than to limit the present disclosure. The scope of the claims therefore should be accorded the broadest interpretation so as to encompass all such modifications.

SEQUENCE LISTING

Sequence total quantity: 18
 SEQ ID NO: 1 moltype = AA length = 1360
 FEATURE Location/Qualifiers

-continued

source	1..1360 mol_type = protein organism = synthetic construct
REGION	1..1360 note = consensus sequence
VARIANT	6..10 note = Xaa can be any naturally occurring amino acid
VARIANT	17..22 note = Xaa can be any naturally occurring amino acid
VARIANT	30..34 note = Xaa can be any naturally occurring amino acid
VARIANT	48..52 note = Xaa can be any naturally occurring amino acid
VARIANT	54..59 note = Xaa can be any naturally occurring amino acid
VARIANT	81..84 note = Xaa can be any naturally occurring amino acid
VARIANT	89..95 note = Xaa can be any naturally occurring amino acid
VARIANT	100..102 note = Xaa can be any naturally occurring amino acid
VARIANT	105..106 note = Xaa can be any naturally occurring amino acid
VARIANT	114..116 note = Xaa can be any naturally occurring amino acid
VARIANT	120..121 note = Xaa can be any naturally occurring amino acid
VARIANT	142..143 note = Xaa can be any naturally occurring amino acid
VARIANT	145..146 note = Xaa can be any naturally occurring amino acid
VARIANT	183..184 note = Xaa can be any naturally occurring amino acid
VARIANT	186..187 note = Xaa can be any naturally occurring amino acid
VARIANT	233..234 note = Xaa can be any naturally occurring amino acid
VARIANT	265..267 note = Xaa can be any naturally occurring amino acid
VARIANT	269..270 note = Xaa can be any naturally occurring amino acid
VARIANT	273..274 note = Xaa can be any naturally occurring amino acid
VARIANT	305..307 note = Xaa can be any naturally occurring amino acid
VARIANT	313..314 note = Xaa can be any naturally occurring amino acid
VARIANT	321..325 note = Xaa can be any naturally occurring amino acid
VARIANT	328..329 note = Xaa can be any naturally occurring amino acid
VARIANT	333..336 note = Xaa can be any naturally occurring amino acid
VARIANT	410..411 note = Xaa can be any naturally occurring amino acid
VARIANT	446..449 note = Xaa can be any naturally occurring amino acid
VARIANT	479..481 note = Xaa can be any naturally occurring amino acid
VARIANT	565..566 note = Xaa can be any naturally occurring amino acid
VARIANT	572..573 note = Xaa can be any naturally occurring amino acid
VARIANT	671..673 note = Xaa can be any naturally occurring amino acid
VARIANT	681..683 note = Xaa can be any naturally occurring amino acid
VARIANT	751..752 note = Xaa can be any naturally occurring amino acid
VARIANT	754..756 note = Xaa can be any naturally occurring amino acid
VARIANT	798..800 note = Xaa can be any naturally occurring amino acid
VARIANT	807..813 note = Xaa can be any naturally occurring amino acid
VARIANT	816..817

-continued

VARIANT	note = Xaa can be any naturally occurring amino acid 819..821
VARIANT	note = Xaa can be any naturally occurring amino acid 823..827
VARIANT	note = Xaa can be any naturally occurring amino acid 841..842
VARIANT	note = Xaa can be any naturally occurring amino acid 856..859
VARIANT	note = Xaa can be any naturally occurring amino acid 924..927
VARIANT	note = Xaa can be any naturally occurring amino acid 929..930
VARIANT	note = Xaa can be any naturally occurring amino acid 932..933
VARIANT	note = Xaa can be any naturally occurring amino acid 937..938
VARIANT	note = Xaa can be any naturally occurring amino acid 963..965
VARIANT	note = Xaa can be any naturally occurring amino acid 981..982
VARIANT	note = Xaa can be any naturally occurring amino acid 995..997
VARIANT	note = Xaa can be any naturally occurring amino acid 1020..1021
VARIANT	note = Xaa can be any naturally occurring amino acid 1041..1042
VARIANT	note = Xaa can be any naturally occurring amino acid 1105..1106
VARIANT	note = Xaa can be any naturally occurring amino acid 1213..1214
VARIANT	note = Xaa can be any naturally occurring amino acid 1231..1235
VARIANT	note = Xaa can be any naturally occurring amino acid 1240..1246
VARIANT	note = Xaa can be any naturally occurring amino acid 1255..1256
VARIANT	note = Xaa can be any naturally occurring amino acid 1260..1263
VARIANT	note = Xaa can be any naturally occurring amino acid 1265..1266
VARIANT	note = Xaa can be any naturally occurring amino acid 1268..1270
VARIANT	note = Xaa can be any naturally occurring amino acid 1272..1273
VARIANT	note = Xaa can be any naturally occurring amino acid 1277..1278
VARIANT	note = Xaa can be any naturally occurring amino acid 1280..1281
VARIANT	note = Xaa can be any naturally occurring amino acid 1283..1284
VARIANT	note = Xaa can be any naturally occurring amino acid 1298..1299
VARIANT	note = Xaa can be any naturally occurring amino acid 1320..1321
VARIANT	note = Xaa can be any naturally occurring amino acid 1325..1328
VARIANT	note = Xaa can be any naturally occurring amino acid 1345..1348
VARIANT	note = Xaa can be any naturally occurring amino acid 1350..1351
SITE	note = Xaa can be any naturally occurring amino acid 12
SITE	note = Xaa can be any naturally occurring amino acid 14
SITE	note = Xaa can be any naturally occurring amino acid 26
SITE	note = Xaa can be any naturally occurring amino acid 28
SITE	note = Xaa can be any naturally occurring amino acid 46
SITE	note = Xaa can be any naturally occurring amino acid 79
SITE	note = Xaa can be any naturally occurring amino acid 86
SITE	note = Xaa can be any naturally occurring amino acid 111

-continued

SITE	note = Xaa can be any naturally occurring amino acid 124
SITE	note = Xaa can be any naturally occurring amino acid 127
SITE	note = Xaa can be any naturally occurring amino acid 134
SITE	note = Xaa can be any naturally occurring amino acid 136
SITE	note = Xaa can be any naturally occurring amino acid 140
SITE	note = Xaa can be any naturally occurring amino acid 157
SITE	note = Xaa can be any naturally occurring amino acid 206
SITE	note = Xaa can be any naturally occurring amino acid 208
SITE	note = Xaa can be any naturally occurring amino acid 210
SITE	note = Xaa can be any naturally occurring amino acid 213
SITE	note = Xaa can be any naturally occurring amino acid 245
SITE	note = Xaa can be any naturally occurring amino acid 299
SITE	note = Xaa can be any naturally occurring amino acid 316
SITE	note = Xaa can be any naturally occurring amino acid 361
SITE	note = Xaa can be any naturally occurring amino acid 373
SITE	note = Xaa can be any naturally occurring amino acid 420
SITE	note = Xaa can be any naturally occurring amino acid 426
SITE	note = Xaa can be any naturally occurring amino acid 428
SITE	note = Xaa can be any naturally occurring amino acid 466
SITE	note = Xaa can be any naturally occurring amino acid 486
SITE	note = Xaa can be any naturally occurring amino acid 492
SITE	note = Xaa can be any naturally occurring amino acid 497
SITE	note = Xaa can be any naturally occurring amino acid 570
SITE	note = Xaa can be any naturally occurring amino acid 581
SITE	note = Xaa can be any naturally occurring amino acid 586
SITE	note = Xaa can be any naturally occurring amino acid 598
SITE	note = Xaa can be any naturally occurring amino acid 605
SITE	note = Xaa can be any naturally occurring amino acid 616
SITE	note = Xaa can be any naturally occurring amino acid 618
SITE	note = Xaa can be any naturally occurring amino acid 626
SITE	note = Xaa can be any naturally occurring amino acid 665
SITE	note = Xaa can be any naturally occurring amino acid 667
SITE	note = Xaa can be any naturally occurring amino acid 700
SITE	note = Xaa can be any naturally occurring amino acid 705
SITE	note = Xaa can be any naturally occurring amino acid 709
SITE	note = Xaa can be any naturally occurring amino acid 715
SITE	note = leucine (Leu) or methionine (Met) 721
SITE	note = Xaa can be any naturally occurring amino acid 737

-continued

SITE	note = Xaa can be any naturally occurring amino acid 741
SITE	note = Xaa can be any naturally occurring amino acid 744
SITE	note = Xaa can be any naturally occurring amino acid 760
SITE	note = Xaa can be any naturally occurring amino acid 763
SITE	note = Xaa can be any naturally occurring amino acid 796
SITE	note = Xaa can be any naturally occurring amino acid 802
SITE	note = Xaa can be any naturally occurring amino acid 804
SITE	note = Xaa can be any naturally occurring amino acid 831
SITE	note = Xaa can be any naturally occurring amino acid 861
SITE	note = Xaa can be any naturally occurring amino acid 863
SITE	note = Xaa can be any naturally occurring amino acid 886
SITE	note = Xaa can be any naturally occurring amino acid 895
SITE	note = Xaa can be any naturally occurring amino acid 903
SITE	note = Xaa can be any naturally occurring amino acid 920
SITE	note = Xaa can be any naturally occurring amino acid 950
SITE	note = Xaa can be any naturally occurring amino acid 957
SITE	note = Xaa can be any naturally occurring amino acid 959
SITE	note = Xaa can be any naturally occurring amino acid 972
SITE	note = Xaa can be any naturally occurring amino acid 975
SITE	note = Xaa can be any naturally occurring amino acid 978
SITE	note = Xaa can be any naturally occurring amino acid 1025
SITE	note = Xaa can be any naturally occurring amino acid 1076
SITE	note = Xaa can be any naturally occurring amino acid 1080
SITE	note = Xaa can be any naturally occurring amino acid 1095
SITE	note = Xaa can be any naturally occurring amino acid 1112
SITE	note = Xaa can be any naturally occurring amino acid 1115
SITE	note = Xaa can be any naturally occurring amino acid 1140
SITE	note = Xaa can be any naturally occurring amino acid 1144
SITE	note = Xaa can be any naturally occurring amino acid 1165
SITE	note = Xaa can be any naturally occurring amino acid 1181
SITE	note = Xaa can be any naturally occurring amino acid 1185
SITE	note = Xaa can be any naturally occurring amino acid 1194
SITE	note = Xaa can be any naturally occurring amino acid 1197
SITE	note = Xaa can be any naturally occurring amino acid 1258
SITE	note = Xaa can be any naturally occurring amino acid 1275
SITE	note = Xaa can be any naturally occurring amino acid 1295
SITE	note = Xaa can be any naturally occurring amino acid 1315
SITE	note = Xaa can be any naturally occurring amino acid 1323

-continued

SITE note = Xaa can be any naturally occurring amino acid
1330

SITE note = Xaa can be any naturally occurring amino acid
1336

SITE note = Xaa can be any naturally occurring amino acid
1339

SITE note = Xaa can be any naturally occurring amino acid
1341

SITE note = Xaa can be any naturally occurring amino acid
1343

SITE note = Xaa can be any naturally occurring amino acid
1359

SITE note = Xaa can be any naturally occurring amino acid

SEQUENCE: 1

MKEKRXXXXX	PXVXPXXXX	XXWDDXDXPX	XXXXIKRQSI	DHGVGXEXXX	XXEXXXXXXF	60
RKYN SQGFKA	KD TDLMGTXL	XXXXXEXELXX	XXXXXADQXQ	XSAXXERKK	XPTXXXLDPX	120
XEPXIFXILD	TDYXIXPAQX	VXXGXDXDTE	GKPVIRXFGK	KENGFEFKIEY	DRTFEPYFYA	180
LLXXGXXKDD	LSAIEEVKKI	TAERHXRXXGX	ETXVRVVDAA	EVVKKFEPLG	LEXXRPIEVW	240
KLYFXHPQDV	PAIRDKIREH	PAVVXXXGXX	DIXXYEYDIP	FAKRYLIDKG	LIPPMWVXM	300
EGDKXXGRL	LNXXSXCQLE	XXXXXXSDXXV	HPXXXXRLEE	LPDYIPPLKV	LAFDIETLYH	360
XELGSFPPEP	KGXRP IIAIT	HYDSIDDRFY	VFDMISSAYA	DEEEARPFLX	XVITWKNIDX	420
LPYEIXDXVD	VVSFDTEREM	LKRFLXXXXE	KDPDVIIGYN	GDNFDXPYLK	KRAEKLKVVX	480
XGIPLXLGRL	SXKRSSXDGS	VFEPKIQRMV	GDRFAVEICL	GYKKGRIHF	DLYPVIRRTI	540
FTNYAFEWNL	PSYTLAEVAE	ALLGXXKEKX	DXXVYAEET	AWNEXGENL	ERLARYSXED	600
AELTXELGAK	KEFLPXEXLI	QLSRVXGQPL	DWDVSRSTG	NLVEWLLLRK	AYERNEVAPN	660
KPSEXEXARR	XXXRESALIK	XXXYPGGYVW	LNDRFKEPEX	GLYEXNIVXL	DPRSLYPSII	720
XTHNLSPDTL	LNREGCXPYH	XYIXDVAPEE	XXVXXXSGHX	FCXTKDFPGF	IPSILIKRGD	780
LLEERKKIKK	SGGKMXAXXX	NXEXIKKXXX	XXLSXXEXX	XVXXXXXFS	XFKKTIDPNT	840
XXIEKKLLDY	RQRAIXXXXT	XQXNRKILAN	SFYGYLGYAK	ARWYXCECAE	SVTAXGFRQY	900
IEXTIREIEE	KFGKVCCTNX	EAFXXXXAXX	CXXNGYXXGF	KVLYGDTDSX	FVTIPGXIKX	960
VGXXXXFADAE	EXKKXILXFK	XXEFLKYINA	WWREXXXVMN	KLPGLLELEY	EGFYKRGFLX	1020
XGRGXVFGTK	KRYAGLIDEM	XXDRFAEGKI	TTKGLTVRR	DWSELAKETQ	ARVLEXILKX	1080
EPEGDVVEAE	ESVXYVKEV	IEKLXXYEVP	PXAGXIEKLV	IHKQLTRDL	DYKATGFPGX	1140
KCPXHVAVAK	RLDRANAARG	RGVXXRPGTV	ISYVILKGGG	XIGDXDRAIP	PDEXDPXKHK	1200
YDAEYYIENQ	VLXXPAVERI	LEAFIGDAPA	XXXXXXRGDHX	XXXXXXTGYR	KEDLXXFXKX	1260
XXXXXXXXXX	LXXGXGXXCX	XCXXRYQKTK	QVGLXAWXXL	KPKGKRLWTQ	CQRCXGNLHX	1320
XVXCCCCXCX	IFYMRKXVXK	XLXDXXXLX	XFGPPGPFEX			1360

SEQ ID NO: 2 moltype = AA length = 773
 FEATURE Location/Qualifiers
 source 1..773
 mol_type = protein
 organism = Thermococcus gorgonarius

SEQUENCE: 2

MILDTDYITE	DGKPVIRIFK	KENGFEKIDY	DRNFEPYIYA	LLKDDSAIED	VKKITAERHG	60
TTVVRVRAEK	VKKKFLGRPI	EVWKLYFTHP	QDVPAIRDKI	KEHPAVVDIY	EYDIPFAKRY	120
LIDKGLIPME	GDEELKMLAF	DIETLYHEGE	EFAEGPILMI	SYADEEGARV	ITWKNIDL PY	180
VDVSTEEKM	IKRFLKVVKE	KDPDLVITYN	GDNFDFAYLK	KRSEKLGKVF	ILGREGSEPK	240
IQRMGDRFAV	EVKGRIHFDL	YPVIRRTINL	PTYTLEAVYE	AIFGQPKKEV	YAEETIAQAW	300
TGEGLERVAR	YSMEDAKVTY	ELGKEFFPME	AQLSRLVGQS	LWDVSRSTG	NLVEWFLLRK	360
AYERNELAPN	KPDERELARR	RESYAGGYVK	EPERGLWENI	VYLDFRSLYP	SIIITHNVSP	420
DTLNREGCEE	YDVAPQVGHR	FCKDFPGFIP	SLGLDLEER	QKVKKKMKAT	IDPIEKKLLD	480
YRQRAIKILA	NSFYGYGYGA	KARWYCKECA	ESVTAWGRQY	IETTIREIEE	KFGFKVLYAD	540
TDGFFATIPG	ADAETVKKKA	KEPLDYINAK	LPGLLELEYE	GFYKRGFFVT	KKKYAVIDEE	600
DKITTRGLEI	VRDWSIEIAK	ETQARVLEAI	LKHGDVEEAV	RIVKEVTEKL	SKYEVPEPEL	660
VIYEQITRDL	KDYKATGPHV	AVAKRLAARG	IKIRPGTVIS	YIVLKGSGRI	GDRAIPFDEF	720
DPAKHKYDAE	YYIENQVLPA	VERILRAFGY	RKEDLRYQKT	RQVGLGAWLK	PKT	773

SEQ ID NO: 3 moltype = AA length = 774
 FEATURE Location/Qualifiers
 source 1..774
 mol_type = protein
 organism = Thermococcus kodakarensis

SEQUENCE: 3

MILDTDYITE	DGKPVIRIFK	KENGFEKIEY	DRTFEPYFYA	LLKDDSAIEE	VKKITAERHG	60
TVVTVKRVKE	VQKFLGRPV	EVWKLYFTHP	QDVPAIRDKI	REHPAVIDIY	EYDIPFAKRY	120
LIDKGLVPM	GDEELKMLAF	DIETLYEEGE	EFAEGPILMI	SYADEEGARV	ITWKNVDLPY	180
VDVSTEREM	IKRFLRVVKE	KDPDLVITYN	GDNFDFAYLK	KRCEKLGINF	ALGRDGESEPK	240
IQRMGDRFAV	EVKGRIHFDL	YPVIRRTINL	PTYTLEAVYE	AVFGQPKKEV	YAEETITAW	300
TGENLERVAR	YSMEDAKVTY	ELGKEFLPME	AQLSRLIGQS	LWDVSRSTG	NLVEWFLLRK	360
AYERNELAPN	KPDEKELARR	RQSYEGGYVK	EPERGLWENI	VYLDFRSLYP	SIIITHNVSP	420
DTLNREGCEE	YDVAPQVGHR	FCKDFPGFIP	SLGLDLEER	QKIKKKMKAT	IDPIERKLLD	480
YRQRAIKILA	NSYGYGYGYA	RARWYCKECA	ESVTAWGREY	ITMTIKEIEE	KYGFVKIYSD	540
TDGFFATIPG	ADAETVKKKA	MEFLKYINAK	LPGALELEYE	GFYERGFFVT	KKKYAVIDEE	600
GKITTRGLEI	VRDWSIEIAK	ETQARVLEAL	LKGDVVEKAV	RIVKEVTEKL	SKYEVPEPEL	660
VIHEQITRDL	KDYKATGPHV	AVAKRLAARG	VKIRPGTVIS	YIVLKGSGRI	GDRAIPFDEF	720

-continued

DPTKHKYDAE YYIENQVLPA VERILRAFYG RKEDLRYQKT RQVGLSAWLK PKGT 774

SEQ ID NO: 4 moltype = AA length = 775
FEATURE Location/Qualifiers
source 1..775
 mol_type = protein
 organism = Thermococcus sp.

SEQUENCE: 4

MILDTDYITE	NGKPVIRVFK	KENGEPKIEY	DRTFEPYFYA	LLKDDSAIED	VKKVTAKRHG	60
TVVKVKRAEK	VQKKFLGRPI	EVWKLIFNHP	QDVPAIRDRI	RAHPAVVDIY	EYDIPFAKRY	120
LIDKGLIPME	GDEELTMLAF	DIETLYHEGE	EFGTGPILMI	SYADGSEARV	ITWKKIDLPI	180
VDVVSSTEKEM	IKRFLRVVRE	KDPDVLITYN	GDNFDFAYLK	KRCEELGIKF	TLGRDGSEPK	240
IQRMGDRFAV	EVKGRIHFDL	YPVIRRTINL	PTYTLEAVYE	AVFGKPKEKV	YAEETIAQAW	300
SGEGLERVAR	YSMEDAKVTY	ELGREFFPME	AQLSRLIGQS	LWDVSRSTG	NLVEWFLLRK	360
AYKRNELAPN	KPDERELARR	RGGYAGGYVK	EPERGLWDNI	VYLDFRSLYP	SIITHNVSP	420
DTLNREGCKE	YDVAVEVGHK	FCKDFPGFIP	SLGDLLEER	QKIKRKMAT	VDPLEKKLLD	480
YRQRAIKILA	NSFYGYGYGA	KARWYCKECA	ESVTAWGREY	IEMVIRELEE	KFGFKVLYAD	540
TDGLHATIPG	ADAETVKKKA	KEFLKYINPK	LPGLLELEYE	GFYVRGPFVT	KKKYAVIDEE	600
GKITTRGLEI	VRRDWEIAK	ETQARVLEAI	LKHGDVEEAV	RIVKEVTEKL	SKYEVPPPEKL	660
VIHEQITRDL	RDYKATGPHV	AVAKRLAARG	VKIRPGTVIS	YIVLKGSGRI	GDRAIPADEF	720
DPTKHRYDAE	YYIENQVLPA	VERILKAFGY	RKEDLRYQKT	KQVGLGAWLK	VKGKK	775

SEQ ID NO: 5 moltype = AA length = 775
FEATURE Location/Qualifiers
source 1..775
 mol_type = protein
 organism = Pyrococcus furiosus

SEQUENCE: 5

MILDVDYITE	EGKPVIRLFK	KENGKFKIEH	DRTFRPIIYA	LLRDDSKIEE	VKKITGERHG	60
KIVRIVDVEK	VEKKFLGKPI	TVWKLYLEHP	QDVPTIREKV	REHPAVVDIF	EYDIPFAKRY	120
LIDKGLIPME	GEEELKILAF	DIETLYHEGE	EFGKGPIMI	SYADENEAKV	ITWKNIDLPI	180
VEVVSSEREM	IKRFLRIIRE	KDPDIIVTYN	GDSDFFPYLA	KRAEKLGIKL	TIGRDGSEPK	240
MQRIGDMTAV	EVKGRIHFDL	YHVITRTINL	PTYTLEAVYE	AIFGKPKEKV	YADEIAKAW	300
SGENLERVAR	YSMEDAKATY	ELGKEFLPME	IQLSRLVGQP	LWDVSRSTG	NLVEWFLLRK	360
AYERNEVAPN	KPSEEEYQRR	LRESYTGGFV	KEPEKGLWEN	IVYLDPRALY	PSIITHNVVS	420
PDTLNLEGCK	NYDIAPQVGH	KPKCDIPGFI	PSLLGHLLEE	RQKIKTKMKE	TQDPIEKILL	480
DYRQKAIKLL	ANSFYGYGY	AKARWYCKEC	AESVTAWGRK	YIELVWKELE	EKFGFKVLYI	540
DTDGLYATIP	GGESEIIEKK	ALEFVKYINS	KLPGLLELEY	EGFYKRGFFV	TKKRYAVIDE	600
EGKVITRGLE	IVRRDWEIA	KETQARVLET	ILKHGDVEEA	VRIVKEVIQK	LANYEIPPEK	660
LAIYEQITRP	LHEYKAIGPH	VAVAKKLAAG	GVIKIPGMVI	GYIVLRGDGP	ISNRILAEE	720
YDPKKHKYDA	EYYIENQVLP	AVLRILEGFG	YRKEDLRYQK	TRQVGLTSLW	NIKKK	775

SEQ ID NO: 6 moltype = AA length = 774
FEATURE Location/Qualifiers
source 1..774
 mol_type = protein
 organism = Thermococcus litoralis

SEQUENCE: 6

MILDTDYITK	DGKPIIRIFK	KENGEPKIEL	DPHFQPIIYA	LLKDDSAIEE	IKAIKGERHG	60
KTVRVLDVAV	VRKKFLGREV	EVWKLIFEHP	QDVPAMRGKI	REHPAVVDIY	EYDIPFAKRY	120
LIDKGLIPME	GDEELKLLAF	DIETFYHEGD	EFGKGEIIMI	SYADEEEARV	ITWKNIDLPI	180
VDVVSNEREM	IKRPFQVVEK	KDPDVIITYN	GDNFDLPLYI	KRAEKLGVRL	VLGRDKHEPE	240
PKIQRMGDSF	AVEIKGRIHF	DLFPVVRRTI	NLPTYTLEAV	YEAVLGKTKS	KLGAEEIAAI	300
WETEESMKKL	AQYSMEDARA	TYELGKEFFP	MEAEALAKLIG	QSVWDVSRSS	TGNLVEWYLL	360
RVAYARNELA	PNKPDEEYK	RRLRTTYLGG	YVKEPEKGLW	ENIYLDPRS	LYPSIIVTHN	420
VSPDTLEKEG	CKNYDVAPIV	GYRFCKDFPG	FIPSILGDLI	AMRQDIKKKM	KSTIDPIEKK	480
MLDYRQRAIK	LLANSYYGYM	GYPKARWYSK	ECAESVTAWG	RHYIEMTIRE	IEEKFGFKVL	540
YADTDGFFAT	IPGKEPELIK	KKAKEFLNYI	NSKLPGLLEL	EYEGFYLRGF	FVTKKRYAVI	600
DEEGRITTRG	LEVVRDWE	IAKETQAKVL	EAILKEGSVE	KAVEVVRDVV	EKIAKYRVPL	660
EKLVIHEQIT	RDLKDYKAIG	PHVAIAKRLA	ARGIKVKPGT	IISYIVLKGS	GKISDRVIL	720
TEYDPRKHXY	DPDYIENQV	LPAVLRILEA	FGYRKEDLRY	QSSKQTGLDA	WLKR	774

SEQ ID NO: 7 moltype = AA length = 937
FEATURE Location/Qualifiers
source 1..937
 mol_type = protein
 organism = Methanosarcina acetivorans

SEQUENCE: 7

MPMDFQILDA	DYEVVNDSGP	VIRLFGRGAD	GKSVCCFPVD	PEPYFYLKAS	GDLHAVARLI	60
KDTFEQVKV	EIVEKFEFVG	YQKTKKEMLR	VTRLPKDVP	EIRDEILKIR	DVLAEGDWQ	120
VYESDILFRN	RFLIDRALGG	MVWVSABGKP	VDPVRYLGAG	SAWRSRCENF	ACDSAVLASG	180
LKRVENLAIA	PLKYLAFDIE	CLPLDGGMPS	PDVSPIMIS	FSFEPEYKGH	KTLLILLAKPA	240
AGMDGDVLS	MDTEMLNKF	FEIICEYDPD	IVAGYNHQDF	DIPYITERVK	ALVAKGETIN	300
SVVGRDGSPI	GYRKFGILTR	TEMKGRVVVD	ALPLVRRAPS	LKQYTLRAVS	KELLSREKLD	360
VPLEMEHWH	NDSDGKFRKF	VDYARRDSEL	ALELVLELRL	LDKYIALAQV	SGSLLEQIVD	420
GGQTSMVETL	LLREFGLKDR	VILPKPGDEL	SAERYDMSSD	LKGGEVLEPK	KGLENNVLIL	480

-continued

DYKSLYPTIM	MAHNLCTYTV	VTRDRPDGKT	IKPPSGGEFV	PPEVFRGIVP	SILEDLLNKR	540
GDTKKRMKRT	SDENEHRVLD	ATQLAIKILL	NSFYGYSGYA	RARLYSLTLA	NAVTSFGRSN	600
ILNTRDLING	RIGKIVLRNS	AALLLEEAGK	LSPQDRIVEL	SVAYGDTDSV	FVHCKAKGDL	660
SLEEVSLVGN	RLSEIVSASL	PDPMELEFES	VAKRALLIAK	KRYALWLFEF	RNSGWENKIK	720
VKGMETVRRD	WCELTSTITLN	RVLEFVLIEG	DVDKAVEHVR	KVVSVDVRNLD	PGKDAGIEIK	780
LVLTRTLTRK	ADSYKNKQPH	LTVAENLKKR	TGIMPSIGTR	IPFVITAGKG	LFVDRAEDPD	840
YVRENNVPID	VDYVVKQIL	PPVERILEVF	GVKMSSLDFF	AKQKGLDFE	VKKPEAKKQE	900
KSSSQKGTNG	KILEKAPEEK	ARYSENGRVE	QRSLEDF			937

SEQ ID NO: 8 moltype = AA length = 785
 FEATURE Location/Qualifiers
 source 1..785
 mol_type = protein
 organism = Pyrobaculum islandicum

SEQUENCE: 8

MELKVMPLDI	TYAVVGSVPE	IRIFGILSSG	ERVVLIDRSF	KPYFYVDCAV	CEPAALKTAL	60
SRVAPIDDDQ	IVERRFLGRS	KKFLKVIKI	PEDVRKLREA	AMSIPRVSGV	YEADIRFYMR	120
YMIDMGVVP	SWNVAEEVEG	GRLGGIPTV	VSQWYGIDEG	FPPSLKVMF	DIEVYNERGS	180
PDPIRDPVVM	LAIKTNDGHE	EVFEASGKDD	RGVVRADFV	IRSYDDPVIV	GYNSNGFDWP	240
YLVERAKAVG	VPLKVDRLSN	PPQQSVYGH	SIVGRANVDL	YNIVEEFPPI	KLKTLDRVAE	300
YFGVMKREER	VLIPGHKIYE	YWKDPNKRPL	LKRYVLDDVR	STLGLADKLL	PFLIQLSSVS	360
GLPLDQVAAA	SVGNRVEWML	LRAYARLGEV	APNREEREYE	PYKGAIVLEP	KPGMYEDVLV	420
LDFSSMYPNI	MMKYNLSPDT	YLEPGEPPDP	EGVNVAVEVG	HRFRRSPPGF	VPQVLKSLVE	480
LRKAVREEAK	KYPPDSPFEK	ILDERQALK	VMANAIYGYL	GWVGARWYKR	EVAESVTAPA	540
RAILKDVIEQ	ARRLGIVVVV	GDTSLEFVKK	HGDVDKLIK	VEEKYGIDIK	VDKDYAKVLF	600
TEAKKRYAGL	LRDGRIDIVG	FEVVRGDWSE	LAKDVQLRVI	EIILKSRLDIV	EARHGVIKVI	660
REIIEERLKNY	KFNIDDLIIV	KTLDKELDEY	KAYPPHVHAA	QILKRHGYRV	GKGTITIGYVI	720
VKGGEKVSER	ALPYILLDDI	KKIDIDYIE	RQIIPAAALRI	AEVIGVKESD	LKTGRMERSL	780
LDLFLS						785

SEQ ID NO: 9 moltype = AA length = 882
 FEATURE Location/Qualifiers
 source 1..882
 mol_type = protein
 organism = Sulfolobus solfataricus

SEQUENCE: 9

MTKQLTLFDI	PSSKPAKSEQ	NTQQSQQSAP	VEEKKVVRRE	WLEEAQENKI	YFLLQVDYDG	60
KKGKAVCKLF	DKETQKIYAL	YDNTGHKPYF	LVDLEPKVKG	KIPKIVRDPS	FDHIETVSKI	120
DPYTWNKFKL	TKIVVRDPLA	VRRLRNDVPK	AYEBAHIKYN	NMYDIGLIP	GMPYVVKNGK	180
LESVYLSLDE	KDVEIKKAF	ADSDENTROM	AVDWLPIFET	EIPKIKRVAI	DIEVYTPVKG	240
RIPDSQKAEF	PIISIALAGS	DGLKKVLVLN	RNDVNEGSVK	LDGISVERFN	TEYELLGRFF	300
DTLLEYPIVL	TFNGDDFDLP	YIYFRALKLG	YFPEEIPIDV	AGKDEAKYLA	GLHIDLYKFF	360
FNKAVRNAYF	EGKYNEYNLD	AVAKALLGTS	KVKVDTLISF	LDVEKLIIEYN	FRDAEITLQL	420
TTFNNDLTMK	LIVLFSRISR	LGIEELTRTE	ISTWVKNLYY	WEHRKRNWLI	PLKEEILAKS	480
SNIRTSALIK	GKGYGKAVVI	DPPAGIFFNI	TVLDFASLYP	SIIRTWNLSY	ETVDIQCKK	540
PYEVKDETEG	VLHIVCMDRP	GITAVITGLL	RDFRVKIYKK	KAKNPNNSEE	QKLLYDVVQR	600
AMKVFINATY	GVFGAETFPL	YAPAVAESVT	ALGRYVITST	VKKAREEGLT	VLYGDTDSL	660
LLNPPKNSLE	NIIKWVKTTF	NLDLEVDKTY	KFVAFSGLKK	NYPGVYQDGK	VDIKGMLVKK	720
RNTPEFVKKV	FNEVKELMIS	INSPNDVKEI	KRKIVDVVKG	SYEKLKNKGY	NLDELAFKVM	780
LSKPLDAYKK	NTPQHVKAAAL	QLRPPGVNVL	PRDIIYVVKV	RSKDGVKPVQ	LAKVTEIDAE	840
KYLEALRSTF	EQILRAFVGS	WDEIAATMSI	DSFFSYPSKG	NS		882

SEQ ID NO: 10 moltype = AA length = 784
 FEATURE Location/Qualifiers
 source 1..784
 mol_type = protein
 organism = Methanococcus maripaludis

SEQUENCE: 10

MESLIDLIDYN	SDDLICIYLYL	INSIIKEKDF	KPYFYVNSTD	KEQILEFLKD	YEKKHKLDSE	60
ISKMIENIET	VKKIVFDENY	QEKELSKVTV	KYPNNVKTVR	EILMEFERLY	EYDIPFVRRY	120
LIDNSVIPTS	TWDFENNKKI	DNKIPDFKT	SPDIEVYCNK	EPNPKKDPII	MASFSSKDFN	180
TVVSTKKFDH	EKLEYVKDEK	ELIKRIIEIL	KEYDIIYTYN	GDNFDFPYLK	KRAESFGLEL	240
KLGNDEKIK	ITKGMMSKS	YIPGRVHIDL	YPIARLLNL	TKYRLNVTE	ALFDVKKVDV	300
GHENIPKMW	NLDETLEVEYS	HQDAYYTQRI	GEQFLPLEIM	FSRVVNQSLY	LDNRMSSSQM	360
VEYLLLLKNSY	KMGVIAPNRP	SGKEYQKRIR	SSYEGGYVKE	PLKGIHEDIV	SMDFLSLYPS	420
IIMSHNLSPE	TIDCTCCSDE	ENGINEEILG	HKFCKKSIGI	IPKTLMDLIN	RRKKVKVKVLR	480
EKAKEGEPDE	EYQILDYEQR	SIKVLANSY	GYLAPPMARW	YSRDCAEITT	HLGRQYIQKT	540
IEEAENFGFK	VIYADTDGFY	SKWADDKEKL	SKYELLEKTR	EFLKNINNTL	PGEMELEFEG	600
YFKRGIFVTK	KKYALIDENE	KITVKGLEVV	RRDWSNVSKN	TQKNVLNALL	KEGSVENAKK	660
VIQDTIKELK	DGKVNNEDLL	IHTQLTKRIE	DYKTTAPHVE	VAKKILKSGN	RVNTGDISY	720
IITSGNKSIS	EBAEILENAK	INDTNYIEN	QILPPVIRLM	EALGITKDEL	KDSKKQYTLH	780
HFLK						784

SEQ ID NO: 11 moltype = AA length = 1107
 FEATURE Location/Qualifiers
 source 1..1107

-continued

mol_type = protein
organism = Homo sapiens

SEQUENCE: 11

MDGKRRPGPG	PGVPPKRARG	GLWDDDDAPR	PSQFEEDLAL	MEEMEAHRL	QEQQEEELQS	60
VLEGVADGQV	PPSAIDPRWL	RPTPPALDPQ	TEPLIFQQLE	IDHYVGPAQP	VPGGPPPSRG	120
SVPVLRAGV	TDEGFSVCCH	IHGFPYFYT	PAPPGFGPEH	MGDLQRELN	AISSRSRGG	180
ELTGPAVLAV	ELCSRESMFG	YHGHPSPFL	RITVALPRLV	APARRLLEQG	IRVAGLGTPS	240
FAPYEANVDF	EIRFMVDTDI	VGCNWLELPA	GKYALRLKEK	ATQCQLEADV	LWSDVVSHPP	300
EGPWQRIAPL	RVLSFDIECA	GRKGIFPEPE	RDVPIQICSL	GLRWGEPEPF	LRLALTLRPC	360
APILGAKVQS	YEKEEDLLQA	WSTFIRIMDP	DVITGYNIQN	FDLPYLISSRA	QTLKVQTFPF	420
LGRVAGLCSN	IRDSFSQSKQ	TGRDRTKVVS	MVGRVQMDML	QVLLREYKLR	SYTLNAVSPH	480
FLGEQKEDVQ	HSIITDLQNG	NDQTRRLAV	YCLKDAYLPL	RLLERLMVLV	NAVEMARVTG	540
VPLSYLLSRG	QQVKVVSQLL	RQAMHEGLLM	PVVKSEGGED	YTGATVIEPL	KGYVDVPIAT	600
LDFSSLYPSI	MMAHNLCYTT	LLRPGTAQKL	GLTEDQFIRT	PTGDEFVKTS	VRKGLLPQIL	660
ENLLSARKRA	KAEALAKETDP	LRRQVLDGRQ	LALKVSANSV	YGFTGAQVKG	LPCLEISQSV	720
TGFGRMIEK	TQLVESKYT	VENGYSYTSK	VYVGDTSVM	CRFGVSSVAE	AMALGREAAD	780
WVSGHFPSP	RLEFEKVYFP	YLLISKRYA	GLLFSSRPDA	HDRMDCKGLE	AVRRDNCPLV	840
ANLVTASLRR	LLIDRDPEGA	VAHAQDVISD	LLCNRIDISQ	LVITKELTRA	ASDYAGKQAH	900
VELAERMKR	DGSAAPSLGD	RVPYVVISAA	KGVAAYMKSE	DPLFVLEHSL	PIDTQYYLEQ	960
QLAKPLLRIF	EPILGEGRAE	AVLLRGDHT	CKTVLTGKVG	GLLAFAKRRN	CCIGCRTVLS	1020
HQGAUCEFCQ	PRESELYQKE	VSHLNALKEE	FSRLWTQCQR	CQGLSHEDVI	CTSRDCPIFY	1080
MRKKVRKDL	DQEQLLRFP	PGPEAW				1107

SEQ ID NO: 12 moltype = AA length = 1097
FEATURE Location/Qualifiers
source 1..1097
 mol_type = protein
 organism = Saccharomyces cerevisiae

SEQUENCE: 12

MSEKRSPLMV	DVKIDDEDTP	QLEKKIKRQS	IDHGVGSEPV	STIEIIPSDS	FRKYNSSQGF	60
AKDIDLMTG	LESTFEQELS	QMEHDMADQE	EHDLSSEFERK	KLPTDFDPSL	YDISFQQIDA	120
EQSVLNGIKD	ENTSTVVRFF	GVTSEGHSLV	CNVTFGKKNL	YVPAPNSSDA	NDQEQINKFV	180
HYLNTFFDHA	IDSIEVVSQK	SIWGYSGDTK	LFPWKIYVTV	PHMVNKLRTA	FERGHLSPNS	240
WFSNGTTTVD	NIATYTLRLM	DCGIVGMSWI	TLPGKGYSMI	EPNNRVSSCQ	LEVSINYNRL	300
IAHPAEGDWS	HTAPLRIMSF	DIECAGRIGV	FPEPEYDPVI	QIANVVSIAQ	AKKPFIRNVF	360
TLNCSPTITG	SMIFSHATEE	EMLSNWRNFI	IKVDPDVIG	YNTTNFDIPY	LLNRAKALKV	420
NDFFPYFGRK	TVKQEKESV	FSSKAYGTRE	TKNVNIDGRL	QLDLLQFIQR	EYKLSYTLN	480
AVSAHFLGEQ	KEDVHYSIIS	DLQNGDSETR	RRLAVYCLKD	AYLPLRLMEK	LMALVNYTEM	540
ARVTGVFFSY	LLARGQQIKV	VSQLFKRCLE	IDTVIPNMQS	QASDDQYEGA	TVIEPIRGYY	600
DVIATLDFN	SLYPSIMMAH	NLCYTTLCNK	ATVERLNLKI	DEDYVITPNG	DYFVTTKRRR	660
GILPIILDEL	ISARKRAKDD	LRDEKDPFKR	DVLNGRQLAL	KISANSVYGF	TGATVGKLPC	720
LAISSSVTAY	GRTMILKTKT	AVQEKYCIKN	GKHKDAVVVY	GDTDSVMVKF	GTTDLKEAMD	780
LGTEAAKYVS	TLFKPHINLE	INKRYAGFL	WTNPDKFDKL	DQKGLASVRR		840
DSCSLVSIVM	NKVLKILIE	RNVGALAFV	RETINDILHN	RVDISKLIIS	KTLPANYTNP	900
QPHAVLAERM	KRREGVGNV	GDRVDVYIIG	GNDKLYNRAE	DPLFVLENNI	QVDSRYLYTN	960
QLQNPISIV	APIIGDKQAN	GMPVVKSIKI	NTGSQKGLM	SFIKKVEACK	SCKGPLRKGE	1020
GPLCSNCLAR	SGELYIKALY	DVRDLEEKYS	RLWTQCQRC	GNLHSEVLCS	NKNCDFIYMR	1080
VKVKELQEK	VEQLSKW					1097

SEQ ID NO: 13 moltype = AA length = 787
FEATURE Location/Qualifiers
source 1..787
 mol_type = protein
 organism = Pseudomonas aeruginosa

SEQUENCE: 13

MELLQGFVLT	RHWRDTPAGT	EVAFWLATEQ	GPRQVRLPPQ	PSVAFVLAEQ	RGRVESLLAG	60
ETGAELRPLA	LRDFQQRVPL	GLYCQQHRQL	MNLEKRLRQA	GVEVFADIR	PPERYLMERF	120
ITAPVSLEAS	VEADGSLAR	RLKPAPDYRP	RLRLVSLDIE	TNARGDLYSI	ALEGCDQRQV	180
YMLGPANGDA	AAVDRLDYC	DSRAGLLERL	NQWLAEHDPD	AIIGWNLVQF	DLRVLHEHAQ	240
RLKVPLRLGR	GGDEMGRWH	GSRNNHFFAA	AAGRLIIDGI	EALRSATWSF	PSFSLENVAR	300
TLLEGKAIID	NPYQRMDEID	RMFAEDKPAL	AHYNLKDCEL	VTRIFARTEL	LDLFLERATV	360
TGLPADRSSG	SVAAFTHLYM	PLMHRAGFVA	PNLGEKRPEA	SPGGFVMDSR	PGLYESVLVL	420
DYKSLYPSII	RTFLIDPVL	VEGLRQPDDE	HSVEGFRGAR	FSRTRHCLPA	IVARVWEGRE	480
AAKRERNQPL	SQALKIMNA	VEGLGSSGSC	RFFDPRLESS	ITLRGHRIMR	RTRELIEAEG	540
YTVIYGDTS	TFVWLGSPPA	EEEAAGIGRA	LVARVNDWWR	EHLKEEGLD	SALELQFETH	600
YRRFLMPTVR	GAEESKRY	AGLVRRADGG	EEMVFKGLET	VRTDWSPLAQ	RFQQLYLRI	660
FNRQPYQDYV	RDYVVRTLAG	ELDLLLVYRK	RLRRRLDDYQ	RNVPPHVRAA	RIADDYNLER	720
GRPRQYQSGG	WISVVISVAG	PERLEARRSA	IDYEHYVGKQ	LQPVADAILP	FVGDDFATLV	780
DRQMALF						787

SEQ ID NO: 14 moltype = AA length = 783
FEATURE Location/Qualifiers
source 1..783
 mol_type = protein
 organism = Escherichia coli

SEQUENCE: 14

-continued

MAQAGFILTR	HWRDTPQGTE	VSPWLATDNG	PLQVTLAPQE	SVAFIPADQV	PRAQHILQGE	60
QGFRLTPLAL	KDFHRQPVYG	LYCRAHRQLM	NYEKRLREGG	VTVYEADVRP	PERYLMERFI	120
TSPVWVEGDM	HNGTIVNARL	KPHPDYRPL	KWVSIDIETT	RHGELYCIGL	EGCGQRIVYM	180
LGPENGDASS	LDFELEYVAS	RPQLLEKLNA	WFANYDPDVI	IGWNVVQFDL	RMLQKHAERY	240
RLPLRLGRDN	SELEWREHGF	KNGVFFAQAK	GRLIIDGIEA	LKSAFWNFSS	FSLETVAQEL	300
LGEKSIDNP	WDRMDEIDRR	FAEDKPALAT	YNLKDCELVT	QIFHKTEIMP	FLLERATVNG	360
LPVDRHGGSV	AAFHGLYFPR	MHRAGYVAPN	LGEVPPHASP	GGYVMDSRPG	LYDSVLVLDY	420
KSLYPSIIRT	FLIDPVGLVE	GMAQPDPEHS	TEGFDAWFS	REKHCLPEIV	TNIWHGRDEA	480
KRQGNKPLSQ	ALKIIMNAFY	GVLGTTACRF	FDPRLASSIT	MRGHQIMRQT	KALIEAQGYD	540
VIYGDSTDSTF	VWLKGAHSEE	EAAKIGRALV	QHVNAWMAET	LQKQRLTSAL	ELEYETHFCR	600
FLMPTIRGAD	TGSKKRYAGL	IQEGDKQRMV	FKGLETVRTD	WTPLAQQFQQ	ELYLRIFRNE	660
PYQEVVRETI	DKLMAGELDA	RLVYRKRLRR	PLSEYQRNVP	PHVRAARLAD	EENQKRGRPL	720
QYQNRGTIKY	VWTTNGPEPL	DYQRSPLDYE	HYLTRQLQPV	AEGILPFIED	NFATLMTGQL	780
GLF						783

SEQ ID NO: 15 moltype = AA length = 903
 FEATURE Location/Qualifiers
 source 1..903
 mol_type = protein
 organism = Escherichia coli phage

SEQUENCE: 15

MKEFYLTVEQ	IGDSIFERYI	DSNGRERTRE	VEYKPSLFAH	CPESQATKYF	DIYGKPCTRK	60
LFANMRDASQ	WIKRMEDIGL	EALGMDDFKL	AYLSDTYNVE	IKYDHTKIRV	ANFDIEVTSP	120
DGFPEPSQAK	HPIDAITHYD	SIDDRFYVFD	LLNSPYGNVE	EWSIEIAAKL	QEQQGDEVPS	180
EIIDKIIYMP	PDNKEKLME	YLNFWQKQTP	VILTGWNVES	FDIPYVYNRI	KNIFGESTAK	240
RLSPHRKTRV	KVIENMYGSR	EIITLFGISV	LDYIDLYKKF	SFTNQPSYSL	DYISEFELNV	300
GKLKYDGPIS	KLRESNHQRY	ISYNIIDVYR	VLQIDAKRQF	INLSLDMGY	AKIQIQSVFS	360
PIKTDWAIIF	NSLKEQNKVI	PQGRSHPVQP	YPGAFVKEPI	PNRYKYVMSF	DLTSLYPSII	420
RQVNISPETI	AGTFKVAPLH	DYINAVAERP	SDVYSCSPNG	MMYYKDRDGV	VPTEITKVFN	480
QRKEHKGYML	AAQNGNEIIK	EALHNPNSLV	DEPLDVDYRF	DFSDEIKEKI	KKLSAKSLNE	540
MLFRAQRTFV	AGMTAQINRK	LLINSLYGAL	GNVWFRYVDL	RNATAITTFG	QMALQWIERK	600
VNEYLNEVCG	TEGEAFVLVG	DTDSIYVSAD	KIIDKVGESK	PRDTNHWVDF	LDKFARERME	660
PAIDRGFREM	CEYMNKQHL	MFMDREAIA	PPLGSKGIGG	FWTGKKRYAL	NVWDMEGTRY	720
AEPKLKIMGL	ETQKSSTPKA	VQKALKECIR	RMLQEGEESL	QEYFKEFEKE	FRQLNYISIA	780
SVSSANNIAK	YDVGGFPQPK	CPFHIRGILT	YNRAIKGNID	APQVVEGEKV	YVLPREGNP	840
FGDKCIAWPS	GTEITDLIKD	DVLHWMDYTV	LLEKTFIKPL	EGFTSAAKLD	YEKKASLFDM	900
FDF						903

SEQ ID NO: 16 moltype = AA length = 898
 FEATURE Location/Qualifiers
 source 1..898
 mol_type = protein
 organism = Escherichia coli phage

SEQUENCE: 16

MKEFYISLET	VGNNIVERYI	DENGKERTRE	VEYLPMTFRH	CKEESKYKDI	YGKNCAPQKF	60
PSMKDARDWM	KRMEDIGLEA	LGMNDFKLAY	ISDTYGSEIV	YDRKFVRVAN	CDIEVTGDKF	120
PDPMAEYEI	DAITHYDSID	DRFYVFDLLN	SMYGSVSKWD	AKLAAKLDCE	GGDEVPPQEL	180
DRVIYMPFDN	ERDMLMEYIN	LWEQKRPAIF	TGWNIEGPDV	PYIMNRVKMI	LGRSMKRFS	240
PIGRVKSCLI	YDMYGSKEIY	SIDGVSILDY	LDLYKKFAFT	NLPSFVLESV	AQHETKKGKL	300
PYDGPINKLR	ETNHQRYISY	NIIDVESVQA	IDKIRGFIDL	VLSMSYYAKM	PFSGVMSPIK	360
TWDAIIFNSL	KGEHKVIPQQ	GSHVQKSPFG	AFVFEKPIA	RRYIMSPDLT	SLYPSIIRQV	420
NISPETIRGQ	FKVHPHIEYI	AGTAPKPSDE	YSCSPNGWMY	DKHQEGIIPK	EIAKVFFQK	480
DWKKMFAEE	MNAEAIKKII	MKGAGSCSTK	PEVERYVKFS	DDFLNELSNY	TESVLNSLIE	540
ECEKAATLAN	TNQLNRKILI	NSLYGALGNI	HFYRYDLRNA	TAITIFGQVG	IQWIARKINE	600
YLNKVCGTND	EDFIAAGTDT	SVYVCVDKVI	EKVGLDRFKE	QNDLVEFMNQ	FGKKMEPMI	660
DVAYRELCDY	MNNREHLMHM	DREAI SCPPL	GSKGVGGFWK	AKKRYALNVY	DMEDKRFAEP	720
HLKIMGMETQ	QSSTPKAVQE	ALEESIRRL	QEGEESVQY	YKNPEKEYRQ	LDYKVI AEVK	780
TANDIAKYDD	KGWPGFKCPF	HIRGVLTYYR	AVSGLGVAPI	LDGNKVMVLP	LREGNPFQDK	840
CIAWPSGTEL	PKEIRSDVLS	WIDHSTLFQK	SPVKPLAGMC	ESAGMDYEEK	ASLDPLFG	898

SEQ ID NO: 17 moltype = AA length = 575
 FEATURE Location/Qualifiers
 source 1..575
 mol_type = protein
 organism = Bacillus phage

SEQUENCE: 17

MKHMPRKMY	CDFETTTKVE	DCRVWAYGYM	NIEDHSEYKI	GNSLDEFMAW	VCLKVQADLYF	60
HNLKFDGAFI	INWLERNGFK	WSADGLPNTY	NTIISRMGQW	YMIDICLYGK	GKRKIHTVIY	120
DSLKLFPFV	KKIAKDFKLT	VLKGDIDYHK	ERPVGKITYP	BEYAYIKNDI	QIIAEALLIQ	180
FKQLDRMTA	GSDSLKGFKD	IITTKKFKKV	FPTLSLGLDK	EVRYAYRGGF	TWLNDRFKKE	240
EIGEGMVFVD	NSLYPAGPIY	RLLPYGEPIV	FEGKYVWDED	YPLHIQHIRC	EFELEKGYIP	300
TIQIKRSRFP	KGNEYLKSSG	GEIADLWLSN	VDLELMKEHY	DLYNVEYISG	LKFKATTGLF	360
KDFIDKWTYI	KTTSEGAIKQ	LAKLMLNSLY	GKFASNPVDT	GKVPYLKENG	ALGFRLEGEE	420
TKDPVYTPMG	VFITAWARYT	TITAAQACYD	RIIYCDTDSI	HLTGTEIPDV	IKDIVDPKKL	480
GYWAHESTFK	RAKYLRLQKTY	IQDIYMKEVD	GKLVEGSPDD	YTDIKFSVKC	AGMTDKIKKE	540
VTFENFKVGF	SRKMKPKPVQ	VPGGVVLVDD	TFTIK			575

-continued

```

SEQ ID NO: 18      moltype = DNA  length = 45
FEATURE            Location/Qualifiers
misc_feature        1..45
                    note = FAM-45-mer DNA initiator
source              1..45
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 18
ctcggcctgg cacaggtccg ttcagtgtg cggcgaccac cgagg

```

What is claimed is:

1. An DNA polymerase variant having a template-independent RNA synthesis function and modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NOs: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17, wherein:

the amino acid L at position 408 of SEQ ID NO: 2, 3, or 4 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid L at position 409 of SEQ ID NO: 5 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid L at position 411 of SEQ ID NO: 6 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid L at position 485 of SEQ ID NO: 7 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid M at position 426 of SEQ ID NO: 8 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid L at position 518 of SEQ ID NO: 9 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

the amino acid L at position 417 of SEQ ID NO: 10 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 606 of SEQ ID NO: 11 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 612 of SEQ ID NO: 12 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 425 of SEQ ID NO: 13 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 423 of SEQ ID NO: 14 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 415 of SEQ ID NO: 15 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y;

an amino acid L at position 412 of SEQ ID NO: 16 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y; or

an amino acid L at position 253 of SEQ ID NO: 17 is substituted with A, C, D, F, G, H, K, M, N, Q, S, W, or Y, and

wherein the RNA polymerase variant has a reduced or deficient 3' to 5' exonuclease activity.

2. The DNA polymerase variant of claim 1, wherein the wild-type B-family DNA polymerase has the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8 and 9.

3. The DNA polymerase variant of claim 2, wherein the wild-type B-family DNA polymerase is *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanosarcina acetivorans* DNA polymerase (Mac), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Methanococcus maripaludis* DNA polymerase (Mma), human DNA polymerase, delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), *Escherichia coli* phage RB69 DNA polymerase (RB69), *Escherichia coli* phage T4 DNA polymerase (T4), or *Bacillus* phage Phi29 DNA polymerase (Phi29).

4. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein:

the amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

5. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus gorgonarius* DNA polymerase (Tgo) having a wild-type amino acid sequence of SEQ ID NO: 2; and wherein:

the amino acid Y at position 409 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 410 of SEQ ID NO: 2 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 485 of SEQ ID NO: 2 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

6. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein:

the amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

7. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO: 3; and wherein:

- i. the amino acid Y at position 409 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, D, G, N, S, T or V;
- ii. the amino acid P at position 410 of SEQ ID NO: 3 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and
- iii. the amino acid A at position 485 of SEQ ID NO: 3 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

8. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein:

the amino acid Y at position 409 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 410 of SEQ ID NO: 4 remains unchanged or is A, C, G, I, L, M, N, S, T or V.

9. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus* sp. (strain 9° N-7) DNA polymerase (9° N) having a wild-type amino acid sequence of SEQ ID NO: 4; and wherein:

the amino acid Y at position 409 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 410 of SEQ ID NO: 4 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 485 of SEQ ID NO: 4 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

10. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein:

the amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

11. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO: 5; and wherein:

the amino acid Y at position 410 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 411 of SEQ ID NO: 5 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 486 of SEQ ID NO: 5 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

12. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus*

litoralis DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein:

the amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

13. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO: 6; and wherein:

the amino acid Y at position 412 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 413 of SEQ ID NO: 6 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 488 of SEQ ID NO: 6 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

14. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein:

the amino acid Y at position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

15. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO: 7; and wherein:

the amino acid Y at position 486 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 487 of SEQ ID NO: 7 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 565 of SEQ ID NO: 7 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

16. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein:

the amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

17. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO: 8; and wherein:

the amino acid Y at position 427 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 428 of SEQ ID NO: 8 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 508 of SEQ ID NO: 8 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

18. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein:

the amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V; and

the amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V.

19. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO: 9; and wherein:

the amino acid Y at position 519 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, D, G, N, S, T or V;

the amino acid P at position 520 of SEQ ID NO: 9 remains unchanged or is substituted with A, C, G, I, L, M, N, S, T or V; and

the amino acid A at position 601 of SEQ ID NO: 9 is substituted with C, D, E, F, G, H, I, K, L, M, N, P, Q, R, T, V, W or Y.

20. The DNA polymerase variant of claim 1, wherein the DNA polymerase variant exhibits an activity of synthesizing nucleic acids in a template-independent manner by adding at least one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator.

21. The DNA polymerase variant of claim 20, wherein the extendible initiator comprises a single-stranded oligonucle-

otide initiator, a blunt-ended double-stranded oligonucleotide initiator, or a mixture thereof.

22. The DNA polymerase variant of claim 20, wherein the extendible initiator is a free form nucleic acid to be reacted in a liquid phase.

23. The DNA polymerase variant of claim 20, wherein the extendible initiator is immobilized on a solid support, wherein the solid support comprises a particle, bead, slide, array surface, membrane, flow cell, well, microwell, nanowell, chamber, microfluidic chamber, channel, or microfluidic channel.

24. The DNA polymerase variant of claim 20, wherein the at least one nucleotide is linked with a detectable label.

25. The DNA polymerase variant of claim 20, wherein the at least one nucleotide comprises a ribose.

26. The DNA polymerase variant of claim 20, wherein the DNA polymerase variant exhibits the activity at reaction temperatures ranging from 10° C. to 100° C.

27. A kit for performing de novo enzymatic nucleic acid synthesis, comprising the DNA polymerase variant of claim 1, wherein the DNA polymerase variant exhibits activity of synthesizing nucleic acids in a template-independent manner by adding at least one nucleotide selected from the group of naturally occurring nucleotide, nucleotide analogue, or a mixture thereof, to an extendible initiator, thereby synthesizing a desired nucleic acid sequence.

28. A method for template-independent synthesis of an RNA oligonucleotide, comprising:

- (a) providing an initiator oligonucleotide;
- (b) providing the DNA polymerase variant of claim 1;
- (c) combining the initiator oligonucleotide, the DNA polymerase variant, and one or more nucleotides under conditions sufficient for the addition of at least one nucleotide to the 3' end of the initiator oligonucleotide.

* * * * *