



[45] **Date of Patent:** **Mar. 23, 1999**

- Page 1

1081/361	ttc	cag	gag	ggg	cgg	gac	atc	cac	acg	gag	acc	gcc	agg	tgg	atg	ttc	ggc	gtc	ccc	cgg	gag	gcc	gtg	gac	ccc	ctg	atg	cgc	cgg	gcg
1171/391	F	Q	E	G	R	D	I	H	T	E	T	A	S	W	M	F	G	V	P	R	E	A	V	D	P	L	M	R	R	A
1261/421	gcc	aag	acc	atc	aac	tac	ggg	gtc	ctc	tac	ggc	atg	tcg	gcc	cac	cgc	ctc	tcc	cag	gag	cta	gcc	atc	cct	tac	gag	gag	gcc	cag	gcc
1351/451	A	K	T	I	N	Y	G	V	L	Y	G	M	S	A	H	R	L	S	Q	E	L	A	I	P	Y	E	E	A	Q	A
1441/481	ttc	att	gag	cgc	tac	ttt	cag	agc	ttc	ccc	aag	gtg	cgg	gcc	tgg	att	gag	aag	acc	ctg	gag	gag	ggc	agg	agg	cgg	ggg	tac	gtg	gag
1531/511	F	I	E	R	Y	F	Q	S	F	P	K	V	R	A	W	I	E	K	T	L	E	G	R	R	G	Y	V	E	E	
1621/541	acc	ctc	ttc	ggc	cgc	cgc	cgc	tac	gtg	cca	gac	cta	gag	gcc	cgg	gtg	aag	agc	gtg	cgg	gag	gag	gcc	cgc	atg	gcc	ttc	aac	atg	
1711/561	T	L	F	G	R	R	Y	V	P	D	L	E	A	R	V	K	S	V	R	E	A	E	R	M	A	F	N	M		
1801/581	ccc	gtc	cag	ggc	acc	gcc	gcc	gac	ctc	atg	aag	ctg	gct	atg	gtg	aag	ctc	ttc	ccc	agg	ctg	gag	gaa	atg	ggg	gcc	agg	atg	ctc	ctt
1891/601	P	V	Q	G	T	A	A	D	L	M	K	L	A	M	V	K	L	F	P	R	L	E	E	M	G	A	R	M	L	L
1981/621	cag	gtc	cac	gac	gag	ctg	gtc	ctc	gag	gcc	cca	aaa	gag	agg	gcg	gag	gcc	gtg	gcc	cgg	ctg	gcc	aag	gag	gtc	atg	gag	ggg	gtg	tat
2071/641	Q	V	H	D	E	L	V	L	E	A	P	K	E	R	A	E	A	V	A	R	L	A	K	E	V	M	E	G	V	Y
2161/661	ccc	ctg	gcc	gtg	ccc	ctg	gag	gtg	gag	gtg	ggg	ata	ggg	gag	gac	tgg	ctc	tcc	gcc	aag	gag	tga								
2251/681	P	L	A	V	P	L	E	V	E	V	G	I	G	E	D	W	L	S	A	K	E	*								

Fig. 1B

1/1 atg gcg atg ctt ccc ctc ttt gag ccc aaa 31/11 61/21 tac cgc acc ttc ttt gcc ctc aag
M A L P L F E P K G R L L V D G H H L A A F A L K
91/31 ggc ctc acc acc agc cgc ggc gaa ccc gtt 121/41 151/51 ggc ctc aag gag gac ggg gac gtg
G L T T S R G E P V Q A V Y G F A K S L L K A L K E D G D V
181/61 gtg gtc ttt gag gcc aag gcc ccc 211/71 241/81 ggc cgg gcc ccc acc ccg gag gac
V V V F D A K A P S F R H E A Y E A Y K A A G G R A P T P E D
271/91 ttt ccc cgg cag ctg gcc ctc atc aag gag ttg gtg gac cta ggc ctt gtg cgg ctg ggt 331/111
F P R Q L A L I K E L V D L L G L V R L L E V E V D R R A A G G G R A P T P E D
361/121 ctg gcc acc ctg gcc aag cgg gcg gaa aag gag ttg gtg gac cta ggc ctt gtg cgg ctg ggt 421/141
L A T L A K R A E K E G Y E V R I L T A D R R A A G G G R A P T P E D
451/151 cgc atc atc ctc cac cct gag ggg tac ctg atc 481/161 511/171 ctg cgc ccg gag cag tgg gtg gac
R I A I L H P E G Y L I T P A W L Y E K Y G L R P E Q W V D
541/181 tac cgg gcc ctg gcg ggg gac ccc tgg gat aac atc 571/191 601/201 gcc cag agg ctc atc cgc gag tgg
Y R A L A G D P S D N I P G V K G I G E K T A Q R L I R E W
631/211 ggg agc ctg gaa aac ctc ttc cag cac ctg gac cag 661/221 691/231 ggc ggc atg gag gcc ctg gcc ctt
G S L E N L F Q H L D Q Q P S L R E K L Q A G M E A L A L
721/241 tcc cgg aag ctt tcc cag gtg cac act gag ccc 751/251 781/261 ccc aac ctg gag ggt ctg cgg gct
S R K L S Q V H T D L P L P L E V D F G R R T Cgc aca
811/271 ttt ttg gag cgg ttg gag ttt gga agc ctc ctc 841/281 871/291 gca gag gag gcc ccc tgg ccc cct
F L E R L E F F G S L L H L L G L L E G P K A A A E A P W P P
901/301 ccg gaa ggg gct ttt ttg ggc ttt tcc cgt 931/311 961/321 ctg gcc ctg gct ggg gcg tgg gag ggg cgc
P E G A F L G F S S R P E F G L L W A E L A A L A G A W E G R
991/331 ctc cat cgg gca caa gac ccc ctt agg ggc ctg agg gac ctt aag ggg gtg cgg gga atc 1051/351
L H R A Q D P L R G L R D L K G V R G I L A K D L A V L A L
ctg

Fig. 3A

1081/361 cgg gag ggc ctg gac ctc ttc cca gag gac 1111/371 gac ccc atg ctc ctg gcc tac ctt ctg gac 1141/381 ccc tcc aac acc acc cct gag ggg gtg gcc
 R E G L D L F P P E D D P M L L A Y L L D P S N T T P E G V A
 1171/391 cgg cgt tac ggg ggg gag tgg acg gag gat 1201/401 gcg ggg gag agg gcc ctc ctg gcc gag cgc 1231/411 ctc ttc cag acc cta aag gag cgc ctt aag
 R R Y G G G E W T T E D A G E R A L L A E R L F Q T L K E R L K
 1261/421 gga gaa gaa cgc ctg ctt tgg ctt tac gag 1291/431 gag ggg gag aag ccg ctt tcc cgg gtg ttg gcc cgg atg gag gcc acg ggg gtc cgg ctg
 G E R L L L W L Y E F V E K P L S R V L A R M E A T G G V R L
 1351/451 gac gtg gcc tac ctc cag gcc ctc tcc ctg gac gtg ggc gag ctg gac 1411/471 ttc cgc ctg gcc ggc cac ccc
 D V A Y L Q A L S L E V E A E V F R L A G H P
 1441/481 ttc aac ctc aac tcc cgc gac cag ctg gag 1471/491 cgg gtg ctc ttt gac gag ctg ggc ctg gac 1501/501 gcc atc ggc aag acg gag aag acg ggg aaa
 F N L N S R D Q L E R V L F D E L G L P A I G K T E K T G K
 1531/511 cgc tcc acc agc gct gcc gtg gac ggc ctg gag gac 1561/521 ctg cga gag gcc cac ccc atc gtg gac cgc 1591/531 atc ctg cag tac cgg gag ctc acc aag ctc
 R S T S A A V L E A L R E A H P I V D R I L Q Y R E L T K L
 1621/541 aag aac acc tac ata gac ccc ctg gcc 1651/551 ctg gtc cac ccc aag acc ggc cgc ctc acc ggc acc ggc acg
 K N T Y I D P L P A L V H P K T G R L H T R F N Q T A T A T
 1711/571 ggc agc ctt tcc agc tcc gac ccc aac ctg gcc 1741/581 cag aac atc ccc gtg cgc acc cct ctg gcc 1771/591 cag cgc atc cgc cga gcc ttc gtg gcc gag
 G R L S S S D P P L P A L Q N I P V R T P L G Q R I R R A F V A E
 1801/601 gag ggc tgg gtg ctg gtg gtc ttg gac tac 1831/611 agc cag att gag ctt cgg gtc ctg gcc cac 1861/621 ctc tcc ggg gac gag aac ctg atc cgg gtc
 E G W V L V V L D Y S Q I E L R V L A H L S G D E N L I R V
 1891/631 ttt cag gag ggg agg gac atc cac acc cag 1921/641 acc gcc agc tgg atg ttc ggc gtt tcc ccc 1951/651 gaa ggg gta gac cct ctg atg cgc cgg gcg
 F Q E G R D I H T Q T A S W M F G V S P E G V D P L M R R A
 1981/661 ggc aag acc atc aac ttc ggg gtg ctc tac 2011/671 ggc atg tcc gcc cac cgc ctc tcc ggg gag 2041/681 ctt tcc atc ccc tac gag gag gcg gtg gcc
 A K T I N F G V L Y G M S A H R L S G E L S I P Y E A V A
 2071/691 ttc att gag cgc tac ttc cag agc tac ccc aag gtg 2101/701 cgg gcc tgg att gag ggg acc ctc 2131/711 gag gag ggc cgc cgg ggg tat gtg gag
 F I E R Y F Q S Y P K V R A W I E G T L E E G R R G Y V E

Fig. 3B

2161/721

acc ctc ttc ggc cgc cgg cgg tat gtg ccc

T L F G R R R Y V P

2251/751

ccg gtc cag ggc acc gcc cgc ctc atg

P V Q G T A A A D L M

2341/781

cag gtg cac gac gag ctg gtc ctc gag gcc

Q V H D E L L V L E A

2431/811

ccc ctg cag gtg ccc ctg gag gtg gag

P L Q V P L L E V E V

2191/731

gac ctc aac gcc cgg gtg aag agc gtg cgc

D L N A R V K S V R

2281/761

aag ctg gcc atg gtg cgg ctt ttc ccc cgg

K L A M V R L F P R

2371/791

ccc aag gac cgg gcg gag agg gta gcc gct

P K D R A E R V A A

2461/821

ggc ctg ggg gag gac tgg ctc tcc gcc aag

G L G E D W L S A K

2221/741

gag gcg gcg gag cgc atg gcc ttc aac atg

E A A E R M A F N M

2311/771

ctt cag gaa ctg ggg gcg agg atg ctt ttg

L Q E L G A R M L L

2401/801

ttg gcc aag gag gtc atg gag ggg gtc tgg

L A K E V M E G V W

2491/831

gag tag

E *

Fig. 3C

1/1	ATG	GAG	GCG	ATG	CTT	CCG	CTC	TTT	GAA	CCC	AAA	GGC	CGG	GTC	CTC	CTG	GTG	GAC	GGC	CAC	CTG	GCC	TAC	CGC	ACC	TTC	TTC	GCC	CTG		
91/31	M	E	A	M	L	P	L	F	E	P	K	G	R	V	L	L	V	D	G	H	L	A	Y	R	T	F	F	A	L		
181/61	AAG	GGC	CTC	ACC	ACG	AGC	CGG	GGC	GAA	CCG	GTG	CAG	GCG	GTC	TAC	GGC	TTC	GCC	AAG	AGC	CTC	CTC	AAG	GCC	CTG	AAG	GAG	GAC	GGG	TAC	
271/91	K	G	L	T	T	S	R	G	E	P	V	Q	A	V	Y	G	F	A	K	S	L	L	K	A	L	K	E	D	G	Y	
361/121	AAG	GCC	GTC	TTC	GTG	GTC	TTT	GAC	GCC	AAG	GCC	CCC	TCC	TTC	CGC	CAC	GAG	GCC	TAC	GAG	GCC	TAC	AAG	GCG	GGG	AGG	GCC	CCG	ACC	CCC	
451/151	K	A	V	F	V	V	F	D	A	K	A	P	S	F	R	H	E	A	Y	E	A	Y	K	A	G	R	A	P	T	P	
541/181	GAG	GAC	TTC	CCC	CGG	CAG	CTC	GCC	CTC	ATC	AAG	GAG	CTG	GTG	GAC	CTC	CTG	GGG	TTT	ACC	CGC	CTC	GAG	GTC	CCC	GGC	TAC	GAG	GCG	GAC	
631/211	E	D	F	P	R	Q	L	A	L	I	K	E	L	V	D	L	L	G	F	T	R	L	E	V	P	G	Y	E	A	D	
721/241	TCC	GAC	CGC	GTC	GCC	GTG	CTC	CAC	CCC	GAG	GGC	CAC	CTC	ATC	ACC	CCG	GAG	TGG	CTT	TGG	GAG	AAG	TAC	GGC	CTC	AGG	CCG	GAG	CAG	TGG	
811/271	S	D	R	V	A	V	L	H	P	E	G	H	L	I	T	P	E	W	L	W	E	K	Y	G	L	R	P	E	Q	W	
901/301	541/181	GTG	GAC	TTC	CGC	GCC	CTC	GTG	GGG	GAC	CCC	TCC	GAC	AAC	CTC	CCC	GGG	GTG	AAG	GGC	ATC	GGG	GAG	AAG	ACC	GCC	CTC	AAG	CTC	AAG	
991/331	V	D	F	R	A	L	V	G	D	P	S	D	N	L	P	G	V	K	G	I	G	E	K	T	A	L	K	L	L	K	
1021/341	631/211	GGA	AGC	CTG	GAA	AAC	CTC	CTC	AAG	AAC	CTG	GAC	CGG	GTA	AAG	CCA	GAA	AAC	GTC	CGG	GAG	AAG	ATC	AAG	GCC	CAC	CTG	GAA	GAC		
1051/351	721/241	CTC	AGG	CTC	TCC	TTG	GAG	CTC	TCC	CGG	GTG	ACC	GAC	CTC	CCC	CTG	GAG	GTG	GAC	CTC	GCC	CAG	GGG	CGG	GAG	CCC	GAC	CGG	GAG	GGG	
1081/361	811/271	L	R	A	F	L	E	R	L	E	F	S	D	L	H	E	F	G	L	L	E	A	Q	G	R	E	P	D	R	E	G
1111/371	901/301	GCC	TTC	CTG	GAG	AGG	CTG	GAG	TTT	GAG	TTT	GAG	CTC	CTC	CAC	GAG	TTC	GGC	CTC	CTG	GAG	GCC	CCC	GCC	CCC	CTG	GAG	GAG	GCC	CCC	
1141/381	991/331	CTT	AGG	GCC	CCG	GAA	GGG	GCC	TTC	GTG	GGC	TTC	GTG	TCC	CGC	CCC	GAG	CCC	ATG	TGG	GCG	GAG	CTT	AAA	GCC	CTG	GCC	GCC	TGC	AGG	
1171/391	D	G	R	V	H	R	A	A	D	P	L	A	G	L	K	D	L	K	E	V	R	G	L	A	K	D	L	A	V		

Fig. 4A

1081/361	TTG GCC TCG AGG GAG GGG CTA GAC CTC GTG CCC GGG GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCC TCC AAC ACC ACC CCC GAG	1111/371	CCC GGG GAC GAC CCC ATG CTC CTC GCC TAC CTC CTG GAC CCC TCC AAC ACC ACC CCC GAG	1141/381	CTC CTC CTG GAC CCC TCC AAC ACC ACC CCC GAG
1171/391	GGG GTG GCG CGG TAC GGG GAG TGG AGG GAG GAC GCG GCC CAC CGG GCC CTC CTC CGG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1201/401	GGG GAG TGG AGG GAG GAG TGG AGG GAG GAC GCG GCC CAC CGG GCC CTC CTC CGG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1231/411	TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG
1261/421	CGC CGG CGG TAC GGG GAG TGG AGG GAG GAC GCG GCC CAC CGG GCC CTC CTC CGG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1291/431	CGC CGG CGG TAC GGG GAG TGG AGG GAG GAC GCG GCC CAC CGG GCC CTC CTC CGG GAG AGG CTC CAT CGG AAC CTC CTT AAG	1321/441	TCG GAG AGG CTC CAT CGG AAC CTC CTT AAG
1351/451	CGC CGG GAG GAG AAG CTC CTT TGG TCG TAC CAC GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG	1381/461	CGC CGG GAG GAG AAG CTC CTT TGG TCG TAC CAC GAG GTG GAA AAG CCC CTC TCC CGG GTC CTG GCC CAC ATG GAG GCC ACC GGG	1411/471	CGC CTC GAG GAG GTG GTC TTC CGC TTG GCG
1441/481	GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC GAC GTG GAA AGG GTG CTC TTT GAC GAG ATC CGC GAG GAG GAG GTC TTC CGC TTG GCG	1471/491	GGC CAC CCC TTC AAC CTC AAC TCC CGG GAC GAC GTG GAA AGG GTG CTC TTT GAC GAG ATC CGC GAG GAG GAG GTC TTC CGC TTG GCG	1501/501	AGG CTT CCC GCC TTG GGG AAG ACG CAA AAG
1531/511	CGC TCC ACC AGC GGC GCG GTG CTG GAG GCG CTA CGG GAG GCC CAC CCC ATC CCC ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC	1561/521	CGC TCC ACC AGC GGC GCG GTG CTG GAG GCG CTA CGG GAG GCC CAC CCC ATC CCC ATC GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC	1591/531	GTG GAG AAG ATC CTC CAG CAC CGG GAG CTC
1621/541	ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG AGG ACG GGC CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC	1651/551	ACC AAG CTC AAG AAC ACC TAC GTG GAC CCC CTC CCA AGC CTC GTC CAC CCG AGG ACG GGC CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC	1681/561	CGC CTC CAC ACC CGC TTC AAC CAG ACG GCC
1711/571	GGG AGG CTT AGT AGC TCC GCG TTG GTG GCG TCG AAC CTC CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1741/581	GGG AGG CTT AGT AGC TCC GCG TTG GTG GCG TCG AAC CTC CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1771/591	CGC TTC AAC CAG ACG GCC
1801/601	GGT TGG GCG TTG GTG GCG TCG AAC CTC CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1831/611	GGT TGG GCG TTG GTG GCG TCG AAC CTC CAG AAC ATC CCC GTC CGC ACC CCC TTG GGC CAG AGG ATC CGC CGG GCC TTC	1861/621	TCC GGG GAC GAA AAC CTG
1891/631	ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTT GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG	1921/641	ATC AGG GTC TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTT GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG	1951/651	GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG
1981/661	CGC CGG GCG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTT GCC ATC CCC TAC GAG GAG	2011/671	CGC CGG GCG GCC AAG ACG GTG AAC TTC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTT GCC ATC CCC TAC GAG GAG	2041/681	TCC CAG GAG CTT GCC ATC CCC TAC GAG GAG
2071/691	TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG GCG TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC	2101/701	TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG GCG TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC	2131/711	AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC

Fig. 4B

2161/721	TAC GTG GAA	ACC CTC TTC GGA AGA AGG CGC	CTC AAC GCC GCG GTG AAG	AGC GTC AGG	2221/741	GAG GCC GCG GAG CGC ATG GCC
Y V E	T L F	R R R	D L N A R V	S V R	E A A E R M A	
2251/751	TTC AAC ATG	CCC GGC ACC GGC	CTC GGC ATG GTG AAG CTC	2311/771	TTC CCC CGC CTC	CGG GAG ATG GGG GCC CGC
F N M	P V Q	A A A	AAG CTC L A M V K L	F P R	L R E M G A R	
2341/781	ATG CTC CTC	CAG GAC GAG CTC	CTC CAA GCG CGG GCC GAG GAG	2401/801	GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG	
M L L	Q V H	D E L L	P Q A R A E E	V A A L A K E A M E		
2431/811	AAG GCC TAT	CCC CTC GCC GTG CCC	CTG GAG GTG GGG ATG GGG GAG GAC TGG CTT	2491/831	TCC GCC AAG GGT TAG	
K A Y	P L A V	P P L	G M G E D W L S A K G *			

Fig. 4C

[illegible]

Fig. 5A

1081/361 TTC CAG GAG GGG AAG GAC ATC CAC ACC CAG ACC GCA AGC TGG ATG TTC GGC GTC CCC CCG GAG GCC GTG GAC CCC CTG ATG CGC CGG GCG
F Q E G K D I H T Q T A S W M F G V P P E A V D G A V GTG GAG GCC GTG GCG CGG R A
1171/391 GCC AAG ACG GTG AAC TAC GGC GTC CTC TAC GGC ATG TCC GCC CAT AGG CTC TCC CAG GAG CTA GCC ATC CCC TAC GAA GAA GCG GTG GCC
A K T V N Y G V L Y . G M S A H R L S Q E L A I P Y E A V A
1261/421 TTT ATA GAG CGC TAC TTC CAA AGC TTC CCC AAG GTG CGG GCC TGG ATA GAA AAG ACC CTG GAG GAG GGG AGG AAG CGG GGC TAC GTG GAA
F I E R Y F Q S F P K V R A W I E K T L E G R K R G Y V E
1351/451 ACC CTC TTC GGA AGA AGG CGC TAC GTG CCC GAC CTC AAC GCC CGG GTG AAG AGC GTC AGG GAG GCC GCG GAG CGC ATG GCC TTC AAC ATG
T L F G R R Y V P D L N A R V K S V R E A A E R M A F N M
1441/481 CCC GTC CAG GGC ACC GCC GAC CTC ATG AAG CTC GCC ATG GTG AAG CTC TTC CCC CCG GAG ATG GGG GCC CGC ATG CTC CTC
P V Q G T A A D L M K L A M V K L F P R L R E M G A R M L L
1531/511 CAG GTC CAC GAC GAG CTC CTC CTG GAG GCC CTC CAA GCG CGG GCC GAG GAG GTG GCG GCT TTG GCC AAG GAG GCC ATG GAG AAG GCC TAT
Q V H D E L L L E A P Q A R A E E V A A L A K E A M E K A Y
1621/541 CCC CTC GCC GTG CCC CTG GAG GTG GAG GTG GGG ATG GGG GAG GAC TGG CTT TCC GCC AAG GGT TAG
P L A V P L E V E V G M G E D W L S A K G *

Fig. 5B

THERMOSTABLE DNA POLYMERASES

This application is a Continuation-in-Part application of U.S. Ser. No. 08/455,686, filed May 31, 1995, now abandoned.

BACKGROUND OF THE INVENTION

The present invention relates to novel thermo-stable DNA polymerases, the genes and vectors encoding them and their use in DNA sequencing.

U.S. Pat. Nos. 4,889,818 and 5,079,352 describe the isolation and expression of a DNA polymerase known as Taq DNA Polymerase (hereinafter referred to as Taq). It is reported that amino-terminal deletions wherein approximately one-third of the coding sequence is absent have resulted in producing a gene product that is quite active in polymerase assays. Taq is described as being of use in PCR (polymerase chain reaction).

U.S. Pat. No. 5,075,216 describes the use of Taq in DNA sequencing.

International patent application WO 92/06/06188 describes a DNA polymerase having an identical amino acid sequence to Taq except that it lacks the N-terminal 235 amino acids of Taq and its use in sequencing. This DNA polymerase is known as Δ Taq.

U.S. Pat. No. 4,795,699 describes the use of T7 type DNA polymerases (T7) in DNA sequencing. These are of great use in DNA sequencing in that they incorporate dideoxy nucleoside triphosphates (NTPs) with an efficiency comparable to the incorporation of deoxy NTPs; other polymerases incorporate dideoxy NTPs far less efficiently which requires comparatively large quantities of these to be present in sequencing reactions.

At the DOE Contractor-Grantee Workshop (Nov. 13-17, 1994, Santa Fe) and the I. Robert Lehman Symposium (Nov. 11-14, 1994, Sonoma), Prof. S. Tabor identified a site in DNA polymerases that can be modified to incorporate dideoxy NTPs more efficiently. He reported that the presence or absence of a single hydroxy group (tyrosine vs. phenylalanine) at a highly conserved position on *E. coli*, DNA Polymerase I, T7, and Taq makes more than a 1000-fold difference in their ability to discriminate against dideoxy NTPs. (See also European Patent Application 94203433.1 published May 31, 1995, Publication No. 0 655 506 A1 and hereby incorporated by reference herein.)

SUMMARY OF THE INVENTION

The present invention provides a DNA polymerase having an amino acid sequence differentiated from Taq in that it lacks the N-terminal 272 amino acids and has the phenylalanine at position 667 (of native Taq) replaced by tyrosine. Preferably, the DNA polymerase has methionine at position 1 (equivalent to position 272 of Taq) (hereinafter referred to as FY2) The full DNA sequence is given as FIG. 1 (SEQ. ID. NO. 1). Included within the scope of the present invention are DNA polymerases having substantially identical amino acid sequences to the above which retain thermostability and efficient incorporation of dideoxy NTPs.

By a substantially identical amino acid sequence is meant a sequence which contains 540 to 582 amino acids that may have conservative amino acid changes compared with Taq which do not significantly influence thermostability or nucleotide incorporation, i.e. other than the phenylalanine to tyrosine conversion. Such changes include substitution of like charged amino acids for one another, or amino acids

with small side chains for other small side chains, e.g., ala for val. More drastic changes may be introduced at noncritical regions where little or no effect on polymerase activity is observed by such a change.

The invention also features DNA polymerases that lack between 251 and 293 (preferably 271 or 272) of the N-terminal amino acids of *Thermus flavus* (Tfl) and have the phenylalanine at position 666 (of native Tfl) replaced by tyrosine; and those that lack between 253 and 295 (preferably 274) of the N-terminal amino acids of *Thermus thermophilus* (Tth) and have the phenylalanine at position 669 (of native Tth) replaced by tyrosine.

By efficient incorporation of dideoxy NTPs is meant the ability of a polymerase to show little, if any, discrimination in the incorporation of ddNTPs when compared with dNTPs. Suitably efficient discrimination is less than 1:10 and preferably less than 1:5. Such discrimination can be measured by procedures known in the art.

One preferred substantially identical amino acid sequence to that given above is that which contains 562 amino acids having methionine at position 1 and alanine at position 2 (corresponding to positions 271 and 272 of native Taq) (hereinafter referred to as FY3). A full DNA sequence is given as FIG. 2. This is a preferred DNA polymerase for expression by a gene of the present invention.

The purified DNA polymerases FY2 and FY3 both give a single polypeptide band on SDS polyacrylamide gels, unlike Δ Taq, having either a phenylalanine or tyrosine at position 667 which forms several polypeptide bands of similar size on SDS polyacrylamide gels.

A second preferred substantially identical amino acid sequence is that which lacks 274 of the N-terminal amino acids of *Thermus thermophilus* having methionine at position 1, and the phenylalanine to tyrosine mutation at position 396 (corresponding to position 669 of native Tth) (hereinafter referred to as FY4). A full DNA sequence is given as FIG. 5 (SEQ. ID. NO. 14).

The present invention also provides a gene encoding a DNA polymerase of the present invention. In order to assist in the expression of the DNA polymerase activity, the modified gene preferably codes for a methionine residue at position 1 of the new DNA polymerase. In addition, in one preferred embodiment of the invention, the modified gene also codes for an alanine at position 2 (corresponding to position 272 of native Taq).

In a further aspect, the present invention provides a vector containing the gene encoding the DNA polymerase activity of the present invention, e.g., encoding an amino acid sequence differentiated from native Taq in that it lacks the N-terminal 272 amino acids and has phenylalanine at position 396 (equivalent to position 667 of Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

In a yet further aspect, the present invention provides a host cell comprising a vector containing the gene encoding the DNA polymerase activity of the present invention, e.g., encoding an amino acid sequence differentiated from native Taq in that it lacks the N-terminal 272 amino acids and has phenylalanine at position 396 (equivalent to position 667 of native Taq) replaced by tyrosine or a substantially identical amino acid sequence thereto.

The DNA polymerases of the present invention are preferably in a purified form. By purified form is meant that the DNA polymerase is isolated from a majority of host cell proteins normally associated with it; preferably the polymerase is at least 10% (w/w) of the protein of a preparation,

even more preferably it is provided as a homogeneous preparation, e.g., a homogeneous solution. Preferably the DNA polymerase is a single polypeptide on an SDS polyacrylamide gel.

The DNA polymerases of the present invention are suitably used in sequencing, preferably in combination with a pyrophosphatase. Accordingly, the present invention provides a composition which comprises a DNA polymerase of the present invention in combination with a pyrophosphatase, preferably a thermostable pyrophosphatase such as *Thermoplasma acidophilum* pyrophosphatase. (Schafer, G. and Richter, O. H. (1992) *Eur. J. Biochem.* 209, 351–355).

The DNA polymerases of the present invention can be constructed using standard techniques. By way of example, mutagenic PCR primers can be designed to incorporate the desired Phe to Tyr amino acid change (FY mutation) in one primer. In our hands these primers also carried restriction sites that are found internally in the sequence of the Taq polymerase gene clone of Delta Taq, pWB253, which was used by us as template DNA. However, the same PCR product can be generated with this primer pair from any clone of Taq or with genomic DNA isolated directly from *Thermus aquaticus*. The PCR product encoding only part of the gene is then digested with the appropriate restriction enzymes and used as a replacement sequence for the clone of Delta Taq digested with the same restriction enzymes. In our hands the resulting plasmid was designated pWB253Y. The presence of the mutation can be verified by DNA sequencing of the amplified region of the gene.

Further primers can be prepared that encode for a methionine residue at the N-terminus that is not found at the corresponding position of Taq, the sequence continuing with amino acid residue 273. These primers can be used with a suitable plasmid, e.g., pWB253Y DNA, as a template for amplification and the amplified gene inserted into a vector, e.g., pRE2, to create a gene, e.g., pRE273Y, encoding the polymerase (FY2). The entire gene can be verified by DNA sequencing.

Improved expression of the DNA polymerases of the present invention in the pRE2 expression vector was obtained by creating further genes, pREFY2pref (encoding a protein identical to FY2) and pREFY3 encoding FY3. A mutagenic PCR primer was used to introduce silent codon changes (i.e., the amino acid encoded is not changed) at the amino terminus of the protein which did not affect the sequence of the polypeptide. These changes led to increased production of FY2 polymerase. FY3 was designed to promote increased translation efficiency in vivo. In addition to the silent codon changes introduced in pREFY2pref, a GCT codon was added in the second position (SEQ. ID. NO. 2), as occurs frequently in strongly expressed genes in *E. coli*. This adds an amino acid to the sequence of FY2, and hence the protein was given its own designation FY3. Both constructs produce more enzyme than pRE273Y.

Silent codon changes such as the following increase protein production in *E. coli*:

- substitution of the codon GAG for GAA;
- substitution of the codon AGG, AGA, CGG or CGA for CGT or CGC;
- substitution of the codon CTT, CTC, CTA, TTG or TTA for CTG;
- substitution of the codon ATA for ATT or ATC;
- substitution of the codon GGG or GGA for GGT or GGC.

The present invention also provides a method for determining the nucleotide base sequence of a DNA molecule.

The method includes providing a DNA molecule annealed with a primer molecule able to hybridize to the DNA molecule; and incubating the annealed molecules in a vessel containing at least one deoxynucleotide triphosphate, and a DNA polymerase of the present invention. Also provided is at least one DNA synthesis terminating agent which terminates DNA synthesis at a specific nucleotide base. The method further includes separating the DNA products of the incubating reaction according to size, whereby at least a part of the nucleotide base sequence of the DNA molecule can be determined.

In preferred embodiments, the sequencing is performed at a temperature above 50° C., 60° C., or 70° C.

In other preferred embodiments, the DNA polymerase has less than 1000, 250, 100, 50, 10 or even 2 units of exonuclease activity per mg of polymerase (measured by standard procedure, see below) and is able to utilize primers having only 4, 6 or 10 bases; and the concentration of all four deoxynucleoside triphosphates at the start of the incubating step is sufficient to allow DNA synthesis to continue until terminated by the agent, e.g., a ddNTP.

For cycle sequencing, the DNA polymerases of the present invention make it possible to use significantly lower amounts of dideoxynucleotides compared to naturally occurring enzymes. That is, the method involves providing an excess amount of deoxynucleotides to all four dideoxynucleotides in a cycle sequencing reaction, and performing the cycle sequencing reaction.

Preferably, more than 2, 5, 10 or even 100 fold excess of a dNTP is provided to the corresponding ddNTP.

In a related aspect, the invention features a kit or solution for DNA sequencing including a DNA polymerase of the present invention and a reagent necessary for the sequencing such as dITP, deaza GTP, a chain terminating agent such as a ddNTP, and a manganese-containing solution or powder and optionally a pyrophosphatase.

In another aspect, the invention features a method for providing a DNA polymerase of the present invention by providing a nucleic acid sequence encoding the modified DNA polymerase, expressing the nucleic acid within a host cell, and purifying the DNA polymerase from the host cell.

In another related aspect, the invention features a method for sequencing a strand of DNA essentially as described above with one or more (preferably 2, 3 or 4) deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and a first chain terminating agent. The DNA polymerase causes the primer to be elongated to form a first series of first DNA products differing in the length of the elongated primer, each first DNA product having a chain terminating agent at its elongated end, and the number of molecules of each first DNA products being approximately the same for substantially all DNA products differing in length by no more than 20 bases. The method also features providing a second chain terminating agent in the hybridized mixture at a concentration different from the first chain terminating agent, wherein the DNA polymerase causes production of a second series of second DNA products differing in the length of the elongated primer, with each second DNA product having the second chain terminating agent at its elongated end. The number of molecules of each second DNA product is approximately the same for substantially all second DNA products differing in length from each other by from 1 to 20 bases, and is distinctly different from the number of molecules of all the first DNA products having a length differing by no more than 20 bases from that of said second DNA products.

In preferred embodiments, three or four such chain terminating agents can be used to make different products and

the sequence reaction is provided with a magnesium ion, or even a manganese or iron ion (e.g., at a concentration between 0.05 and 100 mM, preferably between 1 and 10 mM); and the DNA products are separated according to molecular weight in four or less lanes of a gel.

In another related aspect, the invention features a method for sequencing a nucleic acid by combining an oligonucleotide primer, a nucleic acid to be sequenced, between one and four deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and at least two chain terminating agents in different amounts, under conditions favoring extension of the oligonucleotide primer to form nucleic acid fragments complementary to the nucleic acid to be sequenced. For example, the chain terminating agent may be a dideoxynucleotide terminator for adenine, guanine, cytosine or thymine. The method further includes separating the nucleic acid fragments by size and determining the nucleic acid sequence. The agents are differentiated from each other by intensity of a label in the primer extension products.

While it is common to use gel electrophoresis to separate DNA products of a DNA sequencing reaction, those in the art will recognize that other methods may also be used. Thus, it is possible to detect each of the different fragments using procedures such as time of flight mass spectrometry, electron microscopy, and single molecule detection methods.

The invention also features an automated DNA sequencing apparatus having a reactor including reagents which provide at least two series of DNA products formed from a single primer and a DNA strand. Each DNA product of a series differs in molecular weight and has a chain terminating agent at one end. The reagents include a DNA polymerase of the present invention. The apparatus includes a separating means for separating the DNA product along one axis of the separator to form a series of bands. It also includes a band reading means for determining the position and intensity of each band after separation along the axis, and a computing means that determines the DNA sequence of the DNA strand solely from the position and intensity of the bands along the axis and not from the wavelength of emission of light from any label that may be present in the separating means.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The drawings will first briefly be described.

DRAWINGS

FIGS. 1–4 are the DNA sequences, and corresponding amino acid sequences, of FY2, FY3, and the DNA polymerases of *T. flavus* and *Thermus thermophilus*, respectively. FIG. 5 is the DNA sequence and corresponding amino acid sequence of FY4.

EXAMPLES

The following examples serve to illustrate the DNA polymerases of the present invention and their use in sequencing.

Preparation of FY DNA Polymerases (FY2 and FY3)

Bacterial Strains

E. coli strains: MV1190 [Δ (srl-recA) 306::Tn10, Δ (lac-proAB), thi, supE, F' (traD36 proAB⁺ lacI^q lacZ Δ M15)];

DH λ ⁺ [gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44, λ ⁺ M5248 [λ (bio275, cI857, cIII⁺, N⁺, Δ (H1))]].

PCR

Reaction conditions based on the procedure of Barnes (91 *Proc. Nat'l. Acad. Sci.* 2216–2220, 1994) were as follows: 20 mM Tricine pH8.8, 85 mM KOAc, 200 mM dNTPs, 10% glycerol, 5% DMSO, 0.5 mM each primer, 1.5 mM MgOAc, 2.5 U HotTub (Amersham Life Science Inc.), 0.025 U DeepVent (New England Biolabs), 1–100 ng target DNA per 100 ml reaction. Cycling conditions were 94° C. 30 s, 68° C. 10 m40 s for 8 cycles; then 94° C. 30 s, 68° C. 12 m00 s for 8 cycles; then 94° C. 30 s, 68° C. 13 m20 s for 8 cycles; then 94° C. 30 s, 68° C. 14 m40 s for 8 cycles.

In vitro mutagenesis

Restriction enzyme digestions, plasmid preparations, and other in vitro manipulations of DNA were performed using standard protocols (Sambrook et al., *Molecular Cloning* 2nd Ed. Cold Spring Harbor Press, 1989). PCR (see protocol above) was used to introduce a Phe to Tyr amino acid change at codon 667 of native Taq DNA polymerase (which is codon 396 of FY2). Oligonucleotide primer 1 dGCT-TGGGCAGAGGATCCGCCGGG (SEQ. ID. NO. 3) spans nucleotides 954 to 976 of the coding region of SEQ. ID. NO. 1 including a BamHI restriction site. Mutagenic oligo primer 2 dGGGATGGCTAGCTCCTGGGAGAGGCG-GTGGGCCGACATGCCGTAGA GGACCCCGTAGT-TGATGG (SEQ. ID. NO. 4) spans nucleotides 1178 to 1241 including an NheI site and codon 396 of Sequence ID. NO. 1. A clone of exo⁻ Taq deleted for the first 235 amino acids, pWB253 encoding DeltaTaq polymerase (Barnes, 112 *Gene* 29–35, 1992) was used as template DNA. Any clone of Taq polymerase or genomic DNA from *Thermus aquaticus* could also be utilized to amplify the identical PCR product. The PCR product was digested with BamHI and NheI, and this fragment was ligated to BamHI/NheI digested pWB253 plasmid to replace the corresponding fragment to create pWB253Y, encoding polymerase FY1. Cells of *E. coli* strain MV1190 were used for transformation and induction of protein expression, although any host strain carrying a lac repressor could be substituted. DNA sequencing verified the Phe to Tyr change in the coding region.

PCR primer 3 dGGAATTCCATATGGACGATCT-GAAGCTCTCC (SEQ. ID. NO. 5) spanning the start codon and containing restriction enzyme sites, was used with PCR primer 4 dGGGTACCAAGCTTCACTCCTTGGCG-GAGAG (SEQ. ID. NO. 6) containing restriction sites and spanning the stop codon (codon 562 of Sequence ID. NO. 1). A methionine start codon and restriction enzyme recognition sequences were added to PCR primer 5 dGGAATTC-CATATGCTGGAGAGGCTTGAGTTT (SEQ. ID. NO. 7), which was used with primer 4 above. PCR was performed using the above primer pairs, and plasmid pWB253Y as template. The PCR products were digested with restriction enzymes NdeI and KpnI and ligated to NdeI/KpnI digested vector pRE2 (Reddi et al., 17 *Nucleic Acids Research* 10,473–10,488, 1989) to make plasmids pRE236Y, encoding FY1 polymerase, and pRE273Y encoding FY2 polymerase, respectively. Cells of *E. coli* strain DH λ ⁺ were used for primary transformation with this and all subsequent pRE2 constructions, and strain M5248 (λ cI857) was used for protein expression, although any comparable pair of *E. coli* strains carrying the cI⁺ and cI857 alleles could be utilized. Alternatively, any rec⁺ cI⁺ strain could be induced by chemical agents such as nalidixic acid to produce the polymerase. The sequences of both genes were verified. pRE273Y was found to produce a single polypeptide band on SDS polyacrylamide gels, unlike pRE253Y or pRE236Y.

Primer 6 dGGAATTCCATATGCTGGAACGTCTG-GAGTTTGGCAGCCTC CTC (SEQ. ID. NO. 8) and primer 4 were used to make a PCR product introducing silent changes in codon usage of FY2. The product was digested with NdeI/BamHI and ligated to a pRE2 construct containing the 3' end of FY2 to create pREFY2pref, encoding FY2 DNA polymerase. Primer 7 dGGAATTCCATATG-GCTCTGGAACGTCTGGAGTTTGGCAGCCTCCTC (SEQ. ID. NO. 9) and primer 4 were used to make a PCR product introducing an additional alanine codon commonly occurring at the second position of highly expressed genes. The NdeI/BamHI digested fragment was used as above to create pREFY3, encoding FY3 DNA polymerase.

Preparation of FY4 DNA Polymerase

Bacterial Strains

E. coli strains: DH1 λ^+ [gyrA96, recA1, relA1, endA1, thi-1, hsdR17, supE44, λ^+]; M5248 [λ (bio275, cl857, cIII+, N+, Δ (H1))].

PCR

Genomic DNA was prepared by standard techniques from *Thermus thermophilus*. The DNA polymerase gene of *Thermus thermophilus* is known to reside on a 3 kilobase AlwNI fragment. To enrich for polymerase sequences in some PCR reactions, the genomic DNA was digested prior to PCR with AlwNI, and fragments of approximately 3 kb were selected by agarose gel electrophoresis to be used as template DNA. Reaction conditions were as follows: 10 mM Tris pH8.3, 50 mM KCl, 800 μ M dNTPs, 0.001% gelatin, 1.0 μ M each primer, 1.5 mM MgCl₂, 2.5 U Tth, 0.025 U DeepVent (New England Biolabs), per 100 μ l reaction. Cycling conditions were 94° C. 2 min, then 35 cycles of 94° C. 30 s, 55° C. 30 s, 72° C. 3 min, followed by 72° C. for 7 min.

In vitro mutagenesis

Restriction enzyme digestions, plasmid preparations, and other in vitro manipulations of DNA were performed using standard protocols (Sambrook et al., 1989). Plasmid pMR1 was constructed to encode an exonuclease-free polymerase, with optimized codons for expression in *E. coli* at the 5' end. Primer 8 (SEQ. ID. NO. 10) (GGAATTCCATATGCTGGAACGTCTGGAATTCGGCAGCCTC) was used with Primer 9 (SEQ. ID. NO.11) (GGGGTACCCTAA CCCTTGGCGGAAAGCCAGTC) to create a PCR product from Tth genomic DNA, which was digested with restriction enzymes NdeI and KpnI and inserted into plasmid pRE2 (Reddi et al., 1989, *Nucleic Acids Research* 17, 10473–10488) digested with the same enzymes.

To create the desired F396Y mutation, two PCR products were made from Tth chromosomal DNA. Primer 8 above was used in combination with Primer 10 (SEQ. ID. NO. 12) (GGGATGGCTAGCTCCTGGGAGAGCCTATGGGCG GACAT GCCGTAGAGGACGCCGTAGTTCACCG) to create a portion of the gene containing the F to Y amino acid change as well as a silent change to create an NheI restriction site. Primer 11 (SEQ. ID. NO. 13) (CTAGCTAGCCATCCCTACGAAGAAGCGGT GGCCT) was used in combination with primer 9 above to create a portion of the gene from the introduced NheI site to the stop codon at the 3' end of the coding sequence. The PCR product of Primers 8 and 10 was digested with NdeI and NheI, and the PCR product of Primers 9 and 11 was digested with NheI and KpnI. These were introduced into expression vector pRE2 which was digested with NdeI and KpnI to produce plasmid pMR5. In addition to the desired changes, pMR5 was found to have a spurious change introduced by PCR, which led to an amino acid substitution, K234R. Plasmid pMR8 was created to eliminate this substitution, by replacing the AflIII/BamHI fragment of pMR5 for the cor-

responding fragment from pMR1. The FY4 polymerase encoded by plasmid pMR8 (SEQ. ID. NO. 14) is given in FIG. 5.

Cells of *E. coli* strain DH1 λ^+ were used for primary transformation, and strain M5248 (λ cl857) was used for protein expression, although any comparable pair of *E. coli* strains carrying the cl λ^+ and cl857 alleles could be utilized. Alternatively, any rec λ^+ cl λ^+ strain could be induced by chemical agents such as nalidixic acid to produce the polymerase.

Protein Sequencing

Determinations of amino terminal protein sequences were performed at the W. M. Keck Foundation, Biotechnology Resource Laboratory, New Haven, Conn.

Purification of Polymerases

A 1 liter culture of 2 $\times$ LB (2% Bacto-Tryptone, 1% Bacto-Yeast Extract, 0.5% NaCl)+0.2% Casamino Acids+20 mM KPO₄ pH 7.5+50 μ g/ml Ampicillin was inoculated with a glycerol stock of the appropriate cell strain and grown at 30° C. with agitation until cells were in log phase (0.7–1.0 OD₅₉₀). 9 liters of 2 $\times$ LB+0.2% Casamino Acids +20 mM KPO₄ pH 7.5+0.05% Mazu Anti-foam was inoculated with 1 liter of log phase cells in 10 liter Microferm Fermentors (New Brunswick Scientific Co.). Cells were grown at 30° C. under 15 psi pressure, 350–450 rpm agitation, and an air flow rate of 14,000 cc/min $\pm$ 1000 cc/min. When the OD₅₉₀ reached 1.5–2.0, the cultures were induced by increasing the temperature to 40°–42° C. for 90–120 minutes. The cultures were then cooled to <20° C. and the cells harvested by centrifugation in a Sorvall RC-3B centrifuge at 5000 rpm at 4° C. for 15–20 minutes. Harvested cells were stored at –80° C.

Frozen cells were broken into small pieces and resuspended in pre-warmed (90°–95° C.) Lysis Buffer (20 mM Tris pH 8.5, 1 mM EDTA, 10 mM MgCl₂, 16 mM (NH₄)₂SO₄, 0.1% Tween 20, 0.1% Nonidet P-40, 1 mM PMSF). Resuspended cells were then heated rapidly to 80° C. and incubated at 80° C. for 20 minutes with constant stirring. The suspension was then rapidly cooled on ice. The cell debris was removed by centrifugation using a Sorvall GSA rotor at 10,000 rpm for 20 minutes at 4° C. The NaCl concentration of the supernatant was adjusted to 300 mM. The sample was then passed through a diethylaminoethyl cellulose (Whatman DE-52) column that had been previously equilibrated with Buffer A (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 300 mM NaCl, 10% glycerol, 1 mM DTT), and polymerase collected in the flow through. The sample was then diluted to a concentration of NaCl of 100 mM and applied to a Heparin-sepharose column. The polymerase was eluted from the column with a NaCl gradient (100–500 mM NaCl). The sample was then dialyzed against Buffer B (20 mM Tris pH 8.5, 1 mM EDTA, 0.1% Tween 20, 0.1% Nonidet P-40, 10 mM KCl, 10% glycerol, 1 mM DTT) and further diluted as needed to lower the conductivity of the sample to the conductivity of Buffer B. The sample was then applied to a diethylaminoethyl (Waters DEAE 15 HR) column and eluted with a 10–500 mM KCl gradient. The polymerase was then diluted with an equal volume of Final Buffer (20 mM Tris pH 8.5, 0.1 mM EDTA, 0.5% Tween 20, 0.5% Nonidet P-40, 100 mM KCl, 50% glycerol, 1 mM DTT) and dialyzed against Final Buffer.

Assay of Exonuclease Activity

The exonuclease assay was performed by incubating 5 μ l (25–150 units) of DNA polymerase with 5 μ g of labelled [³H]-pBR322 PCR fragment (1.6 $\times$ 10⁴ cpm/ μ g DNA) in 100 μ l of reaction buffer of 20 mM Tris.HCl pH 8.5, 5 mM

MgCl₂, 10 mM KCl, for 1 hour at 60° C. After this time interval, 200 μ l of 1:1 ratio of 50 μ g/ml salmon sperm DNA with 2 mM EDTA and 20% TCA with 2% sodium pyrophosphate were added into the assay aliquots. The aliquots were put on ice for 10 min and then centrifuged at 12,000 g for 10 min. Acid-soluble radioactivity in 200 μ l of the supernatant was quantitated by liquid scintillation counting. One unit of exonuclease activity was defined as the amount of enzyme that catalyzed the acid solubilization of 10 nmol of total nucleotide in 30 min at 60° C.

Utility in DNA Sequencing

EXAMPLE 1

DNA Sequencing with FY Polymerases (e.g., FY2 and FY3)

The following components were added to a microcentrifuge vial (0.5 ml): 0.4 pmol M13 DNA (e.g., M13 mp18, 1.0 μ g); 2 μ l Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM MgCl₂); 2 μ l of labeling nucleotide mixture (1.5 μ M each of dGTP, dCTP and dTTP); 0.5 μ l (5 μ Ci) of [α -³²P]dATP (about 2000 Ci/mmol); 1 μ l -40 primer (0.5 μ M; 0.5 pmol/ μ l 5'-GTTTCCCAGTCACGAC-3'); 2 μ l of a mixture containing 4 U/ μ l FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophosphatase (32 U/ μ l polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5 μ l. These components (the labeling reaction) were mixed and the vial was placed in a constant-temperature water bath at 45° C. for 5 minutes.

Four vials were labeled A, C, G, and T, and filled with 4 μ l of the corresponding termination mix: ddA termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddATP); ddT termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddTTP); ddC termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddCTP); ddG termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddGTP).

The labeling reaction was divided equally among the four termination vials (4 μ l to each termination reaction vial), and tightly capped.

The four vials were placed in a constant-temperature water bath at 72° C. for 5 minutes. Then 4 μ l of Stop Solution (95% Formamide 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added to each vial, and heated briefly to 70°-80° C. immediately prior to loading on a sequencing gel (8% acrylamide, 8.3M urea). Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using Taq DNA polymerase or Δ Taq DNA polymerase.

EXAMPLE 2

DNA Cycle Sequencing with FY Polymerases

The following components were added to a microcentrifuge vial (0.5 ml) which is suitable for insertion into a thermocycler machine (e.g., Perkin-Elmer DNA Thermal Cycler): 0.05 pmol or more M13 DNA (e.g., M13 mp18, 0.1 μ g), or 0.1 μ g double-stranded plasmid DNA (e.g., pUC19); 2 μ l Reaction Buffer (260 mM Tris-HCl, pH 9.5 65 mM

MgCl₂); 1 μ l 3.0 μ M dGTP; 1 μ l 3.0 μ M dTTP; 0.5 μ l (5 μ Ci) of [α -³²P]dATP (about 2000 Ci/mmol); 1 μ l -40 primer (0.5 μ M; 0.5 pmol/ μ l 5'-GTTTCCCAGTCACGAC-3'); 2 μ l of a mixture containing 4 U/ μ l FY polymerase and 6.6 U/ml *Thermoplasma acidophilum* inorganic pyrophosphatase (32 U/ μ l polymerase and 53 U/ml pyrophosphatase in 20 mM Tris (pH8.5), 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% NP-40, 0.5% TWEEN-20 and 50% glycerol, diluted 8 fold in dilution buffer (10 mM Tris-HCl pH8.0, 1 mM 2-mercaptoethanol, 0.5% TWEEN-20, 0.5% NP-40)); and water to a total volume of 17.5 μ l.

These components (labeling reaction mixture) were mixed and overlaid with 10 μ l light mineral oil (Amersham). The vial was placed in the thermocycler and 30-100 cycles (more than 60 cycles is unnecessary) from 45° C. for 1 minute to 95° C. for 0.5 minute performed. (Temperatures can be cycled from 55°-95° C., if desired) The temperatures may be adjusted if the melting temperature of the primer/template is significantly higher or lower, but these temperatures work well for most primer-templates combinations. This step can be completed in about 3 minutes per cycle.

Four vials were labeled A, C, G, and T, and filled with 4 μ l of the corresponding termination mix: ddA termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddATP); ddT termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddTTP); ddC termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddCTP); ddG termination mix (150 μ M each dATP, dCTP, dGTP, dTTP, 1.5 μ M ddGTP). No additional enzyme is added to the termination vials. The enzyme carried in from the prior (labeling) step is sufficient.

The cycled labeling reaction mixture was divided equally among the four termination vials (4 μ l to each termination reaction vial), and overlaid with 10 μ l of light mineral oil.

The four vials were placed in the thermocycler and 30-200 cycles (more than 60 cycles is unnecessary) performed from 95° C. for 15 seconds, 55° C. for 30 seconds, and 72° C. for 120 seconds. This step was conveniently completed overnight. Other times and temperatures are also effective.

Six μ l of reaction mixture was removed (avoiding oil), 3 μ l of Stop Solution (95% Formamide 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF) added, and heated briefly to 70°-80° C. immediately prior to loading on a sequencing gel. Autoradiograms required an 18-36 hour exposure using Kodak XAR-5 film or Amersham Hyperfilm MP. High-quality sequence results with uniform band intensities were obtained. The band intensities were much more uniform than those obtained with similar protocols using Taq DNA polymerase or Δ Taq DNA polymerase.

EXAMPLE 3

Sequencing with dGTP Analogs to Eliminate Compression Artifacts

For either of the sequencing methods outlined in examples 1 and 2, 7-Deaza-2'-deoxy-GTP can be substituted for dGTP in the labeling and termination mixtures at exactly the same concentration as dGTP. When this substitution is made, secondary structures on the gels are greatly reduced. Similarly, 2'-deoxyinosinetriphosphate can also be substituted for dGTP but its concentration must be 10-fold higher than the corresponding concentration of dGTP. Substitution of dTTP for dGTP is even more effective in eliminating compression artifacts than 7-deaza-dGTP.

EXAMPLE 4

Other Sequencing Methods Using FY Polymerases

FY polymerases have been adapted for use with many other sequencing methods, including the use of fluorescent

primers and fluorescent-dideoxy-terminators for sequencing with the ABI 373A DNA sequencing instrument.

EXAMPLE 5

SDS-Polyacrylamide Gel Electrophoresis

Protein samples were run on a 14×16 mm 7.5 or 10% polyacrylamide gel. (Gels were predominantly 10% Polyacrylamide using a 14×16 mm Hoefer apparatus. Other sizes, apparatuses, and percentage gels are acceptable. Similar results can also be obtained using the Pharmacia Phast Gel system with SDS, 8–25% gradient gels. Reagent grade and ultrapure grade reagents were used.) The stacking gel consisted of 4% acrylamide (30:0.8, acrylamide: bisacrylamide), 125 mM Tris-HCl pH 6.8, 0.1% Sodium Dodecyl Sulfate (SDS). The resolving gel consisted of 7.5 or 10% acrylamide (30:0.8, acrylamide: bisacrylamide), 375

mM Tris-HCl pH 8.8, 0.1% SDS. Running Buffer consisted of 25 mM Tris, 192 mM Glycine and 0.1% SDS. 1× Sample Buffer consisted of 25 mM Tris-HCl pH 6.8, 0.25% SDS, 10% Glycerol, 0.1M Dithiothreitol, 0.1% Bromophenol Blue, and 1 mM EDTA. A ¼ volume of 5× Sample Buffer was added to each sample. Samples were heated in sample buffer to 90°–100° C. for approximately 5 minutes prior to loading. A 1.5 mm thick gel was run at 50–100 mA constant current for 1–3 hours (until bromophenol blue was close to the bottom of the gel). The gel was stained with 0.025% Coomassie Blue R250 in 50% methanol, 10% acetic acid and destained in 5% methanol, 7% acetic acid solution. A record of the gel was made by taking a photograph of the gel, by drying the gel between cellulose film sheets, or by drying the gel onto filter paper under a vacuum.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i i i) NUMBER OF SEQUENCES: 14

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1686 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(i x) FEATURE:
(A) NAME/KEY: FY2
(B) LOCATION: 1...1683

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG	CTG	GAG	AGG	CTT	GAG	TTT	GGC	AGC	CTC	CTC	CAC	GAG	TTC	GGC	CTT	48
Met	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
1				5					10					15		
CTG	GAA	AGC	CCC	AAG	GCC	CTG	GAG	GAG	GCC	CCC	TGG	CCC	CCG	CCG	GAA	96
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
			20					25					30			
GGG	GCC	TTC	GTG	GGC	TTT	GTG	CTT	TCC	CGC	AAG	GAG	CCC	ATG	TGG	GCC	144
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
			35				40					45				
GAT	CTT	CTG	GCC	CTG	GCC	GCC	GCC	AGG	GGG	GGC	CGG	GTC	CAC	CGG	GCC	192
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
			50				55				60					
CCC	GAG	CCT	TAT	AAA	GCC	CTC	AGG	GAC	CTG	AAG	GAG	GCG	CGG	GGG	CTT	240
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
					70					75					80	
CTC	GCC	AAA	GAC	CTG	AGC	GTT	CTG	GCC	CTG	AGG	GAA	GGC	CTT	GGC	CTC	288
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
				85					90				95			
CCG	CCC	GGC	GAC	GAC	CCC	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	CCT	TCC	336
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
			100					105					110			
AAC	ACC	ACC	CCC	GAG	GGG	GTG	GCC	CGG	CGC	TAC	GGC	GGG	GAG	TGG	ACG	384
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
			115				120					125				
GAG	GAG	GCG	GGG	GAG	CGG	GCC	GCC	CTT	TCC	GAG	AGG	CTC	TTC	GCC	AAC	432
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
			130				135					140				

-continued

CTG Leu 145	TGG Trp	GGG Gly	AGG Arg	CTT Leu	GAG Glu 150	GGG Gly	GAG Glu	GAG Glu	AGG Arg	CTC Leu 155	CTT Leu	TGG Trp	CTT Leu	TAC Tyr	CGG Arg 160	480
GAG Glu	GTG Val	GAG Glu	AGG Arg	CCC Pro 165	CTT Leu	TCC Ser	GCT Ala	GTC Val	CTG Leu 170	GCC Ala	CAC His	ATG Met	GAG Glu	GCC Ala 175	ACG Thr	528
GGG Gly	GTG Val	CGC Arg	CTG Leu 180	GAC Asp	GTG Val	GCC Ala	TAT Tyr	CTC Leu 185	AGG Arg	GCC Ala	TTG Leu	TCC Ser	CTG Leu 190	GAG Glu	GTG Val	576
GCC Ala	GAG Glu 195	GAG Glu	ATC Ile	GCC Ala	CGC Arg	CTC Leu	GAG Glu 200	GCC Ala	GAG Glu	GTC Val	TTC Phe	CGC Arg 205	CTG Leu	GCC Ala	GGC Gly	624
CAC His 210	CCC Pro	TTC Phe	AAC Asn	CTC Leu	AAC Asn	TCC Ser 215	CGG Arg	GAC Asp	CAG Gln	CTG Leu	GAA Glu 220	AGG Arg	GTC Val	CTC Leu	TTT Phe	672
GAC Asp 225	GAG Glu	CTA Leu	GGG Gly	CTT Leu	CCC Pro 230	GCC Ala	ATC Ile	GGC Gly	AAG Lys	ACG Thr 235	GAG Glu	AAG Lys	ACC Thr	GGC Gly	AAG Lys 240	720
CGC Arg	TCC Ser	ACC Thr	AGC Ser	GCC Ala 245	GCC Ala	GTC Val	CTG Leu	GAG Glu 250	GCC Ala	CTC Leu	CGC Arg	GAG Glu	GCC Ala	CAC His 255	CCC Pro	768
ATC Ile	GTG Val	GAG Glu	AAG Lys 260	ATC Ile	CTG Leu	CAG Gln	TAC Tyr	CGG Arg 265	GAG Glu	CTC Leu	ACC Thr	AAG Lys	CTG Leu 270	AAG Lys	AGC Ser	816
ACC Thr	TAC Tyr	ATT Ile 275	GAC Asp	CCC Pro	TTG Leu	CCG Pro	GAC Asp 280	CTC Leu	ATC Ile	CAC His	CCC Pro	AGG Arg 285	ACG Thr	GGC Gly	CGC Arg	864
CTC Leu 290	CAC His	ACC Thr	CGC Arg	TTC Phe	AAC Asn	CAG Gln 295	ACG Thr	GCC Ala	ACG Thr	GCC Ala	ACG Thr 300	GGC Gly	AGG Arg	CTA Leu	AGT Ser	912
AGC Ser 305	TCC Ser	GAT Asp	CCC Pro	AAC Asn	CTC Leu 310	CAG Gln	AAC Asn	ATC Ile	CCC Pro	GTC Val 315	CGC Arg	ACC Thr	CCG Pro	CTT Leu	GGG Gly 320	960
CAG Gln	AGG Arg	ATC Ile	CGC Arg	CGG Arg 325	GCC Ala	TTC Phe	ATC Ile	GCC Ala	GAG Glu 330	GAG Glu	GGG Gly	TGG Trp	CTA Leu 335	TTG Leu	GTG Val	1008
GCC Ala	CTG Leu	GAC Asp	TAT Tyr 340	AGC Ser	CAG Gln	ATA Ile	GAG Glu 345	CTC Leu	AGG Arg	GTG Val	CTG Leu	GCC Ala	CAC His 350	CTC Leu	TCC Ser	1056
GGC Gly	GAC Asp 355	GAG Glu	AAC Asn	CTG Leu	ATC Ile	CGG Arg	GTC Val 360	TTC Phe	CAG Gln	GAG Glu	GGG Gly	CGG Arg 365	GAC Asp	ATC Ile	CAC His	1104
ACG Thr 370	GAG Glu	ACC Thr	GCC Ala	AGC Ser	TGG Trp	ATG Met 375	TTC Phe	GGC Gly	GTC Val	CCC Pro	CGG Arg 380	GAG Glu	GCC Ala	GTG Val	GAC Asp	1152
CCC Pro 385	CTG Leu	ATG Met	CGC Arg	CGG Arg	GCG Ala 390	GCC Ala	AAG Lys	ACC Thr	ATC Ile	AAC Asn 395	TAC Tyr	GGG Gly	GTC Val	CTC Leu	TAC Tyr 400	1200
GGC Gly	ATG Met	TCG Ser	GCC Ala	CAC His 405	CGC Arg	CTC Leu	TCC Ser	CAG Gln	GAG Glu 410	CTA Leu	GCC Ala	ATC Ile	CCT Pro	TAC Tyr 415	GAG Glu	1248
GAG Glu	GCC Ala	CAG Gln	GCC Ala 420	TTC Phe	ATT Ile	GAG Glu	CGC Arg	TAC Tyr 425	TTT Phe	CAG Gln	AGC Ser	TTC Phe	CCC Pro 430	AAG Lys	GTG Val	1296
CGG Arg	GCC Ala	TGG Trp 435	ATT Ile	GAG Glu	AAG Lys	ACC Thr	CTG Leu 440	GAG Glu	GAG Glu	GGC Gly	AGG Arg	AGG Arg 445	CGG Arg	GGG Gly	TAC Tyr	1344
GTG Val 450	GAG Glu	ACC Thr	CTC Leu	TTC Phe	GGC Gly	CGC Arg 455	CGC Arg	CGC Arg	TAC Tyr	GTG Val	CCA Pro 460	GAC Asp	CTA Leu	GAG Glu	GCC Ala	1392

-continued

CGG Arg 465	GTG Val	AAG Lys	AGC Ser	GTG Val	CGG Arg 470	GAG Glu	GCG Ala	GCC Ala	GAG Glu	CGC Arg 475	ATG Met	GCC Ala	TTC Phe	AAC Asn	ATG Met 480	1440
CCC Pro	GTC Val	CAG Gln	GGC Gly	ACC Thr 485	GCC Ala	GCC Ala	GAC Asp	CTC Leu	ATG Met 490	AAG Lys	CTG Leu	GCT Ala	ATG Met	GTG Val 495	AAG Lys	1488
CTC Leu	TTC Phe	CCC Pro	AGG Arg 500	CTG Leu	GAG Glu	GAA Glu	ATG Met	GGG Gly 505	GCC Ala	AGG Arg	ATG Met	CTC Leu	CTT Leu 510	CAG Gln	GTC Val	1536
CAC His	GAC Asp	GAG Glu 515	CTG Leu	GTC Val	CTC Leu	GAG Glu	GCC Ala 520	CCA Pro	AAA Lys	GAG Glu	AGG Arg	GCG Ala 525	GAG Glu	GCC Ala	GTG Val	1584
GCC Ala	CGG Arg 530	CTG Leu	GCC Ala	AAG Lys	GAG Glu	GTC Val 535	ATG Met	GAG Glu	GGG Gly	GTG Val	TAT Tyr 540	CCC Pro	CTG Leu	GCC Ala	GTG Val	1632
CCC Pro 545	CTG Leu	GAG Glu	GTG Val	GAG Glu	GTG Val 550	GGG Gly	ATA Ile	GGG Gly	GAG Glu	GAC Asp 555	TGG Trp	CTC Leu	TCC Ser	GCC Ala	AAG Lys 560	1680
GAG Glu	TGA *															1686

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1689 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(i x) FEATURE:

(A) NAME/KEY: FY3

(B) LOCATION: 1...1686

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ATG Met 1	GCT Ala	CTG Leu	GAA Glu	CGT Arg 5	CTG Leu	GAG Glu	TTT Phe	GGC Gly 10	AGC Ser 10	CTC Leu	CTC Leu	CAC His	GAG Glu 15	TTT Phe 15	GGC Gly	48
CTT Leu	CTG Leu	GAA Glu 20	AGC Ser	CCC Pro	AAG Lys	GCC Ala	CTG Leu	GAG Glu 25	GAG Glu	GCC Ala	CCC Pro	TGG Trp	CCC Pro 30	CCG Pro	CCG Pro	96
GAA Glu 35	GGG Gly	GCC Ala	TTC Phe	GTG Val	GGC Gly	TTT Phe	GTG Val 40	CTT Leu	TCC Ser	CGC Arg	AAG Lys	GAG Glu 45	CCC Pro	ATG Met	TGG Trp	144
GCC Ala 50	GAT Asp	CTT Leu	CTG Leu	GCC Ala	CTG Leu	GCC Ala 55	GCC Ala	GCC Ala	AGG Arg	GGG Gly 60	GGC Gly	CGG Arg	GTC Val	CAC His	CGG Arg	192
GCC Ala 65	CCC Pro	GAG Glu	CCT Pro	TAT Tyr	AAA Lys 70	GCC Ala	CTC Leu	AGG Arg	GAC Asp	CTG Leu 75	AAG Lys	GAG Glu	GCG Ala	CGG Arg	GGG Gly 80	240
CTT Leu	CTC Leu	GCC Ala	AAA Lys 85	GAC Asp	CTG Leu	AGC Ser	GTT Val	CTG Leu 90	GCC Ala	CTG Leu	AGG Arg	GAA Glu	GGC Gly 95	CTT Leu 95	GGC Gly	288
CTC Leu	CCG Pro	CCC Pro	GGC Gly 100	GAC Asp	GAC Asp	CCC Pro	ATG Met	CTC Leu 105	CTC Leu	GCC Ala	TAC Tyr	CTC Leu	CTG Leu 110	GAC Asp	CCT Pro	336
TCC Ser	AAC Asn 115	ACC Thr	ACC Thr	CCC Pro	GAG Glu	GGG Gly	GTG Val 120	GCC Ala	CGG Arg	CGC Arg	TAC Tyr	GGC Gly 125	GGG Gly	GAG Glu	TGG Trp	384
ACG Thr 130	GAG Glu	GAG Glu	GCG Ala	GGG Gly	GAG Glu	CGG Arg 135	GCC Ala	GCC Ala	CTT Leu	TCC Ser	GAG Glu 140	AGG Arg	CTC Leu	TTC Phe	GCC Ala	432
AAC	CTG	TGG	GGG	AGG	CTT	GAG	GGG	GAG	GAG	AGG	CTC	CTT	TGG	CTT	TAC	480

5,885,813

17

18

-continued

Asn 145	Leu	Trp	Gly	Arg	Leu 150	Glu	Gly	Glu	Glu	Arg 155	Leu	Leu	Trp	Leu	Tyr 160	
CGG Arg	GAG Glu	GTG Val	GAG Glu	AGG Arg 165	CCC Pro	CTT Leu	TCC Ser	GCT Ala	GTC Val 170	CTG Leu	GCC Ala	CAC His	ATG Met	GAG Glu 175	GCC Ala	528
ACG Thr	GGG Gly	GTG Val	CGC Arg 180	CTG Leu	GAC Asp	GTG Val	GCC Ala	TAT Tyr 185	CTC Leu	AGG Arg	GCC Ala	TTG Leu	TCC Ser 190	CTG Leu	GAG Glu	576
GTG Val	GCC Ala	GAG Glu 195	GAG Glu	ATC Ile	GCC Ala	CGC Arg	CTC Leu 200	GAG Glu	GCC Ala	GAG Glu	GTC Val	TTC Phe 205	CGC Arg	CTG Leu	GCC Ala	624
GGC Gly	CAC His 210	CCC Pro	TTC Phe	AAC Asn	CTC Leu	AAC Asn 215	TCC Ser	CGG Arg	GAC Asp	CAG Gln 220	CTG Leu	GAA Glu	AGG Arg	GTC Val	CTC Leu	672
TTT Phe 225	GAC Asp	GAG Glu	CTA Leu	GGG Gly	CTT Leu 230	CCC Pro	GCC Ala	ATC Ile	GGC Gly	AAG Lys 235	ACG Thr	GAG Glu	AAG Lys	ACC Thr	GGC Gly 240	720
AAG Lys	CGC Arg	TCC Ser	ACC Thr	AGC Ser 245	GCC Ala	GCC Ala	GTC Val	CTG Leu	GAG Glu 250	GCC Ala	CTC Leu	CGC Arg	GAG Glu	GCC Ala 255	CAC His	768
CCC Pro	ATC Ile	GTG Val	GAG Glu 260	AAG Lys	ATC Ile	CTG Leu	CAG Gln	TAC Tyr 265	CGG Arg	GAG Glu	CTC Leu	ACC Thr	AAG Lys 270	CTG Leu	AAG Lys	816
AGC Ser	ACC Thr	TAC Tyr 275	ATT Ile	GAC Asp	CCC Pro	TTG Leu	CCG Pro 280	GAC Asp	CTC Leu	ATC Ile	CAC His	CCC Pro 285	AGG Arg	ACG Thr	GGC Gly	864
CGC Arg	CTC Leu 290	CAC His	ACC Thr	CGC Arg	TTC Phe	AAC Asn 295	CAG Gln	ACG Thr	GCC Ala	ACG Thr	GCC Ala 300	ACG Thr	GGC Gly	AGG Arg	CTA Leu	912
AGT Ser 305	AGC Ser	TCC Ser	GAT Asp	CCC Pro	AAC Asn 310	CTC Leu	CAG Gln	AAC Asn	ATC Ile	CCC Pro 315	GTC Val	CGC Arg	ACC Thr	CCG Pro	CTT Leu 320	960
GGG Gly	CAG Gln	AGG Arg	ATC Ile	CGC Arg 325	CGG Arg	GCC Ala	TTC Phe	ATC Ile	GCC Ala 330	GAG Glu	GAG Glu	GGG Gly	TGG Trp	CTA Leu 335	TTG Leu	1008
GTG Val	GCC Ala	CTG Leu	GAC Asp 340	TAT Tyr	AGC Ser	CAG Gln	ATA Ile	GAG Glu 345	CTC Leu	AGG Arg	GTG Val	CTG Leu	GCC Ala 350	CAC His	CTC Leu	1056
TCC Ser	GGC Gly	GAC Asp 355	GAG Glu	AAC Asn	CTG Leu	ATC Ile	CGG Arg 360	GTC Val	TTC Phe	CAG Gln	GAG Glu	GGG Gly 365	CGG Arg	GAC Asp	ATC Ile	1104
CAC His	ACG Thr 370	GAG Glu	ACC Thr	GCC Ala	AGC Ser	TGG Trp 375	ATG Met	TTC Phe	GGC Gly	GTC Val	CCC Pro 380	CGG Arg	GAG Glu	GCC Ala	GTG Val	1152
GAC Asp 385	CCC Pro	CTG Leu	ATG Met	CGC Arg	CGG Arg 390	GCG Ala	GCC Ala	AAG Lys	ACC Thr	ATC Ile 395	AAC Asn	TAC Tyr	GGG Gly	GTC Val	CTC Leu 400	1200
TAC Tyr	GGC Gly	ATG Met	TCG Ser	GCC Ala 405	CAC His	CGC Arg	CTC Leu	TCC Ser	CAG Gln 410	GAG Glu	CTA Leu	GCC Ala	ATC Ile	CCT Pro 415	TAC Tyr	1248
GAG Glu	GAG Glu	GCC Ala	CAG Gln 420	GCC Ala	TTC Phe	ATT Ile	GAG Glu	CGC Arg 425	TAC Tyr	TTT Phe	CAG Gln	AGC Ser	TTC Phe 430	CCC Pro	AAG Lys	1296
GTG Val	CGG Arg	GCC Ala 435	TGG Trp	ATT Ile	GAG Glu	AAG Lys	ACC Thr 440	CTG Leu	GAG Glu	GAG Glu	GGC Gly	AGG Arg 445	AGG Arg	CGG Arg	GGG Gly	1344
TAC Tyr	GTG Val 450	GAG Glu	ACC Thr	CTC Leu	TTC Phe	GGC Gly 455	CGC Arg	CGC Arg	CGC Arg	TAC Tyr	GTG Val 460	CCA Pro	GAC Asp	CTA Leu	GAG Glu	1392
GCC	CGG	GTG	AAG	AGC	GTG	CGG	GAG	GCG	GCC	GAG	CGC	ATG	GCC	TTC	AAC	1440

-continued

Ala 465	Arg	Val	Lys	Ser	Val 470	Arg	Glu	Ala	Ala	Glu 475	Arg	Met	Ala	Phe	Asn 480		
ATG	CCC	GTC	CAG	GGC	ACC	GCC	GCC	GAC	CTC	ATG	AAG	CTG	GCT	ATG	GTG	1488	
Met	Pro	Val	Gln	Gly 485	Thr	Ala	Ala	Asp	Leu 490	Met	Lys	Leu	Ala	Met	Val 495		
AAG	CTC	TTC	CCC	AGG	CTG	GAG	GAA	ATG	GGG	GCC	AGG	ATG	CTC	CTT	CAG	1536	
Lys	Leu	Phe	Pro 500	Arg	Leu	Glu	Glu	Met 505	Gly	Ala	Arg	Met	Leu 510	Leu	Gln		
GTC	CAC	GAC	GAG	CTG	GTC	CTC	GAG	GCC	CCA	AAA	GAG	AGG	GCG	GAG	GCC	1584	
Val	His	Asp 515	Glu	Leu	Val	Leu	Glu 520	Ala	Pro	Lys	Glu	Arg 525	Ala	Glu	Ala		
GTG	GCC	CGG	CTG	GCC	AAG	GAG	GTC	ATG	GAG	GGG	GTG	TAT	CCC	CTG	GCC	1632	
Val	Ala 530	Arg	Leu	Ala	Lys	Glu 535	Val	Met	Glu	Gly	Val 540	Tyr	Pro	Leu	Ala		
GTG	CCC	CTG	GAG	GTG	GAG	GTG	GGG	ATA	GGG	GAG	GAC	TGG	CTC	TCC	GCC	1680	
Val 545	Pro	Leu	Glu	Val	Glu 550	Val	Gly	Ile	Gly	Glu 555	Asp	Trp	Leu	Ser	Ala 560		
AAG	GAG	TGA														1689	
Lys	Glu	*															

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCTTGGGCAG AGGATCCGCC GGG 23

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 64 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGGATGGCTA GCTCCTGGGA GAGGCGGTGG GCCGACATGC CGTAGAGGAC 50
CCCGTAGTTG ATGG 64

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAATTCCAT ATGGACGATC TGAAGCTCTC C 31

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GGGGTACCAA GCTTCACTCC TTGGCGGAGA G 31

(2) INFORMATION FOR SEQ ID NO: 7:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 31 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 7:	
GGAATTCCAT ATGCTGGAGA GGCTTGAGTT T	3 1
(2) INFORMATION FOR SEQ ID NO: 8:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 43 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 8:	
GGAATTCCAT ATGCTGGAAC GTCTGGAGTT TGGCAGCCTC CTC	4 3
(2) INFORMATION FOR SEQ ID NO: 9:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 46 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 9:	
GGAATTCCAT ATGGCTCTGG AACGTCTGGA GTTTGGCAGC CTCCTC	4 6
(2) INFORMATION FOR SEQ ID NO: 10:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 40 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 10:	
GGAATTCCAT ATGCTGGAAC GTCTGGAATT CGGCAGCCTC	4 0
(2) INFORMATION FOR SEQ ID NO: 11:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 32 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 11:	
GGGGTACCCT AACCCTTGGC GGAAAGCCAG TC	3 2
(2) INFORMATION FOR SEQ ID NO: 12:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 64 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 12:	
GGGATGGCTA GCTCCTGGGA GAGCCTATGG GCGGACATGC CGTAGAGGAC	5 0
GCCGTAGTTC ACCG	6 4

-continued

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

CTAGCTAGCC ATCCCCTACG AAGAAGCGGT GGCCT 3 5

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1686 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (i x) FEATURE:
 - (A) NAME/KEY: FY4
 - (B) LOCATION: 1...1683

(x i) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATG	CTG	GAA	CGT	CTG	GAA	TTC	GGC	AGC	CTC	CTC	CAC	GAG	TTC	GGC	CTC	4 8
Met	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
1				5					10					15		
CTG	GAG	GCC	CCC	GCC	CCC	CTG	GAG	GAG	GCC	CCC	TGG	CCC	CCG	CCG	GAA	9 6
Leu	Glu	Ala	Pro	Ala	Pro	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
			20					25					30			
GGG	GCC	TTC	GTG	GGC	TTC	GTC	CTC	TCC	CGC	CCC	GAG	CCC	ATG	TGG	GCG	1 4 4
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Pro	Glu	Pro	Met	Trp	Ala	
		35					40					45				
GAG	CTT	AAA	GCC	CTG	GCC	GCC	TGC	AGG	GAC	GGC	CGG	GTG	CAC	CGG	GCA	1 9 2
Glu	Leu	Lys	Ala	Leu	Ala	Ala	Cys	Arg	Asp	Gly	Arg	Val	His	Arg	Ala	
	50					55					60					
GCA	GAC	CCC	TTG	GCG	GGG	CTA	AAG	GAC	CTC	AAG	GAG	GTC	CGG	GGC	CTC	2 4 0
Ala	Asp	Pro	Leu	Ala	Gly	Leu	Lys	Asp	Leu	Lys	Glu	Val	Arg	Gly	Leu	
	65				70					75				80		
CTC	GCC	AAG	GAC	CTC	GCC	GTC	TTG	GCC	TCG	AGG	GAG	GGG	CTA	GAC	CTC	2 8 8
Leu	Ala	Lys	Asp	Leu	Ala	Val	Leu	Ala	Ser	Arg	Glu	Gly	Leu	Asp	Leu	
			85						90					95		
GTG	CCC	GGG	GAC	GAC	CCC	ATG	CTC	CTC	GCC	TAC	CTC	CTG	GAC	CCC	TCC	3 3 6
Val	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
			100					105					110			
AAC	ACC	ACC	CCC	GAG	GGG	GTG	GCG	CGG	CGC	TAC	GGG	GGG	GAG	TGG	ACG	3 8 4
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
		115					120					125				
GAG	GAC	GCC	GCC	CAC	CGG	GCC	CTC	CTC	TCG	GAG	AGG	CTC	CAT	CGG	AAC	4 3 2
Glu	Asp	Ala	Ala	His	Arg	Ala	Leu	Leu	Ser	Glu	Arg	Leu	His	Arg	Asn	
	130					135					140					
CTC	CTT	AAG	CGC	CTC	GAG	GGG	GAG	GAG	AAG	CTC	CTT	TGG	CTC	TAC	CAC	4 8 0
Leu	Leu	Lys	Arg	Leu	Glu	Gly	Glu	Glu	Lys	Leu	Leu	Trp	Leu	Tyr	His	
	145				150					155					160	
GAG	GTG	GAA	AAG	CCC	CTC	TCC	CGG	GTC	CTG	GCC	CAC	ATG	GAG	GCC	ACC	5 2 8
Glu	Val	Glu	Lys	Pro	Leu	Ser	Arg	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
				165					170					175		
GGG	GTA	CGG	CTG	GAC	GTG	GCC	TAC	CTT	CAG	GCC	CTT	TCC	CTG	GAG	CTT	5 7 6
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Gln	Ala	Leu	Ser	Leu	Glu	Leu	
			180					185					190			
GCG	GAG	GAG	ATC	CGC	CGC	CTC	GAG	GAG	GAG	GTC	TTC	CGC	TTG	GCG	GGC	6 2 4
Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	Gly	
			195				200					205				

-continued

CAC His 225	CCC Pro 210	TTC Phe	AAC Asn	CTC Leu	AAC Asn	TCC Ser 215	CGG Arg	GAC Asp	CAG Gln	CTG Leu	GAA Glu 220	AGG Arg	GTG Val	CTC Leu	TTT Phe	672
GAC Asp 225	GAG Glu	CTT Leu	AGG Arg	CTT Leu	CCC Pro 230	GCC Ala	TTG Leu	GGG Gly	AAG Lys	ACG Thr 235	CAA Gln	AAG Lys	ACA Thr	GGC Gly	AAG Lys 240	720
CGC Arg	TCC Ser	ACC Thr	AGC Ser	GCC Ala 245	GCG Ala	GTG Val	CTG Leu	GAG Glu	GCC Ala 250	CTA Leu	CGG Arg	GAG Glu	GCC Ala	CAC His 255	CCC Pro	768
ATC Ile	GTG Val	GAG Glu	AAG Lys 260	ATC Ile	CTC Leu	CAG Gln	CAC His	CGG Arg 265	GAG Glu	CTC Leu	ACC Thr	AAG Lys	CTC Leu 270	AAG Lys	AAC Asn	816
ACC Thr	TAC Tyr	GTG Val 275	GAC Asp	CCC Pro	CTC Leu	CCA Pro	AGC Ser 280	CTC Leu	GTC Val	CAC His	CCG Pro	AGG Arg 285	ACG Thr	GGC Gly	CGC Arg	864
CTC Leu	CAC His 290	ACC Thr	CGC Arg	TTC Phe	AAC Asn	CAG Gln 295	ACG Thr	GCC Ala	ACG Thr	GCC Ala	ACG Thr 300	GGG Gly	AGG Arg	CTT Leu	AGT Ser	912
AGC Ser 305	TCC Ser	GAC Asp	CCC Pro	AAC Asn	CTG Leu 310	CAG Gln	AAC Asn	ATC Ile	CCC Pro	GTC Val 315	CGC Arg	ACC Thr	CCC Pro	TTG Leu	GGC Gly 320	960
CAG Gln	AGG Arg	ATC Ile	CGC Arg	CGG Arg 325	GCC Ala	TTC Phe	GTG Val	GCC Ala	GAG Glu 330	GCG Ala	GGT Gly	TGG Trp	GCG Ala	TTG Leu 335	GTG Val	1008
GCC Ala	CTG Leu	GAC Asp	TAT Tyr 340	AGC Ser	CAG Gln	ATA Ile	GAG Glu	CTC Leu 345	CGC Arg	GTC Val	CTC Leu	GCC Ala	CAC His 350	CTC Leu	TCC Ser	1056
GGG Gly	GAC Asp	GAA Glu 355	AAC Asn	CTG Leu	ATC Ile	AGG Arg	GTC Val 360	TTC Phe	CAG Gln	GAG Glu	GGG Gly	AAG Lys 365	GAC Asp	ATC Ile	CAC His	1104
ACC Thr	CAG Gln 370	ACC Thr	GCA Ala	AGC Ser	TGG Trp	ATG Met 375	TTC Phe	GGC Gly	GTC Val	CCC Pro	CCG Pro 380	GAG Glu	GCC Ala	GTG Val	GAC Asp	1152
CCC Pro 385	CTG Leu	ATG Met	CGC Arg	CGG Arg	GCG Ala 390	GCC Ala	AAG Lys	ACG Thr	GTG Val	AAC Asn 395	TAC Tyr	GGC Gly	GTC Val	CTC Leu	TAC Tyr 400	1200
GGC Gly	ATG Met	TCC Ser	GCC Ala	CAT His 405	AGG Arg	CTC Leu	TCC Ser	CAG Gln	GAG Glu 410	CTA Leu	GCC Ala	ATC Ile	CCC Pro	TAC Tyr 415	GAA Glu	1248
GAA Glu	GCG Ala	GTG Val	GCC Ala 420	TTT Phe	ATA Ile	GAG Glu	CGC Arg	TAC Tyr 425	TTC Phe	CAA Gln	AGC Ser	TTC Phe	CCC Pro 430	AAG Lys	GTG Val	1296
CGG Arg	GCC Ala	TGG Trp 435	ATA Ile	GAA Glu	AAG Lys	ACC Thr	CTG Leu 440	GAG Glu	GAG Glu	GGG Gly	AGG Arg	AAG Lys 445	CGG Arg	GGC Gly	TAC Tyr	1344
GTG Val 450	GAA Glu	ACC Thr	CTC Leu	TTC Phe	GGA Gly	AGA Arg 455	AGG Arg	CGC Arg	TAC Tyr	GTG Val	CCC Pro 460	GAC Asp	CTC Leu	AAC Asn	GCC Ala	1392
CGG Arg 465	GTG Val	AAG Lys	AGC Ser	GTC Val	AGG Arg 470	GAG Glu	GCC Ala	GCG Ala	GAG Glu	CGC Arg 475	ATG Met	GCC Ala	TTC Phe	AAC Asn 480	ATG Met	1440
CCC Pro	GTC Val	CAG Gln	GGC Gly	ACC Thr 485	GCC Ala	GCC Ala	GAC Asp	CTC Leu	ATG Met 490	AAG Lys	CTC Leu	GCC Ala	ATG Met	GTG Val 495	AAG Lys	1488
CTC Leu	TTC Phe	CCC Pro	CGC Arg 500	CTC Leu	CGG Arg	GAG Glu	ATG Met	GGG Gly 505	GCC Ala	CGC Arg	ATG Met	CTC Leu	CTC Leu 510	CAG Gln	GTC Val	1536
CAC His	GAC Asp	GAG Glu 515	CTC Leu	CTC Leu	CTG Leu	GAG Glu	GCC Ala 520	CCC Pro	CAA Gln	GCG Ala	CGG Arg	GCC Ala 525	GAG Glu	GAG Glu	GTG Val	1584

-continued

GCG	GCT	TTG	GCC	AAG	GAG	GCC	ATG	GAG	AAG	GCC	TAT	CCC	CTC	GCC	GTG	1632
Ala	Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala	Val	
530						535					540					
CCC	CTG	GAG	GTG	GAG	GTG	GGG	ATG	GGG	GAG	GAC	TGG	CTT	TCC	GCC	AAG	1680
Pro	Leu	Glu	Val	Glu	Val	Gly	Met	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys	
545					550					555					560	
GGT	TAG															1686
Gly	*															

We claim:

1. An enzymatically active DNA polymerase comprising 561 amino acids having a tyrosine residue at an amino acid position corresponding to Taq DNA polymerase residue 667 in its dNMP binding site, wherein said polymerase lacks 5' to 3' exonuclease activity as a result of an N-terminal deletion, and wherein said polymerase is substantially identical to the DNA polymerase of *Thermus aquaticus*, *Thermus flavus* or *Thermus thermophilus*, and wherein said polymerase forms a single polypeptide band on an SDS polyacrylamide gel.

2. An enzymatically active DNA polymerase comprising 562 amino acids having a tyrosine residue at an amino acid position corresponding to Taq DNA polymerase residue 667 in its dNMP binding site, wherein said polymerase lacks 5' to 3' exonuclease activity as a result of an N-terminal deletion, and wherein said polymerase is substantially identical to the DNA polymerase of *Thermus aquaticus*, *Thermus flavus* or *Thermus thermophilus*, and wherein said polymerase forms a single polypeptide band on an SDS polyacrylamide gel.

3. The polymerase of claim 2, wherein the polymerase is FY3 (SEQ ID NO:2).

4. The polymerase of claim 1, wherein the polymerase is FY4 (SEQ ID NO:14).

5. The polymerase of claim 1 or 2 wherein the amino acid sequence of said polymerase includes less than 3 conservative amino acid changes compared to the equivalent region of one said DNA polymerase of said named *Thermus* species.

6. The polymerase of claim 1 or 2 wherein the amino acid sequence of said polymerase includes less than 3 additional amino acids at its N-terminus compared to the equivalent region of one said DNA polymerase of said named *Thermus* species.

7. The polymerase of claim 1 wherein the polymerase is FY2 (SEQ ID NO:1).

8. Purified nucleic acid encoding the DNA polymerase of any of claims 1, 7, 2, 3, or 4.

9. Method for sequencing DNA comprising the step of generating chain terminated fragments from the DNA template to be sequenced with a DNA polymerase of any of claims 1, 7, 2, 3, or 4 in the presence of at least one chain

terminating agent and one or more nucleotide triphosphates, and determining the sequence of said DNA from the sizes of said fragments.

10. Kit for sequencing DNA comprising a DNA polymerase of any of claims 1, 7, 2, 3, or 4 and a pyrophosphatase.

11. The kit of claim 10 wherein said pyrophosphatase is thermostable.

12. Apparatus for DNA sequencing having a reactor comprising a DNA polymerase of any of claims 1, 7, 2, 3, or 4 and a band separator.

13. Purified nucleic acid encoding the DNA polymerase of claim 5.

14. Method for sequencing DNA comprising the step of generating chain terminated fragments from the DNA template to be sequenced with a DNA polymerase of claim 5 in the presence of at least one chain terminating agent and one or more nucleotide triphosphates, and determining the sequence of said DNA from the sizes of said fragments.

15. Kit for sequencing DNA comprising a DNA polymerase of claim 5 and a pyrophosphatase.

16. The kit of claim 15 wherein said pyrophosphatase is thermostable.

17. Apparatus for DNA sequencing having a reactor comprising a DNA polymerase of claim 5 and a band separator.

18. Purified nucleic acid encoding the DNA polymerase of claim 6.

19. Method for sequencing DNA comprising the step of generating chain terminated fragments from the DNA template to be sequenced with a DNA polymerase of claim 6 in the presence of at least one chain terminating agent and one or more nucleotide triphosphates, and determining the sequence of said DNA from the sizes of said fragments.

20. Kit for sequencing DNA comprising a DNA polymerase of claim 6 and a pyrophosphatase.

21. The kit of claim 20 wherein said pyrophosphatase is thermostable.

22. Apparatus for DNA sequencing having a reactor comprising a DNA polymerase of claim 6 and a band separator.

* * * * *