

SEQUENCE LISTING

<110> ALBIREO AB

<120> BENZOTHIADIAZEPINE COMPOUNDS AND THEIR USE AS BILE ACID MODULATORS

<130> NP0461W0

<150> IN 201911004690

<151> 2019-02-06

<150> SE 1950464-6

<151> 2019-04-12

<150> IN 201911049981

<151> 2019-12-04

<160> 4

<170> PatentIn version 3.5

<210> 1

<211> 1251

<212> PRT

<213> Homo sapiens

<400> 1

Met Ser Thr Glu Arg Asp Ser Glu Thr Thr Phe Asp Glu Asp Ser Gln
1 5 10 15

Pro Asn Asp Glu Val Val Pro Tyr Ser Asp Asp Glu Thr Glu Asp Glu
20 25 30

Leu Asp Asp Gln Gly Ser Ala Val Glu Pro Glu Gln Asn Arg Val Asn
35 40 45

Arg Glu Ala Glu Glu Asn Arg Glu Pro Phe Arg Lys Glu Cys Thr Trp
50 55 60

Gln Val Lys Ala Asn Asp Arg Lys Tyr His Glu Gln Pro His Phe Met
65 70 75 80

Asn Thr Lys Phe Leu Cys Ile Lys Glu Ser Lys Tyr Ala Asn Asn Ala

85

90

95

Ile Lys Thr Tyr Lys Tyr Asn Ala Phe Thr Phe Ile Pro Met Asn Leu
 100 105 110

Phe Glu Gln Phe Lys Arg Ala Ala Asn Leu Tyr Phe Leu Ala Leu Leu
 115 120 125

Ile Leu Gln Ala Val Pro Gln Ile Ser Thr Leu Ala Trp Tyr Thr Thr
 130 135 140

Leu Val Pro Leu Leu Val Val Leu Gly Val Thr Ala Ile Lys Asp Leu
 145 150 155 160

Val Asp Asp Val Ala Arg His Lys Met Asp Lys Glu Ile Asn Asn Arg
 165 170 175

Thr Cys Glu Val Ile Lys Asp Gly Arg Phe Lys Val Ala Lys Trp Lys
 180 185 190

Glu Ile Gln Val Gly Asp Val Ile Arg Leu Lys Lys Asn Asp Phe Val
 195 200 205

Pro Ala Asp Ile Leu Leu Leu Ser Ser Ser Glu Pro Asn Ser Leu Cys
 210 215 220

Tyr Val Glu Thr Ala Glu Leu Asp Gly Glu Thr Asn Leu Lys Phe Lys
 225 230 235 240

Met Ser Leu Glu Ile Thr Asp Gln Tyr Leu Gln Arg Glu Asp Thr Leu
 245 250 255

Ala Thr Phe Asp Gly Phe Ile Glu Cys Glu Glu Pro Asn Asn Arg Leu
 260 265 270

Asp Lys Phe Thr Gly Thr Leu Phe Trp Arg Asn Thr Ser Phe Pro Leu
 275 280 285

Asp Ala Asp Lys Ile Leu Leu Arg Gly Cys Val Ile Arg Asn Thr Asp
290 295 300

Phe Cys His Gly Leu Val Ile Phe Ala Gly Ala Asp Thr Lys Ile Met
305 310 315 320

Lys Asn Ser Gly Lys Thr Arg Phe Lys Arg Thr Lys Ile Asp Tyr Leu
325 330 335

Met Asn Tyr Met Val Tyr Thr Ile Phe Val Val Leu Ile Leu Leu Ser
340 345 350

Ala Gly Leu Ala Ile Gly His Ala Tyr Trp Glu Ala Gln Val Gly Asn
355 360 365

Ser Ser Trp Tyr Leu Tyr Asp Gly Glu Asp Asp Thr Pro Ser Tyr Arg
370 375 380

Gly Phe Leu Ile Phe Trp Gly Tyr Ile Ile Val Leu Asn Thr Met Val
385 390 395 400

Pro Ile Ser Leu Tyr Val Ser Val Glu Val Ile Arg Leu Gly Gln Ser
405 410 415

His Phe Ile Asn Trp Asp Leu Gln Met Tyr Tyr Ala Glu Lys Asp Thr
420 425 430

Pro Ala Lys Ala Arg Thr Thr Thr Leu Asn Glu Gln Leu Gly Gln Ile
435 440 445

His Tyr Ile Phe Ser Asp Lys Thr Gly Thr Leu Thr Gln Asn Ile Met
450 455 460

Thr Phe Lys Lys Cys Cys Ile Asn Gly Gln Ile Tyr Gly Asp His Arg
465 470 475 480

Asp Ala Ser Gln His Asn His Asn Lys Ile Glu Gln Val Asp Phe Ser

485

490

495

Trp Asn Thr Tyr Ala Asp Gly Lys Leu Ala Phe Tyr Asp His Tyr Leu
 500 505 510

Ile Glu Gln Ile Gln Ser Gly Lys Glu Pro Glu Val Arg Gln Phe Phe
 515 520 525

Phe Leu Leu Ala Val Cys His Thr Val Met Val Asp Arg Thr Asp Gly
 530 535 540

Gln Leu Asn Tyr Gln Ala Ala Ser Pro Asp Glu Gly Ala Leu Val Asn
 545 550 555 560

Ala Ala Arg Asn Phe Gly Phe Ala Phe Leu Ala Arg Thr Gln Asn Thr
 565 570 575

Ile Thr Ile Ser Glu Leu Gly Thr Glu Arg Thr Tyr Asn Val Leu Ala
 580 585 590

Ile Leu Asp Phe Asn Ser Asp Arg Lys Arg Met Ser Ile Ile Val Arg
 595 600 605

Thr Pro Glu Gly Asn Ile Lys Leu Tyr Cys Lys Gly Ala Asp Thr Val
 610 615 620

Ile Tyr Glu Arg Leu His Arg Met Asn Pro Thr Lys Gln Glu Thr Gln
 625 630 635 640

Asp Ala Leu Asp Ile Phe Ala Asn Glu Thr Leu Arg Thr Leu Cys Leu
 645 650 655

Cys Tyr Lys Glu Ile Glu Glu Lys Glu Phe Thr Glu Trp Asn Lys Lys
 660 665 670

Phe Met Ala Ala Ser Val Ala Ser Thr Asn Arg Asp Glu Ala Leu Asp
 675 680 685

Lys Val Tyr Glu Glu Ile Glu Lys Asp Leu Ile Leu Leu Gly Ala Thr
690 695 700

Ala Ile Glu Asp Lys Leu Gln Asp Gly Val Pro Glu Thr Ile Ser Lys
705 710 715 720

Leu Ala Lys Ala Asp Ile Lys Ile Trp Val Leu Thr Gly Asp Lys Lys
725 730 735

Glu Thr Ala Glu Asn Ile Gly Phe Ala Cys Glu Leu Leu Thr Glu Asp
740 745 750

Thr Thr Ile Cys Tyr Gly Glu Asp Ile Asn Ser Leu Leu His Ala Arg
755 760 765

Met Glu Asn Gln Arg Asn Arg Gly Gly Val Tyr Ala Lys Phe Ala Pro
770 775 780

Pro Val Gln Glu Ser Phe Phe Pro Pro Gly Gly Asn Arg Ala Leu Ile
785 790 795 800

Ile Thr Gly Ser Trp Leu Asn Glu Ile Leu Leu Glu Lys Lys Thr Lys
805 810 815

Arg Asn Lys Ile Leu Lys Leu Lys Phe Pro Arg Thr Glu Glu Glu Arg
820 825 830

Arg Met Arg Thr Gln Ser Lys Arg Arg Leu Glu Ala Lys Lys Glu Gln
835 840 845

Arg Gln Lys Asn Phe Val Asp Leu Ala Cys Glu Cys Ser Ala Val Ile
850 855 860

Cys Cys Arg Val Thr Pro Lys Gln Lys Ala Met Val Val Asp Leu Val
865 870 875 880

Lys Arg Tyr Lys Lys Ala Ile Thr Leu Ala Ile Gly Asp Gly Ala Asn

885

890

895

Asp Val Asn Met Ile Lys Thr Ala His Ile Gly Val Gly Ile Ser Gly
 900 905 910

Gln Glu Gly Met Gln Ala Val Met Ser Ser Asp Tyr Ser Phe Ala Gln
 915 920 925

Phe Arg Tyr Leu Gln Arg Leu Leu Leu Val His Gly Arg Trp Ser Tyr
 930 935 940

Ile Arg Met Cys Lys Phe Leu Arg Tyr Phe Phe Tyr Lys Asn Phe Ala
 945 950 955 960

Phe Thr Leu Val His Phe Trp Tyr Ser Phe Phe Asn Gly Tyr Ser Ala
 965 970 975

Gln Thr Ala Tyr Glu Asp Trp Phe Ile Thr Leu Tyr Asn Val Leu Tyr
 980 985 990

Thr Ser Leu Pro Val Leu Leu Met Gly Leu Leu Asp Gln Asp Val Ser
 995 1000 1005

Asp Lys Leu Ser Leu Arg Phe Pro Gly Leu Tyr Ile Val Gly Gln
 1010 1015 1020

Arg Asp Leu Leu Phe Asn Tyr Lys Arg Phe Phe Val Ser Leu Leu
 1025 1030 1035

His Gly Val Leu Thr Ser Met Ile Leu Phe Phe Ile Pro Leu Gly
 1040 1045 1050

Ala Tyr Leu Gln Thr Val Gly Gln Asp Gly Glu Ala Pro Ser Asp
 1055 1060 1065

Tyr Gln Ser Phe Ala Val Thr Ile Ala Ser Ala Leu Val Ile Thr
 1070 1075 1080

Val Asn Phe Gln Ile Gly Leu Asp Thr Ser Tyr Trp Thr Phe Val
1085 1090 1095

Asn Ala Phe Ser Ile Phe Gly Ser Ile Ala Leu Tyr Phe Gly Ile
1100 1105 1110

Met Phe Asp Phe His Ser Ala Gly Ile His Val Leu Phe Pro Ser
1115 1120 1125

Ala Phe Gln Phe Thr Gly Thr Ala Ser Asn Ala Leu Arg Gln Pro
1130 1135 1140

Tyr Ile Trp Leu Thr Ile Ile Leu Ala Val Ala Val Cys Leu Leu
1145 1150 1155

Pro Val Val Ala Ile Arg Phe Leu Ser Met Thr Ile Trp Pro Ser
1160 1165 1170

Glu Ser Asp Lys Ile Gln Lys His Arg Lys Arg Leu Lys Ala Glu
1175 1180 1185

Glu Gln Trp Gln Arg Arg Gln Gln Val Phe Arg Arg Gly Val Ser
1190 1195 1200

Thr Arg Arg Ser Ala Tyr Ala Phe Ser His Gln Arg Gly Tyr Ala
1205 1210 1215

Asp Leu Ile Ser Ser Gly Arg Ser Ile Arg Lys Lys Arg Ser Pro
1220 1225 1230

Leu Asp Ala Ile Val Ala Asp Gly Thr Ala Glu Tyr Arg Arg Thr
1235 1240 1245

Gly Asp Ser
1250

<210> 2

<211> 3756

<212> DNA

<213> Homo sapiens

<400> 2

atgagtacag aaagagactc agaaacgaca tttgacgagg attctcagcc taatgacgaa	60
gtggttccct acagtgatga tgaaacagaa gatgaacttg atgaccaggg gtctgctggt	120
gaaccagaac aaaaccgagt caacagggaa gcagaggaga accgggagcc attcagaaaa	180
gaatgtacat ggcaagtcaa agcaaacgat cgcaagtacc acgaacaacc tcactttatg	240
aacacaaaat tcttgtgtat taaggagagt aaatatgcga ataatgcaat taaaacatac	300
aagtacaacg catttacctt tataccaatg aatctgtttg agcagtttaa gagagcagcc	360
aatttatatt tcctggctct tcttatctta caggcagttc ctcaaactc taccctggct	420
tggtagacca cactagtgcc cctgcttggt gtgctgggcg tcaactgcaat caaagacctg	480
gtggacgatg tggctcgcca taaaatggat aaggaaatca acaataggac gtgtgaagtc	540
attaaggatg gcaggttcaa agttgctaag tggaaagaaa ttcaagttgg agacgtcatt	600
cgtctgaaaa aaaatgattt tgttccagct gacattctcc tgctgtctag ctctgagcct	660
aacagcctct gctatgtgga aacagcagaa ctggatggag aaaccaattt aaaatttaag	720
atgtcacttg aaatcacaga ccagtacctc caaagagaag atacattggc tacatttgat	780
ggttttattg aatgtgaaga acccaataac agactagata agtttacagg aacactattt	840
tggagaaaca caagttttcc tttggatgct gataaaattt tgttacgtgg ctgtgtaatt	900
aggaacaccg atttctgcca cggcttagtc atttttgcag gtgctgacac taaaataatg	960
aagaatagtg ggaaaaccag atttaaaaga actaaaattg attacttgat gaactacatg	1020
gtttacacga tctttgttgt tcttattctg ctttctgctg gtcttgccat cggccatgct	1080
tattgggaag cacaggtggg caattcctct tggtagctct atgatggaga agacgataca	1140
ccctcctacc gtggattcct catTTTctgg ggctatatca ttgttctcaa caccatggta	1200
cccattctct tctatgtcag cgtggaagtg attcgtcttg gacagagtca cttcatcaac	1260
tgggacctgc aaatgtacta tgctgagaag gacacacccg caaaagctag aaccaccaca	1320
ctcaatgaac agctcgggca gatccattat atcttctctg ataagacggg gacactcaca	1380

caaaatatca tgacctttaa aaagtgctgt atcaacgggc agatatatgg ggaccatcgg	1440
gatgcctctc aacacaacca caacaaaata gagcaagttg attttagctg gaatacatat	1500
gctgatggga agcttgcatt ttatgaccac tatcttattg agcaaatcca gtcagggaaa	1560
gagccagaag tacgacagtt cttcttcttg ctcgcagttt gccacacagt catggtggat	1620
aggactgatg gtcagctcaa ctaccaggca gcctctcccg atgaagggtgc cctggtaaac	1680
gctgccagga actttggctt tgccttcctc gccaggaccc agaacaccat caccatcagt	1740
gaactgggca ctgaaaggac ttacaatgtt cttgccattt tggacttcaa cagtgaccgg	1800
aagcgaatgt ctatcattgt aagaacccca gaaggcaata tcaagcttta ctgtaaaggt	1860
gctgacactg ttatttatga acggttacat cgaatgaatc ctactaagca agaaacacag	1920
gatgccctgg atatctttgc aaatgaaact cttagaaccc tatgcctttg ctacaaggaa	1980
attgaagaaa aagaatttac agaatggaat aaaaagttta tggctgccag tgtggcctcc	2040
accaaccggg acgaagctct ggataaagta tatgaggaga ttgaaaaaga cttattctc	2100
ctgggagcta cagctattga agacaagcta caggatggag ttccagaaac catttcaaaa	2160
cttgcaaaag ctgacattaa gatctgggtg cttactggag acaaaaagga aactgctgaa	2220
aatataggat ttgcttgta acttctgact gaagacacca ccatctgcta tggggaggat	2280
attaattctc ttcttcatgc aaggatggaa aaccagagga atagagggtg cgtctacgca	2340
aagtttgcac ctctgtgca ggaatctttt tttccaccgc gtggaaaccg tgccttaatc	2400
atcactgggt cttggttgaa tgaaattctt ctcgagaaaa agaccaagag aaataagatt	2460
ctgaagctga agttccaag aacagaagaa gaaagacgga tgcggacca aagtaaaagg	2520
aggctagaag ctaagaaaaga gcagcggcag aaaaactttg tggacctggc ctgcgagtgc	2580
agcgcagtca tctgctgccg cgtcaccccc aagcagaagg ccatggtggt ggacctggtg	2640
aagaggtaca agaaagccat cacgctggcc atcggagatg gggccaatga cgtgaacatg	2700
atcaaaaactg cccacattgg cgttggaata agtggacaag aaggaaatgca agctgtcatg	2760
tcgagtgact attcctttgc tcagttccga tatctgcaga ggctactgct ggtgcatggc	2820
cgatggtctt acataaggat gtgcaagttc ctacgatact tcttttaca aaactttgcc	2880

tttacttttg	ttcattttctg	gtactccttc	ttcaatggct	actctgcgca	gactgcatac	2940
gaggattggt	tcatcaccct	ctacaacgtg	ctgtacacca	gcctgcccgt	gctcctcatg	3000
gggctgctcg	accaggatgt	gagtgacaaa	ctgagcctcc	gattccctgg	gttatacata	3060
gtgggacaaa	gagacttact	attcaactat	aagagattct	ttgtaagctt	gttgcatggg	3120
gtcctaacat	cgatgatcct	cttcttcata	cctcttggag	cttatctgca	aaccgtaggg	3180
caggatggag	aggcaccttc	cgactaccag	tcttttgccg	tcaccattgc	ctctgctctt	3240
gtaataacag	tcaattttcca	gattggcttg	gatacttctt	attggacttt	tgtgaatgct	3300
ttttcaattt	ttggaagcat	tgcactttat	tttggcatca	tgtttgactt	tcatagtgt	3360
ggaatacatg	ttctctttcc	atctgcattt	caattttacag	gcacagcttc	aaacgctctg	3420
agacagccat	acatttggtt	aactatcatc	ctggctgttg	ctgtgtgctt	actacccgtc	3480
gttgccattc	gattcctgtc	aatgaccatc	tggccatcag	aaagtgataa	gatccagaag	3540
catcgcaagc	ggttgaaggc	ggaggagcag	tggcagcgac	ggcagcaggt	gttccgccgg	3600
ggcgtgtcaa	cgcggcgctc	ggcctacgcc	ttctcgcacc	agcggggcta	cgcggaacctc	3660
atctcctccg	ggcgcagcat	ccgcaagaag	cgctcgccgc	ttgatgccat	cgtggcggat	3720
ggcaccgcgg	agtacaggcg	caccggggac	agctga			3756

<210> 3
 <211> 1321
 <212> PRT
 <213> Homo sapiens

<400> 3

Met	Ser	Asp	Ser	Val	Ile	Leu	Arg	Ser	Ile	Lys	Lys	Phe	Gly	Glu	Glu
1				5					10					15	

Asn	Asp	Gly	Phe	Glu	Ser	Asp	Lys	Ser	Tyr	Asn	Asn	Asp	Lys	Lys	Ser
		20						25					30		

Arg	Leu	Gln	Asp	Glu	Lys	Lys	Gly	Asp	Gly	Val	Arg	Val	Gly	Phe	Phe
		35					40					45			

Gln Leu Phe Arg Phe Ser Ser Ser Thr Asp Ile Trp Leu Met Phe Val
50 55 60

Gly Ser Leu Cys Ala Phe Leu His Gly Ile Ala Gln Pro Gly Val Leu
65 70 75 80

Leu Ile Phe Gly Thr Met Thr Asp Val Phe Ile Asp Tyr Asp Val Glu
85 90 95

Leu Gln Glu Leu Gln Ile Pro Gly Lys Ala Cys Val Asn Asn Thr Ile
100 105 110

Val Trp Thr Asn Ser Ser Leu Asn Gln Asn Met Thr Asn Gly Thr Arg
115 120 125

Cys Gly Leu Leu Asn Ile Glu Ser Glu Met Ile Lys Phe Ala Ser Tyr
130 135 140

Tyr Ala Gly Ile Ala Val Ala Val Leu Ile Thr Gly Tyr Ile Gln Ile
145 150 155 160

Cys Phe Trp Val Ile Ala Ala Ala Arg Gln Ile Gln Lys Met Arg Lys
165 170 175

Phe Tyr Phe Arg Arg Ile Met Arg Met Glu Ile Gly Trp Phe Asp Cys
180 185 190

Asn Ser Val Gly Glu Leu Asn Thr Arg Phe Ser Asp Asp Ile Asn Lys
195 200 205

Ile Asn Asp Ala Ile Ala Asp Gln Met Ala Leu Phe Ile Gln Arg Met
210 215 220

Thr Ser Thr Ile Cys Gly Phe Leu Leu Gly Phe Phe Arg Gly Trp Lys
225 230 235 240

Leu Thr Leu Val Ile Ile Ser Val Ser Pro Leu Ile Gly Ile Gly Ala

245

250

255

Ala Thr Ile Gly Leu Ser Val Ser Lys Phe Thr Asp Tyr Glu Leu Lys
 260 265 270

Ala Tyr Ala Lys Ala Gly Val Val Ala Asp Glu Val Ile Ser Ser Met
 275 280 285

Arg Thr Val Ala Ala Phe Gly Gly Glu Lys Arg Glu Val Glu Arg Tyr
 290 295 300

Glu Lys Asn Leu Val Phe Ala Gln Arg Trp Gly Ile Arg Lys Gly Ile
 305 310 315 320

Val Met Gly Phe Phe Thr Gly Phe Val Trp Cys Leu Ile Phe Leu Cys
 325 330 335

Tyr Ala Leu Ala Phe Trp Tyr Gly Ser Thr Leu Val Leu Asp Glu Gly
 340 345 350

Glu Tyr Thr Pro Gly Thr Leu Val Gln Ile Phe Leu Ser Val Ile Val
 355 360 365

Gly Ala Leu Asn Leu Gly Asn Ala Ser Pro Cys Leu Glu Ala Phe Ala
 370 375 380

Thr Gly Arg Ala Ala Ala Thr Ser Ile Phe Glu Thr Ile Asp Arg Lys
 385 390 395 400

Pro Ile Ile Asp Cys Met Ser Glu Asp Gly Tyr Lys Leu Asp Arg Ile
 405 410 415

Lys Gly Glu Ile Glu Phe His Asn Val Thr Phe His Tyr Pro Ser Arg
 420 425 430

Pro Glu Val Lys Ile Leu Asn Asp Leu Asn Met Val Ile Lys Pro Gly
 435 440 445

Glu Met Thr Ala Leu Val Gly Pro Ser Gly Ala Gly Lys Ser Thr Ala
450 455 460

Leu Gln Leu Ile Gln Arg Phe Tyr Asp Pro Cys Glu Gly Met Val Thr
465 470 475 480

Val Asp Gly His Asp Ile Arg Ser Leu Asn Ile Gln Trp Leu Arg Asp
485 490 495

Gln Ile Gly Ile Val Glu Gln Glu Pro Val Leu Phe Ser Thr Thr Ile
500 505 510

Ala Glu Asn Ile Arg Tyr Gly Arg Glu Asp Ala Thr Met Glu Asp Ile
515 520 525

Val Gln Ala Ala Lys Glu Ala Asn Ala Tyr Asn Phe Ile Met Asp Leu
530 535 540

Pro Gln Gln Phe Asp Thr Leu Val Gly Glu Gly Gly Gly Gln Met Ser
545 550 555 560

Gly Gly Gln Lys Gln Arg Val Ala Ile Ala Arg Ala Leu Ile Arg Asn
565 570 575

Pro Lys Ile Leu Leu Leu Asp Met Ala Thr Ser Ala Leu Asp Asn Glu
580 585 590

Ser Glu Ala Met Val Gln Glu Val Leu Ser Lys Ile Gln His Gly His
595 600 605

Thr Ile Ile Ser Val Ala His Arg Leu Ser Thr Val Arg Ala Ala Asp
610 615 620

Thr Ile Ile Gly Phe Glu His Gly Thr Ala Val Glu Arg Gly Thr His
625 630 635 640

Glu Glu Leu Leu Glu Arg Lys Gly Val Tyr Phe Thr Leu Val Thr Leu

645

650

655

Gln Ser Gln Gly Asn Gln Ala Leu Asn Glu Glu Asp Ile Lys Asp Ala
 660 665 670

Thr Glu Asp Asp Met Leu Ala Arg Thr Phe Ser Arg Gly Ser Tyr Gln
 675 680 685

Asp Ser Leu Arg Ala Ser Ile Arg Gln Arg Ser Lys Ser Gln Leu Ser
 690 695 700

Tyr Leu Val His Glu Pro Pro Leu Ala Val Val Asp His Lys Ser Thr
 705 710 715 720

Tyr Glu Glu Asp Arg Lys Asp Lys Asp Ile Pro Val Gln Glu Glu Val
 725 730 735

Glu Pro Ala Pro Val Arg Arg Ile Leu Lys Phe Ser Ala Pro Glu Trp
 740 745 750

Pro Tyr Met Leu Val Gly Ser Val Gly Ala Ala Val Asn Gly Thr Val
 755 760 765

Thr Pro Leu Tyr Ala Phe Leu Phe Ser Gln Ile Leu Gly Thr Phe Ser
 770 775 780

Ile Pro Asp Lys Glu Glu Gln Arg Ser Gln Ile Asn Gly Val Cys Leu
 785 790 795 800

Leu Phe Val Ala Met Gly Cys Val Ser Leu Phe Thr Gln Phe Leu Gln
 805 810 815

Gly Tyr Ala Phe Ala Lys Ser Gly Glu Leu Leu Thr Lys Arg Leu Arg
 820 825 830

Lys Phe Gly Phe Arg Ala Met Leu Gly Gln Asp Ile Ala Trp Phe Asp
 835 840 845

Asp Leu Arg Asn Ser Pro Gly Ala Leu Thr Thr Arg Leu Ala Thr Asp
850 855 860

Ala Ser Gln Val Gln Gly Ala Ala Gly Ser Gln Ile Gly Met Ile Val
865 870 875 880

Asn Ser Phe Thr Asn Val Thr Val Ala Met Ile Ile Ala Phe Ser Phe
885 890 895

Ser Trp Lys Leu Ser Leu Val Ile Leu Cys Phe Phe Pro Phe Leu Ala
900 905 910

Leu Ser Gly Ala Thr Gln Thr Arg Met Leu Thr Gly Phe Ala Ser Arg
915 920 925

Asp Lys Gln Ala Leu Glu Met Val Gly Gln Ile Thr Asn Glu Ala Leu
930 935 940

Ser Asn Ile Arg Thr Val Ala Gly Ile Gly Lys Glu Arg Arg Phe Ile
945 950 955 960

Glu Ala Leu Glu Thr Glu Leu Glu Lys Pro Phe Lys Thr Ala Ile Gln
965 970 975

Lys Ala Asn Ile Tyr Gly Phe Cys Phe Ala Phe Ala Gln Cys Ile Met
980 985 990

Phe Ile Ala Asn Ser Ala Ser Tyr Arg Tyr Gly Gly Tyr Leu Ile Ser
995 1000 1005

Asn Glu Gly Leu His Phe Ser Tyr Val Phe Arg Val Ile Ser Ala
1010 1015 1020

Val Val Leu Ser Ala Thr Ala Leu Gly Arg Ala Phe Ser Tyr Thr
1025 1030 1035

Pro Ser Tyr Ala Lys Ala Lys Ile Ser Ala Ala Arg Phe Phe Gln

1040

1045

1050

Leu Leu Asp Arg Gln Pro Pro Ile Ser Val Tyr Asn Thr Ala Gly
 1055 1060 1065

Glu Lys Trp Asp Asn Phe Gln Gly Lys Ile Asp Phe Val Asp Cys
 1070 1075 1080

Lys Phe Thr Tyr Pro Ser Arg Pro Asp Ser Gln Val Leu Asn Gly
 1085 1090 1095

Leu Ser Val Ser Ile Ser Pro Gly Gln Thr Leu Ala Phe Val Gly
 1100 1105 1110

Ser Ser Gly Cys Gly Lys Ser Thr Ser Ile Gln Leu Leu Glu Arg
 1115 1120 1125

Phe Tyr Asp Pro Asp Gln Gly Lys Val Met Ile Asp Gly His Asp
 1130 1135 1140

Ser Lys Lys Val Asn Val Gln Phe Leu Arg Ser Asn Ile Gly Ile
 1145 1150 1155

Val Ser Gln Glu Pro Val Leu Phe Ala Cys Ser Ile Met Asp Asn
 1160 1165 1170

Ile Lys Tyr Gly Asp Asn Thr Lys Glu Ile Pro Met Glu Arg Val
 1175 1180 1185

Ile Ala Ala Ala Lys Gln Ala Gln Leu His Asp Phe Val Met Ser
 1190 1195 1200

Leu Pro Glu Lys Tyr Glu Thr Asn Val Gly Ser Gln Gly Ser Gln
 1205 1210 1215

Leu Ser Arg Gly Glu Lys Gln Arg Ile Ala Ile Ala Arg Ala Ile
 1220 1225 1230

Val	Arg	Asp	Pro	Lys	Ile	Leu	Leu	Leu	Asp	Glu	Ala	Thr	Ser	Ala
1235						1240					1245			

Leu	Asp	Thr	Glu	Ser	Glu	Lys	Thr	Val	Gln	Val	Ala	Leu	Asp	Lys
1250						1255					1260			

Ala	Arg	Glu	Gly	Arg	Thr	Cys	Ile	Val	Ile	Ala	His	Arg	Leu	Ser
1265						1270					1275			

Thr	Ile	Gln	Asn	Ala	Asp	Ile	Ile	Ala	Val	Met	Ala	Gln	Gly	Val
1280						1285					1290			

Val	Ile	Glu	Lys	Gly	Thr	His	Glu	Glu	Leu	Met	Ala	Gln	Lys	Gly
1295						1300					1305			

Ala	Tyr	Tyr	Lys	Leu	Val	Thr	Thr	Gly	Ser	Pro	Ile	Ser
1310						1315					1320	

<210> 4
 <211> 3966
 <212> DNA
 <213> Homo sapiens

<400> 4	
atgtctgact cagtaattct tcgaagtata aagaaatttg gagaggagaa tgatggtttt	60
gagtcagata aatcatataa taatgataag aaatcaaggt tacaagatga gaagaaaggt	120
gatggcgtta gagttggctt ctttcaattg tttcggtttt cttcatcaac tgacatttgg	180
ctgatgtttg tgggaagttt gtgtgcattt ctccatggaa tagcccagcc aggcgtgcta	240
ctcattttttg gcacaatgac agatgttttt attgactacg acgttgagtt acaagaactc	300
cagattccag gaaaagcatg tgtgaataac accattgtat ggactaacag ttccctcaac	360
cagaacatga caaatggaac acgttgtggg ttgctgaaca tcgagagcga aatgatcaaa	420
tttgccagtt actatgctgg aattgctgtc gcagtactta tcacaggata tattcaaata	480
tgcttttggg tcattgccgc agctcgtcag atacagaaaa tgagaaaatt ttactttagg	540
agaataatga gaatggaaat aggggtggtt gactgcaatt cagtggggga gctgaataca	600

agattctctg atgatattaa taaaatcaat gatgccatag ctgaccaa	at ggcccttttc	660
attcagcgca tgacctcgac catctgtggt ttcctgttgg gatttttcag	gggttgga	720
ctgaccttgg ttattatttc tgtcagccct ctcataggga ttggagcagc	caccattggt	780
ctgagtgtgt ccaagtttac ggactatgag ctgaaggcct atgccaaagc	aggggtggtg	840
gctgatgaag tcatttcac aatgagaaca gtggctgctt ttggtggtga	gaaaagagag	900
gttgaaagggt atgagaaaaa tcttgtgttc gcccagcggt	ggggaattag aaaaggaata	960
gtgatgggat tctttactgg attcgtgtgg tgtctcatct ttttgtgtta	tgcactggcc	1020
ttctggtacg gctccacact tgtcctggat gaaggagaat atacaccagg	aacccttgtc	1080
cagattttcc tcagtgtcat agtaggagct ttaaactctt	gcaatgcctc tccttgtttg	1140
gaagcctttg caactggacg tgcagcagcc accagcattt ttgagacaat	agacaggaaa	1200
cccatcattg actgcatgtc agaagatggt tacaagttgg atcgaatcaa	gggtgaaatt	1260
gaattccata atgtgacctt ccattatcct tccagaccag aggtgaagat	tctaaatgac	1320
ctcaacatgg tcattaaacc aggggaaatg acagctctgg taggaccag	tggagctgga	1380
aaaagtacag cactgcaact cattcagcga ttctatgacc cctgtgaagg	aatggtgacc	1440
gtggatggcc atgacattcg ctctcttaac attcagtggc ttagagatca	gattgggata	1500
gtggagcaag agccagttct gttctctacc accattgcag aaaatattcg	ctatggcaga	1560
gaagatgcaa caatggaaga catagtccaa gctgccaaagg aggccaatgc	ctacaacttc	1620
atcatggacc tgccacagca atttgacacc cttgttggag aaggaggagg	ccagatgagt	1680
ggtggccaga aacaaagggt agctatcgcc agagccctca tccgaaatcc	caagattctg	1740
cttttggaca tggccacctc agctctggac aatgagagtg aagccatggt	gcaagaagtg	1800
ctgagtaaga ttcagcatgg gcacacaatc atttcagttg ctcatcgctt	gtctacggtc	1860
agagctgcag ataccatcat tggttttgaa catggcactg cagtggaaag	agggacccat	1920
gaagaattac tggaaggaa aggtgtttac ttactctag tgactttgca	aagccaggga	1980
aatcaagctc ttaatgaaga ggacataaag gatgcaactg aagatgacat	gcttgcgagg	2040
accttagca gagggagcta ccaggatagt ttaagggtt ccatccggca	acgctccaag	2100

tctcagcttt cttacctggg gcacgaacct ccattagctg ttgtagatca taagtctacc	2160
tatgaagaag atagaaagga caaggacatt cctgtgcagg aagaagttga acctgccccca	2220
gtaggagga ttctgaaatt cagtgcctca gaatggccct acatgctggg agggctctgtg	2280
ggcgcagctg tgaacgggac agtcacaccc ttgtatgcct ttttattcag ccagattcctt	2340
gggacttttt caattcctga taaagaggaa caaaggcac agatcaatgg tgtgtgccta	2400
ctttttgtag caatgggctg tgtatctctt ttcacccaat ttctacaggg atatgccttt	2460
gctaaatctg gggagctcct aacaaaaagg ctacgtaaat ttggtttcag ggcaatgctg	2520
gggcaagata ttgcctgggt tgatgacctc agaaatagcc ctggagcatt gacaacaaga	2580
cttgctacag atgcttccca agttcaaggg gctgccggct ctcagatcgg gatgatagtc	2640
aattccttca ctaacgtcac tgtggccatg atcattgcct tctcctttag ctggaagctg	2700
agcctgggtca tcttgtgctt cttccccctt ttggctttat caggagccac acagaccagg	2760
atgttgacag gatttgcctc tcgagataag caggccctgg agatgggtggg acagattaca	2820
aatgaagccc tcagtaacat ccgcactgtt gctggaattg gaaaggagag gcggttcatt	2880
gaagcacttg agactgagct ggagaagccc ttcaagacag ccattcagaa agccaatatt	2940
tacggattct gctttgcctt tgcccagtg atcatgttta ttgcgaattc tgcttcctac	3000
agatatggag gttacttaat ctccaatgag gggctccatt tcagctatgt gttcagggtg	3060
atctctgcag ttgtactgag tgcaacagct cttggaagag ccttctctta ccccccaagt	3120
tatgcaaaag ctaaaatatc agctgcacgc ttttttcaac tgctggaccg acaacccccca	3180
atcagtgtat acaatactgc aggtgaaaaa tgggacaact tccaggggaa gattgatttt	3240
gttgattgta aatttacata tccttctcga cctgactcgc aagttctgaa tggctctctca	3300
gtgtcgatta gtccagggca gacactggcg tttgttggga gcagtggatg tggcaaaagc	3360
actagcattc agctgttggg acgtttctat gatcctgac aagggaaggat gatgatagat	3420
ggcatgaca gcaaaaaagt aaatgtccag ttcctccgct caaacattgg aattgtttcc	3480
caggaaccag tgttgtttgc ctgtagcata atggacaata tcaagtatgg agacaacacc	3540
aaagaaattc ccatggaaag agtcatagca gctgcaaaac aggctcagct gcatgatttt	3600

gtcatgtcac tcccagagaa atatgaaact aacgttgggt cccaggggtc tcaactctct	3660
agagggggaga aacaacgcat tgctattgct cgggccattg tacgagatcc taaaatcttg	3720
ctactagatg aagccacttc tgccttagac acagaaagtg aaaagacggt gcaggttgct	3780
ctagacaaag ccagagaggg tcggacctgc attgtcattg cccatcgctt gtccaccatc	3840
cagaacgcgg atatcattgc tgtcatggca caggggggtg tgattgaaaa ggggacccat	3900
gaagaactga tggcccaaaa aggagcctac tacaaactag tcaccactgg atcccccatc	3960
agttga	3966