



US 20130209436A1

(19) **United States**(12) **Patent Application Publication**
Quinn et al.(10) **Pub. No.: US 2013/0209436 A1**(43) **Pub. Date: Aug. 15, 2013**(54) **LYSOSOMAL STORAGE DISEASE ENZYMES****Publication Classification**(75) Inventors: **Anthony Quinn**, Chestnut Hill, MA
(US); **Alex J. Harvey**, Athens, GA (US)(51) **Int. Cl.**
C12N 9/20 (2006.01)
A61K 38/38 (2006.01)
A61K 31/194 (2006.01)
A61K 38/46 (2006.01)(73) Assignee: **SYNAGEVA BIOPHARMA CORP.**,
Lexington, MA (US)(21) Appl. No.: **13/642,790**(52) **U.S. Cl.**
CPC **C12N 9/20** (2013.01); **A61K 38/465**
(2013.01); **A61K 38/385** (2013.01); **A61K**
31/194 (2013.01)(22) PCT Filed: **Apr. 23, 2011**USPC **424/94.6**; 435/198; 800/19(86) PCT No.: **PCT/US2011/033699**

§ 371 (c)(1),

(2), (4) Date: **Apr. 30, 2013**(57) **ABSTRACT****Related U.S. Application Data**

(60) Provisional application No. 61/343,177, filed on Apr. 23, 2010, provisional application No. 61/396,376, filed on May 26, 2010, provisional application No. 61/403,011, filed on Sep. 9, 2010, provisional application No. 61/456,014, filed on Oct. 29, 2010, provisional application No. 61/432,372, filed on Jan. 13, 2011.

The present invention provides compositions of recombinant human lysosomal acid lipase having particular glycosylation patterns for internalization into target cells, a vector containing the nucleic acid encoding human lysosomal acid lipase, a host cell transformed with the vector, pharmaceutical compositions comprising the recombinant human lysosomal acid lipase and method of treating conditions associated with lysosomal acid lipase deficiency.

	1	60
rhLAL	MKMRFLGLVVCLVLWTLHSEG	<u>SGGKLTAVDPETNMNVSEIISYWGFPSSEYLVETEDGYI</u>
Natural	MKMRFLGLVVCLVLWTLHSEG	<u>SGGKLTAVDPETNMNVSEIISYWGFPSSEYLVETEDGYI</u>
	61	120
rhLAL	<u>LCLNRI PHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS</u>	<u>SLGFILADAGFDVVMGNS</u>
Natural	<u>LCLNRI PHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANS</u>	<u>SLGFILADAGFDVVMGNS</u>
	121	180
rhLAL	<u>RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQE</u>	<u>QVYYVGHSQGTTIG</u>
Natural	<u>RGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFILNKTGQE</u>	<u>QVYYVGHSQGTTIG</u>
	181	240
rhLAL	<u>FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL</u>	
Natural	<u>FIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLFGDKEFLPQSAFL</u>	
	241	300
rhLAL	<u>KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV</u>	
Natural	<u>KWLGTHVCTHVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHSPAGTSVQNMLHWSQAV</u>	
	301	360
rhLAL	<u>KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVDVNILLTQI</u>	
Natural	<u>KFQKFQAFDWGSSAKNYFHYNQSYPPTYNVKDMLVPTAVWSGGHDWLADVDVNILLTQI</u>	
	361	399
rhLAL	<u>TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ</u>	(SEQ ID NO:1)
Natural	<u>TNLVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ</u>	(SEQ ID NO:20)

FIG. 1

```
1   ATGAAAATGC GGTTCCTGGG GTTGGTGGTC TGTTTGGTTC TCTGGACCCCT
51  GCATTCCGAG GGGTCCGGAG GGAAACTGAC AGCTGTGGAT CCTGAAACAA
101 ACATGAATGT CAGTGAAATT ATCTCTTACT GGGGATTCCC TAGTGAGGAA
151 TACCTAGTTG AGACAGAAGA TGGATATATT CTGTGCCTTA ACCGAATTCC
201 TCATGGGAGG AAGAACCATT CTGACAAAGG TCCCAAACCA GTTGTCTTCC
251 TGCAACATGG CTTGCTGGCA GATTCTAGTA ACTGGGTCAC AAACCTTGCC
301 AACAGCAGCC TGGGCTTCAT TCTTGCTGAT GCTGGTTTTG ACGTGTGGAT
351 GGGCAACAGC AGAGGAAATA CCTGGTCTCG GAAACATAAG ACACTCTCAG
401 TTTCTCAGGA TGAATTCTGG GCTTTCAGTT ATGATGAGAT GGCAAAATAT
451 GACCTACCAG CTTCCATTAA CTTCAATTCTG AATAAGACTG GCCAAGAACA
501 AGTGTATTAT GTGGGTCATT CTCAAGGCAC CACTATAGGT TTTATAGCAT
551 TTTCACAGAT CCCTGAGCTG GCTAAAAGGA TTAAAATGTT TTTTGCCCTG
601 GGTCCGTGG CTTCCGTCGC CTTCTGTACT AGCCCTATGG CCAAACCTGGG
651 ACGACTGCCA GATCATCTCA TTAAGGACCT CTTTGAGAGC AAAGAATTTC
701 TTCCCCAGAG TGC GTTTTTG AAGTGGCTGG GTACCCACGT TTGCACTCAT
751 GTCATACTGA AGGAGCTCTG TGGAAATCTC TGTTTTCTTC TGTGTGGATT
801 TAATGAGAGA AATTTAAATA TGTCTAGAGT GGATGTGTAT ACAACACATT
851 CTCCTGCTGG AACTTCTGTG CAAAACATGT TACACTGGAG CCAGGCTGTT
901 AAATTCAAA AGTTTCAAGC CTTTGACTGG GGAAGCAGTG CCAAGAATTA
951 TTTTCATTAC AACCAGAGTT ATCCTCCCAC ATACAATGTG AAGGACATGC
1001 TTGTGCCGAC TGCAGTCTGG AGCGGGGGTC ACGACTGGCT TGCAGATGTC
1051 TACGACGTCA ATATCTTACT GACTCAGATC ACCAACTTGG TGTTCCATGA
1101 GAGCATTCCG GAATGGGAGC ATCTTGACTT CATTTGGGGC CTGGATGCCC
```

(SEQ ID NO:5)

FIG. 2

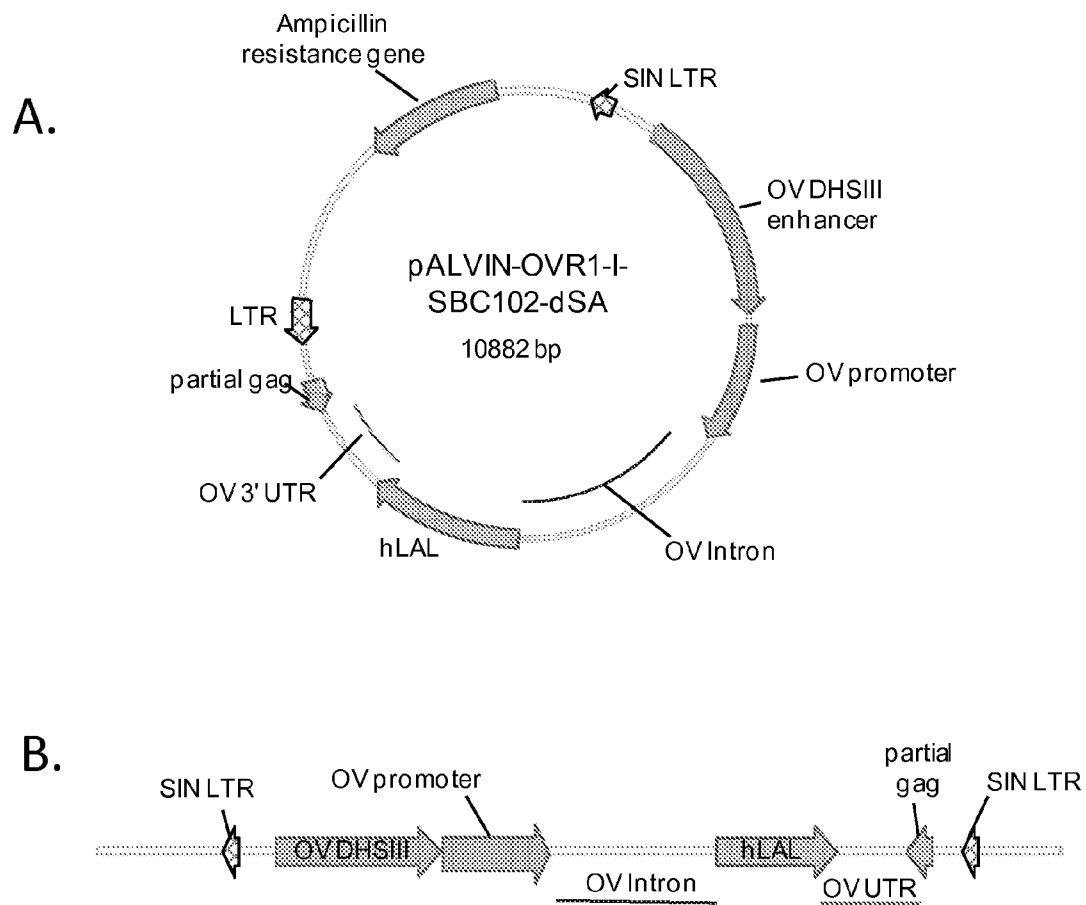


FIG. 3

Sequence of pALVIN-OVR1-I-SBC102-dSA (10882 bp)

Feature	Location (bp)
SIN LTR	521-693 (complement strand)
DHSIII enhancer	1069-2720
OV promoter	2720-3851
Intron	3899-5487
hLAL CDS	5505-6704
OV 3'UTR	6719-7392
partial gag CDS	7404-7657 (complement strand)
LTR	7955-8300 (complement strand)
Ampicillin resistance gene	9764-10621 (complement strand)

```

1   CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT CCCTTTAGGG TTCCGATTTA GTGCTTTACG
61  GCACCTCGAC CCCAAAAAAC TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG
121 ATAGACGGTT TTTCGCCCTT TGACGTGGA GTCCACGTTT TTTAATAGTG GACTCTTGTT
181 CCAAACCTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT
241 GCCGATTTTC GCCTATTGGT TAAAAAATGA GCTGATTAA CAAAAATTTA ACGCGAATTT
301 TAACAAATA TTAACGCTTA CAATTTCCAT TCGCCATTCA GGCTGCGCAA CTGTGGGAA
361 GGGCGATCGG TCGGGGCTC TTCGCTATTA CGCCAGCTGG CGAAAGGGGG ATGTGCTGCA
421 AGGCGATTAA GTTGGGTAAC GCCAGGGTTT TCCCAGTCAC GACGTTGTAA AACGACGGCC
481 AGTGAGCGCG TATTCCTTAA CGATCACGTC GGGGTCACCA AATGAAGCCT TCTGCTTCAT
541 GCATGTGCTC GTAGTCGTCA GGAATCAAC GGTCCGGCCA TCAACCCAGG TGCACACCAA
601 TGTGGTGAAT GGTCAAATGG CGTTTATTGT ATCGAGCTAG GCACTTAAAT ACAATATCTC
661 TGCAATGCGG AATTCAAGTG TTCGTCCAAT CCGTCCCCCT CCCTATGCAA AAGCGAAACT
721 ACTATATCCT GAGGGGACTC CTAACCGCGT ACAACCGAAG CCCCCTTTT CGCCTAAACA
781 TGCTATTGTC CCCTCAGTCA AGCCTTGCCC GTTACAACCC GATTGCAAG CCTTGCCCTC
841 CCCACATTAT CCGTAGCATT ATTTCTAGC AGTCATCAGA GCTACAGAAG ATACTCTATG
901 CTGTAGCCAA GTCTACAAGT TTACTATTCA GCGACCTCCT ATATTCCGCG TGCCAGCCGA
961 TCAATTACCA ATCCAACCAG CTATCACACG GAATACAAGA ACTCGCCTAC GCTCTTCTTT
1021 CGGGCTGCTT ATAAGCCTCC TGTAATTTTT TTATATTCTT CGCTCGAGTC TCTTCAGAAT
1081 GGCACAGCAC CGCTGCAGAA AAATGCCAGG TGGACTATGA ACTCACATCC AAAGGAGCTT
1141 GACCTGATAC CTGATTTTCT TCAAACAGGG GAAACAACAC AATCCCACAA AACAGCTCAG
1201 AGAGAAACCA TCACTGATGG CTACAGCACC AAGGTATGCA ATGGCAATCC ATTTCGACATT
1261 CATCTGTGAC CTGAGCAAAA TGATTTATCT CTCCATGAAT GGTGCTTCT TTCCCTCATG
1321 AAAAGGCAAT TTCCACACTC ACAATATGCA ACAAAGACAA ACAGAGAACA ATTAATGTGC
1381 TCCTTCCTAA TGTAAAAATT GTAGTGGCAA AGAGGAGAAC AAAATCTCAA GTTCTGAGTA
1441 GGTTTTAGTG ATTGGATAAG AGGCTTTGAC CTGTGAGCTC ACCTGGACTT CATATCCTTT
1501 TGGATAAAAA GTGCTTTTAT AACTTTGAGG TCTCCGAGTC TTTATTCATG AGACTGTTTG
1561 TTTAGGGACA GACCCACAAT GAAATGCCTG GCATAGGAAA GGGCAGCAGA GCCTTAGCTG
1621 ACCTTTTCTT GGGACAAGCA TTGTCAAACA ATGTGTGACA AAATATTTG TACTGCTTTG
1681 CACAGCTGTG CTGGGCAGGG CAATCCATTG CCACCTATCC CAGGTAACCT TCCAACGCA
1741 AGAAGATTGT TGCTTACTCT CTCTAGACCC CCAAGTCAAA CCAACTATGC AGGTATGCTG
1801 ACAACACTAT GATGACAGCC TGTTCTGATC AAGATCTCAT TTGTTTCATG ACAATTTTGT
1861 TTGCTTGCAG CTGGTCTTCC ATTGGGAAAG AGTGTAGTAT ATCCTTCTCA TCTGACAGAA
1921 AAGCAGAAAT TCTCATGCTC CACACTTAAT CTACATTGTT TTAAACCAAC GGCTACTTCT
1981 TGGAGAGGAA AAATGGCTTT TATAAGACTC ACAAACAAA GCTCTGCAAG TCAATGTCAT
2041 ACAAACCTGT TCTGTAGGTC TGGAATCAGG AACTATGTG GAAGTCAAA ATAGAGAGCTT
2101 TAAAAAGCCT TTGGGATCAT TCTCATCTTA TATTGTCAGC ACGATACTAT GACAGTGATA
2161 ACTGACATAA CTGCATCAAT TTCCCTTGATA TTTTATTTGT CTTAAAGTAC AAGACATAGA
2221 GATGGACGTA AAGATGGACA TATGACTCAG GTCTGGACAG GTCCGTGGTC CATGTATGAT
2281 AAAAGAGATG AAGGAAGGA GAATTGAGAC TGTCTAAGAA GGGCTTCAGG GACGTTCTGA
2341 AGGCAGATTT GACTGAATCA GATGTACTGT CCAAGTCTCA TATGTAGCAA TGGAAGGCTG
2401 ATATTGGAGA AATATAAAGA AATGGCTGTG AACTCAAAGT GACCCTGAAC AGAAAAGGGA
2461 TATGGAGTTA AAATAATGTC ACAGAACTGA GGTTTATATG ATATACCATG GGCTGCAGAG
2521 GGTCAGAGTG CTCCACCATG GGCTTCTCTT GGGCTGCAGG GAACCTCTGT TCTACACCTG
2581 GAACACCTCC TGCCCTCCTC CGCACTGACC TCAGTGTGAT CAGGGCTGTT TCTCTCACAT
2641 TTTCTCACTC ACCTCTCCCA ACTACCATTG TACAGCAGTT GTTCTTACAT ATTGCTCCTC
2701 CTGAGGTACA TCTAGCATCG TTAAGTCTCT AGACTTGGCA AGGAGAATGT AGATTTCAC
2761 AGTATATATG TTTTCACAAA AGGAAGGAGA GAAACAAAAG AAAATGGCAC TGACTAACT

```

2821	TCAGCTAGTG	GTATAGGAAA	GTAATTCTGC	TTAACAGAGA	TTGCAGTGAT	CTCTATGTAT
2881	GTCCTGAAGA	ATTATGTTGT	ACTTTTTTCC	CCCATTTTTA	AATCAAACAG	TGCTTTACAG
2941	AGGTGAGAAT	GGTTTCTTTA	CTGTTTGTCA	ATTCTATTAT	TTCAATACAG	AACAATAGCT
3001	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	TCCCTCGAAC	CATGAACACT
3061	CCTCCAGCTG	AATTTCACAA	TTCTCTGTGC	ATCTGCCAGG	CCATTAAGTT	ATTCATGGAA
3121	GATCTTTGAG	GAACACTGCA	AGTTCATATC	ATAAACACAT	TTGAAATTGA	GTATTGTTTT
3181	GCATTGTATG	GAGCTATGTT	TTGCTGTATC	CTCAGAATAA	AAGTTTGTTA	TAAAGCATTC
3241	ACACCCATAA	AAAGATAGAT	TTAAATATTC	CAACTATAGG	AAAGAAAGTG	TGCTGTCTCT
3301	TCACTCTAGT	CTCAGTTGGC	TCCTTCACAT	GCACGCTTCT	TTATTTCTCC	TATTTTGTCA
3361	AGAAAATAAT	AGGTCAAGTC	TTGTTCTCAT	TTATGTCTCG	TCTAGCGTGG	CTCAGATGCA
3421	CATTGTACAT	ACAAGAAGGA	TCAAAAGAAA	CAGACTTCTG	GTCTGTTACT	ACAACCATAG
3481	TAATAAGCAC	ACTAACTAAT	AATTGCTAAT	TATGTTTTCC	ATCTCCAAGG	TTCCACATT
3541	TTTCTGTTTT	CTTAAAGATC	CCATTATCTG	GTTGTAACATG	AAGCTCAATG	GAACATGAGC
3601	AATATTTCCC	AGTCTTCTCT	CCCATCCAAC	AGTCCTGATG	GATTAGCAGA	ACAGGCAGAA
3661	AACACATTGT	TACCCAGAAT	TAAAACTAA	TATTTGCTCT	CCATTCAATC	CAAAATGGAC
3721	CTATTGAAAC	TAAAATCTAA	CCCAATCCCA	TTAAATGATT	TCTATGGTGT	CAAAGGTCAA
3781	ACTTCTGAAG	GGAACCTGTG	GGTGGGTGAC	AATTCAGACT	ATATATTCCC	CAGGGCTCAG
3841	CCAGTGCTCG	TACATACAGC	TAGAAAGCTG	TATTGCCTTT	AGCAGTCAAG	CTCGAAAGGT
3901	AAGCAACTCT	CTGGAATTAC	CTTCTCTCTA	TATTAGCTCT	TACTTGCACC	TAAACTTTAA
3961	AAAATTAACA	ATTATTGTGC	TATGTGTTGT	ATCTTTAAGG	GTGAAGTACC	TGCGTGATAC
4021	CCCCATAAAA	AACCTCTCAC	CTGTGTATGC	ATTCTGCACT	ATTTTATTAT	GTGTAAAAGC
4081	TTTGTGTTTG	TTTTCAGGAG	GCTTATTCTT	TGTGCTTAAA	ATATGTTTTT	AATTTTCAGAA
4141	CATCTTATCC	TGTCGTTTAC	TATCTGATAT	GCTTTGCAGT	TTGCTTGATT	AACTTCTAGC
4201	CCTACAGAGT	GCACAGAGAG	CAAAATCATG	GTGTTTCAGT	AATTCTGGGG	AGTTATTTTA
4261	ATGTGAAAAT	TCCTAGAAAG	TTTAAATTCCT	GCAAAGTGCA	GCTGCTGATC	ACTACACAAG
4321	ATAAAAATGT	GGGGGGTGCA	TAAACGTATA	TTCTTACAAT	AATAGATACA	TGTGAACHTA
4381	TATACAGAAA	AGAAAATGAG	AAAAATGTGT	GTGTGTATAC	TCACACACGT	GGTCAGTAAA
4441	AACTTTTGAG	GGGTTTAATA	CAGAAAATCC	AATCCTGAGG	CCCCAGCACT	CAGTACGCAT
4501	ATAAAGGGCT	GGGCTCTGAA	GGACTTCTGA	CTTTCACAGA	TTATATAAAT	CTCAGGAAAG
4561	CAACTAGATT	CATGCTGGCT	CCAAAAGCTG	TGCTTTATAT	AAGCACACTG	GCTATACAAT
4621	AGTTGTACAG	TTCAGCTCTT	TATAATAGAA	ACAGACAGAA	CAAGTATAAA	TCTTCTATTG
4681	GTCTATGTCA	TGAACAAGAA	TTCAATTCAGT	GGCTCTGTTT	TATAGTAAAC	ATTGCTATTT
4741	TATCATGTCT	GCATTTCTCT	TCTGTCTGAA	TGTCACCACCT	AAAATTTAAC	TCCACAGAAA
4801	GTTTATACTA	CAGTACACAT	GCATATCTTT	GAGCAAAGCA	AACCATACCT	GAAAGTGCAA
4861	TAGAGCAGAA	TATGAATTAC	ATGCGTGTCT	TTCTCCTAGA	CTACATGACC	CCATATAAAT
4921	TACATTCCCT	ATCTATTCTG	CCATCACCAA	AACAAAGGTA	AAAATACTTT	TGAAGATCTA
4981	CTCATAGCAA	GTAGTGTGCA	ACAAACAGAT	ATTTCTCTAC	ATTTATTTTT	AGGGAATAAA
5041	AATAAGAAAT	AAAATAGTCA	GCAAGCCTCT	GCTTCTCAT	ATATCTGTCC	AAACCTAAAG
5101	TTTACTGAAA	TTTGCTCTTT	GAATTTCCAG	TTTTGCAAGC	CTATCAGATT	GTGTTTAAAT
5161	CAGAGGTACT	GAAAAGTATC	AATGAATTCT	AGCTTTCAT	GAACAAAAAT	ATGTAGAGGC
5221	AACTGGCTTC	TGGGACAGTT	TGCTACCCAA	AAGACAACATG	AATGCAAATA	CATAAATAGA
5281	TTTATGAATA	TGGTTTTGAA	CATGCACATG	AGAGGTGGAT	ATAGCAACAG	ACACATTACC
5341	ACAGAATTAC	TTTAAAACCTA	CTTGTTAACA	TTTAAATTGCC	TAAAAACTGC	TCGTAATTTA
5401	CTGTTGTAGC	CTACCATAGA	GTACCTTGCA	TGGTACTATG	TACAGCATTC	CATCCTTACA
5461	TTTTCACTGT	TCTGCTGTTT	GCTCTAGACA	ACTCAGAGTT	CACCATGAAA	ATGCGGTTCT
5521	TGGGGTTGGT	GGTCTGTTTG	GTTCTCTGGA	CCCTGCATTG	CGAGGGGTCC	GGAGGGAAAC
5581	TGACAGCTGT	GGATCCAGAA	ACAAACATGA	ATGTCAGTGA	AATTATCTCT	TACTGGGGAT
5641	TCCCTAGTGA	GGAATACCTA	GTTGAGACAG	AAGATGGATA	TATTCTGTGC	CTTAACCGAA
5701	TTCCTCATGG	GAGGAAGAAC	CATTCTGACA	AAGGTCCCAA	ACCAGTTGTC	TTCTTGCAAC
5761	ATGGCTTGCT	GGCAGATTCT	AGTAACTGGG	TCACAAACCT	TGCCAACAGC	AGCCTGGGCT
5821	TCATTCTTGC	TGATGCTGGT	TTTGACGTGT	GGATGGGCAA	CAGCAGAGGA	AATACCTGGT
5881	CTCGGAAACA	TAAGACACTC	TCAGTTTCTC	AGGATGAATT	CTGGGCTTTC	AGTTATGATG
5941	AGATGGCAAA	ATATGACCTA	CCAGCTTCCA	TTAACTTCAT	TCTGAATAAG	ACTGGCCAAAG
6001	AACAAGTGTA	TTATGTGGGT	CATTCTCAAG	GCACCACTAT	AGGTTTTATA	GCATTTTCAC
6061	AGATCCCTGA	GCTGGCTAAA	AGGATTAAAA	TGTTTTTTGC	CCTGGGTCCT	GTGGCTTCCG
6121	TCGCCTTCTG	TACTAGCCCT	ATGGCCAAAC	TGGGACGACT	GCCAGATCAT	CTCATTAAGG
6181	ACCTCTTTGG	AGACAAAGAA	TTTCTTCCCC	AGAGTGCCTT	TTTGAAGTGG	CTGGGTACCC
6241	ACGTTTGCAC	TCATGTCTAT	CTGAAGGAGC	TCTGTGAAAA	TCTCTGTTTT	CTTCTGTGTG
6301	GATTTAATGA	GAGAAATTTA	AATATGTCTA	GAGTGGATGT	GTATACAACA	CATTCTCCTG
6361	CTGGAACCTC	TGTGCAAAAC	ATGTTACACT	GGAGCCAGGC	TGTTAAATTC	CAAAAGTTTC
6421	AAGCCTTTGA	CTGGGGAAGC	AGTGCCAAGA	ATTATTTTCA	TTACAACCAG	AGTTATCCTC

6481 CCACATACAA TGTGAAGGAC ATGCTTGTGC CGACTGCAGT CTGGAGCGGG GGTCACGACT
6541 GGCTTGCGA TGTCTACGAC GTCAATATCT TACTGACTCA GATCACCAAC TTGGTGTTC
6601 ATGAGAGCAT TCCGGAATGG GAGCATCTTG ACTTCATTTG GGGCCTGGAT GCCCCTTGGA
6661 GGCTTTATAA TAAGATTATT AATCTAATGA GGAAATATCA GTGATTCGAA GCGGCCGCAA
6721 GAAGAAAGCT GAAAACTCT GTCCCTTCCA ACAAGACCCA GAGCACTGTA GTATCAGGGG
6781 TAAAAAGAAA AGTATGTTAT CTGCTGCATC CAGACTTCAT AAAAGCTGGA GCTTAATCTA
6841 GAAAAAAAT CAGAAAGAAA TTACACTGTG AGAACAGGTG CAATTCACCTT TTCCTTTACA
6901 CAGAGTAATA CTGGTAACTC ATGGATGAAG GCTTAAGGGA ATGAAATTGG ACTCACAGTA
6961 CTGAGTCATC ACACTGAAAA ATGCAACCTG ATACATCAGC AGAAGGTTTA TGGGGGAAAA
7021 ATGCAGCCTT CCAATTAAGC CAGATATCTG TATGACCAAG CTGCTCCAGA ATTAGTCACT
7081 CAAAATCTCT CAGATTAAAT TATCAACTGT CACCAACCAT TCCTATGCTG ACAAGGCAAT
7141 TGCTTGTTCT CTGTGTTCCCT GATACTACAA GGCTCTTCCT GACTTCCTAA AGATGCATTA
7201 TAAAAATCTT ATAATTCACA TTTCTCCCTA AACTTTGACT CAATCATGGT ATGTTGGCAA
7261 ATATGGTATA TTACTATTCA AATTGTTTTT CTTGTACCCA TATGTAATGG GTCTTGTGAA
7321 TGTGCTCTTT TGTTCTTTTA ATCATAATAA AAACATGTTT AAGCAAACAC TTTTCACTTG
7381 TAGTATTTGA AGGTACCGGA TCTCGAGCCG CCTTCAATGC CCCCCAAACC AATCCCCAGG
7441 TTTTAACTC TCCCGATTTT CCAAGTACCA TAGCCCGCTG AGAGAGCGCC GCGGTAATGG
7501 GATCCAGGA CCCCAGGGAA TATAAGTCTG AGGGGGACGT AAGCAACCTT TCCTTTTGTA
7561 ACAGGGACAA CATAGCCCCCT ATTTCTTCT TAGAAGGAGA GGTTTTCCCG CAATAGGTCT
7621 TACACGCGGA CGAAATCACC TTTATGACGG CTTCCATGCT TGATCCACCG GCGACCGGA
7681 ATCAGCAGA GCAACCGGAA TCACGCCTGG GGTGGACCGC TCAGTCGTCG GGCTTCCTTC
7741 CCGTCTTCCA ACGACTCTCT GAGTTCCTCG TAGGGTATGT TGGCCCCCTG CAGTAGGGCT
7801 CCGTCCGACG CCACTCAGCT TCTGCCCTCC TAAGCCGAG CCCCCTCTAC TAGGGTCACTC
7861 GTCCGCTCCC CGAATAAGCG AGACGGATGA GGACAGGATC GCCACGCCGC CTGTGGCCGA
7921 CCACTATTCC CTAACGATCA CGTCGGGGTC ACCAAATGAA GCCTTCTGCT TCATGCATGT
7981 GCTCGTAGTC GTCAGGGAAT CAACGGTCCG GCCATCAACC CAGGTGCACA CCAATGTGGT
8041 GAATGGTCAA ATGGCGTTTT TGTATCGAG CTAGGCACCT AAATACAATA TCTCTGCAAT
8101 GCGGAATTCA GTGGTTCGTC CAATCCGTGT TAGACCCGTC TGTTGCCTTC CTAACAAGGC
8161 ACGATCATAC CACGATCATA CCACCTTACT CCCACCAATC GGCATGCACG GTGCTTTTTC
8221 TCTCCTTATA AGGCATGTTG CTAACATCAT GTTACATAAG CATGTTGCAA GACTACAAGA
8281 GTATTGCATA AGACTACATT TCCCCCTCCC TATGCAAAAG CGAACTACT ATATCCTGAG
8341 GGGACTCCTA ACCGCGTACA ACCGAAGCCC CGCTTTTCGC CTAAACATGC TATGTCCCC
8401 TCAGTCAAGC CTTGCCCCGT ACAACCCGAT TCGCAAGCCT TGCCCTCCCC ACATTATCCG
8461 TAGCATTATT TCCTAGCAGT CATCAGAGCT ACAGAAGATA CTCTATGCTG TAGCCAAGTC
8521 TACAAGTTTA CTATTACGCG ACCTCCTATA TTCCGCGTGC CAGCCGATCA ATTCCAATG
8581 CGCGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA
8641 TTCCACACAA CATAAGAGCC GGAAGCATAA AGTGTAAGC CTGGGGTGCC TAATGAGTGA
8701 GCTAACTCAC ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCTG
8761 GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT
8821 CTTCCGCTTC CTCGCTCACT GACTCGTCTG GCTCGGTCTG TCGGCTGCGG CGAGCGGTAT
8881 CAGCTCACTC AAAGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA
8941 ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT
9001 TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAATAA TCGACGCTCA AGTCAGAGGT
9061 GCGGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC
9121 GCTCTCCTGT TCCGACCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA
9181 GCGTGGCGCT TTCTCATAGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTGCTG
9241 CCAAGCTGGG CTGTGTGCAC GAACCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA
9301 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG
9361 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC
9421 CTAACACGCG CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA
9481 CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG
9541 GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCTT
9601 TGATCTTTTT TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA GGGATTTTGG
9661 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAATAA TGAAGTTTTA
9721 AATCAATCTA AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG
9781 AGGCACCTAT CTCAGCGATC TGTCTATTTT GTTCATCCAT AGTTGCCTGA CTCGCCGTCG
9841 TGTAAGTAAC TACGATACGG GAGGGCTTAT CATCTGGCCC CAGTGCTGCA ATGATACCGC
9901 GAGACCCACG CTCACCGGCT CCAGATTAT CAGCAATAAA CCAGCCAGCC GGAAGGGCCG
9961 AGCGCAGAAG TGGTCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT TGTGCGGGG
10021 AAGCTAGAGT AAGTAGTTCC CCAGTTAATA GTTTGCGCAA CGTTGTTGCC ATTGCTACAG
10081 GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT

10141	CAAGGCGAGT	TACATGATCC	CCCATGTTGT	GCAAAAAAGC	GGTTAGCTCC	TTCGGTCCTC
10201	CGATCGTTGT	CAGAAATAAG	TTGGCCGCAG	TGTTATCACT	CATGGTTATG	GCAGCACTGC
10261	ATAATTCTCT	TACTGTCATG	CCATCCGTAA	GATGCTTTTC	TGTGACTGGT	GAGTACTCAA
10321	CCAAGTCATT	CTGAGAATAG	TGTATGCGGC	GACCGAGTTG	CTCTTGCCCG	GCGTCAATAC
10381	GGGATAATAC	CGCGCCACAT	AGCAGAACTT	TAAAAGTGCT	CATCATTGGA	AAACGTTCTT
10441	CGGGGCGAAA	ACTCTCAAGG	ATCTTACCGC	TGTTGAGATC	CAGTTCGATG	TAACCCACTC
10501	GTGCACCCAA	CTGATCTTCA	GCATCTTTTA	CTTTCACCAG	CGTTTCTGGG	TGAGCAAAAA
10561	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	TAAGGGCGAC	ACGGAAATGT	TGAATACTCA
10621	TACTCTTCCT	TTTTCAATAT	TATTGAAGCA	TTTATCAGGG	TTATTGTCTC	ATGAGCGGAT
10681	ACATATTTGA	ATGTATTTAG	AAAAATAAAC	AAATAGGGGT	TCCGCGCACA	TTTCCCCGAA
10741	AAGTGCCACC	TGACGCGCCC	TGTAGCGGCG	CATTAAGCGC	GGCGGGTGTG	GTGGTTACGC
10801	GCAGCGTGAC	CGCTACACTT	GCCAGCGCCC	TAGCGCCCGC	TCCTTTCGCT	TTCTTCCCTT
10861	CCTTTCTCGC	CACGTTCGCC	GG			

(SEQ ID NO: 6)

FIG. 4

DNA sequence of the proviral region (7780 bp)

Feature	Location (bp)
SIN LTR	1-173 (complement strand)
DHSIII enhancer	549-2200
OV promoter	2200-3331
Intron	3379-4967
hLAL CDS	4985-6184
OV 3'UTR	6199-6872
partial gag CDS	6884-7137 (complement strand)
LTR	7435-7780 (complement strand)

```

1   AATGAAGCCT TCTGCTTCAT GCATGIGCTC GTAGTCGTCA GGGAAATCAAC GGTCCGGCCA
61  TCAACCCAGG TGCACACCAA TGTGGTGAAT GGTCAAATGG CGTTTATTGT ATCGAGCTAG
121 GCACTTAAAT ACAATATCTC TGCAATGCGG AATTCAGTGG TTCGTCCAAT CCGTCCCCCT
181 CCCTATGCAA AAGCGAAACT ACTATATCCT GAGGGGACTC CTAACCGCGT ACAACCGAAG
241 CCCCCTTTT CGCCTAAACA TGCTATTGTC CCCTCAGTCA AGCCTTGCCC GTTACAACCC
301 GATTGCGAAG CCTTGCCCTC CCCACATTAT CCGTAGCATT ATTTCTTAGC AGTCATCAGA
361 GCTACAGAAG ATACTCTATG CTGTAGCCAA GTCTACAAGT TTACTATTCA GCGACCTCCT
421 ATATTCCGCG TGCCAGCCGA TCAATTACCA ATCCAACCAG CTATCACACG GAATACAAGA
481 ACTCGCCTAC GCTCTTCTTT CGGGCTGCTT ATAAGCCTCC TGTAAATTTT TTATATTCTT
541 CGCTCGAGTC TCTTCAGAA GGCACAGCAC CGCTGCAGAA AAATGCCAGG TGGACTATGA
601 ACTCACATCC AAAGGAGCTT GACCTGATAC CTGATTTTCT TCAAACAGGG GAAACAACAC
661 AATCCCACAA AACAGCTCAG AGAGAAACCA TCACTGATGG CTACAGCACC AAGGTATGCA
721 ATGGCAATCC ATTCGACATT CATCTGTGAC CTGAGCAAAA TGATTTATCT CTCCATGAAT
781 GGTTGCTTCT TTCCCTCATG AAAAGGCCAAT TTCCACACTC ACAATATGCA ACAAGACAA
841 ACAGAGAACA ATTAATGTGC TCCTTCCTAA TGTTAAATTT TGTAGTGCAA AGAGGAGAAC
901 AAAATCTCAA GTTCTGAGTA GGTTTTAGTG ATTGGATAAG AGGCTTTGAC CTGTGAGCTC
961 ACCTGGACTT CATATCCTTT TGGATAAAAA GTGCTTTTAT AACTTTTCAGG TCCTCGAGTC
1021 TTTATTCATG AGACTGTTGG TTTAGGGACA GACCCACAAT GAAATGCCTG GCATAGGAAA
1081 GGGCAGCAGA GCCTTAGCTG ACCTTTTCTT GGGACAAGCA TTGTCAAACA ATGTGTGACA
1141 AAACCTATTG TACTGCTTTG CACAGCTGTG CTGGGCAGGG CAATCCATTG CCACCTATCC
1201 CAGGTAACCT TCCAAC TGCA AGAAGATTGT TGCTTACTCT CTCTAGACCC CCAAGTCAAA
1261 CCAACTATGC AGGTATGCTG ACAACACTAT GATGACAGCC TGTCTGATC AAGATCTCAT
1321 TTGTTTCATG ACAATTTTTG TTGCTTGCGG CTGGTCTTCC ATTGGGAAAG AGTGTAGTAT
1381 ATCCTTCTCA TCTGACAGAA AAGCAGAAAT TCTCATGCTC CACACTTAAT CTACATTGTT
1441 TTAACCCACC GGCTACTTCT TGGAGAGGAA AAATGGCTTT TATAAGACTC ACAAACAAA
1501 GCTCTGCAAG TCAAATGCAT ACAAACTGTG TCTGTAGGTC TGAATCAGG ACCTATGTG
1561 GAAGTCAAAT AGAGCAGCTT TAAAAAGCCT TTGGGATCAT TCTCATCTTA TATTTGCAGC
1621 ACGATACTAT GACAGTGATA ACTGACATAA CTGCATCAAT TTCCTTGATA TTTTATTTGT
1681 CTTAAAGTAC AAGACATAGA GATGGACGTA AAGATGGACA TATGACTCAG GTCTGGACAG
1741 GTCCGTGGTC CATGTATGAT AAAAGAGATG AAGGGAAGGA GAATTGAGAC TGTCTAAGAA
1801 GGGCTTCAGG GACGTTCTGA AGGCAGATTT GACTGAATCA GATGTACTGT CCAAGTCTCA
1861 TATGTAGCAA TGGAAGGCTG ATATTGGAGA AATATAAAGA AATGGCTGTG AACTCAAAGT
1921 GACCCTGAAC AGAAAAGGGA TATGGAGTTA AAATAATGTC ACAGAACTGA GGTTTATATG
1981 ATATAACCATG GGCTGCAGAG GGTCAGAGTG CTCCACCATG GGCTCTCTT GGGCTGCAGG
2041 GAACTTCTGT TCTACACCTG GAACACCTCC TGCCCTCCTC CGCACTGACC TCAGTGTCTAT
2101 CAGGGCTGTT TCTCTCACAT TTTCTCACTC ACCTCTCCCA ACTACCATTTG TACAGCAGTT
2161 GTTCTTACAT ATTGCTCCTC CTGAGGTACA TCTAGCATCG TTAAGTCCTC AGACTTGGCA
2221 AGGAGAAATG AGATTTCCAC AGTATATATG TTTTACAAA AGGAAGGAGA GAAACAAAAG
2281 AAAATGGCAC TGAATAAATC TCAGCTAGTG GTATAGGAAA GTAATTCTGC TTAACAGAGA
2341 TTGCAGTGAT CTCTATGTAT GTCCTGAAGA ATTATGTTGT ACTTTTTTCC CCCATTTTTA
2401 AATCAAACAG TGCTTTACAG AGGTCAGAAT GGTTTCTTTA CTGTTTGTCA ATTCTATTAT
2461 TTCAATACAG AACAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG
2521 TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCCCTCTGTC ATCTGCCAGG
2581 CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA AGTTCATATC ATAAACACAT
2641 TTGAAATTGA GTATTGTTTT GCATTGTATG GAGCTATGTT TTGCTGTATC CTCAGAATAA
2701 AAGTTTGTTA TAAAGCATTC ACACCCATAA AAAGATAGAT TTAAATATTC CAACTATAGG
2761 AAAGAAAGTG TGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCACGCTTCT
2821 TTATTTCTCC TATTTTGTCA AGAAAATAAT AGGTCAAGTC TTGTCTCAT TTATGTCCTG

```

2881	TCTAGCGTGG	CTCAGATGCA	CATTGTACAT	ACAAGAAGGA	TCAAATGAAA	CAGACTTCTG
2941	GTCTGTTACT	ACAACCATAG	TAATAAGCAC	ACTAACTAAT	AATTGCTAAT	TATGTTTTCC
3001	ATCTCCAAGG	TTCCCACATT	TTTCTGTTTT	CTTAAAGATC	CCATTATCTG	GTGTAAGTGG
3061	AAGCTCAATG	GAACATGAGC	AATATTTCCT	AGTCTTCTCT	CCCATCCAAC	AGTCTGTATG
3121	GATTAGCAGA	ACAGGCAGAA	AACACATTGT	TACCCAGAAT	TAAAAACTAA	TATTTGCTCT
3181	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAATCTAA	CCCAATCCCA	TTAAATGATT
3241	TCTATGGTGT	CAAAGGTCAA	ACTTCTGAAG	GGAACTCTGT	GGTGGGTGAC	AATTCAGACT
3301	ATATATTCCC	CAGGGCTCAG	CCAGTGTCTG	TACATACAGC	TAGAAAGCTG	TATTGCCTTT
3361	AGCAGTCAAG	CTCGAAAGGT	AAGCAACTCT	CTGGAATTAC	CTTCTCTCTA	TATTAGCTCT
3421	TACTTGACCC	TAAACTTTAA	AAAATTAACA	ATTATTGTGC	TATGTGTGTG	ATCTTTAAGG
3481	GTGAAGTACC	TGCGTGATAC	CCCCATATAA	AACTTCTCAC	CTGTGTATGC	ATTCTGCACT
3541	ATTTTATTAT	GTGTAAAAGC	TTTGTGTTTG	TTTTCAGGAG	GCTTATTCTT	TGTGCTTAAA
3601	ATATGTTTTT	AATTCAGAA	CATCTTATCC	TGTCGTTTAC	TATCTGATAT	GCTTTGCACT
3661	TTGCTTGATT	AACTTCTAGC	CCTACAGAGT	GCACAGAGAG	CAAAATCATG	GTGTTTCAGT
3721	AAATCTGGGG	AGTTATTTTA	ATGTGAAAAT	TCTCTAGAAG	TTAATTTCCT	GCAAAAGTGA
3781	GCTGCTGATC	ACTACACAAG	ATAAAAATGT	GGGGGGTGCA	TAAACGTATA	TTCTTACAAT
3841	AATAGATACA	TGTGAACCTA	TATACAGAAA	AGAAAATGAG	AAAAATGTGT	GTGTGTATAC
3901	TCACACACGT	GGTCAGTAAA	AACTTTTGAG	GGGTTTAATA	CAGAAAATCC	AATCCTGAGG
3961	CCCCAGCACT	CAGTACGCAT	ATAAAGGGCT	GGGCTCTGAA	GGACTTCTGA	CTTTCACAGA
4021	TTATATAAAT	CTCAGGAAAG	CAACTAGATT	CATGCTGGCT	CCAAAAGCTG	TGCTTTTATAT
4081	AAGCACACTG	GCTATACAAT	AGTTGTACAG	TTCAGCTCTT	TATAATAGAA	ACAGACAGAA
4141	CAAGTATAAA	TCTTCTATTG	GTCTATGTCA	TGAACAAGAA	TTCATTCACT	GGCTCTGTTT
4201	TATAGTAAAC	ATTGCTATTT	TATCATGTCT	GCATTTCTCT	TCTGTCTGAA	TGTCACCACT
4261	AAAATTTAAC	TCCACAGAAA	GTTTATACTA	CAGTACACAT	GCATATCTTT	GAGCAAAGCA
4321	AACCATACCT	GAAAGTGCAA	TAGAGCAGAA	TATGAATTAC	ATGCGTGTCT	TTCTCCTAGA
4381	CTACATGACC	CCATATAAAT	TACATTCCTT	ATCTATTCTG	CCATCACCAA	AACAAGGTA
4441	AAAATCTTTT	TGAAGATCTA	CTCATAGCAA	GTAGTGTGCA	ACAAACAGAT	ATTTCTCTAC
4501	ATTTATTTTT	AGGGAATAAA	AATAAGAAAT	AAAATAGTCA	GCAAGCCTCT	GCTTTCTCAT
4561	ATATCTGTCC	AAACCTAAAG	TTTACTGAAA	TTTGCTCTTT	GAATTTCCAG	TTTTGCAAGC
4621	CTATCAGATT	GTGTTTAAAT	CAGAGGTACT	GAAAAGTATC	AATGAATTCT	AGCTTTCACT
4681	GACAAAAAAT	ATGTAGAGGC	AACTGGCTTC	TGGGACAGTT	TGCTACCCAA	AAGACAAGCT
4741	AATGCAATA	CATAAATAGA	TTTATGAATA	TGGTTTTGAA	CATGCACATG	AGAGGTGGAT
4801	ATAGCAACAG	ACACATTACC	ACAGAATTAC	TTTAAACTA	CTTGTTAACA	TTTAATTGCC
4861	TAAAAACTGC	TCGTAATTTA	CTGTTGTAGC	CTACCATAGA	GTACCCTGCA	TGGTACTATG
4921	TACAGCATTC	CATCCTTACA	TTTTCACTGT	TCTGCTGTTT	GCTCTAGACA	ACTCAGAGTT
4981	CACCATGAAA	ATGCGGTTCT	TGGGGTTGGT	GGTCTGTTTG	GTTCTCTGGA	CCCTGCATTG
5041	CGAGGGGTCC	GGAGGGAAAC	TGACAGCTGT	GGATCCAGAA	ACAAACATGA	ATGTCAGTGA
5101	AATTATCTCT	TACTGGGGAT	TCCCTAGTGA	GGAAATACCTA	GTTGAGACAG	AAGATGGATA
5161	TATTTCTGTG	CTTAACCGAA	TTCTCATAGG	GAGGAAGAAC	CATTCTGACA	AAGTCCCCAA
5221	ACCAGTTGTC	TTCTTGCAAC	ATGGCTTGCT	GGCAGATTCT	AGTAACTGGG	TCACAAACCT
5281	TGCCAACAGC	AGCCTGGGCT	TCATTTCTGC	TGATGCTGGT	TTTGACGTGT	GGATGGGCAA
5341	CAGCAGAGGA	AATACCTGGT	CTCGGAAACA	TAAGACACTC	TCAGTTTCTC	AGGATGAATT
5401	CTGGGCTTTC	AGTTATGATG	AGATGGCAAA	ATATGACCTA	CCAGCTTCCA	TTAACTTCAT
5461	TCTGAATAAG	ACTGGCCAAG	AACAAGTGTA	TTATGTGGGT	CATTCTCAAG	GCACCACTAT
5521	AGGTTTTATA	GCATTTTCAC	AGATCCCTGA	GCTGGCTAAA	AGGATTAATA	TGTTTTTTGC
5581	CCTGGGTCCCT	GTGGCTTCCG	TGCGCTTCTG	TACTAGCCCT	ATGGCCAAAC	TGGGACGACT
5641	GCCAGATCAT	CTCATTAAGG	ACCTCTTTGG	AGACAAAGAA	TTTCTTCCCC	AGAGTGCGTT
5701	TTTGAAGTGG	CTGGGTACCC	ACGTTTGCAC	TCATGTCATA	CTGAAGGAGC	TCTGTGGAAA
5761	TCTCTGTTTT	CTTCTGTGTG	GATTTAATGA	GAGAAATTTA	AATATGTCTA	GAGTGGATGT
5821	GTATACAACA	CATTCTCCTG	CTGGAACCTC	TGTGCAAAAC	ATGTTACACT	GGAGCCAGGC
5881	TGTTAAATTC	CAAAAGTTTC	AAGCCTTTGA	CTGGGGAAGC	AGTGCCAAGA	ATTATTTTCA
5941	TTACAACCAG	AGTTATCCTC	CCACATACAA	TGTGAAGGAC	ATGCTTGTGC	CGACTTCAGT
6001	CTGGAGCGGG	GGTCACGACT	GGCTTGACAG	TGTCTACGAC	GTCAATATCT	TACTGACTCA
6061	GATCACCAAC	TTGGTGTTC	ATGAGAGCAT	TCCGGAATGG	GAGCATCTTG	ACTTCATTTG
6121	GGGCTTGGAT	GCCCCTTGGG	GGCTTTATAA	TAAGATTATT	AATCTAATGA	GGAAATATCA
6181	GTGATTCGAA	GCGGCCGCAA	GAAGAAAGCT	GAAAAACTCT	GTCCCTTCCA	ACAAGACCCA
6241	GAGCACTGTA	GTATCAGGGG	TAAAATGAAA	AGTATGTTAT	CTGCTGCATC	CAGACTTCAT
6301	AAAAGCTGGA	GCTTAATCTA	GAAAAAGAAA	TACACTGTG	TTACACTGTG	AGAAGAGGTG
6361	CAATTTCATT	TTCTTTTACA	CAGAGTAAAT	CTGGTAACTC	ATGGATGAAG	GCTTAAGGGA
6421	ATGAAATTGG	ACTCACAGTA	CTGAGTCATC	ACACTGAAAA	ATGCAACCTG	ATACATCAGC
6481	AGAAGGTTTA	TGGGGGAAAA	ATGCAGCCTT	CCAATTAAGC	CAGATATCTG	TATGACCAAG

6541	CTGCTCCAGA	ATTAGTCACT	CAAAATCTCT	CAGATTAAAT	TATCAACTGT	CACCAACCAT
6601	TCCTATGCTG	ACAAGGCAAT	TGCTTGTTCT	CTGTGTTCCCT	GATACTACAA	GGCTCTTCCT
6661	GACTTCCTAA	AGATGCATTA	TAAAAATCTT	ATAATTCACA	TTTCTCCCTA	AACTTTGACT
6721	CAATCATGGT	ATGTTGGCAA	ATATGGTATA	TTACTATTCA	AATTGTTTTTC	CTTGTACCCA
6781	TATGTAATGG	GTCTTGTA	TGTGCTCTTT	TGTTCCCTTA	ATCATAATAA	AAACATGTTT
6841	AAGCAAACAC	TTTTCACTG	TAGTATTTGA	AGGTACCGGA	TCTCGAGCCG	CCTTCAATGC
6901	CCCCAAAACC	AATCCCCAGG	TTTTTAACTC	TCCCCGATTTT	CCAAGTACCA	TAGCCCCGCTG
6961	AGAGAGCGCC	GCGGTAATGG	GATCCCAGGA	CCCCGGGGAA	TATAAGTCTG	AGGGGGACGT
7021	AAGCAACCTT	TCCTTTTGTA	ACAGGGACAA	CATAGCCCCT	ATTTCTTCTT	TAGAAGGAGA
7081	GGTTTTCCCG	CAATAGGTCT	TACACGCGGA	CGAAATCACC	TTTATGACGG	CTTCCATGCT
7141	TGATCCACCG	GGCGACCGGA	ATCACGCAGA	GCAACCGGAA	TCACGCCTGG	GGTGGACCGC
7201	TCAGTCGTCG	GGCTTCCTTC	CCGTCTTCCA	ACGACTCTCT	GAGTTCTCGG	TAGGGTATGT
7261	TGGCCCCCTG	CAGTAGGGCT	CCCTCCGACG	CCACTCAGCT	TCTGCCCTCC	TAAGCCGCAG
7321	CCCCCTCTAC	TAGGGTCATC	GTCCGCTCCC	CGAATAAGCG	AGACGGATGA	GGACAGGATC
7381	GCCACGCCGC	CTGTGGCCGA	CCACTATTCC	CTAACGATCA	CGTCGGGGTC	ACCAAATGAA
7441	GCCTTCTGCT	TCATGCATGT	GCTCGTAGTC	GTCAGGGAAT	CAACGGTCCG	GCCATCAACC
7501	CAGGTGCACA	CCAATGTGGT	GAATGGTCAA	ATGGCGTTTA	TTGTATCGAG	CTAGGCACTT
7561	AAATACAATA	TCTCTGCAAT	GCGGAATTCA	GTGGTTTCGTC	CAATCCGTGT	TAGACCCGTC
7621	TGTTGCCTTC	CTAACAAGGC	ACGATCATAC	CACGATCATA	CCACCTTACT	CCCACCAATC
7681	GGCATGCACG	GTGCTTTTTC	TCTCCTTATA	AGGCATGTTG	CTAACTCATC	GTTACATAAG
7741	CATGTTGCAA	GACTACAAGA	GTATTGCATA	AGACTACATT		

(SEQ ID NO:7)

FIG. 5

DNA Sequence of pALVIN-OV-1.1-I (10762 bp)

Feature	Location (bp)
Stuffer	1-2734
OV 3'UTR	2750-3423
partial gag CDS	3435-3688 (complement strand)
LTR	3986-4331 (complement strand)
Ampicillin resistance gene	5795-6652 (complement strand)
SIN LTR	7434-7606 (complement strand)
OV promoter	7975-9106
Intron	9154-10742
ATG of ovalbumin or POI	10760-10762

1	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
61	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
121	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
181	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
241	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
301	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
361	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
421	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
481	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
541	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
601	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
661	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
721	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
781	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
841	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
901	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
961	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1021	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1081	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1141	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1201	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1261	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1321	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1381	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1441	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1501	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1561	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1621	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1681	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1741	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1801	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1861	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1921	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
1981	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2041	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2101	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2161	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2221	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2281	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2341	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2401	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2461	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2521	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2581	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2641	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2701	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn
2761	TGAAAAACTC	TGTCCCTTCC	AACAAGACCC	AGAGCACTGT	AGTATCAGGG	GTAAAAATGAA

2821	AAGTATGTTA	TCTGCTGCAT	CCAGACTTCA	TAAAAGCTGG	AGCTTAATCT	AGAAAAAAAA
2881	TCAGAAAGAA	ATTACACTGT	GAGAACAGGT	GCAATTCACT	TTTCCTTTAC	ACAGAGTAAT
2941	ACTGGTAACT	CATGGATGAA	GGCTTAAGGG	AATGAAATTG	GACTCACAGT	ACTGAGTCAT
3001	CACACTGAAA	AATGCAACCT	GATACATCAG	CAGAAGGTTT	ATGGGGGAAA	AATGCAGCCT
3061	TCCAATTAAG	CCAGATATCT	GTATGACCAA	GCTGCTCCAG	AATTAGTCAC	TCAAAATCTC
3121	TCAGATTAAA	TTATCAACTG	TCACCAACCA	TTCCTATGCT	GACAAGGCAA	TTGCTTGTTT
3181	TCTGTGTTCC	TGATACTACA	AGGCTCTTCC	TGACTTCCTA	AAGATGCATT	ATAAAAAATCT
3241	TATAATTAC	ATTTCTCCCT	AAACTTTGAC	TCAATCATGG	TATGTTGGCA	AATATGGTAT
3301	ATTACTATTC	AAATTGTTTT	CCTTGTACCC	ATATGTAATG	GGTCTTGTTA	ATGTGCTCTT
3361	TTGTTCCTTT	AATCATAATA	AAAACATGTT	TAAGCAAACA	CTTTTCACTT	GTAGTATTTG
3421	AAGGTACCGG	ATCTCGAGCC	GCCTTCAATG	CCCCCAAAAC	CAATCCCCAG	GTTTTTAACT
3481	CTCCCGATTT	TCCAAGTACC	ATAGCCCGCT	GAGAGAGCGC	CGCGGTAATG	GGATCCCAGG
3541	ACCCCGGGGA	ATATAAGTCT	GAGGGGGACG	TAAGCAACCC	TTCTTTTGT	AACAGGGACA
3601	ACATAGCCCC	TATTTCTTTC	TTAGAAGGAG	AGGTTTTCCC	GCAATAGGTC	TTACACGCGG
3661	ACGAAATCAC	CTTTATGACG	GCTTCCATGC	TTGATCCACC	GGGCGACCGG	AATCACGCAG
3721	AGCAACCGGA	ATCACGCCTG	GGGTGGACCG	CTCAGTCGTC	GGGCTTCCTT	CCCGTCTTCC
3781	AACGACTCTC	TGAGTTCTCG	GTAGGGTATG	TTGGCCCCCT	GCAGTAGGGC	TCCCTCCGAC
3841	GCCACTCAGC	TTCTGCCCTC	CTAAGCCGCA	GCCCCCTCTA	CTAGGGTCAT	CGTCCGCTCC
3901	CCGAATAAGC	GAGACGGATG	AGGACAGGAT	CGCCACGCCG	CCTGTGGCCG	ACCACATATC
3961	CCTAACGATC	ACGTCGGGGT	CACCAAATGA	AGCCTTCTGC	TTCATGCATG	TGCTCGTAGT
4021	CGTCAGGGAA	TCAACGGTCC	GCCCATCAAC	CCAGGTGCAC	ACCAATGTGG	TGAATGGTCA
4081	AATGGCGTTT	ATTGTATCGA	GCTAGGCACT	TAAATACAAT	ATCTCTGCAA	TGCGGAATTC
4141	AGTGGTTCGT	CCAATCCGTG	TTAGACCCGT	CTGTTGCCTT	CCTAACCAAG	CACGATCATA
4201	CCAGATCAT	ACCACCTTAC	TCCCACCAAT	CGGCATGCAC	GGTGCTTTTT	CTCTCCTTAT
4261	AAGGCATGTT	GCTAACTCAT	CGTTACATAA	GCATGTTGCA	AGACTACAAG	AGTATTGCAT
4321	AAGACTACAT	TTCCCCCTCC	CTATGCAAAA	GCGAAACTAC	TATATCCTGA	GGGGACTCCT
4381	AACCGCGTAC	AACCGAAGCC	CCGCTTTTCG	CCTAAACATG	CTATTGTCCC	CTCAGTCAAG
4441	CCTTGCCCGT	TACAACCCGA	TTCGCAAGCC	TTGCCCTCCC	CACATTATCC	GTAGCATTTAT
4501	TTCTTAGCAG	TCATCAGAGC	TACAGAAGAT	ACTCTATGCT	GTAGCCAAGT	CTACAAGTTT
4561	ACTATTCAGC	GACCTCCTAT	ATTCCGCGTG	CCAGCCGATC	AATTACCAAT	GCGCGCTTGG
4621	CGTAATCATG	GTCATAGCTG	TTTCCTGTGT	GAAATTGTTA	TCCGCTCACA	ATTCCACACA
4681	ACATACGAGC	CGGAAGCATA	AAGTGTAAG	CCTGGGGTGC	CTAATGAGTG	AGCTAACTCA
4741	CATTAATTGC	GTTGCGCTCA	CTGCCCGCTT	TCCAGTCGGG	AAACCTGTGC	TGCCAGCTGC
4801	ATTAATGAAT	CGGCCAACGC	GCGGGGAGAG	GCGGTTTGCG	TATTGGGCGC	TCTTCCGCTT
4861	CCTCGCTCAC	TGACTCGCTG	CGCTCGGTGC	TTCGGCTGCG	GCGAGCGGTA	TCAGCTCACT
4921	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA	CGCAGGAAAG	AACATGTGAG
4981	CAAAGGCCA	GCAAAAGGCC	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA
5041	GGCTCCGCCC	CCCTGACGAG	CATCACAAAA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC
5101	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG
5161	TTCCGACCTT	GCCGCTTACC	GGATACCTGT	CCGCTTTTCT	CCCTTCGGGA	AGCTCTGGCG
5221	TTTCTCATAG	CTCACGCTGT	AGGTATCTCA	GTTGCGGTGA	GGTCGTTCGC	TCCAAGCTGG
5281	GCTGTGTGCA	CGAACCCCCC	GTTTACGCCC	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC
5341	TTGAGTCCAA	CCCGGTAAAG	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA
5401	TTAGCAGAGC	GAGGTATGTA	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG
5461	GCTACACTAG	AAGGACAGTA	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA
5521	AAAGAGTTGG	TAGCTCTTGA	TCCGGCAAAC	AAACCACCGC	TGGTAGCGGT	GGTTTTTTTG
5581	TTTGCAAGCA	GCAGATTACG	CGCAGAAAAA	AAGGATCTCA	AGAAGATCCT	TTGATCTTTT
5641	CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA	AGGGATTTTG	GTCATGAGAT
5701	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAAAA	ATGAAGTTTT	AAATCAATCT
5761	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	CTTAATCAGT	GAGGCACCTA
5821	TCTCAGCGAT	CTGTCTATTT	CGTTCATCCA	TAGTTGCCTG	ACTCCCCGTC	GTGTAGATAA
5881	CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	AATGATACCG	CGAGACCCAC
5941	GCTCACCGGC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	CGGAAGGGCC	GCGAGCAGAA
6001	GTGGTCTGTC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	TTGTTGCCGG	GAAGCTGAGAG
6061	TAAGTAGTTC	GCCAGTTAAT	AGTTTGCAGC	ACGTTGTTGC	CATTGCTACA	GGCATCGTGG
6121	TGTCACGCTC	GTCGTTTGGT	ATGGCTTCAT	TCAGCTCCGG	TTCCCAACGA	TCAAGGCGAG
6181	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGCTC	CTTCGGTCTT	CCGATCGTTG
6241	TCAGAAGTAA	GTTGGCCGCA	GTGTTATCAC	TCATGGTTAT	GGCAGCACTG	CATAATTCTC
6301	TTACTGTCAT	GCCATCCGTA	AGATGCTTTT	CTGTGACTGG	TGAGTACTCA	ACCAAGTCAT
6361	TCTGAGAATA	GTGTATGCGG	CGACCGAGTT	GCTCTTGCCC	GGCGTCAATA	CGGGATAATA
6421	CCGCGCCACA	TAGCAGAACT	TTAAAAGTGC	TCATCATTTG	AAAACGTTCT	TCGGGGCGAA

6481	AACTCTCAAG	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	GTAACCCACT	CGTGCACCCA
6541	ACTGATCTTC	AGCATCTTTT	ACTTTACCCA	CGCTTTCTGG	GTGAGCAAAA	ACAGGAAGGC
6601	AAAATGCCGC	AAAAAAGGGA	ATAAGGGCGA	CACGGAAATG	TTGAATACTC	ATACTCTTCC
6661	TTTTTCAATA	TTATTGAAGC	ATTTATCAGG	GTTATTGTCT	CATGAGCGGA	TACATATTTG
6721	AATGTATTTA	GAAAAATAAA	CAAATAGGGG	TTCCGCGCAC	ATTTCCCCGA	AAAGTGCCAC
6781	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	CGCAGCGTGA
6841	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCC	CTCCTTTCGC	TTTCTTCCCT	TCCTTTCTCG
6901	CCACGTTCCG	CGGCTTTCCC	CGTCAAGCTC	TAAATCGGGG	GCTCCCTTTA	GGGTTCCGAT
6961	TTAGTGCTTT	ACGGCACCTC	GACCCCAAAA	AACTTGATTA	GGGTGATGGT	TCACGTAGTG
7021	GGCCATCGCC	CTGATAGACG	GTTTTTCGCC	CTTTGACGTT	GGAGTCCACG	TTCTTTAATA
7081	GTGGACTCTT	GTTCCAACT	GGAACAACAC	TCAACCCAT	CTCGGTCTAT	TCTTTTGATT
7141	TATAAGGGAT	TTTGCCGATT	TCGGCCTATT	GGTTAAAAAA	TGAGCTGATT	TAACAAAAAT
7201	TTAACGCGAA	TTTTAACAAA	ATATTAACGC	TTACAATTTT	CATTGCCCAT	TCAGGCTGCG
7261	CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG
7321	GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG
7381	TAAAACGACG	GCCAGTGAGC	GCGTATTCCC	TAACGATCAC	GTCGGGGTCA	CCAAATGAAG
7441	CCTTCTGCTT	CATGCATGTG	CTCGTAGTCG	TCAGGGAATC	AACGGTCCGG	CCATCAACCC
7501	AGGTGCACAC	CAATGTGGTG	AATGGTCAAA	TGGCGTTTAT	TGTATCGAGC	TAGGCACTTA
7561	AATACAATAT	CTCTGCAATG	CGGAATTCAG	TGGTTCGTCC	AATCCGTCCC	CCTCCCTATG
7621	CAAAGCGGAA	ACTACTATAT	CCTGAGGGGA	CTCCTAACCG	CGTACAACCG	AAGCCCCGCT
7681	TTTCGCCCTAA	ACATGCTATT	GTCCCCCTCAG	TCAAGCCTTG	CCCGTTACAA	CCCGATTTCG
7741	AAGCCTTGCC	CTCCCCACAT	TATCCGTAGC	ATTATTTCCCT	AGCAGTCATC	AGAGCTACAG
7801	AGGATACTCT	ATGCTGTAGC	CAAGTCTACA	AGTTTACTAT	TCAGCGACCT	CCTATATTCC
7861	CGGTGCCAGC	CGATCAATT	CCAATCCAAC	CAGCTATCAC	ACGGAATACA	AAGACTCGCC
7921	TACGCTCTTC	TTTCGGGCTG	CTTATAAGCC	TCCTGTAATT	TTTTTATATT	CCTCGTTAAG
7981	TCCTCAGACT	TGGCAAGGAG	AATGTAGATT	TCCACAGTAT	ATATGTTTTT	ACAAAAGGAA
8041	GGAGAGAAAC	AAAAGAAAA	GGCACTGACT	AACTTTCAGC	TAGTGGTATA	GGAAAGTAAT
8101	TCTGCTTAAC	AGAGATTGCA	GTGATCTCTA	TGTATGTCCT	GAAGAATTAT	GTTGTACTTT
8161	TTTCCCCCAT	TTTTAAATCA	AACAGTGCTT	TACAGAGGTC	AGAATGGTTT	CTTTACTGTT
8221	TGTCAATTCT	ATTATTTCAA	TACAGAACAA	TAGCTTCTAT	AACGTAAATA	TATTTGCTAT
8281	TGTATATTAT	GATTGTCCCT	CGAACCATGA	ACACTCCTCC	AGCTGAATTT	CACAATTCCT
8341	CTGTCATCTG	CCAGGCCATT	AAGTTATTCA	TGGAAGATCT	TTGAGGAACA	CTGCAAGTTC
8401	ATATCATATA	CACATTTGAA	ATTGAGTATT	GTTTTCGATT	GTATGGAGCT	ATGTTTTGCT
8461	GTATCCTCAG	AATAAAAGTT	TGTTATAAAG	CATTACACCC	CATAAAAAAG	TAGATTTAAA
8521	TATTCCAAC	ATAGGAAAGA	AAGTGCTGCT	GCTCTTCACT	CTAGTCTCAG	TTGGCTCCTT
8581	CACATGCACG	CTTCTTTATT	TCTCCTATTT	TGTCAAGAAA	ATAATAGGTC	AAGTCTTGTT
8641	CTCATTATAT	TCCTGTCTAG	CGTGGCTCAG	ATGCACATTG	TACATACAAG	AAGGATCAAA
8701	TGAAACAGAC	TTCTGGTCTG	TTACTACAAC	CATAGTAATA	AGCACACTAA	CTAATAATTG
8761	CTAATTATGT	TTTCCATCTC	CAAGGTTCCC	ACATTTTTCT	GTTTTCTTAA	AGATCCCATT
8821	ATCTGGTTGT	AACTGAAGCT	CAATGGAACA	TGAGCAATAT	TTCCCAGTCT	TCTCTCCCAT
8881	CCAACAGTCC	TGATGGATTA	GCAGAACAGG	CAGAAAACAC	ATTGTTACCC	AGAATTTAAA
8941	ACTAATATTT	GCTCTCCATT	CAATCCAAAA	TGGACCTATT	GAAACTAAAA	TCTAACCCAA
9001	TCCCATTAAT	TGATTTCTAT	GGTGTCAAAG	GTCAAACCTT	TGAAGGGAAC	CTGTGGGTGG
9061	GTCACAATTC	AGACTATATA	TTCCCCAGGG	CTCAGCCAGT	GTCTGTACAT	ACAGCTAGAA
9121	AGCTGTATTG	CCTTTAGCAG	TCAAGCTCGA	AAGGTAAGCA	ACTCTCTGGA	ATTACCTTCT
9181	CTCTATATTA	GCTCTTACTT	GCACCTAAAC	TTTAAAAAAT	TAACAATTAT	TGTGCTATGT
9241	GTTGTATCTT	TAAGGGTGAA	GTACCTGCGT	GATACCCCTT	ATAAAAACTT	CTCACCTGTG
9301	TATGCATTCT	GCACATTTTT	ATTATGTGTA	AAAGCTTTGT	GTTTGTTTTC	AGGAGGCTTA
9361	TTCTTTGTGC	TTAAAAATAT	TTTTTAATTT	CAGAACATCT	TATCCTGTCT	TTCACTATCT
9421	GATATGCTTT	GCAGTTTGCT	TGATTAACCT	CTAGCCCTAC	AGAGTGCACA	GAGAGCAAAA
9481	TCATGGTGTT	CAGTGAATTC	TGGGGAGTTA	TTTTAATGTG	AAAATTCTCT	AGAAGTTTAA
9541	TTCTTGCAAA	GTGCAGCTGC	TGATCACTAC	ACAAGATAAA	AATGTGGGGG	GTGCATAAAC
9601	GTATATTCTT	ACAATAATAG	ATACATGTGA	ACTTATATAC	AGAAAAAGAA	ATGAGAAAAA
9661	TGTGTGTGTG	TATACTCACA	CACGTGGTCA	GTAAAAACTT	TTGAGGGGTT	TAATACAGAA
9721	AATCCAATCC	TGAGGCCCCCA	GCACCTAGTA	CGCATATAAA	GGGCTGGGCT	CTGAAGGACT
9781	TCTGACTTTC	ACAGATTATA	TAAATCTCAG	GAAAGCAACT	AGATTCATGC	TGGCTCCAAA
9841	AGCTGTGCTT	TATATAAGCA	CACTGGCTAT	ACAATAGTTG	TACAGTTCAG	CTCTTTATAA
9901	TAGAAACAGA	CAGAACAAGT	ATAAATCTTC	TATTGGTCTA	TGTCATGAAC	AAGAATTCAT
9961	CTAGTGGCTC	TGTTTTATAG	TAAACATTGC	TATTTTATCA	TGTCTGCATT	TCTCTTCTGT
10021	TCGAATGTCA	CCACTAAAA	TTAACCCTAC	AGAAAGTTTA	TACTACAGTA	CACATGCATA
10081	TCTTTGAGCA	AAGCAAACCA	TACCTGAAAG	TGCAATAGAG	CAGAATATGA	ATTACATGCG

10141	TGTCTTTCTC	CTAGACTACA	TGACCCCAT	TAAATTACAT	TCCTTATCTA	TTCTGCCATC
10201	ACCAAAACAA	AGGTAAAAAT	ACTTTTGAAG	ATCTACTCAT	AGCAAGTAGT	GTGCAACAAA
10261	CAGATATTTC	TCTACATTTA	TTTTTAGGGA	ATAAAAATAA	GAAATAAAAT	AGTCAGCAAG
10321	CCTCTGCTTT	CTCATATATC	TGTCCAAACC	TAAAGTTTAC	TGAAATTTGC	TCTTTGAATT
10381	TCCAGTTTTG	CAAGCCTATC	AGATTGTGTT	TTAATCAGAG	GTAATGAAAA	GTATCAATGA
10441	ATTCTAGCTT	TCACTGAACA	AAAATATGTA	GAGGCAACTG	GCTTCTGGGA	CAGTTTGCTA
10501	CCCAAAAGAC	AACTGAATGC	AAATACATAA	ATAGATTTAT	GAATATGGTT	TTGAACATGC
10561	ACATGAGAGG	TGGATATAGC	AACAGACACA	TTACCACAGA	ATTACTTTAA	AACTACTTGT
10621	TAACATTTAA	TTGCCTAAAA	ACTGCTCGTA	ATTTACTGTT	GTAGCCTACC	ATAGAGTACC
10681	CTGCATGGTA	CTATGTACAG	CATTCCATCC	TTACATTTTC	ACTGTTCTGC	TGTTTGCTCT
10741	AGACAACTCA	GAGTTCACCA	TG			

(SEQ ID NO:8)

FIG. 6

DNA sequence of SBC102 adaptor (242 bp)

Feature		Location (bp)				
From OV gene		1-149				
Partial hLAL CDS		150-242				
1	CCCGGGTTGT	TAACATTTAA	TTGCCTAAAA	ACTGCTCGTA	ATTTACTGTT	GTAGCCTACC
61	ATAGAGTACC	CTGCATGGTA	CTATGTACAG	CATTCCATCC	TTACATTTTC	ACTGTTCTGC
121	TGTTTGCTCT	AGACAACTCA	GAGTTCACCA	TGAAAATGCG	GTTCTTGGGG	TTGGTGGTCT
181	GTTTGGTTCT	CTGGACCCTG	CATTCCGAGG	GGTCCGGAGG	GAAACTGACA	GCTGTGGATC
241	CT					

(SEQ ID NO:9)

FIG. 7

DNA sequence of Syn SBC102 (1575 bp)

Feature		Location (bp)				
Partial OV promoter		1-355				
hLAL CDS		356-1575				
1	CCATTATCTG	GTTGTAAGTG	AAGCTCAATG	GAACATGAGC	AATATTTCCC	AGTCTTCTCT
61	CCCATCCAAC	AGTCCTGATG	GATTAGCAGA	ACAGGCAGAA	AACACATTGT	TACCCAGAAT
121	TAAAAACTAA	TATTGCTCT	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAACTTAA
181	CCCAATCCCA	TTAAATGATT	TCTATGGTGT	CAAAGGTCAA	ACTTCTGAAG	GGAACCTGTG
241	GGTGGGTCAC	AATTCAGACT	ATATATTTCC	CAGGGCTCAG	CCAGTGTCTG	TACATACAGC
301	TAGAAAGCTG	TATTGCCTTT	AGCAGTCAAG	CTCGAAAGAC	AACTCAGAGT	TCACCATGAA
361	AATGCGGTTC	TTGGGGTTGG	TGGTCTGTGT	GGTTCTCTGG	ACCCTGCATT	CCGAGGGGTC
421	CGGAGGGA	CTGACAGCTG	TGGATCCTGA	AACAAACATG	AATGTCAGTG	AAATTATCTC
481	TTACTGGGGA	TTCCCTAGTG	AGGAATACCT	AGTTGAGACA	GAAGATGGAT	ATATTCTGTG
541	CCTTAACCGA	ATTCTCATG	GGAGGAAGAA	CCATTCTGAC	AAAGGTCCCA	AACCAGTTGT
601	CTTCCTGCAA	CATGGCTTGC	TGGCAGATTC	TAGTAAGTGG	GTCACAAACC	TTGCCAACAG
661	CAGCCTGGGC	TTCATTCTTG	CTGATGCTGG	TTTTGACGTG	TGGATGGGCA	ACAGCAGAGG
721	AAATACCTGG	TCTCGGAAAC	ATAAGACACT	CTCAGTTTCT	CAGGATGAAT	TCTGGGCTTT
781	CAGTTATGAT	GAGATGGCAA	AATATGACCT	ACCAGCTTCC	ATTAACCTCA	TTCTGAATAA
841	GACTGGCCAA	GAACAAGTGT	ATTATGTGGG	TCATTCTCAA	GGCACCCTAA	TAGGTTTTAT
901	AGCATTTTCA	CAGATCCCTG	AGCTGGCTAA	AAGGATTAAA	ATGTTTTTTG	CCCTGGGTCC
961	TGTGGCTTCC	GTCGCCTTCT	GTACTAGCCC	TATGGCCAAA	CTGGGACGAC	TGCCAGATCA
1021	TCTCATTAAG	GACCTCTTTG	GAGACAAAGA	ATTTCTTCCC	CAGAGTGCCT	TTTTGAAGTG
1081	GCTGGGTACC	CACGTTTGCA	CTCATGTCAT	ACTGAAGGAG	CTCTGTGGAA	ATCTCTGTTT
1141	TCTTCTGTGT	GGATTTAATG	AGAGAAATTT	AAATATGTCT	AGAGTGGATG	TGTATACAAC
1201	ACATTCTCCT	GCTGGAACCT	CTGTGCAAAA	CATGTTACAC	TGGAGCCAGG	CTGTTAAATT
1261	CCAAAAGTTT	CAAGCCTTTG	ACTGGGGAAG	CAGTGCCAAG	AATTATTTTC	ATTACAACCA
1321	GAGTTATCCT	CCCACATACA	ATGTGAAGGA	CATGCTTGTG	CCGACTGCAG	TCTGGAGCGG
1381	GGGTACACGAC	TGGCTTGACG	ATGTCTACGA	CGTCAATATC	TTACTGACTC	AGATCACCAA
1441	CTTGGTGTTC	CATGAGAGCA	TTCCGGAATG	GGAGCATCTT	GACTTCATTT	GGGGCCTGGA
1501	TGCCCCTTGG	AGGCTTTATA	ATAAGATTAT	TAATCTAATG	AGGAAATATC	AGTGATTCTGA
1561	ACCGGCCGCC	CCGGG				

(SEQ ID NO:10)

FIG. 8

DNA sequence of OVR1 promoter (2789 bp)

Feature		Location (bp)				
From DHSIII region of OV gene		7-1658				
OV promoter		1658-2789				
1	CTCGAGTCTC	TTCAGAATGG	CACAGCACCG	CTGCAGAAAA	ATGCCAGGTG	GACTATGAAC
61	TCACATCCAA	AGGAGCTTGA	CCTGATACCT	GATTTTCTTC	AAACAGGGGA	AACAACACAA
121	TCCCACAAAA	CAGCTCAGAG	AGAAACCATC	ACTGATGGCT	ACAGCACCAA	GGTATGCAAT
181	GGCAATCCAT	TCGACATTCA	TCTGTGACCT	GAGCAAAATG	ATTTATCTCT	CCATGAATGG
241	TTGCTTCTTT	CCCTCATGAA	AAGGCAATTT	CCACACTCAC	AATATGCAAC	AAAGACAAAC
301	AGAGAACAAT	TAATGTGCTC	CTTCCTAATG	TTAAATTTGT	AGTGGCAAAG	AGGAGAACAA
361	AATCTCAAGT	TCTGAGTAGG	TTTTAGTGAT	TGGATAAGAG	GCTTTGACCT	GTGAGCTCAC
421	CTGGACTTCA	TATCCTTTTG	GATAAAAAGT	GCTTTTATAA	CTTTCAGGTC	TCCGAGTCTT
481	TATTCATGAG	ACTGTTGGTT	TAGGGACAGA	CCCACAATGA	AATGCCTGGC	ATAGGAAAGG
541	GCAGCAGAGC	CTTAGCTGAC	CTTTTCTTGG	GACAAGCATT	GTCAAACAAT	GTGTGACAAA
601	ACTATTTGTA	CTGCTTTGCA	CAGCTGTGCT	GGGCAGGGCA	ATCCATTGCC	ACCTATCCCA
661	GGTAACCTTC	CAACTGCAAG	AAGATTGTTG	CTTACTCTCT	CTAGACCCCC	AAGTCAAACC
721	AACTATGCAG	GTATGCTGAC	AACACTATGA	TGACAGCCTG	TTCTGATCAA	GATCTCATTT
781	GTTTCATGGAC	AATTTTTGTT	GCTTGCAGCT	GGTCTTCCAT	TGGGAAAGAG	TGTAGTATAT
841	CCTTCTCATC	TGACAGAAAA	GCAGAAAATC	TCATGCTCCA	CACCTAATCT	ACATTGTTTT
901	AAACCACCGG	CTACTTCTTG	GAGAGGAAAA	ATGGCTTTTA	TAAGACTCAC	AAAACAAAGC
961	TCTGCAAGTC	AAATGCATAC	AAAAGCTTTC	TGTAGGTCTG	GAATCAGGAC	ACTATGTGGA
1021	AGTCAAATAG	AGCAGCTTTA	AAAAGCCTTT	GGGATCATTG	TCATCTTATA	TTTGCAGCAC
1081	GATACTATGA	CAGTGATAAC	TGACATAACT	GCATCAATTT	CCTTGATATT	TTATTTGTCT
1141	TAAAGTACAA	GACATAGAGA	TGGACGTAAA	GATGGACATA	TGACTCAGGT	CTGGACAGGT
1201	CCGTGGTCCA	TGTATGATAA	AAGAGATGAA	GGGAAGGAGA	ATTGAGACTG	TCTAAGAAGG
1261	GCTTCAGGGA	CGTTCTGAAG	GCAGATTTGA	CTGAATCAGA	TGTACTGTCC	AAGTCTCATA
1321	TGTAGCAATG	GAAGGCTGAT	ATTGGAGAAA	TATAAAGAAA	TGGCTGTGAA	CTCAAAGTGA
1381	CCCTGAACAG	AAAAGGGATA	TGGAGTTAAA	ATAATGTCAC	AGAAGTGAAG	TTTATATGAT
1441	ATACCATGGG	CTGCAGAGGG	TCAGAGTGCT	CCACCATGGG	CCTCTCTTGG	GCTGCAGGGA
1501	ACTTCTGTTT	TACACCTGGA	ACACCTCCTG	CCCTCCTCCG	CACCTGACCT	AGTGTATGTA
1561	GGGCTGTTTT	TCTCACATTT	TCTCACTCAC	CTCTCCCAAC	TACCATGTGA	CAGCAGTTGT
1621	TCTTACATAT	TGCTCCTCCT	GAGGTACATC	TAGCATCGTT	AAGTCTCAG	ACTTGGCAAG
1681	GAGAATGTAG	ATTTCCACAG	TATATATGTT	TTCAAAAAG	GAAGGAGAGA	AACAAAAGAA
1741	AATGGGCACTG	ACTAAACTTC	AGCTAGTGGT	ATAGGAAAGT	AATTCGTGCT	AACAGAGATT
1801	GCAGTGATCT	CTATGTATGT	CCTGAAGAA	TATGTTGTAC	TTTTTTCCCC	CATTTTTTAAA
1861	TCAAACAGTG	CTTTACAGAG	GTCAGAATGG	TTTCTTTACT	GTTTGTCAAT	TCTATTATTT
1921	CAATACAGAA	CAATAGCTTC	TATAACTGAA	ATATATTTGC	TATTGTATAT	TATGATTGTC
1981	CCTCGAACCA	TGAACACTCC	TCCAGCTGAA	TTTCACAATT	CCTCTGTCAT	CTGCCAGGCC
2041	ATTAAGTTAT	TCATGGAAGA	TCTTTGAGGA	ACACTGCAAG	TTTATATCAT	AAACACATTT
2101	GAAATTGAGT	ATTGTTTTGC	ATTGTATGGA	GCTATGTTTT	GCTGTATCCT	CAGAATAAAA
2161	GTTTGTATATA	AAGCATTAC	ACCCATAAAA	AGATAGATTT	AAATATTCCA	ACTATAGGAA
2221	AGAAAGTGTG	TCTGCTCTTC	ACTCTAGTCT	CAGTTGGCTC	CTTCACATGC	ACGCTTCTTT
2281	ATTTCTCCTA	TTTTGTCAAG	AAAATAATAG	GTCAAGTCTT	GTTCTCATTT	ATGTCCTGTC
2341	TAGCGTGGCT	CAGATGCACA	TTGTACATAC	AAGAAGGATC	AAATGAAACA	GACTTCTGGT
2401	CTGTTACTAC	AACCATAGTA	ATAAGCACAC	TAACATAATA	TTGCTAATTA	TGTTTTCCAT
2461	CTCCAAGGTT	CCCACATTTT	TCTGTTTTCT	TAAAGATCCC	ATTATCTGGT	TGTAACAGAA
2521	GCTCAATGGA	ACATGAGCAA	TATTTCCCAG	TCTTCTCTCC	CATCCAACAG	TCCTGATGGA
2581	TTAGCAGAAC	AGGCAGAAAA	CACATTGTGA	CCCAGAAATTA	AAAACATAATA	TTTGCCTTCC
2641	ATTCAATCCA	AAATGGACCT	ATTGAAACTA	AAATCTAACC	CAATCCCATT	AAATGATTTT
2701	TATGGTGTCA	AAGGTCAAAC	TTCTGAAGGG	AACCTGTGGG	TGGGTCACAA	TTTCAACTAT
2761	ATATTCCCCA	GGGCTCAGCC	AGTGTCTGT			

(SEQ ID NO: 11)

FIG. 9

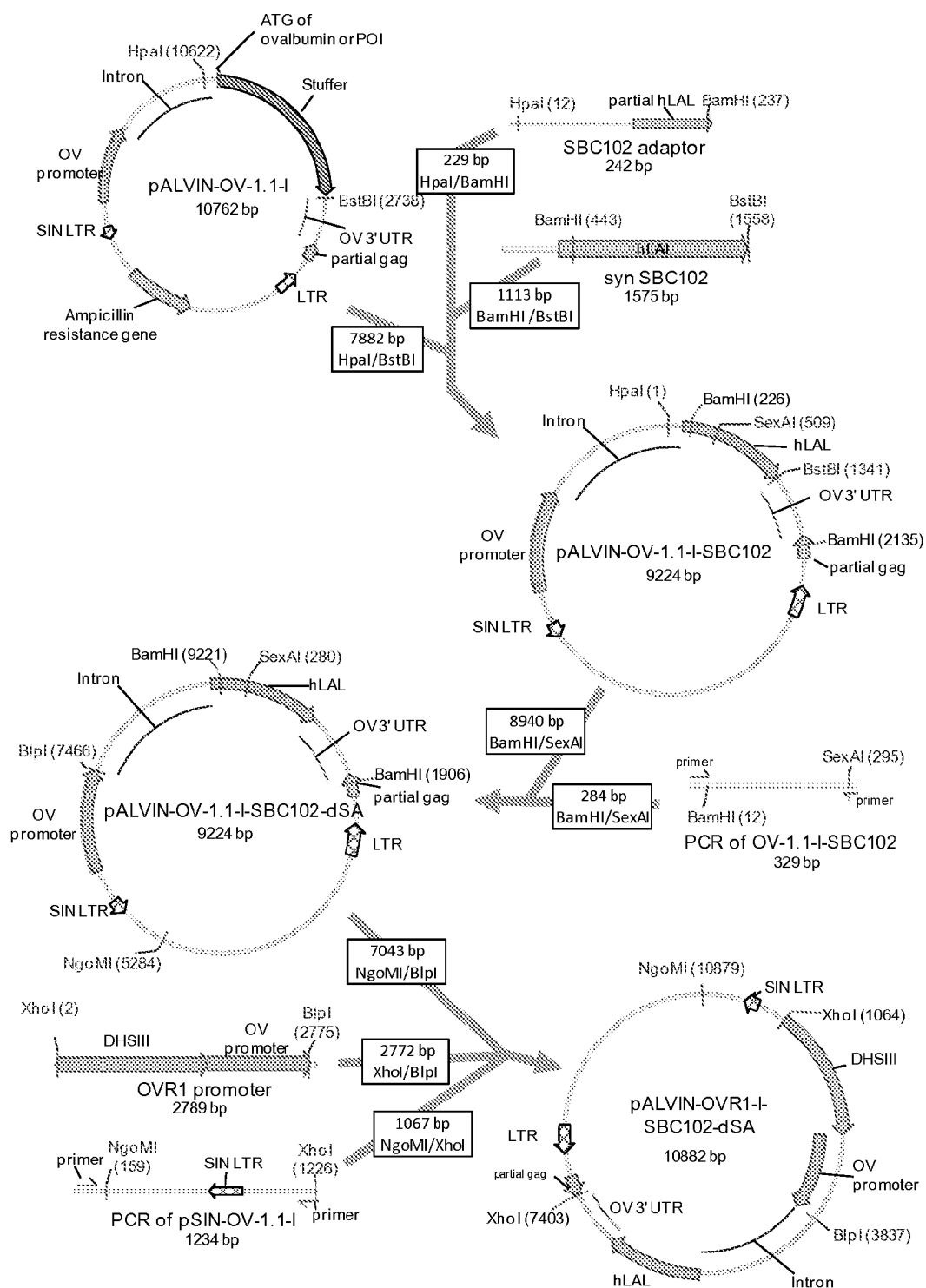


FIG. 10

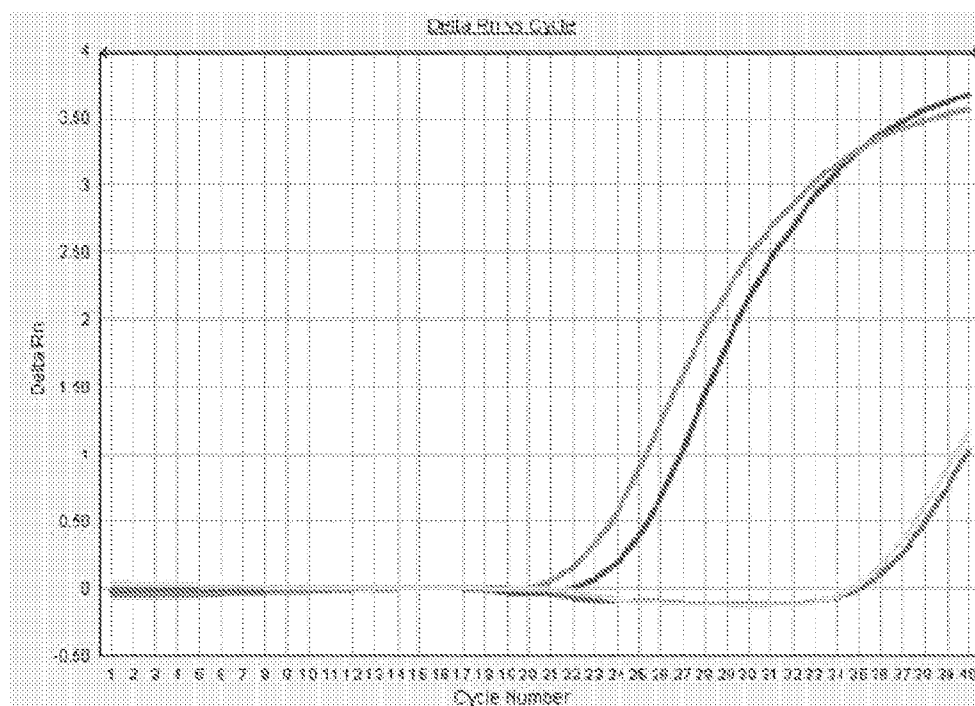


FIG. 11

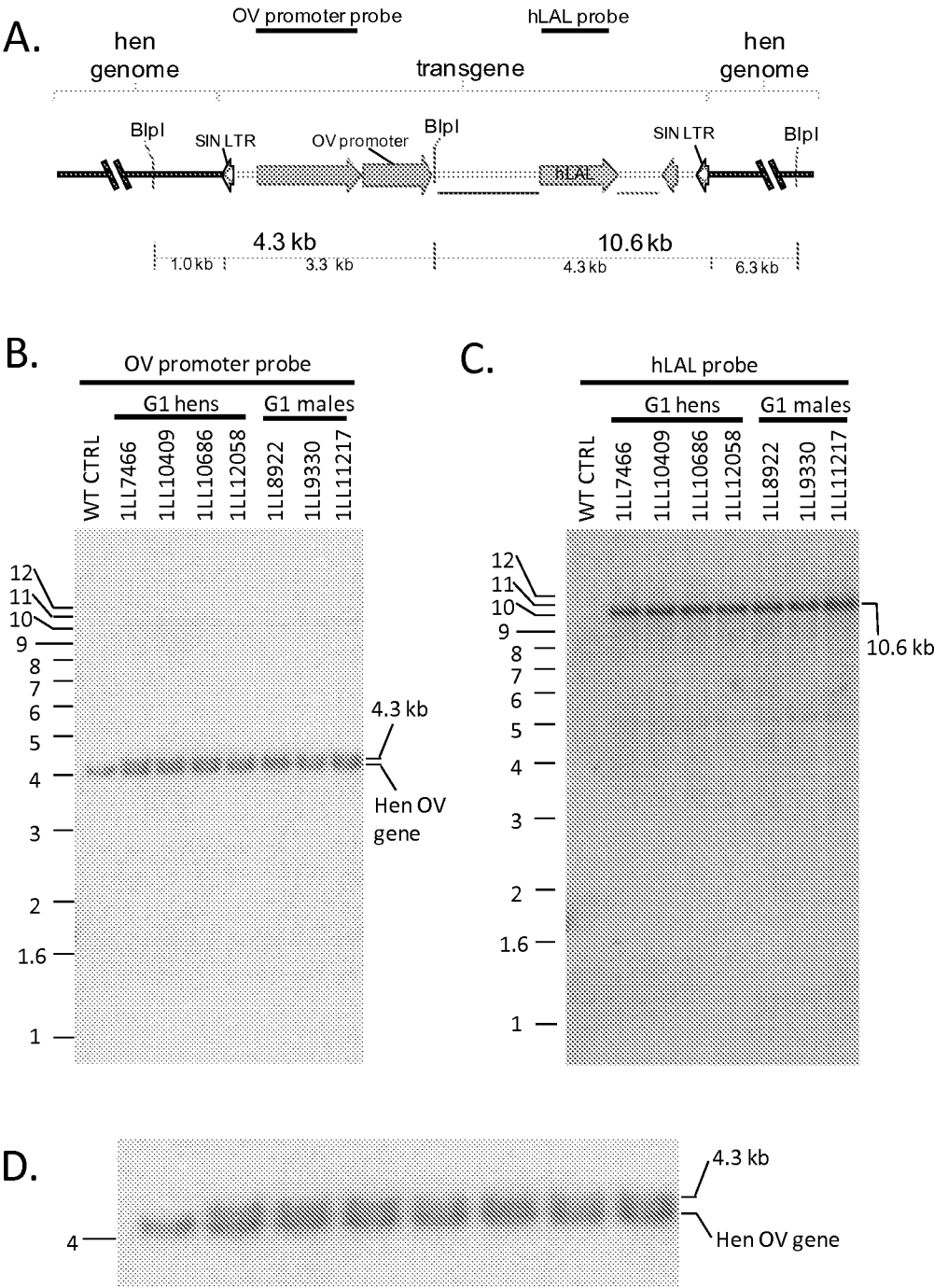


FIG. 12

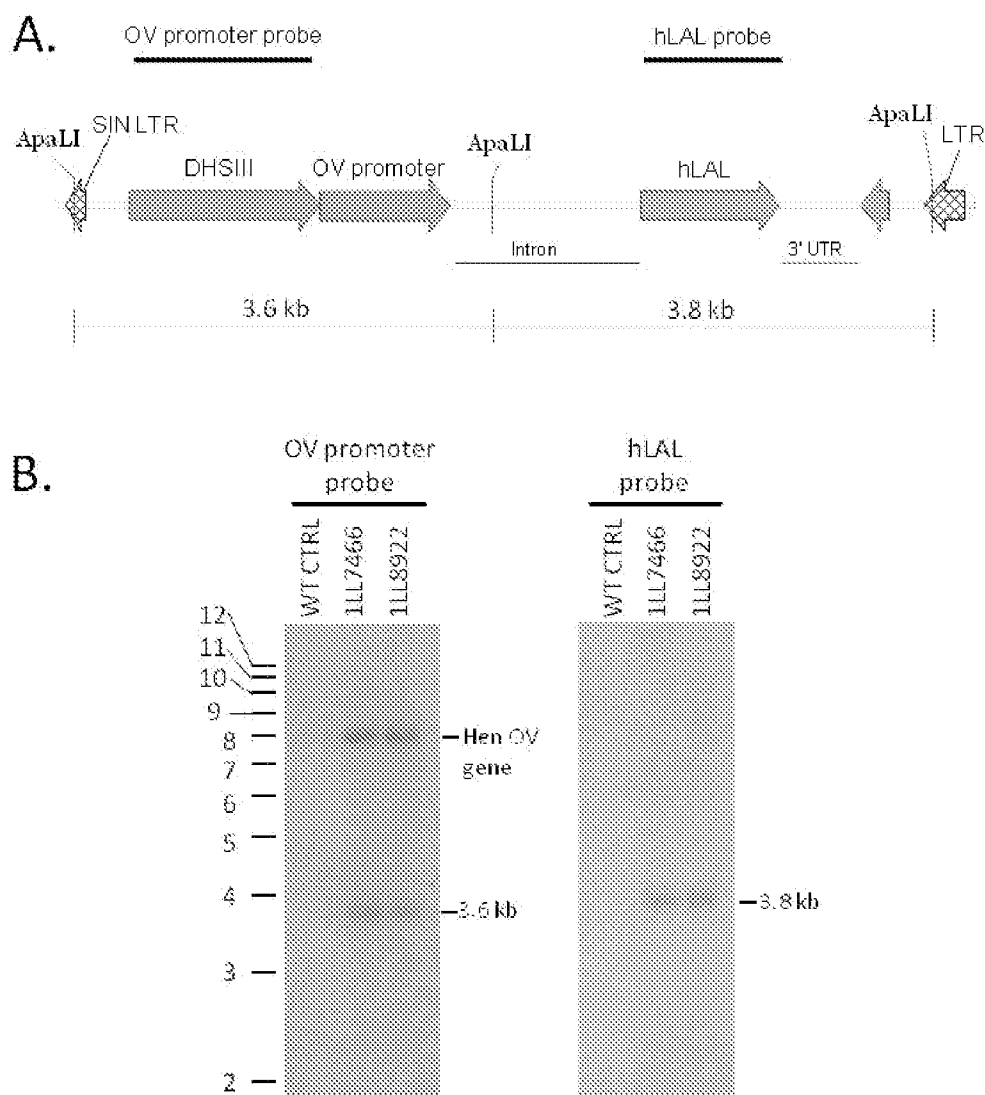


FIG. 13

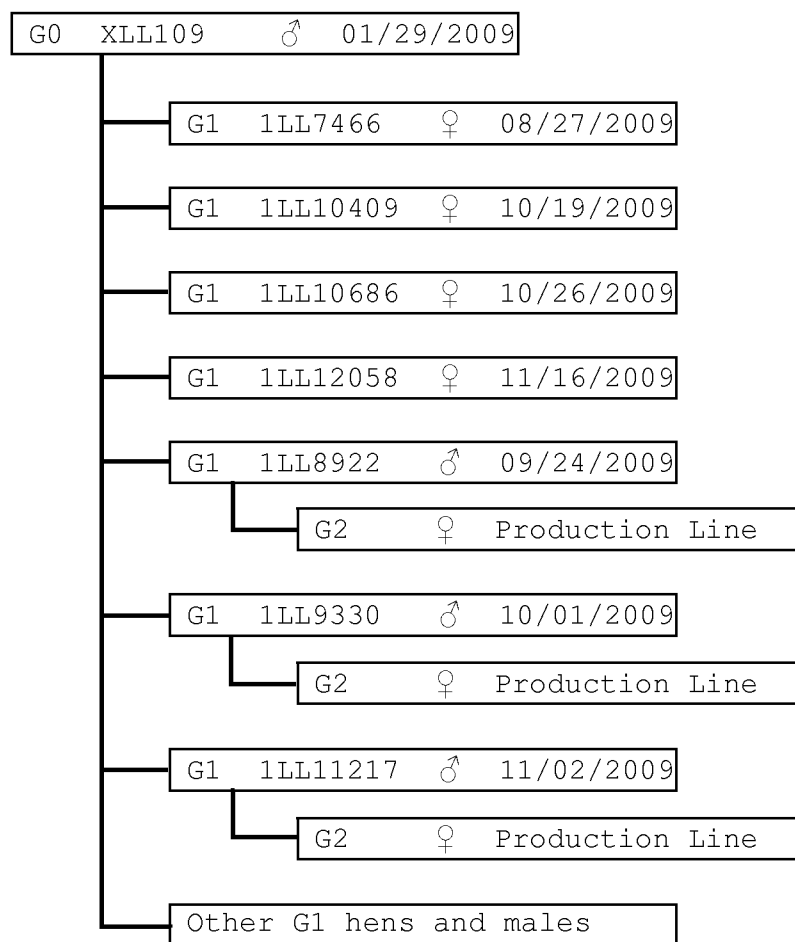


FIG. 14

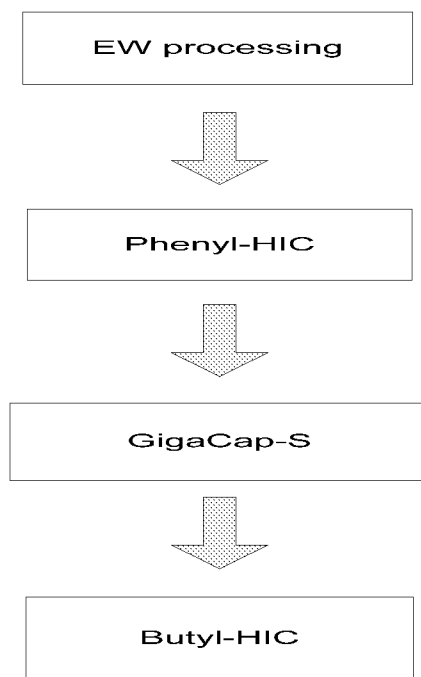
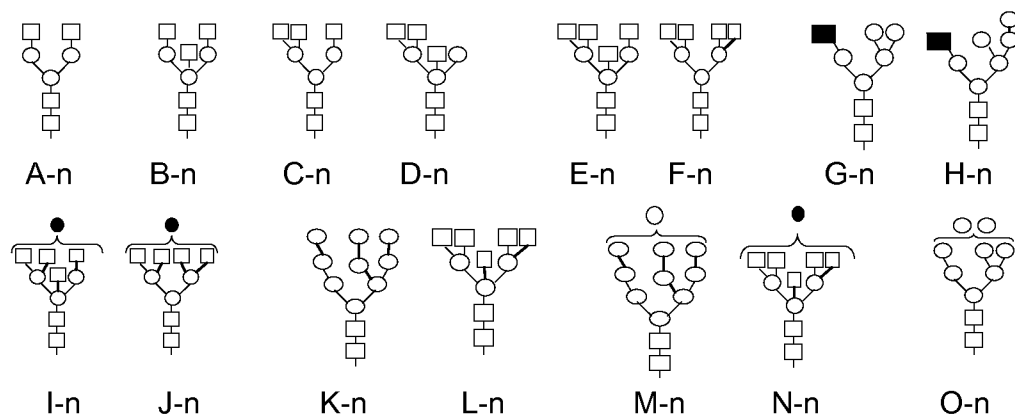


FIG. 15



Square = N-Acetyl Glucosamine

Filled Square = Mannose-6-Phosphate

Circle = Mannose

Filled Circle = Galactose

Filled Triangle = Fucose

FIG. 16

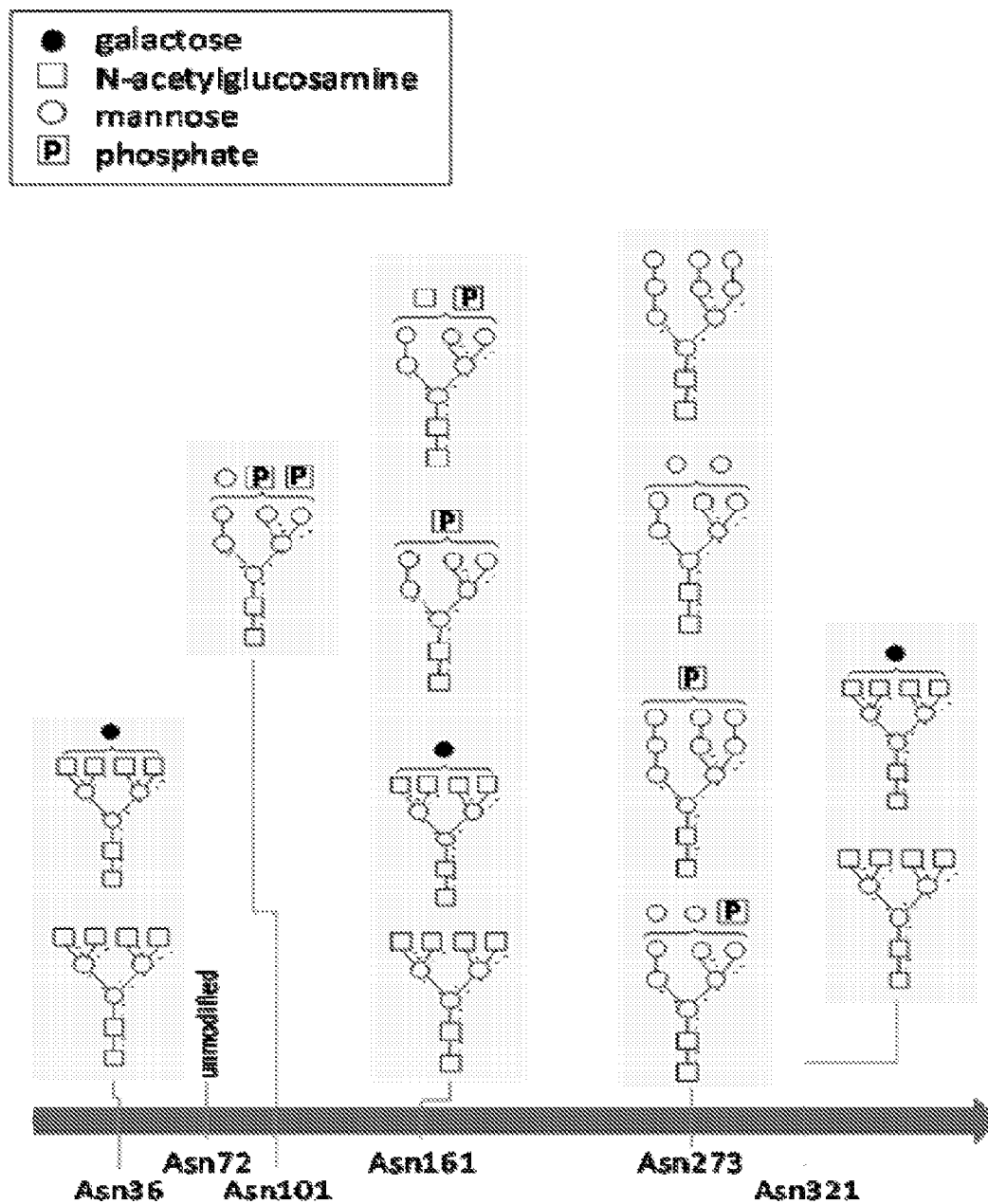


FIG. 17

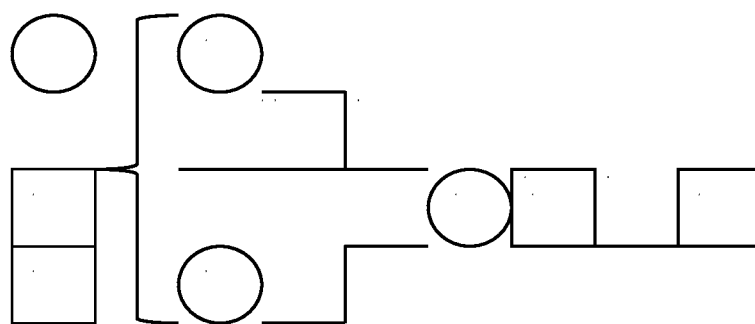
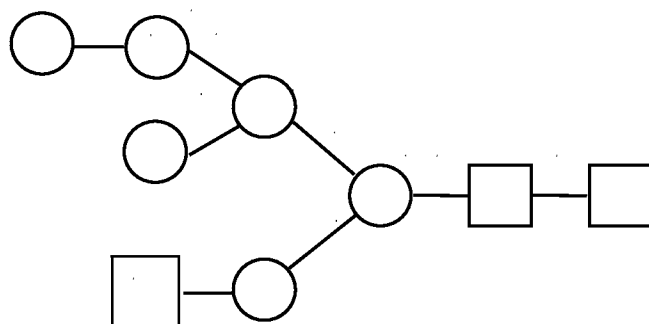
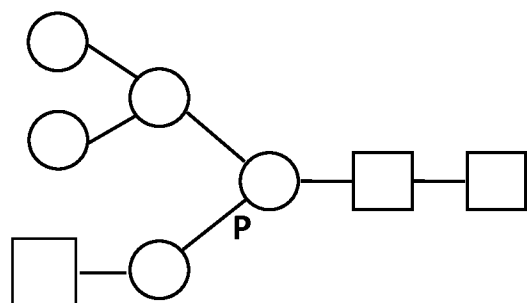


FIG. 18

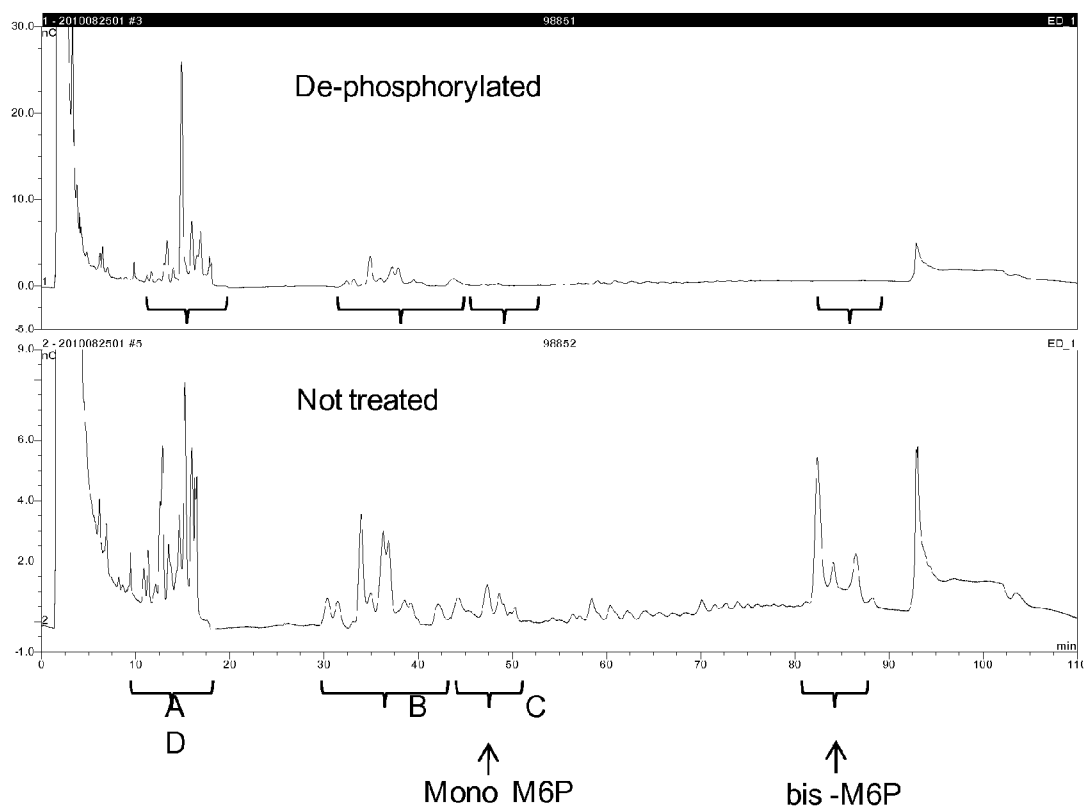
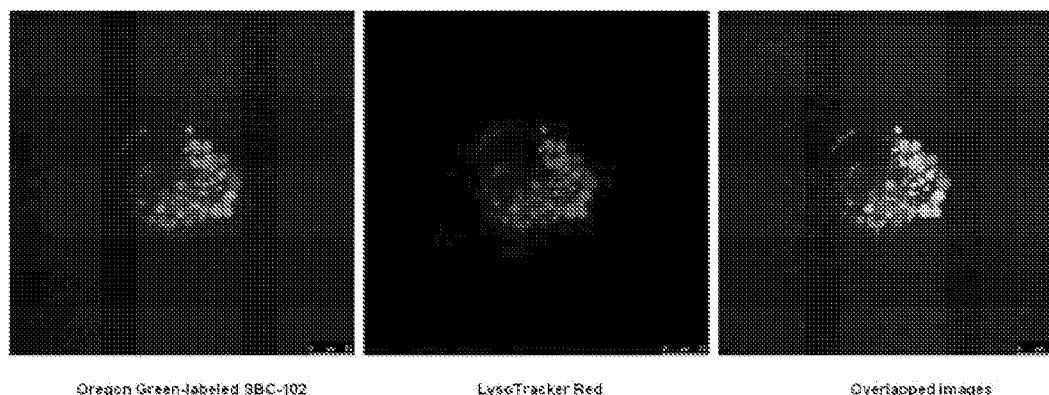


FIG. 19



Oregon Green-labeled SBC-102

LysoTracker Red

Overlapped images

FIG. 20

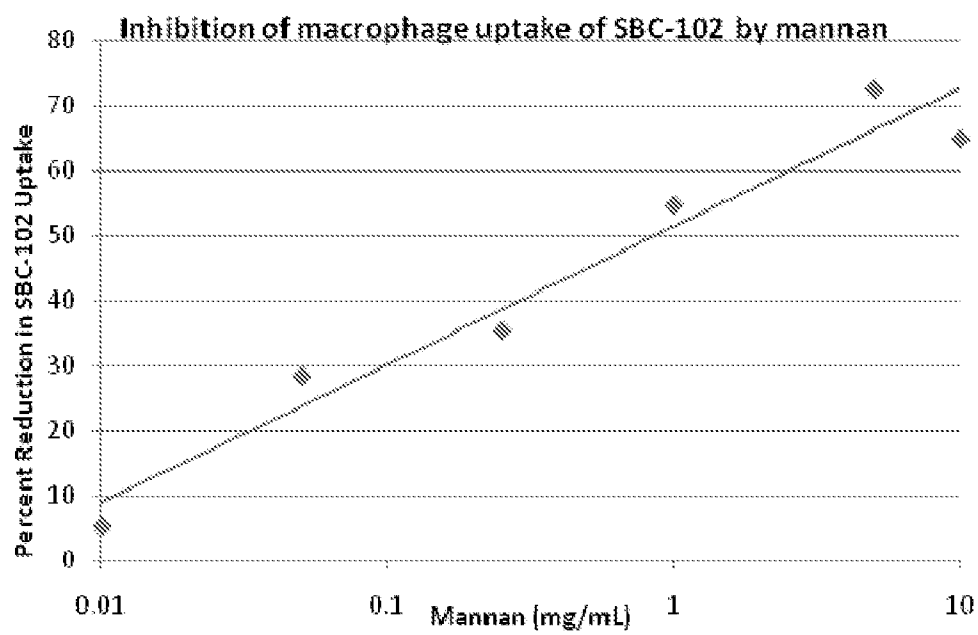


FIG. 21

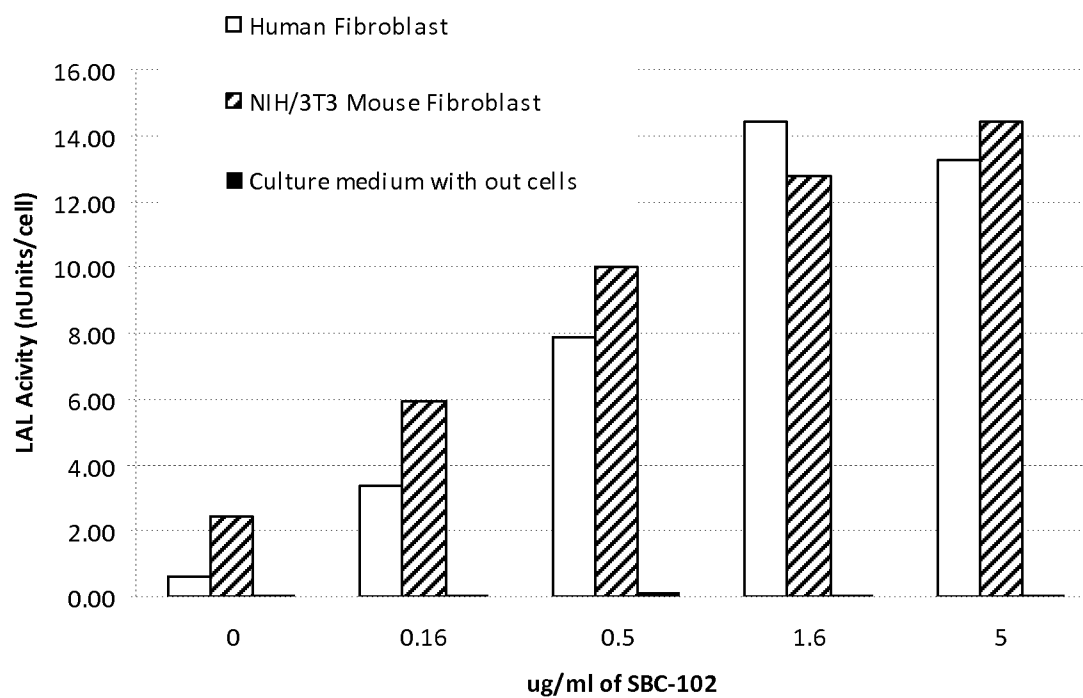


FIG. 22

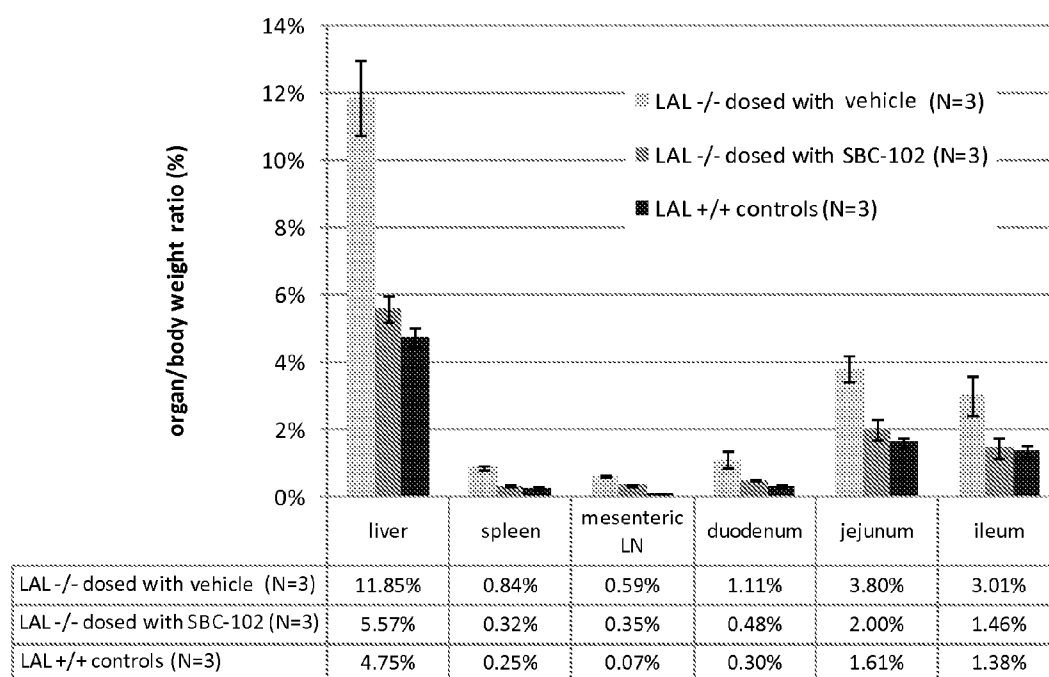


FIG. 23

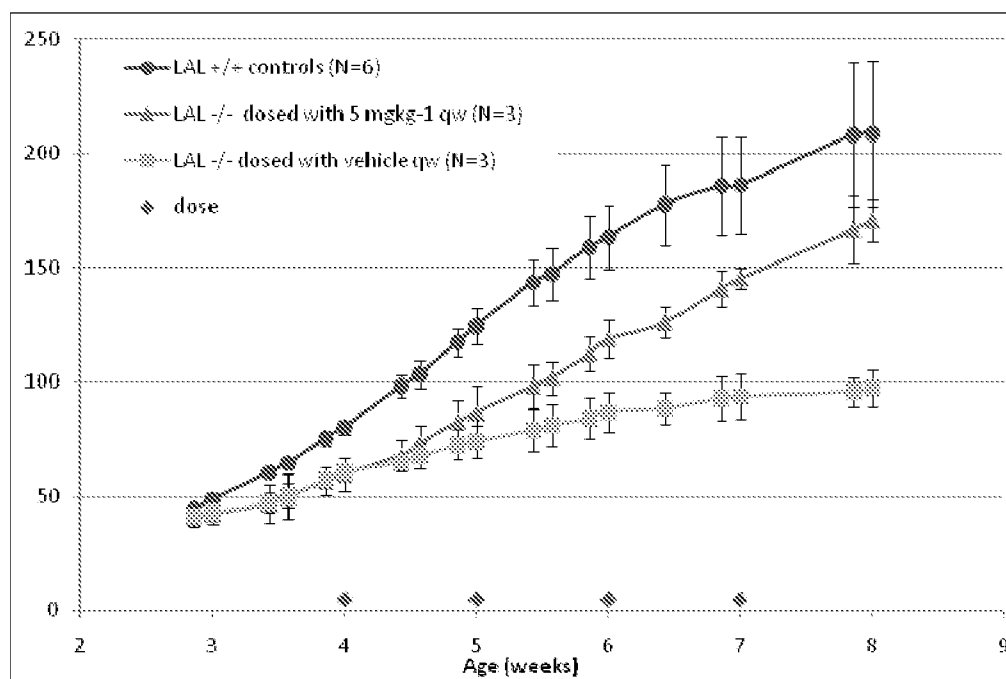


FIG. 24

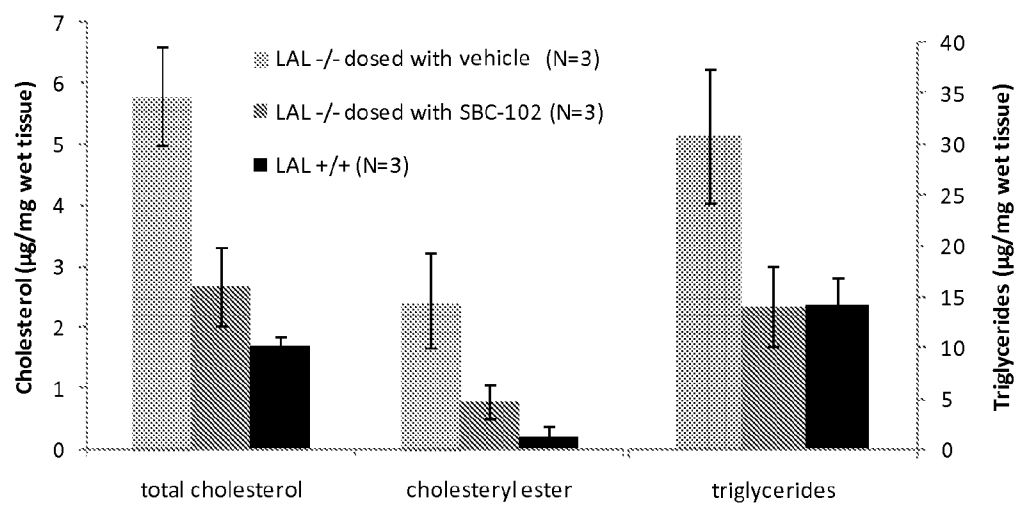


FIG. 25

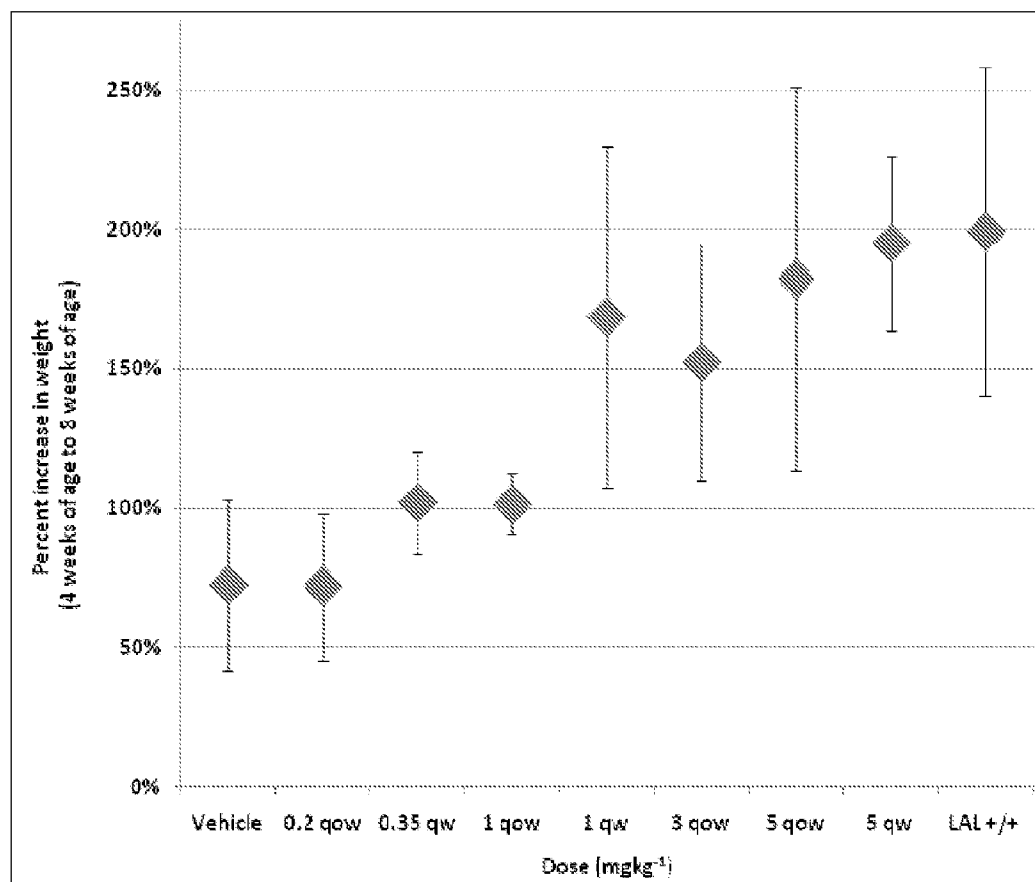


FIG. 26

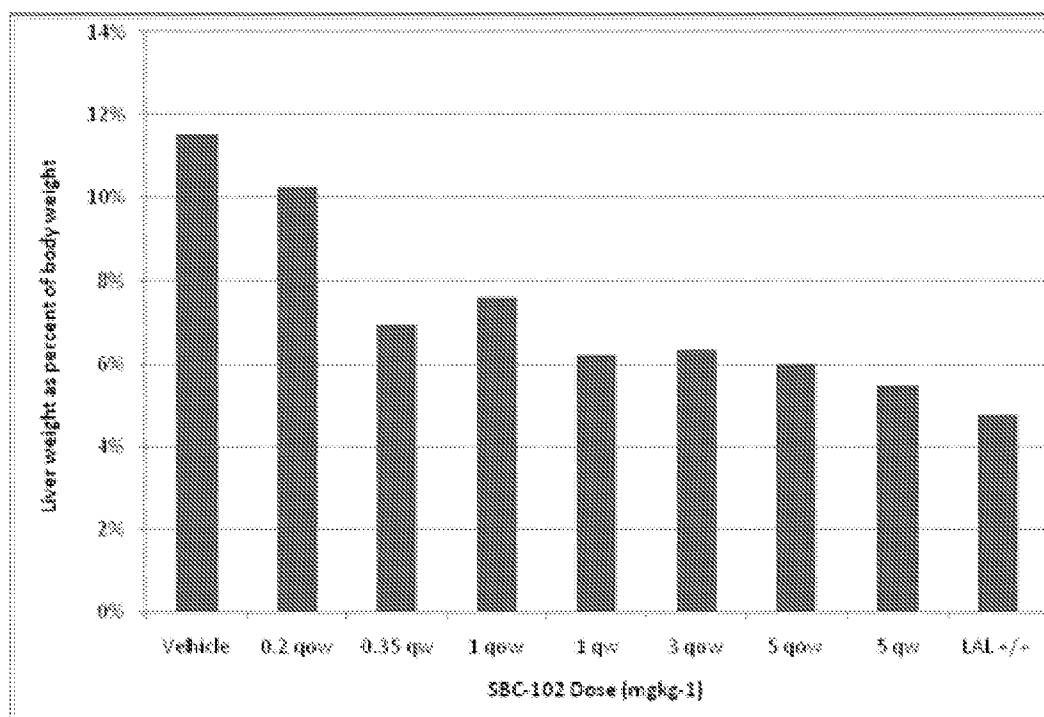


FIG. 27

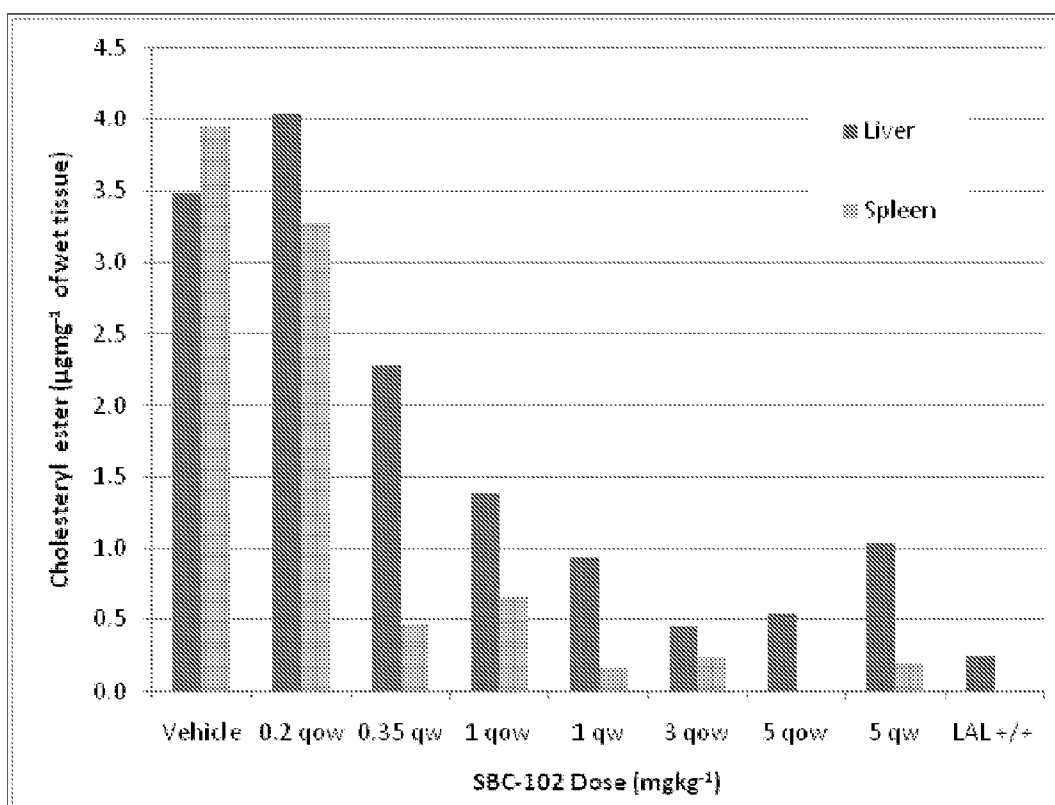


FIG. 28

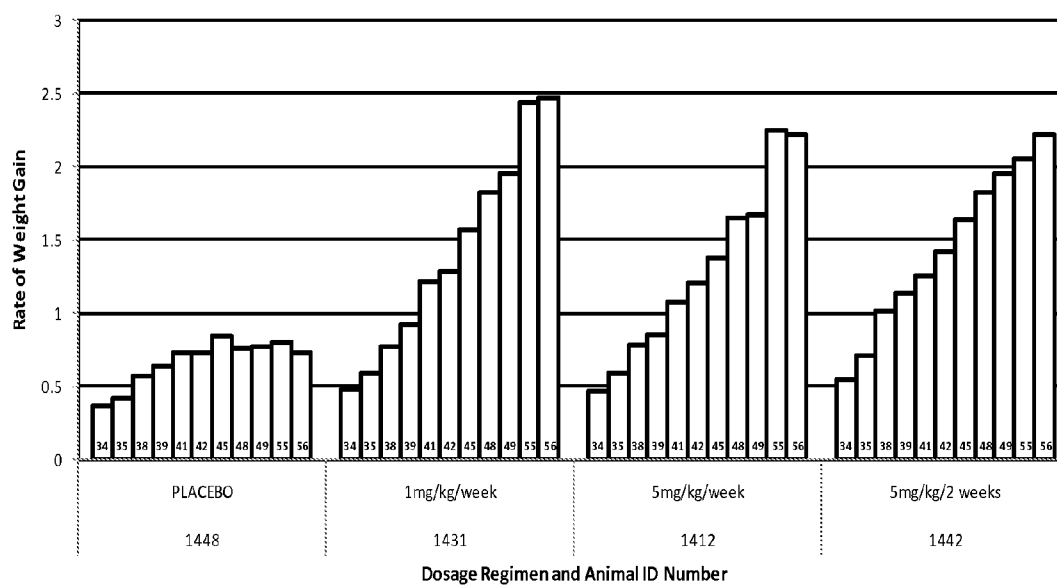


FIG. 29

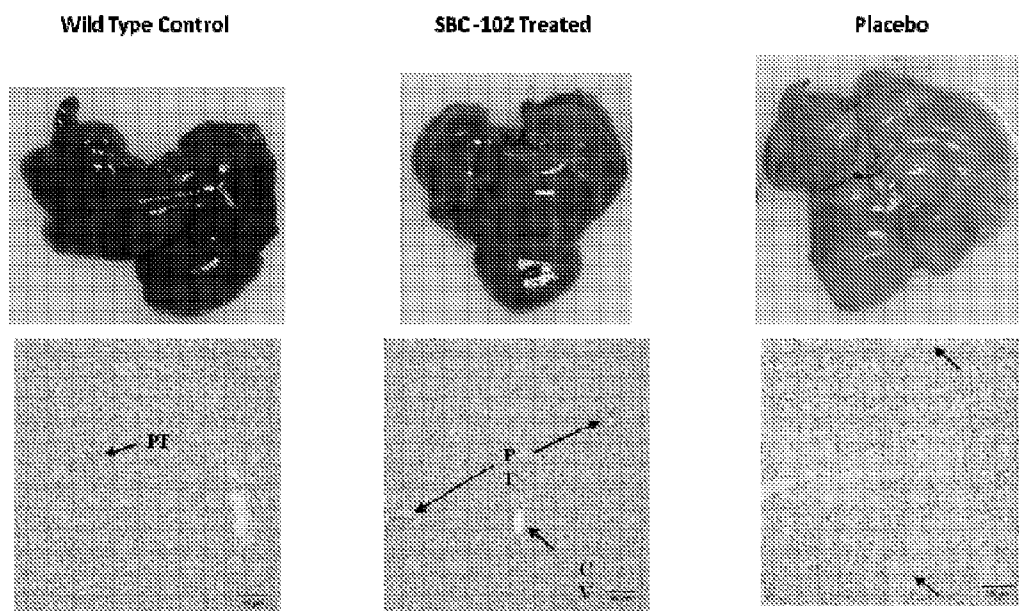


FIG. 30

LYSOSOMAL STORAGE DISEASE ENZYMES

RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/343,177, filed on Apr. 23, 2010, U.S. Provisional Application No. 61/396,376, filed on May 26, 2010, U.S. Provisional Application No. 61/403,011, filed on Sep. 9, 2010, U.S. Provisional Application No. 61/456,014, filed on Oct. 29, 2010, U.S. Provisional Application No. 61/432,372, filed on Jan. 13, 2011. The entire teachings of the above applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Lysosomal Acid Lipase (LAL) deficiency is a very rare lysosomal storage disease (LSD) characterized by a failure to breakdown cholesteryl esters (CE) and triglycerides (TAG) in lysosomes due to a deficiency of the enzyme. LAL deficiency resembles other lysosomal storage disorders with the accumulation of substrate in a number of tissues and cell types. In LAL deficiency substrate accumulation is most marked in cells of the reticuloendothelial system including Kupffer cells in the liver, histiocytes in the spleen and in the lamina propria of the small intestine. Reticuloendothelial cells express the macrophage mannose/N-acetylglucosamine receptor (also known as macrophage mannose receptor or MMR, CD206), which mediates binding, cell uptake and lysosomal internalization of proteins with GlcNAc or mannose terminated N-glycans, and provides a pathway for the potential correction of the enzyme deficiency in these key cell types.

[0003] LAL deficiency is a multi-system disease that most commonly manifests with gastrointestinal, liver and cardiovascular complications and is associated with significant morbidity and mortality. The clinical effects of LAL deficiency are due to a massive accumulation of lipid material in the lysosomes in a number of tissues and a profound disturbance in cholesterol and lipid homeostatic mechanisms, including substantial increases in hepatic cholesterol synthesis. LAL deficiency presents of at least two phenotypes, Wolman Disease (WD) and Cholesteryl Ester Storage Disease (CESD).

[0004] Wolman Disease is the most aggressive presentation of LAL deficiency. This phenotype is characterized by gastrointestinal and hepatic manifestations including growth failure, malabsorption, steatorrhea, profound weight loss and hepatomegaly. Wolman Disease is rapidly progressive and fatal usually within the first year of life. Case report review indicates survival beyond 12 months of age is highly unusual for patients who present with growth failure due to LAL deficiency in the first year of life. In this most aggressive form, growth failure is the predominant clinical feature and is a key contributor to the early mortality. Hepatic involvement as evidenced by liver enlargement and elevation of transaminases is also common in infants. Physical findings include abdominal distention with hepatomegaly and splenomegaly, and radiographic examination often reveals calcification of the adrenal glands. Laboratory evaluations typically reveal elevated levels of serum transaminases and absent or markedly reduced LAL enzyme activity. Elevated blood levels of cholesterol and triglycerides are also seen in patients.

[0005] Current treatment options for Wolman Disease are extremely limited. Antibiotics are administered to infants with pyrexia and/or evidence of infection. Steroid replace-

ment therapy for adrenal insufficiency and specialized nutritional support may be prescribed and while there is no evidence that these interventions prevents death, it is also unclear at present if they have an impact on short term survival. In a series of four patients with LAL deficiency treated with bone marrow transplantation, all four patients died due to complications of the procedure within months of transplantation.

[0006] Patients with LAL deficiency can also present later in life with predominant liver and cardiovascular involvement and this is often called Cholesteryl Ester Storage Disease (CESD). In CESD, the liver is severely affected with marked hepatomegaly, hepatocyte necrosis, elevation of transaminases, cirrhosis and fibrosis. Due to the increased levels of CE and TG, hyperlipidemia and accelerated atherosclerosis are also seen in LAL deficiency. Particularly, an accumulation of fatty deposits on the artery walls is described early in life. The deposits narrow the arterial lumen and can lead to vessel occlusion increasing the risk of significant cardiovascular events including myocardial infarction and strokes. The presentation of CESD is highly variable with some patients going undiagnosed until complications manifest in late adulthood, while others can have liver dysfunction presenting in early childhood. CESD is associated with shortened lifespan and significant ill health; the life expectancy of those with CESD depends on the severity of the associated complications.

[0007] Current treatment options for the CESD phenotype are focused on controlling lipid accumulation through diet that excludes foods rich in cholesterol and triglycerides and suppression of cholesterol synthesis and apolipoprotein B production through administration of cholesterol lowering drugs. Although some clinical improvement may be seen, the underlying disease manifestations persist and disease progression still occurs.

[0008] In most cases, therapy for LAL deficiencies requires life-long treatment. In addition, due to the high cost of protein therapeutics, it is desirable to administer a minimum effective amount of therapeutic to treat LAL deficiency. However, to date, there is no effective therapy for treating LAL deficiency, particularly the patients suffering from Wolman Disease and CESD. Therefore, there is a strong need for an effective therapy with a minimized frequency of administration in order to improve the quality of life for patients. There is also a need for a high expressing and robust protein production platform which can produce LAL proteins that are stable and efficiently targeted to the lysosomal compartment in the affected tissue cells in patients.

SUMMARY OF THE INVENTION

[0009] Disclosed herein are compositions of LAL which are particularly suited for use in therapy, for example, for treatment of conditions associated with LAL deficiency. The LAL molecules described herein contain particular glycan structures which afford efficient and rapid uptake into lysosomes of cells when administered into a subject, for example, a human subject.

[0010] In one aspect, the compositions disclosed herein comprise human LAL wherein a substantial percentage of the human LAL contain at least one mannose-6-phosphate glycan moiety, which can serve as a ligand for internalization by the mannose-6-phosphate receptor on the surface of cells found, for example, on hepatocytes. In one embodiment, 30% or more, for example, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least

95%, at least 97%, or at least 99%, of the LAL contained in the composition contains at least one mannose-6-phosphate moiety. The mannose-6-phosphate moiety can be found, for example, on an N-glycan structure located at one or more residues selected from the group consisting of Asn¹⁵, Asn⁵¹, Asn⁸⁰, Asn¹⁴⁰, Asn²⁵² and Asn³⁰⁰ of SEQ ID NO:2.

[0011] In another aspect, the compositions disclosed herein comprise human LAL wherein a substantial percentage of the human LAL does not contain a sialic acid moiety in any of its N-glycan structures, which can sometimes interfere with internalization of the enzyme into cells. In one embodiment, 15% or less, for example, 10% or less, 5% or less, 2% or less, 1% or less, or essentially none, of the LAL contained in the composition contains a sialic acid moiety in any of its N-glycan structures.

[0012] In another aspect, the compositions disclosed herein comprise human LAL wherein a substantial percentage of the human LAL does not contain a fucose moiety in any of its N-glycan structures. In one embodiment, 50% or less, for example, 50% or less, 40% or less, 30% or less, 20% or less, 10% or less, 5% or less, 2% or less, 1% or less, or essentially none, of the LAL contained in the composition contains a fucose moiety in any of its N-glycan structures.

[0013] In yet another aspect, vectors, host cells, expression systems and associated methods suitable for producing the LAL-containing compositions are described.

[0014] Typically, the LAL of the invention discussed and disclosed herein is human LAL. In one embodiment, the composition comprising LAL includes the mature LAL having the amino acid sequence of:

(SEQ ID NO: 2)
SGGKLTAVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRI PHGRKNHSDKGPKPVVFLQHGL
LADSSNWVTNLANS SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPAS
INFILNKTGQEQVYYVGHSGQTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDH
LIKDLFGDKEFLPQSAFLKWLGVCTHVIKELCGNLCFLLCGFNERNLNMSRVDVYTTTSPAGTSS
VQNMLHWSQAVKFKQAFDWGSSAKNYFHYNQSYPPITYNVKDMLVPTAVWSGGHDWLADVYDVNILL
LTQITNLVPHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ.

[0015] In another embodiment, the mature LAL has the amino acid sequence of:

(SEQ ID NO: 3)
GKLTAVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRI PHGRKNHSDKGPKPVVFLQHGLLA
DSSNWVTNLANS SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASIN
FILNKTGQEQVYYVGHSGQTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLI
KDLFGDKEFLPQSAFLKWLGVCTHVIKELCGNLCFLLCGFNERNLNMSRVDVYTTTSPAGTSSVQ
NMLHWSQAVKFKQAFDWGSSAKNYFHYNQSYPPITYNVKDMLVPTAVWSGGHDWLADVYDVNILLT
QITNLVPHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ.

[0016] In another embodiment, the mature LAL has the amino acid sequence of:

(SEQ ID NO: 4)
TAVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRI PHGRKNHSDKGPKPVVFLQHGLLADSS
NWVTNLANS SLGFILADAGFDVWMGNSRGNTWSRKHKTLSVSQDEFWAFSYDEMAKYDLPASINFIL
NKTGQEQVYYVGHSGQTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDL
FGDKEFLPQSAFLKWLGVCTHVIKELCGNLCFLLCGFNERNLNMSRVDVYTTTSPAGTSSVNML
HWSQAVKFKQAFDWGSSAKNYFHYNQSYPPITYNVKDMLVPTAVWSGGHDWLADVYDVNILLTQIT
NLVPHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ.

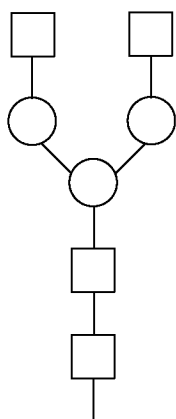
[0017] In another embodiment, the mature LAL has the amino acid sequence of:

(SEQ ID NO: 19)

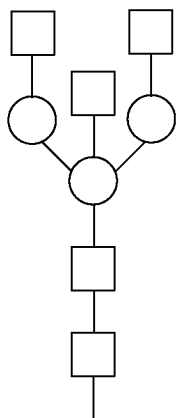
AVDPETNMNVSEIISYWGFPSSEYLVETEDGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSN
WVTNLANSSLGFLADAGFDVWVMGNSRGNTWSRKHKTLVSQDEFWAFSYDEMAKYDLPASINFILN
KTGQEQVYVYVGHSGQTTIGFIAFSQIPELAKRIKMFFALGPVASVAFCTSPMAKLGRLPDHLIKDLF
GDKEFLPQSAFLKWLGTHTVILKELCGNLCFLLCGFNERNLNMSRVDVYTTHTSPAGTSVQNMLH
WSQAVKFKQFQAFDWSGSSAKNYFHYNQSYPTYNVKDMLVPTAVWSGGHDLADVDVNIILLTQITN
LVFHESIPEWEHLDFIWGLDAPWRLYNKIINLMRKYQ.

[0018] In another embodiment, the mature LAL is a mixture of at least two polypeptides selected from the group consisting of: SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and SEQ ID NO:19.

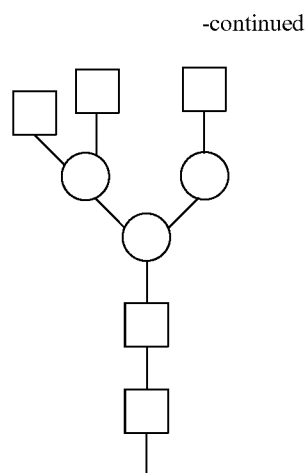
[0019] The invention also provides for compositions which contain isolated mixtures of an individual type of useful protein molecule, such as those proteins disclosed herein, where one or more of the protein molecules contained in the mixture has a specific oligosaccharide structure attached, in particular an oligosaccharide structure disclosed herein. For example, the invention provides for isolated mixtures of LAL molecules, for example, human LAL molecules which contain an LAL molecule glycosylated with one or more of the following structures A-n to O-n:



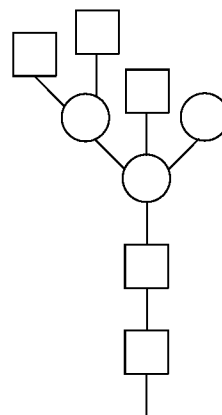
A-n



B-n

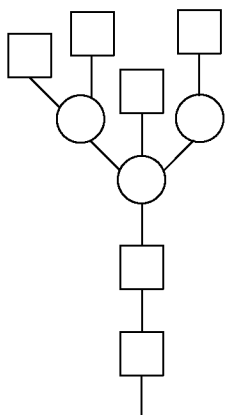


C-n



D-n

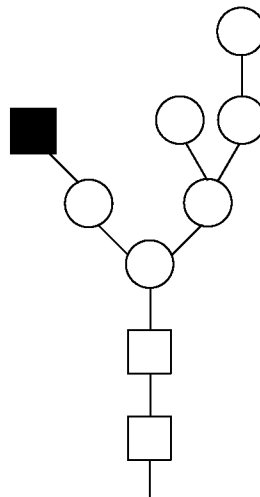
-continued



-continued

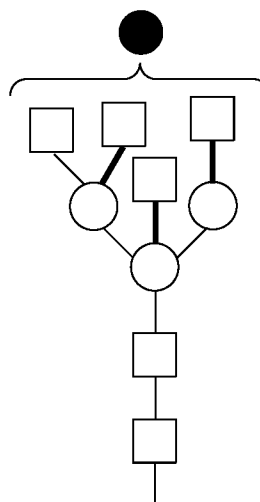
E-n

H-n



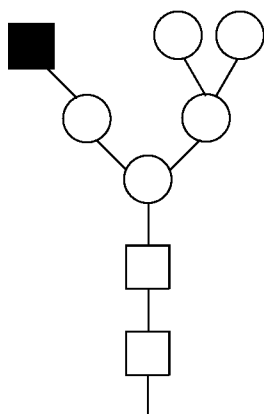
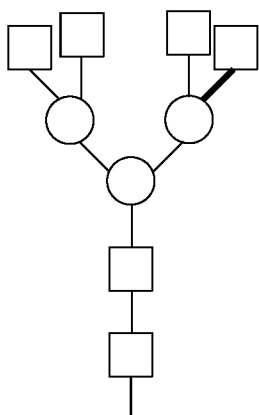
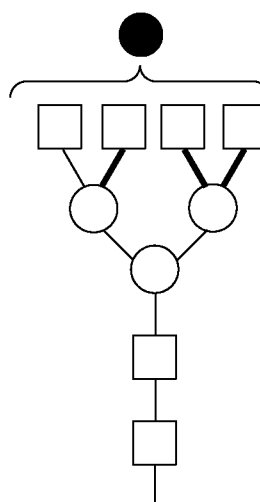
F-n

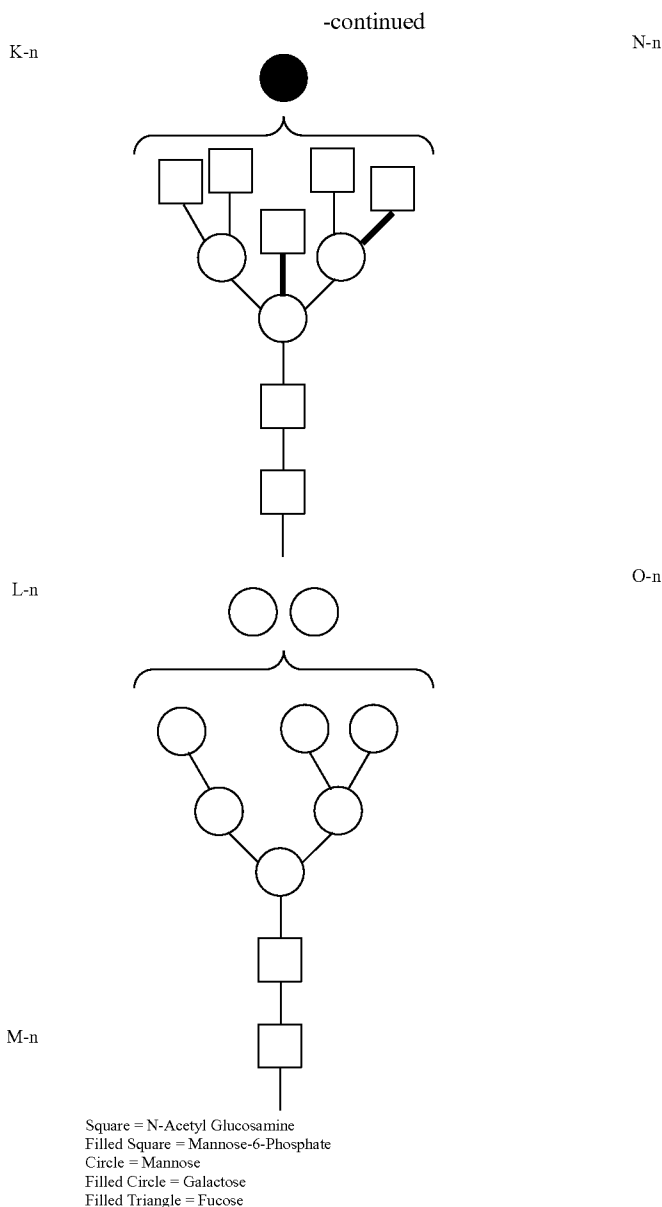
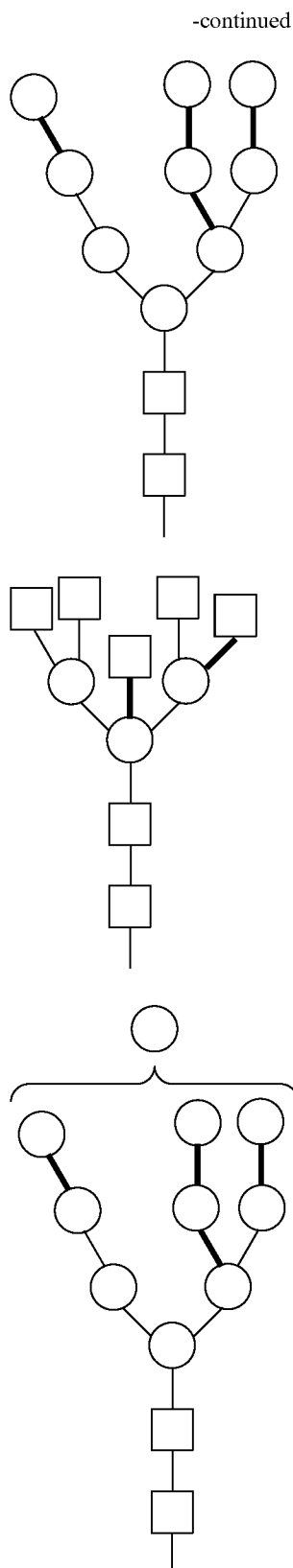
I-n



G-n

J-n





[0020] According to one aspect of the present invention, a composition comprises any isolated individual or combination of the polypeptides described above. In one embodiment, the composition can be a pharmaceutical composition, for example, a formulation that further comprises pharmaceutically acceptable carriers, such that the composition is, for example, suitable for administration into a subject (e.g., a human, particularly a patient suffering from or diagnosed with a condition). The composition can be administered any number of ways, including by intravenous administration. In another embodiment, the composition can further comprise a second agent. Such an agent can be a medicament, or an agent which can influence or modify a biological process when administered into a subject. For example, the second agent can be an immunomodulatory agent. Such immunomodulatory agents can include any agent which, when administered together (i.e., administered at the same time as, or shortly

before or after) with any of the LAL compositions described herein, may have the effect of reducing the immunogenicity of the LAL composition in the subject (e.g., Rituximab, or any other B-cell depleting antibody).

[0021] In a final aspect, methods and compositions for the treatment of symptoms associated with LAL deficiency are disclosed.

[0022] Additional objects and aspects of the present invention will become more apparent upon review of the detailed description set forth below when taken in conjunction with the accompanying figures and sequences.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] FIG. 1 depicts the amino acid sequences of human LAL. The amino acid sequence of the recombinant hLAL shows 100% homologous to that of natural human LAL. The mature form of hLAL is underlined.

[0024] FIG. 2 depicts the nucleotide sequence of recombinant hLAL, the rhLAL transgene of pALVIN-OVR1-I-hLAL-dSA.

[0025] FIGS. 3A and 3B depict diagrams of pALVIN-OVR1-I-hLAL-dSA and its proviral region. FIG. 3A depicts a diagram of human LAL retrovirus expression vector used in the production of transducing particles is diagrammed (the DNA sequence of the plasmid is located in Appendix A). FIG. 3A depicts pALVIN-OVR1-I-hLAL-dSA proviral region that has been integrated into the genome. SIN LTR, self-inactivating long terminal repeat; OV DHSIII enhancer, DNase hypersensitive site III of the ovalbumin gene; OV Intron; ovalbumin 5' untranslated region and intron 1; hLAL, human LAL cDNA; OV 3' UTR, ovalbumin gene 3' untranslated region; partial gag, partial gag gene; LTR, long terminal repeat.

[0026] FIG. 4 depicts a nucleotide sequence of pALVIN-OVR1-I-hLAL-dSA.

[0027] FIG. 5 depicts a nucleotide sequence of pALVIN-OVR1-I-hLAL-dSA proviral region that has been integrated into the genome.

[0028] FIG. 6 depicts a nucleotide sequence of pALVIN-OV-1.1-I vector.

[0029] FIG. 7 depicts a nucleotide sequence of rhLAL adaptor.

[0030] FIG. 8 depicts a nucleotide sequence of rhLAL including the partial ovalbumin promoter.

[0031] FIG. 9 depicts a nucleotide sequence of OVR1 promoter.

[0032] FIG. 10 depicts schematics of the steps used to construct the pALVIN-OVR1-I-hLAL-dSA vector.

[0033] FIG. 11 depicts a real-time PCR analysis of blood DNA samples from a hemizygous transgenic G1 offspring of XLL109. The signals from duplicate DNA samples of hemizygous G1 progeny, 1LL7466, are indicated by the curves that initiate an increase in Delta Rn before cycle 22. The curves for two non-transgenic progeny are shown; these curves stay at or near baseline through at least 34 cycles.

[0034] FIGS. 12A-D depict Southern analysis of G1 chickens carrying the ALVIN-OVR1-I-hLAL-dSA transgene. FIG. 12A illustrates schematic of the integrated transgene and flanking genomic regions is shown with the known position of the transgene BlnI site and predicted position of the flanking genomic BlnI sites. The position of the OV promoter probe and the hLAL coding sequence probe (hLAL probe) are indicated by the black bars. The positions of the 4.3 kb and 10.6 kb bands detected in the Southern analysis are shown as well as the predicted sizes of the genomic and transgene

portions of the 4.3 kb and 10.6 kb bands. FIG. 12B illustrates a Southern blot of genomic DNA digested with BlnI and probed with the OV probe. WT CTRL is genomic DNA isolated from a non-transgenic chicken. The ID numbers of the G1 transgenics are indicated above the lanes. The position and size (kb) of the molecular weight markers are shown to the left of the blot. The position and size of the detected transgene fragment (4.3 kb) and endogenous ovalbumin gene (4.1 kb) are shown to the right of the blot. FIG. 12C depicts a Southern blot was probed with the hLAL probe. The position and size of the detected transgene fragment (10.6 kb) is shown to the right of the blot. FIG. 12D depicts a section of the image shown in FIG. 12B at a larger scale to demonstrate the presence of the 4.1 and 4.3 kb bands.

[0035] FIG. 13A depicts schematic of the ALVIN-OVR1-I-hLAL-dSA transgene. The size of ApaI bands predicted to be detected by the OV probe and hLAL probe are also shown. FIG. 13B depicts schematic of a Southern blot analysis of the ALVIN-OVR1-I-hLAL-dSA transgene for confirmation of transgene size. Southern blot of genomic DNA digested with ApaI and probed with either the OV probe (left panel) or hLAL probe (right panel). WT CTRL is genomic DNA isolated from a non-transgenic chicken. The ID number of the G1s is indicated above each lane. The position and size (kb) of the molecular weight markers are shown to the left of the blots. The position and size of the detected transgene fragments (OV promoter probe, 3.6 kb; hLAL probe, 3.8 kb) and endogenous ovalbumin gene (7.7 kb) are shown to the right of the blots.

[0036] FIG. 14 depicts a lineage of transgenic chickens. Shown for each chicken are the generation number (G0, G1 or G2), identification number, gender and hatch date. Other G1 chickens are those of other lineages.

[0037] FIG. 15 depicts the purification steps of hLAL from egg white.

[0038] FIG. 16 depicts N-glycans found as an N-linked Glycosylation structure in LAL produced in accordance with the invention. Square, N-Acetyl glucosamine; Filled square, mannose-6-phosphate; circle, mannose; filled circle; galactose; and filled triangle, fucose.

[0039] FIG. 17 depicts the relative position of predicted N-glycan sites indicated on the LAL polypeptide (arrow) set forth in SEQ ID NO:1. N-glycans that are structurally representative of those detected at each site are shown. Square, N-Acetyl glucosamine; Filled square, mannose-6-phosphate; circle, mannose; filled circle; galactose; and filled triangle, fucose.

[0040] FIG. 18 depicts phosphorylated N-glycans released by PNGase and analyzed by MALDI-TOF. Structures are shown.

[0041] FIG. 19 depicts the effect of dephosphorylation of LAL on HPAEC-PAD retention time of N-glycans. LAL produced in accordance with the invention was dephosphorylated with bacterial alkaline phosphatase (upper panel) or left untreated (lower panel). Released N-glycans were analyzed by HPAEC-PAD.

[0042] FIG. 20 depicts the co-localization of recombinant human LAL(SBC-102) and lysosomal marker in the lysosomes of these cells examined by confocal fluorescence microscopy using a sequential scanning mode.

[0043] FIG. 21 depicts the binding specificity of recombinant human LAL (SBC-102) to the GlcNAc/mannose receptor assessed by competitive binding assays using the macrophage cell line, NR8383.

[0044] FIG. 22 depicts the activity of recombinant human LAL in cells in normal and LAL-deficient cells in vitro.

[0045] FIG. 23 depicts the effect of recombinant human LAL (SBC-102) treatment on internal organs mass of LAL deficient rats. Organ size is represented as percent of body weight determined at 8 weeks of age, in LAL^{-/-} rats and LAL^{+/-} rats after weekly administration of vehicle or SBC-102 at 5 mg/kg for 4 weeks.

[0046] FIG. 24 depicts body weight in wild type and LAL-deficient rats after weekly administration of vehicle or SBC-102 at 5 mg·kg⁻¹ for 4 weeks. Dose administration is highlighted on X-axis by diamonds starting at 4 week.

[0047] FIG. 25 shows liver cholesterol, cholesteryl ester and triglyceride levels determined at 8 weeks of age in WT and LAL deficient rats after weekly administration of vehicle or recombinant human LAL (SBC-102) at 5 mg·kg⁻¹ for 4 weeks.

[0048] FIG. 26 depicts percent increase in body weight in LAL-deficient rats after 4 weeks administration recombinant human LAL (SBC-102) at the indicated levels and schedules, determined at 8 weeks of age.

[0049] FIG. 27 shows liver weight, as a percent of body weight, in LAL-deficient rats after 4 weeks administration SBC-102 at the indicated levels and schedules, determined at 8 weeks of age.

[0050] FIG. 28 shows tissue cholesteryl ester levels in LAL-deficient rats after 4 weeks administration SBC-102 at the indicated levels and schedules, determined at 8 weeks of age.

[0051] FIG. 29 shows the daily progress in weight gain of rats which were administered either 1 mg/kg of LAL per week or 5 mg/kg of LAL per week or 5 mg/kg of LAL per two weeks.

[0052] FIG. 30 depicts the gross pathological examination of treated animals showing a substantial normalization in liver size and color as can be seen in the dissection at the top panels and histopathology of liver tissue from LAL of treated rats showing normal liver histology in marked contrast to the substantial accumulation of foamy macrophages in the placebo-treated animals at the bottom panels.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0053] Certain definitions are set forth herein to illustrate and define the meaning and scope of the various terms used to describe the invention herein.

[0054] As used herein, the term “acceptable” with respect to a formulation, composition or ingredient, as used herein, means having no persistent detrimental effect on the general health of the subject being treated.

[0055] As used herein, the term “administration” or “administering” refers to providing a recombinant human lysosomal acid lipase of the invention to a subject in need of treatment.

[0056] A “nucleic acid or polynucleotide sequence” includes, but is not limited to, eukaryotic mRNA, cDNA, genomic DNA, and synthetic DNA and RNA sequences, comprising the natural nucleoside bases adenine, guanine, cytosine, thymidine, and uracil. The term also encompasses sequences having one or more modified bases.

[0057] The term “avian” as used herein refers to any species, subspecies or race of organism of the taxonomic class avia, such as, but not limited to chicken, turkey, duck, goose,

quail, pheasants, parrots, finches, hawks, crows and ratites including ostrich, emu and cassowary. The term includes the various known strains of *Gallus gallus*, or chickens, (for example, White Leghorn, Brown Leghorn, Barred-Rock, Sussex, New Hampshire, Rhode Island, Australorp, Minorca, Amroxt, California Gray), as well as strains of turkeys, pheasants, quails, duck, ostriches and other poultry commonly bred in commercial quantities. It also includes an individual avian organism in all stages of development, including embryonic and fetal stages.

[0058] “Therapeutic proteins” or “pharmaceutical proteins” include an amino acid sequence which in whole or in part makes up a drug.

[0059] A “coding sequence” or “open reading frame” refers to a polynucleotide or nucleic acid sequence which can be transcribed and translated (in the case of DNA) or translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A transcription termination sequence is usually located 3' to the coding sequence. A coding sequence may be flanked on the 5' and/or 3' ends by untranslated regions.

[0060] “Exon” refers to that part of a gene which, when transcribed into a nuclear transcript, is “expressed” in the cytoplasmic mRNA after removal of the introns or intervening sequences by nuclear splicing.

[0061] Nucleic acid “control sequences” or “regulatory sequences” refer to promoter sequences, translational start and stop codons, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, as necessary and sufficient for the transcription and translation of a given coding sequence in a defined host cell. Examples of control sequences suitable for eukaryotic cells are promoters, polyadenylation signals, and enhancers. All of these control sequences need not be present in a recombinant vector so long as those necessary and sufficient for the transcription and translation of the desired gene are present.

[0062] “Operably or operatively linked” refers to the configuration of the coding and control sequences so as to perform the desired function. Thus, control sequences operably linked to a coding sequence are capable of effecting the expression of the coding sequence. A coding sequence is operably linked to or under the control of transcriptional regulatory regions in a cell when DNA polymerase binds the promoter sequence and transcribes the coding sequence into mRNA that can be translated into the encoded protein. The control sequences need not be contiguous with the coding sequence, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered “operably linked” to the coding sequence.

[0063] The terms “heterologous” and “exogenous” as they relate to nucleic acid sequences such as coding sequences and control sequences, denote sequences that are not normally associated with a region of a recombinant construct or with a particular chromosomal locus, and/or are not normally associated with a particular cell. Thus, an “exogenous” region of a nucleic acid construct is an identifiable segment of nucleic acid within or attached to another nucleic acid molecule that

is not found in association with the other molecule in nature. For example, an exogenous region of a construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of an exogenous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, a host cell transformed with a construct or nucleic acid which is not normally present in the host cell would be considered exogenous for purposes of this invention.

[0064] As used herein the terms “N-glycan,” “oligosaccharide,” “oligosaccharide structure,” “glycosylation pattern,” “glycosylation profile” and “glycosylation structure” have essentially the same meaning and each refer to one or more structures which are formed from sugar residues and are attached to glycosylated proteins.

[0065] “Exogenous protein” as used herein refers to a protein not naturally present in a particular tissue or cell, a protein that is the expression product of an exogenous expression construct or transgene, or a protein not naturally present in a given quantity in a particular tissue or cell. A protein that is exogenous to an egg is a protein that is not normally found in the egg. For example, a protein exogenous to an egg may be a protein that is present in the egg as a result of the expression of a coding sequence present in a transgene of the animal laying the egg.

[0066] “Endogenous gene” refers to a naturally occurring gene or fragment thereof normally associated with a particular cell.

[0067] “LAL” means “human lysosomal acid lipase,” “SBC-102” or “human lysosomal acid lipase molecule” and these terms are used interchangeably throughout the specification.

[0068] The expression products described herein may consist of proteinaceous material having a defined chemical structure. However, the precise structure depends on a number of factors, particularly chemical modifications common to proteins. For example, since all proteins contain ionizable amino and carboxyl groups, the protein may be obtained in acidic or basic salt form, or in neutral form. The primary amino acid sequence may be derivatized using sugar molecules (glycosylation) or by other chemical derivatizations involving covalent or ionic attachment with, for example, lipids, phosphate, acetyl groups and the like, often occurring through association with saccharides. These modifications may occur in vitro or in vivo, the latter being performed by a host cell through post-translational processing systems. Such modifications may increase or decrease the biological activity of the molecule, and such chemically modified molecules are also intended to come within the scope of the invention.

[0069] Alternative methods of cloning, amplification, expression, and purification will be apparent to the skilled artisan. Representative methods are disclosed in Sambrook, Fritsch, and Maniatis, *Molecular Cloning, a Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory (1989).

[0070] “Vector” means a polynucleotide comprised of single strand, double strand, circular, or supercoiled DNA or RNA. A typical vector may be comprised of the following elements operatively linked at appropriate distances for allowing functional gene expression: replication origin, promoter, enhancer, 5' mRNA leader sequence, ribosomal binding site, nucleic acid cassette, termination and polyadenylation sites, and selectable marker sequences. One or more of

these elements may be omitted in specific applications. The nucleic acid cassette can include a restriction site for insertion of the nucleic acid sequence to be expressed. In a functional vector the nucleic acid cassette contains the nucleic acid sequence to be expressed including translation initiation and termination sites. An intron optionally may be included in the construct, for example, 5' to the coding sequence. A vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the “control” of the control or regulatory sequences. Modification of the sequences encoding the particular protein of interest may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; or to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site which is in reading frame with and under regulatory control of the control sequences.

[0071] A “promoter” is a site on the DNA to which RNA polymerase binds to initiate transcription of a gene. In some embodiments the promoter can be modified by the addition or deletion of sequences, or replaced with alternative sequences, including natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences. Many eukaryotic promoters contain two types of recognition sequences: the TATA box and the upstream promoter elements. The former, located upstream of the transcription initiation site, is involved in directing RNA polymerase to initiate transcription at the correct site, while the latter appears to determine the rate of transcription and is upstream of the TATA box. Enhancer elements can also stimulate transcription from linked promoters, but many function exclusively in a particular cell type. Many enhancer/promoter elements derived from viruses, e.g., the SV40 promoter, the cytomegalovirus (CMV) promoter, the rous-sarcoma virus (RSV) promoter, and the murine leukemia virus (MLV) promoter are all active in a wide array of cell types, and are termed “ubiquitous.” Alternatively, non-constitutive promoters such as the mouse mammary tumor virus (MMTV) promoter may also be used in the present invention. The nucleic acid sequence inserted in the cloning site may have any open reading frame encoding a polypeptide of interest, with the proviso that where the coding sequence encodes a polypeptide of interest, it should lack cryptic splice sites which can block production of appropriate mRNA molecules and/or produce aberrantly spliced or abnormal mRNA molecules.

[0072] As used herein, the term “pharmaceutical composition” refers to a mixture of a compound described herein with other chemical components, such as carriers, stabilizers, diluents, dispersing agents, suspending agents, thickening agents, and/or excipients.

[0073] The term “poultry derived” or “avian derived” refers to a composition or substance produced by or obtained from poultry. “Poultry” refers to avians that can be kept as livestock, including but not limited to, chickens, duck, turkey, quail and ratites. For example, “poultry derived” may refer to chicken derived, turkey derived and/or quail derived.

[0074] A “retroviral particle,” “transducing particle,” or “transduction particle” refers to a replication-defective or replication-competent virus capable of transducing non-viral DNA or RNA into a cell. In one particularly useful embodiment, retroviral particles used to produce transgenic avians in accordance with the invention are made as disclosed in U.S. Pat. No. 7,524,626, issued Apr. 28, 2009, the disclosure of which is incorporated in its entirety herein by reference.

[0075] The terms “transformation,” “transduction” and “transfection” all denote the introduction of a polynucleotide into an avian blastodermal cell. “Magnum” is that part of the oviduct between the infundibulum and the isthmus containing tubular gland cells that synthesize and secrete the egg white proteins of the egg.

[0076] The term “transgene” refers to heterologous nucleotide sequence inserted into an avian genome in accordance with the invention. “Transgene” can specifically refer to an exogenous coding sequence, an exogenous coding sequence linked to an exogenous promoter or other regulatory sequence, all nucleotide sequence between two retroviral LTRs and/or retroviral LTRs and nucleotide sequence between the LTRs wherein the LTRs are of a retrovirus used to introduce the transgene.

[0077] The term “optimized” is used in the context of “optimized coding sequence”, wherein the most frequently used codons for each particular amino acid found in the egg white proteins ovalbumin, lysozyme, ovomucoid, and ovotransferrin are used in the design of the optimized human interferon- α 2b (IFN- α 2b) polynucleotide sequence that is inserted into vectors of the present invention. More specifically, the DNA sequence for optimized human IFN- α 2b is based on the hen oviduct optimized codon usage and is created using the BACKTRANSLATE program of the Wisconsin Package, Version 9.1 (Genetics Computer Group Inc., Madison, Wis.) with a codon usage table compiled from the chicken (*Gallus gallus*) ovalbumin, lysozyme, ovomucoid, and ovotransferrin proteins. For example, the percent usage for the four codons of the amino acid alanine in the four egg white proteins is 34% for GCU, 31% for GCC, 26% for GCA, and 8% for GCG. Therefore, GCU is used as the codon for the majority of alanines in an optimized coding sequence. The vectors containing the gene for the optimized human protein are used to produce transgenic avians that express transgenic poultry derived protein in their tissues and eggs.

[0078] As used herein, the term “subject” encompasses mammals and non-mammals. Examples of mammals include, but are not limited to, humans, chimpanzees, apes, monkeys, cattle, horses, sheep, goats, swine; rabbits, dogs, cats, rats, mice, guinea pigs, and the like.

[0079] As used herein, the term “therapeutically effective amount” refers to any amount of a compound which, as compared to a corresponding subject who has not received such amount, results in improved treatment, healing, prevention, or amelioration of a disease, disorder, or side effect, or a decrease in the rate of advancement of a disease or disorder. The term also includes within its scope amounts effective to enhance normal physiological function.

[0080] The term “treat,” “treating” or “treatment” refers to methods of alleviating, abating or ameliorating a disease or condition symptoms, preventing additional symptoms, ameliorating or preventing the underlying causes of symptoms, inhibiting the disease or condition, arresting the development of the disease or condition, relieving the disease or condition, causing regression of the disease or condition, relieving a

condition caused by the disease or condition, or stopping the symptoms of the disease or condition either prophylactically and/or therapeutically.

LAL Compositions

[0081] The invention is generally drawn to compositions comprising enzymes useful for therapy, for example, in the treatment of lysosomal storage diseases. In one aspect, the invention is drawn to lysosomal storage disease enzymes such as LAL with a glycosylation pattern that renders the molecule amenable for internalization by certain cell types. Also included in the invention are recombinant human proteins including LAL in isolated or purified form. The isolation of the lysosomal storage disease enzymes (such as LAL) can be accomplished by methodologies readily apparent to a practitioner skilled in the art of protein purification.

[0082] In one embodiment, the invention is directed to lysosomal storage disease enzymes including, but not limited to LAL, having an N-linked glycosylation pattern described herein.

[0083] In one aspect, the compositions disclosed herein comprise human LAL wherein a substantial percentage of the human LAL contain a mannose-6-phosphate glycan moiety, which can serve as a ligand for internalization by the mannose-6-phosphate receptor on the surface of cells found, for example, on hepatocytes. In one embodiment, 30% or more, for example, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, or at least 99%, of the LAL contained in the composition contains at least one mannose-6-phosphate moiety. The mannose-6-phosphate moiety can be found, for example, on an N-glycan structure located at one or more residues selected from the group consisting of Asn¹⁵, Asn⁵¹, Asn⁸⁰, Asn¹⁴⁰, Asn²⁵² and Asn³⁰⁰ of SEQ ID NO:2. Glycan structures containing mannose-6-phosphate moieties include, for example, G-n and H-n shown in FIG. 16.

[0084] The recombinant human LAL according to the present invention contains multiple N-linked carbohydrate chains (e.g., about 5 or 6 carbohydrate chains). N-linked glycosylation structures at each of the five or six sites can be selected from one of A-n, B-n, C-n, D-n, E-n, F-n, G-n, H-n, I-n, J-n, K-n, L-n, M-n, N-n and O-n as shown in FIG. 16.

[0085] Also described herein are a mixture of LAL molecules (e.g., more than one LAL molecule can be present in a mixture such as the LAL molecules set forth in SEQ ID NOs: 2, 3, 4 and 19) wherein some or all of the LAL molecules have one or more glycosylation structures selected from Structure A-n, Structure B-n, Structure C-n, Structure D-n, Structure E-n, Structure F-n, Structure G-n, Structure H-n, Structure I-n, Structure J-n, Structure K-n, Structure L-n, Structure M-n, Structure N-n and Structure O-n (FIG. 16). In one embodiment, the mixture of lysosomal acid lipase molecules is purified or isolated, for example, isolated from an egg or purified or isolated from egg white produced in a transgenic avian.

[0086] The invention also includes an individual LAL molecule comprising a Structure A-n. The invention also includes an individual LAL molecule comprising a Structure B-n. The invention also includes an individual LAL molecule comprising a Structure C-n. The invention also includes an individual LAL molecule comprising a Structure D-n. The invention also includes an individual LAL molecule comprising a Structure E-n. The invention also includes an individual LAL molecule comprising a Structure F-n. The invention also

includes an individual LAL molecule comprising a Structure G-n. The invention also includes an individual LAL molecule comprising a Structure H-n. The invention also includes an individual LAL molecule comprising a Structure I-n. The invention also includes an individual LAL molecule comprising a Structure J-n. The invention also includes an individual LAL molecule comprising a Structure K-n. The invention also includes an individual LAL molecule comprising a Structure L-n. The invention also includes an individual lysosomal acid lipase molecule comprising a Structure M-n. The invention also includes an individual LAL molecule comprising a Structure N-n. The invention also includes an individual LAL molecule comprising a Structure O-n.

[0087] N-linked oligosaccharides attached to human LAL according to the present invention have a paucity of terminal sialic acid and galactose residues. That is, only minor amounts of the N-linked oligosaccharide structures are terminally sialylated and few galactose residues are present as well. Further, terminal N-Acetyl Glucosamine (GlcNAc) is present extensively on the N-linked oligosaccharide structures of the LAL described herein. As such, LAL produced in accordance with the invention can be targeted to cells such as monocyte macrophages and Kupffer cells.

[0088] One aspect of the invention provides compositions of LAL having essentially no sialic acid. In another aspect, the compositions disclosed herein comprise recombinant human LAL wherein a substantial percentage of the human LAL does not contain a sialic acid moiety in any of its N-glycan structures, which can interfere with internalization of the enzyme into cells. In one embodiment, 15% or less, for example, 10% or less, 5% or less, 2% or less, 1% or less, or essentially none, of the LAL contained in the composition contains a sialic acid moiety in any of its N-glycan structures.

[0089] In another embodiment, about 95% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain sialic acid. In another embodiment, about 90% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain sialic acid. In another embodiment, about 80% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain sialic acid. In another embodiment, more than about 70% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain sialic acid.

[0090] In still another embodiment, essentially none of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain sialic acid. In another embodiment, about 90% or more of the N-linked oligosaccharides structure types found to be associated with the LAL molecules of the invention do not contain sialic acid. For example, if there are 20 oligosaccharide structure types, then 18 or more of the structure types do not contain sialic acid. In another embodiment, about 80% or more of the N-linked oligosaccharides structure types found to be associated with the LAL molecules of the invention do not contain sialic acid. In another embodiment, about 70% or more of the N-linked oligosaccharides structure types found to be associated with the LAL molecules of the invention do not contain sialic acid. In another embodiment, about 60% or more of the N-linked oligosaccharides structure types found to be associated with the LAL molecules of the invention do not contain sialic acid. In another embodiment, about 50% or more of the N-linked

oligosaccharides structure types found to be associated with the LAL molecules of the invention do not contain sialic acid.

[0091] According to one aspect of the invention, LAL as described herein contain high levels of terminal N-Acetyl Glucosamine. In one aspect, about 95% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 90% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 80% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 70% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 60% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 50% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention contain a terminal N-Acetyl Glucosamine.

[0092] In one embodiment, all of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 90% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine. For example, if there are 20 oligosaccharide structure types, then 18 or more of the structure types do not contain a terminal N-Acetyl Glucosamine. In another embodiment, about 80% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 70% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 60% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine. In another embodiment, about 50% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain a terminal N-Acetyl Glucosamine.

[0093] In another aspect of the invention, the compositions disclosed herein comprise human LAL wherein a substantial percentage of the human LAL does not contain a fucose moiety in any of its N-glycan structure. In one embodiment, 50% or less, for example, 50% or less, 40% or less, 30% or less, 20% or less, 10% or less, 5% or less, 2% or less, 1% or less, or essentially none, of the LAL contained in the composition contains a fucose moiety in any of its N-glycan structure.

[0094] In one embodiment, fucose is essentially not present on the N-linked oligosaccharide structures of the LAL produced in accordance of the invention. In another embodiment, about 95% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain fucose. In another embodiment, about 90% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain fucose. In another embodiment, about 85% or more of the N-linked oligosaccharides present on the individual LAL molecule of the inven-

tion do not contain fucose. In another embodiment, about 80% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain fucose. In another embodiment, about 70% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain fucose. In another embodiment, about 60% or more of the N-linked oligosaccharides present on the individual LAL molecule of the invention do not contain fucose. In another embodiment, about 50% or more of the N-linked oligosaccharides present on the LAL of the invention do not contain fucose.

[0095] In one embodiment, essentially none of the N-linked oligosaccharides structure types present on the LAL molecules of the invention contain fucose. In another embodiment, about 95% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention do not contain fucose. For example, if there are 20 oligosaccharide structure types, then 19 or more of the structure types do not contain fucose. In another embodiment, about 90% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention do not contain fucose. In another embodiment, about 85% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention do not contain fucose. In another embodiment, about 80% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention do not contain fucose. In another embodiment, about 70% or more of the N-linked oligosaccharides structure types present on the LAL molecules of the invention do not contain fucose.

[0096] As discussed above, certain monosaccharides are abundantly present in LAL molecules produced in accordance with the present invention. The total monosaccharide species analyzed includes fucose, N-acetyl galactosamine, N-acetyl glucosamine, galactose, glucose, mannose, mannose-6-phosphate, N-acetyl neuraminic acid and N-glycolyl neuraminic acid. Fucose can be present between about 0% and about 1% of the total monosaccharide composition. N-acetyl galactosamine can be present between about 0% and about 1% of the total monosaccharide composition. N-acetyl glucosamine can be present between about 35% and about 50% of the total monosaccharide composition. Galactose can be present between about 1-10% of the total monosaccharide composition. Glucose is present at 0% of the total monosaccharide composition. Mannose is present between about 32% and about 50% of the total monosaccharide composition. Mannose-6-phosphate is present between about 1% and about 11% of the total monosaccharide composition.

[0097] In one embodiment, LAL produced in accordance with the present invention do not contain any xylose. In addition, because there is essentially no N-acetyl galactosamine (GalNAc) in LAL produced in accordance with the invention, one aspect of the invention includes a composition of LAL having no O-linked glycosylation.

[0098] LAL has 6 potential sites in its amino acid sequence for N-linked glycosylation, for example Asn³⁶, Asn⁷², Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³, and Asn³²¹ as in SEQ ID NO:1. Five of these, Asn³⁶, Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³ and Asn³²¹ are glycosylated while Asn⁷² can be unglycosylated or substantially unglycosylated (substantially unglycosylated means in a mixture of LAL molecules, fewer Asn⁷² are glycosylated than any of Asn³⁶, Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³ and Asn³²¹) (see FIG. 17). Accordingly, one aspect of the invention is a composition of LAL which is unglycosylated and/or substantially ungly-

cosylated at Asn⁷². LAL having a glycosylated Asn⁷² is within the scope of the invention. The positions of Asn described herein are based on the LAL amino acid sequence set forth in SEQ ID NO:1. It will be apparent to those skilled in the art that the numbering of Asn (i.e., the position of asparagine) can vary depending on individual LAL molecule and be readily determined in other LAL molecules such as those whose amino acid sequences are set forth in SEQ ID NOs:2, 3, 4 and 19.

[0099] The LAL molecules produced in accordance with the present invention contain N-glycan structures comprising a mixture of bi-, tri- and tetraantennary structures with N-acetylglucosamine, mannose and mannose-6-phosphate (M6P) as the major sugars (FIGS. 16 and 17). According to one aspect of the invention, M6P-modified N-glycans reside at least at Asn¹⁰¹, Asn¹⁶¹ and Asn²⁷³. Thus, one embodiment of the present invention includes a composition of LAL having M6P-modified N-glycans residing at any one of Asn¹⁰¹, Asn¹⁶¹ or Asn²⁷³. In yet another embodiment, the present invention includes a composition of LAL having M6P-modified N-glycans residing at Asn²⁷³. In another embodiment, the present invention includes a composition of LAL having monophosphorylated N-glycans (M6P) residing at any one of Asn¹⁰¹, Asn¹⁶¹ or Asn²⁷³. In yet another embodiment, the present invention includes a composition of LAL having monophosphorylated N-glycans residing at Asn¹⁶¹ and Asn²⁷³. In yet another embodiment, the present invention includes a composition of LAL having monophosphorylated N-glycans residing at Asn¹⁰¹ and Asn²⁷³. In one specific embodiment, a LAL produced in accordance with the present invention can contain bisphosphorylated mannose (bis-M6P) at Asn¹⁰¹.

[0100] The LAL molecules produced in accordance with the present invention contain reduced levels of galactose (e.g., "Gal"). One aspect of the present invention includes a composition of LAL having terminal galactose at any one of Asn³⁶, Asn¹⁶¹ or Asn³²¹. In yet another embodiment, the present invention includes a composition of LAL having terminal galactose at Asn³⁶ and Asn¹⁶¹. In yet another embodiment, the present invention includes a composition of LAL having terminal galactose at Asn¹⁶¹ and Asn³²¹. In yet another embodiment, the present invention includes a composition of LAL having terminal galactose at Asn³⁶ and Asn³²¹. In yet another embodiment, the present invention includes a composition of LAL having terminal galactose at Asn³⁶, Asn¹⁶¹ and Asn³²¹. In yet another embodiment, the present invention includes a composition of LAL having no terminal galactose.

[0101] Various types of N-glycans were found in LAL at different N-linked glycosylation sites. The N-glycan structures include a mixture of bi-, tri- and tetraantennary structures with N-acetylglucosamine, mannose and mannose-6-phosphate (M6P) as the major sugars. Specifically, in one embodiment of the present invention, LAL contains an N-glycan structure selected from GlcNAc4Man3GlcNAc2 or Gal1GlcNAc4Man3GlcNAc2 at the first N-linked glycosylation site (e.g., Asn³⁶ as in SEQ ID NO:1). In another embodiment, LAL contains no glycosylation or is substantially unglycosylated at the second N-linked glycosylation site (e.g., Asn⁷² as in SEQ ID NO:1). In yet another embodiment, LAL contains Phos2Man7GlcNAc2 at its third N-linked glycosylation site (e.g., Asn¹⁰¹ as in SEQ ID NO:1). In yet another embodiment, LAL contains a N-glycan structure selected from Phos1Man6GlcNAc2, GlcNAc1Phos1Man6GlcNAc2, Man3GlcNAc2, GlcNAc2Man3GlcNAc2,

GlcNAc3Man3GlcNAc2, GlcNAc4Man3GlcNAc2, or Gal1GlcNAc4Man3GlcNAc2 at its fourth N-linked glycosylation site (e.g., Asn¹⁶¹ as in SEQ ID NO:1). In yet another embodiment, LAL contains a N-glycan structure selected from Man7GlcNAc2, Man8GlcNAc2, Man9GlcNAc2, Phos1Man8GlcNAc2, or Phos1Man9GlcNAc2 at its fifth N-linked glycosylation site (e.g., Asn²⁷³ as in SEQ ID NO:1). In yet another embodiment, LAL contains a N-glycan structure selected from GlcNAc2Man3GlcNAc2, GlcNAc3Man3GlcNAc2, GlcNAc4Man3GlcNAc2, Gal1GlcNAc4Man3GlcNAc2, GlcNAc5Man3GlcNAc2, Gal1GlcNAc5Man3GlcNAc2, GlcNAc6Man3GlcNAc2, or Gal1GlcNAc6Man3GlcNAc2 at its sixth N-linked glycosylation site (e.g., Asn³²¹ as in SEQ ID NO:1).

[0102] According to certain aspects of the invention, compositions of LAL include LAL glycosylated at Asn³⁶, Asn⁷², Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³ and Asn³²¹ of SEQ ID NO:1 (or corresponding Asparagine residues within SEQ ID NOs: 2, 3, 4, and 19) with one N-glycan at the designated Asn position as shown below:

[0103] a) at Asn³⁶, GlcNAc4Man3GlcNAc2, or

[0104] Gal1GlcNAc4Man3GlcNAc2;

[0105] b) at Asn⁷², no glycosylation;

[0106] c) at Asn¹⁰¹, Phos2Man7GlcNAc2;

[0107] d) at Asn¹⁶¹, Phos1Man6GlcNAc2,

[0108] GlcNAc1Phos1Man6GlcNAc2;

[0109] Man3GlcNAc2;

[0110] GlcNAc2Man3GlcNAc2;

[0111] GlcNAc3Man3GlcNAc2;

[0112] GlcNAc4Man3GlcNAc2, or

[0113] Gal1GlcNAc4Man3GlcNAc2;

[0114] e) at Asn²⁷³, Man7GlcNAc2,

[0115] Man8GlcNAc2,

[0116] Man9GlcNAc2,

[0117] Phos1Man8GlcNAc2, or

[0118] Phos1Man9GlcNAc2; and

[0119] f) at Asn³²¹, GlcNAc2Man3GlcNAc2,

[0120] GlcNAc3Man3GlcNAc2,

[0121] GlcNAc4Man3GlcNAc2,

[0122] Gal1GlcNAc4Man3GlcNAc2,

[0123] GlcNAc5Man3GlcNAc2,

[0124] Gal1GlcNAc5Man3GlcNAc2,

[0125] GlcNAc6Man3GlcNAc2, or

[0126] Gal1GlcNAc6Man3GlcNAc2,

where Man=mannose,

[0127] GlcNAc=N-Acetyl Glucosamine,

[0128] Phos=phosphate, and

[0129] Gal=galactose.

[0130] In one embodiment, Gal1GlcNAc4Man3GlcNAc2 can be found as a glycan component at any one of Asn³⁶, Asn¹⁶¹ or Asn³²¹ in LAL produced in accordance with the invention. In one specific embodiment, Gal1GlcNAc4Man3GlcNAc2 can be found as a glycan component of Asn³⁶, Asn¹⁶¹ and Asn³²¹.

[0131] In the LAL of the present invention, Asn¹⁰¹ and Asn²⁷³ display the high-mannose-type having about 6 to about 10 mannose molecules (MAN6-MAN10 as described herein) as a major component. Accordingly, one aspect of the present invention includes a composition of LAL having a high mannose structure at Asn¹⁰¹ or Asn²⁷³. In another embodiment, a composition of LAL of the invention can comprise a N-glycan structure having at least 6 mannose at Asn¹⁰¹ or Asn²⁷³. In another embodiment, a composition of LAL contains a N-glycan having 7, 8 or 9 mannose at Asn¹⁰¹

or Asn²⁷³. In yet another embodiment, the present invention includes a composition of LAL having 7, 8 or 9 mannose at Asn¹⁰¹ and Asn²⁷³. In yet another embodiment, the present invention includes a composition of LAL having 7, 8 or 9 mannose at Asn¹⁰¹ and/or Asn²⁷³ and at least one of the mannose is phosphorylated.

[0132] It is to be understood that the glycosylation sites and the numbers associated with Asn described above is based on the amino acid sequence of LAL set forth in SEQ ID NO:1 and that the glycosylation profiles described above in context of SEQ ID NO:1 also apply to LAL molecules set forth in SEQ ID NOs: 2, 3, 4 and 19 though the numbering of corresponding Asn may vary depending on LAL molecule. For example, Asn³⁶ in SEQ ID NO:1 corresponds to Asn¹⁵ in SEQ ID NO:2, Asn¹³ in SEQ ID NO:3, Asn¹⁰ in SEQ ID NO:4 and Asn⁹ in SEQ ID NO:19. Asn⁷² in SEQ ID NO:1 corresponds to Asn⁵¹ in SEQ ID NO:2, Asn⁴⁹ in SEQ ID NO:3, Asn⁴⁶ in SEQ ID NO:4 and Asn⁴⁵ in SEQ ID NO:19. Asn¹⁰¹ in SEQ ID NO:1 corresponds to Asn⁸⁰ in SEQ ID NO:2, Asn⁷⁸ in SEQ ID NO:3, Asn⁷⁵ in SEQ ID NO:4 and Asn⁷⁴ in SEQ ID NO:19. Asn¹⁶¹ in SEQ ID NO:1 corresponds to Asn¹⁴⁰ in SEQ ID NO:2, Asn¹³⁸ in SEQ ID NO:3, Asn¹³⁵ in SEQ ID NO:4 and Asn¹³⁴ in SEQ ID NO:19. Asn²⁷³ of SEQ ID NO:1 corresponds to Asn²⁵² in SEQ ID NO:2, Asn²⁵⁰ in SEQ ID NO:3, Asn²⁴⁷ in SEQ ID NO:4 and Asn²⁴⁶ in SEQ ID NO:19. Asn³²¹ of SEQ ID NO:1 corresponds to Asn³⁰⁰ in SEQ ID NO:2, Asn²⁹⁸ in SEQ ID NO:3, Asn²⁹⁵ in SEQ ID NO:4 and Asn²⁹⁴ in SEQ ID NO:19.

[0133] For example, in one embodiment, the LAL is N-linked glycosylated at least at one position selected from the group consisting of Asn¹⁵, Asn⁵¹, Asn⁸⁰, Asn¹⁴⁰, Asn²⁵² and Asn³⁰⁰ of SEQ ID NO:2. In another embodiment, the LAL is N-linked glycosylated at Asn¹⁵, Asn⁸⁰, Asn¹⁴⁰, Asn²⁵² and Asn³⁰⁰ of SEQ ID NO:2. In yet another embodiment, N-glycan structures of LAL of SEQ ID NO:2 have no xylose while less than 15%, 10%, 5%, or 1% of N-glycan structures contain sialic acid; less than 50%, 40%, 30%, 20%, 10%, 5% or 1% of N-glycan structures contain fucose; and at least 30%, 50%, 60%, 70%, 80%, 90% and 95% of N-glycan structures contain phosphorylated mannose (M6P). S

[0134] In one embodiment, the LAL is N-linked glycosylated at least at one position selected from the group consisting of Asn¹³, Asn⁴⁹, Asn⁷⁸, Asn¹³⁸, Asn²⁵⁰ and Asn²⁹⁸ of SEQ ID NO:3. In another embodiment, the LAL is N-linked glycosylated at Asn¹³, Asn⁷⁸, Asn¹³⁸, Asn²⁵⁰ and Asn²⁹⁸ of SEQ ID NO:3. In yet another embodiment, N-glycan structures of LAL of SEQ ID NO:3 have no xylose while less than 15%, 10%, 5%, or 1% of N-glycan structures contain sialic acid; less than 50%, 40%, 30%, 20%, 10%, 5% or 1% of N-glycan structures contain fucose; and at least 30%, 50%, 60%, 70%, 80%, 90% and 95% of N-glycan structures contain phosphorylated mannose (M6P).

[0135] In one embodiment, the LAL is N-linked glycosylated at least at one position selected from the group consisting of Asn¹⁰, Asn⁴⁶, Asn⁷⁵, Asn¹³⁵, Asn²⁴⁷ and Asn²⁹⁵ of SEQ ID NO:4. In another embodiment, the LAL is N-linked glycosylated at Asn¹⁰, Asn⁷⁵, Asn¹³⁵, Asn²⁴⁷ and Asn²⁹⁵ of SEQ ID NO:4. In yet another embodiment, N-glycan structures of LAL of SEQ ID NO:4 have no xylose while less than 15%, 10%, 5%, or 1% of N-glycan structures contain sialic acid; less than 50%, 40%, 30%, 20%, 10%, 5% or 1% of N-glycan structures contain fucose; and at least 30%, 50%, 60%, 70%, 80%, 90% and 95% of N-glycan structures contain phosphorylated mannose (M6P).

[0136] In one embodiment, the LAL is N-linked glycosylated at least at one position selected from the group consisting of Asn⁹, Asn⁴⁵, Asn⁷⁴, Asn¹³⁴, Asn²⁴⁶ and Asn²⁹⁴ of SEQ ID NO:19. In another embodiment, the LAL is N-linked glycosylated at Asn⁹, Asn⁷⁴, Asn¹³⁴, Asn²⁴⁶ and Asn²⁹⁴ of SEQ ID NO:19. In yet another embodiment, N-glycan structures of LAL of SEQ ID NO:4 have no xylose while less than 15%, 10%, 5%, or 1% of N-glycan structures contain sialic acid; less than 50%, 40%, 30%, 20%, 10%, 5% or 1% of N-glycan structures contain fucose; and at least 30%, 50%, 60%, 70%, 80%, 90% and 95% of N-glycan structures contain phosphorylated mannose (M6P).

[0137] The composition according to the present invention can be produced a number of ways, including by use of transgenic avians, transgenic fish, transgenic mammals, for example, transgenic goats or in transgenic plants, such as tobacco and duck weed (*Lemna minor*) and certain types of cell culture.

[0138] The present invention also contemplates compositions comprising PEGylated LAL. LAL enzyme as described herein can be PEGylated as disclosed, for example, in U.S. Patent publication No. 20070092486, published Apr. 26, 2007, the disclosure of which is incorporated in its entirety herein by reference.

[0139] In one embodiment, the derived glycosylation pattern is obtained through expression specialized expression systems, for example, from avian oviduct cells, for example, tubular gland cells. For example, glycosylation patterns disclosed herein have been demonstrated to be present on lysosomal storage disease enzymes produced in oviduct cells of an avian such as a chicken in accordance with the present invention.

[0140] Proteins produced in accordance with the invention can be purified from egg white by any useful procedure such as those apparent to a practitioner of ordinary skill in the art of protein purification. For example, the human LAL (hLAL) produced in transgenic avians in accordance with the invention can be purified from egg white by methods apparent to practitioners of ordinary skill in the art of protein purification. An example of a purification protocol for LAL present in egg white is described in the Examples.

[0141] The invention includes the eggs and egg white and the avians (e.g., chicken turkey and quail) that lay the eggs and produce the egg white containing lysosomal acid lipase molecules of the invention comprising one or more of the glycosylation structures disclosed herein.

Expression of LAL in Avians

[0142] Disclosed herein are vectors and methods for the stable introduction of exogenous nucleic acid sequences into the genome of avians to express desired proteins such as those which benefit (e.g., attain an increased efficacy) from the addition of mannose-6-phosphatase such as lysosomal enzymes including, without limitation, lysosomal acid lipase (LAL) and other proteins such as those specifically disclosed herein. In particular, transgenic avians are produced which express exogenous sequences in their oviducts and which deposit exogenous proteins, such as pharmaceutical proteins, into their eggs. Avian eggs that contain such exogenous proteins are also described herein. Also disclosed herein are novel forms of LAL which are efficiently expressed in the oviduct of transgenic avians and deposited into avian eggs.

[0143] One aspect of the invention relates to compositions containing LAL, i.e., LAL molecules produced in accordance

with the invention. In a particularly useful embodiment, the LAL is purified or isolated. For example, the LAL has been removed from the contents of a hard shell egg laid by a transgenic avian. In one particularly useful embodiment, the LAL is human LAL. In one embodiment, the LAL of the invention has a glycosylation pattern resulting from the LAL being produced in an oviduct cell of an avian. For example, the compositions can contain a mixture of LAL molecules produced in avians, for example, chickens, in accordance with the invention and isolated from egg white. In one useful embodiment, the LAL containing compositions are pharmaceutical formulations.

[0144] In one aspect, the invention is drawn to compositions containing isolated LAL molecules, for example, human LAL molecules, wherein the LAL is produced in an avian which contains a transgene encoding the LAL. In one embodiment, the LAL is produced in an oviduct cell (e.g., a tubular gland cell) of a transgenic avian (e.g., transgenic chicken) and the LAL is isolated from egg white of the transgenic avian. In one embodiment, the LAL is glycosylated in the oviduct cell (e.g., tubular gland cell) of the bird, for example, a chicken.

[0145] In another aspect, methods for producing exogenous proteins such as lysosomal storage disease enzymes, for example, LAL, in specific tissues of avians, are provided. Such exogenous proteins may be expressed in the oviduct, blood and/or other cells and tissues of the avian. In one embodiment, transgenes are introduced into embryonic blastodermal cells, for example, near stage X, to produce a transgenic avian, such that the protein of interest is expressed in the tubular gland cells of the magnum of the oviduct, secreted into the lumen, and deposited into the egg white of a hard shell egg. A transgenic avian so produced can carry the transgene in its germ line providing transmission of the exogenous transgene to the avian's offspring stably in a Mendelian fashion.

[0146] The present invention encompasses methods of producing exogenous protein such as LAL in an avian oviduct. The methods may include a first step of providing a vector that contains a coding sequence and a promoter operably linked to the coding sequence, so that the promoter can effect expression of the nucleic acid in the avian oviduct. Transgenic cells and/or tissues can be produced, wherein the vector is introduced into avian embryonic blastodermal cells, either freshly isolated, in culture, or in an embryo, so that the vector sequence is inserted into the avian genome. A mature transgenic avian which expresses the exogenous protein such as LAL in its oviduct can be derived from the transgenic cells and/or tissue.

[0147] In one aspect of the invention, production of a transgenic avian is accomplished by transduction of embryonic blastodermal cells with replication-defective or replication-competent retroviral particles carrying the transgene between the 5' and 3' LTRs of the retroviral vector. For instance, an avian leukosis virus (ALV) retroviral vector or a murine leukemia virus (MLV) retroviral vector may be used which comprises a modified pNLB plasmid containing an exogenous gene that is inserted downstream of a segment of a promoter region. An RNA copy of the modified retroviral vector, packaged into viral particles, can be used to infect embryonic blastoderms which develop into transgenic avians.

[0148] Another aspect of the invention provides a vector which includes a coding sequence and a promoter in operational and positional relationship such that the coding sequence is expressed in an avian oviduct. Such vectors

include, but are not limited to, an avian leukosis virus (ALV) retroviral vector, a murine leukemia virus (MLV) retroviral vector, and a lentivirus vector. In addition, the vector may be a nucleic acid sequence which includes an LTR of an avian leukosis virus (ALV) retroviral vector, a murine leukemia virus (MLV) retroviral vector, or a lentivirus vector. The promoter is sufficient for effecting expression of the coding sequence in the avian oviduct. The coding sequence codes for an exogenous protein which is deposited into the egg white of a hard shell egg. As such, the coding sequence codes for exogenous proteins such as transgenic poultry derived proteins such as transgenic poultry derived lysosomal acid lipase (TPD LAL).

[0149] In one embodiment, vectors used in the methods of the invention contain a promoter which is particularly suited for expression of exogenous proteins in avians and their eggs. As such, expression of the exogenous coding sequence may occur in the oviduct and blood of the transgenic avian and in the egg white of its avian egg. The promoters include, but are not limited to, a cytomegalovirus (CMV) promoter, a rous-sarcoma virus (RSV) promoter, a β -actin promoter (e.g., a chicken β -actin promoter), a murine leukemia virus (MLV) promoter, a mouse mammary tumor virus (MMTV) promoter, an ovalbumin promoter, a lysozyme promoter, a conalbumin promoter, an ovomucoid promoter, an ovomucin promoter, and an ovotransferrin promoter. Optionally, the promoter may be a segment of at least one promoter region, such as a segment of the ovalbumin, lysozyme, conalbumin, ovomucoid, ovomucin, and ovotransferrin promoter region. In one embodiment, the promoter is a combination or a fusion of one or more promoters or a fusion of a portion of one or more promoters such as ovalbumin, lysozyme, conalbumin, ovomucoid, ovomucin, and ovotransferrin promoters.

[0150] In one embodiment, the vector includes a signal peptide coding sequence which is operably linked to the coding sequence, so that upon translation in a cell, the signal peptide directs secretion of the exogenous protein expressed by the vector, such as human LAL, into the egg white of a hard shell egg.

[0151] One aspect of the invention provides for coding sequences for exogenous proteins produced as disclosed herein wherein the coding sequence is codon optimized for expression in an avian, for example, in a chicken. Codon optimization may be determined from the codon usage of at least one, and preferably more than one, protein expressed in an avian cell (e.g., a chicken cell). For example, the codon usage may be determined from the nucleic acid sequences encoding the proteins ovalbumin, lysozyme, ovomucin and ovotransferrin of chicken. For example, the DNA coding sequence for the exogenous protein may be codon optimized using the BACKTRANSLATE® program of the Wisconsin Package, version 9.1 (Genetics Computer Group, Inc., Madison, Wis.) with a codon usage table compiled from the chicken (*Gallus gallus*) ovalbumin, lysozyme, ovomucoid, and ovotransferrin proteins.

[0152] One important aspect of the present invention relates to avian hard shell eggs (e.g., chicken hard shell eggs) which contain an exogenous peptide or protein including, but not limited to, a human LAL. The exogenous peptide or protein such as human LAL may be encoded by a transgene of a transgenic avian. Often, the exogenous peptide or protein (e.g., LAL) is glycosylated. The protein may be present in any useful amount. In one embodiment, the protein is present in an amount in a range of between about 0.01 μ g per hard-shell

egg and about 1 gram per hard-shell egg. In another embodiment, the protein is present in an amount in a range of between about 1 μ g per hard-shell egg and about 1 gram per hard-shell egg. For example, the protein may be present in an amount in a range of between about 10 μ g per hard-shell egg and about 1 gram per hard-shell egg (e.g., a range of between about 10 μ g per hard-shell egg and about 400 milligrams per hard-shell egg).

[0153] In one embodiment, the exogenous protein of the invention is present in the egg white of the egg. In one embodiment, the protein is present in an amount in a range of between about 1 ng per milliliter of egg white and about 0.2 gram per milliliter of egg white. For example, the protein may be present in an amount in a range of between about 0.1 μ g per milliliter of egg white and about 0.2 gram per milliliter of egg white (e.g., the protein may be present in an amount in a range of between about 1 μ g per milliliter of egg white and about 100 milligrams per milliliter of egg white. In one embodiment, the protein is present in an amount in a range of between about 1 μ g per milliliter of egg white and about 50 milligrams per milliliter of egg white. For example, the protein may be present in an amount in a range of about 1 μ g per milliliter of egg white and about 10 milligrams per milliliter of egg white (e.g., the protein may be present in an amount in a range of between about 1 μ g per milliliter of egg white and about 1 milligram per milliliter of egg white). In one embodiment, the protein is present in an amount of more than 0.1 μ g per milliliter of egg white. In one embodiment, the protein is present in an amount of more than 0.5 μ g per milliliter of egg white. In one embodiment, the protein is present in an amount of more than 1 μ g per milliliter of egg white. In one embodiment, the protein is present in an amount of more than 1.5 μ g per milliliter of egg white.

[0154] The avians of the invention which produce exogenous proteins disclosed herein (e.g., LAL) which are developed from the blastodermal cells into which the vector has been introduced are the G0 generation and can be referred to as "founders." Founder birds are typically chimeric for each inserted transgene. That is, only some of the cells of the G0 transgenic bird contain the transgene(s). The G0 generation typically is also hemizygous for the transgene(s). The G0 generation may be bred to non-transgenic animals to give rise to G1 transgenic offspring which are also hemizygous for the transgene and contain the transgene(s) in essentially all of the bird's cells. The G1 hemizygous offspring may be bred to non-transgenic animals giving rise to G2 hemizygous offspring or may be bred together to give rise to G2 offspring homozygous for the transgene. Substantially all of the cells of birds which are positive for the transgene that are derived from G1 offspring contain the transgene(s). In one embodiment, hemizygous G2 offspring from the same line can be bred to produce G3 offspring homozygous for the transgene. In one embodiment, hemizygous G0 or G1 animals, for example, are bred together to give rise to homozygous G1 offspring containing two copies of the transgene(s) in each cell of the animal. These are merely examples of certain useful breeding methods and the present invention contemplates the employment of any useful breeding method such as those known to individuals of ordinary skill in the art.

[0155] In one embodiment, the invention provides for the LAL to be isolated. That is, the LAL contained in the composition may be an isolated LAL. For example, the LAL may

be isolated from egg white. The isolated LAL may be LAL molecules having a variety of glycosylation structures among the LAL molecules.

[0156] By the methods of the present invention, transgenes can be introduced into avian embryonic blastodermal cells to produce a transgenic chicken, transgenic turkey, transgenic quail and other avian species, that carry the transgene in the genetic material of its germ-line tissue in order to produce proteins of the invention. The blastodermal cells are typically stage VII-XII cells, or the equivalent thereof, and in one embodiment are near stage X.

[0157] Some vectors useful in carrying out the methods of the present invention are described herein. In one embodiment, the coding sequence and the promoter of the vector are both positioned between 5' and 3' LTRs before introduction into blastodermal cells. In one embodiment, the vector is retroviral and the coding sequence and the promoter are both positioned between the 5' and 3' LTRs of the retroviral vector. In one useful embodiment, the LTRs or retroviral vector is derived from the avian leukosis virus (ALV), murine leukemia virus (MLV), or lentivirus.

[0158] In one embodiment, vectors are used for transfecting blastodermal cells and generating stable integration into the avian genome contain a coding sequence and a promoter in operational and positional relationship to express the coding sequence in the tubular gland cell of the magnum of the avian oviduct, wherein the exogenous protein such as an lysosomal enzyme (e.g., LAL) is deposited in the egg white of a hard shell egg.

[0159] The promoter may optionally be a segment of the ovalbumin promoter region which is sufficiently large to direct expression of the coding sequence in the tubular gland cells. Truncating the ovalbumin promoter and/or condensing the critical regulatory elements of the ovalbumin promoter so that it retains sequences required for expression in the tubular gland cells of the magnum of the oviduct, while being small enough that it can be readily incorporated into vectors is included within the scope of the invention. In one embodiment, a segment of the ovalbumin promoter region may be used. This segment comprises the 5'-flanking region of the ovalbumin gene.

[0160] The promoter may also be a promoter that is largely, but not entirely, specific to the magnum, such as the lysozyme promoter. The promoter may also be a mouse mammary tumor virus (MMTV) promoter. Alternatively, the promoter may be a constitutive promoter (e.g., a cytomegalovirus (CMV) promoter, a rous-sarcoma virus (RSV) promoter, a murine leukemia virus (MLV) promoter, etc.). In one embodiment, the promoter is a cytomegalovirus (CMV) promoter, a MDOT promoter, a rous-sarcoma virus (RSV) promoter, a murine leukemia virus (MLV) promoter, a mouse mammary tumor virus (MMTV) promoter, an ovalbumin promoter, a lysozyme promoter, a conalbumin promoter, an ovomucoid promoter, an ovomucin promoter and/or an ovotransferrin promoter. Optionally, the promoter may be at least one segment of a promoter region, such as a segment of the ovalbumin, lysozyme, conalbumin, ovomucoid, ovomucin, and ovotransferrin promoter region.

[0161] In one method of transfecting blastodermal cells, a packaged retroviral-based vector is used to deliver the vector into embryonic blastodermal cells so that the vector is integrated into the avian genome.

[0162] Useful retrovirus for randomly introducing a transgene into the avian genome is the replication-deficient avian

leucosis virus (ALV), the replication-deficient murine leukemia virus (MLV), or the lentivirus. In one embodiment, a pNLB vector is modified by inserting a region of the ovalbumin promoter and one or more exogenous genes between the 5' and 3' long terminal repeats (LTRs) of the retrovirus genome. The invention contemplates that any coding sequence placed downstream of a promoter that is active in tubular gland cells can be expressed in the tubular gland cells. For example, the ovalbumin promoter can be expressed in the tubular gland cells of the oviduct magnum because the ovalbumin promoter drives the expression of the ovalbumin protein and is active in the oviduct tubular gland cells.

[0163] Any of the vectors described herein can also optionally include a coding sequence encoding a signal peptide that directs secretion of the protein expressed by the vector's coding sequence from the tubular gland cells of the oviduct. This aspect effectively broadens the spectrum of exogenous proteins that may be deposited in avian eggs using the methods described herein. Where an exogenous protein would not otherwise be secreted, the vector containing the coding sequence is modified to comprise a DNA sequence comprising about 60 bp encoding a signal peptide from the lysozyme gene. The DNA sequence encoding the signal peptide is inserted in the vector such that it is located at the N-terminus of the protein encoded by the DNA.

[0164] Another aspect of the invention involves the use of internal ribosome entry site (IRES) elements in any of the vectors of the present invention to allow the translation of two or more proteins from a dicistronic or polycistronic mRNA. The IRES units are fused to 5' ends of one or more additional coding sequences which are then inserted into the vectors at the end of the original coding sequence, so that the coding sequences are separated from one another by an IRES.

[0165] In one embodiment when using an IRES, post-translational modification of the product is facilitated because one coding sequence can encode an enzyme capable of modifying the other coding sequence product. For example, the first coding sequence may encode collagen which would be hydroxylated and made active by the enzyme encoded by the second coding sequence wherein an IRES is employed as is understood in the art.

[0166] In another aspect, the coding sequences of vectors used in any of the methods of the present invention are provided with a 3' untranslated region (3' UTR) to confer stability to the RNA produced. When a 3' UTR is added to a retroviral vector, the orientation of the promoter, gene X and the 3' UTR must be reversed in the construct, so that the addition of the 3' UTR does not interfere with transcription of the full-length genomic RNA. In one embodiment, the 3' UTR may be that of the ovalbumin or lysozyme genes, or any 3' UTR that is functional in a magnum cell, i.e., the SV40 late region.

[0167] In one embodiment, a constitutive promoter is used to express the coding sequence of a transgene in the avian. In this case, expression is not limited to the magnum; expression also occurs in other tissues within the avian (e.g., blood). The use of such a transgene, which includes a constitutive promoter and a coding sequence, is particularly suitable for effecting or driving the expression of a protein in the oviduct and the subsequent secretion of the protein into the egg.

[0168] Transducing particles (i.e., transduction particles) are produced for the vector and titrated to determine the appropriate concentration that can be used to inject embryos. Avian eggs are windowed according to the Speksnijder procedure (U.S. Pat. No. 5,897,998, the disclosure of which is incorpo-

rated in its entirety herein by reference), and eggs are injected with transducing particles. Eggs hatch about 21 days after injection and male birds are selected for breeding. In order to screen for G0 roosters which contain the transgene in their sperm, DNA is extracted from rooster sperm samples. The G0 roosters with the highest levels of the transgene in their sperm samples are bred to nontransgenic hens by artificial insemination. Blood DNA samples are screened for the presence of the transgene. The serum of transgenic roosters is tested for the presence of exogenous protein. If the exogenous protein is confirmed, the sperm of the transgenic roosters is used for artificial insemination of nontransgenic hens. A certain percent of the offspring then contains the transgene (e.g., more than 50%). When exogenous protein is present in eggs produced in accordance with the present invention the protein may be isolated. The protein may also be tested for biological activity.

[0169] The methods of the invention which provide for the production of exogenous protein in the avian oviduct and the production of eggs which contain exogenous protein involve an additional step subsequent to providing a suitable vector and introducing the vector into embryonic blastodermal cells so that the vector is integrated into the avian genome. The subsequent step involves deriving a mature transgenic avian from the transgenic blastodermal cells produced in the previous steps. Mature transgenic avians can be obtained from the cells of a blastodermal embryo which has been transfected or transduced with the vector directly within the embryo. The resulting embryo is allowed to develop and the chick allowed to mature.

[0170] The transgenic avian produced from blastodermal cells is known as a founder. Some founders will carry the transgene in tubular gland cells in the magnum of their oviducts. These avians will express the exogenous protein encoded by the transgene in their oviducts. The exogenous protein may also be expressed in other tissues (e.g., blood) in addition to the oviduct. If the exogenous protein contains the appropriate signal sequence(s), it will be secreted into the lumen of the oviduct and into the egg white of the egg.

[0171] Some founders are germ-line founders. A germ-line founder is a founder that carries the transgene in genetic material of its germ-line tissue, and may also carry the transgene in oviduct magnum tubular gland cells that express the exogenous protein. Therefore, in accordance with the invention, the transgenic avian may have tubular gland cells expressing the exogenous protein, and the offspring of the transgenic avian may also have oviduct magnum tubular gland cells that express the exogenous protein. Alternatively, the offspring express a phenotype determined by expression of the exogenous gene in specific tissue(s) of the avian. In one embodiment, the transgenic avian is a chicken or a turkey.

Pharmaceutical Compositions & Therapeutic Methods

[0172] While it is possible that, for use in therapy, therapeutic proteins produced as described herein may be administered in raw form, it is preferable to administer the therapeutic proteins as part of a pharmaceutical formulation. Therefore, further provided are pharmaceutical formulations comprising poultry derived glycosylated therapeutic proteins such as LAL or a pharmaceutically acceptable derivative thereof together with one or more pharmaceutically acceptable carriers thereof and, optionally, other therapeutic and/or prophylactic ingredients and methods of administering such pharmaceutical formulations. The carrier(s) must be "accept-

able" in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. Methods of treating a patient (e.g., quantity of pharmaceutical protein administered, frequency of administration and duration of treatment period) using pharmaceutical compositions of the invention can be determined using standard methodologies known to physicians of skill in the art.

[0173] Compositions comprising carriers, including composite molecules, are formulated by well-known conventional methods (see, for example, Remington's Pharmaceutical Sciences, 14th Ed., Mack Publishing Co., Easton, Pa.), the entire teachings of which are incorporated herein by reference. The carrier may comprise a diluent. In one embodiment, the pharmaceutical carrier can be a liquid and the recombinant human LAL can be in the form of a solution. The pharmaceutical carrier can be wax, fat, or alcohol. In one embodiment, the wax- or fat-based carrier does not contain ester. In another embodiment, the pharmaceutically acceptable carrier may be a solid in the form of a powder, a lyophilized powder, or a tablet. In one embodiment, the carrier may comprise a liposome or a microcapsule.

[0174] The pharmaceutical formulations include those suitable for administration by injection including intramuscular, sub-cutaneous and intravenous administration. Pharmaceutical formulations include those suitable for oral, rectal, nasal, topical (including buccal and sub-lingual), vaginal or parenteral. The pharmaceutical formulations also include those for administration by inhalation or insufflation. The formulations may, where appropriate, be conveniently presented in discrete dosage units and may be prepared by any of the methods well known in the art of pharmacy. The methods of producing the pharmaceutical formulations typically include the step of bringing the therapeutic protein into association with liquid carriers or finely divided solid carriers or both and then, if necessary, shaping the product into the desired formulation.

[0175] Pharmaceutical formulations suitable for oral administration may conveniently be presented as discrete units such as capsules, cachets or tablets each containing a predetermined amount of the active ingredient; as a powder or granules; as a solution; as a suspension; or as an emulsion. The active ingredient may also be presented as a bolus, electuary or paste. Tablets and capsules for oral administration may contain conventional excipients such as binding agents, fillers, lubricants, disintegrants, or wetting agents. The tablets may be coated according to methods well known in the art. Oral liquid preparations may be in the form of, for example, aqueous or oily suspensions, solutions, emulsions, syrups or elixirs, or may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may contain conventional additives such as suspending agents, emulsifying agents, non-aqueous vehicles (which may include edible oils) or preservatives.

[0176] LAL may also be formulated for parenteral administration (e.g., by injection, for example, bolus injection or continuous infusion) and may be presented in unit dose form in ampoules, pre-filled syringes, small volume infusion or in multi-dose containers with an added preservative. The therapeutic proteins can be injected by, for example, subcutaneous injections, intramuscular injections, and intravenous infusions or injections.

[0177] The LAL may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing

and/or dispersing agents. It is also contemplated that the therapeutic protein may be in powder form, obtained by aseptic isolation of sterile solid or by lyophilization from solution, for constitution with a suitable vehicle, e.g., sterile, pyrogen-free water, before use.

[0178] For intravenous infusions or injection, the LAL produced in accordance of the invention can be formulated as an aqueous suspension or solution. Excipients suitable for the formulation for intravenous infusion or injection can include one of the following: trisodium citrate dehydrate, citric acid and human serum albumin. The pharmaceutical formulation can also include other suitable excipients well known in the art used for other products for lysosomal storage disorders. The pH of LAL produced in accordance with the invention is maintained between about 5.6 and about 6.2. Preferably, the pH of the LAL formulation is maintained at 5.9 ± 0.2 .

[0179] For topical administration to the epidermis, the therapeutic proteins of the invention produced according to the invention may be formulated as ointments, creams or lotions, or as a transdermal patch. Ointments and creams may, for example, be formulated with an aqueous or oily base with the addition of suitable thickening and/or gelling agents. Lotions can be formulated with an aqueous or oily base and can also contain one or more emulsifying agents, stabilizing agents, dispersing agents, suspending agents, thickening agents or coloring agents.

[0180] Formulations suitable for topical administration in the mouth include lozenges comprising active ingredient in a flavored base, usually sucrose and acacia or tragacanth; pastilles comprising the active ingredient in an inert base such as gelatin and glycerin or sucrose and acacia; and mouthwashes comprising the active ingredient in a suitable liquid carrier.

[0181] Pharmaceutical formulations suitable for rectal administration wherein the carrier is a solid are most preferably represented as unit dose suppositories. Suitable carriers include cocoa butter and other materials commonly used in the art, and the suppositories may be conveniently formed by a mixture of the active compound with the softened or melted carrier(s) followed by chilling and shaping in molds.

[0182] Formulations suitable for vaginal administration may be presented as pessaries, tampons, creams, gels, pastes, foams or sprays containing in addition to the active ingredient, such carriers as are known in the art to be appropriate.

[0183] For intra-nasal administration the therapeutic proteins of the invention may be used as a liquid spray or dispersible powder or in the form of drops. Drops may be formulated with an aqueous or non-aqueous base also comprising one or more dispersing agents, solubilizing agents or suspending agents. Liquid sprays are conveniently delivered from pressurized packs.

[0184] For administration by inhalation, therapeutic proteins according to the invention may be conveniently delivered from an insufflator, nebulizer or a pressurized pack or other convenient means of delivering an aerosol spray. Pressurized packs may comprise a suitable propellant such as dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol, the dosage unit may be determined by providing a valve to deliver a metered amount.

[0185] For administration by inhalation or insufflation, the therapeutic proteins according to the invention may take the form of a dry powder composition, for example, a powder mix of the compound and a suitable powder base such as lactose or starch. The powder composition may be presented in unit

dosage form in, for example, capsules or cartridges or, e.g., gelatin or blister packs from which the powder may be administered with the aid of an inhalator or insufflator.

[0186] When desired, the above described formulations adapted to give sustained release of the active ingredient, may be employed.

[0187] The pharmaceutical compositions described herein may also contain other active ingredients such as antimicrobial agents, or preservatives.

[0188] In addition, it is contemplated that the therapeutic proteins disclosed herein may be used in combination with other therapeutic agents. For example, the invention provides methods for pretreatment with a pharmaceutically effective dose of an antihistamine to minimize or prevent potential infusion-related anaphylactic reactions. For example, the antihistamine may be any pharmaceutically acceptable antihistamine (e.g. diphenhydramine) as disclosed herein and as known in the art. In one embodiment, the antihistamine is administered in a dose between about 1 mg and about 10 mg per kilogram of body weight. For example, the antihistamine may be administered in a dose of about 5 mg per kilogram. In one embodiment, the antihistamine is administered between about 10 minutes and about 90 minutes, for example, about 30 minutes to about 60 minutes, prior to administration of lysosomal acid lipase using an ambulatory system connected to the vascular access port. In one embodiment, the dose of diphenhydramine effectively counteracts potential anaphylactic infusion reactions.

[0189] Immunosuppressants such as antihistamines, corticosteroids, sirolimus, voclosporin, ciclosporin, methotrexate, IL-2 receptor directed antibodies, T-cell receptor directed antibodies, TNF- α directed antibodies or fusion proteins (infliximab, etanercept or adalimumab), CTLA4-Ig (e.g., abatacept), anti-OX-40 antibodies can also be administered before, during or after LAL administration if an anaphylactic reaction or adverse immune response is experienced by the patient.

[0190] The invention also contemplates therapy involving administration of LAL-containing compositions in combination with one or more cholesterol lowering agents (e.g., HMG-CoA reductase inhibitors). Non-limiting examples of such agents include: atorvastatin (Lipitor® and Torvast®), fluvastatin (Lescol®), lovastatin (Mevacor®, Altocor®, Altoprev®), pitavastatin (Livalo®, Pitava®), pravastatin (Pravachol®, Selektine®, Lipostat®), rosuvastatin (Crestor®) and simvastatin (Zocor®, Lipex®).

[0191] Compositions or proteins described herein can be used to treat a variety of conditions. For example, there are conditions for which treatment therapies are known to practitioners of skill in the art. The present invention contemplates that the therapeutic proteins (e.g., LAL) produced in an avian system containing a poultry derived glycosylation pattern can be employed to treat such conditions. That is, the treatment of conditions known to be treatable by conventionally produced therapeutic proteins by using therapeutic proteins produced as described herein is also contemplated. For example, LAL produced as described herein can be used to treat conditions resulting from or associated with LAL deficiency or insufficiency (collectively, "LAL deficiency"), such as Wolman disease and cholesteryl ester storage disease (CESD). As described herein, LAL deficiency also contemplates conditions in which expression of LAL is reduced due to a condition (e.g., a genetic mutation), physiological or environmental factors which leads to a reduction or deficiency of LAL

produced in the body. LAL produced as described herein can also be used to treat other conditions such as atherosclerosis, fatty liver disease, non-alcoholic fatty liver disease, nonalcoholic steatohepatitis (NASH) and cirrhosis. LAL produced as described herein can also be used to treat other conditions such as those disclosed in U.S. Pat. No. 6,849,257, issued Feb. 1, 2005, the disclosure of which is incorporated in its entirety herein by reference, U.S. publication No. 2009/0297496, published Dec. 3, 2009; US publication No. 2004/0223960, published Nov. 11, 2004; US publication No. 2007/0264249, published Nov. 15, 2009, the disclosures of which (i.e., the disclosures of each of these four patent publications) are incorporated in their entirety herein by reference.

[0192] It is also contemplated that LAL produced as disclosed herein can be used to treat certain specific conditions including pancreatitis, for example, chronic pancreatitis and/or acute pancreatitis as well as alcohol induced pancreatic injury such as alcohol induced pancreatitis.

[0193] LAL produced by any useful method, such as the ones disclosed herein, is contemplated for use to treat diseases due to alcohol induced cell injury including, but not limited to, those alcohol induced cell injuries that result in accumulation of lipid esters in body tissue such as, but not limited to, liver, spleen, gut and cardiovascular tissue. The invention also contemplates the treating of malabsorption by administering LAL.

[0194] One aspect of the invention is drawn to methods of treating a patient comprising administering to a patient a therapeutically effective amount of a composition comprising recombinant human LAL as described herein. The patient can be suffering or diagnosed with any number of conditions, including those associated with LAL deficiency. In one embodiment, the therapeutically effective amount is an amount that increases the red blood cell count in a patient by a desired amount. It is contemplated that LAL produced in accordance with the invention can be used to treat chronic kidney disease, for example, where tissues fail to sustain production of lysosomal acid lipase.

[0195] It is also contemplated that LAL produced by any useful method may be useful for the treatment of patients with Tangier disease and familial hypoalphalipoproteinemia. Tangier disease/familial hypoalphalipoproteinemia is associated with the accumulation of cholesterol esters in macrophages accompanied by hepatosplenomegaly and/or lymphadenopathy along with low levels of high-density lipoproteins (HDL) which can be treated by the administration of LAL. For example, without wishing to limit the invention to any particular theory or mechanism of operation, it is believed that impaired LAL activity decreases ABCA1 expression and conversely an increased LAL activity obtained by the administration of LAL to a patient with Tangier disease/familial hypoalphalipoproteinemia will increase ABCA1 expression to overcome the effects of an ABCA1 gene with a reduced functional activity as a result of polymorphism.

[0196] For the treatment of a condition, generally, the dosage administered can vary depending upon known factors such as age, health and weight of the recipient, type of concurrent treatment, frequency of treatment, and the like. Usually, a dosage of active ingredient can be between about 0.0001 and about 10 milligrams per kilogram of body weight. Precise dosage, frequency of administration and time span of treatment can be determined by a physician skilled in the art of administration of the respective therapeutic protein.

[0197] In addition, it has been discovered that dosages of 1 mg/kg and less can be effective in treating LAL deficiencies. The present invention provides methods of treating conditions comprising administering to a mammal (e.g. a patient, preferably a human patient) a therapeutically effective dose of lysosomal acid lipase between one time every 5 days and one time every 25 days, for example, between one time every 7 days and one time every 14 days. In one embodiment, the dose of lysosomal acid lipase administered is between about 0.1 mg and about 50 mg per kilogram of body weight, for example, the dose may be between about 1 mg and 5 mg per kilogram.

[0198] In one particularly useful embodiment, the invention provides methods of treating a condition by administering a dose of lysosomal acid lipase of between about 0.1 mg and 1.0 mg per kilogram of body weight in accordance with any therapeutically effective dosage regime such as those described herein.

[0199] The invention provides methods for treating any complication of LAL deficiency which may benefit from administering a therapeutically effective dose of LAL. In one embodiment, malabsorption and growth failure may be treated in accordance with the methods described herein. In another embodiment, complications seen in LAL deficiency patients including but not restricted to hepatomegaly and liver dysfunction may be treated using the methods provided herein.

[0200] The invention provides for treatment with recombinant LAL (e.g. recombinant human LAL) that can be produced by any useful protein expression system, for example, transgenic mammals and avians as is understood in the art. Other protein expression systems may include, but are not limited to, cell culture, bacteria, and plant systems.

[0201] The invention encompasses the administration of recombinant LAL as a part of a pharmaceutically acceptable composition by any route which may achieve the intended therapeutic effect, as determined by a physician skilled in the art. In one embodiment, the LAL may be administered by intravenous infusion over a period of about five hours. For example, the infusion may be facilitated by an ambulatory infusion pump connected to a vascular access port (e.g. a Port-a-Cath).

[0202] The invention also includes monitoring clinical and pathological presentation of the conditions, for example, Wolman Disease and CESD, in the mammal (e.g. the human patient). In one embodiment, the assessments consist of but are not limited to: lipid analysis, chest x-ray, liver function tests, stool chart, plasma mevalonic acid, immunogenicity, plasma lysosomal acid lipase, chitotriosidase, PARC, portal hypertension, anthropometry, volume and characterization of the liver, spleen, and gastrointestinal tract using, for example, imaging technology. For example, the aforementioned imaging technology may consist of ultrasound, magnetic resonance imaging, and nuclear magnetic resonance spectroscopy.

EXAMPLES

[0203] The present invention is further exemplified by the following examples. The examples are for illustrative purpose only and are not intended, nor should they be construed as limiting the invention in any manner.

Example 1

Construction of Vector
(pALVIN-OVR1-I-hLAL-dSA) Carrying
Recombinant Human Lysosomal Acid Lipase
(rhLAL) Coding Sequence

[0204] The nucleotide sequence of the hLAL gene in the pALVIN-OVR1-I-hLAL-dSA vector encodes a protein that is identical to the amino acid sequence of the protein produced by the human lysosomal acid lipase gene (GenBank Accession, NP_000226) (FIG. 1). Transcription of this sequence and subsequent translation of the resultant mRNA produces a 399 amino acid precursor protein, which is processed to a mature 378 amino acid protein identical to human LAL (GenBank Accession, NP_000226) (FIG. 1) as set forth in SEQ ID NO:1. Expression of the hLAL gene (see FIG. 2 for the cDNA sequence) in this Example is controlled by non-coding elements derived from the ovalbumin gene including enhancer, promoter, intronic, and 5' and 3' untranslated sequences. The ovalbumin gene produces ovalbumin, the major protein constituent of egg white. Activity of the chicken ovalbumin promoter is very specific to the cells within the chicken oviduct that produce egg white; expression in other tissues is minimal.

[0205] The plasmid vector pALVIN-OVR1-I-hLAL-dSA (FIG. 3A; the nucleotide sequence of which is shown in FIG. 4) was used to produce a replication-deficient retrovirus (RDR) that stably integrated the hLAL transgene into the genome of the founder (XLL109). This plasmid vector includes retroviral nucleotide sequences required for viral RNA packaging, reverse transcription and integration, but does not contain the intact sequences for the viral gag, pol and env genes. The methods used to generate the retroviral vector and their use in subsequent transgenesis procedures are described herein.

[0206] The retroviral portion of pALVIN-OVR1-I-hLAL-dSA is based on the ALV vector, pNLB. pNLB was modified such that the LTRs would be self-inactivating (SIN) (FIG. 3B). To accomplish this, 273 bp of the 3' LTR was deleted, which includes the enhancer and CAAT box of the U3 region. Because the inactivated U3 region at the 3' end of the retroviral sequence serves as a template for a new U3 region present at the 5' end of an integrated provirus, 5' LTR is normally also inactivated. The deletion of LTR sequences in the SIN construct decreases promoter interference on the internal promoter from the LTR, and minimizes the possibility for recombination of sequences to form a replication competent retrovirus. The new vector is termed pALVIN for ALV inactivation vector.

[0207] Downstream of the 5' LTR are the partial gag and env coding sequences, which were carried over from the pNLB vector. In pALVIN-OVR1-I-hLAL-dSA, a small portion (12%) of the gag protein precursor sequence remains (55% of the p19 mature peptide sequence) and a small portion (1.7%) of the env precursor sequence of RAV2 remains (GenBank Accession, AF033808). These truncated gag and env regions are unable to produce functional proteins needed to create replication competent retrovirus (Cosset, 1991).

[0208] Transcriptional and translational control elements of the chicken ovalbumin gene were inserted into pALVIN to create pALVIN-OV-1.1-I (sequence of which is shown in FIG. 6; SEQ ID NO: 8). The first section of pALVIN-OV-1.1-I is composed of a contiguous section of the chicken ovalbumin gene which includes the 1.1 kb proximal promoter

region, the first exon, first intron and part of the 2nd exon. The next section is a stuffer insert fragment that takes the place of the ovalbumin protein coding sequences. The stuffer is followed by the 3' untranslated region (UTR) of the chicken ovalbumin gene, which includes sequences that facilitate proper processing of the mRNA, including polyadenylation. In general, the stuffer fragment is replaced by DNA fragments encoding the desired protein, in this case hLAL. The result is a vector that has specific elements that promote regulated transcriptional expression and translation of an mRNA in the oviduct of transgenic chickens, that closely mimics regulation of the endogenous ovalbumin mRNA, and that allows high expression of the protein of interest in egg white.

[0209] The pALVIN-OV-1.1-I vector includes the first intron of the ovalbumin gene. Because the intron is susceptible to splicing during the production and packaging of the retroviral RNA genome, we inserted the expression cassette in the opposite orientation relative to the LTRs. In this way the intron is not recognizable in the retroviral RNA and is packaged without splicing. For convenience all maps in this document are drawn with the LTRs in the opposite orientation and the expression cassette in the forward or clockwise orientation.

[0210] pALVIN-OV-1.1-I is the base vector into which the coding sequence (CDS) of hLAL was inserted. Two DNA fragments, hLAL adaptor and Syn hLAL, which make up the hLAL CDS and sequences required for compatibility with pALVIN-OV-1.1-I, were synthesized at Integrated DNA Technologies, Coralville, Iowa, (see FIGS. 7 and 8; SEQ ID NOs: 9 and 10). A 229 bp HpaI/BamHI fragment of hLAL adaptor and a 1113 bp BamHI/BstBI fragment of Syn hLAL were inserted into the 7882 HpaI/BstBI fragment of pALVIN-OV-1.1-I, thereby replacing the stuffer region with the hLAL CDS and creating pALVIN-OV-1.1-I-hLAL.

[0211] It was discovered that there was a cryptic splice site in the antisense strand of the hLAL CDS which prevented packaging of intact retroviral RNA. The cryptic splice site was removed by alteration of the DNA sequence without changing the amino acid sequence of hLAL. This change was performed by polymerase chain amplification of region 232 to 534 of pALVIN-OV-1.1-I-hLAL with primer 5'-AGAACTGAGAGTGTCTTAT-3' (SEQ ID NO: 12) and primer 5'-TGACAGCTGTGGATCCAGAAACAAACATG-3' (SEQ ID NO: 13), creating a 329 bp amplicon. This amplicon was digested with BamHI and SexAI and ligated with the 8940 bp BamHI/SexAI fragment of pALVIN-OV-1.1-I-hLAL to create pALVIN-OV-1.1-I-hLAL-dSA.

[0212] A putative promoter enhancer which contains DNase hypersensitive site III (DHSIII) of the chicken ovalbumin gene (−3819 to −2169 relative to the OV promoter start site) (Kaye, Bellard et al. 1984) was inserted into pALVIN-OV-1.1-I-hLAL-dSA to create pALVIN-OVR1-I-hLAL-dSA. This was performed as follows: a DNA fragment which included the DHSIII enhancer and 1.1 kb proximal OV promoter termed OVR1 promoter (see FIG. 9; and SEQ ID NO: 11 for sequence) was isolated by digestion with XhoI and BlnI. To facilitate subcloning, an adaptor fragment, PCR of pSIN-OV-1.1-I was generated by PCR amplification of region 6752 to 7974 of pALVIN-OV-1.1-I with primers 5'-GCCGCTCGAGCGAGGAATATAAAAAATT-3' (SEQ ID NO: 14) and 5'-TCCGCGCACATTTCCCCGAA-3' (SEQ ID NO: 15) followed by digestion with NgoMI and XhoI. The 2772 bp XhoI/BlnI fragment of OVR1 promoter and 1067 bp NgoMI/XhoI fragment of PCR of pSIN-OV-1.1-I were

inserted into the 7043 bp NgoMI/BlpI fragment of pALVIN-OV-1.1-I-hLAL-dSA, thereby creating pALVIN-OVR1-I-hLAL-dSA (see FIG. 10 for the construction schematics of pALVIN-OVR1-I-hLAL-dSA). The construction of the retroviral vector segment of the vector, denoted as pALVIN (aka pAVIJCR-A395.22.3.1-KM or pALV-SIN), is described in United States Patent Application 2008/0064862.

[0213] In addition, included is the production of LAL in accordance with the invention using a promoter and/or vector disclosed in US patent publication No. 2008/0064862, published Mar. 13, 2008, the disclosure of which is incorporated in its entirety herein by reference.

Example 2

Viral Particle Production

[0214] The G0 founder transgenic male, XLL109, carrying the hLAL transgene in its genome, was created by using a retroviral transgenesis method as follows. Replication-defective viral particles carrying the pALVIN-OVR1-I-hLAL-dSA vector were produced by transient transfection of an immortalized chicken fibroblast cell line. These chicken fibroblast cells were simultaneously transfected with three plasmids, pALVIN-OVR1-I-hLAL-dSA, pCMV-gag-pol and pCMV-VSV-G. pCMV-gag-pol expresses the gag and pol genes of RAV1 strain of the avian leukosis virus. pCMV-VSV-G expresses the envelope protein of the vesicular stomatitis virus. Four hours after transfection, the media was replaced with DMEM supplemented with 10% fetal bovine serum, 100 units/mL penicillin and 100 µg/mL streptomycin. Media was harvested at 48 hours post-transfection, filtered through a 0.45 micron filter (Millipore) and concentrated by ultracentrifugation. Concentrated retrovirus carrying the ALVIN-OVR1-I-hLAL-dSA transgene was collected and used in the transduction of early stage embryos. Note that because “p” is the notation for the plasmid form of vector, the “p” is absent from the transgene designations once the transgene is in the form of packaged vector or integrated transgene.

Example 3

Embryo Transgenesis

[0215] Integration of the ALVIN-OVR1-I-hLAL-dSA expression cassette into the genome of an embryo was achieved by transduction of early stage embryos (Speksnijder and Ivarie, 2000). Freshly laid fertilized White Leghorn eggs were obtained from a breeding colony. An aperture was made in the shell to provide access to the embryo. Seven microliters of concentrated replication deficient retrovirus particles carrying the ALVIN-OVR1-I-hLAL-dSA expression cassette described above were injected into the subgerminal cavity of the embryo. Eggs were sealed with hot glue, and then incubated and hatched under standard conditions. Progeny produced from these injections were given individual identification markers at hatch for identification and traceability. Blood samples from the progeny were transgene positive when analyzed by real-time PCR for the hLAL transgene using PCR primers specific for the hLAL coding sequence (as described below). This gave an indication that the transgenesis procedure was successful. The real-time PCR assay for the hLAL transgene utilizes Taqman® chemistry (Applied Biosystems). The forward and reverse primers were 5'-ACGACTG-GCTTGCAATGCTCT-3' (SEQ ID NO: 16) and 5'-CCCCAAATGAAGTCAAGATGCT-3' (SEQ ID NO: 17),

respectively. The Taqman® probe sequence was 5'-CCG-GAATGCTCTCATGGAACACCAA-3' (SEQ ID NO: 18) and was labeled with FAM (as the emitter) at the 5' end and Iowa Black (as the quencher) at the 3' end. Primers, probe and 1 µl of extracted DNA was added to 30 µl Taqman® Universal Master Mix (Applied Biosystems). Control reactions included various dilutions of a plasmid bearing the hLAL sequence and DNA from wild-type chickens (data not shown). Standard cycling parameters were used on an Applied Biosystems 7500 Fast Real-Time PCR System.

Example 4

Identification of G0 Founder

[0216] Semen was collected from sexually mature males and DNA was extracted and assayed using the hLAL real-time PCR assay. The number of transgene copies in each sample was estimated using known standards (a plasmid bearing the hLAL gene) mixed with negative control semen DNA. The transgene cassette DNA content in male XLL109 was at a level that would allow transmission of the transgene to his progeny, as estimated by real-time PCR. This XLL109 male was the G0 transgenic founder and was bred with non-transgenic chickens to generate the G1 hemizygotic transgenic chickens.

Example 5

Propagation and Characterization of Hemizygotic G1 Avians

[0217] Progeny sired by the transgenic founder XLL109 were tested for the presence of the transgene in blood cell DNA using the hLAL real-time PCR assay. Blood was collected from 1-2 week old progeny and DNA was extracted using a high-throughput technique (Harvey et al., 2002). The DNA solutions were not quantified prior to the Taqman assay to facilitate the high-throughput screen. Typically 1 µl of DNA solution contains 50 to 400 ng of DNA which is sufficient to generate a positive amplification signal. A total of 1,322 chicks sired by XLL109 were tested, and positive progeny were re-bled and tested for confirmation. According to the PCR results, 22 progeny were positive for the ALVIN-OVR1-I-hLAL-dSA transgene. An example of the Taqman results is shown in FIG. 11.

Example 6

Identification and Characterization of High-Expression Line

[0218] One of the G1 chickens, 1LL7466, laid eggs with significantly higher levels of rhLAL protein in the egg white, as compared to the other G1 chickens. Southern blot analysis was performed on 1LL7466 and sibling G1 males to identify which sibling males had the same integration site as the high expressing chicken. Digests were performed with a restriction enzyme that cut only once within the transgene (BlpI), and the Southern blots were probed with a segment of the ovalbumin promoter or the hLAL coding sequence (FIGS. 12A-D). The position of the 2nd restriction site, which resides in the flanking genomic region, varies depending on the site of integration. Thus the size of the BlpI band detected by the OV probe or hLAL probe is unique to each line generated.

[0219] The OV probe detected a single band of 4.1 kb in BlnI-digested DNA from wild-type chickens, which corresponded to the expected size of a BlnI segment within the endogenous ovalbumin gene of the chicken genome (FIGS. 12B and 12D). A second band of 4.3 kb was detected with chicken 1LL7466, which corresponded to the transgene band. Three additional female siblings, 1LL10409, 1LL10686 and 1LL12058 and three additional male siblings, 1LL8922, 1LL9330 and 1LL11217 displayed the 4.3 kb band, indicating that these siblings might be of the same line (FIGS. 12B and 12D).

[0220] As expected the hLAL probe did not detect a band in DNA from wild-type chickens as the DNA sequence of the chicken lysosomal acid lipase gene and the coding sequence for the recombinant human lysosomal acid lipase are sufficiently differentiated to not permit hybridization under the conditions used in these Southern assays (FIG. 12C). The hLAL probe detected a single band of ~10.6 kb in BlnI-digested genomic DNA from the same chickens that were positive for the 4.3 kb band detected by the OV probe, indicating that these 7 G1 chickens have the same integration site and thus are of the same line. No other bands were detected, indicating that 1LL7466, 1LL10409, 1LL10686, 1LL12058, 1LL8922, 1LL9330 and 1LL11217 all had a single integration site.

[0221] The Southern analysis also indicated that the transgene was integrated as the bands detected by the OV and hLAL probes were of different sizes and greater in size than from the transgene alone. A map showing the predicted structure of the integrated transgene and position of BlnI sites in the flanking genomic regions is shown in FIG. 12A.

[0222] To confirm that the transgene is intact, two steps were taken. First, the hLAL coding sequence was isolated by PCR from 1LL7466. The PCR products were sequenced on both strands from the hLAL start codon to the stop codon. The DNA sequence was exactly as expected, indicating no changes in the DNA sequence of coding regions in the transgene. Second, Southern blot analysis was conducted using restriction enzyme ApaLI, which digests intact transgene into 2 segments, 3.6 and 3.8 kb (FIG. 13A). Both the 3.6 and 3.8 kb bands were detected in ApaLI-digested genomic DNA from G1s, indicating that the transgene was integrated in a fully intact form (FIG. 13B).

Example 7

Propagation and Characterization of G2s

[0223] FIG. 14 shows the lineage of the hLAL G2s descended from a single G0 founder, XLL109. At the G1 stage, the transgene was characterized with regard to copy number, integrity, hLAL sequence and integration site—and seven G1 transgenics were identified and characterized (four chickens and three roosters). Propagation of the G2s was accomplished by artificial insemination of non-transgenic chickens with semen collected from the G1 sires 1LL8922, 1LL9330 and 1LL11217 (FIG. 14). Each inseminated chicken, her eggs and subsequent progeny were housed separately from the other progeny. Hatched progeny were tested for presence of the hLAL transgene using the hLAL real-time PCR assay. Because G1 founders were hemizygous with respect to the transgene, half of the progeny were expected to be transgenic G2s. Of 610 G2 progeny analyzed to date, 330 or 54% were transgenic.

Example 8

Genetic Analysis of the hLAL Avians

[0224] After identification of each G2 chicken by the hLAL real-time PCR assay of blood DNA, the production line is subjected to the following genetic assays: the hLAL gene was PCR-amplified from blood DNA and sequenced to confirm 100% homology with the human sequence; the transgene integration site was confirmed by integration site PCR, as described above. The PCR sequencing and integration site analysis was performed on: each chicken in a <10 chicken production line; 10% of chickens (minimum 10) for 11-100 chicken production line; 5% of chickens (minimum 10) for 101-1000 chicken production line; 1% of chickens (minimum 50) for 1001-10,000 chicken production line; 0.1% of chickens (minimum 100) for >10,001 chicken production line. Detailed records were maintained at every step of the growing and production phase.

Example 9

Purification of hLAL from Egg White

[0225] Egg white (EW) containing LAL was solubilized at pH 6 overnight and clarified through centrifugation (or depth filtration) with 0.2 µm filtration. The EW was adjusted with 1 M NaOAc buffer (pH 4) to pH 6.

[0226] The clarified EW was loaded onto a Phenyl-HIC column (EW:column size=2:1) equilibrated with 20 mM phosphate/137 mM NaCl buffer (pH 6). After the completion of loading, the column was washed with equilibration buffer and 5 mM phosphate buffer (pH 6). The LAL was eluted with 30% propylene glycol with 5 mM Tris buffer (pH 7.2).

[0227] The eluted LAL fraction was adjusted to pH 5 with 1 M acid and then loaded onto a GigaCap S column (EW:column size=10:1). The column was equilibrated with 50 mM NaOAc buffer (pH 5). After completion of loading, the column was washed with the equilibration buffer. The LAL was eluted with 50 mM NaOAc/60 mM NaCl (pH 5).

[0228] The LAL fraction off the GigaCap S column was adjusted to pH6 with 1 M Tris buffer and then loaded onto a Butyl-HIC column (EW:column size=10:1). The column was equilibrated with 20 mM phosphate/137 mM NaCl buffer (pH 6). After the completion of loading, the column was washed with equilibration buffer and 5 mM phosphate buffer (pH 6). The pure LAL was eluted with 50% propylene glycol with 5 mM Tris buffer (pH 7.2). FIG. 15 depicts the purification steps of hLAL from egg white.

Example 10

Carbohydrate Analysis of Transgenic Avian Derived hLAL

[0229] The oligosaccharide structures were determined for avian derived human LAL by employing the following analysis techniques as are well known to practitioners of ordinary skill in the art.

[0230] Two hundred micrograms were digested with trypsin and chymotrypsin for 18 h at 37° C. in 0.1 M Tris-HCl, pH 8.2, containing 1 mM CaCl₂. The digestion products were enriched and freed of contaminants by Sep-Pak C18 cartridge column. After enrichment, the glycopeptides were digested with 2 µl of PNGaseF (7.5 unit/ml) in 50 µl of 20 mM sodium phosphate buffer, pH 7.5, for 18 h at 37° C. Released oli-

gosaccharides were separated from peptide and enzyme by passage through a Sep-Pak C18 cartridge column.

[0231] The glycan fraction was dissolved in dimethylsulfoxide and then permethylated based on the method of Anumula and Taylor (Anumula and Taylor, 1992). The reaction was quenched by addition of water and per-O-methylated carbohydrates were extracted with dichloromethane. Per-O-methylated glycans were dried under a stream of nitrogen.

[0232] MALDI/TOF-MS (Matrix assisted laser desorption ionization time-of-flight mass spectrometry) was performed in the reflector positive ion mode using α -dihydroxybenzoic acid (DHBA, 20 mg/mL solution in 50% methanol:water) as a matrix. All spectra were obtained by using a Microflex LRF (Bruker).

[0233] MALDI-TOF-MS analysis and ESI MS/MS (electrospray ionization tandem mass spectrometry) were performed on the oligosaccharides after release from the peptide backbone and purification as is understood in the art. Samples of the individual polysaccharide species were also digested with certain enzymes and the digest products were analyzed by HPLC as is understood in the art.

[0234] It is believed that there are about six N-linked glycosylation sites present on human LAL. See, Zschenker, et al (2005) *J. Biochem.*, Vol 137, p 387-394, the disclosure of which is incorporated in its entirety by reference. This reference also indicates that there may be an O-linked glycosylation site on Human LAL. The N-linked oligosaccharide structures identified are shown in FIG. 16.

[0235] The data revealed that many or all of these structures were found as an N-linked Glycosylation structure in LAL produced in accordance with the invention (FIG. 16). For example, A-n is found attached to LAL produced in accordance with the invention. For example, O-n is found attached to LAL produced in accordance with the invention. For example, at least one of B-n, C-n and D-n is found attached to LAL produced in accordance with the invention. For example, at least one of E-n and F-n is found attached to LAL produced in accordance with the invention. For example, at least one of I-n and J-n is found attached to LAL produced in accordance with the invention. For example, at least one of K-n and L-n is found attached to LAL produced in accordance with the invention. For example, at least one of M-n and N-n is found attached to LAL produced in accordance with the invention. For example, G-n is found attached to LAL produced in accordance with the invention. For example, H-n is found attached to LAL produced in accordance with the invention.

Example 11

N-Glycan Species of Transgenic Avian Derived LAL

[0236] Purified samples of transgenic avian derived hLAL (600 μ g/sample) were dialyzed using a Tube-O-Dialyzer (4.0 kDa cut-off membrane; G BioSciences) against nanopure water at 4° C. for about 24 hours to remove salts and other contaminants. Nanopure water was replaced four times during the entire dialysis period.

[0237] After dialysis, each of the samples was divided into three aliquots: ~1/4 of sample weight for neutral and amino sugars analysis, ~1/4 of sample weight for mannose-6-phosphate analysis, and ~1/2 of sample weight for oligosaccharide profiling. The aliquot intended for neutral and amino sugars analysis was hydrolyzed with 2 N trifluoroacetic acid (TFA) at 100° C. for 4 hours and the aliquot for mannose-6-phos-

phate analysis was hydrolyzed with 6.75 N TFA at 100° C. for 1.5 hours. The hydrolysates were then dried under N₂, redissolved with 50 μ L H₂O, sonicated for 7 min in ice and transferred to an injection vial. However, the neutral and amino sugar samples were diluted 2 times because the peaks produced from the originally dissolved hydrolysates were too large.

[0238] A mix of standards for neutral and amino sugars, and for mannose-6-phosphate with a known number of moles was hydrolyzed in the same manner and at the same time as the sample. Four concentration of the neutral and amino sugar standard mix (Fuc & GalNAc, 0.2, 0.4, 0.8, and 1.6 nmoles per 10 μ L; GlcNAc, 0.5, 1.0, 2.0, and 4.0 nmoles per 10 μ L; Gal & Man, 0.3, 0.6, 1.2, and 2.4 nmoles per 10 μ L; and Glc, 0.1, 0.2, 0.4, and 0.8 nmoles per 10 μ L) and mannose-6-phosphate (640, 1280, 2560, 5120 picomoles per 10 μ L) were prepared to establish a calibration equation. The number of moles of each sugar in the sample was quantified by linear interpolation from the calibration equation.

[0239] The neutral and amino sugars and mannose-6-phosphate were analyzed by HPAEC using a Dionex ICS3000 system equipped with a gradient pump, an electrochemical detector, and an autosampler. The individual neutral and amino sugars, and mannose-6-phosphate were separated by a Dionex CarboPac PA20 (3 $\times$ 150 mm) analytical column with an amino trap. The gradient programs used eluents A, degassed nanopure water and B, 200 mM NaOH for neutral and amino sugars, and C, 100 mM NaOH and D, 1 M sodium acetate in 100 mM NaOH for mannose-6-phosphate. Injection (10 μ L/injection) was made every 40 minutes for neutral and amino sugar determination and every 35 minutes for mannose-6-phosphate determination. All methods were based on protocols described by Hardy and Townsend (Hardy, M. R., and Townsend, R. R., "High-pH anion-exchange chromatography of glycoprotein-derived carbohydrates", 1994, *Methods Enzymol.* 230: 208-225). Instrument control and data acquisition were accomplished using Dionex chromeleon software. Results are shown in Table 1 below. The control sample is ovomucoid purified from EW.

TABLE 1

Monosaccharide composition of control and LAL by HPAEC.

Sample ID	Analyte	nano-moles	nano-moles/ μ g	mole %
Control	Fucose	nd	—	—
	N-acetyl galactosamine	5.066	0.020	9.6
	N-acetyl glucosamine	26.947	0.108	51.4
	Galactose	3.876	0.016	7.4
	Glucose	nd	—	—
	Mannose	16.565	0.066	31.6
	Mannose-6-phosphate	nd	—	—
	N-acetyl neuraminic acid	ndm	—	—
	N-glycolyl neuraminic acid	ndm	—	—
	Fucose	nd	—	—
Transgenic Avian derived hLAL	N-acetyl galactosamine	nd	—	—
	N-acetyl glucosamine	17.932	0.120	37.6
	Galactose	0.879	0.006	1.8
	Glucose	nd	—	—
	Mannose	23.290	0.155	48.8
	Mannose-6-phosphate	5.642	0.038	11.8
	N-acetyl neuraminic acid	ndm	—	—
	N-glycolyl neuraminic acid	ndm	—	—

nd = not detected; ndm = not determined.

Structural Features of LAL

[0240] LAL has 6 potential sites in its amino acid sequence for N-linked glycosylation, Asn³⁶, Asn⁷², Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³, and Asn³²¹. Five of these, Asn³⁶, Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³ and Asn³²¹ were found to be glycosylated while Asn⁷² was unglycosylated or substantially unglycosylated (substantially unglycosylated means in a mixture of LAL molecules, fewer Asn⁷² are glycosylated than any of Asn³⁶, Asn¹⁰¹, Asn¹⁶¹, Asn²⁷³ and Asn³²¹). Accordingly, one aspect of the invention is LAL (e.g., human LAL) which is unglycosylated and/or substantially unglycosylated at Asn⁷², and production and use of such LAL. However, LAL having a glycosylated Asn⁷² is within the scope of the invention. The N-glycan structures primarily consist of a mixture of bi-, tri- and tetra-antennary structures with N-acetylglucosamine, mannose and mannose-6-phosphate (M6P) as the major sugars. Each site appears to have a favored set of structures (Table 2 and FIG. 17) which is one aspect of the invention. For example, M6P-modified N-glycans reside at least at Asn¹⁰¹, Asn¹⁶¹ and Asn²⁷³. The non-phosphorylated structures are typical of N-glycans found on endogenous egg white proteins. No O-linked glycans were detected as determined by lack of N-acetylgalactosamine (GalNAc). No sialic acid was detected which is consistent with previously determined N-glycan structures of other endogenous and exogenous proteins produced in accordance with the invention. The invention includes LAL glycosylated with one or more of the oligosaccharide structures disclosed herein.

TABLE 2

Site residence of LAL glycan structures as determined by LC/MS of glycopeptides.	
Site	Glycan structure
Asn ³⁶	GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2
Asn ⁷²	None detected
Asn ¹⁰¹	Phos2Man7GlcNAc2
Asn ¹⁶¹	Phos1Man6GlcNAc2 GlcNAc1Phos1Man6GlcNAc2 Man3GlcNAc2 GlcNAc2Man3GlcNAc2 GlcNAc3Man3GlcNAc2 GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2
Asn ²⁷³	Man7GlcNAc2 Man8GlcNAc2 Man9GlcNAc2 Phos1Man8GlcNAc2 Phos1Man9GlcNAc2
Asn ³²¹	GlcNAc2Man3GlcNAc2 GlcNAc3Man3GlcNAc2 GlcNAc4Man3GlcNAc2 Hex1GlcNAc4Man3GlcNAc2 GlcNAc5Man3GlcNAc2 Hex1GlcNAc5Man3GlcNAc2 GlcNAc6Man3GlcNAc2 Hex1GlcNAc6Man3GlcNAc2

Hex, galactose; Phos, phosphate; Man, mannose; GlcNAc2, N-acetylglucosamine

Methods

[0241] Monosaccharide composition, including the neutrals, amino and M6P, was determined qualitatively and quantitatively by high pH anion exchange chromatography-pulsed amperometric detection (HPAEC-PAD).

[0242] The structures of the predominant glycans were determined with data from several mass spectrometry methods (MALDI-TOF, NSI-MS/MS and glycopeptide LC-MS). [0243] MALDI-TOF was useful for determination of neutral N-glycans and was able to detect phosphorylated N-glycans (FIG. 18). NSI-MS/MS was employed to determine the nature of minor peaks in the MALDI-TOF spectra, some of which were attributed to phosphorylated N-glycans (FIG. 19). Efforts to improve the ability of MALDI-TOF to detect phosphorylated N-glycans were not fruitful.

[0244] LC/MS of glycopeptides was able to detect neutral and phosphorylated structures and was able to determine the position of specific structures in the amino acid sequence of LAL (data summarized in FIG. 17 and Table 2).

[0245] To determine which peaks in the HPAEC-PAD chromatogram are due to phosphorylated N-glycans, LAL was treated with phosphatase and analyzed (FIG. 3). Peaks in groups C and D decreased in area under the curve (AUC) while a peak in group A became more prominent. Peaks in group B did not change in proportion to the other peaks. Based on the knowledge that retention time is proportional to the degree of charge (either due to phosphorylation or sialylation), it is contemplated that group C is composed of N-glycans with one phosphate (mono M6P) and group D composed of N-glycans with two phosphates (bis-M6P).

[0246] The retention time was also affected by composition and relative structural position of the neutral and amino monosaccharides. Such examples include the presence of galactose, the presence of a bisecting GlcNAc and the degree of GlcNAc substitution. Such factors contribute to the multiplicity of peaks in the HPAEC-PAD chromatogram.

Example 12

In Vitro Enzyme Activity Analysis of Transgenic Avian Derived hLAL in Egg White

[0247] Activity of Lysosomal Acid Lipase in egg white was determined using the fluorogenic substrate 4-methylumbelliferyl-oleate assay essentially as described in Yan et al. (2006), *American Journal of Pathology*, Vol. 169, No. 3, p 916-926, the disclosure of which is incorporated in its entirety herein by reference.

[0248] A stock solution of 4-methylumbelliferyl oleate (4-MUO) was prepared consisting of 2.5 mM 4-MUO in 4% Triton X-100. The assay was performed in a microtiter plate each well containing 62.5 μ l of 0.2 M Sodium Citrate (pH 5.5) in 0.01% Tween80, 12.5 μ l of egg white sample and 25 μ l of the 2.5 mM 4-MUO. Change in fluorescence was monitored for 30 minutes at 37° C. using a Bio-Tek Synergy HT fluorometric microplate reader (excitation 360 nm and emission 460 nm). Prior to assay, egg white containing the hLAL was diluted to an enzyme concentration that resulted in the reaction continuing linearly for at least 30 minutes. The reaction was stopped with 50 μ l of 0.75 M Tris-HCl, pH 8.0 and the endpoint fluorescence signal was measured in the same plate reader used above (excitation 360 nm and emission 460 nm).

[0249] Units of activity were determined using 4-methylumbelliferyl as a standard. One unit (U) is defined as the amount of enzyme which results in the formation of 1 umole of 4-methylumbelliferyl/min under the assay conditions described above. Non-hLAL containing egg white was used as a negative control.

[0250] Egg white samples which were positive for hLAL contained between 1 U and 100 U of activity per ml egg white.

Egg white from 21 G1 chickens was analyzed. Egg white from 10 of the chickens tested positive for hLAL activity.

Example 13

In Vitro Analysis of Transgenic Avian Derived LAL

[0251] The ability of LAL produced in the oviduct cells of transgenic avians (referred to herein as “SBC-102,” “avian derived LAL,” “LAL,” or “hLAL”) to bind to cells and be internalized to the lysosomal compartment, was examined in vitro using macrophage and fibroblast cells. When incubated with macrophage cells, fluorescently-labeled SBC-102 was found to localize to the lysosome. This effect could be attenuated by using a mannose polysaccharide competitor, implicating the N-acetylglucosamine/mannose (GlcNAc/mannose) receptor as a mechanism of recognition and uptake by these cells. SBC-102 increased the cell-associated LAL activity in both LAL-deficient human fibroblasts and normal murine fibroblasts after incubation in vitro, indicating that exposure to SBC-102 can result in substantial replacement of deficient enzymatic activity.

[0252] Mannose-6-phosphate (M6P) is present in the oligosaccharide structures of SBC-102 which have been shown to be involved in the delivery of lysosomal enzymes to a wide variety of cells types via the ubiquitous M6P receptor.

[0253] LAL was purified from the egg white of transgenic hens. Oregon Green NHS was obtained from Invitrogen™ (#0-10241). The rat alveolar macrophage line, NR8383, and the mouse fibroblast line, NIH-3T3, were obtained from ATCC. LAL-deficient Wolman’s fibroblasts were obtained from Coriell Institute for Medical Research and LysoTracker® Red was obtained from Invitrogen™.

[0254] Enzyme Labeling:

[0255] 4 mg of transgenic avian derived LAL in PBS was labeled with Oregon Green, according to the manufacturer’s recommendations and reaction was subsequently dialyzed against PBS then concentrated.

[0256] Macrophage Uptake:

[0257] Fluorescently-labeled transgenic avian derived LAL (5 µg/mL) and LysoTracker® Red were incubated with NR8383 cells for 2 hours. Cells were examined by co-focal fluorescence microscopy using a sequential scanning mode at 488 nm and then 514 nm.

[0258] Competitive Inhibition with Mannan:

[0259] Fluorescently-labeled SBC-102 (5 µg/mL) and mannan were incubated with NR8383 cells for 2 hours. Cells were trypsinized and LAL uptake measured by fluorescence-activated cell sorting using median fluorescence intensity as the endpoint.

[0260] The ability of transgenic avian derived LAL to be taken up and subsequently incorporated into the lysosomes of target cells was examined using the macrophage cell line, NR8383. Fluorescently-labeled transgenic avian derived LAL and the lysosomal marker, “LysoTracker® Red” (Invitrogen™), were incubated with cells for 2 hours. The colocalization of transgenic avian derived LAL and lysosomal marker in the lysosomes of these cells was subsequently examined by confocal fluorescence microscopy using a sequential scanning mode (FIG. 20). The LAL demonstrated localization to lysosomes, which is consistent with similar in vitro studies using rhLAL from a variety of sources.

[0261] The binding specificity of transgenic avian derived LAL to the GlcNAc/mannose receptor has been assessed by competitive binding assays using the macrophage cell line,

NR8383 (FIG. 21). Fluorescently-labeled (Oregon Green) transgenic avian derived LAL at 5 µg/mL and various concentrations of the mannose-containing oligosaccharide, mannan, were co-incubated with cells for 2 hours. The relative inhibition of transgenic avian derived LAL uptake by mannan, as compared with no mannan control, was quantified by fluorescence-activated cell sorting analysis using median fluorescence intensity as the endpoint. A mannose dose dependent inhibition in transgenic avian derived LAL binding/uptake was observed, which is consistent with transgenic avian derived LAL: GlcNAcR interaction.

[0262] In addition, mannose-6-phosphate mediated uptake in fibroblast cells was demonstrated by competition experiments with mannose-6-phosphate (results not shown).

[0263] The ability of transgenic avian derived LAL exposure to increase LAL activity in cells has been examined using both normal and LAL-deficient cells in vitro. Fibroblasts isolated from a Wolman’s patient and normal murine fibroblasts (NIH-3T3) were incubated in the presence of transgenic avian derived LAL at concentrations of either 0, 0.16 or 0.5 µg/mL for 5 hours. Cells were then washed to remove non-specific signal and cell lysates were assayed for LAL activity using 4-MUO substrate. Endogenous cell-associated LAL activity was lower in Wolman’s fibroblasts compared to NIH-3T3 and dose-dependent increases in activity were observed in both cell types after incubation with transgenic avian derived LAL (FIG. 22).

Example 14

In Vivo Analysis of Transgenic Avian Derived LAL

[0264] LAL-deficient Yoshida Rats (i.e., Homozygous) (see Kuriyama et al. (1990), Journal of Lipid Research, vol. 31, p 1605-1611; Nakagawa et al., (1995) Journal of Lipid Research, vol. 36, p 2212-2218; and Yoshida and Kuriyama (1990) Laboratory Animal Science, vol. 40, p 486-489) were treated with either SBC-102 (5 mg/kg, IV) or placebo, once/week for four weeks beginning at four weeks of age. For each administration the SBC-102 was injected into the rat tail vein in two equal doses (2.5 mg/kg) 30 minutes apart. Rats and aged-matched wild-type controls were examined one week after the final dose. Analyses were done in triplicate.

[0265] Gross pathologic examination of the SBC-102 treated animals demonstrated normalization in liver color in addition to reduction in organ size. The SBC-102 treated rats showed essentially normal liver histology in marked contrast to the substantial accumulation of foamy macrophages in the vehicle-treated animals (data not presented). Serum alanine and aspartate transferase levels, which are elevated in LAL^{-/-} rats, were also reduced in SBC-102 treated rats (not shown).

[0266] Mass of internal organs and tissue was determined for each rat and the data is shown in FIG. 23. Organ size is represented as percent of body weight determined at 8 weeks of age, in LAL^{-/-} rats and LAL^{+/+} rats after weekly administration of vehicle or SBC-102 at 5 mg/kg for 4 weeks.

[0267] Body weight of SBC-102- or vehicle-treated Yoshida rats were compared with wild type rats, as is shown in FIG. 24. SBC-102 (5 mg/kg) or vehicle was administered by IV injection either as a single dose or as split doses (given within 4 hour period) to LAL^{-/-} rats. LAL^{+/+} rats are age-matched littermate controls.

Example 15

Triglyceride Analysis

[0268] Triglyceride analysis was performed on liver and spleen tissue from wild type, homozygous placebo and homozygous SBC-102 treated animals. The triglyceride analyses were performed using standard methodologies (i.e., MBL International's Triglyceride Quantification Kit Catalog #JM-K622-100) and were done in triplicate.

TABLE 3

Liver and Spleen Triglyceride levels in wild-type and LAL deficient rats Triglyceride (ug/mg wet tissue)			
	Wild Type (n = 3)	Placebo (n = 3)	SBC-102 (n = 3)
Liver	48	84	57
Spleen	3	22	4

Liver Substrate Levels

[0269] FIG. 25 shows liver cholesterol, cholesteryl ester and triglyceride levels determined at 8 weeks of age, in WT and LAL deficient rats after weekly administration of vehicle or SBC-102 at 5 mg·kg⁻¹ for 4 weeks.

Example 16

Dose Response Study

[0270] Based on the studies performed above, the pharmacodynamic (PD) effects of a range of doses and dose schedules (qw and qow) of LAL ("SBC-102") were examined in LAL^{-/-} rats. In these studies, SBC-102 was administered by IV injections at dosages of 0.2, 1, 3 and 5 mg/kg, qow, or 0.35, 1 and 5 mg/kg, qw, for 1 month, beginning at 4 weeks of age. Results demonstrate improvements in body weight (BW) gain (FIG. 26), organomegaly (FIG. 27), and tissue substrate levels (FIG. 28). Serum transaminase levels were also reduced as the SBC-102 dose increased, with levels reaching essentially wild-type levels at the higher doses.

Example 17

Administration of Recombinant LAL in a Rat Model

[0271] The effects of repeat-dosing with recombinant human lysosomal acid lipase (LAL) on weight, tissue triglycerides and cholesterol, hepatomegaly, splenomegaly, lymphadenopathy, intestinal weight, and other parameters were evaluated in LAL Deficient Donryu rats described in Yoshida and Kuriyama (1990) Laboratory Animal Science, vol. 40, p 486-489 (see also Kuriyama et al. (1990) Journal of Lipid Research, vol. 31, p 1605-1611; Nakagawa et al., (1995) Journal of Lipid Research, vol. 36, p 2212-2218), the disclosure of which is incorporated in its entirety herein by reference.

[0272] At 4 weeks of age, Donryu rats homozygous for the LAL deletion (LAL^{-/-}) were assigned into groups to either be dosed with recombinant human LAL produced in a transgenic chicken oviduct system or a saline placebo. Wild-type, age-matched, littermate rats were used as controls. The LAL^{-/-} rats were dosed once a week for four weeks (four doses total) or once every two weeks for four weeks (two

doses total) by tail-vein injection as a single dose or in two equal doses given 30 minutes apart. Doses of LAL were 1 mg/kg or 5 mg/kg. Dosing schedule is shown in Table 4. The rats were pretreated with diphenhydramine (5 mg/kg) to counteract potential anaphylactic reactions, a procedure which is based on previous experiences in animal models of enzyme replacement therapy for the treatment of lysosomal storage disease (Shull et al. (1994) Proceedings of the National Academy of Science, vol. 91, p. 12937; Bielicki et al. (1999) The Journal of Biological Chemistry, 274, p. 36335; Vogler et al. (1999) Pediatric Research, 45, p. 838.), the disclosure of which is incorporated in its entirety herein by reference.

[0273] FIG. 29 shows the daily progress in weight gain of rats which were administered either 1 mg/kg of LAL per week or 5 mg/kg of LAL per week or 5 mg/kg of LAL per two weeks. It can be seen in the figure that there is little or no difference in therapeutic effect between the two dose sizes and frequencies.

TABLE 4

Weighing and Dosing Schedule	
Day from Birth	Assessments/Injections Performed
Day 13	WEIGHED
Day 14	
Day 20	
Day 21	
Day 24	Pups Weaned
Day 25	
Day 27	
Day 28	First Injection for administration once every week and once every two weeks
Day 31	
Day 32	
Day 34	
Day 35	Second Injection for administration once every week
Day 38	
Day 39	
Day 41	
Day 42	Third Injection for administration once every week; Second administration for once every two weeks
Day 45	
Day 48	
Day 49	
Day 55	Fourth injection for administration once every week
Day 56	
Day 56	Necropsy

Pathologic Examination of LAL^{-/-} Rats Treated with Recombinant LAL

[0274] At the termination of the study described in Example 1, study animals were humanely euthanized and necropsied to examine gross pathology, histopathology, and clinical chemistry. The gross necropsy included examination of the external surface of the body, all orifices, and the cranial, thoracic, and abdominal cavities and their contents. Mass of internal organs and tissues were determined for the rats and the organs and tissues were harvested and fixed in 10% neutral-buffered formalin. Following fixation, the tissues were processed and histological slides of hematoxylin and eosin-stained sections were prepared and evaluated.

[0275] The gross pathological examination of treated animals analyzed showed a substantial normalization in liver

size and color as can be seen in the dissection shown in FIG. 30. Organ-to-body weight ratios were determined and demonstrated a reduction in the relative organ size for liver, spleen, mesenteric tissue, duodenum, jejunum and ileum in successfully treated animals which were dissected, as compared to the placebo treated rats (FIG. 23). Histopathology of liver tissue from LAL of treated rats analyzed shows essentially normal liver histology in marked contrast to the substantial accumulation of foamy macrophages in the placebo-treated animals (FIG. 30).

Example 18

Treatment of Wolman Disease (WD) by Administration of Recombinant LAL

[0276] At 7 weeks of age a female patient is admitted to the hospital because of difficulty in weight gain and poor progress since birth. At the initial physical examination the patient weighs 3.6 kg (birth weight 3.7 kg) and is thin, with loose skin folds. The abdomen is distended, with firm hepatomegaly of 6 cm and firm splenomegaly of about 4 cm. Enlarged lymph nodes are noted in the groin and muscular activity is weak.

[0277] The initial hemoglobin level is 9.2 gm, platelets 506,000, and white blood cells 11,550. Urinalysis is normal, and bone marrow smears reveal vacuolated lymphocytes and numerous foam cells. Serum chemical measurements: total lipids 834 mg/100 ml, phospholipids 176 mg/100 ml, triglycerides 141 mg/100 ml, cholesterol 129 mg/100 ml, bilirubin 0.3 mg/100 ml, alkaline phosphatase 9.0 BU %, SGOT 90 units, SGPT 50 units, cholinesterase 20 units, urea nitrogen 8.3 mg, fasting sugar 45 mg/100 ml. CT scan of the abdomen shows hepatosplenomegaly and bilateral symmetrically enlarged adrenal glands with calcification.

[0278] The patient is surgically implanted with a venous vascular access port for dosing. After connecting the port to an ambulatory infusion machine, the patient is pretreated with 1 mg/kg of diphenhydramine 20 minutes prior to LAL infusion in order to counteract potential anaphylactic infusion reactions. The patient is then administered LAL at 1 mg/kg over the course of 5 hours by intravenous infusion. This therapy is repeated one time every 7 days indefinitely.

[0279] Within two weeks of administering the first dose of LAL, the patient experiences a significant improvement in weight gain and normalization in size of key abdominal organs as determined by ultrasound. Laboratory results demonstrate that infusion of the LAL restores lysosomal acid lipase activity in the patient and leads to correction of related abnormalities.

Example 19

Treatment of Cholesteryl Ester Storage Disease (CESD) by Administration of Recombinant LAL

[0280] A 3-year-old boy with a pruritic abdominal rash is examined by his pediatrician. Upon abdominal examination, hepatomegaly is noted by the physician and confirmed by ultrasound. At this point no diagnosis is made and the patient is monitored periodically.

[0281] At age 8, he is admitted to the hospital with gastroenteritis. Light microscopy of a liver biopsy shows increased intracytoplasmic glycogen and small lipid droplets in hepatocytes. Electron microscopy shows membrane-bound lipid droplets with small electron dense granules. A working diag-

nosis of glycogen storage disease type III (DeBrancher disease) is made, but skin fibroblast Debrancher activity is normal.

[0282] At age 10, hepatomegaly persists and a second liver biopsy is taken. Light microscopy shows altered lobular architecture of the hepatic parenchyma with distended hepatocytes containing cytoplasmic granules and vacuoles with mild periportal fibrosis. Fibroblast acid lipase activity is found to be 7% of normal, confirming the diagnosis of CESD. Plasma concentrations of total cholesterol (TC), triglycerides (TG), low-density lipoprotein cholesterol (LDL-C) are each above the 95th percentile for age and sex at 7.51, 3.24 and 5.58 mmol/L, respectively, while plasma high-density lipoprotein cholesterol (HDL-C) is below the 5th percentile at 0.47 mmol/L; he has combined hyperlipidemia (hypercholesterolemia, hypertriglyceridemia, hypoalphalipoproteinemia and hyperbetalipoproteinemia).

[0283] The patient is surgically implanted with a venous vascular access port for dosing. After connecting the port to an ambulatory infusion machine, the patient is pretreated with 5 mg/kg of diphenhydramine 20 minutes prior to LAL infusion in order to counteract potential anaphylactic infusion reactions. The patient is then administered LAL at 5 mg/kg over the course of 5 hours by intravenous infusion. This therapy is repeated one time every 14 days indefinitely.

[0284] Within two weeks of administering the first dose of LAL, the patient experiences a significant improvement in weight gain and normalization in size of key organs as determined by ultrasound. Laboratory results demonstrate that infusion of the LAL restores lysosomal acid lipase activity in the patient and leads to correction of related abnormalities.

Example 20

Description and Composition of the Medicinal Product

[0285] The drug substance of LAL described herein ("SBC-102") is a recombinant human lysosomal acid lipase (rhLAL) purified from the egg white produced from transgenic Gallus. The excipients used in SBC-102 are similar to those used for other products for lysosomal storage disorders (LSD) currently on the market, and have been selected to maintain stability of the drug product.

[0286] SBC-102 is a clear, colorless, sterile liquid provided in a clear, Type I borosilicate glass vial with a non-natural latex (butyl), FluroTec®-coated stopper and aluminum crimp seal. SBC-102 is provided as an aqueous solution comprised of SBC-102 (2 mg/mL), Trisodium Citrate Dihydrate (13.7 mg/mL, USP), Citric Acid Monohydrate (1.57 mg/mL, USP), Human Serum Albumin (10 mg/mL, USP), and Water for Injection (to final volume, USP). The pH of SBC-102 is 5.9±0.2. SBC-102 contains no preservatives and vials are intended for single use only.

TABLE 5

Excipients in SBC-102 (LAL)			
Excipient	CAS number	Grade	Function
Trisodium Citrate Dihydrate	6132-04-03	USP	Buffer
Citric Acid Monohydrate	5949-29-1	USP	Buffer
Human Serum Albumin	70024-90-7	USP	Stabilizer

Components of the Drug Product

[0287]

TABLE 6

Formulation of SBC-102	
Component	Concentration
SBC-102 (rhLAL)	2 mg/mL*
Trisodium Citrate Dihydrate	13.7 mg/mL
Citric Acid Monohydrate	1.57 mg/mL
Human Serum Albumin	10 mg/mL
Water for Injection, QS to	1.0 mL

[0288] Each example in the above specification is provided by way of explanation of the invention, not limitation of the invention. In fact, it will be apparent to those skilled in the art that various modifications, combinations, additions, deletions and variations can be made in the present invention

without departing from the scope or spirit of the invention. For instance, features illustrated or described as part of one embodiment can be used in another embodiment to yield a still further embodiment. It is intended that the present invention cover such modifications, combinations, additions, deletions, and variations.

[0289] All documents (e.g., U.S. patents, U.S. patent applications, publications) cited in the above specification are incorporated herein by reference. Various modifications and variations of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

[0290] While this invention has been particularly shown and described with references to example embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 20

<210> SEQ ID NO 1

<211> LENGTH: 399

<212> TYPE: PRT

<213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 1

```

Met Lys Met Arg Phe Leu Gly Leu Val Val Cys Leu Val Leu Trp Thr
1          5          10          15

Leu His Ser Glu Gly Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu
20          25          30

Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr Trp Gly Phe Pro Ser
35          40          45

Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr Ile Leu Cys Leu Asn
50          55          60

Arg Ile Pro His Gly Arg Lys Asn His Ser Asp Lys Gly Pro Lys Pro
65          70          75          80

Val Val Phe Leu Gln His Gly Leu Leu Ala Asp Ser Ser Asn Trp Val
85          90          95

Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly
100         105         110

Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn Thr Trp Ser Arg Lys
115         120         125

His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe Trp Ala Phe Ser Tyr
130         135         140

Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser Ile Asn Phe Ile Leu
145         150         155         160

Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val Gly His Ser Gln Gly
165         170         175

Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile Pro Glu Leu Ala Lys
180         185         190

Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val Ala Ser Val Ala Phe
195         200         205

Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu Pro Asp His Leu Ile

```

-continued

210	215	220
Lys Asp Leu Phe Gly	Asp Lys Glu Phe Leu	Pro Gln Ser Ala Phe Leu
225	230	235 240
Lys Trp Leu Gly Thr	His Val Cys Thr His Val	Ile Leu Lys Glu Leu
	245	250 255
Cys Gly Asn Leu Cys	Phe Leu Leu Cys Gly	Phe Asn Glu Arg Asn Leu
	260	265 270
Asn Met Ser Arg Val	Asp Val Tyr Thr Thr His	Ser Pro Ala Gly Thr
	275	280 285
Ser Val Gln Asn Met	Leu His Trp Ser Gln Ala	Val Lys Phe Gln Lys
	290	295 300
Phe Gln Ala Phe Asp	Trp Gly Ser Ser Ala Lys	Asn Tyr Phe His Tyr
305	310	315 320
Asn Gln Ser Tyr Pro	Pro Thr Tyr Asn Val Lys	Asp Met Leu Val Pro
	325	330 335
Thr Ala Val Trp Ser	Gly Gly His Asp Trp Leu	Ala Asp Val Tyr Asp
	340	345 350
Val Asn Ile Leu Leu	Thr Gln Ile Thr Asn Leu	Val Phe His Glu Ser
	355	360 365
Ile Pro Glu Trp Glu	His Leu Asp Phe Ile Trp	Gly Leu Asp Ala Pro
	370	375 380
Trp Arg Leu Tyr Asn	Lys Ile Ile Asn Leu Met	Arg Lys Tyr Gln
385	390	395

<210> SEQ ID NO 2
 <211> LENGTH: 378
 <212> TYPE: PRT
 <213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 2

Ser Gly Gly Lys	Leu Thr Ala Val Asp	Pro Glu Thr Asn Met	Asn Val
1	5	10	15
Ser Glu Ile Ile	Ser Tyr Trp Gly Phe	Pro Ser Glu Glu Tyr	Leu Val
	20	25	30
Glu Thr Glu Asp	Gly Tyr Ile Leu Cys	Leu Asn Arg Ile	Pro His Gly
	35	40	45
Arg Lys Asn His	Ser Asp Lys Gly Pro	Lys Pro Val Val Phe	Leu Gln
	50	55	60
His Gly Leu Leu	Ala Asp Ser Ser Asn	Trp Val Thr Asn Leu	Ala Asn
65	70	75	80
Ser Ser Leu Gly	Phe Ile Leu Ala Asp	Ala Gly Phe Asp	Val Trp Met
	85	90	95
Gly Asn Ser Arg	Gly Asn Thr Trp Ser	Arg Lys His Lys	Thr Leu Ser
	100	105	110
Val Ser Gln Asp	Glu Phe Trp Ala Phe	Ser Tyr Asp Glu	Met Ala Lys
	115	120	125
Tyr Asp Leu Pro	Ala Ser Ile Asn Phe	Ile Leu Asn Lys	Thr Gly Gln
	130	135	140
Glu Gln Val Tyr	Tyr Val Gly His Ser	Gln Gly Thr Thr	Ile Gly Phe
145	150	155	160
Ile Ala Phe Ser	Gln Ile Pro Glu Leu	Ala Lys Arg Ile	Lys Met Phe
	165	170	175
Phe Ala Leu Gly	Pro Val Ala Ser	Val Ala Phe Cys	Thr Ser Pro Met

-continued

180					185					190					
Ala	Lys	Leu	Gly	Arg	Leu	Pro	Asp	His	Leu	Ile	Lys	Asp	Leu	Phe	Gly
		195					200					205			
Asp	Lys	Glu	Phe	Leu	Pro	Gln	Ser	Ala	Phe	Leu	Lys	Trp	Leu	Gly	Thr
		210					215					220			
His	Val	Cys	Thr	His	Val	Ile	Leu	Lys	Glu	Leu	Cys	Gly	Asn	Leu	Cys
		225					230					235			240
Phe	Leu	Leu	Cys	Gly	Phe	Asn	Glu	Arg	Asn	Leu	Asn	Met	Ser	Arg	Val
			245					250						255	
Asp	Val	Tyr	Thr	Thr	His	Ser	Pro	Ala	Gly	Thr	Ser	Val	Gln	Asn	Met
			260					265					270		
Leu	His	Trp	Ser	Gln	Ala	Val	Lys	Phe	Gln	Lys	Phe	Gln	Ala	Phe	Asp
		275					280					285			
Trp	Gly	Ser	Ser	Ala	Lys	Asn	Tyr	Phe	His	Tyr	Asn	Gln	Ser	Tyr	Pro
		290					295					300			
Pro	Thr	Tyr	Asn	Val	Lys	Asp	Met	Leu	Val	Pro	Thr	Ala	Val	Trp	Ser
		305					310					315			320
Gly	Gly	His	Asp	Trp	Leu	Ala	Asp	Val	Tyr	Asp	Val	Asn	Ile	Leu	Leu
			325					330						335	
Thr	Gln	Ile	Thr	Asn	Leu	Val	Phe	His	Glu	Ser	Ile	Pro	Glu	Trp	Glu
			340					345					350		
His	Leu	Asp	Phe	Ile	Trp	Gly	Leu	Asp	Ala	Pro	Trp	Arg	Leu	Tyr	Asn
		355					360					365			
Lys	Ile	Ile	Asn	Leu	Met	Arg	Lys	Tyr	Gln						
		370					375								

<210> SEQ ID NO 3

<211> LENGTH: 376

<212> TYPE: PRT

<213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 3

Gly	Lys	Leu	Thr	Ala	Val	Asp	Pro	Glu	Thr	Asn	Met	Asn	Val	Ser	Glu
1				5					10				15		
Ile	Ile	Ser	Tyr	Trp	Gly	Phe	Pro	Ser	Glu	Glu	Tyr	Leu	Val	Glu	Thr
		20					25					30			
Glu	Asp	Gly	Tyr	Ile	Leu	Cys	Leu	Asn	Arg	Ile	Pro	His	Gly	Arg	Lys
		35				40					45				
Asn	His	Ser	Asp	Lys	Gly	Pro	Lys	Pro	Val	Val	Phe	Leu	Gln	His	Gly
		50			55						60				
Leu	Leu	Ala	Asp	Ser	Ser	Asn	Trp	Val	Thr	Asn	Leu	Ala	Asn	Ser	Ser
		65			70					75				80	
Leu	Gly	Phe	Ile	Leu	Ala	Asp	Ala	Gly	Phe	Asp	Val	Trp	Met	Gly	Asn
			85					90					95		
Ser	Arg	Gly	Asn	Thr	Trp	Ser	Arg	Lys	His	Lys	Thr	Leu	Ser	Val	Ser
		100					105					110			
Gln	Asp	Glu	Phe	Trp	Ala	Phe	Ser	Tyr	Asp	Glu	Met	Ala	Lys	Tyr	Asp
		115				120						125			
Leu	Pro	Ala	Ser	Ile	Asn	Phe	Ile	Leu	Asn	Lys	Thr	Gly	Gln	Glu	Gln
		130				135					140				
Val	Tyr	Tyr	Val	Gly	His	Ser	Gln	Gly	Thr	Thr	Ile	Gly	Phe	Ile	Ala
		145			150				155					160	
Phe	Ser	Gln	Ile	Pro	Glu	Leu	Ala	Lys	Arg	Ile	Lys	Met	Phe	Phe	Ala

-continued

165					170					175					
Leu	Gly	Pro	Val	Ala	Ser	Val	Ala	Phe	Cys	Thr	Ser	Pro	Met	Ala	Lys
			180						185				190		
Leu	Gly	Arg	Leu	Pro	Asp	His	Leu	Ile	Lys	Asp	Leu	Phe	Gly	Asp	Lys
		195					200					205			
Glu	Phe	Leu	Pro	Gln	Ser	Ala	Phe	Leu	Lys	Trp	Leu	Gly	Thr	His	Val
	210					215					220				
Cys	Thr	His	Val	Ile	Leu	Lys	Glu	Leu	Cys	Gly	Asn	Leu	Cys	Phe	Leu
225						230					235				240
Leu	Cys	Gly	Phe	Asn	Glu	Arg	Asn	Leu	Asn	Met	Ser	Arg	Val	Asp	Val
			245						250					255	
Tyr	Thr	Thr	His	Ser	Pro	Ala	Gly	Thr	Ser	Val	Gln	Asn	Met	Leu	His
			260					265					270		
Trp	Ser	Gln	Ala	Val	Lys	Phe	Gln	Lys	Phe	Gln	Ala	Phe	Asp	Trp	Gly
		275					280					285			
Ser	Ser	Ala	Lys	Asn	Tyr	Phe	His	Tyr	Asn	Gln	Ser	Tyr	Pro	Pro	Thr
	290					295					300				
Tyr	Asn	Val	Lys	Asp	Met	Leu	Val	Pro	Thr	Ala	Val	Trp	Ser	Gly	Gly
305						310					315				320
His	Asp	Trp	Leu	Ala	Asp	Val	Tyr	Asp	Val	Asn	Ile	Leu	Leu	Thr	Gln
			325					330						335	
Ile	Thr	Asn	Leu	Val	Phe	His	Glu	Ser	Ile	Pro	Glu	Trp	Glu	His	Leu
		340						345					350		
Asp	Phe	Ile	Trp	Gly	Leu	Asp	Ala	Pro	Trp	Arg	Leu	Tyr	Asn	Lys	Ile
	355					360					365				
Ile	Asn	Leu	Met	Arg	Lys	Tyr	Gln								
	370					375									

<210> SEQ ID NO 4
 <211> LENGTH: 373
 <212> TYPE: PRT
 <213> ORGANISM: HOMO SAPIENS

 <400> SEQUENCE: 4

Thr	Ala	Val	Asp	Pro	Glu	Thr	Asn	Met	Asn	Val	Ser	Glu	Ile	Ile	Ser
1				5					10					15	
Tyr	Trp	Gly	Phe	Pro	Ser	Glu	Glu	Tyr	Leu	Val	Glu	Thr	Glu	Asp	Gly
		20						25					30		
Tyr	Ile	Leu	Cys	Leu	Asn	Arg	Ile	Pro	His	Gly	Arg	Lys	Asn	His	Ser
	35						40					45			
Asp	Lys	Gly	Pro	Lys	Pro	Val	Val	Phe	Leu	Gln	His	Gly	Leu	Leu	Ala
	50					55					60				
Asp	Ser	Ser	Asn	Trp	Val	Thr	Asn	Leu	Ala	Asn	Ser	Ser	Leu	Gly	Phe
65					70					75				80	
Ile	Leu	Ala	Asp	Ala	Gly	Phe	Asp	Val	Trp	Met	Gly	Asn	Ser	Arg	Gly
		85							90					95	
Asn	Thr	Trp	Ser	Arg	Lys	His	Lys	Thr	Leu	Ser	Val	Ser	Gln	Asp	Glu
		100							105				110		
Phe	Trp	Ala	Phe	Ser	Tyr	Asp	Glu	Met	Ala	Lys	Tyr	Asp	Leu	Pro	Ala
		115					120					125			
Ser	Ile	Asn	Phe	Ile	Leu	Asn	Lys	Thr	Gly	Gln	Glu	Gln	Val	Tyr	Tyr
	130					135					140				
Val	Gly	His	Ser	Gln	Gly	Thr	Thr	Ile	Gly	Phe	Ile	Ala	Phe	Ser	Gln

-continued

145	150	155	160
Ile Pro Glu Leu Ala Lys Arg Ile Lys Met Phe Phe Ala Leu Gly Pro	165	170	175
Val Ala Ser Val Ala Phe Cys Thr Ser Pro Met Ala Lys Leu Gly Arg	180	185	190
Leu Pro Asp His Leu Ile Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu	195	200	205
Pro Gln Ser Ala Phe Leu Lys Trp Leu Gly Thr His Val Cys Thr His	210	215	220
Val Ile Leu Lys Glu Leu Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly	225	230	235
Phe Asn Glu Arg Asn Leu Asn Met Ser Arg Val Asp Val Tyr Thr Thr	245	250	255
His Ser Pro Ala Gly Thr Ser Val Gln Asn Met Leu His Trp Ser Gln	260	265	270
Ala Val Lys Phe Gln Lys Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala	275	280	285
Lys Asn Tyr Phe His Tyr Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val	290	295	300
Lys Asp Met Leu Val Pro Thr Ala Val Trp Ser Gly Gly His Asp Trp	305	310	315
Leu Ala Asp Val Tyr Asp Val Asn Ile Leu Leu Thr Gln Ile Thr Asn	325	330	335
Leu Val Phe His Glu Ser Ile Pro Glu Trp Glu His Leu Asp Phe Ile	340	345	350
Trp Gly Leu Asp Ala Pro Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu	355	360	365
Met Arg Lys Tyr Gln	370		

<210> SEQ ID NO 5

<211> LENGTH: 1150

<212> TYPE: DNA

<213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 5

```

atgaaaaatgc ggttcttggg gttggtggtc tgtttggttc tctggaccct gcattccgag      60
gggtccggag ggaaactgac agctgtggat cctgaaacaa acatgaatgt cagtgaatt      120
atctcttact ggggattccc tagtgaggaa tacctagttg agacagaaga tggatatatt      180
ctgtgcctta accgaattcc tcatgggagg aagaaccatt ctgacaaagg tcccaaacca      240
gttgtcttcc tgcaacatgg cttgctggca gattctagta actgggtcac aaaccttgcc      300
aacagcagcc tgggcttcat tcttgctgat gctggttttg acgtgtggat gggcaacagc      360
agaggaaata cctggtctcg gaaacataag aactctcag tttctcagga tgaattctgg      420
gcttttcagtt atgatgagat ggcaaaatat gacctaccag cttccattaa cttcattctg      480
aataagactg gccaaagaaca agtgtattat gtgggtcatt ctcaaggcac cactataggt      540
tttatagcat tttcacagat ccctgagctg gctaaaagga ttaaaatggt ttttgccctg      600
ggctcctgtg cttccgtgcg cttctgtact agccctatgg ccaaactggg acgactgcca      660
gatcatctca ttaaggacct ctttggagac aaagaatttc tccccagag tgcgtttttg      720
aagtggctgg gtaccacgt ttgcactcat gtcatactga aggagctctg tggaaatctc      780

```

-continued

tggtttcttc	tggtggatt	taatgagaga	aatttaaata	tgtctagagt	ggatgtgtat	840
acaacacatt	ctcctgctgg	aacttctgtg	caaaacatgt	tacactggag	ccaggctgtt	900
aaattccaaa	agtttcaagc	ctttgactgg	ggaagcagtg	ccaagaatta	ttttcattac	960
aaccagagtt	atcctccac	atacaatgtg	aaggacatgc	ttgtgccgac	tgcagtctgg	1020
agcgggggtc	acgactggct	tgcagatgto	tacgacgtca	atatottact	gactcagatc	1080
accaacttgg	tgttccatga	gagcattccg	gaatggggagc	atcttgactt	catttggggc	1140
ctggatgcc						1150

<210> SEQ ID NO 6
 <211> LENGTH: 10882
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: pALVIN-OVR1-I-SBC102-dSA

<400> SEQUENCE: 6

ctttcccggt	caagctctaa	atcgggggct	ccctttagggt	ttccgattta	gtgctttacg	60
gcacctcgac	cccaaaaaac	ttgattagggt	tgatgggtca	cgtagtgggc	catcgccctg	120
atagacgggt	tttcgccctt	tgacgttggg	gtccacgttc	tttaaatagtg	gactcttggt	180
ccaaactgga	acaacactca	accctatctc	ggctctattct	tttgatttat	aagggatttt	240
gccgatttcg	gcctattggg	taaaaaatga	gctgatttaa	caaaaattta	acgcgaattt	300
taacaaaata	ttaacgctta	caatttccat	tcgccattca	ggctgcgcaa	ctgttgggaa	360
ggcgatcggt	tgccggcctc	ttcgctatta	cgccagctgg	cgaagggggg	atgtgctgca	420
aggcgattaa	gttgggtaac	gccagggttt	tcccagtcac	gacgttgtaa	aacgacggcc	480
agtgaagcgc	tattccctaa	cgatcacgtc	ggggtcacca	aatgaagcct	tctgcttcat	540
gcatgtgctc	gtagtcgtca	gggaatcaac	gggtccggcca	tcaaccagg	tgcacaccaa	600
tggtgtgaat	gggtcaaatg	cggtttattgt	atcgagctag	gcacttaaat	acaatatctc	660
tgcaatgcgg	aattcagtg	ttcgctcaat	ccgtccccct	ccctatgcaa	aagcgaaact	720
actatctcct	gaggggactc	ctaaccgcgt	acaaccgaag	ccccgctttt	cgccataaca	780
tgctattgtc	ccctcagtc	agccttgccc	gttacaaccc	gattcgcaag	ccttgccctc	840
cccacattat	ccgtagcatt	atttcctagc	agtcacgaga	gctacagaag	atactctatg	900
ctgtagccaa	gtctacaagt	ttactattca	gcgacctcct	atattccgcg	tgccagccga	960
tcaattacca	atccaaccag	ctatcacacg	gaatacaaga	actcgccctac	gctcttcttt	1020
cgggctgctt	ataagcctcc	tgtaattttt	ttatattcct	cgtcgcagtc	tcttcagaat	1080
ggcacagcac	cgtcgagcaa	aaatgccagg	tggactatga	actcacatcc	aaaggagctt	1140
gacctgatac	ctgattttct	tcaaacaggg	gaaacaacac	aatcccacaa	aacagctcag	1200
agagaaacca	tactgatggg	ctacagcacc	aaggatatgca	atggcaatcc	attcgacatt	1260
catctgtgac	ctgagcaaaa	tgatttatct	ctccatgaat	gggtgcttct	ttccctcatg	1320
aaaaggcaat	ttccacactc	acaatatgca	acaaagacaa	acagagaaca	attaatgtgc	1380
tccttcctaa	tgttaaaatt	gtagtggcaa	agaggagaaac	aaaatctcaa	gttctgagta	1440
ggtttttagtg	attggataag	aggctttgac	ctgtgagctc	acctggactt	catatccctt	1500
tggaataaaa	gtgcttttat	aactttcagg	tctccgagtc	tttattcatg	agactgttgg	1560

-continued

tttagggaca gaccacaat gaaatgcctg gcataggaaa gggcagcaga gccttagctg	1620
accttttctt gggacaagca ttgtcaaaca atgtgtgaca aaactatttg tactgctttg	1680
cacagctgtg ctgggcaggg caatccattg ccacctatcc caggtaacct tccaactgca	1740
agaagattgt tgcttactct ctctagacc ccaagtcaaa ccaactatgc aggtatgctg	1800
acaacactat gatgacagcc tgttctgac aagatctcat ttgttcattg acaatttttg	1860
ttgtctgcag ctggctcttc attgggaaag agtgtagtat atccttctca tctgacagaa	1920
aagcagaaat tctcatgctc cacacttaat ctacattggt ttaaaccacc ggctacttct	1980
tgagagaggaa aatggctttt tataagactc aaaaaacaaa gctctgcaag tcaaatgcat	2040
aaaaaactgt tctgtaggct tggaaatcagg aactatgtg gaagtcaaat agagcagctt	2100
taaaaagcct ttgggatcat tctcatctta ttttgcagc acgatactat gacagtgata	2160
actgacataa ctgcacaaat ttccttgata ttttatttgt cttaaagtac aagacataga	2220
gatggacgta aagatggaca tatgactcag gtctggacag gtccgtgggc catgtatgat	2280
aaaagagatg aagggaagga gaattgagac tgtctaagaa gggcttcagg gacgttctga	2340
aggcagattt gactgaatca gatgtactgt ccaagtctca tatgtagcaa tggaaggctg	2400
atattggaga aatataaaga aatggctgtg aactcaaagt gacctgaac agaaaaggga	2460
tatggagtta aaataatgtc acagaactga ggtttatatg atataccatg ggctgcagag	2520
ggtcagagtg ctccaccatg ggcctctctt gggtctgagg gaacttctgt tctacacctg	2580
gaacacctcc tgcctctctc cgcactgacc tcagtgtcat cagggtgtgt tctctcacat	2640
tttctcactc acctctccca actaccattg tacagcagtt gttcttcat attgctctc	2700
ctgaggtaca tctagcatcg ttaagtctc agacttgga aggagaatgt agatttccac	2760
agtatatatg ttttcacaaa aggaaggaga gaaacaaaag aaaatggcac tgactaaact	2820
tcagctagt gatatggaaa gtaattctgc ttaacagaga ttgcagtgat ctctatgtat	2880
gtcctgaaga attatgttgt acttttttcc cccattttta aatcaaacag tgctttacag	2940
aggtcagaat ggtttcttta ctgtttgtca attctattat ttcaatacag aacaatagct	3000
tctataactg aaatatattt gctattgtat attatgattg tccctcgaac catgaacact	3060
cctccagctg aatttcacaa ttcctctgtc atctgccagg ccattaagtt attcatggaa	3120
gatctttgag gaacactgca agttcatatc ataaacacat ttgaaattga gtattgtttt	3180
gcattgtatg gagctatgtt ttgctgtatc ctcaagaataa aagtttgta taaagcattc	3240
acaccataa aaagatagat ttaaatattc caactatagg aaagaaagtg tgtctgctct	3300
tcactctagt ctcaagtggc tccttcacat gcaagcttct tttttctcc tttttgtca	3360
agaaaataat aggtcaagtc ttgttctcat ttatgtctg tctagcgtgg ctcaagtga	3420
cattgtacat acaagaagga tcaaatgaaa cagacttctg gtctgttact acaaccatag	3480
taataagcac actaactaat aattgctaat tatgttttcc atctccaagg tccccacatt	3540
tttctgtttt cttaaagatc ccattatctg gttgtaactg aagctcaatg gaacatgagc	3600
aatatttccc agtcttctct cccatccaac agtcctgatg gattagcaga acaggcagaa	3660
aacacattgt taccagaat taaaaactaa ttttgctct ccattcaatc caaaatggac	3720
ctattgaaac taaaatctaa cccaatccca ttaaatgatt tctatggtgt caaagggtcaa	3780
acttctgaag ggaacctgtg ggtgggtcac aattcagact atatattccc cagggtctag	3840
ccagtgtctg tacatacagc tagaaagctg tattgccttt agcagtcaag ctcgaaaggt	3900

-continued

aagcaactct ctggaattac cttctctcta tattagctct tacttgacc taaacttta	3960
aaaattaaca attattgtgc tatgtgtgt atctttaagg gtgaagtacc tgcgtgatac	4020
cccctataaa aacttctcac ctggtgatgc attctgcact attttattat gtgtaaaagc	4080
tttgtgtttg ttttcaggag gcttattctt tgtgcttaaa atatgttttt aatttcagaa	4140
catcttatcc tgtcgttcac tatctgatat gctttgcagt ttgcttgatt aacttctagc	4200
cctacagagt gcacagagag caaaatcatg gtgttcagtg aattctgggg agttatttta	4260
atgtgaaat tctctagaag ttaattctct gcaaagtga gctgctgatc actacacaag	4320
ataaaaatgt ggggggtgca taaacgtata ttcttacaat aatagatata tgtgaactta	4380
tatacagaaa agaaaatgag aaaaatgtgt gtgtgtatac tcacacacgt ggtcagtaaa	4440
aacttttgag gggtttaata cagaaaatcc aatcctgagg cccagcact cagtacgc	4500
ataaagggct gggctctgaa ggacttctga ctttcacaga ttatataaat ctcaggaaag	4560
caactagatt catgctggct ccaaaagctg tgctttatat aagcactctg gctatacaat	4620
agttgtacag ttcagctctt tataatagaa acagacagaa caagtataaa tcttctattg	4680
gtctatgtca tgaacaagaa ttcattcagt ggctctgttt tatagtaaac attgctat	4740
tatcatgtct gcatttctct tctgtctgaa tgtcaccact aaaatttaac tcacagaaa	4800
gtttatacta cagtacacat gcatacttt gagcaaagca aaccatacct gaaagtga	4860
tagagcagaa tatgaattac atgcgtgtct ttctcctaga ctacatgacc ccatataaat	4920
tacattcctt atctattctg ccatcaccaa acaaaaggta aaaatacttt tgaagatcta	4980
ctcatagcaa gtagtggtga acaaacagat atttctctac atttattttt aggaataaa	5040
aataagaaat aaaatagtca gcaagcctct gctttctcat atatctgtcc aaacctaaag	5100
tttactgaaa tttgctcttt gaatttccag ttttgcaagc ctatcagatt gtgtttta	5160
cagaggtagt gaaaagtatc aatgaattct agctttcact gaacaaaaat atgtagaggc	5220
aactggcttc tgggacagtt tgctacccaa aagacaactg aatgcaaata cataaataga	5280
tttatgaata tggttttgaa catgcacatg agagggtgat atagcaacag acacattacc	5340
acagaattac tttaaaacta cttgttaaca ttaattgcc taaaaactgc tcgtaattta	5400
ctgttgtagc ctaccataga gtacctgca tggtagctatg tacagcatc catccttaca	5460
ttttcactgt tctgctgttt gctctagaca actcagagtt caccatgaaa atgcgggtct	5520
tgggggttgt ggtctgtttg gttctctgga cctgcatc cgaggggtcc ggagggaaac	5580
tgacagctgt ggatccagaa acaaacatga atgtcagtga aattatctct tactggggat	5640
tccttagtga ggaataccta gttgagacag aagatggata tattctgtgc cttaaccgaa	5700
ttcctcatgg gaggaagaac cattctgaca aaggtcccaa accagttgtc ttcctgcaac	5760
atggctgtct ggcagattct agtaactggg tcacaaacct tgccaacagc agcctgggct	5820
tcattcttgc tgatgctggt tttgacgtgt ggatgggcaa cagcagagga aatacctggt	5880
ctcgaaaca taagacactc tcagtttctc aggatgaatt ctgggcttcc agttatgatg	5940
agatggcaaa atatgacctc ccagcttcca ttaacttcat tctgaataag actggccaag	6000
aacaagtgtt ttatgtgggt cattctcaag gcaccactat aggttttata gcattttcac	6060
agatccctga gctggctaaa aggattaaaa tgttttttgc cctgggtcct gtggcttccg	6120
tcgccttctg tactagccct atggccaaac tgggacgact gccagatcat ctcatgaag	6180

-continued

acctcttttg	agacaaagaa	tttcttcccc	agagtgcgtt	tttgaagtgg	ctgggtaccc	6240
acgtttgcac	tcatgtcata	ctgaaggagc	tctgtggaaa	tctctgtttt	cttctgtgtg	6300
gatttaatga	gagaaattta	aatatgtcta	gagtggatgt	gtatacaaca	cattctcctg	6360
ctggaacttc	tgtgcaaac	atgttacact	ggagccaggc	tgttaaattc	caaaagtttc	6420
aagcctttga	ctggggaagc	agtgccaa	attattttca	ttacaaccag	agttatcctc	6480
ccacatacaa	tgtgaaggac	atgcttgtgc	cgactgcagt	ctggagcggg	ggtcacgact	6540
ggcttgacga	tgtctacgac	gtcaatatct	tactgactca	gatcaccaac	ttgggtgttc	6600
atgagagcat	tccggaatgg	gagcatcttg	acttcatttg	gggcctggat	gccccttgga	6660
ggctttataa	taagattatt	aatctaata	ggaaatatca	gtgattcgaa	gcggccgcaa	6720
gaagaaagct	gaaaaactct	gtcccttcca	acaagaccca	gagcactgta	gtatcagggg	6780
taaaatgaaa	agtatgttat	ctgctgcata	cagacttcat	aaaagctgga	gcttaattca	6840
gaaaaaaaaa	cagaaagaaa	ttacactgtg	agaacagggt	caattcactt	ttcctttaca	6900
cagagtaata	ctggtaactc	atggatgaag	gcttaaggga	atgaaattgg	actcacagta	6960
ctgagtcata	acactgaaaa	atgcaacctg	atacatcagc	agaaggttta	tgggggaaaa	7020
atgcagcctt	ccaattaagc	cagatatctg	tatgaccaag	ctgctccaga	attagtcact	7080
caaaatctct	cagattaaat	tatcaactgt	caccaacccat	tcctatgctg	acaaggcaat	7140
tgttggttct	ctgtgttctc	gatactacaa	ggctcttctc	gacttcctaa	agatgcatta	7200
taaaatctct	ataattcaca	tttctcccta	aactttgact	caatcatggg	atgttgga	7260
atatggtata	ttactattca	aattgttttc	ctgtaccca	tatgtaattg	gtcttgta	7320
tgtgtctctt	tgttccttta	atcataataa	aaacatgttt	aagcaaacac	ttttcacttg	7380
tagtatttga	aggtaccgga	tctcgagcgg	ccttcaatgc	ccccaaaacc	aatcccagg	7440
tttttaactc	tcccgattht	ccaagtacca	tagcccgctg	agagagcgcc	gcggtaattg	7500
gatcccagga	ccccggggaa	tataagtctg	agggggacgt	aagcaaccct	tccttttgta	7560
acagggacaa	catagccctc	atttccttct	tagaaggaga	ggttttcccg	caataggctc	7620
tacacgcgga	cgaaatcacc	tttatgacgg	cttccatgct	tgatccaccg	ggcgaccgga	7680
atcacgcaga	gcaaccggaa	tcacgcctgg	gggtggaccgc	tcagtgcctg	ggcttccttc	7740
ccgtcttcca	acgactctct	gagttctcgg	tagggatgtg	tggccccctg	cagtagggct	7800
ccctccgacg	ccactcagct	tctgccctcc	taagcccgag	ccccctctac	tagggtcata	7860
gtccgctccc	cgaataagcg	agacggatga	ggacaggatc	gccacgcgcg	ctgtggccga	7920
ccactattcc	ctaacgatca	cgtcgggggc	accaaataaa	gccttctgct	tcatgcatgt	7980
gctcgtatgc	gtcagggaat	caacgggtccg	gccatcaacc	caggtgcaca	ccaatgtggg	8040
gaatggtcaa	atggcggtta	ttgtatcgag	ctaggcactt	aaatacaata	tctctgcaat	8100
gcggaattca	gtgggttcgc	caatccgtgt	tagaccctgc	tgttgccctc	ctaacaaggc	8160
acgatcatac	cacgatcata	ccaccttact	cccaccaatc	ggcatgcacg	gtgctttttc	8220
tctccttata	aggcatgttg	ctaactcctc	gttacataag	catgttgcaa	gactacaaga	8280
gtattgcata	agactacatt	tccccctccc	tatgcaaaag	cgaactact	atatcctgag	8340
gggactccta	accgcgtaca	accgaagccc	cgcttttcgc	ctaaacatgc	tattgtcccc	8400
tcagtcaagc	cttgcccgtt	acaaccgat	tcgcaagcct	tgcctccccc	acattatccg	8460
tagcattatt	tcctagcagt	catcagagct	acagaagata	ctctatgctg	tagccaagtc	8520

-continued

tacaagtta ctattcagcg acctcctata ttccgcgtgc cagccgatca attaccaatg	8580
cgcgcttggc gtaatcatgg tcatagctgt ttctgtgtg aaattgttat ccgctcacia	8640
ttccacacia catagcagcc ggaagcataa agtgtaaagc ctgggggtgc taatgagtga	8700
gctaactcac attaatgtcg ttgcgctcac tgcccgttt ccagtcggga aacctgtcgt	8760
gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgct attgggcgct	8820
cttcgcttc ctgcgctact gactcgctgc gctcggtcgt tcggctgcgg cgagcggtat	8880
cagctcactc aaaggcggtat atacggttat ccacagaatc aggggataac gcaggaaaga	8940
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt	9000
ttttccatag gctccgccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt	9060
ggcgaaaccc gacaggacta taaagatacc aggcgtttcc cctggaagc tccctcgtgc	9120
gctctcctgt tccgaccctg ccgcttaccg gatacctgct cgctttctc ccttcgggaa	9180
gcgtggcgt ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct	9240
ccaagctggg ctgtgtgcac gaacccccg ttcagcccga ccgctgcgc ttatccggtat	9300
actatcgtct tgagtcacac ccggttaagc acgacttct gccactggca gcagccactg	9360
gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtgggtggc	9420
ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta	9480
ccttcggaag aagagttggt agctcttgat ccggcaaaaa aaccaccgct ggtagcgggtg	9540
gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt	9600
tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa gggatttttg	9660
tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta	9720
aatcaatcta aagtatatat gagtaaaactt ggtctgacag ttaccaatgc ttaatcagt	9780
aggcacctat ctacagcagc tgtctatttc gttcatccat agttgcctga cccccgctg	9840
tgtagataac tacgatacgg gagggcttac catctggccc cagtgcgtga atgataccgc	9900
gagaccacg ctacacggct ccagatttat cagcaataaa ccagccagcc ggaaggcccg	9960
agcgagaaag tggctctgca actttatccg cctccatcca gtctattaat tgttgccggg	10020
aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttggtgct attgctacag	10080
gcacgtggt gtcacgctcg tcggttggtat tggttcatt cagctccggt tcccaacgat	10140
caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggtagctcc ttcggtctc	10200
cgatcggtgt cagaagtaag ttggccgagc tggtatcact catgggtatg gcagcactgc	10260
ataattctct tactgtcatg ccacccgtaa gatgcttttc tgtgactggt gactactcaa	10320
ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgccc gcgtcaatac	10380
gggataatac cgccccacat agcagaactt taaaagtgt catcattgga aaacgttctt	10440
cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg taaccactc	10500
gtgcacccaa ctgatcttca gcacttttta ctttaccag cgtttctggg tgagcaaaaa	10560
caggaaagca aaatgccga aaaaaggaa taaggcgac acggaatgt tgaatactca	10620
tactcttct ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggtat	10680
acataattga atgtatttag aaaaataaac aaatagggt tccgcgcaca tttcccgaa	10740
aagtgcacc tgacgcgccc tgtagcggcg cattaagcgc ggcgggtgtg gtggttacgc	10800

-continued

```
gcagcgtgac cgctacactt gccagcgccc tagcgcccgc tcctttcgct ttcttccctt 10860
cctttctcgc caggttcgcc gg 10882
```

<210> SEQ ID NO 7

<211> LENGTH: 7780

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Proviral sequence of pALVIN-OVR1-I-SBC102-dSA

<400> SEQUENCE: 7

```
aatgaagcct tctgcttcat gcatgtgctc gtagtcgtca gggaatcaac ggtccggcca 60
tcaaccaggc tgcacaccaa tgtgtggaat ggtcaaatgg cgtttattgt atcgagctag 120
gcacttaaat acaatatctc tgcaatgcgg aattcagtgg ttcgtccaat ccgtccccct 180
ccctatgcaa aagcgaaact actatatcct gaggggactc ctaaccgcgt acaaccgaag 240
ccccgctttt cgctaaaca tgctattgtc cctcagtcga agccttgccc gttacaaccc 300
gattcgcaag ccttgccttc ccacattat ccgtagcatt atttcctagc agtcacaga 360
gctacagaag atactctatg ctgtagccaa gtctacaagt ttactattca ggcacctcct 420
atattccgcg tgcacgccga tcaattacca atccaaccag ctatcacacg gaatacaaga 480
actcgccctac gctcttcttt cgggctgctt ataagcctcc tgtaattttt ttatattcct 540
cgctcgagtc tcttcagaat ggcacagcac cgctgcagaa aaatgccagg tggactatga 600
actcacatcc aaaggagctt gacctgatac ctgattttct tcaaacaggg gaaacaacac 660
aatcccacaa aacagctcag agagaaacca tcaactgatg ctacagcacc aagggtatgca 720
atggcaatcc attcgacatt catctgtgac ctgagcaaaa tgatttatct ctccatgaat 780
ggttgcttct ttcctcatg aaaaggcaat ttccacactc acaatatgca acaagacaa 840
acagagaaca attaatgtgc tccttcctaa tgtaaaaatt gtagtggaac agaggagaac 900
aaaaatctca gttctgagta ggttttagtg attggataag aggctttgac ctgtgagctc 960
acctggactt catatccttt tggataaaaa gtgcttttat aactttcagg tctccgagtc 1020
tttattcatg agactgttgg tttagggaca gaccacaaat gaaatgcctg gcataggaaa 1080
gggcagcaga gccttagctg accttttctt gggacaagca ttgtcaaca atgtgtgaca 1140
aaactatttg tactgcttgg cacagctgtg ctgggcaggg caatccattg ccacctatcc 1200
caggtaacct tccaactgca agaagattgt tgcttactct ctctagaccc ccaagtcaaa 1260
ccaactatgc aggtatgctg acaacactat gatgacagcc tgttctgac aagatctcat 1320
ttgttcattg acaatttttg ttgcttgacg ctggtcttcc attgggaaag agtgtagtat 1380
atccttctca tctgacagaa aagcagaaat tctcatgctc cacacttaat ctacattgtt 1440
ttaaaccacc ggctacttct tggagaggaa aaatggcttt tataagactc aaaaaacaaa 1500
gctctgcaag tcaaatgcat acaaaactgt tctgtaggtc tggaatcagg acaactatgtg 1560
gaagtcaaat agagcagctt taaaaagcct ttgggatcat tctcatctta tatttgacg 1620
acgatactat gacagtgata actgacataa ctgcatcaat ttccttgata ttttatttgt 1680
cttaagtagc aagacataga gatggacgta aagatggaca tatgactcag gtctggacag 1740
gtccgtggtc catgtatgat aaaagagatg aaggaagga gaattgagac tgtctaagaa 1800
gggcttcagg gacgttctga aggcagattt gactgaatca gatgtactgt ccaagtctca 1860
tatgtagcaa tggaaggctg atattggaga aatataaaga aatggctgtg aactcaaagt 1920
```

-continued

gacctgaac agaaaagga tatggagta aaataatgtc acagaactga ggtttatatg	1980
atataccatg ggctgcagag ggtcagatg ctcaccatg ggctctctt gggctgcagg	2040
gaacttctgt tctacacctg gaacacctcc tgcctctctc cgcactgacc tcagtgtcat	2100
cagggtgtt tctctcatat tttctcactc acctctccca actaccattg tacagcagtt	2160
gttcttcatat attgtctctc ctgaggtaca tctagcatcg ttaagtctctc agacttggca	2220
aggagaatgt agatttcac agtatatatg ttttcacaaa aggaaggaga gaaacaaaag	2280
aaaatggcac tgactaaact tcagctagtgt gtataggaaa gtaattctgc ttaacagaga	2340
ttgcagtgtat ctctatgtat gtcctgaaga attatgttgt acttttttcc cccattttta	2400
aatcaaacag tgctttacag aggtcagaat ggtttcttta ctgtttgtca attctattat	2460
ttcaatacag aacaatagct tctataactg aaatatattt gctattgtat attatgattg	2520
tcctcgaac catgaacact cctccagctg aatttcacaa tctctctgtc atctgccagg	2580
ccattaagtt attcatggaa gatctttgag gaacactgca agttcatatc ataaacacat	2640
ttgaaattga gtattgtttt gcattgtatg gagctatgtt ttgctgtatc ctcagaataa	2700
aagtttggtta taaagcatc acaccataa aaagatagat ttaaatattc caactatagg	2760
aaagaaagtgt tgtctgtctt tcactctagt ctcagttggc tccttcacat gcacgcttct	2820
ttatttctcc tattttgtca agaaaataat aggtcaagtc ttgttctcat ttatgtctg	2880
tctagcgtgg ctcagatgca cattgtacat acaagaagga tcaaatgaaa cagacttctg	2940
gtctgttact acaaccatag taataagcac actaactaat aattgctaata tatgttttcc	3000
atctccaagg ttcaccattt tttctgtttt cttaaagatc ccattatctg gttgtaactg	3060
aagctcaatg gaacatgagc aatatttccc agtcttctct cccatccaac agtctgatg	3120
gattagcaga acaggcagaa aacacattgt taccagaaat taaaaactaa tatttgcctt	3180
ccattcaatc caaatggac ctattgaaac taaaatctaa ccaatccca ttaaatgatt	3240
tctatgggtgt caaagggtcaa acttctgaag ggaacctgtg ggtgggtcac aattcagact	3300
atatattccc cagggtctag ccagtgctctg tacatacagc tagaaagctg tattgccttt	3360
agcagtcag ctcgaagggt aagcaactct ctggaattac cttctctcta tattagctct	3420
tacttgacc taaactttaa aaaattaaca attattgtgc tatgtgttgt atctttaagg	3480
gtgaagtacc tgctgtatc cccctataaa aacttctcac ctgtgtatgc attctgcact	3540
atatttattat gtgtaaaagc tttgtgtttg ttttcaggag gcttattctt tgtgcttaa	3600
atatgttttt aatttcagaa catcttatcc tgcgttcac tatctgatat gctttgcagt	3660
ttgcttgatt aacttctagc cctacagagt gcacagagag caaaatcatg gtgttcagt	3720
aattctgggg agttatttta atgtgaaaat tctctagaag ttttaattctt gcaaagtga	3780
gctgtgtatc actacacaag ataaaaatgt ggggggtgca taaacgtata ttcttacaat	3840
aatagatata tgtgaactta tatacagaaa agaaaatgag aaaaatgtgt gtgtgtatac	3900
tcacacacgt ggtcagtaaa aacttttgag gggtttaata cagaaaatcc aatcctgagg	3960
ccccagcact cagtagcat ataaagggt gggctctgaa ggacttctga ctttcacaga	4020
ttatataaat ctcaggaaa caactagatt catgctggct ccaaaagctg tgctttatat	4080
aagcacactg gctatacaat agttgtacag ttcagctctt tataatagaa acagacagaa	4140
caagtataaa tcttctattg gtctatgtca tgaacaagaa ttcattcagt ggctctgttt	4200

-continued

tatagtaaac attgctat	tattcatgtct	gcattttctct	tctgtctgaa	tgtcaccact	4260
aaaatttaac tccacagaaa	gtttatacta	cagtacacat	gcatactctt	gagcaaagca	4320
aaccatacct gaaagtgc	aa tagagcagaa	tatgaattac	atgcgtgtct	ttctcctaga	4380
ctacatgacc ccatataaat	tacatttcct	atctattctg	ccatcaccaa	aacaaaggta	4440
aaaaacttt tgaagatcta	ctcatagcaa	gtagtgtgca	acaacagat	atttctctac	4500
atttattttt agggataaaa	aataagaaat	aaaatagtca	gcaagcctct	gctttctcat	4560
atatctgtcc aaacctaaag	tttactgaaa	tttgcctctt	gaatttccag	ttttgcaagc	4620
ctatcagatt gtgtttta	aat cagaggtact	gaaaagtatc	aatgaattct	agctttcact	4680
gaacaaaaat atgtagaggc	aactggcttc	tgggacagtt	tgtaccctca	aagacaactg	4740
aatgcaaata cataaataga	tttatgaata	tggttttgaa	catgcacatg	agagggtgat	4800
atagcaacag acacattacc	acagaattac	tttaaaacta	cttggttaaca	tttaattgcc	4860
taaaaactgc tcgtaattta	ctggtgttagc	ctaccataga	gtaccctgca	tgttactatg	4920
tacagcattc catccttaca	ttttcactgt	tctgctgttt	gctctagaca	actcagagtt	4980
caccatgaaa atgcggttct	tggggttggt	ggtctgtttg	gttctctgga	cctgcatttc	5040
cgaggggtcc ggaggga	aac tgacagctgt	ggatccagaa	acaacatga	atgtcagtga	5100
aattatctct tactggg	gat tccctagtga	ggaataccta	gttgagacag	aagatggata	5160
tattctgtgc cttaaccgaa	ttcctcatgg	gaggaagaac	cattctgaca	aaggtcccaa	5220
accagttgtc ttctgcaac	atggcttgct	ggcagattct	agtaactggg	tcacaaacct	5280
tgccaacagc agcctggg	ct cattcttgc	tgatgctgg	tttgacgtgt	ggatgggcaa	5340
cagcagagga aatacctg	gt ctgggaaaca	taagacactc	tcagtttctc	aggatgaatt	5400
ctgggctttc agttatgat	g agatggcaaa	atatgaccta	ccagcttcca	ttaaacttcat	5460
tctgaataag actggcca	ag aacaagtgt	ttatgtggg	cattctcaag	gcaccactat	5520
aggttttata gcattttc	ac agatccctga	gctggctaaa	aggattaaaa	tgttttttgc	5580
cctgggtcct gtggcttc	ccg tcgccttctg	tactagccct	atggccaaac	tgggacgact	5640
gccagatcat ctcat	taagg acctcttgg	agacaaagaa	tttcttccc	agagtgcgtt	5700
tttgaagtgg ctgggtacc	c acgtttgcac	tcatgtcata	ctgaaggagc	tctgtggaaa	5760
tctctgtttt cttctgtgt	g gatttaatga	gagaaattta	aatatgtcta	gagtggatgt	5820
gtatacaaca cattctc	ctg ctggaacttc	tgtgcaaaac	atgttacact	ggagccaggc	5880
tgttaaatc caaaagt	ttc aagcctttga	ctggggaagc	agtgccaaga	attattttca	5940
ttacaaccag agttatc	ctc ccacatacaa	tgtgaaggac	atgcttgtgc	cgaactgcagt	6000
ctggagcggg ggtcacg	act ggcttgcaga	tgtctacgac	gtcaatatct	tactgactca	6060
gatcaccaac ttggtgt	tc atgagagcat	tccggaatgg	gagcatcttg	acttcatttg	6120
gggcctggat gcccctg	ga ggctttataa	taagattatt	aatctaata	ggaaatatca	6180
gtgattcgaa gcggccg	caa gaagaaagct	gaaaaactct	gtcccttcca	acaagacca	6240
gagcactgta gtatcagg	ggg taaaatgaaa	agtatgttat	ctgctgcac	cagacttcat	6300
aaaagctgga gctta	atctc taaaaaaat	cagaaagaaa	ttacactgtg	agaacagggtg	6360
caattcactt ttcctt	taca cagagtaata	ctggtaactc	atggatgaag	gcttaaggga	6420
atgaaattgg actcac	agta ctgagtcac	acactgaaaa	atgcaacctg	atacatcagc	6480
agaaggttta tggggg	aaaa atgcagcctt	ccaattaagc	cagatatctg	tatgaccaag	6540

-continued

```

ctgtccaga attagtcact caaaatctct cagattaaat tatcaactgt caccaaccat 6600
tcctatgctg acaaggcaat tgcttggtct ctgtgttctt gatactacaa ggctcttcct 6660
gacttcctaa agatgcatta taaaaatctt ataattcaca ttctcccta aactttgact 6720
caatcatggt atgttgga atatggtata ttactattca aattgttttc ctgtaccca 6780
tatgtaatgg gtcttgtaa tgtgtctctt tgttccttta atcataataa aaacatgttt 6840
aagcaaacac ttttcacttg tagtatttga aggtaccgga tctcgagccg cettcaatgc 6900
ccccaaaacc aatccccagg tttttaactc tcccgatttt ccaagtacca tagcccgtg 6960
agagagcgcc gcggtaatgg gatcccagga ccccggggaa tataagtctg agggggacgt 7020
aagcaacctt tccttttgta acagggacaa catagccctt atttccttct tagaaggaga 7080
ggttttcccg caataggtct tacacgcgga cgaaatcacc tttatgacgg ctccatgct 7140
tgatccaccg gcgcaccgga atcacgcaga gcaaccgga tcacgcctgg ggtggaccgc 7200
tcagtcgtcg ggcttccttc ccgtcttcca acgactctct gagttctcgg tagggatgt 7260
tgccccctg cagtagggct ccctccgacg ccactcagct tctgccctcc taagccgcag 7320
ccccctctac tagggtcacg gtccgctccc cgaataagcg agacggatga ggacaggatc 7380
gccacgcgcg ctgtggccga ccactattcc ctaacgatca cgtcggggtc accaaatgaa 7440
gccttctgct tcatgcatgt gctcgtagtc gtcagggaat caacggctcg gccatcaacc 7500
caggtgcaca ccaatgtggt gaatgggtcaa atggcggtta ttgtatcgag ctaggcactt 7560
aaatacaata tctctgcaat gcggaattca gtggttcgtc caatccgtgt tagaccgctc 7620
tgttgccttc ctaacaagge acgatcatc cacgatcata ccaccttact cccaccaatc 7680
ggcatgcacg gtgctttttc tctcctata aggcattgtg ctaactcatc gttacataag 7740
catgttgcaa gactacaaga gtattgcata agactacatt 7780

```

```

<210> SEQ ID NO 8
<211> LENGTH: 10762
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: pALVIN-OV-1.1-I
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2735)
<223> OTHER INFORMATION: n is a, c, g, or t
<400> SEQUENCE: 8

```

```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600

```

-continued

tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	660
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	720
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	780
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	840
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	900
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	960
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1020
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1080
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1140
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1200
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1260
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1320
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1380
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1440
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1500
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1560
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1620
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1680
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1740
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1800
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1860
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1920
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	1980
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2040
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2100
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2160
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2220
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2280
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2340
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2400
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2460
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2520
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2580
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2640
tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	tttttttttttt	2700
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnttcga	agcggccgca	agaagaaagc	2760
tgaaaaaactc	tgtcccttcc	aacaagaccc	agagcactgt	agtatcaggg	gtaaaaatgaa	2820
aagtatgtta	tctgctgcat	ccagacttca	taaaagctgg	agcttaatct	agaaaaaaaa	2880
tcagaaagaa	attacactgt	gagaacaggt	gcaattcact	tttcctttac	acagagtaat	2940

-continued

actggtaact catggatgaa ggcttaaggg aatgaaattg gactcacagt actgagtcac	3000
cacactgaaa aatgcaacct gatacatcag cagaagggtt atgggggaaa aatgcagcct	3060
tccaattaag ccagatatct gtatgaccaa gctgctccag aattagtcac tcaaaatctc	3120
tcagattaaa ttatcaactg tcaccaacca ttcttatgct gacaaggcaa ttgcttggtc	3180
tctgtgttcc tgatactaca aggctcttcc tgacttccta aagatgcatt ataaaaatct	3240
tataattcac atttctccct aaactttgac tcaatcatgg tatgttggca aatatggtat	3300
attactattc aaattgtttt ccttgtagcc atatgtaatg ggtcttgatg atgtgctctt	3360
ttgttccttt aatcataata aaaacatggt taagcaaaca cttttcactt gtagtatttg	3420
aaggtagcgg atctcgagcc gccttcaatg cccccaaaac caatcccag gtttttaact	3480
ctcccgattt tccaagtacc atagcccgct gagagagcgc cgcggtaatg ggatcccagg	3540
accccgggga atataagtct gagggggacg taagcaaccc ttcttttgt aacagggaca	3600
acatagcccc tatttctctc ttagaaggag aggttttccc gcaataggtc ttacacgcgg	3660
acgaatcac ctttatgacg gcttccatgc ttgatccacc gggcgaccgg aatcacgcag	3720
agcaaccgga atcacgcctg ggggtggacc ctcagtcgtc gggcttctt cccgtcttcc	3780
aacgactctc tgagtctctg gtagggtatg ttggccccct gcagtagggc tccctccgac	3840
gccactcagc ttctgcctc ctaagccgca gccccctcta ctagggtcat cgtccgctcc	3900
ccgaataagc gagacggatg aggacaggat cgccacgccg cctgtggccg accactattc	3960
cctaacgac acgtcggggg caccaaatga agccttctgc ttcatgcatg tgctcgtagt	4020
cgtcagggaa tcaacgggtc ggccatcaac ccagggtcac accaatgtgg tgaatgggtc	4080
aatggcggtt attgtatcga gctaggcact taaatacaat atctctgcaa tgcggaattc	4140
agtgttctgt ccaatccgtg ttagaccctg ctgttgctt cctaacaagg cacgatcata	4200
ccacgatcat accaccttac tcccaccaat cggcatgcac ggtgcttttt ctctccttat	4260
aaggcatggt gctaactcat cgttacataa gcatgttgca agactacaag agtattgcat	4320
aagactacat ttccccctcc ctatgcaaaa gcgaaactac tatatcctga ggggactcct	4380
aaccgcgtac aaccgaagcc ccgcttttct cctaacaatg ctattgtccc ctcagtcaag	4440
ccttgccctg tacaaccgga ttcgcaagcc ttgcctccc cacattatcc gtagcattat	4500
ttcttagcag tcatcagagc tacagaagat actctatgct gtagccaagt ctacaagttt	4560
actattcagc gaectcctat attccgctg ccagccgac aattaccaat gcgcgcttg	4620
cgtaatcatg gtcatactg tttctgtgt gaaattgtta tccgctcaca attccacaca	4680
acatacgagc cggaagcata aagtgtaaa cctgggggtgc ctaatgagtg agctaactca	4740
cattaattgc gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg tgccagctgc	4800
attaatgaat cggccaacgc gcggggagag gcggtttgct tattgggcgc tcttccgctt	4860
cctcgctcac tgactcgctg cgctcggtcg ttccggtcg gcgagcggta tcagctcact	4920
caaaggcgg aatacgggta tccacagaat caggggataa cgcaggaaag aacatgtgag	4980
caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg tttttccata	5040
ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg tggcgaaacc	5100
cgacaggact ataaagatac caggcggttc cccctggaag ctcctcgtg cgctctctg	5160
ttccgacct gcgccttacc ggatacctgt ccgcctttct cccttcggga agcgtggcgc	5220

-continued

tttctcatag	ctcacgctgt	aggtatctca	gttcggtgta	ggtcggttcgc	tccaagctgg	5280
gctgtgtgca	cgaaccccc	gttcagcccg	accgctgcgc	cttatccggt	aactatcgtc	5340
ttgagtccaa	cccggttaaga	cacgacttat	cgcactggc	agcagccact	ggtaacagga	5400
ttagcagagc	gaggtatgta	ggcggtgcta	cagagtctct	gaagtggtagg	cctaactacg	5460
gctacactag	aaggacagta	tttggatatc	gcgctctgct	gaagccagtt	accttcggaa	5520
aaagagttgg	tagctcttga	tccggcaaac	aaaccaccgc	tggtagcggg	ggtttttttg	5580
tttgcaagca	gcagattacg	cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	5640
ctacggggtc	tgacgctcag	tggaaacgaaa	actcacgtta	agggattttg	gtcatgagat	5700
tatcaaaaag	gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt	aatcaatct	5760
aaagtatata	tgagtaaact	tggctcgaca	gttaccaatg	cttaatcagt	gaggcaccta	5820
tctcagcgat	ctgtctatct	cgttcaccca	tagttgcttg	actccccgtc	gtgtagataa	5880
ctacgatacg	ggagggtcta	ccatctggcc	ccagtgcctg	aatgataccg	cgagaccac	5940
gctcaccggc	tccagattta	tcagcaataa	accagccagc	cgggaaggcc	gagcgcagaa	6000
gtggtcctgc	aactttatcc	gcctccatcc	agtctattaa	ttgttgccgg	gaagctagag	6060
taagtagttc	gccagttaat	agtttgcgca	acgttggtgc	cattgctaca	ggcatcgtgg	6120
tgtcacgctc	gtcgtttggt	atggcttcat	tcagctccgg	ttcccaacga	tcaaggcgag	6180
ttacatgata	ccccatgttg	tgcaaaaaag	cggttagctc	cttcgggtcct	ccgatcgttg	6240
tcagaagtaa	gttgcccgca	gtgttatcac	tcattggttat	ggcagcactg	cataattctc	6300
ttactgtcat	gccatccgta	agatgctttt	ctgtgactgg	tgagtactca	accaagtcac	6360
tctgagaata	gtgtatgcgg	cgaccgagtt	gctcttgccc	ggcgtcaata	cgggataata	6420
ccgcgccaca	tagcagaact	ttaaaagtgc	tcattcattgg	aaaacgttct	tcggggcgaa	6480
aactctcaag	gatcttaccg	ctgttgagat	ccagttcgat	gtaaccact	cgtgcacca	6540
actgatcttc	agcatctttt	actttcacca	gcgtttctgg	gtgagcaaaa	acaggaaggc	6600
aaaaatgccg	aaaaaaggga	ataagggcga	cacggaaaatg	tgataactc	atactcttcc	6660
tttttcaata	ttattgaagc	atttatcagg	gttattgtct	catgagcgga	tacatatttg	6720
aatgtattta	gaaaaataaa	caaatagggg	ttccgcgcac	atttccccga	aaagtgccac	6780
ctgacgcgcc	ctgtagcggc	gcattaagcg	cggcggtgtg	ggtaggttacg	cgcagcgtga	6840
ccgctacact	tgccagcgcc	ctagcgcgg	ctcctttcgc	tttcttcct	tcctttctcg	6900
ccacgttcgc	cggctttccc	cgtcaagctc	taaatcgggg	gctcccttta	gggttccgat	6960
ttagtgtctt	acggcacctc	gaccccaaaa	aacttgatta	gggtgatggg	tcacgtagtg	7020
ggccatcgcc	ctgatagacg	gtttttcgcc	ctttgacgtt	ggagtccacg	ttctttaata	7080
gtggactctt	gttccaaact	ggaacaacac	tcaaccctat	ctcgggtctat	tcttttgatt	7140
tataagggat	tttgccgatt	tcggcctatt	gggttaaaaa	tgagctgatt	taacaaaaat	7200
ttaacgcgaa	ttttaacaaa	atattaacgc	ttacaatttc	cattcgccat	tcaggctgcg	7260
caactgttgg	gaaggcgcat	cggtcggggc	ctcttcgcta	ttacgccagc	tggcgaaagg	7320
gggatgtgct	gcaaggcgat	taagttgggt	aacgccaggg	ttttccagct	cacgacgttg	7380
taaaacgacg	gccagtgcgc	gcgtattccc	taacgatcac	gtcgggggtca	ccaaatgaag	7440
ccttctgctt	catgcatgtg	ctcgtagtcg	tcagggaatc	aacgggtccg	ccatcaaccc	7500
aggtgcacac	caatgtgggt	aatgggtcaaa	tggcggttat	tgtatcgagc	taggcactta	7560

-continued

aatacaatat	ctctgcaatg	cggaattcag	tggttcgtcc	aatccgtccc	cctccctatg	7620
caaaagcgaa	actactatat	cctgagggga	ctcctaaccg	cgtacaaccg	aagccccgct	7680
tttcgcctaa	acatgctatt	gtcccctcag	tcaagccttg	cccgttacaa	cccgattcgc	7740
aagccttgcc	ctccccacat	tatccgtagc	attatttcct	agcagtcac	agagctacag	7800
aagatactct	atgctgtagc	caagtctaca	agtttactat	tcagcgacct	cctatatctc	7860
gcgtgccagc	cgatcaatta	ccaatccaac	cagctatcac	acggaataca	agaactcgcc	7920
tacgtctctc	tttcgggctg	cttataagcc	tctgttaatt	tttttatatt	cctcgtaag	7980
tcctcagact	tggcaaggag	aatgtagatt	tccacagtat	atatgttttc	acaaaaggaa	8040
ggagagaaac	aaaagaaaat	ggcactgact	aaacttcagc	tagtggtata	ggaaagtaat	8100
tctgcttaac	agagattgca	gtgatctcta	tgtatgtcct	gaagaattat	gttgactctt	8160
ttcccccat	ttttaaatca	aacagtgcct	tacagaggtc	agaatgggtt	ctttactgtt	8220
tgcaattct	attatttcaa	tacagaacaa	tagcttctat	aactgaaata	tatttgcctat	8280
tgtatattat	gattgtccct	cgaacctga	acactcctcc	agctgaattt	cacaattcct	8340
ctgtcatctg	ccaggccatt	aagttattca	tggaagatct	ttgaggaaca	ctgcaagttc	8400
atatcataaa	cacatttgaa	attgagtatt	gttttgctt	gtatggagct	atgttttgct	8460
gtatcctcag	aataaaagt	tggtataaag	cattcacacc	cataaaaaga	tagatttaaa	8520
tattccaact	ataggaaga	aagtgtgtct	gctcttcact	ctagtctcag	ttggctcctt	8580
cacatgcacg	cttctttatt	tctctattt	tgtcaagaaa	ataataggtc	aagtcttggt	8640
ctcatttatg	tctgtctag	cgtggctcag	atgcacattg	tacatacaag	aaggatcaaa	8700
tgaacacagc	ttctggtctg	ttactacaac	catagtaata	agcacactaa	ctaataattg	8760
ctaattatgt	ttccatctc	caaggttccc	acatttttct	gttttcttaa	agatcccat	8820
atctggttgt	aactgaagct	caatggaaca	tgagcaatat	ttcccagtct	tctctcccat	8880
ccaacagtcc	tgatggatta	gcagaacagg	cagaaaacac	attgttacct	agaattaaaa	8940
actaatattt	gctctccatt	caatccaaaa	tggacctatt	gaaactaaaa	tctaaccctaa	9000
ttccattaaa	tgatttctat	ggtgtcaaag	gtcaaacttc	tgaagggaac	ctgtgggtgg	9060
gtcacaattc	agactatata	ttccccaggg	ctcagccagt	gtctgtacat	acagctagaa	9120
agctgtattg	ccttttagcag	tcaagctcga	aaggtaagca	actctctgga	attaccttct	9180
ctctatatta	gctcttactt	gcacctaaac	tttaaaaaat	taacaattat	tgtgctatgt	9240
gttgatctct	taagggtgaa	gtacctgctg	gataccccct	ataaaaactt	ctcacctgtg	9300
tatgcattct	gcactatttt	attatgtgta	aaagctttgt	gtttgttttc	aggaggctta	9360
ttctttgtgc	ttaaaatatg	tttttaattt	cagaacatct	tatcctgtcg	ttcactatct	9420
gatatgcttt	gcagtttgct	tgattaactt	ctagccctac	agagtgcaca	gagagcaaaa	9480
tcatggtgtt	cagtgaattc	tggggagtta	ttttaatgtg	aaaattctct	agaagtttaa	9540
ttctgcgaaa	gtgcagctgc	tgatcactac	acaagataaa	aatgtggggg	gtgcataaac	9600
gtatatctct	acaataatag	atacatgtga	acttatatac	agaaaagaaa	atgagaaaaa	9660
tgtgtgtgtg	tatactcaca	cacgtggtca	gtaaaaactt	ttgaggggtt	taatacagaa	9720
aatccaatcc	tgaggcccca	gcactcagta	cgcataataa	gggctgggct	ctgaaggact	9780
tctgactttc	acagattata	taaattctcag	gaaagcaact	agattcatgc	tggctccaaa	9840

-continued

```

agctgtgctt tatataagca cactggctat acaatagttg tacagttcag ctctttataa 9900
tagaaacaga cagaacaagt ataaatcttc tattgggtcta tgtcatgaac aagaattcat 9960
tcagtggctc tgttttatag taaacattgc tattttatca tgtctgcatt tctcttctgt 10020
ctgaatgtca ccactaaaat ttaactccac agaaagttta tactacagta cacatgcata 10080
tctttgagca aagcaaacca tacctgaaag tgcaatagag cagaatatga attacatgcg 10140
tgtctttctc ctgactaca tgaccccata taaattacat tccttatcta ttctgccatc 10200
accaaacaag aggtaaaaat acttttgaag atctactcat agcaagtagt gtgcaacaaa 10260
cagatatctc tctacattta tttttaggga ataaaaataa gaaataaaat agtcagcaag 10320
cctctgcttt ctcatatata tgtccaaacc taaagtttac tgaaatttgc tctttgaatt 10380
tccagttttg caagcctata agattgtgtt ttaatcagag gtactgaaaa gtatcaatga 10440
attctagctt tcaactgaaca aaaatatgta gaggcaactg gcttctggga cagtttgcta 10500
cccaaaaagc aactgaatgc aaatacataa atagatttat gaatatggtt ttgaacatgc 10560
acatgagagg tggatatagc aacagacaca ttaccacaga attactttaa aactacttgt 10620
taacatttaa ttgcctaaaa actgctcgta atttactgtt gtagcctacc atagagtacc 10680
ctgcatggta ctatgtacag cattccatcc ttacattttc actgttctgc tgtttgctct 10740
agacaactca gagttcacca tg 10762

```

```

<210> SEQ ID NO 9
<211> LENGTH: 242
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Adaptor sequence

```

```

<400> SEQUENCE: 9

```

```

cccgggttgt taacatttaa ttgcctaaaa actgctcgta atttactgtt gtagcctacc 60
atagagtacc ctgcatggta ctatgtacag cattccatcc ttacattttc actgttctgc 120
tgtttgctct agacaactca gagttcacca tgaaaatgcg gttcttgggg ttggtgggtc 180
gtttggttct ctggaccctg cattccgagg ggtccggagg gaaactgaca gctgtggatc 240
ct 242

```

```

<210> SEQ ID NO 10
<211> LENGTH: 1575
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Syn SBC102

```

```

<400> SEQUENCE: 10

```

```

ccattatctg gttgtaactg aagctcaatg gaacatgagc aatatttccc agtcttctct 60
cccatccaac agtcctgatg gattagcaga acaggcagaa aacacattgt taccagaat 120
taaaaactaa tatttgctct ccattcaatc caaaatggac ctattgaaac taaaatctaa 180
cccaatccca ttaaagtatt tctatgggtg caaagggtcaa acttctgaag ggaacctgtg 240
ggtgggtcac aattcagact atatattccc cagggctcag ccagtgtctg tacatacagc 300
tagaaagctg tattgccttt agcagtcaag ctgaaagac aactcagagt tcaccatgaa 360
aatgcgggtc ttggggttgg tgggtctgtt gggtctctgg accctgcatt ccgaggggtc 420
cggaggggaaa ctgacagctg tggatcctga aacaacatg aatgtcagtg aaattatctc 480

```

-continued

ttactgggga	ttccctagtg	aggaatacct	agttgagaca	gaagatggat	atattctgtg	540
ccttaaccga	attcctcatg	ggaggaagaa	ccattctgac	aaagggtccca	aaccagttgt	600
cttctgcaa	catggcttgc	tggcagattc	tagtaactgg	gtcacaaacc	ttgccaacag	660
cagctggggc	ttcattcttg	ctgatgctgg	ttttgacgtg	tggatgggca	acagcagagg	720
aaatacctgg	tctcggaaac	ataagacact	ctcagtttct	caggatgaat	tctgggcttt	780
cagttatgat	gagatggcaa	aatatgacct	accagcttcc	attaacttca	ttctgaataa	840
gactggccaa	gaacaagtgt	attatgtggg	tcattctcaa	ggcaccacta	taggttttat	900
agcattttca	cagatccctg	agctggctaa	aaggattaaa	atgttttttg	ccctgggtcc	960
tgtggcttcc	gtcgcttctt	gtactagccc	tatggccaaa	ctgggacgac	tgccagatca	1020
tctcattaag	gacctctttg	gagacaaaga	atttcttccc	cagagtgcgt	tttgaaagtg	1080
gctgggtacc	cacgtttgca	ctcatgtcat	actgaaggag	ctctgtggaa	atctctgttt	1140
tcttctgtgt	ggatttaatg	agagaaatgt	aaatatgtct	agagtggatg	tgtatacaac	1200
acattctcct	gctggaactt	ctgtgcaaaa	catgttacac	tggagccagg	ctgttaaatt	1260
ccaaaagttt	caagcctttg	actggggaag	cagtgcgaag	aattattttc	attacaacca	1320
gagttatcct	cccacataca	atgtgaagga	catgcttggt	ccgactgcag	tctggagcgg	1380
gggtcacgac	tggtctgcag	atgtctacga	cgtcaatata	ttactgactc	agatcaccaa	1440
cttgggtgtc	catgagagca	ttccggaatg	ggagcatctt	gacttcattt	ggggcctgga	1500
tggcccttgg	aggctttata	ataagattat	taatctaata	aggaaatata	agtgattcga	1560
agcggccgcc	ccggg					1575

<210> SEQ ID NO 11

<211> LENGTH: 2789

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: OVR1 promoter

<400> SEQUENCE: 11

ctcagagtctc	ttcagaatgg	cacagcaccg	ctgcagaaaa	atgccagggtg	gactatgaac	60
tcacatccaa	aggagcttga	cctgatacct	gattttcttc	aaacagggga	aacaacacaa	120
ttccacaaaa	cagctcagag	agaaaccatc	actgatggct	acagcaccaa	ggtatgcaat	180
ggcaatccat	tcgacattca	tctgtgacct	gagcaaaatg	atttatctct	ccatgaatgg	240
ttgtctcttt	ccctcatgaa	aaggcaatgt	ccacactcac	aatatgcaac	aaagacaaac	300
agagaacaat	taatgtgtct	cttcctaata	ttaaaattgt	agtggcaaag	aggagaacaa	360
aatctcaagt	tctgagtagg	ttttagtgtg	tggataagag	gctttgacct	gtgagctcac	420
ctggacttca	tatccttttg	gataaaaagt	gcttttataa	ctttcaggtc	tccgagtctt	480
tattcatgag	actgttggtt	tagggacaga	cccacaatga	aatgcctggc	ataggaaagg	540
gcagcagagc	cttagctgac	cttttcttgg	gacaagcatt	gtcaaacaat	gtgtgacaaa	600
actatttgta	ctgctttgca	cagctgtgct	gggcagggca	atccattgcc	acctatccca	660
ggtaaccttc	caactgcaag	aagattgttg	cttactctct	ctagaccccc	aagtcaaac	720
aactatgcag	gtatgctgac	aacactatga	tgacagcctg	ttctgatcaa	gatctcattt	780
gttcattggac	aatttttgtt	gcttgcagct	ggctctccat	tgggaaagag	tgtagtatat	840

-continued

ccttctcctc	tgacagaaaa	gcagaaatto	tcctgctcca	cacttaattc	acattgtttt	900
aaaccaccgg	ctacttcttg	gagaggaaaa	atggctttta	taagactcac	aaaacaaagc	960
tctgcaagtc	aaatgcatac	aaaactgttc	tgtaggctcg	gaatcaggac	actatgtgga	1020
agtcaaatag	agcagcttta	aaaagccttt	gggatcattc	tcactttata	tttgacgac	1080
gatactatga	cagtataaac	tgacataact	gcatacaatt	ccttgatatt	ttatttgtct	1140
taaagtacaa	gacatagaga	tgagcgtaaa	gatggacata	tgactcaggt	ctggacaggt	1200
ccgtgggtcca	tgtatgataa	aagagatgaa	gggaaggaga	attgagactg	tctaagaagg	1260
gcttcaggga	cgttctgaag	gcagatttga	ctgaatcaga	tgtactgtcc	aagtctcata	1320
tgtagcaatg	gaaggctgat	attggagaaa	tataaagaaa	tggtgtgtaa	ctcaaagtga	1380
ccctgaacag	aaaagggata	tgaggttaaa	ataatgtcac	agaactgagg	tttatatgat	1440
ataccatggg	ctgcagaggg	tcagagtgtc	ccaccatggg	cctctcttgg	gctgcaggga	1500
acttctgttc	tacacctgga	acacctctcg	ccctctctcg	cactgacctc	agtgtcatca	1560
gggctgtttc	tctcacattt	tctcactcac	ctctcccaac	taccattgta	cagcagttgt	1620
tettacatat	tgtctctctc	gaggtagatc	tagcatcggt	aagtcctcag	acttgccaag	1680
gagaatgtag	atttccacag	tatatatggt	ttcacaaaag	gaaggagaga	aacaaaagaa	1740
aatggcactg	actaaacttc	agctagtggg	ataggaaaag	aattctgctt	aacagagatt	1800
gcagtgatct	ctatgtatgt	cctgaagaat	tatgtgttac	ttttttcccc	cattttttaa	1860
tcaaacagtg	ctttacagag	gtcagaatgg	tttctttact	gtttgtcaat	tctattatct	1920
caatacagaa	caatagcttc	tataactgaa	atatatttgc	tattgtatat	tatgattgtc	1980
cctcgaacca	tgaacactcc	tccagctgaa	tttcacaatt	cctctgtcat	ctgccaggcc	2040
attaagttaa	tcagggaaga	tctttgagga	acactgcaag	ttcatatcat	aaacacattt	2100
gaaattgagt	attgtttttg	attgtatgga	gctatgtttt	gctgtatcct	cagaataaaa	2160
gtttgttata	aagcattcac	accataaaaa	agatagattt	aaatattcca	actataggaa	2220
agaaagtgtg	tctgtctctc	actctagtct	cagttggctc	cttcacatgc	acgtctcttt	2280
atttctccta	ttttgtcaag	aaaataatag	gtcaagtctt	gttctcattt	atgtcctgtc	2340
tagcgtggct	cagatgcaca	ttgtacatac	aagaaggatc	aaatgaaaca	gacttctggg	2400
ctgttactac	aaccatagta	ataagcacac	taactaataa	ttgctaatta	tgttttccat	2460
ctccaagggt	cccacatttt	tctgttttct	taaagatccc	attatctggg	tgtaaactgaa	2520
gctcaatgga	acatgagcaa	tatttcccag	tcttctctcc	catccaacag	tcttgatgga	2580
ttagcagaac	aggcagaaaa	cacattgtta	cccagaatta	aaaactaata	tttgctctcc	2640
attcaatcca	aaatggacct	attgaaacta	aaatctaacc	caatcccatt	aaatgatctc	2700
tatggtgtca	aaggctcaaac	ttctgaaggg	aacctgtggg	tgggtcacia	ttcagactat	2760
atattcccca	gggctcagcc	agtgtctgt				2789

<210> SEQ ID NO 12

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 12

agaaactgag agtgtcttat

20

-continued

<210> SEQ ID NO 13
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 13

tgacagctgt ggatccagaa acaaactg

29

<210> SEQ ID NO 14
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 14

gccgctcgag cgaggaatat aaaaaaatt

29

<210> SEQ ID NO 15
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 15

tccgcgcaca tttccccgaa

20

<210> SEQ ID NO 16
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 16

acgactggct tgcagatgct t

21

<210> SEQ ID NO 17
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 17

ccccaaatga agtcaagatg ct

22

<210> SEQ ID NO 18
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide

<400> SEQUENCE: 18

ccggaatgct ctcatggaac accaa

25

<210> SEQ ID NO 19
<211> LENGTH: 372
<212> TYPE: PRT

-continued

<213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 19

Ala Val Asp Pro Glu Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr
 1 5 10 15
 Trp Gly Phe Pro Ser Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr
 20 25 30
 Ile Leu Cys Leu Asn Arg Ile Pro His Gly Arg Lys Asn His Ser Asp
 35 40 45
 Lys Gly Pro Lys Pro Val Val Phe Leu Gln His Gly Leu Leu Ala Asp
 50 55 60
 Ser Ser Asn Trp Val Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile
 65 70 75 80
 Leu Ala Asp Ala Gly Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn
 85 90 95
 Thr Trp Ser Arg Lys His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe
 100 105 110
 Trp Ala Phe Ser Tyr Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser
 115 120 125
 Ile Asn Phe Ile Leu Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val
 130 135 140
 Gly His Ser Gln Gly Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile
 145 150 155 160
 Pro Glu Leu Ala Lys Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val
 165 170 175
 Ala Ser Val Ala Phe Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu
 180 185 190
 Pro Asp His Leu Ile Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro
 195 200 205
 Gln Ser Ala Phe Leu Lys Trp Leu Gly Thr His Val Cys Thr His Val
 210 215 220
 Ile Leu Lys Glu Leu Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe
 225 230 235 240
 Asn Glu Arg Asn Leu Asn Met Ser Arg Val Asp Val Tyr Thr Thr His
 245 250 255
 Ser Pro Ala Gly Thr Ser Val Gln Asn Met Leu His Trp Ser Gln Ala
 260 265 270
 Val Lys Phe Gln Lys Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala Lys
 275 280 285
 Asn Tyr Phe His Tyr Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val Lys
 290 295 300
 Asp Met Leu Val Pro Thr Ala Val Trp Ser Gly Gly His Asp Trp Leu
 305 310 315 320
 Ala Asp Val Tyr Asp Val Asn Ile Leu Leu Thr Gln Ile Thr Asn Leu
 325 330 335

-continued

Val Phe His Glu Ser Ile Pro Glu Trp Glu His Leu Asp Phe Ile Trp
340 345 350

Gly Leu Asp Ala Pro Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu Met
355 360 365

Arg Lys Tyr Gln
370

<210> SEQ ID NO 20

<211> LENGTH: 399

<212> TYPE: PRT

<213> ORGANISM: HOMO SAPIENS

<400> SEQUENCE: 20

Met Lys Met Arg Phe Leu Gly Leu Val Val Cys Leu Val Leu Trp Thr
1 5 10 15

Leu His Ser Glu Gly Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu
20 25 30

Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr Trp Gly Phe Pro Ser
35 40 45

Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr Ile Leu Cys Leu Asn
50 55 60

Arg Ile Pro His Gly Arg Lys Asn His Ser Asp Lys Gly Pro Lys Pro
65 70 75 80

Val Val Phe Leu Gln His Gly Leu Leu Ala Asp Ser Ser Asn Trp Val
85 90 95

Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly
100 105 110

Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn Thr Trp Ser Arg Lys
115 120 125

His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe Trp Ala Phe Ser Tyr
130 135 140

Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser Ile Asn Phe Ile Leu
145 150 155 160

Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val Gly His Ser Gln Gly
165 170 175

Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile Pro Glu Leu Ala Lys
180 185 190

Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val Ala Ser Val Ala Phe
195 200 205

Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu Pro Asp His Leu Ile
210 215 220

Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro Gln Ser Ala Phe Leu
225 230 235 240

Lys Trp Leu Gly Thr His Val Cys Thr His Val Ile Leu Lys Glu Leu
245 250 255

Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe Asn Glu Arg Asn Leu
260 265 270

-continued

Asn	Met	Ser	Arg	Val	Asp	Val	Tyr	Thr	Thr	His	Ser	Pro	Ala	Gly	Thr
	275						280					285			
Ser	Val	Gln	Asn	Met	Leu	His	Trp	Ser	Gln	Ala	Val	Lys	Phe	Gln	Lys
	290					295					300				
Phe	Gln	Ala	Phe	Asp	Trp	Gly	Ser	Ser	Ala	Lys	Asn	Tyr	Phe	His	Tyr
305					310					315				320	
Asn	Gln	Ser	Tyr	Pro	Pro	Thr	Tyr	Asn	Val	Lys	Asp	Met	Leu	Val	Pro
				325					330					335	
Thr	Ala	Val	Trp	Ser	Gly	Gly	His	Asp	Trp	Leu	Ala	Asp	Val	Tyr	Asp
				340				345					350		
Val	Asn	Ile	Leu	Leu	Thr	Gln	Ile	Thr	Asn	Leu	Val	Phe	His	Glu	Ser
	355						360					365			
Ile	Pro	Glu	Trp	Glu	His	Leu	Asp	Phe	Ile	Trp	Gly	Leu	Asp	Ala	Pro
	370					375					380				
Trp	Arg	Leu	Tyr	Asn	Lys	Ile	Ile	Asn	Leu	Met	Arg	Lys	Tyr	Gln	
385					390					395					

1.-72. (canceled)

73. A composition comprising an isolated human recombinant lysosomal acid lipase (rhLAL) comprising one or more N-glycan structures, wherein the rhLAL is N-linked glycosylated at Asn¹⁵, Asn⁸⁰, Asn¹⁴⁰, Asn²⁵² and Asn³⁰⁰ of SEQ ID NO:2.

74. The composition of claim **73**, wherein the LAL comprises a N-glycan structure comprising a phosphorylated mannose at Asn⁸⁰, Asn¹⁴⁰ or Asn²⁵² of SEQ ID NO:2.

75. The composition of claim **74**, wherein the LAL comprises an N-glycan structure comprising a phosphorylated mannose at Asn⁸⁰.

76. The composition of claim **75**, wherein at least 30% of N-glycan structures associated with Asn⁸⁰ have a phosphorylated mannose.

77. The composition of claim **74**, wherein the LAL comprises an N-glycan structure comprising a bisphosphorylated mannose at Asn⁸⁰.

78. The composition of claim **74**, wherein the LAL comprises an N-glycan structure comprising a phosphorylated mannose (M6P) at Asn¹⁴⁰.

79. The composition of claim **78**, wherein between about 10% and about 50% of N-glycan structures associated with Asn¹⁴⁰ comprise a M6P.

80. The composition of claim **74**, wherein the LAL comprises an N-glycan structure comprising a phosphorylated mannose (M6P) at Asn²⁵².

81. The composition of claim **80**, wherein at least 50% of the N-glycan structures at Asn²⁵² comprise a M6P.

82. The composition of claim **73**, wherein the LAL comprises an N-glycan structure comprising a high-mannose group at Asn⁸⁰ or Asn²⁵².

83. The composition of claim **82**, wherein the high mannose-group comprises 6, 7, 8, 9 or 10 mannose.

84. The composition of claim **83**, wherein the LAL comprises an N-glycan structure comprising 7, 8 or 9 mannose at Asn⁸⁰.

85. The composition of claim **83**, wherein the LAL comprises an N-glycan structure comprising 7, 8 or 9 mannose at Asn²⁵².

86. The composition of claim **73**, wherein the LAL comprises an N-glycan structure comprising a terminal galactose at Asn¹⁵, Asn¹⁴⁰ or Asn³⁰⁰.

87. The composition of claim **86**, wherein about 2% to about 10% of N-glycan structures associated with Asn¹⁵ comprise a terminal galactose.

88. The composition of claim **87**, wherein less than 5% of N-glycan structures associated with Asn¹⁴⁰ comprise a terminal galactose.

89. The composition of claim **88**, wherein the LAL comprises a terminal galactose at Asn³⁰⁰.

90. The composition of claim **73**, wherein Asn⁵¹ of the LAL is unglycosylated.

91. The composition of claim **73**, wherein the LAL comprises N-glycan structures having no xylose and less than 15% of the N-glycan structures contain sialic acid.

92. The composition of claim **91**, wherein the LAL comprises N-glycan structures having no xylose and less than 10% of the N-glycan structures contain sialic acid.

93. The composition of claim **92**, wherein the LAL comprises N-glycan structures having no xylose and less than 5% of the N-glycan structures contain sialic acid.

94. The composition of claim **73**, wherein the LAL contains no sialic acid and no xylose.

95. The composition of claim **73**, wherein the LAL contains no fucose.

96. The composition of claim **73**, wherein the LAL comprises a N-linked glycosylation profile as follows:

- a) at Asn¹⁵, GlcNAc4Man3GlcNAc2, or Gal1GlcNAc4Man3GlcNAc2;
- b) at Asn⁸⁰, Phos2Man7GlcNAc2;
- c) at Asn¹⁴⁰, Phos1Man6GlcNAc2, GlcNAc1Phos1Man6GlcNAc2; Man3GlcNAc2; GlcNAc2Man3GlcNAc2; GlcNAc3Man3GlcNAc2; GlcNAc4Man3GlcNAc2, or Gal1GlcNAc4Man3GlcNAc2;
- d) at Asn²⁵², Man7GlcNAc2, Man8GlcNAc2, Man9GlcNAc2,

Phos1Man8GlcNAc2, or
 Phos1Man9GlcNAc2; and
 e) at Asn³⁰⁰, GlcNAc2Man3GlcNAc2,
 GlcNAc3Man3GlcNAc2,
 GlcNAc4Man3GlcNAc2,
 Gal1GlcNAc4Man3GlcNAc2,
 GlcNAc5Man3GlcNAc2,
 Gal1GlcNAc5Man3GlcNAc2,
 GlcNAc6Man3GlcNAc2, or
 Gal1GlcNAc6Man3GlcNAc2,
 wherein Man=mannose,
 GlcNAc=N-Acetyl Glucosamine
 Phos=phosphate
 Gal=galactose.

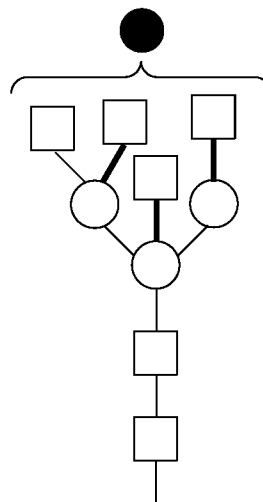
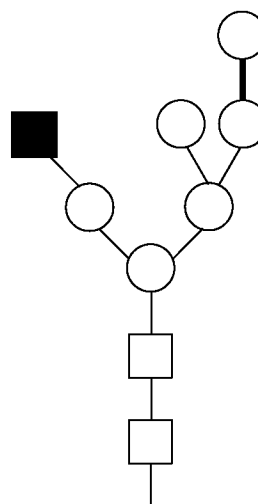
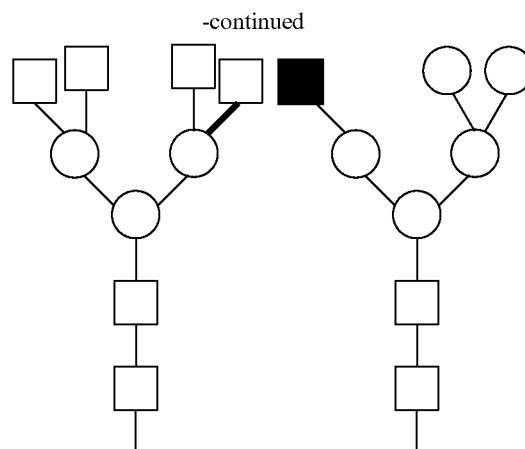
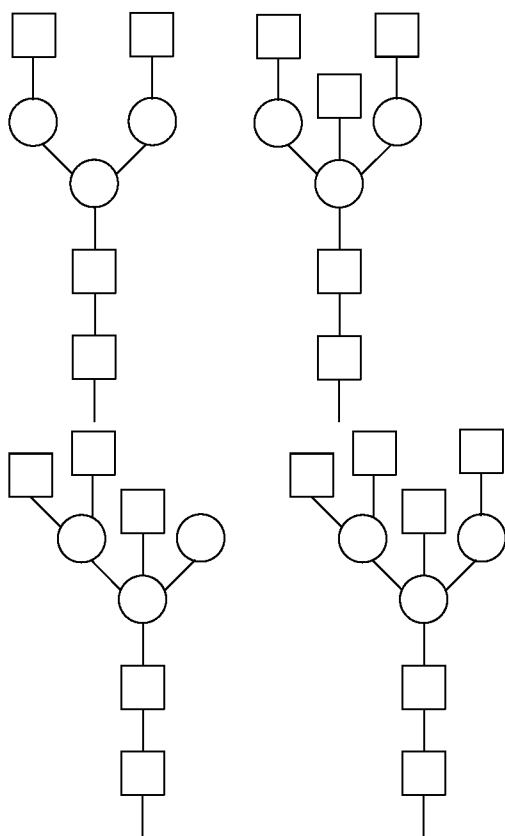
97. The composition of claim **73**, wherein the LAL is produced in a germ-line transgenic avian.

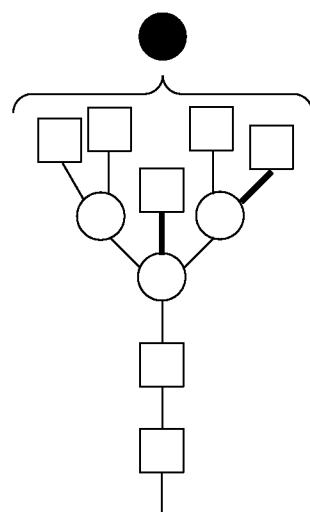
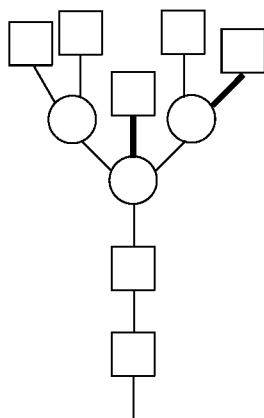
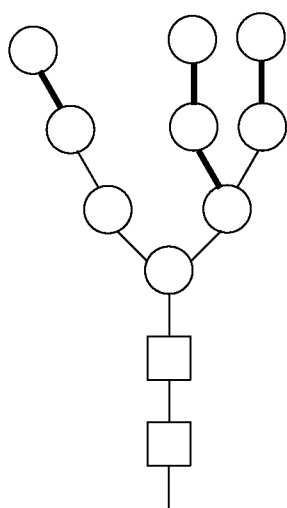
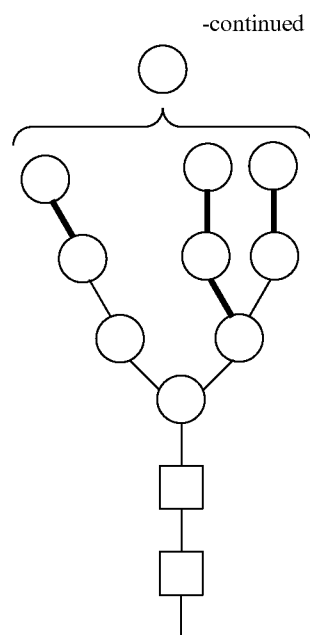
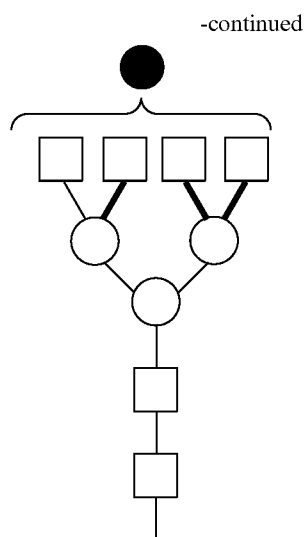
98. The composition of claim **97**, wherein the LAL is produced from an oviduct cell of the germ-line transgenic avian.

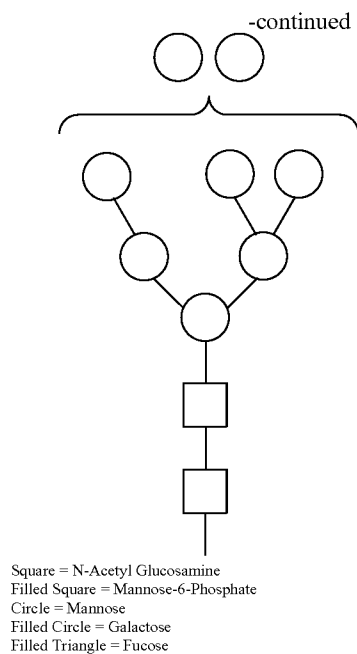
99. The composition of claim **98**, wherein the avian is a chicken.

100. A composition comprising a mixture of recombinant human lysosomal acid lipase (LAL), the mixture comprising at least two human LAL selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4 and SEQ ID NO:19.

101. The composition of claim **100**, wherein the mixture of human LAL comprises N-glycans selected from the following structures:







102. A method of producing glycosylated human lysosomal acid lipase (LAL), comprising creating a germ-line transgenic avian which contains a transgene encoding human LAL operably linked to a promoter and expresses the human LAL in an oviduct cell, wherein the LAL is glycosylated in the oviduct cell and deposited into a hard shell egg laid by the transgenic avian; and obtaining the LAL from egg white of the egg.

103. A transgenic avian which produces glycosylated human lysosomal acid lipase (LAL) according to claim **73** or claim **100**.

104. An egg laid by a transgenic avian according to claim **103**.

105. A pharmaceutical formulation comprising a composition according to claim **73** or claim **100** in combination with a pharmaceutically acceptable carrier, diluent or excipient.

106. The pharmaceutical formulation of claim **105**, wherein the formulation comprises at least one agent selected from the group consisting of trisodium citrate dehydrate, citric acid and human serum albumin.

107. The pharmaceutical formulation of claim **105**, wherein the formulation is provided in an aqueous solution maintained at pH between about 5.6 and about 6.2.

* * * * *