



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C07K	A2	(11) International Publication Number: WO 00/04041 (43) International Publication Date: 27 January 2000 (27.01.00)																		
(21) International Application Number: PCT/US99/16180 (22) International Filing Date: 16 July 1999 (16.07.99) (30) Priority Data: <table style="width: 100%;"> <tr> <td>09/118,205</td> <td>16 July 1998 (16.07.98)</td> <td>US</td> </tr> <tr> <td>09/122,449</td> <td>24 July 1998 (24.07.98)</td> <td>US</td> </tr> <tr> <td>09/244,444</td> <td>4 February 1999 (04.02.99)</td> <td>US</td> </tr> <tr> <td>09/273,447</td> <td>19 March 1999 (19.03.99)</td> <td>US</td> </tr> <tr> <td>09/350,836</td> <td>9 July 1999 (09.07.99)</td> <td>US</td> </tr> </table> (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application <table style="width: 100%;"> <tr> <td>US</td> <td>Not furnished (CIP)</td> </tr> <tr> <td>Filed on</td> <td>Not furnished</td> </tr> </table> (71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): FORD, John [US/US]; 2763 S. Norfolk, #210, San Mateo, CA 94403 (US). MULERO, Julio [US/US]; 892 Southampton Drive, Palo Alto, CA 94303 (US).	09/118,205	16 July 1998 (16.07.98)	US	09/122,449	24 July 1998 (24.07.98)	US	09/244,444	4 February 1999 (04.02.99)	US	09/273,447	19 March 1999 (19.03.99)	US	09/350,836	9 July 1999 (09.07.99)	US	US	Not furnished (CIP)	Filed on	Not furnished	(74) Agent: UHL, Jill, E.; Marshall, O'Toole, Gerstein, Murray & Borun, 6300 Sears Tower, 233 South Wacker Drive, Chicago, IL 60606 (US). (81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i> <i>With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.</i>
09/118,205	16 July 1998 (16.07.98)	US																		
09/122,449	24 July 1998 (24.07.98)	US																		
09/244,444	4 February 1999 (04.02.99)	US																		
09/273,447	19 March 1999 (19.03.99)	US																		
09/350,836	9 July 1999 (09.07.99)	US																		
US	Not furnished (CIP)																			
Filed on	Not furnished																			
(54) Title: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES																				
(57) Abstract The invention provides novel polynucleotides isolated from cDNA libraries of human fetal liver-spleen and macrophage as well as polypeptides encoded by these polynucleotides and mutants or variants thereof. The polypeptides correspond to a novel human CD39-like protein. Other aspects of the invention include vectors containing polynucleotides of the invention and related host cells as well as processes for producing novel CD39-like polypeptides, and antibodies specific for such polypeptides.																				
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 15%; vertical-align: top;"> CD39Human.seq 246 prot </td> <td style="width: 85%; vertical-align: top;"> <pre> M E D T H E S N V K T F E S K N I C A I L G E S S L T A V I L L A L A - - - - - 1 K A T L S W G I T V F - - - - - F H M V V S C V C S A W S H R N O C I A F E T F 2 Q N K - - - - - A L F E N K Y G I V L D A G S S H T E L Y L V Y K E F E E N D 3 L S I S M C P E N V S A L E T L Y G I M F D A G S T G I T R I K V Y T F V C K R P O C 4 I G V V H Q V F E E C R V K Q P G I S K F V O K V N E T I C Y L T T C M F E A R E 114 L P L U E G E V F D S I V E - P G L S A L F V I D Q P K Q G A E T V Q G O L L E V A K L 115 V I P R S C H O E T F V Y L G A T A G M R L L R M E S E E L A F R L I L V E R 184 S L P R S H W K K T P W V L K I A T A G L R L L - - - P I E H K I A K A L F E E K E 185 S I S N Y P F D F O - - - G A R I I T G O E E G A Y Q W I T I N Y L A C K F E E R 192 I F R K S I P E L V P K G S V S L M D L S I D E G I L A W V T V N F I L T S Q L - - - 197 I E W E S I V P Y E T N N Q E T F G A L D L G G A S T Q V T F L P C I N O T T E 231 - - - - - H G H R I O E T V I G T I D L G G A S T O L T E L P C F E R K T L E 236 S P D N A - - - - - L Q F P L Y G K D Y N V Y T H S F L C Y G K D Q A L G C R L A R D 259 Q T P R G Y L T S I E M F N S T L Y K L Y T H S Y L G F G L K A L - - - R L A T L 255 I O V A S N E F I R D P C F H E G Y K K V V N V S D I Y K T P C T R R - F E M T 305 G A L E L T E G - - - - - T D G H T F R I S A L C L P W L E A E 280 I L P F O O F - - - - - E I Q G I G N Y O O C H O S I L E L F N T S Y C P Y S O 342 W I I G G V K Y Q Y G G N Q E I G E V G F E P L Q A E V I L R V V R G K - - - - 314 C A E N G I F L P P I O G T F G A F S A F - - Y F V M K F L N L T S E K V S Q E 380 - - - - - L H Q I P E E V Q R I G S D Y A F S Y L Y D R - - - A V D I T D M I D Y E 345 K V T E M - M K K F C A O P W E - - E I K T S Y A G V K E K Y L S E Y C F S G T 417 G G G I L K V E D I E R K A R L E V C L N L E N F T S G S P - F L - - - C M D L S 382 Y I L S L L L O G Y E F T A D S W E H I H F I G K I O G S D A G W T L G Y H L N 437 Y T I T A L L K D G F G I E N D S I - - - V L Q L T K K V N N I E T G W A L G A T F H 419 L T N M I P A E O P L S T P L S H S T Y V F L H V L F S L V L F T V A I T G L I 497 I L Q S L G I S H 428 I F H K P S Y F W K K M V 510 </pre> </td> </tr> </table>			CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot	<pre> M E D T H E S N V K T F E S K N I C A I L G E S S L T A V I L L A L A - - - - - 1 K A T L S W G I T V F - - - - - F H M V V S C V C S A W S H R N O C I A F E T F 2 Q N K - - - - - A L F E N K Y G I V L D A G S S H T E L Y L V Y K E F E E N D 3 L S I S M C P E N V S A L E T L Y G I M F D A G S T G I T R I K V Y T F V C K R P O C 4 I G V V H Q V F E E C R V K Q P G I S K F V O K V N E T I C Y L T T C M F E A R E 114 L P L U E G E V F D S I V E - P G L S A L F V I D Q P K Q G A E T V Q G O L L E V A K L 115 V I P R S C H O E T F V Y L G A T A G M R L L R M E S E E L A F R L I L V E R 184 S L P R S H W K K T P W V L K I A T A G L R L L - - - P I E H K I A K A L F E E K E 185 S I S N Y P F D F O - - - G A R I I T G O E E G A Y Q W I T I N Y L A C K F E E R 192 I F R K S I P E L V P K G S V S L M D L S I D E G I L A W V T V N F I L T S Q L - - - 197 I E W E S I V P Y E T N N Q E T F G A L D L G G A S T Q V T F L P C I N O T T E 231 - - - - - H G H R I O E T V I G T I D L G G A S T O L T E L P C F E R K T L E 236 S P D N A - - - - - L Q F P L Y G K D Y N V Y T H S F L C Y G K D Q A L G C R L A R D 259 Q T P R G Y L T S I E M F N S T L Y K L Y T H S Y L G F G L K A L - - - R L A T L 255 I O V A S N E F I R D P C F H E G Y K K V V N V S D I Y K T P C T R R - F E M T 305 G A L E L T E G - - - - - T D G H T F R I S A L C L P W L E A E 280 I L P F O O F - - - - - E I Q G I G N Y O O C H O S I L E L F N T S Y C P Y S O 342 W I I G G V K Y Q Y G G N Q E I G E V G F E P L Q A E V I L R V V R G K - - - - 314 C A E N G I F L P P I O G T F G A F S A F - - Y F V M K F L N L T S E K V S Q E 380 - - - - - L H Q I P E E V Q R I G S D Y A F S Y L Y D R - - - A V D I T D M I D Y E 345 K V T E M - M K K F C A O P W E - - E I K T S Y A G V K E K Y L S E Y C F S G T 417 G G G I L K V E D I E R K A R L E V C L N L E N F T S G S P - F L - - - C M D L S 382 Y I L S L L L O G Y E F T A D S W E H I H F I G K I O G S D A G W T L G Y H L N 437 Y T I T A L L K D G F G I E N D S I - - - V L Q L T K K V N N I E T G W A L G A T F H 419 L T N M I P A E O P L S T P L S H S T Y V F L H V L F S L V L F T V A I T G L I 497 I L Q S L G I S H 428 I F H K P S Y F W K K M V 510 </pre>																
CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot CD39Human.seq 246 prot	<pre> M E D T H E S N V K T F E S K N I C A I L G E S S L T A V I L L A L A - - - - - 1 K A T L S W G I T V F - - - - - F H M V V S C V C S A W S H R N O C I A F E T F 2 Q N K - - - - - A L F E N K Y G I V L D A G S S H T E L Y L V Y K E F E E N D 3 L S I S M C P E N V S A L E T L Y G I M F D A G S T G I T R I K V Y T F V C K R P O C 4 I G V V H Q V F E E C R V K Q P G I S K F V O K V N E T I C Y L T T C M F E A R E 114 L P L U E G E V F D S I V E - P G L S A L F V I D Q P K Q G A E T V Q G O L L E V A K L 115 V I P R S C H O E T F V Y L G A T A G M R L L R M E S E E L A F R L I L V E R 184 S L P R S H W K K T P W V L K I A T A G L R L L - - - P I E H K I A K A L F E E K E 185 S I S N Y P F D F O - - - G A R I I T G O E E G A Y Q W I T I N Y L A C K F E E R 192 I F R K S I P E L V P K G S V S L M D L S I D E G I L A W V T V N F I L T S Q L - - - 197 I E W E S I V P Y E T N N Q E T F G A L D L G G A S T Q V T F L P C I N O T T E 231 - - - - - H G H R I O E T V I G T I D L G G A S T O L T E L P C F E R K T L E 236 S P D N A - - - - - L Q F P L Y G K D Y N V Y T H S F L C Y G K D Q A L G C R L A R D 259 Q T P R G Y L T S I E M F N S T L Y K L Y T H S Y L G F G L K A L - - - R L A T L 255 I O V A S N E F I R D P C F H E G Y K K V V N V S D I Y K T P C T R R - F E M T 305 G A L E L T E G - - - - - T D G H T F R I S A L C L P W L E A E 280 I L P F O O F - - - - - E I Q G I G N Y O O C H O S I L E L F N T S Y C P Y S O 342 W I I G G V K Y Q Y G G N Q E I G E V G F E P L Q A E V I L R V V R G K - - - - 314 C A E N G I F L P P I O G T F G A F S A F - - Y F V M K F L N L T S E K V S Q E 380 - - - - - L H Q I P E E V Q R I G S D Y A F S Y L Y D R - - - A V D I T D M I D Y E 345 K V T E M - M K K F C A O P W E - - E I K T S Y A G V K E K Y L S E Y C F S G T 417 G G G I L K V E D I E R K A R L E V C L N L E N F T S G S P - F L - - - C M D L S 382 Y I L S L L L O G Y E F T A D S W E H I H F I G K I O G S D A G W T L G Y H L N 437 Y T I T A L L K D G F G I E N D S I - - - V L Q L T K K V N N I E T G W A L G A T F H 419 L T N M I P A E O P L S T P L S H S T Y V F L H V L F S L V L F T V A I T G L I 497 I L Q S L G I S H 428 I F H K P S Y F W K K M V 510 </pre>																			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES

1. RELATED APPLICATIONS

5 This patent application is a continuation-in-part of U.S. patent application Serial No. 09/350,836 [Attorney Docket No. 28110/35761] filed July 9, 1999 which is a continuation-in-part of U.S. patent application Serial No. 09/273,447 filed March 19, 1999 which is a continuation-in-part of U.S. patent application Serial No. 09/122,449 filed July 24, 1998 and also a
10 continuation-in-part of U.S. patent application Serial No. 09/244,444 filed February 4, 1999, which in turn is a continuation of U.S. patent application Serial No. 09/118,205 filed July 16, 1998, the disclosures of all of which are incorporated by reference herein in their entirety.

15 2. FIELD OF THE INVENTION

 This invention relates in general to novel polynucleotides isolated from cDNA libraries of human fetal liver-spleen and macrophages and to polypeptides encoded by these polynucleotides. In particular, the invention relates to a human CD39-like protein with homologies to ATP
20 diphosphohydrolases and variants thereof.

3. BACKGROUND

 CD39 (cluster of differentiation 39) is a cell-surface molecule recognized by a "cluster" of monoclonal antibodies that can be used to
25 identify the lineage or stage of differentiation of lymphocytes and thus to distinguish one class of lymphocytes from another. This CD39 molecule was originally defined as a B lymphocyte marker (Rowe, M., et al. Int. J. Cancer 29:373 (1982)). Subsequent studies have shown CD39 to be a marker for a distinct subset of activated lymphocytes within the allosensitized
30 CD8-positive cytotoxic cells (Gouttefangeas C., et al., Eur. J. Immunol. 22:2681 (1992)). Outside of lymphoid tissue, CD39 can be found in quiescent vascular endothelial cells (Kansas, G. S., et al., J. Immunol.

146:2235 (1991)) and throughout rat brain in the neurons of the cerebral cortex, hippocampus, and cerebellum, as well as in glial cells (Wang, T-F. and Guidotti, G., Brain Res. 790:318 (1998)).

CD39 is a 510-amino acid protein with a predicted molecular mass of 57 kDa. However, because of heavy glycosylation at asparagine residues (six potential N-glycosylation sites) the molecule displays a mobility closer to 100 kDa (Maliszewski, C. R., et al., J. Immunol. 153:3574 (1994)). CD39 contains two hydrophobic regions, one near the amino terminus and the other near the carboxyl terminus which are believed to be transmembrane regions.

Reports that several ATP Diphosphohydrolases (ATPDases) share amino acid sequence homology with CD39 have been substantiated by showing that CD39 is itself an ATPDase (Wang, T- F., et al., J. Biol. Chem. 271:9898 (1996); Kaczmarek, E., et al., J. Biol. Chem. 271:33116 (1996)). Since CD39 is a plasma membrane-bound enzyme, CD39 has been termed an "ecto-ATPase," but CD39 is more often referred to as an "ecto-apyrase" because of the reduced rate of hydrolysis of ADP when compared with ecto-ATPases.

This activity has shown to modulate platelet reactivity and aggregation in response to vascular injury. During vascular injury, activated platelets aggregate forming an occlusive thrombus. Excessive platelet accumulation at sites of vascular injury can contribute to vessel occlusion. Endothelial cells respond to the potentially occlusive effects of platelet aggregation by several mechanisms. One of these mechanisms results ecto-apyrase-mediated removal of ADP, which in turn eliminates platelet reactivity and recruitment. It is now known that the endothelial ecto-apyrase responsible for this ADP removal is CD39 (Marcus, A. J., et al., J. Clin. Invest. 99:1351 (1997)).

Recently, CD39 was engineered to produce a soluble form of the molecule. This soluble CD39 was shown to display the same

nucleotidase activity as the membrane-bound molecule (Gayle, R. B., et al., J. Clin. Invest. 101:1851 (1998)). Intravenously administered soluble CD39 also remained active in mice for an extensive period of time, indicating that soluble CD39 could be useful as a inhibitor of platelet aggregation in the prophylaxis or treatment of platelet-mediated thrombotic conditions.

Platelet aggregation inhibitors (antithrombotic agents) decrease the formation or the action of chemical signals that promote platelet aggregation. Currently available antithrombotic agents include aspirin, ticlopidine, and dipyridamole. These agents have proven beneficial in the prevention and treatment of occlusive cardiovascular diseases, including myocardial infarction, cerebral ischemia, angina. Antithrombotic therapy has also been used in the maintenance of vascular grafts.

Myocardial infarction is the development of necrosis of the myocardium (the middle muscular layer of the heart wall) due to a critical imbalance between oxygen and myocardial demand. The most common cause of acute myocardium infarction is narrowing of the epicardial blood vessels due to atheromatous plaques. Plaque rupture with subsequent exposure of basement membrane results in platelet aggregation and thrombus formation, which can result in partial or complete occlusion of the vessel and subsequent myocardial ischemia.

In cerebral ischemia, inadequate blood flow results from an occlusion in a blood vessel or hemorrhaging. In the latter case, excessive bleeding in one area of the brain deprives another area of blood. If the damage occurs in a singular small area, "transient" or "focused" cerebral ischemia results. When a major artery is blocked (carotid artery) global or diffused ischemia results. The primary medical strategy for secondary prevention of stroke is antiplatelet therapy. Aspirin is currently employed for reducing the risk of recurrent transient ischemic attacks or stroke in men who have transient ischemia of the brain due to fibrin emboli.

Each year, thousands of patients suffer a decline in blood flow to one or more limbs. Without sufficient blood flow, and, unless blood flow can be restored in time, the limb must be amputated. In some cases, grafts from the patient's veins can be used to form new arteries. However, in cases
5 where the quality of the veins is poor, polymeric vascular grafts are typically used. The polymeric grafts are inherently thrombogenic as the blood constituents passing through the grafts become activated and tend to form clots. Efforts to line the grafts with endothelial cells can reduce blood clotting, but better results are obtained when antithrombotic therapy is
10 employed.

Angina pectoris is a characteristic chest pain caused by inadequate blood flow through the blood vessels of the myocardium. The imbalance between oxygen delivery and utilization may result from a spasm of the vascular smooth muscle or from obstruction of blood vessels caused by
15 atherosclerotic lesions. Three classes of drugs have been shown to be effective in treating angina: nitrates, beta-blockers and calcium channel blockers. Currently, the antithrombotics dipyridamole and aspirin are employed to prophylactically treat angina pectoris.

Ecto-apyrases, such as CD39, offer a number of advantages
20 over several of the standard antithrombotics. For example, aspirin treatment controls the prothrombotic action of thromboxane; however, aspirin also prevents formation of antithrombotic prostacyclin, which limits aspirin's efficacy. Another antithrombotic, endothelium-derived relaxing factor (nitric oxide; "EDRF/NO"), is inhibited *in vitro* and *in vivo* by hemoglobin after its
25 rapid diffusion into erythrocytes. In contrast, CD39 is aspirin-insensitive and completely inhibits platelet reactivity even when eicosanoid and EDRF/NO production are blocked.

CD39's ATPDase activity also implicates CD39 in the modulation of neurotransmission. ATP is a major purinergic neurotransmitter
30 that is often co-released into the synaptic cleft with several neurotransmitters.

Responses to ATP are mediated by specific plasma membrane receptors, called P2 purinergic receptors (Dubyak, G. R. and El-Motassim, C. *Am J. Physiol.* 34:C577-C606 (1993)). The distribution of CD39 in the rat brain indicates that CD39 plays a role in terminating P2 purinergic neurotransmission (Wang, T. F. and Guidotti, G., *Brain Res.* 790:318 (1998)). Furthermore, a decrease in ecto-apyrase activity is believed to lead to an accumulation of the excitatory neurotransmitter, extracellular ATP, as well as a deficiency of the endogenous anticonvulsant extracellular adenosine.

The chromosomal localization of CD39 provides additional support for a role in modulation of neurotransmission. More specifically, CD39 has been mapped to chromosome 10q 23.1-24.1 (Maliszewski, C. R., et al., *J. Immunol.* 153:3574 (1994)), and this site overlaps with the susceptibility locus for human partial epilepsy with audiogenic symptoms (Ottman, R. et al., *Nature Genet.* 10:56 (1995)). This co-localization of the CD39 gene and the susceptibility locus has led to the hypothesis that decrease in ecto-apyrase activity in the brain is the primary cause of partial epilepsy (Wang T-F., et al., *Mol. Brain Res.* 47:295 (1997)).

A screen for human cDNAs that hybridize to cosmids from the human chromosome 9q34 region lead to the identification of a transcript with high homology to a chicken muscle ecto-ATPase (60% identity) and the ecto-apyrase CD39 (41% amino acid identity) (Chadwick, B. P., *Mamm. Genome* 8:668 (1997)). This gene, designated "CD39-like-1 gene" (CD39L1), has a higher degree of homology to CD39 than does chicken muscle ecto-ATPase. The biological activity of this protein has not been tested but on the basis of the high amino acid homology, CD39L1 is believed to be a new member of the ecto-ATPase family. Recently, a mouse gene with homology to NTPases was cloned and sequenced (Acc. No. AF006482) by Chadwick et al. (*Mamm. Gen.* 9:162-164 (1998)).

4. SUMMARY OF THE INVENTION

The invention is based on polynucleotides isolated from cDNA libraries prepared from human fetal liver-spleen and macrophages. The compositions of the present invention include novel isolated polypeptides with apyrase and/or NDPase activity, in particular, novel human CD39-like polypeptides, and active variants thereof, isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The isolated polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. One polynucleotide according to the invention encodes a novel CD39-like protein having the amino acid sequence shown in Figure 2 (SEQ ID NO. 3), which has been designated CD39-L4. In another embodiment, a polynucleotide according to the invention encodes a novel CD39-like protein having the full length or mature amino acid sequence set forth in SEQ ID NO. 25, which has been designated CD39-L66, and is an isoform of CD39-L4. The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequence of SEQ ID NO. 2 or 24. The polynucleotides of the invention also include polynucleotides that encode polypeptides with a biological activity of the polypeptide of SEQ ID NO. 3 (including apyrase or NDPase activity) such as (a) the nucleotide sequence of SEQ ID NO. 2 or 24, or (b) a nucleotide sequence encoding the full length or mature amino acid sequence of SEQ ID NO. 3 or 25; (c) a polynucleotide

which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide that hybridizes under stringent conditions to (a) or (b); (e) or a polynucleotide that encodes a polypeptide comprising at least one CD39-like domain, e.g. catalytic domain.

5 The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above.

 The invention also provides a polynucleotide including a nucleotide sequence that is substantially equivalent to this polynucleotide. Polynucleotides according to the invention can have at least about 80%,
10 more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide of SEQ ID NO. 2 or 24, and specifically include a human polynucleotide which has at least 80% sequence identity to a polynucleotide of SEQ ID NO. 2 or 24; or a polynucleotide which
15 has at least 90% sequence identity to a polynucleotide of SEQ ID NO. 2 or 24. Similarly, polypeptides of the invention include polypeptides having apyrase or NDPase activity and at least about 80%, 90% or 95% sequence identity to SEQ ID NO. 3 or 25.

 A further aspect of the invention is the development of novel CD39-L4 variants which preferably have improved ADPase or NDPase
20 activity compared to wild type CD39-L4 (SEQ ID NO: 5). This aspect of the invention includes polypeptides comprising at least one amino acid substitution selected from the group consisting of: D168-T, S170-Q and L175-F, wherein said substitution(s) result in increased ADPase activity of the polypeptide. One preferred embodiment is the polypeptide having the
25 amino acid sequence set forth in SEQ ID NO: 7 (encoded by the nucleotide sequence of SEQ ID NO. 6), which is a variant CD39-L4 containing all three substitutions that has been designated ACRIII. A plasmid containing this DNA was deposited with the American Type Culture Collection (ATCC),
30 10801 University Avenue, Manassas, Virginia, on July 13, 1999 under the terms of the Budapest Treaty (ATCC accession number PTA 346).

Alternatively, instead of making the specific D168→T, S170→Q and/or L175→F substitution(s), substitution of amino acids with similar properties is contemplated. Additional conservative substitutions at amino acid positions other than D168, S170 and/or L175 are further contemplated. For example,
5 all of the corresponding amino acids from CD39 could be substituted for amino acids 167-181 of CD39-L66 or CD39-L4.

Polynucleotides encoding these polypeptides, vectors and host cells comprising such polynucleotides, methods of using such host cells to produce polypeptides, and other therapeutic products comprising the
10 polypeptides (including fusion proteins in which the CD39-like polypeptide is fused to a heterologous peptide or polypeptide, such as an immunoglobulin constant region, or derivatives in which the CD39-like polypeptide is modified by water soluble polymers to increase its half-life) are also comprehended by the invention, as are methods of treating a subject suffering from a disorder
15 relating to thrombosis, coagulation or platelet aggregation by administering such therapeutic products.

Gene therapy techniques are also provided to modulate disease states associated with CD39-L4 expression and/or biological activity. Delivery of a functional CD39-L4 gene to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors
20 (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments).

The invention also relates to methods for producing polypeptides of the invention comprising growing a culture of cells of the
25 invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the protein from the cells or the culture medium. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Protein compositions of the present invention, including
30 therapeutic compositions, comprise polypeptides of the invention and

optionally an acceptable carrier, such as a hydrophilic (e.g., pharmaceutically acceptable) carrier.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, because the expression of CD39-L4 mRNA is largely restricted to macrophages, polynucleotides of the invention can be used as hybridization probes to detect the presence of macrophage mRNA in a sample using, e.g., *in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath, et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook, J., et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polypeptides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can

be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody which specifically binds the polypeptide. The polypeptides of the invention having ATPDase activity are also useful for inhibiting platelet aggregation and can therefore be employed in the prophylaxis or treatment of pathological conditions caused by the inflammatory response. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Another aspect of the invention is an antibody that specifically binds the polypeptide of the invention. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments thereof and humanized forms or fully human forms, such as those produced in transgenic animals. The invention further provides a hybridoma that produces an antibody according to the invention and anti-idiotypic antibodies.

Antibodies of the invention are useful for detection and/or purification of the polypeptides of the invention.

Methods are also provided for preventing, treating or ameliorating a medical condition, including thrombotic diseases, which comprises administering to a mammalian subject, including but not limited to humans, a therapeutically effective amount of a composition comprising a polypeptide of the invention or a therapeutically effective amount of a composition comprising a binding partner of (e.g., antibody specifically reactive for) CD39-like polypeptides of the invention. The mechanics of the particular condition or pathology will dictate whether the polypeptides of the invention or binding partners (or inhibitors) of these would be beneficial to the individual in need of treatment.

The invention also provides a method of inhibiting platelet function comprising administering a CD39-L4 polypeptide of the invention to a medium comprising platelets. According to this method, polypeptides of the invention can be administered to produce an *in vitro* or *in vivo* inhibition of platelet function. A polypeptide of the invention can be administered *in vivo* as antithrombotic agent alone or as an adjunct to other therapies.

Also provided are methods of hydrolyzing nucleotidediphosphates comprising administering CD39-L4 polypeptides of the invention to a medium comprising nucleotidediphosphates. According to this method, polypeptides of the invention can be administered to produce an *in vitro* or *in vivo* hydrolysis of nucleotidediphosphates. A polypeptide of the invention can be administered *in vivo* alone or as an adjunct to other therapies.

The invention further provides methods for manufacturing medicaments useful in the above described methods relating to platelet aggregation and thrombosis.

The invention also provides methods for detecting or quantitating the presence of the polynucleotides or polypeptides of the invention in a tissue or fluid sample, and corresponding kits that comprise suitable polynucleotide probes or antibodies, together with an optional quantitative standard. Such methods and kits can be utilized as part of prognostic and diagnostic evaluation of patients and for the identification of subjects exhibiting a predisposition to platelet mediated conditions.

The invention also provides methods for the identification of compounds that modulate (i.e. increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds and other substances that interact with (e.g., bind to) the polypeptides of the invention, and assays for identifying compounds and other substances that enhance or inhibit the activity of the polypeptides of the invention, such

assays comprising the step of measuring activity of such polypeptides in the presence and absence of the test compound.

5. BRIEF DESCRIPTION OF THE DRAWINGS

5 FIG. 1 shows polynucleotide sequences according to the invention. SEQ ID NO:1 was obtained from the b2HFLS20W cDNA library using standard PCR, sequencing by hybridization signature analysis, and single pass gel sequencing technology. A- adenosine; C-cytosine; G-guanosine; T-thymine. Ambiguous positions are designated as follows: R indicates A or G; M indicates A or C; W indicates A or T; Y indicates C or T; S indicates C or G; K indicates G or T; V indicates A or C or G; H indicates A or C or T; D indicates A or G or T; B indicates C or G or T; and N indicates any of the four bases.

10 SEQ ID NO:2 is an extended version of SEQ ID NO:1 which was obtained as described in Example 2.

15 FIG. 2 shows an amino acid sequence corresponding to the polynucleotide sequence of SEQ ID NO:2. This sequence is designated as SEQ ID NO:3. The open reading frame encoding SEQ ID NO:3 begins at nucleotide 246 (numbered from the 5' end) of SEQ ID NO:2. A- Alanine; R- Arginine; N- Asparagine; D- Aspartic Acid; C- Cysteine; E- Glutamic Acid; Q- Glutamine; G- Glycine; H- Histidine; I- Isoleucine; L- Leucine; K- Lysine; M- Methionine; F- Phenylalanine; P- Proline; S- Serine; T- Threonine; W- Tryptophan; Y- Tyrosine; V- Valine; X - any of the twenty amino acids.

20 FIG. 3 shows the amino acid sequence alignment of SEQ ID NO:3 (identified as "246 prot") and human CD39 ("CD39Human.seq"). The amino acid residues are designated as for FIG. 2. The alignment was generated using the Jotun Hein method with the PAM250 residue weight table. Gaps are indicated by dashes; residues that are identical between the two sequences (within 1 distance unit) are boxed.

FIG. 4 shows the amino acid sequence alignment of SEQ ID NO:3 (identified as "264 prot") and murine NTPase ("mur ntpase"). The amino acid residues are designated as for FIG. 2. The alignment was generated as discussed for FIG. 3. Gaps are indicated by dashes; residues that are identical between the two sequences (within 1 distance unit) are boxed.

FIG. 5 shows the apyrase conserved regions (ACR) in CD39-L4 in bold. ACR I starts at Phe 53, ACR II starts at Pro 124 and ACR III starts at Met 167. The boxed sections highlight the amino acid substitutions that were made in the wild type CD39-L4 amino acid sequence to form mutants designated ACRI, ACRII and ACRIII.

FIG. 6 (SEQ ID NOS: 6 and 7) shows the nucleotide and corresponding amino acid sequences of a preferred ACRIII mutant containing the following substitutions in the wild type CD39-L4 amino acid sequence: D168→T, S170→Q and L175→F.

FIG. 7 shows the ADPase activity of CD39-L4 variants ACRI, ACRII and ACRIII in comparison to wild type CD39-L4: (1) CD39-L4 ACR I mutant; (2) CD39-L4 ACR II mutant; (3) CD39-L4 ACR III mutant; (4) CD39-L4 wild type; (5) sCD39; and (6) pSecTag2 vector (Invitrogen).

FIG. 8 shows the amino acid sequence alignment of SEQ ID NO. 3, SEQ ID NO. 25 (previously identified as SEQ ID NO. 5 in Fig. 5 of USSN 09/122,449) and human CD39 ("CD39Human.seq"). The alignment was generated using the Jotun Hein method with the PAM250 residue weight table. Gaps are indicated by dashes; residues that are identical between the two sequences (within 1 distance unit) are boxed.

FIG. 9 shows the amino acid sequence alignment of SEQ ID NO. 3, SEQ ID NO. 25 (previously identified as SEQ ID NO. 5 in Fig. 6 of USSN 09/122,449) and the murine NTPase ("mur ntpase"). The alignment was generated as discussed for FIG. 8. Gaps are indicated by dashes;

residues that are identical between the two sequences (within 1 distance unit) are boxed.

6. DETAILED DESCRIPTION

5

6.1 Definitions

The term "nucleotide sequence" refers to a heteropolymer of nucleotides or the sequence of these nucleotides. The terms "nucleic acid" and "polynucleotide" are also used interchangeably herein to refer to a heteropolymer of nucleotides. Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

15

An "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" is a stretch of polypeptide nucleotide residues which is long enough to use in polymerase chain reaction (PCR) or various hybridization procedures to identify or amplify identical or related parts of mRNA or DNA molecules.

20

"Oligonucleotides" or "nucleic acid probes" are prepared based on the cDNA sequence provided in the present invention. Oligonucleotides comprise portions of the DNA sequence having at least about 15 nucleotides and usually at least about 20 nucleotides. Nucleic acid probes comprise portions of the sequence having fewer nucleotides than about 6 kb, usually fewer than about 1 kb. After appropriate testing to eliminate false positives, these probes may be used to determine whether mRNAs are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh, P.S., et al (1992 PCR Methods Appl 1:241-250).

25

The term "probes" includes naturally occurring or recombinant single- or double-stranded nucleic acids or chemically synthesized nucleic acids. They may be labeled by nick translation, Klenow fill-in reaction, PCR or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J., et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M., et al (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both incorporated herein by reference.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. An exemplary set of conditions include a temperature of 60-70 °C, (preferably about 65 °C) and a salt concentration of 0.70 M to 0.80 M (preferably about 0.75M). Further exemplary conditions include, hybridizing conditions that (1) employ low ionic strength and high temperature for washing, for example, 0.015 M NaCl/0.0015 M sodium citrate/0.1% SDS at 50°C.; (2) employ during hybridization a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (0.75 M NaCl, 0.075 M Sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 g/ml), 0.1% SDS, and 10% dextran sulfate at 42°C, with washes at 42°C in 0.2 x SSC and 0.1% SDS.

The term "recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., E. coli, will be free of glycosylation

modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

The term "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The term "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF. As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the

expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotide molecules which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems known in the art.

The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

"Active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide.

"Naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

"Derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

"Recombinant variant" refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and

substitutions, created using recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, such as cellular trafficking, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, such as the replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, or a threonine with a serine, i.e., conservative amino acid replacements. "Insertions" or "deletions" are typically in the range of about 1 to 5 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a mutant sequence varies from one of those listed herein by no more than about 20% (i.e., the number of substitutions, additions, and/or deletions in a mutant sequence, as compared to the corresponding listed sequence, divided by the total number of residues in the mutant sequence is about 0.2 or less). Such a mutant sequence is said to have 80% sequence identity to the listed sequence. In one embodiment, a mutant sequence of the invention varies from a listed sequence by no more than 10% (90% sequence identity), in a variation of this embodiment, by no more than 5% (95% sequence identity), and in a further variation of this embodiment, by no more than 2% (98% sequence identity). Mutant amino acid sequences according to the invention

generally have at least 95% sequence identity with a listed amino acid sequence, whereas mutant nucleotide sequence of the invention can have lower percent sequence identities. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence should be disregarded.

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, often at least about 7 amino acids, typically at least about 9 to 13 amino acids, and, in various embodiments, at least about 17 or more amino acids. To be active, any polypeptide must have sufficient length to display biologic and/or immunologic activity.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polypeptide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

"Activated" cells as used in this application are those which are engaged in extracellular or intracellular membrane trafficking, including the

export of neurosecretory or enzymatic molecules as part of a normal or disease process.

The term "purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, e.g., polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99.8% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration.

The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed.

The term "intermediate fragment" means a nucleic acid between 5 and 1000 bases in length, and preferably between 10 and 40 bp in length.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

6.2 Hybridization Conditions

Suitable hybridization conditions may be routinely determined by optimization procedures or pilot studies. Such procedures and studies are routinely conducted by those skilled in the art to establish protocols for use in a laboratory. See e.g., Ausubel, et al., Current Protocols in Molecular Biology, Vol. 1-2, John Wiley & Sons (1989); Sambrook, et al., Molecular Cloning A Laboratory Manual, 2nd Ed., Vols. 1-3, Cold Springs Harbor Press (1989); and Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Cold Spring Harbor, New York (1982), all of which are incorporated by reference herein. For example, conditions such as temperature, concentration of components, hybridization and washing times, buffer components, and their pH and ionic strength may be varied.

6.3 Nucleic Acids of the Invention

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic variations thereof. Allelic variations can be routinely determined by comparing the sequence provided in SEQ ID NOs: 1-2 or 24, a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to SEQ ID NO:1-2 or 24 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (i.e., sequence both strands).

The present invention further provides recombinant constructs comprising a nucleic acid having the sequence of any one of SEQ ID NOs:

1-2 or 24, the mature protein coding sequence or a fragment thereof. The recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having the sequence of any one of SEQ ID NOs 1-2 or 24 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs and UMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF or UMF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be

derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (e.g., temperature shift or

chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

5 Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequences that hybridize under stringent conditions to a fragment of the DNA sequences in Figure 1, which fragment is greater than about 10 bp, preferably 20-50 bp, and even greater than 100 bp, including 200 bp or greater, 300 bp or greater, 400 bp or greater, and 500 bp or greater .

10 In accordance with the invention, polynucleotide sequences which encode the novel nucleic acids, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells.

15 The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the
20 mutation. The amino acid sequence variants of the nucleic acids are preferably constructed by mutating the polynucleotide to give an amino acid sequence that does not occur in nature. These amino acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such
25 locations will typically be modified in series, e.g., by substituting first with conservative choices (e.g., hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site.

Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells.

In a preferred method, polynucleotides encoding the novel nucleic acids are changed via site-directed mutagenesis. This method uses oligonucleotide sequences that encode the polynucleotide sequence of the desired amino acid variant, as well as a sufficient adjacent nucleotide on both sides of the changed amino acid to form a stable duplex on either side of the site being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., DNA 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, Nucleic Acids Res. 10:6487-6500 (1982).

PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook, et al., *supra*, and *Current Protocols in Molecular Biology*, Ausubel, et al.

Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Furthermore, knowledge of the DNA sequence provided by the present invention allows for the modification of cells to permit, or increase, expression of endogenous CD39-like polypeptides. Cells can be modified (e.g., by homologous recombination) to provide increased CD39-like expression by replacing, in whole or in part, the naturally occurring CD39-like promoter with all or part of a heterologous promoter so that the cells express CD39-like polypeptides at a higher level. The heterologous promoter is inserted in such a manner that it is operatively linked to CD39-like encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., *ada*, *dhfr*, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the CD39-like coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the CD39-like coding sequences in the cells.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express functional CD39-like polypeptides or that express a variant of a CD39-like polypeptide. Such animals are useful as models for studying the *in vivo* activities of CD39-like polypeptides as well as for studying modulators of CD39-like polypeptides.

6.4 Identification of Polymorphisms

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labelled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed.

Alternatively, a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

6.5 Hosts

The present invention further provides host cells containing SEQ ID NOs:1-2 or 24 of the present invention, wherein the nucleic acid has been introduced into the host cell using known transformation, transfection or infection methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran mediated transfection, or electroporation (Davis, L., et al., Basic Methods in Molecular Biology (1986)).

The host cells containing one of SEQ ID NOs:1-2 or 24 of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as E. coli and B. subtilis. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, insect cells or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981), and other cell lines capable of expressing a compatible
5 vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional
10 termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by
15 one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any
20 convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

6.6 Peptides

The present invention further provides isolated polypeptides
25 encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. Fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites. For example, fragments of the protein may be fused through "linker" sequences to
30 the Fc portion of an immunoglobulin. For a bivalent form of the protein, such

a fusion could be to the Fc portion of an IgG molecule. Other immunoglobulin isotypes may also be used to generate such fusions. For example, a protein-IgM fusion would generate a decavalent form of the protein of the invention. Analogs of the polypeptides of the invention can be fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs which encode proteins.

The invention also provides both full length and mature forms (for example, without a signal sequence or precursor sequence) of CD39-like polypeptides. The full length form of such proteins is identified in the sequence listing by translation of the nucleotide sequence of each disclosed clone. The mature form of such protein may be obtained by expression of the full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full length form.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to,

immunochemistry, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, e.g., Scopes, Protein Purification: Principles and Practice, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: A Laboratory Manual; Ausubel, et al., Current Protocols in Molecular Biology.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The purified polypeptides are used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for example, small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the binding molecules may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to the tumor or other cell by the specificity of the binding molecule for SEQ ID NOs:3-4.

6.7 Gene Therapy

Mutations in the CD39-like gene that result in loss of normal function of the CD39-like gene product underlie CD39-related human disease states. The invention comprehends gene therapy to restore CD39-like activity that would thus be indicated in treating those disease states. Delivery of a functional CD39-like gene to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, *Nature*, supplement to vol. 392, no 6679, pp. 25-30 (1998). For additional reviews of gene therapy technology, see Friedmann, *Science*, 244: 1275-1281 (1989); Verma, *Scientific American*: 68-84 (1990); and Miller, *Nature*, 357: 455-460 (1992). Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of CD39-like polypeptides will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of CD39-like polypeptides.

6.8 Deposit of clone

A plasmid containing DNA encoding the ACR III mutant was deposited with the American Type Culture Collection (ATCC), 10801 University Avenue, Manassas, Virginia, on July 13, 1999 under the terms of the Budapest Treaty (ATCC accession no. PTA 346).

6.9 Antibodies

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A.M., *Monoclonal Antibodies Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth,

et al., J. Immunol. 35:1-21 (1990); Kohler and Milstein, Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor, et al., Immunology Today 4:72 (1983); Cole, et al., in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985), pp. 77-96). In addition, techniques for preparing chimeric and humanized antibodies (including polypeptides containing CDR and/or antigen-binding sequences of antibodies) are well known in the art.

Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with a peptide or polypeptide of the invention. Methods for immunization are well known in the art. Such methods include subcutaneous or intraperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling the antigen with a heterologous protein (such as globulin or -galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz, et al., Exp. Cell Research. 175:109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antiserum is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.) fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well-known in the art, for example, see (Sternberger, L.A. et al., J. Histochem. Cytochem. 18:315 (1970); Bayer, E.A. et al., Meth. Enzym. 62:308 (1979); Engval, E. et al., Immunol. 109:129 (1972); Goding, J.W. J. Immunol. Meth. 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays to identify cells or tissues in which a fragment of the polypeptide of interest is expressed. The antibodies may also be used directly in therapies or other diagnostics.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and polyacrylamide and latex beads.

Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D.M. et al., "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W.D. et al., Meth. Enzym. 34 Academic Press, N.Y. (1974)). The
5 immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and *in situ* assays as well as for immuno-affinity purification of the proteins of the present invention.

6.10 Computer Readable Sequences

10 In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium,
15 and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded
20 thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the
25 nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a
30 variety of data processor programs and formats can be used to store the

nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g. text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide sequence of SEQ ID NOs:1-2 or 24, a representative fragment thereof, or a nucleotide sequence at least 99.9% identical to SEQ ID NOs:1-2 or 24 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Software which implements the BLAST (Altschul, et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag, et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system may be used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems is suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used
5 herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs
10 which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a
15 variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software
20 packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is,
25 the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene
30 expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

6.11 Expression Modulating Sequences

EMF sequences can be identified within a genome by their proximity to the ORFs. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken 5' from any ORF will modulate the expression of an operably linked 3' ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to the fragments of a genome which are between two ORF(S) herein described. Alternatively, EMFs can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site 5' to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below. A sequence which is suspected of being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host

using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

5 **6.12 Triplex Helix Formation**

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides
10 suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee, et al., Nucl. Acids Res. 6:3073 (1979); Cooney, et al., Science 15241:456 (1988); and Dervan, et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem.
15 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)).

Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been
20 demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

25 **6.13 Diagnostic Assays and Kits**

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with
30 one or more of the antibodies or one or more of nucleic acid probes of the

present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present

invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

5 In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allow one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be
10 added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody
15 or probe.

Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will
20 readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

6.14 Screening Assays

25 Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs from a nucleic acid with a sequence of one of SEQ ID NOs:1-2 or 24, or to a nucleic acid with a sequence of one of SEQ ID NOs:1-2 or 24.

In detail, said method comprises the steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or nucleic acid of the invention; and (b) determining whether the agent binds to said protein or said nucleic acid.

5 The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

10 For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

15 Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby, et al., Application of
20 Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak, et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

25 In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same
30 EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee, et al., Nucl. Acids Res. 6:3073 (1979); Cooney, et al., Science 241:456 (1988); and Dervan, et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent, in the control of bacterial infection by modulating the activity of the protein encoded by the ORF. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

6.15 Use of Nucleic Acids as Probes

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from the nucleotide sequence of the

SEQ ID NOs:1-2 or 24. Because the corresponding gene is expressed in only one out of 18 tissues tested, namely macrophages, a hybridization probe derived from SEQ ID NOs:1-2 or 24 can be used as an indicator of the presence of macrophage RNA in a sample. Any suitable hybridization
5 technique can be employed, such as, for example, *in situ* hybridization.

PCR as described US Patent Nos 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a
10 discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are
15 commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides.

The nucleotide sequences may be used to construct
20 hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include *in situ* hybridization, linkage analysis against known chromosomal markers,
25 hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent *in situ* hybridization of chromosome spreads has been described, among other places, in Verma, et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY. Fluorescent *in situ*
30 hybridization of chromosomal preparations and other physical chromosome

mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

The nucleotide sequence may be used to produce purified polypeptides using well known methods of recombinant DNA technology. Among the many publications that teach methods for the expression of genes after they have been isolated is Goeddel, (1990) Gene Expression Technology, Methods and Enzymology, Vol 185, Academic Press, San Diego. Polypeptides may be expressed in a variety of host cells, either prokaryotic or eukaryotic. Host cells may be from the same species from which a particular polypeptide nucleotide sequence was isolated or from a different species. Advantages of producing polypeptides by recombinant DNA technology include obtaining adequate amounts of the protein for purification and the availability of simplified purification procedures.

Each sequence so obtained was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 Sequence Analysis System. In this algorithm, Pattern Specification Language (developed by TRW Inc., Los Angeles, CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of

homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT™ 670 Sequence Analysis System in a way similar to that used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

Alternatively, BLAST, which stands for Basic Local Alignment Search Tool, is used to search for local sequence alignments (Altschul, S.F., (1993) J Mol Evol 36:290-300; Altschul, S.F., et al (1990) J Mol Biol 215:403-10). BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. Whereas it is ideal for matches which do not contain gaps, it is inappropriate for performing motif-style searching. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire

database search. Any database sequence whose match satisfies E is reported in the program output.

In addition, BLAST analysis was used to search for related molecules within the libraries of the LIFESEQ™ database. This process, an "electronic northern" analysis is analogous to northern blot analysis in that it uses one cellubrevin sequence at a time to search for identical or homologous molecules at a set stringency. The stringency of the electronic northern is based on "product score". The product score is defined as (% nucleotide or amino acid [between the query and reference sequences] in Blast multiplied by the % maximum possible BLAST score [based on the lengths of query and reference sequences]) divided by 100. At a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous or related molecules can be identified by selecting those which show product scores between approximately 15 and 30.

6.16 SEQ ID NOs:1- 8 and 23 - 24

Referring to Figure 1, SEQ ID NO:1 is the nucleotide sequence of an expressed sequence tag corresponding to a polynucleotide isolated from a cDNA library of human fetal liver-spleen. SEQ ID NO:2 is an extended version of SEQ ID NO:1 obtained as described in Example 2, and the encoded polypeptide in SEQ ID NO: 3 is referred to herein as CD39-L4. SEQ ID NO:2 encodes a polypeptide having the amino acid sequence of SEQ ID NO:3 (shown in Figure 2). The open reading frame corresponding to SEQ ID NO:3 starts at nucleotide 246, as numbered from the 5' end of SEQ ID NO:2. This open reading frame encodes a polypeptide 428 amino acids in length. The estimated molecular weight of the unglycosylated polypeptide is approximately 47.52 kDa.

Protein database searches with the BLAST algorithm indicate that SEQ ID NO:3 is homologous to the CD39 family. Figure 3 shows the amino acid sequence alignment between SEQ ID NO:3 (identified as "246

prot") and human CD39 ("CD39Human.seq"), indicating that the two sequences share 30% amino acid sequence identity. Moreover, a higher degree of homology between the apyrase conserved regions (Kaczmarek et al., J. Biol. Chem. 271:33116-33122 (1996) is observed. In particular, an almost perfect match to a putative ATP-binding region was found from amino acids 54-58, DAGST (DAGSS in CD39). In addition, the DLGGASTQ motif (DLGGASTQ in CD39), which is very well conserved among ATPDases, is found from amino acids 199-206 in SEQ ID NO:3. Other regions conserved in apyrases were found from amino acids 129-134, ATAGLR (ATAGMR in CD39) and from amino acids 169-173, GSDEG (GQEEG in CD39).

SEQ ID NO:3 differs from CD39 in that SEQ ID NO:3 contains a hydrophobic stretch of 22 amino acids at its amino terminus, which is indicative of a leader peptide. SEQ ID NO:3 also lacks the transmembrane domain found at the carboxyl terminus of CD39. These features indicate that SEQ ID NO:3 is a soluble ATPDase.

SEQ ID NO:3 shares an even higher degree of homology (86% identity) with a murine NTPase, as shown in the amino acid sequence alignment presented in Fig. 4 (SEQ ID NO:3 is identified as "246 prot," and mouse CD39 as "mur ntpase").

The message encoding SEQ ID NO:3 is tightly regulated in a tissue-specific manner. An expression study using a semiquantitative PCR/Southern blot approach revealed a significant level of expression in macrophage. In contrast, human CD39 is expressed in tissues such as placenta, lung, skeletal muscle, kidney, and heart.

SEQ ID NO: 4 is the polynucleotide sequence for CD39-L4. SEQ ID NO: 5 is the corresponding amino acid sequence.

SEQ ID NO: 6 is the polynucleotide sequence for a CD39-L4 variant designated ACRIII, wherein the following amino acid substitutions have been made: D168→T, S170→Q and L175→F; SEQ ID NO: 7 is the corresponding amino acid sequence.

SEQ ID NO: 8 is the genomic sequence for the human CD39-L4 gene; exons appear at nucleotides 1-288 (exon 1), 1281-1580 (exon 2), 1820-1855 (exon 3) 2467-2555 (exon 4), 2863-2942 (exon 5), 3889-3950 (exon 6), 4894-4995 (exon 7), 5847-5987 (exon 8), 6966-7138 (exon 9) and 8556-9365 (exon 10).

SEQ ID NO: 24 in the polynucleotide sequence for a CD39-L4 splice variant that creates an isoform designated CD39-L66. SEQ ID NO: 25 is the corresponding amino acid sequence.

6.17 Uses of Novel CD39-Like Polypeptides and Antibodies

Polypeptides of the invention having ATPDase, including NDPase, activity are useful for inhibiting platelet function and can therefore be employed in the prophylaxis or treatment of pathological conditions caused by or involving thrombosis or excessive coagulation or excessive platelet aggregation, such as myocardial infarction, cerebral ischemia, angina, and the like. Polypeptides of the invention can also be used in the maintenance of vascular grafts. Platelet function can be measured by any of a number of standard assays, such as, for example, the platelet aggregation assay described in Example 5.

Such pathological conditions include conditions caused by or involving arterial thrombosis, such as coronary artery thrombosis and resulting myocardial infarction, cerebral artery thrombosis or intracardiac thrombosis (due to, e.g., atrial fibrillation) and resulting stroke, and other peripheral arterial thrombosis and occlusion; conditions associated with venous thrombosis, such as deep venous thrombosis and pulmonary embolism; conditions associated with exposure of the patient's blood to a foreign or injured tissue surface, including diseased heart valves, mechanical heart valves, vascular grafts, and other extracorporeal devices such as intravascular cannulas, vascular access shunts in hemodialysis patients, hemodialysis machines and cardiopulmonary bypass machines; and

conditions associated with coagulopathies, such as hypercoagulability and disseminated intravascular coagulopathy. Co-administration of other agents suitable for treating the pathological condition, e.g., other anti-coagulation agents, is also contemplated.

5 In particular, variants like the ACRIII mutant described herein are expected to be superior therapeutics for treating such pathological conditions because (1) ACRIII exhibits six-fold greater activity compared to wild type CD39-L4, and (2) ACRIII, like CD39-L4, is uniquely specific for ADP and does not hydrolyze ATP. Thus, adverse side effects from hydrolysis of
10 circulating ATP are avoided.

 For instance, ATP is known to act as an extracellular signal in many tissues. In the heart, extracellular ATP modulates ionic processes and contractile function (for review see Burnstock, G., *Neuropharmacology* 36:1127). Recently, it has been shown that extracellular ATP markedly
15 inhibits glucose transport in rat cardiomyocytes (Fisher, Y. et al., *J. Biol. Chem.* 274:755-761. Another source of extracellular ATP is that released from parenchymal cells under hypoxic or ischemic conditions (Skobel, E., and Kammermeier, H. *Biochim. Biophys. Acta* 1362:128-134). ATP is also involved in the modulation of anti-IgE-induced release of histamine from
20 human lung mast cells (Schulman, E. S., et al., *Am. J. Respir. Cell Mol. Biol.* 20:520-537).

 Furthermore, the ability of CD39-L4 to hydrolyze NDPs other than ADP has implications outside the circulatory system. For instance, it has been reported that UDP is the most potent agonist for the human P2Y₆
25 receptor. Communi, et al., *Bioch Bioph Res Com* 222:303-308 (1996). This receptor is expressed in several tissues including infiltrating T cells present in inflammatory bowel disease. Somers, et al., *Lab Invest* 78:1375-1383 (1998). In this microenvironment, a molecule with the enzymatic properties of CD39-L4 could influence T cell responses by modifying the extracellular half-
30 life of UDP. Another role for CD39-L4 has been suggested by the report that

mouse CD39-L4 maps closely to a locus associated with audigenic brain seizures in mice. See Chadwick, et al., Genomics 50:357-367 (1998); Seyfried, et al., Genetics 99:117-126 (1981). This locus, known as Asp-1, is thought to be linked or to correspond to a factor that influences Ca^{2+} -ATPase activity. Neumann, et al., Behav. Genetics 20:307-323 (1990).

Additionally, the polypeptides of the invention can be used as molecular weight markers, and as a food supplement. A polypeptide consisting of SEQ ID NO:3, for example, has a molecular mass of approximately 47.52 kD in its unglycosylated form. Protein food supplements are well known and the formulation of suitable food supplements including polypeptides of the invention is within the level of skill in the food preparation art.

The polypeptides of the invention are also useful for making antibody substances that are specifically immunoreactive with CD39-like proteins. Antibodies and portions thereof (e.g., Fab fragments) which bind to the polypeptides of the invention can be used to identify the presence of such polypeptides in a sample. For example, the level of the native protein corresponding to SEQ ID NO:3 in a blood sample can be determined as an indication of vascular condition. Such determinations are carried out using any suitable immunoassay format, and any polypeptide of the invention that is specifically bound by the antibody can be employed as a positive control.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. For treatment of vascular disease, polypeptides according to the invention are generally administered intravenously. In vivo murine studies with soluble human CD39 have shown that mice injected intravenously with 50 mg recombinant soluble human CD39 in 100 ml sterile saline had biologically active CD39 in their sera for an extended period of time, with an elimination half-life of almost 2

days (Gayle, R.B., et al., J. Clinical Invest. 101:1851-1859 (1998)). Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models.

Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

6.18 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may also contain (in addition to protein and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or compliment its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein of the invention, or to minimize side effects. Conversely, protein of the present invention may be included in formulations of the particular cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor,

or anti-inflammatory agent to minimize side effects of the cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent. A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or
5 other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective
10 dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, e.g., treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers
15 to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a
20 therapeutically effective amount of protein of the present invention is administered to a mammal having a condition to be treated. Protein of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors.
25 When co-administered with one or more cytokines, lymphokines or other hematopoietic factors, protein of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the
30 appropriate sequence of administering protein of the present invention in

combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

6.18.1. ROUTES OF ADMINISTRATION

5 Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein of the present invention used in the
10 pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

15 Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered
20 topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

25 6.18.2. COMPOSITIONS/FORMULATIONS

 Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which
30 can be used pharmaceutically. These pharmaceutical compositions may be

manufactured in a manner that is itself known, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein of the present invention is administered orally, protein of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein of the present invention, and preferably from about 25 to 90% protein of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein of the present invention, and preferably from about 1 to 50% protein of the present invention.

When a therapeutically effective amount of protein of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical

composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders
5 such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such
10 administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a
15 suitable propellant, e.g., 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. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator may be formulated containing a powder mix
20 of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as
25 suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form.

30 Additionally, suspensions of the active compounds may be prepared as

appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a cosolvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The cosolvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be

varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the proteinase inhibiting compounds of the invention may be provided as salts with pharmaceutically compatible counterions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine,

dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention. The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Pat. Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the

amount of protein of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein of the present invention and observe the patient's response. Larger doses of protein of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μ g to about 100 mg (preferably about 0.1 μ g to about 10 mg, more preferably about 0.1 μ g to about 1 mg) of protein of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the

appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalciumphosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalciumphosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor

cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-.alpha. and TGF-.beta.), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, e.g., amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (e.g., bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral

vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

5

6.18.3. EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (i.e., the concentration of the test compound which achieves a half-maximal inhibition of the C-proteinase activity). Such information can be used to more accurately determine useful doses in humans.

10

15

20

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50} . Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used

25

30

in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the C-proteinase inhibiting effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from in vitro data; for example, the concentration necessary to achieve 50-90% inhibition of the C-proteinase using the assays described herein. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

6.18.4. PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the

active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labelled for treatment of an indicated condition.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples.

EXAMPLE 1

Isolation of SEQ ID NO:1 from a cDNA Library of Human Fetal Liver-Spleen

A plurality of novel nucleic acids were obtained from a *b2HFLS20W* cDNA library prepared from human fetal liver-spleen, as described in Bonaldo et al., Genome Res. 6:791-806 (1996), using standard PCR, Sequencing by hybridization sequence signature analysis, and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for vector sequences flanking the inserts. These samples were spotted onto nylon membranes and interrogated with oligonucleotide probes to give sequence signatures. The clones were clustered into groups of similar or identical sequences, and single representative clones were selected from each group for gel sequencing. The 5' sequence of the amplified inserts was then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single-pass gel sequencing was done using a 377 Applied

Biosystems (ABI) sequencer. One of these inserts was identified as a novel sequence not previously obtained from this library and not previously reported in public databases. This sequence is shown in Figure 1 as SEQ ID NO:1.

5

EXAMPLE 2

Isolation of SEQ ID NO:2 and Determination of a Nucleotide Sequence Encoding a 428-Amino Acid Protein with Sequence Homology to CD39

10 The nucleotide sequence shown in Figure 1, and labeled SEQ ID NO:2, encodes the translated amino acid sequence SEQ ID NO:3, which is shown in Figure 2. The extended nucleotide sequence was obtained by isolating colonies generated from pools of clones from a human macrophage cDNA library (Invitrogen, Cat. # A550-25). Briefly, the macrophage cDNA
15 library was plated on LB/Amp plates (containing 100 mg/ml ampicillin) at a density of about 40,000 colonies/plate. The colonies were lifted onto nitrocellulose filters and hybridized with a radiolabeled probe generated from the original clone (i.e., SEQ ID NO:1).

20 That the identified clones corresponded to SEQ ID NOs:1 and 2 was confirmed by using gene-specific primers (5'-GCTACCTCACTTCCTTTGAG-3' [SEQ ID NO: 9] and 5'-CTGGCTGGTGAAGTTTTCCTC-3' [SEQ ID NO: 10]) in a PCR-based assay. Then PCR using vector- and gene-specific primers was employed to amplify the 5' portion of the cDNA. Nested primers were used to generate
25 sequence from the amplified product(s). Laser gene™ software was used to edit and "contig" the partial sequences into a full-length sequence. As discussed above, the amino acid sequence has striking homology to CD39, which is involved in modulating platelet reactivity during vascular inflammation. Based in part on the observed sequence similarity to CD39,
30 the polypeptide encoded by SEQ ID NO: 2 was designated CD39-L4.

EXAMPLE 3

A. Expression of SEQ ID NOS. 3 and 5 in COS-7 cells

COS-7 cells were grown in DMEM (ATCC) and 10% fetal bovine serum (FBS) (Gibco) to 70% confluence. Prior to transfection the media was changed to DMEM and 0.5% FCS. Cells were transfected with cDNAs for SEQ ID NOS. 3 and 5 or with pBGal vector by the FuGENE-6 transfection reagent (Boehringer). In summary, 4 μ l of FuGENE-6 was diluted in 100 μ l of DMEM and incubated for 5 minutes. Then, this was added to 1 μ g of DNA and incubated for 15 minutes before adding it to a 35 mm dish of COS-7 cells. The COS-7 cells were incubated at 37 °C with 5% CO₂. After 24 hours, media and cell lysates were collected, centrifuged and dialyzed against assay buffer (15 mM Tris pH 7.6, 134 mM NaCl, 5 mM glucose, 3 mM CaCl₂ and MgCl₂). More robust expression can be achieved using the protocol described in Example 6 below.

B. Expression Study Using SEQ ID NO:2

The expression of SEQ ID NO. 2 in various tissues was analyzed using a semi-quantitative polymerase chain reaction-based technique. Human cDNA libraries were used as sources of expressed genes from tissues of interest (adult brain, adult heart, adult kidney, adult lymph node, adult liver, adult lung, adult ovary, adult placenta, adult spleen, adult testis, bone marrow, fetal kidney, fetal liver, fetal liver-spleen, fetal skin, fetal brain, fetal leukocyte and macrophage). Gene-specific primers (5'-GCTACCTCACTTCCTTTGAG-3' [SEQ ID NO: 9] and 5'-GCAGGTCTCCAAGGAAGTACG-3' [SEQ ID NO: 11]) were used to amplify portions of the SEQ ID NO:2 sequence from the samples. Amplified products were separated on an agarose gel, transferred and chemically linked to a nylon filter. The filter was then hybridized with a radioactively labeled (α^{33} P-dCTP) double-stranded probe generated from the full-length SEQ ID NO:2 sequence using a Klenow polymerase, random-prime method. The filters were washed (high stringency) and used to expose a

phosphorimaging screen for several hours. Bands indicated the presence of cDNA including SEQ ID NO:2 sequences in a specific library, and thus mRNA expression in the corresponding cell type or tissue.

Of the 18 human tissues tested, macrophage was the only sample that provided a signal, indicating that expression of SEQ ID NO:2 is tightly regulated. In contrast, the CD39 molecule has been found in tissues such as placenta, lung, skeletal muscle, kidney and heart.

EXAMPLE 4

Chromosomal Localization of the Gene Corresponding to SEQ ID NOs:1 and 2

Chromosome mapping technologies allow investigators to link genes to specific regions of chromosomes. Assignment to chromosome 14 was performed with the Coriell cell repository monochromosomal panel #2 (NIGMS cell repository). This human rodent somatic cell hybrid panel consists of DNA isolated from 24 hybrid cell cultures retaining 1 human chromosome each. The panel was screened with gene-specific primers (5'-GCTACCTCACTTCCTTTGAG-3' [SEQ ID NO: 9] and 5'-CTGGCTGGTGAAGTTTTCCTC-3' [SEQ ID NO: 10]) that generated a sequence tag site (STS). The Genebridge 4 radiation hybrid panel was also screened (Research Genetics), and the results of the PCR screening were submitted to the Whitehead/MIT Radiation Hybrid mapping email server at <http://www-genome.wi.mit.edu>.

EXAMPLE 5

Platelet Aggregation Assay

Blood is anticoagulated with 0.1 volume 3.2% sodium citrate. Platelet-rich plasma (PRP) is prepared with an initial whole blood centrifugation (200 x g, 15 min., 25°C) and a second centrifugation of the PRP (90 x g, 10 min.) to eliminate residual erythrocytes and leukocytes. The stock suspension of PRP is maintained at room temperature under 5%

CO₂-air. The platelet aggregation assay uses a two-sample, four-channel Whole Blood Lumi-Aggregometer, model 560 (Chronolog Corp., Havertown, PA). PRP containing 1.22×10^8 platelets is preincubated with the sample to be tested for inhibition of aggregation for 10 min. at 37°C in a siliconized glass cuvette containing a stirring bar, followed by stimulation with either ADP (5 mM), collagen (5 mg/ml), or thrombin (0.1 unit/ml). Platelet aggregation is recorded for at least 10 min. Data are expressed as the percentage of light transmission with platelet-poor plasma equal to 100%.

Example 6

CD39-L4 is a soluble apyrase

The mammalian ectoapyrase CD39 is an integral membrane protein with two transmembrane domains (one at each end of the protein) (Maliszewski, C. R. et al., J. Immunol. 153:3574-3583). The hydrophobicity profiles for the deduced amino acid sequence of other family members, such as CD39L1 and CD39L3, are very similar to CD39 (Chadwick, B. P. and Frischauf A.M., Genomics 50:357-367), suggesting that these proteins also have two membrane spanning domains. However, CD39-L4 does not appear to have a second transmembrane domain at its C-terminus, suggesting that the N-terminus hydrophobic region could code for a secretory signal. To test this hypothesis, CD39-L4 was subcloned into the mammalian expression vector pCDNA3.1 and a 6-Histidine tag was inserted into the coding sequence.

The CD39-L4 cDNA sequence was initially isolated from a macrophage cDNA library (Invitrogen). The sense primer (5'-TTAAAGCTTGGGAAAAGAATGGCCACTTC-3', SEQ ID NO. 20) with a HindIII site and the antisense primer (5'-AGACTCGAGGTGGCTCAATGGGAGATGCC-3', SEQ ID NO. 21) with a XhoI site were used to subclone the coding sequences into the mammalian expression vector pcDNA3.1 (Invitrogen). The nucleotide sequence of the

insert is set forth in SEQ ID NO. 4. In order to immunologically detect the protein, the coding region was further modified so that it would include a Gly-Ser-6His epitope tag immediately following Arg²⁴. Briefly, two partially overlapping complementary oligonucleotides (5'-

5 GCGCTGTCTCCACAGAGGATCGCATCACCATCACCATCACAACCAGCA
GACTTGGTT-3' (SEQ ID. NO. 22) and 5'-

AACCAAGTCTGCTGGTTGTGATGGTGTGATGGTGTGCGATCCTCTGTGGG
AGACAGCGC-3' (SEQ ID NO. 23)) were used on the CD39-L4 pcDNA3.1

template. The primers were extended in opposite directions around the
10 plasmid using a 12 cycle PCR program (95°C, 1 minute; 60°C, 1 minute;
72°C, 15 minutes) (Stratagene). The reaction was treated with DpnI to digest
the methylated parental DNA and then transformed into E. coli. Colonies
were screened for the insert.

To ascertain whether CD39-L4-6His is secreted, the coding
15 region of the CD39-L4-6His protein was inserted into the pcDNA3.1
expression vector and transiently transfected into COS-7 cells. Cos-7 cells
obtained from the American Tissue Type Culture Collection were grown in
DMEM supplemented with 10% FBS and 100 units/ml penicillin G and 100
µg/ml streptomycin sulfate at 37°C in 10% CO₂. Transfections were
20 performed at 75% confluency in 10cm plates with Eugene-6 according to the
manufacturers instructions. The cells in 7 mls of medium were incubated with
16 µl of Eugene-6 and 8 µg of DNA for 14-18 hours. At the end of the
transfection the medium was replaced with DMEM medium containing low
serum (1% FBS). The cells were then incubated for 24-48 hours prior to
25 harvesting.

The CD39-L4-6His was concentrated by treating the cell lysates
and medium with Nickel-NTA agarose (Qiagen) followed by SDS/PAGE and
immunoblot analysis with an antibody against the Arg-Gly-Ser-6His epitope.
Cells were washed twice with PBS containing 0.5 µg/ml leupeptin, 0.7 µg/ml
30 pepstatin and 0.2 µg/ml aprotinin. After a brief sonication and centrifugation

step to clear the lysate, the samples were then incubated with a Nickel-NTA resin at 4°C for 2-3 hours. The histidine-tagged protein complexed to the resin was washed three times with PBS before loading onto a 10% SDS/PAGE gel for Western blot analysis. CD39-L4 was detected in both the cell lysate and the medium from cells transfected with the CD39-L4-6His expression vector, but not from control cells. While the predicted molecular weight of CD39-L4-6His is 46 kDa, the immunoreactive protein exhibited a mobility by SDS/PAGE corresponding to a molecular mass of approximately 51 kDa in the media and approximately 48 kDa in the cell lysate. The difference in apparent molecular weight may be due to posttranslational modifications of three potential N-glycosylation sites in the CD39-L4 predicted amino acid sequence.

Secretion of CD39-L4 was also examined by treatment of the transfected cells with brefeldin A, an inhibitor of translocation of secretory proteins from the endoplasmic reticulum to the Golgi apparatus. Chadwick, et al., Genomics 50:357-367 (1998). Brefeldin A was dissolved in ethanol and added to the transfected cells 48 hours after transfection. Both control and brefeldin A treated cells were washed once with PBS and incubated for 8 hours in medium with none or varying dosages of brefeldin A. Increasing dosages of brefeldin A blocked secretion of CD39-L4-6His and led to massive intracellular accumulation.

Example 7

Assay for ATPase activity

Apyrase activity was determined by measuring the amount of [³³P]P_i released from [γ³³P]ATP. In summary, 50 μl of samples were incubated in the presence of 10 μCi of [γ³³P]ATP for one hour at 37 °C. The [³³P]P_i released and the [γ³³P]ATP were separated by thin layer chromatography (TLC) plates (EM Science). The solvent system consisted of 1 M KH₂PO₄. The separated compounds were scanned for radioactivity with

a Phosphoimager (Molecular Dynamics, Sunnyvale, CA) and quantitated by ImageQuant software. COS-7 cells transfected with SEQ ID NOs. 3 and 25 had at least a four fold increase in activity over cells transfected with the vector alone. Although ATPase activity was present, Example 13 demonstrates that CD39-L4 has significantly more NDPase activity.

EXAMPLE 8

Site-directed mutagenesis of CD39L4

Site directed mutagenesis was employed to increase the enzymatic activity of CD39L4. Amino acid sequence comparisons between CD39 family members reveal four highly homologous regions in all five human members (Chadwick and Frischauf, Genomics 50:357-367, 1998). These regions, termed apyrase-conserved regions (ACRs), are present not only in the CD39 family members but other apyrases from species as distant as yeast and plants. Examination of similarities and differences in the CD39 ACRs led to the design of three CD39L4 mutants (see Figure 5). In these mutants, codons encoding CD39 ACR specific residues were used to replace codons from the CD39L4 wild type ACR sequence. Only residues with significantly different structural or chemical properties were replaced. A PCR based approach was used to produce these mutations.

Briefly, the expression vector pCDNA3.1 (Invitrogen) containing the full coding sequence of the CD39L4 gene (with a 6 Histidine tag inserted after Arg 24 in the coding sequence to allow purification of the secreted mature form of the protein) was subjected to a PCR-based site-directed mutagenesis approach using overlapping oligonucleotides [CD39-L4 ACR I mutant (nt 177-148 and 160-204): 5'-GTG AGT GCT CCC TGC ATC TAA CAT AAT TCC-3' (SEQ ID NO: 12) and 5'-GAT GCA GGG AGC ACT CAC ACT AGT ATT CAT GTT TAC ACC TTT GTG-3' (SEQ ID NO: 13); CD39-L4 ACR II mutant (nt 402-359 and 385-415): 5'-GCG TAG TCC TGC TGT TGC CCC TAG GTA CAC TGG GGT CTT TTT CC-3' (SEQ ID NO: 14) and

5'-GCA ACA GCA GGA CTA CGC TTA CTG CCA GAA C-3' (SEQ ID NO: 15); and CD39-L4 ACR III mutant (nt 532-485 and 513-540): 5'-CCC AAG CGA ATA TGC CTT CGT CTT GTC CAG TCA TGA TGC TAA CAC TGC-3' (SEQ ID NO: 16) and 5'-CGA AGG CAT ATT CGC TTG GGT TAC TGT G-3' (SEQ ID NO:17)]. After amplification of the whole plasmid with Pfu DNA polymerase (Stratagene) (95°C/1 min; 60°C/1 min; 72°C/15 min for 12 cycles), the methylated parental DNA was digested with the restriction enzyme DpnI, leaving only the unmethylated PCR amplified products. The resulting annealed double-stranded nicked products were then transformed into bacteria and the resulting colonies were screened for the desired mutations by sequencing. The subsequent constructs were fully sequenced to verify that the mutations were in fact introduced and that no extraneous mutations were generated.

Example 9

ACR III mutant increases ADPase activity

Plasmids containing the mutated and wild type forms of the CD39L4 gene were transfected into COS-7 cells. After two days, protein was purified from the culture medium using a Nickel-NTA resin approach to concentrate the tagged proteins. These proteins were then assayed for ATPase and ADPase activity by measuring the inorganic phosphate released (Wang, T.F., et al., J. Biol. Chem. 273:24814-24821, 1998). The proteins were incubated in apyrase buffer (15 mM Tris pH 7.4, 135 mM NaCl, 2mM EGTA and 10 mM glucose) for 1 hour at 37 °C with or without 2 mM CaCl₂ or 2 mM MgCl₂. Phosphatase reactions were initiated by the addition of ADP or ATP to a final concentration of 1 mM. The reaction of inorganic phosphorus with ammonium molybdate in the presence of sulfuric acid, produces an unreduced phosphomolybdate complex. The absorbance of this complex at 340 nm is directly proportional to the inorganic phosphorus concentration

(Daly, J. A., and Ertingshausen G., Clin. Chem. 18:263 (1972) (Sigma Diagnostics)).

As seen in Figure 7, mutations in ACR I and II eliminate activity, whereas the mutations in ACR III increase activity six-fold over wild type.

5 This increased activity therefore offers a greater therapeutic potential, as less protein could be administered to offer the same pharmacological effect. The replacement of three amino acids in the III region (amino acids 167 to 181 in CD39-L4) and the resulting increase in ADPase activity predicts that replacement of additional amino acids within this region by amino acids from
10 the equivalent region of CD39 may also enhance the activity of the protein over wild type CD39L4. The increase in ADPase activity over wild type may also be due to the replacement of only one or two of the three amino acids; this can be confirmed by replacing one or two amino acids at a time.

The polynucleotide and amino acid sequences of a CD39-L4
15 variant termed ACRIII and having the amino acid substitutions D168-T, S170-Q and L175-F compared to wild type CD39-L4 (SEQ ID NO: 5) are set forth in SEQ ID NOs: 6 and 7, respectively, and in Figure 6.

Example 10

20 **ACR III mutant and wild type forms are specific for ADP and not ATP**

Both the CD39L4 wild type and the CD39L4 variant with mutations in the ACRIII region hydrolyze ADP. However, when ATP was tested as a substrate, neither the CD39L4 nor the CD39L4 mutant, ACR III, catalyzed hydrolysis. In contrast, CD39 as a membrane bound molecule
25 (Marcus, et al., The Journal of Clinical Investigation, 99: 1351-1360) or as a genetically engineered soluble form (Gayle, et al., The Journal of Clinical Investigation, 101:1851-1858, 1998) is able to hydrolyze both ATP and ADP substrates efficiently. The specificity that both CD39L4 wild type and the CD39L4 ACR III mutant have for ADP is an advantageous feature that makes
30 these CD39L4-type molecules better antiplatelet therapeutic candidates than

CD39, as ADP is the agonist that causes platelet aggregation. Therapeutics that have both ADPase and ATPase activities potentially could create adverse side effects by interfering with levels of ATP in the circulation.

5

Example 11

Organization of the human CD39-L4 gene

A human CITB BAC genomic library (Research Genetics) was screened with gene specific primers [246-I6 (nt 5522-5543), 5'-CTTCCTTCACTGGGAATTCAGG-3' (SEQ ID NO: 18) and 246-K4 (nt 4922-4945), 5'-CTGTTTACCGAGATGGTTGGAAGC-3' (SEQ ID NO: 19)]
10 using a PCR based assay.

Briefly, gene specific primers were used to screen pools of BAC DNAs. BAC pools that produced an amplified DNA fragment of the predicted size were pursued until an individual BAC was identified. BAC63-I18 was
15 isolated and sequenced with gene specific primers for the CD39-L4 cDNA, as well as intron specific primers. The CD39-L4 coding sequence was found to be distributed over 10 exons spanning 9.3 kb of genomic DNA as set out in SEQ ID NO: 8.

20

Example 12

CD39-L4 is stimulated by divalent cations

The high degree of conservation in the apyrase conserved regions of CD39-L4 suggests similar function to other apyrases. To test this hypothesis, COS-7 cells were transfected with the CD39-L4-6His construct as
25 described above. The medium from transfected cells was incubated with Nickel-NTA resin (Qiagen) in order to capture the 6His tagged protein, the resin was washed with assay buffer (buffer A, 15 mM Tris pH 7.5, 134 mM NaCl and 5 mM glucose) and the protein still tethered to the resin in a suspension was assayed for ADPase activity. Nucleotidase activity was
30 determined by measuring the amount of inorganic phosphate released from

nucleotide substrates using the technique of Dlay and Ertingshausen, Clin. Chem. 18:263-265 (1972). In this reaction the complex of inorganic phosphorus with phosphor reagent (ammonium molybdate in the presence of sulfuric acid) produces an unreduced phosphomolybdate compound. The absorbance of this complex at 340 nm is directly proportional to the inorganic phosphorus concentration. The protein still tethered to the resin as a 30% suspension in buffer A was assayed by the addition of the nucleotide to a final concentration of 1 mM and incubated at 37°C for 30 minutes. The reaction was stopped by adding 100 volumes of phosphor reagent. The amount of phosphate released from the reaction was quantified using a calcium/phosphorus combined standard (Sigma). The amount of CD39-L4 protein used in the assays was estimated by comparing the intensity of the CD39-L4 band in Western blots with that of a series of standards of known quantity. CD39-L4 protein from transfected cells displayed a 2.3 fold increase in activity over the cells transfected with the vector alone. When Ca^{2+} and Mg^{2+} were added, the activity increased 3.6 fold and 6 fold, respectively.

Example 13

Characterization of CD39-L4 activity

CD39-L4 protein was assayed for ADPase activity in the presence of different kinds of inhibitors of ADPases. Control ecto-apyrase activity was determined with protein tethered to the Nickel-NTA resin. Both assays were performed as described above except the protein was in buffer A containing 2 mM CaCl_2 and 2 mM MgCl_2 . As shown by Table 1 below, inhibitors of phosphatases (F^-) and adenylate kinase (Ap5A) did not inhibit activity. The inhibitors of vacuolar ATPases (NEM), mitochondrial ATPases (N_3^-) and Na^+ , K^+ , ATPase (ouabain) did not significantly inhibit the Ca^{2+} and Mg^{2+} stimulated activity. However, metal chelators (EDTA and EGTA) significantly inhibited activity. These results show that the overwhelming

majority of the activity in the assays originates from a protein bound to the resin with characteristics of an E-type apyrase.

Table 1

Inhibition of CD39-L4 activity

INHIBITORS	% OF CONTROL
Control	100 ± 7
Ouabain (1 mM)	96 ± 6
NEM (10 mM)	106 ± 5
N ³⁻ (1 mM)	100 ± 12
F ⁻ (10 mM)	113 ± 5
Ap5A (10 μM)	121 ± 9
EGTA (2 mM)	35 ± 3
EDTA (2 mM)	52 ± 3

As shown in Table 2 below, the nucleotide specificity of CD39-L4 was also assayed as described above. The CD39-L4 activity was determined with protein tethered to the Ni-NTA resin. The protein was in buffer A containing 1 mM EGTA, as well as 2 mM CaCl₂ and MgCl₂. The assay was started by adding the nucleotides to a final concentration of 1 mM. The values below are expressed relative to ADP. The relative activity of the nucleotide triphosphates varies almost seven-fold with ATP being the poorest substrate. No phosphate release was detected with AMP and ADP was hydrolyzed at a rate approximately twenty-fold higher than ATP. The other nucleotide diphosphates (GDP and UDP) were also very efficiently hydrolyzed by CD39-L4. These results indicate that CD39-L4 defines a new class of E-type apyrase in humans with a specificity for NDPs as enzymatic substrates.

Table 2
Substrate specificity of CD39-L4

NUCLEOTIDE	% OF CONTROL
ADP	100 ± 15
ATP	5 ± 1
AMP	0
CTP	26 ± 2
GTP	34 ± 1
UTP	12 ± 4
CDP	268 ± 11
GDP	334 ± 38
UDP	408 ± 14

Example 14

Glycosylation is not essential for CD39-L4 activity

Posttranslational modifications such as N-linked glycosylation are common in secreted and membrane-bound mammalian proteins. These modifications may be important for correct protein folding or enzymatic activity and are not easily reproduced when the proteins are expressed in other organisms such as bacteria. In order to test whether CD39-L4 is glycosylated, COS-7 cells, transfected as described in Example 6, were treated with tunicamycin (Sigma), which blocks the formation of N-glycosidic linkages.

COS-7 cells were grown to 75% confluency and transfected with the CD39-L4-6His construct. After 24 hours, a fraction of the COS-7 cells were treated with Tunicamycin at a concentration of 5 µg/ml. The media was replaced again after 24 hours with fresh tunicamycin and harvested after 48 hours. The CD39-L4-6His protein was concentrated by treating the media with Nickel-NTA agarose (Qiagen). The resin was washed with assay buffer

and the protein still tethered to the resin in a suspension was assayed for a shift in electrophoretic mobility as well as its ADPase activity.

Western blot analysis using an antibody against the 6-His epitope revealed that the glycosylated CD39-L4 protein isolated from the control cells had an approximate size of 51 kDa. However, tunicamycin
5 treated cells had a molecular weight of approximately 46 kDa indicating that the protein was deglycosylated.

ADPase activity of the tunicamycin treated cells was assayed as described in Example 12 above. The deglycosylated CD39-L4 protein had
10 ADPase activity comparable to an equal amount of the glycosylated protein isolated from control cells. This demonstrates that glycosylation of the protein is not important for ADPase activity.

The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects
15 of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed
20 upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

CLAIMS

WHAT IS CLAIMED IS:

1. An isolated polynucleotide encoding an apyrase and/or NDPase and comprising a nucleotide sequence having at least about 80% sequence identity to a human polynucleotide selected from the group consisting of:
 - (a) a polynucleotide having the nucleotide sequence of SEQ ID NO. 2; and
 - (b) a polynucleotide having the protein coding nucleotide sequence of the polynucleotide sequence of (a).
2. An isolated polynucleotide encoding an apyrase and comprising a nucleotide sequence having at least about 90% sequence identity to a polynucleotide selected from the group consisting of:
 - (a) a polynucleotide having the nucleotide sequence of SEQ ID NO. 2; and
 - (b) a polynucleotide having the protein coding nucleotide sequence of the polynucleotide sequence of (a).
3. An isolated polynucleotide encoding a polypeptide with apyrase and/or NDPase activity, comprising a polynucleotide selected from the group consisting of:
 - (a) polynucleotides that encode the mature protein coding amino acid sequence of SEQ ID NO. 3.
4. An isolated polynucleotide encoding a polypeptide with apyrase and/or NDPase activity that hybridizes under stringent conditions to the complement of a polynucleotide of SEQ ID NO. 2.

5. The polynucleotide of any one of claims 1 through 4 which is a DNA.
6. The DNA of claim 5 which is a wholly or partially chemically synthesized DNA molecule.
7. An anti-sense polynucleotide which specifically hybridizes with the complement of the polynucleotide of claim 4.
8. The polynucleotide of claim 1 which comprises the nucleotide sequence of SEQ ID NO. 2 or 24 or the mature protein coding portions thereof.
9. An isolated polynucleotide which comprises a complement of the polynucleotide of claim 1.
10. An expression vector comprising the DNA of claim 5.
11. A host cell comprising the DNA of claim 5.
12. A host cell genetically engineered to express the DNA of claim 5.
13. An isolated human polypeptide with apyrase and/or NDPase activity comprising:
 - (a) the mature protein coding sequence of SEQ ID NO. 3; or
 - (b) an amino acid sequence having at least about 80% sequence identity to SEQ ID NO. 3.

14. An isolated polypeptide with apyrase and/or NDPase activity comprising:
- (a) the CD39-like protein coding sequence of SEQ ID NO. 3; or
 - (b) an amino acid sequence having at least about 90% sequence identity to SEQ ID NO. 3.
15. The polypeptide of claim 13 or 14 which comprises the amino acid sequence of SEQ ID NO. 3 or 25 or the mature protein portions thereof.
16. The polypeptide of claim 13 or 14 wherein the polypeptide comprises at least one amino acid substitution selected from the group consisting of: D168→T, S170→Q and L175→F.
17. The polypeptide of claim 16 comprising a polypeptide having the amino acid sequence set forth in SEQ ID NO. 7.
18. A method for producing a CD39-like polypeptide comprising the steps of:
- (a) growing a culture of cells according to claim 11 under conditions permitting expression of a CD39-like polypeptide; and
 - (b) isolating the CD39-like polypeptide from the host cell or its growth medium.
19. A composition comprising the polypeptide of claim 13, 14 or 15 and a pharmaceutically acceptable carrier.
19. An antibody specifically immunoreactive with a polypeptide encoded by the polynucleotide according to claim 1.

20. The antibody according to claim 19 which is a monoclonal antibody.
21. A hybridoma which secretes the antibody according to claim 20.
22. An anti-idiotypic antibody specifically immunoreactive with the antibody according to claim 20.
23. A method for detecting a polynucleotide of claim 1, 2 or 3 in a sample comprising the steps of:
- (a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to detect the complex; and
 - (b) detecting the complex so that if a complex is detected, a polynucleotide of claim 1, 2 or 3 is detected.
24. A method for detecting a polynucleotide of claim 1, 2 or 3 in a sample comprising the steps of:
- (a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of claim 1, 2 or 3 under such conditions; and
 - (b) amplifying the polynucleotides of claim 1, 2 or 3 so that if a polynucleotide is amplified, a polynucleotide of claim 1, 2 or 3 is detected.
25. The method of claim 24 wherein the polynucleotide is an RNA molecule that encodes a polypeptide of claim 13 or 14, and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

26. A method for detecting a polypeptide of claim 13 or 14 in a sample comprising:

- (a) contacting the sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to detect the complex; and
- (b) detecting the complex so that if a complex is detected, a polypeptide of claim 13 or 14 is detected.

27. A method for identifying a compound that binds to a polypeptide of claim 13 or 14 comprising:

- (a) contacting a compound with a polypeptide of claim 13 or 14 for a time sufficient to form a polypeptide/compound complex; and
- (b) detecting the complex so that if a polypeptide/compound complex is detected, a compound that binds to a polypeptide of claim 13 or 14 is detected.

28. A method for identifying a compound that binds to a polypeptide of claim 13 or 14 comprising:

- (a) contacting a compound with a polypeptide of claim 13 or 14, in a cell, for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell, and
- (b) detecting the complex by detecting reporter gene sequence expression so that if a polypeptide/compound complex is detected, a compound that binds to a polypeptide of claim 13 or 14 is identified.

29. A method of identifying a modulator compound of a CD39-like polypeptide with apyrase activity comprising the steps of:

- (a) contacting the CD39-like polypeptide encoded by the polynucleotide of claim 1, 2 or 3 with a substrate in the presence and absence of a test compound;
- (b) comparing apyrase activity of the CD39-like polypeptide in the presence and absence of the test compound; and
- (c) identifying the test compound as a modulator compound when biological activity of the CD39-like polypeptide is increased or decreased in the presence of the test compound.

30. A method of identifying a modulator compound of a CD39-like polypeptide with NDPase activity comprising the steps of:

- (a) contacting the CD39-like polypeptide encoded by the polynucleotide of claim 1, 2 or 3 with a substrate in the presence and absence of a test compound;
- (b) comparing NDPase activity of the CD39-like polypeptide in the presence and absence of the test compound; and
- (c) identifying the test compound as a modulator compound when biological activity of the CD39-like polypeptide is increased or decreased in the presence of the test compound.

31. A chimeric polypeptide comprising one or more domains of a CD39-like polypeptide fused to one or more domains of heterologous peptide or polypeptide, e.g., an immunoglobulin constant region.

32. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 13 or 14 and a pharmaceutically acceptable carrier.

33. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 13 or 14 and a pharmaceutically acceptable carrier.

34. A method of inhibiting platelet function comprising administering the polypeptide of claim 13 or 14 to a medium comprising platelets.

35. A method of treating thrombotic diseases comprising administering a therapeutic amount of the polypeptide of claim 13 or 14 to a mammalian subject in need thereof.

36. A method of hydrolyzing nucleotidediphosphates comprising administering the polypeptide of claim 13 or 14 to a medium comprising nucleotidediphosphates.

37. Use of the polypeptide of claim 13 or 14 in preparation or manufacture of a medicament for use in a method of inhibiting platelet aggregation.

38. Use of the polypeptide of claim 13 or 14 in preparation or manufacture of a medicament for use in treating thrombotic diseases.

1/9

SEQ ID NO:1:

GGCATATTAGCTTGGGTTACTGTGAATTTTCTGACAGGTCAGCTGCATGGCCACAGA
CAGGAGACTGTGGGGACCTTGGACCTAGGGGGAGCCTCCACCCAAATCACGTTCCCT
GCCCCAGTTTGAGAAAACCTCTGGAACAACTCCTAGGGGCTACCTCACTTCCTTTGA
GATGTTTAAACAGCACTTATAAGCTCTATACACATAGTTACCTGGGATTGGAATTGAA
AGCTGCAAGACTAGCAACCCTGGGAGCCCTGGAGACAGAAGGGACTGATGGGCACA
CTTCCGGAGTGCCTGT

SEQ ID NO:2:

GCGGGCTGCCGCGCAAGGGTGGCGCGCGCGCGCTTTTCCTTGTTCCCTGGTCAACAAAG
AAATGTGGAGTGTCTTGGCTGAATCCTCATAACAGACAAGATCATTATGGTGCTGTTA
GGTTGAAAAAGTGATATAATAAAGGAACCAAGGAGAAAATTCAGAAGGAAAGAAA
AAATTGCCCTCTGCAGGTGTGCGAGCAGGATTGCTTCTGCAACAAAAGCCTCCACCCA
GCCACATCTTGGGAAAAGAATGGCCACTTCTTGGGGCACAGTCTTTTTCATGCTGGT
GGTATCCTGTGTTTGCAGCGCTGTCTCCACAGGAACCAGCAGACTTGGTTTGAGGG
TATCTTCTGTCTTCCATGTGCCCCATCAATGTCAGCGCCAGCACCTTGTATGGAATT
ATGTTTGATGCAGGGAGCACTGGAACCTCGAATTCATGTTTACACCTTTGTGCAGAAA
ATGCCAGGACAGCTTCCAATTCTAGAAGGGGAAGTTTTTGATTCTGTGAAGCCAGGA
CTTTCTGCTTTTGTAGATCAACCTAAGCAGGGTGCTGAGACCGTTCAAGGGCTCTTA
GAGGTGGCCAAAGACTCAATCCCCCGAAGTCACTGGAAAAAGACCCCACTGGTCTCT
AAAGGCAACAGCAGGACTACGCTTACTGCCAGAACACAAAGCCAAGGCTCTGCTCT
TTGAGGTAAAGGAGATCTTCAGGAAGTCACCTTTCTGGTACCAAGGGCAGTGTTA
GCATCATGGATGGATCCGACGAAGGCATATTAGCTTGGGTTACTGTGAATTTTCTGA
CAGGTCAGCTGCATGGCCACAGACAGGAGACTGTGGGGACCTTGGACCTAGGGGGA
GCCTCCACCCAAATCACGTTCCCTGCCCCAGTTTGAGAAAACCTCTGGAACAACTCCT
AGGGGCTACCTCACTTCCTTTGAGATGTTTAAACAGCACTTATAAGCTCTATACACAT
AGTTACCTGGGATTGGAATTGAAAGCTGCAAGACTAGCAACCCTGGGAGCCCTGGA
GACAGAAGGGACTGATGGGCACACTTCCGGAGTGCCTGTTTACCGAGATGGTTGG
AAGCAGAGTGGATCTTTGGGGGTGTGAAATACCAGTATGGTGGCAACCAAGAAGGG
GAGGTGGGCTTTGAGCCCTGCTATGCCGAAGTGCTGAGGGTGGTACGAGGAAAACCT
TCACCAGCCAGAGGAGGTCCAGAGAGGTTCTTCTATGCTTTCTCTTACTATTATGA
CCGAGCTGTTGACACAGACATGATTGATTATGAAAAGGGGGGTATTTTAAAAGTTGA
AGATTTTGAAAGAAAAGCCAGGGAAGTGTGTGATAACTTGGAAAACCTCACCTCAG
GCAGTCCTTTCTGTGCATGGATCTCAGCTACATCACAGCCCTGTTAAAGGATGGCT
TTGGCTTTGCAGACAGCACAGTCTTACAGCTCACAAGAAAGTGAACAACATAGAG
ACGGGCTGGGCCTTGGGGGCCACCTTTACCTGTTGCAGTCTCTGGGCATCTCCCAT
TGAGGCCACGTACTTCCTTGGAGACCTGCATTTGCCAACACCTTTTTAAGGGGAGGA
GAGAGCACTTAGTTTCTGAACTAGTCTGGGGACATCCTGGACTTGAGCCTAGAGATT
WRGTTAATTAASCGGCCGAGCTTATCCTTWATRAGGTAATTTACTTGCMTGGCCGCG
TTTACACGTCGTGATGGNAAACCTGCGTCCCAACTAACGCTTGASAMATCCCTTCG
CAGCTGCGATACCAAAAGCCGACGACGCCTTCCACAGTGCCA

Figure 1

SEQ ID NO:3:

MATSWGTVFFMLVVSCVCSAVSHRNQQTWFEGIFLSSMCPINVSASTLYGIMFDAGSTG
TRIHVYTFVQKMPGQLPILEGEVFDVSKPGLSAFVDQPKQGAETVQGLLEVAKDSIPRSH
WKKTPVVLKATAGLRLLPEHKAKALLFEVKEIFRKSPFLVPKGSVSIMDGSDEGILAWV
TVNFLTGQLHGHRQETVGTLDLGGASTQITFLPQFEKTLEQTPRGYLTSFEMFNSTYKLY
THSYLGFGLKAARLATLGALETEGTDGHTFRSACLPRWLEAEWIFGGVKYQYGGNQEG
EVGFEP CYAEVLRVVRGKLHQPEEVQRGSFYAFSYYYDRAVDTDMIDYEKG GILKVED
FERKAREVCDNLENFTSGSPFLCMDLSYITALLKDGF GFADSTVLQLTKKVNNIETGWA
LGATFHLLQSLGISH

Figure 2

3/9

CD39Human..seq	M E D T K E S N V R T F C S K N I C A I L G F S S T A Y T A L L A - - - - -	36
246 prot	M A T L S W G L V F - - - - - F M L V M S C V C S A M S H R N Q Q T A F E E T F	24
CD39Human..seq	Q N K - - - - A L P E N N K Y G I V L D A G S S H T S I V Y A P A E R E N D	74
246 prot	L S I S M C P I N V S A L S T L Y G L M F D A G S T G I R I H V Y T F L Q K M P Q C	74
CD39Human..seq	T G V V H O V E E C R V K Q P G I S K F V O K V N E I G Y L T C O M E P A R E	124
246 prot	L P L L E G E V F D S I V K - P G L S A F V I D Q P K Q G A E T V Q C L L E V A K Q	124
CD39Human..seq	V I P R S Q H Q E T F V Y L G A T A G M R L L R M E S E E L A D E R K L D V E R	154
246 prot	S I P R S H W K K I T P V I V K A T A G L R L L - - - P I E H K I A K A L L F E M K E	150
CD39Human..seq	S L S N Y P F D F Q - - - G A R I T T G Q E E G A Y G W I T E N Y L L Q K F S C F	180
246 prot	I F R K S P E L V P K G S V S L M D G S I D E G I L A W V T V N F L T Q I Q L - - -	187
CD39Human..seq	T E W F S I V P Y E T N N Q E C F G A L D L G G A S T Q V T F V P Q - N Q T T E	231
246 prot	- - - - - H G H R Q E T V G T L D L G G A S T Q T T E L I P Q F E K I T L E	218
CD39Human..seq	S P D N A - - - L Q F R L Y G K D Y N V Y T H S F L C Y G K D O A L W O K L A K D	269
246 prot	Q T P R G Y L T S E E M F N S T Y K L Y T H S Y L G F G L K A A - - - R L A T L	255
CD39Human..seq	I O V A S N E I L R D P C F H P Q Y K K V V N V S D L Y K T P C T K R - F E M T	306
246 prot	G A L E T E G - - - - - T D G H T F R L S A C L P I R W L E A E	280
CD39Human..seq	L P F O O F - - - - - E I Q G I G N Y O O C H O S I L E L E N T S Y C P Y S O	342
246 prot	W I E G G V K Y Q Y G G N Q E I G E V G F E P C Y A E V I R V V R G K - - - - -	314
CD39Human..seq	C A F N G I E L P P L O G E F G A F S A F - - - Y F V M K F L N L T S E K V S O E	380
246 prot	- - - - - L H Q P E E V Q R I G S F Y A F S Y Y Y D R - - - A V D T D M I D Y E	345
CD39Human..seq	K V T E M - M K K F C A O P W E - - - E I K T S Y A G V K E K Y L S E Y C F S G T	417
246 prot	K G G I L K V E D E E R K A R E V C D N L E N F T S G S P - F L - - - C M D L S	381
CD39Human..seq	Y I L S L L L O G Y H F T A D S W E H I H F I G K I O G S D A G W T L G Y M L N	457
246 prot	Y I T A L L K D G F G E N D S T - - - V L Q L T K K V N N I E T G W A L G A T F H	419
CD39Human..seq	L T N M I P A E O P L S T P L S H S T Y V F L M V L F S L V L F T V A I I G L L	497
246 prot	L L Q S L G I S H	428
CD39Human..seq	I F H K P S Y F W K D M V	510
246 prot		428

Figure 3

4/9

1	IMATSWGTVFEMLVVSCVICSAVSHRNQQTWFECIFLSSMCP	246	prot
1	IMATSWGAVF-IMETIACVGSTVFYREOQTWFECIFLSSMCP		mur ntpase
41	INVSASTLYGIMFDAGSTGTRIHVYTFVQAMPGQLPFL	246	prot
40	INVSAQTFLYGIMFDAGSTGTRIHVYTFVOKTAGQLPFL		mur ntpase
81	EVFDSVVKPGLSAFVDQPKQGAETVQGLLEVAKDSIPRS	246	prot
80	EVFDSVVKPGLSAFVDQPKQGAETVQGLLEVAKDSIPRS		mur ntpase
121	KKTPVVLKATAGLRLLPEHKAIAKALLFEVKEEIPRKIS	246	prot
120	ERTPVVLKATAGLRLLPEOKAQALLLEVEEIPKNS		mur ntpase
161	KIGSVSIMDGSDEGILAWVTVNFLTGGQLHGHRIQETVGT	246	prot
50	DGSV SIMDGSYEGILAWVTVNFLTGGQLHGHRIQETVGT		mur ntpase
201	GGASTQITFLPQFEKTLQTPRGYLTSEMFNSTYKCLYTH	246	prot
200	GGASTQITFLPQFEKTLQTPRGYLTSEMFNSTEKLCTH		mur ntpase
241	SYLGFGKKAARLATLGALETETGTDGHTERSACLPRWLEAE	246	prot
240	SYLGFGKKAARLATLGALEAKIGTDGHTERSACLPRWLEAE		mur ntpase
281	WIFGGVKYQYGGNQEGEVGFEP CYAEVLRVVIRGKLHQPEE	246	prot
280	WIFGGVKYQYGGNQEGEMIGFEP CYAEVLRVVIRGKLHQPEE		mur ntpase
321	VQRGSFYAFSYYYDRAVDTDIMIDYEKGGILKVEDFERKAR	246	prot
320	VRGSFYAFSYYYDRAADTHLIDYEKGGVVKVEDFERKAR		mur ntpase
361	EVCDNLENFETSGSPFLCMDLSYITALLKDGFGFADSTVLO	246	prot
360	EVCDNLGSSSSGSPFLCMDLYITALLKDGGLGFAERHPLT		mur ntpase
401	LTAKKVNNIETGW-ALGATF-----HLLQSLGISH	246	prot
400	-AHKWESEQHRDWLQLGGHLSAPAPVSGHHQLRPSSTSEACT		mur ntpase
428		246	prot
439	SEPVFSSQEGVDSETFSCLSGKAWPETR		mur ntpase

Figure 4

Apyrase Conserved Regions in CD39-L4

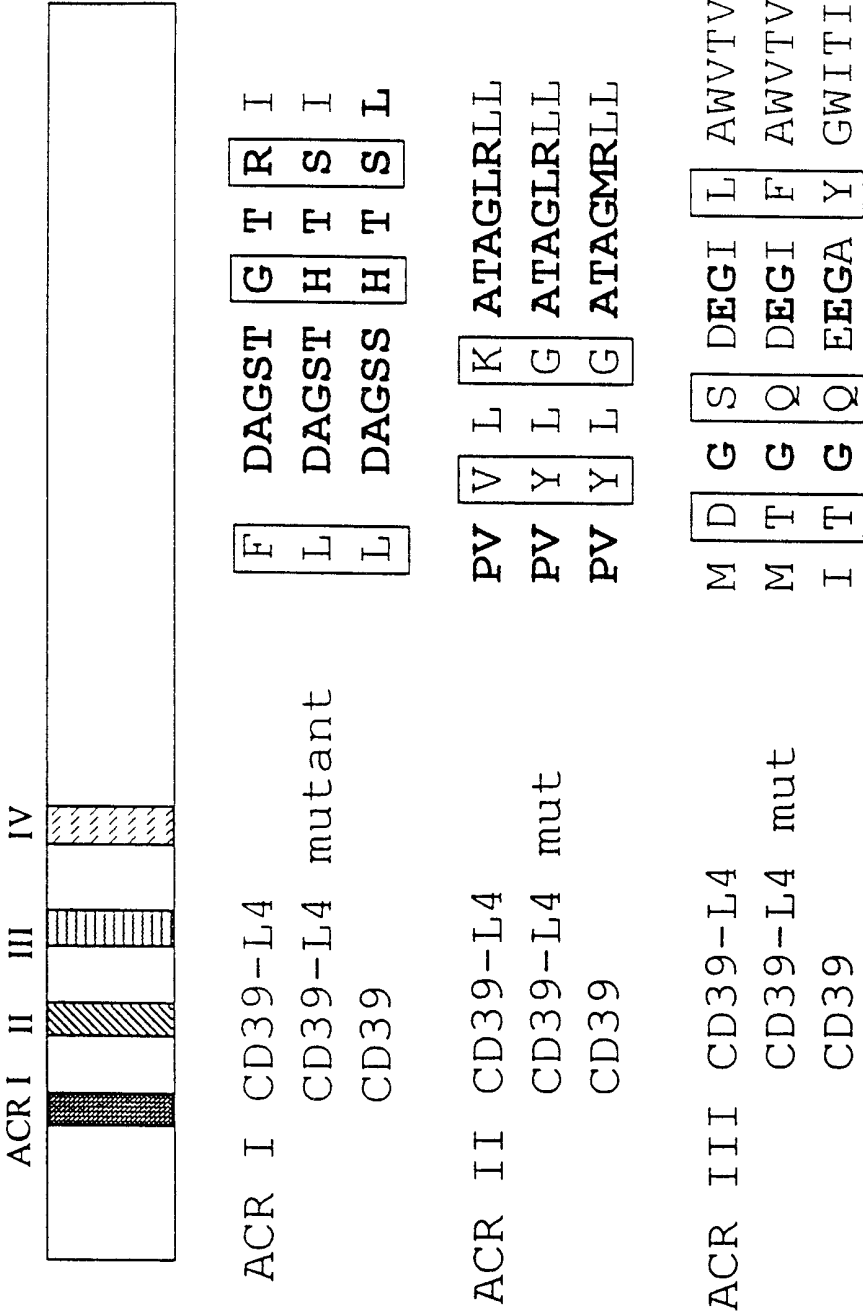


FIG. 5

6/9

Nucleotide sequence of the CD39-L4 mutant ACRIII (SEQ ID NO:6). The nucleotide changes have been highlighted. The G to A and A to C changes at positions 502 and 503 produce a Thr, the T to C, C to A and C to A changes at positions 508-510 result in a Gln and the A to C changed at position 525 result in a Phe.

```

ATGGCCACTTCTTGGGGCACAGTCTTTTTTCATGCTGGTGGTATCCTGTGTTTGCAGCGCTGTCT
CCCACAGGAACCAGCAGACTTGGTTTTGAGGGTATCTTCCTGTCTTCCATGTGCCCCATCAATGT
CAGCGCCAGCACCTTGTATGGAATTATGTTTGTATGCAGGGAGCACTGGAACCTCGAATTCATGTT
TACACCTTTGTGCAGAAAATGCCAGGACAGCTTCCAATTCTAGAAGGGGAAGTTTTTGATTCTG
TGAAGCCAGGACTTTCTGCTTTTGTAGATCAACCTAAGCAGGGTGCTGAGACCGTTCAAGGGCT
CTTAGAGGTGGCCAAAGACTCAATCCCCGAAAGTCACTGGAAAAAGACCCCACTGGTCTCTAAAG
GCAACAGCAGGACTACGCTTACTGCCAGAACACAAAGCCAAGGCTCTGCTCTTTGAGGTAAAGG
AGATCTTCAGGAAGTCACCTTTCCTGGTACCAAAGGGCAGTGTTAGCATCATGACTGGACAAGA
CGAAGGCATATTCGCTTGGGTTACTGTGAATTTTCTGACAGGTCAGCTGCATGGCCACAGACAG
GAGACTGTGGGGACCTTGGACCTAGGGGGAGCCTCCACCCAAATCACGTTCTGCCCCAGTTTG
AGAAAACCTCTGGAACAACTCCTAGGGGCTACCTCACTTCCTTTGAGATGTTTAACAGCACTTA
TAAGCTCTATACACATAGTTACCTGGGATTTGGATTGAAAGCTGCAAGACTAGCAACCCTGGGA
GCCCTGGAGACAGAAGGGACTGATGGGCACACTTTCGGGAGTGCCTGTTTACCGAGATGGTTGG
AAGCAGAGTGGATCTTTGGGGGTGTGAAATACCAGTATGGTGGCAACCAAGAAGGGGAGGTGGG
CTTTGAGCCCTGCTATGCCGAAGTGCTGAGGGTGGTACGAGGAAAACCTTCACCAGCCAGAGGAG
GTCCAGAGAGGTTCTCTTCTATGCTTTCTCTTACTATTATGACCGAGCTGTTGACACAGACATGA
TTGATTATGAAAAGGGGGGTATTTTAAAAGTTGAAGATTTTGAAGAAAAGCCAGGGAAGTGTG
TGATAACTTGAAAACCTTCACC TCAGGCAGTCCTTTCTGTGCATGGATCTCAGCTACATCAC
AGCCCTGTTA AAGGATGGCTTTGGCTTTGCAGACAGCACAGTCTTACAGCTCACAAAGAAAGT
GAAC AACATAG AGACGGGCTGGGCCTTGGGGGCCACCTTTCACCTGTTGCAGTCTCTGGGCA
TCTCCCATTTGA

```

Amino acid sequence of CD39-L4 mutant ACR III (SEQ ID NO:7). The amino acid changes are D to T (a.a. 168), S to Q (a.a. 170) and L to F (a.a. 175). The changes are shown in bold.

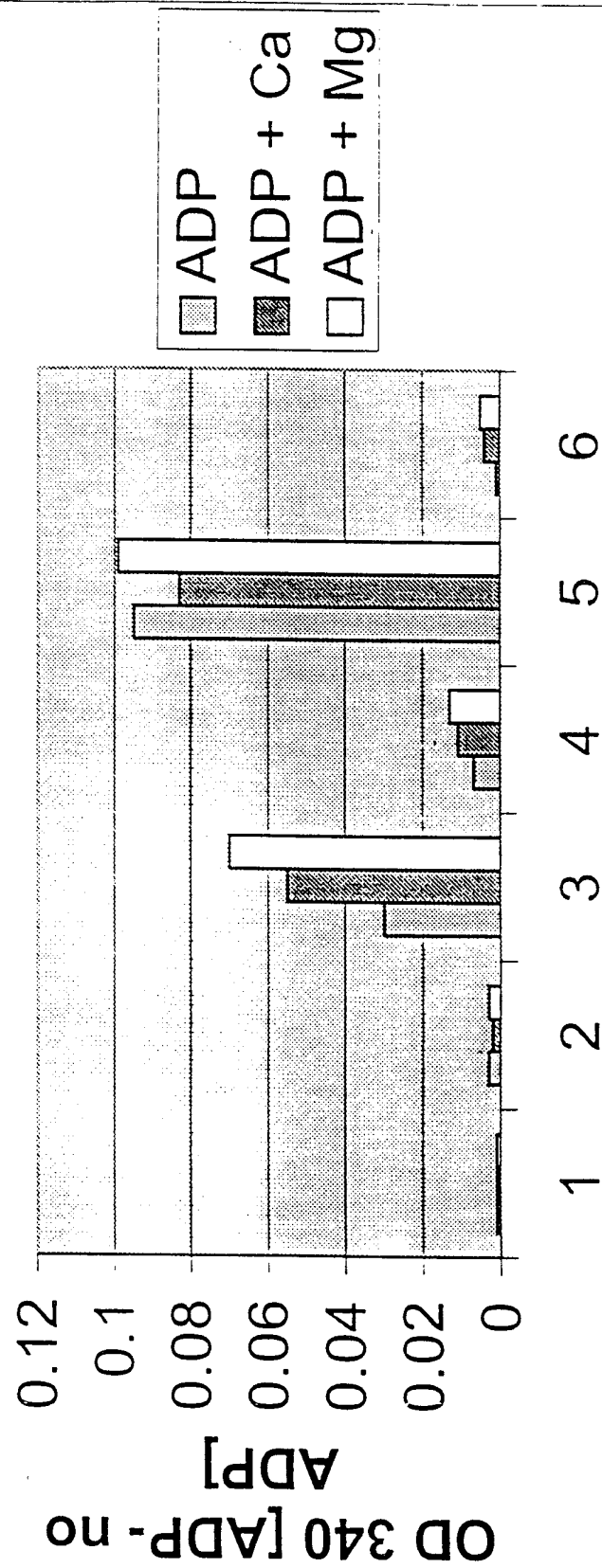
```

MATSYGTVFFMLVVSVCVCSAVSHRNQQTWFEGIFLSSMCPINVSASTLYGIMFDAGSTGT
RIHVYTFVQKMPGQLPILEGEVFDVSKPGLSAFVDQPKQGAETVQGLLEVAKDSIPRSHW
KKTPVVLKATAGLRLLPEHKAKALLFEVKEIFRKSPFLVPKGSVSIMTGQDEGIFAWVTV
NFLTGQLHGHROETVGTLDLGGASTQITFLPQFEKTLQTPRGYLTSFEMFNSTYKLYTH
SYLGFGGLKAARLATLGALETEGTDGHTFRSACLPRWLEAEWIFGGVKYQYGGNQEGEVGF
EPCYAEVLRVVRGKLHQPEEVQRGSFYAFSYYYDRAVD TDMIDYEKGGILKVEDFERKAR
EVCNLENFTSGSPFLCMDLSYITALLKDGFGFADSTVLQLTKKVNNIETGWALGATFHL
LQSLGISH

```

FIG. 6

FIG. 2 ADPase activity of CD39-L4
ACR mutants



1) ACR I; 2) ACR II; 3) ACR III; 4)
Wild Type; 5) CD39; 6) Vector

CD39-L4

FIG. 7

8/9

SEQ ID No-3	MATSWG	46
SEQ ID No-5	MATSWG	46
CD39Human.seq	MEDTKESNVKTFCSKNILAILGFSIIAVIA	46
SEQ ID No-3	TLYGIMFDAGSTGTRIHHVYTFVQKMPGQLPILEGEVFDVSK	100
SEQ ID No-5	TLYGIMFDAGSTGTRIHHVYTFVQKMPGQLPILEGEVFDVSK	100
CD39Human.seq	VKYGIVLDAGSHTSLYIKWPAEKENDTGVVHQVEECRVKGPISKFVQKVNEI	101
SEQ ID No-3	AETVQGLLEVAKDSIPRSHWKKTPVVLKATAGLRLL	152
SEQ ID No-5	AETVQGLLEVAKDSIPRSHWKKTPVVLKATAGLRLL	152
CD39Human.seq	GIYLTDCMERAREVIRPSQHQQETPVYLGATAGMRLLRMESEELADRVLDDVVERSL	156
SEQ ID No-3	RKSPFLVPKGSVSIMDGSDEGILAWVTVNFLTQQL	195
SEQ ID No-5	RKSPFLVPKGSVSIMDGSDEGILAWVTVNFLTQQL	195
CD39Human.seq	SNYPFDFQ--GARITITGQEEGAYGWIITINYLGLGKFSQKTRWFSIVPYETNNQET	209
SEQ ID No-3	GTLDLGGASTQITFLPQFEKTLLEQTPRGYLTSTFEMFNSTYKLYTHSYLGFGLKAA	250
SEQ ID No-5	GTLDLGGASTQITFLPQFEKTLLEQTPRGYLTSTFEMFNSTYKLYTHSYLGFGLKAA	250
CD39Human.seq	GALDLGGASTQITFLPQFEKTLLEQTPRGYLTSTFEMFNSTYKLYTHSYLGFGLKAA	261
SEQ ID No-3	RLATLGALETEG	287
SEQ ID No-5	RLATLGALETEG	287
CD39Human.seq	LWQKLLAKDIQVASNEILRDPCHFPGYKVKVNVSDLYKTPCTKR--FEMTLPFQQF	314
SEQ ID No-3	YQYGGNQEGEVGFEPCCYAEVLRVVRGK	330
SEQ ID No-5	YQYGGNQEGEVGFEPCCYAEVLRVVRGK	330
CD39Human.seq	YQYGGNQEGEVGFEPCCYAEVLRVVRGK	363
SEQ ID No-3	YYYDR	378
SEQ ID No-5	YYYDR	378
CD39Human.seq	YFVVMKFLNLTSEKVSQEKVTEM--MKKECAQPWEE--EIKTSYAGVKEKYLSEYCF	414
SEQ ID No-3	DLSYITALLKDGFGFADST	412
SEQ ID No-5	DLSYITALLKDGFGFADST	402
CD39Human.seq	SGTYILSLQLQGYHFTADSWEHIFIGKIQGSDAGWTLGYMLNLT	459
SEQ ID No-3	ALGATFHLLQSLGISH	428
SEQ ID No-5	ALGATFHLLQSLGISH	405
CD39Human.seq	ALGATFHLLQSLGISH	502

FIGURE 8

FIGURE 9

- 1 -

SEQUENCE LISTING

<110> Hyseq
Ford, John (inventor)
Mulero, Julio (inventor)

<120> METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
POLYPEPTIDES

<130> 28110/35836

<140>

<141>

<150> 09/273,447

<151> 1999-03-19

<150> 09/118,205

<151> 1998-07-16

<150> 09/122,449

<151> 1998-07-24

<150> 09/244,444

<151> 1999-02-04

<150> XX/XXX,XXX

<151> 1999-07-09

<160> 25

<170> PatentIn Ver. 2.0

<210> 1

<211> 300

<212> DNA

<213> Homo sapiens

<400> 1

ggcatattag cttgggttac tgtgaatttt ctgacaggtc agctgcatgg ccacagacag 60

gagactgtgg ggaccttga cctaggggga gcctccaccc aaatcacgtt cctgccccag 120

tttgagaaaa ctctggaaca aactcctagg ggctacctca cttcctttga gatgtttaac 180

agcacttata agctctatac acatagttac ctgggatttg gattgaaagc tgcaagacta 240

gcaaccctgg gagccctgga gacagaaggg actgatgggc acactttccg gagtgcctgt 300

<210> 2

<211> 1799

<212> DNA

<213> Homo sapiens

- 2 -

<220>

<221> CDS

<222> (246)..(1529)

<220>

<221> misc_feature

<222> (1718)

<223> n = adenine or guanine or cytosine or thymine

<400> 2

```

gcgggctgcc gcgcaagggt ggcgcgcgcg cgttttcctt gttcctgggc aacaaagaaa 60
tgtggagtggt cttggctgaa tcctcataca gacaagatca ttatgggtgct gttaggttga 120
aaaagtgata taataaagga accaaggaga aaattcagaa ggaaagaaaa aattgcctct 180
gcagggtgtgc gagcaggatt gcttctgcaa caaaagcctc caccagcca catcttggga 240
aaaga atg gcc act tct tgg ggc aca gtc ttt ttc atg ctg gtg gta tcc 290
      Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser
        1             5             10             15

tgt gtt tgc agc gct gtc tcc cac agg aac cag cag act tgg ttt gag 338
Cys Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu
              20             25             30

ggt atc ttc ctg tct tcc atg tgc ccc atc aat gtc agc gcc agc acc 386
Gly Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr
              35             40             45

ttg tat gga att atg ttt gat gca ggg agc act gga act cga att cat 434
Leu Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His
              50             55             60

gtt tac acc ttt gtg cag aaa atg cca gga cag ctt cca att cta gaa 482
Val Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu
              65             70             75

ggg gaa gtt ttt gat tct gtg aag cca gga ctt tct gct ttt gta gat 530
Gly Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp
              80             85             90             95

caa cct aag cag ggt gct gag acc gtt caa ggg ctc tta gag gtg gcc 578
Gln Pro Lys Gln Gly Ala Glu Thr Val Gln Gly Leu Leu Glu Val Ala
              100            105            110

aaa gac tca atc ccc cga agt cac tgg aaa aag acc cca gtg gtc cta 626
Lys Asp Ser Ile Pro Arg Ser His Trp Lys Lys Thr Pro Val Val Leu
              115            120            125

aag gca aca gca gga cta cgc tta ctg cca gaa cac aaa gcc aag gct 674
Lys Ala Thr Ala Gly Leu Arg Leu Leu Pro Glu His Lys Ala Lys Ala
              130            135            140

```

- 3 -

ctg ctc ttt gag gta aag gag atc ttc agg aag tca cct ttc ctg gta	722
Leu Leu Phe Glu Val Lys Glu Ile Phe Arg Lys Ser Pro Phe Leu Val	
145 150 155	
cca aag ggc agt gtt agc atc atg gat gga tcc gac gaa ggc ata tta	770
Pro Lys Gly Ser Val Ser Ile Met Asp Gly Ser Asp Glu Gly Ile Leu	
160 165 170 175	
gct tgg gtt act gtg aat ttt ctg aca ggt cag ctg cat ggc cac aga	818
Ala Trp Val Thr Val Asn Phe Leu Thr Gly Gln Leu His Gly His Arg	
180 185 190	
cag gag act gtg ggg acc ttg gac cta ggg gga gcc tcc acc caa atc	866
Gln Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile	
195 200 205	
acg ttc ctg ccc cag ttt gag aaa act ctg gaa caa act cct agg ggc	914
Thr Phe Leu Pro Gln Phe Glu Lys Thr Leu Glu Gln Thr Pro Arg Gly	
210 215 220	
tac ctc act tcc ttt gag atg ttt aac agc act tat aag ctc tat aca	962
Tyr Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys Leu Tyr Thr	
225 230 235	
cat agt tac ctg gga ttt gga ttg aaa gct gca aga cta gca acc ctg	1010
His Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr Leu	
240 245 250 255	
gga gcc ctg gag aca gaa ggg act gat ggg cac act ttc cgg agt gcc	1058
Gly Ala Leu Glu Thr Glu Gly Thr Asp Gly His Thr Phe Arg Ser Ala	
260 265 270	
tgt tta ccg aga tgg ttg gaa gca gag tgg atc ttt ggg ggt gtg aaa	1106
Cys Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys	
275 280 285	
tac cag tat ggt ggc aac caa gaa ggg gag gtg ggc ttt gag ccc tgc	1154
Tyr Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys	
290 295 300	
tat gcc gaa gtg ctg agg gtg gta cga gga aaa ctt cac cag cca gag	1202
Tyr Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu	
305 310 315	
gag gtc cag aga ggt tcc ttc tat gct ttc tct tac tat tat gac cga	1250
Glu Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg	
320 325 330 335	
gct gtt gac aca gac atg att gat tat gaa aag ggg ggt att tta aaa	1298
Ala Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys	
340 345 350	
gtt gaa gat ttt gaa aga aaa gcc agg gaa gtg tgt gat aac ttg gaa	1346
Val Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu	
355 360 365	

- 4 -

aac ttc acc tca ggc agt cct ttc ctg tgc atg gat ctc agc tac atc 1394
 Asn Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile
 370 375 380

aca gcc ctg tta aag gat ggc ttt ggc ttt gca gac agc aca gtc tta 1442
 Thr Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu
 385 390 395

cag ctc aca aag aaa gtg aac aac ata gag acg ggc tgg gcc ttg ggg 1490
 Gln Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly
 400 405 410 415

gcc acc ttt cac ctg ttg cag tct ctg ggc atc tcc cat tgaggccacg 1539
 Ala Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
 420 425

tacttccttg gagacctgca ttgccaaca cctttttaag gggaggagag agcacttagt 1599

ttctgaacta gtctggggac atcctggact tgagcctaga gattwrgtta attaascggc 1659

cgagcttata cttwatragg taattttactt gcmtggccgc gtttacacgt cgtgatggna 1719

aacctgcgtc ccaactaacg cttgasamat ccccttcgca gctgcgatac caaaagccga 1779

cgacgccttc cacagtgcc 1799

<210> 3

<211> 428

<212> PRT

<213> Homo sapiens

<400> 3

Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser Cys
 1 5 10 15

Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu Gly
 20 25 30

Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr Leu
 35 40 45

Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His Val
 50 55 60

Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu Gly
 65 70 75 80

Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln
 85 90 95

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

- 5 -

Asp	Ser	Ile	Pro	Arg	Ser	His	Trp	Lys	Lys	Thr	Pro	Val	Val	Leu	Lys	115	120	125
Ala	Thr	Ala	Gly	Leu	Arg	Leu	Leu	Pro	Glu	His	Lys	Ala	Lys	Ala	Leu	130	135	140
Leu	Phe	Glu	Val	Lys	Glu	Ile	Phe	Arg	Lys	Ser	Pro	Phe	Leu	Val	Pro	145	150	155
Lys	Gly	Ser	Val	Ser	Ile	Met	Asp	Gly	Ser	Asp	Glu	Gly	Ile	Leu	Ala	165	170	175
Trp	Val	Thr	Val	Asn	Phe	Leu	Thr	Gly	Gln	Leu	His	Gly	His	Arg	Gln	180	185	190
Glu	Thr	Val	Gly	Thr	Leu	Asp	Leu	Gly	Gly	Ala	Ser	Thr	Gln	Ile	Thr	195	200	205
Phe	Leu	Pro	Gln	Phe	Glu	Lys	Thr	Leu	Glu	Gln	Thr	Pro	Arg	Gly	Tyr	210	215	220
Leu	Thr	Ser	Phe	Glu	Met	Phe	Asn	Ser	Thr	Tyr	Lys	Leu	Tyr	Thr	His	225	230	235
Ser	Tyr	Leu	Gly	Phe	Gly	Leu	Lys	Ala	Ala	Arg	Leu	Ala	Thr	Leu	Gly	245	250	255
Ala	Leu	Glu	Thr	Glu	Gly	Thr	Asp	Gly	His	Thr	Phe	Arg	Ser	Ala	Cys	260	265	270
Leu	Pro	Arg	Trp	Leu	Glu	Ala	Glu	Trp	Ile	Phe	Gly	Gly	Val	Lys	Tyr	275	280	285
Gln	Tyr	Gly	Gly	Asn	Gln	Glu	Gly	Glu	Val	Gly	Phe	Glu	Pro	Cys	Tyr	290	295	300
Ala	Glu	Val	Leu	Arg	Val	Val	Arg	Gly	Lys	Leu	His	Gln	Pro	Glu	Glu	305	310	315
Val	Gln	Arg	Gly	Ser	Phe	Tyr	Ala	Phe	Ser	Tyr	Tyr	Tyr	Asp	Arg	Ala	325	330	335
Val	Asp	Thr	Asp	Met	Ile	Asp	Tyr	Glu	Lys	Gly	Gly	Ile	Leu	Lys	Val	340	345	350
Glu	Asp	Phe	Glu	Arg	Lys	Ala	Arg	Glu	Val	Cys	Asp	Asn	Leu	Glu	Asn	355	360	365
Phe	Thr	Ser	Gly	Ser	Pro	Phe	Leu	Cys	Met	Asp	Leu	Ser	Tyr	Ile	Thr	370	375	380
Ala	Leu	Leu	Lys	Asp	Gly	Phe	Gly	Phe	Ala	Asp	Ser	Thr	Val	Leu	Gln	385	390	395

- 6 -

Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly Ala
 405 410 415

Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
 420 425

<210> 4
 <211> 1287
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1) .. (1284)

<400> 4
 atg gcc act tct tgg ggc aca gtc ttt ttc atg ctg gtg gta tcc tgt 48
 Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser Cys
 1 5 10 15
 gtt tgc agc gct gtc tcc cac agg aac cag cag act tgg ttt gag ggt 96
 Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu Gly
 20 25 30
 atc ttc ctg tct tcc atg tgc ccc atc aat gtc agc gcc agc acc ttg 144
 Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr Leu
 35 40 45
 tat gga att atg ttt gat gca ggg agc act gga act cga att cat gtt 192
 Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His Val
 50 55 60
 tac acc ttt gtg cag aaa atg cca gga cag ctt cca att cta gaa ggg 240
 Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu Gly
 65 70 75 80
 gaa gtt ttt gat tct gtg aag cca gga ctt tct gct ttt gta gat caa 288
 Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln
 85 90 95
 cct aag cag ggt gct gag acc gtt caa ggg ctc tta gag gtg gcc aaa 336
 Pro Lys Gln Gly Ala Glu Thr Val Gln Gly Leu Leu Glu Val Ala Lys
 100 105 110
 gac tca atc ccc cga agt cac tgg aaa aag acc cca gtg gtc cta aag 384
 Asp Ser Ile Pro Arg Ser His Trp Lys Lys Thr Pro Val Val Leu Lys
 115 120 125
 gca aca gca gga cta cgc tta ctg cca gaa cac aaa gcc aag gct ctg 432
 Ala Thr Ala Gly Leu Arg Leu Leu Pro Glu His Lys Ala Lys Ala Leu
 130 135 140

- 7 -

ctc ttt gag gta aag gag atc ttc agg aag tca cct ttc ctg gta cca	480
Leu Phe Glu Val Lys Glu Ile Phe Arg Lys Ser Pro Phe Leu Val Pro	
145 150 155 160	
aag ggc agt gtt agc atc atg gat gga tcc gac gaa ggc ata tta gct	528
Lys Gly Ser Val Ser Ile Met Asp Gly Ser Asp Glu Gly Ile Leu Ala	
165 170 175	
tgg gtt act gtg aat ttt ctg aca ggt cag ctg cat ggc cac aga cag	576
Trp Val Thr Val Asn Phe Leu Thr Gly Gln Leu His Gly His Arg Gln	
180 185 190	
gag act gtg ggg acc ttg gac cta ggg gga gcc tcc acc caa atc acg	624
Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile Thr	
195 200 205	
ttc ctg ccc cag ttt gag aaa act ctg gaa caa act cct agg ggc tac	672
Phe Leu Pro Gln Phe Glu Lys Thr Leu Glu Gln Thr Pro Arg Gly Tyr	
210 215 220	
ctc act tcc ttt gag atg ttt aac agc act tat aag ctc tat aca cat	720
Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys Leu Tyr Thr His	
225 230 235 240	
agt tac ctg gga ttt gga ttg aaa gct gca aga cta gca acc ctg gga	768
Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr Leu Gly	
245 250 255	
gcc ctg gag aca gaa ggg act gat ggg cac act ttc cgg agt gcc tgt	816
Ala Leu Glu Thr Glu Gly Thr Asp Gly His Thr Phe Arg Ser Ala Cys	
260 265 270	
tta ccg aga tgg ttg gaa gca gag tgg atc ttt ggg ggt gtg aaa tac	864
Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr	
275 280 285	
cag tat ggt ggc aac caa gaa ggg gag gtg ggc ttt gag ccc tgc tat	912
Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys Tyr	
290 295 300	
gcc gaa gtg ctg agg gtg gta cga gga aaa ctt cac cag cca gag gag	960
Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu Glu	
305 310 315 320	
gtc cag aga ggt tcc ttc tat gct ttc tct tac tat tat gac cga gct	1008
Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg Ala	
325 330 335	
gtt gac aca gac atg att gat tat gaa aag ggg ggt att tta aaa gtt	1056
Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys Val	
340 345 350	
gaa gat ttt gaa aga aaa gcc agg gaa gtg tgt gat aac ttg gaa aac	1104
Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu Asn	
355 360 365	

- 8 -

ttc acc tca ggc agt cct ttc ctg tgc atg gat ctc agc tac atc aca 1152
 Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile Thr
 370 375 380

gcc ctg tta aag gat ggc ttt ggc ttt gca gac agc aca gtc tta cag 1200
 Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu Gln
 385 390 395 400

ctc aca aag aaa gtg aac aac ata gag acg ggc tgg gcc ttg ggg gcc 1248
 Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly Ala
 405 410 415

acc ttt cac ctg ttg cag tct ctg ggc atc tcc cat tga 1287
 Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
 420 425

<210> 5

<211> 428

<212> PRT

<213> Homo sapiens

<400> 5

Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser Cys
 1 5 10 15

Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu Gly
 20 25 30

Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr Leu
 35 40 45

Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His Val
 50 55 60

Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu Gly
 65 70 75 80

Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln
 85 90 95

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

Asp Ser Ile Pro Arg Ser His Trp Lys Lys Thr Pro Val Val Leu Lys
 115 120 125

Ala Thr Ala Gly Leu Arg Leu Leu Pro Glu His Lys Ala Lys Ala Leu
 130 135 140

Leu Phe Glu Val Lys Glu Ile Phe Arg Lys Ser Pro Phe Leu Val Pro
 145 150 155 160

Lys Gly Ser Val Ser Ile Met Asp Gly Ser Asp Glu Gly Ile Leu Ala
 165 170 175

- 9 -

Trp Val Thr Val Asn Phe Leu Thr Gly Gln Leu His Gly His Arg Gln
 180 185 190
 Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile Thr
 195 200 205
 Phe Leu Pro Gln Phe Glu Lys Thr Leu Glu Gln Thr Pro Arg Gly Tyr
 210 215 220
 Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys Leu Tyr Thr His
 225 230 235 240
 Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr Leu Gly
 245 250 255
 Ala Leu Glu Thr Glu Gly Thr Asp Gly His Thr Phe Arg Ser Ala Cys
 260 265 270
 Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr
 275 280 285
 Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys Tyr
 290 295 300
 Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu Glu
 305 310 315 320
 Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg Ala
 325 330 335
 Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys Val
 340 345 350
 Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu Asn
 355 360 365
 Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile Thr
 370 375 380
 Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu Gln
 385 390 395 400
 Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly Ala
 405 410 415
 Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
 420 425

<210> 6

<211> 1287

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

- 10 -

<222> (1)..(1284)

<400> 6

atg gcc act tct tgg ggc aca gtc ttt ttc atg ctg gtg gta tcc tgt	48
Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser Cys	
1 5 10 15	
ggt tgc agc gct gtc tcc cac agg aac cag cag act tgg ttt gag ggt	96
Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu Gly	
20 25 30	
atc ttc ctg tct tcc atg tgc ccc atc aat gtc agc gcc agc acc ttg	144
Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr Leu	
35 40 45	
tat gga att atg ttt gat gca ggg agc act gga act cga att cat gtt	192
Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His Val	
50 55 60	
tac acc ttt gtg cag aaa atg cca gga cag ctt cca att cta gaa ggg	240
Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu Gly	
65 70 75 80	
gaa gtt ttt gat tct gtg aag cca gga ctt tct gct ttt gta gat caa	288
Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln	
85 90 95	
cct aag cag ggt gct gag acc gtt caa ggg ctc tta gag gtg gcc aaa	336
Pro Lys Gln Gly Ala Glu Thr Val Gln Gly Leu Leu Glu Val Ala Lys	
100 105 110	
gac tca atc ccc cga agt cac tgg aaa aag acc cca gtg gtc cta aag	384
Asp Ser Ile Pro Arg Ser His Trp Lys Lys Thr Pro Val Val Leu Lys	
115 120 125	
gca aca gca gga cta cgc tta ctg cca gaa cac aaa gcc aag gct ctg	432
Ala Thr Ala Gly Leu Arg Leu Leu Pro Glu His Lys Ala Lys Ala Leu	
130 135 140	
ctc ttt gag gta aag gag atc ttc agg aag tca cct ttc ctg gta cca	480
Leu Phe Glu Val Lys Glu Ile Phe Arg Lys Ser Pro Phe Leu Val Pro	
145 150 155 160	
aag ggc agt gtt agc atc atg act gga caa gac gaa ggc ata ttc gct	528
Lys Gly Ser Val Ser Ile Met Thr Gly Gln Asp Glu Gly Ile Phe Ala	
165 170 175	
tgg gtt act gtg aat ttt ctg aca ggt cag ctg cat ggc cac aga cag	576
Trp Val Thr Val Asn Phe Leu Thr Gly Gln Leu His Gly His Arg Gln	
180 185 190	
gag act gtg ggg acc ttg gac cta ggg gga gcc tcc acc caa atc acg	624
Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile Thr	
195 200 205	

- 11 -

ttc ctg ccc cag ttt gag aaa act ctg gaa caa act cct agg ggc tac	672
Phe Leu Pro Gln Phe Glu Lys Thr Leu Glu Gln Thr Pro Arg Gly Tyr	
210 215 220	
ctc act tcc ttt gag atg ttt aac agc act tat aag ctc tat aca cat	720
Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys Leu Tyr Thr His	
225 230 235 240	
agt tac ctg gga ttt gga ttg aaa gct gca aga cta gca acc ctg gga	768
Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr Leu Gly	
245 250 255	
gcc ctg gag aca gaa ggg act gat ggg cac act ttc cgg agt gcc tgt	816
Ala Leu Glu Thr Glu Gly Thr Asp Gly His Thr Phe Arg Ser Ala Cys	
260 265 270	
tta ccg aga tgg ttg gaa gca gag tgg atc ttt ggg ggt gtg aaa tac	864
Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr	
275 280 285	
cag tat ggt ggc aac caa gaa ggg gag gtg ggc ttt gag ccc tgc tat	912
Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys Tyr	
290 295 300	
gcc gaa gtg ctg agg gtg gta cga gga aaa ctt cac cag cca gag gag	960
Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu Glu	
305 310 315 320	
gtc cag aga ggt tcc ttc tat gct ttc tct tac tat tat gac cga gct	1008
Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg Ala	
325 330 335	
gtt gac aca gac atg att gat tat gaa aag ggg ggt att tta aaa gtt	1056
Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys Val	
340 345 350	
gaa gat ttt gaa aga aaa gcc agg gaa gtg tgt gat aac ttg gaa aac	1104
Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu Asn	
355 360 365	
ttc acc tca ggc agt cct ttc ctg tgc atg gat ctc agc tac atc aca	1152
Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile Thr	
370 375 380	
gcc ctg tta aag gat ggc ttt ggc ttt gca gac agc aca gtc tta cag	1200
Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu Gln	
385 390 395 400	
ctc aca aag aaa gtg aac aac ata gag acg ggc tgg gcc ttg ggg gcc	1248
Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly Ala	
405 410 415	
acc ttt cac ctg ttg cag tct ctg ggc atc tcc cat tga	1287
Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His	
420 425	

- 12 -

<210> 7

<211> 428

<212> PRT

<213> Homo sapiens

<400> 7

```

Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val Val Ser Cys
  1              5              10              15

Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp Phe Glu Gly
          20              25              30

Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala Ser Thr Leu
      35              40              45

Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg Ile His Val
      50              55              60

Tyr Thr Phe Val Gln Lys Met Pro Gly Gln Leu Pro Ile Leu Glu Gly
      65              70              75              80

Glu Val Phe Asp Ser Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln
          85              90              95

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

Asp Ser Ile Pro Arg Ser His Trp Lys Lys Thr Pro Val Val Leu Lys
      115              120              125

Ala Thr Ala Gly Leu Arg Leu Leu Pro Glu His Lys Ala Lys Ala Leu
      130              135              140

Leu Phe Glu Val Lys Glu Ile Phe Arg Lys Ser Pro Phe Leu Val Pro
      145              150              155              160

Lys Gly Ser Val Ser Ile Met Thr Gly Gln Asp Glu Gly Ile Phe Ala
          165              170              175

Trp Val Thr Val Asn Phe Leu Thr Gly Gln Leu His Gly His Arg Gln
      180              185              190

Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile Thr
      195              200              205

Phe Leu Pro Gln Phe Glu Lys Thr Leu Glu Gln Thr Pro Arg Gly Tyr
      210              215              220

Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys Leu Tyr Thr His
      225              230              235              240

Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr Leu Gly
          245              250              255

```

- 13 -

Ala Leu Glu Thr Glu Gly Thr Asp Gly His Thr Phe Arg Ser Ala Cys
260 265 270

Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr
275 280 285

Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys Tyr
290 295 300

Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu Glu
305 310 315 320

Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg Ala
325 330 335

Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys Val
340 345 350

Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu Asn
355 360 365

Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile Thr
370 375 380

Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu Gln
385 390 395 400

Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu Gly Ala
405 410 415

Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
420 425

<210> 8
<211> 9365
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (72) .. (287)

<220>
<221> CDS
<222> (1280) .. (1579)

<220>
<221> CDS
<222> (1819) .. (1854)

<220>
<221> CDS
<222> (2466) .. (2555)

- 14 -

<220>
<221> CDS
<222> (2863)..(2940)

<220>
<221> CDS
<222> (3887)..(3952)

<220>
<221> CDS
<222> (4896)..(4994)

<220>
<221> CDS
<222> (5846)..(5986)

<220>
<221> CDS
<222> (6965)..(7138)

<220>
<221> CDS
<222> (8556)..(8639)

<220>
<221> misc_feature
<222> (3409)
<223> n = adenine or guanine or cytosine or thymine

<220>
<221> misc_feature
<222> (9214)
<223> n = adenine or guanine or cytosine or thymine

<220>
<221> misc_feature
<222> (9303)
<223> n = adenine or guanine or cytosine or thymine

<220>
<221> misc_feature
<222> (9311)
<223> n = adenine or guanine or cytosine or thymine

<400> 8
gcctctgcag gtgtgcgagc aggattgctt ctgcaacaaa agcctccacc cagccacatc 60

ttgggaaaag a atg gcc act tct tgg ggc aca gtc ttt ttc atg ctg gtg 110
Met Ala Thr Ser Trp Gly Thr Val Phe Phe Met Leu Val
1 5 10

gta tcc tgt gtt tgc agc gct gtc tcc cac agg aac cag cag act tgg 158
Val Ser Cys Val Cys Ser Ala Val Ser His Arg Asn Gln Gln Thr Trp
15 20 25

- 15 -

ttt gag ggt atc ttc ctg tct tcc atg tgc ccc atc aat gtc agc gcc 206
 Phe Glu Gly Ile Phe Leu Ser Ser Met Cys Pro Ile Asn Val Ser Ala
 30 35 40 45

agc acc ttg tat gga att atg ttt gat gca ggg agc act gga act cga 254
 Ser Thr Leu Tyr Gly Ile Met Phe Asp Ala Gly Ser Thr Gly Thr Arg
 50 55 60

att cat gtt tac acc ttt gtg cag aaa atg cca g gtaagtgc actgggrccc 307
 Ile His Val Tyr Thr Phe Val Gln Lys Met Pro
 65 70

ttagtagagt ctgtaaatcc acacttttagc atctcctccc agaaacaaat atgctgagag 367

tttattatgt gaattacaga atctcacacc tagtggatgt ctttcttcag agaactttgg 427

actacaattg aacatgtggg ttattttattt atttttattt atttgttttg tttttatttt 487

ttaacttttt ttttgagaca aggtcttget ttgttgcccg gtctgtagtg cagtggcatg 547

atgacacatc actgcaacct tgacctctg ggctcaagca gtccttcac ctcagcccc 607

tgagttgttg agactacagg cttgtgccac catgcccagc tcatttttaa attttttat 667

agagacctgc tcagactggc ctcaaactcc taggctcaat tgatcctccc acctcagcct 727

cccaaagtac tgggattata ggtgtaagtc accatgcttg gccagaacac atggcttaat 787

tcaatgtgaa attagaagag agctgggctg tctgtagtct gaaacccatg tgttcaaaaa 847

gaatagttat aatttgttct tcctctttaa acatgggata ctccagggat ccataatatt 907

cagaatatgg ggagtgggtt tgggagaagg atcacatgag aatttcactg ccaccttgg 967

acatgaggct aggaatccct gaagattaac ttttctgaa tttgtcagt tttttctc 1027

aggtcactta tggagcctgg ggaaagggtg aggagttagg tgtccaccag agaaatggta 1087

gcagaaatgg accctcagag gttgctctag tccttcttct cagtactcct gcaagacatt 1147

cctcacaact aggatcattg ggtaacttc agggaagtca taggaaaact tacagagaca 1207

gagcccagca tctgaagcag cctaactttt ggtaaccagc tctctcttct gttttgttcc 1267

atgracaaaa tag ga cag ctt cca att cta gaa ggg gaa gtt ttt gat tct 1318
 Gly Gln Leu Pro Ile Leu Glu Gly Glu Val Phe Asp Ser
 75 80 85

gtg aag cca gga ctt tct gct ttt gta gat caa cct aag cag ggt gct 1366
 Val Lys Pro Gly Leu Ser Ala Phe Val Asp Gln Pro Lys Gln Gly Ala
 90 95 100

gag acc gtt caa ggg ctc tta gag gtg gcc aaa gac tca atc ccc cga 1414
 Glu Thr Val Gln Gly Leu Leu Glu Val Ala Lys Asp Ser Ile Pro Arg
 105 110 115

- 16 -

```

agt cac tgg aaa aag acc cca gtg gtc cta aag gca aca gca gga cta 1462
Ser His Trp Lys Lys Thr Pro Val Val Leu Lys Ala Thr Ala Gly Leu
      120                      125                      130

cgc tta ctg cca gaa cac aaa gcc aag gct ctg ctc ttt gag gta aag 1510
Arg Leu Leu Pro Glu His Lys Ala Lys Ala Leu Leu Phe Glu Val Lys
      135                      140                      145

gag atc ttc agg aag tca cct ttc ctg gta cca aag ggc agt gtt agc 1558
Glu Ile Phe Arg Lys Ser Pro Phe Leu Val Pro Lys Gly Ser Val Ser
      150                      155                      160                      165

atc atg gat gga tcc gac gaa g gtgggagag gtgttgatat gcgttccagg 1609
Ile Met Asp Gly Ser Asp Glu
      170

gggagagggg caggatcagt gaaagatcta actaaaggaa ctggggccag gaataaacag 1669

aaggaatgag atagcaggaa atagaagaca gggagaaggg aacatgtgct ctagacatgg 1729

aathtagaga ggaaaaaaaa aaaacaaggt tggggccagg aaagagaaaa aatgctctgg 1789

gatctaatacc ttgtctttct ttcttttttag gc ata tta gct tgg gtt act gtg 1842
                        Gly Ile Leu Ala Trp Val Thr Val
                        175                      180

aat ttt ctg aca g gtaatacat cctcaagttt atctttagag cttaactagc 1894
Asn Phe Leu Thr

ttttacatgc atagtcagag gagtaaaagc ctcttctttc attctgtatt gtttcttctt 1954

ctttaaaaaa ggaaaagagg ctgggtgtgg cagttcatgc ctgttaattc cagcgctttg 2014

ggaggctgag ttgggcagat cacttgaggc caggagttca agaccagcct ggccaacatg 2074

gcgaaactcc gtctctacca aaaatacaaa aatagctggg catggtggtg tgtacctgta 2134

gtcccagcta ctcaggaggc tggagaatca cttgaaccca ggaggcagag gttgcagtga 2194

gctgagagcc gagattgcgc cactgcactc caggctggat gatagagcaa gactctgtct 2254

ccaaaaaggc cttccaaaaa aaaaaaaaaac acctgccttg aaggcctctg ctgcaacaag 2314

agtccttccg agttgacatt cacctgcagc cttggggctg gggagcagtg gagtatatat 2374

ggaatacctt cagtgtatga taagagcaag agagacaagt gttgggctgc ccaggatgtc 2434

gaggctattt agagctggct ctcatttgac ag gt cag ctg cat ggc cac aga 2486
                        Gly Gln Leu His Gly His Arg
                        185                      190

cag gag act gtg ggg acc ttg gac cta ggg gga gcc tcc acc caa atc 2534
Gln Glu Thr Val Gly Thr Leu Asp Leu Gly Gly Ala Ser Thr Gln Ile
      195                      200                      205

```

acg ttc ctg ccc cag ttt gag gtgagtcatt taatgaagat ctggtttagaa 2585
 Thr Phe Leu Pro Gln Phe Glu
 210

gtgcacttgg caggcgtatc atggtgccaa gaaagaggcg ccccatcttc agccagcagc 2645
 tctaccacgc ttaggcagag tcaagtcaat taataactag gtgaatgttc ccttgccatc 2705
 tcaactgttca gaatcccttc gtttcctcaa gcctagttag attagcccct taatctgtct 2765
 tcatctctga ttttttgctg ggagggacgg gtggtggtgt gaacatcttc aggtaattac 2825
 agatcctgaa tagctttttg ctttttctga tttgcag aaa act ctg gaa caa amt 2880
 Lys Thr Leu Glu Gln Xaa
 215 220

cya trg ggc tac ctc act tcc ttt gag atg ttt aac agc act tat aag 2928
 Xaa Xaa Gly Tyr Leu Thr Ser Phe Glu Met Phe Asn Ser Thr Tyr Lys
 225 230 235

ctc tat aca cat ag gtgaggac ggggacaggg aagaagaataatmtwtkttg 2980
 Leu Tyr Thr His
 240

tatgatksty ytamctktss maagcwtkct caaatctstk aytkyatctg attmgcaaaa 3040
 acaaagdctg tgccaattcc ctaaggccta tcaactgaaa cccggwccac ttacaaagcc 3100
 ggaggagcct aagaggcttc tccattcttg gcctcaaaag cattaatata tgacttaaga 3160
 gtcaaaagtt ttggstgggg cagtggcttc atgcctgtaa tccctgcact ttgggaggcc 3220
 gaggtgggtg ggtcacctga ggtcaggcgt ttragaccag cctggcaaac atggtgaaac 3280
 cccgtctyta ctaaaataca aaaattagct ggatatgaca gcgcacacct gtaatcctag 3340
 ctattcagga ggctgaggca ggagaatcat ttgaaccctg gaggcggaga ttgcagttag 3400
 ccgagatcnc accmctgcac ttcagccgga gcgacagagc aagactcagt ctcaaaaaaa 3460
 aaaaaaaaaa gaatcaaaag ctttctgtag ggagaggaca cttcaagaag gctcaggcaa 3520
 agctccttgc cagctccttt gagctggcct tcagaggcttc agaatccagc ctggaatgtg 3580
 atcccagttg gggctaggag ctaagctaaa gagagctttt ctgggaatgg ttcctagwgt 3640
 gggaccctag gaattgtcac tgtctctggc ctttgaatga taactgtggg gaattcttac 3700
 tgcatagcct tgatccaaac tgtgcagaaa ttacccttg ttgaccacag gagatgaata 3760
 tgtcacagac agaacaaggt tttcatcttt ccagagggac acaggaacaa tgttactttt 3820
 gaaagaggta gctttaggct agagaacttc aggaccagca tgaaattagt caatcctgta 3880

- 18 -

ttttacag t tac ctg gga ttt gga ttg aaa gct gca aga cta gca acc 3928
 Ser Tyr Leu Gly Phe Gly Leu Lys Ala Ala Arg Leu Ala Thr
 245 250

ctg gga gcc ctg gag aca gaa g gtttgtctgggt acctgtgctg gggggggatg 3982
 Leu Gly Ala Leu Glu Thr Glu Gly
 255 260

gtgaggggtga cacagatact ccgcttgctt cttcccttcc ttgatagcca ttctatggag 4042

gaaaagatta tgttgaattg ggaggcaaact gttgtataat ggacctata atggcaaact 4102

ccttttctag tttataagtt cagaagtttt gatgtatatt attagccatt tttagaatga 4162

ggtctacttg ttcaggggta acagcctatg tctaggcagc tgaagtgtct gcagaaatcc 4222

caggctttac gaatacatto agcaggagct tgctcaagcc ctgagcttta cattggaggc 4282

acaggaagca gagtctgttc tacatgcagg tggaacaaca gagtaactcc attgatctct 4342

tcacaggtca ggcagaactg gggtcagtc cagtgttgtg atatgaggcr agtaacctat 4402

ctgtgcccc ttcctcacat taaatgagaa tttgcattta aggcactttg tacagtaatc 4462

tgttattggg atgacatcta ttttgcattt cagagtatac aaaacatctt caagtatatatt 4522

taattgaagc ctctcagcaa ccagtgagga aggtagcata gcatttcttt cctgttttta 4582

taaaggggaa agttgctgta kgaaggttyk rgatctcttw ragatgtgat raaagccatg 4642

gaccctctg acaaaaagcac atatgcatga aaatttgctt ctggtttcag ggggttcacc 4702

aacccacaa agcctatctt tgaacctga gttaaggatt cctgtcacag gatgttgtca 4762

tggaattaat ttcataggat ttttaaggccc agcccccatg gtgaytcttt tccacctcac 4822

tggcttcttg cttgccttcc tccctctctc tcacttactt acctcttacc ttgtgccttg 4882

gattctttcag gg act gat ggg cac act ttc cgg agt gcc tgt tta ccg 4931
 Thr Asp Gly His Thr Phe Arg Ser Ala Cys Leu Pro
 265 270

aga tgg ttg gaa gca gag tgg atc ttt ggg ggt gtg aaa tac cag tat 4979
 Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr Gln Tyr
 275 280 285 290

ggt ggc aac caa gaa g gcaagtgat gttttttcac tggtaaagt tacgtttaca 5034
 Gly Gly Asn Gln Glu
 295

atggaagctc tggaaaagtc ccatgggaaa ctttttccag aactcaagag aagcttatct 5094

tgttgagggg asttattcca aagatcttgg catgcctcca aggactaatg tgaagtgaca 5154

attacaggtg tgtgccacca ctcttgtcta atttttttgt agagacgggg ttttgccatg 7948
 ttgccagggc tgggtcttgaa ctcttgacct caagcaatcc acctgccttg gcctcccaaa 8008
 gtgccaggag tacaggcatg agccactgcg cctggcccca tgtttggtta ttattagtgc 8068
 ttaggaagag gcacttgctt acatagtagg agttgagaag cttggtttgt tctttcctac 8128
 ccctagatct attctcacct cctgacctg ctctttctgc cacatctatt atcattacaa 8188
 gttgccttat ctgaaattag tgaatcagaa aataaagcag gggatacttt gtgtagtttc 8248
 aacgttaggg aaagttcaga atactgtctg tctaaactat ctctctagaa ggccctgatgg 8308
 gccacaacct gggccagaag cattcagttc agatatgaga atgggtgggtg taggggcaat 8368
 ggccaatggg ccatggccgg aaggaaattg ttacagagta gtgggaagcc tgcaaagact 8428
 ggcttctgtc cgttttgctt tggtttgccc atgtggatat tctttgcaa tattttctgc 8488
 ccaagagctg tgcttgctag agttggaaac tggatgaaaa ggtgaagact ttttttcttc 8548
 tcaacag ctc aca aag aaa gtg aac aac ata gag acg ggc tgg gcc ttg 8597
 Leu Thr Lys Lys Val Asn Asn Ile Glu Thr Gly Trp Ala Leu
 405 410
 ggg gcc acc ttt cac ctg ttg cag tct ctg ggc atc tcc cat 8639
 Gly Ala Thr Phe His Leu Leu Gln Ser Leu Gly Ile Ser His
 415 420 425
 tgaggccacg tacttccttg gagacctgca tttgccaaca cctttttaag gggaggagag 8699
 agcacttagt ttctgaacta gtctggggac atcctggact tgagcctaga gatttaggtt 8759
 taattaatth tacacatcta atagtgaact gctgcctaac cactcaagag tacacagctg 8819
 gcaccagagc atcacagaga gccctgtgag ccaaaaagta tagttttgga acttaacctt 8879
 ggagtgagag cccagggaca ggtccctgga aaccaaagaa aaatcgcat tcaacccttt 8939
 gagtgcctca ttccactgaa tatttaaatt ttctcttaa atgggaaact gacttattgc 8999
 aatcccaaga cccatcaata tcagtatttt ttctctcct atacagggcc ctgcccaccc 9059
 ttatctgcac ccacctcccc tgaaaaagag agaaaaaaaa aaamccbgtt tttgctttcc 9119
 wtgtwtaaty camcgacmca aaakgggacc atgtcaaaat ctgtwtgatc ctattytggg 9179
 ttascyccaa tcagccagct garagccttc ctaantttta wtaggatgar agagtaccyc 9239
 ctaactgtgc ataaattcag ccttaaaaaa aaaggcaccc gggctttggg gacatgtttg 9299
 ggangggggg gntgcctcat ataccacct ttggtttaat aacattttat cagcactttg 9359
 ggataa 9365

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 9
gctacctcac ttcctttgag 20

<210> 10
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 10
ctggctggtg aagttttcct c 21

<210> 11
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 11
gcaggtctcc aaggaagtac g 21

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 12
gtgagtgtc cctgcatcta acataattcc 30

<210> 13
<211> 45
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 13

gatgcagggga gcactcacac tagtattcat gtttacacct ttgtg

45

<210> 14

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 14

gcgtagtcct gctgttgccc ctaggtacac tggggtcttt ttcc

44

<210> 15

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 15

gcaacagcag gactacgctt actgccagaa c

31

<210> 16

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 16

cccaagcgaa tatgccttcg tcttgtccag tcatgatgct aacactgc

48

<210> 17

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 17

cgaaggcata ttcgcttggg ttactgtg

28

<210> 18

- 24 -

<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 18
cttccttcac tgggaattca gg 22

<210> 19
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 19
ctgtttaccg agatggttg aagc 24

<210> 20
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 20
ttaaagcttg ggaaaagaat ggccacttc 29

<210> 21
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 21
agactcgagg tggctcaatg ggagatgcc 29

<210> 22
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 22

- 25 -

gcgctgtctc ccacagagga tcgcatcacc atcaccatca caaccagcag acttggtt 58

<210> 23

<211> 58

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 23

aaccaagtct gctggttgtg atggtgatgg tgatgcgatc ctctgtggga gacagcgc 58

<160> 2

<170> PatentIn Ver. 2.0

<210> 24

<211> 1601

<212> DNA

<213> Homo sapiens

<400> 24

```

gcgggctgcc gcgcaagggg ggcgcgcgcg cgttttcctt gttcctgggc aacaaagaaa 60
tgtggagtgt cttggctgaa tcctcataca gacaagatca ttatgggtgt gttaggttga 120
aaaagtgata taataaagga accaaggaga aaattcagaa ggaaagaaaa aattgcctct 180
gcaggtgtgc gagcaggatt gcttctgcaa caaaagcctc caccagcca catcttggga 240
aaagaatggc cacttcttgg ggcacagtct ttttcattgt ggtggtatcc tgtgtttgca 300
gcgctgtctc ccacaggaac cagcagactt ggtttgaggg tatcttcttg tcttccatgt 360
gccccatcaa tgtcagcgcc agcaccttgt atggaattat gtttgatgca gggagcactg 420
gaactcgaat tcatgtttac acctttgtgc agaaaatgcc aggacagctt ccaattctag 480
aaggggaagt ttttgattct gtgaagccag gactttctgc tttttagatg caacctaacg 540
aggggtgctg gaccgttcaa gggctcttag aggtggccaa agactcaatc ccccgaaagc 600
actggaaaaa gacccagtg gtcctaaagg caacagcagg actacgctta ctgccagaac 660
acaaagccaa ggctctgctc tttgaggtaa aggagatctt caggagtgca cctttcctgg 720
taccaaaagg cagtgttagc atcatggatg gatccgacga aggcataatta gcttgggtta 780
ctgtgaatct tctgacaggt cagctgcatg gccacagaca ggagactgtg gggaccttgg 840
acctaggggg agcctccacc caaatcacgt tcctgcccc a gtttgagaaa actctggaac 900
aaactcctag gggctacctc acttcctttg agatgtttaa cagcacttat aagctctata 960
cacatagtta cctgggattt ggattgaaag ctgcaagact agcaaccctg ggagccctgg 1020
agacagaagg gactgatggg cacactttcc ggagtgcctg tttaccgaga tgggtggaag 1080
cagagtggat ctttgggggt gtgaaatacc agtatggtgg caaccaagaa ggggaggtgg 1140
gctttgagcc ctgctatgcc gaagtgtgta ggggtgtacg aggaaaactt caccagccag 1200
aggaggtcca gagaggttcc ttctatgctt tctcttacta ttatgaccga gctgttgaca 1260
cagacatgat tgattatgaa aaggggggta ttttaaaagt tgaagatctt gaaagaaaag 1320
ccagggaagt gtgtgataac ttggaaaact tcacctcagg cagtcctttc ctgtgcatgg 1380
atctcagcta catcacagcc ctgttaaagg atggctttgg ctttgcagac agcacagtct 1440
tacaggctgc cgtactgagg tgatgggcca agctggagat atccccaag cccatgttga 1500
caccctgtcc tgcaagcgga tggactctgt gggctgcac cctaagaata aagcagagtt 1560
caggtgtgac ctctggcagc aaaaaaaaaa aaaaaaaaaa a 1601

```

<210> 25

<211> 405

<212> PRT

- 26 -

<213> Homo sapiens

<400> 25

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

- 27 -

Leu Pro Arg Trp Leu Glu Ala Glu Trp Ile Phe Gly Gly Val Lys Tyr
275 280 285

Gln Tyr Gly Gly Asn Gln Glu Gly Glu Val Gly Phe Glu Pro Cys Tyr
290 295 300

Ala Glu Val Leu Arg Val Val Arg Gly Lys Leu His Gln Pro Glu Glu
305 310 315 320

Val Gln Arg Gly Ser Phe Tyr Ala Phe Ser Tyr Tyr Tyr Asp Arg Ala
325 330 335

Val Asp Thr Asp Met Ile Asp Tyr Glu Lys Gly Gly Ile Leu Lys Val
340 345 350

Glu Asp Phe Glu Arg Lys Ala Arg Glu Val Cys Asp Asn Leu Glu Asn
355 360 365

Phe Thr Ser Gly Ser Pro Phe Leu Cys Met Asp Leu Ser Tyr Ile Thr
370 375 380

Ala Leu Leu Lys Asp Gly Phe Gly Phe Ala Asp Ser Thr Val Leu Gln
385 390 395 400

Ala Ala Val Leu Arg
405