



(72) GOODMAN, COREY, US

(72) KID, THOMAS, US

(72) TEAR, GUY, GB

(72) RUSSELL, CLAIRE, GB

(72) MITCHELL, KEVIN J., US

(71) THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, US

(51) Int.Cl.<sup>6</sup> C12N 15/18, C07K 14/475, A61K 38/18

(30) 1997/11/14 (60/065,543) US

(54) **PROCEDES PERMETTANT DE MODULER UNE FONCTION  
DES CELLULES NERVEUSES**

(54) **METHODS FOR MODULATING NERVE CELL FUNCTION**

(57) La quantité de Robo actif exprimé sur une cellule est modifiée par modulation de la quantité efficace d'un polypeptide Comm mis en contact avec ladite cellule, la quantité de Robo actif exprimé étant modulée inversement à la modulation du polypeptide Comm. Selon un mode de réalisation particulier, le polypeptide Comm est fourni à la cellule, de manière exogène, dans une composition pharmaceutiquement acceptable. Selon un autre mode de réalisation, l'invention concerne également des procédés de criblage des agents qui modulent les interactions Robo-Comm. Ces procédés impliquent généralement la formation d'un mélange d'une cellule exprimant Robo, d'un polypeptide Comm et d'un agent candidat, et la détermination de l'effet de l'agent sur la quantité de Robo exprimé par la cellule.

(57) The amount of active Robo expressed on a cell is modified by modulating the effective amount of a Comm polypeptide in contact with the cell, whereby the amount of expressed active Robo is modulated inversely with the modulation of the effective amount of the Comm polypeptide in contact with the cell. In a particular embodiment, the Comm polypeptide is provided to the cell exogenously in a pharmaceutically acceptable composition. In another aspect, the invention provides methods of screening for agents which modulate Robo-Comm interactions. These methods generally involve forming a mixture of a Robo-expressing cell, a Comm polypeptide and a candidate agent, and determining the effect of the agent on the amount of Robo expressed by the cell.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : C12N 15/18, C07K 14/475, A61K 38/18</p>	<p>A1</p>	<p>(11) International Publication Number: <b>WO 99/25833</b> (43) International Publication Date: 27 May 1999 (27.05.99)</p>
<p>(21) International Application Number: PCT/US98/24327 (22) International Filing Date: 13 November 1998 (13.11.98) (30) Priority Data: 60/065,543 14 November 1997 (14.11.97) US (71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 5th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US). (72) Inventors: GOODMAN, Corey; University of California, Berkeley, Life Sciences Addition #519, Berkeley, CA 94720 (US). KID, Thomas; University of California, Berkeley, Life Sciences Addition #519, Berkeley, CA 94720 (US). TEAR, Guy; Imperial College, Dept. of Biochemistry, Exhibition Road, London SW7 2AZ (GB). RUSSELL, Claire; Imperial College, Dept. of Biochemistry, Exhibition Road, London SW7 2AZ (GB). MITCHELL, Kevin, J.; University of California, Berkeley, Life Sciences Addition #519, Berkeley, CA 94720 (US). (74) Agent: OSMAN, Richard, Aron; Science &amp; Technology Law Group, 75 Denise Drive, Hillsborough, CA 94010 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, IL, 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, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, 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).</p> <p><b>Published</b> <i>With international search report.</i></p>
<p>(54) Title: METHODS FOR MODULATING NERVE CELL FUNCTION</p>		
<p>(57) Abstract</p> <p>The amount of active Robo expressed on a cell is modified by modulating the effective amount of a Comm polypeptide in contact with the cell, whereby the amount of expressed active Robo is modulated inversely with the modulation of the effective amount of the Comm polypeptide in contact with the cell. In a particular embodiment, the Comm polypeptide is provided to the cell exogenously in a pharmaceutically acceptable composition. In another aspect, the invention provides methods of screening for agents which modulate Robo-Comm interactions. These methods generally involve forming a mixture of a Robo-expressing cell, a Comm polypeptide and a candidate agent, and determining the effect of the agent on the amount of Robo expressed by the cell.</p>		

*Methods for Modulating Nerve Cell Function*

5 The research carried out in the subject application was supported in part by NIH grant NS18366. The government may have rights in any patent issuing on this application.

## INTRODUCTION

### Field of the Invention

10 The field of this invention is methods for modulating nerve cell function.

### Background

15 In the developing CNS, most growth cones confront the midline at one or multiple times during their journey and make the decision of whether to cross or not to cross. This decision is not a static one but rather changes according to the growth cone's history. For example, in the *Drosophila* ventral nerve cord, about 10% of the interneurons project their axons only on their own side, in some cases extending near the midline without crossing it. The other 90% of the interneurons first project their axons across the midline and then turn to project longitudinally on the other side, often extending near the midline. These growth cones, having crossed the midline once, never cross it again, in spite of their close proximity to the midline and the many commissural axons crossing it. This decision to cross or not to cross is not unique to *Drosophila* but is common to a variety of midline structures in all bilaterally symmetric nervous systems.

20 What midline signals and growth cone receptors control whether growth cones do or do not cross the midline? After crossing once, what mechanism prevents these growth cones from crossing again? A related issue concerns the nature of the midline as an intermediate target. If so many growth cones find the midline such an attractive structure, why do they cross over it rather than linger? Why do they leave the midline?

30 One approach to find the genes encoding the components of such a system is to screen for mutations in which either too many or too few axons cross the midline. Such a large-scale mutant screen was previously conducted in *Drosophila*, and led to the identification of two key genes: *commissureless* (*comm*) and *roundabout* (*robo*) (Seeger et al., 1993; reviewed by Tear

et al., 1993). In *comm* mutant embryos, commissural growth cones initially orient toward the midline but then fail to cross it and instead recoil and extend on their own side. *robo* mutant embryos, on the other hand, display the opposite phenotype in that too many axons cross the midline; many growth cones that normally extend only on their own side instead now project  
5 across the midline and axons that normally cross the midline only once instead appear to cross and recross multiple times (Seeger et al, 1993; present disclosure). Double mutants of *comm* and *robo* display a *robo*-like phenotype.

How do *comm* and *robo* function to control midline crossing? Neither the initial paper on these genes (Seeger et al., 1993) nor the cloning of *comm* (Tear et al., 1996) resolved this  
10 question. *comm* encodes a novel surface protein expressed on midline cells. In fact, the *comm* paper (Tear et al., 1996) ended with the hope that future work would "... help shed some light on the enigmatic function of Comm."

A copending application (*Robo, A Novel Family of Polypeptides and Nucleic Acids*, by inventors: Corey S. Goodman, Thomas Kidd, Kevin J. Mitchell and Guy Tear, and filed  
15 herewith) discloses the cloning and characterization of *robo* in various species including *Drosophila*. *robo* encodes a new class of guidance receptor with 5 immunoglobulin (Ig) domains, 3 fibronectin type III domains, a transmembrane domain, and a long cytoplasmic domain. Robo defines a new subfamily of Ig superfamily proteins that is highly conserved from fruit flies to mammals. The Robo ectodomains, and in particular the first two Ig  
20 domains, are highly conserved from fruit fly to human, while the cytoplasmic domains are more divergent. Nevertheless, the cytoplasmic domains contain three highly conserved short proline-rich motifs which may represent binding sites for SH3 or other binding domains in linker or signaling molecules.

For those axons that never cross the midline, Robo is expressed on their growth cones  
25 from the outset; for the majority of axons that do cross the midline, Robo is expressed at high levels on their growth cones only after they cross the midline. Transgenic rescue experiments in *Drosophila* reveal that Robo can function in a cell autonomous fashion, consistent with it functioning as a receptor. Thus, in *Drosophila*, Robo appears to function as the gatekeeper controlling midline crossing; growth cones expressing high levels of Robo are prevented from  
30 crossing the midline. Robo proteins in mammals function in a similar manner in controlling axon guidance.

Here we disclose ectopic and overexpression studies revealing that Comm down-regulates Robo expression, demonstrating that Comm functions to suppress the Robo-mediated midline repulsion. These results show that the levels of Comm at the midline and Robo on growth cones are tightly intertwined and dynamically regulated to assure that only certain growth cones cross the midline, that those growth cones that cross do not linger at the midline, and that once they cross they never do so again.

#### SUMMARY OF THE INVENTION

The invention provides methods and compositions for modulating the amount of active Robo expressed on a cell. The general method involves modulating the effective amount of a Comm polypeptide in contact with a cell expressing an amount of active Robo polypeptide, whereby the amount of expressed active Robo is modulated inversely with the modulation of the effective amount of the Comm polypeptide in contact with the cell. For example, where the effective amount of the Comm polypeptide is increased, the amount of expressed Robo is decreased. The Robo polypeptide is preferably a human, mouse, *C. elegans* or *Drosophila* Robo I or II sequence or a polypeptide domain thereof having a Robo-specific activity, and the Comm polypeptide specifically modulates Robo expression and (a) comprises SEQ ID NO:14 or a deletion mutant thereof which specifically modulates Robo expression and/or (b) is encoded by a nucleic acid comprising SEQ ID NO:13 or a nucleic acid which hybridizes with SEQ ID NO:13, preferably under stringent conditions. In a particular embodiment, the Comm polypeptide is provided to the cell exogenously in a pharmaceutically acceptable composition. In another aspect, the invention provides methods of screening for agents which modulate Robo-Comm interactions. These methods generally involve forming a mixture of a Robo-expressing cell, a Comm polypeptide and a candidate agent, and determining the effect of the agent on the amount of Robo expressed by the cell.

#### DETAILED DESCRIPTION OF THE INVENTION

The subject methods involve modulating the effective amount of a Comm polypeptide in contact with a cell expressing an amount of active Robo polypeptide, whereby the amount of expressed active Robo is modified inversely with the modulation of the effective amount of the Comm polypeptide in contact with the cell. Robo expression is found to regulate a wide

variety of cell functions, including cell-cell interactions, cell mobility, morphology, etc. Accordingly, the invention provides methods for modulating targeted cell function comprising the step of modulating Robo expression by contacting the cell with a Comm polypeptide.

The targeted Robo polypeptide is generally naturally expressed on the targeted cells.

5 The nucleotide sequences of exemplary natural cDNAs encoding *drosophila* 1, *drosophila* 2, *C. elegans*, human 1, human 2 and mouse 1 Robo polypeptides are shown as SEQ ID NOS:1, 3, 5, 7, 9 and 11, respectively, and the full conceptual translates are shown as SEQ ID NOS:2, 4, 6, 8, 10 or 12. The targeted Robo polypeptides comprise at least a functional domain of SEQ ID NOS:2, 4, 6, 8, 10 and 12, which domain has Robo-specific amino acid sequence and  
 10 binding specificity or function. Preferred Robo domains comprise at least 8, preferably at least 16, more preferably at least 32, most preferably at least 64 consecutive residues of one of these SEQ ID NOS. In a particular embodiment, the domains comprise one or more structural/functional Robo immunoglobulin, fibronectin or cytoplasmic motif domains described herein. The subject domains provide Robo-specific antigens and/or immunogens,  
 15 especially when coupled to carrier proteins. For example, peptides corresponding to Robo- and human Robo-specific domains are covalently coupled to keyhole limpet antigen (KLH) and the conjugate is emulsified in Freund's complete adjuvant. Laboratory rabbits are immunized according to conventional protocol and bled. The presence of Robo-specific antibodies is assayed by solid phase immunosorbant assays using immobilized Robo  
 20 polypeptides of SEQ ID NO:2, 4, 6, 8, 10 or 12. Generic Robo-specific peptides are readily apparent as conserved regions in the aligned Robo polypeptide sequences of Table 1.

Table 1. Sequence Alignment of Robo Family Members: The complete amino acid alignment of the predicted Robo proteins encoded by *drosophila robo 1* (D1, SEQ ID NO:2) and Human  
 25 *robo 1* (H1, SEQ ID NO:8) are shown. The extracellular domain of *C.elegans robo* (CE, SEQ ID NO:6; Sax-3; Zallen et al., 1997), the extracellular domain of *Drosophila robo 2* (D2, SEQ ID NO:4), and partial sequence of Human *robo 2* (H2, SEQ ID NO:10) are also aligned. The D2 sequence was predicted by the gene-finder program Grail. The position of immunoglobulin domains (Ig), fibronectin domains (FN), the transmembrane domain (TM),  
 30 and conserved cytoplasmic motifs are indicated. The extracellular domain of rat *robo 1* is nearly identical to H1.

WO 99/25833

PCT/US98/24327

	mH.....PMHpENHAIaRSTSTTNNPSrsRSSRMWllpAWLLLVLVASNGLP	47	D1
	m.FNRKTLlCTi.llVlQA.....vIrsFCEDASNlA.....	30	CE
	mKWKHVpFlVMiSlSlSlsNHLFLaQLIPDPEDvErG.NDHGTPIpTSDNDDNSLGYTGS	59	H1
5			
	>IG #1		
	AVrGQYQSpriiehpTdlvvKknepatlnckVegKpEptiewfkdgepvStn..EKKshr	105	D1
	GENpriiehpMdTTvPknDpFtFncQaegNptptiQwfkdgRELkt...dTGshr		D2
	.....pViiehpIdVvvsRgSpatlnCGaK.PStAKiTwykdgQpvItnkeQVNsHR	81	CE
10	RLrQEDFPprievehpSdlIvskgepatlnckaegRptptiewykGgeRvEtDkDdPRshr	119	H1
	>IG #2		
	VQFKDgAlffYrimQgkkeQ..dGgEywcvaknRVgQavsRHaslqIavlrdfrvepKd	163	D1
	iMlpAgGlfflkvIhSrReS..dagTywcEakneFgVaRsrnaTlqvavlrdEfrLepAN		D2
15	iVlDTgslfLlkvNSgkNGKDSdagAyYcvaSneHgeVKsNEGSllKLaMlrEdfrvRpRT	141	CE
	MLlpSgslfflriVhgrkSRP.dEgVyVcvaRnYLgeavshnaslEvaIlrdfrQNPsd	178	H1
	trvaKgeTallecgppKgIpeptLIwIkdgVplddLKAMsFGASSrVrivdggnlLiSNv	223	D1
	trvaQgeValmecgAprgSppeQiswrkNgQTlNL.....VGNKririvdggnlAiQEA		D2
20	vQALGgeMavlecSpprgFpewVvswrkdDKElRI.QDmP.....rYTLHSDgnliiDPv	195	CE
	vMvaVgePavmecQpprgHpeptiswKkdgSpldd.....KDEri.TIRggKlMiTYT	230	H1
	>IG #3		
	EPIdEgNyKcIaQnLvgtresSYaKlIvQvKpYfMkepKdqVMLYgQTaTfHcSvggdpP	283	D1
25	rQsdDgRyqcVKnVvgtresATaFlKvHvrpFLIRGpQnqtAVvgSsvVfQcrIggdpL		D2
	DRsdSgTyqcvaNnmvgerVsNPArIsvFekpKfEQepkdMtvDvgAAvLfDcrvTgdpQ	255	CE
	rKsdAgKyVcvGTnmvgeresEVaElTvLerpSfVkrpSnLAvTvDDsaEfKceARgdpV	290	H1
	pKvlwkk..EEgnIpsrA.....RiLHdEKslEiSNitpTdegTyvceaHnNvg	331	D1
30	pDvlwrrTASGgnmpLRKFSWLHSASGRVHvl.EdrslkLDDvtLEdmgeytceaDnAvg		D2
	pQITwkr..KNEPmpvTra.....YiAKdNrGlRiERvQpSdegeyvcYaRnPaG	303	CE
	pTvRwrk..DDgELpKsrY.....Ei.RddHTlkiRKvtAGdmgSytCvaEnMvg	337	H1
	>IG #4		
35	QisARaSlIvhappNfTKrpSnKKvGlNgVvQLPcMaSgnpPpSvfwTkegVSTlMfpn.	388	D1
	GiTaTGiltvhappKfvIrpKnqLvEIgDEvLfecQaNghRpRtLYwsVegNSSllLpGy		D2
	TLeasaHlRvqappSfQTkpAdqSvPaggTAtfecTLVgQpSpaYfwskegQqDlIfpsy	363	CE

WO 99/25833

PCT/US98/24327

KAeasaTltvqEppHfvVkpRdqVvalgrtvtfQceaTgnppaIfwRRegsqnllf.sy 396 H1  
 qIvaQgrtvtfPceTKgnppavfwQkegsqnllfpn. H2

...SsHGrQYvAADgtlQitDvrqedegyyv.cSaFSvvdSsTVrVFlQvSS..vD.... 440 D1  
 5 RDGRMEVTLTPEGRSVlSiARFaredSgKVvTcNalnAvgsVSsrTVVSvDt..QF.... D2  
 VSADGRTK..vsptgtltiEEvrqVdegAyv.cAGMnSagsslskaAlKvttKAvTGNTP 420 CE  
 qpPQsSsrFsvsQtgdltitnvqrsdVgyyi.cqTlnvagsiITkaYlevtd..vIA... 450 H1  
 qpQQPNsrCsvsptgdltitnIqrsdAgyyi.cqalTvagsilAkaQlevtd..vLT... H2

10 >IG #5

erpppiiQIgpAnqtlpKgsVaTlpcratgNpSpRiKwFHdgHAvQA.GNRYSi.iqG.. 496 D1  
 eLpppiiieggpvntlpvKsIVvlpcrTLgTpvvQVswYLdgIpidVqEHERrNLsDA.. D2  
 AKpppTieHgHQnqtlMvgsSaIlpcQaSgKpTpGiswlRdgLpidITd..sri.sqHST 477 CE  
 drpppViRqgpvnqtVavdgtFvlScVatgSpvpTiLwRkdgVLvSTqd..sriK.qLeN 507 H1  
 15 drpppiiLqgpAnqtlavdgtaLcKcKatgDpLpViswlkEgFTFPGRd..PrATiq.eQ H2

>FN #1

SslRVDdlq.lsdSgtytciasGeRgeTswAaTltveKpgs..TSLHraAdpstypAppg 553 D1  
 gAlTiSdlqrHEdEgLytcvasnRNgKsswsGylRLDTptNpNiKfFrapElstypgppg D2  
 20 gslHiAdl.kKPdtgVytciaKneDgestwsaSlTveDHTsN.AqfVrMpdpsNFpsSpT 535 CE  
 gvlqiR.YAkIGdtgRytciasTPsgeatwsayIEvQeFgVp.VqPPrPTdpNLipsAps 565 H1  
 gTlqiKNl.rIsdtgtytcvaTSSsgeaswsaVlDvTeSgAT.i..SKNYdlsDLpgpps H2

TpKvLnvstrtsISlRwAKSqEKPGAVgpIi.gyTVeyfspdlQTgwIVAaHrvGDtQvti 612 D1  
 25 kpqMvEKGEnsvtlsw...TRSNKVggSSLVgyViemfGKNETDgwVAvGTrvQNttFtQ D2  
 QpIivnvtDtEvElHw...NAPSTsgaGpitgyiiQyYspdlgQTwFNIPDYvASTeYri 592 CE  
 kpEvtdvsrnTvtlsw...qpNLNsgaTp.tSyiieafSHASgSswqtvaENvktEtSAi 621 H1  
 kpqvtdvtKnsvtlsw...qpGTPGTLpA.SAyieafSQSVSNswqtvaNHvktLyTV H2

30 >FN #2

SglTpgtsyVflvraentQgisvpsGLsNViktIEA...DfDAASANdlsAarT.llTg 667 D1  
 TglLpgVNYFfliraenSHgLSLpsPMsEpitVGTR...YfNS..gLdlsEarASllsg D2  
 kglkpSHsyMfViraenEkgiGTpsVSsALvttSKPAAQVALSDKNKmdMAIaEKRLTsE 652 CE  
 kglkpnAiylflvraAnAYgisDpsqIsDpvktQDV.....lPTSQgVdHKQVQRE.lGN 675 H1  
 35 RglRpntiylfMvraInPkV.svT.q H2

KSvelIDasAinAsavrleWMLHvSADEkyvegLRiHyK..DaSVPSAQYHSITvMDAsa 725 D1

WO 99/25833

PCT/US98/24327

DvvelSnasvVDstsMKlTwQI...INGkyvegFyVYArQLpNPLNTKyRMLTILNGGGa D2  
 QLIKlEEVKTinstavrIFwKKR..KLEELiDgyyiKWrGPPRTNDNQyVN...vTSpsT 707 CE  
 AvLHlHnPTvLSsssIEVHwT..vDQOSQyiQgyKiLyrPSGaNHGESDWLVFevRTPAK 733 H1

5 >FN #3

esFvvGnlKkytKyeffLTpf...fETiegQpsnskTaltYedvpsappDNIQiGmYn.. 780 D1  
 SsCTiTGlVQytLyeffIVpf...YKsVegKpsnsRIartledvpsEApYgMEALLln.. D2  
 eNYvvSnlMPFtnyeffVIpYHSGVHsiHgapsnsMDVltAeApsLppEDvRiRmlnL. 766 CE  
 NsVviPDlRkGVnyeIKARpf...fNEFQgaDsEikFaKtleEApSappQgvTVSKNDGN 790 H1

10

QtaGWvRwTpppSQHHngNlygykiEVSAGnTM....KvLAnMtLnaTtTsvLlNnltt 835 D1  
 SsaVFLKwkapELKDRHgVlLNyH.vivRgIDtAHNFSRilTnVtIdaASPTLvlAnlte D2  
 .tTLRIswkapKAdGIngIlKgFQiviv.gQAPNNNR....nItTnERAAsvTlFhlvt 819 CE  
 GtaILvswQpppEdTQngMVQEykv.WCLgnEtr....YHInKtVdGStFsvvIPFlVP 844 H1

15 <

gAVysvrLNSftKagDgpysKpISlFMdpTHHVHPpRAHPsGTHDGRHEGqDLTYHNNGN 895 D1  
 gVMYtvGvaaGNnagvgpyCVpATlRldpITKRLDpFINQRDHVND..... D2  
 gMTyKIrvaARsnGvgv.....ShgTSEVIMNqDTlEKHL.AAQqENESFLYgL 868 CE

20 gIRysvEvaaStGagSgvKsEpQFIQldAhgNPVSpEDqVslAQOI..... 890 H1

> TM <

iPPGDINPTTHKKTtdYlSGpwLMViVCiVlVlVisAAIsM.vyFkrkhQmTKElGHLS 954 D1  
 .....vLTqpwFiiiLgAilavlMLs..fGAMvFVkrkhMm..MkQsAL D2

25 iNK.....SHVpVIViVaILiIFvViiIAY.CYWRNS.rNSD...gkDRSF 909 CE  
 .....SdvVKqp..AFiagiGAaCWiiLMVfsIwLyRHrkKR..NglTsTY 932 H1

VVSDNEIT.....AlniNSKESL.wIDHHRGwRTADTDKD.. 988 D1  
 AGIRKVPSFTFTPTVtyQRGGEAVSSGGRPGLlniSEPAAQPwLAD..TwPNTGNNHNDC 990 H1

30 .....SgLSesKlLSHVNSSQ..SnynnS.....DGGtDyAEvd....TRNL 1024 D1  
 SISCCTAGNgNsDsNlTTYSRPADCIAnynnQLDNKQTNLMLPEStVyGDvdLSNKINEM 1050 H1

CYTOPLASMIC MOTIF #1

35 TtfYNCR.....KSPDNptpyattMIIGTS.....sSETCTkT.TSISADkDSGT 1068 D1  
 KtfNSPNLKDRFVNPSGQptpyattQLiQSNLSNMNNGsGDSGekHWKPLGQQkQEVA 1110 H1

HSPyS.....DAFAGQVPAVpVV..KSnyLqYPVEP..... 1097 D1  
 PVQyNIVEQNKLNKDYRANDTVPPtIPYNQSyDqNTGGSYNSSDRGSSTSGSQGHKKGAR 1170 H1

CYTOPLASMIC MOTIF #2

5 .....InwSEFlppppEhppp...sSTy.....GyAqGSp..... 1124 D1  
 TPKVPKQGGMnwADLlppppAhpppHSNsEEyNISVDESyDqEMpCPVPPARMYLQODEL 1230 H1

.....eSSRKSSKSAGSgISTNQSILNAsIHsSSSGGFsAWGVSPQYAVAcP..... 1171 D1  
 EEeEDERGPTPPVRgAASSPAAVSYsHQsTATLTPsPQEELQPMLQDcpEETGHMQHQPd 1290 H1

10 .....pENVy...sNpl.....SAVAGGTQNRyQITPTNQHPPQl.... 1203 D1  
 RRRQPVSPPPPRPISpPHTyGYIsGplVSDMDTDAPEEEDEADMEVAKMQTRRlLLRG 1350 H1

.....paY.....FATTGPGGAVPPNHLp.....faTQRHaa 1230 D1  
 15 LEQTpaSSVGDLESSVTGSMINGWGSASEEDNISSGRSSVSSSDGSFFTDADfaQAVAAa 1410 H1

SeyQaglNAar.....cAQSRACNsCdALATPSPmq..... 1261 D1  
 Aey.aglKVarRQMqDAAGRRHFHASQcPRPTSPVsTdsNMSAAVmQKTRPAKKLKHQPG 1469 H1

20 style="text-align: center;">CYTOPLASMIC MOTIF #3

.....ppppvpVpEGWYQPVHPNSH.PMHpTS.SNHQIYQCSSECSdHSRSsQS 1307 D1  
 HLRRETYTDDLppppvpPpAIKSPTAQSKTQLEVRpVVVPKLPSMDARTDRsSDRKGsSY 1529 H1

HKrQL.....QLEeHGSSAkQrgGHRRrA.pVVQPCMESeN.....ENM D1  
 25 KGrEVLdGRQVDMRTNPGDPREAQeQQNDGkGrgNKAARdLpPAKTHLIQeDILPYCRPTF H1

LAEYEQRQYTsDCCNssrEGDTC.....SCSeGSCl..yAeAgePAPRQMTAKNT 1395 D1  
 PTSNNPrDPSsSSSMssrGSGSRQREQANVGRRNIAeMQVlGGy.eRgeDNNEELEETES 1651 H1

30 Exemplary such Robo specific immunogenic and/or antigenic peptides are shown in Table 2.

Table 2. Immunogenic Robo polypeptides eliciting Robo-specific rabbit polyclonal antibody:  
 Robo polypeptide-KLH conjugates immunized per protocol described below.

<u>Robo Polypeptide, Sequence</u>	<u>Immunogenicity</u>
35 SEQ ID NO:2, residues 68-77	+++
SEQ ID NO:2, residues 79-94	+++

WO 99/25833

PCT/US98/24327

	SEQ ID NO:2, residues 95-103	+++
	SEQ ID NO:2, residues 122-129	+++
	SEQ ID NO:2, residues 165-176	+++
	SEQ ID NO:2, residues 181-191	+++
5	SEQ ID NO:2, residues 193-204	+++
	SEQ ID NO:2, residues 244-251	+++
	SEQ ID NO:2, residues 274-290	+++
	SEQ ID NO:2, residues 322-331	+++
	SEQ ID NO:2, residues 339-347	+++
10	SEQ ID NO:2, residues 407-417	+++
	SEQ ID NO:2, residues 441-451	+++
	SEQ ID NO:2, residues 453-474	+++
	SEQ ID NO:2, residues 502-516	+++
	SEQ ID NO:2, residues 541-553	+++
15	SEQ ID NO:2, residues 617-629	+++

In addition, species-specific antigenic and/or immunogenic peptides are readily apparent as diverged extracellular or cytosolic regions in Table 1. Human Robo-specific antibodies are characterized as uncross-reactive with non-human Robo polypeptides (SEQ ID NOS:2, 4, 6 and 12). Exemplary such human specific peptides are shown in Table 3.

Table 3. Immunogenic Robo polypeptides eliciting human Robo-specific rabbit polyclonal antibody: Robo polypeptide-KLH conjugates immunized per protocol described below (some antibodies show cross-reactivity with corresponding mouse/rat Robo polypeptides).

	<u>Robo Polypeptide, Sequence</u>	<u>Immunogenicity</u>
25	SEQ ID NO:8, residues 1-12	+++
	SEQ ID NO:8, residues 18-28	+++
	SEQ ID NO:8, residues 31-40	+++
	SEQ ID NO:8, residues 45-65	+++
30	SEQ ID NO:8, residues 106-116	+++
	SEQ ID NO:8, residues 137-145	+++

WO 99/25833

PCT/US98/24327

	SEQ ID NO:8, residues 174-184	+++
	SEQ ID NO:8, residues 214-230	+++
	SEQ ID NO:8, residues 274-286	+++
	SEQ ID NO:8, residues 314-324	+++
5	SEQ ID NO:8, residues 399-412	+++
	SEQ ID NO:8, residues 496-507	+++
	SEQ ID NO:8, residues 548-565	+++
	SEQ ID NO:8, residues 599-611	+++
	SEQ ID NO:8, residues 660-671	+++
10	SEQ ID NO:8, residues 717-730	+++
	SEQ ID NO:8, residues 780-791	+++
	SEQ ID NO:8, residues 835-847	+++
	SEQ ID NO:8, residues 877-891	+++
	SEQ ID NO:8, residues 930-942	+++
15	SEQ ID NO:8, residues 981-998	+++
	SEQ ID NO:8, residues 1040-1051	+++
	SEQ ID NO:8, residues 1080-1090	+++
	SEQ ID NO:8, residues 1154-1168	+++
	SEQ ID NO:8, residues 1215-1231	+++
20	SEQ ID NO:8, residues 1278-1302	+++
	SEQ ID NO:8, residues 1378-1400	+++
	SEQ ID NO:8, residues 1460-1469	+++
	SEQ ID NO:8, residues 1497-1519	+++
	SEQ ID NO:8, residues 1606-1626	+++
25	SEQ ID NO:8, residues 1639-1651	+++
	SEQ ID NO:10, residues 5-16	+++
	SEQ ID NO:10, residues 38-47	+++
	SEQ ID NO:10, residues 83-94	+++
	SEQ ID NO:10, residues 112-125	+++
30	SEQ ID NO:10, residues 168-180	+++
	SEQ ID NO:10, residues 195-209	+++

WO 99/25833

PCT/US98/24327

SEQ ID NO:10, residues 222-235 +++

SEQ ID NO:10, residues 241-254 +++

5 The subject domains provide Robo domain specific activity or function, such as Robo-specific cell, especially neuron modulating or modulating inhibitory activity, Robo-ligand-binding or binding inhibitory activity. Robo-specific activity or function may be determined by convenient *in vitro*, cell-based, or *in vivo* assays: e.g. *in vitro* binding assays, cell culture assays, in animals (e.g. gene therapy, transgenics, etc.), etc. The binding target may be a natural intracellular binding target, a Robo regulating protein or other regulator that directly  
10 modulates Robo activity or its localization; or non-natural binding target such as a specific immune protein such as an antibody, or a Robo specific agent such as those identified in screening assays such as described below. Robo-binding specificity may be assayed by binding equilibrium constants (usually at least about  $10^7 M^{-1}$ , preferably at least about  $10^8 M^{-1}$ , more preferably at least about  $10^9 M^{-1}$ ), by the ability of the subject polypeptide to function as  
15 negative mutants in Robo-expressing cells, to elicit Robo specific antibody in a heterologous host (e.g a rodent or rabbit), etc.

Similarly, the Comm polypeptide is conveniently selected from Comm polypeptides which specifically modulate Robo expression. Exemplary suitable Comm polypeptides (a) comprise SEQ ID NO:14 or a deletion mutant thereof which specifically modulates Comm  
20 expression and/or (b) are encoded by a nucleic acid comprising SEQ ID NO:13 or a nucleic acid which hybridizes with SEQ ID NO:13 under stringent conditions. Suitable deletion mutants are readily screened in Robo down-regulations assays as described below. Preferred Comm domains comprise at least 8, preferably at least 16, more preferably at least 32, most preferably at least 64 consecutive residues of SEQ ID NO:14 and provide a Comm specific  
25 activity, such as Comm-specific antigenicity and/or immunogenicity, especially when coupled to carrier proteins as described above for Robo. Exemplary such Comm specific immunogenic and/or antigenic peptides are shown in Table 4.

30 Table 4. Immunogenic Comm polypeptides eliciting Comm-specific rabbit polyclonal antibody: Comm polypeptide-KLH conjugates immunized per protocol described above.

Comm Polypeptide, SequenceImmunogenicity

WO 99/25833

PCT/US98/24327

	SEQ ID NO:14, residues 1-11	+++
	SEQ ID NO:14, residues 6-17	+++
	SEQ ID NO:14, residues 18-34	+++
	SEQ ID NO:14, residues 35-44	+++
5	SEQ ID NO:14, residues 45-63	+++
	SEQ ID NO:14, residues 64-73	+++
	SEQ ID NO:14, residues 74-89	+++
	SEQ ID NO:14, residues 92-109	+++
	SEQ ID NO:14, residues 110-126	+++
10	SEQ ID NO:14, residues 127-136	+++
	SEQ ID NO:14, residues 137-151	+++
	SEQ ID NO:14, residues 152-171	+++
	SEQ ID NO:14, residues 172-185	+++
	SEQ ID NO:14, residues 186-199	+++
15	SEQ ID NO:14, residues 200-215	+++
	SEQ ID NO:14, residues 216-235	+++
	SEQ ID NO:14, residues 236-250	+++
	SEQ ID NO:14, residues 251-260	+++
	SEQ ID NO:14, residues 261-275	+++
20	SEQ ID NO:14, residues 276-288	+++
	SEQ ID NO:14, residues 289-307	+++
	SEQ ID NO:14, residues 308-317	+++
	SEQ ID NO:14, residues 318-331	+++
	SEQ ID NO:14, residues 332-344	+++
25	SEQ ID NO:14, residues 345-356	+++
	SEQ ID NO:14, residues 357-370	+++
	SEQ ID NO:14, residues 41-153	+++
	SEQ ID NO:14, residues 117-329	+++

30           The subject domains provide Comm domain specific activity or function, such as Comm-specific cell, especially neuron modulating or modulating inhibitory activity, Comm-

ligand-binding or binding inhibitory activity. Comm-specific activity or function may be determined by convenient *in vitro*, cell-based, or *in vivo* assays: e.g. *in vitro* binding assays, cell culture assays, in animals (e.g. gene therapy, transgenics, etc.), etc. The binding target may be a natural intracellular binding target, a Comm regulating protein or other regulator that directly modulates Comm activity or its localization; or non-natural binding target such as a specific immune protein such as an antibody, or a Comm specific agent such as those identified in screening assays such as described below. Comm-binding specificity may be assayed by binding equilibrium constants (usually at least about  $10^7 M^{-1}$ , preferably at least about  $10^8 M^{-1}$ , more preferably at least about  $10^9 M^{-1}$ ), by the ability of the subject polypeptide to function as negative mutants in Comm-expressing cells, to elicit Comm specific antibody in a heterologous host (e.g a rodent or rabbit), etc.

In one embodiment, the Comm polypeptides are encoded by a nucleic acid comprising SEQ ID NO:13 or a nucleic acid which hybridizes with a full-length strand of SEQ ID NO:13, preferably under stringent conditions. Such nucleic acids are at least 36, preferably at least 72, more preferably at least 144 and most preferably at least 288 bases in length. Demonstrating specific hybridization generally requires stringent conditions, for example, hybridizing in a buffer comprising 30% formamide in 5 x SSPE (0.18 M NaCl, 0.01 M NaPO<sub>4</sub>, pH7.7, 0.001 M EDTA) buffer at a temperature of 42°C and remaining bound when subject to washing at 42°C with 0.2 x SSPE (Conditions I); preferably hybridizing in a buffer comprising 50% formamide in 5 x SSPE buffer at a temperature of 42°C and remaining bound when subject to washing at 42°C with 0.2 x SSPE buffer at 42°C (Conditions II). Exemplary nucleic acids which hybridize with a strand of SEQ ID NO:13 are shown in Table 5.

Table 5. Exemplary nucleic acids which hybridize with a strand of SEQ ID NO:13 under Conditions I and/or II.

<u>Comm Nucleic Acids</u>	<u>Hybridization</u>
SEQ ID NO:13, nucleotides 1-47	+
SEQ ID NO:13, nucleotides 58-99	+
SEQ ID NO:13, nucleotides 95-138	+
SEQ ID NO:13, nucleotides 181-220	+
SEQ ID NO:13, nucleotides 261-299	+

WO 99/25833

PCT/US98/24327

	SEQ ID NO:13, nucleotides 274-315	+
	SEQ ID NO:13, nucleotides 351-389	+
	SEQ ID NO:13, nucleotides 450-593	+
	SEQ ID NO:13, nucleotides 524-546	+
5	SEQ ID NO:13, nucleotides 561-608	+
	SEQ ID NO:13, nucleotides 689-727	+
	SEQ ID NO:13, nucleotides 708-737	+
	SEQ ID NO:13, nucleotides 738-801	+
	SEQ ID NO:13, nucleotides 805-854	+
10	SEQ ID NO:13, nucleotides 855-907	+
	SEQ ID NO:13, nucleotides 910-953	+
	SEQ ID NO:13, nucleotides 1007-1059	+

15 A wide variety of cell types express Robo polypeptides subject to regulation by the  
 disclosed methods, including many neuronal cells, transformed cells, infected (e.g. virus) cells,  
 etc. Ascertaining Robo expression is readily effected by antibody staining. Accordingly,  
 indications for the subject methods encompass a wide variety of cell types and function,  
 including axon outgrowth, tumor cell invasion or migration, etc. The target cell may reside in  
 culture or in situ, i.e. within the natural host. For in situ applications, the compositions are  
 20 added to a retained physiological fluid such as blood or synovial fluid. For CNS  
 administration, a variety of techniques are available for promoting transfer of the therapeutic  
 across the blood brain barrier including disruption by surgery or injection, drugs which  
 transiently open adhesion contact between CNS vasculature endothelial cells, and compounds  
 which facilitate translocation through such cells. Comm polypeptides may also be amenable  
 25 to direct injection or infusion, topical, intratracheal/nasal administration e.g. through aerosol,  
 intraocularly, or within/on implants e.g. fibers e.g. collagen, osmotic pumps, grafts comprising  
 appropriately transformed cells, etc. A particular method of administration involves coating,  
 embedding or derivatizing fibers, such as collagen fibers, protein polymers, etc. with  
 therapeutic polypeptides. Other useful approaches are described in Otto et al. (1989) J  
 30 Neuroscience Research 22, 83-91 and Otto and Unsicker (1990) J Neuroscience 10, 1912-  
 1921. Generally, the amount administered will be empirically determined, typically in the

range of about 10 to 1000  $\mu\text{g}/\text{kg}$  of the recipient and the concentration will generally be in the range of about 50 to 500  $\mu\text{g}/\text{ml}$  in the dose administered. Other additives may be included, such as stabilizers, bactericides, etc. will be present in conventional amounts.

In one embodiment, the invention provides administering the subject Comm  
5 polypeptides in combination with a pharmaceutically acceptable excipient such as sterile saline or other medium, gelatin, an oil, etc. to form pharmaceutically acceptable compositions. The compositions and/or compounds may be administered alone or in combination with any convenient carrier, diluent, etc. and such administration may be provided in single or multiple dosages. Useful carriers include solid, semi-solid or liquid media including water and non-  
10 toxic organic solvents. In another embodiment, the invention provides the subject compounds in the form of a pro-drug, which can be metabolically converted to the subject compound by the recipient host. A wide variety of pro-drug formulations for polypeptide-based therapeutics are known in the art. The compositions may be provided in any convenient form including tablets, capsules, troches, powders, sprays, creams, etc. As such the compositions, in  
15 pharmaceutically acceptable dosage units or in bulk, may be incorporated into a wide variety of containers. For example, dosage units may be included in a variety of containers including capsules, pills, etc. The compositions may be advantageously combined and/or used in combination with other therapeutic or prophylactic agents, different from the subject compounds. In many instances, administration in conjunction with the subject compositions  
20 enhances the efficacy of such agents, see e.g. *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 9<sup>th</sup> Ed., 1996, McGraw-Hill.

In another aspect, the invention provides methods of screening for agents which modulate Robo-Comm interactions. These methods generally involve forming a mixture of a Robo-expressing cell, a Comm polypeptide and a candidate agent, and determining the effect  
25 of the agent on the amount of Robo expressed by the cell. The methods are amenable to automated, cost-effective high throughput screening of chemical libraries for lead compounds. Identified reagents find use in the pharmaceutical industries for animal and human trials; for example, the reagents may be derivatized and rescreened in *in vitro* and *in vivo* assays to optimize activity and minimize toxicity for pharmaceutical development. Cell and animal  
30 based neural guidance/repulsion assays are described in detail in the experimental section below.

The following experimental section and examples are offered by way of illustration and not by way of limitation.

## EXPERIMENTAL

5           *roundabout* is Required to Prevent Ipsilateral Axons from Crossing the Midline. Mutations in *robo* lead to an increase in the number of embryonic CNS axons in the commissures, coincident with a reduction of the number of axons in the longitudinal connectives as observed with MAb BP102. The two commissures are thicker than normal and partially fuse as they spill over into one another; the longitudinals are thinner and pulled closer together toward the midline. We analyzed the *robo* mutant phenotype in more detail using the 10 1D4 MAb (anti-Fas II) which at stage 13 stains a subset of growth cones (including aCC, pCC, vMP2, MP1, dMP2) and from stages 14-17 stains three major longitudinal axon tracts, including (from medial to lateral) the pCC pathway (pioneered by the pCC growth cone), the MP1 pathway (pioneered by the MP1 growth cone), and a 3rd lateral pathway (Lin et al., 15 1994; Hidalgo and Brand, 1997). Previous analysis (Seeger et al., 1993) with MAb 1D4 showed that the pCC growth cone, which normally projects anteriorly on its own side near the midline to pioneer the pCC pathway, in *robo* mutant embryos projects across the midline, fasciculating with its contralateral homologue at the midline. The axon pathway it pioneers -- the pCC pathway -- which normally projects longitudinally on its own side near the midline, in 20 *robo* mutant embryos projects back and forth across the midline. The pCC pathway takes on a circular pattern as it joins with the same pathway from the other side and whirls back and forth across the midline, thus defining the phenotype for which the gene was named.

The fuzzy commissure phenotype observed in *robo* mutant embryos does not appear to be due to changes in cell fates at the midline or elsewhere in the CNS. All of the midline cells 25 are present, and their fates appear normal as assayed with a variety of different markers (Seeger et al., 1993). All of the commissural and longitudinal axon pathways begin in their normal location, but the longitudinal pathways are pulled closer at the midline as axon bundles circle around the midline, and the commissures become fused and fuzzy as too many axons cross the midline. In contrast, in mutants in which all or some of the midline cells die or fail 30 to properly differentiate, the longitudinal pathways either collapse onto the midline or from the outset form closer together than normal (Klambt et al., 1991; Mayer and Nüsslein-Volhard,

1988). Thus, the defects observed in *robo* mutant embryos are not due to changes in cell fates but rather result from defects in axon guidance.

We examined in greater detail the behavior of the pCC growth cone in *robo* mutant compared to wild type embryos. In wild type embryos, the vMP2 cell body lies embedded at the edge of the midline. The pCC growth cone extends anterior to a point just lateral to vMP2's cell body. The pCC growth cone is then met by the lateral extension of vMP2's growth cone, and as pCC extends anteriorly and a bit laterally, the vMP2 growth cone wraps around pCC's axons and extends right behind it (Lin et al., 1994). This tight association of vMP2 and pCC is mediated by Fasciclin II (Fas II), a homophilic cell adhesion molecule (CAM) (Grenningloh et al., 1990, 1991) that is expressed from the beginning of axon outgrowth on the cell bodies, axons, and growth cones of a subset of neurons, including pCC and vMP2. In *FasII* mutant embryos, vMP2 and pCC no longer tightly associate, and their axons fail to fasciculate (Lin et al., 1994).

If the pCC and vMP2 neurons express Fas II, and their growth cones and axons are so attracted to each other in a Fas II-mediated fashion, why does not pCC's growth cone initially extend more medially toward vMP2's cell body which is a short distance away? The answer appears to be because vMP2's cell body is partly embedded in other midline cells, and thus vMP2's cell body is partly surrounded by the putative midline repellent. In *robo* mutant embryos, pCC's initial trajectory is directly toward vMP2's cell body, where it adheres to vMP2; pCC's growth cone then crosses the midline, fasciculating with its contralateral homologue at the midline.

*roundabout* is Required to Prevent Commissural Axons from Recrossing the Midline. The circular pathway taken by the pCC pathway as it crosses back and forth across the midline (as visualized with the anti-Fas II MAb) led us to suggest that some axons were freely recrossing the midline. Although Fas II is expressed on a relatively small subset of axons in the early embryo, and thus we can use it to observe pCC's growth cone abnormal crossing of the midline in *robo* mutants, the resulting pattern of expression in older embryos becomes quite complicated and it is difficult to resolve precisely which axons are crossing the midline.

To confirm that axons cross and recross the midline freely in *robo* mutants, we examined the expression of Connectin (Nose et al., 1992), a CAM expressed on a more restricted subset of CNS axons than is Fas II. Connectin is also expressed on motor axons in

the segmental nerve. We used the C1.427 MAb to follow Connectin expression (Meadows et al., 1994). Connectin is expressed on the SP1 neuron whose cell body lies near the midline just anterior to the anterior commissure, and just medial to the longitudinal tracts. SP1's growth cone normally projects across the midline, fasciculating with the axon of its  
 5 contralateral homologue. The growth cone then appears to adhere to the cell body of its contralateral homologue, grows around that cell body, and turns to project anteriorly in a medial sub-fascicle of the pCC pathway.

In *robo* mutant embryos as in wild type embryos, SP1's growth cone extends across the midline, adheres to the axon and then cell body of its contralateral homologue, and turns to  
 10 project anteriorly. However, as it extends anteriorly into the next segment, it typically moves toward the midline, apparently attracted towards and adhering to the axon of its contralateral homologue just on the other side of the midline. The two SP1 axons typically join together around the posterior commissure of the next anterior segment. Sometimes they extend  
 15 together on the left side of the midline and sometimes on the right side, freely crossing and recrossing the midline while fasciculating with the SP1 axons originating from both sides of neighboring segments. These results show that in addition to preventing ipsilaterally projecting axons from crossing the midline, Robo also functions to prevent contralaterally projecting axons from recrossing the midline.

*roundabout* Controls Crossing of the midline in a Dosage Sensitive Manner. Another  
 20 axonal marker which labels a very small subset of axons is the *Tau-β-galactosidase* reporter gene expressed under control of the *apterous* promoter (called *apC*; Lundgren et al., 1995). In wild type embryos, the *apC-tau-lacZ* transgene labels three interneurons per abdominal hemisegment, here called the Ap neurons. The Ap neurons have lateral cell bodies and their growth cones initially project towards the midline. Upon nearing the midline, these growth  
 25 cones then turn to project anteriorly on their own side along the edge of the midline, fasciculating with each other and with their homologues from neighboring segments; in wild type embryos, they never cross the midline in abdominal segments.

In *robo* mutant embryos, the Ap axons cross the midline in every segment, join up with their contralateral homologues, and often project anteriorly in one discrete longitudinal fascicle.  
 30 The Ap fascicle displays two behaviors, usually crossing and recrossing the midline multiple times as a single bundle, or occasionally separating into different bundles of axons which

project on one side or the other and independently cross the midline.

We observed a partially penetrant Ap axon phenotype in *robo* heterozygous embryos. In wild type, none of the 6 Ap axons in each segment ever cross the midline; in *robo* homozygous mutants, all 6 Ap axons cross the midline. In *robo* heterozygous mutant embryos, one of the Ap axons is observed crossing the midline in about 30% of segments, which accounts for a penetrance of about 5% of all Ap axons (Table 1). This partially penetrant crossing with 50% of *robo* indicates a dose requirement for the *robo* gene product in these axons. Moreover, since the Ap axons extend midway through axonogenesis, once many axon pathways have already been pioneered, these results indicate that *robo* is required throughout axonogenesis, not just to establish the initial projections of the pioneer axons.

Underexpression of Comm Leads to Increased Levels of Robo Protein. The *commissureless (comm)* mutant has a complementary phenotype to that of *robo* in that too few axons cross the midline (Seeger et al., 1993). When visualised with MAb BP102, the axon commissures are noticeably absent. In certain hypomorphic *comm* alleles (e.g., *comm*<sup>7</sup>; Tear et al., 1996), the commissures are not completely absent, but rather partial and highly abnormal axon commissures do form in a few segments (particularly in the thorax). We examined the expression of Robo protein in these *comm* hypomorphic alleles using the 13C9 anti-Robo MAb (Kidd et al, 1997). Normally, Robo is expressed at very low levels on commissural axons and at high levels on longitudinal axons. In *comm* mutant embryos, Robo expression in the longitudinal tracts appears even higher than normal. Interestingly, in *comm* hypomorphic alleles, the occasional thin commissures express Robo protein at levels that are higher than normally seen in the commissures and closer to what is typically seen in the longitudinal tracts. This result was our first hint that Comm protein might function by suppressing Robo expression on commissural axons. Previous studies had shown that *comm* encodes a novel transmembrane protein that is expressed by the midline glia and that is apparently transferred to commissural axons (Tear et al., 1996). Given these results, we wondered whether expression of *comm* in all neurons might reduce Robo levels and lead to a *robo* phenotype.

Overexpression of Comm generates a *robo*-Like Phenotype. To test the hypothesis that increased expression of *comm* might lead to a *robo*-like phenotype, we used the UAS-GAL4 system (Brand and Perrimon, 1993) to change the pattern of *comm* expression. We

generated *UAS-comm* transgenic lines and drove expression pan-neurally using the *sca-GAL4* line. Since flies carrying a copy of both the driver and effector transgenes are viable, we used them as parents and examined their progeny. A continuous range of *robo*-like phenotypes was observed with MAbs BP102 and 1D4. The range of phenotypes reveals the *comm* gain-of-function phenotype to be dosage sensitive, as the severity increased in embryos carrying two copies of both transgenes as compared to embryos carrying only one copy of each.

Superficially, the *robo* phenotype can be mimicked by mutants causing inappropriate migration or cell death of the midline glia, both of which result in fuzzy commissures (Klämbt et al., 1991). However, such phenotypes are not visible until midway through axonogenesis, and are easily detected by examining early axon behavior. In addition, we stained the embryos with a MAb raised against 8H11, a protein expressed specifically by the midline glia, and confirmed that the midline glia are still present

In the embryos ectopically expressing *comm*, Fas II positive axons, such as pCC, were found to behave identical to how they behave in *robo* mutants. When *Comm* is overexpressed, the pCC growth cone extends towards the vMP2 cell body, and then across the midline, just as it does in a *robo* mutant. In the *comm* gain-of-function, the pCC fascicle freely crosses and midline and forms the same circles or whirls as it does in the *robo* loss-of-function.

**Overexpression of *Comm* Leads to Reduced Levels of Robo Protein.** Having established that the *comm* overexpression generates a bona fide *robo*-like axon guidance phenotype, we next examined Robo expression in these embryos using the anti-Robo MAb 13C9. The *sca-GAL4* driver begins driving expression in the neuroepithelium before axon outgrowth (~stage 9) has begun and switches off by stage 13; *sca-GAL4* does not express in the epidermis. In wild type embryos, the pattern of Robo protein expression begins in the neuroepithelium, as well in some lateral epidermal stripes, but is conspicuously absent from the midline region. In *comm* gain-of-function embryos, Robo expression in the neuroepithelium is greatly reduced or absent, while the epidermal expression outside the nervous system is maintained. This same pattern can be observed around the time when the first growth cones are extending. In wild type embryos during stages 12 and 13, no Robo is seen at the midline, but there is a high level of Robo expression on ipsilaterally projecting growth cones such as pCC and a significant level throughout the neuroepithelium. In contrast,

in *comm* gain-of-function embryos, the pCC growth cone lacks Robo protein and the neuroepithelium expresses greatly reduced levels of Robo.

The dramatic reduction in the levels of Robo were observed until about stage 14, coincident with the *sca-GAL4* driver ceasing expression. In the *sca-GAL4; UAS-robo* embryos, Robo protein begins to accumulate throughout the CNS after stage 14, reaching significant levels (but still below wild type) by stage 16. Interestingly, in these transgenic embryos, although we observe some Robo-positive axons in the commissures at later stages, Robo expression remains higher in longitudinal tracts. We interpret the Robo-positive axons in the commissures as later axons following misguided pioneer axons; fasciculation with the pioneers allows these Robo-positive axons to cross the midline in spite of modest levels of Robo.

The *elav-GAL4* line also expresses pan-neurally but only in post-mitotic neurons; it begins driving expression of UAS transgenes during stage 12 and remains expressed throughout the rest of embryogenesis. Ectopic expression of *comm* by *elav-GAL4* led to a less severe version of the *robo* phenotype. We interpret this weaker phenotype as being due to either a weaker overall level of Comm expression or because increased Comm initiates after the pioneers have already established the initial pathways. In addition, since *sca-GAL4* drives expression in midline glia, the source of normal *comm* expression, while *elav-GAL4* does not drive expression in the midline glia, the possibility exists that the less severe phenotype of the *elav* transgene is due to this lack of midline *comm* expression.

To address this issue, we attempted to increase the level of Comm specifically at the midline using multiple GAL4 lines, including *sim-GAL4*, *slit-GAL4*, *F63-GAL4*, and *p52A-GAL4*, all of which express at the midline during the period of commissure formation. When *UAS-comm* was expressed by any of these four lines, only very weak BP102 phenotypes were observed, although because most of these inserts are homozygous lethal, we have not been able to easily increase the dosage with these lines to comparable levels as with the *sca-GAL4* line. None of these gain-of-function phenotypes was as strong as that observed with the *sca-GAL4* line. We also cannot rule out that these differences in the strength of the gain-of-function phenotypes using different GAL4 lines do not reflect differences in timing, levels of expression, or location of expression within the CNS.

We conclude that the normal function of *comm* is to down-regulate the low level of

Robo expression present on commissural axons, thereby allowing them to cross the midline. Increasing levels of Comm in the CNS lead to more severe *robo*-like phenotypes, indicating a dosage sensitivity. This sensitivity to dosage is also reflected in the behavior of Ap axons in *robo* heterozygotes, thus showing a parallel dosage sensitivity by either decreasing Robo or  
 5 increasing Comm.

These results indicate that control of Robo expression is complex and highly regulated from transcription to translation to post-translational. We show that there is an inverse correlation between Comm expression and Robo expression. In wild type embryos, Comm is expressed at the midline, and Robo expression is very low on commissural axons crossing the  
 10 midline. In *comm* hypomorphic mutant embryos, those few axons that do cross the midline now express higher levels of Robo protein. In *comm* gain-of-function embryos (using transgenic constructs that drive over- and ectopic expression of *comm*), the overall levels of Robo are dramatically decreased wherever increased Comm expression coincides with Robo expression. Furthermore, using certain GAL4 lines that drive transient *comm* expression, we  
 15 observe that once Comm disappears in older embryos, Robo protein expression begins to increase towards its normal levels. Thus, Comm down-regulates Robo expression in a very tight fashion.

Only a small amount of Comm is normally expressed at the midline. The midline also expresses high levels of a putative repellent that is the ligand for the Robo receptor. Growth  
 20 cones that express high levels of Robo, such as ipsilaterally projecting growth cones from the outset or commissural growth cones once they cross the midline, are relatively immune to significant down-regulation by the normally low levels of midline Comm and thus are prevented from crossing the midline. Only abnormally high levels of Comm (using transgenes that drive overexpression) are sufficient to down-regulate this Robo expression to a level that  
 25 allows these growth cones to cross the midline. In contrast, growth cones that normally express lower levels of Robo (i.e., those commissural growth cones that cross the midline in the presence of Comm) are highly sensitive to Comm, in that the normal low levels of Comm can further reduce their levels of Robo and thus allow them to cross the midline. In the absence of Comm, these growth cones can not cross the midline, due to their low levels of  
 30 Robo; in the *robo; comm* double mutant they all freely cross.

Genetic Stocks. All *robo* alleles were isolated on chromosomes deficient for *Fasciclin*

*III* as described in Seeger et al., 1993. The *robo* phenotype is independent of the absence of *FasIII*.

Protein Immunocytochemistry. Immunocytochemistry was performed as described by Patel (1994). For anti-Robo staining, MAb 13C9 was diluted 1:10 in PBS with 0.1% Tween-20, and the embryos were fixed and cracked so as to minimize exposure to methanol. The presence of triton and storage of embryos in methanol were both found to destroy the activity of MAb 13C9. For anti-Connectin staining with MAb C1.427, the embryos were fixed in 3.7% formaldehyde/PEM buffer (100mM PIPES, 2mM EGTA, 1mM MgSO<sub>4</sub>); C1.427 was diluted 1:10 in PBS with 0.1% Triton. The *apterous-tau-lacZ* embryos were hand devitellinized and dissected on poly-lysine coated slides and subsequently fixed for 20 minutes with 3.7% formaldehyde; rabbit anti- $\beta$ -galactosidase (Cappel) was used at 1:10,000 and biotinylated anti-rabbit secondary was used at 1:1000 and enhanced with the Vectastain Elite ABC kit (Vector Laboratories).

Transformation of *Drosophila*, *robo* rescue and overexpression. The *comm* cDNA was inserted as a 1.7kb XhoI-XbaI fragment into the XhoI and Xba sites of pUAST (Brand and Perrimon, 1993). Transformant lines were generated and mapped by standard procedures.

#### References

- Brand, A. H. and Perrimon, N. (1993) *Development* 118, 401-415.
- Hidalgo, A., and Brand, A.H. (1997) *Development* 124, 3253-3262.
- Kidd, T., Brose, K., Mitchell, K., Fetter, R., Tessier-Lavigne, M., Goodman, C.S., and Tear, G. (1997). Roundabout controls axon crossing of the CNS midline and defines a new subfamily of evolutionarily conserved guidance receptors. *Cell*, in review.
- Klämbt, C., Jacobs, J. R., and Goodman, C. S. (1991) *Cell* 64, 801-815.
- Lundgren, S.E., et al. (1995) *Development* 121, 1769-1773.
- Mayer, U. and Nüsslein-Volhard, C. (1988) *Genes Dev.* 2, 1496 - 1511.
- Meadows, L.A., et al. (1994) *J. Cell Sci.* 107, 321-328.
- Nose, A., Mahajan, V. B., and Goodman, C. S. (1992) *Cell* 70, 553-567.
- Patel, N. H. (1994) In "Methods in Cell Biology, Vol 44. *Drosophila melanogaster*: Practical Uses in Cell Biology" (L. S. B. Goldstein and E. Fyrberg, eds) Academic Press, New York.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Seeger, M., Tear, G., Ferres-Marco, D. and Goodman C.S. (1993) *Neuron* 10, 409 - 426.

Tear G., et al. (1996) *Neuron* 16, 501 - 514.

5 All publications and patent applications cited in this specification are herein  
incorporated by reference as if each individual publication or patent application were  
specifically and individually indicated to be incorporated by reference. Although the  
foregoing invention has been described in some detail by way of illustration and example for  
purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the  
art in light of the teachings of this invention that certain changes and modifications may be  
10 made thereto without departing from the spirit or scope of the appended claims.

WO 99/25833

PCT/US98/24327

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Goodman, Corey S.  
Kidd, Thomas  
5 Russel, Claire  
Tear, Guy  
Mitchell, Kevin

(ii) TITLE OF INVENTION: Methods for Modulating Nerve Cell Function

(iii) NUMBER OF SEQUENCES: 14

10 (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
(B) STREET: 75 DENISE DRIVE  
(C) CITY: HILLSBOROUGH  
(D) STATE: CALIFORNIA  
15 (E) COUNTRY: USA  
(F) ZIP: 94010

(v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
20 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:  
25 (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: OSMAN, RICHARD A  
(B) REGISTRATION NUMBER: 36,627  
(C) REFERENCE/DOCKET NUMBER: B98-006

30 (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: (650) 343-4341  
(B) TELEFAX: (650) 343-4342

(2) INFORMATION FOR SEQ ID NO:1:

35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4188 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGCATCCCA	TGCATCCCGA	AAACCACGCC	ATCGCCCGGA	GCACGAGCAC	CACTAATAAC	60
CCATCTCGCA	GTCGGAGCAG	CAGGATGTGG	CTCCTGCCCG	CCTGGCTGCT	CCTCGTCCTG	120
GTGGCCAGCA	ATGGCCTGCC	AGCAGTCAGA	GGCCAGTACC	AATCGCCACG	TATCATCGAG	180
45 CATCCCACGG	ATCTGGTCGT	TAAGAAGAAT	GAACCCGCCA	CGCTCAACTG	CAAAGTGGAG	240
GGCAAGCCGG	AACCCACCAT	TGAGTGGTTT	AAGGATGGCG	AACCCGTCAG	CACCAACGAA	300
AAGAAATCGC	ACCGCGTCCA	GTTCAAGGAC	GGCGCCCTCT	TCTTTTACAG	GACAATGCAA	360
GGCAAGAAGG	AGCAGGACGG	CGGAGAGTAC	TGGTGCGTGG	CCAAGAACCG	AGTGGGCCAG	420
GCCGTTAGTC	GCCATGCCTC	CCTCCAGATA	GCTGTTTTGC	GCGACGATTT	TCGCGTGGAG	480
50 CCCAAAGACA	CGCGAGTGGC	CAAAGGCGAG	ACGGCTCTGC	TGGAGTGTGG	GCCGCCCAA	540
GGCATTCCAG	AGCCAACGCT	GATTTGGATA	AAGGACGGCG	TTCCCTTGGA	CGACCTGAAA	600
GCCATGTTCG	TTGGCGCCAG	CTCCCGCGTT	CGAATTGTGG	ACGGTGGCAA	CCTGCTGATC	660
AGCAATGTGG	AGCCCATTTA	TGAGGGCAAC	TACAAGTGCA	TTGCCAGAA	TCTGGTAGGC	720
ACCCGCGAGA	GCAGCTATGC	CAAGCTGATT	GTCCAGGTCA	AACCATACTT	TATGAAGGAG	780
55 CCCAAGGATC	AGGTGATGCT	CTACGGCCAG	ACAGCCACTT	TCCACTGCTC	AGTGGGCGGT	840

WO 99/25833

PCT/US98/24327

	GATCCGCCGC	CGAAAGTGTT	GTGGAAAAAG	GAGGAGGGCA	ATATTCCGGT	GTCCAGAGCG	900
	CGAATCCTTC	ACGACGAGAA	AAGTTTAGAG	ATATCCAACA	TAACGCCAC	CGATGAGGGC	960
	ACCTATGTCT	GCGAGGCACA	CAACAATGTC	GGTCAGATCA	GCGCTAGGGC	TTCTCTTATA	1020
	GTCCACGCTC	CGCCGAACCT	TACGAAAAGA	CCCAGTAACA	AGAAAGTGGG	ACTAAATGGG	1080
5	GTTGTCCAAC	TACCTTGCA	GGCCTCCGGA	AACCCTCCGC	CGTCTGTATT	CTGGACCAAG	1140
	GAAGGAGTAT	CCACTCTTAT	GTTCCCAAAT	AGTTCGCACG	GAAGGCAGTA	TGTGGCTGCC	1200
	GATGGAACCT	TGCAGATTAC	GGATGTGCGG	CAGGAAGACG	AAGGCTACTA	TGTGTGTTCC	1260
	GCTTTCAGTG	TAGTCGATTC	CTCTACAGTA	CGGGTTTTCC	TGCAAGTCAG	CTCGGTAGAC	1320
	GAGCGTCCAC	CTCCGATTAT	TCAAATCGGA	CCTGCCAATC	AAACACTGCC	CAAGGGATCA	1380
10	GTTGCTACTT	TACCCTGTCT	GGCCACTGGA	AATCCCAGTC	CCCGTATCAA	GTGGTTCCAC	1440
	GATGGACATG	CCGTACAAGC	GGGCAATCGA	TACAGCATCA	TCCAAGGAAG	CTCACTGAGA	1500
	GTCGATGACC	TTCAACTAAG	TGACTCTGGT	ACCTACACCT	GCACTGCATC	TGGCGAACGA	1560
	GGAGAAACTT	CCTGGGCTGC	CACACTAACG	GTGGAAAAAC	CCGGTTCTAC	ATCTCTTCAC	1620
	CGGGCAGCTG	ATCCTAGCAC	TTATCCTGCT	CCTCCAGGAA	CACCTAAAGT	CCTGAATGTC	1680
15	AGTCGCACCA	GCATTAGTCT	TCGTTGGGCT	AAAAGCCAAG	AGAAACCCGG	AGCTGTGGGC	1740
	CCAATCATTG	GATACACTGT	AGAGTACTTC	AGTCCGGATC	TGCAAACCTGG	TTGGATTGTG	1800
	GCTGCCCATC	GAGTCGGCGA	CACTCAAGTC	ACTATCTCGG	GTCTCACTCC	TGGCACTTCG	1860
	TATGTGTTCC	TAGTTAGAGC	TGAGAATACT	CAGGGTATTT	CTGTGCCTTC	CGGCTTATCA	1920
	AATGTTATTA	AAACCATTGA	GGCAGATTTT	GATGCAGCTT	CTGCCAATGA	TTTGTGAGCA	1980
20	GCTCGAACTT	TGCTGACAGG	AAAGTCGGTG	GAGCTAATAG	ATGCCCTCGGC	TATCAATGCT	2040
	AGTGCCGTTA	GACTTGAGTG	GATGCTCCAC	GTGAGCGCTG	ATGAGAAATA	CGTAGAGGGC	2100
	CTGCGCATA	ACTATAAGGA	TGCCAGTGTA	CCATCCGCAC	AGTATCACTC	GATCACTGTT	2160
	ATGGATGCCT	CTGCAGAATC	GTTTGTGGTG	GGAAACCTTA	AGAAGTACAC	CAAGTATGAG	2220
	TTCTTCCTAA	CACCCTTTTT	TGAGACAATT	GAAGGACAGC	CCAGTAACTC	CAAGACAGCC	2280
25	CTCACCTATG	AAGATGTTCC	CTCCGCACCA	CCGGATAACA	TTGAGATTGG	CATGTACAAC	2340
	CAAACAGCCG	GTTGGGTGCG	TTGGACTCCG	CCACCCTCCC	AGCACCACAA	TGGCAATTTG	2400
	TATGGCTACA	AGATTGAGGT	CAGCGCCGGT	AACACCATGA	AGGTGCTGGC	CAATATGACT	2460
	CTTAATGCTA	CCACCACATC	TGTGCTCCTA	AATAACCTAA	CCACCAGGAGC	TGTGTACAGC	2520
	GTGAGGTTGA	ACTCCTTTAC	CAAGGCAGGA	GATGGACCTT	ACTCCAAACC	GATATCACTA	2580
30	TTCATGGACC	CCACCATCA	TGTGCATCCG	CCACGGGCAC	ATCCAAGCGG	CACCCATGAT	2640
	GGGCGACATG	AGGGACAGGA	TCTCACGTAT	CATAACAATG	GCAACATACC	ACCTGGCGAC	2700
	ATTAATCCCA	CCACTCATAA	AAAGACCACT	GACTACCTAT	CTGGACCGTG	GCTAATGGTG	2760
	CTGGTCTGCA	TCGTTCTTCT	AGTCCTGGTT	ATTTCCGGCGG	CTATTTTCGAT	GGTCTACTTC	2820
	AAGCGCAAGC	ATCAAATGAC	CAAGGAATTG	GGTCACTTAA	GTGTGGTCAG	TGACAACGAA	2880
35	ATAACCGCAT	TAAATATCAA	TAGCAAAGAG	AGCCTTTGGA	TAGACCATCA	TCGTGGATGG	2940
	CGAACTGCCG	ATACTGACAA	AGACTCAGGA	TTAAGCGAAT	CGAAGCTACT	ATCCCACGTT	3000
	AACAGCAGTC	AATCCAACCTA	CAATAACTCC	GATGGAGGAA	CCGATTATGC	AGAAGTTGAC	3060
	ACCCGTAACC	TTACCACCTT	CTACAATTGT	CGCAAGAGCC	CCGATAATCC	CACGCCGTAC	3120
	GCCACCACTA	TGATCATTTG	TACCTCTTCC	AGTGAGACCT	GCACCAAGAC	AACATCTATA	3180
40	AGTGCCGATA	AGGACTCGGG	AACTCATTCG	CCCTATTCTG	ACGCATTTGC	CGGTCAGGTG	3240
	CCAGCGGTTT	CTGTTGTCAA	ATCCAACCTAT	CTTCAGTATC	CGGTTGAACC	GATCAACTGG	3300
	TCAGAGTTTC	TACCCCCGCC	GCCAGAACAC	CCACCTCCGT	CTTCTACCTA	TGGATACGCA	3360
	CAAGGATCTC	CTGAATCTTC	GCGGAAGAGC	TCCAAAAGCG	CAGGTTCCGG	CATTTCTACA	3420
	AATCAAAGCA	TTCTGAACGC	ATCCATACAC	AGCAGCTCCT	CGGGCGGCTT	TTCAGCTTGG	3480
45	GGAGTATCGC	CCCAATATGC	TGTCGCCTGT	CCACCGGAAA	ACGTTTATAG	CAATCCGCTG	3540
	TCGGCAGTGG	CTGGCGGCAC	CCAGAACCGC	TATCAGATAA	CGCCCACAAA	CCAACATCCG	3600
	CCACAGTTAC	CGGCCTACTT	TGCCACCACG	GGTCCAGGAG	GAGCTGTACC	ACCCAACCAC	3660
	CTGCCATTTG	CCACACAGCG	TCATGCAGCC	AGCGAGTACC	AGGCTGGACT	GAATGCAGCG	3720
	CGATGTGCCC	AAAGCCGCGC	CTGCAACAGC	TGCGATGCCT	TGGCCACACC	CTCGCCCATG	3780
50	CAACCCCCAC	CGCCAGTTCC	CGTACCCGAG	GGCTGGTACC	AACCGGTGCA	TCCCAATAGC	3840
	CACCCGATGC	ACCCGACCTC	CTCCAACCAC	CAGATCTACC	AGTGCTCCTC	CGAGTGCTCG	3900
	GATCACTCGA	GGAGCTCGCA	GAGTCACAAG	CGGCAGCTGC	AGCTCGAGGA	GCACGGCAGC	3960
	AGTGCCAAAC	AACGCGGAGG	ACACCACCGT	CGACGAGCCC	CGGTGGTGCA	GCCGTGCATG	4020
	GAGAGCGAGA	ACGAGAACAT	GCTGGCGGAG	TACGAGCAGC	GCCAGTACAC	CAGCGATTGC	4080
55	TGCAATAGCT	CCCGCGAGGG	CGACACCTGC	TCCTGCAGCG	AGGGATCCTG	TCTTTACGCC	4140

WO 99/25833

PCT/US98/24327

GAGGCGGGCG AGCCGGCGCC TCGTCAAATG ACTGCTAAGA ACACCTAA

4188

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1395 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

1	Met	His	Pro	Met	His	Pro	Glu	Asn	His	Ala	Ile	Ala	Arg	Ser	Thr	Ser
				5						10					15	
	Thr	Thr	Asn	Asn	Pro	Ser	Arg	Ser	Arg	Ser	Ser	Arg	Met	Trp	Leu	Leu
				20						25				30		
15	Pro	Ala	Trp	Leu	Leu	Leu	Val	Leu	Val	Ala	Ser	Asn	Gly	Leu	Pro	Ala
				35						40				45		
	Val	Arg	Gly	Gln	Tyr	Gln	Ser	Pro	Arg	Ile	Ile	Glu	His	Pro	Thr	Asp
				50						55			60			
	Leu	Val	Val	Lys	Lys	Asn	Glu	Pro	Ala	Thr	Leu	Asn	Cys	Lys	Val	Glu
20						70										80
	Gly	Lys	Pro	Glu	Pro	Thr	Ile	Glu	Trp	Phe	Lys	Asp	Gly	Glu	Pro	Val
						85				90					95	
	Ser	Thr	Asn	Glu	Lys	Lys	Ser	His	Arg	Val	Gln	Phe	Lys	Asp	Gly	Ala
				100						105					110	
25	Leu	Phe	Phe	Tyr	Arg	Thr	Met	Gln	Gly	Lys	Lys	Glu	Gln	Asp	Gly	Gly
										120					125	
	Glu	Tyr	Trp	Cys	Val	Ala	Lys	Asn	Arg	Val	Gly	Gln	Ala	Val	Ser	Arg
										130					140	
	His	Ala	Ser	Leu	Gln	Ile	Ala	Val	Leu	Arg	Asp	Asp	Phe	Arg	Val	Glu
30																160
	Pro	Lys	Asp	Thr	Arg	Val	Ala	Lys	Gly	Glu	Thr	Ala	Leu	Leu	Glu	Cys
										165						175
	Gly	Pro	Pro	Lys	Gly	Ile	Pro	Glu	Pro	Thr	Leu	Ile	Trp	Ile	Lys	Asp
										180						190
	Gly	Val	Pro	Leu	Asp	Asp	Leu	Lys	Ala	Met	Ser	Phe	Gly	Ala	Ser	Ser
35										200						205
	Arg	Val	Arg	Ile	Val	Asp	Gly	Gly	Asn	Leu	Leu	Ile	Ser	Asn	Val	Glu
										210						220
	Pro	Ile	Asp	Glu	Gly	Asn	Tyr	Lys	Cys	Ile	Ala	Gln	Asn	Leu	Val	Gly
40																240
	Thr	Arg	Glu	Ser	Ser	Tyr	Ala	Lys	Leu	Ile	Val	Gln	Val	Lys	Pro	Tyr
										245						255
	Phe	Met	Lys	Glu	Pro	Lys	Asp	Gln	Val	Met	Leu	Tyr	Gly	Gln	Thr	Ala
										260						270
45	Thr	Phe	His	Cys	Ser	Val	Gly	Gly	Asp	Pro	Pro	Pro	Lys	Val	Leu	Trp
										275						285
	Lys	Lys	Glu	Glu	Gly	Asn	Ile	Pro	Val	Ser	Arg	Ala	Arg	Ile	Leu	His
										290						300
	Asp	Glu	Lys	Ser	Leu	Glu	Ile	Ser	Asn	Ile	Thr	Pro	Thr	Asp	Glu	Gly
50																320
	Thr	Tyr	Val	Cys	Glu	Ala	His	Asn	Asn	Val	Gly	Gln	Ile	Ser	Ala	Arg
										325						335
	Ala	Ser	Leu	Ile	Val	His	Ala	Pro	Pro	Asn	Phe	Thr	Lys	Arg	Pro	Ser
										340						350
55	Asn	Lys	Lys	Val	Gly	Leu	Asn	Gly	Val	Val	Gln	Leu	Pro	Cys	Met	Ala

WO 99/25833

PCT/US98/24327

		355				360					365					
	Ser	Gly	Asn	Pro	Pro	Pro	Ser	Val	Phe	Trp	Thr	Lys	Glu	Gly	Val	Ser
		370					375					380				
5	Thr	Leu	Met	Phe	Pro	Asn	Ser	Ser	His	Gly	Arg	Gln	Tyr	Val	Ala	Ala
	385					390					395					400
	Asp	Gly	Thr	Leu	Gln	Ile	Thr	Asp	Val	Arg	Gln	Glu	Asp	Glu	Gly	Tyr
					405					410					415	
	Tyr	Val	Cys	Ser	Ala	Phe	Ser	Val	Val	Asp	Ser	Ser	Thr	Val	Arg	Val
				420					425					430		
10	Phe	Leu	Gln	Val	Ser	Ser	Val	Asp	Glu	Arg	Pro	Pro	Pro	Ile	Ile	Gln
			435					440					445			
	Ile	Gly	Pro	Ala	Asn	Gln	Thr	Leu	Pro	Lys	Gly	Ser	Val	Ala	Thr	Leu
	450						455					460				
15	Pro	Cys	Arg	Ala	Thr	Gly	Asn	Pro	Ser	Pro	Arg	Ile	Lys	Trp	Phe	His
	465					470					475					480
	Asp	Gly	His	Ala	Val	Gln	Ala	Gly	Asn	Arg	Tyr	Ser	Ile	Ile	Gln	Gly
				485						490					495	
	Ser	Ser	Leu	Arg	Val	Asp	Asp	Leu	Gln	Leu	Ser	Asp	Ser	Gly	Thr	Tyr
				500					505					510		
20	Thr	Cys	Thr	Ala	Ser	Gly	Glu	Arg	Gly	Glu	Thr	Ser	Trp	Ala	Ala	Thr
			515					520					525			
	Leu	Thr	Val	Glu	Lys	Pro	Gly	Ser	Thr	Ser	Leu	His	Arg	Ala	Ala	Asp
	530						535					540				
25	Pro	Ser	Thr	Tyr	Pro	Ala	Pro	Pro	Gly	Thr	Pro	Lys	Val	Leu	Asn	Val
	545					550					555					560
	Ser	Arg	Thr	Ser	Ile	Ser	Leu	Arg	Trp	Ala	Lys	Ser	Gln	Glu	Lys	Pro
				565						570					575	
	Gly	Ala	Val	Gly	Pro	Ile	Ile	Gly	Tyr	Thr	Val	Glu	Tyr	Phe	Ser	Pro
				580					585					590		
30	Asp	Leu	Gln	Thr	Gly	Trp	Ile	Val	Ala	Ala	His	Arg	Val	Gly	Asp	Thr
			595					600					605			
	Gln	Val	Thr	Ile	Ser	Gly	Leu	Thr	Pro	Gly	Thr	Ser	Tyr	Val	Phe	Leu
	610						615					620				
35	Val	Arg	Ala	Glu	Asn	Thr	Gln	Gly	Ile	Ser	Val	Pro	Ser	Gly	Leu	Ser
	625					630					635					640
	Asn	Val	Ile	Lys	Thr	Ile	Glu	Ala	Asp	Phe	Asp	Ala	Ala	Ser	Ala	Asn
				645						650					655	
	Asp	Leu	Ser	Ala	Ala	Arg	Thr	Leu	Leu	Thr	Gly	Lys	Ser	Val	Glu	Leu
				660						665				670		
40	Ile	Asp	Ala	Ser	Ala	Ile	Asn	Ala	Ser	Ala	Val	Arg	Leu	Glu	Trp	Met
			675					680					685			
	Leu	His	Val	Ser	Ala	Asp	Glu	Lys	Tyr	Val	Glu	Gly	Leu	Arg	Ile	His
	690						695					700				
45	Tyr	Lys	Asp	Ala	Ser	Val	Pro	Ser	Ala	Gln	Tyr	His	Ser	Ile	Thr	Val
	705					710					715					720
	Met	Asp	Ala	Ser	Ala	Glu	Ser	Phe	Val	Val	Gly	Asn	Leu	Lys	Lys	Tyr
				725						730					735	
	Thr	Lys	Tyr	Glu	Phe	Phe	Leu	Thr	Pro	Phe	Phe	Glu	Thr	Ile	Glu	Gly
				740					745					750		
50	Gln	Pro	Ser	Asn	Ser	Lys	Thr	Ala	Leu	Thr	Tyr	Glu	Asp	Val	Pro	Ser
			755					760					765			
	Ala	Pro	Pro	Asp	Asn	Ile	Gln	Ile	Gly	Met	Tyr	Asn	Gln	Thr	Ala	Gly
	770						775					780				
55	Trp	Val	Arg	Trp	Thr	Pro	Pro	Pro	Ser	Gln	His	His	Asn	Gly	Asn	Leu
	785					790					795					800

WO 99/25833

PCT/US98/24327

	Tyr	Gly	Tyr	Lys	Ile	Glu	Val	Ser	Ala	Gly	Asn	Thr	Met	Lys	Val	Leu
					805					810					815	
	Ala	Asn	Met	Thr	Leu	Asn	Ala	Thr	Thr	Thr	Ser	Val	Leu	Leu	Asn	Asn
				820					825					830		
5	Leu	Thr	Thr	Gly	Ala	Val	Tyr	Ser	Val	Arg	Leu	Asn	Ser	Phe	Thr	Lys
			835					840					845			
	Ala	Gly	Asp	Gly	Pro	Tyr	Ser	Lys	Pro	Ile	Ser	Leu	Phe	Met	Asp	Pro
		850					855					860				
10	Thr	His	His	Val	His	Pro	Pro	Arg	Ala	His	Pro	Ser	Gly	Thr	His	Asp
	865					870					875					880
	Gly	Arg	His	Glu	Gly	Gln	Asp	Leu	Thr	Tyr	His	Asn	Asn	Gly	Asn	Ile
				885						890					895	
	Pro	Pro	Gly	Asp	Ile	Asn	Pro	Thr	Thr	His	Lys	Lys	Thr	Thr	Asp	Tyr
				900					905					910		
15	Leu	Ser	Gly	Pro	Trp	Leu	Met	Val	Leu	Val	Cys	Ile	Val	Leu	Leu	Val
			915				920						925			
	Leu	Val	Ile	Ser	Ala	Ala	Ile	Ser	Met	Val	Tyr	Phe	Lys	Arg	Lys	His
		930					935					940				
20	Gln	Met	Thr	Lys	Glu	Leu	Gly	His	Leu	Ser	Val	Val	Ser	Asp	Asn	Glu
	945					950					955					960
	Ile	Thr	Ala	Leu	Asn	Ile	Asn	Ser	Lys	Glu	Ser	Leu	Trp	Ile	Asp	His
				965						970					975	
	His	Arg	Gly	Trp	Arg	Thr	Ala	Asp	Thr	Asp	Lys	Asp	Ser	Gly	Leu	Ser
				980					985					990		
25	Glu	Ser	Lys	Leu	Leu	Ser	His	Val	Asn	Ser	Ser	Gln	Ser	Asn	Tyr	Asn
			995					1000					1005			
	Asn	Ser	Asp	Gly	Gly	Thr	Asp	Tyr	Ala	Glu	Val	Asp	Thr	Arg	Asn	Leu
		1010					1015						1020			
30	Thr	Thr	Phe	Tyr	Asn	Cys	Arg	Lys	Ser	Pro	Asp	Asn	Pro	Thr	Pro	Tyr
	1025					1030					1035					1040
	Ala	Thr	Thr	Met	Ile	Ile	Gly	Thr	Ser	Ser	Ser	Glu	Thr	Cys	Thr	Lys
				1045							1050				1055	
	Thr	Thr	Ser	Ile	Ser	Ala	Asp	Lys	Asp	Ser	Gly	Thr	His	Ser	Pro	Tyr
				1060					1065					1070		
35	Ser	Asp	Ala	Phe	Ala	Gly	Gln	Val	Pro	Ala	Val	Pro	Val	Val	Lys	Ser
			1075					1080					1085			
	Asn	Tyr	Leu	Gln	Tyr	Pro	Val	Glu	Pro	Ile	Asn	Trp	Ser	Glu	Phe	Leu
		1090					1095					1100				
40	Pro	Pro	Pro	Pro	Glu	His	Pro	Pro	Pro	Ser	Ser	Thr	Tyr	Gly	Tyr	Ala
	1105					1110						1115				1120
	Gln	Gly	Ser	Pro	Glu	Ser	Ser	Arg	Lys	Ser	Ser	Lys	Ser	Ala	Gly	Ser
				1125							1130				1135	
	Gly	Ile	Ser	Thr	Asn	Gln	Ser	Ile	Leu	Asn	Ala	Ser	Ile	His	Ser	Ser
				1140					1145					1150		
45	Ser	Ser	Gly	Gly	Phe	Ser	Ala	Trp	Gly	Val	Ser	Pro	Gln	Tyr	Ala	Val
			1155					1160					1165			
	Ala	Cys	Pro	Pro	Glu	Asn	Val	Tyr	Ser	Asn	Pro	Leu	Ser	Ala	Val	Ala
		1170					1175					1180				
50	Gly	Gly	Thr	Gln	Asn	Arg	Tyr	Gln	Ile	Thr	Pro	Thr	Asn	Gln	His	Pro
	1185					1190					1195					1200
	Pro	Gln	Leu	Pro	Ala	Tyr	Phe	Ala	Thr	Thr	Gly	Pro	Gly	Gly	Ala	Val
				1205						1210					1215	
	Pro	Pro	Asn	His	Leu	Pro	Phe	Ala	Thr	Gln	Arg	His	Ala	Ala	Ser	Glu
				1220					1225					1230		
55	Tyr	Gln	Ala	Gly	Leu	Asn	Ala	Ala	Arg	Cys	Ala	Gln	Ser	Arg	Ala	Cys

WO 99/25833

PCT/US98/24327

		1235				1240					1245			
		Asn Ser Cys Asp Ala Leu				Ala Thr Pro Ser Pro Met Gln Pro Pro Pro								
		1250				1255					1260			
5		Pro Val Pro Val Pro Glu Gly Trp Tyr Gln Pro Val His Pro Asn Ser												
		1265			1270						1275			1280
		His Pro Met His Pro Thr Ser Ser Asn His Gln Ile Tyr Gln Cys Ser												
					1285					1290				1295
		Ser Glu Cys Ser Asp His Ser Arg Ser Ser Gln Ser His Lys Arg Gln												
					1300					1305				1310
10		Leu Gln Leu Glu Glu His Gly Ser Ser Ala Lys Gln Arg Gly Gly His												
					1315					1320				1325
		His Arg Arg Arg Ala Pro Val Val Gln Pro Cys Met Glu Ser Glu Asn												
					1330					1335				1340
		Glu Asn Met Leu Ala Glu Tyr Glu Gln Arg Gln Tyr Thr Ser Asp Cys												
15					1345					1350				1360
		Cys Asn Ser Ser Arg Glu Gly Asp Thr Cys Ser Cys Ser Glu Gly Ser												
					1365					1370				1375
		Cys Leu Tyr Ala Glu Ala Gly Glu Pro Ala Pro Arg Gln Met Thr Ala												
					1380					1385				1390
20		Lys Asn Thr												
					1395									

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4146 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

30	GGTGAAAATC CACGCATCAT CGAGCATCCC ATGGACACGA CGGTGCCAAA AAATGATCCA	60
	TTTACGTTTA ATTGCCAGGC CGAGGGCAAT CCAACACCAA CCATTCAATG GTTTAAGGAC	120
	GGTCGCGAAC TGAAGACGGA TACGGGTTCG CATCGCATAA TGCTGCCCGC CGGGGGTCTA	180
	TTCTTTCTCA AGGTTATCCA CTCACGTAGA GAGAGCGATG CGGGCACTTA CTGGTGCGAG	240
35	GCCAAAAACG AGTTTGGAGT GGCACGGTCC AGGAATGCAA CGTTGCAAGT GGCAGTTCTC	300
	CGCGACGAAT TCCGTTTGGG GCCGGCAAAT ACCCGCGTGG CCCAAGGCGA GGTGGCCCTG	360
	ATGGAATGCG GTGCCCCCG AGGATCTCCG GAGCCGCAA TCTCGTGGCG CAAGAACGGC	420
	CAGACCCTGA ATCTTGTCGG GAACAAGCGG ATTCGCATTG TCGACGGTGG CAATCTGGCC	480
	ATCCAGGAAG CCCGCCAATC GGACGACGGA CGCTACCAGT GTGTGGTCAA GAATGTGGTT	540
40	GGCACCCGGG AGTCGGCCAC CGCTTTTCTT AAAGTGCATG TACGTCCATT CCTCATCCGA	600
	GGACCCCGA ATCAGACGGC GGTGGTGGGC AGCTCGGTGG TCTTCCAGTG CCGCATCGGA	660
	GGCGATCCCC TGCCTGATGT CCTGTGGCGA CGCACTGCCT CCGGCGCAA TATGCCACTG	720
	CGTAAGTTTT CTTGGCTTCA TTCAGCTTCA GGTCGTGTGC ACGTACTTGA GGACCGCAGT	780
	CTGAAGCTGG ACGACGTTAC TCTGGAGGAC ATGGGCGAGT ACACTTGCGA GGCGGACAAT	840
45	GCGGTGGGCG GCATCACGGC CACTGGCATC CTCACCGTTC ACGTCCCCC CAAATTTGTG	900
	ATACGCCCCA AGAATCAGCT GGTGGAGATC GGTGATGAAG TGCTGTTCGA GTGCCAAGCG	960
	AATGGACATC CCCGACCAAC GCTCTACTGG TCGGTGGAGG GCAACAGCTC CCTGCTGCTC	1020
	CCCGGCTATC GGGATGGCCG CATGGAAGTG ACCCTGACGC CCGAGGGGCG CTCGGTGCTC	1080
	TCGATAGCTC GATTTGCCCG TGAGGATTCC GGAAAGGTGG TCACTTGCAA CGCCCTGAAC	1140
50	GCCGTGGGCA GCGTCAGCAG TCGGACTGTG GTCAGTGTGG ATACGCAATT CGAGCTGCCA	1200
	CCGCCGATTA TCGAACAGGG GCCCGTGAAT CAAACGTTGC CCGTTAAATC AATTGTGGTT	1260
	CTGCCATGCC GAACTCTGGG CACTCCAGTG CCACAGGTCT CTTGGTACCT GGATGGCATA	1320
	CCCATCGATG TGCAGGAGCA CGAGCGGCGG AATCTTTCGG ACGCTGGAGC CTTAACCATT	1380
	TCGGATCTTC AGCGCCACGA GGATGAAGGC TTGTACACCT GCGTGGCCAG CAATCGCAAC	1440
55	GGAAAATCCT CTTGGAGTGG TTACCTTCGT CTGGACACCC CGACAAATCC GAATATCAAG	1500

WO 99/25833

PCT/US98/24327

	TTCTTCAGAG	CCCCAGAACT	TTCCACCTAC	CCAGGGCCGC	CAGGAAAACC	GCAAATGGTG	1560
	GAGAAGGGCG	AAAATTCGGT	GACTCTCAGC	TGGACGAGGA	GCAACAAGGT	GGGCGGCTCC	1620
	AGTCTGGTGG	GCTATGTAAT	CGAGATGTTT	GGCAAAAACG	AAACGGATGG	CTGGGTGGCT	1680
	GTGGGCACTA	GGGTGCAAAA	TACCACGTTT	ACCCAAACGG	GTCTGCTGCC	GGGTGTGAAT	1740
5	TACTTCTTTC	TAATTCGAGC	CGAGAACTCC	CATGGCTTAT	CACTGCCCAG	TCCGATGTCG	1800
	GAACCCATTA	CGGTGGGAAC	GCGCTACTTC	AATAGTGGTC	TGGATCTGAG	CGAGGCTCGT	1860
	GCCAGTCTGC	TGTCCGGAGA	TGTTGTGGAG	CTGAGCAACG	CCAGTGTGGT	GGACTCCACT	1920
	AGCATGAAAC	TCACCTGGCA	GATCATCAAT	GGCAAATACG	TCGAGGGCTT	CTATGTCTAT	1980
	GCGAGACAGT	TGCCAAATCC	AATAGTCAAC	AATCCGGCGC	CCGTTACTAG	CAATACCAAT	2040
10	CCGCTGCTGG	GCTCTACATC	CACATCCGCA	TCCGCATCCG	CCTCGGCATC	GGCATTGATT	2100
	TCGACAAAGC	CAAATATTGC	AGCTGCCGGC	AAACGTGATG	GGGAGACAAA	CCAGAGTGGG	2160
	GGAGGAGCTC	CGACCCCACT	GAACACCAAG	TATCGCATGC	TAACGATTCT	CAATGGCGGT	2220
	GGCGCCTCAT	CCTGCACCAT	CACCGGGCTC	GTCCAGTACA	CGCTGTATGA	ATTTTTCATC	2280
	GTGCCATTTT	ACAAATCCGT	CGAGGGCAAG	CCGTGCAATT	CGCGCATCGC	TCGCACCCTT	2340
15	GAAGATGTTC	CCTCTGAGGC	ACCATATGGA	ATGGAGGCTC	TGCTGTTGAA	CTCCTCCGCG	2400
	GTCTTCCTCA	AATGGAAGGC	ACCAGAACTC	AAGGATCGGC	ATGGTGTTC	CTTGAACTAT	2460
	CATGTTATAG	TCCGAGGTAT	TGACACTGCC	CACAATTTCT	CACGCATTTT	GACAAATGTC	2520
	ACCATCGATG	CCGCTTCGCC	TACTCTGGTT	TTGGCCAATC	TCACCGAAGG	CGTCATGTAC	2580
	ACCGTGGGCG	TGGCGGCCGG	AAATAACGCT	GGAGTTGGTC	CTTATTGTGT	CCCAGCTACT	2640
20	TTGCGTTTGG	ATCCCATCAC	AAAGCGACTC	GATCCGTTCA	TCAATCAGCG	GGACCATGTT	2700
	AACGATGTGC	TGACGCAGCC	CTGGTTCATA	ATACTCCTGG	GCGCCATCCT	GGCCGTTCTT	2760
	ATGCTGTCCT	TTGGCGCAAT	GGTCTTTGTG	AAGCGCAAGC	ACATGATGAT	GAAGCAGTCG	2820
	GCCCTAAATA	CAATGCGTGG	CAATCACACG	AGCGACGTGC	TCAAAATGCC	GAGTCTATCG	2880
	GCGCGCAATG	GAAACGGCTA	CTGGCTGGAC	TCCCTCACCG	GCGGAATGGT	GTGGCGTCCC	2940
25	TCGCCCGGCG	GCGACTCGCT	GGAGATGCAA	AAGGATCACA	TCGCCGACTA	TGCGCCGGTC	3000
	TGCGGTGCCC	CCGGTTCTCC	GGCCGGCGGT	GGCACCTCTT	CCGGTGGATC	CGGTGGCGCG	3060
	GGCAGCGGTG	CCAGCGGCGG	CGATGACATT	CATGGAGGAC	ACGGCAGCGA	ACGCAATCAG	3120
	CAGCGGTACG	TGGGCGAGTA	CTCCAACATA	CCGACCGACT	ATGCAGAGGT	GTCCAGTTTT	3180
	GGCAAGGCAC	CCAGCGAGTA	TGGTCGGCAT	GGCAACGCCT	CCCCGGCCCC	TTATGCCACC	3240
30	TCTTCGATCC	TGAGTCCCCA	CCAGCAGCAA	CAGCAGCAGC	AGCCGCGTTA	TCAACAGCGA	3300
	CCAGTGCCCG	GCTATGGGCT	CCAGCGCCCA	ATGCACCCAC	ACTACCAGCA	GCAGCAGCAT	3360
	CAGCAGCAAC	AGGCGCAGCA	GACGCACCAG	CAACACCAGG	CTCTCCAGCA	GCACCAGCAA	3420
	CTGCCACCCA	GCAACATCTA	CCAGCAGATG	TCCACCACCA	GCGAGATATA	CCCCACGAAC	3480
	ACGGGTCCCT	CGCGCTCTGT	CTACTCTGAG	CAGTATTACT	ACCCCAAGGA	CAAGCAGAGA	3540
35	CACATCCACA	TCACCGAGAA	CAAGCTGAGC	AACTGCCACA	CCTATGAGGC	GGCTCCTGGC	3600
	GCCAAGCAGT	CCTCGCCGAT	ATCCTCGCAG	TTCGCCAGCG	TGAGGCGGCA	GCAGCTGCCG	3660
	CCCAACTGCA	GCATCGGCAG	GGAAAGTGCC	CGCTTCAAGG	TGCTAAACAC	GGATCAGGGC	3720
	AAGAACCAGC	AGAATCTCCT	GGATCTCGAC	GGCTCCTCGA	TGTGCTACAA	CGGTCTGGCA	3780
	GACTCGGGCT	GCGGTGGATC	TCCCTCCCCG	ATGGCCATGC	TGATGTCGCA	CGAGGACGAG	3840
40	CACGCGCTGT	ACCACACGGC	GGATGGGGAT	CTGGACGACA	TGGAACGACT	GTACGTCAAG	3900
	GTGGACGAGC	AGCAGCCTCC	ACAGCAGCAG	CAGCAGCTGA	TTCCCCTGGT	CCCACAGCAT	3960
	CCGGCGGAAG	GTCACCTGCA	GTCCTGGCGG	AATCAGAGCA	CGCGGAGCAG	TCGGAAGAAC	4020
	GGCCAGGAAT	GCATCAAGGA	ACCCAGCGAG	TTGATCTACG	CTCCGGGAAG	CGTGGCCAGC	4080
	GAACGGAGCC	TCCTCAGCAA	CTCGGGTAGC	GGCACCAGCA	GCCAGCCAGC	TGGCCACAAT	4140
45	GTCTGA						4146

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1381 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

55 Gly Glu Asn Pro Arg Ile Ile Glu His Pro Met Asp Thr Thr Val Pro

WO 99/25833

PCT/US98/24327

	1			5					10					15			
	Lys	Asn	Asp	Pro	Phe	Thr	Phe	Asn	Cys	Gln	Ala	Glu	Gly	Asn	Pro	Thr	
				20					25					30			
5	Pro	Thr	Ile	Gln	Trp	Phe	Lys	Asp	Gly	Arg	Glu	Leu	Lys	Thr	Asp	Thr	
			35					40					45				
	Gly	Ser	His	Arg	Ile	Met	Leu	Pro	Ala	Gly	Gly	Leu	Phe	Phe	Leu	Lys	
		50					55					60					
	Val	Ile	His	Ser	Arg	Arg	Glu	Ser	Asp	Ala	Gly	Thr	Tyr	Trp	Cys	Glu	
	65					70					75					80	
10	Ala	Lys	Asn	Glu	Phe	Gly	Val	Ala	Arg	Ser	Arg	Asn	Ala	Thr	Leu	Gln	
					85					90					95		
	Val	Ala	Val	Leu	Arg	Asp	Glu	Phe	Arg	Leu	Glu	Pro	Ala	Asn	Thr	Arg	
				100					105					110			
	Val	Ala	Gln	Gly	Glu	Val	Ala	Leu	Met	Glu	Cys	Gly	Ala	Pro	Arg	Gly	
15			115					120					125				
	Ser	Pro	Glu	Pro	Gln	Ile	Ser	Trp	Arg	Lys	Asn	Gly	Gln	Thr	Leu	Asn	
		130					135					140					
	Leu	Val	Gly	Asn	Lys	Arg	Ile	Arg	Ile	Val	Asp	Gly	Gly	Asn	Leu	Ala	
	145					150					155				160		
20	Ile	Gln	Glu	Ala	Arg	Gln	Ser	Asp	Asp	Gly	Arg	Tyr	Gln	Cys	Val	Val	
					165					170					175		
	Lys	Asn	Val	Val	Gly	Thr	Arg	Glu	Ser	Ala	Thr	Ala	Phe	Leu	Lys	Val	
				180					185					190			
	His	Val	Arg	Pro	Phe	Leu	Ile	Arg	Gly	Pro	Gln	Asn	Gln	Thr	Ala	Val	
25			195					200					205				
	Val	Gly	Ser	Ser	Val	Val	Phe	Gln	Cys	Arg	Ile	Gly	Gly	Asp	Pro	Leu	
		210					215					220					
	Pro	Asp	Val	Leu	Trp	Arg	Arg	Thr	Ala	Ser	Gly	Gly	Asn	Met	Pro	Leu	
	225					230					235				240		
30	Arg	Lys	Phe	Ser	Trp	Leu	His	Ser	Ala	Ser	Gly	Arg	Val	His	Val	Leu	
					245					250				255			
	Glu	Asp	Arg	Ser	Leu	Lys	Leu	Asp	Asp	Val	Thr	Leu	Glu	Asp	Met	Gly	
				260					265					270			
	Glu	Tyr	Thr	Cys	Glu	Ala	Asp	Asn	Ala	Val	Gly	Gly	Ile	Thr	Ala	Thr	
35			275					280					285				
	Gly	Ile	Leu	Thr	Val	His	Ala	Pro	Pro	Lys	Phe	Val	Ile	Arg	Pro	Lys	
		290					295					300					
	Asn	Gln	Leu	Val	Glu	Ile	Gly	Asp	Glu	Val	Leu	Phe	Glu	Cys	Gln	Ala	
	305					310					315					320	
40	Asn	Gly	His	Pro	Arg	Pro	Thr	Leu	Tyr	Trp	Ser	Val	Glu	Gly	Asn	Ser	
					325					330					335		
	Ser	Leu	Leu	Leu	Pro	Gly	Tyr	Arg	Asp	Gly	Arg	Met	Glu	Val	Thr	Leu	
				340					345					350			
	Thr	Pro	Glu	Gly	Arg	Ser	Val	Leu	Ser	Ile	Ala	Arg	Phe	Ala	Arg	Glu	
45			355					360					365				
	Asp	Ser	Gly	Lys	Val	Val	Thr	Cys	Asn	Ala	Leu	Asn	Ala	Val	Gly	Ser	
		370					375					380					
	Val	Ser	Ser	Arg	Thr	Val	Val	Ser	Val	Asp	Thr	Gln	Phe	Glu	Leu	Pro	
	385					390					395					400	
50	Pro	Pro	Ile	Ile	Glu	Gln	Gly	Pro	Val	Asn	Gln	Thr	Leu	Pro	Val	Lys	
					405					410					415		
	Ser	Ile	Val	Val	Leu	Pro	Cys	Arg	Thr	Leu	Gly	Thr	Pro	Val	Pro	Gln	
				420					425					430			
	Val	Ser	Trp	Tyr	Leu	Asp	Gly	Ile	Pro	Ile	Asp	Val	Gln	Glu	His	Glu	
55			435					440					445				

WO 99/25833

PCT/US98/24327

	Arg	Arg	Asn	Leu	Ser	Asp	Ala	Gly	Ala	Leu	Thr	Ile	Ser	Asp	Leu	Gln
	450						455					460				
	Arg	His	Glu	Asp	Glu	Gly	Leu	Tyr	Thr	Cys	Val	Ala	Ser	Asn	Arg	Asn
	465					470					475					480
5	Gly	Lys	Ser	Ser	Trp	Ser	Gly	Tyr	Leu	Arg	Leu	Asp	Thr	Pro	Thr	Asn
					485					490						495
	Pro	Asn	Ile	Lys	Phe	Phe	Arg	Ala	Pro	Glu	Leu	Ser	Thr	Tyr	Pro	Gly
				500					505					510		
10	Pro	Pro	Gly	Lys	Pro	Gln	Met	Val	Glu	Lys	Gly	Glu	Asn	Ser	Val	Thr
			515					520					525			
	Leu	Ser	Trp	Thr	Arg	Ser	Asn	Lys	Val	Gly	Gly	Ser	Ser	Leu	Val	Gly
	530						535						540			
	Tyr	Val	Ile	Glu	Met	Phe	Gly	Lys	Asn	Glu	Thr	Asp	Gly	Trp	Val	Ala
	545					550						555				560
15	Val	Gly	Thr	Arg	Val	Gln	Asn	Thr	Thr	Phe	Thr	Gln	Thr	Gly	Leu	Leu
					565						570					575
	Pro	Gly	Val	Asn	Tyr	Phe	Phe	Leu	Ile	Arg	Ala	Glu	Asn	Ser	His	Gly
				580					585						590	
20	Leu	Ser	Leu	Pro	Ser	Pro	Met	Ser	Glu	Pro	Ile	Thr	Val	Gly	Thr	Arg
			595					600					605			
	Tyr	Phe	Asn	Ser	Gly	Leu	Asp	Leu	Ser	Glu	Ala	Arg	Ala	Ser	Leu	Leu
	610						615						620			
	Ser	Gly	Asp	Val	Val	Glu	Leu	Ser	Asn	Ala	Ser	Val	Val	Asp	Ser	Thr
	625					630					635					640
25	Ser	Met	Lys	Leu	Thr	Trp	Gln	Ile	Ile	Asn	Gly	Lys	Tyr	Val	Glu	Gly
					645					650						655
	Phe	Tyr	Val	Tyr	Ala	Arg	Gln	Leu	Pro	Asn	Pro	Ile	Val	Asn	Asn	Pro
				660					665						670	
30	Ala	Pro	Val	Thr	Ser	Asn	Thr	Asn	Pro	Leu	Leu	Gly	Ser	Thr	Ser	Thr
			675					680						685		
	Ser	Ala	Ser	Ala	Ser	Ala	Ser	Ala	Ser	Ala	Leu	Ile	Ser	Thr	Lys	Pro
	690						695					700				
	Asn	Ile	Ala	Ala	Ala	Gly	Lys	Arg	Asp	Gly	Glu	Thr	Asn	Gln	Ser	Gly
	705					710						715				720
35	Gly	Gly	Ala	Pro	Thr	Pro	Leu	Asn	Thr	Lys	Tyr	Arg	Met	Leu	Thr	Ile
					725						730					735
	Leu	Asn	Gly	Gly	Gly	Ala	Ser	Ser	Cys	Thr	Ile	Thr	Gly	Leu	Val	Gln
				740					745					750		
40	Tyr	Thr	Leu	Tyr	Glu	Phe	Phe	Ile	Val	Pro	Phe	Tyr	Lys	Ser	Val	Glu
			755					760					765			
	Gly	Lys	Pro	Ser	Asn	Ser	Arg	Ile	Ala	Arg	Thr	Leu	Glu	Asp	Val	Pro
			770				775						780			
	Ser	Glu	Ala	Pro	Tyr	Gly	Met	Glu	Ala	Leu	Leu	Leu	Asn	Ser	Ser	Ala
	785					790						795				800
45	Val	Phe	Leu	Lys	Trp	Lys	Ala	Pro	Glu	Leu	Lys	Asp	Arg	His	Gly	Val
					805					810						815
	Leu	Leu	Asn	Tyr	His	Val	Ile	Val	Arg	Gly	Ile	Asp	Thr	Ala	His	Asn
				820					825					830		
50	Phe	Ser	Arg	Ile	Leu	Thr	Asn	Val	Thr	Ile	Asp	Ala	Ala	Ser	Pro	Thr
			835					840						845		
	Leu	Val	Leu	Ala	Asn	Leu	Thr	Glu	Gly	Val	Met	Tyr	Thr	Val	Gly	Val
			850				855					860				
	Ala	Ala	Gly	Asn	Asn	Ala	Gly	Val	Gly	Pro	Tyr	Cys	Val	Pro	Ala	Thr
	865					870						875				880
55	Leu	Arg	Leu	Asp	Pro	Ile	Thr	Lys	Arg	Leu	Asp	Pro	Phe	Ile	Asn	Gln

WO 99/25833

PCT/US98/24327

				885					890					895			
	Arg	Asp	His	Val	Asn	Asp	Val	Leu	Thr	Gln	Pro	Trp	Phe	Ile	Ile	Leu	
				900					905					910			
5	Leu	Gly	Ala	Ile	Leu	Ala	Val	Leu	Met	Leu	Ser	Phe	Gly	Ala	Met	Val	
			915					920					925				
	Phe	Val	Lys	Arg	Lys	His	Met	Met	Met	Lys	Gln	Ser	Ala	Leu	Asn	Thr	
	930						935					940					
	Met	Arg	Gly	Asn	His	Thr	Ser	Asp	Val	Leu	Lys	Met	Pro	Ser	Leu	Ser	
10	945					950						955				960	
	Ala	Arg	Asn	Gly	Asn	Gly	Tyr	Trp	Leu	Asp	Ser	Ser	Thr	Gly	Gly	Met	
				965						970					975		
	Val	Trp	Arg	Pro	Ser	Pro	Gly	Gly	Asp	Ser	Leu	Glu	Met	Gln	Lys	Asp	
				980					985					990			
15	His	Ile	Ala	Asp	Tyr	Ala	Pro	Val	Cys	Gly	Ala	Pro	Gly	Ser	Pro	Ala	
			995					1000					1005				
	Gly	Gly	Gly	Thr	Ser	Ser	Gly	Gly	Ser	Gly	Gly	Ala	Gly	Ser	Gly	Ala	
	1010						1015					1020					
	Ser	Gly	Gly	Asp	Asp	Ile	His	Gly	Gly	His	Gly	Ser	Glu	Arg	Asn	Gln	
20	1025					1030						1035				1040	
	Gln	Arg	Tyr	Val	Gly	Glu	Tyr	Ser	Asn	Ile	Pro	Thr	Asp	Tyr	Ala	Glu	
				1045							1050				1055		
	Val	Ser	Ser	Phe	Gly	Lys	Ala	Pro	Ser	Glu	Tyr	Gly	Arg	His	Gly	Asn	
				1060					1065					1070			
25	Ala	Ser	Pro	Ala	Pro	Tyr	Ala	Thr	Ser	Ser	Ile	Leu	Ser	Pro	His	Gln	
			1075					1080						1085			
	Gln	Gln	Gln	Gln	Gln	Gln	Pro	Arg	Tyr	Gln	Gln	Arg	Pro	Val	Pro	Gly	
	1090						1095						1100				
	Tyr	Gly	Leu	Gln	Arg	Pro	Met	His	Pro	His	Tyr	Gln	Gln	Gln	Gln	His	
30	1105					1110						1115				1120	
	Gln	Gln	Gln	Gln	Ala	Gln	Gln	Thr	His	Gln	Gln	His	Gln	Ala	Leu	Gln	
				1125							1130				1135		
	Gln	His	Gln	Gln	Leu	Pro	Pro	Ser	Asn	Ile	Tyr	Gln	Gln	Met	Ser	Thr	
				1140					1145					1150			
35	Thr	Ser	Glu	Ile	Tyr	Pro	Thr	Asn	Thr	Gly	Pro	Ser	Arg	Ser	Val	Tyr	
			1155					1160					1165				
	Ser	Glu	Gln	Tyr	Tyr	Tyr	Pro	Lys	Asp	Lys	Gln	Arg	His	Ile	His	Ile	
	1170						1175					1180					
	Thr	Glu	Asn	Lys	Leu	Ser	Asn	Cys	His	Thr	Tyr	Glu	Ala	Ala	Pro	Gly	
40	1185					1190						1195				1200	
	Ala	Lys	Gln	Ser	Ser	Pro	Ile	Ser	Ser	Gln	Phe	Ala	Ser	Val	Arg	Arg	
				1205							1210				1215		
	Gln	Gln	Leu	Pro	Pro	Asn	Cys	Ser	Ile	Gly	Arg	Glu	Ser	Ala	Arg	Phe	
				1220						1225				1230			
45	Lys	Val	Leu	Asn	Thr	Asp	Gln	Gly	Lys	Asn	Gln	Gln	Asn	Leu	Leu	Asp	
			1235					1240					1245				
	Leu	Asp	Gly	Ser	Ser	Met	Cys	Tyr	Asn	Gly	Leu	Ala	Asp	Ser	Gly	Cys	
	1250						1255					1260					
	Gly	Gly	Ser	Pro	Ser	Pro	Met	Ala	Met	Leu	Met	Ser	His	Glu	Asp	Glu	
50	1265					1270						1275				1280	
	His	Ala	Leu	Tyr	His	Thr	Ala	Asp	Gly	Asp	Leu	Asp	Asp	Met	Glu	Arg	
				1285							1290				1295		
	Leu	Tyr	Val	Lys	Val	Asp	Glu	Gln	Gln	Pro	Pro	Gln	Gln	Gln	Gln	Gln	
				1300						1305				1310			
55	Leu	Ile	Pro	Leu	Val	Pro	Gln	His	Pro	Ala	Glu	Gly	His	Leu	Gln	Ser	
			1315					1320						1325			

WO 99/25833

PCT/US98/24327

Trp Arg Asn Gln Ser Thr Arg Ser Ser Arg Lys Asn Gly Gln Glu Cys  
 1330 1335 1340  
 Ile Lys Glu Pro Ser Glu Leu Ile Tyr Ala Pro Gly Ser Val Ala Ser  
 1345 1350 1355 1360  
 5 Glu Arg Ser Leu Leu Ser Asn Ser Gly Ser Gly Thr Ser Ser Gln Pro  
 1365 1370 1375  
 Ala Gly His Asn Val  
 1380

10 (2) INFORMATION FOR SEQ ID NO:5:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3894 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

	ATGTACTATC	TAGGTTTTTA	CCACACTCAC	ACACACACAC	ACACATACAT	AAATTTTGAT	60
	AAAATTCCTA	ATGCCTCAAA	TCTCGCTCCC	GTGATAATCG	AACATCCCAT	CGATGTGGTG	120
20	GTATCTAGGG	GATCGCCAGC	AACCCTCAAC	TGTGGTGCAA	AGCCATCTAC	CGCCAAAATC	180
	ACATGGTACA	AGGATGGACA	GCCCGTAATC	ACGAATAAGG	AGCAAGTGAA	CAGCCACCGG	240
	ATTGTTCTCG	ACACGGGATC	CCTGTTTCTT	CTGAAAGTGA	ATAGTGGAAA	AAACGGAAAA	300
	GACAGCGATG	CGGGAGCGTA	CTATTGTGTG	GCCAGCAACG	AGCACGGAGA	AGTGAAGTCG	360
	AACGAAGGAT	CGTTAAAATT	GGCGATGCTT	CGCGAAGACT	TTCGAGTTCG	GCCAAGAACA	420
25	G TTCAGGCTC	TTGGTGGAGA	GATGGCCGTT	CTGGAATGCA	GTCCGCCACG	TGGATTCCCG	480
	GAGCCGGTTG	TGAGCTGGCG	GAAAGACGAC	AAAGAGCTCC	GAATTCAAGA	CATGCCACGA	540
	TACTACTTAC	ACTCTGACGG	AAACCTCATC	ATTGATCCGG	TCGATCGAAG	CGATTCTGGT	600
	ACTTATCAGT	GTGTTGCCAA	CAACATGGTC	GGAGAACGGG	TGTCCAATCC	CGCAAGATTG	660
	AGTGTCTTTG	AGAAACCAAA	GTTTGAGCAA	GAACCCAAGG	ACATGACGGT	CGACGTCGGA	720
30	GCCGCAGTGC	TGTTTGATTG	TCGTGTGACT	GGAGATCCTC	AACCACAAAT	TACGTGGAAA	780
	CGCAAAAATG	AGCCGATGCC	AGTTACACGT	GCATACATTG	CCAAGGATAA	TCGGGGGTTG	840
	AGAATCGAAA	GAGTTCAACC	ATCAGACGAA	GGTGAATACG	TTTGCTATGC	ACGAAATCCA	900
	GCGGGAAC TC	TTGAAGCATC	TGCACATCTT	CGTGTCCAGG	CACCTCCATC	CTTCCAGACA	960
	AAACCAGCAG	ACCAGTCAGT	TCCAGCTGGA	GGCACGGCAA	CTTTTGAATG	CACCTTGGTC	1020
35	GGTCAACCGA	GTCCCGCCTA	TTTTTGGAGC	AAGGAAGGCC	AACAGGATCT	TCTTTTCCCA	1080
	AGTTATGTGT	CCGCTGATGG	TAGAACGAAA	GTTTCACCAA	CTGGAACATT	GACAATTGAG	1140
	GAAGTTCGTC	AAGTTGATGA	GGGAGCTTAT	GTGTGCGCTG	GAATGAACTC	GGCAGGAAGC	1200
	TCGTTGAGCA	AGGCAGCTTT	GAAAGCAACA	TTTGAAACCA	AAGGCCGTGT	CCAAAAAAA	1260
	AAGAGCAAAA	TGGGCAAAACA	GAAACAAAAA	AATGTTCAAT	CAATTATCAA	ATATTTAATT	1320
40	TCAGCCGTGA	CCGGAAACAC	ACCCGCCAAA	CCACCACCAA	CAATCGAGCA	TGGTCATCAA	1380
	AATCAGACCC	TTATGGTTGG	ATCATCAGCC	ATCCTTCCAT	GTCAGGCTAG	CGGAAAACCA	1440
	ACTCCAGGAA	TATCATGGCT	CAGGGATGGG	CTACCTATTG	ACATTACAGA	TAGTCGTATC	1500
	AGTCAACATT	CAACGGGAAG	TCTACATATT	GCCGATTTAA	AGAAACCTGA	CACCGGAGTT	1560
	TACTACTTGA	TTGCGAAGAA	CGAGGATGGA	GAGTCAACAT	GGTCGGCATC	TCTGACTGTT	1620
45	GAAGATCACA	CTAGCAATGC	ACAATTTGTT	CGGATGCCGG	ATCCATCGAA	CTTCCCCTCT	1680
	TCTCCAACGC	AACCCATTAT	TGTCAATGTC	ACTGATACCG	AAGTAGAGCT	CCACTGGAAT	1740
	GCTCCCTCCA	CATCTGGCGC	AGGACCAATC	ACTGGTTATA	TCATTACAGTA	CTACAGTCCA	1800
	GACCTCGGAC	AGACGTGGTT	TAACATTCCA	GACTACGTGG	CATCTACTGA	ATATAGAATA	1860
	AAGGGTCTGA	AACCATCTCA	CTCGTATATG	TTTGTGATTC	GAGCAGAAAA	TGAGAAAGGT	1920
50	ATTGGAACGC	CGAGTGTGTC	GTCGGCTCTC	GTTACCACTA	GCAAGCCAGC	AGCTCAAGTT	1980
	GCGCTTTCTG	ACAAGAACAA	AATGGACATG	GCCATCGCTG	AGAAGAGACT	CACTTCGGAA	2040
	CAACTCATAA	AACTCGAGGA	AGTGAAGACT	ATTAATTCTA	CGGCCGTTCC	TTTGTCTCTG	2100
	AAGAAGAGGA	AACTTGAAGA	GCTGATTGAT	GGTACTTACA	TCAAGTGGAG	AGGGCCTCCA	2160
	AGAACCAATG	ATAATCAATA	CGTGAATGTG	ACCAGCCCTA	GCACCGAAAA	CTATGTTGTT	2220
55	TCAAATTTAA	TGCCATTACAC	CAACTATGAG	TTTTTCGTGA	TTCTTATCA	TTCCGGAGTT	2280

WO 99/25833

PCT/US98/24327

CATAGTATTC ATGGAGCACC GAGTAATTCC ATGGACGTGT TGACCGCCGA AGCTCCACCT 2340  
 TCATTGCCAC CAGAGGATGT GCGAATCCGT ATGCTCAACC TGACCACTCT TCGTATCTCT 2400  
 TGGAAAGCAC CAAAAGCCGA CGGCATCAAC GGAATTCTCA AAGGATTCCA AATTGTTATT 2460  
 GTTGGTCAAG CGCCAACAA CAATCGGAAC ATCACTACAA ACGAGAGAGC TGCCAGTGTT 2520  
 5 ACTCTGTTCC ATTTAGTGAC TGGAAATGACG TATAAAATTC GTGTAGCGGC TAGAAGCAAT 2580  
 GGTGGAGTTG GAGTCTCACA TGGAACGAGT GAAGTCATCA TGAATCAAGA CACGCTGGAA 2640  
 AAACACCTTG CTGCTCAACA AGAAAACGAA TCATTTTTGT ATGGGCTGAT CAATAAATCT 2700  
 CATGTTCCCTG TGATTGTCAT TGTTGCAATT CTGATTATTT TCGTAGTCAT CATTATAGCC 2760  
 TATTGTTACT GGAGGAATAG CAGAAACAGT GATGGAAAGG ATCGAAGTTT TATAAAGATC 2820  
 10 AATGATGGAA GTGTTTCATAT GGCTTCGAAT AATCTTTGGG ATGTTGCACA AAATCCGAAT 2880  
 CAGAATCCAA TGTACAACAC TGCTGGAAGA ATGACTATGA ACAATAGAAA TGGCCAGGCT 2940  
 CTCTATTCGC TGACACCAA TCGCAAGAC TTTTTCACA ATTGTGATGA CTACAGTGGA 3000  
 ACGATGCACA GACCAGGATC CGAGCATCAC TATCATTATG CTCAACTGAC TGGCGGACCT 3060  
 GGTAATGCGA TGTCTACTTT TTATGGAAAC CAATATCAGC ATGATCCATC TCCATATGCC 3120  
 15 ACCACAACAC TGGTCCTGTC GAACCAACAA CCAGCTTGGC TCAATGACAA AATGCTTCGC 3180  
 GCGCCAGCAA TGCCAACAAA TCCCGTGCCA CCAGAGCCAC CGGCGCGATA TGCAGATCAT 3240  
 ACCGCTGGAA GACGATCTCG ATCGAGCCGT GCATCCGATG GGAGAGGAAC TCTGAATGGC 3300  
 GGACTCCATC ACCGGACTAG CGGAAGTCAA CGGTCCGATA GTCCACCTCA CACAGATGTG 3360  
 AGCTATGTTT AGCTTCACTC ATCCGATGGA ACTGGTAGTA GTAAGGAAAG AACTGGGGAG 3420  
 20 CGGAGAACAC CACCGAATAA GACTCTGATG GACTTTATTC CGCCACCACC TTCCAATCCA 3480  
 CCACCACCTG GAGGGCACGT TTATGACACA GCAACTAGGC GTCAGTTGAA TCGTGGAAAGT 3540  
 ACTCCACGAG AAGACACCTA CGATTCGGTC AGTGACGGAG CTTTTGCTCG GGTGATGTG 3600  
 AATGCAAGGC CAACGAGTCG GAATCGGAAT TTGGGAGGAA GGCCGCTGAA AGGGAAACGA 3660  
 GACGACGATA GTCAGCGGTC TTCGTTGATG ATGGACGATG ATGGTGGATC TTCTGAAGCT 3720  
 25 GACGGGGAGA ACTCTGAAGG AGACGTTCCG CGTGGAGGTG TTAGAAAAGC AGTTCCTCGA 3780  
 ATGGGTATCT CTGCAAGTAC GCTGGCTCAT AGTTGTTACG GGACAAACGG CACTGCTCAA 3840  
 CGATTCCGGT CAATTCCACG TAACAATGGA ATCGTCACAC AAGAACAAAC TTGA 3894

(2) INFORMATION FOR SEQ ID NO:6:  
 30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1297 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 35 (ii) MOLECULE TYPE: peptide  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:  
 Met Tyr Tyr Leu Gly Phe Tyr His Thr His Thr His Thr His Thr Tyr  
 1 5 10 15  
 40 Ile Asn Phe Asp Lys Ile Pro Asn Ala Ser Asn Leu Ala Pro Val Ile  
 20 25 30  
 Ile Glu His Pro Ile Asp Val Val Val Ser Arg Gly Ser Pro Ala Thr  
 35 40 45  
 Leu Asn Cys Gly Ala Lys Pro Ser Thr Ala Lys Ile Thr Trp Tyr Lys  
 50 55 60  
 45 Asp Gly Gln Pro Val Ile Thr Asn Lys Glu Gln Val Asn Ser His Arg  
 65 70 75 80  
 Ile Val Leu Asp Thr Gly Ser Leu Phe Leu Leu Lys Val Asn Ser Gly  
 85 90 95  
 50 Lys Asn Gly Lys Asp Ser Asp Ala Gly Ala Tyr Tyr Cys Val Ala Ser  
 100 105 110  
 Asn Glu His Gly Glu Val Lys Ser Asn Glu Gly Ser Leu Lys Leu Ala  
 115 120 125  
 Met Leu Arg Glu Asp Phe Arg Val Arg Pro Arg Thr Val Gln Ala Leu  
 130 135 140  
 55 Gly Gly Glu Met Ala Val Leu Glu Cys Ser Pro Pro Arg Gly Phe Pro

WO 99/25833

PCT/US98/24327

	145				150				155				160			
	Glu	Pro	Val	Val	Ser	Trp	Arg	Lys	Asp	Asp	Lys	Glu	Leu	Arg	Ile	Gln
					165				170						175	
5	Asp	Met	Pro	Arg	Tyr	Thr	Leu	His	Ser	Asp	Gly	Asn	Leu	Ile	Ile	Asp
				180					185					190		
	Pro	Val	Asp	Arg	Ser	Asp	Ser	Gly	Thr	Tyr	Gln	Cys	Val	Ala	Asn	Asn
			195					200					205			
	Met	Val	Gly	Glu	Arg	Val	Ser	Asn	Pro	Ala	Arg	Leu	Ser	Val	Phe	Glu
	210							215				220				
10	Lys	Pro	Lys	Phe	Glu	Gln	Glu	Pro	Lys	Asp	Met	Thr	Val	Asp	Val	Gly
	225					230					235				240	
	Ala	Ala	Val	Leu	Phe	Asp	Cys	Arg	Val	Thr	Gly	Asp	Pro	Gln	Pro	Gln
					245					250					255	
15	Ile	Thr	Trp	Lys	Arg	Lys	Asn	Glu	Pro	Met	Pro	Val	Thr	Arg	Ala	Tyr
				260					265					270		
	Ile	Ala	Lys	Asp	Asn	Arg	Gly	Leu	Arg	Ile	Glu	Arg	Val	Gln	Pro	Ser
			275					280					285			
	Asp	Glu	Gly	Glu	Tyr	Val	Cys	Tyr	Ala	Arg	Asn	Pro	Ala	Gly	Thr	Leu
	290						295					300				
20	Glu	Ala	Ser	Ala	His	Leu	Arg	Val	Gln	Ala	Pro	Pro	Ser	Phe	Gln	Thr
	305					310					315					320
	Lys	Pro	Ala	Asp	Gln	Ser	Val	Pro	Ala	Gly	Gly	Thr	Ala	Thr	Phe	Glu
					325					330					335	
25	Cys	Thr	Leu	Val	Gly	Gln	Pro	Ser	Pro	Ala	Tyr	Phe	Trp	Ser	Lys	Glu
				340					345					350		
	Gly	Gln	Gln	Asp	Leu	Leu	Phe	Pro	Ser	Tyr	Val	Ser	Ala	Asp	Gly	Arg
			355					360					365			
30	Thr	Lys	Val	Ser	Pro	Thr	Gly	Thr	Leu	Thr	Ile	Glu	Glu	Val	Arg	Gln
	370						375					380				
	Val	Asp	Glu	Gly	Ala	Tyr	Val	Cys	Ala	Gly	Met	Asn	Ser	Ala	Gly	Ser
	385					390					395					400
	Ser	Leu	Ser	Lys	Ala	Ala	Leu	Lys	Ala	Thr	Phe	Glu	Thr	Lys	Gly	Arg
				405						410				415		
35	Val	Gln	Lys	Lys	Lys	Ser	Lys	Met	Gly	Lys	Gln	Lys	Gln	Lys	Asn	Val
				420					425					430		
	Gln	Ser	Ile	Ile	Lys	Tyr	Leu	Ile	Ser	Ala	Val	Thr	Gly	Asn	Thr	Pro
			435					440					445			
	Ala	Lys	Pro	Pro	Pro	Thr	Ile	Glu	His	Gly	His	Gln	Asn	Gln	Thr	Leu
	450						455					460				
40	Met	Val	Gly	Ser	Ser	Ala	Ile	Leu	Pro	Cys	Gln	Ala	Ser	Gly	Lys	Pro
	465					470					475					480
	Thr	Pro	Gly	Ile	Ser	Trp	Leu	Arg	Asp	Gly	Leu	Pro	Ile	Asp	Ile	Thr
					485					490				495		
45	Asp	Ser	Arg	Ile	Ser	Gln	His	Ser	Thr	Gly	Ser	Leu	His	Ile	Ala	Asp
				500					505					510		
	Leu	Lys	Lys	Pro	Asp	Thr	Gly	Val	Tyr	Thr	Cys	Ile	Ala	Lys	Asn	Glu
			515					520					525			
	Asp	Gly	Glu	Ser	Thr	Trp	Ser	Ala	Ser	Leu	Thr	Val	Glu	Asp	His	Thr
	530						535					540				
50	Ser	Asn	Ala	Gln	Phe	Val	Arg	Met	Pro	Asp	Pro	Ser	Asn	Phe	Pro	Ser
	545					550					555					560
	Ser	Pro	Thr	Gln	Pro	Ile	Ile	Val	Asn	Val	Thr	Asp	Thr	Glu	Val	Glu
				565						570				575		
55	Leu	His	Trp	Asn	Ala	Pro	Ser	Thr	Ser	Gly	Ala	Gly	Pro	Ile	Thr	Gly
				580					585					590		

WO 99/25833

PCT/US98/24327

	Tyr	Ile	Ile	Gln	Tyr	Tyr	Ser	Pro	Asp	Leu	Gly	Gln	Thr	Trp	Phe	Asn
			595					600					605			
	Ile	Pro	Asp	Tyr	Val	Ala	Ser	Thr	Glu	Tyr	Arg	Ile	Lys	Gly	Leu	Lys
		610					615					620				
5	Pro	Ser	His	Ser	Tyr	Met	Phe	Val	Ile	Arg	Ala	Glu	Asn	Glu	Lys	Gly
	625					630					635					640
	Ile	Gly	Thr	Pro	Ser	Val	Ser	Ser	Ala	Leu	Val	Thr	Thr	Ser	Lys	Pro
					645					650					655	
10	Ala	Ala	Gln	Val	Ala	Leu	Ser	Asp	Lys	Asn	Lys	Met	Asp	Met	Ala	Ile
				660					665					670		
	Ala	Glu	Lys	Arg	Leu	Thr	Ser	Glu	Gln	Leu	Ile	Lys	Leu	Glu	Glu	Val
			675					680					685			
	Lys	Thr	Ile	Asn	Ser	Thr	Ala	Val	Arg	Leu	Phe	Trp	Lys	Lys	Arg	Lys
		690					695					700				
15	Leu	Glu	Glu	Leu	Ile	Asp	Gly	Tyr	Tyr	Ile	Lys	Trp	Arg	Gly	Pro	Pro
	705					710					715					720
	Arg	Thr	Asn	Asp	Asn	Gln	Tyr	Val	Asn	Val	Thr	Ser	Pro	Ser	Thr	Glu
					725					730					735	
20	Asn	Tyr	Val	Val	Ser	Asn	Leu	Met	Pro	Phe	Thr	Asn	Tyr	Glu	Phe	Phe
				740					745					750		
	Val	Ile	Pro	Tyr	His	Ser	Gly	Val	His	Ser	Ile	His	Gly	Ala	Pro	Ser
			755					760					765			
	Asn	Ser	Met	Asp	Val	Leu	Thr	Ala	Glu	Ala	Pro	Pro	Ser	Leu	Pro	Pro
		770					775					780				
25	Glu	Asp	Val	Arg	Ile	Arg	Met	Leu	Asn	Leu	Thr	Thr	Leu	Arg	Ile	Ser
	785					790					795					800
	Trp	Lys	Ala	Pro	Lys	Ala	Asp	Gly	Ile	Asn	Gly	Ile	Leu	Lys	Gly	Phe
					805					810					815	
30	Gln	Ile	Val	Ile	Val	Gly	Gln	Ala	Pro	Asn	Asn	Asn	Arg	Asn	Ile	Thr
				820					825					830		
	Thr	Asn	Glu	Arg	Ala	Ala	Ser	Val	Thr	Leu	Phe	His	Leu	Val	Thr	Gly
			835					840					845			
	Met	Thr	Tyr	Lys	Ile	Arg	Val	Ala	Ala	Arg	Ser	Asn	Gly	Gly	Val	Gly
		850				855						860				
35	Val	Ser	His	Gly	Thr	Ser	Glu	Val	Ile	Met	Asn	Gln	Asp	Thr	Leu	Glu
	865					870					875					880
	Lys	His	Leu	Ala	Ala	Gln	Gln	Glu	Asn	Glu	Ser	Phe	Leu	Tyr	Gly	Leu
					885					890					895	
40	Ile	Asn	Lys	Ser	His	Val	Pro	Val	Ile	Val	Ile	Val	Ala	Ile	Leu	Ile
					900				905					910		
	Ile	Phe	Val	Val	Ile	Ile	Ile	Ala	Tyr	Cys	Tyr	Trp	Arg	Asn	Ser	Arg
			915					920					925			
	Asn	Ser	Asp	Gly	Lys	Asp	Arg	Ser	Phe	Ile	Lys	Ile	Asn	Asp	Gly	Ser
		930					935						940			
45	Val	His	Met	Ala	Ser	Asn	Asn	Leu	Trp	Asp	Val	Ala	Gln	Asn	Pro	Asn
	945					950					955					960
	Gln	Asn	Pro	Met	Tyr	Asn	Thr	Ala	Gly	Arg	Met	Thr	Met	Asn	Asn	Arg
					965					970					975	
50	Asn	Gly	Gln	Ala	Leu	Tyr	Ser	Leu	Thr	Pro	Asn	Ala	Gln	Asp	Phe	Phe
				980					985					990		
	Asn	Asn	Cys	Asp	Asp	Tyr	Ser	Gly	Thr	Met	His	Arg	Pro	Gly	Ser	Glu
			995					1000					1005			
	His	His	Tyr	His	Tyr	Ala	Gln	Leu	Thr	Gly	Gly	Pro	Gly	Asn	Ala	Met
		1010					1015					1020				
55	Ser	Thr	Phe	Tyr	Gly	Asn	Gln	Tyr	His	Asp	Asp	Pro	Ser	Pro	Tyr	Ala

WO 99/25833

PCT/US98/24327

	1025			1030				1035			1040					
	Thr	Thr	Thr	Leu	Val	Leu	Ser	Asn	Gln	Gln	Pro	Ala	Trp	Leu	Asn	Asp
				1045				1050			1055					
5	Lys	Met	Leu	Arg	Ala	Pro	Ala	Met	Pro	Thr	Asn	Pro	Val	Pro	Pro	Glu
			1060					1065			1070					
	Pro	Pro	Ala	Arg	Tyr	Ala	Asp	His	Thr	Ala	Gly	Arg	Arg	Ser	Arg	Ser
			1075					1080			1085					
	Ser	Arg	Ala	Ser	Asp	Gly	Arg	Gly	Thr	Leu	Asn	Gly	Gly	Leu	His	His
		1090					1095				1100					
10	Arg	Thr	Ser	Gly	Ser	Gln	Arg	Ser	Asp	Ser	Pro	Pro	His	Thr	Asp	Val
		1105				1110					1115					1120
	Ser	Tyr	Val	Gln	Leu	His	Ser	Ser	Asp	Gly	Thr	Gly	Ser	Ser	Lys	Glu
			1125							1130						1135
15	Arg	Thr	Gly	Glu	Arg	Arg	Thr	Pro	Pro	Asn	Lys	Thr	Leu	Met	Asp	Phe
			1140							1145						1150
	Ile	Pro	Pro	Pro	Pro	Ser	Asn	Pro	Pro	Pro	Pro	Gly	Gly	His	Val	Tyr
		1155						1160								1165
	Asp	Thr	Ala	Thr	Arg	Arg	Gln	Leu	Asn	Arg	Gly	Ser	Thr	Pro	Arg	Glu
		1170					1175					1180				
20	Asp	Thr	Tyr	Asp	Ser	Val	Ser	Asp	Gly	Ala	Phe	Ala	Arg	Val	Asp	Val
		1185				1190					1195					1200
	Asn	Ala	Arg	Pro	Thr	Ser	Arg	Asn	Arg	Asn	Leu	Gly	Gly	Arg	Pro	Leu
			1205							1210						1215
25	Lys	Gly	Lys	Arg	Asp	Asp	Asp	Ser	Gln	Arg	Ser	Ser	Leu	Met	Met	Asp
			1220							1225						1230
	Asp	Asp	Gly	Gly	Ser	Ser	Glu	Ala	Asp	Gly	Glu	Asn	Ser	Glu	Gly	Asp
			1235					1240					1245			
	Val	Pro	Arg	Gly	Gly	Val	Arg	Lys	Ala	Val	Pro	Arg	Met	Gly	Ile	Ser
		1250					1255						1260			
30	Ala	Ser	Thr	Leu	Ala	His	Ser	Cys	Tyr	Gly	Thr	Asn	Gly	Thr	Ala	Gln
		1265				1270					1275					1280
	Arg	Phe	Arg	Ser	Ile	Pro	Arg	Asn	Asn	Gly	Ile	Val	Thr	Gln	Glu	Gln
			1285							1290						1295
35	Thr															

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4956 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

45	ATGAAATGGA	AACATGTTCC	TTTTTTGGTC	ATGATATCAC	TCCTCAGCTT	ATCCCCAAAT	60
	CACCTGTTTC	TGGCCCAGCT	TATTCCAGAC	CCTGAAGATG	TAGAGAGGGG	GAACGACCAC	120
	GGGACGCCAA	TCCCCACCTC	TGATAACGAT	GACAATTCGC	TGGGCTATAC	AGGCTCCCGT	180
	CTTCGTCAGG	AAGATTTTCC	ACCTCGCATT	GTTGAACACC	CTTCAGACCT	GATTGTCTCA	240
	AAAGGAGAAC	CTGCAACTTT	GAAGTGCAAA	GCTGAAGGCC	GCCCCACACC	CACTATTGAA	300
	TGGTACAAAG	GGGGAGAGAG	AGTGGAGACA	GACAAAGATG	ACCCTCGCTC	ACACCGAATG	360
50	TTGCTGCCGA	GTGGATCTTT	ATTTTCTTCTTA	CGTATAGTAC	ATGGACGGAA	AAGTAGACCT	420
	GATGAAGGAG	TCTATGTCTG	TGTAGCAAGG	AATTACCTTG	GAGAGGCTGT	GAGCCACAAT	480
	GCATCGCTGG	AAGTAGCCAT	ACTTCGGGAT	GACTTCAGAC	AAAACCCTTC	GGATGTCATG	540
	GTTGCAGTAG	GAGAGCCTGC	AGTAATGGAA	TGCCAACCTC	CACGAGGCCA	TCCTGAGCCC	600
	ACCATTTTCAT	GGAAGAAAGA	TGGCTCTCCA	CTGGATGATA	AAGATGAAAG	AATAACTATA	660
55	CGAGGAGGAA	AGCTCATGAT	CACTTACACC	CGTAAAAGTG	ACGCTGGCAA	ATATGTTTGT	720

WO 99/25833

PCT/US98/24327

	GT TGGTACCA	ATATGGTTGG	GGAACGTGAG	AGTGAAGTAG	CCGAGCTGAC	TGTCTTAGAG	780
	AGACCATCAT	TTGTGAAGAG	ACCCAGTAAC	TTGGCAGTAA	CTGTGGATGA	CAGTGCAGAA	840
	TTTAAATGTG	AGGCCCGAGG	TGACCCTGTA	CCTACAGTAC	GATGGAGGAA	AGATGATGGA	900
	GAGCTGCCCA	AATCCAGATA	TGAAATCCGA	GATGATCATA	CCTTGAAAAT	TAGGAAGGTG	960
5	ACAGCTGGTG	ACATGGGTTT	ATACACTTGT	GTTGCAGAAA	ATATGGTGGG	CAAAGCTGAA	1020
	GCATCTGCTA	CTCTGACTGT	TCAAGAACCT	CCACATTTTG	TTGTGAAACC	CCGTGACCAG	1080
	GTTGTTGCTT	TGGGACGGAC	TGTAAC TTTT	CAGTGTGAAG	CAACCGGAAA	TCCTCAACCA	1140
	GCTATTTTCT	GGAGGAGAGA	AGGGAGTCAG	AATCTACTTT	TCTCATATCA	ACCACCACAG	1200
	TCATCCAGCC	GATTTTCAGT	CTCCCAGACT	GGCGACCTCA	CAATTACTAA	TGTCCAGCGA	1260
10	TCTGATGTTG	GTTATTACAT	CTGCCAGACT	TTAAATGTTG	CTGGAAGCAT	CATCACAAAG	1320
	GCATATTTGG	AAGTTACAGA	TGTGATTGCA	GATCGGCCTC	CCCCAGTTAT	TCGACAAGGT	1380
	CCTGTGAATC	AGACTGTAGC	CGTGGATGGC	ACTTTCGTCC	TCAGCTGTGT	GGCCACAGGC	1440
	AGTCCAGTGC	CCACCATTCT	GTGGAGAAAG	GATGGAGTCC	TCGTTTCAAC	CCAAGACTCT	1500
	CGAATCAAAC	AGTTGGAGAA	TGGAGTACTG	CAGATCCGAT	ATGCTAAGCT	GGGTGATACT	1560
15	GGTCGGTACA	CCTGCATTGC	ATCAACCCCC	AGTGGTGAAG	CAACATGGAG	TGCTTACATT	1620
	GAAGTTCAAG	AATTTGGAGT	TCCAGTTCAG	CCTCCAAGAC	CTACTGACCC	AAATTTAATC	1680
	CCTAGTGCCC	CATCAAACCC	TGAAGTGACA	GATGTCAGCA	GAAATACAGT	CACATTATCG	1740
	TGGCAACCAA	ATTTGAATTC	AGGAGCAACT	CCAACATCTT	ATATTATAGA	AGCCTTCAGC	1800
	CATGCATCTG	GTAGCAGCTG	GCAGACCGTA	GCAGAGAATG	TGAAAACAGA	AACATCTGCC	1860
20	ATTAAAGGAC	TCAAACCTAA	TGCAATTTAC	CTTTTCCTTG	TGAGGGCAGC	TAATGCATAT	1920
	GGAATTAGTG	ATCCAAGCCA	AATATCAGAT	CCAGTGAAAA	CACAAGATGT	CCTACCAACA	1980
	AGTCAGGGGG	TGGACCACAA	GCAGGTCCAG	AGAGAGCTGG	GAAATGCTGT	TCTGCACCTC	2040
	CACAACCCCA	CCGTCCTTTC	TTCCTCTTCC	ATCGAAGTGC	ACTGGACAGT	AGATCAACAG	2100
	TCTCAGTATA	TACAAGGATA	TAAAATTCTC	TATCGGCCAT	CTGGAGCCAA	CCACGGAGAA	2160
25	TCAGACTGGT	TAGTTTTTGA	AGTGAGGACG	CCAGCCAAAA	ACAGTGTGGT	AATCCCTGAT	2220
	CTCAGAAAGG	GAGTCAACTA	TGAAATTAAG	GCTCGCCCTT	TTTTTAATGA	ATTTCAAGGA	2280
	GCAGATAGTG	AAATCAAGTT	TGCCAAAACC	CTGGAAGAAG	CACCCAGTGC	CCCACCCCAA	2340
	GGTGTAAGTG	TATCCAAGAA	TGATGGAAAC	GGAAGTCAA	TTCTAGTTAG	TTGGCAGCCA	2400
	CCTCCAGAAG	ACACTCAAAA	TGGAATGGTC	CAAGAGTATA	AGGTTTGGTG	TCTGGGCAAT	2460
30	GAAACTCGAT	ACCACATCAA	CAAAACAGTG	GATGGTTCCA	CCTTTTCCGT	GGTCATTCCC	2520
	TTTCTTGTTT	CTGGAATCCG	ATACAGTGTG	GAAGTGGCAG	CCAGCACTGG	GGCTGGGTCT	2580
	GGGGTAAAGA	GTGAGCCTCA	GTTTATCCAG	CTGGATGCCC	ATGGAAACCC	TGTGTCACCT	2640
	GAGGACCAAG	TCAGCCTCGC	TCAGCAGATT	TCAGATGTGG	TGAAGCAGCC	GGCCTTCATA	2700
	GCAGGTATTG	GAGCAGCCTG	TTGGATCATC	CTCATGGTCT	TCAGCATCTG	GCTTTATCGA	2760
35	CACCGCAAGA	AGAGAAACGG	ACTTACTAGT	ACCTACGCGG	GTATCAGAAA	AGTCCCGTCT	2820
	TTTACCTTCA	CACCAACAGT	AACTTACCAG	AGAGGAGGCG	AAGCTGTCAG	CAGTGGAGGG	2880
	AGGCCTGGAC	TTCTCAACAT	CAGTGAACCT	GCCGCGCAGC	CATGGCTGGC	AGACACGTGG	2940
	CCTAATACTG	GCAACAACCA	CAATGACTGC	TCCATCAGCT	GCTGCACGGC	AGGCAATGGA	3000
	AACAGCGACA	GCAACCTCAC	TACCTACAGT	CGCCCAGCTG	ATTGTATAGC	AAATTATAAC	3060
40	AACCAACTGG	ATAACAAACA	AACAAATCTG	ATGCTCCCTG	AGTCAACTGT	TTATGGTGAT	3120
	GTGGACCTTA	GTAACAAAAT	CAATGAGATG	AAAACCTTCA	ATAGCCCAA	TCTGAAGGAT	3180
	GGGCGTTTTG	TCAATCCATC	AGGGCAGCCT	ACTCCTTACG	CCACCACTCA	GCTCATCCAG	3240
	TCAAACCTCA	GCAACAACAT	GAACAATGGC	AGCGGGGACT	CTGGCGAGAA	GCACTGGAAA	3300
	CCACTGGGAC	AGCAGAAACA	AGAAGTGGCA	CCAGTTCAGT	ACAACATCGT	GGAGCAAAC	3360
45	AAGCTGAACA	AAGATTATCG	AGCAAATGAC	ACAGTTCCTC	CAACTATCCC	ATACAACCAA	3420
	TCATACGACC	AGAACACAGG	AGGATCCTAC	AACAGCTCAG	ACCGGGGCAG	TAGTACATCT	3480
	GGGAGTCAGG	GGCACAAGAA	AGGGGCAAGA	ACACCCAAGG	TACCAAACA	GGGTGGCATG	3540
	AACTGGGCAG	ACCTGCTTCC	TCC TCCCCCA	GCACATCCTC	CTCCACACAG	CAATAGCGAA	3600
	GAGTACAACA	TTTCTGTAGA	TGAAAGCTAT	GACCAAGAAA	TGCCATGTCC	CGTGCCACCA	3660
50	GCAAGGATGT	ATTTGCAACA	AGATGAATTA	GAAGAGGAGG	AAGATGAACG	AGGCCCACT	3720
	CCCCCTGTTC	GGGGAGCAGC	TTCTTCTCCA	GCTGCCGTGT	CCTATAGCCA	TCAGTCCACT	3780
	GCCACTCTGA	CTCCCTCCCC	ACAGGAAGAA	CTCCAGCCCA	TGTTACAGGA	TTGTCCAGAG	3840
	GAGACTGGCC	ACATGCAGCA	CCAGCCCGAC	AGGAGACGGC	AGCCTGTGAG	TCCTCCTCCA	3900
	CCACCACGGC	CGATCTCCCC	TCCACATACC	TATGGCTACA	TTTCAGGACC	CCTGGTCTCA	3960
55	GATATGGATA	CGGATGCGCC	AGAAGAGGAA	GAAGACGAAG	CCGACATGGA	GGTAGCCAAG	4020

WO 99/25833

PCT/US98/24327

ATGCAAACCA GAAGGCTTTT GTTACGTGGG CTTGAGCAGA CACCTGCCTC CAGTGTTGGG 4080  
 GACCTGGAGA GCTCTGTCAC GGGGTCCATG ATCAACGGCT GGGGCTCAGC CTCAGAGGAG 4140  
 GACAACATTT CCAGCGGACG CTCCAGTGTT AGTTCTTCGG ACGGCTCCTT TTTCAGTGAT 4200  
 GCTGACTTTG CCCAGGCAGT CGCAGCAGCG GCAGAGTATG CTGGTCTGAA AGTAGCACGA 4260  
 5 CGGCAAATGC AGGATGCTGC TGGCCGTCGA CATTTTCATG CGTCTCAGTG CCCTAGGCC 4320  
 ACAAGTCCCG TGTCTACAGA CAGCAACATG AGTGCCGCCG TAATGCAGAA AACCAGACCA 4380  
 GCCAAGAAAC TGAAACACCA GCCAGGACAT CTGCGCAGAG AAACCTACAC AGATGATCCT 4440  
 CCACCACCTC CTGTGCCGCC ACCTGCTATA AAGTCACCTA CTGCCCAATC CAAGACACAG 4500  
 CTGGAAGTAC GACCTGTAGT GGTGCCAAA CTCCCTTCTA TGGATGCAAG AACAGACAGA 4560  
 10 TCATCAGACA GAAAAGGAAG CAGTTACAAG GGGAGAGAAG TGTTGGATGG AAGACAGGTT 4620  
 GTTGACATGC GAACAAATCC AGGTGATCCC AGAGAAGCAC AGGAACAGCA AAATGACGGG 4680  
 AAAGGACGTG GAAACAAGGC AGCAAAACGA GACCTTCCAC CAGCAAAGAC TCATCTCATC 4740  
 CAAGAGGATA TTCTACCTTA TTGTAGACCT ACTTTTCCAA CATCAAATAA TCCCAGAGAT 4800  
 CCCAGTTCCT CAAGCTCAAT GTCATCAAGA GGATCAGGAA GCAGACAAAG AGAACAAGCA 4860  
 15 AATGTAGGTC GAAGAAATAT TGCAGAAATG CAGGTAATTG GAGGATATGA AAGAGGAGAA 4920  
 GATAATAATG AAGAATTAGA GGAAACTGAA AGCTGA 4956

## (2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:  
 20 (A) LENGTH: 1651 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: peptide  
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:  
 Met Lys Trp Lys His Val Pro Phe Leu Val Met Ile Ser Leu Leu Ser  
 1 5 10 15  
 Leu Ser Pro Asn His Leu Phe Leu Ala Gln Leu Ile Pro Asp Pro Glu  
 20 25 30  
 30 Asp Val Glu Arg Gly Asn Asp His Gly Thr Pro Ile Pro Thr Ser Asp  
 35 40 45  
 Asn Asp Asp Asn Ser Leu Gly Tyr Thr Gly Ser Arg Leu Arg Gln Glu  
 50 55 60  
 Asp Phe Pro Pro Arg Ile Val Glu His Pro Ser Asp Leu Ile Val Ser  
 35 65 70 75 80  
 Lys Gly Glu Pro Ala Thr Leu Asn Cys Lys Ala Glu Gly Arg Pro Thr  
 85 90 95  
 Pro Thr Ile Glu Trp Tyr Lys Gly Gly Glu Arg Val Glu Thr Asp Lys  
 100 105 110  
 40 Asp Asp Pro Arg Ser His Arg Met Leu Leu Pro Ser Gly Ser Leu Phe  
 115 120 125  
 Phe Leu Arg Ile Val His Gly Arg Lys Ser Arg Pro Asp Glu Gly Val  
 130 135 140  
 Tyr Val Cys Val Ala Arg Asn Tyr Leu Gly Glu Ala Val Ser His Asn  
 45 145 150 155 160  
 Ala Ser Leu Glu Val Ala Ile Leu Arg Asp Asp Phe Arg Gln Asn Pro  
 165 170 175  
 Ser Asp Val Met Val Ala Val Gly Glu Pro Ala Val Met Glu Cys Gln  
 180 185 190  
 50 Pro Pro Arg Gly His Pro Glu Pro Thr Ile Ser Trp Lys Lys Asp Gly  
 195 200 205  
 Ser Pro Leu Asp Asp Lys Asp Glu Arg Ile Thr Ile Arg Gly Gly Lys  
 210 215 220  
 55 Leu Met Ile Thr Tyr Thr Arg Lys Ser Asp Ala Gly Lys Tyr Val Cys  
 225 230 235 240

WO 99/25833

PCT/US98/24327

	Val	Gly	Thr	Asn	Met	Val	Gly	Glu	Arg	Glu	Ser	Glu	Val	Ala	Glu	Leu
					245					250					255	
	Thr	Val	Leu	Glu	Arg	Pro	Ser	Phe	Val	Lys	Arg	Pro	Ser	Asn	Leu	Ala
				260					265					270		
5	Val	Thr	Val	Asp	Asp	Ser	Ala	Glu	Phe	Lys	Cys	Glu	Ala	Arg	Gly	Asp
			275					280					285			
	Pro	Val	Pro	Thr	Val	Arg	Trp	Arg	Lys	Asp	Asp	Gly	Glu	Leu	Pro	Lys
		290					295					300				
	Ser	Arg	Tyr	Glu	Ile	Arg	Asp	Asp	His	Thr	Leu	Lys	Ile	Arg	Lys	Val
10	305					310						315				320
	Thr	Ala	Gly	Asp	Met	Gly	Ser	Tyr	Thr	Cys	Val	Ala	Glu	Asn	Met	Val
					325					330					335	
	Gly	Lys	Ala	Glu	Ala	Ser	Ala	Thr	Leu	Thr	Val	Gln	Glu	Pro	Pro	His
				340					345					350		
15	Phe	Val	Val	Lys	Pro	Arg	Asp	Gln	Val	Val	Ala	Leu	Gly	Arg	Thr	Val
			355					360					365			
	Thr	Phe	Gln	Cys	Glu	Ala	Thr	Gly	Asn	Pro	Gln	Pro	Ala	Ile	Phe	Trp
		370					375					380				
	Arg	Arg	Glu	Gly	Ser	Gln	Asn	Leu	Leu	Phe	Ser	Tyr	Gln	Pro	Pro	Gln
20	385					390						395				400
	Ser	Ser	Ser	Arg	Phe	Ser	Val	Ser	Gln	Thr	Gly	Asp	Leu	Thr	Ile	Thr
					405						410				415	
	Asn	Val	Gln	Arg	Ser	Asp	Val	Gly	Tyr	Tyr	Ile	Cys	Gln	Thr	Leu	Asn
				420					425					430		
25	Val	Ala	Gly	Ser	Ile	Ile	Thr	Lys	Ala	Tyr	Leu	Glu	Val	Thr	Asp	Val
			435					440					445			
	Ile	Ala	Asp	Arg	Pro	Pro	Pro	Val	Ile	Arg	Gln	Gly	Pro	Val	Asn	Gln
		450					455					460				
	Thr	Val	Ala	Val	Asp	Gly	Thr	Phe	Val	Leu	Ser	Cys	Val	Ala	Thr	Gly
30	465					470						475				480
	Ser	Pro	Val	Pro	Thr	Ile	Leu	Trp	Arg	Lys	Asp	Gly	Val	Leu	Val	Ser
					485						490				495	
	Thr	Gln	Asp	Ser	Arg	Ile	Lys	Gln	Leu	Glu	Asn	Gly	Val	Leu	Gln	Ile
				500					505					510		
35	Arg	Tyr	Ala	Lys	Leu	Gly	Asp	Thr	Gly	Arg	Tyr	Thr	Cys	Ile	Ala	Ser
			515					520					525			
	Thr	Pro	Ser	Gly	Glu	Ala	Thr	Trp	Ser	Ala	Tyr	Ile	Glu	Val	Gln	Glu
		530					535					540				
	Phe	Gly	Val	Pro	Val	Gln	Pro	Pro	Arg	Pro	Thr	Asp	Pro	Asn	Leu	Ile
40	545					550						555				560
	Pro	Ser	Ala	Pro	Ser	Lys	Pro	Glu	Val	Thr	Asp	Val	Ser	Arg	Asn	Thr
					565						570				575	
	Val	Thr	Leu	Ser	Trp	Gln	Pro	Asn	Leu	Asn	Ser	Gly	Ala	Thr	Pro	Thr
				580					585					590		
45	Ser	Tyr	Ile	Ile	Glu	Ala	Phe	Ser	His	Ala	Ser	Gly	Ser	Ser	Trp	Gln
			595					600					605			
	Thr	Val	Ala	Glu	Asn	Val	Lys	Thr	Glu	Thr	Ser	Ala	Ile	Lys	Gly	Leu
		610					615					620				
	Lys	Pro	Asn	Ala	Ile	Tyr	Leu	Phe	Leu	Val	Arg	Ala	Ala	Asn	Ala	Tyr
50	625					630						635				640
	Gly	Ile	Ser	Asp	Pro	Ser	Gln	Ile	Ser	Asp	Pro	Val	Lys	Thr	Gln	Asp
					645						650				655	
	Val	Leu	Pro	Thr	Ser	Gln	Gly	Val	Asp	His	Lys	Gln	Val	Gln	Arg	Glu
				660						665				670		
55	Leu	Gly	Asn	Ala	Val	Leu	His	Leu	His	Asn	Pro	Thr	Val	Leu	Ser	Ser

WO 99/25833

PCT/US98/24327

		675				680				685						
	Ser	Ser	Ile	Glu	Val	His	Trp	Thr	Val	Asp	Gln	Gln	Ser	Gln	Tyr	Ile
		690					695					700				
5	Gln	Gly	Tyr	Lys	Ile	Leu	Tyr	Arg	Pro	Ser	Gly	Ala	Asn	His	Gly	Glu
	705					710					715					720
	Ser	Asp	Trp	Leu	Val	Phe	Glu	Val	Arg	Thr	Pro	Ala	Lys	Asn	Ser	Val
					725					730						735
	Val	Ile	Pro	Asp	Leu	Arg	Lys	Gly	Val	Asn	Tyr	Glu	Ile	Lys	Ala	Arg
				740					745							750
10	Pro	Phe	Phe	Asn	Glu	Phe	Gln	Gly	Ala	Asp	Ser	Glu	Ile	Lys	Phe	Ala
			755					760								765
	Lys	Thr	Leu	Glu	Glu	Ala	Pro	Ser	Ala	Pro	Pro	Gln	Gly	Val	Thr	Val
	770						775						780			
	Ser	Lys	Asn	Asp	Gly	Asn	Gly	Thr	Ala	Ile	Leu	Val	Ser	Trp	Gln	Pro
15	785					790						795				800
	Pro	Pro	Glu	Asp	Thr	Gln	Asn	Gly	Met	Val	Gln	Glu	Tyr	Lys	Val	Trp
					805					810						815
	Cys	Leu	Gly	Asn	Glu	Thr	Arg	Tyr	His	Ile	Asn	Lys	Thr	Val	Asp	Gly
				820					825							830
20	Ser	Thr	Phe	Ser	Val	Val	Ile	Pro	Phe	Leu	Val	Pro	Gly	Ile	Arg	Tyr
			835					840						845		
	Ser	Val	Glu	Val	Ala	Ala	Ser	Thr	Gly	Ala	Gly	Ser	Gly	Val	Lys	Ser
	850						855						860			
	Glu	Pro	Gln	Phe	Ile	Gln	Leu	Asp	Ala	His	Gly	Asn	Pro	Val	Ser	Pro
25	865					870						875				880
	Glu	Asp	Gln	Val	Ser	Leu	Ala	Gln	Gln	Ile	Ser	Asp	Val	Val	Lys	Gln
					885					890						895
	Pro	Ala	Phe	Ile	Ala	Gly	Ile	Gly	Ala	Ala	Cys	Trp	Ile	Ile	Leu	Met
				900					905							910
30	Val	Phe	Ser	Ile	Trp	Leu	Tyr	Arg	His	Arg	Lys	Lys	Arg	Asn	Gly	Leu
			915					920								925
	Thr	Ser	Thr	Tyr	Ala	Gly	Ile	Arg	Lys	Val	Pro	Ser	Phe	Thr	Phe	Thr
	930						935						940			
	Pro	Thr	Val	Thr	Tyr	Gln	Arg	Gly	Gly	Glu	Ala	Val	Ser	Ser	Gly	Gly
35	945					950						955				960
	Arg	Pro	Gly	Leu	Leu	Asn	Ile	Ser	Glu	Pro	Ala	Ala	Gln	Pro	Trp	Leu
					965					970						975
	Ala	Asp	Thr	Trp	Pro	Asn	Thr	Gly	Asn	Asn	His	Asn	Asp	Cys	Ser	Ile
				980					985							990
40	Ser	Cys	Cys	Thr	Ala	Gly	Asn	Gly	Asn	Ser	Asp	Ser	Asn	Leu	Thr	Thr
			995					1000					1005			
	Tyr	Ser	Arg	Pro	Ala	Asp	Cys	Ile	Ala	Asn	Tyr	Asn	Asn	Gln	Leu	Asp
	1010						1015						1020			
	Asn	Lys	Gln	Thr	Asn	Leu	Met	Leu	Pro	Glu	Ser	Thr	Val	Tyr	Gly	Asp
45	1025					1030						1035				1040
	Val	Asp	Leu	Ser	Asn	Lys	Ile	Asn	Glu	Met	Lys	Thr	Phe	Asn	Ser	Pro
					1045					1050						1055
	Asn	Leu	Lys	Asp	Gly	Arg	Phe	Val	Asn	Pro	Ser	Gly	Gln	Pro	Thr	Pro
				1060					1065							1070
50	Tyr	Ala	Thr	Thr	Gln	Leu	Ile	Gln	Ser	Asn	Leu	Ser	Asn	Asn	Met	Asn
			1075					1080						1085		
	Asn	Gly	Ser	Gly	Asp	Ser	Gly	Glu	Lys	His	Trp	Lys	Pro	Leu	Gly	Gln
	1090						1095						1100			
	Gln	Lys	Gln	Glu	Val	Ala	Pro	Val	Gln	Tyr	Asn	Ile	Val	Glu	Gln	Asn
55	1105					1110						1115				1120

WO 99/25833

PCT/US98/24327

	Lys	Leu	Asn	Lys	Asp	Tyr	Arg	Ala	Asn	Asp	Thr	Val	Pro	Pro	Thr	Ile
					1125					1130					1135	
	Pro	Tyr	Asn	Gln	Ser	Tyr	Asp	Gln	Asn	Thr	Gly	Gly	Ser	Tyr	Asn	Ser
				1140					1145					1150		
5	Ser	Asp	Arg	Gly	Ser	Ser	Thr	Ser	Gly	Ser	Gln	Gly	His	Lys	Lys	Gly
				1155					1160					1165		
	Ala	Arg	Thr	Pro	Lys	Val	Pro	Lys	Gln	Gly	Gly	Met	Asn	Trp	Ala	Asp
				1170					1175					1180		
10	Leu	Leu	Pro	Pro	Pro	Pro	Ala	His	Pro	Pro	Pro	His	Ser	Asn	Ser	Glu
	1185							1190				1195				1200
	Glu	Tyr	Asn	Ile	Ser	Val	Asp	Glu	Ser	Tyr	Asp	Gln	Glu	Met	Pro	Cys
					1205						1210				1215	
	Pro	Val	Pro	Pro	Ala	Arg	Met	Tyr	Leu	Gln	Gln	Asp	Glu	Leu	Glu	Glu
					1220					1225				1230		
15	Glu	Glu	Asp	Glu	Arg	Gly	Pro	Thr	Pro	Pro	Val	Arg	Gly	Ala	Ala	Ser
				1235					1240					1245		
	Ser	Pro	Ala	Ala	Val	Ser	Tyr	Ser	His	Gln	Ser	Thr	Ala	Thr	Leu	Thr
				1250					1255					1260		
20	Pro	Ser	Pro	Gln	Glu	Glu	Leu	Gln	Pro	Met	Leu	Gln	Asp	Cys	Pro	Glu
	1265					1270						1275				1280
	Glu	Thr	Gly	His	Met	Gln	His	Gln	Pro	Asp	Arg	Arg	Arg	Gln	Pro	Val
					1285						1290				1295	
	Ser	Pro	Pro	Pro	Pro	Pro	Arg	Pro	Ile	Ser	Pro	Pro	His	Thr	Tyr	Gly
					1300					1305				1310		
25	Tyr	Ile	Ser	Gly	Pro	Leu	Val	Ser	Asp	Met	Asp	Thr	Asp	Ala	Pro	Glu
				1315					1320					1325		
	Glu	Glu	Glu	Asp	Glu	Ala	Asp	Met	Glu	Val	Ala	Lys	Met	Gln	Thr	Arg
				1330				1335					1340			
30	Arg	Leu	Leu	Leu	Arg	Gly	Leu	Glu	Gln	Thr	Pro	Ala	Ser	Ser	Val	Gly
	1345					1350						1355				1360
	Asp	Leu	Glu	Ser	Ser	Val	Thr	Gly	Ser	Met	Ile	Asn	Gly	Trp	Gly	Ser
					1365						1370				1375	
	Ala	Ser	Glu	Glu	Asp	Asn	Ile	Ser	Ser	Gly	Arg	Ser	Ser	Val	Ser	Ser
					1380					1385				1390		
35	Ser	Asp	Gly	Ser	Phe	Phe	Thr	Asp	Ala	Asp	Phe	Ala	Gln	Ala	Val	Ala
				1395					1400					1405		
	Ala	Ala	Ala	Glu	Tyr	Ala	Gly	Leu	Lys	Val	Ala	Arg	Arg	Gln	Met	Gln
				1410				1415					1420			
40	Asp	Ala	Ala	Gly	Arg	Arg	His	Phe	His	Ala	Ser	Gln	Cys	Pro	Arg	Pro
	1425					1430						1435				1440
	Thr	Ser	Pro	Val	Ser	Thr	Asp	Ser	Asn	Met	Ser	Ala	Ala	Val	Met	Gln
					1445						1450				1455	
	Lys	Thr	Arg	Pro	Ala	Lys	Lys	Leu	Lys	His	Gln	Pro	Gly	His	Leu	Arg
				1460						1465				1470		
45	Arg	Glu	Thr	Tyr	Thr	Asp	Asp	Leu	Pro	Pro	Pro	Pro	Val	Pro	Pro	Pro
				1475				1480						1485		
	Ala	Ile	Lys	Ser	Pro	Thr	Ala	Gln	Ser	Lys	Thr	Gln	Leu	Glu	Val	Arg
				1490				1495					1500			
50	Pro	Val	Val	Val	Pro	Lys	Leu	Pro	Ser	Met	Asp	Ala	Arg	Thr	Asp	Arg
	1505					1510					1515					1520
	Ser	Ser	Asp	Arg	Lys	Gly	Ser	Ser	Tyr	Lys	Gly	Arg	Glu	Val	Leu	Asp
					1525						1530				1535	
	Gly	Arg	Gln	Val	Val	Asp	Met	Arg	Thr	Asn	Pro	Gly	Asp	Pro	Arg	Glu
				1540						1545				1550		
55	Ala	Gln	Glu	Gln	Gln	Asn	Asp	Gly	Lys	Gly	Arg	Gly	Asn	Lys	Ala	Ala

WO 99/25833

PCT/US98/24327

Lys Arg Asp Leu Pro Pro Ala Lys Thr His Leu Ile Gln Glu Asp Ile  
 1570 1575 1580  
 5 Leu Pro Tyr Cys Arg Pro Thr Phe Pro Thr Ser Asn Asn Pro Arg Asp  
 1585 1590 1595 1600  
 Pro Ser Ser Ser Ser Ser Met Ser Ser Arg Gly Ser Gly Ser Arg Gln  
 1605 1610 1615  
 Arg Glu Gln Ala Asn Val Gly Arg Arg Asn Ile Ala Glu Met Gln Val  
 1620 1625 1630  
 10 Leu Gly Gly Tyr Glu Arg Gly Glu Asp Asn Asn Glu Glu Leu Glu Glu  
 1635 1640 1645  
 Thr Glu Ser  
 1650

15 (2) INFORMATION FOR SEQ ID NO:9:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1300 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (ix) FEATURE:  
 (A) NAME/KEY: misc\_feature  
 (B) LOCATION: 855..1187  
 25 (D) OTHER INFORMATION: /note= "N signifies gap in sequence"  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAGATTGTTG CTCAAGGTCG AACAGTGACA TTTCCCTGTG AACTAAAGG AAACCCACAG 60  
 CCAGCTGTTT TTTGGCAGAA AGAAGGCAGC CAGAACCTAC TTTTCCCAA CCAACCCACAG 120  
 CAGCCCAACA GTAGATGCTC AGTGTCCACCA ACTGGAGACC TCACAATCAC CAACATTCAA 180  
 30 CGTTCCGACG CGGGTTACTA CATCTGCCAG GCTTTAACTG TGGCAGGAAG CATTTTAGCA 240  
 AAAGCTCAAC TGGAGGTTAC TGATGTTTTG ACAGATAGAC CTCCACCTAT AATTCTACAA 300  
 GGCCAGCCA ACCAAACGCT GGCAGTGGAT GGTACAGCGT TACTGAAATG TAAAGCCACT 360  
 GGTGATCCTC TTCCTGTAAT TAGCTGGTTA AAGGAGGGAT TACTTTTCC GGGTAGAGAT 420  
 CCAAGAGCAA CAATTCAAGA GCAAGGCACA CTGCAGATTA AGAATTTACG GATTTCTGAT 480  
 35 ACTGGCACTT ATACTTGTGT GGCTACAAGT TCAAGTGGAG AGGCTTCCTG GAGTGCAGTG 540  
 CTGGATGTGA CAGAGTCTGG AGCAACAATC AGTAAAAACT ATGATTTAAG TGACCTGCCA 600  
 GGGCCACCAT CCAAACCGCA AGTCACTGAT GTTACTAAGA ACAGTGTCCAC CTTGTCCTGG 660  
 CAGCCAGGTA CCCCTGGAAC CCTTCCAGCA AGTGCATATA TCATTGAGGC TTTCCAGCCAA 720  
 TCAGTGAGCA ACAGCTGGCA GACCGTGGCA AACCATGTAA AGACCACCCT CTATACTGTA 780  
 40 AGAGGACTGC GGCCCAATAC AATCTACTTA TTCATGGTCA GAGCGATCAA CCCCAAGGTY 840  
 TCAGTGACCC AAGTNAAC ACAGAAAAAC AATGGATCCA CTTGGGCCAA TGTCCCTCTA 900  
 CCTCCCCCCC CAGTCCAGCC CCTTCTGGC ACGGAGCTGG AACACTATGC AGTGGAAACAA 960  
 CAAGAAAATG GCTATGACAG TGATAGCTGG TGCCCACCAT TGCCAGTACA AACTTACTTA 1020  
 CACCAAGGTC TGGAAGATGA ACTGGAAGAA GATGATGATA GGGTCCCAAC ACCTCCTGTT 1080  
 45 CGAGGCGTGG CTTCTTCTCC TGCTATCTCC TTTGGACAGC AGTCCACTGC AACTCTTACT 1140  
 CCATCCCCAC GGAAGAGAT GCAACCCATG CTGCAGGCTT CACCTNTTTA CCTCCTCTCA 1200  
 AAGACCTCGA CCTACCAGCC CATTTTCTAC TGACAGTAAC ACCAGTGCAG CCCTGAGTCA 1260  
 AAGTCAGAGG CCTCGGCCCA CTAAAAACA CAAGGGAGGG 1300

50 (2) INFORMATION FOR SEQ ID NO:10:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 434 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 55 (D) TOPOLOGY: linear

WO 99/25833

PCT/US98/24327

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 285..396

5 (D) OTHER INFORMATION: /note= "Xaa signifies gap in sequence"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

	Gln	Ile	Val	Ala	Gln	Gly	Arg	Thr	Val	Thr	Phe	Pro	Cys	Glu	Thr	Lys
	1				5					10					15	
10	Gly	Asn	Pro	Gln	Pro	Ala	Val	Phe	Trp	Gln	Lys	Glu	Gly	Ser	Gln	Asn
				20					25					30		
	Leu	Leu	Phe	Pro	Asn	Gln	Pro	Gln	Gln	Pro	Asn	Ser	Arg	Cys	Ser	Val
			35					40					45			
	Ser	Pro	Thr	Gly	Asp	Leu	Thr	Ile	Thr	Asn	Ile	Gln	Arg	Ser	Asp	Ala
		50					55					60				
15	Gly	Tyr	Tyr	Ile	Cys	Gln	Ala	Leu	Thr	Val	Ala	Gly	Ser	Ile	Leu	Ala
	65					70					75					80
	Lys	Ala	Gln	Leu	Glu	Val	Thr	Asp	Val	Leu	Thr	Asp	Arg	Pro	Pro	Pro
				85						90					95	
20	Ile	Ile	Leu	Gln	Gly	Pro	Ala	Asn	Gln	Thr	Leu	Ala	Val	Asp	Gly	Thr
				100						105				110		
	Ala	Leu	Leu	Lys	Cys	Lys	Ala	Thr	Gly	Asp	Pro	Leu	Pro	Val	Ile	Ser
			115					120					125			
	Trp	Leu	Lys	Glu	Gly	Phe	Thr	Phe	Pro	Gly	Arg	Asp	Pro	Arg	Ala	Thr
		130					135					140				
25	Ile	Gln	Glu	Gln	Gly	Thr	Leu	Gln	Ile	Lys	Asn	Leu	Arg	Ile	Ser	Asp
	145					150					155					160
	Thr	Gly	Thr	Tyr	Thr	Cys	Val	Ala	Thr	Ser	Ser	Ser	Gly	Glu	Ala	Ser
				165						170					175	
	Trp	Ser	Ala	Val	Leu	Asp	Val	Thr	Glu	Ser	Gly	Ala	Thr	Ile	Ser	Lys
30				180					185					190		
	Asn	Tyr	Asp	Leu	Ser	Asp	Leu	Pro	Gly	Pro	Pro	Ser	Lys	Pro	Gln	Val
			195					200					205			
	Thr	Asp	Val	Thr	Lys	Asn	Ser	Val	Thr	Leu	Ser	Trp	Gln	Pro	Gly	Thr
		210					215					220				
35	Pro	Gly	Thr	Leu	Pro	Ala	Ser	Ala	Tyr	Ile	Ile	Glu	Ala	Phe	Ser	Gln
	225					230					235					240
	Ser	Val	Ser	Asn	Ser	Trp	Gln	Thr	Val	Ala	Asn	His	Val	Lys	Thr	Thr
				245						250					255	
40	Leu	Tyr	Thr	Val	Arg	Gly	Leu	Arg	Pro	Asn	Thr	Ile	Tyr	Leu	Phe	Met
				260					265					270		
	Val	Arg	Ala	Ile	Asn	Pro	Lys	Val	Ser	Val	Thr	Gln	Xaa	Lys	Pro	Gln
			275					280					285			
	Lys	Asn	Asn	Gly	Ser	Thr	Trp	Ala	Asn	Val	Pro	Leu	Pro	Pro	Pro	Pro
		290					295					300				
45	Val	Gln	Pro	Leu	Pro	Gly	Thr	Glu	Leu	Glu	His	Tyr	Ala	Val	Glu	Gln
	305					310					315					320
	Gln	Glu	Asn	Gly	Tyr	Asp	Ser	Asp	Ser	Trp	Cys	Pro	Pro	Leu	Pro	Val
				325						330					335	
50	Gln	Thr	Tyr	Leu	His	Gln	Gly	Leu	Glu	Asp	Glu	Leu	Glu	Glu	Asp	Asp
				340					345						350	
	Asp	Arg	Val	Pro	Thr	Pro	Pro	Val	Arg	Gly	Val	Ala	Ser	Ser	Pro	Ala
			355					360					365			
	Ile	Ser	Phe	Gly	Gln	Gln	Ser	Thr	Ala	Thr	Leu	Thr	Pro	Ser	Pro	Arg
		370					375					380				
55	Glu	Glu	Met	Gln	Pro	Met	Leu	Gln	Ala	Ser	Pro	Xaa	Phe	Thr	Ser	Ser

WO 99/25833

PCT/US98/24327

385 390 395 400  
 Gln Arg Pro Arg Pro Thr Ser Pro Phe Ser Thr Asp Ser Asn Thr Ser  
 405 410 415  
 Ala Ala Leu Ser Gln Ser Gln Arg Pro Arg Pro Thr Lys Lys His Lys  
 5 420 425 430  
 Gly Gly

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 444 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCCCAGGCAG TTGCTGCAGC TCGGAGTAT GCGGGCCTGA AAGTGGCTCG CCGCCAAATG 60  
 CAAGATGCTG CTGGCCGCCG CCACTTCCAT GCCTCTCAGT GCCCAAGGCC CACGAGTCCT 120  
 GTGTCCACAG ACAGCAACAT GAGTGCTGTT GTGATCCAGA AAGCCAGACC CGCCAAGAAG 180  
 CAGAAACACC AGCCAGGACA TCTGCGCAGG GAAGCCTACG CAGATGATCT TCCACCCCCT 240  
 20 CCAGTGCCAC CACCTGCTAT AAAATCGCCC ACTGTCCAGT CCAAGGCACA GCTGGAGGTA 300  
 CGGCCTGTCA TGGTGCCAAA ACTCGCGTCT ATAGAAGCAA GGACAGATAG ATCGTCAGAC 360  
 AGAAAAGGAG GCAGTTACAA GGGGAGAGAA GCTCTGGATG GAAGACAAGT CACTGACCTG 420  
 CGAACAAATC CAAGTGACCC CAGA 444

25 (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 148 amino acids  
 (B) TYPE: amino acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

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

55 (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:

WO 99/25833

PCT/US98/24327

- (A) LENGTH: 1781 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

	TGAGTGCCTC	GCCACGACT	CACTCAAGAG	CGAGTGAATC	GAGAAAATGA	GTGAGTGGAA	60
	TGGCAAATTT	CCCGCATGGC	ATGGCATTCT	CAGCGACCCG	ATGTAGGTAT	TATGCTCTGC	120
	CAGCTAGCAG	TCAGTCAACT	TATAGCCGCT	CCAGCGAACG	GACGCACCGA	AAATCGAGCT	180
10	ATAGAAACTC	TACGGCAAAT	AAAAAAACCC	GAAATCGAAC	CCATGACTTA	TATATAGAGC	240
	CAAACGATTT	GTTTACGTGT	CGTGACGCGA	GTGTGCCAGC	AGTTGTGTGC	GTGTGCCAGA	300
	CGCAAGCAGA	AAATAAAGTT	TACAAATTAA	AATAAATCAA	AGGCAATCAC	CATGATTAGC	360
	ACCACGGATT	ATCCAACGGT	GGAGACCACC	ACAAC TGCCG	AAGAGCTCTA	TGCGGAGTAC	420
	ATATCCGCGC	CGGCCAGCAG	CATGAGTCCC	GCTGCAATTG	CCGAGCACCT	GCAGCAGAAT	480
15	CAGATCACCT	TCGAGATACC	CAGTGCCAC	GATCTGCGAC	ACATCGACGC	CCTCAACTCC	540
	TTCAACGCCC	TGCTCCAGAG	GATTGGCAAC	GCAGCGGTGT	CCTACGATCC	CGCTCCACCC	600
	AGCGGCTGGT	CTCCGGATGG	CAGCATCAGC	ACCGAGCAGC	TCTCCAAGTC	GGTGGTTCTG	660
	GATCTGGCCG	ACTTGCGGGA	TAGGTCCGAG	GAATCGGGCG	AATCCTCGTG	GTGGAGCCAG	720
	ATCTTCGGGG	ATGCCGACAT	GCATGTGATC	ATCAATTATC	TGTGGATCGG	AGTGGTCAGC	780
20	TCGCTGGTCA	TTCTGTGCGT	GGTCTTCATC	CTCTTCAGCT	GCTACTTCTA	CAGGAAGTTC	840
	CGCACTTGGA	AAAAATGCAA	TAAGGACATA	CGTGCCCGA	TCCATGCGGC	CAGTGAATCG	900
	TACTCCTCGC	ACCTGGTTGG	TTGCGATGCC	AGCAGACTGC	TGTTGCACCA	GCAGATGCAG	960
	CATCCACATC	ATCGGAGCAG	CGAGGCTGGG	TTCTACCAA	TCGAATCGCC	GCCCTGCTAC	1020
	ACAATCGCCA	CCGGATTGCC	CAGCTACGAT	GAGGCACTGC	ATCATCAGCC	TCGGCACTTT	1080
25	GCCTACGGCA	TGAAGTTCGT	TTATCCCTCG	TTGGCGGCCG	TGCATCATCA	TCACCATTGC	1140
	ATTTCCAATT	GGGAGAAGCA	GGAGCCGCTG	AATAAGCTGC	AAAAGTGCAA	GCTGTCAGCG	1200
	GCAGCGGCAG	TGGAGGAGGA	TAAAGCCGAC	TCCTCATCGT	CCACGTCTGC	ATCCGCGTCG	1260
	CCATCCTCTT	CGGAATCCAG	CAACTTGGCC	ACAGCAACGC	CTGCCATTTG	CATTAACATG	1320
	CCAAGTGGGC	GGCAGGATGA	GGAGGTGGAT	AATTCAGATT	CCGATTCCGC	AATTGCAGTT	1380
30	GCAGTGGCAG	TAGCACAAAG	TTTACAGCCG	GCGGCGCCTG	CCGACGATGA	TTGCGCCTCA	1440
	TTGGTTCGTT	TTGTTGCCGC	GTGATTGCC	ACCCAGGGGT	CAGAGGTCAG	GGAATAAGGG	1500
	GTCAGTCGAG	GGTTTAGGGC	CGGGAGTAGC	AGCCCAATTG	CTTAGTCACT	TGTAAATAGT	1560
	TGCATCGTCT	GCTAATGTTG	TTTGCCCCAT	GTAATAACAT	ATATCTCCGA	GAATGTACCC	1620
	TCTCCCTCCC	TCTAAGAAA	TTTAGTTTGT	GTTTTTAGTT	ATCTGTTATA	TATTATTGTA	1680
35	TTACGCAAGA	CTGAGTATTA	TTTATTGTAC	TATGTATAAA	ACCGAATTGT	CTATCGATTA	1740
	AGCCCTAATA	TCTAAGCGAG	AGTTGAAGAA	AAAATATAAA	A		1781

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 370 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

	Met	Ile	Ser	Thr	Thr	Asp	Tyr	Pro	Thr	Val	Glu	Thr	Thr	Thr	Thr	Ala
	1				5					10					15	
	Glu	Glu	Leu	Tyr	Ala	Glu	Tyr	Ile	Ser	Ala	Pro	Ala	Ser	Ser	Met	Ser
				20					25					30		
50	Pro	Ala	Ala	Ile	Ala	Glu	His	Leu	Gln	Gln	Asn	Gln	Ile	Thr	Phe	Glu
				35				40					45			
	Ile	Pro	Ser	Ala	His	Asp	Leu	Arg	His	Ile	Asp	Ala	Leu	Asn	Ser	Phe
		50					55					60				
	Asn	Ala	Leu	Leu	Gln	Arg	Ile	Gly	Asn	Ala	Ala	Val	Ser	Tyr	Asp	Pro
55	65					70					75					80

WO 99/25833

PCT/US98/24327

	Ala	Pro	Pro	Ser	Gly	Trp	Ser	Pro	Asp	Gly	Ser	Ile	Ser	Thr	Glu	Gln
					85					90					95	
	Leu	Ser	Lys	Ser	Val	Val	Leu	Asp	Leu	Ala	Asp	Leu	Arg	Asp	Arg	Ser
				100				105						110		
5	Glu	Glu	Ser	Gly	Glu	Ser	Ser	Trp	Trp	Ser	Gln	Ile	Phe	Gly	Asp	Ala
			115					120					125			
	Asp	Met	His	Val	Ile	Ile	Asn	Tyr	Leu	Trp	Ile	Gly	Val	Val	Ser	Ser
	130						135					140				
10	Leu	Val	Ile	Leu	Ser	Leu	Val	Phe	Ile	Leu	Phe	Ser	Cys	Tyr	Phe	Tyr
	145					150					155					160
	Arg	Lys	Phe	Arg	Thr	Trp	Lys	Lys	Cys	Asn	Lys	Asp	Ile	Arg	Ala	Gln
				165						170					175	
	Ile	His	Ala	Ala	Ser	Asp	Ser	Tyr	Ser	Ser	His	Leu	Val	Gly	Cys	Asp
				180					185					190		
15	Ala	Ser	Arg	Leu	Leu	Leu	His	Gln	Gln	Met	Gln	His	Pro	His	His	Arg
			195					200					205			
	Ser	Ser	Glu	Ala	Gly	Phe	Tyr	Gln	Ile	Glu	Ser	Pro	Pro	Cys	Tyr	Thr
			210				215					220				
20	Ile	Ala	Thr	Gly	Leu	Pro	Ser	Tyr	Asp	Glu	Ala	Leu	His	His	Gln	Pro
						230					235					240
	Arg	His	Phe	Ala	Tyr	Gly	Met	Lys	Phe	Val	Tyr	Pro	Ser	Leu	Ala	Ala
				245						250					255	
	Val	His	His	His	His	His	Cys	Ile	Ser	Asn	Trp	Glu	Lys	Gln	Glu	Pro
				260					265					270		
25	Leu	Asn	Lys	Leu	Gln	Lys	Cys	Lys	Leu	Ser	Ala	Ala	Ala	Ala	Val	Glu
			275					280					285			
	Glu	Asp	Lys	Ala	Asp	Ser	Ser	Ser	Ser	Thr	Ser	Ala	Ser	Ala	Ser	Pro
						295						300				
30	Ser	Ser	Ser	Glu	Ser	Ser	Asn	Leu	Ala	Thr	Ala	Thr	Pro	Ala	Ile	Cys
						310					315					320
	Ile	Asn	Met	Pro	Ser	Gly	Arg	Gln	Asp	Glu	Glu	Val	Asp	Asn	Ser	Asp
					325					330					335	
	Ser	Asp	Ser	Ala	Ile	Ala	Val	Ala	Val	Ala	Val	Ala	Gln	Ser	Leu	Gln
				340					345					350		
35	Pro	Ala	Ala	Pro	Ala	Asp	Asp	Asp	Cys	Ala	Ser	Leu	Val	Val	Val	Val
			355					360					365			
	Ala	Ala														
40																

WO 99/25833

PCT/US98/24327

## WHAT IS CLAIMED IS:

1. A method of modulating an amount of expressed active Robo on a cell, said method comprising the step of modulating the effective amount of a Comm polypeptide in contact with the cell, whereby the amount of expressed active Robo is specifically modulated  
5 inversely with the modulation of the effective amount of the Comm polypeptide in contact with the cell.
2. A method according to claim 1, wherein the effective amount of the Comm polypeptide is increased and the amount of expressed Robo is decreased.  
10
3. The method of claim 1, in which the Comm polypeptide is provided to the cell exogenously in a pharmaceutically acceptable composition.
4. The method of claim 1, further comprising the steps of forming a mixture of the cell,  
15 the Comm polypeptide and a candidate agent, and determining the effect of the agent on the amount of expressed Robo.