PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/00 | A2

(11) International Publication Number: **WO 00/55193**

(43) International Publication Date: 21 September 2000 (21.09.00)

(21) International Application Number: PCT/US00/05537

(22) International Filing Date: 2 March 2000 (02.03.00)

(30) Priority Data:
60/124,118 12 March 1999 (12.03.99) US

(71) Applicant *(for all designated States except US)*: REGENERON PHARMACEUTICALS, INC. [US/US]; 777 Old Saw Mill River Road, Tarrytown, NY 10591-6707 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: ECONOMIDES, Aris, N. [GR/US]; 12 Mt. Morris Park West, New York, NY 10027 (US).

(74) Agents: PALLADINO, Linda, O.; Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY 10591 (US) et al.

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract

DAN/Cerberus Related protein 6 (DCR6) polypeptides and related nucleic acids are provided. Included are natural (DCR6) homologs from several species and polypeptides comprising a (DCR6) domain having specific activity. The polypeptides may be produced recombinantly from transformed host cells with the subject nucleic acids. Also provided are isolated hybridization probes and oligonucleotide primers capable of specifically hybridizing with the disclosed genes, specific binding agents and methods of making and using the subject compositions.

FOR THE PURPOSES OF INFORMATION ONLY

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

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

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

This International Application claims priority of U.S. Provisional Application No. 60/124,118, filed March 12, 1999. All publications and patent applications cited in this specification are hereby incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

INTRODUCTION

Field of the Invention

The field of this invention is polypeptides which regulate cell function and, in particular, antagonize bone morphogenic proteins and which are involved in the development and maintenance of the vascular system.

Background

Natural regulators of cellular growth, differentiation and function have provided important pharmaceuticals, clinical and laboratory tools, and targets for therapeutic intervention. A variety of such regulators have been shown to have profound effects on basic cellular differentiation and developmental pathways. For example, the recently cloned cerberus protein induces the formation of head structures in anterior endoderm of vertebrate embryos. Similarly, the noggin protein induces head structures in vertebrate embryos, and can redirect mesodermal fates from ventral fates, such as blood and mesenchyme, to dorsal fates such as muscle and notochord and can redirect epidermal fates to anterior neural fates. The activities of chordin are similar to those of noggin, reflecting a common mechanism of action - namely antagonizing bone morphogenic proteins (BMPs) and thereby preventing their function. BMPs have diverse biological activities in different biological contexts, including the induction of

cartilage, bone and connective tissue, and roles in kidney, tooth, gut, skin and hair development.

Different members of the TGFβ superfamily can instruct cells to follow different fates, for example TGFβ induces neural crest to form smooth muscle, while BMP2 induces the same cells to become neurons. In Xenopus experiments, dissociated animal cap cells (prospective ectoderm) become epidermis in response to BMP4 but become mesoderm in response to activin.

Since the sequence identity between activin and BMP4 is low, it is not surprising that they induce different fates. It is more surprising that members of the BMP subfamily, which are quite closely related in sequence, can induce distinct fates. A striking example results from implantation of a matrix impregnated with a BMP into muscle; when the effects are monitored histologically, BMP2, 4 and 7 induce endochondral bone formation, whereas a related molecule BMP12/GDF7 induces connective tissue similar to tendon. Similarly, BMP4 can induce cell death in the hindbrain neural crest, while the related protein dorsalin does not.

Since different BMP family members can induce different fates, then BMP antagonists that have specificity in blocking subsets of BMPs could change the balance of BMPs that are presented to a cell, thus altering cell fate. In view of the importance of relative BMP expression in human health and disease, regulators of cellular function and BMP function in particular, such as noggin and cerberus, provide valuable reagents with a host of clinical and biotechnological applications.

The ability of ligands to bind cells and thereby elicit a phenotypic response such as development, differentiation, growth, proliferation, survival and regeneration in such cells is often mediated through transmembrane receptors. The extracellular portion of each receptor is generally the most distinctive portion of the molecule, as it provides the protein with its ligand-recognizing characteristic.

In the case of receptor tyrosine kinases (RTKs), binding of a ligand to the extracellular domain results in signal transduction via an intracellular tyrosine kinase catalytic domain which transmits a biological signal to intracellular target proteins. For example, a gene encoding an endothelial cell transmembrane tyrosine kinase, originally identified by RT-PCR as an unknown tyrosine kinase-homologous cDNA fragment from human leukemia cells, was described by Partanen, et al., Proc. Natl. Acad. Sci. USA, 87: 8913-8917 (1990). This gene and its encoded protein are called "tie" which is an abbreviation for "tyrosine kinase with Ig and EGF homology domains." Partanen, et al. Mol. Cell. Biol. 12: 1698-1707 (1992).

It has been reported that tie mRNA is present in all human fetal and mouse embryonic tissues. Upon inspection, tie message has been localized to the cardiac and vascular endothelial cells. tie mRNA has been localized to the endothelia of blood vessels and endocardium of 9.5 to 18.5 day old mouse embryos. Enhanced tie expression was shown during neovascularization associated with developing ovarian follicles and granulation tissue in skin wounds. Korhonen, et al. Blood 80: 2548-2555 (1992). Thus tie has been suggested to play a role in angiogenesis, which is important for developing treatments for solid tumors and several other angiogenesis-dependent diseases such as diabetic retinopathy, psoriasis, atherosclerosis and arthritis.

Two structurally related rat TIE receptor proteins have been reported to be encoded by distinct genes with related profiles of expression. One gene, termed tie-1, is the rat homolog of human tie. Maisonpierre, et al., Oncogene 8: 1631-1637 (1993). The other gene, tie-2, may be the rat homolog of the murine tek gene, which, like tie, has been reported to be expressed in the mouse exclusively in endothelial cells and their presumptive progenitors. Dumont, et al. Oncogene 8: 1293-1301 (1993). Both genes were found to be widely expressed in endothelial cells of embryonic and postnatal tissues. Significant levels of tie-2 transcripts were also present in other embryonic cell populations, including lens epithelium, heart epicardium and regions of mesenchyme. Maisonpierre, et al.,

Oncogene 8: 1631-1637 (1993). The predominant expression of the TIE receptor in vascular endothelia suggests that TIE plays a role in the development and maintenance of the vascular system. This could include roles in endothelial cell determination, proliferation, differentiation and cell migration and patterning into vascular elements. In the mature vascular system, TIE could function in endothelial cell survival, maintenance and response to pathogenic influences.

An angiogenic factor, which was originally called TIE-2 ligand-1 (TL1) but is also referred to as angiopoietin-1 (Ang1), has been identified that signals through the TIE-2 receptor and is essential for normal vascular development in the mouse. By homology screening, an Ang1 relative has been identified and called TIE-2 ligand-2 (TL2) or angiopoietin-2 (Ang2). Ang2 is a naturally occurring antagonist for Ang1 and the TIE2 receptor. For a description of the cloning and sequencing of TL1 (Ang1) and TL2 (Ang2) as well as for methods of making and uses thereof, reference is hereby made to PCT International Publication No. WO 96/11269 published 18 April 1996 and PCT International Publication No. WO 96/31598 published 10 October 1996 both in the name of Regeneron Pharmaceuticals, Inc.; and S. Davis, et al., Cell 87: 1161-1169 (1996) each of which is hereby incorporated by reference.

The present invention relates to a novel regulator of cellular functions such as antagonizing bone morphogenic proteins and playing a role in the development and maintenance of the vascular system. This novel regulator shares homology with the DAN/cerberus family and is expressed in vascular tissues.

Relevant Literature

Bouwmeester, et al. (1996) Nature 382: 595-601 describe the cloning of Xenopus cerberus gene; Lamb, T. M., et al. (1993) Science 262: 713-718; Smith, W. C., et al. (1992) Cell 70: 829-840; Smith, W. C., et al. (1993) Nature 361: 547-549; and Zimmerman, L. B., et al. (1996) Cell 86: 599-606 describe the isolation and function of the noggin protein. Piccolo, S., et al. (1996) Cell 86: 589-598; Sasai, Y., et al.

(1995) Nature 376: 333-336; and Sasai, Y., et al. (1994) Cell 79: 779-790 relate to the chordin protein. Enomoto et al. (1994) Oncogene 9: 2785-2791 and Ozaki, et al. (1996) Jpn. J. Cancer Res. 87: 58-61 describe human and murine homologs of the DAN gene.

SUMMARY OF THE INVENTION

The invention provides methods and compositions relating to DAN/Cerberus - Related protein 6 (DCR6) polypeptides and related nucleic acids. Included are natural DCR6 homologs from different species, as well as polypeptides comprising a DCR6 domain and having DCR6-specific activity. The polypeptides may be produced recombinantly from transformed host cells with the subject nucleic acids. The invention provides isolated hybridization probes and primers capable of specifically hybridizing with the disclosed genes, specific binding agents such as specific antibodies, and methods of making and using the subject compositions in diagnosis (e.g., genetic hybridization screens for DCR6 transcripts), therapy (e.g., gene therapy to modulate DCR6 gene expression) and in the biopharmaceutical industry (e.g., reagents for screening chemical libraries for lead pharmacological agents).

Preferred applications of the subject DCR6 polypeptides include modifying the physiology of a cell comprising an extracellular surface by contacting the cell or medium surrounding the cell with an exogenous DCR6 polypeptide under conditions whereby the added polypeptide specifically interacts with a component of the medium and/or the extracellular surface to effect a change in the physiology of the cell. Also preferred are methods for screening for biologically active agents, which methods involve incubating a DCR6 polypeptide in the presence of an extracellular DCR6 polypeptide-specific binding target and a candidate agent, under conditions whereby, but for the presence of the agent, the polypeptide specifically binds the binding target at a reference affinity; detecting the binding affinity of the polypeptide to the binding target to determine an agent-biased affinity, wherein a difference between the agent-biased

affinity and the reference affinity indicates that the agent modulates the binding of the polypeptide to the binding target.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-1F. The genomic DNA sequence of vts_hDCR6. The predicted boundaries of exons 1, 2, 3, and 4 are indicated underneath the sequence.

Figure 2A-2B. The nucleic acid and deduced amino acid sequence of vts_hDCR6 that was created by PCR-amplifying the individual exons from human genomic DNA and splicing them together. Silent mutations introduced to facilitate cloning and polypeptide expression are indicated in bold above the nucleic acid sequence and splice-junction sites between adjacent exons are underlined.

Figure 3A-3B. The nucleic acid and deduced amino acid sequence of hDCR6 that was cloned from a human kidney cDNA library having exons 1 and 4.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

An "oligonucleotide" or "oligonucleotide primer" or "primer" is a stretch of nucleotide residues which has a sufficient number of bases to be used in, for example, a polymerase chain reaction (PCR) or in DNA sequencing methodologies. These short sequences are based on (or designed from) genomic or cDNA sequences or back translated from protein sequences and are used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue or to initiate sequencing reactions. Oligonucleotides or oligomers comprise portions of a DNA sequence having at least about 10 nucleotides and as many as about 50 nucleotides, preferably about 15 to 30 nucleotides. They are chemically synthesized and may be used as probes.

"Probes" are nucleic acid sequences of variable length, preferably between at least about 10 and as many as about 6,000 nucleotides, depending on use. They are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligonucleotides. They may be single- or double-stranded and carefully designed to have specificity in PCR, hybridization membrane-based, or ELISA-like technologies.

A "portion" or "fragment" of a polynucleotide or nucleic acid or polypeptide comprises all or any part of the polynucleotide or a polypeptide sequence having fewer nucleotides or amino acids than the complete polynucleotide or nucleic acid or polypeptide.

A "signal sequence" is a short amino acid sequence which can be used, when desired, to direct the polypeptide through a membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous sources by recombinant DNA techniques.

"Animal" as used herein may be defined to include human, domestic (i.e., cats, dogs), agricultural (i.e., cows, horses, sheep, goats, chicken, fish) or test species (i.e., frogs, mice, rats, rabbits, simians).

Since the list of technical and scientific terms cannot be all encompassing, any undefined terms shall be construed to have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. Furthermore, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, reference to a "restriction enzyme" or a "high fidelity enzyme" may include mixtures of such enzymes and any other enzymes fitting the stated criteria, or reference to the method includes reference to one or more methods for obtaining cDNA

sequences which will be known to those skilled in the art or will become known to them upon reading this specification.

Before the present sequences, variants, formulations and methods for making and using the invention are described, it is to be understood that the invention is not to be limited only to the particular sequences, variants, formulations or methods described. The sequences, variants, formulations and methodologies may vary, and the terminology used herein is for the purpose of describing particular embodiments. The terminology and definitions are not intended to be limiting since the scope of protection will ultimately depend upon the claims.

The invention provides DCR6 polypeptides which include natural DCR6 polypeptides and recombinant polypeptides comprising a DCR6 amino acid sequence, or a functional DCR6 polypeptide domain thereof having an assay-discernable DCR6-specific activity. Accordingly, the polypeptides may be deletion mutants of the disclosed natural DCR6 polypeptides and may be provided as fusion products, e.g., with non-DCR6 polypeptides. The subject DCR6 polypeptide domains have DCR6-specific activity or function and are functionally distinct from each other and from DAN/Cerberus family and noggin homologs. Such domains include at least 6 and preferably at least 8 consecutive amino acid residues of a natural DCR6 polypeptide (see human DCR6 sequence disclosed herein). Preferred DCR6 polypeptides comprise a DCR6 sequence conserved across species.

Note that contrary to prior art teachings which state that DAN is an intracellular zinc finger protein, applicants disclose that the natural DAN protein is extracellularly active as an antagonist of certain morphogenic proteins such as BMPs. In addition, the DCR5 sequence, set forth in co-pending US Provisional Application No. 60/097,296, filed August 20, 1998, is also extracellularly active as an antagonist of certain morphogenic proteins such as BMPs. Because DCR-6 is structurally similar to DAN and DCR5, applicants predict that DCR6 will exhibit biological activities similar to these two related proteins. DCR6-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., immune response, gene therapy, transgenics). Binding assays encompass any assay where the specific molecular interaction of a DCR6 polypeptide with a binding target is evaluated. The binding target may be a natural binding target, a chaperon, or other regulator that directly modulates DCR6 activity or its localization; or non-natural binding target such as a specific immune protein such as an antibody, or a DCR6-specific agent such as those identified in assays described below. Generally, binding specificity is assayed by bioassay (e.g., the ability to induce neuronal tissue from injected embryonic ectoderm), target protein 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 DCR6-expressing cells, by the ability to elicit DCR6-specific antibody production in a heterologous host (e.g., a rodent or rabbit).

The claimed polypeptides may be isolated or pure - an "isolated" polypeptide is one that is no longer accompanied by some of the material with which it is associated in its natural state, and that preferably constitutes at least about 0.5%, and more preferably at least about 5% by weight of the total protein in a given sample; a "pure" polypeptide constitutes at least about 90%, and preferably at least about 99% by weight of the total protein in a given sample. The subject polypeptides and polypeptide domains may be synthesized, produced by recombinant technology, or purified from cells. A wide variety of molecular and biochemical methods are available for biochemical synthesis, molecular expression and purification of the subject compositions (see e.g., Molecular Cloning, A Laboratory Manual (Sambrook, et al., Cold Spring Harbor Laboratory), Current Protocols in Molecular Biology (Eds. Ausubel, et al., Greene Publ. Assoc., Wiley-Interscience, NY).

The subject polypeptides find a wide variety of uses including use as immunogens, targets in screening assays, bioactive reagents for modulating cell

growth, differentiation and/or function. For example, the invention provides methods for modifying the physiology of a cell comprising an extracellular surface by contacting the cell or medium surrounding the cell with an exogenous DCR6 polypeptide under conditions whereby the added polypeptide specifically interacts with a component of the medium and/or the extracellular surface to effect a change in the physiology of the cell. According to these methods, the extracellular surface includes plasma membrane-associated receptors; the exogenous DCR6 refers to a polypeptide not made by the cell or, if so, expressed at non-natural levels, times or physiologic locales; and suitable media include in vitro culture media and physiological fluids such as blood, synovial fluid or lymph. Effective administrations of subject polypeptides may be useful in reducing undesirable (e.g., ectopic) bone formation, inhibit the growth of cells that require a morphogenic protein (e.g., BMP-dependent neuroblastomas and gliomas), alter morphogen-dependent cell fate/differentiation in culture, such as with cells for transplantation or infusion. The polypeptides may be may be introduced, expressed, or repressed in specific populations of cells by any convenient way such as microinjection, promoter-specific expression of recombinant enzyme, or targeted delivery of lipid vesicles.

The invention provides natural and non-natural DCR6-specific binding agents, methods of identifying and making such agents, and their use in diagnosis, therapy and pharmaceutical development. DCR6-specific binding agents may include ligands such as BMPs, and receptors, such as somatically recombined protein receptors like specific antibodies or T-cell antigen receptors (See, e.g., Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory) and may also include other natural binding agents identified with assays such as one-, two- and three-hybrid screens, and non-natural binding agents identified in screens of chemical libraries such as described below. Agents of particular interest modulate DCR6 function.

The invention provides DCR6 nucleic acids, which find a wide variety of applications including use as translatable transcripts, hybridization probes, PCR

primers, diagnostic nucleic acids, as well as use in detecting the presence of DCR6 genes and gene transcripts and in detecting or amplifying nucleic acids encoding additional DCR6 homologs and structural analogs.

The subject nucleic acids are of synthetic/non-natural sequences and/or are isolated, i.e., no longer accompanied by some of the material with which it is associated in its natural state, preferably constituting at least about 0.5%, more preferably at least about 5% by weight of total nucleic acid present in a given fraction, and usually recombinant, meaning they comprise a non-natural sequence or a natural sequence joined to nucleotide(s) other than that which it is joined to on a natural chromosome. Nucleic acids comprising the nucleotide sequence of Figure 2A-2B or Figure 3A-3B or fragments thereof, contain such sequences or fragments at a terminus, immediately flanked by a sequence other than that to which it is joined on a natural chromosome, or flanked by a native flanking region fewer than 10 kb, preferably fewer than 2 kb, which is immediately flanked by a sequence other than that to which it is joined on a natural chromosome. While the nucleic acids are usually RNA or DNA, it is often advantageous to use nucleic acids comprising other bases or nucleotide analogs to provide modified stability.

DCR6-encoding nucleic acids may be part of expression vectors and may be incorporated into recombinant host cells, e.g., for expression and screening, for transgenic animals, for functional studies such as the efficacy of candidate drugs for diseases associated with DCR6-mediated signal transduction. Expression systems are selected and/or tailored to effect DCR6 polypeptide structural and functional variants through alternative post-translational processing.

The invention also provides for nucleic acid hybridization probes and replication/amplification primers having a DCR6 cDNA specific sequence and sufficient to effect specific hybridization with the sequences set forth in Figures 1A-1F, 2A-2B, or 3A-3B. 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 $NaPO_4$, 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; 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 0.2x SSPE buffer at 42°C. DCR6 cDNA homologs can also be distinguished from other cDNA-encoding polypeptides using alignment algorithms, such as BLASTX (Altschul, et al. (1990) Basic Local Alignment Search Tool, J. Mol. Biol. 215: 403-410).

DCR6 hybridization probes find use in identifying wild-type and mutant alleles in clinical and laboratory samples. Mutant alleles are used to generate allele-specific oligonucleotide (ASO) probes for high-throughput clinical diagnoses. DCR6 nucleic acids are also used to modulate cellular expression or intracellular concentration or availability of active DCR6. DCR6 inhibitory nucleic acids are typically antisense - single stranded sequences comprising complements of the disclosed natural DCR6 coding sequences. Antisense modulation of the expression of a given DCR6 polypeptide may employ antisense nucleic acids operably linked to gene regulatory sequences. Cells are transfected with a vector comprising a DCR6 sequence with a promoter sequence oriented such that transcription of the gene yields an antisense transcript capable of binding to endogenous DCR6-encoding mRNA. Transcription of the antisense nucleic acid may be constitutive or inducible and the vector may provide for stable extrachromosomal maintenance or integration. Alternatively, single-stranded antisense nucleic acids that bind to genomic DNA or mRNA encoding a given DCR6 polypeptide may be administered to the target cell, in or temporarily isolated from a host, at a concentration that results in a substantial reduction in expression of the targeted protein. An enhancement in DCR6 expression is effected by introducing into the targeted cell type DCR6 nucleic acids which increase the functional expression of the corresponding gene products. Such nucleic acids may be DCR6 expression vectors, vectors which upregulate the functional expression of an endogenous allele, or replacement vectors for

targeted correction of mutant alleles. Techniques for introducing the nucleic acids into viable cells are known in the art and include retroviral-based transfection or viral coat protein-liposome mediated transfection.

The invention provides efficient methods of identifying agents, compounds or lead compounds for agents active at the level of DCR6 modulatable cellular function. Generally, these screening methods involve assaying for compounds which modulate DCR6 interaction with a natural DCR6 binding target. A wide variety of assays for binding agents are provided including protein-protein binding assays, immunoassays or cell based assays. Preferred methods are amenable to automated, cost-effective high throughput screening of chemical libraries for lead compounds.

In vitro binding assays employ a mixture of components including a DCR6 polypeptide, which may be part of a fusion product with another peptide or polypeptide, e.g., a tag for detection or anchoring. The assay mixtures comprise a natural DCR6 binding target. While native binding targets may be used, it is frequently preferred to use portions thereof as long as the portion provides binding affinity and avidity to the subject DCR6 that is conveniently measurable in the assay. The assay mixture also comprises a candidate pharmacological agent. Candidate agents encompass numerous chemical classes, though typically they are organic compounds, preferably small organic compounds, and are obtained from a wide variety of sources including libraries of synthetic or natural compounds. A variety of other reagents such as salts, buffers, neutral proteins, e.g., albumin, detergents, protease inhibitors, nuclease inhibitors or antimicrobial agents, may also be included. The mixture components can be added in any order that provides for the requisite binding and incubations may be performed at any temperature which facilitates optimal binding. The mixture is incubated under conditions whereby, but for the presence of the candidate pharmacological agent, the DCR6 specifically binds the cellular binding target, portion or analog with a reference binding affinity. Incubation periods are chosen for optimal binding but are also minimized to facilitate rapid, high throughput screening.

After incubation, the agent-biased binding between the DCR6 and one or more binding targets is detected by any convenient way. For cell-free binding type assays, a separation step is often used to separate bound from unbound components. Separation may be effected by, for example, precipitation or immobilization, followed by washing by, e.g., membrane filtration or gel chromatography. For cell-free binding assays, one of the components usually comprises or is coupled to a label. The label may provide for direct detection such as radioactivity, luminescence, optical or electron density, or indirect detection such as an epitope tag or an enzyme. A variety of methods may be used to detect the label depending on the nature of the label and other assay components, e.g., through optical or electron density, radiative emissions, nonradiative energy transfers, or indirectly detected with antibody conjugates. A difference in the binding affinity of the DCR6 polypeptide to the target in the absence of the agent as compared with the binding affinity in the presence of the agent indicates that the agent modulates the binding of the DCR6 polypeptide to the corresponding binding target. A difference, as used herein, is statistically significant and preferably represents at least a 50%, more preferably at least a 90% difference.

The invention provides for a method for modifying the physiology of a cell comprising an extracellular surface in contact with a medium, said method comprising the step of contacting said medium with an exogenous DCR6 polypeptide under conditions whereby said polypeptide specifically interacts with at least one of a component of said medium and said extracellular surface to effect a change in the physiology of said cell.

The invention further provides for a method for screening for biologically active agents, said method comprising the steps of a) incubating a DCR6 polypeptide in the presence of an extracellular DCR6 polypeptide specific binding target and a candidate agent, under conditions whereby, but for the presence of said agent, said polypeptide specifically binds said binding target at a reference affinity; b)

detecting the binding affinity of said polypeptide to said binding target to determine an agent-biased affinity, wherein a difference between the agent-biased affinity and the reference affinity indicates that said agent modulates the binding of said polypeptide to said binding target.

The invention provides for an isolated nucleic acid molecule encoding human DCR6.

The invention further provides for an isolated nucleic acid molecule having a sequence selected from the group consisting of (a) the nucleotide sequence comprising the coding region of human DCR6 as set forth in Figure 2A-2B; (b) the nucleotide sequence comprising the coding region of human DCR6 as set forth in Figure 3A-3B; (c) a nucleotide sequence that hybridizes under stringent conditions to the nucleotide sequence of (a) or (b) and which encodes a molecule having the biological activity of human DCR6; or (d) a nucleotide sequence which, but for the degeneracy of the genetic code would hybridize to a nucleotide sequence of (a), (b), or (c) and which encodes a molecule having the biological activity of the human DCR6.

The invention provides for a vector or plasmid wherein the DCR6 nucleic acid molecule is operatively linked to an expression control sequence capable of directing its expression in a host cell.

The invention further provides for isolated human DCR6 polypeptide comprising the amino acid sequence as set forth in Figure 2A-2B or Figure 3A-3B, or a fragment thereof having DCR6-specific activity.

The invention provides for a host-vector system for the production of human DCR6 wherein the host cell is a bacterial, yeast, insect or mammalian cell.

The invention provides for a method of producing human DCR6 which comprises growing cells of a host-vector system under conditions permitting

production of the human DCR6, and recovering the human DCR6 so produced.

The invention also provides for an antibody which specifically binds the human DCR6 polypeptide. The antibody may be a polyclonal antibody or a monoclonal antibody.

The invention provides for a pharmaceutical composition comprising human DCR6 polypeptide and an acceptable carrier as well as a pharmaceutical composition comprising an antibody an acceptable carrier.

The invention further provides for human DCR6 polypeptide, an antibody, or a composition for use in a method of treatment of the human or animal body, or in a method of diagnosis.

The invention provides for a ligandbody which comprises human DCR6 fused to an immunoglobulin constant region and a ligandbody wherein the immunoglobulin constant region is the Fc portion of human IgG1.

The invention provides for a ligandbody for use in a method of treatment of the human or animal body, or in a method of diagnosis.

Another embodiment of the invention is a recombinant nucleic acid encoding DCR6 polypeptide comprising the amino acid sequence as set forth in Figure 2A-2B or Figure 3A-3B or a fragment thereof having DCR6- specific activity.

Still another embodiment is an isolated nucleic acid comprising a nucleotide sequence as set forth in Figure 2A-2B or Figure 3A-3B or a fragment thereof having at least 18 consecutive bases of the sequences set forth in Figure 2A-2B or Figure 3A-3B and sufficient to specifically hybridize with a nucleic acid having the sequences as set forth in Figure 2A-2B or Figure 3A-3B in the presence of natural DCR6 cDNA.

The present invention also provides for antibodies to the DCR6 polypeptide described herein which are useful for detection of the polypeptide in, for example, diagnostic applications. For preparation of monoclonal antibodies directed toward this DCR6 polypeptide, any technique which provides for the production of antibody molecules by continuous cell lines in culture may be used. For example, the hybridoma technique originally developed by Kohler and Milstein (1975, Nature 256:495-497), as well as the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., 1985, in "Monoclonal Antibodies and Cancer Therapy," Alan R. Liss, Inc. pp. 77-96) and the like are within the scope of the present invention.

The monoclonal antibodies for diagnostic or therapeutic use may be human monoclonal antibodies or chimeric human-mouse (or other species) monoclonal antibodies. Human monoclonal antibodies may be made by any of numerous techniques known in the art (e.g., Teng et al., 1983, Proc. Natl. Acad. Sci. U.S.A. 80:7308-7312; Kozbor et al., 1983, Immunology Today 4:72-79; Olsson et al., 1982, Meth. Enzymol. 92:3-16). Chimeric antibody molecules may be prepared containing a mouse antigen-binding domain with human constant regions (Morrison et al., 1984, Proc. Natl. Acad. Sci. U.S.A. 81:6851, Takeda et al., 1985, Nature 314:452).

Various procedures known in the art may be used for the production of polyclonal antibodies to epitopes of the DCR6 polypeptide described herein. For the production of antibody, various host animals can be immunized by injection with the DCR6 polypeptide, or a fragment or derivative thereof, including but not limited to rabbits, mice and rats. Various adjuvants may be used to increase the immunological response, depending on the host species, and including but not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, polypeptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human adjuvants such as BCG (Bacille

Calmette-Guerin) and Corynebacterium parvum.

A molecular clone of an antibody to a selected DCR6 polypeptide epitope can be prepared by known techniques. Recombinant DNA methodology (see e.g., Maniatis et al., 1982, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York) may be used to construct nucleic acid sequences which encode a monoclonal antibody molecule, or antigen binding region thereof.

The present invention provides for antibody molecules as well as fragments of such antibody molecules. Antibody fragments which contain the idiotype of the molecule can be generated by known techniques. For example, such fragments include but are not limited to: the $F(ab')_2$ fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragment, and the Fab fragments which can be generated by treating the antibody molecule with papain and a reducing agent. Antibody molecules may be purified by known techniques, e.g., immunoabsorption or immunoaffinity chromatography, chromatographic methods such as HPLC (high performance liquid chromatography), or a combination thereof.

The invention further provides for a method of using a DCR6 polypeptide or fragment thereof as an antagonist of the activity of a bone morphogenic protein (BMP).

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

EXAMPLES

Example 1: Cloning and Sequencing of "Virtual" Human DCR6

A. "Virtual" cloning

The Human Virtual Transcribed Sequence Database (Kazusa DNA Research Institute, http://zearth.kazusa.or.jp/vts/intro.html), is a database that contains protein sequences that are predicted to be encoded by human genomic sequences. The Human Virtual Transcribed Sequence Project aims to provide candidate transcribed sequences from the available human genome sequencing data by using the gene detection method, GENSCAN (see *infra*) by Chris Burge (cburge@mit.edu). Therefore it is entirely *in silico* gene cloning.

Currently, the database is collecting human genome sequence data from Genbank gss, htg, new, pri1, pri2, entries and from the Web pages of Lawrence Berkeley National Laboratory Human Genome Center, Whitehead Institute/MIT Genome Sequencing Project, The Sanger Centre, Washington University Genome Sequencing Center, Genome Therapeutics Corporation, Japan Science and Technology Corporation, and Yale Center for Medical Informatics.

VTS has been developed by Nobuyuki Miyajima (miyajima@kazusa.or.jp, Kazusa DNA Research Institute) and Toshiyuki Saito (t_saito@nirs.go.jp, National Institute of Radiological Sciences).

GENSCAN is a program designed to predict complete gene structures, including exons, introns, promoter and polyadenylation signals, in genomic sequences. It differs from the majority of existing gene finding algorithms in that it allows for partial genes as well as complete genes and for the occurrence of multiple genes in a single sequence, on either or both DNA strands. The program is based on a probabilistic model of gene structure/compositional properties and does not

make use of protein sequence homology information. The text output of the program is a list of one or more (or possibly zero) predicted genes together with the corresponding peptide sequences. The graphical output (PostScript or gif) is a diagram of the locations of the predicted exons.

In an attempt to clone novel members of the DAN/Cerberus family, the Human Virtual Transcribed Sequence Database was searched by querying with the sequences of several different DAN/Cerberus family members, including the human DCR5 sequence as set forth in co-pending US Provisional Application No. 60/097,296, filed August 20, 1998). A "virtual" predicted polypeptide sequence sharing homology to the human DCR5 query sequence was identified and the corresponding genomic DNA sequence was obtained from the NCBI database (http://www.ncbi.nih.gov; Entrez Search System, nucleotides, Accession #AC003098). This genomic DNA sequence, designated virtual Human DAN/Cerberus related protein 6 (vts_hDCR6) was used to design oligonucleotide primers for use in a PCR-based homology cloning strategy to determine if the "virtual" sequence was in fact transcribed *in vivo*..

Vts_hDCR6 was identified as a predicted open reading frame (ORF) encoding a polypeptide that shares sequence homology with the DAN/Cerberus protein family. The vts_hDCR6 genomic DNA sequence and the regions corresponding to the predicted open reading frame consisting of four exons is set forth in Figure 1A-1F. Because vts_hDCR6 is only a predicted ORF identified by a computer algorithm, it was necessary to (a) show that hDCR6 is expressed in human tissues, (b) determine if the predicted ORF has the same sequence as any actual cDNA clone of hDCR6, and (c) demonstrate that it is a secreted polypeptide.

B. PCR-amplification and cloning of vts hDCR6 exons 1, 2, 3, and 4:

The predicted four exons comprising vts_hDCR6 that are set forth in Figure 1A-1F were each PCR-amplified independently using the following oligonucleotide primers:

Exon 1:

vts_DCR6.ex1 PCR5' (Sal I):

CAG ATA GTC GAC GCC GCC ACC ATG GTG CTC CCA CTG GCC CTG TGT CTC GTC TGC

vts_DCR6.ex1 PCR3' (Spe I):

CTC GAC TAG TGC TTT GGT CTC AAA GGG GTG GTG GGG AGG

Exon 2:

vts_DCR6.ex2 PCR5' (Spe I):

AAA GCA CTA GTC GAG GAA CAG TCT TGC CTG GAG GTG

vts_DCR6.ex2 PCR3' (Eae):

CTC GGC CAC CTT GTT CCC TTC CCA GTG GTA CCA GCA GCT

Exon 3:

vts_DCR6.ex3 PCR5' (Eae):

CAT GTG GCC GAG AAG TCC ACT GCC CAG GCT

vts_DCR6.ex3 PCR3' (Afl 3):

CTC GGA CAC GTA GCC CTT CAG GCA GTC GCT GGA GCC

Exon 4:

vts_DCR6.ex4 PCR5' (Afl 3):

CAG TAC GTG TCC GAG TAC AGC TGC CGC GAG

vts_DCR6.ex4 PCR3' (Not I):

GTA GCG GCC GCC TAG TAG GCG TTC TCC AGC TCG GCC TG

Exons 1, 2, and 3 were PCR-amplified from human genomic DNA using the ExTaq DNA Polymerase PCR system (Panvera, Madison, WI, Cat. #TAKRR001C).

Exon 4 was amplified from human genomic DNA using the ExTaq DNA Polymerase PCR system in conjunction with PCRx Enhancer System (Life Technologies, Inc., Rockville, MD, Cat. # 11495-017). Each PCR-amplified exon was subcloned into the pUC18 vector using the SureClone Ligation Kit (Amersham Pharmacia Biotech AB, Uppsala, Sweden, Cat. #27-9300-01) and standard genetic engineering methodologies (see e.g., Molecular Cloning, A Laboratory Manual (Sambrook, et al., Cold Spring Harbor Laboratory), Current Protocols in Molecular Biology (Eds. Ausubel, et al., Greene Publ. Assoc., Wiley-Interscience, NY). The sequence of each exon was verified using an ABI 373A DNA sequencer and Taq Dideoxy Terminator Cycle Sequencing Kit (Applied Biosystems, Inc., Foster City, CA).

The complete ORF encoding vts_hDCR6 was then genetically engineered by piecing together the four individual exons into the expression vector pCS107 using standard techniques familiar to one of skill in the art. In order to facilitate reconstruction of the vts_hDCR6 ORF into this expression vector, it was necessary to introduce restriction sites between exons to allow for ligating the individual pieces in one unit. However, in each instance, the introduction of restriction sites resulted in silent mutations that did not alter the polypeptide sequence. The sites of exon boundaries are underlined in the sequence set forth in Figure 2A-2B. In addition to the silent mutations described *supra*, the second codon of vts_hDCR6 was changed from CAG to GTG to accommodate a Kozak sequence (Kozak, M., 1987, Nucleic Acids Res. 15:8125-8148) at the 5′ end to promote efficient translational initiation.

Example 2: Northern blot analysis to evaluate the expression profile of hDCR6.

To determine whether vts_hDCR6 is expressed in human tissues, Multiple Tissue Northern blots (Clontech, Palo Alto, CA, Cat. # 7760-1, 7759-1, 7767-1, and 7765-1) were probed using standard Northern blot methodology with a ^{32}P-labeled nucleic acid fragment of vts_hDCR6 consisting of exons 1, 2, and 3. Exon

4 was omitted because its sequence is very GC-rich and as a result is prone to high background levels of non-specific hybridization. The results of the Northern analysis revealed low levels of hDCR6 mRNA expression in the adult kidney and very low levels of expression in heart muscle and colon. The size of the hDCR6 mRNA transcript was approximately 2.4kb.

Example 3: Cloning of hDCR6 by screening human kidney cDNA and a human kidney cDNA library:

Based on the results obtained in the Northern analysis, human kidney cDNA (Clontech, Palo Alto, CA, Cat. #7405-1) was used as a template in the following PCR-based gene cloning strategy. Using the 5′ oligonucleotide primer used to amplify exon 1 of vts_hDCR6 (vts_DCR6.ex1 PCR5' (Sal I)) and the 3′ oligonucleotide primer used to amplify exon 4 of vts_hDCR6 (vts_DCR6.ex4 PCR3' (Not I)) and human kidney cDNA as a template, a PCR reaction was performed. Unexpectedly, the PCR reaction resulted in the amplification of an approximately 0.7kb DNA fragment, rather than the expected 1.2kb fragment predicted by the vts_hDCR6 ORF. Because the size of this fragment was smaller than that expected for vts_hDCR6, it was reasoned that the splicing of the hDCR6 mRNA differed from that of vts_hDCR6. To verify this, the PCR-derived DNA fragment was directly sequenced by standard techniques. The sequence revealed that hDCR6 as expressed in kidney was comprised of exons 1 and 4 of vts_hDCR6 and not any sequence associated with exons 2 and 3.

To obtain a cDNA clone of hDCR6, a human kidney cDNA Rapid-Screen cDNA Library Panel (Origene Technologies, Inc., Rockville, MD, Cat. #LKD-1001) was screened by PCR using the same oligonucleotide primers (vts_DCR6.ex1 PCR5' (Sal I) and vts_DCR6.ex4 PCR3' (Not I)). A full length cDNA clone of hDCR6, comprising only exons 1 and 4 was thus obtained and sequence-verified. The nucleic acid and deduced amino acid sequence of this hDCR6 clone is set forth in Figure 3A-3B. Using the computer program MacVector, it is predicted that the approximately first 20 amino acids encode a signal peptide sequence.

Example 4: Expression pattern of DCR6 in rat tissues.

As described *supra*, Northern analysis revealed that the expression of human DCR6 in adult human tissues is highly restricted to the heart, kidney, and colon (see Table 1).

TABLE 1

Tissue	relative level of expression of hDCR6
adrenal gland	undetectable
bladder (muscle only)	undetectable
bone marrow	undetectable
brain	undetectable
colon (mucosa lining)	**low**
colon (no mucosa) (muscle only)	undetectable
heart	**low**
heart (muscle only)	**medium**
kidney	**high**
liver	undetectable
lung	undetectable
lymph node	undetectable
ovary	undetectable
pancreas	undetectable
peripheral blood leukocytes	undetectable
placenta	undetectable
prostate	undetectable
prostate (muscle only)	undetectable
skeletal muscle	undetectable
sleletal (muscle only)	undetectable
small intestine	undetectable
small intestine (muscle only)	undetectable
spinal chord	undetectable
spleen	undetectable
stomach	undetectable
stomach (muscle only)	undetectable
testis	undetectable
thymus	undetectable
thyroid	undetectable
trachea	undetectable
uterus (no endometrium) (muscle only)	undetectable

Because these data do not yield any information as to which part of the tissue and which cell type(s) human DCR6 is expressed in, the expression of rat DCR6 was also examined in rat embryos at embryonic day 15 (E15) and in adult rat kidneys, using standard *in situ* hybridization techniques. Consecutive sections were hybridized either to a sense or an anti-sense rat DCR6 probe and those tissues that hybridized to the anti-sense but not the sense probe where considered to be positive. By this criteria, rat DCR6 was found to be expressed throughout the choroid plexus (in the brain), in the dorsal surface of the tongue, in the pulmonary artery and aorta, the iliac artery, the lower intestine, and the developing whisker follicles (follicles of vibrissa). There was also expression in the liver either in the lymphatic channels or in the portal veins. In the adult rat kidney, expression of rat DCR6 was restricted to the glomeruli. The association of DCR6 expression with vascular structures indicates that DCR6 may play an important role in the development and homeostasis of these structures. It is also possible that in different diseases (e.g. kidney fibrosis) DCR6 may play an important role.

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 made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule encoding human DCR6.

2. An isolated nucleic acid molecule as in claim 1 having a sequence selected from the group consisting of:
   (a) the nucleotide sequence comprising the coding region of the human DCR6 as set forth in Figure 2A-2B;
   (b) the nucleotide sequence comprising the coding region of the human DCR6 as set forth in Figure 3A-3B;
   (c) a nucleotide sequence that hybridizes under stringent conditions to the nucleotide sequence of (a) or (b) and which encodes DCR6; or
   (d) a nucleotide sequence which, as a result of the degeneracy of the genetic code, differs from the nucleic acid of (a), (b) or (c) and which encodes DCR6.

3. A vector which comprises a nucleic acid molecule of claim 1 or 2.

4. A vector according to claim 3, wherein the nucleic acid molecule is operatively linked to an expression control sequence capable of directing its expression in a host cell.

5. A vector according to claim 3 which is a plasmid.

6. Isolated human DCR6 polypeptide.

7. Isolated human DCR6 polypeptide, having the amino acid sequence as set forth in Figure 2A-2B.

8. Isolated human DCR6 polypeptide, having the amino acid sequence as set forth in Figure 3A-3B.

9. A host-vector system for the production of human DCR6 which comprises a vector of claim 3, in a host cell.

10. A host-vector system according to claim 9, wherein the host cell is a bacterial, yeast, insect or mammalian cell.

11. A method of producing human DCR6 which comprises growing cells of a host-vector system of claim 9, under conditions permitting production of the human DCR6, and recovering the human DCR6 so produced.

12. An antibody which specifically binds the human DCR6 of claim 6, 7, or 8.

13. An antibody according to claim 12, which is a monoclonal antibody.

14. A pharmaceutical composition comprising human DCR6 according to claim 6, 7, or 8, and an acceptable carrier.

15. A pharmaceutical composition comprising an antibody according to claim 12 and an acceptable carrier.

16. Human DCR6 according to claim 6, 7, or 8 for use in a method of treatment of the human or animal body, or in a method of diagnosis.

17. An antibody according to claim 12 for use in a method of treatment of the human or animal body, or in a method of diagnosis.

18. A composition according to claim 14 for use in a method of treatment of the human or animal body, or in a method of diagnosis.

19. A polypeptide produced by the method of claim 11.

20. A ligandbody which comprises human DCR6 fused to an immunoglobulin constant region.

21. The ligandbody of claim 20, wherein the immunoglobulin constant region is the Fc portion of human IgG1.

22. A ligandbody according to claim 20 or 21, for use in a method of treatment of the human or animal body, or in a method of diagnosis.

Figure 1A

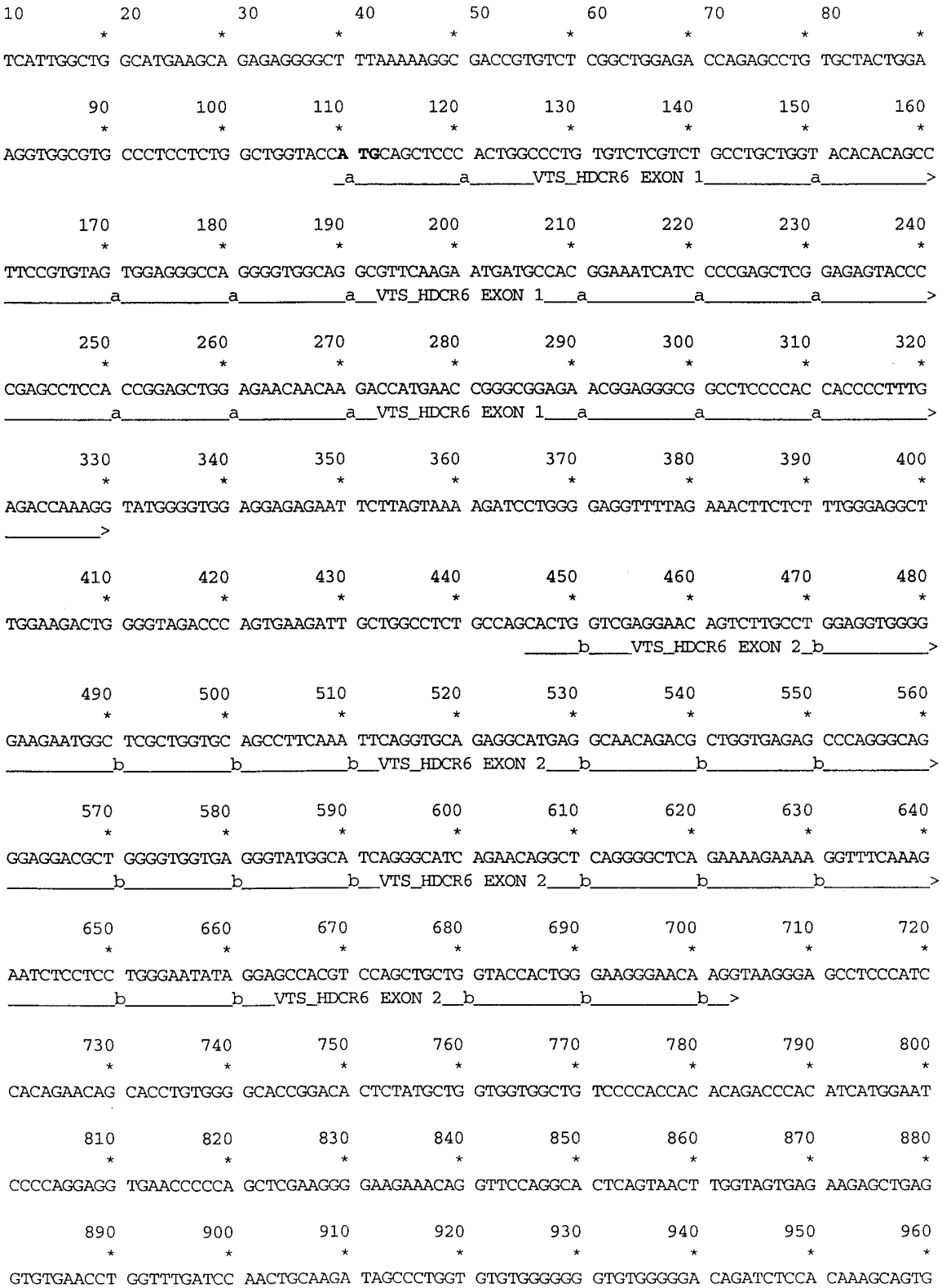
10 20 30 40 50 60 70 80
TCATTGGCTG GCATGAAGCA GAGAGGGGCT TTAAAAAGGC GACCGTGTCT CGGCTGGAGA CCAGAGCCTG TGCTACTGGA
90 100 110 120 130 140 150 160
AGGTGGCGTG CCCTCCTCTG GCTGGTACCA TGCAGCTCCC ACTGGCCCTG TGTCTCGTCT GCCTGCTGGT ACACACAGCC
_a_____a_____VTS_HDCR6 EXON 1_____a_____>
170 180 190 200 210 220 230 240
TTCCGTGTAG TGGAGGGCCA GGGGTGGCAG GCGTTCAAGA ATGATGCCAC GGAAATCATC CCCGAGCTCG GAGAGTACCC
_____a_____a_____a__VTS_HDCR6 EXON 1___a_____a_____a_____>
250 260 270 280 290 300 310 320
CGAGCCTCCA CCGGAGCTGG AGAACAACAA GACCATGAAC CGGGCGGAGA ACGGAGGGCG GCCTCCCCAC CACCCCTTTG
_____a_____a_____a__VTS_HDCR6 EXON 1___a_____a_____a_____>
330 340 350 360 370 380 390 400
AGACCAAAGG TATGGGGTGG AGGAGAGAAT TCTTAGTAAA AGATCCTGGG GAGGTTTTAG AAACTTCTCT TTGGGAGGCT
_____>
410 420 430 440 450 460 470 480
TGGAAGACTG GGGTAGACCC AGTGAAGATT GCTGGCCTCT GCCAGCACTG GTCGAGGAAC AGTCTTGCCT GGAGGTGGGG
_____b____VTS_HDCR6 EXON 2_b_____>
490 500 510 520 530 540 550 560
GAAGAATGGC TCGCTGGTGC AGCCTTCAAA TTCAGGTGCA GAGGCATGAG GCAACAGACG CTGGTGAGAG CCCAGGGCAG
_____b_____b_____b__VTS_HDCR6 EXON 2___b_____b_____b_____>
570 580 590 600 610 620 630 640
GGAGGACGCT GGGGTGGTGA GGGTATGGCA TCAGGGCATC AGAACAGGCT CAGGGGCTCA GAAAAGAAAA GGTTTCAAAG
_____b_____b_____b__VTS_HDCR6 EXON 2___b_____b_____b_____>
650 660 670 680 690 700 710 720
AATCTCCTCC TGGGAATATA GGAGCCACGT CCAGCTGCTG GTACCACTGG GAAGGGAACA AGGTAAGGGA GCCTCCCATC
_____b_____b___VTS_HDCR6 EXON 2__b_____b_____b__>
730 740 750 760 770 780 790 800
CACAGAACAG CACCTGTGGG GCACCGGACA CTCTATGCTG GTGGTGGCTG TCCCCACCAC ACAGACCCAC ATCATGGAAT
810 820 830 840 850 860 870 880
CCCCAGGAGG TGAACCCCCA GCTCGAAGGG GAAGAAACAG GTTCCAGGCA CTCAGTAACT TGGTAGTGAG AAGAGCTGAG
890 900 910 920 930 940 950 960
GTGTGAACCT GGTTTGATCC AACTGCAAGA TAGCCCTGGT GTGTGGGGGG GTGTGGGGGA CAGATCTCCA CAAAGCAGTG

Figure 1B

```
      970        980        990       1000       1010       1020       1030       1040
       *          *          *          *          *          *          *          *
GGGAGGAAGG CCAGAGAGGC ACCCCTGCAG TGTGCATTGC CCATGGCCTG CCCAGGGAGC TGGCACTTGA AGGAATGGGA

     1050       1060       1070       1080       1090       1100       1110       1120
       *          *          *          *          *          *          *          *
GTTTTCGGCA CAGTTTTAGC CCCTGACATG GGTGCAGCTG AGTCCAGGCC CTGGAGGGGA GAGCAGCATC CTCTGTGCAG

     1130       1140       1150       1160       1170       1180       1190       1200
       *          *          *          *          *          *          *          *
GAGTAGGGAC ATCTGTCCTC AGCAGCCACC CCAGTCCCAA CCTTGCCTCA TTCCAGGGGA GGGAGAAGGA AGAGGAACCC

     1210       1220       1230       1240       1250       1260       1270       1280
       *          *          *          *          *          *          *          *
TGGGTTCCTG GTCAGGCCTG CACAGAGAAG CCCAGGTGAC AGTGTGCATC TGGCTCTATA ATTGGCAGGA ATCCTGAGGC

     1290       1300       1310       1320       1330       1340       1350       1360
       *          *          *          *          *          *          *          *
CATGGGGGCG TCTGAAATGA CACTTCAGAC TAAGAGCTTC CCTGTCCTCT GGCCATTATC CAGGTGGCAG AGAAGTCCAC
                                                                    ___VTS_HDCR6 E____>

     1370       1380       1390       1400       1410       1420       1430       1440
       *          *          *          *          *          *          *          *
TGCCCAGGCT CCTGGACCCC AGCCCTCCCC GCCTCACAAC CTGTTGGGAC TATGGGGTGC TAAAAAGGGC AACTGCATGG
__________c__________c__________c__VTS_HDCR6 EXON 3___c__________c__________c__________>

     1450       1460       1470       1480       1490       1500       1510       1520
       *          *          *          *          *          *          *          *
GAGGCCAGCC AGGACCCTCC GTCTTCAAAA TGGAGGACAA GGGCGCCTCC CCCCACAGCT CCCCTTCTAG GCAAGGTCAG
__________c__________c__________c__VTS_HDCR6 EXON 3___c__________c__________c__________>

     1530       1540       1550       1560       1570       1580       1590       1600
       *          *          *          *          *          *          *          *
CTGGGCTCCA GCGACTGCCT GAAGGGCTGT AAGGAACCCA AACACAAAAT GTCCACCTTG CTGGACTCCC ACGAGAGGCC
_______VTS_HDCR6 EXON 3_________>

     1610       1620       1630       1640       1650       1660       1670       1680
       *          *          *          *          *          *          *          *
ACAGCCCCTG AGGAAGCCAC ATGCTCAAAA CAAAGTCATG ATCTGCAGAG GAAGTGCCTG GCCTAGGGGC GCTATTCTCG

     1690       1700       1710       1720       1730       1740       1750       1760
       *          *          *          *          *          *          *          *
AAAAGCCGCA AAATGCCCCC TTCCCTGGGC AAATGCCCCC CTGACCACAC ACACATTCCA GCCCTGCAGA GGTGAGGATG

     1770       1780       1790       1800       1810       1820       1830       1840
       *          *          *          *          *          *          *          *
CAAACCAGCC CACAGACCAG AAAGCAGCCC CAGACGATGG CAGTGGCCAC ATCTCCCCTG CTGTGCTTGC TCTTCAGAGT

     1850       1860       1870       1880       1890       1900       1910       1920
       *          *          *          *          *          *          *          *
GGGGGTGGGG GGTGGCCTTC TCTGTCCCCT CTCTGGTTTG GTCTTAAGAC TATTTTTCAT TCTTTCTTGT CACATTGGAA
```

Figure 1C

```
1930        1940        1950        1960        1970        1980        1990        2000
*           *           *           *           *           *           *           *
CTATCCCCAT GAAACCTTTG GGGGTGGACT GGTACTCACA CGACGACCAG CTATTTAAAA AGCTCCCACC CATCTAAGTC

2010        2020        2030        2040        2050        2060        2070        2080
*           *           *           *           *           *           *           *
CACCATAGGA GACATGGTCA AGGTGTGTGC AGGGGATCAG GCCAGGCCTC GGAGCCCAAT CTCTGCCTGC CCAGGGAGTA

2090        2100        2110        2120        2130        2140        2150        2160
*           *           *           *           *           *           *           *
TCACCATGAG GCGCCCATTC AGATAACACA GAACAAGAAA TGTGCCCAGC AGAGAGCCAG GTCAATGTTT GTGGCAGCTG

2170        2180        2190        2200        2210        2220        2230        2240
*           *           *           *           *           *           *           *
AACCTGTAGG TTTTGGGTCA GAGCTCAGGG CCCCTATGGT AGGAAAGTAA CGACAGTAAA AAGCAGCCCT CAGCTCCATC

2250        2260        2270        2280        2290        2300        2310        2320
*           *           *           *           *           *           *           *
CCCCAGCCCA GCCTCCCATG GATGCTCGAA CGCAGAGCCT CCACTCTTGC CGGAGCCAAA AGGTGCTGGG ACCCCAGGGA

2330        2340        2350        2360        2370        2380        2390        2400
*           *           *           *           *           *           *           *
AGTGGAGTCC GGAGATGCAG CCCAGCCTTT TGGGCAAGTT CTTTTCTCTG GCTGGGCCTC AGTATTCTCA TTGATAATGA

2410        2420        2430        2440        2450        2460        2470        2480
*           *           *           *           *           *           *           *
GGGGGTTGGA CACACTGCCT TTGATTCCTT TCAAGTCTAA TGAATTCCTG TCCTGATCAC CTCCCCTTCA GTCCCTCGCC

2490        2500        2510        2520        2530        2540        2550        2560
*           *           *           *           *           *           *           *
TCCACAGCAG CTGCCCTGAT TTATTACCTT CAATTAACCT CTACTCCTTT CTCCATCCCC TGTCCACCCC TCCCAAGTGG

2570        2580        2590        2600        2610        2620        2630        2640
*           *           *           *           *           *           *           *
CTGGAAAAGG AATTTGGGAG AAGCCAGAGC CAGGCAGAAG GTGTGCTGAG TACTTACCCT GCCCAGGCCA GGGACCCTGC

2650        2660        2670        2680        2690        2700        2710        2720
*           *           *           *           *           *           *           *
GGCACAAGTG TGGCTTAAAT CATAAGAAGA CCCCAGAAGA GAAATGATAA TAATAATACA TAACAGCCGA CGCTTTCAGC

2730        2740        2750        2760        2770        2780        2790        2800
*           *           *           *           *           *           *           *
TATATGTGCC AAATGGTATT TTCTGCATTG CGTGTGTAAT GGATTAACTC GCAATGCTTG GGGCGGCCCA TTTTGCAGAC

2810        2820        2830        2840        2850        2860        2870        2880
*           *           *           *           *           *           *           *
AGGAAGAAGA GAGAGGTTAA GGAACTTGCC CAAGATGACA CCTGCAGTGA GCGATGGAGC CCTGGTGTTT GAACCCCAGC

2890        2900        2910        2920        2930        2940        2950        2960
*           *           *           *           *           *           *           *
AGTCATTTGG CTCCGAGGGG ACAGGGTGCG CAGGAGAGCT TTCCACCAGC TCTAGAGCAT CTGGGACCTT CCTGCAATAG
```

Figure 1D

```
      2970       2980       2990       3000       3010       3020       3030       3040
        *          *          *          *          *          *          *          *
ATGTTCAGGG GCAAAAGCCT CTGGAGACAG GCTTGGCAAA AGCAGGGCTG GGGTGGAGAG AGACGGGCCG GTCCAGGGCA

      3050       3060       3070       3080       3090       3100       3110       3120
        *          *          *          *          *          *          *          *
GGGGTGGCCA GGCGGGCGGC CACCCTCACG CGCGCCTCTC TCCACAGACG TGTCCGAGTA CAGCTGCCGC GAGCTGCACT
                                                  ___d_____VTS_HDCR6 EXON 4___________>

      3130       3140       3150       3160       3170       3180       3190       3200
        *          *          *          *          *          *          *          *
TCACCCGCTA CGTGACCGAT GGGCCGTGCC GCAGCGCCAA GCCGGTCACC GAGCTGGTGT GCTCCGGCCA GTGCGGCCCG
__________d__________d__________d__VTS_HDCR6 EXON 4___d__________d__________d__________>

      3210       3220       3230       3240       3250       3260       3270       3280
        *          *          *          *          *          *          *          *
GCGCGCCTGC TGCCCAACGC CATCGGCCGC GGCAAGTGGT GGCGACCTAG TGGGCCCGAC TTCCGCTGCA TCCCCGACCG
__________d__________d__________d__VTS_HDCR6 EXON 4___d__________d__________d__________>

      3290       3300       3310       3320       3330       3340       3350       3360
        *          *          *          *          *          *          *          *
CTACCGCGCG CAGCGCGTGC AGCTGCTGTG TCCCGGTGGT GAGGCGCCGC GCGCGCGCAA GGTGCGCCTG GTGGCCTCGT
__________d__________d__________d__VTS_HDCR6 EXON 4___d__________d__________d__________>

      3370       3380       3390       3400       3410       3420       3430       3440
        *          *          *          *          *          *          *          *
GCAAGTGCAA GCGCCTCACC CGCTTCCACA ACCAGTCGGA GCTCAAGGAC TTCGGGACCG AGGCCGCTCG GCCGCAGAAG
__________d__________d__________d__VTS_HDCR6 EXON 4___d__________d__________d__________>

      3450       3460       3470       3480       3490       3500       3510       3520
        *          *          *          *          *          *          *          *
GGCCGGAAGC CGCGGCCCCG CGCCCGGAGC GCCAAAGCCA ACCAGGCCGA GCTGGAGAAC GCCTACTAGA GCCCGCCCGC
__________d__________d_______VTS_HDCR6 EXON 4__________d__________d__________>

      3530       3540       3550       3560       3570       3580       3590       3600
        *          *          *          *          *          *          *          *
GCCCCTCCCC ACCGGCGGGC GCCCCGGCCC TGAACCCGCG CCCCACATTT CTGTCCTCTG CGCGTGGTTT GATTGTTTAT

      3610       3620       3630       3640       3650       3660       3670       3680
        *          *          *          *          *          *          *          *
ATTTCATTGT AAATGCCTGC AACCCAGGGC AGGGGGCTGA GACCTTCCAG GCCCTGAGGA ATCCCGGGCG CCGGCAAGGC

      3690       3700       3710       3720       3730       3740       3750       3760
        *          *          *          *          *          *          *          *
CCCCCTCAGC CCGCCAGCTG AGGGGTCCCA CGGGGCAGGG GAGGGAATTG AGAGTCACAG ACACTGAGCC ACGCAGCCCC

      3770       3780       3790       3800       3810       3820       3830       3840
        *          *          *          *          *          *          *          *
GCCTCTGGGG CCGCCTACCT TTGCTGGTCC CACTTCAGAG GAGGCAGAAA TGGAAGCATT TTCACCGCCC TGGGGTTTTA

      3850       3860       3870       3880       3890       3900       3910       3920
        *          *          *          *          *          *          *          *
AGGGAGCGGT GTGGGAGTGG GAAAGTCCAG GGACTGGTTA AGAAAGTTGG ATAAGATTCC CCCTTGCACC TCGCTGCCCA
```

Figure 1E

```
      3930       3940       3950       3960       3970       3980       3990       4000
        *          *          *          *          *          *          *          *
TCAGAAAGCC TGAGGCGTGC CCAGAGCACA AGACTGGGGG CAACTGTAGA TGTGGTTTCT AGTCCTGGCT CTGCCACTAA

4010       4020       4030       4040       4050       4060       4070       4080
        *          *          *          *          *          *          *          *
CTTGCTGTGT AACCTTGAAC TACACAATTC TCCTTCGGGA CCTCAATTTC CACTTTGTAA AATGAGGGTG GAGGTGGGAA

      4090       4100       4110       4120       4130       4140       4150       4160
        *          *          *          *          *          *          *          *
TAGGATCTCG AGGAGACTAT TGGCATATGA TTCCAAGGAC TCCAGTGCCT TTTGAATGGG CAGAGGTGAG AGAGAGAGAG

      4170       4180       4190       4200       4210       4220       4230       4240
        *          *          *          *          *          *          *          *
AGAAAGAGAG AGAATGAATG CAGTTGCATT GATTCAGTGC CAAGGTCACT TCCAGAATTC AGAGTTGTGA TGCTCTCTTC

      4250       4260       4270       4280       4290       4300       4310       4320
        *          *          *          *          *          *          *          *
TGACAGCCAA AGATGAAAAA CAAACAGAAA AAAAAAAGTA AAGAGTCTAT TTATGGCTGA CATATTTACG GCTGACAAAC

      4330       4340       4350       4360       4370       4380       4390       4400
        *          *          *          *          *          *          *          *
TCCTGGAAGA AGCTATGCTG CTTCCCAGCC TGGCTTCCCC GGATGTTTGG CTACCTCCAC CCCTCCATCT CAAAGAAATA

      4410       4420       4430       4440       4450       4460       4470       4480
        *          *          *          *          *          *          *          *
ACATCATCCA TTGGGGTAGA AAAGGAGAGG GTCCGAGGGT GGTGGGAGGG ATAGAAATCA CATCCGCCCC AACTTCCCAA

      4490       4500       4510       4520       4530       4540       4550       4560
        *          *          *          *          *          *          *          *
AGAGCAGCAT CCCTCCCCCG ACCCATAGCC ATGTTTTAAA GTCACCTTCC GAAGAGAAGT GAAAGGTTCA AGGACACTGG

      4570       4580       4590       4600       4610       4620       4630       4640
        *          *          *          *          *          *          *          *
CCTTGCAGGC CCGAGGGAGC AGCCATCACA AACTCACAGA CCAGCACATC CCTTTTGAGA CACCGCCTTC TGCCCACCAC

      4650       4660       4670       4680       4690       4700       4710       4720
        *          *          *          *          *          *          *          *
TCACGGACAC ATTTCTGCCT AGAAAACAGC TTCTTACTGC TCTTACATGT GATGGCATAT CTTACACTAA AAGAATATTA

      4730       4740       4750       4760       4770       4780       4790       4800
        *          *          *          *          *          *          *          *
TTGGGGGAAA AACTACAAGT GCTGTACATA TGCTGAGAAA CTGCAGAGCA TAATAGCTGC CACCCAAAAA TCTTTTTGAA

      4810       4820       4830       4840       4850       4860       4870       4880
        *          *          *          *          *          *          *          *
AATCATTTCC AGACAACCTC TTACTTTCTG TGTAGTTTTT AATTGTTAAA AAAAAAAAGT TTTAAACAGA AGCACATGAC

      4890       4900       4910       4920       4930       4940       4950       4960
        *          *          *          *          *          *          *          *
ATATGAAAGC CTGCAGGACT GGTCGTTTTT TTGGCAATTC TTCCACGTGG GACTTGTCCA CAAGAATGAA AGTAGTGGTT
```

Figure 1F

```
4970        4980        4990        5000        5010        5020        5030        5040
*           *           *           *           *           *           *           *
TTTAAAGAGT TAAGTTACAT ATTTATTTTC TCACTTAAGT TATTTATGCA AAAGTTTTTC TTGTAGAGAA TGACAATGTT

5050        5060        5070        5080        5090        5100        5110        5120
*           *           *           *           *           *           *           *
AATATTGCTT TATGAATTAA CAGTCTGTTC TTCCAGAGTC CAGAGACATT GTTAATAAAG ACAATGAATC ATGACCGAAA

5130        5140        5150        5160        5170        5180        5190        5200
*           *           *           *           *           *           *           *
GGATGTGGTC TCATTTTGTC AACCACACAT GACGTCATTT CTGTCAAAGT TGACACCCTT CTCTTGGTCA CTAGAGCTCC

5210        5220        5230        5240        5250        5260        5270        5280
*           *           *           *           *           *           *           *
AACCTTGGAC ACACCTTTGA CTGCTCTCTG GTGGCCCTTG TGGCAATTAT GTCTTCCTTT GAAAAGTCAT GTTTATCCCT

5290        5300        5310        5320        5330        5340        5350        5360
*           *           *           *           *           *           *           *
TCCTTTCCAA ACCCAGACCG CATTTCTTCA CCCAGGGCAT GGTAATAACC TCAGCCTTGT ATCCTTTTAG CAGCCTCCCC

5370        5380        5390        5400        5410        5420        5430        5440
*           *           *           *           *           *           *           *
TCCATGCTGG CTTCCAAAAT GCTGTTCTCA TTGTATCACT CCCCTGCTCA AAAGCCTTCC ATAGCTCCCC CTTGCCCAGG

5450        5460        5470        5480        5490        5500        5510        5520
*           *           *           *           *           *           *           *
ATCAAGTGCA GTTTCCCTAT CTGACATGGG AGGCCTTCTC TGCTTGACTC CCACCTCCCA CTCCACCAAG CTTCCTACTG

5530        5540        5550        5560        5570        5580        5590        5600
*           *           *           *           *           *           *           *
ACTCCAAATG GTCATGCAGA TCCCTGCTTC CTTAGTTTGC CATCCACACT TAGCACCCCC AATAACTAAT CCTCTTTCTT

5610        5620        5630        5640        5650        5660        5670        5680
*           *           *           *           *           *           *           *
TAGGATTCAC ATTACTTGTC ATCTCTTCCC CTAACCTTCC AGAGATGTTC CAATCTCCCA TGATCCCTCT CTCCTCTGAG
```

Figure 2A

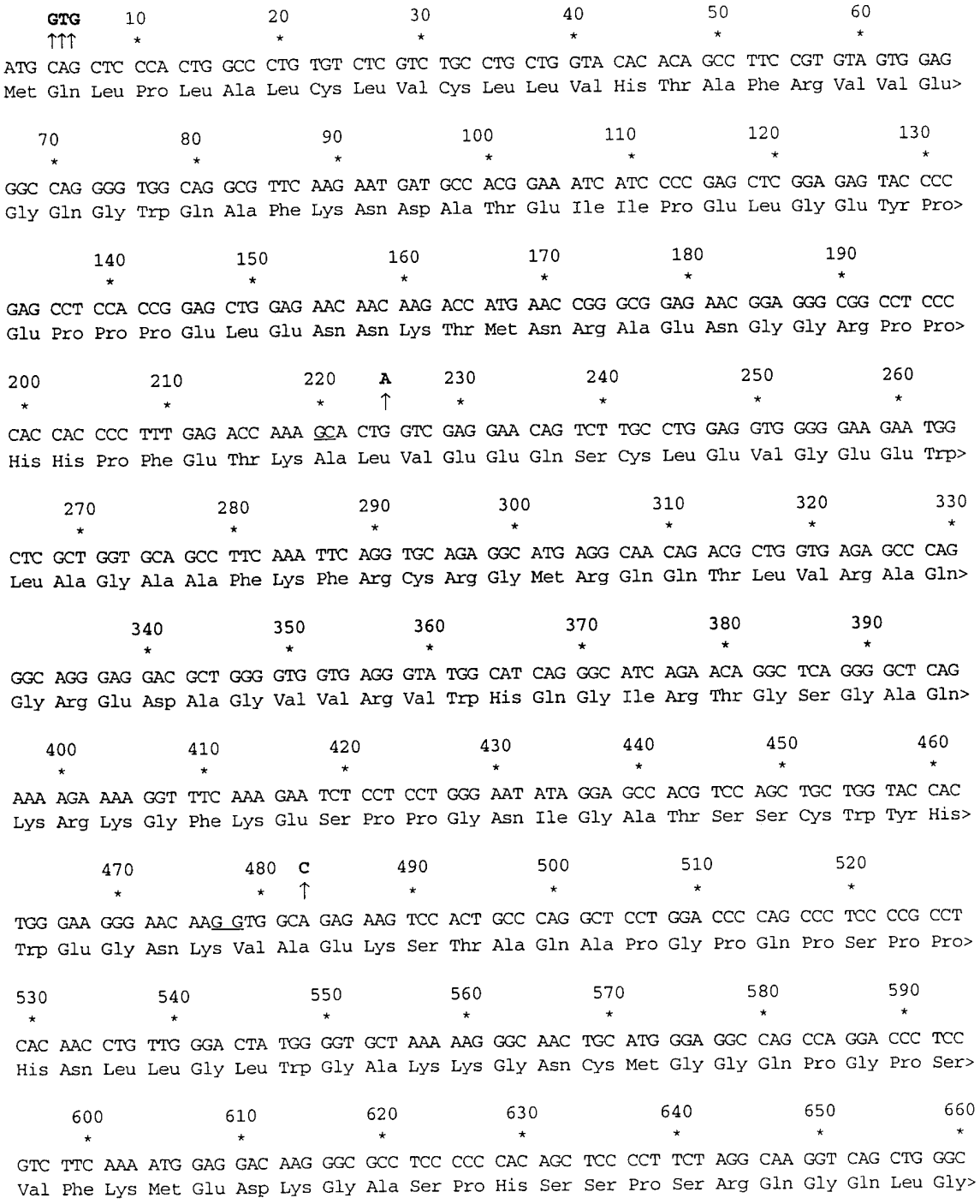
GTG
10 20 30 40 50 60
ATG CAG CTC CCA CTG GCC CTG TGT CTC GTC TGC CTG CTG GTA CAC ACA GCC TTC CGT GTA GTG GAG
Met Gln Leu Pro Leu Ala Leu Cys Leu Val Cys Leu Leu Val His Thr Ala Phe Arg Val Val Glu>
70 80 90 100 110 120 130
GGC CAG GGG TGG CAG GCG TTC AAG AAT GAT GCC ACG GAA ATC ATC CCC GAG CTC GGA GAG TAC CCC
Gly Gln Gly Trp Gln Ala Phe Lys Asn Asp Ala Thr Glu Ile Ile Pro Glu Leu Gly Glu Tyr Pro>
140 150 160 170 180 190
GAG CCT CCA CCG GAG CTG GAG AAC AAC AAG ACC ATG AAC CGG GCG GAG AAC GGA GGG CGG CCT CCC
Glu Pro Pro Pro Glu Leu Glu Asn Asn Lys Thr Met Asn Arg Ala Glu Asn Gly Gly Arg Pro Pro>
200 210 220 A 230 240 250 260
CAC CAC CCC TTT GAG ACC AAA GCA CTG GTC GAG GAA CAG TCT TGC CTG GAG GTG GGG GAA GAA TGG
His His Pro Phe Glu Thr Lys Ala Leu Val Glu Glu Gln Ser Cys Leu Glu Val Gly Glu Glu Trp>
270 280 290 300 310 320 330
CTC GCT GGT GCA GCC TTC AAA TTC AGG TGC AGA GGC ATG AGG CAA CAG ACG CTG GTG AGA GCC CAG
Leu Ala Gly Ala Ala Phe Lys Phe Arg Cys Arg Gly Met Arg Gln Gln Thr Leu Val Arg Ala Gln>
340 350 360 370 380 390
GGC AGG GAG GAC GCT GGG GTG GTG AGG GTA TGG CAT CAG GGC ATC AGA ACA GGC TCA GGG GCT CAG
Gly Arg Glu Asp Ala Gly Val Val Arg Val Trp His Gln Gly Ile Arg Thr Gly Ser Gly Ala Gln>
400 410 420 430 440 450 460
AAA AGA AAA GGT TTC AAA GAA TCT CCT CCT GGG AAT ATA GGA GCC ACG TCC AGC TGC TGG TAC CAC
Lys Arg Lys Gly Phe Lys Glu Ser Pro Pro Gly Asn Ile Gly Ala Thr Ser Ser Cys Trp Tyr His>
470 480 C 490 500 510 520
TGG GAA GGG AAC AAG GTG GCA GAG AAG TCC ACT GCC CAG GCT CCT GGA CCC CAG CCC TCC CCG CCT
Trp Glu Gly Asn Lys Val Ala Glu Lys Ser Thr Ala Gln Ala Pro Gly Pro Gln Pro Ser Pro Pro>
530 540 550 560 570 580 590
CAC AAC CTG TTG GGA CTA TGG GGT GCT AAA AAG GGC AAC TGC ATG GGA GGC CAG CCA GGA CCC TCC
His Asn Leu Leu Gly Leu Trp Gly Ala Lys Lys Gly Asn Cys Met Gly Gly Gln Pro Gly Pro Ser>
600 610 620 630 640 650 660
GTC TTC AAA ATG GAG GAC AAG GGC GCC TCC CCC CAC AGC TCC CCT TCT AGG CAA GGT CAG CTG GGC
Val Phe Lys Met Glu Asp Lys Gly Ala Ser Pro His Ser Ser Pro Ser Arg Gln Gly Gln Leu Gly>

Figure 2B

```
          670         680         690         700         710         720
           *           *           *           *           *           *
TCC AGC GAC TGC CTG AAG GGC TAC GTG TCC GAG TAC AGC TGC CGC GAG CTG CAC TTC ACC CGC TAC
Ser Ser Asp Cys Leu Lys Gly Tyr Val Ser Glu Tyr Ser Cys Arg Glu Leu His Phe Thr Arg Tyr>

   730         740         750         760         770         780         790
    *           *           *           *           *           *           *
GTG ACC GAT GGG CCG TGC CGC AGC GCC AAG CCG GTC ACC GAG CTG GTG TGC TCC GGC CAG TGC GGC
Val Thr Asp Gly Pro Cys Arg Ser Ala Lys Pro Val Thr Glu Leu Val Cys Ser Gly Gln Cys Gly>

        800         810         820         830         840         850
         *           *           *           *           *           *
CCG GCG CGC CTG CTG CCC AAC GCC ATC GGC CGC GGC AAG TGG TGG CGA CCT AGT GGG CCC GAC TTC
Pro Ala Arg Leu Leu Pro Asn Ala Ile Gly Arg Gly Lys Trp Trp Arg Pro Ser Gly Pro Asp Phe>

860         870         880         890         900         910         920
 *           *           *           *           *           *           *
CGC TGC ATC CCC GAC CGC TAC CGC GCG CAG CGC GTG CAG CTG CTG TGT CCC GGT GGT GAG GCG CCG
Arg Cys Ile Pro Asp Arg Tyr Arg Ala Gln Arg Val Gln Leu Leu Cys Pro Gly Gly Glu Ala Pro>

     930         940         950         960         970         980         990
      *           *           *           *           *           *           *
CGC GCG CGC AAG GTG CGC CTG GTG GCC TCG TGC AAG TGC AAG CGC CTC ACC CGC TTC CAC AAC CAG
Arg Ala Arg Lys Val Arg Leu Val Ala Ser Cys Lys Cys Lys Arg Leu Thr Arg Phe His Asn Gln>

          1000        1010        1020        1030        1040        1050
           *           *           *           *           *           *
TCG GAG CTC AAG GAC TTC GGG ACC GAG GCC GCT CGG CCG CAG AAG GGC CGG AAG CCG CGG CCC CGC
Ser Glu Leu Lys Asp Phe Gly Thr Glu Ala Ala Arg Pro Gln Lys Gly Arg Lys Pro Arg Pro Arg>

  1060        1070        1080        1090        1100
   *           *           *           *           *
GCC CGG AGC GCC AAA GCC AAC CAG GCC GAG CTG GAG AAC GCC TAC TAG
Ala Arg Ser Ala Lys Ala Asn Gln Ala Glu Leu Glu Asn Ala Tyr ***>
```

Figure 3A

```
          10          20          30          40          50          60
           *           *           *           *           *           *
ATG CAG CTC CCA CTG GCC CTG TGT CTC GTC TGC CTG CTG GTA CAC ACA GCC TTC CGT GTA GTG GAG
Met Gln Leu Pro Leu Ala Leu Cys Leu Val Cys Leu Leu Val His Thr Ala Phe Arg Val Val Glu>
                                  (10)                                    (20)       22

   70          80          90          100         110         120         130
    *           *           *           *           *           *           *
GGC CAG GGG TGG CAG GCG TTC AAG AAT GAT GCC ACG GAA ATC ATC CCC GAG CTC GGA GAG TAC CCC
Gly Gln Gly Trp Gln Ala Phe Lys Asn Asp Ala Thr Glu Ile Ile Pro Glu Leu Gly Glu Tyr Pro>
                          (30)                                    (40)              44

        140         150         160         170         180         190
         *           *           *           *           *           *
GAG CCT CCA CCG GAG CTG GAG AAC AAC AAG ACC ATG AAC CGG GCG GAG AAC GGA GGG CGG CCT CCC
Glu Pro Pro Pro Glu Leu Glu Asn Asn Lys Thr Met Asn Arg Ala Glu Asn Gly Gly Arg Pro Pro>
                  (50)                                    (60)                      66

200         210         220         230         240         250         260
 *           *           *           *           *           *           *
CAC CAC CCC TTT GAG ACC AAA GAC GTG TCC GAG TAC AGC TGC CGC GAG CTG CAC TTC ACC CGC TAC
His His Pro Phe Glu Thr Lys Asp Val Ser Glu Tyr Ser Cys Arg Glu Leu His Phe Thr Arg Tyr>
           (70)                                    (80)                              88

     270         280         290         300         310         320         330
      *           *           *           *           *           *           *
GTG ACC GAT GGG CCG TGC CGC AGC GCC AAG CCG GTC ACC GAG CTG GTG TGC TCC GGC CAG TGC GGC
Val Thr Asp Gly Pro Cys Arg Ser Ala Lys Pro Val Thr Glu Leu Val Cys Ser Gly Gln Cys Gly>
   (90)                                    (100)                                    110

          340         350         360         370         380         390
           *           *           *           *           *           *
CCG GCG CGC CTG CTG CCC AAC GCC ATC GGC CGC GGC AAG TGG TGG CGA CCT AGT GGG CCC GAC TTC
Pro Ala Arg Leu Leu Pro Asn Ala Ile Gly Arg Gly Lys Trp Trp Arg Pro Ser Gly Pro Asp Phe>
                                  (120)                                   (130)      132

   400         410         420         430         440         450         460
    *           *           *           *           *           *           *
CGC TGC ATC CCC GAC CGC TAC CGC GCG CAG CGC GTG CAG CTG CTG TGT CCC GGT GGT GAG GCG CCG
Arg Cys Ile Pro Asp Arg Tyr Arg Ala Gln Arg Val Gln Leu Leu Cys Pro Gly Gly Glu Ala Pro>
                          (140)                                   (150)              154

        470         480         490         500         510         520
         *           *           *           *           *           *
CGC GCG CGC AAG GTG CGC CTG GTG GCC TCG TGC AAG TGC AAG CGC CTC ACC CGC TTC CAC AAC CAG
Arg Ala Arg Lys Val Arg Leu Val Ala Ser Cys Lys Cys Lys Arg Leu Thr Arg Phe His Asn Gln>
                  (160)                                   (170)                      176

530         540         550         560         570         580         590
 *           *           *           *           *           *           *
TCG GAG CTC AAG GAC TTC GGG ACC GAG GCC GCT CGG CCG CAG AAG GGC CGG AAG CCG CGG CCC CGC
Ser Glu Leu Lys Asp Phe Gly Thr Glu Ala Ala Arg Pro Gln Lys Gly Arg Lys Pro Arg Pro Arg>
           (180)                                   (190)                              198
```

Figure 3B

```
     600           610           620           630           640
      *             *             *             *             *
GCC CGG AGC GCC AAA GCC AAC CAG GCC GAG CTG GAG AAC GCC TAC TAG
Ala Arg Ser Ala Lys Ala Asn Gln Ala Glu Leu Glu Asn Ala Tyr ***>
   (200)                                       (210)       213
```