Ligand binding domains of nuclear receptors in controllable form and methods involving the same

The present invention relates to an isolated protein comprising a ligand binding domain of a nuclear receptor in controllable form, a method of producing the same, its use for the identification of a ligand, a test system comprising the isolated protein and a method for screening for a ligand for a nuclear receptor using the test system.

Nuclear receptors represent a superfamily of proteins which are found within cells and which induce signals of ligands such as hormones and vitamins. In response, agonist-activated nuclear receptors usually increase expression of specific genes upon activation in general together with other proteins.

Thus, nuclear receptors act as agonist-induced transcription factors which directly interact as monomers, homodimers or heterodimers with DNA response element of target genes as well as through signaling pathways. In contrast to membrane receptors and membrane-associated receptors, nuclear receptors reside within cells, either in cytoplasm or in the nucleus. Thus, nuclear receptors comprise a class of intercellular, soluble, ligand-regulated factors which are found in eukaryotic cells. Nuclear receptors have the ability to directly bind to DNA and regulate the expression of adjacent genes; hence these receptors are classified as transcription factors. As detailed above, the regulation of gene expression by nuclear receptor is ligand-dependent, wherein nuclear receptors are normally only active in the presence of an agonist. Ligand binding to a nuclear receptor results in a conformational change in the receptor, which in turn activates the receptor resulting in general in up-regulation of gene expression.

Due to their unique ability to directly interact with and control the expression of genomic DNA, nuclear receptors play a key role in development and homeostasis of organisms.

The members of the superfamily of nuclear receptors display an overall structural motif of four modular domains:

- A variable amino-terminal domain (also referred to as N-terminal regulatory domain), which contains activation function 1 (AF-1), whose action is independent of the presence of a ligand. The transcriptional activation of AF-1 is normally weak,

but synergizes with AF-2 to up- regulate gene expression. This domain is highly variable in sequence between various nuclear receptors.

- A highly conserved DNA-binding domain (DBD) contains two zinc fingers and binds to hormone response elements (HREs).
- A less conserved ligand binding domain (LBD), though only moderately conserved in sequence, is highly conserved in structure among the various nuclear receptors. The structure of the LBD is referred to as an alpha-helical sandwich fold. The ligand binding cavity is within the interior of the LBD just below three anti-parallel alpha helices forming the "sandwich filling". Along with the DBD, the LBD contributes to the dimerization interface of the receptor and, in addition, binds co-activator and co-repressor proteins. Additionally, it contains the activation function 2 (AF-2), whose activation is dependent on the presence of bound ligand and which synergizes with AF-1 (see above).
- A variable carboxy-terminal domain which is variable in sequence between various nuclear receptors.

As an example, the structure of RORα1 is shown in Figure 1A.

Depending on their mechanism of action and subcellular distribution in the absence of ligand, nuclear receptors (NRs) are classified into four classes.

Type I NRs are nuclear receptors located in the cytosol. Binding of a ligand to type I NRs results in dissociation of heat shock proteins, homo-dimerization, translocation to the nucleus and binding to HREs consisting of two half sites separated by variable length of DNA and the second half site having a sequence inverted from the first (inverted repeat). After formation of a nuclear receptor/DNA complex, other proteins are recruited which transcribe DNA downstream from the HRE into mRNA and, eventually, a protein which causes a change in cell function.

Type II NRs remain in the nucleus in the presence and absence of a ligand. They bind as heterodimers (usually with RXR) to DNA. In the absence of a ligand, type II NRs are often complexed with co-repressor proteins. Ligand binding to the nuclear receptor causes dissociation of co-repressors and recruitment of co-activator proteins and further proteins including RNA polymerase, which effects translation of DNA into mRNA.

Type III nuclear receptors are similar to type I NRs, but bind to direct repeat instead of inverted repeat HREs.

Type IV NRs bind either as monomers or dimers, but only a single DNA binding domain of the receptor binds to a half site HRE.

As detailed above, nuclear receptors activated upon ligand binding and bound to HREs recruit a significant number of other proteins which modify transcription of the associated target gene into mRNA. The function of these transcription co-regulators are varied and include chromatin remodeling in order to render the target gene more or less accessible to transcription, or a bridging function to stabilized the binding of other co-regulatory proteins. The co-regulatory protein (also referred to as co-factor) may be a co-activator, which often has an intrinsic histone acetyltransferase (HAT) activity which weakens the association of histones to DNA and, therefore, promotes transcription. In contrast thereto, co-repressors, which are preferably bound upon the binding of an agonist to NR, recruit histone deacetylases (HDACs), which promotes the association of histones to DNA and, therefore, represses transcription.

Members of the nuclear receptor superfamily include receptors such as those for glucocorticoids (GRs), androgens (ARs), mineralocorticoids (MRs), progestins (PRs), estrogens (ERs), thyroid hormones (TRs), vitamin D (VDRs), retinoids (RARs and RXRs), peroxisomes (XPARs and PPARs) and icosanoids (IRs).

Due to their role in development and homeostasis, nuclear receptors are an interesting target for studying their involvement in particular functions. Additionally, some of the nuclear receptors are so-called "orphan receptors", whose natural ligand is still unknown. Accordingly, it is of particular interest to identify these yet unknown natural ligands. Additionally, due to their involvement in physiological and pathophysiological functions of the body, nuclear receptors are an interesting target in pharmacological sciences. Data on functional interactions between nuclear receptors and co-regulators offer new chances in the development of novel pharmaceutical therapies for a wide range of diseases. Clinical strategies addressing the role of co-activators and co-repressors involved in cell proliferation with steroid receptors, may offer new treatments for, e.g. cancer. Furthermore, the functional importance of co-regulators and

signaling receptors involved in energy metabolism may offer new opportunities for diseases with impaired energy metabolism.

However, it was not possible to isolate proteins comprising a ligand binding domain of a nuclear receptor in a controllable form, particularly not for RORalpha.

Surprisingly, the inventor succeeded in providing an isolated protein comprising a ligand binding domain of a nuclear receptor in a controllable form. The protein could be prepared by culturing a cell comprising a nucleic acid coding for the protein under suitable conditions and isolating the protein from the cell culture. Thereafter, the isolated protein was contacted with a detergent, particularly lithium dodecyl sulphate (LDS), which restored controllability of the isolated protein.

Accordingly, a first aspect of the invention relates to an isolated protein comprising a ligand binding domain of a nuclear receptor in controllable form.

The ligand binding domain (see also above) of a nuclear receptor is that domain of the nuclear receptor which acts in response to ligand binding, which causes a conformational change in the nuclear receptor to induce a response, thereby acting as a molecular switch to turn on transcriptional activity. The ligand binding domain is a flexible unit, wherein the binding of a ligand stabilizes its conformation which in turn favors co-factor binding to modify receptor activity. The co-activator may bind to the activator function 2 (AF-2) at the same terminal end of the ligand binding domain. The binding of different ligands may alter the conformation of the ligand binding domain, which ultimately affects the DNA-binding specificity of the DNA binding domain of the nuclear receptor. The ligand binding domains of various nuclear receptors are well known in the art and are summarized, for example, at EMBL-EBI (www.ebi.ac.uk) or InterPro: IPR000536 (see http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?[interpro-AccNumber:IPR000536]+-e).

Examples of suitable ligand binding domains include:

- amino acids 271 to 523 of retinoic acid receptor-related orphan receptor alpha 1 (ROR alpha 1)
- amino acids 267 to 459 of retinoic acid receptor-related orphan receptor beta (ROR beta)

- amino acids 325 to 318 of retinoic acid receptor-related orphan receptor gamma (ROR gamma)
- amino acids 192 to 464 of hepatocyte nuclear factor alpha 1 (HNF4 alpha 1)
- amino acids 192 to 474 of hepatocyte nuclear factor alpha 2 (HNF4 alpha 2)
- amino acids 233 to 423 of estrogen-related receptor alpha (ERR alpha)
- amino acids 248 to 500 of estrogen-related receptor beta (ERR beta)
- amino acids 250 to 435 of estrogen-related receptor gamma (ERR gamma)

The nuclear receptor may be any known nuclear receptor. Depending on their sequence homologies nuclear receptors are divided into seven subfamilies.

Subfamily 1 includes thyroid hormone receptor-like, including thyroid hormone receptor-α and -β, retinoic acid receptor-α, -β and -γ, peroxisome proliferators-activated receptor-α, -β/δ, γ, Rev-ErbA-α and -β, RAR-related orphan receptors α, β and γ, liver X receptor-like α and β, farnesoid X receptor, vitamin D receptor, pregnane X receptor and constitutive androstane receptor.

Subfamily 2 relates to retinoic X receptor-like including, for example, hepatocyte nuclear receptor-4 (α and γ), retinoic X receptor (α, β and γ), testicular receptor (2 and 4), human homologue of the Drosophila tailless gene, photoreceptor cell-specific nuclear receptor, chicken ovalbumin upstream promoter-transcription factor (I and II) and V-erbA-related.

Subfamily 3 relates to estrogen receptor-like including, amongst others, estrogen receptor (α and β), estrogen related receptor (α, β and γ), corticoid receptor, mineralocorticoid receptor, progesterone receptor and androgen receptor.

Subfamily 4 relates to nerve growth factor IB-like including receptors such as nerve growth factor IB, nuclear receptor related 1 and neuron-derived orphan receptor 1.

Subfamily 5 relates to steroidogenic factor-like including, for example, steroidogenic factor 1 and liver receptor homolog-1.

Subfamily 6 relates to germ cell nuclear factor- like including germ cell nuclear factor.

A further subfamily, referred to as subfamily 0, includes miscellaneous receptors such as dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX1), small heterodimer partner and nuclear receptors with two DNA binding domains (2DBD-NR).

According to the present invention, the ligand binding domain of the nuclear receptor is comprised in an isolated protein. An isolated protein in the context of the present invention relates to a protein which is not in its natural environment. Accordingly, the "isolated protein" is not associated with proteins, it is normally found within nature or is isolated from a cell in which it normally occurs or is isolated from a cell in which the nucleic acid coding for the same has been expressed or is essentially free from other proteins from the same cellular source. The protein may be a naturally occurring protein, preferably a naturally occurring nuclear receptor or part thereof, wherein the part encompasses the ligand binding domain. However, the protein may also be artificial in that it does not naturally occur or in that it may encompass one or more sections which are naturally not connected to the ligand binding domain, for example, a fusion protein comprising or consisting of a ligand binding domain of a nuclear receptor and a further protein such as a second domain used for, e.g., purification or detection purposes.

Preferably, the term "isolated protein" means a protein molecule which is essentially separated from other cellular components of its natural environment. However, after isolation of the protein, cellular components may be added again, e.g., for measuring signal transduction pathways. Additionally, the skilled person will understand that the isolated protein is to be kept under suitable conditions allowing activity of the isolated protein, e.g., suitable buffers, pH values, ions, etc.

"Controllable form" in the context of the isolated protein of the invention comprising a ligand binding domain of a nuclear receptor relates to a protein, which is still amendable to activation upon agonist binding to the ligand binding domain. As detailed above, the LBD is activated upon binding of an agonistic ligand to the same, which alters gene expression of a target gene. However, up to now it was not possible to produce RORalpha protein or many other isolated proteins comprising an LBD of a nuclear receptor which could be controlled or regulated, i.e.

there was no significant or only little difference of activity in the presence or absence of an agonistic ligand for the respective LBD.

Accordingly, an isolated protein of the invention in controllable form can be detected by comparing activity in the presence or absence of an agonistic ligand for the respective LBD. Activity of the LBD may be determined in any suitable matter, e.g., by determining influence on the downstream elements of the respective signal transduction pathway, such as binding to any of the downstream components of the respective signal transduction pathway, such as co-regulator and/or target DNA. An example of such a task is described in the Example 2 and illustrated in Figure 1C.

Preferably, the activity of the isolated protein comprising an LBD of an NR in controllable form amounts to at least 1.2, more preferably at least 1.5, still more preferably at least 2, 3, 4 or 5, and most preferably at least 10, if the activity in the presence of an agonistic ligand is compared to that in the absence of an agonistic ligand for the respective LBD.
An isolated protein of the invention particularly relates to an isolated protein comprising a ligand binding domain of an NR in controllable form, wherein the protein is not constitutively active, which means that the protein is not active in the absence of an agonistic ligand for the respective LBD.

In one embodiment of the invention the isolated protein may comprise or consist of the full amino acid sequence of a naturally occurring nuclear receptor. Alternatively, the isolated protein may comprise or consist of a part of a naturally occurring nuclear receptor, provided that the LBD is still present in the part of the nuclear receptor.

The isolated protein may comprise or consist of any of the nuclear receptors as defined above. The nuclear receptor may be the isolated protein of it may be fused to a further domain, e.g., in order to ease purification of the protein or to detect the protein or to measure activity of the protein.

As detailed above, the isolated protein may also comprise or consist of a part of the nuclear receptor as long as the LBD of the nuclear receptor is part of the protein. Accordingly, the isolated protein may also comprise the amino terminal regulatory domain, the DNA binding

domain, a hinge region connecting the DNA binding domain and the ligand binding domain, and/or a carboxy-terminal domain of a nuclear receptor. The additional domains and regions may independent from each other, be derived from the same nuclear receptor as the LBD or from one or more other nuclear receptors.

In a preferred embodiment of the invention, the nuclear receptor is a retinoic acid receptor-related orphan receptor (ROR), particularly RORα, RORβ or RORγ, especially RORα.

The orphan receptors ROR, also referred to as RZR, constitute a subfamily of nuclear receptors for which initially no ligand had been identified. Presently, three subtypes of ROR receptors have been identified - RORα, RORβ and RORγ. ROR receptors bind in monomeric or dimeric form, each to a specific response element consisting of a sequence rich in A/T preceding a sequence of the PuGGTCA type and modulate transcription of the target genes.

Following alternative splicing, the RORα gene leads to four isoforms α1, α2, α3 and α4 RZRA, which differ in their N-terminal domain and show DNA recognition and distinct transactivation properties.

As for nuclear receptors, any mammalian ROR receptor is preferred, and human ROR receptors are even more preferred.

RORα (also referred to as RAR-related orphan receptor A, RZRA, ROR1, ROR2, ROR3, NR1F1) has been sequenced, and its sequence is available from the NCBI (National Center for Biotechnology Information) data bank under accession no. U04897, which provides the human mRNA and protein sequence. Known agonistic ligands for RORα include cholesterol, derivatives thereof and possibly melatonin.

RORβ (also referred to as RAR-related orphan receptor B, RZRB, NR1F2) has been sequenced and its sequence is available from the NCBI (National Center for Biotechnology Information) data bank under accession no. Y08639, which provides the human mRNA and protein sequence. A known agonistic ligand for RORβ is retinoic acid.

RORγ (also referred to as RAR-related orphan receptor C, RZRG, RORG, NR1F3, TOR) has been sequenced and its sequence is available from the NCBI (National Center for Biotechnology Information) data bank under accession no. U16997, which provides the human mRNA and protein sequence.

The three forms of ROR fulfill a number of critical roles including:

- RORα: development of the cerebellum, maintenance of bone, lymph node development, immune response, development of skeletal muscle, differentiation of smooth muscle cells, lipid metabolism (diseases: e.g. cerebellar degeneration, osteoporosis, ischemia-induced angiogenesis, artherosclerosis, inflammatory diseases)
- RORβ: central nervous system
- RORγ: immune response, skeletal muscle, adipocyte differentiation

Particularly preferred is an isolated protein comprising a ligand binding domain of RORα in a controllable form. The full length protein of RORα consists of 523 amino acids, wherein amino acids 271 - 523 code for the ligand binding domain. A particularly preferred protein is shown in SEQ ID NO. 1:

```
AELEHLAQNI SKSHLETCQY LREELQQITW QTFLQEEIEN YQNKQREVMW QLCAIKITEA      60
IQYVVEFAKR IDGFMELCQN DQIVLLKAGS LEVVFIRMCR AFDSQNNTVY FDGKYASPDV     120
FKSLGCEDFI SFVFEFGKSL CSMHLTEDEI ALFSAFVLMS ADRSWLQEKV KIEKLQQKIQ     180
LALQHVLQKN HREDGILTKL ICKVSTLRAL CGRHTEKLMA FKAIYPDIVR LHFPPLYKEL     240
FTSEFEPAMQ IDG
```

(SEQ ID NO:1)

An exemplary sequence comprising the above domain as well as a tag and a cleavage site is shown in the following: reads as follows:

```
MGSSHHHHHH LEVLFQGPAE LEHLAQNISK SHLETCQYLR EELQQITWQT FLQEEIENYQ      60
NKQREVMWQL CAIKITEAIQ YVVEFAKRID GFMELCQNDQ IVLLKAGSLE VVFIRMCRAF     120
DSQNNTVYFD GKYASPDVFK SLGCEDFISF VFEFGKSLCS MHLTEDEIAL FSAFVLMSAD     180
RSWLQEKVKI EKLQQKIQLA LQHVLQKNHR EDGILTKLIC KVSTLRALCG RHTEKLMAFK     240
AIYPDIVRLH FPPLYKELFT SEFEPAMQID G                                     271
```

(SEQ ID NO:2)

The isolated protein encompasses the domain of SEQ ID NO: 1 with a His-**tag (HHHHHH; SEQ ID NO: 3) and PreScission cleavage site (LEVLFQGP; SEQ ID NO: 4) inserted at amino acid 270 of RORα1. However, the His-tag may be substituted with another suitable** tag

e.g. as described herein as well as with another suitable cleavage site e.g. as described below. Examples of those are shown in Fig. 1B.

In one embodiment of the present invention, the isolated protein of the present invention comprises a marker, particularly a tag.

A marker in the context of the present invention may be any kind of molecule which can be easily detected. In the present invention, the molecule is bound to the isolated protein, therefore, the presence of the marker is indicative for the presence of the isolated protein. Markers (also referred to as labels) are known to a skilled person and include, for example, radiolabels (such as ^{3}H, ^{32}P, ^{35}S or ^{14}C), fluorescence markers (such as fluorescein, green fluorescence protein, or DyLight 488), enzymes (such as horse radish oxidase, β-lactamase, alkaline phosphatase or β-glucosidase) or an antigene detectable by a suitable antibody or antibody fragment.

Preferably, the marker is a tag. Tags are usually proteins which are used as biochemical indicators. They may be included into a protein, such as a recombinant, expressed protein and can serve several purposes. Preferably, they are used for purifying the proteins to which they are attached using standard conditions suitable for the particular tag. However, the tags may be also used as indicators in order to detect the presence of a particular protein.

A number of (affinity) tags are known at present. These are usually divided into 3 classes according to their size: small tags have a maximum of 12 amino acids, medium-sized ones have a maximum of 60 and large ones have more than 60. The small tags include the Arg-tag, the His-tag, the Strep-tag, the Flag-tag, the T7-tag, the V5-peptide-tag and the c-Myc-tag, the medium-sized ones include the S-tag, the HAT-tag, the calmodulin-binding peptide, the chitin-binding peptide and some cellulose-binding domains. The latter can contain up to 189 amino acids and are then regarded, like the GST- (glutathione-S-transferase-) and MBP-tag (maltose binding protein-tag), as large affinity tags.

In order to produce especially pure proteins, so-called double tags or tandem tags were developed. In this case the proteins are purified in two separate chromatography steps, in each case utilizing the affinity of a first and then of a second tag. Examples of such double or tandem

tags are the GST-His-tag (glutathione-S- transferase fused to a polyhistidine-tag), the 6xHis-Strep-tag (6 histidine residues fused to a Strep-tag), the 6xHis-tag100-tag (6 histidine residues fused to a 12-amino-acid protein of mammalian MAP-kinase 2), 8xHis-HA-tag (8 histidine residues fused to a haemagglutinin-epitope-tag), His-MBP (His-tag fused to a maltose-binding protein, FLAG-HA-tag (FLAG-tag fused to a hemagglutinin-epitope-tag), and the FLAG-Strep-tag.

Preferably, the isolated protein of the present invention comprises a tag selected from the group consisting of His-tag, Arg-tag, Strep-tag, Flag-tag, T7-tag, V5-peptide-tag, c-Myc-tag, S-tag, HAT-tag, calmodulin-binding peptide-tag, chitin-binding peptide-tag, GST-tag and MBP-tag. However, any other tag may be also used, but some tags such as His-tag, Arg-tag, Strep-tag, Flag-tag or GST-tag are preferred.

In an embodiment of the invention the isolated protein comprises a marker or tag, wherein the marker or tag is removable from the protein by proteolytic cleavage at a specific cleavage site, for example a cleavage site for an enzyme. This may be located between the LBD and the marker or tag. The cleavage site could for example be a protease cleavage site. Examples of proteases are chymotrypsin, trypsin, elastase, and plasmin; the corresponding cleavage sites are known to a person skilled in the art. Since the molecule to be purified is a protein, specific proteases, especially proteases from viruses that normally attack plants, are preferred. Examples of suitable specific proteases are thrombin, factor Xa, Igase, TEV-protease from the "Tobacco Etch Virus", the protease PreScission (Human Rhinovirus 3C Protease), enterokinase or Kex2. TEV-protease and PreScission are especially preferred.

An example of a protein comprising an LBD, a His-tag and a precision cleaving site is disclosed in SEQ ID NO. 2. A suitable nucleic acid and a vector encoding that protein are given in SEQ ID NO: 5 and SEQ ID NO: 6, respectively. Additionally, exemplary isolated proteins of the invention are illustrated in Figure 1B.

Nucleotides 4021 to 5040 of the vector of SEQ ID NO: 6:

- Upper nucleic acid sequence: coding strand (SEQ ID NO: 5)
- Lower nucleic acid sequence: template strand

- Amino acid sequence: LBD (as defined in SEQ ID NO: 1) with His-tag and PreScission cleavage site (SEQ ID NO: 2)
- [ ]: cloning sites (as specified)
- *ATG* and *TAA*: start/stop (each adjacent to cloning site)
- **CATCATCATCATCATCTGGAAGTTCTGTTCCAGGGGCCC:** His-tag and PreScission cleavage site

```
          GATAACCATCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAA
     4021 ---------+---------+---------+---------+---------+---------+ 4080
          CTATTGGTAGAGCGTTTATTTATTCATAAAATGACAAAAGCATTGTCAAAACATTATTTT

                                                          BamHI
                                                            |
          AAACCTATAAATATTCCGGATTATTCATACCGTCCCACCATCGGGCGCGGATCCATGGGA
     4081 ---------+---------+---------+---------+---------+---------+ 4140
          TTTGGATATTTATAAGGCCTAATAAGTATGGCAGGGTGGTAGCCCGCGCCTAGGTACCCT

a                                                               M  G  -

          AGTAGCCATCATCATCATCATCATCTGGAAGTTCTGTTCCAGGGGCCCGCAGAATTAGAA
     4141 ---------+---------+---------+---------+---------+---------+ 4200
          TCATCGGTAGTAGTAGTAGTAGTAGACCTTCAAGACAAGGTCCCCGGGCGTCTTAATCTT

a         S  S  H  H  H  H  H  H  L  E  V  L  F  Q  G  P  A  E  L  E  -

          CACCTTGCACAGAATATATCTAAATCGCATCTGGAAACCTGCCAATACTTGAGAGAAGAG
     4201 ---------+---------+---------+---------+---------+---------+ 4260
          GTGGAACGTGTCTTATATAGATTTAGCGTAGACCTTTGGACGGTTATGAACTCTCTTCTC

a         H  L  A  Q  N  I  S  K  S  H  L  E  T  C  Q  Y  L  R  E  E  -

          CTCCAGCAGATAACGTGGCAGACCTTTTTACAGGAAGAAATTGAGAACTATCAAAACAAG
     4261 ---------+---------+---------+---------+---------+---------+ 4320
          GAGGTCGTCTATTGCACCGTCTGGAAAAATGTCCTTCTTTAACTCTTGATAGTTTTGTTC

a         L  Q  Q  I  T  W  Q  T  F  L  Q  E  E  I  E  N  Y  Q  N  K  -

          CAGCGGGAGGTGATGTGGCAATTGTGTGCCATCAAAATTACAGAAGCTATACAGTATGTG
     4321 ---------+---------+---------+---------+---------+---------+ 4380
          GTCGCCCTCCACTACACCGTTAACACACGGTAGTTTTAATGTCTTCGATATGTCATACAC

a         Q  R  E  V  M  W  Q  L  C  A  I  K  I  T  E  A  I  Q  Y  V  -

          GTGGAGTTTGCCAAACGCATTGATGGATTTATGGAACTGTGTCAAAATGATCAAATTGTG
     4381 ---------+---------+---------+---------+---------+---------+ 4440
          CACCTCAAACGGTTTGCGTAACTACCTAAATACCTTGACACAGTTTTACTAGTTTAACAC

a         V  E  F  A  K  R  I  D  G  F  M  E  L  C  Q  N  D  Q  I  V  -

          CTTCTAAAAGCAGGTTCTCTAGAGGTGGTGTTTATCAGAATGTGCCGTGCCTTTGACTCT
     4441 ---------+---------+---------+---------+---------+---------+ 4500
          GAAGATTTTCGTCCAAGAGATCTCCACCACAAATAGTCTTACACGGCACGGAAACTGAGA

a         L  L  K  A  G  S  L  E  V  V  F  I  R  M  C  R  A  F  D  S  -

          CAGAACAACACCGTGTACTTTGATGGGAAGTATGCCAGCCCCGACGTCTTCAAATCCTTA
     4501 ---------+---------+---------+---------+---------+---------+ 4560
          GTCTTGTTGTGGCACATGAAACTACCCTTCATACGGTCGGGGCTGCAGAAGTTTAGGAAT

a         Q  N  N  T  V  Y  F  D  G  K  Y  A  S  P  D  V  F  K  S  L  -
```

```
GGTTGTGAAGACTTTATTAGCTTTGTGTTTGAATTTGGAAAGAGTTTATGTTCTATGCAC
    4561 ---------+---------+---------+---------+---------+---------+ 4620
         CCAACACTTCTGAAATAATCGAAACACAAACTTAAACCTTTCTCAAATACAAGATACGTG

a        G  C  E  D  F  I  S  F  V  F  E  F  G  K  S  L  C  S  M  H  -

         CTGACTGAAGATGAAATTGCATTATTTTCTGCATTTGTACTGATGTCAGCAGATCGCTCA
    4621 ---------+---------+---------+---------+---------+---------+ 4680
         GACTGACTTCTACTTTAACGTAATAAAAGACGTAAACATGACTACAGTCGTCTAGCGAGT

a        L  T  E  D  E  I  A  L  F  S  A  F  V  L  M  S  A  D  R  S  -

         TGGCTGCAAGAAAAGGTAAAAATTGAAAAACTGCAACAGAAAATTCAGCTAGCTCTTCAA
    4681 ---------+---------+---------+---------+---------+---------+ 4740
         ACCGACGTTCTTTTCCATTTTTAACTTTTTGACGTTGTCTTTTAAGTCGATCGAGAAGTT

a        W  L  Q  E  K  V  K  I  E  K  L  Q  Q  K  I  Q  L  A  L  Q  -

         CACGTCCTACAGAAGAATCACCGAGAAGATGGAATACTAACAAAGTTAATATGCAAGGTG
    4741 ---------+---------+---------+---------+---------+---------+ 4800
         GTGCAGGATGTCTTCTTAGTGGCTCTTCTACCTTATGATTGTTTCAATTATACGTTCCAC

a        H  V  L  Q  K  N  H  R  E  D  G  I  L  T  K  L  I  C  K  V  -

         TCTACATTAAGAGCCTTATGTGGACGACATACAGAAAAGCTAATGGCATTTAAAGCAATA
    4801 ---------+---------+---------+---------+---------+---------+ 4860
         AGATGTAATTCTCGGAATACACCTGCTGTATGTCTTTTCGATTACCGTAAATTTCGTTAT

a        S  T  L  R  A  L  C  G  R  H  T  E  K  L  M  A  F  K  A  I  -

         TACCCAGACATTGTGCGACTTCATTTTCCTCCATTATACAAGGAGTTGTTCACTTCAGAA
    4861 ---------+---------+---------+---------+---------+---------+ 4920
         ATGGGTCTGTAACACGCTGAAGTAAAAGGAGGTAATATGTTCCTCAACAAGTGAAGTCTT

a        Y  P  D  I  V  R  L  H  F  P  P  L  Y  K  E  L  F  T  S  E  -

                                    EcoRI
                                      |
         TTTGAGCCAGCAATGCAAATTGATGGGTAAGAATTCCGGAGCGGCCGCTGCAGATCTGAT
    4921 ---------+---------+---------+---------+---------+---------+ 4980
         AAACTCGGTCGTTACGTTTAACTACCCATTCTTAAGGCCTCGCCGGCGACGTCTAGACTA

a        F  E  P  A  M  Q  I  D  G  *                                 -

         CCTTTCCTGGGACCCGGCAAGAACCAAAAACTCACTCTCTTCAAGGAAATCCGTAATGTT
    4981 ---------+---------+---------+---------+---------+---------+ 5040
         GGAAAGGACCCTGGGCCGTTCTTGGTTTTTGAGTGAGAGAAGTTCCTTTAGGCATTACAA
```

Vector (for details see above) (SEQ ID NO: 6):

_____ : vector insert

```
AAGCTTTACTCGTAAAGCGAGTTGAAGGATCATATTTAGTTGCGTTTATGAGATAAGATTGAAAGCACGTGTAAAA
TGTTTCCCGCGCGTTGGCACAACTATTTACAATGCGGCCAAGTTATAAAAGATTCTAATCTGATATGTTTTAAAAC
ACCTTTGCGGCCCGAGTTGTTTGCGTACGTGACTAGCGAAGAAGATGTGTGGACCGCAGAACAGATAGTAAAACAA
AACCCTAGTATTGGAGCAATAATCGATTTAACCAACACGTCTAAATATTATGATGGTGTGCATTTTTTGCGGGCGG
GCCTGTTATACAAAAAAATTCAAGTACCTGGCCAGACTTTGCCGCCTGAAAGCATAGTTCAAGAATTTATTGACAC
GGTAAAAGAATTTACAGAAAAGTGTCCCGGCATGTTGGTGGGCGTGCACTGCACACACGGTATTAATCGCACCGGT
TACATGGTGTGCAGATATTTAATGCACACCCTGGGTATTGCGCCGCAGGAAGCCATAGATAGATTCGAAAAAGCCA
GAGGTCACAAAATTGAAAGACAAAATTACGTTCAAGATTTATTAATTTAATTAATATTATTTGCATTCTTTAACAA
ATACTTTATCCTATTTTCAAATTGTTGCGCTTCTTCCAGCGAACCAAAACTATGCTTCGCTTGCTCCGTTTAGCTT
```

GTAGCCGATCAGTGGCGTTGTTCCAATCGACGGTAGGATTAGGCCGGATATTCTCCACCACAATGTTGGCAACGTT
GATGTTACGTTTATGCTTTTGGTTTTCCACGTACGTCTTTTGGCCGGTAATAGCCGTAAACGTAGTGCCGTCGCGC
GTCACGCACAACACCGGATGTTTGCGCTTGTCCGCGGGGTATTGAACCGCGCGATCCGACAAATCCACCACTTTGG
CAACTAAATCGGTGACCTGCGCGTCTTTTTTCTGCATTATTTCGTCTTTCTTTTGCATGGTTTCCTGGAAGCCGGT
GTACATGCGGTTTAGATCAGTCATGACGCGCGTGACCTGCAAATCTTTGGCCTCGATCTGCTTGTCCTTGATGGCA
ACGATGCGTTCAATAAACTCTTGTTTTTTTAACAAGTTCCTCGGTTTTTTGCGCCACCACCGCTTGCAGCGCGTTTG
TGTGCTCGGTGAATGTCGCAATCAGCTTAGTCACCAACTGTTTGCTCTCCTCCTCCCGTTGTTTGATCGCGGGATC
GTACTTGCCGGTGCAGAGCACTTGAGGAATTACTTCTTCTAAAAGCCATTCTTGTAATTCTATGGCGTAAGGCAAT
TTGGACTTCATAATCAGCTGAATCACGCCGGATTTAGTAATGAGCACTGTATGCGGCTGCAAATACAGCGGGTCGC
CCCTTTTCACGACGCTGTTAGAGGTAGGGCCCCCATTTTGGATGGTCTGCTCAAATAACGATTTGTATTTATTGTC
TACATGAACACGTATAGCTTTATCACAAACTGTATATTTTAAACTGTTAGCGACGTCCTTGGCCACGAACCGGACC
TGTTGGTCGCGCTCTAGCACGTACCGCAGGTTGAACGTATCTTCTCCAAATTTAAATTCTCCAATTTTAACGCGAG
CCATTTTGATACACGTGTGTCGATTTTGCAACAACTATTGTTTTTTAACGCAAACTAAACTTATTGTGGTAAGCAA
TAATTAAATATGGGGGAACATGCGCCGCTACAACACTCGTCGTTATGAACGCAGACGGCGCCGGTCTCGGCGCAAG
CGGCTAAAACGTGTTGCGCGTTCAACGCGGCAAACATCGCAAAAGCCAATAGTACAGTTTTGATTTGCATATTAAC
GGCGATTTTTTAAATTATCTTATTTAATAAATAGTTATGACGCCTACAACTCCCCGCCCGCGTTGACTCGCTGCAC
CTCGAGCAGTTCGTTGACGCCTTCCTCCGTGTGGCCGAACACGTCGAGCGGGTGGTCGATGACCAGCGGCGTGCCG
CACGCGACGCACAAGTATCTGTACACCGAATGATCGTCGGGCGAAGGCACGTCGGCCTCCAAGTGGCAATATTGGC
AAATTCGAAAATATATACAGTTGGGTTGTTTGCGCATATCTATCGTGGCGTTGGGCATGTACGTCCGAACGTTGAT
TTGCATGCAAGCCGAAATTAAATCATTGCGATTAGTGCGATTAAAACGTTGTACATCCTCGCTTTTAATCATGCCG
TCGATTAAATCGCGCAATCGAGTCAAGTGATCAAAGTGTGGAATAATGTTTTCTTTGTATTCCCGAGTCAAGCGCA
GCGCGTATTTTAACAAACTAGCCATCTTGTAAGTTAGTTTCATTTAATGCAACTTTATCCAATAATATATTATGTA
TCGCACGTCAAGAATTAACAATGCGCCCGTTGTCGCATCTCAACACGACTATGATAGAGATCAAATAAAGCGCGAA
TTAAATAGCTTGCGACGCAACGTGCACGATCTGTGCACGCGTTCCGGCACGAGCTTTGATTGTAATAAGTTTTTAC
GAAGCGATGACATGACCCCCGTAGTGACAACGATCACGCCCAAAAGAACTGCCGACTACAAAATTACCGAGTATGT
CGGTGACGTTAAAACTATTAAGCCATCCAATCGACCGTTAGTCGAATCAGGACCGCTGGTGCGAGAAGCCGCGAAG
TATGGCGAATGCATCGTATAACGTGTGGAGTCCGCTCATTAGAGCGTCATGTTTAGACAAGAAAGCTACATATTTA
ATTGATCCCGATGATTTTATTGATAAATTGACCCTAACTCCATACACGGTATTCTACAATGGCGGGGTTTTGGTCA
AAATTTCCGGACTGCGATTGTACATGCTGTTAACGGCTCCGCCCACTATTAATGAAATTAAAAATTCCAATTTTAA
AAAACGCAGCAAGAGAAACATTTGTATGAAAGAATGCGTAGAAGGAAAGAAAAATGTCGTCGACATGCTGAACAAC
AAGATTAATATGCCTCCGTGTATAAAAAAAATATTGAACGATTTGAAAGAAAACAATGTACCGCGCGGCGGTATGT
ACAGGAAGAGGTTTATACTAAACTGTTACATTGCAAACGTGGTTTCGTGTGCCAAGTGTGAAAACCGATGTTTAAT
CAAGGCTCTGACGCATTTCTACAACCACGACTCCAAGTGTGTGGGTGAAGTCATGCATCTTTTAATCAAATCCCAA
GATGTGTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTCTGTCCGTTTGCTGGCAACTGCAAGG
GTCTCAATCCTATTTGTAATTATTGAATAATAAAACAATTATAAATGCTAAATTTGTTTTTTATTAACGATACAAA
CCAAACGCAACAAGAACATTTGTAGTATTATCTATAATTGAAAACGCGTAGTTATAATCGCTGAGGTAATATTTAA
AATCATTTTCAAATGATTCACAGTTAATTTGCGACAATATAATTTTATTTTCACATAAACTAGACGCCTTGTCGTC
TTCTTCTTCGTATTCCTTCTCTTTTTCATTTTTCTCCTCATAAAAATTAACATAGTTATTATCGTATCCATATATG
TATCTATCGTATAGAGTAAATTTTTTGTTGTCATAAATATATATGTCTTTTTTAATGGGGTGTATAGTACCGCTGC
GCATAGTTTTTCTGTAATTTACAACAGTGCTATTTTCTGGTAGTTCTTCGGAGTGTGTTGCTTTAATTATTAAATT
TATATAATCAATGAATTTGGGATCGTCGGTTTTGTACAATATGTTGCCGGCATAGTACGCAGCTTCTTCTAGTTCA
ATTACACCATTTTTTAGCAGCACCGGATTAACATAACTTTCCAAAATGTTGTACGAACCGTTAAACAAAAACAGTT
CACCTCCCTTTTCTATACTATTGTCTGCGAGCAGTTGTTTGTTGTTAAAAATAACAGCCATTGTAATGAGACGCAC
AAACTAATATCACAAACTGGAAATGTCTATCAATATATAGTTGCTGATATCATGGAGATAATTAAAATGATAACCA
TCTCGCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTAT
TCATACCGTCCCACCATCGGGCGCGGATCC*ATG*GGAAGTAGC**CATCATCATCATCATCATCTGGAAGTTCTGTTCC**
AGGGGCCCGCAGAATTAGAACACCTTGCACAGAATATATCTAAATCGCATCTGGAAACCTGCCAATACTTGAGAGA
AGAGCTCCAGCAGATAACGTGGCAGACCTTTTTACAGGAAGAAATTGAGAACTATCAAAACAAGCAGCGGGAGGTG
ATGTGGCAATTGTGTGCCATCAAAATTACAGAAGCTATACAGTATGTGGTGGAGTTTGCCAAACGCATTGATGGAT
TTATGGAACTGTGTCAAAATGATCAAATTGTGCTTCTAAAAGCAGGTTCTCTAGAGGTGGTGTTTATCAGAATGTG
CCGTGCCTTTGACTCTCAGAACAACACCGTGTACTTTGATGGGAAGTATGCCAGCCCCGACGTCTTCAAATCCTTA
GGTTGTGAAGACTTTATTAGCTTTGTGTTTGAATTTGGAAAGAGTTTATGTTCTATGCACCTGACTGAAGATGAAA
TTGCATTATTTTCTGCATTTGTACTGATGTCAGCAGATCGCTCATGGCTGCAAGAAAAGGTAAAAATTGAAAAACT
GCAACAGAAAATTCAGCTAGCTCTTCAACACGTCCTACAGAAGAATCACCGAGAAGATGGAATACTAACAAAGTTA
ATATGCAAGGTGTCTACATTAAGAGCCTTATGTGGACGACATACAGAAAAGCTAATGGCATTTAAAGCAATATACC
CAGACATTGTGCGACTTCATTTTCCTCCATTATACAAGGAGTTGTTCACTTCAGAATTTGAGCCAGCAATGCAAAT
TGATGGG*TAA*GAATTCCGGAGCGGCCGCTGCAGATCTGATCCTTTCCTGGGACCCGGCAAGAACCAAAAACTCACT
CTCTTCAAGGAAATCCGTAATGTTAAACCCGACACGATGAAGCTTGTCGTTGGATGGAAAGGAAAAGAGTTCTACA
GGGAAACTTGGACCCGCTTCATGGAAGACAGCTTCCCCATTGTTAACGACCAAGAAGTGATGGATGTTTTCCTTGT

TGTCAACATGCGTCCCACTAGACCCAACCGTTGTTACAAATTCCTGGCCCAACACGCTCTGCGTTGCGACCCCGAC
TATGTACCTCATGACGTGATTAGGATCGTCGAGCCTTCATGGGTGGGCAGCAACAACGAGTACCGCATCAGCCTGG
CTAAGAAGGGCGGCGGCTGCCCAATAATGAACCTTCACTCTGAGTACACCAACTCGTTCGAACAGTTCATCGATCG
TGTCATCTGGGAGAACTTCTACAAGCCCATCGTTTACATCGGTACCGACTCTGCTGAAGAGGAGGAAATTCTCCTT
GAAGTTTCCCTGGTGTTCAAAGTAAAGGAGTTTGCACCAGACGCACCTCTGTTCACTGGTCCGGCGTATTAAAACA
CGATACATTGTTATTAGTACATTTATTAAGCGCTAGATTCTGTGCGTTGTTGATTTACAGACAATTGTTGTACGTA
TTTTAATAATTCATTAAATTTATAATCTTTAGGGTGGTATGTTAGAGCGAAAATCAAATGATTTTCAGCGTCTTTA
TATCTGAATTTAAATATTAAATCCTCAATAGATTTGTAAAATAGGTTTCGATTAGTTTCAAACAAGGGTTGTTTTT
CCGAACCGATGGCTGGACTATCTAATGGATTTTCGCTCAACGCCACAAAACTTGCCAAATCTTGTAGCAGCAATCT
AGCTTTGTCGATATTCGTTTGTGTTTTGTTTTGTAATAAAGGTTCGACGTCGTTCAAAATATTATGCGCTTTTGTA
TTTCTTTCATCACTGTCGTTAGTGTACAATTGACTCGACGTAAACACGTTAAATAAAGCTTGGACATATTTAACAT
CGGGCGTGTTAGCTTTATTAGGCCGATTATCGTCGTCGTCCCAACCCTCGTCGTTAGAAGTTGCTTCCGAAGACGA
TTTTGCCATAGCCACACGACGCCTATTAATTGTGTCGGCTAACACGTCCGCGATCAAATTTGTAGTTGAGCTTTTT
GGAATTATTTCTGATTGCGGGCGTTTTTGGGCGGGTTTCAATCTAACTGTGCCCGATTTTAATTCAGACAACACGT
TAGAAAGCGATGGTGCAGGCGGTGGTAACATTTCAGACGGCAAATCTACTAATGGCGGCGGTGGTGGAGCTGATGA
TAAATCTACCATCGGTGGAGGCGCAGGCGGGGCTGGCGGCGGAGGCGGAGGCGGAGGTGGTGGCGGTGATGCAGAC
GGCGGTTTAGGCTCAAATGTCTCTTTAGGCAACACAGTCGGCACCTCAACTATTGTACTGGTTTCGGGCGCCGTTT
TTGGTTTGACCGGTCTGAGACGAGTGCGATTTTTTTCGTTTCTAATAGCTTCCAACAATTGTTGTCTGTCGTCTAA
AGGTGCAGCGGGTTGAGGTTCCGTCGGCATTGGTGGAGCGGGCGGCAATTCAGACATCGATGGTGGTGGTGGTGGT
GGAGGCGCTGGAATGTTAGGCACGGGAGAAGGTGGTGGCGGCGGTGCCGCCGGTATAATTTGTTCTGGTTTAGTTT
GTTCGCGCACGATTGTGGGCACCGGCGCAGGCGCCGCTGGCTGCACAACGGAAGGTCGTCTGCTTCGAGGCAGCGC
TTGGGGTGGTGGCAATTCAATATTATAATTGGAATACAAATCGTAAAAATCTGCTATAAGCATTGTAATTTCGCTA
TCGTTTACCGTGCCGATATTTAACAACCGCTCAATGTAAGCAATTGTATTGTAAAGAGATTGTCTCAAGCTCGCCG
CACGCCGATAACAAGCCTTTTCATTTTTACTACAGCATTGTAGTGGCGAGACACTTCGCTGTCGTCGACGTACATG
TATGCTTTGTTGTCAAAAACGTCGTTGGCAAGCTTTAAAATATTTAAAAGAACATCTCTGTTCAGCACCACTGTGT
TGTCGTAAATGTTGTTTTTGATAATTTGCGCTTCCGCAGTATCGACACGTTCAAAAAATTGATGCGCATCAATTTT
GTTGTTCCTATTATTGAATAAATAAGATTGTACAGATTCATATCTACGATTCGTCATGGCCACCACAAATGCTACG
CTGCAAACGCTGGTACAATTTTACGAAAACTGCAAAAACGTCAAAACTCGGTATAAAATAATCAACGGGCGCTTTG
GCAAAATATCTATTTTATCGCACAAGCCCACTAGCAAATTGTATTTGCAGAAAACAATTTCGGCGCACAATTTTAA
CGCTGACGAAATAAAAGTTCACCAGTTAATGAGCGACCACCCAAATTTTATAAAAATCTATTTTAATCACGGTTCC
ATCAACAACCAAGTGATCGTGATGGACTACATTGACTGTCCCGATTTATTTGAAACACTACAAATTAAAGGCGAGC
TTTCGTACCAACTTGTTAGCAATATTATTAGACAGCTGTGTGAAGCGCTCAACGATTTGCACAAGCACAATTTCAT
ACACAACGACATAAAACTCGAAAATGTCTTATATTTCGAAGCACTTGATCGCGTGTATGTTTGCGATTACGGATTG
TGCAAACACGAAAACTCACTTAGCGTGCACGACGGCACGTTGGAGTATTTTAGTCCGGAAAAAATTCGACACACAA
CTATGCACGTTTCGTTTGACTGGTACGCGGCGTGTTAACATACAAGTTGCTAACGTAATCATGGTCATAGCTGTTT
CCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTG
CCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACT
GACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGC
TTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAG
TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
ACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAA
AAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG
TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCT
GACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGA
CCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCA
ACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG
CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGT

TCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACT
GATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTAT
TGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA
AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTT
TCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
TAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACT
ATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAA
ATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTA
TTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGAC
GTTGTAAAACGACGGCCAGTGCC

A further aspect of the present invention relates to a method of producing the isolated protein of the present invention, comprising the steps of:

a) culturing a cell comprising a nucleic acid coding for the protein of the invention under conditions conducive to the production of the protein,

b) isolating the protein from the cell culture, and

c) contacting the isolated protein of step b) with a detergent, particularly with an alkaline salt of a saturated unbranched C6-C20 alkyl sulphate or carbonate or with an isoprenyl salt.

In general, steps a) and b) of the method of the invention are well known to the skilled person. However, in the following, they will be briefly summarized and illustrated by examples. The skilled person will understand that the described method may be modified depending on the respective protein of the invention to be isolated, the cell used for cell culture, etc.

For the production of recombinant protein, i.e. for synthesis of an exogene gene product in a living cell, a multiplicity of expression systems is available. These include series of well-known organisms and cell lines (bacteria, insect cells, yeasts, mammalian cells, etc.) as well as various expression vectors with different promoters, selection markers and optionally fusion partners. The production of a recombinant protein usually includes the introduction of the coding gene into a plasmid or any other suitable vector. In general, the selection of a suitable vector or plasmid depends on the intended host cell. This vector is then introduced into the chosen target cell (transformation or transfection) and the target cell is cultivated. Depending on the promoter, the expression of gene may occur throughout a period of cultivation or may be induced by an external signal.

As detailed above, it may be necessary to produce a suitable nucleic acid coding for the protein of the invention or a vector containing the same.

The nucleic acid coding for a protein of the invention may be a naturally occurring gene coding for a nuclear receptor or it may be a part thereof coding for a part of the nuclear receptor as defined above. The isolation of nucleic acids coding for naturally occurring proteins or a part thereof is well known to the skilled person and may include the isolation using suitable probes and separation methods, or the nucleic acid may be derived from a commercial supplier. If a nucleic acid is used which is not comprised by a naturally occurring gene (e.g., nucleic acid coding for a fusion protein comprising a full or partial NR and a tag) the respective nucleic acid coding for the fusion protein may be produced in accordance with standard procedures including well-known methods of genetic engineering. Usually, suitable restriction endonucleases are used to cut DNA at specific sites. The fragments formed by restriction enzymes may be joined together with ligase. Thereafter, the DNA may be introduced into a vector suitable for transferring genetic material into a cell. The vector may be a viral vector or plasmid vector. Suitable vectors include adenovirus, adeno-associated virus, cytomegalovirus, etc. Examples of comerially available vectors are pBR322, the pUC series, pBluescript, pTZ, pSP and pGEM. Alternatively, also naked DNA may be introduced into a cell. If the nucleic acid, i.e. DNA or RNA, is used without a vector, transfection may be carried out by mixing it with a cationic lipid to produce liposomes, which fuse with the cell membrane and deposit the nucleic acid inside the cell. The transfection may be also carried out by calcium phosphate, wherein, e.g., HEPES-buffered saline solution containing phosphate ions is combined with calcium chloride solution containing the DNA to be transfected. When the two solutions are combined, a fine precipitate of calcium phosphate is formed, binding the DNA to be transfected on its surface. The suspension of the precipitate is then added to the cell to be transfected (usually a cell culture grown in a monolayer). Other methods for transfection include electroporation, heat shock, magnetofection, nucleofection and the use of transfection agents such as Lipofectamine, Fugene, etc. A further approach is the „gene gun", where the DNA is coupled to a nanoparticle of an inert solid (usually gold) which is then shot directly into the target cells.

The transfection of a cell may be transient or stable. In case the nucleic acid introduced into a cell during a transfection process, is not inserted into the nuclear genome, the foreign nucleic acid may be lost at a later stage, when the cell undergoes mitosis. This is called transient transfection. More preferably, the transfection is a stable transfection, wherein the nucleic acid remains in the genome of the cell. In order to accomplish stable transfection, usually selection

techniques are used, wherein the nucleic acid is co-transfected with another gene, providing for selective selection. The additional gene may confer resistance towards a certain condition or substance, e.g. an antibiotic or metabolic deficiency. Examples of suitable genes include neomycin resistant gene, hygromycin phosphotransferase gene, etc.

After introduction of the nucleic acid coding the protein of the invention into the host, the cell is grown under suitable conditions. A series of different host cells for the production of proteins is known to the skilled person including bacteria, insect cells, yeasts and mammalian cells. Examples of such cells are Sf9, Sf21, HEK 293 cells, CHO cells, HeLa cells, CaCo cells or NIH 3T3 cells.

The host cell may be either a primary cell or it may be a cell line, wherein cell lines are preferred.

The cell comprising the nucleic acid of the invention is grown and maintained under conditions conducive to the production of the cell. This includes an appropriate temperature and gas mixture (typically 37 °C, 5 % CO_2), optionally in a cell incubator. Culture conditions may vary widely for each cell type and are known to the skilled practitioner. The expression may take place for example in insect cells after transformation with suitable baculovirus vector systems. In such a case the temperature is kept at 26 °C wheras control of CO_2 is not required.

Aside from temperature and gas mixture, the most commonly varied factor in cell culture systems is the growth medium. Recipes for growth media can vary in pH, glucose concentration, growth factor and the presence of other nutrient components among others. Growth factors used for supplement media are often derived from animal blood such as calf serum.

A skilled person in the art knows how to derive a nucleic acid sequence, which may be DNA or RNA, from a protein sequence, taking into account the genetic code. He also knows how to produce such a nucleic acid sequence using standard techniques of molecular biology. This can be accomplished, for example, by the use and combination of existing sequences using restriction enzymes. The nucleic acid suitably also contains further elements, e.g., a promoter and a transcription start and stop signal and a translation start and stop signal.

After step a), the protein is isolated from the cell culture by any suitable separation or purification method known to the skilled person. If a sufficient amount of the target protein has been secreted into the medium, the isolation can continue with the same. Otherwise, it may be necessary to disrupt the cells. This can be effected, for example, by lysis of the cell, e.g., by means of ultrasound or hypertonic medium or by shearing. To remove insoluble components, the sample can, for example, be centrifuged, especially at 10,000 x g to 15,000 x g and the supernatant obtained can be used for step c) or may be further purified or concentrated.

The isolation of step b) may alternatively or additionally include well-known purification concentration steps such as extraction, precipitation, electrophoretic methods, chromatographic methods, etc. Examples of those include cell electrophoresis, ion exchange chromatography, size exclusion chromatography, SDS-PAGE chromatography, or affinity chromatography particularly immobilized metal ion affinity chromatography (IMAC).

Affinity purification is particularly envisioned, if the isolated protein of the invention comprises a suitable marker or tag, as defined above. Affinity purification is a special form of adsorption purification, in which there are, on a carrier, groups (binding partners) with high affinity and therefore high binding strength to one of the two domains, so that these can be adsorbed preferentially and thus separated from other substances. Purification can be carried out using a first and a second tag (e.g His-tag and GST-Tag). Purification takes place by specific binding to a suitable binding partner. The binding partner is preferably bound to a solid phase. The solid phase can be usual carrier materials, for example Sepharose, Superflow, Macroprep, POROS 20 or POROS 50. Separation is then carried out for example chromatographically, e.g. by gravity, HPLC or FPLC. The protein of interest may be eluted from the solid phase by altering the conditions, so that the changed conditions no longer permit binding between affinity marker or tag and binding partner (e.g. alteration of the pH value or the ionic strength), or by separating the molecule from the domain bound to the binding partner. Separation can be effected by cleavage of the bond between molecule and binding partner, e.g. by chemical means or using specific enzymes, as was described in detail above. Alternatively, it is also possible to use specific competitors, which are added in excess. Alternatively, the binding partner can also be bound to beads, especially magnetic beads. After adding the beads to the sample, binding takes place between the particular domain and the corresponding binding

partner. The suspension can then be centrifuged for example, so that the labeled molecule sediments with the bead, and other components remain in the supernatant, from where they can be removed. Alternatively, the suspension is separated utilizing the magnetic properties of the beads. In one embodiment, the suspension is applied to a column, which is located in a magnetic field. As the magnetic beads and the molecule bound to them are retained in the magnetic field, other constituents of the sample can be washed out in several washing operations. The molecule of interest can then for example be washed from the beads using a suitable elution buffer, or can be separated from the beads by enzymatic cleavage e.g. at the cleavage site between the LBD and the tag or marker.

After step b), the isolated protein is contacted with a detergent, in order to provide an LBD in a controllable form. Detergents in the biological sense are membrane-active substances commonly used to disrupt the bipolar lipid membrane of cells in order to free and solubilize membrane-bound proteins. The value of the detergents is derived from their amphophilic nature. Each detergent molecule is characterized by a hydrophilic "head" region and a hydrophobic "tail" region. The result of this characteristic is the formation of a thermodynamically stable micelle with hydrophobic course in aqueous media. The hydrophobic core provides an environment that allows for the dissolution of hydrophobic molecules or domains of proteins. The detergent can be anionic, cationic, zwitterionic or non-ionic. Anionic and cationic detergents typically modify protein structure to a greater extent than the other two classes. The degree of modification varies with the individual protein and the particular detergent. Ionic detergents are also more sensitive to pH, ionic strength and the nature of the counter ion and can interfere with charge based analytical methods. Alternatively, most non-ionic detergents are non-denaturating, but are less effective at disrupting protein aggregation. Zwitterionic detergents uniquely offer some intermediate class properties that are superior to the other three detergent types in some applications offering the low denaturating and net zero charge characteristics of non-ionic detergents. Zwitterionics also efficiently disrupt protein aggregations.

Preferably, the detergent is an alkaline salt of a saturated unbranched C6-C20 alkyl sulphate or carbonate or is an isoprenyl salt. From the alkaline salts, lithium, sodium and potassium salt, especially lithium salts, are preferred.

The saturated unbranched C6-C20 alkyl sulphate may be an n-hexyl, n-heptyl, n-octyl, n-nonyl, n-decyl, n-undecyl, n-dodecyl, n-tridecyl, n-tetradecyl, n-pentadecyl, n-hexadecyl, n-heptadecyl, n-octadecyl, n-nonadecyl, n-icosayl sulphate. The saturated unbranched C6-C20 alkyl carbonate may be an n-hexyl, n-heptyl, n-octyl, n-nonyl, n-decyl, n-undecyl, n-dodecyl, n-tridecyl, n-tetradecyl, n-pentadecyl, n-hexadecyl, n-heptadecyl, n-octadecyl, n-nonadecyl, n-icosayl carbonate.

In a further preferred embodiment of the invention alkaline salt is a lithium salt, preferably a lithium salt of n-hexyl, n-heptyl, n-octyl, n-nonyl, n-decyl, n-undecyl, n-dodecyl, n-tridecyl, n-tetradecyl, n-pentadecyl, n-hexadecyl, n-heptadecyl, n-octadecyl, n-nonadecyl, n-icosayl sulphate or a lithium salt of n-hexyl, n-heptyl, n-octyl, n-nonyl, n-decyl, n-undecyl, n-dodecyl, n-tridecyl, n-tetradecyl, n-pentadecyl, n-hexadecyl, n-heptadecyl, n-octadecyl, n-nonadecyl, n-icosayl carbonate.

In a preferred embodiment of the invention, the method of producing an isolated protein of the invention further comprises removing the marker or tag after step b) or step c). The removal of the marker may, for example, be carried out by cleaving off the marker. The marker may be cleaved-off by a suitable enzyme which specifically cleaves proteins at specific a cleaving site. The cleaving site could be a protease cleaving site, which may be located in a spacer between the marker or tag and the LBD. Examples of proteases are chymotrypsin, trypsin, elastase and plasmid; the corresponding cleaving sites are known to a person skilled in the art. Since the molecule to be purified is a protein, specific proteases, especially proteases from viruses that normally are attack plants are preferred. Examples of suitable specific proteases are thrombin, Factor Xa, Igase, TEV-protease from tobacco etch virus, protease PreScission (human rhinovirus 3C protease), enterokinase or Kex2 TEV-protease and PreScission are especially preferred.

In a further preferred embodiment of the invention, the alkaline salt of a saturated unbranched C6-C20 alkyl sulphate or carbonate is an alkaline salt of a saturated unbranched C9-C15 alkyl sufate, preferably an alkaline salt of dodecyl sulphate, more preferably lithium dodecyl sulphate (LDS). In accordance with this, the saturated unbranched C9-C15 alkyl sulphate may be a lithium, sodium or potassium salt of n-nonyl sulphate, n-decyl sulphate, n-undecyl sulphate, n-dodecyl sulphate, n-tridecyl sulphate, n-tetradecyl sulphate, or n-pentadecyl sulphate,

particularly lithium n-nonyl sulphate, lithium n-decyl sulphate, lithium n-undecyl sulphate, lithium n-dodecyl sulphate, lithium n-tridecyl sulphate, lithium n-tetradecyl sulphate, or lithium n-pentadecyl sulphate, more particularly lithium dodecyl sulphate, sodium dodecyl sulphate, potassium dodecyl sulphate, in particular lithium dodecyl sulphate.

In another embodiment of the invention, the detergent is an isoprenyl salt. The isoprenyl salt may be any salt of isoprene, such as isoprenyl acetate, isoprenyl diphosphate, isoprenyl pyrophosphate. In another embodiment of the invention, the detergent is a terpene consisting of 1-4 isoprenyl-units and a hydroxyl-group, or an alkaline or ammounium salt of the corresponding carbonate, sulphate, phosphate or pyrophosphate particularly geraniol, farnesol, geranylpyrophosphate, farnesylpyrophosphate or geranylgeranylpyrophosphate.

In a further embodiment of the invention, steps b) and/or c) may be performed in the presence of an agonist for the LBD. The agonist may be the naturally occurring agonist or a functionally active derivative thereof or it may be an agonist different from the naturally occurring agonist. Naturally occurring agonistic ligands that bind to and activate nuclear receptors include lypophilic substances such as endogenous hormones, vitamins A and B and xenobiotic endocrine disrupters. For example, thyroid hormone receptors are activated by binding of thyroid hormone, particularly, thyroxine (T_4). The naturally occurring ligands for retinoic acid receptors are all-trans retinoic acid and 9-cis retinoic acid. Ligands for peroxisome proliferator-activated receptors are free fatty acids and eicosanoids; PPARγ is activated by PGJ_2 (a prostaglandine) and PPARα is activated by leukotriene B_4. Liver X receptor α and β form heterodimers with the obligate partner RXR. The heterodimer can be activated also with an LXR agonist (e.g. oxysterols) or an RXR agonist (such as 9-cis retinoic acid). Oxysterols are the oxygenated derivatives of cholesterol, such as 22 (R)-hydroxy cholesterol, 24 (S)-hydroxysterol, 27-hydroxy sterol and cholestenoic acid. Other agonists may be vitamin D (vitamin D receptor), steroids (estrogen receptor, progesterone receptor, androgen receptor), cortisol (glucocorticoid receptor), aldosterone (mineral corticoid receptor) or fatty acids (hepatocyte nuclear factor 4). Not naturally occurring agonistic ligands include dexamethasone for glucocorticoid receptor or diethylstilbestrol (DES) for estrogen receptor. Suitable agonistic ligands for the members of the ROR family include, cholesterol and derivatives thereof, such as cholesterol sulfate, or melatonin.

As detailed above, the isolated protein of the invention is produced by culturing a cell comprising nucleic acid coding for the protein under suitable conditions conducive to the production of the protein. In one embodiment of the invention, the cell is selected from the group consisting of an animal cell, a plant cell, a yeast cell. Suitable cells of animal cells include mammalian cells, in particular human cells. Examples of those cells are mentioned above. Alternatively, the cell may be a yeast cell, such as an *E. coli* cell, or an insect cell.

An exemplary method of producing an isolated protein of the invention is described in Example 1.

A further aspect of the present invention relates to an isolated protein comprising a ligand binding domain of a nuclear receptor in controllable form produced according to the method of the invention.

The skilled person will understand that the isolated protein can be as defined above in any of the embodiments of the invention, particularly of the preferred embodiments of the invention.

A further aspect of the invention relates to the use of an isolated protein for the identification of a ligand for a ligand binding domain of a nuclear receptor, particularly of an agonist or antagonist, especially an agonist. In accordance with the present invention, the isolated protein of the invention may be used for the identification of a ligand for LBD of a nuclear receptor, particularly an agonist or an antagonist, e.g., by switching on or switching off the downstream signal transduction. In accordance with this, any downstream signal may be detected or evaluated in order to detect binding of a ligand. If it is searched for an agonist, the agonist may be identified by the activation of the downstream signal pathway. On the other hand, an antagonist might be identified by switching off the downstream signal transduction. In case of an antagonist, it might be necessary to use the combination of an agonist and a potential antagonist in order to detect the deactivation of agonist-induced signal transduction by the antagonist. As detailed above, the signal transduction of nuclear receptor in general includes formation of a multimer, particularly a dimer, binding of a co-activator and/or co-repressor, or alteration in gene transcription, often an induction of gene transcription, accordingly, production of mRNA and a protein and, therefore, changed cell function.

In accordance with the above described signal transduction, changed signal transduction may be assessed at each level of signal transduction including binding of the protein of interest to a second protein (such as co-activators or co-repressors), binding to a target gene, determination of the amount of mRNA or a protein, or altered cell function. Methods of determining binding of a protein to a further protein, or a target gene are well known to the skilled person and include those defined herein. Methods for determining the amount of mRNA or protein are also well known to the skilled person. Methods of observing changed cell function largely depend on the type of cell function and are also well known to the skilled practitioner.

A still further aspect of the present invention relates to a test system comprising

- the isolated protein of the invention,
- a co-factor, and
- means for detecting the interaction between the protein and the co-factor upon binding of a ligand, especially an agonist, to the ligand binding domain of the nuclear receptor.

In accordance with the present invention, the isolated protein of the invention may be any of the proteins as specified in the above aspects and embodiments, particularly preferred embodiments. Additionally, the co-factor (also referred to as co-regulator, co-regulatory protein, or transcription co-regulator including also a co-activator or a co-repressor; see also above) is bound by a nuclear receptor activated by the binding of an agonistic ligand, whereas the co-repressor is bound by a nuclear receptor upon binding of an antagonistic ligand. A common feature of nuclear receptor co-activators is that they contain one or more LXXLL binding motifs (a continuous sequence of five amino acids, where L = leucine and X = any amino acid) referred to as NR (nuclear receptor) boxes. The LXXLL binding motifs have been shown to bind to a structure on the surface of the LBD of nuclear receptors. Examples include:

- NCOA-1 (nuclear receptor co-activator 1) / SRC-1 (steroid receptor co-activator-1)
- NCOA-2 (nuclear receptor co-activator 2) / GRIP-1 (glucocorticoid receptor interacting protein 1)
- NCOA-3 (nuclear receptor co-activator 3) / AIB-1 (amplified in breast)
- NCOA-4 (nuclear receptor co-activator 4) / ARA 70 (androgen receptor associated protein 70)

- NCOA-5 (nuclear receptor co- activator 5)
- NCOA-6 (nuclear receptor co-activator 6)
- NCOA-7 (nuclear receptor co-activator 7)
- PGC-1 (proliferator-activated receptor γ co-activator 1)
- CBP (cAMP responsive element-binding (CREB) protein-binding protein)
- PCAF (p300/CBP associating factor)
- ARA 54 (androgen receptor associated protein 54)
- ARA 55 (androgen receptor associated protein 55)

Co-repressor proteins also bind to the surface of the ligand binding domain of nuclear receptors, but through an LXXXIXXX (I/L) motif of amino acids (where L = leucine, I = isoleucine and X = any amino acid). Additionally, co-repressors preferably bind to the nuclear receptor in inactivated form, free of an agonist or, possibly, in antagonist-bound form. Examples of co-receptors include:

- NCOR-1 (nuclear receptor co-repressor)
- NCOR-2 (nuclear receptor co-repressor) / SMRT (silencing mediator (co-repressor) for retinoid and thyroid hormone receptors)
- LCoR (ligand-dependent co-repressor)
- RCOR (REST co-repressor)
- CtBP 602618
- Rb (retinoblastoma protein)
- SIN3 (SIN3a, SIN3b)

Co-factors with dual function activator/repressor include:

- PELP-1 (proline, glutamic acid and leucine-rich protein 1)
- NSD-1
- RIP-14 (RXR-interacting protein 14)

The co-activator may be chosen depending on the LBD which is encompassed in the isolated protein according to the invention. The skilled person will understand that the co-factor depends on the signal transduction of the nuclear receptor the LBD is derived from and he will

be able to choose a suitable co-factor for the test system of the invention to detect the interaction between the co-factor and the isolated protein comprising the LBD upon binding of a ligand.

In a preferred embodiment of the invention, the co-factor is glucocorticoid receptor-inactivating protein-1 (GRIP-1) or steroid receptor co-activator-1 (SRC-1), optionally, labelled with a marked, preferably a tag.

GRIP-1 is a transcriptional co-regulatory protein which contains several nuclear receptor interacting domains and an intrinsic histone acetyl transferase activity. GRIP-1 is recruited to DNA promotion sites by ligand-activated nuclear receptors, such as ROR, particularly RORα. GRIP-1, in turn, acetylates histones which promotes DNA transcription. GRIP-1 is also referred to as SRC-2 (steroid receptor co-activator-2) or transcriptional mediators / intermediary factor 2 (TIF-2) or nuclear receptor co-activator 2 (NCOA2).

SCR-1 is also a transcriptional co-regulatory protein which also contains several nuclear receptor interaction domains and an instrinsic histone acetyl transferase activity. SRC-1 is recruited to DNA promotion sites by ligand-activated nuclear receptors and, in turn, acetylates histones, which promotes downstream DNA transcription. SRC-1 is also referred to as nuclear receptor co-activator-1 (NCOA-1).

In one embodiment of the invention, the co-factor may be labelled with a marker, preferably with a tag. The marker or tag may be defined as above. The marker may be used for purification of the co-factor or it may be used in order to detect the interaction between the isolated protein and the co-factor. Suitable markers include antibodies, antigens, enzymes, radiolabels, etc. However, it should be understood that the co-factor should be labelled in a manner still allowing interaction with other components of the test system in order to allow detection of a signal indicative of the interaction between the isolated protein and the co-factor.

In a preferred embodiment of the invention, the test system is designed in a manner, wherein the proximity of the protein to the co-factor induces a detectably signal.
The proximity of the isolated protein and the co-factor may be reached by binding of the isolated protein to the co-factor. The induction of a signal may be effected by labeling of each

of the components, wherein the proximity of the labels induces a detectable signal. The induced detectable signal may be a chemiluminescent signal, a change in color, a fluorescence signal, a radiation or any other suitable signal. The signal may be induced by interaction of two labels, wherein each label is bound to one component of the test system, i.e. the isolated protein and the co-factor. Such signals include a radiolabel, such as ^{125}I, one the one hand and a suitable quencher on the other hand in order to detect proximity in a scintillation counter (scintillation proximity assay). The components may encompass antigens accessible to antibodies labeled in a manner to detect FRET (fluorescence resonance energy transfer, see below). In another alternative, one of the proteins has bound to its surface a biomolecule capable of phosphorylation by a kinase and the other component has a trivalent metal ion complexed to its surface, e.g., via a suitable linker such as nitrilotriacetic acid, iminodiacetic acid or an appropriately substituted N-containing heterocycle, for example a triazoheterocycle, for example a triazocyclononaneononane, such as 1-propylamino-4-acetato-1,4,7-triazacyclononane. A chemiluminescent signal is generated when the donor and acceptor particles are in close proximity, which occurs upon binding of the protein to the co-factor (luminescent proximity assay).

In one preferred embodiment of the invention, the means for detecting the interaction between the protein and the co-factor include at least one antibody specific for the protein or the co-factor. As detailed above, the isolated protein may include an antigen for a specific antibody and the detection of the antibody. In an even more preferred embodiment of the invention, the test system comprises two antibodies, wherein the first antibody is specific for the protein and the second antibody is specific for the co-factor. The antibody may be labeled with a suitable marker, which is indicative of the presence of the respective component. Alternatively, the above-mentioned primary antibody may be detected by a suitable secondary antibody directed against the primary antibody. The secondary antibody may be used in order to detected the presence of the primary antibody, e.g. when bound to the secondary antibody. The primary or secondary antibody may be labeled with a marker as defined above, for example, an enzyme, a radiolabel, a fluorescence marker, a chemiluminescent marker, etc. Alternatively, the components (the protein and/or the co-factor) may encompass a tag, which is detectable with a suitable antibody.

The test system may be used in a manner that a purification system for the separation of at least one component of the complex is used and the presence of the complex or of each of the components of the complex of the protein and the co-factor is detected. For example, the complex may be purified by gel electrophoresis, column chromatography, affinity purification, etc. and the complex may be detected by the presence of one signal or two signals indicative of one component of the complex or both components of the complex. For example, if a separation technique is used which is specific for one complex of the component, the detection method may be limited to the other component. Alternatively, the purification method may be not specific for one of the components, such as gel electrophoresis, and the formation of the complex may be detected by two signals, e.g. two antibodies labeled with distinguishable fluorescence markers, wherein the presence of both fluorescence markers e.g. at the same area of a gel is indicative of the complex.

In a still further embodiment of the invention, the test system of the invention is characterized in that

a) the first antibody is labeled with a donor moiety for fluorescence resonance energy transfer (FRET) and the second antibody is labeled with an acceptor moiety for FRET or vice versa; or

b) the first antibody is labeled with a donor moiety for time-resolved fluorescence resonance energy transfer (TR-FRET) and the second antibody is labeled with an acceptor moiety for TR-FRET or vice versa; or

c) the first antibody is labeled with a donor moiety for Amplified Luminescence Proximity Homogeneous Assay (ALPHA) and the second antibody is labeled with an acceptor moiety for ALPHA or vice versa.

Fluorescence resonance energy transfer (FRET) describes a radiation-free energy transfer between two chromophores. A donor chromophore in its excited state can transfer energy by a non-radiative long-range dipole-dipole coupling mechanism to an acceptor fluorophore in close proximity (typically <10 nm). As both molecules are fluorescent, the energy transfer is often referred to as "fluorescence resonance energy transfer", although the energy is not actually transferred by fluorescence. FRET is a useful tool to detect and quantify protein-protein interactions, protein-DNA interactions, and protein-conformational changes. For monitoring binding of one protein to another or one protein to DNA, one of the molecules is labeled with a donor and the other with an acceptor and these fluorophore-labeled molecules are mixed. When

they are present in an unbound state, donor emission is detected upon donor excitation. Upon binding of the molecules, the donor and acceptor are brought in proximity and the acceptor emission is predominantly observed because of the intermolecular FRET from the donor to the acceptor. Suitable neighbors for FRET are known in the art and the skilled practitioner will be able to choose a suitable combination of labels for both antibodies. As used herein with respect to donor and corresponding acceptor, "corresponding" refers to an acceptor fluorescent moiety having an emission spectrum that overlaps with the excitation spectrum of the donor. However, both signals should be separable from each other. Accordingly, the wavelength maximum of the emission spectrum of the acceptor should preferably be at least 30 nm, more preferably at least 50 nm, such as at least 80 nm, at least 100 nm or at least 150 nm greater than the wavelength maximum of the excitation spectrum of the donor.

Representative donor fluorescent moieties that can be used with various acceptor fluorescent moieties in FRET technology include fluorescein, Lucifer Yellow, B-phycoerythrin, 9-acridineisothiocyanate, Lucifer Yellow VS, 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid, 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin, succinimdyl 1-pyrenebutyrate, and 4-acetamido-4'-isothiocyanatostilbene-2,2'-disulfonic acid derivatives. Representative acceptor fluorescent moieties, depending upon the donor fluorescent moiety used, include LC-Red 610, LC -Red 640, LC-Red 670, LC -Red 705, Cy5, Cy5.5, Lissamine rhodamine B sulfonyl chloride, tetramethyl rhodamine isothiocyanate, rhodamine x isothiocyanate, erythrosine isothiocyanate, fluorescein, diethylenetriamine pentaacetate or other chelates of Lanthanide ions (e.g., Europium, or Terbium). Donor and acceptor fluorescent moieties can be obtained, for example, from Molecular Probes (Junction City, OR) or Sigma Chemical Co. (St. Louis, MO).

Alternatively, time-resolved fluorescence resonance energy transfer (TR-FRET) may be used for the test system of the present invention. TR-FRET unites TRF (time-resolved fluorescence) and the FRET principle. This combination combines the low background benefits of TRF and the homogeneous assay format of FRET. While FRET has already been described above, TRF takes advantage of the unique properties of lanthanides or any other donor with long half-life. Suitable donors for TR-FRET include, amongst others, lanthanide chelates (cryptates) and some other metal ligand complexes, which can have fluorescent half-life in the micro- to millisecond time range and which, therefore, also allow the energy transfer to occur in micro-

to millisecond measurements. Fluorescence lanthanide chelates have been used as energy donors in the late seventies. The commonly used lanthanides include samarium (Sm), europium (Eu), terbium (Tb) and dysprosium (Dy). Because of their specific photophysical and spectral properties, complexes of lanthanides are of major interest for fluorescence application in biology. Specifically, they have a large stroke's shift and extremely long emission half-lives (from microseconds to milliseconds) when compared to more traditional fluorophores.

Usually, organic chromophores are used as acceptors. These include allophycocyanin (APC). Suitable details on TR-FRET as well as acceptors are described in WO 98/15830.

In a further embodiment of the invention, the test system of the invention is adapted for an amplified luminescence proximity homogeneous assay (ALPHA). ALPHA is a solution-based assay which was originally developed by Packard BioScience. ALPHA is a luminescence-based proximity assay, wherein one interaction partner is attached to donor beads, while the other is coupled to acceptor beads, both with a diameter of only about 250 nm. A photosensitizer compound is embedded into the donor bead. With this compound upon illumination with laser light at a wavelength of about 680 nm, ambient oxygen is converted into energy-rich, short-life singlet oxygen. When no acceptor bead is in proximity, the singlet oxygen decays without producing a signal. If donor and acceptor bead are brought together (ca. 250 nm) by the biological interaction of the attached biomolecules, the singlet oxygen released by the donor bead initiates a luminescence/fluorescence cascade in the nearby acceptor bead, leading to a highly amplified signal in the 520-620 nm range. The luminescence signal is detected in a suitable reader. For more details regarding ALPHA techniques, see Ullman et al., 1994, Proc. Natl. Acad. Sci., USA 91, 5426-5430.

An exemplary test system and its use are described in Example 2 and illustrated in Figure 1C.

Still a further aspect of the present invention relates to a method of screening a ligand for a ligand binding domain of a nuclear receptor comprising the steps of:

a) contacting the test system according to the invention with a substance and
b) detecting a measurable signal upon binding of the substance to the ligand binding domain, thereby identifying the substance as a ligand for the ligand binding domain.

In accordance with the present invention, the test system may be specified as detailed in the present description of the invention, particularly as detailed in the preferred embodiments. The test system may be encompassed in a cell or it may be a cell-free system. A cell-free system is preferred. The test system may be contacted with a test substance under conditions suitable to detect a measurable signal. This includes a suitable temperature, chemical environment (buffers, pH-value, etc.) as well as a suitable concentration of the substance and an appropriate time of contact.

After or simultaneously with the contacting, a signal is observed wherein the detection of a signal is indicative of a ligand for the LBD. The signal may be any signal as detailed above in the context of the test system of the invention and the isolated protein of the invention and its use.

In a preferred embodiment of the method of the invention, the test system includes a first antibody labeled with a donor for FRET or TR-FRET and a second antibody labeled with an acceptor moiety for FRET or T-FRET or vice versa. Additionally, the presence of FRET is indicative of an agonist.

In a preferred embodiment of the invention, the screening method is used for screening for a medicament for preventing and/or treating a coronary artery disease (CAD), arteriosclerosis, dyslipidemia, a neurodegenerative disease, sleep disorder, a disease of circadian rhythmicity or osteoporosis.

The aforementioned diseases involve ROR, particularly RORα. Accordingly, it is assumed that an agonist or an antagonist of an ROR, particularly an RORα LBD, will have a beneficial influence on these diseases and may therefore be used in order to prevent or treat these diseases.

In the following, the present invention is illustrated by figures and examples which are not intended to limit the scope of the present invention.

FIGURES

Figure 1A shows a schematic illustration of RORα1, indicating the various domain of this nuclear receptor.

Figure 1B shows various constructs of the invention, wherein the LBD of RORα1 is linked to a cleavage site (TEV or PreScission (PreSci)), a His-tag (His) and optionally Glutathione-S-transferase-tag (GST).

Figure 1C shows an exemplary illustration of an FRET assay according to the invention. In this assay, RORα is detected by a specific antibody bound to FRET donor Europium (Eu). The co-activator (CA) is biotin labeled. The biotin label is detecting by streptavidin, which encompasses the FRET acceptor allophycocyanin (APC) marker. If an agonist is bound to RORα, the NR is activated a CA binds to the NR. Accordingly, APC is brought into proximity of Eu, leading to a detecting FRET signal (left panel). In the presence of an antagonist, CA does not bind to RORα, accordingly no FRET signal is produced (right panel).

Figure 2 shows the results of the assay of Example 2. Cholesterol and cholesterol sulfate dose-dependently induce binding of a co-activator peptide SRC1-NR1+2 to RORα as measured by an increase in fluorescence intensity ratio in a fluorescence resonance energy transfer assay. The upper line (+) represents an assay perform in the presence of increasing concentrations of cholesterol sulfate, whereas the middle line (O) shows the assay carried out in the presence of increasing concentrations of cholesterol. The bottom lines (□ and Δ) represent experiments carried out in the presence of solvent only.

EXAMPLES:

Example 1: RORα - Expression, Purification and AD

Expression and purification of RORalpha protein was performed according to Kallen et al.. 2002, Structure, 10 (1697).

For expression and purification, DNA of RORα1 encompassing the ligand binding domain (LBD, amino acids 271-523) was cloned into the pVL1393 vector with an N-terminal stretch of 6 consecutive histidine-residues and a recognition sequence for the HRV 3C protease. Virus

generation and expression in Sf9 insect cells was done following standard procedures. Infected cells were harvested 72 hrs post infection and cell pellets were stored at -80°C.

Frozen cell pellets were resuspended in 500 mM NaCl, 50 mM Tris, pH 8.0, 5 mM β-mercaptoethanol. Protease inhibitors (Complete EDTA-free, Roche Diagnostics) as well as Benzonase (Novagen) were added. The lysate was stirred for 30 minutes on ice and cells were finally disrupted by sonication. After centrifugation, the cleared lysate was loaded onto a Ni^{2+} - charged HisTrap FF (GE Healthcare) column equilibrated with 500 mM NaCl, 50 mM Tris, pH 8.0. Elution was done with a linear gradient over 30 CV to 500 mM NaCl, 500 mM imidazole, pH 8.0. Fractions were analyzed by SDS-PAGE and corresponding fractions were pooled and dialyzed against 30 mM Hepes, pH 7.0, 20 mM NaCl, 2 mM DTT, 5% glycerol.

For further purification, protein was loaded onto a HiTrap Q FF anion exchange column (GE Healthcare), equilibrated with 30 mM Hepes, pH 7.0, 20 mM NaCl, 2 mM DTT. Protein was eluted over a 20 CV gradient to 30 mM Hepes, pH 7.0, 1 M NaCl, 2 mM DTT. Corresponding fractions were pooled and further purified on a Superdex 200 26/60 gelfiltration column (GE Healthcare) equlilibrated in 150 mM NaCl, 5 mM DTT, 50 mM Tris, pH 7.5. Fractions containing RORαLBD were pooled and concentrated. The protein fractions were at least 95% pure as judged by SDS-PAGE and capillary electrophoresis (Agilent 2100 Bioanalyzer).

Protein (2.3 mg/ml) was aliquoted and stored at -80°C. Prior usage for *in vitro* assays, protein was dialyzed against 20 mM Tris, pH 7.5 for 8 hrs at room temperature. Lithium dodecylsulphate (LDS) was added to a final concentration of 3 mM and the mixture was further incubated at room temperature with gently shaking over night. Protein was aliqoted and stored at -80°C.

The purified and LDS-treated RORα LBD protein was shown to interact with glucocorticoid receptor-interacting protein 1 (GRIP1) and steroid receptor co-activator 1 (SRC1) in a co-factor recruitment fluorescence resonance energy transfer assay (FRET). Specifically the peptides GRIP1-NR1, GRIP1-NR2 and SRC1-NR1+2 - derived from the different nuclear receptor boxes (NR) of the two co-factors - were found to interact with RORα.

In all three cases cholersterol sulphate was able to induce co-factor binding dose-dependently with an EC_{50} of 22 µM. 20 nM RORα LBD protein labeled with equimolar amounts of fluorescent anti-6xHis antibody were incubated with 400 nM of the respective biotin-labeled peptide and equimolar amounts of a streptavidin-tagged fluorophore in a buffer containing 50 mM Tris, 100 mM NaCl, 1 mM DTT and 0.1 % BSA for 2 hrs. Time resolved measurement of fluorescence ratios at 665 nm and 612 nm showed dose-dependent co-factor recruitment by known agonists like cholesterol sulphate and cholesterol.

Example 2: TR-FRET assay measuring RORalphaLBD interaction with the LXXLL peptide, SRC1-NR1+2

For the TR-FTET assay the following materials were used:

Proteins:

- 6xHis-RORalpha:
  source : Protein Production (Lab. Thomas Langer) treated with 3 mM LDS, stock solution at 1.5 mg/ml, storage at –20°C, pipetting at room temperature, corresponding to 47,5 µM (with MW 31,6 kD), Concentration in the assay = 20 nM

- SRC1-NR1+2:
  source: JPT Peptide Technologies, stock solution at 250 µM, storage at –20°C, concentration in the assay = 400 nM

Fluorophores/Labels:

- anti-6xHis antibody:
  source: Perkin Elmer # AD0110, stock solutions are at varying concentration with each batch, storage at –20°C, concentration in the assay = 20 nM
- strep-APC:
  source: Perkin Elmer # AD0201, stock solutions are at varying concentration with each batch, storage at +4°C after reconstitution, concentration in the assay = 200 nM (based on streptavidin)

Agonist:

- Cholesterol sulfate:
  source: SIGMA #C9523, stock solution in DMSO (30 mM), storage at –20°C

Buffer:

- 50 mM Tris, 100 mM NaCl, 1 mM DTT, 0,1 % Bovine Serum Albumin (add BSA fresh each day), adjust pH to 7,4

Plates:

- Test plates: CORNING Costar # 3639 (96 well half area)

The assay was carried out as follows:

All proteins, labels and compounds were diluted in the assay buffer just prior to assay setup. This assay was realised in 96 well half area plates. The final test volume was 20 µl. It is divided in 2 pipetting steps :

First, a mix containing 6xHis-RORalphaLBD protein, BiotinSRC1-NR1+2 peptide, anti-6xHis antibody and strep-APC was prepared on ice in the following concentrations:

		Test concentration (final):
6xHis-RORalphaLBD	27 nM	20 nM
BiotinSRC1-NR1+2	533 nM	400 nM
anti-6xHis antibody	27 nM	20 nM
strep-APC	267 nM	200 nM

⇨ 15µl of this mix is prepipetted into a white 96 well plate

In a second step, 5 µl of the agonist compound was added as a 4-fold concentrated solution. Cholesterol sulphate was typically diluted from 300 µM final compound concentration downwards in 2-fold dilution steps and yields an EC_{50} of about 10 µM. Plates were sealed to avoid evaporation.

The plates were incubated for 2h in subdued light at room temperature and read on a Tecan ULTRA at room temperature with 340 nm excitation and 612 nm as Europium reference and 665 nm as APC FRET signal. The results of this test are shown in Fig. 2.

Claims

1. An isolated protein comprising a ligand binding domain of a nuclear receptor in controllable form.

2. The isolated protein of claim 1, wherein the nuclear receptor is a Retinoic acid receptor-related Orphan Receptor (ROR), particularly RORα, RORβ or RORγ, especially RORα.

3. The isolated protein of claim 1 or 2, wherein the ligand binding domain is activatable upon binding of an agonist to the ligand binding domain.

4. The isolated protein of any of claims 1 to 3, wherein the protein further comprises a marker, particularly a tag.

5. The isolated protein of claim 4, wherein the tag is selected from the group consisting of a His-tag, Arg-tag, Strep-tag, Flag-tag, T7-tag, V5-peptide-tag, c-Myc-tag, S-tag, HAT-tag, calmodulin-binding peptide-tag, chitin-binding peptide-tag, GST-tag and MBP-tag.

6. The isolated protein of claim 4 or 5, wherein the marker or tag is removable from the protein by proteolytic cleavage at a specific cleavage site.

7. The isolated protein of any of claims 1 to 6 comprising or consisting of the sequence of SEQ ID NO: 1 or 2.

8. A method of producing the isolated protein of any of claims 1 to 7, comprising the steps of:
   a) culturing a cell comprising a nucleic acid coding for the protein of any of claims 1 to 7 under suitable conditions conducive to the production of the protein,
   b) isolating the protein from the cell culture, and
   c) contacting the isolated protein of step b) with an detergent, particularly with an alkaline salt of a saturated unbranched C6 to C20 alkyl sulphate or carbonate or with an isoprenyl salt.

9. The method of claim 8, wherein the method further comprises removing the marker or tag after step b).

10. The method of claim 9 or 10, wherein the alkaline salt of a saturated unbranched C6 to C20 alkyl sulphate or carbonate is an alkaline salt of a saturated unbranched C9 to C15

alkyl sulphate, preferably an alkaline salt of dodecyl sulphate, more preferably lithium dodecyl sulphate (LDS).

11. The method of any of claims 8 to 10, wherein steps b) and/or c) are performed in the presence of an agonist for the ligand binding domain.

12. The method of any of claims 8 to 11, wherein the cell is selected from the group consisting of an animal cell, a plant cells and a yeast cell, particularly wherein the cell is an insect cell or an *E. coli* cell.

13. An isolated protein comprising a ligand binding domain of a nuclear receptor in controllable form produced according to the method of any of claims 8 to 12.

14. Use of an isolated protein according to any of claims 1 to 7 and/or 13 for the identification of a ligand for a ligand binding domain of a nuclear receptor, particularly an agonist or antagonist, especially an agonist.

15. A test system comprising
- the isolated protein according to any of claims 1 to 7 and/or 13,
- a co-factor, and
- means for detecting the interaction between the protein and the co-factor upon binding of a ligand, especially an agonist, to the ligand binding domain of the nuclear receptor.

16. The test system of claim 15, wherein the co-factor is glucocorticoid receptor-inactivating protein-1 (GRIP-1) or steroid receptor co-activator 1 (SRC1), optionally labeled with a marker, preferably a tag.

17. The test system of claim 15 or 16, wherein the proximity of the protein to the co-activator induces a detectable signal.

18. The test system of any of claims 15 to 17, wherein the means for detecting the interaction between the protein and the co-factor include at least one antibody specific for the protein or the co-factor.

19. The test system of claim 18, wherein the test system comprises two antibodies, wherein the first antibody is specific for the protein and the second antibody is specific for the co-factor.

20. The test system of claim 19, wherein
    (a) the first antibody is labeled with a donor moiety for fluorescence resonance energy transfer (FRET) and the second antibody is labeled with an acceptor moiety for FRET or vice versa; or
    (b) the first antibody is labeled with a donor moiety for time-resolved fluorescence resonance energy transfer (TR-FRET) and the second antibody is labeled with an acceptor moiety for TR-FRET or vice versa; or
    (c) the first antibody is labeled with a donor moiety for Amplified Luminescence Proximity Homogeneous Assay (ALPHA) and the second antibody is labeled with an acceptor moiety for ALPHA or vice versa.

21. A method for screening for a ligand for a ligand binding domain of a nuclear receptor comprising the steps of:
    a) contacting the test system according to any of claims 15 to 20 with a substance and
    b) detecting a measurable signal upon binding of the substance to the ligand binding domain, thereby identifying the substance as a ligand for the ligand binding domain.

22. The method of claim 21, wherein the test system is as defined in claim 20 and wherein the presence of FRET is indicative for an agonist.

23. The method of claims 21 and 22, wherein the method is used for screening for an medicament for preventing and/or treating a coronary artery disease (CAD), atherosclerosis, dyslipidemia, a neurodegenerative disease, sleep disorder, a disease of circadian rhythmicity or osteoporosis.