



## (12)发明专利

(10)授权公告号 CN 105279394 B

(45)授权公告日 2017.12.26

(21)申请号 201510681442.1

(22)申请日 2015.10.13

(65)同一申请的已公布的文献号

申请公布号 CN 105279394 A

(43)申请公布日 2016.01.27

(73)专利权人 山西农业大学

地址 030801 山西省晋中市太谷县铭贤南路1号

(72)发明人 李鹏飞 吕丽华 谢建山 孟金柱

刘岩 姚晓磊 赵妙妙 景灵婕

毕锡麟 王锴

(51)Int.Cl.

G06F 19/16(2011.01)

(56)对比文件

CN 1735595 A,2006.02.15,

CN 1909888 A,2007.02.07,

US 8886505 B2,2014.11.11,

CN 104208069 A,2014.12.17,

WO 2013050481 A1,2013.04.11,

CN 1726044 A,2006.01.25,

审查员 贺馨

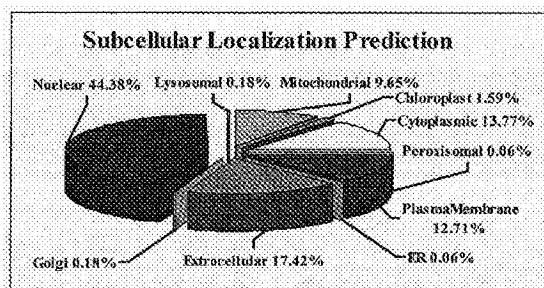
权利要求书1页 说明书4页 附图3页

(54)发明名称

用于神经肽受体筛选的方法

(57)摘要

本发明公开了一种用于神经肽受体筛选的方法,依据统计学原理,要求基因组测序结果FDR校正后的 $P < 0.05$ ,基因差异表达倍数的选择依据是配体-受体结合产生的效应强度与占领受体的数量成正比;采用蛋白同源建模技术建立蛋白模型;综合应用评价体系最为全面的SWISS-MODEL、MODELLER和PHYRE2三项同源建模技术,避免了单一建模技术造成模型质量不高的问题;本具体实施在分子对接过程中,各技术参数设定要求明确,对接区域精准,对接成功后需通过几何匹配和分子动力学方法优化对接,经相互作用预测临界评价和分子力场评价函数获得对接复合体评分,综合表达量验证、神经肽受体的分子特征及功能分析确定候选受体。



1. 用于神经肽受体筛选的方法,其特征在于,包括如下步骤:

S1、明确要研究的神经肽及其分布组织,采集2组不同生理条件下的该组织,分别分离神经肽作用的靶细胞,进行转录组高通量测序;

S2、将步骤S1获得的测序数据设定条件FDR校正后的 $P < 0.05$ ,选择高差异表达基因,表达差异倍数根据测序数据设定;

S3、将步骤S2获得的差异表达基因应用

CELLOV2.5SubcellularLocalizationPrediction数据库进行亚细胞定位;

S4、取步骤S3获得的膜蛋白的FASTA序列,应用HMMTOPV2.0筛选7次跨膜 $\alpha$ 螺旋区段蛋白,即为G蛋白偶联受体;

S5、应用同源建模技术对神经肽和步骤S4中获得的G蛋白偶联受体分别建立蛋白模型并转化为分子坐标结构:

S6、根据步骤S5获得的神经肽和各G蛋白偶联受体的PDBQT文件,应用ZDOCKV3.0.2分子对接技术对神经肽和各G蛋白偶联受体分别进行模型对接;

S7、将步骤S6中配体和各受体对接的蛋白复合体模型和评分函数输出,筛选评分函数值最高的3~5个蛋白复合体,将这3~5个蛋白复合体对应的G蛋白偶联受体作为候选受体;

S8、将步骤S7所得的G蛋白偶联受体分别在293T细胞中过表达,荧光标记该神经肽,通过配体-受体结合实验最终确定受体。

2. 根据权利要求1所述的用于神经肽受体筛选的方法,其特征在于,步骤S3的具体步骤为:首先通过NCBIFASTA搜索引擎分别找出每一个Genesymbol的蛋白序列,选择物种;然后将该基因的FASTA序列输入CELLOV2.5对话框,并对结果进行统计。

3. 根据权利要求1所述的用于神经肽受体筛选的方法,其特征在于,步骤S5的具体步骤为,采用综合应用评价体系最为全面的SWISS-MODEL、MODELLER和PHYRE2三项同源建模技术,通过评分函数、模型的跨膜折叠数和结合域区域分布,对三种建模技术获得的每一个蛋白的分子模型进行对比评价,选择最为可信合理的分子模型并输出PDBQT文件。

## 用于神经肽受体筛选的方法

### 技术领域

[0001] 本发明涉及神经肽生物学和动物生理调控、药物靶标筛选研究技术领域,具体涉及一种用于神经肽受体筛选的方法。

### 背景技术

[0002] 受体相互作用与调控机理的研究开拓了新的思路。神经肽作为神经元间重要的化学信使,执行神经调质、神经递质和神经激素的作用。神经肽在动物体广泛分布,表明生物学功能具有多样性和复杂性,主要体现在:1)神经肽数量众多;2)同一神经肽有多种功能,且对同一靶器官(靶组织或靶细胞)的效应也不相同;3)同一神经肽随动物种属及剂量、作用部位的不同,功能也不一样;4)同一神经肽对不同效应细胞作用不同,可产生多种效应。因此,一直以来,神经肽及其受体的研究受到医学家、生理学家及分子药理学家的高度重视。

[0003] 已发现的神经肽受体中,除心房钠尿肽受体本身具有G蛋白介导的功能外,其它神经肽受体都属于G蛋白偶联的受体,这类受体的特征是受体蛋白的肽链跨膜7次并形成7个 $\alpha$ 螺旋区段。

[0004] 基于神经肽及其受体——G蛋白偶联受体在生理病理过程中的重要生物作用,这一蛋白家族也是目前最重要的药物作用靶标库。因此,筛选神经肽的受体有助于进一步揭示人类及动物相关疾病的发生发展过程,对新药物的研制开发和疾病治疗具有重大意义。

[0005] 目前,神经肽受体的筛选多采用传统的大规模筛选方法,如酵母双杂交、细菌双杂交、蛋白质芯片和免疫亲和技术,共同缺点就是实验过程繁琐、工作量非常大、费用高,实验结果中假阳性、假阴性问题突出,干扰因素多,造成后期受体的确定困难巨大。

### 发明内容

[0006] 为解决上述问题,本发明提供了一种用于神经肽受体筛选的方法。

[0007] 为实现上述目的,本发明采取的技术方案为:

[0008] 用于神经肽受体筛选的方法,其特征在于,包括如下步骤:

[0009] S1、明确要研究的神经肽及其分布组织,采集2组不同生理条件下的该组织,分别分离神经肽作用的靶细胞,进行转录组高通量测序;

[0010] S2、将步骤S1获得的测序数据设定条件FDR校正后的 $P < 0.05$ ,选择高差异表达基因(表达差异倍数根据测序数据设定);

[0011] S3、将步骤S2获得的差异表达基因应用CELLO V2.5 Subcellular Localization Prediction数据库进行亚细胞定位;

[0012] S4、取步骤S3获得的膜蛋白的FASTA序列,应用HMMTOP V2.0 (Prediction of transmembrane helices and topology of proteins Version 2.0)筛选7次跨膜 $\alpha$ 螺旋区段蛋白,即为G蛋白偶联受体;

[0013] S5、应用同源建模技术对神经肽和步骤S4中获得的G蛋白偶联受体分别建立蛋白模型并转化为分子坐标结构;

[0014] S6、根据步骤S5获得的神经肽和各G蛋白偶联受体的PDBQT文件,应用ZDOCK V3.0.2分子对接技术对神经肽和各G蛋白偶联受体分别进行模型对接;关键技术在于参数的设定,包括建立受体、配体结合中心网格,设定蛋白分子结合距离 $< 6 \text{ \AA}$ ,预测结合部位1,000,评分函数值由高到低,显示结合构象为前五个预测结果,其它参数软件自动默认;其中,受体、配体结合中心网格的设定关系到结果的准确性,在设定时,通过实验研究和查阅现有资料,确定受体、配体的结合域氨基酸区段;

[0015] S7、将步骤S6中配体和各受体对接的蛋白复合体模型和评分函数输出,筛选评分函数值最高的3~5个蛋白复合体,将这3~5个G蛋白偶联受体作为候选受体;

[0016] S8、将步骤S7所得的G蛋白偶联受体分别在293T细胞中过表达,荧光标记该神经肽,通过配体-受体结合实验最终确定受体。

[0017] 其中,步骤S3的具体步骤为:首先通过NCBI FASTA搜索引擎分别找出每一个Gene symbol的蛋白序列(越完整越好),选择物种;然后将该基因的FASTA序列输入CELLO V2.5对话框,并对结果进行统计;

[0018] 其中,步骤S5的具体步骤为,采用综合应用评价体系最为全面的SWISS-MODEL、MODELLER和PHYRE2三项同源建模技术,通过评分函数、模型的跨膜折叠数和结合域区域分布(通过Java软件实现),对三种建模技术获得的每一个蛋白的分子模型进行对比评价,选择最为可信合理的分子模型并输出PDBQT文件。

[0019] 本具体实施依据统计学原理,要求基因组测序结果FDR(False Discovery Rate,错误发现率)校正后的 $P < 0.05$ ,基因差异表达倍数的选择依据是受体占领学说——配体-受体结合产生的效应强度与占领受体的数量成正比,故在不同组织或细胞转录组中受体与配体表达趋势相同;采用目前应用较为广泛,也是蛋白质三维结构预测最精确的方法——蛋白同源建模技术建立蛋白模型;同时,综合应用评价体系最为全面的SWISS-MODEL、MODELLER和PHYRE2三项同源建模技术,避免了单一建模技术造成模型质量不高的问题;本具体实施在分子对接过程中,各技术参数设定要求明确,对接区域精准,对接成功后需通过几何匹配和分子动力学方法优化对接,经相互作用预测临界评价和分子力场评价函数获得对接复合体评分,综合表达量验证、神经肽受体的分子特征及功能分析确定候选受体。

[0020] 本发明具有以下有益效果:

[0021] (1) 使用生物信息学技术对神经肽的受体进行规模筛选,减少了繁琐的实验室工作,节约了实验费用,避免了假阳性、假阴性结果的出现。

[0022] (2) 采用本发明方法得到的候选受体数量可控,筛选过程严格按照配体-受体的结构识别理论进行,各技术参数设定要求明确,对接区域精准,分辨率高、准确性和重复性好,为进一步受体-配体结合实验确定受体奠定基础。

## 附图说明

[0023] 图1为1331个高差异表达基因CELLO V2.5蛋白亚细胞定位获得的188个膜蛋白。

[0024] 图2为188个膜蛋白HMMP TOP V2.0跨膜 $\alpha$ 螺旋区预测获得的38个G蛋白偶联受体。

[0025] 图中,横坐标为跨膜螺旋区数量,纵坐标为蛋白数量。

[0026] 图3为经评分函数、模型的跨膜折叠数和结合域区域分布选择,利用SWISS-MODEL建立的CART模型(将PDBQT文件输入Java软件实现)可信合理,左图为立体模型,右图为分子

模型(以CART模型为例,其他38个G蛋白偶联受体模型略)。

[0027] 图4为SWISS-MODEL建立的CART模型的质量评价曲线,横坐标为对应的氨基酸残基,纵坐标为评分值,要求所构建模型的每一个氨基酸残基的评分最好为正值,否则表明该氨基酸残基在模型中的空间位置不可信合理。

[0028] 图5为CART与TEDDM1分子对接后形成的蛋白复合体模型。

### 具体实施方式

[0029] 为了使本发明的目的及优点更加清楚明白,以下结合实施例对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0030] 实施例1

[0031] 用本发明的方法筛选牛卵泡可卡因-苯丙胺调节转录肽(Cocaine-and Amphetamine-Regulated Transcript,CART)的候选受体,采用如下步骤:

[0032] S1、选取8头10月龄青年母牛(海福特杂交后代),同期发情处理,期间每天用B超仪检测并记录卵泡的生长状况。在第一卵泡波的卵泡发育偏离期启动时,屠宰并摘除卵巢,用眼科剪剪下出现偏差时的最大卵泡ODF1(The largest follicle at onset of deviation,发情后的第4~5d)和第二大卵泡ODF2(The second largest follicle at onset of deviation),放入灭菌的DPBS中备用;

[0033] S2、将S1取得的卵泡用眼科剪剪开一小口,放到盛有DPBS的表面皿上,用细胞刮刀轻轻刮取附于卵泡内膜的颗粒细胞(Granulosa Cells,GCs)并在DPBS中清洗,移液枪吸取含有GCs的DPBS混合液于灭菌EP管中离心,弃上清;

[0034] S3、将S2取得的GCs Trizol法分别提取总RNA,Illumina HiSeq 2000测序平台进行测序,分别获得44,189,827和43,826,914个原始“reads”;

[0035] S4、将S3获得的原始“reads”经过滤、装配、数据库比对,最终获到35,325个已经注释的转录本;

[0036] S5、取S4获得的35,325个转录本,设定条件为:ODF1 RPKM/ODF2 RPKM>2(对于CART,ODF1 RPKM/ODF2 RPKM=38.15),FDR校正后的P<0.05,获得符合条件的高差异表达基因1331个;

[0037] S6、通过CELLO V2.5 Subcellular Localization Prediction数据库对S5获得的1331个Gene symbol进行亚细胞定位。首先通过NCBI中的FASTA搜索引擎分别找出每一个蛋白的序列,物种选择牛;然后将FASTA序列分别输入CELLO V2.5对话框,并对结果进行统计,共获得188个膜蛋白;

[0038] S7、应用HMMTOP V2.0对S6获得的188个膜蛋白进行跨膜 $\alpha$ 螺旋区预测,共筛选出38个膜蛋白含有7个跨膜 $\alpha$ 螺旋区,即为G蛋白偶联受体;

[0039] S8、综合应用评价体系最为全面的SWISS-MODEL、MODELLER和PHYRE2三项同源建模技术对S7获得的38个G蛋白偶联受体,通过评分函数、模型的跨膜折叠数和结合区域分布,对三种建模技术获得的每一个蛋白的分子模型进行对比评价,选择最为可信合理的分子模型并输出PDBQT文件;

[0040] S9、以CART做为配体,38个G蛋白偶联受体分别作为受体,ZDOCK V3.0.2大分子对

接软件对S8获得的PDBQT文件进行模拟结合。在这里以CART和TEDDM1对接为例：设定蛋白分子结合距离 $< 6 \text{ \AA}$ ，预测数量1,000，评分函数值由高到低，显示结合构象为前五个预测结果，其它参数软件自动默认；中心网格（结合域）的设定力求短小准确，要求配体/受体活性区域较为明确的，就准确设定；不明确的需依据现有资料模糊设定，过短易将活性肽段排除在模拟结合区域外，影响复合体评分，过长会影响模拟结合次数，造成一些区域错过模拟结合的机会。牛CART全长116个氨基酸，其中CART<sub>1-27</sub>为信号肽，活性肽段为CART<sub>76-116</sub>，牛TEDDM1全长509个氨基酸，活性区域不明确，但跨膜区肽段为TEDDM1<sub>8-250</sub>，因此，TEDDM1的活性区域选择TEDDM1<sub>8-250</sub>；参数设定完成提交运行；

[0041] S10、将S9中CART分别与38个G蛋白偶联受体的对接结果（包括蛋白复合体模型和每一蛋白复合体的评分值）输出，通过相互作用预测临界评价和分子力场评价函数获得评分值，最终选择TEDDM1，CMKLR1，AGTR2和GPR116四个G蛋白偶联受体作为CART的候选受体；

[0042] S11、将S10中的结果经qRT-PCR验证，综合神经肽受体的特征，如：神经肽受体的相对分子质量为40,000~50,000Da，由350~500个氨基酸组成等，最终将TEDDM1，CMKLR1和AGTR2作为CART的候选受体。

[0043] 以上所述仅是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以作出若干改进和润饰，这些改进和润饰也应视为本发明的保护范围。

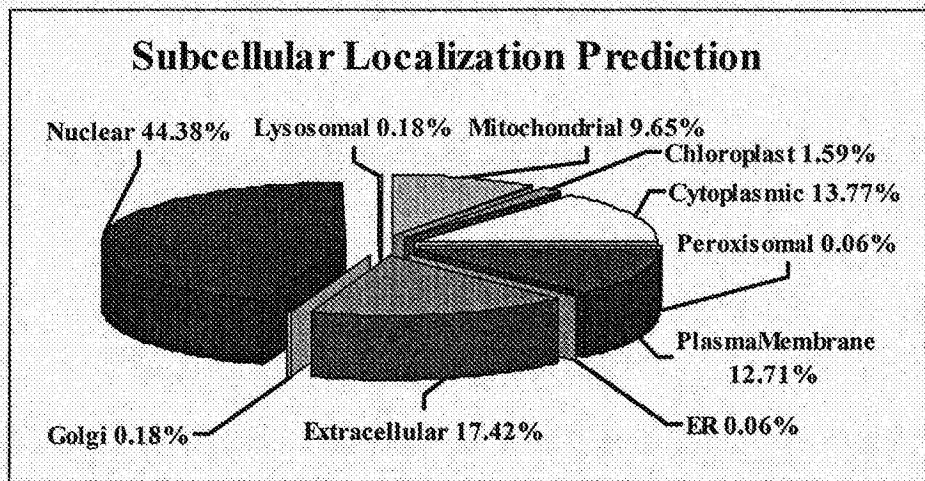


图1

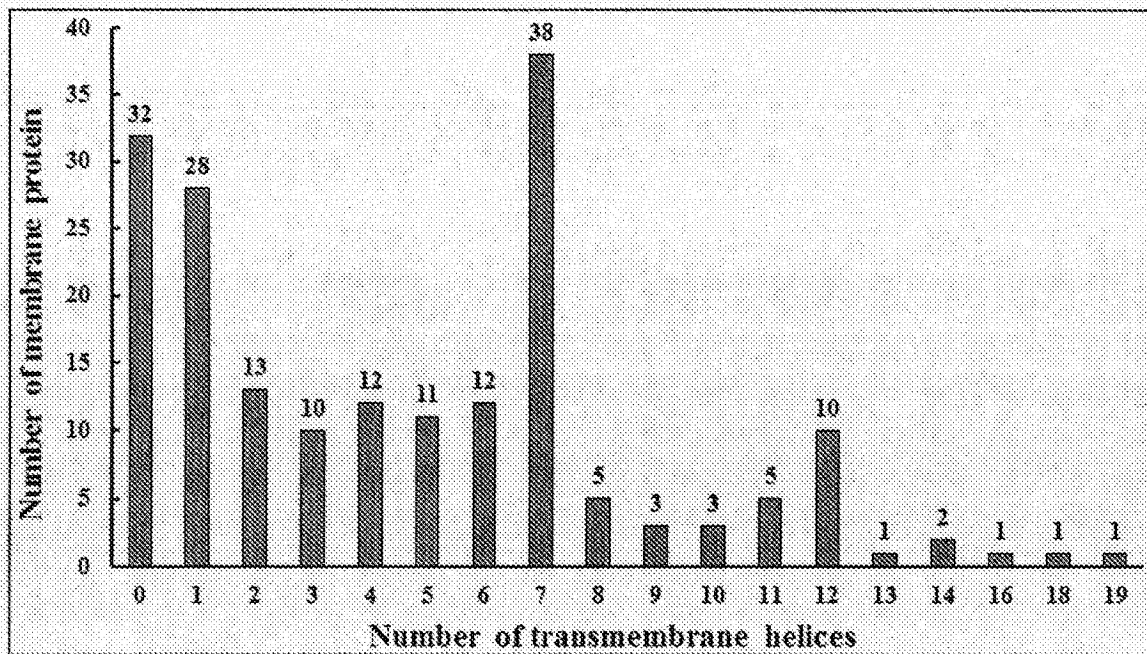


图2

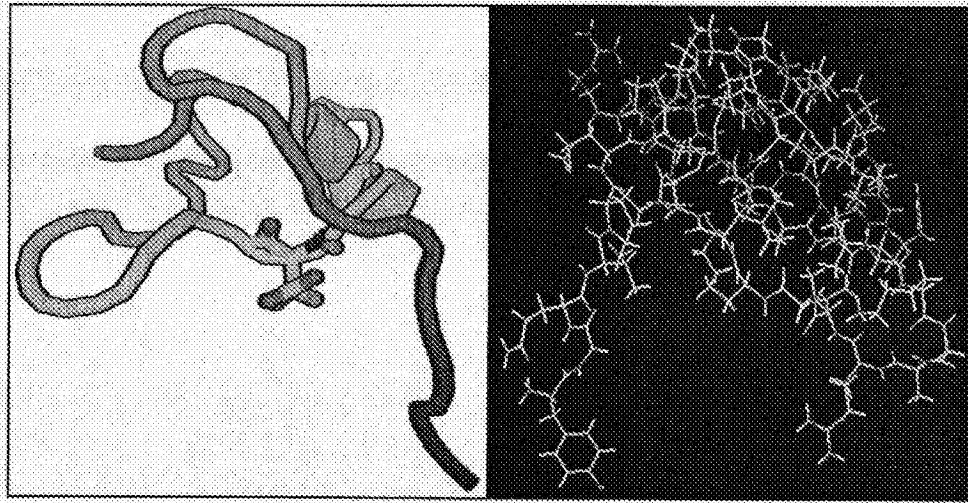


图3

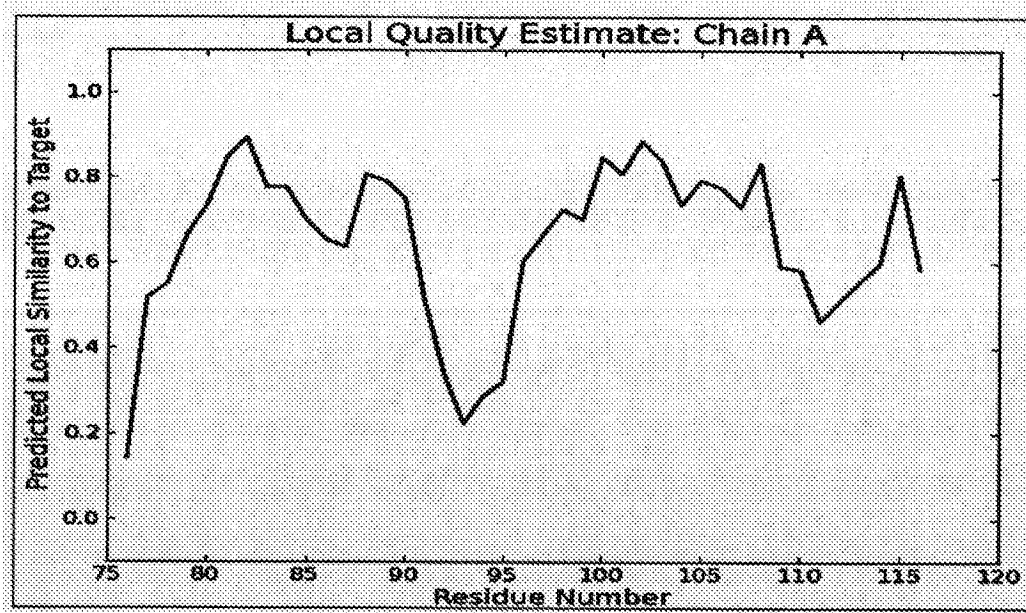


图4



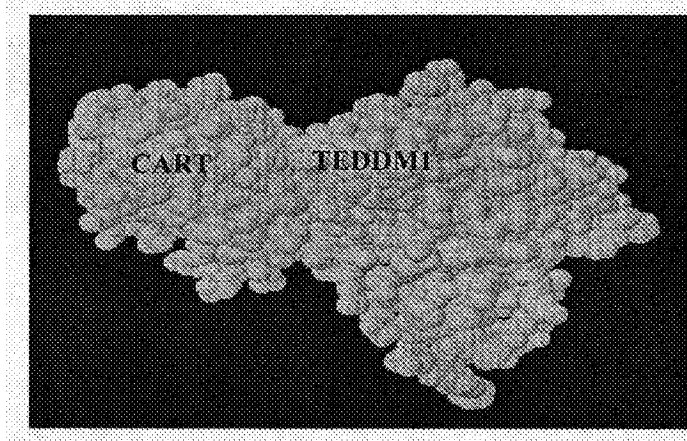


图5