

恒河猴 *PPARgamma* 基因的生物信息学分析

焦传珍

(德州学院生物系, 山东德州 253023)

摘要: 恒河猴 *Macaca mulatta* 是继黑猩猩 *Pan troglodytes* 后第 2 个完成基因组测序的非人类灵长类动物, 在生命科学领域中具有重要意义。本文利用生物信息学方法寻找与人类肥胖密切相关的 *PPARg* 基因在恒河猴中的同源基因, 对该基因的序列、外显子信息、编码蛋白及其理化性质进行分析, 并预测其编码蛋白的结构与功能, 构建其同源基因的系统进化树, 旨在为今后人类肥胖的相关研究提供一定的依据。

关键词: 恒河猴; *PPARgamma*; 生物信息学

中图分类号: Q959.848 文献标识码: A 文章编号: 1000-7083(2009)04-0545-03

Bioinformatic Analysis of the *PPARgamma* Gene from *Macaca mulatta*

JIAO Chuan-zhen

(Department of Biology, Dezhou University, Dezhou, Shandong Province 253023, China)

Abstract: The rhesus monkey is the second animal only the chimpanzee in its importance to life sciences, in particular in regards to determining their complete genome sequence. In this article, the *PPARgamma* gene, which was just discovered and is linked with human obesity, was examined. Its exons, coding proteins, and physical and chemical characteristics were analyzed. The structure and function of the proteins were forecasted, and a phylogenetic tree of the homologous gene was constructed. All of these will provide baseline information on human obesity for the future research.

Key words: *Macaca mulatta*; *PPARgamma*; bioinformatics

恒河猴 *Macaca mulatta* 是人类的近亲, 在形态结构、生理机能和生化代谢方面同人类非常相似, 成为生物医学和药物研究不可缺少的实验动物。测序结果显示恒河猴基因组与人类的基因组序列相似度为 92% ~ 95%, 而与黑猩猩 *Pan troglodytes* 基因组相似程度在 98% 以上。因此恒河猴为这 3 种亲缘关系相近的灵长类动物(人类、黑猩猩和恒河猴)的对比提供了一个理想的参考点(Heidi *et al.*, 2002)。

现代医学已证明, 肥胖与许多慢性非传染性疾病的发生和发展密切相关, 对肥胖的产生原因和预防的研究始终是医学和生物学的热点。*PPARg* 基因是使人体细胞转化为脂肪细胞的最关键的一个环节, *PPARg* 蛋白是一种细胞核激素受体, 负责调节细胞外信号的基因表达。*PPARg* 基因在前脂肪细胞转化成脂肪细胞的过程中起作用(金谷雷等, 2005), 高表达于脂肪组织, 而 *PPARg* 基因的活化伴随噻唑烷二酮类改变脂肪局部分布、脂肪细胞表型, 并且上调多种与脂肪酸代谢和甘油三酯蓄积基因的表达。因此, *PPARg* 基因激活受体改善脂肪组织功能, 对阻止糖尿病胰岛素抵抗的进展和动脉粥样硬化的内皮功

能紊乱的进展具有作用。本文利用生物信息学方法, 分析与人类肥胖密切相关的 *PPARg* 基因在恒河猴中的同源基因 *PPARgamma*, 并预测其结构和功能, 旨在通过对恒河猴 *PPARgamma* 基因的研究为人类肥胖的相关研究提供一定的依据。

1 分析方法

1.1 恒河猴 *PPARgamma* 基因结构分析

在 GenBank 数据库(<http://www.ncbi.nlm.nih.gov/>)中, 搜索人类 *PPARg* 基因在恒河猴中的同源序列, 分析其内含子和外显子信息。

1.2 恒河猴 *PPARgamma* 基因编码蛋白质的理化性质分析

利用互联网上提供的 ExPaSy (<http://us.Expasy.org/tools/>)软件包中的 Computer pI/MW 和 BioEdit 软件进行蛋白质的氨基酸组成、分子质量、等电点以及疏水性分析。

1.3 恒河猴 *PPARgamma* 基因编码蛋白质的结构分析

利用互联网 ExPaSy 数据库, 分析该蛋白质跨膜

结构域。利用 PUMA2 服务器的 SOPM 软件进行蛋白质序列的二级结构分析,利用 ExPaSy 的 3djigsaw 工具 (<http://www.bmm.icnet.uk/servers/3djigsaw/>) 向蛋白质立体结构数据库 PDB (Protein Data Bank) 提交该蛋白质序列,利用 RasMol 软件显示该蛋白的三维分子结构。

1.4 恒河猴 *PPARgamma* 基因编码蛋白质的功能分析

利用 Scansite (<http://scansite.mit.edu/>) 网站工具,分析功能结构域及磷酸化位点。

1.5 构建系统进化树

选取不同进化层次的典型物种相应的同源基因,构建系统进化树。

2 结果与分析

2.1 恒河猴 *PPARgamma* 基因结构分析

通过 NCBI 搜索数据库检索 (http://www.ncbi.nlm.nih.gov/projects/genome/guide/rhesus_macaque/), 发现恒河猴中与人类 *PPARg* 同源的基因为 *PPARgamma*。在 NCBI 上的核苷酸登录号为 NM_001032860, 蛋白质登录号为 NP_001028032, 此记录 NCBI 2007 年 7 月 27 日更新。该基因 mRNA 全长为 1759 bp, 编码全长为 505 aa 氨基酸序列。

恒河猴 *PPARgamma* 基因共有 6 个外显子, 各外显子含有的碱基数分别为 106 bp、229 bp、168 bp、137 bp、199 bp 和 450 bp (表 1)。

表 1 恒河猴 *PPARgamma* 基因外显子信息

序列号	Exon 1	Exon 2	Exon 3	Exon 4	Exon 5	Exon 6
NC_007859	1 ~ 106	30592 ~ 30821	32190 ~ 32358	43661 ~ 43798	57565 ~ 57764	68339 ~ 68789

2.2 恒河猴 *PPARgamma* 基因编码蛋白质的理化性质分析

利用 BioEdit 软件及互联网上工具进行氨基酸组成、分子质量、等电点分析结果显示, 蛋白质分子式为 $C_{2576}H_{4048}N_{670}O_{770}S_{27}$, 原子总数为 8091, 蛋白分

子质量为 57590.1, 理论等电点为 5.61。在氨基酸组成上 Leu 有 52 个, 占 10.3%, 含量最高; 其次, Lys 和 Ser 各有 39 个, 分别占 7.7%; Trp 含量只有 1 个, 占 0.2% (表 2)。

表 2 基因编码蛋白质的氨基酸组成

Amino Acid	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Number	28	10	37	34	27	21	14	36	39	52	17	16	25	24	19	39	25	23	1	18
Mol%	5.5	2.0	7.3	6.7	5.3	4.2	2.8	7.1	7.7	10.3	3.4	3.2	5.0	4.8	3.8	7.7	5.0	4.6	0.2	3.6

对恒河猴 *PPARg* 蛋白进行亲疏水性分析得知该蛋白有 1 个亲水性区域, 4 个疏水性区域。

2.3 恒河猴 *PPARgamma* 基因编码蛋白质结构分析

利用 Tmpred 服务器 (http://www.ch.embnet.org/software/TMPRED_form.html), 通过分析疏水区域预测蛋白质的跨膜区域以及方向, 输出结果包括可能的跨膜螺旋区、相关性列表、建议的跨膜拓扑模型。结果表明该蛋白质从内向外、由外向内分别含有两个可能的跨膜螺旋区, 其中从内向外螺旋的位置在 47 ~ 67、477 ~ 494; 由外向内螺旋的位置在 47 ~ 66、413 ~ 435、472 ~ 494。为了确认以上结果, 再利用 ExPaSy 的 TMHMM 工具预测该蛋白质的跨膜区, 此结果说明该蛋白不是膜蛋白, 无跨膜区 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)

利用 SOPM (<http://compbio.mcs.anl.gov/puma2/>) 软件工具, 预测基因编码蛋白质的二级结构,

α 螺旋 (Alpha helix) 有 196 个, 占 38.81%, β 折叠片 (Beta turn) 39 个, 占 7.72%, 延伸直链 (Extended strand) 有 105 个占 20.79%, 无规则卷曲 (Random coil) 有 165 个, 占 32.67% (图 1)。

利用 ExPaSy 服务器的 3djigsaw (<http://www.bmm.icnet.uk/servers/3djigsaw/>) 工具预测蛋白质的三维结构 (图 2), 二者存在细微差别。

2.4 恒河猴 *PPARgamma* 基因编码蛋白质的功能分析

利用 Scansite 分析 motif (<http://scansite.mit.edu/>), 采用默认参数, 选择 high Stringency, 寻找蛋白质序列中易于被特定的蛋白酶磷酸化的模体和绑定结构域。结果发现, 此蛋白不含同源结构域。其磷酸化位点分析发现 SH2 结构域: 在 Y78 位置发现 Fyn SH2 和 Shc SH2 序列, 在 Y355 位置发现 SHIP SH2 序列; Pro-ST_kin 结构域: 在 S112 位置发现 Erk1 Kinase 序列; Kin_bind 结构域: 在 D69 位置

