



本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

— 包括说明书序列表部分(细则 5.2(a))。

磁感应受体蛋白和其复合物及它们的用途

技术领域

5 本发明涉及生物学和其应用领域。具体地，本发明涉及一种科学上首次研究和制备得到的自身具有磁性的蛋白，即磁感应受体蛋白，以及所述磁感应受体蛋白的纯合蛋白复合物，以及所述磁感应受体蛋白和光感应蛋白形成的杂合蛋白复合物。本发明还提供了所述蛋白或复合物的用途。

背景技术

10 铁是生命体必需的微量元素，和铁结合、转运相关的蛋白质也得到了较多的关注和研究。铁结合蛋白是含铁的金属结合蛋白，其中有一个亚类称之为铁硫蛋白，其结合的铁是以铁硫中心（Iron-sulfur cluster）的形式存在。铁硫蛋白是非常重要的电子传递的载体，主要牵涉到呼吸链和植物的光合作用等生化反应过程。对铁结合蛋白和铁硫蛋白的研究非常广泛和深入，主要集中在和生化代谢相关的各个方面。生物中最重要的两个铁结合蛋白是乳铁蛋白和转铁蛋白，其中转铁蛋白构成的复合笼状结构，能装载大约 3000-5000 个铁原子，是生物中重要的储存铁的机制。一些研究也表明，可以利用装载铁原子的转铁蛋白作为纳米颗粒用于生物医学和其它领域的应用。然而，迄今为止，所有已经发现的铁结合蛋白都不具备自身的磁性，即都不是“磁性蛋白”。例如，上述载铁能力极高的转铁蛋白在磁场下能够被磁场吸引，具备顺磁的特性，但也不具备自身的磁性。该蛋白具有类似于铁球在磁场下的反应，但因为不具备自身的磁性，所以不能作为如同用作指南针的磁铁那样的磁性材料使用。

20 另外，动物的磁感应能力一直是生物学领域中最引人注目的未解之谜。很多年以来，人们已经知道许多动物在进行长距离迁徙时是通过感应地磁场来判断方向和导航的，例如黑脉金斑蝶、家鸽、蜜蜂、龙虾、海龟、鲨鱼、蝙蝠、牛羚羊以及众所周知的候鸟等等。除此之外，还有一些物种利用地磁场来引导自身的朝向和筑巢的走向，例如名为 *Tritonias diomedea* 的海洋软体动物、指南白蚁、裸鼹鼠等。然而，人们对动物体感应磁场的分子机理了解甚少。

25 动物能够通过某种分子机制（例如，通过某种特定的基因或者蛋白质，或者通过某条细胞信号通路）来感应地磁场的这一说法，从无人相信到被学界认可，经历了很长的一段时间。迄今为止，与动物体感应磁场有关的最关键的一个研究是隐花色素蛋白（Cryptochrome，简称为 Cry）的发现和研究。2008 年，在 Gegear, R. J., Casselman, A., Waddell, S. & Reppert, S. M. 等的 Cryptochrome mediates light-dependent magnetosensitivity in *Drosophila*. *Nature* 454, 1014-1018 (2008) 中描述了将模式动物果蝇的 Cry 基因敲除后，果蝇就失去了对磁场的感知能力。这一实验充分说明了动物对磁场的感知和利用磁场导航存在某种分子机制，并且这一过程由特定的基因和蛋白质参与。这表明，动物具有磁感应能力不是伪科学，而是由特定基因决定的，并证明 Cry 蛋白是动物磁感应信号通路中非常关键的一个环节。但是，该发现并不能确认 Cry 蛋白是直接感应磁场的受体蛋白。事实上，其后有很多该领域的相关文献也进一步说明了这一点，如 Gould, J. L. 的 Magnetoreception. *Curr Biol* 20, R431-435

(2010), 其指出当前的几种假说都有缺陷, 尤其是 Cry 作为磁感应蛋白的假说并不能解释蛋白质如何识别磁场的磁极, 而 Cry 蛋白质在磁场下的反应并没有任何直接的实验证据支持。值得注意的是, Cry 蛋白并不是一种含铁的蛋白质, 因此 Cry 蛋白不能提供其具有感磁特性的让人信服的机理。

- 5 本领域尚未发现本身具有磁性和/或光磁性的蛋白或蛋白复合物。本领域亦未得到过本身具有磁性和/或光磁性的蛋白或蛋白复合物以及将其应用于人类生产生活中。

发明概述

- 10 本发明涉及一种在现代生物学研究中新研究和制备得到的自身具有磁性和/或能够形成具有磁性之复合物的蛋白, 即磁感应受体蛋白, 以及所述磁感应受体蛋白的纯合蛋白复合物, 以及所述磁感应受体蛋白和光感应蛋白形成的杂合蛋白复合物。本发明还提供了所述磁感应受体蛋白或复合物的用途。

- 15 本发明提供了一种分离的磁感应受体蛋白。磁感应受体蛋白 (或称“Magnetoreceptor”, 或称“MagR”) 是本申请发明人在现代生物科学研究中首个研究、鉴定并命名的一种其核酸编码序列在生物体基因组中广泛存在并高度保守, 可形成具有生物磁性的蛋白单体或纯合复合物或杂合复合物的蛋白质。本发明提供的磁感应受体蛋白的核酸序列在大量不同物种的基因组中都存在, 其编码的蛋白序列高度保守, 并具有多个高度保守的结构域, 例如特征性的结合铁和/或硫的结构域。如按照本申请图 13 所示的氨基酸位置, 本发明的磁感应受体蛋白的氨基酸序列从 N 端到 C 端依次包含以下结构域: 第 10-13 位的 TVRA (SEQ ID NO: 17)、第 56-61 位的 RGCNGL (SEQ ID NO: 18)、第 80-83 位的 QDGV (SEQ ID NO: 19)、第 89-93 位的 KKAQL (SEQ ID NO: 20)、第 95-101 位的 LLGTEMD (SEQ ID NO: 21) 以及第 115-128 位的 NPNIKGTCGCGESF (SEQ ID NO: 22)。

- 25 进一步地, 本发明的磁感应受体蛋白的氨基酸序列从 N 端到 C 端依次还包含以下结构域: 第 17 位的 R、第 23 位的 R、第 25-26 位的 AL、第 28-29 位的 LT、第 32-33 位的 AV、第 48 位的 G、第 51 位的 V、第 53 位的 V、第 63 位的 Y、第 67 位的 Y、第 71 位的 K、第 75-76 位的 DE、第 78 位的 V、第 87 位的 I、第 104 位的 E、第 106-107 位的 KL、第 110-111 位的 EF 和第 113 位的 F。本领域技术人员应该明白, 图 13 所示的氨基酸位置是将多条序列进行比对后所得的、能够对应大部分磁感应受体蛋白的氨基酸实际位置的结果; 当有新的磁感应受体蛋白发现后, 本领域技术人员能够很容易地通过比对的方式获知新蛋白中与图 13 所示的氨基酸位置相对应的位置。

更进一步地, 本发明的磁感应受体蛋白或其活性片段或衍生物具有 2Fe2S 的铁硫中心。

- 35 本发明的磁感应受体蛋白可为磁感应受体蛋白家族中的其中一种。本发明的磁感应受体蛋白可来自动物, 如脊椎动物或无脊椎动物。示例性的 MagR 的核酸或蛋白序列为本申请发明人从家鸽 (拉丁名 *Columba livia*) 和黑脉金斑蝶 (或称美洲帝王蝶, 拉丁名 *Danaus plexippus*) 得到, 其氨基酸序列分别如 SEQ ID NO: 1 和 SEQ ID NO: 2 所示。这些序列在本发明之前是未被公开的。

本发明提供的磁感应受体蛋白还包括其它生物物种的磁感应受体蛋白，例如来自果蝇(Fruit fly, 拉丁名 *Drosophila melanogaster*)，蜜蜂(Honey bee, 拉丁名 *Apis mellifera*)，文昌鱼(Lancelet, 拉丁名 *Branchiostoma floridae*)，斑马鱼(Zebrafish, 拉丁名 *Danio rerio*)，蟒蛇(Burmese python, 拉丁名 *Python bivittatus*)，草雀(Zebra finch, 拉丁名 *Taeniopygia guttata*)，鸡(Chicken, 拉丁名 *Gallus gallus*)，鲸鱼(Minke whale, 拉丁名 *Balaenoptera acutorostrata*)，蝙蝠(Little brown bat, 拉丁名 *Myotis lucifugus*)，裸鼹鼠(Naked mole rat, 拉丁名 *Heterocephalus glaber*)，小鼠(Mouse, 拉丁名 *Mus musculus*)，人(Human, 拉丁名 *Homo sapiens*)。上述来自果蝇、蜜蜂、文昌鱼、斑马鱼、蟒蛇、草雀、鸡、鲸鱼、蝙蝠、裸鼹鼠、小鼠和人的磁感应受体蛋白的氨基酸序列分别如 SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 和 SEQ ID NO: 14 所示。

在本发明的其中一个方面，提供了一种分离的磁感应受体蛋白或其活性片段或衍生物，其具有：

(1) 如 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 所示的氨基酸序列，或

(2) 在与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 进行比较时具有至少 50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 的序列同一性的氨基酸序列。其中所述序列同一性可通过 Clustal V 比对方法计算得到。

优选地，其中不相同的残基位置的差异为保守性氨基酸取代。“保守性氨基酸取代”指其中氨基酸残基被另一种具有化学性质(例如：电荷或亲水性)类似之侧链 R 基的氨基酸残基取代。通常，保守性氨基酸取代实质上不会改变蛋白质的功能性质。具有相似化学性质之侧链的氨基酸基团实例包括：1) 脂肪族侧链：甘氨酸、丙氨酸、缬氨酸、亮氨酸与异亮氨酸；2) 脂肪族-羟基侧链：丝氨酸与苏氨酸；3) 含酰胺侧链：天冬酰胺与谷氨酰胺；4) 芳香族侧链：苯丙氨酸、酪氨酸与色氨酸；5) 碱性侧链：赖氨酸、精氨酸与组氨酸；6) 酸性侧链：天冬氨酸与谷氨酸；及 7) 含硫侧链：半胱氨酸与甲硫氨酸。优选的保守性氨基酸取代组合为：缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸、谷氨酸-天冬氨酸、及天冬酰胺-谷氨酰胺。

在本发明的其中一个方面，提供了一种分离的磁感应受体蛋白或其活性片段或衍生物，其具有：在与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO:

4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 进行比较时具有一个或多个氨基酸的取代、缺失、添加的氨基酸序列。例如，其中取代、缺失、添加的氨基酸为保守性氨基酸。本发明提供的具有所述取代、缺失、添加的磁感应受体蛋白具有磁感应受体蛋白活性。

在本发明的其中又一个方面，上述分离的磁感应受体蛋白或其活性片段或衍生物具有磁性和/或能够形成具有磁性的蛋白复合物。在本发明的其中又一个方面，上述分离的磁感应受体蛋白或其活性片段或衍生物具有结合铁和/或硫的结构域。在本发明的其中又一个方面，上述分离的磁感应受体蛋白或其活性片段或衍生物能形成具有磁性的纯合蛋白复合物，和/或能与隐花色素蛋白（Cryptochrome，简称为 Cry）形成杂合蛋白复合物。在本发明的其中又一个方面，上述纯合蛋白复合物或杂合蛋白复合物具有特定的蛋白立体结构，例如长条形；在所述纯合蛋白复合物或杂合蛋白复合物中，具有多个铁原子，其在棒状结构蛋白质的内部近中心轴区域也形成有序排列，由此使得所述蛋白复合物具有磁性。

在本发明的其中一个方面，本发明提供了一种重组 DNA 构建体，其包含编码如前所述的磁感应受体蛋白或其活性片段或衍生物的核苷酸序列。在所述构建体中，所述编码如前所述的磁感应受体蛋白或其活性片段或衍生物的核苷酸序列可操作地连接至少一个调控元件，例如可在真核细胞或原核细胞中有功能的启动子。

本发明提供了一种分离的蛋白复合物，其包含两个或两个以上前述的本发明的磁感应受体蛋白或其活性片段或衍生物，即纯合蛋白复合物。

在本发明的其中一个方面，所述蛋白复合物含有 4 个或以上所述磁感应受体蛋白或其活性片段或衍生物，优选为 20 个或以上所述磁感应受体蛋白或其活性片段或衍生物，进一步优选为 20-24 个所述磁感应受体蛋白或其活性片段或衍生物，更优选为 20 个所述磁感应受体蛋白或其活性片段或衍生物。

在本发明的其中一个方面，所述蛋白复合物的立体结构为“长条形”。蛋白质立体结构是指蛋白质分子的空间结构。长条形的蛋白立体结构也可称为“棒状结构”或“棍状结构”。蛋白质的立体结构可以通过电子显微镜观察和测量。在本发明的蛋白复合物中，所述多个磁感应受体蛋白或其活性片段或衍生物发生线性聚合，并进而聚集成棒状结构。在本发明的其中一个方面，所述多个磁感应受体蛋白或其活性片段或衍生物发生线性聚合形成两条蛋白长链（等长的蛋白长链，即两条蛋白单链中含有相同个数的蛋白单体）。在本发明的其中又一个方面，所述两条蛋白长链以螺旋形式结合，形成棒状结构的蛋白复合物。在本发明的所述蛋白复合物中还包含多个铁原子。在本发明的其中一个方面，所述多个铁原子位于所述棒状结构蛋白复合物的内部近中心轴区域。例如，每个形成所述蛋白复合物的单个磁感应受体蛋白都结合铁原子，又例如，每个磁感应受体蛋白单体结合铁原子的位置都相同或相近，由此，在由磁感应受体蛋白单体线性聚合形成的棒状结构蛋白复合物中，所述多个铁原子在棒状结构的内部近中心轴区域也形成有序排列。

在本发明的其中一个方面，所述蛋白复合物中每个磁感应受体蛋白或其活性片段或衍生物含有 1 个 2Fe2S 的铁硫中心，即每个磁感应受体蛋白含有 2 个铁原子和 2

个硫原子。在本发明的其中又一个方面，所述蛋白复合物含有约 20 个磁感应受体蛋白，由此，所述蛋白复合物含有约 40 个铁原子。

在本发明的其中一个方面，前述蛋白复合物还可结合有光感应元件。

因此，本发明还提供了一种分离的杂合蛋白复合物，其包含 1) 前述的本发明的磁
5 感应受体蛋白或其活性片段或衍生物，以及 2) 光感应蛋白（例如隐花色素蛋白）或其活性片段或衍生物。

光感应元件是可感受/传递光信号的部分，如可结合到蛋白上的分子、复合物或蛋白等，例如可以是可感受/传递光信号的蛋白。在生物界已发现多种可感受/传递光信号的蛋白。其中一种是隐花色素蛋白（Cryptochrome，简称为 Cry）或其活性片段
10 或衍生物。在本发明的其中又一个方面，在所述杂合蛋白复合物中，每 2 个磁感应受体蛋白与 1 个隐花色素蛋白结合。

隐花色素蛋白（Cryptochrome，简称为 Cry）是一类能够感受蓝光（400~500nm）和近紫外光（320~400nm）的黄素蛋白（Flavoprotein），分子量为 70~80KD，生色团可为黄素腺嘌呤二核苷酸（FAD）和蝶呤（pterin）。Cry 蛋白及其编码基因普遍
15 存在于植物、动物以及整个高等真核生物中。Cry 蛋白在动物中可能与动物时钟节律和生物钟的控制有关。Cry 蛋白家族包括 Cry、Cry1、Cry2、Cry3 和 Cry4 等。在哺乳动物、昆虫体内都有其同源基因编码隐花色素蛋白。

在本发明的其中一个方面，在本发明的蛋白复合物中，所述隐花色素蛋白可选自 Cry、Cry1、Cry2、Cry3 或 Cry4。在本发明的其中又一个方面，在本发明的蛋白复
20 合物中，所述隐花色素蛋白具有：

（1）如 SEQ ID NO: 15 或 SEQ ID NO: 16 所示的氨基酸序列，或

（2）在与 SEQ ID NO: 15 或 SEQ ID NO: 16 进行比较时具有至少 50%、51%、
52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、
65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、
25 78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、
91%、92%、93%、94%、95%、96%、97%、98%、99% 或 100% 的序列同一性的氨基酸序列，并且具有隐花色素蛋白的活性。其中所述序列同一性可通过 Clustal V 比对方法计算得到。

优选地，其中不相同的残基位置的差异为保守性氨基酸取代。

在本发明的其中又一个方面，所述分离的蛋白复合物中的隐花色素蛋白具有：在
30 与 SEQ ID NO: 15 或 SEQ ID NO: 16 进行比较时具有一个或多个氨基酸的取代、缺失、添加的氨基酸序列。例如，其中取代、缺失、添加的氨基酸为保守性氨基酸。本发明提供的具有所述取代、缺失、添加的磁感应受体蛋白具有磁感应受体蛋白活性。

35 在本发明的其中一个方面，所述杂合蛋白复合物中含有铁硫原子。

在本发明的其中一个方面，所述杂合蛋白复合物包含 4 个或以上、优选 20 个或以上所述磁感应受体蛋白或其活性片段或衍生物，进一步优选为 20-24 个、更优选为 20 个所述磁感应受体蛋白或其活性片段或衍生物。相应地，所述杂合蛋白复合物中包含 2 个或以上、优选 10 个或以上所述隐花色素蛋白，进一步优选为 10-12 个、更

优选为 10 个所述隐花色素蛋白。

在本发明的其中一个方面，前述本发明的蛋白复合物或杂合蛋白复合物具有磁性。“蛋白质具有磁性”或者“磁性蛋白质”是指蛋白质或者蛋白质复合物作为磁性材料的主体存在，自身具备磁性。所述磁性蛋白质可具有类似磁铁的磁极（南北极），
5 可以被外源的磁场牵引运动，也具备吸引铁、钴、镍等铁磁性物质的性质。值得指出的是，本发明所述的蛋白质具有自身的磁性，其磁性不依赖于外界磁场而存在。

在本发明的其中一个方面，前述的本发明蛋白复合物或杂合蛋白复合物可按照本领域的常规方法制备成为晶体。

本发明还提供了生产前述的本发明蛋白复合物的方法，其包含下述步骤：

10 在允许形成蛋白复合物的条件下，使所述磁感应受体蛋白或其活性片段或衍生物聚合，形成同源聚合物；分离该同源聚合物。

必要时，按照本领域的常规技术使该同源聚合物结晶。

在上述生产本发明的蛋白复合物的方法中，生产所述磁感应受体蛋白的环境中具有可生物利用的铁元素和硫元素。

15 在本发明的其中一个方面，在生产前述的本发明蛋白复合物的方法中，其中所述磁感应受体蛋白或其活性片段或衍生物在细胞中表达和形成复合物。所述细胞可以是培养的原核或者真核动物细胞，例如大肠杆菌或者哺乳动物细胞。在本发明的其中一个方面，在培养所述细胞过程中，所述磁感应受体蛋白或其活性片段或衍生物含有铁和/或硫原子，因此培养液中需加入可生物利用的铁元素和/或硫元素（单体或
20 化合物）。

在本发明的其中一个方面，生产前述的本发明蛋白复合物的方法可在存在磁场的条件下进行。例如，其中分离同源聚合物的步骤可在存在磁场的条件下进行。

在本发明的其中又一个方面，所述磁场为外加磁场。在本发明的其中又一个方面，分离蛋白质同源聚合物通过亲和层析法进行，例如通过亲和层析柱；在采用亲和层析法分离所述蛋白质时，提供磁场条件。在本发明的其中一个方面，在生产前述的本发明蛋白复合物的方法中，所述磁场的磁场强度大于约 5 倍地球磁场，例如是地球磁场的约 10-100 倍。在生产前述的本发明蛋白复合物的方法中，所述磁场的磁场强度大于约 100 倍地球磁场也可以获得良好的效果。

同样地，本发明还提供了生产前述的本发明杂合蛋白复合物的方法，其包含下述
30 步骤：

在允许形成杂合蛋白复合物的条件下，使 1) 前述的本发明的磁感应受体蛋白或其活性片段或衍生物与 2) 前述的光感应蛋白（例如隐花色素蛋白）或其活性片段或衍生物接触和聚合，形成杂合聚合物，其中 1)和 2)的蛋白中的至少一种是分离形式或被重组表达；和分离该杂合聚合物。

35 必要时，按照本领域的常规技术使该杂合聚合物结晶。

在本发明的其中一个方面，在生产前述的本发明杂合蛋白复合物的方法中，其中磁感应受体蛋白和光感应蛋白在细胞中表达和形成复合物。所述细胞可以是培养的原核细胞或者真核动物细胞，例如大肠杆菌和哺乳动物细胞。在本发明的其中一个方面，在培养所述原核或真核动物细胞过程中，所述磁感应受体蛋白或其活性片段

或衍生物含有铁和/或硫原子，因此培养液中需加入可生物利用的铁元素和/或硫元素（单体或化合物）。

在本发明的其中一个方面，生产前述的本发明杂合蛋白复合物的方法在存在磁场的条件下进行。例如，其中分离该杂合复合物的步骤在存在磁场的条件下进行。在本发明的其中又一个方面，所述磁场为外加磁场。在本发明的其中又一个方面，分离杂合蛋白复合物通过亲和层析法进行，例如通过亲和层析柱；在采用亲和层析法分离所述蛋白质时，可提供磁场条件。在本发明的其中一个方面，在生产前述的本发明杂合蛋白复合物的方法中，所述磁场的磁场强度大于约 5 倍地球磁场，例如是地球磁场的约 10-100 倍。在生产前述的本发明杂合蛋白复合物的方法中，所述磁场的磁场强度大于约 100 倍地球磁场也可以获得良好的效果。

在本发明的其中一个方面，提供了鉴别前述本发明磁感应受体蛋白可形成之光磁感应复合物的方法，其包含下述步骤：

a. 在允许形成蛋白复合物的条件下，使前述的本发明的磁感应受体蛋白或其活性片段或衍生物与待测蛋白或衍生物接触；

b. 测定所述磁感应受体蛋白或其活性片段或衍生物与待测蛋白或衍生物是否形成了复合物；

c. 测定所形成的复合物内 1) 的蛋白与待测蛋白之间的相互作用。

在本发明的一个方面，提供了前述的本发明磁感应受体蛋白或蛋白复合物或杂合蛋白复合物的应用。

在本发明的其中一个方面，提供了前述的本发明磁感应受体蛋白和/或蛋白复合物和/或杂合蛋白复合物的与磁性相关的应用，例如，其作为磁性材料的应用或者制备这些磁性材料的应用。本发明的磁感应受体蛋白和/或蛋白复合物和/或杂合蛋白复合物具有对磁场的高度敏感，更特殊的是，其存在内在的磁性，可作为各种与磁性相关的应用中的磁性材料。与磁性相关的应用的领域非常广泛，包括但不限于生物医药和医疗器械如磁性生物材料、生物分子提纯、细胞内物质的定向移动、通过磁场控制某些细胞生命活动等，磁记录介质，超导应用如纳米环形电流超导体、磁流变流体等，电器设备如电磁转换设备，通讯和导航设备如分子陀螺仪中的应用等。

本文所引用的全部出版物都并入本文作为参考。除非另有定义，在本文中使用的全部技术和科学术语都具有与本发明所属领域的普通技术人员一般理解的相同的含义。

术语“多肽”、“蛋白”和“肽”在本文中互换地用于指氨基酸链，其中氨基酸残基通过肽键或修饰的肽键相连。氨基酸链可以是大于 2 个氨基酸的任意长度。除非另有说明，术语“多肽”、“蛋白”和“肽”也包括它们的各种修饰形式。这些修饰形式可以是天然存在的修饰形式或化学修饰形式。修饰形式的实例包括但不限于：糖基化形式、磷酸化形式、核糖基化形式、乙酰化形式、泛素化形式等。修饰也包括分子内交联和与各种部分（例如脂质、黄素、生物素、聚乙二醇或其衍生物等）的共价附着。另外，修饰也可以包括环化、分支和交联。此外，在多肽中也可以包含由基因密码子编码的 20 个常见氨基酸以外的氨基酸。

“磁感应受体蛋白”（或称“Magnetoreceptor”，或称“MagR”）是本申请发明人在

现代生物科学研究中首个研究、得到、鉴定并命名的一种其核酸编码序列在生物体基因组中广泛存在并高度保守，可形成具有生物磁性单体或复合物的蛋白质。本发明提供的磁感应受体蛋白的核酸序列在大量不同物种的基因组中都存在，其编码的蛋白序列高度保守，并具有多个高度保守的结构域，如结合铁和/或硫的结构域。磁感应受体蛋白能形成具有特定形态结构、具有自身磁性的纯合蛋白复合物，和/或能与隐花色素蛋白（Cryptochrome，简称为 Cry）形成杂合蛋白复合物。“蛋白家族”是指生物物种中存在的一系列具有同源性结构域或序列、进化上相关的、功能相同或相似的蛋白质。磁感应受体蛋白的蛋白家族中包括家鸽(*Columba livia*)磁感应受体蛋白和黑脉金斑蝶(*Danaus plexippus*)磁感应受体蛋白，其氨基酸序列分别如 SEQ ID NO: 1 和 SEQ ID NO: 2 所示。磁感应受体蛋白的蛋白家族的其它成员包括果蝇(*Drosophila melanogaster*)、蜜蜂(*Apis mellifera*)、文昌鱼 (*Branchiostoma floridae*)、斑马鱼(*Danio rerio*)、蟒蛇(*Python bivittatus*)、草雀(*Taeniopygia guttata*)、鸡(*Gallus gallus*)、鲸鱼(*Balaenoptera acutorostrata*)、蝙蝠(*Myotis lucifugus*)、裸鼯鼠(*Heterocephalus glaber*)、小鼠(*Mus musculus*)、人 (*Homo sapiens*)的磁感应受体蛋白，其氨基酸序列分别如 SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 所示。磁感应受体蛋白的纯合蛋白复合物或磁感应受体蛋白与隐花色素蛋白形成的杂合蛋白复合物具有特定的蛋白立体结构，例如长条形；在所述纯合蛋白复合物或杂合蛋白复合物中，具有多个铁原子，其在棒状结构蛋白质的内部近中心轴区域也形成有序排列，由此使得所述蛋白复合物具有磁性。例如，多个所述磁感应受体蛋白发生线性聚合，并进而聚集形成棒状结构。其中，多个所述磁感应受体蛋白线性聚合形成两条蛋白长链；所述两条蛋白长链以螺旋形式结合，形成棒状结构的蛋白复合物。磁感应受体蛋白的纯合蛋白复合物或磁感应受体蛋白与隐花色素蛋白形成的杂合蛋白复合物中还包含多个铁原子；所述多个铁原子位于所述棒状结构蛋白复合物的内部近中心轴区域。例如，每个形成所述蛋白复合物的单个磁感应受体蛋白都结合铁原子，又例如，每个磁感应受体蛋白单体结合铁原子的位置都相同或相近，由此，在由磁感应受体蛋白单体线性聚合形成的棒状结构蛋白复合物中，所述多个铁原子在棒状结构的内部近中心轴区域也形成有序排列。本文所述的中心轴具有通常理解的含义。“内部近中心轴区域”是指位于棒状结构内部、环绕中心轴的地方，多个铁原子在该区域排列可形成环状电子流。

物质的“磁性”是指能够产生磁场、具有吸引铁磁性物质（如铁、镍、钴等金属）的特性。

“蛋白质具有磁性”或者“磁性蛋白质”是指蛋白质或者蛋白质复合物自身具备磁性，体现出类似磁铁的磁极（南北极），可以被外源的磁场牵引运动，也具备吸引铁、钴、镍等铁磁性物质的性质，可作为磁性材料。

“光磁性”是指物质受到光照后，磁学性质(如磁化率、磁晶各向异性、磁滞回线等)发生变化的现象，包括基于对光信号和磁场信号中的一种感应引发另一种基于光或磁场的作用。

本文使用的“磁感应受体蛋白的活性片段”是指磁感应受体蛋白的部分或片段，其

包含磁感应受体蛋白中维持它以类似于全长磁感应受体蛋白的方式并具有磁感应受体蛋白全部活性或部分活性的部分，或为具有该蛋白的多种活性中的某种或某几种活性的片段，例如其结构域片段。

本文使用的“磁感应受体蛋白的衍生物”是指，具有与本发明的磁感应受体蛋白的天然形式类似的一级和/或三级结构，但与所述天然形式相差一个或多个氨基酸残基（例如一个或多个氨基酸取代、插入和/或缺失）的蛋白（多肽），或是具有/额外具有其它可以与之结合的化学基团或蛋白（多肽）基团；这些衍生物具有或保持磁感应受体蛋白的全部或部分活性。所述可结合的化学基团包括例如多肽的翻译后的衍生化或修饰，例如 PEG 化和/或巯基基团。所述可结合的蛋白（多肽）基团包括例如 His 标签或 IgG 蛋白等。

术语“分离的”用于描述这样的物质（例如核酸和/或蛋白质），该物质不同于自然界中其天然存在或在原始细胞或生物环境中发现的形式物质。该物质基本上不含在天然存在的环境中通常伴随该物质或与其反应的组分，或者说，该物质已从所述组分中移出。

“分离的蛋白”是这样的蛋白：它基本上与蛋白的天然来源中存在的至少一种组分或其它蛋白分离，或当化学合成所述蛋白时，它基本上不含有至少一种化学前体或其它化学药品。当存在小于约 30%、20%、10%或 5%(按干重计)的其它蛋白或其它化学药品(在本文中也称作“污染蛋白”或“污染化学药品”)时，在蛋白的制剂中蛋白“基本上分离于”或“基本上不含有”其它蛋白或其它化学药品。

分离的蛋白可以具有几种不同的物理形式。分离的蛋白可以作为全长新生的或未加工的多肽，或者作为部分加工的多肽，或者作为加工的多肽，或者它们的组合存在。

分离的多肽可以是非天然存在的多肽。例如，“分离的多肽”可以是“杂种多肽（hybrid peptide）”。“分离的多肽”也可以是通过氨基酸的添加、缺失或取代从天然存在的多肽衍生的多肽。分离的多肽也可以是“纯化的多肽”，后者在本文中用于指基本上均质的制剂中基本上不含有其它细胞组分、其它多肽、病毒材料或培养基的特定多肽，或当化学合成所述多肽时，基本上不含有化学前体或与化学合成有关的副产物。“纯化的多肽”可以通过标准的纯化技术或通过化学合成从天然或重组宿主细胞得到，如技术人员显而易见的。

如本文所使用的，术语“相互作用”是指，2 个蛋白结构域、片段或完整蛋白彼此表现出足够的物理亲和力，以便使所述 2 个“相互作用”蛋白结构域、片段或蛋白彼此物理上接近。相互作用可以来自一个或多个化学键的形成，其导致 2 个相互作用实体的连续且稳定的接近。相互作用也可以仅仅基于物理亲和力，这在共同定位(co-localize)2 个蛋白方面同样有效。物理亲和力和化学键的实例包括但不限于，由电荷差异造成的力、疏水性、氢键、范德华力、离子力、共价键以及它们的组合。相互作用域、片段或蛋白之间的接近状态可以是瞬时的或永久的，可逆的或不可逆的。在任何情况下，其与 2 个实体的天然随机运动造成的接触形成对照且可以区分开。通常，通过相互作用域、片段或蛋白之间的结合，表现出“相互作用”。相互作用的实例包括抗原和抗体、配体和受体、酶和底物等之间的特异性相互作用。

如本文所使用的，术语“蛋白复合物”或“多肽复合物”是指一种复合单位，它是通过蛋白之间的相互作用形成的 2 个或更多个蛋白的组合。通常，通过特异性非共价结合亲和力将 2 个或更多个蛋白结合到一起，形成“蛋白复合物”。“纯合蛋白复合物”是指由同一种蛋白质组分通过聚合（或称之为“同源聚合”）形成的多聚物。“杂合蛋白复合物”是指由两种或者多种不同的蛋白质组分通过聚合（或称之为“异源聚合”）形成的多聚物。

术语“分离的蛋白复合物”是指在不同于其在自然界中天然存在或在原始细胞或生物环境中发现的形式的蛋白复合物。该蛋白复合物基本上不含在天然存在的环境中通常伴随该蛋白复合物或与其反应的组分，或者说，该蛋白复合物已从所述组分中移出。

“分离的蛋白复合物”也可以是未在自然界发现的蛋白复合物。

“分离的”核酸分子是这样的，它基本上与核酸的天然来源中存在的至少一种其它核酸分子分离，或当化学合成所述核酸分子时，它基本上不含有至少一种化学前体或其它化学药品。“分离的”核酸分子也可以是，例如，基本上不含有在该核酸的来源生物基因组 DNA 中在 5'和 3'末端天然侧接该核酸分子的至少一个核苷酸序列的核酸分子。在核酸分子的制剂中，当存在小于约 30%、20%、10%或 5%(按干重计)的其它核酸分子或其它化学药品(在本文中也称作“污染核酸分子”或“污染化学药品”)时，核酸分子“基本上分离于”或“基本上不含有”其它核酸分子或其它化学药品。

“序列”是指聚合物中单体出现的线性次序，例如，多肽中氨基酸的次序或多核苷酸中核苷酸的次序。“核苷酸序列”指单链或双链形式的聚合物中脱氧核糖核苷酸或核糖核苷酸残基的排列。核酸序列可以由包含下述碱基的天然核苷酸组成：胸腺嘧啶、腺嘌呤、胞嘧啶、鸟嘌呤和尿嘧啶；分别缩写为 T、A、C、G 和 U，和/或所述天然核苷酸的合成类似物。

核苷酸序列或氨基酸序列比对和序列的同一性百分比可用设计用于鉴别同源序列的多种比较方法来确定，这些方法包括但不限于 LASERGENE[®] 生物信息计算包 (DNASTAR[®] Inc., Madison, WI) 的 Megalign[®] 程序。除非另外说明，否则本文提供的序列多重比对用 Clustal V 比对方法 (Higgins 和 Sharp, 1989, CABIOS. 5: 151-153) 采用默认参数 (空位罚分=10, 空位长度罚分=10) 执行。用 Clustal V 方法进行成对比对和蛋白质序列的同一性百分比计算的默认参数为 KTUPLE=1、空位罚分=3、窗口 (WINDOW)=5 和 DIAGONALS SAVED=5。而对于核酸，这些参数为 KTUPLE=2, 空位罚分=5, 窗口=4 和 DIAGONALS SAVED=4。用 Clustal V 程序比对序列后，可通过查看同一程序中的“序列距离”表来获得“同一性百分比”和“趋异”值。

“重组 DNA 构建体”指在自然界中通常不会一起存在的核酸片段的组合。因此，重组 DNA 构建体可包含源于不同来源的调控序列和编码序列，或源于相同来源但以不同于通常天然存在的方式排列的调控序列 (如启动子等) 和编码序列。

分子生物学和蛋白-蛋白相互作用领域的技术人员使用常规实验，可以在磁感应受体蛋白或其活性片段或衍生物中导入氨基酸序列变异。例如，本领域技术人员可以导入位点特异性的突变，例如将“CTT”密码子变成“ATT”密码子，从而造成天然多肽中的亮氨酸残基被替换为合成同源物中的异亮氨酸残基。

在另一个一般方面，本发明提供了制备本发明蛋白复合物的方法。本发明的蛋白复合物可以通过多种方法来制备。具体而言，蛋白复合物可以直接分离自含有该蛋白复合物的动物组织样品，例如人组织样品。蛋白复合物也可以分离自重组表达蛋白复合物成员的宿主细胞。或者，可以通过组合蛋白复合物的个别成员来体外构建蛋白复合物。

本领域技术人员会明白，为了表达蛋白或蛋白复合物的目的，在本发明中可以使用任何重组表达方法。例如，可以将编码 MagR 和 Cry 的核酸导入合适的宿主细胞中。通常，将核酸（优选地以 DNA 形式）整合入表达载体中，从而在导入宿主细胞后指导生产相互作用蛋白成员。许多类型的载体可以用于本发明。获知本公开内容的技术人员应了解构建用于本发明目的的表达载体的方法。

通常，表达载体包括具有启动子的表达盒，所述启动子可操作地连接到编码相互作用蛋白成员的 DNA 上。所述启动子可以是天然启动子，即在天然存在的细胞中发现的负责在该细胞中表达相互作用蛋白成员的启动子。或者，所述表达盒可以是嵌合表达盒，即，具有不是在天然存在的细胞中负责表达相互作用蛋白成员的天然启动子的异源启动子。表达载体还可以包括用于在宿主细胞中复制载体的 DNA 复制起点。优选地，表达载体包括用于在例如大肠杆菌中扩增载体的复制起点，和用于仅选择和维持携带表达载体的那些宿主细胞的选择标记。另外，表达盒优选地也含有诱导型元件，其功能是控制从编码相互作用蛋白成员的 DNA 的转录。其它调节序列例如转录增强子序列和翻译调节序列（例如，SD 序列）也可以可操作地包含在表达盒中。终止序列（例如来自牛生长激素、SV40 和 lacZ）也可以可操作地与表达盒中编码蛋白成员的 DNA 连接。当需要在单个宿主细胞中表达 2 个或更多个相互作用蛋白成员时，编码相互作用蛋白成员的 DNA 片段可以整合入单个载体或不同载体中。

通过本领域已知的任意技术，例如，通过直接 DNA 转化、显微注射、电穿孔、病毒感染、脂转染、基因枪等，可以将表达载体导入宿主细胞中。或者可通过常规技术例如稳定细胞系的选择或位点特异性的重组，可以将表达载体整合入宿主细胞的染色体中。

可以将载体构建体设计成适合在各种宿主细胞中表达，宿主细胞包括但不限于细菌、酵母细胞、植物细胞、昆虫细胞和哺乳动物和人细胞。

其中，可以将哺乳动物细胞用作用于表达融合蛋白和检测蛋白-蛋白相互作用的宿主细胞。为此，实际上可以使用任何哺乳动物细胞，包括正常组织细胞、稳定的细胞系和转化的肿瘤细胞。方便地，使用哺乳动物细胞系，例如 CHO 细胞、Jurkat T 细胞、NIH 3T3 细胞、HEK-293 细胞、CV-1 细胞、COS-1 细胞、HeLa 细胞、VERO 细胞、MDCK 细胞、WI38 细胞等。哺乳动物表达载体是本领域众所周知的，且许多可商业得到。用于在哺乳动物细胞中转录嵌合基因的合适启动子的实例包括源自下述病毒的病毒转录启动子：腺病毒、猿猴病毒 40(SV40)(例如，SV40 的早期和晚期启动子)、劳斯肉瘤病毒(RSV)和巨细胞病毒(CMV)(例如，CMV 立即早期启动子)、人免疫缺陷病毒(HIV)(例如，长末端重复序列(LTR))、痘苗病毒(例如，7.5K 启动子)和单纯疱疹病毒(HSV)(例如，胸苷激酶启动子)。也可以使用诱导型启动子。合适的诱导型启动子包括，例如，四环素响应元件(TRE)(参见 Gossen 等人, *Proc. Natl. Acad.*

Sci. USA, 89: 5547-5551(1992))、金属硫蛋白 IIA 启动子、蜕皮激素- 响应启动子和热激启动子。用于在哺乳动物细胞中复制和维持表达载体的合适复制起点包括, 例如, 在有 Epstein Barr 核抗原存在下的 Epstein Barr 复制起点(参见 Sugden 等人, *Mole. Cell. Biol.*, 5: 410-413(1985))和在有 SV40 T 抗原(它存在于 COS-1 和 COS-7 细胞中)存在下的 SV40 复制起点(参见 Margolskee 等人, *Mole. Cell. Biol.*, 8: 2837(1988))。合适的选择标记包括, 但不限于, 赋予对新霉素、潮霉素、zeocin 等的抗性的基因。许多可商业得到的哺乳动物表达载体可以用于本发明, 包括, 例如, pCEP4、pcDNA1、pIND、pSecTag2、pVAX1、pcDNA3.1 和 pBI-EGFP 和 pDisplay。使用任何已知的技术, 例如磷酸钙沉淀、脂转染、电穿孔等, 可以将载体导入哺乳动物细胞。可以将诱饵载体和猎物载体共转化进相同细胞, 或者, 导入 2 个不同的细胞, 随后通过细胞融合或其它合适的技术将它们融合到一起。

允许通过病毒感染将重组基因导入细胞中的病毒表达载体, 也可以用于表达融合蛋白。本领域普遍已知的病毒表达载体包括基于腺病毒、牛乳头瘤病毒、鼠干细胞病毒和反转录病毒等的病毒载体。

使用上述的重组方法, 也可以容易地表达天然相互作用蛋白成员的同源物和片段。例如, 为了表达蛋白片段, 可以选择整合入表达载体中的 DNA 片段, 从而使它仅仅编码该蛋白片段。同样, 使用编码杂种蛋白的重组 DNA, 可以表达特定的杂种蛋白。类似地, 同源物蛋白可以从编码该同源物蛋白的 DNA 序列表达。通过使用重组 DNA 技术操作编码天然蛋白的序列, 可以得到编码同源物蛋白的 DNA 序列。为此, 可以使用本领域普遍已知的技术, 进行随机的或定点诱变。

另外, 通过使某些用于修饰蛋白的部分经由化学方式连接到天然蛋白的氨基酸侧链上, 由此制备本发明的蛋白的衍生物。

如果需要, 可以测试这样产生的同源物和衍生物, 以确定它们是否能与它们预期的配体相互作用, 以形成蛋白复合物。测试可以通过例如酵母双杂交系统或本领域已知的检测蛋白-蛋白相互作用的其它方法来进行。

通过本领域技术人员众所周知的常规生化和免疫化学方法, 可以纯化表达的蛋白。转染的真核细胞或生物组织样品可以匀浆, 并在将不同细胞组分进行分离的适宜条件下分级分离。通常, 将细胞裂解物在蔗糖梯度(或基于大小和密度分离细胞组分的其它材料)上运行。使用本领域技术人员众所周知的方法, 例如免疫印迹或免疫沉淀方法, 用适宜的抗体分析目标蛋白的存在。

然后将纯化的蛋白用于亲和层析处理: 将来自培养的细胞或匀浆的组织样品的提取物在适宜的缓冲液中装载到柱上, 目的蛋白或蛋白复合物可与柱上的成分结合; 洗脱非结合蛋白; 然后使用各种方法, 例如 pH 梯度或盐浓度梯度, 洗脱结合蛋白或蛋白复合物; 最后可以通过二维凝胶电泳, 分离洗脱的蛋白, 也可以通过微量测序进行鉴别。所有这些方法是本领域技术人员众所周知的。

纯化的目标蛋白或复合物, 也可以用于在兔子、小鼠、大鼠、鸡、山羊、绵羊、猪、豚鼠、牛和马中制备抗体。用于抗体生成和表征的方法是本领域技术人员众所周知的。

附图说明

图 1: 家鸽 MagR 蛋白的聚丙烯酰胺凝胶电泳分析图。

图 2: 家鸽和黑脉金斑蝶的 MagR-Cry 蛋白复合物的聚丙烯酰胺凝胶电泳分析图。

图 3: 家鸽 MagR 蛋白样品全波长扫描结果图。

5 图 4: 家鸽 MagR 形成的纯合蛋白复合物或杂合蛋白复合物的电镜观察结果图。
A: 家鸽 MagR 形成的纯合蛋白复合物的横截面结构; B: 家鸽 MagR 形成的纯合蛋白复合物的纵截面结构; C: 家鸽 MagR-Cry4 杂合蛋白复合物的横截面结构; D: 家鸽 MagR-Cry4 杂合蛋白复合物的纵截面结构。

10 图 5: 家鸽 MagR 纯合蛋白复合物的结构图。A: 家鸽 MagR 纯合蛋白复合物的纵截面的电镜形态和结构示意图; B: 家鸽 MagR 纯合蛋白复合物的横截面的电镜形态和结构示意图。

图 6: 家鸽 MagR 和 Cry 形成的杂合蛋白复合物的结构图。A: 杂合蛋白复合物的纵截面; B: 杂合蛋白复合物的横截面。

15 图 7: 家鸽 MagR 纯合蛋白复合物晶体受磁场影响的电镜观察结果图。A: 电子显微镜照片; B: 统计结果。

图 8: 用铁粉从细胞裂解液中分离和纯化家鸽 MagR-Cry4 杂合蛋白复合物的实验示意和电泳检测结果图。

图 9: 家鸽 MagR 纯合蛋白复合物晶体在旋转磁场中同步旋转实验结果图。

图 10: 黑脉金斑蝶 MagR 蛋白的聚丙烯酰胺凝胶电泳分析图。

20 图 11: 黑脉金斑蝶 MagR-Cry1 纯合蛋白复合物的晶体在旋转磁场中同步旋转实验结果图。

图 12: 家鸽 MagR 突变体蛋白样品全波长扫描结果图。

图 13: 各物种 MagR 蛋白的序列比对图。

图 14: 家鸽 MagR-Cry4 杂合蛋白复合物在溶液中的室温磁性测量的负对照。

25 图 15: 家鸽 MagR-Cry4 杂合蛋白复合物在溶液中在室温下的 SQUID 磁性测量结果。

具体实施方式

以下实施例中, 如无特殊说明, 所使用的试剂、仪器都是本领域常用试剂、仪器, 可以从化学或生物制品/制剂公司购买; 以下实施例中使用的方法, 如 PCR、亲和纯
30 化、分子筛凝胶层析、离子柱交换等都是本领域常规方法, 本领域技术人员根据如《分子克隆实验指南》等现有技术或根据生产商提供的操作手册可以毫无疑义地知道这些实验的操作过程并获得相应结果。

实施例 1: 家鸽 MagR 蛋白的获得

35 利用同源性较高的鸡 MagR 蛋白质序列 (SEQ ID NO: 9) 在家鸽基因组中通过 blast 寻找同源序列, 并将找到的同源区段提取出来在软件 sequencher 中与鸡的 MagR 编码区序列进行比对, 得知一共四个外显子。其中, 外显子 2、3、4 的序列和鸡的 MagR 序列有较好的匹配, 但家鸽基因组序列中 MagR 基因的外显子 1 序列缺失。为了获得完整的鸽子 MagR 的序列, 根据鸡的 MagR 的外显子 1 的序列设计如下引物,

并从家鸽基因组 DNA 中扩增该外显子:

外显子 1 引物: F: ACGGACCGCTACCCAATAG (正向引物) (SEQ ID NO: 34),
R: ACACTTTTTCAGCGGCTGTG (反向引物) (SEQ ID NO: 35)。

将 PCR 得到的外显子 1 的序列和通过上述基因组匹配获得的外显子 2、3、4 的
5 序列合并, 获得家鸽 MagR 蛋白的编码区全序列:

ATGGCCTCGTCGGCCTCGTCCGTGGTGAGGGCCACGGTGCGCGCCGTCAGC
AAGCGGAAGATCCAGGCTACGCGCGCCGCCCTCACCTGACCCCATCAGCTGTT
CAGAAGATAAAAGAGCTTCTTAAAGATAAACCTGAGCATGTAGGCGTGAAAGTA
GGTGTTTCGCACAAGAGGATGCAATGGACTTTCTTACACATTAGAATATACAAAAT
10 CAAAAGGAGACTCTGATGAAGAAGTAGTTCAAGATGGGGTTAGAGTGTTTATTG
AGAAGAAGGCACAGCTGACGCTTTTAGGCACTGAAATGGACTATGTAGAAGAC
AAACTGTCCAGTGAATTTGTCTTCAATAATCCAAACATCAAAGGAACATGTGGCT
GTGGAGAAAGCTTTAACATCTGA (SEQ ID NO: 36)。

其编码的氨基酸序列如 SEQ ID NO: 1 所示, 命名为 clMagR。

之后在家鸽基因组 DNA 中扩增 MagR 序列, 验证上述序列正确。

实施例 2: 家鸽 MagR 蛋白的表达和纯化

1. 人工合成编码以下氨基酸序列的插入片段(表达框):

MWSHPQFEKGGSTSMASSASSVVRATVRAVSKRKIQATRAALTTPSAVQKIKEL
LKDKPEHVGKVGVRTRGCNGLSYTLEYTKSKGDSDEEVVQDGV RVFIEKKAQLTLL
20 GTEM DYVEDKLSSEFVFNNPNIKGTCGCGESFNI (SEQ ID NO: 23);

插入片段序列:

CATATGTGGAGCCACCCCCAGTTCGAAAAGGGCGGTTCCACTAGTATGGCTTC
TTCTGCTTCTTCTGTTGTTCTGCTACCGTTCGTGCTGTTTCTAAACGTAAAATCCA
GGCTACCCGTGCTGCTCTGACCCTGACCCCGTCTGCTGTTTCAGAAAATCAAAGAAC
25 TGCTGAAAGACAAACCGGAACACGTTGGTGTTAAAGTTGGTGTTCTGCTACCCGTGGT
TGCAACGGTCTGTCTTACACCCTGGAATACACCAAATCTAAAGGTGACTCTGACGA
AGAAGTTGTTTCAGGACGGTGTTCTGCTGTTTTCATCGAAAAAAAAGCTCAGCTGACCC
TGCTGGGTACTGAAATGGACTACGTTGAAGACAAACTGTCTTCTGAATTCGTTTTC
AACAACCCGAACATCAAAGGTACTTGCGGTTGCGGTGAATCTTTCAACATCTAAG
30 GATCC (SEQ ID NO: 24);

其中, 第 2-9 个氨基酸 WSHPQFEK (SEQ ID NO: 26) 为 Strep II 亲和标签序列; 第
10-14 个氨基酸 GGSTS (SEQ ID NO: 25) 为链接区序列; 带下划线的序列为家鸽(*Columba*
livia) 的 MagR 的蛋白质序列, 即如 SEQ ID NO: 1 所示的氨基酸序列。

核苷酸序列中, 第 1-6 位是 NdeI 酶切位点, 7-30 位是 StrepII 表达序列, 31-45 位是
35 GGSTS (SEQ ID NO: 25) 表达序列, 46-441 位是家鸽 MagR 表达序列, 442-444 位是终
止密码子, 445-450 位是 BamHI 酶切位点。该基因在金斯瑞生物科技有限公司进行合成
所得。

2. 将上述合成的编码家鸽 MagR 蛋白的氨基酸序列且 N 端具有 Strep II 亲和标签的
插入片段通过位点 NdeI 和 BamHI 克隆到 pET-21a(+)质粒 (Novagen)。

3. 将上述包含具有编码家鸽 MagR 蛋白的氨基酸序列且 N 端具有 Strep II 亲和标签的插入片段的质粒转化大肠杆菌 BL21(DE3) (商购)。挑取单克隆接种至 LB 液体培养基中扩大培养, 15℃下用 20μM 的 IPTG 诱导目的蛋白的表达, 20-24 小时后收菌。

4. 将上述得到的表达了 MagR 蛋白的转化菌用缓冲液 (20mM Tris, 150mM NaCl, pH8.0, 10mM 巯基乙醇以及蛋白酶抑制剂) 重悬后, 用超声破碎仪裂解 (工作 4s, 暂停 8s, 99 个循环), 完成后进行高速离心 (17000rpm), 取上清液进行亲和纯化。通过 Strep-Tactin 亲和柱 (IBA 公司) 亲和纯化后, 进行阴离子交换柱 (GE Healthcare 公司) 或者分子筛凝胶层析 (GE Healthcare 公司) 进一步纯化, 得到纯度较高的家鸽 MagR 蛋白。纯化全过程在 4℃进行。纯化后得到的 MagR 蛋白在浓度较高时呈现明显的棕黄色或棕褐色, 这提示蛋白质中结合了二价或三价的铁。

5. 对收集的 MagR 蛋白样品进行变性聚丙烯酰胺凝胶电泳(SDS-PAGE)分析。实验结果如图 1 所示, 其中各泳道的样品为: 1、蛋白分子量标准; 2、纯化的含有 Strep II 亲和标签的家鸽 MagR 蛋白, 分子量约为 15KDa。

实施例 3: 家鸽 MagR 蛋白形成二聚体和多聚体的纯合复合物

将实施例 2 中表达和纯化的家鸽 MagR 蛋白, 在通过 Strep-Tactin 亲和柱亲和纯化后进一步用 Superdex 200 10/300 GL 纯化柱 (GE Healthcare) 进行分子筛凝胶层析 (缓冲液: 20mM Tris, 150mM NaCl, pH8.0, 10mM 巯基乙醇), 即凝胶过滤 (Gel-filtration) 或凝胶排阻层析 (Size-exclusion chromatography)。样品的分子筛凝胶层析呈现出单体峰、二聚体峰和多聚体峰, 分别收集三个峰的样品, 获得纯化蛋白。这显示, 纯化的 MagR 蛋白在分子筛凝胶层析中以三种形式存在。与蛋白质分子量标准进行比较可知, 三个峰 (洗脱体积大约为: 17.5ml、16.2ml、9.5ml) 的大致分子量为 15 KDa、30 KDa 和大于 200KDa, 这与 MagR 蛋白的单体、二聚体和多聚体的分子量相吻合。由于 Superdex S200 的分辨率最高为约 200KDa, 所以, 尚无法判断多聚体的分子量和多聚体所含单体个数。

实施例 4: 家鸽 MagR 蛋白和家鸽 Cry4 蛋白形成的杂合蛋白复合物的表达和纯化

1. 人工合成编码以下氨基酸序列的插入片段(表达框):

MHHHHHHHHHHHGGSTSMPHRTIHLFRKGRLRLHDNPTLLAALESSETIYPVYVLDR
RFLASAMHIGALRWHFLLQSLEDLHKNLSRLGARLLVIQGEYESVLRDHSVQKWNITQ
VTLDAEMEPFYKEMEANIRRLGAELGFEVLSRVGHSLYDTKRILDNLNGGSPPLTYKRF
LHILSQLGDPEVPVRNLTAEDFQRCMSPEPGLAERYRVPVPADLEIPPQSLSPWTGGET
EGLRRLEQHLTDQGWVANFTKPRITIPNSLLPSTTGLSPYFSMGCLSVRTFFQRLSNIYA
QAKHHSLLPPVSLOGQLLWREFFYTVASATQNFTQMAGNPICLQIHWYEDAERLHKWK
TAQTGFPWIDAIMTQLRQEGWIHHLARHAVACFLTRGDLWISWEEGMKVFEELLLDA
DYSINAGNWMWLSASAFFHHYTRIFCPVRFGKRTDPEGQYIRKYLPLVKNFPTKYIYE
PWTASEEEQRQAGCIIGRDYPFPMVNHKEASDRNLQLMRRVREEQRGTAQLTR (SEQ ID NO: 27)。

插入片段的序列:

CATATGCATCATCATCATCATCACCATCACCATCACGGCGGTTCCACTAGTAT

GCCGCACCGTACCATCCACCTGTTCCGTAAAGGTCTGCGTCTGCACGACAACCCGA
CCCTGCTGGCTGCTCTGGAATCTTCTGAAACCATCTACCCGGTTTACGTTCTGGACC
GTCGTTTCCTGGCTTCTGCTATGCACATCGGTGCTCTGCGTTGGCACTTCCTGCTGC
AGTCTCTGGAAGACCTGCACAAAAACCTGTCTCGTCTGGGTGCTCGTCTGCTGGTT
5 ATCCAGGGTGAATACGAATCTGTTCTGCGTGACCACGTTTACAGAAATGGAACATCA
CCCAGGTTACCCTGGACGCTGAAATGGAACCGTTCTACAAAGAAATGGAAGCTAA
CATCCGTCGTCTGGGTGCTGAACTGGGTTTCGAAGTTCTGTCTCGTGTGGTCACTC
TCTGTACGACACCAAACGTATCCTGGACCTGAACGGTGGTTCTCCGCCGCTGACCT
ACAAACGTTTCCTGCACATCCTGTCTCAGCTGGGTGACCCGGAAGTTCCGGTTCGT
10 AACCTGACCGCTGAAGACTTCCAGCGTTGCATGTCTCCGGAACCGGGTCTGGCTGA
ACGTTACCGTGTTCGGTTCGGCTGACCTGGAAATCCCGCCGCAGTCTCTGTCTC
CGTGGACCGGTGGTGAAACCGAAGGTCTGCGTCGTCTGGAACAGCACCTGACCGA
CCAGGGTTGGGTTGCTAACTTCACCAAACCGCGTACCATCCCGAACTCTCTGCTGC
CGTCTACCACCGGTCTGTCTCCGTACTTCTCTATGGGTTGCCTGTCTGTTCTGTACCT
15 TCTTCCAGCGTCTGTCTAACATCTACGCTCAGGCTAAACACCACTCTCTGCCGCCG
GTTTCTCTGCAGGGTCAGCTGCTGTGGCGTGAATTCTTCTACACCGTTGCTTCTGCT
ACCCAGAACTTCACCCAGATGGCTGGTAACCCGATCTGCCTGCAGATCCACTGGTA
CGAAGACGCTGAACGTCTGCACAAATGGAACACCGCTCAGACCGGTTTCCCGTGG
ATCGACGCTATCATGACCCAGCTGCGTCAGGAAGGTTGGATCCACCACCTGGCTCG
20 TCACGCTGTTGCTTGCTTCTTCTGACCCGTGGTGACCTGTGGATCTCTTGGGAAGAAG
GTATGAAAGTTTTCGAAGAACTGCTGCTGGACGCTGACTACTCTATCAACGCTGGT
AACTGGATGTGGCTGTCTGCTTCTGCTTTCTTCCACCACTACACCCGTATCTTCTGC
CCGGTTCGTTTCGGTAAACGTACCGACCCGGAAGGTCAGTACATCCGTAAATACCT
GCCGGTTCTGAAAAACTTCCCGACCAAATACATCTACGAACCGTGGACCGCTTCTG
25 AAGAAGAACAGCGTCAGGCTGGTTGCATCATCGGTCGTGACTACCCGTTCCCGAT
GGTTAACCACAAAGAAGCTTCTGACCGTAACCTGCAGCTGATGCGTCGTGTTCTGTG
AAGAACAGCGTGGTACCGCTCAGCTGACCCGTTAACTCGAG (SEQ ID NO: 28)。

其中,第2-11个氨基酸为10个组氨酸的His亲和标签序列;第12-16个氨基酸 GGSTS
 (SEQ ID NO: 25)为链接区序列;带下划线的序列为家鸽的Cry4的氨基酸序列,即如
 30 SEQ ID NO: 15所示的氨基酸序列。

核苷酸序列中,第1-6位是NdeI酶切位点,7-36位是10×His表达序列,37-51位是
 GGSTS (SEQ ID NO: 25)表达序列,52-1542位是家鸽Cry4表达序列,1543-1545位是
 终止密码子,1546-1551位是XhoI酶切位点。该基因在博迈德生物科技有限公司进行合
 成所得。

35 2. 将上述合成的编码家鸽Cry4蛋白的氨基酸序列并且N端具有His亲和标签的插
 入片段通过NdeI和XhoI位点克隆到pET-21a(+)质粒(Novagen)。

3. 将前述具有家鸽Cry4蛋白的编码基因并融合了组氨酸标签(His-tag)的质粒与
 实施例2步骤2得到的具有家鸽MagR蛋白的编码基因并融合了StrepII标签的质粒共同
 转化大肠杆菌BL21(DE3)。表达MagR的质粒为卡那霉素抗性,表达Cry4的质粒为氨苄

青霉素抗性，在双抗培养基（卡那霉素和氨苄青霉素）中挑取含有两种质粒的大肠杆菌单克隆接种至 LB 液体培养基中扩大培养，15℃下用 20μM 的 IPTG 诱导目的蛋白的表达，20-24 小时后收菌。

4. 将上述得到的共表达了家鸽 MagR 蛋白和 Cry4 蛋白的转化菌用缓冲液（20mM Tris, 150mM NaCl, 10mM 咪唑, pH8.0, 10mM 巯基乙醇以及蛋白酶抑制剂）重悬后，用超声破碎仪裂解，完成后进行高速离心，取上清液进行亲和纯化。通过 Ni-NTA 亲和柱（QIAGEN 公司）亲和纯化后，再通过 Strep-Tactin 亲和柱（IBA 公司）纯化，得到家鸽 MagR 蛋白和 Cry4 蛋白形成的 MagR-Cry4 杂合蛋白复合物。进一步通过阴离子交换柱和分子筛凝胶层析（GE Healthcare 公司）纯化，得到纯度较高的杂合蛋白复合物。纯化全过程在 4℃和存在外加磁场的条件下进行。外加磁场通过下述方法施加：当样品过亲和柱时，将两个相同的条形磁铁分别绑在亲和柱的两侧，该两个相同的条形磁铁的南极和北极分别相对，提供一个均匀稳定的磁场。结果显示，在纯化时施加地球磁场 100 倍的外加磁场，纯化蛋白的产率有显著增加；在洗脱纯化蛋白时，撤去磁场，蛋白质洗脱效率更高。

5. 对收集的家鸽 MagR-Cry4 杂合蛋白复合物样品进行变性聚丙烯酰胺凝胶电泳 (SDS-PAGE) 分析。实验结果如图 2 所示，其中从左向右第 4 泳道加载的样品为蛋白质分子量标准，第 5 泳道加载的样品为家鸽 MagR-Cry4 杂合蛋白复合物。结果：由于使用变性电泳分析，泳道 5 中显示出两种蛋白质样品，即含有 Strep II 亲和标签的家鸽 MagR 蛋白（分子量约为 15KDa）以及含有 His 标签的家鸽 Cry4 蛋白（分子量约为 70KDa）。

实施例 5：家鸽 MagR 蛋白以及 MagR 纯合蛋白复合物含有铁的证明实验

将实施例 3 得到的家鸽 MagR 蛋白的单体峰样品用分光光度计（Nanodrop 2000）进行全波长扫描。

如图 3 所示，结果显示在 Fe 元素的特征吸收位置（320 nm 和 410 nm）有吸收峰。第一个吸收峰为 UV280 的蛋白质吸收峰。

将实施例 3 中得到的家鸽 MagR 蛋白形成的二聚物和多聚物样品分别用分光光度计（Nanodrop 2000）进行全波长扫描，结果与图 3 相似。

实施例 6：家鸽 MagR 纯合蛋白复合物和家鸽 MagR-Cry4 杂合蛋白复合物的电子显微镜图像和结构

将实施例 3 中经分子筛凝胶层析纯化后获得的家鸽 MagR 纯合蛋白复合物的多聚体蛋白样品以及实施例 4 中经分子筛凝胶层析纯化后的家鸽 MagR-Cry4 杂合蛋白复合物的蛋白样品采用乙酸铀或甲酸铀负染，制备蛋白质的负染电镜样品。将样品放置于透射电镜下观察并拍照，按照标准的电镜结构分析方法解析其结构，并测量大小（包括蛋白长度和宽度）。所述电镜结构分析方法可参照 Ohi, M., et al. (2004), "Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy." *Biol Proced Online* 6: 23-34。

结果如图 4 所示。其中 A 和 B 是家鸽 MagR 形成的纯合蛋白复合物的电镜结构，其中 A 为横截面结构，B 为纵截面结构。C 和 D 为家鸽 MagR-Cry4 杂合蛋白复合物的电子显微镜结构，其中 C 为横截面结构，D 为纵截面结构。MagR 和/或 Cry 蛋白在复合物中的组织形式和相对位置的示意图如右侧所示，并标出了在电镜下测量的

大小尺寸（单位为纳米）。

如图 4 的 A 和 B 显示，家鸽 MagR 纯合蛋白复合物呈现为长条形的棒状结构，长度为约 20-24 纳米长，宽度为约 7-11 纳米。“宽度”是指电镜下该棒状结构的横截面为圆形或类似圆形的形状时测量到的直径。根据同源晶体结构和分子模型计算，结果表明该复合物具有约 20 个 MagR 蛋白。如图 4 的 C 和 D 显示，家鸽 MagR-Cry4 杂合蛋白复合物呈现为长条形的棒状结构，长度为约 20-24 纳米长，宽度为约 13-18 纳米。根据同源晶体结构和分子模型计算，结果表明该多聚体具有约 20 个 MagR 蛋白和 10 个 Cry4 蛋白。

图 5 为家鸽 MagR 纯合蛋白复合物的结构图。图 5 中，A 显示复合物的纵截面的电镜形态和结构示意，B 显示复合物的横截面的电镜形态和结构示意。每一个 MagR 纯合蛋白复合物包括约 20 个 MagR 蛋白，有序排列形成线性双螺旋聚合体，其中每 10 个 MagR 蛋白单体以线性聚合的形式首尾相接，形成两条蛋白长链，两条蛋白长链以螺旋形式结合，由此形成长条形棒状结构的蛋白复合物。图 5 中的虚线框显示出一个 MagR 四聚体的结构：如图所示的 MagR 纯合蛋白复合物棒状结构由 5 个该四聚体构成。另外，每一个 MagR 蛋白包括一个 $2\text{Fe}2\text{S}$ 的铁硫中心，也即每一个 MagR 蛋白包括一个铁结合位点，结合了 2 个铁原子。整个棒状的纯合蛋白复合物包含约 20 个铁结合位点，结合了约 40 个铁原子。所述多个铁原子位于棒状结构的蛋白复合物的内部近中心轴区域。磁感应受体蛋白单体结合铁原子的位置相同或相近，处于棒状结构的蛋白复合物的立体结构的内部，由此，在磁感应受体蛋白单体线性聚合棒状结构的蛋白复合物中，所述多个铁原子也在蛋白复合物中形成有规律的排列。

图 6 为家鸽 MagR-Cry4 杂合蛋白复合物的结构图。图 6 中，A 显示杂合蛋白复合物的纵截面的电镜形态（左上）、组织形式（左下）和结构示意（右），B 显示杂合蛋白复合物的横截面的电镜形态（左上）、组织形式（左下）和结构示意（右）。在 MagR-Cry4 杂合蛋白复合物中，在由约 20 个 MagR 组成的纯合蛋白复合物所形成的棒状线性双螺旋聚合体的外围，有约 10 个感光的 Cry4 蛋白缠绕排列，从而形成杂合复合物。同样地，整个棒状的杂合蛋白复合物包含约 20 个铁结合位点，结合了约 40 个铁原子。所述多个铁原子位于棒状结构的蛋白复合物的内部近中心轴区域，也形成有规律的排列。

实施例 7：家鸽 MagR 蛋白复合物结晶

将实施例 3 中经分子筛凝胶层析纯化后获得的家鸽 MagR 纯合蛋白复合体的蛋白样品以及实施例 4 中经分子筛凝胶层析纯化后的家鸽 MagR-Cry4 杂合蛋白复合物的蛋白样品浓缩，然后进行结晶。蛋白质结晶方法按照如教科书 McPherson, A. (2009). *Introduction to macromolecular crystallography*. Hoboken, N.J., Wiley-Blackwell 记载的标准实验流程采用悬滴法进行。在硅化过的显微镜盖玻片上将蛋白质溶液与等体积的沉淀剂溶液混合形成液滴；盖玻片被倒置在盘子的空穴上方，空穴中放置了所需的沉淀剂溶液，在盖玻片放置前，在空穴的边缘涂抹油或脂以密封空穴。大约 2-7 天后可以观察到 MagR 蛋白复合物形成的晶体。

图 9 显示了本发明的家鸽 MagR 蛋白形成的纯合复合物的晶体照片。

结果：共观察到两种类型的晶体：

第一种是黑色不透明的长方体形状的晶体（参见图 9 中上排给出的晶体照片；所使用的结晶条件为：0.1M HEPES, pH 7.5, 3.0M 氯化钠），其长度约为 0.5mm，显示铁元素在晶体中已被氧化成均一的三价。第二种晶体为浅黄色或黄色的透明晶体（参见图 9 中下排给出的晶体照片；所使用的结晶条件为：0.1M BIS-TRIS, pH6.5, 3.0M 氯化钠），形状不规则或呈小片状，显示铁元素可能以二价存在。在以下实施

例的实验中证明，两种晶体都表现出非常强的磁性。

本发明的家鸽 MagR 蛋白与 Cry4 蛋白形成的杂合复合物的晶体的形态与图 9 的照片显示的晶体相似。

实施例 8：家鸽 MagR 蛋白复合物的磁性测试

家鸽 MagR 纯合蛋白复合物以及家鸽 MagR-Cry4 杂合蛋白复合物的磁性测试通过以下几个实验来进行：

（1）电子显微镜下的磁性测试：

在实施例 6 中描述的电镜分析制样中，将实施例 3 中经分子筛凝胶层析纯化后获得的家鸽 MagR 纯合蛋白复合物的蛋白样品以及实施例 4 中经分子筛凝胶层析纯化后的家鸽 MagR-Cry4 杂合蛋白复合物的蛋白样品制备负染电镜样品，此时在电镜铜网样品上标记制样时的地球磁场方向或者人工外加的磁场方向，在电镜得到的蛋白颗粒的图像中，统计所有的棍状的蛋白复合物颗粒的方向（蛋白为长条形棒状，其方向是指该蛋白长度方向，即长轴的轴向），按照图 7 所示的三个方向统计（平行于地球或者外加磁场方向的，垂直于地球或者外加磁场方向的，以及在两者之间不能归于上述两个方向的），计算各个方向的棍状蛋白复合物所占的比例。

结果如图 7 所示。图 7 为利用电子显微镜对 MagR 纯合蛋白复合物在磁场下的排列方向进行研究，显示 MagR 蛋白复合物具有自身磁性，达到了类似“蛋白质指南针”的效果。图 7 的 A 为电子显微镜照片，其中用深色方框标示平行于磁场方向（即与磁场方向相同或相差小于等于 30 度）的蛋白质复合物颗粒，浅色方框标示垂直于磁场方向（即与磁场垂直的方向相同或相差小于等于 30 度）的蛋白质复合物颗粒，圆圈显示介于两者之间的其它方向。图 7 的 B 显示对于大样本的颗粒数目按排列方向进行统计的结果。每次实验包括大约 1500 个蛋白复合物颗粒进行统计分析，实验重复三次，取平均值和标准差作图。图 7 的 B 中横坐标为磁场的强度，以 mT 为单位，地球磁场为 0.04 mT（以北京的地理位置为准），试验中所采用的外加的磁场通过永磁铁获得，磁场强度经过磁力计测量为 1 mT。纵坐标为沿着三个不同的方向（平行于地球磁场或者外加磁场的方向，垂直于地球磁场或者外加磁场的方向，介于两者之间的其它方向）排列的 MagR 蛋白复合物在所有蛋白复合物中所占的比例。结果显示，家鸽磁感应受体蛋白 MagR 形成的纯合蛋白复合物有将近或超过一半的蛋白质颗粒顺应（平行于）地球磁场或者外加人工磁场的方向。增加磁场强度，平行于磁场方向的蛋白质颗粒所占比例将随之上升，其它两个方向的比例将相应地下降。这充分说明即使在微观条件下的纳米尺度的蛋白质颗粒也呈现了非常明显的磁性，具备与指南针类似的物理性质。蛋白质颗粒在磁场中的方向不是像宏观世界的指南针一样 100%地顺应磁场方向分布，是因为在微观的纳米尺度下，分子热运动和溶液中的布朗运动严重影响了蛋白质颗粒的取向，使得微观粒子（如本实验中的纳米尺度下的蛋白质颗粒）的取向只遵循统计物理学的规律。

本发明的家鸽 MagR 蛋白与 Cry4 蛋白形成的杂复合物的晶体受磁场影响的实验和电镜观察结果与图 7 的照片显示的相似。

(2) 利用蛋白质的磁性对蛋白质进行分离和纯化:

家鸽 MagR 纯合蛋白复合物和家鸽 MagR-Cry4 杂合蛋白复合物可用铁粉从细胞裂解液中快速分离和纯化出来, 表明它们均具有磁性。

以家鸽 MagR-Cry4 杂合蛋白复合物为例, 具体方法如图 8 所示: 将实施例 4 中共表达 MagR 与 Cry4 蛋白质的大肠杆菌细胞收集起来, 用缓冲液 (20mM Tris, 150mM NaCl, pH7.5) 重悬, 超声破碎后取上清液, 向其加入市售的纳米铁粉颗粒 (海狸生物科技有限公司, #70301), 室温孵育 10min 后, 用标准 TBS 缓冲液 (20mM Tris, 150mM NaCl, pH7.5) 清洗铁粉颗粒, 然后用蛋白电泳上样缓冲液对铁粉进行洗脱, 对洗脱液进行变性聚丙烯酰胺凝胶电泳 (SDS-PAGE), 分析蛋白成分。

结果显示, 经上述方法得到的纳米铁粉洗脱液中含有对应的蛋白。如图 8 所示, 在实施例 4 中共表达 MagR 与 Cry4 蛋白质的大肠杆菌细胞的样品经上述方法得到的纳米铁粉洗脱液经变性聚丙烯酰胺凝胶电泳后显示出同时含有分子量约为 15KDa 的条带 (MagR) 和约为 70KDa 的条带 (Cry4)。

另外, 在实施例 2 中表达家鸽 MagR 的大肠杆菌细胞的样品经上述方法得到的纳米铁粉洗脱液经变性聚丙烯酰胺凝胶电泳后显示出分子量约为 15KDa 的条带 (MagR)。

(3) 蛋白质晶体的磁性检测:

将实施例 7 中获得的 MagR 纯合蛋白复合物的晶体或 MagR-Cry4 杂合蛋白复合物的晶体置于光学显微镜下观察, 在显微镜的载物台上安装旋转的磁场, 保持均匀的磁场强度 (10 高斯)。将晶体生长的液滴 (其中含有蛋白质晶体) 置于旋转磁场的中心。图 9 为家鸽 MagR 纯合蛋白复合物晶体在旋转磁场中同步旋转实验结果图。如图 9 所示, 在磁场 (圆圈中指针方向为磁场方向) 旋转的时候, 可以观察到蛋白质晶体几乎完全同步地沿着磁场旋转, 如同由磁铁构成的指南针一样。当用磁铁去接近蛋白质晶体时, 蛋白质晶体会立即靠近, 当颠倒磁铁的磁极时, 蛋白质晶体会即刻远离并颠倒方向, 然后重新靠近, 显示蛋白质形成的晶体如同磁铁一样, 具有自身的磁性和磁极。

本发明的家鸽 MagR 蛋白与 Cry4 蛋白形成的杂复合物的晶体在旋转磁场中同步旋转实验的结果与图 9 的照片显示的相似。

另外, 当用铁磁性物质 (如铁棒) 接近前述晶体生长的液滴 (其中含有蛋白质晶体) 时, 蛋白质晶体会向该铁磁性物质迅速靠近, 甚至飞出液滴去接近该铁性物质, 显示蛋白质晶体具备非常强的磁性。

(4) 蛋白质的磁性测量

为了在物理学上证明蛋白质是否具有内在的磁性, 在溶液状态下, 对实施例 4 中获得的纯化的家鸽 MagR-Cry4 杂合蛋白复合物的样品进行物理学的磁性测定, 得到了可以重复的非常精确的磁测量结果。所用仪器是超导量子干涉磁强计 ((Superconducting Quantum Interference Device, SQUID), 型号为 MPMS-XL1。所有测量均在室温下进行, 为标准的 SQUID 测量流程, 具体为: 将纯化的家鸽 MagR-Cry4 杂合蛋白复合物保存在缓冲液中 (20mM Tris, pH 7.5, 150mM NaCl), 蛋白质浓度为 3.8 mg/ml; 负对照为: (1) 对蛋白质样品的样品管 (holder) 进行测量, 确认没有磁滞回线后才用于后续的蛋白质磁

性测量；（2）对缓冲液的样品管（holder）进行测量，确认没有磁滞回线后才用于后续的缓冲液磁性测量作为基准；（3）对缓冲液在同样条件下进行测量。最终，蛋白质的测量结果以缓冲液测量结果修正。

测量结果如图 14 和图 15 所示。图 14 为家鸽 MagR-Cry4 杂合蛋白复合物在溶液中的室温磁性测量之负对照，其中，a 为装有家鸽 MagR-Cry4 杂合蛋白复合物的样品管（Holder）的 SQUID 磁性测量结果，显示无磁滞回线；b 为 SQUID 实验中装载蛋白质储存的溶液（即缓冲液）的样品管的 SQUID 磁性测量结果，显示无磁滞回线；c 为溶液对照，即蛋白质储存的溶液（即缓冲液）的 SQUID 磁性测量结果，显示无磁滞回线。证明了实验过程中使用的装载蛋白质或者缓冲液的容器，以及缓冲液均无磁性污染。横坐标为磁场强度，纵坐标为磁矩。

图 15 中，a 为家鸽 MagR-Cry4 杂合蛋白质复合物在溶液中在室温下的 SQUID 磁性测量结果，显示有明显的磁滞回线；b 为同一管家鸽 MagR-Cry4 杂合蛋白质复合物在测量持续 40 小时以后，磁滞回线消失；c 中左侧为测量前的蛋白样品 SDS-PAGE 结果，箭头标示 Cry4（上）和 MagR 蛋白质（下）的位置，右侧为 40 小时后磁滞回线消失后，对同一管样品的 SDS-PAGE 分析，显示 Cry4 蛋白已经降解为几条小分子量的条带，MagR 已经完全降解消失。可以看出家鸽 MagR-Cry4 杂合蛋白复合物在溶液中室温下具有明显的磁滞回线；而蛋白质降解后，磁滞回线消失，证明了 SQUID 实验测到的磁性来自于蛋白质本身，而不是容器（如样品管），也不是来自于溶液（缓冲液）。

实施例 9：黑脉金斑蝶 MagR 蛋白的获得

使用果蝇 MagR 蛋白序列（SEQ ID NO: 3）在美洲帝王蝶基因组中用 blast 寻找同源序列，并将找到的同源区域提取出来在 sequencer 中与果蝇 MagR 编码区序列进行比对，获得 5 个外显子区域位置。将得到的 5 个外显子序列提取出来、拼接、翻译，得到的蛋白序列与其他物种 MagR 序列一致性较高，且含有保守区，因此认定为黑脉金斑蝶 MagR 蛋白序列，具体序列如 SEQ ID NO: 2 所示，其编码区核苷酸序列为：

ATGTCTACTAAACTATAGCAAGTGCAACTGTTAGGGCAGTAAAAAAGCGT
CTGCTACCATCCAGAGCTGCTCTAGTTTTGACTTCTTCAGCCGTAAATAAAGTT
AAGGAAATAATGGCCAAGGAAGAAGGCAAGGGTTATATAGGATTGAAAGTTG
GTGTGCGGCAAAGAGGTTGCAATGGATTGTCATATACCTTAGATTATGCAACAT
CAAAAGGGAACTTGACGAAGAAGTAAAACAGGATGGAGTCACTATAATTATT
GACAAAAAAGCACAGTTGACCTTGTTGGGTACTGAAATGGATTTTGTCTGAAGA
TAAGCTGTCAGCGGAATTTGTGTTTAACAATCCGAATATAAAAGGCACTTGTGG
ATGTGGAGAATCTTTTCAGTATATAA（SEQ ID NO: 37）。

之后在黑脉金斑蝶的基因组 DNA 中扩增 MagR 蛋白序列，验证上述序列正确。

实施例 10：黑脉金斑蝶 MagR 蛋白的表达和纯化

1. 人工合成编码以下氨基酸序列的插入片段(表达框):

MWSHPQFEKGGSTSMSTKTIASATVRAVKKRLLPSRAALVLTSSAVNKKVKEIMA
KEEGKGYIGLKVGVRQRCNGLSYTLDYATSKGKLDEEVKQDGVTHIDKKAQLTLLG
TEMDFVEDKLSAEFVFNNPNIKGTCGCGESFSI（SEQ ID NO: 29）；

插入片段的序列：

CATATGTGGAGCCACCCCCAGTTCGAAAAGGGCGGTTCCACTAGTATGTCTAC
TAAAACTATAGCAAGTGCAACTGTTAGGGCAGTAAAAAAGCGTCTGCTACCATCC
AGAGCTGCTCTAGTTTTGACTTCTTCAGCCGTAAATAAAGTTAAGGAAATAATGGC
CAAGGAAGAAGGCAAGGGTTATATAGGATTGAAAGTTGGTGTGCGGCAAAGAGG
5 TTGCAATGGATTGTCATATACCTTAGATTATGCAACATCAAAAAGGGAACTTGACG
AAGAAGTAAACAGGATGGAGTCACTATAATTATTGACAAAAAAGCACAGTTGAC
CTTGTTGGGTACTGAAATGGATTTTGTCTGAAGATAAGCTGTCAGCGGAATTTGTGT
TTAACAATCCGAATATAAAAAGGCACTTGTGGATGTGGAGAATCTTTCAGTATATAA
GGATCC (SEQ ID NO: 30)。

10 其中，第 2-9 个氨基酸 WSHPQFEK (SEQ ID NO: 26) 为 Strep II 亲和标签序列；第 10-14 个氨基酸 GGSTS (SEQ ID NO: 25) 为链接区序列；带下划线的序列为黑脉金斑蝶 (或称美洲帝王蝶，拉丁名 *Danaus plexippus*，以下也可简称为“蝴蝶”) 的 MagR 的蛋白质序列，即如 SEQ ID NO: 2 所示的氨基酸序列。

15 核苷酸序列中，第 1-6 位是 NdeI 酶切位点，7-30 位是 StrepII 表达序列，31-45 位是 GGSTS (SEQ ID NO: 25) 表达序列，46-438 位是黑脉金斑蝶 MagR 表达序列，439-441 位是终止密码子，442-447 位是 BamHI 酶切位点。该基因在博迈德生物科技有限公司进行合成所得。

2. 将上述合成的编码蝴蝶 MagR 蛋白的氨基酸序列并且 N 端具有 Strep II 亲和标签的插入片段通过 NdeI 和 BamHI 位点克隆到 pET-21a(+)质粒 (Novagen)。

20 3. 将上述包含具有编码蝴蝶 MagR 蛋白的氨基酸序列且 N 端具有 Strep II 亲和标签的插入片段的质粒转化大肠杆菌 BL21(DE3) (商购)。挑取单克隆接种至 LB 液体培养基中扩大培养，15℃下用 20μM 的 IPTG 诱导目的蛋白的表达，20-24 小时后收菌。

25 4. 将上述得到的表达了 MagR 蛋白的转化菌用缓冲液 (20mM Tris, 150mM NaCl, pH8.0, 10mM 巯基乙醇以及蛋白酶抑制剂) 重悬后，用超声破碎仪裂解，完成后进行高速离心，取上清液进行亲和纯化。通过 Strep-Tactin 亲和柱亲和纯化后，进行阴离子交换柱或者分子筛凝胶层析进一步纯化，得到纯度较高的蝴蝶 MagR 蛋白。纯化全过程在 4℃进行。纯化后得到的 MagR 蛋白在浓度较高时可呈现明显的棕黄色或棕褐色，这提示蛋白质中结合了二价或三价的铁。

30 5. 对收集的 MagR 蛋白样品进行变性聚丙烯酰胺凝胶电泳(SDS-PAGE)分析。实验结果如图 10 所示，其中各泳道的样品为：1、蛋白质分子量标准；2、纯化的含有 Strep II 亲和标签的蝴蝶 MagR 蛋白，分子量约为 15KDa。

实施例 11：黑脉金斑蝶 MagR 蛋白形成纯复合物的实验

35 将实施例 10 中表达和纯化的蝴蝶 MagR 蛋白，在通过 Strep-Tactin 亲和柱亲和纯化后进一步用 Superdex 200 10/300 GL 纯化柱 (GE Healthcare) 进行分子筛凝胶层析 (实验过程参考实施例 3)。

样品的分子筛凝胶层析呈现出单体峰、二聚体峰和多聚峰，分别收集三个峰的样品，获得纯化蛋白。这显示，纯化的 MagR 蛋白在分子筛凝胶层析中以三种形式存在。与蛋白质分子量标准进行比较可知，三个峰的大致分子量为 15 KDa、30KDa 和大于 200KDa，这与 MagR 蛋白的单体、二聚体和多聚体的分子量相吻合。由于

Superdex S200 的分辨率最高为约 200KDa, 所以, 尚无法判断多聚体的分子量范围和多聚体所含单体个数。

实施例 12: 黑脉金斑蝶 MagR 蛋白和 Cry1 蛋白形成的杂合蛋白复合物的表达和纯化

5 1. 人工合成编码以下氨基酸序列的核苷酸插入片段(表达框):

MHHHHHHHHHHHGGSTSM LGGNVIWFRHGLRLHDNPSLHSALEDASSPFFPIFID
GETAGTKMVGYNRMRYLLEALNDLDQQFRKYGGKLLMIKGRPDLIFRRLWEEFGIRT
LCFEQDCEPIWRPRDASVRALCRDIGVSCREHVAHTLWNPDTVIKANGGIPPLTYQMF
LHTVEIIGNPPRPVDDVDLNGVNFGSLPESFYREFVFDKAPKPEDLGVFLENEDIRMIR
10 WVGGETAALKQM QERLAVEYETFCRGSYLPHTGNPDLLGPPISLSPALRFGCLSVRRE
YWSLQDLFQQVHQGRLASTQFITGQLIWREYFYTMSVNNPNYAQMSGNPICLDIPWK
EPENDELQRWKEGRTGFPFVDAAMRQLRTEGWLHHVVRNTVASFLTRGTLWLSWEH
GLQHFLKYLLDADWSVCAGNWMWVSSSAFEALLDSGECACPVRLGRRLEPTGHYVR
RYVPELARMPGEYIYEPWRAPLEVQEAAGCVIGRDYPAPVVDHTAAAARNRANMQE
15 LRRLLEKAPPHCCPSEDEVROFMWLGD DSQPELTTT (SEQ ID NO: 31) ;

插入片段的序列:

CATATGCATCATCATCATCATCACCATCACCATCACGGCGGTCCACTAGTAT
GCTTGGTGGTAATGTCATTTGGTTCCGTCACGGTCTCCGTCCTCCACGACAACCCTTC
GCTTCACAGCGCTCTGGAAGATGCAAGCTCACCGTTCTTCCCTATATTCATATTTG
20 ATGGAGAGACAGCTGGTACAAAGATGGTGGGCTACAATCGTATGCGATACCTGCT
GGAGGCGCTGAACGATTTGGACCAGCAGTTCAGGAAGTACGGCGGGAAACTGCTC
ATGATTAAGGGGAGACCTGATTTAATATTCAGGAGGCTGTGGGAGGAATTTGGTA
TACGTACGCTATGCTTCGAGCAGGACTGTGAGCCAATATGGCGTCCGCGCGACGC
GAGCGTGCGTGCTCTGTGCCGCGACATAGGCGTGTCGTGCCGCGAGCACGTCGCA
25 CACACGCTGTGGAACCCGGACACAGTCATCAAGGCCAATGGAGGAATACCGCCGC
TTACGTACCAGATGTTCTCCTGCATACAGTTGAAATCATCGGTAATCCTCCGCGTCCC
GTAGACGACGTCGACCTGAACGGCGTCAACTTTGGCTCGCTGCCTGAGAGCTTCTA
CAGGGAATTCGTTGTCTTTGATAAGGCTCCAAAACCAGAAGATCTGGGTGTGTTCC
TGGAAAACGAAGATATTCGTATGATTCGCTGGGTGGGAGGAGAGACAGCGGCCTT
30 GAAGCAGATGCAGGAGAGATTGGCTGTGGAGTACGAGACATTCTGCAGGGGTTCT
TATTTGCCGACCCATGGCAACCCCGACCTCCTTGGACCGCCGATATCTCTGAGTCC
AGCCTTGCGCTTCGGATGTCTGTCTGTCCGTCGCTTCTACTGGAGTCTCCAGGACCT
GTTCCAGCAGGTGCATCAGGGACGCCTGGCTTCCACTCAGTTTATCACTGGTCAGT
TAATATGGCGGGAGTATTTCTACACGATGAGCGTCAATAACCCCAACTACGCCA
35 AATGTCGGGGAATCCTATCTGCCTGGACATAACCGTGGAAGGAACCGGAAAATGAC
GAGTTACAGAGATGGAAGGAGGGTCGTACGGGGTTCCCATTCGTGGACGCGGCCA
TGCGCCAGCTGCGTACGGAGGGCTGGTTGCATCACGTTGTTTCGGAACACCGTGGCC
TCGTTCCCTACCCGCGGGACCCTGTGGCTGTCCTGGGAACACGGGCTGCAGCACTT
CCTCAAGTATCTGCTGGATGCTGATTGGTTCGGTGTGTGCGGGTAACTGGATGTGGG

TGTCGTCCAGTGCGTTCGAGGCCTTGTTGGACTCTGGCGAGTGCGCGTGTCCCGTC
AGACTGGGCCGAAGACTGGAGCCCACTGGCCATTATGTACGGAGATACGTACCAG
AACTGGCTCGGATGCCCCGAGAGTACATTTACGAGCCGTGGCGTGCCCCGCTCGA
GGTGCAGGAGGCTGCGGGCTGTGTTCATAGGTCGAGACTACCCCGCGCCGGTCTGTC
5 GACCACACAGCTGCGGCCGCCAGGAACAGGGCCAACATGCAGGAGCTGCGCCGCC
TGTTGGAGAAAGCTCCTCCTCACTGCTGTCCGTCATCTGAAGACGAGGTGCGGCAG
TTCATGTGGCTTGGAGACGACTCGCAGCCTGAGCTCACCACCACATGAGGATCC
(SEQ ID NO: 32)。

其中,第2-11个氨基酸为10个组氨酸的His亲和标签序列;第12-16个氨基酸 GGSTS
10 (SEQ ID NO: 25) 为链接区序列;带下划线的序列为蝴蝶的Cry1的氨基酸序列,即如
SEQ ID NO: 16所示的氨基酸序列。

核苷酸序列中,第1-6位是NdeI酶切位点,7-36位是10xHis表达序列,37-51位是
GGSTS (SEQ ID NO: 25) 表达序列,52-1653位是黑脉金斑蝶Cry1表达序列,1654-1656
15 位是终止密码子,1657-1662位是BamHI酶切位点。该基因在博迈德生物科技有限公司
进行合成所得。

2. 将上述合成的编码蝴蝶Cry1的氨基酸序列并且N端具有His亲和标签的插入片
段通过NdeI和BamHI位点克隆到pET-21a(+)质粒(Novagen)。

3. 将前述具有蝴蝶Cry1蛋白的编码基因并融合了组氨酸标签(His-tag)的质粒与
实施例10步骤2得到的具有蝴蝶MagR蛋白的编码基因并融合了StrepII标签的质粒共
20 同转化大肠杆菌BL21(DE3),表达MagR的质粒为卡那霉素抗性,表达Cry1的质粒为氯
苄青霉素抗性,在双抗培养基(卡那霉素和氯苄青霉素)中挑取含有两种质粒的大肠杆
菌单克隆接种至LB液体培养基中扩大培养,15℃下用20μM的IPTG诱导目的蛋白的表
达,20-24小时后收菌。

4. 将上述得到的共表达了蝴蝶MagR蛋白和Cry1蛋白的转化菌用缓冲液(20mM
25 Tris, 150mM NaCl, 10mM 咪唑, pH8.0, 10mM 巯基乙醇以及蛋白酶抑制剂)重悬后,
用超声破碎仪裂解,完成后进行高速离心,取上清液进行亲和纯化。通过Ni-NTA亲和
柱亲和纯化后,再通过Strep-Tactin亲和柱纯化,得到蝴蝶MagR蛋白和Cry1蛋白形成
的MagR-Cry1杂合蛋白复合物。进一步通过阴离子交换柱和分子筛凝胶层析纯化,得到
纯度较高的杂合蛋白复合物。纯化全过程在4℃和存在外加磁场的条件下进行。外加磁场
30 通过下述方法施加:当样品过亲和柱时,将两个相同的条形磁铁分别绑在亲和柱的两侧,
该两个相同的条形磁铁的南极和北极分别相对,提供一个均匀稳定的磁场。结果显示,
在纯化时施加外加磁场(外加磁场的强度是地球磁场的100倍),纯化蛋白的产率有显
著增加;在洗脱纯化蛋白时,撤去外加磁场,蛋白质洗脱效率更高。

5. 对收集的蝴蝶MagR-Cry1蛋白复合物样品进行变性聚丙烯酰胺凝胶电泳
35 (SDS-PAGE)分析。实验结果如图2所示,其中从左向右第2泳道加载的样品为蛋白质分
子量标准,第3泳道加载的样品为蝴蝶MagR-Cry1杂合蛋白复合物。结果:由于使用变
性电泳分析,泳道3中显示出两种蛋白质样品,即含有Strep II亲和标签的蝴蝶MagR
蛋白(分子量约为15KDa)以及含有His标签的蝴蝶Cry1蛋白(分子量约为70KDa)。

实施例13:黑脉金斑蝶MagR蛋白以及MagR纯合蛋白复合物含有铁的证明实

验

将实施例 11 得到的蝴蝶 MagR 蛋白的单体峰样品用分光光度计(Nanodrop 2000)进行全波长扫描。结果与图 3 相似。

5 将实施例 11 中得到的蝴蝶 MagR 蛋白形成的二聚物或多聚物样品, 用分光光度计(Nanodrop 2000)进行全波长扫描。结果与图 3 相似。

实施例 14: 黑脉金斑蝶 MagR 纯合蛋白复合物和黑脉金斑蝶 MagR-Cry1 杂合蛋白复合物的电子显微镜图像和结构

10 将实施例 11 中经分子筛凝胶层析纯化后获得的蝴蝶 MagR 纯合蛋白复合物的蛋白样品以及实施例 12 中经分子筛凝胶层析纯化后的蝴蝶 MagR-Cry1 杂合蛋白复合物的蛋白样品采用乙酸铀或者甲酸铀负染, 制备蛋白质的负染电镜样品, 将样品放置于透射电镜下观察并拍照, 按照标准的电镜结构分析方法解析其结构, 并测量大小(包括蛋白长度和宽度)。

15 电镜结果显示, 蝴蝶 MagR 纯合蛋白复合物和蝴蝶 MagR-Cry1 杂合蛋白复合物的结构形式与实施例 6 图 4 中观测到的家鸽 MagR 纯合蛋白复合物和 MagR-Cry4 杂合蛋白复合物的结构形式相同。

20 电镜观测到的结果显示, 蝴蝶 MagR 纯合蛋白复合物呈现为长条形的棒状结构, 长度为约 20-24 纳米长, 宽度为约 7-11 纳米。根据同源晶体结构和分子模型计算, 结果表明该复合物具有约 20 个 MagR 蛋白。电镜观测到的结果显示, 蝴蝶 MagR-Cry1 杂合蛋白复合物呈现为棒状, 长度为约 20-24 纳米长, 宽度为约 13-18 纳米宽。根据同源晶体结构和分子模型计算, 结果表明该复合物具有约 20 个 MagR 蛋白和 10 个 Cry1 蛋白。

25 电镜结果还显示, 每一个 MagR 纯合蛋白复合物包括约 20 个 MagR 蛋白质, 有序排列形成线性双螺旋聚合体, 其中每 10 个 MagR 蛋白单体以线性聚合的形式首尾相接, 形成两条蛋白长链, 两条蛋白长链以螺旋形式结合, 由此形成长条形棒状结构的蛋白复合物。

电镜结果还显示, 在蝴蝶 MagR-Cry1 杂合蛋白复合物中, 在由约 20 个 MagR 组成的蛋白复合物所形成的棒状线性双螺旋聚合体的外围, 有约 10 个感光的 Cry1 蛋白质缠绕排列, 从而形成杂合蛋白复合物。

实施例 15: 黑脉金斑蝶 MagR 蛋白复合物结晶

30 将实施例 11 中经分子筛凝胶层析纯化后获得的蝴蝶 MagR 纯合蛋白复合物的蛋白样品以及实施例 12 中经分子筛凝胶层析纯化后的蝴蝶 MagR-Cry1 杂合蛋白复合物的蛋白样品浓缩, 然后进行结晶。蛋白质结晶方法按照如教科书 McPherson, A. (2009). *Introduction to macromolecular crystallography*. Hoboken, N.J., Wiley-Blackwell 记载的标准实验流程采用悬滴法进行。

35 蝴蝶 MagR 纯合蛋白复合物的晶体和蝴蝶 MagR-Cry1 杂合蛋白复合物的晶体的形态与对应的家鸽 MagR 纯合蛋白复合物的晶体或家鸽 MagR-Cry4 杂合蛋白复合物的晶体类似, 分别制备和观察到两种类型的晶体: 第一种是黑色不透明的长方体形状的晶体, 提示铁元素在晶体中已被氧化成均一的三价; 第二种晶体为浅黄色或黄色的透明晶体, 形状不规则或呈小片状, 提示铁元素可能以二价存在。图 11 示出了

蝴蝶 MagR-Cry1 杂合蛋白复合物的晶体照片, 可见该晶体具有黑色不透明的长方体形状, 长度约为 0.5mm。

实施例 16: 黑脉金斑蝶 MagR 蛋白复合物的磁性测试

采用与实施例 8 中同样的方法对蝴蝶 MagR 蛋白形成的纯合蛋白复合物以及蝴蝶
5 MagR-Cry1 杂合蛋白复合物的磁性进行测试。结果显示, 蝴蝶 MagR 纯合蛋白复合物和蝴蝶 MagR-Cry1 杂合蛋白复合物的磁性测试结果与家鸽的测试结果 (如实施例 8 中观察到的) 类似。这表明, 蝴蝶 MagR 纯合蛋白复合物和蝴蝶 MagR-Cry1 杂合蛋白复合物具有磁性。

图 11 为蝴蝶 MagR 纯合蛋白复合物晶体在旋转磁场中同步旋转实验结果图。实验方
10 法如实施例 8 的 (3) 所描述。如图 11 所示, 可以观察到蝴蝶 MagR-Cry1 纯合蛋白复合物的晶体为黑色, 并且在旋转磁场中发生同步旋转。同样, 蝴蝶 MagR 杂合蛋白复合物的晶体在旋转磁场中也出现同步旋转的现象。

实施例 17: 家鸽 MagR 蛋白以及 MagR 纯合蛋白复合物具有结合铁的保守结构域的证明实验

15 根据实施例 2、3 和 5 描述的方法和步骤, 测试将家鸽 MagR 的如 SEQ ID NO: 1 所示的氨基酸序列中的三个半胱氨酸 C60、C124、C126 突变为丙氨酸 (A, 密码子为 GCC) 后得到的蛋白结合铁的能力。

即将实施例 2 步骤 1 中人工合成的插入片段所编码的氨基酸序列改变为:

20 MWSHPQFEKGGGTSMASASSVVRATVRAVSKRKIQATRAALTTPSAVQKIKEL
LKDKPEHVGKVGVRTRGANGLSYTLEYTKSKGDSDEEVVQDGVRVFIEKKAQLTLL
GTEM DYVEDKLSSEFVFNNPNIKGTAGAGESFNI (SEQ ID NO: 33)。

根据实施例 2 和 3 描述的方法和步骤表达和纯化具有上述 C60A、C124A、C126A 突变的家鸽 MagR 蛋白以及 MagR 纯合蛋白复合物。突变的家鸽 MagR 蛋白经过分子筛凝胶层析, 也呈现出单体峰、二聚体峰和多聚体峰。分别收集三个峰的样品。将
25 得到的蛋白样品分别根据实施例 5 描述的方法用分光光度计 (Nanodrop 2000) 进行全波长扫描。

如图 12 所示, 结果显示上述突变蛋白形成的家鸽 MagR 单体蛋白的吸收曲线 (实线所示) 在 Fe 元素的特征吸收位置 (320 nm 和 410 nm) 没有吸收峰。作为对照的野生型的家鸽 MagR 蛋白以及 MagR 纯合蛋白复合物的吸收曲线 (虚线所示) 则显
30 示在 320 nm 和 410 nm 有明显的吸收峰。类似的, 根据前述方法经过分子筛凝胶层析和纯化后得到的突变蛋白形成的家鸽 MagR 的二聚物或多聚物的吸收曲线在 Fe 元素的特征吸收位置 (320 nm 和 410 nm) 也没有吸收峰。

结果表明, 家鸽 MagR 蛋白具有结合铁的结构域, 其包括 C60、C124、C126 三个位点中的一个或多个。

35 实施例 18: 各物种 MagR 蛋白的序列比对

测序或根据 NCBI 数据库获得果蝇 (*Drosophila melanogaster*)、黑脉金斑蝶 (*Danaus plexippus*)、蜜蜂 (*Apis mellifera*)、文昌鱼 (*Branchiostoma floridae*)、斑马鱼 (*Danio rerio*)、蟒蛇 (*Python bivittatus*)、家鸽 (*Columba livia*)、草雀 (*Taeniopygia guttata*)、鸡 (*Gallus gallus*)、鲸鱼 (*Balaenoptera acutorostrata*)、蝙蝠 (*Myotis lucifugus*)、裸鼹鼠 (*Heterocephalus glaber*)、

小鼠(*Mus musculus*)、人(*Homo sapiens*)的磁感应受体蛋白的氨基酸序列,其分别如 SEQ ID NO: 3、SEQ ID NO: 2、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 1、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 所示。

5 用 Clustal V 比对方法对获得的氨基酸序列进行比对。

图 13 是上述各物种 MagR 蛋白的序列比对的结果。其中从上至下的来源物种名称以及基因在 NCBI 的序列号(如果有的话)分别为:(1)果蝇(*Drosophila melanogaster*), 基因序列号为 NP_573062.1; (2)黑脉金斑蝶(*Danaus plexippus*); (3)蜜蜂(*Apis mellifera*), 基因序列号为 XP_624993.1; (4)文昌鱼(*Branchiostoma floridae*), 基因序列号为 XP_002589524.1; (5)斑马鱼(*Danio rerio*), 基因序列号为 NP_001020349.1; (6)蟒蛇(*Python bivittatus*), 基因序列号为 XP_007429307.1; (7)家鸽(*Columba livia*); (8)草雀(*Taeniopygia guttata*), 基因序列号为 XP_002194930.1; (9)鸡(*Gallus gallus*), 基因序列号为 XP_003643055.1; (10)鲸鱼(*Balaenoptera acutorostrata*); (11)蝙蝠(*Myotis lucifugus*), 基因序列号为 XP_006102189.1; (12)裸鼹鼠(*Heterocephalus glaber*), 基因序列号为 XP_004879403.1; (13)小鼠(*Mus musculus*), 基因序列号为 NP_081197.1; (14)人(*Homo sapiens*), 基因序列号为 NP_112202.2。其中(2)黑脉金斑蝶、(7)家鸽和(10)鲸鱼的序列为本申请发明人测序获得。

结果显示,本申请发明人首次研究、制得、鉴定并命名的磁感应受体蛋白的核酸编码序列在生物体基因组中广泛存在。本发明提供的磁感应受体蛋白的核酸序列在大量不同物种的基因组中都存在,其编码的蛋白序列高度保守,并具有多个高度保守的结构域,如结合铁和/或硫的结构域,在图 13 中箭头所指的三个 100%保守的半胱氨酸残基(C),其为结合铁(或铁硫中心)的位置。

除非另外指出,本发明的实施方式可使用生物技术、有机化学、无机化学等的常规技术,显然除在上述说明和实施例中所特别描述之外,还可以别的方式实现本发明。其它在本发明范围内的方面与改进将对本发明所属领域的技术人员是显而易见的。根据本发明的教导,许多改变和变化是可行的,因此其在本发明的范围之内。本文所提到的所有专利、专利申请与科技论文均据此通过引用并入本文中。

权利要求书

1. 一种分离的磁感应受体蛋白或其活性片段或衍生物, 优选地, 所述磁感应受体蛋白来自脊椎动物或无脊椎动物; 进一步优选地, 所述脊椎动物为家鸽、斑马鱼、蟒蛇、草雀、鸡、鲸鱼、蝙蝠、裸鼹鼠、小鼠或人, 所述无脊椎动物为果蝇、文昌鱼、蜜蜂或黑脉金斑蝶。

2. 权利要求 1 所述的磁感应受体蛋白或其活性片段或衍生物, 其中, 参照图 13 所示的氨基酸位置, 所述磁感应受体蛋白或其活性片段或衍生物的氨基酸序列从 N 端到 C 端依次包含以下结构域: 第 10-13 位的 TVRA、第 56-61 位的 RGCNGL、第 80-83 位的 QDGV、第 89-93 位的 KKAQL、第 95-101 位的 LLGTEMD 以及第 115-128 位的 NPNIKGTCGCGESF。

3. 权利要求 2 所述的磁感应受体蛋白或其活性片段或衍生物, 还包含以下结构域: 第 17 位的 R、第 23 位的 R、第 25-26 位的 AL、第 28-29 位的 LT、第 32-33 位的 AV、第 48 位的 G、第 51 位的 V、第 53 位的 V、第 63 位的 Y、第 67 位的 Y、第 71 位的 K、第 75-76 位的 DE、第 78 位的 V、第 87 位的 I、第 104 位的 E、第 106-107 位的 KL、第 110-111 位的 EF 和第 113 位的 F。

4. 权利要求 1-3 中任一项所述的磁感应受体蛋白或其活性片段或衍生物, 其具有 2Fe2S 的铁硫中心。

5. 权利要求 1-4 中任一项所述的磁感应受体蛋白或其活性片段或衍生物, 其中所述磁感应受体蛋白或其活性片段或衍生物具有:

(1) 如 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 所示的氨基酸序列; 或

(2) 在与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 进行比较时具有至少 75% 的序列同一性, 优选具有至少 90% 的序列同一性, 最优选具有至少 95% 的序列同一性的氨基酸序列; 或

(3) 在与 SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13 或 SEQ ID NO: 14 进行比较时具有一个或多个氨基酸的取代、缺失、添加的氨基酸序列。

6. 一种分离的蛋白复合物, 包含两个或两个以上如权利要求 1-5 中任一项所述的磁感应受体蛋白或其活性片段或衍生物。

7. 权利要求 6 所述的蛋白复合物, 其中所述蛋白复合物包含 4 个或以上所述磁感应受体蛋白或其活性片段或衍生物, 优选为 20-24 个所述磁感应受体蛋白或其活性片段或衍生物, 更优选为 20 个所述磁感应受体蛋白或其活性片段或衍生物。

8. 权利要求 7 所述的蛋白复合物, 其中, 所包含的多个磁感应受体蛋白或其活性片

段或衍生物线性聚合，形成两条蛋白长链并以螺旋形式结合，形成棒状结构的复合物。

9. 权利要求 8 所述的蛋白复合物，其具有多个铁原子，多个铁原子在棒状结构的内部近中心轴区域形成有序排列。

10. 生产权利要求 6-9 中任一项所述的蛋白复合物的方法，其包含下述步骤：

5 a. 在允许形成蛋白复合物的条件下，使所述磁感应受体蛋白或其活性片段或衍生物聚合，形成同源聚合物；

b. 分离该同源聚合物；优选地，该步骤在存在磁场的条件下进行。

11. 一种分离的杂合蛋白复合物，其包含 1) 如权利要求 1-5 中任一项所述的磁感应受体蛋白或其活性片段或衍生物，以及 2) 光感应蛋白或其活性片段或衍生物，每 2 个所述磁感应受体蛋白或其活性片段或衍生物与 1 个光感应蛋白或其活性片段或衍生物结合。

12. 权利要求 11 所述的杂合蛋白复合物，其中，所述光感应蛋白为隐花色素蛋白；优选地，所述隐花色素蛋白具有：

(1) 如 SEQ ID NO: 15 或 SEQ ID NO: 16 所示的氨基酸序列；或

15 (2) 在与 SEQ ID NO: 15 或 SEQ ID NO: 16 进行比较时具有至少 90%、91%、92%、93%、94%、95%、96%、97%、98% 或 99% 的序列同一性的氨基酸序列，并且具有隐花色素蛋白的活性；或

(3) 在与 SEQ ID NO: 15 或 SEQ ID NO: 16 进行比较时具有一个或多个氨基酸的取代、缺失、添加的氨基酸序列。

13. 权利要求 11-12 中任一项所述的杂合蛋白复合物，其中所述杂合蛋白复合物包含 4 个或以上所述磁感应受体蛋白或其活性片段或衍生物，优选包含 20-24 个所述磁感应受体蛋白或其活性片段或衍生物，更优选包含 20 个所述磁感应受体蛋白或其活性片段或衍生物。

14. 一种生产权利要求 11-13 中任一项所述的杂合蛋白复合物的方法，包括以下步骤：

25 a. 在允许形成蛋白复合物的条件下，使 1) 如权利要求 1-5 中任一项所述的磁感应受体蛋白或其活性片段或衍生物与 2) 光感应蛋白或其活性片段或衍生物接触，形成杂合聚合物，其中 1) 和 2) 的蛋白中的至少一种是分离形式或被重组表达；

b. 分离该杂合聚合物；优选地，该步骤在存在磁场的条件下进行。

15. 权利要求 1-5 中任一项所述的磁感应受体蛋白或其活性片段或衍生物或权利要求 6-9 中任一项所述的蛋白复合物或权利要求 11-13 中任一项所述的杂合蛋白复合物作为磁性材料或在制备磁性材料中的应用。

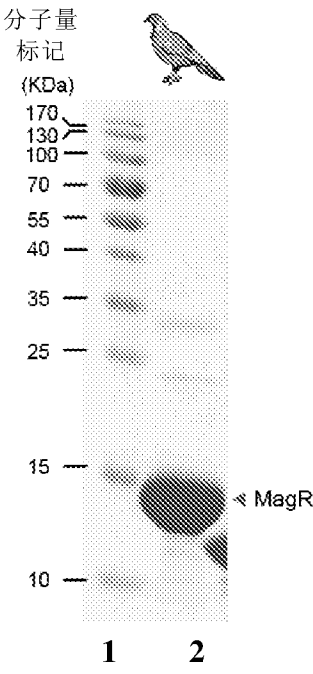


图 1

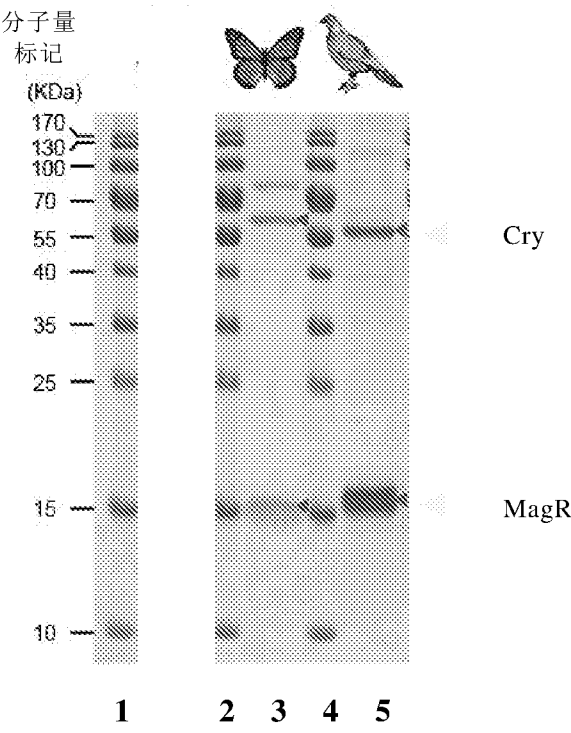


图 2

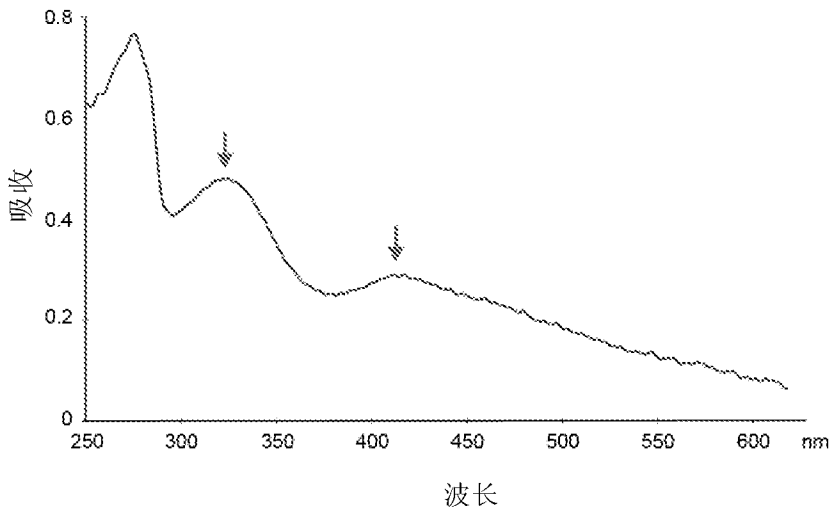


图 3

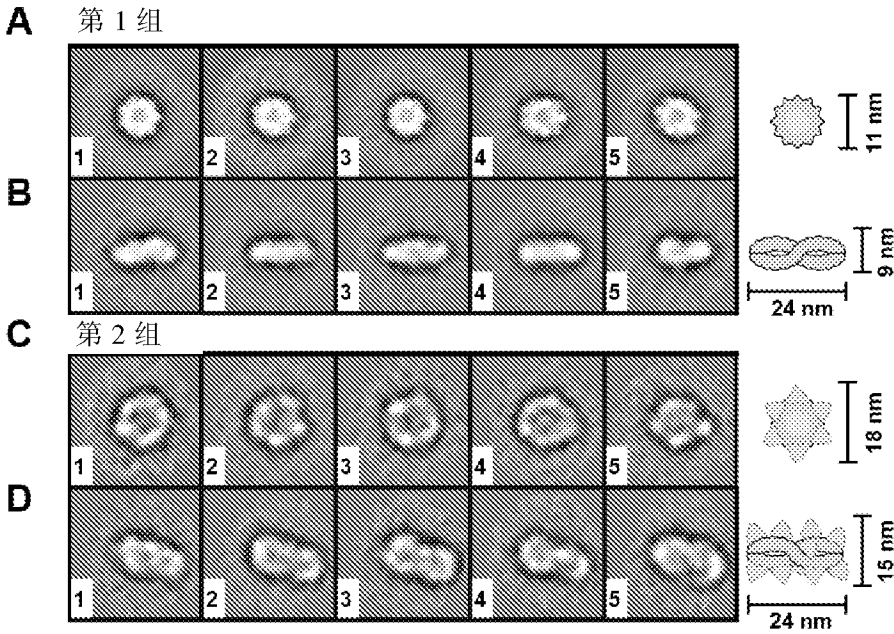


图 4



图 5

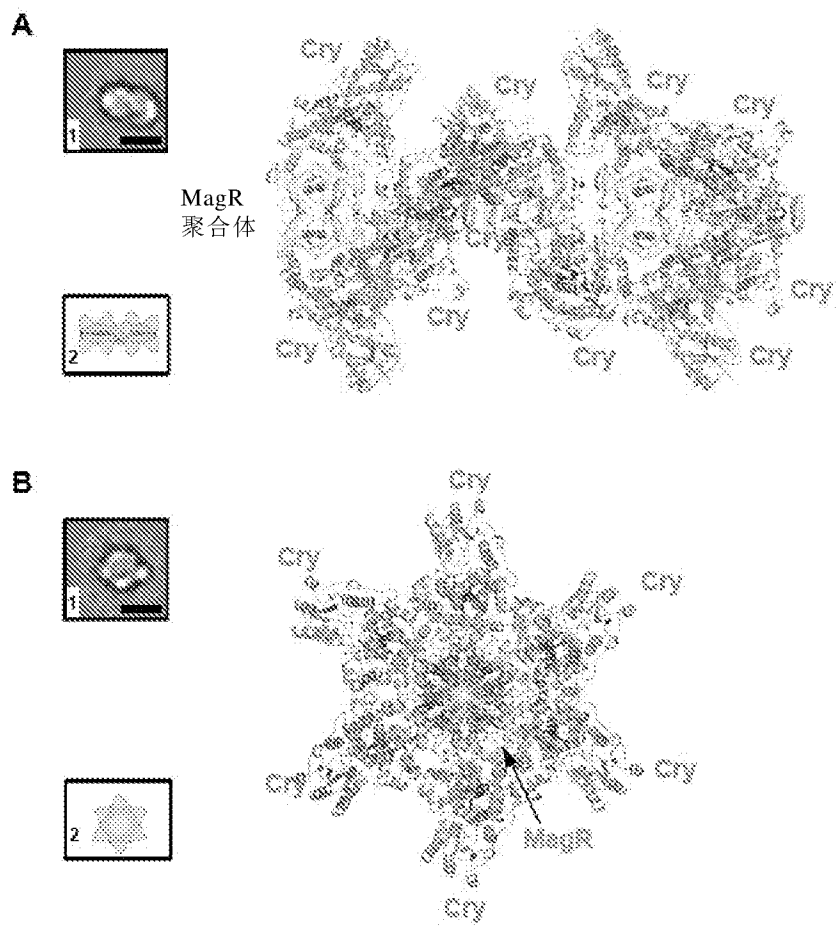


图 6

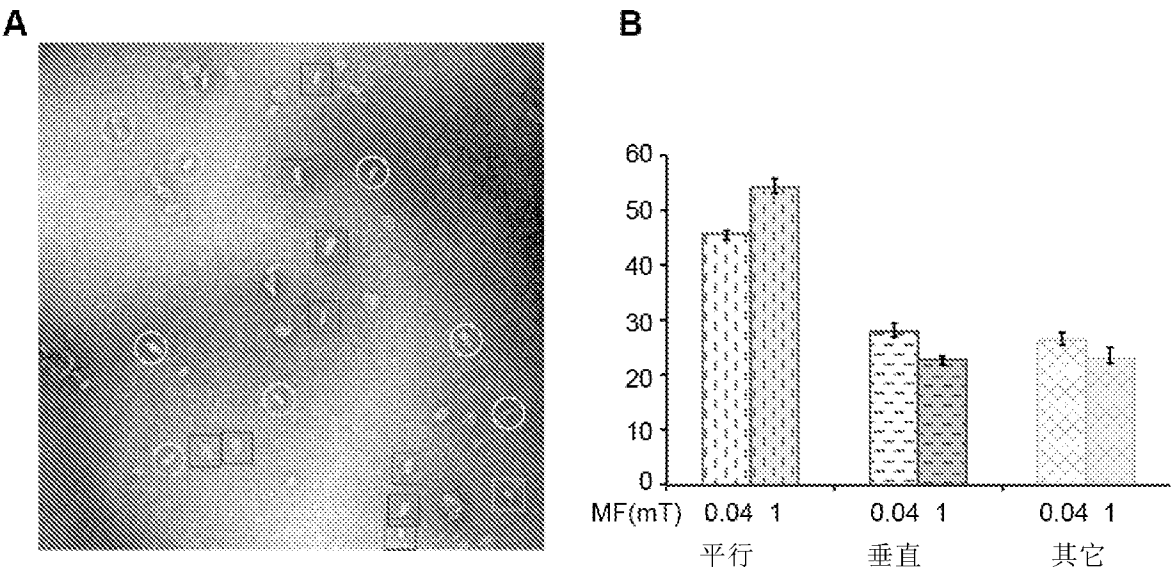


图 7

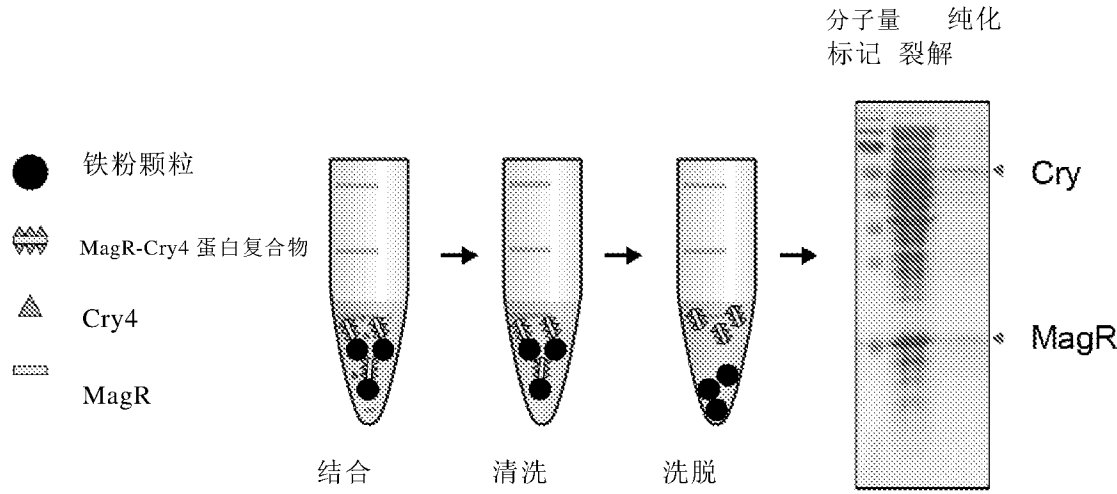


图 8

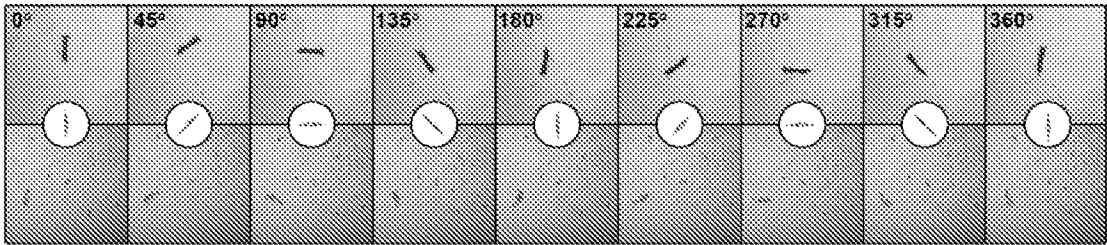


图 9

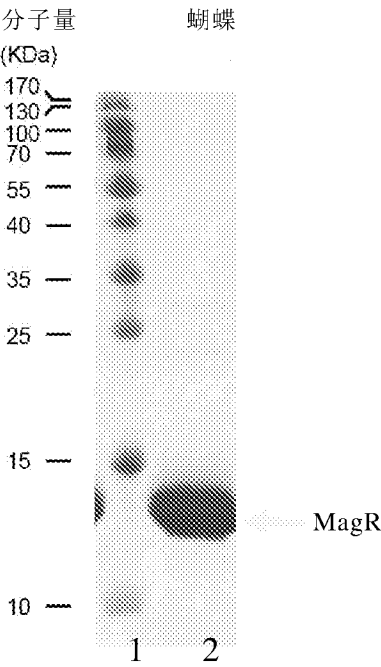


图 10

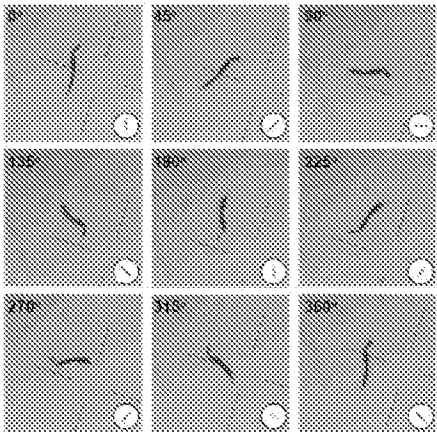


图 11

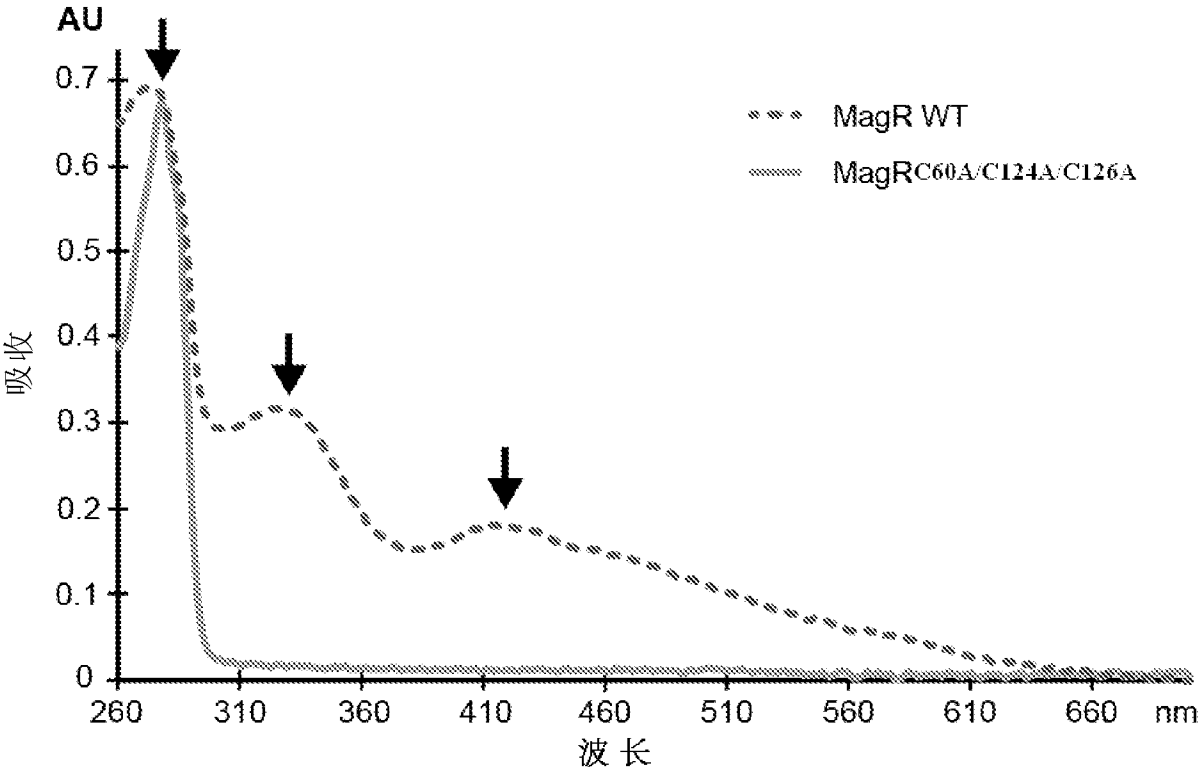


图 12

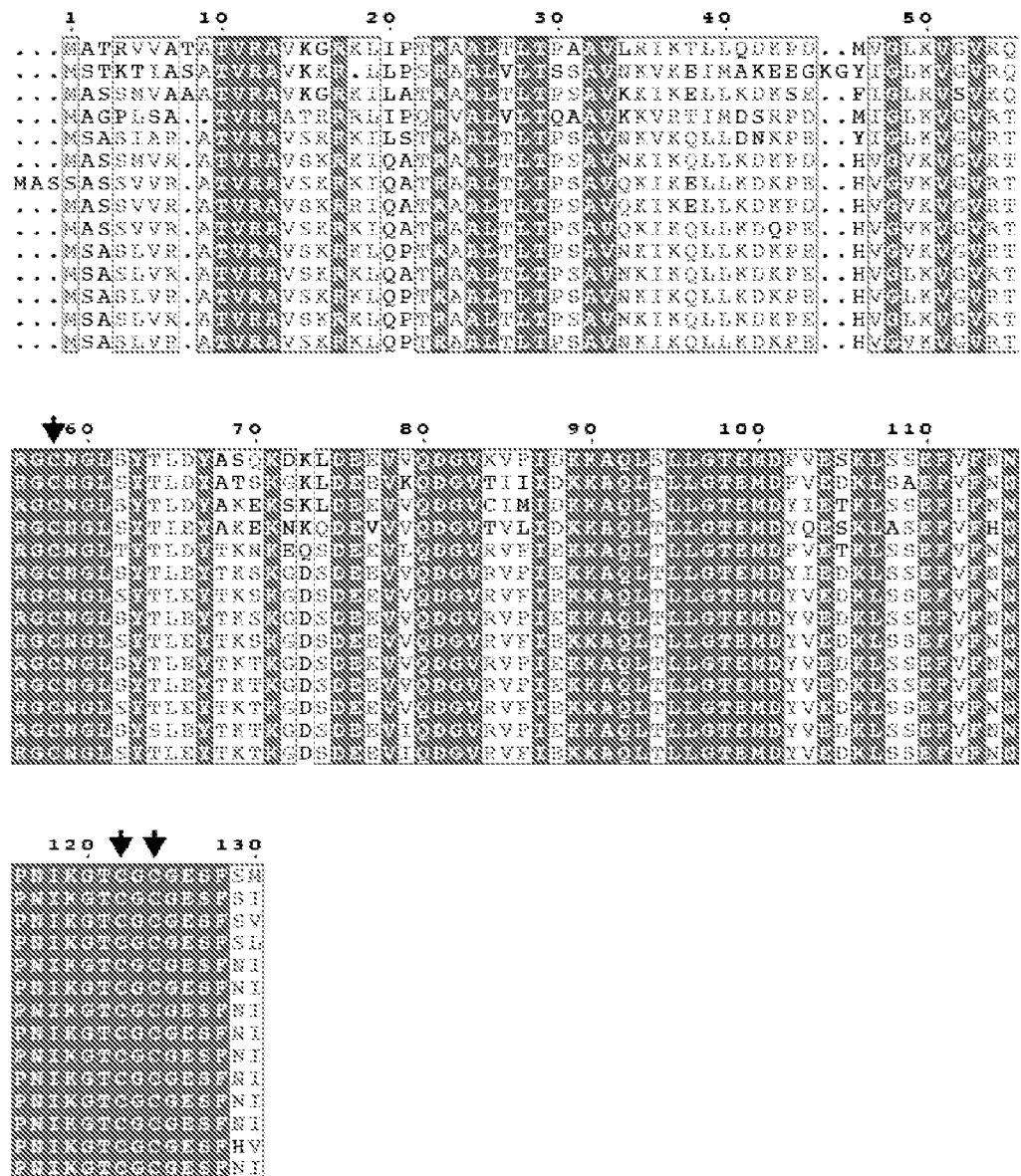


图 13

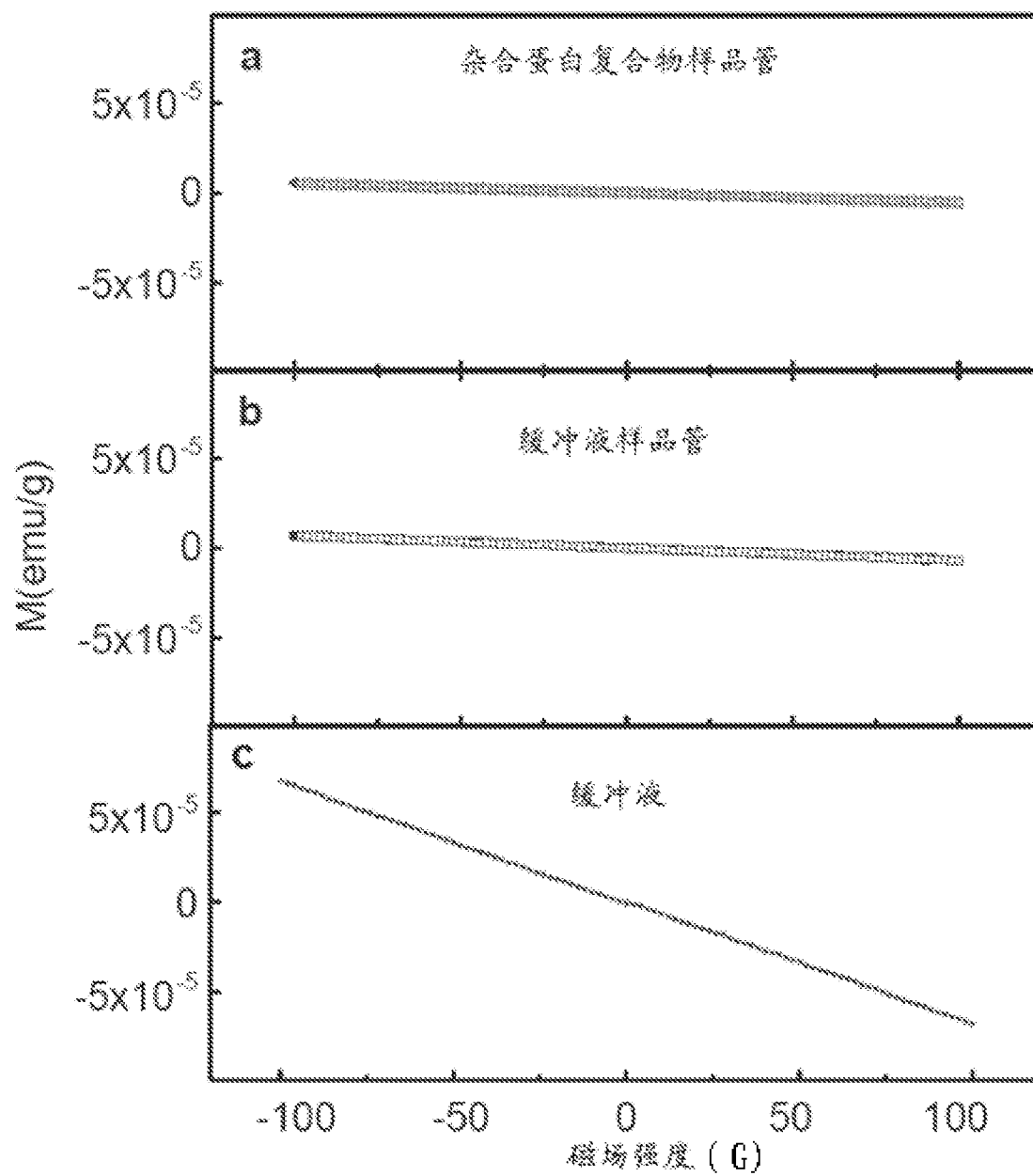


图 14

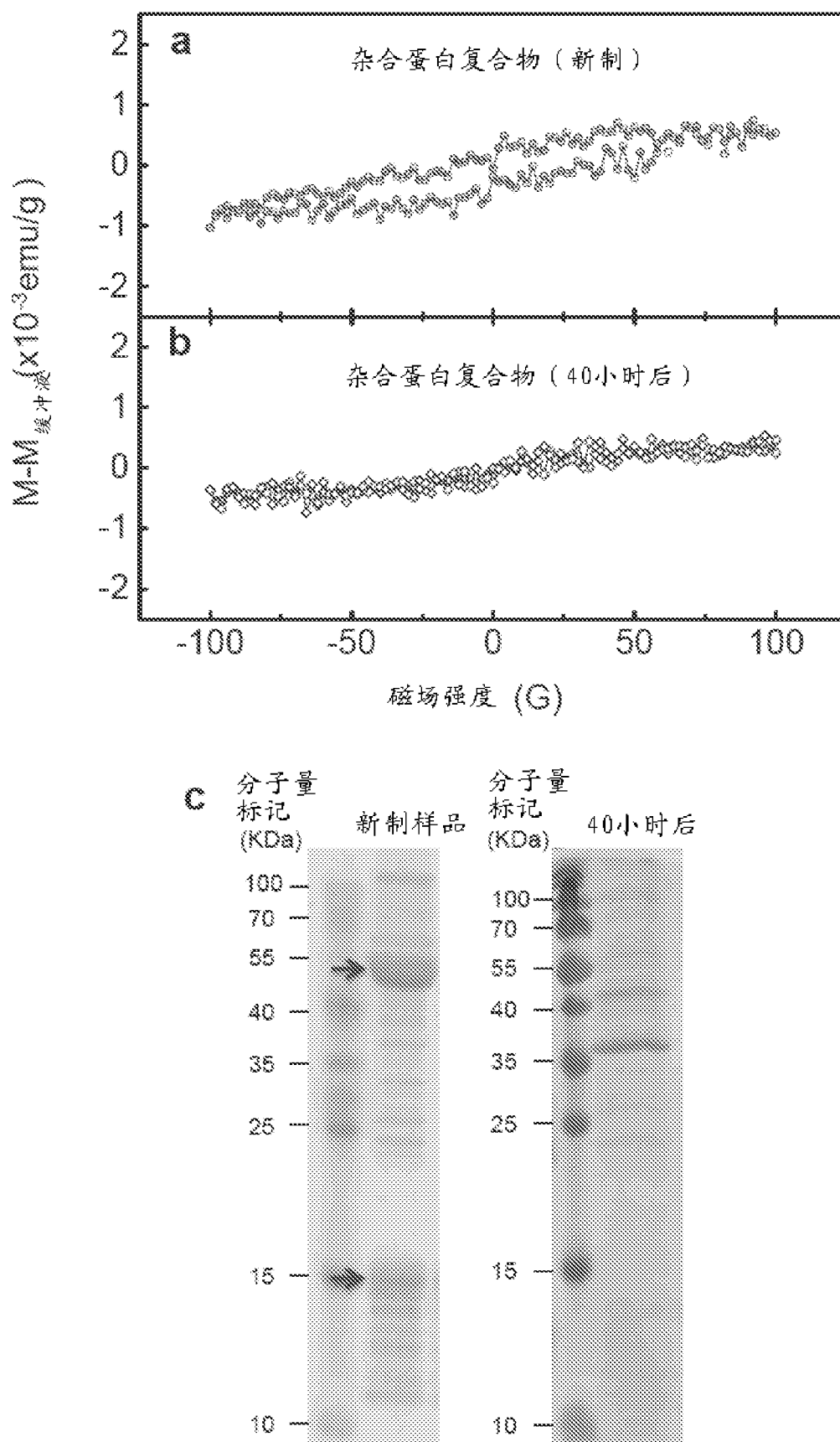


图 15

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/078842

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:
 - a. (means)
 - ☐ on paper
 - ☒ in electronic form
 - b. (time)
 - ☒ in the international application as filed
 - ☐ together with the international application in electronic form
 - ☐ subsequently to this Authority for the purposes of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/078842

A. CLASSIFICATION OF SUBJECT MATTER

C07K 14/705 (2006.01) i; H01F 1/42 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K; H01F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNKI, DWPI, ISI Web of knowledge and Key words: magnet+, receptor?, protein, peptide, iron sulfur, cryptochrome, etc.
EMBL, Genbank and SEQ ID NOs: 1-14 of Amino Acid Sequences

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"GenBank Accession Number XP_005154986", VERSION XP_005154986.1, 26 July 2013(26.07.2013), [retrieved on 06 July 2010 (06.07.2016)], retrieved from NCBI [on-line]: <URL: http://www.ncbi.nlm.nih.gov/protein/527274488 >, see the amino acid sequence	1-5
X	"GenBank Accession Number XP_004926440", VERSION XP_004926440.1, 26 July 2013 (26.07.2013), [retrieved on 06 July 2010 (06.07.2016)], retrieved from NCBI [on-line]: <URL: http://www.ncbi.nlm.nih.gov/protein/512907435 >, see the amino acid sequence	1-5
A	RITZ, T. et al., "Cryptochrome: A Photoreceptor with the Properties of a Magnetoreceptor", COMMUNICATIVE & INTEGRATIVE BIOLOGY, 31 January 2010 (31.01.2010), ISSN: 1942-0889, pages 24-27	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 07 July 2016 (07.07.2016)	Date of mailing of the international search report 20 July 2016 (20.07.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer Jia, Tao Telephone No.: (86-10) 62411993

国际检索报告

国际申请号

PCT/CN2016/078842

A. 主题的分类

C07K 14/705(2006.01)i; H01F 1/42(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C07K; H01F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS, CNKI, DWPI, ISI Web of knowledge和关键词: 磁, 受体, 蛋白, 肽, 铁硫, 隐花色素, magnet+, receptor?, protein, peptide, iron sulfur, cryptochrome, 等 EMBL, Genbank和氨基酸序列SEQ ID NOs:1-14

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	"GenBank登录号XP_005154986" 版本XP_005154986.1, 2013年 7月 26日 (2013 - 07 - 26), [检索于06.7月2010 (06.07.2016)], 检索自NCBI [联机]:<URL: http://www.ncbi.nlm.nih.gov/protein/527274488 >, 见氨基酸序列	1-5
X	"GenBank登录号XP_004926440" 版本XP_004926440.1, 2013年 7月 26日 (2013 - 07 - 26), [检索于06.7月2010 (06.07.2016)], 检索自NCBI [联机]:<URL: http://www.ncbi.nlm.nih.gov/protein/512907435 >, 见氨基酸序列	1-5
A	R. Thorsten等. "Cryptochrome: A photoreceptor with the properties of a magnetoreceptor" Communicative & integrative biology, 第3卷, 第1期, 2010年 1月 31日 (2010 - 01 - 31), ISSN: 1942-0889, 第24-27页	1-15

☐ 其余文件在C栏的续页中列出。☐ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2016年 7月 7日

国际检索报告邮寄日期

2016年 7月 20日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

贾涛

电话号码 (86-10)62411993

第I栏

核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何对要求保护的发明必要的核苷酸和/或氨基酸序列，国际检索是在下列基础上进行的：

a. (提交提供)

☐ 纸件形式

☒ 电子形式

b. (提交时间)

☒ 含在申请提交时的国际申请中

☐ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的申请中的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见：