

人、小鼠、大鼠血红蛋白及其编码基因序列分析

姓名 陈耿佳 学号 1301214752 组号 G01C

1. 认真阅读 NCBI 书架 (Bookshelf)、PDB 分子月报 (Molecule of the Month)、UniProt 蛋白质精选 (Protein Spotlight) 以及维基百科 (Wikipedia) 等网站中有关血红蛋白的介绍, 了解血红蛋白的生理功能、空间结构、亚基组成等基本知识。

血红蛋白是高等生物体内负责运载 O_2 的功能蛋白质, 存在于几乎所有的脊椎动物体内, 在某些无脊椎动物组织也有分布。血红蛋白的特性是: 在 O_2 含量高的地方, 容易与 O_2 结合; 在 O_2 含量低的地方, 又容易与 O_2 分离。血红蛋白的这一特性, 使红细胞具有运输 O_2 的功能。血液中的血红蛋白从呼吸器官中将 O_2 运输到身体其他部位释放, 以满足机体氧化营养物质支持功能运转的需要, 并将由此生成的 CO_2 带回呼吸器官中以排出体外。平均每克血红蛋白可结合 1.34ml 的 O_2 , 是血浆溶氧量的 70 倍。一个哺乳动物血红蛋白分子可以结合最多四个 O_2 分子。血红蛋白也参与其他气体的转运: 它能携带机体的部分 O_2 (大约 10%)。亦可将重要的调节分子 NO 结合在球状蛋白的某个硫醇基团上, 在释放 O_2 的同时将其释放。在红细胞及其祖系细胞以外也发现了血红蛋白, 包括黑质中的 A9 多巴胺神经元、巨噬细胞、肺泡细胞以及肾脏中的系膜细胞。在这些组织中, 血红蛋白作为抗氧化剂和铁代谢 (iron metabolism) 的调节因子存在。

血红蛋白化学式为 $C_{3032}H_{4816}O_{812}N_{780}S_8Fe_4$ 。人体内的血红蛋白由四个亚基构成, 分别为两个 α 亚基和两个 β 亚基, 如图 1 所示, 红色、蓝色分别代表 α 、 β 亚基的肽链, 每条多肽链都由 8 个 α 螺旋组成, 在与人体环境相似的电解质溶液中血红蛋白的四个亚基可以自动组装成 $\alpha_2\beta_2$ 的四聚体。血红蛋白整个分子具有 222 对称性, 即沿 A 轴、B 轴和 C 轴旋转 180° 角, $\alpha\beta$ 对都与另一 $\alpha\beta$ 对相重合。

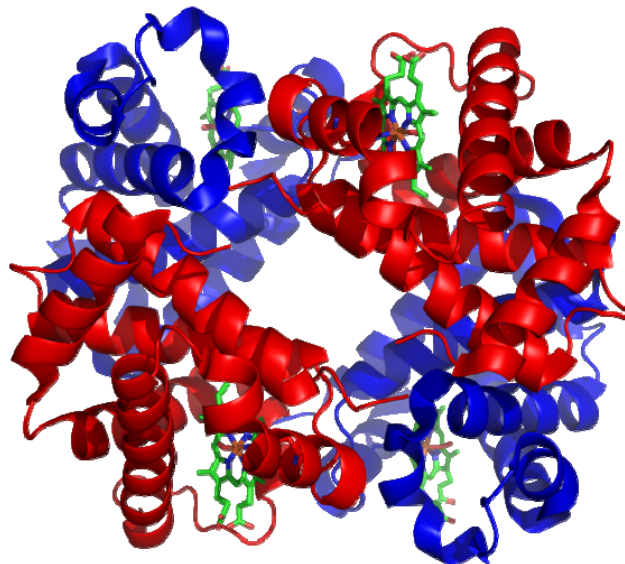


图 1 人血红蛋白空间结构

(图片来自 <http://en.wikipedia.org/wiki/Hemoglobin>)

血红蛋白的每个亚基由一条肽链和一个血红素分子构成, 肽链在生理条件下会盘绕折叠成球形, 把血红素分子抱在里面, 这条肽链盘绕成的球形结构又被称为珠蛋白。血红素分子是一个具有卟啉结构的小分子, 在卟啉分子中心, 由卟啉中四个吡咯环上的氮

原子与一个亚铁离子配位结合，珠蛋白肽链中第 8 位的一个组氨酸残基中的吡咯侧链上的氮原子从卟啉分子平面的上方与亚铁离子配位结合，当血红蛋白不与 O₂ 结合的时候，有一个水分子从卟啉环下方与亚铁离子配位结合，而当血红蛋白载 O₂ 的时候，就由 O₂ 分子顶替水的位置。

2. 查阅 ENSEMBL 基因组数据库中已经或正在进行基因组测序的物种树，了解人、小鼠、大鼠三个物种之间演化关系；检索物种分歧时间数据库 TimeTree，了解人和小鼠、小鼠和大鼠之间的分歧时间。

检索 TimeTree:

人类 (*Homo sapiens*) 和小鼠 (*Mus musculus*) 物种分歧时间为 9.23 千万年前；

人类 (*Homo sapiens*) 和大鼠 (*Rattus norvegicus*) 物种分歧时间为 9.23 千万年前；

小鼠 (*Mus musculus*) 和大鼠 (*Rattus norvegicus*) 物种分歧时间为 2.54 千万年前。

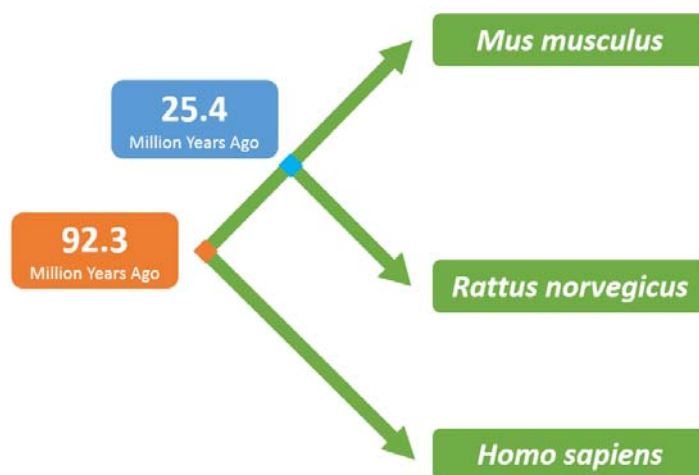


图 2 人、小鼠、大鼠三个之间的演化关系及物种分歧时间

3. 从 UniProt 数据库中检索并提取人、小鼠、大鼠血红蛋白 alpha 亚基蛋白质序列，利用 WebLab 或 JEMBOSS 软件包中的序列比对程序 Needle，选择默认计分矩阵 BLOSUM62 和默认空位罚分进行序列比对，将比对结果填入表 1。分析比对结果，说明得到上述结果的原因和进一步分析思路。

表 1 人、小鼠、大鼠血红蛋白 alpha 亚基蛋白质序列比对结果

| 物种 Species | 登录号 Accession | 得分 Score | 相同氨基酸 Identity | 相同和相似氨基酸 Similarity | 空位 Gaps |
|---------------|-------------------------|-------------|--------------------|------------------------|-----------------|
| 人/小鼠 | NP_000549/ NP_032244 | 653.0 | 123/142 (86.6%) | 131/142 (92.3%) | 0/142 (0.0%) |
| 人/大鼠 | NP_000549/ NP_037228 | 587.0 | 111/142 (78.2%) | 120/142 (84.5%) | 0/142 (0.0%) |

| | | | | | |
|-------|-------------------------|-------|--------------------|-----------------|-----------------|
| 小鼠/大鼠 | NP_032244/ NP_037228 | 632.0 | 120/142 (84.5%) | 127/142 (89.4%) | 0/142 (0.0%) |
|-------|-------------------------|-------|--------------------|-----------------|-----------------|

4. 检索 RefSeq 数据库中人、小鼠、大鼠血红蛋白 alpha 亚基编码基因的 mRNA 序列，下载其编码区序列和所编码的氨基酸序列，进行序列比对，将结果填入下面表 2 中。

表 2 人、小鼠、大鼠血红蛋白编码基因序列比对结果

| 物种 | 基因名 | 编码区序列 登录号 | 编码区序列 差异位点 | 氨基酸序列 登录号 | 氨基酸序列 差异位点 |
|----|---------------|----------------------------|---------------|----------------------------|---------------|
| 人 | HBA1/HBA2 | NM_000558/ NM_000517 | 0/429(0.0%) | NP_000549/ NP_000508 | 0/142(0.0%) |
| 小鼠 | Hba-a1/Hba-a2 | NM_008218/ NM_001083955 | 1/429(0.2%) | NP_032244/ NP_001077424 | 0/142(0.0%) |
| 大鼠 | Hba1/Hba2 | NM_013096/ NM_001007722 | 5/429(1.2%) | NP_037228/ NP_001007723 | 1/142(0.7%) |

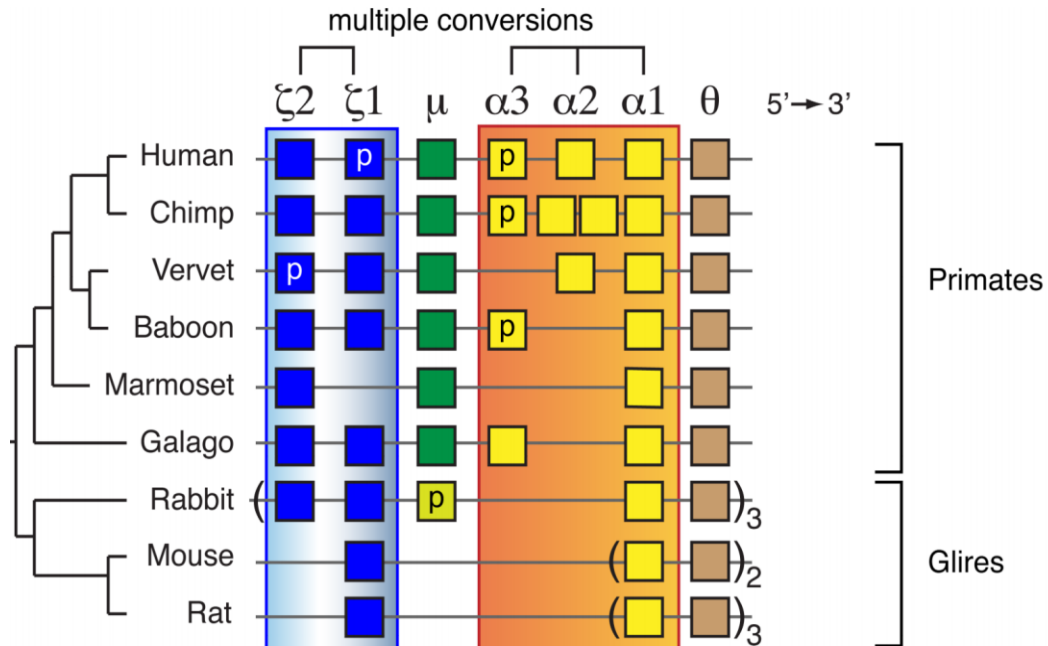
5. 对上述人 HBA1、小鼠 Hba1-a1、大鼠 Hba1 编码区序列进行序列比对，将比对结果填入表 3。注意选择恰当的计分矩阵和空位罚分值。

表 3 人、小鼠、大鼠血红蛋白编码区序列比对结果

| 物种 | 基因名 | 编码区序列 登录号 | 编码区序列 差异位点 | 氨基酸序列 差异位点 |
|-------|-----------------|-------------------------|---------------|---------------|
| 人/小鼠 | HsHBA1/MmHba-a1 | NM_000558/ NM_008218 | 79/429(18.4%) | 19/142(13.4%) |
| 人/大鼠 | HsHBA1/RnHba1 | NM_000558/ NM_013096 | 90/429(21.0%) | 31/142(21.8%) |
| 小鼠/大鼠 | MmHba-a1/RnHba1 | NM_008218/ NM_013096 | 46/429(10.7%) | 22/142(15.5%) |

采用默认计分矩阵 EBLOSUM62 和默认空位罚分 (10) 进行序列比对，系统优先采取错位匹配方式获取最高得分，导致出现空位。提高空位罚分，更改为 15 (或 20、30 等以上数值) 后，则不再出现空位。因为空位罚分值增大后，系统不再优先采用错位方式进行比对，这样可以提高序列比对的准确性。

6. 分析上述人、小鼠、大鼠血红蛋白基因区序列比对结果，阅读血红蛋白基因家族演化论文 (Hardison R, 2012)，说明人、小鼠和大鼠 alpha 血红蛋白基因家族基因结构和演化。



如上图所示，人的 α 血红蛋白基因由 $\alpha 1$ 、 $\alpha 2$ 两个旁系同源基因组成，由演化过程中复制产生。小鼠 α 血红蛋白基因存在两个重复，均对应人的 $\alpha 1$ 亚基。而大鼠 α 血红蛋白基因存在三个重复，也对应于人的 $\alpha 1$ 亚基。人和小鼠、大鼠在 α 血红蛋白基因上的演化很保守，发生变异的位点很少，这与上面的序列比对结果一致。

7. 将上述数据库检索、序列比对、结果分析的方法、思路、策略用于和你研究课题相关或你熟悉的蛋白质及其编码序列，分析比对结果，说明原因。

研究课题主要关注的蛋白质是 MG53 (Mitsugumin-53)，也称为 TRIM72 (Tripartite motif-containing protein 72)。是一种新发现的骨骼肌、心肌特异性表达的多功能蛋白质。已有的研究表明 MG53 参与骨骼肌的细胞膜修复过程；MG53 在心脏缺血预处理和缺血后适应的心肌保护作用中发挥关键作用；最近，我们实验室发现在胰岛素抵抗的动物模型中 MG53 表达量上调，而过表达 MG53 就足以引发肌肉的胰岛素抵抗并产生代谢紊乱；相反，敲除 MG53 使小鼠对高脂饮食诱导的肥胖和代谢综合征产生抵抗。这一新作用的机制是 MG53 具有泛素 E3 连接酶活性，MG53 通过泛素化修饰胰岛素受体 (IR) 及其底物 IRS1，从而显著促进两者的降解，阻断胰岛素信号通路，降低胰岛素敏感性，最终产生胰岛素抵抗。MG53 的泛素 E3 连接酶活性对于胰岛素抵抗的发生是充分且必要的。

从 UniProt 数据库中检索并提取人、小鼠、大鼠 MG53 蛋白质序列，利用 WebLab 中的序列比对程序 Needle，选择默认计分矩阵 BLOSUM62 和默认空位罚分进行序列比对，结果如表 4 所示。

表 4 人、小鼠、大鼠 MG53 蛋白质序列比对结果

| 物种 Species | 登录号 Accession | 得分 Score | 相同氨基酸 Identity | 相同和相似氨基酸 Similarity | 空位 Gaps |
|---------------|-------------------------------|-------------|--------------------|------------------------|-----------------|
| 人/小鼠 | NP_001008275/ NP_001073401 | 2293.0 | 435/477 (91.2%) | 456/477 (95.6%) | 0/477 (0.0%) |

| | | | | | |
|-------|-------------------------------|--------|--------------------|--------------------|-----------------|
| 人/大鼠 | NP_001008275/ NP_001071143 | 2274.0 | 433/477 (90.8%) | 451/477 (94.5%) | 0/477 (0.0%) |
| 小鼠/大鼠 | NP_001073401/ NP_001071143 | 2380.0 | 455/477 (95.4%) | 465/477 (97.5%) | 0/477 (0.0%) |

检索 RefSeq 数据库中人、小鼠、大鼠 MG53 编码基因的 mRNA 序列，下载其编码区序列，利用 WebLab 中的序列比对程序 Needle，选择默认计分矩阵 BLOSUM62，空位罚分增加到 20.0，进行序列比对，结果如表 5 所示。

表 5 人、小鼠、大鼠 MG53 基因编码区序列比对结果

| 物种 Species | 登录号 Accession | 得分 Score | 相同核苷酸 Identity | 编码区序列 差异位点 | 氨基酸序列 差异位点 |
|---------------|-------------------------------|-------------|----------------------|---------------------|------------------|
| 人/小鼠 | NM_001008274/ NM_001079932 | 7704.0 | 1231/1434 (85.8%) | 203/1434 (14.2%) | 42/477 (8.8%) |
| 人/大鼠 | NM_001008274/ NM_001077675 | 7574.0 | 1216/1434 (84.8%) | 218/1434 (15.2%) | 44/477 (9.2%) |
| 小鼠/大鼠 | NM_001079932/ NM_001077675 | 8312.0 | 1338/1434 (93.3%) | 96/1434 (6.7%) | 22/477 (4.6%) |

从以上序列比对结果可以看到，MG53 在人、小鼠和大鼠中较为保守，蛋白质序列两两比对的相似性均在 90%以上，预示该蛋白在机体中可能发挥着重要作用，且通过比对结果可以研究其中的重要功能位点。无论是蛋白质序列还是编码区序列的比对均表明，小鼠/大鼠 MG53 相似性高于人/小鼠，人/大鼠 MG53 相似性最低，这与人、大鼠和小鼠的演化关系相一致，且人和小鼠的 MG53 相似程度高于大鼠，在建立疾病模型时应优先考虑小鼠。对于任一对，编码区序列差异高于蛋白质序列差异，主要是密码子简并性所致。

参考文献:

Hardison RC. Evolution of hemoglobin and its genes. *Cold Spring Harb Perspect Med.* 2012 Dec 1;2(12):a011627.