



(12) 发明专利申请

(10) 申请公布号 CN 114496067 A

(43) 申请公布日 2022. 05. 13

(21) 申请号 202111655356.5

(22) 申请日 2021.12.30

(71) 申请人 深圳市硬核酸生物科技有限公司

地址 518000 广东省深圳市南山区西丽街
道松坪山社区朗山路13号南门西侧清
华信息港科研楼1210

(72) 发明人 谢禹鑫 陈盛培 陈婉婷 潘小瑜

(74) 专利代理机构 北京三高永信知识产权代理
有限责任公司 11138

专利代理师 徐立

(51) Int. Cl.

G16B 15/00 (2019.01)

G16C 10/00 (2019.01)

G06K 9/62 (2022.01)

G06N 10/60 (2022.01)

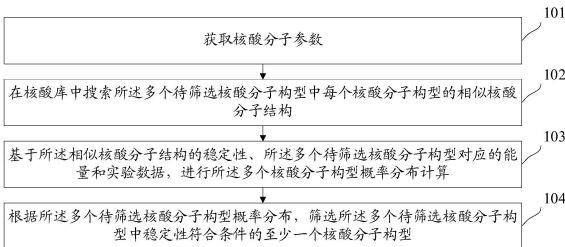
权利要求书3页 说明书10页 附图2页

(54) 发明名称

核酸分子结构筛选方法及装置、计算机设备

(57) 摘要

本公开提供了一种核酸分子结构筛选方法及装置、计算机设备。方法包括：获取核酸分子参数，所述核酸分子参数包括：核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据；在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构；基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据，进行所述多个待筛选核酸分子构型概率分布计算；根据所述多个待筛选核酸分子构型概率分布，筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。



1. 一种核酸分子结构筛选方法,其特征在于,所述方法包括:

获取核酸分子参数,所述核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据;

在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算;

根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

2. 根据权利要求1所述的方法,其特征在于,所述在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构,包括:

将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型,所述相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构;

获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

3. 根据权利要求1所述的方法,其特征在于,所述基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算,包括:

将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型,所述核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率;

获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

4. 根据权利要求1至3任一项所述的方法,其特征在于,所述根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型,包括:

按照所述多个待筛选核酸分子构型的稳定性概率对所述多个待筛选核酸分子构型进

行排序,并将排名转换成第一分数;

按照所述多个待筛选核酸分子构型的药物性质对所述多个待筛选核酸分子构型进行排序,并将排名转换成第二分数;

综合所述第一分数和所述第二分数,从所述多个待筛选核酸分子构型中筛选出分数高的至少一个核酸分子构型。

5. 根据权利要求1至3任一项所述的方法,其特征在于,所述方法还包括:

对所述核酸分子的实验数据进行数据清理和补全;

对清理和补全后的所述核酸分子的实验数据进行归一化,并对非数值实验数据进行数值化和归一化。

6. 根据权利要求1至3任一项所述的方法,其特征在于,所述待筛选核酸分子构型的实验数据包括生物分子实验和动物实验的数据,所述生物分子实验和动物实验的数据包括实验环境、半衰期、退火温度和药物富集度中的至少一项。

7. 一种核酸分子结构筛选装置,其特征在于,所述装置包括:

获取模块,用于获取核酸分子参数,所述核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据;

搜索模块,用于在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

计算模块,用于基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算;

筛选模块,用于根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

8. 根据权利要求7所述的装置,其特征在于,所述搜索模块,用于将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型,所述相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构;

获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

所述计算模块,用于将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型,所述核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率;

获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

9.一种计算机设备,其特征在于,所述计算机设备包括处理器和存储器,所述存储器存储有至少一条程序代码,所述程序代码由所述处理器加载并执行以实现如权利要求1至5任一项所述的核酸分子结构筛选方法。

10.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质中存储有至少一条程序代码,所述程序代码由处理器加载并执行以实现如权利要求1至5任一项所述的核酸分子结构筛选方法。

核酸分子结构筛选方法及装置、计算机设备

技术领域

[0001] 本公开涉及生物技术领域,特别涉及一种核酸分子结构筛选方法及装置、计算机设备。

背景技术

[0002] 核酸是细胞中的基本物质之一,预测核酸的分子结构有助于对于它们功能的研究。

[0003] 相关技术提供了一种基于量子力学和分子动力学的核酸分子结构预测方法,该方法基于量子力学和分子动力学进行模拟计算,确定出多个待筛选核酸分子构型,然后根据这多个待筛选核酸分子构型的能量分布,从其中筛选出稳定性符合条件的一个或多个核酸分子构型进行研究。

[0004] 但是,相关技术筛选核酸分子构型的方法主要是基于基本的核酸分子能量分布,能量分布的计算是通过物理模拟实验仿真自然世界的一种方法,物理模拟实验本身依靠数学方法来近似求解,导致筛选的结果不够精确。

发明内容

[0005] 本公开实施例提供了一种核酸分子结构筛选方法及装置、计算机设备,能够提高核酸分子结构筛选的准确性。所述技术方案如下:

[0006] 本公开至少一实施例提供了一种核酸分子结构筛选方法,所述方法包括:

[0007] 获取核酸分子参数,所述核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据;

[0008] 在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

[0009] 基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算;

[0010] 根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

[0011] 可选地,所述在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构,包括:

[0012] 将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型,所述相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核

酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构;

[0013] 获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

[0014] 可选地,所述基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算,包括:

[0015] 将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型,所述核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率;

[0016] 获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

[0017] 可选地,所述根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型,包括:

[0018] 按照所述多个待筛选核酸分子构型的稳定性概率对所述多个待筛选核酸分子构型进行排序,并将排名转换成第一分数;

[0019] 按照所述多个待筛选核酸分子构型的药物性质对所述多个待筛选核酸分子构型进行排序,并将排名转换成第二分数;

[0020] 综合所述第一分数和所述第二分数,从所述多个待筛选核酸分子构型中筛选出分数高的至少一个核酸分子构型。

[0021] 可选地,所述方法还包括:

[0022] 对所述核酸分子的实验数据进行数据清理和补全;

[0023] 对清理和补全后的所述核酸分子的实验数据进行归一化,并对非数值实验数据进行数值化和归一化。

[0024] 可选地,所述待筛选核酸分子构型的实验数据包括生物分子实验和动物实验的数据,所述生物分子实验和动物实验的数据包括实验环境、半衰期、退火温度和药物富集度中的至少一项。

[0025] 本公开至少一实施例提供了一种核酸分子结构筛选装置,所述装置包括:

[0026] 获取模块,用于获取核酸分子参数,所述核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据;

[0027] 搜索模块,用于在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

[0028] 计算模块,用于基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算;

[0029] 筛选模块,用于根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

[0030] 可选地,所述搜索模块,用于将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型,所述相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构;

[0031] 获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

[0032] 可选地,所述计算模块,用于将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型,所述核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率;

[0033] 获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

[0034] 本公开至少一实施例提供了一种计算机设备,所述计算机设备包括处理器和存储器,所述存储器存储有至少一条程序代码,所述程序代码由所述处理器加载并执行以实现如前所述的核酸分子结构筛选方法。

[0035] 本公开至少一实施例提供了一种计算机可读存储介质,所述计算机可读存储介质中存储有至少一条程序代码,所述程序代码由处理器加载并执行以实现如前任一实施例所述的核酸分子结构筛选方法。

[0036] 本公开实施例提供的技术方案带来的有益效果是:

[0037] 在本公开实施例中,通过在量子力学和分子动力学的计算模拟的基础上,使用多个维度的核酸分子参数在核酸库中搜索相似核酸分子结构,并根据相似核酸分子结构的稳定性以及待筛选核酸分子构型对应的能量和实验数据,进行多个核酸分子构型概率分布计算,进而筛选出稳定性符合条件的至少一个核酸分子构型。由于上述方案参考了相似核酸分子结构的稳定性,并且同时考虑了待筛选核酸分子构型对应的能量和实验数据,而非单纯的依靠模拟实验仿真求解,从而提高了核酸分子结构筛选的准确性。

附图说明

[0038] 为了更清楚地说明本公开实施例中的技术方案,下面将对实施例描述中所需要使

用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0039] 图1是本公开实施例提供的一种核酸分子结构筛选方法的流程图;

[0040] 图2是本公开实施例提供的一种核酸分子结构筛选方法的流程图;

[0041] 图3是本公开实施例提供的一种核酸分子结构筛选装置的结构示意图;

[0042] 图4是本公开实施例提供的一种计算机设备的结构框图。

具体实施方式

[0043] 为使本公开的目的、技术方案和优点更加清楚,下面将结合附图对本公开实施方式作进一步地详细描述。

[0044] 除非另作定义,此处使用的技术术语或者科学术语应当为本公开所属领域内具有一般技能的人士所理解的通常意义。本公开专利申请说明书以及权利要求书中使用的“第一”、“第二”、“第三”以及类似的词语并不表示任何顺序、数量或者重要性,而只是用来区分不同的组成部分。同样,“一个”或者“一”等类似词语也不表示数量限制,而是表示存在至少一个。“包括”或者“包含”等类似的词语意指出现在“包括”或者“包含”前面的元件或者物件涵盖出现在“包括”或者“包含”后面列举的元件或者物件及其等同,并不排除其他元件或者物件。

[0045] 图1是本公开实施例提供的一种核酸分子结构筛选方法的流程图。参见图1,该方法包括:

[0046] 101:获取核酸分子参数。

[0047] 其中,核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据。

[0048] 其中,核酸分子的序列是指核酸分子其碱基序列。序列的长度是指核酸分子碱基的数量或长度。化学修饰是指核酸分子的碱基、中间、五碳糖、两端、磷酸骨架等位置的修饰。基于量子力学和分子动力学的计算模拟得到的数据也可以称为计算物理学结果数据,计算物理学模拟计算核酸分子获得的核酸分子构型(主要包括构型坐标、分子尺寸)、能量大小等数据。实验数据包括实验环境、半衰期、退火温度和药物富集度中的至少一项。

[0049] 在上述参数中,基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量可以通过量子力学和分子动力学模拟试验得到。实验数据通过对待筛选核酸分子构型进行生物分子实验和动物实验得到。核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰为本申请所处理的核酸分子的基本参数。

[0050] 102:在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

[0051] 核算库是指存储已知核酸分子结构数据的数据库,包括核酸分子构型、稳定性、以及核酸分子的各种实验数据。

[0052] 103:基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算。

[0053] 在从核酸库中搜索出相似核酸分子结构后,还可以从核酸库中获取相似核酸分子结构的稳定性,基于相似核酸分子结构的稳定性来进行待筛选核酸分子构型稳定性概率的计算。

[0054] 其中,多个待筛选核酸分子构型概率分布计算是指计算多个待筛选核酸分子构型的稳定性概率。稳定性是指在确定环境下,核酸分子的结构构型将不会或极小概率发生改变。这里的极小概率可以是指概率小于某一阈值,例如小于2%或1%等。稳定性概率是指待筛选核酸分子构型不会发生改变的的概率,稳定性概率越高,则稳定性越好。

[0055] 104:根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

[0056] 这里的稳定性符合条件可以是指稳定性概率高于设定值,或者,稳定性概率在多个待筛选核酸分子构型中排名在设定范围内,例如前3名等。

[0057] 在本公开实施例中,通过在量子力学和分子动力学的计算模拟的基础上,使用多个维度的核酸分子参数在核酸库中搜索相似核酸分子结构,并根据相似核酸分子结构的稳定性以及待筛选核酸分子构型对应的能量和实验数据,进行多个核酸分子构型概率分布计算,进而筛选出稳定性符合条件的至少一个核酸分子构型。由于上述方案参考了相似核酸分子结构的稳定性,并且同时考虑了待筛选核酸分子构型对应的能量和实验数据,而非单纯的依靠模拟实验仿真求解,从而提高了核酸分子结构筛选的准确性。

[0058] 图2是本公开实施例提供的一种核酸分子结构筛选方法的流程图。参见图2,该方法包括:

[0059] 200:训练相似核酸分子搜索模型和核酸分子构型概率分布计算模型。

[0060] 其中,相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构。

[0061] 核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率。

[0062] 示例性地,训练相似核酸分子搜索模型的方案如下:

[0063] 基于所述核酸库获取第一训练集和第一测试集;

[0064] 采用所述第一训练集和所述第一测试集训练所述相似核酸分子搜索模型。

[0065] 其中,训练集和测试集中的数据类型和相似核酸分子搜索模型输入的数据的类型相同。

[0066] 相似核酸分子搜索模型可以实现第一相似度、差值、第二相似度和距离的计算,第一相似度、差值、第二相似度和距离可以按照一定权重计算两个核酸分子构型的相似度。其

中,第一相似度和第二相似度的计算公式可以利用成熟的相似度算法实现。初始时,计算第一相似度的公式中的参数、计算第二相似度的公式中的参数以及计算距离的公式中的参数、以及计算两个核酸分子构型的相似度时的权重都是初始值。在样本集中事先对相似的核酸分子构型进行标注,训练过程中根据标注判断相似核酸分子搜索模型输出的和输入的是否为相似构型,从而实现对上述参数及权重的反馈优化。通过不断反馈优化,完成相似核酸分子搜索模型的训练。当然,在上述过程中,也可以不做样本集的标注,通过人力来判断相似核酸分子搜索模型输出的和输入的是否为相似构型。

[0067] 同样地,核酸分子构型概率分布计算模型也采用类似的方式训练,比如:

[0068] 基于所述核酸库获取第二训练集和第二测试集;

[0069] 采用所述第二训练集和所述第二测试集训练所述核酸分子构型概率分布计算模型。

[0070] 其中,训练集和测试集中的数据类型和核酸分子构型概率分布计算模型输入的数据的类型相同。

[0071] 核酸分子构型概率分布计算模型采用稳定性概率算法,对核酸分子构型的能量、相似核酸分子结构的稳定性和核酸分子构型的实验数据进行计算,得到稳定性概率。该稳定性概率算法可以是将能量、相似核酸分子结构的稳定性、实验数据等参数的归一化数值进行加权求和。初始时,稳定性概率算法中的参数(也即加权的权重)为初始值,通过训练对稳定性概率算法中的参数进行调节,从而使得通过该模型计算出的稳定性概率更加准确。

[0072] 训练时,将所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,将该数组作为输入值输入到核酸分子构型概率分布计算模型中。根据核酸分子构型概率分布计算模型输出的稳定性概率和该核酸分子构型的实际稳定性,对稳定性概率算法的参数进行反馈调整,从而实现对上述模型参数进行优化。通过不断反馈迭代,完成核酸分子构型概率分布计算模型的训练。

[0073] 示例性地,相似核酸分子搜索模型和核酸分子构型概率分布计算模型采用神经网络、决策树、支持向量机,其中,神经网络可以是深度神经网络。

[0074] 201:获取核酸分子参数。

[0075] 其中,核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据。

[0076] 在本公开实施例中,基于量子力学和分子动力学的计算模拟是指通过计算物理学软件进行核酸分子构型预测,在预测过程中利用软件模拟核酸分子构型的物理试验,通过量子力学和分子动力学确定各个核酸分子构型的能量分布。

[0077] 其中,量子力学实现基于从头算法(最原始的输入分子初始构型坐标),用密度泛函或者hartree-fock理论来求解薛定谔方程并获得这个体系的波函数,随着对能量的不断收敛和梯度下降从而找到能量收敛后的核酸分子构型。

[0078] 分子动力学则是基于牛顿力学方程和已知的力场(force field),动态的求解下一时刻每一个原子的速度和位置,当能量稳定或者收敛的时候的坐标即可作为核酸分子构型的预测结果。

[0079] 示例性地,上述计算物理学软件包括但不限于VASP,Gaussian,GAMESS,NWchem,Q-

chem, Spartan, openMM, Amber, NAMD, LAMMPS, Gromacs, CHARMM等。

[0080] 202:将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型。

[0081] 这里,在输入上述参数时,将对这些参数进行特征化,转换成相似核酸分子搜索模型容易识别和处理的格式,例如对这些参数进行数值化和归一化,然后输入到相似核酸分子搜索模型中。

[0082] 203:获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

[0083] 通过相似核酸分子搜索模型确定出和每个核酸分子构型相似的核酸分子结构,从而为确定每个核酸分子构型的稳定性做准备。

[0084] 204:将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型。

[0085] 同样地,在输入上述参数时,可以对这些参数进行特征化,转换成核酸分子构型概率分布计算模型容易识别和处理的格式,例如对这些参数进行数值化和归一化,然后输入到核酸分子构型概率分布计算模型中。

[0086] 也即,该方法还包括:

[0087] 对所述核酸分子的实验数据进行数据清理和补全;

[0088] 对清理和补全后的所述核酸分子的实验数据进行归一化,并对非数值实验数据进行数值化和归一化。

[0089] 其中,待筛选核酸分子构型的实验数据是指针对具有待筛选核酸分子构型的核酸分子进行实验得到的数据,实验数据包括生物分子实验和动物实验的数据,所述生物分子实验和动物实验的数据包括实验环境、半衰期、退火温度和药物富集度中的至少一项。

[0090] 其中,生物分子实验可以是聚合酶链式反应(Polymerase Chain Reaction, PCR)试验,质谱试验,电镜结构解析试验和细胞试验,动物实验可以是老鼠、猴子试验等。实验环境可以是水,生理盐水,血浆等。

[0091] 205:获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

[0092] 206:根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

[0093] 可选地,在进行核酸分子构型筛选时,除了仅按照稳定度进行核酸分子构型排序外,还可以参考核酸分子构型的药物性质。

[0094] 示例性地,步骤206包括:

[0095] 按照所述多个待筛选核酸分子构型的稳定性概率对所述多个待筛选核酸分子构型进行排序,并将排名转换成第一分数;

[0096] 按照所述多个待筛选核酸分子构型的药物性质对所述多个待筛选核酸分子构型进行排序,并将排名转换成第二分数;

[0097] 综合所述第一分数和所述第二分数,从所述多个待筛选核酸分子构型中筛选出分数高的至少一个核酸分子构型。

[0098] 其中,在转换第一分数和第二分数时,排名越高分数越高,例如稳定性概率越大(或越好),则排名越高,对应的第一分数也越高。在综合第一分数和第二分数时,可以采用加权形式计算。

[0099] 其中,药物性质是通过前述生物分子实验和动物实验得到的,可以包括前述药物富集度、半衰期,还可以包括核酸分子特异性、药物有效性和投递效率等。当药物性质包括多个参数时,可以分别排序,得到多个第二分数,然后再与第一分数综合。

[0100] 值得说明的是,上述步骤206也可以采用神经网络模型实现,这里不再赘述。

[0101] 在本公开实施例中,通过在量子力学和分子动力学的计算模拟的基础上,使用多个维度的核酸分子参数在核酸库中搜索相似核酸分子结构,并根据相似核酸分子结构的稳定性以及待筛选核酸分子构型对应的能量和实验数据,进行多个核酸分子构型概率分布计算,进而筛选出稳定性符合条件的至少一个核酸分子构型。由于上述方案参考了相似核酸分子结构的稳定性,并且同时考虑了待筛选核酸分子构型对应的能量和实验数据,从而提高了核酸分子结构筛选的准确性。

[0102] 图3是本公开实施例提供的一种核酸分子结构筛选装置的结构示意图。参见图3,该装置包括:获取模块301、搜索模块302、计算模块303和筛选模块304。

[0103] 其中,获取模块301,用于获取核酸分子参数,所述核酸分子参数包括:核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰、以及基于量子力学和分子动力学的计算模拟得到的多个待筛选核酸分子构型及对应的能量和实验数据;

[0104] 搜索模块302,用于在核酸库中搜索所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构;

[0105] 计算模块303,用于基于所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据,进行所述多个待筛选核酸分子构型概率分布计算;

[0106] 筛选模块304,用于根据所述多个待筛选核酸分子构型概率分布,筛选所述多个待筛选核酸分子构型中稳定性符合条件的至少一个核酸分子构型。

[0107] 示例性地,所述搜索模块302,用于将所述核酸分子序列、核酸分子序列的长度、核酸分子的化学修饰和多个待筛选核酸分子构型输入相似核酸分子搜索模型,所述相似核酸分子搜索模型用于分别确定所述核酸分子序列和所述核酸库中的各个核酸分子序列的第一相似度、确定所述核酸分子序列的长度和所述核酸库中的各个核酸分子序列的长度的差值、确定所述核酸分子的化学修饰和所述核酸库中的各个核酸分子序列的化学修饰的第二相似度,从所述核酸库中选出所述第一相似度、所述差值和所述第二相似度均满足阈值要求的核酸分子,依次确定选出的核酸分子的构型和所述多个待筛选核酸分子构型的距离,基于所述距离得到所述每个核酸分子构型的相似核酸分子结构;

[0108] 获取所述相似核酸分子搜索模型从核酸库中搜索到的所述多个待筛选核酸分子构型中每个核酸分子构型的相似核酸分子结构。

[0109] 示例性地,所述计算模块303,用于将所述相似核酸分子结构的稳定性、所述多个待筛选核酸分子构型对应的能量和实验数据输入核酸分子构型概率分布计算模型,所述核酸分子构型概率分布计算模型用于通过对同一所述核酸分子构型的能量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值进行计算,通过所述计算得到所述每个核酸分子构型的稳定性概率;其中,所述计算过程如下:将所述核酸分子构型的能

量、相似核酸分子结构的稳定性和所述核酸分子构型的实验数据的归一化数值组成一个数组,利于训练得到的稳定性概率算法对所述数组进行求解,得到所述每个核酸分子构型的稳定性概率;

[0110] 获取所述核酸分子构型概率分布计算模型计算出的所述多个待筛选核酸分子构型概率分布,所述多个待筛选核酸分子构型概率分布包括所述每个核酸分子构型的稳定性概率。

[0111] 可选地,筛选模块304,用于按照所述多个待筛选核酸分子构型的稳定性概率对所述多个待筛选核酸分子构型进行排序,并将排名转换成第一分数;按照所述多个待筛选核酸分子构型的药物性质对所述多个待筛选核酸分子构型进行排序,并将排名转换成第二分数;综合所述第一分数和所述第二分数,从所述多个待筛选核酸分子构型中筛选出分数高的至少一个核酸分子构型。

[0112] 可选地,该装置还包括:数据处理模块305,用于对所述核酸分子的实验数据进行数据清理和补全;

[0113] 对清理和补全后的所述核酸分子的实验数据进行归一化,并对非数值实验数据进行数值化和归一化。

[0114] 可选地,所述待筛选核酸分子构型的实验数据包括生物分子实验和动物实验的数据,所述生物分子实验和动物实验的数据包括实验环境、半衰期、退火温度和药物富集度中的至少一项。

[0115] 需要说明的是:上述实施例提供的核酸分子结构筛选装置在筛选核酸分子结构时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即将设备的内部结构划分成不同的功能模块,以完成以上描述的全部或者部分功能。另外,上述实施例提供的核酸分子结构筛选装置与核酸分子结构筛选方法实施例属于同一构思,其具体实现过程详见方法实施例,这里不再赘述。

[0116] 图4是本公开实施例提供的一种计算机设备的结构框图。通常,计算机设备包括有:处理器401和存储器402。

[0117] 处理器401可以包括一个或多个处理核心,比如4核心处理器、8核心处理器等。处理器401可以采用DSP (Digital Signal Processing,数字信号处理)、FPGA (Field-Programmable Gate Array,现场可编程门阵列)、PLA (Programmable Logic Array,可编程逻辑阵列)中的至少一种硬件形式来实现。处理器401也可以包括主处理器和协处理器,主处理器是用于对在唤醒状态下的数据进行处理的处理器,也称CPU (Central Processing Unit,中央处理器);协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。

[0118] 存储器402可以包括一个或多个计算机可读存储介质,该计算机可读存储介质可以是非暂态的。存储器402还可包括高速随机存取存储器,以及非易失性存储器,比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中,存储器402中的非暂态的计算机可读存储介质用于存储至少一个指令,该至少一个指令用于被处理器401所执行以实现本申请中方法实施例提供的核酸分子结构筛选方法。

[0119] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0120] 以上所述仅为本公开的可选实施例,并不用以限制本公开,凡在本公开的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本公开的保护范围之内。

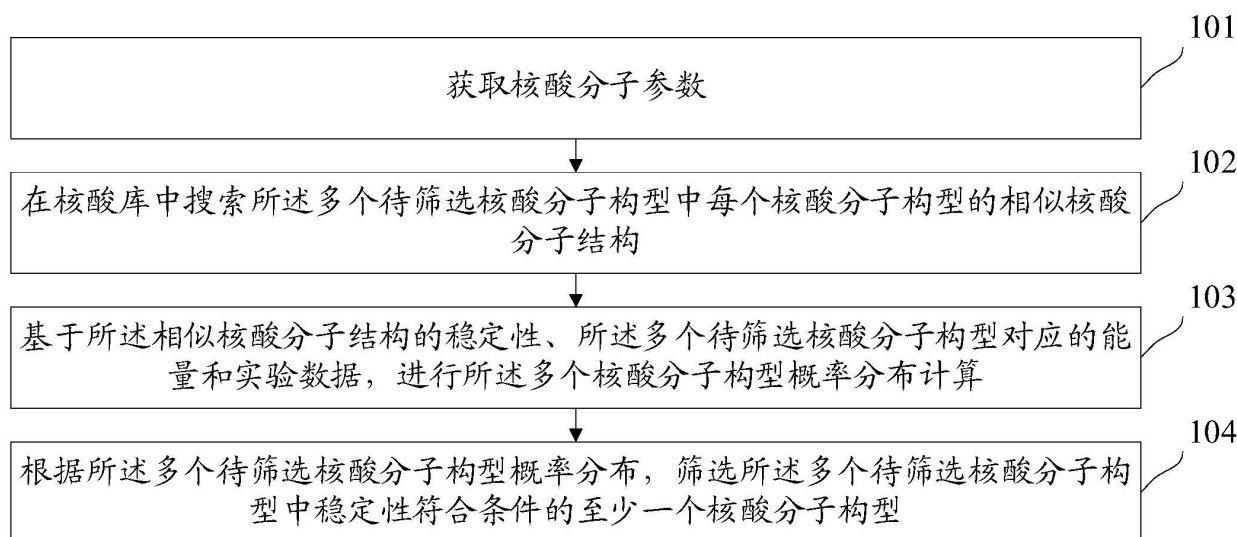


图1

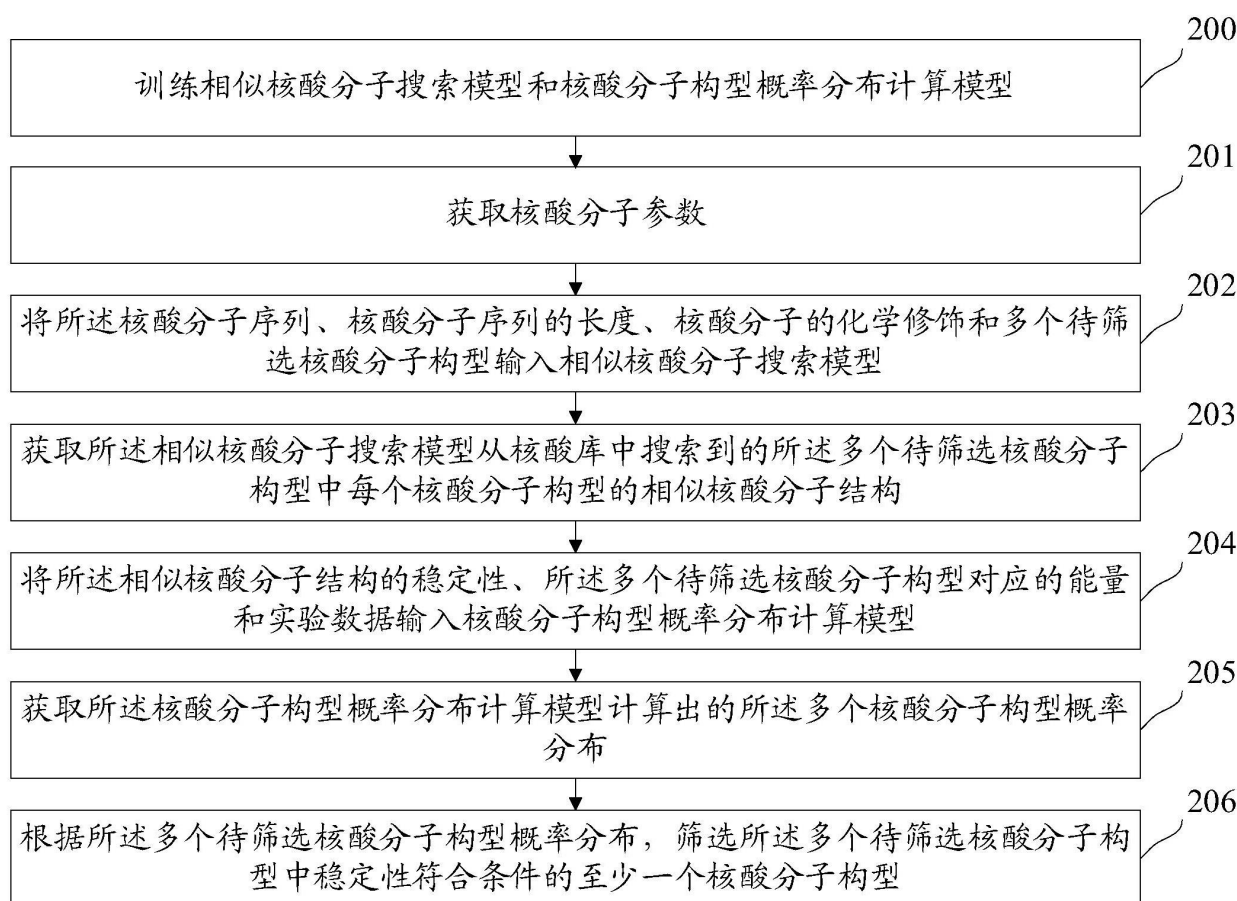


图2

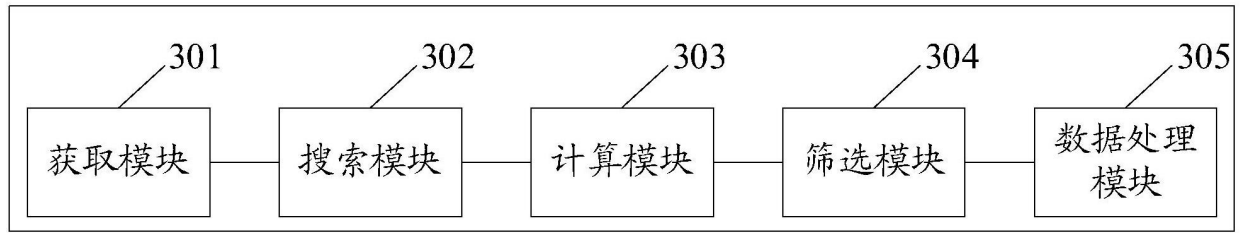


图3

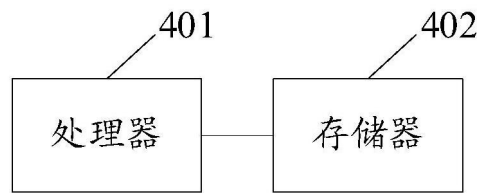


图4