



(10) 申请公布号 CN 117295490 A

(43) 申请公布日 2023. 12. 26

(21) 申请号 202280033865.0

(22) 申请日 2022.05.09

(30) 优先权数据

21173067.6 2021.05.10 EP

(85) PCT国际申请进入国家阶段日

2023.11.08

(86) PCT国际申请的申请数据

PCT/EP2022/062389 2022.05.09

(87) PCT国际申请的公布数据

WO2022/238278 EN 2022.11.17

(71) 申请人 默克专利股份公司

地址 德国达姆施塔特

(72) 发明人 T·罗森克兰兹 M·冯德哈尔

A·索西克 J·G·勃兰登堡

N·巴尼克

(74) 专利代理机构 中国专利代理(香港)有限公司
72001

专利代理师 钟茂建 陈晓

(51) Int.Cl.

A61K 9/00 (2006.01)

权利要求书2页 说明书13页 附图3页

(54) 发明名称

用于蛋白质制剂的粘度降低赋形剂的数字选择

(57) 摘要

用于经由计算机(6)针对包含至少一种未知蛋白质(11)的制剂(8)选择至少一种粘度改变赋形剂(2)的方法,包括以下步骤:提供来自数据库的数据集(1),所述数据集描述了包含至少一种蛋白质和可选的至少一种粘度改变赋形剂(2)的几种已知制剂的粘度;由计算机(6)经由计算机中模拟从赋形剂的列表中生成至少一种赋形剂(2)的表示;使用在计算机(6)上执行的机器学习模型(5),所述机器学习模型使用至少一种赋形剂(2)的所生成的表示来识别数据集(1)中的模式,以通过将所识别的模式应用于所提供的至少一种未知蛋白质(11)的数据上来评估选自赋形剂的列表的至少一种粘度改变赋形剂(2)对包含至少一种未知蛋白质(11)和至少一种粘度改变赋形剂(2)的新制剂(8)的粘度改变效果;取决于评估结果,根据采集标准从列表中选择至少一种赋形剂,并将其应用于未知蛋白质(11),其中所

提供的至少一种未知蛋白质(11)的数据是描述蛋白质组合物的粘度的数据,所述蛋白质组合物包含至少一种未知蛋白质(11)和可选地具有至少一种粘度改变赋形剂(2)。

1. 用于经由计算机 (6) 针对包含至少一种未知蛋白质 (11) 的制剂 (8) 选择至少一种粘度改变赋形剂 (2) 的方法, 包括以下步骤:

- 提供来自数据库的数据集 (1), 所述数据集描述了包含至少一种蛋白质和可选的至少一种粘度改变赋形剂 (2) 的几种已知制剂的粘度;
- 由计算机 (6) 经由计算机中模拟从赋形剂的列表中生成至少一种赋形剂 (2) 的表示;
- 使用在计算机 (6) 上执行的机器学习模型 (5), 所述机器学习模型使用至少一种赋形剂 (2) 的所生成的表示来识别数据集 (1) 中的模式, 以通过将所识别的模式应用于所提供的至少一种未知蛋白质 (11) 的数据上来评估选自赋形剂的列表的至少一种粘度改变赋形剂 (2) 对包含至少一种未知蛋白质 (11) 和至少一种粘度改变赋形剂 (2) 的新制剂 (8) 的粘度改变效果;
- 取决于评估结果, 根据采集标准从列表中选择至少一种赋形剂, 并将其应用于未知蛋白质 (11),

其中所提供的至少一种未知蛋白质 (11) 的数据是描述蛋白质组合物的粘度的数据, 所述蛋白质组合物包含至少一种未知蛋白质 (11) 和可选地具有至少一种粘度改变赋形剂 (2)。

2. 根据权利要求1所述的方法, 其中通过预测包含至少一种未知蛋白质 (11) 和至少一种粘度改变赋形剂 (2) 的新制剂 (8) 的粘度 (3) 来完成评估至少一种粘度改变赋形剂 (2) 的粘度改变效果。

3. 根据权利要求1或2所述的方法, 其中, 所述数据集 (1) 已经通过实验测量 (10) 生成, 并经由计算机 (6) 存储在数据库中。

4. 根据权利要求1至3中的任一项所述的方法, 其中作为来自改变所述新制剂 (8) 的粘度的所述列表的至少一种赋形剂, 使用来自所述列表的两种或更多种赋形剂的最充分组合。

5. 根据权利要求1至4中的任一项所述的方法, 其中, 向制剂专家 (9) 提议至少一个具体的实验测量 (10), 所述制剂专家 (9) 在实验室中进行至少一个相应的实验 (10) 以验证预测粘度 (3), 并且通过经由计算机 (6) 将验证结果添加到数据库中的提供的数据集 (1) 中, 来用所述验证结果训练机器学习模型 (5)。

6. 根据权利要求1至5中的任一项所述的方法, 其中所述机器学习模型 (5) 通过将描述至少一种原型蛋白质制剂 (8) 的粘度的数据集 (1) 与所述至少一种粘度改变赋形剂 (2) 或其组合的表示相结合来创建和训练。

7. 根据权利要求6所述的方法, 其中给定的制剂 (8) 的粘度值经由机器学习模型 (5) 以高斯过程的形式来建模, 并且模型预测 (3) 用于通过贝叶斯最优实验设计来指导制剂专家 (9)。

8. 根据权利要求6或权利要求7所述的方法, 其中, 在计算机 (6) 上训练机器学习模型 (5) 是通过执行至少一次以下步骤来完成的:

- 通过使训练数据的边际可能性最大化, 用来自数据集 (1) 的训练数据优化机器学习模型参数;
- 基于机器学习模型 (5) 来评估未测试赋形剂 (2) 或其组合的粘度值的后验分布, 并且由此预测粘度 (3);

- 通过优化从计算的后验分布获得的采集分数,选择一组新的赋形剂(2)或其组合;
- 向制剂专家(9)提议一组新的赋形剂(2)或其组合,然后制剂专家在实验室中进行相应的实验(10)以确定得到的粘度;
- 将获得的测量值(10)添加到训练数据。

9. 根据权利要求8所述的方法,其中从粘度值(3)的后验分布获得的粘度(3)的预测基于表征在所考虑的制剂(8)中使用的赋形剂(2)的pH依赖性特征向量和所使用的赋形剂浓度水平。

10. 根据权利要求1至9中的任一项所述的方法,其中所述采集标准是哪种粘度改变赋形剂(2)使粘度降低最多。

11. 根据权利要求1至10中的任一项所述的方法,其中所述赋形剂(2)的表示由所述计算机(6)以物理参数以及分子指纹的形式生成。

12. 根据权利要求1至11中的任一项所述的方法,其中,所生成的赋形剂(2)的表示被实验地交叉验证。

13. 根据权利要求1至12中的任一项所述的方法,其中所生成的赋形剂(2)的表示包括量子力学特征,可选地补充有一组拓扑分子指纹。

14. 在根据权利要求5至13创建和训练的计算机上执行的机器学习模型。

15. 根据权利要求14所述的机器学习模型,其中,所述高斯过程用满足相同目的的任何其他模型架构替换,特别是用其他类型的随机过程、深度贝叶斯网络、广义线性模型、神经网络、支持向量机、基于树的模型、集合模型等替换。

用于蛋白质制剂的粘度降低赋形剂的数字选择

[0001] 本发明涉及用于经由计算机来选择用于蛋白质组合物的粘度降低赋形剂 (viscosity reducing excipients) 的方法。

[0002] 现有技术的背景和描述

[0003] 单克隆抗体 (mAB) 和其他蛋白质治疗剂通常通过肠胃外给药。皮下注射由于其简化患者给药 (快速、低体积注射) 和降低治疗成本 (较短的医疗辅助) 的潜力而特别受欢迎用于蛋白质治疗剂的递送。为了确保患者依从性 (compliance), 期望的是皮下注射剂型是等渗的, 并且可以以小体积注射 (每个注射部位 < 2.0ml)。为了减少注射体积, 蛋白质通常以 1mg/ml 至 150mg/ml 的浓度给药。

[0004] 同时, 基于 mAB 的治疗通常需要几 mg/kg 的剂量。因此, 高治疗剂量和低注射体积的组合导致需要高浓缩的治疗性抗体制剂。然而, 作为大蛋白质, 抗体除了具有复杂的三维结构之外, 还具有大量的功能基团。这使得它们的配制困难, 特别是当需要高浓度时。

[0005] 关于高浓度蛋白质溶液的主要问题之一是粘度。在高浓度下, 蛋白质倾向于形成高粘性的溶液, 这主要是由于非天然的自缔合 (non-native self-association)。此外, 在如此高的浓度下, 蛋白质示出增加的聚集和颗粒形成的速率。

[0006] 这些问题涉及制造过程和对患者的给药两者。在制造过程中, 高粘性的高浓缩蛋白制剂给超滤和无菌过滤带来了特别的困难。此外, 切向流过滤通常用于缓冲液交换和增加蛋白质浓度。然而, 因为粘性溶液在注射和过滤期间示出增加的背压和剪切应力, 所以治疗性蛋白质可能不稳定和/或处理时间延长。所述增加的剪切应力频繁地导致产品的损失。这两个方面都不利地影响过程经济性。

[0007] 同时, 在给药时高粘度是不可接受的, 因为它显著限制了蛋白质的可注射性。

[0008] 已经标识了特定的赋形剂和赋形剂组合来降低蛋白质制剂的粘度。然而, 使用筛选方法来标识最佳赋形剂或赋形剂组合是耗时的。特别是考虑到竞争激烈的时间线和有限数量的测试材料 (蛋白质), 用于减少实验的次数和加快赋形剂选择的数据驱动的方法 (data-driven approach) 是非常有益的。

[0009] 在现有技术中, 已知用于蛋白质制剂设计的几种方法, 包括机器学习模型。例如, 文章 “Machine learning models of antibody-exci-pient preferential interactions for use in computational formulation design” (Mol. Pharmaceutics 2020, 17, 3589-3599, DOI.10.1021/acs.molpharmaceut.0c00629) 公开了使用计算机模拟来描述制剂赋形剂与溶液中的蛋白质的相互作用的能力。这使得制剂设计能够在新抗体治疗剂的开发中的早期开始。为此, 它公开了特征集, 用于数字地描述抗体的表面的局部区域, 以用于在机器学习应用中使用。

[0010] 另一种方法在 2020 年 12 月 2 日发表在 “the Journal of Pharmaceutical Sciences” 中的评论 “Prediction Machines: Applied Machine Learning for Therapeutic Protein Design and Development” 中进行了总结。它描述了机器学习模型的应用, 以更好地理解蛋白质溶液的非线性浓度依赖性粘度, 预测蛋白质氧化和脱酰胺速率, 对亚可见颗粒进行分类, 并比较蛋白质的物理稳定性。它进一步提供了使用各种机器学

习方法对先前发布的数据进行回归和分类的改进建模结果。

[0011] 从国际专利申请W0 2021/0413S4 A1已知另一种方法,其公开了一种用于预测潜在蛋白质制剂的性质的方法,其中一组制剂描述符被分类为属于多个预定组中的特定一组,每个预定组对应于所述蛋白质制剂性质的不同值范围;对该组描述符进行分类包括将该组描述符的至少第一部分作为输入应用于第一机器学习模型。该方法还包括基于分类从对应于不同组的多个模型之中选择第二机器学习模型。该方法还包括通过将该组制剂描述符的至少第二部分作为输入应用于所选择的模型来预测对应于该组描述符的蛋白质制剂性质的值。该方法还包括使蛋白质制剂性质的值显示给用户和/或存储在存储器中。

[0012] 然而,尽管那些出版物示出了如何模拟蛋白质溶液的粘度的方式,但它们仍然有一些重要的缺点。首先,它们都需要关于目标蛋白质的先验知识,无论它是以数据、描述符、性质还是以结构细节的形式。这是显著的缺点,因为蛋白质开发者不愿意分享关于他们候选蛋白质的详细信息,或者那些信息根本不可用。用于获得这些参数的实验可能是麻烦的、耗时的,并且因此是成本密集的(cost intensive)。此外,收集蛋白质描述符是复杂的,并且可能没有很好地反映蛋白质特性。关于制剂设计的任务,现有方法的最严重的缺点是它们限于预测固定蛋白质制剂的浓度依赖性粘度,并且没有提供辅助用户探索其他制剂的实验设计。因此,目前的现有技术没有提供直接的方式来使用从其预测中获得的见解来优化涉及其他赋形剂的制剂,更不用说其他蛋白质了。另一个问题是这些文档只考虑了单一赋形剂的使用。组合赋形剂是有益的,因为与具有类似粘度降低效果的单一赋形剂相比,不同的赋形剂可能示出协同粘度降低和/或改善的蛋白质稳定性。赋形剂的组合未被上面提到的参考文献的模型涵盖。

[0013] 本专利申请的任务是解决上面提到的问题。进一步的任务是找到一种更高效的基于机器学习的方法,用于确定和标识改变蛋白质制剂的粘度的最佳赋形剂或赋形剂组合。进一步的任务是找到一种机器学习方法来标识最佳实验设计,以找到改变蛋白质制剂的粘度的最优赋形剂或赋形剂组合。

发明内容

[0014] 该任务已经通过一种用于经由计算机针对包含至少一种未知蛋白质的制剂选择至少一种粘度改变赋形剂的方法,包括以下步骤:提供来自数据库的数据集,所述数据集描述了包含至少一种蛋白质和可选的至少一种粘度改变赋形剂的几种已知制剂的粘度;由计算机经由计算机中模拟(In-Silico-simulation)从赋形剂的列表中生成至少一种赋形剂的表示;使用在计算机上执行的机器学习模型,所述机器学习模型使用至少一种赋形剂的所生成的表示来识别数据集中的模式,以通过将所识别的模式应用于所提供的至少一种未知蛋白质的数据上来评估选自赋形剂的列表的至少一种粘度改变赋形剂对包含至少一种未知蛋白质和至少一种粘度改变赋形剂的新制剂的粘度改变效果;取决于评估结果,根据采集标准从列表中选择至少一种赋形剂,并将其应用于未知蛋白质,其中所提供的至少一种未知蛋白质的数据是描述蛋白质组合物的粘度的数据,所述蛋白质组合物包含至少一种未知蛋白质和可选地具有至少一种粘度改变赋形剂。

[0015] 与最先进的方法相比,该过程提供了一种更高效的方式来探索蛋白质-赋形剂制剂的粘度。主要优点是,通过使用训练的机器学习模型,可以显著减少为了分别确定确认由

所使用的赋形剂达到的得到的粘度而需要进行的真实的实验室测试的次数。为此,所使用的模型评估来自用于创建该表示的数据集的赋形剂的列表中至少一种粘度改变赋形剂的粘度改变效果。评估的一个选项是模型预测所有可能的赋形剂(用于数据集的赋形剂)的得到的粘度,使得可以选择未知蛋白质和(一种或多种)赋形剂的最合适组合,但是模型预测从列表中选择赋形剂的得到的粘度或使用其他评估的方式也是可能的。模型的预测准确性越高,需要执行的实际、耗时和资源消耗的测试就越少。该模型的预测准确性通常随着所提供的粘度测量的次数的增加而增加。如果没有或只有很少的测量值可用,则机器学习模型主要使用赋形剂表示和其他已知制剂的数据,优选与新制剂相似的那些。在机器学习模型已经预测了新制剂的粘度并且已经选择最合适的(一种或多种)赋形剂之后,可以测量制剂的真实粘度(ground truth viscosity),并且可以将相应的数据反馈回到该模型,该模型用于提高后续预测的准确性。为了运行机器学习模型,可以使用具有处理器和相应的工作和存储存储器的每个标准个人或工业计算机。使用同一台计算机来进行计算机中模拟和执行机器学习模型是可能的,但在大多数情况下,使用专门配置成执行相应应用的两台不同的计算机更高效。充分性的标准可以例如建议预期最大程度降低粘度或使得产生最大信息增益的赋形剂。该任务的解决方案还包括软件产品,该软件产品存储在计算机可读存储介质上并且包括指令,当由计算机执行时,该指令使计算机执行如先前章节中公开的方法步骤。

[0016] 如本文中所定义的,“未知蛋白质”意指要通过所描述的方法测试的蛋白质。对于那些蛋白质,性质和/或特性,比如特定的蛋白质描述符,在执行所公开的方法时不一定是已知的。特别地,未知蛋白质是不在如上所述的方法的数据库中的蛋白质。更特别地,除了描述包含至少一种未知蛋白质和可选地具有至少一种粘度改变赋形剂的蛋白质组合物的粘度的数据之外,没有具有或不具有粘度改变赋形剂的综合粘度测量值可用,所述数据是作为如上所述的方法的提供数据所需要的。相对于已知的现有技术,所公开的方法的优点之一确实是不需要知道关于所用蛋白质的具体信息,并且因此也可以使用未知蛋白质。相反的是“包含至少一种蛋白质和可选的至少一种粘度改变赋形剂的已知制剂”,其是指蛋白质组合物,其中蛋白质本身及其一些或全部性质和/或特性是已知的。特别地,那些蛋白质在数据库中,更特别地,具有或不具有粘度改变赋形剂的粘度测量值是可用的。可选地,该制剂可以包含一种或多种已知的粘度改变赋形剂。

[0017] 如本文中所定义的,“包含至少一种未知蛋白质和至少一种粘度改变赋形剂的新制剂”是指包含如上定义的未知蛋白质和至少一种粘度改变赋形剂的蛋白质组合物。该制剂包含一种或多种已知的粘度改变赋形剂。根据本发明,预测如上定义的新制剂的粘度。优选地,预测多于一种制剂的粘度,例如包含至少一种未知蛋白质和至少一种粘度改变赋形剂A的制剂和包含至少一种未知蛋白质和至少一种粘度改变赋形剂B的制剂。更优选地,预测所有可能的制剂的粘度。在该上下文中,“所有可能的组合”意指至少一种未知蛋白质与选自用于生成数据集的赋形剂的列表的至少一种粘度改变赋形剂的所有组合。在进一步的实施例中,一组至少一种粘度改变赋形剂选自赋形剂的列表。

[0018] 如本文中所定义的,“提供的至少一种未知蛋白质的数据”是描述包含至少一种未知蛋白质的蛋白质组合物的粘度的数据,所述蛋白质组合物不具有粘度改变赋形剂或具有至少一种粘度改变赋形剂。在该上下文中,“描述粘度的数据”意指由蛋白质组合物的至少

一次粘度测量产生的数据。蛋白质组合物包含至少一种未知蛋白质,其中蛋白质本身及其性质和特性是未知的。可选地,蛋白质组合物可以包含一种或多种已知的粘度改变赋形剂。一种或多种已知的粘度改变赋形剂是也用于生成数据集的粘度改变赋形剂。

[0019] 所提供的至少一种未知蛋白质的数据不是指该蛋白质的描述符或性质。如在其他方法中使用的蛋白质描述符的收集是复杂的,并且可能没有很好地反映蛋白质特性。

[0020] 此外,对蛋白质开发者而言不需要分享关于他们的候选蛋白质的敏感信息。它还避免了成本密集的MD模拟或同源建模。此外,如其他方法所需的蛋白质的结构细节可能是不可获得的。在本发明的优选实施例中,需要一组有限的粘度测量、需要仅一次粘度测量或不需要粘度测量。

[0021] 本发明的有利且因此优选的进一步发展从相关联的从属权利要求以及从说明书和相关联的附图中显现出来。

[0022] 所公开的方法的那些优选的进一步发展之一包括数据集已经通过实验测量生成并且经由计算机存储在数据库中。真实的实验室测试也是用于生成之后由机器学习模型使用的数据集的优选方式。该数据集越准确且越有代表性,机器学习模型的结果就越好。这一点既考虑了包含已知制剂的粘度和赋形剂的数据集,并且又考虑了新制剂的数据集。

[0023] 所公开的方法的那些优选的进一步发展的另一项包括,作为来自改变新制剂(8)的粘度的列表的至少一种赋形剂,使用来自列表的两种或更多种赋形剂的最充分组合。

[0024] 两种或更多种赋形剂的组合可能是有益的,因为与具有类似粘度降低效果的单一赋形剂相比,不同的赋形剂可能示出协同粘度降低和/或改善的蛋白质稳定性。

[0025] 所公开的方法的那些优选的进一步发展的另一项包括,向制剂专家提议具体的实验测量,所述制剂专家在实验室中进行这些相应的实验以验证预测粘度,并且通过经由计算机将验证结果添加到数据库中的提供的数据集中,来用所述验证结果训练机器学习模型。此外,还可以向制剂专家提议来自机器学习模型的预测粘度值。新制剂中的所得到的粘度的所提及的测量优选地由制剂专家执行。专家由机器人机器和软件支持以执行测量是可能的。如果合适的硬件和软件是可用的,则测量也可以完全自动执行。

[0026] 所公开的方法的那些优选的进一步发展的另一项包括,描述不具有赋形剂和/或已经验证的赋形剂的新制剂的粘度的初始数据被用作新制剂数据的提供数据。更具体地,如果已经有一些关于新制剂的已知数据,例如来自先前的测量或任何其他源,则该数据被提供给机器学习模型,这进一步减少了实现准确预测所需的测试或测量的数量。

[0027] 所公开的方法的那些优选的进一步发展的另一项包括,所述机器学习模型通过将描述至少一种原型蛋白质制剂的粘度的数据集与至少一种赋形剂或其组合的表示相结合来创建和训练。首先用于创建机器学习模型的数据集是包括已知制剂及其赋形剂的数据集。如果机器学习模型然后被用于预测新的、可能未知的制剂的粘度,则通过已知的具有或不具有赋形剂的该制剂的特性(如果可用的话)和/或通过向其馈送由确认实验室测试产生的实验测量数据来进一步训练该机器学习模型。如果已知特性在所需的数字表示形式中不可用,则需要分别对其进行转换。

[0028] 所公开的方法的那些优选的进一步发展的另一项包括,给定的蛋白质制剂的粘度值以高斯过程的形式来建模,并且模型预测用于通过贝叶斯最优实验设计来指导制剂专家。借助该指导,制剂专家然后可以对由机器学习模型建议的赋形剂或其组合执行必要的

测量。

[0029] 所公开的方法的那些优选的进一步发展的另一项包括,在计算机上训练机器学习模型是通过执行至少一次以下步骤来完成的:通过使训练数据的边际可能性最大化,用来自数据集的训练数据优化机器学习模型参数;基于机器学习模型来评估未测试赋形剂或其组合的粘度值的后验分布,并且由此预测粘度;通过优化从计算的后验分布获得的采集分数,选择一组新的赋形剂或其组合;向制剂专家提议一组新的赋形剂或其组合,然后制剂专家在实验室中进行相应的实验以确定得到的粘度;以及将获得的测量值添加到训练数据。

[0030] 所公开的方法的那些优选的进一步发展的另一项包括,从粘度值的后验分布获得的粘度的预测基于表征在所考虑的制剂中使用的赋形剂的pH依赖性特征向量和所使用的赋形剂浓度水平。这表示机器学习模型如何预测粘度的最优选方式。然而,机器学习模型不限于该方法。如果有用于预测粘度值的替代方式,它们可以被实现到机器学习模型中并由机器学习模型执行。

[0031] 所公开的方法的那些优选的进一步发展的另一项包括,所述采集标准评估哪种粘度改变赋形剂预期使粘度降低最多。替代实施例可以包括建议实验的其他采集标准,例如该其他采集标准预期产生最大的信息增益,导致最大的模型变化,提供将制剂粘度提高到超过最佳观察设置的水平的最大可能性,产生超过当前最优制剂的最大预期改进,或者其在制剂搜索空间的探索和迄今收集的知识的利用之间提供任何其他系统的折衷。

[0032] 所公开的方法的那些优选的进一步发展的另一项包括,在水溶液中包含至少蛋白质、至少一种粘度改变剂、至少一种缓冲剂、至少一种稳定剂和至少一种表面活性剂的蛋白质制剂中测量粘度。这种组分的组合是最常见的组合,并且因此是优选使用的。然而,如果存在所要求保护的方法需要的其他组合和/或更适合于所要求保护的方法的其他组合,则也可以使用它们。

[0033] 所公开的方法的那些优选的进一步发展的另一项包括,赋形剂的表示由计算机以物理参数以及分子指纹的形式生成。物理参数描述了赋形剂及其性质,使得机器学习模型可以处理所述参数,并使用它们来预测它们将在特定蛋白质制剂中引起的粘度。可能的参数包括但不限于电荷分布、偶极矩、四极矩轨迹和各向异性、极化率、分子伦敦色散系数(C6)、logP水/己烷分布系数、溶剂可及表面积、分子轨道能量HOMO-LUMO间隙。

[0034] 本专利申请的任务的另一个解决方案是在计算机上执行的机器学习模型,该机器学习模型如先前章节中所述那样被创建和训练。

[0035] 所公开的机器学习模型的那些优选的进一步发展的另一项包括,高斯过程用满足相同目的的任何其他模型架构替换,特别是用其他类型的随机过程、广义线性模型、神经网络、支持向量机、基于树的模型、集合模型(ensemble model)等替换。

[0036] 发明的具体实施方式

[0037] 以下使用至少一个优选示例性实施例参考相关联的附图更详细地描述根据本发明的方法、机器学习模型和软件产品以及它们的功能上有利的开发。在附图中,彼此对应的要素具有相同的参考标号。

[0038] 附图示出了:

[0039] 图1:关于本发明方法的过程概述。

[0040] 图2:所涉及的系统组件的总结。

[0041] 图3:使用的机器学习模型的训练。

[0042] 图4:示出本发明方法的性能的结果图表。

[0043] 针对该问题的解决方案是软件工具,其使得用户能够在数据驱动的决策制定中解决制剂挑战。该工具由三个组件组成:

[0044] 1. 实验数据10:已经测量了各种原型蛋白质制剂的粘度,从而生成了600个数据点的数据集1。

[0045] 2. 以相关物理参数以及分子指纹的形式的赋形剂的表示2。那些都是经由计算机中模拟和实验交叉验证生成的。

[0046] 3. 机器学习模型5,其使用来自步骤2的表示2来识别来自步骤1的数据中的模式,并预测新的蛋白质-赋形剂制剂8的粘度3。

[0047] 在图1中示意性地描述了与开发的软件工具7的预期交互。图2示出了关于参与硬件的概述。除了必要的实验室装备之外,硬件主要由托管软件7的合适的计算机6组成,软件7操作所使用的机器学习模型5。可以使用适合与相应软件7一起使用的每种计算机6,例如标准个人计算机或工业pc。

[0048] 通过测量包含蛋白质的溶液/制剂8的粘度和包含相同蛋白质溶液并另外包含至少一种粘度降低赋形剂2的溶液的粘度生成数据集1。优选地,至少一种粘度降低赋形剂2是单一粘度降低赋形剂或两种粘度降低赋形剂的组合。

[0049] 为了测量粘度降低,将不含粘度降低赋形剂2或粘度降低赋形剂组合的蛋白质组合物的粘度与包含粘度降低赋形剂或粘度降低赋形剂组合的蛋白质组合物的粘度进行比较。

[0050] 测量是用以定义的浓度的不同蛋白质进行的。使用以定义的浓度的不同粘度降低赋形剂2或粘度降低赋形剂组合。

[0051] 通常,蛋白质组合物是液体组合物,并且另外包含至少一种缓冲剂和至少一种稳定剂。根据蛋白质选择缓冲液和pH,并且通常使用NaOH或HCl调节pH。该组合物还可以包括药学上可接受的稀释剂、溶剂、载体、粘合剂、结合剂、防腐剂、增溶剂、稳定剂、表面活性剂、渗透促进剂、乳化剂或生物利用度促进剂。技术人员9知道如何针对液体组合物选择合适的添加剂和参数。

[0052] 在优选的实施例中,根据本发明的组合物是液体制剂8,并且蛋白质是治疗性蛋白质。

[0053] 治疗性蛋白质包括基于抗体的药物、Fc融合蛋白、抗凝血剂、血液因子、骨形态发生蛋白、工程蛋白支架、酶、生长因子、激素、干扰素、白细胞介素、抗体药物偶联物(ADC)和溶栓剂。治疗性蛋白质可以是天然存在的蛋白质或重组蛋白质。它们的序列可以是天然的或工程化的。

[0054] 在特别优选的实施例中,根据本发明的组合物和制剂中的蛋白质是抗体,特别是治疗性抗体。

[0055] 在进一步特别优选的实施例中,根据本发明的组合物和制剂中的蛋白质是血浆衍生的蛋白质,特别是IgG或hyperIgG。一些包含血浆蛋白的药物制剂由不同血浆蛋白的混合物组成。

[0056] 本文中的术语“血浆衍生的蛋白质”是指通过血浆分级分离从供体的血浆衍生的

蛋白质。所述供体可以是人类或非人类。血浆蛋白的一个示例是免疫球蛋白。

[0057] 本文中的术语“IgG”是指G型免疫球蛋白。本文中的术语“IgM”是指M型免疫球蛋白。本文中的术语“IgA”是指A型免疫球蛋白。

[0058] 本文中的术语“hyper-IgG”是指从已经被特定疾病感染或接种了针对特定疾病的疫苗的供体中纯化的IgG的制剂。所述供体可以是人类或非人类。

[0059] 本文中的术语“抗体”是指单克隆抗体(包括全长或完整单克隆抗体)、多克隆抗体、多价抗体、多特异性抗体(例如双特异性抗体)和抗体片段。

[0060] 抗体片段仅包括完整抗体的一部分,通常包括完整抗体的抗原结合位点,并且因此保留了结合抗原的能力。由本定义包含的抗体片段的示例包括:Fab片段、Fab'片段、Fd片段、Fd'片段、Fv片段、dAb片段、分离的CDR区、F(ab')₂片段以及单链抗体分子、双抗体和线性抗体。

[0061] 在一个实施例中,蛋白质是生物仿制药(biosimilar)。“生物仿制药”在本文中被定义为与另一种已经批准的生物药物高度相似的生物药物。在优选的实施例中,生物仿制药是单克隆抗体。

[0062] 在一个实施例中,根据本发明的组合物和制剂包括多于一种蛋白质种类。

[0063] 本发明不限于特定分子量范围的蛋白质。优选地,蛋白质分子量在120kDa和250kDa之间,优选在130kDa和180kDa之间。

[0064] 选择增加溶液8的粘度的一种或多种蛋白质浓度,以便测试粘度降低赋形剂2的粘度降低。所得溶液8的粘度应该具有至少20-25mPas⁻¹的粘度。在优选的实施例中,根据本发明的组合物和制剂中的蛋白质浓度为至少1mg/ml,至少50mg/ml,优选至少75mg/ml,并且更优选至少100mg/ml。在另一个优选的实施例中,蛋白质浓度在90mg/ml和300mg/ml之间,更优选蛋白质浓度在100和250mg/ml之间,甚至更优选在120和210mg/ml之间。本发明对于这些高浓度蛋白质组合物特别有用。

[0065] 对于选择用于生成数据集的蛋白质没有限制。例如,以下蛋白质可以用于建立数据集:Cetuximab、Evolocumab、Infliximab、Reslizumab、Etanercept(融合蛋白)。

[0066] 如本文中所定义的,“粘度”是指物质(通常是液体)流动的阻力。粘度与剪切力的概念有关;它可以被理解为不同流体层在它们相对彼此移动时,在彼此上或在其他表面上施加剪切力的效果。有几种用于表达粘度的方式。粘度的单位是Ns/m²,称为帕斯卡-秒(Pas)。粘度可以是“运动的”或“绝对的”。运动粘度是动量通过流体传递的速率的量度。它是斯托克斯(St)来测量的。运动粘度是流体在重力的影响下的阻力流动的量度。当体积相等但粘度不同的两种流体被放置在相同的毛细管粘度计中并允许通过重力流动时,粘性较大的流体比粘性较小的流体花费更长的时间流过毛细管。例如,如果一种流体花费200秒(s)来完成其流动,而另一种流体花费400s,则在运动粘度标度上,第二种流体被称为第一种流体的两倍粘性。运动粘度的量纲是长度²/时间。通常,运动粘度以厘斯托克斯(cSt)表达。运动粘度的SI是mm²/s,其等于1cSt。有时称为“动态粘度”或“简单粘度”的“绝对粘度”,是运动粘度和流体密度的乘积。绝对粘度以厘泊(cP)为单位表达。绝对粘度的SI是毫帕斯卡-秒(mPas),其中1cP=1mPas。

[0067] 粘度可以通过使用例如粘度计在给定的剪切速率或多个剪切速率下测量。“外推零剪切”粘度可以通过在绝对粘度对剪切速率的图上创建四个最高剪切点的最佳拟合线,

并将粘度线性外推回到零剪切来确定。替代地,对于牛顿流体,粘度可以通过在多个剪切速率下平均粘度值来确定。还可以使用微流体粘度计在单个或多个剪切速率(也称为流速)下测量粘度,其中绝对粘度源自液体流过通道时的压力中的变化。粘度等于剪切应力除以剪切速率。在一些实施例中,用微流体粘度计测量的粘度可以直接与外推的零剪切粘度(例如从使用锥板粘度计在多个剪切速率下测量的粘度外推的那些粘度)进行比较。根据本发明,当上述方法中的至少一种示出稳定效果时,组合物和制剂8的粘度降低。优选地,使用mVROCTM技术在20°C下测量粘度。更优选地,在20°C下使用mVROCTM技术测量粘度。最优选地,使用mVROCTM技术并使用500 μ l注射器、3000 s^{-1} 或2000 s^{-1} 的剪切速率和200 μ l的体积在20°C下测量粘度。本领域中的普通技术人员熟悉使用mVROCTM技术的粘度测量,尤其是熟悉选择上述参数。详细规格、方法和设置可以在901003.5.1-mVROC_User's_Manual中找到。

[0068] 本文中的“剪切速率”是指一层流体以其来通过相邻层的速度的变化率。速度梯度是速度随离板距离的变化率。该简单的情况示出了关于剪切速率($v_1 - v_2$)/h的均匀速度梯度,单位为 $(\text{cm}/\text{sec})/(\text{cm}) = 1/\text{sec}$ 。因此,剪切速率单位是秒的倒数,或者一般来说,是时间的倒数。对于微流体粘度计,压力和流速中的变化与剪切速率相关。“剪切速率”是指材料以其来变形的速度。包含蛋白质和粘度降低剂的制剂8通常在范围从约0.5 s^{-1} 至约200 s^{-1} 的剪切速率下测量,当使用锥板粘度计和由本领域技术人员适当选择的轴来准确测量感兴趣的样品的粘度范围内的粘度时(即,20cP的样品在固定于DV2T粘度计(Brookfield)的CPE40轴上测量最准确);当使用微流体粘度计测量时,大于约20 s^{-1} 至约3000 s^{-1} 。

[0069] 对于如本文中通常使用的经典“牛顿”流体,粘度基本上与剪切速率无关。然而,对于“非牛顿流体”,粘度随着剪切速率的增加而降低或增加,例如,流体分别是“剪切稀化”或“剪切增稠”。在浓缩(即高浓度)蛋白质溶液的情况下,这可能表现为假塑性剪切稀化行为,即粘度随剪切速率中的降低。

[0070] 在一个实施例中,与不包括至少一种第一赋形剂的相同组合物相比,本发明的组合物和制剂示出至少5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%或75%的粘度的降低。

[0071] 在一个实施例中,与不包括至少一种第一和至少一种第二赋形剂2的相同组合物相比,本发明的组合物和制剂示出至少5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%或75%的粘度的降低。

[0072] 本发明进一步提供了根据本发明的药物制剂8,其中粘度在1mPas和60mPas之间,优选在1mPas和50mPas之间,更优选在1mPas和30mPas之间,最优选在1mPas和20mPas之间。

[0073] 组合物通常具有在4和8之间的pH,优选在5和7.2之间的pH。在一个实施例中,组合物和制剂8具有正好为5或正好为7.2的pH。根据蛋白质选择pH,并且通常使用NaOH或HCl调节pH。技术人员知道如何选择蛋白质组合物的pH。

[0074] 至少一种稳定剂是适合于增加蛋白质的稳定性的化合物。合适的稳定剂是本领域中已知的,并且包括合适的糖和/或表面活性剂。合适的糖作为稳定剂是文献中已知的,例如蔗糖或海藻糖。在优选的实施例中,糖是蔗糖。合适的表面活性剂是文献中已知的,例如聚山梨醇酯20或聚山梨醇酯80或泊洛沙姆188。在另一个优选的实施例中,表面活性剂是聚山梨醇酯80。添加另外的稳定剂附加地增强了根据本发明的组合物的稳定效果。优选地,糖具有50至100mg/ml、更优选50mg/ml蔗糖的浓度。优选地,表面活性剂具有0.01至0.2mg/ml、

更优选0.05mg/ml的聚山梨醇酯80的浓度。

[0075] 添加适合于蛋白质溶液的至少一种缓冲剂以制备缓冲溶液。合适的缓冲液是本领域中已知的,例如醋酸盐-柠檬酸盐-或磷酸盐磷酸盐缓冲液。缓冲液通常具有1至50mM的浓度

[0076] 根据本发明,“粘度改变赋形剂2”是可以影响液体制剂的粘度的化合物。该定义包括“粘度降低赋形剂2”,当以如下定义的浓度范围添加到制剂8中时,其适合于降低液体制剂8的粘度。优选地,液体制剂8是蛋白质溶液。

[0077] 对于选择用于生成数据集1的粘度降低赋形剂2没有限制。以下粘度降低赋形剂2可以用于建立数据集:盐酸胍、L-精氨酸、L-肉碱盐酸盐、L-鸟氨酸盐酸盐、L-丝氨酸、赖氨酸、葡甲胺、盐酸奎宁、盐酸硫胺素、抗坏血酸、苯磺酸、樟脑磺酸、焦磷酸硫胺素、琥珀酸二钠、酒石酸二钠、叶酸、葡萄糖酸、葡萄糖醛酸、吡哆醇、对甲苯磺酸钠、硫胺素一磷酸盐、尿素、氨基己酸、咖啡因、氰钴胺、甘氨酸、异亮氨酸、亮氨酸、烟酰胺、苯丙氨酸、脯氨酸、氯化钠、缬氨酸。

[0078] 这样的粘度降低剂以适合于降低蛋白质溶液8的粘度的浓度使用。在优选的实施例中,根据本发明的组合物和制剂中的蛋白质浓度为至少1mg/ml,至少50mg/ml,优选至少75mg/ml,并且更优选至少100mg/ml。在另一个优选的实施例中,蛋白质浓度在90mg/ml和300mg/ml之间,更优选蛋白质浓度在100和250mg/ml之间,甚至更优选在120和210mg/ml之间。优选地,一种粘度改变赋形剂2以最大200mM的浓度、更优选最大150mM、最优选75mM或150mM的浓度使用。在使用两种粘度改变赋形剂2的情况下,优选每种粘度改变赋形剂为最大150mM、更优选每种粘度改变赋形剂为最大100mM、最优选每种粘度改变赋形剂为75mM的浓度。由于超过300mM的所得到的浓度水平是不优选的,因此两种赋形剂2的浓度都不应超过150mM。如果由于任何原因,在优选的两种赋形剂2之间存在不均匀的分布,则比率分别变化。如果使用多于两种赋形剂2,同样的规则也适用。那些浓度的水平对于降低粘度是最有效的,并且因此是优选的。然而,该方法不限于这些特定值。

[0079] 赋形剂数据集1基于简化的分子输入行输入系统(SMILES)表示。每次粘度测量都是在定义的但以它们的pH值为特征的不同环境下执行的。为了结合赋形剂质子化中的变化,使用ChemAxons预测器生成pH依赖性微物种分布。pH依赖性在pH4-pH8之间。使用Marvin molconverter将每个微物种转换成三维结构。根据三维试验结构,计算了室温下构成水溶液中所有构象异构体的集合。为此,采用了CREST算法,该算法是在包括广义玻恩和表面可及区域隐式溶剂化模型(GFN2-xTB+GBSA(水))的扩展紧密结合水平的量子力学势能表面上运行的基于元动力学(meta-dynamics)、结构交叉、模拟退火的全局搜索。经由刚性转子谐振子(RRHO)模型包括零点和热力学贡献。在其2019.0.4参数化的真实溶剂的类导体筛选模型(COSMO-RS)中,已经使用密度泛函近似B97-3c进一步细化各个几何形状。用于构象异构体集合的玻尔兹曼布居的最终单点由电子能量、RRHO贡献和溶剂化自由能组成。忽略具有低于1%的贡献的结构。

[0080] 那些微物种集合是在密度泛函理论水平上进行量子化学计算的基础,以模拟分子可观察量,比如电荷分布、偶极矩、四极矩轨迹和各向异性、极化率、分子伦敦分散系数(C6)、logP水/己烷分布系数、溶剂可及表面积、分子轨道能量HOMO-LUMO间隙。这些量子力学特征补充有一组拓扑分子指纹。这个200个标准化指纹的扩增集是基于使用RDKit的相同

微物种集合生成的。

[0081] 总之,这为每种单一赋形剂2产生了高维pH依赖性特征向量。

[0082] 开发的机器学习模型5将实验室中获得的实验数据与计算的计算机上的赋形剂特征相结合,以构建制剂粘度的预测模型。基于该模型5,提供最优实验调度。制剂专家9使用这些建议并执行推荐的实验10,并且随后将新获得的粘度数据馈入到系统中,如图3示例性所示。通过该过程,实验的执行集中在具有最高的粘度降低的可能性的制剂上。

[0083] 对于给定的蛋白质,粘度值以高斯过程(GP)的形式建模,并且在优选实施例中模型5的预测用于通过贝叶斯最优实验设计来指导制剂专家9。该指导包括几个步骤:

[0084] 1. 给定某些赋形剂/赋形剂组合(=训练数据)的一组可能为空的粘度测量值,通过使训练数据的边际可能性最大化来优化GP模型参数。

[0085] 2. 基于GP模型评估未测试赋形剂/赋形剂组合的粘度值的后验分布。

[0086] 3. 通过优化从计算的后验分布获得的采集分数来选择一组新的赋形剂/赋形剂组合。

[0087] 4. 向制剂专家9提议一组新的赋形剂/赋形剂组合,然后制剂专家9在实验室进行相应的实验以确定所得到的粘度。

[0088] 5. 将获得的测量值添加到训练数据,并从步骤1重复该过程。

[0089] 步骤2中的预测基于表征所考虑的制剂8中使用的赋形剂2的pH依赖性特征向量,以及基于赋形剂浓度水平。

[0090] 在步骤1至5中该过程的挑战是,测量的粘度不仅取决于所选择的赋形剂组合,而且还取决于真实蛋白质浓度,该真实蛋白质浓度在每次测量中可能不同。为了考虑与目标浓度的偏差,GP模型5被设计用于预测粘度中的相对变化,而不是绝对粘度值。更精确地说,它预测了相对于理论粘度水平的相对粘度降低,这将在没有赋形剂的实际蛋白质浓度下实现。所需的理论值从指数回归模型中获得,该指数回归模型根据未配制的蛋白质溶液的浓度依赖性粘度测量值来计算。

[0091] 虽然步骤1至5中的考虑的实验的设计遵循典型的优化过程,但是由于需要编码的特定数据特性,由领先软件套件(诸如GPpyTorch、BoTorch、GPflow)提供的现有黑盒GP模型5不能应用于给定场景。因此,已经设计专门的GP内核结构来考虑以下领域和问题特定的性质。那些性质是:

[0092] • 给定制剂中所含的赋形剂2的组合没有自然的顺序,即添加赋形剂A+赋形剂B相当于添加赋形剂B+赋形剂A。所使用的内核被设计成相对于添加的赋形剂2是排列不变的。

[0093] • 给定的制剂8可以包含不同数量的赋形剂2。内核已被构造以处理灵活的赋形剂数量。

[0094] • 制剂8引起的粘度降低效果取决于给定的蛋白质浓度和应用的赋形剂浓度两者。对这些浓度的依赖性明确地反映在所用内核的结构中。

[0095] • 组合赋形剂2可以导致协同粘度降低效果,这可能无法通过单独的每种赋形剂的特性来描述。基于各个特征维度的自动相关性检测的常见的通用内核结构无法充分捕获这些多元关系。为了从测量数据泛化(generalize)到未测试的赋形剂组合,所使用的内核使用在参数拟合过程期间优化的线性子空间投影。

[0096] 特别具有挑战性的是将粘度预测3泛化到新的蛋白质。这是由于缺少表征蛋白质

的全局以及局部相互作用的化学信息。因此,推荐进一步扩展的优选实施例。它由包含各种制剂8的粘度测量值的数据库组成,所述各种制剂8构成蛋白质和赋形剂2之间的原型相互作用模式。这些相互作用模式可以用作新蛋白质的粘度预测3的先验信息,其形式是使预测偏向匹配的蛋白质-赋形剂模式的那些的附加内核成分。实现这一点的一种方法(尽管其他方法也是可能的)是通过经由多任务内核模型(诸如内在协同区域化模型或其变体)捕获蛋白质影响。在本发明的替代实施例中,其中高斯过程用其他机器学习模型替换,跨蛋白质泛化的任务可以由其他适当的模型组件接管。

[0097] 在下文中,公开了具体的工作示例以展示与执行其中随机选择下一个实验的不知情的(uninformed)搜索相比,使用软件工具7的优点。

[0098] 目标是将蛋白质溶液8的粘度降低到指定阈值以下。鉴于市场上可获得的赋形剂2的广阔前景,难以找到合适的赋形剂组合。为了避免其中对所有候选制剂进行测试的详尽的筛选研究,在所提议的软件工具7的帮助下,执行知情的、数据驱动搜索。

[0099] 为此,执行以下步骤:

[0100] 1) 定义具体的制剂条件,特别是pH值和哪些赋形剂2可以被认为是潜在的候选物。

[0101] 2) 对包含新蛋白质但不含至少一种粘度降低剂的溶液执行少量浓度依赖性粘度测量。该数据被馈入到软件工具7中,以基于粘度预测3估计未配制蛋白质的基础粘度曲线。

[0102] 通过在进行每次测量之后查阅软件工具7,用户被指示接下来考虑哪个蛋白质浓度,并且一旦已经收集了足够量的数据,用户就被告知。

[0103] 3) 软件工具7然后推荐要测试的第一赋形剂2或赋形剂组合。用户在实验室中进行相应的实验10,并将测量的粘度报告回工具。在迭代过程中,响应于报告给工具7的最新测量值,提示用户执行进一步的实验,直到找到具有足够低粘度的制剂8。

[0104] 如果在使用工具7之前已经进行了测量,例如对于不在候选列表上的赋形剂2,则用户可以在启动过程之前报告相应的粘度。这样,工具7可以从一开始就给出改进的推荐。

[0105] 在替代实施例中,用户可以在每次迭代之后查阅软件工具7之前,一次执行几个实验。在这种所谓的“批处理模式”中,用户可以输入在下次迭代期间要并行执行的实验的期望数量,例如为了调度实验室资源的目的。然后,软件工具7将以优化同时进行实验所产生的预期信息增益这样的方式来优化其推荐。

[0106] 图4示出了在有用户进行的多次实验中,两种搜索策略实现的粘度降低。对于给定的示例,总共考虑了629个实验,其涵盖6种蛋白质和33种赋形剂。为了对所有测试的蛋白质的结果进行平均,相对于每个蛋白质观察到的最大降低来报告测量的粘度降低,并且相对于每个蛋白质进行的实验10的总数来示出实验步骤的数量。描绘的是从几次重复实验中获得的得到的平均值(实线)和标准偏差(阴影区域)。这些重复是通过考虑提供给软件工具7的不同组初始测量值和随机基线策略的不同随机实验路径而获得的。

[0107] 与理论所需步骤的数量一致,随机策略在进行50%的所有可能的实验后,在预期中找到最优赋形剂组合。使用本发明的方法,这个数字平均可以减少一半。

[0108] 本发明的进一步的实施例是新制剂8,其包含经由上面提供的方法选择的至少一种粘度改变赋形剂2。

[0109] 本发明的进一步的实施例是一种药物制剂,其包含新制剂8和经由上面提供的方法选择的至少一种粘度改变赋形剂2。

[0110] 本发明的进一步的实施例是一种药物制剂,其包含新制剂8和经由上面提供的方法选择的至少一种粘度改变赋形剂2。

[0111] 示例

[0112] 1.生成实验数据/粘度测量值

[0113] 实验的一般概念

[0114] 为了生成数据集1的实验数据,已经制备了各种蛋白质组合物,并测试了不同粘度降低赋形剂的粘度降低。

[0115] 使用以下商业上可获得的蛋白质:西妥昔单抗(Cetuximab)、依洛尤单抗(Evolocumab)、英夫利昔单抗(Infliximab)、瑞利珠单抗(Reslizumab)、依那西普(Etanercept)。

[0116] 使用以下商业上可获得的粘度降低赋形剂2:盐酸胍、L-精氨酸、L-肉碱盐酸盐、L-鸟氨酸盐酸盐、L-丝氨酸、赖氨酸、葡甲胺、盐酸奎宁、盐酸硫胺素、抗坏血酸、苯磺酸、樟脑磺酸、焦磷酸硫胺素、琥珀酸二钠、酒石酸二钠、叶酸、葡萄糖酸、葡萄糖醛酸、吡哆醇、对甲苯磺酸钠、硫胺素一磷酸盐、尿素、氨基己酸、咖啡因、氰钴胺、甘氨酸、异亮氨酸、亮氨酸、烟酰胺、苯丙氨酸、脯氨酸、氯化钠、缬氨酸及其组合。

[0117] 在下文中,举例说明了作为粘度降低赋形剂2的缬氨酸对英夫利昔单抗溶液的粘度降低的测量。该特定示例的一般概念可以转移到所使用的所有其他蛋白质和粘度降低剂。

[0118] 在使用单一粘度降低剂的情况下,通常在150mM的浓度下测量粘度。在使用两种赋形剂2的组合的情况下,通常针对赋形剂中的每种在75mM的浓度下测量粘度。在一些情况下,根据赋形剂2的溶解度调节粘度降低赋形剂2的浓度。

[0119] 根据所使用的蛋白质,选择缓冲液、pH、蛋白质浓度和可选的稳定剂和/或表面活性剂。通常使用包含蛋白质的商业上可获得的产品的缓冲液、pH、稳定剂和/或表面活性剂。浓缩蛋白质溶液以产生具有至少20mPas⁻¹的粘度的溶液。在一些情况下,粘度是在多于一种蛋白质浓度下测量的。

[0120] 粘度测量

[0121] 缓冲液制备

[0122] 通过适当混合磷酸二氢钠和磷酸氢二钠以产生7.2的pH并将混合物溶解在超纯水中来制备5mM磷酸盐缓冲液。使用Henderson-Hasselbalch方程确定该比率。在必要时使用HCl和NaOH调节pH。添加50mg/ml蔗糖和0.05mg/ml聚山梨醇酯80作为稳定剂。

[0123] 样品制备

[0124] 在pH 7.2的磷酸盐缓冲液中制备150mM缬氨酸的单独赋形剂溶液。在必要时使用HCl或NaOH调节pH。

[0125] 使用离心过滤器(Amicon, 30kDaMWC0)来将原始缓冲液与包含相应赋形剂的缓冲液交换并减少溶液8的体积,来制备包含期望的赋形剂的浓缩英夫利昔单抗溶液。随后将蛋白质分别稀释至122mg/ml和143mg/ml。以类似的方式,制备了不含缬氨酸的其他方面相同的蛋白质溶液。

[0126] 蛋白质浓度测量

[0127] 使用应用Lambert-Beer定律的吸收光谱法确定蛋白质浓度。当赋形剂本身在

280nm处具有强吸光度时,使用Bradford测定法。

[0128] 稀释浓缩的蛋白质溶液,使得它们在测量中的预期浓度将在0.3和1.0mg/mL之间。

[0129] 对于吸收光谱法,使用**BioSpectrometer**[®]kinetic (Eppendorf, Hamburg, 德国) 测量280nm处的吸光度,其中蛋白质消光系数为 $A_{0.1\%}^{280\text{nm}}=1.428$ 。

[0130] 一些赋形剂2本身在280nm处具有强吸光度,这使得有必要使用Bradford测定法来确定浓度。对于Bradford测定法,使用了来自Thermo Scientific TM (Thermo Fisher, Waltham, Massachusetts, 美国) 的试剂盒以及牛丙种球蛋白标准品。使用MultiskanTM Wellplaterereader (Thermo Fisher, Waltham, Massachusetts, 美国) 在595nm处测量吸光度。蛋白质浓度通过从125到1500 $\mu\text{g}/\text{ml}$ 的标准曲线的线性回归来确定。

[0131] 粘度测量

[0132] mVROCTM技术 (Rheo Sense, San Ramon, California, 美国) 用于粘度测量。

[0133] 使用500 μl 注射器和3000 s^{-1} 的剪切速率在20 $^{\circ}\text{C}$ 下执行测量。使用200 μl 的体积。所有样品都一式三份测量。通过比较具有和不具有缬氨酸的蛋白质组合物的绝对粘度来计算粘度降低。

[0134] 2.原始蛋白质溶液 (raw protein solution) 的浓度依赖性粘度测量

[0135] 在下文中,举例说明了英夫利昔单抗的浓度依赖性粘度的测量。该特定示例的一般概念可以转移到所有其他蛋白质。

[0136] 如上所述执行缓冲液和样品制备。随后将蛋白质分别稀释至13、30、42、68、79、80、103、110、117.30、121和148.2 mg/ml 。如上所述,使用应用Lambert-Beer定律的吸收光谱法执行蛋白质浓度测量。使用mVROCTM技术 (RheoSense, San Ramon, California, 美国) 执行不同英夫利昔单抗浓度的粘度测量。

[0137] 参考的列表

[0138] 1 数据集

[0139] 2 赋形剂 (表示)

[0140] 3 预测粘度

[0141] 4 选择的赋形剂 (组合)

[0142] 5 机器学习模型

[0143] 6 使用的计算机

[0144] 7 软件工具

[0145] 8 新制剂

[0146] 9 用户 (制剂专家)

[0147] 10 实验测量数据

[0148] 11 未知蛋白质

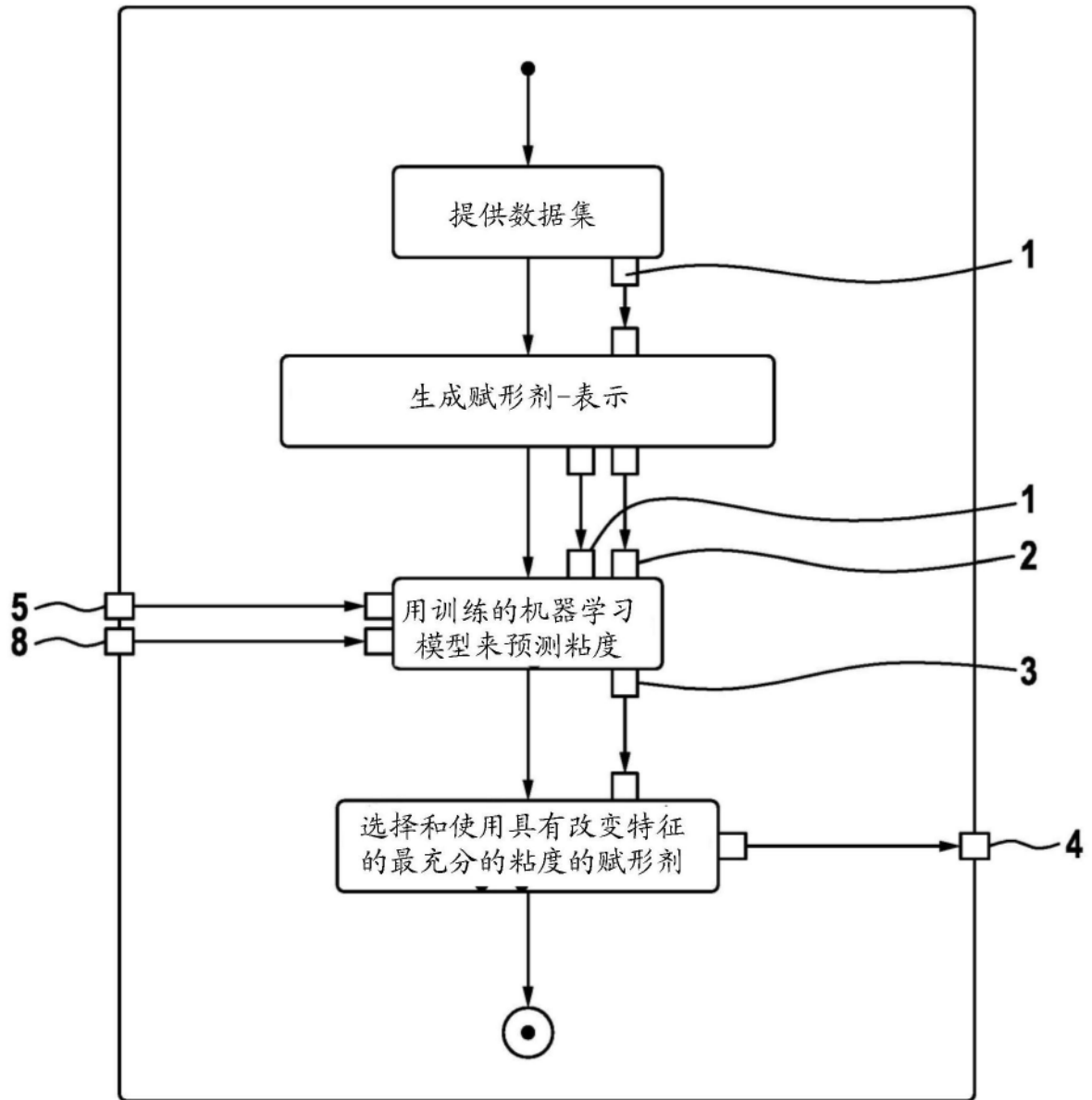


图1

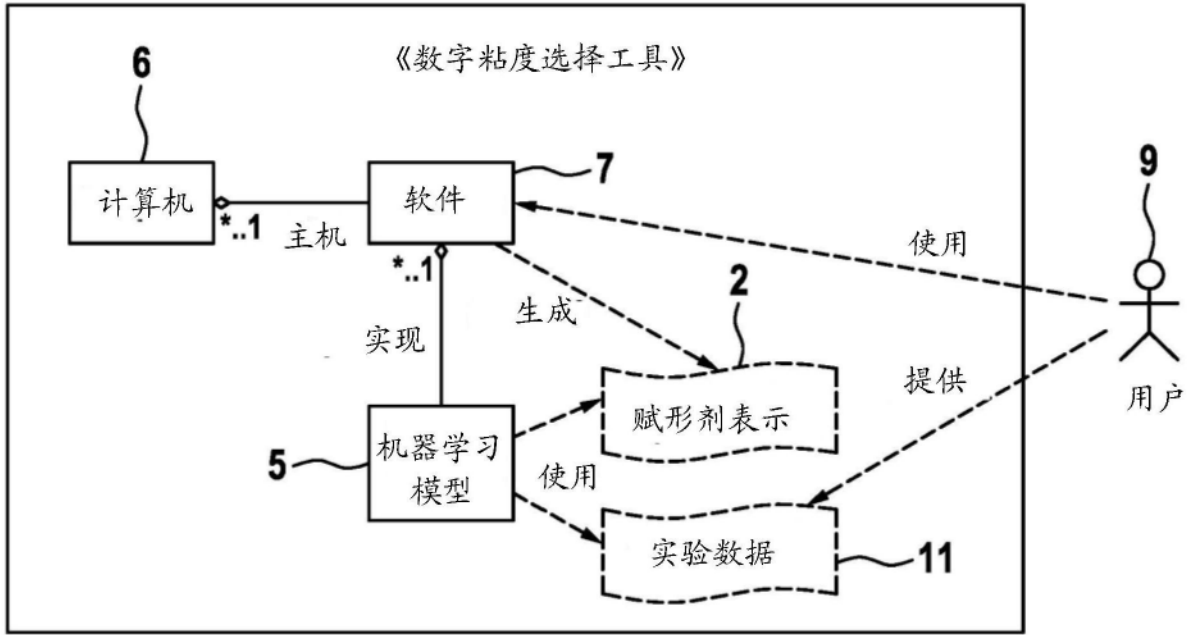


图2

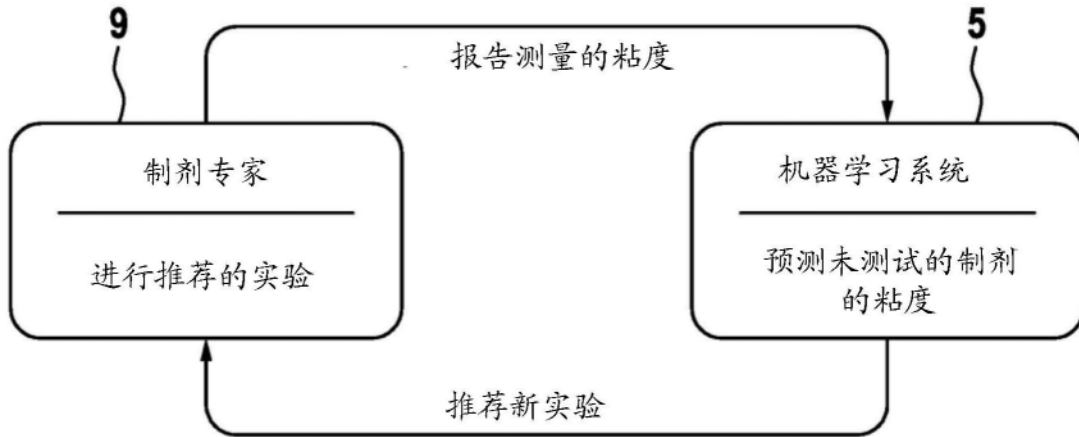


图3

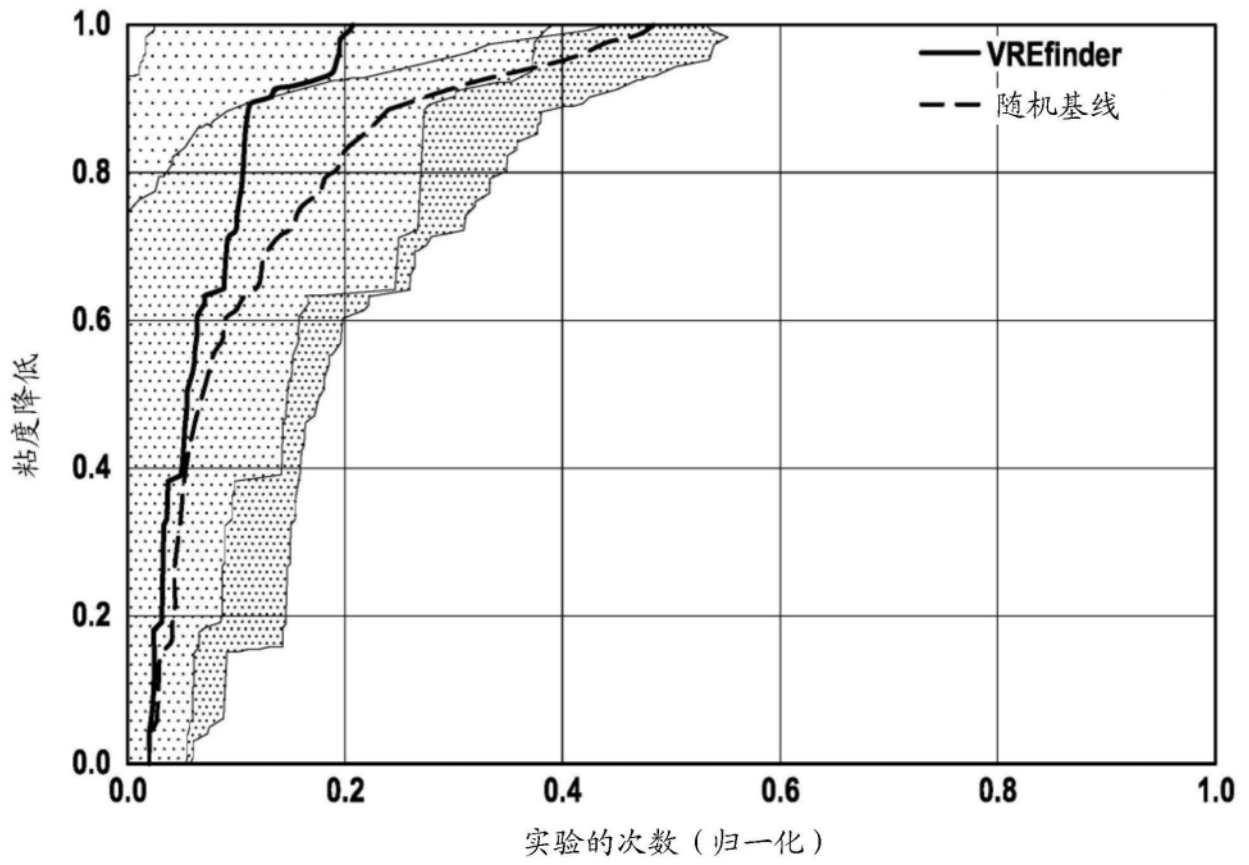


图4