



(12) 发明专利申请

(10) 申请公布号 CN 116034263 A

(43) 申请公布日 2023. 04. 28

(21) 申请号 202180052867.X

(22) 申请日 2021.07.01

(30) 优先权数据

2010104.4 2020.07.01 GB

(85) PCT国际申请进入国家阶段日

2023.02.24

(86) PCT国际申请的申请数据

PCT/GB2021/051673 2021.07.01

(87) PCT国际申请的公布数据

W02022/003359 EN 2022.01.06

(71) 申请人 细胞治疗弹射器有限公司

地址 英国伦敦

(72) 发明人 J·丘奇韦尔 M·O·巴拉迪兹

D·马歇尔

(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205

专利代理师 李艳 黄健

(51) Int.Cl.

G01N 21/65 (2006.01)

权利要求书7页 说明书38页 附图39页

(54) 发明名称

用拉曼光谱分析病毒的方法

(57) 摘要

本发明涉及使用拉曼光谱来监测和评估病毒滴度和/或病毒组分丰度。

1. 一种使用拉曼光谱确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率的方法,所述方法包括以下步骤:

(a) 提供样本并用光源照射所述样本;

(b) (i) 测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第一波数强度数据集,其中所述第一多个波数范围是预先选择的、且是所述样本中的病毒核酸的表征;

(ii) 对所述第一波数强度数据集执行第一组数学数据处理步骤;以及

(iii) 基于所述第一组数学数据处理步骤的输出,确定所述样本中的所述病毒核酸含量;

(c) (i) 测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第二波数强度数据集,其中所述第二多个波数范围是预先选择的且是所述样本中的病毒的所述一个或多个病毒结构分子的表征;

(ii) 对所述第二波数强度数据集执行第二组数学数据处理步骤;以及

(iii) 基于所述第二组数学数据处理步骤的输出,确定所述样本中的包括所述一个或多个病毒结构分子的病毒的含量;以及

(d) 基于步骤(b) (iii) 和(c) (iii) 中确定的值,确定所述样本中的病毒核酸与包括所述一个或多个病毒结构分子的病毒的比率。

2. 根据权利要求1所述的方法,其中,对所述第一波数强度数据集和第二波数强度数据集执行所述第一组数学数据处理步骤和第二组数学数据处理步骤的步骤包括:

(i) 可选地,通过使用一种或多种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量(SNV)法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合(PPF)法或自适应迭代重加权惩罚最小二乘(airPLS)法,对信号强度数据进行预处理,对波数信号强度数据进行归一化;

(ii) 通过将多变量回归算法,例如偏最小二乘(PLS)回归算法,应用于所述波数信号强度数据来获得模型参数,可选地,其中所述PLS算法为非线性迭代偏最小二乘(NIPALS)回归算法或神经网络;以及

(iii) 利用对所述信号强度数据应用所述多变量回归算法而获得的模型参数,确定所述样本中的所述病毒核酸的含量,并确定所述样本中的包括所述一个或多个病毒结构分子的病毒的含量。

3. 根据权利要求1或2所述的方法,其中所述用于照射所述样本的所述光源是激光器,并且所述样本用波长为785nm的光照射。

4. 根据前述权利要求中任一项所述的方法,其中,使用电荷耦合器件(CCD)来检测所述拉曼散射光。

5. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表1中所列出的波数范围1至12中的4个或更多个,并且其中VIP $\geq$ 1.00;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围1至12中的6个或更多个,并且其中所述VIP $\geq$ 1.00;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围1至12中的8个或更多个,并且其中所述VIP $\geq$ 1.00;或者其中所测量的所述拉曼光谱中

的所述多个波数范围包括如表1中所列出的波数范围1至12中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围1至12中的所有12个,并且其中所述VIP ≥ 1.00 。

6. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表1中所列出的波数范围13至22中的4个或更多个,并且其中VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围13至22中的6个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围13至22中的8个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围13至22中的所有10个,并且其中所述VIP ≥ 1.25 。

7. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表1中所列出的波数范围23至30中的4个或更多个,并且其中VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围23至30中的6个或更多个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表1中所列出的波数范围23至30中的所有8个,并且其中所述VIP ≥ 1.50 。

8. 根据权利要求1-7中任一项所述的方法,其中,为获得所述样本的所述第二波数强度数据集所测量的所述拉曼光谱中的所述第二多个波数范围包括如表2中所列出的波数范围1至20中的4个或更多个,并且其中VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围1至20中的6个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围1至20中的8个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围1至20中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围1至20中的12个或更多个、14个或更多个、16个或更多个或18个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围1至20中的所有20个,并且其中所述VIP ≥ 1.00 。

9. 根据权利要求1-7中任一项所述的方法,其中,为获得所述样本的所述第二波数强度数据集所测量的所述拉曼光谱中的所述第二多个波数范围包括如表2中所列出的波数范围21至33中的4个或更多个,并且其中VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围21至33中的6个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围21至33中的8个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括表2中所列出的波数范围21至33中的10个或更多个、11个或更多个或12个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围21至33中的所有13个,并且其中所述VIP ≥ 1.25 。

10. 根据权利要求1-7中任一项所述的方法,其中,为获得所述样本的所述第二波数强度数据集所测量的所述拉曼光谱中的所述第二多个波数范围包括如表2中所列出的波数范

围34至40中的4个或更多个,并且其中VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括表2中所列出的波数范围34至40中的5个或6个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表2中所列出的波数范围34至40中的所有7个,并且其中所述VIP ≥ 1.50 。

11. 根据前述权利要求中任一项所述的方法,其中所述样本中的病毒是腺相关病毒(AAV)。

12. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表3中所列出的波数范围1至28中的5个或更多个,并且其中变量重要性投影(VIP) ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括表3中所列出的波数范围1至28中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围1至28中的15个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围1至28中的20个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围1至28中的25个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围1至28中的所有28个,并且其中所述VIP ≥ 1.00 。

13. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表3中所列出的波数范围29至59中的5个或更多个,并且其中变量重要性投影(VIP) ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括表3中所列出的波数范围29至59中的10个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围29至59中的15个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围29至59中的20个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围29至59中的25个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围29至59中的30个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围29至59中的所有31个,并且其中所述VIP ≥ 1.25 。

14. 根据权利要求1-4中任一项所述的方法,其中,为获得所述样本的所述第一波数强度数据集所测量的所述拉曼光谱中的所述第一多个波数范围包括如表3中所列出的波数范围60至81中的5个或更多个,并且其中变量重要性投影(VIP) ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括表3中所列出的波数范围60至81中的10个或更多个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围60至81中的15个或更多个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围60至81中的20个或更多个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如表3中所列出的波数范围60至81中的所有22个,并且其中所述VIP ≥ 1.50 。

15. 根据权利要求12-14中任一项所述的方法,其中所述样本中的病毒是慢病毒。
16. 根据前述权利要求中任一项所述的方法,其中所述核酸包括病毒DNA基因组或其中所述核酸包括病毒RNA基因组。
17. 根据前述权利要求中任一项所述的方法,其中所述一个或多个病毒结构分子包括一个或多个病毒蛋白,例如一个或多个核蛋白,和/或一个或多个壳粒,一个或多个病毒碳水化合物,一个或多个糖基化病毒分子,例如糖基化病毒蛋白,和/或一个或多个病毒脂质。
18. 根据前述权利要求中任一项所述的方法,其中所述比率提供功能性病毒滴度的量度。
19. 根据前述权利要求中任一项所述的方法,其中所述样本为病毒培养物,可选地,其中所述病毒培养物包括在生物反应器中。
20. 根据权利要求19所述的方法,其中,用光源照射所述病毒培养物并测量所述拉曼散射光的总强度的步骤直接在病毒培养物的介质上(原位)执行。
21. 根据权利要求19所述的方法,其中,用光源照射所述病毒培养物并测量所述拉曼散射光的总强度的步骤直接在取自所述病毒培养物的介质的等分试样上(非原位)执行。
22. 根据前述权利要求中任一项所述的方法,所述方法包括在第一个时间点确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率的第一步,以及在随后的时间点确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率的一个或多个进一步步骤,并且还包括在时间点之间测量所述样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的所述比率的变化,其中每个步骤由根据前述权利要求中任一项所述的方法执行,优选地,其中每一步骤由相同的方法执行。
23. 根据权利要求22所述的方法,其中,在一段时间内重复确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率,以实时测量所述比率的变化。
24. 根据权利要求22或23所述的方法,其中所述样本中的所述比率的变化用于确定病毒生产过程的开始阶段、生产阶段和/或稳定阶段。
25. 根据前述权利要求中任一项所述的方法,其中所述方法用于确定病毒生产过程的最佳条件。
26. 根据前述权利要求中任一项所述的方法,其中所述方法用于评估病毒生产过程的下游过程。
27. 根据前述权利要求中任一项所述的方法,其中所述方法包括将由此获得的所述比率与通过一种替代方法从同一样本中获得的所述比率进行比较的步骤,可选地,其中所述替代方法是qPCR、RT-qPCR、ELISA或通过透射电子显微镜进行目视确定。
28. 一种使用拉曼光谱确定个体中病毒感染程度的方法,所述方法包括通过执行权利要求1-23中任一项所述的方法来确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率,其中所述样本是先前从个体中获得的样本。
29. 根据权利要求28所述的方法,其中所述样本是血液、唾液、痰、血浆、血清、脑脊液、尿液或粪便的样本。
30. 根据权利要求28或29所述的方法,其中将来自受试者的样本中的所述比率与先前针对个体中的感染获得的一个或多个比率测量值进行比较,以便提供个体感染阶段的预后。

31. 一种使用拉曼光谱确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率的方法,所述方法包括以下步骤:

(a) (i) 提供所述样本的第一波数强度数据集,

其中所述第一数据集是通过用光源照射所述样本并测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度而获得的,其中所述拉曼光谱中的所述第一多个波数范围已被选择作为所述样本中的病毒核酸的表征;

(ii) 对所述第一波数强度数据集执行第一组数学数据处理步骤;以及

(iii) 基于所述第一组数学数据处理步骤的输出,确定所述样本中的所述病毒核酸的含量;

(b) (i) 提供所述样本的第二波数强度数据集,

其中所述第二数据集是通过用光源照射所述样本并测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度而获得,其中所述拉曼光谱中的所述第二多个波数范围已被选择作为所述样本中的所述病毒的一个或多个病毒结构分子的表征;

(ii) 对所述第二波数强度数据集执行第二组数学数据处理步骤;以及

(iii) 基于所述第二组数学数据处理步骤的输出,确定所述样本中的包括所述一个或多个病毒结构分子的病毒的含量;

(c) 基于步骤(a) (iii) 和(b) (iii) 中确定的值,确定所述样本中的病毒核酸与包括所述一个或多个病毒结构分子的病毒的所述比率。

32. 根据权利要求31所述的方法,其中所述方法根据权利要求2-30中任一项限定的步骤执行。

33. 拉曼光谱用于确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率的用途。

34. 根据权利要求33所述的用途,其中所述比率是基于在用光源照射所述样本后从所述样本中获得的拉曼散射光的强度的测量来确定的,其中所述拉曼散射光的强度是从拉曼光谱中的第一多个波数范围和拉曼光谱中的第二多个波数范围测量的,所述第一多个波数范围是对所述样本中病毒核酸的表征,所述第二多个波数范围是对所述样本中病毒的所述一个或多个病毒结构分子的表征。

35. 根据权利要求34所述的用途,其中所述样本为病毒培养物,可选地,其中所述病毒培养物包括在生物反应器中。

36. 根据权利要求35所述的用途,其中测量所述拉曼散射光的总强度的步骤直接在病毒培养物的介质上(原位)执行。

37. 根据权利要求35所述的用途,其中测量所述拉曼散射光的总强度的步骤直接在取自所述病毒培养物的介质的等分试样上(非原位)执行。

38. 根据权利要求34-37中任一项所述的用途,其中所述样本中的所述比率是在第一个时间点和之后的一个或多个时间点确定的,其中计算时间点之间的所述样本中的所述比率的变化。

39. 根据权利要求38所述的用途,其中所述样本中的所述比率被重复定量,以实时提供对所述比率的变化测量。

40. 根据权利要求33-39中任一项所述的用途,其中所述样本中的所述比率是通过执行

根据权利要求1-18中任一项所述的方法来确定的。

41. 根据前述权利要求中任一项所述的方法或用途,其中所述样本中的病毒不是HIV-1或HIV-1病毒样粒子(HIV-1VLP)。

42. 根据前述权利要求中任一项所述的方法或用途,其中所述拉曼光谱不是表面增强拉曼光谱。

43. 一种构建多变量数据处理模型的方法,所述模型能够从针对样本获得的拉曼光谱波数强度数据集来确定样本中包括一个或多个病毒结构分子的病毒的含量,所述方法包括:

(a) 提供样本并用光源照射所述样本;

(b) 测量多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的波数强度数据集,其中所述多个波数范围是预先选择的且是所述样本中的所述病毒的一个或多个病毒结构分子的表征;

(c) 通过使用一种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量(SNV)法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合(PPF)法或自适应迭代重加权惩罚最小二乘(airPLS)法,对信号强度数据进行预处理,来获得归一化的波数信号强度数据;

(d) 通过将多变量回归算法,例如偏最小二乘(PLS)回归算法,应用于所述预处理的信号强度数据来获得模型参数,可选地,其中所述PLS算法为非线性迭代偏最小二乘(NIPALS)回归算法或神经网络,其中,执行校准,其中,将所述预处理的信号强度数据与使用非拉曼光谱方法,例如qPCR、RT-qPCR、ELISA,或通过透射电子显微镜进行的目视确定在相同样本条件下获得的病毒滴度数据进行比较;

(e) 利用从所述预处理的数据获得的模型参数来推断响应值;以及

(f) 执行变量选择,可选地,变量重要性投影(VIP),和确定拉曼光谱变量;以及

(g) 可选地,通过重新应用步骤(d)-(f)进行进一步的一轮或多轮建模,其中不重要的变量被移除;

其中,使用源自多变量数据处理模型的针对所确定的拉曼光谱变量获得的模型参数,来确定样本中包括一个或多个病毒结构分子的病毒的含量。

44. 一种构建一个或多个多变量数据处理模型的方法,所述模型能够根据针对样本获得的拉曼光谱波数强度数据集来确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率,所述方法包括:

(a) 提供样本并用光源照射所述样本;

(b) (i) 测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第一波数强度数据集,其中所述第一多个波数范围是预先选择的且是所述样本中的所述病毒核酸的表征;

(ii) 测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第二波数强度数据集,其中所述第一多个波数范围是预先选择的且是所述样本中的所述病毒的一个或多个病毒结构分子的表征;

(c) 通过使用一种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量(SNV)法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合(PPF)法或自适应迭代重加

权惩罚最小二乘 (airPLS) 法, 对信号强度数据进行预处理, 来获得所述第一波数强度数据集和第二波数强度数据集的归一化的波数信号强度数据;

(d) 通过将多变量回归算法, 例如偏最小二乘 (PLS) 回归算法, 应用于所述预处理的信号强度数据集中的每一个来获得待应用于所述第一波数强度数据集和第二波数强度数据集的模型参数, 可选地, 其中所述 PLS 算法为非线性迭代偏最小二乘 (NIPALS) 回归算法或神经网络, 其中, 执行校准, 其中, 将所述预处理的信号强度数据与使用非拉曼光谱方法, 例如 qPCR、RT-qPCR、ELISA, 或通过透射电子显微镜进行的目视确定在相同样本条件下获得的病毒滴度数据进行比较;

(e) 利用从所述预处理的数据集中的每一个获得的模型参数来推断响应值; 以及

(f) 执行变量选择, 可选地, 变量重要性投影 (VIP), 和确定拉曼光谱变量; 以及

(g) 可选地, 通过重新应用步骤 (d) - (f) 对所述数据集中的任一个进行进一步的一轮或多轮建模, 其中不重要的变量被移除;

其中, 使用源自多变量数据处理模型的针对所识别的拉曼光谱变量获得的模型参数, 来确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率。

用拉曼光谱分析病毒的方法

技术领域

[0001] 本发明涉及使用拉曼光谱监测和评估病毒滴度和/或病毒组分丰度。

背景技术

[0002] 病毒载体的制造是许多细胞和基因疗法生产的关键工艺步骤,该行业的发展导致对病毒载体供应的需求增加。鉴于此,重要的是要有分析工具,以确保能够监测和优化病毒载体的生产过程,并确保病毒载体产品能够被量化和表征。对于病毒载体的需求以及生产成本,对在生产过程中实现良好的病毒载体滴度具有特别重要的作用。目前,物理病毒滴度测量通常由标准分析技术进行,例如对于慢病毒滴度,经常使用ELISA来评估p24或qPCR以用于测量病毒RNA,对于AAV,经常使用带有靶向ITR的引物的RT qPCR。然而,这些方法耗时较长,往往不准确,而且需要对介质进行采样。因此,这些方法只能提供病毒浓度的回溯测量。因此,需要新的方法来测量病毒滴度。

[0003] 此外,对于有效的病毒或病毒粒子生产方法,尤其是用于制造用于医疗应用的病毒粒子的方法,有利的是监测病毒核酸与包括一个或多个病毒结构分子的病毒的比率,以最大化所生产功能性病毒粒子的比率。非功能性的病毒粒子,如空粒子,通常被认为是生产过程中的废物,如果给患者服用,可能会引起问题,如不受欢迎的免疫反应。目前,定量病毒核酸与含有一个或多个病毒结构分子的病毒的比率最准确的方法是透射电子显微镜(transmission electron microscopy,TEM)。另一种选择是进行ELISA和qPCR实验,以分别计算病毒蛋白质和核酸的数量。然而,TEM、ELISA和qPCR这些方法只能进行回溯定量。因此,目前还没有方法可以实时量化病毒组分丰度,以便计算样本中功能性病毒粒子的比率。

[0004] 拉曼光谱是一种振动形式的光谱,已被证明在需要分子信息的过程分析技术(process analytical technology,PAT)应用中具有特殊的效用。这项技术是基于对被样本中存在的分子非弹性散射的光子的波数位移的检测(其中,此类光子的波数差异与将特定分子的振动状态从基态改变到第一激发态时从光子损失的能量有关,或与由光子从激发振动态到基态的去激发分子获得的能量有关)。与以前使用的PAT应用方法相比,该技术具有许多优势,最显著的是,与其他系统相比,相对较弱的信号由水产生,这有助于生物过程监测、细胞培养分析和溶液中的蛋白质分析。例如,拉曼光谱已被用于方差测试(variance testing)和确定原材料和细胞培养介质的身份;表征大分子产品;分析药物配方、批次之间的差异性、污染、介质的降解、细胞密度(包括存活细胞密度和总细胞密度)、蛋白质结构和蛋白质稳定性;用于聚合物和纤维分析;用于材料ID测试以及用于葡萄糖、谷氨酰胺、乳酸和氨的定量。Buckley和Ryder (2017)对拉曼光谱的适用性进行了回顾。

[0005] 常规拉曼光谱可能与微弱信号的产生有关,据报道,这会影响生物领域经常需要的用于灵敏定量分析的拉曼光谱的使用。事实上,据报道,使用常规拉曼光谱检测葡萄糖的浓度极限为0.6mM,检测苯丙氨酸的浓度极限为1.1mM(Buckley和Ryder,2017,应用光谱,71,1085-1116页)(估计分别相当于0.11mg/ml和0.18mg/ml)。鉴于此,在本领域中主要使用常规拉曼光谱来评估细胞培养介质和细胞生长,并且在需要更灵敏的测量的情况下使用其

他方法,例如,在检测低于报告的常规拉曼光谱检测极限的实体,例如病毒粒子的情况下。在这方面, Lee等人(2015)描述了使用表面增强拉曼光谱(Surface Enhanced Raman Spectroscopy, SERS)来检测HIV-1。SERS不同于常规拉曼光谱,并且提供了来自纳米尺度的附着在粗糙金属表面上的分子的增强信号,如银、铜或金。在Lee等人中,包括结合的抗gp120抗体片段的Au纳米点制备的氧化铟锡底物被用于对HIV-1病毒样粒子的特异性结合,以确保为HIV-1病毒样粒子检测产生增强的信号。然而, SERS可能会受限,包括要求预先准备具有适当免疫相互作用分子的底物,这限制了该技术的通用应用,以及与此相关的成本增加。此外,由于需要与感兴趣的实体结合, SERS在样本中总是具有侵入性的。

[0006] 因此,需要其他方法来表征和评估可能与敏感信号或者弱信号有关的实体,例如病毒,并原位评估这些实体。此外,还需要实时监测病毒组分丰度的方法,以便监测和最大化生产功能性病毒粒子的比率。

发明内容

[0007] 令人惊讶的是,与本领域的教导直接对比,发明人已经证明了常规拉曼光谱可以用于准确而灵敏地实时定量病毒滴度。这可以通过用光源直接照射病毒培养基以及如本文进一步描述的获取拉曼光谱数据来实现。因此,该方法消除了对病毒培养基加工的需要。发明人已经证明,根据本文描述的方法,由常规拉曼光谱确定的病毒滴度与离线滴度测量(例如,通过诸如RT-qPCR和p24 ELISA等离线分析测量)相当,并且根据本文描述的方法使用的常规拉曼光谱,可以提供另一种快速且可靠的方法来评估病毒滴度。如前所述,考虑到现有技术,这一发现是特别出乎意料的,在现有技术中,常规拉曼光谱与产生弱信号相关联,并且其中溶液中的低浓度病毒将被视为低于使用常规拉曼光谱的典型检测下限,例如,应用于批量溶液体系时,尤其是那些具有大的散射变化的体系,以及存在不需要的荧光背景信号的体系。考虑到现有技术需要处理病毒培养基以允许使用拉曼光谱,例如在使用拉曼光谱分析之前处理病毒培养基以浓缩病毒,或防止生物膜形成,这一发现也特别出乎意料。

[0008] 此外,发明人已经证明,可以使用常规拉曼光谱来实时监测病毒组分丰度,特别是实时量化病毒核酸丰度和病毒结构分子丰度。因此,拉曼光谱可以用来确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率,从而确定产生的功能性病毒粒子的比例。

[0009] 发明人已经确定了一系列光谱变量,这些光谱变量对于利用处理实时拉曼光谱数据的模型进行预测以实现实时病毒滴度和/或病毒核酸丰度和病毒结构分子丰度的测量是重要的。发明人已经建立了具有不同重要性阈值的逐渐增多的变量数量的分层范围,以在使用拉曼光谱时为准确预测病毒滴度和/或病毒核酸丰度和病毒结构分子丰度提供可变范围。发明人已经证明,拉曼光谱可以用来识别病毒生产过程的开始、生产阶段和结束。

[0010] 因此,本发明涉及使用拉曼光谱监测和/或评估样本中的病毒核酸丰度和病毒结构分子丰度。从另一个角度看,本发明涉及一种用于监测和/或评估样本中的病毒核酸丰度和病毒结构分子丰度的方法,包括使用多变量模型分析对包括病毒的样本进行拉曼光谱,以及确定病毒核酸丰度和病毒结构分子丰度的步骤。这种方法还可以包括对样品执行拉曼光谱的步骤,即获得光谱的步骤。

[0011] 本发明还特别涉及使用拉曼光谱监测和/或评估病毒生产过程的开始、生产阶段

和/或结束。发明人已经确认,如上所述,拉曼光谱可以用于确定病毒核酸丰度和病毒结构分子丰度,尤其是,发明人已经确定了为此需要评估的特定波数范围。这种峰的强度可以在本发明的方法中确定,其中这种峰的强度可以导致指纹的产生(其可以用多元模型评估),以确定病毒核酸丰度和病毒结构分子丰度。在本发明的优选实施例中,被监测和/或评估的病毒核酸丰度和病毒结构分子丰度是腺相关病毒的病毒核酸丰度和腺相关病毒的病毒结构分子丰度。

[0012] 本发明的一个优点是可以连续实时监测病毒核酸丰度和病毒结构分子丰度,以及病毒核酸与包括一个或多个病毒结构分子的病毒的比率。不需要处理来自病毒培养基的样本来产生对病毒核酸丰度、病毒结构分子丰度以及病毒核酸与包括一个或多个病毒结构分子的病毒的比率的估计值。如果需要,可以在原位进行测量。换句话说,可以直接在生长培养箱中的病毒培养基上进行测量。如果需要,也可以在非原位进行测量。换句话说,可以直接在病毒培养基中的等分试样的病毒培养基上进行测量,该病毒培养基的等分试样取自生长培养箱或从生长培养箱的主室分离得到的。无论是在原位还是非原位进行测量,都可以直接在病毒培养基上进行测量,而不需要对病毒培养基作进一步处理。这种类型的方法有时被描述为‘在线’分析或‘联机’分析。因此,可以产生更准确的病毒核酸丰度和病毒结构分子丰度的趋势,而不会有污染的风险。因此,本发明的方法比需要处理病毒培养基的常规离线方法更快和更简单。特别是,可以更准确地测量培养物的生产阶段,从而更准确地计时病毒生产的结束/收获阶段,允许在适当的时间点停止过程,潜在地降低生产过程的成本。

[0013] 因此,本发明提供了:一种使用拉曼光谱确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率的方法,该方法包括以下步骤:

[0014] (a) 提供样本并用光源照射所述样本;

[0015] (b) (i) 测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第一波数强度数据集,其中所述第一多个波数范围是预先选择的、且是所述样本中的病毒核酸的表征;

[0016] (ii) 对所述第一波数强度数据集执行第一组数学数据处理步骤;以及

[0017] (iii) 基于所述第一组数学数据处理步骤的输出,确定所述样本中的所述病毒核酸含量;

[0018] (c) (i) 测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第二波数强度数据集,其中所述第二多个波数范围是预先选择的、且是所述样本中的病毒的一个或多个病毒结构分子的表征;

[0019] (ii) 对所述第二波数强度数据集执行第二组数学数据处理步骤;以及

[0020] (iii) 基于所述第二组数学数据处理步骤的输出,确定所述样本中的包括一个或多个病毒结构分子的病毒的含量;以及

[0021] (d) 基于步骤(b) (iii) 和(c) (iii) 中确定的值,确定所述样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的所述比率。

[0022] 在上述定义的方法中,步骤(b) 和(c) 可以按照先(b) 然后(c) 的顺序或者先(c) 然后(b) 的顺序执行。如果步骤(b) (iii) 和(c) (iii) 中确认的值被确定,使得所述比率可以在步骤(d) 中被确定,则执行各步骤的确切顺序并不重要。此外,步骤(b) 和(c) 可以同时执行,因为数学数据处理可以允许同时处理第一波数和第二波数强度数据集,以便提供步骤(b)

(iii) 和 (c) (iii) 中确认的值。

[0023] 在上述定义的方法中,对第一和第二波数强度数据集执行第一和第二组数学数据处理步骤的步骤可以包括:

[0024] (i) 通过使用一种或多种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量 (standard normal variate, SNV) 法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合法 (piecewise polynomial fitting, PPF) 或自适应迭代重加权惩罚最小二乘 (adaptive iteratively reweighted Penalized Least Squares, airPLS) 法等,对信号强度数据进行预处理,可选地对波数信号强度数据进行归一化;

[0025] (ii) 通过将多变量回归算法诸如偏最小二乘 (partial least squares, PLS) 回归算法应用于所述波数信号强度数据来获得模型参数,任选地,其中所述PLS算法为非线性迭代最小二乘 (nonlinear iterative partial least squares, NIPALS) 回归算法或神经网络;以及

[0026] (iii) 利用对所述信号强度数据应用所述多变量回归算法而获得的模型参数,确定所述样本中的所述病毒核酸含量,并确定所述样本中的包括所述一个或多个病毒结构分子的病毒的含量。

[0027] 在上述定义的任何一种方法中,用于照射所述样本的光源可以是激光器,并且所述样本可以用波长为785nm的光照射。

[0028] 在上述定义的任何一种方法中,可以使用电荷耦合器件 (charge-coupled device, CCD) 来检测所述拉曼散射光。

[0029] 在上述定义的方法的一个实施例中,测量所述拉曼光谱中的所述第一多个波数范围以获得所述样本的所述第一波数强度数据集,所述第一多个波数范围可以包括如表1中所列出的波数范围1至12中的4个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围1至12中的6个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围1至12中的8个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围1至12中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围1至12中的所有12个,并且其中所述VIP ≥ 1.00 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒 (adeno-associated virus, AAV)。

[0030] 测量所述拉曼光谱中的所述第一多个波数范围替代地可以包括如表1中所列出的波数范围13至22中的4个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围13至22中的6个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围13至22中的8个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围13至22中的所有10个,并且其中所述VIP ≥ 1.25 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒 (AAV)。

[0031] 测量所述拉曼光谱中的所述第一多个波数范围替代地可以包括如表1中所列出的波数范围23至30中的4个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱

中的所述多个波数范围可以包括如表1中所列出的波数范围23至30中的6个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围23至30中的所有8个,并且其中所述VIP ≥ 1.50 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒(AAV)。

[0032] 在上述定义的方法的相同或其他实施例中,测量所述拉曼光谱中的所述第二多个波数范围以获得所述样本的所述第二波数强度数据集,所述第二多个波数范围可以包括如表2中所列出的波数范围1至20中的4个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围1至20中的6个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围1至20中的8个或更多个,并且所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围1至20中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围1至20中的12个或更多个、14个或更多个、16个或更多个或者18个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围1至20中的所有20个,并且其中所述VIP ≥ 1.00 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒(AAV)。

[0033] 所测量的所述拉曼光谱中的所述第二多个波数范围替代地可包括如表2中所列出的波数范围21至33中的4个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围21至33中的6个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围21至33中的8个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括表2中所列出的波数范围21至33中的10个或更多个、11个或更多个或者12个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围21至33中的所有13个,并且其中所述VIP ≥ 1.25 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒(AAV)。

[0034] 所测量的所述拉曼光谱中的所述第二多个波数范围替代地可包括如表2中所列出的波数范围34至40中的4个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括表2中所列出的波数范围34至40中的5个或6个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表2中所列出的波数范围34至40中的所有7个,并且其中所述VIP ≥ 1.50 。在这些方法中的任何一种中,病毒可以优选为腺相关病毒(AAV)。

[0035] 在上述定义的方法的另一个实施例中,测量所述拉曼光谱中的所述第一多个波数范围以获得所述样本的所述第一波数强度数据集,所述第一多个波数范围可以包括如表3中所列出的波数范围1至28中的5个或更多个,并且其中变量重要性投影(variable importance projection, VIP) ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括表3中所列出的波数范围1至28中的10个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围1至28中的15个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围1至28中的20个或更多个,并且其中所述VIP $\geq$

1.00;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围1至28中的25个或更多个,并且其中所述VIP ≥ 1.00 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围1至28中的所有28个,并且其中所述VIP ≥ 1.00 。在这些方法中的任何一种中,病毒可以优选为慢病毒。

[0036] 测量的所述拉曼光谱中的所述第一多个波数范围替代地可以包括如表3中所列出的波数范围29至59中的5个或更多个,并且其中所述变量重要性投影(VIP) ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括表3中所列出的波数范围29至59中的10个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围29至59中的15个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围29至59中的20个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围29至59中的25个或更多个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围29至59中的30个,并且其中所述VIP ≥ 1.25 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围29至59中的所有31个,并且其中所述VIP ≥ 1.25 。在这些方法中的任何一种中,病毒可以优选为慢病毒。

[0037] 所测量的所述拉曼光谱中的所述第一多个波数范围替代地可以包括如表3中所列出的波数范围60至81中的5个或更多个,并且其中所述变量重要性投影(VIP) ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括表3中所列出的波数范围60至81中的10个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围60至81中的15个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围60至81中的20个或更多个,并且其中所述VIP ≥ 1.50 ;或者所测量的所述拉曼光谱中的所述多个波数范围可以包括如表3中所列出的波数范围60至81中的所有22个,并且其中所述VIP ≥ 1.50 。在这些方法中的任何一种中,病毒可以优选为慢病毒。

[0038] 在上述定义的任何一种方法中,所述核酸可以包括病毒DNA基因组或病毒RNA基因组。

[0039] 在上述定义的任何一种方法中,所述一个或多个病毒结构分子可以包括一个或多个病毒蛋白,如一个或多个核蛋白和/或一个或多个壳粒,一个或多个病毒碳水化合物,一个或多个糖基化病毒分子,如糖基化病毒蛋白和/或一个或多个病毒脂质。

[0040] 在上述定义的任何一种方法中,所述比率可以提供功能性病毒滴度的量度。

[0041] 在上述定义的任何一种方法中,所述样本可以为病毒培养物。所述病毒培养物可以包括在生物反应器中。在任何这样的方法中,用光源照射所述病毒培养物和测量所述拉曼散射光总强度的步骤可以直接在病毒培养物的介质上(原位)执行。替代地,用光源照射所述病毒培养物和测量所述拉曼散射光总强度的步骤可以直接在取自所述病毒培养物中的培养基等分试样上(非原位)执行。

[0042] 上述定义的任何一种方法可以包括在第一个时间点确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率的第一步骤,以及在随后的时间点确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率的一个或多个进一步步骤,并且其中所述方法还包括

测量在各时间点之间所述样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的所述比率的变化,其中每个步骤由根据前述限定的任一项方法执行,优选地,其中每一步骤由相同的方法执行。在任何这样的方法中,在一段时间内反复确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率,以实时测量所述比率的变化。所述样本中的比率的变化可用于确定病毒生产过程的开始阶段、生产阶段和/或稳定阶段。任何此类方法可以用于确定病毒生产过程的最佳条件。任何此类方法可以用于评估病毒生产过程的下游过程。

[0043] 上述定义的任何方法都可以包括将由此获得的比率与通过替代方法从同一样本中获得的比率进行比较的步骤,可选地,所述替代方法是qPCR、RT-qPCR、ELISA或通过透射电子显微镜进行目视确定。

[0044] 本发明还提供一种使用拉曼光谱确定个体中病毒感染程度的方法,所述方法包括通过执行上述定义的任何一种方法来确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率,其中所述样本是先前从个体中获得的样本。所述样本可以是血液、唾液、痰、血浆、血清、脑脊液、尿液或粪便的样本。在任何此类方法中,将来自受试者的样本中的比率与先前针对个体中的感染获得的一个或多个比率测量值进行比较,以便提供个体感染阶段的预后。

[0045] 本发明还提供了一种使用拉曼光谱确定样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率的方法,所述方法包括以下步骤:

[0046] (a) (i) 提供所述样本的第一波数强度数据集,

[0047] 其中所述第一数据集是通过用光源照射所述样本并测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度而获得的,其中所述拉曼光谱中的所述第一多个波数范围已被选择作为所述样本中的病毒核酸的表征;

[0048] (ii) 对所述第一波数强度数据集执行第一组数学数据处理步骤;以及

[0049] (iii) 基于所述第一组数学数据处理步骤的输出,确定所述样本中的所述病毒核酸含量;

[0050] (b) (i) 提供所述样本的第二波数强度数据集,其中所述第二数据集是通过用光源照射所述样本并测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度而获得的,其中所述拉曼光谱中的所述第二多个波数范围已被选择作为所述样本中的所述病毒的一个或多个病毒结构分子的表征;

[0051] (ii) 对所述第二波数强度数据集执行第二组数学数据处理步骤;以及

[0052] (iii) 基于所述第二组数学数据处理步骤的输出,确定所述样本中的包含一个或多个病毒结构分子的病毒的含量;

[0053] (c) 基于步骤(a) (iii) 和(b) (iii) 中确定的值,确定所述样本中的病毒核酸与包括一个或多个病毒结构分子的病毒的所述比率。

[0054] 上述刚刚描述的方法可以根据本文描述和定义的任何一种方法中定义的步骤来执行。

[0055] 在上述刚刚描述的方法中,步骤(a) 和(b) 可以按照(a) 然后(b) 或者(b) 然后(a) 的顺序执行。如果步骤(a) (iii) 和(b) (iii) 中确定的值被确定,使得比率可以在步骤(c) 中被确定,则执行各步骤的确切顺序并不重要。此外,步骤(a) 和(b) 可以同时执行,因为数学数据处理可以允许同时处理第一波数和第二波数强度数据集,以便提供步骤(a) (iii) 和(b)

(iii)中定义的值。

[0056] 本发明还提供使用拉曼光谱来确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率。在任何此类用途中,所述比率可以基于在用光源照射所述样本后从所述样本中获得的拉曼散射光的强度的测量来确定,其中所述拉曼散射光的强度是从拉曼光谱中的第一个波数范围和拉曼光谱中的第二多个波数范围测量的,所述第一个波数范围是所述样本中的病毒核酸的表征,所述第二多个波数范围是所述样本中的病毒的一个或多个病毒结构分子的表征。

[0057] 在任何这种用途中,所述样本可以为病毒培养物,可选地,其中所述病毒培养物包括在生物反应器中。测量所述拉曼散射光的总强度的步骤可直接在病毒培养物的介质上(原位)执行。或者测量所述拉曼散射光的总强度的步骤可直接在取自所述病毒培养物的培养基的等分试样上(非原位)执行。

[0058] 在任何这种用途中,所述样本中的比率是在第一个时间点和在一个或多个稍后时间点确定的,其中计算在各时间点之间所述样本中比率的变化。在任何这种用途中,所述样本中的所述比率被反复定量,以实时提供对所述比率变化的测量。在任何这种用途中,所述样本中的病毒滴度可通过执行本文描述和定义的任一方法来定量。

[0059] 在上述定义的任何方法或用途中,所述样本中的病毒可以不是HIV-1或HIV-1病毒样粒子(HIV-1VLP)。

[0060] 在上述定义的任何方法或用途中,所述拉曼光谱可以不是表面增强拉曼光谱。

[0061] 本发明还提供了一种构建多变量数据处理模型的方法,所述多变量数据处理模型能够从针对样本获得的拉曼光谱波数强度数据集来确定样本中包括一个或多个病毒结构分子的病毒的含量,所述方法包括:

[0062] (a) 提供样本并用光源照射所述样本;

[0063] (b) 测量多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的波数强度数据集,其中所述多个波数范围是预先选择的且是所述样本中的所述病毒的一个或多个病毒结构分子的表征;

[0064] (c) 通过使用一种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量(SNV)法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合(PPF)法或自适应迭代重加权惩罚最小二乘(airPLS)法等,对信号强度数据进行预处理,来获得归一化的波数信号强度数据;

[0065] (d) 通过将多变量回归算法诸如偏最小二乘(PLS)回归算法应用于所述预处理的信号强度数据来获得模型参数,可选地,其中所述PLS算法为非线性迭代偏最小二乘(NIPALS)回归算法或神经网络,其中执行校准,其中,将所述经预处理的信号强度数据与使用诸如qPCR、RT-qPCR、ELISA等非拉曼光谱方法或通过由透射电子显微镜进行目测确定在相同样本条件下获得的病毒滴度数据进行比较;

[0066] (e) 利用经预处理数据获得的模型参数来推断响应值;以及

[0067] (f) 执行变量选择,可选地,变量重要性投影(VIP),和确定拉曼光谱变量;以及

[0068] (g) 可选地,通过重新应用步骤(d)-(f)进行进一步的一轮或多轮建模,其中不重要的变量被除去;以及

[0069] 其中,使用源自多变量数据处理模型的针对所识别的拉曼光谱变量获得的模型参

数,来确定样本中包括一个或多个病毒结构分子的病毒的含量。

[0070] 本发明还提供了一种构建一个或多个多变量数据处理模型的方法,所述一个或多个多变量数据处理模型能够从针对样本获得的拉曼光谱波数强度数据集来确定样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率,所述方法包括:

[0071] (a) 提供样本并用光源照射所述样本;

[0072] (b) (i) 测量第一多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第一波数强度数据集,其中所述第一多个波数范围是预先选择的且是所述样本中病毒核酸的表征;

[0073] (ii) 测量第二多个波数范围中的每个波数范围内的拉曼散射光的总强度,以获得所述样本的第二波数强度数据集,其中所述第一多个波数范围是预先选择的且是所述样本中病毒的一个或多个病毒结构分子的表征;

[0074] (c) 通过使用一种预处理分析方法,例如一阶导数法、二阶导数法、标准正态变量(SNV)法、多项式拟合法、多多项式拟合法、软化子法、分段多项式拟合(PPF)法或自适应迭代重加权惩罚最小二乘(airPLS)法等,对信号强度数据进行预处理,来获得第一和第二波数强度数据集的归一化的波数信号强度数据;

[0075] (d) 通过将多变量回归算法诸如偏最小二乘(PLS)回归算法应用于所述预处理的信号强度数据集的每一个来获得待应用于第一和第二波数强度数据集的模型参数,可选地,其中所述PLS算法为非线性迭代偏最小二乘(NIPALS)回归算法或神经网络,其中执行校准,其中,将经预处理的信号强度数据与使用诸如qPCR、RT-qPCR、ELISA等非拉曼光谱方法或通过由透射电子显微镜进行目测确定在相同样本条件下获得的病毒滴度数据进行比较;

[0076] (e) 利用从经预处理数据集的每一个获得的模型参数来推断响应值;以及

[0077] (f) 执行变量选择,可选地,变量重要性投影(VIP)和确定拉曼光谱变量;以及

[0078] (g) 可选地,通过重新应用步骤(d)-(f)对任一数据集进行进一步的一轮或多轮建模,其中不重要的变量被移除;

[0079] 其中,使用源自多变量数据处理模型的针对所识别的拉曼光谱变量获得的模型参数,来确定样本中病毒核算与包括一个或多个病毒结构分子的病毒的比率。

附图说明

[0080] 图1:显示红外吸收/发射的量子能量转变的雅布隆斯基(Jablonski)图。该图显示了具有斯托克斯跃迁和反斯托克斯跃迁的瑞利(弹性散射)和拉曼(非弹性散射)。

[0081] 图2:来自VV拉曼项目慢病毒生物反应器运行的示例性预处理拉曼光谱。插图是 1000cm^{-1} 区域的放大图。在CCD读出时间约15分钟后,采集光谱10秒,累计75次,总集成时间约12分30秒。

[0082] 图3:条形图显示了在病毒载体拉曼项目中,用于生物反应器转染的代表性qPCR慢病毒滴度结果,cp数=复制数。

[0083] 图4:显示了在经过10倍交叉验证和20次蒙特卡洛重复之后,预测的均方误差(MSEPCV)如何随PLS潜变量/分量数的变化(在光谱变量选择之前)而变化的图。

[0084] 图5(A):从初始10分量的PLS-R模型计算的变量重要性投影(VIP)图。圆圈表示VIP>1.5的光谱变量,该信息用于确定图5(B)所示的范围。图5(C):VIP>1.5的光谱变量表,按重

要性顺序排列。

[0085] 图6:显示了交叉验证后预测的均方误差随潜变量或分量数变化的图。用VIP>1.5进行保守光谱变量约简后得到。

[0086] 图7:显示了在光谱变量约简 (VIP>1.5) 后,来自慢病毒运行1个生物反应器4的拉曼复制数/mL PLS-R预测 (10LV) 与离线qPCR数据的图。T表示“转染”,NT表示“未转染”。

[0087] 图8:显示了在光谱变量约简 (VIP>1.5) 后,来自慢病毒运行2个生物反应器1-4的拉曼复制数PLS-R预测 (10LV) 与离线qPCR数据的图。

[0088] 图9:显示了在光谱变量约简后,来自慢病毒运行3个生物反应器1-4的拉曼复制数PLS-R预测 (10LV) 与离线qPCR数据的图。

[0089] 图10:RT-qPCR和p24 ELISA结果的比较。P24 ELISA试验只用于在少数离线样本上获得慢病毒滴度。

[0090] 图11:病毒滴度的实时拉曼衍生模型的应用实例,以识别病毒生产过程的开始、生产阶段和结束。根据模型(黑实线),还实时计算了一组3个指标。与病毒滴度的模型估计值一起,指示曲线的形状说明了病毒产生的各个阶段(如覆盖至图上的文字所述)。使用RT-qPCR和/或ELISA法回溯计算的物理滴度叠加在图(方块和虚线)上,以显示实时数据和回溯离线数据之间的总体一致性。从拉曼光谱引入额外的峰可以提高拉曼模型的质量,但该图是使用识别病毒生产的主要阶段所需的最小区域数生成的。

[0091] 图12:显示了定量病毒滴度的公式的概要示意图。该公式应用于模型参数(在本案中为回归系数),该模型参数由应用于归一化拉曼信号强度数据的多变量回归算法获得。

[0092] 图13:显示了 R^2 (1-残差平方和/总平方和)随波数范围数的变化,以说明提供慢病毒滴度估计值所需的最小波数范围数的图。

[0093] 图14:来自AAV拉曼项目生物反应器运行的预处理拉曼光谱实例。插图是1000-1200 cm^{-1} 区域的放大图。在CCD读出时间约15分钟后,采集光谱10秒,累计75次,总集成时间约12分30秒。

[0094] 图15:条形图显示了在AAV拉曼项目中,生物反应器转染的有代表性的qPCR AAV病毒滴度结果。

[0095] 图16:显示了在经过10倍交叉验证和20次蒙特卡洛重复之后,预测的均方误差(MSEPCV)如何随PLS潜变量/分量数的变化(在光谱变量选择之前)而变化的图。

[0096] 图17(A):从初始15分量的PLS-R模型计算的变量重要性投影(VIP)图。圆圈表示VIP ≥ 1.0 的光谱变量,该信息用于确定图17(B)所示的范围。图17(C):VIP ≥ 1.0 的光谱变量表,按重要性顺序排列。

[0097] 图18:显示了交叉验证后预测的均方误差随潜变量或分量数变化的图。用VIP ≥ 1.0 进行光谱变量约简后得到。

[0098] 图19:显示了在光谱变量约简后,来自运行4个生物反应器1-4的拉曼复制数PLS-R预测 (9LV) 与离线qPCR数据的图。

[0099] 图20:显示了 R^2 (1-残差平方和/总平方和)随波数范围数的变化,以说明提供AAV病毒滴度估计值所需的最小波数范围数的图。

[0100] 图21:来自AAV拉曼项目生物反应器运行的预处理拉曼光谱实例。插图是1000-1200 cm^{-1} 区域的放大图。在CCD读出时间约15分钟后,采集光谱10秒,累计75次,总集成时间

约12分30秒。

[0101] 图22:条形图显示了在AAV拉曼项目中,生物反应器转染的有代表性的qPCR AAV病毒滴度结果。

[0102] 图23:条形图显示了在AAV拉曼项目中,生物反应器转染的有代表性的ELISA AAV病毒滴度结果。

[0103] 图24:显示了在经过10倍交叉验证和20次蒙特卡洛重复之后,每毫升RT-qPCR复制数的预测的均方误差(MSEPCV)如何随PLS潜变量/分量数的变化(在光谱变量选择之前)而变化的图。

[0104] 图25:显示了在经过10倍交叉验证和20次蒙特卡洛重复之后,每毫升RT-qPCR复制数的预测的均方误差(MSEPCV)如何随PLS潜变量/分量数的变化(在光谱变量选择之前)而变化的图。

[0105] 图26:(A)对于每毫升复制数(基于RT-qPCR)的PLS-R模型,从初始15分量的PLS-R模型计算的变量重要性投影(VIP)图。圆圈表示 $VIP \geq 1.0$ 的光谱变量,该信息用于确定图26(B)所示的范围。图26(C): $VIP \geq 1.0$ 的光谱变量表,按重要性顺序排列。

[0106] 图27:(A)对于每毫升总粒子数(基于ELISA)的PLS-R模型,从初始14分量的PLS-R模型计算的变量重要性投影(VIP)图。圆圈表示 $VIP \geq 1.0$ 的光谱变量,该信息用于确定图27(B)所示的范围。图27(C): $VIP \geq 1.0$ 的光谱变量表,按重要性顺序排列。

[0107] 图28:(A)对于每毫升总粒子数(基于ELISA)的PLS-R模型,从初始14分量的PLS-R模型计算的变量重要性投影(VIP)图。圆圈表示 $VIP \geq 1.0$ 的光谱变量,该信息用于确定图27(B)所示的范围。图27(C): $VIP \geq 1.0$ 的光谱变量表,按重要性顺序排列。

[0108] 图29:显示了交叉验证后每毫升ELISA总病毒粒子数的预测的均方误差随潜变量或分量数变化的图。用 $VIP \geq 1.0$ 进行光谱变量约简后得到。

[0109] 图30:显示了在光谱变量约简后,来自生物反应器1-4的拉曼复制数PLS-R预测值(10LV)以及离线qPCR数据的图。

[0110] 图31:显示了在光谱变量约简后,来自生物反应器5-8的拉曼复制数PLS-R预测(10LV)以及离线qPCR数据的图。

[0111] 图32:显示了在光谱变量约简后,来自生物反应器1-4的拉曼总粒子数PLS-R预测(10LV)以及离线ELISA数据的图。

[0112] 图33:显示了在光谱变量约简后,来自生物反应器5-8的拉曼总粒子数PLS-R预测(10LV)以及离线ELISA数据的图。

[0113] 图34:显示了从拉曼PLS-R模型预测中计算出的基因组复制数(RT-qPCR)和总粒子数(ELISA)的空满比(%)的图。对于生物反应器2-4,从转染后24小时开始显示。

[0114] 图35:显示了从拉曼PLS-R模型预测中计算出的基因组复制数(RT-qPCR)和总粒子数(ELISA)的空满比(%)的图。对于生物反应器5-7,从转染后24小时开始显示。

具体实施方式

[0115] 应当理解,所公开的方法的不同应用可以针对本领域中的特定需要而定制。还应当理解,本文使用的术语仅用于描述本发明的特定实施例,而不是意在限制。

[0116] 此外,如在本说明书和所附权利要求书中所使用的,除非内容另有明确说明,否则

单数形式“一”、“一个”和“所述”包括复数指代。

[0117] 在此引用的所有出版物、专利和专利申请,无论是上文还是下文,在此都通过引用完整地并入本文。

[0118] 利用拉曼光谱进行病毒分析

[0119] 本发明包括使用拉曼光谱来监测和评估病毒滴度和/或病毒组分丰度。本发明包括使用拉曼光谱来监测和评估病毒核酸丰度和病毒结构分子丰度。本发明包括使用拉曼光谱来监测和评估样本中病毒核酸与包括一个或多个病毒结构分子的病毒的比率。

[0120] 本文定义的“病毒滴度”是指在给定的体积中存在的病毒量。利用本发明可以评估任何类型的病毒滴度,例如,可以评估物理病毒滴度、功能性病毒滴度(也称为感染性病毒滴度)或转导病毒滴度。在特定实施方案中,可以评估物理病毒滴度。物理病毒滴度是对样本(例如病毒培养基)中病毒粒子浓度的测量,通常以病毒蛋白(如p24)或病毒核酸的存在为基础。物理滴度可表示为每毫升病毒粒子(VP/mL)、每毫升病毒基因组(vg/mL)、每毫升病毒复制数或每毫升RNA复制数,用于测量物理滴度的现有技术方法包括p24的ELISA(例如Lenti-X p24快速滴定试剂盒(Takara)或慢病毒相关p24的ELISA试剂盒(Cell Biolabs, Inc.))、qPCR或ddPCR(例如AAV实时PCR滴定试剂盒(Takara)或Adeno X qPCR滴定试剂盒(Takara))。物理滴度测量并不总是区分空的或有缺陷的病毒粒子和能够感染细胞的粒子。因此,物理病毒滴度可以与功能性滴度或感染性滴度区分开来,感染性滴度确定所产生的粒子有多少能感染细胞,而转导病毒滴度确定有多少功能性病毒粒子包括感兴趣的基因(例如,对于病毒载体的生产,转导病毒滴度可能是相关的)。因此,物理滴度的确定并不等同于功能性滴度的确定,除非样品中的所有粒子都是功能性的。事实上,功能性滴度通常比物理滴度低100到1000倍。

[0121] 或者,如上所述,可以利用本发明测量或评估功能性或感染性滴度,其中功能性或感染性滴度是对特定体积中存在的能够感染靶细胞的病毒粒子的量的测量。功能性滴度可表示为每毫升斑块形成单位(pfu/mL)或每毫升感染单位(ifu/mL)。可用于测量功能性或感染性滴度的离线分析方法包括斑块分析、病灶形成分析、终点稀释分析或流式细胞术。如上所述,转导滴度是对特定体积中存在的能够感染靶细胞并包括感兴趣基因的病毒粒子数量的测量。转导滴度可表示为转导单位/mL,并可使用上述用于评估功能性滴度的分析方法,以及任何可确定感兴趣基因存在的已知方法,例如PCR,来进行评估。本领域技术人员将认识到,功能性滴度或转导滴度可以通过按比例缩小针对物理滴度所获得的任何值来确定。如上所述,物理滴度和功能性滴度或转导滴度之间的倍数差异在本领域中是众所周知的。因此,在本发明的一个方面,可以通过本发明的方法(例如,通过按比例缩小针对物理滴度所获得的值)间接确定功能性滴度或转导滴度。因此,本发明的方法可以包括额外步骤,即按比例缩小物理滴度的确定以确定功能性滴度或转导滴度。

[0122] 本发明的方法可用于监测和评估样本(例如病毒培养基)中的病毒核酸丰度和病毒结构分子丰度。本发明的方法可用于确定样本(例如病毒培养基)中病毒核酸与包括一个或多个病毒结构分子的病毒的比率。这一比率可以用来确定样本中同时含有核酸组分和结构组分的病毒粒子的比率。这一比率可以用作功能性滴度的估计值。因此,本发明的方法可用于实时估计功能性滴度,而先前已知的用于估计功能性滴度的方法是回溯性的和离线的。

[0123] 技术人员将认识到,本发明的方法能够确定任何血清类型的任何病毒的病毒滴度和/或病毒组分丰度,例如,逆转录病毒,如慢病毒(例如,HIV-1和HIV-2)和 γ 逆转录病毒;腺病毒和腺相关病毒(例如,AAV1-11,特别是AAV1、AAV2、AAV5和AAV8,以及自身互补的AAV)。因此,如本文进一步描述和定义的,本发明的方法能够应用于哺乳动物病毒。如本文进一步描述和定义的,本发明的方法还能够应用于非哺乳动物病毒,包括植物病毒,如烟草花叶病毒、藻类病毒、酵母病毒和昆虫病毒,包括杆状病毒。虽然通常情况下,在本发明中可以评估单一病毒类型的病毒滴度和/或病毒组分丰度,但是本发明的方法将能够评估混合病毒样本(例如包括两种或更多病毒类型的样本)的滴度和/或病毒组分丰度。

[0124] 在这方面,根据本发明产生的拉曼光谱可以直接检测病毒或病毒组分,间接检测病毒或病毒组分,或者同时直接和间接检测病毒或病毒组分。因此,在拉曼光谱上产生和显示的波数峰可指示与病毒或病毒组分相关的若干不同化合物或分子中的任何一种。本领域技术人员将认识到,对于本发明的实施,不需要准确地识别所识别的每个波数峰与哪些化合物/分子相关。在特定条件下(例如培养条件、病毒、生产者细胞系和待测量的病毒滴度类型等)光谱反而充当指纹,其中病毒滴度和/或病毒组分丰度可以通过利用特定的多变量模型(通常是在相同或相似的条件产生的模型),分析本文所披露的每个波数范围的信号强度来确定。多变量模型将在本文中进行更详细的描述。因此,对于特定的拉曼光谱,在特定的波数下获得的峰可以对应于病毒(例如衣壳蛋白等)的分子/化合物,或者可以对应于非病毒(例如培养物中的代谢物,如由病毒生产细胞产生的)的分子/化合物。以这种方式,本发明中使用的拉曼光谱可以检测与病毒滴度和/或病毒组分丰度间接相关的化合物,或者代替检测与病毒滴度和/或病毒组分丰度直接相关的化合物。

[0125] 可以预期到,虽然每个波数范围下的信号强度可能不同,并且可能不同地对应于在不同条件下(例如,用于生产不同的病毒,使用不同的生产者细胞系或使用不同的培养基或生物反应器)获得的拉曼光谱中的病毒滴度和/或病毒组分丰度,但可以评估的波数范围(例如,已经被确定为与病毒滴度和/或病毒组分丰度相关)将保持相同。此外,应当认识到,对于任何给定的条件集,用户将能够基于本文描述的和领域中已知的原理来创建多变量模型,并在分析来自本文描述的不同波数范围的不同信号强度数据时应用该模型,所述数据随后使用相同的条件集来生成。这意味着用户可以使用本文描述的波数范围,通过拉曼光谱计算病毒滴度和/或病毒组分丰度,其中数据在应用不同条件的系统中生成。

[0126] 本文使用的对波数范围的慢病毒评估,已经识别了在病毒滴度和/或病毒组分丰度评估中重要的范围。由于已知慢病毒就其化学成分而言,是一种特别复杂的病毒,因此拥有慢病毒中存在的部分组分的较简单病毒的化学组分可能会产生落在所识别的波数范围的一部分(如果任何一个或多个波数范围直接检测病毒)内的相关信号,例如至少5个所识别的波数范围。因此,本文提供的波数范围的子集可用于评估除慢病毒以外的病毒的滴度。进一步地,如果识别到的与病毒滴度和/或病毒组分丰度相关的一个或多个波数范围间接检测病毒,那么,由于任何病毒转染都可能导致培养物中类似的代谢组学变化,这样的波数范围可能有助于评估任何系统中的病毒滴度和/或病毒组分丰度。

[0127] 在本发明的特定实施方案中,监测和评估慢病毒滴度和/或慢病毒组分丰度。一方面,病毒可能不是HIV-1或HIV-1病毒样粒子。

[0128] “病毒”通常是一种只能在另一种生物的活细胞内复制的小的感染性病原体(通常

比细菌小)。病毒可具有基于RNA或DNA的基因组。本文所用的“病毒”是指任何病毒:修饰病毒、病毒粒子、病毒衍生粒子或病毒载体。因此,尽管可以根据本发明评估任何野生型病毒的病毒滴度和/或病毒组分丰度,但是可以理解,本发明的用途可以特别扩展到对突变体或修饰病毒(即,与野生型或自然存在的病毒相比包括一个或多个核酸替换、插入、缺失或易位,或缺少编码病毒蛋白的大部分遗传物质)或病毒载体的病毒滴度和/或病毒组分丰度的评估。突变体或修饰病毒或病毒粒子经常用于生产疫苗,预计本发明的方法在监测和评估用于疫苗的功能性病毒或病毒粒子的生产效率方面将特别有效。

[0129] 虽然病毒载体可以野生型病毒为基础,但与野生型病毒相比,它们通常是经过修饰的,并且通常用于将遗传物质引入靶细胞(例如,治疗用的基因)。因此,病毒载体具有特殊的用途,例如用于基因治疗、细胞治疗或其他分子应用,它们的生产对基因治疗和细胞治疗行业具有巨大的重要性。将会很好地理解的是,例如,可以进行修饰以提高用于基因和/或细胞治疗的病毒载体的安全性,或者例如改善可被载体携带的基因的大小。为创建病毒载体而进行的修饰可包括删除对复制至关重要的病毒基因组的一部分,从而产生病毒载体,其能够感染细胞但需要辅助病毒的存在来提供产生新病毒粒子所需的缺失蛋白质。其他修饰可包括降低病毒载体对其靶细胞的毒性和/或提高病毒稳定性的修饰,例如减少基因组重排。

[0130] 病毒载体通常可以在包装细胞系(如HEK293细胞)中产生,方法是用编码病毒蛋白并携带所需遗传物质的一个或多个质粒转导包装细胞系。例如,对于慢病毒生产,可以用一个或多个质粒转导HEK293细胞,例如3个或4个编码病毒蛋白(如衣壳和逆转录酶)并携带待由载体传递的遗传物质的质粒。这被转录以产生单链RNA病毒基因组,并以psi序列的存在为标志,这确保了基因组随后被包装成病毒粒子。因此,特别地,慢病毒载体可以通过在生产者细胞系中转化和表达3个(对于第二代系统)或4个(对于第三代系统)质粒来产生。用于生产病毒载体的质粒是可以商业获得的,例如Lenti-Pac和AAV Prime (GeneCopoeia)。

[0131] 特别地,可以通过本发明的方法监测或评估通过包装细胞系产生的病毒载体的滴度和/或病毒组分丰度。如前所述,在基因治疗和细胞治疗领域中,能够以灵敏的方式测量产生的滴度和/或病毒组分丰度特别重要,例如,这样可以准确地监测和管理生产过程,例如生产者细胞系的生产,并且可以实时计算功能载体的比例。病毒不需要是全功能性的或野生型的,就可以通过本发明的方法进行监测或评估。

[0132] 本文中的“病毒组分”被认为是病毒、病毒粒子或病毒载体的任何部分。

[0133] 病毒粒子或“病毒粒子”通常被理解为由以下组成:(i) 病毒的遗传物质,即编码病毒赖以活动的蛋白质结构的DNA或RNA分子;(ii) 由壳粒形成的内部蛋白质外壳,称为衣壳,它围绕和保护病毒的遗传物质;以及在某些情况下,(iii) 脂质的外部包膜,其可包括包膜蛋白。

[0134] 病毒组分包括病毒核酸和病毒结构组分(或病毒结构分子)。本文的病毒核酸被认为包括病毒RNA、病毒DNA、病毒DNA基因组和病毒RNA基因组。病毒核酸被包装在病毒粒子内。

[0135] 本文使用的病毒结构组分或病毒结构分子应被理解为对病毒结构有贡献的任何分子。这里使用的病毒结构组分或病毒结构分子可以排除病毒的遗传物质,即编码病毒赖以活动的蛋白质结构的DNA或RNA分子。

[0136] 病毒结构组分或病毒结构分子在本文中被认为包括病毒蛋白质,如核蛋白、衣壳蛋白、原体、衣壳亚单位、衣壳单体、衣壳单体的组合、壳粒、六位体、五位体、病毒外壳蛋白(viral coat protein,VCP)、病毒外表面糖蛋白、病毒跨膜蛋白、病毒功能所必需的蛋白、病毒粒子或病毒载体、病毒碳水化合物、糖基化病毒分子(如糖基化病毒蛋白)和/或病毒脂质(包括病毒磷脂),或其组合。

[0137] 如上所述,本发明的方法可以“监测或评估”病毒滴度和/或病毒组分丰度。因此,本发明的方法能够确定病毒滴度和/或病毒组分丰度,例如样本中存在的病毒核酸和病毒结构分子的水平、量或浓度。特别地,该方法因此可以确定病毒核酸和病毒结构分子的水平、量或浓度是否随着时间的推移而彼此增加或平稳(例如,通过在不同时间点分析样本),或者与不同样本(例如,在相同或同等的时间点分析)相比是否变化(例如,增加、减少或相等)。以此方式,本发明的方法可用于例如评估病毒的生产方法的效率,例如其中病毒核酸与包括一个或多个病毒结构分子的病毒的比率的检测或确定可指示有效的方法或次优的生产方法),或者可用于例如通过与在其他改进的生产方法期间测量的病毒滴度和/或病毒组分丰度的比较来确定所述病毒生产方法中特定因素的重要性(对于相同或不同的病毒)。期望通过本文描述的方法检测的物理滴度值在 1×10^{10} 到 1×10^{11} 粒子/mL的范围内。预期通过本文描述的方法检测的感染性滴度值在 1×10^8 至 1×10^9 粒子/mL的范围内。因此,可确定对导致所测病毒滴度和/或病毒组分丰度差异的因子的修饰对于生产方法是重要的(例如,对导致所测病毒滴度至少相差5、10、20、30、40或50%的因子的修饰)。这一因素可以包括培养温度、所用的培养基、培养基中所用的葡萄糖或氨基酸的百分比、所用搅拌的存在、不存在或量,或所用的培养瓶或体积等。

[0138] 因此,本发明的方法可用于确定病毒生产的最佳条件,包括对可用于培养可产生病毒的生产者细胞的不同系统的评估,例如摇瓶、量子系统(Terumo)、Ambr系统,例如Ambr15或250(TAP生物系统)。

[0139] 本发明的方法还可用于评估病毒生产过程下游的任何过程,例如确定任一这样的过程是否影响了病毒滴度和/或病毒组分丰度。特别地,本发明的方法可用于评估可采用的纯化方法,例如确定这种纯化方法是否对滴度和/或病毒组分丰度有任何影响,例如,与纯化前存在于样品中的病毒核酸与包括一个或多个病毒结构分子的病毒的比率相比,病毒核酸与包括一个或多个病毒结构分子的病毒的比率在这样的纯化后是否增加、降低或保持相等。本发明的方法还可用于评估病毒的大规模制造,例如用于疫苗或病毒载体的病毒粒子的大规模制造,这对于用于基因治疗的病毒载体的制造可能特别重要。

[0140] 本文使用的病毒滴度和/或病毒组分丰度的增加可以是与正被比较的测量相比较,病毒滴度和/或病毒组分丰度具有多于5、10、20、30、40、50、60、70、80或90%的增加,而本文所使用的病毒滴度和/或病毒组分丰度的减少可以是与正被比较的测量相比较,病毒滴度和/或病毒组分丰度的具有减少或多于5、10、20、30、40、50、60、70、80或90%的减少。相当的病毒滴度和/或病毒组分丰度可在与测量结果进行比较的病毒滴度的5%以内。

[0141] 在这方面,应当认识到,出于某些目的,可能希望在实施方法之前以及在方法之后和/或方法期间评估病毒滴度和/或病毒组分丰度,以便确定病毒滴度和/或病毒组分丰度是否发生了任何改变或变化。本发明的方法还可进一步包括例如将病毒滴度和/或病毒组分丰度例如与不同样本内(在相同或不同的时间点)的或在不同时间点的同一样本内的病

病毒滴度和/或病毒组分丰度进行比较的步骤。

[0142] 在本发明的另一实施例中,该方法可用于确定受试者中病毒感染的程度,例如确定感染是否被成功治疗或减少。在这种方法中,可能希望比较样本(例如来自不同时间点的受试者的相同类型的样本)中的病毒滴度和/或病毒组分丰度,以确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率是否随时间增加、减少或保持相等。替代地或另外地,可能希望将来自个体的样本中的病毒滴度和/或病毒组分丰度与先前针对一种情况获得的病毒滴度和/或病毒组分丰度测量进行比较,该测量例如可以指示感染的阶段和/或预后。

[0143] 或者,本发明的方法可以不确定样本中病毒组分丰度的实际量、水平或浓度,但可以确定该量、水平或浓度是否高于或低于可接受的阈值,例如,对于生产方法,该阈值可以确定样本中是否存在可接受的功能性病毒粒子水平。如上所述,本发明的方法可以确定病毒核酸与包括一个或多个病毒结构分子的病毒的比率是否已经增加、减少或与先前检测的样本的比率相当,因此可以理解,对于特定应用,可能不需要确定实际的病毒滴度和/或病毒组分丰度(例如,存在的病毒的量或浓度)。

[0144] 本发明包括使用拉曼光谱来监测和/或评估样本中的病毒滴度和/或病毒组分丰度,从而可以识别病毒生产过程的开始、生产阶段和结束中的任何一个。应当认识到,在生产的不同阶段,样品中将存在不同数量或浓度的病毒和/或代谢物。例如,在生产开始时,病毒滴度,特别是物理病毒滴度可以在 $0-10^5$ 的范围内,在活跃的病毒生产期间,病毒滴度,特别是物理病毒滴度,可以在 10^5-10^9 的范围内,并且在病毒生产结束时,病毒滴度,特别是物理病毒滴度,可以在 10^9-10^{12} 的范围内。然而,通常,在例如特定的包装细胞系中,随着时间的推移对病毒滴度和/或病毒组分丰度的监测可以识别特定病毒的生产的不同阶段,其中增加的量或峰量可能与该过程的生产阶段相关,早期的低量可能与该过程的开始相关,而之后数量的平稳可能与该过程的结束相关。如前所述,该信息可用于特定过程,以确保在生产已经平稳、减少(例如,与最高生产点相比至少50%)或终止之后不保持培养。

[0145] 本发明的方法还可用于支持适应性制造,并进一步增加系统中的病毒滴度和/或病毒组分丰度生产。在这方面,可以比较在不同条件下和使用不同系统获得的病毒核酸与包括一个或多个病毒结构分子的病毒的比率,以确定产生病毒的最佳条件。

[0146] 本文使用的术语“样本”是指含有病毒的任何样本(例如,包括病毒载体的任何样本)。“样本”优选是病毒培养液,也就是培养病毒的液体。因此,病毒培养基可被直接照射以获得用于本文进一步描述的本方法的拉曼光谱数据。在本发明的优选实施方案中,样品来自工业病毒生产方法或病毒载体制造方法。特别地,在本发明中,样品中的病毒滴度可以通过拉曼光谱实时、原位进行测量,或在样品上非原位地进行。

[0147] “原位”指的是,在能够产生病毒粒子的培养物中获得拉曼散射光强度的测量取自产生病毒粒子的初级培养环境,而不是取自从初级培养环境中提取的样本。因此,通过“原位”测量,对液体处理步骤没有要求。因此,对于本发明的特定应用,可能不需要将样品从其环境中移除,并且可以优选对样品进行原位测量。样品的原位测量可允许定期评估样品中的病毒滴度和/或病毒组分丰度,而不需要实际的取样步骤,其中,样品(例如病毒培养基)的一部分,被从例如病毒生长培养箱的初级培养环境中移除。在这方面,病毒滴度和/或病毒组分丰度评估可以实时准确和灵敏地测量,而不需要可能引入成本和误差的额外步骤。如下面更详细地讨论的,例如,拉曼光谱提供了可以放置在样品内部或外部的探针,从而允

许在需要的地方进行原位测量。原位测量特别适用于“在线”过程分析技术。

[0148] 替代地,本发明的方法可以在样品上非原位执行。“非原位”指的是,在能够产生病毒粒子的培养物中获得拉曼散射光强度的测量是从产生病毒粒子的初级培养环境(例如病毒生长培养箱)中提取的样本(例如病毒培养基)的等分试样中提取的,并直接进行分析。这种非原位测量适用于“在线”或回溯过程分析技术。无论是在原地还是非原位进行测量,都可以直接在样品上进行测量,例如直接在病毒培养基上进行测量,而不需要对样品进行进一步的处理。

[0149] 本发明方法中使用的样本的来源可以是在其中产生病毒的细胞培养物。因此,样本可以是培养基(例如DMEM、MEM或SFII,可选地包括血清、L-谷氨酰胺和/或其他组分)中的一种,其可以另外包括包装细胞(例如HEK293细胞),例如如果是在病毒生产过程中采集的;或者样本可以是用于医疗用途的病毒样本,例如需要质量测试,例如在营销、销售或使用之前。样本还可以是疑似感染病毒的受试者(例如人类或哺乳动物受试者)的样本,例如血液、唾液、痰、血浆、血清、脑脊液、尿液或粪便样本。其他样本来源包括开放水域或公共供水。

[0150] 拉曼光谱

[0151] 拉曼光谱测量被样品散射的单色光的波数的变化,以提供关于样品的化学成分、物理状态和环境的信息。这是可能的,因为入射光光子与组成样品的分子中存在的振动模式相互作用的方式。这些模式在一组给定的物理条件下具有特定的振动频率和散射强度,这使得可以定量所感兴趣的给定分析物的量。与红外吸收光谱不同(在红外吸收光谱中,测量的是来自宽带光源的不同能量的光的吸收),在拉曼光谱中,测量的是单色入射光与散射光的能量差(图1);这称为拉曼位移。

[0152] 通常,对斯托克斯散射光进行监测,因为在环境温度下测量的信号更强烈。图2显示了一些实例光谱;不同的峰表示存在不同的振动模式;一些波段是几个基础峰的重叠区域。对于简单的混合物,可以识别出 $n-1$ 个分析物特有的特定波段,从而在适当的校准后,测量其在固定温度和压力下的强度变化,以对组成完全定量。然而,对于生物介质等复杂系统,不可能严格应用这种简单的方法,因为有太多重叠的波段,这阻碍了直接分配。替代地,除了对原始数据矩阵进行建模外,还必须使用先进的化学计量学模型来获得变量(波长/波数)的线性组合,从而最大化与感兴趣的分析物浓度的协方差。然后可以预测新样品的组成。

[0153] 拉曼光谱提供了一个“分子指纹”,可以对样品例如生物样品进行定性和定量分析。拉曼光谱通常对温度和pH等物理条件的变化很敏感。通常,从生物样品获得的拉曼光谱可包括背景荧光信号,因为此类样品经常包括天然荧光团。在传统的拉曼光谱中,这种背景应该受到最佳激光波数选择的限制,并使用几种常规可用的算法中的一种来去除任何剩余的荧光。

[0154] 拉曼光谱是本领域公知的技术。在本发明中,可以如上所述地在原位使用实时拉曼光谱,从而允许连续测量病毒滴度和/或病毒组分丰度。

[0155] 本文使用的“拉曼光谱”可以指在底物和感兴趣的目标分子(例如,待被拉曼检测的分子)之间不需要结合(例如免疫相互作用)的所有类型的拉曼光谱。底物和目标分子之间的结合可以直接发生,或使用任何类型的结合分子、链霉亲和素/生物素等或抗体片段间接发生。“免疫相互作用”包括使用抗体或抗体片段(如scFv等),这些抗体或抗体片段可以

附着在底物上,特异性结合样品中的感兴趣分子。特别地,本文使用的“拉曼光谱”可以排除表面增强拉曼光谱(surface enhanced Raman spectroscopy,SERS),例如需要底物(例如其可以包括例如Au的金属纳米点)之间免疫相互作用的SERS。SERS要求待检测的分析物被固定在表面上。在本发明方法的一个实施方案中,不执行拉曼光谱来检测已经固定在表面上的样品中的分析物。特别地,在本发明的一个实施方案中,不执行拉曼光谱来检测样本中或已经固定在表面上的样本中的病毒粒子。

[0156] 根据本发明的优选实施方案,SERS不同于拉曼光谱,特别地,SERS需要特定的实验设计来将感兴趣的分析物固定或结合到表面,这导致使用SERS方法的增强的信号强度。然而,分析物的这种固定化需要对样品进行处理,这可导致污染或干扰生物反应器内的条件,包括在样品从这样的系统中提取的情况下。一般来说,SERS更适合于下述方法,即该方法涉及包括感兴趣的分析物、样品中几乎没有污染物的简单样品,而不是根据本文定义的方法在用于处理病毒的生物反应器或生物样品中发现的各组分的复杂混合物。因此,传统的SERS不适合于直接监测或者在线或原位监测复杂样品,包括用于评估病毒滴度的含有病毒粒子的样品,而是更典型地适用于对经过处理且其中待检测分析物已被提纯然后固定在表面上的样品的分析。

[0157] 相反,使用本文定义的拉曼光谱,本发明的方法的优选实施方案可以在线/原位检测样本中的分析物,特别是病毒粒子,而不需要样品中存在的任何分析物的表面附着,特别是不需要病毒粒子的表面附着或固定化。

[0158] 本发明中定义的拉曼光谱特别包括常规类型的拉曼光谱和其他类型的拉曼光谱,例如受激拉曼光谱(stimulated Raman spectroscopy,SRS)、皮科(pico)拉曼、空间偏移拉曼(spatially offset Raman,SORS)、逆SORS、透明拉曼光谱、相干反斯托克拉曼光谱(coherent anti-Stoke Raman spectroscopy,CARS)、相干斯托克斯拉曼光谱(coherent Stokes Raman spectroscopy,CSRS)、共振拉曼光谱(resonance Raman spectroscopy,RR光谱)和全内反射拉曼光谱(total internal reflection Raman spectroscopy,TIR)拉曼。拉曼光谱设备可从Renishaw、WITec、Horiba和ThermoFisher Science等多个供应商处获得。另请参阅:<http://www.optiqgain.com/>,<https://www.timegate.com/>和<https://www.newport.com/>。

[0159] 数据处理与多变量建模

[0160] 本文使用的“多变量”数据是指其中为每个样本测量多个变量的数据,“多变量模型”是使用这种多变量数据建立的模型。拉曼光谱(例如,在细胞培养的一段时间内产生的)包括多变量数据,其中对于测量的每个样本或时间点,可以记录多个波数处的强度。

[0161] 在本发明中,对拉曼光谱和包括光谱的多变量数据(所述拉曼光谱和数据来自对培养物中的病毒生产的原位监测)进行了分析,以识别在使模型能够预测以实现实时病毒滴度和/或病毒组分丰度测量中最重要的一系列光谱变量。特别是,模型预测实现了病毒核酸丰度和病毒结构分子丰度的测量,以便评估病毒核酸与包含培养物中一个或多个病毒结构分子的病毒的比率。创建了从10或12分量多变量模型计算的变量重要性投影(VIP)的图,并且建立了波数变量的重要性,例如关于表3中列出的数据。还创建了从8或15分量多变量模型计算的变量重要性投影(VIP)的图,并且建立了波数变量的重要性,例如关于表1中列出的数据。因此,发明人使用来自拉曼光谱测量的多变量数据,连同来自现有技术分析的与

病毒滴度和病毒核酸丰度或病毒结构分子丰度有关的离线数据,来识别在确定病毒滴度和/或病毒组分丰度时可以评估的波数范围,并进一步建立多变量模型,该多变量模型能够分析在特定条件下针对含有病毒的样本实现的任何拉曼光谱的指定波数范围处的信号强度,以确定病毒滴度和/或病毒组分丰度,以便评估病毒核酸与含有一个或多个病毒结构分子的病毒的比率。可以使用本发明的方法来确定本文描述的任何类型的病毒滴度和/或病毒组分丰度,条件是已经使用与相关类型的病毒组分丰度相关的离线数据建立了用于预测病毒组分丰度的多变量模型。

[0162] 在复杂的生物样品中,明确地将波数范围分配给特定的分析物往往是困难或不可能的。还有一个问题是分析物具有较低浓度,因为信号可能会被细胞培养基中的其他高浓度化合物如水或葡萄糖所遮蔽。如果遮蔽信号变得太强烈,则与它们相关的噪声大于所测量的源自感兴趣的基础分析物的变化,并且感兴趣的组分的信号强度降到检测极限以下。通过对拉曼光谱的分析、对所获得数据的建模以及用病毒滴度和病毒核酸丰度或病毒结构分子丰度的离线测量来校准数据,本发明人已经确定了一系列波数范围,该波数范围与样本中的病毒滴度和/或病毒组分丰度的增加相关,特别是与组分样本中病毒核酸与包含一个或多个病毒结构分子的病毒的比率相关。由于这些波数范围随着时间的推移与病毒滴度和/或病毒组分丰度显示出一致和强烈的相关性,因此没有必要将所述范围分配给特定的分析物。因此,信号被高浓度化合物遮挡的问题对于本发明方法来说不是一个问题。

[0163] 为了识别在本发明中使用的波数范围,获得了多变量模型参数。然后在随后的分析中使用这些参数来推断响应值,并选择用于计算病毒滴度和/或病毒核酸丰度和病毒结构分子丰度的重要变量。在这种情况下,对从含有病毒的样本获得的拉曼数据进行回归。

[0164] 在这种特定情况下,发明人对经过预处理的拉曼数据进行回归。在这方面,可以理解,通常从光谱仪获取的原始拉曼光谱数据可能需要对几个干扰信号进行校正,例如背景荧光,并且对为样本采集的原始光谱进行归一化以校正绝对强度的总变化通常是重要的。因此,技术人员将认识到,可能需要执行光谱预处理来处理此类问题。

[0165] 获得的拉曼光谱的预处理可以使用科学文献中可用的许多算法中的任何一种来执行。特别是,可以使用例如一阶导数、二阶导数(Savitzky等人,1964)和标准正态变量(standard normal variate,SNV)归一化以及多项式背景拟合和去除。Barnes等人(1989)描述了一种标准的正态变量方法。Lieber和Mahadevan-Jansen(2003)描述了一种生物拉曼光谱荧光减法的自动化方法,该方法基于对最小二乘多项式曲线拟合的修正。Zhao等人(2007)描述了一种改进的荧光去除自动算法,该方法基于改进的多多项式拟合。Koch等人(2017)描述了一种基于“软化子”的拉曼光谱预处理基线校正算法。Hu等人(2018)描述了一种基于分段多项式拟合(PPF)法的基线校正方法。Zhang等人(2010)描述了一种自适应迭代重加权惩罚最小二乘(airPLS)算法方法。另见Huang等人(2010)。

[0166] 一旦采取步骤预处理拉曼光谱,如果需要,得到数据建模的参数。在这方面,拉曼数据的回归,特别是预处理拉曼数据可以基于其他技术获得的离线响应进行,例如可以确定病毒滴度的qPCR和p24 ELISA、斑块分析等,或可以用于分析代谢标记物的CuBiAn和LC-MS等。因此,回归可能涉及将预处理的拉曼光谱和离线数据进行比较。多变量回归的典型方法是使用偏最小二乘回归(partial least squares regression,PLS-R)。

[0167] 具体来说,在标准的正交分数PLS-R分析中,寻求拉曼光谱X阵列与响应y之间的线

性关系(例如,y可以是一个向量,其中每个元素代表每个样本的滴度或病毒组分丰度值:

$$[0168] \quad y = \alpha + X\beta + E$$

[0169] 其中 α 和 β 为未知参数,E为误差强度或残差的矩阵)。可以使用不同的基本PLS-R算法,这取决于y对每个样本是包括单个响应值还是多个响应值,其中PLS1算法可以用于单个响应值,PLS2算法可以用于多个响应值。在从拉曼光谱预测病毒滴度和/或病毒组分丰度的情况下,y通常是每个样本的单变量参数,因此通常可以使用PLS1。

[0170] 简单地说,PLS1算法的功能如下。最初,X和y的变量是以均值为中心的(可以从X列的每个元素中减去每个变量的均值,从y的每个元素中减去y的均值。可以为模型选择一些基础因素A,这些因素是可以线性组合使用的因素,以对X建模。第一步,可以将X投影到y上,以找到权重;这些权重定义了X的与y有最大协方差的向量/因子空间的方向。然后这些权重可以归一化为具有单位长度。随后,可以通过将X投影到这些归一化权重上来计算X分数。可以通过将X投影到分数上来计算X负载。类似地,y负载可以通过在这些分数上投影y的转置来计算。当前分量的贡献可以通过减去给定分量的贡献来压缩X和y而从X和y中去除。这可以通过将分量各自的分数和负载向量相乘,并分别从运行的X阵列和y向量中减去所得阵列或向量来进行。

[0171] 然后,在迭代过程中对于每个后续分量,可以以相同的方式再次使用压缩的X和y。即依次确定权重、分数和负载,直到所有A分量耗尽,不再进行进一步的压缩。这种PLS方法和NIPALS算法可以在Wold等人(2001)中找到参考。

[0172] 在每次迭代中,计算的权重、分数和负载向量和标量可以按顺序存储在它们自己的阵列或向量中,即,对于每次迭代,相关向量或标量可以作为阵列中的新列或新行或向量中的新元素放置,其中现有的向量或标量可能是从前一次迭代中获得的。如果要将回归系数用作后续数据建模的参数,则回归系数 β ,可通过乘以最终X区块负载阵列的转置在最终权重矩阵上的投影的倒数得到。通过研究未在迭代模型构建过程中使用的预处理拉曼光谱测试集的预测误差,可以确定最佳分量数量。

[0173] 可以如上所述使用模型参数,如回归系数 β ,并通过与离线分析数据如qPCR/P24Elisa/斑块分析的均方误差(mean squared errors,MSE)进行比较,来进行病毒滴度和/或病毒组分丰度的预测。当预测的MSE达到最小值时,可以选择最优数量的基础分量A。对于病毒核酸丰度,可以使用离线分析,包括qPCR或RT-qPCR。对于病毒结构分子丰度,可以使用离线检测,如ELISA。

[0174] 在例如使用上述过程建立初始的PLS-R模型之后,可以使用许多可用变量选择方法中的一个来选择最重要的变量,例如变量重要性投影(VIP),其识别在y的预测中可能最重要的变量以及解释X的方差。VIP可以生成与X的行相同长度的VIP向量,即VIP向量包含对应于X的每个变量的元素,其中每个元素的数值可以是该变量的重要性的测量。一种常见的方法是,只用由VIP或其他变量选择方法确定的最重要的变量重建上面的初始模型来细化该模型。为了确定哪些变量是重要的,选择了阈值方法。通常,VIP阈值可以被设置为1,因为这是VIP参数的平均值,但是技术人员将理解,这本质上是任意的,并且可以选择其他更高的阈值,例如1.5。

[0175] 如下所述,在本发明中识别的对于确定病毒滴度和/或病毒组分丰度来说重要的波数范围是基于将VIP参数设置为至少1.00或更高。因此,在本发明中,在此阶段不产生大

于1.00的峰值强度的波数范围被排除在外。

[0176] 技术人员会认识到,PLS1算法可以在选择VIP参数后再次运行,但在这种情况下,可以删除VIP阈值以下的X的变量,生成一个新的多变量模型,具有更短的负载向量和更短的回归系数 β 向量。

[0177] 下面列出的波数范围是用波长785nm的激光进行拉曼光谱得到的结果。本发明包括,还可以使用除了785nm以外的不同波长的激光器。使用不同波长的激光获得的波数范围将是相同的,因为拉曼位移(即用于诱导拉曼散射的单色激光束的非弹性散射拉曼光的波长差异)在很大程度上与波长无关。

[0178] 此处使用的“波数”($\tilde{\nu}$)是每单位距离的波长数,单位为 cm^{-1} 。通常, $\tilde{\nu}=1/\lambda$,其中 λ 为波长。

[0179] 在本发明的优选实施方案中,超过1.5的峰(变量重要性投影(VIP))表示波数范围可能对确定病毒的存在和确定病毒滴度和/或病毒组分丰度很重要。确定特定波数范围的信号强度对于通过本发明的方法预测病毒滴度是必不可少的。在本发明实施例中使用1.5VIP阈值来确定哪些波数范围对于病毒滴度和/或病毒组分丰度的预测是重要的。因此,在优选实施方案中,本发明包括一种使用拉曼光谱评估或预测病毒滴度和/或病毒组分丰度的方法,该方法包括从获得的拉曼光谱中确定在这些波数范围中的5个或更多个范围下的信号强度的步骤,所述拉曼光谱从所述样本获得。用测得的预处理信号进行预测。任何后续光谱,即模型建立后的光谱(从中进行预测),都需要使用与用于训练和构建PLS模型的数据完全相同的方法进行预处理。

[0180] 本发明的任何一种方法都涉及测量多个波数范围内每个波数范围内的拉曼散射光总强度以获得样本的波数强度数据集的步骤,其中多个波数范围是预先选定的并且是样本中病毒组分的表征。

[0181] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表1中所列出的波数范围1至12中的4个或更多个,并且其中所述 $\text{VIP} \geq 1.00$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围1至12中的6个或更多个,并且其中所述 $\text{VIP} \geq 1.00$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围1至12中的8个或更多个,并且其中所述 $\text{VIP} \geq 1.00$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围1至12中的10个或更多个,并且其中所述 $\text{VIP} \geq 1.00$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围1至12中的所有12个,并且其中所述 $\text{VIP} \geq 1.00$ 。

[0182] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表1中所列出的波数范围13至22中的4个或更多个,并且其中所述 $\text{VIP} \geq 1.25$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围13至22中的6个或更多个,并且其中所述 $\text{VIP} \geq 1.25$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围13至22中的8个或更多个,并且其中所述 $\text{VIP} \geq 1.25$ 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如表1中所列出的波数范围13至22中的所有10个,并且其中所述 $\text{VIP} \geq 1.25$ 。

[0183] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表1中所列出的波数范围23至30中的4个或更多个,并且其中所述 $\text{VIP} \geq 1.50$ 。所测量的所述

拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围23至30中的6个或更多个,并且其中所述VIP ≥ 1.50 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表1中所列出的波数范围23至30中的所有8个,并且其中所述VIP ≥ 1.50 。

[0184] 表1

	AAV 载体生产								
	VIP ≥ 1.00			VIP ≥ 1.25			VIP ≥ 1.50		
	#	从	至	#	从	至	#	从	至
[0185] 波数范围 / cm^{-1}	1	420	420	13	512	515	23	848	861
	2	510	517	14	846	862	24	994	1035
	3	844	863	15	993	1036	25	1119	1129
	4	992	1037	16	1060	1066	26	1355	1363
	5	1057	1069	17	1115	1134	27	1425	1431
	6	1112	1137	18	1352	1376	28	1597	1608
	7	1182	1184	19	1415	1455	29	1638	1644
	8	1193	1199	20	1596	1611	30	1652	1658
	9	1333	1380	21	1626	1626			
	10	1410	1461	22	1635	1678			
	11	1583	1586						
	12	1594	1692						

[0186] 可以使用如上参考表1描述的拉曼光谱中的多个波数范围来测量病毒滴度和/或病毒组分丰度。在本发明的优选实施方案中,如上参考表1所述的使用拉曼光谱中的多个波数范围测量的病毒滴度和/或病毒组分丰度是腺相关病毒(AAV)滴度。在本发明的一个特别优选实施方案中,如上参考表1描述的使用拉曼光谱中的多个波数范围测量的病毒滴度和/或病毒组分丰度是腺相关病毒血清型8(AAV8)滴度。

[0187] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表2中所列出的波数范围1至20中的4个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围1至20中的6个或更多个,并且其中所述VIP ≥ 1.00 ;或其中所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围1至20中的8个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围1至20中的10个或更多个,并且其中所述VIP ≥ 1.00 。所测量的拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围1至20中的12个或更多个、14个或更多个、16个或更多个或18个或更多个,并且其中所述VIP ≥ 1.00 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围1至20中的所有20个,并且其中所述VIP ≥ 1.00 。

[0188] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表2中所列出的波数范围21至33中的4个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如下表2中所列出的波数范围21至33中的6个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括如下表2中所列出的波数范围21至33中的8个或更多个,并且其中所述VIP ≥ 1.25 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围包括下表2中所列出的波数范围21至33中的10个或更多个、11个或更多个或12个,并且其中所述VIP ≥ 1.25 ;或者其中所测量

的所述拉曼光谱中的所述多个波数范围包括如下表2中所列出的波数范围21至33中的所有13个,并且其中所述VIP ≥ 1.25 。

[0189] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表2中所列出的波数范围34至40中的4个或更多个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围可以包括下表2中所列出的波数范围34至40中的5个或6个,并且其中所述VIP ≥ 1.50 ;或者其中所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表2中所列出的波数范围34至40中的所有7个,并且其中所述VIP ≥ 1.50 。

[0190] 表2

AAV 载体 CAPSID ELISA									
VIP ≥ 1.00			VIP ≥ 1.25			VIP ≥ 1.50			
#	从	至	#	从	至	#	从	至	
波数范围/ cm ⁻¹	1	420	426	21	422	423	34	845	862
	2	447	448	22	843	864	35	995	1010
	3	514	519	23	994	1019	36	1028	1034
	4	824	833	24	1026	1035	37	1060	1066
	5	838	866	25	1057	1069	38	1113	1135
	6	879	884	26	1110	1137	39	1535	1539
	7	993	1037	27	1356	1362	40	1599	1607
	8	1055	1074	28	1415	1420			
	9	1107	1140	29	1450	1452			
	10	1332	1338	30	1530	1543			
	11	1350	1376	31	1598	1607			
	12	1412	1429	32	1675	1675			
	13	1438	1441	33	1689	1690			
	14	1445	1464						
	15	1471	1475						
	16	1486	1506						
	17	1513	1546						
	18	1558	1562						
	19	1597	1609						
	20	1671	1703						

[0192] 病毒滴度和/或病毒组分丰度可以使用拉曼光谱中的多个波数范围进行测量,如上文关于表2所述。在本发明的优选实施方案中,使用如上文关于表2所述的拉曼光谱中多个波数范围测量的病毒滴度和/或病毒组分丰度是腺相关病毒(AAV)滴度。在本发明的一个特别优选实施方案中,使用如上文关于表2所述的拉曼光谱中多个波数范围测量的病毒滴度和/或病毒组分丰度是腺相关病毒血清型8(AAV8)滴度。

[0193] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表3中所列出的波数范围1至28中的5个或更多个,并且其中所述变量重要性投影(VIP) ≥ 1.00 。所测量的所述拉曼光谱中的所述多个波数范围可以包括下表3中所列出的波数范围1至28中的10个或更多个,并且其中所述VIP ≥ 1.00 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围1至28中的15个或更多个,并且其中所述VIP ≥ 1.00 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围1至28中的20个或更多个,并且其中所述VIP ≥ 1.00 。所测量的所述拉曼光谱中的所述

多个波数范围可以包括如下表3中所列出的波数范围1至28中的25个或更多个,并且其中所述VIP ≥ 1.00 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围1至28中的所有28个,并且其中所述VIP ≥ 1.00 。

[0194] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表3中所列出的波数范围29至59中的5个或更多个,并且其中所述变量重要性投影(VIP) ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括下表3中所列出的波数范围29至59中的10个或更多个,并且其中所述VIP ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围29至59中的15个或更多个,并且其中所述VIP ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围29至59中的20个或更多个,并且其中所述VIP ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围29至59中的25个或更多个,并且其中所述VIP ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围29至59中的30个,并且其中所述VIP ≥ 1.25 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围29至59中的所有31个,并且其中所述VIP ≥ 1.25 。

[0195] 在本发明的任何一种方法中,所测量的拉曼光谱中的多个波数范围可以包括如下表3中所列出的波数范围60至81中的5个或更多个,并且其中VIP ≥ 1.50 。所测量的所述拉曼光谱中的所述多个波数范围可以包括下表3中所列出的波数范围60至81中的10个或更多个,并且其中所述VIP ≥ 1.50 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围60至81中的15个或更多个,并且其中所述VIP ≥ 1.50 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围60至81中的20个或更多个,并且其中所述VIP ≥ 1.50 。所测量的所述拉曼光谱中的所述多个波数范围可以包括如下表3中所列出的波数范围60至81中的所有22个,并且其中所述VIP ≥ 1.50 。

[0196] 表3

慢病毒载体生产									
VIP ≥ 1.00			VIP ≥ 1.25			VIP ≥ 1.50			
#	从	至	#	从	至	#	从	至	
波数范围 / cm^{-1}	1	420	438	29	420	421	60	420	420
	2	457	497	30	426	429	61	467	471
	3	503	552	31	434	436	62	474	481
	4	576	580	32	459	486	63	505	529
	5	588	589	33	490	496	64	537	543
	6	604	608	34	504	549	65	836	884
	7	617	621	35	798	800	66	897	902
	8	796	805	36	834	885	67	919	937
	9	808	809	37	892	907	68	995	1043
	10	824	911	38	919	938	69	1046	1046
	11	918	939	39	973	973	70	1049	1071
	12	971	1168	40	981	983	71	1084	1144

[0198]	13	1191	1197	41	990	1145	72	1209	1210
	14	1206	1212	42	1207	1211	73	1271	1273
	15	1234	1237	43	1248	1250	74	1277	1302
	16	1246	1252	44	1270	1322	75	1347	1366
	17	1259	1481	45	1328	1331	76	1386	1433
	18	1497	1500	46	1346	1380	77	1444	1461
	19	1526	1540	47	1383	1473	78	1467	1469
	20	1545	1550	48	1476	1478	79	1610	1612
	21	1584	1591	49	1498	1499	80	1629	1630
	22	1598	1685	50	1528	1528	81	1655	1671
	23	1699	1699	51	1590	1590			
	24	1717	1719	52	1599	1602			
	25	1754	1754	53	1609	1613			
	26	1768	1771	54	1616	1620			
	27	1782	1783	55	1628	1634			
	28	1798	1800	56	1640	1672			
				57	1678	1679			
				58	1769	1769			
				59	1800	1800			

[0199] 可以使用如上关于表3描述的拉曼光谱中的多个波数范围来测量病毒滴度和/或病毒组分丰度。在本发明的优选实施方案中,使用如上关于表3描述的拉曼光谱中的多个波数范围测量的病毒滴度和/或病毒组分丰度是慢病毒滴度和/或病毒组分丰度。

[0200] 在本发明的另一个优选实施方案中,超过1.5的峰(变量重要性投影(VIP))表示波数范围可能对于确定病毒的存在和确定病毒滴度和/或病毒组分丰度是重要的。确定特定波数范围的信号强度对于通过本发明的方法预测病毒滴度和/或病毒组分丰度是必不可少的。在本发明实施方案中使用1.5VIP阈值来确定哪些波数范围对于病毒滴度和/或病毒组分丰度的预测是重要的。因此,在优选实施方案中,本发明包括一种使用拉曼光谱评估或预测病毒滴度和/或病毒组分丰度的方法,该方法包括从获得的拉曼光谱中确定在这些波数范围中的4个或更多个范围处的信号强度的步骤,所述拉曼光谱从所述样本获得。用测得的预处理信号进行预测。任何后续光谱,即模型建立后的光谱(从中进行预测),都需要使用与用于训练和构建PLS模型的数据完全相同的方法进行预处理。

[0201] 本发明的任何方法包括测量多个波数范围中的每个波数范围内的拉曼散射光的总强度以获得样本的波数强度数据集的步骤,其中多个波数范围是预先选择的并且是样本中的病毒组分的表征。

[0202] 确定每个期望的波数范围内的信号强度的步骤需要确定在期望的波数范围内识别的任何峰的水平。技术人员将认识到,如上所述,被认为与病毒滴度和/或病毒组分丰度相关的任何波数范围内的信号强度可以是任何水平,并且当用适当的多变量模型分析时测量这些强度将允许确定病毒滴度和/或病毒组分丰度。如下文更详细讨论的,本发明因此可以包括通过分析使用多变量模型测量的信号强度来评估或计算病毒滴度和/或病毒组分丰度的步骤。这样的多变量模型可以在实施本发明之前准备,或者替代地作为本发明方法的一部分来准备,其中这些方法可以另外包括建立多变量模型的步骤。

[0203] 本发明的方法可以包括在除了本文指定的波长范围之外的其他波数范围确定信

号强度。

[0204] 神经网络也可以用于基于拉曼光谱进行分类,例如在分析病理上的病变组织和健康组织时。神经网络也可用于回归问题,如在应用拉曼数据监测病毒产生时所面临的回归问题,如本文所述。

[0205] 更复杂的方法利用所谓的卷积神经网络(CNN)(深度学习),通常使用Google的TensorFlow后端和Kera API来编写面向对象的Python编程语言的脚本。使用卷积层的优点是,由于网络基本上“学习”了对光谱本身进行预处理以进行最佳滴定/浓度预测的完美方法,因此,预处理变得越来越不必要。在本文中描述的数据处理步骤中使用的这种神经网络是本领域技术人员公知的。欲了解更多信息,请访问例如以下地址:

[0206] <https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#2c9442a91245>

[0207] <http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>

[0208] 如上所述,可以获得多变量数据模型参数,并将其用于本文所定义的病毒滴度和/或病毒组分丰度的定量方法中。如果需要,这些模型参数也可以用于构建本文定义的替代多变量数据模型。本发明人将多变量算法应用于拉曼光谱波数信号强度数据,以获得模型参数,然后将其用于病毒滴度和/或病毒组分丰度的定量。具体而言,本发明人应用了一种多变量算法来获取回归系数,并将其用作模型参数。然而,技术人员应了解,可依据所选模型的性质获得和使用替代的模型参数。因此,在本文中定义的任何方法中,可以适当地选择多变量数据模型参数,并且可以可选地包括回归系数。任何合适的多变量算法都可以应用于拉曼光谱波数信号强度数据来获得模型参数。可以使用多变量回归算法,例如偏最小二乘(PLS)回归算法,可选地,其中PLS算法为非线性迭代偏最小二乘(nonlinear iterative partial least squares, NIPALS)回归算法。还可以使用涉及神经网络的算法来获得模型参数。

[0209] 技术人员还将认识到,本发明的方法可以包括额外的数学数据处理和建模步骤,以定量感兴趣的病毒组分。

[0210] 预估病毒滴度的模型

[0211] 如本文所述,进行了拉曼光谱的化学计量学建模,以识别实时病毒滴度和/或病毒组分丰度的增加与拉曼光谱中看到的波数范围的同一性和强度之间的相关性。所识别的波数范围如上所述。

[0212] 如前所述,本发明需要评估在评估病毒滴度和/或病毒组分丰度时被确定为重要的波数范围中的4或5个或更多个范围的信号强度,并使用已建立的多变量模型(无论是校准的还是未校准的多变量模型)进一步分析强度。

[0213] 技术人员将认识到,不同的多变量模型可以依据待分析的样本而建立,并且构建多变量模型的方法在本领域是众所周知的(例如,参见本文引用的参考文献)。因此,在包括不同类型病毒、不同细胞培养基或不同产生者细胞的样品中,可能需要不同的多变量模型来确定病毒滴度和/或病毒组分丰度。

[0214] 在本发明的一个实施方案中,可以使用以下方法构建多变量模型:

[0215] i) 对使用其他技术(如qPCR和p24 ELISA、斑块分析等)获得的离线响应的预处理

拉曼数据进行回归,如前所述,回归可能涉及比较所述预处理拉曼光谱,

[0216] ii) 使用获得的回归系数来利用预处理数据预测响应值,其中可以通过调整用于多变量回归的分量/因素的基础数量来优化这些预测的质量,

[0217] iii) 使用任何已知的方法进行变量选择,例如变量重要性投影(VIP),它除了解释X外,识别出对预测Y有强大/重要作用的变量,

[0218] iv) 建立细化模型,其中在识别重要的光谱变量之后,可以使用步骤(ii)中描述的相同方法进行另一轮的建模,但其中在模型建立之前,将预处理拉曼光谱阵列中被视为与变量选择不相关的变量或列删除。这导致了一个建立在数据上的更简单的模型,删除了许多不相关的变化。如在(ii)中,分量的基础数量可以通过选择使用最少数量的基础因素构建的模型来优化,使分量之间的变化具有最低的预测误差。

[0219] v) 使用(iv)中确定的回归系数进行未来预测,其中新的预处理拉曼光谱可以乘以获得的回归系数以生成估计值。

[0220] 应认识到,一旦识别了用于分析的波数范围,就可能没有必要重复上述步骤(i)至(iii)。特别是,在这种情况下,构建模型可能只需要步骤iv)。此外,可以使用回归系数以外的数据建模参数。

[0221] 在本发明的另一方面,所述方法可包括制备或构建多变量模型的额外步骤。

[0222] 如上述步骤v)所述,生成的多变量模型的回归系数可用于从一个或多个样本的拉曼光谱中获得病毒滴度和/或病毒组分丰度的估计值。因此,在本发明中,该方法可以包括使用来自多变量模型的回归系数确定病毒滴度和/或病毒组分丰度的步骤。同样的预处理方法用于模型的训练/构建。

[0223] 本发明由以下实施例进一步说明,这不应被解释为进一步限制。在整个本申请中引用的所有图表和所有参考文献、专利和公开的专利申请的内容都通过引用明确地纳入本申请中。

[0224] 实施例

[0225] 实施例1-生产过程中不同阶段慢病毒浓度的计算

[0226] 进行计算,以提供在一个实例病毒生产过程中的不同阶段预期的以mg/ml计的估计慢病毒浓度。使用已知的慢病毒浮力密度和物理滴度来计算浓度。结果如下表所示。

参数	数值	单位	备注
浮力密度	1.15E+00	g/cm ³	实际范围为 1.15 - 1.19 g/ml, 使用较低的值进行保守估计。
粒子半径	7.50E-08	m	
粒子直径	1.50E-07	m	慢病毒的直径约为 0.15-0.2μm, 使用较低的值作保守估计。

[0228]

粒子体积	1.77E-21	m ³	$\frac{4}{3} \times \pi \times r^3$
粒子体积	1.77E-15	cm ³	
粒子质量	2.03E-15	g	
典型物理滴定 (a)	1.00E+08	粒子/ ml	物理滴度大于 qPCR 的复制数 100-1000 倍。
典型物理滴定 (b)	1.00E+07	粒子/ ml	
典型物理滴定 (c)	1.00E+09	粒子/ ml	
典型物理滴定 (d)	1.00E+10	粒子/ ml	
浓度 (a)	2.03E-07	g / ml	
浓度 (b)	2.03E-08	g / ml	
浓度 (c)	2.03E-06	g/ ml	
浓度 (d)	2.03E-05	g / ml	
浓度 (a)	2.03E-04	mg/ml	
浓度 (b)	2.03E-05	mg/ml	
浓度 (c)	2.03E-03	mg/ml	
浓度 (d)	2.03E-02	mg/ml	

[0229] 如所示,慢病毒的估计浓度被发现在2.03E-02到2.03E-05mg/ml的范围内。

[0230] 为了进行比较,下面显示的是基于Buckley和Ryder (2017,应用光谱学,71,1085-1116页)的葡萄糖和苯丙氨酸检测极限值的一些有指导意义的计算。

[0231] 葡萄糖

[0232] mw=180.156g/mol

[0233] 检出限0.6mM (见Buckley和Ryder)

[0234] 摩尔数=体积/1000*摩尔浓度

[0235] $1/1000 \times 0.0006$ —在1ml中

[0236] =0.0000006mol

[0237] 质量=摩尔数*mw

[0238] =0.0000006*180.156

[0239] =0.00011g/ml

[0240] 估计浓度=0.11mg/ml

- [0241] 苯丙氨酸
- [0242] $mw = 165.19 \text{ g/mol}$
- [0243] $LOD = 1.1 \text{ mM}$ (见Buckley和Ryder)
- [0244] 摩尔数 = 体积/1000*摩尔浓度
- [0245] $1/1000 * 0.0011$ — 在1ml中
- [0246] $= 0.0000011 \text{ mol}$
- [0247] 质量 = 摩尔* mw
- [0248] $= 0.0000011 \text{ mol} * 165.19 \text{ g/mol}$
- [0249] $= 0.00018 \text{ g}$
- [0250] 估计浓度 = 0.18 mg/ml
- [0251] 葡萄糖和苯丙氨酸的以 mg/ml 为单位的检出限浓度,比与保守的物理滴定相称的乐观估计浓度高5-10倍。
- [0252] 基于上述,预计培养基中慢病毒的近似浓度(以 mg/mL 为单位)将低于使用拉曼光谱的检测限。
- [0253] 实施例2-HEK 293瞬时过程产生慢病毒
- [0254] 实验方法
- [0255] 细胞培养和瞬时转染
- [0256] HEK293培养物的扩增在 37°C 下在Eppendorf DASbox BioBLU 300生物反应器中,使用FreeStyle 293表达介质(ThermoFisher)进行,不添加任何添加剂。在瞬时转染之前,将细胞搅拌并扩增2天以产生慢病毒。用gag-pol、vsv-g和编码eGFP的基因组转染细胞,从而用Polypplus公司的PEIPro生产LV粒子。
- [0257] 在整个过程中,从每个生物反应器中获取10或12个样本,使用qRT-PCR测量病毒滴度,并通过P24 ELISA进行确认。在整个扩增和病毒生产阶段获得拉曼光谱。
- [0258] RT-qPCR
- [0259] 使用PCR试剂盒:Lenti-X™ qRT-PCR滴定试剂盒(由Takara, Cat#631235)
- [0260] 供应商:Clontech
- [0261] 作用方法:本试剂盒为进一步逆转录PCR扩增试剂盒。该试剂盒的引物靶向HIV-1基因组中与包装信号相邻的保守区域。扩增子通过SYBR绿色荧光检测,最终滴度由用于生成标准曲线的ssRNA标准品确定。病毒滴度的最终定量以病毒基因组/ml提供。
- [0262] P24-ELISA
- [0263] ELISA试剂盒:QuickTiter™慢病毒滴度试剂盒(慢病毒相关HIV p24) (Cat#VPK-107)
- [0264] 供应商:Cell Biolabs, Inc
- [0265] 作用方法:该试剂盒是开发的一种酶免疫测定法,仅用于检测和定量慢病毒相关HIV-1p24核心蛋白。病毒相关p24可以用p24滴度(ng/ml)或粒子/ml来定量,假设每个慢病毒粒子大约有2000个p24分子。
- [0266] 拉曼光谱
- [0267] 使用Kasier Optics RxN2拉曼光谱仪进行拉曼测量。该光谱仪能够连续监测4个探针通道。RxN2激发源为785nm近红外二极管激光器,每个探针头的标称输出功率为约

270mW。样品包括4个Eppendorf,dasBox BioBLU一次性使用系统的内容物。光束通过4个Kaiser Optics过滤光纤MR探针和BioOptic 220's(每个生物反应器一套)传输到每个样品生物反应器。在进行过程测量之前,RxN2系统稳定1小时,然后使用RxN2的内部自动校准标准对4个探针通道中的每个通道进行校准,此外,使用国家标准与技术研究所(National Institute of Standards and Technology,NIST)认证的光源(HCA)对每个探针通道进行CCD灵敏度校正。散射光是使用与光束传输中所用相同的BioOptic 220's和MR探针收集的。在每个MR探针中,散射光通过第二光纤传送到RxN2 f\1.8成像光谱仪。利用全息陷波滤波器滤除瑞利散射光后,将拉曼散射光被引导至凯撒光学全息透射光栅,然后成像到热电冷却的1024像素CCD检测器上。该系统有效带宽为100-3425 cm^{-1} ,分辨率为4 cm^{-1} 。拉曼光谱采集自100-3425 cm^{-1} ,积分时间约为15分钟/通道,包括CCD读出时间,将累计75次的10s采集平均化,生成每个测量光谱。依次测量每个通道。在整个过程的不同时间,从每个生物反应器中获得液体样品,并注明时间点,以使离线分析数据与相应的拉曼光谱能够进行事后匹配。

[0268] 拉曼数据分析

[0269] 所有数据分析均在MATLAB(The MathWorks,MA,USA)版本R2017b中进行。通过将整个光谱归一化至 $\sim 3000\text{cm}^{-1}$ 处的水波段的峰强度对原始拉曼光谱进行预处理。除去420-1800 cm^{-1} 区域的中等荧光背景信号。选择这个范围的低端是为了避免这样的拉曼波段:该拉曼波段可能源自BioOptic-220的蓝宝石窗口,或者是拉曼仪器和探针的光学设计的伪影。然后约减归一化光谱,检查明显的异常值和伪影。识别与离线采样时间点相关的光谱,并创建预处理光谱的模型训练子集。然后将预处理光谱的训练集用于化学计量学建模。在化学计量学建模之前,光谱是以平均值为中心的。建立了几个初步的对关键分析物和病毒滴度的潜在结构预测-回归(PLS-R)模型。这些模型允许您对含有已知浓度的感兴趣分析物(病毒滴度)的样品进行多变量拉曼光谱回归。基于这些计算,将来可以预测分析物的浓度。在训练数据上使用10倍交叉验证程序来准备模型,即将1/10的数据从训练数据中随机选择和删除,并将其用于评估模型性能,这样做了10次,误差值、模型精度/性能统计数据是针对每个10倍训练集获得的平均值。选择基础分量或基向量的数量是构建受监督线性模型(如PLS-R)的重要步骤。在这项工作中,通过检查交叉验证后预测的均方误差(MSECV)作为分量数的函数的图来识别最优的基础分量数;最小值表示PLS分量的最优数量。变量选择的第二阶段需要通过只选择对预测最重要的波数/变量来优化所建立的模型。这是使用变量重要性投影(VIP)方法进行的。然而,存在许多进行变量选择的方法。一个典型的VIP图如图5所示。通常,VIP值大于1的变量用于最终模型。然而,在本文中,我们使用几个VIP阈值来构建和评估模型,以确定做出良好预测所需的最小谱变量数量,以及无法对离线RT-qPCR数据建模的阈值。随着VIP阈值的增加,识别的谱变量数量减少。一旦确定了每个VIP阈值的显著变量,就建立了最终模型。随后,这些模型被用于预测中间病毒滴度值,即所有可用运行的每个离线数据点之间的值。

[0270] 初步病毒滴度模型评价及范围选择

[0271] 用于化学计量学建模的预处理拉曼光谱示例如图2所示。

[0272] 在整个项目中对病毒滴度进行了监测,通过RT-qPCR获得的具有代表性的滴度如图3所示。

[0273] 初始PLS-R模型交叉验证后,预测的均方误差图如图4所示。当使用所有光谱变量

或通道时,当使用12个PLS分量时,有效预测误差最小。

[0274] 因此,从这个图(图4)可以得出结论,预测误差最小化和模型简单化之间的最佳折衷在于12分量模型。通过增加分量的数量可以获得预测能力的小幅改进,但这可能只是纳入噪声和模型过度拟合到训练集的结果。也就是说,建立了一个模型,该模型仅对训练数据具有预测性,而不是新的未见过的数据,并且是基于测量变量和因变量/响应变量之间的伪相关性。

[0275] 在初始12分量模型准备之后,对不同的变量选择方法进行了评估,以选择最终模型的最佳/最具预测性的光谱变量。本文的目的是从模型中去除不必要的光谱通道/变量,以增强其简约性,并且只包括物理上有意义的信息。最后计算变量重要性投影(VIP),以确定哪些光谱变量在预测病毒复制数方面最重要(图5A)。为了评估和识别做出可接受的物理滴度预测所需的光谱变量的最小数量,研究了几个变量重要性阈值作为保留变量的标准;一般情况下,使用的VIP阈值为1—研究了1.00-1.75的阈值。图5B示出了VIP算法识别为被视作最重要区域的变量或波数范围,即那些大于选定阈值的区域,在本例中为1.5,图5C按重要性顺序示出了这些波数范围。

[0276] 在光谱变量的数量约简后,进一步评估基础潜变量的数量。PLS分量的最佳数量可以随模型中使用的光谱变量或波数的数量而变化。在使用1.5阈值进行光谱变量约简后,发现PLS潜变量的最佳数量为10。图6显示了具有不同数量的基础分量的细化模型的MSECV图。预测的均方误差随着基础分量数量的增加而增加,这一事实表明,在包括超过10个PLS分量的情况下,所产生的模型过度拟合了训练集。

[0277] 利用从10-潜变量和 $VIP \geq 1.5$ 选择的光谱变量(保守)模型得到的回归系数估计的3次运行中的每一次运行的RT-qPCR病毒复制数的模型预测分别如下述图7、8和9所示。结果表明,使用拉曼光谱数据的模型与病毒滴度的离线测量随时间是一致的。由RT-qPCR法和P24 ELISA法得到的滴度的比较见图10。

[0278] 图11描述了如何使用本发明的方法来监测病毒生产培养的阶段。由于拉曼光谱是实时、连续使用的,在本发明的方法中,可以准确地跟踪病毒生产速率的变化。因此,可以识别从开始到生产阶段,以及从生产阶段到结束阶段的变化。

[0279] 实施例3-精细化的病毒滴度模型评估和范围选择

[0280] 为了进一步精细化用于病毒滴度评估的波数范围选择,使用上述实施例2中所述的相同方法进行了对使用其他样本的进一步研究。使用这种方法,通过应用 ≥ 1.00 的变量重要性投影(VIP)阈值,来识别用于计算病毒滴度的各种波数范围。通过应用 ≥ 1.25 的变量重要性投影(VIP)阈值来识别额外的波数范围,并通过应用 ≥ 1.50 的变量重要性投影(VIP)阈值来识别进一步的波数范围。结果如上表3所示。

[0281] 图12显示了用于定量病毒滴度的公式的概要示意图。该公式应用于由多变量回归算法得到的回归系数,回归算法应用于归一化的拉曼信号强度数据。

[0282] 进行了进一步的分析,以分析可用于准确估计病毒滴度的波数范围的数量。

[0283] 识别了对病毒载体生产重要的范围,即在使用扩展光谱范围($\sim 420 - 1800\text{cm}^{-1}$)进行初始PLS建模后,通过 ≥ 1.00 的变量重要性投影(VIP)识别的重要范围(即如上表3所列的波数范围1至28),并进行了进一步分析。

[0284] 数据以4:1的比率被分成随机选择的成对的训练和测试数据块,即,用于模型构建

(80%)和模型测试(20%)的拉曼光谱及其相关的离线病毒滴度数据。

[0285] 对于不同的训练和测试对,随机评估通过VIP认为重要的范围的不同组合,即对于范围1-28中的每r个总数,基于模型性能 R^2 统计量(注, $R^2=1-\text{残差平方和}/\text{总平方和}$)评估许多组合,并评估不同模型性能的标准偏差以生成置信区间。通过选择几个训练/测试数据对的 R^2 均值的平均值约为0.5的范围数来识别最小范围数。图13显示了 R^2 作为波数范围数的函数的图。

[0286] 该分析确定5是提供病毒滴度估计所需的最小波数范围数。

[0287] 因此,在本发明的任何一种方法中,如表3所示,在 ≥ 1.00 的VIP阈值下识别的波数范围1至28中的5个或更多个可用于计算病毒滴度,如本文中更详细地描述。在本发明的任一方法中,优选地,如表3所示,在 ≥ 1.25 的VIP阈值下识别的波数范围29至59中的5个或更多个可以用于计算病毒滴度,如本文中更详细地描述。在本发明的任何一种方法中,更优选地,如表3所示,在 ≥ 1.50 的VIP阈值下识别的波数范围60至81中的5个或更多个可用于计算病毒滴度,如本文中更详细地描述。在任一这些方法中,如本文更详细描述的,优选10个或更多个波数范围可用于计算病毒滴度,更优选的是15个或更多个,或还更优选的是20个或更多个。

[0288] 实施例4—从HEK293瞬时过程生产AAV8

[0289] 实验方法

[0290] 细胞培养和瞬时转染

[0291] HEK 293F培养物在37℃下在Eppendorf DASbox BioBLU 300生物反应器中以BalanCD培养基(Irvine Scientific)和4mM GlutaMAX (Fisher)进行扩增。在瞬时转染之前,将细胞搅拌并扩增24小时以产生AAV8。用rep,cap基因组编码的eGFP质粒和辅助质粒(E2A,E4)在无血清的Opti-MEM (Gibco)中转染细胞,从而用PEIPro (Polypplus转染)产生AAV8粒子。

[0292] 在整个过程中,从每个生物反应器中获取12个样本,使用qRT-PCR测量病毒滴度。在整个扩增和病毒生产阶段获得拉曼光谱。

[0293] RT-qPCR

[0294] 采用基于TaqMan™的实时qPCR测量AAV8样本的病毒滴度,最终定量以病毒基因组/mL (VG/mL) 来提供。该分析实验的引物靶向AAV8病毒基因组中的ITR2序列。扩增子用TaqMan™荧光探针检测。病毒滴度由线性化质粒生成的标准曲线确定。

[0295] 拉曼光谱

[0296] 使用Kasier Optics RxN2拉曼光谱仪进行拉曼测量。该光谱仪能够连续监测4个探针通道。RxN2激发源为785nm近红外二极管激光器,每个探针头的标称输出功率为约270mW。样品包括4个Eppendorf,dasBox BioBLU一次性使用系统的内容物。光束通过4个Kaiser Optics过滤光纤MR探针和BioOptic 220探针(每个生物反应器一套)传输到每个样品生物反应器。在进行过程测量之前,RxN2系统稳定1小时,然后使用RxN2的内部自动校准标准对4个探针通道中的每个通道进行校准,此外,使用国家标准与技术研究所(National Institute of Standards and Technology,NIST)认证的光源(HCA)对每个探针通道进行CCD灵敏度校正。散射光是使用与光束传输中所用相同的BioOptic 220和MR探针收集的。在每个MR探针中,散射光通过第二光纤传送到RxN2 f\1.8成像光谱仪。利用全息陷波滤波器

滤除瑞利散射光后,将拉曼散射光被引导至凯撒光学全息透射光栅,然后成像到热电冷却的1024像素CCD检测器上。该系统有效带宽为 $100\text{-}3425\text{cm}^{-1}$,分辨率为 4cm^{-1} 。拉曼光谱采集自 $100\text{-}3425\text{cm}^{-1}$,积分时间约为15分钟/通道,包括CCD读出时间,将累计75次的10s采集平均化,生成每个测量光谱。依次测量每个通道。在整个过程的不同时间,从每个生物反应器中获得液体样品,并注明时间点,以使离线分析数据与相应的拉曼光谱能够进行事后匹配。

[0297] 拉曼数据分析

[0298] 所有数据分析均在MATLAB(The MathWorks,MA,USA)版本R2019b中进行。通过将整个光谱归一化至 $\sim 3000\text{cm}^{-1}$ 处的水波段的峰强度对原始拉曼光谱进行预处理。除去 $420\text{-}1800\text{cm}^{-1}$ 区域的中等荧光背景信号。选择这个范围的低端是为了避免这样的拉曼波段:该拉曼波段可能源自BioOptic-220的蓝宝石窗口,或者是拉曼仪器和探针的光学设计的伪影。然后约减归一化光谱,检查明显的异常值和伪影。识别与离线采样时间点相关的光谱,并创建预处理光谱的模型训练子集。然后将预处理光谱的训练集用于化学计量学建模。在化学计量学建模之前,光谱是以平均值为中心的。建立了几个初步的对关键分析物和病毒滴度的潜在结构预测-回归(PLS-R)模型。这些模型允许您对含有已知浓度的感兴趣分析物(病毒滴度)的样品进行多变量拉曼光谱回归。基于这些计算,将来可以预测分析物的浓度。在训练数据上使用10倍交叉验证程序来准备模型,即将1/10的数据从训练数据中随机选择和删除,并将其用于评估模型性能,这样做了10次,误差值、模型精度/性能统计数据是针对每个10倍训练集获得的平均值。选择基础分量或基向量的数量是构建受监督线性模型(如PLS-R)的重要步骤。在这项工作中,通过检查交叉验证后预测的均方误差(MSECV)作为分量数的函数的图来识别最优的基础分量数;最小值表示PLS分量的最优数量。变量选择的第二阶段需要通过只选择对预测最重要的波数/变量来优化所建立的模型。这是使用变量重要性投影(VIP)方法进行的。然而,存在许多进行变量选择的方法。一个典型的VIP图如图18所示。通常,VIP值大于1的变量用于最终模型。然而,在本文中,我们使用几个VIP阈值来构建和评估模型,以确定做出良好预测所需的最小谱变量数量,以及无法对离线RT-qPCR数据建模的阈值。随着VIP阈值的增加,识别的谱变量数量减少。一旦确定了每个VIP阈值的显著变量,就建立了最终模型。随后,这些模型被用于预测中间病毒滴度值,即所有可用运行的每个离线数据点之间的值。

[0299] 初步病毒滴度模型评价及范围选择

[0300] 用于AAV化学计量学建模的预处理拉曼光谱示例如图14所示。

[0301] 在整个项目中监测AAV滴度,通过RT-qPCR获得的代表性滴度如图15所示。

[0302] 初始PLS-R模型交叉验证后,预测的均方误差图如图16所示。当使用所有光谱变量或通道时,当使用15个PLS分量时,有效预测误差最小。

[0303] 因此,从这个图(图16)可以得出结论,预测误差最小化和模型简单化之间的最佳折衷在于15分量模型。在准备了初始的15分量模型之后,评估了不同的变量选择方法,以为最终模型选择最优/最具预测性的光谱变量。本文的目的是从模型中去除不必要的光谱通道/变量,以增强其简约性,并且只包括物理上有意义的信息。最后计算变量重要性投影(VIP),以确定哪些谱变量在预测病毒复制数方面具有最大的重要性(图17A)。为了评估和确定做出可接受的物理滴度预测所需的光谱变量的最小数量,研究了几个变量重要性阈值作为保留变量的标准;一般情况下,使用的VIP阈值为1—研究了1.00-1.75的阈值。图17B示

出了VIP算法识别为被认为最重要区域的变量或波数范围,即那些大于选定阈值的区域,在本例中为1.0,图17C按重要性顺序示出了这些波数范围。

[0304] 在光谱变量的数量约简后,进一步评估基础潜变量的数量。PLS分量的最佳数量可以随模型中使用的光谱变量或波数的数量而变化。在使用阈值为1.0的光谱变量约简后,PLS潜变量的最佳数量为9。图18显示了具有不同数量的基础分量的细化模型的MSECV图。预测的均方误差随着基础分量数量的增加而增加,这一事实表明,在包括多于9个PLS分量的情况下,所产生的模型过度拟合了训练集。

[0305] 利用从9-潜变量和VIP>1.0选择的光谱变量(保守)模型得出的回归系数,对4个生物反应器的示例运行的RT-qPCR病毒复制数的模型预测如下图19所示。结果表明,使用拉曼光谱数据的模型与离线测量的病毒滴度随时间是一致的。

[0306] 实施例5—精细化的AAV病毒滴度模型评估和范围选择

[0307] 对实施例4中的描述进行进一步分析,以分析可用于提供对AAV病毒滴度准确估计的波数范围数。

[0308] 确定了对病毒载体生产重要的范围,即在使用扩展光谱范围($\sim 420 - 1800\text{cm}^{-1}$)进行初始PLS建模后,通过变量重要性投影(VIP) ≥ 1.00 确定的重要范围(即如上表1所列的波数范围1至12),并进行了进一步分析。

[0309] 数据以4:1的比率被分成随机选择的成对的训练和测试数据块,即,用于模型构建(80%)和模型测试(20%)的拉曼光谱及其相关的离线病毒滴度数据。

[0310] 对于不同的训练和测试对,随机评估通过VIP认为重要的范围的不同组合,即对于范围1-(12)中的每r个总数,基于模型性能 R^2 统计量(注, $R^2 = 1 - \text{残差平方和} / \text{总平方和}$)评估许多组合,并评估不同模型性能的标准偏差以生成置信区间。通过选择几个训练/测试数据对的 R^2 均值的平均值约为0.5的范围数来确定最小的范围数。图20显示了 R^2 作为波数范围数的函数的图。

[0311] 该分析确定了4是提供AAV病毒滴度估计所需的最小波数范围数。

[0312] 因此,在本发明的任何一种方法中,如表1所示,在 ≥ 1.00 的VIP阈值下识别的波数范围1至12中4个或更多个可用于计算病毒滴度,如本文更详细地描述的。在任一这些方法中,如本文更详细描述的,优选地可以使用6个或更多个波数范围来计算病毒滴度,更优选8个或更多个,或者仍更优选10个或更多个,或者最优选全部12个。如本文更详细描述的,在本发明的任何一种方法中,如表1所示,在 ≥ 1.25 的VIP阈值下识别的波数范围13至22中的4个或更多个可用于计算病毒滴度。在任一这些方法中,如本文更详细描述的,优选可以使用6个或更多个波数范围来计算病毒滴度,更优选8个或更多个,或最优选全部10个。在本发明的任何一种方法中,如本文更详细描述的,如表1所示,在 ≥ 1.50 的VIP阈值下识别的波数范围23至30中的4个或更多个可用于计算病毒滴度。在这些方法中的任何一种中,如本文更详细描述的,优选可以使用6个或更多个波数范围来计算病毒滴度,或最优选全部8个。

[0313] 实施例7-AAV8由HEK 293瞬时过程产生,空满比的确定

[0314] 实验方法

[0315] 细胞培养和瞬时转染

[0316] HEK 293F培养物在37°C在Eppendorf DASbox BioBLU 300生物反应器中以BalanCD培养基(Irvine Scientific)和4mM GlutaMAX(Fisher)进行扩增。在瞬时转染之

前,将细胞搅拌并扩增24小时以产生AAV8。用rep, cap, 基因组编码的eGFP质粒和辅助质粒(E2A, E4) 在无血清的Opti-MEM (Gibco) 中转染细胞,从而用PEIPro (Polypplus转染) 产生AAV8粒子。

[0317] 在整个过程中,从每个生物反应器中获取11个样本,使用RT-qPCR (基因组复制数/ml) 测量病毒滴度,其中5个样本额外用于ELISA (总粒子数/ml)。在整个扩增和病毒生产阶段获得拉曼光谱。

[0318] RT-qPCR

[0319] 采用基于TaqMan™的实时qPCR方法测定AAV8样本的病毒滴度,最终定量以病毒基因组/mL (VG/mL) 的形式提供。该分析实验的引物靶向AAV8病毒基因组中的ITR2序列。扩增子用TaqMan™荧光探针检测。病毒滴度由线性化质粒生成的标准曲线确定。

[0320] ELISA

[0321] 通过ELISA确定AAV8细胞外样品中AAV8总衣壳滴度,最终定量为总粒子/mL (TP/mL)。为了准确定量每个样品中的TP/mL,使用已知粒子浓度的重建AAV8标准以生成标准曲线。为了进行ELISA,一种针对组装的AAV8衣壳 (克隆ADK8) 上的一个构象表位特异性的小鼠单克隆抗体被涂覆在微量滴度板的条带上,并用于捕获样本中的AAV8粒子。通过两个步骤检测捕获的AAV8粒子:1) 生物素偶联的抗AAV8抗体与免疫复合物结合,2) 链霉亲和素过氧化物酶偶联物与生物素分子反应。添加四甲基联苯胺 (TMB) 底物溶液后发生显色反应,显色反应与特异性结合病毒粒子的数量成正比。然后在450nm处用光度法测量吸光度。

[0322] 拉曼光谱

[0323] 使用Kaiser Optics RxN2拉曼光谱仪进行拉曼测量。该光谱仪能够连续监测4个探针通道。RxN2激发源为785nm近红外二极管激光器,每个探针头的标称输出功率为约270mW。样品包括4个Eppendorf, dasBox BioBLU一次性使用系统的内容物。光束通过4个Kaiser Optics过滤光纤MR探针和BioOptic 220's (每个生物反应器一套) 传输到每个样品生物反应器。在进行过程测量之前, RxN2系统稳定1小时,然后使用RxN2的内部自动校准标准对4个探针通道中的每个通道进行校准,此外,使用国家标准与技术研究所 (National Institute of Standards and Technology, NIST) 认证的光源 (HCA) 对每个探针通道进行CCD灵敏度校正。散射光是使用与光束传输中所用相同的BioOptic 220's和MR探针收集的。在每个MR探针中,散射光通过第二光纤传送到RxN2 f\1.8成像光谱仪。利用全息陷波滤波器滤除瑞利散射光后,将拉曼散射光被引导至凯撒光学全息透射光栅,然后成像到热电冷却的1024像素CCD检测器上。该系统有效带宽为100-3425cm⁻¹,分辨率为4cm⁻¹。拉曼光谱采集自100-3425cm⁻¹,积分时间约为15分钟/通道,包括CCD读出时间,将累计75次的10s采集平均化,生成每个测量光谱。依次测量每个通道。在整个过程的不同时间,从每个生物反应器中获得液体样品,并注明时间点,以使离线分析数据与相应的拉曼光谱能够进行事后匹配。

[0324] 拉曼数据分析

[0325] 所有数据分析均在MATLAB (The MathWorks, MA, USA) 版本R2019b中进行。通过将整个光谱归一化至~3000cm⁻¹处的水波段的峰强度对原始拉曼光谱进行预处理。除去420-1800cm⁻¹区域的中等荧光背景信号。选择这个范围的低端是为了避免这样的拉曼波段:该拉曼波段可能源自BioOptic-220的蓝宝石窗口,或者是拉曼仪器和探针的光学设计的伪影。然后约减归一化光谱,检查明显的异常值和伪影。识别与离线采样时间点相关的光谱,并创

建预处理光谱的模型训练子集。然后将预处理光谱的训练集用于化学计量学建模。在化学计量学建模之前,光谱是以平均值为中心的。建立了几个初步的对关键分析物和病毒滴度的潜在结构预测-回归 (PLS-R) 模型 (一个是基于RT-qPCR的病毒滴度模型,校准为基因组复制数/ml,一个是基于AAV8 ELISA的病毒滴度模型,校准为总粒子/ml)。这些模型允许您对含有已知浓度的感兴趣分析物 (病毒滴度,由不同的测定方法确定,在本实施例中为RT-qPCR和ELISA) 的样品进行多变量拉曼光谱回归。基于这些计算,将来可以预测分析物的浓度。在训练数据上使用10倍交叉验证程序来准备模型,即将1/10的数据从训练数据中随机选择和删除,并将其用于评估模型性能,这样做了10次,误差值、模型精度/性能统计数据是针对每个10倍训练集获得的平均值。选择基础分量或基向量的数量是构建受监督线性模型 (如PLS-R) 的重要步骤。在这项工作中,通过检查交叉验证后预测的均方误差 (MSECV) 作为分量数的函数的图来识别每个模型的最优的基础分量数;最小值表示对于给定模型的PLS分量的最优数量。变量选择的第二阶段需要通过只选择对预测最重要的波数/变量来优化所建立的模型。这是使用变量重要性投影 (VIP) 方法进行的。然而,存在许多进行变量选择的方法。一个典型的VIP图如图26A所示。通常,VIP值大于1的变量用于最终模型。然而,在本文中,我们使用几个VIP阈值来构建和评估模型,以确定做出良好预测所需的最小谱变量数量,以及无法对感兴趣的离线数据建模的阈值。随着VIP阈值的增加,识别的谱变量数量减少。一旦确定了每个VIP阈值的显著变量,就建立了最终模型。随后,这些模型被用于预测中间病毒滴度值,以基因组复制数/ml和来自拉曼光谱的总粒子/ml的形式,即所有可用运行的每个离线数据点之间的值。

[0326] 病毒滴度模型初步评估和范围选择

[0327] 用于AAV滴度 (基因组复制数/ml和总粒子数/ml) 化学计量学建模的预处理拉曼光谱示例如图21所示。

[0328] 在整个项目中监测了AAV滴度,图22总结了通过RT-qPCR获得的代表性滴度 (基因组复制数/ml),图23显示了通过ELISA获得的每毫升的代表性总粒子。

[0329] 图24显示了基因组复制数/ml的初始PLS-R模型交叉验证后的预测的均方误差图。当使用所有光谱变量或通道时,当使用15个PLS分量时,有效预测误差最小。图25显示了从ELISA数据校准的总粒子数/ml的初始PLS-R模型交叉验证后的预测的另一均方误差图。当使用所有光谱变量或通道时,当使用14个PLS分量时,有效预测误差最小。这些分量数量的选择在预测误差最小化和模型简单化之间提供了一个很好的折衷。

[0330] 在准备了初始的15和14分量模型之后,评估了不同的变量选择方法,以分别为最终的RT-qPCR和ELISA校准模型选择最佳/最具预测性的光谱变量。本文的目的是从两个模型中删除不必要的光谱通道/变量,以增强它们的简洁性,并且只包括物理上有意义的信息。最后计算变量重要性投影 (VIP) 以确定哪些光谱变量在预测病毒复制数方面具有最大重要性 (图26A)。为了评估和确定做出可接受的物理滴度预测所需的光谱变量的最小数量,研究了几个变量重要性阈值作为保留变量的标准;通常,使用的VIP阈值为1—研究了1.00-1.75的阈值。图26B显示可变或波数范围,VIP算法将其识别为对于预测每毫升基因组复制最重要的区域,即大于选定阈值的区域,在本例中为1.0,图26C按重要性顺序显示这些波数范围。

[0331] 此外,还进行了类似的分析,以确定使用上述ELISA预测每毫升粒子数的最重要的

光谱变量。计算变量重要性投影 (VIP) 以确定哪些光谱变量在预测病毒粒子数量方面具有最大重要性 (图27A)。为了评估和确定做出可接受的物理滴度预测所需的光谱变量的最小数量,研究了几个变量重要性阈值作为保留变量的标准;通常使用VIP阈值1—研究了1.00-1.75的阈值。图27B显示VIP算法识别为最重要的区域的变量或波数范围,即大于所选阈值的区域,在本情况下为1.0,图27C按重要性顺序显示这些波数范围。

[0332] 在减少光谱变量的数量之后,分别对病毒复制数和病毒粒子数模型的基础潜变量的数量进行了进一步的评估。最佳PLS分量的数目可以随模型中使用的光谱变量或波数的数目而变化。在使用阈值1.0进行光谱变量约简后,发现对于每毫升基因组复制数和每毫升粒子数模型,最佳的PLS潜变量数均为10。图28和29显示了具有不同基础分量数量的精细化模型的MSECV图。

[0333] 下面的图30和图31显示了使用从10潜变量和VIP>1.0选定的光谱变量(保守)模型获得的回归系数估计的8个生物反应器的示例运行的RT-qPCR病毒复制数的模型预测。结果表明,使用拉曼光谱数据的模型与病毒滴度(基因组复制数/ml)的离线测量随时间是一致的。图32和图33显示了使用从10个潜变量和VIP>1.0选定的光谱变量(保守)模型获得的回归系数估算的8个生物反应器的示例运行的ELISA总粒子数的类似预测。结果表明,使用拉曼光谱数据的模型与病毒滴度(粒子数/ml)的离线测量随时间是一致的。

[0334] 一种估计单个AAV样本空满比的百分比的方法是将每毫升的基因组复制数(RT-qPCR)除以每毫升的总粒子数(ELISA),然后将该数字乘以100。基于这两种病毒滴度确定方法,可以使用上面开发的预测模型(图30-33)的输出进行类似的计算。这些计算的结果如图34和图35所示,从这些转染的培养物中可以看出,空满比与使用离线RT-qPCR数据和ELISA数据所做的估计很好地匹配。

[0335] 实施例8—AAV的精细化的ELISA病毒滴度模型评估和范围选择

[0336] 可以对AAV8的ELISA模型训练进行进一步的分析,例如在上述实施例3和5中描述的,以计算提供AAV病毒滴度估计所需的波数范围数,具体为总粒子/ml。

[0337] 将使用对AAV8ELISA确定为重要的范围,即使用扩展光谱范围($\sim 420-1800\text{cm}^{-1}$)进行初始PLS建模后,由变量重要性投影(VIP) ≥ 1.00 确定为重要的范围(即,如上图27B和图27C所示的波数范围1至20),并将进行进一步分析。

[0338] 数据将以4:1的比率分成被随机选择的成对的训练和测试数据块,即模型构建(80%)和模型测试(20%)的拉曼光谱及其相关的离线病毒滴度数据。

[0339] 对于不同的训练和测试对,随机评估被VIP认为重要的范围的不同组合,即对于范围1-(20)中的每r个总数,基于模型性能统计量例如 R^2 统计量(注, $R^2=1-\text{残差平方和}/\text{总平方和}$)评估许多组合,并评估不同模型性能的标准偏差以生成置信区间。最小范围数将通过选择若干训练/测试数据对的 R^2 均值的平均值约为0.5的范围数来确定。除 R^2 之外的其他性能统计信息也可用于此方法。

[0340] 参考文献

[0341] Barnes, R.J., Dhanoa, M.S. and Lister, S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. Applied Spectroscopy. 1989, 43 (5): pp772-777.

[0342] Buckley, K. and Ryder, A.G. Applications of Raman Spectroscopy in

Biopharmaceutical Manufacturing:A Short Review.Applied Spectroscopy.2017,71 (6):pp1085-1116.

[0343] Hu H.,Bai,J.Xia,G.Zhang,W.Ma,Y.Improved Baseline Correction Method Based on Polynomial Fitting for Raman Spectroscopy.Photonic Sensors.2018,8 (4):pp332-340.

[0344] Huang,J.,Romero-Torres,S.and Moshgbar,M.Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy.American Pharmaceutical Review.2010.

[0345] Koch,M.,Suhr,C.,Roth,B.and Meinhardt-Wollweber,M.Iterative morphological and mollifier-based baseline correction for Raman spectra.Journal of Raman Spectroscopy.2017,48 (2):pp336-342.

[0346] Lee,J.H.,Kim,B.C.,Oh,B.K.and Choi,J.W.Rapid and Sensitive Determination of HIV-1 Virus Based on Surface Enhanced Raman Spectroscopy.J. Biomed.Nanotechnol.2015,11 (12):pp2223-2230.

[0347] Lieber,C.A.and Mahadevan-Jansen,A.Automated method for subtraction of fluorescence from biological Raman spectra.Applied Spectroscopy.2003,57 (11): pp1363-1367.

[0348] Savitzky,A.and Golay,M.J.E.Smoothing and Differentiation of Data by Simplified Least Squares Procedures.Analytical Chemistry.1964,36 (8):pp1627-1639,

[0349] Wold,S.,Sjostrom,M and Eriksson,L.PLS-regression:a basic tool of chemometrics.Chemometrics and Intelligent Laboratory Systems.2001,58:pp109-130.

[0350] Zhang,Z.M.,Chen,S.and Liang,Y.Z.Baseline correction using adaptive iteratively reweighted penalized least squares.Analyst.2010,135 (5):pp1138-1146.

[0351] Zhao,J.,Lui,H.,McLean,D.I.and Zeng,H.Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy.Applied Spectroscopy.2007,61 (11):pp1225-32.

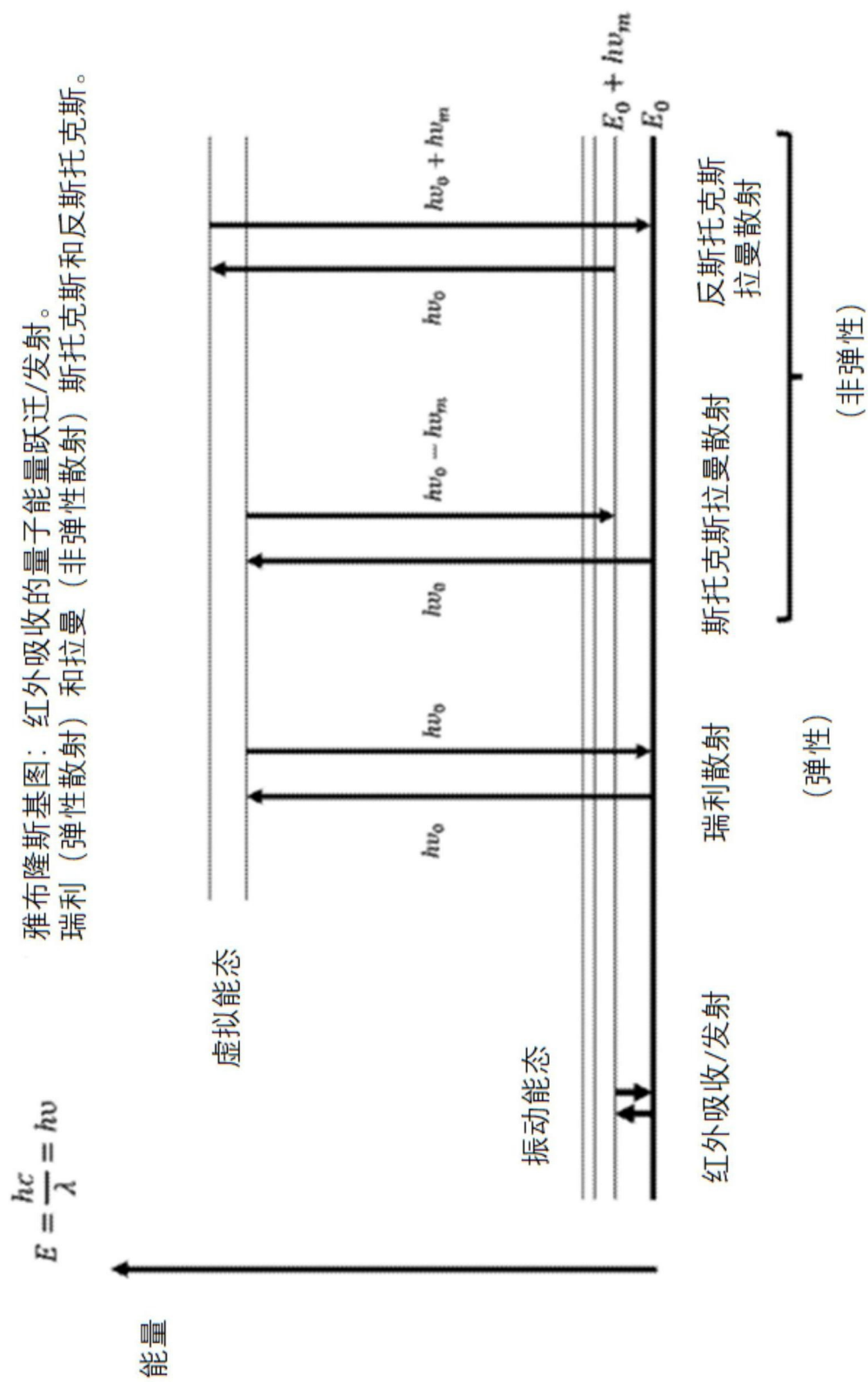


图1

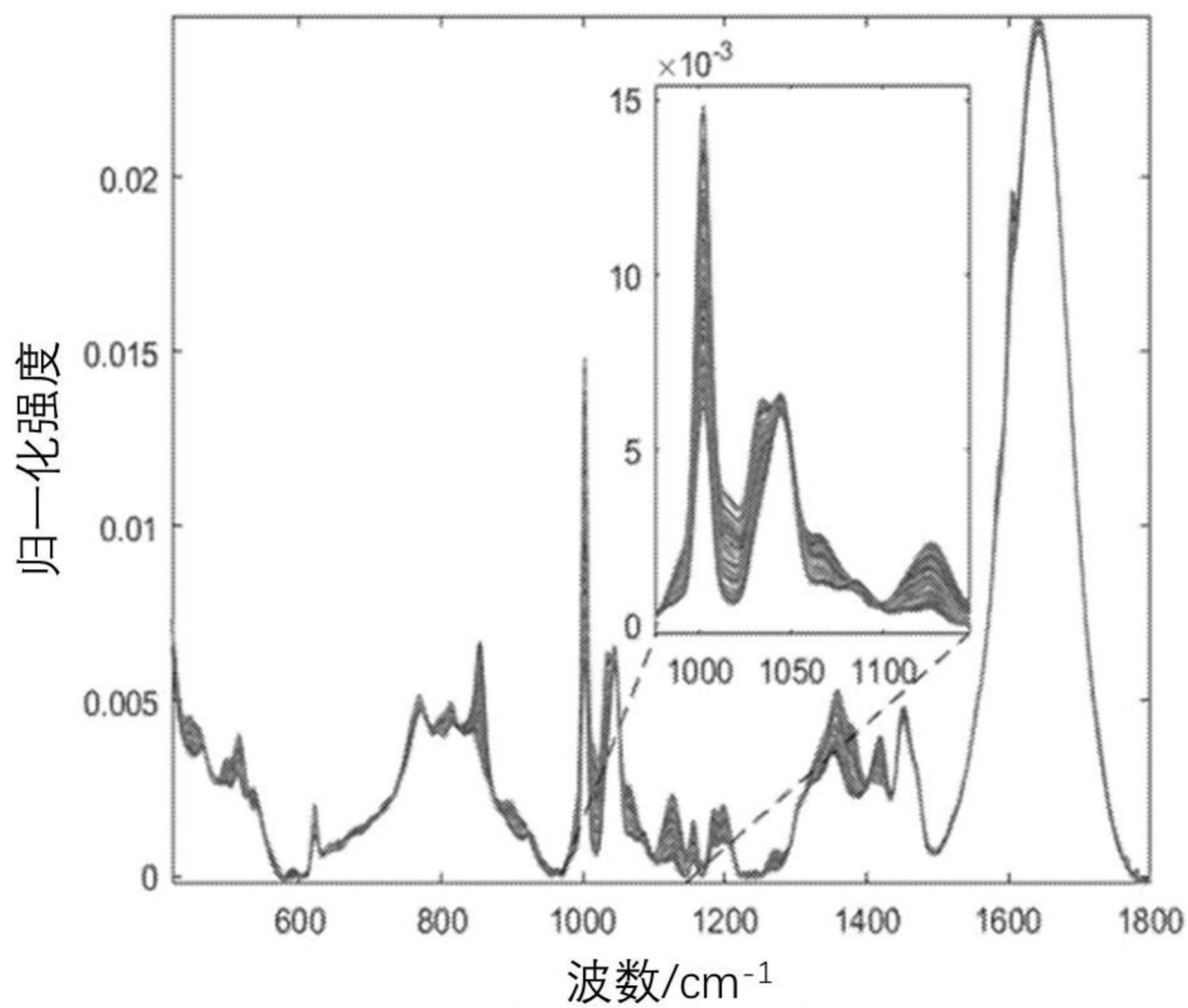


图2

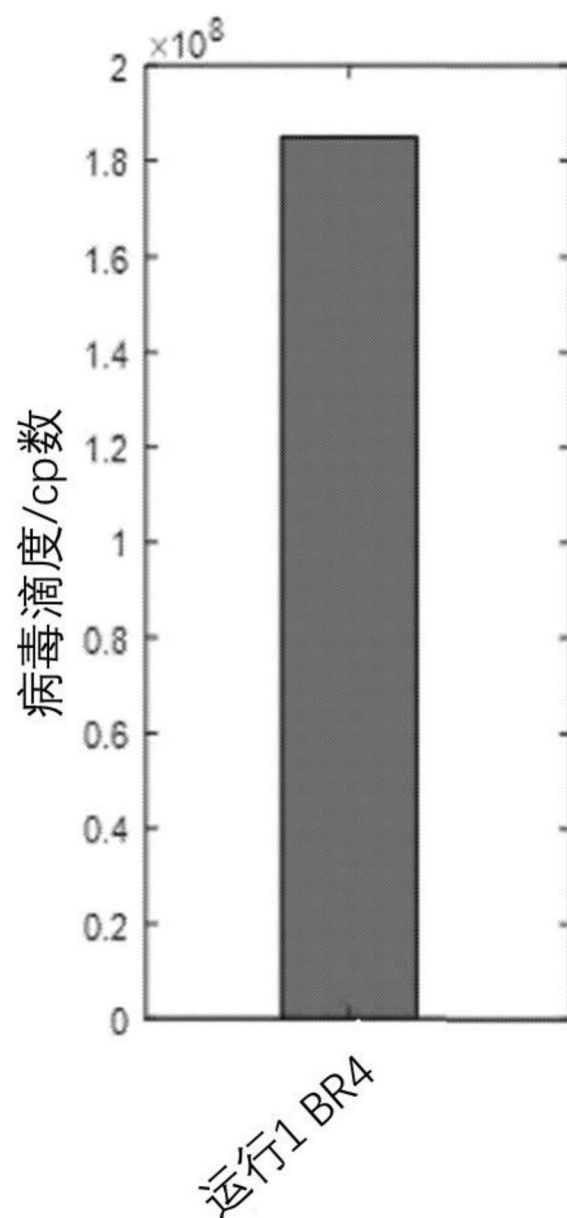


图3

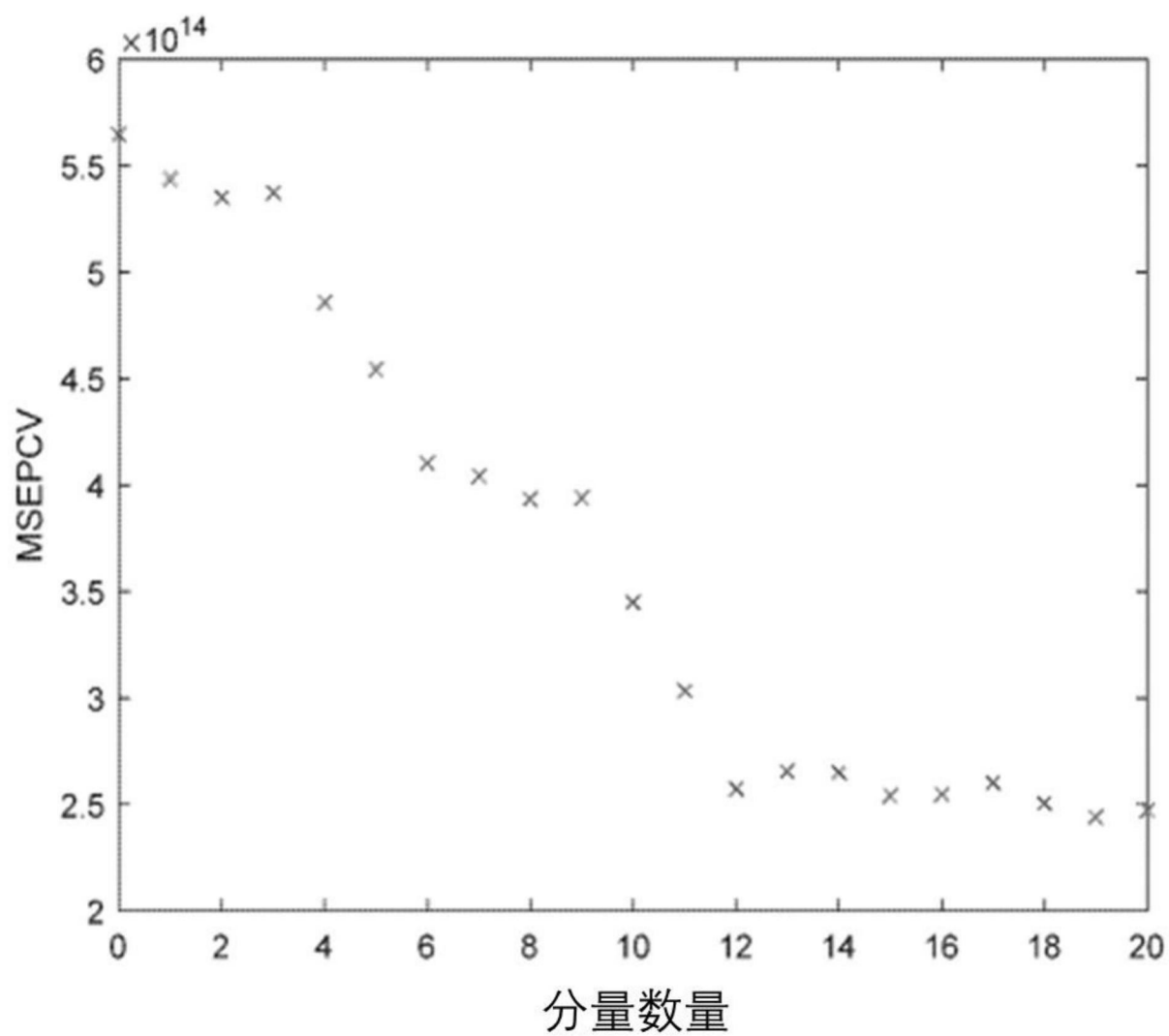


图4

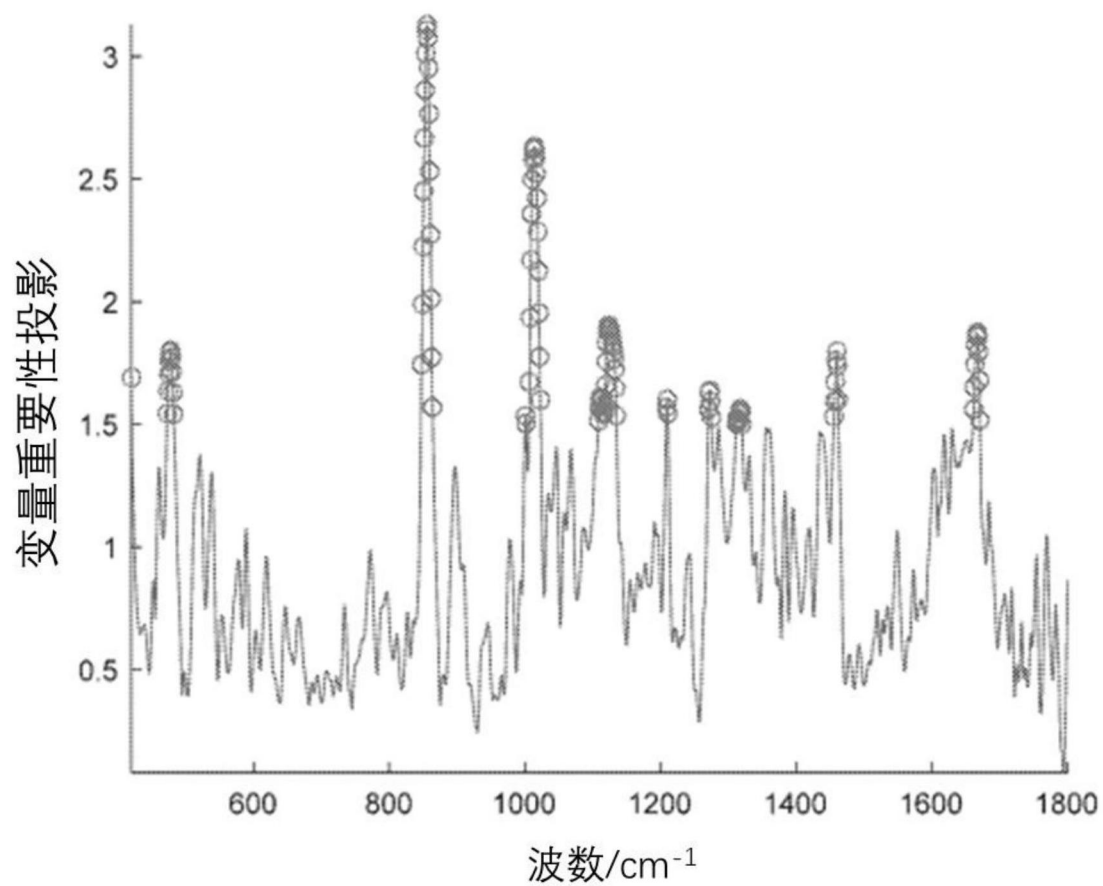


图5A

变量范围 /cm ⁻¹
420
473-481
848-863
1001
1007-1022
1109-1134
1209-1210
1271-1274
1312-1319
1456-1461
1662-1671

图5B

<u>波数范围/cm⁻¹</u>	<u>峰VIP</u>	<u>积分的VIP</u>	<u>积分的VIP占总 积分的VIP的%</u>
1109 - 1134	1.9030	44.7182	24
848 - 863	3.1279	40.5701	21
1007-1022	2.6319	36.7527	19
1662 - 1671	1.8745	17.3666	9
473 - 481	1.8006	15.3503	8
1312 - 1319	1.5638	12.2383	6
1456 - 1461	1.7977	10.0990	5
1271 - 1274	1.6370	6.4248	3
1209 - 1210	1.6015	3.1708	2
420	1.6907	1.6907	1
1001	1.5333	1.5333	1

图5C

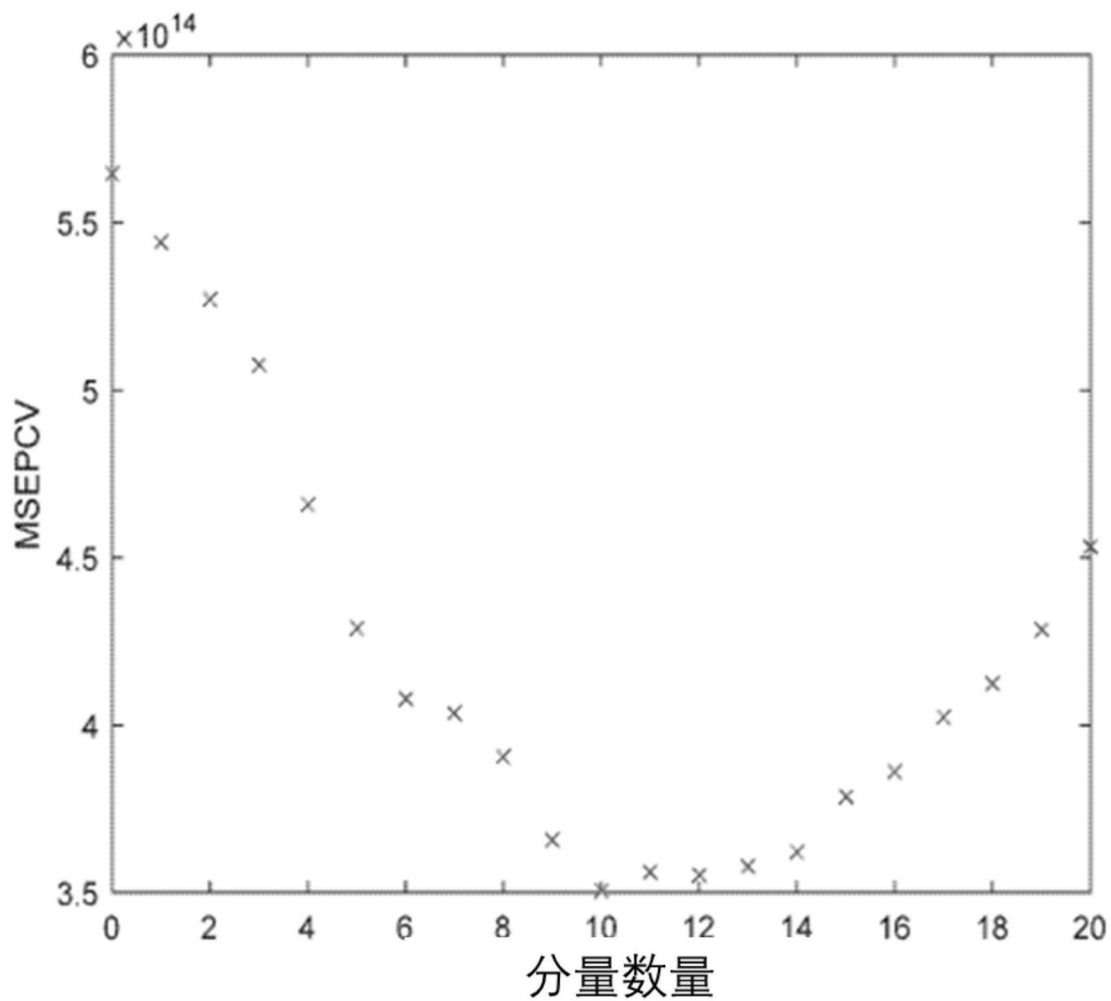


图6

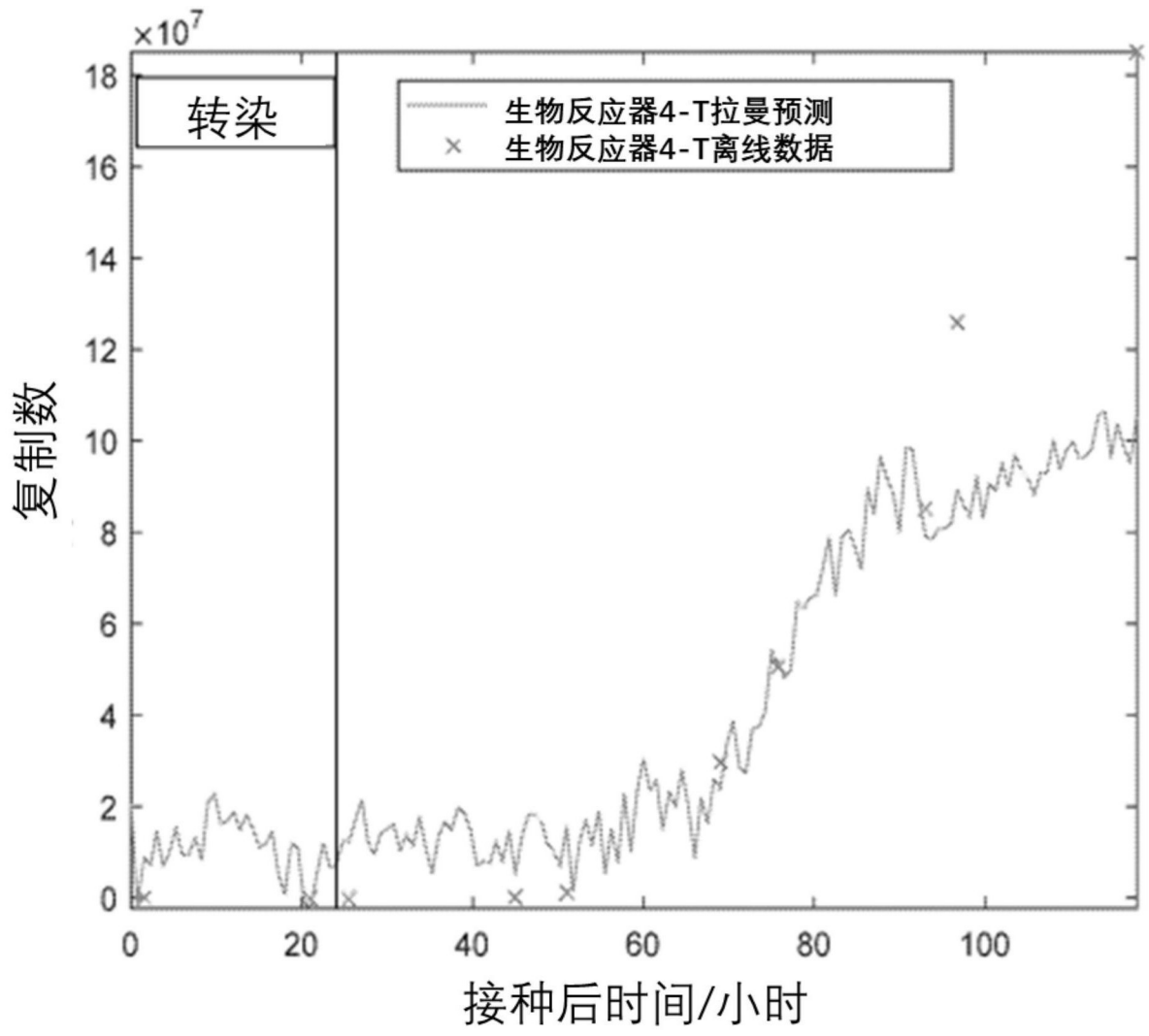


图7

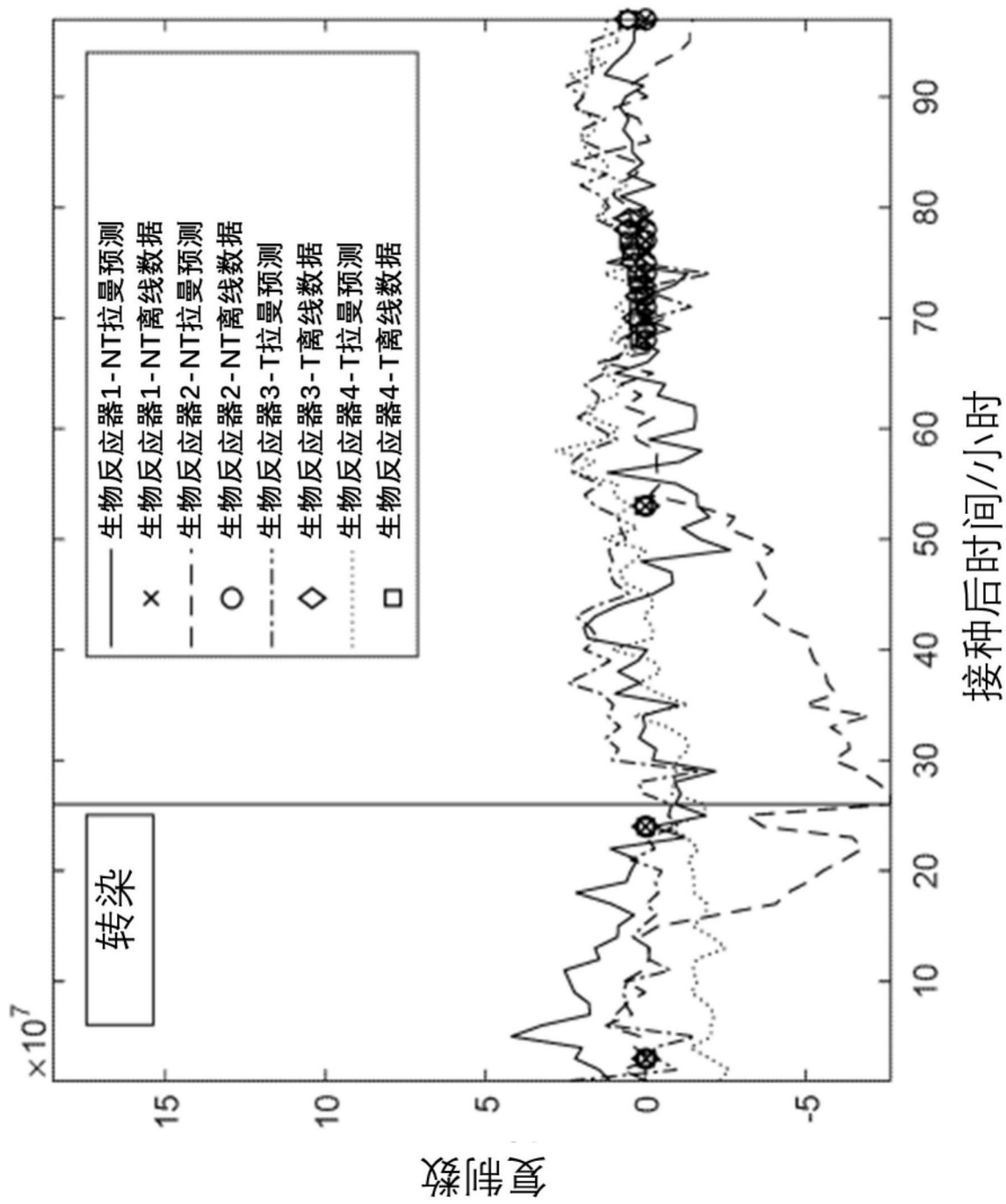


图8

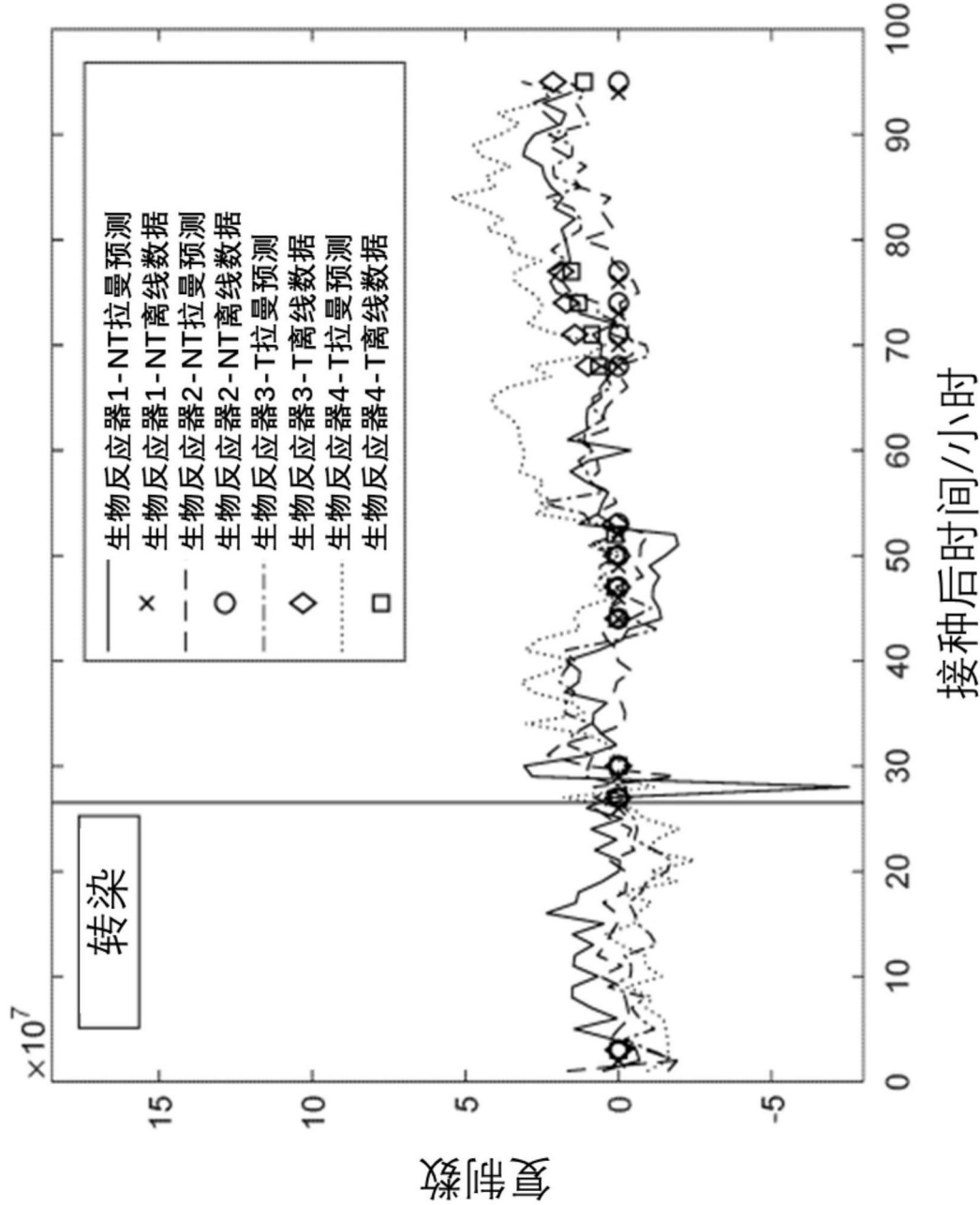


图9

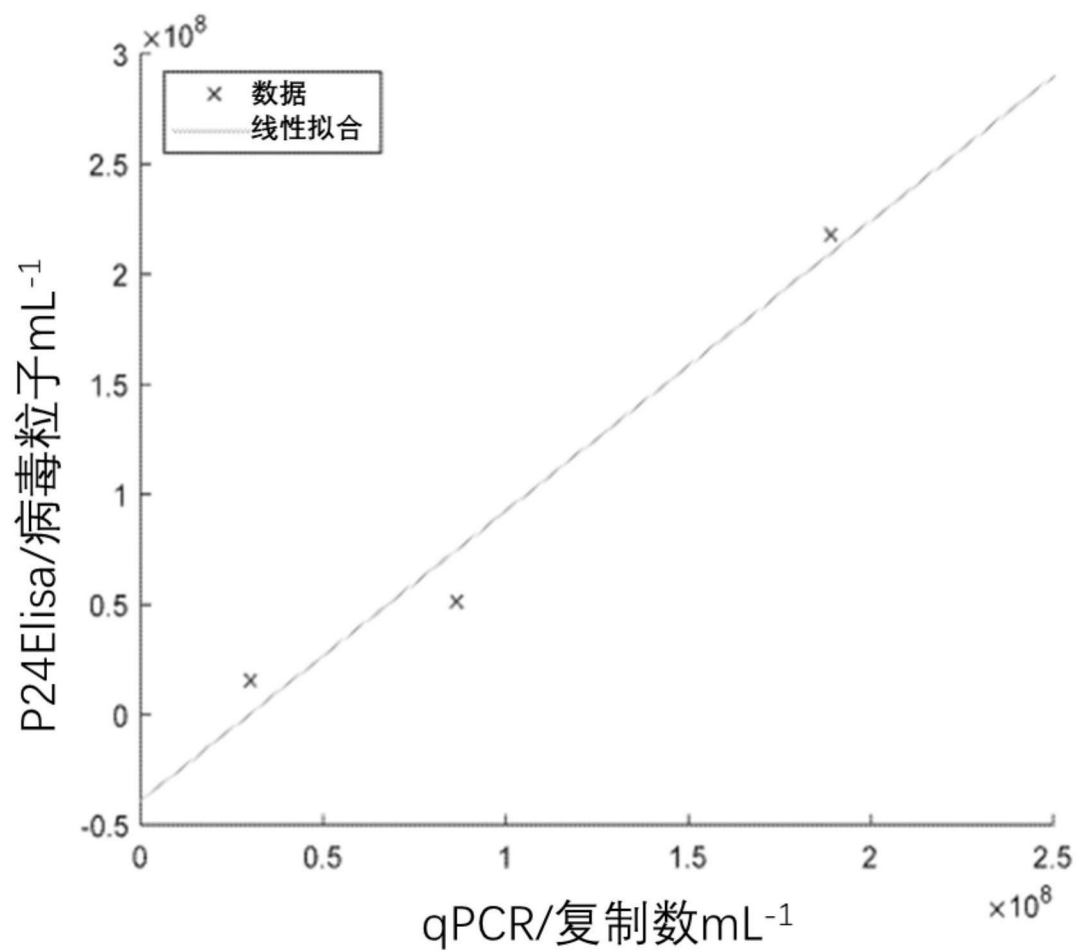


图10

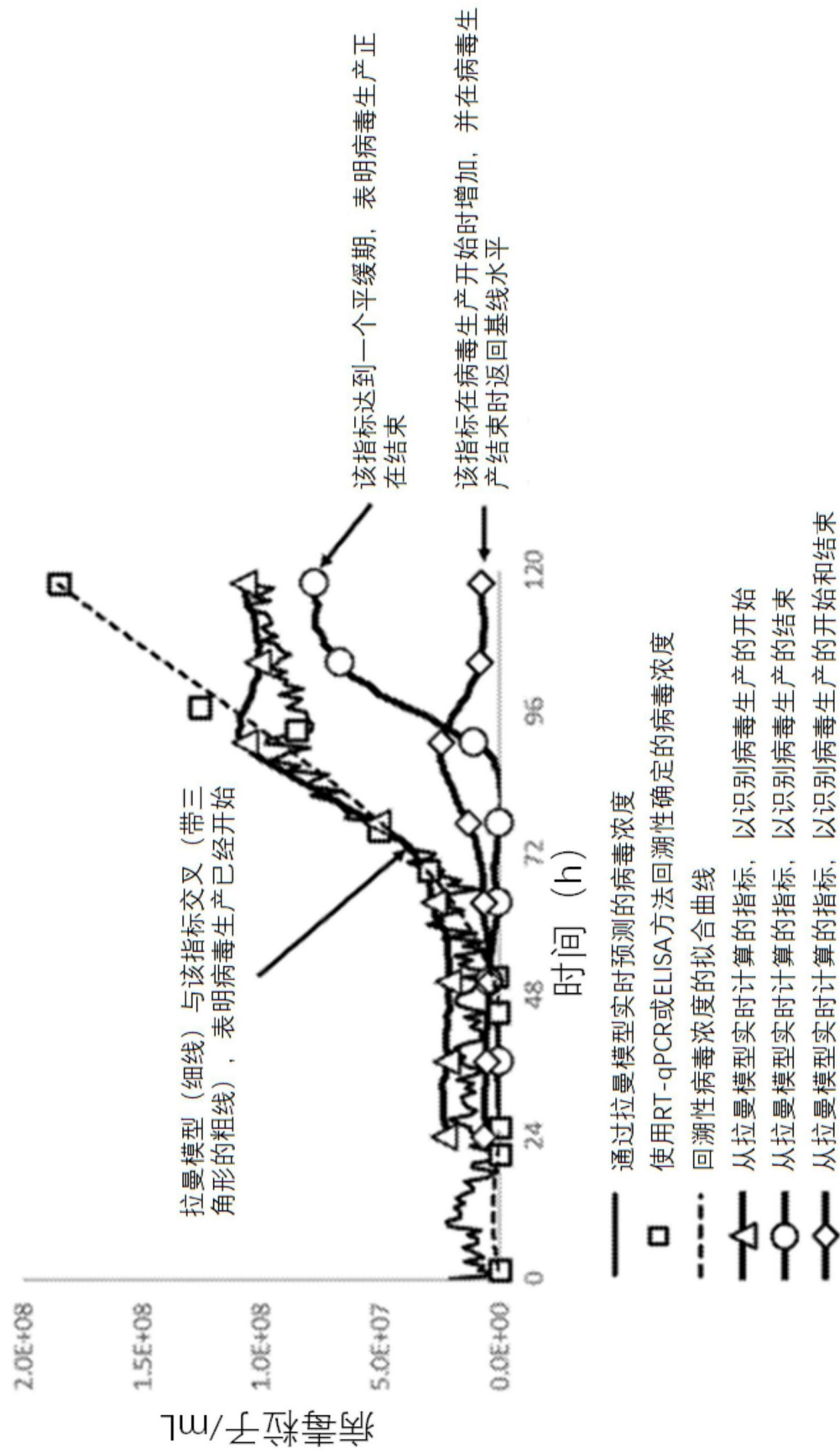


图11

回归 系数(β)
β_1
β_2
β_3
...
β_n

波数	w_1	w_2	w_3	...	w_n
拉曼光谱 强度	A_1	A_2	A_3	...	A_n

滴度预测 $= (A_1 \times \beta_1) + (A_2 \times \beta_2) + (A_3 \times \beta_3) + \dots (A_n \times \beta_n)$

图12

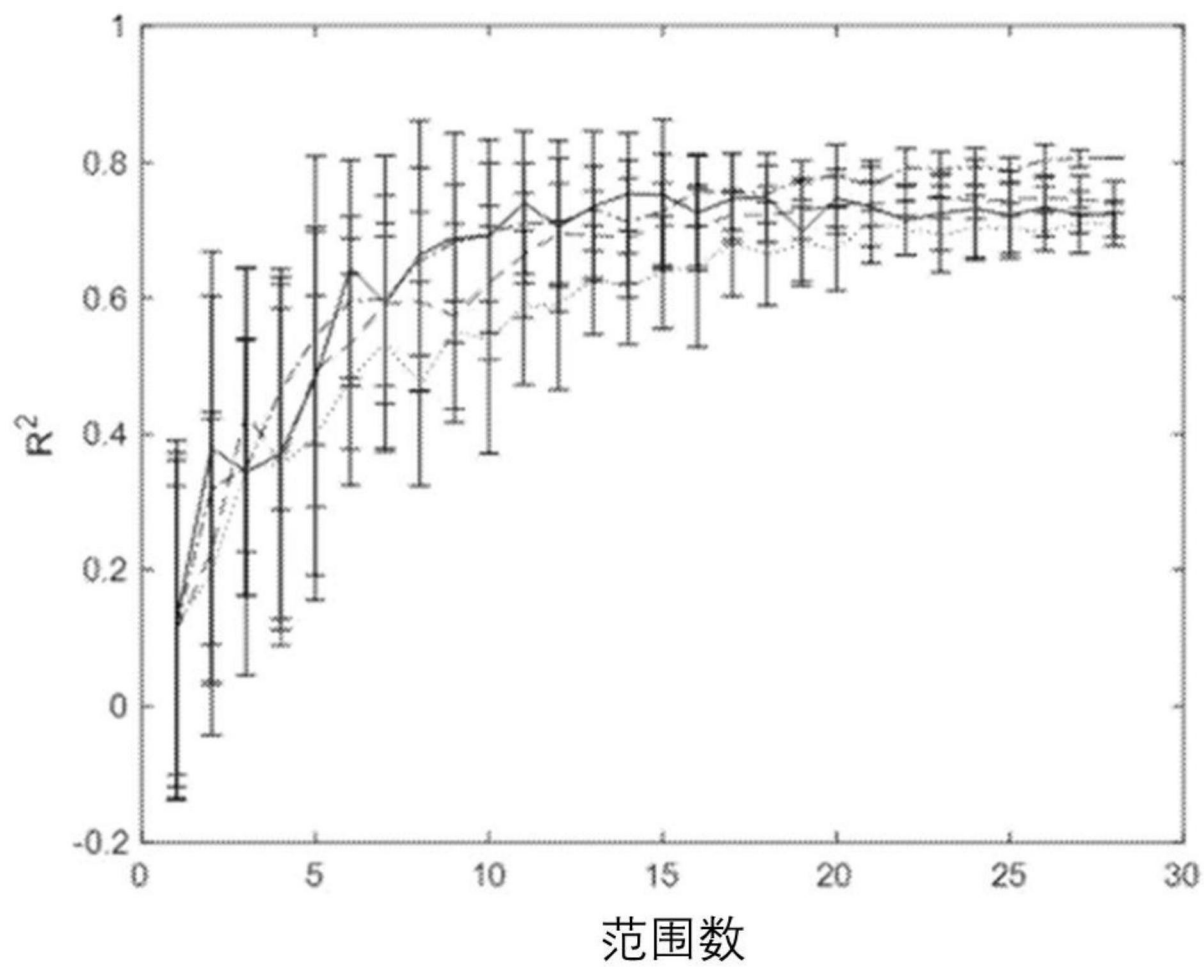


图13

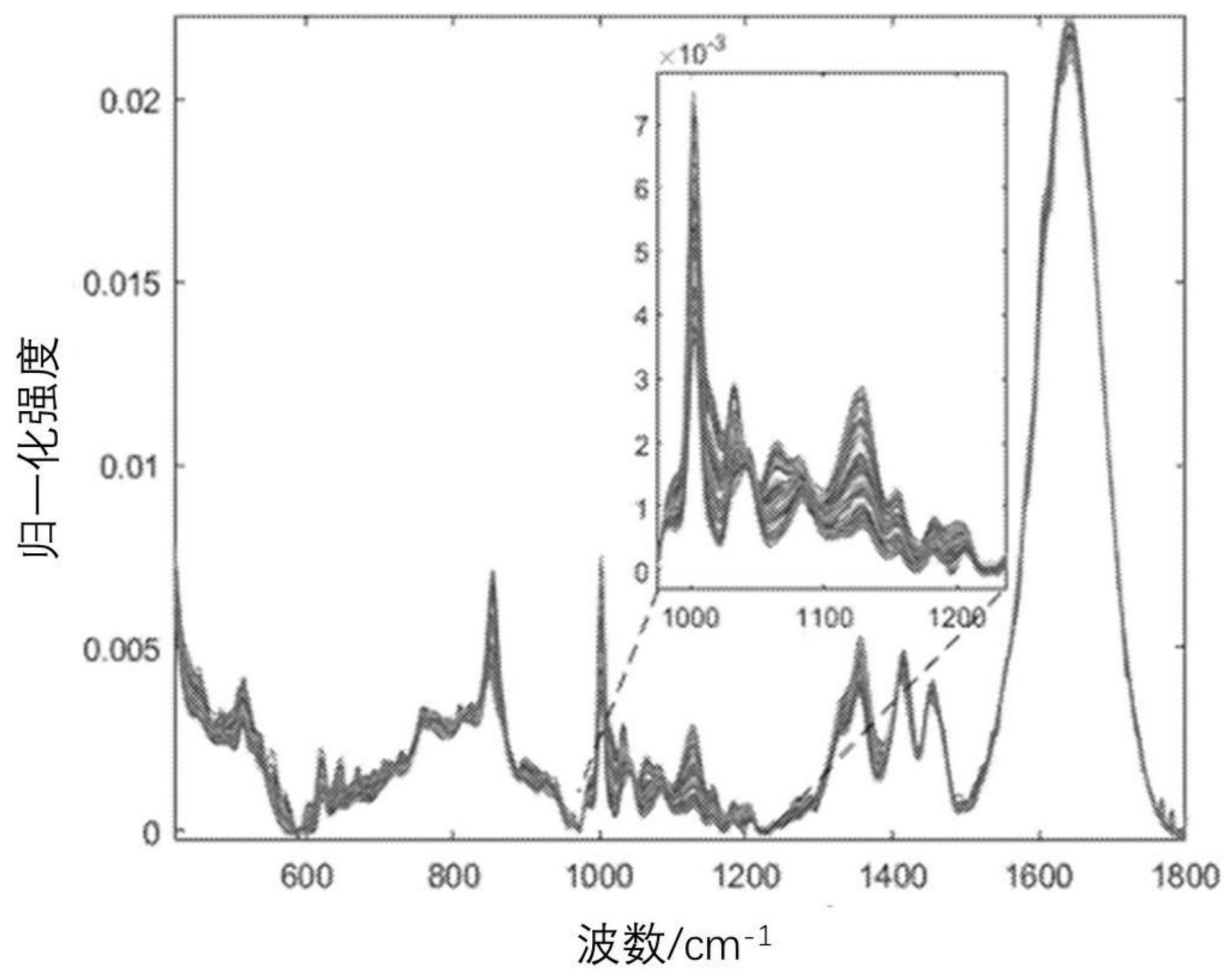


图14

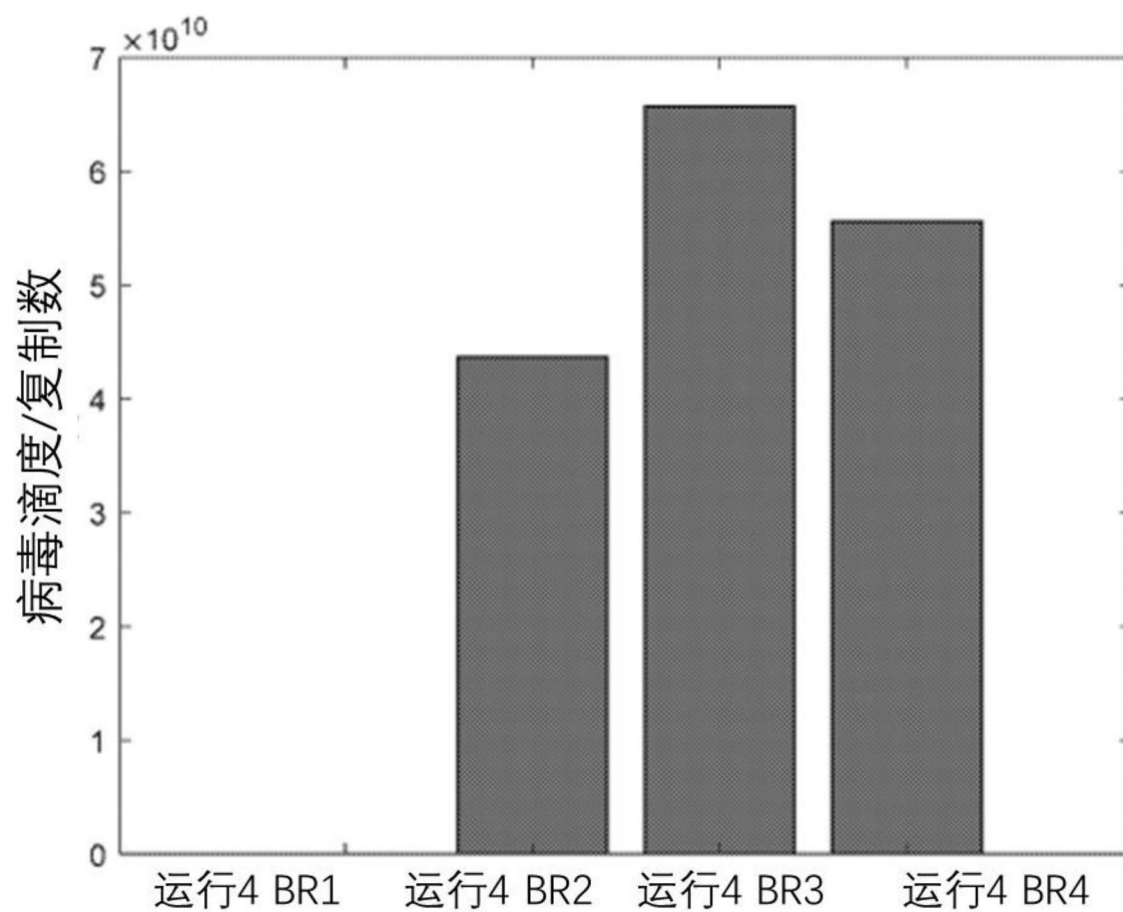


图15

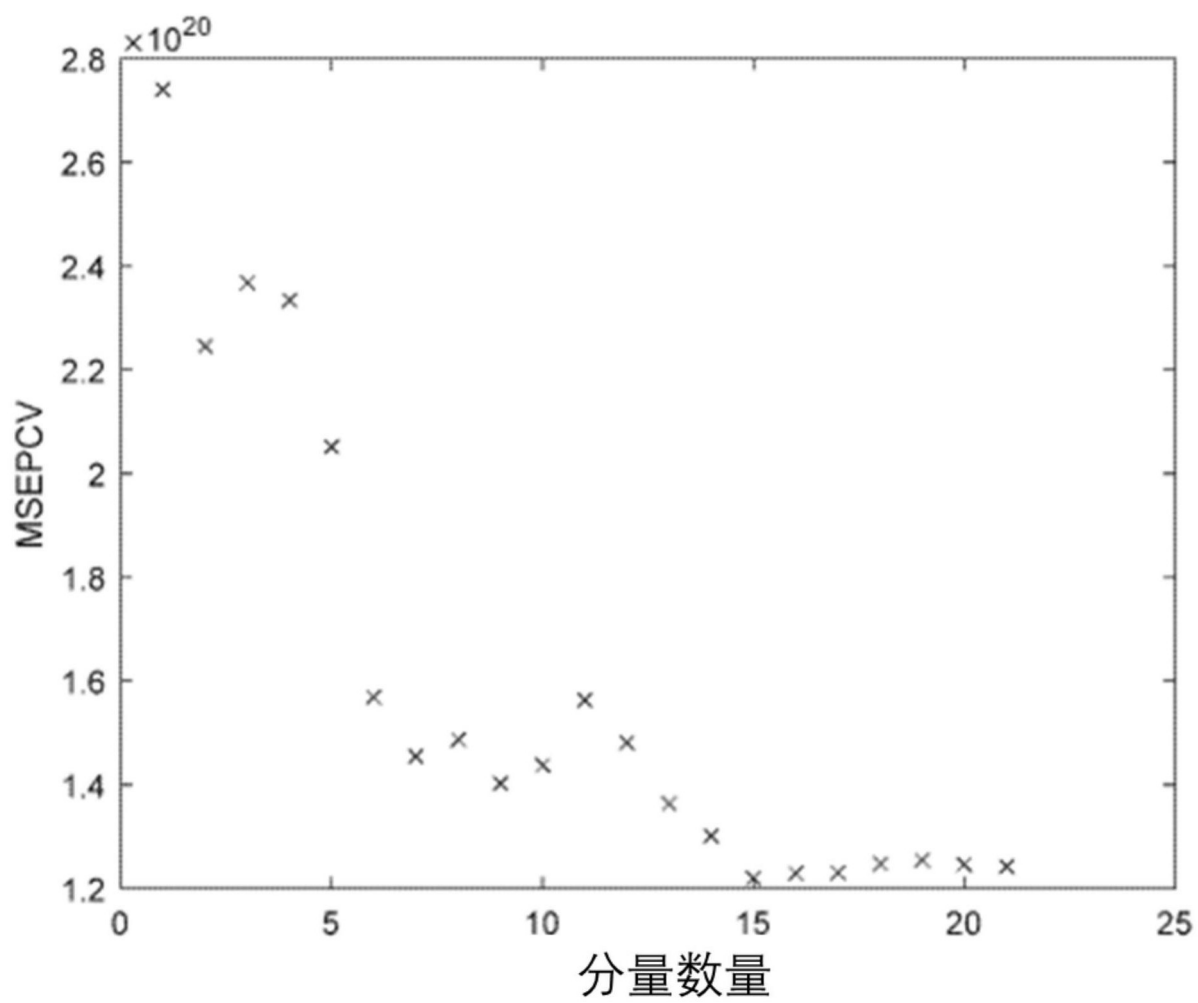


图16

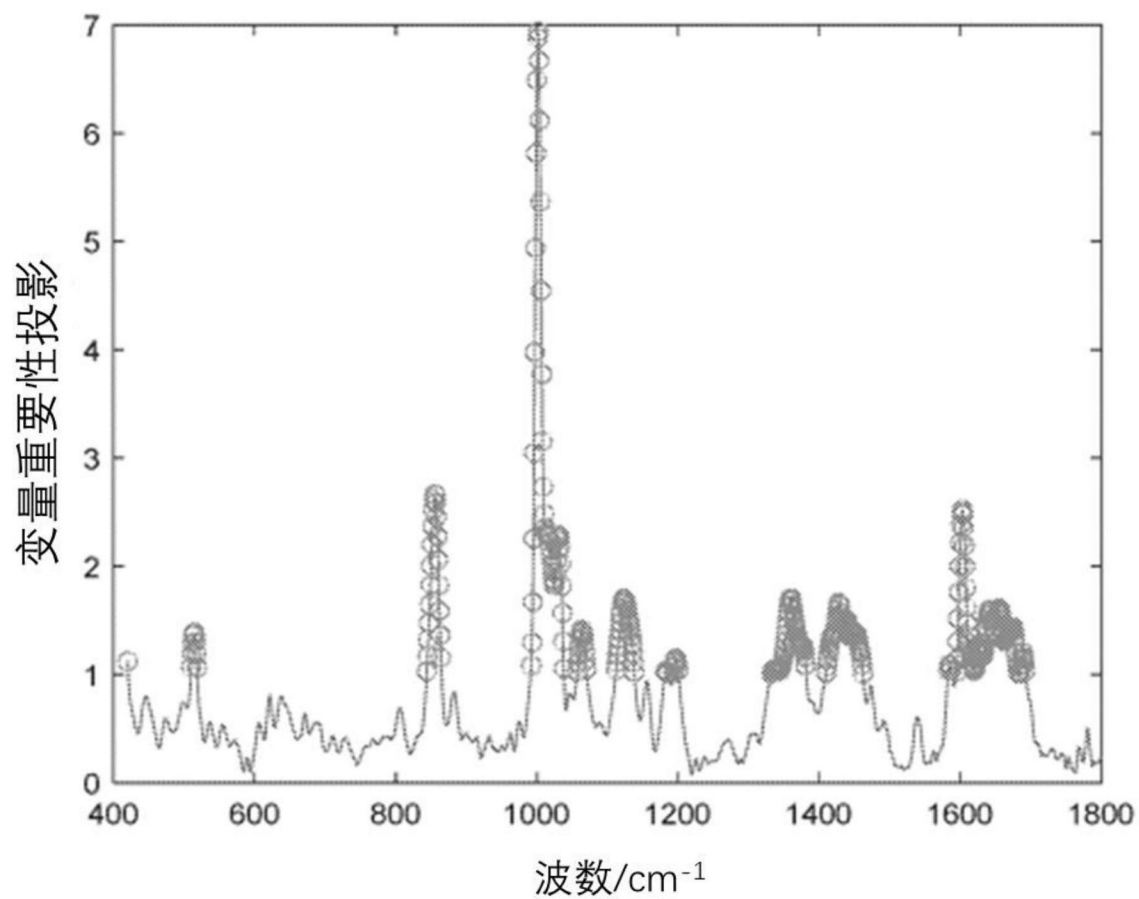


图17A

变量范围 /cm ⁻¹
420
510-517
844-863
992-1037
1057-1069
1112-1137
1182-1184
1193-1199
1333-1380
1410-1461
1583-1586
1594-1692

图17B

波数范围/cm ⁻¹	峰VIP	积分的VIP	积分的VIP占总 积分的VIP的%
1594-1692	2.5216	138.2853	26.80
992-1037	6.9381	133.8646	25.94
1410-1461	1.6650	70.7048	13.70
1333-1380	1.7030	60.6275	11.75
844-863	2.6652	38.4608	7.45
1112-1137	1.7035	36.5806	7.09
1057-1069	1.4161	15.3906	2.98
510-517	1.3884	9.0333	1.75
1193-1199	1.1516	6.6665	1.29
1583-1586	1.0933	3.2301	0.63
1182-1184	1.0359	2.0611	0.40
420	1.1229	1.1229	0.22

图17C

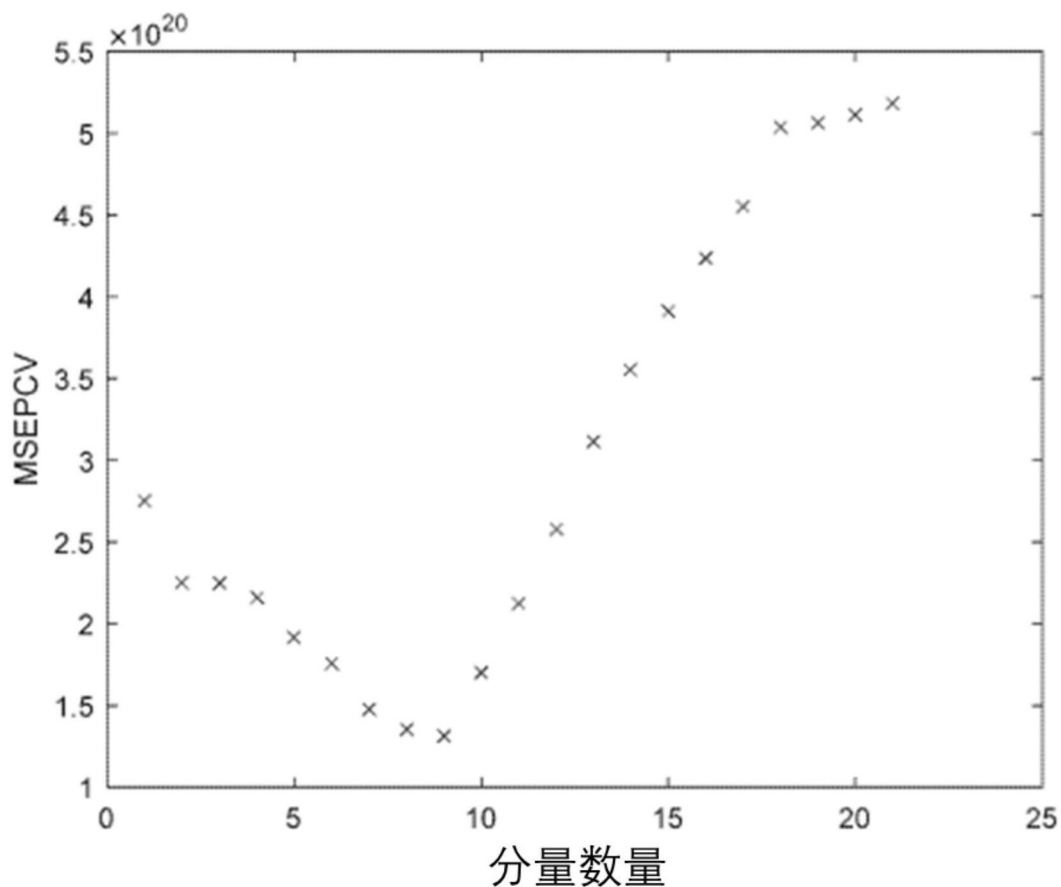


图18

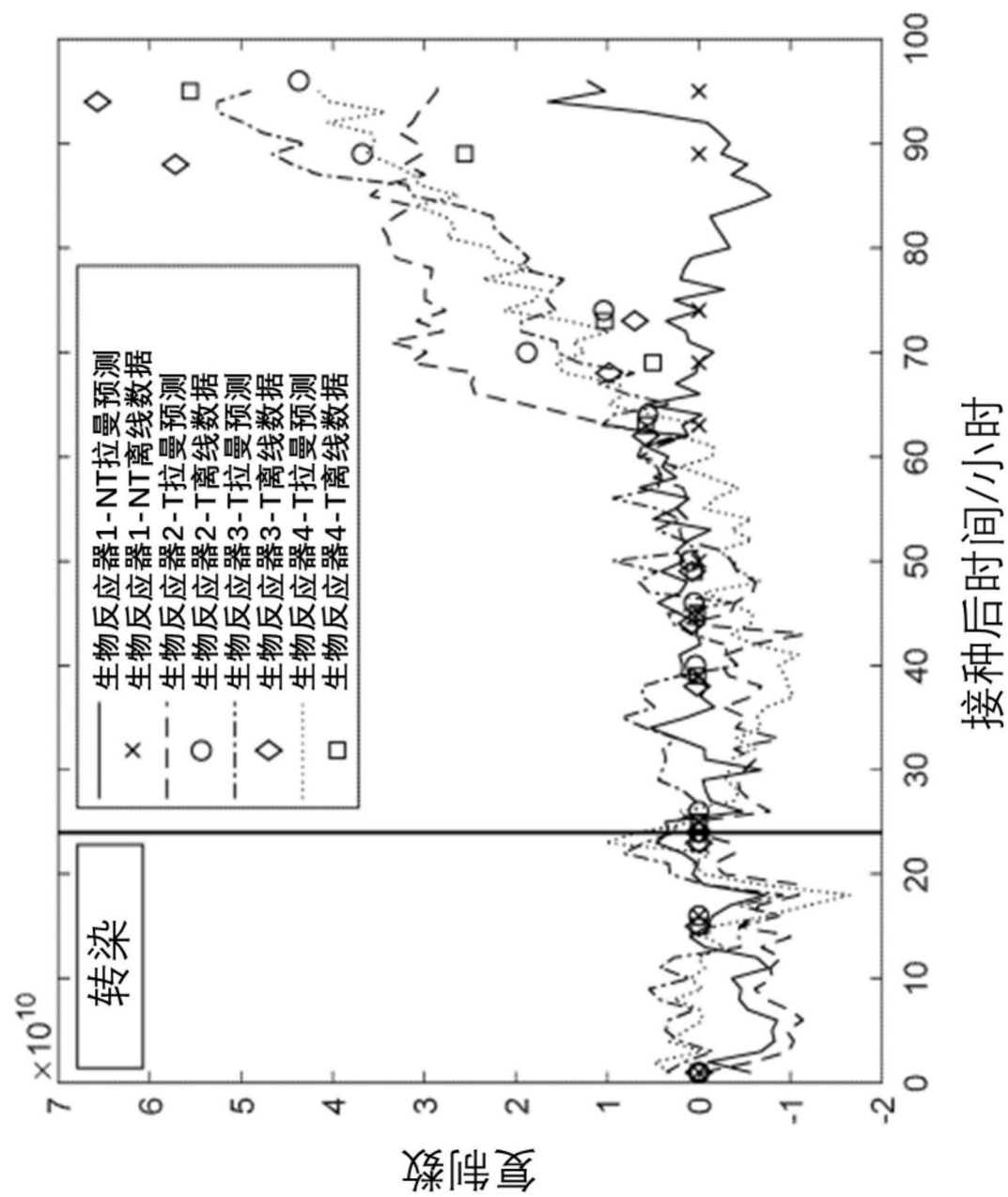


图19

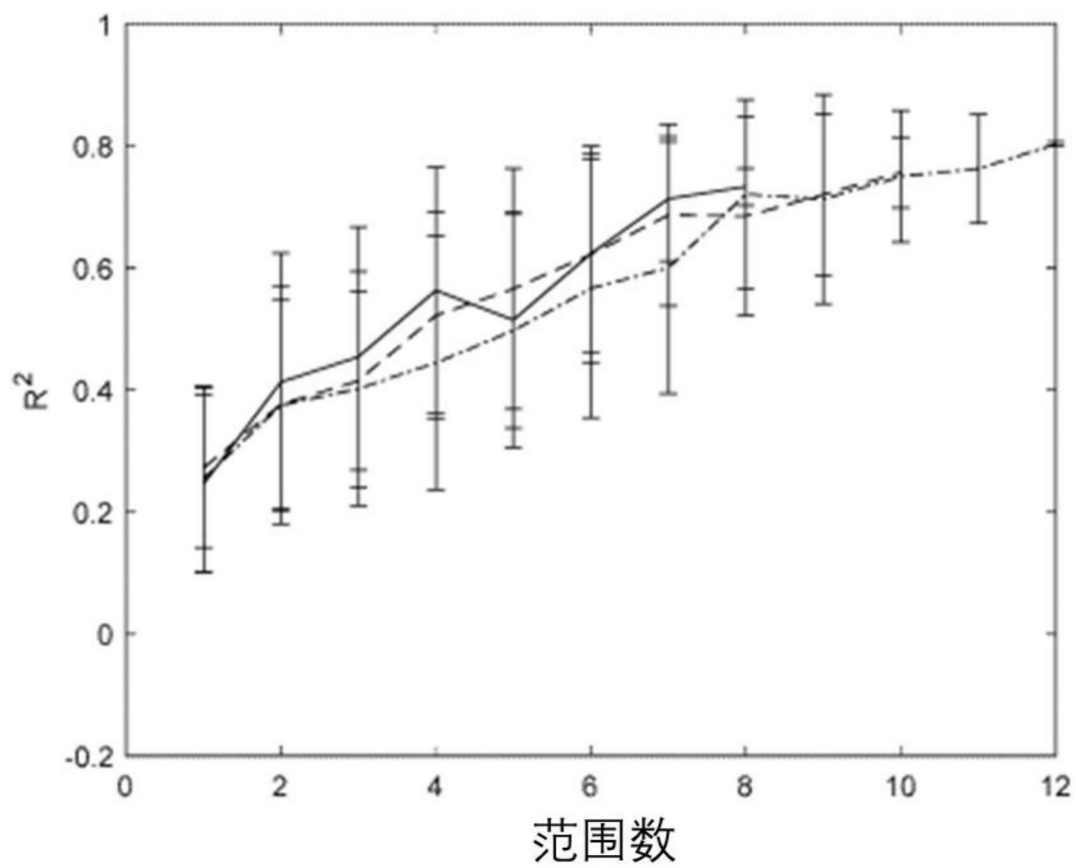


图20

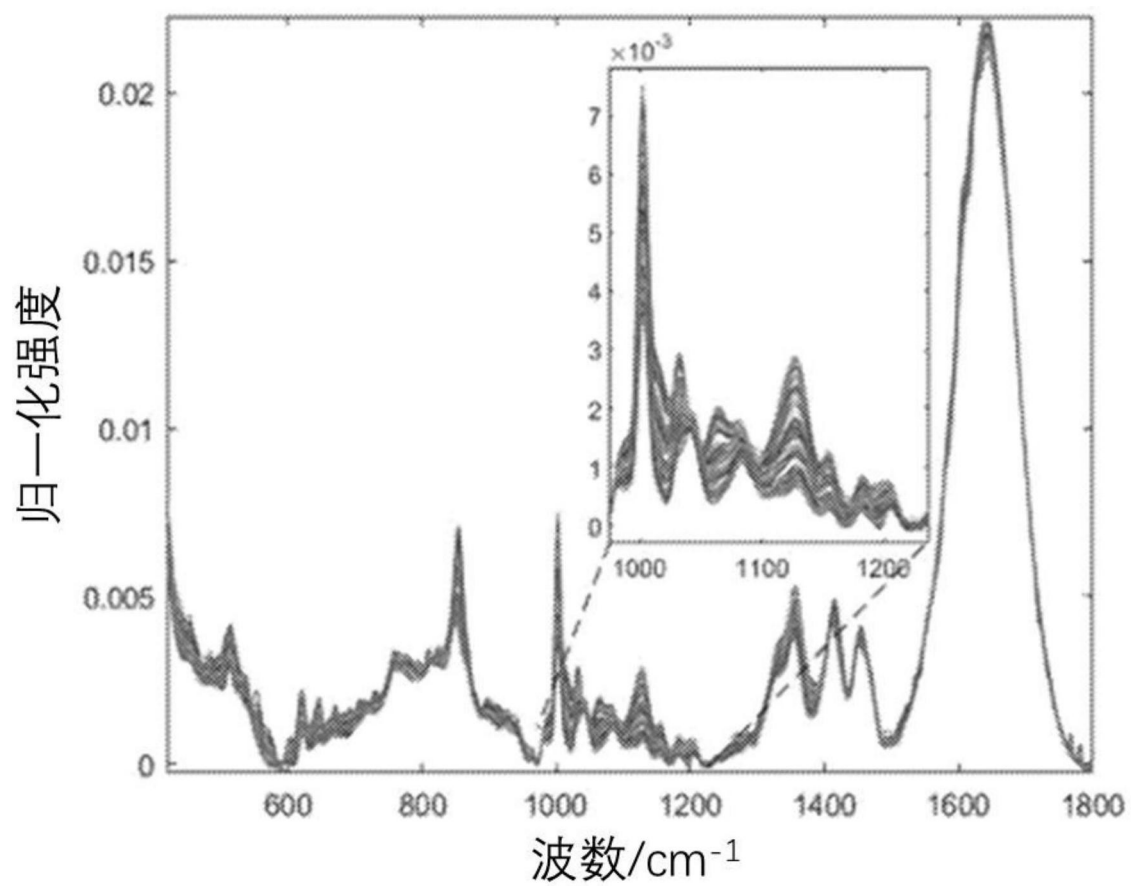


图21

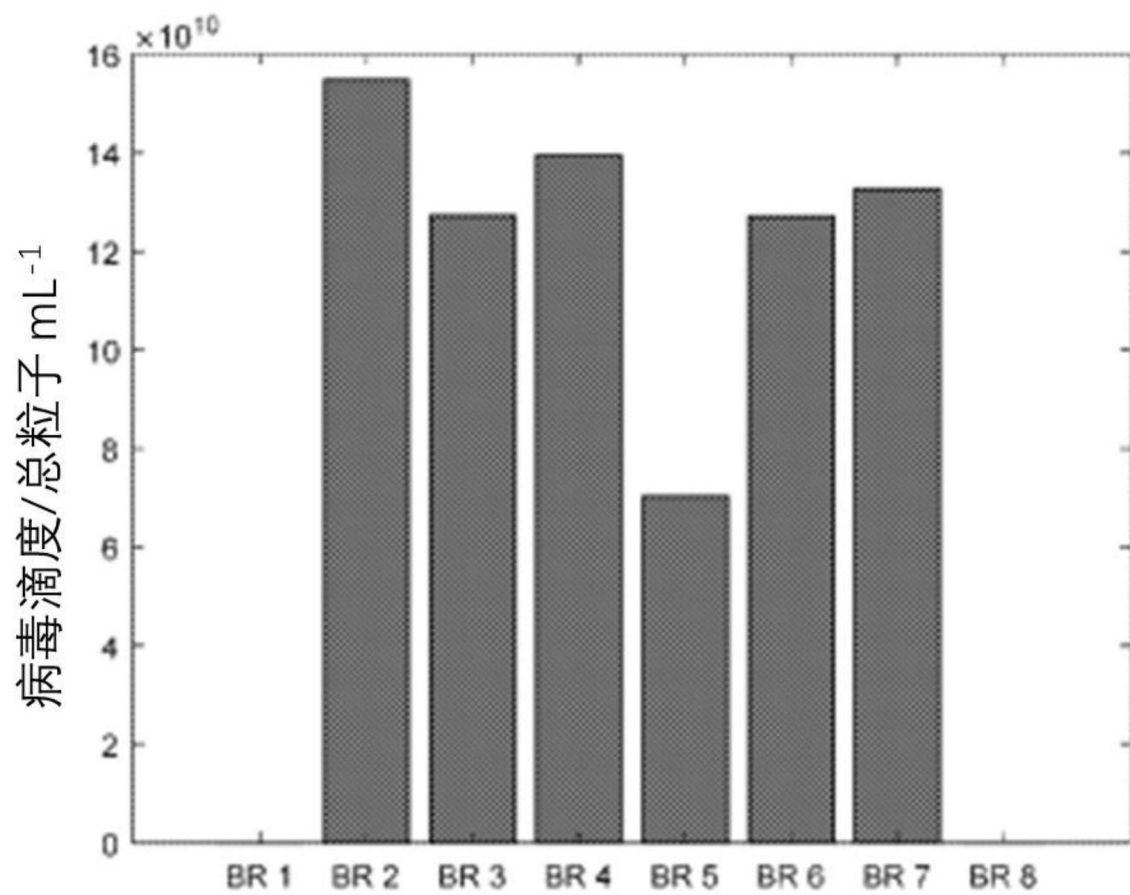


图22

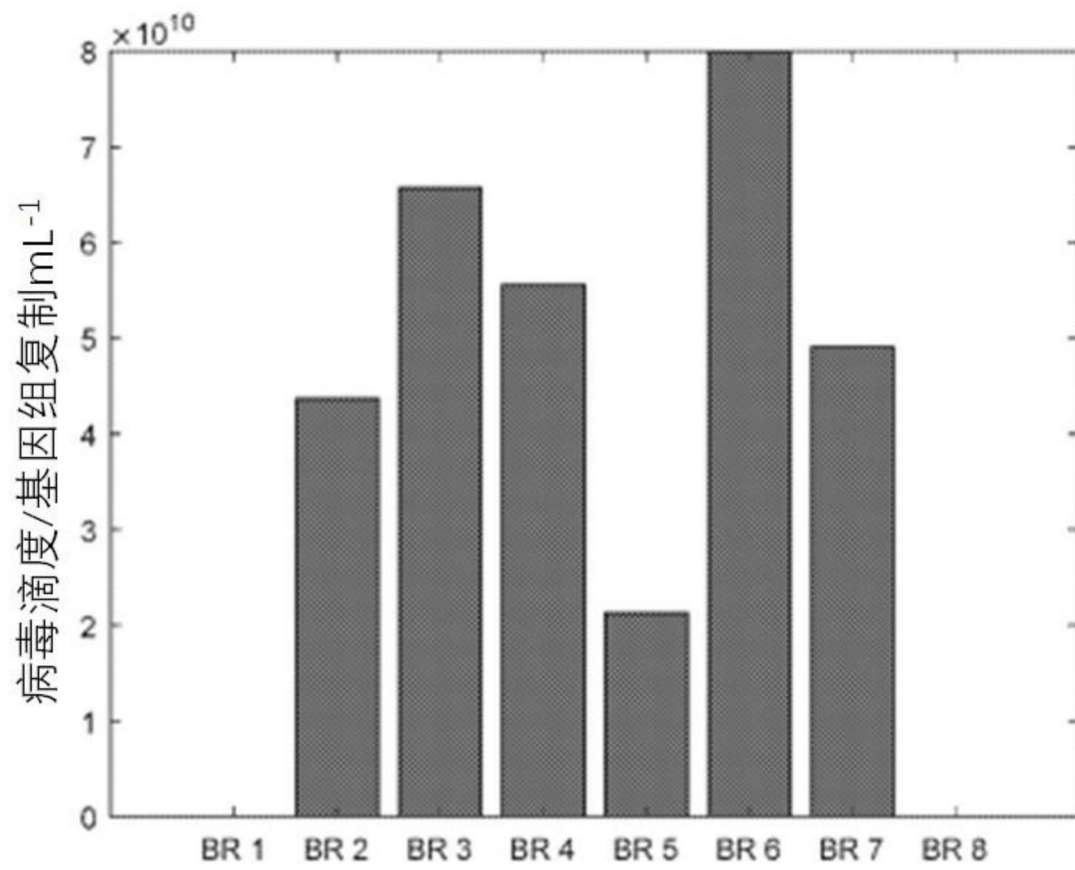


图23

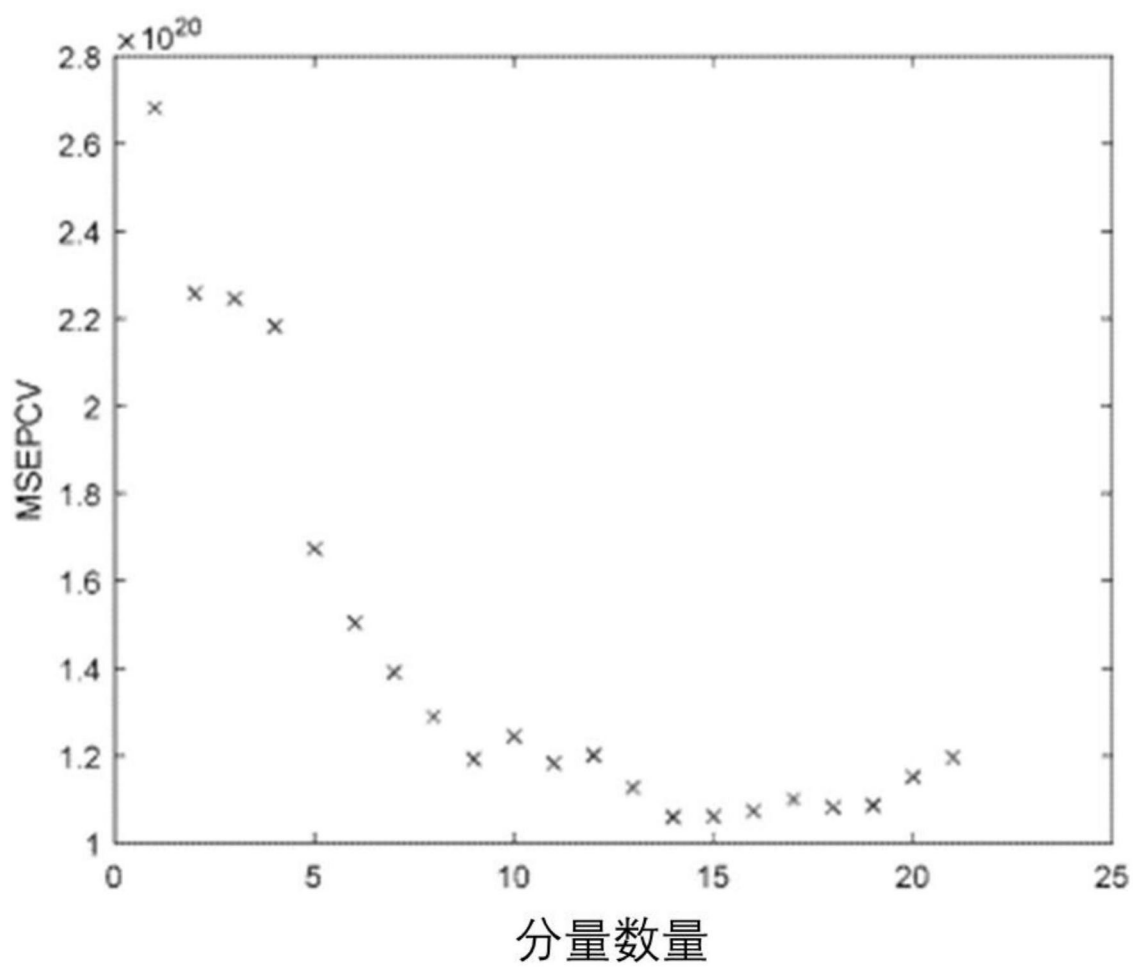


图24

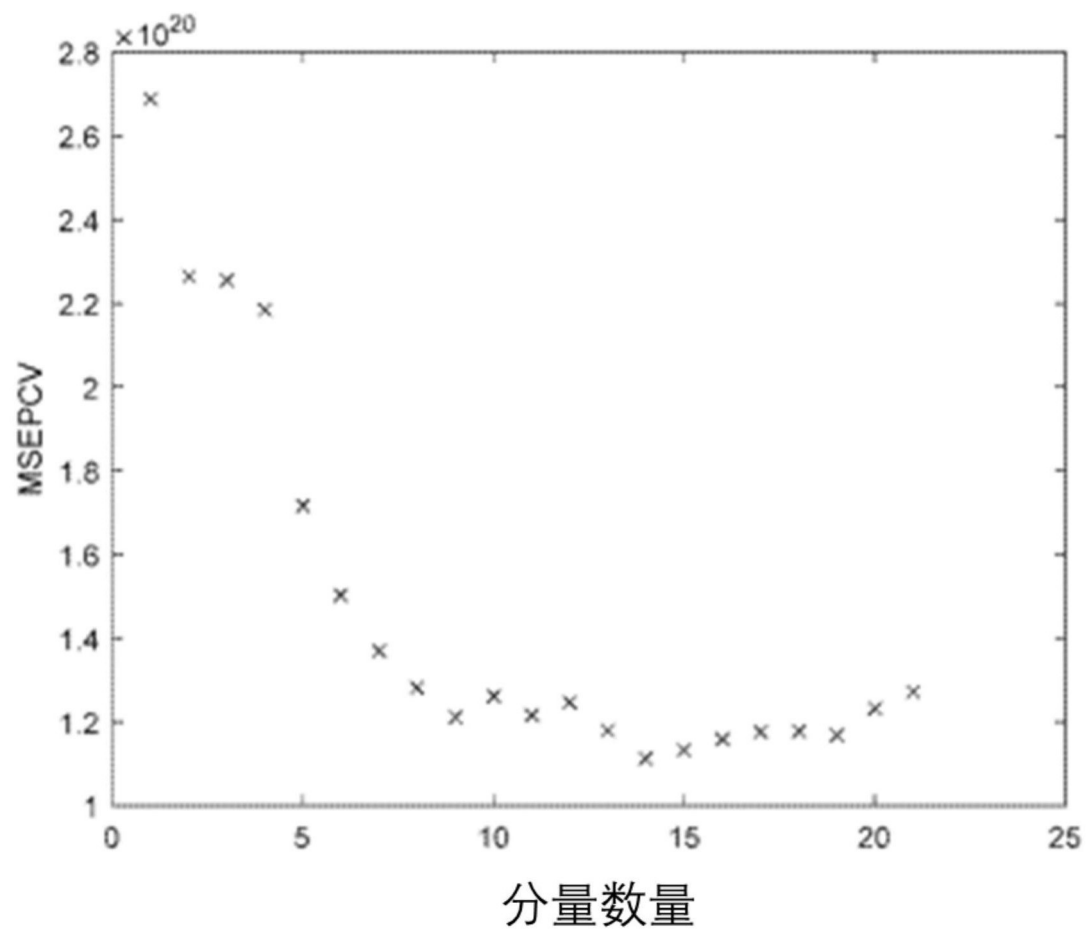


图25

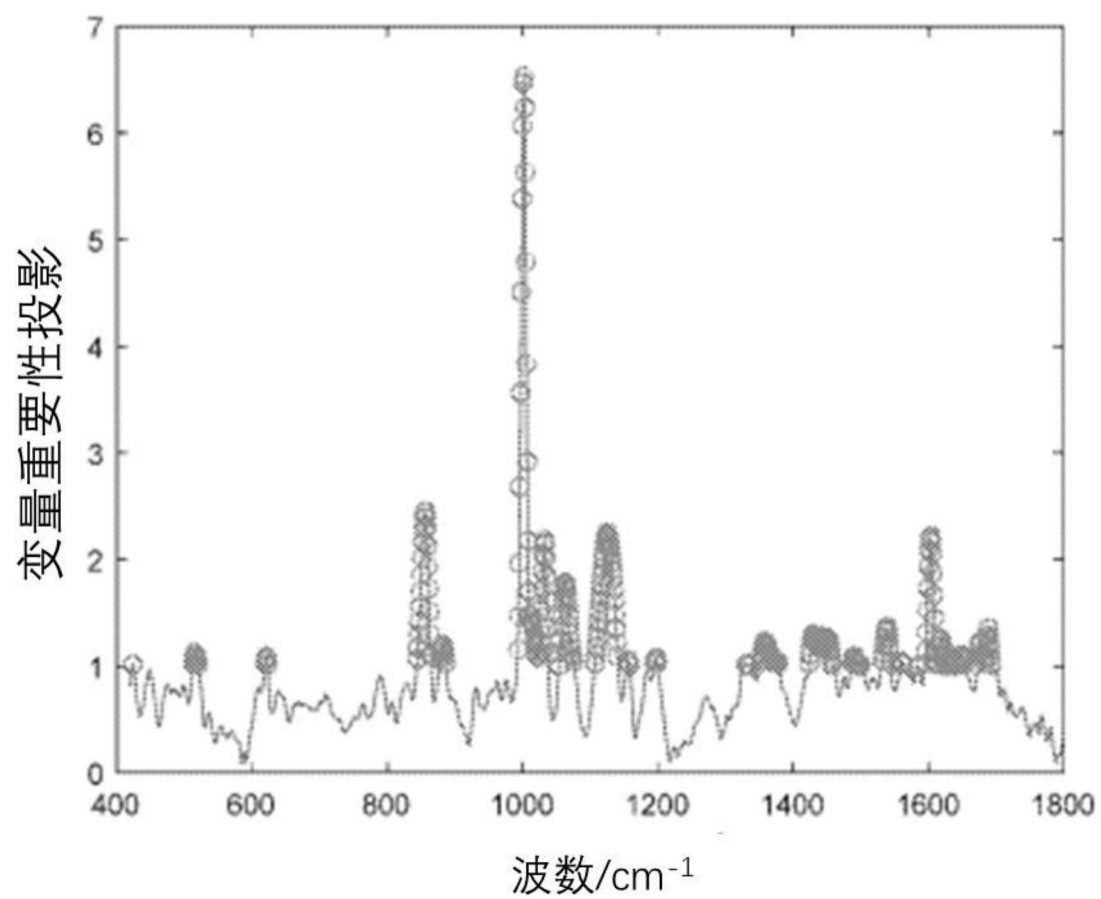


图26A

变量范围 /cm ⁻¹
423
513-519
619-6223
844-863
877-885
993-1037
1054-1074
1107-1139
1154-1157
1195-1198
1330-1335
1352-1377
1424-1457
1487-1493
1496-1500
1532-1543
1559-1562
1583
1595-1610
1614-1622
1626-1634
1646-1651
1673-1678
1684-1693

图26B

<u>波数范围/cm⁻¹</u>	<u>峰VIP</u>	<u>积分的VIP</u>	<u>积分的VIP占总 积分的VIP的%</u>
993-1037	6.5282	107.9935	25.4655
1107-1139	2.2502	57.6637	13.5975
1424-1457	1.2990	40.0296	9.4392
844-863	2.4516	35.9060	8.4669
1054-1074	1.7864	29.1864	6.8823
1352-1377	1.2330	27.7755	6.5496
1595-1610	2.2190	25.6717	6.0535
1532-1543	1.3651	13.8427	3.2642
1684-1693	1.3534	10.8253	2.5527
1614-1622	1.2634	9.4275	2.2231
877-885	1.2026	9.2663	2.1850
1626-1634	1.0945	8.4667	1.9965
513-519	1.1255	6.5514	1.5449
1487-1493	1.0950	6.4289	1.5160
1673-1678	1.2333	5.8411	1.3774
1646-1651	1.1042	5.3947	1.2721
1330-1335	1.0255	5.0958	1.2016
1496-1500	1.0188	4.0542	0.9560
619-6223	1.0892	3.2173	0.7587
1195-1198	1.0766	3.1871	0.7515
1154-1157	1.0585	3.1354	0.7394
1559-1562	1.0462	3.1098	0.7333
423	1.0063	1.0063	0.2373
1583	1.0002	1.0002	0.2358

图26C

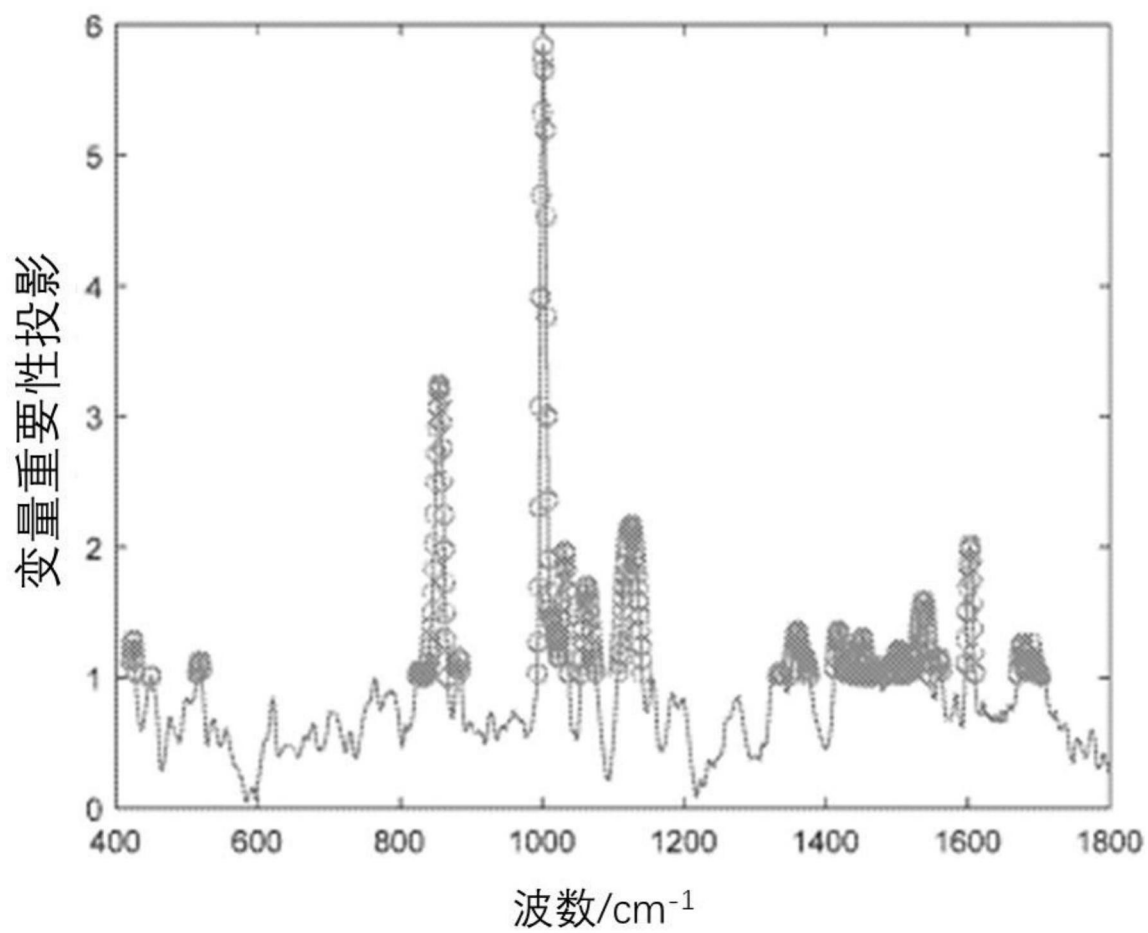


图27A

变量范围 /cm ⁻¹
420-426
447-448
541-519
824-833
838-866
879-884
933-1037
1055-1074
1107-1140
1332-1338
1350-1376
1412-1429
1438-1441
1445-1464
1471-1475
1486-1506
1513-1546
1558-1562
1597-1609
1671-1703

图27B

<u>波数范围/cm⁻¹</u>	<u>峰VIP</u>	<u>积分的VIP</u>	<u>积分的VIP占总 积分的VIP的%</u>
933-1037	5.8401	101.7239	20.9756
838-866	3.2433	58.5054	12.0639
1107-1140	2.1703	57.9222	11.9436
1513-1546	1.5896	42.063	8.6734
1671-1703	1.2635	35.7815	7.3782
1350-1376	1.3605	31.1909	6.4316
1055-1074	1.6992	26.8308	5.5325
1486-1506	1.2119	22.1012	4.5573
1445-1464	1.3060	21.37	4.4065
1412-1429	1.3611	20.4891	4.2249
1597-1609	2.0128	20.0685	4.1381
824-833	1.0571	9.3968	1.9376
420-426	1.2872	7.3876	1.5233
1332-1338	1.0403	6.1995	1.2783
879-884	1.1546	5.5705	1.1486
541-519	1.1249	5.4956	1.1332
1558-1562	1.1539	4.5103	0.93
1471-1475	1.0868	4.2639	0.8792
1438-1441	1.0348	3.0875	0.6366
447-448	1.0116	1.0059	0.2074

图27C

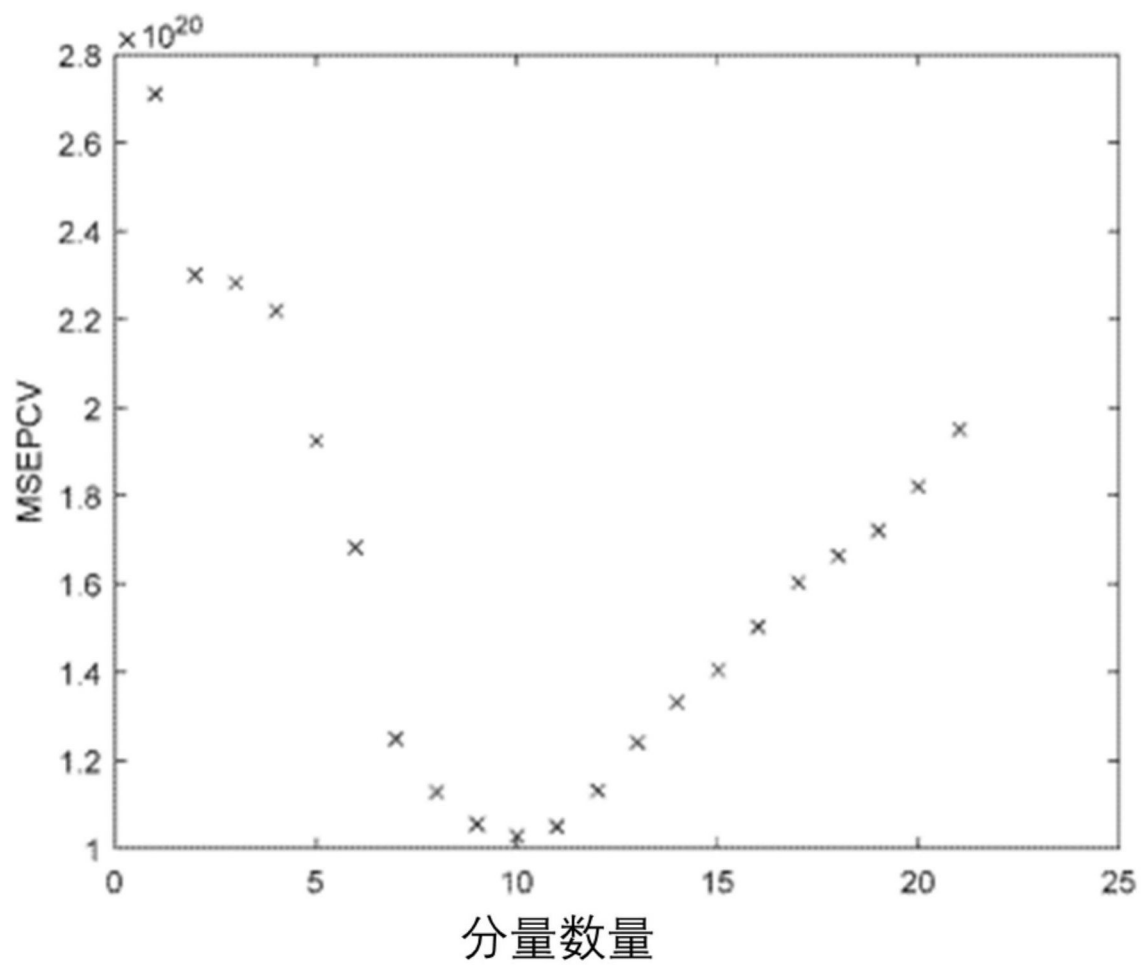


图28

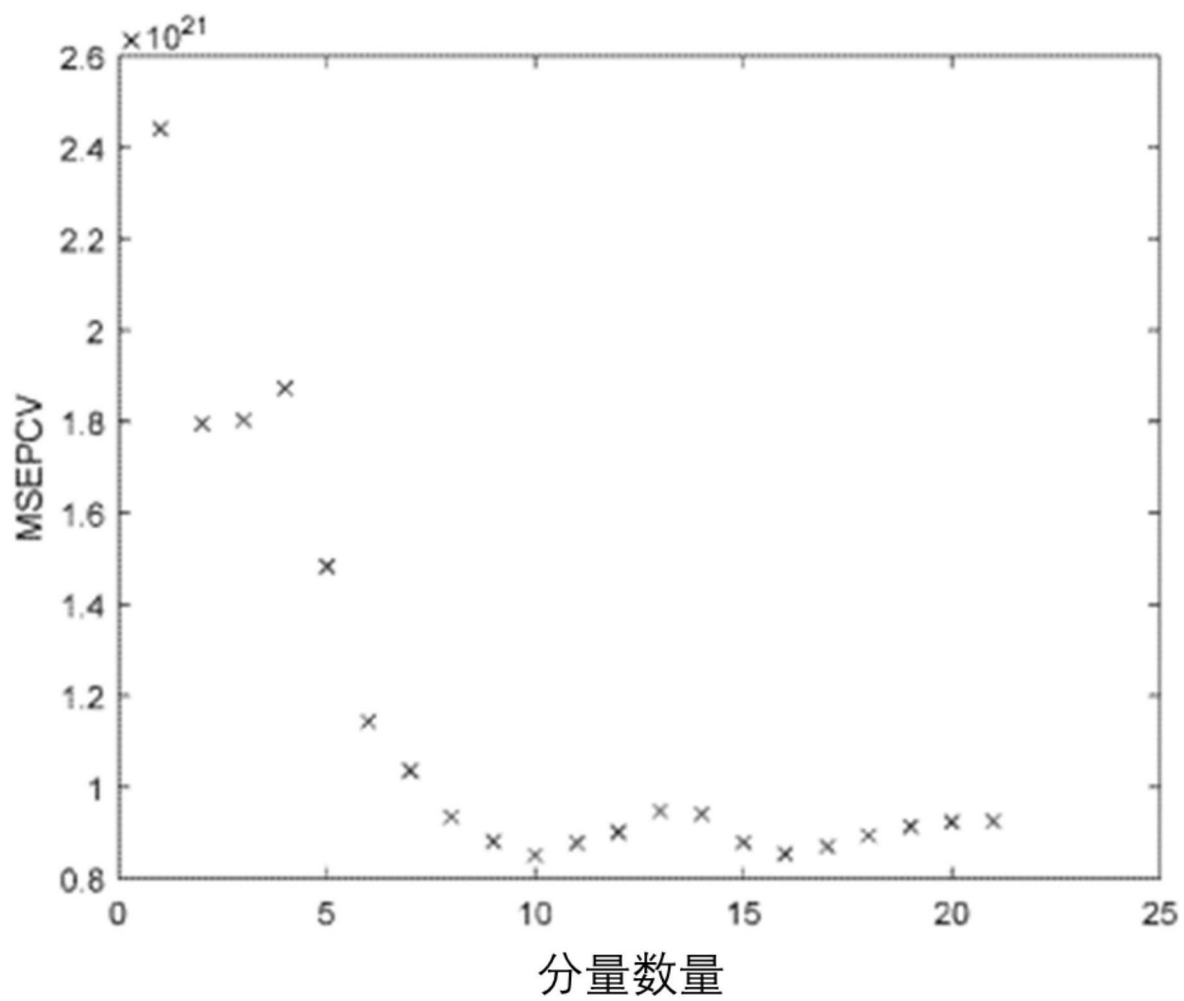


图29

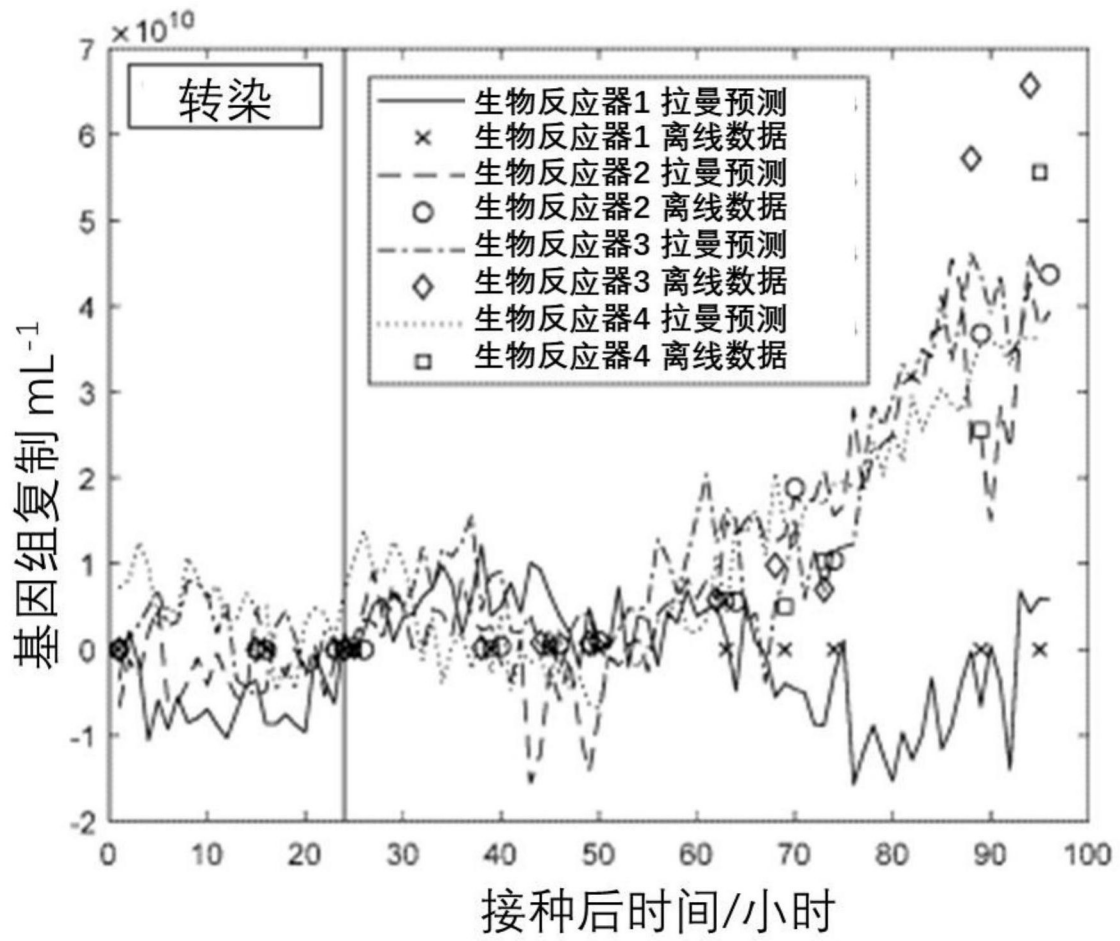


图30

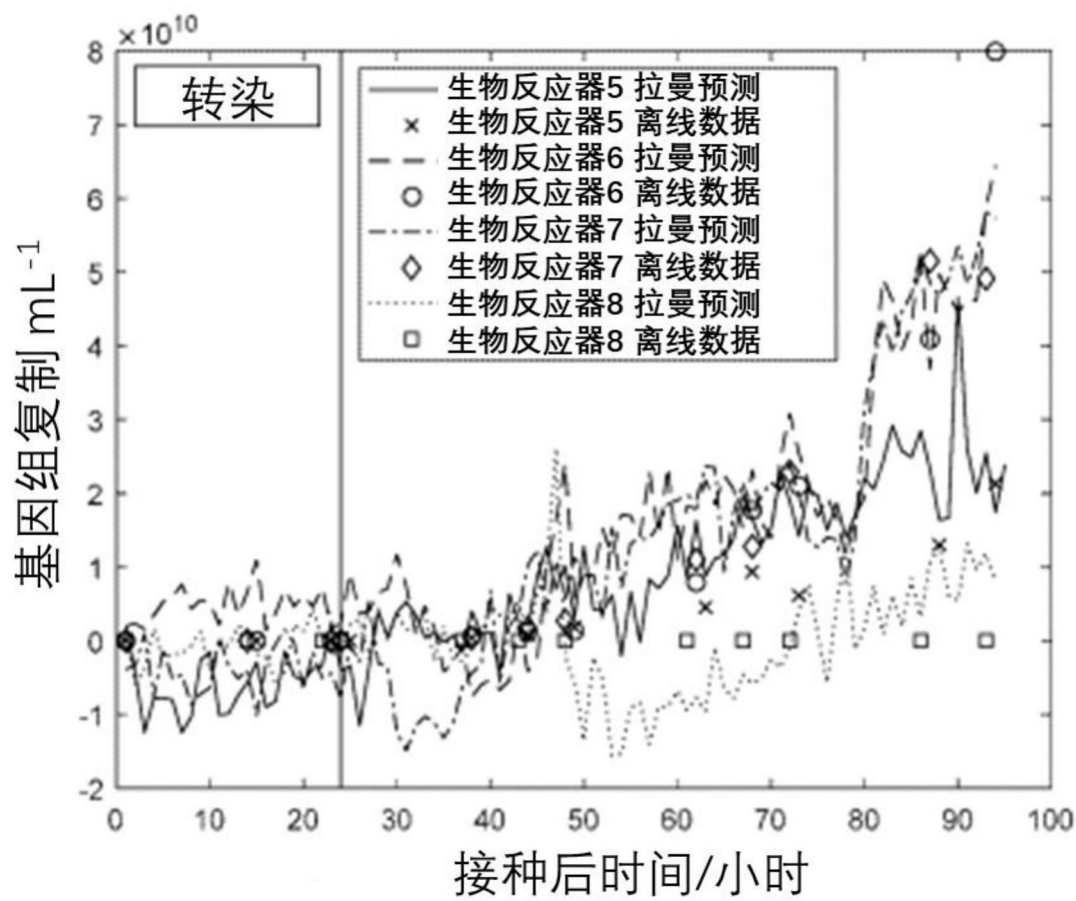


图31

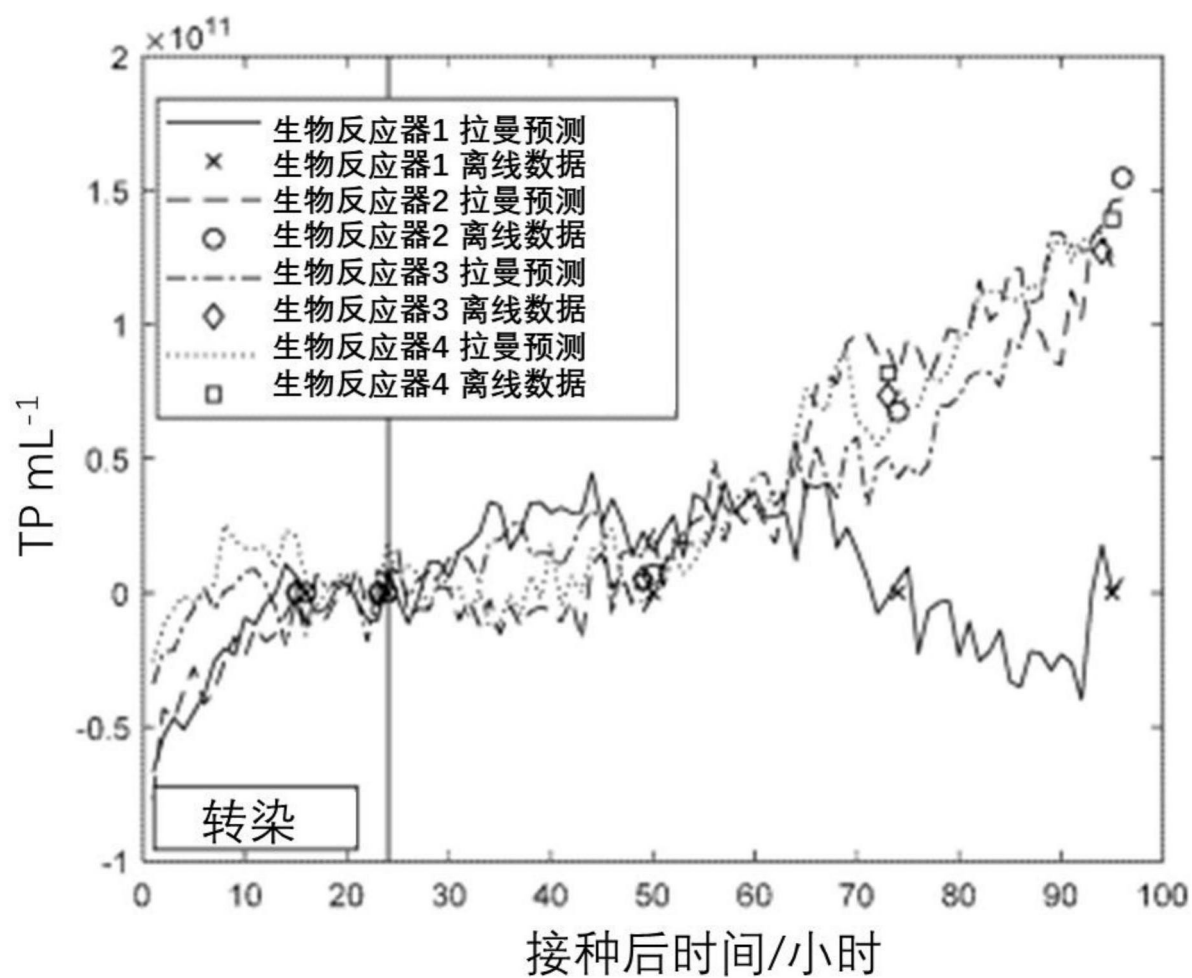


图32

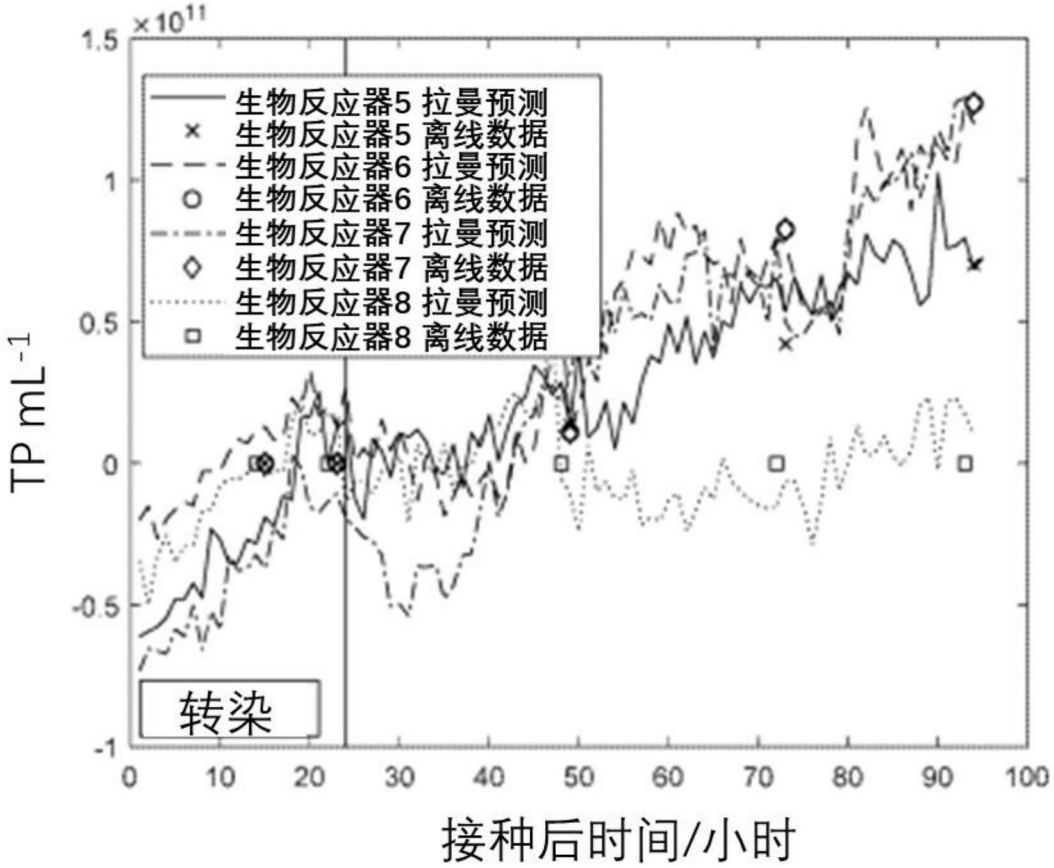


图33

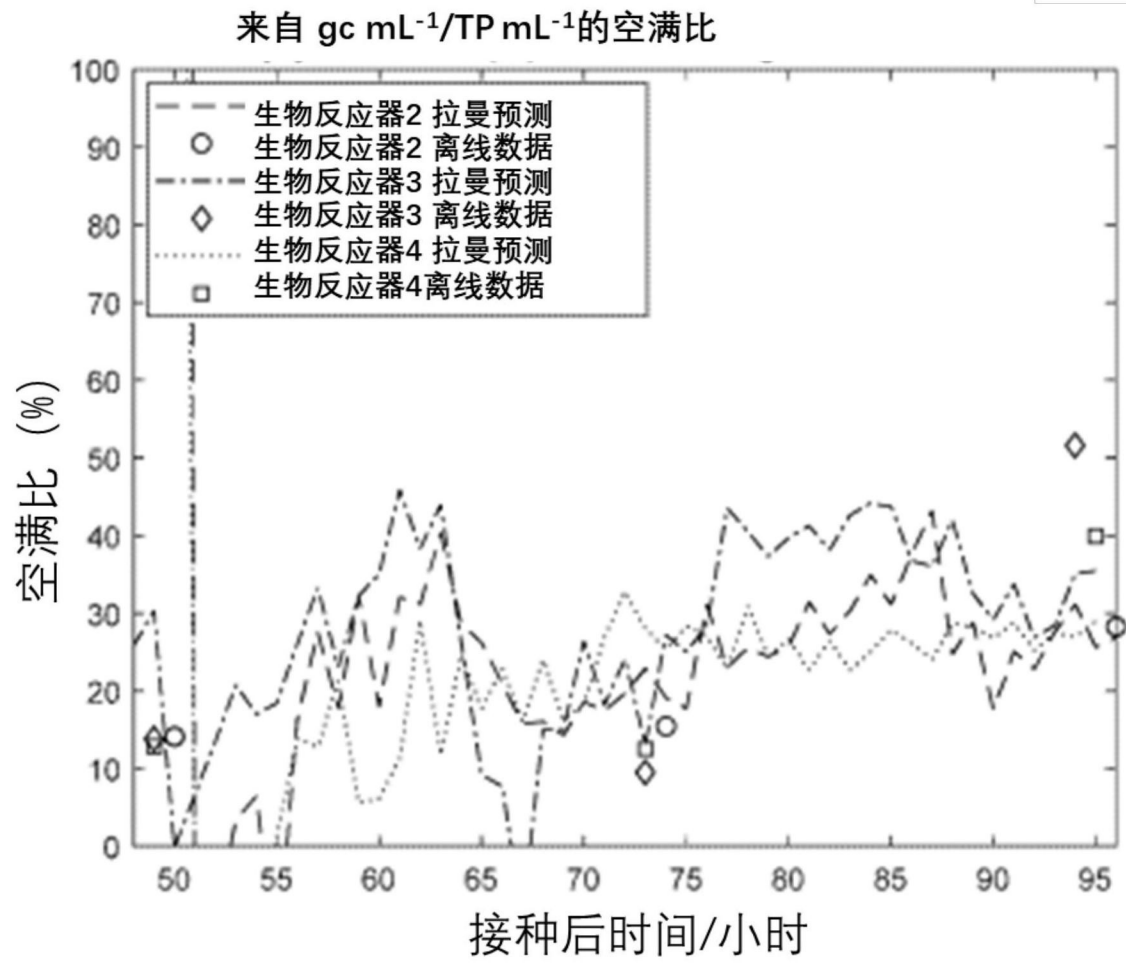


图34

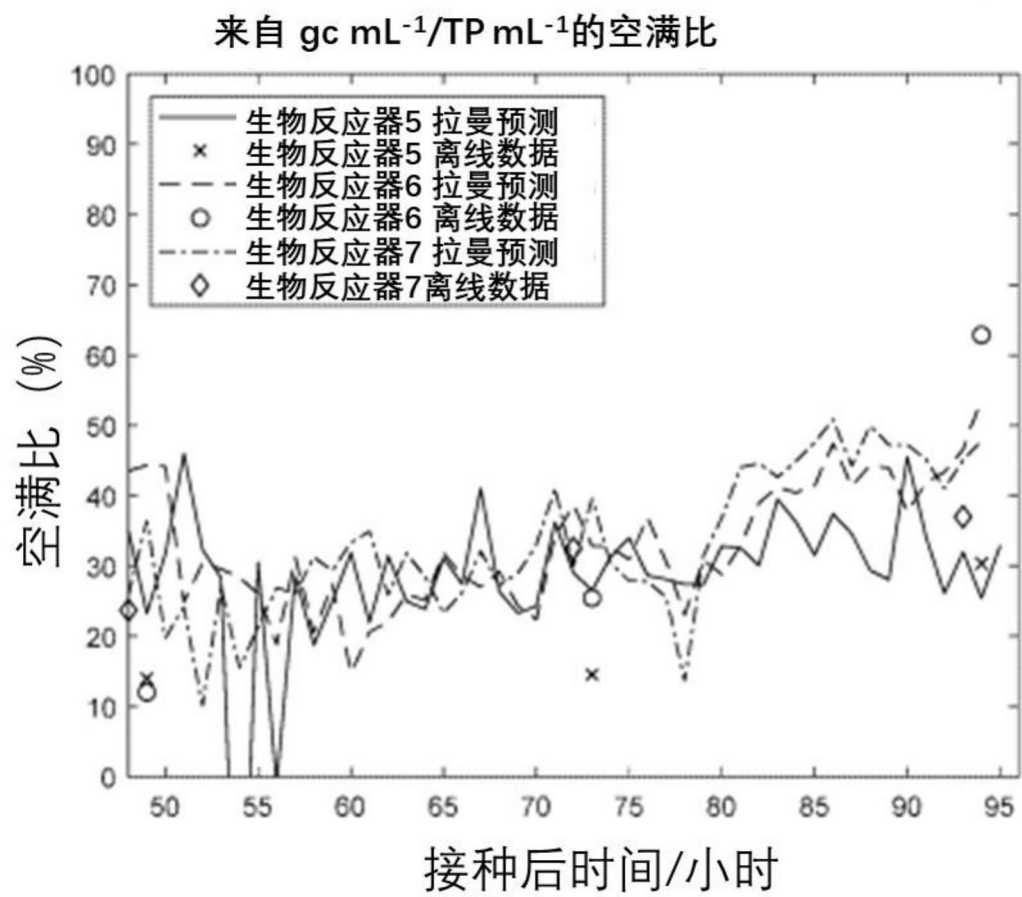


图35