



(12) 发明专利申请

(10) 申请公布号 CN 103201616 A

(43) 申请公布日 2013. 07. 10

(21) 申请号 201180053140. X

(74) 专利代理机构 北京市中咨律师事务所
11247

(22) 申请日 2011. 11. 03

代理人 黄草生 林柏楠

(30) 优先权数据

10190193. 2 2010. 11. 05 EP

(51) Int. Cl.

G01N 21/35(2006. 01)

G01N 33/68(2006. 01)

C12M 1/34(2006. 01)

C12M 3/00(2006. 01)

(85) PCT申请进入国家阶段日

2013. 05. 03

G01N 21/64(2006. 01)

G01J 3/44(2006. 01)

(86) PCT申请的申请数据

PCT/EP2011/069267 2011. 11. 03

(87) PCT申请的公布数据

W02012/059520 EN 2012. 05. 10

(71) 申请人 弗·哈夫曼-拉罗切有限公司

地址 瑞士巴塞尔

(72) 发明人 J·卡多苏-曼内斯 C·哈克迈尔

G·E·乔斯 U·施特劳斯 S·韦茨

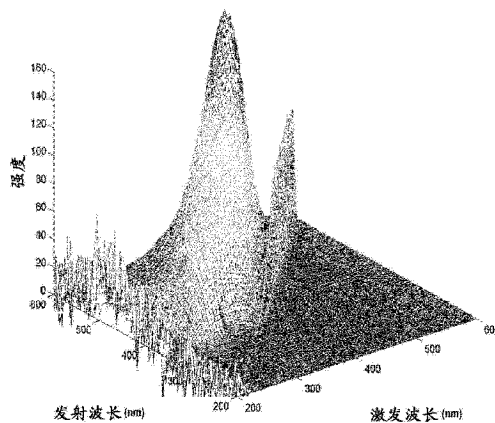
权利要求书1页 说明书14页 附图16页

(54) 发明名称

原料的光谱指纹识别

(57) 摘要

本文报道了选择待用于培养表达目标蛋白质的哺乳动物细胞的培养组分批次的方法, 其中培养中至少使用两种不同组分, 所述方法包括以下步骤: a) 提供用第一种光谱学方法获得的不同批次的第一种组分的光谱和用第二种光谱学方法获得的不同批次的第二种组分的光谱, 以及在使用这些不同批次的第一种和第二种组分的组合的培养中获得的目标蛋白质的培养上清产量, b) 处理光谱, 过滤光谱, 平滑光谱, 并将光谱转化为其一阶导数, c) 鉴定光谱中的图式, d) 鉴定 d) 中鉴定的图式与培养产量的关系, e) 提供用第一种光谱学方法获得的其他批次的第一种组分的光谱和用第二种光谱学方法获得的其他批次的第二种组分的光谱, f) 处理光谱, 过滤光谱, 平滑光谱, 并将光谱转化为其一阶导数, g) 如果基于 d) 中鉴定的关系预测的培养上清产量在 a) 中提供的平均产量的 +/-10% 内, 选择提供的第一种组分和提供的第二种组分的组合。



1. 选择待用于培养表达目标蛋白质的哺乳动物细胞的培养组分批次的方法,其中培养中使用至少两种不同的组分,所述方法包括以下步骤:

a) 提供用第一种光谱学方法获得的不同批次的的第一种组分的光谱和用第二种不同的光谱学方法获得的第二种组分的光谱,以及在使用这些不同批次的的第一种和第二种组分的组合的培养中获得的目标蛋白质的培养上清产量,

b) 鉴定计算光谱 PCA 分数后融合的光谱与培养产量的关系,

c) 提供用第一种光谱学方法获得的其他批次的的第一种组分的光谱和用第二种光谱学方法获得的其他批次的第二种组分的光谱,

d) 如果基于 b) 中鉴定的计算光谱 PCA 分数后融合的光谱关系预测的培养上清产量在 a) 中提供的平均产量的 $\pm 10\%$ 内,选择提供的的第一种组分和提供的的第二种组分的组合。

2. 根据权利要求 1 的方法,所述方法特征在于第一种和第二种光谱学方法选自 NIR 光谱学、MIR 光谱学和 2D 荧光光谱学。

3. 根据前述权利要求中任一项的方法,所述方法特征在于通过主成分分析进行鉴定。

4. 根据权利要求 3 的方法,所述方法特征在于主成分分析是展开的主成分分析。

5. 根据权利要求 4 的方法,所述方法特征在于展开保留第一种模式(样品)的信息。

6. 根据前述权利要求中任一项的方法,所述方法特征在于用偏最小二乘分析鉴定用 PCA 分数融合并压缩的光谱和收获时培养产量之间的关系。

7. 根据前述权利要求中任一项的方法,所述方法特征在于目标蛋白质是抗体或抗体片段或抗体缀合物。

原料的光谱指纹识别

[0001] 本文报道了在接到培养材料组分时,在进行测试培养前并且不需要进行测试培养,对其产品产量方面进行评估的方法。

[0002] 发明背景

[0003] 自从 20 世纪 80 年代早期,重组 DNA 技术使得在不同类型的微生物例如细菌、酵母或哺乳动物细胞中表达重组蛋白质变得可能后,重组生物药物产品市场在持续增长。从那时起,这类蛋白质产品被用于多种诊断和药物应用。

[0004] 由于对重组蛋白质的需求增加,对高效和稳健的生产方法的需求极为迫切。对稳健和可重现的生产方法最重要的影响因素之一是起始材料例如培养基的组分。大部分培养基是除了别的之外无机盐、糖、氨基酸、维生素、有机酸和缓冲剂的复杂混合物。许多情况下,使用复杂的,化学上未限定的原料例如源自植物或细菌的蛋白质水解物来促进细胞生长和蛋白质生产。

[0005] 通常,原料以粉末混合物的形式供应,然后溶于水以形成培养基。许多情况下,对于化学上未限定的蛋白质水解物和化学上限定的基础培养基混合物,可观察到显著的批次间变异,导致重组生产的治疗蛋白质产量的巨大变化。

[0006] 快速光谱“指纹识别”技术,例如近红外光谱、中红外光谱、拉曼光谱或 2D 荧光光谱相对便宜,并非常适于分析复杂混合物。这些方法生成非常大量的高维度数据,所述数据只能用化学计量学方法例如主成分分析(PCA)或偏最小二乘法(PLS)建模处理。复杂光谱学方法和化学计量学的组合通常用于原料的身份测试或用作原料分类的工具。

[0007] **Næs, T.** 等人报道了主成分分析(PCA)和偏最小二乘法(PLS)用于处理和建模复杂数据的用途(**Næs, T.** 等人, NIR Publications, (2002))。在 W02009/086083 中报道了用 PLS 分等级地组织数据的方法。在 W02008/146059 中报道了用于确定生物混合物的级分的相对重要性的分析仪和方法。在 W02009/061326 中报道了对色谱材料的评价。

[0008] 在 US2009/0306932 中报道了多变量数据阵列的快速分类方法。在 EP2128599 中报道了分析光谱数据用于选择校准模型。在 US5, 498, 875 中报道了样品化学分析的信号处理。在 US2008/0177481 中报道了分类科学材料例如硅酸盐材料、聚合物材料和 / 或纳米材料的方法。在 US2010/0129857 中报道了分离和鉴定微生物的方法。

[0009] 发明概述

[0010] 已经发现,能够根据培养基组分(例如蛋白质水解物和 / 或用作复杂培养基的组分的化学上限定的培养基制剂)的 NIR 和 2D 荧光光谱的组合预测重组蛋白质生产方法的性能。

[0011] 如本文报道的一方面是选择待用于培养表达目标蛋白质的哺乳动物细胞的培养基组分批或批次的方法,其中培养使用至少两种不同组分,用两种不同光谱技术的融合的光谱数据进行此类选择。

[0012] 在一个实施方式中,选择待用于培养表达目标蛋白质的哺乳动物细胞的培养组分批次的方法包括以下步骤,其中培养中使用至少两种不同的培养组分:

[0013] a) 提供用第一种光谱学方法获得的不同批次的第一种组分的光谱、用不同于第一

种光谱学方法的第二种光谱学方法获得的不同批次的第二种组分的光谱,以及在使用这些不同批次的的第一种和第二种组分的组合的培养中获得的目标蛋白质的培养上清产量,

[0014] b) 鉴定计算光谱 PCA 分数后融合的光谱与培养产量的关系,

[0015] c) 提供用第一种光谱学方法获得的其他批次的的第一种组分的光谱和 / 或用第二种光谱学方法获得的其他批次的第二种组分的光谱,和

[0016] d) 如果基于 b) 中鉴定的计算光谱 PCA 分数后融合的光谱关系预测的培养上清产量在 a) 中提供的平均产量的 $\pm 10\%$ 内,选择提供的第一种组分和提供的第二种组分的组合。

[0017] 在一个实施方式中,选择待用于培养表达目标蛋白质的哺乳动物细胞的培养组分批次的方法包括以下步骤,其中培养中使用至少两种不同培养组分:

[0018] a) 提供用第一种光谱学方法获得的不同批次的的第一种组分的光谱、用不同于第一种光谱学方法的第二种光谱学方法获得的不同批次的第二种组分的光谱,以及在使用这些不同批次的的第一种和第二种组分的组合的培养中获得的目标蛋白质的培养上清产量,

[0019] b) 处理光谱,过滤光谱,平滑光谱,并将光谱转化为其一阶导数,

[0020] c) 鉴定光谱中的图式,

[0021] d) 鉴定 c) 中鉴定的图式与培养的产量的关系,

[0022] e) 提供用第一种光谱学方法获得的其他批次的的第一种组分的光谱和 / 或用第二种光谱学方法获得的其他批次的第二种组分的光谱,

[0023] f) 处理光谱,过滤光谱,平滑光谱,并将光谱转化为其一阶导数,

[0024] g) 如果基于 d) 中鉴定的关系预测的培养上清产量在 a) 中提供的平均产量的 $\pm 10\%$ 内,选择提供的第一种组分和提供的第二种组分的组合。

[0025] 在一个实施方式中,第一种和第二种光谱学方法选自 NIR 光谱学、MIR 光谱学和 2D 荧光光谱学。

[0026] 在一个实施方式中,光谱的处理包含移除水吸收区和应用多元散射校正,和 / 或过滤包含 Savitzky-Golay 过滤。

[0027] 在一个实施方式中通过主成分分析鉴定光谱中的图式。在一个实施方式中主成分分析是展开主成分分析。在一个实施方式中展开保留第一模式(样品)的信息。在一个实施方式中 Savitzky-Golay 平滑具有 19 点的窗口和二阶多项式。在一个实施方式中数据是以平均数为中心的,并且主成分的最优数是用留一(leave-one-out)交叉验证法选择的。

[0028] 在一个实施方式中处理包括排除散射区和移除点的插值。在一个实施方式中终光谱是用 290nm 至 594nm 的发射波长范围和 230nm 至 575nm 的激发波长范围构成的。

[0029] 在一个实施方式中,用偏最小二乘分析鉴定用 PCA 分数融合并压缩的光谱和收获时培养产量之间的关系。

[0030] 在一个实施方式中,在 4784cm^{-1} 至 8936cm^{-1} 的波数区收集 NIR 光谱。

[0031] 在一个实施方式中,光谱维数从 1039 波数减至 3 个主成分。

[0032] 在一个实施方式中,目标蛋白质是抗体、或抗体片段或抗体缀合物。

[0033] 发明详述

[0034] 已经发现,可以基于培养基组分(例如蛋白质水解物和 / 或用作复杂培养基的组分的化学限定的培养基制剂)的 NIR 和 2D 荧光光谱中所含有的组合的信息预测重组蛋白质

的生产方法的性能。

[0035] 本文报道了这样的方法,其中由用于重组生物药物的发酵的两种培养基组分获得的、来自两种不同(正交)光谱学技术的光谱 -- 在处理使其通过变量减少至主成分分析(PCA)分数而加和后 -- 被组合,并使用此类经转换的光谱(输入)模型,基于具有在不同发酵性能方面的批次间变异的所研究的培养基组分混合物,来预测生物药物产品培养收获时的产量(输出)。

[0036] 通过组合使用不同(正交)的光谱学和 PCA 方法(以确保其加和性),以及产生此类培养基混合物对主发酵收获时产量影响的方法模型,建立了预测能力,允许选择最适于方法目标的每种原料和 / 或配制混合物的培养基批次。

[0037] 形成完全培养基的单独的组分的不同批次在其具体组分中有轻微差异,但仍在制造商提供的说明书范围内。在一些情况下,可跟踪此变化至单个成分,但通常批次间变异性不能用分析方法检测。为评估不同的单独的组分批次对产品产量的影响,可重复进行相同哺乳动物细胞系的可比较培养。

[0038] 本文报道了 56 个培养,其中发酵和种子培养基分别使用了 9 个不同批次的大豆蛋白水解物、2 个不同大豆蛋白水解物批次的 2 种混合物、5 个批次的稻蛋白水解物,和 6 个批次的化学上限定的基础培养基粉。

[0039] 为评估不同大豆蛋白水解物批次在产品产量方面的影响,进行了可比较培养,其中在发酵和种子培养基中使用相同批次的化学上限定的基础培养基和稻蛋白水解物。可根据使用的不同大豆蛋白水解物批次对结果分组。基于处于相似的平均接种细胞密度(ICD)值的产品产量,评估不同批次的性能(表 1)。

[0040] 表 1

[0041]

批次	大豆蛋白水解物批号	化学上限定的培养基批号	稻蛋白水解物批号	ICD	330h 时产品[mg/l]
D45KD11	1	1	1	5.7	1319
D45KD12				5.3	1234
D45KD13				5.6	1305
D45KD22	2			5.3	1023
D45KD23				5.1	1070
D45KD31	3			4.8	1008
D45KD32				4.9	991
D45KD33				5.3	978

[0042] 第 2 组培养获得的结果列于表 2。

[0043] 表 2

[0044]

批次	大豆蛋白水解物批号	化学上限定的培养基批号	稻蛋白水解物批号	ICD	330h 时产品[mg/l]
D52KD11	1	2	2	6.1	1434
D52KD12				5.0	1411
D52KD13				5.6	1459
D52KD21	4			5.0	1213
D52KD22				5.3	1243
D52KD23				5.4	1163
D55KD11	5			5.0	1409
D55KD12				5.4	1426
D55KD13				5.7	1430
D55KD21	2			6.8	1263
D55KD22				6.8	1256
D55KD23				6.8	1278
D55KD31	6			6.1	1269
D55KD32				6.1	1262
D55KD33				5.8	1265

[0045] 可以看出,不同批次的单独的组分导致不同的产品产量。在这一系列培养中也使用了不同的平均 ICD 值。尽管具有较低的 ICD 值,使用批次 1 和批次 5 的培养比具有较高 ICD 值(批次 3 和批次 6)的培养给出显著更高的产品产量。因此,不同的大豆蛋白水解物批次导致不同的生产性能。类似地,可以评估稻蛋白水解物对方法性能的影响(表 3)。

[0046] 表 3

[0047]

批次	大豆蛋白水解物批号	化学上限定的培养基批号	稻蛋白水解物批号	ICD	330h 时产品[mg/l]
D61KD11	3	3	2	5.9	1132
D61KD12				6.0	1085
D61KD13				5.3	1101
D61KD21			3	6.1	1062
D61KD22				6.1	1056
D61KD23				5.6	1043

[0048] 进行了 6 个培养,并可根据每个培养中使用的不同批次的稻蛋白水解物对其分组。可基于平均产品产量评价不同稻蛋白水解物批次的性能。两组,即两个稻蛋白水解物

批次具有类似的 ICD 值。

[0049] 为评估化学上限定的基础培养基对产品产量的影响,可使用在发酵起始培养基制剂和种子培养基中的相同批次的大豆蛋白水解物和稻蛋白水解物进行培养。进行 3 次系列实验(表 4、5 和 6)。

[0050] 第一系列包含 6 个培养,所述培养发酵和种子培养基中含大豆蛋白水解物批次 3 (如表 3)和稻蛋白水解物批次 2 (如表 2)。根据使用的化学上限定的基础培养基批次将培养分组。基于产品产量评估不同的化学上限定的基础培养基批次的性能。两组的平均 ICD 和平均产品产量都有轻微差异。具有较低的 ICD 可获得较低的产品形成。因此,化学上限定的基础培养基批次对产品产量几乎或完全没有影响。

[0051] 表 4

[0052]

批次	大豆蛋白水解物批号	化学上限定的培养基批号	稻蛋白水解物批号	ICD	330h 时产品[mg/l]
D55KD21	3	2	2	6.8	1263
D55KD22				6.8	1256
D55KD23				6.8	1278
D61KD11		3		5.9	1132
D61KD12				6.0	1085
D61KD13				5.3	1101

[0053] 第二系列涉及 6 个培养,所述培养在发酵起始培养基制剂和种子培养基中使用大豆蛋白水解物批次 1 (如表 2 中)。根据使用的化学上限定的基础培养基批次将实验分组。不存在显著的 ICD 差异。因此,产品产量的差异是由于使用的化学上限定的基础培养基批次的差异。

[0054] 表 5

批次	大豆蛋白水解物批号	化学上限定的培养基批号	ICD	330h 时产品[mg/l]
D45KD11	1	1	5.7	1319
D45KD12			5.3	1234
D45KD13			5.6	1205
D52KD11		2	6.1	1434
D52KD12			5.0	1411
D52KD13			5.6	1459

[0056] 第三系列涉及 5 个培养,所述培养在发酵起始培养基配方和种子培养基中含有大豆蛋白水解物批次 2。根据使用的化学上限定的基础培养基批次将实验分组。两组中使用的 ICD 和获得的产品浓度都有差别。

[0057] 表 6

批次	大豆蛋白 水解物批 号	化学上限定的培养 基批号	ICD	330h 时产品 [mg/l]
[0058] D45KD22	2	1	5.3	1023
D45KD23			5.1	1070
D73KD11		4	4.9	1062
D73KD12			4.3	1112
D73KD13			4.4	1121

[0059] 从以上可以看出,存在表征原料批次的需求和提供这样的方法的需求,所述方法中获得的数据可用于预测哪批原料生产较高产量的产品,而不需要进行发酵实验。

[0060] 可获得三种不同培养基组分的全部批次的 NIR、MIR 和 2D 荧光光谱。然后可用已建立的化学计量学方法进行光谱分析。本文报道了分析用这些不同源获得的光谱信息的新方法,其可用于预测性建模目的。

[0061] 在不同时间段获得一式三份的原料批次的 NIR 光谱。对于粉末和异质粗糙样品, NIR 光谱在重复间不同。可基于其在 PCA 分数曲线空间上的相对位置(欧几里得距离)剔除此类偏远的重复。

[0062] 在全部提供的测量中选择大豆蛋白水解物的 18 个批次、稻蛋白水解物的 12 个批次和化学上限定的基础培养基的 14 个批次的 NIR 光谱。在 4784cm^{-1} 和 8936cm^{-1} 之间收集 NIR 光谱。此光谱区不包含噪声区。观察到的强基线迁移是由于光散射,所述光散射与平均颗粒尺寸分布(粒度)具有差异的不同原料批次相关。不进行基线修正的原光谱分析允许集中在主要由物理效应导致的变化。对每种原料分别进行原光谱 PCA 分析。

[0063] 图 1 显示了不同的受试大豆蛋白水解物批次在 2 维空间上的分布,所述 2 维空间是通过基于原始 NIR 光谱、捕获 94% 的 NIR 光谱变量的 PCA 构建的。光谱维度从 1039 个波数降低至 3 个显著主成分。不能基于此分析区分产生高产品产量的批次与产生低产品产量的批次。此外,样品的粒度(如不同的 NIR 光谱基线所示,图 2)和湿度含量(根据 Karl Fischer 测量)也不同,使得根据任何单个性聚聚类批次非常困难。

[0064] 图 3 显示了不同的受试稻蛋白水解物批次如何在 2 维空间上分布,所述 2 维空间是通过基于原始 NIR 光谱、捕获 92% 的 NIR 光谱变量的 PCA 构建的。对于大豆蛋白水解物,不能仅基于此分析区分产生高产品产量的批次与产生低产品产量的批次。同样,样品粒度和湿度在批次间不同,影响聚类。

[0065] 图 4 显示了化学上限定的基础培养基的批次在 2 维空间上的分布,所述 2 维空间是通过基于原始 NIR 光谱、捕获 98% 的 NIR 光谱变量的 PCA 构建的。对于大豆和稻蛋白水解物,不能仅基于此分析区分产生高产品产量的批次与产生低产品产量的批次。

[0066] 如由获得的 NIR 光谱所示,所分析的三种培养基组分的粒度和湿度含量显示显著的批次间变异。NIR 对这两种因素都非常灵敏。此外,这两种因素都超过其他可能存在的较

小但仍显著的化学组成差异。在 PCA 分析前,须通过光谱预处理移除物理信息。

[0067] 水在 NIR 区,特别是在从 $6,900\text{cm}^{-1}$ 到 $7,150\text{cm}^{-1}$ 和从 $5,160\text{cm}^{-1}$ 到 $5,270\text{cm}^{-1}$ 的范围内吸收非常强烈。这些吸收区分别由 O-H 伸展键的第一泛频峰(overtone)和 O-H 伸展与 O-H 弯曲键的组合造成。水吸收区可被移除。此外,基线迁移可通过应用多元散射校正(MSC)消除。为增强样品间的变异,可应用 Savitzky-Golay 过滤和平滑方法,并且光谱可被转化为其一阶导数(25 点的窗口)。

[0068] 对先前经预处理的大豆蛋白水解物光谱进行 PCA 分析(图 5)。在方法产量方面表现非常优秀至优秀的几乎全部批次聚集于 PCA 曲线的左侧(负 PC1 分值)。相反,看起来表现差的批次 4 占据了曲线右侧的空间。

[0069] 对先前经预处理的稻蛋白水解物光谱进行 PCA 分析(图 6)。产生非常相似产量的批次聚类到一起,因此,显示经预处理的曲线的 PCA 是充分的,并且已经有一些可追踪至此组分原料的化学组成的批次间变异,其与粒度或湿度水平无关。

[0070] 化学上限定的基础培养基的经预处理的光谱的 PCA 分析(图 7)显示,总体上表现非常优秀至优秀的全部批次聚集在 PCA 曲线的左侧(负 PC1 分值)。相反,看起来表现差的批次 3 占据了曲线右侧的空间。这些结果与从蛋白水解物批次获得的结果是可比较的。

[0071] 除了 NIR 光谱,可分析获得的不同水溶性发酵原料的荧光激发-发射光谱(EEM)。可建立以激发波长沿 x 轴,发射波长沿 y 轴,强度沿 z 轴的三向数据阵列。图 8 中显示了大豆蛋白水解物批次样品的荧光 EEM 图。

[0072] 获得了 19 个批次的大豆蛋白水解物、12 个批次的稻蛋白水解物和 14 个批次的化学上限定的基础培养基的 2D 荧光光谱。光谱是使用从 200nm 至 600nm,间隔 5nm 的激发波长和从 200nm 至 600nm,间隔 2nm 的发射波长获得的,共有 81 个激发波长,201 个发射波长。

[0073] 为允许基于对原料的分析预测培养产量,可从各个矩阵生成每种原料的三向阵列。

[0074] 典型的 EEM 光谱可被瑞利和拉曼散射效应影响,它们影响荧光图的信息内容。为克服瑞利效应,可使用若干策略和技术:

[0075] - 将小于激发波长的发射波长归零;

[0076] - 插入散射区的缺失值;

[0077] - 排除散射区并对移除的点进行插值;或

[0078] - 减去背景光谱。

[0079] 发现排除散射区和对移除的点进行插值最适于本文报道的方法。因此可使用

Matlab[®] 算法 EEMscat。此算法可从以下网址免费下载:http://www.models.kvl.dk/source/EEM_correction/。用此处理可完全除去散射。光谱也在第一激发波长内沿着整个发射轴显示显著的噪音。此区域(200nm 至 225nm),以及无信息发射波长(200nm 至 315nm 和 596nm 至 600nm)和激发波长(580nm 至 600nm)被从光谱排除。获得的光谱如图 9 所示。

[0080] 终大豆蛋白水解物光谱是由 320nm 至 594nm 的发射波长范围和 230nm 至 575nm 的激发波长范围组成的,获得 $19 \times 138 \times 70$ 元素的阵列。可对稻蛋白水解物和化学上限定的基础培养基数据集进行相同的步骤。因此,终稻蛋白水解物光谱是由分别为 290nm 至 594nm 和 230nm 至 550nm 的发射和激发波长范围组成的,获得 $12 \times 153 \times 65$ 元素的阵列。终化学上限定的基础培养基光谱是由 290nm 至 594nm 的发射波长范围和 230nm 至 550nm 的激发波长

范围组成的,获得 14x162x60 元素的阵列。

[0081] 总之,可对每种原料数据集进行 EEM 光谱预处理以增强信号噪音比例。因此可清楚地看到每种原料之间的差异:大豆蛋白水解物包含 2 个或 3 个荧光团,稻蛋白水解物包含 3 个荧光团,而化学上限定的基础培养基包含多于 4 个荧光团。

[0082] 为获得原料批次间变异的概况,可对每种组分原料进行展开的荧光数据阵列的 PCA。可在三向阵列的三个模式中的任一个中应用展开步骤。为增强批次间差异,可采用第一模式(样品)的展开保留信息。这样,荧光图可展开为一排相继的发射光谱(图 10)。

[0083] 大豆蛋白水解物阵列的维度是 19x138x70(批次 x 发射波长 x 激发波长)。在展开策略后,可获得大小为 19x9960 的二向矩阵。图 11 显示了获得的三个不同批次的大豆蛋白水解物的光谱的一小部分。可看到极端激发波长中的噪音。

[0084] 为克服这些偏差,可使用若干策略。发现使用 19 点窗口和二阶多项式的 Savitzky-Golay 平滑来移除噪声最合适,并且多元散射校准(MSC)最适于消除基线漂移。

[0085] 对大豆蛋白水解物经预处理的矩阵进行展开 PCA。数据是以平均数为中心的,使用留一交叉验证法选择最优主成分数。图 12 显示了覆盖整个展开的 EEM 图上所见的 96% 的方差的 PCA 的 PC1xPC2 的分数曲线。

[0086] 在展开后,获得的稻蛋白水解物矩阵大小为 12x9945。应用了与大豆蛋白水解物相同的预处理。图 13 显示了使用三个主成分,覆盖展开的 EEM 光谱中超过 98% 的方差的 PCA 的 PC1xPC2 的分数曲线。

[0087] 展开的化学上限定的基础培养基矩阵的大小为 14x9600。使用了与应用于另两种培养基组分相同的 EEM 光谱预处理步骤。图 14 显示了使用两个主成分,覆盖展开的 EEM 光谱中超过 92% 的总方差的 PCA 的 PC1xPC2 的分数曲线。与前述,使用相同培养基组分的 NIR 光谱,发现产生较高产量的批次与产生较低产量的批次在 EEM 展开光谱的 PCA 分数曲线上分开。

[0088] 可建立 PLS 模型用于基于从不同批次的每种培养基组分和 / 或其组合获得的 NIR 和 / 或荧光光谱来预测方法最后的产品产量。PLS 算法是给出 X 模块(经预处理的光谱,有或没有变量选择)和 Y 模块(产品参数),通过找到 X 中导致 Y 中改变的变异将二者相互关联(即最大化两个模块之间的协方差)。可限定基础集,其中可包括大部分不同批次原料。在具有相同批次组合的重复批中,选择产生最高产品产量的一个用于校准数据集(表 7)。

[0089] 表 7

批次	大豆蛋白水解物 F/ZF 批号	330h 时产品 [mg/l]
D52KD13	1	1458
D52KD22	4	1232
D55KD13	5	1430
D55KD23	3	1257
D55KD31	6	1263
D73KD13	2	1120
D73KD33	7	1044
D79KD22	8	1162

[0090]

[0091] 可按如上所述经预处理的光谱以移除源自不同颗粒尺寸分布的物理效应的

影响。由于未使用重复光谱,用留一交叉验证法作为内部验证策略。

[0092] 获得的模型只由 2 个 LV 组成,但获得了 0.139 的非显著的 R^2 。测量值与交叉验证预测的曲线如图 15 所示。

[0093] 可用如表 8 所示的校准数据集建立将不同批次的化学上限定的基础培养基与产品产量相关联的 PLS 模型。

[0094] 表 8

批次	化学上限定的培养基 F/ZF 批号	330h 时产品 [mg/l]
D45KD11	1	1314
D52KD13	2	1458
D61KD12	3	1134
D73KD21	4	1147
D79KD22	5	1162

[0096] 获得的模型只由 2 个 LV 组成,但再次获得了 0.04 的非显著的 R^2 (图 16)。

[0097] 考虑到不只一种培养基组分,而是两个最相关的组分影响产量,也考虑到使用的各种不同的光谱学方法捕获不同的化学信息,可在相同的光谱学 / 不同培养基组分和不同的光谱学 / 不同的培养基组分之间使用组合策略。

[0098] 用于选择校准和验证批次的标准基于在校准过程中获得可能的最宽范围(表 9)。

[0099] 表 9

	批次	大豆蛋白水解物 F/ZF 批号	化学上限定的培养基 F/ZF 批次	330h 时产品 [mg/l]
[0100] 校准	D45KD11	1	1	1314
	D45KD31	3	1	999
	D52KD13	1	2	1458
	D52KD22	4	2	1232
	D55KD13	5	2	1430
	D55KD31	6	2	1263
	D61KD12	3	3	1134
	D73KD13	2	4	1120
	D73KD33	7	4	1044
	D79KD22	8	5	1162
验证	D45KD23	2	1	1061
	D55KD23	3	2	1257
	D73KD21	8	4	1147

[0101] 用三分之一的数据集进行外部验证。用如上所述的方法对校准和验证数据(NIR 光谱)进行预处理。获得的预测模型基于 3 个 LV,并且获得的 R^2 达到 0.88 的显著值。

[0102] 模型的精确性和长期稳健性反映为高 R^2 值和低校准和验证误差,在 RMSECV 和 RMSEP 之间有小差别(图 17)。在上述情况下,预测误差低(RMSEP=36mg/l),与 RMSECV (126mg/l) 差别不大。

[0103] 因此,已经发现产品产量可与来自培养基的不同化合物的光谱学数据相关联,所

述数据是用两种(最重要的)方法原料或培养基组分的相同性质的光谱学信息(NIR)的组合获得的。每个光谱具有 944 个波数,并且包括于模型中的整个校准数据集由 18880 个变量(10 个样品 x2 种原料 x 变量选择后 944 个波数)代表。为降低所需工作量,进行基于光谱的 PCA 分析,所述光谱首先通过将所含信息转化为少数几个非相关联的变量来压缩。以此获得的模型更简单,并且只含 2 个潜在变量(LV),获得 0.81 的 R^2 。

[0104] 不同的光谱学方法捕获互补的化学信息。使用两种不同类型的光谱学信息改善模型的预测质量。因此,使用大豆蛋白水解物的荧光光谱和化学上限定的基础培养基的 NIR 光谱(表 10)。

[0105] 表 10

[0106]

	批次	大豆蛋白水解物 F/ZF 批号	化学上限定的培养 基 F/ZF 批号	330h 时产 品[mg/l]
校准	D45KD11	1	1	1314
	D45KD31	3	1	999
	D52KD13	1	2	1458
	D52KD22	4	2	1232
	D55KD13	5	2	1430
	D55KD31	6	2	1263
	D61KD12	3	3	1134
	D73KD13	2	4	1120
	D73KD33	7	4	1044
	D79KD22	8	5	1162
验证	D45KD23	2	1	1061
	D55KD23	3	2	1257
	D73KD21	8	4	1147

[0107] 在如上所述的预处理后,将荧光光谱和 NIR 光谱压缩为几个主成分。获得的模型只有 3 个潜在变量,并且获得 0.90 的 R^2 (图 18)。当与先前的模型相比较时,此模型具有更好的性能并更稳健,因为其不仅有更高的 R^2 值,也具有更低的 RMSECV 和 RMSEP 值(约 90mg/l),二者之间差异非常小。

[0108] 用化学上限定的基础培养基的 MIR 而不是 NIR 进行其他测试。所使用的校准和验证数据集与上文所呈现的相同(见表 10)。按上文所述预处理荧光和 MIR 光谱。获得的模型有 3 个潜在变量,0.88 的 R^2 ,低 RMSECV 和 RMSEP 值,二者之间没有差别(都为约 100mg/l),因此与用化学上限定的基础培养基的 NIR 数据获得的模型没有显著差异(图 19)。

[0109] 将大豆蛋白水解物的 NIR 光谱和化学上确定的基础培养基的荧光光谱结合并评价获得的模型。用于建立模型的校准和验证数据集与上文相同(见表 10)。获得的模型有 3 个潜在变量和非常相似的 R^2 值(0.87)(图 20)和 RMSECV 和 RMSEP 值(分别为 124mg/l 和 60mg/l)。

[0110] 产品的参考分析物的分析方差处于约 60mg/l 左右(平均产品浓度 1200mg/l 的 5%),建立的大部分模型显示非常接近实验极限的预测精确度。

[0111] 总之,为实现对在 330h 时产品产量的预测,必须使用大豆蛋白水解物和化学上限定的基础培养基的光谱信息。化学上限定的基础培养基的荧光光谱学数据的使用给出了比

基于化学上限定的基础培养基的 NIR 光谱学数据和大豆蛋白水解物的 2D 荧光光谱学数据的模型略低的(但尽管如此是可比较的) 预测误差。

[0112] 如本文报道的方法涉及不同性质的光谱(荧光光谱和 IR 光谱)的组合,其从本质上具有不同维度(分别为 2 维(2D)和 1 维(1D)),并需要将每个光谱首先压缩为主成分分析得分并其次产生每个光谱得分的线性组合的操作。不同性质的光谱是通过维度降低和那些降低的转化的变量(通过压缩每个光谱获得的 PCA 得分)的线性组合来组合的。

[0113] 因此,在本文报道的方法中,不同维度和性质的光谱被用于捕获在两种不同发酵原料的混合物中的对所述原料的发酵性能负责的组分,并对特别的批次组合的发酵产量进行预测。

[0114] 用本文报道的方法能够基于待用于发酵方法的两种不同原料的光谱,通过确定发酵收获时的条件,提前 10 至 14 天预测性能。

[0115] 以下实施例和附图用于帮助理解本发明,本发明的实际范围由权利要求书限定。应当理解,可以对示出的程序进行修改而不背离本发明的精神。

附图说明

[0116] 图 1 不同的受试大豆蛋白水解物批次在通过基于原始 NIR 光谱的 PCA 建立的 2 维空间上的分布。

[0117] 图 2 不同大豆蛋白水解物批次的 NIR 光谱。

[0118] 图 3 不同的受试稻蛋白水解物批次在通过基于原始 NIR 光谱的 PCA 建立的 2 维空间上的分布。

[0119] 图 4 不同的受试化学上限定的基础培养基批次在通过基于原始 NIR 光谱的 PCA 建立的 2 维空间上的分布。

[0120] 图 5 基于大豆蛋白水解物批次的经预处理的光谱的 PCA 分析。

[0121] 图 6 基于稻蛋白水解物批次的经预处理的光谱的 PCA 分析。

[0122] 图 7 基于化学上限定的基础培养基批次的经预处理的光谱的 PCA 分析。

[0123] 图 8 大豆蛋白水解物批次样品的荧光 EEM 图。

[0124] 图 9 经处理的大豆蛋白水解物批次样品的荧光 EEM 图谱。

[0125] 图 10 将荧光图展开为一排发射光谱。

[0126] 图 11 三个不同批次的大豆蛋白水解物的展开光谱的引用。

[0127] 图 12 展开的 EEM 图的大豆蛋白水解物的 PCA 的 PC1xPC2 分数曲线。

[0128] 图 13 展开的 EEM 图的稻蛋白水解物的 PCA 的 PC1xPC2 分数曲线。

[0129] 图 14 展开的 EEM 图的化学上限定的基础培养基的 PCA 的 PC1xPC2 分数曲线。

[0130] 图 15 测量的与交叉验证预测的曲线。

[0131] 图 16 关联不同批次的化学上限定的基础培养基的 NIR 光谱与产品产量的 PLS 模型。

[0132] 图 17 关联不同批次的大豆蛋白水解物和化学上限定的基础培养基的 NIR 光谱与产品产量的 PLS 模型。

[0133] 图 18 关联不同批次的大豆蛋白水解物的荧光光谱和不同批次的化学上限定的基础培养基的 NIR 光谱与产品产量的 PLS 模型。

[0134] 图 19 关联不同批次的大豆蛋白水解物的荧光光谱和不同批次的化学上限定的基础培养基的 MIR 光谱与产品产量的 PLS 模型。

[0135] 图 20 关联不同批次的大豆蛋白水解物的 NIR 光谱和不同批次的化学上限定的基础培养基的荧光光谱与产品产量的 PLS 模型。

[0136] 图 21 泛频峰和共价键有机分子的组合带的 NIR 吸收放射。

实施例

[0137] 材料与amp;方法

[0138] 细胞培养：

[0139] 将细胞在温度、湿度和二氧化碳受控环境中的摇瓶中培养。为比较不同批次，用这些批次制备培养基，并在含有这些培养基的摇瓶中接种细胞。每天向摇瓶培养物加入一定体积的饲养培养基，以延长细胞生长并获得更高的产品浓度。

[0140] 近红外光谱(NIR)：

[0141] 在美国农业部的 Karl Norris 工作下，NIR 在 20 世纪 60 年代进入分析界(Siesler 等人，2002)。在电磁光谱中，NIR 区域位于中红外和可见区之间。在 $4,000-14,000\text{cm}^{-1}$ 波数(分别为 $700-2,500\text{nm}$ 的波长)范围内，泛频峰和有机分子共价键例如 N-H、O-H 和 C-H 的组合带的吸收放射(图 21)。

[0142] 用平底闪烁瓶在装备了钨 - 卤素源和 InAs 探测器的 **Bruker**[®] MPA FT-NIR 中收集 NIR 光谱。在 $4,999$ 至 $9,003\text{cm}^{-1}$ 的波数范围中记录每个光谱，平均 32 次扫描并且光谱解析度为 8cm^{-1} 。

[0143] 中红外光谱(MIR)：

[0144] 用石英玻璃杯在 Avatar370FT-IR, Thermo Fischer, Diamant ATR 中获得中红外光谱。每个光谱在 $4,000$ 至 400cm^{-1} 的波数范围内记录。

[0145] 荧光光谱：

[0146] 荧光光谱使用特定波长的照射激发分子，分子随后发射不同波长的放射。此技术通常用于研究大分子的结构和功能，特别是蛋白质相互作用。见于蛋白质和核酸的发色团的荧光特征的试验性设定如下表所示。

[0147]

物质	吸收		荧光	
	$I_{\text{最大}}(\text{nm})$	$\epsilon_{\text{max}}(10^{-3})$	$I_{\text{最大}}(\text{nm})$	f_{F}
色氨酸	280	5.60	348	0.20
酪氨酸	274	1.40	393	0.14
苯丙氨酸	257	0.20	282	0.04
腺嘌呤	260	13.40	321	2.60×10^{-4}
鸟嘌呤	275	8.10	329	2.60×10^{-4}
胞嘧啶	267	6.10	313	0.80×10^{-4}
尿嘧啶	260	9.50	308	0.40×10^{-4}
NADH	340	6.20	470	0.02

[0148] 使用从 200nm 至 600nm, 间隔 5nm 的激发波长和从 200nm 至 600nm, 间隔 2nm 的发射波长(给出共计 81 个激发波长, 201 个发射波长), 获得细胞培养原料的 2D 荧光光谱。使用 Varian Cary Eclipse Spectrometer, 在从 200nm 至 600nm, 间隔 5nm 的激发波长和从 200nm 至 600nm, 间隔 2nm 的发射波长上(共有 81 个激发波长, 201 个发射波长), 测量发射 - 激发荧光光谱。用软件 Cary Eclipse Bio, Package 1.1 收集数据。

[0149] 光谱处理和化学计量学分析:

[0150] 在 **Matlab**[®] 7.2(MathWorks, U. S. A.) 中用 PLS toolbox 5.5(Eigenvector, U. S. A.) 和 Simca P+12.01(Umetrics, Sweden) 进行光谱预处理和化学计量学计算。用 EEMscat 算法(Bahram 等人, 2006) 移除瑞利和拉曼散射。

[0151] 用 PCA(主成分分析)和 PLS(偏最小二乘)进行多变量数据分析。这些技术基于降低数据中存在的维度, 允许取得隐藏于大量数据中的相关信息。将原始测量的变量转化为新变量, 称为主成分。用 PCA 分析寻找光谱中的图式。为将这些图式与特定参数联系, 进行 PLS 分析以建立数学模型, 所述模型能只用光谱信息预测未来样品中的此参数值。

[0152] 为建立可靠的模型, 分析测量的质量极为重要。由于噪音和不需要的信息是测量固有的, 必须对获得的光谱进行预处理。

[0153] 处理 NIR 光谱中此类问题最常用的技术是 Savitzky-Golay 平滑过滤器(Savitzky, A. 和 Golay, M. J. E., Anal. Chem., 36(1964)1627-1639), 它通常与导数共用, 其具有减小基线迁移和增强光谱显著性质的优势。

[0154] 对于荧光光谱, 主要问题与拉曼和瑞利散射有关, 它们由光的偏差造成, 所述偏差与样品的荧光性质无关。由于受散射影响的波长区已知, 在此类特别区测量的强度可通过用插值点代替移除。

[0155] 为获得适于 PLS 和 PCA 分析的矩阵, 将三向发射 - 激发光谱展开。出于校准目的, 也进行了基于平行因子(PARAFAC)分析的三向分析。(Bahram, M. 等人, J. Chemometrics, 20(2006)99-105)。展开方法在于将此三个维度的两个连接, 保持另一个固

定。在此情况下,发射和激发轴被连接,维持样品的信息。

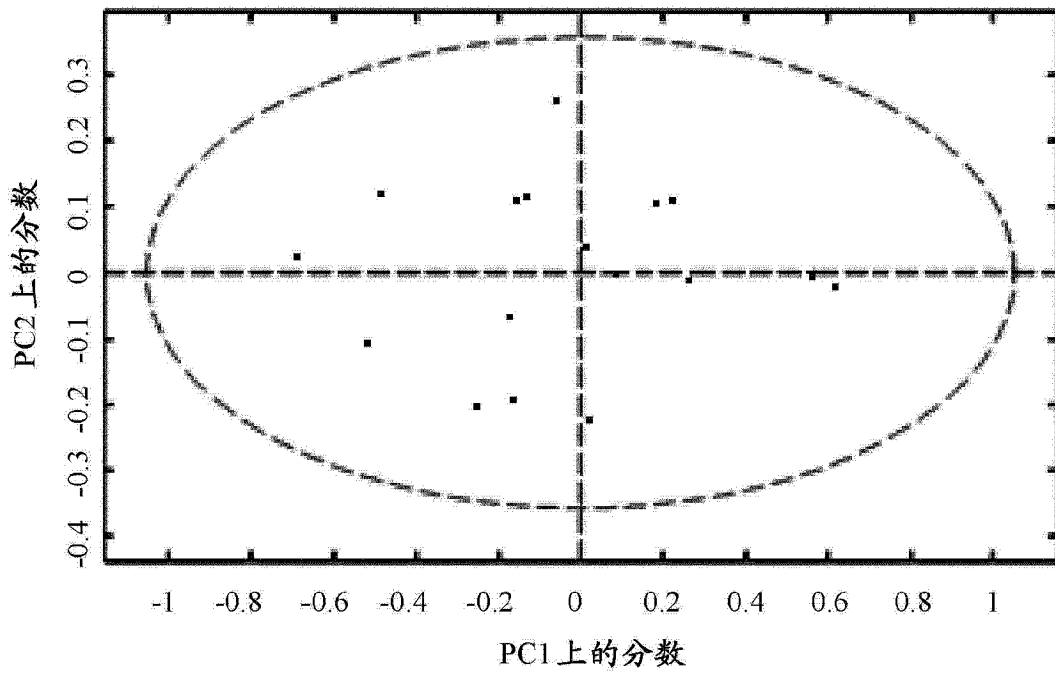


图 1

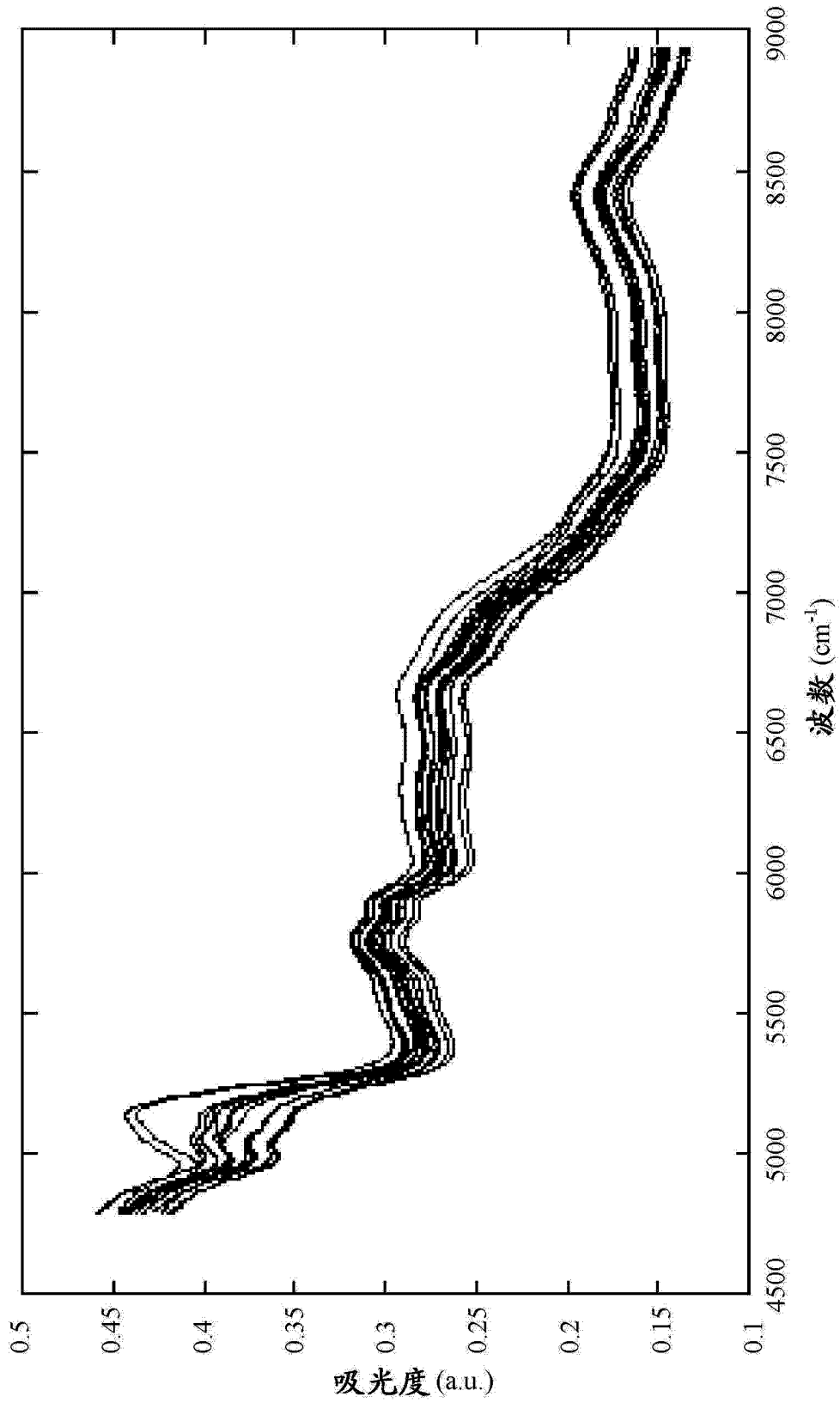


图 2

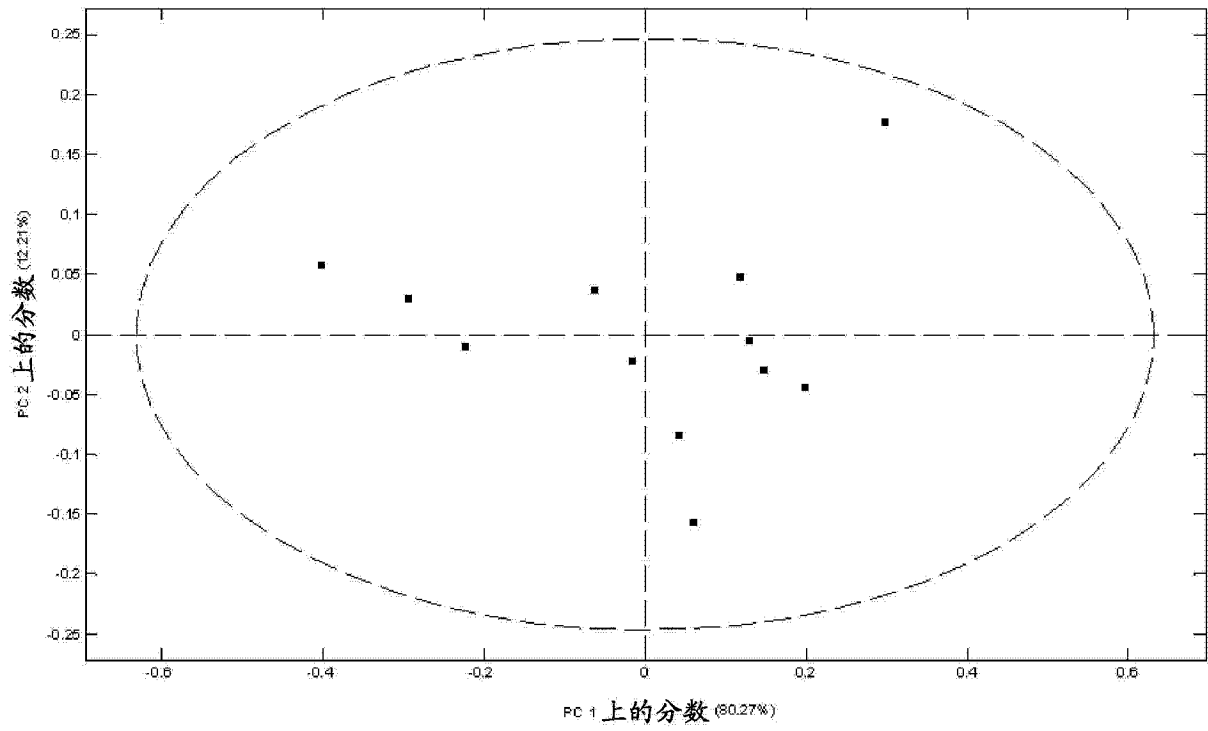


图 3

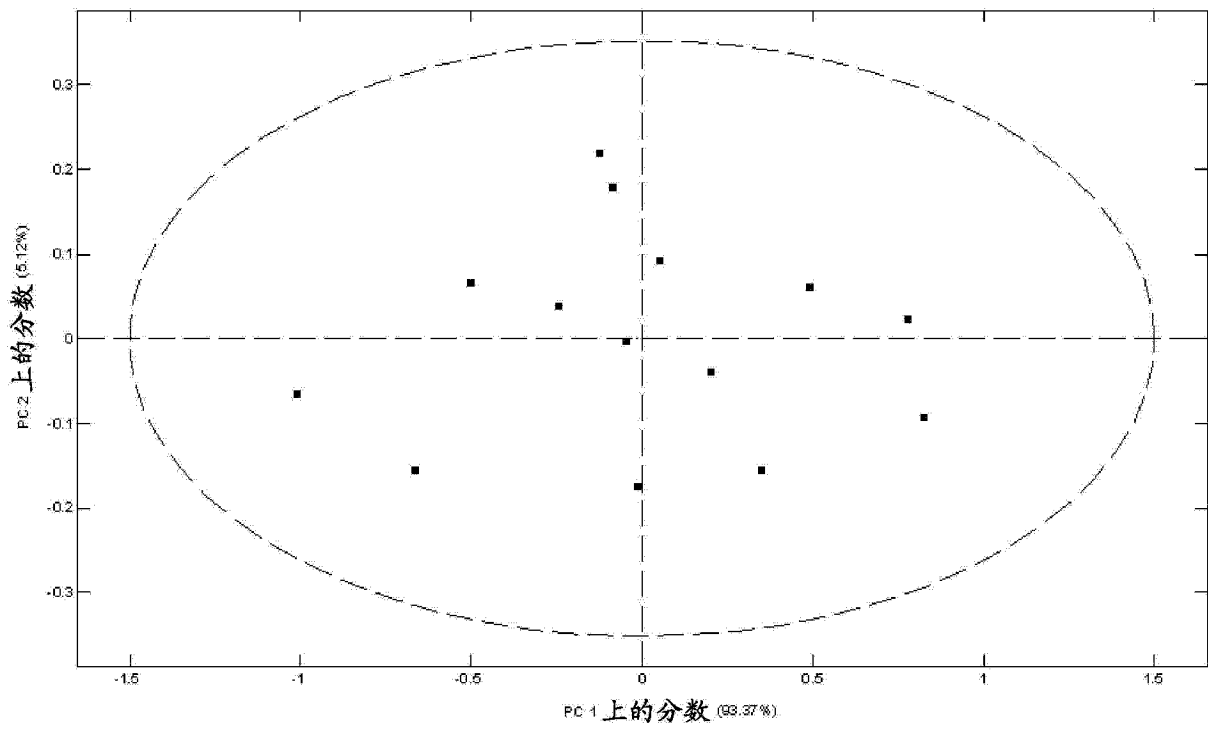


图 4

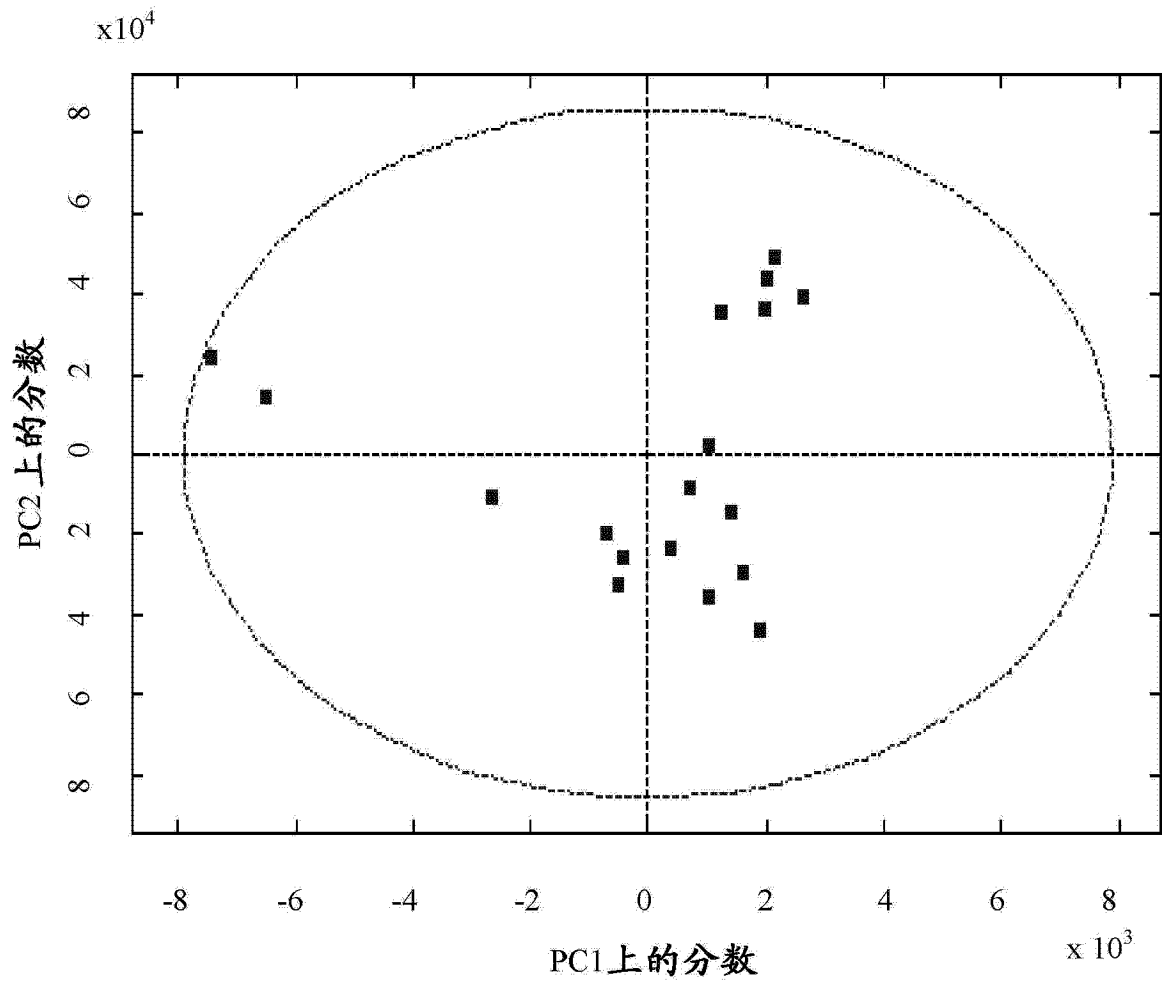


图 5

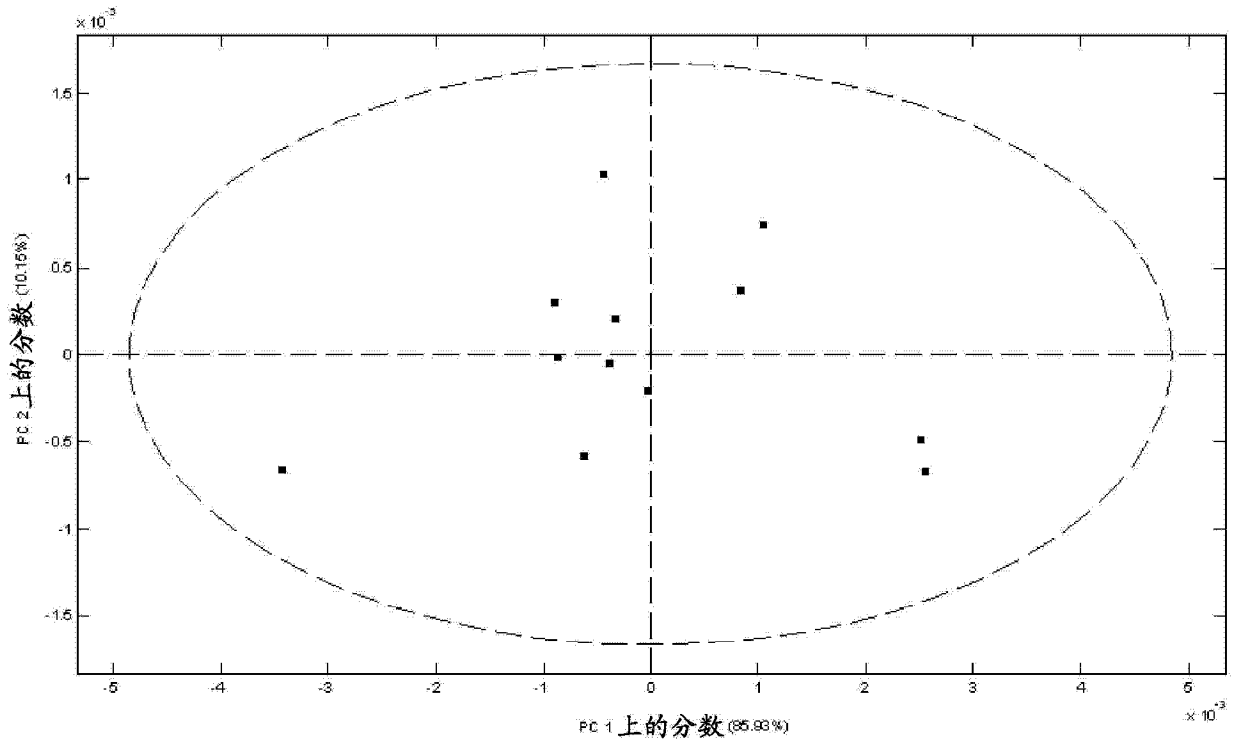


图 6

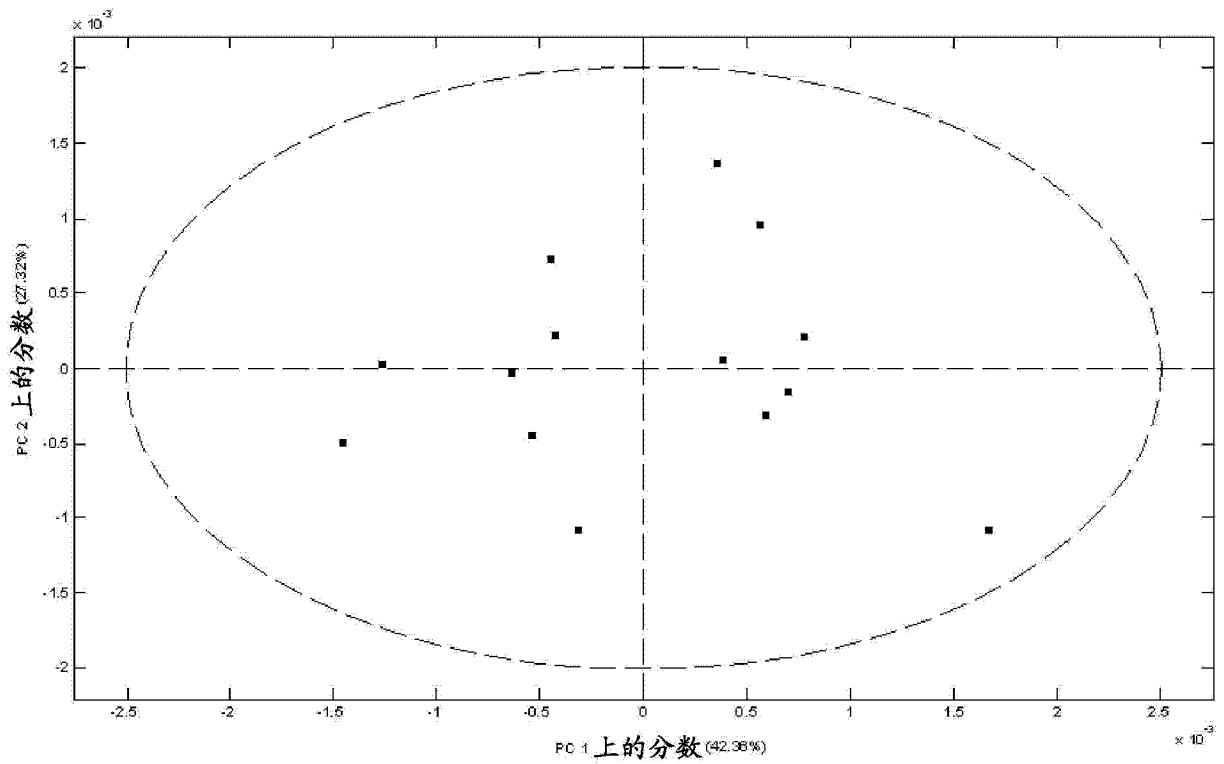


图 7

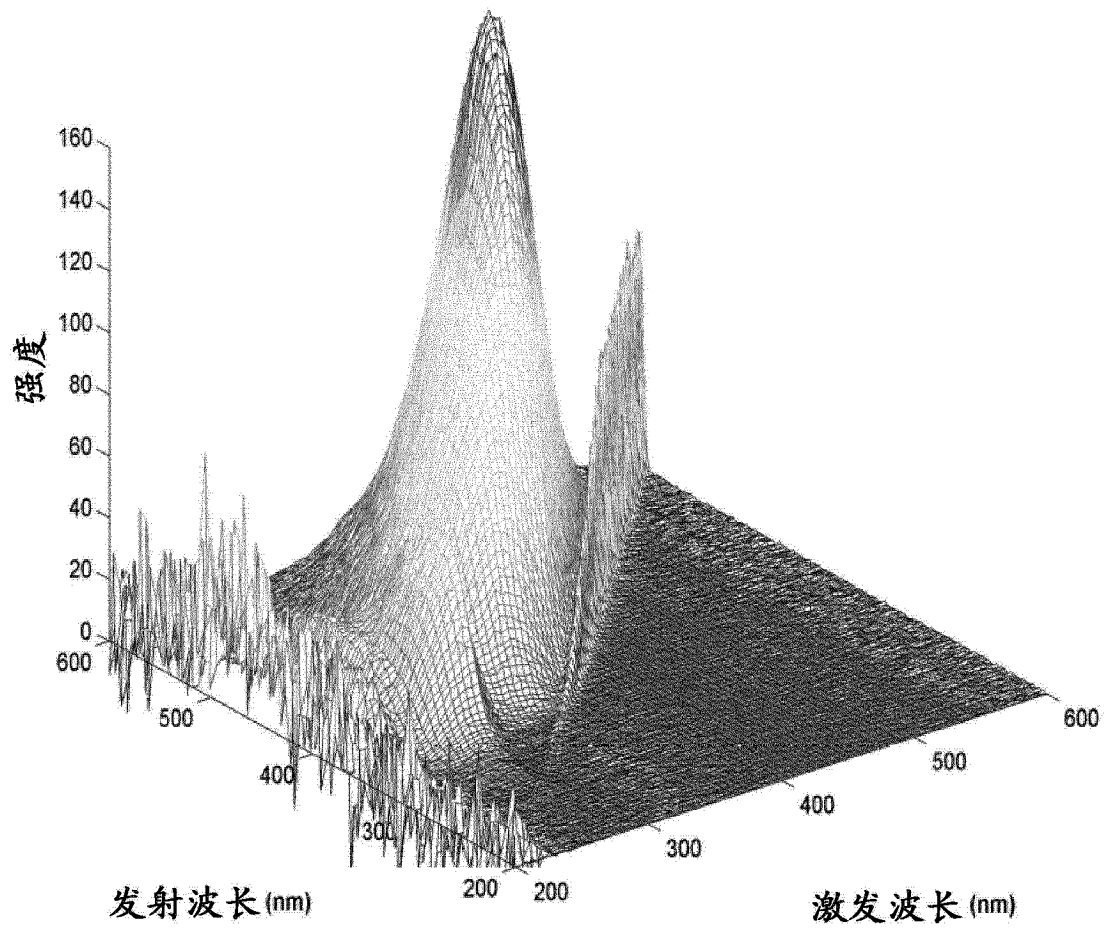


图 8

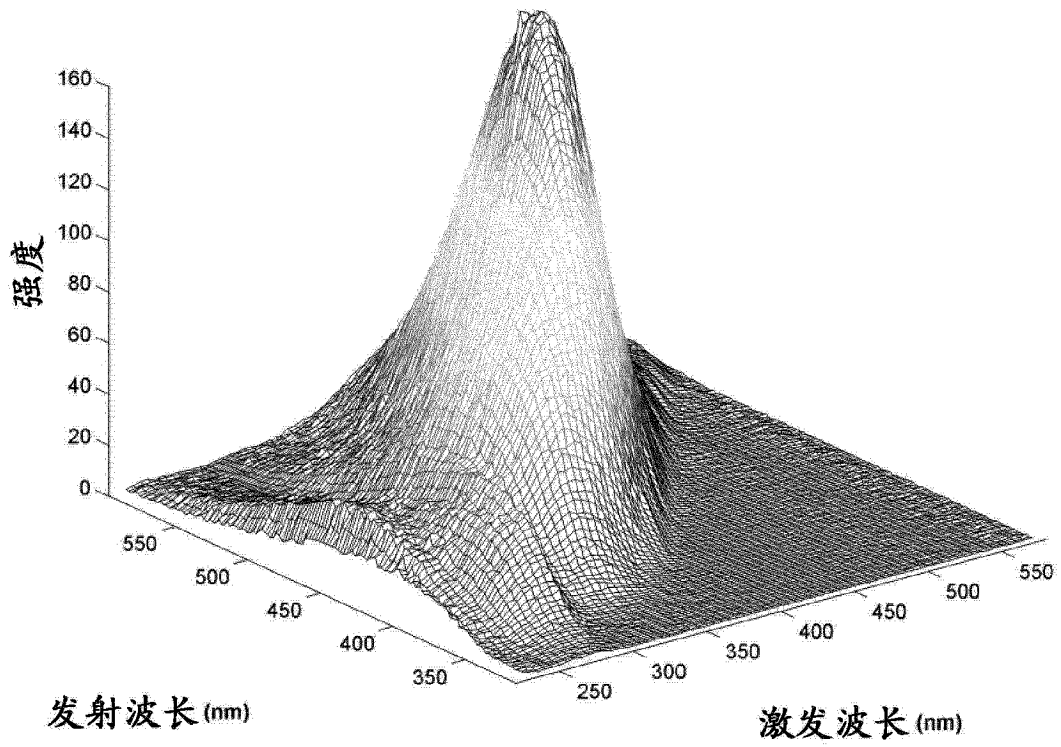


图 9

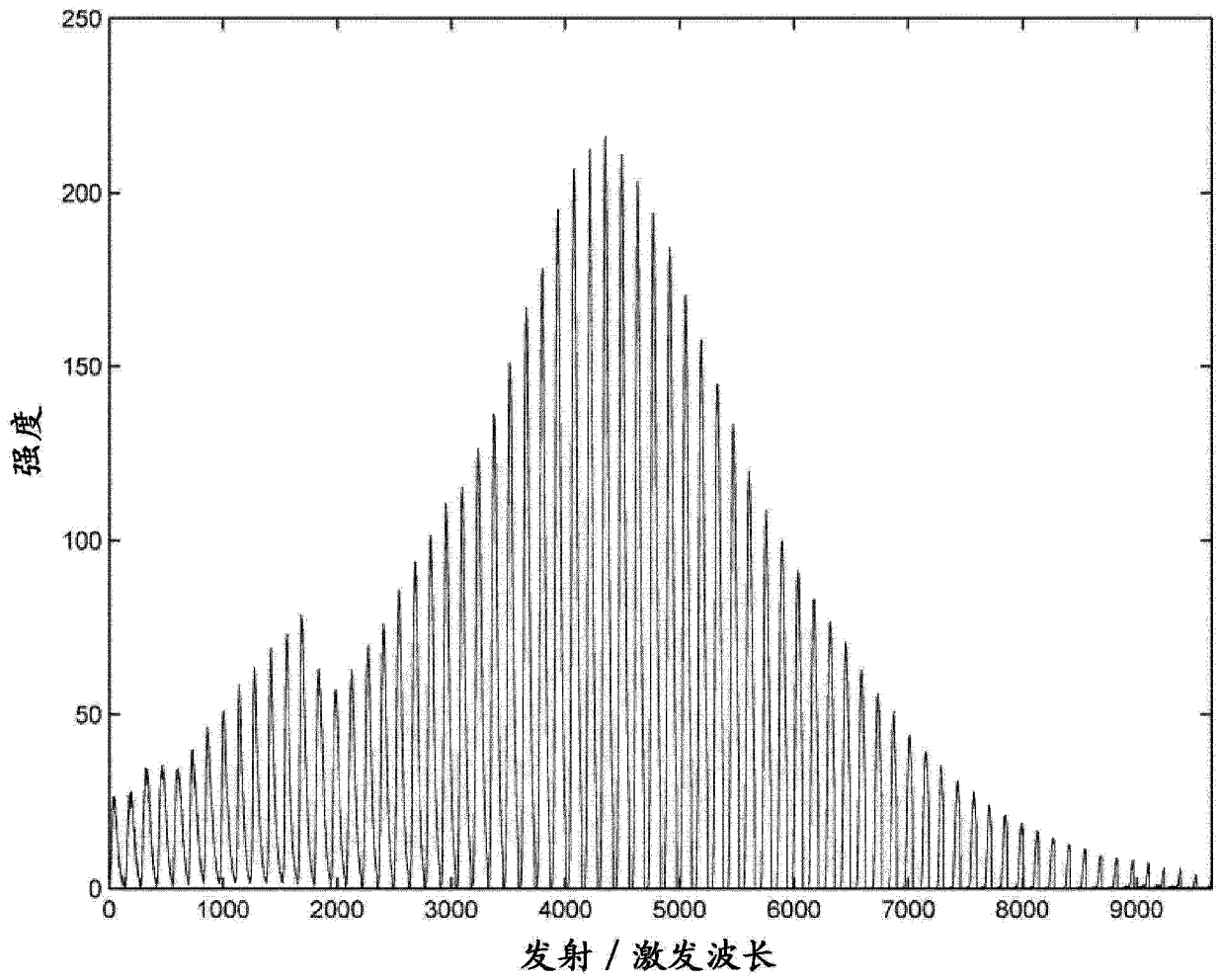


图 10

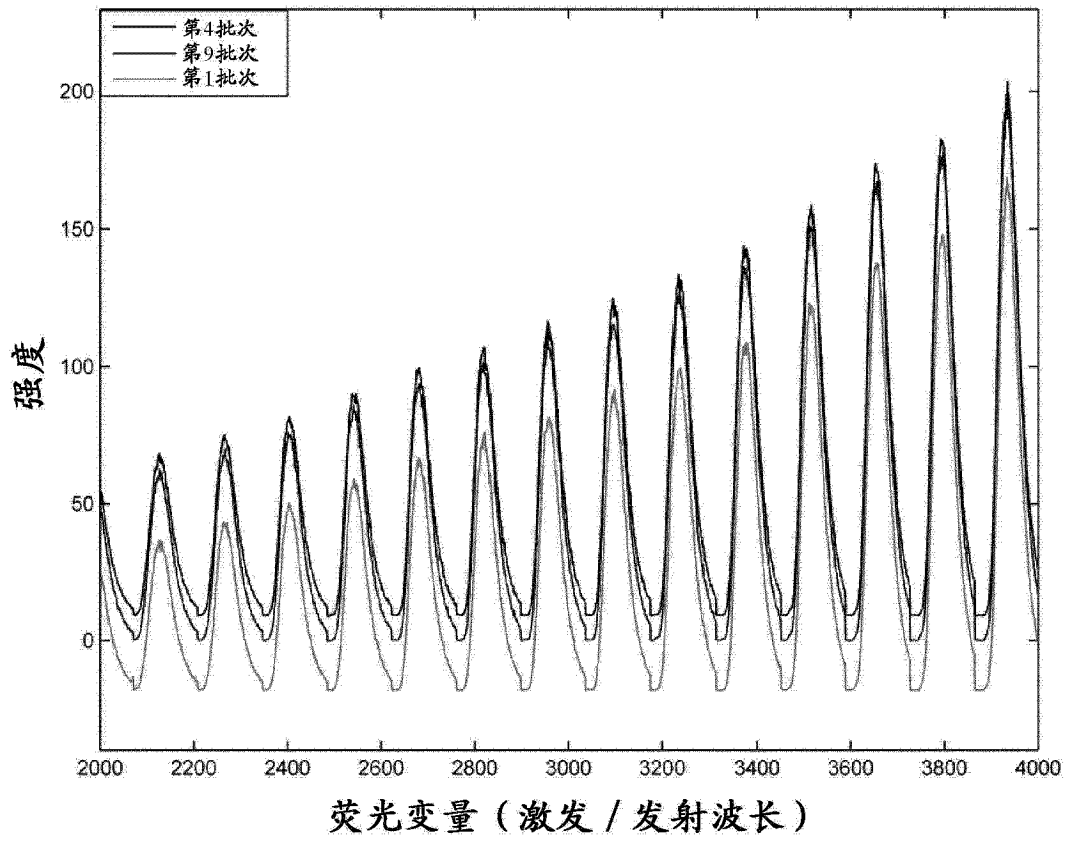


图 11

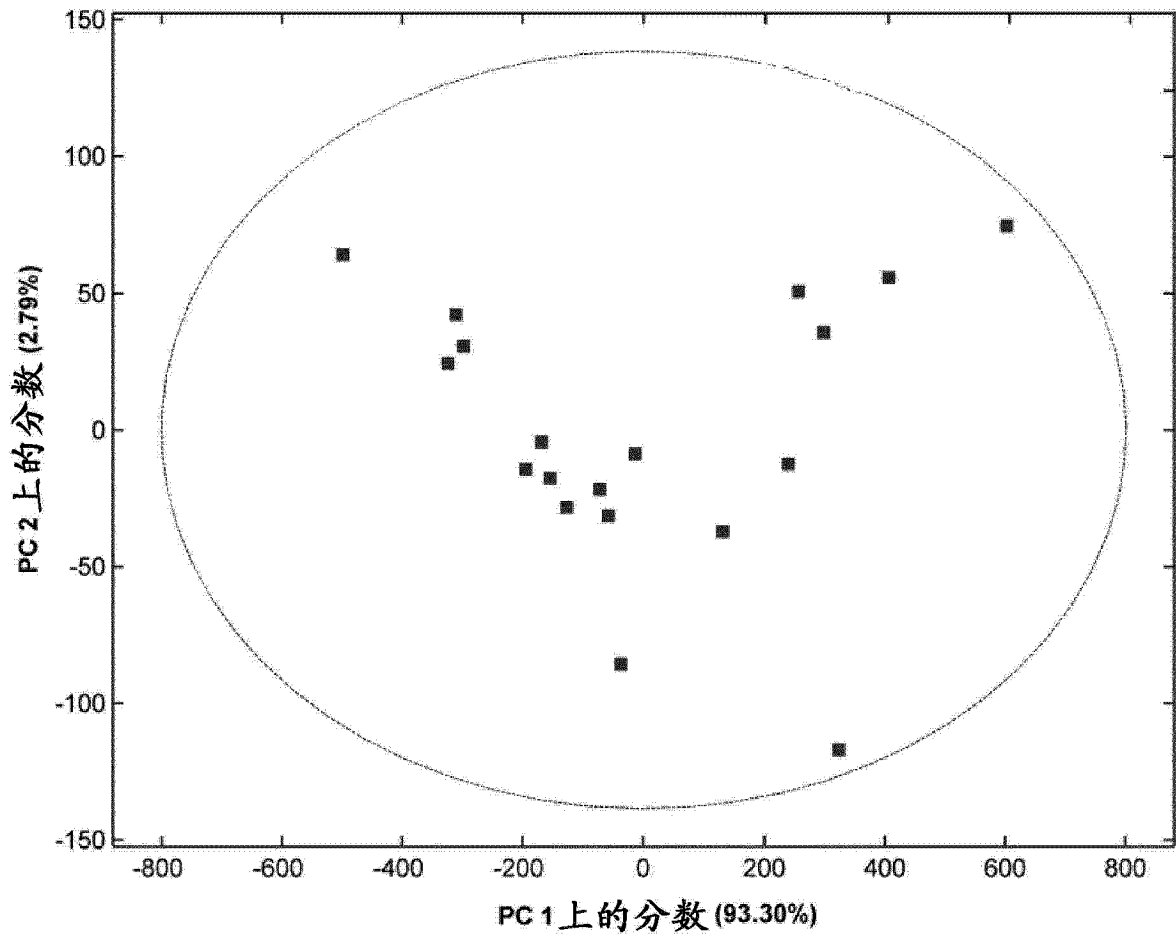


图 12

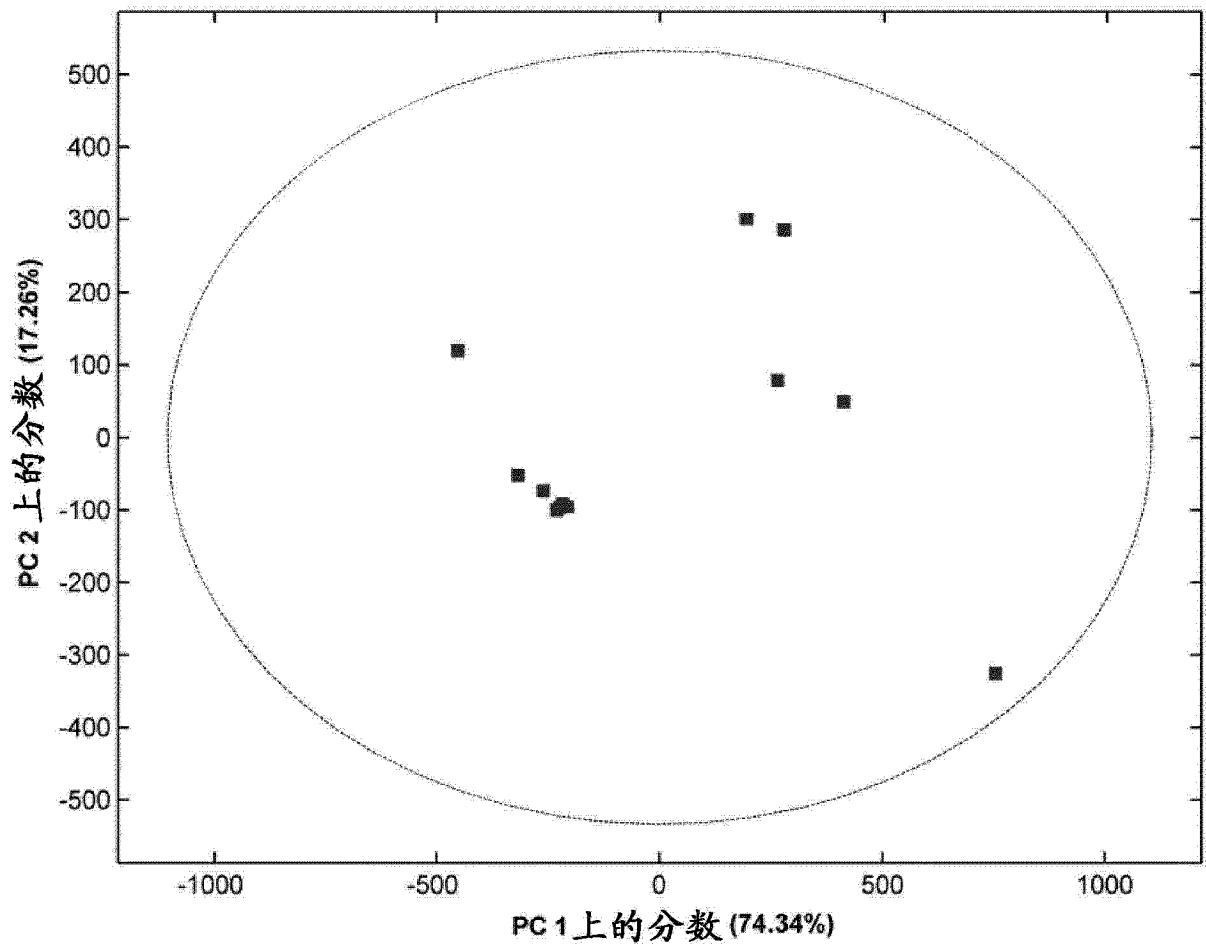


图 13

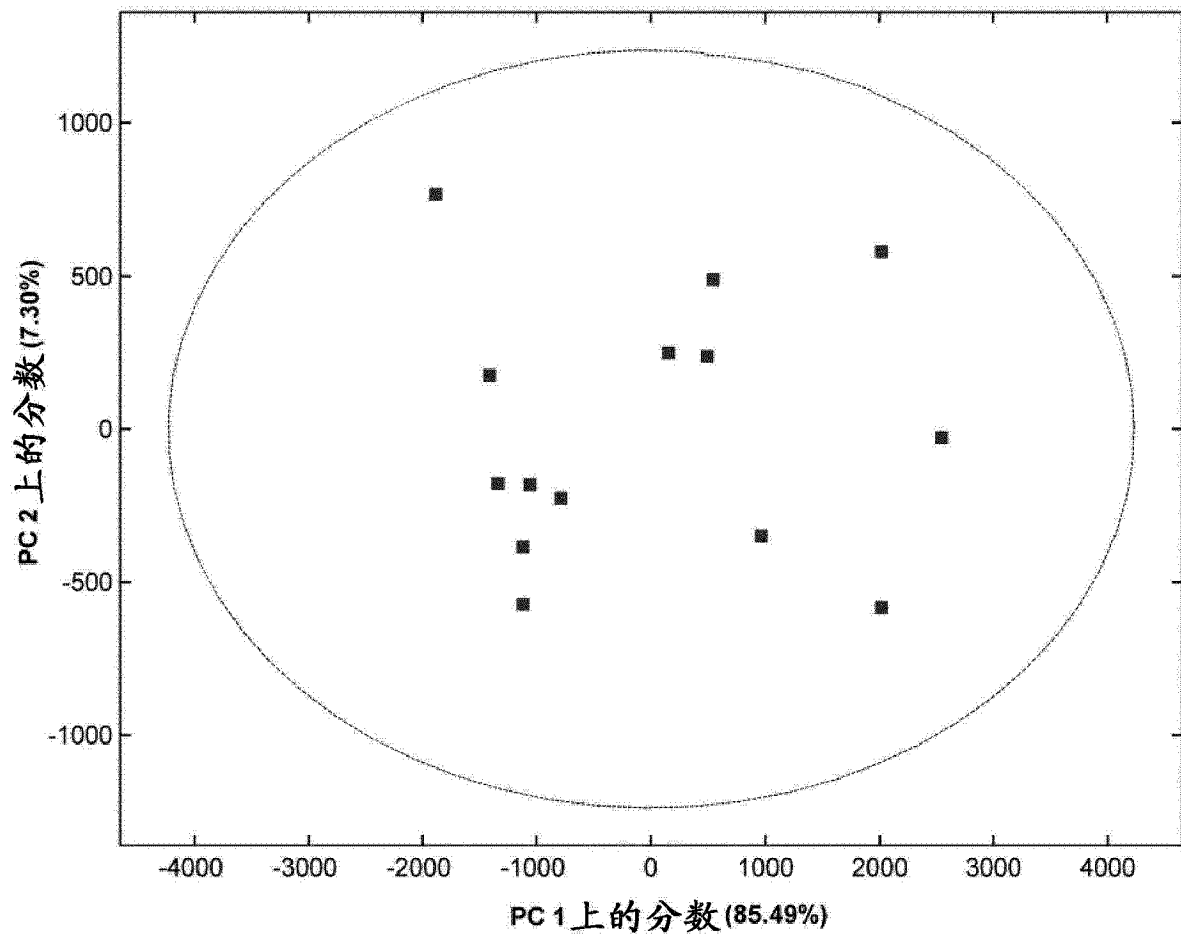


图 14

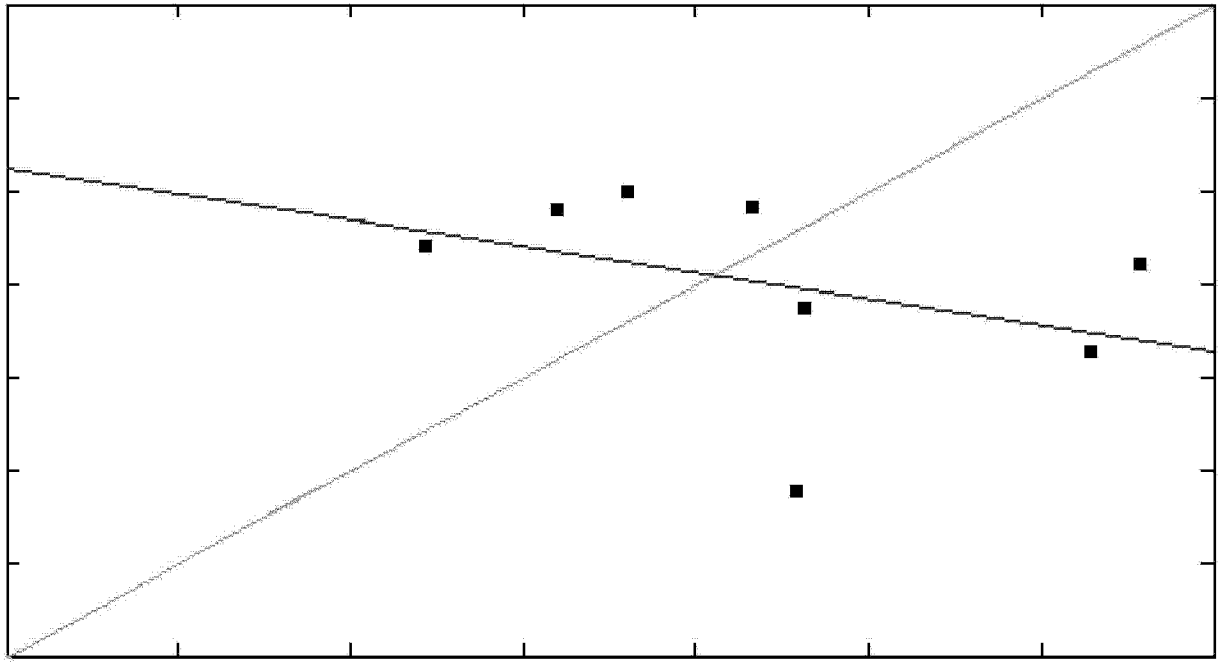


图 15

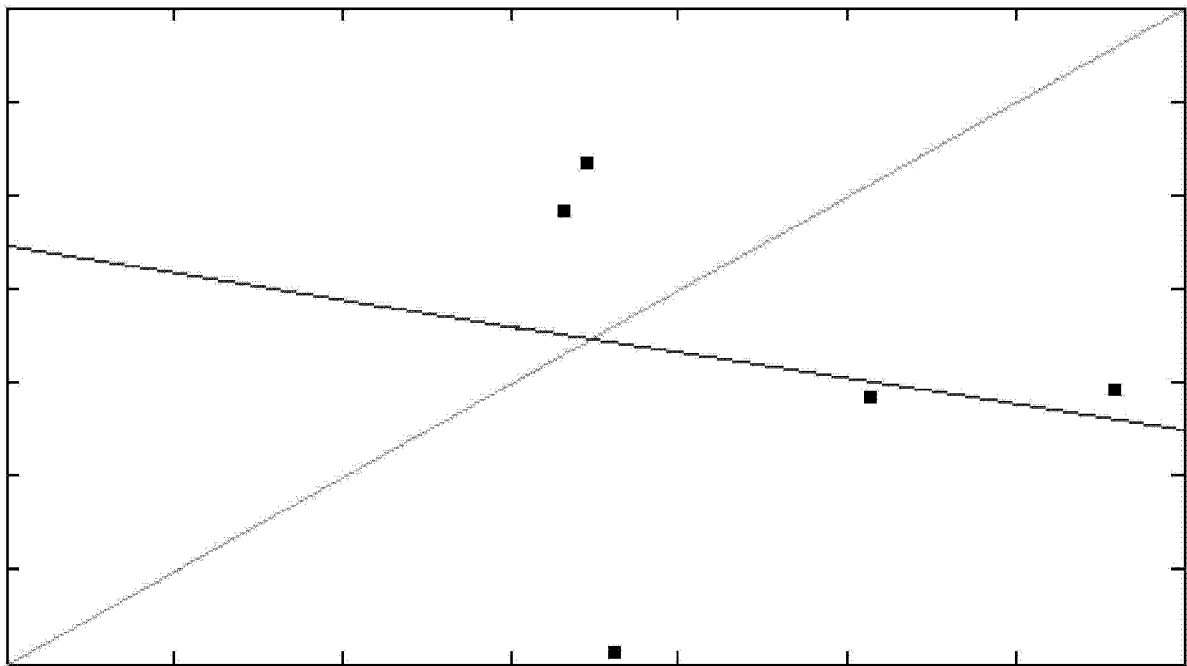


图 16

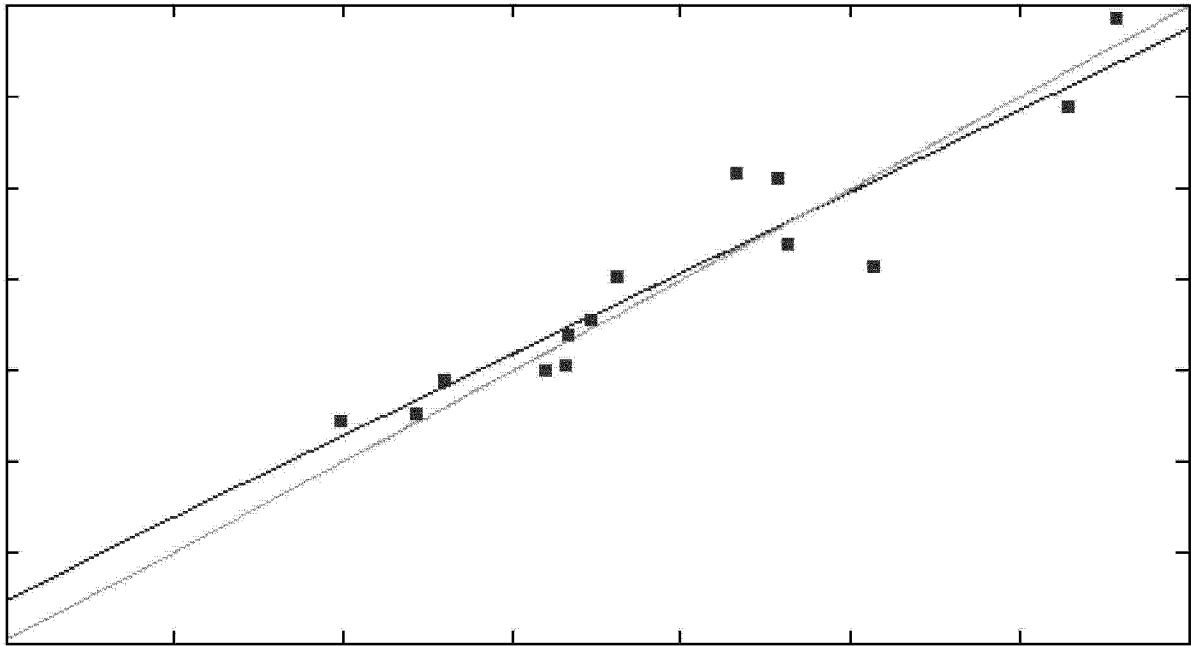


图 17

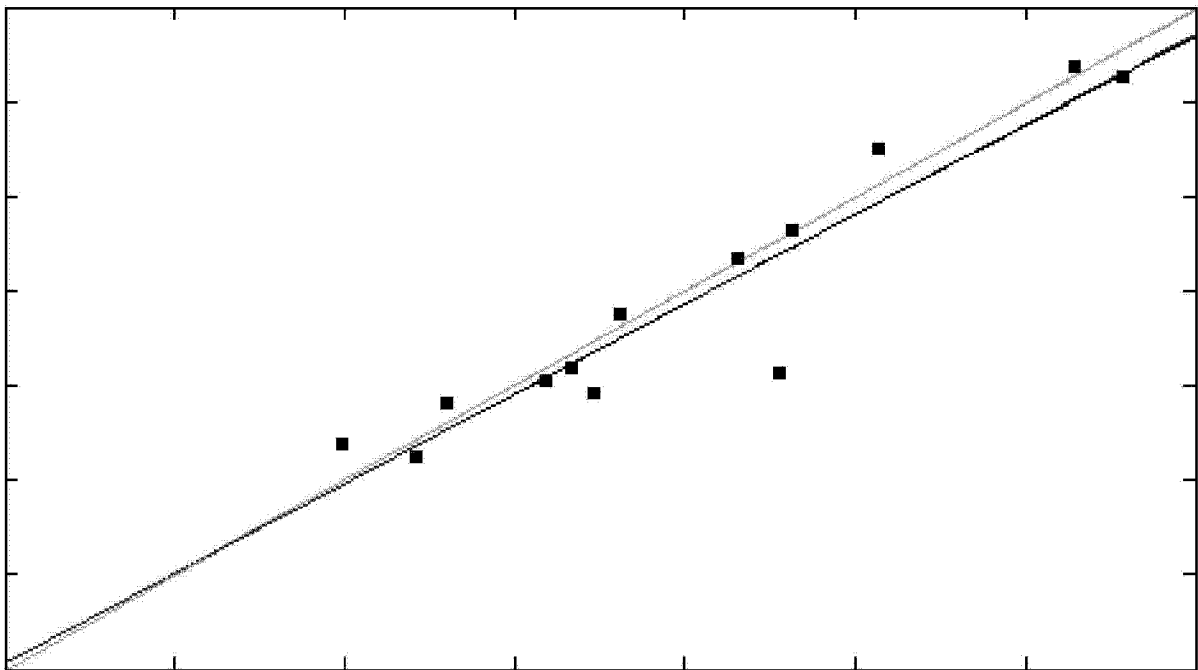


图 18

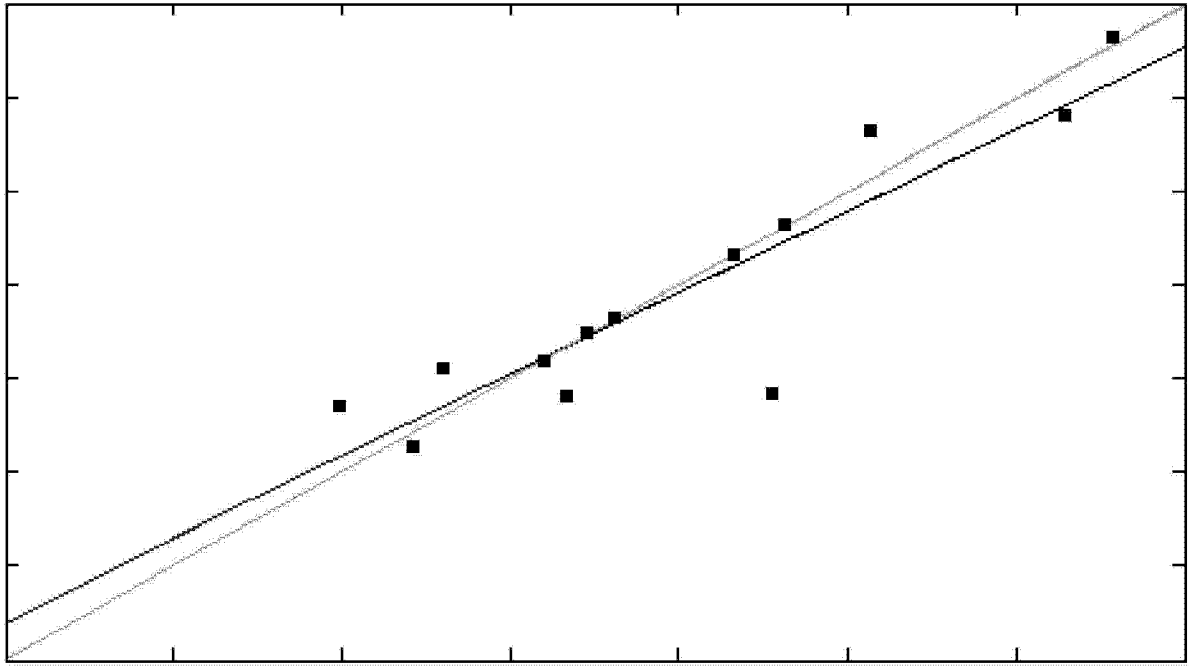


图 19

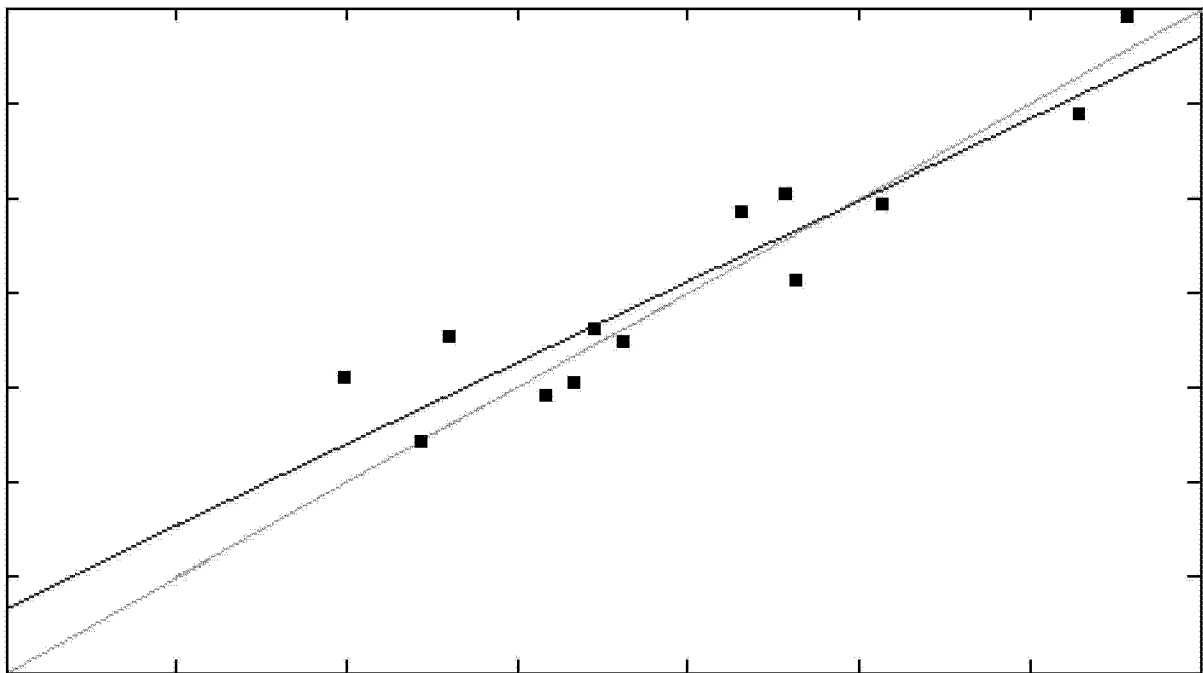


图 20

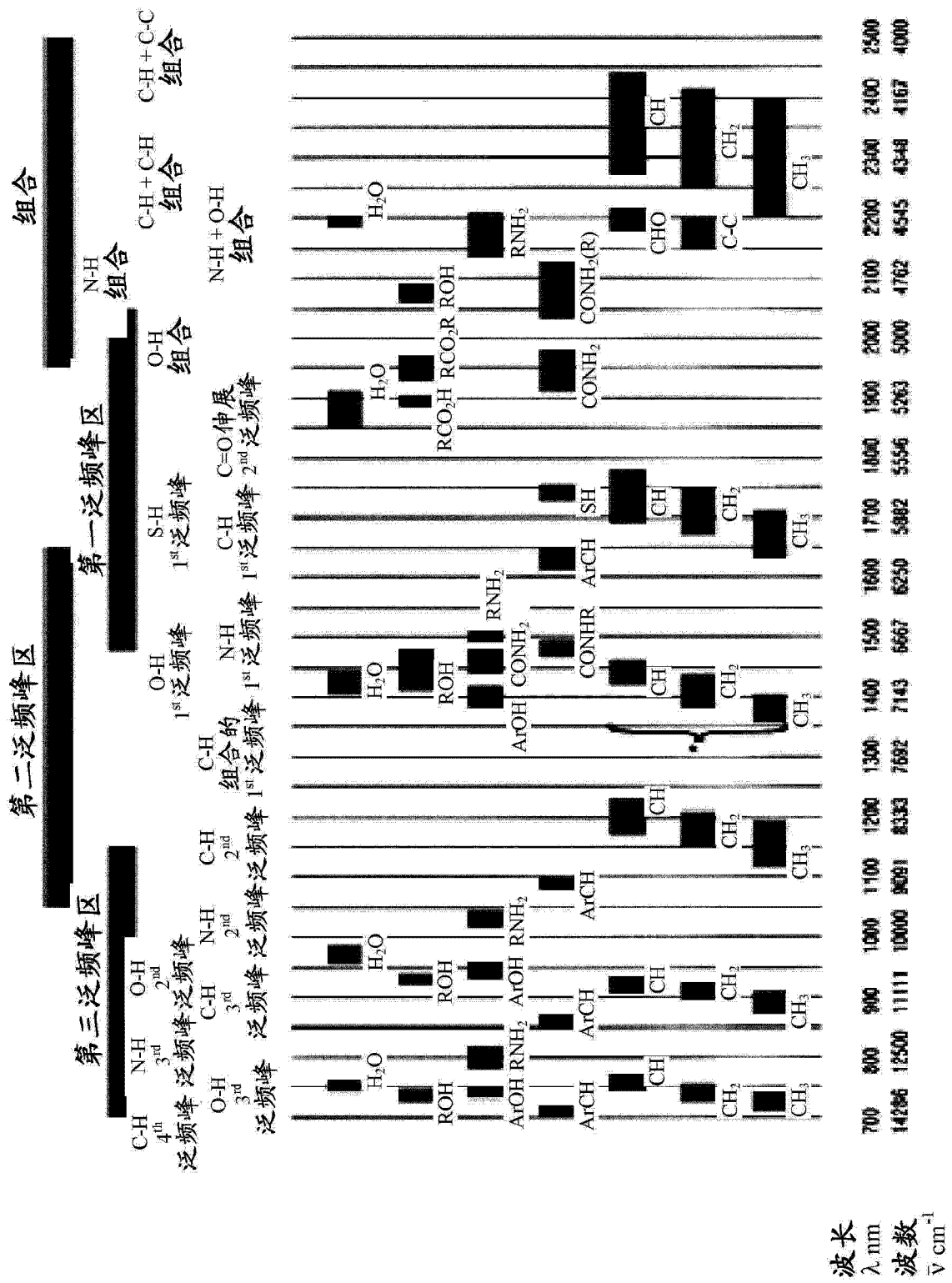


图 21