## 六味地黄(LWDH)中不确定化合物的来源调研

祖松鹏

2015/6/9

## 问题描述：

组内收集的六味地黄化合物共有311种化合物，其中不确定来源的有61种化合物。期望根据化学结构信息，判断这61种化合物是否来自其余化合物中，如是否为其结构的一部分，是否有结构相似的化合物。

## 方法：

利用ChemmineR，fmcsR等化学计算工具包，计算该61种化合物与其余250种化合物的最大共有子结构，以推测其可能的来源。

## 结果：

我们从250种化合物中，找出和61种化合物最大共有子结构相似性最高的相应61种化合物。从结构上判断，**61种化合物中共有33对可以匹配到相应的化合物**，见表格1. 这其中，1. 绝大部分的匹配为两种化合物只是在某个局部位置的官能团不一致，其可能为官能团的缺失，添加或变换。

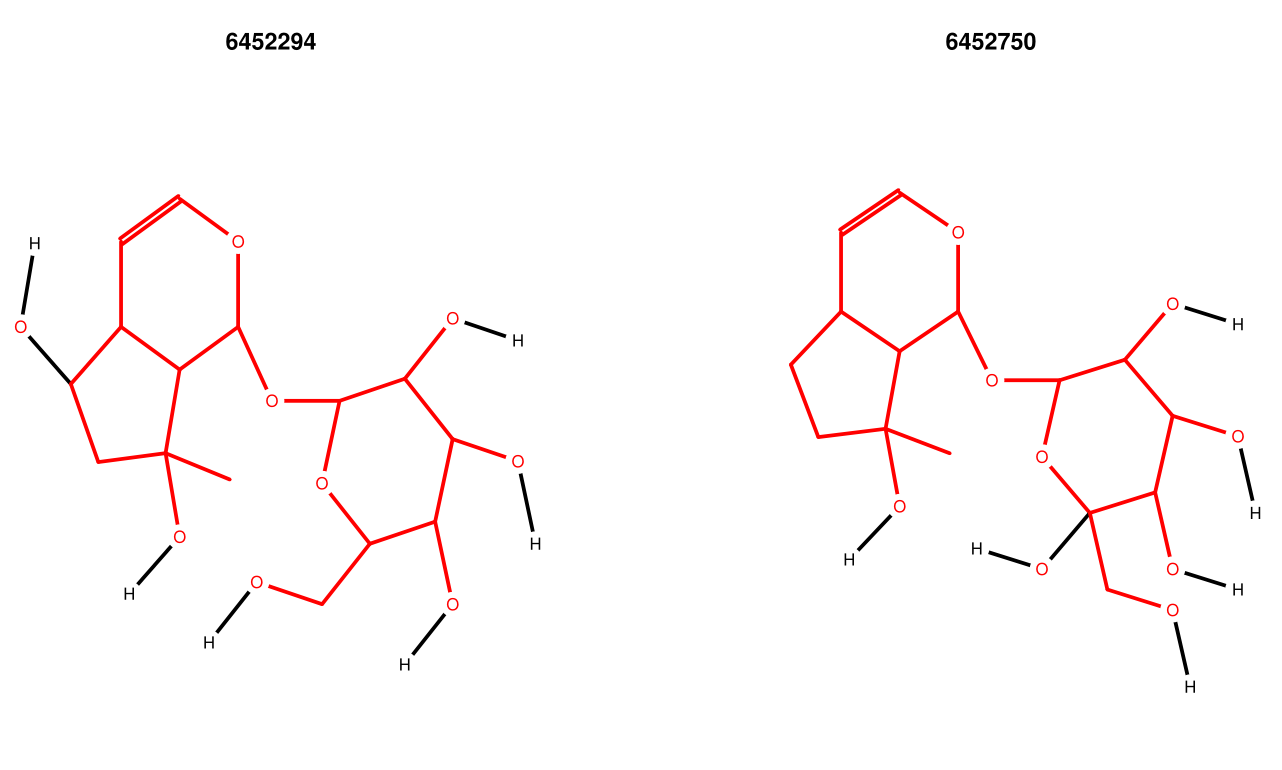
2. 而极少出现一个化合物为另一个化合物的片段。

3. 有个别化合物为另一种化合物含HCL的形式。

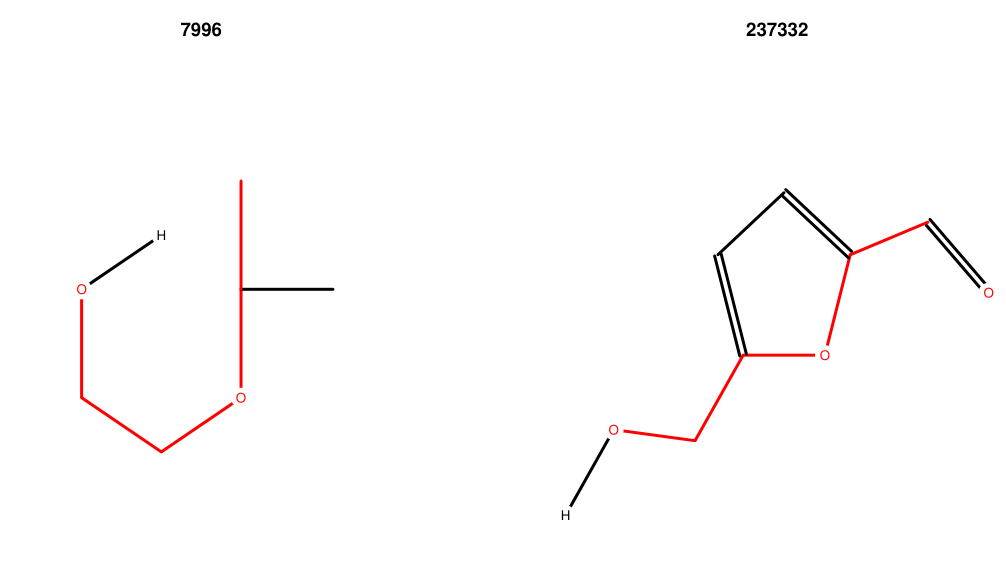
4. 有个别完全匹配。

图一给出了匹配的一种情况：两个化合物，红色为完全匹配的位置，二者仅仅为一羟基位置的改变。图二为一种不匹配的例子。在这种情况下，我们不能确定该化合物的来源。

在附件中，给出了，全部61对局部化学结构匹配结果图示。



图表 1 匹配的例子。数字为相应的CID编号。左边来自61种化合物的一种。



图表 2 不匹配的例子。数字为相应的CID编号。左边来自61中化合物的一种。

表格 1 可以匹配的33对化合物

|  |  |
| --- | --- |
| 化合物CID编号(来自61种不确定化合物) | 化合物CID编号(来自250种确定化合物) |
| 122159 | 442834 |
| 100199 | 442834 |
| 10458876 | 15558616 |
| 5316550 | 10005131 |
| 5317294 | 15558616 |
| 5318708 | 442420 |
| 5318707 | 442420 |
| 11968648 | 6325881 |
| 6452294 | 6452750 |
| 11968737 | 6325884 |
| 59707781 | 439242 |
| 5318130 | 3086648 |
| 6744 | 222284 |
| 15391340 | 56841132 |
| 69423 | 12401 |
| 12523 | 521556 |
| 11287 | 7002 |
| 65340 | 681 |
| 7269 | 13629 |
| 568 | 6853 |
| 123409 | 12366 |
| 3036237 | 56841132 |
| 158193 | 471122 |
| 8221 | 11635 |
| 124024 | 65238 |
| 5283387 | 445639 |
| 107905 | 9064 |
| 294705 | 12620 |
| 15979 | 19775 |
| 73179 | 16131156 |
| 46882884 | 46882880 |
| 22116985 | 3036133 |
| 7361 | 237332 |