

虚拟活性化合物的自动生成

陈海峰 荔建锋 姚建华 袁身刚* 郑崇直

(中国科学院上海有机化学研究所 计算机化学开放实验室 上海 200032)

摘要 虚拟活性化合物的自动生成是从药效团和预先设定的结构碎片出发,通过碎片选择、碎片组装和柔性构象搜索来获得虚拟活性结构.经过对 HIV-1 蛋白酶抑制剂药效团进行虚拟活性化合物生成,得到了 16 个虚拟活性化合物,通过构象分析发现生成的化合物满足药效团的限制条件.说明这一方法能够有效的生成虚拟活性结构,与药效团检索结果对比发现生成的虚拟活性结构新颖易于合成.

关键词 虚拟活性化合物,自动生成,药效团,HIV-1 蛋白酶抑制剂

Automatic Generation of Virtual Bio-Active Compounds

CHEN Hai - Feng LI Jiang - Feng YAO Jian - Hua

YUAN Shen - Gang* ZHENG Chong - Zhi

(Laboratory of Computer Chemistry, Shanghai Institute of Organic Chemistry,
Chinese Academy of Sciences, Shanghai, 200032)

Abstract The described automatic generation of virtual bio-active compounds starts from pharmacophore and pre-selected fragments. The generation process consists of three stages: selection of fragments, assemblage of fragments and conformation flexible searching. This method is demonstrated by an example that sixteen virtual bio-active compounds were generated starting from HIV-1 protease inhibitors pharmacophore in this work. The results of conformation analysis showed that the pharmacophore is contained in the favorite conformers of the generated structure. It was showed that this method could generate virtual bio-active structure effectively. In comparing of the results from 3D structure searching, the generated structures have novel structural features and they are easier to be synthesized.

Keywords virtual bio-active compounds, automatic generation, pharmacophore, HIV-1 protease inhibitors

组合化学和高通量筛选技术的出现,大大加速了新药的研制与开发.但是这类方法在获得活性化合物的同时会合成出许多无用化合物,造成资源浪费和对环境的污染.如果能采用某种理性的方法先找到具有某种特定生物活性的化合物的结构,再合成这个化合物和药理测定,而不是先合成再确定它

有无活性,将有可能摒弃传统的先大量合成再筛选的低效方法,从而大大提高新药的发现效率.虚拟活性化合物的自动生成就是这类理性方法之一.虚拟活性化合物的生成一般有两种方法^[1],一是当受体结构与作用位点已知时,可采用生成与受体互补结构的方法,这种方法可以使虚拟配体与受体配合良

* E-mail: yuansg@pub. sioc. ac. cn

收稿日期:1999-10-27,修回日期:1999-12-07,定稿日期:1999-12-26,国家科技部“九五”重点攻关项目(96-547-01-01,96-547-01-02)、国家基础研究发展规划(G1998051115)和中法先进研究计划(PRA 98M-A036),国家自然科学基金重点项目(29832050),国家自然科学基金(29872048),国家教委留学回国人员科研基金项目,上海市科技开发基金(98BSH1426)资助项目.

(Received October 27, 1999. Revised December 7, 1999. Accepted December 26, 1999)

好;二是当仅知道药物的药效团结构时,可生成包含药效团的虚拟化合物结构.一般具有相同作用机理的一类药物都包含一些特定的起药效作用的基团,并且满足一定的空间限制条件.如果从这些条件出发,进行计算机的虚拟生成,有可能得到有前途的活性化合物的先导结构.由于结构生成通常是在一个组合空间中进行,需要在几乎无穷多个方向进行搜索,如果不通过约束来限制搜索空间将是不可能实现的.因此,避免“组合爆炸”是虚拟化合物生成的关键.本系统从碎片的选择,碎片的组装以及柔性构象搜索等方面实现了虚拟活性结构的自动生成.在此基础上,尝试了由 HIV-1 蛋白酶抑制剂药效团进行虚拟活性结构生成,得到了一些结构新颖的化合物,而且与药效团检索^[2]结果进行了对比,以考察虚拟活性结构生成系统的实用性.

1 生成方法

1.1 碎片选择

碎片分为疏水碎片和亲水碎片两类,分别采用 SYBYL 6.4^[3] 软件进行结构优化,得到它们的优势三维结构,存放于基本碎片库中.为了保证虚拟活性结构的多样性,对库中碎片不做预先筛选,而是采用随机数发生器来选择碎片的种类与数量.此外,还建立了药效团的结构库,与基本碎片库分别存放,以利于碎片的管理与扩充.

1.2 碎片组装

对确定种类与数量的结构碎片,采用遗传算法决定碎片间的连接关系,然后根据连接关系进行碎片的拼装.如果能对这些碎片进行合理的拼装,就可以得到完整的三维结构,视作虚拟活性结构.

拼装碎片时,根据连接顺序,依次把后一个碎片拼接到前一个碎片上去,相邻的两个碎片其拼接部分的键性必须相同,基本过程如图 1 所示.经过旋转和平移,先将 C(2)—H(3) 与 H(1')—C(2') 重合,然后按照成键的规则旋转 C[2(1')]—C[3(2')] 键,得到一个稳定构象.

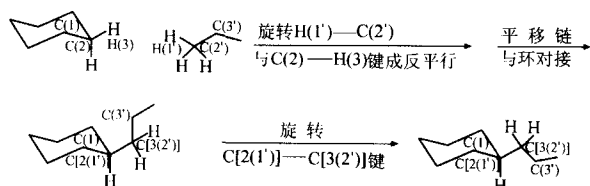


图 1 碎片拼装示意图

在拼装过程中必须尽可能使构象的能量处于较低的状态,例如把取代基放在环的平伏键上,大基团间处于反式等.拼装完毕后再对整个分子的构象进行合理性检查,用范德华能量(van der Waals)作能量限制.以 Lennard-Jones 6-12 势函数计算分子中原子之间的范德华相互作用能,并且忽略最短路径上连接键少于 4 个的原子对.如果能量比较高,对于上图中的分子就旋转 C[2(1')]—C[3(2')] 键,直至得到比较合理的构象.

$$E_{\text{vdw}} = \sum_{i=1}^{S_{\text{atom}}} \sum_{j=1} E_{ij} (1.0/a_{ij}^{12} - 2.0/a_{ij}^6) \quad (1)$$

其中, S_{atom} 是原子总数, E_{ij} 是范德华常数,其值等于 $\sqrt{E_i E_j}$, E_i, E_j 是力场参数, $a_{ij} = r_{ij}/(R_i + R_j)$, r_{ij} 是原子 i 和 j 的距离(nm), R_i 是原子的范德华半径(nm).

1.3 柔性构象搜索

对于拼接成功的三维结构,再次利用遗传算法^[4].遗传算法包括染色体的复制、杂交、变异等操作.对于上述结构随机产生 50 个构象作为一个群体,根据可扭转键的数目确定染色体串的长度,然后从这 50 个构象群体中,任选两个个体进行杂交操作,以产生新的构象(染色体),构象是否被保留要根据新生成构象的适应值是否比父代构象高来取舍,适应值高则取代父代构象,否则保留父代构象.然后再从这 50 个构象中任选 5 个构象体进行变异操作,使整个群体既保持相对稳定,又能向最优解逼近.对于变异的子代构象的取舍,依然依赖于适应值的高低.设定进化代数为 20 代.当任一构象与我们给定的药效团的限制条件满足 RMS 阈值小于 0.002 nm 时就跳出循环,否则直至最大代数.如果遗传代数已达 20 代, RMS 仍然不满足给定条件,就结束本次遗传算法,重新选择碎片,进行下一次虚拟结构生成.

1.4 活性评估

通过柔性构象搜索得到的结构,需要采用空间效应、疏水参数等来评价所得结构的可能活性.如果包含输入药效团的结构在某一药效团位置附近没有大的基团,而生成的结构如果在该位置有一个较大的基团,那么该结构的活性就认为较低.同时,由于疏水参数影响肌体对药物的吸收,所以适中的疏水参数也是评价所得结构活性大小的一个重要参数.如果有构效关系模型,也可通过适用的模型预测虚拟化合物的活性,从而选择最优者.同样,在作用机理明确的情况下,还可通过更精确的量子化学方法计算虚拟化合物与配体的结合能,从而达到筛选的目的.本例中采用了药效团模型来评价虚拟化合物

的活性.

2 实验

文献[5]报道的 HIV-1 蛋白酶抑制剂的药效团如图 2 所示, 包含一个羰基氧作为氢键受体, 两个羟基可分别作为氢键给体. 距离限制的单位为 nm.

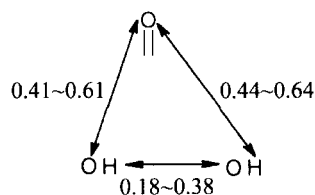


图 2 HIV-1 蛋白酶抑制剂药效团

输入上面的药效团及限制条件进行结构生成, 共得到 16 个虚拟活性结构, 耗时 56min. 图 3 所示列出了其中 10 个生成结构.

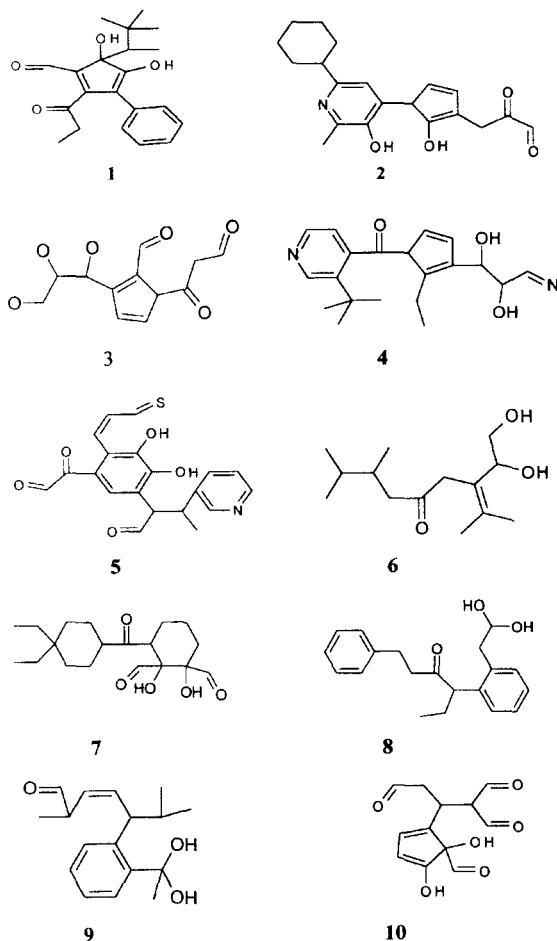


图 3 生成的虚拟活性化合物

为了检验生成的化合物是否满足输入药效团的限制条件, 以图 3 中的结构 5, 6 为例, 采用 SYBYL

6.4 软件研究它们的三维结构, 它们最低能构象的三维结构如图 4 所示. 羰基、两个羟基的三点药效团之间的距离分别为 0.514, 0.638, 0.283 及 0.542, 0.568, 0.293 nm, 都分别满足图 2 所示药效团的限制条件.

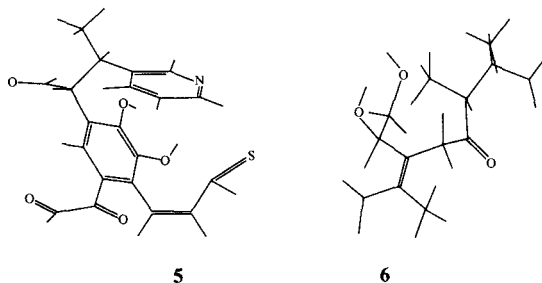


图 4 结构 5, 6 的最低能构象

3 药效团检索结果

对图 2 所示药效团在我们自己的未开发化合物库中进行药效团检索, 共得到 960 个命中结构, 其中图 5 是一个报道的具有抗 HIV 活性的化合物[6].

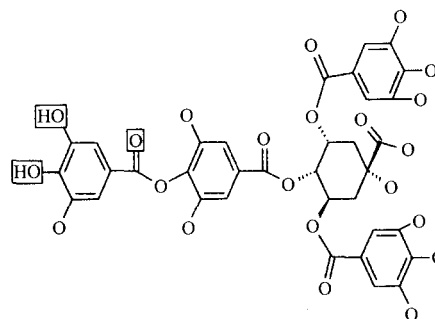


图 5 药效团检索的命中结构

比较图 5 的结构与图 3 中的结构 5, 可以发现它们都包含苯环及与苯环直接连接的羟基与羰基, 而这些基团对 HIV-1 蛋白酶的抑制活性可能具有比较重要的作用[7]. 说明药效团检索与基于药效团的虚拟活性结构生成, 都可以达到寻找活性先导化合物的目的. 而且通过结构生成所得的结构更加简单和更容易合成, 也可能获得新颖的结构.

4 结果与讨论

虚拟活性化合物的自动生成是从药效团和预先设定的碎片出发, 通过碎片选择、碎片组装和柔性构象搜索来获得虚拟活性结构. 经过实验证实, 这一方

法能够高效的生成虚拟活性结构,生成的结构能满足药效团的限制条件.而且通过与药效团的检索结果对比发现虚拟生成的结构更为简单和易于合成.更重要的是有时能产生全新的先导结构,这是另一常用的药物设计方法,是三维结构检索不具备的特别有意义的优点.

致谢 陈敏伯教授、董喜城博士、柴鸽庆硕士等对本文提出了许多宝贵意见,在此一并感谢.

References and Notes

- 1 V. J. Gillet, W. Newell, P. Mata, G. Myatt, S. Sike, Z. Zsoldos, A. P. Johnson, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 207.
- 2 CHEN Hai - Feng, YAO Jian - Hua, YUAN Shen - Gang, LI Jian - Feng, YANG Shuo, ZHENG Chong - Zhi, FAN Bo - Tao, A. Panaye, J. P. Doucet, *Computer and Applied Chemistry*, **1999**, *16* (2), 101 (in Chinese).
- 3 SYBYL [Computer Program], Version 6. 4, St. Louis (MO): Tripos Associates Inc., **1993**.
- 4 YUAN Shen - Gang, YAO Jian - Hua, CHEN Hai - Feng, ZHENG Chong - Zhi, *Computer and Applied Chemistry*, **1999**, *16* (1), 77 (in Chinese).
- 5 S. Wang, G. W. A. Milne, X. J. Yan, I. J. Posey, M. C. Nicklaus, L. Graham, G. Rice, *J. Med. Chem.*, **1996**, *39*, 2047.
- 6 M. Nishizawa, T. Yamagishi, G. E. Dutschman, W. B. Parker, A. J. Bodner, R. E. Kilkuskie, Y. C. Cheng, K. H. Lee, *J. Nat. Prod.*, **1989**, *52*, 762.
- 7 CHEN Hai - Feng, YUAN Shen - Gang, LUO Shi - Wei, DONG Xi - Cheng, YAO Jian - Hua, YANG Shuo, ZHENG Chong - Zhi, *Acta Chimica Sinica*, in Press (in Chinese).

(Ed. CHENG Biao)

(ZHENG Guo - Cheng)