磷脂酶 A₂ 家族的功能性分类研究*

刘振明 李 博 来鲁华

(北京大学化学与分子工程学院, 分子动态与稳态结构国家重点实验室, 北京 100871)

摘要 采用“结合强度指纹图谱分析”方法, 通过对多重分子对接得到的作用强度数据进行聚类矩阵分析对蛋白质进行功能分类. 着重研究了磷脂酶 A₂ 家族基于抑制剂作用强度的功能分类, 并且与基于序列的聚类结果进行比较, 成功地解决了序列比对方法不能处理的远源蛋白(cPLA₂)的分类问题.

关键词: 指纹图谱, 多重分子对接, 功能性蛋白质分类

中图分类号: O641

蛋白质功能预测与分类研究是生命科学与结构功能组学研究的一个热点领域, 对于药物靶点的识别、蛋白质功能注释以及蛋白质家族分类具有重要的意义^[1-2]. 目前, 利用计算手段进行蛋白质功能预测和分类的方法主要有两大类: 蛋白质序列同源性预测以及基于蛋白质结构域比对的功能预测方法. 序列同源性预测^[3]被证明是一类有效的研究方法并且得到了广泛的应用; 但对于某些进化上亲源性较远、序列有较大差别, 却具有相似功能的蛋白质无法做出正确的预测^[4]. 通过蛋白质结构域的比对来研究蛋白质的功能是一类基于三级结构的分析方法^[5], 可以用于序列相似性很差的蛋白质功能的相关性分析, 从而得到蛋白质结构域整体上的相似性信息以及功能相似性信息.

利用蛋白质结合部位的特征预测蛋白质的功能和进行蛋白质的分类是近年引起重视的一种蛋白质功能研究方法^[6-9]. 伴随着化学基因组学的发展, 利用小分子功能探针研究蛋白质并进行功能注释已经成为蛋白质功能研究的重要辅助手段^[10]. Greenbaum等^[11]利用结合强度指纹图谱方法进行了酶的分类研究, 主要是利用一系列的多肽类似物与木瓜水解酶家族中的十七种酶共价结合的实验数据建立亲和性指纹图谱, 并进行聚类分析. 他们也初步探索了利用分子对接计算建立亲和性指纹图谱并进行聚类分析的方法, 但与实验结果的一致性有待提高.

本研究针对磷脂酶 A₂(PLA₂)家族, 利用分子对接方法建立磷脂酶 A₂ 家族与结构多样性抑制剂的

亲和性指纹图谱, 并进行酶及抑制剂的聚类分析. 磷脂酶 A₂ 在自然界中分布广泛, 在蛇毒、蜂毒、动物胰脏、人的关节滑液以及某些组织细胞的细胞液中均存在^[12-13], 是研究脂代谢、凝血溶血机制的重要工具酶^[14-15]. 通常所说的 PLA₂ 一般是指 II 型 PLA₂, 相对分子质量较小, 一般在 1.4×10⁴(14 kD)左右; 来源不同, 序列略有差异. 另一类是 VI 型 PLA₂, 其代表为细胞液 PLA₂(cPLA₂), 特点是相对分子质量较大, 在 8.5×10⁴(85 kD)左右. 我们拟通过研究磷脂酶 A₂ 家族的功能聚类, 检验结合强度指纹图谱分析方法在蛋白质功能分类上的作用, 同时探讨蛋白质-配基多重相互作用矩阵的数据分析方法.

1 方法与流程

蛋白质结构的选取根据分辨率最优原则和计算的需要, 从来源(人、其他哺乳动物、蛇、昆虫)和类型(胰脏磷脂酶 A₂、关节滑液磷脂酶 A₂、细胞质磷脂酶 A₂、蛇毒磷脂酶 A₂、蜂毒磷脂酶 A₂)出发, 最终在 PDB 数据库中挑选了 12 个 PLA₂ 复合物晶体结构(见表 1). 不同类型和来源的磷脂酶 A₂ 在序列上和结构上有一定的差别, 特别是细胞质磷脂酶 A₂, 它本身相对分子质量较大, 无法在序列和结构上和其他磷脂酶 A₂ 进行归类. 其次是探针分子的选取, 研究中使用的是 MDDR(MDL Drug Database Report)数据库^[16]中有确切抑制活性数据的各类磷脂酶 A₂ 抑制剂以及文献报道的 Lilly 公司合成的磷脂酶 A₂ 抑制剂分子^[17-18], 共计 84 种.

2005-05-17 收到初稿, 2005-06-08 收到修改稿. 联系人: 来鲁华(E-mail: lhlai@pku.edu.cn; Tel: 010-62757486; Fax: 010-62751725). * 国家科技部及国家自然科学基金(20173001, 20228306, 90403001, 30490240)资助项目

表 1 选取的 12 种磷脂酶 A₂

Table 1 Twelve selected PLA ₂ s from PDB ^[19]			
ID	PDB ID	Source	Resolution(nm)
01	IMF4	Andaman Cobra	0.190
02	1SV3	Russell's Viper	0.135
03	1CJY	Human Cytosolic PLA ₂	0.250
04	1OXL	Russell's Viper	0.180
05	1SV9	Russell's Viper	0.271
06	1AE7	Mainland Tiger Snake	0.200
07	1POC	Bee Venom	0.200
08	5P2P	Porcine Pancreas	0.240
09	1KVO	Human Nonpancreatic sPLA ₂	0.200
10	1DB4	Human Nonpancreatic sPLA ₂	0.220
11	1BP2	Bovine Pancreas	0.170
12	1G4I	Bovine Pancreas	0.097

计算和数据处理采用的是多重分子对接策略和聚类分析. 多重对接是蛋白质与配基分子多相互作用的分子对接研究方法, 可以给出比单一对接过程更多的信息; 聚类分析是一种应用十分广泛的数据分析多元统计方法, 利用这种方法研究蛋白质相互作用的工作近年来也有报道^[20]. 研究中利用Dock 4.0^[21]进行分子对接计算(Maximum_orientations=500, Maximum_iterations=500, 每个对接过程输出排名靠前的 500 个构象, 其他参数默认), 采用 Dock4.0 提供的力场能量打分函数和 Score2.0^[22-23]评价不同结合构象的能量以及小分子与蛋白酶之间的结合常数. 将上述 12 个蛋白晶体复合物结构与 84 种抑制剂小分子进行对接后得到的能量打分建立指纹图谱矩阵(见图 1), 然后以小分子为变量对于蛋白质进行聚类分析, 或者以蛋白质为变量对于抑制剂进行聚类分析. 本研究中选用的数据聚类分析方法是 HCA(hierarchical cluster analysis)^[24].

2 结果与讨论

将得到的聚类树状图与应用序列比对得到的树状图(见图 2)比较, 发现在功能上蛇毒和细胞液

图 1 对各类磷脂酶 A₂ 抑制剂聚类的结果

Fig.1 Clustering result of PLA₂ inhibitors

PLA₂ 被分为一支, 人类关节滑液 PLA₂ 与从猪胰中提取出的 PLA₂ 被合并在一处. 细胞液 PLA₂ 由于其与其它 PLA₂ 的远缘性关系, 无法从序列比对中得出有用信息. 但从分子对接计算, 我们推测其底物结合部位的性质应该与蛇毒 PLA₂ 更为类似. 另外值得一提的是, 在序列比对的聚类结果中来自不同种蛇类的 PLA₂ 并没有像它们在功能分类中那样很好地归为一类, 说明它们在整体序列上有一定的差异, 但其底物结合位点又有很好的保守性, 因此表现出相似的功能活性. 这也从一个侧面说明功能性分类更有利于进行蛋白质功能研究. 对蛋白质的多重对接分类很好地再现了各种不同来源的 PLA₂ 在功能上的亲疏关系, 并且解决了序列比对方法不能处理的远源蛋白分类问题(细胞液 cPLA₂).

这种基于分子对接的指纹图谱分类方法有着很好的容错性. 我们在选取蛋白时, 特意加入了对于同一种蛋白的不同分辨率结构, 在聚类结构中这些结构也聚在了一起, 说明我们不必强求选取的结构具有一致的分辨率. 由于在蛋白晶体结构数据库中, 各种蛋白结构的分辨率千差万别, 这种宽容的选择尺度就给实际操作带来了很大便利. 另外从统计方法的角度看, 只要变量数目足够多, 个别数据的错误并不会对聚类结果产生很大影响, 这就在很大程度上减轻了分子对接计算的假阳性对聚类的影响, 也说明了利用计算得到的亲合性指纹图谱进行分类研究是可行的.

此外, 以 PLA₂ 作为分类变量, 对 84 个抑制剂小分子也进行了多重对接聚类分析(图 1), 结果区分出了一类对于人类非胰脏磷脂酶 A₂ 专一性抑制较好的小分子群. 在这 84 个小分子当中, 蛇毒和 cPLA₂ 抑制剂的选择性通常不好, 而哺乳动物

图 2 磷脂酶 A₂ 家族聚类结果

Fig.2 Clustering result of PLA₂s family

sPLA₂ 抑制剂则具有较好的专一性.

3 结 论

利用“结合强度指纹图谱分析”方法,成功解决了磷脂酶 A₂ 家族中序列比对方法不能处理的远源蛋白(cPLA₂)的分类问题,同时得到了基于相互作用的 PLA₂ 家族的功能性分类.相同来源但不同分辨率结构的相同分类说明了我们的分类方法具有相当好的宽容性,可以容忍不同分辨率甚至计算结果假阳性的影响.最后我们还对小分子抑制剂进行了类似的分类研究,成功地聚类得到了对人类 sPLA₂ 有选择性抑制作用的小分子集合.

References

- Gabalton, T.; Huynen, M. A. *Cell Mol. Life Sci.*, **2004**, **61**(7-8): 930
- Whisstock, J. C.; Lesk, A. M. *Q. Rev. Biophys.*, **2003**, **36**(3): 307
- Baxter, S. M.; Fetrow, J. S. *Curr. Opin. Drug Discov. Devel.*, **2001**, **4**(3): 291
- Dobson, P. D.; Cai, Y. D.; Stapley, B. J.; Doig, A. J. *Curr. Med. Chem.*, **2004**, **11**(16): 2134
- Pazos, F.; Sternberg, M. J. *Proc. Natl. Acad. Sci.*, **2004**, **101**(41): 14754
- Arnold, J. R.; Burdick, K. W.; Pegg, S. C.; Toba, S.; Lamb, M. L.; Kuntz, I. D. *J. Chem. Inf. Comput. Sci.*, **2004**, **44**(6): 2190
- Zrailev, S.; Farnum, M. A. *Proteins*, **2004**, **57**(4): 711
- Chalk, A. J.; Worth, C. L.; Overington, J. P.; Chan, A. W. *J. Med. Chem.*, **2004**, **47**(15): 3807
- Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. *J. Mol. Biol.*, **2004**, **339**(3): 607
- Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. *Chem. Biol.*, **1995**, **2**(2): 107
- Greenbaum, D. C.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Bromme, D.; Kuntz, I. D.; Bogoy, M. *Chem. Biol.*, **2002**, **9**(10): 1085
- Yedgar, S.; Lichtenberg, D.; Schntizer, E. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids*, **2000**, **1488**: 182
- Bernard, P.; Scior, T.; Didier, B.; Hibert, M.; Berthon, J. Y. *Phytochem.*, **2001**, **58**(6): 865
- Six, D. A.; Dennis, E. A. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids*, **2000**, **1488**: 1
- Murakami, M.; Kudo, I. *J. Biochem. (Tokyo)*, **2002**, **131**(3): 285
- <http://www.mdl.com/>
- Dillard, R. D.; Bach, N. J.; Draheim, S. E.; Berry, D. R.; Carlson, D. G.; Chirgadze, N. Y.; Clawson, D. K.; Hartley, L. W.; Johnson, L. M.; Jones, N. D.; McKinney, E. R.; Mihelich, E. D.; Olkowski, J. L.; Schevitz, R. W.; Smith, A. C.; Snyder, D. W.; Sommers, C. D.; Wery, J. P. *J. Med. Chem.*, **1996**, **39**(26): 5119
- Dillard, R. D.; Bach, N. J.; Draheim, S. E.; Berry, D. R.; Carlson, D. G.; Chirgadze, N. Y.; Clawson, D. K.; Hartley, L. W.; Johnson, L. M.; Jones, N. D.; McKinney, E. R.; Mihelich, E. D.; Olkowski, J. L.; Schevitz, R. W.; Smith, A. C.; Snyder, D. W.; Sommers, C. D.; Wery, J. P. *J. Med. Chem.*, **1996**, **39**(26): 5137
- <http://www.rcsb.org>
- Arnau, V.; Mars, S.; Marin, I. *Bioinformatics*, **2005**, **21**(3): 364
- Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided. Mol. Des.*, **2001**, **15**(5): 411
- Pei, J. F.; Wang, Q.; Zhou, J. J.; Lai, L. H. *Proteins*, **2004**, **57**(4): 651
- Wang, R. X.; Liu, L.; Lai, L. H.; Tang, Y. Q. *J. Mol. Mod.*, **1998**, **4**(12): 379
- Johnson, S. C. *Psychometrika*, **1967**, **32**(3): 241

Functional Classification of Phospholipase A₂ Family^{*}

LIU, Zhen-Ming LI, Bo LAI, Lu-Hua

(State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871)

Abstract The affinity fingerprint method was used to functionally classify proteins based on the binding data matrix produced by multiple-docking. The method has been applied to the classification of phospholipase A₂ enzymes. Though many of the classifications are similar to those from sequence alignment, the human cPLA₂ was successfully grouped with bee PLA₂ based on the binding site similarity, which can not be handled by sequence alignment due to cPLA₂'s low homology with other PLA₂s.

Keyword: Affinity fingerprint, Multiple docking, Function-based protein classification

Received: May 17, 2005; Revised: June 8, 2005. Correspondent: LAI, Lu-Hua(E-mail: lhlai@pku.edu.cn; Tel: 010-62757486; Fax: 010-62751725). ^{*} The Project Supported by the Ministry of Science and Technology of China and NSFC(20173001, 20228306, 90403001, 30490240)