

doi:10.3969/j.issn.1672-5565.2013.01.08

基于结构的单绕蛋白聚类图构建与分析

宋新蕊, 李晓琴*

(北京工业大学生命科学与生物工程学院, 北京 100124)

摘要:蛋白质分子进化规律研究是分子进化研究的重点,对揭示生命起源与进化机制有重要意义。本文对已知空间结构及物种信息的单绕蛋白,利用结构比对信息,构建了不同层次单绕样本系统聚类图。分析发现:功能相似蛋白存在明显聚集现象,同一超家族样本基本聚在一个大支中,同一家族样本集中在所属超家族下的小支中,功能约束下单绕样本聚类图与物种进化图有较好对应关系。结果表明:单绕蛋白的结构演化反映了蛋白质功能的约束,特定功能单绕样本的结构差异具有种属特异性,结构演化包含了物种进化信息。

关键词:单绕,结构比对,系统聚类,物种进化

中图分类号:Q51 **文献标识码:**C **文章编号:**1672-5565(2013)-01-044-06

Clustering dendrograms construction and analysis of TIM – barrel proteins based on structure information

SONG Xin – rui, LI Xiao – qin *

(Bioengineering Center, Beijing University of Technology, Beijing 100124, China)

Abstract: The regularity study of protein molecular evolution is the key to molecular evolution and it has important meaning to reveal life origin and the evolutionary mechanism. In this paper, TIM – barrel proteins with the known spatial structure material and species information were selected as different level datasets to construct clustering dendrograms using their structure alignment information. The analysis of clustering dendrograms shows that proteins with similar function obviously tend to cluster. Members with the same superfamily are almost clustered in a large branch members and the same family with Members are clustered in a small branch which belonged to their super-families. Under the constraints of function, clustering dendrograms of TIM – barrel proteins have a good corresponding relation whie with species phylogenies. Our results indicate that the structure evolution of TIM – barrel protein reflects the constraints of protein function, and the structural differences between specific functional TIM – barrel proteins show species specificity. The structure evolution implies the evolution information of species.

Key words: TIM – barrel Proteins, Structure Alignment, Hierarchical Cluster, Species Evolution

蛋白质分子进化规律研究是分子进化研究的重点,是生物学重大基础研究课题之一,对揭示生命起源与进化机制有重要意义。蛋白质分子的空间结构复杂而不规则,但其对应的蛋白折叠类型却只有数百到数千种^[1],比较分析有限数目的蛋白质折叠类型,可以更本质的反映蛋白质分子的种属进化关系。

蛋白质分子进化中,片段插入、缺失、替换、循环

排列以及 β 片层在拓扑结构中的重新排布等机制解释了蛋白质结构的不规则变化^[2],但复杂而不规则的蛋白质分子所对应的数百到数千种折叠类型是一步形成的,还是逐步形成的并不很清楚^[3]。目前,关于蛋白质进化研究主要集中在蛋白质家族、结构类进化、蛋白质进化速率影响因素、蛋白质结构进化单位、蛋白质趋同进化还是趋异进化等方面。例

收稿日期:2012-04-24;修回日期:2012-05-24.

基金项目:北京市自然科学基金(4112010)及北京市教委科技发展计划面上项目资助。

作者简介:宋新蕊,女,硕士研究生,主要研究方向:生物信息学。E-mail: songxinrui@emails.bjut.edu.cn.

* 通讯作者:李晓琴,女,教授。Tel: 010-67391610, E-mail: lxq0811@bjut.edu.cn.

如, Buljan 等^[4]利用 Pfam 数据库研究特定家族的进化,发现免疫球蛋白家族可以追溯到 5 亿年前。Choi 等^[5]分析了蛋白质共同结构祖先在 α 、 β 、 $\alpha + \beta$ 、 α/β 结构类中的分布,提出蛋白质结构类进化的一个方案。最近, Begum 等^[6]研究了二级结构片段在进化中的作用,他们以人的全 α 、全 β 类蛋白为研究对象,分析溶剂可及性、蛋白质二级结构组分、蛋白质聚集性等进化速率的影响因素,结果表明蛋白质的二级结构 - α 螺旋和 β 片层在控制蛋白质进化速率上起重要作用。Setiyaputra 等人围绕 ($\beta\alpha$) 8 筒结构的形成,建立了统一的模型,结果表明 ($\beta\alpha$) 4 等亚结构的组合形成了 ($\beta\alpha$) 8 筒结构蛋白^[7]。Wilmanns 等人介绍了组氨酸和色氨酸生化合成中涉及到的单绕结构酶,并从结构上、生化途径、进化数据中给出在色氨酸生物合成中涉及到的 TrpF, TrpC, TrpA, 组氨酸合成中的 HisF, HisA, 以及双功能酶 PriA 可能的进化祖先,并表明 TrpF, TrpC, TrpA 有共同的进化起源^[8]。而有关同一种折叠类型所属蛋白的进化研究却少有报道。分析同一折叠类型所属蛋白的结构演化关系,有助于我们了解特定折叠类型所属家族、超家族间的亲疏关系,发现结构演变中一些可能的限制性因素,对探索结构演变与功能多样性间关系有重要意义。

单绕结构是一种典型的蛋白质折叠类型,经过

长期进化,单绕蛋白的序列差异已经达到 70% 以上,使得基于序列的单绕蛋白进化研究困难重重。尽管单绕蛋白的序列差异很大,但其拓扑结构却依然保持不变,使得基于结构的单绕蛋白进化研究成为可能。特别是随着蛋白质结构数据的增多,结构比对算法的发展^[9-10],使得蛋白质分子的结构比较凸显更多的优越性^[11-12],也推动了样本量大、功能分布广泛的典型蛋白质折叠类型的进化研究。本文以单绕结构为研究对象,依据样本间结构比对信息构建系统聚类图,并对聚类结果进行分析,探索单绕结构演化与功能及物种进化的关系。

1 数据

蛋白质结构依据二级结构含量主要分为 5 类:即 α 类、 β 类、 α/β 类、 $\alpha + \beta$ 类和 ζ 类,进一步,根据蛋白质二级结构片段的排列、取向等将每一类蛋白进一步细分为折叠类型。单绕结构 (TIM - barrel 桶结构) 是 α/β 类蛋白中样本量最多的一种折叠类型, Astral 1.65^[13] 中相似性小于 95% 的数据集中有物种信息的单绕样本 326,其中 315 个样本属于酶类,分布在氧化还原酶、转移酶、水解酶、裂合酶、异构酶 5 大酶类中,326 个单绕结构蛋白在 Astral 1.65 数据库中的编号见表 1。

表 1 326 个单绕结构蛋白 Astral ID

Table 1 Astral IDs of 326 TIM - barrel proteins

Astral ID								
d1 ads_	d1 afsa_	d1 ah4_	d1 ak5_1	d1 b5ta_	d1 c9wa_	d1 ct5a_	d1 d3ga_	d1 ea0a2
d1 a3wa2	d1 a49a2	d1 ad1a_	d1 aj2_	d1 b9ba_	d1 cgt_4	d1 ciu_4	d1 cyg_4	d1 d8ca_
d1 a4ma_	d1 aq0a_	d1 avaa2	d1 b1ya_	d1 bag_2	d1 bf2_3	d1 bf6a_	d1 bg4_	d1 bhga3
d1 a53_	d1 a5ca_	d1 adoa_	d1 bqg_1	d1 bwva1	d1 bxna1	d1 d7ka2	d1 dbta_	d1 dhpa_
d1 a0ca_	d1 a0da_	d1 a0ea_	d1 aw1a_	d1 bd0a2	d1 bxba_	d1 bxca_	d1 ccwb_	d1 d8wa_
d1 fvpa_	d1 nfp_	d1 jnda1	d1 hjxa1	d1 e9la1	d1 exba_	d1 nar_	d1 env_	d1 ljya1
d1 ezwa_	d1 f07a_	d1 f76a_	d1 frb_	d1 gox_	d1 gtea2	d1 gvea_	d1 gvoo_	d1 gwja_
d1 eswa_	d1 eyea_	d1 f05a_	d1 f6ya_	d1 g5aa2	d1 gjwa2	d1 h6za1	d1 hfba_	d1 i2oa_
d1 cbg_	d1 ceo_	d1 clxa_	d1 cz1a_	d1 e43a2	d1 e4ia_	d1 e4mm	d1 e9yb2	d1 ea9e3
d1 dqwa_	d1 dvja_	d1 dxea_	d1 e51a_	d1 e9ia1	d1 ec7a1	d1 eexa_	d1 eixa_	d1 epxa_
d1 htia_	d1 i60a_	d1 iv8a2	d1 j5sa_	d1 jpdx1	d1 jpma1	d1 kv5a_	d1 lyxa_	d1 m53a2
d1 lqaa_	d1 luca_	d1 luch_	d1 m41a	d1 o94a1	d1 ofda2	d1 og6a_	d1 oyb_	d1 zfja1
d1 m7xa3	d1 n8fa_	d1 n8ia_	d1 ntha_	d1 o4ua1	d1 ot1a4	d1 oy0a_	d1 pkla2	d1 qapa1
d1 fh9a_	d1 foba_	d1 g0ca_	d1 g94a2	d1 gcya2	d1 ghsa_	d1 gkpa2	d1 gkra2	d1 gnxa_
d1 gvfa_	d1 gzga_	d1 h5ya_	d1 hkva2	d1 hl2a_	d1 i4na_	d1 igwa_	d1 iwpa_	d1 iyxa1
d1 pyma_	d1 qo2a_	d1 qt1a_	d1 rpxa_	d1 teda_	d1 tph1_	d1 trea_	d1 ttj_	d1 xima_
d1 hw6a_	d1 icpa_	d1 j96a_	d1 jena1	d1 jeza_	d1 jr1a1	d1 k87a2	d1 hjsa_	d1 ht6a2
d1 jvna1	d1 k4ga_	d1 kb1a1	d1 liua2	d1 lt8a_	d1 lwha2	d1 m3ua_	d1 kkoa1	d1 km3a_
d1 edqa2	d1 edt_	d1 egza_	d1 eh9a3	d1 ejxc2	d1 eoka_	d1 fa2a_	d1 j0ha3	d1 j18a2
d1 f74a_	d1 f8ma_	d1 fbaa_	d1 fdja_	d1 fhua1	d1 geha1	d1 geqa_	d1 o0ya_	d1 oeapa1
d1 mo0a_	d1 muca1	d1 muwa_	d1 n55a_	d1 neya_	d1 nsj_	d1 p0ka_	d1 xyfa2	d1 xyza_
d1 h1na_	d1 h41a1	d1 hiza_	d1 hjqa_	d1 qopa_	d1 rbla1	d1 kbia1	d1 k1da2	d1 k6wa2
d1 jqna_	d1 jqoa_	d1 ka9f_	d1 kcza1	d1 ll7a1	d1 m7ja3	d1 m5wa_	d1 thfd_	d1 ub3a_
d7reqa1	d7reqb1	d1 itua_	d1 itxa1	d7odca2	d8ruca1	d1 fcqa_	d1 mxga2	d1 nfga2
d1 ilwa_	d1 iexa1	d1 n7ka_	d1 nvma2	d1 onwa2	d1 plma2	d1 gk8a1	d1 nowa1	d1 o12a2

(续表 1)

Astral ID								
d1 muma_	d1 mzha_	d1 ji2a3	d1 jz8a5	d1 jfxa_	d1 ji1 a3	d1 pii_2	d1 pbga	_d1 qasa3
d1 qo5b_	d7 taa_2	d1 h09a2	d1 jcla_	d4 xiaa_	d1 i0da_	d1 losc_	d1 jaka1	d1 pii_1
d1 kwga2	d1 qfea_	d7 a3ha_	d1 gvia3	d1 j93a_	d2 mnr_1	d1 hxja_	d1 l6wa_	d1 jae_2
d1 ep3a_	d1 e0ta2	d1 byb__	d1 dqua_	d1 hg3a_	d1 om0a_	d1 huva_	d1 jxa_	d1 edg__
d2 dora_	d2 tpsa_	d1 gowa_	d1 j5ta_	d2 chr_1	d1 hx0a2	d1 l6sa_	d1 j79a_	d1 onea1
d1 lg2a1	d5 ruba1	d1 ogsa2	d1 uok_2	d1 qvba_	d1 l8na1	d3 rubl1	d1 odza	_d1 ug6a_
d1 krma_	d2 plc__	d1 ob0a2	d1 tux__	d1 qhoa4	d2 toda2	d1 od8a_	d1 ud2a2	d1 qoxa_
d2 ptd__	d1 od0a_	d1 uasa2	d1 qnra_	d2 hvm__	d1 ktba2	d1 eepa_	d1 d9ea_	d1 bqca_
d1 h1ya_	d1 k77a_	d1 hqta_	d1 iq8a1	d1 ecea_	d1 euua_	d1 m6ja_	d2 alr__	d1 dosa_
d1 smd_2	d1 qba_3	d4 ubpc2	d1 pdz_1	d1 f2ja_	d2 ebn__	d1 qtw_	d1 mza_	d1 o1za_
d1 qp0a1	d1 goia2	d1 izca_	d2 btma_	d1 hvxa2	d1 kv8a_	d1 j6oa_	d1 ohla_	d2 aaa_2
d1 kfwa1	d1 uroa_							

326 个单绕样本分布在 26 个超家族、73 个家族中,其中最大的一个超家族包含 107 个样本,对应的 SCOP^[14-15] 数据库编号为 c. 1. 8; 最大的家族包含 42 个样本,对应的 SCOP 数据库编号为 c. 1. 8. 1; 其中 α -淀粉酶(α -amylase)、丙糖磷酸异构酶

(Triosephosphate isomerase)、D-木糖异构酶(D-xylose isomerase) - 这三类酶包含的样本最多,分别含 14、15、10 个样本,不同数据集样本的信息统计见表 2。

表 2 不同数据集样本的信息统计

Table 2 Information summary of different datasets samples

数据集名称	SCOP 类别	样本数目	平均序列相似性	物种数目	物种分布
单绕数据集	c. 1	326	6. 12%	154	细菌,古生菌,真核生物,病毒
α -amylase	c. 1. 8. 1	14	22. 81%	13	细菌,古生菌,真核生物,
Triosephosphate isomerase	c. 1. 1. 1	15	42. 18%	14	细菌,古生菌,真核生物,
D-xylose isomerase	c. 1. 15. 3	10	45. 18%	10	细菌

2 方法

2.1 系统聚类图构建方法

对已知空间结构信息的任意两个单绕结构样本,利用 TM-align^[10] 算法进行结构比对,获得两样本的结构比对结果,利用比对结果中匹配残基的坐标信息计算 RMSD。RMSD 反映了蛋白质 C α 原子之间的匹配程度,是衡量蛋白质分子骨架之间平均距离的度量标准,计算公式为: $RMSD(v, w) =$

$\sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2}$ 其中, n 为两个蛋白样本匹配的残基总数, v_i, w_i 分别为第 i 对匹配残基的空间坐标, $\|v_i - w_i\|$ 为第 i 对匹配残基 C α 原子间距离。

对给定数据集,应用 TM-align 算法进行任意两样本的结构比对,利用结构比对结果计算两样本的 RMSD 值,形成该数据集对应的 RMSD 矩阵,进一步利用 SPSS 软件提供的系统聚类方法,选择组间联接、欧式距离构建数据集所属样本对应的系统聚类图。

TM-align 是一种基于模板的结构比对方法,该方法结合了 TM-score 旋转矩阵与动态规划算法,运算速度比 DALI 和 SAL 快 20 倍,比 CE 快 4 倍,是

一种多结构比对方法,其双结构比对结果与 DALI 相当。

2.2 物种进化树构建方法

对给定的数据集,提取数据集中每一个样本对应的物种名称,保留全部不同物种名称,形成数据集对应的物种名称集。将数据集对应的物种名称提交到 NCBI 的 Taxonomy Common Tree,即可得到该数据集样本对应的物种进化树。

3 结果与分析

3.1 折叠类型层次的系统聚类图及分析

对单绕数据集的 326 个样本,利用系统聚类方法得到该数据集样本的系统聚类图,聚类图的拓扑结构见图 1。图中部分分支末端有 SCOP 家族分类号标记,表示该分支含 2 个以上该家族所属样本,分支末端有 SCOP 超家族分类号标记,表示该分支含 2 个以上该超家族所属家族成员;图中没有标记的分支表示该分支包含 1 个或 2 个同一家族的样本或 2 个同一超家族的家族成员,且该分支不与相邻分支属于同一个家族或超家族。系统聚类图大体可以分为 4 部分,如图 1 数字标注。

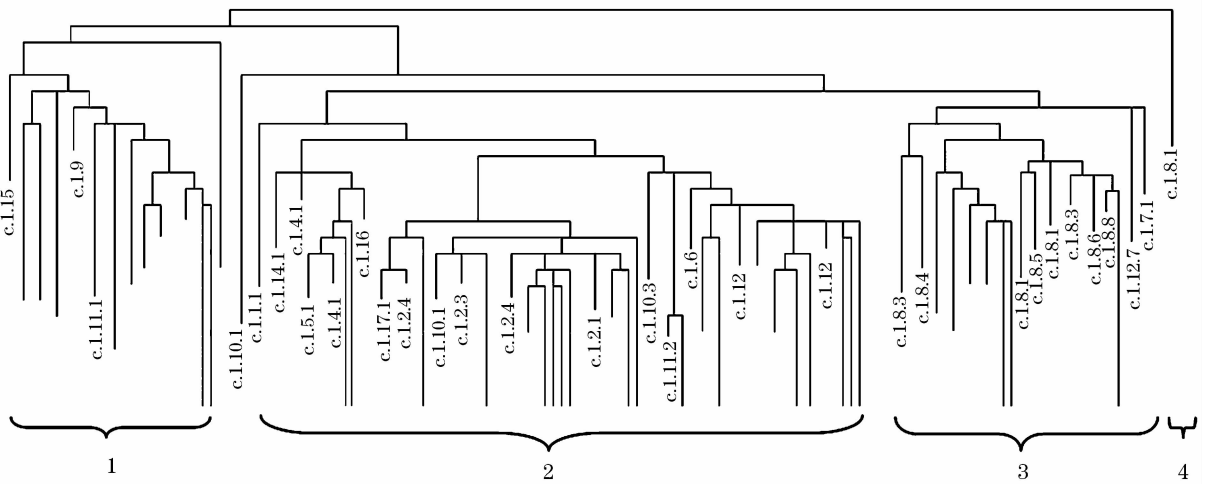


图1 单绕数据集样本系统聚类图

Fig. 1 Clustering dendrograms of TIM-barrel protein dataset samples

聚类图的第1部分包含了59个样本,主要分布在c.1.15、c.1.9、c.1.11等3个超家族中。c.1.15超家族共包含5个家族、14个样本,并且这14个样本都聚集在图1中标注c.1.15的分支中;c.1.9超家族包含13个家族、20个样本,且这20个样本都聚集标注该超家族的分支中;c.1.11超家族包含2个家族,家族c.1.11.1中的所有5个样本都聚集标注的分支中。分析上述3个超家族样本的所属功能,发现这3个超家族都有与金属离子结合的位点,相似的功能使得这3个超家族聚集。

聚类图的第2部分包含128个样本,其中68%的样本包含与磷酸结合相关的位点,包括c.1.1、c.1.2、c.1.4、c.1.5、c.1.6等超家族。该部分包含的家族多,但大多数同一家族的样本仍然能够聚集在一个分支中,如c.1.1.1家族中有15个样本,其中14个聚集在同一个分支中,c.1.14.1、c.1.5.1、c.1.17.1、c.1.2.2各家族包含的样本全部聚集在同一个分支中。比较分散的超家族有c.1.10、c.1.2、c.1.12,在这些超家族中,有些属于同一个家族的样本也没有聚集在一个分支中,如c.1.10.1家族,共有20个样本,分属多个分支,其中分支中样本数目最多的包含5个样本,其它分支包含样本数目在3个以下;c.1.2超家族中各家族分布也很分散,且分散的大多数家族是被c.1.10超家族样本分割开的;c.1.12超家族中的家族也被分到2个分支中,且2个分支之间也主要是c.1.10超家族的样本。进一步从功能上看,c.1.10超家族是醛缩酶超家族,该超家族中的c.1.10.1家族是I类醛缩酶,可能起到了连接醛缩酶超家族和磷酸结合 α/β 桶结构的作用;c.1.2超家族是核酮糖磷酸结合桶超家族,该超家族

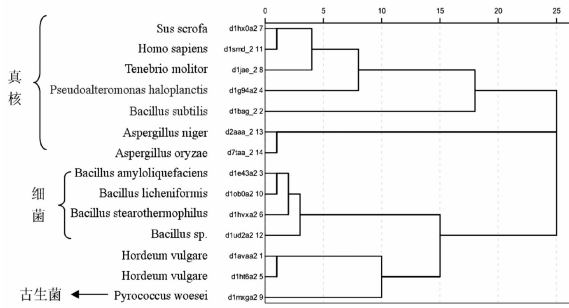
中的样本包含磷酸结合位点;c.1.12超家族是磷酸烯醇式丙酮酸/丙酮酸超家族,大部分家族都有与磷酸结合相关的结构。c.1.10超家族功能与c.1.2和c.1.12两个超家族功能之间的关联,使得在聚类图上c.1.2超家族及c.1.12超家族中间频繁插入c.1.10超家族样本,显示了基于结构的聚类与功能的密切关系。

聚类图的第3部分包含样本数95,主要是c.1.8超家族,几乎94%的样本为参与糖代谢过程中的水解酶,各个家族聚集程度较高。c.1.7.1家族也聚集在第三部分,该家族14个样本都聚集在同一个分支中,且大部分酶也参与了糖的代谢过程。聚类图的第4部分为c.1.8.1家族,这个家族共包含42个结构,除了分布于第三部分的7个样本外,其余35个样本都聚集在这一分支中,这个家族中的一些样本在结构上可能插入了一些亚结构域,使得该家族在聚类图中被独立分到了最外层。

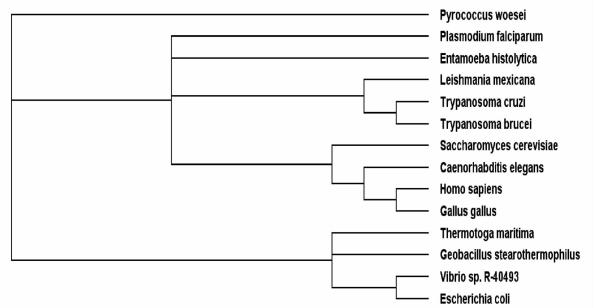
此外,我们分析了326个样本的系统聚类图与物种进化图之间的关系,发现系统聚类图与物种进化图之间并没呈现明显的对应关系,从折叠类型层次上不能反映结构演变与物种进化之间的关系。

3.2 功能约束的单绕样本系统聚类图及分析

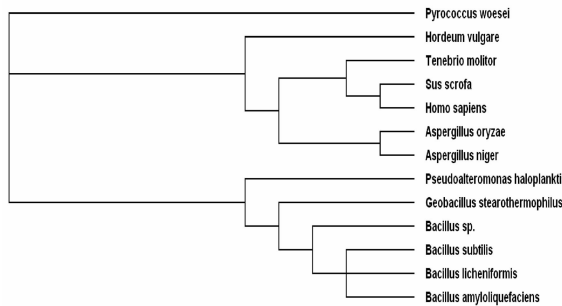
为了探索单绕结构类型演变与物种进化之间的关系,我们对选取的单绕结构样本的功能进行了限制,选取 α -淀粉酶(alpha-amylase)、丙糖磷酸异构酶(Triosephosphate isomerase)、D-木糖异构酶(D-xylose isomerase)三种功能蛋白,分别对三个功能蛋白样本进行系统聚类,并与该功能蛋白所含物种分类结果进行比较。图2为[ABD1]3个功能蛋白的系统聚类结果及其对应样本的物种分类结果。



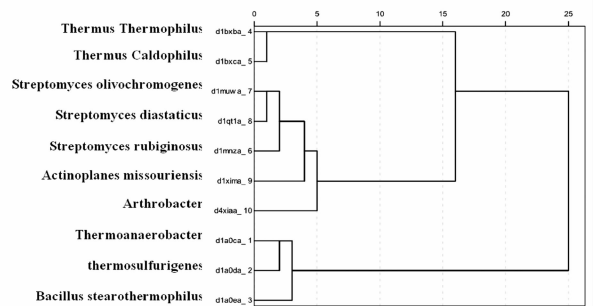
(a) alpha-amylase 的系统聚类图



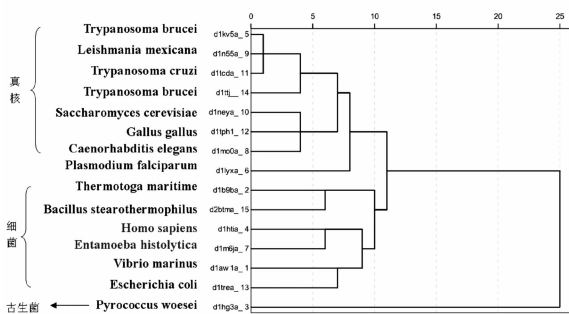
(d) Triosephosphate isomerase 的物种进化树



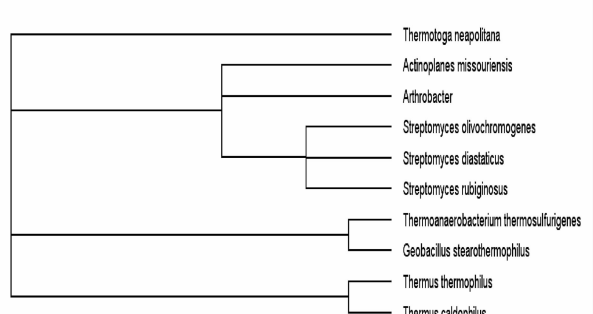
(b) alpha-amylase 的物种进化树



(e) D-xylose isomerase 的系统聚类图



(c) Triosephosphate isomerase 的系统聚类图



(f) D-xylose isomerase 的物种进化树

图 2 不同功能蛋白的系统聚类图与物种进化树

Fig. 2 Clustering dendrograms and species phylogenies of different functional proteins

图 2(a) 为基于 alpha - amylase 数据集中样本结构的系统聚类图, 图 2(b) 为其样本对应的物种进化树, 这一数据集包含 14 个样本, 功能都是 alpha - amylase, 这 14 个样本属于 13 个不同的物种, 总体上物种分类为真核部分、细菌部分和古生菌部分。由图 2(a) 我们可以看出, d1hx0a2 与 d1smd_2 聚为一类, 再与 d1jae_2 聚集, 这与其所对应的物种 Sus Scrofa、Homo Sapiens、Tenebrio molitor 的聚类结果是相对应的; d2aaa_2 与 d7taa_2 聚为一类, 这两个样本所对应的物种 Aspergillus niger 和 Aspergillus oryzae 同样也聚集在一个分支中。以上 5 个样本所对应的物种属于生命三域中的真核生物部分, 结构聚类与物种分类基本上是对应的, 只是结构聚类中混入了 d1g94a2 和 d1bag_2, 这两个样本属于物种分类中细菌部分, 并且原来位于物种分类真核部分的物种

Hordeum vulgare 对应的样本 d1avaa2 和 d1ht6a2 与古生菌 Pyrococcus woesei 对应的样本 d1mxga2 聚在了一起, 这是结构聚类图与物种分类图不一致的地方, 这可能与古生菌更接近于真核生物有关系。结构聚类图中的其它分支与物种分类图中的细菌部分几乎是一一对应的。尽管部分结构与物种分类存在不一致对应, 总体上同一功能蛋白质结构的差异是可以反映物种之间亲缘关系的远近的。

同样, 图 2(c) 中属于真核部分的样本 d1kv5a_、d1n55a_、d1tceda_、d1ttj_、d1neya_、d1tph1_、d1mo0a_、d1lyxa_ 聚在一个较大的分支中, 与物种分类图 2(d) 中的真核部分是相互对应的; 属于细菌部分的样本 d1b9ba_、d2btma_、d1aw1a_、d1trea_ 聚在一个较大的分支中, 与物种分类对应, 只是在结构分支中混入了 d1b9ba_ 和 d2btma_, 这两个样本属于真核部分;

d1hg3a 独自形成一支,与物种分类中古生菌相互对应。

图 2(e) 中的结构都是来自细菌中的样本,结构聚类图与物种进化树形成了很好的对应,尤其是结构聚类图中位于中间的一个分支中的 5 个样本 d1muwa_、d1qt1a_、d1mnza_、d1xima_、d4xiaa_、这 5 个样本与其所对应的物种聚类几乎是一一对应;另外 d1bxba_ 和 d1bxca_ 聚为一个分支,这与其对应的物种 *Thermus Thermophilus* 和 *Thermus Caldophilus* 在物种进化树中的聚集情况是一样的;对于结构图中的 d1a0ca_、d1a0da_、d1a0ea_、这 3 个样本聚为一个分支,而其对应的物种则分散在 2 个不同的分支中。虽然结构聚类图和物种进化树的部分分支的对应情况有差异,但总体结构聚类可以反映物种之间的差异。

为了检验系统聚类结果与物种分类结果对应关系的普适性,利用 Astral 1.75 相似性小于 100% 数据库,对上述三个功能蛋白对应的数据集进行扩充,扩充后得样本量分别为 52, 55, 39, 对应的物种数目分别为 16, 16, 12。以同样的方法进行系统聚类,结果与图 2 中结果是一致的,依然是相同物种的样本首先聚到一起,再次,相近的物种样本聚到一起。系统聚类结果与物种分类结果有较好的对应,说明在同一种折叠类型中,相同功能蛋白样本的结构差异具有种属特异性,这种特异性能很好的对应于物种分类。

4 结论

基于单绕蛋白结构信息构建的系统聚类图,在折叠类型层次下,326 个样本的系统聚类图反映了功能相近超家族和家族的聚集,但是没有显示出折叠类型与物种的对应关系;对同一种功能的蛋白而言,基于结构的系统聚类能够反映物种间的差异。因此,单绕蛋白的结构演化反映了蛋白质功能的约束,特定功能单绕蛋白的结构差异具有种属特异性。结构演化包含了物种进化信息,基于蛋白质结构信息的聚类分析方法对于分析序列低相似的单绕折叠类型是可行的。

参考文献 (References)

- [1] Chothia C. One thousand families for the molecular biologist[J]. Nature, 1992, 357(6379): 543–544.
- [2] Grishin N V. Fold Change in Evolution of Protein Structures[J]. Journal of Structural Biology, 2001, 134(2–3): 167–185.
- [3] Lupas A N, Ponting C P, Russell R B. On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World? [J]. Journal of Structural Biology, 2001, 134(2–3): 191–203.
- [4] Buljan M, Bateman A. The evolution of protein domain families [J]. Biochemical Society Transactions, 2009, 37(4): 751–755.
- [5] Choi I, Kim S. Evolution of protein structural classes and protein sequence families[J]. Proceedings of the National Academy of Sciences, 2006, 103(38): 14056–14061.
- [6] Begum T, Ghosh T C. Understanding the Effect of Secondary Structures and Aggregation on Human Protein Folding Class Evolution[J]. Journal of Molecular Evolution, 2010, 71(1): 60–69.
- [7] Setiyaputra S, Mackay J P, Patrick W M. The Structure of a Truncated Phosphoribosylanthranilate Isomerase Suggests a Unified Model for Evolution of the ($\beta\alpha$)₈ Barrel Fold[J]. Journal of Molecular Biology, 2011, 408(2): 291–303.
- [8] List F, Sterner R, Wilmanns M. Related ($\beta\alpha$)₈–Barrel Proteins in Histidine and Tryptophan Biosynthesis: A Paradigm to Study Enzyme Evolution[J]. Chembiochem, 2011, 12(10): 1487–1494.
- [9] Konagurthu A S, Whistock J C, Stuckey P J, Lesk A M. MUS-TANG: A multiple structural alignment algorithm [J]. Proteins: Structure, Function, and Bioinformatics, 2006, 64(3): 559–574.
- [10] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score[J]. Nucleic Acids Research, 2005, 33(7): 2302–2309.
- [11] Taylor W R, Flores T P, Orengo C A. Multiple protein structure alignment[J]. Protein Science, 1994, 3(10): 1858–1870.
- [12] Taylor W R, Orengo C A. Protein structure alignment[J]. Journal of Molecular Biology, 1989, 208(1): 1–22.
- [13] Chandonia J M, Hon G, Walker N S, Lo Conte L, Koehl P, Levitt M, Brenner S E. The ASTRAL Compendium in 2004[J]. Nucleic Acids Research, 2004, 32(suppl 1): D189–D192.
- [14] Andreeva A, Howorth D, Chandonia J, Brenner S E, Hubbard T J P, Chothia C, Murzin A G. Data growth and its impact on the SCOP database: new developments[J]. Nucleic Acids Research, 2008, 36(suppl 1): D419–D425.
- [15] Murzin A G, Brenner S E, Hubbard T, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures[J]. Journal of Molecular Biology, 1995, 247(4): 536–540.