

研究报告

Research Report

醛脱氢酶(ALDH)蛋白家族的信息学比较分析与修饰预测

常凯¹ 童立冬² 江忠勇¹ 熊怡淞¹ 张睿¹ 熊杰^{1*}

1 中国人民解放军成都军区总医院检验科, 成都, 610083; 2 中国人民解放军成都军区总医院放射科, 成都, 610083

*通讯作者, xiongjie1969@126.com

摘要 以 ALDH1~ALDH4 为核心,利用生物信息学方法对比分析 ALDH 家族成员的蛋白理化性质、亚细胞定位、构建分析蛋白三维模型并绘制家族系统进化树。对不稳定蛋白进行修饰预测。结果表明:ALDH 家族间的序列相似性较高但同源性相对较低,氨基酸序列差异较大。蛋白的 GRAVY 值(-0.284~0.01)和分子量(47 974.0~61 810.5)没有典型的规律性分布,但不稳定指数、脂肪系数和等电点差异显著。结构分析催化位点和结合区域有较大差异,这赋予了各家族成员不同的催化活性。最后就不稳定蛋白进行了理想的模拟突变与预测分析。分析结果可为 ALDH 蛋白的研究提供有价值信息,为进一步研究 ALDH 在人体内的疾病发生途径与作用机理提供依据。

关键词 醛脱氢酶, 生物信息学, 对比分析, 蛋白特征, 修饰预测

Comparative Analysis and Mutation Prediction of Aldehyde Dehydrogenase Based on the Bioinformatics Methods

Chang Kai¹ Tong Lidong² Jiang Zhongyong¹ Xiong Yisong¹ Zhang Rui¹ Xiong Jie^{1*}

1 Department of Medical Laboratory, Chengdu Military General Hospital, Chengdu, 610083; 2 Department of Radiology, Chengdu Military General Hospital, Chengdu, 610083

* Corresponding author, xiongjie1969@126.com

DOI: 10.13417/j.gab.035.000028

Abstract In order to modify and predict the unstable protein, the physical and chemical properties of protein and subcellular localization were analyzed, and the tertiary structure of protein was built and analyzed, then the phylogenetic tree of the ALDH family was drew by using bioinformatics methods, which was in the center of ALDH1~ALDH4. The result showed that the similarity was higher but the homology was relatively low among the sequence of ALDH family, while acid sequence had more difference. There was no special regularity of the distribution in GRAVY values (-0.284~0.01) and molecular weight of protein (47 974.0~61 810.5), but instability index, aliphatic index and isoelectric point were significantly different. Catalytic site and binding domain were analyzed to have more differences, which gave different catalytic activity to members in every family. At last, the stability of unstable protein was predicted after modification. The results of the analysis might provide valuable information for the study of ALDH protein, and provide the basis for further study of the ways and the mechanism of ALDH on how it works in human body.

Keywords Aldehyde dehydrogenase, Bioinformation, Comparative analysis, Protein properties, Mutation prediction

醛作为一种高活性的化合物,能够诱导包括 DNA 和蛋白质形成在内的多种细胞毒性。醛脱氢酶(aldehyde dehydrogenase, ALDH)超家族是一类依赖 NAD(P)⁺的多功能蛋白。人 ALDH 超家族成员至少存在 19 种,根据进化距离和氨基酸相似性对其命名。ALDH 家族成员可通过氧化反应催化内源性和外源性醛,使之不可逆的转变为相应的酸,保护机体免受醛类物质的毒害(Fontaine et al., 2002)。虽然许

多 ALDH 酶具有广泛的底物活性且能氧化多种脂肪族和芳香醛,但是仍有一些 ALDH 具有唯一的底物偏好性。例如 ALDH2 的水解活性与硝酸甘油的联系密切且能被乙酰氨基酚的结合蛋白所识别;ALDH1 能够结合雄激素、胆固醇和甲状腺激素,醛脱氢酶在维持细胞内环境稳定中起着至关重要的作用。ALDH 能够通过生成 NADH 再生还原型谷胱甘肽(GSH)以促进细胞的抗氧化能力。此外,ALDH3 可以通过还原半胱氨酸和甲硫氨酸巯基清除羟自由基(Mann and Weiner, 1999)。

目前发现的 ALDH 蛋白均具有较为保守的 NAD(P)⁺结合区域和催化区域。但由于其构象的不同赋予了家族成员不同的职能(Izaguirre et al., 2002)。ALDH1 和 ALDH2 研究相对较多,但 ALDH3 和 ALDH4 的研究并不深入,这在极大程度上是由于这两个家族的蛋白在体外的稳定性不理想所致。本研究通过生物信息学方法对已克隆出的多组 ALDH 超家族成员进行比对研究,系统比较 ALDH 家族的各项理化系数与结构差异,基于蛋白质工程原理对不

稳定蛋白进行模拟突变和预测分析,为进一步深入研究 ALDH 奠定基础。

1 结果与分析

1.1 氨基酸序列的聚类对比分析

将来源于人、牛等不同物种的多个 ALDH 亚群的氨基酸序列在 CLUSTALX 和 MEGA 5.0 软件中进行分子进化树分析,基于 Neighbor-Joining 的原理建树(图 1),结果显示 ALDH1~ALDH7 家族分别归属于 7 个不同的分支,且每个分支聚类明确(Lan et al., 2013)。由此可见,该蛋白的聚类分析不是由物种进行分类,而是由 ALDH 的各个亚家族的差异进行聚类,进而说明虽然物种间差异巨大,但 ALDH 亚家族内部的氨基酸序列相似性极高且足以稀释物种间的差异。各分支内部的不同物种聚类进化距离分析发现,人类的 ALDH 基因均处于进化树远端,可见人类 ALDH 的进化发生较晚,功能也较为复杂。

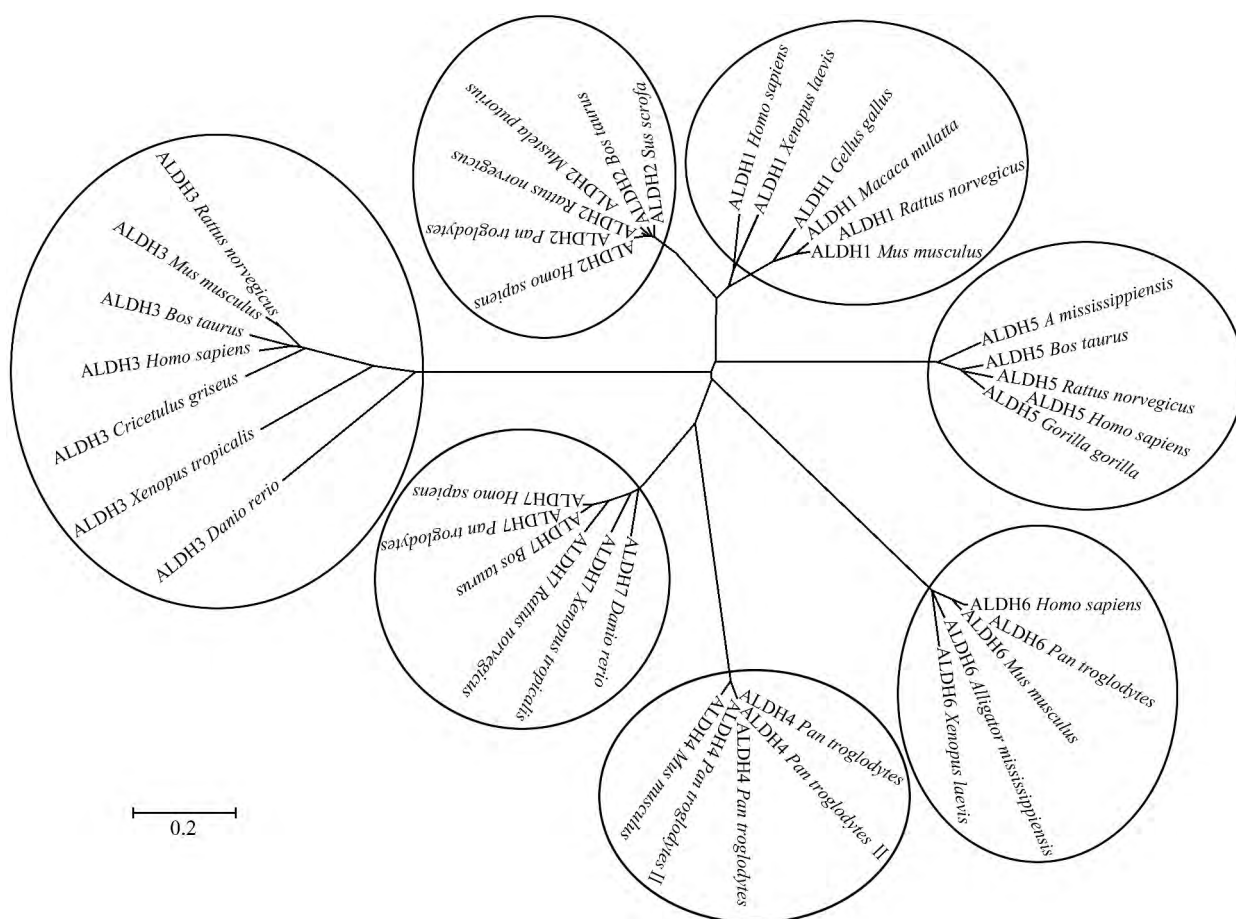


图 1 基于 MEGA5.0 软件平台用 neighbor-joining 方法构建的 ALDH 进化树

Figure 1 A phylogenetic tree of ALDH constructed by neighbor-joining method on MEGA 5.0

1.2 ALDH 蛋白特性对比分析

利用 SignalP 3.0 Server 和 TMpred 网络服务器对 ALDH 进行跨膜区分析,应用 Wolf-psort 软件进行蛋白的亚细胞定位分析(Chang et al., 2014)。结果发现在 ALDH1 氨基酸序列中具有跨膜区(183~203, 223~244, 457~480)定位不明确;ALDH2 中具有跨膜区(178~208, 228~248, 462~485)定位于线粒体;ALDH3 中具有跨膜区(118~137, 391~410)定位于线粒体、胞质和细胞核;ALDH4 中具有跨膜区(210~232)定位于线粒体;ALDH5 中具有跨膜区(188~210, 211~230)定位于线粒体;ALDH6 中具有跨膜区(25~44, 188~208, 306~327)定位于线粒体;ALDH7 中具有跨膜区(157~175, 191~210, 283~305)定位不明确。在上诉 7 个家族的定位中,除 ALDH1 和 ALDH7 定位不明确外,均定位于线粒体中。各 ALDH 家族蛋白均在 170~220 的氨基酸间形成跨膜区域。这与一些文献报道的亚细胞定位实验结果一致。

在 ExPASy Proteomics Server 网站上用 ProtParam 软件对 ALDH 蛋白进行预测,统计 Instability index、Ext. coefficient、Aliphatic index 值、GRAVY 值、pI 值等一系列数据,分析后发现 GRAVY 值(-0.284~0.01)和分子量(47 974.0~61 810.5)在不同物种、不同 ALDH 亚群间没有典型的规律性分布;但不同物种间的同一 ALDH 亚群的 pI 值较为稳定,不会因为分子量的大小和氨基酸的分布而改变。各 ALDH 家族的蛋白质不稳定指数同样差异显著,ALDH3 和 ALDH4 的不稳定指数>40,表明这两种家族的成员在体外稳定性很差。脂肪指数结果显示:ALDH3 和 ALDH7 家族脂肪指数(91.29±1.78)远高于其他家族(84.37±2.38)。GRAVY 值均为负值,表明 ALDH 蛋白均为亲水性蛋白,均有助于与 NAD(P)⁺和醛类的醛

基结合(图 2)。

1.3 氨基酸序列比对与高级结构比对

利用 Vector NTI、NCBI BLAST 软件对 ALDH 蛋白进行比对分析,进行 ALDH 全家族氨基酸的同源性为 44.4% 相似性为 2.4%。ALDH 亚家族成员内的比对分析,发现 ALDH1 同源性为 99.4%,相似性为 62.5%;ALDH2 同源性为 99.8%,相似性为 86%;ALDH3 同源性为 96.6%,相似性为 40.7%;ALDH4 同源性为 99.3% 相似性为 70.5%;ALDH5 同源性为 95.1% 相似性为 63.7%;ALDH6 同源性为 98.0% 相似性为 72.5%;ALDH7 同源性为 94.6%,相似性为 54.9%。上述比对结果提示,ALDH 家族间的序列相似性较高但同源性相对较低,氨基酸序列差异较大。ALDH 超家族的同源性在 45%左右,相似性却只有 2.4%左右,说明 ALDH 的各个亚家族虽同属于醛脱氢酶,但其亚群间的氨基酸序列具有较大差异。

在 PBIL 网站上使用 SOPMA 软件对 ALDH 进行二级结构预测表明:ALDH1 包含 42.77%的 α -螺旋、17.97%延伸带、11.13% β 转角和 28.12%随意卷曲构成;ALDH2 包含 43.52%的 α -螺旋、18.38%延伸带、9.67% β 转角和 28.43%随意卷曲构成;ALDH3 包含 46.15%的 α -螺旋、17.31%延伸链、10.04% β 转角和 26.5%随意卷曲构成;ALDH4 包含 37.3%的 α -螺旋、20.96%延伸链、8.17% β 转角和 33.57%随意卷曲构成;ALDH5 包含 43.18%的 α -螺旋、15.33%延伸链、9.35% β 转角和 32.15%随意卷曲构成;ALDH6 包含 37.76%的 α -螺旋、19.81%延伸带、10.09% β 转角和 32.34%随意卷曲构成;ALDH7 包含 39.33%的 α -螺旋、22.31%延伸链、8.61% β 转角和 29.75%随意卷曲构成(图 3)。

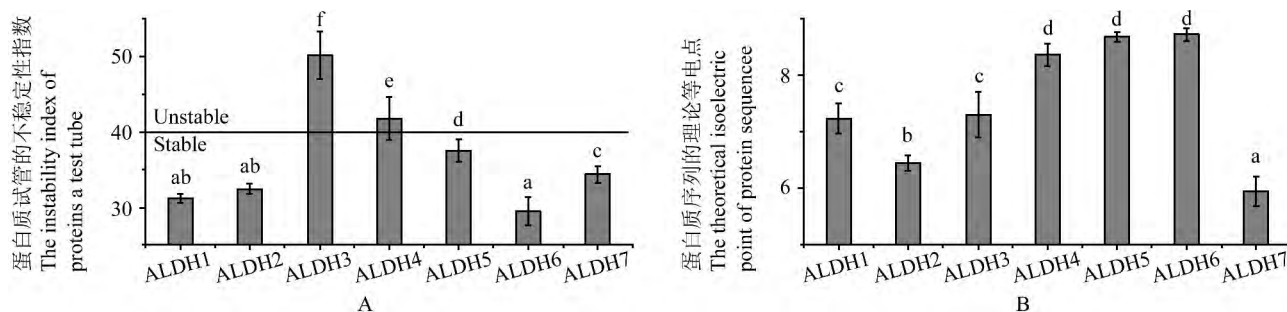


图 2 ALDH 蛋白序列的稳定性分析与等电点统计分析

注: A: 稳定性分析; B: 等电点分析; 统计分析应用 SPSS 邓肯氏复极差测验, 显著性差异由字母标示($p < 0.01$)

Figure 2 The instability index and the pI point of ALDH protein sequence

Note: A: The instability index analysis; B: The pI point analysis; The statistical analysis was performed by Duncan's multiple range tests; The different letters indicate significant difference ($p < 0.01$)

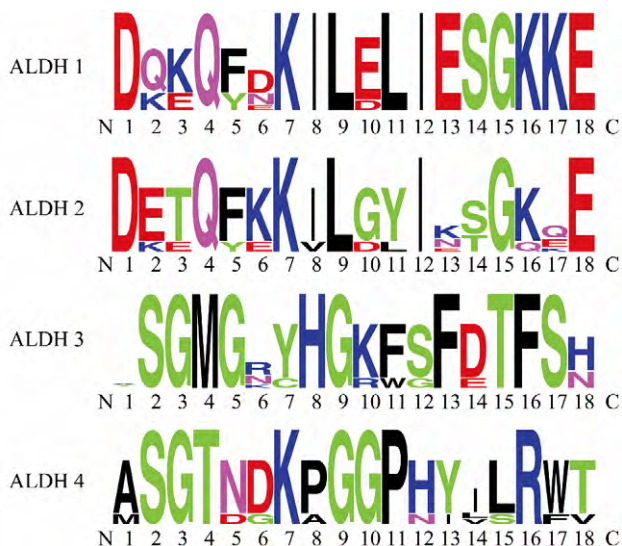
图 3 NAD(P)⁺结合区域的氨基酸序列保守性分析

Figure 3 The Weblogo of ALDH proteins NAD(P)⁺-binding conservative domain

基于日益庞大的蛋白质结构数据库,对蛋白质的高级结构进行预测和分析,有助于深入剖析其结构及功能。该研究用同源建模法,对蛋白质分子的三维结构进行初步模拟。应用 SWISS-MODEL 和 WebLab ViewerLite 软件构建 ALDH1~4 四级结构和三级结构进行功能位点与结构的比较分析,结合 Swiss-PdbViewer 拉式构象图进行蛋白构象合理性评估(Biasini et al., 2014)。分析 Psi 和 Phi 角分布,超过 90%的氨基酸残基位于理想 Psi 和 Phi 角分布区域内,可见预测结构的排布稳定,同源建模结果可靠(Johansson et al., 2012)。

ALDH1 (除 ALDH1A3)和 ALDH2 为同源四聚体,ALDH3 和 ALDH4 为同源二聚体,但各 ALDH 亚家族单体均具有典型的催化区域、NAD(P)⁺结合域、齐聚反应域和保守的功能位点(图 4)。ALDH1 被视为催化视黄醛转化为视黄酸的重要蛋白,具有活性位点 Glu267 和 Cys301。ALDH2 具有硝酸甘油的结合域和乙醛结合域和活性位点 Glu284 和 Cys318,主要担负催化硝酸甘油和乙醛的职责。ALDH3 具有保护视力不受紫外线伤害的作用,含 NAD(P)⁺结合域和活性位点 Asn115 和 Tyr116。ALDH4 含有醛结合环和活性位点 Glu447 和 Phe449。蛋白高级结构比对可知每个亚家族都有相对保守的功能域和保守位点,这是进行配体结合与催化选择的关键,有助于通过查询 ALDH 家族成员保守位点预测蛋白调节功能。

1.4 不稳定蛋白的修饰与预测

蛋白质的稳定性直接影响生物分子的功能和活性,它是评价蛋白的一个基本特征。蛋白质的稳定性不仅取决于外部环境更取决于蛋白质的内在属性。因此对于结构不稳定的蛋白质可以进行蛋白质工程方法修饰蛋白,从而达到增强稳定性的目的。ALDH3 和 ALDH4 的不稳定指数>40,稳定性较差。基于疏水相互作用和二硫键的应用,设计两个 ALDH3 突变位点 Y365C 和 E328C,ALDH4 突变位点 K347M 和 T213M (Elsohly and Francis, 2015)。应用 I-mutant2.0 软件计算 DDG 值(Kcal/mol)如表 1,

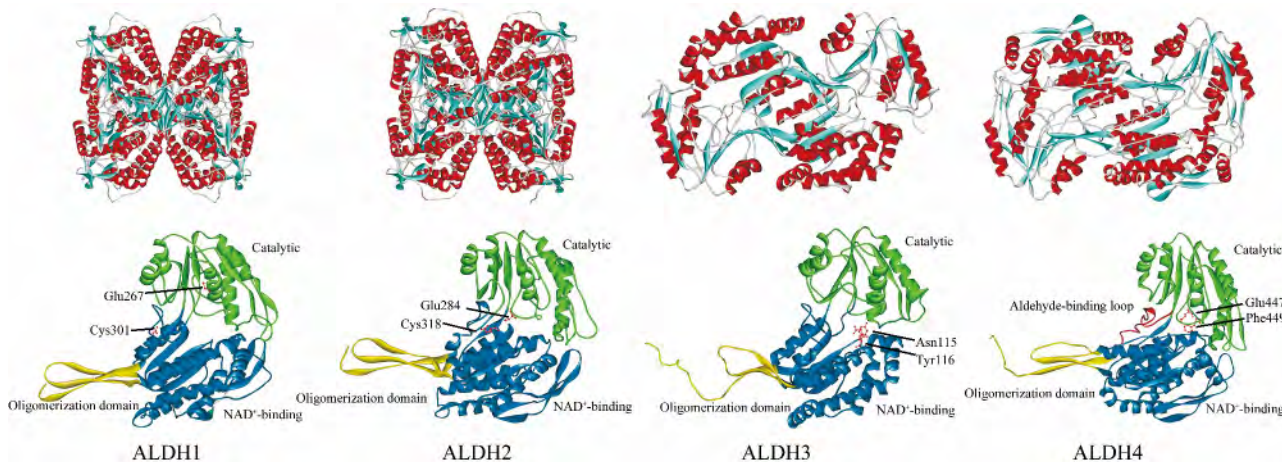


图 4 ALDH 家族四级结构建模与分析

注: 绿色: 催化区域; 蓝色: NAD⁺结合区域; 金色: 齐聚反应域; 红色: 醛结合环

Figure 4 The primary structure of ALDH protein established by homology-based modeling

Note: Green: Promoter with the catalytic domain; Blue: The NAD⁺-binding domain; Gold: The oligomerization domain; Red: Aldehyde-binding loop

表 1 人不稳定 ALDH 蛋白突变后的稳定性预测
Table 1 Stability prediction of human unstable ALDH protein after mutation

蛋白名称 Protein name	突变位点 Mutation site	DDG (10°C)	DDG (25°C)	DDG (40°C)	DDG (55°C)	DDG (70°C)	pH 值 pH value
ALDH3	Y365C	0.76	1.0	1.22	1.41	1.55	7.0
ALDH3	E328C	0.2	0.39	0.58	0.76	0.91	7.0
ALDH4	K347M	0.42	0.42	0.41	0.39	0.34	7.0
ALDH4	T213M	0.52	0.55	0.56	0.55	0.53	7.0

该预测方法基于 SEEF 的方法模拟程度较高, DDG 值 >0 表示稳定性增减, DDG <0 表示稳定性降低 (Capriotti et al., 2005)。结果显示:再 pH 为 7.0 时, ALDH3 蛋白 Y365C 和 E328C 突变随温度增高稳定性增加, ALDH4 蛋白的 K347M 和 T213M 突变最稳定环境出现在 40°C。该修饰预测结果有助于进一步研究该蛋白的特性。

2 讨论

醛脱氢酶超家族蛋白存在于多个亚细胞区域, 包括线粒体、细胞质、内质网和细胞核。超家族各成员单体均具有典型的催化区域、NAD (P)⁺结合域、齐聚反应域和保守的功能位点, 其中均含有一个催化口袋, 由于该口袋的结构限制的配体差异赋予了各成员具有不同的催化活性。

ALDH 同工酶无论在动物、植物还是微生物中都起着至关重要的作用。ALDH 与人类多种疾病呈正相关, 如 ALDH1 与结直肠癌、乳腺癌、非小细胞肺癌的发生相关。ALDH2 与酒精性肝病、消化道肿瘤、白血病、先天性心脏病等密切相关。该家族许多蛋白质已经被定义, 但缺乏相应的系统化归类研究 (Olen et al., 2012)。ALDH1 被认为是参与视黄醛转化为视黄酸代谢的重要蛋白 (Resetskova et al., 2010)。ALDH2 线粒体同工酶在乙醇代谢中将乙醛转化为乙酸并参与硝酸甘油的代谢 (Lang et al., 2012)。在眼角膜中, 紫外诱导脂质过氧化产生醛, ALDH3 能够催化该醛起到解毒作用 (Perozich et al., 2000)。ALDH4 与脯氨酸的降解有关, 能够将脂肪和芳香族醛催化为酸进行脱毒 (Pemberton et al., 2014)。ALDH5 为四聚体催化 γ -氨基丁酸分解代谢的最后一步催化琥珀半醛转化为琥珀酸盐。ALDH6 是唯一已知的依赖于辅酶 A 的人 ALDH。ALDH7 可催化 α -氨基己二半醛氧化成 α -氨基酸, 在嘧啶酸通路中发挥主要作用 (Luo et al., 2015)。本研究通过对 ALDH 超家族进行生物信息学比较分析。发现虽同属 ALDH 家族且蛋白结构

相似, 但由于催化结构、作用位点、亚细胞定位和组织定位的差异赋予了各家族成员不同的催化活性。同时通过信息学方法对不稳定的 ALDH3 和 ALDH4 蛋白进行蛋白修饰和预测, 为深入开展 ALDH 相关研究奠定基础。

3 材料与方法

3.1 数据来源

以 ALDH1~ALDH4 家族为出发点, 在 NCBI 及 PDB 数据库中运行 BLAST 程序, 找出 ALDH 家族核酸序列及蛋白序列各 121 个作为候选, 应用 Vector NTI 软件剔除序列过长或过短的干扰蛋白。最终确定涵盖 ALDH1~ALDH4 及典型 ALDH 亚家族的核酸及蛋白序列 82 个。

3.2 ALDH 亚家族的核酸及蛋白序列的生物信息学分析

应用 Vector NTI 对 ALDH 家族的核酸序列进行对比分析, 应用 ExPasy 数据库中 ProtParam 软件 (<http://www.expasy.org/tools/ProtParam.html>) 对比分析并统计目的蛋白的理化特性包括: 等电点、分子量、氨基酸数目、脂肪系数、亲疏水系数等。在 ExPASy Proteomics Server 网站上用 ProtScale 软件 (<http://web.expasy.org/protscale>) 对 ALDH 蛋白的氨基酸序列进行亲/疏水性预测。Vector NTI 8.0 软件对其进行保守序列分析比对。应用 WoLF PSORT 在线软件 (http://www.genscript.com/psort/wolf_psort.html) 对 ALDH 蛋白进行亚细胞定位预测。TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) 和 Tmpred 软件对蛋白的跨膜区域进行预测。在 PBIL 网站上使用 MLRC 软件对 ALDH 进行二级结构预测。蛋白结构域建模比对分析和分析结果评估通过 Clustal X、MEGA5.0、Swiss Modle、Swiss-PdbView 和 WebLab ViewLite 完成 (常凯等, 2015)。在 I-Mutant 网址进行 SEEF 方法的稳定性模拟。

作者贡献

常凯完成论文构思、数据分析统计和写作,童立冬、江忠勇、张睿和熊怡淞参与材料收集与整理,熊杰完成论文修改。

致谢

本研究由四川省科技厅应用基础研究项目(2013JY0173)基金资助。

参考文献

- Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L., and Schwede T., 2014, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Research*, 42(Web Server issue): W252-W258
- Capriotti E., Fariselli P., and Casadio R., 2005, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.*, 33(Web Server issue): W306-W310
- Chang K., Qiu F., Chen M., Zeng L., Liu X., Yang C., Lan X., Wang Q., and Liao Z., 2014, Engineering the MEP pathway enhanced ajmalicine biosynthesis, *Biotechnol. Appl. Biochem.*, 61(3): 249-255
- Chang K., Xiong Y.S., Qu Y.Q., Wu A.L., Chen Q.H., and Wu L. J., 2015 Comparative analysis of *FOX* family genes based on the bioinformatics methods, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 34(9): 1818-1825 (常凯,熊怡淞,曲远青,吴艾霖,陈清红,吴丽娟,2015,基于生物信息学方法的 *FOX* 基因家族对比分析, *基因组学与应用生物学*, 34(9): 1818-1825)
- Elsohly A.M., and Francis M.B., 2015, Development of oxidative coupling strategies for site-selective protein modification, *Acc. Chem. Res.*, 48(7): 1971-1978
- Fontaine F.R., Dunlop R.A., Petersen D.R., and Burcham P.C., 2002, Oxidative bioactivation of crotyl alcohol to the toxic endogenous aldehyde crotonaldehyde: association of protein carbonylation with toxicity in mouse hepatocytes, *Chem. Res. Toxicol.*, 15(8): 1051-1058
- Izaguirre G., Pietruszko R., Cho S., and MacKerell A.D.J. 2002, Human aldehyde dehydrogenase catalytic activity and structural interactions with coenzyme analogs, *Journal of Biomolecular Structure & Dynamics*, 19(3): 429-447
- Johansson M.U., Zoete V., Michielin O., and Guex N., 2012, Defining and searching for structural motifs using DeepView/Swiss-PdbViewer, *BMC Bioinformatics*, 13: 173
- Lan X., Chang K., Zeng L., Liu X., Qiu F., Zheng W., Quan H., Liao Z., Chen M., Huang W., Liu W., and Wang Q., 2013, Engineering salidroside biosynthetic pathway in hairy root cultures of *Rhodiola crenulata* based on metabolic characterization of tyrosine decarboxylase, *PLoS One*, 8(10): 108-110
- Lang B.S., Gorren A.C.F., Oberdorfer G., Wenzl M.V., Furdui C. M., Poole L.B., Mayer B., and Gruber K., 2012, Vascular bioactivation of nitroglycerin by aldehyde dehydrogenase-2, *J. Biol. Chem.*, 287(45): 38124-38134
- Luo M., and Tanner J.J., 2015, Structural basis of substrate recognition by aldehyde dehydrogenase 7A1, *Biochemistry*, 54(35): 5513-5522
- Mann C.J., and Weiner H., 1999, Differences in the roles of conserved glutamic acid residues in the active site of human class 3 and class 2 aldehyde dehydrogenases, *Protein Science*, 8(10): 1922-1929
- Olen O., Gudjónsdóttir A.H., Browaldh L., Hessami M., Elvin K., Liedberg A.S., Neovius M., and Grahnquist L., 2012, Antibodies against deamidated gliadin peptides and tissue transglutaminase for diagnosis of pediatric celiac disease, *J. Pediatr. Gastroenterol. Nutr.*, 55(6): 695-700
- Pemberton T.A., Srivastava D., Sanyal N., Henzl M.T., Becker D. F., and Tanner J.J., 2014, Structural studies of yeast $\Delta(1)$ -pyrroline-5-carboxylate dehydrogenase (ALDH4A1): active site flexibility and oligomeric state, *Biochemistry*, 53(8): 1350-1359
- Perozich J., Kuo I., Wang B.C., Boesch J.S., Lindahl R., and Hempel J., 2000, Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase, *Eur. J. of Biochem.*, 267(20): 6197-203
- Resetskova E., Reis-Filho J.S., Jain R.K., Mehta R., Thorat M.A., Nakshatri H., and Badve S., 2010, Prognostic impact of ALDH1 in breast cancer: a story of stem cells and tumor microenvironment, *Breast Cancer Res. Treat.*, 123(1): 97-108