

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 ✉ 作者

计算分子生物学, 2016 年, 第 5 卷, 第 5 篇 doi: 10.5376/cmb.cn.2015.05.0005

本文首次发表在《基因组学与应用生物学》(2016 年第 35 卷第 1 期 28-33 页)上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议最佳引用格式:

引用格式(中文):

常凯等, 2016, 醛脱氢酶(ALDH)蛋白家族的信息学比较分析与修饰预测, 计算分子生物学(online), 5(5): 1-7 (doi: 10.5376/cmb.cn.2014.05.0005)

引用格式(英文):

Chang et al., 2016, Comparative Analysis and Mutation Prediction of Aldehyde Dehydrogenase Based on the Bioinformatics Methods, Computational Molecular Biology (online), 5(5): 1-7 (doi: 10.5376/cmb.cn.2014.05.0005)

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 Jie¹

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 ✉ Authors

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.

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

收稿日期: 2016 年 07 月 10 日

接受日期: 2016 年 07 月 10 日

发表日期: 2016 年 07 月 10 日

基金项目: 本研究由四川省科技厅应用基础研究项目(2013JY0173)基金资助。

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

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, 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。

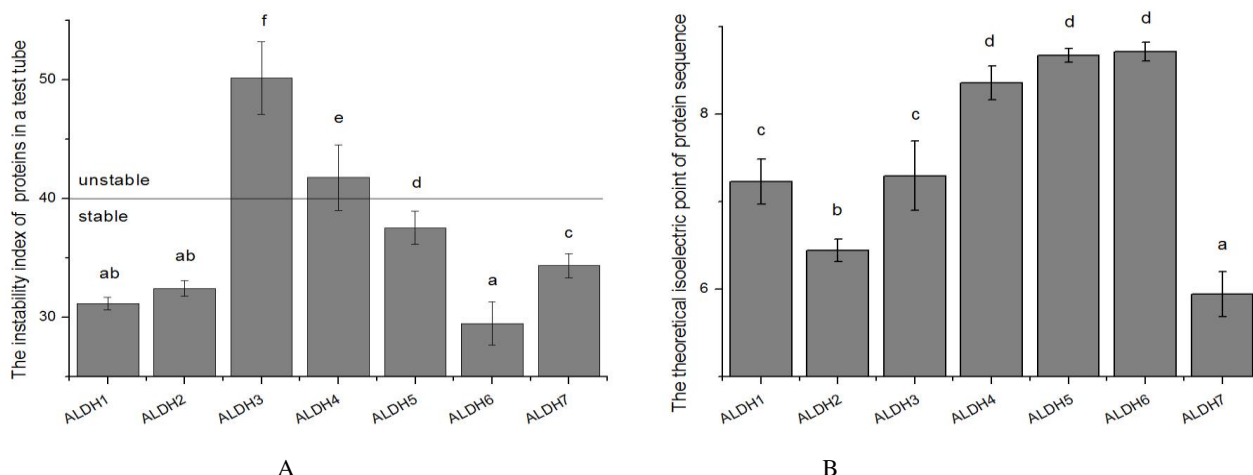


图2 ALDH蛋白序列的稳定性分析与等电点统计分析

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 ($p < 0.01$)

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%延伸带、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%延伸带、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%延伸带、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%延伸带、8.61% β 转角和29.75%随意卷曲构成(图3)。

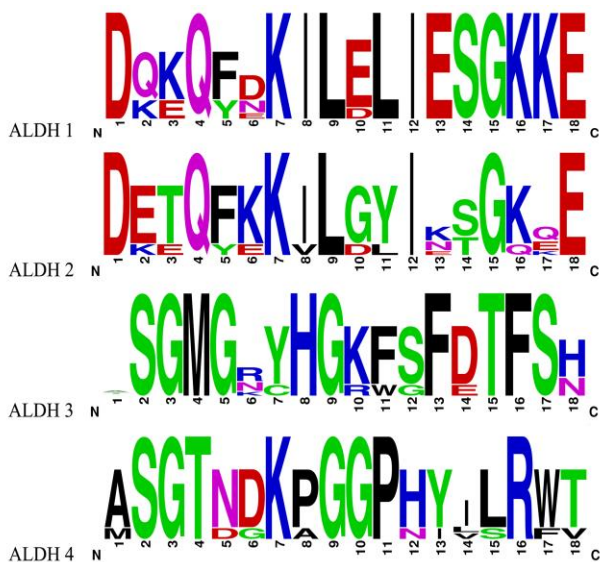


图3 NAD(P)⁺结合区域的氨基酸序列保守性分析

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-MODLE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想合理性评估(Biasini, 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, 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, 2015)。应用I-mutant2.0软件计算DDG值(Kcal/mol)如表1，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(Emidio, 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℃。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。

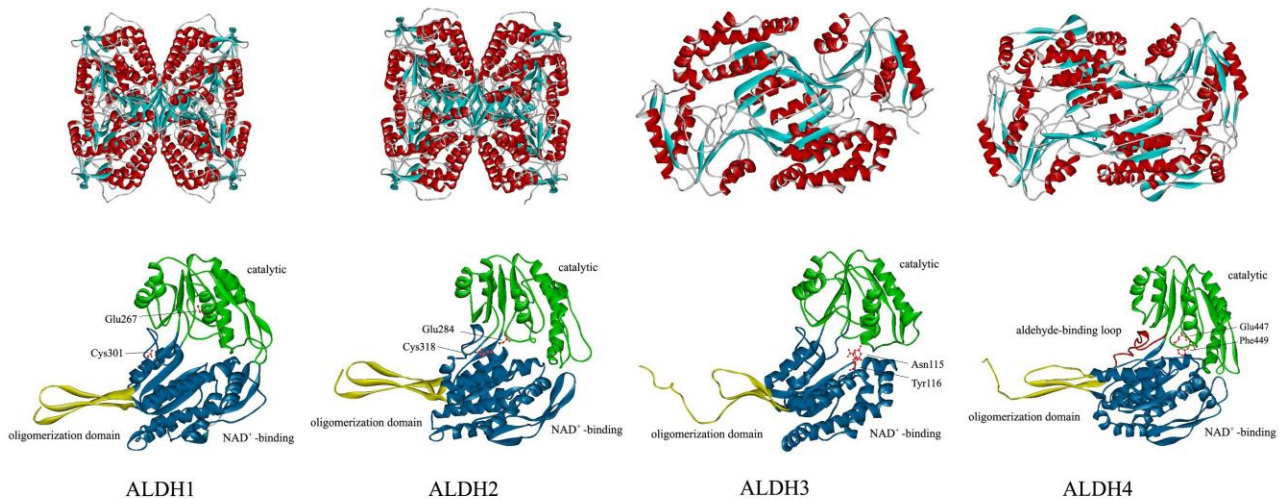


图4 ALDH家族四级结构建模与分析

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 sequence 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	突变位点	DDG (10°C)	DDG (25°C)	DDG (40°C)	DDG (55°C)	DDG (70°C)	pH值
Names	Mutation site						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

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, 2012)。ALDH1被认为是参与视黄醛转化为视黄酸代谢的重要蛋白(Erika, 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, 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, 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, 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, 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 NTI8.0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 Model、Swiss-PdbView和WebLab ViewLite完成(常凯, 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、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, *Jiyinzhexue 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 Deep-View/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