

DOI:10.12113/202203013

基于生物信息学方法分析人 CREB 结合蛋白的结构与功能

葛永飞, 骆洁雅, 叶子, 林文娅, 黄劲*

(贵州医科大学 基础医学院, 贵阳 550025)

摘要:为了预测分析人 CREB 结合蛋白(CREB-binding protein, CBP)的结构和功能。本研究利用生物信息学相关数据库及软件对人 CBP 蛋白的理化性质、保守性、亚细胞定位、信号肽、跨膜结构域、二级结构、三级结构、相互作用蛋白及功能进行预测。结果表明,人 CBP 蛋白是一种定位于核内的不稳定亲水性蛋白质,无跨膜区和信号肽。其二级结构以无规卷曲和 α -螺旋为主,并且该蛋白质 HAT 结构域在各物种间高度保守,推测与其酶活性密切相关的氨基酸残基为 Tyr1433、Leu1434、Asp1435、Arg1664。此外,人 CBP 蛋白能够与 TP53、CREB1、NCAO3 等多种转录因子或转录辅激活因子发生相互作用,主要参与转录调控、细胞分化、组织发育、信号转导及细胞凋亡等生物学过程。本研究为进一步研究 CBP 在恶性肿瘤发生发展中的作用机制提供了理论依据。

关键词:人 CBP 蛋白;理化性质;结构;生物信息学

中图分类号:Q51 文献标志码:A 文章编号:1672-5565(2023)02-121-07

Analysis of structure and function of human CREB-binding protein based on bioinformatics

GE Yongfei, LUO Jieya, YE Zi, LIN Wenya, HUANG Jin*

(School of Basic Medicine, Guizhou Medical University, Guiyang 550025, China)

Abstract:In order to predict and analyze the structure and function of human CREB-binding protein (CBP), bioinformatics methods were used to predict the physical and chemical properties, conservation, subcellular localization, signal peptide, transmembrane domain, secondary structure, tertiary structure, interacting protein, and function of human CBP. Results showed that human CBP was an unstable and hydrophilic protein located in the nucleus without transmembrane region and signal peptide. The secondary structure was characterized by random coil and α -helix. The HAT domain of the protein was highly conserved among species. It was speculated that the amino acid residues closely related to its enzyme activity were Tyr1433, Leu1434, Asp1435, and Arg1664. In addition, human CBP could interact with transcription factors and transcription coactivators, such as TP53, CREB1, and NCAO3, and was mainly involved in biological processes such as transcription regulation, cell differentiation, tissue development, signal transduction, and apoptosis. This study provides a theoretical basis for further studying the mechanism of CBP in the occurrence and development of malignant tumors.

Keywords:Human CBP; Physicochemical properties; Structure; Bioinformatics

CREB 结合蛋白(CREB-binding protein, CBP)于1993年被 Chrivia 等首次报道,因其能够特异性地结合磷酸化的环腺苷酸应答元件结合蛋白(cAMP-response element binding protein, CREB)而得名^[1]。CBP 作为一种重要的组蛋白乙酰基转移酶

(Histone acetyltransferase, HAT)和转录辅助因子,通过对乙酰化以及转录进程的调节,在体内诸多生理过程中发挥着重要作用^[2-4],包括细胞分化、细胞增殖、细胞凋亡以及细胞自噬。并通过调节细胞周期,参与多种肿瘤的恶性转化及转移^[5]。

收稿日期:2022-03-23;修回日期:2022-05-24;网络首发日期:2022-06-10.

网络首发地址:<https://kns.cnki.net/kcms/detail/23.1513.Q20220610.0819.002.html>

基金项目:贵州省2017年大学生创新创业训练计划项目(No.DC201710660028).

*通信作者:黄劲,女,副教授,研究方向:生物化学与分子生物学.E-mail: huangjin@gmc.edu.cn.

研究发现, CBP 蛋白的功能与其结构中的多个保守区域密切相关^[5], 如转录适配器锌指结构域 (Transcriptional adapter zinc finger domain, TAZ)、激酶诱导的 CREB 相互作用结构域 (Kinase inducible domain of CREB interacting domain, KIX)、溴域 (Bromodomain, BRD)、ZZ 型锌指结构域 (ZZ-type zinc finger domain, ZZ) 和 HAT 催化结构域 (HAT domain)。通过这些功能结构域的介导, CBP 可与许多转录激活因子或转录因子相互作用, 并乙酰化组蛋白, 从而重塑染色质结构, 参与基因转录的调控。越来越多的研究证据表明^[6-8], CBP 基因的突变以及 HAT 酶活性的异常与多种肿瘤疾病密切相关。然而, CBP 参与肿瘤发生发展的作用机制尚不清晰。

生物信息学作为现今生命科学领域的有力工具, 利用其相关方法分析蛋白质的信息使之助力相关研究的开展已经成为一种常态^[9-10]。本研究基于生物信息学对 CBP 蛋白的理化性质、亲/疏水性、亚细胞定位、信号肽、跨膜区域、二级结构、三级结构进行分析并构建 CBP 蛋白相互作用网络与分子进化

树, 为 CBP 参与恶性肿瘤发生发展调控机制的深入研究奠定基础。

1 材料和方法

1.1 材料

自 UniProt 网站下载人 (*Homo sapiens*, Q92793)、黑猩猩 (*Pan troglodytes*, K7DBC8)、恒河猴 (*Macaca mulatta*, F6RRA8)、牛 (*Bos taurus*, F1MD32)、马 (*Equus caballus*, F6PZC1)、狗 (*Canis familiaris*, F1PY87)、小型棕褐蝠 (*Myotis lucifugus*, G1NVG2)、小鼠 (*Mus musculus*, P45481)、大鼠 (*Rattus norvegicus*, Q6JHU9)、爪蟾 (*Xenopus laevis*, Q5U248)、斑马鱼 (*Danio rerio*, A0A2R8Q4T2)、果蝇 (*Drosophila melanogaster*, O01368)、秀丽隐杆线虫 (*Caenorhabditis elegans*, P34545) 的 CBP 蛋白序列。

1.2 方法

运用生物信息学相关数据库及分析软件对人 CBP 蛋白有关信息进行分析与预测, 各数据库及分析软件信息见表 1。

表 1 本文应用的数据库及分析软件

Table 1 Database and software for protein analysis

数据库/分析软件	网址	主要功能
ProtParam	https://web.expasy.org/protparam/	蛋白质理化性质分析
ProtScale	https://web.expasy.org/protscale/	蛋白质亲/疏水性分析
PSORT II	https://psort.hgc.jp/form2.html	蛋白质亚细胞定位分析
SignalP-5.0 server	http://services.healthtech.dtu.dk/service.php?SignalP-5.0	蛋白质信号肽序列分析
TMHMM Server v.2.0	http://services.healthtech.dtu.dk/service.php?TMHMM-2.0	蛋白质跨膜结构分析
SOPMA	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html	蛋白质二级结构分析
SWISS-MODEL	https://swissmodel.expasy.org/	蛋白质三级结构建模
WHAT-IF	https://swift.cmbi.umcn.nl/servers/html/index.html	蛋白质三级结构优化
The Structure Analysis and Verification Server	https://saves.mbi.ucla.edu/	蛋白质结构拉式图分析
String	https://string-db.org/	构建蛋白质相互作用网络
ClustalX2.1	离线软件	蛋白质同源序列分析
MEGA6	离线软件	构建系统进化树

2 结果

2.1 人 CBP 蛋白理化性质分析

采用 ProtParam 工具对人 CBP 蛋白的理化性质进行分析, 结果显示人 CBP 蛋白共有 2 442 个氨基酸残基, 总分子式为 $C_{11451}H_{18225}N_{3381}O_{3590}S_{145}$, 分子质量为 265 350.87, 理论等电点为 8.83。在构成 CBP 的氨基酸中, 脯氨酸 (Pro, 276 个)、谷氨酰胺 (Gln, 268 个) 以及丝氨酸 (Ser, 218 个) 的含量较高, 并且带负

电 (Asp+Glu) 和正电 (Arg+Lys) 氨基酸残基的总数分别是 183 和 214。此外, 其不稳定系数、脂肪系数和平均总亲水值分别为 65.91、60.95 和 -0.690, 提示人 CBP 为不稳定亲水蛋白质。

2.2 人 CBP 蛋白的同源性比较分析

HAT 结构域是 CBP 行使乙酰化催化功能的区域, 该区域的保守氨基酸可能与其酶活性密切相关。CBP 蛋白的 HAT 结构域氨基酸的多序列比对分析 (见图 1) 显示, 人 CBP 蛋白与黑猩猩、恒河猴、牛、马、狗、小型棕褐蝠、小鼠、大鼠、爪蟾、斑马鱼、果蝇

以及秀丽隐杆线虫中的同源蛋白相似度分别为 99.96%、99.47%、94.48%、95.83%、93.38%、94.13%、95.26%、95.71%、74.31%、69.76%、68.83%、48.71%。HAT 结构域的保守氨基酸集中于 1 394 位~1 400 位、1 407 位~1 415 位、1 428 位~1 440 位、1 449~1 457 位、1 470 位~1 509 位、

1 537 位~1 548 位、1 626 位~1 633 位、1 659 位~1 665 位、1 677 位~1 683 位等区段。另外,CBP 蛋白同源进化树分析结果(见图 2),亲缘关系与人类最为接近的是黑猩猩和恒河猴,其次为牛、马、狗等其它哺乳类动物,而果蝇、线虫与人类的亲缘关系较远。

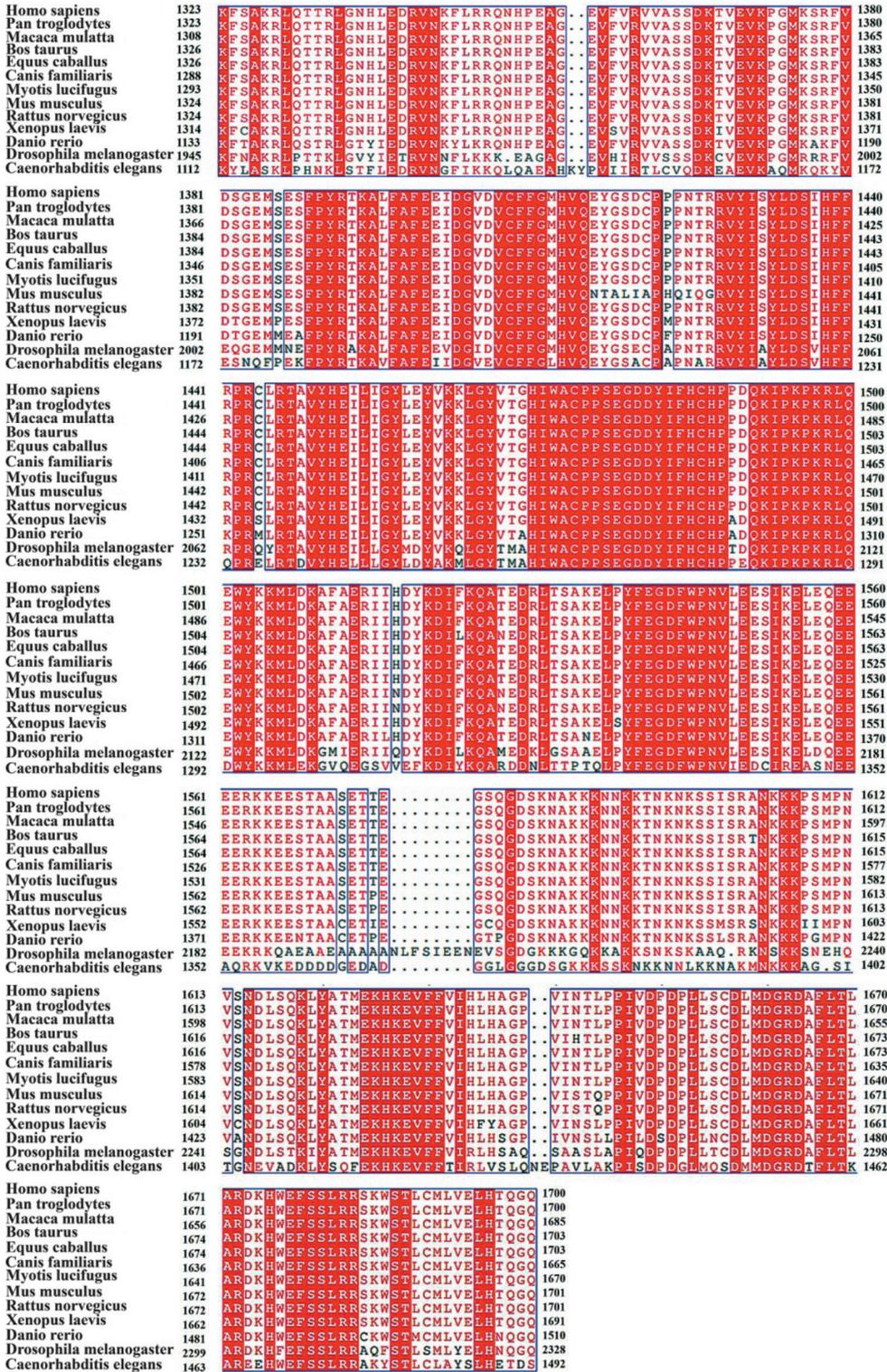


图 1 不同物种 CBP 蛋白 HAT 结构域多序列比对

Fig. 1 Multiple sequence alignment of HAT domain of CBP in different species

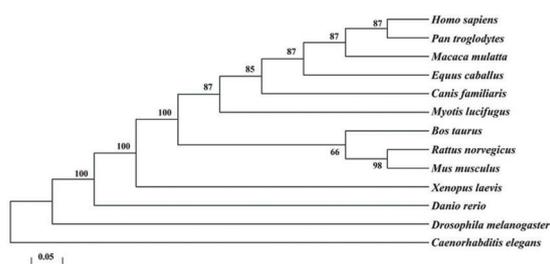


图 2 CBP 蛋白 HAT 结构域同源进化树

Fig. 2 Homology evolutionary tree of CBP based on HAT domain

2.3 人 CBP 蛋白的亲水性/疏水性预测分析

通过在线分析软件 ProtScale 的 Hphob/Kyte&Doolittle 算法分析人 CBP 蛋白的亲疏水性(见图 3)。人 CBP 蛋白中位于负值区的氨基酸占 77.71% (1 980/2 432), 其总分为 $-1\,944.087$, 亲水性最强的位点为 1 561 位、1 562 位的谷氨酸以及 1 563 位的精氨酸, 其分值为 -3.7 ; 位于正值区的氨基酸占 21.8% (532/2 432), 其总分为 254.966 , 疏水性最强的位点为 1 635 位的异亮氨酸, 其分值为 1.967 。预测结果表明, 人 CBP 蛋白的大部分氨基酸处于亲水区, 故推测该蛋白为亲水蛋白质。

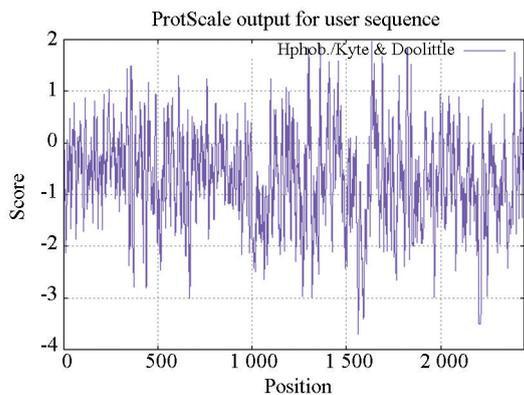


图 3 人 CBP 蛋白亲/疏水性分析

Fig. 3 Analysis of CBP hydrophilicity/hydrophobicity in human

2.4 人 CBP 蛋白的亚细胞定位分析

通过在线分析软件 PSORT II 对人 CBP 蛋白的亚细胞定位进行预测, 结果显示人 CBP 蛋白定位于细胞核、细胞质、细胞膜上的概率分别为 82.6%、8.7%、8.7%, 表明人 CBP 蛋白存在于细胞核内的概率较大。

2.5 人 CBP 蛋白的信号肽及跨膜区域分析

通过在线分析软件 SignalP5.0^[11] 对人 CBP 蛋白的信号肽序列进行分析(见图 4), 其中 OTHER 的数值为 0.999 6, SP 的数值为 0.000 4, 表明人 CBP 蛋白不存在信号肽序列, 可以推测人 CBP 蛋白不是

分泌蛋白。进一步利用在线分析软件 TMHMM Server v.2.0 对人 CBP 蛋白的跨膜结构进行分析(见图 5), 表明人 CBP 蛋白不存在跨膜区域, 提示人 CBP 蛋白可能为非跨膜蛋白质。

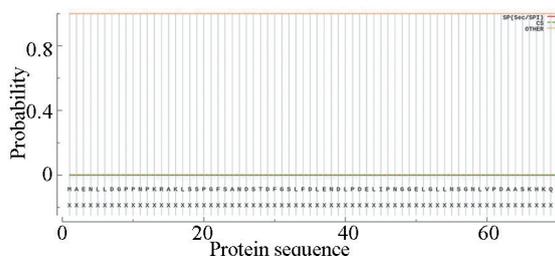


图 4 人 CBP 蛋白信号肽序列预测

Fig. 4 Signal peptide prediction of CBP in human

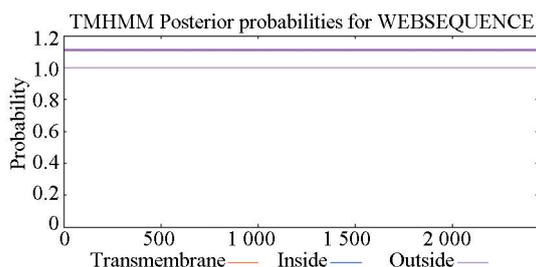


图 5 人 CBP 蛋白跨膜区域预测

Fig. 5 Transmembrane structure prediction of CBP in human

2.6 人 CBP 蛋白的二级结构分析

蛋白质序列二级结构预测软件 SMOPA 的分析结果(见图 6)显示, 人 CBP 蛋白中 α -螺旋、 β -折叠、 β -转角和无规卷曲所占比例分别为 30.47%、9.95%、6.43% 和 53.15%, 其中无规卷曲和 α -螺旋是人 CBP 蛋白二级结构中数量较多的结构。

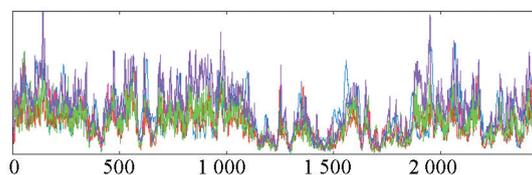


图 6 人 CBP 蛋白二级结构预测

Fig. 6 Secondary structure prediction of CBP in human

注: 蓝色, α -螺旋; 紫色, 无规卷曲; 红色, β -折叠; 绿色, β -转角。

2.7 人 CBP 蛋白三级结构模型的构建及优化

将人 CBP 蛋白氨基酸序列提交至 SWISS-MODEL^[12] 中, 选取序列一致性为 87.05%, 相似性为 59%, 包含整个 HAT 结构域的模板 (*Homo sapiens*, PDB ID: 6gyr) 进行同源建模(见图 7)。根据相似波形图(见图 8)显示, 大部分氨基酸残基得分大于 0.6, 表明该模型具有一定的可靠性。

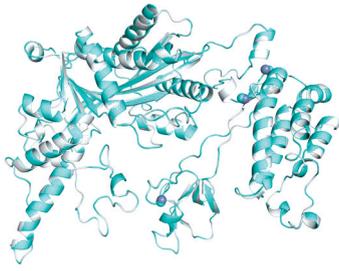


图 7 人 CBP 蛋白优化前后的三级结构比较图

Fig. 7 Comparison of tertiary structure of human CBP before and after optimization

注:白色为优化前结构;蓝色为优化后结构。

为提高酶活性中心关键氨基酸残基预测的准确性,采用 WHAT-IF 优化该蛋白三级结构^[13],并利用拉式图分析网站 The Structure Analysis and Verification Server 对优化前后模型进行分析以评估该模型的可靠度。经优化后,落在拉氏图的红色以及黄色区域外的氨基酸残基由 2% 下降到了 0.7%,而落在不合理区域氨基酸的数量明显减少,表明优

化后模型的稳定性和可信度均有较大提高(见图 9)。

进一步结合序列比对结果(见图 1)分析该模型,发现 HAT 结构域保守序列共同组成了一个圆形的凹陷结构,推测该结构内的 Tyr1433、Leu1434、Asp1435、Arg1664 为酶活性中心的关键氨基酸位点。

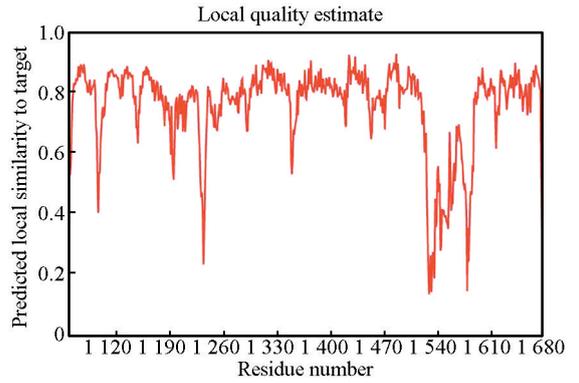
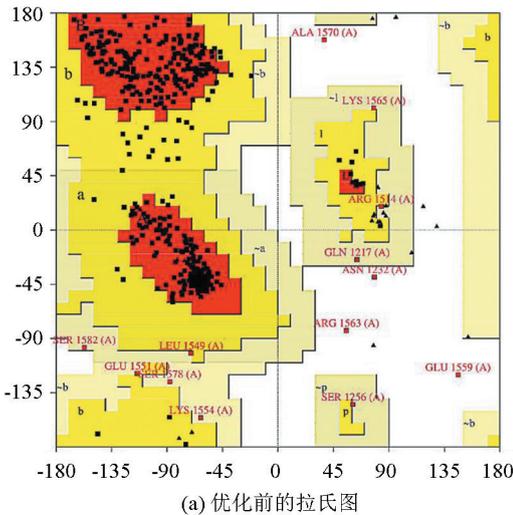
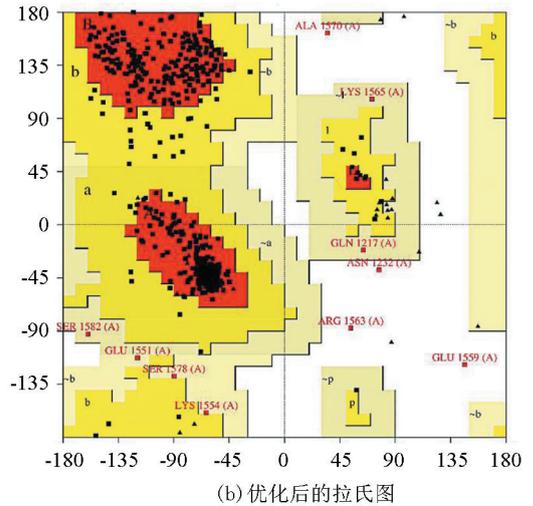


图 8 人 CBP 蛋白三级结构与模板相似波形图

Fig. 8 Waveform of similarity between tertiary structure of human CBP and template



(a) 优化前的拉氏图



(b) 优化后的拉氏图

图 9 人 CBP 蛋白 HAT 结构域三级结构优化前后拉氏图

Fig. 9 Ramachandran plot of tertiary structure of HAT domain of human CBP before and after optimization

2.8 人 CBP 相互作用蛋白质预测及其功能富集分析

采用String 11.0 蛋白质相互作用预测软件^[14]分析与人 CBP 互作的蛋白网络结果(见图 10)所示,相互作用的蛋白质分别为 CREB1、NCOA3、TP53、KAT2B、RELA/P65、SMAD3、MYB、PPARG、CTNNB1、HTT、STAT1。进一步采用 Gene Ontology 功能富集分析,预测结果(见图 11)显示,人 CBP 蛋白主要参与转录调控、细胞分化、组织发育、信号传导及细胞凋亡等生物学过程,具有与转录因子和染

色质结合、转录辅激活因子和乙酰转移酶等生物学功能,而细胞组分分析显示为染色质、核质及染色体等。KEGG 通路富集分析(见图 12)显示,人 CBP 蛋白主要涉及的通路,包括甲状腺激素和 Wnt 信号通路、细胞周期、及病毒感染和肿瘤等疾病。上述预测结果表明,人 CBP 蛋白能够通过转录因子和转录辅激活因子在内的多种转录调节因子相结合,从而调控转录,参与细胞分化、细胞凋亡、组织发育及细胞周期等生物学过程。

胞定位、信号肽、跨膜区域、二级结构、三级结构以及相互作用蛋白质进行分析和预测,研究结果为深入探究人 CBP 蛋白在肿瘤恶性化进程中的作用机制提供了理论依据。

参考文献(References)

- [1] CHRIVIA J C, KWOK R P, LAMB N, et al. Phosphorylated CREB binds specifically to the nuclear protein CBP[J]. *Nature*, 1993, 365(6449):855-859. DOI: 10.1038/365855a0.
- [2] MEDRANO-FERNANDEZ A, DELGADO-GARCIA J M, DEL-BLANCO B, et al. The epigenetic factor CBP is required for the differentiation and function of medial ganglionic eminence-derived interneurons[J]. *Molecular Neurobiology*, 2019, 56(6):4440-4454. DOI: 10.1007/s12035-018-1382-4.
- [3] WONG C K, WADE-VALLANCE A K, LUCIANI D S, et al. The p300 and CBP transcriptional coactivators are required for β -cell and α -cell proliferation[J]. *Diabetes*, 2018, 67(3):412-422. DOI: 10.2337/db17-0237.
- [4] XU T J, LIU Y, LI P, et al. Insulin in combination with selenium inhibits HG/Pal-induced cardiomyocyte apoptosis by Cbl-b regulating p38MAPK/CBP/Ku70 pathway[J]. *European Review for Medical and Pharmacological Sciences*, 2016, 20(15):3297-3303. PMID: 27467007.
- [5] ZHANG X F, WEI B F, ZHANG X, et al. Current development of CBP/p300 inhibitors in the last decade[J]. *European Journal of Medicinal Chemistry*, 2021, 209:112861. DOI:10.1016/j.ejmech.2020.112861.
- [6] KIM M S, YOO N J, LEE S H. Expressional and mutational analysis of CREBBP gene in gastric and colorectal cancers with microsatellite instability[J]. *Pathology & Oncology Research*, 2014, 20(1):221-222. DOI: 10.1007/s12253-013-9668-3.
- [7] PARK S, STANFIELD R L, MARTINEZ-YAMOUT M A, et al. Role of the CBP catalytic core in intramolecular SUMOylation and control of histone H3 acetylation[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(27):e5335-e5342. DOI: 10.1073/pnas.1703105114.
- [8] ZHANG J, VLASEVSKA S, WELLS V A, et al. The CREBBP acetyltransferase is a haploinsufficient tumor suppressor in B-cell lymphoma[J]. *Cancer Discovery*, 2017, 7(3):322-337. DOI:10.1158/2159-8290.CD-16-1417.
- [9] 张献, 彭涛, 张耀, 等. 氨基甲酸乙酯水解酶家族生物信息学分析[J]. *生物信息学*, 2022, 20(1):64-74. DOI:10.12113/202011002.
ZHANG Xian, PENG Tao, ZHANG Yao, et al. Bioinformatics analysis of urethane hydrolase family[J]. *Chinese Journal of Bioinformatics*, 2022, 20(1):64-74. DOI: 10.12113/202011002.
- [10] 黄劲, 王洁, 牛雪可, 等. 结核分枝杆菌 pyrF 基因的生物信息学分析[J]. *中国病原生物学杂志*, 2019, 14(2):125-130+136. DOI: 10.13350/j.cjpb.190201.
HUANG Jin, WANG Jie, NIU Xueke, et al. Bioinformatics analysis of the pyrF gene encoding protein orotidine 5'-monophosphate decarboxylase in *Mycobacterium tuberculosis*[J]. *Journal of Pathogen Biology*, 2019, 14(2):125-130+136. DOI: 10.13350/j.cjpb.190201.
- [11] ALMAGRO J J, TSIRIGOS K D, SØNDERBY C K, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks[J]. *Nature Biotechnology*, 2019, 37(4):420-423. DOI: 10.1038/s41587-019-0036-z.
- [12] BIENERT S, WATERHOUSE A, DE BEER T A P, et al. The SWISS-MODEL Repository-new features and functionality[J]. *Nucleic Acids Research*, 2017, 45(D1):D313-d319. DOI: 10.1093/nar/gkw1132.
- [13] LANGE J, BAAKMAN C, PISTORIUS A, et al. Facilities that make the PDB data collection more powerful[J]. *Protein Science*, 2020, 29(1):330-344. DOI:10.1002/pro.3788.
- [14] SZKLARCZYK D, MORRIS J H, COOK H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible[J]. *Nucleic Acids Research*, 2017, 45:D362-D368. DOI: 10.1093/nar/gkw937.
- [15] WANG L, ZHANG L, GONG X D, et al. PP-1 β and PP-2A α modulate cAMP response element-binding protein (CREB) functions in aging control and stress response through de-regulation of α B-crystallin gene and p300-p53 signaling axis[J]. *Aging Cell*, 2021, 20(9):e13458. DOI: 10.1111/ace1.13458.
- [16] WEINERT B T, NARITA T, SATPATHY S, et al. Time-resolved analysis reveals rapid dynamics and broad scope of the CBP/p300 acetylome[J]. *Cell*, 2018, 174(1):231-244.e12. DOI: 10.1016/j.cell.2018.04.033.
- [17] 吴小霞, 许天委, 李国寅. 小鼠 APOBEC3 的结构与功能分析[J]. *生物信息学*, 2021, 19(1):66-74. DOI:10.12113/202005006.
WU Xiaoxia, XU Tianwei, LI Guoyin. Structural and functional analysis of mouse APOBEC3[J]. *Chinese Journal of Bioinformatics*, 2021, 19(1):66-74. DOI: 10.12113/202005006.
- [18] WADDELL A R, HUANG H, LIAO D, et al. CBP/p300: critical co-activators for nuclear steroid hormone receptors and emerging therapeutic targets in prostate and breast cancers[J]. *Cancers (Basel)*, 2021, 13(12):2872. DOI: 10.3390/cancers13122872.
- [19] LASKO L M, JAKOB C G, EDALJI R P, et al. Discovery of a selective catalytic p300/CBP inhibitor that targets lineage-specific tumours[J]. *Nature*, 2017, 550(7674):128-132. DOI: 10.1038/nature24028.
- [20] LEE S Y, KIM J J, MILLER K M. Bromodomain proteins: protectors against endogenous DNA damage and facilitators of genome integrity[J]. *Experiment and Molecular Medicine*, 2021, 53(9):1268-1277. DOI: 10.1038/s12276-021-00673-0.
- [21] ZHANG F C, SUN Z Y, LIAO L P, et al. Discovery of novel CBP bromodomain inhibitors through TR-FRET-based high-throughput screening[J]. *Acta Pharmacologica Sinica*, 2020, 41(2):286-292. DOI: 10.1038/s41401-019-0256-2.
- [22] NARITA T, WEINERT B T, CHOUDHARY C. Functions and mechanisms of non-histone protein acetylation[J]. *Nature Reviews Molecular Cell Biology*, 2019, 20(3):156-174. DOI: 10.1038/s41580-018-0081-3.