

DOI:10.13350/j.cjpb.230201

• 论著 •

新型冠状病毒相关蛋白 DPP1 的生物信息学分析及分子对接研究*

钱思彤^{1,2}, 李琳^{1,2}, 杨婷婷^{1,2}, 刘艺^{3**}

(1. 徐州医科大学药学院, 江苏徐州 221004; 2. 江苏省新药研究与临床药学重点实验室;
3. 长江大学附属第一医院口腔科)

【摘要】 **目的** 利用生物信息学和分子对接方法分析预测新冠病毒相关蛋白二肽基肽酶 1(Dipeptidyl peptidase 1, DPP1)的特征信息,为探究其作用机制和靶向药物的研发提供理论依据。 **方法** 通过 NCBI 获得 DPP1 氨基酸序列,利用 ProtParam、PredictProtein、PSORT、ProtScale、SignalP、TMHMM、SOMPA、PROSITE、SWISS-MODEL、SYBYL-X、STRING 和 IEBD 等软件或服务对其理化性质、亚细胞定位、信号肽、跨膜区、二、三级结构、活性中心、对接结合口袋、B/T 细胞抗原表位及相互作用蛋白等进行预测分析。 **结果** DPP1 由 463 个氨基酸残基构成,有 1 个信号肽序列,无跨膜区,是稳定的亲水性分泌蛋白。二级结构以无规卷曲为主,存在 3 个活性中心和 2 个功能结构域,涉及去除底物末端二肽的肽酶功能。三级结构模型显示其为四聚体结构,蛋白单体与小分子抑制剂的活性结合口袋由 VAL40、LYS63、VAL65、THR67、VAL142、GLY143、THR144、ALA145、ASN148 和 THR150 等核心氨基酸组成。DPP1 主要与 10 个蛋白存在相互作用,具有 10 个 B 细胞抗原表位和 14 个 T 细胞抗原表位。 **结论** DPP1 为含有信号肽序列的稳定亲水分泌蛋白,该蛋白含有 B、T 细胞抗原表位,该研究为 DPP1 在 COVID-19 重症化中的作用机制和开发抗新冠靶向药物奠定了基础。

【关键词】 新型冠状病毒;二肽基肽酶 1;生物信息学;分子对接

【中图分类号】 R373.1

【文献标识码】 A

【文章编号】 1673-5234(2023)02-0125-07

[*Journal of Pathogen Biology*. 2023 Feb;18(2):125-131.]

Bioinformatic analysis and molecular docking studies of SARS-CoV-2 related protein DPP1

QIAN Si-tong^{1,2}, LI Lin^{1,2}, YANG Ting-ting^{1,2}, LIU Yi³ (1. College of Pharmacy, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China; 2. Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy; 3. Department of stomatology, The First Affiliated Hospital of Changjiang University)***

【Abstract】 **Objective** The characteristic information of SARS-CoV-2 related protein DPP1 was analyzed and predicted by bioinformatics and molecular docking method, providing theoretical basis for exploring its mechanism and developing targeted drugs. **Methods** According to the amino acid sequence of DPP1, the physical and chemical properties, subcellular localization, signal peptide, transmembrane region, secondary structure, active center, B cell and T cell epitopes and interacting protein were predicted and analyzed by ProtParam, PredictProtein, PSORT, ProtScale, SignalP, TMHMM, SOMPA, PROSITE, STRING, IEBD etc software or on-line servers. The tertiary structure of DPP1 was built by SWISS-MODEL. Active binding pockets of DPP1 were defined according to the docking results by SYBYL-X. **Results** Being composed of 463 amino acid residues, DPP1 was a stable hydrophilic secretory protein which had one signal peptide sequence, but no transmembrane region. Random coil was main secondary structure component. DPP1 contained 3 active centers and 2 functional domains which referred to the peptidase function by removing N-terminal dipeptide of substrate. The 3 active centers were belonged to the active site of sulfhydryl protease. The tertiary structure simulation showed that DPP1 was a tetramer. The docking results showed that the binding pockets were composed of 10 amino acids, which were described as VAL40, LYS63, VAL65, THR67, VAL142, GLY143, THR144, ALA145, ASN148 and THR150. Among them, LYS63 and THR144 were believed to play an important role in binding the small molecule inhibitors by the formation of hydrogen bond. DPP1 mainly interacted with 10 proteins. It was predicted that DPP1 had 10 B cell epitopes and 14 T cell epitopes. **Conclusion** DPP1 was a stable hydrophilic secretory protein with signal peptide sequence, and it had B and T cell epitopes. These results laid a foundation for further research on the mechanism

* **【基金项目】** 国家自然科学基金项目(No. 81903689)。

** **【通讯作者】** 刘艺, E-mail: xiaoyi_99@163.com

【作者简介】 钱思彤(1992-), 女, 湖北天门人, 硕士, 实验师, 主要从事病原微生物相关研究。E-mail: 1779127130@qq.com

of DPP1 in COVID-19 severity and the development of targeted drugs against COVID-19.

【Key words】 SARS-CoV-2; dipeptidyl peptidase 1; bioinformatics; molecular docking

由新型冠状病毒(SARS-CoV-2)引起的新型冠状病毒肺炎(COVID-19)疫情正处于全球大流行阶段,其传播速度快,传播隐蔽,感染范围广^[1-2],给世界公共卫生事业带来巨大挑战,严重威胁人类的生命健康。人感染新冠病毒后会出现不同的临床转归,根据症状可分为无症状感染、轻症和重症等^[3]。重症患者出现急性呼吸窘迫综合征、脓毒症、休克甚至死亡^[4]。有研究发现,机体的遗传因素在新冠病毒感染和病程进展中发挥重要作用,新冠肺炎的重症化也与之密切相关^[5-9]。针对中国人群新冠肺炎的全基因组测序结果和关联性分析揭示了多种影响新冠肺炎重症化的遗传易感因素,其中多个重症易感基因可能成为新冠治疗的潜在药物靶点^[9-10]。

已有研究表明,位于人类染色体 11q14.2 的 Cathepsin C(CTSC)基因是新冠重症化的遗传易感候选基因^[10],它编码一种溶酶体半胱氨酸蛋白酶,又称二肽基肽酶 1(Dipeptidyl peptidase 1, DPP1)。DPP1 可以去除丝氨酸蛋白酶 N 端抑制性二肽,激活该酶,进而参与多种免疫和炎症反应^[11-12]。同时,它还是多种炎症和免疫相关疾病的治疗靶点^[12]。DPP1 抑制剂 Brensocatib 也被考虑为新冠治疗候选药物,用于治疗重症新冠患者^[13]。然而,DPP1 在 COVID-19 重症化及其在 SARS-CoV-2 入侵人体过程所发挥的作用尚未完全明确,以上因素极大制约了 DPP1 作为潜在新冠治疗靶点的发展。本研究利用生物信息学和分子对接的方法对 CTSC 基因及编码的 DPP1 蛋白进行系统性分析,深入阐述其结构和功能特性,为进一步揭示其在 COVID-19 重症化中的作用机制奠定基础,对于研发靶向 DPP1 的抗新冠药物具有重要意义。

材料与方法

1 材料

Cathepsin C(CTSC)基因序列(Gene ID:1075)和 Dipeptidyl peptidase 1(DPP1)氨基酸序列(GenBank 登录号:NP_001805.4)均下载自 NCBI 网站。

2 方法

2.1 DPP1 编码基因 CTSC 的一般特性分析 利用 NCBI 网站检索 CTSC 基因,分析其一般特性,包括位置、长度及转录产物等。

2.2 DPP1 一级结构与理化性质分析 利用 ProtParam 服务器分析 DPP1 的一级结构与理化性质,利用 PredictProtein 服务器验证上述结果。

2.3 DPP1 的亚细胞定位、跨膜区、信号肽及亲疏水

性分析 利用 PSORT 服务器预测 DPP1 的细胞亚定位,ProtScale 服务器预测目标蛋白的亲疏水性。利用 SignalP 4.1 和 TMHMM 2.0 服务器分别预测 DPP1 的氨基酸序列是否含有信号肽和跨膜区结构。

2.4 DPP1 的二级结构分析 利用 SOMPA 服务器预测 DPP1 的二级结构,利用 COILS 预测 DPP1 的卷曲螺旋域,利用 PROSITE 和 NCBI 保守性结构域检索工具预测分析 DPP1 的功能位点。

2.5 DPP1 三级结构建模 利用 SWISS-MODEL 线上服务器对 DPP1 进行蛋白三级结构同源建模,并用拉式图评价模型效果。

2.6 DPP1 与小分子抑制剂结合模式分析 通过文献调研^[14]发现 4 个已知 DPP1 小分子抑制剂,利用 ChemDraw 软件将小分子化合物转变为 SDF 格式,并保存。利用 SYBYL-X 2.0 软件将上述小分子对接到 SWISS-MODEL 所建的 DPP1 蛋白模型中:先用 Minimize 模块将小分子配体能量最小化,Gradient 设定为 0.005 kcal/(mol * Å),Max Iterations 设定为 1000,然后预处理 DPP1 蛋白单体,最后利用 Docking 模块进行对接操作。对接结果使用 PyMOL 软件呈现。

2.7 预测 DPP1 的相互作用蛋白 将 DPP1 蛋白名称输入 STRING 网站,Organisms 选择 Homo sapiens,分析与 DPP1 相互作用的蛋白及蛋白互作网络的功能富集结果。

2.8 预测 B、T 细胞抗原表位 将 DPP1 的氨基酸序列输入 IEBD 线上服务器,分析预测 DPP1 的 B/T 细胞优势抗原表位,B 细胞表位分析筛选阈值设定为 0.5,T 细胞表位分析筛选阈值设定为 0.95。

结果

1 CTSC 基因的一般特性

CTSC 基因位于人类染色体 11q14.2 上,包含 9 个外显子,可编码 3 种可变剪切产物,转录产物全长 1 870 bp,编码的(DPP1)为 CTSC 的共识编码序列,是由 463 个氨基酸组成的多肽(表 1)。

表 1 CTSC 基因可变剪切产物列表
Table 1 List of variable splicing product of CTSC

mRNA 编号 mRNA ID	蛋白产物编号 Protein ID	氨基酸数 Amino acid number
NM_001814.6	NP_001805.4	463
NM_148170.5	NP_680475.1	137
NM_001114173.3	NP_001107645.1	141

2 DPP1 的一级结构与理化性质

ProtParam 分析显示 DPP1 由 463 个氨基酸残基组成,其中 Gly(9.3%)占比最高,Trp(2.2%)最低(表 2,图 1)。氨基酸残基中带正电的 Arg 和 Lys 共 42 个,带负电的 Asp 和 Glu 共 45 个,理论等电点(pI)为 6.53,推测 DPP1 为偏酸性蛋白。DPP1 相对分子质量为 51.85382×10^3 ,分子式为 $C_{2332}H_{3521}N_{615}O_{680}S_{25}$,半衰期大于 30 h,不稳定系数为 36.05,表明 DPP1 蛋白性质稳定。PredictProtein 预测结果与 ProtParam 基本相符。

表 2 DPP1 的氨基酸组成
Table 2 Amino acid composition of DPP1

氨基酸残基 Amino acid residue	个数 Number	比例(%) Proportion
Ala (A)	31	6.7
Arg (R)	17	3.7
Asn (N)	23	5.0
Asp (D)	21	4.5
Cys (C)	14	3.0
Gln (Q)	15	3.2
Glu (E)	24	5.2
Gly (G)	43	9.3
His (H)	14	3.0
Ile (I)	21	8.0
Leu (L)	25	5.4
Lys (K)	18	4.8
Met (M)	11	2.4
Phe (F)	20	4.3
Pro (P)	19	4.1
Ser (S)	31	6.7
Thr (T)	28	6.0
Trp (W)	10	2.2
Tyr (Y)	29	6.3
Val (V)	30	6.5

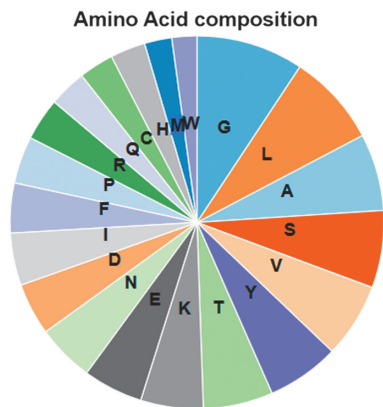


图 1 PredictProtein 预测 DPP1 的氨基酸组成
Fig. 1 Prediction of DPP1 amino acid composition by PredictProtein

3 DPP1 的亚细胞定位、跨膜区、信号肽及亲疏水性

PSORT 预测 DPP1 主要位于细胞质中 (Reliability:94.1),具体分布为:内质网 33.3%,细胞外(包括细胞壁)33.3%,液泡 22.2%,线粒体 11.1%。将 DPP1 的氨基酸序列输入 SignalP 4.1 服务器,结果

如图 2。DPP1 存在 1 个信号肽序列,由 1-24 位的氨基酸组成,后续 25-463 位氨基酸组成成熟链。TMHMM 2.0 服务器预测 DPP1 不存在跨膜序列(图 3),推测其可能是分泌性蛋白。利用 ProtScale 对 DPP1 的亲疏水性进行预测分析,结果显示 DPP1 多肽链中第 11 位的丙氨酸和 12 位的亮氨酸疏水值均为 3.36,是疏水性最强的氨基酸,而第 162 位赖氨酸的疏水值为-2.81,是亲水性最强的氨基酸。整条多肽链中亲水性氨基酸残基数量明显多于疏水性氨基酸残基(图 4),推测 DPP1 为亲水性蛋白。

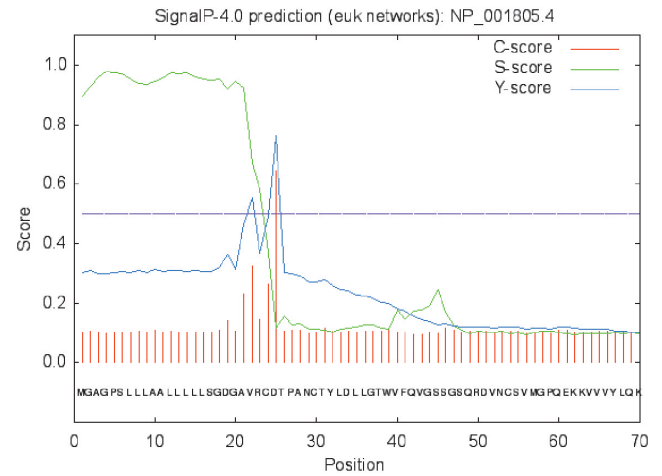


图 2 SignalP 预测 DPP1 信号肽结构
Fig. 2 Prediction of DPP1 signal peptide structure by SignalP

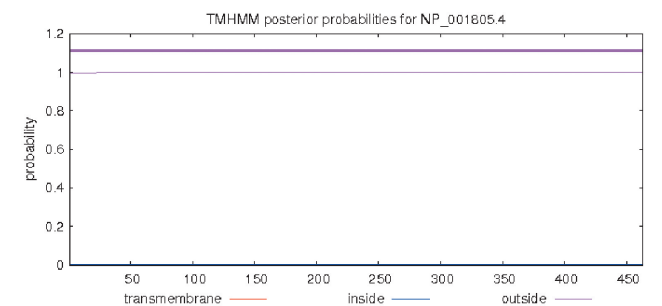


图 3 TMHMM 预测 DPP1 跨膜结构
Fig. 3 Prediction of DPP1 transmembrane structure by TMHMM

4 DPP1 的二级结构

利用 SOMPA 预测分析 DPP1 的二级结构中 α 螺旋 (Alpha helix) 占 22.68% (105 个氨基酸残基),无规卷曲 (Random coil) 占 52.70% (244 个氨基酸残基),延伸链 (Extended strand) 占 18.57% (86 个氨基酸残基), β 转角 Beta turn 占 6.05% (28 个氨基酸残基),表明 DPP1 二级结构的主要构成原件为无规卷曲,其次是 α 螺旋(图 5)。

PROSITE 预测 DPP1 包含 3 个活性中心,分别是 巯基(半胱氨酸)蛋白酶半胱氨酸活性位点 (252-263 位氨基酸),巯基(半胱氨酸)蛋白酶组氨酸活性位点

(403-413 位氨基酸), 巯基(半胱氨酸)蛋白酶天冬酰胺活性位点(422-441 位氨基酸)(图 6)。利用 NCBI 分析 DPP1 存在 2 个保守结构域, 分别属于 Cathepsin C exclusion 超家族和肽酶 C1 超家族结构域(图 7)。

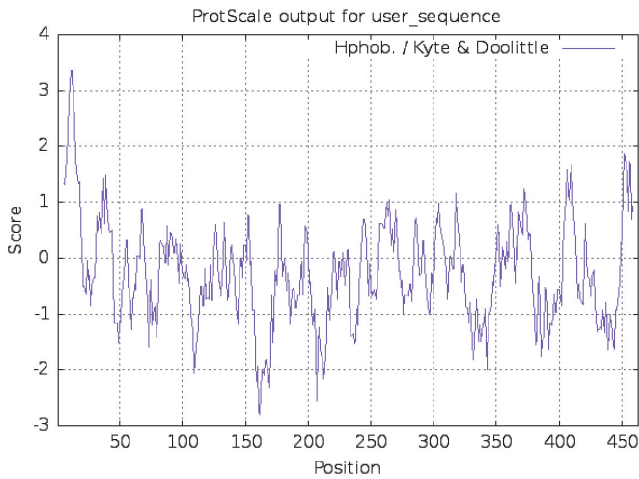


图 4 ProtScale 预测 DPP1 的亲疏水性
Fig. 4 Prediction of DPP1 hydrophilicity by TMHMM

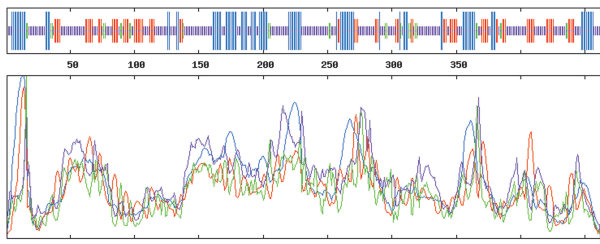


图 5 SOPMA 预测 DPP1 二级结构
Fig. 5 Prediction of DPP1 Secondary structure by SOPMA



图 6 PROSITE 预测 DPP1 活性中心
Fig. 6 Prediction of DPP1 functional site by PROSITE

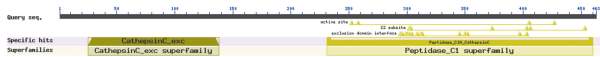


图 7 NCBI 分析 DPP1 功能结构域
Fig. 7 Analysis of DPP1 functional domain by NCBI

5 DPP1 的卷曲螺旋域分析

COILS 预测结果如图 8 所示, DPP1 在 Window = 14 时卷曲概率较高, 但数值仍小于 0.5, 推测 DPP1 不含卷曲螺旋结构。

6 DPP1 三级结构建模

利用 SWISS-MODEL 以 PDB ID 3pdf.1 为模板对 DPP1 进行同源建模, 序列一致性为 99.54%, 相似度为 0.63, 覆盖率为 0.95, 功能预测为二肽基肽酶 1。模型如图 9 所示, 其总体评分为 0.58。利用

Ramachandran 图评价所建模型, 有 85.31% (395/463) 的氨基酸位点位于较佳区域, 96.97% (449/463) 的氨基酸位点位于允许区域(图 10), 表明所建模型准确可靠。

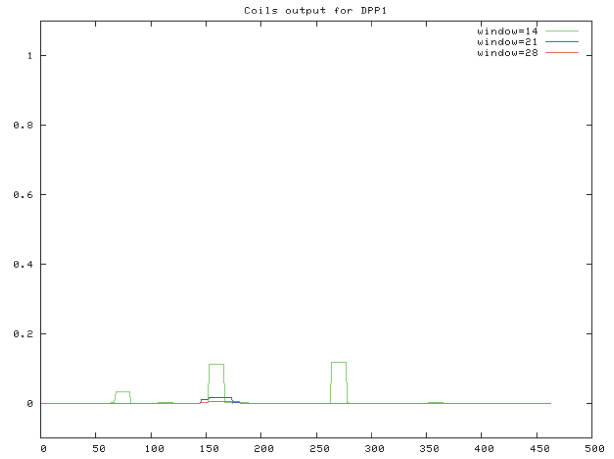


图 8 COILS 分析 DPP1 卷曲螺旋结构
Fig. 8 Analysis of DPP1 coiled helix structure by COILS

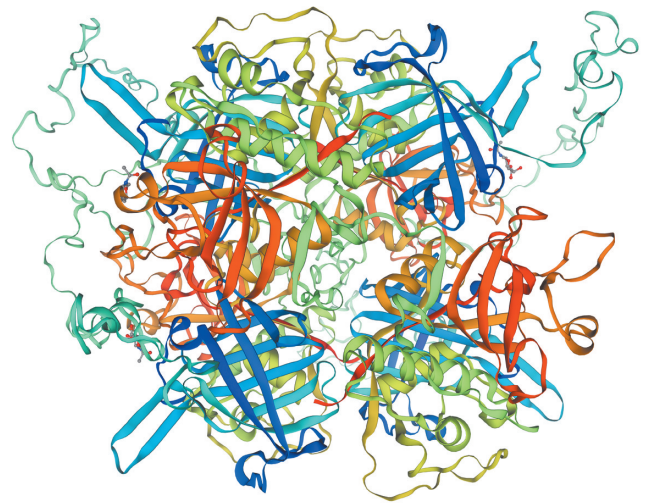


图 9 SWISS-MODEL 预测 DPP1 三级结构
Fig. 9 Prediction of DPP1 tertiary structure by SWISS-MODEL

7 DPP1 与小分子抑制剂结合模式分析

SWISS-MODEL 预测 DPP1 模型为四聚体结构, 拆出其单体。4 种已知的小分子抑制剂化学结构如图 11 所示。利用 SYBYL-X 2.0 软件将小分子抑制剂对接到蛋白单体中, 分析二者之间结合模式, 结果如图 12 和表 3。DPP1 单体与化合物 1 的活性结合口袋由 12 个核心氨基酸组成, 二者之间存在 3 个氢键作用: THR38 与化合物 1 五元环氧原子形成氢键, LYS63 与七元环氧原子形成氢键, THR144 与七元环氨基氢形成氢键; DPP1 单体与化合物 2 的活性结合口袋由 14 个核心氨基酸组成, 存在 2 个氢键作用: LYS63 与羰基氧形成氢键, THR144 与氨基氢形成氢键; DPP1 单体与化合物 3 的活性结合口袋由 12 个核

心氨基酸组成,存在3个氢键作用:THR144分别与羰基氧和氨基氢形成2个氢键,LYS63与羰基氧原子形成氢键;DPP1单体与化合物4的活性结合口袋由13个核心氨基酸组成,存在3个氢键作用:THR144与2个氨基氢形成2个氢键,LYS63与羰基氧原子形成氢键。通过分析DPP1单体与小分子抑制剂的结合模式,显示其活性口袋的组成包含10种共性氨基酸,分别为VAL40、LYS63、VAL65、THR67、VAL142、GLY143、THR144、ALA145、ASN148和THR150,其中LYS63和THR144作为可以与小分子抑制剂形成氢键的关键氨基酸,需要特别关注。

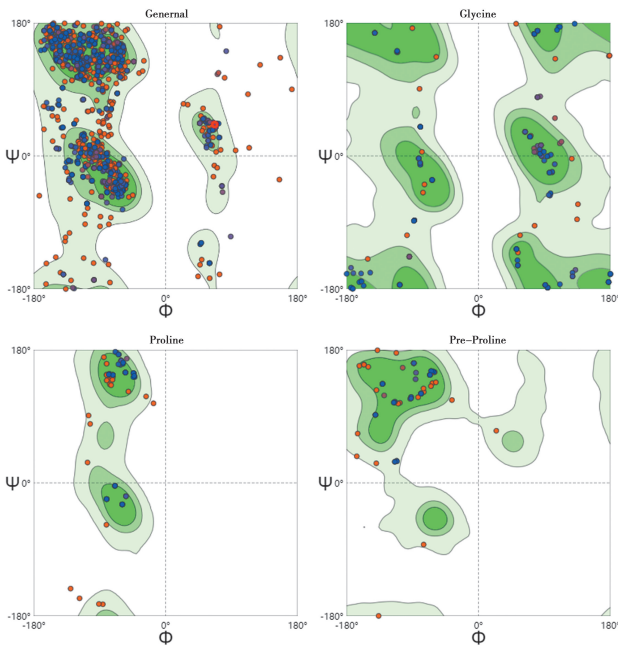


图10 Ramachandran评价DPP1的三级结构模型

Fig. 10 Ramachandran evaluation of DPP1 tertiary structure model

表3 DPP1单体与小分子抑制剂对接结果

Table 3 The details of small molecule inhibitors docked to DPP1 monomer

配体 Ligand	结合口袋组成氨基酸 Amino acids of binding pocket	形成氢键关键氨基酸 Core amino acids of hydrogen bond	对接 打分 Score
化合物1	THR38 VAL40 LYS63 VAL65 THR67 ASP77 VAL142	THR38	4.9596
	GLY143 THR144 ALA145 ASN148	LYS63	
	THR150	THR144	
化合物2	THR38 VAL40 LYS63 VAL65 THR67 ASP77	LYS63	5.2567
	VAL142 GLY143 THR144 ALA145 GLU147 ASN148	THR144	
	THR150		
化合物3	VAL151	LYS63	5.8411
	VAL40 GLN42 LYS63 VAL65 LYS140 VAL142 GLY143	THR144	
	THR144 ALA145 ASN148	THR150 VAL151	
化合物4	THR38 VAL40 LYS63 VAL65 THR67 ASP76	THR38	5.5665
	ASP77 VAL142 GLY143 THR144 ALA145	LYS63	
	ASN148 THR150	THR144	

8 DPP1相互作用蛋白

STRING预测DPP1与其他10个蛋白存在相互作用,且均为人类白细胞II类抗原(HLA Class II),结

果如图13。对此作用网进行功能富集(GO分析),生物过程(BP)主要集中在肽抗原与主要组织相容性复合体(MHC)结合、正调控病毒进入宿主细胞和单核细胞分化;分子功能(MF)集中在MHC II类受体活性,结合肽抗原和CD4受体;细胞成分(CC)为MHC II类蛋白复合体和内质网侧腔的组成成分。

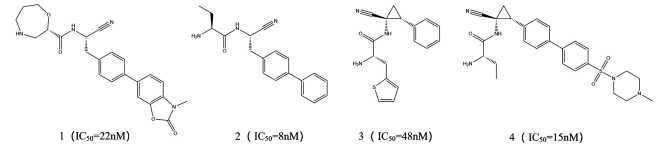
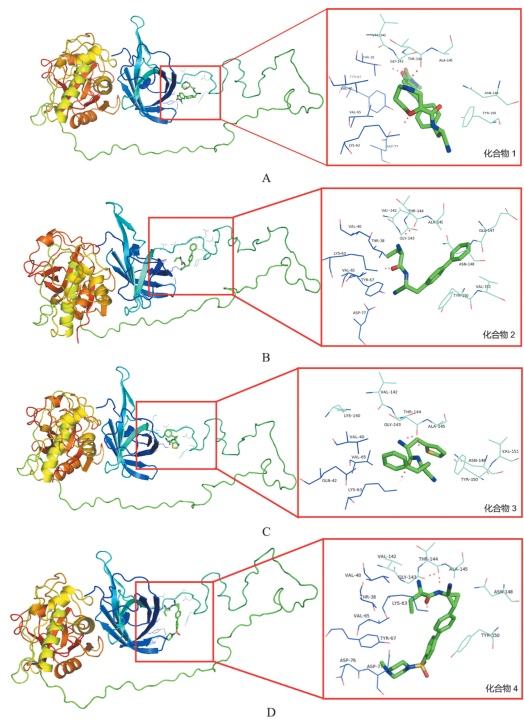


图11 DPP1小分子抑制剂的化学结构及IC₅₀值

Fig. 11 Chemical structure of small molecule inhibitors of DPP1. IC₅₀ values were mentioned in brackets.



A 化合物1 B 化合物2 C 化合物3 D 化合物4

图12 4种小分子抑制剂与DPP1单体的对接结合模式

A Compound 1 B Compound 2 C Compound 3 D

Compound 4

Fig. 12 Binding modes of four small molecule inhibitors docked to DPP1 monomer

9 B、T细胞抗原表位

为了更好地评价DPP1的抗原属性,利用IEBD服务器分析预测DPP1蛋白的B细胞表位,结果如图14和表4。以0.5为阈值可以预测到10个B细胞抗原表位,总体结果平均得分为0.488,最高得分为0.693,最低得分为0.154。利用IEBD服务器预测DPP1的T细胞表位,从MHC结合结果看,以0.95为阈值可预测到14个T细胞表位,其中最高得分为0.997462,最低得分为0.955764(表5)。

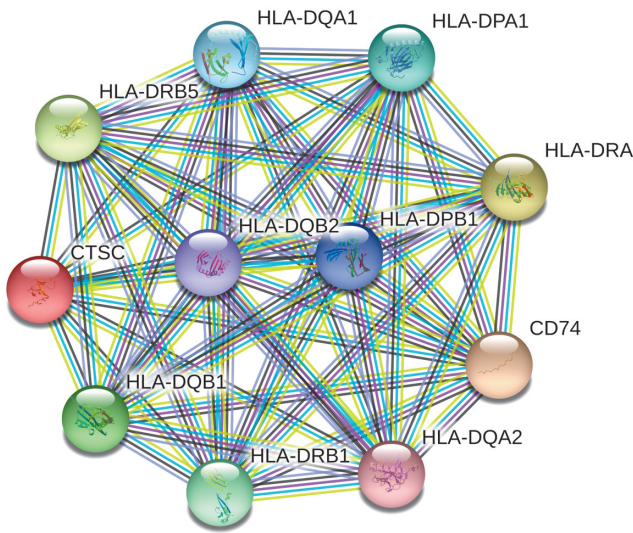


图 13 STRING 预测 DPP1 相互作用蛋白
Fig. 13 Prediction of interaction protein with DPP1 by STRING

表 4 IEBD 预测 DPP1 的 B 细胞表位信息
Table 4 B cell epitopes analysis results of DPP1 via IEBD

序号 No.	始于 Start	终于 End	氨基酸序列 Peptide	长度 Length
1	46	61	SGSQRDVNCSVMGPQE	16
2	71	80	LDTAYDDLGN	10
3	108	170	KEEGSKVTTYCNETMTGWVHVDVL GRNWACFTGKKVGTASENVVYVNIA HLKNSQEKYSNRLYKY	63
4	173	173	N	1
5	178	253	INAIQKSWTATTYMEYETLTLGDMI RRSGGHSRKIPRKPAPLTAIEIQQKIL HLPTSWDWRNVHGINFVSPVRNQA	76
6	275	283	TNNSQTPIIL	9
7	313	357	DFGLVEEACFPYTGTDSPCKMKED CFRYSSEYHYVGGFYGGCNE	45
8	379	402	DFLHYKKGIIYHHTGLRDPFNPTEL	24
9	415	421	DSASGMD	7
10	431	436	TGWGEN	6

表 5 IEBD 预测 DPP1 的 T 细胞表位信息
Table 5 T cell epitopes analysis results of DPP1 via IEBD

等位基因 Allele	始于 Start	终于 End	长度 Length	氨基酸序列 Peptide	得分 Score
HLA-B * 07:02	212	220	9	IPRPKAPL	0.997462
HLA-A * 26:01	375	383	9	EVYDDLFLHY	0.993868
HLA-B * 57:01	176	185	10	KAINAIQKSW	0.991947
HLA-B * 44:03	109	117	9	EEGSKVTTY	0.981308
HLA-B * 55:01	214	222	9	RPKPAPLTA	0.980793
HLA-B * 15:01	181	190	10	IQKSWTATTY	0.978694
HLA-B * 40:02	60	68	9	QEKKVVVYL	0.976387
HLA-A * 11:01	55	63	9	SVMGPQEKK	0.975487
HLA-B * 15:02	67	75	9	YLQKLDLAY	0.970264
HLA-B * 38:01	366	374	9	HHGPMVAVAF	0.967974
HLA-B * 35:01	302	310	9	FPYLIAGKY	0.962888
HLA-A * 32:01	227	235	9	KILHLPTSW	0.960746
HLA-A * 01:01	413	422	10	GTDSASGMDY	0.957409
HLA-B * 18:01	109	117	9	EEGSKVTTY	0.955764

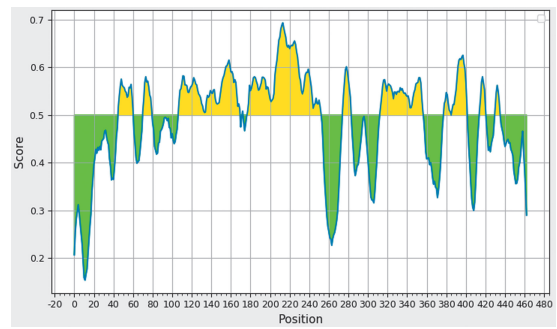


图 14 IEBD 预测 DPP1 的 B 细胞表位
Fig. 14 Prediction of B cell epitopes of DPP1 via IEBD

讨论

研究表明, SARS-CoV-2 感染人体后, 因为个体间的性别、年龄、基础健康状况等差异因素会造成不同的临床转归, 具体表现为无症状、轻症和重症等^[15]。同时, 新冠病毒感染和病程进展与机体的遗传因素亦密切相关。针对人群进行全基因组遗传关联性分析, 筛选与 COVID-19 重症相关的易感基因, 极大助力于新冠重症患者的早期发现、早期预测和早期干预, 具有重大的现实意义^[16-17]。而利用生物信息学手段提前对易感基因和相应蛋白进行系统分析, 不仅节省了实验时间和成本, 缩小研究范围, 更为探索其在 COVID-19 疾病进程中所发挥作用及抗新冠靶向药物的开发打下坚实基础。

Li 等^[10]报道位于人类染色体 11q14.2 的 CTSC 基因是新冠重症化的遗传易感基因之一, 其编码的 DPP1 蛋白是一种二肽基肽酶, 是某些炎症和免疫疾病的重要治疗靶点。分析蛋白一级结构, DPP1 由 463 个氨基酸组成, 理论等电点为 6.53, 不稳定系数为 36.05, 多肽链中亲水性氨基酸残基数量明显多于疏水性氨基酸残基, 推测 DPP1 是性质稳定的亲水性蛋白, 且偏酸性, 这些特性参数为 DPP1 的体外表达、蛋白纯化和功能研究奠定了基础。DPP1 二级结构中无规卷曲占比最多, 该结构易形成潜在的抗原表位, 且有利于维持蛋白构象的稳定性^[18]。DPP1 主要定位于细胞质, 以内质网居多, 含有 1 个信号肽序列, 无跨膜结构, 推测其属于分泌蛋白。此外, DPP1 含有 3 个活性中心, 属于巯基(半胱氨酸)蛋白酶的不同氨基酸活性位点, 可以协调激活一系列丝氨酸蛋白酶, 参与免疫反应。DPP1 存在 2 个保守结构域, 分别属于 Cathepsin C exclusion 超家族和肽酶 C1 超家族, 其中 Cathepsin C exclusion 是由 8 条反向平行的 β 链组成的 β 桶结构, 属于排除结构域, 排斥多肽链除末端以外的接触^[19]。以上结果进一步证实了 DPP1 的肽酶属性, 并为 DPP1 去除底物 N 端二肽的酶学功能提供了结构依据。

SWISS-MODEL 建模分析显示, DPP1 由 4 条多

肽链组成,是典型的四聚体结构。将4种结构不同的DPP1小分子抑制剂对接到DPP1单体中,发现活性结合口袋包含10种共性氨基酸,分别为VAL40、LYS63、VAL65、THR67、VAL142、GLY143、THR144、ALA145、ASN148和THR150,其中LYS63和THR144可以与小分子抑制剂形成氢键作用,属于结合口袋中的核心氨基酸。明确结合口袋关键氨基酸组成和蛋白与小分子抑制剂之间的重要作用力为进一步筛选和研发新的DPP1抑制剂奠定了基础,可极大节省研发时间和研发成本。对DPP1相互作用的蛋白网络进行功能富集,发现其生物过程、分子功能及细胞成分均与MHC相关,推测DPP1可能参与外源性抗原的递呈过程,与机体免疫应答相关^[20]。DPP1免疫细胞表位预测显示其具有10个B细胞表位和14个T细胞表位,这些抗原表位的挖掘有利于找出具有类似免疫保护功能及抗原特性但毒性相对较低的抗原片段,助力开发免疫原性较强的高效疫苗。

本研究利用生物信息学和分子对接的方法系统分析了DPP1的特性、结构与功能,可为进一步探究DPP1在新冠重症化中的作用机制提供参考,对发掘其潜在抑制剂和研发靶向抗新冠药物具有重要意义。

【参考文献】

- [1] Zu ZY, Jiang MD, Xu PP, et al. Coronavirus Disease 2019 (COVID-19): A Perspective from China[J]. Radiology, 2020, 296(2):E15-E25.
- [2] Habas K, Nganwuchu C, Shahzad F, et al. Resolution of coronavirus disease 2019 (COVID-19)[J]. Expert Rev Anti-Infect Ther, 2020, 18(12):1201-1211.
- [3] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China [J]. Lancet, 2020, 395(10223):497-506.
- [4] Wu, Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus Disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention[J]. JAMA, 2020, 323(13):1239-1242.
- [5] Zhang Q, Bastard P, Liu Z, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19 [J]. Science, 2020, 370(6515):4570-4585.
- [6] Van der Made CI, Simons A, Schuur-Hoeijmakers J, et al. Presence of Genetic Variants Among Young Men With Severe COVID-19[J]. JAMA, 2020, 324(7):663-673.
- [7] Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in COVID-19[J]. Nature, 2021, 591(7848):92-98.
- [8] Shelton JF, Shastri AJ, Ye C, et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity[J]. Nat Genet, 2021, 53(6):801-808.
- [9] Wang F, Huang S, Gao R, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility[J]. Cell Discov, 2020, 6(1):83.
- [10] Li Y, Ke Y, Xia X, et al. Genome-wide association study of COVID-19 severity among the Chinese population [J]. Cell Discov, 2021, 7(1):76.
- [11] Mallen-St Clair J, Pham CT, Villalta, et al. Mast cell dipeptidyl peptidase I mediates survival from sepsis[J]. J Clin Invest, 2004, 113(4):628-634.
- [12] Korkmaz B, Caughey GH, Chapple I, et al. Therapeutic targeting of cathepsin C: from pathophysiology to treatment [J]. Pharmacol Ther, 2018(190):202-236.
- [13] Korkmaz B, Lesner A, Marchand-Adam S, et al. Lung protection by cathepsin C inhibition: A new hope for COVID-19 and ARDS? [J]. J Med Chem, 2020, 63(22):13258-13265.
- [14] Korkmaz B, Lesner A, Wysocka M, et al. Structure-based design and in vivo anti-arthritis activity evaluation of a potent dipeptidyl cyclopropyl nitrile inhibitor of cathepsin C [J]. Biochem Pharmacol, 2019(164):349-367.
- [15] Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage[J]. J Transl Med, 2020, 18(1):206.
- [16] Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, et al. Genomewide association study of severe Covid-19 with respiratory failure[J]. N Engl J Med, 2020, 383(16):1522-1534.
- [17] Namkoong H, Eda Hiro R, Takano T, et al. DOCK2 is involved in the host genetics and biology of severe COVID-19[J]. Nature, 2022, 609(7928):754-760.
- [18] 戴姿薇, 唐标. 新型冠状病毒相关 TMPRSS2 蛋白结构特征和抗原表位分析[J]. 微生物学杂志, 2021, 41(1):58-68.
- [19] Dufour E. Sequence homologies, hydrophobic profiles and secondary structures of cathepsins B, H and L: comparison with papain and actinidin[J]. Biochimie, 1988, 70(10):1335-1342.
- [20] Milinski M. A review of suggested mechanisms of MHC odor signaling[J]. Biology (Basel), 2022, 11(8):1187.

【收稿日期】 2022-09-15 【修回日期】 2022-12-01