

新冠病毒相关蛋白 *TMPRSS2* 结构与功能的生信分析及表达载体构建

徐本锦¹ 李卓禧² 范 蕾² 李 璟² 严荣荣³ 杜 森¹ 宣 焱¹ 门 杰² 陈晓聪² ,
汤文婷² 侯艳香¹ 宋彬妤⁴ 杨娅男¹ 刘晓梁⁴ 余骏骁² 刘 玲¹

(山西医科大学汾阳学院 1. 医学检验系、2. 基础医学部、3. 山西省汾阳医院检验科、4. 科技中心 山西 汾阳 032200)

doi: 10.12360/CPB202108035

文献标志码: A 文章编号: 1001-1978(2022)08-1218-09

中国图书分类号: R341.81; R373; R392.11; R394; R977.6

摘要: 目的 人源 TMPRSS2 是一种跨膜丝氨酸蛋白酶, 本文对该蛋白结构与功能进行系统生物信息学分析, 优化密码子并构建原核表达载体, 以探究其参与 SARS-CoV-2 感染宿主细胞的分子机制。方法 利用分子克隆技术构建重组表达

载体 pET-22b-1; 采用 ProtParam、NetPhos3.1、Blast、Clustal X2 和 MEGA7.0 等分析工具对 TMPRSS2 蛋白的同源性、功能位点、亚细胞定位、三维结构和进化特点等进行生信分析。结果 原核表达质粒构建正确, TMPRSS2 蛋白属于中等分子量蛋白,

收稿日期: 2022-04-11 修回日期: 2022-06-22

基金项目: 山西省基础研究计划(自由探索类)项目(No 20210302123397); 山西省高等学校科技创新项目(No 2020L0749); 国家级大学生创新创业训练计划项目(No 202117114001); 山西省高等学校大学生创新创业训练计划重点项目(No S202117114008); 吕梁市科技计划项目(No 2020SHFZ29); 山西医科大学汾阳学院引进人才启动金项目(No 2020A01)

作者简介: 徐本锦(1987-), 男, 博士, 讲师, 研究方向: 基因工程和蛋白质翻译调控, E-mail: bj0726@sxmu.edu.cn;
李卓禧(2002-), 女, 本科生, 研究方向: 病原分子生物学, E-mail: 2813533607@qq.com;
刘 玲(1983-), 女, 硕士, 助教, 研究方向: 病原分子生物学, 通信作者, E-mail: ll772x@sxmu.edu.cn

OGD/R model was significantly increased by !-asarone, with the increase of administration dose, the cells in the low, medium and high dose groups of !-asarone decreased, and the "amoeba-like" cells and the cell body were gradually became stereoscopic and full. From the results of cell morphology, it could be seen that !-asarone had a certain proliferative effect on normal cells; the release was significantly reduced of proinflammatory factor IL-1', IL-18 and TNF-! in OGD/R injured BV2 cells pretreated with !-asarone, also increased the release of IL-10, IL-4 and TGF-! , with a dose-effect relationship, and the high dose (16 -mol · L⁻¹) was the best; the expressions of inflammatory related protein NLRP3, caspase 1, NF-(B and

ROS activity in injured cells of OGD/R model were significantly reduced after pretreatment with !-asarone.

Conclusions !-asarone has a significant protective effect on cerebral ischemia/reperfusion injury, mainly by regulating ROS activity and inhibiting phosphorylation of NF-(B, in order to reduce the excessive activation of NLRP3 inflammatory corpuscles reducing the secretion of proinflammatory factor IL-1' and IL-18, promoting the secretion of anti-inflammatory factor IL-10 and IL-4, so as to protect cerebral ischemia/reperfusion injury by anti-inflammatory reaction.

Key words: !-asarone; cerebral ischemia/reperfusion injury; microglia; inflammatory reaction; NLRP3; NF-(B

2019年新冠病毒(SARS-CoV-2)的出现及其在世界范围内的迅速传播造成了严重的全球公共卫生安全事件^[1]。截至2021年6月17日7时01分,全球累计确诊病例超过1778万例,累计死亡病例384.8万例。SARS-CoV-2是单股正链RNA病毒,颗粒呈球形或多形性,直径80~120 nm,其核酸与核衣壳蛋白紧密包装在病毒颗粒中,病毒膜表面有刺突^[2]。人类感染该病毒后常表现为上呼吸道感染、干咳、中低度发热、呼吸困难、肺炎、多脏器衰竭,严重者可致死。新冠肺炎疫情(COVID-19)在全球范围的大肆传播,迫切需要深入研究病毒传播和感染的分子机制。

跨膜丝氨酸蛋白酶2(transmembrane serine proteinase 2, TMPS2)属于II型丝氨酸蛋白酶(type II transmembrane serine protease, TTSPs)家族成员^[3]。人类TMPS2基因定位于染色体21q22.3, N端有跨膜区, C端有“催化三联体”(丝氨酸-组氨酸-天冬氨酸)结构域,此外,还包括低密度脂蛋白受体结构域(low-density lipoprotein receptor domain, LDLR)和清道夫受体结构域(scavenger receptor domain, SR)^[4]。人类TMPS2蛋白在呼吸道细胞表面大量表达,可切割ACE2蛋白(angiotensin-converting enzyme 2)和刺突蛋白(spike glycoprotein),对于冠状病毒侵染宿主细胞起着至关重要的作用^[5],是开发新冠病毒抑制剂和治疗药物的关键靶点。SARS-CoV-2的感染依赖于TMPS2的活性,高表达TMPS2的细胞在病毒侵染时会产生更高的病毒损伤^[6]。TMPS2通过切割ACE2和S蛋白,激活了S蛋白,使SARS-CoV-2更容易进入细胞。研究发现, TMPS2的关键残基(H296、S441和S460)与SARS-CoV-2 S蛋白裂解位点的侧翼残基相互作用^[1,7],因此, TMPS2在SARS-CoV-2进入宿主细胞过程中发挥重要作用。TMPS2在肺和肠道中的高表达使得这两个器官更容易受到SARS-CoV-2的侵袭^[8]。

为进一步揭示TMPS2在病毒入侵过程中的作用机制及其结构与功能特性,本文利用生物信息分析手段对TMPS2蛋白的生物学特性进行了系统分析,同时对该蛋白进行了同源性和进化分析,此外,还构建了TMPS2蛋白的原核表达载体。本研究有助于阐明TMPS2在SARS-CoV-2侵染人体细胞中的作用机制,对于研发靶向该蛋白的抑制剂和抗病毒药物具有重要意义。

1 材料与方法

1.1 材料 质粒pET-22b为本室保存, DNA分子量标准、感受态细胞及质粒DNA提取试剂购自全式

金;限制性内切酶购自NEB;氨苄西林和IPTG等常规生化试剂购自国药。

1.2 运用相关网站分析 TMPS2蛋白的理化性质、翻译后修饰位点、信号序列、分子构型及B/T细胞表位等功能特点。

1.3 多序列比对与进化分析 从UniProt网站下载12条TMPS2蛋白序列,用Clustal X 2及MEGA7.0进行序列比对及系统进化树分析。

1.4 载体构建及蛋白表达 对质粒pET-22b和TMPS2基因用XhoI与NdeI同步双酶切,回收酶切产物,将目的基因与载体进行连接,将连接产物转化Top10和BL21感受态细胞后进行PCR验证。

2 结果

2.1 TMPS2蛋白的理化性质 TMPS2由492个氨基酸组成,其中酸性氨基酸(Asp + Glu)有35个,碱性的氨基酸(Arg + Lys)有39个, Ser含量最高(0.089),其次是Gly(0.087)(Tab 1),分子量53.86 ku,理论等电点8.12,是一个碱性蛋白。该蛋白分子式为C₂₃₈₇H₃₆₅₄N₆₅₀O₇₀₉S₃₃,原子总数为7433,分子消光系数为118 145 L · mol⁻¹ · cm⁻¹,不稳定系数为41.94;在哺乳动物网织红细胞中的半衰期约为30 h,在酵母中>20 h,在大肠杆菌内>10 h,脂肪族氨基酸指数72.70,总平均亲水系数-0.248。

Tab 1 Amino acid composition of TMPS2 protein

Amino acid	Number	Frequency/%	Amino acid	Number	Frequency/%
Ala(A)	33	6.70	Leu(L)	33	6.70
Arg(R)	15	3.00	Lys(K)	24	4.90
Asn(N)	32	6.50	Met(M)	11	2.20
Asp(D)	19	3.90	Phe(F)	11	2.20
Cys(C)	22	4.50	Pro(P)	34	6.90
Gln(Q)	22	4.50	Ser(S)	44	8.90
Glu(E)	16	3.30	Thr(T)	25	5.10
Gly(G)	43	8.70	Trp(W)	15	3.00
His(H)	10	2.00	Tyr(Y)	23	4.70
Ile(I)	22	4.50	Val(V)	38	7.70

2.2 TMPS2蛋白跨膜结构预测 结果显示,该蛋白第M1-K83肽段位于细胞膜内, A84-W106肽段处于跨膜区, K107-G492肽段位于细胞膜外。跨膜螺旋残基数的期望值约21.818, N-term位于膜细胞质侧的总概率为0.948。因此, TMPS2蛋白存在1个跨膜螺旋区,属于跨膜蛋白(Fig 1)。

2.3 TMPS2蛋白卷曲螺旋预测 结果显示,基于3种不同的窗口宽度(14、21、28)没有检测到卷曲螺旋结构(Fig 2)。

2.4 TMPS2蛋白亲/疏水性分析 预测结果显

Fig 1 Transmembrane structure prediction of *TMPRSS2* protein

Fig 2 Coiled coil analysis of *TMPRSS2* protein

示 ! " # \$ % 2 蛋白的平均亲水系数为 -0.248 , 残基 E178 , K340 和 T341 的亲水性最强

Fig 5 Glycosylation sites prediction of *TMPRSS2* protein

2.7 *TMPRSS2* 蛋白的功能位点分析 *TMPRSS2* 蛋白有 3 个活性位点,分别是 H296、D345 和 S441,它们构成了一个电荷中继系统。通过查阅文献发现,目前已报道多个人源 *TMPRSS2* 蛋白的突变位点 (Tab 3)。

Tab 3 Mutation sites of *TMPRSS2* protein

Mutation type	Mutation site	Description
Point mutation	R255Q	Loss of cleavage ^[9]
Point mutation	S441A	Loss of activity ^[9]
Point mutation	A28T, G74R, V160M	Natural variant
Point mutation	S254C, E329Q, D491N	Natural variant
Point mutation	K449N	Natural variant
Point mutation	R255S, S441G	Decreased binding affinity with S protein ^[10]
Point mutation	V160M, A28T	Increased binding affinity with S protein ^[10, 11]
Nucleotide deletion	rs35074065 (del C)	Facilitates entry of 614G subtype into host cells

2.8 *TMPRSS2* 蛋白的亚细胞定位和信号序列预测 亚细胞定位分析显示,人源 *TMPRSS2* 蛋白主要表达在细胞的内质网膜 (0.444)、线粒体 (0.333)、细胞质 (0.111) 和细胞核 (0.111)。信号序列预测显示,该蛋白不含信号序列,说明它不属于分泌蛋白。

2.9 *TMPRSS2* 蛋白多序列对比 在线检索与人源 *TMPRSS2* 蛋白序列同源性较高的其它 11 种动物:大猩猩,黑猩猩,南方豚尾猕猴,东非狒狒,家猫,山羊,鸭嘴兽,袋熊,树鼩,西班牙猓獾以及中华鳖。结果显示,*TMPRSS2* 蛋白在哺乳动物间的保守性很

高,12 条序列中完全相同的残基 (* 标示) 占 0.329,性质相似的 (: 标示) 占 0.128,性质微弱相近的 (. 标示) 占 0.073。与人源 *TMPRSS2* 序列一致性最高的是黑猩猩,高达 0.980,其次是大猩猩,为 0.974;人与东非狒狒和南方豚尾猕猴的序列一致性分别为 0.886 和 0.878,与中华鳖的序列一致性最低,为 0.616 (Fig 6)。

2.10 系统进化分析 对上述“2.9”构建进化树,结果显示,黑猩猩和大猩猩与人三者聚为一支,置信度为 100;东非狒狒与南方豚尾猕猴聚为一支,置信度为 100;家猫与西班牙猓獾聚为一支,置信度为 100。山羊单独聚为一支,与人的亲缘关系最远 (Fig 7)。

2.11 *TMPRSS2* 蛋白的二级结构预测 *TMPRSS2* 蛋白含有 α-螺旋 (Hh) 80 个,占 0.1626;延长链 (Ee) 124 个,占 0.2520;β-转角 (Tt) 65 个,占 0.1321;无规卷曲 (Cc) 223 个,占 0.4533 (Fig 8)。

2.12 *TMPRSS2* 蛋白三级结构建模 利用相关构建 *TMPRSS2* 蛋白的三级结构模型,结果只显示 D144-D491 位、G103-A490 位,以及 S167-R489 位的三级结构 (Fig 9)。

2.13 B/T 细胞抗原表位预测

2.13.1 B 细胞表位分析 B 细胞表位分析有助于减少和合成多肽片段的范围与数量,提高试验效率。利用 IEDB 对该蛋白进行 B 细胞抗原表位预测,筛选阈值设定为 0.50。结果显示,可能含有 13 个 B 细胞抗原表位,分别为 S5-K80、L105-S122、T125-P154、Y161-I256、A262-L263、E299-P305、Q317-A324、H334-N343、M372-L373、E388-S394、T407-L419、Q431-S441、G462-Y469 (Fig 10)。

2.13.2 T 细胞表位分析 本文运用以下两种方法对 *TMPRSS2* 蛋白的 T 细胞表位进行分析:其中 HLA-A* 02:01 限制性 CTL 表位,阈值为 25,检测出 1 个限制性 CTL 表位;HLA-DRB1* 0401 辅助性 T 细胞表位,阈值设为 26,检测出 6 个 HLA-DRB1* 0401 辅助性 T 细胞表位。

2.14 *TMPRSS2* 的蛋白-蛋白互体分析 *TMPRSS2* 与其他 10

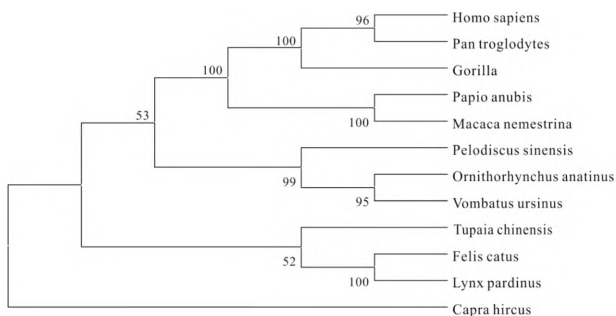


Fig 7 Phylogenetic tree based on *TMPRSS2* protein sequence

Note: The entry numbers are the same as those in Figure 6

Fig 10 Prediction of B cell antigen epitopes of *TMPRSS2* protein

Fig 8 Prediction of secondary structure of *TMPRSS2* protein

! Helix (Hh) 0. 162 6; ' Turn (Tt) 0. 132 1; Extended strand (Ee) 0. 252 0; Random coil (Cc) 0. 453 3

Fig 11 Analysis of interaction network of *TMPRSS2* protein

2. 15 pET-22b-*TMPRSS2* 原核表达载体的构建

2. 15. 1 质粒 pET-22b 与 ! " # \$ % 2 基因片段的酶切及菌落 PCR 验证 对质粒 pET-22b 单酶切(688* ") 及同步双酶切 , 反应体系为 100 -L , 电泳结果与预期相符(Fig 12A) 。 双切酶 ! " # \$ % 2 基因片段进行 68* " 和 O' 0 " 双酶切(Fig 12B) , 通过菌落 PCR 鉴定阳性克隆(Fig 12C) 。

2. 15. 2 pET-22b-! " # \$ % 2 质粒的验证及转化 用 68*M 和 O' OM 对重组质粒进行双酶切 , 电泳检测后结果符合预期(Fig 13A) 。 将重组质粒转化感受态细胞 Top10 , 涂布 LB 平板(Amp⁺) 后 37 °C 培养 12 h(Fig 13B 左) 。 将验证后的重组质粒转化感受态细胞 BL21(Fig 13B 右) , 用于表达目标蛋白。

3 讨论

SARS-CoV-2 入胞的第一步是刺突蛋白 S 与受体 ACE2 识别 , 在宿主蛋白 ! " # \$ % 2 的催化下完成

Fig 9 Schematic diagram of tertiary structure and similarity waveform of homologous proteins

Fig 12 Double enzyme digestion of pET-22b and TMPRSS2 gene fragment and colony PCR verification of positive clones

A: Electrophoretic detection of the enzyme digestion products of pET-22b. 1: No digestion, 2-3: 68 ℃ digestion, 4-5: 68 ℃ and 0 ℃ digestion, M: Trans 2k Plus DNA marker; B: Electrophoretic detection of the double enzyme digestion products of TMPRSS2 gene fragment. M: 2-log DNA marker, 1: double enzyme digestion products of TMPRSS2 gene fragment; C: Positive clones were verified by colony PCR. 1-2: positive clones.

Fig 13 Double enzyme digestion verification and transformation of recombinant plasmid

A: Verification of recombinant plasmid by double enzyme digestion. 1: The recombinant plasmid was not digested, 2: Double enzyme digestion of recombinant plasmid, M: 2-log DNA marker; B: Transformation of Top10 (left) and BL21 (right) competent cells with recombinant plasmid.

病毒粒子与宿主细胞的膜融合。研究表明, SARS-CoV-2 之所以不同于 SARS-CoV, 就在于它能更快地利用 TMPRSS2。当 SARS-CoV-2 入侵时, TMPRSS2 对 S 蛋白的 S2 亚基上进行酶切^[1], 这个切割点后露出一段疏水残基, 并迅速嵌入宿主细胞膜。随后, 伸展的 S 蛋白会折叠起来, 像拉链一样迫使病毒的外膜与细胞膜融合。

本文生信分析显示, TMPRSS2 为亲水性跨膜蛋白, 说明其可能作为膜受体或膜离子通道起作用。TMPRSS2 由 492 个氨基酸组成, 分子量 53.86 ku, 等电点 8.12, 是一个碱性蛋白, 这与其含有较多正电荷残基相契合。TMPRSS2 蛋白的不稳定性与亲水性, 为离子交换分离纯化、体外探究该蛋白的功能奠定了理论基础^[11-12]。TMPRSS2 蛋白在哺乳动物间保守性很高, 与人源 TMPRSS2 序列一致性最高的

是黑猩猩 (0.980 0), 次是大猩猩 (0.974 0), 暗示该蛋白在功能上的重要性。进化分析显示, 黑猩猩、大猩猩与人亲缘关系最近, 三者聚为一支, 置信度为 100; 山羊单独聚为一支, 与人的亲缘关系最远。

糖基化和磷酸化是两种重要的蛋白质翻译后修饰方式。TMPRSS2 是抗病毒药物设计的重要靶点蛋白, 有 3 个潜在的 N-糖基化位点和 12 个 O-糖基化位点。高度亲水的糖基, 对于蛋白质的理化性质和生理功能具有重要影响。TMPRSS2 蛋白糖基化位点的改造可进一步影响该蛋白的半衰期、靶向性、以及稳定性, 有助于研制针对 TMPRSS2 蛋白的抗病毒药物。蛋白磷酸化修饰主要参与细胞内信号转导等生物学过程, 而糖基化修饰可使不同蛋白质拥有不同的标记, 从而改变多肽的构象, 因此通过干预磷酸化和糖基化修饰位点将有望抑制病毒对宿主的侵袭^[13]。

TMPRSS2 具有 3 个功能结构域: N 端 LDL 受体 A 类结构域 (C113-C148), 其次是 SRCR (G153-V246), 最后是跨越 I256-Q487 的 C 端肽酶^[7]。当 TMPRSS2 与 ACE2 共表达时, TMPRSS2 能够切割 ACE2 并激活 S 蛋白, 促进 S 蛋白与 ACE2 相互识别从而介导病毒入侵^[14], 因此, TMPRSS2 在病毒入侵过程中发挥了“助攻”作用, TMPRSS2 与 ACE2 的高表达可提高 SARS-CoV-2 的入侵能力^[15]。研究发现^[1, 16], TMPRSS2 抑制剂可明显抑制 SARS-CoV-2 S 蛋白进入表达 TMPRSS2 的细胞系, 而促进 TMPRSS2 的表达则可取消这种抑制作用, 这就说明 SARS-CoV-2 的 S 蛋白依赖于 TMPRSS2 启动, TMPRSS2 的表达能够促进 SARS-CoV-2 的摄取^[17]。

二级结构预测发现, TMPRSS2 二级结构中占比最高的是无规则卷曲, 该结构对 TMPRSS2 的整体构象和活性有着极其重要的作用, 是构成酶活性部位和特异功能部位的重要结构^[18]。本研究还建立了 TMPRSS2 蛋白在 D144-D491、G103-A490、S167-R489 位点的三级结构模型, 为进一步设计和筛选 TMPRSS2 蛋白的潜在抑制剂和靶向药物奠定了结构基础^[3]。表位预测显示, TMPRSS2 蛋白具有 13 个 B 细胞抗原表位, 这一结果对于找到有着相似免疫保护功能及抗原特异性却无完整蛋白毒性的抗原片段, 具有很大帮助。此外, TMPRSS2 的 T 细胞表位, 这有助于细胞免疫机制研究及基因疫苗或亚单位多肽的开发。

功能位点分析显示, TMPRSS2 蛋白有多个重要位点, 例如 R255、I256 和 S441, 三者均处于预测的

Bioinformatics analysis on structure and function and expression vector construction of SARS-CoV-2 related protein TMPRSS2

XU Ben-jin¹, LI Zhuo-xi², FAN Lei², LI Jing², YAN Rong-rong³, DU Miao¹, XUAN Yan¹, MEN Jie², CHEN Xiao-cong², TANG Wen-ting², HOU Yan-xiang¹, SONG Bin-yu⁴, YANG Ya-nan¹, LIU Xiao-liang⁴, YU Jun-xiao², LIU Ling¹

(1J @*E5 06 " *8(/, 1 ; , <0-, 50-C , 2J @*E5 06 3, +(/ " *8(/()* , 3J ; , <0-, 50-C @*E5 06 %' ,)F(B*)C,) 2 D0+E(5, 1 , 4J %/(*)/* ,)8 !*/')0102C &*)5-* , B*)C,)2 &011*2* 06 %' ,)F(" *8(/, 1 9)(?*-(5C , B*)C,)2 , %' ,)F(032200 , &' () ,)

Abstract: Aim Human TMPRSS2 is a transmembrane serine protease. In this paper , the structure and function of the protein were systematically analyzed by bioinformatics , the codon was optimized and the prokaryotic expression vector was constructed to explore the molecular mechanism of SARS-CoV-2 infecting host cells. **Methods** The recombinant expression vector pET-22b-TMPRSS2 was generated by molecular cloning technology. The homology , functional sites , sub-cellular localization , three-dimensional structure and evolutionary characteristics of TMPRSS2 protein were systematically analyzed by using analytical tools such as Protparam , NetPhos3. 1 , Blast , Clustal X2 and MEGA7.0. **Results** The prokaryotic expression plasmid was constructed correctly; TMPRSS2 belongs to medium molecular weight protein , which is composed of 492 amino acid residues. The theoretical isoelectric point is 8. 12 , the molecular extinction coefficient is 118 145 L • mol⁻¹ • cm⁻¹ , and the half-life is 30 h; TMPRSS2 has 15 potential glycosylation sites and 49

possible phosphorylation sites. It is a transmembrane hydrophilic protein without signal sequence. In addition , the protein has 13 potential B-cell epitopes and 7 T-cell epitopes. Secondary structure analysis showed that random coil accounted for the highest proportion of TMPRSS2 protein (0. 453 3) , followed by extended strand (0. 252 0) . Sequence comparison and evolutionary analysis showed that the highest sequence consistency and closest genetic relationship with human TMPRSS2 was Pan troglodytes , followed by gorilla. **Conclusions** Human-derived TMPRSS2 protein is evolutionarily conserved and functionally important. The results of this study can help to reveal the structure and mechanism of action of TMPRSS2 protein , provide ideas for the diagnosis and treatment of COVID-19 , and accelerate the research and development process of new drugs targeting TMPRSS2 protein. **Key words:** SARS-CoV-2; ! " # \$ % & ; bioinformatics analysis; vector construction; structure and function; evolutionary analysis



欢迎投稿 , 欢迎订阅!