

冠状病毒S蛋白N-端结构域在病毒入侵过程中的作用机制

王嘉君¹ 关洪鑫¹ 欧阳松应^{1,2*}

(¹福建师范大学生命科学学院, 福州 350117; ²福建师范大学南方生物医学研究中心, 福州 350117)

摘要 冠状病毒(coronavirus)是单股正链RNA病毒, 可以引起包括人类在内的多种动物的呼吸道、胃肠道和中枢神经系统疾病。病毒刺突蛋白(spike protein, S蛋白)的S1亚基的N-端结构域(N-terminal domain, NTD)和C-端结构域(C-terminal domain, CTD)都可以作为受体结合域(receptor-binding domain, RBD), 且是病毒入侵宿主细胞的关键因素。一般认为, 在病毒入侵过程中, S1-NTD主要通过识别并结合糖类受体(attachment receptors)来辅助S1-CTD特异性识别蛋白质受体[小鼠肝炎病毒(mouse hepatitis virus, MHV)除外]。然而, 随着对新冠病毒的深入研究, 发现S1-NTD也可以识别多种蛋白受体, 其作用机理与特点也逐渐被揭示。该文综述了冠状病毒的S1-NTD与受体识别的结构基础, 总结了冠状病毒S1-NTD的进化过程, 有利于深入理解冠状病毒入侵宿主细胞机制和病毒跨物种传播机制, 并为基于NTD的药物及疫苗的开发提供参考。

关键词 冠状病毒; 刺突蛋白; S1-NTD; 蛋白质受体; 糖类受体

The Mechanism of N-Terminal Domain of Coronavirus S Protein in the Process of Virus Invasion

WANG Jiajun¹, GUAN Hongxin¹, OUYANG Songying^{1,2*}

(¹College of Life Sciences, Fujian Normal University, Fuzhou 350117, China;

²Biomedical Research Center of South China, Fujian Normal University, Fuzhou 350117, China)

Abstract Coronavirus is a single-stranded plus strand RNA virus, which can cause respiratory tract, gastrointestinal tract and central nervous system diseases in many animals, including humans. Both NTD (N-terminal domain) and CTD (C-terminal domain) of the S1 subunit of viral S protein (spike protein) can act as RBD (receptor-binding domain), which is the key factor of virus invasion into host cells. Previous studies generally believe that S1-NTD mainly recognizes and binds carbohydrate receptors (also known as attachment receptors) to assist S1-CTD in specific recognition of protein receptors (except mouse hepatitis virus) for virus invasion. With the in-depth study of novel coronavirus, it is found that S1-NTD can also recognize a variety of protein receptors, and its mechanism, characteristics and rules of action are gradually revealed. This article summarizes the structural basis of coronavirus S1-NTD and receptor recognition, and systematically summarizes the evolution process of coronavirus S1-NTD, which is conducive to comprehensive understanding of the mechanism of coronavirus invading host cells

收稿日期: 2022-01-14 接受日期: 2022-02-28

科技部国家重点研发项目(批准号: 2021YFC2301403)、国家自然科学基金(批准号: 82172287)、福建省科技厅项目(批准号: 2020Y4007)和福建省海洋经济发展补助项目(批准号: FJHJF-L-2020-2)资助的课题

*通讯作者。Tel: 0591-22868199, E-mail: ouyangsy@fjnu.edu.cn

Received: January 14, 2022 Accepted: February 28, 2022

This work was supported by the National Key Research and Development Project of Ministry of Science and Technology (Grant No.2021YFC2301403), the National Natural Science Foundation of China (Grant No.82172287), the Fujian Science and Technology Department Project (Grant No.2020Y4007), and the Marine Economic Development Subsidy Project of Fujian Province (Grant No.FJHJF-L-2020-2)

*Corresponding author. Tel: +86-591-22868199, E-mail: ouyangsy@fjnu.edu.cn

and cross-species transmission. At the same time, this paper will provide reference for the development of NTD-based drugs and vaccines.

Keywords coronavirus; spike protein; S1-NTD; protein receptor; sugar receptor

在过去的20年里,新型冠状病毒(coronavirus, CoV)的不断涌现,对人类和其他动物的健康构成了严重威胁。2002—2003年间,严重急性呼吸综合征冠状病毒(severe acute respiratory syndrome coronavirus, SARS-CoV)感染了8 000多人,死亡率为9.6%^[1]。2012年,中东呼吸综合征冠状病毒(middle east respiratory syndrome coronavirus, MERS-CoV)确诊病例2 229例,死亡率高达35.5%^[2]。2013年,猪流行性腹泻冠状病毒(porcine epidemic diarrhea virus, PEDV)席卷美国^[3],造成仔猪近100%的死亡率^[4]。2019年,严重急性呼吸综合征冠状病毒2(severe acute respiratory syndrome coronavirus 2, SARS-CoV-2)迅速蔓延全球^[5],根据世界卫生组织数据统计,截至2021年11月2日,感染冠状病毒的病例数为2.47亿,确认死亡人数为500万。随着时间推移,SARS-CoV-2发生了进化和基因突变,出现了数千种变异株。世界卫生组织将冠状病毒变异株分为两大类:值得关注的变异株(variants of concern, VOC)毒株[包括Alpha(B.1.1.7)、Beta(B.1.351)、Gamma(P.1)、Delta(B.1.617.2)和Omicron(B.1.1.529)毒株]和待观察突变株(variant of interest, VOI)毒株[包括Epsilon(B.1.427和B.1.429)、Lambda(C.37)毒株和Mu(B.1.621)毒株等],其中Lambda毒株和Delta毒株危险性极强,在毒性、传播力以及抵御中和抗体的能力等方面显著增强,Omicron毒株具有极强传染性和免疫逃逸机制。冠状病毒的爆发是对世界公共卫生的重大威胁,严重影响人类的生命安全和可持续的经济发展。因此,了解和控制冠状病毒的传播对全球人类健康和经济稳定具有重要意义。

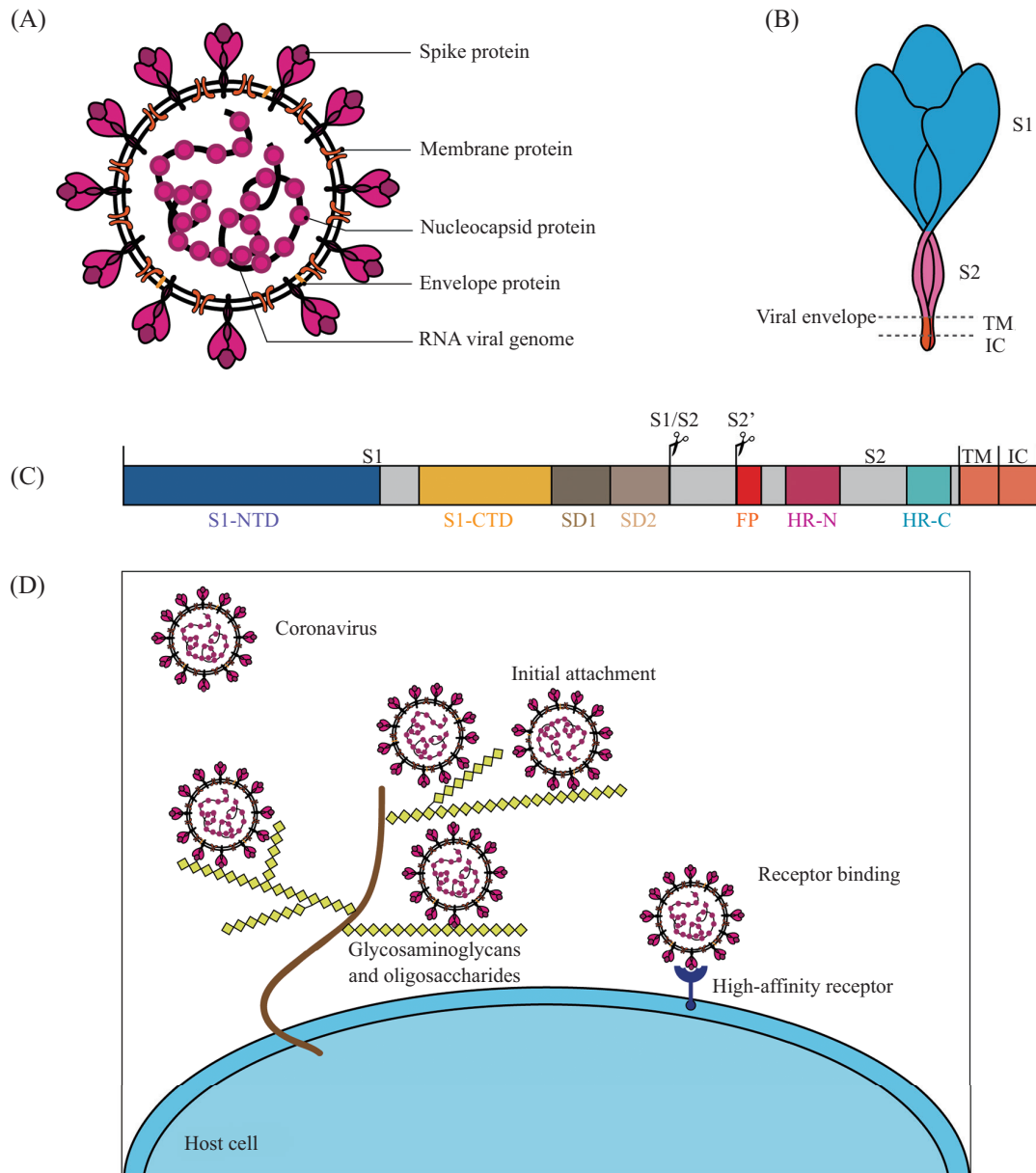
冠状病毒属于巢病毒目(Nidovirales)冠状病毒科(Coronaviridae)正冠状病毒亚科(Orthocoronavirinae),病毒粒子外表面为球形(图1A),基因组是单股正链RNA^[6]。在所有RNA病毒中,冠状病毒的基因组最大,其基因组大小为26~32 Kb^[5]。根据基因组和系统发育分析,冠状病毒可分为4个属:α属冠状病毒、β属冠状病毒、γ属冠状病毒和δ属冠状病毒。通常,α和β属冠状病毒感染哺乳动物,γ属冠状病毒感染禽类,δ属冠状病毒可以同时感染哺乳动物和禽类。病

毒感染主要会引起宿主呼吸系统、肝脏、肠道和神经系统的急性或慢性症状。冠状病毒至少有四种主要的结构蛋白:刺突蛋白(spike protein, S蛋白)、包膜蛋白(envelope protein, E蛋白)、膜蛋白(membrane protein, M蛋白)和核蛋白(nucleocapsid protein, N蛋白)(图1A)。除了这四种主要的结构蛋白之外,一些冠状病毒还编码一种特殊的结构蛋白,即包膜相关的血凝素-酯酶(hemagglutinin-esterase, HE)蛋白。这些结构蛋白在病毒组装和感染过程中起着至关重要的作用,其中S蛋白通过其N-端结构域(N-terminal domain, NTD)和C-端结构域(C-terminal domain, CTD)识别并结合宿主细胞表面受体介导病毒入侵,是决定病毒宿主范围和病毒嗜组织性的关键因素,也是宿主免疫反应的主要诱因^[7]。

通常来说,病毒最初接触宿主细胞是通过S1-NTD(图1B和图1C)与宿主细胞表面的聚糖[包括糖胺聚糖(glycosaminoglycan, GAG)和含唾液酸的低聚糖等]非特异性相互作用^[8],从而将冠状病毒富集到细胞表面的,这有助于病毒S1-NTD/CTD与高亲和力的特异性蛋白受体结合,促进病毒进入宿主细胞(图1D)。目前,该领域已有研究阐明了S1-CTD在病毒入侵过程中的作用机制^[9],但关于S1-NTD识别受体机制的研究较少。本文将从冠状病毒S蛋白结构特征、S1-NTD结构特点及其与受体相互作用机理等方面对S1-NTD在病毒入侵宿主细胞过程中的作用机制进行系统阐述,本文总结的内容将为进一步研究冠状病毒入侵机制提供有力的参考,有助于人们深入理解冠状病毒的发病机制,为控制病毒的跨物种传播以及为未来可能发生的人畜共患病疫情做好准备^[10-11]。

1 冠状病毒S蛋白的结构特征

冠状病毒S蛋白由1 100~1 600个氨基酸残基组成,含有21~35个N-糖基化位点,是高度糖基化的糖蛋白。冠状病毒S蛋白由3个部分组成:1个大的膜外结构域、1个单向跨膜域和1个短的细胞内尾巴区^[12](图1B)。S蛋白作为I类病毒融合蛋白,类似于丝状病毒、副黏病毒、流感病毒和逆转录病毒的融合蛋白,在翻译后折叠成亚稳态融合前构象。冠状病毒S蛋白以三



A: 冠状病毒结构示意图; B: 刺突蛋白结构示意图; C: 刺突蛋白的组成; D: 冠状病毒入侵宿主细胞的可能机制。TM: 跨膜域; IC: 胞内结构; SD1: 子结构域1; SD2: 子结构域2; FP: 融合肽; HR: 七肽重复序列; S1/S2: S1/S2酶切位点; S2': S2'酶切位点。

A: a schematic diagram of the structure of the coronavirus; B: a schematic diagram of the structure of the spike protein; C: the composition of spike protein; D: possible mechanisms of coronavirus infection of host cells. TM: transmembrane domain; IC: intracellular structure; SD1: subdomain 1; SD2: subdomain 2; FP: fusion peptide; HR: heptad repeat; S1/S2: S1/S2 cleavage sites; S2': S2' cleavage sites.

图1 冠状病毒和刺突蛋白的组成以及冠状病毒入侵宿主细胞的过程

Fig.1 Composition of coronavirus and spike protein and the process of coronavirus infection

聚体形式分布于病毒表面, 每个单体的膜外结构域又由1个负责与宿主细胞受体结合的N-端S1亚单位和驱动膜融合的C-端S2亚单位组成。冠状病毒在入侵宿主细胞过程中, 利用宿主蛋白酶(furin样蛋白酶、丝氨酸蛋白酶和胰蛋白酶等)在S蛋白的某个或某几个位置[例如: S1/S2或S2'位点(图1C)]进行裂解^[13], 以达到促进病毒与宿主细胞膜融合的目的, 但不同类型冠

状病毒的S蛋白是否被切割, 以及其被切割的时间和方式是不同的。S1亚基主要包括2个重要结构域NTD和CTD, 这2个结构域都可以作为RBD发挥作用^[9], S1-CTD主要负责识别蛋白质受体(表1), 如血管紧张素转换酶2(angiotensin-converting enzyme 2, ACE2)^[14]、氨基肽酶(amino peptidase N, APN)^[15]和丝氨酸蛋白酶(dipeptidyl peptidase 4, DPP4)^[16]等。之前的研究发现, S1-

表1 冠状病毒S蛋白目前已知的受体和受体结合域总结

Table 1 Overview of currently known receptors for coronavirus S protein and their binding domains

属	物种	S1-NTD的受体	S1-CTD的受体	参考文献
Genus	Species	Receptors of S1-NTD	Receptors of S1-CTD	References
α -CoV (α -coronavirus)	HCoV-229E	—	APN	[20]
	HCoV-NL63	—	ACE2	[21]
	PEDV	Neu5Ac	—	[22]
	PRCV	—	APN	[23]
	TGEV	Neu5Ac, Neu5Gc	APN	[24]
β -CoV (β -coronavirus)	MHV	CEACAM1	—	[17]
	BCoV	Neu5,9Ac ₂	HLA-1*	[25-26]
	OC43	Neu5,9Ac ₂	—	[27]
	HKU1	Neu5,9Ac ₂	HLA-1*	[25,28]
	PHEV	Neu5,9Ac ₂	—	[26]
	SARS-CoV	—	ACE2	[29]
	SARS-CoV-2	AXL, LDLRAD3, TMEM30A, CLEC4M, sialic acid	ACE2	[14,18-19]
γ -CoV (γ -coronavirus)	MERS-CoV	Sialic acid	DPP4	[30-31]
	IBV	Neu5Gc	—	[27,32]
δ -CoV (δ -coronavirus)	PDCV	—	APN	[33]

*: 有待验证。—: 受体未知。

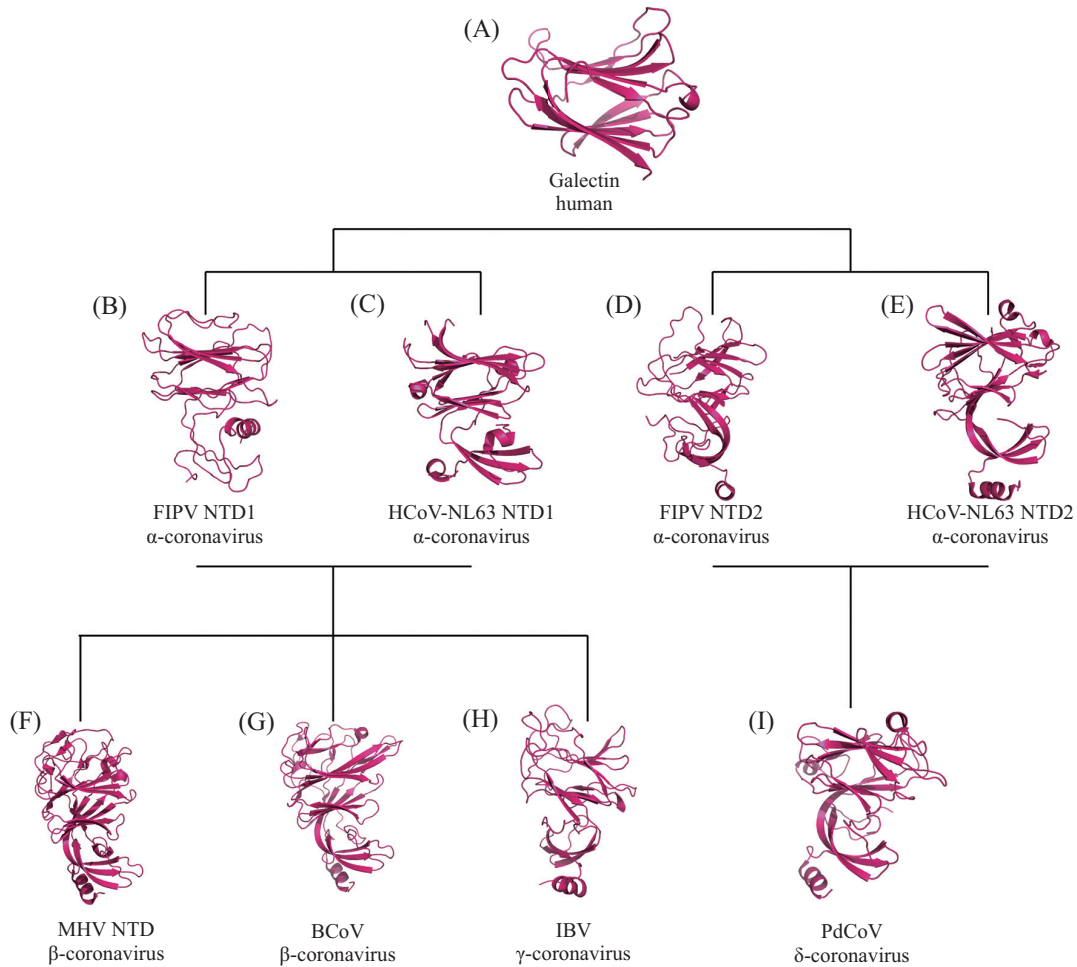
*: need further examining. —: receptor is unknown.

NTD主要结合糖类受体(表1), 如唾液酸(sialic acid)等, 小鼠肝炎病毒(mouse hepatitis virus, MHV)比较特殊, 它通过S1-NTD识别蛋白受体癌胚抗原相关细胞黏附分子1(carcinoembryonic antigen-related cell adhesion molecule 1, CEACAM1)^[17]。最新研究表明, SARS-CoV-2的NTD可以识别蛋白受体(表1), 如受体酪氨酸激酶AXL(也称UFO)、低密度脂蛋白受体结构域蛋白3(low-density lipoprotein receptor class A domain-containing 3, LDLRAD3)、P4型ATP酶的 β 亚基TMEM30A和C型凝集素CLEC4M, 并促进病毒的入侵^[18-19]。总之, 冠状病毒已经进化出了一套复杂的受体识别模式: (1) 冠状病毒可以利用S蛋白的1个或2个不同的结构域识别受体; (2) 同一属的冠状病毒S蛋白可以识别不同的受体, 而不同属冠状病毒也可识别相同受体; (3) 冠状病毒S1-NTD既可以识别蛋白质受体也可以识别糖受体^[9]。

2 NTD结构域

2021年微软亚洲研究院和清华大学的科研人员^[34]通过分子动力学模拟实验提出S1-NTD利用“楔形”模型在病毒侵染中发挥调控作用: 当S1-NTD向S蛋白的中心轴运动时, S1-NTD会像“楔子”一样插入到与其相连的RBD和S蛋白中心轴之

间的下方区域中, 从而阻碍RBD向下运动, 维持RBD向上的状态, 保证RBD与ACE2结合。当S1-NTD向外运动时, 向上构象的RBD转变为向下的构象, S蛋白结构为闭合状态, 冠状病毒完成了S蛋白从部分开放构象到闭合构象的转变。这一过程证实了NTD在病毒感染中参与S蛋白的激活。大多数已知的冠状病毒S蛋白包含1个NTD, 但某些 α 属冠状病毒S蛋白包含2个NTD亚型^[35]: 亚型I(S1-NTD1)和亚型II(S1-NTD2)(图2)^[36-37]。这两种亚型的NTD在特征性二硫键的分布上具有明显的差异, 此类二硫键在I型NTD中保守, 而在II型NTD中并不保守。目前有两种说法来解释这2个NTD的存在: (1) WALLS团队^[35]通过DALI搜索发现HCoV-NL63(human coronavirus NL63)的2个S1-NTD的Z值和RMSD(root-mean-square deviation)值分别为6.7和3.8, 因此推断HCoV-NL63中的2个NTD是进化过程中基因复制的结果; (2) 其他一些研究人员则认为这2个NTD是由病毒基因组中2个独立的原始域重组而来的^[35,38-39]。从结构上看, 这两种亚型的S1-NTD都包含1个与人类半乳糖凝集素结构折叠类似的核心结构域(图2A)^[40]。根据NTD结构比对分析, α 属冠状病毒的NTDs(I型NTD和II型NTD)和 δ 属冠状病毒具有最简单的折叠, 其核心结构顶



A: 人半乳糖凝集素结构图(PDB: 1A3K); B: α 属 FIPV NTD1 结构图(PDB: 6JX7); C: α 属 HCoV-NL63 NTD1 结构图(PDB: 5SZS); D: α 属 FIPV NTD2 结构图(PDB: 6JX7); E: α 属 HCoV-NL63 NTD2 结构图(PDB: 5SZS); F: β 属 MHV NTD 结构图(PDB: 3R4D); G: β 属 BCoV 结构图(PDB: 4H14); H: β 属 IBV 结构图(PDB: 6CV0); I: δ 属 PdCoV NTD 结构图(PDB: 6BFU)。

A: structure diagram of human galactose agglutinin (PDB: 1A3K); B: structure diagram of FIPV NTD1 (α -CoV) (PDB: 6JX7); C: structure diagram of HCoV-NL63 NTD1 (α -CoV) (PDB: 5SZS); D: structure diagram of FIPV NTD2 (α -CoV) (PDB: 6JX7); E: structure diagram of HCoV-NL63 NTD2 (α -CoV) (PDB: 5SZS); F: structure diagram of MHV NTD (β -CoV) (PDB: 3R4D); G: structure diagram of BCoV (β -CoV) (PDB: 4H14); H: structure diagram of IBV (β -CoV) (PDB: 6CV0); I: structure diagram of PdCoV NTD (δ -CoV) (PDB: 6BFU)。

图2 人的半乳糖凝集素和不同属冠状病毒的S1-NTD的结构示意图

Fig.2 Structure diagrams of human galectin and S1-NTD of different genera of coronavirus

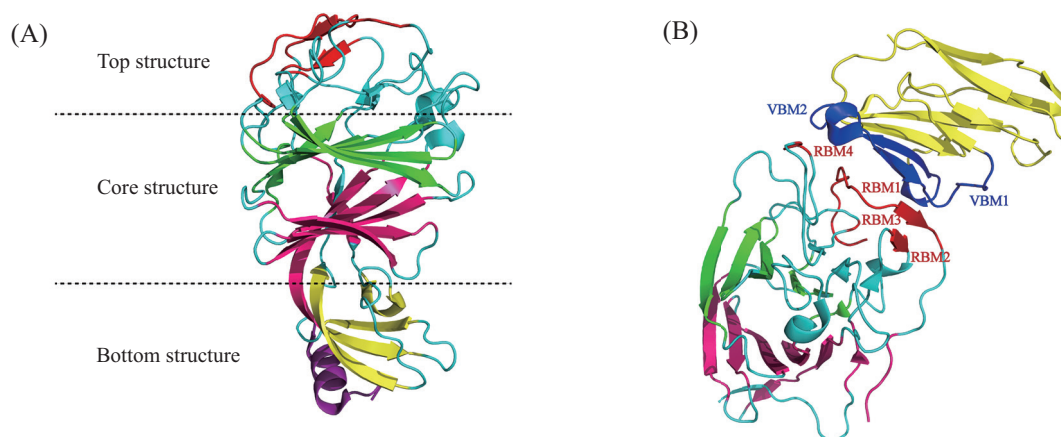
部几乎无其他元件(图2D、图2E和图2I), 而 γ 属冠状病毒S1-NTD核心结构域顶部则由少量的短 loop 等元件组成(图2B、图2C和图2H)。 β 属冠状病毒S1-NTD通过进化出更加复杂的顶部结构以实现识别糖类和蛋白质受体的功能(图2F和图2G)。下文将从NTD结合不同受体的角度分析比较NTD的结构特点。

3 冠状病毒S1-NTD识别细胞表面膜蛋白受体的机制

冠状病毒S1-NTD与宿主细胞表面蛋白受体的结合在很大程度上决定了宿主细胞的易感染性、宿

主趋向性以及病毒跨物种传播的能力, 也是成功启动感染的前提和关键步骤。目前, MHV是已知的第一个利用S1-NTD识别蛋白受体的冠状病毒。

β 属冠状病毒MHV的S1-NTD与受体CEACAM1a的复合物结构首次提供了冠状病毒S1-NTD与蛋白受体相互作用的细节。MHV S1-NTD由1个核心结构域和其顶部结构以及底部外围元件组成^[41](图3A), 该核心结构域包含2个反向平行的 β -sheet层, 其中一个 β -sheet层有6条 β 链, 另一个 β -sheet层有7条 β 链。MHV S1-NTD核心结构域与人半乳糖结合凝集素(半乳糖结合凝集素)有相同的结构折叠, 但MHV S1-NTD在核心结构顶部包含额外的结



A: MHV NTD的整体结构(PDB: 3R4D); 该结构包括顶部结构、核心结构和底部结构; NTD核心的两片 β 片分别为绿色和洋红色, NTD的其他部分为青色, RBM为红色; B: MHV NTD/mCEACAM1a接口结构细节。RBM是红色的, VBM是蓝色的。

A: overall structure of MHV NTD (PDB: 3R4D). The structure includes top structure, core structure and bottom structure. Two β -sheets of NTD core are colored green and hotpink, respectively. Other parts of NTD are colored cyan, and RBM is red; B: structural details of MHV NTD/mCEACAM1a interface. RBM is red, and VBM is blue.

图3 MHV NTD的晶体结构以及与CEACAM1a形成的复合物的晶体结构

Fig.3 The crystal structures of MHV NTD and its complex with CEACAM1a

构, 其顶部 β -sheet层与周围3个loop结构在N末端汇聚, 形成独特的结构, 这4个结构元件也被称为受体结合基序(receptor binding motif, RBM), 它可以与CEACAM1a的D1结构域上的病毒结合基序(virus-binding motif, VBM)相互识别并结合(图3B)。小鼠、牛和人类的CEACAM1a分子在几个关键的VBM残基上也有所不同, 这也是MHV不能跨物种传播的主要原因^[41]。

除了MHV的S1-NTD可以识别蛋白质受体之外, 2021年, WEI等^[18]基于CRRISPR全基因组敲除技术发现膜蛋白LDLRAD3、TMEM30A、CLEC4G和AXL可以通过与新型冠状病毒S1-NTD特异性结合, 以不依赖于ACE2的方式, 有效介导病毒入侵宿主细胞。ACE2作为介导SARS-CoV-2入侵宿主细胞的高亲和力受体主要通过识别S1-CTD发挥作用, 并在肾脏和消化系统中高表达, 而在人体其他组织如大脑和呼吸道以及多种免疫细胞中的表达极低甚至不表达。研究表明, LDLRAD3在神经元中高度表达, CLEC4G在肝脏、淋巴结和单核细胞中高度表达, AXL在几乎所有人体器官中广泛表达, 特别是在肺和支气管上皮组织和细胞中, AXL的表达远远高于ACE2的表达。进一步证实了SARS-CoV-2 S蛋白的CTD和NTD结构域通过与组织特异性受体结合, 从而导致SARS-CoV-2的多器官嗜性并提高其在不同器官中的感染效率, 推动新冠病毒在人群中快速传播。这些具有组织特异

性的受体为理解SARS-CoV-2的器官倾向性提供了更多的线索, 也为寻找冠状病毒受体和研究入侵过程提供新的思路 and 方向。目前缺乏SARS-CoV-2 S1-NTD与其蛋白质受体的复合物结构数据, 因此无法阐述NTD与蛋白质受体的分子作用机制。但以往的文章有报道过NTD与糖类受体的复合物结构, 阐述了NTD通过结合糖类受体辅助病毒富集在宿主细胞表面。

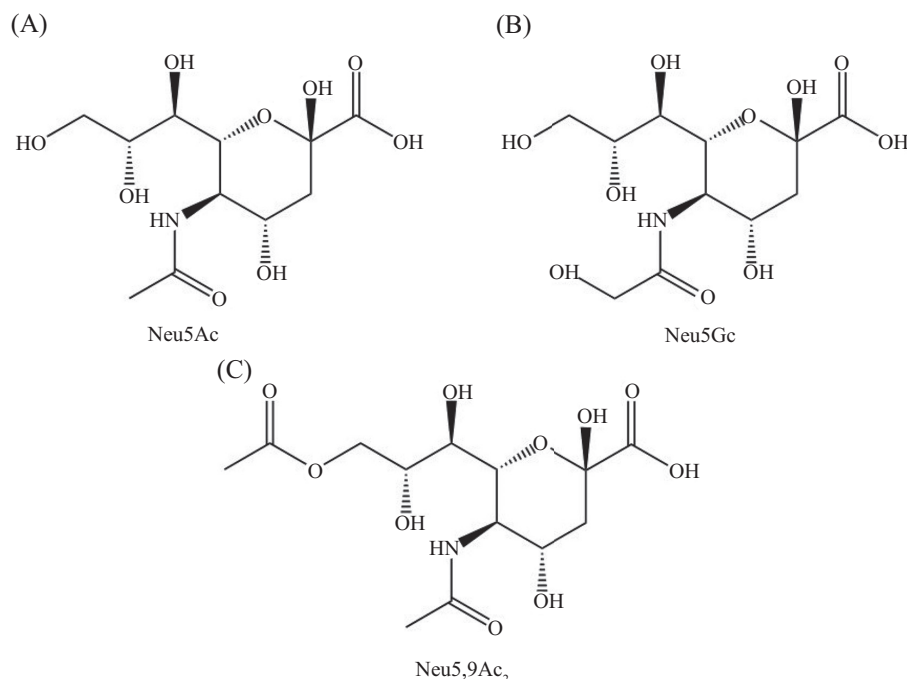
4 冠状病毒S1-NTD识别糖受体的机制

许多冠状病毒S蛋白S1-NTD除了与蛋白质受体结合外, 还能与糖类受体结合。目前, 已经确定的糖类受体主要是唾液酸及唾液酸苷。唾液酸是由神经氨酸衍生的具有9个碳骨架的单糖, 通常以 α -糖苷键的形式连接到糖蛋白、糖脂和寡糖的末端^[25]。唾液酸具有多样性和复杂性, 这种现象主要是由N-乙酰神经氨酸和N-羟神经氨酸的修饰以及唾液酸和寡糖链之间不同的糖苷连接形式造成的。唾液酸经常被各种包膜病毒(如冠状病毒和副流感病毒等)和非包膜病毒(如轮状病毒和腺病毒等)用于附着和入侵宿主细胞。据报道, 有三种可充当冠状病毒受体或受体结合辅助因子的唾液酸(图4), 即N-乙酰神经氨酸(N-acetylneuraminic acid, Neu5Ac)^[26]、N-羟乙酰神经氨酸(N-glycolylneuraminic acid, Neu5Gc)^[26]和N-乙酰基-9-O-乙酰神经氨酸(9-O-acetyl-N-acetylneuraminic acid, Neu5,9Ac₂)^[26]。研究表明唾液酸分

子与其他糖分子的连接方式是决定病毒的宿主嗜性的主要因素^[42],例如人类呼吸道细胞上广泛分布着 α -2,6-连接形式的唾液酸,而禽类消化道细胞上则主要分布着 α -2,3-连接形式的唾液酸,因此,偏好 α -2,6-连接形式唾液酸的人类甲型流感病毒可感染人类,而偏好 α -2,3-连接形式唾液酸的甲型禽流感病毒则将禽类作为宿主^[43]。冠状病毒中,BCoV(bovine coronavirus)、HCoV-OC43(human coronavirus OC43)、PHEV(porcine hemagglutinating encephalomyelitis virus)、HCoV-HKU1(human coronavirus HKU1)和HCoV-HKU23(human coronavirus HKU23)等都可以通过S1-NTD与宿主细胞表面的Neu5,9Ac₂结合,进而辅助其S1-CTD与蛋白受体的后续识别。S1-NTD具有1个由核心结构和顶部结构包围而成的空腔,该空腔包含了1个糖基结合口袋用于结合糖受体。糖基阵列筛选技术往往被用于鉴定不同冠状病毒的糖受体,例如研究发现Neu5,9Ac₂是BCoV NTD的首选糖底物^[44]。

尽管NTD具有保守的人半乳糖结合凝集素样折叠,但是有些可以单一结合蛋白受体或糖类受体,还有些既可以结合糖基受体又可以结合蛋白受体。例如BCoV的NTD与MHV的NTD结构高度相

似,其RMSD值为0.585,前者结合Neu5,9Ac₂,而后者则结合MCEACAM1。结构生物学及突变实验表明,BCoV NTD的Y162、E182、W184和H185(图5A)是负责糖结合的关键四残基基序(tetra-residue motif, TRM)^[45]。TRM分布在S1-NTD的2个糖结合环上: Tyr162位于loop11-12(连接 β 11、 β 12)上,而Glu182、Trp184和His185在loop12-13(连接 β 12、 β 13)上。BCoV和MHV NTD的结构比较可以解释两者为什么可以识别不同类型的受体。首先, MHV NTD的RBM4(loop12-13)作为MCEACAM1的关键结合位点(图5B),与BCoV NTD相比具有明显的构象差异,并导致两者结合CEACAM1a的能力不同(图5C);其次, MHV S1-NTD中loop10-11明显缩短,而BCoV NTD loop10-11与其他loop在核心结构区上有广泛的接触,BCoV NTD通过这些loop间的相互作用力稳定糖结合口袋的空间结构从而间接促进糖基结合。最后, MHV NTD的182位氨基酸是甘氨酸(Gly182),相比于BCoV NTD的182位的谷氨酸,甘氨酸没有侧链,从而导致MHV NTD和糖基之间的相互作用减弱。因此,糖结合环的构象改变和关键的糖结合氨基酸的改变都将导致NTD获得或丧失糖结合能力。2019年



A: Neu5Ac的分子结构; B: Neu5Gc的分子结构; C: Neu5,9Ac₂的分子结构。

A: the molecular structure of Neu5Ac; B: the molecular structure of Neu5Gc; C: the molecular structure of Neu5,9Ac₂.

图4 细胞外表面存在的三种唾液酸的分子结构

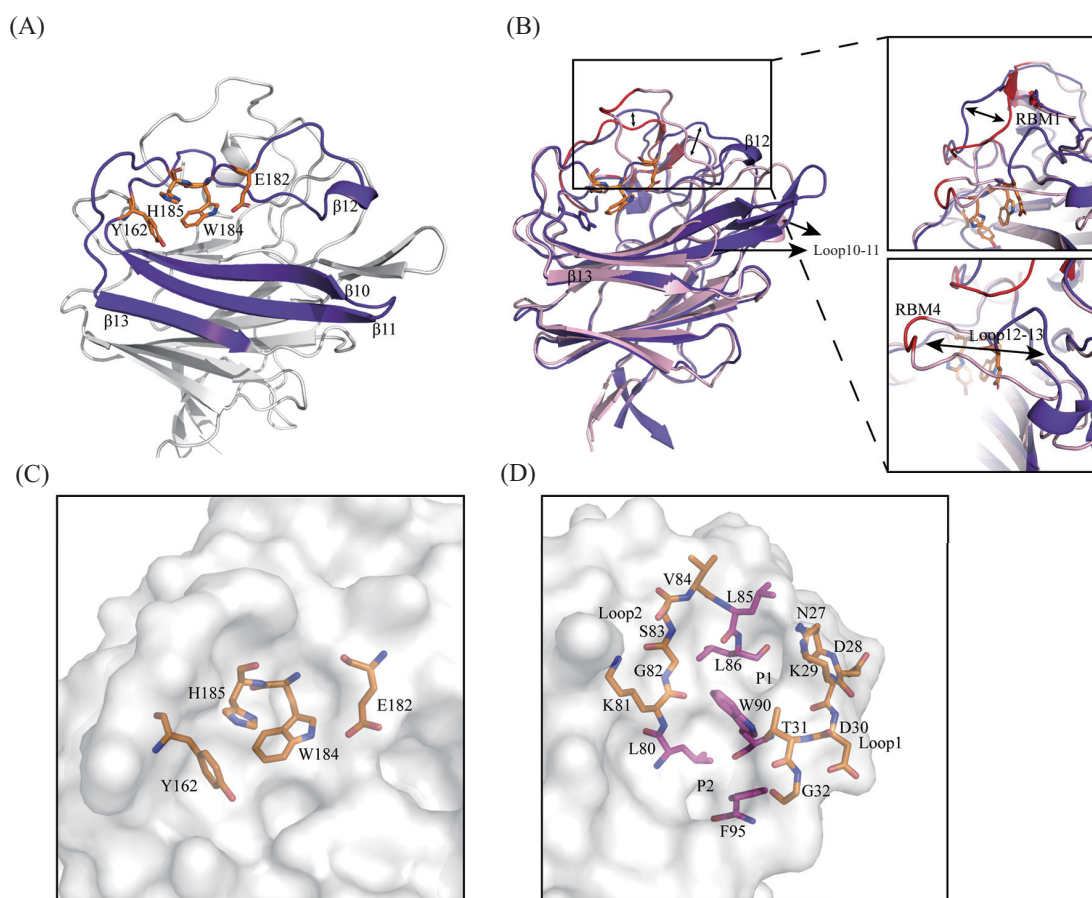
Fig.4 The molecular structures of three sialic acids present on the extracellular surface

底爆发的 SARS-CoV-2 的 S1-NTD 具有结合糖类受体和蛋白质受体的能力。2020 年 FANTINI 等^[46]发现 SARS-CoV-2 S1-NTD 可以与含有唾液酸的神经节苷脂结合, 2021 年 WEI 等^[18]发现了 SARS-CoV-2 S1-NTD 的蛋白质受体: LDLRAD3、TMEM30A、CLEC4G 和 AXL。

据报道, 部分冠状病毒 (BCoV、PHEV、OC43 和 HKU23 等) 的 NTD 有 2 个口袋具有结合糖受体能力。2012 年 PENG 等^[45]发现了第一个糖结合口袋 (口袋 A, pocket A), 其位于 NTD 顶部结构和核心结构之间, 如上面提到的 BCoV 糖结合口袋。2019 年 VEESLER 等^[47]利用冷冻电镜技术在 NTD 顶部发现了第二个糖结合口袋 (口袋 B, pocket B)。HULSWIT

等^[26]认为口袋 B 由 2 个与唾液酸相互作用的疏水口袋 (pocket 1, P1; pocket 2, P2) 组成。以 OC43 为例 (图 5D), P1 口袋由 L85、L86 和 W90 残基组成, P2 口袋由 L80、W90 和 F95 残基组成^[26]。P1 和 P2 口袋由 W90 吡啶侧链分隔, 并由形成结合位点边缘的 2 个环 [loop1 (N27-D28-K29-D30-T31-G32)] 和 [loop2 (L80-K81-G82-S83-V84-L85-L86)] 组成。研究表明, 任何一个口袋中的残基突变都会影响 NTD 与糖受体结合^[26]。与 A 口袋相比, B 口袋明显体积较大, 而 B 口袋是否可以提高糖结合效率以及容纳不同种类的糖还有待研究。

然而, 糖结合口袋 A 和 B 在 MERS-CoV、IBV、PEDV 和 TGEV 等冠状病毒的 S1-NTD 结构域中都不保守, 这导致口袋构象发生改变影响 S1-NTD 结构域



A: BCoV 的关键糖结合残基; 残基标记为黄色杆状 (PDB: 4H14); B: BCoV 和 MHV NTD 叠加结构图像; BCoV NTD 标记为蓝色, MHV NTD 标记为粉色。MHV NTD 中的 RBM1 和 RBM4 标记为红色。双向箭头表示 2 个 NTD 中受体结合环的不同构象; C: BCoV NTD 的糖结合口袋 A (pocket A); D: OC43 NTD 的糖结合口袋 B (pocket B), 包括 P1 和 P2 (PDB: 6NZK)。

A: the key sugar binding residue of BCoV. The residue is yellow rod (PDB: 4H14); B: image of the superimposed structures of BCoV and MHV NTDs. BCoV NTD is colored blue, and MHV NTD is colored pink. RBM1 and RBM4 are marked in red in MHV NTD. Bidirectional arrows indicate different conformations of the receptor-binding loops in the two NTDs; C: the sugar-binding pocket A of BCoV NTD; D: the sugar-binding pocket B of OC43 NTD includes P1 and P2 (PDB: 6NZK).

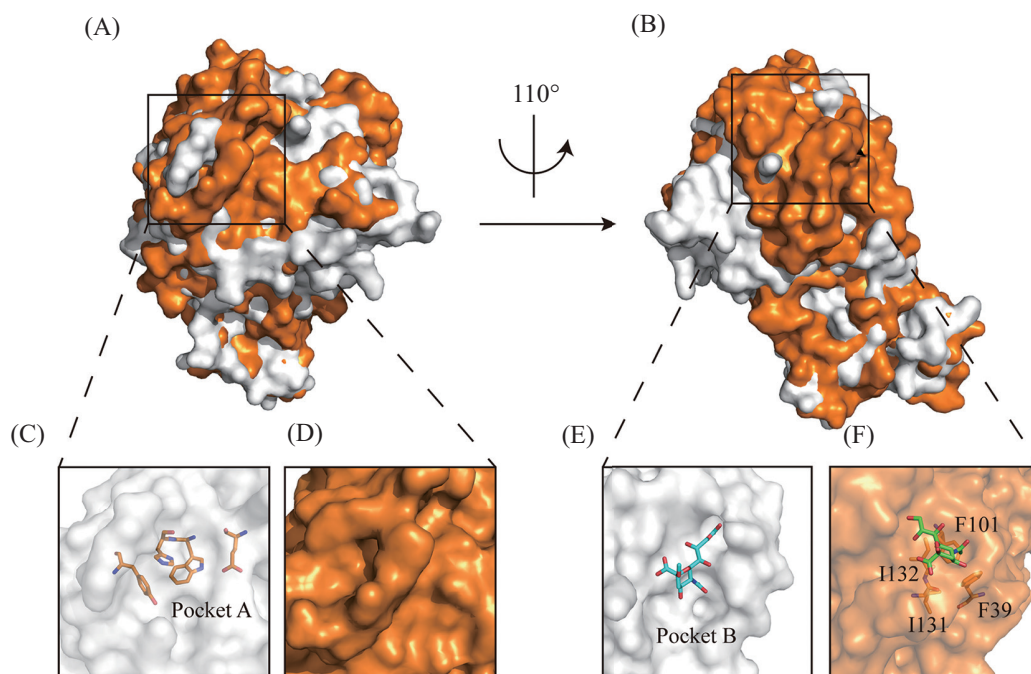
图5 BCoV、MHV和OC43的结构图

Fig.5 The structure diagrams of BCoV, MHV and OC43

与糖受体的结合^[31]。以MERS-CoV为例,与OC43 S1-NTD的口袋A结构相比,MERS-CoV S1-NTD在该结构位置没有保守的Y-E-W-H基序,分子构象发生变化,没有形成糖结合口袋,因此推测其没有糖结合能力(图6A、图6C和图6D)。VEESLER团队^[48]在2019年6月通过冷冻电镜解析了OC43 S1-NTD与Neu5,9Ac₂的复合物结构,同年12月他们利用相同的技术得到了MERS-CoV S1-NTD分别与Neu5Ac和Neu5Gc的复合物结构^[27]。通过结构比较,MERS-CoV S1-NTD识别结合Neu5Ac的凹槽与OC43 S1-NTD上结合Neu5,9Ac₂的凹槽(口袋B)不同,识别结合Neu5Ac的凹槽更浅显,由F39、F101、I131和I132 4个保守氨基酸残基组成。研究统计发现,以 α -2,3-方式连接的唾液酸(Neu5Ac)比以 α -2,6-方式连接的唾液酸(Neu5Gc)具有明显的与糖结合口袋结合的优势,5-*N*-羟乙酰基阻碍了Neu5Gc与糖结合口袋结合。由于在不同宿主的细胞以及同一宿主不同组织细胞上的唾液酸苷的修饰和连接方式不同,因此冠状病毒对宿主和组织嗜性是否起决定

性影响还有待继续探究。

斑点杂交实验、点突变和黏蛋白结合实验等技术证实^[22,32,48], α 属冠状病毒TGEV、PEDV和 γ 属冠状病毒IBV的S1-NTD也能识别糖受体,但由于目前TGEV、PEDV和IBV都没有明确的复合物结构,因此可以通过多序列比对及NTD结构比对分析的方式推测冠状病毒是否与唾液酸结合以及结合唾液酸的类型。如,SARS-CoV-2 NTD似乎保留了BCoV和其他 β 冠状病毒属成员NTD的许多核心结构特征,我们推测其可以结合唾液酸。最近,GIBSON等^[49]基于含唾液酸衍生物的纳米颗粒的研究表明,SARS-CoV-2的刺突蛋白可以与 α -*N*-乙酰神经氨酸结合。对于某些冠状病毒来说,结合唾液酸的能力是其入侵宿主细胞必不可少的。PDCoV通过与黏蛋白的糖链部分结合而穿过厚厚的黏液屏障,进入肠道上皮细胞,从而启动感染。TGEV冠状病毒通过与唾液酸结合,表现出猪肠道致病性,而与其关系密切的PRCoV由于编码S蛋白N末端的基因存在大量缺失,NTD无法结合唾液酸,



A: OC43 S1-NTD和MERS-CoV S1-NTD叠加结构图; OC43 S1-NTD标记为白色, MERS-CoV S1-NTD标记为橘黄色(PDB: 6Q04); B: A图结构绕中心轴顺时针旋转110°后的结构图像; C、D: OC43 S1-NTD的糖结合口袋A与MERS-CoV S1-NTD相对应位置结构比较图; E、F: OC43 S1-NTD糖结合口袋B结构与MERS-CoV S1-NTD中糖结合口袋的比较, 蓝色棍棒模型为Neu5,9Ac₂, 绿色棍棒模型为Neu5Ac, 黄色杆状模型标记氨基酸残基。

A: alignment of OC43 and MERS-CoV S1-NTDs. OC43 S1-NTD is colored white, MERS S1-NTD is colored orange (PDB: 6Q04); B: the structure image of figure A after 110° clockwise rotation around the central axis; C,D: structural comparison of OC43 S1-NTD sugar-binding pocket A with its corresponding position in MERS S1-NTD; E,F: comparison of structure between OC43 S1-NTD sugar binding pocket B and MERS S1-NTD sugar binding pocket. Blue stick model is Neu5,9Ac₂, green stick model is Neu5Ac, and the residue is yellow rod.

图6 OC43和MERS-CoV的结构图

Fig.6 The structure diagrams of OC43 and MERS-CoV

导致PRCoV在肠道中复制效率低,而在呼吸道中复制效率高。总之,唾液酸在调节冠状病毒在肠胃感染的趋向性方面起着重要作用,然而目前研究数据无法明确阐述唾液酸结合活性如何影响冠状病毒的肠道致病性,因此这个领域还需要进一步研究与探索。

5 总结

本文从分子生物学和结构生物学的角度综述了近年来冠状病毒S1-NTD识别、结合宿主受体等过程的研究进展。在冠状病毒漫长的进化过程中,不同属的冠状病毒逐渐进化并获得了识别不同类型受体(蛋白质和糖类)的能力,同时进化出了利用S蛋白不同结构域识别受体的能力,从而最大限度地拓展了病毒的组织嗜性和感染能力。例如冠状病毒S1-NTD或者S1-CTD编码序列的突变,会致使其编码的氨基酸序列发生相应变化,引起S蛋白结构的轻微变化,最终导致其识别受体的种类和能力发生明显变化。例如, MHV和BCoV的S1-NTD整体上非常相似,但是其受体结合环的细微构象变化(图5C)导致MHV的S1-NTD可以识别蛋白受体CEACAM1a,而BCoV的S1-NTD则识别糖受体。此外,冠状病毒也可以通过窃取宿主蛋白基因并将其转化为自身基因,从而使其自身具有新的特异性识别蛋白受体的能力。大量研究表明,冠状病毒NTD的核心结构与宿主半乳糖凝集素相似,是祖先半乳糖凝集素结构域在不同方向上进化的结果。祖先冠状病毒可能没有编码NTD核心结构的基因,它通过基因捕获,将外源宿主序列插入至编码S蛋白的基因中,从而获得了结合糖类受体的能力。在宿主免疫压力下,冠状病毒S1-NTD核心结构上面的顶部结构不断进化,进而更好地保护糖结合位点免受宿主免疫监视。因为半乳糖凝集素作为宿主蛋白可以不被宿主免疫系统识别,所以它没有类似的顶部结构。

有趣的是,冠状病毒也能够以不依赖CTD结合蛋白质受体的方式,利用NTD结合蛋白受体(例如: AXL)介导病毒感染,这些研究发现极大地拓展了人们对冠状病毒入侵宿主细胞的认知。冠状病毒复杂的进化史反映了它们适应新环境和新宿主的倾向性,并为深入了解新型冠状病毒是如何出现以及跨越物种障碍从而征服全新的宿主群体提供了思路。冠状病毒S1-NTD的受体识别机制可以为阐明病毒致病机理提供依据,同时我们可以根据S1-NTD的受体结合特征开发相应的阻断药物或抗体。

参考文献 (References)

- [1] DELLANNO C, VEGA Q, BOESENBERG D. The antiviral action of common household disinfectants and antiseptics against murine hepatitis virus, a potential surrogate for SARS coronavirus [J]. *Am J Infect Control*, 2009, 37(8): 649-52.
- [2] HEMIDA M G, ALI A M, ALNAEEM A. The middle east respiratory syndrome coronavirus (MERS-CoV) nucleic acids detected in the saliva and conjunctiva of some naturally infected dromedary camels in Saudi Arabia-2019 [J]. *Zoonoses Public Health*, 2021, 68(4): 353-7.
- [3] JUNG K, SAIF L J, WANG Q. Porcine epidemic diarrhea virus (PEDV): an update on etiology, transmission, pathogenesis, and prevention and control [J]. *Virus Res*, 2020, 286: 198045.
- [4] KOONPAEW S, TEERAVECHYAN S, FRANTZ P N, et al. PEDV and PDCoV pathogenesis: the interplay between host innate immune responses and porcine enteric coronaviruses [J]. *Front Vet Sci*, 2019, 6: 34.
- [5] ANAND K B, KARADE S, SEN S, et al. SARS-CoV-2: camazotz's curse [J]. *Med J Armed Forces India*, 2020, 76(2): 136-41.
- [6] CUI J, LI F, SHI Z L. Origin and evolution of pathogenic coronaviruses [J]. *Nat Rev Microbiol*, 2019, 17(3): 181-92.
- [7] KADAM S B, SUKHRAMANI G S, BISHNOI P, et al. SARS-CoV-2, the pandemic coronavirus: molecular and structural insights [J]. *J Basic Microbiol*, 2021, 61(3): 180-202.
- [8] HAO W, MA B, LI Z, et al. Binding of the SARS-CoV-2 spike protein to glycans [J]. *Sci Bull*, 2021, 66(12): 1205-14.
- [9] LI F. Structure, function, and evolution of coronavirus spike proteins [J]. *Annu Rev Virol*, 2016, 3(1): 237-61.
- [10] SCHOEMAN D, FIELDING B C. Coronavirus envelope protein: current knowledge [J]. *Virol J*, 2019, 16(1): 69.
- [11] BARANOWSKI E, RUIZ-JARABO C M, DOMINGO E. Evolution of cell recognition by viruses [J]. *Science*, 2001, 292(5519): 1102-5.
- [12] DU L, HE Y, ZHOU Y, et al. The spike protein of SARS-CoV—a target for vaccine and therapeutic development [J]. *Nat Rev Microbiol*, 2009, 7(3): 226-36.
- [13] OU X, LIU Y, LEI X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV [J]. *Nat Commun*, 2020, 11(1): 1620.
- [14] YAN R, ZHANG Y, LI Y, et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2 [J]. *Science*, 2020, 367(6485): 1444-8.
- [15] LI F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies [J]. *J Virol*, 2015, 89(4): 1954-64.
- [16] WANG N, SHI X, JIANG L, et al. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4 [J]. *Cell Res*, 2013, 23(8): 986-93.
- [17] HIRAI A, OHTSUKA N, IKEDA T, et al. Role of mouse hepatitis virus (MHV) receptor murine CEACAM1 in the resistance of mice to MHV infection: studies of mice with chimeric mCEACAM1a and mCEACAM1b [J]. *J Virol*, 2010, 84(13): 6654-66.
- [18] ZHU S, LIU Y, ZHOU Z, et al. Genome-wide CRISPR activation screen identifies candidate receptors for SARS-CoV-2 entry [J]. *Sci China Life Sci*, 2021, doi: 10.1007/s11427-021-1990-5.
- [19] WANG S, QIU Z, HOU Y, et al. AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bron-

- chial epithelial cells [J]. *Cell Res*, 2021, 31(2): 126-40.
- [20] LI Z, TOMLINSON A C, WONG A H, et al. The human coronavirus HCoV-229E S-protein structure and receptor binding [J]. *eLife*, 2019, 8: e51230.
- [21] HOFMANN H, PYRC K, VAN DER HOEK L, et al. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry [J]. *Proc Natl Acad Sci USA*, 2005, 102(22): 7988-93.
- [22] LIU C, TANG J, MA Y, et al. Receptor usage and cell entry of porcine epidemic diarrhea coronavirus [J]. *J Virol*, 2015, 89(11): 6121-5.
- [23] PENG J Y, PUNYADARSANIYA D, SHIN D L, et al. The cell tropism of porcine respiratory coronavirus for airway epithelial cells is determined by the expression of porcine aminopeptidase N [J]. *Viruses*, 2020, 12(11): 1211.
- [24] YUAN P, YANG Z, SONG H, et al. Three main inducers of alphacoronavirus infection of enterocytes: sialic acid, proteases, and low pH [J]. *Intervirology*, 2018, 61(2): 53-63.
- [25] SZCZEPANSKI A, OWCZAREK K, BZOWSKA M, et al. Canine respiratory coronavirus, bovine coronavirus, and human coronavirus OC43: receptors and attachment factors [J]. *Viruses*, 2019, 11(4): 328.
- [26] HULSWIT R J G, LANG Y, BAKKERS M J G, et al. Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A [J]. *Proc Natl Acad Sci USA*, 2019, 116(7): 2681-90.
- [27] TORTORICI M A, WALLS A C, LANG Y, et al. Structural basis for human coronavirus attachment to sialic acid receptors [J]. *Nat Struct Mol Biol*, 2019, 26(6): 481-9.
- [28] HUANG X, DONG W, MILEWSKA A, et al. Human coronavirus HKU1 spike protein uses *O*-acetylated sialic acid as an attachment receptor determinant and employs hemagglutinin-esterase protein as a receptor-destroying enzyme [J]. *J Virol*, 2015, 89(14): 7202-13.
- [29] HOU Y X, PENG C, YU M, et al. Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry [J]. *Arch Virol*, 2010, 155(10): 1563-9.
- [30] 沈娟, 陈冰清, 于瑞嵩, 等. 冠状病毒S蛋白及其受体的结构和功能 [J]. *微生物学通报* (SHEN M, CHEN B Q, YU R S, et al. Structure and function of coronaviral S proteins and their receptors [J]. *Microbiol China*), 2017, 44(10): 2452-62.
- [31] LI W, HULSWIT R J G, WIDJAJA I, et al. Identification of sialic acid-binding function for the middle east respiratory syndrome coronavirus spike glycoprotein [J]. *Proc Natl Acad Sci USA*, 2017, 114(40): E8508-E17.
- [32] WINTER C, SCHWEGMANN-WESSELS C, CAVANAGH D, et al. Sialic acid is a receptor determinant for infection of cells by avian infectious bronchitis virus [J]. *J Gen Virol*, 2006, 87(Pt 5): 1209-16.
- [33] JI W, PENG Q, FANG X, et al. Structures of a deltacoronavirus spike protein bound to porcine and human receptors [J]. *Nat Commun*, 2022, 13(1): 1467.
- [34] LI Y, WANG T, ZHANG J, et al. Exploring the regulatory function of the N-terminal domain of SARS-CoV-2 spike protein through molecular dynamics simulation [J]. *Adv Theory Simul*, 2021, 4(10): 2100152.
- [35] WALLS A C, TORTORICI M A, FRENZ B, et al. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy [J]. *Nat Struct Mol Biol*, 2016, 23(10): 899-905.
- [36] GUAN H, WANG Y, PERCULIJA V, et al. Cryo-electron microscopy structure of the swine acute diarrhea syndrome coronavirus spike glycoprotein provides insights into evolution of unique coronavirus spike proteins [J]. *J Virol*, 2020, 94(22): e01301-20.
- [37] YU J, QIAO S, GUO R, et al. Cryo-EM structures of HKU2 and SADS-CoV spike glycoproteins provide insights into coronavirus evolution [J]. *Nat Commun*, 2020, 11(1): 3070.
- [38] WRAPP D, MCLELLAN J S. The 3.1-angstrom cryo-electron microscopy structure of the porcine epidemic diarrhea virus spike protein in the prefusion conformation [J]. *J Virol*, 2019, 93(23): e00923-19.
- [39] SHANG J, ZHENG Y, YANG Y, et al. Cryo-electron microscopy structure of porcine deltacoronavirus spike protein in the prefusion state [J]. *J Virol*, 2018, 92(4): e01556-17.
- [40] MILLET J K, JAIME S J A, WHITTAKER G R. Molecular diversity of coronavirus host cell entry receptors [J]. *FEMS Microbiol Rev*, 2021, 45(3): fuaa057.
- [41] SHANG J, WAN Y, LIU C, et al. Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry [J]. *PLoS Pathog*, 2020, 16(3): e1008392.
- [42] XIONG X, COOMBS P J, MARTIN S R, et al. Receptor binding by a ferret-transmissible H5 avian influenza virus [J]. *Nature*, 2013, 497(7449): 392-6.
- [43] UJIE M, TAKADA K, KISO M, et al. Long-term culture of human lung adenocarcinoma A549 cells enhances the replication of human influenza A viruses [J]. *J Gen Virol*, 2019, 100(10): 1345-9.
- [44] SEYRAN M, TAKAYAMA K, UVERSKY V N, et al. The structural basis of accelerated host cell entry by SARS-CoV-2 dagger [J]. *FEBS J*, 2021, 288(17): 5010-20.
- [45] PENG G, XU L, LIN Y L, et al. Crystal structure of bovine coronavirus spike protein lectin domain [J]. *J Biol Chem*, 2012, 287(50): 41931-8.
- [46] FANTINI J, DI SCALA C, CHAHINIAN H, et al. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection [J]. *Int J Antimicrob Agents*, 2020, 55(5): 105960.
- [47] PARK Y J, WALLS A C, WANG Z, et al. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors [J]. *Nat Struct Mol Biol*, 2019, 26(12): 1151-7.
- [48] SCHWEGMANN-WESSELS C, HERRLER G. Sialic acids as receptor determinants for coronaviruses [J]. *Glycoconj J*, 2006, 23(1/2): 51-8.
- [49] BAKER A N, RICHARDS S J, GUY C S, et al. The SARS-CoV-2 spike protein binds sialic acids and enables rapid detection in a lateral flow point of care diagnostic device [J]. *ACS Cent Sci*, 2020, 6(11): 2046-52.