(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0153750
(43) 공개일자 2024년10월24일

(51) 국제특허분류(Int. Cl.)
G16B 15/30 (2019.01) *G06N 3/0464* (2023.01)
G16B 15/10 (2019.01) *G16B 15/20* (2019.01)
G16B 30/00 (2019.01) *G16B 40/20* (2019.01)
G16B 45/00 (2019.01) *G16B 50/40* (2019.01)
G16H 50/80 (2018.01)
(52) CPC특허분류
G16B 15/30 (2019.02)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0050043
(22) 출원일자 2023년04월17일
심사청구일자 2023년04월17일

(71) 출원인
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
(72) 발명자
이상수
강원도 춘천시 후석로 459, 101동 1504호
(74) 대리인
특허법인본

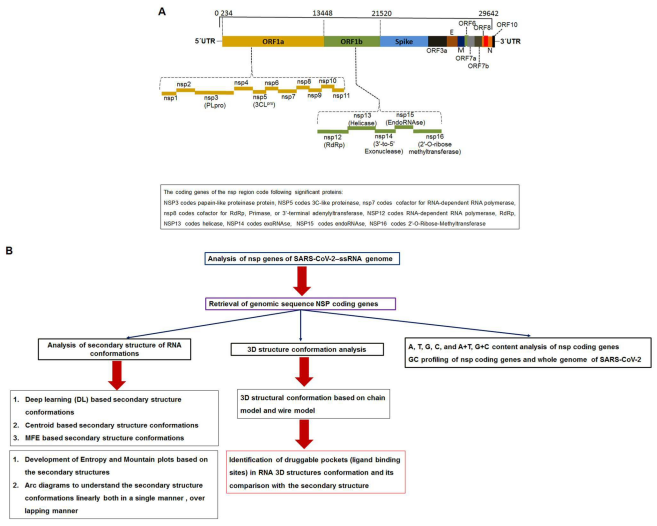
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **SARS-CoV-2 RNA 게놈의 2차원 및 3차원 구조 분석 방법**

(57) 요 약

본 발명은 SARS-CoV-2 RNA 게놈의 2차원 및 3차원 구조 분석 방법, 그리고 약물 결합 가능 포켓 분석 방법에 관한 것이다. 발명은 SARS-CoV-2 게놈 서열과 비구조단백질(non structural protein) 코딩 유전자를 검색하는 단계; SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계; SARS-CoV-2의 RNA의 3차원 구조를 분석하는 단계; SARS-CoV-2의 RNA의 GC 함량을 분석하는 단계;를 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법을 제공한다.

대 표 도 - 도1



(52) CPC특허분류
G16B 15/10 (2019.02)
G16B 15/20 (2019.02)
G16B 30/00 (2019.02)
G16B 40/20 (2019.02)
G16B 45/00 (2019.02)
G16B 50/40 (2019.02)
G16H 50/80 (2018.01)

명 세 서

청구범위

청구항 1

SARS-CoV-2 게놈 서열과 비구조단백질(non structural protein) 코딩 유전자를 검색하는 단계;

SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계;

SARS-CoV-2의 RNA의 3차원 구조를 분석하는 단계;

SARS-CoV-2의 RNA의 GC 함량을 분석하는 단계;를 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 2

제1항에 있어서,

2차원 구조를 분석하는 단계는,

완전 합성곱 신경망(fully convolutional networks)을 이용하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계인, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 3

제2항에 있어서,

2차원 구조를 분석하는 단계는,

최소 자유 에너지 기반으로 하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 4

제3항에 있어서,

2차원 구조를 분석하는 단계는,

최소 염기쌍 거리를 갖는 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 5

제4항에 있어서,

3차원 구조를 분석하는 단계는,

웹서버를 통해 SARS-CoV-2의 RNA의 3차원 구조를 라인 체인 모델 및 와이어 모델 기반으로 분석하는 단계인, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 6

제1항에 있어서,

GC 함량을 분석하는 단계는,

SARS-CoV-2 게놈의 비구조단백질 코딩 게놈 영역의 GC 함량과 SARS-CoV-2 게놈의 전체 GC 함량을 비교하는 단계를 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법.

청구항 7

nsp7, nsp12 및 nsp13을 코딩하는 SARS-CoV-2 게놈 영역의 3차원 구조를 분석하는 단계;

보로노이 셀 기반 방법으로 단백질 패킹 밀도를 분석하여 약물 결합 가능 포켓 후보를 선정하는 단계를 포함하는, 약물 결합 가능 포켓의 분석 방법.

청구항 8

제7항에 있어서,

단백질 패킹 밀도는 아래 식에 의해 계산되는, 약물 결합 가능 포켓의 분석 방법:

PD(패킹 밀도; packing density) = V_{vdW} (V_{vdW} +V_{se})

여기서, V_{vdW}는 원자의 반데르발스 필드 바로 내부에 있는 각 원자에 할당된 부피이며, V_{se}는 나머지 부피이다.

청구항 9

제7항에 있어서,

nsp7을 코딩하는 영역은 RNA의존성 RNA 중합효소에 대한 보조인자 단백질 코딩하는 영역이고, nsp12를 코딩하는 영역은 RNA 의존성 RNA 중합효소 단백질을 코딩하는 영역이며, nsp13을 코딩하는 영역은 헬리카아제(helicase) 단백질을 코딩하는 영역인, 약물 결합 가능 포켓의 분석 방법.

청구항 10

전자 장치의 프로세서에 의해 실행되어, 상기 전자 장치로 하여금,

SARS-CoV-2 게놈 서열과 비구조단백질(non structural protein) 코딩 유전자를 검색하는 단계;

SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계;

SARS-CoV-2의 RNA의 3차원 구조를 분석하는 단계;

SARS-CoV-2의 RNA의 GC 함량을 분석하는 단계;를 포함하는 방법을 수행하도록 하는, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램.

청구항 11

제10항에 있어서,

2차원 구조를 분석하는 단계는,

완전 합성곱 신경망(fully convolutional networks)을 이용하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계인, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램.

청구항 12

제11항에 있어서,

2차원 구조를 분석하는 단계는,

최소 자유 에너지 기반으로 하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함하는, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램.

청구항 13

제12항에 있어서,

2차원 구조를 분석하는 단계는,

최소 염기쌍 거리를 갖는 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함하는, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 SARS-CoV-2 RNA 게놈의 2차원 및 3차원 구조 분석 방법, 그리고 약물 결합 가능 포켓 분석 방법에 관한 것이다.

배 경 기 술

[0002] SARS-CoV-2는 코로나바이러스과(Coronaviridae family)의 RNA 바이러스로 현재 대유행의 원인 병원체이다. 이 바이러스는 전체 RNA 바이러스 계열 중에서 가장 큰 RNA 중 하나를 구성하는 약 ~30kb 게놈을 가진 단일 가닥 RNA(ssRNA) 게놈을 포함한다. 게놈은 5' 캡 구조 및 3' 폴리(A) 꼬리와 같은 게놈의 몇 가지 중요한 구조적 부분을 포함하는 양성 센스 RNA를 포함한다. 더욱이 게놈 구조는 약 29개의 단백질을 암호화한다. 여러 연구자들이 게놈 구조를 설명하려고 시도했다. 기존 연구자들은 새로운 치료법을 개발하기 위한 전략을 제공하기 위해 SARS-CoV-2 게놈 구조에 대한 통찰력 있는 관점을 제공하려고 시도했다. 이 방향에서 그들은 5' UTR 구조, ribosomal FSE(frameshifting element), 3' UTR 구조와 같은 다른 게놈 구조 요소를 설명했다. 유사하게 SARS-CoV-2 RNA 게놈에서 여러 구조적 요소가 예상되었고 게놈에 대해 생체 내 구조적 특성 분석이 수행되었다. 그들은 42개의 숙주 단백질이 SARS-CoV-2 RNA에서 투사된 다양한 요소와 상호 작용할 수 있음을 확인했다. 바이러스에 대한 새로운 치료법을 개발하는 데 도움이 될 수 있다. 유사하게, 기존 연구자들은 SARS-CoV-2 RNA의 게놈 구조를 개발했다. 연구자들은 RNA 2차 구조를 기반으로 SARS-CoV-2 게놈의 패킹을 이해하려고 노력했다. 이 연구는 바이러스의 구조, 기능 및 역학에 대한 게놈의 프레임워크를 이해하려고 했다. 마지막으로 그들은 RNA 기반 약물을 위한 siRNA를 설계하려고 시도했다.

[0003] RNA 2차 구조의 대규모 특성화는 SARS-CoV-2 게놈 전체 조직에 대한 통찰력을 제공할 수 있다. 기존 연구자들은 RNA 2차 구조를 통해 SARS-CoV-2 게놈 구성을 설명하고 이를 다른 코로나바이러스와 비교했다. 여러 연구에서 바이러스 복제와 전사 및 번역 측면에서 RNA 이차 구조의 중요성을 설명한다. SARS-CoV-2의 게놈은 구조 및 비구조 단백질(nsps)의 두 가지 유형의 단백질을 암호화한다(도 1a). nsps는 전사, 외피 단백질 생산 및 바이러스 복제와 같은 바이러스의 기본 과정을 돕는 중요한 단백질이다. 이러한 모든 NSP는 복제를 수행하기 위해 RNA 의존성 RNA 중합효소(RdRp)를 지원한다. 따라서 복제 기계 역할을 한다. 과학자들은 약물 표적으로서 비구조적 단백질(nsps)을 사용했다. SARS-CoV-2의 NSP 코딩 유전자는 이러한 관점에서 중요한다. 2개의 개방 판독 프레임(ORF1a 및 ORF1b)은 nsp1에서 nsp16인 16개의 NSP의 코딩과 관련되어 있음이 주목되었다(도 1a). ORF1a는 nsp1에서 nsp11로 코딩하고 ORF1b는 nsp12에서 nsp16 단백질로 코딩한다. 주요 단백질은 nsp3(papain-like proteinase protein), nsp5(3C-like proteinase), nsp12(RNA-dependent RNA polymerase, RdRp), nsp13(helicase, HEL), NSP14(exoRNAse), nsp15(endoRNAse), nsp16 (2'-O-Ribose-Methyltransferase) 및 기타 비구조 단백질을 포함한다. 다른 중요한 nsp는 nsp7(RNA 의존성 RNA 폴리머라제에 대한 보조인자), nsp8(RdRp, primase 또는 3'-말단 아데닐릴트랜스퍼라제에 대한 보조인자)이다. 대부분의 연구 작업은 nsps(nsp1-16)의 구조적 및 기능적 관점을 설명하기 위해 수행되었다. 그러나 유전자(ORF1a 및 ORF1b)의 nsp 코딩과 게놈 구조 배열에 대한 통찰력을 보여주는 증거는 적다. 따라서 nsp 코딩 게놈 영역(ORF1a 및 ORF1b)의 구조적 통찰력과 구조 기반 게놈 배열을 이해하는 것이 시급한다.

[0004] 2차 및 3차 RNA 구조는 SARS-CoV-2의 수명 주기에서 중요한 기능적 역할을 한다. 연구자들은 3D 구조를 사용하여 이 바이러스에 대한 구조 기반 약물 발견을 도울 수 있는 구조의 구성 요소를 분석하고 있다. SARS-CoV-2의 RNA 게놈 구조를 설명하려는 기존 연구는 거의 없다. 기존 연구자들은 SARS-CoV-2 게놈의 RNA 서열 보존을 평가하고 잠재적인 항바이러스 약물 표적으로 작용하는 106개의 RNA 기반 보존 영역을 발견했다. 기존 연구자들은 NSP 사이의 상호 작용을 발견한 RNA 구조 기반 복제 기계를 이해하려고 노력했다. 기존 연구자들은 바이러스 감염에 중요한 역할을 하는 SARS-CoV-2 RNA 게놈의 RNA 2차 구조 요소를 분석했다. 그들의 연구는 RNA 기반 구조 요소를 사용하여 표적화하는 안티센스 올리고뉴클레오티드를 평가하는 데 도움이 될 수 있으며 이 바이러스에 대한 치료제 개발에 도움이 될 것이다. 기존 연구자들은 치료 개발을 도울 수 있는 RNA 구조 기반 요소를 식별하는 데 기여했다. 예를 들어 몇몇 연구자는 SARS-CoV-2 게놈 전체에 걸쳐 RNA 구조를 매핑했다. RNA 요소는 게놈에서 확인되었다. SARS-CoV-2 게놈을 설명하기 위해 2차 구조 패턴이 생성되었다. 그들은 RNA 기반 치료 혁신을 도울 수 있는 RNA의 3D 구조를 모델링했다. 그러나 nsp 유전자의 2차 및 3D 구조적 RNA 형태, 약물 결합 가능한 게놈 및 바이러스 게놈의 RNA 기반 치료 표적은 대부분 미개척 상태이다. 따라서 SARS-CoV-2 게놈의 2차 3D RNA 구조 형태와 바이러스에 대한 RNA 표적 약물 발견을 시작할 수 있는 약물 가능 게놈의 기본 단위에 대한

심층 분석을 이해하는 것이 필요하다.

[0005] 본 발명자들은 SARS-CoV-2의 RNA 게놈의 NSP 코딩 유전자의 2차 구조 형태, 엔트로피 및 마운틴 플롯, 선형 방식의 RNA 2차 구조 및 3D 구조를 평가하려고 시도했다. 최근 알고리즘 기반(딥 러닝 기반) 2차 구조, 기존 알고리즘 기반(센트로이드 및 MFE 기반) 2차 구조 형태, 2차 구조 기반 엔트로피 및 마운틴 플롯, 아크 다이어그램의 단일 및 중첩 구조를 선형 방식으로 예측했다. 모든 nsp 코딩 게놈 영역(nsp1-16)의 마지막으로, RNA 3D 구조 모델은 이러한 2차 구조 형태로부터 예측되었다. 또한 모든 NSP 코딩 유전자의 GC 함량을 이해하기 위해 A, T, G, C 및 A+T, G+C 함량 및 GC 프로파일링을 분석했다. 마지막으로, 본 발명자들은 RNA 3D 구조에서 2차 구조 유도 약물 결합 가능 포켓(리간드 결합 부위)을 확인했다.

선행기술문헌

[0006] Umair, M., Ikram, A., Salman, M., Khurshid, A., Alam, M., Badar, N., Suleman, R., Tahir, F., Sharif, S., Montgomery, J., & Whitmer, S. (2021). Whole-genome sequencing of SARS-CoV-2 reveals the detection of G614 variant in Pakistan. PLoS ONE, 16(3), e0248371.

[0007] 2. De Wit, E., Van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. Nature Reviews Microbiology, 14(8), 523-534.

[0008] 3. Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z.W., Tian, J. H., Pei, Y. Y., & Yuan, M. L. (2020). A new coronavirus associated with human respiratory disease in China. Nature, 579(7798), 265-269.

[0009] 4. Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X., & Gong, M. (2021). In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. Cell, 184(7), 1865-1883. e20.

[0010] 5. Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., & Tummino, T. A. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature, 583(7816), 459-468.

[0011] 6. Manfredonia, I., & Incarnato, D. (2021). Structure and regulation of coronavirus genomes: State-of-the-art and novel insights from SARS-CoV-2 studies. Biochemical Society Transactions, 49(1), 341-352.

[0012] 7. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., Yang, M., Xing, X., Wang, Y., Li, M,. & Zhou, B. (2021). The architecture of the SARS-CoV-2 RNA genome inside virion. Nature Communications, 12(1), 1-14.

[0013] 8. Simmonds, P. (2020). Pervasive RNA secondary structure in the genomes of SARS-CoV-2 and other coronaviruses. MBio, 11(6), e01661-e1720.

[0014] 9. Zust, R., Miller, T. B., Goebel, S. J., Thiel, V., & Masters, P. S. (2008). Genetic interactions between an essential 3′ cis-acting RNA pseudoknot, replicase gene products, and the extreme 3′ end of the mouse coronavirus genome. Journal of Virology, 82(3), 1214-1228.

[0015] 10. Yang, D., & Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic 3′ and 5′ ends. Virus Research, 206, 120-133.

[0016] 11. Zhang, Y., Huang, K., Xie, D., Lau, J. Y., Shen, W., Li, P., Wang, D., Zou, Z., Shi, S., Ren, H., & Wang, Y. (2021). In vivo structure and dynamics of the SARS-CoV-2 RNA genome. Nature Communications, 12(1), 1-12.

[0017] 12. Mousavizadeh, L., & Ghasemi, S. (2021). Genotype and phenotype of COVID-19: Their roles in pathogenesis. Journal of Microbiology, Immunology and Infection, 54(2), 159-163.

[0018] 13. Mariano, G., Farthing, R. J., Lale-Farjat, S. L., & Bergeron, J. R. (2020). Structural characterization of SARS-CoV-2: Where we are, and where we need to be. Frontiers in Molecular Biosciences, 7, 344.

[0019] 14. Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. Nature Reviews Molecular Cell Biology, 23(1), 21-39.

[0020] 15. Raj, R. (2021). Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. Biochemistry and Biophysics Reports, 25, 100847.

[0021] 16. Rahimi, A., Mirzazadeh, A., & Tavakolpour, S. (2021). Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. Genomics, 113(1), 1221-1232.

[0022] 17. Subissi, L., Imbert, I., Ferron, F., Collet, A., Coutard, B., Decroly, E., & Canard, B. (2014). SARS-CoV ORF1b-encoded non-structural proteins 12-16: Replicative enzymes as antiviral targets. Antiviral Research, 101, 122-130.

[0023] 18. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews Microbiology, 19(3), 155-170.

[0024] 19. Rangan, R., Zheludev, I. N., Hagey, R. J., Pham, E. A., WaymentSteele, H. K., Glenn, J. S., & Das, R. (2020). RNA genome conservation and secondary structure in SARS-CoV-2 and SARSrelated viruses: A first look. RNA, 26(8), 937-959.

[0025] 20. Romano, M., Ruggiero, A., Squeglia, F., Maga, G., & Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. Cells, 9(5), 1267.

[0026] 21. Manfredonia, I., Nithin, C., Ponce-Salvatierra, A., Ghosh, P., Wirecki, T. K., Marinus, T., Ogando, N. S., Snijder, E. J., van Hemert, M. J., Bujnicki, J. M., & Incarnato, D. (2020). Genomewide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. Nucleic Acids Research, 48(22), 12436-12452.

[0027] 22. Fu, L., Cao, Y., Wu, J., Peng, Q., Nie, Q., & Xie, X. (2022). Ufold: Fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Research, 50(3), e14-e14.

[0028] 23. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubӧck, R., & Hofacker, I. L. (2008). The vienna RNA websuite. Nucleic Acids Research, 36(suppl_2), W70-W74.

[0029] 24. Ding, Y., Chan, C. Y., & Lawrence, C. E. (2005). RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA, 11(8), 1157-1166. https://doi.org/10.1261/rna. 2500605

[0030] 25. Li, Y., & Zhang, S. (2011). Finding stable local optimal RNA secondary structures. Bioinformatics, 27(21), 2994-3001. https:// doi.org/10.1093/bioinformatics/btr510

[0031] 26. Lai, D., Proctor, J. R., Zhu, J. Y. A., & Meyer, I. M. (2012). R-CHIE: A web server and R package for visualizing RNA secondary structures. Nucleic Acids Research, 40(12), e95-e95.

[0032] 27. Tsybulskyi, V., Mounir, M., & Meyer, I. M. (2020). R-chie: A web server and R package for visualizing cis and trans RNA-RNA, RNA-DNA and DNA-DNA interactions. Nucleic Acids Research, 48(18), e105-e105.

[0033] 28. Biesiada, M., Purzycka, K. J., Szachniuk, M., Blazewicz, J., & Adamiak, R. W. (2016). Automated RNA 3D structure prediction with RNAComposer. RNA Structure Determination (pp. 199- 215). New York, NY: Springer.

[0034] 29. Biesiada, M., Pachulska-Wieczorek, K., Adamiak, R. W., & Purzycka, K. J. (2016). RNAComposer and RNA 3D structure prediction for nanotechnology. Methods, 103, 120-127. https://doi.org/ 10.1016/j.ymeth.2016.03.010

[0035] 30. Pérot, S., Sperandio, O., Miteva, M. A., Camproux, A. C., & Villoutreix, B. O. (2010).

Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. Drug Discovery Today, 15(15-16), 656-667.

[0036] 31. Staritzbichler, R., Ristic, N., Goede, A., Preissner, R., & Hildebrand, P. W. (2021). Voronoia 4-ever. Nucleic Acids Research, 49(W1), W685-W690.

[0037] 32. Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M. A., Murakami, J., Kamahora, T., Shiraki, K., & Hino, S. (1999). Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. Journal of Virology, 73(5), 3582-3586.

[0038] 33. Gao, F., & Zhang, C. T. (2006). GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. Nucleic Acids Research, 34(Web Server), W686- W691. https://doi.org/10.1093/nar/gkl040

[0039] 34. Yesselman, J. D., Denny, S. K., Bisaria, N., Herschlag, D., Greenleaf, W. J., & Das, R. (2019). Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. Proceedings of the National Academy of Sciences, 116(34), 16847-16855.

[0040] 35. Abraham, M., Dror, O., Nussinov, R., & Wolfson, H. J. (2008). Analysis and classification of RNA tertiary structures. RNA, 14(11), 2274-2289.

[0041] 36. Shawan, M. M. A. K., Sharma, A. R., Bhattacharya, M., Mallik, B., Akhter, F., Shakil, M. S., Hossain, M. M., Banik, S., Lee, S. S., Hasan, M. A., & Chakraborty, C. (2021). Designing an effective therapeutic siRNA to silence RdRp gene of SARS-CoV-2. Infection, Genetics and Evolution, 93, 104951.

[0042] 37. Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., & Chang, H. (2020). The architecture of SARS-CoV-2 transcriptome. Cell, 181(4), 914-921.e10.

[0043] 38. Smyth, R. P., Negroni, M., Lever, A. M., Mak, J., & Kenyon, J. C. (2018). RNA structure—A neglected puppet master for the evolution of virus and host immunity. Frontiers in Immunology., 9, 2097.

[0044] 39. Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., & Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature, 584(7819), 154-156.

[0045] 40. Warner, K. D., Hajdin, C. E., & Weeks, K. M. (2018). Principles for targeting RNA with drug-like small molecules. Nature Reviews Drug Discovery, 17(8), 547-558.

[0046] 41. Scott, H. L., Buckner, N., Fernandez-Albert, F., Pedone, E., Postiglione, L., Shi, G., Allen, N., Wong, L. F., Magini, L., Marucci, L., & O'Sullivan, G. A. (2020). A dual druggable genome-wide siRNA and compound library screening approach identifies modulators of parkin recruitment to mitochondria. Journal of Biological Chemistry, 295(10), 3285-3300.

[0047] 42. Moumné, L., Marie, A.-C., & Crouvezier, N. (2022). Oligonucleotide therapeutics: from discovery and development to patentability. Pharmaceutics, 14(2), 260.

[0048] 43. Hu, B., Zhong, L., Weng, Y., Peng, L., Huang, Y., Zhao, Y., & Liang, X. J. (2020). Therapeutic siRNA: State of the art. Signal Transduction and Targeted Therapy, 5(1), 1-25.

[0049] 44. Cavasotto, C. N., Lamas, M. S., & Maggini, J. (2021). Functional and druggability analysis of the SARS-CoV-2 proteome. European journal of pharmacology, 890, 173705.

[0050] 45. Chaicumpa, W., Chaisri, U., Tapchaisri, P., Chongsa-Nguan, M.,& Pongponratn, E. (1987). Oral vaccine against cholera prepared from Vibrio cholerae antigen(s). Southeast Asian Journal of Tropical Medicine and Public Health, 18(2), 142-148.

[0051] 46. Koumakis, L. (2020). Deep learning models in genomics; are we there yet Computational and Structural Biotechnology Journal, 18, 1466-1473.

[0052] 47. Sato, K., Akiyama, M., & Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. Nature communications, 12(1), 1-9.

[0053] 48. Li, Y., Yang, X., Wang, N., Wang, H., Yin, B., Yang, X., & Jiang, W. (2020). GC usage of SARS-CoV-2 genes might adapt to the environment of human lung expressed genes. Molecular Genetics and Genomics, 295(6), 1537-1546.

[0054] 49. Hagen, J. B. (2000). The origins of bioinformatics. Nature Reviews Genetics, 1(3), 231-236. https://doi.org/10.1038/35042090

[0055] 50. Hogeweg, P. (2011). The roots of bioinformatics in theoretical biology. PLoS Computational Biology, 7(3), e1002021. https://doi.org/10.1371/journal.pcbi.1002021

발명의 내용

해결하려는 과제

[0056] 본 발명은 SARS-CoV-2 RNA 게놈의 2차원 및 3차원 구조 분석 방법, 그리고 약물 결합 가능 포켓 분석 방법을 제공하는 것을 목적으로 한다.

[0057] 본 개시의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 개시의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시 예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0058] 본 발명은 SARS-CoV-2 게놈 서열과 비구조단백질(non structural protein) 코딩 유전자를 검색하는 단계; SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계; SARS-CoV-2의 RNA의 3차원 구조를 분석하는 단계; SARS-CoV-2의 RNA의 GC 함량을 분석하는 단계;를 포함하는, SARS-CoV-2의 RNA의 구조 분석 방법을 제공한다.

[0059] 2차원 구조를 분석하는 단계는, 완전 합성곱 신경망(fully convolutional networks)을 이용하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계이다.

[0060] 2차원 구조를 분석하는 단계는, 최소 자유 에너지 기반으로 하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함한다.

[0061] 2차원 구조를 분석하는 단계는, 최소 염기쌍 거리를 갖는 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함한다.

[0062] 3차원 구조를 분석하는 단계는, 웹서버를 통해 SARS-CoV-2의 RNA의 3차원 구조를 라인 체인 모델 및 와이어 모델 기반으로 분석하는 단계를 포함한다.

[0063] GC 함량을 분석하는 단계는, SARS-CoV-2 게놈의 비구조단백질 코딩 게놈 영역의 GC 함량과 SARS-CoV-2 게놈의 전체 GC 함량을 비교하는 단계를 포함한다.

[0064] 본 발명은 또한, nsp7, nsp12 및 nsp13을 코딩하는 SARS-CoV-2 게놈 영역의 3차원 구조를 분석하는 단계; 보로노이 셀 기반 방법으로 단백질 패킹 밀도를 분석하여 약물 결합 가능 포켓 후보를 선정하는 단계를 포함하는, 약물 결합 가능 포켓의 분석 방법을 제공한다.

[0065] 단백질 패킹 밀도는 아래 식에 의해 계산된다: PD(패킹 밀도; packing density) = V_{vdW} (V_{vdW} +V_{se}) 여기서, V_{vdW}는 원자의 반데르발스 필드 바로 내부에 있는 각 원자에 할당된 부피이며, V_{se}는 나머지 부피이다.

[0066] nsp7을 코딩하는 영역은 RNA의존성 RNA 중합효소에 대한 보조인자 단백질 코딩하는 영역이고, nsp12를 코딩하는 영역은 RNA 의존성 RNA 중합효소 단백질을 코딩하는 영역이며, nsp13을 코딩하는 영역은 헬리카아제(helicase) 단백질을 코딩하는 영역이다.

[0067] 본 발명은 또한, 전자 장치의 프로세서에 의해 실행되어, 상기 전자 장치로 하여금, SARS-CoV-2 게놈 서열과 비구조단백질(non structural protein) 코딩 유전자를 검색하는 단계; SARS-CoV-2의 RNA의 2차원 구조를 분석하

는 단계; SARS-CoV-2의 RNA의 3차원 구조를 분석하는 단계; SARS-CoV-2의 RNA의 GC 함량을 분석하는 단계;를 포함하는 방법을 수행하도록 하는, 컴퓨터 판독 가능 매체에 저장된 컴퓨터 프로그램을 제공한다.

[0068] 2차원 구조를 분석하는 단계는, 완전 합성곱 신경망(fully convolutional networks)을 이용하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계이다.

[0069] 2차원 구조를 분석하는 단계는, 최소 자유 에너지 기반으로 하여 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함한다.

[0070] 2차원 구조를 분석하는 단계는, 최소 염기쌍 거리를 갖는 SARS-CoV-2의 RNA의 2차원 구조를 분석하는 단계를 추가로 포함한다.

발명의 효과

[0071] 본 발명은 SARS-CoV-2의 RNA 게놈의 NSP 코딩 유전자의 2차 구조, 3차 구조의 분석 방법, 그리고 약물 결합 가능 포켓의 분석 방법을 제공한다.

도면의 간단한 설명

[0072] 도 1은 SARS-CoV-3 게놈의 RNA 및 본 발명의 흐름도를 나타낸다. (a)는 SARS-CoV-2 ssRNA 게놈에는 비구조 단백질 코딩 유전자(nsp)(nsp1~16)와 구조 단백질 코딩 유전자가 포함되어 있음을 나타낸 것이고, (b)는 본 발명의 흐름도를 나타낸다.

도 2a 내지 2c는 스템 루프(SL), 다중 분기 루프, 슈도노트 및 벌지 등과 같은 구조의 다양한 구성 요소를 보여주는 모든 nsp 코딩 영역(nsp1-nsp16)의 딥러닝(DL)에 의해 예측된 RNA 2차 구조 모델을 나타낸다.

도 3a 내지 3c는 모든 nsp 코딩 영역(nsp1-nsp16)의 단일 방식, 중첩 방식 모두에서 선형으로 2차 구조 형태의 아크 다이어그램을 나타낸다.

도 4a 내지 4c는 nsp 코딩 영역(nsp1-nsp16)의 RNA의 3D 모델을 나타낸다.

도 5는 RNA 3차원 구조에서 식별된 리간드 결합 포켓(약물 결합 가능 포켓)을 나타낸다. A는 nsp7 코딩 구조에서의 리간드 결합 포켓, B는 nsp12 코딩 구조에서의 리간드 결합 포켓, C는 nsp13 코딩 구조에서의 리간드 결합 포켓을 나타낸다.

도 6은 nsp 코딩 영역의 GC 프로파일링을 나타낸다. A는 ORF1a 및 ORF1b 영역의 GC 프로파일링, B는 전체 게놈에서의 GC 프로파일링을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0074] 이하 본원을 실시예 및 실험예를 통하여 보다 상세하게 설명한다. 그러나 이들 실시예 및 시험예는 본 발명을 예시적으로 설명하기 위한 것으로 본원의 범위가 이들 실시예 및 시험예에 한정되는 것은 아니다.

[0076] **방법**

[0077] 본 발명자들의 분석은 2차 구조 형태 분석, 3D 분석 및 nsp 코딩 게놈 영역(nsp1-16)의 안정성의 세 부분으로 분류된다. 이 게놈 영역의 RNA 2차 구조, 3D 구조 및 안정성을 분석하기 위해 몇 가지 최근 방법과 본 발명의 알고리즘이 사용되었다.

[0078] SARSCoV2 게놈의 게놈 서열과 nsp 코딩 유전자 검색

[0079] NCBI(NCBI GenBank Id: MZ054889.1)에서 SARS-CoV-2 게놈 서열과 nsp 코딩 유전자를 검색했다. nsp 코딩 유전자와 해당 서열의 위치는 표 1에 나열되어 있다. nsp 코딩 유전자 서열(nsp1-16)은 추가 분석에 사용되었다.

표 1

Sl. No.	Gene	Coded protein	Position (in bp)
1.	nsp1	–	234-773
2.	nsp2	–	774-2687
3.	nsp3	Papain-like proteinase	2688-8522
4.	nsp4	–	8523-10022
5.	nsp5	3C-like proteinase	10023-10940
6.	nsp6	–	10941-11810
7.	nsp7	Cofactor for RNA-dependent RNA polymerase (RdRp)	11811-12059
8.	nsp8	Cofactor for RdRp, primase, or 3′-terminal adenylyltransferase	12060-12653
9.	nsp9	–	12654-12992
10.	nsp10	–	12993-13409
11.	nsp11	–	13410-13448
12.	nsp12	RNA-dependent RNA polymerase	13410-16204
13.	nsp13	Helicase	16205-18007
14.	nsp14	3′-to-5′ exonuclease	18008-19588
15.	nsp15	EndoRNAse	19589-20626
16.	nsp16	2′-O-ribose methyltransferase	20627-21520

[0080]

[0081] 2차 구조 형태 분석

[0082] 딥 러닝(DL) 기반 구조 형태 분석

[0083] DL은 핵산에서 단백질에 이르는 구조 분석 분야의 연구자들을 탐구하는 새로운 접근 방식이다. 본 발명자들은 RNA의 DL 기반 2차 구조를 분석하기 위해 DL 기반 UFold 서버를 사용했다. 인코더-디코더 아키텍처를 사용하는 이 서버는 FCN(완전 합성곱 신경망; Fully Convolutional Networks)을 사용하여 빠르고 효율적인 RNA 2차 구조를 생성한다. 구조 생성을 위해 U-Net 모델을 사용한다. 17×L×L 텐서는 U-Net 모델의 입력으로 사용된다. 모델은 연락처 맵을 통한 출력에 대해 L×L 대칭 점수를 사용하는 점수 행렬 Y를 예측했다. 전통적인 RNA 2차 구조 예측 방법의 경우 알고리즘은 주로 자유 에너지 최소화를 통한 열역학 모델을 기반으로 하며, 이는 강력한 사전 가정과 느린 실행 속도를 실행한다. UFold 서버는 주석이 달린 데이터와 뉴클레오티드 염기쌍 규칙을 직접 기반으로 RNA 2차 구조를 예측하기 위한 심층 학습 기반 방법을 제안한다. 이 접근법은 RNA의 2차 구조를 계산적으로 예측하는 데 더 적합할 수 있다.

[0084] 센트로이드 기반 2차 구조 형태 분석

[0085] 또 다른 2차 구조 형태는 최소 염기쌍 거리를 갖는 센트로이드 구조에 기반한 nsp 코딩 유전자에 대해 예측되었다. 센트로이드 기반 2차 구조는 RNA 서버(Vienna RNA 웹사이트)를 사용하여 예측되었다. 다양한 유형의 구조적 RNA에서 RNA 앙상블의 센트로이드 구조는 양의 예측 값으로 측정되는 예측 오류를 30.0% 적게 만듭니다. 또한 연구된 RNA의 고유한 고유 구조를 명백히 포함하는 완전한 모델을 사용할 수 없다고 예측했다.

[0086] MFE(minimum free energy) 기반 2차 구조 형태 분석

[0087] MFE 기반 2차 구조는 nsp 코딩 게놈 영역에 대해 생성되었다. 이 구조는 최소한의 자유 에너지를 활용하는 MFE

기반 2차 구조에 의해 예측되었다. MFE 구조는 Vienna RNA 웹사이트 서버를 사용하여 예측되었다. 서버는 루프 기반 에너지 모델 고유의 분해된 루프인 외부 베이스를 사용하는 최적의 컴퓨터 폴딩을 사용한다. MFE 구조를 예측하기 위해, 대안적으로, 각각의 허용가능한 염기쌍(분리된 염기쌍 및 스태킹 염기쌍)에 대해 주어진 서열의 RNA 천연 구조가 생성된다. 에너지적으로 가장 좋은 구조는 최소 자유 에너지 환경의 맥락에서 해당 염기쌍을 포함한다. 이 구조는 루프 기반 에너지 모델과 동적 프로그래밍 알고리즘을 사용하여 예측된다. RNA 2차 구조는 고유하게 루프 및 외부 염기로 분해될 수 있으므로, 루프 기반 에너지 모델은 RNA 2차 구조 s의 자유 에너지 F(s)를 s에 포함된 루프 L의 기여 자유 에너지 FL의 합으로 계산된다. 선택한 에너지 매개변수 세트와 주어진 온도(기본값은 37° C)에 따라 F(s)를 최소화하는 2차 구조 s가 계산됩니다.

[0088] 2차 구조에 기반한 엔트로피 및 마운틴 플롯 분석

[0089] 엔트로피 및 마운틴 플롯은 2차 구조를 기반으로 생성되었다. 엔트로피 플롯은 RNA 이차 구조의 열역학적 어셈블리와 RNA 이차 구조의 위치 엔트로피를 나타낸다. 엔트로피 플롯은 Vienna RNA 웹사이트 서버를 사용하여 예측되었다. 엔트로피 및 마운틴 플롯은 기존의 구조 도시화가 혼동이 되는 긴 뉴클레오티드 시퀀스에 독점적으로 유용하다. 이는 축(xy-다이어그램) 지향 플로팅으로, 서열 위치 대 뉴클레오타이드 위치를 도는 염기쌍의 수이다. 컴퓨터 프로그램 세트는 도트 플롯을 그래픽 사용자 인터페이스가 시각화할 수 있는 마운틴 플롯 좌표로 변환한다. 결과 플롯은 MFE 구조(빨간색)에서 파생된 두 개의 마운틴 플롯, 위치 엔트로피 곡선(녹색) 및 페어링 확률(검은색)의 세 가지 중요한 곡선을 표시한다.

[0090] 마운틴 플롯은 동일한 서버를 사용하여 생성되었다. 높이를 통해 2차 구조를 나타내며 RNA 2차 구조의 위치 높이를 알려준다. 높이 m(k)는 위치 k의 염기에 포함된 염기쌍의 수에 의해 할당된다.

[0091] 단일 방식, 중첩 방식 및 TransRNA-RNA 상호 작용 모두에서 선형으로 2차 구조 형태를 이해하기 위한 아크 다이어그램

[0092] Arc 다이어그램은 R-chie 서버를 통해 개발되었다. 아크 다이어그램은 단일 및 중첩 방식 모두에서 선형으로 2차 구조 형태를 설명하는 데 사용되었다. Arc 다이어그램을 개발하기 위해 서버는 R 패키지를 통해 시각화되고 패키지(R4RNA)를 사용하여 라인 다이어그램을 통해 그래픽으로 표현되는 나선 예측 방법 TRANSAT(TRANSAT 알고리즘)를 사용했다. 마지막으로, 그 결과는 본 발명자들이 trans-RNA-RNA 상호작용을 이해하는 데 도움이 될 것이다. 서버를 사용하여 본 발명자들은 trans-RNA-RNA 상호 작용을 설명했다.

[0093] 3D 구조적 형태 분석

[0094] 라인 체인 모델과 와이어 모델 기반의 3D 구조적 형태

[0095] RNA의 3D 구조를 예측하고 이에 대한 적절한 이해는 여전히 중요한 과제로 남아 있다. 3D 구조는 RNA 구조의 구성 요소를 돕고 RNA의 나선형 영역을 배치하는 데 도움을 줄 수 있다. nsp 코딩 게놈 영역(nsp1-16)의 3D 구조 형태는 RNAComposer에 의해 개발되었다. 여기에서 본 발명자들은 라인 체인 3D 모델과 와이어 3D 모델의 두 가지 유형의 3D 모델을 개발했다. RNAComposer 서버는 RNA(최대 500개 뉴클레오티드)의 3D 구조 예측을 위한 사용자 친화적이고 자유롭게 액세스할 수 있는 서버 대상이다. 사용자가 자세히 설명하는 RNA 2차 구조에 대한 완전 자동화된 조각 조립 기반 방법을 사용한다. 현재, 계산 생물학 분야에서 거리 제한 결과 전략의 통합을 기반으로 하는 이러한 새로운 계산 방법이 등장하고 있다.

[0096] RNA 3D 구조 형태에서 약물 결합 가능 포켓(리간드 결합 부위) 식별 및 이차 구조와의 비교

[0097] 약물 결합 가능 포켓의 식별은 구조 기반 약물 설계의 흥미로운 영역이다. Voronoia 4-ever 서버를 사용하여 nsp 코딩 유전자에서 RNA 3D 구조적 형태의 약물 결합 가능 포켓을 식별한다. 서버는 약물 결합 가능 포켓, 용매가 없는 표면, 패킹 결함 등을 계산한다. 또한 다음 공식을 사용하여 계산되는 PD로 표시되는 모든 원자 패킹 밀도(PD; packing density)에 대한 패킹 밀도를 계산한다.

[0098] $PD = V_{vdW} (V_{vdW} + V_{se})$

[0099] 여기서 보로노이 셀 방법을 사용하며, V_{vdW}는 원자의 $_{vd}W$(반데르발스) 필드 바로 내부에 있는 각 원자에 할당된 부피이다. 여기서 V_{se}는 나머지 부피로 정의되며 용매에서 제외된다.

[0100] nsp 코딩 게놈 전반에 걸친 GC 함량 및 GC-프로필 분석

[0101] A, T, G, C 및 A+T, G+C 함량 통한 안정성 분석

[0102] 기존 연구자들은 바이러스 게놈의 GC가 풍부한 영역을 분석하려고 시도했다. GC 함량 바이어스 영역을 이해하는 데 도움이 될 수 있다. GC 함량 계산기를 사용하고 검색된 게놈 시퀀스를 사용하여, nsp 코딩 게놈 영역(nsp1-16) SARS-CoV-2 게놈의 A, T, G, C 및 A+T, G+C 함량을 계산했다.

[0103] 함량 이해를 위한 GC-프로파일링

[0104] GCProfile 서버를 통한 GC-프로파일링 플롯을 통해 SARS-CoV-2의 전체 게놈에 대한 GC-프로파일링을 분석했다. 본 발명자들은 GC-Profile 서버를 통한 GC-프로파일링 플롯을 통해 바이러스의 nsp 코딩 게놈 영역의 GC-프로파일링을 평가하고 최종적으로 두 영역(nsp 코딩 게놈 영역 및 전체 게놈)의 GC-프로파일을 비교했다. GC-Profile 서버는 분할 알고리즘을 사용하여 GC-Profile을 추정한다. 분할 알고리즘에서 S로 표시되는 게놈 순서 지수가 계산된다. S는 다음과 같이 정의된다.

[0105] $S \equiv S(P) = a^2 + c^2 + g^2 + t^2$

[0106] GC-Profile은 주로 진핵 및 원핵 게놈을 분석하는 데 사용되는 웹 기반 상호작용 소프트웨어 시스템이다. 유기체 게놈 구성에 대한 정성적 및 정량적 보기를 제공한다. 서버 기반 결과는 또한 G+C 함량과 기타 중요한 게놈 환경(예: CpG 섬 및 유전자의 분포) 간의 관계를 재현하며, 이는 또한 포괄적으로 분석된다. 이것은 GC-profile이 고등 진핵 게놈의 isochore 구조를 분석하고 원핵 게놈에서 게놈 섬을 식별하는 데 적절한 초기 지점임을 입증했다.

[0107] 마지막으로, 서버는 게놈 또는 nsp 코딩 게놈 영역의 누적 GC 프로파일을 제시했다.

[0108] 본 발명자들 작업의 전체 방법론은 순서도를 통해 설명되었다(도 1b).

[0110] **결과**

[0111] 2차 구조 형태

[0112] 딥 러닝(DL) 기반 2차 구조 형태

[0113] 딥 러닝(DL) 기반 2차 구조는 SARS-CoV-2 게놈의 다른 nsp(nsp1-16) 코딩 영역을 개발했다. DL 기반 2차 구조 ORF1a 및 ORF1b 영역을 설명했다. ORF1a의 DL 기반 2차 구조는 nsp1-11의 코딩 게놈 영역을 통해 설명되었다(도 2 A1-A11). 마찬가지로, ORF1b의 DL 기반 2차 구조는 nsp12-16의 코딩 게놈 영역을 통해 설명되었다(도 2B1-B5). nsp1 코딩 영역은 14개 이상의 내부 루프를 보여준다. 본 발명자들은 또한 2~3개의 더 큰 내부 루프를 발견했다(도 2 A1). 이 지역에서는 다른 지역에 비해 긴 줄기가 발견된다. 마찬가지로 nsp2 코딩 영역은 14개 이상의 내부 루프를 보여준다(도 2 A2). 본 발명자들은 또한 2~3개의 큰 내부 루프를 발견했다. 벌지 루프, 다중 분기, 스템, 헤어핀 루프 및 내부 루프도 이 영역에서 발견되었다. 유사하게, nsp3 코딩 영역은 더 많은 수의 다중 분기 루프를 보여준다. 본 발명자들은 또한 이 영역에서 벌지 루프, 스템, 헤어핀 루프 및 내부 루프를 발견했다(도 2 A3). nsp4 코딩 영역은 3개의 더 큰 다중 분기 루프와 여러 개의 작은 다중 분기 루프를 보여준다(도 2 A4). 다시, nsp5 코딩 영역은 여러 개의 작은 다중 분기 루프와 함께 하나의 더 큰 다중 분기 루프를 나타냅니다(도 2 A5). nsp6 코딩 영역은 하나의 더 큰 다중 분기 루프를 나타냅니다. 4개의 중간 크기 다중 분기 루프와 다른 몇 개의 작은 다중 분기 루프(도 2 A6). 또한 nsp7 코딩 영역에서 2개의 중간 크기의 다중 분기 루프와 함께 하나의 더 큰 다중 분기 루프를 발견했다(도 2 A7). nsp8 코딩 영역에서 벌지 루프, 스템, 헤어핀 루프 및 내부 루프와 함께 큰 다중 분기 루프가 주목되었다(도 2 A8). 유사하게, nsp9 코딩 영역에서 벌지 루프, 스템, 헤어핀 루프 및 내부 루프와 함께 하나의 더 큰 다중 분기 루프를 발견했다(도 2 A9). nsp10 코딩 영역에서 구조가 복잡한 더 큰 다중 분기 루프를 발견했다(도 2 A10). 그러나 nsp11 코딩 영역의 구조는 벌지 루프, 내부 루프 및 스템을 보여주는 다른 영역보다 단순한다(도 2 A11). ORF1b 영역에는 nsp12 코딩 영역이 3개의 다중 분기 루프가 있는 다른 영역에 비해 더 복잡한다(도 2 B1). 유사하게, nsp13 코딩 영역에서 다중 분기 루프와 함께 여러 내부 루프를 발견했다(도 2 B2). 다시 nsp14 코딩 영역은 벌지 루프, 헤어핀 루프 및 내부 루프와 함께 5개의 다중 분기 루프를 나타낸다 (도 2 B3). 유사하게, nsp15 코딩 영역은 벌지 루프, 헤어핀 루프 및 내부 루프와 같은 다른 구조와 함께 세 개의 다중 분기 루프를 보여준다(도 2 B4). 마지막으로 nsp15 코딩 영역에서 내부 루프, 헤어핀 루프 및 벌지와 같은 다른 구조와 함께 두 개의 다중 분기 루프를 발견했다(도 2 B5).

[0114] 센트로이드 기반 2차 구조 형태

[0115] DL 기반 2차 구조와 마찬가지로 SARS-CoV-2 게놈의 nsps(nsp1-16)에 대한 코딩 영역의 센트로이드 기반 2차 구

조 형태도 개발했다. 본 발명자들은 nsp1-11의 코딩 게놈 영역을 통해 설명된 ORF1a의 센트로이드 기반 2차 구조를 개발했다(도 S1 A1-A11). 마찬가지로, ORF1b의 센트로이드 기반 2차 구조는 nsp12-16의 코딩 게놈 영역을 통해 입증되었다(도 S1 B1-B5). 두 경우(ORF1a 및 ORF1b)에서 nsp1-16의 코딩 게놈 영역 구조는 DL 기반 2차 구조에 비해 더 단순하다. nsp1은 3개의 다중 분기 루프 구조를 보여준다(도 S1A1). 마찬가지로 nsp2는 3개의 다중 분기 루프 구조를 보여준다(도 S1A2). 마찬가지로 nsp3은 선형 구조를 나타내며 선형 구조에서 여러 줄기 및 루프 구조가 생성되었다(도 S1A3). 유사하게, nsp4에서 더 큰 루프 구조를 가진 3개의 다중 분기 루프가 주목된다(도 S1A4). nsp5에서는 5~6개의 다중 분기 루프가 서로 다른 팽창으로 관찰된다. 또한 구조에서 헤어핀 루프와 내부 루프를 발견했다(도 S1A5). nsp6에서 주요 다중 분기 루프가 관찰된다(도 S1A6). 마찬가지로 nsp7에서 두 개의 주요 다중 분기 루프가 관찰된다(도 S1A7). 또한 nsp8에서는 전체 구조에서 여러 헤어핀 루프 및 내부 루프와 함께 3개의 주요 다중 분기 루프를 확인했다(도 S1A8). 마찬가지로 nsp9에서 두 개의 주요 다중 분기 루프가 표시되고 하나의 구조가 돌출로 표시된다(도 S1A9). 동시에 nsp10에서 두 개의 다중 분기 루프와 함께 여러 스템, 헤어핀 루프 및 내부 루프가 발견되었다(도 S1A10). nsp11 코딩 영역은 짧은 시퀀스 길이로 인해 작은 구조를 형성했다(도 S1A11). 마찬가지로 nsp12에서도 영역 구조에 약 9개의 다중 분기 루프가 표시된다. 본 발명자들은 또한 이 영역에서 여러 헤어핀 루프와 내부 루프를 발견했다(도 S1B1). 유사하게, nsp13에서 함께 연관되는 두 개의 중요한 다중 분기 루프가 주목된다(도 S1B2). nsp14는 6개의 다중 분기 루프가 있는 선형 구조를 개발했다(도 S1B3). 마찬가지로 nsp15는 4개의 다중 분기 루프를 포함하고 중간 다중 분기 루프는 2개의 주요 벌지 구조를 둘러싼다(도 S1B4). 마지막으로 nsp16은 두 개의 다중 분기 루프를 형성하고 마지막 다중 분기 루프는 하나의 주요 벌지 구조를 둘러싼다. 첫 번째는 여러 스템, 루프 및 헤어핀 구조를 포함한다(도 S1B5).

[0116] MFE 기반 2차 구조 형태

[0117] DL 기반 및 센트로이드 기반 2차 구조와 마찬가지로 바이러스 게놈의 nsps(nsp1-16) 코딩 영역의 MFE 기반 구조도 개발했다. 본 발명자들은 nsp1-11의 코딩 게놈 영역으로 ORF1a의 MFE 기반 2차 구조를 개발했다(도 S2A1-A11). 마찬가지로, ORF1b는 nsp12-16의 코딩 게놈 영역의 MFE 기반 구조를 통해 입증되었다(도 S2B1-B5). nsp1은 두 개의 주요 다중 분기 루프를 보여준다(도 S2A1). 마찬가지로 nsp2는 6개의 다중 분기 루프 구조를 보여준다(도 S2A2). nsp3는 여러 스템과 루프 구조가 있는 선형 구조를 나타낸다(도 S2A3). nsp3의 이 MFE 기반 보조 구조는 센트로이드 기반 보조 구조와 유사한다. 마찬가지로 게놈은 nsp4에서 5개의 다중 분기 루프에 취약한 숙주 단백질을 식별한다(도 S2A4). 유사한 유형의 센트로이드 기반 2차 구조인 5~6개의 다중 분기 루프가 관찰된다(도 S2A5). nsp6에서도 5~6개의 다중 분기 루프가 관찰되었다(도 S2A6). 마찬가지로 nsp7에서 하나의 주요 다중 분기 루프가 관찰된다. 본 발명자들은 구조물의 팽창을 확인했다(도 S2A7). 유사하게, nsp8에서 본 발명자들은 RNA 2차 구조의 다른 구성 요소를 가진 세 가지 주요 다중 분기 루프에 주목했다. 첫 번째와 마지막 다중 분기 구조(도 S2A8)에서 몇 가지 구조적 구성 요소(줄기, 루프, 헤어핀 루프)를 확인했다. 동시에 nsp9에는 두 개의 주요 다중 분기 루프가 표시되고 마지막 구조는 다른 구조 구성 요소로 표시된다(도 S2A9). 동시에 nsp10에서 두 개의 다중 분기 루프를 발견했다(도 4A10). 마찬가지로 영역 구조에서 약 8개의 다중 분기 루프를 포함하는 선형 구조를 발견했다(도 S2B1). 같은 방식으로 nsp13의 선형 구조를 찾았다(도 S2B2). 동시에 본 발명자들은 nsp14가 5개의 다중 분기 루프가 있는 선형 구조를 생성했음을 확인했다(도 S2B3). nsp15는 약 3개의 다중 분기 루프를 포함하고 중간 및 마지막 다중 분기 루프는 여러 스템, 루프 및 헤어핀 루프 구조를 둘러쌉니다(도 S2B4). 마침내 본 발명자들은 nsp16이 두 개의 다중 분기 루프를 형성하고 다중 분기 루프가 여러 스템, 루프 및 헤어핀 구조를 포함한다는 것을 발견했다(도 S2B5).

[0118] 2차 구조에 기반한 엔트로피 및 마운틴 플롯

[0119] 본 발명자들은 모든 nsp 코딩 영역의 엔트로피 및 마운틴 플롯을 묘사했다(도 S3A1-A11, B1-B5). nsp1은 엔트로피와 산도를 보여준다. 최대 엔트로피는 2.0(300~500개 위치)을 나타냈고 높이는 nsp1 영역에서 약 40이다(도 S3A1). 마찬가지로 nsp2가 표시된다. 1에서 500 위치 사이의 최대 엔트로피는 2.0인치이고 높이는 약 100이다(도 S3A2). nsp3은 최대 엔트로피 2.0(1~1000개 위치)과 최대 높이 200을 나타낸다(도 S3A3). 마찬가지로 nsp12에서 값이 3(1000-1500과 2000-2500 위치 사이)인 에너지 피크를 기록했다. 본 발명자들은 해당 지역에서 몇 가지 에너지 피크를 확인했다. 마운틴 플롯은 첫 번째 부분에 위치한 약 120의 최대 높이를 보여준다(도 S3B1). nsp15 영역에서 최대 엔트로피를 보였고 3.0(약 1000 위치 부근)으로 나타났으며 산 플롯의 높이는 약 60~90 정도이다(도 S3B4). 다시, nsp16 영역에서 최대 엔트로피를 확인했으며, 2.0~2.5(200~400 위치 사이, 약 600 근처)를 나타냈으며, 마운틴 플롯의 높이는 약 40~60이다(도 S3B5).

[0120] 단일 방식, 중복 방식 및 TransRNA-RNA 상호 작용 모두에서 선형으로 이차 구조 형태의 아크 다이어그램

[0121] 아크 다이어그램은 관련 RNA 서열을 2차 구조에 연결하는 RNA 구조의 가능한 염기쌍을 보여준다. 왼쪽에서 오른쪽으로(5'에서 3'으로 끝남) 수평선으로 묘사된다.

[0122] 본 발명자들은 모든 nsp 코딩 영역(nsp1-16)에 대해 단일 및 중첩 방식 모두에서 2차 구조 형태의 아크 다이어그램을 선형으로 묘사했다(도 3A1-A11, B1-B5). TRANSAT 알고리즘과 관련하여 nsp1 코딩 영역은 단일 라인의 경우 [1e-06, 1e-05] 값으로 높은 기본 페어링을 나타낸다. 마찬가지로 중첩 방식의 경우에도 염기쌍 형성 가능성이 높다는 것을 보여준다. ORF1a의 nsp1, nsp3, nsp4, nsp7, nsp9, nsp10의 코딩 영역에서 더 많은 선형 아크 구조가 발견되었다(도 3A1, A3, A4, A7, A9, A10). 마찬가지로 ORF1b의 nsp12 및 nsp15에서 더 많은 선형 아크 구조를 발견했다(도 3B1, B4). nsp7의 코딩 영역은 ORF1b 및 ORF1b의 모든 nsps 중에서 단일 중첩 패션의 두 경우 모두에서 더 조밀한 선형 아크 구조를 개발한 것으로 나타났다.

[0124] 3차원 구조적 형태 해석

[0125] 라인 체인 모델과 와이어 모델 기반의 3차원 구조적 형태

[0126] 본 발명자들은 모든 nsp 코딩 영역(nsp1-16)에 대한 폴딩을 이해하기 위해 2차 구조의 3D 모델을 개발했다(도 4A1-A11, B1-B5). 본 발명자들은 3D 구조 공간에서 RNA 구조의 적절한 폴딩을 이해했다. 여기서 본 발명자들은 모든 nsp 코딩 영역에 대한 라인 체인 및 와이어 모델의 스템 및 루프 구조를 확인했다. ORF1a 영역에서 nsp7 코딩 영역이 더 많은 스템 구조를 포함하는 것으로 밝혀졌다(도 4A7). 동시에 ORF1b 영역에서 본 발명자들은 nsp13 코딩 영역이 더 많은 스템 구조를 포함한다는 것을 발견했다(도 4B2).

[0127] 현재 서열 관련 RNA 나선 형태를 이해하는 것은 흥미로운 분야이다. 기존 연구자들은 또한 RNA 구조의 3D 모델의 RNA 나선 패턴을 이해하려고 노력하고 있다. 기존 연구자들이 RNA 3D 모델에서 결합된 형태와 결합되지 않은 형태를 평가하려고 시도하고 있었다. 서로 다른 결합 및 결합되지 않은 형태의 RNA 3D 모델은 이 형태의 RNA 나선 내에서 염기 쌍 형성의 경우에 나타났다. 본 발명자들은 또한 3D 형태의 RNA 패킹을 이해했다. 여기서 본 발명자들은 모든 nsp 코딩 영역(nsp1-16)에서 이중, 분기 구조, 3방향 접합 및 4방향 접합을 확인했다.

[0128] RNA 3D 구조 형태에서 약물 결합 가능 포켓(리간드 결합 부위) 식별 및 이차 구조와의 비교

[0129] 마지막으로 모든 nsp 코딩 영역(nsp1-16)의 약물 결합 포켓을 확인했다. 여기서 본 발명자들은 세 가지 nsp 코딩 구조(nsp7, nsp12 및 nsp13)를 표현했다(도 5a-c). 모든 경우에, 본 발명자들은 각 RNA 3D 구조 구조에서 두 개의 약물 결합 포켓을 보여주었다. 각 RNA 3D 구조의 각 약물 결합 포켓은 nsp 코딩 구조에서 예측된 RNA의 2차 구조와 비교되었다. 이 3개의 nsp 영역(nsp7, nsp12 및 nsp13)은 RNA의존성 RNA 중합효소(nsp7), RNA 의존성 RNA 중합효소 또는 RdRp(nsp12), 헬리카제 또는 HEL(nsp13)에 대한 보조 인자인 3개의 중요한 단백질을 코딩한다. 이 모든 단백질은 바이러스 수명 주기와 관련이 있다. 따라서, 약물 결합 가능 포켓(리간드 결합 부위)은 바이러스 수명 주기에 영향을 미치는 이러한 영역을 표적으로 하는 약물 후보를 식별하는 데 도움이 될 것이다. siRNA 기반 분자는 이러한 약물 결합 가능 포켓을 사용하여 예측할 수 있다. 이전에 본 발명자들은 RdRp를 표적으로 하는 siRNA 기반 치료제 후보를 확인했다.

[0130] nsp 코딩 게놈 전반에 걸친 GC 함량 및 GC-프로필 분석

[0131] nsp 코딩 게놈 영역의 A, T, G, C 및 A+T, G+C 함량(nsp1-16)

[0132] 여기에서 nsp 코딩 게놈 영역의 내용을 이해하기 위해 시퀀스 공간 GC-프로파일링에서 nsp 코딩 게놈 영역의 A, T, G, C 및 A+T, G+C 함량을 계산했다(nsp1-16) (도 S4 A1-A11 및 도 S4 B1-B5). 본 발명자들은 A+T가 G+C 함량보다 모든 nsp 코딩 영역에서 더 높다는 것을 발견했다. ORF1a의 모든 nsp(nsp1 내지 nsp11)의 G+C 함량은 48.52%, 39.76%, 35.75%, 36.53%, 38.13%, 36.21%, 37.39%, 38.22%, 40.12%, 42.45%, 53.80%로 기록된다. 동시에, ORF1b의 모든 nsp(nsp12 내지 nsp16)의 G+C 함량은 37.18%, 38.21%, 38.08%, 34.23%, 35.46%로 기록된다.

[0133] nsp 코딩 게놈 영역(nsp1-16)의 GC프로파일링 및 SARSCoV2의 전체 게놈과의 비교

[0134] ORF1a 및 ORF1b의 nsp 코딩 영역의 GC 프로파일링을 분석했다. 그래프는 4000bp 및 20,000bp 영역에서 가장 높은 피크를 보여준다. 동시에 그래프는 9000~12,000bp 영역에서 가장 낮은 피크를 보여준다(도 6a).

[0135] 전체 게놈의 GC-profiling은 4000 bp와 20,000 bp 영역에서 가장 높은 피크를 보였으며 갑자기 게놈의 마지막 부분에서 매우 높은 피크가 나타났다. 동시에 그래프는 9000~12,000 bp와 약 23,000 bp 영역에서 두 개의 가장 낮은 피크를 보여준다(도 6b). 그러나 nsp 코딩 게놈 영역(nsp1-16)의 GC-프로파일 값은 전체 게놈(약 0.38)에 비해 적다(약 0.375).

[0137] **논의**

[0138] 현재, 연구자들은 RNA 구조, 특히 게놈의 RNA 구조를 이해하려고 노력하고 있다. RNA 구조는 다른 중요한 과학적 사건(숙주 면역 및 바이러스의 진화)과 같은 다른 조건에서 활용될 수 있다. 동시에 복제 과정을 설명하기 위해서는 2차 구조 형태와 3차원 구조 형태의 RNA 구조에 대한 구조를 이해하는 것이 중요한다. 동시에 연구자들은 구조 기반 약물 설계의 초점 영역 중 하나인 약물 가능 게놈을 이해하려고 노력하고 있다. 따라서 게놈 구성 요소의 RNA 구조는 RNA 표적 약물 설계에 도움이 될 것이다. 이 방향에서 연구자들은 SARS-CoV-2 게놈에서 RNA 구조를 풀려고 노력하고 있다. SARS-CoV-2-ssRNA 게놈의 RNA 2차 및 3D 구조를 이해하는 것은 이 대유행 시나리오에서 중요하며 RNA 표적 치료 설계에 도움이 될 것이다. 기존 연구자들은 SARS Cov-2 게놈의 구조적 특징을 평가하고 치료적 의미가 있는 RNA 구조적 구성 요소를 확인하려고 시도했다. 이 방향에서 본 발명자들은 SARS-CoV-2 게놈의 nsp 코딩 영역의 2차 구조, 3D 구조 등급 및 GC 프로파일을 설명했다. 마지막으로 nsp 코딩 영역의 리간드 결합 포켓을 확인했다. 따라서 본 발명자들의 연구는 매우 중요하며 SARS-CoV-2에 대한 RNA 기반 치료법의 발견을 시작할 것이다.

[0139] Druggable genome-wide 치료용 siRNA 발견은 안티센스 올리고뉴클레오티드 치료제 개발을 위한 효과적인 접근법 중 하나이다. 기존 연구자들은 SARS-CoV-2 RNA 게놈 구조와 siRNA 기반 치료 설계를 연구했다. 그들은 관찰된 4개의 siRNA를 확인했으며 siRNA 처리 후 세포에서 감염이 발견되지 않았다. 기존 연구자들은 SARS-CoV-2의 RNA 게놈 구조를 설명하고 siRNA 기반 치료 설계를 추가로 탐구했다. 이 연구에서 연구자들은 siRNA의 잠재적인 표적이 되는 다양한 게놈 영역을 식별하려고 노력했다. 이 연구는 SARS-CoV-2에 대한 siRNA 기반 치료법을 개발하는 데 도움이 될 것이다. 이전에 본 발명자들은 SARS-CoV-2에 대한 RNA 기반 치료 발견의 예인 RdRp의 mRNA를 표적으로 하는 siRNA도 확인했다. 그러나 본 발명자들은 SARS-CoV-2의 RdRp mRNA를 표적으로 삼았다. 몇몇 기존 연구자들은 SARS-CoV-2의 전사체 프로필을 이해하려고 노력했다. SARS-CoV-2 수명 주기와 그 병원성을 이해하는 데 도움이 될 것이다. 동시에 이 연구는 RNA 기반 치료 설계를 탐색하는 데 도움이 될 것이다. SARS-CoV-2 프로테옴 내 약물 가능성 특성화는 바이러스 약물 표적 기반 치료를 위한 상당한 노력이다. 기존 연구자들은 전체 RNA 게놈에서 인코딩된 각 SARS-CoV-2 단백질의 기능적 역할을 측정했다. 그 후, 그들은 바이러스 프로테옴의 심층 in silico 약물 가능성 평가를 사용하여 실험적 상동성 모델 또는 구조 데이터를 사용하여 주요 구조적 환경을 분석했다.

[0140] 딥 러닝 기반 알고리즘과 도구는 게놈 구조를 설명하는 데 중요하며 RNA 이차 구조 생성에 정기적으로 사용된다. 기존 연구자들은 딥 러닝 도구를 사용하여 SARS-CoV-2의 RNA 게놈 구조를 설명하고 이 바이러스의 RNA 2차 구조를 개발했다. 기존 방법과 함께 본 발명자들 연구에서는 딥 러닝 기반 최신 도구를 사용하여 SARS-CoV-2 RNA 게놈의 nsp 코딩 영역의 2차 구조를 개발했다. 본 발명은 예측된 딥러닝 기반 2차 구조와 기존 2차 구조(Centroid 및 MFE 기반)를 비교했다(표 2 및 3). 동시에 nsp 코딩 영역의 GC 함량을 계산했다. GC 함량의 범위는 34.23%에서 48.52%로 나타났다. 하지만 본 발명자들은 nsp11 코딩 영역이 약 53.80%임을 발견했다. 그러나 이 부분의 시퀀스 길이가 너무 짧아서 이 영역을 고려하지 않았다. 다시 nsp 코딩 영역의 GC 프로파일을 평가하고 전체 SARS CoV-2 게놈과 비교했다. 몇몇 다른 연구자들은 SARS-CoV-2의 GC 사용량을 계산했다. 기존 연구자들은 SARS-CoV-2 유전자의 GC 사용량을 계산하고 게놈의 낮은 GC 함량을 발견했다. 바이러스 코딩 서열의 GC 함량은 약 38%인 것으로 보고되었다. 동시에 기존 연구원들은 인간과 함께 척추동물의 GC 함량이 약 60%임을 발견했다. 또한, 본 발명자들의 연구는 RNA 게놈에서 약물 결합 포켓을 확인했으며 SARS-CoV-2 또는 기타 감염성 RNA 바이러스의 치료 개발을 위한 유망한 접근 방식이다.

표 2

Sl No	Genomic regions	nsp coding genes	Deep learning-based secondary structures	Centroid-based secondary structures	MFE-based secondary structures
1	ORF1a	nsp1	Approximately 11 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model which is developed by us is conventional	Stem-loop, multi-branch loop noted. This structure is conventional
2		nsp2	Approximately 6 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model found as the conventional form	Stem-loop, multi-branch loop noted. This secondary structure is conventional
3		nsp3	Approximately 6 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model is conventional	Stem-loop, multi-branch loop noted. The secondary structure model is conventional
4		nsp4	Approximately 6 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. This secondary structure is found as conventional conformation	Stem-loop, multi-branch loop noted. It seems to be a conventional
5		nsp5	Approximately 9 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. We found the secondary structure as a conventional model	Stem-loop, multi-branch loop noted. We noted as conventional model
6		nsp6	Approximately one complicated is found with pseudoknots	Stem-loop, multi-branch loop noted. The model is shown as a conventional model	Stem-loop, multi-branch loop noted. We found the structure as a conventional model
7		nsp7	Approximately 2 stem-loop structure are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The structure is the conventional type of structure	Stem-loop, multi-branch loop noted. We observed the secondary structure as a conventional model
8		nsp8	Approximately 4 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. This secondary structure looks as conventional	Stem-loop, multi-branch loop noted. The secondary structure model is conventional
9		nsp9	Approximately 7 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model is conventional	Stem-loop, multi-branch loop noted. The secondary structure model found as the conventional form
10		nsp10	Approximately one complicated is found with pseudoknots	Stem-loop, multi-branch loop noted. The secondary structure model is conventional	Stem-loop, multi-branch loop noted. The secondary structure model found as a conventional form
11		nsp11	One small stem-loop structure is found	Stem-loop, multi-branch loop noted. We found the structure as a conventional model	Stem-loop, multi-branch loop noted. The secondary structure model found as a conventional form

[0142]

표 3

Sl No	Genomic regions	nsp coding genes	Deep learning-based secondary structures	Centroid-based secondary structures	MFE-based secondary structures
12	ORF1b	nsp12	Approximately 6 stem-loop structure are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model is conventional	Stem-loop, multi-branch loop noted. The secondary structure model found as a typical form
13		nsp13	Approximately one complicated is found with pseudoknots	Stem-loop, multi-branch loop noted. The secondary structure model is typical	Stem-loop, multi-branch loop noted. The secondary structure model found as the typical form
14		nsp14	Approximately 5 stem-loop structure are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. This secondary structure looks like a typical model	Stem-loop, multi-branch loop noted. This secondary structure is a typical form
15		nsp15	Approximately 4 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. The secondary structure model which is developed by us is typical	Stem-loop, multi-branch loop noted. This structure found as the typical form
16		nsp16	Approximately 8 stem-loop structures are found and pseudoknots are noted	Stem-loop, multi-branch loop noted. We found the structure as a conventional model	Stem-loop, multi-branch loop noted. The secondary structure model found as a conventional form

[0143]

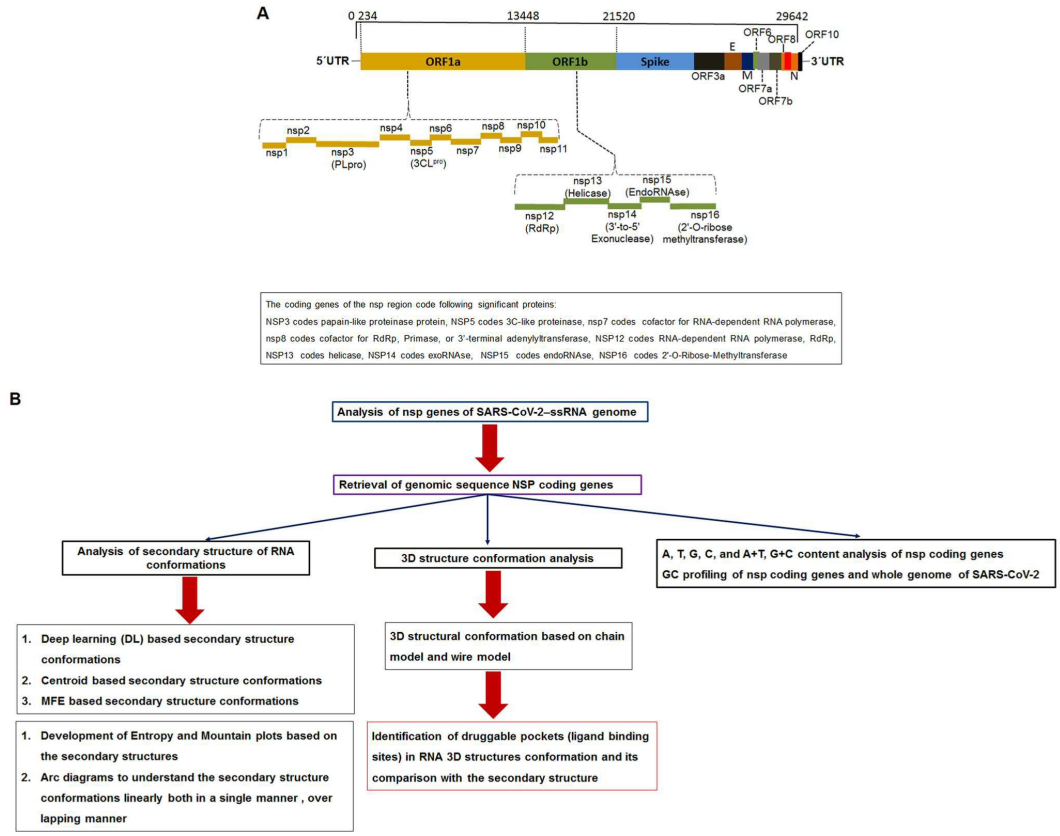
[0145] **결론**

[0146] 본 발명자들의 연구는 이차 및 3D 구조 개발을 강조했을 뿐만 아니라 nsp 코딩 유전자의 이차 구조에서 스템-루프 구조, 다중 분기 루프, 슈도노트 구조 및 벌지 구조 구성 요소와 같은 이차 구조의 모든 요소를 식별했다. 또한 3D 구조에서 언바운드 구조, 분기 구조, 듀플렉스 구조, 3방향 접합, 4방향 접합 등을 식별했다. 이 연구는 SARS-COV-2 게놈의 구조적 요소를 식별하는 데 도움이 될 것이다. 바인딩 포켓 식별에 대한 본 발명자들의 데이터는 RNA 표적 치료 개발을 위한 기본 플랫폼을 제공할 것이다. 그것은 SARS-CoV-2에 대한 구조 기반 약물 설계를 도울 약물 결합 가능한 게놈과 고유한 RNA 기반 치료 표적을 식별하는 데 도움이 될 것이다. 마지막으로, nsp 코딩 영역에서 식별된 결합 포켓은 여러 소분자 기반 리간드 발견과 바이러스에 대항하기 위해

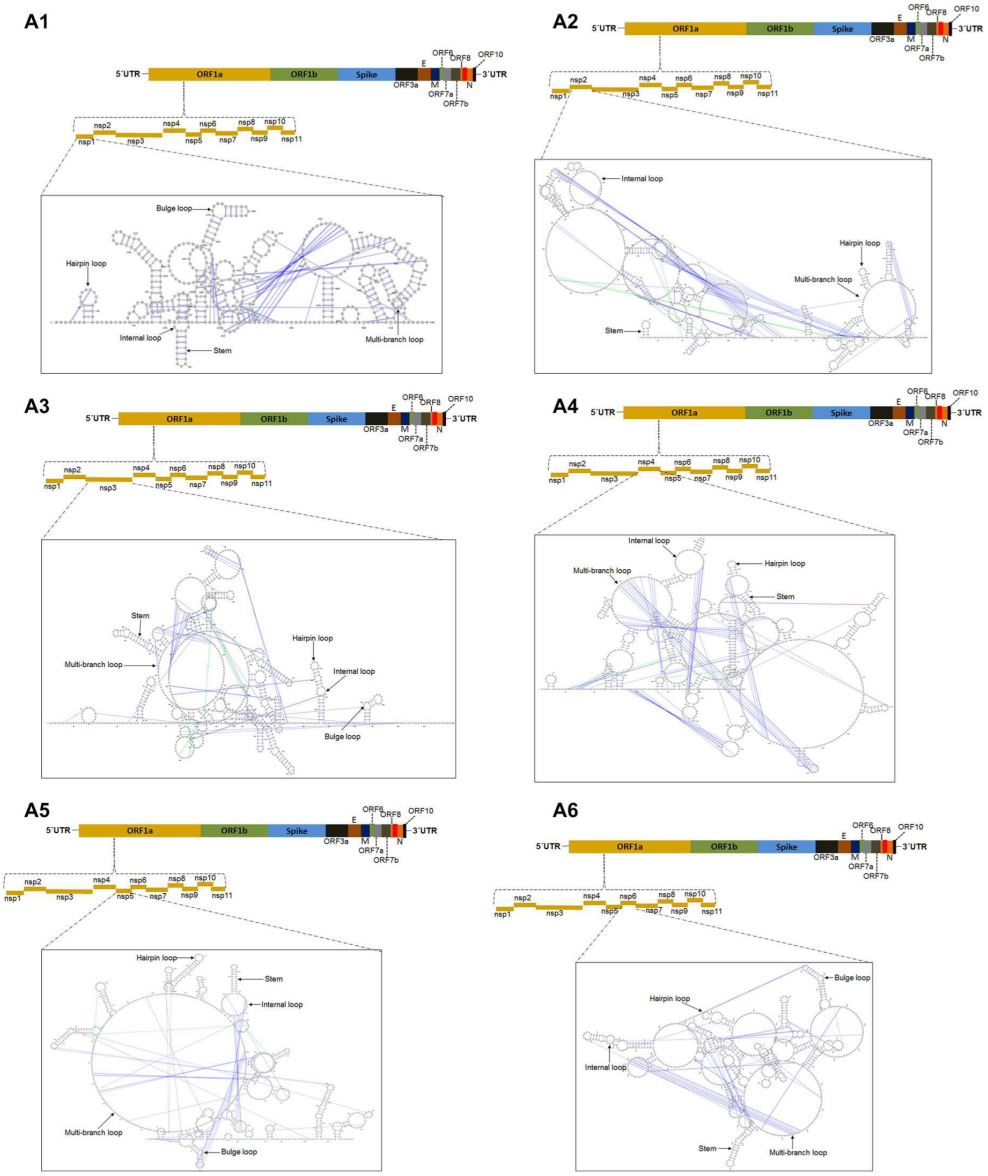
해당 게놈 영역을 표적으로 하는 안티센스 올리고뉴클레오티드 치료제 발견에 도움이 될 것이다.

도면

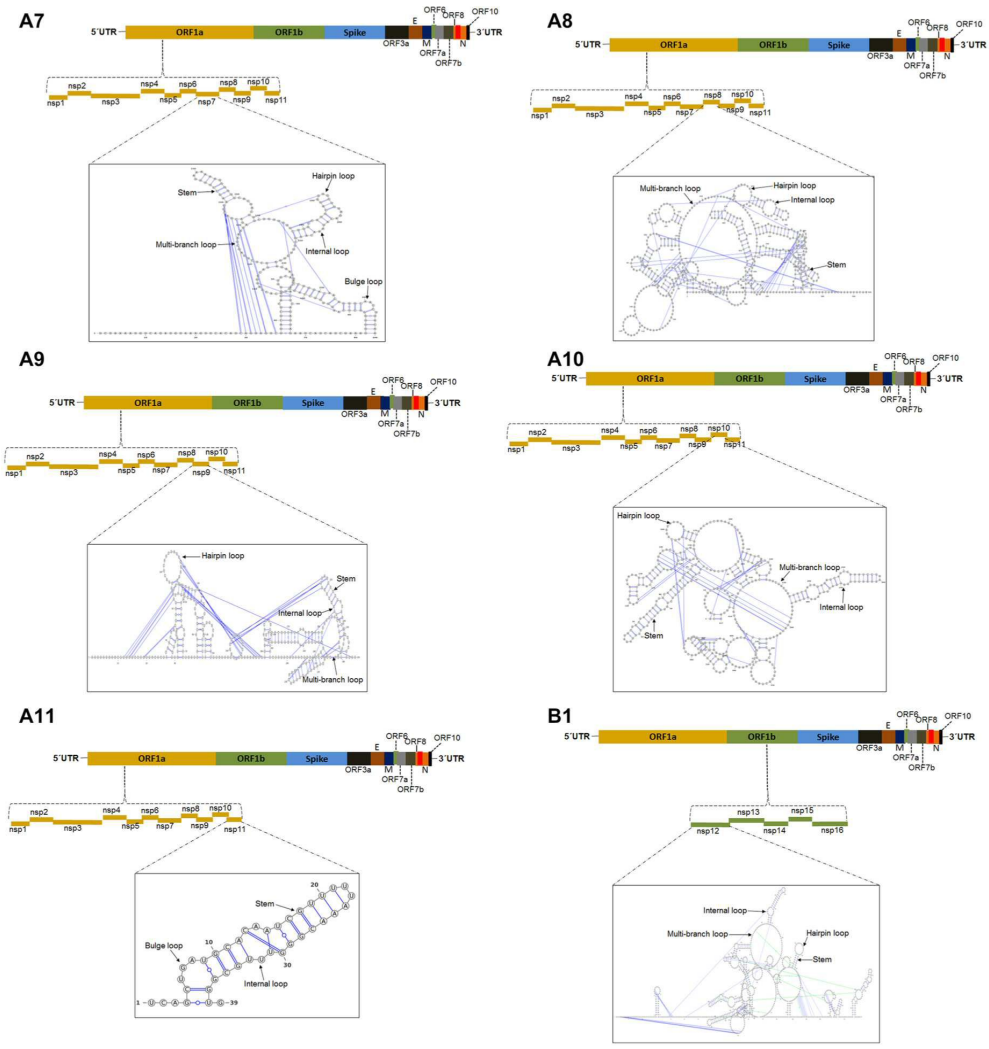
도면1



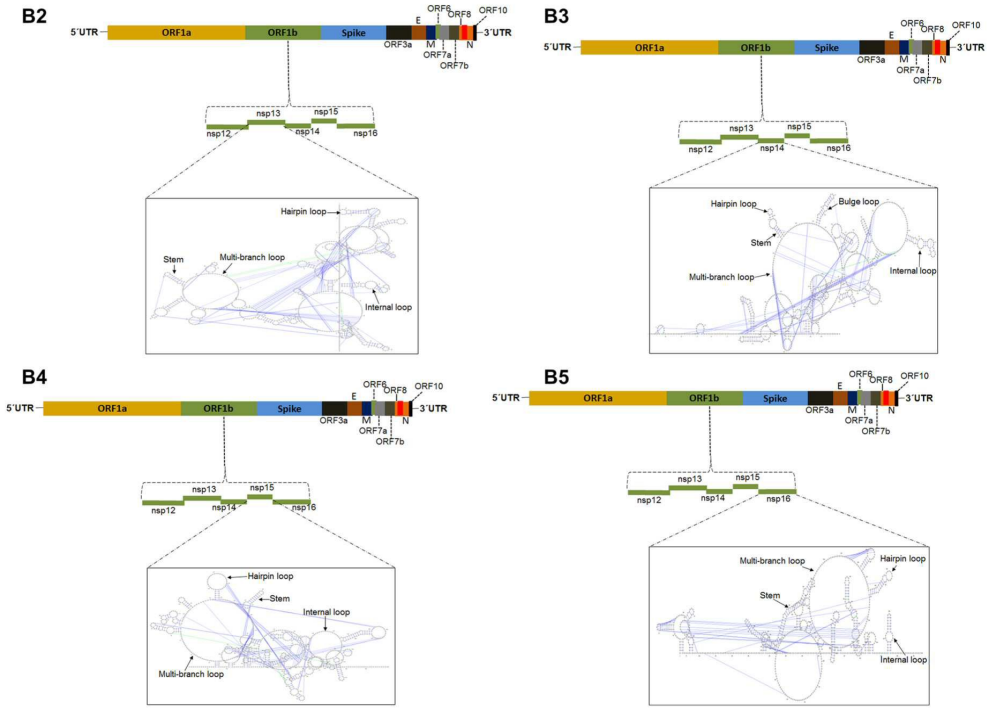
도면2


A1
5'UTR
ORF1a
ORF1b
Spike
ORF3a
E
M
ORF6
ORF8
ORF10
N
ORF7a
ORF7b
3'UTR
nsp1
nsp2
nsp3
nsp4
nsp5
nsp6
nsp7
nsp8
nsp9
nsp10
nsp11
Bulge loop
Hairpin loop
Internal loop
Stem
Multi-branch loop
A2
A3
A4
A5
A6

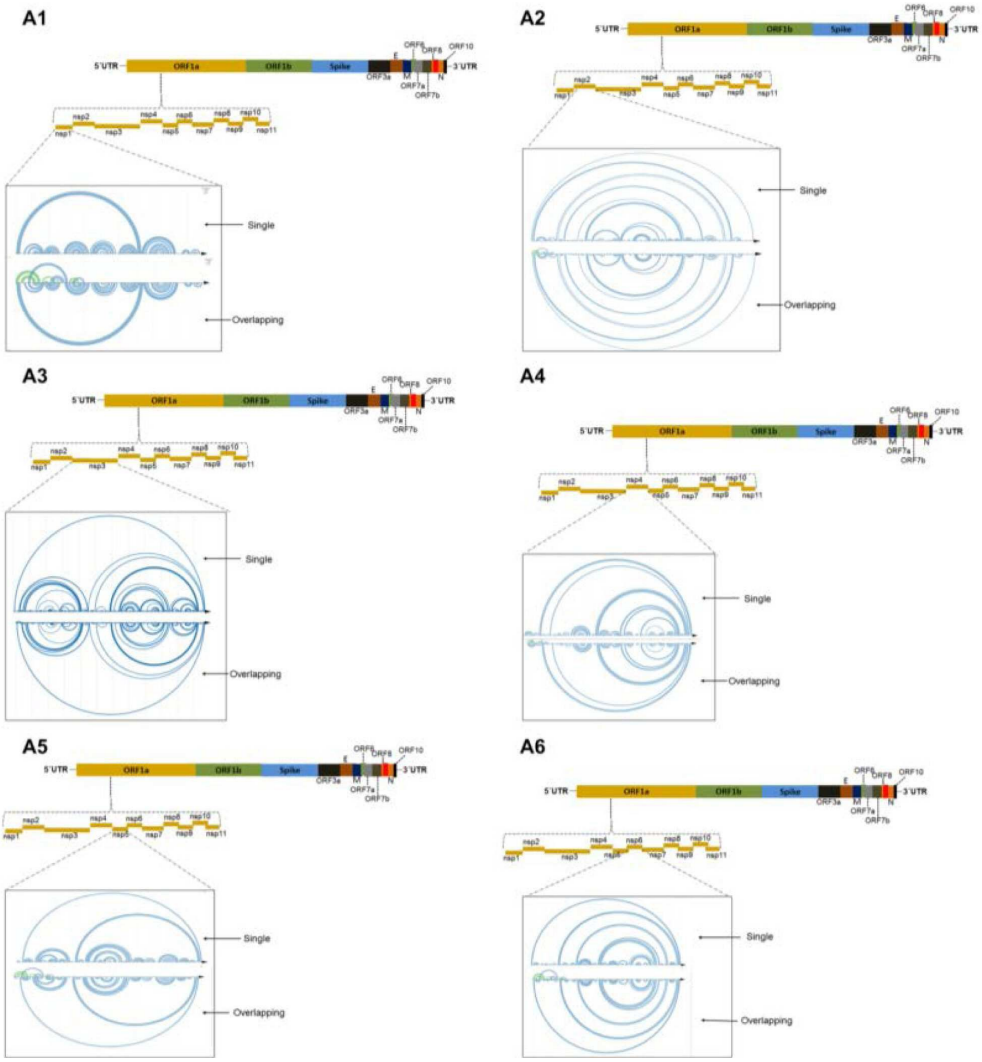
도면2b


A7
5'UTR
ORF1a
ORF1b
Spike
ORF3a
E
M
ORF6
ORF8
ORF10
ORF7a
ORF7b
N
3'UTR
nsp1
nsp2
nsp3
nsp4
nsp5
nsp6
nsp7
nsp8
nsp9
nsp10
nsp11
Hairpin loop
Stem
Multi-branch loop
Internal loop
Bulge loop
A8
Multi-branch loop
Hairpin loop
Internal loop
Stem
A9
Hairpin loop
Stem
Internal loop
Multi-branch loop
A10
Hairpin loop
Multi-branch loop
Internal loop
Stem
A11
Stem
Bulge loop
Internal loop
B1
nsp12
nsp13
nsp14
nsp15
nsp16
Internal loop
Multi-branch loop
Hairpin loop
Stem

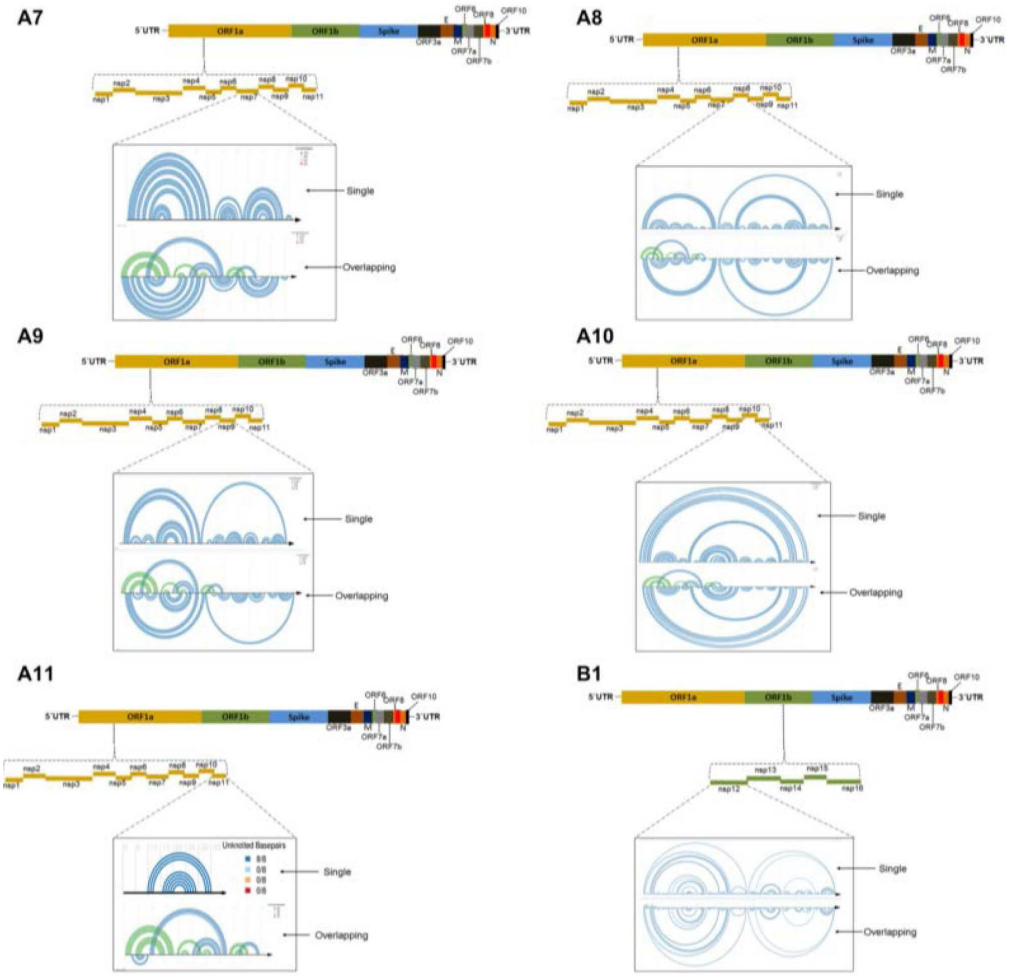
도면2c


B2
5'UTR
ORF1a
ORF1b
Spike
ORF3a
E
M
ORF6
ORF8
ORF10
ORF7a
ORF7b
N
3'UTR
nsp12
nsp13
nsp14
nsp15
nsp16
Hairpin loop
Stem
Multi-branch loop
Internal loop
B3
Bulge loop
B4
B5

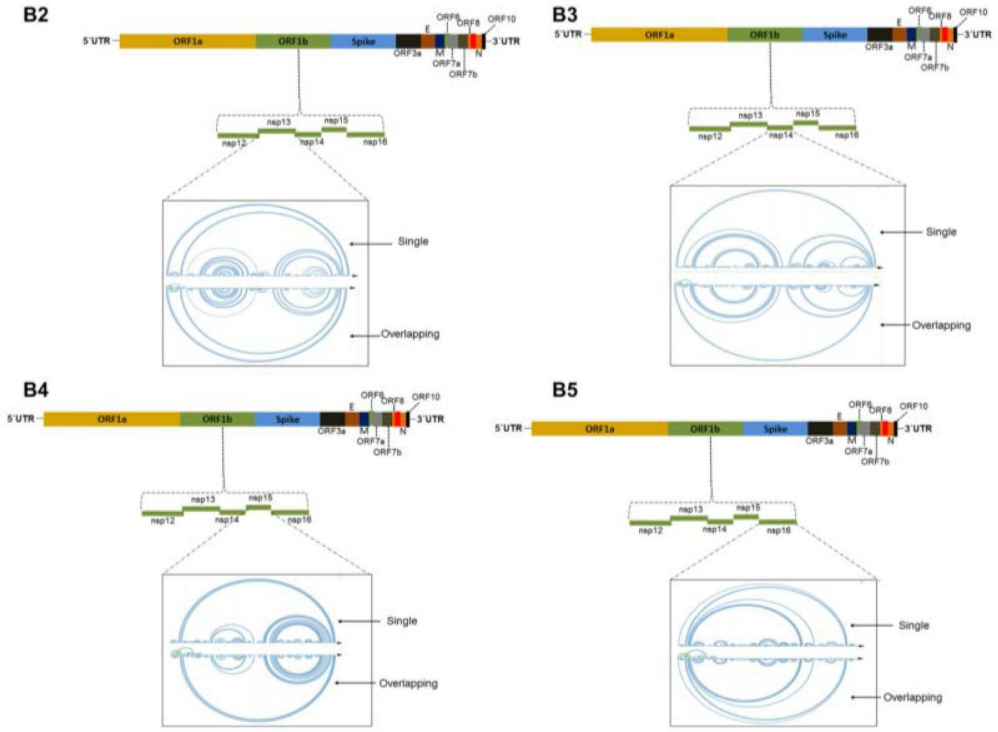
도면3a



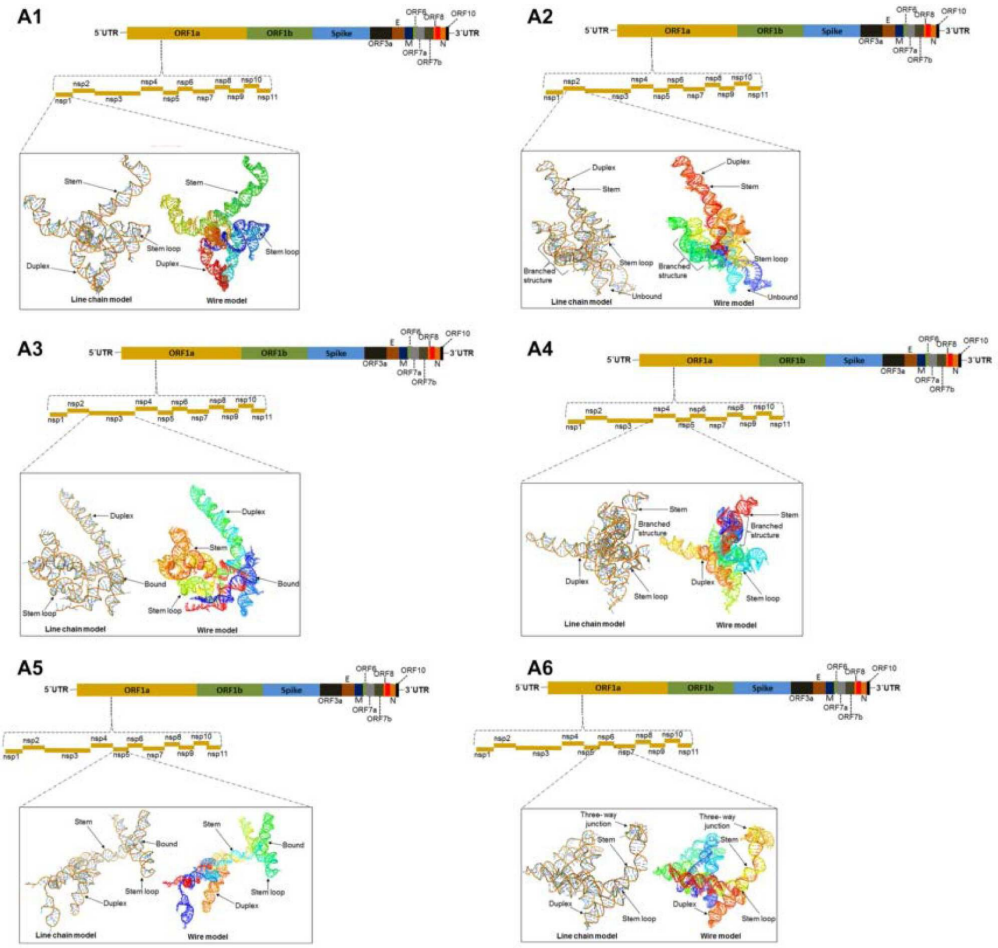
도면3b



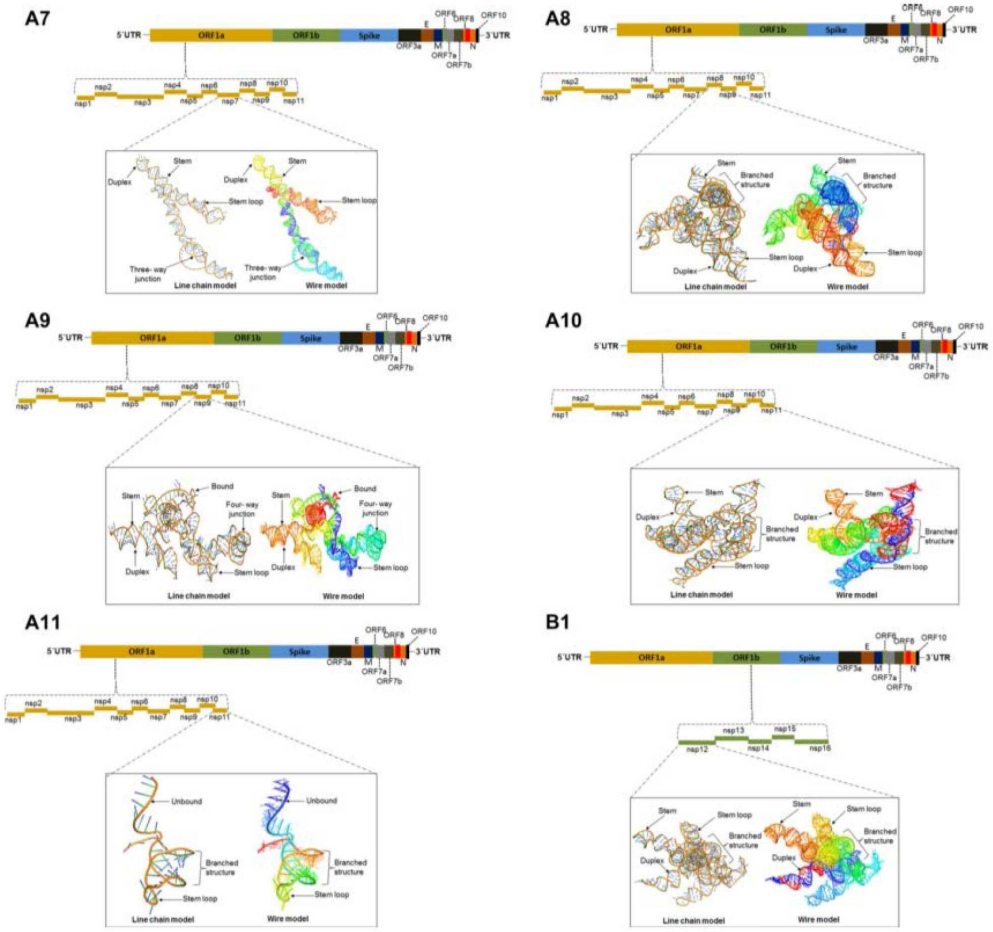
도면3c


B2
5'UTR
ORF1a
ORF1b
Spike
ORF3a
E
M
ORF6
ORF7a
ORF7b
ORF8
N
ORF10
3'UTR
nsp12
nsp13
nsp14
nsp15
nsp16
Single
Overlapping
B3
B4
B5

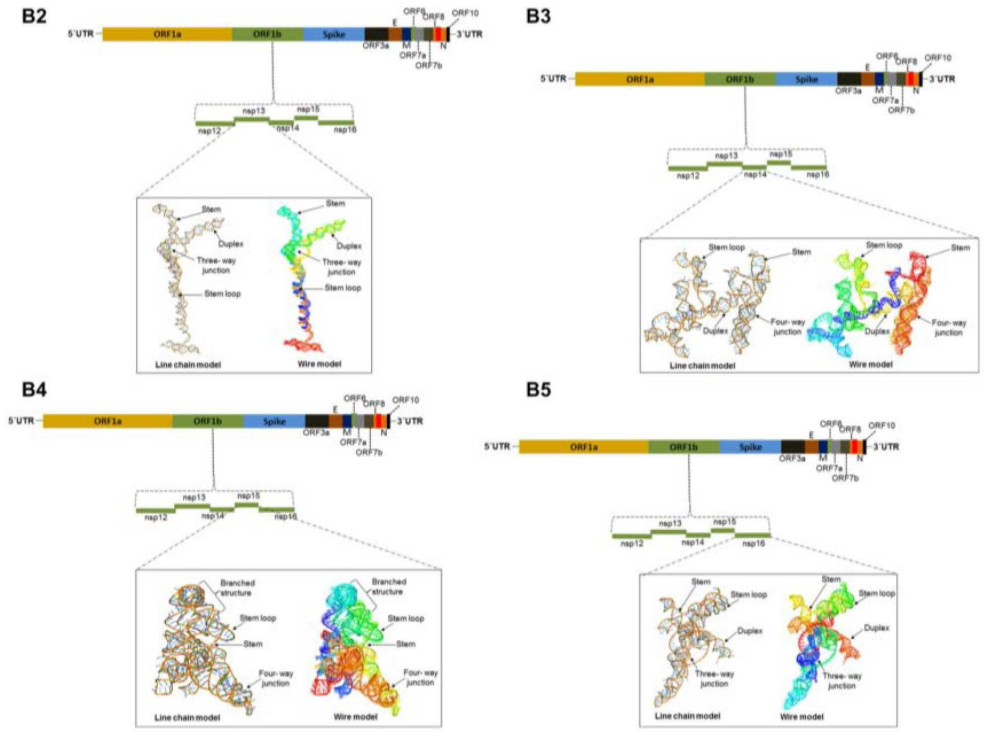
도면4a


A1
A2
A3
A4
A5
A6
5'UTR
ORF1a
ORF1b
Spike
3'UTR
Line chain model
Wire model

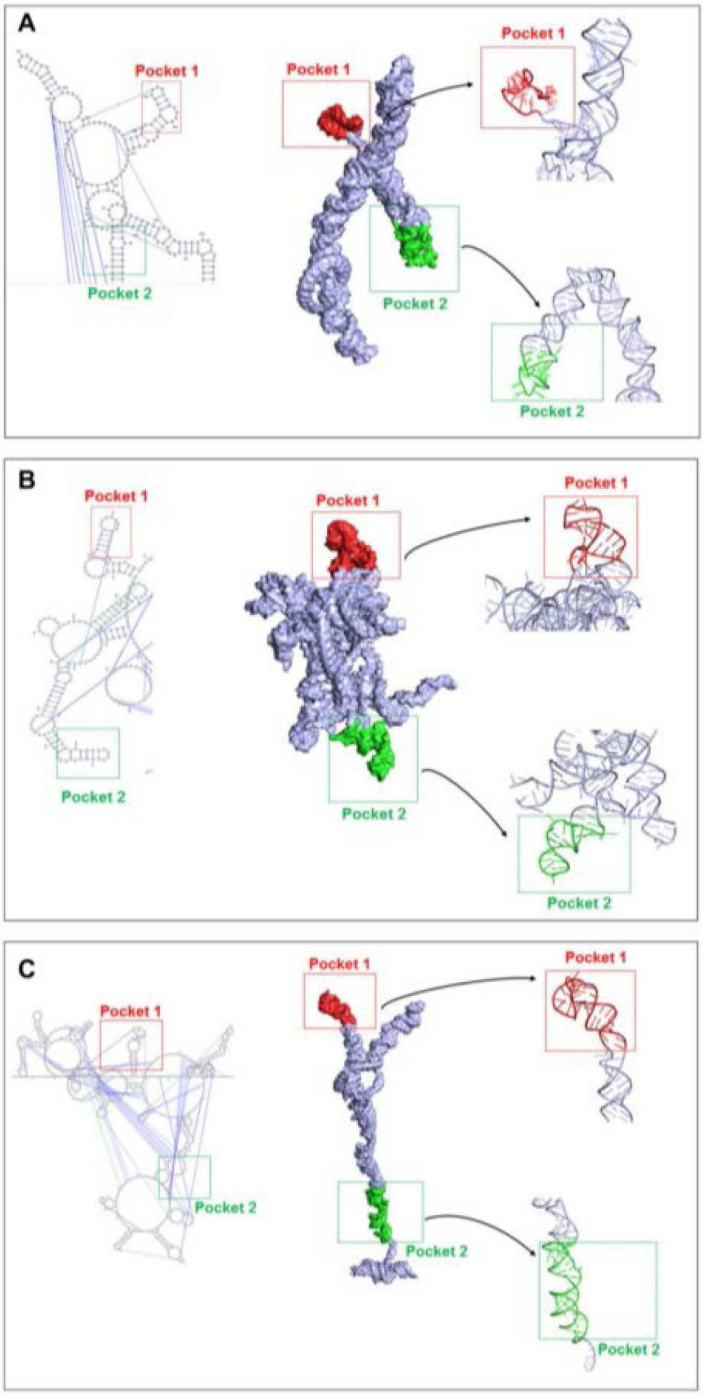
도면4b


A7
A8
A9
A10
A11
B1
5'UTR
ORF1a
ORF1b
Spike
3'UTR
Line chain model
Wire model

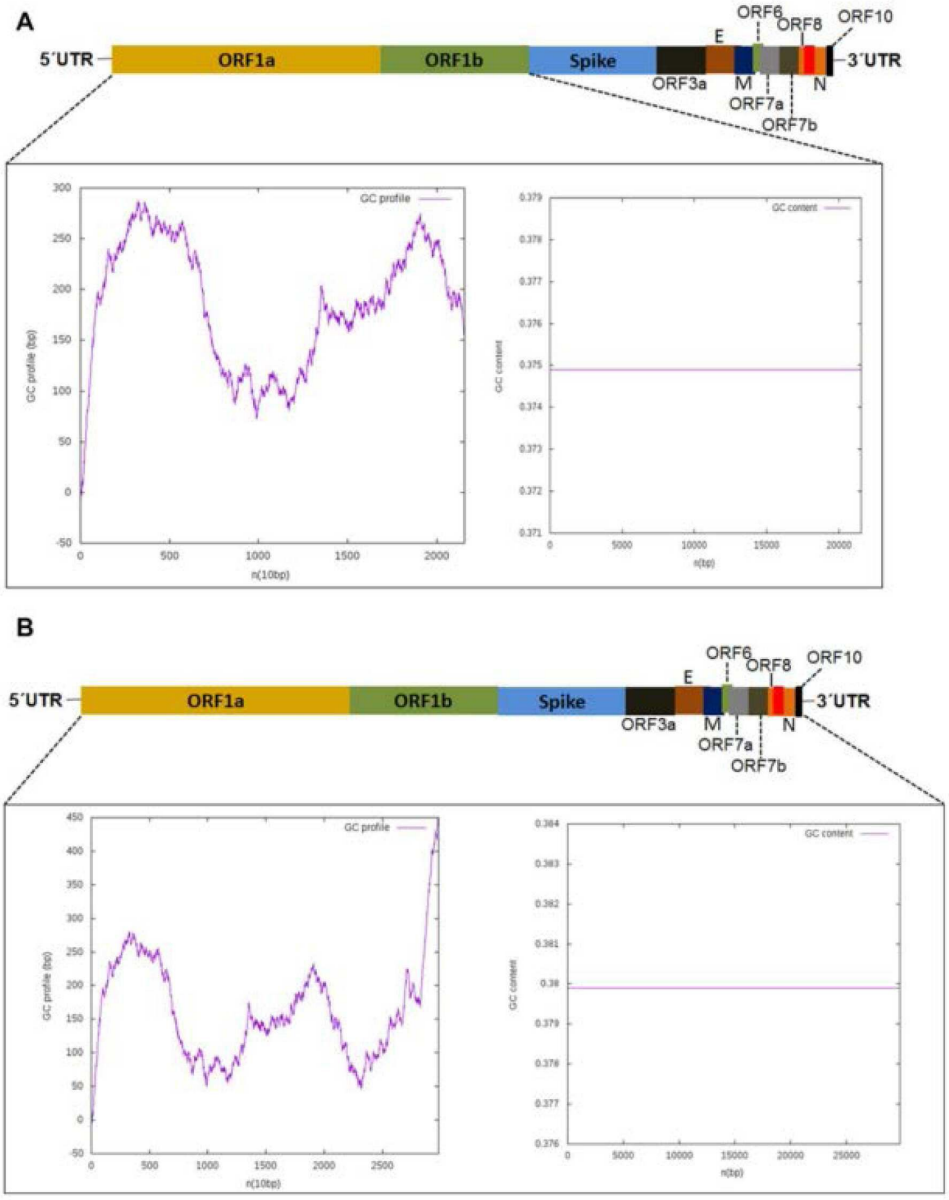
도면4c


B2
B3
B4
B5
5'UTR
ORF1a
ORF1b
Spike
3'UTR
Line chain model
Wire model

도면5


A
Pocket 1
Pocket 2
Pocket 1
Pocket 2
Pocket 1
Pocket 2
B
Pocket 1
Pocket 2
Pocket 1
Pocket 2
Pocket 1
Pocket 2
C
Pocket 1
Pocket 2
Pocket 1
Pocket 2
Pocket 1
Pocket 2

도면6


A
5´UTR
ORF1a
ORF1b
Spike
E
ORF6
ORF8
ORF10
3´UTR
ORF3a
M
ORF7a
N
ORF7b
GC profile
GC profile (bp)
n(10bp)
GC content
n(bp)
B
5´UTR
ORF1a
ORF1b
Spike
E
ORF6
ORF8
ORF10
3´UTR
ORF3a
M
ORF7a
N
ORF7b
GC profile
GC profile (bp)
n(10bp)
GC content
n(bp)