Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach

Ahmad Abu Turab Naqvi, Kisa Fatima, Taj Mohammad, Urooj Fatima, Indrakant K Singh, Archana Singh, Shaikh Muhammad Atif, Gururao Hariprasad, Gulam Mustafa Hasan, Md Imtaiyaz Hassan

Considerações a respeito do genoma, da estrutura, da evolução, da patogênese e do tratamento do SARS-CoV-2: abordagem genômico-estrutural

October, 2021
Outubro, 2021
FMRP - USP

Ana Luisa da Cunha Almeida (12536960)
Eduardo Barbosa de Lima (12565820)
Gabriel Silva Rio Lima Costa (12536869)
Isabela Nardo (10755662)
Covid-19 is an infectious disease caused by SARS-CoV-2 virus;
- Infected >219 mi and killed > 4,55 mi people worldwide;
- Coronaviruses (CoVs) are a group of enveloped viruses, having a positive single-stranded RNA.
- SARS-CoV-2 is more pathogenic than SARS-CoV (2002) and MERS-CoV (2013);
- The SARS-CoV-2 genome share about 82% sequence identity with SARS-CoV and MERS-CoV;
- SARS-CoV-2 contains four structural proteins, that include spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins.
• Three main enzymes: 3CLpro, PLpro and RdRp - potential drug targets;
• Catalytic pockets as important regions for the development of inhibitors;
• SARS-CoV-2 shares a high sequence identity in their RdRp and 3CLpro proteins to the SARS-CoV and MERS-CoV, but their RBD are significantly different.
The genome of SARS-CoV-2 is comprised of a single-stranded positive sense (+) mRNA.

Source: https://www.researchgate.net/figure/Relationship-between-the-negative-and-positive-RNA-strands-of-ISAV-Schematic_fig4_253698153

Eduardo Barbosa de Lima
The genetic makeup of SARS-COV 2 is composed of 13-15 ORFs (Open Reading Frames), containing ~30 000 nucleotides.

- 38% of GC content.
- 11 protein-coding genes, with 12 expressed proteins.

Observation: The ORFs represents the part of the reading frame that has the ability to be translated. So, it represents a valid frame between a start (AUG) and a stop codon (usually UAA, UAG, UGA).

ORF sample.

Source: https://en.wikipedia.org/wiki/Open_reading_frame#/media/File:Sampleorf.png
The ORFs are arranged as replicase (ORF-1a) and protease (ORF-1b) and major S, E, M and N proteins.

- ORF-1a = encodes polyprotein pp1a
- ORF-1b = encodes polyprotein pp1ab (ribosomal frameshift mechanism of gene 1b)

These polyproteins: further processed by proteinases = produce 16 proteins (NSPs, conserved in all CoV belonging to the same family).

The SARS-CoV-2 whole genome encodes about 7096 residues long polyprotein.
Non-structural proteins (NSPs)

In addition to the capsid-forming structural proteins, the viral genome encodes many NSPs that perform numerous roles in the replication and virus assembly processes.

These proteins participate in viral pathogenesis by:
1. Modulating early transcription regulation;
2. Helicase activity;
3. Immunomodulation;
4. Countering the antiviral response;
5. Other functions.
Non-structural proteins (NSPs)
Proteínas Não-Estruturais

Eduardo Barbosa de Lima

Table 1
List of non-structural proteins in SARS-CoV-2 and their molecular functions.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Range</th>
<th>Protein name and ID</th>
<th>Description</th>
<th>Proposed function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 1-180</td>
<td>Nsp1</td>
<td>YP_009725297.1</td>
<td>Nsp1 is the N-terminal product of the viral replicase</td>
<td>Leader protein host translation inhibitor. Mediates RNA replication and processing. Involved in mRNA degradation [41].</td>
</tr>
<tr>
<td>2. 181-818</td>
<td>Nsp2</td>
<td>YP_009725298.1</td>
<td>Nsp2 is a replicase product essential for proofreading viral replication.</td>
<td>Modulation of host cell survival signalling pathway by interacting with host PHB and PHB2 [42].</td>
</tr>
<tr>
<td>3. 819-2763</td>
<td>Nsp3</td>
<td>YP_009725299.1</td>
<td>Nsp3 is a papain-like protease contains several domains.</td>
<td>Functions as a protease to separate the translated polyprotein into its distinct proteins [43,44].</td>
</tr>
<tr>
<td>4. 2764-3263</td>
<td>Nsp4</td>
<td>YP_009725300.1</td>
<td>A membrane-spanning protein contains transmembrane domain 2 (TM2)</td>
<td>Believed to anchor the viral replication-transcription complex to modified ER membranes [45].</td>
</tr>
<tr>
<td>5. 3264-3569</td>
<td>Nsp5</td>
<td>YP_009725301.1</td>
<td>3C-like proteinase and main protease.</td>
<td>Involved in viral polypeptide processing during replication [46].</td>
</tr>
<tr>
<td>6. 3570-3859</td>
<td>Nsp6</td>
<td>YP_009725302.1</td>
<td>Putative transmembrane domain.</td>
<td>Plays a role in the initial induction of autophagosomes from host endoplasmic reticulum.</td>
</tr>
<tr>
<td>7. 3860-3942</td>
<td>Nsp7</td>
<td>YP_009725303.1</td>
<td>Nsp7 is an RNA-dependent RNA polymerase.</td>
<td>It forms a hexadecameric super-complex with nsp8 that adopts a hollow cylinder-like structure implicated replication [47,48].</td>
</tr>
<tr>
<td>8. 3943-4140</td>
<td>Nsp8</td>
<td>YP_009725304.1</td>
<td>Multimeric RNA polymerase; replicase.</td>
<td>It forms a hexadecamer super-complex with nsp7 that adopts a hollow cylinder-like structure implicated replication [47,48].</td>
</tr>
<tr>
<td>9. 4141-4253</td>
<td>Nsp9</td>
<td>YP_009725305.1</td>
<td>A single-stranded RNA-binding viral protein.</td>
<td>Participate in viral replication by acting as an ssRNA-binding protein [49].</td>
</tr>
<tr>
<td>10. 4254-4392</td>
<td>Nsp10</td>
<td>YP_009725306.1</td>
<td>Growth-factor-like protein contains two zinc-binding motifs.</td>
<td>In viral transcription by stimulating both nsp14 3'-5' exoribonuclease and nsp16 2'-O-methyltransferase activities. Therefore plays an essential role in viral mRNAs cap methylation [50].</td>
</tr>
<tr>
<td>11. 4393-5324</td>
<td>Nsp12</td>
<td>YP_009725307.1</td>
<td>RNA-dependent RNA polymerase (Pol/RdRp).</td>
<td>Responsible for replication and transcription of the viral RNA genome [51].</td>
</tr>
<tr>
<td>12. 5325-5925</td>
<td>Nsp13</td>
<td>YP_009725308.1</td>
<td>Zinc-binding domain, NTPase/helicase domain, RNA 5'-triphosphatase.</td>
<td>A helicase core domain that binds ATP. Zinc-binding domain is involved in replication and transcription [52,53].</td>
</tr>
<tr>
<td>13. 5926-6452</td>
<td>Nsp14</td>
<td>YP_009725309.1</td>
<td>Proofreading Exoribonuclease domain (ExoN/nsp14).</td>
<td>Exoribonuclease activity acting in a 3' to 5' direction and N7-guanine methyltransferase activity.</td>
</tr>
<tr>
<td>14. 6453-6798</td>
<td>Nsp15</td>
<td>YP_009725310.1</td>
<td>EndoRNase; nsp15A1 and nsp15B-NendoU</td>
<td>Mn(2+) dependent endoribonuclease activity.</td>
</tr>
<tr>
<td>15. 6799-7096</td>
<td>Nsp16</td>
<td>YP_009725311.1</td>
<td>2'-O-ribose methyltransferase.</td>
<td>Methyltransferase that mediates mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs [54].</td>
</tr>
<tr>
<td>16. 4393-4405</td>
<td>Nsp11</td>
<td>YP_009725312.1</td>
<td>Made of 13 amino acids (sdaqgsflngav) and identical to the first segment of Nsp12.</td>
<td>Unknown</td>
</tr>
</tbody>
</table>


Tendency of disorder: it’s a concept that reflects the amount of intrinsically unstructured regions in proteins.
- Proteins with a high disorder ratio lack a fixed or ordered three-dimensional structure.
- In SARS-Cov-2: some NSPs have a high disorder ratio (NSP-3, for example).
- Viral proteins with a high disorder ratio difficult the structure-based drug discovery.

The graph in the next slide shows the disordering tendency of each residue in SARS-Cov-2 polyprotein, considering a reference score (S)
- $S > 0.5 = \text{residues are intrinsically disordered}$;
- $0.2 < S < 0.5 = \text{residues are flexible}$.
Non-structural proteins (NSPs)

Proteínas Não-Estruturais

VLXT, VL3, VSL2 and IUPred are different types of predictors used on the analysis;
The interval of high scores 920/1020 shows the intrinsically unstructured regions of SARS-CoV-2 NSP-3.

Source: the article.
Coronaviruses are a large family of viruses that usually cause mild to moderate upper-respiratory tract illnesses.

Sometimes those viruses, commonly found on animals, jump to humans and can cause disease, as the SARS-CoV-2, virus responsible for COVID-19 disease.

Viruses in the same family share much of their genetic identity.

Proteins loaded and translated from their genetic material will also be similar.

So, with similar target proteins, will the therapeutic methods, as the use of drugs, be the same as well?

Viruses used in this comparative research:

- SARS-CoV-2
- SARS-CoV
- MERS-CoV
- BAT-CoV
What will be the focus of the drugs used?

The small machinery of proteins that contributes to the pathogenesis by helping the vírus entry into the host cell:

1. spike glycoprotein (S)
2. envelope protein (E)
3. membrane protein (M)
4. nucleoprotein (N)
5. and two isoforms of replicase polyprotein
1. SPIKE PROTEIN - main protein to be studied

- The S protein initiates the infection by sticking the virion to the host cell: the receptor-binding domain (RBD), in S1 subunit of SPIKE PROTEIN, binds strongly to human angiotensin-converting enzyme 2 (ACE2) receptors, generating the viral attachment, fusion and entry.

- The RBD of spike protein is the most variable part of the SARS-CoV-2:
  Of 6 residues of RBD of SARS-CoV2 that are critical for binding to ACE 2, 5 are different from SARS-CoV.

Comparison: aminoacid sequence of spike RBD protein
Fig. 6. Structural comparison of RBDs of S protein for all four strains. A. Superposed image of BAT-CoV S RBD protein (lime green) and SARS-CoV-2 (red) (RMSD: 2.3 Å). B. Surface representation of superimposed RBD of BAT-CoV and SARS-CoV-2. C. Superposed image of RBD of MERS-CoV S protein (light orange) and SARS-CoV-2 (light blue) (RMSD: 8.6 Å). D. Surface representation of the RBD of MERS-CoV and SARS-CoV-2. E. Superposed image of RBD of SARS-CoV S protein (warm pink) and SARS-CoV-2 (slate) (RMSD: 1.5 Å). F. Surface representation of the RBD of SARS-CoV and SARS-CoV-2.
2. **Envelope (E)** is only present in small quantities and most likely forms ion channels.  
   ● The sequences of E protein for all four strains are **highly conserved**.

3. **Matrix/Membrane (M)** is the most abundant structural protein of the virus. M proteins are responsible for membrane curvature of the viral envelope.  
   ● **Higher sequence conservation** among BAT-CoV, SARS-CoV, and SARS-CoV-2.
4. **Nucleocapsid (N)** binds to the viral RNA genome and ensures the maintenance of the RNA in a ‘beads-on-a-string’ conformation

- MSA profile of N protein from the BAT-CoV, SARS-CoV, and SARS-CoV-2 show **highly conserved** regions.
5. The two isoforms of replicase polyprotein: The coronavirus replicase is a polyprotein that mediate the synthesis of genome in the virus-infected cell.

- MSA of replicase polyprotein la shows a **conserved region** among all three domains except for the main protease where regions of least sequence conservation are presente.
Results of MSA (multiple sequence alignment):

- Membrane proteins (M)
- Envelope proteins (E)
- Nucleoproteins (N)
- The two replicase isoforms

Did not show any significant structural differences between the compared coronaviruses

As observed, the S protein (spike protein), more specifically the RBD domain, contains the most structurally diverse regions as compared to other target proteins.
CONCLUSION:

In spite of The SARS-CoV-2 genome share about 82% sequence identity with SARS-CoV and MERS-CoV and > 90% sequence identity for essential enzymes and structural proteins, those variations on the RBD domain of Spike protein change completely the mechanism of the pathology, making each of the coronaviruses unique.

And this uniqueness can only be treated with specific drugs.
Inflamatory Pathways and Cytokine Response
Via Inflamatória e Resposta de Citocinas

Innate immune response: first line of defense
- If dysregulated → excessive inflammation/death
- CoV infections → cytokine storm
  - intracellular pattern recognition receptors → activation of signalling cascades → release of cytokines and chemokines → recruitment of immune cells to the site of infection

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF8a and ORF9b</td>
<td>trigger cellular apoptosis</td>
</tr>
<tr>
<td>ORF8b</td>
<td>induces DNA synthesis and suppressing viral envelope protein expression</td>
</tr>
<tr>
<td>ORF7a</td>
<td>activates nuclear factor-κB (NF-κB)</td>
</tr>
<tr>
<td>ORF6</td>
<td>limits interferon production</td>
</tr>
<tr>
<td>ORF3a</td>
<td>induces necrotic cell death</td>
</tr>
<tr>
<td>ORF9b</td>
<td>alters interferon responses by promoting the degradation of mitochondrial antiviral signalling protein</td>
</tr>
</tbody>
</table>
Comparing the inflammatory pathways during SARS-CoV, MERS-CoV and SARS-CoV-2 infections:

**MERS-CoV**

- **N-protein** of MERS-CoV binds to the E3 ubiquitin ligase of triple motif protein 25 → prevention of the interaction between the triple motif protein 25 and retinoic acid-inducible gene I → **blocking the ubiquitination and activation of the retinoic acid-inducible gene I** mediated by triple motif protein 25 ultimately → **inhibition of type-I IFN production** (the N protein of CoV regulates the host's immune response against the virus).
Comparing the inflammatory pathways during SARS-CoV, MERS-CoV and SARS-CoV-2 infections:

**SARS-CoV and MERS-CoV**

- MERS-CoV and SARS-CoV: induce similar pathogen recognition receptor genes and pro-inflammatory cytokine genes related to interleukin 17 (IL-17) signalling by IL-17A and IL-17 F cytokines, but MERS-CoV infection downregulates the genes involved in antigen presentation pathway.
Comparing the inflammatory pathways during SARS-CoV, MERS-CoV and SARS-CoV-2 infections:

**SARS-CoV-2**
- Cytokine dysregulation similar to SARS-CoV and MERS-CoV
- Abnormally low levels of antiviral cytokines and [high levels](#) of [pro-inflammatory cytokines](#) including IL-1, IL-2, IL-6, IL8, IL-17, G-CSF, GM-CSF and chemokines such as IP-10 and MCP-1 → lung dysfunction by leading to the accumulation of immune cells within the lungs
- Increased concentrations of IL-6 are associated with increased viral load and the recruitment of inflammatory monocytes. **Suppressor of cytokine signalling 3 (SOCS3)** regulates the negative feedback mechanism of IL-6, which is found to be reduced in the patients with COVID-19.

*In summary, the three viral infections have marginal difference in terms of innate immune responses but greatly differ in terms of morbidity and mortality.*
Therapeutics Approaches

Divided into acting:
- on targets of the human immune system or human cells
- on the virus itself

Management of patients:
- ventilation assistance
- symptomatic approach

Prevention:
- hand washing
- avoiding touching of the face
- social distancing

Several drug molecules have been tried in the last couple of months as a treatment strategy.
List of drugs that have been found to have clinical effectiveness in COVID-19 therapy.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Drug</th>
<th>Target</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Remdesivir</td>
<td>Broad-spectrum antiviral inhibits RdRP of RNA viruses, including SARS-CoV and MERS-CoV</td>
<td>Adenosine analog, which incorporates into the nascent viral RNA chains during synthesis and causes premature termination.</td>
</tr>
<tr>
<td>2.</td>
<td>Chloroquine</td>
<td>Anti-malarial drug. Works at entry and post-entry stages of viral infection.</td>
<td>Increases endosomal pH required for virus/cell fusion. Interferes with the glycosylation of cellular receptors of SARS-CoV2.</td>
</tr>
<tr>
<td>3.</td>
<td>Hydroxychloroquine</td>
<td>JAK2 and FMS-like tyrosine kinase 3 (JAK) inhibitor</td>
<td>Inhibition of JAK2 inhibits phosphorylation of STAT 3 and 5, which prevents cell division and induces apoptosis.</td>
</tr>
<tr>
<td>4.</td>
<td>Lopinavir</td>
<td>Protease inhibitor have in vitro antiviral activity against SARS associated coronavirus</td>
<td>Inhibition of coronavirus main protease interferes in the processing of polypeptide translation products.</td>
</tr>
<tr>
<td>5.</td>
<td>Oseltamivir</td>
<td>Neuraminidase inhibitor</td>
<td>Inhibits the neuraminidase activity of the virus subsequently prevents viral replication.</td>
</tr>
<tr>
<td>6.</td>
<td>Favipiravir</td>
<td>RNA-dependent RNA polymerase</td>
<td>A guanine analog inhibits the RdRP activity of several RNA viruses (influenza, Ebola, Yellow fever and Chikungunya)</td>
</tr>
<tr>
<td>7.</td>
<td>Ribavirin</td>
<td>Nucleoside inhibitor An approved of HCV and RSV patients with SARS and MERS</td>
<td>A nucleoside inhibitor that interferes with viral RNA synthesis and miRNA capping.</td>
</tr>
<tr>
<td>8.</td>
<td>Gilead</td>
<td>RNA polymerase</td>
<td>Disrupts RNA polymerase activity causes premature termination of the elongating viral RNA strand</td>
</tr>
<tr>
<td>9.</td>
<td>Nafamostat</td>
<td>Serine proteases inhibitor</td>
<td>Prevents membrane fusion by reducing the release of cathepsin B.</td>
</tr>
<tr>
<td>11.</td>
<td>Baricitinib</td>
<td>Janus kinase (JAK) inhibitor</td>
<td>May block viral entry by inhibiting adaptor associated protein kinase 1 and cyclin G-associated kinase</td>
</tr>
<tr>
<td>12.</td>
<td>Tocilizumab (mAB)</td>
<td>IL-6 inhibitor</td>
<td>Inhibition of IL-6 may attenuate pulmonary inflammation by controlling cytokine storm.</td>
</tr>
<tr>
<td>13.</td>
<td>Anti TNF alpha agents</td>
<td>TNF alpha</td>
<td>TNF-α promotes the production of other chemokines and cytokines, controls endotoxin induced septic shock.</td>
</tr>
</tbody>
</table>
Remdesivir

- adenosine analog
- terminates the synthesis of viral RNA chains by incorporating in place of real nucleotide
- has been effective against single-stranded RNA viruses including MERS and SARS-CoV. Effectively inhibited SARS-CoV-2 \textit{in vitro}.

Chloroquine

- antimalarial
- antiviral effects by increasing endosomal and lysosomal pH causes an impaired release of the virus from the endosome or lysosome and thus recommended to handle severe COVID-19 patients
- Hydroxy-chloroquine, a less toxic derivative of chloroquine, also found is effective in inhibiting SARS-CoV-2 infection
Renin–angiotensin–aldosterone system (RAAS)

- Central to SARS-CoV-2 infection
- Angiotensinogen: converted into angiotensin I by plasma renin released by the liver
- Angiotensin I: converted to angiotensin II by the ACE found on the surface of vascular endothelial cells of lungs
  Angiotensin II acts on the AT1 receptor to cause vasoconstriction and also stimulates the secretion of the hormone aldosterone, which increases the reabsorption of sodium, thereby increasing blood pressure

- Angiotensin I and II are degraded by ACE2 to angiotensin (1–9) and angiotensin (1–7), respectively.
- Angiotensin (1–9) and angiotensin (1–7) act on the Mas receptor to effect vaso-dilation.
Renin-angiotensin-aldosterone system (RAAS)

SARS-CoV-2 infection requires the binding of the virus to the membrane-bound form of ACE2 and internalization of the S protein of the virus to the extracellular domain of ACE2, a membrane receptor.
Association between COVID-19 infection and Cardiovascular Diseases

- Many patients with hypertension or other cardiovascular diseases are routinely treated with RAAS blockers and statins.
- Clinical concerns remain whether these patients are at greater risk for SARS-CoV-2 infection due to enhanced ACE2 expression.
- More laboratory and clinical evidence are required to establish the roles of antihypertensive agents, ACE2 expression outcomes of COVID-19 in patients with cardiovascular disease.
Mesenchymal Stem Cell (MSC) Therapy

- **Promising option** (immunomodulatory function)
- Studies suggested that **intravenous transplantation of MSCs** into patients is effective in the treatment of COVID-19 with **lesser side effects**.
- The successful infusion of MSCs resulted in increased pulmonary function of the subjected patients to this therapy.

Convalescent Plasma

- Potential therapy for **severe** COVID-19 patients
- The convalescent plasma is retrieved from the **fully recovered patients** of viral disease and is **transfused in the infected** person as a treatment strategy
- During the time of the COVID-19 pandemic, the successful application of this therapy is effective in some of the patients
this study provides a possible hypothesis for the pathogenesis and transmission of the disease in humans

efforts in the short term should be focused on developing a vaccine or inhibitors that help to prevent the infection by targeting the major viral proteins such as S, E, M, N, proteins, RdRP and proteases

the three infections have marginal differences in terms of innate immune responses but greatly differ in terms of morbidity and mortality