Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Identification and classification of differentially expressed genes reveal potential molecular signature associated with SARS-CoV-2 infection in lung adenocarcinomia cells

Opeyemi S. Soremekun, Kehinde F. Omolabi, Mahmoud E.S. Soliman

Molecular Bio-computation and Drug Design Laboratory, School of Health Sciences, University of KwaZulu-Natal, Westville Campus, Durban, 4001, South Africa

ARTICLE INFO

Keywords: Differentially expressed genes; SARS-CoV-2; COVID-19; Enrichment analysis; RNAseq

ABSTRACT

Genomic techniques such as next-generation sequencing and microarrays have facilitated the identification and classification of molecular signatures inherent in cells upon viral infection, for possible therapeutic targets. Therefore, in this study, we performed a differential gene expression analysis, pathway enrichment analysis, and gene ontology on RNAseq data obtained from SARS-CoV-2 infected A549 cells. Differential expression analysis revealed that 753 genes were up-regulated while 746 down-regulated. SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B are the top 5 upregulated genes upon SARS-Cov-2 infection. Expectedly, these genes have been implicated in the immune response to viral assaults. In the Ontology of protein classification, a high percentage of the genes are classified as Gene-specific transcriptional regulator, metabolite interconversion enzyme, and Protein modifying enzymes. Twenty pathways with P-value lower than 0.05 were enriched in the up-regulated genes while 18 pathways are enriched in the down-regulated DEGs. The toll-like receptor signalling pathway is one of the major pathways enriched. This pathway plays an important role in the innate immune system by identifying the pathogen-associated molecular signature emanating from various microorganisms. Taken together, our results present a novel understanding of genes and corresponding pathways upon SARS-CoV-2 infection, and could facilitate the identification of novel therapeutic targets and biomarkers in the treatment of COVID-19.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been implicated as the causative agent of the recent global pandemic disease named Coronavirus disease 2019 (COVID-19). SARS-CoV-2 belongs to a larger family of Coronaviridae (CoVs) which causes gastrointestinal, respiratory, and neurological diseases to a varying degree of severity [1,2]. CoVs are non-segmented positive-sense RNA viruses that are zoonotically introduced into the human population [3]. They are divided into four groups: the α, β, γ, and δ-CoVs [4]. SARS-CoV-2, SARS-CoV and MERS-CoV are β-coronaviruses that cause lethal respiratory infections [5]. It has been estimated that deaths from the ravage of SARS-CoV-2 in a short space of time, has been more than that which resulted from SARS-CoV, and MERS-CoV combined [6] (Fig. 1). The method of transmission of SARS-CoV-2 is mainly human to human especially by close contact with an infected host’s aerosols and also the surfaces these have contaminated [7,8]. To date, there is no effective antiviral treatment for SARS-CoV-2 [9]. Instead, existing drugs are being repurposed for its treatment. Examples of these drugs among many others include serine protease inhibitors and a promising drug candidate, remdesivir, which has been found to have broad-spectrum activity against RNA viruses by hampering the mechanism of replication [10]. Other examples include favipiravir, hydroxychloroquine, co-administration of lopinavir and ritonavir, baricitinib, etc. [11–13]. These drugs and other supplemental therapies have functioned to resolve the infection to varying degrees of success [14–17]. Different approaches have been employed in the design of CoV vaccine, a majority of these strategies are developed to target the surface-exposed spike (S) glycoprotein, such as targeting the full S protein or S1-receptor-binding domain [43]. The similarity in the T-cell epitopes of MERS-CoV, SARS- and SARS-CoV-2 has been leveraged upon to examine the cross-reactivity of MERS-CoV and SARS-CoV vaccines [44]. A comparative study on the sequence length of S protein between SARS-CoV-2 and SARS-CoV showed that the variable amino acids were found on the S1

* Corresponding author.
E-mail address: soliman@ukzn.ac.za (M.E.S. Soliman).

https://doi.org/10.1016/j.imu.2020.100384
Received 26 April 2020; Received in revised form 20 June 2020; Accepted 21 June 2020
Available online 23 June 2020
© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
subunit of S protein, the major CoV vaccine target [44]. Furthermore, immune-informatics approaches have also been employed in the discovery of candidate epitopes as vaccine targets [45]. Vaccine candidates in clinical trials include INO-4800, mRNA-1273 and ChAdOx1 nCoV-19 [18–20].

Over the years, genome sequencing (whole and targeted) has evolved as a viable and efficient route for drug discovery. It has been employed in understanding the molecular basis of the pathogenesis of diseases, the mechanism of drug actions/resistance, and also for the identification of chemotherapeutic targets in organisms [21]. The complete genome of the Wuhan-Hu-1 coronavirus (WHCV) a strain of SARS-CoV-2, has been sequenced. From this, a clearer insight into the molecular basis of its virulence, its establishment and main target in human hosts is being propounded.

High throughput sequencing technologies have facilitated the increase in the number of genetic biomarkers, many techniques such as protein-arrays, mass-spectrometry, and gene-expression have been employed in the discovery of genomic biomarkers. These techniques encourage the analysis of expressions and also enable the ability to determine the activity of these genes in different conditions. In this study, the RNAseq dataset of mock-treated A549 (MT_A549) cells and SARS-CoV-2 infected A549 (SI_A549) cells retrieved from the Gene expression Omnibus (GEO) database were analysed to identify differentially expressed genes between MT_A549 cells and SI_A549 cells. Furthermore, the objective of this study is to identify the dysregulated pathways between these two conditions.

2. Materials and methods

2.1. Data retrieval, processing, and differential expression analysis

We downloaded RNAseq datasets of MT_A549 cells (SRR11412249) and SI_A549 cells (SRR11412251) from the study with accession number GSE147507 [22]. In this study, transformed lung alveolar (A549) cells and human lung epithelium cells were mock treated with SARS-CoV-2.

Fig. 1. COVID-19 infection rate and diversity based on geographical location.
Fig. 2. Top 10 most up-regulated DEGs (red arrow) and 10 most down-regulated DEGs (green arrow) in the samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
Fig. 3. Functional Classification of DEGs in MT_A549 cells and SI_A549 cells. (A) Molecular function ontology of overexpressed genes (red) and down regulated genes (black). (B) Biological process ontology of overexpressed genes (red) and down regulated genes (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
5

Interacting Genes (STRING) version 11.0, covering 24584628 proteins from 5090 species was employed to determine Protein-protein Interaction (PPIs). The interactions retrieved in STRING had a confidence score. During the prediction, we used experimentally validated interactions possessing a confidence score 0.4 and maximum additional interactor set at 0 to develop the protein-protein interaction network with the aid of Cytoscape software version 3.8.0 [46]. The String Enrichment plugin in Cytoscape was used to retrieve the functional enrichment. The statistically significant difference was set at P < 0.05.

2.3. 3D structure modeling and binding site characterisation

Homology modeling is a technique used in computational biology for the determination of protein 3D structure using its amino acid sequence as building blocks [31]. It is adjudged the most accurate and effective method of computational structure prediction [31]. The protein sequence of OAS2, SYCP2L, GNB3, and ALOX12B with accession number P29728, QST4T6, P16520, and O75342 respectively, were retrieved from UniProt [32]. These amino acids sequences were used as input in the online template-based homology modeling server; SwissModel [33]. Validation of the built models was carried out with the aid of RAMPAGE [34], the ProSA web server [35], and Verify-3D [36]. Furthermore, after the structures were modeled, the binding sites needed to be identified; we therefore used metaPocket [37] to identify binding sites for potential drug binding sites. metaPocket is a webserver that uses different predictors (LIGSITE, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity, and POCASA) to determine a protein’s binding site.

3. Results and discussion

3.1. Differential gene expression between MT_A549 cells and the SI_A549 cells

The DESeq2 package was employed to determine DEGs in the MT_A549 cells and the SI_A549 cells. At the Trimming stage, 7,461,512 (94.42%) reads of the (infected sample) passed the trim filter, 469, 643 (5.58%) were removed and 7,461,512 (99.11%) were aligned. For the mock-treated samples, 2440, 387 (89.20%) of the total reads passed the trim filter; out of this, 99.12% were aligned. At a P-value of 0.05 and 1% FDR (Q < 0.01), a total of 17179 transcripts and genes were differentially expressed between MT_A549 cells and the SI_A549 cells. Out of the 17179 expressed genes, 753 were up-regulated while 746 were down-regulated. However, the analysed result indicated that the level and range of expression of the down-regulated genes and transcripts were lower than that of the overexpressed genes. The top 5 overexpressed genes are SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B while the top 5 down-regulated genes are GNB3, TMED10P1, ALOX12B, MAFG-AS1 and SAYSD1 (Fig. 2). SNORA81, and SNORD35B belong to a highly expressed group of non-coding RNAs called Small nucleolar RNAs and C/D box snoRNAs, respectively. These proteins have been implicated in the regulation of pre-mRNA splicing, polyadenylation of genes,

| Protein Class | Up-regulated genes | Down-regulated genes |
|---------------|--------------------|----------------------|
| Extracellular matrix protein | 2 | 3 |
| Cytoskeletal protein | 16 | 12 |
| Transporter | 20 | 29 |
| Scaffold/adaptor protein | 10 | 13 |
| Nucleic acid binding protein | 30 | |
| Intercellular signal molecule | 12 | 16 |
| Protein-binding activity modifier | 18 | 19 |
| Calcium-binding protein | 5 | 1 |
| Gene-specific transcriptional regulator | 45 | 20 |
| Defense/immunity protein | 2 | 5 |
| Translational protein | 4 | – |
| Metabolite interconversion enzyme | 34 | 26 |
| Protein modifying enzyme | 40 | 23 |
| Chromatin/chromatin-binding, or -regulatory protein | 8 | 2 |
| Transfer/carrier protein | 1 | – |
| Membrane traffic protein | 11 | 7 |
| Chaperone | 1 | |
| Transmembrane signal receptor | 11 | 15 |
| Cell adhesion | – | 6 |
| Structural protein | – | 3 |
| Cell junction | – | 7 |

Fig. 4. Pathway enrichment analysis of the overexpressed DEGs using WebGestalt.
the replication of RNA viruses and DNA viruses [38] (see Fig. 3).

suggested that the SNORDs, and not the host genes, were responsible for identified as major factors required for viral infectivity. This study indicated that the SNORDs, and not the host genes, were responsible for the replication of RNA viruses and DNA viruses [38] (see Fig. 3).

The second most upregulated gene OAS2 encodes the enzyme 2’-5’-oligoadenylate synthases 2; these enzymes are reputed for their ability to enhance intracellular antiviral mechanisms. After induction by type 1 interferons, 2’-5’-oligoadenylate synthases 2 facilitates the polymerization of ATP to 2’-5’-linked adenosine oligomers (2-5As). 2-5As subsequently activates the RNase L degradative pathway responsible for cleaving viral RNA and mitigates against infection [39].

3.2. Classification and overrepresentation analysis

To explore the DEGs at an in-depth functional level, the predicted DEGs were mapped onto the PANTHER database. Over-represented GO terms were identified in three ontologies. In the ontology of molecular function, the result indicated that the up-regulated genes were involved in majorly 6 molecular functions with the majority of the genes involved in binding (149 genes), catalytic activity (116 genes), transcription regulatory activity (27 genes), etc. Similarly, some down-regulated genes were also predicted to be implicated in catalytic activity, binding, and transcription regulatory activity. The abundance of genes in these processes is due to the fact that they contain genes that are necessary for basic life functionality. Granulocyte-macrophage colony stimulating factor is one of the genes involved in cellular process. This gene is a monomeric glycoprotein that serves as a cytokine that facilitates the development of the immune system and promotes defense against infections [40]. The cellular processes include cell communication, cell cycle, gene expression, protein targeting, metabolism, etc. Some genes classified under metabolic processes are involved in energy-producing pathways. These genes include PDK4, NLK, BATF2, ATP6AP1L, GRK2, etc. Biological regulation describes genes that are involved in metabolism, apoptotic regulation, homeostasis, catalytic activity, and translation. The most abundant protein classification of the DEGs are Gene-specific transcriptional regulator, protein modifying enzyme, and metabolite interconversion enzyme (Table 1).

3.3. Signalling pathway analysis

To identify the dysregulated pathways on infection with SARS-CoV-2, pathway enrichment analysis was carried out with the aid of WebGestalt. 20 pathways with a P-value lower than 0.05 were enriched in the up-regulated genes while 18 pathways were enriched in the down-regulated DEGs. The Toll-like receptor (TLR) signalling pathway is one of the major pathways enriched among the DEGs (Fig. 4). TLRs play a very crucial function in the innate immune system by identifying the pathogen-associated molecular signature emanating from various microorganism [41]. Genes implicated in this pathway include C-C motif chemokine ligand 5, interferon regulatory factor 7, signal transducer and activator of transcription 1, C-X-C motif chemokine ligand 8, toll-like receptor 1 and Fos proto-oncogene (Fig. 4B). Furthermore, other pathways enriched were Leukocyte transendothelial migration (LEM), and Cytokine-cytokine receptor interaction. LEM movement from the blood vessel plays a crucial role in inflammation and immune surveillance. Leukocytes interact with endothelial cell adhesion molecules (CAM), and subsequently move across the vascular endothelium [42]. Some of the genes enriched in the cytokine-cytokine receptor interaction include C-C motif chemokine ligand 12, thrombopoietin, TNF superfamily member 8, X-C motif chemokine ligand 2, and interleukin 20 receptor subunit alpha (Fig. 5B).

3.4. Protein-protein interaction network analysis

PPI network analysis is crucial in the understanding of the biological responses of cells to infection. Based on the interactions result from STRING, we built a protein-protein interaction network using Cytoscape [46] (Fig. 6). A total of 810 nodes and 2300 edges were identified in the network. The STRING online database was used in the analysis of the predicted DEGs. The top 15 hub genes with the highest node degrees were ISG15, STAT1, HERC6, IRF7, HERCS, IFIH1, DDX58, GBP1, IRF9, OAS1, MX1, OAS2, DDX60, OAS1, and IFIT1 (Table 2). The accuracy of the network was ascertained by the clustering coefficient (0.247), network density (0.247), and network centralization (0.063). These top 15 hub genes were investigated for betweenness centrality, MNC.
centrality, stress centrality, and degree centrality. ISG15 showed the highest node degree of 52 and betweenness centrality of 0.0639. OAS2 is strongly correlated with MX1; this connectivity could probably be due to the function of MX1 as protective genes against influenza virus infection [43] by interrupting viral ribonucleoprotein complex assembly [44].

3.5. Gene-products structure modeling

Homology modeling plays a crucial role in facilitating the drug discovery and development pipeline. We selected two representative genes from the up-regulated and down-regulated genes to be modeled. The protein sequence of OAS2, SYCP2L, GNB3, and ALOX12B with accession numbers P29728, Q5T4T6, P16520, and O75342 respectively, were retrieved from UniProt [32]. OAS2 has 719 amino acids; this sequence was used to build a model for OAS2, based on the in-built alignment algorithm in SwissModel [33]. 2'-5'-oligoadenylate synthase 1 (4rwn) was used as template; there was 52.33% sequence identity between OAS2 and 2'-5'-oligoadenylate synthase 1. Similarly, this process was repeated for SYCP2L, GNB3, and ALOX12B. Synaptonemal complex

| Genes | Betweenness Centrality | Degree | Number Of Undirected Edges | Closeness Centrality |
|-------|------------------------|--------|----------------------------|----------------------|
| ISG15 | 0.0639                 | 52     | 52                         | 0.2075               |
| STAT1 | 0.0530                 | 50     | 50                         | 0.2053               |
| HERC6 | 0.0068                 | 45     | 45                         | 0.1841               |
| IFR7  | 0.0114                 | 45     | 45                         | 0.1873               |
| HERC5 | 0.0081                 | 45     | 45                         | 0.1843               |
| IFIH1 | 0.0057                 | 41     | 41                         | 0.1862               |
| DDX58 | 0.0135                 | 41     | 41                         | 0.1862               |
| GBP1  | 0.0051                 | 40     | 40                         | 0.1853               |
| IFR9  | 0.0110                 | 40     | 40                         | 0.1875               |
| OASL  | 0.0073                 | 40     | 40                         | 0.1865               |
| MX1   | 0.0020                 | 40     | 40                         | 0.1857               |
| OAS2  | 0.0022                 | 39     | 39                         | 0.1852               |
| DDX60 | 0.0041                 | 37     | 37                         | 0.1832               |
| OAS1  | 0.0013                 | 37     | 37                         | 0.1843               |
| IFIT1 | 9.0918E-4              | 37     | 37                         | 0.1839               |

Fig. 6. Network analysis of the Differentially Expressed Genes based on betweenness centrality.
protein (5iwz) was used as the template to model SYCP2L; there was a 40.21% similarity between Synaptonemal complex protein and SYCP2L. An 83.24% sequence identity was found between Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 and GNB3. Lastly, in modeling ALOX12B, Arachidonate 15- lipooxygnase B was used as a template; these two proteins have 51.04% similarity (Fig. 7).

The quality of the selected model was validated by ProSA (z-score), Ramachandran plots, and Verify-3D. According to the RAMPAGE algorithm, 97.1% of OAS2 residues are found in the favoured region, 1.7% are in the allowed regions while 1.2% are located in the outlier region. 97.1%, 94.7%, and 94.1% amino acid residues of SYCP2L, GNB3, and ALOX12B respectively, were found in the favoured region (Table 3). Validation by ProSA revealed that the model structures are of good quality judging by the values of the z-score (Table 3). Furthermore, the ProSA plots indicate that all of the modeled structures fall between the range of scores characteristic of proteins of similar size. Assessment of the model by Verify-3D showed that for OAS2, the profile score was 96.56%; this indicates that 96.56% of the residues have an average 3D-1D score of >0.2 (Fig. 8). The 3D-1D score >0.2 of SYCP2L, GNB3, and ALOX12B are 93.75, 88.20, and 93.71, respectively. Taken together, results obtained from these three servers suggest that the model OAS2, SYCP2L, GNB3 and ALOX12B structures are of high quality and could be used for further study.

### Table 3

Assessment and validation of OAS2, SYCP2L, GNB3, and ALOX12B using RAMPAGE and ProSA.

| Genes  | Favoured Region (%) | Allowed Region (%) | Outlier Region (%) | ProSA z-score |
|--------|---------------------|--------------------|--------------------|---------------|
| OAS2   | 97.1                | 1.7                | 1.2                | -9.86         |
| SYCP2L | 97.1                | 2.1                | 0.8                | -8.04         |
| GNB3   | 94.7                | 3.6                | 1.8                | -7.34         |
| ALOX12B| 94.1                | 4.6                | 1.3                | -11.61        |

3.5.1. Binding site characterization and potential for drug targeting

The function of a protein is dependent on its interaction with other biomolecules; these biomolecules could be called ligands [45]. The interaction between these ligands and proteins occur at a specific region of the protein called ligand binding sites; binding is facilitated by amino acid residues located in this ligand binding site [45]. Ligand binding sites have gained remarkable attention in drug-target interaction, molecular docking, drug design, etc. [46]. Hence, the identification of ligand binding sites aids in the understanding of the mechanism of intermolecular interaction between ligand and protein and hence its implication in the pathogenesis or modulation of disease condition [47]. We used metaPocket, which is a traditional machine learning-based ligand binding site server. metaPocket uses different predictors (LIGSITE, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity, and POCASA) to determine a protein’s binding site. Binding site
prediction using metaPocket based on z-score and clustering highlighted two potential binding sites in the proteins. Fig. 9 highlights the two binding sites of OAS2, SYCP2, GNB3, and ALOX12B. Having highlighted these sites, they could be leveraged in the drug design and development pipeline. Most importantly, structure-based virtual screening (SBVS), SBVS is a method used in screening small molecule libraries or database for novel bioactive compounds against a certain protein [48]. This method makes use of the 3-dimensional structure, obtained from X-ray crystallography, Nuclear Magnetic Resonance, or homology modeling to dock a collection of small molecules into the binding pocket of proteins and subsequently select a subset of these compounds based on their binding pose [48]. Different online servers could be used in SBVS; for instance, when GNB3 was queried on the online Drug-Gene Interaction Database (DGIdb 3.0) [49], Lovastatin, Simvastatin, Olanzapine, Cerivastatin, and Hydrochlorothiazide were returned as potential drugs that could be used to target GNB3.

4. Conclusion

This research presents a differentially expressed gene analysis of RNAseq data of SARS-CoV-2 infected A549 cells. This expression profile was retrieved from Gene Expression Omnibus. We performed differential gene expression within the scope of gene sets and pathway analysis to explore genes and molecular-signature pathways that may be peculiar to Covid-19 infection. SNORA81, OAS2, SYCP2, LOC100506985, and SNORD35B are the top upregulated genes upon infection. To explore the DEGs at an in-depth functional level, the predicted DEGs were mapped onto the PANTHER database. Over-represented GO terms were identified in three ontologies. PPI network analysis is crucial in the understanding of the biological responses of cells to infection. The STRING online database was used in the analysis of the predicted DEGs. The top 20 overexpressed genes were investigated for betweenness centrality, MNC centrality, stress centrality, and degree centrality.

Ethical statement

This research did not involve animal model, hence ethical statement is not required.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence
Acknowledgement

The authors acknowledge the College of Health Sciences, UKZN, for their financial and infrastructural support, and at the same time thank the Centre for High Performance Computing (CHPC, www.chpc.ac.za), Cape Town, for computational resources.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2020.100384.

References

[1] Morfopoulou S, Brown JR, Davies EG, Anderson G, Virasami A, Qasim W, Breuer J. Human coronavirus OC43 associated with fatal encephalitis. N Engl J Med 2016,
Andrews S. FastQC 1.1 what is FastQC 2. Basic operations 2.1 opening a sequence file. Retrieved from, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; 2010.

St-Jean JR, Jacomy H, Desforges M, Vabret A, Freymuth F, Talbot PJ. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. J Virol 2004;78(10):5233–42. https://doi.org/10.1128/JVI.00863-04.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Li H, Durbin R. Fast and accurate read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp684.

Anders S, Pyl PT, Huber W. HTSeq–A Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31(12):166–9. https://doi.org/10.1093/bioinformatics/btu383.

Gene T, Consortium O. The gene ontology Resource: 20 years and still GOing strong. Vol. 47; 2013. p. 330–4. https://doi.org/10.1093/g knitting of functional gene expression patterns. Bioinformatics 2011;27(7):941–7. https://doi.org/10.1093/bioinformatics/btr088.

Mahase E. Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate, 641; 2020. p. 2020. https://doi.org/10.1136/bmj.m641.

Liao Y, Wang J, Jaehng EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res 2019;47(W1):W199–205. https://doi.org/10.1093/nar/gkz401.

Panda SK, Saxena S, Guruprasad L. Homology modeling, docking and structure-based virtual screening for new inhibitor identification of Klebsiella pneumoniae heparitinase-III. J Biomol Struct Dyn 2020;38(7):1887–902. https://doi.org/10.1080/07391102.2019.1624296.

Bateman A, Martin MJ, O’Donovan C, Magrane M, Alpi E, Antonius R, Zhang J. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2017;45(D1):D158–69. https://doi.org/10.1093/nar/gkw1099.

Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381–5. https://doi.org/10.1093/nar/gkg090.

Leisching G, Cole V, Ali AT, Baker B. International Journal of Infectious Diseases. OSA1 , OAS2 and OAS3 restrict intracellular M . tb replication and enhance oxidative stress. J Infect Dis 2019;209(suppl 1):S78–84. https://doi.org/10.1093/infdis/jiy181.

Donovan C, Magrane M, Alpi E, Antunes R, Zhang J. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2017;45(D1):D158–69. https://doi.org/10.1093/nar/gkw1099.

Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 2003;31(13):3381–5. https://doi.org/10.1093/nar/gkg090.