

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Utilization from Computational Methods and Omics Data for Antiviral Drug Discovery to Control of SARS-CoV-2

Ömür Baysal and Ragıp Soner Silme

Abstract

SARS-CoV-2 pandemic issue threatening world health and economy became a major problem with its destructive impact. The researchers have seen that conventional methods related to medicine and immunological background do not resolve this disease by gained knowledge of viruses previously studied. Advances in computational biology comprising bioinformatics, simulation, and yielded databases have accelerated and strengthened our facilities to predict some cases related to the biological complex by comparison with the use of artificial intelligence. Various novel drugs by using *in silico* resources and *in vivo* imaging techniques associated with high-resolution technologies can cause the confidential development of methods for the detection of antiviral drugs and the production of diagnosis kits. In the future, we will start seeing these novel techniques' positive reflection and their advantages in cost/time effective profits. This chapter highlights these approaches and addresses updated knowledge currently used for research and development.

Keywords: Computational biology, Drug discovery, Genomics, Omics science, SARS-CoV-2

1. Introduction

Coronaviruses (CoVs) are positive-strand RNA viruses belonging to the order of *Nidovirales* including three families *Arteriviridae*, *Coronaviridae*, and *Roniviridae* [1]. Relied on the genetic studies, they classify CoVs into four genera as alpha, beta, gamma, and delta CoVs [2]. The diameter of CoVs is between 80 to 120 nm and their shapes are spherical. The fundamental structural proteins of CoVs are envelope (E), membrane (M), nucleocapsid (N), and spike (S) [1]. Its RNA genome composes of six to ten open reading frames (ORFs) [3].

The new studies will fill the knowledge gaps to reveal how the virus is evolving and adapting to new conditions. In recent years, the advanced findings on nucleic acid amplification technologies have been the reason for improving of automated DNA sequencing with the help of bioinformatics tools to characterize and classification of all kinds of infectious disease agents. One of the single-stranded RNA viruses, the Coronaviruses, has been classified using molecular tools. On sequence analysis, the genomes were identified by direct RNA extraction of the clinical

specimens isolated from nasopharyngeal aspirate or stool, as the template, since the viruses are non-cultivable [4–11]. The collection of SARS-CoV-2 sequences has been started in 2020 under the GISAID database. The analysis of viral genomes provided the preventing of possible viral mutations during *in vitro* viral replication. These provided data helped us to understand the virus, threatening the world health, at genomic and *in silico* levels, which gave rise to new experiments carried out in the laboratory. Protein function prediction methods mainly fall into sequence- and structure-based approaches. Using precisely important databases and tools relied on comparison for sequence, structural differentiation, and gene ontology enables us to find exact protein function annotation [12]. Given the destructive effect of the virus SARS-CoV-2 on human health and its contagious virulence, it has attracted the attention of researchers to find its efficiently curative method. We have realized that antiviral chemotherapy with small molecules for their properties as nucleoside analogs can identify new uncharacterized viral genes for producing antiviral drugs related to viral glycoproteins to cellular receptors, viral regulatory proteins. These drugs may block the synthesis and replication of the viral genome that induce the host immune response [13, 14].

These targeted regions can reduce the crucial function required for survival of the virus, by polymerase and/or protease assay enabled us getting of high throughput screening for identification of inhibitor small molecules and enzymes, which can be beneficial for the development of effective antiviral components and synthesized novel molecules [15]. Understanding of viral gene products could overcome the challenge of the development of antiviral drugs contributing to essential functions on the virus, with assays carried out *in vitro* considering molecular mechanisms involved in gene products and biochemical processes. New technology related to omics science is suggesting novel possibilities to find the right answers to inquiries resulting from unexplored pathogenic behavior of the virus. Bioinformatics promises to generate new knowledge on virus and host interaction that can help drive the efforts in more detail to the discovery and development of antiviral therapies [16]. Genomics, proteomics, and related technologies will also be beneficial in molecular virology as suited techniques and approaches for big data.

2. Bioinformatics and computational tools

The progress in the fields of genomics and proteomics are encouraging biological studies on the virus. Genomic sequences and bioinformatics are also major tools in this field and quantities of raw data which has tremendously increased besides their complexity. Therefore, significant computational resources required to manage the volumes of data and their manipulation, researchers studying in these fields for any future drug discovery projects are using these new technologies. Bioinformatics resources (GISAID; NCBI) required to analyze the data, identify patterns and display the patterns help to investigators for understanding the problem, testing, and confirmation of their hypothesis in the laboratory to focus on prioritized compounds or genes [17]. Computational methods applied in the study of SARS-CoV-2 could be paved for the characterization of the virus collected from unique specimens and comparison with similar genomes resulted from sequence similarity. Comprehensively studied investigations on the characterization of the viruses to set a unique set of well-described genomes compared within each other have been reported [18]. Bioinformatics workflows and tools related to SARS-CoV-2 to the detection of potential drug targets and providing beneficial knowledge on

therapeutic strategies have also been recently acknowledged [19]. New bioinformatics tools applied to these genomes to test their ability and to predict the organization of viral genes involving coding capacity and the function of the viral proteins are commonly used [20]. These tools are assisting for the confirmation of transcriptional patterns, gene expression, and gene function which are essential studies earlier than *in vitro* studies not to lose time and labor cost [21].

Data relevant to the discovery of new drugs contain information related to biological function, chemical structure, and the biologic activity of small molecules that all findings can help for the searching for new compounds. Even the nature of this problem is inherently complex, bioinformatics is a useful tool to handle the volumes of data required with databases. Small molecule inhibitors could target the computational methods, suggesting aspects of the viral genomic property. Then they may be the reason for identification of the small molecules well-described with their biological effects, which could be used to probe for following of the cellular functions related to chemical structure, protein structure, biochemical activity, and biologic activity of the virus. As another branch of data mining on whole existing data shows a way for screening on inhibitory chemicals with known biochemical activities according to their chemical classes.

3. Impact of omics science and related fields on SARS-CoV-2 research

With advanced techniques and bioinformatics tools, the scientific landscape has dramatically changed in recent years. The huge data yielding on omics science plays an important role in the steps related to the biology of SARS-CoV-2 infections towards understanding more [22]. These resources are tremendously necessary to scientists studying SARS-CoV-2 infections and provide a map and common reference points to reach the data for describing precisely viral transcripts and ORFs. Comparison of different genomic organizations among all the SARS-CoV-2 isolates forms a starting point to determine the evolutionary relationships in this virus family. The most important point that should not be missed is the instability of the nucleotide sequences in the virus genome, which causes high ratio mutations. Viral genomes data inherently in GenBank involves missing annotated parts that sequence needs to be corrected. Annotations of viral genomes conducted with the best tools are available to test gene prediction with precise algorithms to identify new genes [23]. The annotation process may cause inconsistent findings for different genomes as the terminology used to describe gene function [24]. Viral genomes need to be updated and re-annotated as additional strains are of importance for comparing sequences for the continual annotation process considering analogous by released versions in NCBI [25].

The RNA sequence and structure of the genomes could be easily sequenced, but to predict their role in infection with any certainty seems difficult at the course of an infection. *In vitro* experimentation results should prove the different ORFs identified through algorithms such as codon/pair usage, dinucleotide/junction usage, RNA structure differentiation which are detected by bioinformatics tools on a viral infection [26]. Even microarray using oligonucleotide probes to hybridize with putative exons and splice junctions could be beneficial for following the expression of predicted transcripts and splice variants in virus genome [27], single-cell RNA sequencing analysis of SARS-CoV-2 will help define how the virus integrates into a human as use host cell organization to regulate and code for all the required biologic process [28]. As this knowledge with different biological assays increasingly supports findings on SARS-CoV-2 and its pathogenic behavior, the proteomics data

obtained on up/down-regulated expression levels expressed by the virus reflects ongoing RNA transcripts that can be evaluated as biological cases related to post-translational processing playing role in protein formation complexity. Proteomics methods have also the potential to follow the modification of viral gene products during viral infection, which will help to characterize how post-translation modification that affects viral replication. Since omics technology maybe not sufficient alone to find effective compounds inhibiting viral replication and invasive negative effects that occurred on the human body, we should consider the detection of the genomic parts showing stability without a high mutation ratio to design targeted molecules with inhibitory potential [13]. Genomic screening using specific algorithms to identify conserved motifs and to predict protein structure could be an efficient way to understand protein functions [29]. In immunological studies, model organism, yeast, thanks to its two-hybrid (Bait and Prey) system that can be suggested to prove the protein–protein interactions among viral–cellular proteins and potential gene products cooperating in biological processes can be clarified by the construction of protein–protein interaction maps [30].

4. Drug discovery by means of omics data on SARS-CoV-2

Genomics and proteomics are promising new areas affecting apparently whole biological fields with widespread data and tools provided by databases.

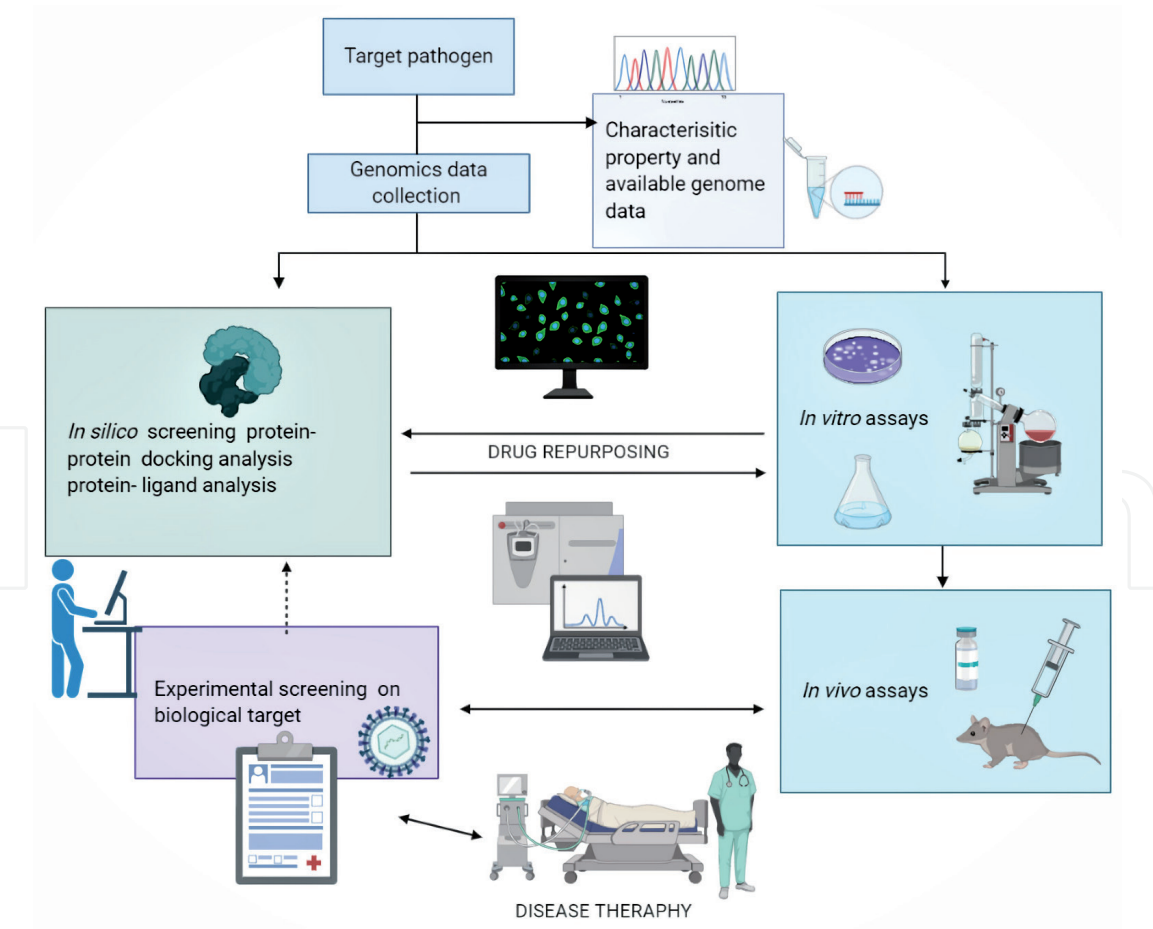


Figure 1. The experiment-based approach is activity-based repositioning of original drugs for new pharmacological indications based on experimental assays, which involves protein target-based and cell/organism-based screening in vitro and/or in vivo assays. These studies are followed with cell assay, animal model approach and clinical approach. Illustration was created with BioRender.com by the authors of this chapter.

The DNA Data Bank of Japan (DDBJ) (<https://www.ddbj.nig.ac.jp/>), GISAID initiative (<https://www.gisaid.org/>), National Center for Biotechnology Information (The GenBank) (<http://www.ncbi.nlm.nih.gov/>) and The European Bioinformatics Institute (EMBL-EBI) (<http://www.ebi.ac.uk/>) are important resources to researchers as nucleotide databases provided on the web. Most functions in micro/ macro organism are directed by interactions of proteins and ligands. Hence, computational techniques comprising *in silico* techniques to predict the protein complex formed can be remarkably cheaper and quicker than experimental methods. They are being a guide for subsequent targeted experiments before initiating *in vitro* studies cause of their predictive capability. Predicting the binding possibilities of multiple proteins is critical for understanding their biological function in any target organism to design of drugs addressing the impairment of biological processes (**Figure 1**). Many solutions generated from a pair of static molecular structures with scoring function comprise the specific position of each atom, giving rise to the simulation of modeling that is seriously sensitive to the specific packing of atoms at the interface [31–33]. For modeling the protein, dynamics and correct protein arrangement are required, considering scoring functions related to the feature of docking poses using techniques such as molecular dynamics (MD) [34, 35].

5. Importance of drug discovery and molecular docking

The docking method relies on steric complementarity at the protein–protein interface level. These interfaces are observed in co-crystallized complexes available in the Protein Data Bank (PDB). They have been the major driving force in the development of docking with the addition of physicochemical and statistics-based properties [36, 37].

Homology modeling and protein prediction analysis enables us to test different proteins on SARS-CoV-2 with various ligands. Analysis by protein-ligand docking servers (**Table 1**) is available for geometric shape complementarity score (GSC score) and approximate interface area (AI area). Additionally, different software-based tools for molecular dynamic analysis could be used. The interaction analysis of protein-ligand complexes and their amino acid position with bond distances calculated and visualized through the software provides an opportunity for molecular docking simulations. Protein docking servers can confirm the results within the protein and ligand [44]. They can get an insight into their all binding preferences within the active site of the protein and ligand (**Figure 2**).

Program	Country	Year	Reference
AADS	India	2011	[38]
AutoDock Vin	USA	2010	[39]
BetaDock	South Korea	2011	[40]
LigDockCSA	South Korea	2011	[41]
PythDock	South Korea	2011	[42]
VoteDock	Poland	2011	[43]

Table 1.
List of most commonly used protein-ligand softwares, comprising the updated ones.

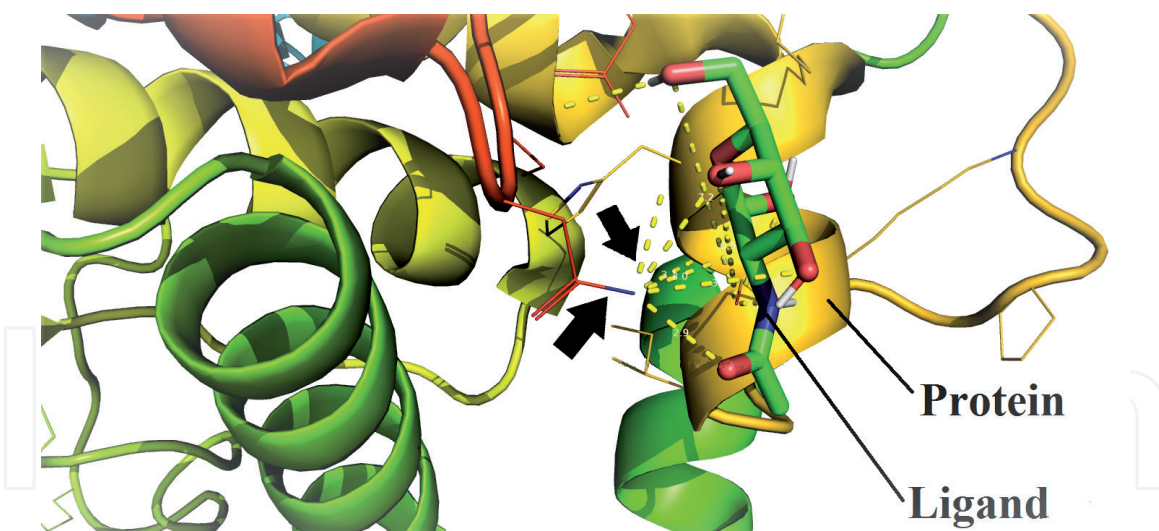


Figure 2.
Protein- ligand interaction illustration from Baysal et al. [13]. Arrows indicate the binding possibilities.

6. The effect of circadian rhythm on SARS-CoV-2-infection and host immune response

Many physiological processes influenced by 24-hour circadian rhythms has been known within individual cells. Even virology and circadian rhythms seem like different fields of biology, studies have suggested a novel interaction between viral infection and the circadian clock [45]. Edgar et al. showed 10 times higher viral replication in the mice infected at the start of their resting phase, likely as late evening for humans [45]. The finding confirmed the relationship between the circadian phase and virus infection. This data shows viral infection can directly affect the advantage of physiological rhythms in combating viruses. Noteworthy, a new study also indicated that the expression of the genes exhibited circadian rhythm in monocytes are shifting between two-time points showing the active and the resting periods in human individuals, which also indirectly affects SARS-CoV-2 replication [46]. The drug intake should also be adjusted considering the circadian rhythm of the body and virus to increase the efficacious of inhibitory compounds (**Figure 3**).

Another study also shows circadian clock has a central role in coordinating daily physiological processes involving immunity and biological process that humans are more susceptible to infections at certain times of the day cause of the function related to defense systems (**Figure 3**) with a daily rhythmic pattern [47]. In viral diseases, deciphering the complex relationships between circadian timekeeping, host immunity, and host-virus interactions has a great potential to unravel the complexity of severe acute respiratory syndrome coronavirus pathogenicity. Infection severity which may be regulated by the circadian clock affecting therapy positively against the novel pathogen may also result in recession of the pathogen. Therefore, circadian nature accounted for responses needs urgent studies on clock-infection biology in SARS-CoV-2. Even the pathophysiology of SARS-CoV-2 infection and its severe complications are not well understood, SARS-CoV-2 severity may also depend on day-night cycle (**Figure 3**). The battle between virus replication and its neutralization by the host immune system could move simultaneously with the circadian activity phase of the host. Accelerating the activity of circadian immunity factors may help to control virus replication, as circadian clocks provide a competitive advantage to the host against SARS-CoV-2.

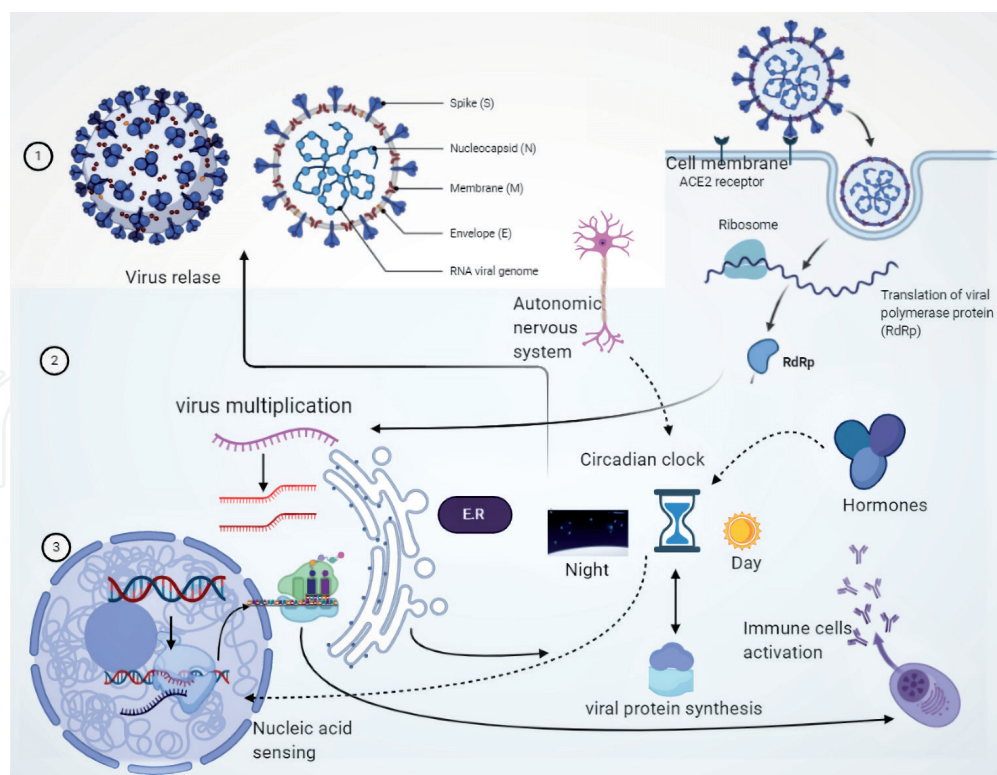


Figure 3.
Cellular clock affects the virus life cycle (1) which directly/ indirectly influenced by circadian clock of the host cell. These cases trigger multiple steps in regulation of different pathways (2) against virus and its replication phases involving particle genesis. In addition, immune responses combating viral infections regulated by circadian clocks could change depending on day and night (3). Illustration was created with BioRender.com by the authors of this chapter.

As known defect on the circadian clock in hosts causes increasing of pathogen replication and invasion, which indicates that the severity of infection influenced by circadian rhythms. All these cases can be followed by omics technologies comprising genomics, transcriptomics and proteomics data [48, 49]. We stress the combination of the whole omics data to understand all biological cases related to SARS-CoV-2 and host defense mechanism. In consequence, exact treatment times aiming to control of virus will provide higher success in managing the disease. Circadian Expression Profiles Database (CircaDB) (<http://circadb.hogeneschlab.org/>) could be beneficial for the evaluation of omics data with expression profiles related to the circadian clock [50]. Notably, the expression profiles of potential drug targets considering the circadian rhythm of the host could provide a new strategy for effective compounds depending on application doses, which may affect the efficiency of tested pre-purposed or novel compounds. We need to better understand how the circadian clock affects SARS-CoV-2 infection to optimal clinical management of the virus.

Another highlighted point for drawing the attention of the scientists, virulent agents causing infections could help to shape our genome that is also responsible for various polymorphic structure involving many loci related to MHC antigen processing [51]. There is also strong evidence on the effect of the virus infection depending on HLA allele's expression [52], even their influence is moderate level. Accordingly, the resistance to pathogens comprising viruses conferred by several genes acts at different stages of the host-pathogen interaction. The finding on another virus HIV at least 250 genes affecting the success of the infection has been demonstrated [53]. Consequently, the lack of genes or its impairment could have a negative or less effect on the virus invasion. These cases arise from deficient genes which lead to increased susceptibility to different virus infections [54].

7. Mutational changes in virus genome and traceability

As the inherent property of the virus genome, there is a higher mutational tendency compared to other micro/macro-organisms. The RNA virus replication does not comprise proofreading, as in the DNA cycle that this case renders genetic material to convenient for missing in the transcriptional phase [55]. This feature of RNA virus replication gives rise to high mutation possibility and formation of high yields, occurring of replication in a short time. RNA virus replication involves a complex and dynamic mutant formation ratio in certain genomic sites affecting the nucleotide sequence, caused by environmental factors. The model described for the evolution of virus shows comprehensible pathogenic behavior as quasispecies that have special population structure with a huge number of variant genomes relied on mutations [56]. These high mutation rates raised continually are changing in the relative frequency of the replication and selection. This process is the adoption of primitive replicons involving mutant distributions as seen in RNA viruses within their host [56, 57].

This mutational tendency depends on the population size of the virus involved in the infection. Therefore, a large population results in rapid fitness for cellular organisms. An important challenge in studies on RNA virus evolution is the differentiation, depending on phenotypic traits with ongoing specific mutations. They may associate different mutations with the biological behavior of the virus, which may be existence for the expression of phenotypic traits. These cases are the reason for the formation of restricted types. The findings on epidemiological, functional genomics, and structural studies showed the tolerance of the genetic changes on RNA viruses which are indispensable characteristic properties stemming from the virus evolution. However, the extinction of the viral infection cannot be estimated just from the characteristics of the existing sequence that is an unpredictable transitional phase of the genetic information based on lethal mutagenesis. Relied on the genomic data, the mutational ratio on viral sequences can be easily followed, but the effect of the mutation resulted in less epidemical and pathogenic behavior cannot be determined without clinical studies and monitoring on the host. Even the omics science provides predication data on the virus, this is not enough alone if not supported by filiation studies on epidemic cases. This mutation limiting the pathogenicity of the virus may result in alternative solutions occurring spontaneously in nature for ending the viral infections.

8. Future outlook

According to our current knowledge on SARS-CoV-2, our facilities limit to the efficient management of the disease and do not be enough to cut down the severity of the pathogen invasion except for protective methods relied on vaccination. Even it seems a major alternative method within other possibilities we are not sure how long the virus will keep its genetic stability without the mutation, which will not render all developed vaccines possible and effective approach for further infection waves. Particularly, we now need to determine whether SARS-CoV-2 is also more severe at certain times of the day or not, which is directly related with crosstalk between the circadian clocks and viral infections besides immunological strategies based on vaccination. Drug designing or testing of pre-purposed compounds with high potential inhibitory on viral replication should be accelerated without wasting time. We are not sure what will tomorrow bring us and how other biotic and abiotic reasons will affect the pathogenicity and genetically that may change the SARS-CoV-2 and other viruses.

9. Conclusions

Given the existing advanced techniques present today for following the genetic structure of the RNA viruses, we are able to find a solution for combating them. But it is possible to face the novel viruses that appeared in further periods in the world cause of shifting in ecological balance and the negative effect of global warming. We should be prepared for the further worst epidemic scenarios resulted from not only due to the virus but also by other microorganisms. It appears this kind of devastating case will be not the last one if the human being continues wasting of the irreversible property of nature and ecological biodiversity. As the genetic material RNA, viruses have their own unique repair process that emerged as early as 3.5 to 2.5 billion years ago in the crust of the world. The uncovered genomic data puts insights on many biological processes for deciphering the dramatic scientific cases threatening world health, environmental issues. We believe that the post-genomic area at which we have completed genomic characterization of whole macro/micro-organisms will serve for the harvesting of the fruits, which will be useful for the scientists. The data on the genome sequences available, already /or will be soon, will offer all the information concerning the threats such as SARS-CoV-2. Bioinformatics will have a dramatic impact on improving our understanding of this kind of unclear cases. Omics science and yielded purified data are expected to be an important contributor to the global issues waiting for outbreaks cause of pandemic cases. Researches in this field will play a major role and will impact drug discovery and pharmaceutical development comprising health care and the environment.

We stress in this chapter; bioinformatics tools will increase the potential of curing the diseases and producing new effective solutions besides accurately correlated clinical parameters of patient responsiveness to therapy. Bioinformatics used in the building of global databases in molecular microbiology to enhance the accumulative knowledge in the purpose of the experimental data and meta-data about microorganisms. Drastically, whole bioinformatics tools and data yielding with omics science involving data mining will establish dynamically updated and flexible portals upon the novel microbial diversity with biotechnological innovations by our efforts aimed to reach end-products.

Acknowledgements

The authors wish to thank to Prof. Dr. Nazlı Arda for her valuable support on our studies related to SARS-CoV-2. Moreover, the authors dedicate this chapter to scientists who think outside the box.

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

Author details

Ömür Baysal^{1*} and Ragıp Soner Silme²

1 Faculty of Science, Molecular Microbiology Unit, Department of Molecular Biology and Genetics, Muğla Sıtkı Koçman University, Muğla, Turkey

2 Center for Research and Practice in Biotechnology and Genetic Engineering, Istanbul University, Istanbul, Turkey

*Address all correspondence to: omurbaysal@mu.edu.tr

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Masters PS. The molecular biology of coronaviruses. *Advances in Virus Research*. 2006;**66**:193-292. DOI: 10.1016/S0065-3527(06)66005-3
- [2] de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ. Host factors in coronavirus replication. *Current Topics in Microbiology and Immunology*. 2018;**419**:1-42. DOI: 10.1007/82_2017_25
- [3] Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses*. 2012;**4**(6):1011-1033. DOI: 10.3390/v4061011
- [4] Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, Luk GS, Dyrting KC, Chan KH, Yuen KY. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *Journal of Virology*. 2009;**83**(2):908-917. DOI: 10.1128/JVI.01977-08
- [5] Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, Wong BH, Gao K, Tsoi HW, Huang Y, Li KS, Lam CS, Chan KH, Zheng BJ, Yuen KY. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *Journal of Virology*. 2007;**81**(4):1574-1585. DOI: 10.1128/JVI.02182-06
- [6] Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, Li G, Dong BQ, Liu W, Cheung CL, Xu KM, Song WJ, Vijaykrishna D, Poon LL, Peiris JS, Smith GJ, Chen H, Guan Y. Prevalence and genetic diversity of coronaviruses in bats from China. *Journal of Virology*. 2006;**80**(15):7481-7490. DOI: 10.1128/JVI.00697-06
- [7] Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;**102**(39):14040-14045. DOI: 10.1073/pnas.0506735102
- [8] Chu DK, Peiris JS, Chen H, Guan Y, Poon LL. Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in *Miniopterus* bats. *The Journal of General Virology*. 2008;**89**(5):1282-1287. DOI: 10.1099/vir.0.83605-0
- [9] Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *Journal of Virology*. 2008;**82**(10):5084-5088. DOI: 10.1128/JVI.02722-07
- [10] Zhang J, Guy JS, Snijder EJ, Denniston DA, Timoney PJ, Balasuriya UB. Genomic characterization of equine coronavirus. *Virology*. 2007;**369**(1):92-104. DOI: 10.1016/j.virol.2007.06.035
- [11] Lau SK, Woo PC, Li KS, Huang Y, Wang M, Lam CS, Xu H, Guo R, Chan KH, Zheng BJ, Yuen KY. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology*. 2007;**367**(2):428-439. DOI: 10.1016/j.virol.2007.06.009
- [12] Grant MA. Integrating computational protein function prediction into drug discovery initiatives. *Drug Development Research*. 2011;**72**(1):4-16. DOI: 10.1002/ddr.20397
- [13] Baysal O, Silme RS, Karaaslan C, Ignatov A. Genetic uniformity of a specific region in SARS-CoV-2 genome

and repurposing of N-Acetyl-D-Glucosamine. *Fresenius Environmental Bulletin*. 2021;**30**(3):2848-2857. DOI: 10.5281/zenodo.4621319

[14] Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*. 2020;**12**(3):254. DOI: 10.3390/v12030254

[15] Dooley AJ, Shindo N, Taggart B, Park JG, Pang YP. From genome to drug lead: Identification of a small-molecule inhibitor of the SARS virus. *Bioorganic & Medicinal Chemistry Letters*. 2006;**16**(4):830-833. DOI: 10.1016/j.bmcl.2005.11.018

[16] Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature*. 2003;**422**:835-847. DOI: 10.1038/nature01626

[17] Venter JC, Levy S, Stockwell T, Remington K, Halpern A. Massive parallelism, randomness and genomic advances. *Nature Genetics*. 2003;**33**:219-227. DOI: 10.1038/ng1114

[18] Chaitanya KV. Structure and organization of virus genomes. In: *Genome and Genomics*. 1st ed. Singapore: Springer; 2019. p. 1-30. DOI: 10.1007/978-981-15-0702-1_1

[19] Hufsky F, Lamkiewicz K, Almeida A, Aouacheria A, Arighi C, Bateman A, Baumbach J, Beerenwinkel N, Brandt C, Cacciabue M, Chuguransky S, Drechsel O, Finn RD, Fritz A, Fuchs S, Hattab G, Hauschild AC, Heider D, Hoffmann M, Hölzer M, Hoops S, Kaderali L, Kalvari I, von Kleist M, Kmiecinski R, Kühnert D, Lasso G, Libin P, List M, Löchel HF, Martin MJ, Martin R, Matschinske J, McHardy AC, Mendes P, Mistry J, Navratil V, Nawrocki EP, O'Toole ÁN,

Ontiveros-Palacios N, Petrov AI, Rangel-Pineros G, Redaschi N, Reimering S, Reinert K, Reyes A, Richardson L, Robertson DL, Sadegh S, Singer JB, Theys K, Upton C, Welzel M, Williams L, Marz M. Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Briefings in Bioinformatics*. 2020;**bbaa232**:1-22. DOI: 10.1093/bib/bbaa232

[20] Pappas N, Roux S, Hölzer M, Lamkiewicz K, Mock F, Marz M, Dutilh BE. *Virus bioinformatics. Reference Module in Life Sciences*. 2020:1-9. DOI: 10.1016/B978-0-12-814515-9.00034-5

[21] Hölzer M, Marz M. Software dedicated to virus sequence analysis “bioinformatics goes viral”. *Advances in Virus Research*. 2017;**99**:233-257. DOI: 10.1016/bs.aivir.2017.08.004

[22] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – From vision to reality. *EuroSurveillance*. 2017;**22**(13):30494. DOI: 10.2807/1560-7917.ES.2017.22.13.30494

[23] Mills R, Rozanov M, Lomsadze A, Tatusova T, Borodovsky M. Improving gene annotation of complete viral genomes. *Nucleic Acids Research*. 2003;**31**(23):7041-7055. DOI: 10.1093/nar/gkg878

[24] Koonin EV, Galperin MY. *Genome Annotation and Analysis. Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. Boston: Kluwer Academic; 2003. Chapter 5, Available from: <https://www.ncbi.nlm.nih.gov/books/NBK20253/> [Accessed 2020-12-12]

[25] Holmes EC. RNA virus genomics: A world of possibilities. *The Journal of Clinical Investigation*. 2009;**119**(9): 2488-2495. DOI: 10.1172/JCI38050

- [26] Kames J, Holcomb DD, Kimchi O, DiCuccio M, Hamasaki-Katagiri N, Wang T, Komar AA, Alexaki A, Kimchi-Sarfaty C. Sequence analysis of SARS-CoV-2 genome reveals features important for vaccine design. *Scientific Reports*. 2020;**10**:15643. DOI: 10.1038/s41598-020-72533-2
- [27] Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scherer S, Boguski MS. Experimental annotation of the human genome using microarray technology. *Nature*. 2001;**409**:922-927. DOI: 10.1038/35057141
- [28] Mahalingam R, Dharmalingam P, Santhanam A, Kotla S, Davuluri G, Karmouty-Quintana H, Ashrith G, Thandavarayan RA. Single-cell RNA sequencing analysis of SARS-CoV-2 entry receptors in human organoids. *Journal of Cellular Physiology*. 2020;**1-9**. DOI: 10.1002/jcp.30054
- [29] Grant MA. Integrating computational protein function prediction into drug discovery initiatives. *Drug Development Research*. 2011;**72**(1):4-16. DOI: 10.1002/ddr.20397
- [30] Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. *International Journal of Molecular Sciences*. 2009;**10**(6):2763-2788. DOI: 10.3390/ijms10062763
- [31] Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein docking algorithm. *Proteins*. 2003;**52**(1):80-87. DOI: 10.1002/prot.10389
- [32] Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*. 2006;**34**(suppl_2):W310-W314. DOI: 10.1093/nar/gkl206
- [33] Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins*. 2006;**65**(2):392-406. DOI: 10.1002/prot.21117
- [34] Jackson RM, Gabb HA, Sternberg MJE. Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *Journal of Molecular Biology*. 1998;**276**(1):265-285. DOI: 10.1006/jmbi.1997.1519
- [35] Krol M, Tournier AL, Bates PA. Flexible relaxation of rigid-body docking solutions. *Proteins*. 2007;**68**(1):159-169. DOI: 10.1002/prot.21391
- [36] Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*. 2003;**331**(1):281-299. DOI: 10.1016/s0022-2836(03)00670-3
- [37] Zacharias M. Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Science*. 2003;**12**(6):1271-1282. DOI: 10.1110/ps.0239303
- [38] Singh T, Biswas D, Jayaram B. AADS - an automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of Chemical Information and Modeling*. 2011;**51**(10):2515-2527. DOI: 10.1021/ci200193z
- [39] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of*

- Computational Chemistry. 2010;**31**(2):455-461. DOI: 10.1002/jcc.21334
- [40] Kim DS, Kim CM, Won CI, Kim JK, Ryu J, Cho Y, Lee C, Bhak J. BetaDock: Shape-priority docking method based on beta-complex. *Journal of Biomolecular Structure & Dynamics*. 2011;**29**(1):219-242. DOI: 10.1080/07391102.2011.10507384
- [41] Shin WH, Heo L, Lee J, Ko J, Seok C, Lee J. LigDockCSA: Protein-ligand docking using conformational space annealing. *Journal of Computational Chemistry*. 2011;**32**(15):3226-3232. DOI: 10.1002/jcc.21905
- [42] Chung JY, Cho SJ, Hah JM. A python-based docking program utilizing a receptor bound ligand shape: PythDock. *Archives of Pharmacal Research*. 2011;**34**:1451-1458. DOI: 10.1007/s12272-011-0906-5
- [43] Plewczynski D, Lazniewski M, Von Grotthuss M, Rychlewski L, Ginalski K. VoteDock: Consensus docking method for prediction of protein-ligand interactions. *Journal of Computational Chemistry*. 2011;**32**(4):568-581. DOI: 10.1002/jcc.21642
- [44] Mashich E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. *Proteins*. 2010;**78**(15):3197-3204. DOI: 10.1002/prot.22790
- [45] Edgar RS, Stangherlin A, Nagy AD, Nicoll MP, Efsthathiou S, O'Neill JS, Reddy AB. Cell autonomous regulation of herpes and influenza virus infection by the circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;**113**(36):10085-10090. DOI: 10.1073/pnas.1601895113
- [46] Diallo AB, Gay L, Coiffard B, Leone M, Mezouar S, Mege JL. Daytime variation in SARS-CoV-2 infection and cytokine production. *bioRxiv*. 2020.09.09.290718. DOI: 10.1101/2020.09.09.290718
- [47] Man K, Loudon A, Chawla A. Immunity around the clock. *Science*. 2016;**354**(6315):999-1003. DOI: 10.1126/science.aah4966
- [48] Mazzocchi G, Vinciguerra M, Carbone A, Relógio A. The circadian clock, the immune system, and viral infections: The intricate relationship between biological time and host-virus interaction. *Pathogens*. 2020;**9**(2):83. DOI: 10.3390/pathogens9020083
- [49] Borrmann H, McKeating JA, Zhuang X. The circadian clock and viral infections. *Journal of Biological Rhythms*. 2020. DOI: 10.1177/0748730420967768
- [50] Pizarro A, Hayer K, Lahens NF, Hogenesch JB. CircaDB: A database of mammalian circadian gene expression profiles. *Nucleic Acids Research*. 2013;**41**(D1):D1009-D1013. DOI: 10.1093/nar/gks1161
- [51] Segal S, Hill AV. Genetic susceptibility to infectious disease. *Trends in Microbiology*. 2003;**11**(9):445-448. DOI: 10.1016/s0966-842x(03)00207-5
- [52] Goulder PJ, Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology*. 2008;**8**:619-630. DOI: 10.1038/nri2357
- [53] Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ. Identification of host proteins required for HIV infection through a functional genomic screen. *Science*. 2008;**319**(5865):921-926. DOI: 10.1126/science.1152725
- [54] Kenney AD, Dowdle JA, Bozzacco L, McMichael TM, St Gelais C, Panfil AR,

Sun Y, Schlesinger LS, Anderson MZ, Green PL, López CB, Rosenberg BR, Wu L, Yount JS. Human genetic determinants of viral diseases. *Annual Review of Genetics*. 2017;**51**:241-263. DOI: 10.1146/annurev-genet-120116-023425

[55] Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;**96**(24):13910-13913. DOI: 10.1073/pnas.96.24.13910

[56] Eigen M, McCaskill J, Schuster P. Molecular quasi-species. *Journal of Physical Chemistry*. 1988;**92**(24):6881-6891. DOI: 10.1021/j100335a010

[57] Nowak MA. What is a quasispecies? *Trends in Ecology & Evolution*. 1992;**7**(4):118-121. DOI: 10.1016/0169-5347(92)90145-2