Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

Genomic Surveillance of SARS-Cov-2 in Iraq and *In Silico* Prediction of the Effect of Amino Acid Changes on the Protein Function

Nihad A. M. Al-Rashedi[1,]*, Murad G. Munahi[b], Zainab I. Kuwait[2]

[1] Department of Biology, College of Science, Al-Muthanna University, Samawah, Iraq
[2] Department of Chemistry, College of Science, Al-Muthanna University, Samawah, Iraq

***Corresponding author:**
Nihad A. M. Al-Rashedi, PhD
Department of Biology
College of Science
Al-Muthanna University
Samawah, Al Muthanna Province, Iraq
Telephone: +9647801165803
E-mail address: nhidaee@mu.edu.iq

Abstract

The emergence and rapid outbreak of novel coronavirus-19 show the significance of viral genomic surveillance, which provides a key insight into the tracking and pathogenicity of this infectious disease. Although data of complete genome sequences of SARS-CoV-2 is important to monitor the spreading of the ongoing pandemic, the number of genome sequences from Iraq is still limited to date. However, our efforts have been focused on establishing the first Iraqi genomic surveillance. The whole-genome sequences were uploaded from GISAID EpiCoV database to perform phylogenetic analysis and to investigate the effect of some amino acid variations on protein stability and function. The model structure of the mutant spike (A522V) was built and then used in molecular docking to assess binding affinity toward ACE-2 receptor and compared to Wuhan Spike protein sequence as wild type. The phylogenetic analysis was revealed that most of the genome sequences of the Iraqi isolates belonged to clades GH (GISAID), B.1 (Pangolin) and 20C (Nextstrain). Furthermore, the common variations D614G (spike), Q57H (NS3) and P323L (NSP12) were detected in high frequency. Moreover, there was no novel amino acid variation and most of them were found to be stabilizing the proteins of SARS-CoV-2. It has been noticed that the amino acid variation A522V has no significant effect on spike-ACE2 interactions. Finally, SARS-CoV-2 in Iraq is still virulent and has not been attenuated by the reported mutations until this time.

Keywords: COVID-19; SARS-CoV-2; Genome sequencing; Spike; ACE2; Molecular docking

Introduction

In December 2019, a severe acute respiratory disease was emerged in Wuhan, China which was caused by SARS-CoV-2 (Lorenzo-redondo et al., 2020), the World Health Organization was declared the outbreak of this disease as a pandemic on March 11, 2020 (WHO, 2020). 127 Million confirmed cases have now reached by this pandemic in 188 countries over the world as of 28 March 2021 M30. Genome sequencing of SARS-CoV-2 has reported that this virus belong to a new betacoronavirus-Sarbecovirus and closely similar to SARS-CoV (Maitra et al., 2020), the genome sequencing has been showed that SARS-CoV-2 is one of the RNA viruses having nucleotides length of 29,903 (Zhu et al., 2020). Currently, more than 900,000 complete genome sequences of SARS-CoV-2 isolated from different countries are available on GISAID database,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

however, the number of shared sequences isolated from Iraq in this database is still limited (GISAID, https://www.gisaid.org/). Mutations of SARS-CoV-2 may occur over time to escape the human immune system responses, especially with the strong immunological pressure in human (Molina-Mora et al., 2020). This possibly leads to the generation of new virus variants with potential different infectivity, transmissibility and virulence. For this reason, monitoring and analyzing mutation of the viral genomes can be helpful(Nguyen et al., 2021).

Among all the mutations, non-synonymous mutations leading to changing amino acid that could affect the proteins structure of the virus, thus RNA virus variants cause concerns that may be subject to severe pathogenicity, more transmission and escape from antibodies neutralization during vaccination as well as the misleading diagnosis of viral RNA (Wu et al., 2021). It was reported that, the structure of SARS-CoV-2 Orf6 was dramatically altered when depletion of 9 amino acids occurred (Riojas et al., 2020). Furthermore, the structure of SARS-CoV main protease was significantly changed and its enzymatic activity was decreased by the presence of F140A and S139A mutations (Hu et al., 2009). Previous study was observed that among 10333 spike proteins, 9654 mutations were reported(Guruprasad, 2021), this protein play a critical role in the viral replication cycle via its interaction with the human ACE2. The interaction between the spike protein of SARS-CoV-2 and its receptor (ACE2) represents a critical step in the viral replication cycle (Akisawa et al., 2021). Moreover, the process of viral infection is significantly dependent on this protein. Spike-ACE2 interactions are associated with several physicochemical properties. These properties are determined by the nature and type of residues that occurring between spike and its receptor(Ortega et al., 2020). Hence, the presence of mutation in spike may produce a lower free energy (an energetically favored binding) which drive interaction kinetics and eventually lead to the binding event. Depending on that, this work aimed to investigate the effect of the identified amino acid variation in Iraqi isolates on the stability and function of spike protein as well as other proteins of SARS-CoV-2.

Materials & methods

Data collection and sequence analysis

To investigate the SARS-CoV-2 genomes from Iraqi cases of COVID-19, all the available complete genome sequences were 33 genome sequences collected from GISAID database (www.gisaid.org). *Covserver* (Shu and McCauley, 2017) which is powered *by GISAID* in addition to a web application called *Coronapp* (http://giorgilab.unibo.it/coronannotator/) were used to identify mutations as well as amino acid variants of the SARS-CoV genome sequences from Iraq in comparison to the reference sequence (Wuhan-Hu-1; NC_045512.2). Also, the collected sequences were subjected to Nextclade 0.13.0 (Hadfield et al., 2018).

Prediction of mutation effect on protein stability and function

The web servers *DUET* (Pires et al., 2014) and *PROVEAN* (Choi and Chan, 2015) were used to predict the effect of mutations on the stability and function of SARS-CoV-2 proteins of the collected Iraqi isolates. The 3D-structures and sequences of SARS-CoV-2 proteins were obtained from PDB and/or GenBank. The 3D-model of the mutant spike RBD was built by I-TASSER web server (Zhang, 2008), where, spike RBD (pdb, 6m0j) was used as a template. The quality analysis of the built model and structural validation were performed using QMEAN 4.2.0 (Benkert et al., 2011) and MolProbity (Chen et al., 2010). *FUpred* web server (Zheng et al.,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

2020) was used to predict the effect of amino acid variation (A522) on the secondary structure of the spike protein. To investigate the effect of amino acid variation (A522) on the protein-protein interaction of spike-ACE2 complex, the obtained model of mutant spike RBD was subjected to HADDOCK 2.4 web server (Van Zundert et al., 2016) **.**In this work, all the web servers and tools were used with default parameters.

Results

Phylogenetic analysis

The non-synonymous mutations were the most common mutations in the collected genome sequences. The mutation C241T (at 5′ UTR region) was found in 32 isolates among 33. Moreover, the previously reported mutations 1059 (C→T), 1302(C→T), 3037 (C→T), 10834 (C→T) and 14408 (C→T) were identified at high frequency in the Iraqi isolates at ORF1ab region (266–21,555) as depicted in Figure 2. In the coding region of the S protein, two non-synonymous frequent mutations were detected C23127T and A23403G. Also, one non-synonymous frequent mutation was observed in both ORF3a (G25536T) and N regions (C28887T). In addition two infrequent non-synonymous mutations G1397A (Orf1ab region) and C28854T (N region) were found in *hCoV-19/Iraq/BAS-2/2020* and *hCoV-19/Iraq/NN_Erbil/2021*, respectively.

A total of 6 to 25 variants per genome were identified in the collected sequences (Table 1). The first two registered genome sequences of SARS-CoV-2 in Iraq were from the 20A Nextstrain clade, while the later sequenced genomes were from the 20C clade, only one isolate was belong to 19A clade (*hCoV-19/Iraq/BAS-2/2020*). The GISAID clade (GH) is presented in 31 isolates, while the remaining two isolate belonged to the O clade. According to PANGO Lineage clade, B.1 clade was characterized in 31 isolate, also, B.1.36 and B.4 clades were characterized in two strains (*hCoV-19/Iraq/NN_Erbil/2021*and *hCoV-19/Iraq/BAS-2/2020*, respectively). However, the most dominant clade in the Iraqi isolates is 20C, GH and B.1 as illustrated in Figure 1.

Table 1: Mutational and clade classification of Iraqi SARS-CoV-2 complete genomes.

Virus name	Collection date	GISAID accession ID	GISAID clade	PANGO Lineage clade	Nextstrain clade	Mutations
hCoV-19/Iraq/USAFSAM-S061/2020	2020-06-06	EPI_ISL_812258	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C12784T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S065/2020	2020-06-07	EPI_ISL_812262	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S072/2020	2020-06-07	EPI_ISL_812267	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

						C23127T, A23403G, G25563T, C26537T, C28695T, C28887T
hCoV-19/Iraq/USAFSAM-S075/2020	2020-06-07	EPI_ISL_812270	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S077/2020	2020-06-07	EPI_ISL_812272	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10440T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S079/2020	2020-06-07	EPI_ISL_812274	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, A10479G, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S080/2020	2020-06-07	EPI_ISL_812275	GH	B.1	20C	C241T, C1059T, C1302T C2695T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T, G29179T
hCoV-19/Iraq/USAFSAM-S081/2020	2020-06-07	EPI_ISL_812276	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, G6710T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S085/2020	2020-06-07	EPI_ISL_812280	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T C23127T, A23403G, G25563T, C26537T, C28695T,C28887T
hCoV-19/Iraq/USAFSAM-S086/2020	2020-06-07	EPI_ISL_812281	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S063/2020	2020-06-11	EPI_ISL_812260	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-	2020-06-	EPI_ISL_812259	GH	B.1	20C	C241T, C1059T,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

S062/2020	16					C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S064/2020	2020-06-16	EPI_ISL_812261	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S068/2020	2020-06-16	EPI_ISL_812263	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10440T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S069/2020	2020-06-16	EPI_ISL_812264	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T, G29543T
hCoV-19/Iraq/USAFSAM-S070/2020	2020-06-16	EPI_ISL_812265	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S071/2020	2020-06-16	EPI_ISL_812266	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S073/2020	2020-06-16	EPI_ISL_812268	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S074/2020	2020-06-16	EPI_ISL_812269	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S076/2020	2020-06-16	EPI_ISL_812271	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C6120T, C10834T, G11083T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S078/2020	2020-06-16	EPI_ISL_812273	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

						G25563T, C28887T, G29179T
hCoV-19/Iraq/USAFSAM-S082/2020	2020-06-16	EPI_ISL_812277	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S083/2020	2020-06-16	EPI_ISL_812278	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S084/2020	2020-06-16	EPI_ISL_812279	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10440T, C10834T, C14408T, C23127T, A23403G, G25563T C28887T
hCoV-19/Iraq/USAFSAM-S088/2020	2020-06-16	EPI_ISL_812282	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14256T, C14408T, C23127T, A23403G, G25563T C28887T
hCoV-19/Iraq/USAFSAM-S090/2020	2020-06-17	EPI_ISL_812283	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T
hCoV-19/Iraq/USAFSAM-S091/2020	2020-06-17	EPI_ISL_812284	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T, G29179T
hCoV-19/Iraq/USAFSAM-S092/2020	2020-06-17	EPI_ISL_812285	GH	B.1	20C	C241T, C1059T, C1302T, C3037T, C10834T, C14408T, C23127T, A23403G, G25563T, C28887T, C29719T
hCoV-19/Iraq/ICGEB-5T/2020	2020-06-30	EPI_ISL_582030	GH	B.1	20A	C241T, C3037T, C10078T, C12318T,C18877T, A23403G, G25563T, G28916A
hCoV-19/Iraq/ICGEB-2T/2020	2020-06-30	EPI_ISL_582029	GH	B.1	20A	C241T, C1684T, C3037T, T6294C, C14408T, C18877T, A23403G, T23509C,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

						G25563T
hCoV-19/Iraq/BAS-2/2020	2020-10-01	EPI_ISL_907075	O	B.4	19A	G45C,G1397A, G11083T, C18377T, G29374A, G29742T
hCoV-19/Iraq/BAS-1/2020	2020-10-01	EPI_ISL_956332	GH	B.1	20C	C241T, C1059T, C1302T,C3037T, C10834T, C14408T, C23127T,A23403G, G25563T, C26537T, C28695T, C28887T
hCoV-19/Iraq/NN_Erbil/2021	2021-01-20	EPI_ISL_885143	O	B.1.36	20A	C241T, C527T, C3037T, G3470A, T4579A, C8299T, C9679T, C11224T, C11941T, C12043T, C12242T, C12400T, C14408T, C18877T, C19186T, C22444T, G22604T,C22995A, A23403G,G25563T, C26735T,G26814T, A26990G,C28854T, C29762T

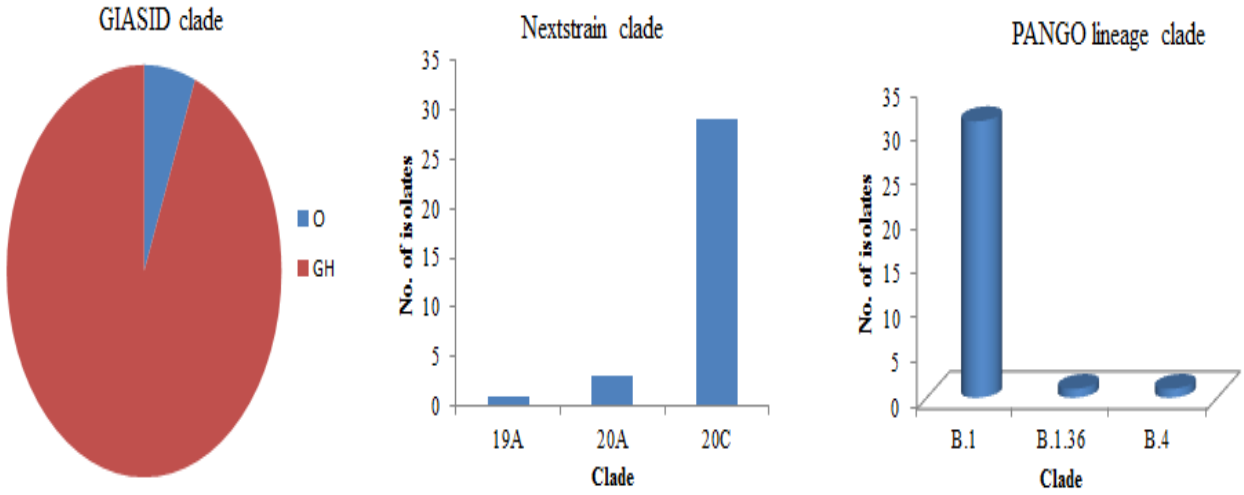

Figure 1: Clade distribution of the Iraqi isolates according to three different classifications used for SARS-CoV-2.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

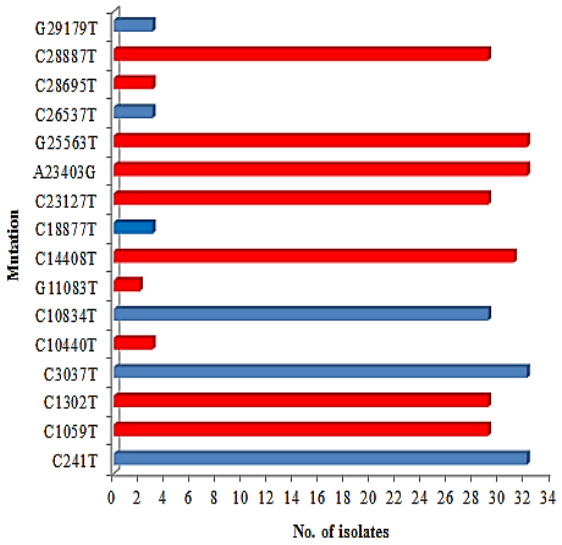

Figure 2: Mutation frequency which is detected in at least two Iraqi isolates (non-synonymous mutations are highlighted in red color).

Amino acids substitution

Depending on the analysis of amino acids substitution, the marker amino acid variations D614G (spike), Q57H (NS3), and P323L (NSP12) were found in more than 93 % of the Iraqi isolates. Also, the geographically ubiquitous mutations T205I (N) and T85I (NSP2) were observed in more than 26 isolates (87 % and 78 %, respectively). On the other hand, rare amino acid changes (T166I in NSP2 and A522V in Spike) were observed frequently in most of the Iraqi isolates (>80 %) as depicted in Figure 3. Moreover, a rare amino acid change in spike protein (T478K) was detected in one Iraqi isolate (*hCoV-19/Iraq/NN_Erbil/2021*). In addition to, infrequent amino acid changes were found individually in some isolates as illustrated in Table 2. Noteworthly, unique amino acid variations were not identified in all of the Iraqi isolates.

Table 2: Amino acid changes of SARS-CoV-2 proteins of the Iraqi isolates (rare amino acid changes are indicated by Blue color).

Virus name	NSP2 85	NSP2 166	NSP5 129	NSP6 37	NSP12 323	Spike 522	Spike 614	NS3 57	N 141	N 205	Infrequent variations
hCoV-19/Wuhan/WIV04/2019	T	T	A	L	P	A	D	Q	T	T	
hCoV-19/Iraq/USAFSAM-S061/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S065/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S072/2020	I	**I**	A	L	L	**V**	G	H	**I**	I	-

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

hCoV-19/Iraq/USAFSAM-S075/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S077/2020	I	**I**	**V**	L	L	**V**	G	H	T	I	
hCoV-19/Iraq/USAFSAM-S079/2020	I	**I**	A	L	L	**V**	G	H	T	I	**N142S**(NSP5)
hCoV-19/Iraq/USAFSAM-S080/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S081/2020	I	**I**	A	L	L	**V**	G	H	T	I	**V1331F**(NSP3)
hCoV-19/Iraq/USAFSAM-S085/2020	I	**I**	A	L	L	**V**	G	H	**I**	I	-
hCoV-19/Iraq/USAFSAM-S086/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S063/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S062/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S064/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S068/2020	I	**I**	**V**	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S069/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S070/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S071/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S073/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-	I	**I**	A	L	L	**V**	G	H	T	I	-

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

S074/2020											
hCoV-19/Iraq/USAFSAM-S076/2020	I	**I**	A	**F**	L	**V**	G	H	T	I	**S1134L**(NSP3**)**
hCoV-19/Iraq/USAFSAM-S078/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S082/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S083/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S084/2020	I	**I**	**V**	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S088/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S090/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S091/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/USAFSAM-S092/2020	I	**I**	A	L	L	**V**	G	H	T	I	-
hCoV-19/Iraq/ICGEB-5T/2020	T	T	A	L	P	A	G	H	T	T	**S76F**(NSP8) **G215S**(N)
hCoV-19/Iraq/ICGEB-2T/2020	T	T	A	L	L	A	G	H	T	T	**I1192T** (NSP3)
hCoV-19/Iraq/BAS-2/2020	T	T	A	**F**	P	A	D	Q	T	T	**V198I**(NSP2) **T113I**(NSP14)
hCoV-19/Iraq/BAS-1/2020	I	**I**	A	L	L	**V**	G	H	**I**	I	-
hCoV-19/Iraq/NN_Erbil/2021	T	T	A	L	L		G	H	T	T	**G251R**(NSP3), **R51C**(NSP8), **A348S**, **T478K,**

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

											(Spike), **S194L**(N)

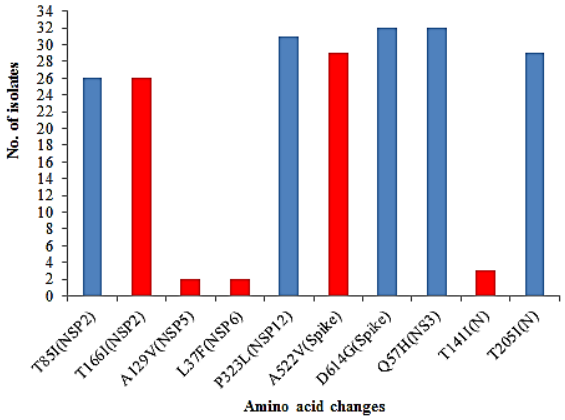

Figure 3: Frequency of amino acid changes which is identified in at least two Iraqi isolates (rare amino acid changes are highlighted in red color).

Effect of amino acids substitution on the stability and function of proteins

The thermodynamic stability of proteins is quit important for their structure and function, and it may be affected by protein mutations (Tokuriki et al., 2008). Therefore, the effect of the most frequent amino acid changes on SARS-CoV-2 proteins of the Iraqi isolate was studied by utilizing DUET and PROVEAN web servers. The stability analysis of the amino acid changes was showed that most mutations lead to stabilize the structure of the given protein (Table 3). Only NSP6 protein was showed to be destabilized by the amino acid change L37F, noteworthly, the latter is a rare and infrequent amino acid change and found in only two Iraqi isolate. Moreover, the analysis of protein function was showed that the effect of amino acid changes T85I and Q57H are deleterious, while all the other ones are neutral (Table 3).

Table 3: Prediction of amino acid-changes effect on stability and function of SARS-CoV-2 proteins of the Iraqi isolates (> 5% frequent identified amino acid changes).

Variant	Protein stability		Protein function	
	Stabilizing	destabilizing	Variation effect on protein (cutoff= -2.5)	PROVEAN score
T85I (NSP2)	√		Deleterious	-4.090
T166I (NSP2)	√		Neutral	-2.027
A129V (NSP5)	√		Neutral	-0.802
L37F (NSP6)		√	Neutral	-1.369

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

P323L (NSP12)	√		Neutral	-1.030
A522V (Spike)	√		Neutral	-0.823
D614G (Spike)	√		Neutral	0.625
Q57H (NS3)	√		Deleterious	-3.286
T141I (N)	√		Neutral	-2.446
T205I (N)	√		Neutral	-1.511

Spike structure and function

Previous studies have suggested that SARS-CoV-2 binds to the ACE2 of the host cell via spike protein, and the molecular interactions of these two proteins are explored (Lan et al., 2020) (Yan et al., 2020). The spike protein of the Iraqi isolates was accompanied by two amino acid changes; D614G and A522V, the latter is scarcely reported around the world and its possible effect on SARS-Cov-2 spike is unknown. To investigate the effect of this amino acid change on the spike-ACE2 interaction, we have performed a protein-protein interaction simulation using HADDOCK 2.4 web server, the simulation of spike RBD-ACE2 interaction was carried out for both wild and mutant spike RBD. The structure of the mutant spike RBD was modeled by I-TASSER web server, where the spike RBD (pdb, 6m0j) was utilized as a template. The validation and quality of the model structure were evaluated using MolProbity (score: 1.65) and QMEAN (score: 0.75). The effect of amino acid change (A522V) on the secondary structure of mutant spike was predicted via using FUpred, the result was showed that there is no effect on the secondary structure as shown in Figure 4. Furthermore, the obtained model structure of the mutant spike RBD and the reference spike RBD were subjected to protein-protein interaction simulation to investigate their affinity in binding with ACE2, the HADDOCK score was -134.2 for the mutant spike RBD, while the score of the reference spike RBD was -134.8.

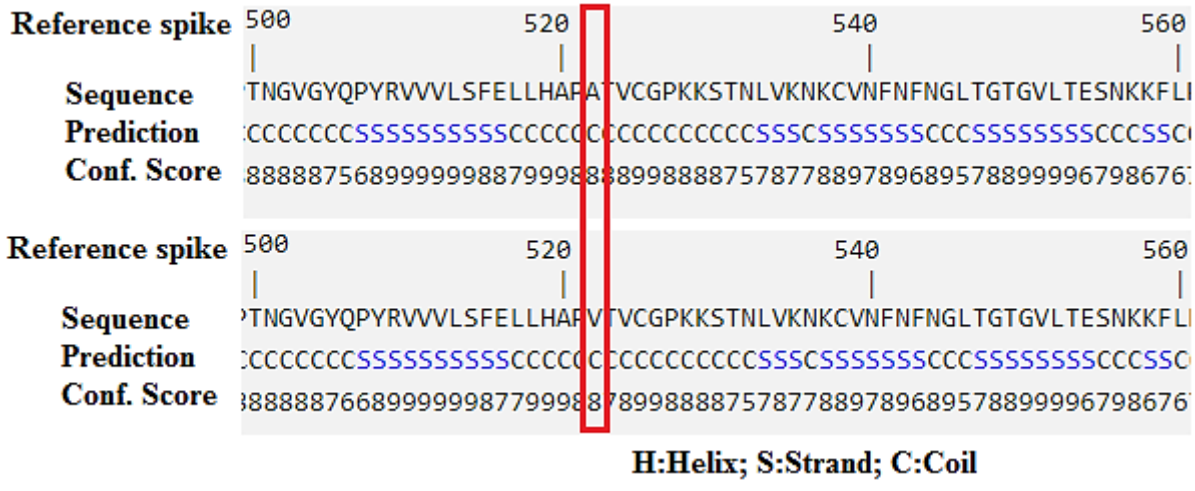

Figure 4: Secondary structure of SARS-CoV-2 mutant reference spike.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4950 - 4964
Received 05 March 2021; Accepted 01 April 2021.

Discussion

In the present study we have compared the genome sequences of SARS-CoV-2 Iraqi isolates which were retrieved from the GISAID database with the SARS-CoV-2 reference genome sequence to gain an important insight into genomes mutations. Most of the collected genome sequences of Iraqi isolates were identified with the mutations A23403G and C14408T which are associated with Nextstrain clade 20 (Zuckerman et al., 2020), also, the Nextstrain 20C clade-defining mutations (G25563T and C1059T) were detected in most of the analyzed Iraqi isolates. The majority of Iraqi isolate of SARS-CoV-2 belonged to the B.1 lineage (93 %) which is currently fast-spreading and worldwide distributed. This lineage was initially reported as the most common one in Europe and currently it's also predominant in the Americas. (Sjaarda et al., 2021; Resende et al., 2021). We also analyzed the GISAID clade distribution in Iraq, and the results was displayed that GH clade is the most prevalent with 93 %, followed by O clade 6 %. The former is the most common clade in North America, and it was associated with the most deaths in Europe as reported by the literature (Hamed et al., 2020).

In this work, the amino acid changes which occurred in more than one isolates are highlighted, and their effect on stability and function of SARS-CoV-2 proteins, obviously, there is no effect associated with the studied amino acid variations except for those L37F, T85I and Q57H. Where, L37F (NSP6) is destabilizes the structure of NSP6 but it's not deleterious, while, the amino acid variations T85I (NSP2) and Q57H (NS3) are deleterious but they are not change the stability of the corresponding proteins. However, the majority of variations taken place in the Iraqi isolates are neutral not deleterious in nature. Previous studies were showed that the amino acid variations D614G and P323L are located together in most cases and have a global prevalence ~70%, as described in results (Table 2), these two variants are located in 96 % of Iraqi isolates. V. Elizondo *et.* al, (Elizondo et al., 2021) were hypothesized that the occurrence of D614G and P323L together may possess a synergetic effect which influence the virus epidemiological success, the synergetic effect is represented by a balance between the efficient transmission that comes from D614G variation and the decreased production of the viral RNA that comes from P323L variation, thereby attenuating the virulence of the virus.

The drug-resistance phenomena could be potentially occurred with the naturally generated mutations in spike protein, as already described previously (Pachetti et al., 2020). Also, it might induce a significant decrease in spike-ACE2 complex binding affinity. Interestingly, the variant A522V in spike is a scarcely reported variant in the world (0.05%, GISAID) which was reported in some isolates in North America (Guruprasad, 2021), nevertheless this variant was found in high frequency in Iraqi isolates (87 %). For this reason, the potential effect of this amino acid change on the function of the spike was *in silico* predicted and it was found that there is no a considerable effect on its binding affinity to ACE2 as well as its secondary structure and this variant is not likely to disrupt the interaction between the spike RBD of SARS-CoV-2 and ACE2. This result is expected, especially both amino acids have a hydrophobic side chain and similar physio-chemical properties.

Conclusion

The analysis of complete genome of SARS-CoV-2 is important to shed light on its transmission and virulence. In this work we investigated the mutation, amino acid variation and clade distribution of all the available Iraqi whole genomes of SARS-CoV-2 in GISAID database.

Also, the effect of the reported variations on stability and function of SARS-CoV-2 proteins were studied. Besides, spike protein was highlighted for its key role in SARS-CoV-2 life cycle, where protein-protein interactions of the mutant spike and ACE2 was evaluated. However, the current number of the documented Iraqi complete genome sequences of SARS-CoV-2 is limited and represents <0.005 % of the total infected patients in Iraq, on the other hand, the disease is spreading fast. Therefore, this study represents the current picture, and the collected data should be considered as that.

Acknowledgments
We are thankful to all of the heroes who are fighting on the front line against COVID-19.

Conflict of Interest
The authors declare that they have no conflict of interest.

Funding
None.

References

1. Akisawa, K., Hatada, R., Okuwaki, K., Mochizuki, Y., Fukuzawa, K., Komeiji, Y., Tanaka, S.: Interaction analyses of SARS-CoV-2 spike protein based on fragment molecular orbital calculations. *RSC Advances*. *11*, 3272, 2020
2. Benkert, P., Biasini, M., Schwede, T.: Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. *27*(3), 343–350, 2011.
3. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Richardson, D. C.: MolProbity: All-atom structure validation for macromolecular crystallography. *Biological Crystallography*. *66*, 12–21, 2010.
4. Choi, Y., Chan, A. P.: PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745–2747, 2015.
5. Elizondo, V., Harkins, G. W., Mabvakure, B., Smidt, S., Zappile, P., Marier, C., Duerr, R.: SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Emerging Microbes and Infections*. *10*(1), 51–65, 2021.
6. Guruprasad, L.: Human SARS CoV-2 spike protein mutations. *Proteins*. *9*(10), 1–8, 2021.
7. Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Neher, R. A.: NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*. *34*(23), 4121–4123, 2018.
8. Hamed, S. M., Elkhatib, W. F., Khairallah, A. S., Noreddin, A. M.: Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Research Square*. *Preprint, 2020*.
9. Hu, T., Zhang, Y., Li, L., Wang, K., Chen, S., Chen, J., Shen, X.: Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. *Virology*. *388*, 324–334, 2009.
10. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Wang, X.: Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. *581*, 215–220, 2020.
11. Lorenzo-redondo, R., Nam, H. H., Roberts, S. C., Simons, L. M., Jennings, L. J., Qi, C., Ozer, E. A.: A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. *EBioMedicine*. *62*, 103112, 2020.
12. Maitra, A., Raheja, H., Biswas, N. K., Chakraborti, S.: Mutations in SARS-CoV-2 viral RNA identified in Eastern India : Possible implications for the ongoing outbreak in India and impact on viral structure and host susceptibility. *J Biosci. 45*(76), 1–18, 2020.
13. Molina-Mora, J. A., Cordero-Laurent, E., Godínez, A., Calderón, M., Porras, H. B., Soto-Garita, C., Calderón Guardia, R. A.: SARS-CoV-2 genomic surveillance in costa rica: Evidence of a divergent population and an increased detection of a spike T1117I mutation. *BioRxiv*. *Preprint, 2020*.
14. Nguyen, T. T., Pathirana, P. N., Nguyen, T., Nguyen, Q. V. H., Bhatti, A., Nguyen, D. C., Abdelrazek, M.: Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2

(COVID-19 virus). *Scientific Reports*. *11*, 3487, 2021.
15. Ortega, J. T., Serrano, M. L., Pujol, F. H., Rangel, H. R.: Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: An in silico analysis. *EXCLI Journal. 19*, 410–417, 2020.
16. Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Ippodrino, R.: Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*. *18*, 1–9, 2020.
17. Pires, D. E. V., Ascher, D. B., Blundell, T. L.: DUET: A server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*. *42*, 314–319, 2014.
18. Resende, P. C., Delatorre, E., Gräf, T., Mir, D., Motta, F. C., Appolinario, L. R., Siqueira, M. M.: Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. *Frontiers in Microbiology*. *11*, 615280, 2021.
19. Riojas, M., Frank, A., Puthuveetil, N., Flores, B., Parker, M., King, S., Rashid, S.: A Rare Deletion in SARS-CoV-2 ORF6 Dramatically Alters the Predicted Three-Dimensional Structure of the Resultant Protein. *BioRxiv*. *Preprint*, 2020.
20. Shu, Y., McCauley, J.: GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveillance*, *22*(13), 30494, 2017.
21. Sjaarda, C. P., Rustom, N., Evans, G. A., Huang, D., Perez-Patrigeon, S., Hudson, M. L., Sheth, P. M.: Phylogenomics reveals viral sources, transmission, and potential superinfection in early-stage COVID-19 patients in Ontario, Canada. *Scientific Reports*. *11*, 3697, 2021.
22. Tokuriki, N., Stricher, F., Serrano, L., Tawfik, D. S.: How protein stability and new functions trade off. *PLoS Computational Biology*. *4*(2), e1000002, 2008.
23. Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Bonvin, A. M. J. J.: The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*. *428*, 720–725, 2016.
24. World Health Organization.: WHO. WHO Director-General's Opening remarks at the media briefing on COVID-19, 11 March 2020, 2020, Available at : https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
25. Wu, S., Tian, C., Liu, P., Guo, D., Zheng, W., Huang, X., Liu, L.: Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions. *Journal of Medical Virology*. *93*, 2132–2140, 2021.
26. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q.: Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *BioRxiv*. *Preprint*, 2020.
27. Zhang, Y.: I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. *9*(40), 1–8, 2008.
28. Zheng, W., Zhou, X., Wuyun, Q., Pearce, R., Li, Y., Zhang, Y.: FUpred: detecting protein domains through deep-learning-based contact map prediction. *Bioinformatics*. *36*(12), 3749–3757, 2020.
29. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Tan, W.: A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*. *382*, 727–733, 2020.
30. Zuckerman, N. S., Pando, R., Bucris, E., Drori, Y., Lustig, Y., Erster, O., Mandelboim, M.: Transmission in a Public Health Virology Laboratory. *Viruses*. *12*(8), 854, 2020.