DNA Mutations via Chern-Simons Currents

Francesco Bajardi [1,2], Lucia Altucci [3,4*], Rosaria Benedetti [3], Salvatore Capozziello [1,2,5*], Maria Rosaria Del Sorbo [6,7], Gianluigi Franci [8,9], Carlo Altucci [2,10*]

1 Dipartimento di Fisica "Ettore Pancini", Università degli Studi di Napoli "Federico II", Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
2 INFN Sezione di Napoli, Compl. Univ. di Monte S. Angelo, Edificio G, Via Cinthia, I-80126, Napoli, Italy
3 Dipartimento di Medicina di Precisione, Università degli Studi della Campania "L. Vanvitelli", Napoli, Italy.
4 Biogem "Istituto di Biologia molecolare e genetica", 83031 Ariano Irpino, Italy
5 Scuola Superiore Meridionale, Largo San Marcellino 10, 80138 Napoli, Italy
6 Istituto Statale d'Istruzione Superiore "Leonardo da Vinci", via F. Turati Poggiomarino, Naples, Italy
7 Dipartimento di Ingegneria Industriale, Università degli Studi di Napoli "Federico II", Via Claudio n.21, I-80125, Napoli, Italy
8 Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", University of Salerno, 84081 Baronissi SA, Italy
9 AOU "San Giovanni di Dio e Ruggi d'Aragona", Salerno
10 Dipartimento di Scienze Biomediche Avanzate, Università degli Studi di Napoli "Federico II", via Pansini 5, Napoli, Italy

*lucia.altucci@unicampania.it *capozziello@na.infn.it *carlo.altucci@unina.it

Abstract

We test the validity of a possible schematization of DNA structure and dynamics based on the Chern-Simons theory, that is a topological field theory mostly considered in the context of effective gravity theories. By means of the expectation value of the Wilson Loop, derived from this analogue gravity approach, we find the point-like curvature of genomic strings in KRAS human gene and COVID-19 sequences, correlating this curvature with the genetic mutations. The point-like curvature profile, obtained by means of the Chern-Simons currents, can be used to infer the position of the given mutations within the genetic string. Generally, mutations take place in the highest Chern-Simons current gradient locations and subsequent mutated sequences appear to have a smoother curvature than the initial ones, in agreement with a free energy minimization argument.

1 Introduction

Genomic strings schematization methods represent one of the most controversial and discussed branch of science. In this scenario, the application of those methods to DNA alignment is still not fully uncovered. Several approaches aim to exhaustively predict the evolution of macro molecules, in order to get information regarding their spatial configuration [1–4]. However, a complete theory capable of predicting the interactions occuring among macro molecules and the corresponding biological implications is still missing. Biological systems, such as nucleic acids or protein, often exhibit complicated

topological structures, since several parts of the same molecule may assume a non-trivial three-dimensional shape, called tertiary structure. When two or more tertiary structures interact, the resulting system fold into a quaternary structure. In this framework, schematization approaches are particularly important in view of understanding the spatial configuration assumed by the system and, consequently, the interactions occurring among neighboring elements which may be located hundreds of kilobases away from each other and, in some cases, also in different chromosomes [5].

As an example, from the spatial configuration assumed by the DNA, it is possible to infer the place in which genomic mutations might occur, as well as the consequent difference among phenotypes. Schematization approaches can also help to provide the genetic (and epigenetic) probability to develop a certain disease. Another example is given by the interaction between proteins and virus genome which, if well described, can lead to a comprehension of the corresponding infection evolution. Standard modeling techniques are mostly based on probability considerations, aimed at outlining the many body interactions by means of statistical mechanics [6–8].

In this paper we want to test an innovative method for the schematization of biomolecule configurations, based on the topological Chern-Simons theory. It mainly relies on the curvature assumed by biological systems, using the numerical value of the Chern-Simons current, namely the expectation value of the Wilson loop [9–12].

Indeed, from very general and basic theories such as classical and quantum theories of gravity, ideas can lead to far beyond closely related fields, such as theoetical physics, cosmology and astrophysics, to push concepts and applications to complex systems, there including the interactions between biomolecules, such as nucleic acids and proteins. This model is significant because it introduces a new approach to treat biological systems, which differs from standard bioinformatics methods as it is not based on approaches typical of statistical mechanics applied to complex systems, but rather on first principles of field theories of physics. This novel point of view might be used completing outputs derived from statistical methods, to address issues of biological and medical sciences, such as preventing diseases, predicting the evolution of a genetic string or investigating the docking among biological large molecules, potentially implementing the nowadays knowledge of the biological scenario. The link between gravitational theories and the dynamics/interactions of complex biomolecules is the topological nature of the former which can be essential to describe the complicated physical-chemical and biological behavior of the latter, very much relying on their topology. Basically, the main idea is to describe the DNA curvature by using the same formalism used for the space-time, treating the interactions occurring in biological systems as driven by the same general principles that govern the gravitational interactions.

Moreover, the deterministic approach based on Chern-Simons gravity can be also merged with the intrinsic probabilistic aspect of standard bioinformatic techniques in different ways. As an example, using topological field theories to describe DNA configuration can provide the exact position in which mutations take place, by means of the comparison between two sequence curvatures. Once the position of the mutation is identified, bioinformatics is able to predict the probabilistic evolution and the clinical impact of that mutation. Another potential application which can be considered in the context of Chern-Simons formalism, is the docking between macro-molecules [13]. This latter can be understood as the interaction between points with different curvatures, which tend to attract each other only where the corresponding curvatures are similar (by analogy with the gravitational interaction). Also in this regard, the probabilistic vision provided by bioinformatic techniques can be combined with the prediction given by topological field theories, in order to develop a coherent scheme capable of predicting where and when a disease could manifest.

Although the application of Chern-Simons gravity [14–18] to complex systems seems

to be unusual, topological field theories are deeply studied in several branches of physics, due to their suitability at ultraviolet (UV) and infrared (IR) scales. In general, they involve *Topological Invariants*, namely quantities which are conserved under homeomorphism transformations. They indeed only depends on the spacetime topology, regardless of the point-like geometry [19]. Topological Invariants find their best application in the description of the gravitational interaction, with the aim of finding alternative theories to General Relativity (GR) which better adapt to the quantum formalism [20, 21].

Moreover, although the theoretical predictions of GR are perfectly consistent with observations at the level of solar system, the theory suffers some shortcoming at larger scales. As an example, the late-time exponential expansion of the universe is nowadays addressed to a never detected form of energy, called Dark Energy. Similarly, incompatibilities in the galaxy rotation curve led to the introduction of Dark Matter, which is supposed to account for the 85% of matter in the universe and to have had a high influence in the evolution of this latter. These are two of the biggest problems suffered by GR, for a complete discussion see *e.g.* [22–26]

With the aim to solve part of these issues, mainly those related to a self-consistent quantization of gravitational interaction, in the first half of twentieth century, S.S. Chern and J.H. Simons developed a topological field theory capable of describing gravity as a gauge invariant theory of different gauge groups [27]. It turns out that the n-dimensional Lagrangians whose exterior derivative gives $n+1$-dimensional topological invariants, are *quasi*-gauge invariant , *i.e* they only change by a surface term after performing a gauge transformation. However, the lack of non-trivial topological invariants in even dimensions, allows to develop the formalism in odd dimensions, only. This is the main obstacle toward the construction of a 3+1 dimensional topological theory of gravity, though odd-dimensional topological theories find large applications in several fields. See *e.g.* [14–18] for basic foundations of Chern-Simons gravity and [28–31] for applications.

Due to the applications to three-dimensional electromagnetic theory [32, 33], one of the most studied Chern-Simons Lagrangian is the 2+1 dimensional $U(1)$-invariant Lagrangian, namely:

$$\mathscr{L}_{CS}^{(3)} = \mathbf{A}\mathrm{d}\mathbf{A}, \tag{1}$$

with $\mathbf{A}$ being the one-form connection and $\mathrm{d}\mathbf{A}$ the two-form exterior derivative. Notice that the exterior derivative of $\mathscr{L}_{CS}^{(3)}$ provides the four-dimensional Pontryagin density, namely $P^{(4)} = F \wedge F$, where F represents the two-form curvature defined as $F = \mathrm{d}\mathbf{A}$.

Another well studied Lagrangian is the $SU(N)$-invariant three-dimensional Chern-Simons Lagrangian

$$\mathscr{L}_{CS}^{(3)} = \mathrm{tr}\left[\mathbf{A}\mathrm{d}\mathbf{A} + \frac{2}{3}\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right], \tag{2}$$

whose exterior derivative yields the $SU(4)$ invariant Pontryagin density $P^{(4)} = \mathrm{tr}\,[F \wedge F]$. It is mostly studied due to the applications to supergravity and string theory [18, 34, 35].

The Chern-Simons approach, as we are going to discuss, can represent a starting point for the analysis of biological systems which can be considered dynamical structures describable under the standard of GR and, furthermore, topological theories of gravity. From a conceptual point of view, the issue comes out because Quantum Mechanics, being a linear theory, could not be sufficient to approach biological systems which are highly non-linear. Due to this, non-linear theories like GR or Chern-Simons could be suitable to describe dynamics of biological systems.

For instance, by means of the Chern-Simons formalism, some important biological problems can be addressed, such as the presence of knotted DNAs and their interactions

with proteins [36]. Furthermore, in [37] the interactions of unknotted RNAs with knotted proteins have been analyzed in the process of codon and correction of RNA in methil transfer, as well as a general equation to solve the dynamics of knotted proteins has been proposed by Lin and Zewail [38], based on the Wilson loop operator for gene expression with a boundary phase condition. The basic foundations lying behind such an application can be found in [39] and [13], where some of the authors of this paper develop the formal structure of the theory, by applying it to unveil the mechanism of DNA-RNA transcriptions and providing some insights to specifically describe the junk area within the DNA sequence [39]. In [39], the theory is applied to the docking mechanism of biological macro-molecules, such as the configurational dynamics occurring in protein-protein interactions. Without claiming completeness, in Sec. 2 we outline the main properties of the theory, with the aim to test its validity by considering DNA sequences and introducing known mutations. The introduction of a mutation yields a change in the point-like curvature of the given sequence, which may give important information regarding the biological impact that such mutation may have. From the mutated sequence, it is possible to infer the frequency/probability of the mutation to occur, as well as to predict the evolution of the system towards a given configuration.

The paper is organized as follows: in Sec. 2 we briefly review the application of Chern-Simons theory to DNA and RNA systems; in Sec. 3 the formalism is then applied to different strings of KRAS human gene and to COVID-19 virus sequences. In the former case, we apply the model to analyze the mutations of a few region of the KRAS human gene, a gene acting as an on/off switch in cell signaling and, among its functions, controls cell proliferation. When KRAS is mutated, negative signaling is disrupted, with the consequence that cells can continuously proliferate, often degenerating into tumors [40, 41].

In our analysis KRAS sequences with mutations are thus compared with reference sequences, with the aim to use Chern-Simons theory to infer predictions of biological interest. As for the latter case, which is naturally one of the most studied RNA sequence to date due to pandemic, using a genome wide approach, Bobay *et al.* [42] examined SARS-CoV-2 RNA, observing that recombination events account for approximately 40% of the polymorphisms, and gene exchange occurs only within strains of the same subgenus (Sarbeco virus). Moreover, frequent mutations tend to increase the likelihood of convergent mutations, in regions exposed to a major positive selection, causing analogies in the sequences that could be misinterpreted as it was a recombination, and introduce new diversifying mutations which might accumulate, hiding past recombination events [42].

Genomic sequences of various SARS-CoV-2 strains from all over the world are available on specific platforms (eg. GISAID) and increasingly monitored to timely track SARS-CoV-2 variants [43]; as large databases and systematic sequencing are required, irregular sampling in time and space represents a crucial limitation to track pandemic evolution. Genetic diversity observed in SARS-CoV-2 populations across distinct geographic areas suggests independent events of SARS-CoV-2 introduction occurred, with few exceptions including China, being the original source, and, to a lesser extent, the early involved Italy [44]. Quantitatively, amino acid mutations were found to be significantly more frequent over the entire viral sequence in SARS-COV-2 genomes tracked in Europe (43.07%), than in Asia (38.08%) and in North America (29.64%) [43].

Here we compare sequences of single filament RNA SARS CoV-2 viruses coming from different countries, using Chern-Simons currents to potentially explain the reason why SARS-CoV-2 variants seem to exhibit a higher incidence during the 2020/2021 pandemic. Finally in Sec. 4 we conclude the work discussing results and future perspectives.

2 The Chern-Simons Theory for DNA Systems

In this section we review the application of Chern-Simons theory to DNA/RNA systems, outlining the main results obtained in [39]. The first step is to use quaternion fields to define a set of Nitrogen Bases over the DNA or RNA, namely

$$\begin{cases} A_{DNA} := e^{\frac{\pi}{2} i\beta_n} & A_{RNA} := e^{\frac{\pi}{2} j\alpha_n} \\ T_{DNA} := i\, e^{-\frac{\pi}{2} i\beta_n} & U_{RNA} := i\, e^{-\frac{\pi}{2} j\alpha_n} \\ C_{DNA} := j\, e^{i\pi\beta_n} & C_{RNA} := j\, e^{j\pi\alpha_n} \\ G_{DNA} := k\, e^{2\pi i\beta_n} & G_{RNA} := k\, e^{2\pi j\alpha_n}, \end{cases} \tag{3}$$

being $[h] \in \mathbb{H}$: $[h] = a + b\,i + c\,j + d\,k$ and $a, b, c, d \in \mathbb{R}$. The one-form connection $\mathbf{A}$ can be thought as a state of the above written nitrogen bases, namely $\mathbf{A} \in \{A, T/U, C, G\}$); consequently the DNA curvature in the configuration space of nitrogen bases is represented by the two-form curvature $F = d\mathbf{A}$, which in coordinates representation can be written as:

$$F_{\mu\nu} = \partial_{[\mu} A_{\nu]} + A_{[\mu} A_{\nu]}. \tag{4}$$

Therefore, taking into account the $SU(2)$-invariant Chern-Simons three-dimensional action

$$S^{SU(2)} = \int \mathrm{Tr}\left[\mathbf{A}\mathrm{d}\mathbf{A} + \frac{2}{3}\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right], \tag{5}$$

it is possible to define the *Chern-Simons current* as the measurable, gauge invariant quantity that can be obtained from the expectation value of the Wilson loop:

$$J = <[W(\mathbf{A})]> = \frac{\int \mathcal{D}A\, e^{iS} \Pi_n W(A_n)}{\int \mathcal{D}A\, e^{iS}}. \tag{6}$$

Wilson loop is the trace of a path-ordered exponential of the gauge connection and represents the only gauge invariant of the theory:

$$W(\mathbf{A}) = \mathrm{tr}\left[\exp\left\{\mathcal{P}\oint \mathbf{A}\right\}\right]. \tag{7}$$

They can be obtained from the holonomy of the gauge connection around a given loop and are mainly used in gauge lattice theories and quantum chromodynamics [9–12]. They have been formerly introduced to address a nonperturbative formulation of quantum chromodynamics [45] but nowadays play an important role in the formulation of loop quantum gravity, particle physics and String Theory.

The choice of the three-dimensional action is the key point of the method: standard biology suggests that nitrogen bases combine each other in triplets, and therefore form a three-dimensional space of configurations that can be described by means of the Chern-Simons three form. Any point of the space is, thus, labeled by a given triplet. Sixty-four possible combinations arise after combining the four nitrogen bases in triplets, and correspond to the combinations occurring in the genetic code. For this reason the space turns out to be discrete and finite.

By means of Eq. (3), it is possible to define a discrete superstate of configurations, in which the nitrogen bases represent the dynamical variables, so that the genetic code is labeled by the Chern-Simons currents only. After some calculations the curvature spectrum of the genetic code can be obtained [39], as reported in **Table 1**.

Table 1. Value of Chern-Simons current for the triplets of the genetic code.

Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current
Phe (UUU)	0.7071	Ser (UCU)	0.0534	Tyr (UAU)	0.0214	Cys (UGU)	0.0122
Phe (UUC)	0.5000	Ser (UCC)	0.0495	Tyr (UAC)	0.0205	Cys (UGC)	0.0118
Leu (UUA)	0.3717	Ser (UCA)	0.0460	Sto (UAA)	0.0197	Sto (UGA)	0.0115
Leu (UUG)	0.2887	Ser (UCG)	0.0429	Sto (UAG)	0.0189	Trp (UGG)	0.0112
Leu (CUU)	0.2319	Pro (CCU)	0.0402	His (CAU)	0.0182	Arg (CGU)	0.0109
Leu (CUC)	0.1913	Pro (CCC)	0.0377	His (CAC)	0.0175	Arg (CGC)	0.0106
Leu (CUA)	0.1612	Pro (CCA)	0.0354	Gin (CAA)	0.0169	Arg (CGA)	0.0103
Leu (CUG)	0.1382	Pro (CCG)	0.0334	Gin (CAG)	0.0163	Arg (CGG)	0.0010
Ile (AUU)	0.1201	Thr (ACU)	0.0316	Asn (AAU)	0.0157	Ser (AGU)	0.0098
Ile (AUC)	0.1057	Thr (ACC)	0.0299	Asn (AAC)	0.0152	Ser (AGC)	0.0096
Ile (AUA)	0.0939	Thr (ACA)	0.0284	Lys (AAA)	0.0147	Arg (AGA)	0.0093
Met (AUG)	0.0841	Thr (ACG)	0.0270	Lys (AAG)	0.0142	Arg (AGG)	0.0091
Val (GUU)	0.0759	Ala (GCU)	0.0257	Asp (GAU)	0.0138	Gly (GGU)	0.0089
Val (GUC)	0.0690	Ala (GCC)	0.0245	Asp (GAC)	0.0134	Gly (GGC)	0.0087
Val (GUA)	0.0630	Ala (GCA)	0.0234	Glu (GAA)	0.0129	Gly (GGA)	0.0085
Val (GUG)	0.0579	Ala (GCG)	0.0224	Glu (GAG)	0.0126	Gly (GGG)	0.0083

The same analysis can be also pursued by considering the amino acids, so that the genetic code is equivalently described by 21 different Chern-Simons currents. The simplest way to construct a curvature spectrum with respect to amino acids, is to take the average values of the Chern-Simons currents which refer to triplets coding for the same amino acid. The Chern-Simons currents of the amino acids are listed in **Table 2**.

Table 2. Value of Chern-Simons current for the amino acids.

Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current
Phe (F)	0.60355	Ser (S)	0.0352	His (H)	0.01785	Giu (E)	0.01275
Leu (L)	0.2305	Pro (P)	0.036675	Gin (Q)	0.0166	Cys (C)	0.012
Ile (I)	0.106567	Thr (T)	0.029225	Asn (N)	0.01545	Trp (W)	0.0112
Met (M)	0.0841	Ala (A)	0.024	Lys (K)	0.01445	Arg (R)	0.01005
Val (V)	0.06645	Tyr (Y)	0.02095	Asp (D)	0.0136	Gly (G)	0.0086

Notice that the formalism permit to assign a numerical value to each component of the genetic code, finding a point by point correspondence between triplets and curvature. Such a curvature of the DNA is the key parameter of our approach, as it may provide several predictions about the docking between two different parts of DNA or between DNA and RNA. The genomic curvature can be also used to find out those positions having highest probability to exhibit a mutation. The introduction of the mutation, indeed, leads to a local variation of the curvature, whose value might suggest the clinic importance and the impact of the corresponding disease. Moreover, the curvature spectrum can provide important insights regarding the evolution of the genomic strings: those points with highest curvature are the best candidates to evolve toward a more stable configuration, making the entire sequence more uniform in the configuration space of all the possible triplets.

3 Application of the Chern-Simons Theory to Biological Systems

3.1 The Chern-Simons Current in Mutated KRAS Human Gene

The first application of the above described method, is focused on the comparison between mutated DNA and standard DNA sequences. In particular, first we consider the KRAS gene, whose details are reported in App. A. It is located in the 12th chromosome, from the base 25,205,246 to 25,250,929 and represents one of the most mutated human genes [40, 46, 47]. Then we introduce some known mutations into the

original sequence, yielding a change in the Chern-Simons current. Being the current linked to the curvature of the DNA, the configuration space made of nitrogen bases changes the point-like curvature whereas a mutation is introduced.

By means of physical considerations, we theoretically expect the mutation to level out the graph, providing slighter variations of the current with respect to the original sequence. In analogy with other physical systems, the curved point is surrounded by a non-equilibrium region, which in turn tends to mutate in order to reach a minimum free energy state.

Moreover, this prescription is in agreement with the general criterion which governs thermodynamic transformations, according to which any spontaneous transformations must minimize the Gibbs free energy. This statement can be simply proved by considering the definition of the Gibbs free energy $\mathcal{G}$, that is

$$\mathcal{G} = U - TS + pV, \tag{8}$$

with p being the pressure, V the volume, T the temperature, S the entropy and U the free energy. Neglecting the contribution of p and setting $T = \text{const.}$ (as standard for biological systems), it turns out that for the system to undergo a spontaneous transformation, the entropy must increase as the free energy must decrease. This latter can be thought as the expectation value of the Hamiltonian of the system, which includes potential and kinetic energies. Therefore, requiring the Gibbs free energy to decrease spontaneously is equivalent to require the gravitational potential to decrease spontaneously. This means that as the system evolves toward a configuration with $\Delta\mathcal{G} < 0$, the potential energy decreases. By applying these considerations to the formalism developed in Sec. 2, a spontaneous transformation must yield an evolution of the system toward a flat regions in the configuration space.

For these reasons, mutations of DNA/RNA sequences occur to render the graph slighter and to bring the general state toward an equilibrium configuration. Reversing the argument, those mutations which make the sequence more peaked than the original one, are supposed to occur less frequently, since they lead to a higher free energy configuration. Therefore, significant variation should not occur in flat regions of the curvature spectrum, which are closer to an equilibrium state. The result of the analysis in KRAS human gene via Chern-Simons current method is reported in **Fig. 1a**.

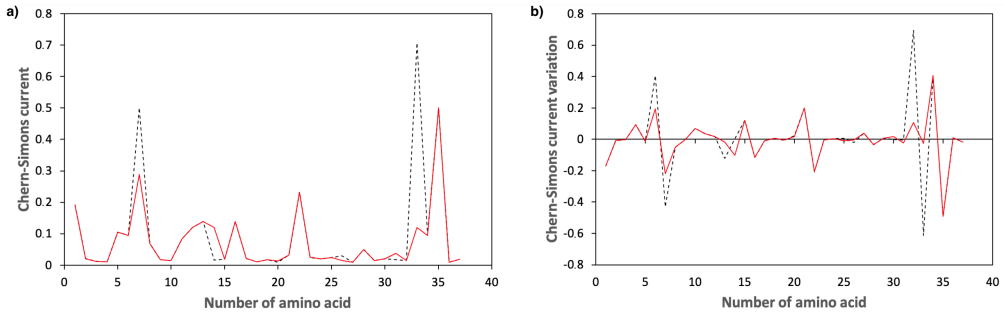


Figure 1. Chern-Simons current in KRAS gene. Figure 1a shows the comparison between the original sequence (black dashed line) and the mutated one (red solid line), while Figure 1b shows the Chern-Simons current variation, obtained comparing the point-like differences between contiguous points of the original and mutated sequences. The region considered is 25,245,274 - 25,245,384 of the 12th chromosome.

Most significant mutations occur in the regions comprised between the 5th and the 15th amino acid, and between the 30th and the 35th. Further details are reported in App. A. As expected by the free energy minimization argument, the mutations occur whereas the curvature is most peaked, providing a smoother general trend, with respect to the

original one. Notice, however, that mutations are not directly correlated to peaks, but rather to curvature gradients, namely they are mostly located near those points whose curvature is very much higher (or lower) with respect to a contiguous point. By computing the differences between contiguous points, it is possible to associate mutations to peaks, as reported in **Fig. 1b**.

In the same region of the twelfth chromosome, another set of mutations occurs (**Fig. 2**)

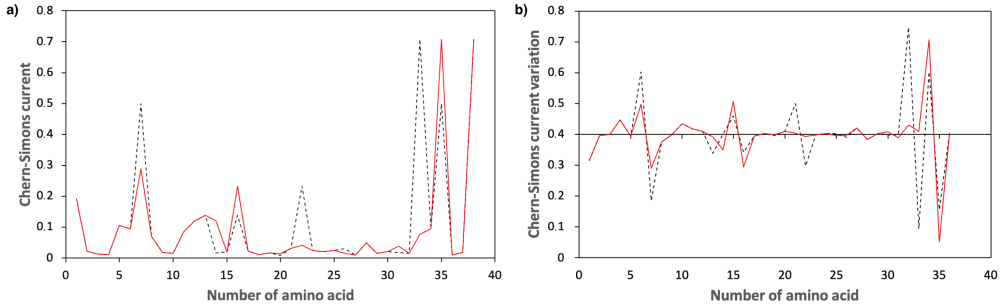


Figure 2. Chern-Simons current in KRAS gene. Figure 2a shows the comparison between the original sequence (black dashed line) and the mutated one (red solid line), while Figure 2b shows the Chern-Simons current variation, obtained comparing the point-like differences between contiguous points of the original and mutated sequences. The region considered is 25,245,274 - 25,245,384 of the 12th chromosome.

Fig. 1 and **Fig. 2** refer to the same region of KRAS, though different mutations are introduced in the two cases. More precisely, mutations occurring in these selected regions are split in two different sets, in order to facilitate reading and visualizing the curvature spectrum. It is worth noticing that even in this case, a mutation corresponds to each peak, as theoretically inferred. Moreover, the mutated sequence makes the overall trend smoother than the original one, in agreement with theoretical predictions. To confirm this result, two other different regions of human KRAS are analyzed in **Fig. 3** and **Fig. 4**, where the original sequences are again compared with the corresponding mutated. Mutations are carefully chosen according to the database BioMuta. Also in this case, further details can be found in App. A.

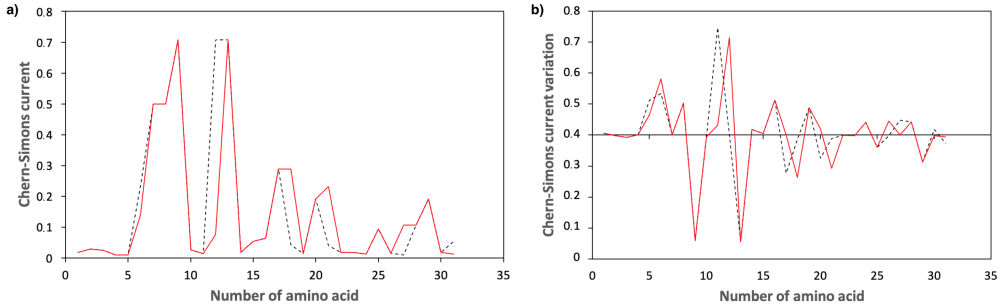


Figure 3. Chern-Simons current in KRAS gene. Figure 3a shows the comparison between the original sequence (black dashed line) and the mutated one (red solid line), while Figure 3b shows the Chern-Simons current variation, obtained comparing the point-like differences between contiguous points of the original and mutated sequences. The region considered is 25,215,468 - 25,215,560 of the 12th chromosome.

The last region analyzed, corresponding to the region 25,227,263-25,227,379 of the 12th chromosome, yields the graph in **Fig. 4**.

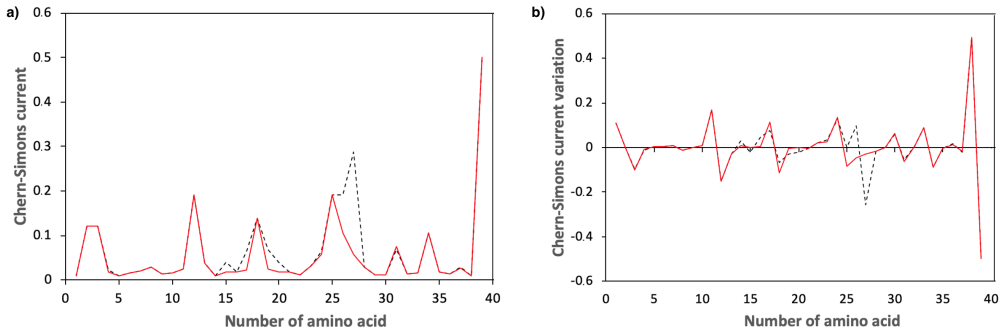


Figure 4. Chern-Simons current in KRAS gene. Figure 4a shows the comparison between the original sequence (black dashed line) and the mutated one (red solid line), while Figure 4b the Chern-Simons current variation, obtained comparing the point-like differences between contiguous points of the original and mutated sequences. The region considered is 25,227,263-25,227,379 of the 12th chromosome.

Notice that in both cases the mutations occur where the sequence is peaked, in agreement with theoretical predictions. This is particularly evident in the former case (**Fig. 3**), where almost all peaks correspond to a mutation (see also **Table 6**). Moreover, the introduction of the mutations has the effect to avoid abrupt differences in the overall trend of the curvature spectrum.

On the contrary, well known mutations may occur also in flat regions of sequences with no peak in the Chern-Simons current values. This nay be due to other factors that induce mutations, not taken into account in our model at the moment, where we basically rely on an argument based on the curvature gradient variation and free energy minimization.

3.2 The Chern-Simons Current in Mutated COVID-19 Sequences

In this subsection we discuss the results provided by the applications of Chern-Simons formalism to different variants of SARS-CoV-2 virus. The S glycoprotein is a Class I fusion protein, composed by two subunits (S1,S2) [48]; the S1 subunit contains the receptor binding domain (RBD), directly binding to the main receptor human angiotensin-converting enzyme 2 (hACE2) and determinant for both host range and cellular tropism [49]; the S2 subunit is directly involved in membrane fusion and virus endocytosis [50, 51]. Receptor binding triggers conformational changes; specifically, host proteases (such as furin) will mediate its functional transition by cleaving the interface between the two subunits (S1, S2). Additionally, the RBDs of SARS-CoV and SARS-CoV-2 are highly similar, despite few key residues, appearing to enhance the transmissibility of the novel CoV [52, 53]. The spike glycoprotein is the main inducer for neutralizing antibodies [54]; unwillingly, it shows the highest mutation rate among SARS-CoV-2 proteins [55, 56], and a variable glycosylation can create novel CTL epitopes, possibly altering hACE2 binding and accessibility to proteases and neutralizing antibodies [50, 57].

The purpose here is to find a correlation in terms of Chern-Simons current among the mutations of the sequences, a correlation that could possibly give insights aiming at localizing and predicting mutation sites in the new variants of the virus. We analyze eleven strings, which underwent mutations with respect to the original sequence of SARS-CoV-2, firstly detected in Wuhan at the end of 2019. They all correspond to the same RNA region and was selected according to **Fig. 5**. In particular, we compare the difference of Chern-Simons currents, considering variants from Asia, Europe, Oceania

and North America. Specifically, sequence 19A is the first one which arose in Wuhan and have been spreading during the initial 2020 outbreak; 19B is the first detected variant in China; 20A dominated mostly in Europe from march 2020, to subsequently spreading out globally; 20B and 20C are variants of 20A which mainly spread in the early 2020; finally, 20D, 20E, 20F, 20G, 20H, 20I occurred on summer 2020 as variants of 20B, 20C and 20A. Among them, 20I and 20H are English and south-African variants. To be more precise, we used the tool Nextclade, yielding the graph of **Fig. 5**. This figure shows the aforementioned evolution of the sequences (https://github.com/nextstrain/ncov/blob/master/defaults/clades.tsv).

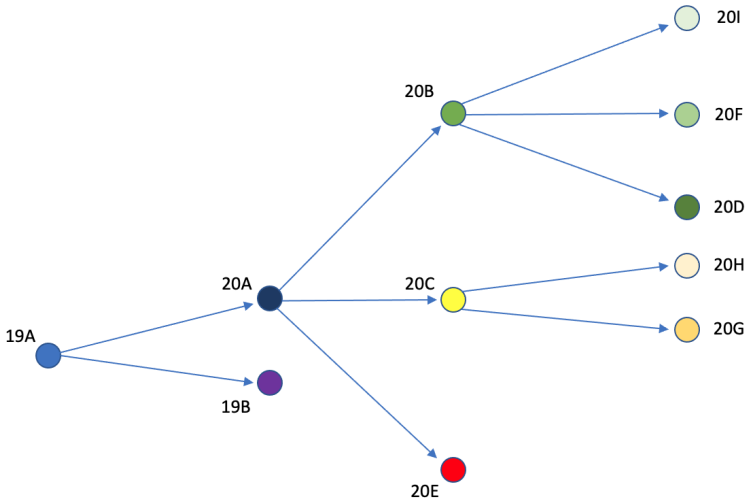


Figure 5. Evolution of the first-detected Wuhan sequence (19A) to other variants which spread out during the 2020 pandemic.

Mutations of the triplets which caused the occurrence of variants are reported in App. B. In our analysis, because of the large amount of nitrogen bases, we only compute the difference of Chern-Simons currents between the original sequence and the mutated one. Specifically, we consider the slope of the current for each mutation, namely the number

$$\text{Slope} = \frac{\text{Mutated Seq.} - \text{Original Seq.}}{\text{Original Seq.}}. \tag{9}$$

Specifically, high values of the slope represents a large discrepancy between the original sequence and the mutated one in the curvature spectrum, while lower values account for small differences. We perform the one-to-one comparison between contiguous sequences (showed in **Fig. 5**), with the aim to find out a correlation between slopes and mutations. Each variant is compared with the corresponding predecessor, so that no comparison is carried on between sequences which are not directly evolving from one another, according to **Fig. 5**. For example, sequence 19A is not compared with 20I, as well as 20D is not compared with 20H.

The analysis shows that mutations occur with highest probability where the slope (as defined in Eq. (9)) of Chern-Simons current assumes extreme values, namely when its modulus is extremely high or extremely low[1].

This means that even those mutations which do not cause significant current variations can support variants. In particular, the one-to-one comparison between the

[1]As reported at the beginning of App. B, we define current variations as "low" if they are comprehended in the range [-11%,11%], and as "high" if they are $> 100\%$ or $< -100\%$. Also notice that there is no upper limit to the modulus of the current variation, since it represents the percentage of current increase with respect to surrounding points

original and the corresponding mutated sequences shows that approx 70% of mutations corresponds to extreme values of current. Such percentage increases up to 80% if we consider only those mutations which will effectively spread out (denoted in italic bold and highlighted in light yellow), as showed in App. B, **Figs. 7-17**. Consequently, this statistic can be used to point out which occurred mutation of the sequence can be more likely to evolve in a real, spread out variant of the virus. To be more precise, once we know the position of a given mutation, Chern–Simons currents can suggest which type of triplets will arise from such mutation. In particular, as provided by the analysis, the mutated sequences should exhibit mutations whose related Chern-Simons currents provide extremely high or extremely low percentage variations, with respect to the original ones. Therefore, we do not expect the sequence to evolve such that mutations cause intermediate values of current variations; rather, if the position of the mutation is known, we expect the triplet to mutate towards those possible configurations whose Chern-Simons current is either very close or very far from the initial one (in terms of percentage). This means that from a given triplet we can select a set of possible mutations, namely those which cause either high or low current variations.

The above results constitute a part of the analysis of SARS-Cov-2 virus, which mainly relies on the evolution of given sequences towards mutated configurations. As mentioned above, this first part turns out to be useful to restrict all possible mutations within a given range, but can provide suitable information only if the position of the mutation is known *a priori*. From this point of view, no information regarding the mutation position can be provided. Now, in the next part, we use Chern-Simons formalism to select regions where mutations are most likely to occur.

With the aim to link the currents with the probability to exhibit mutations, we separately analyzed only those sequences which generate variants, *i.e.* 19A, 20A, 20B and 20C. Specifically, as we can infer from **Fig. 5**, 19A generates 19B and 20A; 20A generates 20B, 20C and 20E; 20C generates 20H and 20C. Similarly to the previous analysis of KRAS human gene, we aim to relate the curvature spectrum with the likelihood to find out mutations. To this purpose, we calculated the Chern–Simons currents of 19A, 20A, 20B and 20C sequences and computed the current variations in those points affected by known mutations. Specifically, let n be the position of a given mutation along the sequence and j_n the corresponding Chern–Simons current. The normalized current variations are computed according to the formulas:

$$\text{Variation } (\%)_1 = \frac{j_{n+1} - j_n}{j_n} \tag{10}$$

and

$$\text{Variation } (\%)_2 = \frac{j_n - j_{n-1}}{j_{n-1}}. \tag{11}$$

This means that we are investigating the current variations where the mutations occur, with respect to the previous and the subsequent points, respectively. The comparison between these values calculated for the triplets affected by mutations and the surrounding points can be used to relate the current variation with RNA mutations.

This prescription is suggested by the analysis performed on human KRAS regions, where it turns out that points far from the equilibrium state in the curvature spectrum are the best candidates to provide mutations. Here, given the large amount of amino acids, the curvature spectrum cannot be compute entirely. For this reason, we only focused on noticeable mutations, namely preferred points which exhibit known triplet variations.

The analysis again shows that mutations mostly occur where the current variation, as calculated in Eqs. (10) and (10), is high-valued. More precisely, in a set of 125 total mutations, 59% of them (74/125, see **Tables 7-10**) are located in points where the

curvature undergoes abrupt variations. This percentage increases up to 69%, if only noticeable mutations which had more impact in the development of the corresponding variants are considered. Indeed, among 25 mutations with the greatest impact in generating the variants, 17 exhibit high percentage variations of current with respect to surroundings points. These results are reported in App. B, **Tables 7-10**.

This result can be explained based on the achievements of the previous section, where non-equilibrium points turned out to be best candidates to provide nitrogen bases mutations. More precisely, large values of the current variations account for peaked regions, which tend to evolve to a lower curvature, that is a lower current. Reversing the argument, large variations of current are exhibited by points which are far from the minimum of energy, which is supposed to occur where the trend is constant.

In this framework, the application of Chern-Simons theory to DNA/RNA systems such as SARS-CoV-2 or KRAS, can give important information about the positions where the mutation is more likely to manifest. The consequent biological impact naturally follows, since this prediction can be used to prevent the occurrence of variants or to know in advance the probability for the sequence to evolve towards another configuration.

Taking into account these results, let us evaluate the spike region of SARS-CoV-2 virus only, with the aim to analyze the tertiary structure. In particular, we rely on the interaction points reported Ref. [58], according to which the amino acids of the spike protein are interact as reported in **Fig. 6** [2].

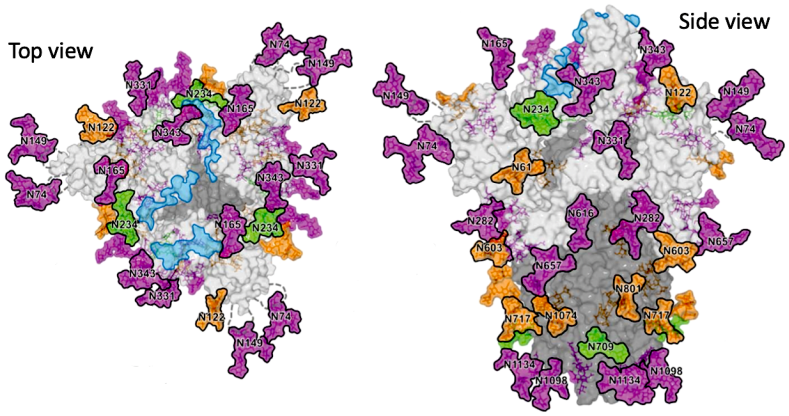


Figure 6. Tertiary structure of the spike protein of SARS-CoV-2 virus (as taken from [58], Fig. 3 therein). Green, orange and pink colors refer to the oligomannose content. Specifically, glycan sites labeled in green contain 80-100% of oligommannose, those labeled in orange 30-79% and those labeled in pink 0-29%. Light blue denotes ACE2 binding sites.

In light of the results provided by Ref. [58], we analyze 11 contact points, namely 22 corresponding amino acids. The features of these latter, such as position, current or percentage variation with respect to the surrounding triplets are reported in **Table 11**.

We considered 22 sites and calculated the Chern-Simons current variation of each amino acid with respect to the surrounding points in the linear structure. Beside the first amino acid (position 19), none of them is affected by known mutations. It is interesting to observe that the percentage of large variations in those sites which are not affected by mutations is 7/21, namely 33%. Note that such a percentage is quite lower than the previously discussed one, which is of the order of 72%. This confirms that Chern-Simons current variations is high-valued whereas mutations occur. Moreover,

[2] Numbers refer to the positions on the spike protein only

these seven sites which undergo large percentage variations are oligomannose-type, as pointed out in Ref. [58]. This, in principle, could be the reason of such large values. For instance, the high value of current variation in position 234 might be due to the proximity of the site with ACE-2, or to the high percentage of glycosylation occurring in such amino acid.

Moreover, it turns out that the docking points have same or similar values of current, which means low percentage variation. This is expected from a physical point of view, since those points with same curvature tend to interact in order to reach a stabler configuration. Also here, the analogy with gravitational interaction is simply understood.

4 Conclusions and Perspectives

In this work we apply the schematization method of the nucleic acids representation, based on the Chern-Simons theory as developed in [13, 39], to analyze some DNA sequences, such as those contained in the KRAS human gene, and some RNA noticeable sequences such as those of the most known SARS-COV variants. In particular, we compare known windows of the reference sequences with the corresponding noticeable mutations, reported in well-known and reputed genetic databases. To develop the formalism, the nitrogen bases are recast as quaternion fields, combined in triplets as dictated by biology golden rules. These triplets form a three-dimensional space of configuration that can be described through the Chern-Simons three form. The expectation value of the only observable of the theory, the Wilson Loop, provides the so called Chern-Simons current. This latter gives point-like information of the curvature of the genetic code, and can be used to compute the curvature spectrum of a given genetic string. If some triplet of the initial sequence changes due for example to the replacement of a nitrogen base, the point-like curvature changes accordingly. Therefore, the introduction of some mutations yields a variation in the Chern-Simons current. The difference between the original and the mutated sequence can be used to infer where DNA-DNA (or DNA-RNA) interactions take place, or to predict the evolution probability toward a given configuration.

On the one hand, this latter application of our method can shed light on the possibility to develop proper vaccination strategies against, for instance, SARS-CoV-2 virus; on the other hand it can potentially be used to monitor pharmacological therapies and to quantify the risk of developing DNA/RNA mutations between remission and relapsing phases.

The result of the analysis of four different regions of KRAS human gene, an important gene acting as on/off switch in cell signaling and controlling cell proliferation, shows that common features are shared in all analyzed cases. Specifically, in almost all cases, a curvature peak of the regions corresponds to a known mutation, which often yields to a new smoother curvature spectrum with respect to the reference. This can be theoretically motivated by physical considerations: the most peaked regions represent non-equilibrium points, which tend to evolve toward stabler configuration of minimum free energy.

Consequently, it follows that the variations in the curvature spectrum, leading to genetic mutations, likely take place in those regions with higher curvature. This means that as an effect of the mutations the overall trend of the curvature spectrum of the sequence tend to become smoother and smoother with no avoid abrupt variations, making nearby points to have similar values of current. As mentioned above, this happens for most of the analyzed cases; however, DNA and RNA evolution can certainly also depend on many other factors that cannot be taken into account by this method. The application of Chern–Simons theory to DNA systems, indeed, only relies on the

intrinsic curvature calculation assumed by biological systems in the configuration space made of nitrogen bases. A free energy minimum principle, then, leads the evolution of the configurations and may suggest likely position for possible mutations.

We utilize our method also to analyze RNA sequences: in this case we pick the COVID-19 virus, a striking example of the present time, and apply the same prescription for more than 20Kbases of the COVID-19 virus, coming from different countries. Due to the intrinsic attitude of RNA viruses to change their sequence with replication, mutations of various types can occur such as recombination and reassortment, rendering more complex the related genomic analyses.

Rather than analyzing the entire RNA sequence of the virus, which is very long, we prefer to focus on the regions that are reported to exhibit the most significant mutations, such as the region coding for the SARS-CoV-2 spike protein. Interestingly, the analysis shows that most of mutations occur where the slope of the Chern–Simons current takes extremely high values, which accounts for peaked regions in the curvature spectrum. This result can be explained again considering the principle of minimum free energy, according to which amino acids in correspondence of peaks of the Chern-Simons current value are intrinsically unstable and therefore tend to evolve towards a stabler configuration. Furthermore, we note that a few mutations are also exhibited in correspondence of low current values. This may happen because some regions with low current values, namely having a small curvature and being rather flat, often are the border with areas with steep gradients of the current value denoting high curvature. Then, in some cases even regions with very small curvature may be affected by a close instability due to the presence of a current gradient nearby and this cause the occurrence of a mutation. By comparing low current variations listed in **Figs. 7-16** with **Tables 7-10**, it turns out that 47% of points which exhibit low current variations between mutated and original sequences, are unstable due to the presence of a current gradient nearby.

As a final remark, the importance of the applications here discussed is twofold. On the one hand, it tests the capability of a topological theory in schematizing DNA and RNA configurations to correctly represent their interactions and mutations. On the other hand, it suggests a general criterion to predict the location in genetic sequences where it could be most likely a mutation to take place with the sequence evolution. This novel method, based on analogue gravity, can be helpful in addressing biological issues, especially when combined with standard bioinformatic approaches. For instance, the probable evolution of a given string, provided by the Chern-Simons formalism, can be approached to mathematical and statistical techniques to increase the likelihood to localize the mutations. In this sense, the approach is deterministic and based on the dynamics of structures and not only on their description. In future works we plan to provide further confirmation of the validity of our approach, by extending it to the analyses other genetic sequences both for DNA- and RNA-based systems. We also aim to study the interactions between macro molecules, in order to check whether their point-like curvature values can provide information regarding the docking probability or predict the points where the interactions occur.

Acknowledgements

C.A., F.B. and S.C. acknowledge the support of *Istituto Nazionale di Fisica Nucleare* (INFN) (*iniziative specifiche* MOONLIGHT2 and GINGER). C.A. acknowledges support from the Italian Ministry for Research under the Project PRIN - Predicting and controlling the fate of bio-molecules driven by extreme-ultraviolet radiation - Prot. Nr.20173B72NB. C.A. and F.B. aknowledge the project PON (Programma Operativo Nazionale Ricerca e Innovazione) 2014-2020 (CCI 2014IT16M2OP005), "Tecnologie innovative per lo studio di interazioni tra acidi nucleici e proteine: metodi sperimentali e modelli"; project code DOT1318991. G.F. acknowledges Giorgio Giurato for the useful discussions and suggestions regarding bioinformatic data. L.A. and R.B. aknowledge financial support by the "Associazione Italiana per la Ricerca sul Cancro" (AIRC IG17217 to L.A.), the Italian Ministry for University and Research (PRIN2015-20152TE5PK, to L.A.), the project “Epigenetic Hallmarks of Multiple Sclerosis” (acronym Epi-MS) (id:415, Merit Ranking Area ERC LS) in VALERE 2019 Program (to R.B.) and Programma V:ALERE 2020 - Progetto competitivo “CIRCE” in risposta al bando D.R. n. 138 del 17/02/2020 (to R.B.); Blueprint 282510 (to L.A.); EPICHEMBIO CM1406 (to L.A.); Campania Regional Government Technology Platform Lotta alle Patologie Oncologiche: iCURE (to LA); Campania Regional Government FASE2: IDEAL (to L.A.); MIUR, Proof of Concept POC01_00043 (to L.A.); POR Campania FSE 2014-2020 ASSE III (to L.A.).

Author contributions statement

Conceptualization, F.B., C.A., L.A. and S.C.; Formal analysis, F.B., C.A. and M.D.S.; Methodology, L.A., R.B., M.D.S. and G.F.; Supervision, C.A, L.A., S.C.; Writing–original draft, F.B., G.F. and R.B.; All authors have read and agreed to the published version of the manuscript. All authors reviewed the manuscript.

A Sequences used and Corresponding Mutations in KRAS

KRAS HUMAN
SOURCE FOR THE SEQUENCES: Genome Browser
SOURCE FOR THE MUTATIONS: BioMuta

ORIGINAL SEQUENCE 1: Chr12: 25,245,274 - 25,245,384

CUCUAUUGUUGGAUCAUAUUCGUCCACAAAAUGAUUCUGAAUUAGCUGU
AUCGUCAAGGCACUCUUGCCUACGCCACCAGCUCCAACUACCACAAGUUU
AUAUUCAGUCAU

First set of mutations (Fig. 1)

Table 3. Comparison between original and mutated sequences in KRAS. Chr12: 25,245,274 - 25,245,384

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,245,279	UAU	UAA	Y	Stop	0.0214	0.0197	-8
25,245,294	UUC	UUG	F	L	0.5	0.2887	-42
25,245,314	AAU	AUU	N	I	0.0157	0.1201	665
25,245,332	GGC	GAC	G	D	0.0087	0.0134	54
25,245,342	GCC	GCU	A	A	0.0245	0.0257	5
25,245,350	ACC	AAC	T	N	0.0299	0.0152	-49
25,245,365	CAC	CCC	H	P	0.0175	0.0377	115
25,245,370	UUU	AUU	F	I	0.7071	0.1201	-83

CUCUAAUGUUGGAUCAUAUUGGUCCACAAAAUGAUUCUGAUUUAGCUGU
AUCGUCAAGACACUCUUGCUUACGCCAACAGCUCCAACUACCCCAAGAUU
AUAUUCAGUCAU

Second set of mutations (Fig. 2)

Table 4. Comparison between original and mutated sequences in KRAS. Chr12: 25,245,274 - 25,245,384

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,245,294	UUC	UUG	F	L	0.5	0.2887	-42
25,245,314	AAU	AUU	N	I	0.0157	0.1201	665
25,245,321	CUG	CUU	L	L	0.1382	0.2319	68
25,245,332	GGC	GAC	G	D	0.0087	0.0134	54
25,245,338	CUU	CCU	L	P	0.2319	0.0402	-83
25,245,350	ACC	AAC	T	N	0.0299	0.0152	-49
25,245,365	CAC	CCC	H	P	0.0175	0.0377	115
25,245,370	UUU	GUU	F	V	0.7071	0.0759	-89
25,245,378	UUC	UUU	F	F	0.5	0.7071	41

CUCUAUUGUUGGAUCAUAUUGGUCCACAAAAUGAUUCUGAUUUAGCUUUA
UCGUCAAGACACUCCUGCCUACGCCAACAGCUCCAACUACCCCAAGGU
UAUAUUUAGUCAU

ORIGINAL SEQUENCE 2: Chr12: 25,215,468 - 25,215,560

CACACAGCCAGGAGUCUUUUCUUCUUUGCUGAUUUUUUUCAAUCUGUAUU

GUCGGAUCUCCCUCACCAAU GUAUAAAAAGCAUCCUCCACUCU

Third set of mutations (Fig. 3)

Table 5. Comparison between original and mutated sequences in KRAS. Chr12: 25,215,468 - 25,215,560

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,215,485	CUU	CUG	L	L	0.2319	0.1382	-40
25,215,501	UUU	GUU	F	V	0.7071	0.0759	-89
25,215,520	UCG	UUG	S	L	0.0429	0.2887	573
25,215,529	CCU	CUU	P	L	0.0402	0.2319	477
25,215,539	UGU	UGC	C	C	0.0122	0.0118	-3
25,215,547	AGC	AUC	S	I	0.0096	0.1057	1001
25,215,559	UCU	UGU	S	C	0.0534	0.0122	-77

CACACAGCCAGGAGUCUGUUCUUCUUUGCUGAUGUUUUUCAAUCUGUAUU
GUUGGAUCUCCUUCACCAAUGCAUAAAAAUCAUCCUCCACUGU

ORIGINAL SEQUENCE 3: Chr12: 25,227,263 - 25,227,379

AGUAUUAUUUAUGGCAAAUACACAAAGAAAGCCCUCCCCAGUCCUCAUGUA
CUGGUCCCUCAUUGCACUGUACUCCUCUUGACCUGCUGUGUCGAGAAUAUC
CAAGAGACAGGUUUC

Fourth set of mutations (Fig. 4)

Table 6. Comparison between original and mutated sequences in KRAS. Chr12: 25,227,263 - 25,227,379

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,227,272	UAU	CAU	Y	H	0.0214	0.0182	-15
25,227,306	CCU	CAU	P	H	0.0402	0.0182	-55
25,227,312	GUA	GCA	V	A	0.063	0.0234	-63
25,227,318	GUC	GCC	V	A	0.069	0.0245	-64
25,227,321	CCU	CAU	P	H	0.0402	0.0182	-55
25,227,334	GUA	GUG	V	V	0.063	0.0579	-8
25,227,338	CUC	AUC	L	I	0.1913	0.1057	-45
25,227,341	UUG	GUG	L	V	0.2887	0.0579	-80
25,227,355	GUC	GUU	V	V	0.069	0.0759	10
25,227,373	ACA	ACG	T	T	0.0284	0.027	-5

AGUAUUAUUCAUGGCAAAUACACAAAGAAAGCCCUCCCCAGUCAUCAUGCA
CUGGCCCAUCAUUGCACUGUGCUCAUCGUGACCUGCUGUGUUGAGAAUAUC
CAAGAGACGGGUUUC

B Mutations SARS-CoV-2

Comparison Between Original Sequences and Mutated Ones

The pie graphs of **Figs. 7-17** show the percentage of large and small values of current variations; large variations ($> 100\% \vee < -100\%$) are labeled by light blue squares, small variations ([-11%;11%]) by solid red, other intermediate values by grey lines.

Position	19A Triplet	19A Current	19B Triplet	19B Current	Current Variation (%)
2840	***AGC***	0.0096	***AG T***	0.0098	2
3607	TTT	0.7071	TTG	0.2887	-59
5829	CTG	0.1382	TTG	0.2887	109
5866	ATG	0.0841	GTG	0.0579	-31
5933	TCT	0.0534	TTT	0.7071	1224
9294	TTT	0.7071	TTC	0.5	-29
9697	***TAT***	0.0214	***TA C***	0.0205	-4
9762	TAC	0.0205	CAC	0.0175	-15

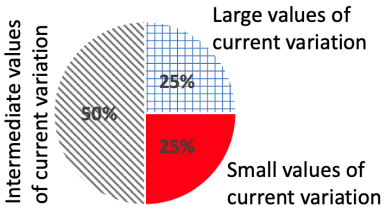


Figure 7. Comparison between 19A and 19B sequences, with related Chern-Simons current and percentage variation.

Position	19A Triplet	19A Current	20A Triplet	20A Current	Current Variation (%)
925	TTC	0.5	TTT	0.7071	41
3607	TTT	0.7071	TTG	0.2887	-59
3840	AAA	0.0147	AAG	0.0142	-3
4716	***CTA***	0.1612	***T TA***	0.3717	131
7714	***GAT***	0.0138	***G G T***	0.0089	-36
8847	CAG	0.0163	GAG	0.0126	-23
9697	TAT	0.0214	TAC	0.0205	-4
9762	TAC	0.0205	CAC	0.0175	-15

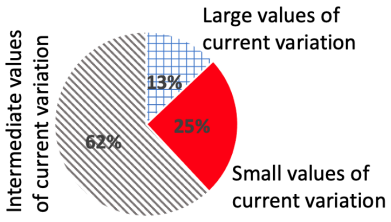


Figure 8. Comparison between 19A and 20A sequences, with related Chern-Simons current and percentage variation.

Position	20A Triplet	20A Current	20B Triplet	20B Current	Current Variation (%)
3840	AAG	0.0142	AAA	0.0147	4
6590	CCC	0.0377	CTC	0.1913	407
8036	GAC	0.0134	GAT	0.0138	3
8847	GAG	0.0126	CAG	0.0163	29
9516	GGC	0.0087	GGT	0.0089	2
9540	***AGG***	0.0091	***A AA***	0.0147	62
9541	GGA	0.0085	CGA	0.0103	21

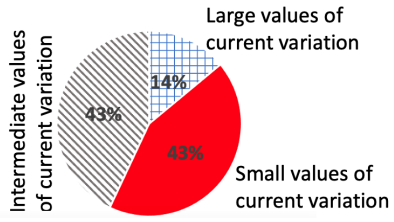


Figure 9. Comparison between 20A and 20B sequences, with related Chern-Simons current and percentage variation.

Position	20A Triplet	20A Current	20C Triplet	20C Current	Current Variation (%)
266	***ACC***	0.0299	***A T C***	0.1057	254
2130	GCT	0.0257	GTT	0.0759	195
3840	AAG	0.0142	AAA	0.0147	4
6098	ACA	0.0284	ATA	0.0939	231
6161	ACG	0.027	ATG	0.0841	211
6773	TGA	0.0115	TTA	0.3717	3132
8434	***AGA***	0.0093	***A T A***	0.0939	910
8437	CCA	0.0354	CTA	0.1612	355
8847	GAG	0.0126	CAG	0.0163	29

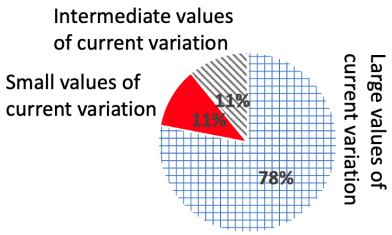


Figure 10. Comparison between 20A and 20C sequences, with related Chern-Simons current and percentage variation.

Position	20A Triplet	20A Current	20E Triplet	20E Current	Current Variation (%)
61	GTT	0.0759	GTC	0.069	-9
2008	ACC	0.0299	ACT	0.0316	6
3840	AAG	0.0142	AAA	0.0147	4
6800	CTA	0.1612	CCA	0.0354	-78
6998	CGT	0.0109	CCT	0.0402	269
7069	TGA	0.0115	TTA	0.3717	3132
7322	***GCT***	0.0257	***GTT***	0.0759	195
9546	AGA	0.0093	AAA	0.0147	58
9557	GCT	0.0257	GTT	0.0759	195
9708	GAC	0.0134	GAT	0.0138	3
9795	GTA	0.063	TTA	0.3717	490

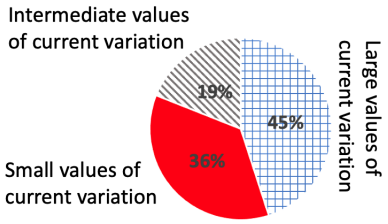


Figure 11. Comparison between 20A and 20E sequences, with related Chern-Simons current and percentage variation.

Position	20B Triplet	20B Current	20D Triplet	20D Current	Current Variation (%)
1247	***ACT***	0.0316	***ATT***	0.1201	280
1306	AAG	0.0142	AAT	0.0157	11
2148	AAC	0.0152	AAT	0.0157	3
3279	***GGT***	0.0089	***AGT***	0.0098	10
3892	TGT	0.0122	TGC	0.0118	-3
4425	***ACA***	0.0284	***ATA***	0.0939	231
4993	ACA	0.0284	ATA	0.0939	231
6299	ATT	0.1201	ACT	0.0316	-74
6479	TCA	0.046	TCG	0.0429	-7
6590	CTC	0.1913	CCC	0.0377	-80
7120	ACC	0.0299	ATC	0.1057	254
7823	***ACC***	0.0299	***ACT***	0.0316	6
8036	GAT	0.0138	GAC	0.0134	-3
8081	CTT	0.2319	CTC	0.1913	-18
9516	GGT	0.0089	GGC	0.0087	-2
9571	ATG	0.0841	ATT	0.1201	43

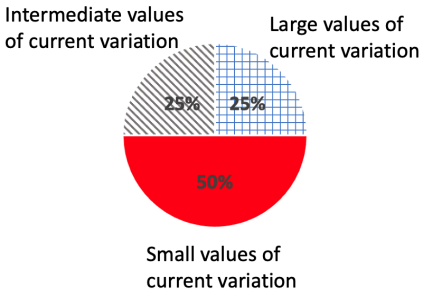


Figure 12. Comparison between 20B and 20D sequences, with related Chern-Simons current and percentage variation.

Position	20B Triplet	20B Current	20F Triplet	20F Current	Current Variation (%)
301	***ATT***	0.1201	***TTT***	0.7071	489
2426	***ACT***	0.0316	***ACC***	0.0299	-5
5462	***CGC***	0.0106	***CTC***	0.1913	1705
6098	***ACA***	0.0284	***ATA***	0.0939	231
6590	CTC	0.1913	CCC	0.0377	-80
7577	***AGC***	0.0096	***AAC***	0.0152	58
7713	***CAG***	0.0163	***CAA***	0.0169	4
8036	GAT	0.0138	GAC	0.0134	-3
9516	GGT	0.0089	GGC	0.0087	-2

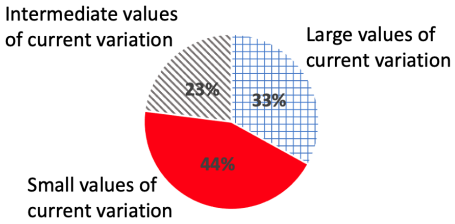


Figure 13. Comparison between 20B and 20F sequences, with related Chern-Simons current and percentage variation.

Position	20B Triplet	20B Current	20I Triplet	20I Current	Current Variation (%)
217	TCC	0.0495	TCT	0.0534	8
1002	ACT	0.0316	ATT	0.1201	280
1668	ATT	0.1201	ATA	0.0939	-22
1709	GCT	0.0257	GAT	0.0138	-46
1908	TTC	0.5	TTT	0.7071	41
2231	ATA	0.0939	ACA	0.0284	-70
4805	***CCG***	0.0334	***CTG***	0.1382	314
5006	***ACC***	0.0299	***ATC***	0.1057	254
5305	CTT	0.2319	CCT	0.0402	-83
5785	AGC	0.0096	GGC	0.0087	-9
6590	CTC	0.1913	CCC	0.0377	-80
7170	GTC	0.069	ATC	0.1057	53
7601	***AAT***	0.0157	***TAT***	0.0214	36
7670	GCT	0.0257	GAT	0.0138	-46
7781	CCT	0.0402	CAT	0.0182	-55
7816	ACA	0.0284	ATA	0.0939	231
8036	GAT	0.0138	GAC	0.0134	-3
8082	TCA	0.046	GCA	0.0234	-49
8218	GAC	0.0134	CAC	0.0175	31
9237	TCA	0.046	TTA	0.3717	708
9262	TAG	0.0189	TAT	0.0214	13
9283	GTA	0.063	GTG	0.0579	-8

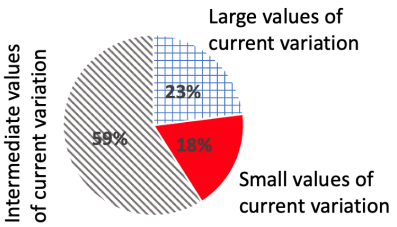


Figure 14. Comparison between 20B and 20I sequences, with related Chern-Simons current and percentage variation.

Position	20C Triplet	20C Current	20G Triplet	20G Current	Current Variation (%)
220	CTG	0.1382	TTG	0.2887	109
555	ACT	0.0316	ACC	0.0299	-5
1323	ACA	0.0284	ACC	0.0299	5
1978	CCC	0.0377	CCT	0.0402	7
2130	GTT	0.0759	GCT	0.0257	-66
3353	***CTT***	0.2319	***TTT***	0.7071	205
4236	AAC	0.0152	AAT	0.0157	3
5168	TAG	0.0189	TAT	0.0214	13
6054	***CTA***	0.1612	***CTG***	0.1382	-14
6092	TTG	0.2887	TTT	0.7071	145
6098	ATA	0.0939	ACA	0.0284	-70
6129	CTG	0.1382	TTG	0.2887	109
6161	ATG	0.0841	ACG	0.027	-68
6773	TTA	0.3717	TGA	0.0115	-97
7331	ATA	0.0939	ATT	0.1201	28
7620	GCA	0.0234	TCA	0.046	97
8437	CTA	0.1612	CCA	0.0354	-78
8549	***GTG***	0.0579	***TTG***	0.2887	399
8556	TTT	0.7071	TTC	0.5	-29
8897	GAT	0.0138	TAT	0.0214	55
9234	***GTC***	0.069	***GTT***	0.0759	10
9404	CCT	0.0402	TCT	0.0534	33
9536	***CCA***	0.0354	***CTA***	0.1612	355
9726	***CAG***	0.0163	***CTG***	0.1382	748

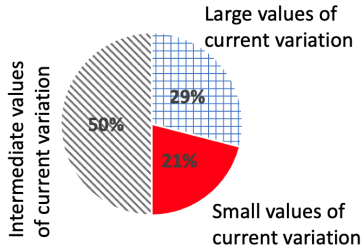


Figure 15. Comparison between 20C and 20G sequences, with related Chern-Simons current and percentage variation.

Position	20C Triplet	20C Current	20H Triplet	20H Current	Current Variation (%)
810	ACA	0.0284	ACT	0.0316	11
1656	AAG	0.0142	AAT	0.0157	11
2047	CCA	0.0354	CTA	0.1612	355
2130	GTT	0.0759	GCT	0.0257	-66
2454	GTT	0.0759	GTC	0.069	-9
2597	AAT	0.0157	AGT	0.0098	-38
3354	AAG	0.0142	AGG	0.0091	-36
3378	GTG	0.0579	GTT	0.0759	31
3451	GAC	0.0134	GAT	0.0138	3
5475	GTA	0.063	GTG	0.0579	-8
6098	ATA	0.0939	ACA	0.0284	-70
6161	ATG	0.0841	ACG	0.027	-68
6398	CGT	0.0109	CAT	0.0182	67
6773	TTA	0.3717	TGA	0.0115	-97
7118	CTT	0.2319	TTT	0.7071	205
7167	GCT	0.0257	GTT	0.0759	195
7180	GAT	0.0138	GCT	0.0257	86
7315	GAT	0.0138	GGT	0.0089	-36
7517	AAG	0.0142	AAT	0.0157	11
7584	***GAA***	0.0129	***AAA***	0.0147	14
7601	***AAT***	0.0157	***TAT***	0.0214	36
7801	GCA	0.0234	GTA	0.063	169
8437	CTA	0.1612	CCA	0.0354	-78
8548	CAG	0.0163	TAG	0.0189	16
8732	CTG	0.1382	TTG	0.2887	109
9331	CAT	0.0182	TAT	0.0214	18
9542	ACT	0.0316	ATT	0.1201	280
9809	AGT	0.0098	ATT	0.1201	1126

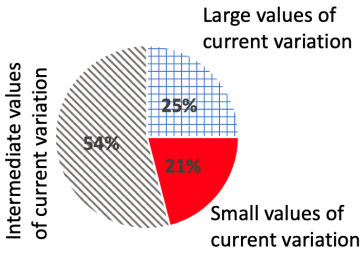


Figure 16. Comparison between 20C and 20H sequences, with related Chern-Simons current and percentage variation.

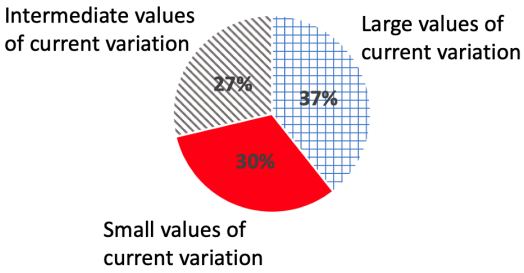


Figure 17. Details of the whole set of mutations occurring in all sequences.

Chern-Simons Current Variations in the Surroundings of Expected Mutations

Table 7. Chern-Simons currents and their corresponding percentage variations (with respect to the surrounding points) in 19A sequence of SARS-CoV-2 virus. Large values are highlighted in red.

19A Mutations				
Position	19A Triplet	19A Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
925	TTC	0.5	836	-96
2840	***AGC***	0.0096	6213	70
3607	TTT	0.7071	0	-97
3840	AAA	0.0147	-97	1201
4716	***CTA***	0.1612	821	-90
5829	CTG	0.1382	362	68
5866	ATG	0.0841	436	-62
5933	TCT	0.0534	299	596
7714	***GAT***	0.0138	-15	450
8847	CAG	0.0163	-37	889
9294	TTT	0.7071	412	-96
9697	TAT	0.0214	-9	-31
9762	TAC	0.0205	53	39

Table 8. Chern-Simons currents and their corresponding percentage variations (with respect to the surrounding points) in 20A sequence of SARS-CoV-2 virus. Large values are highlighted in red.

20A Mutations				
Position	20A Triplet	20A Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
61	GTT	0.0759	772	280
266	***ACC***	0.0299	123	1572
2008	ACC	0.0299	90	-63
2130	GCT	0.0257	0	195
3840	AAG	0.0142	-97	1247
6098	ACA	0.0284	-51	-38
6161	ACG	0.027	-78	98
6590	CCC	0.0377	-55	67
6773	TGA	0.0115	-19	1302
6800	CTA	0.1612	927	-57
6998	CGT	0.0109	-87	67
7069	TGA	0.0115	-88	6049
7322	***GCT***	0.0257	-40	1346
8036	GAC	0.0134	-21	243
8434	***AGA***	0.0093	-98	141
8437	CCA	0.0354	-85	-58
8847	CAG	0.0163	-51	1179
9516	GGC	0.0087	0	13
9540	***AGG***	0.0091	-7	272
9546	AGA	0.0093	-64	804
9557	GCT	0.0257	-89	1023
9708	GAC	0.0134	-6	10
9795	GTA	0.063	273	-78

Table 9. Chern-Simons currents and their corresponding percentage variations (with respect to the surrounding points) in 20B sequence of SARS-CoV-2 virus. Large values are highlighted in red.

20B Mutations				
Position	20B Triplet	20B Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
217	TCC	0.0495	-83	-74
301	***ATT***	0.1201	1191	-91
1002	ACT	0.0316	0	280
1247	***ACT***	0.0316	145	-55
1306	AAG	0.0142	-3	81
1668	ATT	0.1201	125	-88
1709	GCT	0.0257	0	-41
1908	TTC	0.5	2236	-94
2148	AAC	0.0152	-93	-3
2231	ATA	0.0939	-22	28
2426	***ACT***	0.0316	11	280
3279	***GGT***	0.0089	-29	37
3892	TGT	0.0122	-28	466
4425	***ACA***	0.0284	33	-71
4805	***CCG***	0.0334	120	89
4993	ACA	0.0284	0	-57
5006	***ACC***	0.0299	-84	1143
5305	CTT	0.2319	717	-93
5462	***CGC***	0.0106	-28	334
5785	AGC	0.0096	-90	3772
6098	***ACA***	0.0284	-51	-38
6299	ATT	0.1201	586	15
6479	TCA	0.046	-71	708
6590	CTC	0.1913	127	-67
7120	ACC	0.0299	5	-69
7170	GTC	0.069	279	-23
7577	***AGC***	0.0096	8	196
7601	***AAT***	0.0157	-50	-43
7670	GCT	0.0257	-79	-48
7713	***CAG***	0.0163	-24	-45
7781	CCT	0.0402	-25	-75
7816	ACA	0.0284	-25	-45
7823	***ACC***	0.0299	-61	-5
8036	GAT	0.0138	-19	233
8081	CTT	0.2319	119	-80
8082	TCA	0.046	-80	-76
8218	GAC	0.0134	-53	13
9237	TCA	0.046	124	-38
9262	TAG	0.0189	97	-22
9283	GTA	0.063	37	-71
9516	GGT	0.0089	2	10
9571	ATG	0.0841	472	-37

Table 10. Chern-Simons currents and their corresponding percentage variations (with respect to the surrounding points) in 20C sequence of SARS-CoV-2 virus. Large values are highlighted in red.

20C Mutations				
Position	20C Triplet	20C Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
220	CTG	0.1382	717	-90
555	ACT	0.0316	145	-19
810	ACA	0.0284	-55	-46
1323	ACA	0.0284	-20	-53
1656	AAG	0.0142	-3	-21
1978	CCC	0.0377	33	42
2047	CCA	0.0354	141	95
2130	GTT	0.0759	195	0
2454	GTT	0.0759	522	-79
2597	AAT	0.0157	-79	72
3353	***CTT***	0.2319	1533	-94
3354	AAG	0.0142	-94	435
3378	GTG	0.0579	26	542
3451	GAC	0.0134	-82	-32
4236	AAC	0.0152	-96	0
5168	TAG	0.0189	-73	-41
5475	GTA	0.063	-25	358
6054	***CTA***	0.1612	821	-42
6092	TTG	0.2887	3004	-93
6098	ATA	0.0939	62	-81
6129	CTG	0.1382	931	-93
6161	ATG	0.0841	-30	-37
6398	CGT	0.0109	-93	478
6773	TTA	0.3717	2518	-57
7118	CTT	0.2319	1377	-88
7167	GCT	0.0257	41	265
7180	GAT	0.0138	-98	10
7315	GAT	0.0138	27	1286
7331	ATA	0.0939	165	-91
7517	AAG	0.0142	67	746
7584	***GAA***	0.0129	-83	-31
7601	AAT	0.0157	-50	-43
7620	GCA	0.0234	29	51
7801	GCA	0.0234	163	-45
8437	CTA	0.1612	-30	-91
8548	CAG	0.0163	-93	255
8549	***GTG***	0.0579	255	45
8556	TTT	0.7071	339	-80
8732	CTG	0.1382	-72	-24
8897	GAT	0.0138	13	191
9234	***GTC***	0.069	143	22
9331	CAT	0.0182	-97	786
9404	CCT	0.0402	-92	-74
9542	ACT	0.0316	272	69
9536	***CCA***	0.0354	12	-75
9809	AGT	0.0098	-91	491

Table 11. List of amino acids of 19A sequence in the spike protein, with corresponding positions, Chern-Simons currents and their variations with respect to surrounding positions. Listed amino acid are those involved in forming the tertiary structure, according to Ref. [58].

Position	19A		Variation (%) with respect to	Variation (%) with respect to
	Triplets	Currents	the previous position	the subsequent position
17	AAT	0.0157	-79	1377
61	AAT	0.0157	-68	383
74	AAT	0.0157	-47	-43
122	AAC	0.0152	-3	69
149	AAC	0.0152	0	-3
165	AAT	0.0157	0	-25
234	AAC	0.0152	-87	595
282	AAT	0.0157	22	-46
331	AAT	0.0157	-61	665
343	AAC	0.0152	-98	61
603	AAT	0.0157	-45	101
616	AAC	0.0152	-80	-22
657	AAC	0.0152	-78	0
709	AAT	0.0157	-71	-3
717	AAT	0.0157	-45	4404
801	AAT	0.0157	-98	4404
1074	AAC	0.0152	7	3189
1098	AAT	0.0157	-66	-45
1134	AAC	0.0152	-78	0
1158	AAT	0.0157	11	16
1173	AAT	0.0157	-87	64
1194	AAT	0.0157	-96	-18

References

1. P. Baldi and S. Branak, "Bioinformatics, the machine learning approach," Second edition, The MIT Press, Cambridge (USA) - London (UK), 2001,

2. M. Waterman, "Introduction to Computational Biology," Maps, Sequences and Genomes, Chapman Hall, 1995,

3. D. Gusfield, "Algorithms on Strings, Trees and Sequences: Computer Science and computational Biology," Cambridge University Press, Cambridge, 1997

4. S. Vinga and J. Almeida, "Alignment-Free Sequence Comparison: A Review," Bioinformatics **19**(3) (2003), 513-523

5. J. E. Phillips, V. G. Corces "Master weaver of the genome," Cell **137**(7) (2009), 1194-211.

6. T. Lengauer, M. Rarey "Computational methods for biomolecular docking," Curr. Opin. Struct. Biol. **6** (1996), 402-406.

7. R.J. Morris, R.J. Najmanovich, A. Kahraman, J.M. Thornton "Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons," Bioinformatics **21**(2005), 2347-55.

8. A. Kahraman, R.J. Morris, R.A. Laskowski, J.M. Thornton "Shape variation in protein binding pockets and their ligands," J. Mol. Biol. **368** (2007), 283-301.

9. J. M. Maldacena, "Wilson loops in large N field theories," Phys. Rev. Lett. **80** (1998), 4859-4862

10. G. P. Lepage and P. B. Mackenzie, "On the viability of lattice perturbation theory," Phys. Rev. D **48** (1993), 2250-2264

11. Y. Aoki, Z. Fodor, S. D. Katz and K. K. Szabo, "The QCD transition temperature: Results with physical masses in the continuum limit," Phys. Lett. B **643** (2006), 46-54

12. T. Appelquist and R. D. Pisarski, "High-Temperature Yang-Mills Theories and Three-Dimensional Quantum Chromodynamics," Phys. Rev. D **23** (1981), 2305

13. S. Capozziello and R. Pincak, "The Chern–Simons current in time series of knots and links in proteins," Annals Phys. **393** (2018), 413-446

14. J. Zanelli, "Lecture notes on Chern-Simons (super-)gravities. Second edition (February 2008)," In Proceedings, 7th Mexican Workshop on Particles and Fields (MWPF 1999): Merida, Mexico, November 10-17, 1999, 2005. [arXiv:hep-th/0502193 [hep-th]].

15. O. Aharony, O. Bergman, D. L. Jafferis and J. Maldacena, "N=6 superconformal Chern-Simons-matter theories, M2-branes and their gravity duals," JHEP **10** (2008), 091

16. E. Witten, "Quantum Field Theory and the Jones Polynomial," Commun. Math. Phys. **121** (1989), 351-399

17. D. T. Son and A. O. Starinets, "Minkowski space correlators in AdS / CFT correspondence: Recipe and applications," JHEP **09** (2002), 042

18. A. Achucarro and P. K. Townsend, "A Chern-Simons Action for Three-Dimensional anti-De Sitter Supergravity Theories," Phys. Lett. B **180** (1986), 89

19. M. Nakahara, "Geometry, topology and physics," Boca Raton, USA: Taylor & Francis (2003)

20. F. Bajardi and S. Capozziello, "$f(\mathcal{G})$ Noether cosmology," Eur. Phys. J. C **80** (2020) no.8, 704

21. F. Bajardi, K. F. Dialektopoulos and S. Capozziello, "Higher Dimensional Static and Spherically Symmetric Solutions in Extended Gauss–Bonnet Gravity," Symmetry **12** (2020) no.3, 372

22. T. Padmanabhan, "Dark energy and gravity," Gen. Rel. Grav. **40** (2008), 529-564

23. L. Barack, V. Cardoso, S. Nissanke, T. P. Sotiriou, A. Askar, C. Belczynski, G. Bertone, E. Bon, D. Blas and R. Brito, *et al.* "Black holes, gravitational waves and fundamental physics: a roadmap," Class. Quant. Grav. **36** (2019) no.14, 143001

24. P. Bull, Y. Akrami, J. Adamek, T. Baker, E. Bellini, J. Beltran Jimenez, E. Bentivegna, S. Camera, S. Clesse and J. H. Davis, *et al.* "Beyond ΛCDM: Problems, solutions, and the road ahead," Phys. Dark Univ. **12** (2016), 56-99

25. S. Capozziello and M. De Laurentis, "Extended Theories of Gravity," Phys. Rept. **509** (2011), 167-321

26. S. Capozziello and F. Bajardi, "Gravitational waves in modified gravity," Int. J. Mod. Phys. D **28** (2019) no.05, 1942002

27. S. S. Chern and J. Simons, "Characteristic forms and geometric invariants," Annals Math. **99** (1974), 48-69

28. R. G. Cai and K. S. Soh, "Topological black holes in the dimensionally continued gravity," Phys. Rev. D **59** (1999), 044013

29. L. Smolin and C. Soo, "The Chern-Simons invariant as the natural time variable for classical and quantum cosmology," Nucl. Phys. B **449** (1995), 289-316

30. S. Giombi, S. Minwalla, S. Prakash, S. P. Trivedi, S. R. Wadia and X. Yin, "Chern-Simons Theory with Vector Fermion Matter," Eur. Phys. J. C **72** (2012), 2112

31. L. Susskind, "The Quantum Hall fluid and noncommutative Chern-Simons theory," [arXiv:hep-th/0101029 [hep-th]].

32. S. Carlip, S. Deser, A. Waldron and D. K. Wise, "Cosmological Topologically Massive Gravitons and Photons," Class. Quant. Grav. **26** (2009), 075008

33. R. Jackiw, "Chern-Simons violation of Lorentz and PCT symmetries in electrodynamics," Comments Mod. Phys. A **1** (1999), 1-9

34. E. Witten, "Chern-Simons gauge theory as a string theory," Prog. Math. **133** (1995), 637-678

35. H. Ooguri and C. Vafa, Nucl. Phys. B **577** (2000), 419-438

36. P. Dabrowski-Tumanski, J.I. Sulkowska "Topological knots and links in proteins," Proc. Natl. Acad. Sci. **114** (2017), 3415-3420

37. M. Lachner, T. Jenuwein "The many faces of histone lysine methylation," Curr. Opin. Cell. Biol. **14** (2002), 286-98

38. M. Lin, A. Zewail "Protein folding – simplicity in complexity," Ann. Phys. **524** (2012), 379-391

39. S. Capozziello, R. Pincak, K. Kanjamapornkul and E. N. Saridakis, "The Chern-Simons Current in Systems of DNA-RNA Transcriptions," Annalen Phys. **530** (2018) no.4, 1700271

40. D.J. Hartman, J.M. Davison, T.J. Foxwell, M.N. Nikiforova, S.I. Chiosea "Mutant allele-specific imbalance modulates prognostic impact of KRAS mutations in colorectal adenocarcinoma and is associated with worse overall survival," Int. J. Cancer **131** (2012), 1810-1817.

41. O. Kranenburg "The KRAS Oncogene," Biochimica et Biophysica Acta **1756** (2005), 81-82.

42. L.M. Bobay, A.C. O'Donnell, H. Ochman, "Recombination events are concentrated in the spike protein region of Betacoronaviruses," PLoS Genet. **16** (2020) e1009272

43. M.R. Islam, M.N. Hoque, M.S. Rahman, A. Alam, M. Akther, J.A. Puspo, S. Akter, M. Sultana, K.A. Crandall, M.A. Hossain, "Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity," Sci. Rep. **10** (2020) 14004

44. L. van Dorp, D. Richard, C.C.S. Tan, L.P. Shaw, M. Acman, F. Balloux, "No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2," Nat. Commun. **11** (2020) 5986

45. K. G. Wilson, "Confinement of Quarks," Phys. Rev. D **10** (1974), 2445-2459

46. S. Misale, R. Yaeger, S. Hobor, E. Scala, M. Janakiraman, D. Liska, E. Valtorta, R. Schiavo, M. Buscarino, G. Siravegna, K. Bencardino, A. Cercek, C.T. Chen, S. Veronese, C. Zanon, A. Sartore-Bianchi, M. Gambacorta, M. Gallicchio, E. Vakiani, V. Boscaro, E. Medico, M. Weiser, S. Siena, F. Di Nicolantonio, D. Solit, A. Bardelli, "Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer," Nature **486** (7404) (2012), 532-6

47. A. Liévre, J.B. Bachet, D. Le Corre, V. Boige, B. Landi, J.F. Emile, J.F. Coté, G. Tomasic, C. Penna, M. Ducreux, P. Rougier, F. Penault-Llorca, P. Laurent-Puig, "KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer," Cancer Research **66** (8) (2006), 3992-5

48. Y. Yuan Y, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, X. Zhang, G.F. Gao, "Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains," Nat Commun. **8** (2017) 15092

49. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, "Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2," Science **367** (2020) 6485

50. A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein," Cell **181** (2020) 281-292

51. X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, Z. Qian, "Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV," Nat. Commun. **11** (2020) 1620

52. R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," Lancet **395** (2020) 10224

53. D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," Science **367** (2020) 6483

54. C.A. de Haan, L. Kuo, P.S. Masters, H. Vennema, P.J. Rottier, "Coronavirus particle assembly: primary structure requirements of the membrane protein," J. Virol. **72** (1998) 6838-50

55. S.M. Lokman, M. Rasheduzzaman, A. Salauddin, R. Barua, A.Y. Tanzina, M.H. Rumi, M.I. Hossain, A. Siddiki, A. Mannan, M.M. Hasan, "Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: A computational biology approach," Infect. Genet. Evol. **84** (2020) 104389

56. S. Kumar, V.K. Maurya, A.K. Prasad, M.L.B. Bhatt, S.K. Saxena, "Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV)," Virusdisease **31** (2020) 13-21

57. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, "Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor," Nature **581** (2020) 215-220

58. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, "Site-specific glycan analysis of the SARS-CoV-2 spike," Science **369** (2020), 330-333

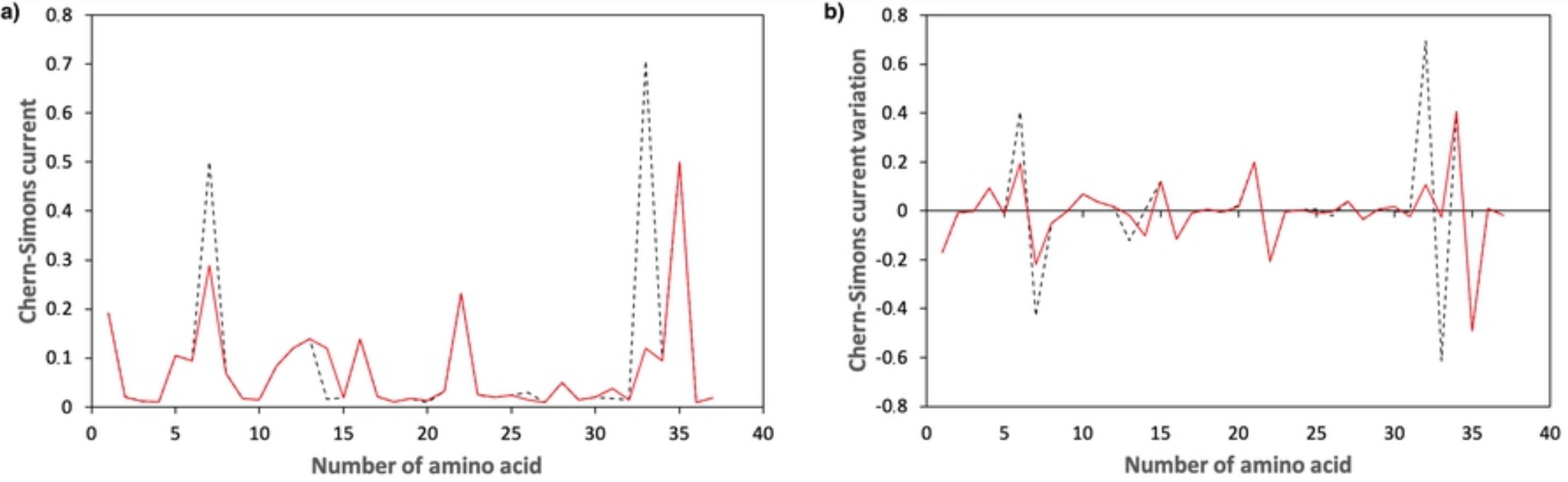

a)
Chern-Simons current
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
Number of amino acid
b)
Chern-Simons current variation
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
5
10
15
20
25
30
35
40
Number of amino acid

Figure 1

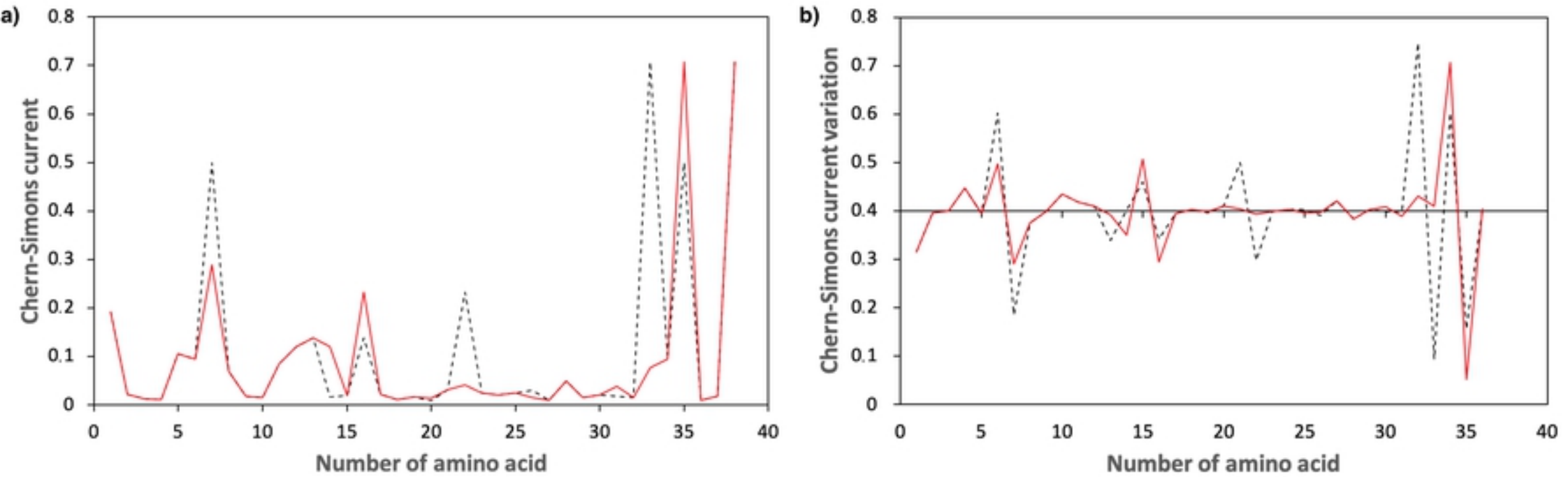
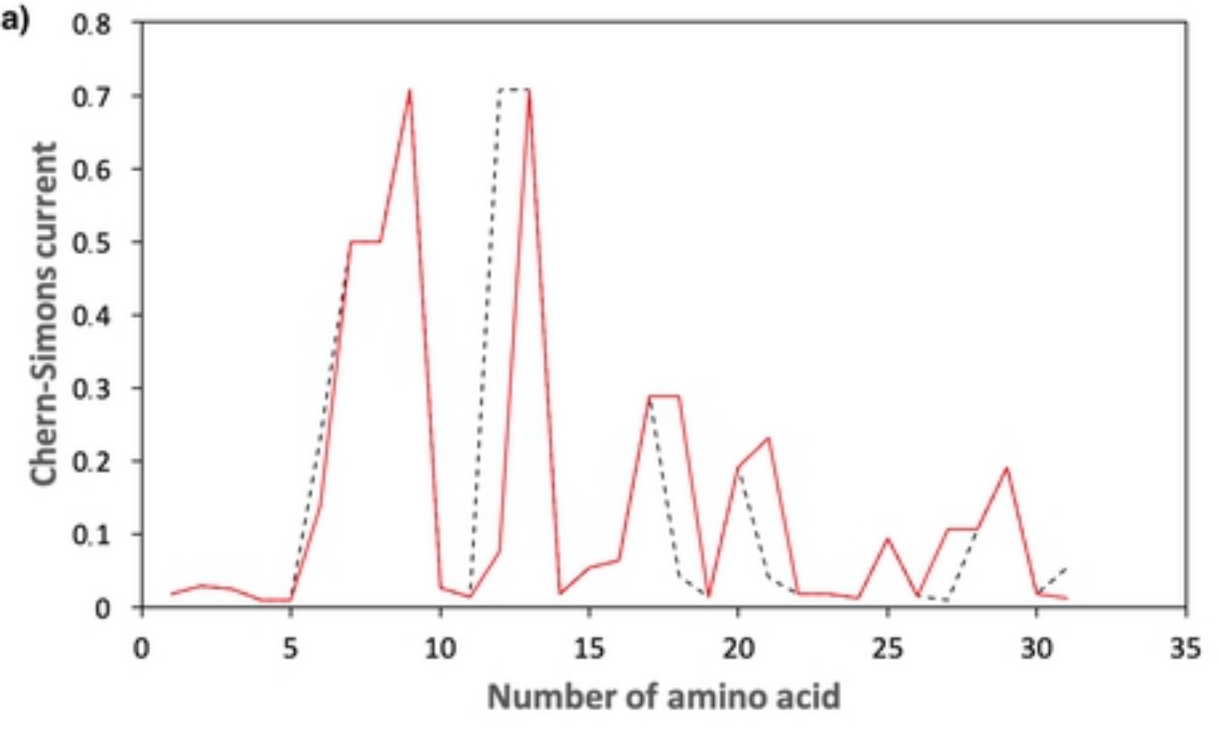

a)
Chern-Simons current
Number of amino acid
b)
Chern-Simons current variation
Number of amino acid

Figure 2


a)
Chern-Simons current
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
5
10
15
20
25
30
35
Number of amino acid

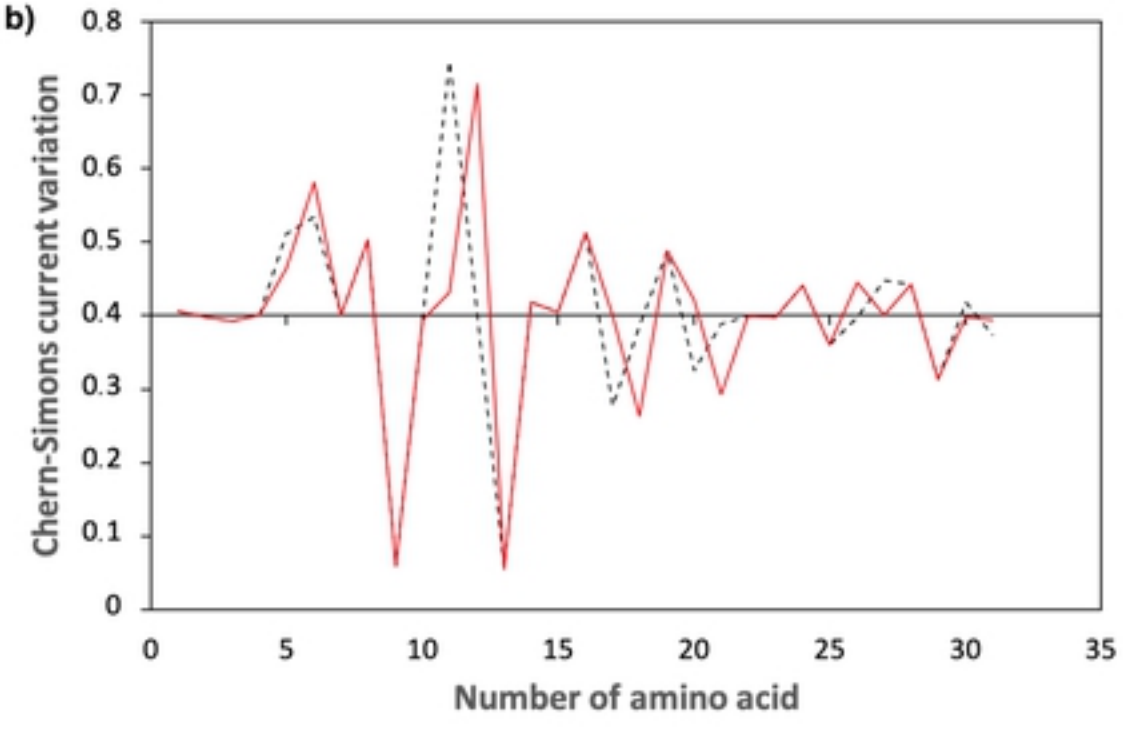
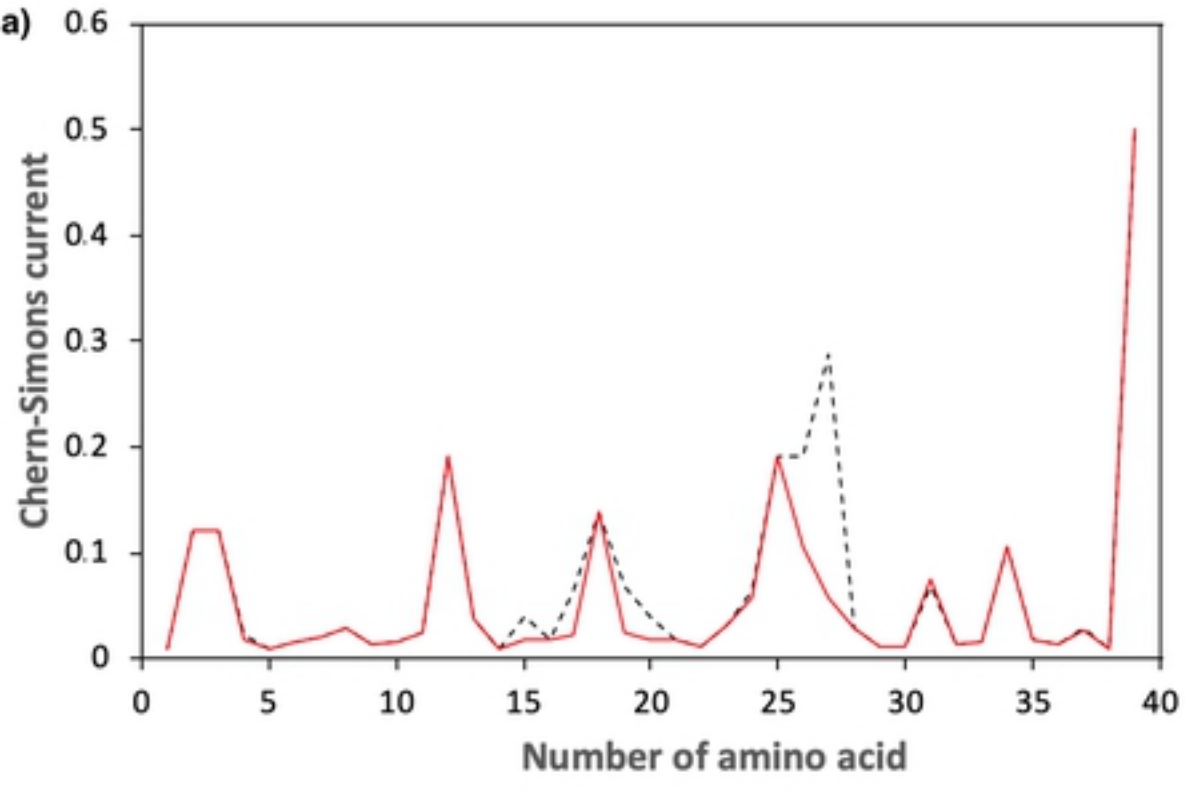

b)
Chern-Simons current variation
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
5
10
15
20
25
30
35
Number of amino acid

Figure 3


a)
Chern-Simons current
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40
Number of amino acid

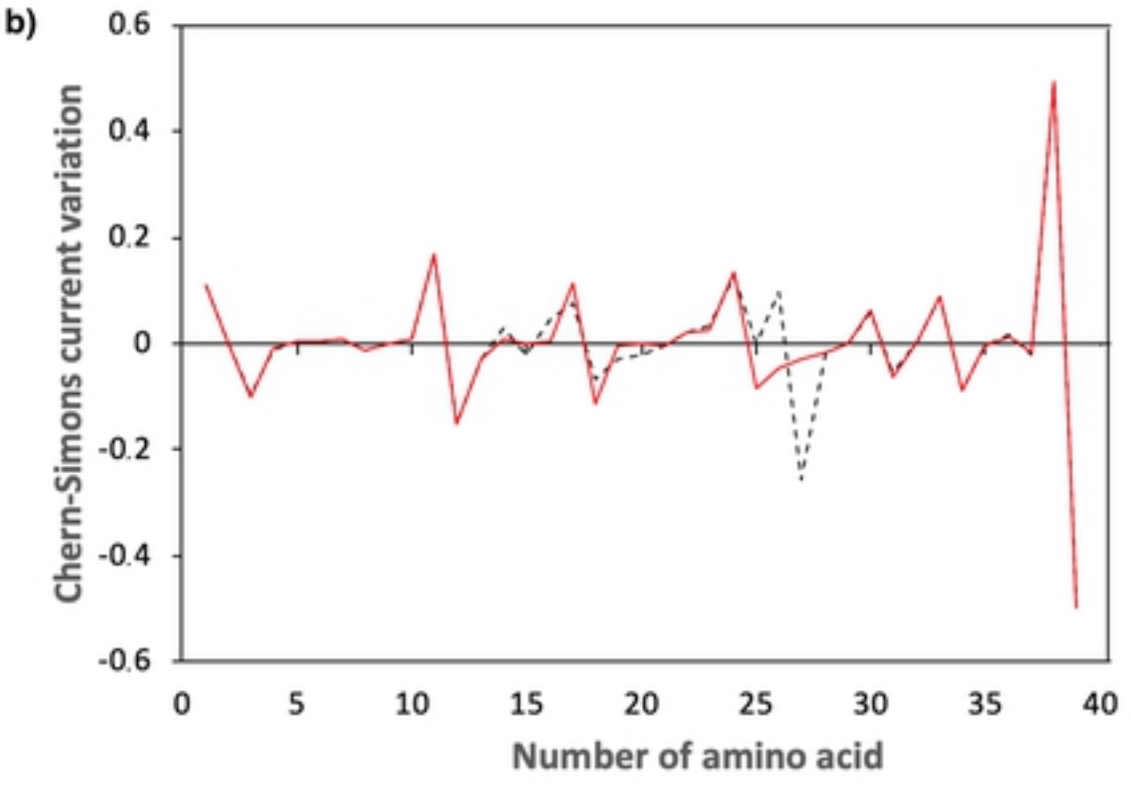

b)
Chern-Simons current variation
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
5
10
15
20
25
30
35
40
Number of amino acid

Figure 4

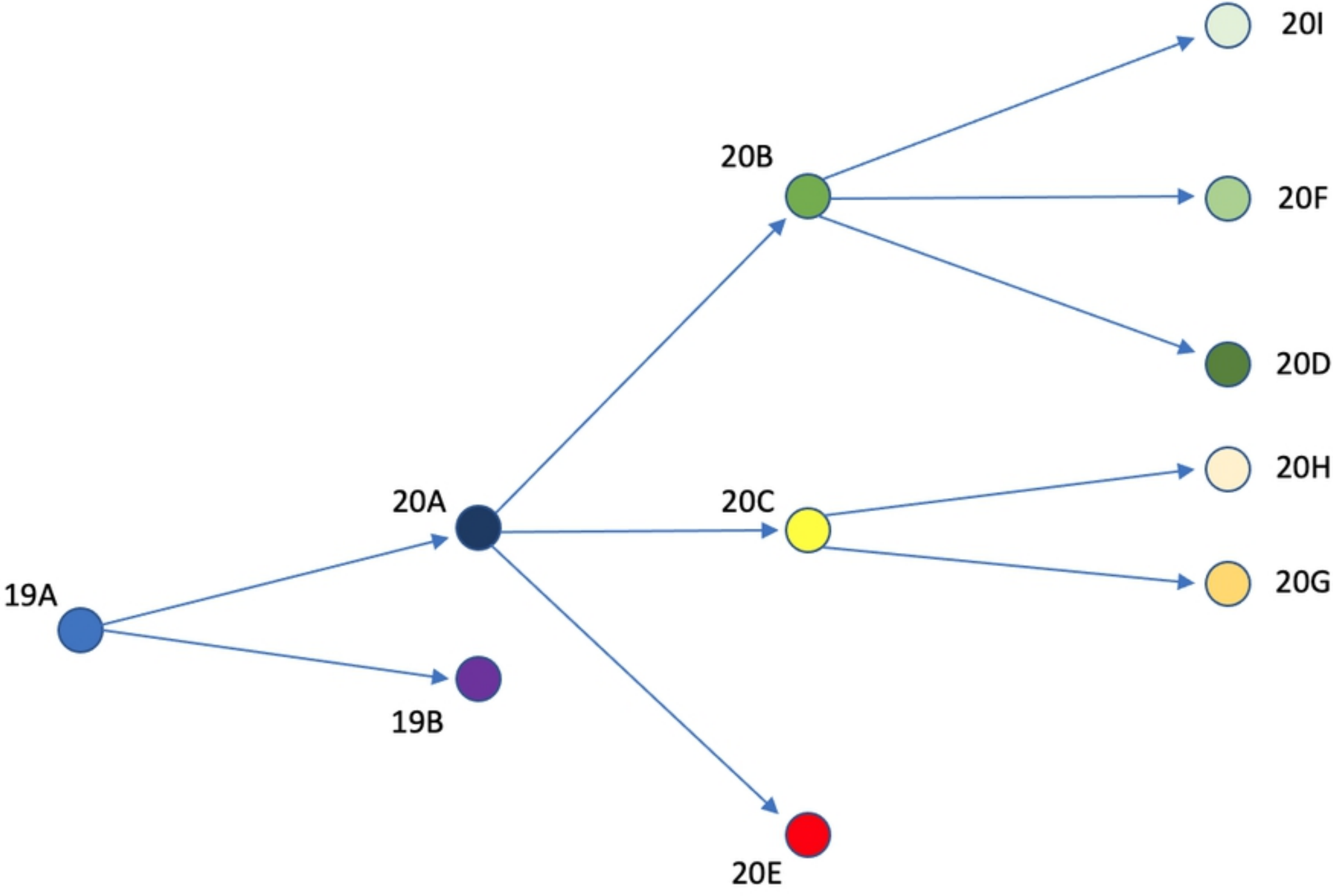


Figure 5

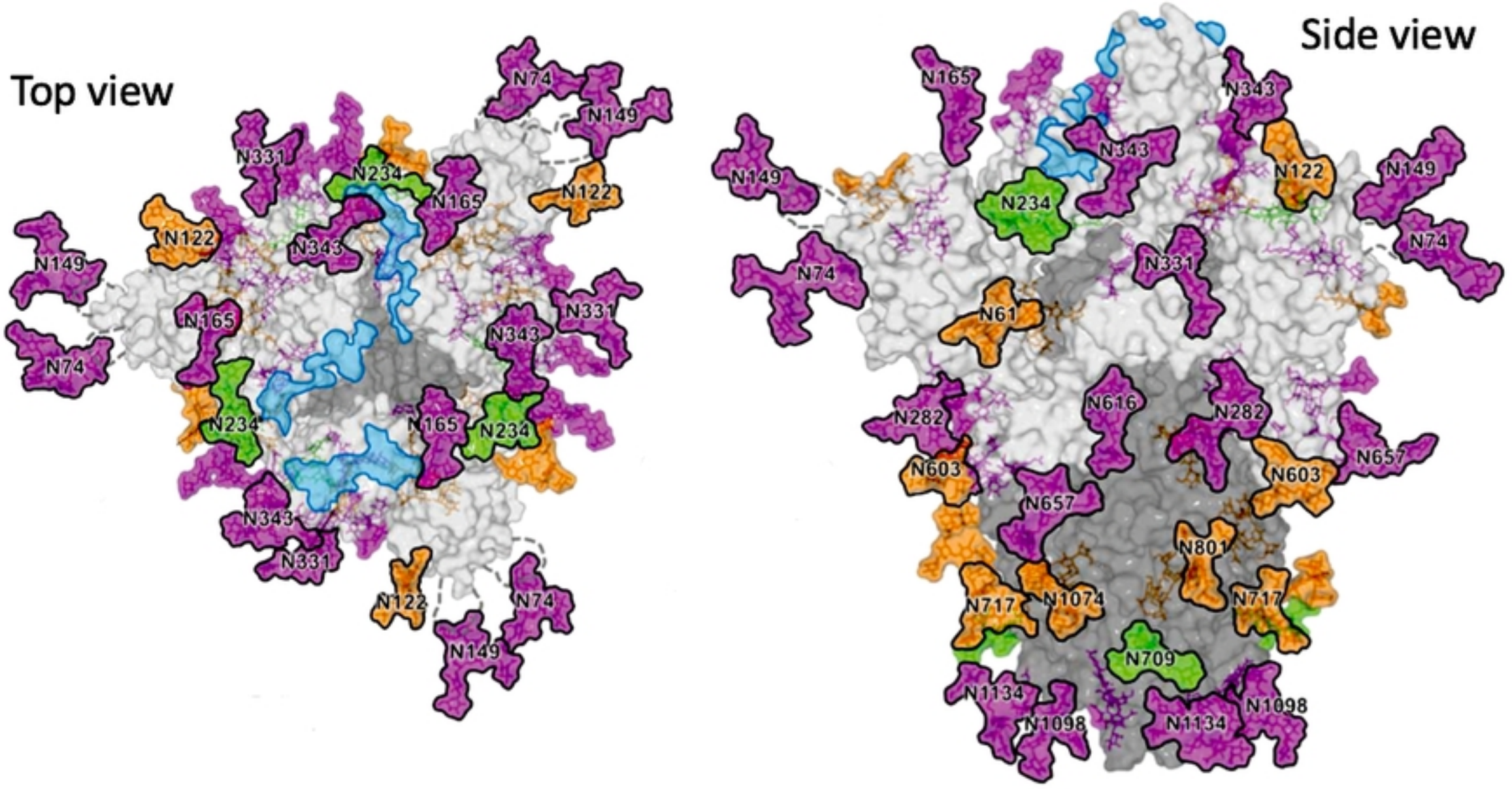

Top view
N74
N149
N331
N234
N165
N122
N122
N343
N149
N165
N331
N74
N343
N234
N165
N234
N343
N331
N122
N74
N149
Side view
N165
N343
N343
N122
N149
N149
N234
N74
N74
N331
N61
N282
N616
N282
N657
N603
N603
N657
N801
N717
N1074
N717
N709
N1134
N1098
N1134
N1098

Figure 6

Position	19A Triplet	19A Current	19B Triplet	19B Current	Current Variation (%)
2840	***AGC***	0.0096	***AG T***	0.0098	2
3607	TTT	0.7071	TTG	0.2887	-59
5829	CTG	0.1382	TTG	0.2887	109
5866	ATG	0.0841	GTG	0.0579	-31
5933	TCT	0.0534	TTT	0.7071	1224
9294	TTT	0.7071	TTC	0.5	-29
9697	***TAT***	0.0214	***TA C***	0.0205	-4
9762	TAC	0.0205	CAC	0.0175	-15

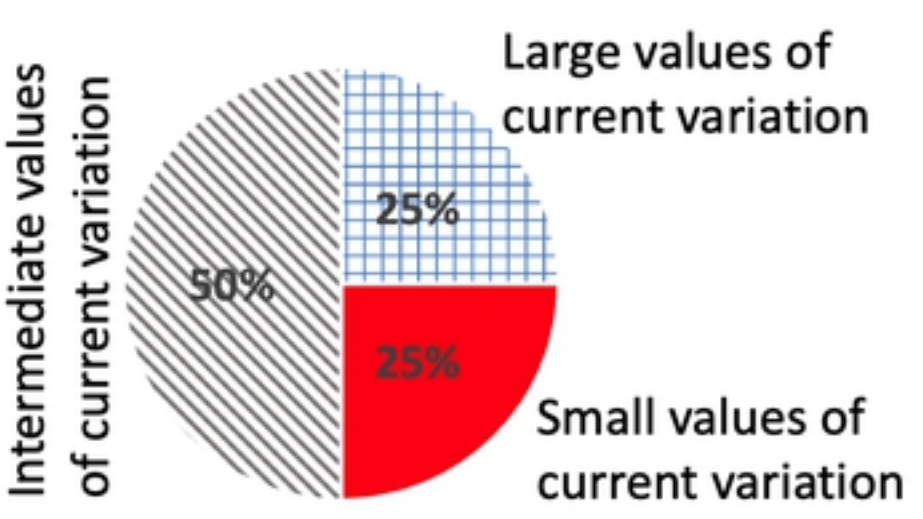


Figure 7

Position	19A Triplet	19A Current	20A Triplet	20A Current	Current Variation (%)
925	TTC	0.5	TTT	0.7071	41
3607	TTT	0.7071	TTG	0.2887	-59
3840	AAA	0.0147	AAG	0.0142	-3
4716	***CTA***	0.1612	***TTA***	0.3717	131
7714	***GAT***	0.0138	***GGT***	0.0089	-36
8847	CAG	0.0163	GAG	0.0126	-23
9697	TAT	0.0214	TAC	0.0205	-4
9762	TAC	0.0205	CAC	0.0175	-15

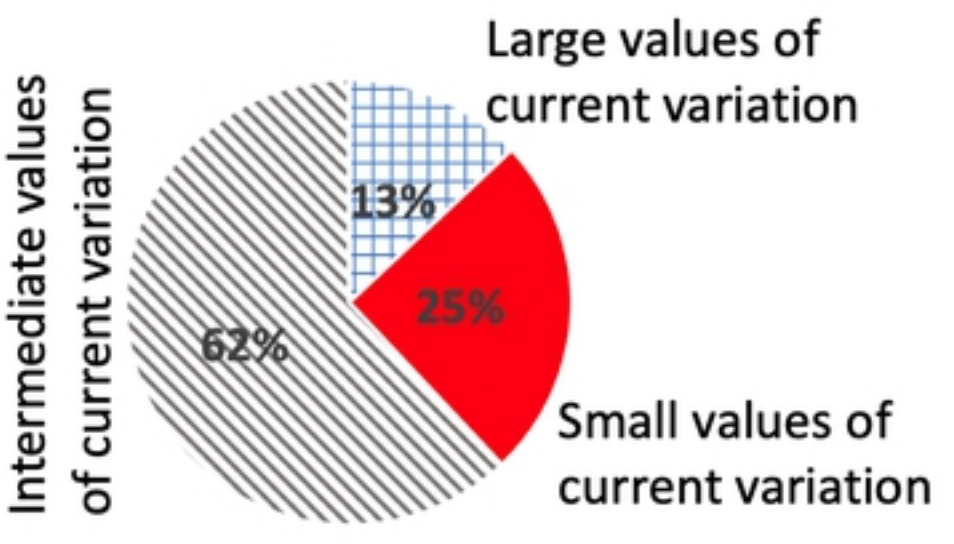


Figure 8

Position	20A Triplet	20A Current	20B Triplet	20B Current	Current Variation (%)
3840	AAG	0.0142	AAA	0.0147	4
6590	CCC	0.0377	CTC	0.1913	407
8036	GAC	0.0134	GAT	0.0138	3
8847	GAG	0.0126	CAG	0.0163	29
9516	GGC	0.0087	GGT	0.0089	2
9540	***AGG***	0.0091	***AAA***	0.0147	62
9541	GGA	0.0085	CGA	0.0103	21

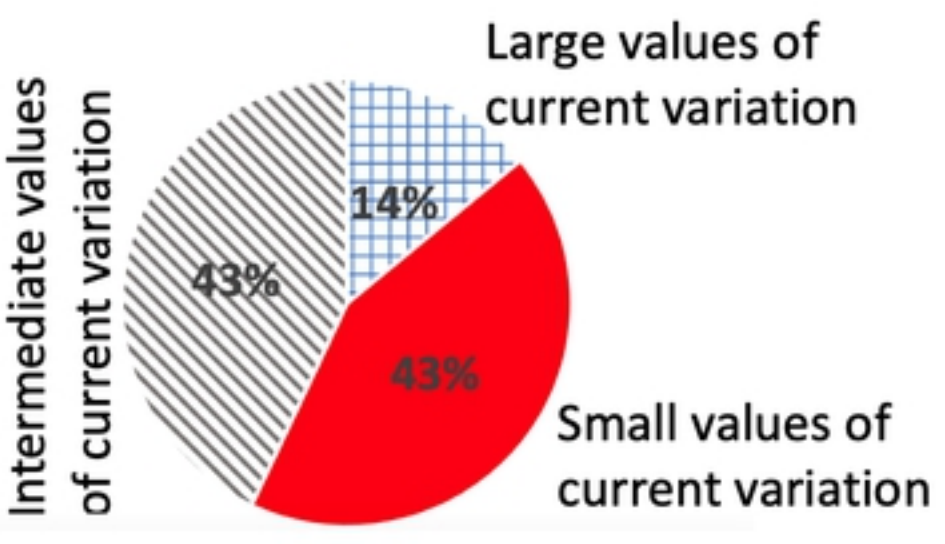


Figure 9

Position	20A Triplet	20A Current	20C Triplet	20C Current	Current Variation (%)
266	***ACC***	0.0299	***ATC***	0.1057	254
2130	GCT	0.0257	GTT	0.0759	195
3840	AAG	0.0142	AAA	0.0147	4
6098	ACA	0.0284	ATA	0.0939	231
6161	ACG	0.027	ATG	0.0841	211
6773	TGA	0.0115	TTA	0.3717	3132
8434	***AGA***	0.0093	***ATA***	0.0939	910
8437	CCA	0.0354	CTA	0.1612	355
8847	GAG	0.0126	CAG	0.0163	29

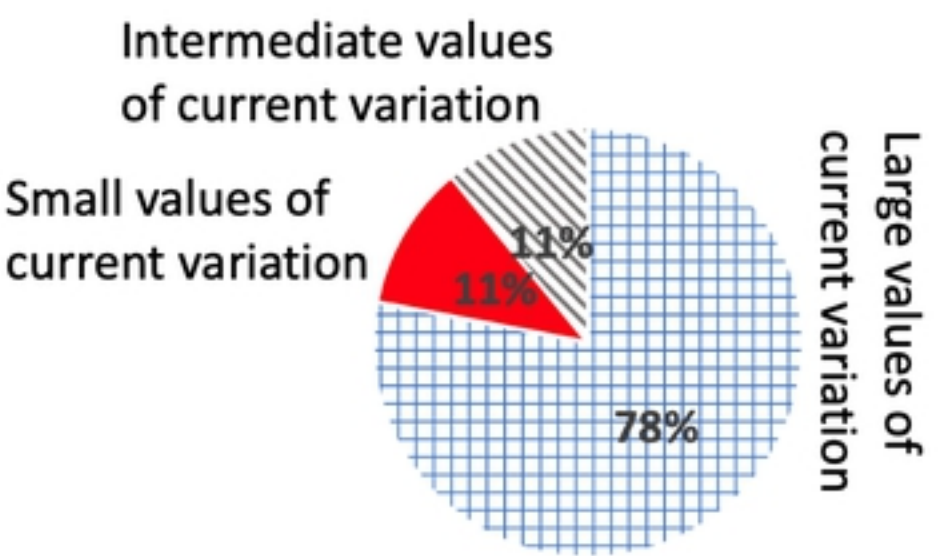


Figure 10

Position	20A Triplet	20A Current	20E Triplet	20E Current	Current Variation (%)
61	GTT	0.0759	GTC	0.069	-9
2008	ACC	0.0299	ACT	0.0316	6
3840	AAG	0.0142	AAA	0.0147	4
6800	CTA	0.1612	CCA	0.0354	-78
6998	CGT	0.0109	CCT	0.0402	269
7069	TGA	0.0115	TTA	0.3717	3132
7322	***GCT***	0.0257	***GTT***	0.0759	195
9546	AGA	0.0093	AAA	0.0147	58
9557	GCT	0.0257	GTT	0.0759	195
9708	GAC	0.0134	GAT	0.0138	3
9795	GTA	0.063	TTA	0.3717	490

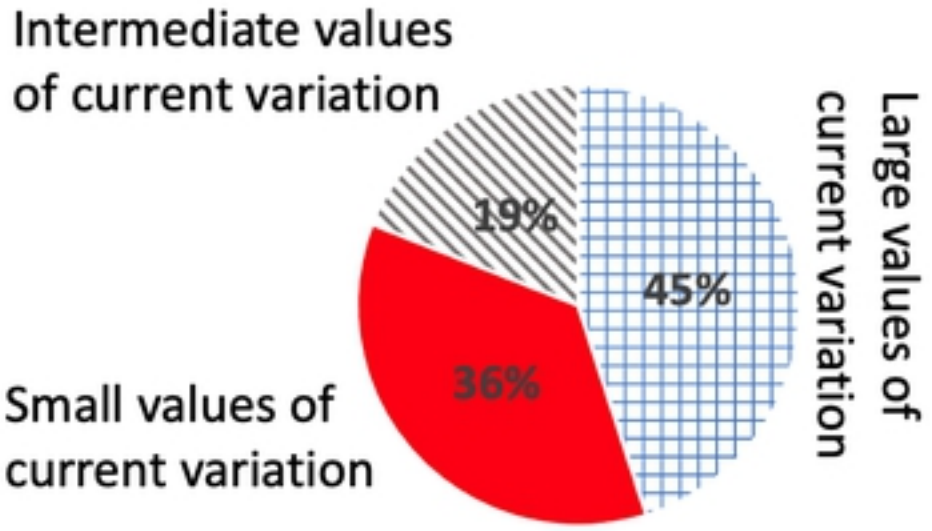


Figure 11

Position	20B Triplet	20B Current	20D Triplet	20D Current	Current Variation (%)
1247	***ACT***	0.0316	***ATT***	0.1201	280
1306	AAG	0.0142	AAT	0.0157	11
2148	AAC	0.0152	AAT	0.0157	3
3279	***GGT***	0.0089	***AGT***	0.0098	10
3892	TGT	0.0122	TGC	0.0118	-3
4425	***ACA***	0.0284	***ATA***	0.0939	231
4993	ACA	0.0284	ATA	0.0939	231
6299	ATT	0.1201	ACT	0.0316	-74
6479	TCA	0.046	TCG	0.0429	-7
6590	CTC	0.1913	CCC	0.0377	-80
7120	ACC	0.0299	ATC	0.1057	254
7823	***ACC***	0.0299	***ACT***	0.0316	6
8036	GAT	0.0138	GAC	0.0134	-3
8081	CTT	0.2319	CTC	0.1913	-18
9516	GGT	0.0089	GGC	0.0087	-2
9571	ATG	0.0841	ATT	0.1201	43

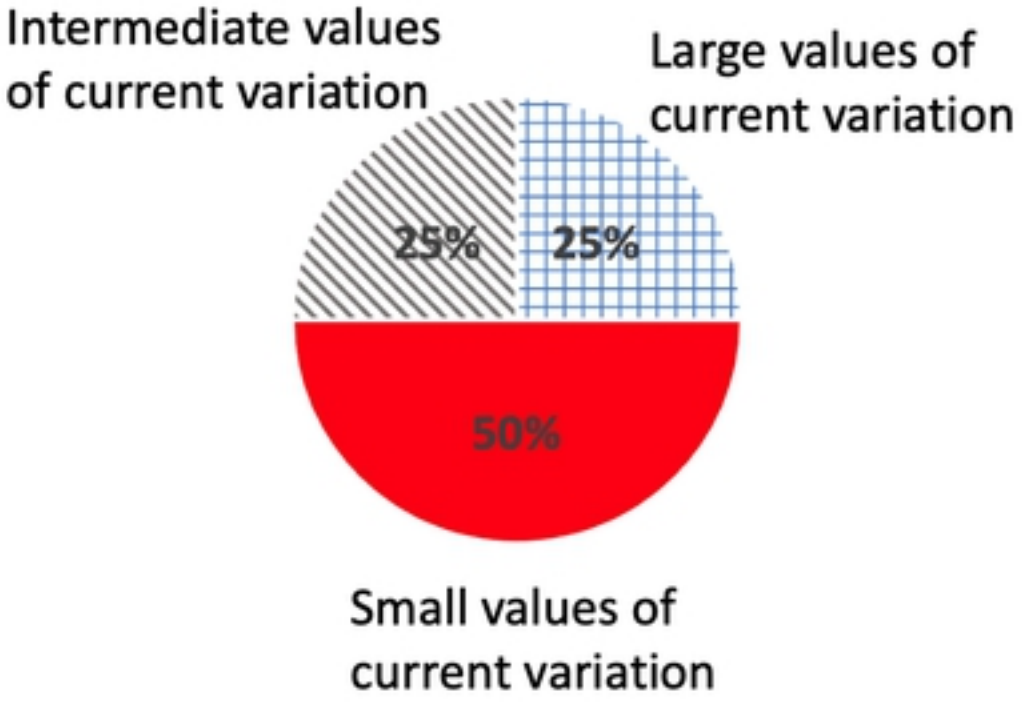


Figure 12

Position	20B Triplet	20B Current	20F Triplet	20F Current	Current Variation (%)
301	***ATT***	0.1201	***TTT***	0.7071	489
2426	***ACT***	0.0316	***ACC***	0.0299	-5
5462	***CGC***	0.0106	***CTC***	0.1913	1705
6098	***ACA***	0.0284	***ATA***	0.0939	231
6590	CTC	0.1913	CCC	0.0377	-80
7577	***AGC***	0.0096	***AAC***	0.0152	58
7713	***CAG***	0.0163	***CAA***	0.0169	4
8036	GAT	0.0138	GAC	0.0134	-3
9516	GGT	0.0089	GGC	0.0087	-2

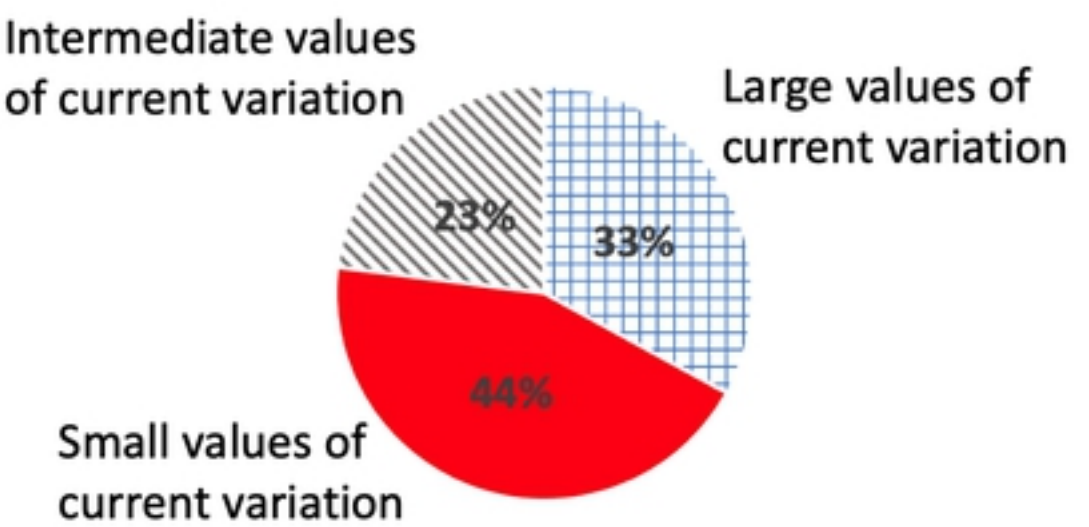


Figure 13

Position	20B Triplet	20B Current	20I Triplet	20I Current	Current Variation (%)
217	TCC	0.0495	TCT	0.0534	8
1002	ACT	0.0316	ATT	0.1201	280
1668	ATT	0.1201	ATA	0.0939	-22
1709	GCT	0.0257	GAT	0.0138	-46
1908	TTC	0.5	TTT	0.7071	41
2231	ATA	0.0939	ACA	0.0284	-70
4805	***CCG***	0.0334	***CTG***	0.1382	314
5006	***ACC***	0.0299	***ATC***	0.1057	254
5305	CTT	0.2319	CCT	0.0402	-83
5785	AGC	0.0096	GGC	0.0087	-9
6590	CTC	0.1913	CCC	0.0377	-80
7170	GTC	0.069	ATC	0.1057	53
7601	***AAT***	0.0157	***TAT***	0.0214	36
7670	GCT	0.0257	GAT	0.0138	-46
7781	CCT	0.0402	CAT	0.0182	-55
7816	ACA	0.0284	ATA	0.0939	231
8036	GAT	0.0138	GAC	0.0134	-3
8082	TCA	0.046	GCA	0.0234	-49
8218	GAC	0.0134	CAC	0.0175	31
9237	TCA	0.046	TTA	0.3717	708
9262	TAG	0.0189	TAT	0.0214	13
9283	GTA	0.063	GTG	0.0579	-8

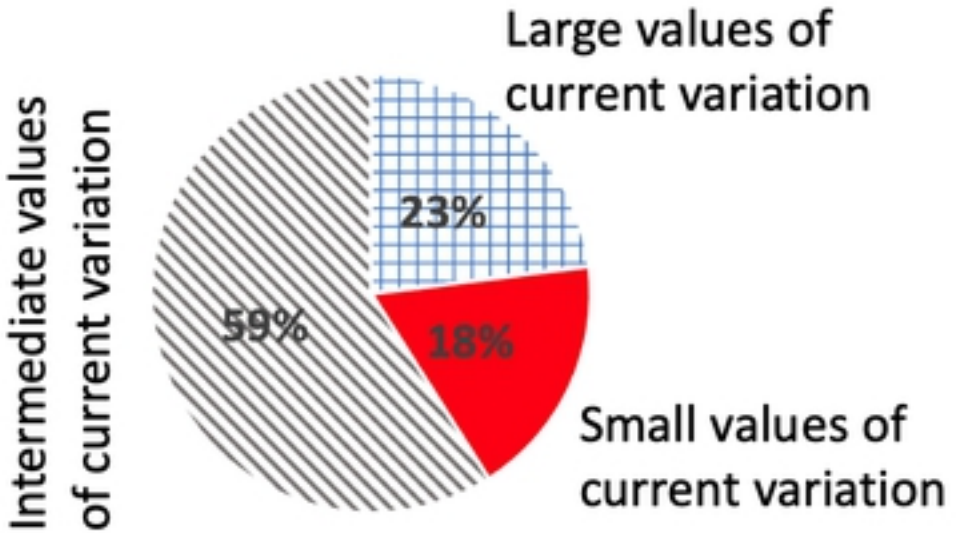


Figure 14

Position	20C Triplet	20C Current	20G Triplet	20G Current	Current Variation (%)
220	CTG	0.1382	TTG	0.2887	109
555	ACT	0.0316	ACC	0.0299	-5
1323	ACA	0.0284	ACC	0.0299	5
1978	CCC	0.0377	CCT	0.0402	7
2130	GTT	0.0759	GCT	0.0257	-66
3353	***CTT***	0.2319	***TTT***	0.7071	205
4236	AAC	0.0152	AAT	0.0157	3
5168	TAG	0.0189	TAT	0.0214	13
6054	***CTA***	0.1612	***CTG***	0.1382	-14
6092	TTG	0.2887	TTT	0.7071	145
6098	ATA	0.0939	ACA	0.0284	-70
6129	CTG	0.1382	TTG	0.2887	109
6161	ATG	0.0841	ACG	0.027	-68
6773	TTA	0.3717	TGA	0.0115	-97
7331	ATA	0.0939	ATT	0.1201	28
7620	GCA	0.0234	TCA	0.046	97
8437	CTA	0.1612	CCA	0.0354	-78
8549	***GTG***	0.0579	***TTG***	0.2887	399
8556	TTT	0.7071	TTC	0.5	-29
8897	GAT	0.0138	TAT	0.0214	55
9234	***GTC***	0.069	***GTT***	0.0759	10
9404	CCT	0.0402	TCT	0.0534	33
9536	***CCA***	0.0354	***CTA***	0.1612	355
9726	***CAG***	0.0163	***CTG***	0.1382	748

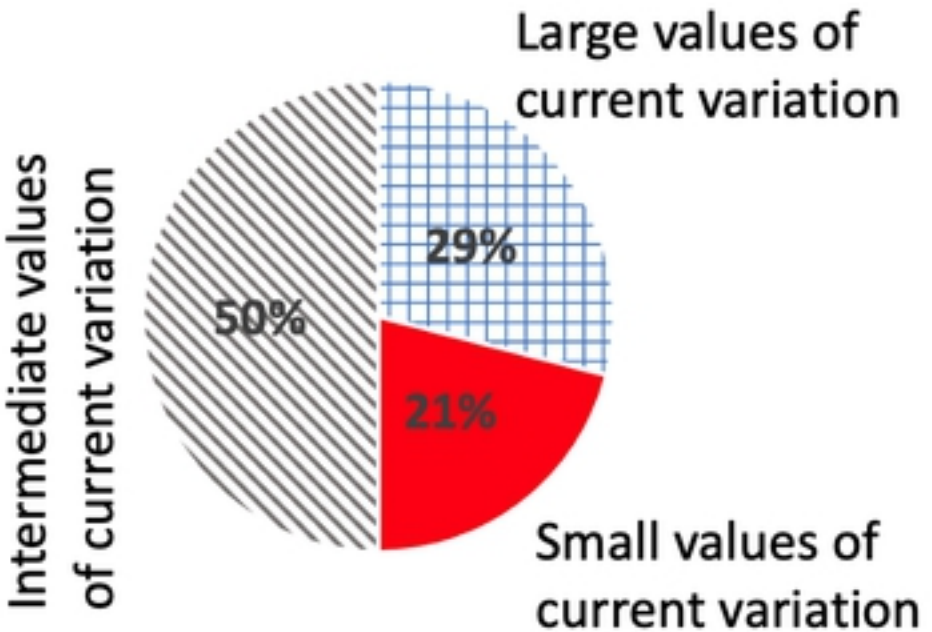


Figure 15

Position	20C Triplet	20C Current	20H Triplet	20H Current	Current Variation (%)
810	ACA	0.0284	ACT	0.0316	11
1656	AAG	0.0142	AAT	0.0157	11
2047	CCA	0.0354	CTA	0.1612	355
2130	GTT	0.0759	GCT	0.0257	-66
2454	GTT	0.0759	GTC	0.069	-9
2597	AAT	0.0157	AGT	0.0098	-38
3354	AAG	0.0142	AGG	0.0091	-36
3378	GTG	0.0579	GTT	0.0759	31
3451	GAC	0.0134	GAT	0.0138	3
5475	GTA	0.063	GTG	0.0579	-8
6098	ATA	0.0939	ACA	0.0284	-70
6161	ATG	0.0841	ACG	0.027	-68
6398	CGT	0.0109	CAT	0.0182	67
6773	TTA	0.3717	TGA	0.0115	-97
7118	CTT	0.2319	TTT	0.7071	205
7167	GCT	0.0257	GTT	0.0759	195
7180	GAT	0.0138	GCT	0.0257	86
7315	GAT	0.0138	GGT	0.0089	-36
7517	AAG	0.0142	AAT	0.0157	11
7584	***GAA***	0.0129	***AAA***	0.0147	14
7601	***AAT***	0.0157	***TAT***	0.0214	36
7801	GCA	0.0234	GTA	0.063	169
8437	CTA	0.1612	CCA	0.0354	-78
8548	CAG	0.0163	TAG	0.0189	16
8732	CTG	0.1382	TTG	0.2887	109
9331	CAT	0.0182	TAT	0.0214	18
9542	ACT	0.0316	ATT	0.1201	280
9809	AGT	0.0098	ATT	0.1201	1126

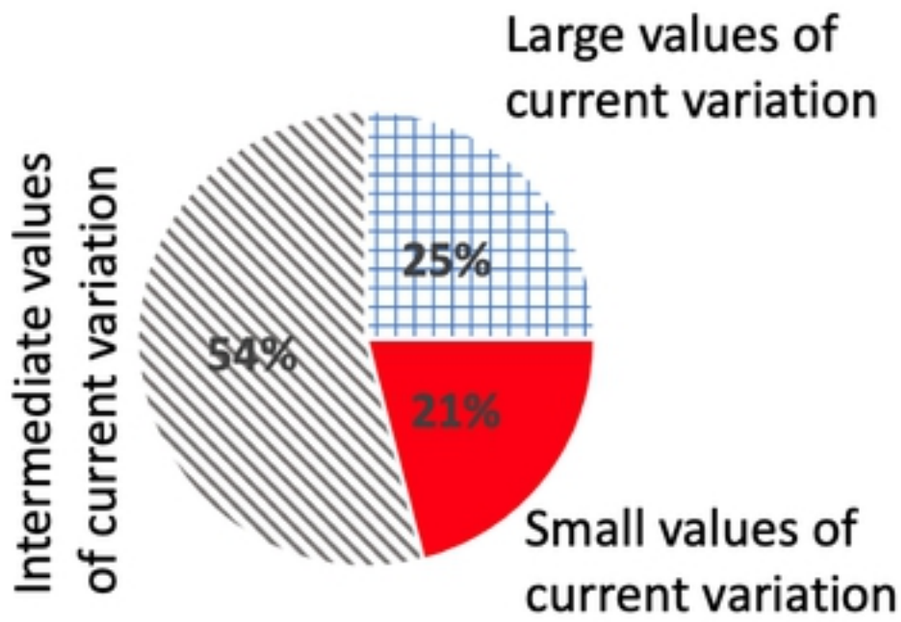


Figure 16

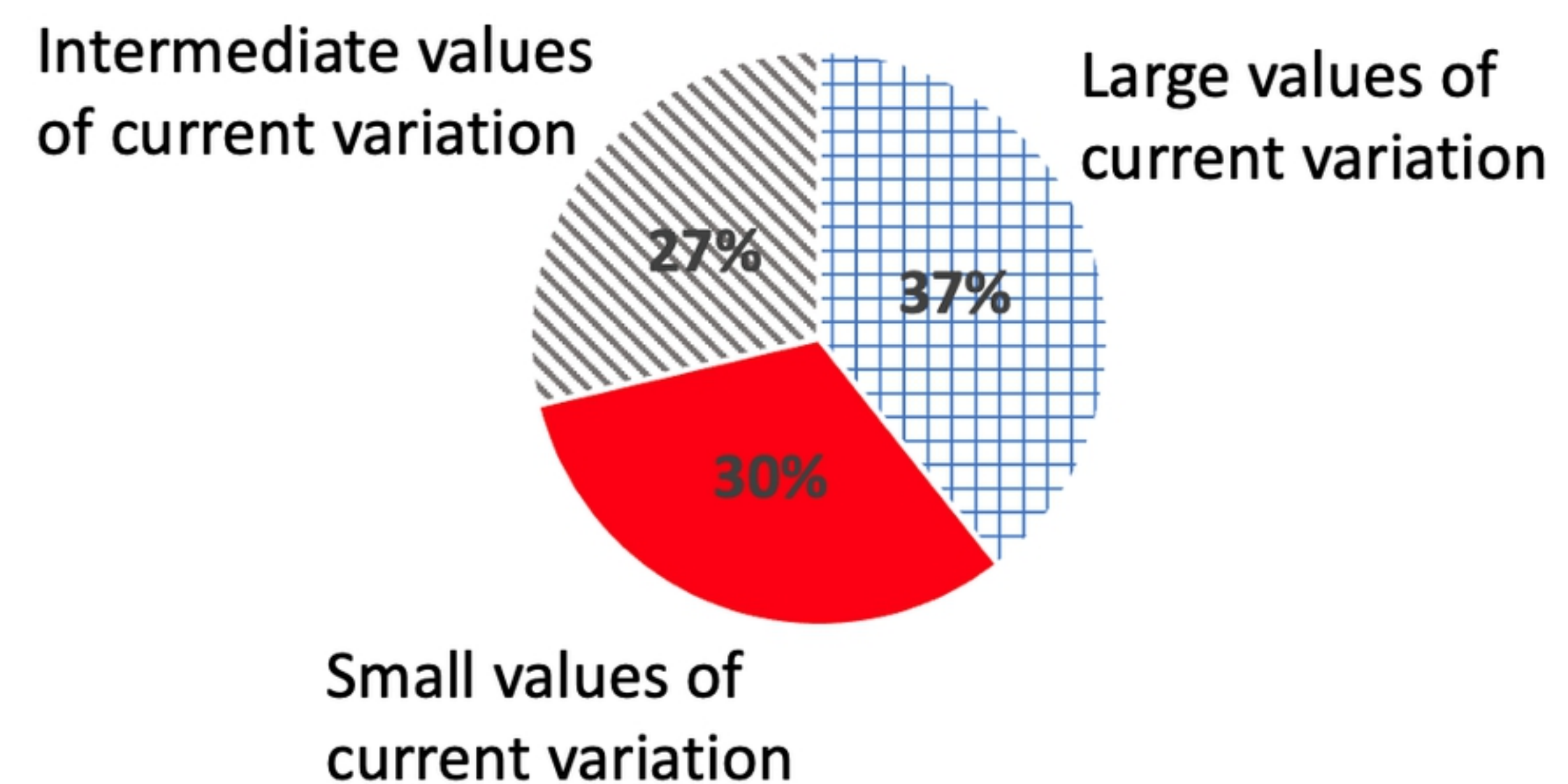

Intermediate values of current variation
Large values of current variation
27%
37%
30%
Small values of current variation

Figure 17

Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current
Phe (UUU)	0.7071	Ser (UCU)	0.0534	Tyr (UAU)	0.0214	Cys (UGU)	0.0122
Phe (UUC)	0.5000	Ser (UCC)	0.0495	Tyr (UAC)	0.0205	Cys (UGC)	0.0118
Leu (UUA)	0.3717	Ser (UCA)	0.0460	Sto (UAA)	0.0197	Sto (UGA)	0.0115
Leu (UUG)	0.2887	Ser (UCG)	0.0429	Sto (UAG)	0.0189	Trp (UGG)	0.0112
Leu (CUU)	0.2319	Pro (CCU)	0.0402	His (CAU)	0.0182	Arg (CGU)	0.0109
Leu (CUC)	0.1913	Pro (CCC)	0.0377	His (CAC)	0.0175	Arg (CGC)	0.0106
Leu (CUA)	0.1612	Pro (CCA)	0.0354	Gin (CAA)	0.0169	Arg (CGA)	0.0103
Leu (CUG)	0.1382	Pro (CCG)	0.0334	Gin (CAG)	0.0163	Arg (CGG)	0.0010
Ile (AUU)	0.1201	Thr (ACU)	0.0316	Asn (AAU)	0.0157	Ser (AGU)	0.0098
Ile (AUC)	0.1057	Thr (ACC)	0.0299	Asn (AAC)	0.0152	Ser (AGC)	0.0096
Ile (AUA)	0.0939	Thr (ACA)	0.0284	Lys (AAA)	0.0147	Arg (AGA)	0.0093
Met (AUG)	0.0841	Thr (ACG)	0.0270	Lys (AAG)	0.0142	Arg (AGG)	0.0091
Val (GUU)	0.0759	Ala (GCU)	0.0257	Asp (GAU)	0.0138	Gly (GGU)	0.0089
Val (GUC)	0.0690	Ala (GCC)	0.0245	Asp (GAC)	0.0134	Gly (GGC)	0.0087
Val (GUA)	0.0630	Ala (GCA)	0.0234	Glu (GAA)	0.0129	Gly (GGA)	0.0085
Val (GUG)	0.0579	Ala (GCG)	0.0224	Glu (GAG)	0.0126	Gly (GGG)	0.0083

Table 1

Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current	Amino acid	CS Current
Phe (F)	0.60355	Ser (S)	0.0352	His (H)	0.01785	Giu (E)	0.01275
Leu (L)	0.2305	Pro (P)	0.036675	Gin (Q)	0.0166	Cys (C)	0.012
Ile (I)	0.106567	Thr (T)	0.029225	Asn (N)	0.01545	Trp (W)	0.0112
Met (M)	0.0841	Ala (A)	0.024	Lys (K)	0.01445	Arg (R)	0.01005
Val (V)	0.06645	Tyr (Y)	0.02095	Asp (D)	0.0136	Gly (G)	0.0086

Table 2

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,245,279	UAU	UAA	Y	Stop	0.0214	0.0197	-8
25,245,294	UUC	UUG	F	L	0.5	0.2887	-42
25,245,314	AAU	AUU	N	I	0.0157	0.1201	665
25,245,332	GGC	GAC	G	D	0.0087	0.0134	54
25,245,342	GCC	GCU	A	A	0.0245	0.0257	5
25,245,350	ACC	AAC	T	N	0.0299	0.0152	-49
25,245,365	CAC	CCC	H	P	0.0175	0.0377	115
25,245,370	UUU	AUU	F	I	0.7071	0.1201	-83

Table 3

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,245,294	UUC	UUG	F	L	0.5	0.2887	-42
25,245,314	AAU	AUU	N	I	0.0157	0.1201	665
25,245,321	CUG	CUU	L	L	0.1382	0.2319	68
25,245,332	GGC	GAC	G	D	0.0087	0.0134	54
25,245,338	CUU	CCU	L	P	0.2319	0.0402	-83
25,245,350	ACC	AAC	T	N	0.0299	0.0152	-49
25,245,365	CAC	CCC	H	P	0.0175	0.0377	115
25,245,370	UUU	GUU	F	V	0.7071	0.0759	-89
25,245,378	UUC	UUU	F	F	0.5	0.7071	41

Table 4

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,215,485	CUU	CUG	L	L	0.2319	0.1382	-40
25,215,501	UUU	GUU	F	V	0.7071	0.0759	-89
25,215,520	UCG	UUG	S	L	0.0429	0.2887	573
25,215,529	CCU	CUU	P	L	0.0402	0.2319	477
25,215,539	UGU	UGC	C	C	0.0122	0.0118	-3
25,215,547	AGC	AUC	S	I	0.0096	0.1057	1001
25,215,559	UCU	UGU	S	C	0.0534	0.0122	-77

Table 5

Position	Ref. Base	Mutation	Ref. Amino	Mutation	Initial CS current	Mutated CS current	Current Variation (%)
25,227,272	UAU	CAU	Y	H	0.0214	0.0182	-15
25,227,306	CCU	CAU	P	H	0.0402	0.0182	-55
25,227,312	GUA	GCA	V	A	0.063	0.0234	-63
25,227,318	GUC	GCC	V	A	0.069	0.0245	-64
25,227,321	CCU	CAU	P	H	0.0402	0.0182	-55
25,227,334	GUA	GUG	V	V	0.063	0.0579	-8
25,227,338	CUC	AUC	L	I	0.1913	0.1057	-45
25,227,341	UUG	GUG	L	V	0.2887	0.0579	-80
25,227,355	GUC	GUU	V	V	0.069	0.0759	10
25,227,373	ACA	ACG	T	T	0.0284	0.027	-5

Table 6

19A Mutations				
Position	19A Triplet	19A Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
925	TTC	0.5	836	-96
2840	***AGC***	0.0096	6213	70
3607	TTT	0.7071	0	-97
3840	AAA	0.0147	-97	1201
4716	***CTA***	0.1612	821	-90
5829	CTG	0.1382	362	68
5866	ATG	0.0841	436	-62
5933	TCT	0.0534	299	596
7714	***GAT***	0.0138	-15	450
8847	CAG	0.0163	-37	889
9294	TTT	0.7071	412	-96
9697	TAT	0.0214	-9	-31
9762	TAC	0.0205	53	39

Table 7

20A Mutations				
Position	20A Triplet	20A Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
61	GTT	0.0759	772	280
266	***ACC***	0.0299	123	1572
2008	ACC	0.0299	90	-63
2130	GCT	0.0257	0	195
3840	AAG	0.0142	-97	1247
6098	ACA	0.0284	-51	-38
6161	ACG	0.027	-78	98
6590	CCC	0.0377	-55	67
6773	TGA	0.0115	-19	1302
6800	CTA	0.1612	927	-57
6998	CGT	0.0109	-87	67
7069	TGA	0.0115	-88	6049
7322	***GCT***	0.0257	-40	1346
8036	GAC	0.0134	-21	243
8434	***AGA***	0.0093	-98	141
8437	CCA	0.0354	-85	-58
8847	CAG	0.0163	-51	1179
9516	GGC	0.0087	0	13
9540	***AGG***	0.0091	-7	272
9546	AGA	0.0093	-64	804
9557	GCT	0.0257	-89	1023
9708	GAC	0.0134	-6	10
9795	GTA	0.063	273	-78

Table 8

20B Mutations				
Position	20B Triplet	20B Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
217	TCC	0.0495	-83	-74
301	***ATT***	0.1201	1191	-91
1002	ACT	0.0316	0	280
1247	***ACT***	0.0316	145	-55
1306	AAG	0.0142	-3	81
1668	ATT	0.1201	125	-88
1709	GCT	0.0257	0	-41
1908	TTC	0.5	2236	-94
2148	AAC	0.0152	-93	-3
2231	ATA	0.0939	-22	28
2426	***ACT***	0.0316	11	280
3279	***GGT***	0.0089	-29	37
3892	TGT	0.0122	-28	466
4425	***ACA***	0.0284	33	-71
4805	***CGC***	0.0334	130	89
4993	ACA	0.0284	0	-57
5006	***ACC***	0.0299	-84	1143
5305	CTT	0.2319	717	-93
5462	***CGC***	0.0106	-28	334
5785	AGC	0.0096	-90	3772
6098	***ACA***	0.0284	-51	-38
6299	ATT	0.1201	586	15
6479	TCA	0.046	-71	708
6590	CTC	0.1913	127	-67
7120	ACC	0.0299	5	-69
7170	GTC	0.069	279	-23
7577	***AGC***	0.0096	8	196
7601	***AAT***	0.0157	-50	-43
7670	GCT	0.0257	-79	-48
7713	***CAG***	0.0163	-24	-45
7781	CCT	0.0402	-25	-75
7816	ACA	0.0284	-25	-45
7823	***ACC***	0.0299	-61	-5
8036	GAT	0.0138	-19	233
8081	CTT	0.2319	119	-80
8082	TCA	0.046	-80	-76
8218	GAC	0.0134	-53	13
9237	TCA	0.046	124	-38
9262	TAG	0.0189	97	-22
9283	GTA	0.063	37	-71
9516	GGT	0.0089	2	10
9571	ATG	0.0841	472	-37

Table 9

20C Mutations				
Position	20C Triplet	20C Current	Variation (%) with respect to previous position	Variation (%) with respect to subsequent position
220	CTG	0.1382	717	-90
555	ACT	0.0316	145	-19
810	ACA	0.0284	-55	-46
1323	ACA	0.0284	-20	-53
1656	AAG	0.0142	-3	-21
1978	CCC	0.0377	33	42
2047	CCA	0.0354	141	95
2130	GTT	0.0759	195	0
2454	GTT	0.0759	522	-79
2597	AAT	0.0157	-79	72
3353	***CTT***	0.2319	1533	-94
3354	AAG	0.0142	-94	435
3378	GTG	0.0579	26	542
3451	GAC	0.0134	-82	-32
4236	AAC	0.0152	-96	0
5168	TAG	0.0189	-73	-41
5475	GTA	0.0363	-25	358
6054	***CTA***	0.1612	821	-42
6092	TTG	0.2887	3004	-93
6098	ATA	0.0939	62	-81
6129	CTG	0.1382	931	-93
6161	ATG	0.0841	-30	-37
6398	CGT	0.0109	-93	478
6773	TTA	0.3717	2518	-57
7118	CTT	0.2319	1377	-88
7167	GCT	0.0257	41	265
7180	GAT	0.0138	-98	10
7315	GAT	0.0138	27	1286
7331	ATA	0.0939	165	-91
7517	AAG	0.0142	67	746
7584	***GAA***	0.0129	-83	-31
7601	AAT	0.0157	-50	-43
7620	GCA	0.0234	29	51
7801	GCA	0.0234	163	-45
8437	CTA	0.1612	-30	-91
8548	CAG	0.0163	-93	255
8549	***GTG***	0.0579	255	45
8556	TTT	0.7071	339	-80
8732	CTG	0.1382	-72	-24
8897	GAT	0.0138	13	191
9234	***GTC***	0.069	143	22
9331	CAT	0.0182	-97	786
9404	CCT	0.0402	-92	-74
9542	ACT	0.0316	272	69
9536	***CCA***	0.0354	12	-75
9809	AGT	0.0098	-91	491

Table 10

Position	19A		Variation (%) with respect to	Variation (%) with respect to
	Triplets	Current	the previous position	the subsequent position
17	AAT	0.0157	-79	1377
61	AAT	0.0157	-68	383
74	AAT	0.0157	-47	-43
122	AAC	0.0152	-3	69
149	AAC	0.0152	0	-3
165	AAT	0.0157	0	-25
234	AAC	0.0152	-87	595
282	AAT	0.0157	22	-46
331	AAT	0.0157	-61	665
343	AAC	0.0152	-98	61
603	AAT	0.0157	-45	101
616	AAC	0.0152	-80	-22
657	AAC	0.0152	-78	0
709	AAT	0.0157	-71	-3
717	AAT	0.0157	-45	4404
801	AAT	0.0157	-98	4404
1074	AAC	0.0152	7	3189
1098	AAT	0.0157	-66	-45
1134	AAC	0.0152	-78	0
1158	AAT	0.0157	11	16
1173	AAT	0.0157	-87	64
1194	AAT	0.0157	-96	-18

Table 11