

## Article

# The Promise of Mutation Resistant Drugs for SARS-CoV-2 That Interdict in the Folding of the Spike Protein Receptor Binding Domain

Fernando Bergasa-Caceres \* and Herschel A. Rabitz 

Department of Chemistry, Princeton University, Princeton, NJ 08544, USA; hrabitz@princeton.edu

\* Correspondence: bergasa@princeton.edu

**Abstract:** In recent work, we proposed that effective therapeutic drugs aimed at treating the SARS-CoV-2 infection could be developed based on interdicting in the early steps of the folding pathway of key viral proteins, including the receptor binding domain (RBD) of the spike protein. In order to provide for a drug target on the protein, the earliest contact-formation event along the dominant folding pathway of the RBD spike protein was predicted employing the Sequential Collapse Model (SCM). The segments involved in the predicted earliest contact were suggested to provide optimal folding interdiction target regions (FITRs) for potential therapeutic drugs, with a focus on folding interdicting peptides (FIPs). In this paper, we extend our analysis to include 13 known single mutations of the RBD spike protein as well as the triple mutation B1.351 and the recent double mutation B1.617.2. The results show that the location of the FITR does not change in any of the 15 studied mutations, providing for a mutation-resistant drug design strategy for the RBD-spike protein.

**Keywords:** SARS; CoV; COVID; therapeutic; peptide; design



**Citation:** Bergasa-Caceres, F.; Rabitz, H.A. The Promise of Mutation Resistant Drugs for SARS-CoV-2 That Interdict in the Folding of the Spike Protein Receptor Binding Domain. *COVID* **2021**, *1*, 288–302.

<https://doi.org/10.3390/covid1010023>

Academic Editor: Tohru Suzuki

Received: 21 July 2021

Accepted: 19 August 2021

Published: 24 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Now in its second year, the pandemic generated by the coronavirus SARS-CoV-2 has produced more than four million deaths, as well as massive economic and social damage [1–3]. A matter of current concern is the long-term suitability of the newly developed vaccines especially under continuing mutations of the coronavirus [4–6]. Many such mutations are already known [7,8] and in recent times, new strains combining several mutations and deletions were identified in the UK, South Africa and India with enhanced infectivity and, possibly, lethality [9–11]. The high rate of mutation, common in RNA viruses [12], raises the concern that the coronavirus SARS-CoV-2 might become endemic within the human population [9], and that more lethal variants might evolve over time. Thus, developing treatments whose efficacy is minimally affected by viral mutation has become a pressing goal.

In recent work, we proposed that effective therapeutic drugs aimed to treat SARS-CoV-2, could be designed by aiming to interdict in the early folding stages of the folding pathway of critical viral proteins [13]. In particular, we suggested that the folding of the receptor binding domain (RBD) of the spike protein of SARS-CoV-2, (i.e., a protein essential for viral entry in the cell through interaction with angiotensin-converting enzyme 2 (ACE2)) [14–18], could be pre-empted by employing specific peptide drugs [13]. Such peptide drugs can be designed relying on the predictions of the Sequential Collapse Model (SCM) for early folding contact formation [19,20]. The SCM allows for a prediction of the earliest contact-formation events along the folding pathway of globular proteins [19]. The earliest contacts along the multi-state folding pathway of proteins longer than ~100 amino acids, are called primary contacts in the SCM. These primary contacts are expected within the model to nucleate the subsequent folding steps, and thus should be key steps

along “Nature’s Shortcut” to protein folding [20]. The specific protein segments involved in these early primary contact-formation events can then be employed to identify folding interdiction target regions (FITRs), as well as the starting sequence to design possible folding interdicting peptide (FIP) drugs [13]. The FITR concept may also be receptive to other drugs besides peptides, although the latter choice is natural and remains the focus of this work.

The purpose of this paper is to show that the proposed FITR-based strategy for the design of therapeutic drugs is robust to a broad sample of the known mutations of the RBD-spike protein. In order to do so, the SCM will be applied to the sequences of 13 single position mutants of RBD-spike (V341I, A348T, N354D, D364Y, V367F, K378R, Q409E, A435S, W436R, G476S, V483A, Y508H and N501Y) as well as the triple mutation associated with the B.1.351 lineage (K417N, E484K and N501Y), and the double mutation known as the delta variant B.1.617.2 (E484Q and L432R). The goal of examining these mutants is to determine whether the location along the sequence of the FITR changes, and also if the predicted starting sequence of the FIPs should be altered as a result of any of the mutations. The FITR and FIP are generally found to be robust with only two mutations showing significant deviations in the stabilizing energy of the early contacts with respect to the predictions for the native sequence. These deviations, however, do not alter the predicted location of the FITR and have a limited effect on the possible FIP sequences.

The concepts presented in this paper are similar to earlier proposals to develop drugs that might be therapeutic by interfering with the folding process [21,22]. There has been a growing interest in the possibility of modulating the folding process through the employment of specific molecules, and theoretical and experimental results have been recently reported that support the view that targeting folding dynamics is a promising path for therapeutic development [23–25]. The main difference between the proposal presented here and other approaches is that the FITR strategy aims to interdict in the folding process while the protein is still completely unfolded [13]. Thus, the FITR strategy is particularly well suited to target proteins as they are translated by the ribosome as we will explain below. Due to this feature, the FITR strategy holds the promise of being a very effective viral suppressor. Moreover, if shown to properly function, the FITR strategy could be readily extended to other viruses.

## 2. Results

### 2.1. The Receptor Binding Domain of the Spike Protein of SARS-CoV-2

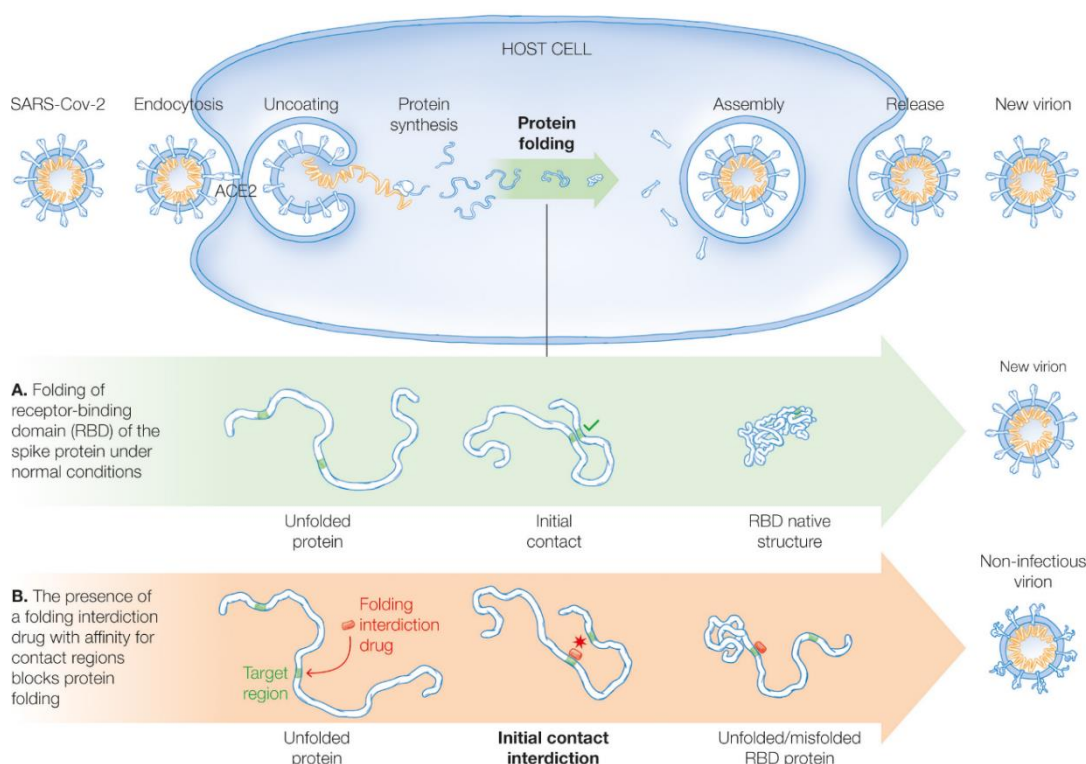
The single-strand RNA genome of SARS-CoV-2 encodes four structural proteins. These are the small protein (E), the nucleocapsid (N), the matrix protein (M) and the spike protein (S) [26–28]. The spike protein is a fusion I type protein that forms trimers on the protein surface [26–28]. It has two subunits: S1 and S2, responsible for receptor-binding and membrane fusion, respectively [26–28].

The receptor recognition mechanisms of coronaviruses have been extensively studied [14,15,28]. In particular, for SARS-CoV-2 [18] and its earlier viral variant SARS-CoV [29], entry into the host’s cells is mediated by the spike protein that includes a specific receptor-binding domain (RBD) [18,30]. The RBD recognizes the angiotensin-converting enzyme 2 (ACE2) as its receptor [15,16,18]. The structure of SARS-CoV RBD is well known [16], and the structure of SARS-CoV-2 RBD is similar [18], albeit with some specific mutations in the ACE2 binding ridge that enhance the ability of SARS-CoV-2 to bind human ACE2 [18]. Due to its role in receptor recognition and binding, the RBD of the spike protein is the main determinant of SARS-CoV-2 infectivity [16,31], and also the major antigen for immune response [32]. Thus, the RBD of the spike protein of SARS-CoV-2 is a natural therapeutic target given its critical biological role in facilitating the virus entry in a cell. The search for inhibitors, including peptides, that actively impede the RBD-ACE2 binding process of the SARS-CoV-1 coronavirus has been an active field of investigation for many years [17,33,34], and the current pandemic due to the closely related SARS-CoV-2 has led to greatly renewed interest [35–41].

A considerable number of mutations of the spike protein of SARS-CoV-2 have been characterized, as the virus mutates rapidly under selective pressure as is commonly observed in RNA viruses [12,42]. Several of these mutations are located in the RBD domain [7–9]. While most of the mutations of the RBD known to date tend to reduce the virus' infectivity [43], several of them (for example V367F, W436R and N354D) have been predicted to have higher binding affinity to human ACE2 on the basis of molecular dynamics simulations, thus probably being more infectious in vivo [44]. Some of the mutations reduce the sensitivity of the virus to specific monoclonal antibodies [45]. For example, A475V, Y508H and F490L reduce the sensitivity to several monoclonal antibodies, while V483A becomes resistant to some as well [43]. Conversely, some mutations in the RBD region, including V367F, Q409E, Q414E, I468F, I468T, Y508H, and A522V, have been observed to make the spike protein more susceptible to neutralization by monoclonal antibodies [44]. Importantly, mutation N501Y of the RBD appears in two recently identified high infectivity strains [29,46,47]. Due to the high rate of mutation of CoV-2, considerable effort has been devoted to identify conserved epitopes in the RBD that might provide for mutation-resistant antibody targets [48–51].

In recent work, we proposed that an alternative strategy for the development of therapeutic drugs for SARS-CoV-2 could target critical regions of the virus proteins involved in folding initiation [13]. Specifically, we proposed that peptide drugs of defined sequence could function as naturally designed inhibitors of the folding of RBD-spike thus leading to virus inactivation [13]. The proposed strategy is graphically described in Figure 1.

#### Interdiction of protein folding for therapeutic purposes in SARS-CoV-2



**Figure 1.** The FITR strategy for SARS-CoV-2: (A) under normal in vivo conditions, the cellular protein synthesis machinery produces all the functional proteins required for the virus to reproduce effectively creating new infective virions; (B) in the presence of a folding inhibitor precisely targeted to inhibit the initial folding stage of RBD-spike, the resulting virion bears misfolded RBD-spike proteins and its infectivity should be much reduced.

For such a strategy to be efficient, it is important that the drugs developed from it are sufficiently effective against mutated viruses. In order to investigate whether that is the case, the following section explores the effects, if any, on the earliest folding events

of 13 known single mutations of RBD-spike (V31I, A348T, N354D, D364Y, V367F, K378R, Q409E, A435S, G476S, W436R, V483A, Y508H and N501Y), and also the impact of the triple mutation associated with the B.1.351 lineage (K417N, E484K and N501Y), and the double mutant B.1.617.2 known as the delta variant (E484Q and L432R). Our predictions rely on a combination of SCM energy calculations from primary sequence information following the procedure described in previous work, and the crystal tertiary structure of RBD-spike [52] to identify the viable primary contacts and their stabilization energies.

## 2.2. Primary Contacts for RBD-Spike Mutants

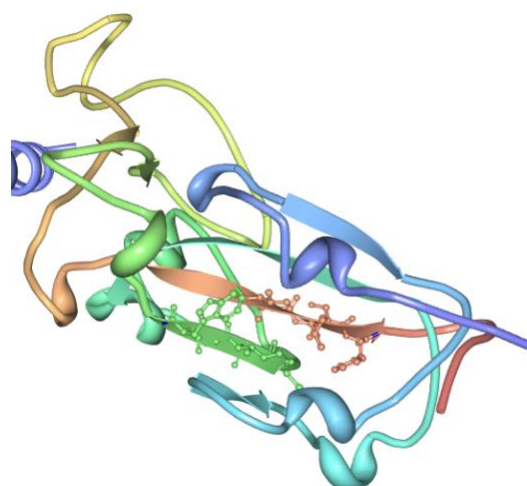
In previous work, we calculated the possible primary contacts (i.e., the earliest contact-formation events along the folding pathway) for native RBD-spike [13] and we show the predicted contacts within  $\sim 6$  kT (i.e., three orders of magnitude difference in population) of the most stable one in Table 1, labeled in order of decreasing energies as C1, C2 ... Cn. The calculations were completed following the methodology employed in the SCM before [19,20], entailing a search for the most stable possible hydrophobic contacts between pairs of 5-amino acid segments  $i$  and  $j$ , located at a distance  $n_{ij}$  along the sequence such that  $65 \lesssim n_{ij} \lesssim 100$  amino acids (see Section 4.2 in Methods for a complete explanation).

**Table 1.** Predicted possible primary contacts of native RBD-spike and their correlation with the native 3D structure, only native-like contacts are expected to lead to native folding pathways. The relative populations are expressed in percentages. Estimated errors are standard deviations.

	Contact	Stability (kT)	Population (%)	On Structure
C1	<sup>432</sup> CVIAW <sup>436</sup> on <sup>511</sup> VVLSF <sup>515</sup>	$-11.5 \pm 0.3$	59	Native-like
C2	<sup>432</sup> CVIAW <sup>436</sup> on <sup>335</sup> LCPFG <sup>339</sup>	$-10.8 \pm 0.2$	29.3	Non-native
C3	<sup>432</sup> CVIAW <sup>436</sup> on <sup>365</sup> YSVLY <sup>369</sup>	$-9.3 \pm 0.4$	6.5	Non-native
C4	<sup>432</sup> CVIAW <sup>436</sup> on <sup>349</sup> SVYAW <sup>353</sup>	$-8.7 \pm 0.3$	3.6	Non-native
C5	<sup>390</sup> LCFTN <sup>394</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-7.6 \pm 0.4$	1.2	Non-native
C6	<sup>398</sup> DSFVI <sup>402</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-6.2 \pm 0.4$	0.3	Non-native
C7	<sup>365</sup> YSVLY <sup>369</sup> on <sup>452</sup> LYRLF <sup>456</sup>	$-4.5 \pm 0.5$	0.1	Non-native

Our calculations predict that the best possible primary contact C1 is established between segments (<sup>432</sup>CVIAW<sup>436</sup>) centered at I434, and (<sup>511</sup>VVLSF<sup>515</sup>), centered at L513, 79 residues apart, with a stability of  $\Delta G_{\text{cont}}(\text{C1}) \approx -11.5$  kT\*, representing  $\sim 59\%$  of the early primary contacts. The two components of the predicted contact C1 are in clear proximity in the crystal structure of the protein (PDB code 6M0J) [52]. The second-best possible contact C2 is defined by segments (<sup>335</sup>LCPFG<sup>340</sup>), centered at P337, and (<sup>432</sup>CVIAW<sup>436</sup>), centered at I434, 97 residues apart, with a stability of  $\Delta G_{\text{cont}} \approx -10.8$  kT. Contact C1 is shown on the crystal structure in Figure 2.

Contacts C2–C4 are not native-like in the crystal structure, although its constituent segments are generally in the same region of RBD-spike. Contacts C5–C7 are clearly non-native. Thus, as explained in previous work [13], our result implies that the best primary contact C1 predicted here constitutes the entry point to “Nature’s Shortcut” [18] to the native structure. As segments (<sup>432</sup>CVIAW<sup>436</sup>) and (<sup>511</sup>VVLSF<sup>515</sup>) define C1, they constitute the natural FITR/FIP pair for consideration of developing a therapeutic peptide. Moreover, because contacts C1–C4 all include segment (<sup>432</sup>CVIAW<sup>436</sup>), a FIP designed from sequence (<sup>511</sup>VVLSF<sup>515</sup>) should be able to inhibit not just folding through the dominant primary contact, but also any added folding that might occur through C2–C4, thus maximizing the folding interdiction effect.



**Figure 2.** The predicted best primary contact for the receptor binding domain from the SARS Coronavirus-2 (PDB: 6M0J). Formation of the primary contact is the folding initiation event in the SCM. The color code reflects the location from the N-terminus (dark blue) to the C-terminus (yellow). Only the side chains corresponding to the segments that define the primary contacts are shown.

Our calculations considered the mutants V341I, A348T, N354D, D364Y, V367F, K378R, Q409E, G476S, V483A, Y508H and N501Y as well as the triple mutation B1.351, K417N, E484K and N501Y, besides the double delta variant B1.617.2, E484Q and L432R. Importantly, none of these variants were found to bear a significantly altered native primary contact as identified in the non-mutated sequence. Thus, the predictions for possible interdicting peptides made for the native sequence apply to the considered mutants. There are some small effects of these mutations on some of the non-dominant, non-native primary contacts that do not affect the dominant native pathway. These effects are shown in Table 2.

**Table 2.** Segment shifts and stability changes induced by the mutations in the non-native contacts.

Mutation	Contact	Segment Shift	Stability (kT)
D364Y	C3	<sup>365</sup> YSVLY <sup>369</sup> to <sup>361</sup> CVAYY <sup>365</sup>	$-9.6 \pm 0.5$
V367F	C3	No	$-10.5 \pm 0.5$

The other two mutations, A435S and W436R are located in segment (<sup>432</sup>CVIAW<sup>436</sup>) which is, as explained above, involved in all of the predicted primary contacts. Due to their location within segment (<sup>432</sup>CVIAW<sup>436</sup>), both mutations induce an energy shift  $\Delta\Delta G_{\text{cont}}$ , in the stability of the native primary contact. The effects of these shifts on the energy of the predicted native primary contact and the possible FITRs are analyzed in Sections 2.3 and 2.4.

### 2.3. Effects of Mutation A435S

Mutation A435S induces an energy shift of magnitude  $\Delta\Delta G_{\text{cont}} \approx 0.8 \text{ kT}$  on contacts C1-C4 that now involve the mutated segment (<sup>432</sup>CVISW<sup>436</sup>), instead of (<sup>432</sup>CVIAW<sup>436</sup>), including the dominant native primary contact. The results obtained for the possible contacts of this mutant are shown in Table 3.

**Table 3.** Predicted possible primary contacts of the A435S RBD-spike mutant and their correlation with the native 3D structure, only native-like contacts are expected to lead to native folding pathways. The relative populations are expressed in percentages.

	Contact	Stability (kT)	Population (%)	On Structure
C1	<sup>432</sup> CVISW <sup>436</sup> on <sup>511</sup> VVLSF <sup>515</sup>	$-10.7 \pm 0.3$	57.7	Native-like
C2	<sup>432</sup> CVISW <sup>436</sup> on <sup>335</sup> LCPFG <sup>339</sup>	$-10 \pm 0.3$	28.7	Non-native
C3	<sup>432</sup> CVISW <sup>436</sup> on <sup>365</sup> YSVLY <sup>369</sup>	$-8.5 \pm 0.4$	6.4	Non-native
C4	<sup>432</sup> CVISW <sup>436</sup> on <sup>349</sup> SVYAW <sup>353</sup>	$-8.0 \pm 0.4$	3.9	Non-native
C5	<sup>390</sup> LCFTN <sup>394</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-7.6 \pm 0.4$	2.6	Non-native
C6	<sup>398</sup> DSFVI <sup>402</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-6.2 \pm 0.4$	0.6	Non-native
C7	<sup>365</sup> YSVLY <sup>369</sup> on <sup>452</sup> LYRLF <sup>456</sup>	$-4.5 \pm 0.5$	0.1	Non-native

With these results, the predictions made for the native chain remain the same as mutation A435S induces the same energy shift on all of the other possible contacts that involve segment (<sup>432</sup>CVIAW<sup>436</sup>) as shown in Table 2. The changes in relative population are also very small, around ~1%. Thus, the dominant primary contact remains the same as for the native sequence C1, with C2 relatively close in energy, and a FITR strategy focused on interdicting the formation of C1 should work equally well as with the native protein.

The mutation A435S could have an effect on the design of an optimal FIP. When searching for an optimal FIP, it is likely that detailed interactions play a significant role. Ideally a FIP is to be found that has greater affinity for the FITR than the naturally occurring segments of the primary contacts. Such an optimized FIP could more efficiently compete with the intramolecular formation of the primary contact, thus reducing the drug concentration required to interdict in the folding of RBD-spike. The search for such an optimal FIP is likely to be greatly facilitated by employing molecular dynamics techniques that allow for an atomic-level resolution assessment of the relative binding affinities of different FIPs [53,54].

#### 2.4. Effects of Mutation W436R

Mutation W436R induces a 1-amino acid shift for contacts C1-C4 involving segment (<sup>432</sup>CVIAW<sup>436</sup>), that now involves segment (<sup>431</sup>GCVIA<sup>435</sup>) instead. This segment shift is the natural consequence of a hydrophobic tryptophan, being replaced at the end of the segment by a charged arginine. From a structural perspective, the segment shift does not significantly change the native classification of the predicted contacts as shown in Table 4.

**Table 4.** Predicted possible primary contacts of the W436R RBD-spike mutant and their correlation with the native 3D structure, only native-like contacts are expected to lead to native folding pathways. The relative populations are expressed in percentages. The mutation W436R has made the non-native C5 more stable than the native-like C1. The location of C1 is shifted by one amino acid with respect to native RBD-spike (see Section 2.2).

	Contact	Stability (kT)	Population (%)	On Structure
C1	<sup>431</sup> GCVIA <sup>435</sup> on <sup>511</sup> VVLSF <sup>515</sup>	$-6.5 \pm 0.2$	16.9	Native-like
C2	<sup>431</sup> GCVIA <sup>435</sup> on <sup>335</sup> LCPFG <sup>339</sup>	$-5.8 \pm 0.2$	9.3	Non-native
C3	<sup>431</sup> GCVIA <sup>435</sup> on <sup>365</sup> YSVLY <sup>369</sup>	$-5.6 \pm 0.4$	6.9	Non-native
C4	<sup>431</sup> GCVIA <sup>435</sup> on <sup>349</sup> SVYAW <sup>353</sup>	$-3.8 \pm 0.3$	1.1	Non-native
C5	<sup>390</sup> LCFTN <sup>394</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-7.6 \pm 0.4$	50.9	Non-native
C6	<sup>398</sup> DSFVI <sup>402</sup> on <sup>488</sup> CYFPL <sup>492</sup>	$-6.2 \pm 0.4$	12.6	Non-native
C7	<sup>365</sup> YSVLY <sup>369</sup> on <sup>452</sup> LYRLF <sup>456</sup>	$-4.8 \pm 0.5$	2.3	Non-native

Thus, it is reasonable to treat all the contacts involving (<sup>431</sup>GCVIA<sup>435</sup>) in the W436R mutant as equivalent to those involving (<sup>432</sup>CVIAW<sup>436</sup>) for the native sequence. From the stability standpoint however, mutation W436R represents a different case from the mutations considered above because it introduces a large change in the stabilization energy of the dominant contact C1, and also contacts C2–C4, all including segment (<sup>432</sup>CVIAW<sup>436</sup>) in the native protein, such that  $\Delta\Delta G_{C1-C4}(W436R) \approx 5$  kT. The new set of predicted contacts and their stabilities is shown in Table 3. The native primary contact location is shifted by one residue; the native segment (<sup>432</sup>CVIAW<sup>436</sup>) is shifted to (<sup>431</sup>GCVIA<sup>435</sup>).

This energy shift induces a large change in the relative populations of the possible primary contacts, which is large enough that the most stable possible primary contact is now C5, defined by segments (<sup>390</sup>LCFTN<sup>394</sup>) and (<sup>488</sup>CYFPL<sup>492</sup>), 98 amino acids apart, with  $\Delta G_{\text{cont}} \approx 7.6$  kT, 1.1 kT more stable than the predicted native-like best mutated contact C1(W436R). This contact is completely non-native on the tertiary structure of RBD-spike, so within the current model, it is not expected to nucleate a native-like folding pathway. The native-like contact C1 represents for this mutant, just ~16.4% of the total population of primary contacts. As contacts that are not native-like are not expected to lead to successful folding, they would have to break up before those molecules bearing them could form a native-like alternative contact. Thus, formation of the non-native contacts is suggested to constitute a kinetic trap along the folding pathway, and it is expected within the model that mutant W436R folds less efficiently than the native RBD-spike, taking a longer time for most of the molecules to reach the native 3D structure. There are no detailed folding experiments on RBD-spike and its mutants to directly test this hypothesis. However, when expression of the mutant W436R was attempted experimentally, it was observed that the yield of functional protein is much less than that of native RBD-spike and also several mutants [39]. Finally, because virus assembly time scales within the cell are much longer than protein folding times, the less kinetically efficient folding of the RBD-spike mutant W436R need not significantly hamper the formation and spreading of the virus in vivo.

The results for mutant W436R suggest that a possible interdiction strategy could employ the equivalent native protein primary contact peptide (<sup>432</sup>CVIAW<sup>436</sup>) as the starting point to design an FIP. This suggestion is made because the interaction of (<sup>432</sup>CVIAW<sup>436</sup>) with (<sup>511</sup>VVLSF<sup>515</sup>) is much stronger than that of segment (<sup>431</sup>GCVIA<sup>435</sup>) included in the native primary contact in mutant W436R. Thus, peptide (<sup>432</sup>CVIAW<sup>436</sup>) is a natural choice as an initial design of a therapeutic peptide drug that interdicts in the formation of the native primary contact of mutant W436R. Finally, because the initial population of the native primary contact is lower than that in the native protein, the FITR strategy is likely to be more efficient for W436R at the same concentrations of folding interdicting peptides.

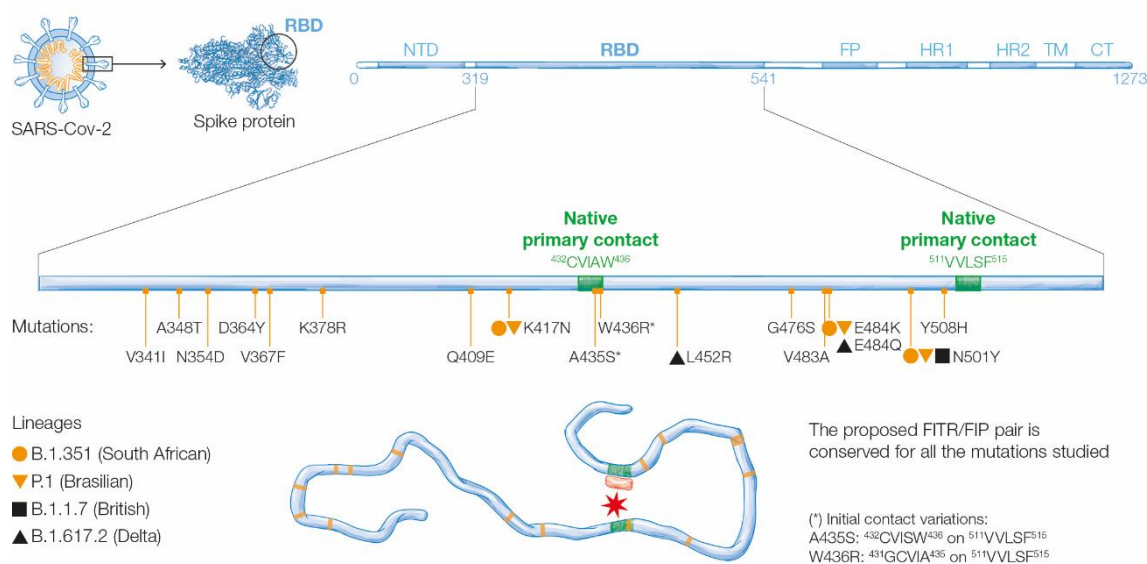
For the purpose of designing an experimental test of the FITR strategy, it is important to bear in mind that, in order to successfully interdict the folding process, the FIP must interact with the unfolded protein. This specific feature of the FITR strategy suggests that the optimal step along the protein synthesis cycle to attempt interdiction of the subsequent folding process is when the protein is still being translated in the ribosome. Thus, a possible proof-of-concept experiment could rely on employing ribosomes to synthesize spike-RBD in the presence of an appropriate FIP designed following the criteria laid out previously. If the FITR strategy operates as expected, the yield of properly folded spike-RBD should decrease, and that can be measured both through standard structure characterization techniques such as circular dichroism (CD) or nuclear magnetic resonance (NMR) [55], or potentially through measures of activity such as the ability to bind ACE-2 [56]. An additional advantage of this experimental approach is that it allows for the molecule being tested to interdict the folding of RBD-spike on much longer time scales than early folding time scales. The earliest folding time scales can be as short as microseconds [57], while the translation process in the ribosomes is much slower [58]. Although such an experiment is certainly within the purview of current techniques [59], it is important to keep in mind that the process of drug discovery is fraught with practical difficulties, and that success in vitro is only the first step on the way to applicability in vivo.

### 3. Discussion

For the 15 SARS-CoV-2 RBD-spike mutants studied here, the only native-like primary contact and thus “Nature’s shortcut” to the native structure remains defined by segment ( $^{432}\text{CVIAW}^{436}$ ) (or the equivalent ( $^{431}\text{GCVIA}^{435}$ ) for the W436R mutant), and ( $^{511}\text{VVLFS}^{515}$ ). The robustness of the location of the native-like primary contact to mutation is a natural consequence of its key role in nucleating the subsequent folding of the protein chain [19]. Thus, although these conclusions await experimental verification, these results also serve to bolster the foundations of the SCM and its prediction capability. It was noted additionally in previous work on adenylate kinase [19] that the segments involved in the predicted best primary contact were very conserved, and their mutation tended to impede the folding of the protein. It would be important to experimentally test further effects of conservative and non-conservative mutations on the predicted RBD-spike primary contacts. In particular, it would be very interesting to test the prediction made here of slower folding for the mutant W436R.

In previous work, we proposed that the identification of key protein segments involved in the earliest folding nucleation events provided for natural folding interdiction target regions (FITRs), against which specific folding interdicting peptides (FIPs) could be identified as leading therapeutic candidates [13]. Moreover, because two protein segments are involved in every contact, in each case one segment can be used as a template for the FITR and the other for the FIP. In the case of RBD-spike, the FITR would be either one of the dominant primary contact segments ( $^{432}\text{CVIAW}^{436}$ ) or ( $^{511}\text{VVLFS}^{515}$ ). Depending upon the FITR actually aimed for, the FIP could be chosen employing the other segment as a template, so if for example ( $^{432}\text{CVIAW}^{436}$ ) was the FITR, the starting point for the design of the FIP would be the segment ( $^{511}\text{VVLFS}^{515}$ ). While these choices are natural within the SCM model, careful mutation of these choices might perform even better. These conclusions are depicted graphically in Figure 3.

#### Conservation of the folding interdiction target regions upon mutation of RBD-spike



**Figure 3.** The predicted native primary contact that initiates the folding pathway of RBD-spike remains viable for all studied mutants. Thus, the proposed FITR/FIP strategy for therapeutic drug development should apply to all mutants.

In this paper, we showed that such a FITR drug development strategy is likely to be quite robust to mutations of the selected target virus proteins. The expected and observed robustness within the model is the natural consequence of the primary contact being an essential nucleating event for the folding of the protein [18].

Robustness to mutation of the FITR strategy is suggested to be relevant for the successful design of effective drugs against viruses with high rates of mutation, and whose

mutations regularly enable them to overcome previously developed treatments, a common occurrence in RNA viruses [13]. It remains to be seen whether a pattern of vaccine effectiveness reduction through mutation establishes itself for SARS-CoV-2 but the possibility cannot be discounted [60]. Thus, a mutation-resistant strategy such as FITR might be very useful for therapeutics whose effectiveness is difficult for the virus to overcome.

An interesting related question is whether mutations that make the primary contact more stable relative the native protein could increase the infectivity and/or lethality of SARS-CoV-2 by increasing the initial population of molecules entering the dominant folding pathway. The efficacy of such a possible gain-of-function mechanism cannot be ascertained on the basis of the mutants of RBD-spike analyzed here.

#### 4. Methods: The SCM Model

The physical basis of the SCM and its most up-to-date formulation have been recently explained in full detail [19,20]. Here, only a brief summary of the main concepts is presented that are relevant to the issues investigated in the present paper.

##### 4.1. The SCM Entropic Cost of Loop Formation

The SCM considers early non-local contacts based on the entropy of formation of the resultant protein loops. The SCM has successfully predicted many of the observed features of protein folding pathways [20]. In the SCM two different loop regimes are considered when analyzing early non-local contacts: short loops for which the gyration radius  $R_g(n)$ , is smaller than the average side chain length  $l(n)$  (i.e.,  $R_g(n) < l(n)$ ), and long loops for which  $R_g(n) > l(n)$ . The loop length at which the transition between the short and long loops takes place (i.e., the length for which  $R_g(n) \approx l(n)$ ) is called the optimal loop length  $n_{op}$ . The optimal loop length has been estimated to be  $n_{op} \approx 65$  amino acids for typical protein sequences for  $l(n) \approx 7.9$  Å [20], although some sequence variability exists and  $n_{op}$  is expected to be shorter for highly disordered proteins that contain few of the bulky hydrophobic amino acids that must be buried in any contact and thus should dominate the effective  $l(n)$  in most globular proteins [61]. This value for  $n_{op}$  is consistent with experimental data showing the behavior of poly-alanine, a polypeptide with smaller side chains than the average globular protein; in this case  $l(n) \approx 6.7$  Å [20], which exhibits deviations from Gaussian statistics due to steric hindrance when  $n < 50$  amino acids [62].

The long loop regime is physically equivalent to the classical Flory–Jacobson–Stockmayer (FSJ) picture and the entropic cost of forming protein loops is well represented, assuming that the amino acids can be taken to be solid ball-like, by a simple logarithmic function of the form [63]:

$$\Delta S_{loop} (n > n_{op}) \approx -3/2 \ln (n) \quad (1)$$

This is clearly an approximation, as for example, the side chains would be better represented by solid spheres of different sizes according to the primary sequence. In the SCM short loop regime, however, the internal degrees of freedom of the side chains cannot be neglected, and the entropic cost of forming short loops must be higher than when the amino acids are taken to be solid spheres. Moreover, because most of the degrees of freedom are in the side chains, we expect the contribution of the side chains to the overall entropic cost to be dominant with respect to that of any constraints imposed by loop formation on the backbone.

Thus, in the SCM it is expected that for a short loop, the entropic cost of loop formation  $\Delta S_{loop}$  approximately becomes:

$$\Delta S_{loop} (n < n_{op}) \approx -3/2 \ln (n) + \Delta S_{side-chain-crowding}(n, l), \quad (2)$$

with  $\Delta S_{side-chain-crowding} \ll 0$ , opposing folding. When  $R_g(n) \geq l(n)$ , we have  $\Delta S_{side-chain-crowding} \approx 0$  and the standard FJS regime is recovered. The side chain crowding term  $\Delta S_{side-chain-crowding}$  will appear as a correction to the JS results for shorter loops. It is

extremely difficult to obtain an analytical expression for the side chain crowding term, and in the SCM it has been presented in generic Boltzmann–Gibbs form [20]:

$$\Delta S_{\text{side-chain-crowding}}(n, l) = n \ln [f_0(n, l)/f_{\text{loop}}(n, l)], \quad (3)$$

where  $f_0(n, l)$  is the average configurational freedom per amino acid in the unfolded chain, and  $f_{\text{loop}}(n, l)$  is the average configurational freedom of an amino acid in a loop. Consideration of modifying the homogeneous Flory-like representation of the protein chain to take into fuller account the microscopic details of the protein–solvent system is not exclusive to the SCM and has been employed before to account, for example, for the effects of the solvent [64].

Based on the model developed above, in the SCM, the folding of proteins with more than ~100 amino acids likely involve the formation of an early non-local contact (i.e., the primary contact within the SCM), that defines the earliest folding phase with  $n \geq n_{\text{op}} \approx 65$  amino acids. As only a few primary contacts can be established at most in proteins of length  $n \geq n_{\text{op}}$ , most of the tertiary structure contacts will still be defined by contacts at shorter range established in later folding phases [20,65]. Formation of the primary contact in the SCM defines the primary loop, which subsequently collapses through two-state kinetics [20,60]. As proteins longer than ~100 amino acids do not generally undergo two-state collapse [20] but rather fold through multi-step pathways, consistent simple physical reasoning implies that there is a limit to the size of the primary loop that can successfully lead to the native SCM folding pathway of ~100 amino acids.

The concept of folding nucleated by non-local contacts is not exclusive of the SCM, having arisen earlier in the context of the diffusion-collision model [66] and in the energy landscape picture [67]. It also has appeared in simulations of the transition state of two-state folding proteins [68]. Additionally, protein topology has been considered an essential element of folding mechanisms in a number of theoretical efforts [69–73]. The particular feature in the SCM is that the early non-local contacts are highly specific as in the loop hypothesis [74], and a methodology is developed to derive their location from primary sequence information [19,20].

#### 4.2. Determining the Primary Contact

Based on the model presented in the previous sections, whether there is a non-local contact in an otherwise unfolded state is dependent upon the stability of the potential contact candidates at loop lengths of  $n \geq n_{\text{op}}$  amino acids. In the SCM, the stability of a contact formed by the number  $n_{\text{cont}}$  of amino acids,  $\Delta G_{\text{contact}}(n_{\text{cont}}, n_{\text{loop}})$ , can be written as:

$$\Delta G_{\text{contact}}(n_{\text{cont}}, n_{\text{loop}}) \approx \Delta G_{\text{int,H}}(n_{\text{cont}}) + \Delta G_{\text{loop}}(n_{\text{loop}}) + \Delta G_{\text{cont,S}}(n_{\text{cont}}) \quad (4)$$

Here  $\Delta G_{\text{loop}}$  represents the entropic free energy cost of the loop as discussed in Section 4.1. The term  $\Delta G_{\text{int,H}}$  denotes all the enthalpic interactions that help stabilize the contact, possibly including hydrophobic interactions, van der Waals interactions, hydrogen bonds, disulfide bonds, and salt bridges [75], and its value satisfies  $\Delta G_{\text{int}} < 0$ . The term  $\Delta G_{\text{cont,S}} > 0$  represents the entropic cost of constraining the side chains of the amino acids defining the contact and it opposes contact formation. A segment-specific determination of the value  $\Delta G_{\text{cont,S}}(n_{\text{cont}})$  for a given contact would require detailed MD techniques. However, a heuristic estimate can be made from earlier work within the SCM which showed that the average entropic cost of folding per amino acid for a sample of 13 proteins was  $\Delta G_{\text{folding/residue,S}} \approx 0.85 \text{ kT/residue}$  [65], and the maximum was  $\Delta G_{\text{folding/residue,S}} \approx 1.09 \text{ kT/residue}$ . As these are estimates for the entropic cost for folding per residue of complete proteins that include highly buried as well as flexible exposed regions, it is then reasonable to expect that the entropic cost of a contact-forming region must be closer to the highest calculated values for  $\Delta G_{\text{folding/residue,S}}$ . Here, we will assume that  $\Delta G_{\text{contact,S}}(n_{\text{contact}})$  for a contact including  $n_{\text{cont}}$  amino acids is approximately  $\Delta G_{\text{folding/residue,S}}$  determined by the number of residues defining the contact, such that

$\Delta G_{\text{cont},S}(n_{\text{cont}}) \approx 1.09 n_{\text{cont}}$ . This result is clearly an approximation, but it suffices for establishing a cut-off in the number of possible contacts that is consistent with the available structural data [65].

Hydrophobic interactions are well understood to constitute the main driving force of the folding process [70]. Other interactions such as hydrogen bonds are weaker [76], or like disulfide bonds and salt bridges form later along the folding pathway [75]. Thus, for an early contact forming from the unfolded state, we can take  $\Delta G_{\text{int}}(n_{\text{op}}) \approx \Delta G_{\text{hyd}}(n_{\text{op}})$ , where  $\Delta G_{\text{hyd}}(n_{\text{op}})$  represents the stabilizing effect of hydrophobicity in the early contacts, and Equation (4) can be written as

$$\Delta G_{\text{contact}}(n_{\text{cont}}, n_{\text{loop}}) \approx \Delta G_{\text{hyd}}(n_{\text{cont}}) + \Delta G_{\text{loop}}(n_{\text{loop}}) + \Delta G_{\text{contact},S}(n_{\text{contact}}) \quad (5)$$

Since the hydrophobic stabilization energy of the contact  $\Delta G_{\text{hyd}}$  is determined by the hydrophobicity of the segments involved, the hydrophobicity values  $h_k$  are obtained from the Fauchere–Pliska scale [77] and assigned to each residue in accordance with previous calculations within the SCM.

As the amino acid side chains are significantly larger than the typical peptide bond length, early contacts between two hydrophobic amino acids will inherently involve segments including several amino acids, adjacent to the initial contact. The stability of this early hydrophobic contact will determine where the folding process is initiated. This picture is not unlike the zapping model of Dill and collaborators [78]. Here, the typical early contact segment size will be taken to be  $\sim 5$  amino acids in line with previous calculations within the SCM [19,20]. The 5-amino acid window size is based on the geometric considerations underlying the SCM: with an average effective fluctuating width of the unfolded protein chain of  $w \sim 2 \ell(n) \approx 15.8 \text{ \AA}$ , and a peptide bond length of  $3.5 \text{ \AA}$ , the minimum number  $n_{\text{cont}}$  of amino acids that can define a contact in the open fluctuating chain should be  $n_{\text{cont}} \sim \text{int}[2 \ell(n)/3.5] = 5$  amino acids. In practice within the SCM, the hydrophobicity  $h_k$  of each residue is added over a segment contact window of five amino acids centered at residue  $i$ , resulting in a segment hydrophobicity  $h_{i,5}$  (a value of  $\sim 0.45$  is equivalent to a change in energy of  $kT$ ).

In order to determine the best contact, the  $h_{i,5}$  values of a segment centered at residue  $i$  is added to the  $h_j$  value of a segment centered at residue  $j$ , located at a distance  $n_{ij}$  at least  $n_{\text{op}}$  amino acids apart along the sequence, and no longer than the maximum primary loop length of  $\sim 100$  amino acids, to give a contact stability of:

$$\Delta G_{\text{cont}}(n_{\text{cont}}, n_{\text{loop}}) \approx kT [-(h_{i,5} + h_{j,5})/0.45 + 3/2 \ln n_{ij} + 10.9] \quad 100 \geq n_{ij} \geq 65 \quad (6)$$

Finally, in order to estimate the uncertainty in  $\Delta G_{\text{cont}}$ , we carried out a standard deviation analysis, relying on the experimental uncertainties reported for the  $h_i$  values for each residue type [77], and the experimental data leading to the calculated value of  $\Delta G_{\text{folding/residue},S} \approx 1.09 kT/\text{residue}$  [65].

## 5. Conclusions

In this paper, we showed that a hypothesized FITR drug design strategy appears robust for 15 mutants of SARS-CoV-2. It was further explained that such robustness is an attractive asset for the development of treatments whose effectiveness does not diminish in the face of natural virus mutation.

The FITR strategy could be moreover extended to other viral and bacterial protein targets. For example, it would be very interesting to design mutation-resistant drugs against Influenza and many other infectious diseases [79]. As the FITR aims to induce therapeutic effects by interfering with the protein folding process in a specific fashion, it is possible that its application could be further extended to diseases which are related to protein folding, including misfolding-related neurodegenerative conditions [80], and maybe even cancer [81,82]. In these scenarios, FITR would serve to interdict in the misfolding process. Work is underway to explore these prospects.

**Author Contributions:** Conceptualization, F.B.-C.; Data curation, F.B.-C.; Formal analysis, F.B.-C.; Funding acquisition, F.B.-C. and H.A.R.; Investigation, F.B.-C. and H.A.R.; Methodology, F.B.-C.; Project administration, F.B.-C. and H.A.R.; Supervision, H.A.R.; Validation, F.B.-C. and H.A.R.; Visualization, F.B.-C.; Writing—original draft, F.B.-C.; Writing—review & editing, F.B.-C. and H.A.R. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded partially by the U.S. Department of Energy (grant DE-FG02-02ER15344).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** Figures were commissioned from Fundamentum S.L., except Figure 2 that was prepared with Protein Workshop [83].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Carlson, C.J.; Phelan, A.L. A choice between two futures for pandemic recovery. *Lancet Planet. Health* **2020**, *4*, E545–E546. [\[CrossRef\]](#)
- Hiscott, J.; Alexandridi, M.; Muscolini, M.; Tassone, E.; Palermo, R.; Soultioti, M.; Zevini, A. The global impact of the coronavirus pandemic. *Cytokine Growth Factor Revs.* **2020**, *53*, 1–9. [\[CrossRef\]](#)
- Impact of the COVID-19 pandemic on trade and development: Transitioning to a new normal. In Proceedings of the United Nations Conference on Trade and Development, New York, NY, USA, 19 November 2020.
- Kemp, S.A.; Collier, D.A.; Gupta, R.V. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **2021**, *592*, 277–282. [\[CrossRef\]](#) [\[PubMed\]](#)
- Williams, T.C.; Burgers, W.A. SARS-CoV-2 evolution and vaccines: Cause for concern? *Lancet Respir. Med.* **2021**, *9*, 333–335. [\[CrossRef\]](#)
- Moore, J.P.; Offit, P.A. SARS-CoV-2 vaccines and the growing threat of viral variants. *JAMA* **2021**, *325*, 821–822. [\[CrossRef\]](#)
- Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812–827. [\[CrossRef\]](#)
- Islam, M.R.; Hoque, M.N.; Rahman, M.S.; Rubayet Ul Alam, A.S.M.; Akther, M.; Puspo, J.A.; Akter, S.; Sultana, M.; Crandall, K.A.; Hossain, M.A. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Nature Sci. Reps.* **2020**, *10*, 14004. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tang, J.W.; Tambyah, P.A.; Hui, D.S. Emergence of a new SARS-CoV-2 variant in the UK. *J. Infect.* **2021**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Emergence and rapid spread of a new severe acute respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *Medrxiv* **2020**. [\[CrossRef\]](#)
- Bernal, J.L.; Andrews, N.; Gower, C.; Gallagher, E.; Simmons, R.; Thelwall, S.; Stowe, J.; Tessler, E.; Groves, N.; Dabrera, G.; et al. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *Medrxiv* **2021**. [\[CrossRef\]](#)
- Holland, J.J.; Drake, J.W. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13910–13913.
- Bergasa-Caceres, F.; Rabitz, H.A. Interdiction of protein folding for therapeutic drug development in SARS-CoV-2. *J. Phys. Chem. B* **2020**, *124*, 8201–8208. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, F. Receptor recognition mechanisms of coronaviruses: A decade of structural studies. *J. Virol.* **2015**, *89*, 1954–1964. [\[CrossRef\]](#)
- Li, W.H.; Greenough, T.C.; Moore, M.J.; Vasilieva, N.; Somasundaran, M.; Sullivan, J.L.; Farzan, M.; Choe, H. Efficient replication of severe acute respiratory syndrome coronavirus in mouse cells is limited by murine angiotensin-converting enzyme. *J. Virol.* **2004**, *78*, 11429–11433. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, F.; Li, W.H.; Farzan, M.; Harrison, S.C. Structure of SARS coronavirus spike receptor binding domain complexed with receptor. *Science* **2005**, *309*, 1664–1668. [\[CrossRef\]](#) [\[PubMed\]](#)
- Du, L.; He, Y.; Zhou, Y.; Liu, S.; Zheng, B.-J.; Jiang, S. The spike protein of SARS-CoV: A target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **2009**, *7*, 226–236. [\[CrossRef\]](#)
- Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581*, 221–224. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bergasa-Caceres, F.; Ronneberg, T.A.; Rabitz, H.A. Sequential collapse model for protein folding pathways. *J. Phys. Chem. B* **1999**, *103*, 9749–9758. [\[CrossRef\]](#)
- Bergasa-Caceres, F.; Haas, E.; Rabitz, H.A. Nature's shortcut to protein folding. *J. Phys. Chem. B* **2019**, *123*, 4463–4476. [\[CrossRef\]](#)
- Motta, A.; Reches, M.; Pappalardo, L.; Andreotti, G.; Gazit, E. The preferred conformation of the tripeptide ala-phe-ala in water is an inverse  $\gamma$ -turn: Implications for protein folding and drug design. *Biochemistry* **2005**, *44*, 14170–14178. [\[CrossRef\]](#)

22. Broglia, R.A.; Serrano, L.; Tiana, G. (Eds.) Protein Folding and Drug Design. In *Proceedings of the International School of Physics "Enrico Fermi"*, 165; IOS Press: Amsterdam, The Netherlands, 2007.
23. Das, A.; Yadav, A.; Gupta, M.; Purushotham, R.; Terse, V.L.; Vishvakarma, V.; Singh, S.; Nandi, T.; Mandal, K.; Gosavi, S.; et al. Rational design of protein-specific folding modifiers. *bioRxiv* **2020**. [[CrossRef](#)]
24. Spagnoli, G.; Massignan, T.; Astolfi, A.; Biggi, S.; Rigoli, M.; Brunelli, P.; Libergoli, M.; Ianeselli, A.; Orioli, S.; Boldrini, A. Pharmacological inactivation of the prion protein by targeting a folding intermediate. *Nat. Comm. Biol.* **2021**, *4*, 62. [[CrossRef](#)]
25. Massignan, T.; Boldrini, A.; Terruzzi, L.; Spagnoli, G.; Astolfi, A.; Bonaldo, V.; Pischedda, F.; Pizzato, M.; Lolli, G.; Barreca, M.L. Antimalarial Artefenomel Inhibits Human sars-cov-2 Replication In Cells While Suppressing The Receptor ACE2. *arXiv* **2021**, 2004, 13493v4.
26. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *181*, 281–292. [[CrossRef](#)] [[PubMed](#)]
27. Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T.S.; Herrler, G.; Wu, N.H.; Nitsche, A.; et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**, *181*, 271–280. [[CrossRef](#)] [[PubMed](#)]
28. Tortorici, M.A.; Veesler, D. Structural insights into coronavirus entry. *Adv. Virus Res.* **2019**, *105*, 93–116. [[PubMed](#)]
29. Greaney, A.J.; Starr, T.N.; Gilchuk, P.; Zost, S.J.; Binshtein, E.; Loes, A.N.; Hilton, S.K.; Hudleston, J.; Eguia, R.; Crawford, K.H.D.; et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **2021**, *29*, 44–57. [[CrossRef](#)]
30. Li, W.H.; Moore, M.J.; Vasilieva, N.; Sui, J.H.; Wong, S.K.; Berne, M.A.; Somasundaran, M.; Sullivan, J.L.; Luzuriaga, K.; Greenough, T.C.; et al. Angiotensin-converting Enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **2003**, *426*, 450–454. [[CrossRef](#)]
31. Hatmal, M.M.; Alshaer, W.; Hatamleh, M.A.I.; Hatmal, M.; Smadi, O.; Taha, M.O.; Oweida, A.J.; Boer, J.C.; Mohamud, R.; Plebanski, M. Comprehensive Structural and Molecular Comparison of Spike Proteins of SARS-CoV-2, SARS-CoV and MERS-CoV and Their Interaction with ACE2. *Cells* **2020**, *9*, 2638. [[CrossRef](#)]
32. Brown, E.E.F.; Rezaei, R.; Jamieson, T.R.; Dave, J.; Martin, N.T.; Singaravelu, R.; Crupi, M.J.F.; Boulton, S.; Tucker, S.; Duong, J.; et al. Characterization of critical determinants of ACE2-SARS CoV-2 RBD interaction. *Int. J. Mol. Sci.* **2021**, *22*, 2268. [[CrossRef](#)]
33. Premkumar, L.; Segovia-Chumbez, B.; Jadi, R.; Martinez, D.R.; Raut, R.; Markmann, A.J.; Cornaby, C.; Bartelt, L.; Weiss, S.; Park, Y.; et al. The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci. Immunol.* **2020**, *5*, eabc8413. [[CrossRef](#)] [[PubMed](#)]
34. Hu, H.; Li, L.; Kao, R.Y.; Kou, B.; Wang, Z.; Zhang, L.; Zhang, H.; Hao, Z.; Tsui, W.H.; Ni, A.; et al. Screening and identification of linear B-cell epitopes and entry-blocking peptide of Severe Acute Respiratory Syndrome (SARS)-associated coronavirus using synthetic overlapping peptide library. *J. Comb. Chem.* **2005**, *7*, 648–656. [[CrossRef](#)]
35. Keng, C.T.; Zhang, A.; Shen, S.; Lip, K.M.; Fielding, B.C.; Tan, T.H.; Chou, C.-F.; Loh, C.B.; Wang, S.; Fu, J.; et al. Amino acids 1055 to 1192 in the S2 region of severe acute respiratory syndrome coronavirus S protein induce neutralizing antibodies: Implications for the development of vaccines and antiviral agents. *J. Virol.* **2005**, *79*, 3289–3296. [[CrossRef](#)]
36. Wu, Y.; Wang, C.; Shen, W.; Peng, D.; Li, C.; Zhao, Z.; Li, S.; Li, Y.; Bi, Y.; Yang, Y.; et al. A non-competing pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **2020**, *368*, 1274–1278. [[CrossRef](#)] [[PubMed](#)]
37. Yuan, M.; Liu, H.; Wu, N.C.; Lee, C.-C.D.; Zhu, X.; Zhao, F.; Wang, D.; Yu, W.; Hua, Y.; Tien, H.; et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science* **2020**, *369*, 1119–1123. [[CrossRef](#)] [[PubMed](#)]
38. Esparza, T.J.; Martin, N.P.; Anderson, G.P.; Goldman, E.R.; Brody, D.L. High affinity nanobodies block SARS-CoV-2 spike receptor binding domain interaction with human angiotensin converting enzyme. *Nature Sci. Rep.* **2020**, *10*, 22370.
39. Kim, S.I.; Noh, J.; Kim, S.; Choi, Y.; Yoo, D.K.; Lee, Y.; Lee, H.; Jung, J.; Kang, C.K.; Song, K.-H.; et al. Stereotypic neutralizing antibodies against SARS-CoV-2 spike protein receptor binding domain in patients with COVID-19 and healthy individuals. *Sci. Trans. Med.* **2021**, *13*, eabd6990. [[CrossRef](#)] [[PubMed](#)]
40. Han, Y.; Král, P. Computational design of ACE-2-based peptide inhibitors of SARS-CoV-2. *ACS Nano* **2020**, *14*, 5143–5147. [[CrossRef](#)] [[PubMed](#)]
41. Maiti, B.K. Potential role of peptide-based antiviral therapy against SARS-CoV-2 infection. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 783–785.
42. Karoyan, P.; Vieillard, V.; Gómez-Morales, L.; Odile, E.; Guihot, A.; Luyt, C.-E.; Denis, A.; Grondin, P.; Lequin, O. Human ACE2 peptide-mimics block SARS-CoV-2 pulmonary cells infection. *Nat. Commun. Biol.* **2021**, *4*, 197. [[CrossRef](#)]
43. Padhi, A.K.; Tripathi, T. Can SARS-CoV-2 accumulate mutations in the S-protein to increase pathogenicity? *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1023–1026. [[CrossRef](#)]
44. Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **2020**, *182*, 1284–1294. [[CrossRef](#)] [[PubMed](#)]
45. Ou, J.; Zhou, Z.; Dai, R.; Zhao, S.; Wu, X.; Zhang, J.; Lan, W.; Cui, L.; Wu, J.; Seto, D.; et al. Emergence of SARS-CoV-2 spike RBD mutants that enhance viral infectivity through increased human ACE2 receptor binding affinity. *bioRxiv* **2020**. [[CrossRef](#)]

46. Rambaut, A.; Loman, N.; Pybus, O.; Barclay, W.; Barrett, J.; Carabelli, A.; Connor, T.; Peacock, T.; Robertson, D.L.; Volz, E.; et al. Preliminary genomic characterization of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Available online: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (accessed on 9 December 2020).
47. Khan, A.; Zia, T.; Suleman, M.; Khan, T.; Ali, S.S.; Abbasi, A.A.; Mohammad, A.; Wei, D.-Q. Higher infectivity of the SARS-CoV-2 new variant is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J. Cell. Physiol.* **2021**, 1–13. [\[CrossRef\]](#)
48. Luan, B.; Wang, H.; Huynh, T. Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: Insights from molecular dynamics simulations. *FEBS Lett.* **2021**, 595, 1454–1461. [\[CrossRef\]](#)
49. Sikora, M.; von Bülow, S.; Blanc, F.E.C.; Gecht, M.; Covino, R.; Hummer, G. Computational epitope map of SARS-CoV-2 spike protein. *PLoS Comp. Biol.* **2021**, 17, e1008790. [\[CrossRef\]](#)
50. Yuan, M.; Wu, N.C.; Zhu, X.; Lee, C.C.D.; So, R.T.; Lv, H.; Mok, C.K.; Wilson, I.A. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, 368, 630–633. [\[CrossRef\]](#)
51. Yi, C.; Sun, X.; Ye, J.; Ding, L.; Liu, M.; Yang, Z.; Lu, X.; Zhang, Y.; Ma, L.; Gu, W. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Nat. Cell. Mol. Immunol.* **2020**, 17, 621–630. [\[CrossRef\]](#)
52. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, 581, 215–220. [\[CrossRef\]](#)
53. Georgoulia, P.S.; Glykos, N.M. Molecular simulation of peptides coming of age: Accurate prediction of folding, dynamics and structure. *Arch. Biochem. Biophys.* **2019**, 664, 76–88. [\[CrossRef\]](#)
54. Geng, H.; Chen, F.; Ye, J.; Jiang, F. Applications of molecular dynamics simulation in structure prediction of peptides and proteins. *Comp. Struct. Biotech. J.* **2019**, 17, 1162–1170. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Wen, J.; Batabyal, D.; Knutson, N.; Lord, H.; Wikström, M. A comparison between emerging and current biophysical methods for the assessment of higher-order structure of biopharmaceuticals. *J. Pharma. Sci.* **2020**, 109, 247–253. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Chan, K.K.; Dorosky, D.; Sharma, P.; Abbasi, S.A.; Dye, J.M.; Kranz, D.M.; Herbert, A.S.; Procko, E. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **2020**, 369, 1261–1265. [\[CrossRef\]](#)
57. Yang, Y.W.; Gruebele, M. Folding at the speed limit. *Nature* **2003**, 423, 193–197. [\[CrossRef\]](#)
58. Samatova, E.; Dabberger, J.; Liutkut, M.; Rodnina, M.V. Translational control by ribosome pausing in bacteria: How a non-uniform pace of translation affects protein production and folding. *Front. Microbiol.* **2021**, 11, 619430. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Vanzi, F.; Vladimirov, S.; Knudsen, C.R.; Goldman, Y.E.; Cooperman, B.S. Protein synthesis by single ribosomes. *RNA* **2003**, 9, 1174–1179. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Aschwanden, C. Why herd immunity for covid is probably impossible. *Nature* **2021**, 591, 520–522. [\[CrossRef\]](#)
61. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Higgs, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, 19, 26–59. [\[CrossRef\]](#)
62. Conrad, J.C.; Flory, P.J. Moments and distribution functions for polypeptide chains. Poly-L-alanine. *Macromolecules* **1976**, 9, 41–47. [\[CrossRef\]](#)
63. Jacobson, H.; Stockmayer, W.H. Intramolecular reactions in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **1950**, 18, 1600–1607. [\[CrossRef\]](#)
64. Plotkin, S.S.; Onuchic, J.N. Structural and energetic heterogeneity in protein folding. I. Theory. *J. Chem. Phys.* **2002**, 116, 5263–5283. [\[CrossRef\]](#)
65. Bergasa-Caceres, F.; Rabitz, H.A. Low entropic barrier to the hydrophobic collapse of the prion protein: Effects of intermediate states and conformational flexibility. *J. Phys. Chem. A* **2010**, 114, 6978–6982. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Karplus, M.; Weaver, D.L. Diffusion-collision model for protein folding. *Biopolymers* **1979**, 18, 1421–1437. [\[CrossRef\]](#)
67. Sali, A.; Shakhnovich, E.; Karplus, M. How does a protein fold? *Nature* **1994**, 369, 248–251. [\[PubMed\]](#)
68. Vendruscolo, M.; Paci, E.; Dobson, C.M.; Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001**, 409, 641–645. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Alm, E.; Baker, D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA* **1999**, 96, 11305–11310. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Muñoz, V.; Eaton, W.A. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA* **1999**, 96, 11311–11316. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Clementi, C.; Jennings, P.A.; Onuchic, J.N. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 $\beta$ . *Proc. Natl. Acad. Sci. USA* **2000**, 97, 5871–5876. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Makarov, D.E.; Plaxco, K.W. The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.* **2003**, 12, 17–26. [\[CrossRef\]](#)
73. Berezovsky, I.N.; Trifonov, E.N. Van der Waals locks: Loop-n-lock structure of globular proteins. *J. Mol. Biol.* **2001**, 307, 1419–1426. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Ittah, V.; Haas, E. Nonlocal interactions stabilize long range loops in the initial folding intermediates of reduced bovine pancreatic trypsin inhibitor. *Biochemistry* **1995**, 34, 4493–4506. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Dill, K.A. Dominant forces in protein folding. *Biochemistry* **1990**, 29, 7133–7155. [\[CrossRef\]](#)

- 
76. Pace, C.N.; Fu, H.; Fryar, K.L.; Landua, J.; Trevino, S.R.; Schell, D.; Thurlkill, R.L.; Imura, S.; Scholtz, J.M.; Gajiwala, K.; et al. Contribution of hydrogen bonds to protein stability. *Protein Sci.* **2014**, *23*, 652–661. [[CrossRef](#)]
  77. Fauchere, J.L.; Pliska, V. Hydrophobic parameters II of amino-acid side chains from the partitioning of N-Acetyl-Amino-Acid amides. *Eur. J. Med. Chem.* **1983**, *18*, 369–375.
  78. Ozkan, S.B.; Wu, G.A.; Chodera, J.D.; Dill, K.A. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11987–11992. [[CrossRef](#)]
  79. Das, K. Antivirals targeting influenza A virus. *J. Med. Chem.* **2012**, *55*, 6263–6277. [[CrossRef](#)] [[PubMed](#)]
  80. Singh, J.; Udgaonkar, J.B. Molecular mechanism of the misfolding and oligomerization of the prion protein: Current understanding and its implication. *J. Phys. Chem. B* **2015**, *54*, 4431–4442. [[CrossRef](#)] [[PubMed](#)]
  81. Navalkar, A.; Ghosh, S.; Pandey, S.; Paul, A.; Datta, D.; Maji, S.K. Prion-like p53 amyloids in cancer. *Biochemistry* **2020**, *59*, 146–155. [[CrossRef](#)]
  82. Friedler, A.; Hansson, L.O.; Veprintsev, S.B.; Freund, S.M.V.; Rippin, T.M.; Nikolova, P.V.; Proctor, M.R.; Rudiger, S.; Fersht, A.R. A Peptide that binds and stabilizes p53 core domain: Chaperone strategy for rescue of oncogenic mutants. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 937–942. [[CrossRef](#)]
  83. Moreland, J.L.; Gramada, A.; Buzko, O.V.; Zhang, Q.; Bourne, P.E. The Molecular Biology Toolkit (MBT): A modular platform for developing molecular visualization applications. *BMC Bioinform.* **2005**, *6*, 21. [[CrossRef](#)]