Journal of Proteins and Proteomics (2021) 12:1–13
https://doi.org/10.1007/s42485-021-00057-y

SHORT COMMUNICATION

Aggregation hot spots in the SARS-CoV-2 proteome may constitute potential therapeutic targets for the suppression of the viral replication and multiplication

Shalini Gour[1] · Jay Kant Yadav[1]

Received: 10 October 2020 / Revised: 23 January 2021 / Accepted: 27 January 2021 / Published online: 13 February 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. part of Springer Nature 2021

Abstract
The emergence of novel coronavirus SARS-CoV-2 is responsible for causing coronavirus disease-19 (COVID-19) imposing serious threat to global public health. Infection of SARS-CoV-2 to the host cell is characterized by direct translation of positive single stranded (+ss) RNA to form large polyprotein polymerase 1ab (pp1ab), which acts as precursor for a number of nonstructural and structural proteins that play vital roles in replication of viral genome and biosynthesis of new virus particles. The maintenance of viral protein homeostasis is essential for continuation of viral life cycle in the host cell. To test whether the protein homeostasis of SARS-CoV-2 can be disrupted by inducing specific protein aggregation, we made an effort to examine whether the viral proteome contains any aggregation prone regions (APRs) that can be explored for inducing toxic protein aggregation specifically in viral proteins and without affecting the host cell. This curiosity leads to the identification of several (>70) potential APRs in SARS-CoV-2 proteome. The length of the APRs ranges from 5 to 25 amino acid residues. Nearly 70% of total APRs investigated are relatively smaller and found to be in the range of 5–10 amino acids. The maximum number of ARPs (>50) was observed in pp1ab. On the other hand, the structural proteins such as, spike (S), nucleoprotein (N), membrane (M) and envelope (E) proteins also possess APRs in their primary structures which altogether constitute 30% of the total APRs identified. Our findings may provide new windows of opportunities to design specific peptide-based, anti-SARS-CoV-2 therapeutic molecules against COVID-19.

Keywords SARS-CoV-2 · COVID-19 · Aggregation prone regions · Protein aggregations

Introduction

The coronavirus disease 19 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first emerged in Wuhan, China now imposing a serious threat to human life all across the globe and responsible for disruption of social and economic integrity worldwide (Arabi et al. 2020; Cucinotta and Vanelli 2020; Nicola et al. 2020; Tandon 2020). So far, patients are mostly being managed by supportive treatment using lopinavir/ritonavir, ribavirin, beta-interferon, glucocorticoid and remdesivir (Antinori et al. 2020;Chan et al. 2020; Jean et al. 2020; Salvi and Patankar 2020;Srinivas et al. 2020;Sternberg et al. 2020]. In the meantime, some novel vaccine candidates and different pharmacological approaches are under investigation [as reviewed in (Scarabelet al. 2021)]. The BNT162b2- BioNTech/Pfizer and mRNA-1273-Moderna vaccines have completed their trials and been approved by FDA and EMA (Scarabelet al. 2021). In India, the use of Covishield (developed by University of Oxford, AstraZeneca and produced by SII, India) has been approved and the other indigenous vaccine candidate Covaxin (ICMR-NIV-Bharat Biotech) is in phase-III trials (ICMR report). Various other strategies are under investigation (Sanders et al. 2020) and almost 100 different vaccine candidates have been proposed (Zhang et al. 2020) and these strategies are being validated through clinical studies and trials (Cao et al. 2020; Hung et al. 2020; Wang et al. 2020; Chen et al. 2021). However, the post-efficacy strategies for the successful vaccine candidates are the prime requirement for the mass use of

✉ Jay Kant Yadav
jky.bio@gmail.com

[1] Department of Biotechnology, Central University of Rajasthan, NH-8 Bandersindri, Kishangarh, Ajmer, Rajasthan 305817, India

these vaccines (Kim et al. 2021). Therefore, in addition to the available options, it becomes imperative to search novel therapeutic targets to curtail virus infection and multiplication. The replication cycle of SARS-CoV-2 in host cell is marked with highly synchronized processes of protein expression, protein folding, and assembly of viral genome along with structural proteins lead to formation of new virus particles (Sims et al. 2008; Fehr and Perlman 2015; Chen et al. 2020; Lukassen et al. 2020; Lunget al. 2020; Malik 2020). Maintenance of protein homeostasis in a eukaryotic cell is achieved by an integrated mechanism of protein biosynthesis, folding and attainment of native structure, and the degradation of misfolded proteins (Balchin et al. 2016; Chiti and Dobson 2017; Klaips et al. 2018; Zhong et al. 2019). After the entry into the host cell, viruses employ various strategies to hijack and regulate various biochemical and molecular activities, such as transcription and translation machineries, of the host cell to produce new viral proteins and enzymes essential for multiplication of the virus (Chen et al. 2020; Malik 2020; Salvi and Patankar 2020). Translation of viral genome represents a key event required for the establishment of infection and multiplication of SARS-CoV-2. We started our prediction using the primary structures of proteins emerging from all the known open reading frames (ORFs) of the SARS-CoV-2. The genome structure of SARS-CoV-2 contains at least six ORFs. The first ORF (known as ORF1a/b) constitutes approximately two-thirds of the total genome length and encodes 16 nonstructural proteins (NSPS1-16) (Gordonet al. 2020; Malik 2020). There is a − 1 frame shift between ORF1a and ORF1b, leading to production of two polypeptides: polypeptide 1a (pp1a) and polypeptide 1ab(pp1ab) having 7096 amino acid residues. These polypeptides are proteolytically cleaved to form 16 polypeptides segments that ultimately give rise nonstructural proteins (NSPS). Chymotrypsin-like protease (3CLpro) which is virally encoded act at specific sites and help in the formation of NSPS. Other ORFs are situated at the 3′-terminus of ORF1 constitute just 1/3rd of the viral genome and encode four major structural proteins namely, spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins. The NSPS play specific roles during infection such as, degradation of host cell mRNA, inhibition of interferon (IFN) signaling, blocking the host innate immune response, promoting cytokine expression, etc. These biochemical functions of NSPS are crucial for establishment of viral infection and multiplication. The four structural proteins are vital for virion assembly and formation of new viral particles. The S protein forms a homotrimer and then form spikes on the viral surface that are responsible for initial attachment to the host receptors (Pillay 2020). The M protein has three trans-membrane domains and it shapes the virions, promotes membrane curvature, and binds to the nucleocapsid. The E protein plays a role in virus assembly and release, and it is involved in viral pathogenesis. The N protein contains two domains, which bind with virus RNA genome through an integrated action S, E and M proteins.

It has often been observed that the protein aggregation frequently disrupts the protein homeostasis leading to development of various disease conditions. Protein aggregation is generally driven by specific amino acid sequences which are interspersed within the primary structure of proteins and polypeptides, known as aggregation-prone regions (APRs). The synthetic analogs of such APRs sequences contain the ability to self-assemble to form aggregates rich β-sheet structures. Further, these APRs are shown to interact with similar sequences present in parent proteins and peptides through homologous interaction and induce aggregation. Hence, these APRs have been successfully explored for the targeted disruption of protein homeostasis. Several recent studies have confirmed that the presence of synthetic analogs of these sequence-stretches (i.e., APRs) effectively block the folding of the original proteins and render them for degradation by the proteasomal degradation machinery of the host cell (Beerten et al. 2012; De Baets et al. 2014; Gallardo et al. 2016; Ganesan et al. 2016; Khodaparast et al. 2018).To explore the possibility of targeted protein aggregation to curtail SARS-CoV-2 infection, we screened the viral proteome to find out presence of APRs. Our initial studies suggest that the primary structures of many of the key proteins such as, polyprotein polymerase 1ab (pp1ab), envelope protein (E), nucleoprotein (N), membrane (M) protein, etc., are marked by the presence of small amino acid sequence-stretches possessing high aggregation propensity. On the other hand, it has been observed that the peptide (APR)-induced protein aggregation turns out to be a highly ordered and specific process. Since these APRs form essential elements of the native proteins, in unfolded state (immediately after translation), can interact with synthetic analogs of APRs and induce aggregation of entire proteins and finally subject the protein molecules for degradation rather than their folding into functional proteins.

Methods

Prediction of the potential aggregation prone regions (APRs) in the SARS-CoV-2 proteome

The complete genome of Wuhan-Hu-1 (NC_045512.2) was downloaded from NCBI nucleotide database. The

aggregation propensity of all the SARS-CoV-2 proteins primary structure was assessed by using in silico predictions. These primary structures of the proteins were sequentially submitted to different computation algorithms namely FoldAmyloid (http://bioinfo.protres.ru/FoldAmyloid/) (Garbuzynskiy et al. 2010), TANGO (http://tango.crg.es/) (Fernandez-Escamilla et al. 2004), AGGRESCAN (http://bioinf.uab.es/aggrescan/) (Conchillo-Soleet al. 2007), and AMYLPRED (http://aias.biol.uoa.gr/AMYLPRED/input.php) (Frousios et al. 2009) with the default setting. The scores were compared with classical aggregating peptide i.e., Amyloid beta (Aβ) peptide.

Result and discussion

Prediction of APRs in different proteins emerging from different ORFs of SARS-CoV-2

Table 1 summarizes the locations of the predicted APRs in different structural and nonstructural proteins of SARS-CoV-2. The APRs are found to be asymmetrically distributed in the different regions of all the proteins investigated. As mentioned earlier, pp1a and pp1ab are the two large polypeptides that formed from direct translation of virus genome after its entry into host cells. Given the fact that 2/3rd proportion of the total virus genome is utilized for the synthesis of NSPS, they are very crucial for the continuation of virus replication cycle (Masters 2006; Chen et al. 2020). NSP-1 is the first non-structural protein formed from pp1ab, obstructs translation of host mRNA by interfering with the 40S ribosomal subunit (Raj 2021). The primary structure of NSP-1 contains 180 residues and it was found to be free from any APRs in it. Similarly, the region corresponding to NSP-13 spanning from 5325 to 5925 residues does not contain any aggregation prone regions in it. In the polypeptide segments corresponding to NSP-2 to NSP-12 and NSP-14 to NSP-16 contain several APRs. NSP-2, 637 residues in its primary structure, is the second nonstructural protein and found to have 6 potential APRs ranging from 6 to 13 residues in length. The maximum numbers of APRs are found in the segment spanning from 3570 to 3859 residues, which corresponds to NSP-6. The total APRs in this region constitute > 35% residues of the total protein. Along with NSP-3 and NSP-4, NSP-6 plays vital role in creation of cytoplasmic double-membrane vesicles essential for viral replication. On the other hand, NSP-6 also plays important role in preventing delivery of the viral components to lysosomes of the host cell and hence protects the virus from lysosomal inactivation (Gordon et al. 2020). Hence, truncating NSP-6 would be helpful in enhancing host mediated destruction of the virus. Similarly, the polypeptide regions corresponding to NSP-3 and NSP-4 consist of large number of APRs (Fig. 1).

Apart from pp1ab, the structural proteins also contain sequence stretches of high significant aggregation score. There were six potential APRs identified in S protein, however, the length of APRs use to be relatively shorter except the APR present at N-terminus of the protein. The E-protein is the smallest structural protein (75 residues) and known to play essential role in the virus morphogenesis (Liu et al. 2007), consists of a single potential APR. The M-protein constitutes an essential component of virus along with other structural proteins and plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins (Neuman et al. 2011). It consists of five APRs ranging from 7 to 18 amino acid residues. The N-protein consists of relatively less number of shorter APRs compared to other structural proteins. Among all the four structural proteins the S-protein and M-proteins are comparatively richer in the APRs compared to N and E-proteins. The score of individual APRs range from 20 to 100. However, most of the APRs have aggregation score above 50, indicative of less chance to give false positive values.

The lengths of most of the APRs identified in the SARS-CoV-2 proteome are in the range of 5–8 residues (Table 1). Most of the APRs in pp1ab possess are found to be relatively shorter in length compared to the one observed in structural proteins. It is observed that the shorter APR peptides (of ≈ 6 residues) found to be giving better prediction reliability compared to the larger one. On the other hand, it has also been established that the longer APRs possesses greater tendency to display false positives compared to the shorter ones. It has been observed that APRs of shorter length possess high aggregation propensity and interact more efficiently with the identical sequences in the large peptides or proteins compared to longer APRs.

The legitimacy of the predicted APRs is based on the reliability of mathematical and statistical lucidities. The computational algorithm TANGO uses a statistical mechanics approach to make predictions of different secondary structures present in different regions for a given proteins (Pande 2004). The algorithm assumes a particular amino acid sequence (of at least five consecutive residues) is aggregation-prone if it has high propensity to form β-sheet structure and when this sequence form aggregate all the residues of the β-region are buried in the hydrophobic interior. It predicts the aggregation propensity in a sequence specific manner and presents the data in the form of beta-aggregation score and its value range from 1 to 100. It is reported that the TANGO score of 5 per residue gives a Matthews

correlation coefficient between prediction and experiment of 0.92 (Fernandez-Escamillaet al. 2004). Further, it has been shown that the false-positive rate of TANGO is below 5% for a TANGO score of more than 15 (Bednarskaet al. 2016). Most of the classical amyloidogenic peptides possess the aggregation score above 50 and hence we gathered all the sequence stretches displaying the score above it. The overall score above 90 suggest high aggregation propensity with less probability of getting false positive. The data obtained from Tango were further analyzed by using other analogous algorithms such as Aggrescan, AmylPred and FoldAmyoid. In all the predictions we used amyloid beta (Aβ1-42) peptide as a reference due to its ability to form classical aggregates rich in β-structures. The AGGRESCAN program predicts the aggregation prone regions in a protein as "hot spot" sequences of 5 to 11 residues that can nucleate aggregation in peptides and proteins. The aggregation propensity of the hot spots is determined largely by amino acid composition, which is based on the experimentally determined aggregation propensity scale for individual amino acids. The FoldAmyloid program predicts short amino acid sequences (≥ 5 residues) based on the contacts, packing density, backbone H-bonds of acceptors or donors for prediction of aggregation prone regions. AmylPred combines the data from SecStr, a secondary structure prediction tool, to predict the amino acid sequence in protein that can act as potential conformational switch. As shown in Fig. 2, the APRs identified in different viral proteins by all the four algorithms are found to be unanimous.

Mechanistic outlook of APRs-induced disruption SARS-CoV-2 protein homeostasis

For the first time the mechanism of APR-induced disruption of protein homeostasis action was proposed by Balch et al. (2008). They showed that the disruption of bacterial protein homeostasis can be induced by small aggregating peptides resulting into formation of toxic protein aggregates in the bacterial cell. Generally, the ordered protein aggregation is facilitated through the formation of intermolecular β-structures by short polypeptide sequence with high aggregation propensity. Presence of such sequences define the basis of amyloid formation in various disease conditions, particularly the most debilitating Alzheimer's and Parkinson's diseases. Similar sequences are commonly present in various globular proteins that constitute their hydrophobic core and confer structural stability. They also assist oligomeric proteins by forming protein–protein interfaces.

Despite the fact that these sequences participate in providing stability to the native proteins, they can self-assemble with identical sequences to form β-structured aggregates in unfolded state. While forming the β-structured aggregates, it is often observed that their interactions with identical sequences in denatured proteins use to be more efficient than

Table 1 Location of newly identified aggregation prone regions in different proteins of SARS-CoV-2

	Amino acid sequences	Positions	Residues	Amino acid Sequence of APRs	Length of APRs
Polyprotein polymerase 1ab					
Nsp1	MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDS- VEEVLSEARQHLKDGTCGLVEVEKGVLPQLEQPY- VFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETL- GVLVPHVGEIPVAYRKVLLRKNGNKGAGGH- SYGADLKSFDLGDELGTDPYEDFQENWNTKHSSG- VTRELMRELNGG	1–180	180	Nil	
Nsp2	AVTRYVDNNFCGPDGYPLDCIKDFLARAGKSMCTLSE- QLDYIESKRGVYCCRDHEHEIAWFTERSDKSYEHQTP- FEIKSAKKFDTFKGECPKFVFPLNSKVKVIQPRVEKK- KTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHCDE- VSWQTCDFLKATCEHCGTENLVIEGPTTCGYLPTNAV- VKMPCPACQDPEIGPEHSVADYHNHSNIETRLRKGGR TRCFGG**CVFAYV**GCYNKRAYWVPRASADIGSGHT- GITGDNVETLNEDLLEILSRERVNINIVGDFHLNEE**VAI-** **ILASF**SASTSAFIDTIKSLDYKSFKTIVESCGNYKVTK- GKPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFAR- TLDAANHSIPDLQR**AAVTIL**DGISEQSLRLVDAMVYTS- DLLTNS**VIIMAYVTG**GLVQQTSQWLSNLLGTTVEKL- RPIFEWIEAKLSAGVEFLKD**AWEILKFLITGVF**DIVKG- QIQVASDNIKDC**VKCFIDVV**NKALEMCIDQVTIAGAK- LRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMPLKAP- KEVTFLEGDSHDTVLTSEEVVLKNGELEALETPVDS- FTNGAIVGTPVCVNGLMLLEIKDKEQYCALSPGL- LATNNVFRLKGG	181–818	637	$_{409}$CVFAYV$_{415}$ $_{473}$VAIILASF$_{480}$ 565AAVTIL$_{570}$ 595VIIMAYVTG$_{603}$ 645AWEILKFLITGVF$_{657}$ 675VKCFIDVV$_{682}$	6 8 6 9 13 8

Table 1 (continued)

	Amino acid sequences	Positions	Residues	Amino acid Sequence of APRs	Length of APRs
Nsp3	APTKVTFGDDTVIEVQGYKSVNITFELDERIDKVL- NEKCSAYTVELGTEVNEFACVVADAVIKTLQPV- SELLTPLGIDLDEWSMATYYLFDESGEFKLASH- MYCSFYPPDEDEEEGDCEEEEFEPSTQYEYGTED- DYQGKPLEFGATSAALQPEEEQEEDWLDDDSQQT- VGQQDGSEDNQTTTIQTIVEVQPQLEMELTPVVQ- TIEVNSFSGYLKLTDNVYIKNADIVEEAKKVKPTV- VVNAANVYLKHGGGVAGALNKATNNAMQVESD- DYIATNGPLKVGGSCVLSGHNLAKHCLHVVGP- NVNKGEDIQLLKSAYENFNQHEVLLAPLLSAGIF- GADPIHSLRVCVDTVRTN**VYLAVF**DKNLYDKLVS- SFLEMKSEKQVEQKIAEIPKEEVKPFITESKPS- VEQRKQDDKKIKACVEEVTTTLEETKFLTENLL- LYIDINGNLHPDSATLVSDIDITFLKKDAPYIVGDV- VQEG**VLTAVV**IPTKKAGGTTEMLAKALRKVPTDNY- ITTYPGQGLNGYTVEEAKTVLKKCKSAFYILPSIIS- NEKQEILGTVSWNLREMLAHAEETRKLMPVCVET- KAIVSTIQRKYKGIKIQEGVVDYGARFYFYT- SKTTVASLINTLNDLNETLVTMPLGYVTHGLN- LEEAARYMRSLKVPATVSVSSPDAVTAYNGYLTSSSK- TPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRG- DKSVYYYTSNPTTFHLDGEVITFDNLKTLLSLREVR- TIK**VFTTV**DNINLHTQVVDMSMTYGQQFGPTYLD- GADVTKIKPHNSHEGKTFYVLPNDDTLRVEAFEYY- HTTDPSFLGRYMSALNHTKKWKYPQVNGLTSIK- WADNNCY**LATALLT**LQQIELKFNPPALQDAYY RARAGEAAN**FCALILAY**CNKTVGELGDVRETM- SYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAV- MYMGTLSYEQFKKGVQIPCTCGKQATKYLVQQESP- FVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKH- ITSKETLYCIDGALLTKSSEYKGPITDVFYKENSYTT- TIKPVTYKLDGVVCTEIDPKLDNYYKKDNSYFTE- QPIDLVPNQPYPNASFDNFKFVCDNIKFADDLNQLT- GYKKPASRELKVTFFPDLNGDVVAIDYKHYTPSFK- KGAKLLHKPIVWHVNNATNKATYKPNTWCIR- CLWSTKPVETSNSFDVLKSEDAQGMDNLACEDLK- PVSEEVVENPTIQKDVLECNVKTTEVVGDIILK- PANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRV- LGLKTLATHGLAAVNSVPWDTIANYAKPFLNKV- VSTTTNIVTRCLNRVCTNYMP**YFFTLLL**QLCT- FTRSTNSRIKASMPTTIAKNTVKSVGKFCLEASF- NYLKSPNFSKLIN**IIIWFLLLSVCLGSLI**YSTAAL- GVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTG SIPCSVCLSGLDSLDTYPSLETIQITISSFKWDLTAFGL- **VAEWFLAYILFTRFFYV**LGLAAIMQLFFSYFAVHFIS- NS**WLMWLIINLV**QMAPISAMVRM**YIFFASFYYVW**K- SYVHVVDGCNSSTCMMCYKRNRATRVECTTIVNG- VRRSFYVYANGGKGFCKLHNWNCVNCDTFCAGST- FISDEVARDLSLQFKRPINPTDQSSYIVDSVTVKNG- SIHLYFDKAGQKTYERHSLSHFVNLDNLRANNT- KGSLP**INVIVF**DGKSKCEESSAKSASVYYSQLMC- QPILLLDQALVSDVGDSAEVAVKMFDAYVNTFSSTFN- VPMEKLKTLVATAEAELAKNVSLDNVLSTFISAARQG- FVDSDVETKDVVECLKLSHQSDIEVTGDSCNNYML- TYNKVENMTPRDLGACIDCSARHINAQVAKSHN**IALI- WNV**KDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTR QVVNVVTTKIALKGG	819–2763	1945	$_{1173}$VYLAVF$_{1178}$ $_{1295}$VLTAVV$_{1300}$ 1570VFTTV$_{1574}$ 1676LATALLT$_{1682}$ 1710FCALILAY$_{1717}$ 2171YFFTLLL$_{2177}$ 2229IIIWFLLLSVCLGSLI$_{2244}$ 2324VAEWFLAYILFTRFFYV$_{2340}$ 2363WLMWLIINLV$_{2372}$ 2384YIFFASFYYVW$_{2394}$ 2538INVIVF$_{2543}$ 2709IALIWNV$_{2715}$	6 6 5 7 8 7 16 17 10 11 6 7

Table 1 (continued)

	Amino acid sequences	Positions	Residues	Amino acid Sequence of APRs	Length of APRs
Nsp4	KIVNNWLKQLIK**VTLVFLFVAAIFYLI**TPVH- VMSKHTDFSSEIIGYKAIDGGVTRDIASTDTC- FANKHADFDTWFSQRGGSYTNDKACP**LIAAVIT-** REVGFVVPGLPGTILRTTNGDFLHFLPRVF- SAVGNICYTPSKLIEYTDFATSACVLAAECTIFK- DASGKPVPYCYDTNVLEGSVAYESLRPDTRY- VLMDGSIIQFPNTYLEGSVR**VVTTF**DSEYCRH- GTCERSEAGVCVSTSGRWVLNNDYYRSLPG- VFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGG **IVAIVVTCLAYYF**MRFRRAFGEYSHV**VAFNTLL-** **FLMSFTVLCL**TPVYSFLPG**VYSVIYLYLTFYL**TNDVS- FLAHIQWMVMFTPLVP**FWITIAYIICI**STKH**FYWFF-** SNYLKRRVVFNGVSFSTFEEAA**LCTFLL**NKEMYLK- LRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSY- REAACCHLAKALNDFSNSGSDVLYQPPQTSITSAVLQ	2764–3263	500	$_{2776}$VTLVFLFVAAIFYLI$_{2790}$ $_{2853}$LIAAVIT$_{2859}$ 2975VVTTF$_{2979}$ 3052IVAIVVTCLAYYF$_{3064}$ 3077VAFNTLLFLMSFTVLCL$_{3094}$ 3104VYSVIYLYLTFYL$_{3116}$ 3138FWITIAYIICI$_{3148}$ 3153FYWFF$_{3157}$ 3180LCTFLL$_{3185}$	15 7 5 13 17 13 11 5 6
Nsp5	SGFRKMAFPSGKVEGCMVQVTCGTTT LNGLWLDDVVYCPRHVICTSEDML- NPNYEDLLIRKSNHNFLVQAGNVQL- RVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTF- SVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVG- FNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGP- FVDRQTAQAAGTDTT**ITVNVLAWLYAAVI**NGDRWFL- NRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQT- GIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTP- FDVVRQCSGVTFQ	3264–3569	306	$_{3463}$ITVNVLAWLYAAVI$_{3476}$	14
Nsp6	SAVKRTIKGTHH**WLLLTILTSLLVLV**QSTQWSLFFF- LYENAFLPFA**MGIIAMSAFAMMFV**KHKHA**FL-** **CLFL**LPS**LATVAYFNMVY**MPASWVMRIMTWLDM- VDTSLSGFKLKDC**VMYASAVVLLILMT**ARTVYDD- GARRV**WTLMNVLTLVY**KVYYGNALDQAIS**MWALI-** **ISV**TSNYSG**VVTTVMFLA**RG**IVFMCV**EYCPIFFIT- GNTLQC**IMLVYCFLGYFCTCYF**GLFCLLNRYFRLTLG- VYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGV- GGKPCIKVATVQ	3570–3859	290	$_{3582}$WLLLTILTSLLVLV$_{3595}$ $_{3616}$MGIIAMSAFAMMFV$_{3629}$ 3635FLCLFL$_{3640}$ 3644LATVAYFNMVY$_{3654}$ 3683VMYASAVVLLILMT$_{3696}$ 3709WTLMNVLTLVY$_{3719}$ 3733MWALIISV$_{3740}$ 3747VVTTVMFLA$_{3755}$ 3758IVFMCV$_{3763}$ 3779IMLVYCFLGYFCTCYF$_{3794}$	14 14 6 11 14 11 8 9 6 16
Nsp7	SKMSDVKCTS**VVLLSVL**QQLRVESSSKLWAQCVQL- HNDILLAKDTTEAFEK**MVSLLSVLL**SMQGAVDINKL- CEEMLDNRATLQ	3860–3942	83	$_{3870}$VVLLSVL$_{3876}$ $_{3911}$MVSLLSVLL$_{3919}$	7 9
Nsp8	AIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLK- KLKKSLNVAKSEFDRDAAMQRKLEKMADQAMTQ- MYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALN- NIINNARDGCVPLNIIPLTTAAK**LMVVI**PDYNTYK- NTCDGTTFTYASALWEIQQVVDADSKIVQLSEISM- DNSPNLAWP**LIVTAL**RANSAVKLQ	3943–4140	198	$_{128}$LMVVI$_{132}$ $_{184}$LIVTAL$_{189}$	5 6
Nsp9	NNELSPVALRQMSCAAGTTQTACTDDNALAYYNTTK- GGR**FVLALL**SDLQDLKWARFPKSDGTGTIYTELEP- PCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVLGS- LAATVRLQ	4141–4253	113	$_{4180}$FVLALL$_{4185}$	6
Nsp10	AGNATEVPANST**VLSFCAFA**VDAAKAYKDY- LASGGQPITNCVKMLCTHTGTGQAITVTPEANM- DQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYV- QIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQL- REPMLQ	4254–4392	139	$_{4266}$VLSFCAFA$_{4273}$	8

Table 1 (continued)

	Amino acid sequences	Positions	Residues	Amino acid Sequence of APRs	Length of APRs
Nsp12	SADAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDI- YNDKVAGFAKFLKTNCCRFQEKDEDDNLIDSYFV- VKRHTFSNYQHEETIYNLLKDCPAVAKHDFFK- FRIDGDMVPHISRQRLTKYTMADLVYALRHFDE- GNCDTLKEILVTYNCCDDDYFNKKDWYDFVENP- DILRVYANLGERVRQALLKTVQFCDAMRNAG**IVG-** **VL**TLDNQDLNGNWYDFGDFIQTTPGSGVPV- VDSYYSLLMPILTLTRALTAESHVDTDLTKPYIK- WDLLKYDFTEERLKLFDRYFKYWDQTYHPNCVN- CLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDG- VPFVVSTGYHFRELGVVHNQDVNLHSSRLSFKE**LLVY-** **A**ADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQT- VKPGNFNKDFYDFAVSKGFFKEGSSVELKHFF- FAQDGNAAISDYDYYRYNLPTMCDIRQ**LLFVV**EVVD- KYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKAR- LYYDSMSYEDQDALFAYTKRNVIPTITQMNLKYAI- SAKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGA TVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYP- KCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYR- LANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYA NSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHR- LYECLYRNRDVDTDFVNEFYAYLRKHFSMMILSD- DAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFM- SEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVY- LPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAID- AYPLTKHPNQEYADVFHLYLQYIRKLHDELTGHMLD- MYSVMLTNDNTSRYWEPEFYEAMYTPHTVLQ	4393–5324	932	$_{4593}$IVGVL$_{4597}$ $_{4763}$LLVYA$_{4767}$ 4861LLFVV$_{4865}$	5 5 5
Nsp13	AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVIST- SHKLVLSVNPYVCNAPGCDVTDVTQLYLGGMSYY- CKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTD- FNAIATCDWTNAGDYILANTCTERLKLFAAETLKA- TEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLN- RNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVVYRG TTTYKLNVGDYFVLTSHTVMPLSAPTLVPQEHY- VRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGP- PGTGKSHFAIGLALYYPSARIVYTACSHAAVDAL- CEKALKYLPIDKCSRIIPARARVECFDKFKVN- STLEQYVFCTVNALPETTADIVVFDEISMATNY- DLSVVNARLRAKHYVYIGDPAQLPAPRTLLT- KGTLEPEYFNSVCRLMKTIGPDMFLGTCRRC PAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKG- VITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYN- SQNAVASKILGLPTQTVDSSQGSEYDYVIFTQTTE- TAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTS- LEIPRRNVATLQ	5325–5925	601		
Nsp14	AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGL- CVDIPGIPKDMTYRRLISMMGFKMNYQVNGYPN- MFITREEAIRHVRAWIGFDVEGCHATREAVGTNL- PLQLGFSTGVNLVAVPTGYVDTPNNTDFSRVSAKPP- PGDQFKHLIPLMYKGLPWNVVRIKIVQMLSDTLKN- LSDR**VVFVLW**AHGFELTSMKYFVKIGPERTCCLCDR- RATCFSTASDTYACWHHSIGFDYVYNPFMIDVQQW- GFTGNLQSNHDLYCQVHGNAHVASCDAIMTR- CLAVHECFVKRVDWTIEYPIIGDELKINAACRKVQHM- VVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFY- DAQPCSDKAYKIEELFYSYATHSDKFTDG**VCLFW**NC- NVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLY- VNKHAFHTPAFDKSAFVNLKQLPFFYYSDSPCESHG- KQVVSDIDYVPLKSATCITRCNLGGAVCRHHANEYR- LYLDAYNMMISAG**FSLWVY**KQFDTYNLWNTFTRLQ	5926–6452	527	$_{6106}$VVFVLW$_{6111}$ $_{6306}$VCLFW$_{6310}$ 6431FSLWVY$_{6436}$	6 5 6
Nsp15	SLEN**VAFNVV**NKGHFDGQQGEVPVSIINNTVYT- KVDGVDVELFENKTTLPVNVAFELWAKRNIKPVPE- VKILNNLGVDIAANTVIWDYKRDAPAHISTIGVCSMT- DIAKKPTETICAP**LTVFF**DGRVDGQVDLFRNARNG- VLITEGSVKGLQPSVGPKQASLNGVTLIGEAVK- TQFNYYKKVDGVVQQLPETYFTQSRNLQEFKPR- SQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSH- SQLGGLHLLIGLAKRFKESPFELEDFIPMDSTVKNY- FITDAQTGSSKCVCSVIDLLLDDFVEIIKSQDLSVVSKV- VKVTIDYTE**ISFMLW**CKDGHVETFYPKLQ	6453–6798	346	$_{6457}$VAFNVV$_{6462}$ $_{6571}$LTVFF$_{6575}$ 6779ISFMLW$_{6784}$	6 5 6

Table 1 (continued)

	Amino acid sequences	Positions	Residues	Amino acid Sequence of APRs	Length of APRs
Nsp16	SSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDS- ATLPKGIMMNVAKYTQLCQYLNTLTLAVPYNMRVI- HFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLND- FVSDADSTLIGDCATVHTANKWDLIISDMYDPKT- KNVTKENDSKEG**FFTYICGFI**QQKLALGGSVAI- KITEHSWNADLYKLMGH**FAWWTAFV**TNVNASS- SEAFLIGCNYLGKPREQIDGYVMHANYIFWRN- TNPIQLSSYSLFDMSKFPLKLRGTAVMSLKEGQINDM- **ILSLL**SKGRLIIRENNRVVISSDVLVNN	6799–7096	298	$_{6947}$FFTYICGFI$_{6955}$ $_{6985}$FAWWTAFV$_{6992}$ 7069ILSLL$_{7073}$	9 8 5
Spike Protein	M**FVFLVL**LPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYP- DKVFRSSVHSTQDLFLPFFSNVTWFHAIHVSGTNGT- KRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSK- TQSLLIVNNATNVVIKVCEFQFCNDP**FLGVYY**HKNNK- SWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGN- FKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSA- LEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWT AGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPL- SETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN- LCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYN- SASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEV- RQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLD- SKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAG- STPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYR**V-** **VVLSF**ELLHAPATVCGPKKSTNLVKNKCVNFN- FNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRD- PQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQD- VNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGC- LIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARS- VASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVT- TEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSF- CTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIK- DFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAG- FIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMI- AQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYR- FNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTA- SALGKLQDVVNQNAQALNTLVKQLSSNFGAISS- VLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQ- LIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGY- HLMSFPQSAPHG**VVFL**HVTYVPAQEKNFTTAPAICH- DGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNT- FVSGNCD**VVIGIV**NNTVYDPLQPELDSFKEELDKY- FKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNL- NESLIDLQELGKYEQYIKWPW**YIWLGFIAGLIAIVM-** **VTI**MLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLK- GVKLHYT	1–1273	1273	$_{2}$FVFLVL$_{7}$ $_{140}$FLGVYY$_{145}$ 510VVVLSF$_{515}$ 1060VVFL$_{1063}$ 1128VVIGIV$_{1133}$ 1215YIWLGFIAGLIAIVMVTI$_{1232}$	6 6 6 4 6 18
E-protein	MYSFVSEETGTLIVNS**VLLFLAFVVFLLVTLAIL**TALRL- CAYCCNIVNVSLVKPSFYVYSRVKNLNSSRVPDLLV	1–75	75	$_{17}$VLLFLAFVVFLLVTLAIL$_{34}$	18
M-protein	MADSNGTITVEELKKLLEQWN**LVIGFL-** **FLTWICLLQFA**YANRNRFLYIIKL**IFL-** **WLL**WP**VTLACFVLAAVY**RINWITGG **IAIAMACLVGLMWLSYFI**ASFRLFARTRSMWSFNPET- NILLNVPLHGTILTRPLLESE**LVIGAVIL**RGHLRIAGH- HLGRCDIKDLPKEITVATSRTLSYYKLGASQRVAG- DSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ	1–222	222	$_{22}$LVIGFLFLTWICLLQFA$_{38}$ $_{51}$LIFLWLL$_{57}$ 60VTLACFVLAAVY$_{71}$ 80IAIAMACLVGLMWLSYFI$_{97}$ 138LVIGAVIL$_{145}$	17 7 12 18 8
N-protein	MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGAR- SKQRRPQGLPNNTASWFTALTQHGKEDLKFPRGQG- VPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPR**W-** **YFYYL**GTGPEAGLPYGANKD**GIIWV**ATEGALNT- PKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEG- SRGGSQASSRSSSRSRNSSRNSTPGSSRGTSPARMA- GNGGDAA**LALLLL**DRLNQLESKMSGKGQQQQGQT- VTKKSAAEASKKPRQKRTATKAYNVTQAFGRRG- PEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASA- FFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFK- DQVILLNKHIDAYKTFPPTEPKKDKKKKADETQAL- PQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA	1–419	419	$_{108}$WYFYYL$_{113}$ $_{129}$GIIWV$_{133}$ 219LALLLL$_{224}$	6 5 6
Aβ (1–42) peptide	DAEFRHDSGYEVHHQK**LVFFA**EDVGSNKG**AIIGLMVG-** **GVVI**A	1–42	42	$_{17}$LVFFA$_{21}$ $_{30}$AIIGLMVGGVVI$_{41}$	5 12

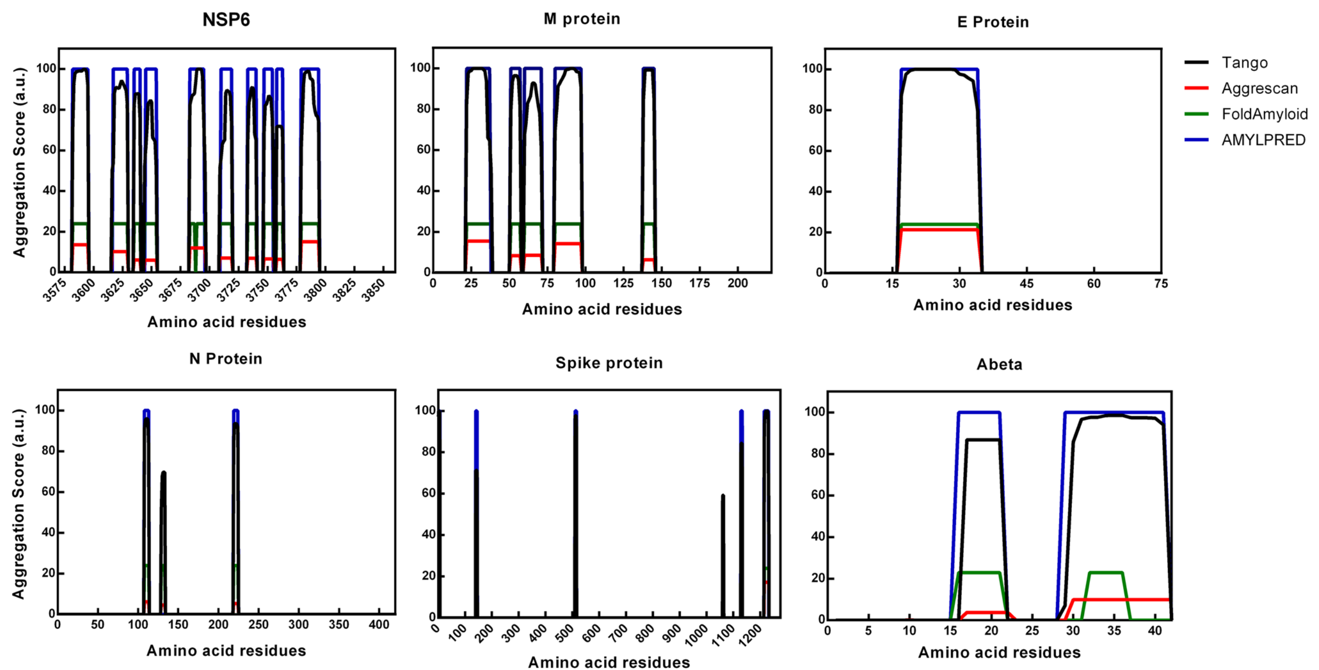

Fig. 1 Identification of aggregation prone regions (APRs) in the major proteins of SARS-CoV-2. The aggregation score and propensity in the predicted APRs found to be equivalent to the Abeta peptide, which serves as a classical β-structured aggregates

the partially similar sequences. Therefore, we hypothesize that the amino acid sequence stretches with high aggregation propensity derived from SARS-CoV-2 proteins could be able to induce specific protein aggregation leading to virucidal activity against the virus. Although, these APRs self-assemble to form β-structured aggregates and initiate seeding of identical peptides or the denatured proteins who's APRs are exposed. Therefore, to target a specific protein using APRs, it is a prerequisite that the protein must remain in unfolded state so that the interaction between homologous APRs becomes feasible. Following the viral entry, the direct translation of the pp1ab is one of the most essential steps for initiation of virus replication cycle. The APRs present in the polypeptide remain transiently exposed during translation. It is that time point when the synthetic analogs of APRs can be effectively be used to target specific proteins by interfering the protein folding reactions of polypeptide chains into functional proteins. As shown in Fig. 3, the repeated interruption of protein folding, aggregation and degradation will lead to deprivation of key proteins leading to suppression of the viral replication and multiplication. On the other hand the proteome of SARS-CoV-2 is highly specific and hence these APRs are not likely to interfere with the protein homeostasis of the host cells.

Conclusion

The maintenance of viral protein homeostasis remains as one of the most crucial steps for continuation of viral life cycle. The presence of APRs in the SARS-CoV-2 proteins constitutes susceptible proteomic segments that might act as hot spots for the commencement of the viral protein homeostasis failure. Taking the advantage of distinctive viral protein expression, folding and assembly of viral proteins, we propose a hypothesis that the disruption of protein homeostasis during viral replication will be able to prevent formation of new viral particles. Maintenance of integrated protein homeostasis is essential and remains at the highest risk during translation. Targeted aggregation of viral proteins, specifically during translation, would be able deplete the functional proteins and imposes an explicit inhibitory effect on viral replication and multiplication. The primary structures of SARS-CoV-2 proteins are marked by the presence of small unique sequences that would play vital role in inhibiting the formation of functional proteins and hence prevent the viral replication and multiplication. A recent study has shown that viral translation, splicing and nucleic

Fig. 2 Consensus among different methods for the prediction of aggregation prone regions in the different proteins of SARS-CoV-2

Amino acid sequence APRs	**Prediction tools that shows aggregation score**			
	TANGO	AGGRESCAN	FoldAmyloid	AmylPred
pp1ab				
$_{409}$CVFAYV$_{415}$	✓	✓	✓	✓
$_{473}$VAIILASF$_{480}$	✓	✓	✓	✓
$_{565}$AAVTIL$_{570}$	✓	✓	✓	✓
$_{595}$VIIMAYVTG$_{603}$	✓	✓	✓	✓
$_{645}$AWEILKFLITGVF$_{657}$	✓	✓	✓	✓
$_{675}$VKCFIDVV$_{682}$	✓	✓	✓	✓
$_{1173}$VYLAVF$_{1178}$	✓	✓	✓	✓
$_{1295}$VLTAVV$_{1300}$	✓	✓	✓	✓
$_{1570}$VFTTV$_{1574}$	✓	✓	X	X
$_{1676}$LATALLT$_{1682}$	✓	✓	✓	✓
$_{1710}$FCALILAY$_{1717}$	✓	✓	✓	✓
$_{2171}$YFFTLLL$_{2177}$	✓	✓	✓	✓
$_{2229}$IIIWFLLLSVCLGSLI$_{2244}$	✓	✓	✓	✓
$_{2324}$VAEWFLAYILFTRFFYV$_{2340}$	✓	✓	✓	✓
$_{2363}$WLMWLIINLV$_{2372}$	✓	✓	✓	✓
$_{2384}$YIFFASFYYVW$_{2394}$	✓	✓	✓	✓
$_{2538}$INVIVF$_{2543}$	✓	✓	✓	✓
$_{2709}$IALIWNV$_{2715}$	✓	✓	✓	✓
$_{2776}$VTLVFLFVAAIFYLI$_{2790}$	✓	✓	✓	✓
$_{2853}$LIAAVIT$_{2859}$	✓	✓	✓	✓
$_{2975}$VVTTF$_{2979}$	✓	✓	✓	X
$_{3052}$IVAIVVTCLAYYF$_{3064}$	✓	✓	✓	✓
$_{3077}$VAFNTLLFLMSFTVLCL$_{3094}$	✓	✓	✓	✓
$_{3104}$VYSVIYLYLTFYL$_{3116}$	✓	✓	✓	✓
$_{3138}$FWITIAYIICI$_{3148}$	✓	✓	✓	✓
$_{3153}$FYWFF$_{3157}$	✓	✓	✓	✓
$_{3180}$LCTFLL$_{3185}$	✓	✓	✓	✓
$_{3463}$ITVNVLAWLYAAVI$_{3476}$	✓	✓	✓	✓
$_{3582}$WLLLTILTSLLVLV$_{3595}$	✓	✓	✓	✓
$_{3616}$MGIIAMSAFAMMFV$_{3629}$	✓	✓	✓	✓
$_{3635}$FLCLFL$_{3640}$	✓	✓	✓	✓
$_{3644}$LATVAYFNMVY$_{3654}$	✓	✓	✓	✓
$_{3683}$VMYASAVVLLILMT$_{3696}$	✓	✓	✓	✓
$_{3709}$WTLMNVLTLVY$_{3719}$	✓	✓	✓	✓
$_{3733}$MWALIISV$_{3740}$	✓	✓	✓	✓
$_{3747}$VVTTVMFLA$_{3755}$	✓	✓	✓	✓
$_{3758}$IVFMCV$_{3763}$	✓	✓	✓	✓
$_{3779}$IMLVYCFLGYFCTCYF$_{3794}$	✓	✓	✓	✓
$_{3870}$VVLLSVL$_{3876}$	✓	✓	✓	✓
$_{3911}$MVSLLSVLL$_{3919}$	✓	✓	✓	✓
$_{4180}$FVLALL$_{4185}$	✓	✓	✓	✓
$_{4266}$VLSFCAFA$_{4273}$	✓	✓	✓	✓
$_{4593}$IVGVL$_{4597}$	✓	✓	✓	X
$_{4763}$LLVYA$_{4767}$	✓	✓	✓	X
$_{4861}$LLFVV$_{4865}$	✓	✓	✓	✓

acid metabolism constitute viable therapeutic targets for COVID-19 (Bojkova et al. 2020). Hence, the development of an exclusive and multi-target strategy to disrupt the protein homeostasis will represent an attractive and potential anti-SARS-CoV-2 strategy. At present, we are actively engaged in synthesizing all peptides analogous to the identified APRs and characterizing their biophysical characteristics and we hope that the APR-induced proteostatic disruptions will provide an innovative approach to fight with COVID-19.

Fig. 2 (continued)

$_{6106}$VVFVLW$_{6111}$	✓	✓	✓	✓
$_{6306}$VCLFW$_{6310}$	✓	✓	✓	✓
$_{6431}$FSLWVY$_{6436}$	✓	✓	✓	X
$_{6457}$VAFNVV$_{6462}$	✓	✓	✓	X
$_{6571}$LTVFF$_{6575}$	✓	✓	✓	X
$_{6779}$ISFMLW$_{6784}$	✓	✓	✓	X
$_{6947}$FFTYICGFI$_{6955}$	✓	✓	✓	✓
$_{6985}$FAWWTAFV$_{6992}$	✓	✓	✓	✓
$_{7069}$ILSLL$_{7073}$	✓	✓	✓	X
Spike protein				
$_{2}$FVFLVL$_{7}$	✓	✓	✓	✓
$_{140}$FLGVYY$_{145}$	✓	✓	✓	✓
$_{510}$VVVLSF$_{515}$	✓	✓	✓	✓
$_{1060}$VVFL$_{1063}$	✓	X	X	X
$_{1128}$VVIGIV$_{1133}$	✓	✓	✓	✓
$_{1215}$YIWLGFIAGLIAIVMVTI$_{1232}$	✓	✓	✓	✓
E-protein				
$_{17}$VLLFLAFVVFLLVTLAIL$_{34}$	✓	✓	✓	✓
M-protein				
$_{22}$LVIGFLFLTWICLLQFA$_{38}$	✓	✓	✓	✓
$_{51}$LIFLWLL$_{57}$	✓	✓	✓	✓
$_{60}$VTLACFVLAAVY$_{71}$	✓	✓	✓	✓
$_{80}$IAIAMACLVGLMWLSYFI$_{97}$	✓	✓	✓	✓
$_{138}$LVIGAVIL$_{145}$	✓	✓	✓	✓
N-protein				
$_{108}$WYFYYL$_{113}$	✓	✓	✓	✓
$_{129}$GIIWV$_{133}$	✓	✓	✓	X
$_{219}$LALLLL$_{224}$	✓	✓	✓	✓

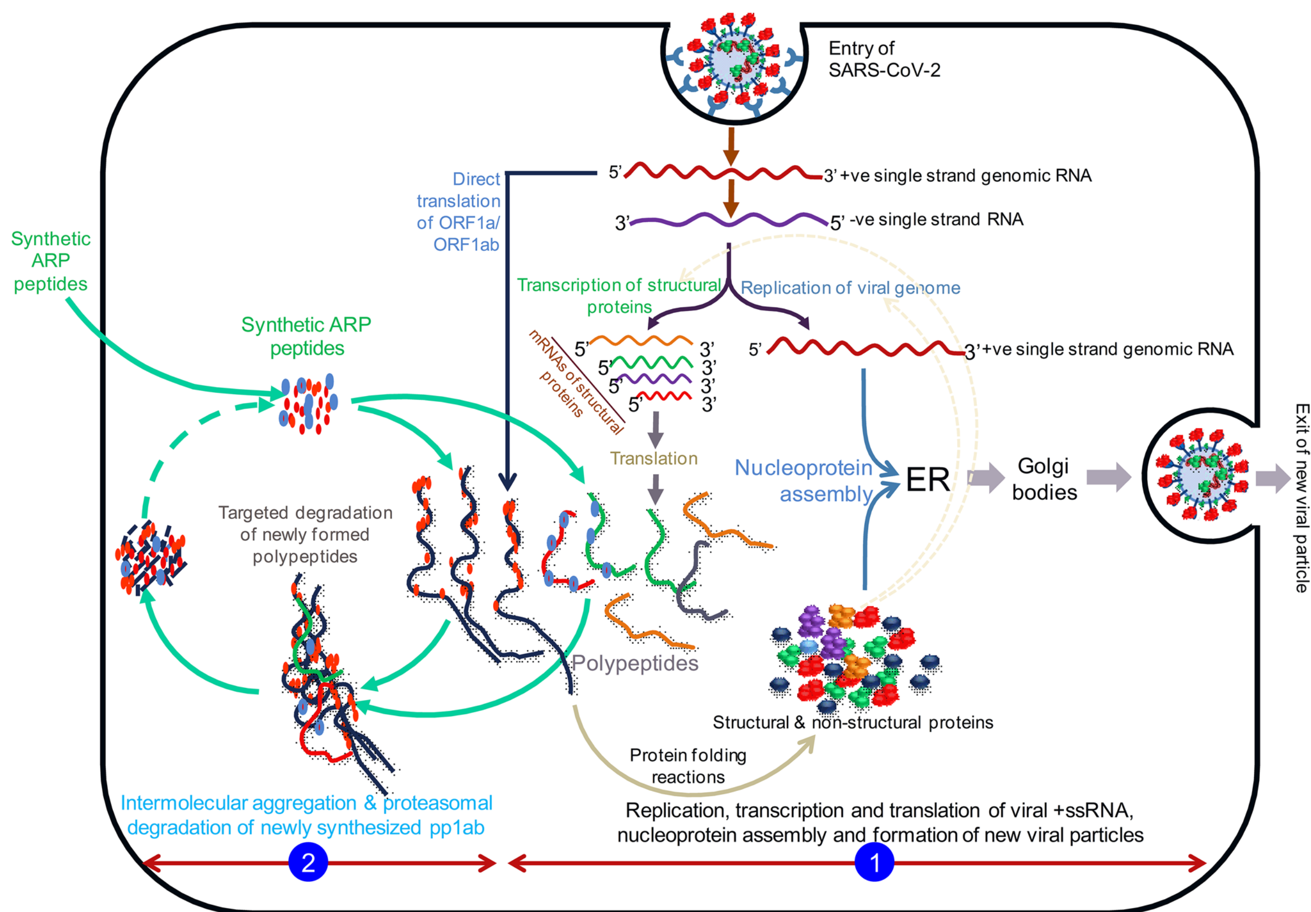

Fig. 3 Schematic representation of APR peptide-based inhibition of viral replication. The events in the region one represents usual cycle of infection, release of viral + ssRNA AND its direct translation to form pp1ab which subsequently forms all the nonstructural proteins (NSPS). The NSPS, are used in amplification of viral genomic + ssRNA, formation of structural and other accessory proteins. At the end genomic + ssRNA assemble with structural proteins to form new viral particles. The events depicted in the region 2 (left side) depict the events leading to APR peptide-based targeting of proteins formed from ORF1a/ORF1ab (pp1ab). Addition of APR peptides will interfere the protein folding reaction of viral proteins and subject them for proteasomal degradation in the host cell. Depletion of essential viral proteins will lead to complete halt of the viral replication and formation of new viral particles

References

Antinori S, Cossu MV, Ridolfo AL, Rech R, Bonazzetti C, Pagani G et al (2020) Compassionate remdesivir treatment of severe Covid-19 pneumonia in intensive care unit (ICU) and Non-ICU patients: clinical outcome and differences in post-treatment hospitalisation status. Pharmacol Res 11:104899

Arabi YM, Murthy S, Webb S (2020) COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med 46(5):833–836

Balchin D, Hayer-Hartl M, Hartl FU (2016) In vivo aspects of protein folding and quality control. Science 353(6294):aac4354

Bednarska NG, van Eldere J, Gallardo R, Ganesan A, Ramakers M, Vogel I et al (2016) Protein aggregation as an antibiotic design strategy. Mol Microbiol 99(5):849–865

Beerten J, Schymkowitz J, Rousseau F (2012) Aggregation prone regions and gatekeeping residues in protein sequences. Curr Top Med Chem 12(22):2470–2478

Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S et al (2020) Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature 583(7816):469–472

Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G et al (2020) A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19. New England J Med 382(19):1787–1799

Chan KW, Wong VT, Tang SCW (2020) COVID-19: an update on the epidemiological, clinical, preventive and therapeutic evidence and guidelines of integrative Chinese-Western medicine for the management of 2019 novel coronavirus disease. Am J Chinese Med 48(3):737–762

Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol 92(4):418–423

Chen P, Nirula A, Heller B, Gottlieb RL (2021) SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with Covid-19. N Engl J Med 384(3):229–237

Chiti F, Dobson CM (2017) Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. Annu Rev Biochem 86:27–68

Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 27(8):65

Cucinotta D, Vanelli M (2020) WHO declares COVID-19 a pandemic. Acta Bio-med Atenei Parmensis 91(1):157–160

De Baets G, Schymkowitz J, Rousseau F (2014) Predicting aggregation-prone sequences in proteins. Essays Biochem 56:41–52

Fehr AR, Perlman S (2015) Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 1282:1–23

Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22(10):1302–1306

Frousios KK, Iconomidou VA, Karletidi C-M, Hamodrakas SJ (2009) Amyloidogenic determinants are usually not buried. BMC Struct Biol 9(1):44

Gallardo R, Ramakers M, De Smet F, Claes F, Khodaparast L, Khodaparast L et al (2016) De novo design of a biologically active amyloid. Science 354(6313):18

Ganesan A, Siekierska A, Beerten J, Brams M, Van Durme J, De Baets G et al (2016) Structural hot spots for the solubility of globular proteins. Nature Commun 7:10816

Garbuzynskiy SO, Lobanov MY, Galzitskaya OV (2010) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics (Oxford, England) 26(3):326–332

Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 10:26

Hung IF, Lung KC, Tso EY, Liu R, Chung TW, Chu MY et al (2020) Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. Lancet (London, England) 395(10238):1695–1704

Jean SS, Lee PI, Hsueh PR (2020) Treatment options for COVID-19: the reality and challenges. J Microbiol Immunol Infect 53(3):436–443

Khodaparast L, Khodaparast L, Gallardo R, Louros NN, Michiels E, Ramakrishnan R et al (2018) Aggregating sequences that occur in many proteins constitute weak spots of bacterial proteostasis. Nat Commun 9(1):866

Kim JH, Marks F, Clemens JD (2021) Looking beyond COVID-19 vaccine phase 3 trials. Nat Med. https://doi.org/10.1038/s41591-021-01230-y

Klaips CL, Jayaraj GG, Hartl FU (2018) Pathways of cellular proteostasis in aging and disease. J Cell Biol 217(1):51–63

Liu DX, Yuan Q, Liao Y (2007) Coronavirus envelope protein: a small membrane protein with multiple functions. Cell Mole life Sci CMLS 64(16):2043–2048

Lukassen S, Chua RL, Trefzer T, Kahn NC, Schneider MA, Muley T et al (2020) SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J 4:e105114

Lung J, Lin YS, Yang YH, Chou YL, Shu LH, Cheng YC et al (2020) The potential chemical structure of anti-SARS-CoV-2 RNA-dependent RNA polymerase. J Med Virol 92(6):693–697

Malik YA (2020) Properties of coronavirus and SARS-CoV-2. Malaysian J Pathol 42(1):3–11

Masters PS (2006) The molecular biology of coronaviruses. Adv Virus Res 66:193–292

Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S et al (2011) A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol 174(1):11–22

Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C et al (2020) The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. Internat J Surg 78:185–193

Pande VS (2004) A universal TANGO? Nat Biotechnol 22(10):1240–1241

Pillay TS (2020) Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. J Clin Pathol 73(7):366–369

Raj R (2021) Analysis of non-structural proteins, NSPs of SARS-CoV-2 as targets for computational drug designing. Biochem Biophys Rep 25:100847

Salvi R, Patankar P (2020) Emerging pharmacotherapies for COVID-19. Biomed Pharma 14:110267

Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB (2020) Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. JAMA 323(18):1824–1836

Scarabel L, Guardascione M, Dal Bo M, Toffoli G (2021) Pharmacological strategies to prevent SARS-CoV-2 infection and to treat the early phases of COVID-19 disease. Internat Soc Infect Dis 10:26

Sims AC, Burkett SE, Yount B, Pickles RJ (2008) SARS-CoV replication and pathogenesis in an in vitro model of the human conducting airway epithelium. Virus Res 133(1):33–44

Srinivas P, Sacha G, Koval C (2020) Antivirals for COVID-19. Cleveland Clinic J Med 10:18–24

Sternberg A, McKee DL, Naujokat C (2020) Novel drugs targeting the SARS-CoV-2/COVID-19 machinery. Curr Top Med Chem 10:25

Tandon PN (2020) COVID-19: Impact on health of people & wealth of nations. Indian J Med Res 151(2 & 3):121–123

Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y et al (2020) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet (London, England) 395(10236):1569–1578

Zhang J, Zeng H, Gu J, Li H, Zheng L, Zou Q (2020) Progress and Prospects on Vaccine Development against SARS-CoV-2. Vaccines 8(2):9

Zhong M, Lee GM, Sijbesma E, Ottmann C, Arkin MR (2019) Modulating protein-protein interaction networks in protein homeostasis. Curr Opin Chem Biol 50:55–65

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.