THE JOURNAL OF PHYSICAL CHEMISTRY B

pubs.acs.org/JPCB Article

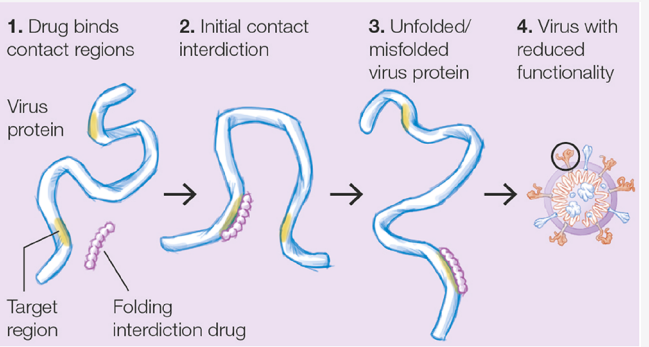
Interdiction of Protein Folding for Therapeutic Drug Development in SARS CoV-2

Fernando Bergasa-Caceres* and Herschel A. Rabitz

Cite This: https://dx.doi.org/10.1021/acs.jpcb.0c03716 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In this article, we predict the folding initiation events of the ribose phosphatase domain of protein Nsp3 and the receptor binding domain of the spike protein from the severe acute respiratory syndrome (SARS) coronavirus-2. The calculations employ the sequential collapse model and the crystal structures to identify the segments involved in the initial contact formation events of both viral proteins. The initial contact locations may provide good targets for therapeutic drug development. The proposed strategy is based on a drug binding to the contact location, thereby aiming to prevent protein folding. Peptides are suggested as a natural choice for such protein folding interdiction drugs.



1. INTRODUCTION

Understanding the biochemistry of the new coronavirus SARS-CoV-2 has become an issue of prime importance and urgency as the virus spread has triggered an ongoing pandemic that has already cost thousands of lives and large economic disruption.[1−3] Many therapeutic strategies are being considered including monoclonal antibodies[4,5] that rely on targeting selected virus proteins based on their native structure. Considerable work has been developed against coronaviruses in the past in this direction, and the experience gained is now being deployed against SARS-CoV-2 based on the crystal structures already available of several of the virus' proteins.[6−9] The purpose of this paper is to introduce a distinct possible new therapy route by (a) presenting predictions of the earliest events along the folding pathway of two of the virus' proteins and (b) building on this foundation to propose an alternative drug development strategy based on reducing the functionality of the virus by interdicting in the folding process of its proteins.

In order to fulfill goal (a) of the article, we will focus on the folding initiation events of two of the SARS-CoV-2 proteins: (1) the ADP ribose phosphatase domain of the nonstructural NsP3 protein[10,11] and (2) the receptor binding domain of the spike protein.[7,12,13] The earliest folding initiation event in both cases will be predicted employing the sequential collapse model (SCM).[14,15] In the SCM, the multistate folding process of proteins longer than ∼100 amino acids is initiated by formation of specific nonlocal contacts called primary contacts. These primary contacts help constrain the folding process by dividing the protein into several smaller domains. In this way, overall folding becomes vastly more efficient than a purely random statistical search, resulting in what we have referred to as "Nature's shortcut" to the native structure.[15] Nucleation of the folding process by the primary contacts then constitutes a potential set of emergent natural physical constraints that sidestep Levinthal's paradox.[15] In the case of misfolded protein-based diseases, the SCM has been applied to investigate some general properties of the folding dynamics of neuropathogenic proteins.[16,17] In order to unequivocally determine the folding initiation events, we will support the SCM predictions with structural information from the crystal structures of both proteins.

Goal (b) of this article will be addressed utilizing the SCM predictions to provide potential target (intraprotein) regions for the development of therapeutic drugs able to interdict the folding initiation event. Various possible therapeutic drugs could be considered, but peptides[18] form a natural class for target region binding, ideally preventing subsequent folding, although other molecular categories could also be similarly employed. This therapeutic drug development strategy based on folding interdiction of target regions (FITRs) is similar to an earlier proposal to develop drugs to interfere in protein folding.[19,20] The novelty in the present work lies in the SCM's ability to predict critical target regions for folding initiation

Received: April 27, 2020
Revised: August 7, 2020
Published: August 10, 2020

from the primary sequence. The broader potential of the FITR strategy and possible development hurdles will also be discussed. In particular, it will be explained that the proposed FITR drug strategy could be extended to other proteins of SARS-CoV-2 as well as to other diseases in which the presence of specific proteins plays a decisive role.

2. METHODS: THE SCM MODEL

The physical basis of the SCM and its most up-to-date formulation have been recently explained in full detail.[15,17] Here, only a brief summary of the main concepts is presented that are relevant to the issues investigated in the present paper.

2.1. SCM Entropic Cost of Loop Formation. The SCM considers early specific nonlocal 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.[15] In the SCM, two different loop regimes are considered when analyzing early nonlocal contacts: short loops for which the gyration radius, $R_g(n)$, is smaller than the average side chain length $£(n)$ [i.e., $R_g(n) < £(n)$], and long loops for which $R_g(n) > £(n)$. The loop length at which the transition between the short and long loops takes place [i.e., the length for which $R_g(n) \approx £(n)$] 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 $£(n) \approx 7.9$ Å,[17] 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.[21] 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 $£(n) \approx 6.7$ Å,[17] which exhibits deviations from Gaussian statistics because of steric hindrance when $n < 50$ amino acids.[22]

The long loop regime is physically equivalent to the classical Flory–Jacobson–Stockmayer (FJS) 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[23]

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

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 ΔS_{loop} approximately becomes

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

with $\Delta S_{side\text{-}chain\text{-}crowding} \ll 0$, opposing folding. When $R_g(n) \geq £(n)$, we have $\Delta S_{side\text{-}chain\text{-}crowding} \approx 0$ and the standard FJS regime is recovered. The side chain crowding term $\Delta S_{side\text{-}chain\text{-}crowding}$ 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[15]

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

where $f_0(n,£)$ is the average configurational freedom per amino acid in the unfolded chain and $f_{loop}(n,£)$ 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.[24]

Based on the model developed above, in the SCM, the folding of proteins with more than ∼100 amino acids likely involves the formation of an early nonlocal contact, called the primary contact within the SCM, that defines the earliest folding phase with $n \geq n_{op} \approx 65$ amino acids. As only a few primary contacts can be established at most in proteins of length $n \geq n_{op}$, most of the tertiary structure contacts will still be defined by contacts at a shorter range established in later folding phases.[14] Formation of the primary contact in the SCM defines the primary loop, which subsequently collapses through two-state kinetics.[15] Because proteins longer than ∼100 amino acids do not generally undergo two-state collapse[15] but rather fold through multistep 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 nonlocal contacts is not exclusive to the SCM, having arisen earlier in the context of the diffusion–collision model[25] and in the energy landscape picture.[26] It also has appeared in simulations of the transition state of two-state folding proteins.[27] Also, protein topology has been considered an essential element of folding mechanisms in a number of theoretical efforts.[28−32] The particular feature in the SCM is that the early nonlocal contacts are highly specific as in the loop hypothesis,[33] and a methodology is developed to derive their location from primary sequence information.[15]

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

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

Here, ΔG_{loop} represents the entropic free energy cost of the loop as discussed in Section 2.1. The term $\Delta G_{int,H}$ 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,[34] and its value satisfies $\Delta G_{int} < 0$. The term $\Delta G_{cont,S} > 0$ represents the entropic cost of constraining the side chains of the amino acids defining the contact such that the contact is stable and it opposes contact formation. A segment-specific determination of the value $\Delta G_{cont,S}(n_{cont})$ for a given contact would require detailed molecular dynamics techniques. However, a heuristic estimate can be made from earlier work within the SCM, which showed that the average entropic cost

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX

of folding per amino acid for a sample of 13 proteins was $\Delta G_{\text{folding/residue,S}} \approx 0.85kT$/residue,[35] and the maximum was $\Delta G_{\text{folding/residue,S}} \approx 1.09kT$/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_{cont} amino acids is approximately $\Delta G_{\text{folding/residue,S}}$ determined by the number of residues defining the contact, such that $\Delta G_{\text{cont,S}}(n_{\text{cont}}) \approx 1.09n_{\text{cont}}$. This result is clearly an approximation, but in Section 3 it will be shown to suffice for establishing a cut-off in the number of possible contacts that is consistent with the available structural data.

Hydrophobic interactions are well understood to constitute the main driving force of the folding process.[34] Other interactions such as hydrogen bonds are weaker[34] or like disulfide bonds and salt bridges form later along the folding pathway.[36] 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 eq 4 can be written as

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

As the hydrophobic stabilization energy of the contact ΔG_{hyd} is determined by the hydrophobicity of the segments involved, the hydrophobicity values h_k are obtained from the Fauchere–Pliska scale[37] and assigned to each residue in accord with previous calculations within the SCM.

Because 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.[38] Here the typical early contact segment size will be taken to be ∼5 amino acids in line with previous calculations within the SCM.[14] 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 $w \sim 2£(n) \approx$ 15.8 Å, and a peptide bond length of 3.5 Å, the minimum number n_{cont} of amino acids that can define a contact in the open fluctuating chain should be $n_{\text{cont}} \sim \text{int}[2£(n)/3.5] = 5$ amino acids. The results for the location of the most stable primary contact were seen to be robust to the employment of five and six-amino acid windows, while some deviations were observed when the window was reduced to four 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 ∼0.45 is equivalent to a change in energy of kT, with the margin of error being $\sim 0.1kT$[35]).

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_{op} amino acids apart along the sequence, and no longer than the maximum primary loop length of ∼100 amino acids, to give a contact stability of

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

3. RESULTS

We have chosen to focus in this paper on two domains of two major functional proteins of SARS-CoV-2: (a) the non-structural ADP ribose phosphatase domain of protein NsP3;[10] and (b) the structural receptor binding domain of the spike protein.[12] This choice was made based on two distinct considerations: (1) to study both a structural and a non-structural protein that have a direct involvement in the viral infection mechanism, thus providing options for drug discovery; and (2) to employ the SCM within the boundaries of its demonstrated applicability, that is, on proteins long enough that a multi-state folding pathway is expected, but not so long that degeneracies in the results might cloud any definite conclusions.[15]

3.1. Primary Contact for the ADP Ribose Phosphatase Domain (X Domain) of Protein NsP3 of SARS-CoV-2.

Non-structural proteins of coronaviruses have been the object of intense study, concerning both their structures and their functionality.[39,40] The multi-domain non-structural protein 3 (Nsp3) is the largest protein encoded by the coronavirus' genome.[10] It includes up to sixteen domains, of which eight domains and two transmembrane regions are conserved.[41,42] One of the conserved domains is the macrodomain (also called the X domain), which includes 173 amino acids.[7] The first available crystal structure of a NsP3 domain of any coronavirus was the unliganded X domain of SARS-CoV, obtained in 2005.[43] The crystal structure for the SARS-CoV-2 variant of the X domain is known.[11] It has the typical X domain organization, with seven β-strands defining a central β-sheet, surrounded by six α-helices.[11] One of the functions of the X domain is to bind ADP-ribose and poly-(ADP-ribose).[43−45] The X domains of coronaviruses, also show ADP-ribose-1″-phosphate phosphatase activity.[43,44,46] It has been observed that this property is linked to the ability of the virus to compromise the immune system of the host.[47,48]

Our calculations predict that the best possible primary contact for the X domain (i.e., the best folding initiation contact) is established between segments (^{35}PTVVV^{39}) centered at V37, and (^{125}LLAPL^{129}) centered at A127, with a stability of $\Delta G_{\text{cont}} \approx -6.3kT$. The predicted primary contact is in clear proximity in the crystal structure of the protein (PDB code 6VXS), and we have represented the two segments in Figure 1a.[49] The next-best possible contacts with energies within $\sim 1kT$ of the best primary contact (37, 127) are shown in Table 1. None of the possible alternatives to contact (37, 127) is a good native contact on the crystal structure. None of the side chains of the two segments defining these alternative contacts appear within the van der Waals interaction distance in the crystal structure. The issue of the multiplicity of possible primary contacts in the SCM has been considered in previous work.[50] Because no major rearrangements of the protein core are expected post-collapse, it is generally assumed within the model that primary contacts that are non-native on the 3D folded structure are likely not to correspond to native pathways leading to the functional folded structure.[51,52] In all the proteins studied to date within the SCM it has been observed that the most stable primary contact is native-like.[15] Thus, our result implies that the majority of the protein molecules are

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX

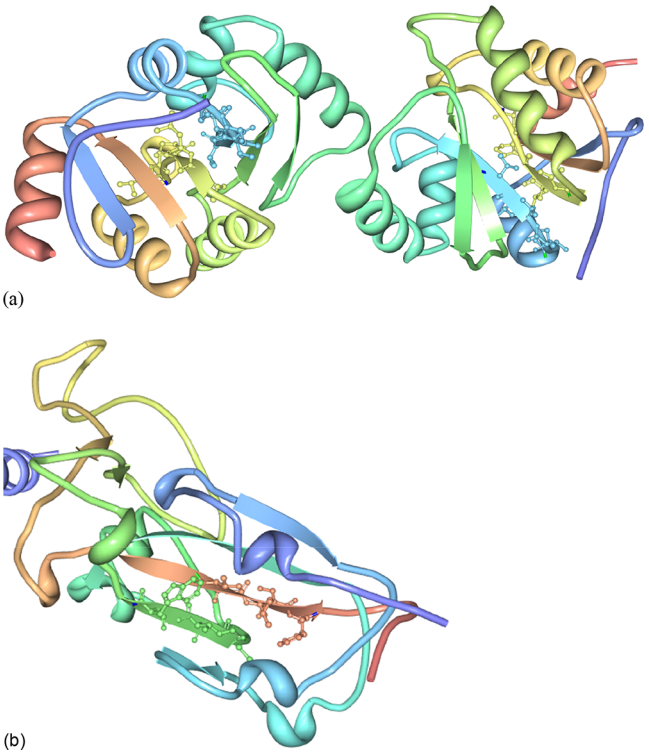


Figure 1. (a) Best primary contact on the dimeric crystal structure of the ribose phosphatase of Nsp3 from the SARS coronavirus-2 (PDB: 6VXS), represented on both identical monomers of the crystal structure to show different perspectives; (b) predicted best primary contact for the receptor binding domain from the SARS coronavirus-2 (PDB: 6M0J). Formation of the primary contact is the folding initiation event in the SCM. The figure has been produced employing Protein Workshop.[49] 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.

Table 1. Possible Primary Contacts of the Ribose Phosphatase of SARS-CoV-2 within ~1*kT* of the Best Primary Contact (37, 127)[a]

contact	ΔG_{cont}	position in the structure
37–127	−6.3	native
86–157	−5.8	non-native
97–169	−5.9	non-native

[a]Only the best primary contact is native on the 3D structure.

expected to fold through the initiation event defined by the primary contact predicted here. In this regard, it is a prediction of the model that primary contact (37−127) is the gateway to "Nature's shortcut"[15] to the folding of the ADP ribose phosphatase domain of SARS-CoV-2.

3.2. Primary Contact of the Receptor Binding Domain of the Spike Protein of SARS-CoV-2. The receptor recognition mechanisms of coronaviruses have been extensively studied.[53,54] In particular, for SARS-CoV-2[7] and its earlier viral variant SARS-CoV,[53] entry into the host's cells is mediated by a virus-surface spike protein that includes a specific receptor-binding domain (RBD).[54−56] The RBD recognizes angiotensin-converting enzyme 2 (ACE2) as its specific receptor.[55] The structure of SARS-CoV RBD is well known,[56] and the structure of SARS-CoV-2 RBD is similar,[7] albeit with some specific mutations in the ACE2 binding ridge that enhance the ability of SARS-CoV-2 to bind human ACE2.[7] The spike protein of SARS-CoV-2 is a natural therapeutic target given its critical biological role in facilitating the virus entry in the cell. The search for inhibitors, including peptides, that actively block RBD-ACE2 binding of coronavirus has been an active field of investigation for many years.[57−59] The RBD of SARS-CoV-2 is 223 amino acids long, making it the longest protein investigated within the SCM to date.[14,15,17]

The best possible contacts with energies within ~1*kT* of the best primary contact (37, 127) are shown in Table 2. Our calculations predict that the best possible primary contact is established between segments (^{114}CVIAW^{118}) centered at I116, and (^{193}VVLSF^{197}) centered at L195, 79 residues apart, with a stability of $\Delta G_{cont} \approx -11.5kT$. The predicted contact is in clear proximity in the crystal structure of the protein (PDB code 6M0J), and we have represented the two segments defining the contact in Figure 1b.[49] The second-best possible contact is defined by segments (^{17}LCPFG^{21}), centered at P19, and (^{114}CVIAW^{118}), 97 residues apart, with a stability of $\Delta G_{cont} \approx -10.8kT$. Contact (19, 116) is not as good a contact on the native structure. Thus, our result implies that the majority of the protein molecules will fold through the initiation event defined by the best primary contact (116, 195) predicted here.

Table 2. Possible Primary Contacts of the Receptor Binding Domain of the Spike Protein of SARS-CoV-2 within ~1*kT* of the Best Primary Contact (116, 195)[a]

contact	ΔG_{cont}	position in the structure
116–195	−11.5	native
19–116	−10.8	non-native

[a]Only the best primary contact is native on the 3D structure.

4. DISCUSSION: A POSSIBLE AVENUE TO PROTEIN FOLDING-INTERDICTING THERAPEUTIC DRUGS AGAINST SARS-COV-2

The identification of the primary contacts along the folding pathway of viral proteins constitutes an important result for at least two reasons: (a) the sequences of the specific segments involved in the primary contacts provide a template to specify candidate peptide drugs of inhibitory effect with the maximum possible contact affinity to compete with the natural folding mechanism; and (b) it provides insight for further investigation into the subsequent folding steps leading to a fully functional viral protein, potentially providing for additional FITRs.

The fact that the primary contact is defined by the interaction between two well defined amino acid sequences suggests that a strategy to develop FITR-based therapeutic drugs could be one utilizing trial peptide drugs as suggested above. Peptide drugs offer several advantages versus other more classical approaches such as function-blocking monoclonal antibodies. In particular, being much smaller and flexible, peptide drugs can much more easily cross the cellular membrane to reach their intended targets.[60] However, designing therapeutically effective peptide drugs remains an important challenge.[61] There are several reasons why effective peptide drugs are hard to discover: (1) the potential space of peptide candidates is very large; (2) ensuring delivery at the right location on the target molecule is a considerable challenge; and (3) making a peptide drug that is contact-site

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX

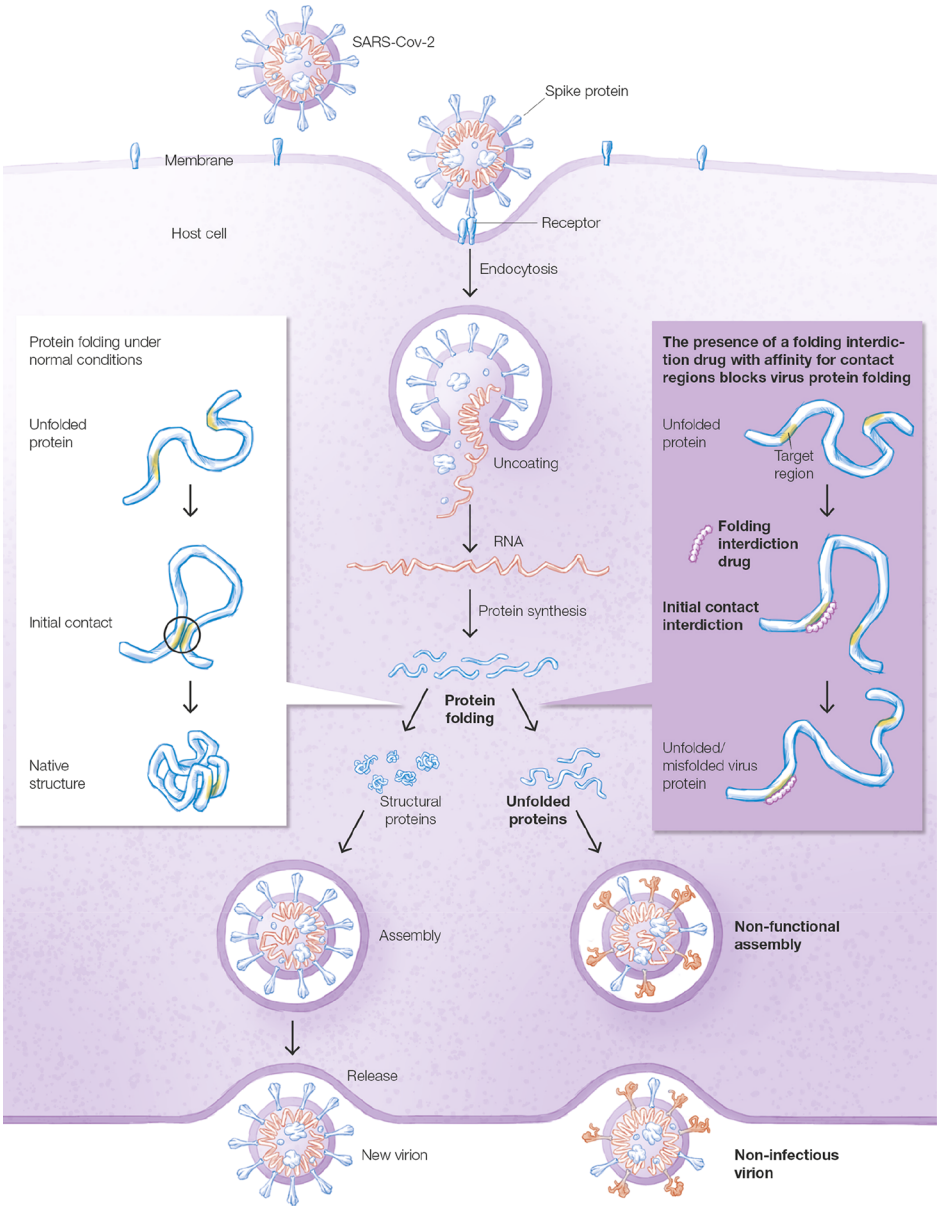


Figure 2. Proposed inhibitory mechanism of viral functionality based on the employment of specific peptides in order to interdict the initial folding event of the viral proteins. The effect on the viral structure of the employment of such folding inhibitory peptides on the spike proteins of SARS-Cov-2 is illustrated in generic form as its precise effect on the overall configuration of the virion is not known. Additionally, multiple spike proteins would need to be inhibited simultaneously to fully disrupt the functionality of the virion.

specific is not an easy task. As a consequence, no more than ~60 effective peptide drugs are in use today,[18] although there is active investigation of many more.[18] The SCM might prove of assistance in addressing at least difficulty (1) above, and maybe also (2) and (3), through the utilization of the predicted primary contact sequence as a template to search for an effective therapeutic peptide capable of inhibiting primary contact formation. Also, other non-peptide molecules have shown potentially therapeutic capabilities by binding to mostly unfolded states of proteins. For example, ceftriaxone binds specifically to the C-terminal region of the intrinsically disordered protein α-synuclein,[62] generally understood to be involved in triggering Parkinson's disease,[63] and has shown therapeutic potential. Also, recent experimental results suggest that the folding kinetics of proteins with two-state transitions can be modulated by the employment of suitable peptides that mimic specific segments of the protein chain.[64] Although the purpose of the experiment was the opposite of the one sought for here, as the goal was to speed up the folding transition, the results provide general support for the contention presented that specific peptide molecules can interfere and alter the folding kinetics of globular proteins. The proposed therapeutic strategy is depicted in Figure 2.

To summarize, we have presented a target-specific strategy to develop folding-interdicting drugs. Such folding-interdicting drugs would function by specifically inhibiting the earliest folding events. In order to do so, we propose to rely on the a priori capabilities to identify FITRs embedded within the SCM. This strategy is generally similar to earlier proposals to develop therapeutic peptide folding inhibitors relying on protein folding information.[65] Our expectation is that by developing such SCM-based folding-interdicting drugs as

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX

proposed here, a new avenue to alleviate the consequences of disease might be open to further research and development. Whether folding interdiction through inhibition of just the dominant folding initiation event suffices to preempt the onset of disease is a matter that calls for experimental assessment.

5. CONCLUSIONS

In this article we presented theoretical predictions for the folding initiation events of two functionally relevant proteins of the SARS-CoV-2 virus. The predicted folding initiation events were shown to map into good contacts on the 3D structure of the proteins. We proposed that knowledge of the protein segments involved in the folding initiation event, opens an attractive route to developing new therapeutic drugs intended to prevent the successful folding of key viral proteins. Such reduction of the population of properly folded viral proteins could lead to a decreased level of viral spread, thus reducing the lethality of the infection.

On the basis of these findings, we proposed a general therapeutic drug development strategy based on interdicting the folding process through the identification of SCM-predicted FITRs. The FITR strategy could be generally used to approach other diseases where specific proteins play an essential role.

AUTHOR INFORMATION

Corresponding Author

Fernando Bergasa-Caceres − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0002-3683-2883; Email: Bergasa@Princeton.edu, f_bergasa@yahoo.es

Author

Herschel A. Rabitz − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0002-4433-6142

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c03716

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge support from NSF grant CHE-1763198.

REFERENCES

(1) Zhou, P.; Yang, X.-L.; Wang, X.-G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.-R.; Zhu, Y.; Li, B.; Huang, C.-L.; et al. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature* **2020**, *579*, 270−273.

(2) Wu, F.; Zhao, S.; Yu, b.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579*, 265−269.

(3) Dong, E.; Du, H.; Gardner, L. An Interactive Web-based Dashboard to Track COVID-19 in Real Time. *Lancet Infect. Dis.* **2020**, *20*, 533−534.

(4) Press, O. W.; Appelbaum, F.; Martin, P. J.; Matthews, D. C.; Bernstein, I. D.; Eary, J. F.; Nelp, W. B.; Gooley, T.; Glenn, S.; Porter, B.; et al. Phase II Trial of 1311-B1 Antibody Therapy with Autologous Stem Cell and Transplantation for Relapsed B Cell Lymphomas. *Lancet* **1995**, *346*, 336−340.

(5) Goldenberg, D. M.; DeLand, F.; Kim, E.; Bennett, S.; Primus, F. J.; van Nagell, J. R., Jr.; Estes, N.; DeSimone, P.; Rayburn, P. Use of Radiolabeled Antibodies to Carcinoembryonic Antigen for the Detection and Localization of Diverse Cancers by External Photoscanning. *N. Engl. J. Med.* **1978**, *298*, 1384−1388.

(6) Elshabrawy, H. A.; Coughlin, M. M.; Baker, S. C.; Prabhakar, B. S. Human Monoclonal Antibodies against Highly Conserved HR1 and HR2 Domains of the SARS-CoV Spike Protein Are More Broadly Neutralizing. *PLoS One* **2012**, *7*, No. e50366.

(7) 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.

(8) 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.

(9) Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; et al. Analysis of Therapeutic Targets for SARS-CoV-2 and Discovery of Potential Drugs by Computational Methods. *Acta Pharm. Sin. B* **2020**, *10*, 766−788.

(10) Lei, J.; Kusov, Y.; Hilgenfeld, R. Nsp3 of Coronaviruses: Structures and Functions of a Large Multi-domain Protein. *Antiviral Res.* **2018**, *149*, 58−74.

(11) Kim, Y.; Jedrzejczak, R.; Maltseva, N.; Endres, M.; Mececar, A.; Michalska, K.; Jochimiak, A. Crystal Structure of ADP Ribose Phosphatase of NSP3 from SARS CoV-2, **2020**. DOI: 10.2210/pdb6VXS/pdb

(12) 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.

(13) 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.

(14) Bergasa-Caceres, F.; Ronneberg, T. A.; Rabitz, H. A. Sequential Collapse Model for Protein Folding Pathways. *J. Phys. Chem. B* **1999**, *103*, 9749−9758.

(15) Bergasa-Caceres, F.; Haas, E.; Rabitz, H. A. Nature's Shortcut to Protein Folding. *J. Phys. Chem. B* **2019**, *123*, 4463−4476.

(16) Bergasa-Caceres, F.; Rabitz, H. A. Macromolecular Crowding Facilitates the Conformational Transition of Molten Globule States of the Prion Protein. *J. Phys. Chem. B* **2016**, *120*, 11093−11101.

(17) Bergasa-Caceres, F.; Rabitz, H. A. Predicting the Location of the Non-local Contacts in Alpha-synuclein Biochimica et Biophysica Acta Proteins and Proteomics. *Biochim. Biophys. Acta, Proteins Proteomics* **2018**, *1866*, 1201−1208.

(18) Lee, A. C.-L.; Harris, J. L.; Khanna, K. K.; Hong, J.-H. A Comprehensive Review of Current Advances in Peptide Drug Development and Design. *Int. J. Mol. Sci.* **2019**, *20*, 2383.

(19) Motta, A.; Reches, M.; Pappalardo, L.; Andreotti, G.; Gazit, E. The Preferred Conformation of the Tripeptide Ala-Phe-Ala in Water is an Inverse γ-Turn: Implications for Protein Folding and Drug Design. *Biochemistry* **2005**, *44*, 14170−14178.

(20) Broglia, R. A.; Serrano, L.; Tiana, G. *Protein Folding and Drug Design*; IOS Press, 2007.

(21) 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.; Hipps, K. W.; et al. Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.

(22) Conrad, J. C.; Flory, P. J. Moments and Distribution Functions for Polypeptide Chains. Poly-L-alanine. *Macromolecules* **1976**, *9*, 41−47.

(23) Jacobson, H.; Stockmayer, W. H. Intramolecular Reactions in Polycondensations. I. The Theory of Linear Systems. *J. Chem. Phys.* **1950**, *18*, 1600−1606.

(24) Plotkin, S. S.; Onuchic, J. N. Structural and Energetic Heterogeneity in Protein Folding. I. Theory. *J. Chem. Phys.* **2002**, *116*, 5263−5283.

(25) Karplus, M.; Weaver, D. L. Diffusion−collision Model for Protein Folding. *Biopolymers* **1979**, *18*, 1421−1437.
(26) Šali, A.; Shakhnovich, E.; Karplus, M. How Does a Protein Fold? *Nature* **1994**, *369*, 248−251.
(27) 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.
(28) Alm, E.; Baker, D. Prediction of Protein-Folding Mechanisms from Free-energy Landscapes Derived from Native Structures. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11305−11310.
(29) Munoz, V.; Eaton, W. A. A Simple Model for Calculating the Kinetics of Protein Folding from Three-dimensional Structures. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311−11316.
(30) Clementi, C.; Jennings, P. A.; Onuchic, J. N. How Native-state Topology Affects the Folding of Dihydrofolate Reductase and Interleukin-1ß. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871−5876.
(31) 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.
(32) Berezovsky, I. N.; Trifonov, E. N. Van der Waals Locks: Loop-n-Lock Structure of Globular Proteins. *J. Mol. Biol.* **2001**, *307*, 1419−1426.
(33) 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.
(34) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, *29*, 7133−7155.
(35) 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.
(36) Shiraki, K.; Nishikawa, K.; Goto, Y. Trifluoroethanol-induced Stabilization of the α-Helical Structure of β-Lactoglobulin: Implication for Non-hierarchical Protein Folding. *J. Mol. Biol.* **1995**, *245*, 180−194.
(37) 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.
(38) Ozkan, S. B.; Wu, G. A.; Chodera, J. D.; Dill, K. A. Protein Folding by Zipping and Assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11987−11992.
(39) Neuman, B. W.; Joseph, J. S.; Saikatendu, K. S.; Serrano, P.; Chatterjee, A.; Johnson, M. A.; Liao, L.; Klaus, J. P.; Yates, J. R., III; Wüthrich, K.; et al. Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3. *J. Virol.* **2008**, *82*, 5279−5294.
(40) Neuman, B. W. Bioinformatics and Functional Analyses of Coronavirus Nonstructural Proteins Involved in the Formation of Replicative Organelles. *Antiviral Res.* **2016**, *135*, 97−107.
(41) Gorbalenya, A. E.; Koonin, E. V.; Lai, M. M.-C. Putative Papain-related Thiol Proteases of Positive-strand RNA Viruses. Identification of Rubi- and Aphthovirus Proteases and Delineation of a Novel Conserved Domain Associated with Proteases of Rubi-, Alpha- and Coronaviruses. *FEBS Lett.* **1991**, *288*, 201−205.
(42) Han, W.; Li, X.; Fu, X. The Macro-domain Protein Family: Structure, Functions and their Potential Therapeutic Implications. *Mutat. Res.* **2011**, *727*, 86−103.
(43) Saikatendu, K. S.; et al. Structural Basis of Severe Acute Respiratory Syndrome Coronavirus ADP-ribose-1"-phosphate Dephosphorylation by a Conserved Domain of nsP3. *Structure* **2005**, *13*, 1665−1675.
(44) Egloff, M.-P.; Malet, H.; Putics, A.; Heinonen, M.; Dutartre, H.; Frangeul, A.; Gruez, A.; Campanacci, V.; Cambillau, C.; Ziebuhr, J.; et al. Structural and Functional Basis for ADP-ribose and Poly(ADP-ribose) Binding by Viral Macro Domains. *J. Virol.* **2006**, *80*, 8493−8502.
(45) Xu, Y.; Cong, L.; Chen, C.; Wei, L.; Zhao, Q.; Xu, X.; Ma, Y.; Bartlam, M.; Rao, Z. Crystal Structures of Two Coronavirus ADP-ribose-1"-monophosphatases and Their Complexes with ADP-ribose: a Systematic Structural Analysis of the Viral ADRP Domain. *J. Virol.* **2009**, *83*, 1083−1092.
(46) Putics, A.; Filipowicz, W.; Hall, J.; Gorbalenya, A. E.; Ziebuhr, J. ADP-ribose-1"-monophosphatase: a Conserved Coronavirus Enzyme that is Dispensable for Viral Replication in Tissue Culture. *J. Virol.* **2005**, *79*, 12721−12731.
(47) Fehr, A. R.; Channappanavar, R.; Jankevicius, G.; Fett, C.; Zhao, J.; Athmer, J.; Meyerholz, D. K.; Ahel, I.; Perlman, S. The Conserved Coronavirus Macrodomain Promotes Virulence and Suppresses the Innate Immune Response During Severe Acute Respiratory Syndrome Coronavirus Infection. *mBio* **2016**, *7*, No. e01721-16.
(48) Eriksson, K. K.; Cervantes-Barragán, L.; Ludewig, B.; Thiel, V. Mouse Hepatitis Virus Liver Pathology is Dependent on ADP-ribose-1"-phosphatase, a Viral Function Conserved in the Alpha-like Supergroup. *J. Virol.* **2008**, *82*, 12325−12334.
(49) 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 Bioinf.* **2005**, *6*, 21.
(50) Bergasa-Caceres, F.; Rabitz, H. A. Sequential Collapse Folding Pathway of β-Lactoglobulin: Parallel Pathways and Non-Native Secondary Structure. *J. Phys. Chem. B* **2003**, *107*, 3606−3612.
(51) Bergasa-Caceres, F.; Rabitz, H. A. Role of Topology in the Cooperative Collapse of the Protein Core in the Sequential Collapse Model. Folding Pathway of r-Lactalbumin and Hen Lysoz. *J. Phys. Chem. B* **2001**, *105*, 2874−2880.
(52) Bergasa-Caceres, F.; Rabitz, H. A. Macromolecular Crowding Facilitates the Conformational Transition of Molten Globule States of the Prion Protein. *J. Phys. Chem. B* **2016**, *120*, 11093−11101.
(53) Li, W.; Moore, M. J.; Vasilieva, N.; Sui, J.; 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.
(54) Li, F. Receptor Recognition Mechanisms of Coronaviruses: A Decade of Structural Studies. *J. Virol.* **2015**, *89*, 1954−1964.
(55) Li, W.; 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.
(56) Li, F.; Li, W. H.; Farzan, M.; Harrison, S. C. Structure of SARS Coronavirus Spike Receptor Binding Domain Complexed with Receptor. *Science* **2005**, *309*, 1864−1868.
(57) 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.
(58) 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.
(59) Han, D. P.; Lohani, M.; Cho, M. W. Specific Asparagine-linked Glycosylation Sites Are Critical for DC-SIGN and L-SIGN-mediated Severe Acute Respiratory Syndrome Coronavirus Entry. *J. Virol.* **2007**, *81*, 12029−12039.
(60) Yang, N. J.; Hinner, M. J. Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins. *Methods Mol. Biol* **2015**, *1266*, 29.
(61) Wells, J. A.; McClendon, C. L. Reaching for High-hanging Fruit in Drug Discovery at Protein-Protein Interfaces. *Nature* **2007**, *450*, 1001−1009.
(62) Ruzza, P.; Siligardi, G.; Hussain, R.; Marchiani, A.; Islami, M.; Bubacco, L.; Delogu, G.; Fabbri, D.; Dettori, M. A.; Sechi, M.; et al. Ceftriaxone Blocks the Polymerization of α-Synuclein and Exerts Neuroprotective Effects in Vitro. *ACS Chem. Neurosci.* **2014**, *5*, 30−38.

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(63) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.-Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. α-Synuclein in Lewy Bodies. *Nature* **1997**, *388*, 839−840.
(64) Das, A.; Yadav, A.; Gupta, M.; Purushotham, R.; Terse, V. L.; Vishvakarma, V.; Singh, S.; Nandi, T.; Mandal, K. et al. Rational Design of Protein-specific Folding Modifiers. **2020**, bioRxiv:2020.04.28.064113.
(65) Broglia, R. A.; Tiana, G.; Sutto, L.; Provasi, D.; Simona, F. Design of HIV-1-PR Inhibitors that do not Create Resistance: Blocking the Folding of Single Monomers. *Prot. Sci.* **2005**, *14*, 2668−2681.

https://dx.doi.org/10.1021/acs.jpcb.0c03716
J. Phys. Chem. B XXXX, XXX, XXX−XXX