

Flawed methods in “COVID-19: Attacks the 1-Beta Chain of Hemoglobin and Captures the Porphyrin to Inhibit Human Heme Metabolism”

Randy J Read, Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, The Keith Peters Building, Hills Road, Cambridge CB2 0XY, UK.

Abstract

The preprint from Liu & Li (<https://doi.org/10.26434/chemrxiv.11938173.v7>) puts forward hypotheses about a proposed role for proteins of SARS-CoV-2, the virus associated with Covid-19, in directly attacking haemoglobin in patients' blood. Arguments for the hypotheses are based on computational methods: bioinformatics calculations searching for evidence that viral proteins share functional domains related to haem binding with human proteins, molecular modeling of viral proteins, and computational docking of these protein models with models of haem, porphyrin and haemoglobin. No experimental evidence is provided to support any of the conclusions. When interpreted according to accepted standards, these computational results do not hold up and do not provide support for the hypotheses. The interpretation of the search for shared functional domains suffers from a fundamental error in how the significance of the results is judged; when interpreted correctly, there is no evidence for these shared functional domains. Molecular modeling is carried out with tools that are easy to use but not best-in-class, and no allowance is made for uncertainty in the resulting atomic coordinates. Finally, the docking results are invalidated by a catastrophic error in their interpretation: the authors choose the docking trials that have the highest energies, whereas the most stable complexes are actually the ones that have the lowest energies and are therefore least strained. An addendum addresses flaws in a new version 8 from Liu & Li (<https://doi.org/10.26434/chemrxiv.11938173.v8>), which retracts most of their results from earlier versions but nonetheless continues to put forward the same conclusions on the basis of poorly-controlled docking calculations.

Introduction

The present international crisis triggered by the Covid-19 pandemic has stimulated rapid action by the scientific community, with many results being reported in the form of preprints to avoid any delays caused by the peer review and publication processes. Where the results provide real insight, this can help a great deal to accelerate progress in understanding this disease. However, conclusions from such preprints must be treated very cautiously prior to peer review.

Here we consider a preprint submitted to ChemRxiv by (Liu & Li, 2020), which makes strong claims for biological and medical significance. It has accordingly generated a very high level of interest. At the time of writing, versions of this preprint have been viewed over 1.7 million times, downloaded nearly 150,000 times and linked to over 11,000 tweets. Anecdotally, patient groups for patients of relevant metabolic diseases have become aware of this paper, as well as the clinicians who care for them. It is important therefore to assess the quality of the science and the strength of the evidence for the hypotheses put forward. This turns out, unfortunately, to be very poor.

The paper starts from the clinical observation that haemoglobin counts in affected patients are decreased from normal, with an increase in serum ferritin. The authors propose that this implies some sort of direct interaction of viral proteins with haemoglobin and haem; they do not consider alternative explanations such as side-effects of inflammation or tissue damage. The computational experiments are designed to look for features that might be consistent with a direct interaction, but without assessing significance by statistical analysis or by introducing positive or negative

controls. As shown below, none of the results hold up to close inspection, and it can be concluded that this work does not add any evidence in support of the hypotheses it puts forward.

It should be noted that the process of assessing the science has been assisted by cooperation from the authors, who have answered some questions and have made all their computational results openly available to download. Most of this discussion concerns version 7 of the preprint, with an addendum discussing the changes made in version 8 in response to the criticism raised here.

Sequence motif search results lack statistical significance

The authors used a bioinformatics evaluation of sequence motifs to test the hypothesis that viral proteins might have functional domains associated with haem or porphyrin binding or with haem oxygenase activity. This was carried out using the MEME server (Bailey & Elkan, 1994) (<http://meme-suite.org/tools/meme>) to compare the sequences of representatives of each category of protein (haem-binding or haem oxygenase activity) with different viral proteins in turn, looking for conserved sequences. The authors assert that a p-value significantly below 1 demonstrates the significance of the hits. However, the p-value reported by the MEME server evaluates the probability that particular segments of each of the protein sequences could resemble each other to the extent observed if they were all drawn independently from amino acids reflecting the overall protein composition. An extremely large number of such comparisons are performed, selecting different subsequences of each of the proteins and of different lengths from 6-50 residues, so the p-values need to be corrected for the effects of multiple comparisons. This is evaluated with the E-value, *i.e.* the number of hits of a certain score or greater that would be expected to be found in the whole pool of multiple comparisons. The authors of MEME suggest that an E-value of less than 0.05 is required to indicate significance for a match involving a particular motif. However, the E-values were not reported by Liu & Li and were not recorded in the results they provided. These calculations were therefore repeated for selected cases using the sequence files provided by the authors. In the MEME server, defaults were chosen for all parameters (classic mode, zero or one occurrence per sequence, search for 3 motifs) and the new calculations yielded the same results reported by the authors. The results for the 6 searches yielding the lowest p-values reported by Liu & Li are given in Table 1.

Table 1. Results of search for conserved motifs using the MEME server

Function of target proteins	Viral protein	p-value for hit	Lowest motif E-value
Haem-binding	orf1ab	4.40e-36	0.29
Haem-binding	nucleocapsid	5.74e-31	0.03
Haem-binding	ORF3a	1.41e-19	0.18
Haem-binding	ORF7a	1.10e-16	0.09
Haem oxygenase*	ORF10	1.31e-14	0.003
Haem-binding	surface glycoprotein	9.63e-13	0.43

* In the reported search, two copies of haem oxygenase-2 were included in the input sequences, which will inflate the apparent significance of the hits. The search was repeated including only one copy of haem oxygenase-2 as well as one copy of haem oxygenase-1.

Although these are the searches with the lowest p-values, most of them fail to achieve the target E-value of less than 0.05. Given that there were 20 such motif searches, it would not be surprising for the 0.05 threshold to be passed even with random sequences.

The potential value of motif searches is in identifying proteins that may have common functions because they share similar sequences *in parts of the proteins that have similar structure*. When structures are not known, the existence of similar structures might be inferred from strong hits in the motif searches. Given that a number of these proteins have known structures (some of which have appeared since the preprint was first submitted by Liu & Li), it is possible to assess the validity of the motif searches independently by comparing the structures globally and in the region of the motif alignment. On this basis, both of the hits with E-values below 0.05 are revealed to be false positives. Figure 1 shows PDB entry 3r8j (a structure of human haem-binding protein 2) superimposed on PDB entry 6vyo (a structure of the SARS-CoV-2 nucleocapsid protein), using a least-squares superposition of the top motif identified by MEME. It is clear that the structures are neither locally nor globally similar.



Figure 1. Superposition of chain A of 3r8j (human haem-binding protein-2, light blue) on chain A of 6vyo (SARS-CoV-2 nucleocapsid, orange), using least-squares superposition of the top motif identified using MEME. The RMSD achieved in the superposition is 7.7 Å over 27 alpha-carbon atoms. This rules out the conclusion from motif searches that these proteins have shared function. All figures were prepared with ChimeraX (Goddard *et al.*, 2018).

Although the potential haem oxygenase motif identified for the viral ORF10 protein has an apparently significant E-value of 0.003, it is unlikely to reflect functional significance. First, ORF10 is an extremely small protein of 38 amino acids, much smaller than typical enzymes such as the haem oxygenases (about 300 residues). It could only have some function as part of a larger assembly. Second, the motif that has been identified is not associated with haem binding or the haem oxygenase catalytic mechanism. Figure 2 shows that this motif is on the opposite side of the protein from the catalytic site.

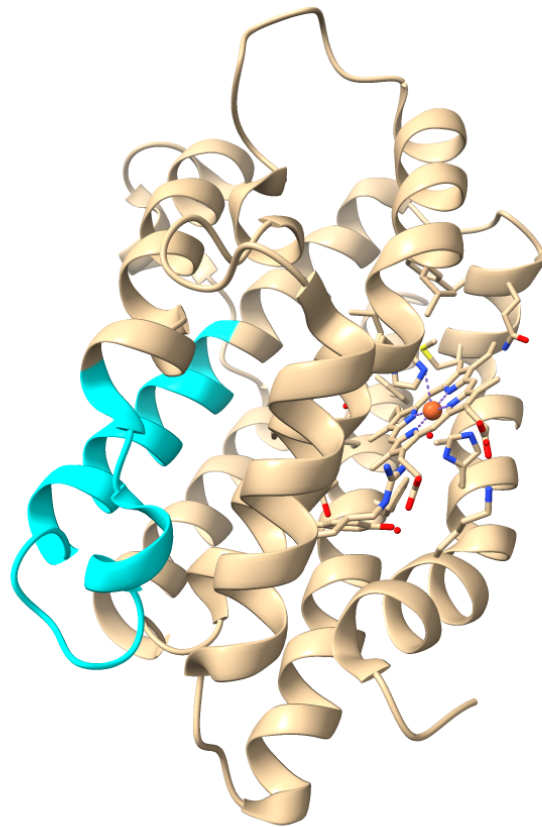


Figure 2. Cartoon view of PDB entry 6eha (human haem oxygenase-1), a target for a search for motifs associated with catalytic action on haem. The haem group bound in the active site on the right is shown in sticks, while the motif identified by MEME is highlighted in cyan on the opposite side of the molecule, where it is clearly not involved in haem binding or catalysis.

Molecular modeling

Apart from two of the viral genes studied, all of the models discussed by Liu & Li were constructed using the online SWISS-MODEL server. SWISS-MODEL is an efficient and convenient server for obtaining template-based (homology) models, yielding good results for reasonably close relatives (Waterhouse *et al.*, 2018). However, other servers perform better on distant homologues (Waterhouse *et al.*, 2018), and it is not an appropriate tool for free (*ab initio*) modeling of the structures of proteins that lack any identifiable homologues of known structure. In fact, 4 of the 10 proteins modeled by the authors, using SWISS-MODEL, have been identified by the CASP organisers as proteins lacking significant homology in proteins of known structure: ORF3a, ORF6, ORF8 and the membrane glycoprotein (<http://predictioncenter.org/caspcommons/targetlist.cgi?view=round2>). An additional protein identified by the CASP organisers, ORF10, was modeled by the authors using Discovery Studio, but based on an alignment that was only obtained by doing a targeted search for similar sequence motifs between ORF10 and haem oxygenases; this alignment is not detected using unbiased searches such as the extremely sensitive searches performed by HHpred (Zimmermann *et al.*, 2018). The SWISS-MODEL authors have carried out their own modeling exercise on all proteins from SARS-CoV-2, with the results summarised at <https://swissmodel.expasy.org/repository/species/2697049>. In this summary, the SWISS-MODEL authors themselves state that the model for ORF3a is expected to be of limited quality and that only models of low quality (not presented, therefore, on the main summary page) could be made for ORF6, ORF8 and the membrane glycoprotein.

Independent blinded tests of different modeling categories have been carried out in the Critical Assessment of Structure Prediction (CASP) set of experiments. The most recent edition of this exercise, CASP13, evaluated methods for a variety of modeling tasks including template-based modeling (Croll *et al.*, 2019) and free modeling (Abriata *et al.*, 2019). Two of the methods that scored very highly in these evaluations have been used to generate models of a number of SARS-CoV-2 proteins: the Robetta server version of Rosetta ([https://www.ssgcid.org/cttdb/molecularmodel list/?organism icontains=COVID-19](https://www.ssgcid.org/cttdb/molecularmodel%20list/?organism%20icontains=COVID-19)) and AlphaFold (<https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>). One of the structures predicted using AlphaFold is ORF3a, so it is of interest to see what a comparison with the SWISS-MODEL prediction of the same protein illustrates about the large uncertainties of modeling such proteins with limited or no homology to proteins of known structure. Liu & Li chose to use model 8 from the SWISS-MODEL prediction for docking studies. A comparison of this model with the one from AlphaFold is shown in Figure 3. SWISS-MODEL has predicted the structure of only a small fragment comprising 30 amino acid residues from position 129 to 158 of the full sequence. For the figure, least-squares has been used to superimpose it on the corresponding portion of the AlphaFold model, yielding an RMSD over alpha-carbon atoms of 6.2 Å, and it is clear that these models bear very little resemblance to each other.

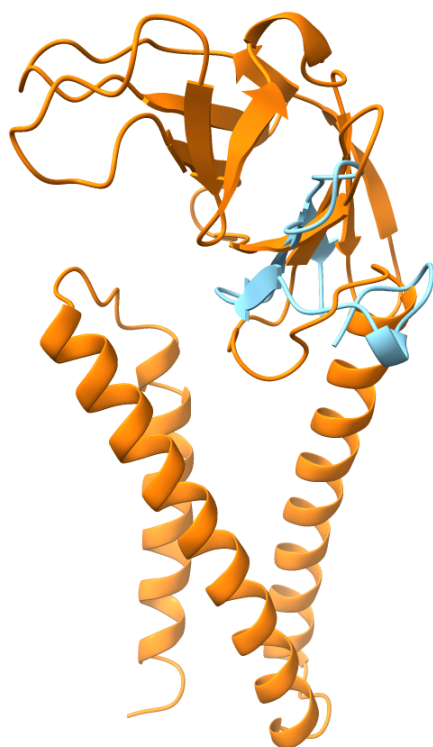


Figure 3. Superposition of the fragmentary SWISS-MODEL model 8 (light blue) of ORF3a on the AlphaFold model of the same protein, shown in orange.

Considering that the energy calculations used for protein-ligand and protein-protein docking are sensitive to shifts of fractions of Å units, any docking results obtained from such models would have to be interpreted very cautiously.

Modeling of orf1ab

The modeling of the gene product of orf1ab is a special case. Although the authors do not appear to take account of this, orf1ab is a polyprotein that is cleaved into a number of smaller proteins by

the actions of viral proteases. From analysis of the genome sequence, there are predicted to be 16 cleavage products (Wu *et al.*, 2020). A BLAST search of the orf1ab sequence looking for matches with PDB entries shows that 11 of these cleavage products have known structures, some as parts of complexes such as PDB entries 5c8t (complex of nsp10 with nsp14 from SARS-CoV) or 6w4h (complex of nsp10 with nsp16 from SARS-CoV-2). The model the authors present for orf1ab covers only the nsp16 component, generated by homology to the nsp16 component from SARS-CoV in PDB entry 5yn5, which is also a complex with nsp10. Prior to determination of the actual SARS-CoV-2 structure (PDB entry 6w4h was released on 18 March 2020), it could have been modeled using SWISS-MODEL, as it appears on the SWISS-MODEL website. The 5000 nucleotide limit that prompted the authors to use Discovery Studio would not apply if they had attempted to model individual proteins or complexes derived from the polyprotein.

Docking calculations

The docking calculations were carried out in Discovery Studio 2015. The result files provided by the authors were opened and inspected in the Discovery Studio Visualizer software (<https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/visualization.html>). Docking calculations are scored by a variety of criteria, of which the binding energy (defined as the total computed energy of the complex minus the sum of the energy computed for the receptor and the energy computed for the ligand) is the one chosen by the authors. For these calculations, the authors used the CHARMM molecular mechanics forcefield. Low-energy states are more stable than high-energy states, so the lowest binding energy scores should reflect the best docking poses. Ideally, favourable docking configurations should show a decrease in energy (negative value) compared to the isolated components. Unexpectedly, the authors have chosen the highest-energy poses instead of the lowest-energy poses as the best docking results. Their conclusions are therefore based on taking the most strained, least favourable docking configurations that the software could generate and treating them as the best results.

For example, the docking of porphyrin into the modelled structure the E2 glycoprotein is discussed as a successful result that is interpreted as showing that this protein is involved in a viral mechanism to attack haemoglobin. For this docking calculation, the binding energy is reported as greater than 7.5 billion kcal/mole, which actually indicates exceptionally high strain. This amount of energy, in fact, is comparable to the energy released by the explosion of 7 kilotons of TNT (https://en.wikipedia.org/wiki/TNT_equivalent), equivalent to more than 30 million moles of TNT. Inspection of the coordinates shows that there are extremely unfavourable clashes in the chosen docking pose, with the closest interatomic contact being 0.31 Å.

Simple docking calculations such as those reported by Liu & Li could in principle be useful in generating hypotheses for binding modes, when a ligand is already known to bind. However, to assess binding affinity on the absolute scale needed to predict whether binding will actually occur, much more sophisticated calculations of the relative free energy of binding must be carried out. These require extremely accurate models of the target, sensitive even to errors in the orientations of side chains (Cournia *et al.*, 2017).

“Validation” of docking by drug docking

The authors present docking results for chloroquine and favipiravir and conclude that these docking experiments validate the haem and porphyrin docking by finding interactions in the same binding pockets. Firstly, the docking is again judged incorrectly by the highest energy returned by the docking program, not the lowest energy. Secondly, even if the docking results yielded plausible

interactions with the same amino acids, this would not constitute independent validation because the docking is constrained to position the ligand in the same pre-selected pocket that was used for the initial docking.

Discussion

In their preprint, Liu & Li describe a set of computational experiments designed to produce positive results but without appropriately assessing significance or considering alternative explanations. The first step is to test the idea that a particular viral protein could share a sequence motif with other proteins that bind or react with haem or porphyrin. These calculations incorrectly use the p-value, not the E-value, to assess significance. When judged appropriately they provide no evidence for the viral proteins possessing the functions the authors wish to show. Even if motifs were identified with high levels of significance, their identity would only imply similar function if they were found within proteins of overall similar structure, but where the structures are known the haem-binding and viral proteins differ dramatically.

After constructing computational models of viral proteins (using tools that are not suited to distant homology modeling), the authors carry out docking studies in an attempt to show that the molecular interactions they propose could take place. However, in this step there is a catastrophic error: the docking poses are evaluated on the basis of which one is associated with the highest energy. Since the most stable molecules and complexes are the ones with lowest energy, this is precisely the wrong way around.

Finally, the authors build up a series of speculations about viral evolution, viral biology, disease pathogenesis and potential therapies. They propose, for instance, that viral proteins bind porphyrin, that the bound porphyrin collects energy from light (although how light penetrates into the lungs of patients is not addressed), this energy is then transferred by the process of fluorescence resonance energy transfer (FRET) and used for processes such as waking the virus from hibernation. More worryingly, they assert that their computational results provide evidence supporting the therapeutic effect of drugs that are only just entering serious clinical trials.

Addendum

The first version of this critique was submitted on 13 April 2020 and posted the next day. Liu & Li submitted a revised version of their preprint on 25 April 2020, only 11 days after this critique appeared. In their revised work, they seem to accept most of the criticisms made by the first version of this critique, yet they continue to put forward the same hypotheses about the virus attacking haemoglobin.

The relevant observations from their revision are as follows:

- They accept that there is no evidence that SARS-CoV-2 proteins possess conserved functional domains from proteins that bind or alter haem or porphyrins. This was a keystone of their original argument.
- They accept that the protein structure modeling techniques used in the original work are not the best available, and have changed to some of the newer models from the Robetta server and the AlphaFold team, referred to above.
- They accept that they had completely misinterpreted the nature of binding energies, believing incorrectly that higher energies were better and thereby choosing the very worst possible binding modes found in any of the docking calculations. The new version invokes completely new binding modes with the lowest energies found.

- They accept that it was not appropriate to use constrained docking of drug candidates to the same pockets as validating evidence that these drugs interfere with binding by viral proteins, and have removed this key section from their preprint.

In effect, they have abandoned all of the key results that they used to argue for their hypotheses. This suggests that even Liu and Li accept the conclusion that they lacked basic competence in the tools they were using. Under these circumstances, many researchers would choose to retract their work to avoid misleading others.

Given that they have not chosen to retract their work, it is fair to ask: is less than two weeks sufficient both to gain and to apply the expertise in bioinformatics and molecular simulations required to put forward paradigm-changing ideas?

Examining the results of the new simulations suggests that the answer to this question is no. The new work suffers from the same problems with application of the scientific method. The presence of docking poses in the results of docking calculations is taken as proof that binding takes place in the real world, but the authors do not acknowledge that docking calculations almost always yield docking poses whether they are correct or not. There is no attempt to disprove the hypotheses. No control experiments are carried out to test whether other random molecules might yield docking poses with similar or even better scores. Finally, the quality of the protein target models and the sophistication of the docking calculations are still insufficient to assess binding affinity on an absolute scale. The revision has not succeeded in providing any valid evidence for the hypotheses advanced.

Acknowledgements

Comments from Airlie McCoy and Tristan Croll are gratefully acknowledged. RJR is supported by a Principal Research Fellowship from the Wellcome Trust UK (grant 209407/Z/17/Z).

References

- Abriata, L. A., Tamò, G. E. & Dal Peraro, M. (2019). *Proteins Struct. Funct. Bioinforma.* **87**, 1100–1112.
- Bailey, T. L. & Elkan, C. (1994). *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36.
- Cournia, Z., Allen, B. & Sherman, W. (2017). *J. Chem. Inf. Model.*
- Croll, T. I., Sammito, M. D., Kryshchuk, A. & Read, R. J. (2019). *Proteins Struct. Funct. Bioinforma.* **87**, 1113–1127.
- Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. (2018). *Protein Sci.* **27**, 14–25.
- Liu, W. & Li, H. (2020). <https://doi.org/10.26434/chemrxiv.11938173>.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. (2018). *Nucleic Acids Res.* **46**, W296–W303.
- 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., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C. & Zhang, Y.-Z. (2020). *Nature.* **579**, 265–269.
- Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N. & Alva, V. (2018). *J. Mol. Biol.* **430**, 2237–2243.