2019-nCoV vs. SARS-CoV: Which Truly Has a Higher ACE2 Affinity? A Quantum Chemical Perspective on Virus-Receptor Noncovalent Interactions

by Nitai Sylvetsky[*,§]

Email: nitai.sylvetsky@weizmann.ac.il

[§] Department of Organic Chemistry, Weizmann Institute of Science, 76100 Reḥovot, Israel

ABSTRACT While the title question is still a matter for ongoing debate, current hypotheses suggest that a few particular amino acid substitutions are responsible for the ACE2 affinity differences between the two viruses (the spike proteins of which exhibit 76% sequence identity, and are therefore assumed to adopt similar folded structures). In the present paper, noncovalent interaction energetics associated with said substitutions are assessed by means of B3LYP/TZVP electronic structure calculations on a representative set of geometries – chosen to reflect different biochemically-significant spatial organizations of the amino acid residues at hand. We found that the sum and (unbiased-)average of calculated interaction energetics under consideration are larger in SARS-CoV compared to 2019-nCoV. Thus, we hereby challenge earlier predictions claiming a higher ACE2 affinity for 2019-nCoV employing "chemical-intuition"-based analyses of said substitutions alone. We also demonstrate that the latter predictions are potentially somewhat-biased for being based on a SARS-CoV–ACE2 crystal structure – which should not be expected to represent a bound spike-protein–receptor complex in physiological conditions. By comparing electronic-structure-based results with ones obtained using the widely-used MMFF94 molecular mechanics force-field, we show that despite being specifically parametrized for van der Waals interactions – such classical force fields might prove inadequate in cases where only several, well-defined noncovalent binding factors are assumed to play crucial roles in biochemically-significant binding events.

Introduction

2019-nCoV is a newly emerged, human-infectious coronavirus (CoV) that has recently spread to become a global epidemiological problem. As of 12.03.20, there have been more than 134,509 diagnosed cases and more than 4,981 confirmed deaths worldwide (ABC news, USA). Due to the fact that the pathogenesis of this virus has not been thoroughly investigated, healthcare professionals are forced to rely on *ad hoc* practical treatment options. A long-term, economically-sustainable battle against 2019-nCoV, however, would require the development of effective medicinal measures (e.g., drugs, vaccines and the like).

One important starting point towards such development is the virus' genome sequence, which is currently available (GenBank ID: MN908947.3).[1] It has been realized that 2019-nCoV shares 82 % sequence identity with severe acute respiratory syndrome-related coronavirus (SARS-CoV). Thus, past findings from medicinal chemistry studies on SARS-CoV may directly be used in current research efforts on 2019-nCoV.

It is well-known that CoV rely on their spike proteins for binding to a host cell surface receptor – which, in turn, leads to cell invasion. As for SARS-CoV, the particular receptor to which 2019-nCoV binds is angiotensin-converting enzyme 2 (ACE2).[2] Since the spike proteins for the two viruses exhibit no less than 76% sequence identity, they are broadly assumed to adopt similar folded structures. Again, in SARS-CoV, a receptor binding domain – responsible for high-affinity interactions with ACE2 – has already been identified in region S1 of the spike protein (see Ref.[2] for the appropriate nomenclature).

The current hypothesis is that such receptor binding domain is also employed by 2019-nCoV.[3] That being said, non-conserved mutations (that is, deviations from the known SARS-CoV amino-acid sequence in these regions) were recognized in structural regions that interact directly with ACE2.[4] Thus, it was suggested that some particular amino acid substitutions are to be held responsible for overall cell-affinity differences between the two viruses. Needless to say, identifying particular amino acids that significantly contribute to ACE2-binding in 2019-nCoV is expected to lead to the development of effective vaccines.[5,6]

Based mostly on "chemical intuition", Morse *et. al.* have predicted that several amino-acid substitutions would result in a greater affinity for 2019-nCoV, due to stronger noncovalent interactions with specific, corresponding residues in ACE2.[3] In this context, the following substitutions were addressed (SARS-CoV/2019-nCoV): V404/K417, Y484/Q498, L472/F486, and C474/Q492. However, reliable assessments of the relevant noncovalent

interactions – that may rise from subtle electronic effects which must be explicitly considered – are to be obtained in order to confirm/refute predictions of this sort.

In this context, electronic structure methods constitute a precious (and almost exclusive) source of information on biochemically-relevant noncovalent interactions between individual molecules – which, in turn, often used for the parametrization and calibration of more approximate computational modeling techniques (such as classical molecular mechanics force fields).[7–11] Having been involved in a few recent quantum chemical investigations of noncovalent complexes,[12–16] we shall now attempt to employ B3LYP/TZVP calculations – which have been shown to be capable of providing reliable noncovalent interaction energetics[17] – for the purpose of calculating energy differences associated with the aforementioned substitutions. In such manner, reliable energetic "gains/losses" which result from these particular substitutions may be quantitatively assessed, so that important practical conclusions regarding the above prediction may be drawn. This general strategy is concisely summarized in Figure 1.
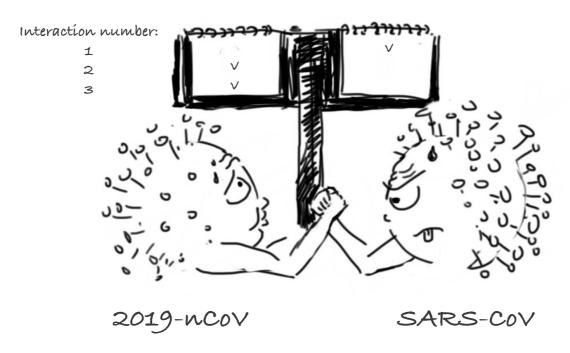

Figure 1. An illustration of the basic analysis strategy employed in this paper; relevant noncovalent interaction energetics will be compared "across" the two viruses, such that their contribution to the overall virus-ACE2 affinity may be assessed.

It should clearly be considered that binding regions in both spike protein and ACE2 may exhibit substantial conformational flexibility in physiological conditions (see, for instance, a

representative discussions on protein flexibility in Refs.[18,19]). Thus, crystal structure geometries should not be considered as an ultimate source of structural information on the systems at hand (this fact seems to have been overlooked in recent research efforts). We will therefore mostly restrict our discussion to particular spatial organizations of the relevant amino acid dimers/clusters, chosen to represent "upper bounds" for said interactions. In such manner, we would be able to use statistical inference tools and critically answer whether said substitutions actually lead to a substantial, potential increase in interaction energies – or, perhaps, suggest that their implications are not quite as straightforward. Still, and in order to compare our results with those of Morse *et. al.*, we will also consider geometries derived directly from the crystal structure in Ref.[2]

Since some recent/contemporary research ventures have involved indirect assessments of the aforementioned noncovalent binding factors by means of classical molecular mechanics force fields,[20–22] we shall also compare our electronic-structure-based energetics with ones obtained using the widely-applicable MMFF94 force-field – which has been specifically parametrized for van der Waals (vdW) interactions[23] and is often used for investigations of organic molecules and polypeptides.[24,25]

Computational Methods

Electronic structure calculations were performed using the Gaussian16[26] software suite and ran on the Faculty of Chemistry HPC cluster at the Weizmann institute of science. The B3LYP functional,[27] which was recommended for biochemically-relevant noncovalent systems in Ref.[17], as well as Grimme's DFT-D3 atom-pairwise dispersion corrections[28] and the TZVP basis set[29] were used throughout.

In addition, MMFF94[23,30] geometry optimizations and energy calculations were performed using the Avogadro software (Macintosh version 1.2).[31]

The sixteen geometries considered in this work were obtained in the following manner: first, four amino-acid dimers, corresponding to the noncovalent interactions between SARS-CoV and ACE2 addressed by Morse *et. al.* (see introduction)*,* were directly extracted from the crystal structure considered in Ref.[2] [PDB ID: 2AJF]. These dimers were then spatially reorganized (based on chemical intuition and bearing in mind that the polypeptide backbone must remain intact) to represent significant binding interactions, and subsequently optimized using the MMFF94 force-field to produce five additional structures. Finally, substitutions

associated with 2019-nCoV were performed and reorganized-optimized in a similar manner – resulting in seven more structures (sixteen in total). For the sake of consistency, we have chosen to carry out MMFF94 geometry optimizations on the original, crystal-structure-derived dimers – despite the fact no noticeable structural changes have seemed to result from such choice. All final geometries used in our interaction energy calculations are provided in .xyz format in the electronic supporting information.

Results and Discussion

Before we begin our analysis, a few notes about nomenclature are in place. As can be seen in Table 1, we labeled all noncovalent interactions under consideration based on the participating residues in ACE2. For instance, the particular interaction involving residues D30 and H34 in ACE2, as well as residue V404 in SARS-CoV, will be denoted 1-SARS-n (n denotes a particular molecular structure representing the latter interaction; specifically, 1-SARS-a corresponds to the SARS-CoV–ACE2 crystal structure from Ref.[2]). Note that for the sake of good order, structure illustrations and names are also presented in Figure 2.

Table 1. Shorthand notation used for interactions considered in this work (see also Figure 2).

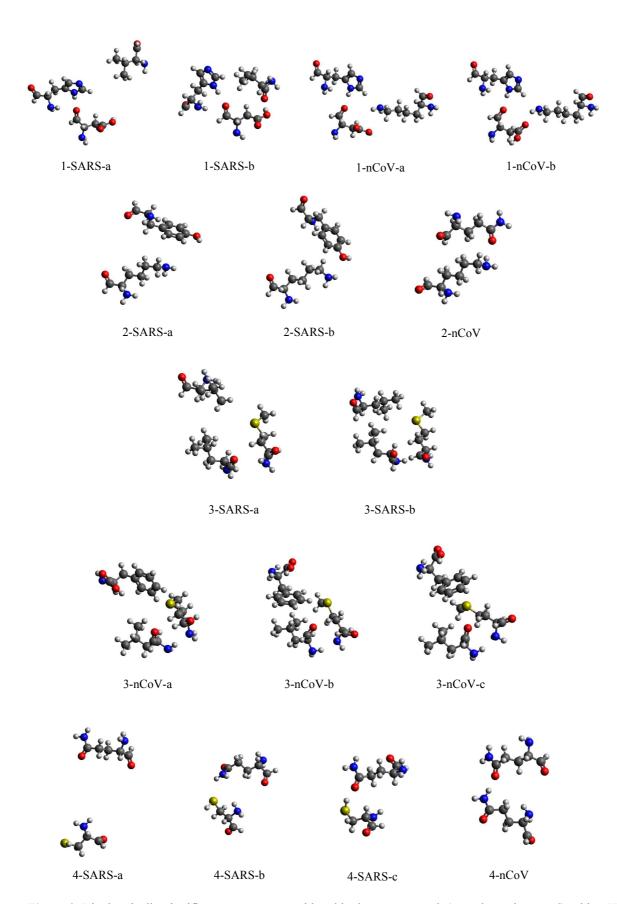| Interaction number (n)	ACE2	SARS/nCoV	Notes
1	D30, H34	V404/K417	*a,b,c* are used to denote different (chemically-significant) structures, which differ in their noncovalent nature (see Table 2).
2	K353	Y484/Q498	
3	L79, M82	L472/F486	n-SARS-a (n=1-4) correspond to interactions derived from the SARS-ACE2 crystal structure in Ref.[2]
4	Q24	C474/Q492	

Figure 2. Biochemically-significant structures considered in the present work (atom legend: gray=C, white=H, Red=O, Blue=N, Yellow=S).

Our calculated B3LYP/def2-TZVP noncovalent interaction energetics, as well as their deviation from ones calculated using the classical MMFF94 force-field, are presented in Table 2.

Table 2. Calculated B3LYP/def2-TZVP interaction energetics, and their deviations from those calculated using the MMFF94 force-field. The noncovalent character of each interaction is also indicated (second column from the left). Statistical dispersion measures for virus-specific interactions are given in the bottom three rows (s=sample standard deviation, MAD=mean absolute deviation); SARS-CoV* denotes the removal of n-SARS-a structures (n=1-4) from the statistical sample.

	Character (in order of significance)	IE (kcal/mol)	dMMFF94
1-SARS-a	vdW (weak)	-0.632	0.055
1-SARS-b	1HB, vdW	-13.723	1.013
1-nCoV-a	vdW	-1.587	-1.944
1-nCoV-b	1HB	-3.712	-0.328
2-SARS-a	vdW	-1.210	-0.143
2-SARS-b	vdW	-3.215	-1.647
2-nCoV	1HB, vdW	-6.389	-1.702
3-SARS-a	vdW, hydrophobic	-3.217	-1.060
3-SARS-b	vdW, hydrophobic	-3.229	-0.695
3-nCoV-a	vdW(S-π), hydrophobic	-2.879	-0.396
3-nCoV-b	Hydrophobic, vdW(S-π)	-7.510	-2.831
3-nCoV-c	vdW(S-π), hydrophobic	-3.744	-3.624
4-SARS-a	vdW (very weak)	0.016	-0.116
4-SARS-b	1HB (S=HB donor)	-5.664	-1.003
4-SARS-c	1HB (S=HB acceptor)	-2.116	-0.502
4-nCoV	1HB, vdW	-6.660	1.712

	Range	s	MAD
SARS-CoV	13.739	4.137	2.679
SARS-CoV*	11.607	4.729	3.283
2019-nCoV	6.676	2.216	1.897

First, let us compare our DFT results to those obtained using the MMFF94 force-field. Not quite surprisingly, exceptionally large deviations are observed for structures reflecting significant vdW contributions (note, in particular, those associated with the S-π stabilized 3-nCoV-b/c). Generally-smaller deviations are accordingly found for hydrophobic or "pure" hydrogen bonding interactions (which do not involve additional significant vdW contributions). It is important to note that for some structures, said deviations are similar in magnitude to the calculated interaction strength; this should indeed indicate that classical force fields of this sort might prove problematic in cases where just a few, well-defined

noncovalent binding factors are considered. In the present such case, MMFF94 clearly fails to provide accurate energetics for many structures, and should thus not be relied upon for answering whether the particular interactions under consideration indicate higher ACE2 affinity for either SARS-CoV or 2019-nCoV. Electronic structure calculations capable of accurately describing noncovalent binding factors should therefore be used as a reliable source of information for such purposes.

We can now move on and focus on our B3LYP/def2-TZVP results. Excluding interaction 1, for which SARS-CoV exhibits the largest interaction energy (1-SARS-b), 2019-nCoV indeed seems to be capable of participating in stronger noncovalent interactions (2-nCoV, 3-nCoV-b, and 4-nCoV correspond to the strongest interactions for n=2-4). As somewhat briefly addressed by Morse *et. al.*, most energetic "gains" associated with 2019-nCoV residues may be rationalized using chemical intuition. For instance, in interactions 2 and 4, 2019-nCoV clearly relies on hydrogen bonding while SARS-CoV is only capable of participating in either van der Waals (vdW), or sulfur-containing hydrogen bonding interactions – both can be expected to be weaker. In addition, 2019-nCoV apparently benefits from S-π stabilization in interaction 2, which can clearly turn significant in an appropriate spatial organization (3-nCoV-b). That being said, and contrary to the prediction by Morse *et. al.*, 1-SARS-b is clearly capable of benefitting from a combination of vdW and hydrogen bond stabilization, which do not exist in the 1-nCoV structures. As shown in Figure 2, this latter finding may, in fact, be attributed to the spatial properties of residue V404 in SARS-CoV, which enable it to interact significantly with both D30 and H34 in ACE2 (compare with residue K417 in 2019-nCoV; see structures 1-nCoV-a/b).

An important thing to note here, however, is that the largest calculated energies for interactions n=1-4 should not, on their own, be simply considered as a measure for the overall interaction strength between SARS-CoV/2019-nCoV and ACE2. As mentioned in the introduction, such macro-scaled protein-protein interactions are driven by a great many noncovalent binding factors, and are characterized with significant conformational flexibility. Thus, one should not expect all possible noncovalent binding factors to be "fully utilized" – as the maximization of one factor might lead to geometric constraints which inhibit maximizations of others. Bearing this in mind, we shall accordingly consider the entire calculated interaction space (as opposed to the aforementioned, maximum-interaction-strength data points in isolation) in the remainder of the present discussion.

A few representative statistical dispersion measures describing the interaction space at hand are also presented in the lower part of Table 2. It can clearly be seen that the range of interaction energetics (from smallest to largest) in SARS-CoV is about as twice as large as that of 2019-nCoV. In addition, both standard deviation and mean absolute deviation are significantly larger in SARS-CoV compared to 2019-nCoV (by ~87% and ~41%, respectively). Indeed, it can easily be inferred that based on the four noncovalent interactions considered in this work, SARS-CoV is more dependent on specific energetically-favored spatial organizations (i.e., n-SARS-b, n=1-4) for effective binding. In particular, interactions 1 and 4 may be identified as potentially-significant binding factors for SARS-CoV, as their potential magnitudes are larger than those of 2 and 3. 2019-nCoV, on the other hand, exhibits rather balanced dependence on the four considered interactions (though interaction 1 is indeed a bit weaker than its counterparts).

Again, based on the present interaction space, and considering that both receptor binding domains in the virus' spike protein and their corresponding binding sites in ACE2 may exhibit significant conformational flexibility, we may now come up with a few possible interaction "scenarios" for the two viruses – which will, in turn, help us establish important practical conclusions regarding the latter's overall ACE2 affinities (Table 3). First, if we assume that all interactions may indeed reach their maximum value for both viruses (scenario I), then SARS-CoV can be expected to bind to ACE2 with a slightly greater (~6%) overall affinity; however, and since this scenario should clearly not be perceived as having exclusive biochemical significance – additional scenarios must obviously be considered as well. In scenario II, minimum values are considered for all investigated interactions. 2019-nCoV, in this case, is clearly the "winner" – as the sum of its interactions amounts to ~350% of the corresponding sum for SARS-CoV. That being said, it should be stressed that interactions derived from the crystal structure of SARS-CoV–ACE2 (that is, n-SARS-a) – which can be expected to represent lower bounds for interactions 1, 2 and 4 – are likely to bias this scenario in favor of 2019-nCoV. Still, and since the statistical dispersion measures are larger for SARS-CoV even in case where crystal structure geometries are not considered (see Table 2) – it is possible to expect 2019-nCoV to bind to ACE2 with higher affinity in the current scenario. If we consider the average value of all investigated interactions for each virus as a measure for overall ACE2 affinity (scenario III), then 2019-nCoV is shown to have the upper hand (larger by ~26% than the average interaction for SARS-CoV). Similarly, summing the average value for each individual interaction (scenario IV) gives 2019-nCoV an advantage of

~34%. However, removing the aforementioned, somewhat-biased crystal-structure-derived energetics from our averaging (V and VI) – SARS-CoV clearly exhibits superior statistics (values are larger than those for 2019-nCoV by ~20% and ~18%, respectively).

Table 3. Six interaction scenarios based upon our calculated energetics (Table 2). ΣMax_n = the sum of maximum calculated values for each interaction (n=1-4); ΣMin_n = the sum of minimum such values; $Average_{tot}$ = the average of all calculated interaction energetics for a given virus; $\Sigma Average_n$ = the sum of average values for each interaction. As in Table 2, * denotes the removal of n-SARS-a structures (n=1-4) from the statistical sample.

Scenario	I	II	III	IV	V	VI
Description:	ΣMax_n	ΣMin_n	$Average_{tot}$	$\Sigma Average_n$	$Average_{tot}$*	$\Sigma Average_n$*
SARS-CoV	**-25.830**	-5.043	-3.665	-15.200	**-5.589**	**-24.056**
2019-nCoV	-24.270	**-17.515**	**-4.640**	**-20.409**	-4.640	-20.409

The above findings lead us to suspect that previous comparisons of individual noncovalent binding factors such as those carried out by Morse *et. al.* are somewhat biased towards the crystal structure of SARS-CoV–ACE2 – which, as said, does not trivially represent binding interactions in physiological conditions. Indeed, crystal structures constitute an important source of information on biochemically-significant structures and corresponding interactions; the information derived from them, however, should be critically evaluated and not be taken "too literally". Bearing this in mind, we suspect that the interactions considered in this work favor SARS-CoV over 2019-nCoV in terms of their overall contribution to the virus' ACE2 affinity. Nevertheless, additional crucial binding factors should certainly be considered for the purpose of arriving at unambiguous, practically-useful predictions of such affinities.

Before ending this discussion, we would like to clarify that we do not at all wish to imply that realistic ACE2 affinities can be trivially reduced (in the statistical sense) to the four noncovalent interactions considered above. What we do mean to demonstrate is that (a) hypotheses regarding implications of particular amino acid substitutions on overall receptor affinity must be confirmed/refuted by means of appropriate information on relevant noncovalent binding factors, and that (b) in the absence of such information, classical molecular dynamics simulations, as well as insights drawn from "chemical intuition", are prone to lead to erroneous conclusions. Alas, claims according to which 2019-nCoV has, in fact, a higher affinity towards ACE2 compared to SARS-CoV should not be based upon shots in the dark and questionable expectations.

Summary and Conclusions

Based on our analysis of the noncovalent interaction space under consideration, we were able to draw the following conclusions:

- The sum of the strongest calculated noncovalent interactions is larger in SARS than in 2019. Thus, the former is, in principle, capable of stronger binding to ACE2 – contrary to former predictions.

- If only significantly interacting structures are considered (that is, if extremely weak interactions derived from the SARS-ACE2 crystal structure are removed from the investigated sample) then the average of all calculated noncovalent interaction energies and sum of average energies for interactions 1-4 are also larger in SARS-CoV than in 2019-nCoV (see scenarios V and VI in Table 3). This indicates that earlier predictions (suggesting the opposite) are potentially biased towards crystal structure geometries – which might indeed be unrepresentative of bound structures in physiological conditions.

- The MMFF94 force-field is clearly inadequate for calculating realistic noncovalent energetics for the interactions at hand. Classical molecular mechanics force-fields of this sort are thus prone to produce unreliable results in cases where a small number of noncovalent factors are considered as biochemically-significant.

In addition, we suggest that the following considerations should generally be considered in future investigations of biochemically-significant noncovalent binding factors, and particularly in virus-receptor affinity estimates:

- Conformational flexibly for considered proteins in physiological conditions is cannot trivially be inferred from crystal-structure geometries (a fact that seems to have been overlooked in previous investigations); thus –
  - Virus-receptor crystal structures should not be regarded as an unquestionable source of structural information on the bound structure at hand.
  - A biochemically-significant interaction space should be considered, such that practical conclusions may statistically be inferred from it.

- Hypotheses suggesting biochemically-significant implications of particular noncovalent binding factors on macromolecular binding events should be confirmed/refuted using appropriate means and techniques – such as electronic

structure methods proven reliable for the systems considered. In such manner, some (unfortunately) flawed inferences may indeed be averted.

As a final remark, we would like to express our hopes for a growing number of practical-problem-solving quantum chemical research attempts. Indeed, even technically unsophisticated applications of electronic structure methods can be used for providing surprisingly-useful answers to important, large-scale (bio)chemical questions – as long as the latter are phrased and examined carefully and critically.

<u>Acknowledgments</u>

<u>Supporting information</u>

All geometries used in this work are provided in the ESI (in .xyz format), alongside a dedicated spreadsheet containing our raw and calculated data, and an example Gaussian input for the DFT calculations used here.

<u>References</u>

[1] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, and Y.-Z. Zhang, Nature **579**, 265 (2020).

[2] F. Li, Science (80-.). **309**, 1864 (2005).

[3] J.S. Morse, T. Lalonde, S. Xu, and W.R. Liu, ChemBioChem **21**, 730 (2020).

[4] N. Dong, X. Yang, L. Ye, K. Chen, E.W.-C. Chan, M. Yang, and S. Chen, BioRxiv 2020.01.20.913368 (2020).

[5] L. Du, Y. He, Y. Zhou, S. Liu, B.-J. Zheng, and S. Jiang, Nat. Rev. Microbiol. **7**, 226 (2009).

[6] S. Xia, Y. Zhu, M. Liu, Q. Lan, W. Xu, Y. Wu, T. Ying, S. Liu, Z. Shi, S. Jiang, and L. Lu, Cell. Mol. Immunol. 1 (2020).

[7] J. Řezáč and P. Hobza, Chem. Rev. **116**, 5038 (2016).

[8] P. Hobza, Acc. Chem. Res. **45**, 663 (2012).

[9] L.A. Burns, M.S. Marshall, and C.D. Sherrill, J. Chem. Theory Comput. **10**, 49 (2014).

[10] M.J. Gillan, D. Alfè, P.J. Bygrave, C.R. Taylor, and F.R. Manby, J. Chem. Phys. **139**, 114101 (2013).

[11] K.E. Riley and P. Hobza, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 3 (2011).

[12] N. Sylvetsky, M.K. Kesharwani, and J.M.L. Martin, J. Chem. Phys. **147**, 134106 (2017).

[13] N. Sylvetsky, M.K. Kesharwani, and J.M.L. Martin, in *Proc. ICCMSE-2017, Thessaloniki, Greece, April 21-26, 2017* (2017), p. http://arxiv.org/abs/1705.01891.

[14] D. Manna, M.K. Kesharwani, N. Sylvetsky, and J.M.L. Martin, J. Chem. Theory Comput. **13**, 3136 (2017).

[15] M.K. Kesharwani, A. Karton, N. Sylvetsky, and J.M.L. Martin, Aust. J. Chem. **EarlyView**, 10.1071/CH17588 (2018).

[16] M.K. Kesharwani, D. Manna, N. Sylvetsky, and J.M.L. Martin, J. Phys. Chem. A **122**, 2184 (2018).

[17] B. Brauer, M.K. Kesharwani, S. Kozuch, and J.M.L. Martin, Phys. Chem. Chem. Phys. **18**, 20905 (2016).

[18] T. Blundell and S. Wood, Annu. Rev. Biochem. **51**, 123 (1982).

[19] M. Amaral, D.B. Kokh, J. Bomke, A. Wegener, H.P. Buchstaller, H.M. Eggenweiler, P. Matias, C. Sirrenberg, R.C. Wade, and M. Frech, Nat. Commun. **8**, 2276 (2017).

[20] C. Peng, Z. Zhu, Y. Shi, X. Wang, K. Mu, Y. Yang, X. Zhang, Z. Xu, and W. Zhu, (2020).

[21] J.T. Ortega, M.L. Serrano, F.H. Pujol, and H.R. Rangel, EXCLI J. **19**, 410 (2020).

[22] E.S. Brielle, D. Schneidman, and M. Linial, BioRxiv 2020.03.10.986398 (2020).

[23] T.A. Halgren, J. Comput. Chem. **17**, 520 (1996).

[24] F.L. Tobiason, R.W. Hemingway, and G. Vergoten, in *Plant Polyphenols 2* (Springer US, Boston, MA, 1999), pp. 527–544.

[25] Á. Jász, Á. Rák, I. Ladjánszki, and G. Cserey, J. Mol. Struct. **1188**, 227 (2019).

[26] Gaussian 16 Revision C.01, M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. Hratchian, A. Izmaylov, J. Bloino, G. Zheng, J. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery, J. Peralta, F. Ogliaro, M. Bearpark, J. Heyd, E. Brothers, K. Kudin,

V. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, P. Rendell, J. Burant, S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. Millam, M. Klene, J. Knox, J. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. Stratmann, O. Yazyev, A. Austin, R. Cammi, C. Pomelli, J. Ochterski, R. Martin, K. Morokuma, V. Zakrzewski, G. Voth, P. Salvador, J. Dannenberg, S. Dapprich, A. Daniels, O. Farkas, J. Foresman, J. Ortiz, J. Cioslowski, and D. Fox, Gaussian Inc., Wallingford CT (2016).

[27] A.D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[28] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[29] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[30] T.A. Halgren, J. Comput. Chem. **17**, 490 (1996).

[31] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeerschd, E. Zurek, and G.R. Hutchison, J. Cheminform. **4**, 17 (2012).