VTR: an algorithm for identifying analogous contacts on protein

structures and their complexes

Vitor Pimentel^{1†}, Diego Mariano^{1†*}, Letícia Xavier Silva Cantão¹, Luana Luiza Bastos¹, Pedro Fischer², Leonardo Henrique França de Lima², Alexandre Victor Fassio¹, Raquel Cardoso de Melo-Minardi^{1*}

¹ Laboratory of Bioinformatics and Systems. Department of Computer Science. Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Brazil.

² Laboratory of Molecular Modelling and Bioinformatics (LAMMB), Department of Physical and Biological Sciences, Universidade Federal de São João Del-Rei, Campus Sete Lagoas, Sete Lagoas, 35701-970, Brazil.

† These authors contributed equally to this work.

*To whom correspondence should be addressed: RCMM: raquelcm@dcc.ufmg.br; DM: diegomariano@ufmg.br

Abstract

Evolutionarily related proteins can present similar structures but very dissimilar sequences. Hence, understanding the role of the inter-residues contacts for the protein structure has been the target of many studies. Contacts comprise non-covalent interactions, which are essential to stabilize macromolecular structures such as proteins. Here we show VTR, a new method for the detection of analogous contacts in protein pairs. VTR performs structural alignment between proteins and detects interactions that occur in similar regions. To evaluate our tool, we proposed three case studies: (i) we compared a vertebrate myoglobin and a truncated invertebrate hemoglobin; (ii) analyzed interactions between the spike protein RBD of SARS-CoV-2 and the cell receptor ACE2; and (iii) compared a glucose-tolerant and a non-tolerant β -glucosidase enzyme used for biofuel production. The case studies demonstrate the potential of VTR for the understanding of functional similarities between distantly sequence-related proteins, as well as the exploration of important drug targets and rational design of enzymes for industrial applications. We envision VTR as a promising tool for understanding differences and similarities between homologous proteins with similar 3D structures but different sequences. VTR is available at <http://bioinfo.dcc.ufmg.br/vtr>.

Keywords: protein interactions; structural bioinformatics; structural alignment; contacts; rational design of enzymes

Introduction

Proteins are macromolecules responsible for most functions in living beings, such as transport, immune protection, control growth, and so on ¹. The function of a protein is directly related to their three-dimensional structure. Previous studies have demonstrated that structural changes are directly related to sequence changes ². Even small modifications, such as mutations, insertions, or deletions, can change the structure ³⁻⁵. However, evolutionarily related proteins may present similar structures but different sequences ⁶. Sequence alignments can unambiguously distinguish similar and non-similar structures when the identity is upper to 40%. Even sequences with identities between 20-35% may generate false negatives for homology identification ⁷. Also, studies have reported similarities in structures with sequence identity lower than 20% ². Until now, the understanding of how the polypeptide sequences fold into a particular three-dimensional shape after synthesis remains a mystery ^{8,9}. It has motivated the search for computational algorithms to predict protein structures from their sequences ^{10,11} or even detect and annotate correctly protein functions ¹²⁻¹⁴. Besides, understanding protein structure and their interactions accurately are crucial to the rational design of molecules for several applications, including discovering novel drugs and improving enzymes for the biotechnological industry ^{15,16}.

Thus, understanding the role of inter-residues contacts in protein folding, stabilization, and function has been the goal of several studies ^{17,18}. Contacts are weak and potentially stabilizing interactions in the structure of macromolecules, such as proteins ^{18,19}. They can be hydrophobic interactions, electrostatic attractive or repulsive interactions, disulfide bonds, aromatic stacking, hydrogen bonds, and so on ^{17,20-22}. Contacts also have been used to compare protein structures, for instance, from contact maps ¹ or graph-based structural signatures ²³. Recent computational approaches have suggested that the substitution of non-interacting residues for interacting ones can improve protein stability, which highlights the importance of

computation of contacts¹⁶. In addition, pairwise comparisons between proteins can be performed using visual and structural alignment tools, such as PyMOL²⁴. However, comparisons between contacts are usually performed individually, which makes comparisons between a considerable number of contacts toilsome. To the best of our knowledge, there is no tool for systematic structural comparisons between contacts in a protein pair.

Therefore, herein, we propose a novel approach to detect and align contacts for protein structure analysis and pairwise comparison. Our algorithm aims to detect differences and similarities in amino acid residues pairs in contact when compared to analogous positions. For this purpose, we first perform a structural superposition between two proteins, detect contacts through a cutoff-based approach, and compare contacts in analogous positions using a score. We also developed a user-friendly web tool called VTR to facilitate the use of our method. As a proof of concept, we propose three case studies: (i) a comparison between a myoglobin (PDB ID: 1a6m) and a hemoglobin (PDB ID: 1dlw), proteins with similar structure but low sequence identity (18%); (ii) a comparison of interactions among the spike protein RBD of SARS-CoV-2, SARS-CoV, and the cell receptor ACE2; and (iii) a comparison between a glucose-tolerant β -glucosidase enzyme efficient for biofuel production (Bgl1A) and a less efficient non-tolerant β -glucosidase (Bgl1B).

Results and discussion

VTR tool workflow

VTR receives as input two PDB files (hereafter called A and B). We suggest that PDBs present similar folding, but if structures with different folds were used, VTR would try to perform the best structural alignment using the TM-align tool. It rotates and translates the coordinates of the protein A considering the alignment with B. Then, VTR uses in-house Python scripts to calculate contacts: (i) hydrogen bonds; (ii) salt bridge (hydrogen bonds and attractive); (iii)

81 ionic interaction (attractive); (iv) repulsive; (v) hydrophobic interactions; (vi) aromatic
 82 stacking; and (vii) disulfide bonds. Then, VTR determines contacts in analogous positions
 83 using AVD score (average distance between the coordinates of atoms in contact), performs
 84 comparative statistical, and returns the results for the VTR interface (Figure 1 A-C).

85

86 **Figure 1.** VTR workflow. **(a)** VTR web tool receives as input two PDB files. **(b)** VTR pipeline analyzes
 87 the PDB files in three steps: (i) structural superposition between the PDB files using TM-align; (ii)
 88 contacts calculation using cutoff definitions obtained from the literature; and (iii) search for analogous
 89 contacts using AVD strategy. **(c)** VTR returns the contact matches, and users can use them in the web
 90 interface. Also, VTR determines the statistics of differences between amino acids in contact.

91

92

Case studies

To evaluate the VTR tool, we proposed three pairwise comparative case studies: (i) eukaryotic myoglobin and a truncated non-vertebrate hemoglobin chain; (ii) interactions among the cell receptor ACE2 and both SARS-CoV-2 and SARS-CoV; and (iii) glucose-tolerant and non-tolerant β -glucosidases. Contact maps generated by VTR demonstrate similarities in the contact pattern between protein pairs in each case study. For instance, Figure 2A is more similar to Figure 2B, Figure 2C to D, and Figure 2E to F. We aim with these case studies to demonstrate a simple use for VTR (case study 1) and show the tool's effectiveness in finding shared and unshared contacts in two systems already described in the literature (case study 2). In the third case study, we propose a real use of VTR at detecting possible mutation sites for enhancing enzymes with biotechnological applications.

Figure 2. Contact maps for case studies 1 (a-b), 2 (c-d), and 3 (e-f). Each point represents a contact. Contacts of similar types are shown in the same colors. In the x-axes and y-axes are shown the residue numbers. (a) PDB ID: 1DLW; (b) PDB ID: 1A6M; (c) PDB ID: 2AJF; (d) PDB ID: 6M0J; (e) Bgl1A; (f) Bgl1B. For C and D, we show only contact maps for chain A. We consider all contacts for determination of the contact maps, including contacts between atoms of the main-chain of closer residues (such as those present in alpha-helices).

Case study 1: Myoglobin and hemoglobin

In the first case study, we performed the contacts alignment between a sperm whale myoglobin (PDB ID: 1a6m) and a truncated single-chain hemoglobin from *Paramecium caudatum* ciliated protozoan (PDB ID: 1dlw). Myoglobins and hemoglobins are both oxygen-binding proteins, belonging to the widespread and distantly related globin family²⁵. For this case study, the evaluated myoglobin structure (1a6m) was at the oxy state (i.e., oxygen bound), while the hemoglobin structure (1dlw) was at the de-oxy one (i.e., oxygen unbound). 1a6m presents a sequence length of 151 amino acids, while 1dlw, as a typical non-vertebrate truncated hemoglobin chain, has a minor amino acid content, with just 116 residues. Both proteins present a similar folding but a sequence identity of only 18%. The literature has described that structures of homologous sequences with an identity lower than 20% may present large structural differences². However, the discrepancy about this typical strong relationship between sequence and folding similarity for the globin family has long been known. In fact, globins form a family substantially conserved in structural topology, despite the distant sequence relationship, being this one of the higher conundrums in biochemistry^{3,25}. Hence, we believe these proteins to be potential targets for comparison of analogous contacts using VTR.

After processing both structures with default parameters, VTR detected the following contacts for 1a6m: 85 hydrogen bonds, 81 attractive interactions, 32 repulsive interactions, nine aromatic stacking, and 364 hydrophobic interactions. We found the following contacts in 1dlw: 65 hydrogen bonds, 20 attractive interactions, one repulsive interaction, two aromatic stacking, and 221 hydrophobic interactions. However, we obtained only 13 main contact matches in analogous positions using a 2 Å AVD cutoff (Table S2). The contact matches increase to 60 when considering conserved contact matches of hydrophobic type.

From the contacts predicted in analogous positions, we found 12 of different types (Figure 3; Table S2-4), such as V21-V26 and Q13-N43 (Figure 3A). For 1a6m, VTR detected

an attractive contact between H36 and E38. However, it detected a hydrogen bond between D26 and T28 in analogous positions of 1dlw (Figure 3B). On the other hand, in 1a6m, I75 performs hydrophobic interactions with L86, while F48 performs the same type of interaction with V109 (Figure 3C). Interestingly, I75 and L86 are located at a distance of 11 amino acids, while F48 and V109 are at a distance of 61 positions. This highlights the VTR's ability to detect contacts even in different sequence positions. Most of the contact matches are located at compatible distances of 2 to 4 amino acids, such as K102-F106 and A76-T80 (Figure 3D), L104-S108 and F78-I82 (Figure 3E), L89-H93 and L64-H68 (Figure 3F). Only the hydrogen bond detected in the contact L89-H93 of 1a6m was considered a match with Q13-N43. This may suggest that both contacts present similar importance for the stability of both proteins.

Figure 3. Six analogous contacts between 1a6m (cyan sticks) and 1dlw (green sticks). [Hy] hydrophobic; [Hb] Hydrogen bond; [At] Attractive ionic interaction. **(a)** V21-V66 (1a6m) and Q13-N43 (1dlw); **(b)** H36-E38 (1a6m) and D26-T28 (1dlw); **(c)** I75-L89 (1dlw) and F48-L64 (1dlw); **(d)** K102-F106 (1a1m) and A76-T80 (1dlw); **(e)** L104-S108 (1a6m) and F78-I82 (1dlw); and **(f)** L89-H93 (1a6m) and L64-H68 (1dlw). Sticks were colored using the cyan/green color scheme.

It is essential to highlight that for this analysis, we did not consider the interactions between atoms of the main-chain of closer residues. We did it to reduce the number of contacts detected in alpha-helices, which can explain the low number of hydrogen bonds conserved. Enabling the advanced option "detection of structural contacts in α -helices", the number of hydrogen bond contact matches increases to 129. This option is disabled by default because VTR desires highlight conserved interactions with the side-chain atoms. Also, enabling this option increases the computational cost once that all contacts of similar secondary structures in close regions will be computed.

Case study 2: Analyzing interactions between the spike protein RBD of SARS-CoV-2 and the cell receptor ACE2

Recently, a new coronavirus (SARS-CoV-2) was related to severe acute respiratory syndrome (COVID-19), spreading rapidly worldwide and causing a pandemic situation ²⁶. A sequence comparison demonstrates that SARS-CoV-2 structures share approximately 80% of sequence identity with the SARS-CoV ^{26,27}. Like SARS-CoV, SARS-CoV2 recognizes the ACE2 (Angiotensin-Converting Enzyme 2) receptor in humans. Hence, understanding the binding mechanism of the spike RBD (Receptor-Binding Domain) of SARS-CoV-2 with ACE2 may help to shed some light on the mechanism of recognition of virus receptors and the initial infection process. In a recent study, Lan et al. ²⁷ identified the relevant residues to the interaction of SARS-CoV-2 with the receptor. It was noted that most of these residues are highly conserved and shared with SARS-CoV. Here, we verified the ability of VTR to find the known contacts between different chains of both structures. We compared the structures of SARS-CoV-2 spike RBD (PDB ID 6m0j) and SARS-CoV spike RBD (PDB ID 2ajf), both in complex with the ACE2 receptor. We maintain the standard 2 Å AVD cutoff, and we use chain A (ACE2) and chain E (RBD portion of the virus) for both structures.

Compared to the Lan et al.²⁷ study, the VTR could find all 16 atomic contacts between SARS-CoV and ACE2 (3 salt bridges and 13 hydrogen bonds). VTR could also find 15 contacts between SARS-CoV-2 and ACE2 (2 salt bridges and 13 hydrogen bonds). Of these, VTR calculated and presented which contacts were shared between receptor and SARS-CoV or SARS-CoV-2. We detected all contacts described in²⁷ (Table 1).

Table 1. Analogous contacts between proteins with PDB entries SARS-CoV (6m0j) and 2ajf. R1: residue 1; R2: residue 2; HB: hydrogen bonds, HY: hydrophobic. In the last column, we highlighted lines where the matches were reported in²⁷. Residues in contact, but atom interactions were not described in²⁷, are presented as "-".

	Chain:Residue (atom)				Contact type		Reported in ²⁷
	SARS-CoV-2 (R1)	SARS-CoV-2 (R2)	SARS-CoV (R1)	SARS-CoV (R2)	SARS-CoV-2 R1-R2	SARS-CoV R1-R2	
1	A:Y41 (OH)	E:N501 (N)	A:Y41 (OH)	E:T487 (N)	HB	HB	✓
2	A:Y41 (OH)	E:N501 (OD1)	A:Y41 (CE2)	E:T487 (CG2)	HB	HY	-
3	A:D38 (OD2)	E:Y449 (OH)	A:D38 (OD2)	E:Y436 (OH)	HB	HB	✓
4	A:Y83 (OH)	E:N487 (OD1)	A:Y83 (OH)	E:N473 (ND2)	HB	HB	✓
5	A:H34 (NE2)	E:Y453 (OH)	A:H34 (NE2)	E:Y440 (OH)	HB	HB	
6	A:E37 (OE2)	E:Y505 (OH)	A:E37 (OE1)	E:Y491 (OH)	HB	HB	
7	A:D38 (OD1)	E:Y449 (OH)	A:D38 (OD1)	E:Y436 (OH)	HB	HB	-
8	A:G354 (O)	E:G502 (N)	A:G354 (O)	E:G488 (N)	HB	HB	
9	A:F28 (N)	E:Y489 (OH)	A:F28 (N)	E:Y475 (OH)	HB	HB	
10	A:Y83 (OH)	E:Y489 (OH)	A:Y83 (OH)	E:Y475 (OH)	HB	HB	✓
11	A:Q42 (NE2)	E:Y449 (OH)	A:Q42 (OE1)	E:Y436 (OH)	HB	HB	
12	A:Y41 (OH)	E:T500 (OG1)	A:Y41 (OH)	E:T486 (OG1)	HB	HB	✓
13	A:Q24 (OE1)	E:N487 (ND2)	A:Q24 (OE1)	E:N473 (ND2)	HB	HB	✓
14	A:K353 (O)	E:G502 (N)	A:K353 (O)	E:G488 (N)	HB	HB	✓

VTR calculated six shared hydrogen bonds and one hydrophobic interaction between SARS-CoV and a receptor that are not present in²⁷ (Table 1). Although the hydrophobic interaction between Y41 (CE2) and T487 (CG2) was not described in²⁷, the hydrogen bond

191 between Y41 and T487 was presented (Table 1; lines 1 and 2). VTR also detected a permuted
 192 interaction between atoms of D38 and Y449 (6m0j) and D38 and Y436 (2ajf) (Table 1; lines 3
 193 and 7). Moreover, the interactions between G354 (O) and G502 (N) of 6m0j and their
 194 equivalents in 2ajf appear to be probably detected due to the cutoff-based strategy (Table 1;
 195 line 8). Also, VTR detected that G502 interacted with K353: interaction described in ²⁷. Among
 196 the conserved hydrogen bond contacts, four called our attention as they were described in the
 197 literature (Figure 4 A-D).

198

199 **Figure 4.** Four analogous contacts between SARS-CoV and ACE2 (purple sticks), and SARS-CoV-2
 200 and ACE2 (yellow sticks). VTR suggests that (a) H34-Y453 (6M0J) is equivalent to H34-Y440 (2AJF);
 201 (b) E37-Y505 (6M0J) is equivalent to E37-Y491 (2AJF); (c) F28-Y489 (6M0J) is equivalent to F28-Y475
 202 (2AJF); and (d) Q42-Y449 (6M0J) is equivalent to Q42-Y436 (2AJF). Sticks were colored using the
 203 yellow/purple color scheme.

204

205 For SARS-CoV-2, VTR detected a hydrogen bond between H34 and Y453 that seems
 206 to be equivalent to H34 and Y440 in SARS-CoV (Figure 4A). The same applies to the

following contact pairs: E37-Y505 (6M0J) and E37-Y491 (2AJF) (Figure 4B); F28-Y489 (6M0J) and F28-436 (2AJF) (Figure 4C); and Q42-Y449 (6M0J) and Q42-Y436 (2AJF) (Figure 4D). It is important to highlight that besides the interaction between E37 and Y505 (Figure 4B), the residue E37 possibly interacts with R403 in 6M0J, but there is no equivalent contact in 2AJF. Also, Lan et al.²⁷ described an interaction between Y83 (OH) and Y489 (OH) of 6M0J. Besides this interaction, VTR also pointed out that Y489 (OH) interacts with the main-chain atoms of F28. It demonstrates that the cutoff-based strategy used by VTR also has advantages when compared to more restrictive methods. This possible interaction could, in future studies, be better comprehended through molecular dynamics experiments.

Case study 3 - Glucose-tolerant β -glucosidases

β -glucosidases (E.C. 3.2.1.21) are enzymes that act in the last step of the saccharification process, cleaving cellobiose into two molecules of glucose^{28,29}. Hence, they are considered very important for the second-generation biofuel industrial applications^{30–32}. Besides, the literature has reported that most β -glucosidases are inhibited in high glucose concentrations^{33,34}. Therefore, the design of enzymes more resistant to glucose inhibition has motivated research around the world³⁵. Recently, two β -glucosidases extracted from the marine metagenome of the South China Sea were characterized and evaluated³⁶. The first one (Bgl1A) was able to keep its activity even in glucose concentrations up to 1000 mM³⁷, being described as glucose-tolerant (a class of β -glucosidase enzymes with high potential for industrial use). On the other hand, the second one (Bgl1B) was inhibited in concentrations up to 50 mM³⁸. Both enzymes have a similar TIM-barrel folding, belonging to the first family of glycoside hydrolases (GH1)³⁹. Moreover, they present a sequence similarity higher than 50%⁴⁰. Thus, we decided to submit to VTR the structures of Bgl1A and Bgl1B to evaluate similar contacts and identify possible differences.

VTR found 375 main contact matches (984 considering hydrophobics), an average RMSD (for contact matches) of 0.96, and a VTR score of 0.19. From 375 matches, 327 keep the contact type, and 48 change the contact type. The matches that change the contact type are interesting targets for detection of differences in the structures, and potentially can be used to suggest mutations. Thus, such information may guide a biotechnological industry at providing glucose tolerance characteristics to Bgl1B through the rational design of enzymes using site-directed mutagenesis.

After analyzing the results (available in the supplementary material), one contact match called our attention: D41-H57 of Bgl1A (Figure 5A) that corresponds to D41-D57 of Bgl1B (Figure 5B). While Bgl1A presents an attractive contact, Bgl1b presents a repulsive interaction. These contacts are located in the extremities of loop A. Furthermore, this loop has been reported in the literature as necessary for restricting the entrance and exit of the active site in glucose-tolerant β -glucosidases³². Hence, changes in this region could modify the mobility of the protein and could explain the differences in the glucose tolerance of both enzymes.

The literature has reported implications of the topology and dynamics of loop A for the differences in glucose tolerance and inhibition between Bgl1A and Bgl1B^{32,36}. To probe the potential of our method for protein engineering, we have checked how considerable were the physical-chemical differences attributed by this point modification at loop A as a whole.

Firstly, we modeled *in silico* a mutant of Bgl1A (D57H) and a mutant of Bgl1B (H57D). We expected that Bgl1B's mutant present characteristics similar to Bgl1A (*i.e.*, characteristics that could lead to glucose tolerance). As a control, we modeled a Bgl1A's mutant that we expected present characteristics like Bgl1B (*i.e.*, negative characteristics for biofuel production). Then, we inspected the mesoscopic influence of such single amino acid modification at the regional electrostatic surface by Poisson-Boltzmann analysis.

Figure 5. Important contacts in analogous positions between Bgl1A and Bgl1B. **(a)** Attractive interaction D41-H57 in loop A of Bgl1A. **(b)** Repulsive interaction D41-D57 in loop A of Bgl1B. **(a-b)** Protein backbones are shown as grey cartoons. Amino acid residues are shown as cyan (Bgl1A) and green (Bgl1B) cartoons. In both proteins, D41 also interacts with R35. **(c-d)** Illustrative scheme of the importance of these contacts. **(c)** Bgl1A: H57 performs attractive interactions with D41. **(d)** Bgl1B: D41 performs repulsive interactions with D57.

PBSA points substantial electrostatic differences

The estimation of the protonation states at pH 7.0 has recovered the usual HSE protonation for the H57 in Bgl1A (*i.e.*, a neutral histidine protonated just at the ϵ nitrogen atom). This is consonant with the position of the side chain of this histidine in loop A relatively turned to the solvent, with just a marginal contact with the D41 residue at the opposite side. Also, the presence of the positively charged R35 prevents that the H57 local pKa be enough changed by the next D41 to induce a bi-protonation at this histidine (Figure 5). Hence, while the D41-D57 in Bgl1B can be classified as an electrically repulsive contact, the D41-H57 interaction in Bgl1A is not a charge to charge, but a charge-dipole attractive interaction. Even so, the Poisson-

274 Boltzmann surface analysis (PBSA) indicates a substantial difference between Bgl1A and
 275 Bgl1B in electrostatic surface potential in loop A (Figure 6 A-B).

276
 277 **Figure 6.** Poisson-Boltzmann surface potential at the vicinity between positions 41 and 57 in Bgl1A,
 278 Bgl1B, and mutants. **(a)** Bgl1A; **(b)** Bgl1B; **(c)** Bgl1A's H57D mutant; **(d)** Bgl1B's D57H mutant. The
 279 PBSA colors are on a red: white: blue scale for electrostatic potentials around $-10.00 : -5.00 : 0.00 \text{ } k_b T / \beta$
 280 unities (being k_b the Boltzmann constant, T the 300 K temperature, and β the electron coulomb charge).
 281 Residues 41 and 57 are depicted as spheres. Regions whose electrostatic potential is affected by the
 282 H57D or D57D permutations are depicted as "x" for Bgl1A or "*" for Bgl1B.
 283

284 Indeed, all the topological context around loop A in Bgl1B accumulates a substantial
 285 amount of negative electrical potential (red region in Figure 6B). On the other hand, Bgl1A
 286 presents a surface with a more neutral electrical potential (blue and white regions in Figure
 287 6A). While the D41-D57 contact appears to be a hot spot of negative charge accumulation in
 288 Bgl1B, the correspondent D41-H57 in Bgl1A shares similar neutrality of the surrounding
 289 region. The substitutions H57D (Bgl1A) and D57H (Bgl1B) introduce a point inversion of this
 290 behavior in both proteins (Figure 6 C-D). Also, these permutations cause electrostatic
 291 modifications at specific points of this region. Comparing Figure 6A and Figure 6C (Bgl1A

and its mutant), we can observe an increase of red regions. Contrarily, comparing Figure 6B and Figure 6D (Bgl1B and its mutant), we can observe a reduction of the red regions.

Furthermore, the sites with local potential affected by the permutations appear to be the same in Bgl1A and Bgl1B, spread around loop A. The dynamics and topology of this region are strongly correlated to the functional differences between both proteins^{32,36}. In addition, the charge inversions in this region have been correlated to changes in activity and stability^{41,42}. All these factors led us to look for a glimpse of the influence of the D41-H57D permutation in the vibrational dynamics of loop A.

Vibrational modulation of loop A

The representativity of the eigenvectors (Supplementary Figure S1) demonstrates the differences between the PCA recovered fluctuation between loops A of Bgl1A and Bgl1B. Bgl1A fluctuation is more homogeneously represented by the two first eigenvectors (PCs) of loop A, with a fractional eigenvalue of 63% for the PC1 and 29% for PC2. On the other hand, the fluctuations at the Bgl1B's loop A are strongly located at PC1, that alone accounts for 82% of the sum of all eigenvalues. This can indicate that a single kind of movement, with a substantial fugacity from the middle structure at the minima, account for most of the vibratory mobility of Bgl1B's loop A. This is consonant with a local instability caused by the highly repulsive environment at this loop (Figure 6B).

In Figure S1, we can see higher proximity among the lines that represent wild Bgl1B, Bgl1B's mutant, and Bgl1A's mutant. The loop A of the wild Bgl1A presents a more equilibrated and distributed fluctuation between different modes (in concurrence with the more neutral profile in Figure 6 A). In addition, while the H57D exchange is enough to invert the vibrational eigenvalue distribution of the Bgl1A's loop A to the Bgl1B pattern, the same does not occur for Bgl1B with the D57H permutation (Figure S1 A). Hence, the more neutralized electrostatic surface for Bgl1A (Figure S1 A) is sensitive to the point insertion of a repulsive

contact at the loop A basis. On the other hand, the strongly negatively charged Bgl1B's loop A is less responsive to neutralization at this single point.

Furthermore, the analysis of the distribution of the fluctuations allocated at the two first eigenvectors (PC1+2) of loop A shows also considerable differences (Figure S1 B-F). Bgl1A presents higher apparent vibrational mobility concentrated at the C-terminal portion of the N-terminal helix of this loop (residues 45 to 50). In contrast, Bgl1B has its high mobility equally spread along the entire medial portion of loop A (residues 45 to 55). The different electrostatic context (Figure 6 A-B) may explain the vibrational distinctions (Figure S1 A-B). Despite this, the impact in contact patterns, caused by changes in position 57, promotes at least qualitatively standardized modifications around half of loop A (region indicated by lines in Figure S1 C-F). The presence of a negatively charged amino acid in position 57 promotes, both in Bgl1A and Bgl1B's environments, an increase in mobility of positions 44, 50, 51, and 57 itself. It also promotes a reduction in the mobility of the region between residues 45 to 49.

For the D41-D57 repulsive contact, the vibrational movement of both acids is unbalanced, with one moving more or faster than the other ("*" in Figure S1 D-E). On the other hand, the movement involving the D41-H57 contact is more balanced (mainly for Bgl1A's context), indicating a more equilibrated vibration again.

Implications for protein engineering

The intensity of some of these modifications or changes at the rest of loop A seems to be context-dependent. This agrees with the fact that the permutation of the entire loop A between Bgl1A and Bgl1B was able to revert the glucose tolerant/inhibited behavior, but with poor results for local single amino acid substitutions³⁶. Despite this, a single substitution detected by VTR was able to promote vibrational changes with the same pattern at almost half of the extension of this functionally crucial loop.

All these mentioned facts illustrate a potential use for this tool. If properly combined with electrostatic and mobility patterns detection techniques, such as the APBS and molecular mechanics/PCA here employed, VTR can be useful for rational protein engineering. The integrated use of VTR with computational or experimental tools can be promising to find the topologically minimum modifications that need to be carried at proteins to introduce some desirable characteristics found in other homologous.

For GH1 β -glucosidases, a protein class extensively used in industry and with a strong interest in the second-generation bioethanol production ²⁹, this approach shows itself encouraging. Overall, minimal topological modifications represent a promising strategy for suggesting mutations, especially in the beta-glucosidase loop regions that surround the active site. The topology and dynamics of these loops can allow or restrict movements involved in glucose entrance and exit (*i.e.*, glucose tolerance) ^{30,32,36,43} or also can affect the thermostability ^{41–43}. These are both examples of industrially desirable characteristics for these and other proteins.

Methods

VTR algorithm receives as input two PDB files and processes them in the back-end using in-house scripts. The files are analyzed in three steps: (i) structural superposition; (ii) contacts computation; and (iii) search for contact matches.

Structural superposition

VTR performs structural alignments between protein pairs using the default parameters of the TM-align algorithm (<https://zhanglab.ccmb.med.umich.edu/TM-align>). TM-align receives as input two PDB files and returns the coordinates of a superimposed structure. TM-align will return the best alignment possible, and VTR will return a warning informing if contact matches could not be found.

Contacts computation

We compute the Euclidean distance among all-atom coordinates using in-house Python scripts. The tool identifies five types of possible interactions: hydrophobic, attractive/repulsive, hydrogen bonds, aromatic stacking, and disulfide bonds. A pair of residues are in contact if any atom pair meets the cutoff ranges presented in Table 2.

Table 2. Cutoff distances for each contact type. Cutoff values obtained from ^{1,20,21,44,45}.

Contact type	Atom classes (residue 1-2)	Cutoff (min-max)
Hydrophobic	Hydrophobic - Hydrophobic	2 - 4.5 Å
Attractive/repulsive	Positively - negatively charged (attractive) Positively - positively charged (repulsive) Negatively - negatively charged (repulsive)	2 - 6 Å
Hydrogen bonds	Acceptor - Donor	≤ 3.9 Å
Aromatic stacking	Ring centroid - ring centroid	2 - 6 Å
Disulfide bonds	CYS - CYS	1.5 - 2.8 Å

Contact must occur between atoms of two different residues. The atoms involved are described in Table S1. For the detection of aromatic stacking, we determined the centroids of the rings (cutoff 2-6Å). For phenylalanine and tyrosine, we calculate the median coordinate between the atoms CG and CZ for determining the ring centroid. For tryptophan, we used the median coordinate between the atoms CD2 and CE2. For histidine, we determined the median coordinate of the atoms CG, ND1, CE1, NE2, and CD2.

By default, VTR ignores contacts between atoms of the main-chain of neighborhood residues (until four positions) to remove contacts that compose the structures of α -helices. However, through the web tool, users can enable the detection of these contacts (increase processing time).

Search for contact matches

We used in-house scripts to compare the contacts in analogous positions. We proposed the AVD (average vector distance) metric to measure the distance between contacts and detect contacts matches. AVD is calculated by equation (1):

$$AVD(P, Q) = \min\left(\frac{D(p1, q1) + D(p2, q2)}{2}, \frac{D(p1, q2) + D(p2, q1)}{2}\right) \quad (1)$$

where P represents the contact between atoms $p1$ and $p2$ of protein A ; Q represents the contact between the atoms $q1$ and $q2$ of protein B ; and D is a function that returns the Euclidean distance between atomic coordinates. To detect a contact match between P and Q , the $AVD(P, Q)$ should be the lowest value. We determine the VTR score based on equation (2):

$$VTR(A, B) = \frac{\sum_{i=1}^n AVD_i (d_A + d_B)}{C n (c_A + c_B)} \quad (2)$$

where A and B are the proteins analyzed; n is the number of matches found; C is the cutoff determined for the AVD; d_A and d_B are the number of contacts without match in both proteins; c_A and c_B are the number of contacts found in each protein. The VTR score is a value between 0 and 1, where the lower the value, the more similar are two proteins in terms of inter-residue contacts.

Web-based tool

The VTR web tool was developed using CodeIgniter (<https://codeigniter.com/>), D3 (<https://d3js.org>), jQuery (<https://jquery.com>), DataTables (<https://datatables.net>), and Bootstrap CSS and JavaScript library (<https://getbootstrap.com>). 3D structure visualizations are presented using 3Dmol.js⁴⁶.

Case studies details

For case study 1, we collected the PDBs entries of a sperm whale myoglobin (PDB ID: 1a6m)⁴⁷ and the *Paramecium caudatum* single-chain and truncated hemoglobin (PDB ID: 1dlw)⁴⁸. For the second case study, we selected SARS-CoV-2 (PDB ID: 6m0j)²⁷ and SARS-CoV (PDB ID: 2ajf)⁴⁹ structures in the RCSB PDB. Each PDB file contains two chains A and E, where A represents the cell receptor ACE2 and E the receptor-binding domain (RBD) portion of the virus.

For the third case study, we selected the sequences of the glucose-tolerant GH1 β -glucosidase of a South China Sea metagenome (Bgl1A; UniProt ID: D5KX75)³⁷ and the non-tolerant GH1 β -glucosidase of a South China Sea metagenome (Bgl1B; UniProt ID: D0VEC8)³⁸ from Glutantbase⁵⁰. We also constructed two mutants, H57D (Bgl1A) and D57H (Bgl1B), to evaluate VTR ability to propose mutations for enzymes based on differences of contacts (Tables S5-6). The 3D structures of wild (Bgl1A and Bgl1B) and mutant (Bgl1A: H57D and Bgl1B: D57H) structures were constructed using SWISS-MODEL^{51,52} (Tables S7-8). To evaluate the impact of mutations, we estimated the Poisson-Boltzmann surface area (PBSA) with the integrated use of the online versions of the H++, PDB2PQR, and the Adaptive Poisson-Boltzmann Solver (APBS) tools⁵³⁻⁵⁷. Details were included in the supplementary material (Additional text S1).

Conclusion

Herein, we presented VTR, a novel approach with a visual web interface that can be used to analyze, compare, and scrutinize analogous contacts in protein pairs. We explored the tool features with three case studies, where we demonstrated the potential of VTR to shed some light on the mechanisms of topology conservation on phylogenetically related but sequentially distant proteins. We also evaluated contact similarities and dissimilarities on

pharmacologically relevant targets. We suggested the use of our tool to the rational design of proteins with biotechnological applications. Concerning the second case study, both the confirmation of contacts already reported in the literature and the finding of four hydrogen bond matches still not described are promising finds for the future rational design of anti-Sars-Cov-2 drugs. In the last case study, we compared a glucose-tolerant beta-glucosidase enzyme with a non-tolerant one. VTR detected several changes in contact types of analogous positions. Called our attention a change in an attractive contact by a repulsive: D41-H57 of Bgl1A that corresponds to D41-D57 of Bgl1B. We explored the importance of this contact using molecular mechanics minimization and vibrational inspection by principal component analysis. Our results demonstrate that the presence of this contact is vital for the stability of Bgl1A. However, Bgl1B's mutant with a similar interaction was not enough to present similar characteristics of the glucose-tolerant one on a substantial extension. Nevertheless, this case study illustrates the potential of the VTR tool for the rational design of industrial enzymes and gives some glimpses about electrostatic and vibrational aspects of β -glucosidase enzymes. We hope VTR can be used for understanding differences and similarities between homologous proteins with similar 3D structure but differences in sequences. VTR is available at <http://bioinfo.dcc.ufmg.br/vtr>.

Author contributions

VP and DM developed the webtool and performed the first case study. LXSC and LLB performed the second case study. LHFL and PF performed APBS and molecular dynamics analysis for the third case study. DM wrote the manuscript. VP, PF, LHFL, AVF, and RCMM revised the manuscript. Project design and funding acquisition: RCMM. All authors read and approved the manuscript.

Availability of data and material

The data underlying this article are available in the article and in its online supplementary material.

Supplementary material

Additional text S1. Material and methods (case study 3)

Figure S1. Vibrational mobility pattern recovered by minimization and PCA for loop A of Bgl1A, Bgl1B, and mutants.

Table S1. Atoms classes for each type of contact.

Table S2. Analogous contacts between proteins with PDB entries 1a6m and 1dlw.

Table S3. 1A6M Contacts (Chain A).

Table S4. 1DLW Contacts (Chain A).

Table S5. Models Assessment (BGL1B).

Table S6. Models Assessment (BGL1A).

Table S7. Models Assessment (D57H_BGL1B).

Table S8. Models Assessment (H57D_BGL1A).

Competing interests

The author(s) declare no competing interests.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Project grant number 51/2013 - 23038.004007/2014-82.

Acknowledgments

The authors thank the funding agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

1. de Melo, R. C. *et al.* A contact map matching approach to protein structure similarity analysis. *Genet Mol Res* **5**, 284–308 (2006).
2. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823–826 (1986).
3. Lesk, A. M. & Chothia, C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225–270 (1980).
4. Lesk, A. M. & Chothia, C. Evolution of proteins formed by beta-sheets. II. The core of the immunoglobulin domains. *J. Mol. Biol.* **160**, 325–342 (1982).
5. Almassy, R. J. & Dickerson, R. E. Pseudomonas cytochrome c551 at 2.0 Å resolution: enlargement of the cytochrome c family. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 2674–2678 (1978).
6. Gan, H. H. *et al.* Analysis of protein sequence/structure similarity relationships. *Biophys. J.* **83**, 2781–2791 (2002).
7. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12**, 85–94 (1999).
8. Upadhyay, A. Structure of proteins: Evolution with unsolved mysteries. *Prog. Biophys. Mol. Biol.* **149**, 160–172 (2019).
9. Science, A. A. for the A. of. So Much More to Know … *Science* **309**, 78–102 (2005).

- 498 10. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The Protein Folding Problem.
 499 *Annu. Rev. Biophys.* **37**, 289–316 (2008).
- 500 11. Dill, K. A. & MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **338**,
 501 1042–1046 (2012).
- 502 12. Veloso, C. *et al.* On the characterization of energy networks of proteins. *Genet Mol Res*
 503 **6**, 799–820 (2007).
- 504 13. Silveira, S. de A., de Melo-Minardi, R. C., da Silveira, C. H., Santoro, M. M. & Meira Jr,
 505 W. ENZYMAP: Exploiting Protein Annotation for Modeling and Predicting EC Number
 506 Changes in UniProt/Swiss-Prot. *PLoS ONE* **9**, (2014).
- 507 14. Franciscani Jr, G. *et al.* An annotation process for data visualization techniques. in.
- 508 15. Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **257**, 1078–
 509 1082 (1992).
- 510 16. Barroso, J. R. M. S. *et al.* Proteus: An algorithm for proposing stabilizing mutation pairs
 511 based on interactions observed in known protein 3D structures. *BMC Bioinformatics* **21**,
 512 275 (2020).
- 513 17. Melo, R. *et al.* Finding protein-protein interaction patterns by contact map matching.
 514 *Embrapa Informática Agropecuária-Artigo Em Períód. Indexado ALICE* (2007).
- 515 18. Silva, M. F. M. *et al.* Proteingo: Motivation, user experience, and learning of molecular
 516 interactions in biological complexes. *Entertain. Comput.* **29**, 31–42 (2019).
- 517 19. Martins, P. M. *et al.* How to Compute Protein Residue Contacts More Accurately? in
 518 *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* 60–67 (ACM,
 519 2018). doi:10.1145/3167132.3167136.
- 520 20. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. Automated analysis
 521 of interatomic contacts in proteins. *Bioinformatics* **15**, 327–332 (1999).
- 522 21. Neshich, G. *et al.* STING Millennium: a web-based suite of programs for comprehensive

- 523 and simultaneous analysis of protein structure and sequence. *Nucleic Acids Res.* **31**,
524 3386–3392 (2003).
- 525 22. Mancini, A. L. *et al.* STING Contacts: a web-based application for identification and
526 analysis of amino acid contacts within protein structure and across protein interfaces.
527 *Bioinformatics* **20**, 2145–2147 (2004).
- 528 23. Pires, D. E. V., de Melo-Minardi, R. C., da Silveira, C. H., Campos, F. F. & Meira, W.
529 aCSM: noise-free graph-based signatures to large-scale receptor-based ligand prediction.
530 *Bioinforma. Oxf. Engl.* **29**, 855–861 (2013).
- 531 24. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
- 532 25. Hardison, R. C. Evolution of Hemoglobin and Its Genes. *Cold Spring Harb. Perspect.*
533 *Med.* **2**, (2012).
- 534 26. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat
535 origin. *Nature* **579**, 270–273 (2020).
- 536 27. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the
537 ACE2 receptor. *Nature* **581**, 215–220 (2020).
- 538 28. Cairns, J. R. K. & Esen, A. β -Glucosidases. *Cell. Mol. Life Sci.* **67**, 3389–3405 (2010).
- 539 29. Mariano, D. C. B. *et al.* Characterization of glucose-tolerant β -glucosidases used in
540 biofuel production under the bioinformatics perspective: a systematic review. *Genet. Mol.*
541 *Res.* **16**, (2017).
- 542 30. Costa, L. S. C. *et al.* Molecular Dynamics Gives New Insights into the Glucose Tolerance
543 and Inhibition Mechanisms on β -Glucosidases. *Molecules* **24**, 3215 (2019).
- 544 31. Bergmann, J. C. *et al.* Discovery of two novel β -glucosidases from an Amazon soil
545 metagenomic library. *FEMS Microbiol. Lett.* **351**, 147–155 (2014).
- 546 32. Lima, L. H. F. de *et al.* Conformational flexibility correlates with glucose tolerance for
547 point mutations in β -glucosidases – A computational study. *J. Biomol. Struct. Dyn.* **0**, 1–

- 548 20 (2020).
- 549 33. Teugjas, H. & Våljamäe, P. Selecting β -glucosidases to support cellulases in cellulose
550 saccharification. *Biotechnol. Biofuels* **6**, 1 (2013).
- 551 34. de Giuseppe, P. O. *et al.* Structural basis for glucose tolerance in GH1 β -glucosidases.
552 *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1631–1639 (2014).
- 553 35. Salgado, J. C. S., Meleiro, L. P., Carli, S. & Ward, R. J. Glucose tolerant and glucose
554 stimulated β -glucosidases - A review. *Bioresour. Technol.* **267**, 704–713 (2018).
- 555 36. Yang, Y. *et al.* A mechanism of glucose tolerance and stimulation of GH1 β -
556 glucosidases. *Sci. Rep.* **5**, 17296 (2015).
- 557 37. Fang, Z. *et al.* Cloning and Characterization of a β -Glucosidase from Marine Microbial
558 Metagenome with Excellent Glucose Tolerance. *J. Microbiol. Biotechnol.* **20**, 1351–1358
559 (2010).
- 560 38. Fang, W. *et al.* Cloning and characterization of a beta-glucosidase from marine
561 metagenome. *Sheng Wu Gong Cheng Xue Bao Chin. J. Biotechnol.* **25**, 1914–1920
562 (2009).
- 563 39. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert
564 resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
- 565 40. Mariano, D. *et al.* A Computational Method to Propose Mutations in Enzymes Based on
566 Structural Signature Variation (SSV). *Int. J. Mol. Sci.* **20**, (2019).
- 567 41. Tamaki, F. K., Textor, L. C., Polikarpov, I. & Marana, S. R. Sets of covariant residues
568 modulate the activity and thermal stability of GH1 β -glucosidases. *PloS One* **9**, e96627–
569 e96627 (2014).
- 570 42. González-Blasco, G., Sanz-Aparicio, J., González, B., Hermoso, J. A. & Polaina, J.
571 Directed evolution of β -glucosidase A from *Paenibacillus polymyxa* to thermal
572 resistance. *J. Biol. Chem.* **275**, 13708–13712 (2000).

- 573 43. Konar, S., Sinha, S. K., Datta, S. & Ghorai, P. K. Probing the Effect of Glucose on the
574 Activity and Stability of β -Glucosidase: An All-Atom Molecular Dynamics Simulation
575 Investigation. *ACS Omega* **4**, 11189–11196 (2019).
- 576 44. Bickerton, G. R., Higuieruelo, A. P. & Blundell, T. L. Comprehensive, atomic-level
577 characterization of structurally characterized protein-protein interactions: the PICCOLO
578 database. *BMC Bioinformatics* **12**, 313 (2011).
- 579 45. Fassio, A. V., Santos, L. H., Silveira, S. A., Ferreira, R. S. & de Melo-Minardi, R. C.
580 nAPOLI: a graph-based strategy to detect and visualize conserved protein-ligand
581 interactions in large-scale. *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2019)
582 doi:10.1109/TCBB.2019.2892099.
- 583 46. Rego, N. & Koes, D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**,
584 1322–1324 (2015).
- 585 47. Vojtechovský, J., Chu, K., Berendzen, J., Sweet, R. M. & Schlichting, I. Crystal
586 structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.* **77**,
587 2153–2174 (1999).
- 588 48. Pesce, A. *et al.* A novel two-over-two alpha-helical sandwich fold is characteristic of the
589 truncated hemoglobin family. *EMBO J.* **19**, 2424–2434 (2000).
- 590 49. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike
591 receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
- 592 50. Mariano, D. *et al.* Glutant β ase: a database for improving the rational design of glucose-
593 tolerant β -glucosidases. *BMC Mol. Cell Biol.* **21**, 50 (2020).
- 594 51. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure
595 using evolutionary information. *Nucleic Acids Res.* **42**, W252-258 (2014).
- 596 52. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-
597 based environment for protein structure homology modelling. *Bioinforma. Oxf. Engl.* **22**,

- 598 195–201 (2006).
- 599 53. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite.
- 600 *Protein Sci.* **27**, 112–128 (2018).
- 601 54. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of
- 602 nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**,
- 603 10037–10041 (2001).
- 604 55. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an
- 605 automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations.
- 606 *Nucleic Acids Res.* **32**, W665–W667 (2004).
- 607 56. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction
- 608 and the preparation of biomolecular structures for atomistic molecular modeling and
- 609 simulations. *Nucleic Acids Res.* **40**, W537–541 (2012).
- 610 57. Unni, S. *et al.* Web servers and services for electrostatics calculations with APBS and
- 611 PDB2PQR. *J. Comput. Chem.* **32**, 1488–1491 (2011).

612

Figure legends

Figure 1. VTR workflow. **(a)** VTR web tool receives as input two PDB files. **(b)** VTR pipeline analyzes the PDB files in three steps: (i) structural superposition between the PDB files using TM-align; (ii) contacts calculation using cutoff definitions obtained from the literature; and (iii) search for analogous contacts using AVD strategy. **(c)** VTR returns the contact matches, and users can use them in the web interface. Also, VTR determines the statistics of differences between amino acids in contact.

Figure 2. Contact maps for case studies 1 **(a-b)**, 2 **(c-d)**, and 3 **(e-f)**. Each point represents a contact. Contacts of similar types are shown in the same colors. In the x-axes and y-axes are shown the residue numbers. **(a)** PDB ID: 1DLW; **(b)** PDB ID: 1A6M; **(c)** PDB ID: 2AJF; **(d)** PDB ID: 6M0J; **(e)** Bgl1A; **(f)** Bgl1B. For C and D, we show only contact maps for chain A. We consider all contacts for determination of the contact maps, including contacts between atoms of the main-chain of closer residues (such as those present in alpha-helices).

Figure 3. Six analogous contacts between 1a6m (cyan sticks) and 1dlw (green sticks). [Hy] hydrophobic; [Hb] Hydrogen bond; [At] Attractive ionic interaction. **(a)** V21-V66 (1a6m) and Q13-N43 (1dlw); **(b)** H36-E38 (1a6m) and D26-T28 (1dlw); **(c)** I75-L89 (1dlw) and F48-L64 (1dlw); **(d)** K102-F106 (1a1m) and A76-T80 (1dlw); **(e)** L104-S108 (1a6m) and F78-I82 (1dlw); and **(f)** L89-H93 (1a6m) and L64-H68 (1dlw). Sticks were colored using the cyan/green color scheme.

Figure 4. Four analogous contacts between SARS-CoV and ACE2 (purple sticks), and SARS-CoV-2 and ACE2 (yellow sticks). VTR suggests that **(a)** H34-Y453 (6M0J) is equivalent to H34-Y440 (2AJF); **(b)** E37-Y505 (6M0J) is equivalent to E37-Y491 (2AJF); **(c)** F28-Y489 (6M0J) is equivalent to F28-Y475 (2AJF); and **(d)** Q42-Y449 (6M0J) is equivalent to Q42-Y436 (2AJF). Sticks were colored using the yellow/purple color scheme.

Figure 5. Important contacts in analogous positions between Bgl1A and Bgl1B. **(a)** Attractive interaction D41-H57 in loop A of Bgl1A. **(b)** Repulsive interaction D41-D57 in loop A of Bgl1B. **(a-b)** Protein backbones are shown as grey cartoons. Amino acid residues are shown as cyan (Bgl1A) and green (Bgl1B) cartoons. In both proteins, D41 also interacts with R35. **(c-d)** Illustrative scheme of the importance of these contacts. **(c)** Bgl1A: H57 performs attractive interactions with D41. **(d)** Bgl1B: D41 performs repulsive interactions with D57.

Figure 6. Poisson-Boltzmann surface potential at the vicinity between positions 41 and 57 in Bgl1A, Bgl1B, and mutants. **(a)** Bgl1A; **(b)** Bgl1B; **(c)** Bgl1A's H57D mutant; **(d)** Bgl1B's D57H mutant. The PBSA colors are on a red: white: blue scale for electrostatic potentials around -10.00 : -5.00 : 0.00 $k_b T / \beta$ unities (being k_b the Boltzmann constant, T the 300 K temperature, and β the electron coulomb charge). Residues 41 and 57 are depicted as spheres. Regions whose electrostatic potential is affected by the H57D or D57D permutations are depicted as "x" for Bgl1A or "*" for Bgl1B.