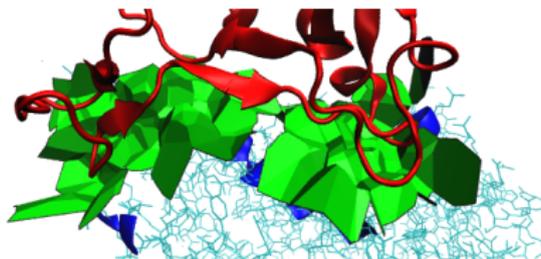


# Multiple Interface String Alignments

## Boosting the analysis of protein interfaces



MISA SSE for MISA-ID SARS-CoV-2-RBD\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
 Bend - Hydrogen bonded turn - Other - Missing Residue -

Residue Index

bound-6e0j-E, res :2.45 Å, 26 interf res  
 bound-6l2g-B, res :2.5 Å, 39 interf res  
 unbound-closed-6vxx-A, res :2.8 Å, 25 interf res  
 unbound-closed-6vyb-A, res :3.2 Å, 26 interf res

Residue Index

bound-6e0j-E, res :2.45 Å, 26 interf res  
 bound-6l2g-B, res :2.5 Å, 39 interf res  
 unbound-closed-6vxx-A, res :2.8 Å, 26 interf res  
 unbound-closed-6vyb-A, res :3.2 Å, 26 interf res

403---410---420---430---440---450---460  
 R-DE-----KJ-Y-----N-----VGG-Y--Y-LF---  
 R-DE-----KJ-Y-----N-----VGG-Y--Y-LF---  
 R-DE-----KJ-Y-----N-----\*\*G-Y--Y\*\*---  
 R-DE-----KJ-Y-----N---\*\*\*-Y--Y-LF---

463---470---480---490---500---  
 ---Y-AGS-----EGFN-YF-LQSYGFQPTN\*GVGYG  
 ---Y-AGS-----EGFN-YF-LQSYGFQPTN\*GVGYG  
 -----\* \*\*\* \*\*\*\*-YF-LQSYGFQPTN\*GVGYG  
 -----\* \*\*\* \*\*\*\*-YF-LQSYGFQPTN\*GVGYG

- ▶ F. Cazals, Inria Sophia Antipolis
- ▶ S. Béreux, Ecole Polytechnique
- ▶ B. Delmas, INRAE, Virology

# MISAs

Context and motivation

MISAs: construction

Software

# Insights into binding: sequences versus structures

▷ **Binding affinity and average residence time:** sure, but...

▶  $\Delta G_a^0 = RT \log c^0 K_a = RT \log K_d / c^0$

▶  $\tau = 1 / K_{off}$

▷ **Ad-hoc analysis based on various proxys**

• Harrison et al, Science 309, 2005: complex SARS-Cov-1 x various ACE2

**Table 1.** Contacts between ACE2 and SARS-CoV RBD. Residues in ACE2 that contact the RBD are listed by their position (numbers across the top of each column) and by their single-letter identity (36) in the palm-civet, mouse, rat, and human receptors. The residues they contact in the structure described here and their position numbers in the spike proteins from human isolates are shown at the bottom of each column.

	24	27	31	34	37	38	41	42	45	79	82	83	90	325	329	330	353	354	
L	T	T	Y	Q	E	Y	Q	V	L	T	T	Y	D	Q	E	N	K	G	civet ACE2
N	T	N	Q	E	D	Y	Q	L	T	S	F	T	Q	A	N	H	G	G	mouse ACE2
K	S	K	Q	E	D	Y	Q	L	I	N	F	N	P	T	N	H	G	G	rat ACE2
Q	T	K	H	E	D	Y	Q	L	M	Y	N	Q	E	N	K	G	G	G	human ACE2
	N473	Y475	Y475	Y440	Y491	Y436	Y484	Y436	Y484	L472	L472	N473	T402	R426	R426	T486	G488	Y491	human SARS
			Y442	N479			T486	Y484				Y475					G488	G488	
							T487										Y491		

• Wilson et al, Science 368, 2020: SARS-Cov-1 x ACE2 vs SARS-Cov-2 x ACE2

**A**

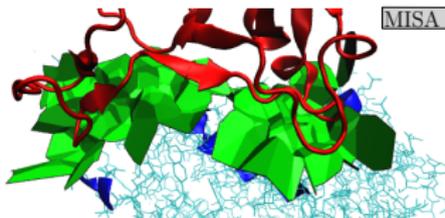
SARS-CoV RBD	306	RVVP	SGD	VVRF	PNIT	NLC	PF	GE	VFN	AT	KFP	SV	IA	MER	KK	I	S	N	C	V	A	D	S	V	L	355	
SARS-CoV-2 RBD	319	R	V	Q	P	T	S	I	V	R	F	P	N	I	T	N	L	C	P	F	G	E	V	F	N	A	368
SARS-CoV RBD	356	Y	N	S	T	P	F	S	T	F	K	C	Y	G	S	A	T	L	N	D	L	C	F	S	N	V	405
SARS-CoV-2 RBD	369	Y	N	S	A	S	F	S	T	F	K	C	Y	G	S	A	T	L	N	D	L	C	F	S	N	V	418
SARS-CoV RBD	406	A	D	Y	N	K	L	P	D	F	F	M	G	C	V	L	A	M	N	T	N	I	D	A	T	S	455
SARS-CoV-2 RBD	419	A	D	Y	N	K	L	P	D	F	F	M	G	C	V	I	A	M	N	S	N	I	D	S	K	V	468
SARS-CoV RBD	456	S	H	V	F	S	I	D	G	K	P	C	T	P	-	P	A	L	S	C	W	F	L	D	E	504	
SARS-CoV-2 RBD	469	S	T	E	I	Q	A	S	T	P	C	N	G	V	E	P	-	P	A	L	S	C	W	F	L	D	518
SARS-CoV RBD	505	H	A	P	A	T	V	C	G	P	K	S	T	L	D	L	I	K	H	Q	C	V	N	P	538		
SARS-CoV-2 RBD	519	H	A	P	A	T	V	C	G	P	K	S	T	L	D	L	I	K	H	Q	C	V	N	P	541		

▷ **Interfaces reported by hand:** often incomplete (down to 50% of interface a.a.); no generic alignment; no highlighting of specific properties

# Rationale for a more powerful

## representation/comparison of interfaces

- ▶ **Wishlist:** standard interface model; generic sequence alignment; highlighting specific properties
- ▶ **Highlighting of specific properties:**
  - ▶ MISA/SSE: coloring based on Secondary Structure Elements
    - ▶ Stability of SSE - hydrogen bonding network
  - ▶ MISA/BSA: coloring based on Buried Surface Area
    - ▶ BSA often correlates with  $\Delta\Delta G$
  - ▶ MISA/ $\Delta$ ASA: coloring based on the variation of accessible surface area.
    - ▶ Stressing local conformational changes upon binding



MISA SSE for MISA-ID SARS-CoV-2-RBD\_0

2 turn helix • 2 chain helix • 3 turn helix • isolated alpha-helix residue • Extended strand •  
loop • Hydrogen bonded turn • Other • Missing Residue

Residue Index

bound-440j-e, res :2.40 Å, 26 interf res  
bound-414g-d, res :2.5 Å, 39 interf res  
unbound-closed-460a-k, res :2.8 Å, 26 interf res  
unbound-closed-461b-k, res :3.2 Å, 25 interf res

Residue Index

bound-440j-e, res :2.40 Å, 26 interf res  
bound-414g-d, res :2.5 Å, 39 interf res  
unbound-closed-460a-k, res :2.8 Å, 26 interf res  
unbound-closed-461b-k, res :3.2 Å, 25 interf res

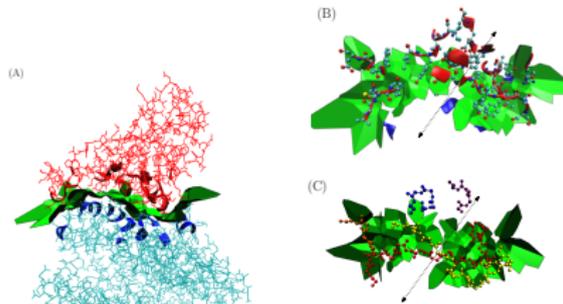
483---419---420---430---440---439---460  
R-DE---K2-Y-----N-----VGG-Y-Y-LR-  
R-DE---K2-Y-----N-----VGG-Y-Y-LR-  
R-DE---K2-Y-----N-----VGG-Y-Y-LR-  
R-DE---K2-Y-----N-----VGG-Y-Y-LR-

483---479---480---490---510---  
I---I---I---I---I---  
---Y-AGS---EGFN-Y-LSYVGQPTNIGVGYO  
---Y-AGS---EGFN-Y-LSYVGQPTNIGVGYO  
---Y-AGS---EGFN-Y-LSYVGQPTNIGVGYO  
---Y-AGS---EGFN-Y-LSYVGQPTNIGVGYO

# Example 1: RBD (SARS-Cov-1 | SARS-Cov-2) x ACE2

PDB 2ajf:

- ▶ Cyan: ACE2
- ▶ red: RBD



MISA SSE for MISA-ID SARS-CoV-1-RBD\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
 Bend - Hydrogen bonded turn - Other - Missing Residue -

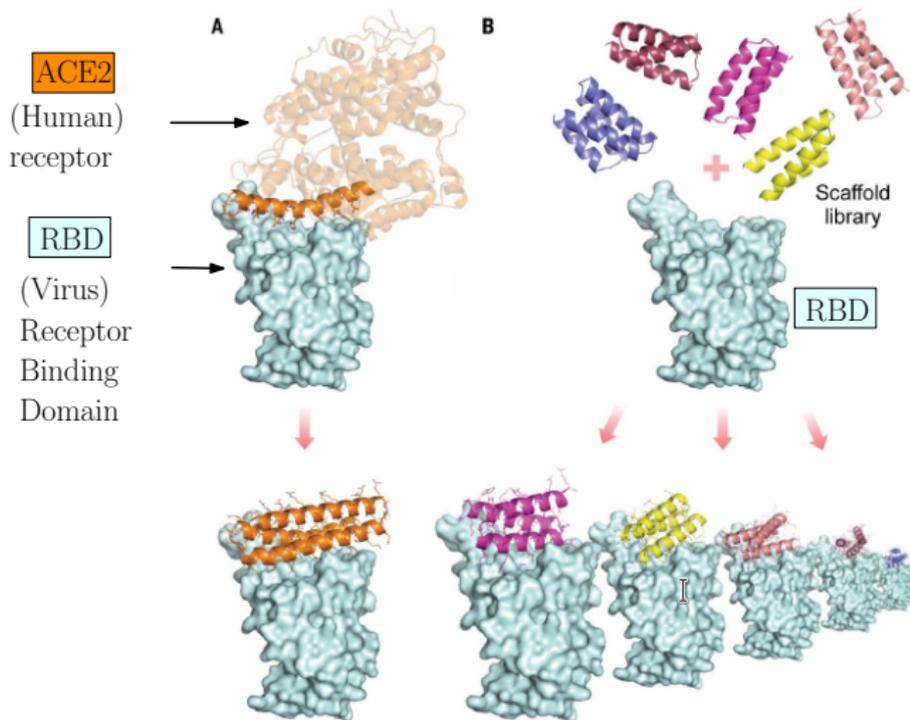
Residue Index	390	400	410	420	430	440	450	460	470	480	490
bound-2ajf-E, res :2.9 Å, 29 interf res	K	-D--Q	--VI--Y	-----R	---S--Y	--Y-YL	-----F	-PD	---P-LNCY	--NDYG	-YTTTGI-YQ
bound-2ajf-F, res :2.9 Å, 29 interf res	K	-D--Q	--VI--Y	-----R	---S--Y	--Y-YL	-----F	-PD	---P-LNCY	--NDYG	-YTTTGI-YQ
unbound-closed-5x58-A, res :3.2 Å, 0 interf res	K	-D--Q	--VI--Y	-----R	---S--Y	--Y-YL	-----F	-PD	---P-LNCY	--NDYG	-YTTTGI-YQ
unbound-closed-6crz-C, res :3.3 Å, 0 interf res	K	-D--Q	--VI--Y	-----R	---S--Y	--Y-YL	-----F	-PD	---P-LNCY	--NDYG	-YTTTGI-YQ

MISA SSE for MISA-ID SARS-CoV-2-RBD\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
 Bend - Hydrogen bonded turn - Other - Missing Residue -

Residue Index	403	410	420	430	440	450	460	470	480	490	500	
bound-6m0j-E, res :2.45 Å, 27 interf res	R	-DE	-----KI--Y	-----H	---VGG	-Y--Y-LF	-----Y	-AGS	-----EGFN	-YF-LQSYGFQPTN	GVGYQ	
bound-6l2g-B, res :2.5 Å, 39 interf res	R	-DE	-----KI--Y	-----H	---VGG	-Y--Y-LF	-----Y	-AGS	-----EGFN	-YF-LQSYGFQPTN	GVGYQ	
unbound-closed-6vxx-A, res :2.8 Å, 0 interf res	R	-DE	-----KI--Y	-----H	---**G	-Y--Y**	-----	---	***	****	YF-LQSYGFQPTN	GVGYQ
unbound-closed-6vyb-A, res :3.2 Å, 0 interf res	R	-DE	-----KI--Y	-----H	---**	-Y--Y-LF	-----	---	***	****	*F-LQSYGFQPTN	GVGYQ

## Example 2: blocking the RBD using designed miniproteins



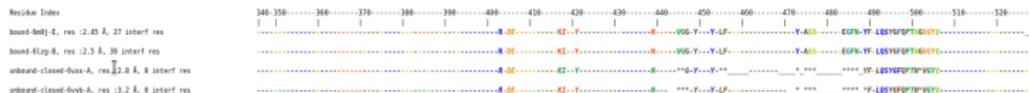
▷Ref: Baker et al., Science 370, 2020

# Example 2: MISAs involving the RBD + { ACE2, antibodies, miniblockers}

## ▷ Footprint on the RBD of various molecules: ACE2, antibody, minibinders

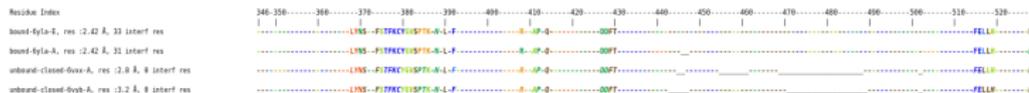
MISA SSE for MISA-10 SARS-CoV-2-RBD-bound-to-ACE2\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
Bond - Hydrogen bonded turn - Other - Missing Residue -



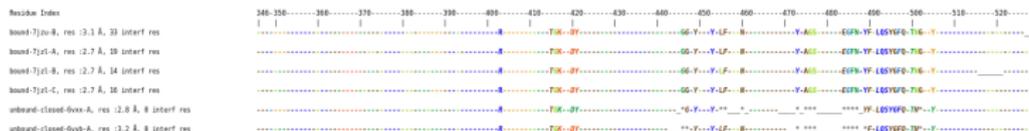
MISA SSE for MISA-10 SARS-CoV-2-RBD-bound-to-00020\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
Bond - Hydrogen bonded turn - Other - Missing Residue -



MISA SSE for MISA-10 SARS-CoV-2-RBD-bound-to-L0B1\_0

3-turn helix - 4-turn helix - 5-turn helix - Isolated beta-bridge residue - Extended strand -  
Bond - Hydrogen bonded turn - Other - Missing Residue -



▷ Ref: Baker al al, Science 370, 2020

# MISAs

Context and motivation

**MISAs: construction**

Software

# MISA construction: overview

▷ **MISA:** Multiple Interface String Alignments

- ▶ Step 1: for each polypeptide chain: compute a so-called *interface string*
- ▶ Step 2: perform the proper coloring

▷ **Practically:** input consists of crystal structures + chains of interest

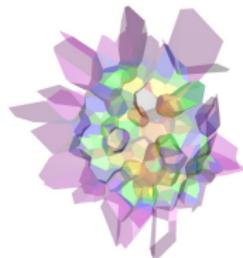
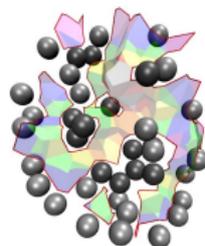
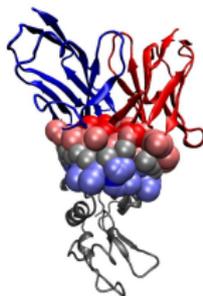
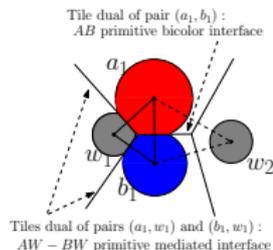
```
# Windows for ACE2-bound-to-SARS-CoV-1
[ACE2-bound-to-SARS-CoV-1_0 (19, 83) (321,393)]

./pdb/2ajf.pdb (A, E, SARS-CoV-1-RBD, bound) (B, A, ACE2-bound-to-SARS-CoV-1, bound)
./pdb/2ajf.pdb (A, F, SARS-CoV-1-RBD, bound) (B, B, ACE2-bound-to-SARS-CoV-1, bound)
./pdb/5x58.pdb (A, A, SARS-CoV-1-RBD, unbound-closed)
./pdb/6crz.pdb (A, C, SARS-CoV-1-RBD, unbound-closed)

# Specification for SARS-CoV-2
./pdb/6m0j.pdb (C, E, SARS-CoV-2-RBD, bound) (D, A, ACE2-bound-to-SARS-CoV-2, bound)
./pdb/6lzg.pdb (C, B, SARS-CoV-2-RBD, bound) (D, A, ACE2-bound-to-SARS-CoV-2, bound)
./pdb/6vxx.pdb (C, A, SARS-CoV-2-RBD, unbound-closed)
./pdb/6vyb.pdb (C, A, SARS-CoV-2-RBD, unbound-closed)
```

# Support, on a per complex basis: the Voronoi interface

## ▷ Voronoi interface



## ▷ Interface string of a chain, derived from the Voronoi interface:

-----N-HNY-----YY-TT--D-----FWST-R-----

## ▷ Consensus interface of a set of complexes : { consensus residues (CR) }

- ▶ Align all interface strings
- ▶ Consensus residue (CR) at a given position: most frequent one in all complexes studied

▷Ref: Lorient, Cazals; Bioinformatics, 2010

# MISA ids and MISA strings

▷ **Given:** a collection of complexes  $\mathcal{C} = \{C_i\}$ , and optionally a collection of unbound structures  $\mathcal{U} = \{U_j\}$ .

▷ **Def.:** **MISA id**, illustrated on IG-Ag complexes

$C_1 = (\{H, L\}, \{A\})$ ,  $C_2 = (\{M, N\}, \{B\})$

▶ Three polypeptide chains: heavy chain, light chain, antigen.

▶ MISA ids: heavy chain: IG\_0; light chain: IG\_1; antigen: Ag\_0.

▷ **Def.:** **interface string of chain**: the string with one character per amino acid, defined as follows:

1. Residue not part of the consensus interface :

▶ Displayed with a dash "-" if it is part of the crystal structure, and underscore "\_" otherwise.

2. Residue part of the consensus interface:

▶ a.a. not in crystal: '\*'.

▶ a.a. at the interface for this chain: uppercase one letter code if  $\equiv$  CR; lowercase letter otherwise.

▶ a.a. not found at the interface for this particular chain, even though the corresponding position contributes to the consensus interface (for other chains): displayed in an italicized uppercase/lowercase letter.

## One specific coloring: MISA/ $\Delta$ ASA

- ▶ **MISA/BSA:** uses the geometry of the bound structure only  $\Rightarrow$  oblivious to conformational changes which may be at play in case of induced fit or conformer selection.
- ▶ **Remedy:**  $\Delta$  ASA coloring scheme.
  - ▶ Consider an interface partner  $P(= A \text{ or } B)$ , in the complex, and consider the  $i$ -th residue of one of its chains.
  - ▶ Let  $ASA_i^b$  be the ASA of this  $i$ -th residue in the structure involving only the chains defining partner  $P$ .
  - ▶ Let  $\overline{ASA}_i^u$  the average ASA of the  $i$ -th residue in unbound structures with the same MISA id.
  - ▶ Displayed for the  $i$ -th residue of a bound structure:  $ASA_i^b - \overline{ASA}_i^u$ .

# MISAs

Context and motivation

MISAs: construction

Software

# Availability in the SBL

- ▷ **SBL package:** python scripts + jupyter notebook
- ▷ **Four scripts:**
  - ▶ [https://sbl.inria.fr/doc/Multiple\\_interface\\_string\\_alignment-user-manual.html](https://sbl.inria.fr/doc/Multiple_interface_string_alignment-user-manual.html)
  - ▶ `sbl-misa.py`: building MISAs from a description of complexes and possibly unbound structures.
  - ▶ `sbl-misa-mix.py`: mixing selected colored MISAs into a single (html) file.
  - ▶ `sbl-misa-bsa.py`: displaying the BSA of all (or selected user defined) residues.
  - ▶ `sbl-misa-diff.py`: comparing i-strings and associated properties (in particular BSA) of two interfaces.
- ▷ **Preprint in biorxiv:**  
<https://www.biorxiv.org/content/10.1101/2020.09.03.281600v1>