Multiple Interface String Alignments
Boosting the analysis of protein interfaces

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...
  - \( \Delta G^0_a = RT \log c^0 K_a = RT \log K_d/c^0 \)
  - \( \tau = 1/K_{off} \)

- Ad-hoc analysis based on various proxys
  - Harrison et al, Science 309, 2005: complex SARS-Cov-1 x various ACE2
  - Wilson et al, Science 368, 2020: SARS-Cov-1 x ACE2 vs SARS-Cov-2 x ACE2

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

PDB 2ajf:
- Cyan: ACE2
- red: RBD
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

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

```plaintext
# 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: Loriot, Cazals; Bioinformatics, 2010
MISA ids and MISA strings

▶ Given: a collection of complexes \( C = \{C_i\} \), and optionally a collection of unbound structures \( U = \{U_i\} \).

▶ 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 \( \equiv \text{CR} \); 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/Δ 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**: Δ ASA coloring scheme.
  - Consider an interface partner $P(=A$ or $B)$, in the complex, and consider the $i$-th residue of one of its chains.
  - Let $ASA^b_i$ be the ASA of this $i$-th residue in the structure involving only the chains defining partner $P$.
  - Let $\overline{ASA}^u_i$ 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: $ASA^b_i - \overline{ASA}^u_i$. 
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](https://www.biorxiv.org/content/10.1101/2020.09.03.281600v1)