

## **Accurate representation of Protein-Ligand structural diversity in the PDB**

Nicolas K. Shinada <sup>1</sup>, Peter Schmidtke<sup>2</sup> and Alexandre G. de Brevern <sup>3,\*</sup>

<sup>1</sup> SBX Corp., Tokyo-to, Shinagawa-ku, Tokyo, Japan; Discngine SAS, Paris France ;  
INSERM, UMR\_S 1134, DSIMB, Univ Paris, INTS, Laboratoire d'Excellence GR-Ex,  
Paris, France; shinada@sbx-corp.com

<sup>2</sup> Discngine SAS, Paris France; peter.schmidtke@discngine.com

<sup>3</sup> INSERM, UMR\_S 1134, DSIMB, Univ Paris, INTS, Laboratoire d'Excellence GR-Ex,  
Paris, France; alexandre.debrevern@univ-paris-diderot.fr

## Appendix A

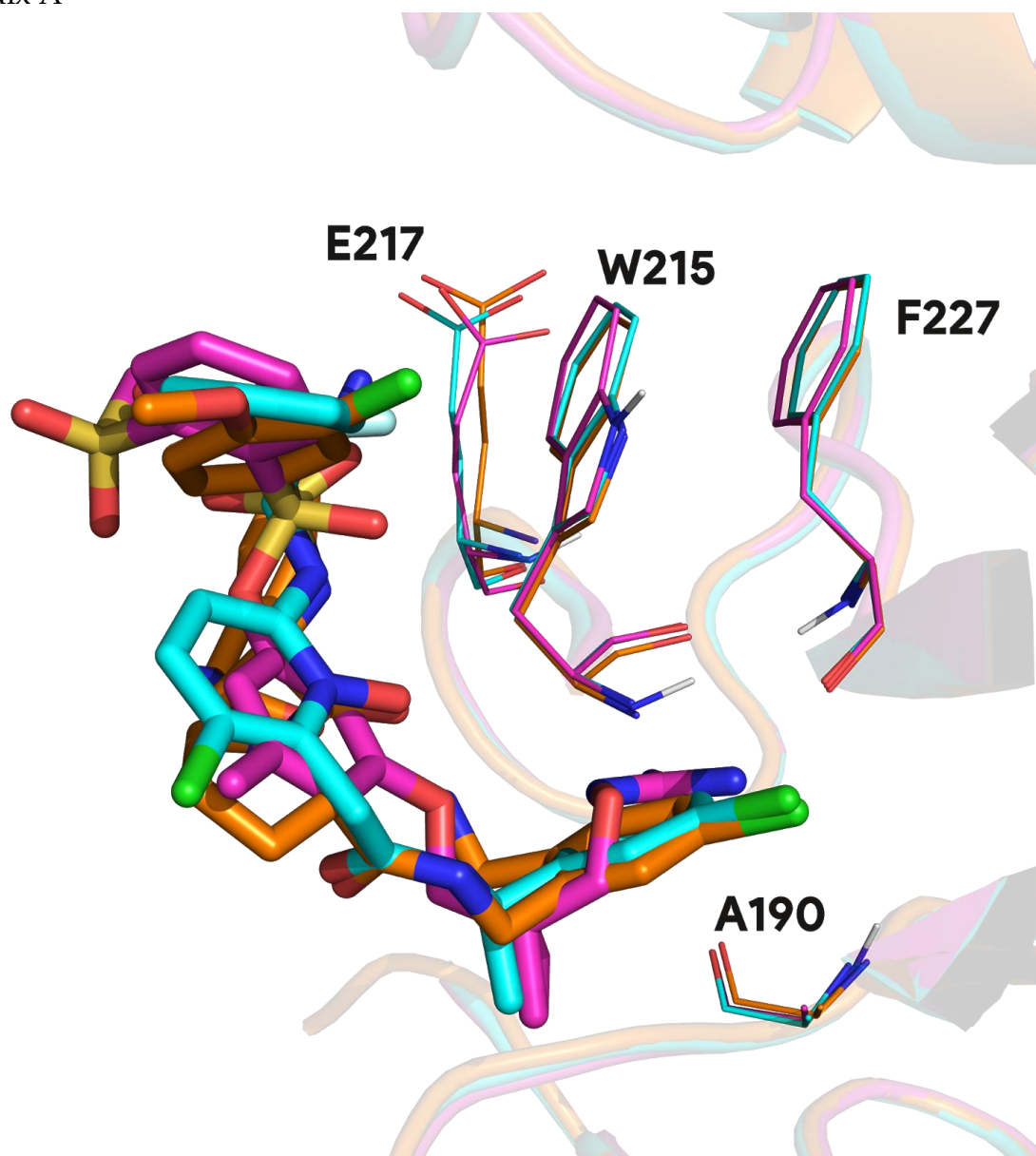


Figure A1. 3D visualization of conserved residues (lines) in the prothrombin ligand (sticks) binding process (PDB ids 1t4u in pink, 1z71 in blue and 3u9a in orange).

The complete list of complexes can be found in the file `PDB_prot-lig_representatives.txt` that is located at DSIMB website ([https://www.dsimb.inserm.fr/~debrevn/PDB\\_prot-lig\\_representatives/](https://www.dsimb.inserm.fr/~debrevn/PDB_prot-lig_representatives/)) and at Discngine website (<https://www.discngine.com>).

Please also notice that a free trial version and a REST API are available ([https://github.com/Discngine/api\\_python\\_3decision](https://github.com/Discngine/api_python_3decision)). The end-goal is to empower the user by allowing him/her to create dynamically a dataset, whatever the scale, with a wide range of metrics (B-factor, ligand size) and comparison tools (sequence, molecular, structural similarity). Thus, every group will be able to compare and easily confront the dataset they used based on the parameters chosen.