# Tools integration to address the importance of water molecules in protein-ligand complexes

Claudia Beato, Iuni Margaret Laura Trist, Andrea Cristiani, Aldo Feriani, Alfonso Pozzan

Research Informatics – Computational Chemistry, Integrated Drug Discovery, Aptuit, an Evotec Company – Via A. Fleming 4, 37135 Verona (Italy)

## Introduction

*In vivo*, macro- and small molecules are surrounded by an aqueous milieu. Water participates in virtually all biomolecular processes and plays a crucial role in the formation of complexes, such as enzyme-substrate or protein-inhibitor complexes. During these events, solvent water molecules need to be rearranged, impacting on the thermodynamics of the binding process.[1]

Water molecules have a key role in trypsin-like Serine Proteases protein family, in which a water molecule located in the proximity of the catalytic triad takes part in the last step of the catalytic machinery, promoting the release of the acylated product and the regeneration of the active site. In addition to this, individual water molecules are located in the S1 pocket and can be displaced upon ligand binding: depending on ligand classes, different water molecules can be displaced determining different inhibition and selectivity profiles (Figure 1A).[2] Also in ATP pocket of protein kinases there are differences in water networks depending on the kinase system. Displacement of or interaction with a specific water molecule are strategies that have been successfully applied to improve activity and selectivity of ligand series towards a specific kinase (Figure 1B).[3] Higgs C. et al have also investigated the role of water molecules in GPCR, revealing that proper understating of the hydration state of these receptors can be important to understand the ligand SAR trends.[4]

In the development of novel compounds, carefully addressing the impact of the new ligand on the pre-existing water network is important for improving affinity and also selectivity against similar proteins. 3D-RISM is useful to understand the hydration state of a complex, but does not allow to characterize the water reorganization upon side chain movements induced by ligand binding. Therefore we developed a workflow combining MOE tools[5] with NAMD MD package[6], AmberTools[7] and python scripts to analyse the reorganization of the water network around the ligand and protein-ligand interaction energy.

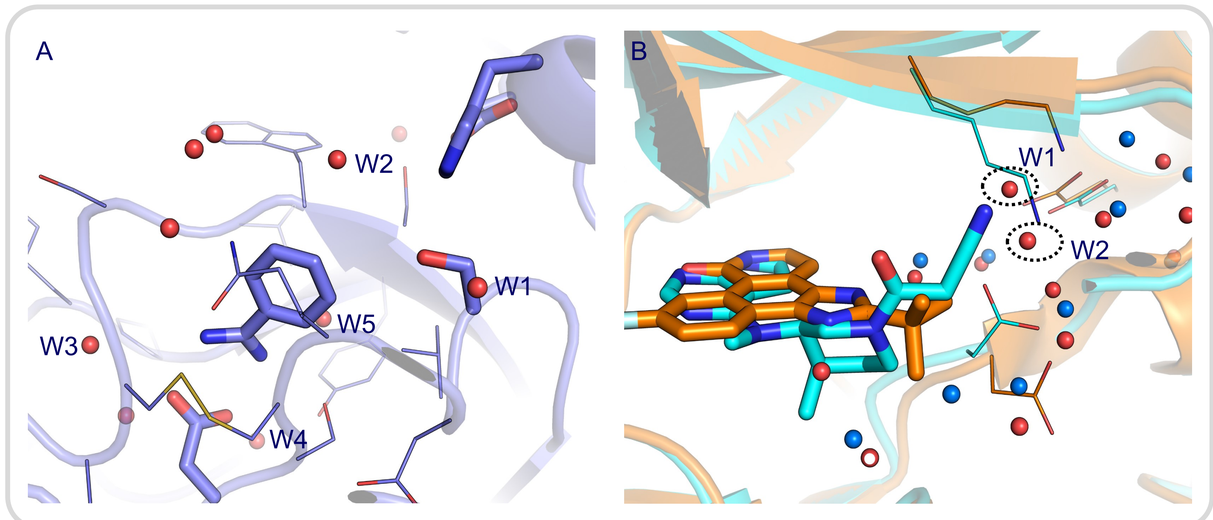
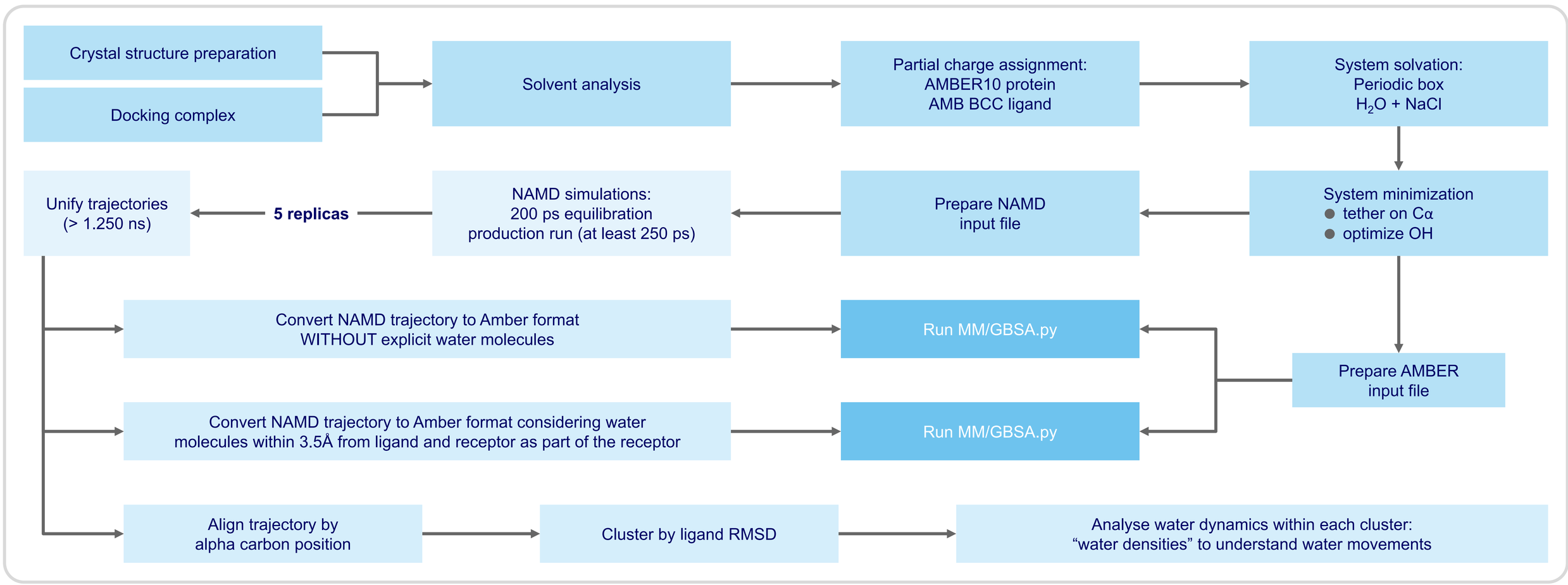



Figure 1: A) bovine trypsin crystal structure 5MNG in complex with benzamidine. The high resolution of the structure allows to clearly identify water molecules present in serine protease binding site. W1 and W2 are close to the active site residues and can take part in the catalytic process. W3-5 are water molecules present in S1 pocket, that in some proteases can be displaced by inhibitors, contributing to increase the selectivity of the series against other proteases. B) Tofacitinib, a selective JAK3 inhibitor, in complex with JAK3 (PDB ID:3LXK, cyan color, ligand in sticks, water molecules represented by blue spheres) superposed to CP-6, an unselective JAK1,2,3 inhibitor, in complex with JAK1 (PDB ID: 3LXL, orange color, ligand in sticks, water molecules as red spheres). The high selectivity displayed by tofacitinib against other members of the JAK family was attributed to the ability to displace two water molecules, W1 and W2 highlighted in the picture, and interact with residues of the glycine-rich loop in JAk3. This is not happening in JAK1 and JAK2, because of different side chains and water network organization.




## Workflow details

The different colors in the workflow represent the different tools used in the specific steps.

- System preparation: preparation of the crystal structure or docking complex using MOE (GUI and SVL scripts). Solvent analysis of the complex is performed with 3D-RISM to define the best possible starting point for the binding site solvation. AMBER10 force field is used for assigning the partial charges to the protein. AM1 BCC partial charges are assigned to ligand. The protein-ligand complex with key water molecules in the binding site is then fully solvated in a cubic box of 12 Å from the closest atom (TIP3 water) with periodic boundary conditions. Na+ and Cl– ions are added mimicking their physiological concentration. The solvated system is then minimized, enabling optimizeOH option in MOE. Input files for NAMD and AMBER are created with MOE MD interface.
- MD simulation: 5 replicas of short NAMD runs are performed using Amber force field. Optimal simulation length and number of replicas are addressed at the beginning of the project. Cα tether are applied during the simulations to minimize the background noise due to backbone movements, since we are only interested in ligand, side chains and water movements. The five trajectories of the production runs are merged into a single trajectory file used for determining the binding energy and the to analyse water movements.
- Binding energy: estimation of the binding energy is performed using mmpbsa.py script provided within AmberTools for all the trajectory frames. Preliminary conversion of the NAMD dcd trajectory to Amber file format is performed with ad-hoc python scripts. In particular to include specific water molecules in the MM/GBSA calculation as part of the receptor, the trajectory is evaluated frame by frame: for each frame water molecules within 3.5Å distance from ligand and receptor (generally less than 10) are retained in the complex, while others are deleted; then mmpbsa.py script is run. For comparison mmpbsa.py script is also run for the trajectory without considering explicit water molecules.
- Analysis of water movements during the MD is performed using ad-hoc python scripts, importing the freely available MDAnalysis library tools. Trajectory is clustered by ligand RMSD (generally with a cuttoff <0.5Å). For each cluster, the hydration state of the protein-ligand binding site is analysed, therefore water molecules within a radius of 8Å from the ligand are dynamically selected frame by frame to derive the solvent density. The centroid is used as reference for the other frames of the trajectory cluster. The positions of all water oxygens are histogrammed on a grid (grid spacing = 1Å). Density is then converted from $Å^{-3}$ to the density relative to the literature value of the TIP3P water model at ambient conditions. Finally, the density is written as an OpenDX compatible file that can be read in PyMOL (Figure 2).

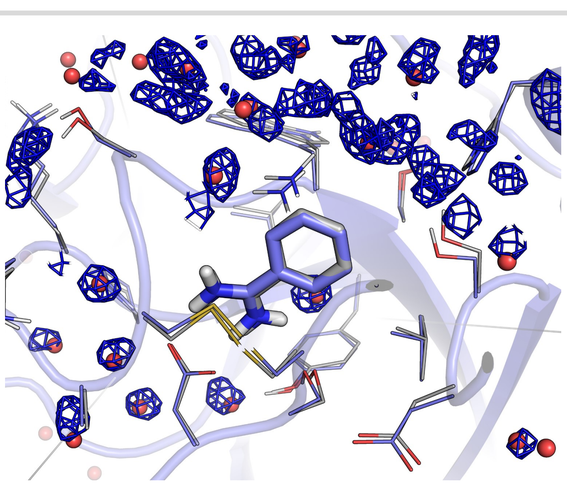


Figure 2: Representative view of the water densities obtain from MD clusters, summarizing water movements during the simulation (5MNG crystal structure).

## References

1. Schiebel J. et al, Nat Commun 2018, 9(1), 3559
2. a) Steiner J.L.R. et al, J Am Chem Soc 1998, 120 (3), 597;
   b) Rühmann E et al, J Med Chem. 2015, 58(17), 6960;
   c) Katz BA, Chem Biol. 2001, 8(11), 1107
3. a) Chrencik JE et al, J Mol Biol 2010, 400, 413;
   b) Robison DD et al, ChemMedChem. 2010, 5(4), 618
4. Higgs C. et al, ACS Med Chem Lett 2010, 1, 160
5. Molecular Operating Environment (MOE), 2018.01;
   Chemical Computing Group ULC, Montreal, QC, Canada
6. Phillips J.C. Et al, Journal of Computational Chemistry
   2005, 26, 1781
7. Case D.A. et al, AMBER 2018,
   University of California, San Francisco

## Conclusions and future work

This workflow is currently being applied to prioritize compounds for synthesis, by predicting binding affinity and analyzing the stability of water mediated hydrogen bonding network.

- The analysis of the water densities in the different clusters helps to identify water molecules involved in stable hydrogen bonding network and to highlight water molecules reorganization due to ligand movements
- MM/GBSA ligand binding energy is estimated considering water molecules close to ligand and binding site as part of the protein, slightly improving the correlation with experimental values compared to MM/GBSA prediction without considering water molecules
- Future work will focus on the ligand binding energy prediction evaluating other approaches, like FMO-DFTB, and more rigorous estimation of hydration free energy