

Subtract v1.0: Measuring the volume of protein binding site cavities

Panos Kakoulidis, Ioannis Galdadas and Zoe Cournia

Biomedical Research Foundation, 4 Soranou Ephessiou, 11527, Athens, Greece

Researchers engaged in computer-aided drug design often require an accurate measurement of the volume of a ligand-binding pocket in order to evaluate pharmacology. The volume of a binding site is of great pharmacological significance, both as a structural feature building its pharmacophoric characteristics and as one of the characterizations used in Quantitative Structure-Activity Relationships (QSAR). When the volume of a binding site is known, potential ligands that are too large to fit in that volume can be eliminated early in the drug discovery process, prior to virtual or high-throughput screening. Moreover, the volume of a binding site is also connected with its viability, giving valuable information to researchers as to whether it can be utilized as target for a small organic molecule that can act as a candidate drug. Moreover, variations in pocket size, when multiple structures of the same protein are considered, can provide pharmacologically useful insights into protein dynamics. Finally, oncogenic mutations in proteins often have an effect in binding sites, which needs to be evaluated accurately in order to draw conclusions of what is the mutant effect on normal (wild type) proteins.

Subtract is a tool that has been created to solve the problem of accurate measurement of the protein binding sites, and works both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories.

Subtract accepts an atom selection in the form of a PDB file and computes the three-dimensional convex hull of the atoms points with the help of SciPy library. The next step of the algorithm is to compute the volume of the convex hull and the volume of the atoms that are included in the solid based on their van der Waals radii. The subtraction of those two volumes yields the volume of the investigated cavity. The algorithm computes cavity volumes of trajectory frames in parallel for maximum efficiency and speed. It requires minimal usage of memory due to the fact that it follows a buffering strategy of reading file chunks and therefore there is no need to load the entire file into memory. There is a wide support of trajectory formats like Gromacs trajectory files and multi-model PDB files due to its dependency to the MDTraj library.

This tool enables access to the users to be able to calculate the volume of binding sites of proteins of interest in a fast and efficient way. Subtract is a tool that has been created to solve this problem, both for crystal structures downloaded from the protein data bank and for trajectories generated as Molecular Dynamics simulations.

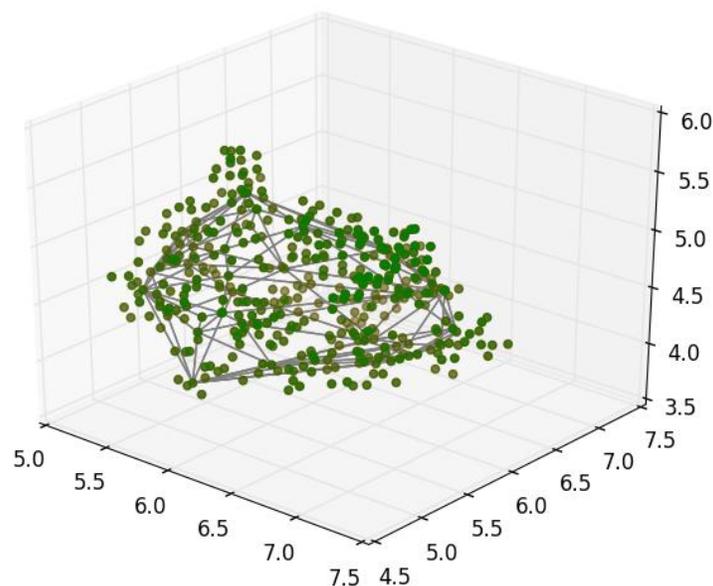


Figure 1 : The convex hull that includes the cavity points

Input Requirements

- The tool processes topology (.pdb) along with trajectory (.xtc) files
- Maximum trajectory filesize : 2 GB
- Trajectory must be fitted before it is provided to Subtract.
- Currently only Gromacs format is supported (.xtc)

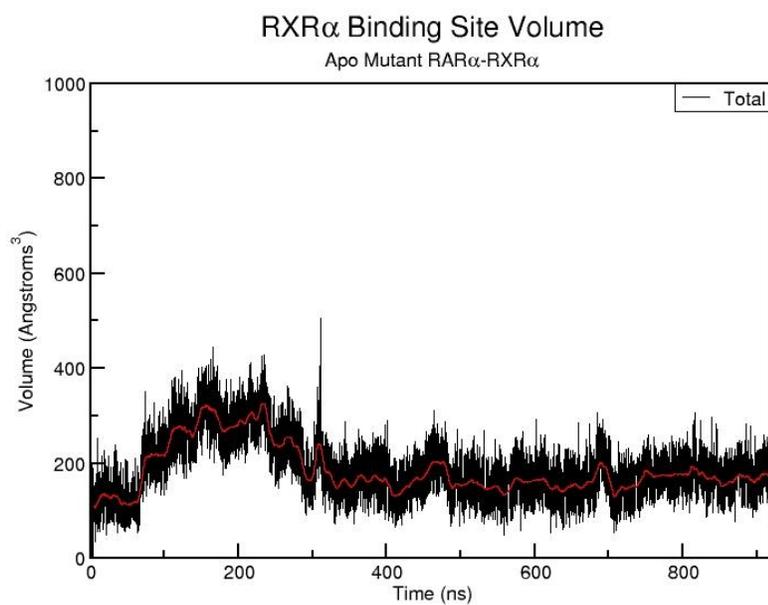
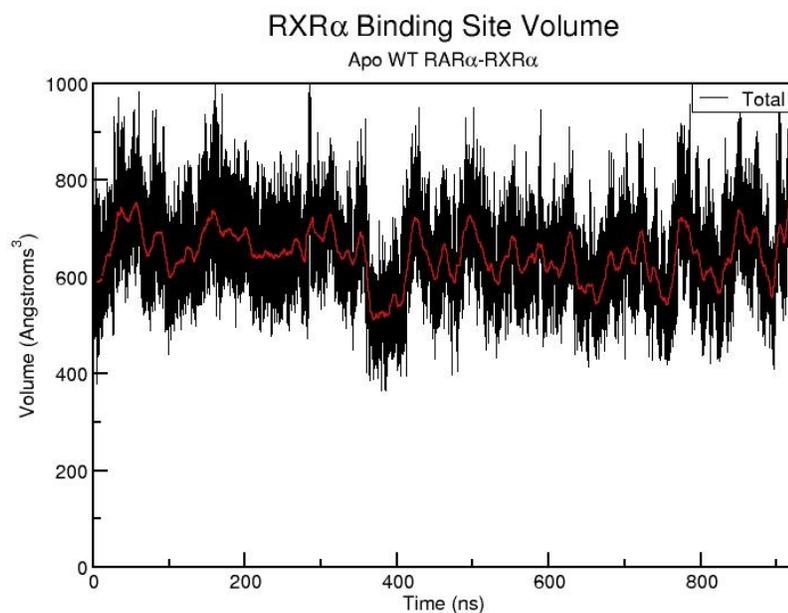
• The cavity is determined by the user in the following format:

<chain numerical index>:<residue number>:: <next chain numerical index>:<residue number>::... e.g."1:110,220,34,678,800::2:345,110,23,67"

The chain is specified by its ascending position in the PDB file. Residue numbers are comma separated. They are stated after the chain index and a colon separator. The information for each chain is separated by double colons (if there is more than one chain participating in the cavity formation).

Therefore, the input consists of a topology file, a trajectory file and a cavity specification by indices. The submission of the input creates an entry in the queuing system of the server. Depending on the available resources and the server load, the request will be processed in an asynchronous manner. The user should return at a later time and check the status of her request. When the volume calculation is completed, a download link of an .xvg file, which contains the result, becomes available. [Note: the input is **not** maintained in the server.]

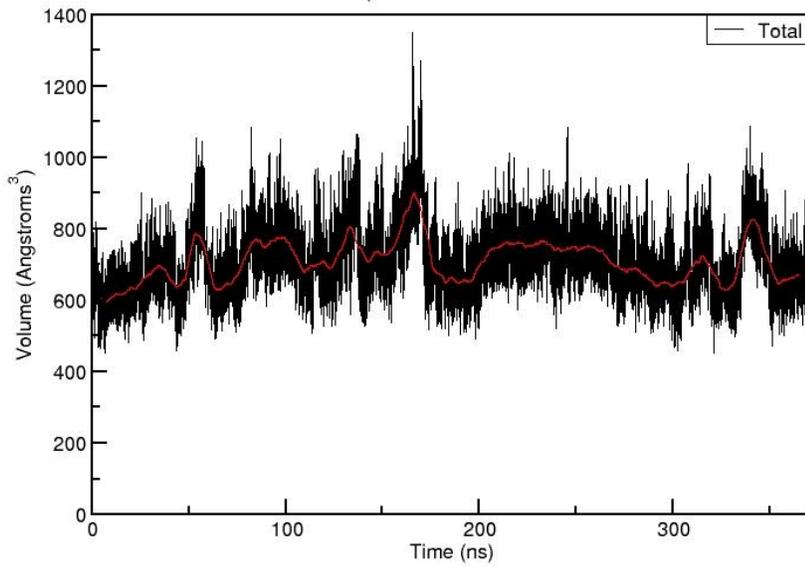
Example Application of Subtract output on the RXRa-RAR Molecular Dynamics Simulation



	Average (angstroms ³)	Standard Deviation (angstroms ³)
Apo WT	646.8466	80.68737
Apo Mutant	185.9875	58.77739

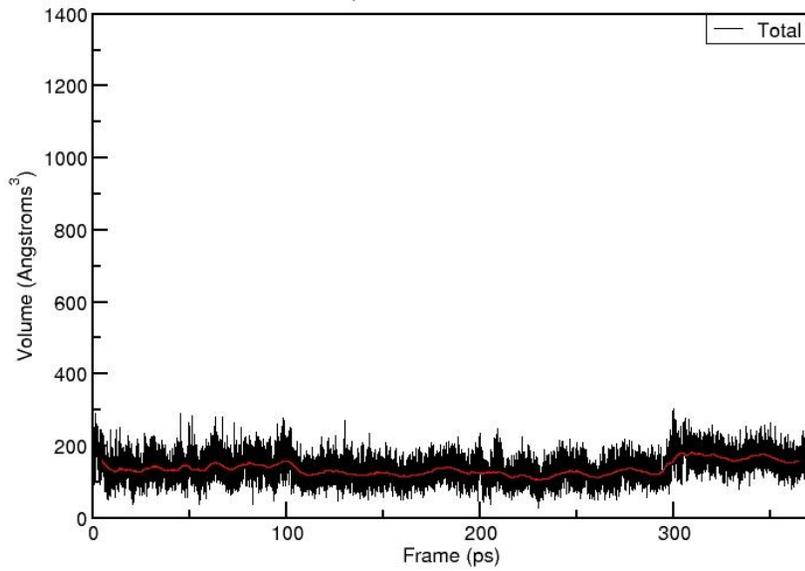
RXR α Binding Site Volume (Replica 2)

Apo WT RAR α -RXR α



RXR α Binding Site Volume (Replica 2)

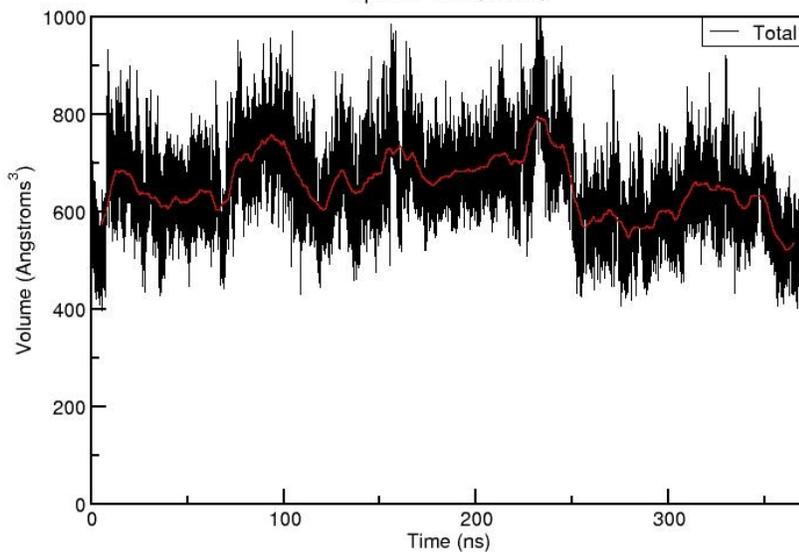
Apo Mutant RAR α -RXR α



	Average (angstroms ³)	Standard Deviation (angstroms ³)
Apo WT (Rep. 2)	708.4732	94.18520
Apo Mutant (Rep. 2)	138.3553	34.19066

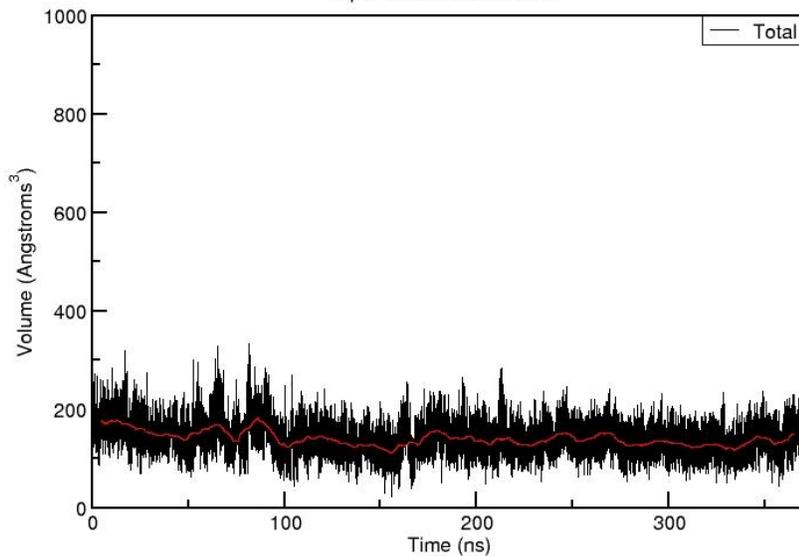
RXR α Binding Site Volume (Replica 3)

Apo WT RAR α -RXR α



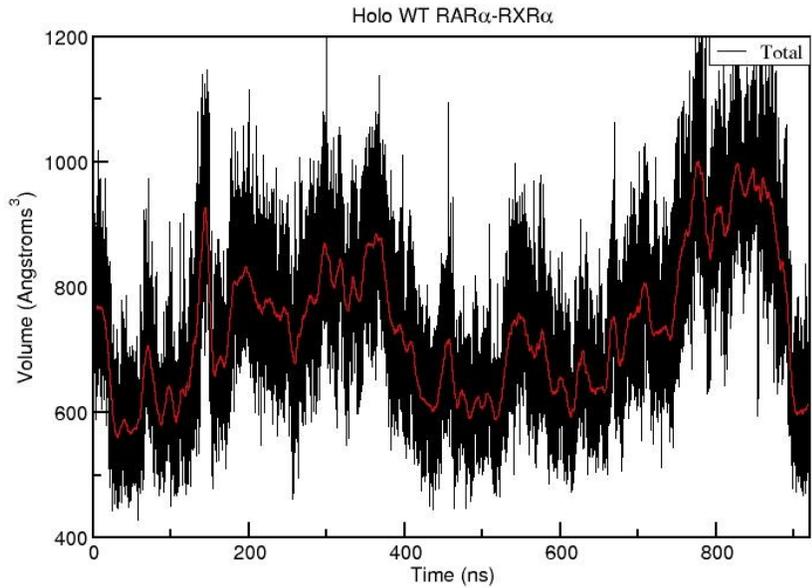
RXR α Binding Site Volume (Replica 3)

Apo Mutant RXR α -RAR α

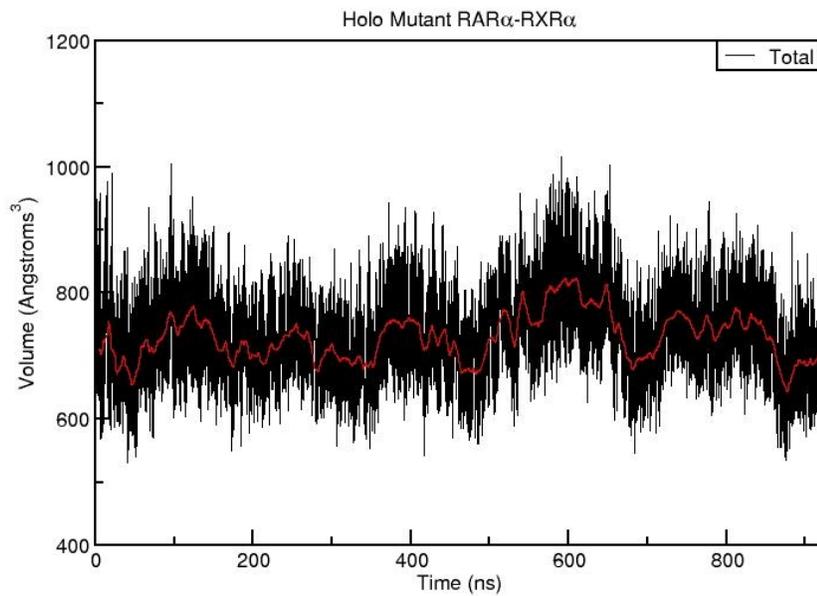


	Average (angstroms ³)	Standard Deviation (angstroms ³)
Apo WT (Rep. 3)	651.6106	87.28659
Apo Mutant (Rep. 3)	141.6229	32.75786

RXR α Binding Site Volume



RXR α Binding Site Volume



	Average (angstroms ³)	Standard Deviation (angstroms ³)
Holo WT	738.3533	129.8462
Holo Mutant	729.9974	59.08058