OXFORD

Structural bioinformatics

AQUA-DUCT: a ligands tracking tool

Tomasz Magdziarz[1], Karolina Mitusińska[1,2], Sandra Gołdowska[1,2], Alicja Płuciennik[1,2], Michał Stolarczyk[1,2], Magdalena Ługowska[1,2] and Artur Góra[1,*]

[1]Tunneling Group, Biotechnology Centre, Silesian University of Technology, Krzywoustego 8, 44-100 Gliwice, Poland and [2]Institute of Automatic Control, Silesian University of Technology, Akademicka 16, Gliwice 44-100, Poland

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Abstract

Motivation: The identification and tracking of molecules which enter active site cavity requires screening the positions of thousands of single molecules along several thousand molecular dynamic steps. To fill the existing gap between tools searching for tunnels and pathways and advanced tools employed for accelerated water flux investigations, we have developed AQUA-DUCT.

Results: AQUA-DUCT is an easy-to-use tool that facilitates analysis of the behaviour of molecules that penetrate any selected region in a protein. It can be used for any type of molecules, e.g. water, oxygen, carbon dioxide, organic solvents, ions.

Availability and Implementation: Linux, Windows, macOS, OpenBSD, http://www.aquaduct.pl.

Contact: a.gora@tunnelinggroup.pl or info@aquaduct.pl

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Enzymes provide their service immersed in a solvent that contributes to catalyst stability, activity and selectivity (Hendil-Forssell *et al.*, 2015; Richard *et al.*, 2014). In enzymes with a buried active site and connected to the surrounding solvent by tunnels, the solvent flux can be controlled much more precisely by the molecular properties of amino acids constituting tunnels, or in more sophisticated enzymes by gates controlling the opening and closing of the access pathways (Gora *et al.*, 2013). Information regarding the exchange of solvent molecules between the protein core and a bulky solvent is embodied in the results of molecular dynamic (MD) simulations. However, due to the requirements of the simultaneous tracking of many thousand interactions, the results for particular molecules are difficult to access. To the best of our knowledge, there is no universal software that could facilitate this type of analysis. Existing methods provide only visual information regarding water traffic (Vassiliev *et al.*, 2010) or use an advanced approach to direct the water flow in a particular direction (Benson and Pleiss, 2014). Other existing tools can provide information about the geometry of possible tunnels, however, since there are overlooked chemical properties of amino acids, the detected pathways do not have

to correspond to their actual usage (Brezovsky *et al.*, 2013, 2016). AQUA-DUCT (AQ) is filling this gap and facilitates the analysis of the movement of any small ligand(s) during MD simulations.

2 How does it work?

AQ is a tool for tracing, analyzing and visualizing molecular trajectories throughout MD simulations. Calculations in AQ are divided into 6 stages, which include statistical analyses and subsequent visualization (Supplementary Fig. S1). During the first stage, a list of all the molecules to trace is created. This is performed by screening the entire trajectory for molecules that enter a user-defined region of interest, the so-called *object* region of the macromolecule. AQ then traces molecules only within the second user-defined region, i.e. the *scope* region which usually encompasses the entire macromolecule (Fig. 1). The second stage of calculations relates the list of all traceable molecules to abstract coordinates at the center of masses for each molecule in all frames of the MD simulation. This result, stored as a list of raw paths, is used in the third stage of calculations to create a list of separate paths. Each path comprises data for one molecule. A molecule may enter and leave the *scope* and the *object* areas many times over an entire MD simulation. Each

event is considered by AQ to be a separate phenomenon and is thus represented by a distinct separate path. However, the ID and number of occurrence are retained and can easily be retrieved.

Each separate path can be divided into three components: incoming, object and outgoing. The incoming component corresponds to the path that runs from the point in which a molecule enters the *scope* to the point in which it enters the *object* for the first time. The outgoing component corresponds to the path that runs from the point at which the molecule finally leaves the *object* to the point that it leaves the *scope* and thus the separate path ends. The remainder of the separate path corresponds to the object component in which a molecule may leave and enter the *object* many times but it never leaves the *scope*.

Each of the separate paths has a beginning and an end. If either or both of these are at the boundaries of the *scope*, they are considered as inlets, i.e. points that mark where the traceable molecules enter or leave the *scope*. The fourth stage of AQ calculations uses inlets to find clusters that are anticipated to mark the endings of tunnels in the macromolecule (Fig. 1). Next, clusters are used to arrange separate paths into groups that correspond to the flux of the examined molecules through the macromolecule inside separated tunnels or between different tunnels. Optionally, paths belonging to the particular transition can be averaged and one mean path is created thereby allowing for convenient visualization and calculation of the general transitional properties.

The last two stages involve statistical analyses (Supplementary Fig. S8) and visualization of the results (Fig. 1).

Input data: The compulsory user input is minimal; by default AQ requires only MD data and definition of the *scope* and *object* areas.

Parsing of the MD data is implemented using the MDAnalysis library (Michaud-Agrawal *et al*., 2011) and therefore AQ uses MDAnalysis Selection Commands for definitions of the *scope* and *object*. Each *object* definition should be comprised of two elements: name(s) of the molecules that will be traced (e.g. WAT for water molecules) and spatial constraints that limit the area of the *object*. The *scope* can be defined in a similar manner, i.e. by providing the name of the molecules and some spatial constraints that should be broad enough to encompass the *object*. Another possibility is to define the *scope* as an interior of a convex hull of some atoms, for example, the macromolecule backbone atoms.
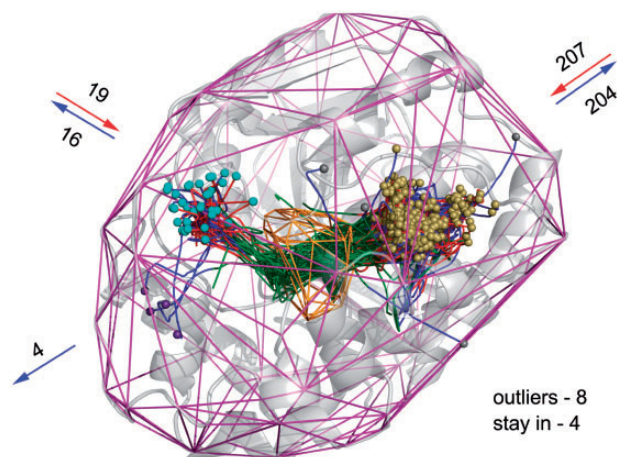

Fig. 1. Summary of AQ results. View of the protein (frame 1) shown in gray. The convex hull of alpha carbon atoms (frame 1) is displayed as a magenta frame. The *object* shape (colored orange) is visualized as a convex hull of atoms in the *object* area. Clusters are displayed as scatterings of sand, cyan, purple and gray spheres. Smooth paths of traced water molecules are displayed as lines. Red, green and blue lines correspond to incoming, object and outgoing parts, respectively. The number of molecules which enter or egress a specific tunnel, as well as those which permanently occupy the active site, are provided. Numbers and arrows were added in post-processing

Statistics: The results of AQ calculations are summarized in a collection of tables and saved into a separate text file (Supplementary Fig. S8). The tables begin with some general data such as the number of traced molecules, number of paths, inlets and clusters. Then, more specific statistics are presented, such as sizes of clusters, mean lengths of paths belonging to particular cluster–cluster transitions, and exact lengths of all paths.

AQ can optionally smooth raw paths. This smoothing can be used for visualization purposes, or can also be effective for statistical calculations. In the latter case, the original (raw) traces are replaced by the smoothed ones to diminish the Brownian character of the paths. This decline of Brownian character depends on the smoothing method employed and consequently impacts the results of statistical analyses.

Visualization: AQ uses PyMOL (PyMOL, Schrödinger (2014)) for visualization purposes and can prepare visualization objects in two different ways. It can either start PyMOL, provide it with visualization data, and save a session file, or it can prepare a visualization script that can be later used to load data into PyMOL. The latter method can be very useful in command line only environments and does not require access to PyMOL or graphic mode.

AQ offers visualization of all found paths including averaged paths generated for transition phenomena. Paths can be visualized in a raw form or after optional smoothing (Fig. 1). Incoming, outgoing and object components of all paths are displayed in different colors. Clusters are visualized as scatters of spheres. It is also possible to display only selected frames of MD simulation and convex hull shapes.

Data analysis: Analysis of statistical data allows the user to gather precise information regarding tunnel network usage, the number of molecules that enter and egress from a specific region of the protein, and information about ligands that traverse a protein i.e. enters and leaves by a different exit (Supplementary Table S3). Analysis of visualization can identify the paths and cavities penetrated by ligands (Supplementary Figs S5 and S9) and key residues that control or block the flow of molecules through the proteins tunnel networks (Supplementary Figs S10–S12).

Other: AQ implements several algorithms for smoothing paths, and it also has a method for trimming paths down to the surface of the macromolecule (Supplementary Figs S6 and S7).

Funding

References

Benson,S.P. and Pleiss,J. (2014) Solvent flux method (SFM): a case study of water access to *Candida antarctica* lipase B. *J. Chem. Theory Comput*., **10**, 5206–5214.

Brezovsky,J. *et al*. (2013) Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnol. Adv*., **31**, 38–49.

Brezovsky,J. *et al*. (2016) Engineering a de novo transport tunnel. *ACS Catal*., **6**, 7597–7610.

Gora,A. *et al*. (2013) Gates of enzymes. *Chem. Rev*., **113**, 5871–5923.

Hendil-Forssell,P. *et al*. (2015) Exploring water as building bricks in enzyme engineering. *Chem. Commun*., **51**, 17221–17224.

Michaud-Agrawal,N. *et al*. (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem*., **32**, 2319–2327.

Richard,J.P. *et al*. (2014) Enzyme architecture: on the importance of being in a protein cage. *Curr. Opin. Chem. Biol*., **21**, 1–10.

The PyMOL Molecular Graphics System, (2014) Version 1.7 Schrödinger, LLC.

Vassiliev,S. *et al*. (2010) Tracking the flow of water through photosystem II using molecular dynamics and streamline tracing. *Biochemistry (Mosc.)*, **49**, 1873–1881.