

# Fpocket : an Open Source Platform for Ligand Pocket Detection

Vincent Le Guilloux<sup>1</sup>, Peter Schmidtke<sup>2</sup> and Pierre Tuffery<sup>3, 4</sup>

<sup>1</sup> Institut de Chimie Organique et Analytique - UMR CNRS 6005, Div. of chemoinformatics and molecular modeling, Univ. Orléans, France

<sup>2</sup> Dpto. Fisicoquímica, Fac. Farmacia, Univ. Barcelona, Barcelona, Spain

<sup>3</sup> Molécules Thérapeutiques in silico, INSERM, UMR-S 936, Univ. Paris Diderot - Paris 7, Paris, France

<sup>4</sup> Ressource Parisienne en Bioinformatique Structurale, Univ. Paris Diderot - Paris 7, Paris, France

Prix SFCi 2009

## CONTEXT and PACKAGE

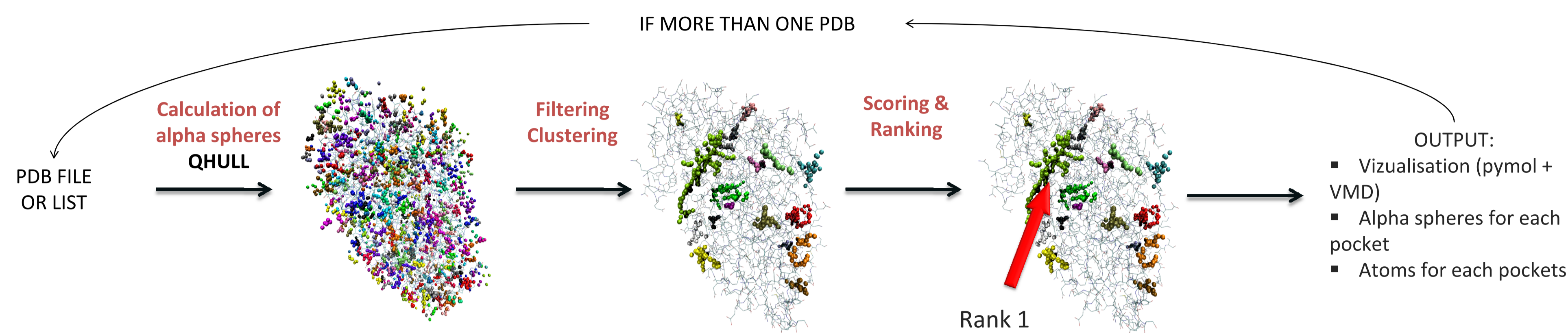
Structure based virtual screening (SBVS) and molecular docking methods start to be well established as effective approaches to identify hits, candidates and leads for drug discovery research. For molecular docking, the identification of candidate pockets in protein structures is a key feature, and the recent years have seen increasing interest in developing methods for pockets and cavities detection on protein surfaces.

Fpocket<sup>1</sup> is a fast, scalable and **open source** pocket detection package based on Voronoï tessellation built on top of the publicly available package Qhull<sup>2</sup>. Three main programs are available:

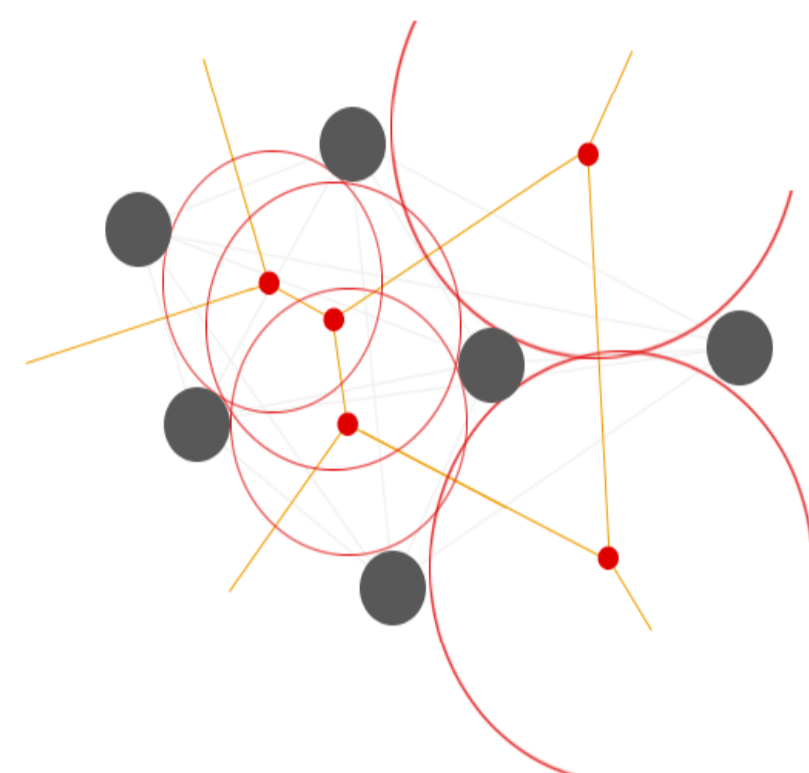
- Fpocket**, to perform pocket identification
- Tpocket** to test fpocket detection performances
- Dpocket** to extract pocket descriptors

Fpocket package can be downloaded for free at <http://sourceforge.net/projects/fpocket/>

## FPOCKET METHODOLOGY



### The alpha sphere concept



- ✓ All spheres that contact exactly 4 vertices (atoms)

### Filtering & Clustering

- Only alpha sphere having a radius  $3\text{\AA} < r < 6\text{\AA}$  are kept
- Clustering (multiple linkage algo.)
- Removal of very small clusters
- Fusion of closed pockets using multiple linkage and barycenter based clustering algorithms
- Removal of small pockets

### Scoring function

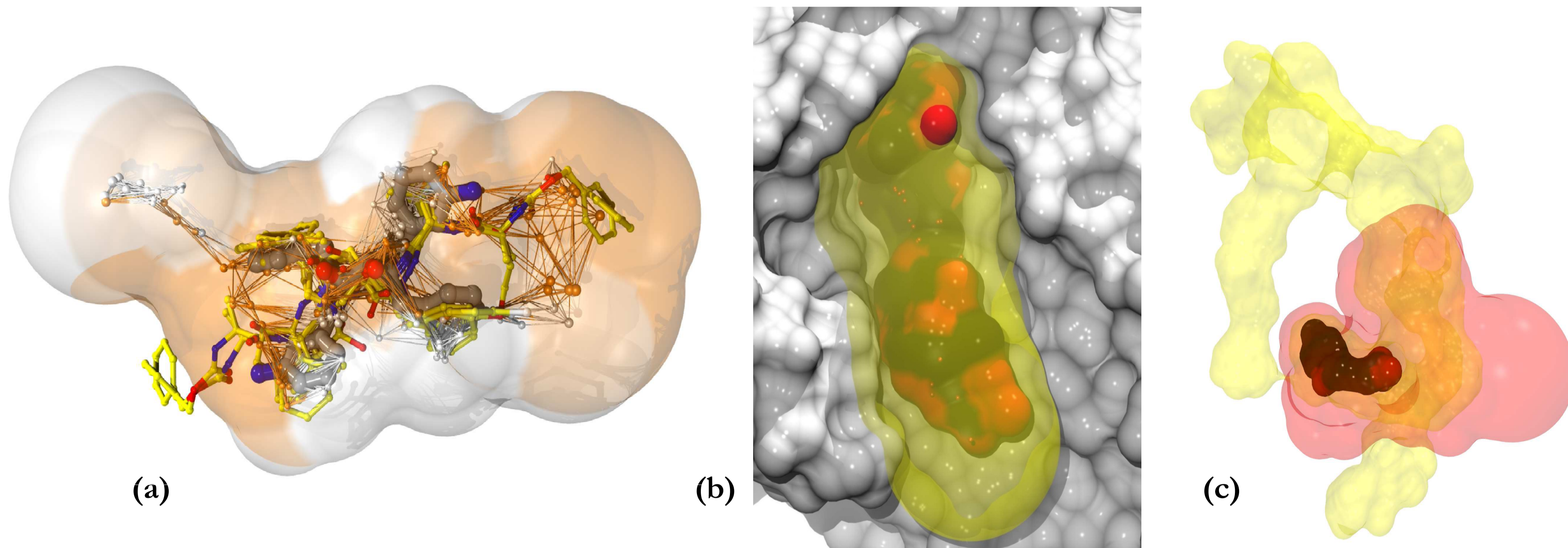
- ✓ Makes use of 5 Dpocket descriptors related to pocket size, alpha spheres density, hydrophobicity & polarity
- ✓ Established using a training set based on a dataset used to evaluate PocketFinder<sup>3</sup>

## SOME EXAMPLES

(a) Rank 1 pocket (transparent halo with connected alpha spheres) of the HIV1 Protease DMP450 complex (grey CPK), with other known inhibitors superposed (yellow).

(b) Rank 1 pocket of the Alpha amylase (yellow transparent halo). Acarbose is represented in black & red.

(c) The Cyclooxygenase-2 indomethacin pocket found by fpocket (red) compared to that found by PocketPicker<sup>4</sup> (yellow)



## EVALUATION

Fpocket has been evaluated on 4 external datasets and compared to other free methods.

This table shows the % of pocket found by fpocket at rank 1 and 3 compared to other methods. In green is highlighted the best score for a given dataset.

### Four external datasets - 192 proteins

- Three Datasets of bound proteins
- One dataset of unbound proteins

### Two evaluation criteria:

- One common evaluation criterion
- One in-house evaluation criterion (in brackets)

Dataset	Algorithm	Rank 1		Rank 3	
		Unbound	Bound	Unbound	Bound
PocketPicker <sup>4</sup> (45 proteins)	Fpocket	69 (67)	83 (85)	94 (92)	92 (92)
	PocketPicker	69	72	85	85
	CAST	58	67	75	83
	PASS	60	63	71	81
	SURFNET	52	54	75	78
	LIGSITE	58	69	75	87
	LIGSITE(CS)	60	69	77	87
Cheng <sup>5</sup> (20 proteins)	Fpocket	-	75 (70)	-	95 (90)
	PocketPicker	-	70	-	80
Astex <sup>6</sup> (82 proteins)	Fpocket	-	67 (73)	-	82 (88)
	PocketPicker	-	59	-	67

## CONCLUSION

### Major advantages of the package:

- ✓ Algorithms and code are fully scalable
- ✓ Easy evaluation of fpocket (new scoring function)
- ✓ Easy extraction of pocket descriptors
- ✓ Good scoring function
- ✓ Very fast (0.3 to 2 second per protein)

### Limitations

- ✓ Currently only available on Linux & Mac OSX
- ✓ Supervised learning on a low resolution training set

## REFERENCES

- V. Le Guilloux, P. Schmidtke, P. Tuffery, BMC Bioinformatics. 2009;**10**(1):168
- Barber & al (1996), *ACM Trans. on Mathematical Software* **22**(4):469-483, <http://www.qhull.org/>

- Weisel & al (2007), Chem Central Journal, **1**:7
- An & al (2005), Mol Cell Proteomics **4**:752-761
- Cheng & al (2007), Nat Biotechnol **27**:71-5
- Hartshorn & al. (2007), J Med Chem **50**, 726-741

