PockMan: Protein Pocket Detection Tool

User manual

🎓 MSc Bioinformatics Project

Author: Narine Fischer, Radostina Kisleva, Wael Badr
Date: ***April 2025***

1. What is PockMan?

PockMan is a bioinformatics tool that detects and analyzes potential binding pockets on protein structures. It uses a voxel-based grid projection system to analyze protein surface features and detect potential binding sites. It works with local `.pdb` files or fetches them directly from the RCSB Protein Data Bank using a PDB ID. The tool outputs structural and visual data suitable for downstream analysis in PyMOL or Chimera.

---

2. Installation Instructions

2.1 Install dependencies

Make sure Python 3 is installed. Recommended dependencies are automatically installed (e.g. `numpy`, `tqdm`, `biopython, requests`). In case they are not automatically installed run this on the terminal:

```
> pip install numpy biopython tqdm requests setuptools
```

2.2 Clone from GitHub

```
> git clone https://github.com/rsk170/PockMan.git
```

```
> cd PockMan
```

2.3 Build the package and install it

Option A:

```
> python3 setup.py sdist

> pip install dist/PockMan-1.0.tar.gz
```

Option B:

```
> pip install .
```

2.4 (Optional) Upgrading the package

To update to the latest version after making changes:

```
> pip install --upgrade dist/PockMan-1.0.tar.gz
```

2.5 (Optional) Uninstalling the package

```
> pip uninstall Pockman
```

---

3. Using PockMan

3.1 Arguments

To run PockMan from the command line, use the following command structure:

```
> pockman [OPTIONS] <pdb_input>
```

Positional Argument

pdb_input	Path to a local .pdb file or a valid PDB ID (e.g., 2RH1)

This is a required argument. The input can be a local .pdb file or a valid 4-character PDB ID (e.g., 1abc), which the program will fetch from the Protein Data Bank.If a local file is provided, it will be parsed directly. If a PDB ID is provided, PockMan will automatically download the corresponding structure. The `pdb` files can be downloaded manually from RCSB PDB.

Optional Arguments

Option	Description
-h, -help	Show a help message and exit
-- grid_size GRID_SIZE	Sets the resolution of the grid in Angstroms. If not set, it defaults to 1.0
– border BORDER	Adds extra space around the protein for grid calculation. If not set, it defaults to 5.0.
– diagonals	Include diagonal pocket detection for a more comprehensive search. Defaults to False.

These optional flags allow you to customize the pocket detection algorithm, especially when working with large or complex proteins.

3.2 Interactive mode (preferred)

If no command-line arguments are provided, the script will prompt for input:

Run:

```
> pockman
```

File selection:

- **Work with local pdb file:**

```
Enter the path to the input PDB file (or a PDB ID to use): 1e28.pdb
```

- **Use a PDB ID and download it automatically:**

```
Enter the path to the input PDB file (or a PDB ID to use): 1e28
```

Prompts for the user:

```
Enter grid size in Angstroms (default: 1.0): 1 (or leave blank)
```

```
Enter border size in Angstroms (default: 5.0): 5 (or leave blank)

Would you like to include diagonal pocket detection? (yes/no): yes
```

3.3 Command Line Execution

Instead of interactive mode, you could also execute the command in the following way.

Work with a local pdb file:

```
> pockman 1a28.pdb --grid_size 1.0 --border 5.0
```

Use a PDB ID and download it automatically:

```
> pockman 1a28 --grid_size 1.0 --border 5.0
```

To enable diagonal pocket detection, add the `--diagonals` flag:

```
> pockman 1a28.pdb --grid_size 1.0 --border 5.0 --diagonals
```

---

4. Output Structure

4.1 Example Run

If a PDB ID has been provided, the program will download the specified PDB files in a folder `pdb_files/.`

```
> pockman

Enter the path to the input PDB file (or a PDB ID to use): 3htb

Enter grid size in Angstroms (default: 1.0):

Enter border size in Angstroms (default: 5.0):

Would you like to include diagonal pocket detection? (yes/no):
```

In this example run, this is what we are inputting:

- **Protein structure**: 4-character PDB ID (e.g., `3htb`) which will be downloaded by

the program.

- **Grid resolution**: Left blank to use the default value (default: `1.0 Å`).
- **Border size**: Left blank to use the default value (default: `5.0 Å`).
- **Diagonal detection**: Left blank to use the default value (no).

4.2 Processing Steps

4.2.1 PDB download and cleaning

```
Local file not found. Assuming '3htb' is a PDB ID and attempting to
download...

Downloaded PDB 3htb → pdb_files/3htb_raw.pdb

Cleaned PDB file saved to: pdb_files/3htb.pdb
```

Since a local file is not found, PockMan attempts to fetch the PDB entry online. The downloaded file is cleaned and saved locally in the `pdb_files/` directory. The `3htb_raw.pdb` is rewritten so that it contains only lines that begin with ATOM, i.e. the standard protein (and nucleic-acid) coordinates. Everything else - HETATM records such as waters, ligands, metal ions, and all remark/header lines, is stripped out, leaving a file with just the protein backbone and side-chain atoms.

4.2.2 Grid projection

```
Protein grid shape: (50, 51, 60)

Pocket grid shape: (50, 51, 60)

✅ Protein grid projection completed.
```

A 3D voxel grid is created to represent the protein's spatial volume. The dimensions `(x, y, z)` reflect the size of the protein plus the border. Protein atoms are projected into this grid for structural analysis.

4.2.3 Pocket detection

```
Detecting pockets:
100%|████████████████████████████████████████████████████████████|
153000/153000 [00:11<00:00, 12946.67it/s]

✅ Protein detection completed.
```

The program evaluates each voxel for pocket likelihood. A progress bar indicates real-time progress so that the user is informed of the time it will take.

4.2.4 Pocket clustering and scoring

```
✅ Pocket clustering and scoring completed.
```

In this step, Identified pockets are ranked by structural and geometric features.

4.2.5 Nearby atom detection and results.

```
✅ Nearby atom detection completed.

📄 Results written to Ligand_binding_sites2_3htb.txt
```

Nearby residues and atoms are associated with each pocket. A .txt file is generated with possible binding sites.

4.2.6 Motivational quote

```
💬 "You cannot escape the responsibility of tomorrow by evading it
today."

   – Abraham Lincoln
```

At the end of each session, PockMan provides a random motivational quote—because science should also lift your spirits. ✨

4.3 Output Files

The following files are generated during a typical run:

Filename	Description
`pdb_files/3htb_raw.pdb`	Raw structure downloaded from the RCSB PDB
`pdb_files/3htb.pdb`	Cleaned PDB file used for analysis
`Ligand_binding_sites2_3htb.txt`	Summary of detected pockets and their associated residues

---

5. Visualizing the Results

Pending
