Istituto di Scienza e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

ISTI Technical Reports

Implementation of a drug discovery pipeline on the D4Science platform

Alessandro Orro, ITB-CNR, Milano, Italy

Pasqualina D'Ursi, ITB-CNR, Milano, Italy

Paola Fossa, Università di Genova, Genova, Italy

Leonardo Candela, ISTI-CNR, Pisa, Italy

Giancarlo Panichi, ISTI-CNR, Pisa, Italy

ISTI-TR-2023/001

Istituto di Scienza e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

Implementation of a drug discovery pipeline on the D4Science platform

Orro A., D'Ursi P., Fossa P., Candela L., Panichi G.

ISTI-TR-2023/001

This report documents the implementation of drug discovery pipeline in the D4Science platform realised in the context of the EOSC-Pillar project. In particular, it documents the pipeline and its constituents. Moreover, it describes how this pipeline has been integrated into the D4Science platform and exploited to create a dedicated Virtual Research Environment facilitating its exploitation and promoting a collaborative oriented approach for screening activities.

Keywords: Drug Discovery, D4Science, Virtual Research Environment.

Citation

Orro A., D'Ursi P., Fossa P., Candela L., Panichi G., *Implementation of a drug discovery pipeline on the D4Science platform*. ISTI Technical Reports 2023/001. DOI: 10.32079/ISTI-TR-2023/001.

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1
56124 Pisa Italy
http://www.isti.cnr.it

Technical Report, January 2023
Consiglio Nazionale delle Ricerche

Implementation of a Drug Discovery Pipeline on the D4Science Platform

Alessandro Orro[1]*, Pasqualina D'Ursi[1], Paola Fossa[2], Leonardo Candela[3], Giancarlo Panichi[3]

Abstract
This report documents the implementation of drug discovery pipeline in the D4Science platform realised in the context of the EOSC-Pillar project. In particular, it documents the pipeline and its constituents. Moreover, it describes how this pipeline has been integrated into the D4Science platform and exploited to create a dedicated Virtual Research Environment facilitating its exploitation and promoting a collaborative oriented approach for screening activities.

Keywords
Drug Discovery — D4Science — Virtual Research Environment

[1] *Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Via Fratelli Cervi, Segrate (MI), Italy*
[1] *Dipartimento di Farmacia, Università di Genova, Viale Cembrano 4, 16148, Genova, Italy*
[3] *Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche, Via G. Moruzzi 1, 56124, Pisa, Italy*
***Corresponding author**: alessandro.orro@itb.cnr.it

Contents

1. Introduction

The European Open Science Cloud (EOSC) is a European Commission funded program aiming at developing a shared cloud infrastructure following the system-of-systems approach and providing services for several research fields and communities with the final goal of encouraging data sharing and open science [1].

Open Science is an international movement aiming at making open every aspect of science and making science results accessible not only to researchers but to all citizens [6, 9]. Six basic principles has been identified as fundamentals of Open Science [7]: (*i*) *Open Data* - data should be licensed under an open license granting anyone can access, exploit, edit and share the data for any purpose; (*ii*) *Open Source* - software specifically written in the context of a scientific research should be freely available for possible modification and redistribution; (*iii*) *Open Methodology* - methods used by researchers to achieve scientific results should be publicly available; (*iv*) *Open Peer Review* - the review process of scientific articles should be revised to permit wider community (and not just invited reviewers) to participate to the revision process and make public the identity and reviewer during the entire process; (*v*) *Open Access* - scientific publications should always be open and available without charges or other factors that could hinder the their diffusion; and, (*vi*) *Open Educational Resources* - promote the creation and diffusion of teaching materials that are open to any form of use, distribution, re-mix, improvement and redistribution.

In this context, the goal of ESOC-Pillar project is to coordinate national initiatives and projects that are relevant to EOSC including a wide range of research communities in the countries of the partners: Italy, France, Germany, Austria and Belgium [4]. EOSC-Pillar operates in two ways: top-down, by promoting harmonization and adoption of standards and by translating national strategies to work plan, and bottom-up, by considering needs and requirements of specific research communities and engaging them for specific projects. The work plan includes research pilots from several scientific domains: natural sciences, engineering and technologies, humanities, medical & health science, social sciences, agricultural sciences.

CNR-ITB participated to the EOSC-Pillar project by providing a use case of Drug Discovery against COVID-19 aiming at the implementation of a large scale drug virtual screening pipeline and the deployment of the pipeline in the D4Science Platform [2, 3] that is one of the Cloud Infrastructure of the EOSC ecosystem.

The coronavirus outbreak has shown that our societies, governments, and healthcare systems are not prepared to handle emerging diseases. Many people are critical of our governing authorities, but the truth is that academia and science in general have a fragmented and unstructured approach to addressing pandemics. Funding agencies offer grants for projects that address specific issues, but there are no comprehensive services in place to support the scientists on the front lines. While various e-infrastructure resources and services are available, many of them are not tailored to the specific needs of front-line scientists who are under tremendous pressure to produce results and may not have the time or resources to train others to help them. The EOSC-Pillar use case on COVID-19 was conceived to develop a solution for rapid screening of chemical compounds against the known COVID-19 targets leveraging the service offering of the project (in terms of services and computing capacity) and the Open Science principles. In this way a speed-up can be achieved in the search of new potential drugs when new pathogenic variants will be discovered in the future. The solution was meant to be integrated into classic drug discovery pipelines for the screening of well characterized and approved drugs.

2. D4Science Infrastructure

D4Science[1] [2] is a mature data infrastructure developed with the technical and scientific leadership of CNR-ISTI and with the support of several EU-funder projects. This infrastructure proved to be suitable for the creation and operation of hundreds of Virtual Research Environments (VRE) in order to provide diverse community of practices with data and computational services promoting the implementation of Open Science practices [3]. It provides data hosting and sharing capabilities together with seamless access to distributed computational facilities for data analysis. D4Science is among the infrastructures EOSC-Pillar project offers for use cases implementation.

It was selected to provide services and to host the development and deployment of COVID-19 Use Cases proposed during the project. In particular, the *co-creation options* [3] supported are relevant for the rapid development and deployment of a working environment promoting open science practices and facilitating the integration of community specific pipelines.

3. The EOSC-Pillar4COVID-19 VRE

A dedicated Virtual Research Environment is created to support the use case and to provide both end-user and developers with an environment to share data and methods as well as to use them for drug discovery activity[2].

This VRE provides its users with a number of services including a rich portfolio of analytics solutions including: (*i*) DataMiner [5], a computing platform enabling its users to integrate analytics methods and execute them via a distributed computing infrastructure; (*ii*) JupyterLab, the platform for implementing and executing notebooks, and (*iii*) RStudio, the R-based development environment. Moreover, it is equipped with a *workspace* to facilitate the sharing of any data or other material of interest for the designated community (cf. Fig. 1), a *social networking* area that can be used by the members of the environment to communicate and collaborate by posts and replies, and a *catalogue* that can be used to publish and share any worth sharing artefact including research objects representing the results of a drug screening activity.

Fig. 2 displays the data miner analytics environment where analytics pipelines integrated in the platform are made available by a user friendly user graphical user interface where the users can execute a pipeline, monitor the execution and access the results.

4. Drug discovery pipeline

The software for Virtual Screening was designed and developed to perform thousands of docking simulations of a set of targets against a set of compounds and to assemble the single results to produce a final tabular results integrating annotation of the poses. The pipeline is described in Fig. 3. Each computational step is represented as a blue box (serial programs) or yellow box (parallel programs).

The model of the receptors and a compound library are the inputs of the software provided by users. The receptor is a model of the target protein of interest that needs to be activated or deactivated (depending on the application and specific disease) by a small molecule (drug). The compound library is a set of potential drugs. Both receptor and compound are represented in PDBQT format, which is a well-known and useful way to represent molecules for docking application.

The first step is the computation of a set of *druggable pockets* of the receptor (using the `fpocket` program [8]). In drug discovery, pocket is a term used to indicate a 3D region of the receptor that, due to specific atoms and bonds, is potentially able to bind a small molecule.

The next step is the *docking simulation* of each target pocket against each molecule of the compound library (based on `Autodock Vina` software [11]).

The result of the docking, represented as a protein-ligand complex, is then processed (by using the Plip software [10]) to extract the *interaction profile and binding properties*.

The last step gathers all results and builds a final report of all hits.

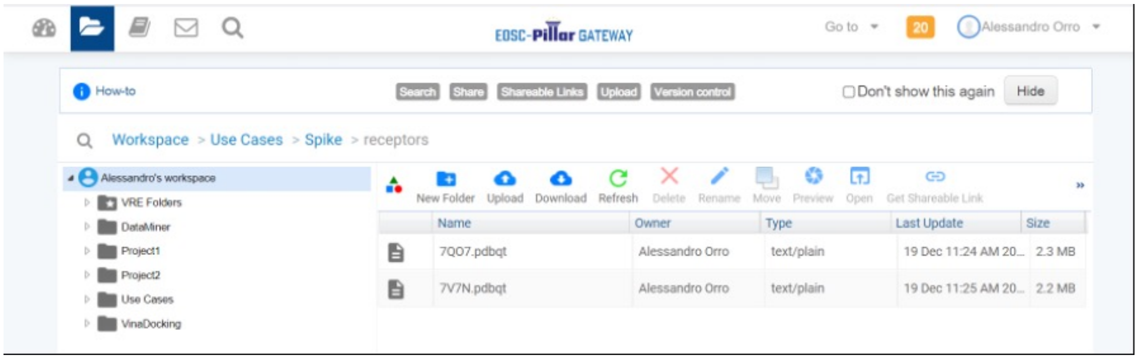
In the following paragraphs, we will provide details about each step of the pipeline, each program used and information about input and outputs

4.1 fpocket

fpocket [8] is an open source protein pocket prediction program based on Voronoi tessellation. It takes as input the structure of a protein receptor and searches for all druggable

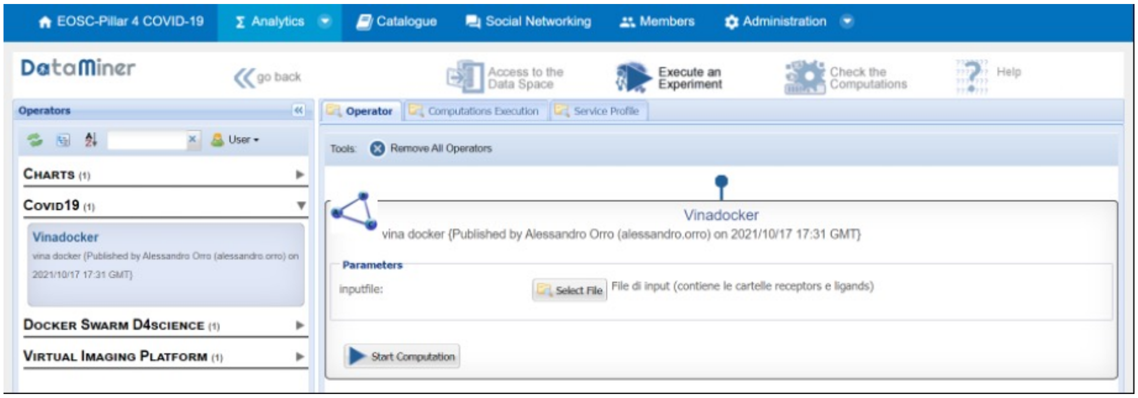
[1] D4Science website `www.d4science.org`

[2] EOSC-Pillar 4 Covid-19 VRE URL `https://eosc-pillar.d4science.org/web/eoscpillar_covid-19`



Name	Owner	Type	Last Update	Size
7QO7.pdbqt	Alessandro Orro	text/plain	19 Dec 11:24 AM 20...	2.3 MB
7V7N.pdbqt	Alessandro Orro	text/plain	19 Dec 11:25 AM 20...	2.2 MB

Figure 1. D4Science User Interface: workspace component for uploading and sharing files



Vinadocker

vina docker (Published by Alessandro Orro (alessandro.orro) on 2021/10/17 17:31 GMT)

inputfile: File di input (contiene le cartelle receptors e ligands)

Figure 2. D4Science User Interface: DataMiner component for importing and running analysis tool

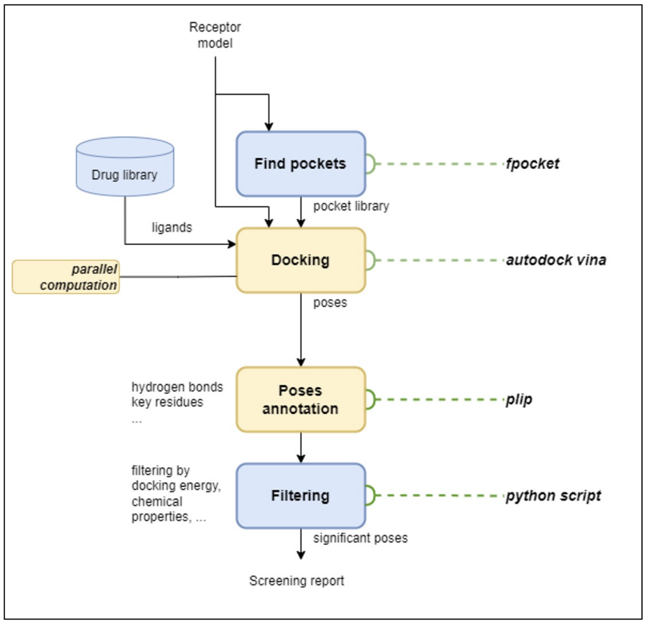


Figure 3. Drug Discovery Pipeline

pockets that are the regions of the protein surfaces most suitable for forming bonds with potential drugs. fpocket supports many parameters to configure the analysis, but using the default values is usually a good choice. Details about fpocket parameter are shown in Listing 1.

Listing 1. fpocket Parameters

```
Parameters
     -f (file)    : File of the protein structure
     -m (float)   : Minimum radius of an alpha-sphere.   (3.0)
     -M (float)   : Maximum radius of an alpha-sphere.   (6.0)
     -A (int)     : Minimum number of apolar neighbor for
                       an a-sphere to be considered as apolar.(3)
     -i (int)     : Minimum number of a-sphere per pocket.(30)
     -D (float)   : Maximum distance for first clustering
                       algorithm.                            (1.73)
     -s (float)   : Maximum distance for single linkage
                       clustering                             (2.5)
     -n (integer) : Minimum number of neighbor close from
                       each other (not merged otherwise).      (3)
     -r (float)   : Maximum distance between two pockets
                       barycenter (merged otherwise).         (4.5)
     -p (float)   : Minimum proportion of apolar sphere in
                       a pocket (remove otherwise)            (0.0)
     -v (integer) : Number of Monte-Carlo iteration for the
                       calculation of each pocket volume.   (2500)
     -b (integer) : Space approximation for the basic method
                       of the volume calculation. Not used by
                       default (Monte Carlo approximation is)
     -d flag      : Put this flag if you want to run fpocket for
                       database creation
```

The output of the program is (for each pocket) the list of atoms and some physical and chemical information describing the pocket (e.g., Listing 2).

Listing 2. List of atoms and some physical and chemical information describing the pocket.

```
Pocket 1 :
        Score :                         42.286
        Druggability Score :            0.077
        Number of Alpha Spheres :       321
        Total SASA :                    1267.531
        Polar SASA :                    634.440
        Apolar SASA :                   633.091
        Volume :                        2533.741
        Mean local hydrophobic density : 17.233
        Mean alpha sphere radius :      3.693
        Mean alp. sph. solvent access :  0.477
        Apolar alpha sphere proportion : 0.227
        Hydrophobicity score:           17.435
        Volume score:                    3.891
        Polarity score:                  27
        Charge score :                   0
        Proportion of polar atoms:      41.615
        Alpha sphere density :          11.097
        Cent. of mass - Alpha Sphere max dist: 32.089
        Flexibility :                    0.000

...
```

Fig. 4 shows protein receptors together with the predicted pockets. Each pocket is represented in different colors as a set of points near to the atoms of the corresponding pocket.

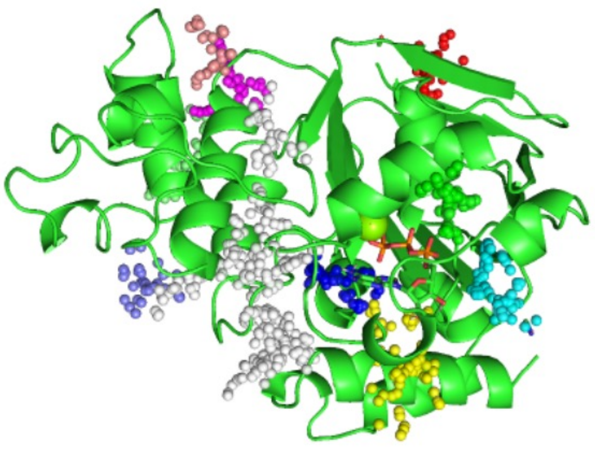


Figure 4. Pockets predicted with fpocket

4.2 autodock vina

`Autodock Vina` is one the most used docking programs due to its computational speed and the accuracy of the binding mode prediction. It takes both the receptor and the ligand as inputs together with the spatial region in which the ligand should be positioned, represented as a rectangular cuboid (center position and dimension). Other parameters are the randomized seed and the exhaustiveness that indicates how much time the search optimal pose configuration should last.

Details of parameters are shown in Listing 3.

Listing 3. Autodock Vina Parameters

```
Input:
  --receptor arg            rigid part of the receptor (PDBQT)
  --flex arg                flexible side chains, if any (PDBQT)
  --ligand arg              ligand (PDBQT)

Search space (required):
  --center_x arg            X coordinate of the center
  --center_y arg            Y coordinate of the center
  --center_z arg            Z coordinate of the center
  --size_x arg              size in the X dimension (Angstroms)
  --size_y arg              size in the Y dimension (Angstroms)
  --size_z arg              size in the Z dimension (Angstroms)

Output (optional):
  --out arg                 output models (PDBQT)
  --log arg                 optionally, write log file

Misc (optional):
  --cpu arg                 the number of CPUs to use
  --seed arg                explicit random seed
  --exhaustiveness arg (=8) exhaustiveness of the global search
                            (roughly proportional to time): 1+
  --num_modes arg (=9)      maximum number of binding modes to
                                            generate
  --energy_range arg (=3)   maximum energy difference between the
                            best binding mode and the worst
                            one displayed (kcal/mol)
```

The PDBQT format is a well known format to represent molecular structures in computational chemistry. It lists all atoms of the structure with their coordinates, atom type, atom name and partial charge. Moreover, the torsional flexibility of the molecule is represented with specific syntax. An example of PDBQT file is given in Listing 4.

Listing 4. Example of a PDBQT file

```
REMARK 4 active torsions:
REMARK status: ('A' for Active; 'I' for Inactive)
REMARK    1 A between atoms: N_1 and CA_5
REMARK    2 A between atoms: CA_5 and CB_6
REMARK    3 A between atoms: CA_5 and C_13
REMARK    4 A between atoms: CB_6 and CG_7
ROOT
ATOM      1 CA  PHE A 1   25.412  19.595  12.578  1.00 12.96
0.287 C
ENDROOT
BRANCH 1 2
ATOM      2 N   PHE A 1   25.225  18.394  13.381
1.00 13.04 -0.065 N
ATOM      3 HN3 PHE A 1   25.856  17.643  13.100  1.00  0.00
0.275 HD
ATOM      4 HN2 PHE A 1   25.558  18.517  14.337  1.00  0.00
0.275 HD
ATOM      5 HN1 PHE A 1   24.247  18.105  13.350  1.00  0.00
0.275 HD
ENDBRANCH 1 2
BRANCH 1 6
...
TORSDOF 4
```

The output of the program is a set of ligand poses in PDBQT format with optimal binding with the receptor together with a log file that list the energy of each pose (Listing 5).

Listing 5. Autodock Vina output log file example

```
mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1         -6.1      0.000      0.000
   2         -5.7      3.920     10.592
   3         -5.5      4.157     11.400
   4         -5.0      8.399     14.052
   5         -4.8     10.194     15.354
   6         -4.6     20.969     27.348
   7         -3.2     20.170     29.318
```

Fig. 5 displays an example of results of vina docking process.

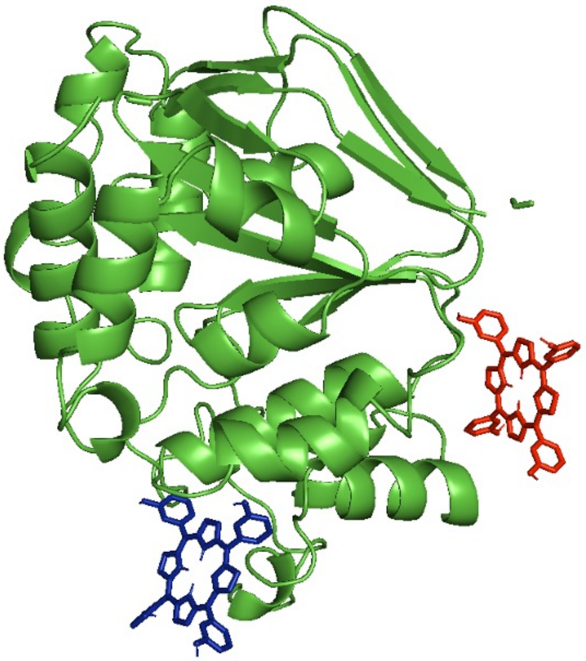


Figure 5. Example of results of vina docking: two binding poses (red and blue) of the same ligand

4.3 plip

The complex obtained by combining a pose ligand with its receptor can be analyzed with software like plip [10] in order to find out noncovalent protein-ligand interactions. Using geometrics and chemical models, plip is able to compute a set of properties related to the binding of the receptor with the corresponding ligand.

The output of plip is a XML file listing all binding and interaction properties (e.g., a portion of a plip output is reported in Listing 6 while Fig. 6 displays an example of the bonds found by plip).

Listing 6. A portion of a plip output.

```
...
       <hydrophobic_interaction id="2">
          <resnr>825</resnr>
          <restype>LEU</restype>
          <reschain>A</reschain>
          <resnr_lig>9999</resnr_lig>
          <restype_lig>LIG</restype_lig>
          <reschain_lig>9</reschain_lig>
          <dist>3.44</dist>
          <ligcarbonidx>25834</ligcarbonidx>
          <protcarbonidx>6207</protcarbonidx>
          <ligcoo>
            <x>160.771</x>
            <y>206.176</y>
            <z>195.815</z>
          </ligcoo>
          <protcoo>
            <x>161.371</x>
            <y>204.978</y>
            <z>192.649</z>
          </protcoo>
        </hydrophobic_interaction>
...
```

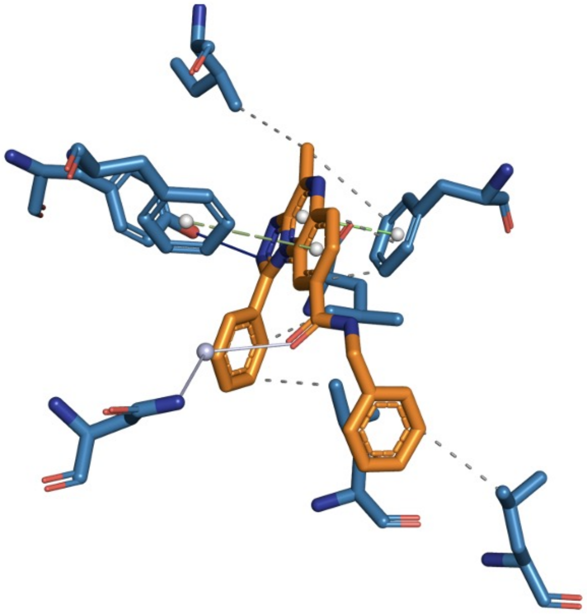


Figure 6. Bonds found by plip. Source: `https://github.com/pharmai/plip`

5. Results

The pipeline has been implemented as a single Python script that, in the simpler form, takes an input zip file that contains both the receptors and ligands and an output folder where the results will be generated (see Listing 7). The input and output can also be expressed as a unique identifier UUID relative to the D4Science platform. In this case, an authentication token is required and appropriate APIs are used to read and write data from D4Science. The –pocket parameter allow to specify the strategy to find pockets before the docking. The implementation supports multi cores architecture by using parallelization of the vina docking simulation and plip analysis.

Listing 7. Pipeline parameters

```
d4pipeline
  --input        input.zip or d4science://UUID
  --output       outfolder or d4science://UUID
  --token        D4Science token
  --pocket       fpocket or global
  --max-pockets  max number of pocket to uwe
```

The input directory is structured as follows:

```
ligands/
    ligand1.pdbqt
    ligand2.pdbqt
    ligand3.pdbqt
    ...
receptors/
    receptor1.pdbqt
    receptor2.pdbqt
    ...
```

The output directory is structured in a similar way contains the all the output results of the vina, fpocket and plip analysis as follows:

```
ligands/
    ligand1.pdbqt
    ligand2.pdbqt
    ligand3.pdbqt
    ...
receptors/
    receptor1.pdbqt
    receptor2.pdbqt
    receptor1_out/
        receptor1_info.txt
        pockets/
            pocket9_atm.pdb
            pocket9_atm.pqr
            ...
    ...
vinaout/
    7V7N-DB14658_splitted-pocket1.vinaout
    7V7N-DB14658_splitted-pocket1.pdbqt
    ...
plipout/
    7QO7-ZINC000261515675_splitted-pocket1-pose7_protonated.pdb
    7QO7-ZINC000261515675_splitted-pocket1-pose7.xml
    ...
```

The final report, assembled in the last step of the pipeline. It is a TSV file (Tab Separated Values) that contains information about docking simulation for each tuple (receptor, pocket, ligand, pose) and the binding information.

The list of all fields is reported Table 1.

The integration and deployment of this pipeline into the D4Science is achieved by using two approaches: (*i*) a first version of the pipeline is integrated by onboarding the Python code into the DataMiner platform [3]; (*ii*) a complete version of the pipeline is integrated by creating a docker image of the application and register it into the Docker Hub[3] with the name `cnritb/d4pipeline:0.1` and exploiting the Docker Image Executor made available by the DataMiner platform.

6. Conclusion

This brief report documents the activity leading to the creation of a Virtual Research Environment supporting a use case of Drug Discovery against COVID-19 aiming at the implementation of a large scale drug virtual screening pipeline and the deployment of the pipeline in the D4Science Platform. In particular, it provides the detail on the pipeline implementation and describes how this pipeline has been easily integrated into a web-based working environment benefitting from a distributed computing infrastructure for its execution and from user friendly facilities for collaboration.

Acknowledgments

This work received funding from the European Union's Horizon 2020 research and innovation programme under EOSC-Pillar project (grant agreement No. 857650).

References

[1] European open science cloud. URL `https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en`.

[2] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, L. Frosini, L. Lelii, F. Mangiacrapa, P. Pagano, G. Panichi, and F. Sinibaldi. Enacting open science by D4Science. *Future Generation Computer Systems*, 101: 555–563, Dec. 2019. ISSN 0167739X. doi: 10.1016/j.future.2019.05.063.

[3] M. Assante, L. Candela, D. Castelli, R. Cirillo, G. Coro, A. Dell'Amico, L. Frosini, L. Lelii, M. Lettere, F. Mangiacrapa, P. Pagano, G. Panichi, T. Piccioli, and F. Sinibaldi. Virtual research environments co-creation: The D4Science experience. *Concurrency and Computation: Practice and Experience*, Mar. 2022. ISSN 1532-0626, 1532-0634. doi: 10.1002/cpe.6925.

[4] E.-P. Consortium. Eosc-pillar project. URL `https://www.eosc-pillar.eu/`.

[5] G. Coro, G. Panichi, P. Scarponi, and P. Pagano. Cloud computing in a distributed e-infrastructure using the web processing service standard: Cloud computing in a distributed e-infrastructure using the web processing service standard. *Concurrency and Computation: Practice and Experience*, 29(18):e4219, Sept. 2017. ISSN 15320626. doi: 10.1002/cpe.4219.

[6] B. Fecher and S. Friesike. Open Science: One Term, Five Schools of Thought. In S. Bartling and S. Friesike, editors, *Opening Science*, pages 17–47. Springer International Publishing, Cham, 2014. ISBN 978-3-319-00025-1 978-3-319-00026-8. doi: 10.1007/978-3-319-00026-8_2.

[7] R. V. Gallagher, D. S. Falster, B. S. Maitner, R. Salguero-Gómez, V. Vandvik, W. D. Pearse, F. D. Schneider,

[3] Docker Hub website `https://hub.docker.com/`

Table 1. Pipeline report fields

Field	Description
receptor	The identifier of the receptor;
pocket	The identifier of the pocket.
ligand	The identifier of the ligand.
pose	The index of the pose (ordered by decreasing affinity);
affinity	The binding affinity of the pose+receptor complex;
rmsd_lb	The lower bound of the set of RMSD values of the pose cluster;
rmsd_ub	The upper bound of the set of RMSD values of the pose cluster;
n_hydrophobic_interactions	The number of hydrophobic bonds;
n_hydrogen_bonds	The number of hydrogen bonds;
n_water_bridges	The number of water bridges;
n_salt_bridges	The number of salt bridges;
n_pi_stacks	The number of pi orbital overlapping interactions;
n_pi_cation_interactions	The number of Cation–π interactions found in the pose+receptor complex;
n_halogen_bonds	The number of halogen bonds;
n_metal_complexes	The number of metal complex specific interactions;

J. Kattge, J. H. Poelen, J. S. Madin, M. J. Ankenbrand, C. Penone, X. Feng, V. M. Adams, J. Alroy, S. C. Andrew, M. A. Balk, L. M. Bland, B. L. Boyle, C. H. Bravo-Avila, I. Brennan, A. J. R. Carthey, R. Catullo, B. R. Cavazos, D. A. Conde, S. L. Chown, B. Fadrique, H. Gibb, A. H. Halbritter, J. Hammock, J. A. Hogan, H. Holewa, M. Hope, C. M. Iversen, M. Jochum, M. Kearney, A. Keller, P. Mabee, P. Manning, L. McCormack, S. T. Michaletz, D. S. Park, T. M. Perez, S. Pineda-Munoz, C. A. Ray, M. Rossetto, H. Sauquet, B. Sparrow, M. J. Spasojevic, R. J. Telford, J. A. Tobias, C. Violle, R. Walls, K. C. B. Weiss, M. Westoby, I. J. Wright, and B. J. Enquist. Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology & Evolution*, 4(3):294–303, Feb. 2020. ISSN 2397-334X. doi: 10.1038/s41559-020-1109-6.

[8] V. Le Guilloux, P. Schmidtke, and P. Tuffery. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1):168, Dec. 2009. ISSN 1471-2105. doi: 10.1186/1471-2105-10-168.

[9] B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. L. Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, and T. Yarkoni. Promoting an open research culture. *Science*, 348(6242):1422–1425, June 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aab2374.

[10] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder. PLIP: fully automated protein–ligand interaction profiler. *Nucleic Acids Research*, 43(W1): W443–W447, July 2015. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkv315.

[11] O. Trott and A. J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, pages NA–NA, 2009. ISSN 01928651, 1096987X. doi: 10.1002/jcc.21334.