

# Retour d'expériences sur la publication de données en biologie

Pierre Poulain

Institut Jacques Monod, CNRS & Université Paris Diderot

[pierre.poulain@univ-paris-diderot.fr](mailto:pierre.poulain@univ-paris-diderot.fr)  
[@pierrepo](#)

27/11/2018



# Quelques données en biologie :

- **Tableaux de données**
- **Paramètres (de simulation)**
- **Programmes informatiques**
- **Données massives**

**Pourquoi publier tout ça ?**  
**Comment faire ?**

# On exclut tout de suite

- Site web perso de chercheur
- Site web de labo / d'institut

# Cas n° 1 : tableau de données propriétés d'une famille de protéines

OPEN ACCESS Freely available online



## Detection and Architecture of Small Heat Shock Protein Monomers

Pierre Poulain\*, Jean-Christophe Gelly, Delphine Flatters\*

DSIMB, Inserm UMR-S 665 and Université Paris Diderot - Paris 7, INTS, Paris, France

### Supporting Information

**Dataset S1** List of the 3787 sequences constituting the sHSPdata09 dataset. First column is the UniProt accession, second column is the corresponding group and third column is the length of the detected ACD.

Found at: [doi:10.1371/journal.pone.0009990.s001](https://doi.org/10.1371/journal.pone.0009990.s001) (0.10 MB)

# Cas n° 1 : tableau de données

## propriétés d'une famille de protéines

accession	group	ACD_size	accession	group	ACD_size	accession	group	ACD_size
B9SA05	plant	91	B2AV59	fungi	158	C4ANX2	bacOther	88
C0QVM3	bacOther	88	Q5ZTH1	bacA	87	Q16S84	animal	83
O01718	animal	101	Q4WYW9	fungi	124	A8EX24	bacOther	89
A1E463	plant	90	Q3R3S4	bacOther	89	B9S3B4	plant	90
B0EHH3	other	87	A8NZF2	fungi	120	C4JZE7	fungi	146
A0K2I6	bacOther	90	Q8Z2L8	bacA	86	B0TJF1	bacA	86
C1C473	animal	83	B6TD78	plant	97	B9S3B3	plant	90
Q1W289	animal	83	A9RDF3	plant	80	B9S3B2	plant	90
B6T6N6	plant	90	C0PTB9	plant	92	A7P4P7	plant	90
Q3STL4	bacOther	86	Q86GU1	animal	135	B8FLL8	bacOther	88
Q8DDJ3	bacA	85	A5BAN4	plant	90	A1CE36	fungi	139
Q061S4	bacOther	89	B5YBV6	bacOther	89	B9S1R8	plant	90
Q99RP2	bacOther	86	B9SIR7	plant	91	B1L7F8	archae	69
Q16S85	animal	91	A0YRM5	bacOther	89	C2WLY4	bacOther	90
B7P7F7	animal	84	A5JBH0	bacOther	90	B8FLL7	bacOther	88
Q97W19	archae	75	Q82Q29	bacOther	89	A7P4P8	plant	90

# Cas n° 1 : tableau de données

## propriétés d'une famille de protéines

accession	group	ACD_size	accession	group	ACD_size	accession	group	ACD_size
B9SA05	plant	91	B2AV59	fungi	158	C4ANX2	bacOther	88
C0QVM3	bacOther	88	Q5ZTH1	bacA	87	Q16S84	animal	83
O01718	animal	101	Q4WYW9	fungi	124	A8EX24	bacOther	89
A1E463	plant	90	Q3R3S4	bacOther	89	B9S3B4	plant	90
B0EHH3	other	87	A8NZF2	fungi	120	C4JZE7	fungi	146
A0K2I6	bacOther	90	Q8Z2L8	bacA	86	B0TJF1	bacA	86
C1C473	animal	83	B6TD78	plant	97	B9S3B3	plant	90
Q1W289	animal	83	A9RDF3	plant	80	B9S3B2	plant	90
B6T6N6	plant	90	C0PTB9	plant	92	A7P4P7	plant	90
Q3STL4	bacOther	86	Q86GU1	animal	135	B8FLL8	bacOther	88
Q8DDJ3	bacA	85	A5BAN4	plant	90	A1CE36	fungi	139
Q061S4	bacOther	89	B5YBV6	bacOther	89	B9S1R8	plant	90
Q99RP2	bacOther	86	B9SIR7	plant	91	B1L7F8	archae	69
Q16S85	animal	91	A0YRM5	bacOther	89	C2WLY4	bacOther	90
B7P7F7	animal	84	A5JBH0	bacOther	90	B8FLL7	bacOther	88
Q97W19	archae	75	Q82Q29	bacOther	89	A7P4P8	plant	90

en PDF !





# Cas n° 2 : fichiers de paramètres

## Simulations numériques

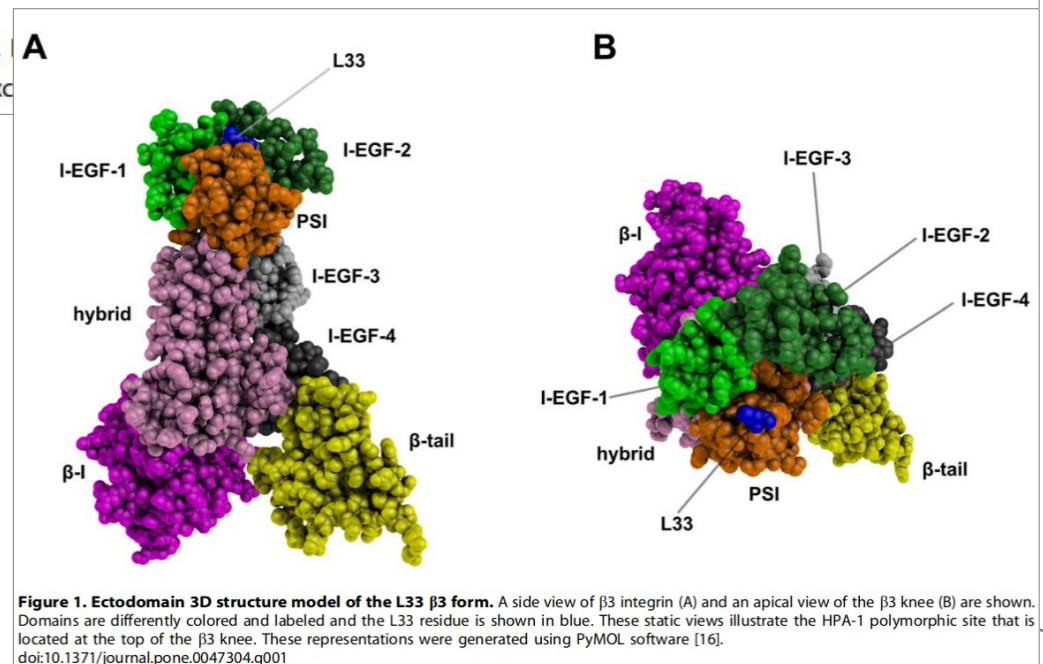
OPEN ACCESS Freely available online



### Modeling and Molecular Dynamics of HPA-1a and -1b Polymorphisms: Effects on the Structure of the $\beta 3$ Subunit of the $\alpha \text{IIb}\beta 3$ Integrin

Vincent Jallu<sup>1,9</sup>, Pierre Poulain<sup>2,3,4,5,9</sup>, Patrick F. J. Fuchs<sup>2,3,4,5</sup>, Cecile Kaplan<sup>1</sup>, Alexandre G. de Brevern<sup>2,3,4,5\*</sup>

1 Laboratoire d'Immunologie Plaquettaire, INTS, Paris, France, 2 INSERM, U665, Paris, France, 3 INSERM, U1133, Paris, France, 4 Institut National de la Transfusion Sanguine, Paris, France, 5 Laboratoire d'Exc



# Cas n° 2 : fichiers de paramètres


md.mdp (3.35 kB)

md\_eq.mdp (3.29 kB)

md\_protocol.sh (3.39 kB)

mini1.mdp (1.25 kB)



mini2.mdp (1.33 kB)




Cite

Download all (12.62 kB)


Share Embed + Collect (you need to log in first)

5 files  

## Molecular Dynamics Protocol with Gromacs 4.0.7

Version 3  Fileset posted on 18.12.2012, 15:29 by [Pierre Poulain](#), Alexandre G de Brevem, Patrick F J Fuchs

771 views | 62 downloads | 0 citations



**Molecular Dynamics Protocol with Gromacs 4.0.7** This protocol had been used in the paper *Modeling and Molecular Dynamics of HPA-1a and -1b Polymorphisms: Effects on the Structure of the  $\beta 3$  Subunit of the  $\alpha \text{IIb}\beta 3$  Integrin* (PLoS ONE 7(11): e47304, 2012).

REFERENCES

- <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0047304>
- <http://dx.doi.org/10.6084/m9.figshare.104602>


CATEGORIES

- Bioinformatics

KEYWORD(S)

molecular dynamics gromacs

LICENCE

 CC BY 4.0





# Figshare ?



(2011 / Marc Hahnel)



**Digital Science**  
(Overleaf, Altmetric)



**Holtzbrinck**



**Springer Nature**

# Figshare / FAIR ?



**figshare** ✓  
@figshare

Abonné



Figshare and the **#FAIR** data principles:

[figshare.com/articles/Figshare ... #rdm](https://figshare.com/articles/Figshare_and_FAIR_data_principles)

🌐 Traduire le Tweet



## **Figshare and FAIR data principles**

This webinar was held on 13 June 2018 and looked at the FAIR data principles and how they are applied to Figshare. For further information please contact [patrick@figshare.com](mailto:patrick@figshare.com).

[figshare.com](https://figshare.com)

12:36 - 14 juin 2018

# Cas n° 3 : tableau de données

## ***Plasmodium falciparum* infection in febrile Congolese children: prevalence of clinical malaria 10 years after introduction of artemisinin-combination therapies**

**Mandingha Kosso Etoaka-Beka<sup>1,2</sup>, Francine Ntoumi<sup>1,2,3</sup>, Michael Kombo<sup>1</sup>, Julia Deibert<sup>3</sup>, Pierre Poulain<sup>1,4,5,6,7</sup>, Christevy Vouvongui<sup>1</sup>, Simon Charles Kobawila<sup>2</sup> and Felix Koukouikila-Koussounda<sup>1,2</sup>**


1 *Fondation Congolaise pour la Recherche et l'Innovation*  
 2 *Faculté des Sciences et Techniques, Université de la République*  
 3 *Institute for Tropical Medicine, University of Bonn*  
 4 *Institut National de la Santé et de la Recherche Scientifique*  
 5 *UMR\_S 1134, DSIMB, Sorbonne Paris Cité*  
 6 *Institut National de la Transfusion Sanguine*  
 7 *UMR\_S 1134, Laboratory of Excellence*

**Table 1** Characteristics of Congolese febrile patients recruited in Brazzaville

Variables (%)	All ( <i>n</i> = 229) 100	AA ( <i>n</i> = 199) 87	AS ( <i>n</i> = 30) 13
Age, years, mean ± SD [min–max]	3.1 ± 2.5 [1–10]	3.1 ± 2.5 [1–10]	2.9 ± 2.3 [1–8]
Sex ratio F/M (%)	105/124 46/54	88/111 44/56	17/13 57/43
Axillary temp. °C, mean ± SD [min–max]	37.9 ± 1.1 [34.6–40.8]	37.9 ± 1.1 [35.8–40.8]	37.8 ± 1.2 [34.6–40.0]
Haemoglobin, g/dl, mean ± SD [min–max]	11.3 ± 4.2 [4.9–69.4]	11.3 ± 4.4 [4.9–69.4]	11.2 ± 1.4 [7.4–13.9]
Bed net use, <i>n</i> (%)	219 (96)	189 (95)	30 (100)
Treated bed net use, <i>n</i> (%)	101 (44)	2 (46)	9 (30)
Previous antimalarial self-treatment, <i>n</i> (%)	71 (31)	60 (30)	11 (37)
Anaemia, <i>n</i> (%)	96 (42)	86 (43)	10 (33)
<i>P. f.</i> uncomplicated malaria, <i>n</i> (%)*	22 (10)	18 (9)	4 (13)
<i>P. f.</i> submicroscopic infection, <i>n</i> (%)	49 (21)	41 (21)	8 (27)

\*Mann–Whitney *P* = 0.277

# Cas n° 3 : tableau de données




September 19, 2016 Dataset Open Access

## Plasmodium falciparum infection in febrile Congolese children: prevalence of clinical malaria ten years after introduction of Artemisinin-combination therapies

Etoka-Beka, Mandingha Kosso; Ntoumi, Francine; Kombo, Michael; Deibert, Julia; Poulain, Pierre; Vouvougui, Christevy; Kobawila, Simon Charles; Koukouikila-Koussounda, Felix

dataset used in the paper.

33 views 55 downloads [See more details...](#)

Indexed in 

**Publication date:** September 19, 2016  
**DOI:** DOI 10.5281/zenodo.154453  
**Keyword(s):** malaria congo  
**Published in:** Tropical Medicine & International Health:  
**Related identifiers:** Referenced by: [10.1111/tmi.12786](https://doi.org/10.1111/tmi.12786)  
**License (for files):** [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Preview

patient_id	sampling_date	age	sex	weight	temp	hb_profile	hb_conc	bed_net_use
MNG001	2014/09/03	9	female	25.0	38.0	AA	8.7	True
MNG002	2014/09/03	6	female	24.4	35.8	AA	12.8	True
MNG003	2014/09/03	3	female	15.5	39.3	AA	10.5	True
MNG004	2014/09/04	1	male	12.0	39.7	AS	9.8	True
MNG005	2014/09/04	1	male	9.0	37.0	AA	12.5	True
MNG007	2014/09/04	3	female	11.8	37.1	AA	11.2	True
MNG008	2014/09/05	5	male	20.3	37.0	AA	9.5	True
MNG009	2014/09/05	3	male	21.0	36.7	AA	12.3	True
MNG010	2014/09/05	5	female	20.0	37.1	AS	9.0	True



# OpenAIRE + CERN (2013)

## FAIR Principles

FAIR Principles definition as referenced from: *Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).*

### To be Findable:

- **F1:** (meta)data are assigned a globally unique and persistent identifier
  - A DOI is issued to every published record on Zenodo.
- **F2:** data are described with rich metadata (defined by R1 below)
  - Zenodo's metadata is compliant with [DataCite's Metadata Schema](#) minimum and recommended terms, with a few additional enrichments.
- **F3:** metadata clearly and explicitly include the identifier of the data it describes
  - The DOI is a top-level and a mandatory field in the metadata of each record.
- **F4:** (meta)data are registered or indexed in a searchable resource
  - Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
  - Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

### To be Accessible:

- **A1:** (meta)data are retrievable by their identifier using a standardized communications protocol
  - Metadata for individual records as well as record collections are harvestable using the [OAI-PMH](#) protocol by the record identifier and the collection name.
  - Metadata is also retrievable through the public [REST API](#).

# Cas n°4 : code informatique



PeerJ

## PBxplore: a tool to analyze local protein structure and deformability with Protein Blocks

Jonathan Barnoud<sup>1,2,3,4,6,\*</sup>, Hubert Santuz<sup>1,2,3,4,7,\*</sup>, Pierrick Craveur<sup>1,2,3,4,8</sup>, Agnel Praveen Joseph<sup>1,2,3,4,9</sup>, Vincent Jallu<sup>5</sup>, Alexandre G. de Brevern<sup>1,2,3,4,\*</sup> and Pierre Poulain<sup>1,2,3,4,10,\*</sup>

<sup>1</sup>INSERM, U 1134, DSIMB, Paris, France

<sup>2</sup>Univ. Paris Diderot, Sorbonne Paris Cité, Univ de la Réunion, Univ des Antilles, UMR-S 1134, Paris, France

<sup>3</sup>Institut National de la Transfusion Sanguine (INTS), Paris, France

<sup>4</sup>Laboratoire d'Excellence GR-Ex, Paris, France

<sup>5</sup>Platelet Unit, INTS, Paris, France

<sup>6</sup>Current affiliation: Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

<sup>7</sup>Current affiliation: Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-Chimique, Paris, France

<sup>8</sup>Current affiliation: Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, United States of America

<sup>9</sup>Current affiliation: Birkbeck College, University of London, London, UK

<sup>10</sup>Current affiliation: University of London, London, UK  
\*Thesis

### Data Availability

The following information was supplied regarding data availability:

GitHub: Available at <https://github.com/pierrepo/PBxplore>

Zenodo: Available at <https://dx.doi.org/10.5281/zenodo.1016257>.

# Cas n°4 : code informatique

The screenshot shows the GitHub repository page for 'pierrepo / PBxplore'. At the top, it displays 'Watch 8', 'Star 14', and 'Fork 8'. Below this, there are navigation tabs for 'Code', 'Issues 4', 'Pull requests 2', 'Projects 0', and 'Insights'. A description reads: 'A suite of tools to explore protein structures with Protein Blocks' with a link to the documentation. There are tags for 'python', 'protein', 'structure', and 'bioinformatics-analysis'. Statistics show '612 commits', '4 branches', '11 releases', '3 contributors', and 'MIT' license. A 'Branch: master' dropdown and a 'New pull request' button are visible. A 'Find file' button and a 'Clone or download' button are also present. A recent commit by 'jbarnoud' is highlighted, with a message 'Merge pull request #171 from pierrepo/add-citation-info' and the latest commit 'ba12999 on 5 Dec 2017'. Below this, a list of folders and their commit messages is shown: 'demo\_doc' (Add an example plot with RMSF and Neq superposed), 'demo\_paper' (Add example logo), 'devtools' (Bump version: 1.3.7 -> 1.3.8), 'doc' (Set citation as text block), and 'pbxplore' (Bump version: 1.3.7 -> 1.3.8).

The screenshot shows a Zenodo record for 'PBxplore (v1.3.8): A program to explore protein structures with Protein Blocks'. The record is dated 'October 17, 2017' and is categorized as 'Software' with 'Open Access'. It shows '34 views' and '4 downloads'. The authors listed are Barnoud, Jonathan; Santuz, Hubert; de Brevern, Alexandre G.; and Poulain, Pierre. A note states: 'This repository contains the archived release for the software PBxplore'. A 'Preview' section shows a file tree for 'PBxplore-v1.3.8.zip' with a subdirectory 'pierrepo-PBxplore-3449404' containing various files like 'coveragerc', 'gitignore', 'travis.yml', 'zenodo.json', 'CHANGELOG', 'LICENSE', 'MANIFEST.in', 'README.rst', and a 'demo\_doc' folder with sub-files like 'psi\_md\_traj.PB.fasta', 'psi\_md\_traj.gro', 'psi\_md\_traj.xtc', etc. The 'Available in' section features a large 'GitHub' logo. The 'Publication date' is 'October 17, 2017' and the 'DOI' is '10.5281/zenodo.1016257'. 'Keyword(s)' include 'Python', 'Protein Blocks', 'Proteins', 'Structure', 'Flexibility', and 'Deformability'. 'Related identifiers' include a link to the GitHub repository. The 'License (for files)' is 'MIT License'.

# GitHub



# Zenodo

# Cas n°4 : code informatique

GitHub repository page for 'pierrepo / PBxplore'. The page shows repository statistics: 612 commits, 4 branches, 11 releases, 3 contributors, and MIT license. A list of files is visible, including 'demo\_doc', 'demo\_paper', 'devtools', 'doc', and 'pbxplore'. A pull request by 'jbarnoud' is also shown.

## GitHub



Software Heritage

## Zenodo

Software Heritage archived directory for the GitHub repository 'https://github.com/pierrepo/PBxplore'. The page shows a file listing with columns for File, Mode, and Size. The files listed are 'demo\_doc', 'demo\_paper', 'devtools', 'doc', and 'pbxplore', all with mode 'd-----'. The page also shows 'Branch: HEAD' and '4260527 /'. Other information includes 'Visits', 'Snapshot date: 21 March 2018, 21:49 UTC', 'Branches (104)', and 'Releases (4)'. A 'DOI' is provided: '10.5281/zenodo.1016257'. Keywords include 'Python', 'Protein Blocks', 'Proteins', 'Structure', and 'Flexibility'. The license is 'MIT License'.

Zenodo repository page for 'PBxplore (v1.3.8): A program to explore protein structures with Protein Blocks'. The page shows 34 views and 4 downloads. The repository contains the archived release for the software PBxplore. A file listing is visible, including 'pierrepo-PBxplore-3449404', 'coveragerc', 'gitignore', 'travis.yml', 'zenodo.json', 'CHANGELOG', 'LICENSE', 'MANIFEST.in', 'README.rst', 'demo\_doc', 'psl\_md\_traj\_PB.fasta', 'psl\_md\_traj\_gro', 'psl\_md\_traj\_xtc', 'psl\_md\_traj\_1\_PB.count', 'psl\_md\_traj\_1\_PB.fasta', 'psl\_md\_traj\_1\_gro', 'psl\_md\_traj\_1\_xtc', and 'psl\_md\_traj\_1\_chiffart\_PR.count'. The page also shows a 'Publication date' of October 17, 2017, and a 'DOI' of '10.5281/zenodo.1016257'. The license is 'MIT License'.

# Cas n°5 : données massives « omiques »

## Séquençage haut débit (ADN / ARN) :

- **Gene Expression Omnibus (GEO)**

<https://www.ncbi.nlm.nih.gov/geo/>

- **Sequence Read Archive (SRA)**

<https://www.ncbi.nlm.nih.gov/sra>

## Protéomique par spectrométrie de masse :

- **PRoteomics IDEntifications (PRIDE)**

<https://www.ebi.ac.uk/pride/archive/>