> Nantes
2—5
juillet

JOBIM 2019

JOURNÉES OUVERTES DE BIOLOGIE INFORMATIQUE & MATHÉMATIQUES

Thèmatiques :

Biologie structurale
Biologie des systèmes
Epidémiologie Génétique
Evolution/Phylogénie
Génomique/Métagénomique
Sciences des données

Abstracts

Keynotes :

Chloé-Agathe Azencott, Paris
Alexander Bockmayr, Berlin
Alessandra Carbone, Paris
Olivier Delaneau, Lausanne
Christophe Dessimoz, Lausanne
Juliette Martin, Lyon

https://jobim2019.sciencesconf.org

Dear attendees of the 20th edition of JOBIM, welcome in Nantes !

JOBIM is the French national conference dedicated to promoting an active interface between Biology, Computer Sciences, and Mathematics. After a previous visit to Nantes in 2009, JOBIM comes back this year in this same city from western France. Since the last visit, the bioinformatics community has impressively grown, and new fields are today covered. The impressive number of submissions at JOBIM 2019 reflects such an increase in our community. The Program Committee received a total of 260 submissions deciphered as 29 long presentations, 11 flash presentations, 15 demos and 204 posters. As a main novelty of the 2019th edition, JOBIM 2019 will present five additional thematic sessions that will cover particular topics more specialized.

We sincerely thank all the members of the Program Committee who helped us to set up a great scientific program by reviewing all submissions in time. This task would have been impossible without them! We also are grateful to the six invited speakers that have accepted to contribute to the success of the JOBIM edition in Nantes.

We are indebted to the organizing institutions, the SFBI, the GDR BIM, and the IFB. We are also grateful to all our partners and sponsors for their financial support.

Finally, we could also not forget to warmly thank Sophie Girault, Elodie Guidon, Aurore Morvan, and Jérémie Ségard as well as all the members of the organizing committee who worked collectively without counting sweat and tears to welcome the cream of bioinformaticians today in the best conditions.

Damien Eveillard and Audrey Bihouée
Organizing committee

Jérémie Bourdon and Richard Redon
Program committee

Program committee

Chloé-Agathe Azencott
Benjamin Bardiaux
Anais Baudot
Benoit Bely
Séverine Bérard
Camille Berthelot
Jérémie Bourdon
Laurence Calzone
Samuel Chaffron
Madalena Chaves
Hélène Chiapello
Sarah Cohen-Boulakia
Erwan Corre
Alexandre G. De Brevern
Damien Eveillard
Christine Froidevaux
Olivier Gandrillon
Christine Gaspin
Pierre-Antoine Gourraud
Carrier Gregory
Carito Guziolowski
Carl Herrmann
Claudine Landès
Solena Le Scouarnec
Claire Lemaitre
Cedric Lhoussaine
Pierre Lindenbaum
Anne Lopes
Morgan Magnin
Anthony Mathelier
Macha Nikolski
Eric Pelletier
Guy Perrière
Ovidiu Radulescu
Richard Redon
Eric Rivals
Frédéric Saubion
Sophie Schbath
Aurélien Sérandour
Anne Siegel
Christine Sinoquet
Stéphane Téletchéa
Dominique Tessier
Denis Thieffry
Morgane Thomas-Chollier
Hélène Touzet
Jacques van Helden

Local organizing committee

Audrey Bihouée (l'institut du thorax)
Damien Eveillard (LS2N)
Isabelle Alves (l'institut du thorax)
Agnes Basseville (CRCINA)
Jérémie Bourdon (LS2N)
Karine Cantele (LS2N)
Grégory Carrier (IFREMER)
Samuel Chaffron (LS2N)
Eric Charpentier (l'institut du thorax)
Benjamin Churcheward (LS2N)
Hélène Chiapello (SFBI)
Erwan Delage(LS2N)
Christian Dina (l'institut du thorax)
Solenne Dumont (l'institut du thorax)
Axelle Durand (CRTI)
Guillaume Fertin (LS2N)
Adrien Foucal (l'institut du thorax)
Victor Gaborit (CRCINA)
Alban Gaignard (l'institut du thorax)
Matthieu Giraud (CRTI)
Sophie Girault (LS2N)
Pierre-Antoine Gourraud (CRTI)
Elodie Guidon (LS2N)
Carito Guziolowski (LS2N)
Julie Haguait (LS2N)
Géraldine Jean (LS2N)
Matilde Karakachoff (CHU)
Abdelhalim Larhlimi (LS2N)
Solena Le Scouarnec (l'institut du thorax)
Sébastien Leuillet (Biofortis)
Sophie Limou (CRTI)
Pierre Lindenbaum (l'institut du thorax)
Dimitri Meistermann (CRTI)
Stéphane Minvielle (CRCINA)
Jérémie Poschmann (CRTI)
Richard Redon (l'institut du thorax)
Jennifer Rondineau (CRCINA)
Bruno Saint-Jean (IFREMER)
Aurélien Serandour (CRCINA)
Floriane Simonet (l'institut du thorax)
Dominique Tessier (INRA BIA)
Raluca Teusan (l'institut du thorax)
Camille Trottier (LS2N)
Nicolas Vince (CRTI)

Our Partners and Sponsors

Redesign of iPPI-DB, a database for modulators of Protein-Protein Interactions

Rachel TORCHET[1,†], Alexandra MOINE-FRANEL[1,2,3,†], Hélène BORGES[1,2,3], Bryan BRANCOTTE[1], Olivia DOPPELT-AZEROUAL[1], Fabien MAREUIL[1], Hervé MÉNAGER[1], and Olivier SPERANDIO[1,2,3*]

[1] C3BI, Institut Pasteur 28 rue du Dr Roux, 7015 Paris
[2] Structural Bioinformatics Unit, Institut Pasteur 28 rue du Dr Roux, 7015 Paris
[3] CNRS UMR 3528, Institut Pasteur 28 rue du Dr Roux, 7015 Paris

† These authors have equally contributed
Corresponding Author: olivier.sperandio@pasteur.fr

iPPI-DB, for inhibitors of Protein-Protein Interaction DataBase, is a web application first released in 2012, which stores physicochemical and pharmacological data about PPI modulators and their targets. Users can query the database using either pharmacological criteria or chemical similarity with a user-defined query compound. The database is manually curated from the scientific literature and contains more than a thousand non-peptide inhibitors (iPPI) across 18 families of Protein-Protein Interactions. In the initial version, The chemical structures, as well as the physicochemical and the pharmacological profiles of these compounds and their targets, were extracted from the literature, computed and retrieved using numerous manual steps. This rather tedious procedure was seriously hindering the updates of the database.

For this project, we applied a combination of Agile methods and a User-Centered Design (UCD) approach to completely redesign the iPPI-DB web application. The main goal of this redesign is to focus on the needs of the user, ensuring that the end product fits the purpose, increasing the number of entries in the database and easing the query process. We adopted an iterative approach, interleaving successive series of design, tests and implementation steps, involving users in each iteration. This process, although it required an important involvement from the users during the project, has been extremely beneficial, as it allowed us to build a constructive dialog between scientists and the development team, and quickly validate or ask for corrections in the software.

The resulting web application provides a rich, robust, and innovative software environment to facilitate the growth and the maintenance of the database, and to query it using a highly intuitive and extremely powerful user interface.

Keywords Protein-Protein interaction, Database, Web interface, UX Design

Introduction

Pharmaceutical innovation is still impaired by the paucity of clinically testable targets and by the fact that only a few are successfully exploited in each therapeutic area [1]. This stands in sharp contrast with the number and diversity of roles of Protein-Protein Interactions. Indeed, with about 130,000 binary PPIs and possibly more just in humans [2], the development of drugs targeting these systems, represents a significant step toward expanding the druggable genome [3] and a possible leverage on the pharmacological modulation of disease-associated cellular pathways.

Historically, the design of small molecular drugs targeting PPIs has been extremely challenging, such that it seems there is a pharmacological cost to pay when choosing such target: selecting the right PPI and the right drug chemotype to work with. Yet, a growing number of successful examples is demonstrating on a daily basis that such an endeavor is accessible when deploying the necessary means: solid knowledge of the biological pathways around the chosen target and extensive medicinal chemistry to identify and optimize new chemical probes. The recent approval of Venetoclax [4], as a small molecular drug targeting some of the anti-apoptotic members of the Bcl-2 family in specific types of Lymphoma, is the perfect example of this virtuous combination that can be learned from.

In the light of this context, there is great value in storing in organized databases the knowledge coming from such studies. It is the purpose of several consortia such as ChemBL [5] or Pubchem [6]. In the field of PPIs, two databases have paved the way either by automatically deriving chemical data

from ChemBL like TIMBAL [7], or by focussing on co-crystallized compounds like the 2P2I database [8]. But none of them has a thorough modeling of the data, nor an elaborated web application to query them.

When we developed iPPI-DB, we decided to use a complementary approach vis a vis the existing iPPI databases. First, we chose to manually store a substantial number of metadata about the PPI targets, the chemical compounds and their activities, as well as the experimental assays that produced those activities. Second, we designed an intuitive web application to allow users to efficiently access the desired data. We first reported iPPI-DB in 2013 [9] and significant improvements were subsequently made to add more PPI modulators and targets, as well as a chemical similarity query mode [10].

In those initial versions of the database, the addition of new data was fastidious.

The new version of iPPI-DB has been created to ease the addition of new entries through a convivial interface and to improve query capabilities for data retrieval. The resulting database is available through a powerful web application that will enable users to query and navigate the contents of the database in a multitude of ways, but also a contribution wizard that guides them through the process of suggesting new entries.

We here describe the organisation and approaches we adopted during the project, focusing on two points in particular: the project management, and the user centered design methodologies we used.

2. Project management and coordination

The size and ambition of the iPPI-DB project required the combined efforts of a research group and a software engineering team, mobilizing an important number of different expertises over the course of two years. We describe here the main guidelines that we adopted to facilitate the development of this new version, which can all be linked to the Agile methods [11], a set of practices that have been increasingly adopted in Bioinformatics software development [12]. This approach focuses on collaboration, communication, and interaction between the different stakeholders.

2.1. Iterative approach

Given the complexity of this project, which includes contributions from experts in Structural Bioinformatics, Software and Database Development, User Interface Design, we structured the project around an iterative approach, interleaving successive series of design, software development, and user tests. Such iterations initially focused on specific topics, such as the analysis of the existing version of iPPI-DB, the redesign of the database, or the design or the web interfaces. This approach, although it required a significant involvement from the researchers, has been extremely beneficial, as it allowed us to build a constructive dialog between scientists and the development team, and quickly validate or correct the software when needed.

2.2. Supporting infrastructure

To support the development of this project, we heavily relied on the infrastructure provided by the IT department. It includes (1) a gitlab server that provides version control and sharing for the source code of the application and other capabilities [13] such as issue tracking and release management, (2) a virtual machine that hosts the system, (3) and a GitLab CI/CD server to automatically run tests and deploy the latest version. This infrastructure enables:

- Collaborative software development, enabling all partners to access the source code and contribute through source modifications or issue reports,
- Quality monitoring through continuous testing
- Automated deployment of new versions.

This infrastructure enabled us to adopt "DevOps" practices[1] to build and share the source code and accelerate the deployment of corrections and new features.

3. User-Centered Design Approach and Technical architecture

We adopted a User-Centered Design (UCD) approach where the needs of users are taken into account all along the project. During the early stages of this process, we focus on understanding user

[1] https://en.wikipedia.org/wiki/DevOps

behaviors, needs, and goals. This results in the identification of three kinds of users: (1) The common users who for instance search for compounds based on chemical similarity or PPI target; (2) The external contributors who suggest new entries based on data from in publications; and (3) The core curators who both enter new data and validate external contributions. Along with these different types of users we defined different goals, expectations, and needs. We used specific UX methodologies to answer questions and to design the different interfaces.

3.1. Query interface revisited

The query interface allows selecting and visualizing the different compounds available, based on biological, chemical, and pharmaceutical criteria. Based on the needs expressed by the users, providing a convivial and efficient interface was mandatory to extract the best of the available data.

We invited the users to *Six Up and One Up* workshops [14], to design mockups and prototypes for the different pages. During such workshops, each participant receives six templates of an empty screen and has to draw six different versions of the user interface. These different prototypes are then presented and compared. This allows identifying the redundant functionalities and needs, the eventual pain points and to bring up new design questions. As a result, all prototypes are summarized in a final accurate version. This method allows us to create a consensus prototype within two meetings. Although this approach has been previously applied in some bioinformatics projects [15], it remains largely unusual. The process is easy to set up with biologists and engineers and is highly effective. Using this methodology helps to generate many ideas over a short time, and gives to all participants the opportunity to contribute. Additionally, since these query capabilities already existed in the previous version of iPPI-DB, providing many different points of view enabled us to avoid retaining the same user interface with which many participants (but not all) were already familiar.

The revisited query interface now lets users select, filter and visualize the different compounds available, based on biological, chemical, and pharmaceutical criteria. It also allows to refine and combine multiple filters to build complex queries, share them easily as URLs with collaborators, and download corresponding data. Query results can be displayed with different layouts (thumbnails, list, or table), and all of them can be sorted according to different parameters.

3.2. New contribution interface

iPPI-DB was developed as a manually curated database from the scientific literature that contains the structure, some physicochemical characteristics, the pharmacological data and the profile of the PPI targets of several hundred modulators of protein-protein interactions. The main limitation of this system is the addition of new entries: the full process relies on several disconnected scripts, different languages and processes, namely in R, Java, Perl, python, starting from simple data sheets compiling the data. This makes the update process complex and error-prone.

We designed a new contribution interface to ease this task, in close collaboration between the developers and the users. To that end, we ran prototyping meetings with users, in order to create a convivial interface which uses a step-by-step approach to lower the workload for the experts. During these focus groups, we discussed openly between all the participants about the functionalities of the interface to provide simple, scaled down versions of it. The interfaces were designed as wireframes, and later refined as interactive prototypes, which users tested to validate the usability of the interface. We iterated our design through different rounds of tests.

The resulting *contribution interface* is wizard-based, i.e, it is a succession of screens that guide users to enter the data needed to populate the database. Users provide the architecture of the PPI complex(es), the chemical compounds tested for modulation, and the various assays in which those compounds were tested. The interface requests minimal participation from users to reduce the risks for errors and facilitate contributions: whenever contributors provide some information, the server automatically retrieves additional details from other reference databases, such as Pubmed, Uniprot, or the PDB.

Conclusion

The upcoming iPPI-DB web application provides a rich, robust, and innovative software environment to facilitate the growth and the curation of the database, as well as to query it using a highly interactive and powerful user interface.

Acknowledgements

We wish to thank the IT department of the Institut Pasteur, especially Eric Deveaud, Thomas Menard, Jean-Baptiste Denis, Emmanuel Guichard and Youssef Ghorbal, as well as Quang Tru Huynh from the Structural Bioinformatics unit for helping us set up the infrastructure for this project. We also thank Chemaxon (http://www.chemaxon.com) for providing us with JChem (JChem 19.8.0, 2019), and MarvinsJS (Marvin 19.13.0, 2019) licenses.

References

1. Teague SJ. Learning lessons from drugs that have recently entered the market. Drug Discov Today. 2011;16: 398–411. doi:10.1016/j.drudis.2011.03.003

2. Venkatesan K, Rual J-F, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, et al. An empirical framework for binary interactome mapping. Nat Methods. 2009;6: 83–90. doi:10.1038/nmeth.1280

3. Laraia L, McKenzie G, Spring DR, Venkitaraman AR, Huggins DJ. Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions. Chem Biol. 2015;22: 689–703. doi:10.1016/j.chembiol.2015.04.019

4. Leverson JD, Phillips DC, Mitten MJ, Boghaert ER, Diaz D, Tahir SK, et al. Exploiting selective BCL-2 family inhibitors to dissect cell survival dependencies and define improved strategies for cancer therapy. Sci Transl Med. 2015;7: 279ra40. doi:10.1126/scitranslmed.aaa4642

5. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45: D945–D954. doi:10.1093/nar/gkw1074

6. Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, et al. PubChem BioAssay: 2017 update. Nucleic Acids Res. 2017;45: D955–D963. doi:10.1093/nar/gkw1118

7. Higueruelo AP, Jubb H, Blundell TL. TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. Database . 2013;2013: bat039. doi:10.1093/database/bat039

8. Basse M-J, Betzi S, Morelli X, Roche P. 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. Database . 2016;2016. doi:10.1093/database/baw007

9. Labbé CM, Laconde G, Kuenemann MA, Villoutreix BO, Sperandio O. iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein–protein interactions. Drug Discov Today. 2013;18: 958–968. doi:10.1016/j.drudis.2013.05.003

10. Labbé CM, Kuenemann MA, Zarzycka B, Vriend G, Nicolaes GAF, Lagorce D, et al. iPPI-DB: an online database of modulators of protein–protein interactions. Nucleic Acids Res. Narnia; 2016;44: D542–D547. doi:10.1093/nar/gkv982

11. Kane DW, Hohman MM, Cerami EG, McCormick MW, Kuhlmman KF, Byrd JA. Agile methods in biomedical software development: a multi-site experience report. BMC Bioinformatics. 2006;7: 273. doi:10.1186/1471-2105-7-273

12. Prlić A, Procter JB. Ten simple rules for the open development of scientific software. PLoS Comput Biol. 2012;8: e1002802. doi:10.1371/journal.pcbi.1002802

13. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F da V, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLoS Comput Biol. 2016;12: e1004947. doi:10.1371/journal.pcbi.1004947

14. Bowles C, Box J. Undercover User Experience Design [Internet]. Pearson Education; 2010. Available:

https://market.android.com/details?id=book-9fwTj1japZcC

15. Karamanis N, Pignatelli M, Carvalho-Silva D, Rowland F, Cham JA, Dunham I. Designing an intuitive web application for drug discovery scientists. Drug Discov Today. 2018;23: 1169–1174. doi:10.1016/j.drudis.2018.01.032