

DOI: 10.1002/minf.201200070

PharmaTrek: A Semantic Web Explorer for Open Innovation in Multitarget Drug Discovery

Maria C. Carrascosa,^[a] Oriol L. Massaguer,^[a] and Jordi Mestres^{*[a]}**Keywords:** Semantic web · Chemogenomics · Target profile · Polypharmacology

The realisation that vast amounts of pharmacological data for small molecules are continuously being reported in bibliographic sources has promoted in recent years the rise of initiatives aiming at collecting, organising, and storing these data together with chemical structures. Today, there are numerous databases that connect hundreds of thousands of small molecules to thousands of biological responses of their interaction with macromolecules. Some of these repositories, such as GLIDA, PDSP, BindingDB, IUPHARdb, PubChem, ChEMBL, and DrugBank, make all data available in the public domain.^[1] In addition, some others, such as BioPrint, Integrity, Wombat, and GOSTAR, offer access to their data only through licensing from the respective commercial providers.^[2]

This wide diversity of sources does not facilitate direct access and interrogation of the entire contents covered by all of them. Integrating all these repositories into a single accessible resource is not trivial, mainly due to issues related with the use of different vocabularies and ontologies for the various domain entities which makes cross-referencing among multiple sources a challenging task.^[3] But even if some degree of integration is accomplished, managing and updating such an integrated framework in an efficient manner may require significant human resources and be extremely time consuming and difficult to fully automate.^[4] Managing chemical structures, for instance, involves taking into consideration a fair amount of detailed aspects such as salt formulation and isomerism (tautomerism, regioisomerism, and optical and geometrical isomerisms), and it has been reported that different molecular identifiers may actually lead to an essentially different number of unique chemical structures depending on the user criteria for defining uniqueness.^[5] On the other hand, managing pharmacological data across databases is also complicated, as one may encounter different values for the same molecule – protein interaction obtained from different laboratories, from the same protein but different species, or from the same protein and species but different settings and conditions.^[6]

In parallel, there have been some recent initiatives to provide some conceptual meaning to the connections established between objects from different domains, so the links are stored in such a way that become more understandable to computers. This is the main goal of applying

semantic web technology to drug discovery.^[7] Semantic web,^[8] also known as Web 3.0, is a web of data that provides tools to unify them in a consistent way and gives access to them through standardized query methods. The main difference with the so-called Web 2.0 is that instead of dealing with a huge amount of dispersed data that requires some level of human interpretation to understand it, data is integrated and conceptualised in a way that computers themselves can “understand” and extract new knowledge from them.

Several recent projects have implemented semantic web technologies in a life science environment. Among them, of mention are Bio2RDF,^[9] that codifies the contents of different public biological databases into a resource description framework (RDF), Linking Open Drug Data (LODD),^[10] that makes a similar task but focussed mainly on drug data, and Chem2Bio2RDF,^[11] that integrates small molecule and drug information with protein targets, genes, and pathways, and allows cross-source linking with LODD and Bio2RDF.

Along these lines, Open PHACTS is a recently funded European project that applies semantic web standards and technologies to create an integrated open pharmacological space (OPS) aiming at facilitating open innovation in drug discovery research.^[12] With this semantic approach, Open PHACTS aspires to solve some of the main bottlenecks of current data access and knowledge generation in drug discovery, namely, access to multiple disparate heterogenic information sources, lack of standards and common identifiers for domain entities, and ability to interrogate the system with complex research questions. At present, OPS offers access to ChEMBL v1.3,^[13] one of the largest public repositories of chemical structures annotated with pharmacological data that has recently integrated in it the contents of other individual sources. With respect to identifiers, vocabularies, and ontologies, OPS uses ConceptWiki,^[14]

[a] M. C. Carrascosa, O. L. Massaguer, J. Mestres
Research Programme on Biomedical Informatics (GRIB), IMIM
Hospital del Mar Research Institute and University Pompeu Fabra
Parc de Recerca Biomèdica, Doctor Aiguader 88, 08003
Barcelona, Catalonia, Spain
*e-mail: jmestres@imim.es

Re-use of this article is permitted in accordance with the Terms and Conditions set out at [http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1868-1751/homepage/2022_onlineopen.html](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1868-1751/homepage/2022_onlineopen.html).

Communication

Maria C. Carrascosa et al.

a collaborative knowledge resource for the life sciences that provides a mapping between scientific textual representations of concepts and database and ontology identifiers. Finally, it is envisaged that OPS provides the framework on which external applications may be developed to allow users to address complex research questions to the system and display the results in an interactive environment that facilitates knowledge extraction.

With this purpose in mind, we introduce PharmaTrek (<http://cgl.imim.es/pharmatrek>), an interactive semantic web explorer purposely designed for researchers in the field of multitarget pharmacology to address complex queries in a most simple and intuitive manner. Access to the RDF {chemical object}–{predicate}–{protein object} triple store of ChEMBL v1.3 is currently managed by an application webserver that retrieves data from OPS through a SPARQL endpoint, but also through an application programming interface (API) provided by the Open PHACTS system. A scheme of the application architecture used is provided in Figure 1.

As an example of the type of complex queries that can be addressed, we will ask PharmaTrek to retrieve all ligands having a $-\log(\text{Activity})$ value (Activity being generally defined here as any of the interaction types available in OPS, such as K_i , K_d , IC_{50} or EC_{50}) larger than or equal to 7.5 (that is, more potent than 31.62 nM) for coagulation factor Xa (EC 3.4.21.6) and being at least two orders of magnitude selective against trypsin (EC 3.4.21.4) and thrombin (EC 3.4.21.5), two phylogenetically related serine proteases. Af-

terwards, we will show how this first query can be further refined by adding other potentially relevant proteins that were not taken into consideration when defining the original target profile.

A typical query in a multitarget drug discovery project requires first to define an objective target profile. In order to do that, type the name of the first target in the prompt of the field located in the upper-left workspace, labeled as “Enter target name”. After typing the first characters of the target name, the prompt will start suggesting names of proteins matching that string. You can then press “Enter” to retrieve the list of suggested names in the space located right below the prompt. To obtain protein name suggestions, the field currently generates a SPARQL query that searches the text entered into the OPS repository. If your intended target appears in the list, you can then simply put the mouse pointer on top the target name and drag and drop it into the target profile basket located on the right hand side. You can then repeat the process for every target in your objective profile. In our case study, the target profile basket should contain the names of Thrombin, Trypsin I, and Coagulation factor X.

Once the target profile has been defined, you can click on the “Show heatmap” button and an interaction map will appear in the largest workspace available. At this stage, it contains 7488 molecules with activity data for any of the three targets in the profile. Now, you can apply affinity filters to each individual target to meet certain selectivity criteria. You can do that by just clicking on the arrow next to

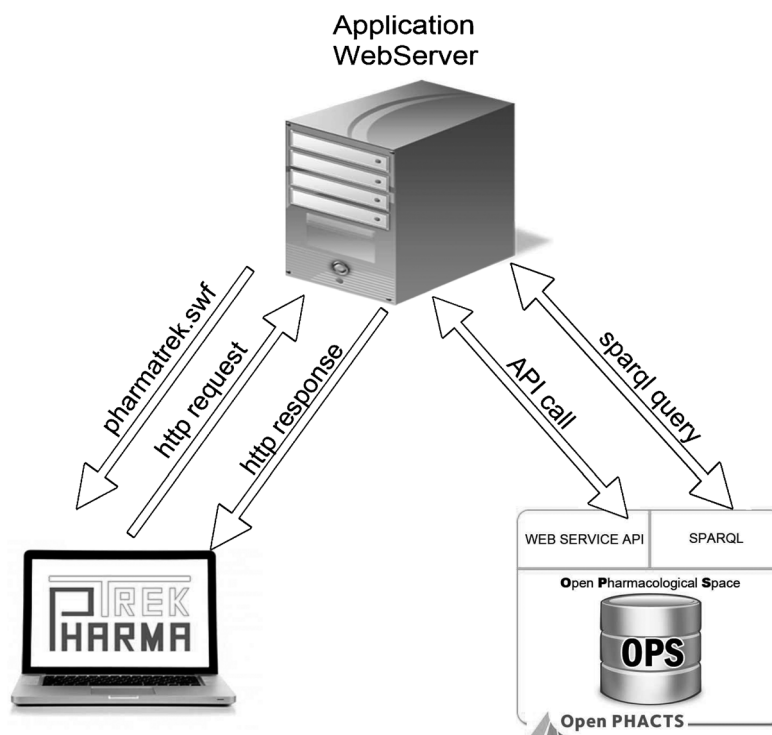


Figure 1. Application architecture.

the target name. By doing that, two fields will appear that will allow you to enter a minimum and a maximum activity value. In our case, since we are looking for potent and selective factor Xa inhibitors, we will enter a minimum affinity value of 7.5 for Coagulation factor X and maximum affinity values of 5.5 for both Thrombin and Trypsin and we will press the "Show heatmap" button again. Additional filters that affect all protein entries in the target profile basket can be defined in the "General filters" space. There are currently four general filters that can be defined to further refine your queries on protein species, interaction type, and general minimum and maximum affinity values.

The results of the query with individual target affinity filters are shown in Figure 2. As can be observed, the number of molecules meeting those affinity criteria has now been reduced to 3966 molecules, represented as rows in the interaction map. Note that summary information on the size of the heatmap (molecules and targets) and overall minimum and maximum affinity values can be found on the right-hand side of the heatmap, above which there is also an interactive small-size overall viewer that allows you to zoom in and out on different regions of the interaction map. The default colour gradation used in the heatmap is green for lack of information about the molecule – protein interaction, yellow for the minimum interaction value, and dark red for the maximum interaction value. Colour gradation is adapted as subsequent filters are applied and new minimum and maximum values are present. For the sake of convenience, one can also customise the colours of the

heatmap with the colour selectors that are located at the bottom-right corner of the heatmap.

However, what is currently shown in Figure 2 is still not the final answer to our intended query. To keep only those ligands that have interaction values with all the proteins defined in the target profile, you ought to click on the "connect" check box located right next to the "Show heatmap" button. Effectively, clicking on "connect" applies a logic AND to all targets and filters defined in the basket. This action results in a final number of 101 potent and selective factor Xa inhibitors, relative to thrombin and trypsin. Visual quantitative confirmation of potency and selectivity for each molecule, can be obtained by passing the mouse over the interactive heatmap. A tooltip will then appear with the pActivity value of the interaction between a ligand (in the row) and a target (in the column).

At this stage, the original question has been answered, and molecules meeting all potency and selectivity criteria across the target profile defined have been identified. However, one could go one step further and check whether any additional targets should be added to the original definition of the target profile on the basis of the information contained in OPS. In the bottom-right corner of the application, you will find an "Expand target space" check box that meets this precise need. By clicking on the "Expand target space" check box, PharmaTrek makes a request to expand the target space of the 101 potent and selective factor Xa ligands with any activity data on additional targets not included originally in the target profile defined in

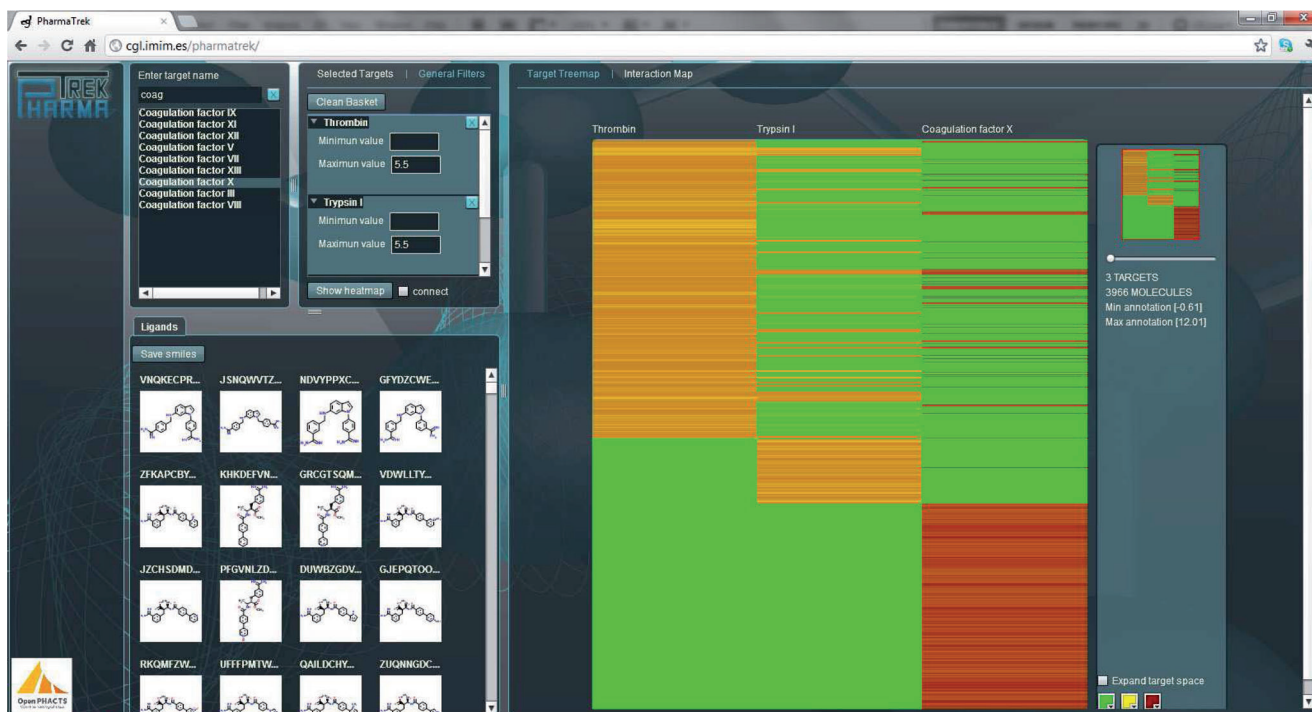


Figure 2. PharmaTrek layout showing the results of a query on ligands having a defined affinity profile on three serine proteases (see text for details).

the basket. In this particular case, interaction data for 16 additional targets are retrieved. Among them, tissue-type plasminogen activator is one of the targets showing a high degree of cross-pharmacology with the three targets in the original profile. Based on these findings, one may now decide that this target should be included in the objective target profile of the multitarget drug discovery project. To do that, simply put the mouse on top of the target name appearing in the labels of the columns (zoom in if necessary) and drag and drop the name into the target profile basket. You can then apply new filters and perform a new ligand extraction request.

At any point during a PharmaTrek session, the structures of all ligands contained in the interaction map appear in the "Ligands" workspace, right below the target profile basket. By clicking on the image of any chemical structure, a ligand card is displayed with all the information about the ligand present in OPS. One can always save the Smiles of all chemical structures present in the workspace by pressing the "Save smiles" button. This is a useful feature to import any ligand selection to another external application.

We have introduced PharmaTrek v1.0, a semantic web explorer of pharmacological space for open innovation in multitarget drug discovery. Other existing applications, such as SuperTarget, STITCH, DrugViz, and iPHACE, provide means to access and visualise drug-target interactions.^[15] PharmaTrek differs conceptually from those tools by the way the user submits complex multitarget queries to the single largest open pharmacology space available to date (ChEMBL v1.3) and visualises the results in an unique interactive manner that allows taking informed decisions on the original objective multitarget queries. Further development is currently underway in our laboratory.

Computational Methods

PharmaTrek is implemented using Flex 4.^[16] Flex is a free, open source framework for building and maintaining highly interactive, expressive Rich Internet Applications (RIA) that deploy consistently on all major browsers. Flex uses two languages to write applications: MXML^[17] and ActionScript.^[18] MXML is an XML markup language used mainly to lay out user interface components, but also to implement the visual aspects of an application. ActionScript is an object-oriented programming language. ActionScript 3.0 is designed to facilitate the creation of highly complex applications with large data sets and object-oriented reusable code bases. We also use FlashDevelop^[19] as Integrated Development Environment (IDE) to build our application. FlashDevelop is a free and open source (MIT license) code editor. Finally, PharmaTrek was developed following the Model-View-Controller^[20] design pattern to facilitate the reusability and maintainability of the application. Accordingly, the application is partitioned into three categories of com-

ponents: model components that encapsulate data and behaviors related to the data processed by the application, view components that define the application's user interface, and the user's view of the data, and controller components that handle data interconnectivity in the application.

Acknowledgements

This project was developed under the *Innovative Medicines Initiative Joint Undertaking Open PHACTS Project*, Grant Agreement Number 115191, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and members of the *European Federation of Pharmaceutical Industries and Associations (EFPIA)*.

References

- [1] a) Y. Okuno, A. Tamon, H. Yabuuchi, S. Nijima, Y. Minowa, K. Tonomura, R. Kunimoto, C. Feng, *Nucl. Acids Res.* **2008**, *36*, D907–D912; b) N. H. Jensen, B. L. Roth, *Comb. Chem. High Throughput Screen.* **2008**, *11*, 420–427; c) T. Liu, Y. Lin, X. Wen, R. N. Jorissen, M. K. Gilson, *Nucl. Acids Res.* **2007**, *35*, D198–D201; d) A. J. Harmar, R. A. Hills, E. M. Rosser, M. Jones, O. P. Buneman, D. R. Dunbar, S. D. Greenhill, V. A. Hale, J. L. Sharman, T. I. Bonner, W. A. Catterall, A. P. Davenport, P. Delagrangé, C. T. Dollery, S. M. Foord, G. A. Gutman, V. Laudet, R. R. Neubig, E. H. Ohlstein, R. W. Olsen, J. Peters, J. P. Pin, R. R. Ruffolo, D. B. Searls, M. W. Wright, M. Spedding, *Nucl. Acids Res.* **2009**, *37*, D680–D685; e) Y. Wang, E. Bolton, S. Dracheva, K. Karapetyan, B. A. Shoemaker, T. O. Suzek, J. Wang, J. Xiao, J. Zhang, S. H. Bryant, *Nucl. Acids Res.* **2010**, *38*, D255–D266; f) A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucl. Acids Res.* **2012**, *40*, D1100–D1107; g) D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, *Nucl. Acids Res.* **2008**, *36*, D901–D906.
- [2] a) C. M. Krejsa, D. Horvath, S. L. Rogalski, J. E. Penzotti, B. Mao, F. Barbosa, J. C. Migeon, *Curr. Opin. Drug Discov. Devel.* **2003**, *6*, 470–480; b) M. Olah, R. Rad, L. Ostopovici, A. Bora, N. Hadaruga, D. Hadaruga, R. Moldovan, A. Fulias, M. Mracec, T. I. Oprea, in *Chemical Biology: From Small Molecules to Systems Biology and Drug Design* (Eds: S. L. Schreiber, T. M. Kapoor, G. Wess), Wiley-VCH, New York, **2007**, pp. 760–786; c) C. Southan, P. Várkonyi, S. Muresan, *J. Cheminf.* **2009**, *1*, 10.
- [3] a) K. Degtyarenko, P. de Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcántara, M. Darsow, M. Guedj, M. Ashburner, *Nucl. Acids Res.* **2008**, *36*, D344–D350; b) N. Skunca, A. Altenhoff, C. Dessimoz, *PLoS Comput. Biol.* **2012**, *8*, e1002533; c) U. Visser, S. Abeyruwan, U. Vempati, R. P. Smith, V. Lemmon, S. C. Schürer, *BMC Bioinformatics* **2011**, *12*, 257; d) P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, M. A. Musen, *Nucl. Acids Res.* **2011**, *39*, W541–W545.
- [4] a) N. Baurin, R. Baker, C. Richardson, I. Chen, N. Foloppe, A. Potter, A. Jordan, S. Roughley, M. Parratt, P. Greaney, D. Morley, R. E. Hubbard, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 643–651; b) P. Tiikkainen, L. Franke, *J. Chem. Inf. Model.* **2012**, *52*, 319–326.

- [5] E. Gregori-Puigjané, R. Garriga-Sust, J. Mestres, *J. Comput. Chem.* **2011**, *32*, 2638–2646.
- [6] B. Zdrazil, M. Pinto, P. Vasanthanathan, A. J. Williams, L. Z. Balderud, O. Engkvist, C. Chichester, A. Hersey, J. P. Overington, G. F. Ecker, *Mol. Inf.* **2012**, *31*, 599–609.
- [7] D. J. Wild, Y. Ding, A. P. Sheth, L. Harland, E. M. Gifford, M. S. Lajiness, *Drug Discov. Today* **2012**, *17*, 469–474.
- [8] T. Berners-Lee, J. Hendler, *Nature* **2001**, *410*, 1023–1024.
- [9] F. Belleau, M. A. Nolin, N. Tourigny, P. Rigault, J. Morissette, *J. Biomed. Inform.* **2008**, *41*, 706–716.
- [10] M. Samwald, A. Jentzsch, C. Bouton, C. Kallesoe, E. Willighagen, J. Hajagos, M. Marshall, E. Prud'hommeaux, O. Hassanzadeh, E. Pichler, S. Stephens, *J. Cheminf.* **2011**, *3*, 19.
- [11] B. Chen, X. Dong, D. Jiao, H. Wang, Q. Zhu, Y. Ding, D. J. Wild, *BMC Bioinformatics* **2010**, *11*, 255.
- [12] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, C. T. Evelo, N. Blomberg, G. Ecker, C. Goble, B. Mons, *Drug Discov. Today* **2012**, <http://dx.doi.org/10.1016/j.drudis.2012.05.016>, in press.
- [13] M. P. Gleeson MP, A. Hersey, D. Montanari, J. Overington, *Nat. Rev. Drug Discov.* **2011**, *10*, 197–208.
- [14] *ConceptWiki, a semantic wiki for science*, <http://www.concept-wiki.org/>, last accessed on July 5th, **2012**.
- [15] a) N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, L. Macha, A. Eckert, M. K. Gilson, P. E. Bourne, R. Preissner, *Nucl. Acids Res.* **2012**, *40*, D1113–1117; b) M. Kuhn, D. Szklarczyk, A. Franceschini, C. von Mering, L. J. Jensen, P. Bork, *Nucl. Acids Res.* **2012**, *40*, D876–D880; c) B. Xiong, K. Liu, J. Wu, D. L. Burk, H. Jiang, J. Shen, *Bioinformatics* **2008**, *24*, 2117–2118; d) R. Garcia-Serna, O. Ursu, T. I. Oprea, J. Mestres, *Bioinformatics* **2008**, *26*, 985–986.
- [16] *Flex*, <http://www.adobe.com/products/flex.html>, last accessed on July 5th, **2012**.
- [17] *MXML*, http://help.adobe.com/en_US/Flex/4.0/UsingSDK/WS2db454920e96a9e51e63e3d11c0bf5f39f-7fff.html, last accessed on July 5th, **2012**.
- [18] *ActionScript*, http://help.adobe.com/en_US/as3/dev/index.html, last accessed on July 5th, **2012**.
- [19] *FlashDevelop*, <http://www.flashdevelop.org/>, last accessed on July 5th, **2012**.
- [20] S. Burbeck, in *Applications Programming in Smalltalk-80(TM): How to use Model-View-Controller (MVC)*, **1986**. <http://st-www.cs.uiuc.edu/users/smarch/st-docs/mvc.html>, last accessed on July 5th, **2012**.

Received: July 5, 2012
Accepted: July 13, 2012
Published online: August 7, 2012