

The EBI enzyme portal

Rafael Alcántara*, Joseph Onwubiko*, Hong Cao, Paula de Matos, Jennifer A. Cham, Jules Jacobsen, Gemma L. Holliday, Julia D. Fischer, Syed Asad Rahman, Bijay Jassal, Mikael Goujon, Francis Rowland, Sameer Velankar, Rodrigo López, John P. Overington, Gerard J. Kleywegt, Henning Hermjakob, Claire O'Donovan, María Jesús Martín, Janet M. Thornton and Christoph Steinbeck*

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

Received September 3, 2012; Revised October 19, 2012; Accepted October 20, 2012

ABSTRACT

The availability of comprehensive information about enzymes plays an important role in answering questions relevant to interdisciplinary fields such as biochemistry, enzymology, biofuels, bioengineering and drug discovery. At the EMBL European Bioinformatics Institute, we have developed an enzyme portal (<http://www.ebi.ac.uk/enzymeportal>) to provide this wealth of information on enzymes from multiple in-house resources addressing particular data classes: protein sequence and structure, reactions, pathways and small molecules. The fact that these data reside in separate databases makes information discovery cumbersome. The main goal of the portal is to simplify this process for end users.

INTRODUCTION

The number of registered databases in the NAR Database Issue during the last year has increased 7% to 1380 (1) with coverage of a great variety of fields and scopes. Within this landscape, users usually have to navigate from one resource to another in order to gather the information they need. That is where portals designed for a particular community can play an important role, simplifying the search process and presenting views of data from different databases.

Many portals already exist which integrate data from different resources. Biomart Central Portal (2) uses the generic Biomart federation technology to present combined data from virtually any biology database.

BioPortal (3) gives access to biomedical ontologies and tools to work with them. BioProject and BioSample (4) include metadata annotations in their registers, which allow the aggregated search of experimental data submitted to different NCBI, EBI and DDBJ databases.

Other portals are focused on more specific scientific fields. The Technology Portal of PSI-SBKB (5) integrates technology summaries and tools for structural biology, along with videos and social networks. InterStoreDB (6) provides plant phenotypic and genomic data from different sequence, crop and alignment databases. The IKMC web portal is a friendly interface to search many different BioMarts for data about targeted and trapped mouse knockout availability and structure (7).

Enzymes—biomolecules that catalyse specific chemical reactions—are of central importance in many fields of science, and their study is the foundation of the entire field of biochemistry. Enzymes have key regulatory and metabolic roles and are of large commercial and healthcare importance. The engineering of new functions, and the development of specific inhibitors for enzymes, is becoming more important in the post-genomic era. The EBI hosts a number of resources providing data on enzymes. One of them is IntEnz, the Integrated Enzyme relational database that provides the NC-IUBMB nomenclature and classification and facilitates downloads in different formats. Other EBI databases that provide enzyme data, such as UniProtKB and PDBe, are cross-referenced from IntEnz. However, these hyperlinks do not provide a good overview of the information users require and make it very difficult for users to navigate. In addition, there are other resources—Reactome, ChEBI, ChEMBL,

*To whom correspondence should be addressed. Tel: +44 1223 494 444; Fax: +44 1223 494 468; Email: rafael.alcantara@ebi.ac.uk
Correspondence may also be addressed to Joseph Onwubiko. Tel: +44 1223 494 640; Fax: +44 1223 494 468; Email: joseph@ebi.ac.uk
Correspondence may also be addressed to Christoph Steinbeck. Tel: +44 1223 492 640; Fax: +44 1223 494 468; Email: steinbeck@ebi.ac.uk
Present addresses:

Hong Cao, European Commission, European Research Council Executive Agency, Brussels, Belgium.

Gemma L. Holliday, Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA.

The authors wish it to be known that, in their opinion, the first five authors should be regarded as joint First Authors.

© The Author(s) 2012. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

Rhea—containing detailed data about the biochemistry of enzymes which could also be integrated for the benefit of users. It was with these aims in mind that the enzyme portal was launched.

The enzyme portal is a free resource that summarizes publicly available information about enzymes, including small-molecule chemistry, biochemical pathways and drug compounds. It provides a concise summary of information from:

- UniProt Knowledgebase (8);
- Protein Data Bank in Europe (PDBe) (9);
- Rhea, a database of enzyme-catalysed reactions (10);
- Reactome, a database of biochemical pathways (11);
- IntEnz, a resource with enzyme nomenclature information (12);
- ChEBI (13) and ChEMBL (14), which contain information about small-molecule chemistry and bioactivity;
- MACiE (15) for highly detailed, curated information about reaction mechanisms;
- EFO (16), the Experimental Factor Ontology, a system for annotation of experiments from which the enzyme portal retrieves disease-related information, concretely from children of its ‘disease’ entry.

The enzyme portal collates diverse information about enzymes and displays it in an organized overview. It covers many species, including mammals, invertebrates and plants, and provides a simple way to compare orthologues.

MATERIALS AND METHODS

User-centred design was applied to improve usability and meet the expectations of users

The portal was designed by following a user-centred design lifecycle (17), whereby decisions regarding the design of the portal were made based on evidence gathered from users representing our target audience, namely, scientists working in enzyme-related research, such as biomarker discovery, enzymology and drug discovery (see also Supplementary Methods).

The portal was implemented using Java technologies

We used Java technologies—Spring Web MVC, JAXB, JAX-WS—and several Web Service APIs either SOAP—EB-Eye, ChEBI, CiteXplore—or REST—UniProt,

The screenshot displays the EMBL-EBI Enzyme Portal search results for the query 'CFTR'. The search bar at the top contains 'CFTR' and a 'Search' button. Below the search bar, an example list includes 'sildenafil, Insulin receptor, Ceramide glucosyltransferase, Phenylalanine-4-hydroxylase, Cytochrome P450 3A4, CFTR, Q13423, REACT_1400.4'. The search results are displayed in a table with 12 results found for 'CFTR', showing results 1 to 10. The results are grouped by orthologues. The left sidebar shows search filters for Species and Chemical Compounds. The Species filter includes Human, Mouse, Rat, Baker's yeast, and African elephant. The Chemical Compounds filter includes (E)-4-oxonon-2-enal, 2'3'-cyclic AMP, ADP, and acetyl. The search results table shows the following entries:

Structure	Orthologue	Function	Synonyms	Disease	Species
	Cystic fibrosis transmembrane conductance regulator [Human]	Involved in the transport of chloride ions. May regulate bicarbonate secretion and salvage in epithelial cells by regulating the SLC4A7 transporter.	ATP-binding cassette sub-family C member 7; CFTR; Channel conductance-controlling ATPase; Cystic fibrosis transmembrane conductance regulator; Cystic-fibrosis membrane-conductance-regulating protein; Show more synonyms	cystic fibrosis	[Human] [Ma's night monkey] [Middle-African hedgehog] [Black-handed spider monkey] [Bovine] Show more species
	Enoyl-CoA delta isomerase 2, mitochondrial [Human]	Able to isomerize both 3-cis and 3-trans double bonds into the 2-trans form in a range of enoyl-CoA species. Has a preference for 3-trans substrates. (By similarity).	3,2-trans-enoyl-CoA isomerase; Acetylene-allene isomerase; D3,D2-enoyl-CoA isomerase; DBI-related protein 1; DRS-1; Show more synonyms		[Human] [Yeast] [Aspergillus clavatus (strain ATCC 1007 / CBS 513.65 / DSM 816 / NCTC 3887 / NRRL 1)] [Aspergillus fumigatus] [Aspergillus niger (strain CBS 513.88 / FGSC A1513)] Show more species
No structure available	Proteasome-activating nucleotidase 2 [Halobacterium halobium]	ATPase which is responsible for recognizing, binding, unfolding and translocation of substrate proteins into the archaeal 20S proteasome core particle. Is essential for opening the gate of the 20S proteasome via an interaction with its C-terminus, thereby allowing substrate entry and access to the site of proteolysis. Thus, the C-termini of the proteasomal ATPase function like a 'key' in a	Show more about function		[Halobacterium halobium] [Baker's yeast]
No structure available	Guanylate cyclase [Spiny dogfish]		GTP diphosphate-lyase (cyclizing); Guanyl cyclase; Guanylate cyclase; Guanylyl cyclase;		[Spiny dogfish]

Figure 1. Search results are grouped as orthologues. Note the paging (top right of the table) and the filter facets (left).

The screenshot shows the EMBL-EBI Enzyme Portal interface. At the top, there is a search bar with the text 'CFTR' and a 'Search' button. Below the search bar, an example list of enzymes is shown: sildenafil, insulin receptor, Ceramide glucosyltransferase, Phenylalanine-4-hydroxylase, Cytochrome P450 3A4, CFTR, Q13423, REACT_1400.4. The search results section shows '12 results found for CFTR, filtered to 1, displaying 1 - 1'. On the left, there is a 'Search Filters' panel with a 'Species:' dropdown menu. The 'Species:' menu is open, showing a list of species with checkboxes. 'Ma's night monkey [Aotus nancymaae]' is selected with a checked box. Other species listed include Human, Mouse, Rat, Baker's yeast, Aerobacter aerogenes, and African clawed frog. The main content area displays the details for 'Cystic fibrosis transmembrane conductance regulator [Ma's night monkey]'. It includes a 3D protein structure image, a 'Function' description, 'Synonyms', 'Disease', and 'Species' information.

Figure 2. Search results for 'CFTR' filtered to show only enzymes known in Ma's night monkey.

The screenshot shows the enzyme summary page for 'Cystic fibrosis transmembrane conductance regulator'. At the top, there is a breadcrumb trail: 'EBI > Enzyme Portal > Search for "CFTR" > Q07DV2 > P13569'. Below the breadcrumb trail, there is an 'ORGANISMS' dropdown menu set to 'Human Homo sapiens'. The main content area is titled 'Cystic fibrosis transmembrane conductance regulator' and includes sections for 'Enzyme', 'Function', 'EC Classification', 'Other names', 'Protein Sequence', 'Data Source', and 'Literature'. The 'Function' section describes its role in chloride ion transport. The 'EC Classification' is listed as 'Hydrolases > Acting on Acid Anhydrides > Acting on acid anhydrides; catalyzing transmembrane movement of substances > 3.6.3.49 - Channel-conductance-controlling ATPase'. The 'Protein Sequence' section states 'This sequence has 1480 amino acids' and provides a link to 'View Sequence in UniProt'. The 'Data Source' section mentions 'IntEnz & UniProt' and provides a brief description of these databases.

Figure 3. Enzyme summary. Orthologues can be selected using the 'organisms' drop-down menu at the top. Notice the breadcrumbs at the top left, which include links to the search results and any other orthologues visited previously.

Reactome, Rhea, ChEMBL, BioMart, DAS—to build the web application. XML schemas describe the underlying model.

For quick cross-referencing and building filters for search results, an Oracle database was populated with

cross-references to/from UniProtKB identifiers to other databases, including those not directly cross-referenced currently in the UniProt Knowledgebase such as ChEBI and ChEMBL. UniProtKB accessions are used throughout the enzyme portal as enzyme identifiers.

SEARCHING THE ENZYME PORTAL

Searches in the enzyme portal are based on the powerful EB-eye search engine (18), which indexes many EBI resources, updating for every release and provides a web service API.

From the website homepage, a free text search can use any relevant query terms, such as enzyme names, EC numbers, UniProtKB accessions, gene names and small-molecule names.

Search results are shown in a table (Figure 1) with orthologues grouped and split into separate pages where appropriate. The results list the enzyme name, a description of its function and any synonyms as well as the list of species in which the enzyme is found. If there are any related diseases, these will also be displayed.

Each result shows on its left side a fully coloured thumbnail of the protein structure or a greyed image if none is found for any of the grouped orthologues.

Search filters

On the left-hand side, there is a list of species, compounds and diseases which are related to the search results. Users can filter these simply by clicking the corresponding checkbox so that only enzymes matching the checked items will be shown, hiding the others.

Species

A list of species where enzymes included in the search results are found is shown so that users can narrow the search to any specie(s) of interest (Figure 2).

Compound

Any small molecules known to interact in some way with the enzymes in the list are shown here and can also be used as filters. This includes cofactors, activators, inhibitors and drugs.

Disease

Search results can also be filtered according to any diseases associated to them. For example, checking a

The screenshot shows the EMBL-EBI Enzyme Portal interface. At the top, there is a search bar with the text "Enter Text Here" and a "Find" button. Below the search bar, there are navigation tabs for "Databases", "Tools", "Research", "Training", "Industry", "About Us", and "Help". The main content area is titled "Enzyme Portal - Your portal to enzyme-related information at the EBI." and features a dropdown menu for "ORGANISMS" set to "Human (Homo sapiens)". The selected enzyme is "Cystic fibrosis transmembrane conductance regulator" (PDZ Domain of CAL). The page displays a 3D ribbon structure of the protein in green and blue. The left sidebar contains navigation icons for "Enzyme", "Protein Structure", "Reactions & pathways", "Small Molecules", "Disease", and "Literature". The main content area provides detailed information about the protein structure, including a description, method (solution nmr), dates (deposited, released, revised), and primary citation. The protein structure is identified as "Ribbon structure of 2lob".

Figure 4. 3D structure tab showing basic information about experimental models of the protein structure.

box labelled 'stroke' will display only those enzymes which have been related to this disease.

Note that several filters within the same section—species, compounds or diseases—have the effect of union of results (boolean OR), while filters from different sections result in intersection of results (boolean AND). For example, checking *Xenopus laevis* and *Drosophila melanogaster* will display only enzymes present in one species or the other; checking *Rattus norvegicus* and GMP3- will display only enzymes present in rat and known to interact with the nucleotide.

ENZYME DATA

Clicking on an enzyme name (or one of its orthologues) in the list of search results takes to the enzyme page, organized into tabs (on the left-hand side).

Enzyme summary

The first one (Figure 3) is the enzyme summary tab, which contains the description of the enzyme function, its classification in the EC hierarchy, any synonyms and some information about the protein sequence. An 'organisms' drop-down menu at the top can be used to switch between the known orthologues of the enzyme, from any of the available tabs.

While navigating search results and orthologues, a history of the users' navigation is kept in the form of breadcrumbs.

Throughout the portal, all data sources are acknowledged and linked, so that users can access in-depth information (such as the protein sequence from UniProtKB or the enzyme EC classification from IntEnz).

Protein 3D structure

The protein structure tab (Figure 4) shows any experimental 3D models of the enzyme. If there are several of them, any one can be selected from the drop-down menu, which indicates the number of structures available.

Only basic information on the model is displayed: again, users can navigate to the PDB web pages that describe the structure.

Reactions and pathways

The reactions and pathways tab (Figure 5) shows the biochemical reaction(s) catalysed by the enzyme—with the chemical structures of the participants linked to ChEBI—as well as any metabolic pathways in which the enzyme may be involved.

Only summarized information is shown in this tab. For more information—additional data about the reaction, reaction mechanisms and context within pathways—users are referred to the data source (in this case, Rhea,

The screenshot shows the EMBL-EBI Enzyme Portal interface. At the top, there is a search bar and navigation links. The main content area is titled 'Enzyme Portal - Your portal to enzyme-related information at the EBI.' Below this, there is a dropdown menu for 'ORGANISMS' set to 'Human (Homo sapiens)'. The enzyme being viewed is 'Cystic fibrosis transmembrane conductance regulator'. The 'Reactions & pathways' tab is active, showing the reaction: $\text{ATP} + \text{H}_2\text{O} \rightleftharpoons \text{ADP} + \text{H}^+ + \text{phosphate}$. Chemical structures for ATP, ADP, and phosphate are displayed. Below the reaction, there are links to 'View reaction in Rhea', 'View reaction in Reactome', and 'View reaction in MACIE'. A section titled 'Transmembrane transport of small molecules' lists pathways, and another section titled 'ABC-family proteins mediated transport' provides a detailed description of the ABC transporter family.

Figure 5. Reactions and pathways tab: chemical structures are hyperlinks to the corresponding entities in ChEBI. Additional information about the reaction available from Rhea, Reactome and MACIE is also linked from here. This tab also includes a list of pathways in with every reaction can be involved, including descriptions and graphics when available.

EMBL-EBI [Terms of Use](#) | [Privacy](#) | [Cookies](#)

Databases Tools Research Training Industry About Us Help [Site Index](#)

EBI > Enzyme Portal > Search for 'CFTR' > G07DV2 > P13569

Enzyme Portal - Your portal to enzyme-related information at the EBI.

ORGANISMS
Human
Homo sapiens

Cystic fibrosis transmembrane conductance regulator

Bioactive compounds that might bind to this enzyme

The following compound(s) have been classified as Bioactive compounds. They have been associated with this enzyme, but it is uncertain whether they bind to this enzyme.

CHEMBL452820 CHEMBL191797 CHEMBL1081471

Chiral

See all 198 bioactive compounds in ChEMBL

Drugs that might interact with this enzyme

The following compound(s) have been classified as drugs. They have been associated with this enzyme, but it is uncertain whether they interact with this enzyme.

bumetanide glyburide

4-Phenoxybenzoic acid in which the hydrogens ortho to the phenoxy group are substituted by butylamino and sulfamoyl groups. Bumetanide is a diuretic, and is used for treatment of oedema associated with congestive heart failure, hepatic and renal disease.
Formula: C17H20N2O5S

Formula: C23H28ClN3O5S

Figure 6. Small molecules that have been associated in some way to the enzyme, including drugs, inhibitors and activators.

MACiE and Reactome, respectively) with the provided hyperlinks when available.

Small molecules

The small molecules tab (Figure 6) includes any available information from UniProtKB about cofactors, activators, inhibitors and drugs, and also any bioactive compounds from the ChEMBL database which have been associated with the enzyme. The chemical structures are links to navigate to ChEBI (with its focus on structure and nomenclature) or ChEMBL (with its focus on bioactivity and function).

Diseases

The disease tab (Supplementary Figure S1) lists any diseases that are related to the selected enzyme—OMIM entries and MeSH terms cross-referenced from UniProtKB, translated into EFO identifiers, with a short description of the disease and how it relates to the enzyme.

Literature

The literature tab (Supplementary Figure S2) lists bibliographic citations relevant to the enzyme. Article titles are

linked to the EBI's CiteXplore bibliography database. If an abstract is available, clicking on the 'toggle abstract' link will show it.

Citations can be filtered according to the different aspects of the enzyme, i.e. the other tabs.

DISCUSSION

The enzyme portal integrates many resources, most of them hosted by EBI and also external ones such as BioPortal. Its main goal is to provide information about enzymes in a suitable format, with a usable interface designed for intended users. Instead of reinventing the wheel, it makes use of available and reliable resources to that end.

Although some of these resources already incorporate real semantics, others do not, which makes it difficult to extract meaningful information on enzymes using existing semantic web technologies. This portal fills these gaps using the well-known relationships between the EBI databases.

The EB-eye tool (18) provides gene and protein summaries for search results. The enzyme portal complements this with additional metabolic information: catalytic

activity, pathways, regulation by small molecules and related diseases. It also keeps a look and feel consistent with that of the EB-eye summaries. The original data sources are always acknowledged and linked from the enzyme portal pages. This portal offers useful overviews on enzymes, but users are referred to the original databases in order to get in-depth information.

The data provided by the enzyme portal are live in the sense that they are not stored in a data warehouse, but retrieved on demand from the different data sources. This way, maintenance is reduced to a minimum and the most recent data are guaranteed by relying on web services from the provider databases, which is an advantage over the data warehouse approach.

The enzyme portal is a one-stop shop for enzyme-related information in resources developed at the EBI. It accumulates this information and aims to present it to the scientist with a unified user experience. The enzyme portal team does not curate enzyme information and therefore is a secondary information resource or portal. At some point, a user interested in more detail will always leave its pages and refer to the information in the underlying primary database (UniProtKB, PDBe, etc.) directly.

BRENDA (19) is the most comprehensive resource about enzymes worldwide and has invested a great amount into the abstraction and curation of enzymes and their related information. BRENDA contains valuable information that cannot be found in the enzyme portal at the moment, such as substrate, kinetic, specificity, stability, application, disease-related and engineering data. As a primary resource, BRENDA could be a candidate for an information source for the enzyme portal in the future.

Future technical developments

Programmatic access through web services will be provided in the future, making use of the existing XML schemas defining the underlying object model. Other features demanded during the design process were customized downloads—users will be able to save search results and enzyme summaries in the format of their choice—and side-by-side enzyme comparison. Additional means to browse the data through enzyme classification, compound classification and disease annotation has also been requested.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1 and 2, Supplementary Methods and Supplementary References [20–22].

ACKNOWLEDGEMENTS

The authors wish to thank the invaluable help received from all the participants involved in the user-centred design process (workshops and user testing) for the design of the portal and also from the data providers.

FUNDING

The European Molecular Biology Laboratory (core funding). Funding for open access charge: European Molecular Biology Laboratory—European Bioinformatics Institute.

Conflict of interest statement. None declared.

REFERENCES

- Galperin, M.Y. and Fernández-Suárez, X.M. (2012) The 2012 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Res.*, **40**, D1–D8.
- Guberman, J.M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R., Chelala, C., Croft, D., Cros, A., Cutts, R.J. *et al.* (2011) BioMart Central Portal: an open database network for the biological community. *Database*, **2011**, bar041.
- Musen, M.A., Noy, N.F., Shah, N.H., Whetzel, P.L., Chute, C.G., Story, M.A., Smith, B. and NCBO team (2012) The National Center for Biomedical Ontology. *J. Am. Med. Inform. Assoc.*, **19**, 190–195.
- Barrett, T., Clark, K., Gevorgyan, R., Gorenkov, V., Gribov, E., Karsch-Mizrachi, I., Kimelman, M., Pruitt, K.D., Resenchuk, S., Tatusova, T. *et al.* (2012) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.*, **40**, D57–D63.
- Gifford, L.K., Carter, L.G., Gabanyi, M.J., Berman, H.M. and Adams, P.D. (2012) The Protein Structure Initiative Structural Biology Knowledgebase Technology Portal: a structural biology web resource. *J. Struct. Funct. Genomics*, **13**, 57–62.
- Love, C.G., Andongabo, A.E., Wang, J., Carion, P.W., Rawlings, C.J. and King, G.J. (2012) InterStoreDB: a generic integration resource for genetic and genomic data. *J. Integr. Plant Biol.*, **54**, 345–355.
- Ringwald, M., Iyer, V., Mason, J.C., Stone, K.R., Tadepally, H.D., Kadin, J.A., Bult, C.J., Eppig, J.T., Oakley, D.J., Briois, S. *et al.* (2011) The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.*, **39**, D849–D855.
- The UniProt Consortium. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **40**, D71–D75.
- Velankar, S., Alhroub, Y., Best, C., Caboche, S., Conroy, M.J., Dana, J.M., Fernandez Montecelo, M.A., van Ginkel, G., Golovin, A., Gore, S.P. *et al.* (2012) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **40**, D445–D452.
- Alcántara, R., Axelsen, K.B., Morgat, A., Belda, E., Coudert, E., Bridge, A., Cao, H., de Matos, P., Ennis, M., Turner, S. *et al.* (2012) Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **40**, D754–D760.
- Haw, R. and Stein, L. (2012) Using the reactome database. *Curr. Protoc. Bioinformatics*, Chapter 8, unit8.7.
- Alcántara, R., Ast, V., Axelsen, K.B., Darsow, M., de Matos, P., Ennis, M., Morgat, A. and Degtyarenko, K. (2007) IntEnz. Molecular Biology Database Collection entry number 508. *Nucleic Acids Res.*
- de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S. and Steinbeck, C. (2010) Chemical entities of biological interest: an update. *Nucleic Acids Res.*, **38**, D249–D254.
- Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. *et al.* (2012) ChEMBL: a large-scale bioactivity database for chemical biology and drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
- Holliday, G.L., Andreini, C., Fischer, J.D., Rahman, S.A., Almonacid, D.E., Williams, S.T. and Pearson, W.R. (2012) MACiE: exploring the diversity of biochemical reactions. *Nucleic Acids Res.*, **40**, D783–D789.
- Malone, J., Holloway, E., Adamusiak, T., Kapushesky, M., Zheng, J., Kolesnikov, N., Zhukova, A., Brazma, A. and Parkinson, H. (2010)

- Modeling sample variables with an experimental factor ontology. *Bioinformatics*, **26**, 1112–1118.
17. Pavelin,K., Cham,J.A., de Matos,P., Brooksbank,C., Cameron,G. and Steinbeck,C. (2012) Bioinformatics meets user-centred design: a perspective. *PLoS Comput. Biol.*, **8**, e1002554.
 18. Valentin,F., Squizzato,S., Goujon,M., McWilliam,H., Paern,J. and López,R. (2010) Fast and efficient searching of biological data resources—using EB-eye. *Brief. Bioinform.*, **11**, 375–384.
 19. Scheer,M., Grote,A., Chang,A., Schomburg,I., Munaretto,C., Rother,M., Söhngen,C., Stelzer,M., Thiele,J. and Schomburg,D. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.*, **39**, 670–676.
 20. Pruitt,J. and Adlin,T. (2006) *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*. Morgan Kaufmann, San Francisco, CA.
 21. Gray,D., Brown,S. and Macanujo,J. (2010) *Game Storming: A Playbook for Innovators, Rulebreakers, and Changemakers*. O'Reilly, Sebastopol, CA.
 22. Snyder,C. (2003) *Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces*. Morgan Kaufmann, San Francisco, CA.