

Original Article

# An update on the Enzyme Portal: an integrative approach for exploring enzyme knowledge

S. Pundir\*, J. Onwubiko, R. Zaru, S. Rosanoff, R. Antunes, M. Bingley, X. Watkins, C. O'Donovan, and M. J. Martin

EMBL– European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

\*To whom correspondence should be addressed. E-mail: spundir@ebi.ac.uk

Edited by Dagmar Ringe

Received 17 October 2016; Revised 13 December 2016; Editorial Decision 14 December 2016; Accepted 14 December 2016

## Abstract

Enzymes are a key part of life processes and are increasingly important for various areas of research such as medicine, biotechnology, bioprocessing and drug research. The goal of the Enzyme Portal is to provide an interface to all European Bioinformatics Institute (EMBL-EBI) data about enzymes (de Matos, P., *et al.*, (2013), *BMC Bioinformatics*, **14** (1), 103). These data include enzyme function, sequence features and family classification, protein structure, reactions, pathways, small molecules, diseases and the associated literature. The sources of enzyme data are: the UniProt Knowledgebase (UniProtKB) (UniProt Consortium, 2015), the Protein Data Bank in Europe (PDBe), (Valenkar, S., *et al.*, *Nucleic Acids Res.* 2016; **44**, D385–D395) Rhea—a database of enzyme-catalysed reactions (Morgat, A., *et al.*, *Nucleic Acids Res.* 2015; **43**, D459–D464), Reactome—a database of biochemical pathways (Fabregat, A., *et al.*, *Nucleic Acids Res.* 2016; **44**, D481–D487), IntEnz—a resource with enzyme nomenclature information (Fleischmann, A., *et al.*, *Nucleic Acids Res.* 2004 **32**, D434–D437) and ChEBI (Hastings, J., *et al.*, *Nucleic Acids Res.* 2013) and ChEMBL (Bento, A. P., *et al.*, *Nucleic Acids Res.* 2014 **42**, 1083–1090)—resources which contain information about small-molecule chemistry and bioactivity. This article describes the redesign of Enzyme Portal and the increased functionality added to maximise integration and interpretation of these data. Use case examples of the Enzyme Portal and the versatile workflows its supports are illustrated. We welcome the suggestion of new resources for integration.

**Key words:** Enzymes, integration, proteins, search, services

## Introduction

Enzymes play a vital role in all life processes and are used extensively in biomedicine and biotechnology. Information about enzymes can be found in various disparate resources, each of which have been built with different communities in mind. This makes exploration of enzyme knowledge cumbersome. Researchers may not always be aware of the data available for their specific requirements or in which resources they can best access these data. Hence they might miss out on potentially valuable information. The Enzyme Portal brings all of the relevant European Molecular Biology Laboratory (EMBL)-EBI information together in one place—making it a unique resource for biomedical and industrial researchers. It has recently been redesigned with an improved interface to enable and enhance such user

workflows. The back-end has also been updated to provide better performance and allow easier integration of further resources. This article describes the methodology followed for the redesign and the key new functionality. Highlights include comprehensive enzyme summaries, enzyme comparison, sequence search and search entry points by disease, pathway, taxonomy and enzyme classification.

## Materials and Methods

### UCD for user interface improvements

The Enzyme Portal was initially developed in 2012 following a User-Centred Design (UCD) process (de Matos *et al.*, 2013) to

ensure that user requirements were understood from the early stages onwards. The redesign began with the evaluation of how user requirements had moved forward since then as well as what we could do to improve and optimise the Enzyme Portal's functionality. The process followed a continuation of the UCD approach and involved consultation with users from the different communities that form the Enzyme Portal's target audience. The user groups identified for consultation were enzymologists, drug discovery scientists, immunologists, biochemists, biocurators and researchers working on enzymes. Representative users were selected from both academia and industry to help balance our findings. Mockups and prototypes were created to use with techniques such as click testing, usability testing and impression testing. Multiple design options for some key features were evolved to the final design specifications through iteration rounds. As development progressed, we continued to validate design elements and decisions with users at key stages. This process highlighted the following areas for improvement and development.

- (i) Providing ways of accessing and searching the data for users coming from different perspectives such as diseases, pathways, EC hierarchy and taxonomy.
- (ii) Facilitating the download of customised search results.
- (iii) Altering the search results to provide an overview of all the proteins available for a particular enzymatic activity.
- (iv) Enhancing the enzyme summary with a sequence feature overview.

## Technology Brief

Integrating data from different resources are a challenging task. To meet this challenge, the Enzyme Portal employs a lightweight architecture consisting of a core database of enzyme metadata including enzyme function and cross-references to relevant source databases. Based on the cross-references, the Enzyme Portal utilises resource APIs to retrieve on-the-fly summaries of specialised data such as structure, chemistry and literature data. This enables the representation of the data in the same way as the source databases, allowing an easy transition to the sources where needed. This greatly reduces the burden of synchronisation with the underlying data. Java and related open source technologies/frameworks such as Spring

Framework (data and web MVC), Java Persistence API (Hibernate), QueryDSL and web technologies (D3, AngularJS and BioJS) were used in the development of the Enzyme Portal.

## Enzyme Portal architecture

The Enzyme Portal architecture can be described in the following three steps:

### Enzyme Portal database

Key enzyme data regarding function and cross-references are collated from selected publicly available resources (UniProt Knowledgebase (UniProtKB), Protein Data Bank in Europe (PDBe), Rhea, Reactome, IntEnz, ChEBI and ChEMBL) into a core Enzyme Portal database. Some of the data are sourced via a direct database connection (e.g. UniProtKB), some are collected via web service requests to the resources and the remainder of the enzyme metadata are sourced by parsing provided enzyme data flat files, as shown in Fig. 1. Using these key data and cross-references, the Enzyme Portal database is able to establish relationships between enzyme-related data from different resources.

### Data indexes

From the database, XML data are generated and provided to the EMBL-EBI Search (Squizzato *et al.*, 2015) for indexing, as shown in Fig. 1. The main purpose of the EBI search here is to provide a full-text search service for the Enzyme Portal.

### Web interface for enzyme information

The web interface at <http://www.ebi.ac.uk/enzymeportal/> is a single entry point that accepts a user's request, analyses it and delivers up-to-date enzyme information.

The search result is presented to the end user in simple and navigable web pages. When the user drills down to a specific enzyme entry, the Enzyme Portal makes web service requests to APIs of the enzyme-related resources, as shown in Fig. 1, based on the core data relationships established in the Enzyme Portal database. These data are collated into a single entry and presented to the user.

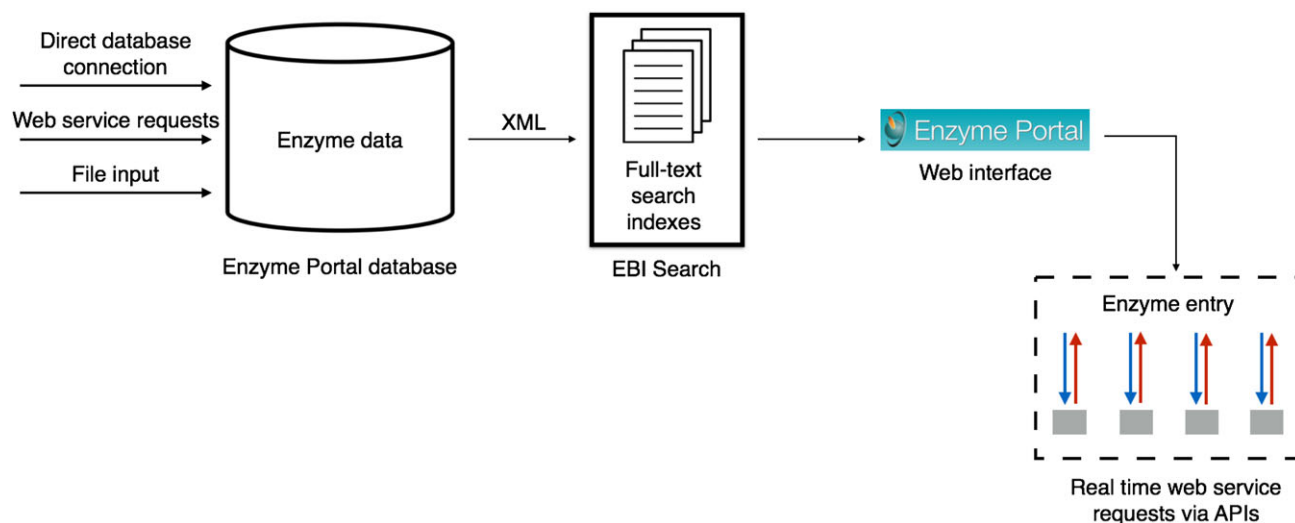


Fig. 1 Enzyme Portal architecture.

## Results

The Enzyme Portal can be accessed through <http://www.ebi.ac.uk/enzymeportal/> with an updated design and home page, as shown in Fig. 2. It provides a unified search for enzymes, specific browsing options by diseases/enzyme classification/taxonomy/pathways, a sequence search tool, as shown in Fig. 2. It also provides a basket that allows you to compare two enzyme entries side by side as shown in Fig. 3. Through the Enzyme Portal search workflow, you can find enzyme entries that compile information about associated proteins by organism, so that you can switch your view by selecting from a list of orthologs for the specific entry. The enzyme entry provides a summary of the enzyme function and sequence features overview, protein structure, reaction and pathways, small molecules, associated diseases and associated literature.

## Enzyme Portal usage examples

### Searching for an enzyme associated with an inhibitor in order to find other possible inhibitors

The Enzyme Portal allows you to carry out a free text search using terms such as an enzyme name and inhibitor. Erlotinib is a well-known inhibitor of Epidermal Growth Factor (EGF) receptor (UniProtKB P00533) used to treat non-small cell lung cancer, pancreatic cancer and several other types of cancer (Schettino *et al.*, 2008). It would be of scientific interest to find information about the EGF receptor, reviewing Erlotinib as an inhibitor and discovering other inhibitors for the EGF receptor. The workflow would begin with searching for Erlotinib in the main search box on the home page. The results page displays the top result as the Receptor protein-tyrosine kinase enzyme. Clicking on the enzyme name

The screenshot shows the Enzyme Portal home page with the following elements:

- Search Bar:** A search box with the text "search for enzymes" and examples: "sildenafil, mTOR, cathepsin, P27361, pyruvate kinase". A "Search" button is next to it.
- Navigation:** "Home", "FAQ", "About Enzyme Portal", "Basket (0)", "Feedback", and "Video" links.
- Specific browsing options:** Four buttons: "Diseases", "Enzyme Classification", "Taxonomy", and "Pathways".
- Sequence search (BLAST):** A button labeled "Sequence search (BLAST)".
- Content Sections:**
  - About Enzyme Portal:** "The Enzyme Portal integrates publicly available information about enzymes, such as small-molecule chemistry, biochemical pathways and drug compounds. See walk-through videos to learn more more..."
  - Enzyme Portal Resources:** "To build its reports, the Enzyme Portal uses data from UniProtKB, PDB, ChEMBL, ChEBI, REACTOME, Rhea, IntEnz and many more resources. more..."
  - Technical documents:** "The Enzyme Portal is an open source project developed at the EMBL-EBI and the source code is freely available and can be downloaded from GitHub, an online project hosting service. Additionally, the technical documentation (Javadoc, project information) is also available online."
  - Tools:**
    - Protein sequence search:** "search Enzyme Portal with a sequence by clicking on the Sequence search link in the search area."
    - Compare enzymes:** "the Enzyme Portal basket provides a tool for comparing two enzymes. Select your enzymes from search results and add them to your basket to access the tool."
- Footer:**
  - Services:** By topic, By name (A-Z), Help & Support
  - Research:** Overview, Publications, Research groups, Postdocs & PhDs
  - Training:** Overview, Train at EBI, Train outside EBI, Train online, Contact organisers
  - Industry:** Overview, Members Area, Workshops, SME Forum, Contact Industry programme
  - About us:** Overview, Leadership, Funding, Background, Collaboration

Fig. 2 Updated Enzyme Portal home page.

The screenshot shows the "Comparing enzymes" page with the following details:

Pyruvate kinase II (Escherichia coli (strain K12))	Pyruvate kinase I (Escherichia coli (strain K12))
<b>Name</b> Pyruvate kinase II	<b>Name</b> Pyruvate kinase I
<b>Species</b> Escherichia coli (strain K12) (Escherichia coli (strain K12))	<b>Species</b> Escherichia coli (strain K12) (Escherichia coli (strain K12))
<b>Function</b> No data available.	<b>Function</b> No data available.
<b>EC classification</b> EC 2.7.1.40	<b>EC classification</b> EC 2.7.1.40
<b>Sequence</b> 480 amino acids.	<b>Sequence</b> 470 amino acids.

Below the table, there is a "Compare protein sequences" button. Under the "Protein structures" section, a 3D ribbon diagram of the Pyruvate kinase I structure is shown, with the following identifiers: 1e0t R292D MUTANT OF E. COLI PYRUVATE KINASE, 1e0u STRUCTURE R271L MUTANT OF E. COLI PYRUVATE KINASE, 1pkv PYRUVATE KINASE FROM E. COLI IN THE T-STATE, and 4yng Twinned pyruvate kinase from E. coli in the T-state.

Fig. 3 Comparing enzymes through the basket.

expands the row to show the individual associated entries. An arrow pointing downwards indicates that the result row has been expanded, as shown in Fig. 4.

In the expanded result row for Receptor protein-tyrosine kinase, we see two associated proteins, including the EGF receptor. Clicking on the EGF receptor protein leads us to the next view consisting of a detailed entry page. This page has tabs on the left hand side that present information about the enzyme, protein structures, reactions and pathways, small molecules, diseases and literature about this enzyme entry as shown in Fig. 5.

We click on the small Molecules tab as we are interested in finding out more about potential inhibitors. Here, we find 12 inhibitors displayed for this enzyme and a link to expand this view to see all 169 inhibitors found by the Enzyme Portal. Clicking on this link displays all inhibitors found, including Erlotinib as shown in Fig. 6.

#### Searching for information about an enzyme involved in a pathway

The Enzyme Portal allows exploration from a specific area of interest including pathways, disease, taxonomy and EC number. This is an example for users interested in finding information about enzymes involved in xenobiotic processes (Hakkola et al., 1998). Clicking on the Pathways box on the home page (shown in Fig. 2) targets the search for enzymes involved in such pathways. In the search box within the resulting Pathways page, type xenobiotics and click on the auto-complete suggestion, as shown in Fig. 7.

This brings us to enzymes associated with the xenobiotic pathway. The 'Enzyme family' filter on the left hand side of the page shows that all enzyme results found are oxidoreductases. Clicking on the row with the top enzyme result 'Cholesterol 25-hydroxylase' expands the result row to show associated protein Cytochrome P450 2C9 as shown in Fig. 8.

The screenshot shows the Enzyme Portal interface. At the top, there is a search bar with the text "search for enzymes" and a search button. Below the search bar, there are navigation links: Home, FAQ, About Enzyme Portal, Basket (0), Feedback, and Video. The main heading is "Enzyme Results" followed by "1 enzymatic activities found for 'erlotinib'". On the left, there are "Active filters" and an "Organism" section with a search box and a list of organisms. The main table has columns: Name, Protein Hits, Enzyme Family, EC, Catalytic Activity, and Species. The first row is for "Receptor protein-tyrosine kinase" with 2 protein hits. The "Associated Proteins" section is expanded, showing "Receptor tyrosine-protein kinase erbB-2" and "Epidermal growth factor receptor".

Name	Protein Hits	Enzyme Family	EC	Catalytic Activity	Species
Receptor protein-tyrosine kinase	2	Transferases	2.7.10.1	[ATP + a [protein]-L-tyrosine = ADP + a [protein]-L-tyrosine phosphate.]	<ul style="list-style-type: none"> <li>Human</li> <li>Mouse</li> <li>Fruit fly</li> <li>Rat</li> <li>golden-collared manakin</li> <li>287 others</li> </ul>

**Associated Proteins:**

- Receptor tyrosine-protein kinase erbB-2 (Homo sapiens)
- Epidermal growth factor receptor (Homo sapiens)

Fig. 4 Results for a keyword search with 'Erlotinib' showing a result for Receptor protein-tyrosine kinase expanded to show associated proteins.

The screenshot shows the detailed entry page for the Epidermal growth factor receptor. The page title is "Epidermal growth factor receptor" and the Enzyme ID is "P00533". The "Organism" is set to "Human (Homo sapiens)". The "Function" section describes the receptor's role in signaling cascades. The "EC Classification" is "Transferases > Transferring phosphorus-containing groups > Protein-tyrosine kinases > 2.7.10.1 - Receptor protein-tyrosine kinase". The "Other names" section lists "Receptor tyrosine-protein kinase erbB-1". The "Protein Sequence" section shows the sequence length (1210 amino acids) and a link to "View Sequence in UniProt". Below the sequence, there are visualizations for "Domains & sites", "Molecule processing", "Post translational modifica...", and "Sequence information".

Fig. 5 Enzyme entry page for EGF Receptor showing tabs on the left hand side.

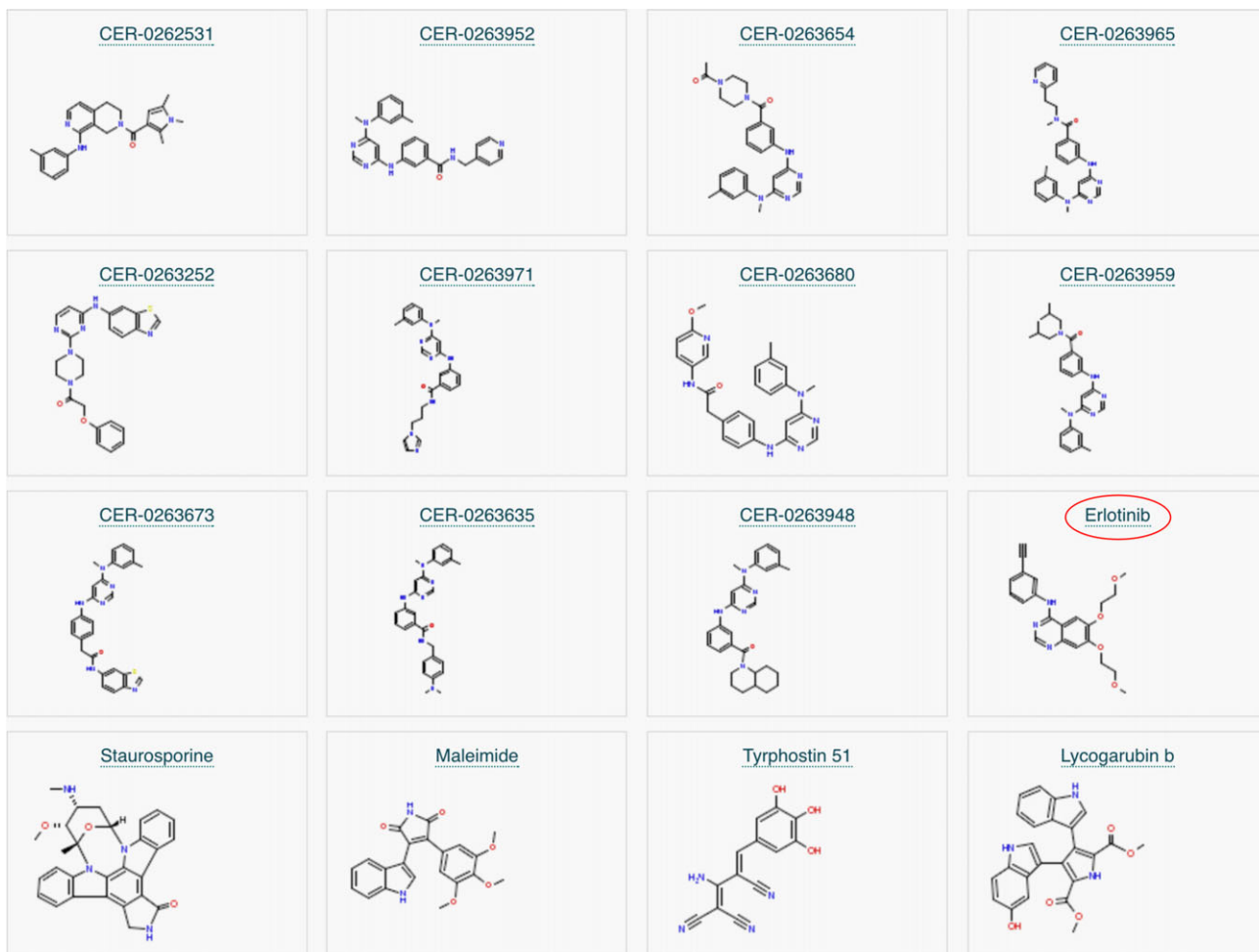


Fig. 6 Enzyme entry for the EGF receptor in the Enzyme Portal with small molecules listed including inhibitors.

Fig. 7 Browsing Enzyme Portal by pathways and searching for xenobiotics within the pathways page.

## Enzyme Results

11 enzymatic activities found for "xenobiotics"

Page 1 of 2

Feedback

**Active filters**

**Organism**

Search for Organism

- Homo sapiens (11)
- Sus scrofa (5)
- Rattus norvegicus (4)
- Mus musculus (4)
- Ixodes scapularis (3)
- Cryptolagus curticulus (3)
- Pediculus humanus corporis (3)
- Leptospira interrogans (2)
- Desulfosporosinus acididurans (2)
- Peptococcaceae bacterium CEB3 (2)
- Mycobacterium smegmatis MKD8 (2)
- Mesocricetus auratus (2)
- Bifidobacterium bifidum DSM 23969 (2)
- Bifidobacterium catenulatum DSM 23973 (2)
- Canis lupus familiaris (2)
- Macaca fascicularis (2)

**Enzyme Family**

- Oxidoreductases (11)

Name	Protein Hits	Enzyme Family	EC	Catalytic Activity	Species
Cholesterol 25-hydroxylase	1	Oxidoreductases	1.14.99.38	[Cholesterol + AH(2) + O(2) = 25-hydroxycholesterol + A + H(2)O.]	<ul style="list-style-type: none"> <li>o Human</li> <li>o Mouse</li> <li>o Rat</li> <li>o Pig</li> </ul>
<div style="border: 1px solid #ccc; border-radius: 50%; padding: 5px; display: inline-block;"> <b>Associated Proteins:</b>            Cytochrome P450 2C9         </div>					
(R)-limonene 6-monoxygenase	2	Oxidoreductases	1.14.13.80	[(+)-R]-limonene + NADPH + O(2) = (+)-trans-carveol + NADP(+) + H(2)O.]	<ul style="list-style-type: none"> <li>o Human</li> </ul>
Taurochenodeoxycholate 6-alpha-hydroxylase	1	Oxidoreductases	1.14.13.97	[(1) Taurochenodeoxycholate + NADPH + O(2) = taurohydrocholate + NADP(+) + H(2)O. (2) Lithocholate + NADPH + O(2) = hydroxychocholate + NADP(+) + H(2)O.]	<ul style="list-style-type: none"> <li>o Human</li> <li>o Body louse</li> <li>o Lettuce bottom rot fungus</li> <li>o Pig</li> </ul>
4-nitrophenol 2-hydroxylase	1	Oxidoreductases	1.14.13.n7	[4-nitrophenol + NADPH + O(2) = 4-nitrocatechol + NADP(+) + H(2)O.]	<ul style="list-style-type: none"> <li>o Human</li> <li>o Mouse</li> <li>o Rat</li> <li>o Dog</li> <li>o Crab-eating macaque</li> <li>o 6 others</li> </ul>
Cholestanetriol 26-	1	Oxidoreductases	1.14.15.15	5-beta-cholestane-3-alpha,7-alpha,12-alpha-	<ul style="list-style-type: none"> <li>o Human</li> </ul>

Fig. 8 Enzyme results for xenobiotics pathway search.

This enzyme might be present in 4 pathway(s):

**Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE)**

Similar to the lipoxygenases, cytochrome P450 (CYP) enzymes catalyse the hydroxylation and epoxidation of arachidonic acid. However, whereas lipoxygenases use an active non-heme iron to abstract hydrogen directly from arachidonic acid, CYPs contain a heme-iron active site that oxidizes its substrate by a different mechanism. They hydroxylate arachidonic acid between C-5 and C-15 to produce lipoxygenase-like hydroxyeicosatetraenoic acids (HETEs) and add a hydroxyl moiety to the sp<sup>3</sup>-hybridized omega-carbons to form a unique class of HETEs. The transfer of oxygen to the unstable arachidonic acid intermediate terminates the reaction by forming HETE or epoxy-eicosatrienoic acid (EETs), respectively (Capdevila et al. 2000, Buczynski et al. 2009, Vance & Vance 2008).

[View pathway in Reactome](#)

**Xenobiotics**

Of the 50 microsomal CYPs, 15 act on xenobiotics. They all possess wide substrate specificity to cater for most foreign compounds that find their way into the body.

[View pathway in Reactome](#)

**Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET)**

The epoxidation of arachidonic acid by cytochrome P450s (CYPs) results in the formation of unique bioactive lipid mediators termed epoxyeicosatrienoic acids (EETs). Each double bond has been shown to be susceptible to oxidation, resulting in 5,6-EET, 8,9-EET, 11,12-EET, and 14,15-EET. The majority of the EET biological activities are diminished by the hydrolysis to the corresponding dihydroxyeicosatrienoic acids (DHET) (Capdevila et al. 2000, Buczynski et al. 2009, Vance & Vance 2008).

[View pathway in Reactome](#)

**CYP2E1 reactions**

CYP2E1 can metabolize and activate a large number of solvents and industrial monomers as well as drugs. This quality of CYP2E1 may make it an important determinant of human susceptibility to the toxic effects of industrial and environmental chemicals. Typical CYP2E1 substrates include acetaminophen, benzene, CCl<sub>4</sub>, halothane, ethanol and vinyl chloride. CYP2E1 contributes to oxidative stress by producing oxidising species called reactive oxygen species (ROS) which can lead to damage to mitochondria, DNA and initiate lipid peroxidation or even cell death.

[View pathway in Reactome](#)

Fig. 9 Pathway data from the Cytochrome P450 2C9 enzyme entry.

We click on the associated protein entry Cytochrome P450 2C9 to find more details about it. The entry page includes on the left a tab for reactions and pathways that, once selected, contains a summary of the pathways. This includes the xenobiotics pathway as well as useful information about the three other pathways that this enzyme is involved in, as shown in Fig. 9.

### Running a sequence search

The Enzyme Portal offers a link to a specialised sequence search from under the 'Search' button in the header of the home page, as shown in Fig. 2. This sequence search provides an EMBL-EBI interface to NCBI BLAST search (Camacho et al., 2008), as shown in

Fig. 10. You can input a sequence in FASTA format, select the target data set (for example UniProtKB) and other optional variables and run a BLAST search to find matches to known proteins.

### Discussion

The integration of enzyme data from different specialised resources is critical for the full exploitation of this data in research and industry. The Enzyme Portal (<http://wwwdev.ebi.ac.uk/enzymeportal/>) provides a simple but effective approach to the challenge of integrating data from different resources. It utilises the richness of the cross-references between the EMBL-EBI resources to provide enzyme summaries for users while maintaining the community-specificity and

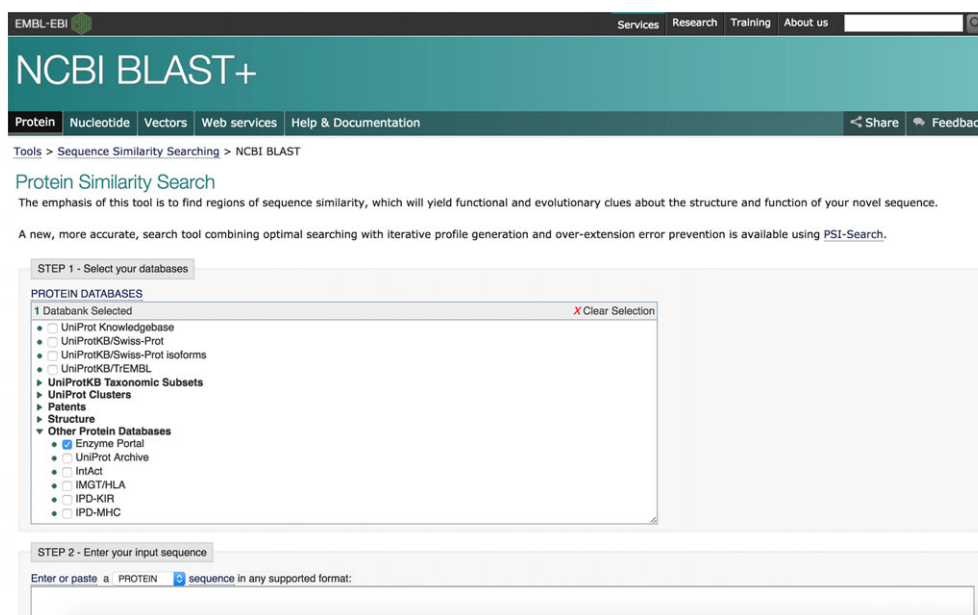


Fig. 10 Sequence search interface linked from the Enzyme Portal home page.

current knowledge from the individual resources. By collating information about enzyme name, function, catalytic activity, sequence features, disease involvement, pathways and reactions, 3D structure, small molecules and literature, it creates an essential hub for researchers. It can be used to explore data at various levels, from the summaries within the Enzyme Portal through to the detailed domain-specific data in the underlying resources that it links to. The described infrastructure allows for both high performance and integration of further resources. This has allowed the Enzyme Portal to serve different purposes for different scientific communities interested in enzymes and to help answer their questions accurately. Future work includes adding analysis tools such as EC Blast (<http://www.ebi.ac.uk/thornton-srv/software/rbl/>) (Rahman *et al.*, 2014). We welcome feedback from the community to help grow the Enzyme Portal with new functionality and further resources.

## Acknowledgements

The UniProt Consortium [www.uniprot.org](http://www.uniprot.org) (UniProt Consortium, 2015), the PDBe team <http://www.ebi.ac.uk/pdbe/> (Valenkar *et al.*, 2016), the Reactome team [www.reactome.org](http://www.reactome.org) (Fabregat *et al.*, 2016), the Rhea team <http://www.rhea-db.org/> (Morgat *et al.*, 2015), the OMIM team <http://www.omim.org/> (McKusick, 2007), the IntEnz team <http://www.ebi.ac.uk/intenz/> (Fleischmann *et al.*, 2004), the ChEBI team <https://www.ebi.ac.uk/chebi/> (Hastings *et al.*, 2013), the ChEMBL team <https://www.ebi.ac.uk/chembl/> (Bento *et al.*, 2014) and the EMBL-EBI search team <https://www.ebi.ac.uk/ebisearch/> (Squizzato *et al.*, 2015).

## Funding

Funding comes from European Molecular Biology Laboratory core funds.

## References

- Bento,A.P., Gaulton,A., Hersey,A. *et al.* (2014) *Nucleic Acids Res.*, **42**, 1083–1090.
- Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2008) *BMC Bioinformatics*, **10**, 421.
- de Matos,P., Cham,J., Cao,H., Alcantara,R., Rowland,F., Lopez,R. and Steinbeck,C. (2013) *BMC Bioinformatics*, **14**, 103.
- Fabregat,A., Sidiropoulos,K., Garapati,P. *et al.* (2016) *Nucleic Acids Res.*, **44**, D481–D487.
- Fleischmann,A., Darsow,M., Degtyarenko,K. *et al.* (2004) *Nucleic Acids Res.*, **32**, D434–D437.
- Hakkola,J., Pelkonen,O., Pasanen,M. and Raunio,H. (1998) *Crit. Rev. Toxicol.*, **28**, 35–72.
- Hastings,J., de Matos,P., Dekker,A. *et al.* (2013) *Nucleic Acids Res.*, **41**, D-456–D463.
- McKusick,V.A. (2007) *Am. J. Hum. Genet.*, **80**, 588–604.
- Morgat,A., Axelsen,K.B., Lombardot,T. *et al.* (2015) *Nucleic Acids Res.*, **43**, D459–D464.
- Rahman,S.A., Cuesta,S.M., Furnham,N., Holliday,G.L. and Thornton,J.M. (2014) *Nat. Methods.*, **11**, 171–174.
- Schettino,C., Bareschino,M.A., Ricci,V. and Ciardiello,F. (2008) *Exp. Rev. Res. Med.*, **2**, 167–178.
- Squizzato,S., Park,Y.M., Buso,N., Serrano,R.L. *et al.* 2015 *Nucleic Acids Res.*, **43**, W585–W588.
- UniProt Consortium (2015) *Nucleic Acids Res.*, **43**, D204–D212.
- Valenkar,S., van Ginkel,G., Alhroub,Y. *et al.* (2016) *Nucleic Acids Res.*, **44**, D385–D395.