

Complementary Sources of Protein Functional Information: The Far Side of GO

Nicholas Furnham

Abstract

The GO captures many aspects of functional annotations, but there are other alternative complementary sources of protein function information. For example, enzyme functional annotations are described in a range of resources from the Enzyme Commission (E.C.) hierarchical classification to the Kyoto Encyclopedia of Genes and Genomes (KEGG) to the Catalytic Site Atlas amongst many others. This chapter describes some of the main resources available and how they can be used in conjunction with GO.

Key words Function similarity, Protein domain functions, Enzyme Commission (EC), Pathway annotation

1 Introduction

The Gene Ontology (GO) offers experimental and computational biology researchers an accessible range of controlled vocabulary annotations to describe protein function. This allows detailed as well as large-scale analyses to be conducted. There is, however, a range of other sources of functional annotations, which in combination with GO provide enhanced function descriptions. Examples of such complementary resources include the Enzyme Commission's classification of enzyme reactions [1], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [2], BRENDA [3], CSA [4], MACiE [5], MetaCyc database of enzyme and pathways [6], amongst many others. Most of these resources include GO terms within their own annotations or their definitions are included within the Gene Ontology. Mapping terms between resources offers enhanced descriptions and relationships between them not readily captured solely within GO. The Gene Ontology provides many of these mappings through its website (<http://geneontology.org/page/download-mappings>), which are automatically updated with various periodicities depending on how often the

corresponding resource is updated. This chapter describes some of these complementary resources focusing mainly on enzymes.

2 Annotating Enzymes

Due to the over 100 years of experimental biochemical data, one of the richest areas for complementary functional annotations are for enzymes. Historically, naming conventions for enzymes have been confused and haphazard, with several names being given to one enzyme and one name being given to several enzymes. Often the names bear little information as to the reaction the enzyme is undertaking. This led to the development of the Enzyme Classification (E.C.) system by the International Commission on Enzymes founded in 1956 by the International Union of Biochemistry [1]. The E.C. number is a hierarchal system consisting of four levels. The first level has six divisions giving a broad description of the overall chemical transformation (enzyme class): Oxidoreductases, Transferases, Hydrolases, Lyases, Isomerases and Ligases. The next two levels (sub class and sub-subclass) generally describe the reactive species and the type of bond being acted upon. The meaning of these numbers is class dependent. The final level is a serial number for the overall reaction of that sub-subclass. The overall reactions described are mass-balanced, as much as possible, though they are not necessarily charge-balanced, nor are they meant to represent the equilibrium position or reaction direction with a convention for writing the reaction in the same direction for all reactions within a given sub-subclass even if their physiological direction is different. General reactions, where the enzyme has broad specificity, are given as single generic reactions and alternative reactions with specific metabolites are also given. Some reactions are incomplete, while others are combinations of successive reactions [7]. Thus it is possible that one enzyme E.C. number might have a multiple number of reactions associated with it and for many reactions to be assigned to the same E.C. number (*see* Fig. 1a).

Currently there are 6510 E.C. numbers approved, with 5560 of them in active use. Of these active annotations only 3924 (70%) have an equivalent GO term. A full list of E.C. to GO cross-references can be found on the GO website (<http://geneontology.org/external2go/ec2go>). There are a number of reasons why a mapping between E.C. and GO cannot be made. Most likely is that GO does not yet have a term that covers the EC term, e.g. E.C. 1.1.1.287 (d-arabinitol dehydrogenase). An automatic pipeline updates the cross-reference file after each GO release with any new terms that are created. Other reasons why E.C. and GO terms cannot be mapped are because of E.C. entries being transferred from one term to another or the E.C. number has yet to be associated with a gene product (termed orphaned E.C. terms).

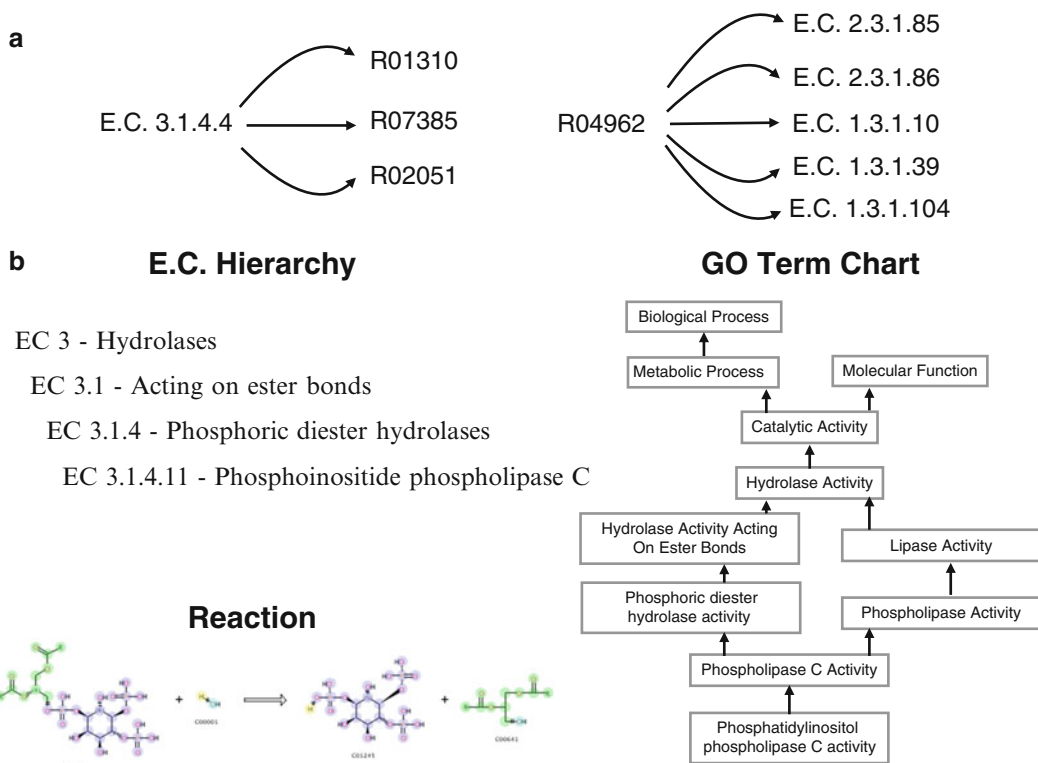


Fig. 1 (a) Examples of ambiguity in the E.C. classification, where one E.C. number can represent many reactions and where many E.C. numbers are describing one reaction. (b) The two representations of the same enzyme (phosphoinositide phospholipase C) in E.C. and GO, with the overall chemical reaction also shown. The reaction diagram is highlighted to show sub-structures across the reaction used in the determination of bond changes and reaction centers in EC-Blast

Additionally, there are “pseudo” E.C. terms created by UniProt that describe an overall reaction derived from the literature but have yet to be included in the E.C. These are easily identifiable as they have a letter n in the fourth level of the hierarchy, e.g. 1.1.1.n5 (3-methylmalate dehydrogenase).

Databases such as KEGG and BRENDA hold details of alternative reactions and data relating to physiological function. Other resources hold more specific functional annotations such as the catalytic residues and how they function in the overall reactions, as cataloged by the Catalytic Site Atlas (CSA), or MACiE that annotates the steps in an enzyme’s reaction, the order in which bonds are broken and formed, the role of cofactors and the function of protein residues at each step. To bridge the gap between these more chemical descriptors and the biological descriptors associated with a protein a new ontology, the Enzyme Mechanism Ontology (EMO), has been developed [4]. Though not directly linked to GO, EMO terms can be determined through links with GOA terms of the UniProtKB record for a particular enzyme.

3 Comparing Enzyme Annotations

Unlike GO, the E.C. number cannot be used to make automated quantitative comparisons between annotations. There are a number of measures of annotation similarity that can be made based on the GO ontological graph. The most basic similarity measure is based on the length of the common path between two terms to the ontology root and has been enhanced to overcome the fact that the depth of a term within the ontology is not necessarily indicative of its specificity, termed information content (IC). Further enhancements normalize the IC measure (Lin score) and use semantic similarity (Wang score) [8, 9]. To overcome the deficiencies of E.C. as a means to measure functional similarity and to capture detailed reaction information not encapsulated in GO, new methods have been developed. Efforts to compare reactions based on their overall reaction chemistry have met with only moderate success, limited by their reliance upon the consistency and reliability of the underlying reaction data and the ability of the algorithm used to process a diverse range of reactions. The latest method called EC-Blast [10] has proven more successful. It uses an atom-atom mapping approach to automatically assign bond changes and reaction centers (the atom and bond type in the immediate region of the metabolite where the bonds are broken/formed). This allows for the reaction to be described in a set of fingerprints that in composite can be used to compare reactions. Taking all available E.C. numbers and equivalent GO terms that can be compared to each other, the difference between the two ways of measuring functional similarity is shown in Fig. 2. Though many comparisons result in similar scores, a substantial number diverge significantly. For example, E.C. 2.1.2.9 when compared to E.C. 2.1.2.11, based on bond order changes, the similarity score as calculated by EC-Blast is 0.22, where as the semantic similarity between the equivalent GO terms is 0.73. The low similarity from EC-Blast encapsulates the differences in bonds cleaved (two C-N bonds and 2 H-N bonds for E.C. 2.1.2.9; compared to one C-C, one H-O and one C-H for E.C. 2.1.2.11 as well as differences in stereochemistry changes and bond order rearrangements.) Thus, care needs to be taken in choosing the best measure of functional similarity, a widely used technique in functional inference (*see* Chap. 12 [26]).

4 Annotating Domains

One of the challenges of functional annotation is the granularity to which an annotation can be attached. Most genomic annotations are assigned to whole protein translations, i.e. the gene, but for many functions it is a protein domain that can be considered the

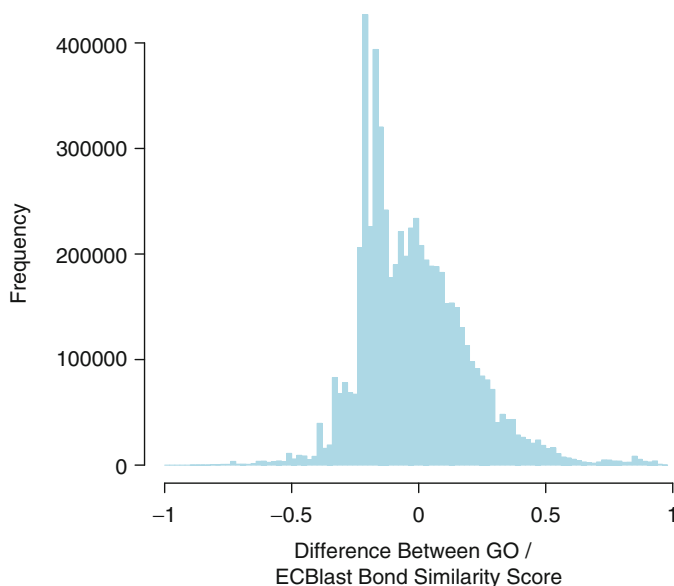


Fig. 2 Differences between GO and EC measures of functional similarity. A frequency histogram showing the difference between the similarity scores of all-by-all pairs of E.C. numbers calculated using EC-Blast bond similarity measure and the equivalent GO term. GO similarity scores are calculated using the Wang semantic similarity method. Not all E.C. numbers are used as: EC-Blast requires fully balanced reactions, and not all E.C. numbers have a GO term equivalents

functional unit. Of course functions are not solely confined to a single domain and many functions are a product of multiple domains in combination. Many domains are combined with others in increasingly complex combinations and arrangements (*see* Fig. 3). This biological complexity adds considerable complexity to functional annotations, where a function can be assigned to complete gene products and other functional annotation to just one component domain or multi-domain combinations. There are a number of domain and motif databases that provide functional annotations, many of which are mapped to GO via the InterPro [11] proteins family database, that integrates predictive models from a range of different protein family databases. One of the main sequence based domain protein family databases is PFam [12], with the goal of creating a collection of functionally annotated families that is representative as much as possible of protein-sequence space. PFam curators provide functional annotations, but in recent releases these annotations have been outsourced to the community via the use of Wikipedia allowing anyone to freely edit and improve the content, with the original curator annotations maintained. By their very nature these annotations do not conform to a controlled vocabulary, but it is possible for PFam annotations to be mapped back to GO terms; this is provided by the InterPro group and is available via the GO website.

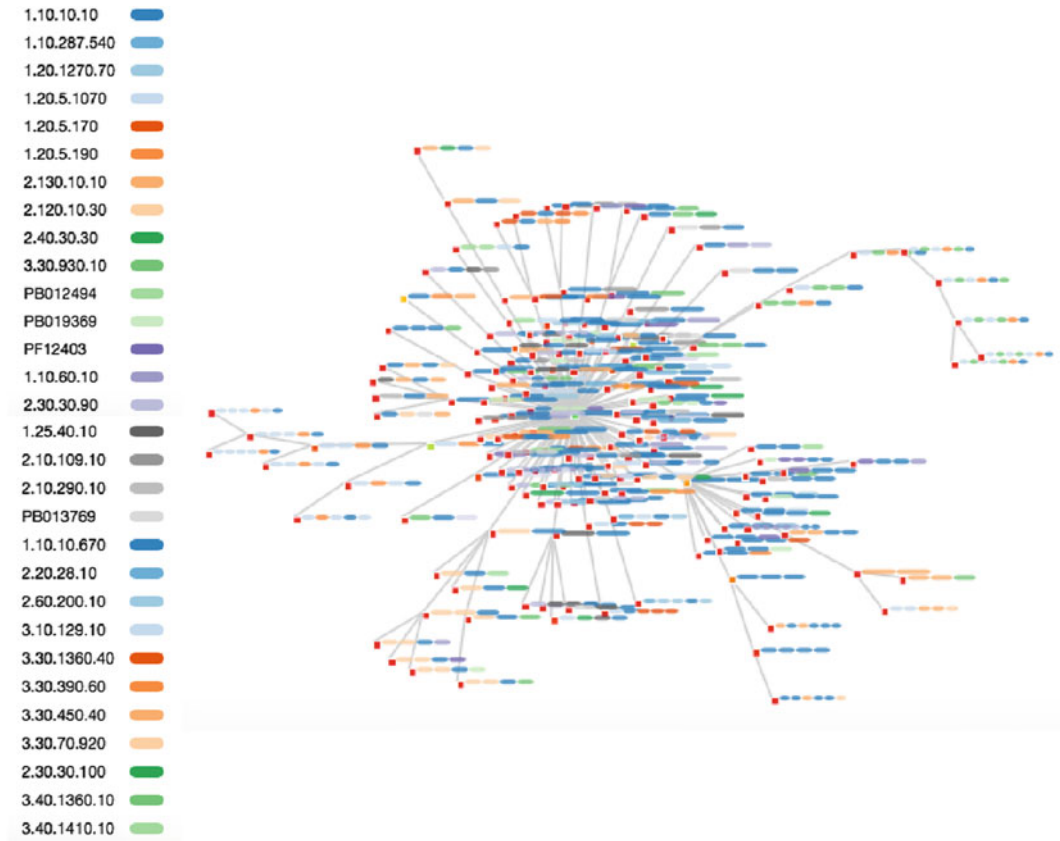


Fig. 3 Biological complexity generated by multi-domain architectures. A force-directed graph of the multi-domain architectures associated with a domain superfamily (“winged helix” repressor DNA binding domain). The graph is centered on architecture containing just the single domain with nodes (*red boxes*) radiating from this representing ever-increasing multi-domain architecture (shown to the right of the node). A key to the domains in these multi-domain architectures is shown on the left identified by PFM codes (starting PF or PB) or CATH codes. Functions are associated with the whole gene product as well as for single domains within the multi-domain architecture. An interactive version of this graph can be found at <http://www.funtree.info/templates/showArch.php?cathcode=00001.00010.00010.00010&cathmethod=&cathcluster=&type=AS>

The CATH [13] resource, which uses protein structures to define domains both within known protein structures and sequences where there is no structural information, uses the GO terms associated with a sequence to define functionally coherent clusters (termed FunFams) within the superfamily division of the classification. The functional annotation provided is derived from the predominant GO term found within the FunFam. These terms though are assigned to the whole sequence and not the domain and therefore may not directly relate to the specific function the domain is participating in. In the SFLD [14] domains that are critical for function are determined (often being used to define the superfamily), thereby linking the functional annotation to a domain or

combination of domains within a multi-domain architecture (*see* Chap. 9 [27]). SUPERFAMILY [15], a domain centric resource that uses an alternative structure based domain classification called SCOP, attempts to assign functional annotations specifically to a domain. Using the GO semantic structure and the proteins multi domain architecture, domain-centric functional annotations are statistically inferred based on the assumption that if a GO term is annotated to proteins that contain a shared domain then that term should also confer functional indicators for that domain. The SUPERFAMILY developers have generated a reduced version of GO for annotating domains and forms part of a structural domain functional ontology (SDFO) [16]. The approach of linking ontological terms to a domain can be generalized to other ontologies, most notably for phenotypic annotations. For example SUPERFAMILY integrates mammalian phenotype ontology (MPO) [17] from the mouse genome informatics (MGI) and the Human Phenotype Ontology (HPO) from the (OMIM) [18] resource.

5 Pathways and Interactions

Individual components of a pathway or groups of interacting proteins are described by the molecular function set of GO terms, while the pathways and interactions these components participate in are captured in the biological process GO terms. These provide overall descriptions of a biological process, such as signal transduction, or more specific terms such as thiamine metabolism. GO does not try to represent the dynamics or dependencies that are equivalent to a signal or metabolic pathway, though the GO consortium has recognized the importance of contextualizing gene product annotations and had begun to add some directional information (*see* Chap. 17 [28]). To be able to put the components into the context of a metabolic pathway for example, the use of specialist databases such as KEGG, BioCarta, MetaCyc, Pathway Interaction Database [19] and Reactome [20] is required (*see* Table 1). These provide curated and computationally derived descriptions of overall topologies and interactions, often displayed as pathway diagrams and maps. Many of these data resources are able to map terms back to GO. IntAct [21], which is a molecular interaction database curated from the literature or by data depositors, scores and filters interaction evidences to generate a high confidence subset of molecular interactions that are exported to GO.

Combinations of GO terms and pathway/interactions databases can be used in the analysis of proteomics data for functional annotation. This can be achieved either using methods for GO enrichment analysis and subsequently linking the results to external pathway resources [22] or by dynamically constructing the

Table 1
A summary of the data resources mentioned

Resource	Data curated	Relationship to GO	URL
ENZYMES			
E.C.	A four level hierarchical classification of enzyme functions curated by the IUBMB Enzyme Commission	Levels have equivalent GO terms in the molecular function ontology. Not all E.C. numbers have equivalent GO terms.	http://www.ebi.ac.uk/intenz
EC-Blast	A database of reaction fingerprints that are used to compare enzyme reactions defined by the E.C.	Provides function similarity scores as an alternative to GO semantic similarity measures.	https://www.ebi.ac.uk/thornton-srv/software/rbl
KEGG	A curated database of genes, enzyme reactions, metabolites, drugs and metabolic pathways.	Reactions, via E.C. numbers, are mapped to GO as well as pathways.	http://www.kegg.jp
BRENDA	A curated database of enzyme functions.	Catalogs GO terms amongst a number of other ontologies and can be used to provide a wide range of enzyme related data such as metabolic rates, molecular interactions, etc.	http://www.brenda-enzymes.org
CSA	A curated and computationally inferred database of residues in protein structures that make up an enzyme's catalytic site.	Can be used to extend GO functional annotations to include residue specific functions involved in catalysis for known structures. Can be linked to GO via EMO ontology.	https://www.ebi.ac.uk/thornton-srv/databases/CSA
MACiE	A curated database of enzyme mechanisms.	Can be used to extend GO functional annotations to include enzyme mechanistic steps. Can be linked to GO via EMO ontology.	https://www.ebi.ac.uk/thornton-srv/databases/MACiE
DOMAINS			
InterPro	Integrates predictive models from a range of different protein family databases.	Provides mappings between GO terms and many domain-centered data resources.	https://www.ebi.ac.uk/interpro

Pfam	A sequence based classification of domains.	Provides curated and community sourced functional annotations of domain families that can be mapped to GO.	http://pfam.xfam.org
CATH	A structure based classification of domains.	Provides functionally coherent clusters (FunFams) annotated by GO terms.	http://www.cathdb.info
SUPERFAMILY	A structure based classification of domains based on SCOP.	Provides domain centered functional annotations in a structural domain functional ontology that contains a slimmed down version of GO.	http://supfam.org
PATHWAYS & INTERACTIONS			
Reactome	A database of molecular pathways.	Provides descriptions of overall topologies and interactions and can be used in conjunction with GO molecular process terms.	http://www.reactome.org
IntAct	A database of molecular interactions.	Provides curated and depositor annotations of molecular interactions a high confidence subset of which is exported to GO.	https://www.ebi.ac.uk/intact

This is not an exhaustive list, but a representation of some of the alternative resources available and how they relate to GO. Many of the mappings between these data resources are available on the GO website and are updated on a daily, weekly, or monthly basis depending on the resource release cycle (<http://geneontology.org/page/download-mappings>)

pathway/interaction network based on the gene list of interest to create a functionally organized GO/pathway term network [23]. Additionally proteins participating in common biological processes or sharing molecular functions are predictive of interactions [24]. Many methods that combine semantic similarity and machine learning techniques have been developed to use GO to predict PPIs (*see* ref. 25 and references therein).

6 Conclusions

The Gene Ontology provides a rich set of ontological terms to describe many aspects of a protein's function. Many of these terms have equivalences in more specialist resources that like the Gene Ontology collate primary data derived from the literature. Often these resources include functional annotations that are not directly captured in GO or allow for annotations to be collated around a different functional unit, as in the case of protein domain centered functional annotations. Other types of functional descriptors such as the dependencies in metabolic pathways and protein-protein interactions are not explicitly captured in GO (though this is currently being addressed through GO annotation extensions), but in combination with other resources can be used to provide and enhance functional annotation of proteins.

Acknowledgements

NF is supported by a MRC Research Methodology Fellowship (MR/K020420/1). Open Access charges were funded by the University College London Library, the Swiss Institute of Bioinformatics, the Agassiz Foundation, and the Foundation for the University of Lausanne.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

References

1. International Union of Biochemistry and Molecular Biology. Nomenclature C, Webb EC (1992) Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes/prepared for NC-IUBMB by Edwin C. Webb. Published for the International Union of Biochemistry and Molecular Biology by Academic Press, San Diego
2. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue):D109–D114. doi:[10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988)
3. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, Sensen CW, Schomburg D (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res* 43(Database issue):D439–D446. doi:[10.1093/nar/gku1068](https://doi.org/10.1093/nar/gku1068)
4. Furnham N, Holliday GL, de Beer TA, Jacobsen JO, Pearson WR, Thornton JM (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res* 42(Database issue):D485–D489. doi:[10.1093/nar/gkt1243](https://doi.org/10.1093/nar/gkt1243)
5. Holliday GL, Andreini C, Fischer JD, Rahman SA, Almonacid DE, Williams ST, Pearson WR (2012) MACiE: exploring the diversity of biochemical reactions. *Nucleic Acids Res* 40(Database issue):D783–D789
6. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 42(Database issue):D459–D471. doi:[10.1093/nar/gkt1103](https://doi.org/10.1093/nar/gkt1103)
7. McDonald AG, Tipton KF (2014) Fifty-five years of enzyme classification: advances and difficulties. *FEBS J* 281(2):583–592. doi:[10.1111/febs.12530](https://doi.org/10.1111/febs.12530)
8. du Plessis L, Skunca N, Dessimoz C (2011) The what, where, how and why of gene ontology--a primer for bioinformaticians. *Brief Bioinform* 12(6):723–735. doi:[10.1093/bib/bbr002](https://doi.org/10.1093/bib/bbr002)
9. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26(7):976–978. doi:[10.1093/bioinformatics/btq064](https://doi.org/10.1093/bioinformatics/btq064)
10. Rahman SA, Cuesta SM, Furnham N, Holliday GL, Thornton JM (2014) EC-BLAST: a tool to automatically search and compare enzyme reactions. *Nat Methods* 11(2):171–174
11. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 43(Database issue):D213–D221. doi:[10.1093/nar/gku1243](https://doi.org/10.1093/nar/gku1243)
12. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Res* 42(Database issue):D222–D230. doi:[10.1093/nar/gkt1223](https://doi.org/10.1093/nar/gkt1223)
13. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, Orengo CA (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* 43(Database issue):D376–D381. doi:[10.1093/nar/gku947](https://doi.org/10.1093/nar/gku947)
14. Akiva E, Brown S, Almonacid DE, Barber AE, Custer AF, Hicks MA, Huang CC, Lauck F, Mashiyama ST, Meng EC, Mischel D, Morris JH, Ojha S, Schnoes AM, Stryke D, Yunes JM, Ferrin TE, Holliday GL, Babbitt PC (2014) The structure-function linkage database. *Nucleic Acids Res* 42(Database issue):D521–D530. doi:[10.1093/nar/gkt1130](https://doi.org/10.1093/nar/gkt1130)
15. Oates ME, Stahlhake J, Vavoulis DV, Smithers B, Rackham OJ, Sardar AJ, Zaucha J, Thurlby N, Fang H, Gough J (2015) The SUPERFAMILY 1.75 database in 2014: a doubling of data. *Nucleic Acids Res* 43(Database issue):D227–D233. doi:[10.1093/nar/gku1041](https://doi.org/10.1093/nar/gku1041)
16. de Lima Morais DA, Fang H, Rackham OJ, Wilson D, Pethica R, Chothia C, Gough J (2011) SUPERFAMILY 1.75 including a domain-centric gene ontology method. *Nucleic Acids Res* 39(Database issue):D427–D434. doi:[10.1093/nar/gkq1130](https://doi.org/10.1093/nar/gkq1130)
17. Smith CL, Eppig JT (2009) The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* 1(3):390–399. doi:[10.1002/wsbm.44](https://doi.org/10.1002/wsbm.44)
18. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A (2015) OMIM.org:

- Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 43(Database issue):D789–D798. doi:[10.1093/nar/gku1205](https://doi.org/10.1093/nar/gku1205)
19. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH (2009) PID: the Pathway Interaction Database. *Nucleic Acids Res* 37(Database issue):D674–D679. doi:[10.1093/nar/gkn653](https://doi.org/10.1093/nar/gkn653)
 20. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42(Database issue):D472–D477. doi:[10.1093/nar/gkt1102](https://doi.org/10.1093/nar/gkt1102)
 21. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeifferberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 40(Database issue):D841–D846. doi:[10.1093/nar/gkr1088](https://doi.org/10.1093/nar/gkr1088)
 22. Ramos H, Shannon P, Aebersold R (2008) The protein information and property explorer: an easy-to-use, rich-client web application for the management and functional analysis of proteomic data. *Bioinformatics* 24(18):2110–2111. doi:[10.1093/bioinformatics/btn363](https://doi.org/10.1093/bioinformatics/btn363)
 23. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8):1091–1093. doi:[10.1093/bioinformatics/btp101](https://doi.org/10.1093/bioinformatics/btp101)
 24. Qi Y, Bar-Joseph Z, Klein-Seetharaman J (2006) Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins* 63(3):490–500. doi:[10.1002/prot.20865](https://doi.org/10.1002/prot.20865)
 25. Maetschke SR, Simonsen M, Davis MJ, Ragan MA (2012) Gene Ontology-driven inference of protein-protein interactions using inducers. *Bioinformatics* 28(1):69–75. doi:[10.1093/bioinformatics/btr610](https://doi.org/10.1093/bioinformatics/btr610)
 26. Pesquita C (2016) Semantic similarity in the gene ontology. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 12
 27. Holliday GL, Davidson R, Akiva E, Babbitt PC (2016) Evaluating functional annotations of enzymes using the gene ontology. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 9
 28. Huntley RP, Lovering RC (2016) Annotation extensions. In: Dessimoz C, Škunca N (eds) *The gene ontology handbook. Methods in molecular biology*, vol 1446. Humana Press. Chapter 17