

IUPHAR-DB: updated database content and new features

Joanna L. Sharman¹, Helen E. Benson¹, Adam J. Pawson¹, Veny Lukito¹, Chidochangu P. Mpamhanga¹, Vincent Bombail¹, Anthony P. Davenport², John A. Peters³, Michael Spedding⁴, Anthony J. Harmar^{1,*} and NC-IUPHAR

¹University/BHF Centre for Cardiovascular Science, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh EH16 4TJ, ²Clinical Pharmacology Unit, Level 6, Centre for Clinical Investigation, Box 110, Addenbrooke's Hospital, University of Cambridge, Cambridge CB2 0QQ, ³Neuroscience Division, Medical Education Institute, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, UK and ⁴Les laboratoires Servier, 50 Rue Carnot, Suresnes 92284, France

Received August 10, 2012; Revised September 10, 2012; Accepted September 22, 2012

ABSTRACT

The International Union of Basic and Clinical Pharmacology (IUPHAR) database, IUPHAR-DB (<http://www.iuphar-db.org>) is an open access, online database providing detailed, expert-driven annotation of the primary literature on human and rodent receptors and other drug targets, together with the substances that act on them. The present release includes information on the products of 646 genes from four major protein classes (G protein-coupled receptors, nuclear hormone receptors, voltage- and ligand-gated ion channels) and ~3180 bioactive molecules (endogenous ligands, licensed drugs and key pharmacological tools) that interact with them. We have described previously the classification and curation of data for small molecule ligands in the database; in this update we have annotated 366 endogenous peptide ligands with their amino acid sequences, post-translational modifications, links to precursor genes, species differences and relationships with other molecules in the database (e.g. those derived from the same precursor). We have also matched targets with their endogenous ligands (peptides and small molecules), with particular attention paid to identifying bioactive peptide ligands generated by post-translational modification of precursor proteins. Other improvements to the database include enhanced information on the clinical relevance of targets and ligands in the database, more extensive links to other databases and a

pilot project for the curation of enzymes as drug targets.

INTRODUCTION

The International Union of Basic and Clinical Pharmacology (IUPHAR) database, IUPHAR-DB (<http://www.iuphar-db.org/>), is now in its 10th year (1,2), and provides expert-driven, manually curated annotation of the pharmacological, physiological and genetic properties of human and rodent receptors and other drug targets, together with the substances that act on them. Developed under the auspices of the IUPHAR Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR), IUPHAR-DB is an authoritative reference and educational resource for pharmacologists, clinicians and allied disciplines, receiving over 7000 unique visitors from 115 countries monthly. Here, we review enhancements to the data content and website since the last published update (2), including the addition of information on endogenous ligands of receptors, annotation of structural data for endogenous peptide ligands, a pilot project to curate enzymes as drug targets, increased database links and expansion of data on the clinical relevance of targets and ligands in the database.

CONTENT

In its present release, IUPHAR-DB covers the complete genomic complement of 4 protein superfamilies in human, rat and mouse: non-sensory G protein-coupled receptors (GPCRs), nuclear hormone receptors (NHRs), voltage-gated ion channels (VGICs) and ligand-gated ion channels (LGICs), as well as pilot data on 10 enzymes (from the

*To whom correspondence should be addressed. Tel: +44 131 242 6693; Fax: +44 131 242 6782; Email: tony.harmar@ed.ac.uk; tharmar@mac.com

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

lanosterol biosynthesis pathway). All data are manually curated by expert scientists and curators and referenced to primary literature citations in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed>). Information provided about targets includes their nomenclature, structure, function, physiology, clinical relevance and key ligands (endogenous ligands, licensed drugs, pharmacological tools and radioligands). Targets are linked to genomic databases such as Entrez Gene (3), Ensembl (4) and OMIM (Online Mendelian Inheritance in Man) (5), protein databases such as UniProt (6) and specialist pharmacological and medicinal chemistry databases such as DrugBank (7) and ChEMBL (8). The database presently includes information on over 8000 interactions between these targets and ~3180 small molecules, natural products, peptides and inorganic compounds. Database pages for ligands provide information on their structures, physicochemical properties, synonyms, binding affinities and actions at targets, lists of similar substances and links to chemistry databases including PubChem (9), ChemSpider (10) and Chemical Entities of Biological Interest (ChEBI) (11). All data are subject to yearly review and updated as required to reflect advances in the field. Table 1 lists the receptor and ion channel database pages that have been updated since the last published report (2).

NEW FEATURES

Annotation of endogenous ligands

IUPHAR-DB target pages list ligands recommended for experimental use including endogenous ligands, synthetic

drugs and radiolabelled forms, together with their actions and quantitative data. Since the lists can be quite extensive, NC-IUPHAR is working to identify a subset of ligands with optimal properties for preclinical and translational research. A crucial first step is to identify the principal endogenous ligands with activity at targets. To this end IUPHAR-DB curators have identified through literature and database searches 863 endogenous ligands for GPCRs and NHRs, and highlighted these on target pages together with additional comments. Likewise, the database pages for endogenous ligands clearly identify their natural targets under the 'Biological Activity' tab.

Linking endogenous small molecule ligands such as monoamines to their cognate receptors was straightforward because most were already described in IUPHAR-DB. However, the task proved more challenging for peptides due to the difficulty in distinguishing between endogenous peptides synthesized by post-translational processing of polypeptide precursors and peptide fragments produced synthetically in order to study structure-activity relationships. It can be very difficult to prove the existence, or otherwise, of particular peptide fragments *in vivo*, many of which may be present at only low concentrations, in distinct tissues or under specific physiological conditions. As a result there is a disjunction between sequences listed in online protein databases and information in the literature on their activities at targets. We sought to bridge this divide by annotating endogenous peptides described in pharmacological literature with structural information and by defining their relationships with other peptides. Endogenous ligands were identified

Table 1. IUPHAR-DB receptor and ion channel pages that have been updated since the last published database report

Receptor family	Details	Date of public release
Dopamine	Annotation added	August 2010
Formylpeptide	Annotation added	August 2010
Leukotriene	Annotation added	August 2010
Motilin	Annotation added	August 2010
Tachykinin	Annotation added	August 2010
Vasopressin and oxytocin	Annotation added	August 2010
Hydrocarboxylic acid	Annotation added	February 2011
Lysophospholipid	Annotation added	February 2011
Frizzled	Annotation added (FZD3 and FZD4)	February 2011
Orphan GPCRs	Annotation added (<i>GPR39</i> , <i>LPAR6</i> , <i>GPR35</i> , <i>GPR87</i> and <i>P2RY10</i>)	February 2011
Ionotropic glutamate	Annotation added (GluA2)	February 2011
GABA _A	Annotation added ($\delta, \epsilon, \theta, \pi$)	February 2011
P2X	Annotation added (P2X7)	February 2011
P2X	Annotation added (P2X3)	May 2011
GABA _A	Annotation added ($\beta 2, \gamma 3$)	May 2011
Nicotinic acetylcholine	Annotation of all subunits	September 2011
VIP and PACAP	Updated	September 2011
α -adrenoceptors	Updated	February 2012
β -adrenoceptors	Updated	February 2012
Calcitonin	Updated	February 2012
Calcium sensing	Updated	February 2012
Cannabinoid	Updated	February 2012
Orexin	Updated	February 2012
GABA _A	Updated ($\alpha 3, \alpha 4, \alpha 5, \beta 3, \gamma 1, \gamma 2, \delta, \epsilon, \theta$)	February 2012
Neuropeptide FF/neuropeptide AF	Updated	July 2012
Neurotensin	Updated	July 2012
Vasopressin and oxytocin	Updated	July 2012
P2X	Updated (P2X3)	July 2012

A

Ligand Id: 1697	
Ligand name	orexin-A
Species	Human, Mouse, Rat

Summary | Biological activity | References | Structure | Similar ligands | Radio analogues

Classification

Compound class: Endogenous peptide in human, mouse or rat

Synonyms

hypocretin-1

Gene/Precursor

Gene name	Species	Protein name	Previous gene names
<i>HCRT</i>	Human	prepro-orexin	
<i>Hcr</i>	Mouse	prepro-orexin	
<i>Hcr</i>	Rat	prepro-orexin	HcRT, orexin-A

Database Links

CAS Registry No.: 205640-90-0

UniProt: O43612 (Hs), O55241 (Mm), O55232 (Rn)

Please note: some links may refer to related isomers. For more details please contact: curators@iuphar-db.org

B

Ligand Id: 3732	
Ligand name	LH β subunit
Species	Human

Summary | Biological activity | References | Structure | Similar ligands

Is a component of

LH

Peptide Sequence

SREPLRPWHPINAILAVEKEGCPVITVTTICAGYCTHMRVLQAVLPPLQVQVCTYRDVRFESIRLPGCRGVDPVV
SFPVALSCRCGCRSRSTSDCGGPKDHLTCDHPQLSGLFL

Ser-Arg-Glu-Pro-Leu-Arg-Pro-Tyr-Cys-His-Pro-Ile-Asn-Ala-Ile-Leu-Ala-Val-Glu-Lys-Glu-Gly-Cys-Pro-Val-Cys-Ile-Thr-Val-Asn-Thr-Thr-Ile-Cys-Ala-Gly-Tyr-Cys-Pro-Thr-Met-Arg-Val-Leu-Gln-Ala-Val-Leu-Pro-Pro-Leu-Pro-Gln-Val-Val-Cys-Thr-Tyr-Arg-Asp-Val-Arg-Phe-Glu-Ser-Ile-Arg-Leu-Pro-Gly-Cys-Pro-Arg-Gly-Val-Asp-Pro-Val-Ser-Phe-Pro-Val-Ala-Leu-Ser-Cys-Arg-Cys-Gly-Pro-Cys-Arg-Ser-Thr-Ser-Asp-Cys-Gly-Gly-Pro-Lys-Asp-His-Pro-Leu-Thr-Cys-Arg-His-Pro-Gln-Leu-Ser-Gly-Leu-Leu-Phe-Leu

Search for 3D structures on the PDB

Search by keyword: LH β subunit Search by accession number: P01229

Post-translational Modification

N-linked glycosylation at residue 50; disulphide bonds between residues 29 and 77, 43 and 92, 46 and 130, 54 and 108, 59 and 110 and 113 and 120. The production of biologically active LH requires appropriately folded and glycosylated LH β and common α subunits that assemble to form the active LH heterodimer [1].

Download 2D Structure

No information available.

C

Ligand Id: 1152	
Ligand name	VIP
Species	Human, Mouse, Rat

Summary | Biological activity | References | Structure | Similar ligands | Radio analogues

For advanced searching click here to open chemical structure editor

Related Sequences

Sequence	IUPHAR-DB targets	GRAC targets
PHI {Sp: Pig}	VPAC ₁ , VPAC ₂	VPAC ₁ , VPAC ₂
PHM {Sp: Human}	PAC ₁ , VPAC ₁	
PHV {Sp: Rat}	PAC ₁ , VPAC ₁ , VPAC ₂	
PHV {Sp: Human}	PAC ₁	
PHI {Sp: Mouse, Rat}	PAC ₁ , VPAC ₁ , VPAC ₂	

Other Similar Sequences

Sequence	IUPHAR-DB targets	GRAC targets
N-stearyl-[¹²⁵ I] neurotensin(6-11)/VIP(7-28)	PAC ₁ , VPAC ₁ , VPAC ₂	
BAY 55-9837	PAC ₁ , VPAC ₁ , VPAC ₂	
PG 99-465	PAC ₁ , VPAC ₁ , VPAC ₂	
[¹²⁵ I]VIP	PAC ₁ , VPAC ₁ , VPAC ₂	VPAC ₁ , VPAC ₂
[Ala ^{11,22,28}]VIP	PAC ₁ , VPAC ₁ , VPAC ₂	VPAC ₁

Figure 1. Panel (A) shows a screenshot of the ligand page for the endogenous peptide orexin-A open at the ‘Summary’ tab. The sequence of orexin-A is identical in human, mouse and rat so the information on this peptide is grouped into a single database entry. Panel (B) shows a screenshot of the LH β subunit ligand page, open at the ‘Structure’ tab and showing the peptide sequence, post-translational modifications and a link to the ligand page for the LH heterodimer. Panel (C) shows a screenshot of the ligand page for vasoactive intestinal peptide (VIP),

through detailed reviews of the literature and interrogation of sequence databases such as UniProt. In total, 366 distinct endogenous peptide sequences were identified as ligands of human, rat and mouse GPCRs.

Reorganized ligand classification

The IUPHAR-DB ligand classification has been reorganized. The ‘Peptide’ category is now divided into ‘Endogenous Peptides’ and ‘Other Peptides’. For this purpose, endogenous peptides were defined as peptides encoded in the genomes of the three species covered by IUPHAR-DB, human, rat and mouse. All other peptides including synthetic peptides, semi-synthetic modified forms and peptides endogenous to other species (e.g. conotoxins) are now classified as ‘Other Peptides’. Endogenous small molecules such as hormones, neurotransmitters and metabolites have been placed into a category named generally as ‘Metabolites’ (formerly ‘Small Organics’), which also includes very close analogues such as radiolabelled forms. All other synthetic compounds may be found under the ‘Synthetic Organics’ category.

Enhanced peptide information

In the last published update (2) we described the curation of structural information and descriptors for ~2000 small molecule ligands in IUPHAR-DB. This information has now been enhanced by the curation of structural data for the 366 human, rat and mouse endogenous peptide sequences described earlier. For each endogenous peptide ligand, a range of information has been collected, including nomenclature, single- and three-letter amino acid sequences, post-translational modifications, encoding genes and precursor proteins for human, rat and mouse. Where the mature peptide was identical in structure between species, the information on all the encoding genes and precursor proteins was collated under the same database entry (Figure 1A). Where the mature peptide differed in sequence between species, separate database entries were created for each species with the relevant structural, precursor and genetic information included. Database links are provided to genetic, protein and chemistry databases including the HUGO Gene Nomenclature Committee (HGNC) database (12), Rat Genome Database (RGD) (13), Mouse Genome Database (MGD) at the Mouse Genome Informatics (MGI) website (14), Ensembl, UniProt, PubChem, ChEBI, and, where available, to experimental 3D structures in the RCSB Protein Data Bank (15).

In some cases, IUPHAR-DB also provides information on endogenous peptides in species other than human, mouse and rat where these are commonly used as experimental tools at human or rodent targets. Ambiguities in

Figure 1. Continued

open at the ‘Similar Ligands’ tab. The ‘Related Sequences’ table lists other peptides derived from the same precursor or orthologues in other species, whereas the ‘Other Similar Sequences’ table lists other peptides that have similar sequences to VIP, including synthetic and modified forms of VIP.

the literature on the peptide hormone secretin illustrate the need for this work. Secretin was first isolated from the pig and porcine secretin is frequently used in assays of secretin receptor function, even though it differs in amino acid sequence from the endogenous human and rodent peptides (16). Re-evaluation of the literature on secretin in the course of this project enabled us to provide separate data on the activities of pig, human and rodent secretins.

Some hormones are formed by the covalent association of multiple peptide fragments of a single polypeptide precursor (e.g. the A and B chains of insulin (17) and related hormones). In other cases, polypeptide subunits encoded by more than one gene combine to form the active ligand; the exact combination can determine target specificity. For example, the glycoprotein hormones follicle-stimulating hormone (FSH), thyroid-stimulating hormone (TSH), luteinizing hormone (LH) and human chorionic gonadotropin (hCG) all share a common α subunit with different β subunits conferring their target specificity (18). The database now supports multi-subunit polypeptide ligands, with distinct entries created for the intact hormone and for each component polypeptide chain. Cross-links between functional ligand and component subunit pages are provided *via* the 'Structure' tab. Sequence and precursor information is provided under the subunit entry (e.g. LH β subunit, Figure 1B) and biological activity data can be found on the page for the functional ligand (e.g. LH).

A new 'Related Sequences' table has been added to endogenous peptide pages, which lists other peptides in the database that have the same parent precursor polypeptide, as well as corresponding peptides from other species. The 'Other Similar Sequences' table lists additional endogenous and synthetic peptides with similar sequences, identified by clustering with the h-cd-hit program, part of the CD-HIT Suite (19). Together, these tables allow users to easily identify similar peptides in the database without any prior knowledge of their structures (Figure 1C).

Overall, this is a unique resource bringing together endogenous peptide sequences with quantitative data on their actions at targets, which complements detailed structural data available in UniProt, and pathway information available in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (20). Work is currently under way to curate sequence and structural information on modified peptides containing non-natural amino acids, pseudo-peptide bonds, and/or cyclization and link these to activity data in IUPHAR-DB.

Clinical relevance of drugs and targets

Information on the clinical relevance of targets and ligands in IUPHAR-DB has been enhanced by the addition of data on the status and nomenclature of approved drugs. By matching structures to entries in DrugBank, 300 small molecule drugs with United States Food and Drug Administration (FDA) approval for clinical use in humans, and 6 drugs now withdrawn from clinical use, were identified and marked as 'approved' or 'withdrawn', respectively. DrugBank includes a range of useful data on

clinically used drugs such as indication, mechanism of action, absorption, metabolism, toxicity, drug–drug and drug–target interactions. Links are provided to DrugBank target pages and DrugBank chemical classifications and drug page links provided for ligands. Structure-based searching also enabled us to assign International Nonproprietary Names (INNs) to 420 of the ligands in IUPHAR-DB. INNs are the official non-proprietary or generic names given to pharmaceutical substances, as designated by the World Health Organization (WHO).

A collaboration with Orphanet, the portal for rare diseases and orphan drugs (21), has led to the addition of data on rare diseases which are associated with targets in IUPHAR-DB. Each rare disease listing in IUPHAR-DB includes a brief description and links to expert-curated summaries in OMIM and Orphanet.

Radiolabelled analogues

Information on assays involving radiolabelled ligands can be difficult to find in literature as labelled forms are not often mentioned in abstracts. IUPHAR-DB ligand pages now feature a quick way to identify radiolabelled forms described in the database *via* a new 'Radio Analogues' tab. Labelled ligands are linked to relevant literature references, accessible *via* a 'References' tab on ligand pages. This feature provides a useful complement to a longer list of structurally similar ligands that may be found under the 'Similar Ligands' tab along with links to their targets.

Enzymes as drug targets

As a pilot study for the future curation of enzymes as drug targets in IUPHAR-DB, we have curated data for the 10 enzymes of the lanosterol biosynthesis pathway, including HMG-CoA reductase, the rate-limiting enzyme in cholesterol synthesis and the target of the statin drugs used to treat hypercholesterolemia. Information provided about enzymes includes structural and genomic data, catalysed reactions, Enzyme Commission (EC) numbers, kinetic data, substrates, cofactors, selected inhibitors and their affinities. Links are provided to other databases including BRENDA, the Comprehensive Enzyme Information System (22), the International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme Nomenclature Database (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>) and KEGG. The information in IUPHAR-DB complements that available in enzyme databases such as BRENDA, which organizes enzymes by EC number across the tree of life, and bioactivity databases such as ChEMBL, which catalogues the results of published medicinal chemistry screens. The IUPHAR-DB approach is focussed on curation of data relevant to individual enzymes as human drug targets.

New database links

New database links have been provided from target pages to Ensembl, a genomic database, DrugBank, a database of drugs and their targets, Orphanet, the portal for rare diseases and orphan drugs, InterPro (23), a protein signature database, ChEMBL, a database of bioactive drug-like

small molecules from medicinal chemistry literature, and TreeFam (24), a database of phylogenetic trees. Ligands are now linked to DrugBank, ChEMBL, ChemSpider, the free chemical database from the Royal Society of Chemistry, BindingDB (25), a database of measured binding affinities, ZINC (26), a database of commercially available compounds for virtual screening, and the Human Metabolome Database (27), a database of small molecule metabolites in humans. Additional new database links for endogenous peptide ligands are described above in the section on 'Enhanced peptide information'.

Links with Guide to PHARMACOLOGY

NC-IUPHAR has established a collaboration with the editors and publishers of the British Pharmacological Society (BPS) Guide to Receptors and Channels (GRAC) (28) to make IUPHAR-DB data available through a new portal, Guide to PHARMACOLOGY (<http://www.guidetopharmacology.org>). The Guide to PHARMACOLOGY includes a new open-access database of information from GRAC and IUPHAR-DB with search tools and guidelines for pharmacology. GRAC covers a wider range of targets than IUPHAR-DB, and for each target lists key selective ligands and probes, recommended literature references and expert summaries of their properties. The IUPHAR-DB website includes links to Guide to PHARMACOLOGY pages and ligand pages include a summary of the data in GRAC.

FUTURE PERSPECTIVES

Work is currently under way to curate structural information for the last remaining group of ligands in IUPHAR-DB that currently lack this information: the 696 peptides not endogenous to human, rat and mouse species, which includes toxins, synthetic, radiolabelled and modified peptides. In the near future, we plan to add a sequence-based search facility for peptides to complement existing structure-based search tools for small molecules. The range of targets will be expanded to encompass all the human targets of current licensed drugs and those of probable future interest, with initial effort towards curation of enzymes and receptor tyrosine kinases. NC-IUPHAR and the BPS are also working towards better integration between IUPHAR-DB and GRAC, ensuring consistency of information but retaining their distinctive focus, with GRAC providing a summary of the key properties and ligands for a wide range of targets and IUPHAR-DB providing detailed annotation for a narrower range of important targets.

DATA ACCESS

IUPHAR-DB is freely available to access online at <http://www.iuphar-db.org>. The website includes downloadable files containing current receptor and channel lists, NC-IUPHAR nomenclature, synonyms, genetic information, HGNC gene nomenclature and identifiers and other database accessions. Other file formats are available by

emailing curators@iuphar-db.org. Information on linking to IUPHAR-DB pages is provided at <http://www.iuphar-db.org/linking.jsp>. In the near future, we plan to enable computational access to the database via Web services and to extend the range of options for data download via the website.

CITING IUPHAR-DB

For a general citation of the resource we recommend citing this article. For citing specific target pages we recommend a format similar to the following example: A. P. Davenport, E. J. Mead. Kisspeptin receptor. Last modified on <date>. Accessed on <date>. IUPHAR database (IUPHAR-DB), <http://www.iuphar-db.org/DATABASE/FamilyMenuForward?familyId=34>.

ACKNOWLEDGEMENTS

The authors thank all contributors and members of NC-IUPHAR and its subcommittees for their ongoing support. NC-IUPHAR members: S.P.H. Alexander, T.I. Bonner, W.A. Catterall, A. Christopoulos, A.P. Davenport, C.T. Dollery, S. Enna, A.J. Harmar, K. Kaibuchi, Y. Kanai, V. Laudet, R.R. Neubig, E.H. Ohlstein, J.A. Peters, J.P. Pin, U. Rugg, P. du Souich, M. Spedding and M.W. Wright. The work of NC-IUPHAR is supported by the American Society for Pharmacology and Experimental Therapeutics, Servier, GlaxoSmithKline, Pfizer, Actelion, AstraZeneca, DiscoveRx, Lundbeck, Abbott and Merck Millipore. The authors also acknowledge the support of the British Heart Foundation Centre of Research Excellence Award (RE/08/001).

FUNDING

International Union of Basic and Clinical Pharmacology, British Pharmacological Society and The Wellcome Trust (Biomedical Resources Grant 099156/Z/12/Z and a Biomedical Vacation Scholarship to V.L.). Funding for open access charge: The Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

- Harmar, A.J., Hills, R.A., Rosser, E.M., Jones, M., Buneman, O.P., Dunbar, D.R., Greenhill, S.D., Hale, V.A., Sharman, J.L., Bonner, T.I. *et al.* (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res.*, **37**, D680–D685.
- Sharman, J.L., Mpamhanga, C.P., Spedding, M., Germain, P., Staels, B., Dacquet, C., Laudet, V. and Harmar, A.J. (2011) IUPHAR-DB: new receptors and tools for easy searching and visualization of pharmacological data. *Nucleic Acids Res.*, **39**, D534–D538.
- Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
- Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.

5. Amberger, J., Bocchini, C. and Hamosh, A. (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum. Mutat.*, **32**, 564–567.
6. UniProt Consortium. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **40**, D71–D75.
7. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
8. 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 drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
9. Bolton, E., Wang, Y., Thiessen, P.A. and Bryant, S.H. (2008) PubChem: Integrated Platform of Small Molecules and Biological Activities. In *Annual Reports in Computational Chemistry*, Vol. 4. American Chemical Society, Washington, DC, pp. 217–241.
10. Pence, H.E. and Williams, A. (2010) ChemSpider: an online chemical information resource. *J. Chem. Edu.*, **87**, 1123–1124.
11. de Matos, P., Alcantara, 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.
12. Seal, R.L., Gordon, S.M., Lush, M.J., Wright, M.W. and Bruford, E.A. (2011) genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
13. Dwinell, M.R., Worthey, E.A., Shimoyama, M., Bakir-Gungor, B., DePons, J., Laulederkind, S., Lowry, T., Nigram, R., Petri, V., Smith, J. *et al.* (2009) The Rat Genome Database 2009: variation, ontologies and pathways. *Nucleic Acids Res.*, **37**, D744–D749.
14. Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A. and Richardson, J.E. (2012) The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.*, **40**, D881–D886.
15. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
16. Mayo, K.E., Miller, L.J., Bataille, D., Dalle, S., Goke, B., Thorens, B. and Drucker, D.J. (2003) International Union of Pharmacology. XXXV. The glucagon receptor family. *Pharmacol. Rev.*, **55**, 167–194.
17. Shabanpoor, F., Separovic, F. and Wade, J.D. (2009) The human insulin superfamily of polypeptide hormones. *Vitam. Horm.*, **80**, 1–31.
18. Pierce, J.G. and Parsons, T.F. (1981) Glycoprotein hormones: structure and function. *Annu. Rev. Biochem.*, **50**, 465–495.
19. Huang, Y., Niu, B., Gao, Y., Fu, L. and Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.
20. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
21. Rath, A., Olry, A., Dhombres, F., Brandt, M.M., Urbero, B. and Ayme, S. (2012) Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum. Mutat.*, **33**, 803–808.
22. Scheer, M., Grote, A., Chang, A., Schomburg, I., Munnaretto, C., Rother, M., Sohngen, C., Stelzer, M., Thiele, J. and Schomburg, D. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.*, **39**, D670–D676.
23. Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S. *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*, **40**, D306–D312.
24. Ruan, J., Li, H., Chen, Z., Coghlan, A., Coin, L.J., Guo, Y., Hering, J.K., Hu, Y., Kristiansen, K., Li, R. *et al.* (2008) TreeFam: 2008 update. *Nucleic Acids Res.*, **36**, D735–D740.
25. Liu, T., Lin, Y., Wen, X., Jorissen, R.N. and Gilson, M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.*, **35**, D198–D201.
26. Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S. and Coleman, R.G. (2012) ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.*, **52**, 1757–1768.
27. Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., Young, N., Gautam, B., Hau, D.D., Psychogios, N., Dong, E., Bouatra, S. *et al.* (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**, D603–D610.
28. Alexander, S.P., Mathie, A. and Peters, J.A. (2011) Guide to receptors and channels (GRAC), 5th edn. *Br. J. Pharmacol.*, **164** (Suppl. 1), S1–S324.