Nucleic Acids Research, 2011, Vol. 39, Database issue ***D1–D6***
doi:10.1093/nar/gkq1243

The 2011 *Nucleic Acids Research* Database Issue and the online Molecular Biology Database Collection

Michael Y. Galperin[1,*] and Guy R. Cochrane[2]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA and [2]EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Received November 12, 2010; Accepted November 15, 2010

ABSTRACT

The current 18th Database Issue of *Nucleic Acids Research* features descriptions of 96 new and 83 updated online databases covering various areas of molecular biology. It includes two editorials, one that discusses COMBREX, a new exciting project aimed at figuring out the functions of the 'conserved hypothetical' proteins, and one concerning BioDBcore, a proposed description of the 'minimal information about a biological database'. Papers from the members of the International Nucleotide Sequence Database collaboration (INSDC) describe each of the participating databases, DDBJ, ENA and GenBank, principles of data exchange within the collaboration, and the recently established Sequence Read Archive. A testament to the longevity of databases, this issue includes updates on the RNA modification database, Definition of Secondary Structure of Proteins (DSSP) and Homology-derived Secondary Structure of Proteins (HSSP) databases, which have not been featured here in >12 years. There is also a block of papers describing recent progress in protein structure databases, such as Protein DataBank (PDB), PDB in Europe (PDBe), CATH, SUPERFAMILY and others, as well as databases on protein structure modeling, protein–protein interactions and the organization of inter-protein contact sites. Other highlights include updates of the popular gene expression databases, GEO and ArrayExpress, several cancer gene databases and a detailed description of the UK PubMed Central project. The *Nucleic Acids Research* online Database Collection, available at: http://www.oxfordjournals.org/nar/database/a/, now lists 1330 carefully selected molecular biology databases. The full content of the Database Issue is freely available online at the *Nucleic Acids Research* web site (http://nar.oxfordjournals.org/).

COMMENTARY

This current, 18th annual Database Issue of *Nucleic Acids Research* (NAR) features descriptions of 96 new (Table 1) online databases covering a variety of molecular biology data and 83 data resources that have previously been published in NAR or other journals. The accompanying NAR online Molecular Biology Database Collection (http://www.oxfordjournals.org/nar/database/a/) now includes 1330 data sources.

In addition to this editorial comment, the current issue includes two more editorials. The first of them (1) is a collective statement by a large consortium of scientists, including the authors of this article, who are concerned with the proliferation of new databases that are rarely able to talk to each other. As a result, instead of contributing to building a single body of knowledge, these databases risk functioning increasingly as isolated islands in a sea of disparate biological data. This article proposes creating a community-defined, uniform, generic description of the core attributes of biological databases, BioDBcore, a kind of 'minimal information about a biological database', and provides a preliminary checklist to describe basic specifications of each new database (1). We would ask the authors of future submissions to the NAR Database Issue to fill out that checklist (or its latest version posted at http://biocurator.org/biodbcore.shtml) and provide it as Supplementary Data to their manuscripts. In addition, we will explore ways in which the NAR online Molecular Biology Database Collection might ultimately support the standard.

Another editorial (2) describes COMBREX, an exciting project that is aimed at figuring out the functions of the

*To whom correspondence should be addressed. Tel: +1 301 435 5910; Fax: +1 301 435 7793; Email: galperin@ncbi.nlm.nih.gov

Table 1. New molecular biology databases featured in the 2011 NAR Database Issue

Database name	URL	Brief description
Allele Frequency Net	http://www.allelefrequencies.net	Immunogenetic gene frequencies in worldwide populations
AmoebaDB	http://amoebadb.org/amoeba/	Functional genomics resource for Amoebazoa
ArachnoServer[a]	http://archnoserver.org	Sequence, structure and activity of protein toxins from spider venom
AREsite	http://rna.tbi.univie.ac.at/AREsite	AU-Rich Elements in vertebrate mRNA UTR sequences
ASD	http://mdl.shsmu.edu.cn/ASD/	Allosteric Site Database
ASPicDB[a]	http://www.caspur.it/ASPicDB/	Alternative Splicing Prediction DataBase
Autophagy Database	http://tp-apg.genes.nig.ac.jp/autophagy/	Proteins involved in autophagy (self-digestion of eukaryotic cells)
BISC	http://bisc.soe.ucsc.edu	BInary SubComplexes in Proteins
Bovine Genome	http://BovineGenome.org	Bovine Genome database
BriX	http://brix.switchlab.org/	Protein building blocks for structural analysis
BSDB	http://www.ifpan.edu.pl/BSDB	Biomolecule Stretching Database
BRENDA-BTO	http://www.brenda-enzymes.org/BTO	BRENDA Tissue Ontology database
CADgene	http://www.bioguo.org/CADgene/	Coronary artery disease gene database
CAMERA[a]	http://camera.calit2.net/	Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis
CancerResource	http://bioinformatics.charite.de/cancerresource/	Cancer-related proteins and compounds
CaSNP	http://cistrome.dfci.harvard.edu/snparray/	Copy number alterations in cancer genomes
cBARBEL	http://www.catfishgenome.org/	Catfish genome database
CCDB	http://crdd.osdd.net/raghava/ccdb/	Cervical cancer gene database
CDDB	http://www.cdyn.org/	Conformational Dynamics Data Bank of proteins and protein assemblies
ChemProt	http://www.cbs.dtu.dk/services/ChemProt/	Annotated and predicted interactions of chemicals with proteins
CLIPZ	http://www.clipz.unibas.ch	Experimentally-determined binding sites of RNA-binding proteins
COMBREX	http://www.combrex.org/	COMputational BRidges to EXperiments
CPLA	http://cpla.biocuckoo.org/	Compendium of protein lysine acetylation
DAnCER	http://wodaklab.org/dancer/	Disease Annotated Chromatin Epigenetic Resource
DBASS5/3	http://www.dbass.org.uk	Database of Aberrant Splice Sites: 5′ and 3′ splice sites
dbCRID	http://dbcrid.biolead.org	Database of Chromosomal Rearrangements In Diseases
dbDNV	http://140.109.42.20/DNVs	Database of Duplicated-gene Nucleotide Variants
dbSNP-Q	http://cgsmd.isi.edu/dbsnpq	GWAS prioritization tool
DDPC	http://apps.sanbi.ac.za/ddpc/	Database of Genes Associated with Prostate Cancer
EDULISS	http://eduliss.bch.ed.ac.uk/	EDinburgh University Ligand Selection System
Effective	http://www.effectors.org	Predicted secreted bacterial proteins
EMDataBank	http://emdatabank.org	3D cryo-electron microscopy maps, models and metadata
FlyFactorSurvey	http://pgfe.umassmed.edu/TFDBS/	Drosophila transcription factor and their binding specificities
FragmentStore	http://bioinformatics.charite.de/fragment_store	Compound fragment library for fragment-based drug design
FusariumDB	http://www.fusariumdb.org/	Comparative genomics of Fusarium strains
GET-Evidence	http://get-evidence.org	A system for analyzing non-synonymous SNPs in human genes
GlycomeDB[a]	http://www.glycome-db.org	Carbohydrate structures
Herb Ingredient Targets	http://lifecenter.sgst.cn/hit	Protein targets for active compounds from Chinese herbs
HitPredict	http://hintdb.hgc.jp/htp/	High-confidence protein–protein interactions
Hymenoptera Genome	http://HymenopteraGenome.org	Genome sequences for honey bee and the wasp *Nasonia vitripennis*
IGDD	http://115.248.74.248/igdd/home.aspx	Indian Genetic Disease Database
IGRhCellID	http://igrcid.ibms.sinica.edu.tw	Integrated Genomic Resources of Human Cell Lines for Identification
IKMC	http://www.knockoutmouse.org	The International Knockout Mouse Consortium database
Isobase	http://isobase.csail.mit.edu	IsoRank PPI Network Alignment Based Ortholog Database
KaPPA-View	http://kpv.kazusa.or.jp/kpv4	Kazusa Plant Pathway Viewer
KUPS	http://www.ittc.ku.edu/chenlab/	University of Kansas Proteomics Service: protein–protein interaction
Laminin Database	http://www.lm.lncc.br	Laminin Database
lncRNAdb	http://www.lncrnadb.com	Long Non-Coding RNA Database
LocDB	http://www.rostlab.org/services/locDB	Protein localization data for human and Arabidopsis
LSD	http://www.eplantsenescence.org	Leaf Senescence Database

(continued)

Table 1. Continued

Database name	URL	Brief description
MatrixDB	http://matrixdb.ibcp.fr	Extracellular matrix proteins and their interactions
mESAdb	http://konulab.fen.bilikent.edu.tr/mirna	microRNA Expression and Sequence Analysis Database
MicrosporidiaDB	http://microsporidiadb.org	Functional genomics resource for Microsporidia
miRTarBase	http://mirtarbase.mbc.nctu.edu.tw	Experimentally validated interactions of microRNA with their targets
MitoGenesisDB	http://www.dsimb.inserm.fr/dsimb_tools/mitgene	Gene expression in mitochondrial biogenesis
NCBI Epigenomics	http://www.ncbi.nlm.nih.gov/epigenomics/	Genomic maps of nuclear changes that control gene expression
NGSmethDB	http://bioinfo2.ugr.es/meth/NGSmethDB.php	Next-generation sequencing DNA methylation data
NIAS GeneBank	http://www.gene.affrc.go.jp/databases_en.php	Plant genetic resources at the National Institute of Agrobiological Sciences in Tsukuba, Japan
non-B DB	http://nonb.abcc.ncifcrf.gov	Non-B DNA forming motifs in mammalian genomes
OMA browser[a]	http://www.omabrowser.org	Orthology Matrix
OMPdb	http://bioinformatics.biol.uoa.gr/OMPdb	Outer membrane proteins from Gram-negative bacteria
P2CS[a]	http://www.p2cs.org	Prokaryotic 2-Component Systems database
PAIR	http://www.cls.zju.edu.cn/pair/	Predicted Arabidopsis Interactome Resource
Pancreas Expression[a]	http://www.pancreasexpression.org	Pancreatic gene Expression database
Pathway Commons	http://www.pathwaycommons.org/pc/	Metabolic and signaling pathways from multiple organisms
PCDB	http://pcdb.unq.edu.ar/	Protein Conformational Diversity database
PCDDB	http://pcddb.cryst.bbk.ac.uk	Protein Circular Dichroism Database
PCRPi-DB	http://www.bioinsilico.org/PCRPIDB	Presaging Critical Residues in Protein interface-DataBase
PhEVER	http://pbil.univ-lyon1.fr/databases/phever/index.php	Phylogenetic Exploration of Viruses and their Evolutionary Relationships
PHOSIDA[a]	http://www.phosida.com	Posttranslational modification sites identified by mass spectrometry
PmiRKB	http://bis.zju.edu.cn/pmirkb	Plant microRNA knowledge base
PolyQ	http://pxgrid.med.monash.edu.au/polyq2i/	Polyglutamine Repeats in Proteins
PREX	http://csb.wfu.edu/PREX	PeroxiRedoxin classification indEX
PRIDB	http://bindr.gdcb.iastate.edu/PRIDB	Protein–RNA Interface Database
PRO	http://pir.georgetown.edu/pro	Protein Ontology based on evolutionary relatedness
PROMISCUOUS	http://bioinformatics.charite.de/promiscuous	Protein interactions data for studies of drug repositioning
ProtCID	http://dunbrack2.fccc.edu/protcid	Protein Common Interface Database
PSSRdb	http://210.212.215.200/PSSR/pssr_frame.html	Polymorphic Simple Sequence Repeats in bacteria
RBPDB	http://rbpdb.ccbr.utoronto.ca/	RNA-binding proteins and their specificities
RegPhos	http://RegPhos.mbc.nctu.edu.tw	Regulatory Network in Protein Phosphorylation
REPAIRtoire	http://repairtoire.genesilico.pl	DNA repair pathways of human, yeast and *E. coli*
RepTar	http://reptar.ekmd.huji.ac.il/	Predicted targets of host and viral miRNAs
RiceXPro	http://ricexpro.dna.affrc.go.jp/	High-resolution analysis of rice transcriptome
RIKEN mammals	http://scines.org/db/mammal	
SAHG	http://bird.cbrc.jp/sahg	Structural Atlas of Human Genome
SCLD	http://scld.mcb.uconn.edu	Stem Cell Lineage Database
SolGenomics[a]	http://solgenomics.net/	Solanaceae Genomics Network
SPIKE	http://www.cs.tau.ac.il/~spike/	Signaling Pathways Integrated Knowledge Engine
Starbase	http://starbase.sysu.edu.cn/	microRNA–mRNA interaction maps
SuperSweet	http://bioinformatics.charite.de/sweet	Natural and artificial sweetening agents
TADB	http://bioinfo-mml.sjtu.edu.cn/TADB/	Type II Toxin-Antitoxin loci in bacteria and archaea
TcoF-DB	http://cbrc.kaust.edu.sa/tcof	Database for Human Transcription Co-Factors
TFGD	http://ted.bti.cornell.edu	Tomato Functional Genomics Database
ThYme	http://www.enzyme.cbirc.iastate.edu	Thioester-active enzymes
TIARA	http://www.gmi.ac.kr	Totally Integrated Array data and high-throughput sequencing Read data Archive for human whole genomes
TMPad	http://bio-cluster.iis.sinica.edu.tw/TMPad/	Helix-packing folds in transmembrane proteins
TOPSAN	http://www.topsan.org	The Open Protein Structure Annotation Network
TRIP	http://www.trpchannel.org	Protein–protein interactions in mammalian TRP channels
UCSC Cancer Genomics Browser	http://genome-cancer.cse.ucsc.edu	Web-based tools to integrate, visualize and analyze cancer genomics and clinical data
UK PubMed Central	http://ukpmc.ac.uk/	UK PubMed Central database
ViralZone	http://www.expasy.org/viralzone	Molecular and epidemiological data on viral genera and families
VnD	http://210.218.222.221:8080/VnD/	Variation and Disease: disease-related SNPs and drugs
WebGeSTer DB	http://pallab.serc.iisc.ernet.in/gester/	Genome Scanner for bacterial transcriptional Terminators
YPA	http://service.csbb.ntu.edu.tw/ypa/	Yeast promoter atlas

[a]A description of this database has been previously published elsewhere.

'conserved hypothetical' and poorly or incorrectly annotated proteins, identified through genome sequencing [see also refs (3,4)]. This project is designed to serve as a clearinghouse, collecting functional predictions from specialists in bioinformatics and functional genomics and then sending these predictions for testing by experimentalists. COMBREX offers an entirely new arrangement for research funding, whereby relatively small amounts of money are offered on a competitive basis to the experimental groups that are willing to test those predictions, employing the techniques and equipment that already exist in their laboratories. This arrangement dramatically decreases the costs of functional analysis of the uncharacterized proteins and gives hope that many of them could be assigned a biochemical—and/or general biological—function.

A bright example of databases that do talk to each other is the International Nucleotide Sequence Database Collaboration (INSDC), which consists of three participating databases, the DNA Data Bank of Japan (DDBJ), the European Nucleotide Archive (ENA) at the European Bioinformatics Institute (EMBL-EBI), and GenBank at the US National Center for Biotechnology Information (NCBI). This issue features separate papers from each of these three databases (5–7), as well as a joint paper describing the principles of data maintenance and exchange within the collaboration (8). A separate paper describes the functioning of the Sequence Read Archive (SRA), recently established by the three INSDC partners (9).

Another area where database collaboration proved extremely successful is storage and dissemination of published research. This issue features a detailed description of the UK PubMed Central, an extremely important project that, in collaboration with PubMed Central projects in USA and Canada, provides a permanent online record for the research sponsored by British funding agencies, such as MRC, BBSRC, Wellcome Trust and the National Institute for Health Research (10).

In addition to the archival databases such as those of the INSDC, this issue includes curated databases of DNA sequence motifs, such as ARESite, a collection of AU-rich elements in vertebrate mRNA UTR sequences, and non-B DB, a repository of DNA sequences that form cruciform, triplex, slipped (hairpin) structures, tetraplex (G-quadruplex), left-handed Z-DNA and other DNA structures (11,12).

The RNA database papers featured in this issue include updates on Rfam and miRBase, two gold-standard databases of RNA sequences (13,14), a description of lncRNAdb, a new resource on experimentally characterized long non-coding RNA (15), as well as descriptions of several databases of predicted and/or experimentally validated microRNA targets (16–21). This issue also includes an update on the status of the RNA Modification Database, which was regularly featured in the NAR Database Issue in the 1990s (22–25) but not in the past 12 years. The current version lists 107 types of posttranscriptional modifications of nucleosides in RNA, primarily in various tRNAs (26). Two new databases present data on the RNA-binding proteins [RBPDB, http://rbpdb.ccbr.utoronto.ca/ (27)] and the specific structures of their RNA-binding sites [PRIDB, http://bindr.gdcb.iastate.edu/PRIDB (28)].

This issue also features a block of 15 papers describing recent progress in protein structure databases, such as Protein DataBank (PDB), PDB in Europe (PDBe), CATH, SUPERFAMILY (29–32), as well as a selection of databases on protein building blocks, protein–protein interactions, protein structure modeling, and the organization of inter-protein contact sites (33–38). Among new databases, it is worth mentioning EMDataBank.org, a database of 3D cryo-electron microscopy maps (39), a database of protein circular dichroism data (40) and three databases that are dedicated to the conformational dynamics of proteins (41–43). In addition, a paper from Gert Vriend's group (44) presents their PDB-facilities web site with several useful PDB-derived databases for the analysis of protein structures. These include the famous Definition of Secondary Structure of Proteins (DSSP) and Homology-derived Secondary Structure of Proteins (HSSP) databases, which were last featured in the NAR Database Issue >12 years ago (45,46).

Progress in the analysis of the human genome prompted the creation of databases that list genes implicated in a variety of human diseases, including coronary artery disease (47), type I diabetes (48) and cancer. Cancer databases in this issue are represented by an update paper on the Catalogue of Somatic Mutations In Cancer [COSMIC, http://www.sanger.ac.uk/cosmic (49)], a description of the University of California Santa Cruz (UCSC) Cancer Genomics Browser [http://genome-cancer.cse.ucsc.edu (50)], a new resource tightly integrated with the popular UCSC Genome Browser and the ENCODE database (51,52), and three more databases, dedicated, respectively, to cervical cancer, prostate cancer and potential cancer drug targets (53–55).

There are many other excellent databases that could not be mentioned here because of the space restrictions. In fact, we expect every single database featured in this issue to be useful to a wide audience of students and researchers in various areas of molecular biology.

As explained in last year's editorial (56), moving to an online-only format for the NAR Database Issue has allowed us to accommodate longer papers and to offer the authors of the most popular data resources an opportunity to describe their resources in more detail, providing a deeper insight into the organization and goals of their respective resources and putting the recent updates of these resources into a broader context. This year, such extended papers were invited for a much larger number of databases, resulting in comprehensive descriptions of the PDB, PDBe, EMDataBank, MODBASE, GPCRDB, RegulonDB, STRING and other well-known databases (29,30,35,39,57–59). In some cases, longer descriptions were accepted for first-time descriptions of several new databases (36,60,61). We intend to continue accepting long(er) database papers in the future.

ACKNOWLEDGEMENTS

The authors thank Sir Richard Roberts and Dr Alex Bateman, Dr David Landsman and Dr Francis Ouellette for helpful comments; Patricia Anderson, Dr Martine Bernardes-Silva and Gail Welsh for excellent editorial assistance, the Oxford University Press team lead by Claire Bird and Jennifer Boyd and Sheila Plaister at EMBL-EBI for their help in compiling this issue and the online Molecular Biology Database Collection.

FUNDING

Intramural Research Program of the US National Institutes of Health (to M.Y.G.); European Molecular Biology Laboratory (to G.R.C.). Funding for open access charge: Waived by Oxford University Press.

Conflict of interest statement. The authors' opinions do not necessarily reflect the views of their respective institutions.

REFERENCES

1. Gaudet,P., Bairoch,A., Field,D., Sansone,S.-A., Taylor,C., Attwood,T.K., Bateman,A., Blake,J.A., Bult,C.J., Cherry,J.M. *et al*. (2011) Towards BioDBcore: a community-defined information specification for biological databases. *Nucleic Acids Res*., **39**, D7–D10.
2. Roberts,R.J., Chang,Y.-C., Hu,Z., Rachlin,J., Anton,B., Pokrzywa,R., Choi,H.-P., Faller,L., Guleria,J., Housman,G. *et al*. (2011) COMBREX: a project to accelerate the functional annotation of prokaryotic genomes. *Nucleic Acids Res*., **39**, D11–D14.
3. Galperin,M.Y. and Koonin,E.V. (2004) 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res*., **32**, 5452–5463.
4. Galperin,M.Y. and Koonin,E.V. (2010) From complete genome sequence to 'complete' understanding? *Trends Biotechnol*., **28**, 398–406.
5. Kaminuma,E., Mashima,J., Kodama,Y., Gojobori,T., Ogasawara,O., Okubo,K., Takagi,T. and Nakamura,Y. (2011) DDBJ Progress Report. *Nucleic Acids Res*., **39**, D22–D27.
6. Leinonen,R., Akhtar,R., Birney,E., Bonfield,J., Bower,L., Corbett,M., Cheng,Y., Demiralp,F., Faruque,N., Goodgame,N. *et al*. (2011) The European nucleotide archive. *Nucleic Acids Res*., **39**, D28–D31.
7. Benson,D., Karsch-Mizrachi,I., Lipman,D., Ostell,J. and Sayers,E.W. (2011) GenBank. *Nucleic Acids Res*., **39**, D32–D37.
8. Cochrane,G., Karsch-Mizrachi,I. and Nakamura,Y. (2011) The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res*, **39**, D15–D18.
9. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The Sequence Read Archive. *Nucleic Acids Res*., **39**, D19–D21.
10. McEntyre,J.R., Ananiadou,S., Andrews,S., Black,W.J., Boulderstone,R., Buttery,P., Chaplin,D., Chevuru,S., Cobley,N., Coleman,L.-A. *et al*. (2011) UKPMC: a full text article resource for the life sciences. *Nucleic Acids Res*., **39**, D58–D65.
11. Gruber,A., Fallmann,J., Kratochvill,F., Kovarik,P. and Hofacker,I.L. (2011) AREsite: a database for the comprehensive investigation of AU-rich elements. *Nucleic Acids Res*., **39**, D66–D69.
12. Stephens,R., Cer,R., Bruce,K., Mudunuri,U., Yi,M., Volfovsky,N., Luke,B., Bacolla,A. and Collins,J. (2011) Non-B DB - A database of predicted non-B DNA forming motifs in mammalian genomes. *Nucleic Acids Res*., **39**, D383–D391.
13. Gardner,P., Daub,J., Tate,J., Moore,B., Osuch,I., Griffiths-Jones,S., Finn,R., Nawrocki,E., Kolbe,D., Eddy,S. *et al*. (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res*., **39**, D141–D145.
14. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*., **39**, D152–D157.
15. Amaral,P.P., Clark,M.B., Gascoigne,D.K., Dinger,M.E. and Mattick,J.S. (2011) lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res*., **39**, D146–D151.
16. Mewes,H.W., Ruepp,A., Theis,F., Rattei,T., Walter,M., Frishman,D., Suhre,K., Mayer,K., Stümpflen,V. and Antonov,A. (2011) MIPS: curated databases and comprehensive secondary data resources in 2010. *Nucleic Acids Res*., **39**, D220–D224.
17. Yang,J.-H., Li,J.-H., Shao,P., Zhou,H., Chen,Y.-Q. and Qu,L.-H. (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-seq and Degradome-seq data. *Nucleic Acids Res*., **39**, D202–D209.
18. Elefant,N., Berger,A., Shein,H., Hofree,M., Margalit,H. and Altuvia,Y. (2011) RepTar: a database of predicted cellular targets of host and viral miRNAs. *Nucleic Acids Res*., **39**, D188–D194.
19. Meng,Y., Gou,L., Chen,D., Mao,C., Jin,Y., Wu,P. and Chen,M. (2011) PmiRKB: a plant microRNA knowledge base. *Nucleic Acids Res*., **39**, D181–D187.
20. Hsu,S.-D., Lin,F.-M., Wu,W.-Y., Liang,C., Huang,W.-C., Chan,W.-L., Tsai,W.-T., Chen,G.-Z., Lee,C.-J., Chiu,C.-M. *et al*. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res*., **39**, D163–D169.
21. Cho,S., Jun,Y., Lee,S., Choi,H., Jung,S., Jang,Y., Lee,S., Kim,S., Lee,S. and Kim,W.K. (2011) miRGator v2.0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Res*., **39**, D158–D162.
22. Crain,P.F. and McCloskey,J.A. (1996) The RNA modification database. *Nucleic Acids Res*., **24**, 98–99.
23. Crain,P.F. and McCloskey,J.A. (1997) The RNA modification database. *Nucleic Acids Res*., **25**, 126–127.
24. McCloskey,J.A. and Crain,P.F. (1998) The RNA modification database–1998. *Nucleic Acids Res*., **26**, 196–197.
25. Rozenski,J., Crain,P.F. and McCloskey,J.A. (1999) The RNA modification database: 1999 update. *Nucleic Acids Res*., **27**, 196–197.
26. Agris,P., Cantara,W., Crain,P., Rozenski,J., McCloskey,J., Harris,K., Zhang,X., Vendeix,F. and Fabris,D. (2011) The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res*., **39**, D195–D201.
27. Cook,K.B., Kazan,H., Zuberi,K., Morris,Q. and Hughes,T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*., **39**, D301–D308.
28. Lewis,B., Walia,R., Terribilini,M., Ferguson,J., Zheng,C., Honavar,V. and Dobbs,D. (2011) PRIDB: a protein-RNA interface database. *Nucleic Acids Res*., **39**, D277–D282.
29. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al*. (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res*., **39**, D392–D401.
30. Velankar,S., Alhroub,Y., Alili,A., Best,C., Boutselakis,H.C., Caboche,S., Conroy,M.J., Dana,J.M., van Ginkel,G., Golovin,A. *et al*. (2011) PDBe: protein data bank in Europe. *Nucleic Acids Res*., **39**, D402–D410.
31. Cuff,A., Sillitoe,I., Lewis,T., Clegg,A., Rentzsch,R., Furnham,N., Pellegrini-Calace,M., Jones,D.T., Thornton,J. and Orengo,C. (2011) Extending CATH: Increasing coverage of the protein structure universe and linking structure with function. *Nucleic Acids Res*., **39**, D420–D426.
32. de Lima Morais,D., Fang,H., Rackham,O., Wilson,D., Pethica,R., Chothia,C. and Gough,J. (2011) SUPERFAMILY 1.75 including a domain-centric Gene Ontology method. *Nucleic Acids Res*., **39**, D427–D434.
33. Vanhee,P., Verschueren,E., Baeten,L., Stricher,F., Serrano,L., Rousseau,F. and Schymkowitz,J. (2011) BriX: a database of protein building blocks for structural analysis, modeling and design. *Nucleic Acids Res*., **39**, D435–D442.
34. Stein,A., Ceol,A. and Aloy,P. (2011) 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res*., **39**, D718–D723.

35. Pieper,U., Webb,B.M., Barkan,D.T., Schneidman-Duhovny,D., Schlessinger,A., Braberg,H., Yang, Meng,E., Pettersen,E., Huang,C. *et al.* (2011) ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **39**, D465–D474.
36. Xu,Q. and Dunbrack,R.L. Jr (2011) The protein common interface database (ProtCID)–a comprehensive database of interactions of homologous proteins in multiple crystal forms. *Nucleic Acids Res.*, **39**, D761–D770.
37. Luo,Q., Pagel,P., Vilne,B. and Frishman,D. (2011) DIMA 3.0: domain interaction map. *Nucleic Acids Res.*, **39**, D724–D729.
38. Yellaboina,S., Tasneem,A., Zaykin,D., Raghavachari,B. and Jothi,R. (2011) DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.*, **39**, D730–D735.
39. Lawson,C.L., Baker,M.L., Best,C., Bi,C., Dougherty,M., Feng,P., van Ginkel,G., Devkota,B., Lagerstedt,I., Ludtke,S.J. *et al.* (2011) EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res.*, **39**, D456–D464.
40. Whitmore, Woollett,B., Miles,A., Klose,D., Janes,R. and Wallace,B. (2011) PCDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata. *Nucleic Acids Res.*, **39**, D480–D486.
41. Kim,D.N., Altschuler,J., Strong,C., McGill,G. and Bathe,M. (2011) Conformational dynamics data bank: a database for conformational dynamics of proteins and supramolecular protein assemblies. *Nucleic Acids Res.*, **39**, D451–D455.
42. Juritz,E., Fernandez Alberti,S. and Parisi,G. (2011) PCDB: a database of protein conformational diversity. *Nucleic Acids Res.*, **39**, D475–D479.
43. Sikora,M., Sulkowska,J.I., Witkowski,B.S. and Cieplak,M. (2011) BSDB: the biomolecule stretching database. *Nucleic Acids Res.*, **39**, D443–D450.
44. Joosten,R., te Beek,T., Krieger,E., Hooft,R., Schneider,R., Sander,C. and Vriend,G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **39**, D411–D419.
45. Hooft,R.W., Sander,C., Scharf,M. and Vriend,G. (1996) The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. *Comput. Appl. Biosci.*, **12**, 525–529.
46. Dodge,C., Schneider,R. and Sander,C. (1998) The HSSP database of protein structure-sequence alignments and family profiles. *Nucleic Acids Res.*, **26**, 313–315.
47. Liu,H., Liu,W., Liao,Y., Cheng,L., Liu,Q., Ren,X., Shi,L., Tu,X., Wang,Q.K. and Guo,A.Y. (2011) CADgene: a comprehensive database for coronary artery disease genes. *Nucleic Acids Res.*, **39**, D991–D996.
48. Burren,O.S., Adlem,E.C., Achuthan,P., Christensen,M., Coulson,R.M. and Todd,J.A. (2011) T1DBase: update 2011, organization and presentation of large-scale data sets for type 1 diabetes research. *Nucleic Acids Res.*, **39**, D997–D1001.
49. Forbes,S.A., Bindal,N., Bamford,S., Cole,C., Kok,C.Y., Beare,D., Jia,M., Shepherd,R., Leung,K., Menzies,A. *et al.* (2011) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **39**, D945–D950.
50. Sanborn,J.Z., Benz,S.C., Craft,B., Szeto,C., Kober,K.M., Meyer,M., Vaske,C.J., Goldman,M., Smith,K.E., Kuhn,R.M. *et al.* (2011) The UCSC cancer genomics browser: update 2011. *Nucleic Acids Res.*, **39**, D951–D959.
51. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.
52. Raney,B.J., Cline,M.S., Rosenbloom,K.R., Dreszer,T.R., Learned,K., Barber,G.P., Meyer,L.R., Sloan,C.A., Malladi,V.S., Roskin,K.M. *et al.* (2011) ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res.*, **39**, D871–D875.
53. Agarwal,S.M., Raghav,D., Singh,H. and Raghava,G.P. (2011) CCDB: a curated database of genes involved in cervix cancer. *Nucleic Acids Res.*, **39**, D975–D979.
54. Maqungo,M., Kaur,M., Kwofie,S.K., Radovanovic,A., Schaefer,U., Schmeier,S., Oppon,E., Christoffels,A. and Bajic,V.B. (2011) DDPC: Dragon Database of Genes associated with Prostate Cancer. *Nucleic Acids Res.*, **39**, D980–D985.
55. Ahmed,J., Meinel,T., Dunkel,M., Murgueitio,M.S., Adams,R., Blasse,C., Eckert,A., Preissner,S. and Preissner,R. (2011) CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.*, **39**, D960–D967.
56. Cochrane,G.R. and Galperin,M.Y. (2010) The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. *Nucleic Acids Res.*, **38**, D1–D4.
57. Gama-Castro,S., Salgado,H., Peralta-Gil,M., Santos-Zavaleta,A., Muniz-Rascado,L., Solano-Lira,H., Jimenez-Jacinto,V., Weiss,V., Garcia-Sotelo,J.S., Lopez-Fuentes,A. *et al.* (2011) RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.*, **39**, D98–D105.
58. Vroling,B., Sanders,M., Baakman,C., Borrmann,A., Verhoeven,S., Klomp,J., Oliveira,L., de Vlieg,J. and Vriend,G. (2011) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res.*, **39**, D309–D319.
59. Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
60. Masuya,H., Makita,Y., Kobayashi,N., Nishikata,K., Yoshida,Y., Mochizuki,Y., Doi,K., Takatsuki,T., Waki,K., Tanaka,N. *et al.* (2011) The RIKEN integrated database of mammals. *Nucleic Acids Res.*, **39**, D861–D870.
61. Lee,T.Y., Bo-Kai Hsu,J., Chang,W.C. and Huang,H.D. (2010) RegPhos: a system to explore the protein kinase-substrate phosphorylation network in humans. *Nucleic Acids Res.*, **39**, D777–D787.