Nucleic Acids Research, 2003, Vol. 31, No. 1 331–333
DOI: 10.1093/nar/gkg122

NRMD: Nuclear Receptor Mutation Database

Joost J. J. van Durme, Emmanuel Bettler, Simon Folkertsma, Florence Horn[1] and Gert Vriend*

CMBI, University of Nijmegen, Nijmegen, The Netherlands and [1]Department of Cellular and Molecular Pharmacology, UCSF, Box 0450, San Francisco, CA 94143-0450, USA

Received August 19, 2002; Revised and Accepted October 30, 2002

ABSTRACT

The NRMD is a database for nuclear receptor mutation information. It includes mutation information from SWISS-PROT/TrEMBL, several web-based mutation data resources, and data extracted from the literature in a fully automatic manner. Because it is also possible to add mutations manually, a hundred mutations were added for completeness. At present, the NRMD contains information about 893 mutations in 54 nuclear receptors. A common numbering scheme for all nuclear receptors eases the use of the information for many kinds of studies. The NRMD is freely available to academia and industry as a stand-alone version at: www.receptors.org/NR/.

INTRODUCTION

Nuclear receptors (NRs) play a crucial role in the regulation of gene expression, and are thus an important target for the pharmaceutical industry. NRs consist of multiple domains, among which are a DNA-binding domain and a ligand-binding domain (LBD). The NRMD deals mainly with mutations in the LBD.

Binding of hormones such as testosterone, vitamin D3 or retinoic acid to LBDs leads to dimerization and binding of a co-activator or co-repressor, which in turn leads to transcription regulation. A good understanding of this process at the molecular level is important for the pharmaceutical industry. Mutation studies are an important source of data on the role of individual amino acids, and together with structural data about amino acid–ligand interactions, they play a central role in the rational drug design process.

The function of residues in NRs is mainly determined by their location (Folkertsma *et al.* in preparation; www.receptors.org/NR/struct/alignmt.html). Therefore, a mutation of a residue at a certain position in one receptor is likely to have a similar effect as the mutation of a different residue at the equivalent position in another receptor. The structural equivalence of residue positions can thus be used to transfer information about mutations in one NR to all other NRs. To aid this transfer of information, we introduced a common structure-based numbering scheme for all NRs.

The NRMD is part of a larger project aimed at the design of Molecular Class Specific Information Systems (MCSISs). Well-known examples of MCSISs are the GPCRDB (www.gpcr.org) (1) and the NucleaRDB (www.receptors.org/NR/) (2). The NRMD currently exists as a stand-alone mutation information resource, but full integration in the NucleaRDB is well underway.

RESULTS

Several mutation information resources are available. SWISS-PROT/TrEMBL (3) contain sequences of about 1300 NRs, and from their annotation, information about 359 variants can easily be extracted. A fully automatic search of OMIM (4) for NR mutations is difficult, but a human-aided computer script could extract 156 mutants from this resource. The Vitamin D receptor (VDR) pages contain well-documented information about vitamin D receptors. The Photoreceptor cell-specific Nuclear Receptor pages (PNR) (5) specialize in the photoreceptor cell-specific nuclear receptor, and the Glucocorticoid Receptor Resource (GRR) (6) contains a lot of information about glucocorticoid receptors. The NucleaRDB provides point mutation data (www.receptors.org/NR/mutation/) automatically extracted from the literature (Horn and Cohen, in preparation). Using pattern matching, the method 'MuteXt' retrieves articles and extracts point mutations, which are then validated by plausibility filters. These filters use the sequence data and the in-text distances between receptor names, organism types and mutants. The preliminary evaluation of MuteXt yields a recall of 85%, a precision of 90% and a coverage of 70%. The recall is the percentage of point mutations that are correctly extracted [true positives/(true positives + false negatives)], the precision is the percentage of validated point mutations that are correct [true positives/(true positives + false positives)] and the coverage is the percentage of relevant point mutations that the system extracted (true and false positives/point mutations present in the documents). Table 1 summarizes the resources at the basis for the NRMD, and Table 2 summarizes the information stored per mutation.

A WWW-based form gives access to a system that allows the user to query the information. Figure 1 shows the layout of this form.

*To whom correspondence should be addressed. Email: vriend@cmbi.kun.nl

Table 1. Mutation information resources

Name	# Mutants found	Address	Ref.
SWISS-PROT/TrEMBL	359	www.expasy.org	3
OMIM	156	www.ncbi.nih.gov/Omim	4
VDR	48	vdr.bu.edu	
PNR	21	www.retina-international.com/sci-news/nr2e3mut.htm	5
GRR	256	nrr.georgetown.edu/GRR/mutation/mutation.html	6
MuteXt[a]	378	www.receptors.org/NR/mutation	
Manual	136		

Nuclear receptor information resources. The seven tables hold the name of the resource, the number of NR mutations we found in the resource, its location in the WWW, and the reference. Due to redundancy, the total number of mutations given by the table is not 893.
[a]This is a still to be published attempt by FH to extract mutation information from electronic articles automatically, using computer programs that work without human intervention.

Table 2. Information per mutant

Field	Example	Description
Mutation	D351Y	Mutation notation
Accession	463	Unique identifier
Pentapeptide	LADRE	Pentapeptide sequence with mutated residue in third position
Link	www.receptors.org/NR/mutation/D351_ESR1_HUMAN.html	URL of the original mutation resource
Source	NucleaRDB	Name of the original resource
V/M	Variant	Indicates whether a mutation is a natural occurring variant or an experimental mutagen
Receptor description	ER alpha	Name of the nuclear receptor
Species	*Homo sapiens* (Human)	Species name
Id	P03372	SWISS-PROT accession code
Gene name	NR3A1	Nuclear receptor gene name
PubMed	10815929	Link to PubMed abstract
Position in alignment	340	Residue position according to the structural alignment
Effect	Cadmium did not activate mutants E523Q, E523A, H524A, or D538N but activated E380Q...	Description of the observed effect of the mutant

Three columns are given. Left: the name of the data field. Middle: example of actual data stored in NRMD. Right: description. Mutations can also be insertions, chimers, deletions, or combinations. Bibliographic information is the PubMed index. Effects include all described effects ranging from expression pattern to antagonism, but effects are only stored if available in the original resource from which the mutation was retrieved.

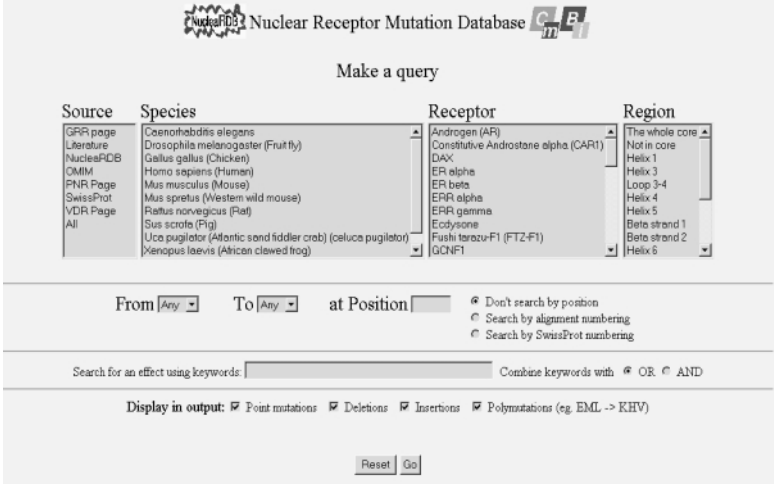

Figure 1. Front page of the NRMD. Source: origin of the mutation information. Species and Receptor: obvious. Region: indicates secondary structure element in which the mutation should reside. From, To: allow for selection of original and introduced residue type. Position allows searching for residues at a given position using the structure-based common numbering scheme or the SWISS-PROT numbering, to choice. The top four tables are multiple selection fields. The bottom selection field allows for searching the mutational effect fields using keywords.

In summary, the NRMD combines NR mutation information from all known sources. Most data collection is fully automatic for the web-based resources so that updates will automatically lead to an update of the NRMD. Moreover, mutation information from the literature continues to be added manually. To ensure high quality data, mutations are only accepted if they are made in a sequence available from SWISS-PROT or TrEMBL, and if the mutation is annotated correctly (i.e. residue number, residue type, sequence name, accession code, etc. all are the same in SWISS-PROT or TrEMBL and in the mutation information resource). Additional information about the NRMD is available at: www.receptors.org/NR/.

REFERENCES

1. Horn,F., Weare,J., Beukers,M.W., Horsch,S., Bairoch,A., Chen,W., Edvardsen,O., Campagne,F. and Vriend,G. (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.*, **26**, 275–279.
2. Horn,F., Vriend,G. and Cohen,F.E. (2001) Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic Acids Res.*, **29**, 346–349.
3. Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
4. McKusick,V.A. (1998) Mendelian Inheritance in Man. *Catalogs of Human Genes and Genetic Disorders*, 12th edn. The Johns Hopkins University Press, Baltimore, MD.
5. Haider,N.B., Jacobson,S.G., Cideciyan,A.V., Swiderski,R., Streb,L.M., Searby,C., Beck,G., Hockey,R., Hanna,D.B., Gorman,S. *et al.* (2000) Mutation of a nuclear receptor gene, NR2E3, causes enhanced S cone syndrome, a disorder of retinal cell fate. *Nature Genet.*, **24**, 127–131.
6. Martinez,E., Moore,D.D., Keller,E., Pearce,D., Vanden Heuvel,J.P., Robinson,V., Gottlieb,B., MacDonald,P., Simons,S.Jr., Sanchez,E. *et al.* (1998) The Nuclear Receptor Resource: a growing family. *Nucleic Acids Res.*, **26**, 239–241.