

The RNA helicase database

Anja Jankowsky^{1,2}, Ulf-Peter Guenther^{1,2} and Eckhard Jankowsky^{1,2,*}

¹Center for RNA Molecular Biology and ²Department of Biochemistry, School of Medicine, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, USA

Received September 15, 2010; Accepted October 6, 2010

ABSTRACT

RNA helicases are ubiquitous and essential enzymes that function in nearly all aspects of RNA metabolism. The RNA helicase database (www.rnahelicase.org) integrates the wealth of accumulating information on RNA helicases in a readily accessible format. The database is a portal that allows straightforward retrieval of comprehensive information on sequence, structure and on biochemical and cellular functions of all RNA helicases from the most widely used model organisms *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, mouse and human. Also included are RNA helicases from other organisms that are subject to specific investigation. The database is structured according to the most recent helicase classification into helicase superfamilies (SFs) and families, and thus emphasizes phylogenetic relations between RNA helicases as well. Information on individual RNA helicases can be accessed through various browsing routes or through text-based searches of the database.

INTRODUCTION

Nearly all aspects of RNA metabolism involve RNA helicases, enzymes that use ATP to bind or remodel RNA and RNA–protein complexes (1). As one of the largest class of enzymes in RNA metabolism, RNA helicases are encoded by organisms from all kingdoms of life and by many viruses (2). In structure and sequence, RNA helicases are closely related to DNA helicases (3). However, RNA helicases outnumber their DNA bound cousins by a considerable margin and RNA helicases often perform functions substantially different from those attributed to DNA helicases (1–3).

RNA helicases are not only essential for most processes of RNA metabolism including ribosome biogenesis, pre-mRNA splicing and translation initiation (1–3), but also for sensing viral RNAs in the context of the innate

immune system and for the biogenesis and function of miRNAs (4,5). Defects or misregulation of certain RNA helicases have been linked to numerous health issues including cancer, neuro-degenerative disorders and infectious diseases (6–9). Given their central biological roles, RNA helicases are subject to intensive ongoing research in diverse fields. To date, more than 7000 articles and roughly 500 reviews have been published on subjects related to RNA helicases. Currently, about 500 publications appear per year, with an upward trajectory.

AIMS AND SCOPE OF THE RNA HELICASE DATABASE

With increasing volume and diversity of the accumulating data on RNA helicases the need arises to integrate the wealth of information in a timely and readily accessible format. To address this issue, we have compiled the RNA helicase database (www.rnahelicase.org). This database is a completely restructured version of the DExH/D protein database from 1999, which covered only a subset of RNA helicases (10). We have now expanded the scope to RNA helicases in general. We approached the design of the RNA helicase database from the view of user who is interested in the fast retrieval of comprehensive information on one or more specific RNA helicases, and in examining phylogenetic relations of RNA helicases to each other. Therefore, the RNA helicase database aims foremost to enable researchers to locate and retrieve comprehensive information about sequence, structure, biochemical and cellular function of RNA helicases. The database provides ready access to sequence, structure, biochemical and cellular function of all RNA helicases from the most widely used model organisms *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, mouse and human. Also included are RNA helicases from other organisms that are subject to specific investigation, e.g. NS3 from hepatitis C virus (11), the P4 RNA packaging motor from Φ 12 (12) or RNA helicase A from bovine (13). Focus on these RNA helicases covers the vast majority of published data on RNA helicases (estimated to >95%), while providing an enduring structure for the database.

*To whom correspondence should be addressed. Tel: +1 216 368 3336; Fax: +1 216 368 3419; Email: exj13@case.edu

Notwithstanding, the database can be readily expanded to other organisms, if this need arises in the future.

In addition to enabling access to information on individual RNA helicases, the database emphasizes phylogenetic relations between related RNA helicases. As briefly outlined below and in more detail in several recent reviews, RNA helicases are categorized in superfamilies and families, and many features are shared between

RNA helicases of a given family (3,14,15). Functions and features of one RNA helicase often provide clues for other, less well-characterized enzymes from the same family (3). However, family relations are often not apparent to investigators outside the RNA helicase field. We therefore believe that emphasizing family relations is particularly useful for database users new to research on RNA helicases.

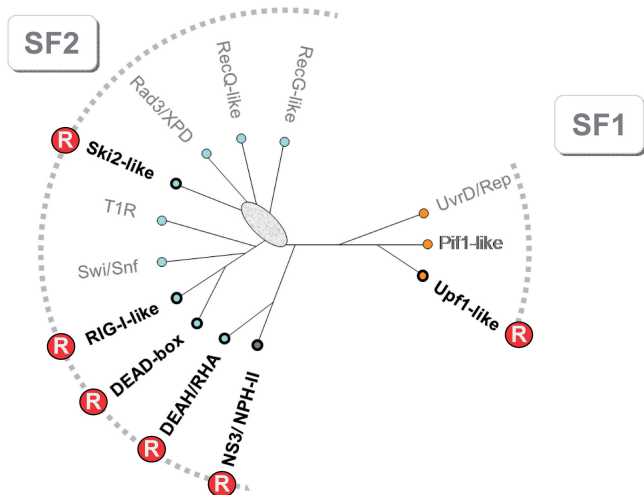


Figure 1. The families of the SF1 and SF2. The cladogram was generated from an structure-aided sequence alignment of all SF1 and 2 proteins in yeast and human, described in (3). The alignment can also be found on the RNA helicase primer page of the database. Branch lengths of the cladogram are not to scale. The oval marks uncertainty in the topology. Families containing RNA helicases are marked with a circled R, the other families contain DNA helicases. (T1R, Type 1 restriction enzymes; RHA, RNA helicase A).

STRUCTURE OF THE DATABASE

The structure of the RNA helicase database is based on the most recent helicase classification into helicase superfamilies (SFs) and families (3,14,15). According to sequence and protein structures, all RNA and DNA helicases are classified into six SFs (15). SFs 1 and 2 contain non-ring-forming helicases, SFs 3–6 contain ring-forming enzymes. RNA helicases are found in SFs 1–5. All eukaryotic RNA helicases identified to date belong to the non-ring-forming SFs 1 and 2 (3). Ring-forming RNA helicases are found in bacteria and viruses (12,16). SFs 1 and 2 are further divided into families (Figure 1). Although all SF1 and SF2 proteins share extensive structural similarities in the overall fold of their helicase core, the respective families have distinct structure and sequence characteristics that are reflected in specific functional features (3).

The user enters the database through the homepage (Figure 2). From this page, via the main navigation (Figure 3A), pages for the helicase SFs 3–5, as well as pages for the SF1 and 2 families can be accessed (Figures 2 and 3B). From these pages, one can directly

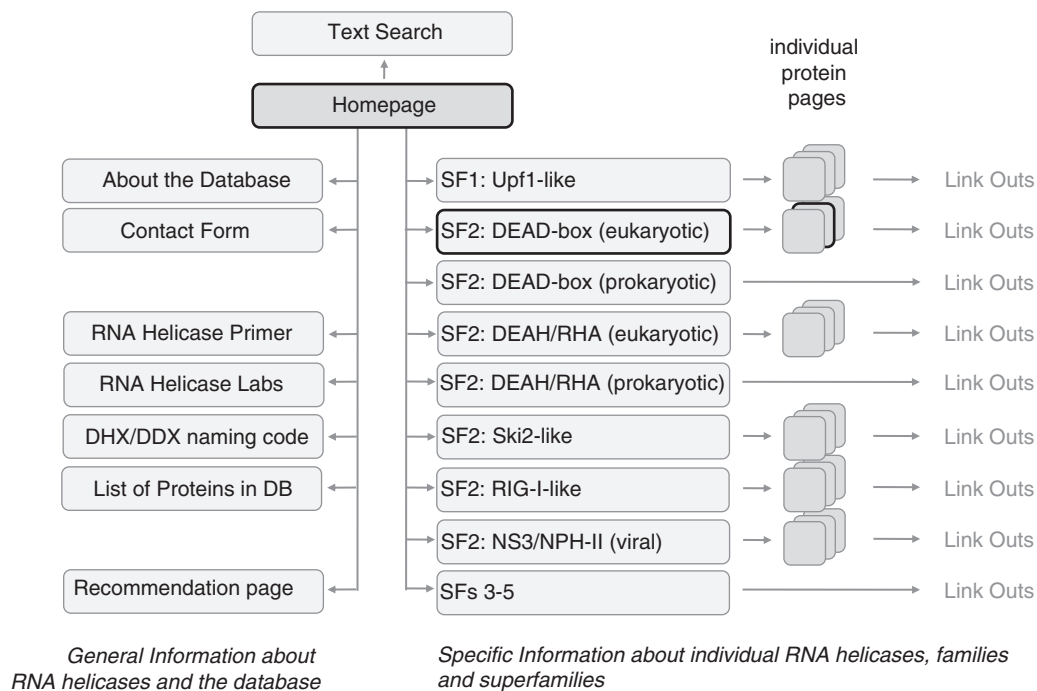


Figure 2. Structure of the RNA helicase database. Gray fields represent database pages, lines show the main connections. Link out indicates links to external sequence databases (see text for details). The bold lines around the Homepage, the SF2: DEAD-box page and one individual DEAD-box protein page mark pages shown as screenshots in Figure 3.

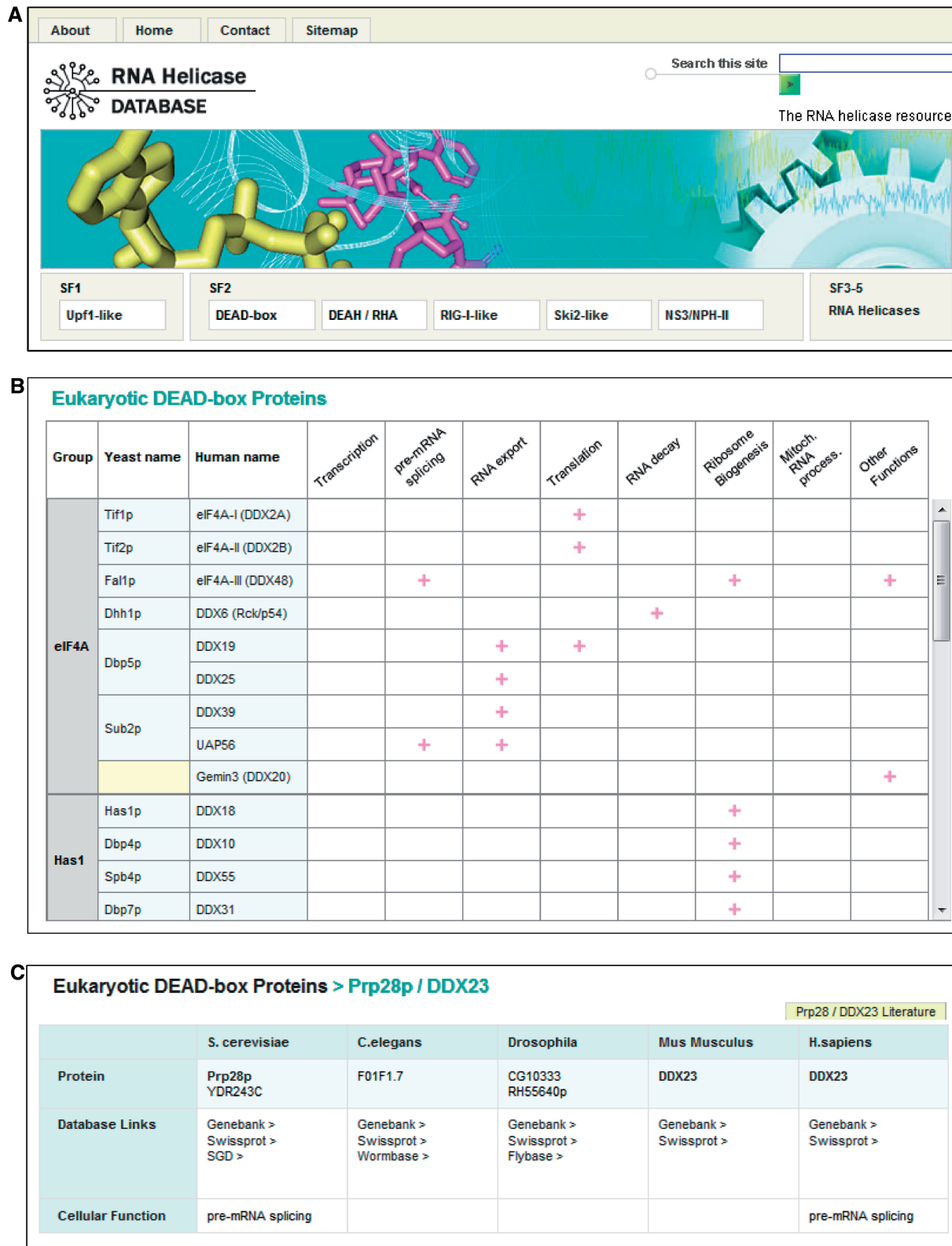


Figure 3. Screenshots of representative pages. (A) The main navigation is identical on all pages and provides immediate access to the SF3–5 and SF1 and 2 family pages, as well as to the search function. (B) Screenshot of the page for eukaryotic DEAD-box proteins. The pages for individual proteins are accessed through both yeast and human names. Cellular function marks the demonstrated involvement of one ortholog from either organism in the indicated process. (C) Page for an individual protein (Prp28p). Database links connect directly to the specific protein in the sequence databases indicated. Cellular function is only marked if involvement in a specific process has been directly demonstrated for a given protein. The (Prp28/DDX23) literature button provides a direct link to a customized PubMed search for all orthologs listed on the page.

reach 'individual protein pages' (Figure 2). These pages contain information on orthologs of a given RNA helicase from *S. cerevisiae*, *C. elegans*, *D. melanogaster*, mouse and human (Figure 3C). DEAD-box and DEAH/RHA helicases from *E. coli* are listed on separate pages (Figure 2), because there are considerably fewer enzymes than in eukaryotes and because the structure of the individual protein pages, which emphasizes orthologs of multiple organisms, appears not suitable for representing only one protein from *E. coli*.

Individual protein pages provide links to various sequence databases with original information about a particular RNA helicase. Links for each protein exist to Genbank and Uniprot, as well as to organism-specific databases SGD (*S. cerevisiae*), Wormbase (*C. elegans*) and Flybase (*D. melanogaster*) (Figure 3C). If structural information is available, respective links to the PDB database are given. The user also has the option to perform a targeted PubMed literature search for the helicase orthologs (Figure 3C). The cellular function of each protein is given only where explicitly tested, although it is generally believed that the cellular functions are largely conserved for orthologs. Finally, a curated alignment of the listed orthologs is provided. This alignment enables the user to identify regions of conservation among orthologs that extend beyond the characteristic helicase motifs.

Individual protein pages were designed based on our roughly decade-long experience with the DExH/D database. Guided by this experience, we chose to directly refer to the large sequence databases for detailed protein information, and not to copy this information into our database. Directly linking to comprehensive, constantly updated sequence databases ensures that information is current without direct intervention from our side.

Further features of the RNA helicase database include full-search ability by text-based queries that are commonplace from the ubiquitous Google search engine (Figure 3A). The search function can be accessed from any page and provides an alternative, yet important route to individual protein pages. A third route of access to individual protein pages is established by a list of proteins in the database (Figure 2). This list shows all RNA helicases in our database, ordered by SF and family and by the organism of origin. Our rationale for incorporating the various routes of access to the individual protein pages is the heterogeneity in the database users, which ranges from researchers looking for updates on information on a helicase that they have long been working on, to those who retrieve the first comprehensive information on a protein they just came across in their research. Finally, the database also contains other information of interest for research on RNA helicases, such as an RNA helicase primer, the naming code for DDX/DHX RNA helicases and a list of laboratories with specific interest in RNA helicases.

ACKNOWLEDGEMENTS

We thank the many users of previous versions of our database for constructive feedback and valuable suggestions.

FUNDING

Funding for open access charge: National Institutes of Health (GM067700 to E.J.).

Conflict of interest statement. None declared.

REFERENCES

- Jankowsky,E. (2010) RNA Helicases at work: binding and rearranging. *Trends Biochem. Sci.*, in press.
- Jankowsky,E. and Fairman-Williams,M.E. (2010) An introduction to RNA helicases: superfamilies, families, and major themes. In Jankowsky,E. (ed.), *RNA helicases*, Vol. 19. RSC Biomolecular Sciences, London, pp. 1–31.
- Fairman-Williams,M.E., Guenther,U.P. and Jankowsky,E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
- Yoneyama,M. and Fujita,T. (2010) Recognition of viral nucleic acids in innate immunity. *Rev. Med. Virol.*, **20**, 4–22.
- Linder,P. and Owttrim,G.W. (2009) Plant RNA helicases: linking aberrant and silencing RNA. *Trends Plant Sci.*, **14**, 344–352.
- Abdelhaleem,M. (2004) Do human RNA helicases have a role in cancer? *Biochim. Biophys. Acta*, **1704**, 37–46.
- Fogel,B.L. and Perlman,S. (2006) Novel mutations in the senataxin DNA/RNA helicase domain in ataxia with oculomotor apraxia 2. *Neurology*, **67**, 2083–2084.
- Nelson,P.T. and Keller,J.N. (2007) RNA in brain disease: no longer just “the messenger in the middle”. *J. Neuropathol. Exp. Neurol.*, **66**, 461–468.
- Kolb,S.J., Sutton,S. and Schoenberg,D.R. (2010) RNA processing defects associated with diseases of the motor neuron. *Muscle Nerve*, **41**, 5–17.
- Jankowsky,E. and Jankowsky,A. (2000) The DExH/D protein family database. *Nucleic Acids Res.*, **28**, 333–334.
- Frick,D.N. and Lam,A.M. (2006) Understanding helicases as a means of virus control. *Curr. Pharm. Des.*, **12**, 1315–1338.
- Tuma,R. (2010) Hexameric Viral RNA Helicases. In Jankowsky,E. (ed.), *RNA Helicases*, Vol. 19. RSC Biomolecular Sciences, London, pp. 213–242.
- Zhang,S. and Grosse,F. (2004) Multiple functions of nuclear DNA helicase II (RNA helicase A) in nucleic acid metabolism. *Acta Biochim. Biophys. Sin.*, **36**, 177–183.
- Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases: amino acid comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, **3**, 419–429.
- Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Ann. Rev. Biochem.*, **76**, 23–50.
- Rabhi,M., Rahmouni,R. and Boudvillain,M. (2010) Transcription termination factor rho: a ring-shaped RNA helicase from bacteria. In Jankowsky,E. (ed.), *RNA helicases*, Vol. 19. RSC Biomolecular Sciences, London, pp. 243–271.