ENCYCLOPEDIA OF MOLECULAR BIOLOGY

VOLUME 2

ENCYCLOPEDIA OF

MOLECULAR BIOLOGY

VOLUME 2

Thomas E. Creighton

European Molecular Biology Laboratory
London, England

A Wiley-Interscience Publication
John Wiley & Sons, Inc.
New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This book is printed on acid-free paper. ♾

Library of Congress Cataloging-in-Publication Data:

Creighton, Thomas E., 1940–
The encyclopedia of molecular biology / Thomas E. Creighton.
p. cm.
Includes index.
ISBN 0-471-15302-8 (alk. paper)
1. Molecular biology–Encyclopedias. I. Title.
QH506.C74 1999
572.8′03—dc21
99-11575
CIP

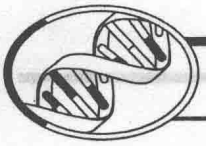
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

ENCYCLOPEDIA OF

MOLECULAR BIOLOGY

VOLUME 2



WILEY BIOTECHNOLOGY ENCYCLOPEDIAS

Encyclopedia of Bioprocess Technology: Fermentation, Biocatalysis, and Bioseparation
Edited by Michael C. Flickinger and Stephen W. Drew

Encyclopedia of Molecular Biology
Edited by Thomas E. Creighton

Encyclopedia of Cell Technology
Edited by Raymond E. Spier

Encyclopedia of Ethical, Legal, and Policy Issues in Biotechnology
Edited by Thomas J. Murray and Maxwell J. Mehlman

ENCYCLOPEDIA OF MOLECULAR BIOLOGY

ENCYCLOPEDIA OF

MOLECULAR BIOLOGY

VOLUME 2

D

D GENE SEGMENT

MICHEL FOUGEREAU

The existence of D genes encoding a short region of the third complementarity determining region (CDR3) of **immunoglobulin** heavy chains was first postulated when it was apparent that expressed H chains between the sequences encoded by the **variable** (V) and **J (joining)** genes contained a segment variable in length that had no counterpart in either gene. It turned out that variability of this region resulted from 2 causes: One was the existence of a set of D genes (D for Diversity); the second was due to the presence of nucleotides that were not germline encoded but added enzymatically at both the VD and the DJ junctions by terminal-deoxynucleotidyltransferase, which is specifically expressed in lymphocytes. As the result of these combined events, heavy-chain CDR3 appeared as the most hypervariable region of Ig molecules, and therefore of prime importance as a basis for antibody specificity.

D gene segments appear to be present in all heavy-chain gene loci, from the lowest vertebrate species. In humans, about 30 D gene segments have been identified, clustered between the IgVH locus and the JH genes. The various D genes differ in length, but are always limited to a small number of potential **codons**. They are flanked on both their 5′ and 3′ ends by the conventional recombination signal sequences recognized by the RAG1 and RAG2 **recombinases**. The spacers are of different length on each side, so they meet with the 12/23 alternate size rule, to recombine with VH and JH gene segments. D genes make a major contribution to heavy-chain variability of CDR3, not only because of their number and sequence diversity, but also because they are used in many alternate possibilities: (a) They may be inserted in either orientation; (b) any reading frame may be used (in humans, not in the mouse; see text below), although one is more frequently encountered; (c) they may join to each other, so that examples of as many as four fused D genes have been described, thus providing a CDR3 of unusual length; and (d) they may be partially deleted at both ends before being inserted between the V and J genes. All of this clearly can generate a huge diversity potential.

In the mouse, minor differences from the above situation in D gene use and function have been reported, due to the fact that there is essentially only one reading frame used by murine D genes. There are two reasons for this. One **gene rearrangement** introduces a **stop codon** in the 3′ JC sequence. The second reason is more complicated. Some mouse D genes have a **promoter**-like region that can be activated after the first gene rearrangement making the DJ joint has been performed. In that case, a D-J-Cμ protein is synthesized, which becomes exposed at the surface of the preB cell and blocks any further gene rearrangement (see **B cell**), but is unable to contribute a functional immunoglobulin.

The genes for **T-cell receptors** (TCRs) have also D genes on the β and δ chains. There are only a number of D genes, which are used in either orientation and reading frame. They have 12-bp spacers of identical length on both 5′ and 3′ flanking recombination signal sequences, corresponding to the 23-bp spacers of the V and J genes. TCR D genes also contribute to CDR3, a region that is most variable, as for immunoglobulins.

See also entries **Gene rearrangement** and **Recombinase**.

Suggestions for Further Reading

F. Matsuda and T. Honjo (1994) Organization of the human immunoglobulin heavy-chain locus. *Adv. Immunol.* **62**, 1–29.

H. Sakano, Y. Kurosawa, M. Weigert, and S. Tonegawa (1982) Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy chain genes. *Nature* **290**, 562–565.

H. Gu, D. Kitamura, and K. Rajewsky (1991) B-cell development regulated by gene rearrangement—arrest of maturation of membrane bound Dμ protein and selection of DH element reading frames. *Cell*, **65**, 47–54.

DANSYL CHLORIDE

HANS JÖRNVALL
WILLIAM GRIFFITHS
TOMAS BERGMAN

The repetitive nature of the **Edman degradation** reaction opened the way to **protein sequencing** and to automation in the form of protein sequencers, but manual Edman degradations also became popular during much of the 1960s and 1970s.

One major manual method used dansyl (1-dimethylaminonaphthalene-5-sulphonyl) chloride (Fig. 1, see top of next page) to detect the new N-terminus after each Edman cycle (1). Combined with purification of peptides on paper by high-voltage **electrophoresis** and **chromatography** (see **Peptide mapping**), it made protein sequence analysis accessible to many laboratories without the need for large equipment. It also avoided the difficulties in identification of the phenylthiohydantoin (PTH)-amino acids that existed early, and replaced PTH-identification with dansyl-amino acid identification. The latter became rapid and reliable by **thin-layer chromatography** (2). Hence, many of the original protein **primary structures** determined used this methodology, originating from the MRC Laboratory of Molecular Biology, Cambridge, England. Together with the Sanger method of DNA sequencing later developed there, the "dansyl-Edman/paper purification" technique made this laboratory the leader in sequencing technology for a long time in the 1960s and 1970s.

Dansyl chloride reacts with protein N-termini much like the original Sanger reagent, fluorodinitrobenzene (FDNB) (3), but it is easier to use, more sensitive, and fluorescent, making it easily detectable. Dansyl chloride is still used for protein labeling in many cases where fluorescence is desirable.

BIBLIOGRAPHY

1. W. R. Gray (1967) *Methods Enzymol.* **11**, 469–475 and (1972) **25**, 121–138.

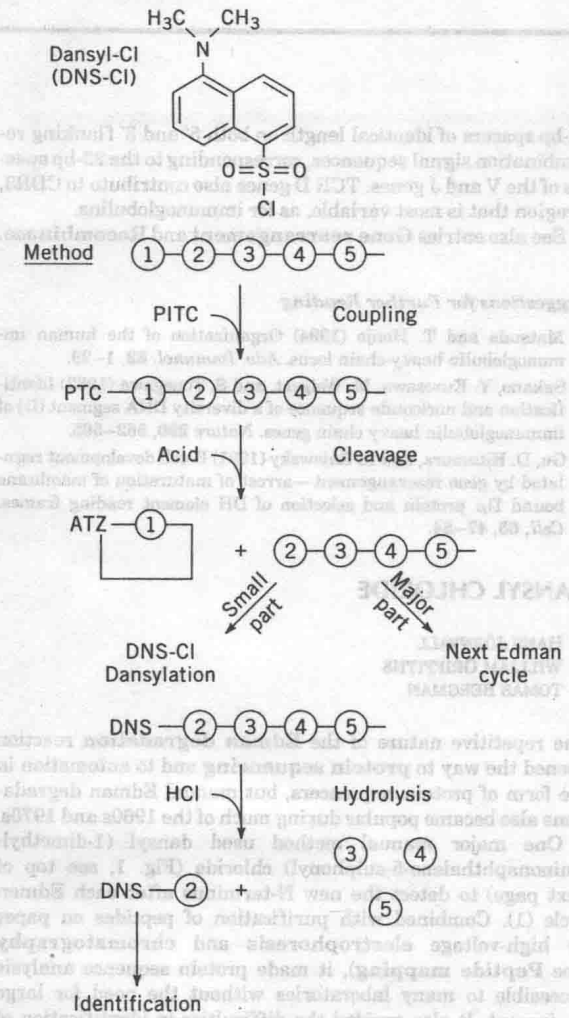


Figure 1. Principle of the "dansyl-Edman" modification of the Edman degradation reaction, with N-terminal analysis of separate samples. PITC, phenylisothiocyanate; ATZ, anilinothiazolinone; PTH, phenylthiohydantoin; Dansyl, 1-dimethylaminonaphthalene-5-sulphonyl.

2. K. R. Woods and K.-T. Wang (1967) *Biochim. Biophys. Acta* **133**, 369–370.

3. F. Sanger (1945) *Biochem. J.* **39**, 507–515.

DATABASES

MINORU KANEHISA

Molecular biology is an empirical discipline that requires observing and understanding different types of data. For example, a **gene** may imply a physical location on the **chromosome**, a **nucleotide sequence**, an **amino acid** sequence, a three-dimensional **protein structure**, a molecular component of cellular function, a regulatory mechanism of **gene expression**, or even a **phenotypic** difference caused by **mutation**. In addition to this variability, the quantity of molecular biology data is increasing rapidly, especially for gene and protein sequences and 3-D structures, due to advances in experimental technologies. Molecular biology databases are a number of resources available over the Internet that comprise a **bioinformatics** infrastructure for biomedical sciences. Each database contains a specific type of data that has cross-references to other databases, which can be used for integrated information retrieval. This is possible because a database is generally organized as a collection of entries, and connections can be made at the level of entries without standardizing how data items should be organized within an entry. The advent of the World Wide Web (WWW) in the early 1990s was a boon to molecular biology databases because the concept of hyperlinks is fully compatible with the practice of cross-references. The WWW also dramatically increased the accessibility of computers to biologists. The primary resources of molecular biology databases are bibliographic databases, **sequence databases**, and **structure databases**.

Although bibliographic, sequence, and structural aspects of molecular biology are relatively easy to computerize, the next step is to organize their functional aspects. There are resources in that direction, such as motif libraries that contain higher level knowledge abstracted from sets of functionally related sequences and pathway databases that contain computerized knowledge of molecular interactions and biochemical pathways. Representative examples of these resources are shown in Table 1.

Abstraction of a real problem is made through a data model. For example, data are organized in two-dimensional tables in the relational data model. The relational database based on the relational model has been widely used in a number of applications, including some of the sequence databases. Although in principle all different types of molecular biology data can be stored in a single, unified, relational database, this is impossible in practice because of the varying views of how data items should be organized and related. In the current web of molecular biology databases, different types of data are integrated by a loose coupling based on links (cross-references), rather than a tight coupling based on unified schema. This approach is extended to include other types of links, especially similarity

Table 1. Selected List of Molecular Biology Databases

Data Type	Database	Organization
Biomedical literature	Medline	National Center for Biotechnology Information (NCBI)
Nucleotide sequence	GenBank	
	EMBL	European Bioinformatics Institute (EBI)
	DDBJ	National Institute of Genetics, Japan
Amino acid sequence	PIR	National Biomedical Research Foundation
	SWISS-PROT	Swiss Institute of Bioinformatics (SIB)
Sequence motif	PROSITE	
3-D structure	PDB	Brookhaven National Laboratory
Biochemical pathway	KEGG	Kyoto University

Table 2. WWW Addresses for the Major Bioinformatics Servers

Server	Address
NCBI	www.ncbi.nlm.nih.gov
EBI	www.ebi.ac.uk
ExPASy	www.expasy.ch
GenomeNet	www.genome.ad.jp

links computed by similarity search algorithms and biological links representing molecular interactions, which can also be integrated for biological reasoning (1). The major bioinformatics servers shown in Table 2 provide link-based database retrieval systems, such as Entrez at NCBI, SRS at EBI, and DBGET/LinkDB at GenomeNet in Kyoto.

In scientific disciplines, the merit of storing and managing information in a computer was first realized in bibliographic databases, which were designed for humans to read and understand. In the next step, factual data reported in the literature were computerized in factual databases, such as in sequence databases, which made more sophisticated retrieval available, for example, sequence similarity searches. Even in this case, however, the database is still a static resource to be retrieved, and it is up to humans to make sense out of the retrieved data. In contrast, the knowledge of links or relationships is more dynamic in nature. For example, ancestors in a family can be retrieved from a "deductive" database that contains parent–child relationships and the rules for combining them. Thus, knowledge is different from data or information, in that new knowledge can be generated dynamically from existing knowledge by logical reasoning. In the era of mass data production, molecular biology requires logical computation based on empirical knowledge rather than numerical computation based on first principles (see **Bioinformatics**). The web of molecular biology databases also requires a new generation of knowledge bases.

BIBLIOGRAPHY

1. M. Kanehisa (1997) *Trends Biochem. Sci.* **22**, 442–444.

DAUNOMYCIN

J. B. CHAIRES

Daunomycin (synonym: daunorubicin) is the prototype of the clinically important anthracycline antibiotics that are widely used in cancer chemotherapy. Daunomycin and its close relative **adriamycin** are natural products isolated from various *Streptomyces* species. These two anthracyclines are probably the most widely used antitumor agents worldwide (1,2). Daunomycin is primarily used in the treatment of acute leukemias. The structure of daunomycin is shown in Figure 1; it is composed of two major parts: an anthraquinone ring system and the pendant daunosamine, an amino sugar.

Daunomycin rapidly permeates into cells, passing through the **membrane** into the cytoplasm, and then accumulates in the nucleus, where it binds avidly to **DNA**. It binds to DNA by the process of **intercalation**, in which the anthraquinone ring inserts between adjacent DNA base pairs. The daunosamine moiety provides additional interactions that

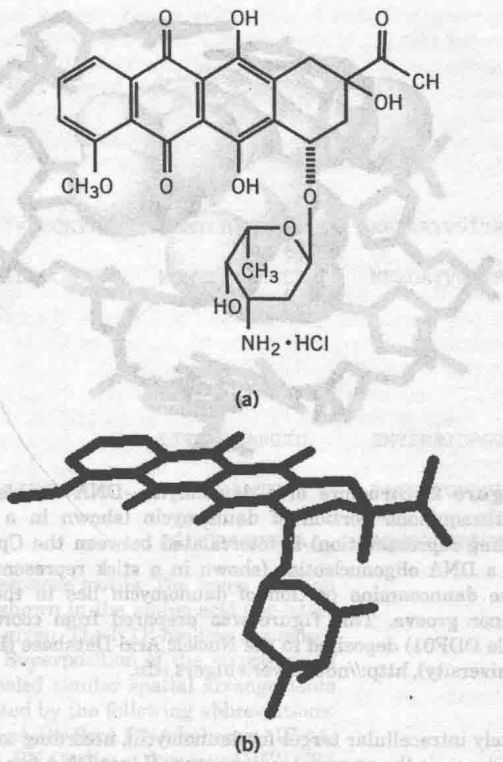


Figure 1. The chemical structure (a) and molecular model (b) of the anthracycline antibiotic daunomycin.

stabilize the complex by fitting into the minor groove of the DNA double helix, where it participates in **van der Waals interactions**. These additional interactions in the minor groove distinguish daunomycin from simple intercalators like **ethidium bromide**, which are stabilized only by the interactions within the intercalation site. Figure 2 shows the structure of a daunomycin–DNA complex determined by **X-ray crystallography** (3). Intercalation simultaneously lengthens and unwinds the DNA duplex, and it stabilizes DNA against thermal denaturation, increasing its **Tm (melting temperature)**.

Daunomycin, despite its small size (527.5 MW), exhibits several levels of **molecular recognition** (4). It is structurally specific and strongly prefers to bind to right-handed, B-form DNA over alternate helical forms. Daunomycin will **allosterically** convert left-handed Z DNA to an intercalated, righted-handed form (5). There is pronounced sequence selectivity in the binding of daunomcyin to DNA, which is revealed by DNase I **footprinting** experiments. Daunomycin binds preferentially to the triplet sequences 5′(A/T)GC and 5′(A/T)CG, where the notation (A/T) indicates that either A or T may occupy the sequence position (6,7).

The mechanism by which daunomycin acts to kill cancer cells remains incompletely defined, and it may involve inhibition of several different cellular functions. Daunomycin, like most intercalators, is an effective inhibitor of **DNA replication** and of **transcription**, both *in vitro* and *in vivo*. The most

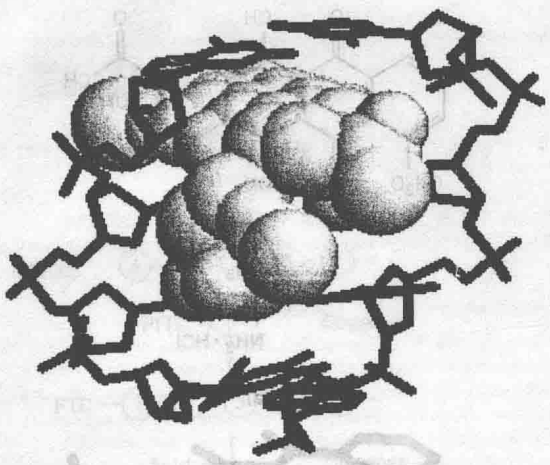


Figure 2. Structure of a daunomycin–DNA complex. The anthraquinone portion of daunomycin (shown in a space-filling representation) is intercalated between the CpG step in a DNA oligonucleotide (shown in a stick representation). The daunosamine portion of daunomycin lies in the DNA minor groove. This figure was prepared from coordinates (file DDF01) deposited in the Nucleic Acid Database (Rutgers University), http://ndbserver.rutgers.edu.

likely intracellular target for daunomycin, according to recent evidence, is the enzyme topoisomerase II (see **DNA topology**). Daunomycin poisons this key enzyme, by trapping an intermediate enzymatic species in which cleaved DNA strands are covalently attached to topoisomerase II, by preventing the normal resealing of the duplex DNA after strand passage (8). A comprehensive study of several anthracycline derivatives, including daunomycin, showed that intercalation is necessary, but not sufficient, for inhibition of topoisomerase II (9).

The anthraquinone moiety of daunomycin readily undergoes a number of oxidation–reduction reactions, with the production of oxygen free radicals. These free radicals can damage cells, and they may represent another important mechanism in the cytotoxicity of daunomycin. A serious, complicating side effect in the clinical use of daunomycin in cancer chemotherapy is its cardiotoxicity (1,2). Free-radical production has been implicated as the primary cause of such cardiotoxicity.

Daunomycin was one of the first effective cancer chemotherapeutic agents identified, and it remains clinically important after over 30 years of use. It is one of the best characterized and understood intercalators and serves as an important model for understanding how small molecules bind to DNA.

BIBLIOGRAPHY

1. R. B. Weiss (1992) *Semin. Oncol.* **19**, 670–686.
2. G. N. Hortobagyi (1997) *Drugs* **54**, 1–7.
3. A. H.-J. Wang (1992) *Curr. Opin. Struct. Biol.* **2**, 361–368.
4. J. B. Chaires (1996) in *Advances in DNA Sequence Specific Agents*, Vol. 2 (L. H. Hurley and J. B. Chaires, eds.), JAI Press, Greenwich, CT, pp. 141–167.
5. J. B. Chaires (1986) *J. Biol. Chem.* **261**, 8899–8907.
6. B. Pullman (1990) *Anti-cancer Drug Design* **7**, 96–105.
7. J. B. Chaires, J. E. Herrera, and M. J. Waring (1990) *Biochemistry* **29**, 6145–6153.
8. F. Zunnino and G. Capranico (1990) *Anti-cancer Drug Design* **5**, 307–317.
9. A. Bodley et al. (1989) *Cancer Res.* **49**, 5969–5978.

Suggestions for Further Reading

F. Arcamone (1981) *Doxorubicin*, Academic Press, New York.

J. W. Lown, ed. (1988) *Anthraacycline and Antrhacenedione Based Anticancer Agents*, Elsevier, Amsterdam.

W. Priebe, ed. (1995) *Anthracycline Antibiotics: New Analogues, Methods of Delivery, and Mechanism of Action*, ACS Symposium Series 574, American Chemical Society, Washington, DC.

DEAD AND DEAH DOMAINS

NANHUA YAO
PATRICIA C. WEBER

DEAD and DEAH domains contain a characteristic Asp-Glu-Ala-Asp/His sequence that is usually abbreviated DEAD or DEAH, using the amino-acid one-letter code. The **domains** belong to a large **superfamily** of **enzymes** involved in nucleic acid metabolism. Both DEAD and DEAH domains alter nucleic acid secondary structure in a nucleoside triphosphate (NTP)-dependent manner. Many function as either **DNA helicases** or **RNA helicases**, which unwind DNA and RNA duplexes, respectively.

The general properties of DEAD/DEAH domains are summarized, and the functions of several well-characterized DEAD and DEAH domains are reviewed. Included are the DEAD domain of eukaryotic initiation factor 4a, eIF-4A, and the precursor RNA processing (PRP) DEAH domains found in the **spliceosome**, a multiprotein complex responsible for processing pre-**messenger RNA** into mRNA (see **RNA splicing**). Finally, the structural basis for sequence conservation will be discussed using the **X-ray crystallographic** structures of the DExx DNA helicases from *Bacillus stearothermophilus* and *Escherichia coli* and the DExH RNA helicase from the hepatitis C virus (HCV).

IDENTIFICATION AND CLASSIFICATION OF DEAD AND DEAH DOMAINS

Helicases are enzymes that remove double-stranded regions of nucleic acid when single-stranded RNA and DNA are required. Regulated production of single-stranded nucleic acids occurs during many cellular processes, including replication, **recombination**, and **transcription**. The energy required for unwinding is derived from the hydrolysis of NTPs such as ATP.

Amino acid sequences can be used to identify and classify helicases. Helicase sequences contain seven regions or motifs of conserved residues (1). On the basis of the sequences of the individual motifs, helicases have been broadly grouped into two major **superfamilies**, SFI and SFII, and several smaller families [(2), Fig. 1). With the increased availability of **genome** information, helicases are frequently identified from their sequences alone, prior to the demonstration of helicase activity *in vitro*.

The energy for helicase-mediated unwinding of nucleic acid duplexes is derived from the hydrolysis of NTPs. Consequently,

	Motif						
	I	Ia	II	III	IV	V	VI
Protein							
SFI							
Rep	AGAGSGKT	FTNKAA	DEYQ	VGDDDQS	IKLEQNYRS	QLMTLHASKGLE	ERRLAYVGITRAQ
PCRA	AGAGSGKT	FTNKAA	DEYQ	VGDADQS	ILLEQNYRS	MLMTLHAAKGLE	ERRLAYVGITRAE
SF2							
eIf-4A	AQSGTGKT	PTRELA	DEAD	LSATMP		LITTDLLARGID	ENYIHRIGRGGRF
HCV	APTGSGKS	PSVAAT	DECH	ATATPP	LIFCHSKKK	VVSTDALMTGFT	DAVSRTQRRGRTG
PRP22	GETGSGKT	PRRVAA	DEAH	VTSATL		DGIYYVVDPGFA	AQNQRKGRAGRTG

Figure 1. Representative examples of conserved sequence motifs. The seven common sequence motifs of DEAD and DEAH domains are shown in the amino acid one-letter code. Helicases from both superfamilies I and II are given. The NTP-binding sequences within motifs I and II are most highly conserved. Superposition of the structures of the *E. coli* Rep DNA and HCV RNA helicases revealed similar spatial arrangements of all motifs except motif IV (22). Proteins are denoted by the following abbreviations: Rep, *E. coli* Rep DNA helicase; PcrA, *B. stearothermophilus* PcrA DNA helicase; eIF-4A, elongation of initiation factor 4A RNA helicase; HCV, RNA helicase domain of HCV NS3 protein, and PRP22, *S. cerevisiae* precursor RNA processing protein 22.

amino acid residues known to bind NTPs in other proteins are among the most conserved across the helicase families. By analogy with structures of ATP:protein complexes, the conserved lysine residue embedded in the Ala-x-x-Gly-x-Gly-Lys-Thr sequence of motif I probably interacts with the β and γ phosphates of NTP (Fig. 2). Similarly, NTPases share an aspartic acid residue in motif II that ligates the magnesium of the NTP: Mg^{+2} complex (Fig. 2). In the DEAD and DEAH domains, this conserved aspartic acid occurs within the title sequences Asp-Glu-Ala-Asp/His. In 1982, Walker et al. (3) identified the functional significance of these motifs through sequence comparisons of ATP-utilizing enzymes. For this reason, motifs I and II are also referred to as Walker motifs A and B. In the literature, DEAD domains are occasionally referred to as DEAD-box proteins.

Initially, the classification of helicases by residues within motif II grouped functionally related proteins. However, as more helicase sequences became known, many exceptions to the strict conservation of motif II residues within functional families have been found. Indeed, a new DExH domain family

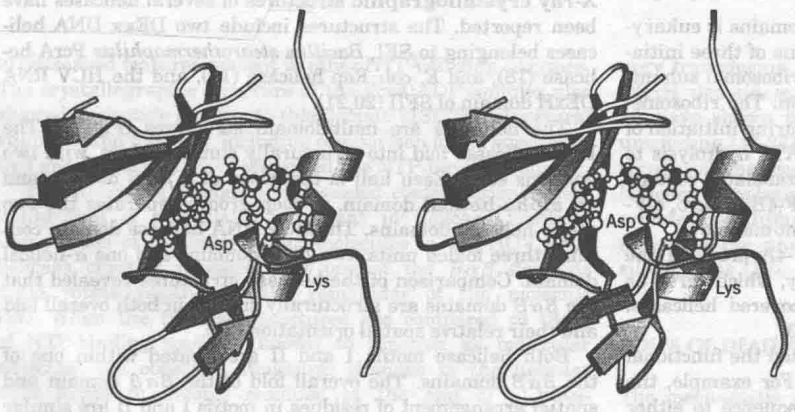


Figure 2. Stereoscopic views of an enzyme:ATP complex. Interactions made by the conserved lysine and aspartic acid residues of the protein with the ATP phosphates and Mg^{+2}, respectively, are highlighted. The protein backbone is shown as ribbons and the bound ATP in the ball-and-stick representation. Phosphate atoms are shown as solid spheres. The Mg^{+2} ion is situated between the phosphate oxygen and aspartic acid. For clarity, the complete coordination of the Mg^{+2} is not shown. Coordinates were taken from the crystal structure of the human cyclin-dependent kinase 2:ATP complex [(28), PDB entry 1HCK].

containing helicases from positive-strand viruses has been proposed (4). For these enzymes, a variety of residues can be accommodated at the third position. Overall, amino acid sequence comparisons suggest that the DEAD and DEAH domains belong to a larger family of DExx domains, where x indicates sequence variability at the third and fourth positions of the DEAD/DEAH motif.

GENERAL PROPERTIES OF DEAD AND DEAH DOMAINS

Isolated DEAD and DEAH domains typically exhibit NTPase activity that can be stimulated by the addition of nucleic acid polymers. The domains bind NTPs with **dissociation constants** in the micromolar range. K_m values for ATP of 80, 260, and 250 μM were measured for eIF-4A (5), PRP16 (6) and HCV helicases (7), respectively. Unlike eIF-4A, which is an obligate **ATPase** (8), little specificity for NTP is observed for some DEAD and DEAH domains. K_m values for ATP, CTP, GTP, and UTP ranged from 260 to 580 μM for the PRP16 DEAH domain (6) and from 250 to 1300 μM for the HCV DExH helicase (7). For the HCV helicase, the addition of poly U increases its ATPase activity approximately 10-fold (9). In the case of eIF-4A, the addition of poly U increases the ATPase activity approximately 8-fold (8).

The mechanism by which NTP hydrolysis is coupled to the modification of a nucleic acid secondary structure remains one of the most interesting questions regarding the function of DEAD and DEAH domains. As outlined below, several lines of evidence support the hypothesis that the hydrolysis reaction produces conformational changes in the domain. The changes alter the stability of the bound nucleic acid to favor the single-stranded form.

DEAD and DEAH domains typically contain ~300 residues and frequently reside within a longer sequence, or as subunits of a larger protein complex. For example, each of the seven *Saccharomyces cerevisiae* PRP splicing factors contains a DEAD/H domain. One of the larger splicing factors, PRP22, also contains an amino-terminal RNA-binding domain (6). The human DNA helicase II has two double-stranded **RNA-binding** domains, a DExH domain, and a glycine-rich domain (10). The HCV nonstructural protein 3 (NS3) is a bifunctional enzyme with an amino-terminal **serine proteinase** activity and a helicase activity in the carboxy-terminal DExH domain (11).

DEAD DOMAINS

One of the most well-characterized DEAD domains is eukaryotic initiation factor 4A, eIF-4A. eIF-4A is one of three initiation factors required for binding of the 40S ribosomal subunit to mRNA during the initiation of translation. The ribosome-binding reaction requires ATP hydrolysis. During initiation of translation, eIF-4A may use the energy of ATP hydrolysis to disentangle structured regions in the 5′-untranslated regions of messenger RNA. When complexed with eIF-4B *in vitro*, eIF-4A unwinds duplex RNA in an ATP-dependent manner.

The biochemical characterization of eIF-4A provided the first detailed views of RNA helicase activity, which serve as a framework for comparison of newly discovered helicases. Studies of eIF-4A are reviewed in (12) and (13).

Mutational studies on eIF-4A demonstrated the functional importance of residues in motifs I and II. For example, the conversion of residues within the DEAD sequence to either NEAD or DQAD (substitution of the first conserved aspartic acid with asparagine, or the second residue with glutamine) abolished ATPase activity (12). The same result was observed when the conserved lysine residue in motif I was changed to asparagine. Similar experiments with other DExH domains have confirmed that these amino acid residues are essential.

Studies of the eIF-4A mechanism support the hypothesis that alterations in RNA structure are induced by conformational changes in the DEAD domain. By monitoring the domain's susceptibility to **proteolysis**, ADP was found to cause conformational changes in both the free eIF-4A and eIF-4A:RNA complex (14). Conformational changes were also observed on the binding of an ATP analog to the eIF-4A:RNA complex. Similarly, the observed nucleic acid-dependent differences in eIF-4A affinity for either $ADP.Mg^{+2}$ or $ATP.Mg^{+2}$ were also thought to reflect structural changes in eIF-4A.

DEAH DOMAINS

The DEAH domains are involved in RNA processing and display many features common among helicases. Originally classified as DEAH domains, new sequence information has expanded the classification to DEAD/H domains. DEAD/H helicases are associated with the large **ribonucleoprotein** (RNP) complex called the spliceosome. The spliceosome processes pre-messenger RNA to remove intervening untranslated sequences called **introns**, so the resultant messenger RNA contains only expressed sequences, or exons.

A number of precursor RNA processing (PRP) proteins associate with the spliceosome. The proteins contain either DEAD or DEAH domains and are also called splicing factors. Similar to many helicases, the isolated DEAH domain protein, PRP22, exhibits ATPase activity that is stimulated by the addition of poly(U) RNA (15). Mutagenesis studies of PRP22 variants demonstrated that the ATPase activity requires the conserved lysine residue of motif I. Although helicase activity has been demonstrated for the DEAH splicing factor PRP16 *in vitro*, the functional significance of these data is unclear (6). Most evidence suggests that the PRP splicing factors interact directly with the RNA in the spliceosome, to modulate RNA structure for efficient processing (15–17).

STRUCTURES OF DEXX AND DEXH DOMAINS

X-ray crystallographic structures of several helicases have been reported. The structures include two DExx DNA helicases belonging to SFI, *Bacillus stearothermophilus* PcrA helicase (18), and *E. coli* Rep helicase (19), and the HCV RNA DExH domain of SFII (20,21).

The helicases are multidomain structures (Fig. 3). The DNA helicases fold into structurally similar halves, with two domains each. Each half is composed of a $\beta\alpha\beta$ domain and an **alpha-helical** domain. A deep groove separates the two major helicase domains. The HCV RNA helicase domain contains three folded units: two $\beta\alpha\beta$ domains and one α-helical domain. Comparison of the helicase structures revealed that the $\beta\alpha\beta$ domains are structurally similar in both overall fold and their relative spatial orientation (22).

Both helicase motifs I and II are located within one of the $\beta\alpha\beta$ domains. The overall fold of the $\beta\alpha\beta$ domain and spatial arrangement of residues in motifs I and II are similar

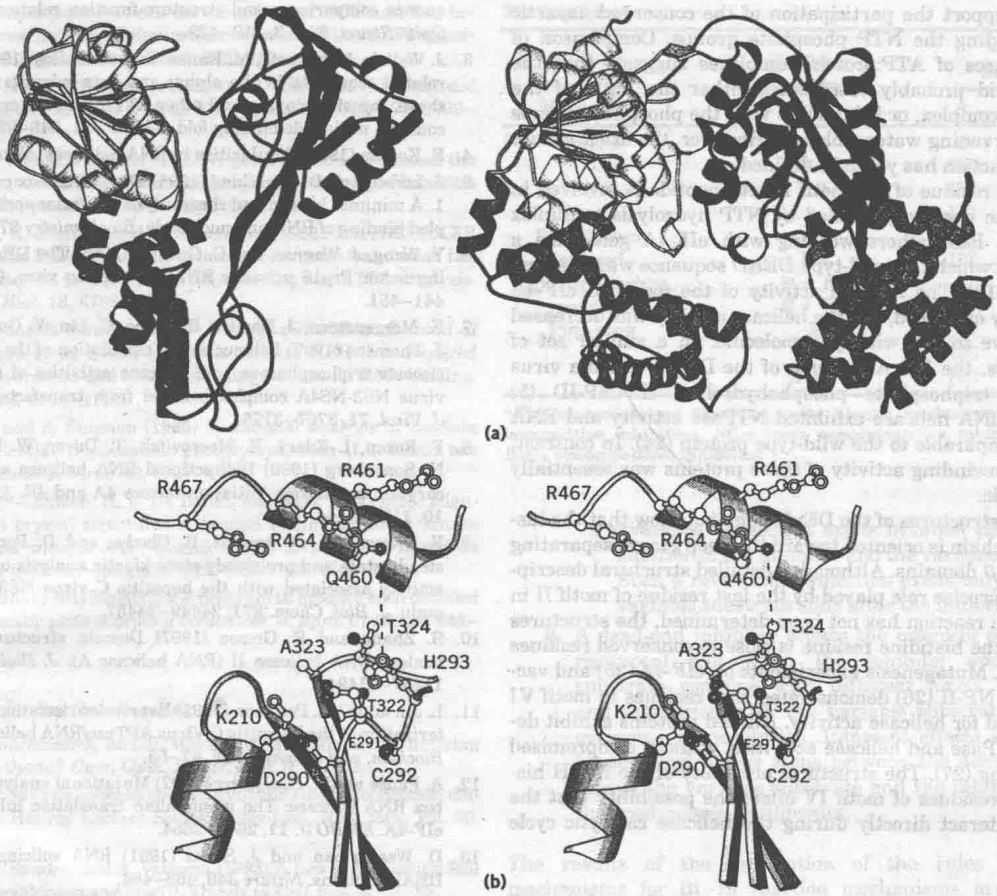


Figure 3. Structural features of DExx and DExH domains. The ribbons trace the polypeptide backbones of the HCV RNA helicase (left drawing, part **a**) and *E. coli* Rep DNA helicase (right drawing, part **a**). The upper left domain in each structure is the NTP-binding domain (shown in the lightest ribbon). The NTP-binding domain and second domain (shown in the darkest ribbon) share a common $\beta\alpha\beta$ fold. Many of the conserved sequence motifs are located near the deep groove that separates the $\beta\alpha\beta$ domains. A detailed view of motif II, III, and VI residues of the HCV RNA helicase is shown in stereoscopic representation (part **b**). Figures prepared using Molscript (29).

to conserved features present in other NTP-utilizing enzymes. The crystallographic structure of the PcrA:ADP complex also showed nucleotide binding to this domain (18). The structural studies taken together, including comparison of the superfamily I and II proteins, have identified the NTP-binding domain.

The structural studies also revealed conformational flexibility that is potentially important in helicase function. Comparison of the independent molecules in the HCV RNA helicase crystals revealed that one of the $\beta\alpha\beta$ domains could rotate relative to the remainder of the molecule (21). When the helicase structures are compared to those of NTP-binding proteins known to function as molecular switches, the observed structural changes in the helicase were found to involve residues at the end of a **beta-strand** previously shown to constitute part of the energy transducing mechanism (18,21). These residues in the DExH domain map to motif III, which mutagenesis studies have shown is required for coupling NTP hydrolysis to unwinding (12,23). Some indication of the extent of possible structural changes is found in the structure of the Rep DNA helicase complexed with a short oligonucleotide, where two orientations differing by a 130-deg rotation about a hinge region of the protein were observed for one of the helical subdomains (19).

ROLES OF DEAD AND DEAH RESIDUES

The mechanistic roles of the first and last residues of the DEAD and DEAH have been largely deciphered. Structural

studies on the free and ADP-bound forms of DExx and DExH domains support the participation of the conserved aspartic acid in binding the NTP phosphate groups. Comparison of the structures of ATP:protein complexes suggests that the aspartic acid probably coordinates either the Mg^{+2} of the NTP.Mg^{+2} complex, or it interacts with the phosphate groups via an intervening water molecule. However, the exact nature of the interaction has yet to be defined.

The last residue of the motif II tetrapeptide is involved in coupling the energy generated by NTP hydrolysis to duplex unwinding. Researchers working with eIF-4A generated a molecule in which the wild-type DEAD sequence was changed to DEAH (12). The ATPase activity of the modified eIF-4A was slightly enhanced, but the helicase activity was decreased 90% relative to the wild-type molecule. In a similar set of experiments, the DExA mutants of the DExH vaccinia virus nucleoside triphosphate phosphohydrolase II (NP-II) (5) and HCV RNA helicase exhibited NTPase activity and RNA binding comparable to the wild-type protein (24). In contrast, the RNA unwinding activity of these proteins was essentially undetectable.

Crystal structures of the DExH domains show that the histidine side chain is oriented toward the deep groove separating the two $\beta\alpha\beta$ domains. Although a detailed structural description of the precise role played by the last residue of motif II in the helicase reaction has not been determined, the structures show that the histidine residue is close to conserved residues in motif VI. Mutagenesis experiments on eIF-4A (25) and vaccinia virus NP-II (26) demonstrated that residues in motif VI are required for helicase activity. Altered proteins exhibit decreased ATPase and helicase activities, without compromised RNA binding (27). The structural proximity of the DExH histidine and residues of motif IV offers the possibility that the residues interact directly during the helicase catalytic cycle (Fig. 3).

CONCLUSION

The well-characterized DEAD and DEAH domains play essential roles in nucleic acid metabolism. Structural and mutagenesis studies, and amino acid sequence comparisons, have revealed the functions of the first D residue and the terminal D and H residues, although the mechanistic details of their interactions have not been determined. The DEAD and DEAH sequences are required for nucleoside triphosphate hydrolysis and, along with residues of motifs Ia, III, and VI, are required for coupling hydrolysis to mechanical action.

DEAD and DEAH domains occur widely in nature and exhibit a variety of biological functions. The functional diversity of these domains constitutes a challenge for classification by sequence-based approaches. Overall, the DEAD, DEAH, and DExH domains appear to be required in processes involving large-scale conformational changes of domain and substrate. Thus, their utilization as molecular motors is emerging as a unifying theme (14).

BIBLIOGRAPHY

1. S. Schmid and P. Linder (1992) D-E-A-D protein family of putative RNA helicases. *Mol. Microbiol.* **6**, 283–291.
2. A. Gorbalenya and E. Koonin (1993) Helicase: Amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429.
3. J. Walker, M. Saraste, M. Runswick, and N. Gay (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.
4. E. Koonin (1991) Similarities in RNA helicases. *Nature* **352**, 290.
5. J. Lorsch and D. Herschlag (1998) The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. *Biochemistry* **37**, 2180–2193.
6. Y. Wang, J. Wagner, and C. Guthrie (1998) The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. *Curr. Biol.* **8**, 441–451.
7. K. Morgenstern, J. Landro, K. Hsiao, C. Lin, Y. Gu, M. Su, and J. Thomson (1997) Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3-NS4A complex isolated from transfected COS cells. *J. Virol.* **71**, 3767–3775.
8. F. Rozen, I. Edery, K. Meerovitch, T. Dever, W. Merrick, and N. Sonenberg (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. *Mol. Cell Biol.* **10**, 1134–1144.
9. F. Preugschat, D. Averett, B. Clarke, and D. Porter (1996) A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain. *J. Biol. Chem.* **271**, 24449–24457.
10. S. Zhang and F. Grosse (1997) Domain structure of human nuclear DNA helicase II (RNA helicase A). *J. Biol. Chem.* **272**, 11487–11494.
11. L. Jin and D. L. Peterson (1995) Expression, isolation, and characterization of the hepatitis C Virus APTase/RNA helicase. *Archives Biochem. and Biophys.* **323**, 47–53.
12. A. Pause and N. Sonenberg (1992) Mutational analysis of a DEAD box RNA helicase: The mammalian translation initiation factor eIF-4A. *EMBO J.* **11**, 2643–2654.
13. D. Wassarman and J. Steitz (1991) RNA splicing. Alive with DEAD proteins. *Nature* **349**, 463–464.
14. J. Lorsch and D. Herschlag (1998) The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. *Biochemistry* **37**, 2194–2206.
15. J. Wagner, E. Jankowsky, M. Company, A. Pyle, and J. Abelson (1998) The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *EMBO J.* **17**, 2926–2937.
16. S. Teigelkamp, M. McGarvey, M. Plumpton, and J. Beggs (1994) The splicing factor PRP2, a putative RNA helicase, interacts directly with pre-mRNA. *EMBO J.* **13**, 888–897.)
17. S. Kim and R. Lin (1996) Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell Biol.* **16**, 6810–6819.
18. H. Subramanya, L. Bird, J. Brannigan, and D. Wigley (1996) Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379–383.
19. S. Korolev, J. Hsieh, G. Gauss, T. Lohman, and G. Waksman (1997) Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635–647.
20. H. Cho, H., N. Ha, L. Kang, K. Chung, S. Back, S. Jang, and B. Oh (1998) Crystal structure of RNA helicase from genotype 1b hepatitis C virus. A feasible mechanism of unwinding duplex RNA. *J. Biol. Chem.* **273**, 15045–15052.
21. N. Yao, T. Hesson, M. Cable, Z. Hong, A. Kwong, H. V. Le, and P. C. Weber (1997) Structure of the hepatitis C virus RNA helicase domain. *Nature Struct. Biol.* **4**, 463–467.

22. S. Korolev, N. Yao, T. Lohman, P. C. Weber, and G. Waksman (1998) Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 superfamilies of helicases. *Prot. Sci.* **7**, 605–610.

23. D. Kim, J. Kim, Y. Gwack, J. Han, and J. Choe (1997) Mutational analysis of the hepatitis C virus RNA helicase. *J. Virol.* **71**, 9400–9409.

24. G. Heilek and M. Peterson (1997) A point mutation abolishes the helicase but not the nucleoside triphosphatase activity of hepatitis C virus NS3 protein. *J. Virol.* **71**, 6264–6266.

25. A. Pause, N. M'ethot, and N. Sonenberg (1993) The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol. Cell Biol.* **13**, 6789–6798.

26. C. Gross and S. Shuman (1996) The QRxGRxGRxxxG motif of the vaccinia virus DExH box RNA helicase NPH-II is required for ATP hydrolysis and RNA unwinding but not for RNA binding. *J. Virol.* **70**, 1706–1713.

27. C. Gross and S. Shuman (1995) Mutational analysis of vaccinia virus nucleoside triphosphate phosphohydrolase II, a DExH box RNA helicase. *J. Virol.* **69**, 4727–4736.

28. U. Schulze-Gahmen, H. L. De Bondt, and S.-H. Kim (1996) High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: Bound waters and natural ligand as guides for inhibitor design. *J. Med. Chem.* **39**, 4540–4546.

29. P. J. Kraulis (1991) Molscript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.

Suggestions for Further Reading

L. Bird, H. Subramanya, and D. Wigley (1998) Helicases: A unifying structural theme? *Curr. Opin. Struct. Biol.* **8**, 14–18.

C. Guthrie (1996) The spliceosome is a dynamic ribonucleoprotein machine. The Harvey Lecture Series, Wiley-Liss, New York, Vol. 90, pp. 59–80.

K. Holmes, C. Sander, and A. Valencia (1993) A new ATP-binding fold in actin, hexokinase and Hsc70. *Trends in Cell Biol.* **3**, 53–59.

G. Kadaré and A. Haenni (1997) Virus-encoded RNA helicases. *J. Virol.* **71**, 2583–2590.

T. Lohman (1992) Escherichia coli DNA helicases: mechanisms of DNA unwinding. *Mol. Microbiol.* **6**, 5–14.

DEAD-END INHIBITION

JOHN F. MORRISON

Dead-end inhibition of an **enzyme** occurs when a compound, usually a structural analogue of a substrate, combines at the **active site** of the enzyme, but does not undergo reaction. The resulting complex can only dissociate back to the components from which it was formed. Dead-end inhibitors are useful for determining kinetic reaction mechanisms because dead-end inhibition patterns can be characteristic of a particular mechanism. Rules have been formulated for the qualitative prediction of such inhibition patterns (1) and these rules can be summarized as follows:

1. A dead-end inhibitor affects the *slope* of a double-reciprocal plot (see **Lineweaver–Burk plot**) when the inhibitor and the variable substrate combine with the same enzyme form or a reversible connection exists between the points of addition of the inhibitor and the variable substrate. Such a reversible connection exists only when the variable substrate adds after the inhibitor.
2. A dead-end inhibitor affects the *intercept* of a double-reciprocal plot when the inhibitor and variable substrate combine with different forms of enzyme and saturation with the variable substrate does not overcome the inhibition. Intercept effects will always be observed unless steps between the points of addition of the variable substrate and the inhibitor are at thermodynamic equilibrium.

Table 1. Dead-end Inhibition Patterns for Bi–Bi Reaction Mechanisms[a]

Mechanism	Inhibitor Combines with	Variable Substrate A	Variable Substrate B
Ordered	E	C	NC
	EA	UC	C
	EQ	UC	UC
	EA + EQ	UC	NC
Equilibrium-ordered	E	C	C
	EA	UC	C
Rapid equilibrium, random	E + EB	C	NC
	E + EA	NC	C
Ping-pong	E	C	UC
	F	UC	C
	E + F	NC	NC

[a] The mechanisms are illustrated under **Kinetic mechanisms**. C = linear **competitive inhibition**. NC = linear **noncompetitive inhibition**. UC = linear **uncompetitive inhibition**.

The results of the application of the rules to kinetic mechanisms for Bi–Bi reaction mechanisms are given in Table 1. Further examples are given in **Competitive inhibition**, **Noncompetitive inhibition**, **Product inhibition**, **Substrate inhibition**, and **Uncompetitive inhibition**.

BIBLIOGRAPHY

1. W. W. Cleland (1963) *Biochim. Biophys. Acta* **67**, 188–196.

DEAMIDATION

T. IMOTO

Deamidation is the reverse reaction of amidation (see **Carboxyl groups**) and involves release of an amine from an amide. Deamidation occurs spontaneously in proteins at **asparagine** and **glutamine** residues, which causes heterogeneity, instability, and sometimes defects in the function of the proteins.

Deamidation in proteins occurs enzymatically or chemically. **Enzymes** that participate in deamidating of proteins are transglutaminase and peptidoglutaminase. An amide bond in an asparagine or glutamine residue is relatively labile, especially in acidic media, and is easily hydrolyzed chemically. The rate of deamidation is sensitive to the environment of the

amide. Exposed amides are more readily deamidated than those buried in the interior of a protein. Thus, denatured proteins often have high potential for deamidation.

The Asn-Gly sequence is especially notorious for rapid deamidation. The reaction shown below proceeds readily in this sequence. Additional reactions such as **racemization**, transpeptidation, and cleavage of peptide bonds take place during this reaction. These reactions also occur with the Asp-Gly sequence and pivotal components of protein deterioration.

The rates of deamidation vary with different peptides. The **half-times** for deamidation of various asparagine residues under the same conditions range from 3.3 to 277 days. Deamidation causes conformational destabilization of proteins and accelerates their degradation. Thus the deamidation of Asn or Gln residues has been regarded as a timer of protein turnover (1).

Asparaginase hydrolyzes asparagine to aspartic acid (deamidation). Because asparagine in the blood is an essential nutrient for the growth of malignant white cells, the intravenous administration of asparaginase suppresses the growth of some types of animal and human leukemias.

BIBLIOGRAPHY

1. A.B. Robinson, J.H. McKerrow, and P. Cary (1970) *Proc. Natl. Acad. Sci. USA* **665**, 753–757.

Suggestion for Further Reading

H.T. Wright (1991) Nonenzymatic deamidation of asparagine and glutamine residues in proteins. *Crit. Rev. Biochem. Mol. Biol.* **26**, 1–52.

DECAPENTAPLEGIC

RAHUL WARRIOR
KAVITA ARORA

The *decapentaplegic* (*dpp*) gene plays a key role in patterning both embryonic and adult structures in the fruitfly *Drosophila*. The name of the **locus** underscores the multiple requirements for *dpp* activity during development and refers to the wide range of defects caused by **mutations** in the **gene** (*decapentaplegic* = 15 defects) (1). *dpp* is one of the few haplo-insufficient loci in *Drosophila*; consequently, even a 50% reduction in *dpp* activity leads to embryonic lethality. Less severe alleles of *dpp* are viable but result in adults that have defects in structures derived from one or more **imaginal discs**. The *dpp* gene encodes a **protein** that is **homologous** to secreted **growth factors** of the **transforming growth factor-β** (TGF-β) **superfamily**. Ligands that belong to this group of signaling molecules have been identified in animals across the phylogenetic spectrum and are involved in cellular and developmental processes ranging from regulation of the **cell cycle** and the **extracellular matrix**, to establishment of embryonic pattern and aging. Genetic and molecular studies of *dpp* and the genes involved in *dpp* **signal transduction** have proven to be of broad general interest, because the basic mechanisms involved in generating and receiving TGF-β signals are evolutionarily conserved.

MOLECULAR FEATURES

The *dpp* gene spans ~55 kb of DNA, the bulk of which consists of a large array of *cis*-regulatory elements (2). *dpp* **messenger RNA** is expressed in a complex spatial and temporal pattern that reflects the multiple roles the ligand plays during development. Three major ***cis*-acting** domains regulate transcription of the gene: an upstream *dpp*-shv region required for expression in the embryonic gut and pupal wing, a central *dpp*-Hin region that directs embryonic expression, and a downstream *dpp*-disk region that regulates transcription in imaginal discs (2). These dispersed regulatory elements contribute to the genetic complexity of the locus and mediate *dpp* transcription in response to inputs from a variety of signaling pathways. For example, *dpp* expression in the blastoderm stage embryo is regulated by **Dorsal**, a **transcription factor** homologous to mammalian NFκB. Later in embryogenesis, *dpp* transcription in the dorsal-most cells of the germband is controlled by genes acting in the Jun kinase signaling pathway, while expression of *dpp* in the midgut is dependent on multiple factors, including **homeobox** DNA-binding proteins, the **growth factor** Wingless, and Dpp itself. During imaginal disc development, localized expression of *dpp* in the wing and leg discs is regulated by the ***hedgehog*** and ***wingless*** genes.

The *dpp* locus gives rise to three major and two minor transcripts through the use of alternative **promoters** that are utilized at different stages of development (2). While the mRNAs have unique 5′-untranslated sequences, they share common second and third exons that contain an open reading frame coding for a single 588-amino-acid-residue protein (see **Introns/exons** and **RNA splicing**). Dpp is most closely related to the vertebrate bone morphogenetic proteins BMP-2 and BMP-4, and it shares ~75% identity with these proteins in the carboxy-terminal region that constitutes the mature ligand **domain**. While the vertebrate BMPs were first identified by their ability to induce ectopic bone, they are now known to have important roles in embryonic development in a number of organisms, including frogs, mice, and humans. Dpp and BMP-4 can substitute for one another functionally, because a human BMP-4 transgene can rescue patterning defects in a *Drosophila* embryo lacking Dpp (3). Conversely, the fly protein can induce the formation of ectopic bone when injected subcutaneously into rats. Like other ligands belonging to the TGF-β superfamily, Dpp is processed and secreted as a **disulfide**-linked dimer of ~30 kDa.

DPP SIGNAL TRANSDUCTION

Dpp acts as a secreted ligand to influence the developmental fate of cells that receive the signal. Genes involved in Dpp signal transduction have been primarily identified using two strategies. Some genes have been recovered in genetic screens to isolate **enhancers** of weak *dpp* **alleles**, while others have been isolated in low stringency **hybridization** screens, based on their homology to TGF-β signaling components identified in other organisms.

According to the current paradigm for BMP signaling, the ligand binds a heteromeric complex of two structurally related transmembrane **serine-threonine kinases**, called the type I and type II receptors (4,5). Formation of the ligand–receptor complex allows the type II kinase to **phosphorylate** and activate the type I receptor. A type II receptor, Punt, as well as

two type I receptors, Thick veins (Tkv) and Saxophone (Sax), have been implicated in Dpp signaling. Although the role of Punt and Tkv as receptors for Dpp is well established, it appears that Sax may primarily mediate the response to other BMP-related ligands in Drosophila. Activation of Tkv results in the direct phosphorylation of a cytoplasmic protein encoded by *mothers against dpp* (*mad*). This modification triggers **Mad** to form a complex with Medea (a structurally related protein), and it enables their translocation from the cytoplasm into the **nucleus**. Mad and Medea contain DNA-binding domains and are thought to regulate the expression of downstream target genes in association with other transcription factors (4,6). The Mad family of proteins is evolutionarily conserved. The human homologue of Medea (DPC4) has been identified as a **tumor-suppressor** gene, a result that appears logical in light of the known antiproliferative effects of TGF-β.

BIOLOGICAL ROLE OF DPP

Dpp plays an important role in the specification of cell fate and morphogenesis in a number of tissues (1). Mutations in *dpp* that interfere with production of an active ligand affect all developmental events regulated by the gene, while mutations that disrupt specific enhancer elements affect *dpp* function in a tissue- or stage-specific manner. Among the processes that require *dpp* signaling are: oogenesis; establishment of dorsal–ventral pattern during embryogenesis; morphogenetic movements of dorsal closure; subdivision of the mesoderm along the dorsal–ventral axis; specification of the visceral mesoderm and endoderm in the embryonic gut; development of the heart, gastric cecae, salivary glands, and the trachea; and growth and patterning of imaginal discs.

The ability of Dpp to trigger distinct responses in a single field of cells has generated a great deal of interest in understanding the mechanisms underlying *dpp* function. Dpp has been shown to specify cell fate in a concentration-dependent manner; that is cells respond to different thresholds of Dpp by following distinct pathways of differentiation. A critical issue that arises is how gradients of *dpp* activity are established, and how such gradients are interpreted. Based on recent studies, two alternative mechanisms have evolved: one is based on **diffusion** of Dpp from its site of synthesis to generate a protein concentration gradient, and the second involves an inhibitor that diffuses into the domain where *dpp* is expressed and interferes with Dpp signaling in a graded manner. This results in a gradient of ligand activity, rather than concentration. Both types of gradients are discussed further below.

Adult structures in the fruitfly arise from imaginal discs, small groups of epithelial cells that are set aside during embryogenesis. These discs grow and are patterned during the larval and pupal stages in response to different signals. In the wing imaginal disc, Dpp acts as a long-range **morphogen** to specify cell fate along the anterior–posterior axis. *dpp* is expressed in a narrow domain of cells at the anterior–posterior compartment boundary, from where it diffuses to generate a gradient of protein in the surrounding tissue. Cells up to 20 cell diameters away respond to different threshold concentrations of Dpp by activating the transcription of target genes like *spalt* and *optomotor blind* (7,8). Expression of *spalt* occurs close to the source of *dpp* protein, while *optomotor blind* expression overlaps with, and extends further than, *spalt* expression. The nested domains of Dpp target gene expression further subdivide the wing disc into distinct regions that are specified by the combination of genes expressed.

During early development, a gradient of Dpp signaling is required to establish cell fates within the dorsal half of the embryo. Peak levels of Dpp signaling specify the dorsal-most amnioserosa tissue, while lower levels are required to specify the dorsal ectoderm (9,10). Reduction in the level of Dpp signaling results in progressive loss of dorsal structures, while increasing concentrations of Dpp can induce dorsal cell fates. Since *dpp* mRNA is expressed uniformly in all dorsal cells, it is generally believed that Dpp activity, rather than its concentration, is graded. A number of extracellular proteins are involved in generating a gradient of Dpp signaling in the embryo. A second BMP ligand, Screw, acts synergistically with Dpp to enhance signaling in the dorsal-most cells. The activity of these ligands is antagonized by a secreted factor, Short gastrulation (Sog), that can prevent ligand binding to the receptor. The inhibitor Sog is expressed in ventral cells and diffuses dorsally to generate a ligand gradient of the opposite polarity. In addition, a **metalloproteinase**, Tolloid (Tld), promotes signaling in dorsal cells by cleaving Sog and releasing the ligand. Thus modulation of ligand activity at multiple levels contributes to establishment of a gradient of BMP signaling in the embryo (11–14).

Similar antagonistic interactions involving homologous proteins are involved in patterning the dorsal–ventral axis in vertebrate embryos. In *Xenopus*, BMP-4 promotes ventral development, while a homologue of Sog (Chordin) promotes dorsal cell fates. Recently an amphibian homologue of Tld (Xolloid) has been shown to cleave Chordin (12). These and other studies suggest that the dorsal–ventral axes in *Drosophila* and vertebrates are specified by a similar mechanism, although they are inverted relative to one another. The extensive parallels between TGF-β/BMP signaling in vertebrates and invertebrates allow one to extend the insights gained from studying Dpp signaling in *Drosophila* to other organisms, including humans.

BIBLIOGRAPHY

1. F. A. Spencer, F. M. Hoffman, and W. M. Gelbart (1982) Decapentaplegic: a gene complex affecting morphogenesis in *Drosophila melanogaster*. *Cell* **28**, 451–461.
2. R. D. St. Johnston, F. M. Hoffman, R. K. Blackman, D. Segal, R. Grimaila, R. W. Padgett, H. A. Irick, and W. M. Gelbart (1990) The molecular organization of the *decapentaplegic* gene in *Drosophila melanogaster*. *Genes Dev.* **4**, 1114–1127.
3. R. W. Padgett, J. M. Wozney, and W. M. Gelbart (1993) Human BMP sequences can confer normal dorsal–ventral patterning in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* **90**, 2905–2909.
4. C.-H. Heldin, K. Miyazono, and P. ten Dijke (1997) TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471.
5. J. Massague (1996) TGF-β signaling: receptors, transducers, and Mad proteins. *Cell* **85**, 947–950.
6. K. Arora, H. Dai, S. G. Kazuko, J. Jamal, M. B. O'Connor, A. Letsou, and R. Warrior (1995) The *Drosophila schnurri* gene acts in the Dpp/TGF-β signaling pathway and encodes a transcription factor homologous to the human MBP family. *Cell* **81**, 781–790.
7. T. Lecuit, W. J. Brook, M. Ng, M. Calleja, H. Sun, and S. M. Cohen (1996) Two distinct mechanisms for long-range patterning