The G_{M2} gangliosidoses databases: Allelic variation at the *HEXA*, *HEXB*, and *GM2A* gene loci

Paulo Cordeiro, MSc¹, Peter Hechtman, PhD¹⁻³, and Feige Kaplan, PhD^{2,3}

The G_{M2} gangliosidoses are a group of recessive disorders characterized by accumulation of G_{M2} ganglioside in neuronal cells. The genes responsible for these disorders are *HEXA* (Tay-Sachs disease and variants), *HEXB* (Sandhoff disease and variants), and *GM2A* (AB variant of G_{M2} gangliosidosis). We report the establishment of three relational locus-specific databases recording allelic variation at the *HEXA*, *HEXB*, and *GM2A* genes and accessed at the G_{M2} gangliosidoses home page (<http://data.mch.mcgill.ca/gm2-gangliosidoses>). Submission forms are available for the addition of new mutations to the databases. The databases are available online for users to search and retrieve information about specific alleles by a number of fields describing mutations, phenotypes, or author(s). **Genetics in Medicine, 2000;2(6):319–327.**

Key Words: G_{M2} gangliosidoses, locus-specific databases, hexosaminidase

The genetic basis of the G_{M2} gangliosidoses is complex. Three lysosomal polypeptides, encoded by three distinct genes, are essential for normal catabolism of the G_{M2} ganglioside substrate in vivo: the α and β subunits of β -hexosaminidase A (Hex A) and the G_{M2} activator protein. Tay-Sachs disease (OMIM: 278200) and its variants, caused by deficiency of Hex A activity, results from mutations at the *HEXA* gene (15q23-q24), which encodes the α subunit of Hex A. Sandhoff disease (OMIM: 268800) and variants, characterized by combined deficiency of Hex A and Hex B activity, is caused by mutation in the *HEXB* gene (5q13) encoding the β subunit common to Hex A and Hex B. The AB variant of G_{M2} gangliosidosis (OMIM: 272750) is caused by mutations in the *GM2A* gene (5q31.3-q33.1) encoding the G_{M2} activator protein. In AB variant patients, Hex A and Hex B are structurally normal, and enzyme activities are in the normal range when measured in the presence of synthetic, hydrophilic substrates. However, hydrolysis of G_{M2} ganglioside is impaired by absent or defective formation of the G_{M2} ganglioside/G_{M2} activator/Hex A complex. All known variants of G_{M2} gangliosidosis display autosomal recessive inheritance.

The clinical phenotypes associated with G_{M2} gangliosidoses variants vary widely. They range from infantile-onset, rapidly progressive neurodegenerative disease culminating in death before age 4 years (classical Tay-Sachs and Sandhoff diseases and G_{M2} activator deficiency) to late-onset, slowly progressive neurologic conditions compatible with long-term survival and with little or no effect on intellect.¹⁻⁵ Phenotypic variants of

G_{M2} gangliosidosis have been classified according to age of the onset of symptoms into infantile, juvenile, and adult forms of the disease. However, these classifications are often arbitrary.

Recent advances in computer technology have created a new area of research and development “bioinformatics,” which integrates biology and informatics. This area of study has provided invaluable information on newly discovered genes: localization of genes influencing complex phenotypes, molecular variation (mutations) influencing human diseases, and identification of functional DNA variants. These technological advances are likely to have a profound impact on current knowledge of complex disease etiology and may reveal novel approaches to disease treatment and prevention. Locus-specific mutation databases record genetic variation and are relevant to human genetic research, genetic epidemiology, and health care.

The exponential growth of information regarding allelic variation accounting for cause of and/or susceptibility to disease mandates the creation of a formal record of “mutation” at gene loci. The current pace of mutation discovery far exceeds the capacity to publish the data by conventional means. Databases provide the most efficient format to maintain an up-to-date record of locus-specific allelic variation. A designated curator or curator consortium evaluates information. Curators receive and distribute information upon request.

The Human Genome Organization (HUGO) provides guidelines and recommendations for content and structure of mutation databases.⁶ These will provide the means to link locus-specific databases into a centralized human mutation database and to ensure cohesiveness in a community of users.

According to HUGO guidelines, mutation databases should be relational and have a minimum core of content in a shared syntax.^{7,8} The core should include (1) a unique identifier for each allele, (2) a source/report for data, (3) a context for alleles (patient genotype or in vitro expression system used in stud-

From the Departments of ¹Biology, ²Pediatrics, and ³Human Genetics, McGill University, Montreal, Quebec, Canada.

Feige Kaplan, PhD, McGill University, Montreal Children's Hospital Research Institute, 4060 St. Catherine Street West, Room 236, Montreal, Quebec, Canada H3Z 2Z3.

Received: August 29, 2000.

Accepted: September 28, 2000.

ies), and (4) a description of alleles. The allele description should be verified by a group or consortium. Relational databases enable users to perform queries on any aspect of the database content.

We report herein the establishment of three relational databases for the *HEXA*, *HEXB*, and *GM2A* genes accessed at the *G_{M2}* gangliosidoses home page (<http://data.mch.mcgill.ca/gm2-gangliosidoses/>). **HEXAdb** (<http://data.mch.mcgill.ca/hexadb/>), **HEXBdb** (<http://data.mch.mcgill.ca/hexbdb/>), and **GM2Adb** (<http://data.mch.mcgill.ca/gm2a/>) are relational databases and are currently divided into four tables each: Mutation, Phenotype, Reference, and Association (Tables 1 and 2). The Mutation table contains descriptions of the following fields: mutation name, alternative names, mutation type, nucleotide changes (location and length), as well as a description of the mRNA and the protein size and abundance. The Phenotype table contains details on clinical and biochemical phenotype, age of onset, diagnosis, genotype, and ethnicity of patients. The Reference table has information on types of publication, authors, journals, titles, etc. It also includes PubMed IDs to enable the user to look up references in MEDLINE. Finally, the Association (also called Resource) table links, i.e. relates, the tables together and provides information on the submitter (the individual who identified the mutation). The purpose of these databases is to collect and distribute information on mutations in the genes responsible for *G_{M2}* gangliosidosis. **HEXAdb** currently records 100 alleles. Among 12 alleles for which no associated clinical phenotype has been described, 6 are predicted to be disease-causing. Such predictions are possible because information in the Phenotype table includes any in vitro gene expression studies that have been performed on the mutant allele. **HEXBdb** currently records 25 alleles; **GM2Adb**, 5 alleles.

Table 1.
Mutation search page

Field ^a	Description
Mutation name	The official name for the mutation given current nomenclature rules
Alternative name	A common name, usually the codon change, used to refer to the mutation
Nucleotide change	Substitution, insertion, deletion
Mutation type	Missense, nonsense, abnormal splicing, silent (polymorphism), frameshift
Nucleotide number	cDNA position where mutation occurs
mRNA description	Information regarding the mRNA transcribed, if any, as a result of the mutation
Protein description	Information regarding the protein translated, if any, as a result of the mutation
Region	Exon (e.g. EO1), intron (e.g. IVS-2)

Also includes submitter (from Association table) and short citations (from reference table).

^aStructure of the **HEXAdb**, **HEXBdb**, and **GM2Adb** tables showing fields that can be used when interrogating using the Mutation Page search engine.

Table 2
Phenotype search page

Field ^a	Description
Clinical phenotype	Acute, subacute, chronic, benign, asymptomatic (none), unknown
Age of onset	Infantile, late infantile, juvenile, late juvenile, and adult
Biochemical phenotype	Biochemical variants (B1, hexosaminidase Paris, neutral polymorphisms); information regarding enzyme activity, processing, and submit association
Genotype	Information regarding the second allele, in patients where identified
Diagnostic	Method used to identify the presence of a mutation (gain or loss of restriction enzyme sites, heteroduplex analysis, allele-specific amplification or allele-specific hybridization)
Ethnic origin	The ethnicity of patients associated with a certain mutation

Also includes mutation name and nucleotide number (from Mutation table), submitter (from Association table) and short citations (from Reference table).
^aStructure of **HEXAdb**, **HEXBdb**, and **GM2Adb** tables showing fields that can be used when interrogating using the Phenotype Page search engine.

METHODS

Databases: software and search engine

HEXAdb, **HEXBdb**, and **GM2Adb** were constructed using MySQL v3.21.33c® (T.c.X. DataKonsultAB) on the Red Hat Linux v5.1® (Red Hat, Inc.) operating system. MySQL is a true multiuser, multithreaded SQL database server that enables one to access, store, and update information. The Web connection exists through a Perl DBI interface. Modifications and additions to the databases are immediately undertaken by the curator using Microsoft Access 97® (Microsoft Corp.) through an ODBC (Open Database Connectivity) interface or directly using MySQL command scripts. The Perl scripts that make up the search engine were written in Wordpad 4.0® (Microsoft Corp.) and then uploaded to the Debelle server (<http://data.mch.mcgill.ca>) using WS_FTP95 LE 4.6© (Ipswitch, Inc., 1992–1998). WS_FTP LE is a File Transfer Protocol (FTP) client application that can connect to any system that has a valid Internet address and contains an FTP server program, allowing one to transfer files between a wide variety of systems, including Windows, OS/2, and UNIX systems.

The Web pages were written in HTML 4.0 (Hyper Text Markup Language) programming code, by using a normal text editor, Wordpad 4.0® (Microsoft Corp.). HTML files were then uploaded to the Debelle server (<http://data.mch.mcgill.ca>) using WS_FTP95 LE 4.6© (Ipswitch, Inc., 1992–1998).

Logos and graphics for the Web pages were designed in Microsoft PowerPoint 97® (Microsoft Corp.) and then edited in Microsoft Photo Editor 3.0® (Microsoft Corp.).

RESULTS

Mutation table

The Mutation table (see Table 1) provides specific molecular information on mutations identified. Each mutation has an ID

(identification number, the primary key), which is assigned by its order of entry, a systematic name (usually referring to the cDNA mutation), an alternative name (the corresponding codon change), as well as information on the nucleotide change (substitution, insertion, deletion), the mutation type (nonsense, missense, silent, abnormal splicing, frameshift), the nucleotide position, the region of the gene involved (exon, intron, UTR) and a description of mRNA and protein phenotypes. Also provided are information on the submitter, which refers to the investigator(s) who identified the mutation either in a peer reviewed publication or by direct submission. Citations of published articles are indicated.

The search engine also allows one to query the database in any of the listed fields. For example, Figure 1A illustrates the mutation data retrieval search page for *HEXBdb*. Figure 1B shows the output of a query on the "Mutation type" field for the value "frameshift" in the *HEXB* gene. The user can also select a table that displays mutation information for all the records in the database. The *HEXA* database records 100 mutations (51 missense, 19 abnormal splicing, 12 frameshifts, 8 nonsense, 6 deletions, and 4 silent mutations). *HEXBdb* records 25 mutations (10 missense, 6 frameshifts, 6 abnormal splicing, 1 nonsense, 1 deletion, and 1 silent) and *GM2Adb* records 5 mutations (2 missense, 2 frameshifts, and 1 nonsense).

Phenotype table

The Phenotype table (see Table 2) is designed to provide details on clinical and biochemical phenotype, age of onset, and genotype and ethnicity of patients harboring specific mutations. Each record in the Phenotype table has an associated mutation ID number. Also included in this table (where available) is a molecular diagnostic method used to detect the mutation. Abnormalities of enzyme properties or specific activity are also recorded.

The Phenotype table can be searched on any of these fields, and a complete table of phenotypes is available online. Figure 2A and B illustrate data retrieval following a search under "Clinical phenotype" for mutations leading to subacute *HEXB* (Sandhoff) variants. To date, clinical phenotypes are known for 88 of the 100 entries in *HEXAdb*. There are 56 acute, 15 subacute, and 7 chronic TSD variants. Three Hex A pseudo-deficient alleles and seven benign alleles are also recorded. Twenty-four of 25 entries in *HEXBdb* are associated with known clinical phenotypes (15 acute, 4 subacute, 3 chronic, 1 asymptomatic, 1 benign). All five of the *GM2Adb* entries are associated with an acute phenotype.

Reference table

The Reference table provides a listing of articles reporting identification or analysis of specific mutations. In the case of published articles, the full reference to the article is specified as well as the PubMed ID, enabling the user to link to references containing the abstract in MEDLINE. The Reference table offers a search tool by author(s), journal, title, and year of publication.

A table with all references on record is available on the Web page. Currently 132 references are listed for *HEXAdb*, 32 for *HEXBdb*, and 7 for *GM2Adb*.

Association table

Finally, the Association (also called Resource) table links (i.e., relates) the main tables together and records the "submitter" who identified the mutation(s) either by publication or by personal submission.

Submission form

All three databases permit online submission of new mutations (see Fig. 3). The submitters are requested to provide data for all fields. The aim is to obtain as complete a record as possible, including clinical and biochemical data for each new submission. The submission form, which will permit a faster and more efficient means to regroup information, is directed to the curator, whose role it is to review the information and determine whether it is sufficient for inclusion.

The clinical descriptions requested in the submission form are limited to two categories, which lend themselves to standardization: age of onset and an index of severity with only six gradations. The criteria for application of these terms are found on the information page "Tay Sachs Disease, Gene Clinics, NIH." A direct link to this page is found at the top of the Phenotype Information section of the submission form. This form also contains a direct link to the email address of the curators, permitting the submitter to request assistance if necessary.

Additional clinical terms such as *ataxic*, *dystonic*, *psychotic*, *muscle-wasting*, etc., frequently appear in the clinical literature describing the presentation of nonclassical variants of these disorders. Submitters may wish to include this information in the "Comments" field. Terms used in this field are also available for search.

Other features

Relevant information pages, provided at the database Web site, include clinical information on Tay-Sachs and Sandhoff disease and links to other sources of information on the Web including Online Mendelian Inheritance in Man (<http://www.ncbi.nlm.nih.gov/Omim/>), LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/>), GeneCards (<http://bioinformatics.weizmann.ac.il/cards/>), and GeneClinics (<http://www.geneclinics.org/>).

Modeling of *HEXA* mutant alleles: genotype/phenotype correlations

To provide information concerning the effects of mutations on the manifestations of disease, a feature for viewing mapped positions of selected *HEXA* mutations onto the homology-modeled hexosaminidase structure was developed. The structural model of the Hex A α subunit was restricted to the catalytic domain (region of the protein containing a $(\beta\alpha)_8$ barrel motif, residues 160–487) of the wild-type protein. This is the region that has been modeled onto the template of the crystallized *Serratia marcescens* chitobiase.⁹ Theoretical models of *HEXA* missense mutations are being added to the Web page.

A.

HEXBdb Mutation Data Retrieval

Please choose search field: frameshift

Enter text or a value: frameshift

Alternative Name
 Nucleotide Change
 Mutation Type
 mRNA Description
 Protein Description
 Nucleotide Number
 Region Affected
 Submitter
 Reference

HELP

Mutation name: For a nucleotide replacement, eg. c.379.C->T, the notation consists of c (indicates cDNA) followed by a nucleotide number (preceded and followed by a period), then the original nucleotide and finally the mutant nucleotide.
 For a deletion or insertion, eg. c.547.insA, the notation consists of c followed by a nucleotide number (preceded and followed by a period), then ins or del and finally the nucleotide(s) that are inserted or deleted.

Alternate name: For an amino acid change, eg. G269S (Gly269Ser), the notation consists of the amino acid followed by the number and then the mutant amino acid. One or three letter abbreviations for the amino acid can be used.
 For an intronic mutation, eg. IVS2+1g->a, the notation consists of IVS followed by intron number, then intronic position (can be +/-), the original nucleotide and finally the mutant nucleotide.

Nucleotide change: Substitution, insertion and deletion.

Mutation type: Missense, nonsense, abnormal splicing, silent (polymorphism) and frameshifts.

mRNA description: Information regarding the mRNA synthesized, if any, as a result of the mutation.

Protein description: Information regarding the protein synthesized, if any, as a result of the mutation.

Nucleotide number: Position where mutation occurs.

Region affected: Exon or intron, eg. E01 or IVS-04, where mutation occurs.

Submitter: Consists of the first author's name and date of publication, if published.

References: Consists of the short citation (first author's name and date of publication) for articles containing information on this mutation.

B.

MUTATION LISTING												
THE DATE AND TIME IS: Mon Sep 25 21:26:48 2000												
FIELD next_type												
VALUE frameshift												
UID	Mutation Name	Alternative Name	Nucleotide Change(s)	Mutation Type	mRNA (Description)	Protein (Description)	Nucleotide Number	Length Affected	Intronic Position	Region Affected	Submitter	References
2	c.76.delA	del A76	Deletion	Frameshift	Reduced mRNA of normal size	-	76	1	0	E01	Banerjee et al.	Banerjee et al. (1994) Hern et al. (1994) Redonnet-Vernhet et al. (1996)
7	c.772.delG	del G772	Deletion	Frameshift	Absent mRNA by Northern blotting (total Hex B mRNA 1.6% of normal, with second allele 1985delAG)	-	772	1	0	E07	O'Dowd et al.	McInnes et al. (1992) O'Dowd et al. (1996) Zhang et al. (1994)
8	c.782.delCTTT	del CTTT782	Deletion	Frameshift	Absent mRNA	-	782	4	0	E07	Kleinman et al.	Kleinman et al. (1994)
15	c.1385.delAG	del AG1385	Deletion	Frameshift	Absent mRNA by Northern blotting (total Hex B mRNA 1.6%, with second allele also frameshift)	-	1385	2	0	E11	O'Dowd et al.	McInnes et al. (1992) O'Dowd et al. (1996) Zhang et al. (1994)
16	c.1344.delTT	del TT1344	Deletion	Frameshift	mRNA detectable by Northern blotting; total mRNA is 18% (36% of single gene dose as second allele null)	-	1344	1	0	E11	Zhang et al.	Zhang et al. (1994)
25	c.1751.delTG	del TG1751	Deletion	Frameshift	Reduction in level of mRNA containing the mutation	-	1751	2	0	E14	Kleinman et al.	Kleinman et al. (1996)
There were 6 repeat(s) found.												

Fig. 1 A: Example of a mutation search performed on HEXBdb. The mutation data table has been selected for query. Fields are selected from a drop down menu and field values are typed in by the user. Only one field can be searched in each query. Note that paragraph 4 of the Help section provides value names for the "Mutation Type" field. B: *Mutation Query Output*. The table shown has been returned in response to the query in A. Note that all fields are displayed. Reference table must be accessed in order to obtain reference links.

A.

HEXBdb Phenotype Data Retrieval

Please choose search field: Clinical phenotype ▼

Enter text or a value: subacute

Nucleotide number
Clinical phenotype
Onset
BioChemical phenotype
Diagnostic
Genotype
Ethnic origin
Submitter
References

HELP

Mutation name: For an amino acid change, eg. G269S (Gly269Ser), the notation consists of the amino acid followed by the number and then the mutant amino acid. One or three letter abbreviations for amino acid can be used.
For an intronic mutation, eg. IVS2+1g->a, the notation consists of IVS followed by intron number, then intronic position (can be +/-), the original nucleotide and finally the mutant nucleotide.
For a deletion or insertion, eg. c.547.insA, the notation consists of c followed by a nucleotide number (preceded and followed by a period), then ins or del and finally the nucleotide(s) that are inserted or deleted.

Nucleotide number: Position where mutation occurs.

Clinical phenotype: Acute, subacute, chronic, mild chronic, benign, asymptomatic (none) and unknown.

Onset: Infantile, late infantile, juvenile, late juvenile and adult.

Biochemical phenotype: Information regarding protein activity, processing and subunit association. Variants (Hexosaminidase Paris) and Neutral Polymorphisms are also indicated.

Genotype: Information regarding the second allele, in people where the mutation was identified, if available.

Diagnostic: Gain or loss of restriction enzyme sites, heteroduplex analysis, allele-specific amplification or allele-specific hybridization.

Ethnic origin: Ethnicity of patients or carriers.

Submitter: Consists of the first author's last name and date of publication, if published.

References: Consists of the short citation (first author's name and date of publication) for articles containing information on this mutation.

B.

PHENOTYPE LISTING											
THE DATE AND TIME IS: Mon Sep 25 21:27:38 2000											
TYPE: all											
VALUE: subacute											
UID	Nucleotide Number	Mutation (Alternative Name)	Clinical Phenotype	Onset	Biochemical Phenotype	Comments	Genotype	Diagnostic	Ethnicity	Submitter	Reference
5	619	D287V (De287Val)	Subacute	NA	30% (MUG), 21% (MUGS) Hex A, absent Hex B, S chains only in Hex A suggested defect in GM2 activator association for this allele (see Tyr466Ser).		Second allele Tyr466Ser			Cashman et al.	Banerjee et al. (1991) Cashman et al. (1996)
14	1260	P417L (Pro417Leu)	Subacute or mild chronic	Infantile (in some cases, see comments)	Mutant Hex B expressed in COS cells shows 78% activity of normal enzyme. When coupled with Ser255Arg in cis, Hex B activity reduced to 36% of normal.	Infantile phenotype when P417L is in cis with S255R in homozygote			Japanese, French Canadian, Italian	Wakamatsu et al.	Fujimaru et al. (1998) Gomez-Lira et al. (1999) McNamee et al. (1992) Wakamatsu et al. (1992)
17	1367	V466S (Tyr466Ser)	Subacute	NA	30% (MUG) and 21% (MUGS) Hex A, absent Hex B, S chains only in Hex A suggested defect in S-S association for this allele (see De287Val).		Second allele is De287Val			Cashman et al.	Banerjee et al. (1991) Banerjee et al. (1994) Cashman et al. (1996)
18	1368	IVS12:24g->a	Subacute or Asymptomatic	NA	Hexosaminidase Paris (i.e. Hex A+ and Hex B-) early defect in B subunit processing (due to fold properly, no association with alpha, no sheep erythrocyte). Reduced Hex A activity likely due to small amount of normal B to give 3-5% Hex A activity in subacute patients to 9-10% Hex A activity in asymptomatic subjects. Hex B activity absent.				French, Japanese, Canadian	Dreyfus et al.	Dietl et al. (1999) Dreyfus et al. (1977) McNamee et al. (1992) Wakamatsu and Suzuki (1989)

There were 4 record(s) found.

Fig. 2 A: Example of a phenotype search performed on HEXBdb. The phenotype data table has been selected for query. Fields are selected from a drop down menu and field values are typed in by the user. Only one field can be searched in each query. Note that paragraph 3 of the Help section provides value names for the "Clinical phenotype" field. B: *Phenotype Query Output*. The table shown has been returned in response to the query in A. Note that all fields are displayed. Reference table must be accessed in order to obtain reference links.

<div style="background-color: black; color: white; border-radius: 15px; padding: 5px; display: inline-block;">HEXAdb</div>	
HEXAdb Mutation Submission Form	
Please use this form to submit mutation data to the HEXA Mutation Analysis Database (HEXAdb). For more information please contact the curators.	
Submitter Information	
Name:	<input type="text"/>
Organization:	<input type="text"/>
E-mail:	<input type="text"/>
Mutation Information	
Mutation name: (eg. c.1236 A→G)	<input type="text"/>
Alternative name: (eg. G269S)	<input type="text"/>
Nucleotide number:	<input type="text"/>
Location:	<input type="text" value="select"/>
Length of lesion: (number of bases):	<input type="text"/>
Intronic position:	<input type="text"/>
Nucleotide change:	<input type="text" value="select"/>
Mutation type:	<input type="text" value="select"/>
Diagnostic: (eg. gain or loss of restriction sites)	<input type="text"/>
mRNA description: (abundance and size of transcript(s))	<input type="text"/>
Protein description: (abundance and size of polypeptide(s))	<input type="text"/>
Ethnic origin(s):	<input type="text"/>
Phenotype Information	
For information on the classification of TSD according to age of onset or clinical phenotype go to Tay-Sachs Disease GeneCensus, NIH	
Age of onset:	<input type="text" value="select"/>
Clinical phenotype:	<input type="text" value="select"/>
Biochemical phenotype: (%Hex A activity)	<input type="text"/>
Genotype: (second allele, if available)	<input type="text"/>
Comments:	<input type="text"/>
Ethnic origin(s): (only if you haven't entered it above)	<input type="text"/>
References:	<input type="text"/>
• Publication status	<input type="text" value="select"/>
• Full citation	<input type="text"/>
HEXAdb DeBelle	
HEXAdb data maintained by: Dr. F. Kaplan , P. Cordeiro , E. Novacki and Xing Zeng . The HEXAdb database and WEB site are maintained by the DeBelle Data Group .	
©1999 DeBelle Data Group	

Fig. 3 Mutation submission form.

Currently, the 3D structures of the catalytic domains of both the human β -hexosaminidase α and β subunits can be viewed online, as can the localization of putative active site residues and some missense mutations (only for the α subunit). Preselected groups of residues at which mutations lead to similar clinical phenotypes or functional impairments (i.e., catalytic site) can be viewed or the viewer may place the arrow on a section of the protein backbone and left-click the mouse to find the amino acid name and position number displayed on the status line. Viewers may also select different rotational views of the protein and different modes of presentation of the peptide backbone. No theoretical models exist, as of yet, for the G_{M2} activator protein.

Figure 4 illustrates the mapping of a subset of infantile acute and adult *HEXA* mutations onto the hexosaminidase model. Mutations leading to acute TSD cluster at or near the active site. Adult mutations more often involve residues on the protein surface. It is important to note that not all mutations conform to these general principles. It is anticipated that correlation of new data on the biochemical phenotypes and structural localization of specific mutant residues may lead to the definition of new functional domains on the Hex A protein such as the activator binding or the dimerization regions.

TSD screening mutation database development

This table will represent a collaborative effort and will include information regarding allele frequencies and numbers of alleles screened annually according to ethnic/geographic origin, number of affected fetuses identified by prenatal testing according to ethnic/geographic origin, screening tests for specific alleles, NTSAD-approved screening centers, etc. The planned format is illustrated in Figure 5. Mutation data will be recorded separately for Ashkenazi Jews and other individuals. The form is intended for use by laboratories conducting community-based Tay Sachs carrier screening. Only those alleles known to be relatively frequent are listed. Information requested includes allele frequency data and ethnic origin of carriers. This information is expected to be useful in selecting appropriate diagnostic tests for local populations/families.

DISCUSSION

The characterization of *HEXA*, *HEXB*, and *GM2A* gene mutations contributed to the understanding of the clinical heterogeneity of the G_{M2} gangliosidoses, the ethnic distribution of mutant alleles, and the structure-function relationships of the Hex isozymes. Information about *HEXA*, *HEXB*, and *GM2A* mutations has also provided for more accurate carrier detection, pre- and postnatal diagnosis, and prediction of outcome for affected individuals. The inclusion of this type of comprehensive information is only possible in databases which seek to provide in-depth characterization of one or a limited number of gene loci. These locus-specific databases can also include detailed information on diagnostic tests and ethnic distribution of alleles. By contrast, databases such as the Human Gene Mutation Database (HGMD) provide more limited informa-

Fig. 4 Viewing of Modeling Data. Selection box (lower right) permits selection of display of mutation groups. Further scrolling reveals controls for rotation of model and choices of backbone views (not shown). Following placement of arrow on peptide chain, left-clicking on mouse displays name and position of amino acid in status line.

tion on *all* known human mutations and are useful to physicians or teachers who require only identification of a specific gene locus. HGMD is limited to five useful fields: codon, nucleotide change, amino acid change, name of associated disorder, and linkage to the abstract of a reference.

Mutations in the *HEXA* gene

The majority of *HEXA* mutations cause complete loss of Hex A and lead to severe, infantile-onset neurodegenerative disease culminating in death by age 4. Later onset G_{M2} gangliosidoses are associated with 20 alleles producing varying amounts of residual enzymatic activity. Patients displaying the B1 variant phenotype are believed to be deficient in α -subunit catalytic activity.¹⁰ Two mutations cause Hex A pseudodeficiency. Of particular interest to laboratories engaged in allele testing are the six alleles identified only in carriers, which are presumed to be disease-causing based on the effect of mutation on enzyme activity in *in vitro* assays. Late- or adult-onset TSD is characterized by extreme variability of phenotype. This variation occurs within the patient population expressing the common G269S allele (variable expressivity) but also among patients expressing five other alleles leading to a late-onset (chronic) phenotype (heteroallelism). Knowledge of the frequency and distribution of these alleles should improve the efficacy of TSD carrier screening.

Ethnic distribution of TSD alleles

A 7.6-kb deletion in the first exon and flanking sequences is the most common allele in French Canadians.^{11,12} All other *HEXA* mutations are nucleotide substitutions: ($n = 81$), small insertions ($n = 4$), or small deletions ($n = 14$). Three alleles account for 92–98% of *HEXA* gene pathology in Ashkenazi Jews.^{13,14} The most common, the 1278 + TATC insertion, also

accounts for 20% of heterozygote alleles in non-Jewish European-derived populations. Carriers of TSD from other geographic isolates often harbor a specific subset of mutations even if the disease gene frequency is not elevated in the population. An IVS 9 + 1 G>A mutation is the most frequent TSD allele in carriers of northern European origin,¹⁵ while in the Japanese population, 79% of mutant alleles result from a C178Y substitution not found in other populations.¹⁶ The two pseudodeficiency mutations account for 36% of alleles in non-Jewish Americans of European descent.¹⁷ The remainder of TSD mutations are considered to be “rare” or “private.”

Genotype/phenotype correlations

HEXA gene mutations associated with variant forms of TSD have been mapped onto a modeled structure of the catalytic domain of Hex A.⁹ Mutations leading to acute TSD were identified in two clusters: one located at or near the active site and including mutations associated with the B1 phenotype, the second cluster involving residues far from the active site which are associated with effects on enzyme stability and protein processing. Mutations leading to a less severe phenotype were also divided into two groups. Benign and adult mutations most often involved residues on the protein surface where a change does not exert a significant effect on protein stability. More severe late infantile and juvenile mutations affected residues which are located closer to the protein core. Not all phenotypes could be interpreted based on these general principles.

Mutations in the *HEXB* gene

Like TSD, the clinical severity of Sandhoff disease correlates with residual activity of Hex A. The most common Sandhoff mutation is a 16-kb deletion including exons 1–5 and flanking sequence.¹⁸ All point mutations are associated with subacute

HEXAdb

Tay-Sachs Disease Screening Data Submission Form

Please use this form to submit Tay-Sachs disease screening data to the
HEXA Mutation Analysis Database (HexAdb).
For more information please contact the [hexadb](mailto:hexadb@mcgill.ca).

Submitter Information

Name: _____
 Organization: _____
 E-mail: _____

Screening Information

Year: _____
 Country: _____
 State / Province: _____
 City: _____

Alleles and Individuals	Individuals pre-screened with serum assay		Individuals pre-screened with leukocyte assay	
	Number of individuals tested	Number of carriers identified	Number of individuals tested	Number of carriers identified
Enzyme Assay				
G805A				
1278 +TATC				
IVR12+1G--C				
C*39T				
C*45T				
IVS9+1 G--A				
*632b (5') deletion				
Other mutation (specify):				

Are all the carriers unrelated?
(Independently segregating carrier chromosomes)

☐ Yes ☐ No ☐ Don't know

Other Individuals	Individuals pre-screened with serum assay		Individuals pre-screened with leukocyte assay		Ethnicity of carriers (if known)
	Number of individuals tested	Number of carriers identified	Number of individuals tested	Number of carriers identified	
Enzyme Assay					
G805A					
1278 +TATC					
IVR12+1G--C					
C*39T					
C*45T					
IVS9+1 G--A					
*632b (5') deletion					
Other mutation (specify):					

Are all the carriers unrelated?
(Independently segregating carrier chromosomes)

☐ Yes ☐ No ☐ Don't know

[Submit Form](#) [Reset](#)

| HEXAdb | DeReBe |

HEXAdb data submitted by: Dr. F. Kanaan, Peter M. Hershfield, Emily Cordeiro and Alan Zimm.
The HEXAdb database and WEB site are maintained by the DeReBe Data Group.

©2000 DeReBe Data Group

Fig. 5 Tay-Sachs disease screening data submission form. This form is intended for use by laboratories conducting community-based Tay Sachs carrier screening. The form lists only those alleles known to be relatively frequent and asks for allele frequency data and ethnic origin of carriers. This information is expected to be useful in selecting appropriate diagnostic tests for local populations/families.

or chronic forms of disease. Mutations with unusual splicing abnormalities have been identified.¹⁹ Asymptomatic individuals usually show some Hex A activity and no Hex B activity, a biochemical phenotype referred to as Hexosaminidase Paris. Benign mutations resulting in heat labile Hex B have the potential to compromise the accuracy of carrier detection and prenatal diagnosis.

Mutations in the *GM2A* gene

Only five mutations have been documented in the *GM2A* gene encoding the G_{M2} activator protein, and all patients have been homozygous for individual mutant alleles associated with complete absence of G_{M2} ganglioside cleavage. The range of defects includes defective Hex A binding,^{20–22} enzyme instability,²³ and degradation in the ER or Golgi.²⁴

Conclusion

The *HEXAdb*, *HEXBdb*, and *GM2Adb* online databases assist in regrouping scattered information into a centralized location that is easily accessible to the scientific community. The ability to submit new mutations online provides an efficient method for collecting and disseminating new information about these genes and their associated phenotypes. Protein modeling of individual mutations may also assist in developing a better understanding of genotype/phenotype correlations. The G_{M2} gangliosidoses databases and others like it serve the interests of genomics, human genetic diversity, medicine, and the needs of patients, families, and communities.

Acknowledgments

This work was supported by grants from the Medical Research Council of Canada to Peter Hechtman and Feige Kaplan and from the Réseau Génétique Humaine Appliquée to Feige Kaplan. The G_{M2} gangliosidoses home page is located at the DeReBe server of the Charles R. Scriver Biochemical Genetics Unit in the Montreal Children's Hospital.

References

1. Johnson WG. The clinical spectrum of hexosaminidase deficiency diseases. *Neurology* 1981;31:1453–1456.
2. Kolodny E, Raghavan S. G_{M2} -gangliosidosis. Hexosaminidase mutations not of the Tay-Sachs type produce unusual clinical variants. *Trends Neurosci* 1983;6:16–20.
3. Navon R, Argov Z, Frisch A. Hexosaminidase A deficiency in adults. *Am J Med Genet* 1986;24:179–196.
4. Navon R. Molecular and clinical heterogeneity of adult G_{M2} gangliosidosis. *Dev Neurosci* 1991;13:295–298.
5. Federico A, Palmeri S, Malandrini A, Fabrizi G, Mondelli M, Guazzi GC. The clinical aspects of adult hexosaminidase deficiencies. *Dev Neurosci* 1991;13:280–287.
6. Cotton RG, McKusick V, Scriver CR. The HUGO Mutation Database Initiative. *Science* 1998;279:10–11.
7. Scriver CR, Nowacki PM, Lehvaslaiho H. Guidelines and recommendations for content, structure, and deployment of mutation databases. *Hum Mutat* 1999;13:344–350.
8. Scriver CR, Nowacki PM, Lehvaslaiho H. Guidelines and recommendations for content, structure, and deployment of mutation databases, II: journey in progress. *Hum Mutat* 2000;15:13–15.
9. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. Bacterial chitinase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat Struct Biol* 1996;3:638–648.

10. Bayleran J, Hechtman P, Kolodny E, Kaback M. Tay-Sachs disease with hexosaminidase A: characterization of the defective enzyme in two patients. *Am J Hum Genet* 1987;41:532–548.
11. Myerowitz R, Hogikyan ND. Different mutations in Ashkenazi Jewish and non-Jewish French Canadians with Tay-Sachs disease. *Science* 1986;232:1646–1648.
12. Myerowitz R, Hogikyan ND. A deletion involving Alu sequences in the beta-hexosaminidase alpha-chain gene of French Canadians with Tay-Sachs disease. *J Biol Chem* 1987;262:15396–15399.
13. Triggs-Raine BL, Feigenbaum AS, Natowicz M, Skomorowski MA, Schuster SM, Clarke JT, Mahuran DJ, Kolodny EH, Gravel RA. Screening for carriers of Tay-Sachs disease among Ashkenazi Jews: a comparison of DNA-based and enzyme-based tests. *N Engl J Med* 1990;323:6–12.
14. Fernandes MJ, Kaplan F, Clow CL, Hechtman P, Scriver CR. Specificity and sensitivity of hexosaminidase assays and DNA analysis for the detection of Tay-Sachs disease gene carriers among Ashkenazic Jews. *Gen Epidemiol* 1992;9:169–175.
15. Akerman BR, Zielenski J, Triggs-Raine BL, Prencce EM, Natowicz MR, Lim-Steele JS, Kaback MM, Mules EH, Thomas GH, Clarke JT. A mutation common in non-Jewish Tay-Sachs disease: frequency and RNA studies. *Hum Mutat* 1992;1:303–309.
16. Tanaka A, Sakazaki H, Murakami H, Isshiki G, Suzuki K. Molecular genetics of Tay-Sachs disease in Japan. *J Inherit Metab Dis* 1994;17:593–600.
17. Kaback M, Lim-Steele J, Dabholkar D, Brown D, Levy N, Zeiger K. Tay-Sachs disease—carrier screening, prenatal diagnosis, and the molecular era: an international perspective, 1970 to 1993. *The International TSD Data Collection Network. JAMA* 1993;270:2307–2315.
18. Neote K, McInnes B, Mahuran DJ, Gravel RA. Structure and distribution of an Alu-type deletion mutation in Sandhoff disease. *J Clin Invest* 1990;86:1524–1531.
19. Wakamatsu N, Kobayashi H, Miyatake T, Tsuji S. A novel exon mutation in the human beta-hexosaminidase beta subunit gene affects 3' splice site selection. *J Biol Chem* 1992;267:2406–2413.
20. Schroder M, Schnabel D, Suzuki K, Sandhoff K. A mutation in the gene of a glycolipid-binding protein (G_{M2} activator) that causes G_{M2}-gangliosidosis variant AB. *FEBS Lett* 1991;290:1–3.
21. Xie B, Wang W, Mahuran DJ. A Cys138-to-Arg substitution in the G_{M2} activator protein is associated with the AB variant form of GM2 gangliosidosis. *Am J Hum Genet* 1992;50:1046–1052.
22. Xie B, Rigat B, Smiljanic-Georgijev N, Deng H, Mahuran D. Biochemical characterization of the Cys138Arg substitution associated with the AB variant form of G_{M2} gangliosidosis: evidence that Cys138 is required for the recognition of the G_{M2} activator/G_{M2} ganglioside complex by beta-hexosaminidase A. *Biochemistry* 1998;37:814–821.
23. Schroder M, Schnabel D, Hurwitz R, Young E, Suzuki K, Sandhoff K. Molecular genetics of G_{M2}-gangliosidosis AB variant: a novel mutation and expression in BHK cells. *Hum Genet* 1993;92:437–440.
24. Schepers U, Glombitza G, Lemm T, Hoffmann A, Chabas A, Ozand P, Sandhoff K. Molecular analysis of a G_{M2}-activator deficiency in two patients with G_{M2}-gangliosidosis AB variant. *Am J Hum Genet* 1996;59:1048–1056.