Journal of Genetics (2020)99:42
https://doi.org/10.1007/s12041-020-01208-8

RESEARCH ARTICLE

In silico analysis of the effects of disease-associated mutations of β-hexosaminidase A in Tay–Sachs disease

MOHAMMAD IHSAN FAZAL[1], RAFAL KACPRZYK[2] and DAVID J. TIMSON[3]*

[1]*Brighton and Sussex Medical School, University of Sussex, Falmer, Brighton BN1 9PX, UK*
[2]*School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast BT9 7BL, UK*
[3]*School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton BN2 4GJ, UK*
*For correspondence. E-mail: d.timson@brighton.ac.uk.

Received 6 September 2019; revised 25 February 2020; accepted 24 April 2020

Abstract. Tay–Sachs disease (TSD), a deficiency of β-hexosaminidase A (Hex A), is a rare but debilitating hereditary metabolic disorder. Symptoms include extensive neurodegeneration and often result in death in infancy. We report an *in silico* study of 42 Hex A variants associated with the disease. Variants were separated into three groups according to the age of onset: infantile (n=28), juvenile (n=9) and adult (n=5). Protein stability, aggregation potential and the degree of conservation of residues were predicted using a range of *in silico* tools. We explored the relationship between these properties and the age of onset of TSD. There was no significant relationship between protein stability and disease severity or between protein aggregation and disease severity. Infantile TSD had a significantly higher mean conservation score than nondisease associated variants. This was not seen in either juvenile or adult TSD. This study has established that the degree of residue conservation may be predictive of infantile TSD. It is possible that these more highly conserved residues are involved in trafficking of the protein to the lysosome. In addition, we developed and validated software tools to automate the process of *in silico* analysis of proteins involved in inherited metabolic diseases. Further work is required to identify the function of well-conserved residues to establish an *in silico* predictive model of TSD severity.

Keywords. GM2 gangliosidosis; sphingolipidosis; inherited metabolic disease; β-hexosaminidase A; protein stability; Tay–Sachs disease.

Introduction

Tay–Sachs disease (TSD; type I G_{M2}-gangliosidosis; OMIM: 272800) is a rare, progressive autosomal recessive disease with an incidence of one in 320,000 births in the general populations and a carrier frequency of one in 250 (Lew *et al.* 2015). Commonly fatal, caused by the abnormal metabolism of GM1 ganglioside, a molecule that accounts for 10–12% of the total lipid content of neuronal cellular membranes and contributes to ~1.5% of dry brain weight (Svennerholm and Fredman 1980; Tettamanti 2004). In healthy individuals, degradation of GM1 ganglioside occurs in the acidic confines of the lysosome (Mark *et al.* 2003). Subunits are removed sequentially from the ganglioside with each step being catalysed by specific enzymes (figure 1a). The conversion of GM2 to GM3 is catalysed by β-hexosaminidase A (HexA/EC 3.2.1.52), a heterodimeric enzyme composed of two subunits, α and β, encoded by the genes *HexA* and *HexB*, respectively (Henrissat and Davies 1997). The α-subunit is uniquely able to bind the negatively charged GM2 ganglioside/activator protein complex and hydrolyse the terminal *N*-acetylgalactosamine (GalNAc) (Mark *et al.* 2003; Sharma *et al.* 2003). If a mutation in *HexA* causes HexA to dysfunction, GM2 accumulates within the neuron resulting in TSD.

The human enzyme is a heterodimer of one α-subunit and one β-subunit encoded by the *HEXA* and *HEXB* genes, respectively, β-hexosaminidase A (Korneluk *et al.* 1986; Neote *et al.* 1988). Two other isoforms of the

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12041-020-01208-8) contains supplementary material, which is available to authorized users.

Published online: 15 May 2020

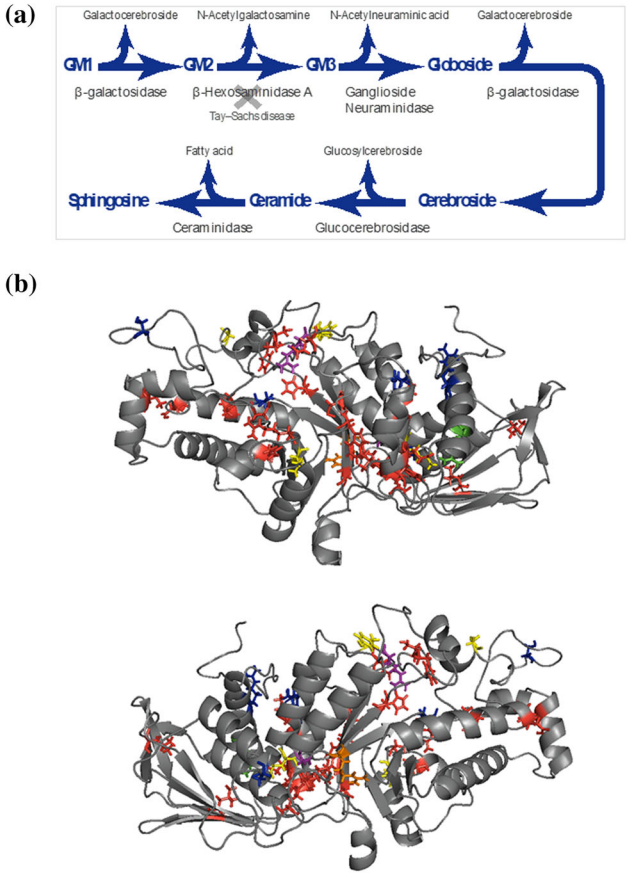

Figure 1. The pathway of GM1 ganglioside degradation and structure of β-hexosaminidase A α-subunit (HexA). (a) TSD occurs when a mutated α-subunit causes β-hexosaminidase A to dysfunction resulting in the accumulation of GM2 and eventual neuronal cell death. (b) The structure of one monomer of the HexA protein with the location of residues known to be affected by disease-mutations mapped. The backbone of the protein is shown in cartoon format in grey. Affected residues are shown in stick format coloured red for those associated with the infantile onset form of the disease, blue for those associated with the juvenile form and yellow for the adult form. A small number of residues were associated with both the infantile and juvenile forms (purple), the infantile and adult forms (orange) and the juvenile and adult forms (green). The images represent a single subunit from the crystal structure of HexA (PDB: 2GJX) which had been subjected to energy minimization. The lower image related to the upper one by a rotation of 180° around the *y*-axis. The images were generated using PyMol (www.pymol.org).

β-hexosaminidase protein are present in humans—an αα homodimer (β-hexosaminidase S) and a ββ homodimer (β-hexosaminidase B) (Robinson and Stirling 1968; Hayase and Kritchevsky 1973; Ikonne *et al.* 1975; Aruna and Basu 1976; Hepbildikler *et al.* 2002). Structurally, the α-subunits and β-subunits have a similar overall fold (root mean square deviation of ~0.7 Å over ~460 equivalent C_α atoms) (Lemieux *et al.* 2006). Both subunits contain an active site which catalyses the hydrolysis of terminal *N*-acetylgalactosamine and *N*-acetylglucoasamine residues from complex polysaccharide moieties. The catalytic mechanism is postulated to involve two residues—Glu-323 which acts as a base and protonates the glycosidic oxygen atom and Asp-322 which acts as a base, polarising the carbonyl in the N-acetyl group and stabilising the transition state (Lemieux *et al.* 2006; Passos *et al.* 2011). However, only the α-subunit can catalyse the hydrolysis of polysaccharides containing sialic acid groups (Kytzia and Sandhoff 1985). This specificity arises from different active site structures: the α-subunit has a positively charged arginine residue which is able to bind to the negatively charged sialic acid (Mark *et al.* 2001; Sharma *et al.* 2003; Lemieux *et al.* 2006).

TSD occurs in four forms: infantile, juvenile, adult and the B1 variant. In infantile TSD, the most severe form, patients appear normal until, within the first six months of life, there is developmental retardation, muscular weakness and an increased startle response. As the disease progresses, there is a loss of hearing and vision, cognitive impairment, seizures and paralysis resulting in death between 3 and 5 years of age (Fernandes Filho and Shapiro 2004; Maegawa *et al.* 2006; Masingue *et al.* 2020). In juvenile TSD, onset occurs in childhood with one cohort study of 21 patients stating a mean onset time of 5.3 ± 4.1 years with symptoms such as gait disturbances, ataxia, dysarthria and developmental delay with a median survival time of 14.5 years (Maegawa *et al.* 2006). As the disease progresses cognitive and motor skills deteriorate and anarthria, muscle wasting and incontinence occurs (Specola *et al.* 1990; Maegawa *et al.* 2006). Adult TSD has the most variable time of onset and is the least severe of the three as it is not always fatal (Masingue *et al.* 2020). One case series of 21 patients reported a mean onset age of 18.1 years (Neudorfer *et al.* 2005), whereas another study reported onsets of 15 and 20 years (Hurowitz *et al.* 1993). A separate case report reported onset age of 47 years (Steiner *et al.* 2016). Symptoms involve psychiatric disturbance, dysarthria, gait abnormalities, and leg weakness. As the disease progresses, the patient may require a wheelchair and experience dysphagia and worsening of psychiatric state (Neudorfer *et al.* 2005). The B1 variant can present as any of the other variants. However, it exhibits a unique biochemical profile. HexA in B1 TSD is able to catalyse the conversion of the neutral synthetic substrate 4-methylumbelliferyl-*N*-acetyl-fl-D-glucosamine, which is sometimes used in diagnostic tests, but is unable to hydrolyse GM2 ganglioside *in vivo* or the acidic synthetic substrate 4-methylumbelliferyl-β-D-*N*-acetylglucosamine-6-sulphate (Gordon *et al.* 1988). Although the neurological symptoms can be partially mitigated with drugs and other therapies, there are no effective treatments available for TSD. Potential treatments under development include inhibition of substrate synthesis (thus reducing the accumulation of nondegraded G_{M2} ganglioside), gene therapy, enzyme replacement therapy, neuronal stem cell transplants, pharmacological chaperones and gene editing (Desnick and Kaback 2001; Platt *et al.* 2001; Tropak *et al.* 2004; Cachon-Gonzalez *et al.* 2006; Akeboshi *et al.* 2007; Maegawa *et al.* 2007; Rountree *et al.* 2009; Tropak *et al.*

2010; Tsuji *et al.* 2011; Gray-Edwards *et al.* 2018; Karumuthil-Melethil *et al.* 2016; Kato *et al.* 2017; Ornaghi *et al.* 2020; Ou *et al.* 2020).

Here, we report an *in silico* investigation to explore how molecular changes in the HexA protein relate to the severity of TSD. The aim was to build the foundation for a predictive framework. Similar work has been done in the past to successfully develop predictive frameworks for several other diseases such as mevalonate kinase deficiency (Browne and Timson 2015), type III galactosaemia (McCorvie and Timson 2013) and triose phosphate isomerase deficiency (Oliver and Timson 2017). Developing a predictive framework that is reliable, accurate and effective could assist in the genetic counselling of a TSD carrier or in guiding prognoses for a TSD sufferer. As all studies were to be performed *in silico*, the opportunity arose to build software tools that could aid the accuracy and speed and reduce the cost of future work in this field.

Materials and methods

Datasets

The identification of mutations known to be pathogenically associated with TSD was performed by a literature search using the NCBI PubMed database (http://www.ncbi.nlm.nih.gov/PubMed). Only mutations reported and categorized as pathogenic in the NCBI clinvar database were included (http://www.ncbi.nlm.nih.gov/clinvar). The classification of the disease was noted along with molecular consequence. Mutants were omitted if they resulted in early termination of the peptide, frameshift or were compound heterozygous. The crystal structure of wild-type human HexA (PDB: 2GJX) (Lemieux *et al.* 2006) was obtained from the Protein Data bank (https://www.rcsb.org/) (Berman *et al.* 2000). The amino acid sequence of wild-type HexA was obtained from the UniProt database (http://www.uniprot.org/uniprot/P06865) (Pundir *et al.* 2017; UniProt Consortium 2018).

Structural analysis

YASARA (http://www.yasara.org) (Krieger *et al.* 2009) is based on the AMBER force-field. It computationally solvates and energy minimises protein structures. All PDB files were processed through YASARA. Where possible, these energy-minimized structures were used for any analysis involving protein structure. Superpose (http://wishart.biology.ualberta.ca/superpose/) (Maiti *et al.* 2004) calculates protein structure superpositions of two or more PDB files based on a modified quaternion eigenvalue approach. Comparisons were made between the wild-type PDB and each mutant variant PDB in turn, the root-mean-square deviation (RMSD) was recorded.

Prediction of biochemical effects of each missense mutation on HexA aggregation

Predicting the effect of a mutation on the aggregation propensity of the protein was done by using freely available software. Fluorescence probe measurement of live macrophages has shown intralysosomal pH to be 4.7 to 4.8 with other studies placing estimates between 4.5 and 5 (Ohkuma and Poole 1978; Lange *et al.* 2006; Tabeta *et al.* 2006; Coen *et al.* 2012). With this in mind, a pH of 5 was selected as the webservers only accepted pH as an integer. Additionally, a temperature of 310 K and an ionic strength of 0.15 M were used to replicate the human lysosomal conditions as best as the software allowed. All HexA sequences had the first 22 residues removed to simulate the initial proteolytic cleavage that occurs in the endoplasmic reticulum before HexA is trafficked to the lysosome (Little *et al.* 1988).

TANGO (http://tango.crg.es/) uses a statistical mechanics algorithm to estimate the propensity for a protein to aggregate (Fernandez-Escamilla *et al.* 2004). The C-terminus and N-terminus status was set to free. Zyggregator (http://www-mvsoftware.ch.cam.ac.uk/index.php/zyggregator) is a web server that uses the Zyggregator algorithm to make predictions of the propensity of polypeptide chains to form protofibrillar assemblies (Tartaglia and Vendruscolo 2008). CamSol (http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic) is a web server that predicts the solubility and the generic aggregation propensity of a protein from a given sequence (Sormanni *et al.* 2015).

Prediction of biochemical effects of each missense mutation on HexA stability and structure

Prediction of the effect of mutations on the stability of the protein was done by using several freely available web servers. I-Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) predicts the effect of a mutation on the stability of a protein (Capriotti *et al.* 2008). It performs a requested point mutation on a provided PDB file or a single letter code amino acid sequence and returns the change in the Gibbs free energy change of unfolding compared to the wild-type protein (ΔΔG). The wild-type HexA PDB file was used as an input and a pH of 5 and a temperature of 37°C were chosen to replicate human lysosomal conditions. The Cologne University protein stability analysis tool (CUPSAT) available from (http://cupsat.tu-bs.de/) predicts the effects of point mutations on the overall protein stability. It returns the ΔΔG for every possible point mutation at a requested location using a PDB file (Parthiban *et al.* 2006). The SDM server (http://131.111.43.103/) uses a knowledge-based system to predict the effect of a given list of mutations on a wild-type PDB file and returns the ΔΔG (Pandurangan *et al.* 2017). The mutation cutoff scanning matrix (mCSM) (http://biosig.unimelb.edu.au/mcsm/stability) (Pires *et al.* 2014) uses predictive models

trained with graph-based signatures to predict the effects of a given list of point mutations on the stability of a wild-type PDB file and returns the ΔΔG.

Multiple sequence alignment and comparison of residue conservation of HexA variants

Wild-type HexA amino acid sequences were obtained from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein). Only complete sequences of HexA found in the RefSeq database were included. A total of 101 sequences were identified including 79 mammals, 14 fish, five birds, one lizard, one nematode and one plant. After elimination of duplicates, 73 homologous sequences remained (for a full list, see table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). These were input to a multiple sequence alignment server Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers *et al.* 2011) with human HexA added to the first position. The resultant multiple aligned sequence was entered into the Scorecons server to output Valdar conservation scores, an estimate of a residue's conservation throughout multiple sequences (Valdar 2002).

Development of software tools to simplify and increase accuracy of data acquisition

Python scripts were created using PyCharm (JetBrains 2017) with manipulation of protein structures (PDB files) performed by communicating with PyMol (Schrödinger 2018) via its application programming interface (API). The openpyxl module (Gazoni and Clark 2018) was used to read/write Microsoft Excel files (ISO/IEC 29500-1:2016) in python. A python script loaded the energy minimized wild-type HexA structure and a Microsoft Excel file containing a list of mutations stored as comma-separated values (e.g. Arg,252,His). The comma-separated values were split and input sequentially to the mutagenesis function within PyMol to produce PDB files for all mutants.

A command line based program was written in Pascal using Lazarus (Team 2018) to facilitate the production of mutant sequences and to output them as text files formatted for further analysis. The program held a copy of the wildtype HexA sequence in memory. The user selects a residue and mutates it to an amino acid single letter code of choice and saves it to a text file – the sequence is reset to wildtype after every save. The program also allows for deletions, early terminations and can output the length of the current sequence. A batch file was created in Notepad++ (Ho 2018) on which the user can drag and drop text files to input it directly into TANGO thereby bypassing any command line interaction.

R scripts were created using RStudio (Team 2016). The RSelenium package (Harrison 2017) was used to automate interaction with and to extract data from webservers. The xlsx package (Dragulescu 2014) was used to read/write Microsoft Excel files in R. An R script was created to automate user interaction with the Zyggregator web server. The script holds a copy of the wildtype HexA sequence in memory and loads a Microsoft Excel file containing a list of mutations stored as comma-separated values. The comma-separated values were split to yield the position of the mutation and the amino acid to mutate to. The amino acids were converted from three letter code to single letter code (i.e. Arg to R) and the wild-type HexA sequence was modified accordingly before being input to Zygreggator along with the experimental conditions. The script is able to identify when the Zygreggator calculation is complete and navigate to the results page. The relevant data is captured and stored in memory before being saved to a Microsoft Excel file. An identical approach was used to create an R script for CamSoli.

An R script was created to automate the user interaction of iMutant 3.0. The script contains the file path to the energy minimized wild-type HexA structure and loads a comma separated values file (CSV) containing a list of mutations. The comma-separated values are split to yield the location of mutation and desired amino acid. The wild-type HexA structure is uploaded to iMutant 3.0 and experimental conditions are entered. The script waits for iMutant 3.0 to finish its calculations before extracting the ΔΔG values. An identical approach was used to create an R script for CUPSAT. An R script was also created to automate the user interaction of superpose. An approach similar to the iMutant 3.0 script was taken. However, key differences were the need to upload both the wild-type and mutant PDB rather than entering mutant locations and amino acids.

An important step in the automation of webservers was the identification of the desired result. Superpose was the only webserver that consistently presented the result in the same cell of an html table. In this case, it was straightforward to use RSelenium's getElementText request. The other webservers presented the result in either plain text (Zyggregator, Camsoli, Imutant 3.0) or as a table where the result could be in any row (CUPSAT). In this case a pattern recognition method (R-Project 2018) was used to identify the text containing desired result (for the patterns used, see supplements E-H under the code comment '#Define patterns'). All code was run with test data and samples of the final outputs were manually examined to assess the level of function of the software tools.

Data analysis

All data analysis and graph production was carried out using Graphpad Prism (GraphPad Software 2017) Where data normalization was performed, the equation $x_{new} = x_n/x_{wildtype}$. was used to normalize data with respect to the wild type. Where no wild-type value was available, the equation $x_{new} = (x_n - x_{min})/(x_{max} - x_{min})$ was used to

Table 1. The variants included in this study, their phenotype and references to the studies describing them.

Variant	TSD variant	Reference	Provean score	SIFT score
Leu39Arg	Infantile	Akli *et al.* 1993	− 4.49 (Deleterious)	0.001 (Damaging)
Glu114Lys	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 3.68 (Deleterious)	0.001 (Damaging)
Leu127Arg	Infantile	Akli *et al.* 1993; Montalvo *et al.* 2005	− 5.48 (Deleterious)	0.000 (Damaging)
Arg170Gln	Infantile	Drucker *et al.* 1992; Kaufman *et al.* 1997; Nakano *et al.* 1990	− 3.85 (Deleterious)	0.000 (Damaging)
Arg170Trp	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014; Triggs-Raine *et al.* 1991	− 7.71 (Deleterious)	0.000 (Damaging)
Arg178Cys	B1 (Infantile)	Navon *et al.* 1995; Tanaka *et al.* 1990a	− 7.71 (Deleterious)	0.000 (Damaging)
Arg178Leu	B1 (Infantile)	Triggs-Raine *et al.* 1991	− 6.74 (Deleterious)	0.000 (Damaging)
Val192Leu	Infantile	Ainsworth and Coulter-Mackie 1992; Hou *et al.* 1996	− 1.26 (Neutral)	0.002 (Damaging)
His204Arg	Infantile	Akli *et al.* 1993	− 7.8 (Deleterious)	0.000 (Damaging)
Ser210Phe	Infantile	Akli *et al.* 1991	− 5.82 (Deleterious)	0.000 (Damaging)
Phe211Ser	Infantile	Akli *et al.* 1993; Brown *et al.* 1989; Montalvo *et al.* 2005	− 7.73 (Deleterious)	0.000 (Damaging)
Arg252Leu	Infantile	Tanaka *et al.* 2003	− 6.83 (Deleterious)	0.002 (Damaging)
Asn295Ser	Infantile	Tanaka *et al.* 2003	− 4.19 (Deleterious)	0.05 (Damaging)
Met301Arg	Infantile	Akli *et al.* 1993	− 4.87 (Deleterious)	0.001 (Damaging)
Asp322Asn	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 4.88 (Deleterious)	0.000 (Damaging)
Asp322Tyr	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 8.73 (Deleterious)	0.000 (Damaging)
Gln374Arg	Infantile	Montalvo *et al.* 2005	− 3.92 (Deleterious)	0.001 (Damaging)
Gln374Pro	Infantile	Sheth *et al.* 2014	− 5.88 (Deleterious)	0.001 (Damaging)
Arg393Pro	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 2.57 (Deleterious)	0.005 (Damaging)
Trp420Cys	Infantile	Matsuzawa *et al.* 2003; Tanaka *et al.* 2003; Tanaka *et al.* 1990b	− 12.44 (Deleterious)	0.000 (Damaging)
Cys458Tyr	Infantile	Matsuzawa *et al.* 2003; Tanaka *et al.* 1994	− 9.5 (Deleterious)	0.000 (Damaging)
Glu462Val	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 6.62 (Deleterious)	0.000 (Damaging)
Gly478Arg	Infantile	Mistri *et al.* 2012; Sheth *et al.* 2014	− 2.16 (Neutral)	0.012 (Damaging)
Glu482Lys	Infantile	Dersh *et al.* 2016; Nakano *et al.* 1988; Proia and Neufeld 1982	− 3.71 (Deleterious)	0.000 (Damaging)
Leu484Gln	Infantile	Tanaka *et al.* 1994	− 5.57 (Deleterious)	0.000 (Damaging)
Leu484Pro	Infantile	Tanaka *et al.* 1994	− 6.5 (Deleterious)	0.000 (Damaging)
Arg504Cys	Infantile	Akli *et al.* 1991; Boonyawat *et al.* 2016; Montalvo *et al.* 2005; Neudorfer *et al.* 2005; Paw *et al.* 1990; Raghavan *et al.* 1985; Shapiro and Natowicz 2009	− 7.43 (Deleterious)	0.003 (Damaging)
Trp485Arg	Infantile	Akalin *et al.* 1992	− 13 (Deleterious)	0.000 (Damaging)
Cys58Tyr	Juvenile	Najmabadi *et al.* 2011	− 8.58 (Deleterious)	0.001 (Damaging)
Arg178His	B1 (Juvenile)	Brown *et al.* 1989; Giraud *et al.* 2010; Maegawa *et al.* 2006; Montalvo *et al.* 2005; Ou *et al.* 2019; Tanaka *et al.* 1990a	− 4.82 (Deleterious)	0.000 (Damaging)
Gly250Asp	Juvenile	Trop *et al.* 1992	− 6.83 (Deleterious)	0.001 (Damaging)
Asp258His	B1 (Juvenile)	Fernandes *et al.* 1992; Fernandes *et al.* 1997	− 6.83 (Deleterious)	0.000 (Damaging)
Ser279Pro	Juvenile	Drucker *et al.* 1997	− 2.89 (Deleterious)	0.06 (Tolerated)
Trp474Cys	Juvenile	Maegawa *et al.* 2006; Neudorfer *et al.* 2005; Petroulakis *et al.* 1998	− 12.21 (Deleterious)	0.000 (Damaging)
Arg499His	Juvenile	Maegawa *et al.* 2006; Ou *et al.* 2019; Paw *et al.* 1990; Shapiro and Natowicz 2009; Tanaka *et al.* 2003; Zampieri *et al.* 2012	− 4.64 (Deleterious)	0.000 (Damaging)
Arg499Cys	Juvenile	Akli *et al.* 1993; Mules *et al.* 1992; Tanaka *et al.* 2003	− 7.43 (Deleterious)	0.000 (Damaging)
Arg504His	Juvenile	Giraud *et al.* 2010; Paw *et al.* 1990	− 4.64 (Deleterious)	0.001 (Damaging)
Tyr180His	Adult	De Gasperi *et al.* 1996	− 4.45 (Deleterious)	0.000 (Damaging)
Lys197Thr	Adult	Akli *et al.* 1993	− 5.85 (Deleterious)	0.000 (Damaging)
Arg252His	Adult	Ribeiro *et al.* 1996	− 4.88 (Deleterious)	0.001 (Damaging)
Gly269Ser	Adult	Maegawa *et al.* 2006; Navon and Proia, 1989; Neudorfer *et al.* 2005; Shapiro and Natowicz 2009	− 5.51 (Deleterious)	0.177 (Tolerated)
Val391Met	Adult	Navon *et al.* 1995	− 2.79 (Deleterious)	0.001 (Damaging)

normalize the data with respect to the maximum and minimum values. One-way ANOVA was used to assess whether there was any statistically significant difference between the means of groups. Tukey post-hoc tests were done to identify as which groups were significantly different to each other. All graphs were plotted as the mean ± standard deviation. The significance threshold was set at $P<0.05$.

Results

Identification of point mutations pathogenic for TSD

A total of 42 pathogenic variants were identified and included in this study. Of these, five exhibited an adult TSD phenotype, 27 exhibited an infantile TSD phenotype and six exhibited a juvenile phenotype. Additionally, two B1 variants exhibited an infantile phenotype and were included in the infantile group and two B1 variants exhibited a juvenile phenotype and were included with this group. Table 1 outlines the details of the variants identified. Infantile mutants were regarded as the most severe form; adult, the least severe and juvenile as intermediate. The positions of the residues affected were mapped onto the HexA crystal structure (figure 1b). These positions were located throughout the structure and did not appear to cluster in any specific region.

Effect of point mutations on aggregation potential

The results from three aggregation calculators (Zyggregator, Camsoli and TANGO) were normalized with respect to the wild type. A mean aggregation score for each variant was then calculated. The mean for each phenotype was then calculated and the results plotted in figure 2a, along with the wild type. There was no significant difference between any of the groups.

Effect of point mutations on protein stability and structure

The ΔΔG for all variants were obtained from four webservers (iMutant 3.0, CUPSAT, SDM and mCSM) and normalized with respect to the minimum and maximum values. The mean ΔΔG for each variant was calculated and an overall mean for each phenotype was plotted in figure 2b. There was no significant difference in mean ΔΔG between groups. The root mean square deviation (RMSD) for the energy minimized model of each variant compared to wild type was calculated using the Superpose webserver. Mean RMSD was calculated for each phenotype and plotted in figure 2c. There was no significant difference in mean RMSD between the groups.

Degree of conservation of residues affected by mutations

The mean conservation scores of nondisease associated residues were compared to the mean conservation scores of each phenotype and plotted in figure 2d. There was a highly significant difference ($P=0.001$) in mean conservation scores between the nondisease associated residues and the infantile phenotype residues. There were no significant differences seen with the adult or juvenile groups.

Software tools to simplify and increase accuracy of data acquisition

A number of tools were created to help automate these analyses (and similar analyses for other diseases). These tools have been made publicly available at the open science framework (www.osf.io/a5gsb) and examples of their implementation are shown in figure 3. A python script was created to automate the production of variant PDB files. This enabled the automated generation of (unminimized) variant models of the protein in pdb file format from an Excel file containing details of the variants (figure 3a). To automate analysis using the TANGO webserver, a pascal command line program was created along with a system to enable operation via a batch file (figure 3, b&c). R-scripts were generated to automate the use of Zyggregator (figure 3d), Camsoli (figure 1a in electronic supplementary material), iMutant 3.0 (figure 1b in electronic supplementary material), CUPSAT (figure 1c in electronic supplementary material) and Superpose (figure 1d in electronic supplementary material). Data will be available from the authors on request and will be placed in the University of Brighton online repository (research.brighton.ac.uk).

Discussion

The 45 variants that were included in this study represent all point mutations known to be pathogenic for TSD. The separation of the variants into each phenotype and the placement of the phenotypes on a spectrum from severe to least severe is appropriate when considering the description of each clinical case (see table 1). The exclusion criteria omitted those mutations where the cause of reduced HexA activity is obvious and where the approach adopted in this paper would not be appropriate. For example, the most common infantile TSD mutation is a 4-bp insertion causing a frame-shift and early termination resulting in a protein considerably different to wild type (Boles and Proia 1995).

The webservers and software used in this project are manageable when used individually or for small numbers of variants. However, when used in combination the manual entry of data becomes a time-consuming and error prone process. A study using 195 computer literate participants, found that manually entering data and visually checking it

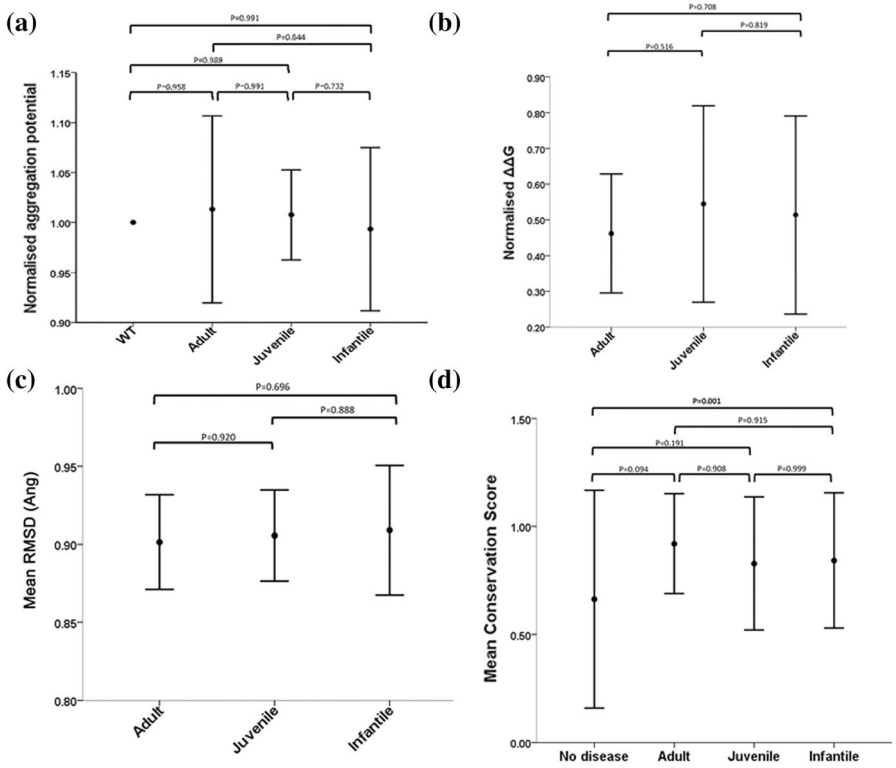

Figure 2. *In silico* analysis of HexA variants associated with TSD. (a) Mean and SD of normalized aggregation potential of the three TSD phenotype groups and wild type calculated from three aggregation webservers (Zyggregator, Camsoli and TANGO). *P* values obtained from Tukey post hoc test. There was no significant difference in aggregation potential between the groups. (b) Mean and SD of normalised ΔΔG of the three TSD phenotype groups calculated from four protein stability web servers (iMutant 3.0, CUPSAT, SDM and mCSM). *P* values obtained from Tukey post hoc test. There was no significant difference in protein stability between the groups. (c) Mean and SD of RMSD of the three TSD phenotype groups calculated from a RMSD web server (Superpose). *P* values obtained from Tukey post hoc test. There was no significant difference in protein RMSD between the groups. (d) Mean and SD for conservation score for disease associated and nondisease associated HexA residues calculated by the Scorecons server using a human HexA sequence aligned with HexA sequences from 74 other species. There is a highly significant difference between no disease and the severe infantile form.

against the source resulted in an error rate nearly three times as high as when using computer-assisted double entry data validation (Barchard and Pace 2011). Given the multiple entry of data was not an option here, the automation of data entry proved an attractive alternative. The scripts and software that we have developed have been shown to work correctly.

Some HexA mutants have been reported to have a propensity to aggregate (Proia and Neufeld 1982; Dersh *et al.* 2016) and it has been shown that aggregation plays a role in other metabolic disorders (Bang *et al.* 2009; McCorvie and Timson 2013). Considering this, we explored whether aggregation plays a role in determining TSD severity. After comparing the mean aggregation scores, there were no significant differences in predicted aggregation between the TSD phenotypes or wild type, suggesting that in any form of TSD, HexA is no more likely to aggregate than wild type and so the loss of HexA function lies elsewhere.

ΔΔG is a useful indicator of protein stability, the higher it is the less stable the mutant protein compared to the wild type (Funahashi *et al.* 2003). Some webservers predicted much larger ΔΔG values than others did, for e.g., Lys197Thr was predicted to have a ΔΔG of –0.94 kcal/mol by iMutant 3.0 but –11.01 kcal/mol by CUPSAT. Considering this, it was necessary to normalize the data with respect to the minimum and maximum values to make it easier to draw comparisons from each server. It is important to note that in doing this we assumed that the servers are equally accurate. From calculating the mean ΔΔG from all four servers and comparing the mean ΔΔG of each phenotype, there was no significant difference between any groups. This suggests that changes in enzyme stability are not responsible for the increased loss of HexA function seen in more severe TSD.

Conservation analysis relies on the principle that in homologous proteins, residues that are more conserved throughout different organisms play a more important role in the function of the enzyme. It has been used to identify the functionally important residues in protein interaction interfaces (Mintseris and Weng 2005), ligand-binding sites (Liang *et al.* 2006) and in maintaining overall protein structure (Valdar and Thornton 2001). In this study, we

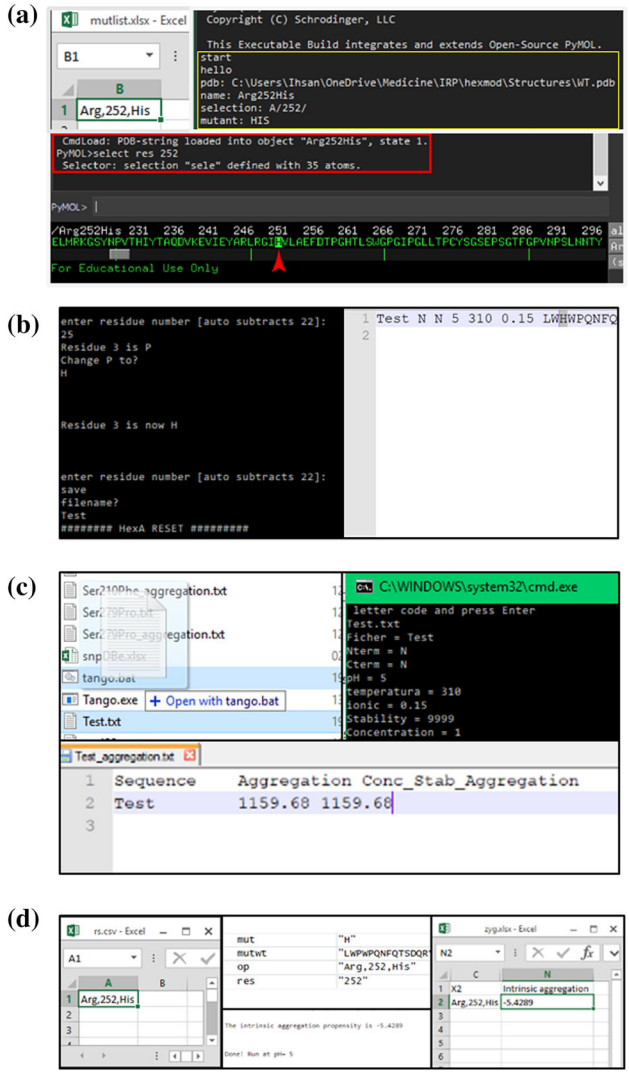

Figure 3. Examples of the implementation of software tools developed in this study. (a) PDB file produced by python script for the mutant Arg252His visualized in PyMol. A command to select residue 252 has been executed (red box) to highlight position 252 in the single code sequence (red arrow). Residue 252 is Histamine rather than the wild type arginine demonstrating the mutation has taken place and the python script functions correctly. (b) Input (left) and output (right) of the pascal program for automating analysis by TANGO. A mutation at position 25 (correlating to position 3 after post-translational cleavage) has been carried out. Position 3 of the output is histidine as expected and the sequence is preceded with the correct parameters for TANGO. (c) Dragging Test.txt onto the batch file (upper left) results in execution of TANGO analysis of the correct file (upper right) and as expected outputs an aggregation score in Test_Aggregation.txt (bottom) matching the output seen when running TANGO manually (not shown). (d) A test input of Arg,252,His (left) is correctly read and split by the R script. This displays variables in memory during execution of the code, note 'res' and 'mut' correctly correspond to the input (upper middle) and the file output from the script (right). The aggregation score is the same as that obtained from the input being entered manually (lower middle).

carried out a multiple protein sequence alignment of human HexA with the enzyme from 73 other species. By comparing the mean Valdar conservation score of disease-associated residues, there was no significant difference between TSD phenotypes. However, when residues associated with TSD phenotypes were compared with nondisease associated residues there was a highly significant difference in conservation (P=0.001) between infantile TSD and the nondisease associated residues. Infantile TSD had a greater mean conservation score than wild type (0.8425±0.1564 vs 0.6635±0.2517). This suggests that the location of the mutation plays an important role in determining the severe loss of HexA function in infantile TSD but not in the other phenotypes.

Considering this in the context of the predictions of protein stability it is unlikely these conserved residues play a role in maintaining stability. Mutations of the active site of HexA are regarded as rare, only seen in the B1 variant (Lemieux *et al.* 2006; Ohno *et al.* 2008). It is more likely that the affected residues are involved in the trafficking of the protein into the lysosome or the interaction between the α-subunits and β-subunits. In fact, Dersh *et al.* (2016) observed that an infantile TSD variant, Glu482Lys, fails to associate with the β-subunit while an adult TSD variant, Gly269Ser, does so correctly. Lemieux *et al.* (2006) reiterate this and additionally state the variant is unable to exit the endoplasmic reticulum. When the HEXB crystal structure was compared with HexA molecular models, the interaction of the α-subunit and β-subunit creates the docking site for the substrate/GM2 activator complex so that a correct protein–protein interface here is critical for enzyme function (Mark *et al.* 2003). Immunoblotting using an anti-α subunit antibody in whole-cell lysates of HEK293T cells expressing either wild type, Glu482Lys (infantile TSD) and Gly269Ser (adult TSD) HexA showed that only the wild type was proteolytically cleaved to the final mature form suggesting either the variants were unable to be cleaved or that they never reached the lysosome. Although these experiments focussed on only two variants, they demonstrate that several factors influence the severity of disease. HexA leaving the endoplasmic reticulum, being trafficked to the lysosome, being cleaved correctly, correct interaction between the α-subunits and β-subunits and having an intact active site capable of interacting with the substrate are all important considerations.

This *in silico* study has established that protein stability and aggregation potential of HexA variants do not appear to play significant roles in determining the severity of TSD. Residue conservation is implicated in infantile TSD. It is possible that these more highly conserved residues are involved in trafficking of the protein to the lysosome but experimental studies are required to confirm this. It appears that TSD has a different underlying molecular pathology compared to a number of other inherited metabolic diseases. A recent study demonstrated that lysosomal proteins, like HexA, are unusually thermally stable compared to those in other cellular compartments (Collier *et al.* 2020). This

suggests that small changes in thermal stability are unlikely to have a significant effect on *in vivo* function.

In several other cases, *in silico* studies have predicted that reduced protein stability is a key factor in the loss of enzymatic activity (Timson 2015). The majority of these studies concerned enzymes which are synthesized and which function in the cellular cytoplasm. In contrast HexA is trafficked to the lysosome, normally a much more acidic environment. This trafficking involves the recognition and cleavage of a signal sequence. These events provide additional steps which are vulnerable to point mutations and structural changes. This study suggests that these steps are the ones critically affected by the majority of mutations associated with TSD.

Acknowledgements

RK was supported by a summer studentship funded by Queen's University, Belfast. This research did not receive any other specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ainsworth P. J. and Coulter-Mackie M. B. 1992 A double mutation in exon 6 of the beta-hexosaminidase alpha subunit in a patient with the B1 variant of Tay-Sachs disease. *Am. J. Hum. Genet.* **51**, 802–809.

Akalin N., Shi H. P., Vavougios G., Hechtman P., Lo W., Scriver C. R. *et al.* 1992 Novel Tay-Sachs disease mutations from China. *Hum. Mutat.* **1**, 40–46.

Akeboshi H., Chiba Y., Kasahara Y., Takashiba M., Takaoka Y. Ohsawa M. *et al*. 2007 Production of recombinant β-hexosaminidase A, a potential enzyme for replacement therapy for Tay-Sachs and Sandhoff diseases, in the methylotrophic yeast Ogataea minuta. *Appl. Environ. Microbiol*. **73**, 4805–4812.

Akli S., Chelly J., Lacorte J. M., Poenaru L. and Kahn A. 1991 Seven novel Tay-Sachs mutations detected by chemical mismatch cleavage of PCR-amplified cDNA fragments. *Genomics* **11**, 124–134.

Akli S., Chomel J. C., Lacorte J. M., Bachner L., Kahn A. and Poenaru L. 1993 Ten novel mutations in the HEXA gene in non-Jewish Tay-Sachs patients. *Hum. Mol. Genet*. **2**, 61–67.

Aruna R. M. and Basu D. 1976 Purification and properties of β-hexosaminidase B from monkey brain. *J. Neurochem*. **27**, 337–339.

Bang Y. L., Nguyen T. T., Trinh T. T., Kim, Y. J., Song J. and Song Y. H. 2009 Functional analysis of mutations in UDP-galactose-4-epimerase GALE associated with galactosemia in Korean patients using mammalian GALE-null cells. *FEBS J*. **276**, 1952–1961.

Barchard K. A. and Pace L. A. 2011 Preventing human error: the impact of data entry methods on data accuracy and statistical results. *Comput. Hum. Behav.* **27**, 1834–1839.

Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. *et al*. 2000 The protein data bank. *Nucleic Acids Res.* **28**, 235–242.

Boles D. J. and Proia R. L. 1995 The molecular basis of HEXA mRNA deficiency caused by the most common Tay-Sachs disease mutation. *Am. J. Hum. Genet*. **56**, 716–724.

Boonyawat B. P., Tim N., Charcrin S. and Suwanpakdee P. 2016 A novel frameshift mutation of HEXA gene in the first family with classical infantile Tay-Sachs disease in Thailand. *Neurol. Asia* **21**, 281–285.

Brown C. A., Neote K., Leung A., Gravel R. A. and Mahuran D. J. 1989 Introduction of the alpha subunit mutation associated with the B1 variant of Tay-Sachs disease into the beta subunit produces a beta-hexosaminidase B without catalytic activity. *J. Biol. Chem.* **264**, 21705–21710.

Browne C. and Timson D. J. 2015 In silico prediction of the effects of mutations in the human mevalonate kinase gene: towards a predictive framework for mevalonate kinase deficiency. *Ann. Hum. Genet.* **79**, 451–459.

Cachon-Gonzalez M. B., Wang S. Z., Lynch A., Ziegler R., Cheng S. H. and Cox T. M. 2006 Effective gene therapy in an authentic model of Tay-Sachs-related diseases. *Proc. Natl. Acad. Sci. USA* **103**, 10373–10378.

Capriotti E., Fariselli P., Rossi I. and Casadio R. 2008 A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics.* **9** Suppl 2, S6.

Coen K., Flannagan R. S., Baron S., Carraro-Lacroix L. R., Wang D., Vermeire W. *et al*. 2012 Lysosomal calcium homeostasis defects, not proton pump defects, cause endo-lysosomal dysfunction in PSEN-deficient cells. *J. Cell Biol.* **198**, 23–35.

Collier A. M., Nemtsova Y., Kuber N., Banach-Petrosky W., Modak A., Sleat D. E. *et al*. 2020 Lysosomal protein thermal stability does not correlate with cellular half-life: global observations and a case study of tripeptidyl-peptidase 1. *Biochem. J.* **477**, 727–745.

De Gasperi R., Gama Sosa M. A., Battistini S., Yeretsian J., Raghavan S., Zelnik N. *et al.* 1996 Late-onset GM2 gangliosidosis: Ashkenazi Jewish family with an exon 5 mutation Tyr180–>His in the Hex A alpha-chain gene. *Neurology* **47**, 547–552.

Dersh D., Iwamoto Y. and Argon Y. 2016 Tay-Sachs disease mutations in HEXA target the alpha chain of hexosaminidase A to endoplasmic reticulum-associated degradation. *Mol. Biol. Cell* **27**, 3813–3827.

Desnick R. J. and Kaback M. M. 2001 Future perspectives for Tay-Sachs disease. *Adv. Genet*. **44**, 349–356.

Dragulescu A. A. 2014 xlsx 0.5.7: read, write, format Excel 2007 and Excel 97/2000/XP/2003 files.

Drucker L., Proia R. L. and Navon R. 1992 Identification and rapid detection of three Tay-Sachs mutations in the Moroccan Jewish population. *Am. J. Hum. Genet.* **51**, 371–377.

Drucker L., Hemli J. A. and Navon R. 1997 Two mutated HEXA alleles in a Druze patient with late-infantile Tay-Sachs disease. *Hum. Mutat*. **10**, 451–457.

Fernandes Filho J. A. and Shapiro B. E. 2004 Tay-Sachs disease. *Arch. Neurol.* **61**, 1466–1468.

Fernandes M., Kaplan F., Natowicz M., Prence E., Kolodny E., Kaback M. *et al.* 1992 A new Tay-Sachs disease B1 allele in exon 7 in two compound heterozygotes each with a second novel mutation. *Hum. Mol. Genet.* **1**, 759–761.

Fernandes M. J., Hechtman P., Boulay B. and Kaplan F. 1997 A chronic GM2 gangliosidosis variant with a HEXA splicing defect: quantitation of HEXA mRNAs in normal and mutant fibroblasts. *Eur. J. Hum. Genet.* **5**, 129–136.

Fernandez-Escamilla A. M., Rousseau F., Schymkowitz J. and Serrano L. 2004 Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306.

Funahashi J., Sugita Y., Kitao A. and Yutani K. 2003 How can free energy component analysis explain the difference in protein stability caused by amino acid substitutions? Effect of three hydrophobic mutations at the 56th residue on the stability of human lysozyme. *Protein Eng.* **16**, 665–671.

Gazoni E. and Clark C. 2018 Openpyxl 2.5.0—a Python library to read/write Excel 2010 xlsx/xlsm files.

Giraud C., Dussau J., Azouguene E., Feillet F., Puech J. P. and Caillaud C. 2010 Rapid identification of HEXA mutations in Tay-Sachs patients. *Biochem. Biophys. Res. Commun.* **392**, 599–602.

Gordon B. A., Gordon K. E., Hinton G. G., Cadera W., Feleki V., Bayleran J. *et al.* 1988 Tay-Sachs disease: B1 variant. *Pediatr. Neurol.* **4**, 54–57.

GraphPad Software Inc. 2017. GraphPad Prism 7.04.

Gray-Edwards H. L., Randle A. N., Maitland S. A., Benatti H. R., Hubbard S. M., Canning P. F. *et al*. 2018 Adeno-associated virus gene therapy in a sheep model of tay-sachs disease. *Hum. Gene Ther.* **29**, 312–326.

Harrison J. 2017 RSelenium 1.71: R Bindings for 'Selenium WebDriver'.

Hayase K. and Kritchevsky D. 1973 Separation and comparison of isoenzymes of N-acetyl-β-D-hexosaminidase of pregnancy serum by polyacrylamide gel electrofocusing. *Int. J. Clin. Chem.* (*Clinica Chimica Acta*) **46**, 455–464.

Henrissat B. and Davies G. 1997 Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644.

Hepbildikler S. T., Sandhoff R., Kolzer M., Proia R. L. and Sandhoff K. 2002 Physiological substrates for human lysosomal β-hexosaminidase S. *J. Biol. Chem*. **277**, 2562–2572.

Ho D. 2018 Notepad++ 7.5.5.

Hou Y., Vavougios G., Hinek A., Wu K. K., Hechtman P., Kaplan F. *et al*. 1996 The Val192Leu mutation in the alpha-subunit of beta-hexosaminidase A is not associated with the B1-variant form of Tay-Sachs disease. *Am. J. Hum. Genet.* **59**, 52–58.

Hurowitz G. I., Silver J. M., Brin M. F., Williams D. T. and Johnson W. G. 1993 Neuropsychiatric aspects of adult-onset Tay-Sachs disease: two case reports with several new findings. *J. Neuropsychiatry Clin. Neurosci.* **5**, 30–36.

Ikonne J. U., Rattazzi M. C. and Desnick R. J. 1975 Characterization of Hex S, the major residual β-hexosaminidase activity in type O Gm2 gangliosidosis Sandhoff-Jatzkewitz disease. *Am. J. Hum. Genet.* **27**, 639–650.

JetBrains 2017 PyCharm.

Karumuthil-Melethil S., Nagabhushan Kalburgi S., Thompson P., Tropak M., Kaytor M. D., Keimel J. G. *et al.* 2016 Novel vector design and hexosaminidase variant enabling self-complementary adeno-associated virus for the treatment of Tay-Sachs disease. *Hum. Gene Ther.* **27**, 509–521.

Kato A., Nakagome I., Nakagawa S., Kinami K., Adachi I., Jenkinson S. F. *et al.* 2017 In silico analyses of essential interactions of iminosugars with the Hex A active site and evaluation of their pharmacological chaperone effects for Tay-Sachs disease. *Org. Biomol. Chem.* **15**, 9297–9304.

Kaufman M., Grinshpun-Cohen J., Karpati M., Peleg L., Goldman B., Akstein E. *et al.* 1997 Tay-Sachs disease and HEXA mutations among Moroccan Jews. *Hum. Mutat.* **10**, 295–300.

Korneluk R. G., Mahuran D. J., Neote K., Klavins M. H., O'Dowd B. F., Tropak M. *et al.* 1986 Isolation of cDNA clones coding for the α-subunit of human β-hexosaminidase. Extensive homology between the α- and β-subunits and studies on Tay-Sachs disease. *J. Biol. Chem*. **261**, 8407–8413.

Krieger E., Joo K., Lee J., Lee J., Raman S., Thompson J. *et al*. 2009 Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* **77** Suppl 9, 114–122.

Kytzia H. J. and Sandhoff K. 1985 Evidence for two different active sites on human β-hexosaminidase A. Interaction of GM2 activator protein with β-hexosaminidase A. *J. Biol. Chem*. **260**, 7568–7572.

Lange P. F., Wartosch L., Jentsch T. J. and Fuhrmann, J. C. 2006 ClC-7 requires Ostm1 as a beta-subunit to support bone resorption and lysosomal function. *Nature* **440**, 220–223.

Lemieux M. J., Mark B. L., Cherney M. M., Withers S. G., Mahuran D. J. and James M. N. 2006 Crystallographic structure of human beta-hexosaminidase A: interpretation of Tay-Sachs mutations and loss of GM2 ganglioside hydrolysis. *J. Mol. Biol.* **359**, 913–929.

Lew R. M., Burnett L., Proos A. L. and Delatycki M. B. 2015 Tay-Sachs disease: current perspectives from Australia. *Appl. Clin. Genet.* **8**, 19–25.

Liang S., Zhang C., Liu S. and Zhou Y. 2006. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res*. **34**, 3698–3707.

Little L. E., Lau M. M., Quon D. V., Fowler A. V. and Neufeld E. F. 1988. Proteolytic processing of the alpha-chain of the lysosomal enzyme, beta-hexosaminidase, in normal human fibroblasts. *J. Biol. Chem.* **263**, 4288–4292.

Maegawa G. H., Stockley T., Tropak M., Banwell B., Blaser S., Kok F. *et al.* 2006 The natural history of juvenile or subacute GM2 gangliosidosis: 21 new cases and literature review of 134 previously reported. *Pediatrics* **118**, e1550–e15562.

Maegawa G. H., Tropak M., Buttner J., Stockley T., Kok F., Clarke J. T. *et al*. 2007 Pyrimethamine as a potential pharmacological chaperone for late-onset forms of GM2 gangliosidosis. *J. Biol. Chem*. **282**, 9150–9161.

Maiti R., Van Domselaar G. H., Zhang H. and Wishart D. S. 2004 SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res.* **32**, W590–W5904.

Mark B. L., Mahuran D. J., Cherney M. M., Zhao D., Knapp S. and James M. N. 2003 Crystal structure of human beta-hexosaminidase B: understanding the molecular basis of Sandhoff and Tay-Sachs disease. *J. Mol. Biol.* **327**, 1093–1109.

Mark B. L., Vocadlo D. J., Knapp S., Triggs-Raine B. L., Withers S. G. and James M. N. 2001 Crystallographic evidence for substrate-assisted catalysis in a bacterial β-hexosaminidase. *J. Biol. Chem*. **276**, 10330–10337.

Masingue M., Dufour L., Lenglet T., Saleille L., Goizet C., Ayrignac X. *et al*. 2020 Natural history of adult patients with gm2 gangliosidosis. *Annl. Neurol*. https://doi.org/10.1002/ana.25689.

Matsuzawa F., Aikawa S., Sakuraba H., Lan H. T., Tanaka A., Ohno K. *et al*. 2003 Structural basis of the GM2 gangliosidosis B variant. *J. Hum. Genet.* **48**, 582–589.

McCorvie T. J. and Timson D. J. 2013 In silico prediction of the effects of mutations in the human UDP-galactose 4'-epimerase gene: towards a predictive framework for type III galactosemia. *Gene* **524**, 95–104.

Mintseris J. and Weng Z. 2005 Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **102**, 10930–10935.

Mistri M., Tamhankar P. M., Sheth F., Sanghavi D., Kondurkar P., Patil S. *et al*. 2012 Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India. *PLoS One* **7**, e39122.

Montalvo A. L., Filocamo M., Vlahovicek K., Dardis A., Lualdi S., Corsolini F. *et al.* 2005 Molecular analysis of the HEXA gene in Italian patients with infantile and late onset Tay-Sachs disease: detection of fourteen novel alleles. *Hum. Mutat*. **26**, 282.

Mules E. H., Hayflick S., Miller C. S., Reynolds L. W. and Thomas G. H. 1992 Six novel deleterious and three neutral mutations in the gene encoding the alpha-subunit of hexosaminidase A in non-Jewish individuals. *Am. J. Hum. Genet.* **50**, 834–841.

Najmabadi H., Hu H., Garshasbi M., Zemojtel T., Abedini S. S., Chen W. *et al*. 2011 Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* **478**, 57–63.

Nakano T., Muscillo M., Ohno K., Hoffman A. J. and Suzuki K. 1988 A point mutation in the coding sequence of the beta-hexosaminidase alpha gene results in defective processing of the enzyme protein in an unusual GM2-gangliosidosis variant. *J. Neurochem.* **51**, 984–987.

Nakano T., Nanba E., Tanaka A., Ohno K., Suzuki Y. and Suzuki K. 1990 A new point mutation within exon 5 of beta-hexosaminidase alpha gene in a Japanese infant with Tay-Sachs disease. *Annl. Neurol.* **27**, 465–473.

Navon R. and Proia R. L. 1989 The mutations in Ashkenazi Jews with adult GM2 gangliosidosis, the adult form of Tay-Sachs disease. *Science* **243**, 1471–1474.

Navon R., Khosravi R., Korczyn T., Masson M., Sonnino S., Fardeau M. *et al.* 1995 A new mutation in the HEXA gene associated with a spinal muscular atrophy phenotype. *Neurology* **45**, 539–543.

Neote K., Bapat B., Dumbrille-Ross A., Troxel C., Schuster S. M., Mahuran D. J. *et al*. 1988 Characterization of the human HEXB gene encoding lysosomal β-hexosaminidase. *Genomics* **3**, 279–286.

Neudorfer O., Pastores G. M., Zeng B. J., Gianutsos J., Zaroff C. M. and Kolodny E. H. 2005 Late-onset Tay-Sachs disease: phenotypic characterization and genotypic correlations in 21 affected patients. *Genet. Med.* **7**, 119–123.

Ohkuma S. and Poole B. 1978 Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proc. Natl. Acad. Sci. USA* **75**, 3327–3331.

Ohno K., Saito S., Sugawara K. and Sakuraba H. 2008 Structural consequences of amino acid substitutions causing Tay-Sachs disease. *Mol. Genet. Metab.* **94**, 462–468.

Oliver C. and Timson D. J. 2017 In silico prediction of the effects of mutations in the human triose phosphate isomerase gene: towards a predictive framework for TPI deficiency. *Eur. J. Med. Genet.* **60**, 289–298.

Ornaghi F., Sala D., Tedeschi F., Maffia M. C., Bazzucchi M., Morena F. *et al*. 2020 Novel bicistronic lentiviral vectors correct β-Hexosaminidase deficiency in neural and hematopoietic stem cells and progeny: implications for in vivo and ex vivo gene therapy of GM2 gangliosidosis. *Neurobiol. Dis*. **134**, 104667.

Ou L., Kim S., Whitley C. B. and Jarnes-Utz J. R. 2019 Genotype-phenotype correlation of gangliosidosis mutations using in silico tools and homology modeling. *Mol. Genet. Metab. Rep.* **20**, 100495.

Ou L., Przybilla M. J., Tăbăran A.-F., Overn P., O'Sullivan M. G., Jiang X. *et al*. 2020 A novel gene editing system to treat both Tay-Sachs and Sandhoff diseases. *Gene Ther.* https://doi.org/10.1038/s41434-019-0120-5.

Pandurangan A. P., Ochoa-Montano B., Ascher D. B. and Blundell T. L. 2017 SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* **45**, W229–W235.

Parthiban V., Gromiha M. M. and Schomburg D. 2006 CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* **34**, W239–W242.

Passos O., Fernandes P. A. and Ramos M. J. 2011 QM/MM study of the catalytic mechanism of GalNAc removal from GM2 ganglioside catalyzed by human β-hexosaminidaseA. *J. Phys. Chem*. B **115**, 14751–14759.

Paw B. H., Moskowitz S. M., Uhrhammer N., Wright N., Kaback M. M. and Neufeld E. F. 1990 Juvenile GM2 gangliosidosis caused by substitution of histidine for arginine at position 499 or 504 of the alpha-subunit of beta-hexosaminidase. *J. Biol. Chem*. **265**, 9452–9457.

Petroulakis E., Cao Z., Clarke J. T., Mahuran D. J., Lee G. and Triggs-Raine B. 1998 W474C amino acid substitution affects early processing of the alpha-subunit of beta-hexosaminidase A and is associated with subacute G M2 gangliosidosis. *Hum. Mutat.* **11**, 432–442.

Pires D. E., Ascher D. B. and Blundell T. L. 2014 mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342.

Platt F. M., Jeyakumar M., Andersson U., Priestman D. A., Dwek R. A., Butters T. D. *et al*. 2001 Inhibition of substrate synthesis as a strategy for glycolipid lysosomal storage disease therapy. *J. Inher. Metab. Dis.* **24**, 275–290.

Proia R. L. and Neufeld E. F. 1982 Synthesis of beta-hexosaminidase in cell-free translation and in intact fibroblasts: an insoluble precursor alpha chain in a rare form of Tay-Sachs disease. *Proc. Natl. Acad. Sci. USA* **79**, 6360–6364.

Pundir S., Martin M. J. and O'Donovan C. 2017 UniProt protein knowledgebase. *Methods Mol. Biol.* **1558**, 41–55.

R-Project. 2018 grep—pattern matching and replacement.

Raghavan S. S., Krusell A., Krusell J., Lyerla T. A. and Kolodny E. H. 1985 GM2-ganglioside metabolism in hexosaminidase A deficiency states: determination in situ using labeled GM2 added to fibroblast cultures. *Am. J. Hum. Genet.* **37**, 1071–1082.

Ribeiro M. G., Sonin T., Pinto R. A., Fontes A., Ribeiro H., Pinto E. *et al*. 1996 Clinical, enzymatic, and molecular characterisation of a Portuguese family with a chronic form of GM2-gangliosidosis B1 variant. *J. Med. Genet.* **33**, 341–343.

Robinson D. and Stirling J. L. 1968 N-Acetyl-β-glucosaminidases in human spleen. *Biochem. J.* **107**, 321–327.

Rountree J. S., Butters T. D., Wormald M. R., Boomkamp S. D., Dwek R. A., Asano N. *et al*. 2009 Design, synthesis, and biological evaluation of enantiomeric β-N-acetylhexosaminidase inhibitors LABNAc and DABNAc as potential agents against Tay-Sachs and Sandhoff disease. *Chem. Med. Chem.* **4**, 378–392.

Schrödinger L. 2018 The PyMOL molecular graphics system, version 2.0

Shapiro B. E. and Natowicz M. R. 2009 Late-onset Tay-Sachs disease presenting as a childhood stutter. *J. Neurol. Neurosurg. Psychiatry* **80**, 94–95.

Sharma R., Bukovac S., Callahan J. and Mahuran D. 2003 A single site in human beta-hexosaminidase A binds both 6-sulfate-groups on hexosamines and the sialic acid moiety of GM2 ganglioside. *Biochim. Biophys. Acta* **1637**, 113–118.

Sheth J., Mistri M., Sheth F., Shah R., Bavdekar A., Godbole K. *et al*. 2014 Burden of lysosomal storage disorders in India: experience of 387 affected children from a single diagnostic facility. *JIMD Rep*. **12**, 51–63.

Sievers F., Wilm A., Dineen D., Gibson T. J., Karplus K., Li W. *et al*. 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

Sormanni P., Aprile F. A. and Vendruscolo M. 2015 The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478–490.

Specola N., Vanier M. T., Goutieres F., Mikol J. and Aicardi J. 1990 The juvenile and chronic forms of GM2 gangliosidosis: clinical and enzymatic heterogeneity. *Neurology* **40**, 145–150.

Steiner K. M., Brenck J., Goericke S. and Timmann D. 2016 Cerebellar atrophy and muscle weakness: late-onset Tay-Sachs disease outside Jewish populations. *BMJ Case Rep.* **2016**, bcr2016214634.

Svennerholm L. and Fredman P. 1980 A procedure for the quantitative isolation of brain gangliosides. *Biochim. Biophys. Acta* **617**, 97–109.

Tabeta K., Hoebe K., Janssen E. M., Du X., Georgel P., Crozat K. *et al.* 2006 The Unc93b1 mutation 3d disrupts exogenous antigen presentation and signaling via Toll-like receptors 3, 7 and 9. *Nat. Immunol*. **7**, 156–164.

Tanaka A., Hoang L. T., Nishi Y., Maniwa S., Oka M. and Yamano T. 2003 Different attenuated phenotypes of GM2 gangliosidosis variant B in Japanese patients with HEXA mutations at codon 499, and five novel mutations responsible for infantile acute form. *J. Hum. Genet.* **48**, 571–574.

Tanaka A., Ohno K., Sandhoff K., Maire I., Kolodny E. H., Brown A. *et al*. 1990a GM2-gangliosidosis B1 variant: analysis of beta-

hexosaminidase alpha gene abnormalities in seven patients. *Am. J. Hum. Genet*. **46**, 329–339.

Tanaka A., Punnett H. H. and Suzuki K. 1990b A new point mutation in the beta-hexosaminidase alpha subunit gene responsible for infantile Tay-Sachs disease in a non-Jewish Caucasian patient a Kpn mutant. *Am. J. Hum. Genet.* **47**, 568–574.

Tanaka A., Sakazaki H., Murakami H., Isshiki G. and Suzuki K. 1994 Molecular genetics of Tay-Sachs disease in Japan. *J. Inherit. Metab. Dis*. **17**, 593–600.

Tartaglia G. G. and Vendruscolo M. 2008 The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395–1401.

Team L. 2018 Lazarus v1.82: The professional Free Pascal RAD IDE.

Team R. 2016 RStudio: integrated development environment for R.

Tettamanti G. 2004 Ganglioside/glycosphingolipid turnover: new concepts. *Glycoconj J*. **20**, 301–317.

Timson D. J. 2015 Value of predictive bioinformatics in inherited metabolic diseases. *World J. Med. Genet*. **5**, 46–51.

Triggs-Raine B. L., Akerman B. R., Clarke J. T. and Gravel R. A. 1991 Sequence of DNA flanking the exons of the HEXA gene, and identification of mutations in Tay-Sachs disease. *Am. J. Hum. Genet*. **49**, 1041–1054.

Trop I., Kaplan F., Brown C., Mahuran D. and Hechtman P. 1992 A glycine250–> aspartate substitution in the alpha-subunit of hexosaminidase A causes juvenile-onset Tay-Sachs disease in a Lebanese-Canadian family. *Hum. Mutat*. **1**, 35–39.

Tropak M. B., Reid S. P., Guiral M., Withers S. G. and Mahuran D. 2004 Pharmacological enhancement of β-hexosaminidase activity in fibroblasts from adult Tay-Sachs and Sandhoff Patients. *J. Biol. Chem*. **279**, 13478–13487.

Tropak M. B., Bukovac S. W., Rigat B. A., Yonekawa S., Wakarchuk W. and Mahuran D. J. 2010 A sensitive fluorescence-based assay for monitoring GM2 ganglioside hydrolysis in live patient cells and their lysates. *Glycobiology* **20**, 356–365.

Tsuji D., Akeboshi H., Matsuoka K., Yasuoka H., Miyasaki E., Kasahara Y. *et al.* 2011 Highly phosphomannosylated enzyme replacement therapy for GM2 gangliosidosis. *Annl. Neurol.* **69**, 691–701.

UniProt Consortium T. 2018 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699.

Valdar W. S. 2002 Scoring residue conservation. *Proteins* **48**, 227–241.

Valdar W. S. and Thornton J. M. 2001 Conservation helps to identify biologically relevant crystal contacts. *J. Mol. Biol.* **313**, 399–416.

Zampieri S., Montalvo A., Blanco M., Zanin I., Amartino H., Vlahovicek K. *et al.* 2012 Molecular analysis of HEXA gene in Argentinean patients affected with Tay-Sachs disease: possible common origin of the prevalent c.459+5A>G mutation. *Gene* **499**, 262–265.

Corresponding editor: H. A. Ranganath