World Journal of Medical Genetics

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.5496/wjmg.v5.i3.46

World J Med Genet 2015 August 27; 5(3): 46-51
ISSN 2220-3184 (online)

EDITORIAL

Value of predictive bioinformatics in inherited metabolic diseases

David J Timson

David J Timson, School of Biological Sciences and Institute for Global Food Security, Queen's University Belfast, BT9 7BL Belfast, United Kingdom

Author contributions: Timson DJ conceived and wrote the paper

Conflict-of-interest statement: The author has no conflicts of interest to declare.

Correspondence to: Dr. David J Timson, School of Biological Sciences and Institute for Global Food Security, Queen's University Belfast, 97 Lisburn Road, BT9 7BL Belfast, United Kingdom. d.timson@qub.ac.uk
Telephone: +44-028-90975875
Fax: +44-028-90975877

Received: February 27, 2015
Peer-review started: March 2, 2015
First decision: April 27, 2015
Revised: April 28, 2015
Accepted: May 16, 2015
Article in press: May 18, 2015
Published online: August 27, 2015

Abstract

Typically, inherited metabolic diseases arise from point mutations in genes encoding metabolic enzymes. Although some of these mutations directly affect amino acid residues in the active sites of these enzymes, the majority do not. It is now well accepted that the majority of these disease-associated mutations exert their effects through alteration of protein stability, which causes a reduction in enzymatic activity. This finding suggests a way to predict the severity of newly discovered mutations. *In silico* prediction of the effects of amino acid sequence alterations on protein stability often correlates with disease severity. However, no stability prediction tool is perfect and, in general, better results are obtained if the predictions from a variety of tools are combined and then interpreted. In addition to predicted alterations to stability, the degree of conservation of a particular residue can also be a factor which needs to be taken into account: alterations to highly conserved residues are more likely to be associated with severe forms of the disease. The approach has been successfully applied in a variety of inherited metabolic diseases, but further improvements are necessary to enable robust translation into clinically useful tools.

Key words: Genetic disease; Metabolism; *In silico* method; Protein stability; Disease-associated mutation

Core tip: Bioinformatics and other *in silico* methods are increasingly being used to predict the severity of disease-associated mutations in inherited metabolic diseases. In general, severity correlates with altered protein stability and the best predictions occur when a variety of tools are applied.

Timson DJ. Value of predictive bioinformatics in inherited metabolic diseases. *World J Med Genet* 2015; 5(3): 46-51 Available from: URL: http://www.wjgnet.com/2220-3184/full/v5/i3/46.htm DOI: http://dx.doi.org/10.5496/wjmg.v5.i3.46

INTRODUCTION

Inherited metabolic diseases result from mutations in the genes encoding enzymes involved in intermediary metabolism. Well characterised examples include galactosemia, lysosomal storage diseases and phenylketonuria. Typically these diseases manifest with effects at the whole organism level, despite their origins at the metabolic pathway level. Physical and cognitive disabilities are associated with many inherited metabolic diseases. While individual diseases are generally rare, the cumulative effect of many of these diseases has a significant effect on societies and economies[1-4]. Furthermore, the burdens on patients, their families and their immediate communities can be devastating since many of these diseases result in progressive deterioration of the patient resulting, in some cases, in death in childhood or early adulthood. Very few of these diseases have effective therapies (*i.e.*, treatments which restore normal, or near-normal, functioning to the patient). One barrier to the development of therapies is the rareness of the diseases: there is limited incentive to the development of drugs or other treatments which would only be applicable to a small number of patients worldwide[2,5,6]. Where therapies do exist, they tend to be extremely expensive, often exceeding United States $100000 per patient per year (for example, see[7,8]).

Biochemical studies on the underlying molecular pathology of a range of inherited metabolic diseases have revealed some common themes. In particular, mutations associated with these diseases often cause changes which destabilise the corresponding protein (for examples, see[9-15]). Very few disease-associated mutations directly affect the residues in the active site of the enzyme; the majority affect residues elsewhere in the protein. A common molecular mechanism of disease causation is that the altered amino acid residue causes a global reduction in the enzyme's stability resulting in reduced catalytic activity[16]. The loss of stability can also be associated with reduced affinity for essential cofactors or increased aggregation of the partially folded protein. It is, of course, the loss of enzymatic activity which commonly leads to disease, for example by reducing the amount of product made or causing a build-up of toxic intermediates. In other cases the accumulation of aggregated protein results in a breakdown of cellular homeostasis. Nevertheless, partial protein misfolding lies at the base of these problems and is the fundamental cause of the disease in these cases.

It is also apparent that, in many inherited metabolic diseases, there is a range of possible symptoms. This is particularly stark in diseases like type Ⅲ galactosemia and mevalonate kinase deficiency. In these diseases the manifestations range from near-normal physiology with some alterations in blood chemistry to highly disabling, life-threatening conditions[17,18]. The experience of each patient will be determined by his/her genetic background, lifestyle and environment. Critical elements include the patient's diet, activity levels and access to good quality medical care. However, the most important factor in determining the severity of symptoms is normally the exact mutation(s) that the patient has. Most inherited metabolic diseases are not caused by one, single mutation. The majority have many possible mutations which are associated with the disease. For example, there are almost 250 mutations in galactose 1-phosphate uridylyltransferase which are associated with type Ⅰ galactosemia[19,20]. Since these mutations alter different amino acid residues, it follows that they will have different effects on the protein. Some will have relatively minor effects on the protein's overall structure and stability whereas others may render the protein essentially non-functional.

Novel, disease-associated mutations continue to be discovered. Indeed, with the decreasing price of whole exome sequencing, we should expect that the rate of discovery of novel mutations will increase in the next few years[21,22]. In some inherited metabolic diseases problems are apparent within a few days of birth; however, in other cases, babies are born with near normal physiology but progressively decline over the following years. Given the range of possible severities associated with some inherited metabolic diseases it is a challenge to physicians and scientists to predict the likely symptoms of an individual patient and to plan treatment accordingly. This is particularly the case for newly discovered disease-associated mutations.

THE CHALLENGE OF PREDICTION

The link between protein stability and severity of disease suggests a way in which predictions might be made. There are a variety of software packages, many freely available online, which claim to predict the stability of proteins (Table 1). In theory if a range of disease-associated variant proteins are compared to the wild-type in one of these packages then the greater the predicted instability, the more severe the disease is likely to be. In practice the situation is more complex. No prediction software is 100% accurate and different packages can give different results for the same variant. This problem can be overcome by using a variety of different programs and aggregating the results together to obtain a consensus. However, problems still remain. In some diseases, both decreases and increases in stability can be associated with disease (for examples see[9,23]). Indeed, it appears that as well as an optimum structure for activity, proteins also require an optimal degree of flexibility and stability. Protein flexibility is inextricably linked to ligand binding and catalysis[24]. Thus, increased stability can lead to a more rigidified, less flexible protein which is less able to bind substrates and catalyse the reaction.

While protein stability is undoubtedly a key factor in disease causation, there are other factors which need to be considered. Alterations in key residues involved in binding substrates or in the chemistry of catalysis will

Table 1 Examples of freely available, online tools for predicting the properties of variant proteins

Category	Name	Weblink	Ref.
Structural analysis	YASARA energy minimisation	www.yasara.org/minimizationserver.htm	[29]
	LS-SNP	ls-snp.icm.jhu.edu/ls-snp-pdb/main	[30]
	GETAREA	curie.utmb.edu/getarea.html	[31]
Stability prediction	I-Mutant 3.0	gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi	[32,33]
	mCSM	bleoberis.bioc.cam.ac.uk/mcsm/	[34]
	SDM score	mordred.bioc.cam.ac.uk/~sdm/sdm.php	[35,36]
	Mupro	mupro.proteomics.ics.uci.edu	[37]
	iStable	predictor.nchu.edu.tw/iStable/	[38]
	PredictSNP 1.0	loschmidt.chemi.muni.cz/predictsnp/	[39]
	Meta-SNP	snps.biofold.org/meta-snp/	[40]
	KD4V	decrypthon.igbmc.fr/kd4v	[41]
	Fold-X	foldx.crg.es	[42]
	PoPMuSiC	dezyme.com/	[43]
	CUPSAT	cupsat.tu-bs.de	[44,45]
	GETAREA	curie.utmb.edu/getarea.html	[31]
Binding affinity changes	BeAtMuSiC	babylone.ulb.ac.be/beatmusic	[46]
Aggregation tendency, amyloid formation and chaperone binding	TANGO	tango.crg.es/	[47]
	WALTZ	www.switchlab.org/bioinformatics/waltz	[48]
	LIMBO	www.switchlab.org/bioinformatics/limbo	[49]
Sequence conservation	Clustal Omega	www.ebi.ac.uk/Tools/msa/clustalo/	[50]
	Scorecons	www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl	[51]
	SIFT	sift.jcvi.org/	[52]
	PROVEAN	provean.jcvi.org/index.php	[53]
	LS-SNP	ls-snp.icm.jhu.edu/ls-snp-pdb/	[30]
	SNPs and GO	snps.biofold.org/snps-and-go/pages/help.html	[54]
	PANTHER	www.pantherdb.org/tools/csnpScoreForm.jsp	[55]
	GenMAPP	www.genmapp.org	[56]
	PolyPhen 2	genetics.bwh.harvard.edu/pph2/	[57,58]
	nsSNP Analyzer	snpanalyzer.uthsc.edu	[59]
	FI mutation assessor	mutationassessor.org/v1	[60]
	YALE MU2A	krauthammerlab.med.yale.edu/mu2a	[61]

lead to direct loss of activity. A failure to interact with cellular chaperones may impede folding. Disruptions to other protein-protein interactions may also affect the enzyme's function. Residues involved in catalysis and protein-protein interactions are generally well conserved through evolution. Therefore, we might expect that mutations which alter highly conserved residues might also lead to more severe forms of disease. Therefore, many predictions incorporate measures of sequence conservation and propensity to interact with cellular chaperones (Table 1). Overall, it is accepted that the best predictions will result from using a variety of different software packages which address different aspects of the protein's structure and function[25-27]. Furthermore, any output requires intelligent and critical analysis by the users.

APPLICATION TO INHERITED METABOLIC DISEASES

These approaches have been employed in a number of inherited metabolic diseases (Table 2). Typically, a set of known mutations and their associated protein variants are identified from the literature and classified according to their association with different severities of the disease. Other information from the literature is required - most importantly an experimental demonstration that protein misfolding is an important factor in disease causation. Using the known variants, a range of prediction tools are applied and the combination which best predicts the known outcomes are then selected. This can then be applied to uncharacterised mutations or to polymorphisms identified through genome and exome sequencing projects. In general, the severity of disease correlates with the predicted loss of stability of the protein. The degree of conservation of the residue(s) affected is also important in some conditions (Table 2). Most studies employ a range of different prediction tools and aggregate results together to make informed predictions (for example see[28]).

CONCLUSION AND FUTURE PERSPECTIVES

To date, no prediction protocol has achieved complete accuracy and it is unlikely that physicians would be confident to rely on them to guide treatment of their patients. In addition, the prediction protocols published so far mostly require extensive bioinformatics analysis using a number of different tools on separate websites. Ideally these would be integrated into a single web-based package which enabled the user to submit a

Table 2 Examples of bioinformatics based predictions of the severity of variants associated with inherited metabolic diseases

Disease	Protein	Comments	Ref.
Alkaptonuria	Homogentisate 1,2-dioxygenase	Combining a variety of computational approaches gave rise to the most accurate predictions	[62]
Apparent mineralocorticoid excess	11βHSD2	The predicted degree of structural change in the enzyme correlates with disease severity	[63]
Fabry disease	GLA	A purpose built program designed to detect protein instability outperformed existing, generic tools	[64]
Fabry disease	GLA	A purpose built web interface allows prediction of a patient's responsiveness to pharmacological chaperone therapy	[65]
Gaucher disease	GBA	Slightly different results were obtained with different programs; however, 22 out of 47 variants were predicted to be harmful by all seven programs used	[28]
Glucose 6-phosphate dehydrogenase deficiency	G6PDH	A combination of prediction tools suggested that protein stability is an important factor in this disease; novel potentially disease-associated variants were identified	[66]
Hyperargininemia	ARG1	Mutations affect residues in the active site, or protein stability, or quaternary structure	[67]
MODY 2	GCK	Variations which decrease protein stability and/or occur in highly conserved regions of the protein are associated with disease	[68]
Niemann-pick disease type C	NPC1 and NPC2	The majority of disease-associated variants were predicted to be less stable than wild-type	[69]
Phenylketonuria	PAH	Protein stability predicted to be most important factor in disease causation	[10]
Pyruvate kinase deficiency	PK1 and PK2	A combination of prediction tools suggested that protein stability is an important factor in this disease; novel potentially disease-associated variants were identified	[66]
Type Ⅰ galactosemia	GALT	Main predicted effect is the loss of stability of GALT	[70]
Type Ⅲ galactosemia	GALE	Effects on protein stability and degree of sequence conservation combined were required for good predictions	[71]

11βHSD2: 11β-hydroxysteroid dehydrogenase type 2; GLA: α-galactosidase A; GBA: Glucocerebrosidase; G6PDH: Glucose 6-phosphate dehydrogenase; ARG1: Arginase 1; GCK: Glucokinase; PAH: Phenylalanine hydroxylase; PK1 and PK2: Pyruvate kinase isoforms 1 and 2; GALT: Galactose 1-phosphate uridylyltransferase; GALE: UDP-galactose 4'-epimerase; MODY 2: Maturity-onset diabetes of the young, type 2.

novel variant and receive a prediction. This would require the software to submit the new variant to the various online tools and integrate the responses into a single prediction. This, in turn, requires the software tools to remain "live" and at the same, stable url - something which cannot be guaranteed when each tool is provided by different organisations. One solution would be to use the source code for the prediction tools directly, but this is not always freely available.

Nevertheless carefully designed prediction methods can already provide some guidance in the assessment of novel mutations and have also had some success in classifying mutations identified through genome and exome sequencing projects. Interestingly, these tend to suggest that there are a number of mutations in the human population which were not previously associated with disease. In the case of type Ⅲ galactosemia, several of these were predicted to be associated with severe forms of the disease. Further experimental testing will be required to see if these predictions are valid. In conclusion, the best predictions will be made when a variety of tools are used and where they are supported by strong experimental and clinical studies. It is highly unlikely that bioinformatics predictions will completely supplant experimental studies; however, with continuing improvements in the tools used they have the potential to be a useful tool for both clinicians and scientists who work on inherited metabolic diseases.

ACKNOWLEDGMENTS

I wish to thank Dr. Thomas McCorvie (University of Oxford, United Kingdom) for introducing me to this field.

REFERENCES

1 **Angelis A**, Tordrup D, Kanavos P. Socio-economic burden of rare diseases: A systematic review of cost of illness evidence. *Health Policy* 2015; **119**: 964-979 [PMID: 25661982]

2 **Kontoghiorghe CN**, Andreou N, Constantinou K, Kontoghiorghes GJ. World health dilemmas: Orphan and rare diseases, orphan drugs and orphan patients. *World J Methodol* 2014; **4**: 163-188 [PMID: 25332915 DOI: 10.5662/wjm.v4.i3.163]

3 **López-Bastida J**, Oliva-Moreno J. Cost of illness and economic evaluation in rare diseases. *Adv Exp Med Biol* 2010; **686**: 273-282 [PMID: 20824451 DOI: 10.1007/978-90-481-9485-8_16]

4 **Pampols T**. Inherited metabolic rare disease. *Adv Exp Med Biol* 2010; **686**: 397-431 [PMID: 20824458 DOI: 10.1007/978-90-481-9485-8_23]

5 **Denis A**, Mergaert L, Fostier C, Cleemput I, Simoens S. A comparative study of European rare disease and orphan drug markets. *Health Policy* 2010; **97**: 173-179 [PMID: 20800761 DOI: 10.1016/j.healthpol.2010.05.017]

6 **Heemstra HE**, van Weely S, Büller HA, Leufkens HG, de Vrueh RL. Translation of rare disease research into orphan drug development: disease matters. *Drug Discov Today* 2009; **14**: 1166-1173 [PMID: 19818412 DOI: 10.1016/j.drudis.2009.09.008]

7 **Connock M**, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, Wailoo A, Abrams K, Cooper N, Sutton A, O'Hagan A, Moore D. The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review. *Health Technol Assess* 2006; **10**: iii-iv, ix-136 [PMID: 16796930 DOI: 10.3310/hta10240]

8 **Kanters TA**, Hoogenboom-Plug I, Rutten-Van Mölken MP, Redekop WK, van der Ploeg AT, Hakkaart L. Cost-effectiveness of enzyme replacement therapy with alglucosidase alfa in classic-infantile patients with Pompe disease. *Orphanet J Rare Dis* 2014; **9**: 75 [PMID: 24884717 DOI: 10.1186/1750-1172-9-75]

9 **McCorvie TJ**, Gleason TJ, Fridovich-Keil JL, Timson DJ. Misfolding of galactose 1-phosphate uridylyltransferase can result in

type I galactosemia. *Biochim Biophys Acta* 2013; **1832**: 1279-1293 [PMID: 23583749 DOI: 10.1016/j.bbadis.2013.04.004]
10 **Pey AL**, Stricher F, Serrano L, Martinez A. Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases. *Am J Hum Genet* 2007; **81**: 1006-1024 [PMID: 17924342 DOI: 10.1086/521879]
11 **Shi Z**, Sellers J, Moult J. Protein stability and in vivo concentration of missense mutations in phenylalanine hydroxylase. *Proteins* 2012; **80**: 61-70 [PMID: 21953985 DOI: 10.1002/prot.23159]
12 **Pey AL**, Maggi M, Valentini G. Insights into human phosphoglycerate kinase 1 deficiency as a conformational disease from biochemical, biophysical, and in vitro expression analyses. *J Inherit Metab Dis* 2014; **37**: 909-916 [PMID: 24838780 DOI: 10.1007/s10545-014-9721-8]
13 **Pey AL**, Padín-Gonzalez E, Mesa-Torres N, Timson DJ. The metastability of human UDP-galactose 4'-epimerase (GALE) is increased by variants associated with type III galactosemia but decreased by substrate and cofactor binding. *Arch Biochem Biophys* 2014; **562**: 103-114 [PMID: 25150110]
14 **Mesa-Torres N**, Fabelo-Rosa I, Riverol D, Yunta C, Albert A, Salido E, Pey AL. The role of protein denaturation energetics and molecular chaperones in the aggregation and mistargeting of mutants causing primary hyperoxaluria type I. *PLoS One* 2013; **8**: e71963 [PMID: 24205397 DOI: 10.1371/journal.pone.0071963]
15 **Henriques BJ**, Bross P, Gomes CM. Mutational hotspots in electron transfer flavoprotein underlie defective folding and function in multiple acyl-CoA dehydrogenase deficiency. *Biochim Biophys Acta* 2010; **1802**: 1070-1077 [PMID: 20674745 DOI: 10.1016/j.bbadis.2010.07.015]
16 **Yue P**, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol* 2005; **353**: 459-473 [PMID: 16169011 DOI: 10.1016/j.jmb.2005.08.020]
17 **Timson DJ**. The structural and molecular biology of type III galactosemia. *IUBMB Life* 2006; **58**: 83-89 [PMID: 16611573]
18 **van der Burgh R**, Ter Haar NM, Boes ML, Frenkel J. Mevalonate kinase deficiency, a metabolic autoinflammatory disease. *Clin Immunol* 2013; **147**: 197-206 [PMID: 23110805 DOI: 10.1016/j.clim.2012.09.011]
19 **d'Acierno A**, Facchiano A, Marabotti A. GALT protein database, a bioinformatics resource for the management and analysis of structural features of a galactosemia-related protein and its mutants. *Genomics Proteomics Bioinformatics* 2009; **7**: 71-76 [PMID: 19591794 DOI: 10.1016/S1672-0229(08)60035-2]
20 **d'Acierno A**, Facchiano A, Marabotti A. GALT protein database: querying structural and functional features of GALT enzyme. *Hum Mutat* 2014; **35**: 1060-1067 [PMID: 24990533 DOI: 10.1002/humu.22613]
21 **Boycott KM**, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013; **14**: 681-691 [PMID: 23999272 DOI: 10.1038/nrg3555]
22 **Danielsson K**, Mun LJ, Lordemann A, Mao J, Lin CH. Next-generation sequencing applied to rare diseases genomics. *Expert Rev Mol Diagn* 2014; **14**: 469-487 [PMID: 24702023 DOI: 10.1586/14737159.2014.904749]
23 **Wohlers TM**, Fridovich-Keil JL. Studies of the V94M-substituted human UDPgalactose-4-epimerase enzyme associated with generalized epimerase-deficiency galactosaemia. *J Inherit Metab Dis* 2000; **23**: 713-729 [PMID: 11117433 DOI: 10.1023/A:1005682913784]
24 **Goodey NM**, Benkovic SJ. Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 2008; **4**: 474-482 [PMID: 18641628 DOI: 10.1038/nchembio.98]
25 **Frédéric MY**, Lalande M, Boileau C, Hamroun D, Claustres M, Béroud C, Collod-Béroud G. UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity -- application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Hum Mutat* 2009; **30**: 952-959 [PMID: 19370756 DOI: 10.1002/humu.20970]
26 **McGee TL**, Seyedahmadi BJ, Sweeney MO, Dryja TP, Berson EL. Novel mutations in the long isoform of the USH2A gene in patients with Usher syndrome type II or non-syndromic retinitis pigmentosa. *J Med Genet* 2010; **47**: 499-506 [PMID: 20507924 DOI: 10.1136/jmg.2009.075143]
27 **Bergman JE**, Janssen N, van der Sloot AM, de Walle HE, Schoots J, Rendtorff ND, Tranebjaerg L, Hoefsloot LH, van Ravenswaaij-Arts CM, Hofstra RM. A novel classification system to predict the pathogenic effects of CHD7 missense variants in CHARGE syndrome. *Hum Mutat* 2012; **33**: 1251-1260 [PMID: 22539353 DOI: 10.1002/humu.22106]
28 **Manickam M**, Ravanan P, Singh P, Talwar P. In silico identification of genetic variants in glucocerebrosidase (GBA) gene involved in Gaucher's disease using multiple software tools. *Front Genet* 2014; **5**: 148 [PMID: 24904648 DOI: 10.3389/fgene.2014.00148]
29 **Krieger E**, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 2009; **77** Suppl 9: 114-122 [PMID: 19768677 DOI: 10.1002/prot.22570]
30 **Ryan M**, Diekhans M, Lien S, Liu Y, Karchin R. LS-SNP/PDB: annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics* 2009; **25**: 1431-1432 [PMID: 19369493 DOI: 10.1093/bioinformatics/btp242]
31 **Fraczkiewicz R**, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* 1998; **19**: 319-333 [DOI: 10.1002/(SICI)1096-987X(199802)19:3<319::AID-JCC6>3.0.CO;2-W]
32 **Capriotti E**, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 2005; **33**: W306-W310 [PMID: 15980478 DOI: 10.1093/nar/gki375]
33 **Capriotti E**, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* 2006; **22**: 2729-2734 [PMID: 16895930 DOI: 10.1093/bioinformatics/btl423]
34 **Pires DE**, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 2014; **30**: 335-342 [PMID: 24281696 DOI: 10.1093/bioinformatics/btt691]
35 **Worth CL**, Preissner R, Blundell TL. SDM--a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res* 2011; **39**: W215-W222 [PMID: 21593128 DOI: 10.1093/nar/gkr363]
36 **Worth CL**, Bickerton GR, Schreyer A, Forman JR, Cheng TM, Lee S, Gong S, Burke DF, Blundell TL. A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. *J Bioinform Comput Biol* 2007; **5**: 1297-1318 [PMID: 18172930]
37 **Cheng J**, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* 2006; **62**: 1125-1132 [PMID: 16372356 DOI: 10.1002/prot.20810]
38 **Chen CW**, Lin J, Chu YW. iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics* 2013; **14** Suppl 2: S5 [PMID: 23369171 DOI: 10.1186/1471-2105-14-S2-S5]
39 **Bendl J**, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, Brezovsky J, Damborsky J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* 2014; **10**: e1003440 [PMID: 24453961 DOI: 10.1371/journal.pcbi.1003440]
40 **Capriotti E**, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* 2013; **14** Suppl 3: S2 [PMID: 23819846 DOI: 10.1186/1471-2164-14-S3-S2]
41 **Luu TD**, Rusu A, Walter V, Linard B, Poidevin L, Ripp R, Moulinier L, Muller J, Raffelsberger W, Wicker N, Lecompte O, Thompson JD, Poch O, Nguyen H. KD4v: Comprehensible Knowledge Discovery System for Missense Variant. *Nucleic Acids Res* 2012; **40**: W71-W75 [PMID: 22641855 DOI: 10.1093/nar/gks474]
42 **Schymkowitz J**, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids*

Res 2005; **33**: W382-W388 [PMID: 15980494 DOI: 10.1093/nar/gki387]
43 **Dehouck Y**, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 2011; **12**: 151 [PMID: 21569468 DOI: 10.1186/1471-2105-12-151]
44 **Parthiban V**, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res* 2006; **34**: W239-W242 [PMID: 16845001 DOI: 10.1093/nar/gkl190]
45 **Parthiban V**, Gromiha MM, Abhinandan M, Schomburg D. Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. *BMC Struct Biol* 2007; **7**: 54 [PMID: 17705837]
46 **Dehouck Y**, Kwasigroch JM, Rooman M, Gilis D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res* 2013; **41**: W333-W339 [PMID: 23723246 DOI: 10.1093/nar/gkt450]
47 **Fernandez-Escamilla AM**, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 2004; **22**: 1302-1306 [PMID: 15361882 DOI: 10.1038/nbt1012]
48 **Maurer-Stroh S**, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JW, Rousseau F. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* 2010; **7**: 237-242 [PMID: 20154676 DOI: 10.1038/nmeth.1432]
49 **Van Durme J**, Maurer-Stroh S, Gallardo R, Wilkinson H, Rousseau F, Schymkowitz J. Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol* 2009; **5**: e1000475 [PMID: 19696878 DOI: 10.1371/journal.pcbi.1000475]
50 **Sievers F**, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011; **7**: 539 [PMID: 21988835 DOI: 10.1038/msb.2011.75]
51 **Valdar WS**. Scoring residue conservation. *Proteins* 2002; **48**: 227-241 [PMID: 12112692 DOI: 10.1002/prot.10146]
52 **Kumar P**, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073-1081 [PMID: 19561590 DOI: 10.1038/nprot.2009.86]
53 **Choi Y**, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 2012; **7**: e46688 [PMID: 23056405 DOI: 10.1371/journal.pone.0046688]
54 **Calabrese R**, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* 2009; **30**: 1237-1244 [PMID: 19514061 DOI: 10.1002/humu.21047]
55 **Brunham LR**, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet* 2005; **1**: e83 [PMID: 16429166 DOI: 10.1371/journal.pgen.0010083]
56 **Salomonis N**, Hanspers K, Zambon AC, Vranizan K, Lawlor SC, Dahlquist KD, Doniger SW, Stuart J, Conklin BR, Pico AR. GenMAPP 2: new features and resources for pathway analysis. *BMC Bioinformatics* 2007; **8**: 217 [PMID: 17588266]
57 **Adzhubei IA**, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248-249 [PMID: 20354512 DOI: 10.1038/nmeth0410-248]
58 **Adzhubei I**, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013; **Chapter 7**: Unit7.20 [PMID: 23315928 DOI: 10.1002/0471142905.hg0720s76]
59 **Bao L**, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res* 2005; **33**: W480-W482 [PMID: 15980516]
60 **Reva B**, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011; **39**: e118 [PMID: 21727090 DOI: 10.1093/nar/gkr407]
61 **Garla V**, Kong Y, Szpakowski S, Krauthammer M. MU2A--reconciling the genome and transcriptome to determine the effects of base substitutions. *Bioinformatics* 2011; **27**: 416-418 [PMID: 21149339 DOI: 10.1093/bioinformatics/btq658]
62 **Magesh R**, George Priya Doss C. Computational methods to work as first-pass filter in deleterious SNP analysis of alkaptonuria. *ScientificWorldJournal* 2012; **2012**: 738423 [PMID: 22606059 DOI: 10.1100/2012/738423]
63 **Manning JR**, Bailey MA, Soares DC, Dunbar DR, Mullins JJ. In silico structure-function analysis of pathological variation in the HSD11B2 gene sequence. *Physiol Genomics* 2010; **42**: 319-330 [PMID: 20571110 DOI: 10.1152/physiolgenomics.00053.2010]
64 **Riera C**, Lois S, Domínguez C, Fernandez-Cadenas I, Montaner J, Rodríguez-Sureda V, de la Cruz X. Molecular damage in Fabry disease: characterization and prediction of alpha-galactosidase A pathological mutations. *Proteins* 2015; **83**: 91-104 [PMID: 25382311 DOI: 10.1002/prot.24708]
65 **Cammisa M**, Correra A, Andreotti G, Cubellis MV. Fabry_CEP: a tool to identify Fabry mutations responsive to pharmacological chaperones. *Orphanet J Rare Dis* 2013; **8**: 111 [PMID: 23883437 DOI: 10.1186/1750-1172-8-111]
66 **B R**, C GP. Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders--a computational approach. *PLoS One* 2011; **6**: e24607 [PMID: 21931771 DOI: 10.1371/journal.pone.0024607]
67 **Carvalho DR**, Brand GD, Brum JM, Takata RI, Speck-Martins CE, Pratesi R. Analysis of novel ARG1 mutations causing hyperarginemia and correlation with arginase I activity in erythrocytes. *Gene* 2012; **509**: 124-130 [PMID: 22959135 DOI: 10.1016/j.gene.2012.08.003]
68 **George DC**, Chakraborty C, Haneef SA, Nagasundaram N, Chen L, Zhu H. Evolution- and structure-based computational strategy reveals the impact of deleterious missense mutations on MODY 2 (maturity-onset diabetes of the young, type 2). *Theranostics* 2014; **4**: 366-385 [PMID: 24578721 DOI: 10.7150/thno.7473]
69 **Fancello T**, Dardis A, Rosano C, Tarugi P, Tappino B, Zampieri S, Pinotti E, Corsolini F, Fecarotta S, D'Amico A, Di Rocco M, Uziel G, Calandra S, Bembi B, Filocamo M. Molecular analysis of NPC1 and NPC2 gene in 34 Niemann-Pick C Italian patients: identification and structural modeling of novel mutations. *Neurogenetics* 2009; **10**: 229-239 [PMID: 19252935 DOI: 10.1007/s10048-009-0175-3]
70 **Facchiano A**, Marabotti A. Analysis of galactosemia-linked mutations of GALT enzyme using a computational biology approach. *Protein Eng Des Sel* 2010; **23**: 103-113 [PMID: 20008339 DOI: 10.1093/protein/gzp076]
71 **McCorvie TJ**, Timson DJ. 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* 2013; **524**: 95-104 [PMID: 23644136 DOI: 10.1016/j.gene.2013.04.061]

P- Reviewer: McEachin RC, Wei QZ **S- Editor**: Ji FF
L- Editor: A **E- Editor**: Liu SQ

Published by **Baishideng Publishing Group Inc**
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com