

***In Silico* Prediction of the Effects of Mutations in the Human Mevalonate Kinase Gene: Towards a Predictive Framework for Mevalonate Kinase Deficiency**

Claire Browne¹ and David J. Timson^{1,2*}

¹*School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast, BT9 7BL, UK*

²*Institute for Global Food Security, Queen's University Belfast, 18-30 Malone Road, Belfast, BT9 5BN, UK*

Summary

Mevalonate kinase (MVK) catalyses the phosphorylation of mevalonate. Deficiency of MVK is associated with two rare periodic fever syndromes, mevalonic aciduria (MA), a severe form and hyper-immunoglobulin-D syndrome (HIDS), a milder form. An *in silico* approach was used to analyse the physicochemical and structural effects of 47 disease-associated variants of MVK. A further 20 variants, which are present in human genome databases, were also analysed. Variants associated with MA are clustered into a “hotspot” consisting of residues 8–35 and 234–338 and tended to result in a prediction of severely reduced protein stability. Four of the uncharacterised variants, p.H24P, p.G198R, p.R253W, and p.G335S, were likely to be associated with MA. This method could be used as the basis for initial predictions of severity when new MVK variants are discovered.

Keywords: Periodic fever syndrome, mevalonic aciduria, hyper-immunoglobulin-D syndrome, HIDS, protein stability, disease prediction

Introduction

Mevalonate kinase (MVK) deficiency is associated with a wide range of systemic diseases ranging from the most severe form, mevalonic aciduria (MA; OMIM#610377) to a milder form, hyper-immunoglobulin D syndrome (HIDS; OMIM#60920) (Drenth et al., 1999; Houten et al., 2000c; Haas & Hoffmann 2006). These rare autosomal recessive disorders are caused by a mutation in the *MVK* gene, which is found on the long arm of chromosome 12 (12q24) (Haas & Hoffmann 2006, Houten et al. 2000c, Houten et al. 2000b). Mevalonate kinase (EC 2.7.1.36) is an essential enzyme in the cholesterol pathway (Buhaescu & Izzedine 2007). It is a dimeric protein that catalyses the ATP-dependent phosphorylation of mevalonate to 5-phosphomevalonate (Potter & Mizioro 1997). This step follows the one catalysed by the highly regulated enzyme and statin target, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase (EC

1.1.1.34). Although there is no effective treatment for MVK deficiency, preliminary evidence has shown that Simvastatin, an HMG-CoA reductase inhibitor, may be useful in treating inflammatory attacks in HIDS (Simon et al. 2004).

MVK is part of the GHMP family of kinases (Bork et al. 1993, Timson 2007). Each subunit has two domains with the active site located in a cleft at the domain interface (Yang et al. 2002, Fu et al. 2002, Sgraja et al. 2007, Fu et al. 2008). The enzyme is dimeric. Thus, mutations causing amino acid changes at the dimer interface may disrupt dimer formation leading to an inactive or unstable protein (Fu et al. 2002). The crystal structure of rat mevalonate kinase in complex with MgATP revealed that an Mg²⁺ ion is co-ordinated by both β - and γ -phosphates of ATP and side chains of Glu-193 and Ser-146 (Fu et al. 2002). Asp204 was found to make a salt bridge with Lys-13, which interacts with the γ -phosphate (Fu et al. 2002). Lys-13 affects the pK_a of the C5 hydroxyl of mevalonate, while Asp-204 abstracts the proton from this hydroxyl. The resulting penta-coordinated γ -phosphoryl group may be stabilized by Mg²⁺, Lys-13, and Glu-193 (Fu et al. 2002). The structure of the human MVK enzyme is very similar to the rat one (rmsd 0.81 Å) and the mechanism of reaction is expected to be identical with Asp-204 acting as a base in the active site (Fu et al. 2008).

*Corresponding author: DAVID J. TIMSON, School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast, BT9 7BL, UK. Tel: +44(0)28 9097 5875; Fax: +44(0)28 9097 5877; E-mail: d.timson@qub.ac.uk

MA and HIDS are rare autosomal recessive disorders characterised by recurrent periodic fevers and generalised inflammation (Haas & Hoffmann 2006, Frenkel et al. 2000). Typical symptoms of the severe form, MA, can include psychomotor retardation, ataxia, failure to thrive, dysmorphic features with recurrent episodes of fevers, lymphadenopathy and rashes (Frenkel et al. 2000, van der Burgh et al. 2013). These symptoms typically present in infancy and patients usually die in childhood (Frenkel et al. 2000). Sufferers of MA can be confirmed based on their biochemical, clinical and genetic data (Drenth et al. 1999). The residual activity of MVK in MA sufferers is below 0.5%, leading to a characteristic build-up of mevalonic acid (Prietsch et al. 2003).

HIDS is a milder phenotype of MVK deficiency with more documented cases (Haas & Hoffmann 2006, Frenkel et al. 2000, van der Burgh et al. 2013, Grose 2005). Symptoms usually present in the first year of life and include recurrent lifelong episodes of fever, rashes, abdominal pain, and lymphadenopathy (Haas & Hoffmann 2006, Frenkel et al. 2000, McDermott & Frenkel 2001, Haas et al. 2001, Hoffman et al. 2001). A high level of IgD and IgA accompanies these symptoms and is often used as a diagnostic confirmation of HIDS (Drenth et al. 1999, Houten et al. 2000c, Buhaescu & Izzedine 2007). Moreover, the residual activity of MVK in HIDS is around 1%–20% (Houten et al. 2000c). HIDS can be distinguished from MA by the lack of abnormal levels of mevalonic acid present as well as the lack of neurological manifestations which feature in MA (Hoffman et al. 2001). Furthermore, unlike individuals with MA, people suffering with HIDS do not experience any symptoms between fever episodes and typically have a normal life expectancy (Haas & Hoffmann 2006).

The most frequent mutation associated with HIDS is c.1129G>A (rs28934897) resulting in an amino acid change from valine to isoleucine at position 377 (p.V377I) (Cuisset et al. 2001). This is found in approximately 80% of HIDS sufferers and has never been reported in MA (Cuisset et al. 2001). MVK deficiency is particularly common in northern Europe (especially The Netherlands and France) but cases have been reported worldwide (Prietsch et al. 2003). Approximately 80 mutations have been reported to cause mevalonate kinase deficiency (Haas & Hoffmann 2006).

Overlapping phenotypes of MA and HIDS have been observed in an intermediate form, although few cases have been documented (Hoffman et al. 2001). However, it is important to note that the relationship between mutation and reported phenotype may not always be straightforward. Phenotypes are classified based on individual clinical observation and can be influenced by environmental factors; for example, the symptoms of MVK deficiency appear to be amplified by increased temperature (Houten et al. 2002a, Tricarico et al. 2013).

It is not known precisely how each disease-associated mutation alters the enzyme structure or how this contributes to the different levels of severity or pathology. It is assumed that the residual activity of MVK determines whether or not a patient has MA or HIDS, therefore a mutation associated with the severe form must disrupt the enzyme's structure and function significantly to decrease its activity below 0.5%. At the biochemical level this is likely to result from effects on protein folding and stability. A bioinformatics approach was taken to understand the functional and structural contribution of molecular alterations in MVK and how these correlate to the associated severity. This was then applied to variants of MVK whose disease association is currently unknown. The approaches can be broadly arranged into three approaches: sequence and evolutionary conservation based methods, protein sequence and structure-based methods, and supervised learning methods. Similar approaches have been successful in the investigation of type I galactosemia (galactose 1-phosphate uridylyltransferase deficiency), type III galactosemia (UDP-galactose 4'-epimerase deficiency), hyperargininemia (arginase 1 deficiency), and apparent mineralocorticoid excess (11 β -hydroxysteroid dehydrogenase type 2 deficiency) to provide the basis for predicting the severity of newly discovered mutations (Carvalho et al. 2012, Manning et al. 2010, Facchiano & Marabotti 2010, d'Acierno et al. 2009, d'Acierno et al. 2014, McCorvie & Timson 2013).

Materials and Methods

Datasets, Sequence Information, and Data Analysis

Literature searches were used to identify disease-associated variants. Where available, the clinical symptoms, observed residual activity of the MVK variant, any biochemical analysis (including any experimental conditions), environmental factors, and the genetic background were noted. Characterised variants and uncharacterised missense mutations from exome sequencing were identified from the databases: NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>), UniProt (<http://www.uniprot.org/>), and Infevers (<http://fmf.igh.cnrs.fr/ISSAID/infevers/>). Excel (Microsoft) and GraphPad Prism 5.0 (GraphPad Software) were used to record scores, perform statistical analysis, and obtain graphical representations of the results.

Secondary and Tertiary Structural Investigations

Structural studies were based on the crystal structure of human MVK (PDB: 2R3V) (Fu et al. 2008). Structures

were visualised using PyMol (<http://www.pymol.com>) and were computationally solvated and energy minimized using YASARA (<http://www.yasara.org>) (Krieger et al. 2009). Sequence variations were introduced *in silico* using the Mutate function in PyMol and these variant forms of the protein were minimised using YASARA. Thus, all structural comparisons were made using minimised structures. Each variant's structure was further analysed using LS-SNP/PDB (<http://ls-snp.icm.jhu.edu/ls-snp-pdb/main>). This tool provided solvent accessibility scores, variant's position at domain interface, location in the 3D structure, if it is exposed or buried, the secondary structure where the variant is found and a 3D representation of the variant in the structure. The GETAREA server (<http://curie.utmb.edu/getarea.html>) was used to determine the change in surface and buried atoms from the variant compared to the wild type (Fraczkiewicz & Braun 1998).

Prediction of the Physicochemical Effect of Each Missense Mutation on MVK Structure, Stability, and Function

Bioinformatics tools were used to investigate a range of physicochemical effects of each variant on MVK's structure and function. The following tools were used to estimate the change in protein stability: I-Mutant 3.0 Binary and Ternary classification (<http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi>) (Capriotti et al. 2005, Capriotti et al. 2006); mCSM (<http://bleoberis.bioc.cam.ac.uk/mcsm/>) (Pires et al. 2014); SDM score (<http://mordred.bioc.cam.ac.uk/~sdm/sdm.php>) (Worth et al. 2011, Worth et al. 2007); Mupro (<http://mupro.proteomics.ics.uci.edu/>) (Cheng et al. 2006); iStable score (<http://predictor.nchu.edu.tw/iStable/>) (Chen et al. 2013); PredictSNP 1.0 (<http://loschmidt.chemi.muni.cz/predictsnp/>) (Bendl et al. 2014); Meta-SNP (<http://snps.biofold.org/meta-snp/>) (Capriotti et al. 2013); KD4V score (<http://decryphon.igbmc.fr/kd4v>) (Luu et al. 2012).

The tools used to estimate the change in free energy resulting from the point mutations (all calculated at pH 7.5, 30 °C) were: I-Mutant 3.0 DDG, SVM3 and SVM2 values (Capriotti et al. 2006); Fold-X DDG value (<http://foldx.crg.es>) (Schymkowitz et al. 2005); PoPMuSiC 2.1 (<http://dezyme.com/>) (Dehouck et al. 2011); CUPSAT DDG (<http://cupsat.tu-bs.de>) (Parthiban et al. 2006, Parthiban et al. 2007); iStable DDG (Chen et al. 2013); GETAREA energy change (<http://curie.utmb.edu/getarea.html>) (Fraczkiewicz & Braun 1998).

The tools used to predict change in binding affinity were: mCSM Protein-Protein affinity prediction (Pires et al. 2014); BeAtMuSiC binding affinity prediction and score (<http://babylone.ulb.ac.be/beatmusic>) (Dehouck et al. 2013).

The tools used to predict the change in solvent accessibility were: PoPMuSiC 2.1 (Dehouck et al. 2011); CUPSAT RSA (<http://cupsat.tu-bs.de>) (Parthiban et al. 2007); SDM score (Worth et al. 2011, Worth et al. 2007); BeAtMuSiC (Dehouck et al. 2013).

In addition, further physicochemical analysis was carried out using SNP effect 4.0 server (<http://snpeffect.switchlab.org>) (De Baets et al. 2012) which uses four programs to predict the biochemical effects of missense mutations; TANGO (<http://tango.crg.es/>) (Fernandez-Escamilla et al. 2004) which predicts the tendency to aggregate; Fold-X (Schymkowitz et al. 2005) which predicts changes in stability; WALTZ (<http://www.switchlab.org/bioinformatics/waltz>) (Maurer-Stroh et al. 2010) which predicts the tendency to form amyloids and LIMBO (<http://www.switchlab.org/bioinformatics/limbo>) (Van Durme et al. 2009) which predicts the ability to bind chaperones. An estimation of the change in hydrogen bond satisfaction was also carried out using SDM (Worth et al. 2011, Worth et al. 2007). The KD4V server (Luu et al. 2012) provided additional scores in the change in size, charge polarity, modifications, and hydrophobicity. Change in hydrophobicity scores were calculated using data from (Black & Mould 1991).

Multiple Sequence Alignment and Evolutionary Conservation-Based Methods

Multiple sequence alignment was carried out using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) (Sievers et al. 2011) from sequences gathered from UniProt database (<http://www.uniprot.org>). Only those classified as MVK and with "reviewed" status were used. A total of 19 sequences were obtained, which included one from yeast, four from vertebrates, one from slime mould, one from plants, one from Archaea, and eleven from eubacteria. The resulting alignment (Fig. S1) was used with the Scorecons server (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl) based on the Valdar01 score (Valdar 2002). The alignment was also used to calculate the degree of tolerance of each amino acid alteration using the SIFT server (<http://sift.jcvi.org/>) (Kumar et al. 2009) and the PROVEAN server (<http://provean.jcvi.org/index.php>) (Choi et al. 2012), both of which are based on updated versions of Ensembl gene annotation (GRCh37 Ensembl 66) and NCBI dbSNP database (Build 137).

Additional investigations of the tolerance to change at positions which are altered in the disease-associated variants was carried out using the tools: LS-SNP/PDB (<http://ls-snp.icm.jhu.edu/ls-snp-pdb/>) (Ryan et al. 2009); SNPs&GO (<http://snps.biofold.org/snps-and-go/pages/help.html>) (Calabrese et al. 2009); PhD-SNP

(<http://snps.biofold.org/phd-snp/phd-snp.html>) (Capriotti et al. 2006); PANTHER (<http://www.pantherdb.org/tools/csnpscoreForm.jsp>) (Brunham et al. 2005); GenMAPP (<http://www.genmapp.org>) (Salomonis et al. 2007); PolyPhen 2 (<http://genetics.bwh.harvard.edu/pph2/>) (Adzhubei et al. 2010, Adzhubei et al. 2013); nsSNP Analyzer (<http://snpanalyzer.uthsc.edu>) (Bao et al. 2005); FI mutation assessor (<http://mutationassessor.org/v1>) (Reva et al. 2011); KD4V (Luu et al. 2012); YALE MU2A (<http://krauthammerlab.med.yale.edu/mu2a>) (Garla et al. 2011).

Supervised Learning Methods for Combined Overall Predictions of Each Variant Severity

Supervised learning methods, including a variety of bioinformatics tools, were used to compare variants and provide an overall prediction on the effect of the variant on the protein's structure, function, and pathology. Programs that provided this overall score, predicting if the mutation was disease causing, damaging, destabilizing or deleterious were: I-Mutant 3.0 Prediction of Disease (Capriotti et al. 2006); SIFT (Kumar et al. 2009); PROVEAN prediction (Choi et al. 2012); PolyPhen 2 prediction (Adzhubei et al. 2013); mCSM prediction (Pires et al. 2014); SNAP prediction (<https://roslab.org/services/snap2web/>) (Bromberg & Rost 2007); PANTHER (Brunham et al. 2005); PhD-SNP predictor (Capriotti et al. 2006); nsSNP Analyzer (Bao et al. 2005); Mupro prediction (Cheng et al. 2006); PoPMuSiC 2.1 (Dehouck et al. 2011); CUPSAT (Parthiban et al. 2007); iStable (Chen et al. 2013); PredictSNP (Bendl et al. 2014); MAPP (Salomonis et al. 2007); Meta-SNP (Capriotti et al. 2013); FI mutation assessor (Reva et al. 2011); BeAtMuSiC (Dehouck et al. 2013).

Results and Discussion

MVK Alterations Associated with MA Tend to be Located in the Central Cleft, Near the Active Site

Literature and database searches identified 47 MVK mutations for which there was sufficient biochemical and clinical information to be included in this study. Four of these were associated with the intermediate severity MVK deficiency, 12 with the severe form (MA), and the remainder with the mild form (HIDS). The location of the changes in the protein sequence showed that there was a relationship between the position of the variant and associated severity, with severe and intermediate mutations only occurring between amino acid positions 8–35 and 234–338 (Fig. 1). These two sequences are located around the inside of the protein's cleft close to the active site and around the domain interface. Four of the

variations associated with the severe form (p.L264F, p.L265P, p.I268H, and p.N301T) occur in α -helices at the dimer interface. To our knowledge, there is no published information on the effect of these (or other) mutations on the ability of MVK to dimerise; nor is there any information on whether, or not, dimerization is required for full activity of human MVK. However, MVK from some bacteria is monomeric, suggesting that the oligomeric state may have limited impact on activity (Hedl & Rodwell 2004, Voynova et al. 2004). The alterations associated with HIDS were more widespread throughout the MVK structure (Fig. 1). The position of the variants in the 3D structure from LS-SNP/PDB showed that the majority of variants associated with severe and intermediate forms of the disease affected buried residues, while the variants associated with HIDS affected a mixture of buried, intermediate, and exposed residues.

Computational Analysis Tends to Place the MVK Variants into Two Groups

All the results from the physicochemical analysis of the disease-associated and uncharacterised variants are given in Table S1. There was a tendency to predict decreasing structural stability with increasing severity of the disease-associated variants. The programs that identified this trend were: I-Mutant DDG, solvent accessibility, FOLD-X, mCSM stability, Mupro, PoPMuSiC 2.1, BeAtMuSiC, CUPSAT (RSA), iStable DDG, GETAREA, SDM, BETAMUSIC, and KD4V. The change in hydrophobicity tended to increase with severity (data not shown). In the majority of cases, however, there was no statistical significance ($P > 0.05$) when the three groups (mild, intermediate, and severe) were compared. In general, alterations at more highly conserved residues were more likely to result in severe forms of MVK deficiency. This result was supported by predictions from: POLYPHEN-2, SAM, SCORECONS, PHD-SNP, SNP&GO, PANTHER, MAPP, FI-mutant, and 4D4V WT residue representation. However, like the physicochemical scores, there was no significant difference between the scores when comparing the results from the three groups with the majority of the programs (data not shown).

Supervised Learning Methods Predict Effects on MVK Stability Using a Combination of Methods

Programs that used supervised learning methods for an overall prediction were used to predict each variant's effect on the stability of MVK. It should be noted that this approach is based on the assumption that all the prediction tools are equally good. However, the percentage score provides an overall consensus which has been found to be more accurate than only using one or two tools (Zhao et al., 2014). The results in

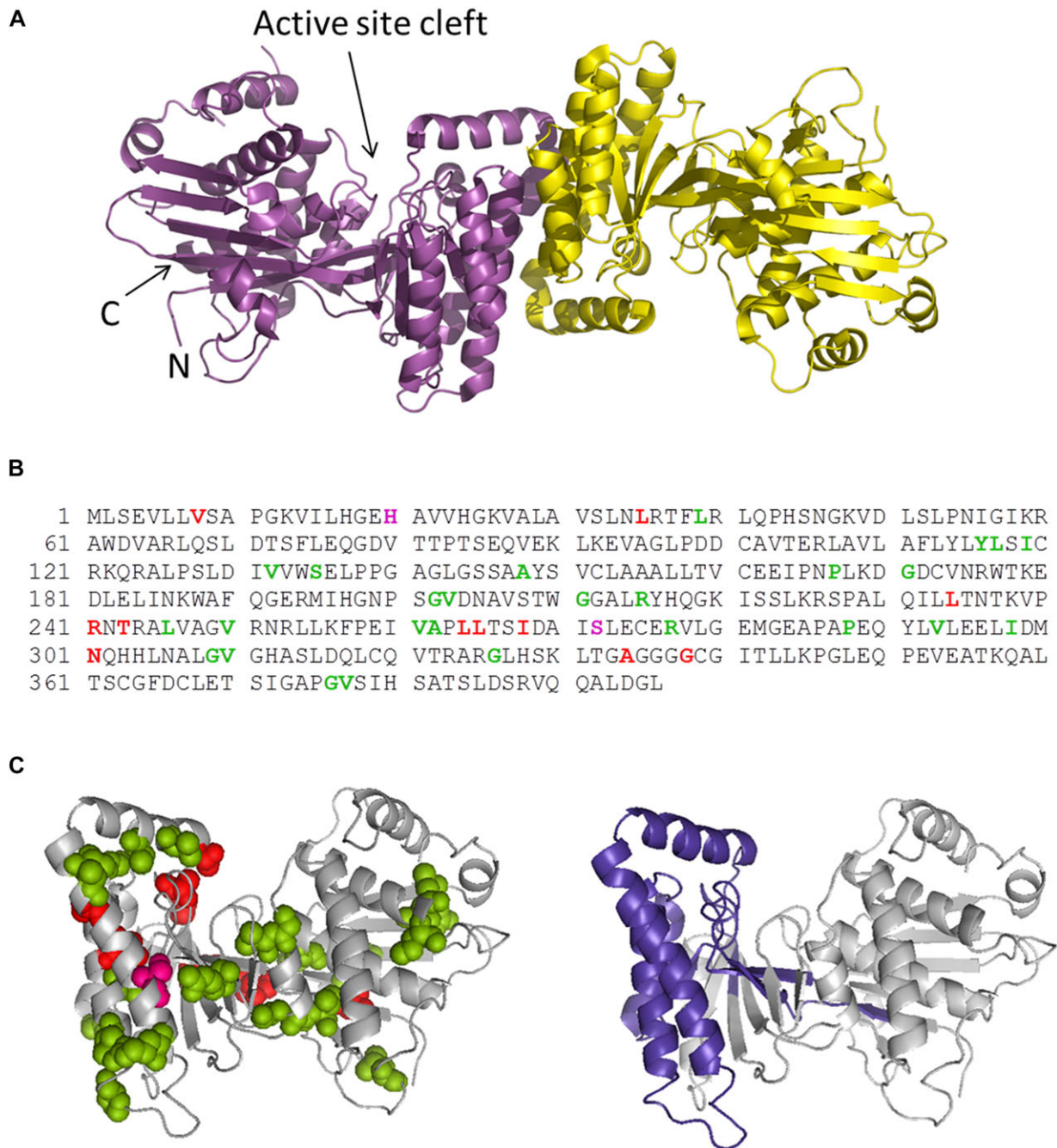


Figure 1 Structure of human mevalonate kinase showing the location of changes resulting from disease-associated mutations. (A) The dimeric structure of human MVK showing the N- and C-termini and active site cleft of one of the subunits. (B) The sequence of human MVK (GenBank: NP_001107657) showing residues which when altered are associated with mild (green), intermediate (magenta), and severe (red) forms of MVK deficiency. (C) These residues mapped onto the three-dimensional structure of a single MVK subunit (left). Each residue associated with disease is shown as a space-filled model. Those associated with the severe form (MA) are coloured red. Magenta indicates residues which are altered in the intermediate form and green residues altered in the mild form (HIDS). The “hotspot” which includes all the alterations associated with MA is coloured blue in the right-hand figure. Parts (A) and (C) were produced using PyMol and (PDB: 2R3V) (Fu et al. 2008).

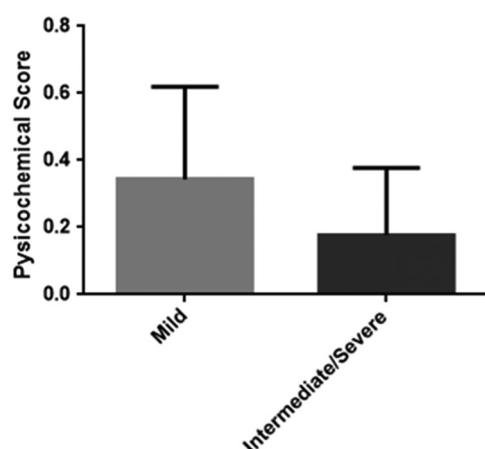


Figure 2 When combined, the intermediate and severe groups of variants show a statistically significant difference from the mild variants in the mean physicochemical score ($p = 0.038$).

Table S2 are based on 28 prediction tools and show that, in general, the percentage of tools predicting a reduction in MVK stability slightly increases with increasing severity. For each group the percentage predicting a decrease ranged from: 28.6 to 89.3 (mild), 53.6 to 89.3 (intermediate), and 64.3 to 92.9 (severe). Although there is an increasing trend, the discrimination between groups was not sufficient to make meaningful predictions about variants of unknown severity.

The Intermediate Form Was Not Well Detected by These Methods

Across the three different prediction methods, the intermediate group was not well discriminated. In general, it showed results similar to, or greater than, the severe group (Table S1). This suggests that neither these methods are sufficiently discriminating or that the intermediate group would be better considered as part of the severe group. Combining these two groups enabled discrimination between the combined intermediate/severe group and the mild group on the basis of the physicochemical scores (Fig. 2). Therefore, in developing a predictive framework, these two groups were considered together.

Towards a Predictive Framework: Four Key Postulates

Based on the data (summarised in Table S1), four postulates were made:

1. The decrease in enzyme activity is correlated with severity of disease. In the majority of cases, this decrease results from improper folding of the protein.
2. Changes to highly conserved residues are more likely to result in improper folding.
3. Intermediate mutations should be treated with the severe group due to their similarity to this group.
4. For a severe or intermediate phenotype to occur, the mutation is likely to cause a change in the “hotspot” region, identified from structural analysis, i.e., around residues 8–35 and 234–338 in the protein sequence.

Prediction of the Likely Clinical Severity Associated with Uncharacterised Variants

There are a number of variants which have been identified through genome and exome sequencing projects, but which have not yet been characterised in terms of their associated disease severity. Applying the postulates above, four of these variants are predicted to be associated with the intermediate or severe form (MA). These are p.H24P, p.G198R, p.R253W, and p.G335S (Table S3). Three of these uncharacterised variants were predicted to cause either the mild form (HIDS) or no disease and the rest were predicted to be associated with HIDS (Table S3). Experimental testing will be necessary to confirm these predictions. Molecular dynamics simulations may be valuable in explaining why disease-associated variants affect protein stability and would also be useful in investigating the properties of these uncharacterised variants. It is further hypothesised that the MVK variants which are predicted to be associated with MA will have very low activities and stabilities *in vitro*. It is also expected that cells homozygous for the corresponding mutations are likely to be deficient in the mevalonate pathway and will accumulate mevalonate.

Conclusions

The analyses presented here suggest that point mutations in the *MVK* gene which result in changes to the protein coding sequence subtly alter the structure and stability of MVK, thus affecting its activity. These effects are greater when the change occurs in a well-conserved residue. Variants associated with MA generally result in changes to the highly conserved “hotspot” region and also result in greater changes to stability and other biophysical parameters. In addition to providing a predictive framework, these results also suggest that pharmacological interventions to stabilise the MVK protein (“small molecule chaperone therapy”) may be possible in the treatment of MA and in the alleviation of the feverish episodes in HIDS.

Acknowledgements

We thank Dr Thomas J. McCorvie (Structural Genomics Centre, University of Oxford, UK) for helpful discussions. This work received no funding from any external source.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

- Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7, Unit7.20.
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248–249.
- Bao, L., Zhou, M. & Cui, Y. (2005) nsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res* **33**, W480–W482.
- Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J. & Damborsky, J. (2014) PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol* **10**, e1003440.
- Black, S. D. & Mould, D. R. (1991) Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal Biochem* **193**, 72–82.
- Bork, P., Sander, C. & Valencia, A. (1993) Convergent evolution of similar enzymatic function on different protein folds: The hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci* **2**, 31–40.
- Bromberg, Y. & Rost, B. (2007) SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res* **35**, 3823–3835.
- Brunham, L. R., Singaraja, R. R., Pape, T. D., Kejariwal, A., Thomas, P. D. & Hayden, M. R. (2005) Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. *PLoS Genet* **1**, e83.
- Buhaescu, I. & Izzedine, H. (2007) Mevalonate pathway: a review of clinical and therapeutical implications. *Clin Biochem* **40**, 575–584.
- Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* **30**, 1237–1244.
- Capriotti, E., Altman, R. B. & Bromberg, Y. (2013) Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* **14** (Suppl 3), S2–2164–14–S3–S2. Epub 2013 May 28.
- Capriotti, E., Calabrese, R. & Casadio, R. (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **22**, 2729–2734.
- Capriotti, E., Fariselli, P. & Casadio, R. (2005) I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* **33**, W306–W310.
- Carvalho, D. R., Brand, G. D., Brum, J. M., Takata, R. I., Speck-Martins, C. E. & Pratesi, R. (2012) Analysis of novel ARG1 mutations causing hyperargininemia and correlation with arginase I activity in erythrocytes. *Gene* **509**, 124–130.
- Chen, C. W., Lin, J. & Chu, Y. W. (2013) iStable: Off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics* **14** (Suppl 2), S5–2105–14–S2–S5. Epub 2013 Jan 21.
- Cheng, J., Randall, A. & Baldi, P. (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* **62**, 1125–1132.
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688.
- Cuisset, L., Drenth, J. P., Simon, A., Vincent, M. F., vander Velde Visser, S., vander Meer, J. W., Grateau, G., Delpech, M. & International Hyper-IgD Study Group. (2001) Molecular analysis of MVK mutations and enzymatic activity in hyper-IgD and periodic fever syndrome. *Eur J Hum Genet* **9**, 260–266.
- d'Acerno, A., Facchiano, A. & Marabotti, A. (2009) 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* **7**, 71–76.
- d'Acerno, A., Facchiano, A. & Marabotti, A. (2014) GALT protein database: Querying structural and functional features of GALT enzyme. *Hum Mutat* **35**, 1060–1067.
- DeBaets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., Schymkowitz, J. & Rousseau, F. (2012) SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res* **40**, D935–D939.
- Dehouck, Y., Kwasigroch, J. M., Gilis, D. & Rooman, M. (2011) PoPMuSiC 2.1: A web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* **12**, 151.
- Dehouck, Y., Kwasigroch, J. M., Rooman, M. & Gilis, D. (2013) BeAtMuSiC: Prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res* **41**, W333–W339.
- Drenth, J. P., Cuisset, L., Grateau, G., Vasseur, C., vande Velde-Visser, S. D., deJong, J. G., Beckmann, J. S., vander Meer, J. W. & Delpech, M. (1999) Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. International Hyper-IgD Study Group. *Nat Genet* **22**, 178–181.
- Facchiano, A. & Marabotti, A. (2010) Analysis of galactosemia-linked mutations of GALT enzyme using a computational biology approach. *Protein Eng Des Sel* **23**, 103–113.
- Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**, 1302–1306.
- Frackiewicz, R. & Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comput Chem* **19**, 319–333.
- Frenkel, J., Houten, S. M., Waterham, H. R., Wanders, R. J., Rijkers, G. T., Kimpfen, J. L., Duran, R., Poll-The, B. T. & Kuis, W. (2000) Mevalonate kinase deficiency and Dutch type periodic fever. *Clin Exp Rheumatol* **18**, 525–532.
- Fu, Z., Voynova, N. E., Herdendorf, T. J., Mizioroko, H. M. & Kim, J. J. (2008) Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism. *Biochemistry* **47**, 3715–3724.
- Fu, Z., Wang, M., Potter, D., Mizioroko, H. M. & Kim, J. J. (2002) The structure of a binary complex between a mammalian mevalonate kinase and ATP: Insights into the reaction mechanism and human inherited disease. *J Biol Chem* **277**, 18134–18142.

- Garla, V., Kong, Y., Szpakowski, S. & Krauthammer, M. (2011) MU2A—reconciling the genome and transcriptome to determine the effects of base substitutions. *Bioinformatics* **27**, 416–418.
- Grose, C. (2005) Periodic fever in children with hyperimmunoglobulinemia D and mevalonate kinase mutations. *Pediatr Infect Dis J* **24**, 573–574.
- Haas, D. & Hoffmann, G. F. (2006) Mevalonate kinase deficiencies: From mevalonic aciduria to hyperimmunoglobulinemia D syndrome. *Orphanet J Rare Dis* **1**, 13.
- Haas, D., Kelley, R. I. & Hoffmann, G. F. (2001) Inherited disorders of cholesterol biosynthesis. *Neuropediatrics* **32**, 113–122.
- Hedl, M. & Rodwell, V. W. (2004) *Enterococcus faecalis* mevalonate kinase. *Protein Sci* **13**, 687–693.
- Hoffman, H. M., Wanderer, A. A. & Broide, D. H. (2001) Familial cold autoinflammatory syndrome: Phenotype and genotype of an autosomal dominant periodic fever. *J Allergy Clin Immunol* **108**, 615–620.
- Houten, S. M., Frenkel, J., Rijkers, G. T., Wanders, R. J., Kuis, W. & Waterham, H. R. (2002a) Temperature dependence of mutant mevalonate kinase activity as a pathogenic factor in hyper-IgD and periodic fever syndrome. *Hum Mol Genet* **11**, 3115–3124.
- Houten, S. M., Frenkel, J., Kuis, W., Wanders, R. J., Poll-The, B. T. & Waterham, H. R. (2000b) Molecular basis of classical mevalonic aciduria and the hyperimmunoglobulinaemia D and periodic fever syndrome: High frequency of 3 mutations in the mevalonate kinase gene. *J Inher Metab Dis* **23**, 367–370.
- Houten, S. M., Wanders, R. J. & Waterham, H. R. (2000c) Biochemical and genetic aspects of mevalonate kinase and its deficiency. *Biochim Biophys Acta* **1529**, 19–32.
- Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. & Karplus, K. (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.
- Kumar, P., Henikoff, S. & Ng, P. C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073–1081.
- Luu, T. D., Rusu, A., Walter, V., Linard, B., Poidevin, L., Ripp, R., Moulinier, L., Muller, J., Raffelsberger, W., Wicker, N., Lecompte, O., Thompson, J. D., Poch, O. & Nguyen, H. (2012) KD4v: Comprehensive knowledge discovery system for missense variant. *Nucleic Acids Res* **40**, W71–W75.
- Manning, J. R., Bailey, M. A., Soares, D. C., Dunbar, D. R. & Mullins, J. J. (2010) *In silico* structure–function analysis of pathological variation in the *HSD11B2* gene sequence. *Physiol Genomics* **42**, 319–330.
- Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W. & Rousseau, F. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237–242.
- McCorvie, T. J. & 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.
- McDermott, M. F. & Frenkel, J. (2001) Hereditary periodic fever syndromes. *Neth J Med* **59**, 118–125.
- Parthiban, V., Gromiha, M. M., Abhinandan, M. & Schomburg, D. (2007) Computational modeling of protein mutant stability: Analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. *BMC Struct Biol* **7**, 54.
- Parthiban, V., Gromiha, M. M. & Schomburg, D. (2006) CUPSAT: Prediction of protein stability upon point mutations. *Nucleic Acids Res* **34**, W239–W242.
- Pires, D. E., Ascher, D. B. & Blundell, T. L. (2014) mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342.
- Potter, D. & Mizioro, H. M. (1997) Identification of catalytic residues in human mevalonate kinase. *J Biol Chem* **272**, 25449–25454.
- Prietsch, V., Mayatepek, E., Krastel, H., Haas, D., Zundel, D., Waterham, H. R., Wanders, R. J., Gibson, K. M. & Hoffmann, G. F. (2003) Mevalonate kinase deficiency: Enlarging the clinical and biochemical spectrum. *Pediatrics* **111**, 258–261.
- Reva, B., Antipin, Y. & Sander, C. (2011) Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res* **39**, e118.
- Ryan, M., Diekhans, M., Lien, S., Liu, Y. & Karchin, R. (2009) LS-SNP/PDB: Annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics* **25**, 1431–1432.
- Salomonis, N., Hanspers, K., Zambon, A. C., Vranizan, K., Lawlor, S. C., Dahlquist, K. D., Doniger, S. W., Stuart, J., Conklin, B. R. & Pico, A. R. (2007) GenMAPP 2: New features and resources for pathway analysis. *BMC Bioinformatics* **8**, 217.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. (2005) The FoldX web server: An online force field. *Nucleic Acids Res* **33**, W382–W388.
- Sgraja, T., Smith, T. K. & Hunter, W. N. (2007) Structure, substrate recognition and reactivity of *Leishmania major* mevalonate kinase. *BMC Struct Biol* **7**, 20.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. & Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.
- Simon, A., Drewe, E., vander Meer, J. W., Powell, R. J., Kelley, R. I., Stalenhoef, A. F. & Drenth, J. P. (2004) Simvastatin treatment for inflammatory attacks of the hyperimmunoglobulinemia D and periodic fever syndrome. *Clin Pharmacol Ther* **75**, 476–483.
- Timson, D. J. (2007) GHMP kinases – structures, mechanisms and potential for therapeutically relevant inhibition. *Curr Enz Inhib* **3**, 77–94.
- Tricarico, P. M., Kleiner, G., Piscianz, E., Zanin, V., Monasta, L., Crovella, S. & Marcuzzi, A. (2013) Temperature and drug treatments in mevalonate kinase deficiency: An *ex vivo* study. *Biomed Res Int* **2013**, 715465.
- Valdar, W. S. (2002) Scoring residue conservation. *Proteins* **48**, 227–241.
- vander Burgh, R., Ter Haar, N. M., Boes, M. L. & Frenkel, J. (2013) Mevalonate kinase deficiency, a metabolic autoinflammatory disease. *Clin Immunol* **147**, 197–206.
- Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F. & Schymkowitz, J. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol* **5**, e1000475.
- Voyanova, N. E., Rios, S. E. & Mizioro, H. M. (2004) *Staphylococcus aureus* mevalonate kinase: Isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. *J Bacteriol* **186**, 61–67.
- Worth, C. L., Bickerton, G. R., Schreyer, A., Forman, J. R., Cheng, T. M., Lee, S., Gong, S., Burke, D. F. & Blundell, T. L. (2007)

- A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. *J Bioinform Comput Biol* **5**, 1297–1318.
- Worth, C. L., Preissner, R. & Blundell, T. L. (2011) SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res* **39**, W215–W222.
- Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I. & Sacchettini, J. C. (2002) Structure of the *Methanococcus jannaschii* mevalonate kinase, a member of the GHMP kinase superfamily. *J Biol Chem* **277**, 9462–9467.
- Zhao, N., Han, J.G., Shyu, C.R., Korkin, D. (2014) Determining effects of non-synonymous SNPs on protein–protein interactions using supervised and semi-supervised learning. *PLoS Comput Biol* **10**, e1003592.

Supporting Information

Additional supporting information may be found in the online version of this article:

Figure S1: A sequence alignment generated in Clustal Omega (see Materials and Methods) of 19 verified MVK

protein sequences. This alignment was used in analysis of sequence conservation.

Table S1: The complete dataset from the various prediction tools used in this study.

Table S2: The percentage of predictions from 28 prediction tools (see Materials and Methods) for MVK variants predicting a decrease, increase or neutral change in stability. Whether or not the residue occurs in the structural “hotspot” is also tabulated.

Table S3: Predicted severities associated with currently uncharacterised variants. Variants were predicted to be intermediate/severe (I/S), mild (M), and mild/neutral (M/N) based on whether, or not, the altered residue was in the “hotspot” and whether the variant was predicted to be substantially less stable than wild-type.

Received: 16 March 2015

Accepted: 21 May 2015