Title: Using structural analysis to assess the impact of missense variants in *MEN1*

Authors:

Richard C. Caswell PhD* (1), Martina M. Owens MSc (2), Adam C. Gunning MRes (1), Sian Ellard PhD (2) & Caroline F. Wright PhD (1).

Author affiliations:

1. Institute of Biomedical and Clinical Science, College of Medicine & Health, University of Exeter, Exeter, United Kingdom.
2. Department of Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter, United Kingdom.

* corresponding author: email r.caswell@exeter.ac.uk; tel: +44 1392 408506

ABSTRACT

Previous studies have shown that thermodynamic analysis of protein structure *in silico* can discriminate between groups of benign and pathogenic missense variants. However, although structures exist for many human disease-associated proteins, such analysis remains largely unexploited in clinical laboratories. Here, we analysed the predicted effect of 338 known missense variants on the structure of Menin, the *MEN1* gene product. Results provided strong discrimination between pathogenic and benign variants, with a threshold of >4 kcal/mol for the predicted change in stability providing a strong indicator of pathogenicity. Subsequent analysis of 7 novel missense variants identified during clinical testing of MEN1 patients showed that all 7 were predicted to destabilise Menin by >4 kcal/mol. We conclude that structural analysis provides a useful tool in understanding the impact of missense variants in *MEN1*, and that integration of proteomic with genomic data could potentially contribute to the classification of novel variants in this disease.

INTRODUCTION

The rapid expansion in recent years of genomic data from both patient and control groups has vastly improved the quantity and quality of information that is available to clinicians in attempting to classify novel genetic variants. While it is often straightforward to interpret likely loss-of-function variants such as stop-gain or frameshift variants, the same is not true of missense variants, where the effect of an amino acid substitution is likely to be specific to its context in the protein of interest. Moreover, such variants are often rare or unique, and thus must be interpreted on a case-by-case basis.

Numerous methods have been developed for predicting the phenotypic effect of missense variants. As has been comprehensively reviewed elsewhere (1), these methods rely either on analysis of DNA and protein conservation, protein structure-based analysis, or a combination of the two. In the case of the latter, widely used tools such as PolyPhen are able to incorporate information on the nature of the amino acid change itself (e.g. Grantham distance between native and variant amino acids, changes in polarity or charge), effects on predicted secondary structure and, where available, data derived from the structural context, such as changes in hydrogen bonding or atomic crowding. However, such data is used in a qualitative, rule-based manner in the final prediction (1), and the tools which are most widely used in the clinical setting do not specifically address the quantitative effects of missense variants on protein stability. Nevertheless, these effects can be calculated where there is an experimental or modeled 3D structure for the protein of interest, and programs such as FoldX (2), Rosetta (3, 4) or other computational methods have been widely used by structural biologists to investigate the effects of missense variants on protein folding and stability (5, 6). Despite this, few studies have sought to address whether there is a direct clinical application of such an approach, i.e. whether pathogenic and benign variants can be distinguished on the basis of their predicted effects on thermodynamic stability.

The potential utility of protein stability data towards the analysis of missense variants has recently been demonstrated in studies of the Lynch syndrome protein, MSH2 (7), and in phenylalanine hydroxylase (PAH) (8), in which pathogenic variants result in phenylketonuria. Both these studies combined *in silico* analysis with extensive functional analysis of a number of *MSH2* and *PAH* variants; however, resources for the latter are unlikely to be routinely available in clinical genetics laboratories. We have therefore asked whether *in silico* analysis, based predominantly on the predicted effects of missense variants on protein stability, can help discriminate between pathogenic and benign variation in the context of clinical testing of the *MEN1* gene.

Pathogenic variants in the *MEN1* gene cause Multiple Endocrine Neoplasia type I, an autosomal dominant disorder, in which patients develop neoplastic lesions in various endocrine tissues, principally the parathyroids, pituitary and pancreas (9). Pathogenic variants may either be inherited or acquired, but in both cases development of disease requires loss of heterozygosity consistent with a role for the *MEN1* gene product as a tumor suppressor. The most common presenting feature of MEN1 is hyperparathyroidism, which occurs in ~95% of patients due to tumors of the parathyroid gland; however, tumors are also frequently observed in the pancreatic islets (40-70%) and pituitary (30-40%) (10). Patients may also develop tumors of the adrenal cortex, carcinoid tumors and non-endocrine tumors, including lipomas, angiofibromas, collagenomas and meningiomas (11), resulting in a range of clinical symptoms which may overlap with other diseases of different genetic etiology (12-14). This overlap presents one of the key problems in assessing genetic variants in cases of MEN1. While a large number of pathogenic variants in *MEN1* have been reported, genetic testing continues to uncover novel missense substitutions which require assessment of their potential pathogenicity. A further confounding issue is the often later onset of disease, with reported age-related penetrance of 10-43% at 20 years and 81- 94% by 50 years (10, 15), which may lead to apparent non-segregation of a variant with disease within a family pedigree.

The identification of a genetic etiology has important implications for the patient and for their family members. With the exception of pituitary neuroendocrine tumors, MEN1-associated tumors are usually multiple and treatment is therefore challenging, requiring a multi-disciplinary team of experts to reduce morbidity and mortality (16). The identification of the familial disease-causing variant enables the identification of carriers when they are still asymptomatic. Clinical surveillance in these individuals allows early recognition of the clinical manifestations and therapeutic intervention. For example, primary hyperparathyroidism often remains asymptomatic in many patients but prolonged hypercalcemia usually results in bone loss and/or nephrocalcinosis (17).

Approximately 20% of the variants identified in the *MEN1* gene are missense variants (18). The standards and guidelines published by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) describe a framework for the classification of sequence variants (19). Adjustments to this framework for the interpretation of *MEN1* missense variants has been proposed (20). However both agree that variants of uncertain significance should not be used to guide the clinical management of patients. This could lead to an under-diagnosis of MEN1 and a lost opportunity for screening at-risk relatives. For these reasons, methods to assist the classification of variants in *MEN1* would be of clinical value. The availability of a number of experimental structures for Menin, the *MEN1* gene product, raises the possibility that structural analysis may provide such clinical utility.

We report here that thermodynamic analysis of *MEN1* variants *in silico* provides a very strong positive predictive value for pathogenicity, thereby helping to assess the impact of novel missense variants on protein function and potentially allowing its use as an aid to variant classification, and discuss briefly the scope for wider application of this approach to other diseases.

METHODS

Variant groups, transcripts and numbering. Previously-reported missense SNVs in *MEN1* were downloaded from the Human Gene Mutation Database, Professional version (HGMD Pro) (21), the Genome Aggregation Database (gnomAD) (22) and the Sydney Genomics Collaborative Database (SGCD) (23). For the purposes of this analysis, variants were divided into groups as follows: pathogenic: DM ('disease mutation') class variants reported in HGMD Pro but not in gnomAD or SGCD (n=162); benign: variants reported in gnomAD or SGCD but not as DM class in HGMD Pro (n=206); uncertain: variants reported as DM in HGMD Pro and present in gnomAD and/or SGCD (n=14). Different nucleotide substitutions resulting in the same coding change were regarded as a single missense substitution. In addition to these previously-reported variants, analysis was performed on seven novel missense variants: H46P; A164P; L175P; A345P; I360F; F364S; and G419D (see Table 1 for details). These variants were identified in our laboratory as part of the NHS (England) Genetic Testing service for rare inherited diseases. The patients tested fulfilled the criteria for a clinical diagnosis of MEN1 (10), presenting with at least two out of the three main MEN1-associated endocrine lesions or one typical MEN1-associated tumour and a first-degree relative with MEN1 or MEN1-associated lesion at a young age. For patients with a family history, the relevant variants (H46P, A164P, I360F and F364S) were shown to co-segregate with disease in the family.

Table 1: Details of 7 novel missense variants in *MEN1*

All variants refer to *MEN1* transcript NM_130799.2, protein NP_570711.1 (610 amino acid isoform)

variant #	**1**	**2**	**3**	**4**	**5**	**6**	**7**
HGVS c. notation	c.137A>C	c.490G>C	c.524T>C	c.1033G>C	c.1078A>T	c.1091T>C	c.1256G>A
HGVS p. notation	p.(His46Pro)	p.(Ala164Pro)	p.(Leu175Pro)	p.(Ala345Pro)	p.(Ile360Phe)	p.(Phe364Ser)	p.(Gly419Asp)
genomic variant (GRCh37/hg19)	chr11:64577445T>G	chr11:64575527C>G	chr11:64575493A>G	chr11:64573720C>G	chr11:64573214T>A	chr11:64573201A>G	chr11:64572600C>T
reported in gnomAD?	no	no	no	no	no	no	no
SIFT prediction	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging
PROVEAN prediction	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious	Deleterious
PolyPhen prediction	probably damaging	probably damaging	probably damaging	probably damaging	probably damaging	probably damaging	probably damaging
REVEL score	0.894	0.925	0.965	0.909	0.883	0.945	0.912

***In silico* mutagenesis and thermodynamic analysis**. Prior to *in silico* mutagenesis, sidechain repair and energy minimization was performed on all 31 Menin chains in isolation, using the RepairPDB function of the FoldX modeling suite, version 4 (33). The FoldX BuildModel function was then used to introduce individual substitutions into each of the repaired PDB structures. Of the 389 unique missense variants, 338 were covered by at least one PDB structure (pathogenic, n=161; benign, n=161; uncertain, n=9; novel, n=7). For each substitution, FoldX reported a change in free energy ($\Delta\Delta G$) resulting from the substitution; from this, an average $\Delta\Delta G$ value was calculated for each variant across all structures containing the relevant position. In total, all 31 structures were used for 308/338 variants (mean for all variants = 29), whereas due to differences in coverage in individual PDB files, analysis was possible using only a single structure for 7 variants. A full list of variants, the number of PDB structures analysed for each and average $\Delta\Delta G$ values for each variant is shown in Supplementary table 3. All structures were visualized in PyMOL (34).

Calculation of solvent accessibility. The absolute area accessible to solvent (ASA) was calculated on a residue-by-residue basis for 7 representative structures of Menin using DSSP (35, 36) version 3.0.0 (37). After calculating an average ASA value for each residue, relative solvent accessibility (RSA) was derived using the theoretical scale described by Tien *et al*. (38). A list of structures used for DSSP analysis is included in Table 2.

RESULTS

Pathogenic variants in MEN1 are predicted to be destabilizing. Over 30 crystal structures have previously been reported for Menin; most of these contain the protein in isolation or bound to a small (drug) ligand, while others show Menin in complex with peptides from JunD, KMT2A or PSIP (Figure 1; Table 2). Although all structures have been derived from expression of full-length (or near full-length) Menin, a number of regions remain unresolved in crystal structures. These regions predominantly lie in the C-terminal of the protein and correspond to stretches of predicted intrinsic disorder (39) in the protein (Figure 1C, D), presumably resulting in high mobility within crystals. Interestingly, while these regions contain a similar

Table 2: **MEN1 crystal structures used in FoldX analysis**: A total of 29 PDB structures containing 31 Menin chains were used for thermodynamic analysis using FoldX; 7 representative structures were also used for relative solvent accessibility (RSA) analysis.

PDB ID	Title	resolution (Å)	release date	Menin chain(s)	used for RSA analysis?	Reference
3u84	Crystal structure of Human Menin	2.50	15/02/2012	A, B	Yes (chain A)	25
3u85	Crystal structure of human menin in complex with MLL1 (KMT2A)	3.00	15/02/2012	A	Yes	
3u86	Crystal structure of human menin in complex with JunD	2.84	15/02/2012	A		
3u88	Crystal structure of human menin in complex with MLL1 (KMT2A) and LEDGF (PSIP)	3.00	15/02/2012	A, B	Yes (chain B)	
4gpq	Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia	1.46	19/09/2012	A		26
4gq3	Human menin with bound inhibitor MI-2	1.56	19/09/2012	A		
4gq4	Human menin with bound inhibitor MI-2-2	1.27	19/09/2012	A		
4gq6	Human menin in complex with MLL (KMT2A) peptide	1.55	19/09/2012	A		
4i80	Crystal structure of human menin in complex with a high-affinity macrocyclic peptidomimetics	3.10	06/03/2013	A	Yes	27
4og3	Human menin with bound inhibitor MIV-3R	2.01	05/03/2014	A		28
4og4	Human menin with bound inhibitor MIV-3S	1.45	05/03/2014	A		
4og5	Human menin with bound inhibitor MIV-5	1.63	05/03/2014	A		
4og6	Human menin with bound inhibitor MIV-4	1.49	05/03/2014	A		
4og7	Human menin with bound inhibitor MIV-7	2.08	05/03/2014	A		
4og8	Human menin with bound inhibitor MIV-6R	1.53	05/03/2014	A		
4x5y	Menin in complex with MI-503	1.59	15/04/2015	A		29
4x5z	Menin in complex with MI-136	1.86	15/04/2015	A		
5db0	Menin in complex with MI-352	1.50	30/03/2016	A		30
5db1	Menin in complex with MI-336	1.86	30/03/2016	A		
5db2	Menin in complex with MI-389	1.54	30/03/2016	A		
5db3	Menin in complex with MI-574	1.71	30/03/2016	A		
5dd9	Menin in complex with MI-326	1.62	09/09/2015	A		31
5dda	Menin in complex with MI-333	1.83	09/09/2015	A	Yes	
5ddb	Menin in complex with MI-319	1.54	09/09/2015	A		
5ddc	Menin in complex with MI-2-3	1.62	06/07/2016	A		
5ddd	Menin in complex with MI-836	2.14	09/09/2015	A		
5dde	Menin in complex with MI-859	1.78	09/09/2015	A		
5ddf	Menin in complex with MI-273	1.66	09/09/2015	A	Yes	
6b41	Menin bound to M-525	2.61	24/01/2018	A	Yes	32

***In silico* mutagenesis and thermodynamic analysis**. Prior to *in silico* mutagenesis, sidechain repair and energy minimization was performed on all 31 Menin chains in isolation, using the RepairPDB function of the FoldX modeling suite, version 4 (33). The FoldX BuildModel function was then used to introduce individual substitutions into each of the repaired PDB structures. Of the 389 unique missense variants, 338 were covered by at least one PDB structure (pathogenic, n=161; benign, n=161; uncertain, n=9; novel, n=7). For each substitution, FoldX reported a change in free energy ($\Delta\Delta G$) resulting from the substitution; from this, an average $\Delta\Delta G$ value was calculated for each variant across all structures containing the relevant position. In total, all 31 structures were used for 308/338 variants (mean for all variants = 29), whereas due to differences in coverage in individual PDB files, analysis was possible using only a single structure for 7 variants. A full list of variants, the number of PDB structures analysed for each and average $\Delta\Delta G$ values for each variant is shown in Supplementary table 3. All structures were visualized in PyMOL (34).

Calculation of solvent accessibility. The absolute area accessible to solvent (ASA) was calculated on a residue-by-residue basis for 7 representative structures of Menin using DSSP (35, 36) version 3.0.0 (37). After calculating an average ASA value for each residue, relative solvent accessibility (RSA) was derived using the theoretical scale described by Tien *et al*. (38). A list of structures used for DSSP analysis is included in Table 2.

RESULTS

Pathogenic variants in MEN1 are predicted to be destabilizing. Over 30 crystal structures have previously been reported for Menin; most of these contain the protein in isolation or bound to a small (drug) ligand, while others show Menin in complex with peptides from JunD, KMT2A or PSIP (Figure 1; Table 2). Although all structures have been derived from expression of full-length (or near full-length) Menin, a number of regions remain unresolved in crystal structures. These regions predominantly lie in the C-terminal of the protein and correspond to stretches of predicted intrinsic disorder (39) in the protein (Figure 1C, D), presumably resulting in high mobility within crystals. Interestingly, while these regions contain a similar

distribution of benign variants as that seen in the protein as a whole, pathogenic variants are rare in regions of predicted disorder (Figure 1D); however, we cannot rule out the possibility that the lack of pathogenic variants in disordered regions is due to reporting bias towards variants which lie close to those already known. As a result of this distribution of pathogenic variants almost entirely within ordered regions, the vast majority (161/162) are covered by one or more PDB entries and are thus amenable to structural analysis.

Figure 1

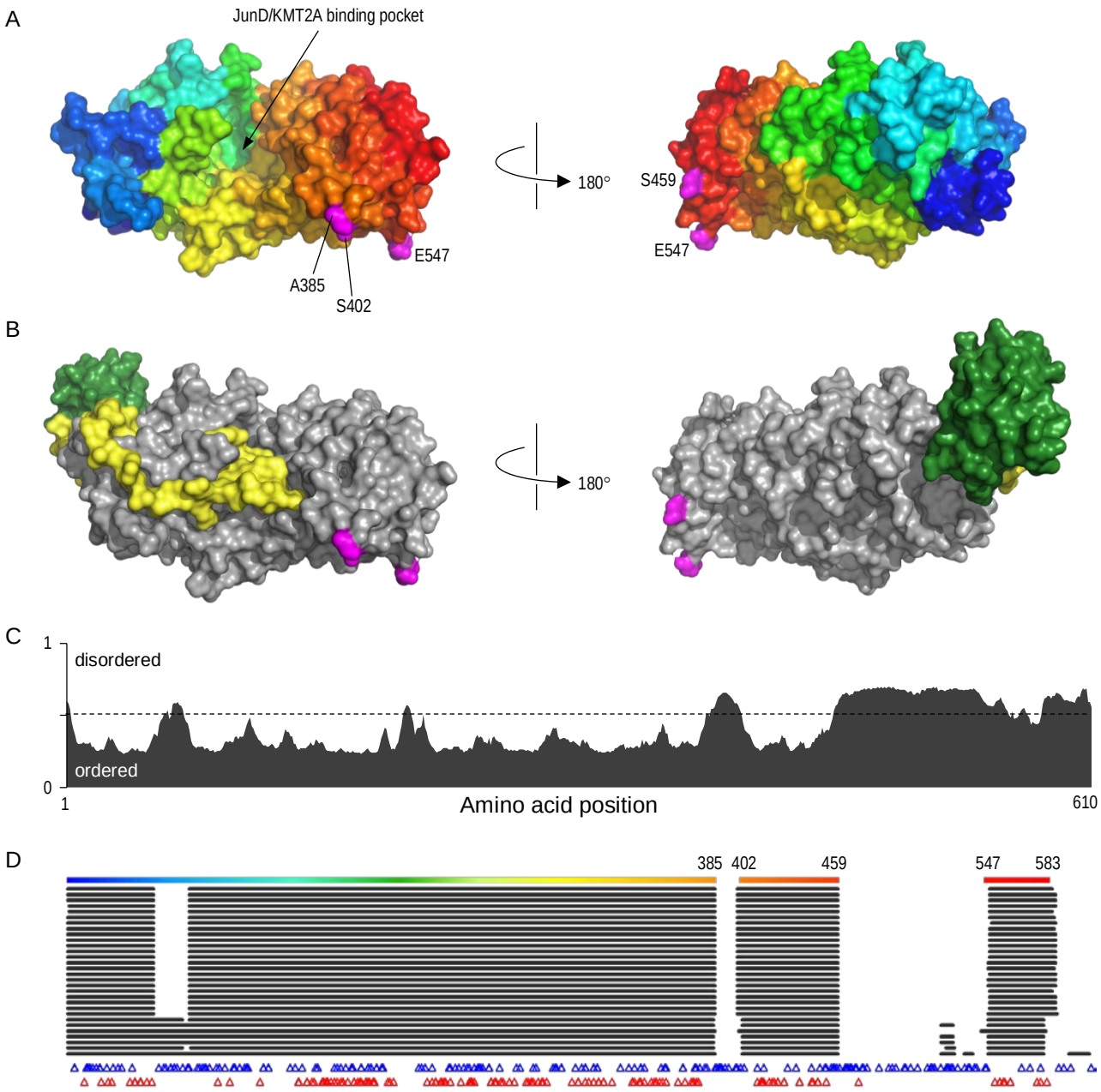

Figure 1: Structure and disorder in Menin. A) The structure of Menin, as represented by PDB entry 3u88 chain A; protein surface is coloured from blue, N-terminal to red, C-terminal; the position of the binding pocket for JunD and KMT2A is indicated; numbered residues coloured magenta indicate positions flanking disordered loops which are not resolved in the crystal structure. B) Menin (grey) in complex with KMT2A (yellow) and PSIP (green), as determined in PDB 3u88; note that while one end

of KMT2A occupies the binding pocket, interaction with PSIP and other regions of KMT2A extends over a wider region of the Menin surface. C) Probability of intrinsic disorder in Menin, as calculated by the MetaDisorder predictor, plotted against amino acid position; extended regions of probability >0.5 are considered to be disordered. D) Coverage of Menin residues in the 31 PDB structures used in this analysis, aligned against amino acid position as in part C. The top line shows coverage in PDB 3u88A, coloured as in 1A; numbering indicates residues flanking unstructured regions missing from the crystal structure. Below this, black horizontal lines show coverage for the 30 remaining PDB structures, while positions of benign and pathogenic variants are indicated by blue or red triangles indicate respectively. Note that regions of predicted intrinsic disorder are absent from the majority, if not all crystal structures, consistent with greater mobility of these residues within the crystal, and that pathogenic variants are almost entirely absent from these regions.

The overall structure of Menin is highly comparable within all reported PDB structures (alignment to PDB 6b41 yields an average RMSD value of 0.65 Å, range 0.55-1.10 Å). Moreover, there is no significant effect of ligand binding on Menin structure (Figure 2). Since different PDB files contain slightly different numbers of amino acids but there are no obvious structural outliers, all available structures were used for thermodynamic analysis of missense variants *in silico* using FoldX.

Figure 2

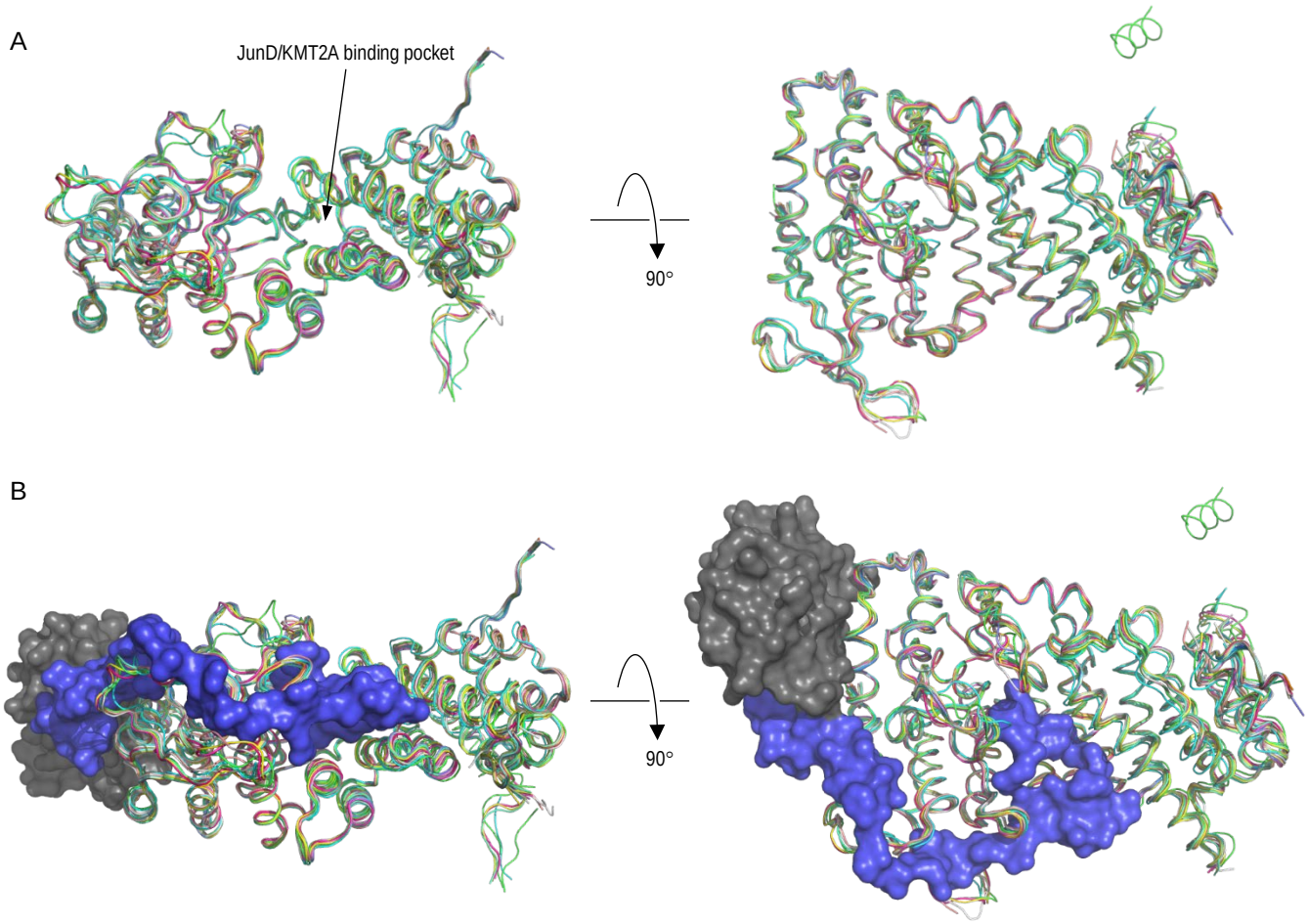

Figure 2: Alignment of Menin structures. A) The α carbon atoms of the 31 Menin structures used in this study were aligned to that of PDB 6b41; each chain is shown in ribbon format, coloured by PDB and chain identifier; the position of the JunD/KMT2A binding pocket is indicated; the short helix visible at the top right of the rotated figure corresponds to residues 596-608 at the extreme C-terminal of Menin, which were resolved only in PDB 3u84 chain A. B) As A, but superimposed with the structures of MLL (blue) and PSIP (grey) from PDB 3u88.

Variant groups were highly distinguishable by their predicted effect on thermodynamic stability, as represented by average $\Delta\Delta G$ value calculated across all structures, with most putatively benign (gnomAD and SGCD) variants having little or no effect (average $\Delta\Delta G$ for all variants, 1.13 kcal/mol; SD, 1.46 kcal/mol),

whereas pathogenic (HGMD) variants were predicted to be strongly destabilizing (average $\Delta\Delta G$, 5.06 kcal/mol; SD 4.25 kcal/mol) (Figure 3A). Notably, the seven novel missense variants were also predicted to be strongly destabilising (average $\Delta\Delta G$, 7.67 kcal/mol; SD 3.14 kcal/mol). Analysis of $\Delta\Delta G$ values for individual PDB structures showed a similar separation of putative benign and pathogenic variant groups, with the vast majority of variants falling into a similar range for all structures (Figure 3B). We further compared the effect at multi-allelic sites where different benign and pathogenic missense variants occur at the same position. Analysis of 27 benign and 23 pathogenic variants co-occurring at 22 residues again showed that the difference between the two groups was highly significant (p = 0.0002), and that pathogenic missense changes were more strongly destabilizing than benign ones at the same position (average $\Delta\Delta G$ values by group = 6.81 kcal/mol and 2.18 kcal/mol respectively) (Figure 3C).

Figure 3

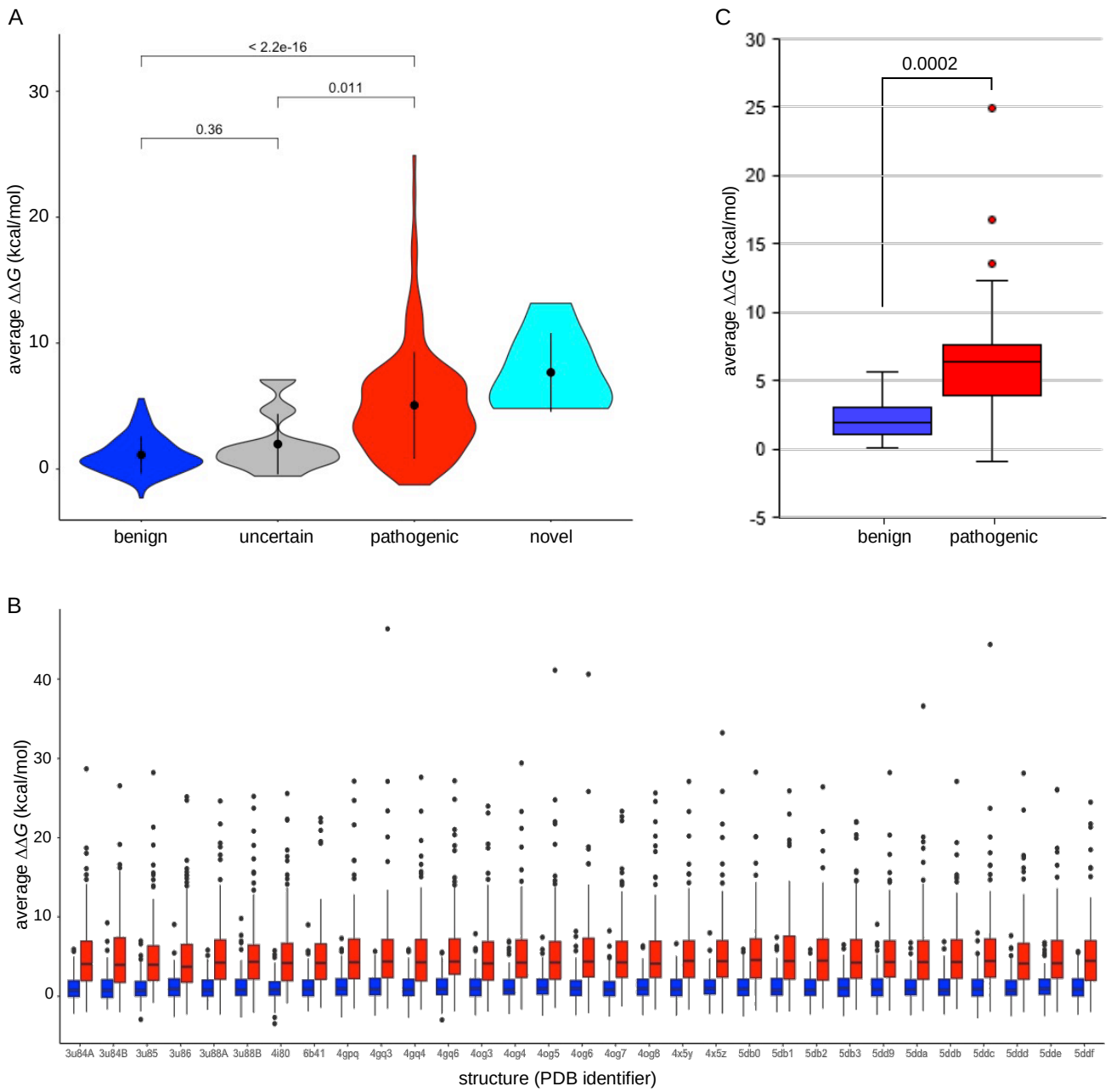

Figure 3: Pathogenic variants are predicted to destabilise Menin structure. A) *In silico* mutagenesis and thermodynamic analysis of Menin variants; for each variant, the average change in thermodynamic stability, $\Delta\Delta G$, was calculated across all structures contained the relevant residue, then plotted by variant group; black circles and vertical lines within each data area represent median and upper and lower quartiles respectively. Numbering above data points shows p values (Student's

Two-tailed t-test) between groups as indicated. B) $\Delta\Delta G$ values for benign (blue) and pathogenic (red) variant groups calculated for 31 individual PDB structures as shown on the x-axis. C) Average $\Delta\Delta G$ values for benign and pathogenic variants occurring at the same amino acid position (residues with one benign and one pathogenic variant, n=16; residues with two benign and one pathogenic variants, n=5; residues with one benign and two pathogenic variants, n=1); coloured boxes show the range between upper and lower quartiles; horizontal lines within each data box show median value; data points are shown for outliers only. The difference in the average $\Delta\Delta G$ value between groups was highly significant (p=0.0002).

If variants which destabilize Menin structure do indeed have a greater tendency to be pathogenic, it might be expected that variants most frequently observed in the general population would have the least destabilizing effect. This appears to be the case, as variants with the highest population frequency had average predicted $\Delta\Delta G$ values in the range -1 to +1 (Figure 4); as the error in FoldX calculations is approximately ±0.8 kcal/mol (2), this suggests little or no effect of these variants on protein stability. Notably, those variants which have also been observed in an aging healthy population, as represented by the SGCD cohort (median age, 80-85 years) and are therefore most likely to be truly benign, all occur within this range of $\Delta\Delta G$ values. This group includes the only commonly-occurring missense *MEN1* variant, R171Q, which has an average $\Delta\Delta G$ value of 0.15 kcal/mol. Conversely, we note that some variants reported in gnomAD have $\Delta\Delta G$ values >4 kcal/mol, and in fact 2/9 of these variants (S38P, D315Y) have also been reported as disease-causing in HGMD Pro. This may reflect the confounding effect of late onset of symptoms in MEN1 on apparent constraint against coding variation, whereby some variants reported in gnomAD may in fact lead to disease in later life.

Figure 4

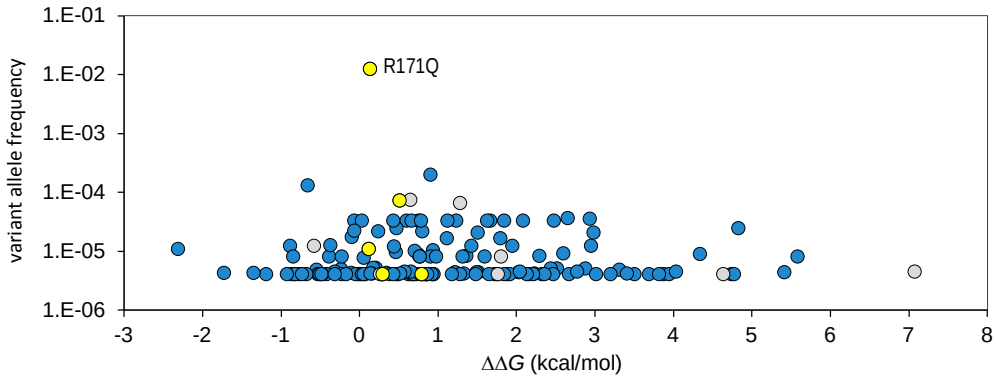

Figure 4: Population frequency of *MEN1* variants. The frequency of benign and uncertain missense variants in the gnomAD database plotted against $\Delta\Delta G$ value; blue fill: variants occurring in the gnomAD database only; yellow fill: variants reported in the gnomAD and SGCD databases; grey fill: variants in both gnomAD and HGMD Pro (DM class) databases. In cases where different nucleotide substitutions give rise to the same amino acid change, frequency is shown as a total for all variant alleles.

Most pathogenic variants are buried in the Menin structure. To examine whether there are differences in the spatial distribution of benign and pathogenic variants, we calculated the relative solvent accessibility (RSA) of wild-type residues at all positions of missense substitutions. This showed that while positions of benign variants are distributed throughout the volume of the protein, 86.3% of pathogenic variants occur in solvent-inaccessible (i.e. buried) regions of RSA <0.2 (Figure 5A). Notably, this is also true for the 7 novel variants, 6 of which had an RSA value <0.02. Plotting RSA against $\Delta\Delta G$ showed that variants at buried positions were also likely to be the most strongly destabilizing to protein structure (Figure 5B). Nevertheless, we observed that a significant number of pathogenic variants exhibited both accessibility to solvent (RSA>0.2) and relatively low $\Delta\Delta G$. Mapping the positions of solvent-accessible variants onto the surface of Menin showed that, as for distribution throughout the internal volume of the protein, benign variants

tended to be distributed across the surface. In contrast, pathogenic variants appeared to occur in clusters, one of which corresponds to binding surfaces for JunD, KMT2A and PSIP (Figure 5C, D), while another occurs on the opposite surface of Menin to the JunD binding pocket. It is possible therefore that the latter region represents the site of an as-yet uncharacterized interaction of Menin. As described above, 6/7 novel missense variants occur at positions which are buried in the interior of the protein, whereas the only solvent-accessible variant, H46A, occurs at the interface with KMT2A and presumably acts to impair this interaction (Figure 5E).

Figure 5

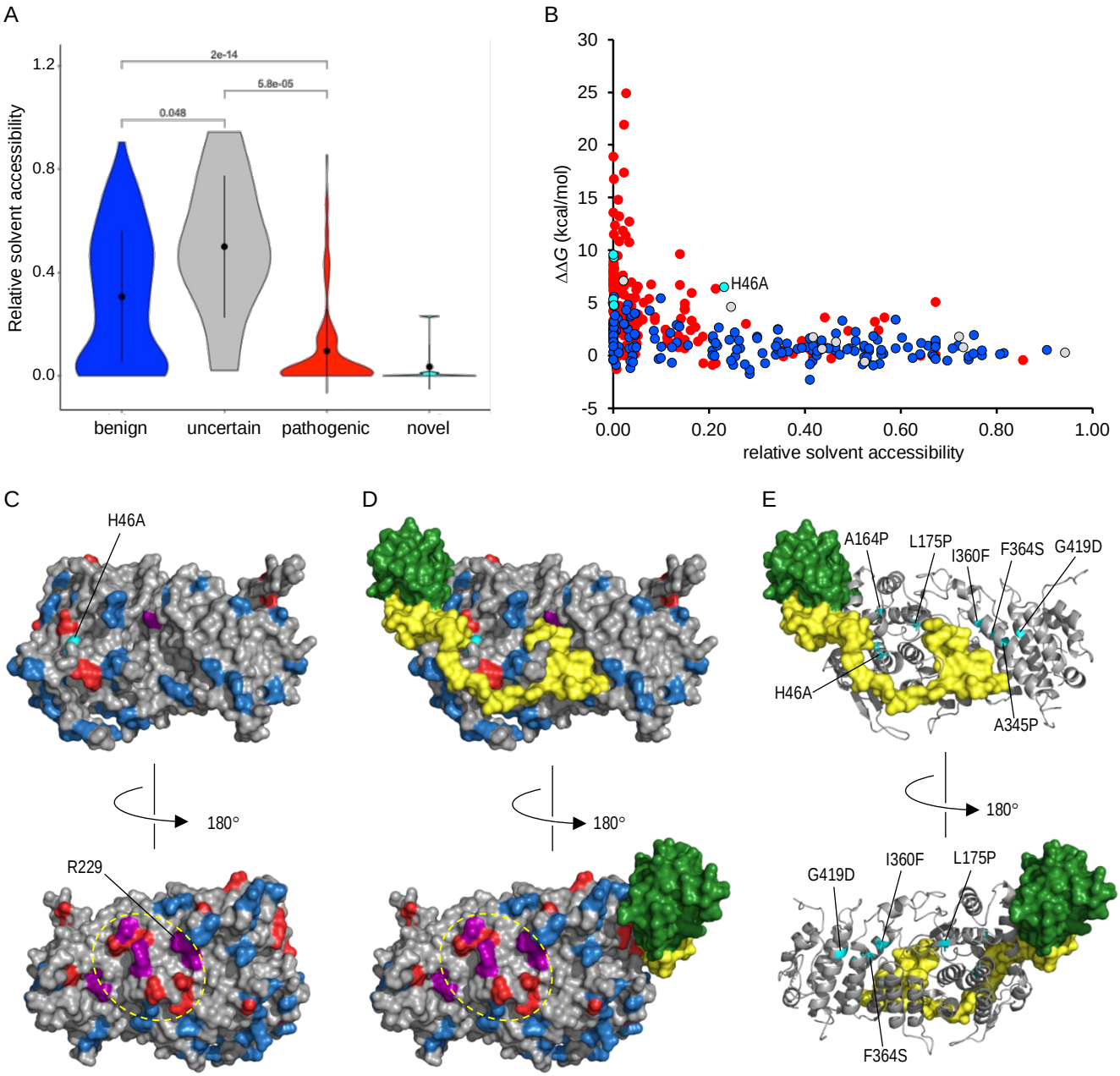

Figure 5: Molecular distribution of pathogenic and benign variants. A) Relative solvent accessibility was calculated for each variant group; black circles and vertical lines within each data area represent median and upper and lower quartiles respectively. Numbering above data points shows p values (Student's Two-tailed t-test) between groups as indicated. B) Buried pathogenic variants are predicted to be the most destabilising to Menin structure; note that 6/7 of the novel missense variants reported

here are deeply buried within the protein (RSA < 0.02), while only novel variant H46A is solvent accessible. C, D) Surface distribution of solvent-accessible variants. The surface of Menin (grey), either alone (C) or in complex (D) with KMT2A (yellow) and PSIP (green) shows all variants with RSA>0.2: blue, benign; red, pathogenic; purple colouring show positions at which different pathogenic and benign variants have been observed; the novel H46A variant is coloured cyan. The broken yellow oval indicates a cluster of pathogenic variants which may constitute an as yet unidentified interface for protein-protein interactions. E) Menin is shown as a grey ribbon; novel missense variants are coloured cyan with sidechains displayed in stick format; KMT2A and DSIP are shown as in D.

To investigate the effects of protein interactions on the thermodynamic effects of *MEN1* variants further, we compared $\Delta\Delta G$ values for variants in PDB structure 3u88 (Menin complexed with KMT2A and PSIP peptides) by analysis both of Menin chains in isolation (chains A, B) and complexed to KMT2A and PSIP. As expected, regions of decreased solvent accessibility in the complexes aligned with residues annotated as forming protein-protein contacts (Figure 6). However, the presence of bound peptides had little effect on $\Delta\Delta G$ values of benign variants, indicating that these have a neutral effect on protein binding. Conversely, protein binding had a large effect on $\Delta\Delta G$ values of a number of pathogenic variants; again, these predominantly occurred at or close to protein interfaces, indicating that these variants are likely to have a direct effect on ligand binding by Menin.

Figure 6

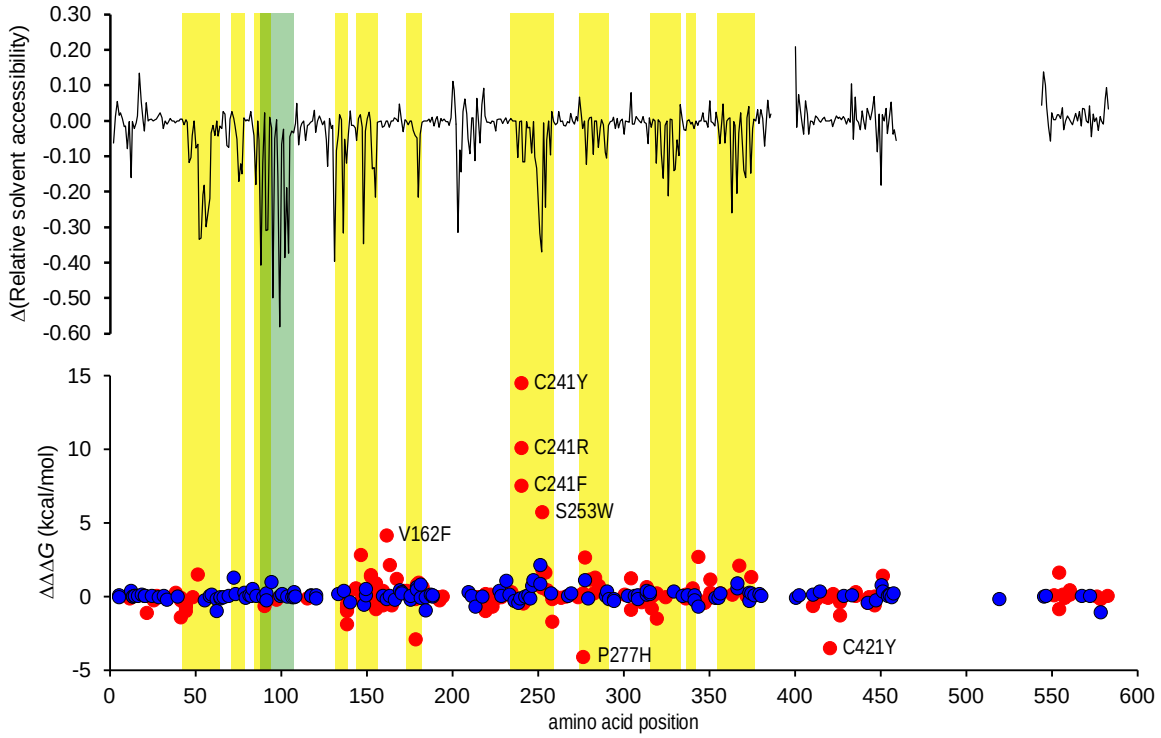

Figure 6: Effect of protein-protein interaction on ΔΔ*G*. Analysis of solvent accessibility and thermodynamic effect of variants was performed on PDB 3u88 (Menin:KMT2A:DSIP complex), both on Menin chains in isolation (chains A, B) and as part of the complex. The upper graph shows the average difference in solvent accessibility by position in the complexed and isolated Menin chains respectively (ΔRSA = RSA [complex] – RSA [isolated]); the lower graph shows the equivalent difference in average ΔΔ*G* value at each position (i.e. ΔΔΔ*G*); data points are labelled for variants where ΔΔΔ*G* 3 kcal/mol; background shading indicates positions of Menin residues forming contacts with KMT2A (yellow) or DSIP (green) in PDB 3u88.

Destabilizing variants reduce levels of functional Menin protein. Previous reports studying the effects of missense variants on levels of functional Menin within the cell have shown that pathogenic variants have a tendency to increase protein turnover and/or reduce the steady-state level of protein, while benign variants

tend to have no such effect (40, 41). We correlated the previously-reported effects of variants on levels of steady-state protein with average $\Delta\Delta G$ values, and observed that variants which were predicted to be strongly destabilizing *in silico* ($\Delta\Delta G$ >3 kcal/mol) exhibited significantly lower levels of steady-state protein in cell-based assays (p=0.0001,Figure 7), consistent with the hypothesis that variants with high $\Delta\Delta G$ values reduce the biological activity of Menin.

Figure 7

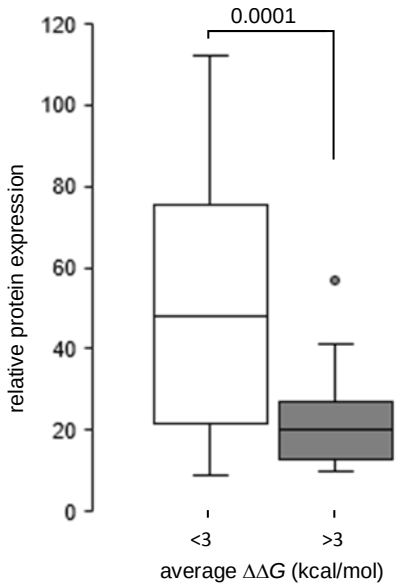

Figure 7. Predicted thermodynamic stability correlates with observed expression. A) Steady-state expression levels have been reported for a number of Menin variants; relative expression level data was sorted into two groups according to $\Delta\Delta G$ value as calculated in this study (neutral or weakly destabilising: $\Delta\Delta G$ <3 kcal/mol [n=14]; strongly destabilising: >3 kcal/mol [n=27]); boxes show the range between upper and lower quartiles; horizontal lines within each data box show median value; data points are shown for outliers only. The difference in relative expression between the two groups was highly significant (p=0.0001).

Can ΔΔG value be used as an aid to variant classification? To evaluate the clinical validity of $\Delta\Delta G$ values, we performed Receiver Operating Characteristic (ROC) curve analysis for the groups of benign and pathogenic variants and compared the results with the outputs from a number of commonly-used phenotypic predictions tools: SIFT (42), PolyPhen (43) and REVEL (44). All methods yielded areas under the curve (AUC) of 0.819-0.864, indicating that all have clinical validity (Figure 8A). However $\Delta\Delta G$ analysis resulted in the highest specificity and lowest sensitivity. Values of $\Delta\Delta G > 3$ kcal/mol are generally regarded as being strongly destabilizing (45); taking this as a threshold for variant classification gives sensitivity and specificity of 67.1% and 89.4% (positive predictive value, 86.4%) while setting a more conservative threshold of $\geq$4 kcal/mol yields increased the specificity to 95.0%, though with a concomitant loss of sensitivity (54.0%; positive predictive value, 90.6%). A marginal increase in positive predictive value (PPV) could be obtained by combining $\Delta\Delta G$ thresholds with a cut-off in the REVEL score of 0.7, which has been reported to exclude 95% of false positive calls (46), yielding PPV's of 87.7% at $\Delta\Delta G \geq 3$ kcal/mol and 91.5% at $\Delta\Delta G \geq 4$ kcal/mol. Notably, all seven novel missense variants reported here cluster within the upper right quadrant (Figure 8B), consistent with a severe impact on protein stability and suggesting that $\Delta\Delta G$ values can potentially be used to provide evidence towards variant classification in MEN1. Full data for average $\Delta\Delta G$ and RSA values, and phenotypic predictions for all previously reported *MEN1* missense variants is shown in Table 3.

Table 3: Unique missense variants, FoldX analysis (DDG) and relative solvent accessibility (RSA): Details are shown for unique missense variants in pathogenic (n=161), benign (n=161) and uncertain (n=9) groups as defined in Materials & Methods. Where the source database included gnomAD, frequency is shown for the variant allele. Phenotypic predictions for each variant show prediction and probability data for PolyPhen2, prediction and score for SIFT and score for REVEL prediction. Results of thermodynamic analysis are shown as average DDG value and standard deviation, derived from FoldX calculation using the number of PDB structures indicated (#PDBs). 'Protein interaction' columns indicate residues annotated in relevant PDB entries as interacting directly either with JunD and/or KMT2A, both of which bind to the JunD binding pocket of Menin, or to PSIP. Results of relative solvent accessibility (RSA) analysis are shown as average values derived from DSSP analysis of 7 representative structures.

Variant details and group					source database				Phenotypic predictions					ΔΔ*G* (kcal/mol)			protein interaction		
position	aa_ref	aa_alt	variant	analysis group	HGMD Pro	gnomAD	allele_freq	SGCD	pred_PolyPhen	prob_PolyPhen	pred_SIFT	score_SIFT	REVEL score	ave	SD	# PDBs	JunD/KMT2A	PSIP	RSA_ave
6	A	P	A6P	benign		Y	4.69E-06		possibly damaging	0.582	BENIGN	0.260	0.659	-0.5340	0.3238	31			0.4164
6	A	S	A6S	benign		Y	9.37E-06		benign	0.053	BENIGN	0.700	0.486	0.4777	0.2745	31			0.4164
12	P	L	P12L	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.892	3.0160	0.4390	31			0.2767
13	L	P	L13P	benign		Y	4.31E-06		probably damaging	0.952	PATHOGENIC	0.000	0.952	5.4231	1.2934	31			0.0995
14	R	C	R14C	benign		Y	4.27E-06		probably damaging	0.931	PATHOGENIC	0.010	0.829	1.2309	0.5364	31			0.4338
14	R	S	R14S	benign		Y	4.27E-06		possibly damaging	0.542	BENIGN	0.130	0.797	0.5869	0.4838	31			0.4338
15	S	C	S15C	benign		Y	4.26E-06		probably damaging	0.927	BENIGN	0.050	0.617	1.5295	0.4617	31			0.2083
15	S	F	S15F	benign		Y	3.19E-05		possibly damaging	0.904	PATHOGENIC	0.040	0.584	1.8519	1.0156	31			0.2083
17	D	N	D17N	benign		Y	3.19E-05		benign	0.344	BENIGN	0.410	0.634	0.7679	0.2995	31			0.4027
19	V	M	V19M	benign		Y	4.20E-06		probably damaging	0.982	PATHOGENIC	0.010	0.850	0.3578	0.7024	31			0.0000
21	R	H	R21H	benign		Y	8.41E-06		possibly damaging	0.731	BENIGN	0.070	0.592	0.7673	0.2664	31			0.5235
21	R	L	R21L	benign		Y	4.21E-06		benign	0.398	BENIGN	0.170	0.642	-0.0928	0.2754	31			0.5235
21	R	S	R21S	benign		Y	1.93E-04		benign	0.166	BENIGN	0.560	0.577	0.9204	0.4025	31			0.5235
22	L	R	L22R	pathogenic	Y				probably damaging	0.980	PATHOGENIC	0.000	0.914	4.1552	1.2317	31			0.0043
25	A	V	A25V	benign		Y	7.42E-06		benign	0.155	BENIGN	0.280	0.547	0.0684	0.3997	31			0.3389
26	E	K	E26K	pathogenic	Y				benign	0.309	BENIGN	0.090	0.826	2.7855	0.6609	31			0.0096
28	G	A	G28A	pathogenic	Y				benign	0.043	BENIGN	0.690	0.638	-0.0115	0.2675	31			0.6415
29	R	G	R29G	benign		Y	4.23E-06		benign	0.007	BENIGN	0.380	0.454	1.5044	0.3584	31			0.2430
32	P	S	P32S	benign		Y	3.42E-05		probably damaging	1.000	PATHOGENIC	0.000	0.848	2.9465	0.5281	31			0.0000
34	L	V	L34V	benign		Y	4.31E-06		probably damaging	0.997	PATHOGENIC	0.000	0.836	2.0395	0.4632	31			0.0000
38	S	P	S38P	uncertain	Y	Y	4.40E-06		probably damaging	0.999	PATHOGENIC	0.000	0.951	7.0856	1.1401	31			0.0212
39	L	W	L39W	pathogenic	Y				probably damaging	0.965	PATHOGENIC	0.000	0.973	11.4983	2.1044	31			0.0007
40	V	A	V40A	benign		Y	9.01E-06		possibly damaging	0.697	PATHOGENIC	0.030	0.881	2.6114	0.1572	31			0.0000
42	G	D	G42D	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.911	10.7359	2.8194	31			0.0330
42	G	S	G42S	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.935	5.4387	1.1875	31			0.0330
42	G	V	G42V	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.937	12.6954	1.9046	31			0.0330
45	E	A	E45A	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.940	1.9144	0.3253	31			0.0224
45	E	D	E45D	pathogenic	Y				probably damaging	0.996	PATHOGENIC	0.000	0.902	3.7314	0.5378	31			0.0224
45	E	G	E45G	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.938	3.2448	0.4001	31			0.0224
45	E	K	E45K	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.895	2.3318	1.0020	31			0.0224
45	E	Q	E45Q	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.878	0.0126	0.6065	31			0.0224
45	E	V	E45V	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.959	0.5279	0.3788	31			0.0224
49	A	V	A49V	pathogenic	Y				possibly damaging	0.465	PATHOGENIC	0.020	0.836	1.1529	0.4904	31			0.0487
52	R	G	R52G	pathogenic	Y				possibly damaging	0.618	PATHOGENIC	0.040	0.865	2.3807	0.8073	31	Y		0.4901
56	T	A	T56A	benign		Y	5.04E-06		benign	0.025	BENIGN	0.320	0.482	0.2220	0.6072	8	Y		0.8151
59	P	L	P59L	benign		Y	1.01E-05		probably damaging	0.998	BENIGN	0.380	0.832	0.9508	0.2605	8			0.7195
60	E	Q	E60Q	benign		Y	5.06E-06		possibly damaging	0.596	BENIGN	0.540	0.447	0.1899	0.2162	8			0.5641
63	F	L	F63L	benign		Y	4.99E-06		benign	0.186	BENIGN	1.000	0.651	2.8941	0.5955	8	Y		0.1325
63	F	Y	F63Y	benign		Y	3.19E-05		possibly damaging	0.579	BENIGN	0.990	0.654	1.2459	0.2243	8	Y		0.1325
65	P	H	P65H	benign		Y	4.96E-06		benign	0.321	BENIGN	0.160	0.619	2.5332	0.3993	8	Y		0.3434
65	P	S	P65S	benign		Y	5.00E-06		benign	0.043	BENIGN	0.600	0.534	2.4520	0.4708	8	Y		0.3434
67	P	L	P67L	benign		Y	9.95E-06		benign	0.012	BENIGN	0.480	0.636	0.7152	0.2640	8			0.7082
70	D	E	D70E	benign		Y	4.82E-06		benign	0.026	BENIGN	0.580	0.479	-0.2189	0.2592	8			0.6715

73	G	D	G73D	benign			Y	4.71E-06		benign	0.391	BENIGN	0.620	0.662	3.3262	2.1752	6					0.4784
74	G	D	G74D	benign			Y	3.20E-05		benign	0.009	BENIGN	0.170	0.547	1.6804	1.3179	30					0.6030
79	P	A	P79A	benign			Y	3.20E-05		probably damaging	1.000	PATHOGENIC	0.000	0.947	2.4887	0.1880	31					0.0305
80	V	M	V80M	benign			Y	4.34E-06		probably damaging	0.933	BENIGN	0.120	0.594	-0.3031	0.3893	31					0.0944
82	D	E	D82E	benign					Y	benign	0.038	BENIGN	1.000	0.504	1.4358	0.5530	31					0.2628
83	L	P	L83P	benign			Y	1.68E-05		probably damaging	0.929	PATHOGENIC	0.020	0.839	-0.0824	0.4428	31					0.2552
84	S	P	S84P	benign			Y	4.17E-06		benign	0.020	BENIGN	0.290	0.706	0.6858	0.7237	31					0.4922
86	I	V	I86V	benign			Y	4.10E-06		benign	0.038	BENIGN	1.000	0.520	1.3403	0.1094	31					0.0015
90	Y	C	Y90C	benign					Y	benign	0.346	BENIGN	0.070	0.835	4.2845	0.3601	31					0.0755
91	A	V	A91V	pathogenic		Y				benign	0.056	BENIGN	0.290	0.667	0.4713	0.1697	31		Y			0.4131
92	R	G	R92G	benign			Y	4.03E-06		possibly damaging	0.826	PATHOGENIC	0.010	0.844	2.2437	0.2400	31		Y		Y	0.5057
92	R	H	R92H	benign			Y	8.03E-06		probably damaging	0.933	PATHOGENIC	0.010	0.856	1.3761	0.2515	31		Y		Y	0.5057
95	A	G	A95G	benign					Y	possibly damaging	0.547	PATHOGENIC	0.040	0.731	0.6061	0.0835	31				Y	0.4884
98	R	L	R98L	pathogenic		Y				possibly damaging	0.560	BENIGN	0.120	0.817	0.3349	0.3609	31				Y	0.4630
100	A	V	A100V	benign			Y	1.60E-05		benign	0.082	BENIGN	0.130	0.700	1.1253	0.5296	31					0.3621
101	V	A	V101A	benign			Y	3.99E-06		possibly damaging	0.779	BENIGN	0.070	0.906	3.7036	0.5968	31				Y	0.0361
101	V	I	V101I	benign			Y	3.99E-06		possibly damaging	0.851	BENIGN	0.240	0.711	-1.1771	0.1592	31				Y	0.0361
104	S	A	S104A	benign			Y	3.98E-06		possibly damaging	0.465	BENIGN	0.220	0.797	0.7002	0.3028	31				Y	0.5889
107	P	S	P107S	benign			Y	3.98E-06		benign	0.389	BENIGN	0.750	0.616	0.9495	0.2669	31					0.6577
108	R	G	R108G	benign			Y	7.96E-06		benign	0.007	BENIGN	0.360	0.571	1.3232	0.5531	31					0.3436
108	R	Q	R108Q	benign			Y	3.19E-05		benign	0.344	BENIGN	0.610	0.524	0.6114	0.3788	31					0.3436
109	E	D	E109D	benign			Y	3.98E-06		benign	0.009	BENIGN	0.910	0.292	0.0795	0.2337	31					0.6662
110	G	E	G110E	uncertain		Y	Y	7.96E-06		benign	0.292	BENIGN	0.450	0.671	1.8165	1.1920	31					0.7212
116	E	G	E116G	pathogenic		Y				possibly damaging	0.882	PATHOGENIC	0.010	0.942	2.4131	0.2424	31					0.5535
118	V	A	V118A	benign			Y	3.98E-06		possibly damaging	0.795	PATHOGENIC	0.010	0.942	2.3403	0.1584	31					0.0057
118	V	L	V118L	benign			Y	3.98E-06		possibly damaging	0.591	PATHOGENIC	0.040	0.823	-0.8529	0.3980	31					0.0057
121	V	A	V121A	benign			Y	3.98E-06		probably damaging	0.947	PATHOGENIC	0.010	0.968	2.4811	0.1451	31					0.0000
121	V	I	V121I	benign			Y	3.98E-06		possibly damaging	0.612	BENIGN	0.120	0.773	-0.4038	0.4065	31					0.0000
134	F	L	F134L	benign			Y	3.18E-05		benign	0.349	PATHOGENIC	0.030	0.866	-0.0536	0.2579	31					0.6690
137	R	G	R137G	benign			Y	3.98E-06		probably damaging	0.939	PATHOGENIC	0.010	0.850	0.6389	0.4195	31		Y			0.3843
139	H	D	H139D	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.936	4.3772	0.6484	31					0.1486
139	H	N	H139N	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.867	0.6698	0.3868	31					0.1486
139	H	P	H139P	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.862	4.9157	0.7108	31					0.1486
139	H	Q	H139Q	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.825	1.4896	0.3918	31					0.1486
139	H	R	H139R	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.855	1.0663	0.7935	31					0.1486
139	H	Y	H139Y	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.953	4.4160	1.1166	31					0.1486
141	Q	R	Q141R	benign			Y	3.98E-06		probably damaging	0.994	PATHOGENIC	0.000	0.968	0.8635	0.7206	31					0.0997
144	F	C	F144C	pathogenic		Y				probably damaging	0.977	PATHOGENIC	0.000	0.960	3.6593	0.3648	31					0.0863
144	F	V	F144V	pathogenic		Y				possibly damaging	0.736	PATHOGENIC	0.010	0.973	3.3642	0.3531	31					0.0863
147	I	F	I147F	pathogenic		Y				possibly damaging	0.533	PATHOGENIC	0.020	0.822	6.5569	2.3759	31					0.0515
148	T	P	T148P	pathogenic		Y				probably damaging	0.956	PATHOGENIC	0.040	0.859	3.6064	1.0523	31					0.4410
149	G	R	G149R	benign			Y	1.19E-05		probably damaging	0.992	PATHOGENIC	0.000	0.661	-0.8710	0.5338	31		Y			0.2500
150	T	N	T150N	benign			Y	4.16E-06		benign	0.010	BENIGN	1.000	0.592	-1.7141	0.6731	31					0.2849
150	T	S	T150S	benign			Y	4.17E-06		benign	0.041	BENIGN	0.600	0.588	-1.3414	0.4688	31					0.2849
153	D	E	D153E	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.940	0.7734	1.5967	31		Y			0.1355
153	D	V	D153V	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.953	1.1639	0.9191	31		Y			0.1355
153	D	Y	D153Y	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.941	1.8806	3.0010	31		Y			0.1355
154	S	I	S154I	pathogenic		Y				possibly damaging	0.723	PATHOGENIC	0.010	0.848	1.3508	0.8337	31		Y			0.1770
155	S	F	S155F	pathogenic		Y				benign	0.005	BENIGN	1.000	0.687	-0.6995	2.4066	31		Y			0.1880
156	G	C	G156C	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.932	3.3558	0.9748	31		Y			0.0124
156	G	D	G156D	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.941	13.2142	1.8654	31		Y			0.0124
156	G	R	G156R	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.932	10.8362	2.4296	31		Y			0.0124
156	G	S	G156S	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.972	2.1774	0.6195	31		Y			0.0124
156	G	V	G156V	pathogenic		Y				probably damaging	1.000	PATHOGENIC	0.000	0.926	6.5116	1.6951	31		Y			0.0124
158	A	D	A158D	pathogenic		Y				probably damaging	0.999	PATHOGENIC	0.000	0.972	7.3823	0.8188	31					0.0000
160	A	G	A160G	benign			Y	4.00E-06		benign	0.162	BENIGN	0.100	0.837	2.1783	0.0977	31					0.0011
160	A	P	A160P	pathogenic		Y				probably damaging	0.982	PATHOGENIC	0.010	0.954	5.5304	0.8689	31					0.0011
160	A	T	A160T	pathogenic		Y				possibly damaging	0.670	PATHOGENIC	0.010	0.936	0.8832	0.6791	31					0.0011

162	V	A	V162A	benign		Y	8.00E-06		probably damaging	0.948	PATHOGENIC	0.000	0.966	2.3030	0.1388	31			0.0000
162	V	F	V162F	pathogenic	Y				probably damaging	0.992	PATHOGENIC	0.000	0.957	13.5779	2.6237	31			0.0000
164	A	D	A164D	pathogenic	Y				probably damaging	0.953	PATHOGENIC	0.010	0.971	7.9503	1.4645	31			0.0000
164	A	S	A164S	benign		Y	4.00E-06		possibly damaging	0.857	BENIGN	0.080	0.852	1.7389	0.3372	31			0.0000
165	C	R	C165R	pathogenic	Y				probably damaging	0.982	PATHOGENIC	0.000	0.959	7.1080	1.4289	31			0.0000
165	C	Y	C165Y	pathogenic	Y				probably damaging	0.991	PATHOGENIC	0.000	0.953	9.2796	1.3621	31			0.0000
167	A	V	A167V	benign		Y	3.99E-06		benign	0.058	BENIGN	1.000	0.488	-0.2112	0.6056	31			0.0986
168	L	P	L168P	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.956	6.7526	0.9219	31			0.0796
170	L	V	L170V	benign		Y	3.98E-06		benign	0.352	BENIGN	0.180	0.720	3.5141	0.7089	31			0.0092
171	R	Q	R171Q	benign		Y	1.22E-02	Y	benign	0.444	BENIGN	0.150	0.434	0.1461	0.3020	31			0.5594
171	R	W	R171W	benign		Y	7.08E-05	Y	benign	0.154	PATHOGENIC	0.010	0.657	0.5226	0.4582	31			0.5594
172	D	V	D172V	pathogenic	Y				possibly damaging	0.615	PATHOGENIC	0.000	0.938	2.4490	0.7182	31			0.1095
172	D	Y	D172Y	pathogenic	Y				probably damaging	0.987	PATHOGENIC	0.000	0.943	2.4656	0.9278	31			0.1095
174	H	P	H174P	pathogenic	Y				possibly damaging	0.905	PATHOGENIC	0.000	0.944	5.9444	0.6535	31			0.1639
175	L	R	L175R	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.955	5.9130	0.7482	31			0.0007
176	A	P	A176P	pathogenic	Y				probably damaging	0.983	PATHOGENIC	0.010	0.942	6.8178	0.8335	31			0.0000
176	A	S	A176S	benign		Y	1.99E-05		possibly damaging	0.622	BENIGN	0.100	0.868	1.5213	0.6669	31	Y		0.0000
176	A	T	A176T	benign		Y	3.98E-06		possibly damaging	0.572	PATHOGENIC	0.010	0.943	3.0286	0.8171	31	Y		0.0000
179	E	D	E179D	pathogenic	Y				probably damaging	0.917	PATHOGENIC	0.000	0.893	4.7402	0.6227	31	Y		0.0378
179	E	K	E179K	pathogenic	Y				possibly damaging	0.852	PATHOGENIC	0.000	0.937	6.9708	2.1935	31	Y		0.0378
179	E	Q	E179Q	pathogenic	Y				possibly damaging	0.730	PATHOGENIC	0.000	0.924	2.9444	0.7707	31	Y		0.0378
180	D	A	D180A	pathogenic	Y				probably damaging	0.976	PATHOGENIC	0.000	0.951	-0.8830	0.4597	31	Y		0.2073
180	D	E	D180E	benign		Y	3.98E-06		probably damaging	0.971	PATHOGENIC	0.000	0.864	0.4977	0.5395	31	Y		0.2073
180	D	V	D180V	benign		Y	3.19E-05		probably damaging	0.969	PATHOGENIC	0.000	0.949	0.0389	0.6952	31	Y		0.2073
181	H	R	H181R	pathogenic	Y				probably damaging	0.939	PATHOGENIC	0.000	0.930	0.1134	1.7384	31	Y		0.0459
182	A	T	A182T	benign		Y	3.98E-06		possibly damaging	0.615	PATHOGENIC	0.030	0.905	0.8528	0.6471	31			0.0111
182	A	V	A182V	benign		Y	3.98E-06		possibly damaging	0.615	BENIGN	0.280	0.888	0.2929	0.7907	31			0.0111
183	W	R	W183R	pathogenic	Y				probably damaging	0.975	PATHOGENIC	0.000	0.929	5.0966	1.2239	31			0.0000
183	W	S	W183S	pathogenic	Y				possibly damaging	0.615	PATHOGENIC	0.000	0.903	4.8391	0.7219	31			0.0000
184	V	E	V184E	pathogenic	Y				possibly damaging	0.682	PATHOGENIC	0.000	0.933	6.7501	1.1294	31			0.0000
185	V	A	V185A	benign		Y	3.98E-06		benign	0.132	PATHOGENIC	0.020	0.775	1.7730	0.2750	31			0.0337
185	V	M	V185M	benign		Y	3.19E-05		benign	0.384	BENIGN	0.080	0.689	0.6731	0.6493	31			0.0337
188	P	L	P188L	uncertain	Y	Y	6.37E-05		benign	0.406	PATHOGENIC	0.040	0.678	1.2977	0.3977	31			0.4645
188	P	S	P188S	benign		Y	3.19E-05		benign	0.107	BENIGN	0.100	0.448	2.0940	0.3936	31			0.4645
189	N	K	N189K	benign		Y	3.98E-06		benign	0.016	BENIGN	0.600	0.405	0.0201	0.2781	31			0.7692
189	N	S	N189S	benign		Y	2.39E-05		benign	0.135	BENIGN	0.430	0.419	0.4844	0.2157	31			0.7692
192	Q	K	Q192K	pathogenic	Y				benign	0.232	PATHOGENIC	0.030	0.549	-0.0430	0.3342	31			0.3543
193	T	I	T193I	pathogenic	Y				possibly damaging	0.899	PATHOGENIC	0.000	0.974	0.9707	0.5893	31			0.0274
195	E	G	E195G	pathogenic	Y				probably damaging	0.982	PATHOGENIC	0.000	0.955	3.4234	0.4687	31			0.0243
210	T	I	T210I	benign		Y	3.98E-06		benign	0.266	PATHOGENIC	0.040	0.752	0.6793	0.3545	31			0.2650
212	N	D	N212D	benign		Y	3.98E-06		benign	0.009	BENIGN	0.830	0.436	-0.8192	0.5023	31			0.5201
214	G	S	G214S	benign		Y	2.39E-05		possibly damaging	0.518	BENIGN	0.100	0.879	4.8407	1.4332	31			0.0302
215	V	M	V215M	pathogenic	Y				possibly damaging	0.896	PATHOGENIC	0.020	0.799	-0.7025	0.1471	31			0.2135
218	R	Q	R218Q	benign		Y	3.99E-06		benign	0.020	BENIGN	0.400	0.557	0.7401	0.2913	31			0.3942
218	R	W	R218W	pathogenic	Y				possibly damaging	0.755	PATHOGENIC	0.000	0.753	1.4165	0.4539	31			0.3942
220	W	L	W220L	pathogenic	Y				probably damaging	0.997	PATHOGENIC	0.000	0.922	2.9227	0.5116	31			0.0015
220	W	R	W220R	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.921	4.0051	0.9144	31			0.0015
220	W	S	W220S	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.901	6.7655	0.6669	31			0.0015
223	L	P	L223P	pathogenic	Y				probably damaging	0.935	PATHOGENIC	0.010	0.941	7.5767	0.8001	31			0.0014
224	K	T	K224T	pathogenic	Y				benign	0.284	BENIGN	0.380	0.758	3.3049	0.9831	31			0.1634
225	G	R	G225R	pathogenic	Y				probably damaging	0.995	PATHOGENIC	0.010	0.942	2.7029	0.3718	31			0.1717
226	S	P	S226P	pathogenic	Y				possibly damaging	0.628	BENIGN	0.270	0.777	3.4129	1.2351	31			0.0618
228	M	I	M228I	benign		Y	3.98E-06		benign	0.018	BENIGN	0.110	0.577	0.3038	0.3738	31			0.0108
229	R	C	R229C	benign		Y	7.96E-06		benign	0.013	PATHOGENIC	0.030	0.624	0.9193	0.2721	31			0.4166
229	R	H	R229H	uncertain	Y	Y	3.98E-06		benign	0.041	BENIGN	0.140	0.622	1.7757	0.3599	31			0.4166
229	R	L	R229L	pathogenic	Y				benign	0.054	BENIGN	0.240	0.598	-0.0602	0.3253	31			0.4166
232	R	C	R232C	benign		Y	3.98E-06		probably damaging	0.999	PATHOGENIC	0.000	0.932	3.8895	0.7070	31			0.0381
234	M	L	M234L	benign		Y	3.98E-06		benign	0.130	BENIGN	0.180	0.780	-0.0072	0.2227	31			0.0402

235	E	K	E235K	pathogenic	Y				probably damaging	0.993	PATHOGENIC	0.000	0.948	8.2199	1.4759	31			0.0000
237	A	V	A237V	benign		Y	7.96E-06		possibly damaging	0.864	PATHOGENIC	0.010	0.931	-0.3751	0.8466	31			0.0000
239	M	T	M239T	benign		Y	3.98E-06		possibly damaging	0.526	PATHOGENIC	0.010	0.928	3.9556	0.7284	31			0.0459
240	V	E	V240E	pathogenic	Y				probably damaging	0.989	PATHOGENIC	0.000	0.956	6.8257	0.9850	31			0.0008
240	V	M	V240M	benign		Y	3.98E-06		probably damaging	0.992	PATHOGENIC	0.000	0.912	2.6809	0.7526	31			0.0008
241	C	F	C241F	pathogenic	Y				possibly damaging	0.628	PATHOGENIC	0.030	0.920	1.7271	3.2359	31	Y		0.1129
241	C	R	C241R	pathogenic	Y				benign	0.429	PATHOGENIC	0.050	0.881	1.2109	3.7403	31	Y		0.1129
241	C	Y	C241Y	pathogenic	Y				possibly damaging	0.628	PATHOGENIC	0.030	0.918	2.1882	4.7994	31	Y		0.1129
242	A	V	A242V	pathogenic	Y				probably damaging	0.962	PATHOGENIC	0.010	0.960	1.8258	0.3744	31	Y		0.1872
243	I	N	I243N	pathogenic	Y				probably damaging	0.968	PATHOGENIC	0.000	0.926	2.6044	0.2112	31			0.0442
243	I	V	I243V	benign		Y	3.98E-06		possibly damaging	0.672	BENIGN	0.060	0.811	0.9559	0.1608	31			0.0442
245	P	A	P245A	benign		Y	3.98E-06		possibly damaging	0.526	PATHOGENIC	0.000	0.909	2.3674	0.3147	31			0.0845
246	S	A	S246A	benign		Y	3.98E-06		benign	0.260	PATHOGENIC	0.020	0.820	-0.0399	0.1525	31	Y		0.2111
247	I	T	I247T	benign		Y	3.98E-06		probably damaging	0.924	PATHOGENIC	0.010	0.876	2.2288	0.2262	31			0.2531
248	D	Y	D248Y	benign		Y	3.98E-06		possibly damaging	0.904	PATHOGENIC	0.050	0.922	1.6610	0.6179	31	Y		0.4108
252	D	H	D252H	benign		Y	7.96E-06		probably damaging	0.997	PATHOGENIC	0.020	0.942	-0.8362	0.2275	31	Y		0.4108
252	D	Y	D252Y	benign		Y	1.06E-05		probably damaging	0.997	PATHOGENIC	0.020	0.947	-2.3017	0.3169	31	Y		0.4108
253	S	L	S253L	pathogenic	Y				possibly damaging	0.819	BENIGN	0.070	0.913	4.3682	2.7866	31	Y		0.0221
253	S	P	S253P	pathogenic	Y				probably damaging	0.968	BENIGN	0.160	0.872	5.8208	0.7425	31	Y		0.0221
253	S	W	S253W	pathogenic	Y				probably damaging	0.994	PATHOGENIC	0.010	0.951	17.3666	12.2514	31	Y		0.0221
255	E	K	E255K	pathogenic	Y				possibly damaging	0.791	PATHOGENIC	0.000	0.937	2.3874	0.7359	31			0.1608
256	L	F	L256F	pathogenic	Y				probably damaging	0.931	PATHOGENIC	0.040	0.898	7.1340	1.6983	31	Y		0.0050
258	Q	H	Q258H	benign		Y	2.12E-05		possibly damaging	0.796	BENIGN	0.120	0.815	0.2473	0.3907	31			0.3117
259	L	P	L259P	pathogenic	Y				probably damaging	0.992	PATHOGENIC	0.000	0.962	9.4625	0.7395	31			0.0085
259	L	R	L259R	pathogenic	Y				probably damaging	0.992	PATHOGENIC	0.000	0.975	5.9008	1.5954	31			0.0085
260	Q	P	Q260P	pathogenic	Y				probably damaging	0.983	PATHOGENIC	0.010	0.962	5.4437	0.4938	31			0.0413
264	L	P	L264P	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.931	7.6574	0.7237	31			0.0000
267	L	P	L267P	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.936	7.2362	0.6006	31			0.0057
268	Y	C	Y268C	benign		Y	3.55E-05		probably damaging	0.987	PATHOGENIC	0.000	0.884	2.6638	0.1960	31			0.2200
270	L	Q	L270Q	benign				Y	possibly damaging	0.721	BENIGN	0.560	0.574	1.4489	0.2796	31			0.4350
274	E	A	E274A	pathogenic	Y				benign	0.018	BENIGN	0.730	0.687	0.7569	0.1576	31			0.4471
275	R	K	R275K	pathogenic	Y				benign	0.028	BENIGN	1.000	0.802	-0.2607	0.1339	31			0.4546
277	P	H	P277H	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.979	14.7731	4.8586	31			0.0099
277	P	L	P277L	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.980	7.2659	1.6875	31			0.0099
278	M	I	M278I	pathogenic	Y				benign	0.406	PATHOGENIC	0.030	0.940	1.2111	0.7490	31	Y		0.1193
278	M	V	M278V	benign		Y	3.98E-06		possibly damaging	0.492	PATHOGENIC	0.020	0.929	2.1412	0.4918	31	Y		0.1193
280	L	S	L280S	benign		Y	1.19E-05		probably damaging	0.992	PATHOGENIC	0.000	0.944	2.9643	0.2609	31			0.0235
281	G	R	G281R	pathogenic	Y				probably damaging	0.983	PATHOGENIC	0.000	0.967	21.9039	3.6992	31			0.0220
283	L	P	L283P	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.970	8.5846	0.6836	31			0.0000
284	A	E	A284E	pathogenic	Y				probably damaging	0.953	PATHOGENIC	0.010	0.967	18.8858	2.3100	31			0.0000
286	L	P	L286P	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.966	5.0184	0.6794	31			0.0633
290	E	A	E290A	benign		Y	3.98E-06		benign	0.018	BENIGN	0.730	0.517	-0.5217	0.2693	31			0.5298
291	P	R	P291R	benign		Y	3.98E-06		benign	0.396	PATHOGENIC	0.010	0.612	-0.7489	0.8665	31			0.5562
292	T	I	T292I	benign		Y	3.98E-06		probably damaging	0.974	BENIGN	0.110	0.651	-0.4668	0.3403	31			0.2458
295	R	Q	R295Q	benign		Y	1.59E-05		benign	0.405	PATHOGENIC	0.020	0.588	1.8114	0.2081	31			0.3008
295	R	W	R295W	benign		Y	3.19E-05		probably damaging	0.986	PATHOGENIC	0.000	0.778	1.6398	0.9589	31			0.3008
301	L	R	L301R	pathogenic	Y				probably damaging	0.994	PATHOGENIC	0.000	0.930	7.7554	1.7455	31			0.0014
303	H	R	H303R	benign		Y	3.99E-06		benign	0.018	BENIGN	0.490	0.581	0.0421	0.2505	31			0.4247
305	G	D	G305D	pathogenic	Y				possibly damaging	0.780	PATHOGENIC	0.010	0.853	7.6302	1.0743	31			0.0206
305	G	R	G305R	pathogenic	Y				possibly damaging	0.897	PATHOGENIC	0.010	0.934	11.8527	1.6470	31			0.0206
307	A	D	A307D	benign		Y	3.98E-06		benign	0.035	BENIGN	0.590	0.471	0.3974	0.0932	31			0.3643
309	A	G	A309G	benign		Y	3.98E-06		benign	0.056	BENIGN	0.180	0.753	1.8990	0.0900	31			0.0000
309	A	P	A309P	pathogenic	Y				probably damaging	0.965	PATHOGENIC	0.020	0.924	7.2060	0.9740	31			0.0000
309	A	T	A309T	benign		Y	3.98E-06		possibly damaging	0.740	PATHOGENIC	0.010	0.878	1.2968	0.4150	31			0.0000
311	T	P	T311P	pathogenic	Y				possibly damaging	0.777	BENIGN	0.280	0.720	3.6314	0.9974	31			0.5656
314	R	P	R314P	pathogenic	Y				benign	0.196	BENIGN	0.260	0.682	5.1090	0.8558	31			0.6731
314	R	Q	R314Q	benign		Y	3.18E-05		benign	0.003	BENIGN	0.580	0.370	1.1370	0.6143	31			0.6731
314	R	W	R314W	benign		Y	1.19E-05		possibly damaging	0.491	PATHOGENIC	0.020	0.608	1.4379	0.5502	31			0.6731

315	D	Y	D315Y	uncertain	Y	Y	3.98E-06		probably damaging	0.974	PATHOGENIC	0.000	0.875	4.6447	0.6823	31			0.2457
316	E	D	E316D	benign		Y	7.95E-06		benign	0.295	BENIGN	0.500	0.591	0.7773	0.5619	31			0.1224
316	E	K	E316K	benign		Y	3.98E-06		benign	0.135	BENIGN	0.670	0.690	-0.6663	0.8476	31			0.1224
317	H	R	H317R	pathogenic	Y				probably damaging	0.958	PATHOGENIC	0.020	0.947	0.9688	0.8820	31			0.0064
317	H	Y	H317Y	pathogenic	Y				possibly damaging	0.872	PATHOGENIC	0.010	0.956	-1.2610	1.1099	31			0.0064
320	P	L	P320L	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.959	4.5393	0.9875	31			0.0261
320	P	R	P320R	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.958	11.4020	3.2030	31			0.0261
325	A	P	A325P	pathogenic	Y				possibly damaging	0.897	PATHOGENIC	0.020	0.910	7.9524	0.5275	31			0.0000
330	R	H	R330H	benign		Y	3.19E-05		probably damaging	0.999	PATHOGENIC	0.000	0.923	0.4408	0.2214	31	Y		0.4254
335	R	Q	R335Q	benign		Y	3.98E-06	Y	possibly damaging	0.642	BENIGN	0.240	0.531	0.3039	0.3378	31			0.3519
337	A	D	A337D	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.948	8.8807	1.3733	31			0.0000
337	A	P	A337P	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.952	8.5759	0.8298	31			0.0000
338	L	P	L338P	pathogenic	Y				probably damaging	0.929	PATHOGENIC	0.010	0.937	6.6133	0.5421	31			0.0227
338	L	V	L338V	benign		Y	1.99E-05		benign	0.384	BENIGN	0.090	0.795	2.9960	0.3553	31			0.0227
341	W	R	W341R	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.935	4.1895	0.6008	31	Y		0.0326
342	A	P	A342P	pathogenic	Y				probably damaging	0.988	PATHOGENIC	0.000	0.955	6.3422	0.5649	31			0.0011
342	A	T	A342T	benign		Y	3.98E-06		probably damaging	0.943	PATHOGENIC	0.000	0.891	4.7506	1.3042	31			0.0011
342	A	V	A342V	benign		Y	7.95E-06		possibly damaging	0.605	PATHOGENIC	0.000	0.865	5.5973	1.4952	31			0.0011
344	T	M	T344M	benign		Y	3.98E-06		possibly damaging	0.902	PATHOGENIC	0.010	0.733	0.8694	1.6832	31			0.0000
344	T	R	T344R	pathogenic	Y				possibly damaging	0.822	PATHOGENIC	0.010	0.812	5.1084	3.2709	31			0.0000
348	I	N	I348N	pathogenic	Y				probably damaging	0.952	PATHOGENIC	0.000	0.867	4.2577	0.4717	31			0.0000
350	D	V	D350V	pathogenic	Y				benign	0.129	BENIGN	0.260	0.688	0.1842	0.4611	31			0.3745
351	Y	H	Y351H	pathogenic	Y				probably damaging	0.994	PATHOGENIC	0.000	0.962	3.0878	0.2902	31			0.0043
351	Y	N	Y351N	pathogenic	Y				probably damaging	0.994	PATHOGENIC	0.000	0.944	3.9461	0.4693	31			0.0043
353	Y	D	Y353D	pathogenic	Y				probably damaging	0.988	PATHOGENIC	0.000	0.942	5.4037	0.5484	31			0.0744
354	C	R	C354R	benign		Y	3.98E-06		benign	0.331	BENIGN	0.150	0.661	-0.4974	0.5042	31			0.0359
355	R	W	R355W	pathogenic	Y				probably damaging	0.989	PATHOGENIC	0.000	0.937	0.0721	0.3291	31			0.6726
356	E	D	E356D	benign		Y	3.98E-06		benign	0.044	BENIGN	0.830	0.524	0.9321	0.8629	31			0.2172
357	D	H	D357H	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.908	16.7655	2.7987	31			0.0007
357	D	N	D357N	benign		Y	3.98E-06		probably damaging	0.997	PATHOGENIC	0.000	0.797	3.2138	0.8205	31			0.0007
364	F	C	F364C	pathogenic	Y				probably damaging	0.946	PATHOGENIC	0.000	0.966	4.2783	0.3465	31			0.0012
366	E	D	E366D	uncertain	Y	Y	7.16E-05		benign	0.044	BENIGN	0.800	0.611	0.6620	0.1874	31	Y		0.4356
367	V	I	V367I	benign		Y	1.27E-04		benign	0.057	BENIGN	1.000	0.422	-0.6475	0.2535	31			0.0411
367	V	L	V367L	benign		Y	7.96E-06		benign	0.280	PATHOGENIC	0.020	0.604	0.7840	0.9305	31			0.0411
368	A	D	A368D	pathogenic	Y				probably damaging	0.987	PATHOGENIC	0.000	0.961	3.2704	1.2425	31			0.0000
372	I	M	I372M	pathogenic	Y				possibly damaging	0.779	PATHOGENIC	0.010	0.837	-0.0463	0.2576	31			0.0094
373	P	A	P373A	pathogenic	Y				probably damaging	0.997	PATHOGENIC	0.000	0.959	3.0938	0.2602	31			0.0512
373	P	L	P373L	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.958	1.9255	0.4669	31			0.0512
373	P	S	P373S	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.972	3.9140	0.3660	31			0.0512
374	N	S	N374S	benign		Y	3.98E-06		benign	0.059	BENIGN	0.380	0.468	0.4952	0.2448	31			0.4637
374	N	T	N374T	benign		Y	2.12E-05		benign	0.059	BENIGN	0.500	0.459	0.8145	0.3527	31			0.4637
375	L	P	L375P	pathogenic	Y				possibly damaging	0.683	PATHOGENIC	0.020	0.910	5.5098	1.1145	31	Y		0.0455
375	L	V	L375V	benign		Y	3.98E-06		benign	0.016	BENIGN	1.000	0.528	1.9319	0.3863	31	Y		0.0455
376	L	P	L376P	pathogenic	Y				probably damaging	0.952	PATHOGENIC	0.010	0.940	6.7157	0.6123	31			0.0085
377	K	R	K377R	benign		Y	3.98E-06		benign	0.262	BENIGN	0.170	0.787	0.0647	0.2166	31			0.4728
380	A	T	A380T	benign		Y	7.95E-06		benign	0.067	BENIGN	0.430	0.437	0.9858	0.4990	31			0.1561
381	S	G	S381G	benign		Y	3.98E-06		possibly damaging	0.730	BENIGN	0.050	0.444	0.6972	0.2501	31			0.4488
385	A	V	A385V	uncertain	Y	Y	3.98E-06	Y	benign	0.024	BENIGN	0.210	0.608	0.8026	0.3563	31			0.7300
386	G	D	G386D	benign		Y	1.19E-05		benign	0.081	BENIGN	0.210	0.434	1.9634	0.2748	24			0.7524
399	S	N	S399N	benign		Y	3.99E-06		benign	0.000	BENIGN	0.230	0.358	-0.8044	0.1721	23			0.2742
401	G	D	G401D	benign		Y	3.99E-06		possibly damaging	0.653	BENIGN	0.620	0.588	1.4955	0.4410	24			0.5513
403	A	T	A403T	benign		Y	7.98E-06		possibly damaging	0.541	BENIGN	0.130	0.729	1.6106	0.8856	31			0.0866
411	A	P	A411P	pathogenic	Y				probably damaging	0.990	PATHOGENIC	0.010	0.922	6.2657	1.1989	31			0.0476
411	A	T	A411T	benign		Y	3.98E-06		probably damaging	0.973	PATHOGENIC	0.010	0.870	0.7892	0.4894	31			0.0476
413	L	P	L413P	pathogenic	Y				probably damaging	0.965	PATHOGENIC	0.000	0.944	7.0953	0.7535	31			0.0036
413	L	R	L413R	pathogenic	Y				probably damaging	0.965	PATHOGENIC	0.000	0.950	2.7864	0.8182	31			0.0036
414	L	P	L414P	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.918	7.7383	0.6482	31			0.0021
414	L	Q	L414Q	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.934	4.0294	0.3152	31			0.0021

415	R	P	R415P	pathogenic	Y				probably damaging	0.942	PATHOGENIC	0.010	0.912	6.3502	1.1223	31			0.2138
415	R	Q	R415Q	benign		Y	3.98E-06		benign	0.139	BENIGN	0.110	0.757	1.1921	0.3848	31			0.2138
418	D	N	D418N	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.891	2.5381	0.8784	31			0.0215
420	I	N	I420N	pathogenic	Y				probably damaging	0.986	PATHOGENIC	0.000	0.921	3.3948	0.3424	31			0.0015
421	C	Y	C421Y	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.952	7.0866	2.9568	31			0.0077
423	W	R	W423R	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.945	4.1542	0.5664	31			0.0000
423	W	S	W423S	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.925	6.5756	0.5713	31			0.0000
427	S	I	S427I	pathogenic	Y				probably damaging	0.972	PATHOGENIC	0.000	0.960	6.7257	1.2151	31			0.1392
427	S	R	S427R	pathogenic	Y				probably damaging	0.981	PATHOGENIC	0.010	0.867	9.6584	2.7474	31			0.1392
429	T	M	T429M	benign		Y	3.98E-06		probably damaging	0.999	PATHOGENIC	0.000	0.951	-0.7217	0.4454	31			0.3164
434	V	M	V434M	benign		Y	3.98E-06		probably damaging	0.929	PATHOGENIC	0.010	0.830	-0.9110	0.3113	31			0.5427
436	W	C	W436C	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.932	4.9191	0.4623	31			0.0000
436	W	R	W436R	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.933	5.3112	0.5725	31			0.0000
443	S	P	S443P	benign		Y	3.98E-06		probably damaging	0.930	PATHOGENIC	0.010	0.907	3.8273	0.8430	31			0.0037
443	S	Y	S443Y	pathogenic	Y				probably damaging	0.952	PATHOGENIC	0.000	0.938	12.3672	2.3746	31			0.0037
444	L	P	L444P	pathogenic	Y				probably damaging	0.929	PATHOGENIC	0.000	0.967	7.1602	0.7193	31			0.0000
447	F	L	F447L	pathogenic	Y				probably damaging	0.998	PATHOGENIC	0.000	0.945	2.3714	0.5948	31			0.0125
447	F	S	F447S	pathogenic	Y				probably damaging	0.999	PATHOGENIC	0.000	0.927	4.7555	0.3058	31			0.0125
448	E	A	E448A	benign		Y	3.99E-06		benign	0.016	PATHOGENIC	0.010	0.601	1.8534	0.7884	31			0.5426
451	V	E	V451E	benign		Y	4.37E-06		possibly damaging	0.865	PATHOGENIC	0.000	0.940	2.0544	0.4323	31			0.0443
452	R	Q	R452Q	benign		Y	8.74E-06		probably damaging	0.999	PATHOGENIC	0.000	0.932	4.3475	1.1805	31			0.0266
452	R	W	R452W	pathogenic	Y				probably damaging	1.000	PATHOGENIC	0.000	0.961	24.8968	2.8975	31			0.0266
455	V	G	V455G	benign		Y	4.38E-06		probably damaging	0.929	PATHOGENIC	0.000	0.944	4.0447	0.3325	31			0.0443
456	R	C	R456C	benign		Y	1.16E-05		benign	0.340	BENIGN	0.060	0.477	0.4542	0.1573	31			0.5339
456	R	H	R456H	benign		Y	4.38E-06		benign	0.001	BENIGN	0.130	0.358	0.6709	0.1510	31			0.5339
457	I	T	I457T	benign		Y	4.38E-06		possibly damaging	0.844	PATHOGENIC	0.000	0.883	2.7840	0.3016	31			0.1385
458	V	A	V458A	benign		Y	4.39E-06		benign	0.058	BENIGN	0.760	0.487	0.5864	0.5185	31			0.2578
520	G	R	G520R	benign		Y	3.19E-05		benign	0.001	BENIGN	0.470	0.400	0.8000	0.7326	5			0.6154
523	A	P	A523P	benign		Y	1.22E-05		benign	0.000	BENIGN	0.280	0.282	-0.3561	0.1075	5			0.7416
523	A	T	A523T	benign		Y	2.16E-05		benign	0.002	BENIGN	0.620	0.258	-0.0516	0.3239	5			0.7416
524	G	S	G524S	benign		Y	4.05E-06		benign	0.001	BENIGN	0.540	0.254	3.4229	0.5713	5			0.5897
525	T	I	T525I	benign		Y	4.04E-06		benign	0.006	BENIGN	0.240	0.296	0.1707	0.3734	5			0.7519
526	A	T	A526T	benign		Y	4.03E-06		benign	0.001	BENIGN	0.320	0.244	0.1557	1.0816	5			0.6408
527	R	Q	R527Q	benign		Y	4.02E-06		benign	0.001	BENIGN	0.480	0.349	0.5328	0.8541	5			0.9051
534	T	M	T534M	benign		Y	4.00E-06		benign	0.058	PATHOGENIC	0.050	0.354	-0.4702	0.0000	1			0.6977
536	Q	R	Q536R	benign		Y	3.99E-06		benign	0.012	BENIGN	0.520	0.398	-0.1550	0.0000	1			0.5289
537	V	G	V537G	benign		Y	3.99E-06		benign	0.000	PATHOGENIC	0.020	0.344	0.8241	0.0000	1			0.6724
537	V	L	V537L	benign		Y	3.99E-06		benign	0.000	BENIGN	0.490	0.346	-0.3012	0.0000	1			0.6724
544	P	S	P544S	uncertain	Y	Y	3.98E-06		benign	0.395	BENIGN	0.310	0.696	0.2736	0.0000	1			0.9434
546	P	Q	P546Q	benign		Y	3.98E-06		benign	0.010	BENIGN	0.360	0.360	0.8759	0.0000	1			0.3396
547	E	G	E547G	benign		Y	1.06E-05	Y	benign	0.001	BENIGN	0.290	0.470	0.1271	0.0042	2			0.8072
550	V	L	V550L	uncertain	Y	Y	1.19E-05		benign	0.071	BENIGN	0.350	0.708	-0.5712	0.2827	31			0.5246
552	T	S	T552S	pathogenic	Y				benign	0.017	BENIGN	0.640	0.417	0.4612	0.2239	31			0.3530
555	S	N	S555N	pathogenic	Y				probably damaging	0.969	PATHOGENIC	0.000	0.747	6.7505	1.7890	31			0.0000
555	S	R	S555R	pathogenic	Y				probably damaging	0.986	PATHOGENIC	0.000	0.878	7.8921	2.4523	31			0.0000
557	K	E	K557E	pathogenic	Y				probably damaging	0.981	PATHOGENIC	0.000	0.879	3.5375	0.6159	31			0.1416
558	M	K	M558K	pathogenic	Y				probably damaging	0.923	PATHOGENIC	0.000	0.945	3.3583	0.3057	31			0.0019
561	M	K	M561K	pathogenic	Y				probably damaging	0.923	PATHOGENIC	0.000	0.916	3.3142	0.7366	31			0.0000
561	M	R	M561R	pathogenic	Y				probably damaging	0.944	PATHOGENIC	0.000	0.920	6.1973	1.3920	31			0.0000
568	T	I	T568I	benign		Y	7.96E-06		benign	0.217	BENIGN	0.100	0.524	-0.2177	0.1399	31			0.6254
573	S	N	S573N	benign		Y	3.98E-06		probably damaging	0.969	BENIGN	0.930	0.642	0.4434	0.2037	31			0.5041
577	L	P	L577P	pathogenic	Y				probably damaging	0.997	PATHOGENIC	0.000	0.954	3.2416	0.9816	31			0.5458
579	L	F	L579F	benign		Y	3.98E-06		probably damaging	0.994	PATHOGENIC	0.020	0.798	4.7813	2.7921	31			0.0021
579	L	P	L579P	pathogenic	Y				probably damaging	0.997	PATHOGENIC	0.000	0.979	8.2627	1.0957	31			0.0021
583	S	P	S583P	pathogenic	Y				probably damaging	0.986	PATHOGENIC	0.000	0.963	-0.4088	0.9754	24			0.8548
597	D	A	D597A	benign				Y	probably damaging	0.994	PATHOGENIC	0.030	0.827	1.2074	0.0000	1			0.6425

DISCUSSION

Previous work has shown that predicted thermodynamic destabilization of protein structure, as measured by $\Delta\Delta G$ values calculated by FoldX, can be used as a predictor of pathogenicity in *MSH2* and *PAH* variants (7, 8). Our data indicates that the same is true for variants in *MEN1*, and that a high predicted $\Delta\Delta G$ value is a strong positive predictor for pathogenicity. Using a threshold of only 3 kcal/mol, specificity for variant classification was 89.4%, rising to 95.0% for a more conservative threshold of 4 kcal/mol. By contrast, using a proposed threshold of 0.7 for the phenotypic meta-prediction tool REVEL yielded a specificity of only 53%. Since MEN1 has variable penetrance and often late onset, the identification of likely pathogenic variants has significant implications for patient surveillance and genetic testing of family members. With respect to the seven novel missense variants reported here, all had high average predicted $\Delta\Delta G$ values (range, 4.81-13.16 kcal/mol) and six were deeply buried within the protein, strongly supporting pathogenicity. All these cases were also predicted as deleterious or probably pathogenic by commonly-used tools for *in silico* pathogenicity prediction; however, the comparatively low specificity of all these tools for variants in *MEN1* highlights the value of thermodynamic analysis as a means of reducing false positive calls.

As might be expected, our analysis shows that variants which are buried within the Menin structure are those that are predicted to result in greatest structural destabilization. In fact, the majority of reported pathogenic variants in *MEN1* are buried, suggesting that any novel variant which is solvent inaccessible (RSA<0.2) and has a predicted $\Delta\Delta G$ >4 kcal/mol is also highly likely to be pathogenic. Nevertheless, a number of pathogenic variants lie on the surface of Menin, and many of these have relatively low $\Delta\Delta G$ values. A number of these variants lie at or close to positions of known interactions with binding partners such as JunD, KMT2A or PSIP, where they presumably have an adverse effect on binding of these factors, emphasizing the value of integrating all known structural annotation into a final classification of the likely effect of a variant. Our data also suggests the possible existence of an as-yet unidentified interaction of Menin, as evidenced by the cluster of pathogenic variants lying on the protein surface opposite the JunD

binding pocket. Notably, *MEN1* has recently been identified as one of the genes exhibiting significant spatial clustering of pathogenic variants (47); our analysis suggests that this clustering is likely to apply both to regions of structural importance, which are buried in the interior of the protein, and to surface regions which form essential interactions with binding partners.

In terms of broader applicability of this approach, our work builds upon the reported analysis of *MSH2* and *PAH* variants and applies it to the classification of novel clinical variants. Whether the same approach can be used for other proteins remains to be determined. One obvious limitation of structural analysis is, by definition, the need for a suitable structural model. However, even where no experimental structures are available for a protein of interest, it may still be possible to use comparative modelling to generate a reliable model of regions or domains which can be used for structural analysis. Another likely limitation is the architecture of the protein itself. Both Menin and MSH2 are relatively compact, globular proteins, with low surface area to volume ratio and a high proportion of amino acids in regions of secondary structure. As a result, the effect of missense variants on the proteins' internal geometry and thermodynamic stability is amenable to *in silico* predictions, particularly given the availability of suitable high-quality PDB structures. However, less well-structured proteins, or fibrillar proteins where a greater proportion of amino acids are exposed to solvent, are likely to be less amenable to such study as the confidence with which the structural and thermodynamic effects of missense variants can be predicted will be much lower. Such rules are likely to be revealed only by proteome-wide study which is beyond the scope of this manuscript.

In summary, we have shown that structural analysis of missense substitutions in *MEN1* can be used to identify variants likely to destabilize the protein and thus potentially as an aid in variant classification. Given that all analysis described herein used publicly-available data and does not require specialist bioinformatic skills or infrastructure, such analysis lies within the capability of any genetics laboratory or testing service. As

such, there is significant scope to make greater use of protein structural data in the routine interpretation of genetic variation.

ACKNOWLEDGMENTS

The authors wish to acknowledge support from the Wellcome Trust (grant no. 200990).

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

1. Tang H, Thomas PD. Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. *Genetics* 2016; 203:635-647. doi: 10.1534/genetics.116.190033.

2. 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(Web Server issue):W382-388. doi: 10.1093/nar/gki387.

3. Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 2011; 79:830-838. doi: 10.1002/prot.22921.

4. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YE, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popović Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 2011; 487:545-574. doi: 10.1016/B978-0-12-381270-4.00019-6.

5. Compiani M, Capriotti E. Computational and theoretical methods for protein folding. *Biochemistry* 2013; 52:8601-8624. doi: 10.1021/bi4001529.

6. Masso M, Vaisman II. AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. *Protein Eng Des Sel* 2010; 23:683-687. doi: 10.1093/protein/gzq042.

7. Nielsen SV, Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG, Kassem MM, Rasmussen LJ, Lindorff-Larsen K, Hartmann-Petersen R. Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLoS Genet* 2017; 13(4):e1006739. doi: 10.1371/journal.pgen.1006739.

8. Scheller R, Stein A, Nielsen SV, Marin FI, Gerdes AM, Di Marco M, Papaleo E, Lindorff-Larsen K, Hartmann-Petersen R. Toward mechanistic models for genotype-phenotype correlations in

phenylketonuria using protein stability calculations. *Hum Mutat* 2019; 40:444-457. doi: 10.1002/humu.23707.

9. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (MEN1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat* 2008; 29:22-32. doi: 10.1002/humu.20605.

10. Thakker RV, Newey PJ, Walls GV, Bilezikian J, Dralle H, Ebeling PR, Melmed S, Sakurai A, Tonelli F, Brandi ML; Endocrine Society. Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab* 2012; 97:2990-3011. doi: 10.1210/jc.2012-1230.

11. Ellard S, Hattersley AT, Brewer CM, Vaidya B. Detection of an *MEN1* gene mutation depends on clinical features and supports current referral criteria for diagnostic molecular genetic testing. *Clin Endocrinol (Oxf)* 2005; 62:169-175. doi: 10.1111/j.1365-2265.2005.02190.x.

12. Alberto F. Genetics of parathyroids disorders: Overview. *Best Pract Res Clin Endocrinol Metab* 2018; 32:781-790. doi: 10.1016/j.beem.2018.09.011.

13. Beckers A, Aaltonen LA, Daly AF, Karhu A. Familial isolated pituitary adenomas (FIPA) and the pituitary adenoma predisposition due to mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene. *Endocr Rev* 2013; 34:239-277. doi: 10.1210/er.2012-1013.

14. Kövesdi A, Tóth M, Butz H, Szücs N, Sármán B, Pusztai P, Tőke J, Reismann P, Fáklya M, Tóth G, Somogyi A, Borka K, Erdei A, Nagy EV, Deák V, Valkusz Z, Igaz P, Patócs A, Grolmusz VK.. True MEN1 or phenocopy? Evidence for geno-phenotypic correlations in MEN1 syndrome. *Endocrine* 2019; doi:10.1007/s12020-019-01932-x. [Epub ahead of print].

15. Schaaf L, Pickel J, Zinner K, Hering U, Höfler M, Goretzki PE, Spelsberg F, Raue F, von zur Mühlen A, Gerl H, Hensen J, Bartsch DK, Rothmund M, Schneyer U, Dralle H, Engelbach M, Karges W, Stalla GK, Höppner W. Developing effective screening strategies in multiple endocrine neoplasia type 1 (MEN 1) on the basis

of clinical and sequencing data of German patients with MEN 1. *Exp Clin Endocrinol Diabetes* 2007; 115:509-517. doi: 10.1055/s-2007-970160.

16. Thakker RV. Multiple endocrine neoplasia type 1 (MEN1) and type 4 (MEN4). *Mol Cell Endocrinol* 2014; 386:2-15. doi: 10.1016/j.mce.2013.08.002.

17. Marini F, Giusti F, Brandi ML. Multiple endocrine neoplasia type 1: extensive analysis of a large database of Florentine patients. *Orphanet J Rare Dis* 2018; 13:205. doi: 10.1186/s13023-018-0938-8.

18. Concolino P, Costella A, Capoluongo E. Multiple endocrine neoplasia type 1 (MEN1): An update of 208 new germline variants reported in the last nine years. *Cancer Genet* 2016; 209:36-41. doi: 10.1016/j.cancergen.2015.12.002.

19. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17:405-424. doi: 10.1038/gim.2015.30.

20. Romanet P, Odou MF, North MO, Saveanu A, Coppin L, Pasmant E, Mohamed A, Goudet P, Borson-Chazot F, Calender A, Béroud C, Lévy N, Giraud S, Barlier A. Proposition of adjustments to the ACMG-AMP framework for the interpretation of *MEN1* missense variants. *Hum Mutat* 2019; 40:661-674. doi: 10.1002/humu.23746.

21. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 2017; 136:665-677. doi: 10.1007/s00439-017-1779-6. https://portal.biobase-international.com/cgi-bin/portal/. Accessed October 25 2018.

22. gnomAD (The Genome Aggregation Database). Cambridge, MA: Broad Institute, 2018. https://gnomad.broadinstitute.org. Accessed October 25 2018.

23. Sydney Genomics Collaborative Database. Sydney, Australia: Garvan Institute of Medical Research, 2018. https://sgc.garvan.org.au. Accessed October 25 2018.

24. wwPDB consortium (2018) Protein Data Bank: the single global archive for 3D macromolecular structure data Nucleic Acids Research, gky949. http://www.wwpdb.org. Accessed October 25 2018.

25. Huang J, Gurung B, Wan B, Matkar S, Veniaminova NA, Wan K, Merchant JL, Hua X, Lei M. The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* 2012; 482:542-546. doi: 10.1038/nature10806.

26. Shi A, Murai MJ, He S, Lund G, Hartley T, Purohit T, Reddy G, Chruszcz M, Grembecka J, Cierpicki T. Structural insights into inhibition of the bivalent menin-MLL interaction by small molecules in leukemia. *Blood* 2012; 120:4461-4469. doi: 10.1182/blood-2012-05-429274.

27. Zhou H, Liu L, Huang J, Bernard D, Karatas H, Navarro A, Lei M, Wang S. Structure-based design of high-affinity macrocyclic peptidomimetics to block the menin-mixed lineage leukemia 1 (MLL1) protein-protein interaction. *J Med Chem* 2013; 56:1113-1123. doi: 10.1021/jm3015298.

28. He S, Senter TJ, Pollock J, Han C, Upadhyay SK, Purohit T, Gogliotti RD, Lindsley CW, Cierpicki T, Stauffer SR, Grembecka J. High-affinity small-molecule inhibitors of the menin-mixed lineage leukemia (MLL) interaction closely mimic a natural protein-protein interaction. *J Med Chem* 2014; 57:1543-1556. doi: 10.1021/jm401868d.

29. Borkin D, He S, Miao H, Kempinska K, Pollock J, Chase J, Purohit T, Malik B, Zhao T, Wang J, Wen B, Zong H, Jones M, Danet-Desnoyers G, Guzman ML, Talpaz M, Bixby DL, Sun D, Hess JL, Muntean AG, Maillard I, Cierpicki T, Grembecka J. Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell* 2015; 27:589-602. doi: 10.1016/j.ccell.2015.02.016.

30. Borkin D, Pollock J, Kempinska K, Purohit T, Li X, Wen B, Zhao T, Miao H, Shukla S, He M, Sun D, Cierpicki T, Grembecka J. Property Focused Structure-Based Optimization of Small Molecule Inhibitors of the Protein-Protein Interaction between Menin and Mixed Lineage Leukemia (MLL). *J Med Chem* 2016; 59:892-913. doi: 10.1021/acs.jmedchem.5b01305.

31. Pollock J, Borkin D, Lund G, Purohit T, Dyguda-Kazimierowicz E, Grembecka J, Cierpickiet T. Rational Design of Orthogonal Multipolar Interactions with Fluorine in Protein-Ligand Complexes. *J Med Chem* 2015; 58:7465-7474. doi: 10.1021/acs.jmedchem.5b00975.

32. Xu S, Aguilar A, Xu T, Zheng K, Huang L, Stuckey J, Chinnaswamy K, Bernard D, Fernández-Salas E, Liu L, Wang M, McEachern D, Przybranowski S, Foster C, Wang S. Design of the First-in-Class, Highly Potent Irreversible Inhibitor Targeting the Menin-MLL Protein-Protein Interaction. *Angew Chem Int Ed Engl*. 2018; 57:1601-1605. doi: 10.1002/anie.201711828.

33. FoldX Suite. Version 4. Available at http://foldxsuite.crg.eu/. The FoldX Consortium, Centre for Genomic Regulation, 2018.

34. The PyMOL Molecular Graphics System, Version 2.2. Schrödinger, LLC, 2018.

35. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983; 22:2577-2637. doi: 10.1002/bip.360221211.

36. Touw WG, Baakman C, Black J, te Beek TA, Krieger E, Joosten RP, Vriend G. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res* 2015; 43(Database issue):D364-368. doi: 10.1093/nar/gku1028.

37. DSSP-3.0.0. Nijmegen, Netherlands: Centre for Molecular and Biomolecular Informatics, Radboud University Medical Centre, 2018. https://swift.cmbi.umcn.nl/gv/dssp/index.html

38. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, Wilke CO. Maximum allowed solvent accessibilities of residues in proteins. *PLoS One* 2013; 8:e80635. doi: 10.1371/journal.pone.0080635.

39. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* 2012; 13:111. http://genesilico.pl/metadisorder/. Accessed December 31 2018.

40. Shimazu S, Nagamura Y, Yaguchi H, Ohkura N, Tsukada T. Correlation of mutant menin stability with clinical expression of multiple endocrine neoplasia type 1 and its incomplete forms. *Cancer Sci* 2011; 102:2097-2102. doi: 10.1111/j.1349-7006.2011.02055.x.

41. Yaguchi H, Ohkura N, Takahashi M, Nagamura Y, Kitabayashi I, Tsukada T. Menin missense mutants encoded by the *MEN1* gene that are targeted to the proteasome: restoration of expression and activity by CHIP siRNA. *J Clin Endocrinol Metab* 2012; 97:E282-291. doi: 10.1128/MCB.24.15.6569-6580.2004.

42. SIFT (Sorts Intolerant from Tolerant). La Jolla, CA: J. Craig Venter Institute. http://provean.jcvi.org/index.php. Accessed March 03 2019.

43. 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. doi: 10.1038/nmeth0410-248. http://genetics.bwh.harvard.edu/pph2/. Accessed March 03 2019.

44. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016;

99:877-885. doi: 10.1016/j.ajhg.2016.08.016. https://sites.google.com/site/revelgenomics/downloads. Accessed March 03 2019.

45. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 2009; 19:596-604. doi: 10.1016/j.sbi.2009.08.003.

46. Oza AM, DiStefano MT, Hemphill SE, Cushman BJ, Grant AR, Siegert RK, Shen J, Chapin A, Boczek NJ, Schimmenti LA, Murry JB, Hasadsri L, Nara K, Kenna M, Booth KT, Azaiez H, Griffith A, Avraham KB, Kremer H, Rehm HL, Amr SS, Abou Tayoun AN; ClinGen Hearing Loss Clinical Domain Working Group.. Expert specification of the ACMG/AMP variant interpretation guidelines for genetic hearing loss. *Hum Mutat* 2018; 39:1593-1613. doi: 10.1002/humu.23630.

47. Sivley RM, Dou X, Meiler J, Bush WS, Capra JA. Comprehensive Analysis of Constraint on the Spatial Distribution of Missense Variants in Human Protein Structures. *Am J Hum Genet* 2018; 102:415-426. doi: 10.1016/j.ajhg.2018.01.017.