Types and effects of protein variations.

Vihinen, Mauno

Published in:
Human Genetics

DOI:
10.1007/s00439-015-1529-6

2015

TYPES AND EFFECTS OF PROTEIN VARIATIONS

Mauno Vihinen

Department of Experimental Medical Science, Lund University, BMC D10, SE-22184 Lund, Sweden

Correspondence to Mauno Vihinen, Department of Experimental Medical Science, Lund University, BMC D10, SE-22184 Lund, Sweden, tel. +46 72 526 0022, fax +46 46 222 0296, e-mail mauno.vihinen@med.lu.se

ABSTRACT

Variations in proteins have very large number of diverse effects affecting sequence, structure, stability, interactions, activity, abundance and other properties. Although protein-coding exons cover just over 1 % of the human genome they harbor an disproportionately large portion of disease-causing variants. Variation ontology (VariO) has been developed for annotation and description of variation effects, mechanisms and consequences. A holistic view for variations in proteins is made available along with examples of real cases. Protein variants can be of genetic origin or emerge at protein level. Systematic names are provided

for all variation types, a more detailed description can be made by explaining changes to protein function, structure and properties. Examples are provided for the effects and mechanisms, usually in relation to human diseases. In addition, the examples are selected so that protein 3D structural changes, when relevant, are included and visualized. Here, systematics is described for protein variants based on VariO. It will benefit the unequivocal description of variations and their effects and further reuse and integration of data from different sources.

Keywords: Variation Ontology; VariO; variation annotation; mutation; protein variant; effects, consequences and mechanisms of variants

INTRODUCTION

Variations manifest their effects at different ways. Of the identified disease-causing cases large proportion appears in protein coding DNA and RNA sequences, although the protein-coding regions constitute only about 1.3 % of human genome. Due to the large number of different functions in which proteins are involved, also the effects of variants are widely different. In addition to functions, protein variants can change structures, properties, interactions and other characteristics in several ways. List of the effects would be really long and include e.g. interactions, stability, electrostatic effects, protein packing, local and global structural changes and so on.

Many deleterious variants are straightforward to explain, such as large deletions, protein truncations, amphigoric amino acid insertions, deletions and indels, and many other types. Usually the most difficult ones are minor alterations, most often amino acid substitutions. The effects on proteins are more difficult to study experimentally than those on DNA or RNA. Hybridization and sequence complementarity-based methods work (almost) equally well for all nucleotide sequences, whereas a dedicated analysis protocol is needed for every individual protein and often even for individual variants [see e.g. (Perniola and Musco 2014; Storz and Zera 2011; Yates and Sternberg 2013)].

Experimental methods are the first choice to study the detailed effects of variants; however, these kinds of data are often missing. In such cases, bioinformatics predictors may be useful (Kucukkal et al. 2014; Thusberg and Vihinen 2009; Zhang et al. 2012). Several prediction methods have been developed based on e.g. evolutionary information, protein physical and chemical features, and Gene Ontology annotations to detect function affecting or harmful variants (Calabrese et al. 2009; Kircher et al. 2014; Li et al. 2009; Niroula et al. 2015; Olatubosun et al. 2012; Schwarz et al. 2014) or those having more specific effects such as in aggregation (Conchillo-Sole et al. 2007; Maurer-Stroh et al. 2010; Trovato et al. 2007) or stability (Capriotti et al. 2005; Cheng et al. 2006; Yin et al. 2007) or affecting specific proteins such as mismatch repair system (Ali et al. 2012), protein kinases (Izarzugaza et al. 2013) or hemophilia-related proteins (Hamasaki-Katagiri et al. 2013). The majority of these kinds of methods are dedicated for amino acid substitutions. Assessments of performance of these tools indicated that tolerance (Thusberg et al. 2011), protein stability (Khan and Vihinen 2010), protein disorder (Ali et al. 2014) and localization (Laurila and Vihinen 2011) predictors, and likely also other kinds of methods, which have not yet been assessed, have widely varying performance.

Three-dimensional structures of proteins are very helpful for interpretation of the effects and mechanisms of variations. If experimental structures obtained with X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy are not available, it may be possible to model the protein structure. Retrospective analysis of computer models indicated that biological and medical inferences can be quite reliable even for difficult modeling cases with low sequence identities (Khan and Vihinen 2010).

Statistical studies of variation types and protein structures have provided useful insight. Disease-causing variants have a clearly distinct amino acid distribution compared to benign variants (de Beer et al. 2013; Schaafsma and Vihinen 2014; Steward et al. 2003; Vitkup et al. 2003), also in protein secondary structural elements (Khan and Vihinen 2007). Protein stability is frequently altered due to variations (Ferrer-Costa et al. 2002; Wang and Moult 2001; Yue et al. 2005).

The majority of the harmful variants are loss-of-function variants; however, gain-of-function variants are identified at increasing rate. Usually the latter ones have constitutive activity and have harmful effects due to not being regulated.

All the experimental and computational approaches mentioned above can be combined to explain and annotate variations and their consequences. In this article, effects of protein variations are discussed mainly in relation to diseases and in the framework of the Variation Ontology (VariO) (Vihinen 2014a), which allows systematic description of variation effects, consequences and mechanisms whether of experimental or of predicted origin. Examples are included to highlight the different features of variants. Some previous studies have reviewed protein variation effects, mainly in relation to protein three-dimensional (3D) structures (Ferrer-Costa et al. 2002; Furnham et al. 2012; Stefl et al. 2013; Steward et al. 2003; Wang and Moult 2001; Yue et al. 2005), but without systematic coverage and systematics.

DATABASES FOR PROTEIN VARIATIONS

Several freely available and commercial databases contain information about proteins and their variants. Swiss-Var (http://www.swissvar.expasy.org) at UniProtKB/Swiss-Prot (http://www.uniprot.org/) contains notions of disease-causing variants. Variation information is available in locus-specific databases (LSDBs), which are collected for individual genes/diseases or groups of them, or from central databases such as the Human Gene Mutation Database (HGMD, http://www.hgmd.org/ or the commercial version with up-to-date information at http://www.biobaseinternational.com/), dbSNP (http://www.ncbi.nlm.nih.gov/

projects/SNP/) and ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/). LSDBs can be searched either from the Human Genome Variation Society web site (http://www.hgvs.org/dblist/glsdb.html), the LOVD listing (http://grenada.lumc.nl/LSDB_list/), or from the GEN2PHEN project (http://www.gen2phen.org/data/lsdbs). In addition, WAVE lists databases and also (partial) lists of variants in them (http://www.bioinformatics.ua.pt/WAVe/). Protein 3D structures are collected to the Protein Data Bank (PDB, http://www.rcsb.org/pdb/). An example of an effect-specific database is ProTherm (http://www.abren.net/protherm/) for protein thermodynamic changes upon variation often for other organisms than human and in many cases including engineered variations.

VARIATION ONTOLOGY

To allow efficient use, reuse, search and integration of information it is essential to describe it in a systematic way. This applies also to variation information. Recently, VariO (http://www.variationontology.org/) was introduced for the systematic description of variation effects, consequences and mechanisms (Vihinen 2014a). The ontology is used to annotate information in databases at the three molecular levels: DNA, RNA and protein. Each of these levels contains further terms for variation type, function, structure and various properties. The discussion of variation effects will be made based on VariO. VariO annotations are always made in relation to a reference state, e.g. reference sequence or wild-type enzyme activity.

VariO annotations consist of two parts including the VariO prefix and a number such as VariO:0012 and name, in this case “protein variation type”. The number with the prefix is mandatory for annotations. This article is organized according to the VariO and is thus divided into the four major sublevel variation types, protein function, structure and properties. VariO terms are used in the subtitles and otherwise written in quotation marks. For the sake of

clarity, in this article, the prefix and the number have been omitted. Detailed guidelines are available for VariO annotations (Vihinen 2014b). Annotations for databases can be made with VariOtator annotation tool. For the examples in the text, full VariO annotations together with Evidence Code (ECO, http://www.evidenceontology.org/) annotations are available at http://www.variationontology.org/annotation/protein_examples. The number of databases utilizing VariO annotations is increasing, including Uni-ProtKB/Swiss-Prot (Famiglietti et al. 2014).

PROTEIN VARIATION TYPE

Variation type in VariO provides a description for a variation in English (see Fig. 1). Human Genome Variation Society (HGVS) nomenclature provides systematic description of variations (den Dunnen and Antonarakis 2001); however, these names may be difficult to interpret. For example p.(K28_K29delinsW) means predicted deletion of two lysines (K28 and K29) and their replacement by a single tryptophan without experimental evidence at protein level. Variation type terms of VariO provide a brief description with commonly used terms and are not intended to replace HGVS names, instead to provide easily understandable description for human readers and computer applications.

There are two types of "protein variation origin", namely "protein variation of genetic origin" and "variation emerging at protein level". Variants of genetic origin have appeared on DNA (or RNA) level and therefore directly affect the protein. An example is R525Q substitution in Bruton tyrosine kinase (BTK) causing X-linked agammaglobulinemia (XLA) (Vihinen et al. 1994), which is caused by g to a transition in the coding gene. The variations emerging at protein level are one of four types. They are either "artificial protein variation", "epigenetic protein variation", "mistranslated protein" or "post-translational modification". Phosphorylation of BTK tyrosine kinase domain at Y551 is a "post-translational modification" (Mahajan et al. 1995). Post-translational modifications are frequent; however, only a few of the 600 known modifications in RESID database (http://www.pir.georgetown.edu/resid) are frequent and not all of them appear in human. "Mistranslated proteins" due to editing-defective tRNA synthetase cause neurodegeneration because of misfolding protein (Lee et al. 2006). The epigenetic variations include "protein structural inheritance" and "proteinaceous infection" of prions. Structural inheritance with unknown mechanism has been noted in centrosomes, organelles which are the main microtubule-organizing center and regulator of cell cycle progression (Wilson 2008). Asymmetric cell division or centrosome inheritance may have effect on cancer (Izumi and Kaneko 2012). Prions are discussed in detail later in the text.

Protein engineering has been widely used in an attempt to modify properties of proteins, especially those having commercial utility e.g. to alter their stability or to increase activity or specificity. A prime example of an engineered protein is subtilisin, a protease that digests proteins, properties of which have been extensively modified with random and targeted mutagenesis [see (Bryan 2000)]. Engineering of the NEMO protein at N-terminus affects thermal stability and ligand binding (Guo et al. 2014) is an example of "artificial protein variation".

"Protein variation classification" in VariO divides the variants into "amino acid insertion", "amino acid deletion", "amino acid indel", "amino acid substitution" or "missing protein". All the following variants in BTK lead to XLA: R525Q "amino acid substitution" (Vihinen et al. 1994), deletion of 28 amino acids from SH3 domain and a linker region cause "amino acid deletion" (Zhu et al. 1994), and "amino acid insertion" at S604 (Holinski-Feder et al. 1998). Insertions and indels can be of two types, either the sequence is retained after the

insertion/indel site or it has changed, in which case they are called as amphigoric variations. “Protein truncation” is a special case of deletion where deletion occurs either in N- or C-terminus. Indels originate due to both insertion and deletion.

The different types of protein variations in a short protein sequence for alanine, alanine, serine… in one letter code AASEQWENCE are depicted in Fig. 2. There are altogether 8 categories of variation types. “Sequence retaining amino acid deletion” in Bruton tyrosine kinase (BTK) deletes, due to skipping of exon 8, altogether 21 amino acids from the C-terminus of Src homology 3 (SH3) domain and linker connecting to SH2 domain (Zhu et al. 1994). The protein is produced but non-functional and is a cause for XLA, a rare primary immunodeficiency, due to block in B cell maturation. The truncated domain alone does not fold correctly (Chen et al. 1996). “Amphigoric amino acid insertion” because of a single-base insertion to the *BTK* gene region encoding PH domain changes the sequence after the insertion site at E59 and creates new premature stop codon truncating the modified C-terminus of the protein (Holinski-Feder et al. 1998). The truncated protein has no activity and cannot be detected.

ClC5 variations lead to Dent’s disease, X-linked disorder with low molecular weight proteinuria, hypercalciuria, nephrocalcinosis, nephrolithiasis, and often also renal failure. One of the variants, R718X, deletes 28 C-terminal amino acids (D’Antonio et al. 2013). Although the truncated protein is produced, it has enhanced protease susceptibility suggesting an at least partly misfolded and destabilized protein. This variant is a “protein truncation”.

Eight amino acid “sequence-retaining amino acid indel” in SERPINC1 is a complex variant that consists of a two amino acid deletion, a seven residue inserted repeat and a three amino acid insertion (Martínez-Martínez et al. 2012). It causes type II antithrombin deficiency due to missing serine protease inhibitory activity. Patients suffer from impaired blood clotting when having vascular damage. The indel appears in the highly conserved β-strand forming part of a β-sheet (Martínez-Martínez et al. 2012). The hyperstable form of the variant has a new β-strand that displaces the reaction center loop (Fig. 3a) thereby inactivating the protein. A “sequence-retaining amino acid insertion” of two residues in the BTK kinase domain at position 604 (Holinski-Feder et al. 1998) affects α-helix G and causes XLA (Mao et al. 2001).

“Amino acid substitutions” describe changes of individual amino acids. These are typically caused by nucleotide substitutions at DNA, and called for missense variants at RNA level. “Missing protein” describes a situation where no protein is produced. This can be due to e.g. nonsense mediated decay (NMD) destroying RNA molecules containing premature stop codons, or variants to start codons preventing translation. The E31K substitution and R41X truncation in Wiskott–Aldrich syndrome (WAS) protein (WASP) are examples of “missing protein” (Jin et al. 2004). WAS is a rare recessive X-linked primary immunodeficiency. WAS patients suffer from thrombocytopenia, eczema and diarrhea in addition to immunodeficiency.

VARIATION AFFECTING PROTEIN FUNCTION

Protein functions are presented in Fig. 4. Changes to the function are divided into seven categories. “Effect on protein catalytic function” describes changes to enzyme activity. Nomenclature Committee of the International Union of Biochemistry and Molecular Biology EC numbers (http://www.chem.qmul.ac.uk/iubmb/enzyme/) can be used for more detailed description of affected enzyme activity. The R525Q substitution at the catalytic site inactivates BTK kinase catalytic function (Vihinen et al. 1994). “Effect on protein information transfer” can be used to describe e.g. hormone effects. Insulin is a peptide

hormone in which variants lead to for example neonatal diabetes due to impaired regulation of carbohydrate and fat metabolism (Støy et al. 2007).

"Effect on protein movement" is used to describe changes to movement of contractile proteins such as actin and myosin. Several myopathies with muscle weakness and muscle structural changes emerge due to some 200 known variants in α-actin (Nowak et al. 2013). These diseases are related to muscle weakness and skeletal muscle structural lesions. "Effect on protein recognition" term describes selective and non-covalent interactions of proteins e.g. for immunoglobulins and other immunological recognition molecules. Insertion to immunoglobulin gene for common γ2 chain (Cγ2) causes IgG2 deficiency (Tashita et al. 1998). The patients are susceptible for sinopulmonary infections.

"Effect on protein storage" describes changes to proteins acting as amino acid or metal cation stores, such as ferritin and casein. Ferritin light chain variants lead to autosomal dominant neuroferritinopathy due to iron deposition to brains (Lehn et al. 2012). The accumulation of iron to brain causes adulthood disease with slow progression. "Effect on transport function of protein" can be used to annotate changes that affect active transport functions of proteins such as membrane transporters. G87R substitution in zinc transporter ZnT-2 (SLC30A2) impairs zinc transport because the protein is mislocalized and has decreased stability (Lasry et al. 2012). The variant has ER retention. The disease affects many cellular processes as Zn^{2+} is required for activity of several enzymes. Infants are especially vulnerable for Zn deficiency, as the prolonged metal deprivation causes growth retardation and neurophysiological changes. "Effect on structural protein" defines alterations in proteins providing support for cells and tissues including keratin, collagen and others. Variations to type II collagen affect the structural protein of connective tissue and result in conditions with varying severity ranging from lethal to relatively mild (Kannu et al. 2012).

The terms for protein function are for fundamental protein activities in cells. One protein may have several functions. The term describing the function(s) affected by the variation should be used for annotation.

VARIATION AFFECTING PROTEIN STRUCTURE

Folding is a prerequisite for activity of most proteins. Structural changes due to variations have a wide range from minor local alterations to drastic global changes. Structure terms are first divided into two categories, those having "effect on protein 3D structure" and those of "epigenetic protein modification" (Fig. 5). The epigenetic modifications include "epigenetic protein complex structure" and "prion formation".

The effect on protein 3D structure is further divided into terms "complex 3D structural change", "effect on protein dynamics", "effect on protein quaternary structure" and "effect on protein tertiary structure". Terms for protein dynamics cover those affecting allosteric site, induced fit, structural disorder, and structural flexibility. The largest number of terms relates to tertiary structure including e.g. amino acid size, interaction site, secondary structural element and protein folding. Structure and property terms in VariO can be modified by attributes to provide more detailed descriptions. ECO terms are used to indicate the evidence for the annotations. VariO website includes examples for each of the terms along with ECO annotations.

Effect on protein dynamics

Proteins are dynamic molecules and in constant motion but the basic scaffolding usually remains the same during the fluctuations. Variations can affect normal dynamics in several

ways. M257Y is a constitutively active variant of G-protein-coupled receptor (GPCR) that affects the dynamics of structural alteration between active and inactive conformation (Tsukamoto and Farrens 2013). The transition is essential for the protein function. Protein with the gain-of function variant is in the active conformation even in the absence of the activating external signal.

P275L substitution in phenylalanine hydroxylase (PAH) has an "effect on allosteric site" regulation by cofactor (Gersting et al. 2008) causing phenylketonuria (PKU), a metabolic disorder, which, if untreated, leads to intellectual disability, seizures and other problems. PAH is needed to degrade amino acid phenylalanine to tyrosine. The variant has decreased affinity for tetrahydrobiopterin, the cofactor, thereby disturbing allostery. The substrate affinity, however, is even increased. PKU is the most common inborn error in amino acid metabolism in European-descent populations.

Apurinic/apyrimidinic endonuclease 1 (APE1) is required for DNA damage correction by base excision and nucleotide incision repair mechanisms. The enzyme–substrate complex undergoes induced fit upon binding to lesion. Substitution K98A in APE1 has an "effect on induced fit" preventing it and decreasing the stability of enzyme–substrate complex (Timofeyeva et al. 2011).

Protein structures typically fold to ordered 3D structures. Exceptions are intrinsically disordered proteins and regions. Disordered regions can adopt diverse temporary conformations and have large numbers of binding partners (Oldfield et al. 2008). An example is tumor suppressor p53, which contains large disordered regions that interact with several different partners. Variations at p53 are hallmarks in numerous cancers. "Effect on structural disorder" status is related to certain diseases. Misfolded proteins have four possible fates in cells. They are either degraded, refolded, they form aggregates or they are sequestered. Single amino acid replacement D54G in creatine kinase leads to muscle creatine kinase deficiency (Feng et al. 2007). The variant has partially misfolded structure, and although the protein is produced, the activity is impaired. Creatine kinase catalyzes the production of ATP by transferring phosphate group from phosphoserine. Muscles, especially heart, consume lots of high-energy ATP molecules.

Structural flexibility is essential for many protein functions (Vihinen 1987). Laforin regulates cellular autophagy. Harmful variations are responsible for progressive form of myoclonus epilepsy called Lafora disease. K87A substitution in laforin has reduced stability due to "effect on structural flexibility" (Srikumar and Rohini 2013). Laforin has two activities, protein dual specificity phosphatase and glycogen binding. K87A eliminates the glycogen-binding activity without effect on phosphatase activity. This is because the activities are in different domains.

Effect on quaternary structure

Protein quaternary structures are formed of interacting subunits. Variations with "effect on quaternary structure forming interaction" can be harmful by preventing or weakening complex formation. Congenital cataract-causing A2V substitution in the eye lens βB2-crystallin folds normally and forms βB2/βA3-crystallin dimer; however, it prevents tetramer formation, the protein complex in eyes (Xu et al. 2012). The tetramer structure is shown in Fig. 3b. The patients have from birth cloudy of the lens of the eye.

Effect on tertiary structure

This is the largest category for protein structural variations depicting the numerous ways protein 3D structures can be altered

Effect on amino acid size

Cores of globular proteins are typically tightly packed and therefore amino acid substitutions by larger or smaller side chains can be harmful. “Effect on protein packing” can be due to “protein overpacking” or “protein cavity formation”. An example of the latter is I2A substitution of invariant residue in chain A of insulin, which generates a cavity into the hydrophobic protein core and leads to reduced thermal stability (Xu et al. 2002). The structure of insulin is shown in Fig. 3c both for the normal and the variant form. Overpacking appears when the variant is larger than the original residue which leads to sterical clashes and at least local structural change. Figure 3d shows a situation where the larger side chain of G584 to W variant (Futatani et al. 2001) does not fit into the core of the BTK kinase domain structure without structural alterations due to clashes with a large number of amino acids.

Effect on protein folding rate

Protein folding is a complex process, the rate of which can be altered by variations. Even minor amino acid substitutions including Y25F and V68A have major effects to the folding rate of muscle acylphosphatase (AcP) (Chiti et al. 1999). AcP is a widely studied model because it folds slowly and is thus easy to investigate.

Effect on protein interaction site

Proteins function via interactions with other molecules. Variations affecting interactions can be described very detailed using VariO attributes.

Type I interferons (IFNs) are cytokines having immunomodulatory, antiviral and antiproliferative activities. Substitution R120A and variations at surrounding amino acids in the α-helix D in human IFNα2 have “effect on protein-binding site” by decreasing interaction with receptor subdomain IFNAR-1 while the affinity for IFNAR-2 subunit remains unchanged (Pan et al. 2008). The sites in IFN-α2 structure are shown in Fig. 3e. An example of variation having an “effect on protein catalytic site” is BTK in which variants R525Q and D430E at catalytic amino acids cause XLA (Fig. 3f) (Vihinen et al. 1994).

Effect on post‐translational modification

Numerous proteins are modified post-translationally after synthesis. These modifications often have important regulatory functions. “Deletion of post-translational modification site” can occur due to a variation. BTK kinase domain is activated by Y551 phosphorylation at activation loop. Variation Y551H inactivates the protein by preventing activation although the protein is expressed (Aghamohammadi et al. 2006).

T168N substitution leads to “generation of novel posttranslational modification site”, a new N-glycosylation site, to interferon receptor IFNγR2 causing severe mycobacterial disease (Vogt et al. 2005). Because the variant protein contains extra carbohydrate moiety it has increased molecular weight.

Proteolytic processing is a common mechanism in protein activation and thus variations can have an “effect on protein processing”. “Effect on protein splicing” does not occur in human, but in some other organisms. Indel I143NT in the Ca^{2+}-binding loop of guanylate cyclase-activating protein 1 (GCAP1) confers the protein susceptible for proteolysis (Nishiguchi et al. 2004) and is thus annotated as “variation generating a novel protein-processing site”. The variant leads to autosomal-dominant cone degenerate in retinal diseases. Variation R628P in integrin α3β1, which acts as receptor for laminins, is a “variation preventing polypeptide processing” of seven-bladed β-propeller structure of the protein (Yamada and Sekiguchi 2013). Upon maturation, the propeller structure is cleaved to produce heavy and light chains,

which remain connected by a disulfide bond. The variant protein transport to cell surface is deficit from endoplasmic reticulum (ER) to Golgi apparatus. The variant is associated with disorders in the lung, kidney, and skin due to the protein's wide expression during the development in epithelial organs.

Effect on protein secondary structural element

Protein 3D folds contain secondary structural elements including helices, strands, coils and various turn motifs. Disease-related variations emerge often in secondary structural elements and have typical characteristics (Khan and Vihinen 2010). Secondary structural elements are usually less tolerant for substitutions than coils or irregular loops. Examples of secondary structural element-affecting variants are H362Q and L358F in BTK Src homology 2 (SH2) domain, which impair the structure of the domain based on functional and structural studies (Mattsson et al. 2000). The structure of SH2 domain is shown in Fig. 3g. These variants have description "effect on protein helix" and more specifically "effect on right-handed protein helix" of type "effect on alpha helix". Examples are not provided for all the possible secondary structural elements as there are so many of them (see Fig. 5) and they all follow the same principles of annotation.

Variations with "effect on transmembrane polypeptide" have either "effect on membrane protein extramembrane region" or "effect on membrane protein intramembrane region". Proteolipid protein (PLP1) variants are causative of hereditary spastic paraplegia (HSP) type 2. The disease is characterized by progressive leukodystrophy and dysmyelination leading to axonal degeneration. PLP1 has four transmembrane helices connected by extramembrane loops. A30P alteration in the transmembrane helix 1 leads to the retaining of the protein in ER instead of transport to cellular membrane (Noetzli et al. 2014). The second extracellular domain of PLP1 contains two disulfide bridges. C184R substitution breaks the bridge and leads to Pelizaeus–Merzbacher disease, another dysmyelination disorder (Fukumura et al. 2011). This is an "effect on membrane protein extramembrane region".

Effect on protein fold

Protein folds are in constant motion and dynamics is essential also for functions. The folds tolerate changes at some positions but are sensitive at other parts. The type of the variation is important for the obtained effect.

"Changed domain orientation" in adenylate kinase 4 (AK4) caused by L171P variation leads to a movement of LID domain (Liu et al. 2009). AK4 has two conformations, open and closed, dependent on ligand binding (Fig. 3h). The L171P variant reorients the LID domain and has effect on ligand binding. AK4 catalyzes the reversible ATP (GTP) production from ADP (GDP).

Variation C260Y in hereditary hemochromatosis (HFE) protein is one of the most common variants causing hereditary hemochromatosis where patients have excessive intestinal absorption of dietary iron. C260Y alteration prevents disulfide bridge formation with structural consequences and thus has "effect on protein disulfide formation" (Bennett et al. 2000) (Fig. 3i).

"Protein conformational change" can be either minor (local) or global. An example of "local conformation change" is caused by ΔF508 deletion in the cystic fibrosis transmembrane conductance regulator (CFTR). The deletion causes a localized conformational change at residues 509–511 (Lewis et al. 2010) (Fig. 3j). Since the region is thought to be in contact with transmembrane domain of the protein it is harmful and causes cystic fibrosis, an

autosomal recessive disease, because of abnormal transport of chloride and sodium across an epithelium affecting lungs and digestive system.

Protein elongation

Elongation of the protein chain can appear at both ends. Variation at the terminal codon for Bβ chain for fibrinogen in a patient with fibrinogen Kyoto VI causes dysfibrinogenemia (defective fibrinogen clot formation). The variation leads to addition of 12 extra C-terminal amino acids (Okumura et al. 2012) and is "C-terminal protein elongation".

c.-14C>T transition leads to generation of upstream start codon in the *IFITM5* gene for interferon-induced transmembrane protein 5 (Semler et al. 2012). The new start site is in frame and it leads to "N-terminal protein elongation" of the produced protein by five additional residues. The elongation is harmful and causes osteogenesis imperfecta with a bone fragility and susceptibility to fractures even after minor trauma.

"Protein fusion" is quite common variant in cancers but appears also in some other diseases. Philadelphia chromosome originates from reciprocal translocation of chromosomes 9 and 22 fusing genes *Abl1* for Abelson murine leukemia viral oncogene homolog 1 and *BCR* coding for breakpoint cluster region protein (Advani and Pendergast 2002). The fusion is characteristic for chronic myelogenous leukemia (CML), 95 % of patients carrying the variant. The fusion protein varies between patients depending on the actual fusion location. The variant protein is constitutively active protein tyrosine kinase, which acts unregulated and contributes to the disease.

Epigenetic protein modification

Epigenetic changes are common at DNA and RNA levels and affect also proteins. "Epigenetic protein complex formation" may appear even in man. "Prion formation" occurs in transmissible spongiform encephalopathies in man in Creutzfeldt–Jakob disease (CJD), kuru, Gerstmann–Sträussler–Scheinker (GSS) syndrome and familial insomnia (FFI). There are other forms in cattle and in sheep and goats. Common to these diseases is progressive degeneration of the central nervous system resulting in dementia, motor dysfunction and death.

Amyloid fibrils accumulate in brains in the prion diseases caused by irreversible structural conformational change of prion protein. The formation of pathogenic conformation that is insoluble and resistant to proteolytic degradation is a signature for prion diseases. The pathogenic form stimulates the conversion of normal proteins, thus prions are called infectious proteins. The prion protein has normal functions and the protein is abundant in brain (Roucou et al. 2004).

Lots of studies have been devoted for the prion conformational conversion. More than 40 variants, including G131V, S132I and A133V, in human prion are associated with the disease (Chen et al. 2010). The details of the conversion are not yet fully revealed. The native prion protein structure is shown in Fig. 3k. The disease form cannot be studied with experimental structure determination methods.

VARIATION AFFECTING PROTEIN PROPERTY

In addition to variation type, functional and structural alterations, variations affect properties of proteins. The properties range from pathogenicity to protein physical characteristics. Many of the properties are further defined in VariO with attributes, especially with quality attributes.

Association of protein variation to pathogenicity

The association of a variation to pathogenicity varies depending on the type of variation, its location etc. An example of "disease-causing" variation is R525Q in BTK (Vihinen et al. 1994). Along with the other variants listed in BTKbase (Piirilä et al. 2006), it is associated with the disease, XLA. The other attributes of this property are "disease associated" and "not related to clinical phenotype".

Conservation of protein variation site

Sequence conservation at the variation site can be defined by conservation attributes such as "conserved region", "covariant position" and "invariant region". An example of the latter is W124C variation in BTK pleckstrin homology (PH) domain (Fiorini et al. 2004). PH domains share the same overall fold but have very low sequence identity of 17 % (Shen and Vihinen 2004). W124 is the only nearly invariant residue in the entire protein domain family. Alterations are deleterious in BTK in this αA helix position.

An example of "conserved region" is BTK SH2 domain R288. It is conserved in a large protein family. Alterations are disease-causing also in this hotspot site which accumulates altogether 8 R288Q cases, 29 patients in 21 families have R288W alteration, and in addition there are two patients with deletions starting at this site [see BTKbase and SH2base (Lappalainen et al. 2008; Piirilä et al. 2006)]. The invariant and conserved regions are depicted in the BTK SH2 domain in Fig. 3l.

Covariant amino acids are coevolving, meaning that viable variation in one site requires also the other site to change. Covarying amino acids are typically interacting with each other, but may have also other reason for conservation as seen in BTK PH domain-binding site (Shen and Vihinen 2004). Analysis of neutrophil elastase (ELANE) homologs revealed a network of covarying amino acids (see Fig. 3m) several of which are involved in congenital or cyclic neutropenia (lack of neutrophils) (Thusberg and Vihinen 2006).

Effect on protein abundance

Alterations to produced protein amounts are frequent in many diseases. Some variations in Wiskott–Aldrich syndrome protein (WASp) lead either to decreased protein abundance (e.g. L39P) in X-linked thrombocytopenia (XLT) or missing protein (e.g. R13X) in Wiskott–Aldrich Syndrome (WAS) (Jin et al. 2004). XLT is a milder disease than WAS.

Effect on protein accessibility

The localization of an amino acid on the surface or in protein core is one of the essential functional and physical properties of residues within proteins. Variants can affect protein accessibility, such as Y279C in SH2-domain-containing protein-tyrosine phosphatase (SHP2) that causes LEOPARD syndrome due to altered accessibility of the substrate recognition surface (Yu et al. 2013), see Fig. 3n. Although the phosphatase activity is reduced there is gain-of-function phenotype because the variants are hypersensitive for upstream activators and they have prolonged substrate turnover.

R234W is a "variation burying exposed region" in cellular retinaldehyde-binding protein (CRALBP) causing a domino-like conformational rearrangement of a number of side chains in the protein structure (He et al. 2009). The variant reduces the rate of pigment regeneration in eye and leads to Bothnia dystrophy, an autosomal recessive disease with high prevalence in northern Sweden. The structural changes are shown in Fig. 3o.

Δ508 deletion in CFTR is an example of a "variation exposing buried region", see Fig. 3j. Due to the deletion, the conformation of the loop is changed so that the side chain of V510 is exposed compared to the normal protein (Lewis et al. 2010) (Fig. 6a).

Effect on protein activity

Numerous variations and variation mechanisms affect protein activity. It is obvious that activity alterations are largely harmful. However, the degree of activity required normally is dependent on the protein. Adenosine deaminase deficiency is a defect in purine nucleotide metabolism and most harmful for T cells. The defect causes severe combined immunodeficiency (SCID), but only when the enzyme activity is below 0.1 % of that in normal enzyme (Hershfield 2003). Thus, it is important to study which level of change is harmful, not just to look at the extent of change. This is important to keep in mind also in regard to other protein properties.

"Effect on protein affinity" has an example of immunoglobulin E (IgE) binding to CD23 receptor on B cells. Ca^{2+} binding to CD23, lectin-like head domain, triggers a conformational change in a loop via P250 *trans* to *cis* conformational change (Fig. 6b) (Yuan et al. 2013). The conformational change increases IgE affinity by 30-fold to the lectin-like head domain of CD23 and variations in Ca-binding residues (E249A and D270A) reduce the binding. Note that although *K*m for reactions following Michaelis–Menten (M–M) kinetics is often thought to be associated with affinity of enzyme to substrate, it is not always the case. Therefore, changes of *K*m and related features should be annotated with "effect on enzyme reaction kinetics".

"Effect on enzyme reaction kinetics" altering in case of M–M kinetics k_{cat}, K_m, V_{max}, k_{cat}/K_m etc., is often harmful. Examples are N158D and E232K in multifunctional enzyme (MFE) type 2 involved in β-oxidation and causing D-bifunctional protein (D-BP) deficiency when altered. N158D variant decreases *V*max and *k*cat/K_m ratio and increases both K_m and k_{cat} values (Mehtälä et al. 2013). E232K alteration has quite normal K_m value; however, the other kinetic characteristics are changed. N158 is involved in substrate binding while K232 is participating in protein dimerization (Fig. 6c).

Variations in the DEF site interaction pocket in p38α MAP kinase decrease phosphorylation of some substrates but not others (Tzarum et al. 2013). The Y258A variant has such "effect on protein specificity" (see Fig. 6d). Variants at the DEF pocket significantly reduce the phosphorylation of ATF-II, Elk-1 and MBP without the effect on MK2 kinase phosphorylation.

Effect on protein charge

Charge and polarity of protein surface are important characteristics for molecular interactions. Change of M712T in UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetyl-mannosamine kinase leads to hereditary inclusion body myopathy because of "effect on protein isoelectric point" lowering the IEP of the protein (Weidemann et al. 2011). The variant generates a novel phosphorylation site which contributes to the surface charge.

"Effect on protein electrostatics" has been shown to be the major effect for several XLA-causing variants in the BTK PH domain (Okoh and Vihinen 1999). Figure 6e shows all the currently known PH domain variation positions (Piirilä et al. 2006) and effects to electrostatics. A large number of the variants change the electrostatics of the domain and thereby affects its interactions with cellular membrane and protein partners.

Effect on protein degradation

Degradation is a normal fate of proteins once they are misfolded. Variant E130D in formylglycine-generating enzyme (FGE) is rapidly degraded although it is correctly localized and relatively active, leading to metabolic disorder multiple sulfatase deficiency (MSD) (Schlotawa et al. 2013). Posttranslational activation of sulfatases is reduced because the variant protein has low stability and it is therefore degraded.

Effect on protein interaction

Variations frequently affect protein interactions, including those in $2Cl^-/H^+$ exchanger ClC-5 responsible for X-linked recessive Dent's disease. A large number of 39 studied variants modify the protein dimer interface, including G513R, I524K and G512R (Lourdel et al. 2012).

Extensive analysis of variation effects in human neutrophil elastase (HNE)-causing variants either in congenital or cyclic neutropenia include several which have an "effect on residue contact energy", the strongest ones including ELANE variants A57T, L121P, C151S and C151Y (Thusberg and Vihinen 2006) (Fig. 6f).

More detailed description of interaction effects can be made using the interaction attributes in VariO.

Effect on protein solubility

Variations reducing protein solubility have either "effect on protein aggregation" or "effect on protein inclusion body formation". Aggregation is a common mechanism in neurodegenerative diseases. L55P variation in transthyretin (TTR, thyrosine-binding prealbumin) causes early-onset familial amyloidotic polyneuropathy (McCutchen et al. 1993). Comparison of crystal structures for the normal and the variant-containing forms indicates disruption of hydrogen bonds between β-strands leading to altered surface that can explain the aggregation of the protein (Sebastião et al. 1998) (Fig. 6g).

Inclusion body formation is the mechanism of P56S variant in vesicle-associated membrane protein (VAPB) major sperm protein domain (MSP) leading to amyotrophic lateral sclerosis (ALS) (Shi et al. 2010). P56 has a *cis* conformation i.e. the side chain is on the same side of protein backbone as the preceding and following amino acids. This conformation is critical for the loop structure. The variant protein is unstructured and insoluble and forms inclusion bodies (Shi et al. 2010).

Effect on protein stability

Stability is a fundamental protein property. Variations can either increase of decrease stability. The most commonly studied form of stability is thermal stability; however, there are also other forms of stability e.g. toward denaturants, proteases or organic solvents. Study of 20 expressed variant proteins of mitochondrial DNA (mtDNA) helicase, implicated in a number of mitochondrial diseases, revealed a spectrum of thermal stability affecting variants (Longley et al. 2010). R354P increases the stability, while A359T and F485L have the most reduced stability.

Effect on protein subcellular localization

For the proper function of a protein it is essential that it is correctly localized within the cell (or outside of it). Proteins of cellular compartments have targeting signals which can be altered by variations or the localization can be otherwise impaired. Mislocalized proteins can be harmful due to being active in wrong compartment and due to lacking activity in the normal environment.

Amino acid substitutions, including R529Q, in the SH3 domain of tetratricopeptide repeats-containing protein 2 (SH3TC2) affect the localization to cell membrane and cause Charcot–Marie Tooth (CMT) disease type 4, a demyelinating neuropathy (Lupo et al. 2009).

Protein function change

Changes to protein function, both "protein gain of function" and "protein loss of function", are frequent consequences of variations. M257Y substitution is constitutively active "protein gain of function" form of rhodopsin, a GPCR family member. The variant alters the dynamics

of conformational change between the active and inactive forms (Tsukamoto and Farrens 2013). "Antimorphic protein variation" has antagonizing activity, i.e. opposite activity compared to normal state. The dominant negative effect appears for example in 637E and K673R variants in signal transducer and activator of transcription 1 (STAT1) SH2 domain resulting in impaired STAT1-mediated responses to IFN-γ and IL27 leading to susceptibility to mycobacterial disease (Tsumura et al. 2012).

"Neomorphic protein variation" has novel activities. D444V substitution of mitochondrial dilipoamide dehydrogenase (DLD) activates cryptic protease activity (Babady et al. 2007). The protein forms normally a homodimer and variants into the dimer interface enhance proteolytic activity and partial or complete inactivation of DLD activity. D444V variant causing severe metabolic disorder has reduced DLD activity and the novel protease activity. The variant misses hydrogen bonds between the monomers and exposes the catalytic site for the protease (Fig. 6h).

There are numerous examples above about "protein loss of function" variants such as BTK R525Q in BTK and E31K in WAS (Jin et al. 2004; Vihinen et al. 1994). Usually it is not necessary to annotate loss of function as it is evident already by functional annotations.

CONCLUSIONS

A comprehensive overview of variation effects on proteins ranging from variation types to functional and structural effects is presented. The effects, mechanisms and consequences are discussed based on VariO annotations and examples often include protein 3D structures for such cases. VariO provides a kind of periodic system for these descriptions. Systematic description of variants allows unequivocal description, easy reuse, integration, and further studies of observed variants and effects.

REFERENCES

Advani AS, Pendergast AM (2002) Bcr-Abl variants: biological and clinical aspects. Leuk Res 26: 713-20.

Aghamohammadi A, Fiorini M, Moin M, Parvaneh N, Teimourian S, Yeganeh M, Goffi F, Kanegane H, Amirzargar AA, Pourpak Z, Rezaei N, Salavati A, Pouladi N, Abdollahzade S, Notarangelo LD, Miyawaki T, Plebani A (2006) Clinical, immunological and molecular characteristics of 37 Iranian patients with X-linked agammaglobulinemia. Int Arch Allergy Immunol 141: 408-14. doi: 10.1159/000095469

Ali H, Olatubosun A, Vihinen M (2012) Classification of mismatch repair gene missense variants with PON-MMR. Hum Mutat 33: 642-50. doi: 10.1002/humu.22038

Ali H, Urolagin S, Gurarslan O, Vihinen M (2014) Performance of protein disorder prediction programs on amino acid substitutions. Hum Mutat 35: 794-804. doi: 10.1002/humu.22564

Babady NE, Pang YP, Elpeleg O, Isaya G (2007) Cryptic proteolytic activity of dihydrolipoamide dehydrogenase. Proc Natl Acad Sci U S A 104: 6158-63. doi: 10.1073/pnas.0610618104

Bennett MJ, Lebron JA, Bjorkman PJ (2000) Crystal structure of the hereditary haemochromatosis protein HFE complexed with transferrin receptor. Nature 403: 46-53. doi: 10.1038/47417

Bryan PN (2000) Protein engineering of subtilisin. Biochim Biophys Acta 1543: 203-222.

Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30: 1237-44. doi: 10.1002/humu.21047

Capriotti E, Fariselli P, Calabrese R, Casadio R (2005) Predicting protein stability changes from sequences using support vector machines. Bioinformatics 21 Suppl 2: ii54-8. doi: 10.1093/bioinformatics/bti1109

Chen W, van der Kamp MW, Daggett V (2010) Diverse effects on the native β-sheet of the human prion protein due to disease-associated mutations. Biochemistry 49: 9874-81. doi: 10.1021/bi101449f

Chen YJ, Lin SC, Tzeng SR, Patel HV, Lyu PC, Cheng JW (1996) Stability and folding of the SH3 domain of Bruton's tyrosine kinase. Proteins 26: 465-71. doi: 10.1002/(sici)1097-0134(199612)26:4<465::aid-prot7>3.0.co;2-a

Cheng J, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62: 1125-32. doi: 10.1002/prot.20810

Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol 6: 1005-9. doi: 10.1038/14890

Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 8: 65. doi: 10.1186/1471-2105-8-65

D'Antonio C, Molinski S, Ahmadi S, Huan LJ, Wellhauser L, Bear CE (2013) Conformational defects underlie proteasomal degradation of Dent's disease-causing mutants of ClC-5. Biochem J 452: 391-400. doi: 10.1042/bj20121848

Danielian S, El-Hakeh J, Basilico G, Oleastro M, Rosenzweig S, Feldman G, Berozdnik L, Galicchio M, Gallardo A, Giraudi V, Liberatore D, Rivas EM, Zelazko M (2003) Bruton tyrosine kinase gene mutations in Argentina. Hum Mutat 21: 451. doi: 10.1002/humu.9131

de Beer TA, Laskowski RA, Parks SL, Sipos B, Goldman N, Thornton JM (2013) Amino acid changes in disease-associated variants differ radically from variants observed in the 1000 genomes project dataset. PLoS Comput Biol 9: e1003382. doi: 10.1371/journal.pcbi.1003382

den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109: 121-4.

Famiglietti ML, Estreicher A, Gos A, Bolleman J, Gehant S, Breuza L, Bridge A, Poux S, Redaschi N, Bougueleret L, Xenarios I (2014) Genetic variations and diseases in UniProtKB/Swiss-Prot: the ins and outs of expert manual curation. Hum Mutat 35: 927-35. doi: 10.1002/humu.22594

Feng S, Zhao TJ, Zhou HM, Yan YB (2007) Effects of the single point genetic mutation D54G on muscle creatine kinase activity, structure and stability. Int J Biochem Cell Biol 39: 392-401. doi: 10.1016/j.biocel.2006.09.004

Ferrer-Costa C, Orozco M, de la Cruz X (2002) Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J Mol Biol 315: 771-86. doi: 10.1006/jmbi.2001.5255

Fiorini M, Franceschini R, Soresina A, Schumacher RF, Ugazio AG, Rossi P, Plebani A, Notarangelo LD (2004) BTK: 22 novel and 25 recurrent mutations in European patients with X-linked agammaglobulinemia. Hum Mutat 23: 286. doi: 10.1002/humu.9219

Fukumura S, Adachi N, Nagao M, Tsutsumi H (2011) A novel *proteolipid protein 1* gene mutation causing classical type Pelizaeus-Merzbacher disease. Brain Dev 33: 697-9. doi: 10.1016/j.braindev.2010.11.010

Furnham N, de Beer TA, Thornton JM (2012) Current challenges in genome annotation through structural biology and bioinformatics. Curr Opin Struct Biol 22: 594-601. doi: 10.1016/j.sbi.2012.07.005

Futatani T, Watanabe C, Baba Y, Tsukada S, Ochs HD (2001) Bruton's tyrosine kinase is present in normal platelets and its absence identifies patients with X-linked agammaglobulinaemia and carrier females. Br J Haematol 114: 141-9.

Gersting SW, Kemter KF, Staudigl M, Messing DD, Danecka MK, Lagler FB, Sommerhoff CP, Roscher AA, Muntau AC (2008) Loss of function in phenylketonuria is caused by impaired molecular motions and conformational instability. Am J Hum Genet 83: 5-17. doi: 10.1016/j.ajhg.2008.05.013

Guo B, Audu CO, Cochran JC, Mierke DF, Pellegrini M (2014) Protein engineering of the N-terminus of NEMO: Structure stabilization and rescue of IKKβ binding. Biochemistry 53: 6776-85. doi: 10.1021/bi500861x

Hamasaki-Katagiri N, Salari R, Wu A, Qi Y, Schiller T, Filiberto AC, Schisterman EF, Komar AA, Przytycka TM, Kimchi-Sarfaty C (2013) A gene-specific method for predicting hemophilia-causing point mutations. J Mol Biol 425: 4023-33. doi: 10.1016/j.jmb.2013.07.037

He X, Lobsiger J, Stocker A (2009) Bothnia dystrophy is caused by domino-like rearrangements in cellular retinaldehyde-binding protein mutant R234W. Proc Natl Acad Sci U S A 106: 18545-50. doi: 10.1073/pnas.0907454106

Hershfield MS (2003) Genotype is an important determinant of phenotype in adenosine deaminase deficiency. Curr Opin Immunol 15: 571-7.

Holinski-Feder E, Weiss M, Brandau O, Jedele KB, Nore B, Backesjo CM, Vihinen M, Hubbard SR, Belohradsky BH, Smith CI, Meindl A (1998) Mutation screening of the BTK gene in 56 families with X-linked agammaglobulinemia (XLA): 47 unique mutations without correlation to clinical course. Pediatrics 101: 276-84.

Izarzugaza JM, Vazquez M, del Pozo A, Valencia A (2013) wKinMut: an integrated tool for the analysis and interpretation of mutations in human protein kinases. BMC Bioinformatics 14: 345. doi: 10.1186/1471-2105-14-345

Izumi H, Kaneko Y (2012) Evidence of asymmetric cell division and centrosome inheritance in human neuroblastoma cells. Proc Natl Acad Sci U S A 109: 18048-53. doi: 10.1073/pnas.1205525109

Jin Y, Mazza C, Christie JR, Giliani S, Fiorini M, Mella P, Gandellini F, Stewart DM, Zhu Q, Nelson DL, Notarangelo LD, Ochs HD (2004) Mutations of the Wiskott-Aldrich Syndrome Protein (WASP): hotspots, effect on transcription, and translation and phenotype/genotype correlation. Blood 104: 4010-9. doi: 10.1182/blood-2003-05-1592

Kannu P, Bateman J, Savarirayan R (2012) Clinical phenotypes associated with type II collagen mutations. J Paediatr Child Health 48: E38-43. doi: 10.1111/j.1440-1754.2010.01979.x

Khan S, Vihinen M (2007) Spectrum of disease-causing mutations in protein secondary structures. BMC Struct Biol 7: 56. doi: 10.1186/1472-6807-7-56

Khan S, Vihinen M (2010) Performance of protein stability predictors. Hum Mutat 31: 675-84. doi: 10.1002/humu.21242

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM (2014) A general framework for estimating the relative pathogenicity of human genetic variants. 46: 310-5. doi: 10.1038/ng.2892

Kucukkal TG, Yang Y, Chapman SC, Cao W, Alexov E (2014) Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. Int J Mol Sci 15: 9670-717. doi: 10.3390/ijms15069670

Lappalainen I, Thusberg J, Shen B, Vihinen M (2008) Genome wide analysis of pathogenic SH2 domain mutations. Proteins 72: 779-92. doi: 10.1002/prot.21970

Lasry I, Seo YA, Ityel H, Shalva N, Pode-Shakked B, Glaser F, Berman B, Berezovsky I, Goncearenco A, Klar A, Levy J, Anikster Y, Kelleher SL, Assaraf YG (2012) A dominant negative heterozygous G87R mutation in the zinc transporter, ZnT-2 (SLC30A2), results in transient neonatal zinc deficiency. J Biol Chem 287: 29348-61. doi: 10.1074/jbc.M112.368159

Laurila K, Vihinen M (2011) PROlocalizer: integrated web service for protein subcellular localization prediction. Amino Acids 40: 975-80. doi: 10.1007/s00726-010-0724-y

Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, Davisson MT, Sundberg JP, Schimmel P, Ackerman SL (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature 443: 50-5. doi: 10.1038/nature05096

Lehn A, Boyle R, Brown H, Airey C, Mellick G (2012) Neuroferritinopathy. Parkinsonism Relat Disord 18: 909-15. doi: 10.1016/j.parkreldis.2012.06.021

Lewis HA, Wang C, Zhao X, Hamuro Y, Conners K, Kearins MC, Lu F, Sauder JM, Molnar KS, Coales SJ, Maloney PC, Guggino WB, Wetmore DR, Weber PC, Hunt JF (2010) Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. J Mol Biol 396: 406-30. doi: 10.1016/j.jmb.2009.11.051

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25: 2744-50. doi: 10.1093/bioinformatics/btp528

Liu R, Xu H, Wei Z, Wang Y, Lin Y, Gong W (2009) Crystal structure of human adenylate kinase 4 (L171P) suggests the role of hinge region in protein domain motion. Biochem Biophys Res Commun 379: 92-7. doi: 10.1016/j.bbrc.2008.12.012

Longley MJ, Humble MM, Sharief FS, Copeland WC (2010) Disease variants of the human mitochondrial DNA helicase encoded by *C10orf2* differentially alter protein stability, nucleotide hydrolysis, and helicase activity. J Biol Chem 285: 29690-702. doi: 10.1074/jbc.M110.151795

Lourdel S, Grand T, Burgos J, González W, Sepulveda FV, Teulon J (2012) ClC-5 mutations associated with Dent's disease: a major role of the dimer interface. Pflugers Arch 463: 247-56. doi: 10.1007/s00424-011-1052-0

Lupo V, Galindo MI, Martinez-Rubio D, Sevilla T, Vilchez JJ, Palau F, Espinos C (2009) Missense mutations in the SH3TC2 protein causing Charcot-Marie-Tooth disease type 4C affect its localization in the plasma membrane and endocytic pathway. Hum Mol Genet 18: 4603-14. doi: 10.1093/hmg/ddp427

Mahajan S, Fargnoli J, Burkhardt AL, Kut SA, Saouaf SJ, Bolen JB (1995) Src family protein tyrosine kinases induce autoactivation of Bruton's tyrosine kinase. Mol Cell Biol 15: 5304-11.

Mao C, Zhou M, Uckun FM (2001) Crystal structure of Bruton's tyrosine kinase domain suggests a novel pathway for activation and provides insights into the molecular basis of X-linked agammaglobulinemia. J Biol Chem 276: 41435-43. doi: 10.1074/jbc.M104828200

Martínez-Martínez I, Johnson DJ, Yamasaki M, Navarro-Fernández J, Ordóñez A, Vicente V, Huntington JA, Corral J (2012) Type II antithrombin deficiency caused by a large in-

frame insertion: structural, functional and pathological relevance. J Thromb Haemost 10: 1859-66. doi: 10.1111/j.1538-7836.2012.04839.x

Mattsson PT, Lappalainen I, Bäckesjö CM, Brockmann E, Lauren S, Vihinen M, Smith CIE (2000) Six X-linked agammaglobulinemia-causing missense mutations in the Src homology 2 domain of Bruton's tyrosine kinase: phosphotyrosine-binding and circular dichroism analysis. J Immunol 164: 4170-7.

Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JW, Rousseau F (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 7: 237-42. doi: 10.1038/nmeth.1432

McCutchen SL, Colon W, Kelly JW (1993) Transthyretin mutation Leu-55-Pro significantly alters tetramer stability and increases amyloidogenicity. Biochemistry 32: 12119-27.

Mehtälä ML, Lensink MF, Pietikäinen LP, Hiltunen JK, Glumoff T (2013) On the molecular basis of D-bifunctional protein deficiency type III. PLoS One 8: e53688. doi: 10.1371/journal.pone.0053688

Niroula A, Urolagin S, Vihinen M (2015) PON-P2: Prediction method for fast and reliable identification of harmful variants. PLoS ONE (in press).

Nishiguchi KM, Sokal I, Yang L, Roychowdhury N, Palczewski K, Berson EL, Dryja TP, Baehr W (2004) A novel mutation (I143NT) in guanylate cyclase-activating protein 1 (GCAP1) associated with autosomal dominant cone degeneration. Invest Ophthalmol Vis Sci 45: 3863-70. doi: 10.1167/iovs.04-0590

Noetzli L, Sanz PG, Brodsky GL, Hinckley JD, Giugni JC, Giannaula RJ, Gonzalez-Alegre P, Di Paola J (2014) A novel mutation in *PLP1* causes severe hereditary spastic paraplegia type 2. Gene 533: 447-50. doi: 10.1016/j.gene.2013.09.076

Nowak KJ, Ravenscroft G, Laing NG (2013) Skeletal muscle α-actin diseases (actinopathies): pathology and mechanisms. Acta Neuropathol 125: 19-32. doi: 10.1007/s00401-012-1019-z

Okoh MP, Vihinen M (1999) Pleckstrin homology domains of tec family protein kinases. Biochem Biophys Res Commun 265: 151-7. doi: 10.1006/bbrc.1999.1407

Okumura N, Terasawa F, Takezawa Y, Hirota-Kawadobora M, Inaba T, Fujita N, Saito M, Sugano M, Honda T (2012) Heterozygous Bβ-chain C-terminal 12 amino acid elongation variant, BβX462W (Kyoto VI), showed dysfibrinogenemia. Blood Coagul Fibrinolysis 23: 87-90. doi: 10.1097/MBC.0b013e32834cb243

Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M (2012) PON-P: integrated predictor for pathogenicity of missense variants. Hum Mutat 33: 1166-74. doi: 10.1002/humu.22102

Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics 9 Suppl 1: S1. doi: 10.1186/1471-2164-9-s1-s1

Pan M, Kalie E, Scaglione BJ, Raveche ES, Schreiber G, Langer JA (2008) Mutation of the IFNAR-1 receptor binding site of human IFN-α2 generates type I IFN competitive antagonists. Biochemistry 47: 12018-27. doi: 10.1021/bi801588g

Perniola R, Musco G (2014) The biophysical and biochemical properties of the autoimmune regulator (AIRE) protein. Biochim Biophys Acta 1842: 326-37. doi: 10.1016/j.bbadis.2013.11.020

Piirilä H, Väliaho J, Vihinen M (2006) Immunodeficiency mutation databases (IDbases). Hum Mutat 27: 1200-8. doi: 10.1002/humu.20405

Roucou X, Gains M, LeBlanc AC (2004) Neuroprotective functions of prion protein. J Neurosci Res 75: 153-61. doi: 10.1002/jnr.10864

Schaafsma G, Vihinen M (2014) VariSNP, a benchmark database for variations from dbSNP. Hum Mutat (in press).

Schlotawa L, Radhakrishnan K, Baumgartner M, Schmid R, Schmidt B, Dierks T, Gartner J (2013) Rapid degradation of an active formylglycine generating enzyme variant leads to a late infantile severe form of multiple sulfatase deficiency. Eur J Hum Genet 21: 1020-3. doi: 10.1038/ejhg.2012.291

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11: 361-2. doi: 10.1038/nmeth.2890

Sebastião MP, Saraiva MJ, Damas AM (1998) The crystal structure of amyloidogenic Leu^{55} -> Pro transthyretin variant reveals a possible pathway for transthyretin polymerization into amyloid fibrils. J Biol Chem 273: 24715-22.

Semler O, Garbes L, Keupp K, Swan D, Zimmermann K, Becker J, Iden S, Wirth B, Eysel P, Koerber F, Schoenau E, Bohlander SK, Wollnik B, Netzer C (2012) A mutation in the 5'-UTR of *IFITM5* creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. Am J Hum Genet 91: 349-57. doi: 10.1016/j.ajhg.2012.06.011

Shen B, Vihinen M (2004) Conservation and covariance in PH domain sequences: physicochemical profile and information theoretical analysis of XLA-causing mutations in the Btk PH domain. Protein Eng Des Sel 17: 267-76. doi: 10.1093/protein/gzh030

Shi J, Lua S, Tong JS, Song J (2010) Elimination of the native structure and solubility of the hVAPB MSP domain by the Pro56Ser mutation that causes amyotrophic lateral sclerosis. Biochemistry 49: 3887-97. doi: 10.1021/bi902057a

Srikumar PS, Rohini K (2013) Exploring the structural insights on human laforin mutation K87A in Lafora disease - a molecular dynamics study. Appl Biochem Biotechnol 171: 874-82. doi: 10.1007/s12010-013-0393-x

Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E (2013) Molecular mechanisms of disease-causing missense mutations. J Mol Biol 425: 3919-36. doi: 10.1016/j.jmb.2013.07.014

Steward RE, MacArthur MW, Laskowski RA, Thornton JM (2003) Molecular basis of inherited diseases: a structural perspective. Trends Genet 19: 505-13. doi: 10.1016/s0168-9525(03)00195-1

Storz JF, Zera AJ (2011) Experimental approaches to evaluate the contributions of candidate protein-coding mutations to phenotypic evolution. Methods Mol Biol 772: 377-96. doi: 10.1007/978-1-61779-228-1_22

Støy J, Edghill EL, Flanagan SE, Ye H, Paz VP, Pluzhnikov A, Below JE, Hayes MG, Cox NJ, Lipkind GM, Lipton RB, Greeley SA, Patch AM, Ellard S, Steiner DF, Hattersley AT, Philipson LH, Bell GI (2007) Insulin gene mutations as a cause of permanent neonatal diabetes. Proc Natl Acad Sci U S A 104: 15040-4. doi: 10.1073/pnas.0707291104

Tashita H, Fukao T, Kaneko H, Teramoto T, Inoue R, Kasahara K, Kondo N (1998) Molecular basis of selective IgG2 deficiency. The mutated membrane-bound form of gamma2 heavy chain caused complete IGG2 deficiency in two Japanese siblings. J Clin Invest 101: 677-81. doi: 10.1172/jci1672

Thusberg J, Olatubosun A, Vihinen M (2011) Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat 32: 358-68. doi: 10.1002/humu.21445

Thusberg J, Vihinen M (2006) Bioinformatic analysis of protein structure-function relationships: case study of leukocyte elastase (ELA2) missense mutations. Hum Mutat 27: 1230-43. doi: 10.1002/humu.20407

Thusberg J, Vihinen M (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum Mutat 30: 703-14. doi: 10.1002/humu.20938

Timofeyeva NA, Koval VV, Ishchenko AA, Saparbaev MK, Fedorova OS (2011) Lys98 substitution in human AP endonuclease 1 affects the kinetic mechanism of enzyme action in base excision and nucleotide incision repair pathways. PLoS One 6: e24063. doi: 10.1371/journal.pone.0024063

Trovato A, Seno F, Tosatto SC (2007) The PASTA server for protein aggregation prediction. Protein Eng Des Sel 20: 521-3. doi: 10.1093/protein/gzm042

Tsukamoto H, Farrens DL (2013) A constitutively activating mutation alters the dynamics and energetics of a key conformational change in a ligand-free G protein-coupled receptor. J Biol Chem 288: 28207-16. doi: 10.1074/jbc.M113.472464

Tsumura M, Okada S, Sakai H, Yasunaga S, Ohtsubo M, Murata T, Obata H, Yasumi T, Kong XF, Abhyankar A, Heike T, Nakahata T, Nishikomori R, Al-Muhsen S, Boisson-Dupuis S, Casanova JL, Alzahrani M, Shehri MA, Elghazali G, Takihara Y, Kobayashi M (2012) Dominant-negative STAT1 SH2 domain mutations in unrelated patients with Mendelian susceptibility to mycobacterial disease. Hum Mutat 33: 1377-87. doi: 10.1002/humu.22113

Tzarum N, Komornik N, Ben Chetrit D, Engelberg D, Livnah O (2013) DEF pocket in p38α facilitates substrate selectivity and mediates autophosphorylation. J Biol Chem 288: 19537-47. doi: 10.1074/jbc.M113.464511

Wang Z, Moult J (2001) SNPs, protein structure, and disease. Hum Mutat 17: 263-70. doi: 10.1002/humu.22

Weidemann W, Reinhardt A, Thate A, Horstkorte R (2011) Biochemical characterization of the M712T-mutation of the UDP-*N*-acetylglucosamine 2-epimerase/N-acetyl-mannosaminekinase in hereditary inclusion body myopathy. Neuromuscul Disord 21: 824-31. doi: 10.1016/j.nmd.2011.06.004

Vihinen M (1987) Relationship of protein flexibility to thermostability. Protein Eng 1: 477-80.

Vihinen M (2014a) Variation Ontology for annotation of variation effects and mechanisms. Genome Res 24: 356-64. doi: 10.1101/gr.157495.113

Vihinen M (2014b) Variation Ontology: annotator guide. J Biomed Semantics 5: 9. doi: 10.1186/2041-1480-5-9

Vihinen M, Vetrie D, Maniar HS, Ochs HD, Zhu Q, Vorechovsky I, Webster AD, Notarangelo LD, Nilsson L, Sowadski JM, et al. (1994) Structural basis for chromosome X-linked agammaglobulinemia: a tyrosine kinase disease. Proc Natl Acad Sci U S A 91: 12803-7.

Wilson PG (2008) Centriole inheritance. Prion 2: 9-16.

Vitkup D, Sander C, Church GM (2003) The amino-acid mutational spectrum of human genetic disease. Genome Biol 4: R72. doi: 10.1186/gb-2003-4-11-r72

Vogt G, Chapgier A, Yang K, Chuzhanova N, Feinberg J, Fieschi C, Boisson-Dupuis S, Alcais A, Filipe-Santos O, Bustamante J, de Beaucoudrey L, Al-Mohsen I, Al-Hajjar S, Al-Ghonaium A, Adimi P, Mirsaeidi M, Khalilzadeh S, Rosenzweig S, de la Calle Martin O, Bauer TR, Puck JM, Ochs HD, Furthner D, Engelhorn C, Belohradsky B, Mansouri D, Holland SM, Schreiber RD, Abel L, Cooper DN, Soudais C, Casanova JL (2005) Gains of glycosylation comprise an unexpectedly large group of pathogenic mutations. Nat Genet 37: 692-700. doi: 10.1038/ng1581

Xu B, Hua QX, Nakagawa SH, Jia W, Chu YC, Katsoyannis PG, Weiss MA (2002) A cavity-forming mutation in insulin induces segmental unfolding of a surrounding α-helix. Protein Sci 11: 104-16. doi: 10.1110/ps.32102

Xu J, Wang S, Zhao WJ, Xi YB, Yan YB, Yao K (2012) The congenital cataract-linked A2V mutation impairs tetramer formation and promotes aggregation of βB2-crystallin. PLoS One 7: e51200. doi: 10.1371/journal.pone.0051200

Yamada M, Sekiguchi K (2013) Disease-associated single amino acid mutation in the calf-1 domain of integrin α3 leads to defects in its processing and cell surface expression. Biochem Biophys Res Commun 441: 988-93. doi: 10.1016/j.bbrc.2013.11.003

Yates CM, Sternberg MJ (2013) The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J Mol Biol 425: 3949-63. doi: 10.1016/j.jmb.2013.07.012

Yin S, Ding F, Dokholyan NV (2007) Eris: an automated estimator of protein stability. Nat Methods 4: 466-7. doi: 10.1038/nmeth0607-466

Yu ZH, Xu J, Walls CD, Chen L, Zhang S, Zhang R, Wu L, Wang L, Liu S, Zhang ZY (2013) Structural and mechanistic insights into LEOPARD syndrome-associated SHP2 mutations. J Biol Chem 288: 10472-82. doi: 10.1074/jbc.M113.450023

Yuan D, Keeble AH, Hibbert RG, Fabiane S, Gould HJ, McDonnell JM, Beavil AJ, Sutton BJ, Dhaliwal B (2013) Ca^{2+}-dependent structural changes in the B-cell receptor CD23 increase its affinity for human immunoglobulin E. J Biol Chem 288: 21667-77. doi: 10.1074/jbc.M113.480657

Yue P, Li Z, Moult J (2005) Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol 353: 459-73. doi: 10.1016/j.jmb.2005.08.020

Zhang Z, Miteva MA, Wang L, Alexov E (2012) Analyzing effects of naturally occurring missense mutations. Comput Math Methods Med 2012: 805827. doi: 10.1155/2012/805827

Zhu Q, Zhang M, Rawlings DJ, Vihinen M, Hagemann T, Saffran DC, Kwan SP, Nilsson L, Smith CI, Witte ON, Chen SH, Ochs HD (1994) Deletion within the Src homology domain 3 of Bruton's tyrosine kinase resulting in X-linked agammaglobulinemia (XLA). J Exp Med 180: 461-70.

FIGURE LEGENDS

Fig. 1. Protein variation types provide information for variation classification and origin.

Fig. 2 Amino acid variation types. The possible variation types with examples in a short protein sequence. Amino acid deletions, indels and insertions are of two types, sequence retaining or amphigoric. Protein truncations can apper either at N- or C-terminus. Original sequence left in the variants is *underlined*

Fig. 3 Examples of variation effects at protein structures. **a** The normal (*left* PDB entry 1E04) and hyperstable (*right* 4EB1) forms of SERPINC1. The β-sheet A is in *yellow* except for the reactive center loop (RCL) in *red*. In the hyperstable form, the RCL is in *blue*, the insertion in *red* and the surface loop with altered conformation in *orange*. **b** Eye lens βB2-crystallin (2BB2) tetramer structure for bovine protein. The monomers are coded with *different colors*. The variant affects tetramer formation. **c** Normal insulin structure, *left* (4INS), and cavity forming variant, *right* (1K3M). For NMR structures the best representative structure, in this case chain number 1, is used, otherwise the first one was taken. The two chains after proteolytic cleavage are color coded as *yellow* and *cyan* in the *left*. The structural alterations generating a cavity are shown with *magenta* to the *right*. **d** Substitution G584W (*red*) causes a number of sterical clashes (indicated by *green lines* for *red* colored atoms) when introduced into the core of BTK kinase domain (1K2P). The variant is harmful because it causes structural alterations. **e** IFN-α2 (1ITF) showing residue R120 (*red*) in which variations affect binding and residues N65, L80, Y85 and Y89 (*yellow*) essential for IFNAR-1

binding. **f** BTK kinase domain structure (1K2P). The catalytic residues K430 (*top*) and R525 (*below*) are in *red* and causative of XLA when substituted. ATP binding amino acids are in *cyan*. **g** BTK SH2 domain (2GE9). The secondary structural elements are color coded, α-helices in *yellow* and β-strands in *cyan*. Amino acid substitutions at L358 (*top*) and H362 (*below*) affect α-helix. **h** The open (*left* 2AR7) and closed forms (*right* 3NDP) of adenylate kinase 4. Variations at L171 (*red*) alter the conformation of LID domain (*yellow*) in the structure to the right. **i** Variation at C260 (*red*) prevents formation of disulfide bridge with C203 (*yellow*) in haemochromatosis HFE protein (1A6Z). **j** Local conformational change at CFTR protein. Residues 509–511 in *yellow* have altered surface accessibility when the normal form (*left* 2BBO) is compared to ΔF508-containing form (*right* 2BBT). **k** Native structure of prion protein (1QLX). The α-helices are in *cyan* and β-strands in *yellow*. Irreversible structural conformational change turns the protein infectious and capable of converting structures of native prion proteins. **l** Sequence conservation in the 58 seed sequences for SH2 domain in Pfam (PF00017). The most conserved sites are *blue* and the most variable ones *red*. **m** Coevolving ELANE positions (3Q76). Those amino acids involved in disease are in *red* others are in *yellow*. These sites appear in crucial positions at protein core. **n** Surface of the normal (*left* 2SHP) and Y279C variant (*right* 4DGX) forms of SHP2. The variant has major impact on the accessibility of residues. The central P-loop, pTyr-loop, Q-loop, D′ E-loop and WPD-loop are in *yellow* and residue 279 in *red*. **o** Variation R234W (*right* 3HX3) buries exposed region of normal CRALBP (*left* 3HY3). Residue 234 is in *red* (color figure online)

Fig. 4. Protein functional variations, clockwise from *top*: protein catalytic function, protein information transfer, protein movement, protein recognition, protein storage, structural protein, and transport function of protein

Fig. 5. Protein structural variation. Organization of the descriptive VariO terms, which facilitate very detailed annotation of observed effects.

Fig. 6. Examples of variation effects at protein structures. **a** Variation ΔF508 in CFTR exposes the side chain of V510 (*right* 2BBT) when compared to the normal protein (*left* 2BBO). **b** Conformational transition of P250 (*red*) in CD23 (4G9A and 4G96) lectin-like head domain has an effect on protein affinity. Surrounding Ca^{2+} binding residues E249 and T251 are in *yellow*. The Ca^{2+} ion is shown in *magenta*. **c** Multifunctional enzyme type 2 (1ZBQ). The two chains of the dimer are in *cyan* and *blue*. Variation at E232 (*magenta*) has an effect on enzyme kinetics. Catalytic site residues are in *yellow*, residues at the lip of the cavity leading to active site in *pink*, and K246 forming salt bridge together with E232 in *green*. This salt bridge stabilizes the dimer as it appears between the monomers. **d** Variation at Y258 (*red*) has an effect on protein specificity by altering the conformation of the DEF site interaction pocket (*yellow*) in p38 MAP kinase (4GEO). **e** Several variants in BTK PH domain (1BTK) have an effect on protein electrostatics. The surface of the normal (*left*) and variant version of the domain (*right*) are colored according to the electrostatics. To the structure right, all PH domain changes affecting polarity/charge are combined. **f** Variations for A57T, L121P and C151S or Y (*red*) in ELANE (3Q76) have effects on residue contact energies. **g** Substitution of L55 in *red* (*left* 1TTA) by proline (*right* 5TTR) alters the surface and structure of transthyretin and leads to protein aggregation. **h** Substitution at residue 444 (*red*) affects the dimer formation of mitochondrial dilipoamide dehydrogenase (1ZMC). The conformational change alters the catalytic site (*yellow*) formed of residues from both the subunits (*cyan* and *gray*) and exposes catalytic site for neomorphic protease activity (*magenta*) (color figure online)

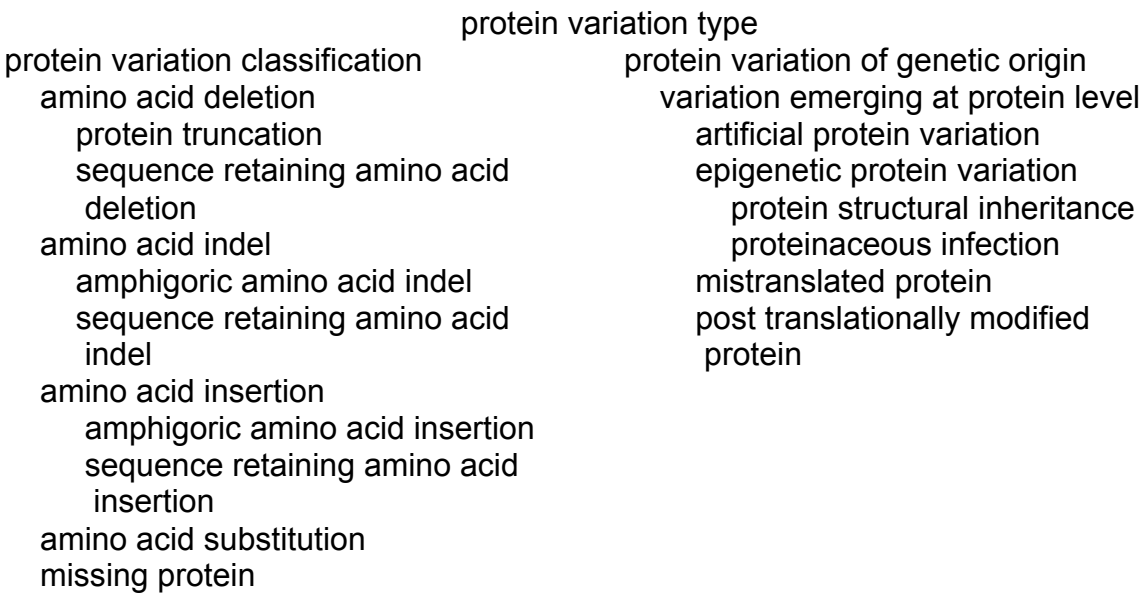
protein variation type
protein variation classification
amino acid deletion
protein truncation
sequence retaining amino acid deletion
amino acid indel
amphigoric amino acid indel
sequence retaining amino acid indel
amino acid insertion
amphigoric amino acid insertion
sequence retaining amino acid insertion
amino acid substitution
missing protein
protein variation of genetic origin
variation emerging at protein level
artificial protein variation
epigenetic protein variation
protein structural inheritance
proteinaceous infection
mistranslated protein
post translationally modified protein

Fig 1.

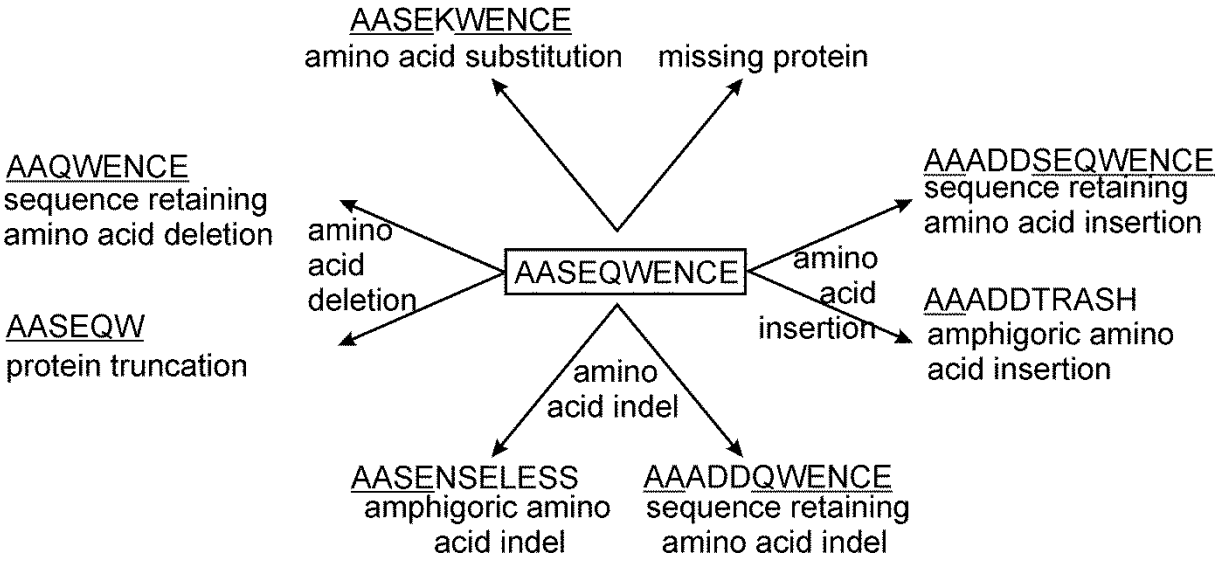
AASEKWENCE
amino acid substitution
missing protein
AAQWENCE
sequence retaining
amino acid deletion
AASEQW
protein truncation
amino
acid
deletion
AASEQWENCE
amino
acid
insertion
AAADDSEQWENCE
sequence retaining
amino acid insertion
AAADDTRASH
amphigoric amino
acid insertion
amino
acid indel
AASENSELESS
amphigoric amino
acid indel
AAADDQWENCE
sequence retaining
amino acid indel

Fig. 2

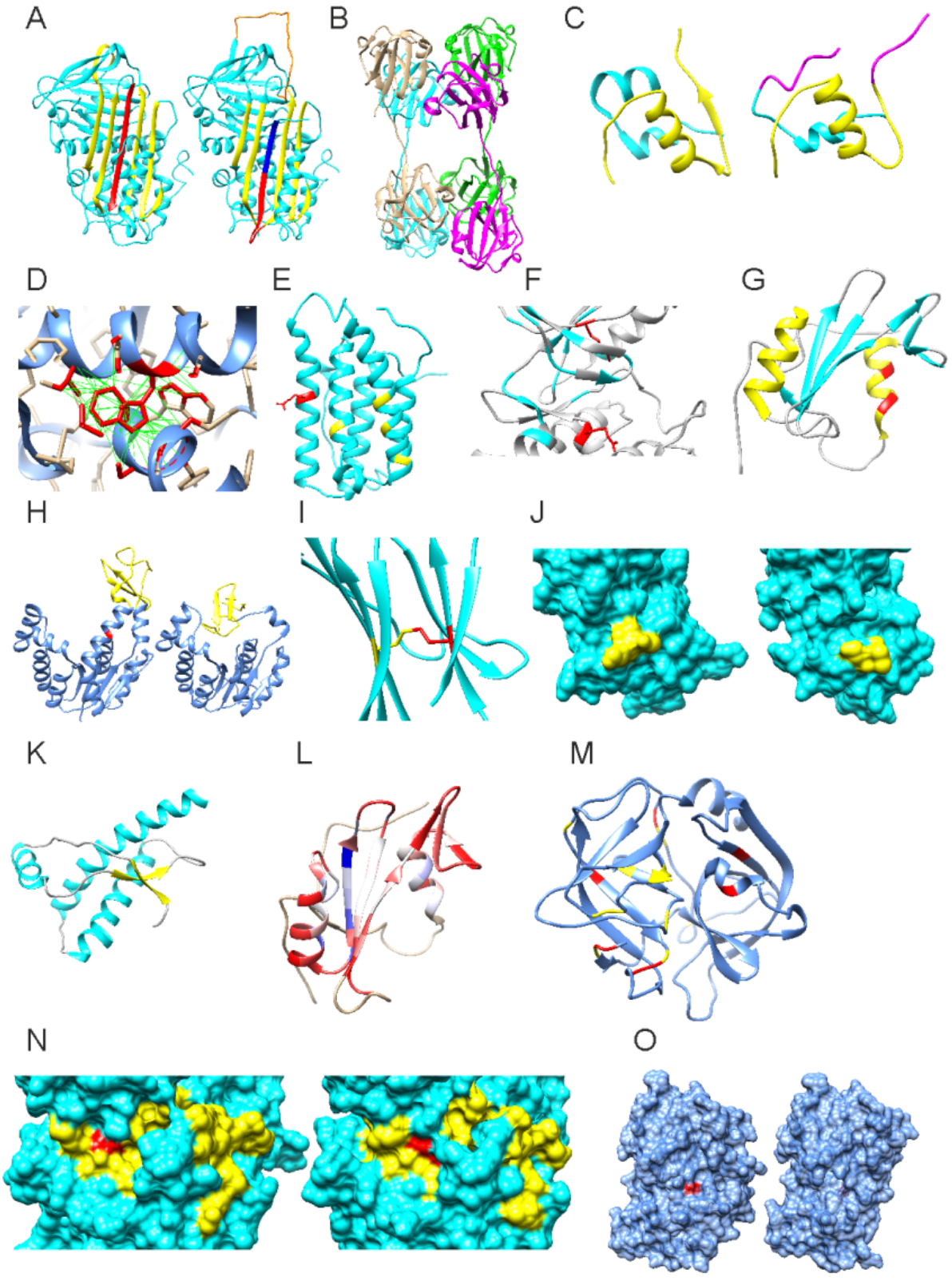

Fig 3.

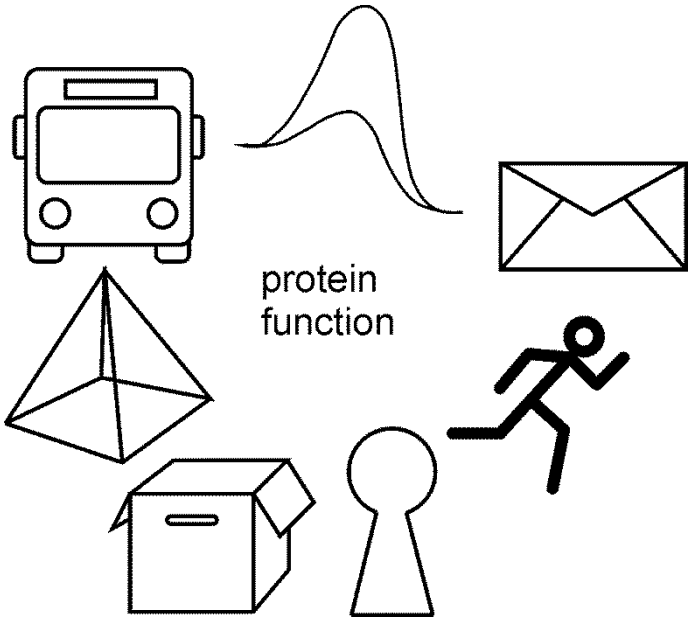

Fig. 4.


variation affecting protein structure
effect on protein 3D structure
epigenetic protein modification
effect on protein structural flexibility
effect on protein structural disorder
effect on protein packing
effect on protein binding site
effect on protein processing
effect on protein interaction site
effect on protein post translational modification
effect on protein secondary structural element
effect on protein fold
effect on protein quaternary structure forming interaction
effect on protein quaternary structure
effect of protein tertiary structure
effect on protein amino acid size
effect on protein folding rate
effect on protein complex 3D structural change
effect on protein protein dynamics
effect on protein induced fit
effect on protein allosteric site
effect on protein catalytic site
effect on protein splicing
effect on protein variation generating a novel prot. processing site
effect on protein variation preventing polypeptide processing
effect on protein prot. cavity formation
effect on protein protein overpacking
effect on protein C-terminal protein elongation
effect on protein N-terminal protein elongation
effect on protein protein fusion
effect on protein global protein conformation change
effect on protein local protein conformation change
effect on protein prot. conform. allonal change
effect on protein eff. on prot. disulphide formation
effect on protein changed domain orientation
effect on protein eff. on prot. st turn
effect on protein eff. on prot. st staple
effect on protein eff. on prot. st loop
effect on protein eff. on prot. schellman loop
effect on protein eff. on prot. niche motif
effect on protein eff. on prot. nest motif
effect on protein eff. on prot. gamma turn
effect on protein eff. on prot. beta turn
effect on protein eff. on prot. beta bulge loop
effect on protein eff. on prot. beta bulge
effect on protein eff. on prot. asx turn
effect on protein eff. on prot. asx motif
effect on protein eff. on prot. alpha beta motif
effect on protein eff. on left handed protein helix
effect on protein eff. on right handed protein helix
effect on protein eff. on prot. anti-parallel beta strand
effect on protein eff. on prot. parallel beta strand
effect on protein deletion or additional modification site
effect on protein generation of post translational modification site
effect on protein eff. on membrane prot. intramembrane region
effect on protein eff. on membrane prot. extramembrane region
effect on protein eff. on prot. turn motif
effect on protein eff. on prot. beta strand
effect on protein eff. on prot. prot. coil
effect on protein eff. on prot. prot. helix
effect on protein effect on trans-membrane polypeptide
effect on protein epigenetic protein complex structure
effect on protein prion formation
effect on protein effect on alpha helix
effect on protein effect on pi helix
effect on protein effect on three ten helix

Fig. 5.

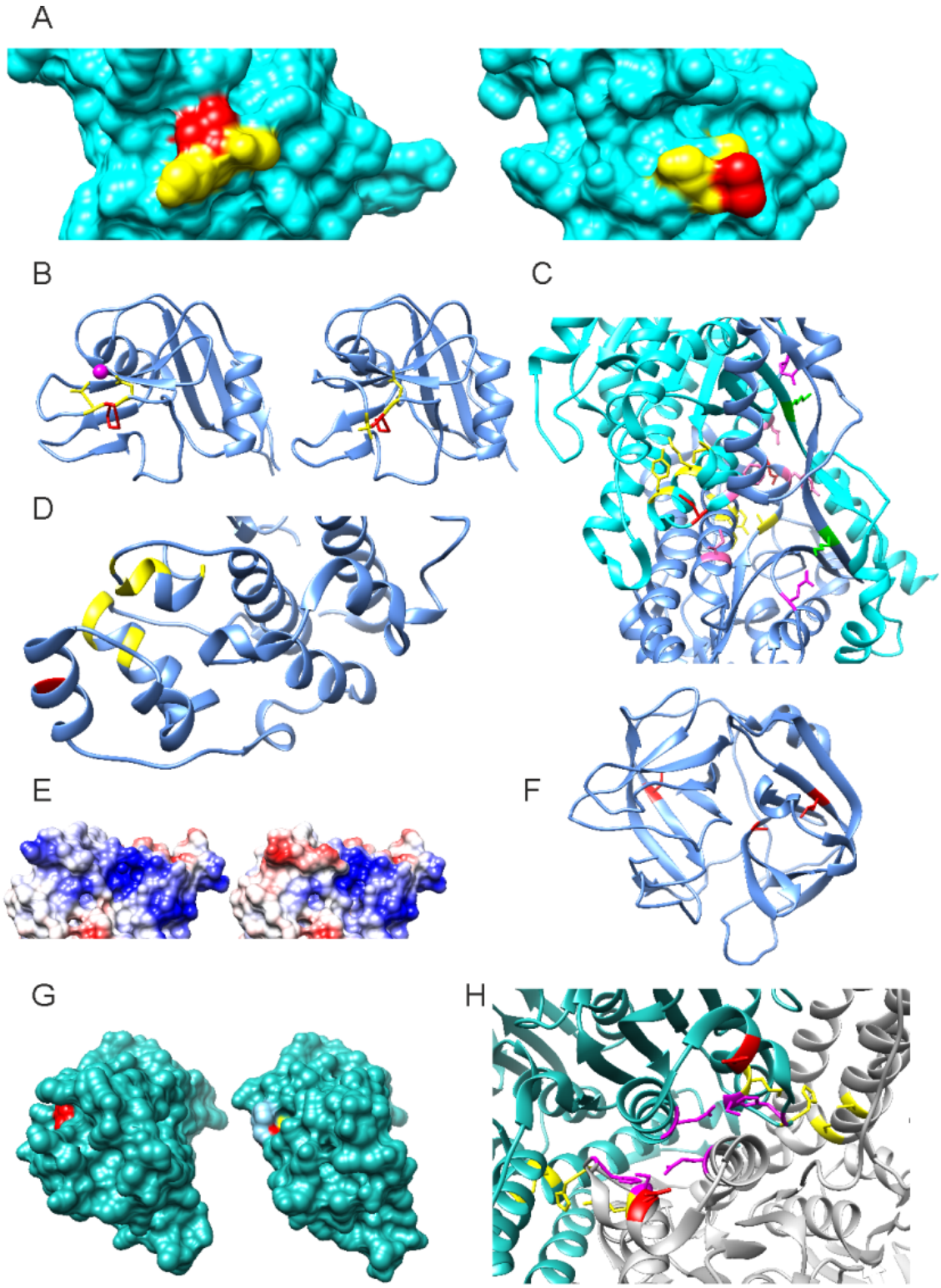

Fig. 6.