Structural Impact of SNPs

Affiliation

Andrew C.R. Martin -and- Anja Baresic

Institute of Structural and Molecular Biology, Division of Biosciences

University College London

London

U.K.

andrew@bioinf.org.uk

Synonyms

- Structural Impact of Single Amino Acid Polymorphisms (SAAPs)
- Structural Impact of Single Amino Acid Mutations

Definition

The simplest form of mutation is a single DNA base change, frequently referred to as a 'single nucleotide polymorphism' (SNP). Strictly, this term should only be applied to single base changes that are observed in at least 1% of a 'normal' population. However, it is frequently used to refer to any single base mutation and is used in that context here. Many SNPs occur in non-coding regions of DNA, where they may affect transcription, mRNA splicing, or mRNA stability. When a single base change occurs in an exon, it will fall into one of three classes: (i) a '**synonymous**' mutation which does not change the amino acid sequence of the resultant protein (although this may still affect expression, splicing or mRNA stability), (ii) a '**nonsense**' mutation resulting in premature termination of the protein sequence, or (iii) a '**non-synonymous**' (or 'missense') mutation (an nsSNP) resulting in a single amino acid change. At the protein level, an nsSNP results in a 'single amino acid polymorphism' (SAAP, a term we use regardless of frequency in the population). Any amino acid change will have some structural effect on the protein, possibly leading to misfolding.

Introduction

SAAPs result in a phenotype which can be classified in terms of its effect on the fitness of the individual: (i) **beneficial** – resulting in an increase in fitness, (ii) **neutral** – having no visible effect on fitness, and (iii) **deleterious** or **pathogenic** – resulting in a decrease in overall fitness. Interestingly some pathogenic mutations introduce an amino acid which is the native amino acid in another species, a concept known as a 'compensated pathogenic deviation' (Baresic & Martin, 2011).

The fraction of individuals having the mutation who exhibit the phenotype is referred to as the '**penetrance**' – 100% penetrance mutations are inherited in a Mendelian fashion, while lower penetrance mutations often express their phenotype as a result of epistatic interactions with other mutations, or with the environment.

Any change in phenotype is likely to result from structural effects that the mutation has on the protein structure. Such effects fall into three broad categories: (i) **functional** mutations – those which affect residues directly involved in protein function, (ii) **fold-preventing** mutations – those which physically stop the protein from being able to fold up in the correct way, (iii) **fold-destabilizing** mutations – those

which do not stop the protein from being able to fold up in the correct way, but which destabilize the correct fold with respect to misfolded or unfolded conformations. This is the most interesting category as it may be possible to design drugs to bind to the correctly folded form thus stabilizing it and restoring activity through the Law of Mass Action. Since most proteins have a surprisingly small thermodynamic stability (between −3 and −10 kcal/mol) and most SAAPs lead to a ΔΔG of 0.5 − 5 kcal/mol, it is not surprising that many SAAPs will destabilize the protein (DePristo et al., 2005).

Structural Effects

Structural effects of mutations include:

- **Introducing large residues that clash** - Clearly, if a small residue is replaced by a larger (or more bulky) residue, there is a strong possibility that the replacement sidechain will clash with the surrounding amino acids. This will destabilize the protein and is likely to prevent it from folding correctly.
- **Introducing a void** - Conversely, if a buried large sidechain is replaced by a smaller sidechain, a void may be left in the protein. Voids are unfavourable both for enthalpic and entropic reasons and consequently the protein will be destabilized and is likely to change its conformation to fill the void.
- **Mutations to proline** - Proline is unique amongst the amino acids as its cyclic sidechain, which links back onto the backbone nitrogen (strictly making it an 'imino' acid), restricts the rotational freedom of the bond between the backbone nitrogen and the Cα (the φ angle). Consequently, if another amino acid, adopting a φ angle not allowed for proline, is mutated to a proline, some structural rearrangement will have to occur to accommodate the proline.
- **Mutations from glycine** - Glycine lacks a sidechain. This means that there is less conformational restriction on combinations of backbone φ and ψ angles; in particular, positive φ angles are largely inaccessible to the other amino acids. If a glycine adopting a backbone conformation not accessible to other amino acids is mutated to any other amino acid then some structural rearrangement of the backbone will have to occur.
- **Disruption of *cis*-prolines** - The unique nature of proline also means that, unlike the other amino acids, the *cis*-isomer of the peptide bond preceding the proline has approximately the same stability as the *trans*-isomer. Consequently, a mutation of a *cis*-proline to any other amino acid will destabilize the protein structure probably leading to a change in backbone conformation from the *cis*- to *trans*-isomer.
- **Introduction of hydrophobic residues on the surface** - Protein folding is driven by the hydrophobic effect – an entropy-driven need to bury hydrophobic residues and expose hydrophilic residues on the surface. Thus a mutation which introduces a hydrophobic residue on the surface will destabilize the protein and may result in some change in folding to bury the residue.
- **Disruption of sidechain hydrogen bonds** - When hydrophilic (hydrogen-bond capable) residues do occur inside proteins, hydrogen bonding capability is almost always fully satisfied; in other words, a buried sidechain capable of making a hydrogen-bond interaction almost always *will* be making a hydrogen bond (McDonald & Thornton, 1994). Consequently, if a sidechain is involved in a hydrogen bond (with another sidechain, or with backbone) and the resultant sidechain after a mutation cannot maintain that hydrogen bond, then the protein will be destabilized.
- **Introducing a hydrophilic or charged residue in the protein core** - For the same reasons that disrupting a hydrogen bond will destabilize a protein, introducing a hydrophilic sidechain into the

core will result in an unsatisfied hydrogen bond, thus destabilizing the protein. If the residue is charged, then it will be further destabilizing as the charge will not be satisfied.

- **Mutations to residues involved in an interface or binding** - Proteins function through their interactions with ligands or other proteins; additionally many proteins function as protein complexes (homo- or hetero-multimers). Consequently a mutation to a residue involved in an interface, or making specific binding interactions with another protein or ligand, is likely to affect function.
- **Mutations to other functional residues** - In addition to binding, residues may be involved in catalytic mechanisms or post-translational modifications. Mutations to such residues may affect function.
- **Mutations to or from cysteine** - Cysteines may be involved in disulphide bonds. Mutations to these may destabilize the protein. If a protein contains disulphides, then introducing an additional cysteine may lead to scrambling of disulphides and misfolding.

Residues which are highly conserved probably have some structural importance even if this cannot be determined directly. At the structural level, residues that are close to highly-conserved patches may be functionally important even if they are not highly conserved themselves.

Predicting and Visualizing Structural Effects

A number of tools have been developed to predict and visualize the structural effects of mutations including:

- **TopoSNP** (http://gila.bioengr.uic.edu/snp/toposnp/) provides sequence-based metrics, but maps mutations to structure.
- **ModSNP** accessible via Swiss-Prot web pages, provides information from Swiss-Prot features and maps the mutation to structure using homology models where necessary.
- **stSNP** (http://ilyinlab.org/StSNP/) links SNPs with pathway information from KEGG and maps mutations to structure.
- **SNPeffect** (http://snpeffect.vib.be/) focuses on information related to catalytic active sites, secondary structure formation, aggregation, amyloid formation, solvent-accessibility and trans-membrane or cellular localization as well as phosphorylation and glycosylation sites.
- **SAAPdb** (http://www.bioinf.org.uk/saap/db/) provides a detailed, pre-calculated, analysis of structural effects of mutations as described above. A server for novel mutations is under development.

Prediction of Phenotypic Effects

The importance of mutations in disease has led to a number of methods for predicting whether mutations will be tolerated. Many of these rely only on sequence information (e.g. SIFT (http://sift.bii.a-star.edu.sg/) and Polyphen (http://genetics.bwh.harvard.edu/pph/)), but some make use of structural information including:

- **nsSNPAnalyzer** (http://snpanalyzer.uthsc.edu/) uses solvent-accessibility and secondary structure elements.
- **SNAP** (http://www.rostlab.org/services/SNAP/) uses solvent accessibility to optimize sets of training features.

- **PMUT** (http://mmb2.pcb.ub.es:8080/PMut/) uses secondary structure to select between optimized neural networks.
- **SNPs3D** (http://www.snps3d.org/) uses structural effects such as solvent accessibility, hydrophobicity, electrostatic interactions and atomic packing to estimate thermodynamic stability.
- **PolyPhen-2** (http://genetics.bwh.harvard.edu/pph2/) uses a combination of 8 sequence-based and 3 structure-based features and contains models optimized for Mendelian-inherited and low-penetrance nsSNPs.

Cross-References

→ Amyloid formation
→ Homology modelling of protein structures
→ Law of Mass Action
→ Misfolding and aggregation
→ Misfolding of proteins
→ Protein-protein complexes
→ Splicing

References

Baresic, A. and Martin, A.C.R. (2011) Compensated pathogenic deviations. BioMolecular Concepts, 2: 281–292.

DePristo, M.A., Weinreich, D.M. and Hartl, D.L. (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. Nature Reviews Genetics, 6: 678–687.

McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen-bonding potential in proteins. J. Mol. Biol., 238: 777–793.