Literature Survey - Effects of In-frame Single Codon Deletions on Protein Function

Elizabeth Lagesse
June 1, 2016

Introduction

My main focus in this survey will be deletion mutations involving exactly three nucleic acids - which result in the removal of exactly one amino acid from the resulting protein. However, this specific phenomenon is somewhat rare, and is difficult to detect through traditional sequencing - so I will touch on some relevant similar mutations.

One might expect that perfect deletions of three nucleotides would be statistically rare - and they are. Mutation is often caused by DNA damage (and/or a failure of cellular machinery) - leading to a greater proportion of frame shifting mutations, such as single nucleotide polymorphisms. Imperfect removal of introns during transcript processing can cause similar sequence problems at the RNA level - and concomitant disease. All of these situations are statistically more likely to result in frame shifting mutations (multiples of 1 or 2, rather than 3), and mutations of greater or fewer than 3 bases.

A countervailing force is that frame-shift mutations are usually catastrophically detrimental to the function of the protein (and often the fitness of the organism) - and are thus eliminated quickly through negative selection. This would lead to a greater number of in-frame mutations remaining in the population once they occur.

Like most mutations, in-frame deletion mutants are most often deleterious to the organism in question, but beneficial effects - like antibiotic resistance, disease resistance, or viral antigenic drift - can occur ‘by accident’.

Deletions in the Human Genome

Insertions and deletions are often referred to as INDELs, but this term has different specific meanings within different branches of evolutionary biology. For the purposes of this discussion, I will be using the term to simply refer to a small insertion or deletion of nucleotides - in or out of frame. Because of limitations inherent to the technology used to perform genomic sequencing, INDELs are less well understood than single nucleotide polymorphisms. However, it is possible that they account for as much or more human variation.

When the genomes of healthy humans were surveyed for INDELs, several hundred were found in coding regions. (Note: Insertions and deletions are thought to be more common in non-coding regions for both mechanistic and mutation tolerance reasons. Despite the diverse and continuously emerging regulatory functions of these regions, this discussion will focus on protein function - and thus on coding regions.) Of these, 61% were in-frame, and the remaining 39% caused frame shifts. (Mullaney, 2010)

Deletion Effects in Clinical and Industrial Settings

Despite the expectation that frame-shift mutations would be the most disruptive to protein structure and function - studies of tumor suppressors have found that even single amino acid deletions can greatly disrupt suppressive function. (Guan, 2012)

There are are also multiple clinically severe human diseases thought to be caused by a single codon deletion - ranging from cystic fibrosis (one of the most common genetic diseases), to relatively rare metabolic disorders.

The most common cause of cystic fibrosis is a 50,000-year-old, in-frame, single codon deletion mutation in the cystic fibrosis transmembrane conductance regulator (CFTCR) gene (ΔF508), which prevents the protein from folding properly. CFTCR is mainly involved in ion transfer and regulation of cellular water content. Homozygotic mutants cannot regulate the water content of the mucus lining many interior surfaces of the body. This leads to problems of the lung, liver, and various parts of the digestive tract.

However, despite the dire consequences for homozygotes, the ΔF508 mutation has remained present in some populations for tens of thousands of years. This is attributed to increased resistance to cholera, tuberculosis, and cholera among heterozygotes. It is therefore unsurprising that the mutation became widespread among European populations, where these diseases were historically endemic.

Fascinatingly, the genetic change that causes ΔF508 is not actually a clean removal of three adjacent nucleotides to preserve the reading frame. It involves the deletion of one base, causing a shift which leads to a synonymous mutation (a different codon for the same amino acid) in an adjacent residue. A separate, 2 base deletion also occurs nearby, causing a net loss of exactly one codon. It seems likely that combination events such as this increase the overall likelihood of single codon mutations over that of perfect, in-frame, three-nucleotide deletion events.

A potentially unique mutation involving single codon deletion has even been implicated in a rare case of hemophilia in a female. (Vencesla, 2008) Since full gene sequencing is not yet a common part of clinical diagnosis (and the difficulty of finding INDELs using some sequencing methods), cases like this imply that we will continue to discover more conditions with single codon deletions at their root as personal medicine becomes widespread.

It may be possible to attribute the advent of the H1 influenza strains, such as the infamous H1N1, to a single amino acid deletion - ΔK134 - from hemagglutinin. Genetic analysis has found this mutation at the root of an Asian branch of the influenza lineage, which lead eventually to current H1 strains. (McDonald, 2007)

Interestingly, in-frame single codon deletions can also cause confer resistance to a specific class of herbicides - to which resistance is otherwise quite rare. (Patzoldt, 2006) They work by inhibiting protoporphyrinigen oxidase, which has multiple isoforms targeted by the herbicide. The codon deletion is in the nuclear DNA, in a position that affects all isoforms - conferring resistance to multiple herbicide targets simultaneously.

β-lactamase Specific Deletion Effects

TEM-1 β-lactamase is an antibiotic resistance factor involved in the metabolism of the β-lactam class of drugs. These drugs include some of the most widely used antibiotics in humans (penicillin, ampicillin, amoxicillin, etc), those used most frequently in animal agriculture (penicillin, cephalosporin, etc), as well as more recently developed drugs of 'last resort" such as methicillin and the carbapenems.

There are also many types of β-lactamase - all of which share the trait that they break the characteristic 4-member ring of β-lactams. TEM-1 β-lactamase is the most common sub type found in Gram negative bacteria.

Experimental groups have used β-lactamase as an model for developing deletion mutation methods capable of deleting exactly three nucleotides (Jones, 2005) - which have been applied to the study of deletion in other systems. This process was not able to obtain the kind of full-sequence data we expect from the Ostermeier lab's current work, but they were able to show that TEM-1 β-lactamase is tolerant of deletions in several regions - especially loops. These results accord with what is already known about the regions where proteins tend to be tolerant of deletions.

Computational Prediction of Deletion Effects

Rosetta has been used to predict deletion effects using structure predictions - most notably Berrondo, 2011 - on which my first method was based. Her work focused on understanding the fitness landscape of Ricin deletion mutants.

Other computational groups have taken a sequence alignment based approach, where sequences are aligned and scored using measures that take deletions and insertions into account. (Choi, 2012) Previous efforts have largely involved comparisons of homologues to find evolutionarily conserved residues, with the understanding that these are likely to be functionally important.

Unfortunately, all of these techniques fail to take secondary and tertiary structural characteristics into account - and thus are useful in addition to, but not as a substitute for, structure prediction based methods.

References:

- Vencesla, et al. (2008) Severe hemophilia A in a female resulting from an inherited gross deletion and de novo codon deletion on the F8 gene. *Haemophilia*,14, 1094-1098
- Uphadhyaya, et al. (2007) An Absence of Cutaneous Neurofibromas Associated with a 3-bp Inframe Deletion in Exon 17 of the NF1 Gene: Evidence of a Clinically Significant NF1 Genotype-Phenotype Correlation. *Am. J. Hum. Genet.* 80, 140-151
- Weedon, et al. (2013) An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. *Nat. Genet.* 45, 8, 947-951
- Patzoldt, et al. (2006) A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *PNAS*, 103, 33, 12329-12334
- Guan, et al. (2012) Functional Analysis of In-frame Indel ARID1A Mutations Reveals New Regulatory Mechanisms of Its Tumor Suppressor Functions. *Neoplasia*, 14,10, 986-993
- McDonald, et al. (2007) Antigenic drift in the evolution of H1N1 influenza A viruses resulting from deletion of a single amino acid in the haemagglutinin gene. *J. Gen. Virol.* 88, 3209-3213
- Jones. (2005) Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 b-lactamase to an amino acid deletion. *Nuc. Acids Res.*, 33, 9
- Hall, Barlow. (2004) Evolution of the serine beta-lactamases: past, present, and future. *Drug Resist. Updat.* 7, 2, 111-123
- Mullaney, et al. (2010) Small insertions and deletions (INDELs) in human genomes. *Hum. Molec. Genet.* 19, Review Issue 2
- Berrondo, Gray. (2011) Computed structures of point deletion mutants and their enzymatic activities. *Proteins*, 79, 2844-2860
- Choi, et al. (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLOS one*, 7, 10
- Pascarella, Argos. (1992) Analysis of insertions/deletions in protein structures. *J. Mol. Biol.* 224, 461–471.