Scores & output values

AmyScan v1.0

Method:
AmyScan is run on the SNP and WT sequences and outputs the amylogenic regions in the sequences.

Significant change

Reported when an additional amylogenic region is detected in either WT or SNP sequence due to the amino acid change.

Prediction of amino acid sequences that can be bound by the Hsp70 chaperone

Method:
The output of this method are the regions in both WT and SNP sequences that are likely to be Hsp70 binding sites.

Significant change

Reported when an additional Hsp70 is detected in either WT or SNP sequence due to the amino acid change.

Conservation

Method:
The amino acid sequences of the wild type are blasted against the SwissProt database with an e-value of 10^{-4}. After that a detailed view of the column in the alignment where the SNP amino acid change occurs is taken.
The metrics that are used are:
the number of times the wild type amino acid residue occurs in the column
the number of times the SNP amino acid occurs in the column
the number of gaps in the column
After that, the Zvelebil method for counting amino acid properties is used. A truth table (Table 1) is used to map amino acids to 10 properties.

For an alignment column, we use a weighted average for each property by counting the truth values for each entry in the alignment column.
In SNPeffect, the properties that are conserved in the alignment are reported. If the average for a property is >0.66, it is considered that it is conserved to have that property.

An example is given below.

Table 1. Zvelebil truth table for amino acid properties.

	I	L	V	C	A	G	M	F	Y	W	H	K	R	E	Q	D	N	S	T	P	B	Z	X	gap
Hydrophobic	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	1	1
Polar	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
Small	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	1	1
Proline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1
Tiny	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
Aliphatic	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Aromatic	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1
Positive	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1
Negative	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1
Charged	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	1	1

Example:

WT amino acid: I
SNP amino acid: R

Alignment column:
I
L
-
V
T
R

Number of times the WT residue occurs in the alignment : 1
Number of times the SNP residue occurs in the alignment: 1
Number of sequences in the alignment: 6
Number of gaps: 1

Zvelebil method:

	I	L	-	V	T	R	average

Hydrophobic	1	1	1	1	1	0	0.8
Polar	0	0	1	0	1	1	0.5
Small	0	0	1	1	1	0	0.5
Proline	0	0	1	0	0	0	0.2
Tiny	0	0	1	0	0	0	0.2
Aliphatic	1	1	1	1	0	0	0.7
Aromatic	0	0	1	0	0	0	0.2
Positive	0	0	1	0	0	1	0.3
Negative	0	0	1	0	0	0	0.2
Charged	0	0	1	0	0	1	0.3

The alignment position is then said to be hydrophobic and aliphatic, and the differences with the SNP is that there the position is polar, positive and charged.

Significant change

Reported which properties the SNP amino acid possesses that are not conserved in the alignment.

CSA - Catalytic Site Atlas

The SNP entries that can be linked to a PDB file are queried against the CSA, and if known catalytic sites are disrupted it is reported.

Significant change

Reported when the amino acid change of the SNP occurs at a known catalytic site.

FoldX output

FoldX calculates free energy changes upon mutation on protein stability or ligand binding. The output is the total free energy change, relative to the wild type in kcal.mol^{-1}, as well as a breakdown of the change into a set of empirical energy terms (all in kcal.mol^{-1}).

Total Energy

Total free energy change of the protein stability caused by the mutation in kcal.mol^{-1}

Backbone H-bond & Sidechain H-bond

H-bonds are classified into two categories: 1) those that involve only backbone atoms and define secondary structure elements and 2) other hbonds, involving side-chain atoms. Bonds are detected based on a set of geometric constraints, energies are assigned based on the protein engineering experiments on the contribution of an Hbond to protein stability.

Van der Waals, Solvation Polar & Solvation Hydrophobic

The Van der Waals term characterises packing inside the protein, whereas the other terms characterise the difference in solvation of hydrophobic and polar residues between the unfolded and the folded state of a protein. These three terms are calculated based on experimentaly calculated transfer energies of model compounds (amino acids) from water to vapour, in which the amino acid loses both packing and solvation, and from water to organic solvents in which the amino acid loses solvation only, but packing remains similar. The Van der Waals term can thus be extracted from the difference of the two experiments. These transfers are thought to parallel the transfer of an amino acid during the folding reaction from the unfolded state where it is exposed to water to the hydrophobic interior of a protein. In FoldX the experimental values for an amino acid are recalculated to obtain a contribution per atom and scaled with the solvent accessibility.

Water bridge

Whereas the solvation terms are a good and computationaly inexpensive way to characterise the general effect of the bulk solvent on the protein stability, this representation is not capable to capture specific interactions between water molecules and protein atoms. To allow the details of such interactions to be included in the energy calculation, FoldX calulcates so-called water bridges, i.e. water molecules that make more than 1 Hbond with a protein atom. The interaction energy of these water bridges are taken into account.

Van der Waals Clashes & Torsional clash

Each atom has a Van der Waals radius, if the distance between two atoms is less than the sum of their Van der Waals radii, minus a term accounting for the interaction between atoms, they interpenetrate and a strong repulsive energy is calculated. The distinction is made between clashes between atoms from different side-chains and clashes inside a side-chain.

Electrostatics & helix dipole

The electrostatics are calculated according to Coulomb's law and includes partial charges on some amino acids such as arginine or on aromatic rings. Helix dipoles are taken into account by charges placed at the N- and C-termini of every helix that are simply included in the electrostatic calculation.

Entropy Main Chain & Entropy Side Chain

These terms characterise the entropic cost of fixing an amino acid in a certain conformation in the folded protein. The distinction is made between the main chain entropy, which is dependent on the

intrinsic tendencies of an amino acid to adopt certain dihedral angles, and the side chain entropy, which depends how restricted a given side chain is in the folded structure.

Disulfide

An energy term estimating the loss of entropy due to disulfide bridges.

Partial covalent bonds

Although the interaction between protein atoms and metal ions are often strongly electrostatic in nature, FoldX also takes a specific partial covalent term into consideration.

Cis Bond

A term penalizing *cis*-peptide bonds.

Significant change

Reported when the $\Delta\Delta G$ value for the amino acid change of the SNP has an absolute value of more than 0.7

$$|\Delta\Delta G_{SNP}| \geq 0.7 \text{ kcal.mol}^{-1}$$

O-GlycBase

SNP entries that can be linked to a SwissProt identifier are queried against the O-GlycBase.

Significant change

Reported when the amino acid change of the SNP occurs at a known glycosylation site.

PENCE Proteome Analyst Subcellular

SNP and WT sequences are submitted to the PA server. The output of this server contains information on the subcellular localisation (Mitochondrion, Nucleus, Endoplasmic reticulum, Extracellular, Cytoplasm, Plasma membrane, Golgi, Lysosome, or Peroxisome) and the probability (%) that the protein can be found in that location.

Significant change

Reported when a different subcellular localisation is reported for SNP and WT.

PhosphoBase

SNP entries that can be linked to a SwissProt identifier are queried against the O-GlycBase.

Significant change

Reported when the amino acid change of the SNP occurs at a known phosphorylation site.

PROF

The three tools comprised in PROF are PROFsec for secondary structure prediction, PROFacc for solvent accessibility prediction and PROFhtm fro prediction of transmembrane regions. The output for these files contains a prediction for each residue:

PHDsec	H=helix, E=extended (sheet), blank=other (loop)
PHDacc	relative solvent accessibility (acc) in 3 states: b = 0-9%, i = 9-36%, e = 36-100%.
PHDhtm	T=helical transmembrane region, blank=other (loop)

together with a reliability score (0-9, 9 being the most reliable score).

Significant change

Reported when the prediction for the SNP amino acid position is different for wild type and SNP sequence.

Psort II

SNP and WT sequences are submitted to the PA server. The output of this server contains information on the subcellular localisation (cytoskeletal, cytoplasmic, nuclear, mitochondrial, vesicles of secretory system, endoplasmic reticulum, Golgi, vacuolar, plasma membrane, peroxisomal, extracellular, including cell wall) and the probability (%) that the protein can be found in that location.

Significant change

Reported when a different subcellular localisation is predicted for SNP and WT.

Tango Score

Tango returns a score that represents the tendency for each amino acid to be part of a beta-sheet aggregate, it ranges from 0 for no aggregation to 100 full strongest aggregation tendency. Extensive comparison with experimental data revealed that a peptide with a stretch of more than 5 consecutive residues with a tango score of more than 5% have an 85% chance on average to aggregate *in vitro*. This tendency is further modulated by other factors such as protein stability and dynamics since a stretch with a strong TANGO tendency cannot participate in aggregation as long as it is buried inside a folded protein.

Significant change

Reported when the absolute value of difference between the Tango score of the SNP and the wild type is larger than 50.

$$|\text{Tango}_{SNP} - \text{Tango}_{WT}| \geq 50$$