# SNPeffect overview

November 17, 2011

# Chapter 1

# Introduction

In this analysis, we focus on the effect of mutation on aggregation propensity (TANGO), amyloid propensity (WALTZ) and chaperone binding (LIMBO). It is important to point out that beta aggregation is mediated by short stretches that need to become exposed by (partial) unfolding before they can actually nucleate protein aggregation. Therefore, when analysing mutations, we consider two important effects.

- the intrinsic aggregation propensity, and
- the stability of the protein.

As protein folding involves a large proportion of amino acid residues, most mutations have their effect by acting on stability, i.e. by exposing aggregation prone regions. In order to estimate the likelihood that a given short stretch may become exposed, we employ FoldX to caclulate the effect on the structural stability.
So, the presence of TANGO, WALTZ or LIMBO regions does not necessarily implies that the protein readily forms aggregates, amyloid or exposes chaperone binding regions respectively. Such regions are normally buried in the protein core, but when these regions become exposed due to other factors (e.g. by a structurally destabilizing mutation), it can become more prominent.

# Chapter 2

# Phenotypic analysis of S59I in fasta sequence sodc_humanse

## 2.1 TANGO prediction

TANGO predicts the aggregation prone regions in a protein sequence. The total TANGO score for you protein is 87.32. Mutations can increase (dTANGO >50), decrease (dTANGO <-50) or not affect aggregation propensity (dTANGO between -50 and 50). For this mutation, dTANGO equals 0.00 which means that the mutation does not affect the aggregation tendency of your protein.

In figure 2.1 and 2.2 the position of the TANGO stretches in the wild type and variant

**Figure 2.1: Bar representation of the TANGO windows present in the wild type (top) and mutant protein (bottom).** In the bar representation, the position of the aggregating stretches is visualised in red, and the dashed vertical line in the variant indicates the position of the variant residue.

**Table 2.1: TANGO regions in variant and wild type.** For each TANGO region, the start, end, sequence and score is given.

| Number | Start | End | Stretch | Score |
|---|---|---|---|---|
| **Wild Type** | | | | |
| 1 | 143 | 152 | LACGVIGIA | 8.86 |
| **Mutant** | | | | |
| 1 | 143 | 152 | LACGVIGIA | 8.86 |

protein are visualized, represented by respectively a bar or profile representation. In table 2.1, the short stretches are listed for both wild type and mutant. To compare the effect of the mutation to the WT, we also show a Difference profile (Figure 2.3), that plots the difference between WT protein and the variant.

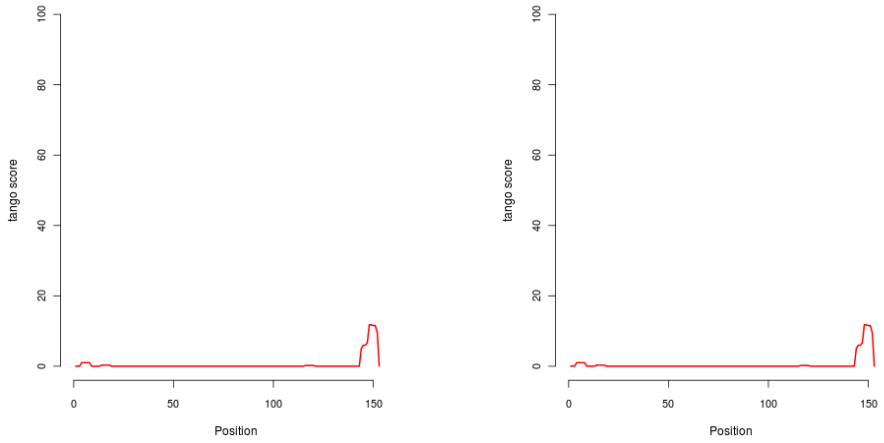


**Figure 2.2: Profile representation of the TANGO stretches in the wild type (left) and mutant (right) protein.** This graph plots the per-residue TANGO aggregation score of the wild type and variant protein. From left to right, all residue scores from the N-terminus to the C-terminus are plotted.

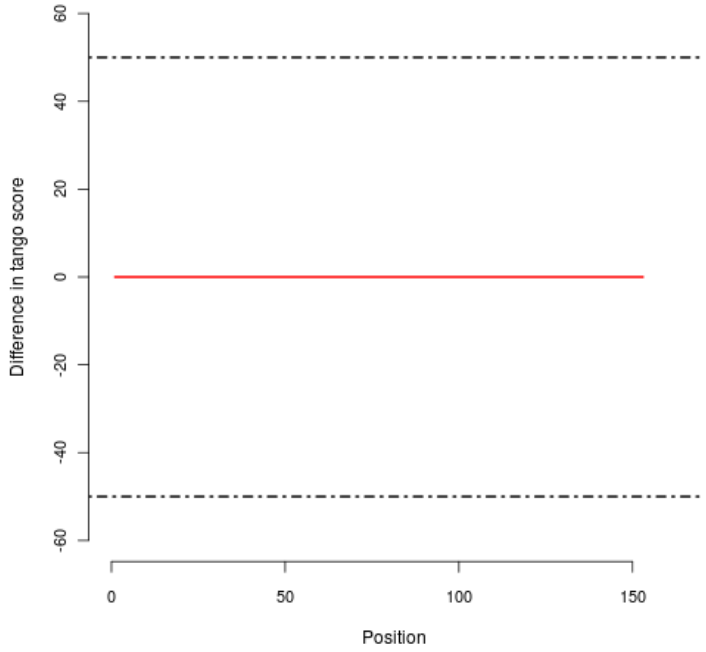


**Figure 2.3: Difference in TANGO aggregation between WT and variant.** This graph plots the per-residue TANGO aggregation score diference between WT protein and the variant. From left to right, all TANGO score differences from the N-terminus to the C-terminus are plotted. A flat line indicates that the variant does not alter the aggregation profile of the protein. Positive peaks indicate increased aggregation tendency due to this mutation. Negative peaks indicate decreased aggregation tendency due to this mutation.

We also retrieved structural information using BLAST. The obtained PDB structure was 1hl4, which shows 100.00 percent homology between the protein sequence and the used struc-

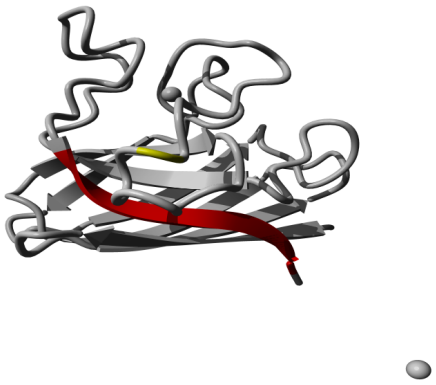

**Figure 2.4: Molecular visualisation of TANGO aggregation-prone regions**. This molecular image shows the TANGO aggregation-prone regions as red colored segments. The structural location of the variant residue is colored in yellow.

ture. The aggregating stretches predicted by TANGO are also visualized in this protein structure (Figure 2.4).

## 2.2 WALTZ prediction

WALTZ is an algorithm that accurately and specifically predicts amyloid-forming regions in protein sequences. It is thus more specific in terms of aggregate morphology than TANGO. The total WALTZ score for your protein is 511.30 and mutations can increase (dWALTZ >50), decrease (dWALTZ <-50) or not affect amyloid propensity (dWALTZ between -50 and 50). In this case, dWALTZ equals 4.38 which means that the mutation does not affect the amyloid propensity of your protein.

In figure 2.5 and 2.6 the position of the WALTZ stretches in the wild type and variant

**Figure 2.5: Bar representation of the WALTZ windows present in the wild type (top) and mutant protein (bottom).** In the bar representation, the position of the aggregating stretches is visualised in blue, and the dashed vertical line in the variant indicates the position of the variant residue.

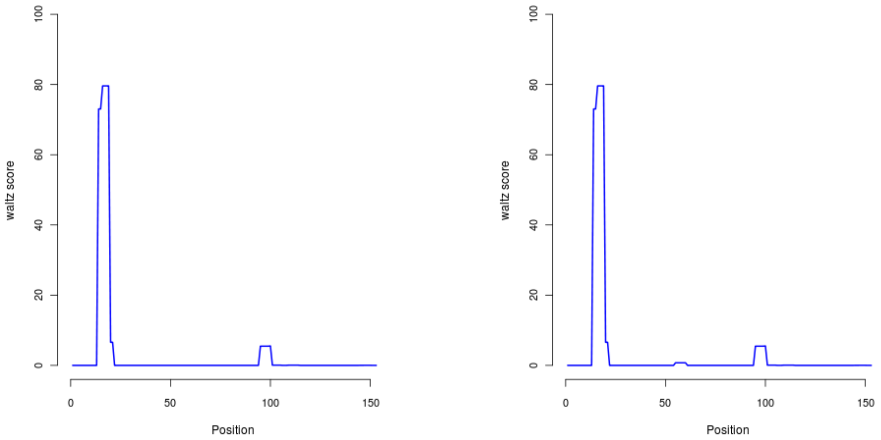


**Figure 2.6: Profile representation of the WALTZ stretches in the wild type (left) and mutant (right) protein.** This graph plots the per-residue WALTZ aggregation score of the wild type and variant protein. From left to right, all residue scores from the N-terminus to the C-terminus are plotted.

protein are visualized, represented by respectively a bar or profile representation. In table 2.2 , the short stretches are listed for both wild type and mutant. To compare the effect of the mutation to the WT, we also show a Difference profile (Figure 2.7), that plots the difference between WT protein and the variant.

We also retrieved structural information using BLAST. The obtained PDB structure was 1hl4, which shows 100.00 percent homology between the protein sequence and the used structure. The amyloid-forming regions predicted by WALTZ are also visualized in this protein structure, Figure 2.8

**Table 2.2: WALTZ regions in variant and wild type.** For each WALTZ region, the start, end, sequence and score is given.

| Number | Start | End | Stretch | Score |
|---|---|---|---|---|
| **Wild Type** | | | | |
| 1 | 13 | 21 | VQGIINFE | 59.70 |
| 2 | 94 | 100 | ADVSIE | 5.49 |
| **Mutant** | | | | |
| 1 | 13 | 21 | VQGIINFE | 59.70 |
| 2 | 94 | 100 | ADVSIE | 5.49 |

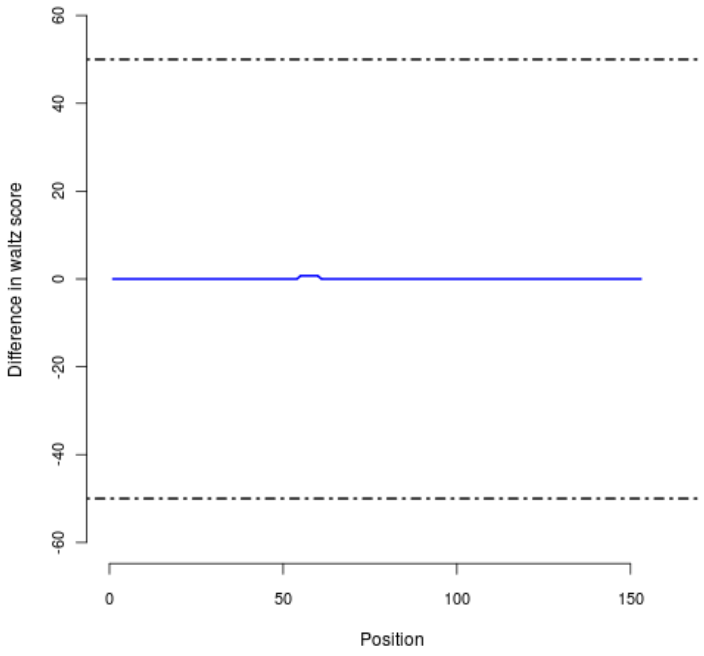


**Figure 2.7: Difference in WALTZ amyloid propensity between WT and variant.** This graph plots the per-residue WALTZ aggregation score diference between WT protein and the variant. From left to right, all WALTZ score differences from the N-terminus to the C-terminus are plotted. A flat line indicates that the variant does not alter the aggregation profile of the protein. Positive peaks indicate increased amyloid propensity due to this mutation. Negative peaks indicate decreased amyloid propensity due to this mutation.

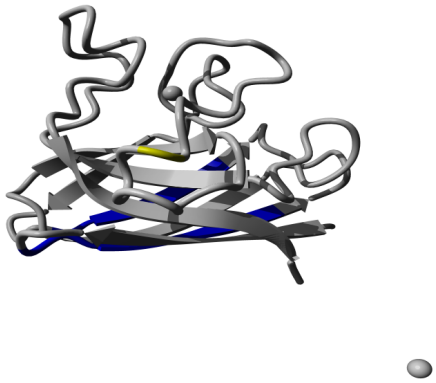

**Figure 2.8: Molecular visualisation of WALTZ amyloid-forming regions**. This molecular image swhos the WALTZ aggregation-prone regions as blue colored segments. The structural location of the variant residue is colored in yellow.

## 2.3 Limbo prediction

LIMBO is a chaperone binding site predictor for the Hsp70 chaperones, trained from peptide binding data and structural modeling. The total LIMBO score for your protein is 1599.59 and a mutation can increase (dLIMBO >50), decrease (dLIMBO <-50) or not affect chaperone binding (dLIMBO between -50 and 50). In this case, dLIMBO equals 0.00 which means that the mutation does not affect the chaperone binding tendency of your protein.
In figure 2.9 and 2.10 the position of the LIMBO stretches in the wild type and variant protein are visualized, represented by respectively a bar or profile representation. In table 2.2, the short stretches are listed for both wild type and mutant. To compare the effect of the mutation to the WT, we also show a Difference profile (Figure 2.11), that plots the difference between WT protein and the variant.

We also retrieved structural information using BLAST. The obtained PDB structure

**Figure 2.9: Bar representation of the LIMBO windows present in the wild type (top) and mutant protein (bottom).** In the bar representation, the position of the aggregating stretches is visualised in pink, and the dashed vertical line in the variant indicates the position of the variant residue.

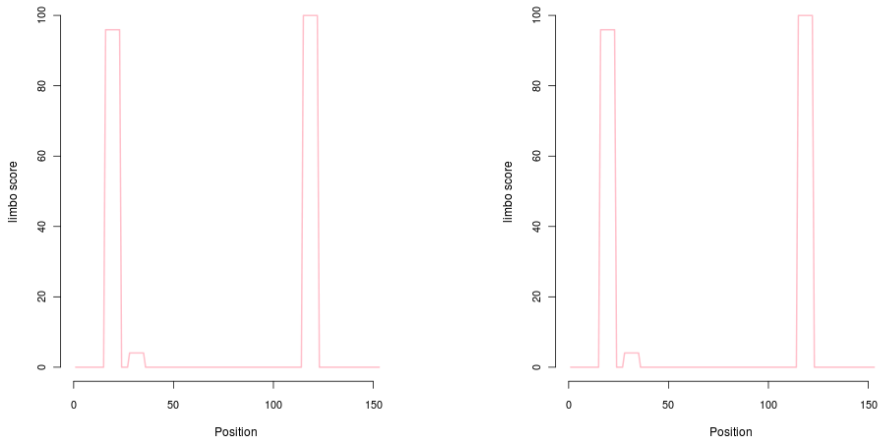


**Figure 2.10: Profile representation of the LIMBO stretches in the wild type (left) and mutant (right) protein.** This graph plots the per-residue LIMBO aggregation score of the wild type and variant protein. From left to right, all residue scores from the N-terminus to the C-terminus are plotted.

was 1hl4, which shows 100.00 percent homology between the protein sequence and the used structure. The chaperone-binding sites predicted by LIMBO are also visualized in this protein structure, Figure 2.8

**Table 2.3: LIMBO regions in variant and wild type.** For each LIMBO region, the start, end, sequence and score is given.

| Number | Start | End | Stretch | Score |
|---|---|---|---|---|
| **Wild Type** | | | | |
| 1 | 15 | 23 | GIINFEQK | 95.92 |
| 2 | 114 | 122 | RTLVVHEK | 99.98 |
| **Mutant** | | | | |
| 1 | 15 | 23 | GIINFEQK | 95.92 |
| 2 | 114 | 122 | RTLVVHEK | 99.98 |

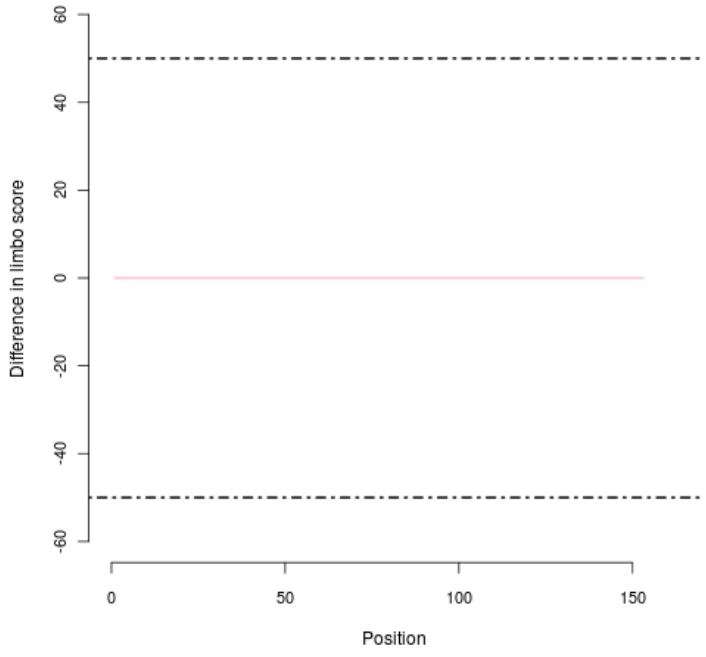


**Figure 2.11: Difference in LIMBO chaperone binding propensity between WT and variant.** This graph plots the per-residue LIMBO chaperone binding score diference between WT protein and the variant. From left to right, all LIMBO score differences from the N-terminus to the C-terminus are plotted. A flat line indicates that the variant does not affect the chaperone-binding sites of the protein. Positive peaks indicate increased chaperone binding due to this mutation. Negative peaks indicate decreased chaperone binding due to this mutation.

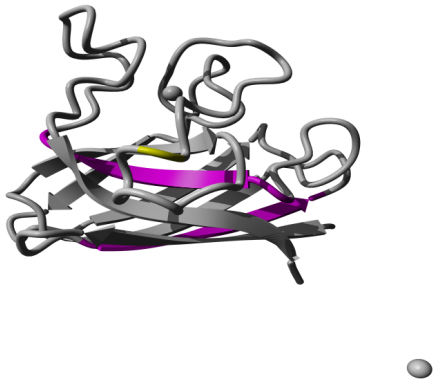

**Figure 2.12: Molecular visualisation of LIMBO chaperone-binding sites**. This molecular image swhos the LIMBO chaperone-binding sites as pink colored segments. The structural location of the variant residue is colored in yellow.

## 2.4 FoldX prediction

The empirical protein design forcefield FoldX is used to calculate the difference in free energy of the mutation: ddG (delta delta G). If the mutation destabilizes the structure, ddG is increased, whereas stabilizing mutations decrease the ddG. Since the FoldX error margin is around 0.5 kcal/mol, changes in this range are considered insignificant.
1hl4 has 100.00 percent homology with the submitted sequence.This pdb is then used to get some more information on the structural effect.The mutation from SER to ILE at position 59 results in a ddG of 4.92 kcal/mol. This implies that the mutation reduces the protein stability.

**Figure 2.13: Molecular visualization of the WT (left) and variant (right) amino acid.** The residues colored in red represents the wild type (SER) and variant residue (ILE).

## 2.5 Conclusion

Finally we can conlude that:

- Based on TANGO, the mutation does not affect the aggregation tendency of your protein.
- Based on WALTZ, the mutation does not affect the amyloid propensity of your protein.
- Based on LIMBO, the mutation does not affect the chaperone binding tendency of your protein.
- Based on FoldX, the mutation reduces the protein stability.

## 2.6 Explanation of all files in the .zip package

- WT_tango.png: TANGO aggregation profile score plot for the wild type sequence
- WT_tango_bar.png: Visual summary of the TANGO short stretches for the wild type sequence
- WT_waltz.png: WALTZ amylogenic profile score plot for the wild type sequence
- WT_waltz_bar.png: Visual summary of the WALTZ short stretches for the wild type sequence
- WT_limbo.png: LIMBO chaperone binding profile score plot for the wild type sequence
- WT_limbo_bar.png: Visual summary of the LIMBO short stretches for the wild type sequence
- MT_tango.png: TANGO aggregation profile score plot for the mutated sequence
- MT_tango_bar.png: Visual summary of the TANGO short stretches for the mutated sequence
- MT_waltz.png: WALTZ amylogenic profile score plot for the mutated sequence
- MT_waltz_bar.png: Visual summary of the WALTZ short stretches for the mutated sequence
- MT_limbo.png: LIMBO chaperone binding profile score plot for the mutated sequence
- MT_limbo_bar.png: Visual summary of the LIMBO short stretches for the mutated sequence
- MT_WT_tango_dif.png: Difference in TANGO aggregation between wild type and mutant
- MT_WT_waltz_dif.png: Difference in WALTZ amylogenicity between wild type and mutant
- MT_WT_limbo_dif.png: Difference in LIMBO chaperone binding between wild type and mutant
- MT_1hl4_tango_colored.png: Molecular visualization of TANGO aggregation-prone regions in the variant structure
- MT_1hl4_waltz_colored.png: Molecular visualization of WALTZ amylogenic regions in the variant structure
- MT_1hl4_limbo_colored.png: Molecular visualization of LIMBO chaperone binding regions in the variant structure
- WT_1hl4_colored_zoom.png: Molecular visualization of the structural environment of the wild type amino acid
- WT_1hl4.pdb: PDB of the wild type

- MT_1hl4_colored_zoom.png: Molecular visualization of the structural environment of the variant amino acid
- MT_1hl4.pdb: PDB of the variant