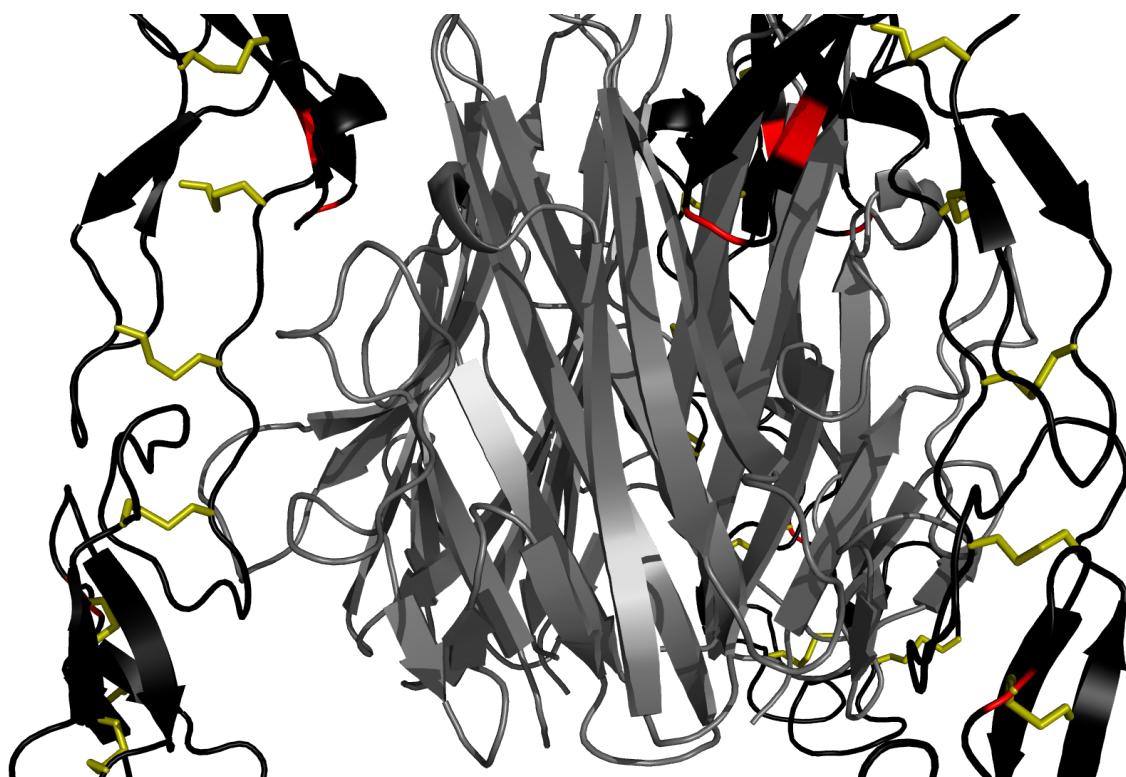


Protein structure modeling for variant pathogenicity prediction

June 13, 2019



Author: Sylt Schuurmans

Student number: 333332

Study: Bio-informatics

Hanze hogeschool: Institute for Life Science and Technology

Supervisor Hanze: Martijn Herber

Supervisor UMCG: Joeri van der Velde

Protein structure modeling for variant pathogenicity prediction
Model of TNF β wild type structures with mutations that will be mutated highlighted in red.
Universitary Medical Center Groningen: Department of Genetics (UMCG GCC)

Abstract

Around 1 in 17 people is affected by one of 7,000 known rare diseases. Most of these patients do not receive a diagnosis, which means they remain in uncertainty without a prognosis, are unable to join specific patient support groups, and do not receive the most appropriate treatment. Next-generation sequencing (NGS) of DNA promises to establish a molecular diagnosis and help these patients but many challenges still stand in the way of maximum success. Recent years have seen great advances in computational tools that quickly reduce the amount of DNA variants to be interpreted by a human expert for potentially pathogenic effects. But, the current tools that rely on features such as evolutionary conservation, annotation of regulatory genomics elements and structural DNA features have been already optimized over many years and significant improvements are not expected. Here we tried to introduce structural features of proteins into diagnostics based on the methods used by VIPUR. Through the difficulties of protein modeling and experts knowledge it was discovered that methods used by VIPUR are not features that can help in diagnosis with machine learning. Structural data of proteins is often incomplete and is highly dependent on experimentally determined structures which are expensive to make. The methods that VIPUR uses to standardize protein structures for machine learning removes the context and treats them like they are in a vacuum. To gain a more realistic view on structural information we chose to use the web service HOPE. We also developed a method to gain insight in the structural features of a protein and several of its variants. With these methods we tried to collect structural information which is usable for diagnosis.

Samenvatting

Ongeveer 1 op de 17 mensen is slachtoffer van een 7.000 bekende zeldzame ziekte. De meeste van deze patiënten krijgen geen diagnose, wat betekent dat ze geen duidelijke prognose krijgen en niet in staat zijn om zich aan te sluiten bij specifieke patiëntondersteuningsgroepen en daarmee geen geschikte behandeling krijgen. Next-generation sequencing (NGS) van DNA geeft de mogelijkheid tot moleculaire diagnose om deze patiënten te helpen, maar vele uitdagingen staan nog steeds in de weg. De afgelopen jaren zijn er grote vorderingen gemaakt in machine learning methoden die de hoeveelheid DNA-varianten verminderen die door menselijke experts worden vastgesteld als potentieel pathogeen. Maar de huidige software die afhankelijk is van eigenschappen zoals evolutionair behoud, annotatie van regulatorische genomische-elementen en structurele DNA-kenmerken zijn over vele jaren geoptimaliseerd en er worden geen significante verbeteringen meer verwacht. Hier hebben we geprobeerd structurele kenmerken van eiwitten in de diagnostiek te introduceren op basis van de methoden die door VIPUR werden gebruikt. Door de moeilijkheden van eiwitmodellering en de kennis van deskundigen hebben we ontdekt dat de door VIPUR gebruikte methoden niet bruikbaar zijn bij de diagnose met machine learning. Structurele gegevens van eiwitten zijn vaak onvolledig en zijn in hoge mate afhankelijk van experimenteel bepaalde structuren die duur zijn om te maken. De methoden die VIPUR gebruikt voor het standaardiseren van eiwitstructuren voor machine learning, verwijderen de context en behandelen eiwitten alsof ze in een vacuüm zitten. Om een meer realistisch beeld te krijgen van structurele kenmerken hebben we ervoor gekozen om de webservice HOPE te gebruiken. Ook is er een methode ontwikkeld om inzicht te krijgen in de structurele kenmerken van een eiwit en zijn varianten. Met deze method hebben we geprobeerd om structurele informatie bruikbaar te maken voor diagnose.

Acknowledgment

This graduation project has taken place at Universitary Medical Center Groningen (UMCG) at the Department of Genetics within the genomics coordination center (GCC), this report has been written for the GCC to gain insight into structural data to improve the GAVIN variant predictor. I want to thank Joeri van der Velde for supervising me during this project and helping me with writing this report. I also would like to thank Benjamin Kant and Marielle van Gijn for helping me decide to focus on assessing tumor necrosis factor associated receptor-associated periodic syndrome (TRAPS). And I especially would like to thank Tsjerk Wassenaar for informing us about the function and structure of TNFRSF1A, giving us the appropriate protein structures to work with and steering this project into a meaningful direction.

Abbreviations

3D Three Dimensional
ACCP Solvent Accessible Surface Area
API Application Programming Interface
Bash Bourne Again Shell
CPU Central Processing Unit
CSV Comma Separated Values
DNA Deoxyribonucleic Acid
FADD Fas Associated Death Domain protein
FEM Fixed End Move
FHF Familial Hibernian Fever
GAVIN Gene-Aware Variant INterpretation
GCC Genomic Coordination Center
GRCh/hg Genome Reference Consortium Human Human Genome
HOPE Have yOur Protein Explained
LOMETS Local Meta-threading Server
MD Molecular Dynamics
MPI Message Parsing Interface
NCBI National Center for Biotechnology Information
NF- κ B Nuclear Factor kappa-light-chain-enhancer of activated B cells
OpenGL Open Graphics Library
OS Operating System
OSF Open Science Framework
PDB Protein Data Bank
PM Pivot Movement
PSI-BLAST Position Specific Iterative BLAST
PSSM Position Specific Scoring Matrix
RCSB Research Collaboratory for Structural Bioinformatics
REU Rosetta Energy Unit
RNA Ribonucleic acid
RMSD Root Mean Square Deviation
SASA Solvent Accessible Surface Area
SCOP The Structural Classification of Proteins
SLURM Simple Linux Utility for Resource Management
SODD Silencer of Death Domain
SPVAA Simple Protein Variant Analysis Approach
TNF Tumor Necrosis Factor
TNF α Tumor Necrosis Factor Alpha
TNF α Tumor Necrosis Factor Alpha
TNF β Tumor Necrosis Factor Beta
TNFB Tumor Necrosis Factor Beta
TNFR1 Tumor Necrosis Factor Receptor Superfamily Member 1A TNFRSF1A Tumor Necrosis Factor Receptor Superfamily Member 1A
TRADD Tumor Necrosis Factor Receptor type 1-Associated DEATH Domain protein
TRAPS Tumor necrosis factor associated Receptor-Associated Periodic Syndrome
VIPUR Variant Interpretation Using Rosetta
VTS VIPUR Training Set

Contents

1	Introduction	1
2	Theory	2
2.1	Mutations and its effects in the central dogma of molecular biology	2
2.2	A general concept of structural levels within proteins and the effect of mutations	2
2.3	Addition of structural data to diagnosis and treatment in healthcare	2
2.4	Protein modeling techniques	3
2.5	A theoretical large scale implementation of structural protein variant assessment	3
2.6	Monte Carlo method	4
2.7	The use of the Monte Carlo method and its pitfalls	4
2.8	Tumor Necrosis Factor Receptor Associated Syndrome	5
2.9	Tumor Necrosis Factor Receptor Super Family Member 1A	5
2.10	Tumor Necrosis Factor Alpha and Beta	5
3	Materials and methods	6
3.1	VIPUR approach	6
3.2	VIPUR specific tools	6
3.2.1	PSI-BLAST	6
3.2.2	Probe	6
3.3	Rosetta	6
3.3.1	Relax	7
3.3.2	DDG Monomer	7
3.3.3	Rescore	7
3.3.4	Backrub	7
3.4	Structure prediction web services	7
3.4.1	Robetta prediction server	7
3.4.2	I-TASSER prediction server	8
3.4.3	HOPE	8
3.5	Structural modification and visualization software	8
3.5.1	Modeller	8
3.5.2	PyMOL	8
3.5.3	PyRosetta	8
3.6	Amino acid sequence variant tables	9
3.6.1	GAVIN Machine Learning Data Table	9
3.6.2	gnomAD	9
3.6.3	Infevers	9
3.7	Protein functional and structural databases	9
3.7.1	Research Collaboratory for Structural Bioinformatics	9
3.7.2	Uniprot	9
3.8	Scripting languages	9
3.8.1	Python	9
3.8.2	R scripting language	10
3.9	Utility software	10
3.9.1	Bash	10
3.9.2	SLURM	10
3.9.3	MPI	10

4 Results	11
4.1 Reviving the VIPUR approach to expand rare disease diagnostics	11
4.1.1 Preparatory steps for using the VIPUR approach	11
4.1.2 Resolving VIPUR system incompatibilities	11
4.1.3 Expanding the VIPUR training set with data from TNFRSF1A by homology modeling and protein threading	12
4.1.4 Practical VIPUR usage	12
4.2 Analyses of proteins variants TNFRSF1A	13
4.2.1 Requirements for determining structural and binding effects of protein variants . .	13
4.2.2 Introduction of the simple protein variant analysis approach	13
4.2.3 Carrying out SPVAA on TNFRSF1A	14
4.3 Finding mutation information with HOPE	20
5 Discussion	21
5.1 VIPUR	21
5.2 SPVAA	22
5.3 HOPE	22
6 Conclusion	23
7 Future work	23

List of Figures

1	Flowcharts VIPUR pipeline and altered VIPUR pipeline	11
2	I-TASSER and Robetta models with and without templates	12
3	Flowchart SPVAA pipeline	14
4	TNFRSF1A homotrimer with TNF α homotrimer relax density plots	16
5	TNFRSF1A homotrimer with TNF β homotrimer relax density plots	17
6	TNFRSF1A homotrimer with TNF α homo trimers wild type and mutated relaxed models	18
7	TNFRSF1A homotrimer with TNF β homo trimers wild type and mutated relaxed models	19

List of Tables

1	Sample from the combined observed TNFRSF1A mutations table	14
2	Sample of the TNFRSF1A PDB residue mutation table	15

1 Introduction

Around 1 in 17 people is affected by one of 7,000 known rare diseases. Most of these patients do not receive a diagnosis, which means they remain in uncertainty without a prognosis, are unable to join specific patient support groups, and do not receive the most appropriate treatment. Next-generation sequencing (NGS) of DNA promises to establish a molecular diagnosis and help these patients but many challenges still stand in the way of maximum success. Recent years have seen great advances in computational tools that quickly reduce the amount of DNA variants to be interpreted by a human expert for potentially pathogenic effects [1]. Although algorithms can now safely remove around 95% of the harmless variants, this still leaves hundreds of variants to be investigated for a whole-exome sequenced patient, which is far too many for a quick and clear diagnosis. Current tools to predict variant pathogenicity rely on features such as evolutionary conservation, annotation of regulatory genomics elements or structural DNA features. These tools have already been optimized over many years and further significant improvements are not expected. Therefore there is still a great need for even more powerful variant prioritization tools. A refreshing alternative was presented by variant interpretation using Rosetta (VIPUR) [2] which shows the potential of structural modeling of proteins to predict the actual effect of a specific variant on the function of a protein. However the publication did not result in a piece of high quality software that is usable for routine diagnostics. Therefore we want to test if the structural information used by VIPUR is useful to diagnostics with the difficult to assess gene tumor necrosis factor receptor 1 alpha (TNFRSF1a) to see if the prediction has any meaning. To test this we use two validated mutations of TNFRSF1A and one provisional mutation in combination with the tool have your protein explained (HOPE) [3] and a self developed method Simple Protein Variant Analyses Approach (SPVAA). Through testing these methods we will explore the potential pitfalls of protein modeling and discover if it is possible to add structural information to routine diagnostics.

2 Variant prediction in genome diagnostics and the addition of protein modeling

2.1 Mutations and its effects in the central dogma of molecular biology

Within the human genome mutations occur continuously by internal and external factors that: insert, remove, substitute or alter the reading frame in a nucleotide sequence. Mutations are not without consequences and can be protective [4], benign or harmful by altering the deoxyribonucleic acid (DNA) order. From a sequence of DNA genes are transcribed into ribonucleic acid (RNA) which can work as machinery or translates into an amino acid sequence to form a protein. Mutations outside a gene could lead to lowered or heightened transcription of a protein, when a mutation resides inside a gene it could lead to proteins that are unstable during or after formation, perform less optimal or are not functional [5–7].

2.2 A general concept of structural levels within proteins and the effect of mutations

The formation of protein structures is classified in different levels, distinctions are made based on bindings and structures that arise with the interaction of bonds. The order in which amino acids appear in a sequence is called the primary structure, in this level amino acids are only bound to each other by peptide bonds. Within a primary structure amino acids can form new peptide bonds between the N-terminus and C-terminus of an amino acid, with these bonds 3D structures are made called α -helices and β -sheets that together make up the secondary structure. The tertiary structure gives further rise to the 3D shape of a polypeptide by making disulfide bridges, ion and hydrogen -bonds, hydrophobic and hydrophilic -interactions between amino acids By combining multiple tertiary structures the quaternary structure of a protein can be formed out of the mentioned bonds, bridges and interactions [8, 9].

Mutations within proteins can have different effects to protein structures, often single missense mutations often have minimal effect on the backbone of a protein [10, 11] but can result in destabilization of the structure when assembled or can disrupt the active site. Frameshift mutations on the other hand can cause large differences in the primary structure and have therefore a higher chance of an altered sequence that leads to deformation or stop codon introduction [12].

2.3 Addition of structural data to diagnosis and treatment in healthcare

Acquiring information about DNA sequences depends on sequencing, which became cheaper over the years [13], and found its use in diagnosing patients within the healthcare sector [1]. From the collected data by genome sequencing experiments most of the analysis is handled in-silico due to the quantities of data that are produced [1]. Proteins often find their use in diagnosing diseases experimentally [9, 14], however in-silico it is often limited to information about conservation in the amino acid sequence [15]. Yet, the 3D shape of proteins defines their function [16] and by assessing structures it can become possible to determine changes in function that are caused by mutations that might not be discoverable through conservation and are therefore unclassifiable. Another advantage of structural information is that it becomes possible to develop treatment with diagnostic information for diseases that are caused by mutations [17]. With experimental methods such as X-ray crystallography and nuclear magnetic resonance (NMR) more than 158000 structures [18] have been completely revealed, however it is only a tiny fraction of the potential possible proteins [19] (especially without the inclusion of all folds). Making 3D structures is currently not common for diagnosis because it is relative expensive and is difficult to perform, some structures contain flexible regions which makes it hard to determine the exact position of some atoms and can cause information loss about the structure [20, 21].

2.4 Protein modeling techniques

An alternative approach for determining structures, compared to experimental methods, is protein modeling wherein structures are generated computationally. One of the benefits from making a structure in such a way is that laws of physics do not hinder the collection of structural information, but the lack of physics is also its weakness since the structures tend to be less accurate representations of proteins.

With the method homology modeling the amino acid sequences of proteins are aligned to sequences of experimental determined structures, based on these alignments a template is formed whereon structural fragments are built. It is not recommended to use this strategy if the sequence identity is less than 20% since there might not be any structural relation [11]. Another approach is protein threading which relies on the observation of folds in previously determined experimental structures, with the occurrence of specific folds a probability is predicted that determines how a protein folds in a specific manner.

Strategies are continuously being developed and improved to determine unknown structures, but all have partially similar guidelines wherein the avoidance of steric hindrance [22] and low energies, determined by scoring systems [23], are important. From the computer generated models many are less accurate than those experimentally determined structures, yet computational models could potentially gain the upper hand in solving membrane proteins[24].

2.5 A theoretical large scale implementation of structural protein variant assessment

With the wide spectrum of potential different proteins it can be difficult to produce any form of universal protein assessment tool, that is able to determine if a mutation is harmful or not based on structural information. However a first step to solve such a complex problem would be by determining the correct approach. In this case it is assumed that a machine learning would be the best method for detecting patterns in structures and classifying the effect of structural changes, because it has the ability to learn from structural mutations currently available and develop new insights [25].

Since the problem is so complex it should be divided into smaller more feasible problems, beginning by separating the different protein classes, which for example can be done according to The Structural Classification of Proteins database (SCOP) [26]. A first discrimination between the proteins could be made based on protein type/fold class (membrane, globular, fibrous and disordered -proteins) because these differences already predetermine some functions and locations for certain proteins in a cell [27–30]. After formation of these classes each should have its own machine learning method applied so their features can be analyzed within context of where and how they function. The next set of discriminators is highly dependent on the variations in classes, but all have features in the end describing bonds, interactions and movement of complexes in protein structures. When for each of the main classes methods have been developed a meta classifier determines which method should be applied to determine the effect of a mutation.

2.6 Monte Carlo method

There are complex problems in a variety of research fields which could take up years or even centuries to compute with simple deterministic methods. For some problems there is an algorithmic solution which makes it possible to cut down computation time significantly, but when no deterministic algorithm is available to speed up the process an empirical probabilistic method might be able to approximate the desired results.

With the Monte Carlo method random samples are taken from the parameter space ,that describe a data set, and fed into a model which produces a potential outcome. By repeating the process more results are generated until at some point the data can display a pattern that describes the outcome. The result is a quantified probability which describes the chance that something might occur based on the quantity of occurrence generated by the model [31–33].

The Monte Carlo methods can differ depending on the algorithm and application in which it is used, but in summary most implementations will follow a general pattern [32].

0. Construct a model which is able to describe an outcome of the problem.
1. Define the space of which inputs can be used by the model to get an outcome (creating a parameter space).
2. Use the model to generate results based on random sampled input from the parameter space.
3. Order and determine which results are part of a certain outcome and draw conclusions on the generated statistical evidence.

2.7 The use of the Monte Carlo method and its pitfalls

The Monte Carlo method is widely used within various applications in different fields of science but is suitable for problems. Where all inputs are known but it is too inefficient to compute the problem with a deterministic method. Problems that require uncertainty to be incorporated into the analysis or prediction and is suitable for discovering new parameters in a model to improve the current model. Monte Carlo is not good at solving problems where answers are known but the inputs are not and in solving problems where no uncertainty should be in the answer produced.

All mentioned problems it can solve all tend to rely on significant quantities of data which makes it a relative time consuming process for generating results. Meaning of the generated result is highly depended on the model and random sampling technique that is used and both contribute to biases in the results [31, 32, 34].

2.8 Tumor Necrosis Factor Receptor Associated Syndrome

Tumor necrosis factor receptor-associated periodic syndrome (TRAPS) is a rare (1 : 1,000,000) hereditary autosomal dominant disease , formerly known as Familial Hibernian Fever (FHF) [35], that can cause recurring fevers with durations from days up to months. Symptoms during these fevers are: skin rash, swelling, inflammatory reactions across the whole body and pain in the abdomen, muscles and/or joints. A long term and lasting effect is the accumulation of amyloids within the kidneys and may result in other diseases [7]. TRAPS is known to be caused by mutations within the gene tumor necrosis factor receptor 1 (TNFRSF1A/TNRF1) (Section 2.9), of which few mutations are known to hinder the transport of proteins to the cell surface and trap them in cell causing the activation of inflammatory response [7, 36]. So far 158 mutations have been associated with the disease [37], but more mutations have been identified in TNFRSF1A wherein some might be pathogenic (Sections 3.6.1, 3.6.2).

2.9 Tumor Necrosis Factor Receptor Super Family Member 1A

Tumor Necrosis Factor Receptor Super Family Member 1A (TNFRSF1A/TNFR1) is a gene located on chromosome 12 region 1 band 3 and sub-band 31. The gene produces a trans-membrane receptor consisting of 445 residues divided into 221 residue cytoplasmic section and a 171 extracellular part that consists of 4 conserved cysteine rich domains [38–40]. The receptor is ubiquitous across most cell surfaces , but is not found on erythrocytes [41], and can form two different types of unbound hexagonal clusters depending on the dimer formation [42]. When the structures are dimers the binding sites are exposed making it possible for tumor necrosis factor (TNF) α and β (Section 2.10) to bind in homotrimeric form, by binding of TNF the dimers disconnect and three TNFR1s interact with the TNF homotrimer [42]. With the interaction of TNF homotrimers with TNFR1 it can activate several pathways such as the nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B), which enhances the transcription of various genes during inflammation, infection or other forms of external stress. Also it is able to activate the extrinsic pathway of apoptosis by releasing the silencer of death domain (SODD) proteins, on the cytoplasmic site of the cell, which will activate the formation of Tumor Necrosis Factor Receptor type 1-Associated DEATH Domain protein (TRADD) [43] that forms a complex that attracts Fas associated death domain (FADD) [44]. On binding of FADD initiator caspase 8 starts a cascade wherein caspase 3 is activated and will cleave aspartate out of proteins and thereby disrupting the metabolism[45–47].

2.10 Tumor Necrosis Factor Alpha and Beta

The proteins TNF α and β are both pro-inflammatory cytokines that are produced as response to an infection or when a cell is damaged. Both are transcribed from their genes that reside in chromosome 6 in the p-arm at region 2 band 1 and sub-band 3. TNF α and β are 35% identical and 50% homologous to each other consisting out of 233 and 205 amino acid residues. Both are able to form a homotrimeric structures that can bind to the dimeric form TNFR1 (Section 2.9) to activate the extrinsic pathway[39, 48–50].

3 Materials and methods

3.1 VIPUR approach

Variant interpretation using Rosetta (VIPUR) is a machine learning approach for predicting deleteriousness of proteins (loss of function) and uses sequential and structural information. To train it the VIPUR training set (VTS) was made which contained sequential and structural features from protein structures that were acquired from Modbase [51] and SWISS-MODEL [52–56]. Proteins that did not have a structure in modbase or SWISS-MODEL were modeled with Modeller (Section 3.5.1) from proteins that had the highest amino acid sequence identity to the desired protein. based on protein fragments that had the highest amino acid sequence identity to the protein Some structures from the databases had: duplicate chains, ligands, metals and non-standard amino acids which were removed to avoid inconsistencies for the generated features by tools. For all proteins a mutation file was made that described which and how many residues had to be mutated, with this file DDG monomer (Section 3.3.2) could determine changes in the protein structure, most results by ddg were used as features. Structural mutations of proteins that are in the VTS were introduced by a script using PyMOL (Section 3.5.2) by default or PyRosetta (Section 3.5.3) when PyMOL was not available. After a mutation the structure was optimized by the relax application (Section 3.3.1) that produced 50 structures of a single variant, for each of these structures property scores where available of which all quartiles have been used as a machine learning feature. of which the quartiles are used as a learning feature. Probe (Section 3.2.2) calculated the solvent accessible surface area (SASA/ACCP) of a protein in square Ångstrom (\AA^2) which was used as a structural machine learning feature. The sequence features of VIPUR were produced by PSI-blast (Section 3.2.1) on non mutated sequences and blasted against the NCBI protein database (nr) which resulted in a position specific scoring matrix (PSSM). From the PSSM scores of the information content, non-mutated, mutated, the difference in scores between non-mutated and mutated and the difference between groups [2, 57] used as sequential feature for VIPUR. With 106 features generated by the mentioned tools deleteriousness of a protein variant is determined with sparse logistic regression. The term sparse implies that a limited set of features was used because the weights "shrink" to 0 with regularization [58].

3.2 VIPUR specific tools

3.2.1 PSI-BLAST

Position specific iterative basic local alignment search tool (PSI-BLAST) focuses on distant relatives of proteins by making a profile of the sequence and querying it at a protein sequence database. With the generated results a new profile is constructed which is queried to the same database as the previous query. These steps are repeated several times to determine which residues are found in distant relatives and results is a position specific scoring matrix (PSSM) that describes the frequency of which residues are substituted by a specific other residue [59–61]. From the PSSMs sequences features were acquired for the VIPUR machine learning method.

Position-Specific Iterated BLAST 2.7.1+

3.2.2 Probe

Probe is able to evaluate atom packing for a single protein or for interacting proteins. It does this by creating a probe, which is described as a sphere like object, that marks an area with dots when at least two non-covalent atoms are in contact with the probe. [62, 63]. VIPUR used this tool to calculate solvent accessible surface area (SASA or ACCP).

version 2.16.130520

3.3 Rosetta

Rosetta is a software suite that has a variety of tools that are developed to aid in macro molecular and antibody, analysis, design and prediction [64]. However no tools in the suite have been encountered

that could introduce missense mutations in the proteins and was done by other software (Sections 3.5.2, 3.5.3, 3.5.1). With the introduced mutations water had to be removed because some tools cannot predict structures well with: water, metals and amino acids that are no part of the standard (20) amino acids [65].

Within the tools from Rosetta various scores are assigned to different properties related to bonds, interactions, energies and geometries within structures and are written into a score file. From all different scoring metrics the Rosetta score, Rosetta energy unit (REU) or total_score in the score files, can be used to compare models from the same protein with the same tool. Not only is the score based on energy but also it has statistical terms which influence the score based on known favorable folds from existing structures that reside in the curated Rosetta database [23]. In summary to asses models, a lower Rosetta score makes a more natural model.

Rosetta software suite Version 3.10

3.3.1 Relax

The Relax application was used by VIPUR and by SPVAA to relax the side chains to minimize energy levels within the local conformational search space [66] of the structure. It determines the energy levels with a Monte Carlo method (Section 2.7) and after a certain set of moves it produces a structure and starts anew [67, 68].

3.3.2 DDG Monomer

DDG monomer is meant to predict energetic stability of a point mutation in monomeric protein. The application was used by VIPUR to collect features related to energies bonds, bridges and constraints differences between the wild type and a mutated protein. To execute the tool a script had to be ran which renumbered the wild type pdb file and it required a "mutation file" that described the changes of a residues based on name and position[69].

3.3.3 Rescore

With this tool Rosetta scores can be calculated based on PDB files proteins structures [70] , the output is identical to that is written within the score files produced by Relax (Section 3.3.1).

3.3.4 Backrub

The backrub application is based on the Monte Carlo method (Section 2.7), and alters a protein by moving its backbone residues with a strategy called fix end move (FEM). With this strategy, groups of residues are selected at random from the structure that can contain up to: four dihedral, two bond angles and two end points. Both ends of a group are fixated at their position in which a new angle α arises, within this angle residues are pivoted in their natural occurring maximum range of $\pm 10^\circ$ [71, 72]. With this application backbones of newly introduced mutations were altered, for each attempt a new file was generated and the scores were written to a score file. The lowest Rosetta scoring model was selected to undergo side chain relaxation with the Relax application (Section 3.3.1).

3.4 Structure prediction web services

3.4.1 Robetta prediction server

The web tool Robetta integrates several tools to form protein structures with homology modeling (Section 2.4). Its only requirement is an amino acid sequence, but optionally constrains and fragments can be added to disallow movement of certain structures or add known fragments to avoid calculating pieces that are already known. With this information Robetta searches with the help of sequence aligners for known fragments and tries to incorporate them into a single protein structure [73–77]. The known fragments of TNFRSF1A (Section 2.4) were given as a template to Robetta and modeled into a whole

protein to make it possible to introduce mutations and predict pathogenicity of variants.

<http://new.rosettaweb.org/>

3.4.2 I-TASSER prediction server

The I-TASSER web server is a tool that is able to predict protein structures with a FASTA sequence. The first step it takes is finding structural templates which resemble the sequence by local meta-threading server (LOMETS). LOMETS starts with multiple sequence alignment of which several sequences will undergo protein threading by different programs to form structural templates. The templates are assessed made from: the highest alignment Z-score, a program within LOMETS specific confidence score and sequence identity [78, 79]. The known fragments of TNFRSF1A (Section 2.4) were given as a template to I-TASSER and modeled into a whole protein to make it possible to introduce mutations and predict pathogenicity of variants.

Server version, <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>

3.4.3 HOPE

Have yOur Protein Explained (HOPE) is a web service that collects information of about a user specified missense mutation in a protein and comes from various sources. Uniprot (Section 3.7.2) is queried with BLAST to find homologous sequences and structures, other features that are found on Uniprot are active sites, domains and various other sequence features that help to identify the function of a region. From the BLAST results homology models are made with Yasara that are sent off to WHAT IF web services that calculate structural information about the protein. Before the formation of a report all information is put into a decision tree to assess mutational effects in context of: contacts, structural locations, non-structural features, previous variant information and amino acid properties. [3, 80–82]. With this method it is not possible to assess ligands and complexes at once but only a single missense mutation within a protein.
Version 1.1.1, <https://www3.cmbi.umcn.nl/hope/>

3.5 Structural modification and visualization software

3.5.1 Modeller

The Modeller software that is developed for homology modeling but it was used for its utilities. Which allowed to complete protein data bank (PDB) structures with missing atoms, predict disulfide bonds that were missing and mutate protein residues [83–86].

Version 9.21

3.5.2 PyMOL

Visualization of 3D structures, making images of proteins, putting monomers in the correct position, replacing TNF β structure with a TNF α in the bound structure and aligning the structures to measure the distance between models and X-ray structures were done with PyMOL [87]. PyMOL was in VIPUR used in combination with Python (Section 3.8.1) to perform mutagenesis on the protein structures.

Version 2.2.3

3.5.3 PyRosetta

Is an application programming (API) which has Python bindings (Section 3.8.1) for the Rosetta software suite (Section 3.3) and finds its use in VIPUR when no PyMOL (Section 3.5.2) was available to mutate residues in a structure [88].

Version 4

3.6 Amino acid sequence variant tables

3.6.1 GAVIN Machine Learning Data Table

Is a collection of nucleotide mutations from rare diseases used by the GAVIN [1] machine learning approach. From this set the genes of TNFRSF1A (Section 2.10) with a missense mutation were filtered (Section 3.8.2) and written into a format which the variant effect predictor could (VEP) [89] could read and translate from nucleotide to protein mutations. The classification of the variants was done by experts based on the five tier IARC classification system [90].

3.6.2 gnomAD

The gnomAD database consists of unified data from large scale genome sequencing data projects and is based on genome reference consortium human genome build 37 human genome 19 (GRCh37/hg19). From this database missense mutations were collected for TNFRSF1A (Section 2.9), no classification was known from these mutations [91].

3.6.3 Infevers

Infevers is a website about auto hereditary inflammatory diseases for which each are tables that contain information about mutations and their classification. The table for TRAPS (Section 2.9) was used to collect missense mutations of TNFRSF1A gene [92].

3.7 Protein functional and structural databases

3.7.1 Research Collaboratory for Structural Bioinformatics

Research Collaboratory for Structural Bioinformatics (RCSB) is a database where whole or fragmented experimentally determined proteins structures, that are published, can be found and downloaded. The Fragments for modeling (Sections 3.4.1, 3.4.2) whole TNFRSF1A (Section 2.9) (1EXT [93]) and determining the differences in energy levels (Section 3.3.1) with TNF α - β (1TNR [40]) with the interaction site were acquired from this database [94].

3.7.2 Uniprot

Knowledge from various omic domains about proteins have been linked together into single database called Uniprot which makes all information accessible at once. For TNFRSF1A (Section 2.10) the FASTA sequences were collected from Uniprot and for structures it redirected to (Section 3.7.1) [95].

3.8 Scripting languages

3.8.1 Python

Both VIPUR and the single protein variant analysis approach (SPVAA) were written in Python. SPVAA was written in Python because of its ease of use and the modeller bindings (Section 3.5.1) that were available. The mutations that were put together from the different tables (Sections 3.6.2, 3.6.3, 3.6.1) with R (Section 3.8.2) were filtered by a Python script. From the mutations a compact list was made by a different script that described the chains that had to be altered by modeller to introduce the appropriate mutation into a PDB file. From these PDB files several were selected to be optimized by pipeline that used backrub (Section 3.3.4) and Relax (Section 3.3.1) to optimize the structure.

Laptop version 2.7.15

Server version 2.7.11

3.8.2 R scripting language

With R the tables from gnomAD, GAVIN and Infevers (Sections 3.6.2, 3.6.1, 3.6.3) of TNFRSF1A missense mutations (Section 2.9) were merged together in a new comma seperated values file with their known classifications. Ordering and filtering the double mutations and removing double classifications where done with Python (Section 3.8.1). It has also been used in combination ggplot2 [96] and data.table [97] to make density plots of all scores acquired from Rosetta Backrub and Relax (See the supplementary for R package versions).

R scripting front-end version 3.5.2 (2018-12-20)

3.9 Utility software

3.9.1 Bash

Unix like operating systems (OS) have a shell which allows users to interact with programs on a computer or with the computer itself based on commands submitted. The default shell for MacOS and also for several Linux distributions is the Bourne again shell (Bash) which was used to launch Python scripts (Section 3.8.1) and submit jobs to the SLURM workload manager (Section 3.9.2).

Laptop Version GNU bash, version 3.2.57(1)-release (x86_64-apple-darwin18)

Server Version GNU bash, version 4.1.2(2)-release (x86_64-redhat-linux-gnu)

3.9.2 SLURM

For computational jobs where a laptop or desktop does not suffice because due to the lack computational resources a computer cluster could come to aid. These clusters consist out of several computers that execute resource intensive tasks, to manage these systems as optimal and fair as possible a workload manager like simple Linux utility resource management (SLURM), is installed. Jobs are submitted that request resources for execution and are scheduled on the systems queue.

3.9.3 MPI

Some tools from the Rosetta software suite (Sections 3.3) have the ability to use multiple central processing unit (CPU) cores from a single computer or from multiple computers. With a message parsing interface (MPI) it is possible for software to communicate between CPU cores on the same and on different computers to exchange information about processes giving the ability to share work between computers and CPUs.

OpenMPI/1.8.8-GNU-4.9.3-2.25

4 Results

Determining the effects of protein variants is not a trivial task and therefore VIPUR, SPVAA and HOPE have described and employed within this order.

The SPVAA scripts can be found at: <https://github.com/Sylt-CSI/variant-protein-prediction>
The VIPUR scripts with applications and data can be found at: <https://osf.io/bd2h4/>

4.1 Reviving the VIPUR approach to expand rare disease diagnostics

4.1.1 Preparatory steps for using the VIPUR approach

After the publication of VIPUR the tools, data and applications became available at the open science framework (OSF) [98] which were downloaded and studied. All applications from the Rosetta software suite available through OSF (Section 3.3) were pre-compiled without support for MPI (Section 3.9.3) and with that could not benefit from multiple CPUs. To solve this a new version of the Rosetta software suite was built with MPI support in a slurm job where the compilation could benefit from multiple CPU cores.

4.1.2 Resolving VIPUR system incompatibilities

Within the VIPUR pipeline residues were mutated to determine the effects of a structural mutation, by default missense mutations were introduced with PyMOL (Section 3.5.2), an alternative from VIPUR was Pyrosetta (Section 3.5.3) when PyMOL was not available. Neither of these programs could be built or compiled because the lack of open graphics library (OpenGL) for PyMOL and having the incorrect C++ and C libraries for PyRosetta. To bypass both programs modeller was introduced which is also capable of replacing residues(Secton 3.5.1).

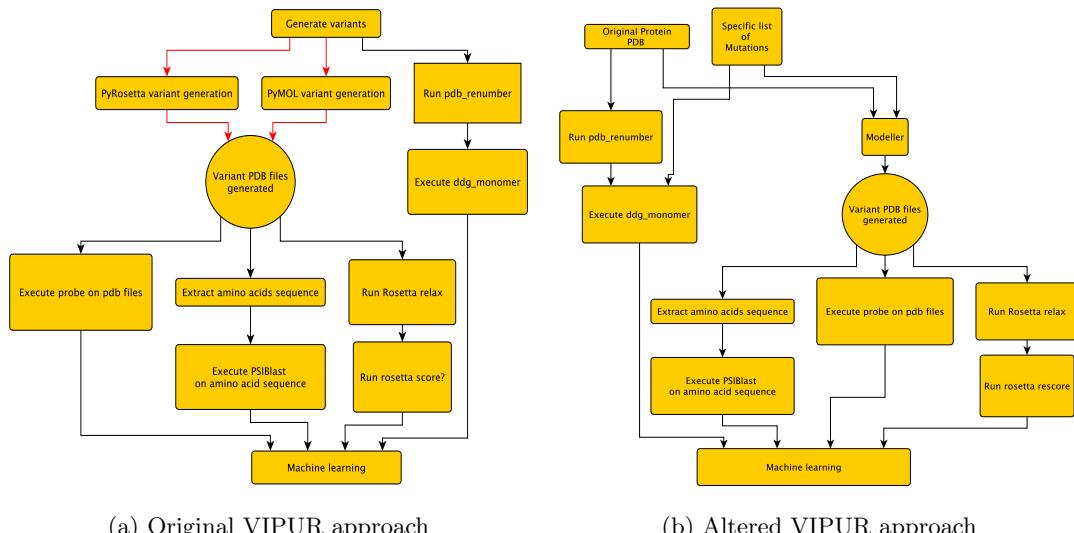


Figure 1: Both flowcharts illustrate the VIPUR pipeline wherein each block is a procedure and each arrows is the path to the up following step. The central circle was the problem that had to be fixed by a different program than PyMOL or API than PyRosetta. Figure 1a has red arrows that indicate that both methods were incapable to produce the mutated PDB files. Within figure 1b the alternative method is proposed wherein PyMOL and PyRosetta (Sections 3.5.2, 3.5.3) are substituted by Modeller (Section 3.5.1) to acquire the mutated protein structures. (To zoom in on the details within the figure it is recommend to look at the PDF version.)

4.1.3 Expanding the VIPUR training set with data from TNFRSF1A by homology modeling and protein threading

Since the VTS did not have any features of TNFRSF1A (Section 2.9) the amino acid sequence was collected from Uniprot (Section 3.7.2) and the structure from RCSB (Section 3.7.1). The structures available of TNFRSF1A were incomplete, fragments for the binding site [40] were available and its death domain that interacts with TRADD [99] (Section 2.9). To acquire a monomeric structure of TNFRSF1A two ab initio modeling web services I-TASSER and Robetta (Sections 3.4.2, 3.4.1) had been employed. Both were given the task to model the whole protein with and without a binding site template to determine how well they could model the structure. Determination of the best model was based on the lowest root mean square deviation distance (RMSD)in Å between produces models and X-ray crystallographic structures of the TNFRSF1A binding site.

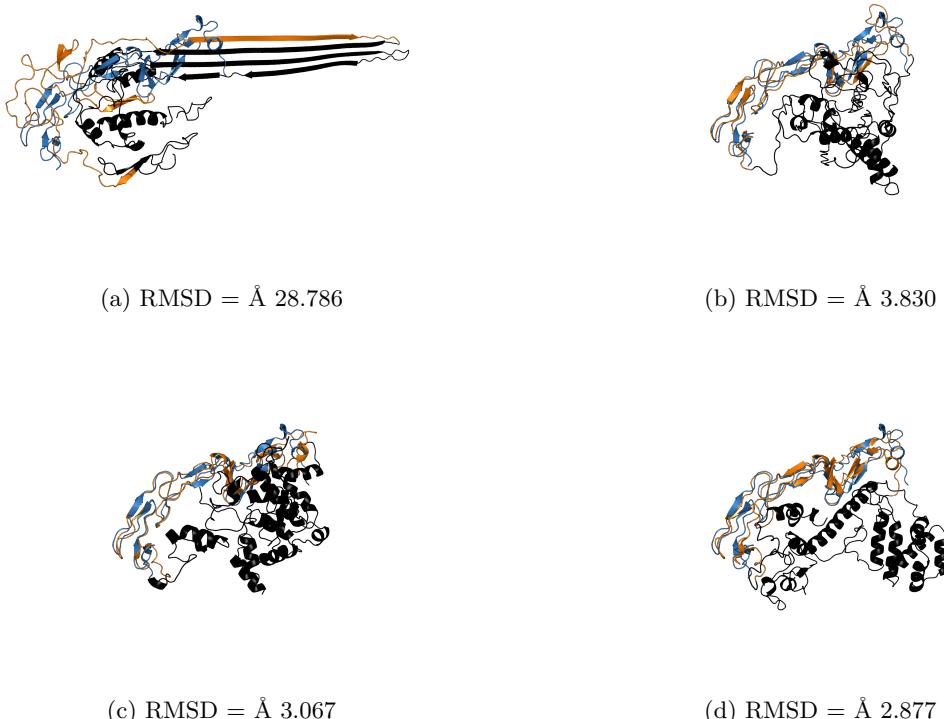


Figure 2: 3D structures of TNFRSF1A (2a, 2b: I-TASSER, 2c, 2d: Robetta) without (left: 2a, 2c) and with templates (right: 2b, 2d). The sky blue colored structure in each figure is an X-ray crystallographic model (1EXT) of the binding site of TNFRSF1A and the orange structure is the representation of that identical fragment in the model made by the web services. (To zoom in on the details within the figure it is recommend to look at the PDF version.)

4.1.4 Practical VIPUR usage

With the hindrance of software incompatibility on the cluster, difference in produced models between the web services, discovery of consequences by removing elements from structures, and the time it would take to reverse engineer VIPUR a new decision was formed. VIPUR would be set aside for now and if time was left it would be modified so that it could evaluate structures.

4.2 Analyses of proteins variants TNFRSF1A

4.2.1 Requirements for determining structural and binding effects of protein variants

Protein variants can be assessed from multiple perspectives and together they can form a holistic view on how a protein works and how mutations affect its workings. However adding perspectives to the protein assessment makes it complex and requires expertise to determine its validity and contribution, therefore the analysis has been limited to basic structural information.

Various proteins consist of multiple chains that can be identical or different depending on their function [100] and should be taken into account when assessing protein variants since one residue might alter the binding between chains and might alter the proteins formation. Different molecules and atoms that do not make up a protein but play a role in a pathway and function (ligands and co-receptors) are able to affect a proteins shape [101] and can behave differently when a residue is mutated.

A different aspect that can change with mutations is alteration in motions between structures which allow or disallow certain movements to occur and inhibit or promote binding.

4.2.2 Introduction of the simple protein variant analysis approach

A different method for analyzing proteins variants than VIPUR is by determining the changes in energy levels (Section 3.3) between a wild type and a variant. To make a variant of the wild type, a structure was required wherein a missense mutation could be introduced with Modeller (Section 3.5.1). The backbone structure of the variant was modified with the backrub application (Section 3.3.4) to make it better interact with other amino acid backbones in the structure resulting in 1000 models. The lowest Rosetta scoring (Section 3.3) structure would be selected to further improve the side chains with the Relax application (Section 3.3.1). With the application 64 models where made of which all scores were plotted in a density plot with R (Section 3.8.2) and the lowest scoring model was visualized with PyMOL (Section 3.5.2). This method is called the simple protein variant analysis approach and was only test on TNFRSF1A (Section 2.9) and its ligands TNF α and β . It shows similarities to VIPUR in the sense that it has to remove waters and metals, but it can keep its ligands and duplicate chains.

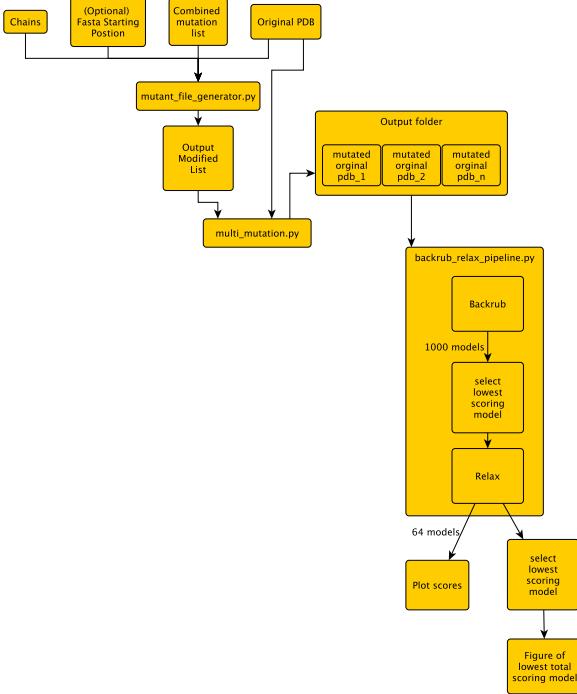


Figure 3: Flowchart of SPVAA wherein a list of known mutations generate the appropriate information for modeller to mutate residues in the original PDB. The mutated models are fed into the backrub relax pipeline to transform the structures into a low energy state. (To zoom in on the details within the figure it is recommended to look at the PDF version.)

4.2.3 Carrying out SPVAA on TNFRSF1A

Before introducing mutations into a protein structure it is helpful to know if a mutation has been observed to avoid allocating resources to mutations that do not occur. Therefore three tables with observed TNFRSF1A mutations (Sections 3.6.1, 3.6.2, 3.6.3) have been combined with an R script (Section 3.8.2) into a single table consisting of two columns. The first column (split into three columns table 1) contains strings that describes the: original residue, position and new residue, the second column describes whether a formed mutation is pathogenic or benign.

Original residue	Position in the protein sequence	New residue	Classification
Cys	44	Tyr	PATHOGENIC
Thr	44	Pro	PATHOGENIC
Thr	44	Ser	PATHOGENIC

Table 1: The format wherein mutations were filtered from the GAVIN, gnomAD and Infevers tables (Sections 3.6.1, 3.6.2, 3.6.3), describe whether a structural mutation is harmful or not, for many mutations in the table the effect is unknown. (To view the whole table visit the supplementary.)

For assessing variants in TNFRSF1A a structural fragment was used that contained TNF β (1TNR) [40] which was made homotrimeric with PyMOL (Section 3.5.2). It resulted in six chains that emulate a bound TNFRSF1A with TNF β . The first column of the mutation table did not contain sufficient information to apply mutations correctly and within the PDB different numbering is used than in the amino acid sequence. To bundle the information and make it usable for introducing mutations a Python script (Section 3.8.1) has been written that combines the mutation table, PDB chains and the correct

position within the sequence into a type of table which has sufficient information to mutate structures.
The script that reads table 1 and produces 2 in combination with a PDB, respective chains and an optional starting position is the mutant_file_generator.py.

Iteration number	Filename	Chain	Residue index in chain	New residue
34	1tnr3_TNFA	R	0	TYR
34	1tnr3_TNFA	T	0	TYR
34	1tnr3_TNFA	S	0	TYR
35	1tnr3_TNFA	R	0	PRO
35	1tnr3_TNFA	T	0	PRO
35	1tnr3_TNFA	S	0	PRO
36	1tnr3_TNFA	R	0	SER
36	1tnr3_TNFA	T	0	SER
36	1tnr3_TNFA	S	0	SER

Table 2: The format that describes the mutations that should be made by Modeller (Section 3.5.1), with specifications of the: model, file, chain, residue index and the new residue. The whole table for TNFA and TNFB are visible within the supplementary.

To introduce mutations within PDB structures a Python script (Section 3.8.1) was written which used the generated mutation table (Table: 2) and a matching PDB structure, from the table. The iteration number specifies if a mutation has to be stored in a single file or across multiple files. The filename serves as key that determines the PDB that should be used. The letters specify chains, numbers are indices within a chain (Python list) and the last column states the three letter code of the new residue. When a structure is read in through the Python bindings of Modeller (Section 3.5.1) all non standard atoms and molecules are removed because Rosetta (Section 3.3) is not able to deal with those atoms. Just before mutagenesis takes place missing atoms are added to the structure that were difficult to determine with experimental methods(Section 2.3). After the introduction of a last attempt was made by modeller to add disulfide bridges on places where a potential connection could be made based on distances between cysteine residues.

The script that is able to produce mutations into PDBs and uses table 2 is multi_mutation.py

With many protein variants generated and limited resources available a few mutations had to be picked to undergo the backrub relax pipeline. All mutations that were analyzed were picked from the Infevers table because these variants are from a single isoform and had a protein structure (1TNR [40]) available that interacted with TNF β . Mutations cysteine 62 to glycine and phenylalanine acid 141 to isoleucine were validated as pathogenic mutations within Infevers table and were used to determine the effectiveness of SPVAA. Within the infevers table no benign validated missense mutations were available [102], but to still have the opportunity to asses a likely benign mutation, the mutation glutamic acid 138 to alanine had been chosen to be assessed with SPVAA.

In the attempt to make mutated structures behave more natural two tools from the Rosetta software suite (Section 3.3) had been used to minimize energies within protein structures. With the Backrub application (Section 3.3.4) 1000 altered backbone models have been produced each with 10000 Monte Carlo moves (Sections 2.7). For each model that Backrub generated a set of scores were assigned to the properties, which together formed a collective score that described energy and bond occurrence in nature (Section 3.3). Models of the wildtypes and mutants with the lowest collective score ,the total score, were chosen to undergo further side chain optimization within the Relax (Section 3.3.1). 64 different relaxed models were produced and with various scores related to the properties of which the one with lowest total score would be chosen to visualize.

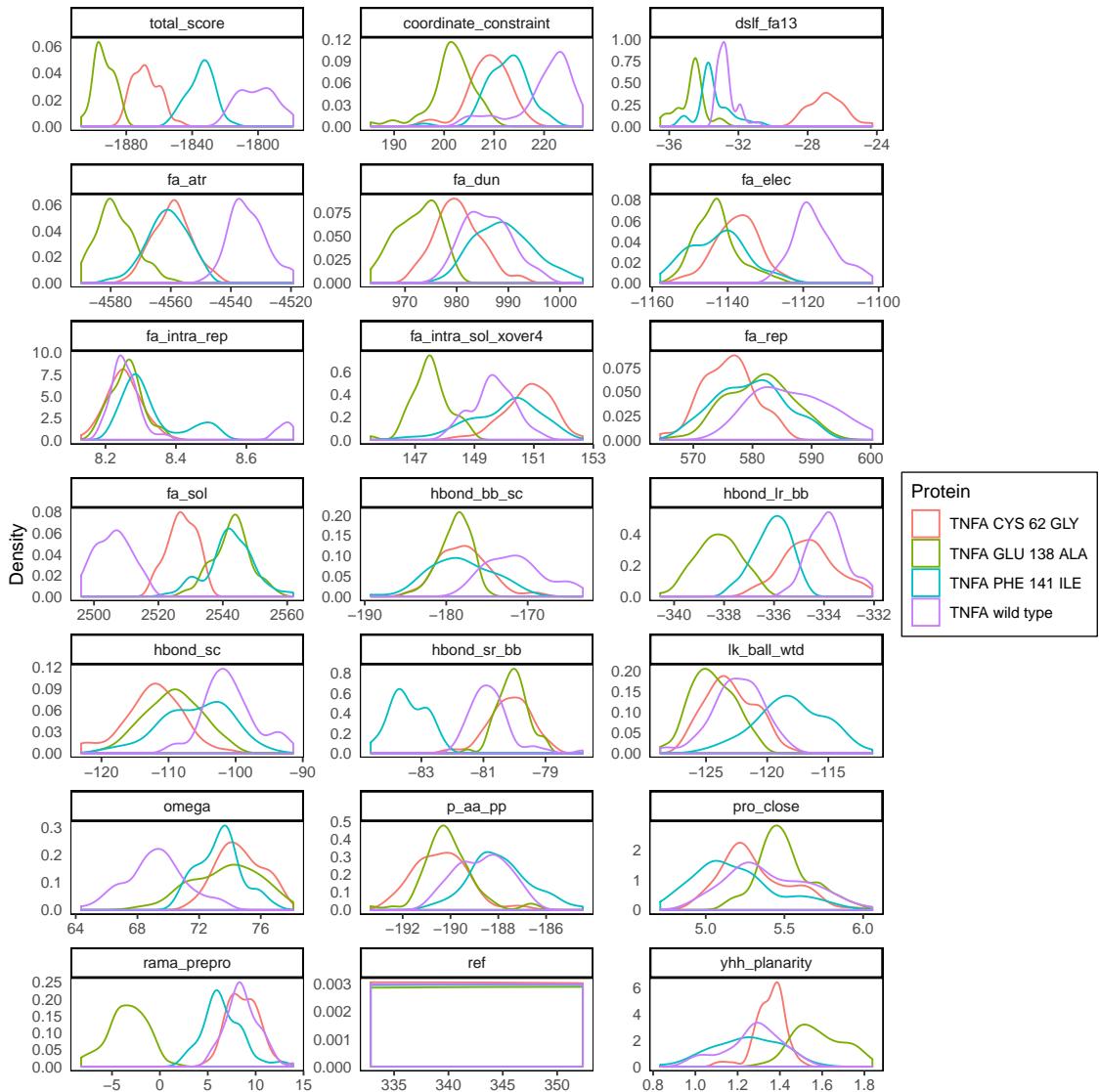


Figure 4: Density of different scoring metrics from the models produced with relax of the wild type and all mutants that interacted with TNF α . Most of the total scores, fa_atr values and fa_ele values of the mutated model are lower than the wild type that is bound to TNF α . The fa_sol values are higher of the wild type than the mutants. TNFA CYS 62 GLY has more higher values at the dslf_fa13 (disulfide geometry potential) than all other mutations. (Plots of backrub TNF α scores are in the supplementary.) (To zoom in on the details within the figure it is recommend to look at the PDF version.)

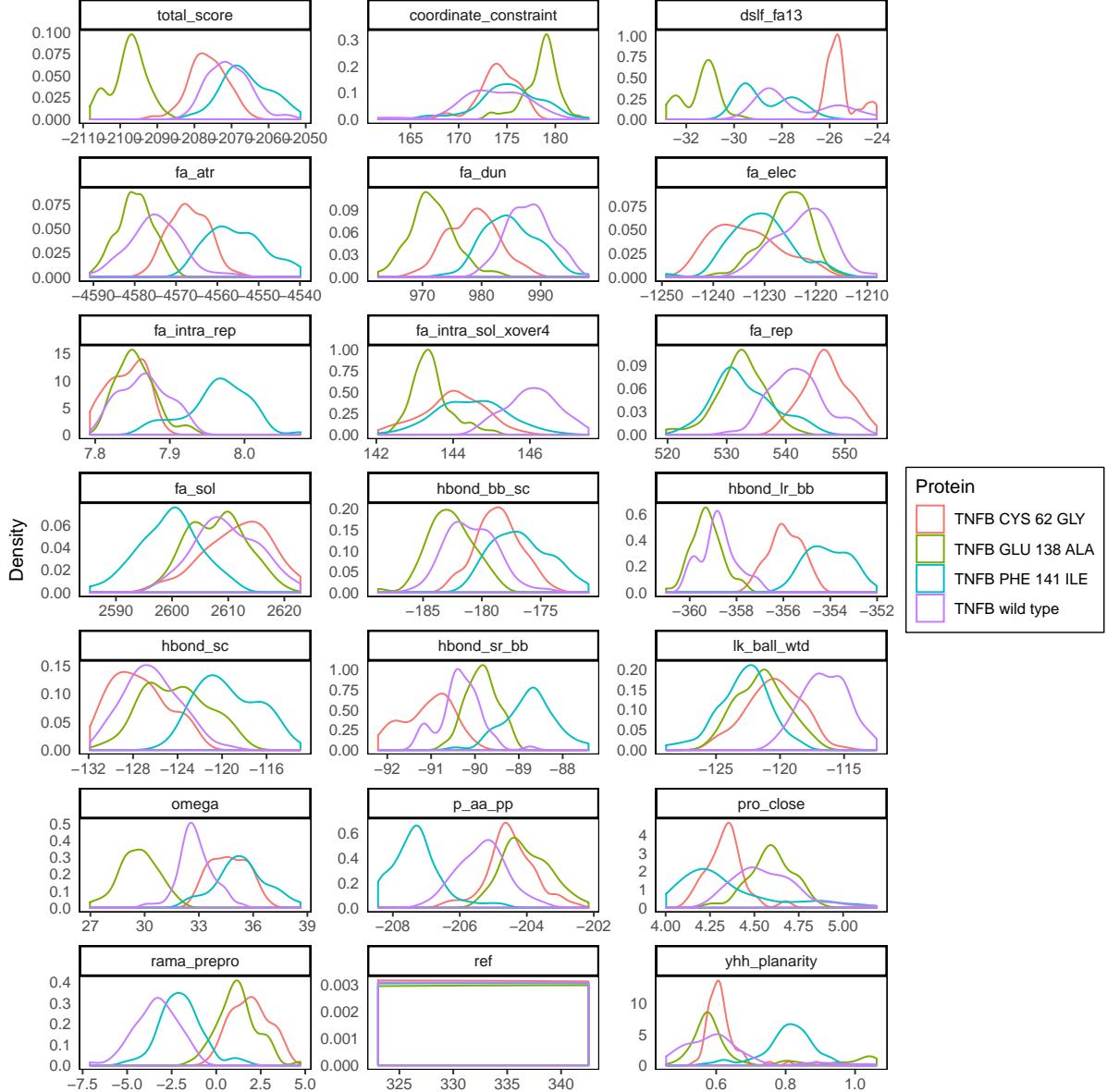


Figure 5: Density distributions of the scores generated by relax for the wild type and mutations of TNFRSF1A that interact with $\text{TNF}\beta$. The total score values of TNFB GLU 138 ALA are lower than all other models and dslf_fa13 (disulfide geometry potential) values are higher at TNFB CYS 62 GLY than the other models. fa_intra_rep (Lennard-Jones repulsive between atoms in the same residue) is higher within the models of PHE 141 ILE and hbond_lr_bb (Backbone-backbone hbonds distant in primary sequence) is higher at CYS 62 GLY and PHE 141 ILE. (Plots of backrub $\text{TNF}\beta$ scores are in the supplementary.) (To zoom in on the details within the figure it is recommended to look at the PDF version.)

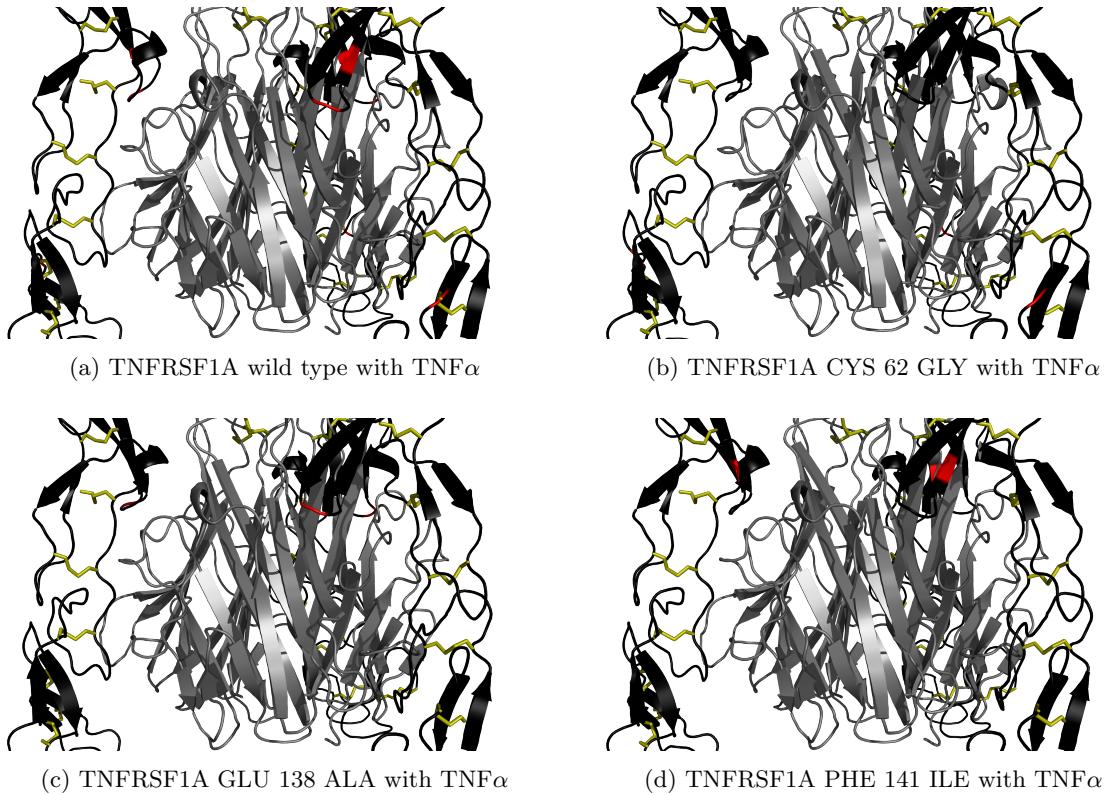


Figure 6: 3D structures of a homotrimer TNFRSF1As (black) with a homotrimer TNF α s (gray) and disulfide bridges (dark yellow). The wild type (6a) has three red colored areas which are the original residues of the protein before any form of mutation. Within CYS 62 GLY (6b) it is visible that at the position where a mutation is introduced (red) a disulfide bridge is missing. The mutations of GLU 138 ALA (6c) and PHE 141 ILE (6d) show no large differences at the mutated spots (red). (To zoom in on the details within the figure it is recommend to look at the PDF version.)

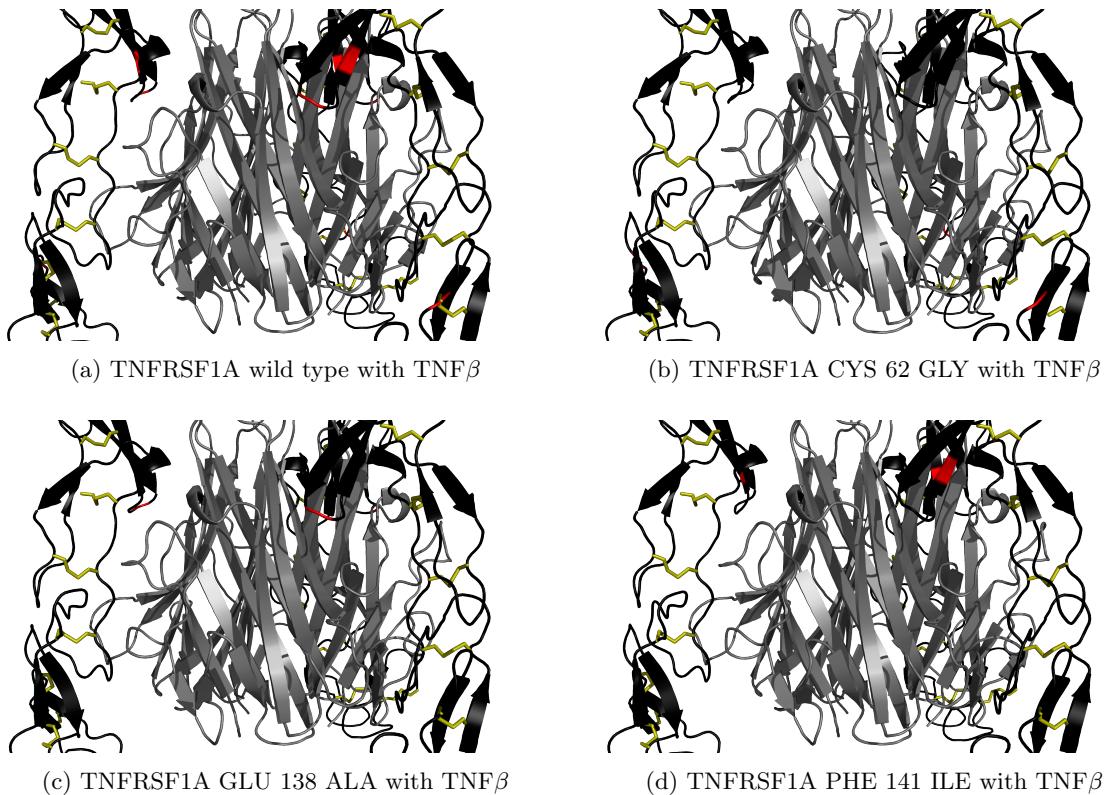


Figure 7: 3D structures of a homotrimer TNFRSF1As (black) with a homotrimer TNF β s (gray) and disulfide bridges (dark yellow). The wild type (7a) has three red colored areas which are the original residues of the protein before any form of mutation. Within CYS 62 GLY (7b) it is visible that at the position where a mutation is introduced (red) a disulfide bridge is missing. The mutations of GLU 138 ALA (7c) and PHE 141 ILE (7d) show no large differences at the mutated spots (red). (To zoom in on the details within the figure it is recommend to look at the PDF version.)

4.3 Finding mutation information with HOPE

A more textual informative approach is web service HOPE (Section 3.4.3) which makes a report about the mutation. The mutations that were known from Infevers (Section 3.6.3) and were also used with SPVAA were tested by HOPE (CYS62GLY, GLU138ALA and PHE141IIE). The reports are visible within the supplementary.

HOPEs first test was the mutation of Cysteine 62 to glycine, which is known within the Infevers table as pathogenic and was validated. It discovered that the residue was involved in a disulfide bridge and was 100% conserved in related protein sequences, based on the observation that cysteine formed a disulfide bridge it expected that with the replacement of it glycine would make the whole structures less rigid. HOPE predicted that mutation is pathogenic because of the high conservation of the residue, which is further confirmed by its search results in which it found the original publication of the discovery that it has been associated to TRAPS [37].

According to Infevers is the mutation of glutamic acid at position 138 mutated to alanine classified as likely benign and was not validated yet. HOPE discovered with a BLAST query that glutamic acid occurs often at position but other residues such as alanine have been observed at the position. Structurally glutamic acid forms salt bridges with proline 368 and leucine 390 and is found in a sequence of amino acids that is repeated through out TNFRSF1A. The amino acid lies within a domain where it interacts with other domains and is important for the proteins activity, with this mutation it might already perturb the binding capabilities according to HOPE.

The last mutation that was tested with HOPE was phenylalanine 141 to Isoleucine and was according to Infevers pathogenic and has been validated. Phenylalanine is conserved at this positions and few other residues have been seen at the position, it is a member of the identical domain as glutamic acid 138. HOPEs prediction was not clear if this mutation could cause any harm.

5 Discussion

Currently, most rare disease patients do not receive a molecular diagnosis. Despite machine learning methods such as GAVIN that can remove 95% of benign variation from the genome[1], it is still very difficult to pin-point causal variants in the genome. Such methods rely mostly on evolutionary conservation and have been heavily optimized over the years. Therefore we need new refreshing approaches such as VIPUR, which uses sequential and structural data instead, that have the exciting potential to help us diagnose more patients.

5.1 VIPUR

Within the attempt to make VIPUR usable for diagnostics it was discovered that some questionable steps were taken to make it applicable for diagnosis but also to determine deleteriousness ; (i) "All protein models were standardized to remove unwanted components (duplicate chains, ligands, metals and non-standard amino acids)" [2]. Standardizing data can be beneficial to avoid learning features from proteins that are available to some models but should not be the determining factor for classification. However any form of context to the protein is removed and might therefore make incorrect assumptions about how: a monomer interacts with other monomers, ligands, metals, non-standard amino acids and water which can all have an effect on how proteins shape and interact [101]. (ii) With the utilization of Rosetta's Relax application different models are formed based on the Monte Carlo method (Section 2.7). VIPUR produces 50 structures with Relax per protein which is a tiny amount of the potential search space of possible folds that could have made changes in a mutated protein, which is also visible in the scores of the models made from of TNFRSF1A with TNF α & β (Figures 4, 5). Rosetta itself suggests to make sufficient models, starting with a minimum of 5000[103]. (iii) The features acquired with probe in combination with the models that were produced, within the publication of Probe is mentioned: "It requires both highly accurate structures and also the explicit inclusion of all hydrogen atoms and their van der Waals interactions." [62]. It is not possible to determine if the structures were accurate. However is it likely that no loose hydrogen atoms were included within the structures because all were standardized. To make the outcome of Probe useful to VIPUR the program Reduce should have been executed first, which adds hydrogen atoms to the structure. It is recommended on the site of Probe to execute Reduce on the structure before using VIPUR [63].

More questions arise when further investigating the publication. Within the figures (4, 5) [2] and supplementary figures (10, 11) [58] are heatmaps of PSSMs added that display the values of the natural and mutated residues. In combination with the methods used on standardizing structures and collecting features a suspicion arises that there is little contribution from the structural features and that prediction depends on PSI-BLAST results. Figure 4 within the publication shows a protein wherein serine 204 which part of an α helix is mutated to proline and is predicted as deleterioueness by VIPUR. Which is logical even without predictions because prolines are known to be α helix stoppers [104] and therefor affect the form of an α helix (which is also not visible in the figure.).

VIPUR has not been used for various reasons. The models that were predicted by the web services (Figures 2) had a decent accuracy for the binding site, but the rest of structures differed and made it too difficult to determine which model was an accurate representation of TNFRSF1A. VIPUR was built on and for a single system and required reverse engineering to make it work in any form even without the substitution of PyMOL and PyRosetta (Sections 3.5.2, 3.5.3) with Modeller (Section 3.5.1). The differences in models is probably due to the fact that TNFRSF1A is transmembrane protein which is hard to acquire structures from with experimental methods [24], even though $\sim 57\%$ of TNFRSF1As structure was known (1EXT[93] , 1ICH[99]).

Although it is not part of developing a technique that can help diagnosing rare disease variants but the publication contains a claim ("VIPUR can be applied to mutations in any organism's proteome...." [2]) which contradicts with its methods: "remove unwanted components (duplicate chains, ligands, metals and non-standard amino acids)" [2]. Currently there are more than 140 amino acids found in natural proteins of which 22 are part of the amino acid alphabet and 20 of those are classified as standard [105].

By removing the non-standard amino acids from proteins it becomes impossible to analyze mutations in every organisms proteome.

5.2 SPVAA

SPVAA did not assess whole complex and neither did became a machine learning tool ready to use in diagnosis to make automated predictions. SPVAA has similar weaknesses as VIPUR wherein: water, metals and other molecules are removed from its structure. However the proteins can keep their extra monomers and protein ligands, even when the structure requires identical or different ones they can be added manually into the structure and analyzed.

For three variants it has been attempted to acquire structural information to determine their differences with the wild type structure, in two of the three structures hardly any changes were visible. From the models that were assessed only homotrimeric TNF α - β bound mutations were processed, not the unboud dimeric structures. From the proteins that were modeled too few were produced of each structure. Backrub used 10000 Monte Carlo moves, which is very little compared for the amount of residues it has and should have had more to make larger changes in the structure. The relax application made 64 models per mutation but Rosetta itself suggest to make at least 5000[103]. The only models with mutations where pathogenicity was highly likely visible were the CYS 62 GLY models, that had broken disulfide bridges that could lead to instability in the protein. Based on the scores produced by Relax little could be discovered except the higher scores in the disulfide geometry potential (dslf_fa13) of CYS 62 GLY. Many of the scores show overlap and are difficult to relate with the applied structural changes. It could have been that many of the relax score distributions should have had more overlap with each because too few models were made.

The models could have had in other situations better disulfide bridge formation which did not rely on the guess of distance between cysteine residues. A better method which Modeller has is to form disulfide bridges based on other protein data that is available in some situations. Within TNFRSF1A it likely did not matter too much because except CYS 62 GLY all other methods show identical disulfide bridges.

5.3 HOPE

HOPE is an informative tool that is easy to use, fast and makes structural problems within proteins understandable when a missense mutation is discovered. However it does not draws a solid conclusions and the information it collects depends on: previous publications, conservation and experimental structures. HOPE has a disadvantage when limited knowledge is available. Also it does not asses a complex but it can describe binding sites from the monomer when previously discovered. For the mutation CYS 62 GLY it was very clear based on conservation and the publication that it was pathogenic. However for the pathogenic mutation PHE 141 ILE the information was less clear. Glutamic acid at position 138 makes salt bridges according to HOPE which are in general strong bonds and can be important for internal structures and binding. With this information GLU 138 ALA would be likely pathogenic, however it does not draw a conclusion. Also no change in structure or interaction has been observed with HOPE or SPVAA and therefore can not be classified pathogenic with certainty and would therefor remain likely benign.

6 Conclusion

All used methods use structural information acquired from experimentally determine structures which are mostly fragments due to the difficulty and expenses of determining structures. Because they are mostly fragments and miss context it can be hard to make relevant assumptions about some protein structures or its general effect in a cell.

At its current state VIPUR is not usable for diagnosing defects within proteins and most likely also not for predicting deleteriousness. To make it usable for predicting effects of variants or deleteriousness it could benefit of manual curation of all wild types, which is partially fulfilled by using proteins from the SWISS-MODEL database. A different addition would be looking at the structure in context of its environment since that is where the wild type should behave properly. With that information proper predictions can be made about the deleteriousness of a protein or its pathogenicity.

SPVAA analyses proteins more into a natural context than VIPUR because it allows the use of more chains and ligands. However the assessment of structures made by SPVAA require expertise to determine the effects of mutations and at its current state it is not user friendly or helpful to inexperienced users. CYS 62 GLY was the only mutation that without prior knowledge could be assumed pathogenic based on the plots and models (Figures 4, 5, 6b, 7b) that were produced. None of the other mutations contained clear information whether they would be pathogenic or not and hardly any differences have been observed between TNF α - β .

HOPE is an informative tool that gave new insight in the mutation GLU 138 ALA, in some cases it can be very clear and almost form a conclusion but in other situations it is unable to discover effects of a mutation to elucidates its user and makes the dependence of previously investigated knowledge visible.

7 Future work

The VIPUR approach could be investigated to test whether the features generated by PSI-BLAST are the main predictors. To measure PSI-BLAST feature importance within VIPUR it first needs to be reverse engineered to make it work with Modeller or another tool which is able to implement mutations in PDB files. Once the reverse engineering is finished VIPURs feature importance can be tested with shap [106–112] that explains the output of machine learning models and or with similar methods and software on the VTS.

SPVAA is currently highly dependent on the resources that are available and comes short to produce enough models [103]. One option to improve the quantity of models is by using different software that is less resource intensive. Another option would be to run SPVAA on a different cluster that has more nodes and allows to setup jobs that use multiple nodes. SPVAA is not a prediction method and has to be modified and expanded become a variant predictor that is able to predict pathogenicity or deleteriouness. A good starting point for such a predictor would be according to the guidelines in section 2.5.

VIPUR and SPVAA could both be improved in various ways, one of them would be by doing molecular dynamic simulations on the mutated structures to determine the effects of structural changes. With SPVAA it would have most likely become clear if the loss of the disulfide bridge from CYS 62 GLY in TNFRSF1A would have caused structural issues. VIPUR could benefit from molecular dynamics as new machine learning feature in situations where limited movement is observed and it changes tremendously when a missense mutation occurred in a protein increases with a mutation or vice versa.

References

1. Van der Velde, K. J. *et al.* GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome Biology* **18**. ISSN: 1474-7596. doi:10.1186/s13059-016-1141-7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5240400/> (2019) (Jan. 16, 2017).
2. Baugh, E. H. *et al.* Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Research* **44**, 2501–2513. ISSN: 0305-1048 (Apr. 7, 2016).
3. Venselaar, H., te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548. ISSN: 1471-2105 (Dec. 2010).
4. Harper, A. R., Nayee, S. & Topol, E. J. Protective alleles and modifier variants in human health and disease. *Nature Reviews Genetics* **16**, 689–701. ISSN: 1471-0064 (Dec. 2015).
5. NIH. *Sickle cell disease* Genetics Home Reference. <https://ghr.nlm.nih.gov/condition/sickle-cell-disease> (2019).
6. NIH. *Cystic fibrosis* Genetics Home Reference. <https://ghr.nlm.nih.gov/condition/cystic-fibrosis> (2019).
7. NIH, R. G. H. *TRAPS* Genetics Home Reference. <https://ghr.nlm.nih.gov/condition/tumor-necrosis-factor-receptor-associated-periodic-syndrome> (2019).
8. Wikipedia. in *Wikipedia* Page Version ID: 891541555 (Apr. 8, 2019). https://en.wikipedia.org/w/index.php?title=Protein_structure&oldid=891541555 (2019).
9. Bennion, B. J. & Daggett, V. Protein Conformation and Diagnostic Tests: The Prion Protein. *Clinical Chemistry* **48**, 2105–2114. ISSN: 0009-9147, 1530-8561 (Dec. 1, 2002).
10. Feyfant, E., Sali, A. & Fiser, A. Modeling mutations in protein structures. *Protein Science : A Publication of the Protein Society* **16**, 2030–2041. ISSN: 0961-8368 (Sept. 2007).
11. Chothia, C. & Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *The EMBO Journal* **5**, 823–826. ISSN: 0261-4189 (Apr. 1986).
12. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603. ISSN: 1476-4687 (May 2001).
13. NIH. *The Cost of Sequencing a Human Genome* Genome.gov. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> (2019).
14. Hortin, G. L., Carr, S. A. & Anderson, N. L. Introduction: Advances in Protein Analysis for the Clinical Laboratory. *Clinical chemistry* **56**, 149–151. ISSN: 0009-9147 (Feb. 2010).
15. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812–3814. ISSN: 0305-1048 (July 1, 2003).
16. Nanev, C. N. How do crystal lattice contacts reveal protein crystallization mechanism? *Crystal Research and Technology* **43**, 914–920. ISSN: 1521-4079 (2008).
17. Niu, B. *et al.* Protein-structure-guided discovery of functional mutations across 19 cancer types. *Nature Genetics* **48**, 827–837. ISSN: 1546-1718 (2016).
18. wwPDB. *wwPDB: Deposition Statistics* <http://www.wwpdb.org/stats/deposition> (2019).
19. Cantrill, S. *Chemiotics: How many proteins can we make? : The Sceptical Chymist* http://blogs.nature.com/thescepticalchymist/2008/04/chemiotics_how_many_proteins_c.html (2019).
20. PDB101. *PDB101: Learn: Guide to Understanding PDB Data: Missing Coordinates and Biological Assemblies* RCSB: PDB-101. <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/missing-coordinates-and-biological-assemblies> (2019).
21. Ridgen, D. J. in *From Protein Structure to Function With Bioinformatics* 2nd ed., 503 (Springer). ISBN: ISBN 978-94-024-1069-3.

22. Wikipedia. in *Wikipedia* Page Version ID: 895739876 (May 6, 2019). https://en.wikipedia.org/w/index.php?title=Ramachandran_plot&oldid=895739876 (2019).
23. Shourya, S., Burman, R. & Mulligan, V. K. *Scoring Tutorial* <https://rosettacommons.org/demos/latest/tutorials/scoring/scoring#comparing-rosetta-scores-to-real-life-energies> (2019).
24. Yonath, A. X-ray crystallography at the heart of life science. *Current Opinion in Structural Biology. Carbohydrates and glycoconjugates/Biophysical methods* **21**, 622–626. ISSN: 0959-440X (Oct. 1, 2011).
25. Evans, R., Jumper, J., Kirkpatrick, J. & Sifre, L. *AlphaFold: Using AI for scientific discovery* DeepMind. <https://deepmind.com/blog/alphafold/> (2019).
26. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. & Murzin, A. G. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research* **42**, D310–D314. ISSN: 0305-1048 (Database issue Jan. 1, 2014).
27. Wikipedia. in *Wikipedia* Page Version ID: 898321911 (May 22, 2019). https://en.wikipedia.org/w/index.php?title=Membrane_protein&oldid=898321911 (2019).
28. Wikipedia. in *Wikipedia* Page Version ID: 898320912 (May 22, 2019). https://en.wikipedia.org/w/index.php?title=Globular_protein&oldid=898320912 (2019).
29. Wikipedia. in *Wikipedia* Page Version ID: 867647761 (Nov. 7, 2018). <https://en.wikipedia.org/w/index.php?title=Scleroprotein&oldid=867647761> (2019).
30. Wikipedia. in *Wikipedia* Page Version ID: 891043075 (Apr. 5, 2019). https://en.wikipedia.org/w/index.php?title=Intrinsically_disordered_proteins&oldid=891043075 (2019).
31. Stephanie. *Monte Carlo Simulation / Method* Statistics How To. <https://www.statisticshowto.datasciencecentral.com/monte-carlo-simulation/> (2019).
32. Wikipedia. in *Wikipedia* Page Version ID: 896113843 (May 8, 2019). https://en.wikipedia.org/w/index.php?title=Monte_Carlo_method&oldid=896113843 (2019).
33. Wikipedia. in *Wikipedia* Page Version ID: 51889441 (July 2, 2018). <https://nl.wikipedia.org/w/index.php?title=Monte-Carlosimulatie&oldid=51889441> (2019).
34. Alon Honig. *Introduction to Monte Carlo Methods* <https://www.youtube.com/watch?v=t0F3S-46bIQ> (2019).
35. Roth-Wojcicki, E. *Tumor Necrosis Factor Receptor Associated Periodic Syndrome (Juvenile)* <https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Tumor-Necrosis-Factor-Receptor-Associated-Periodic-Syndrome-Juvenile> (2019).
36. Kimberley, F. C., Lobito, A. A., Siegel, R. M. & Screaton, G. R. Falling into TRAPS-receptor misfolding in the TNF receptor 1-associated periodic fever syndrome. *Arthritis Research & Therapy* **9**, 217. ISSN: 1478-6362 (2007).
37. Aksentijevich, I. et al. The tumor-necrosis-factor receptor-associated periodic syndrome: new mutations in TNFRSF1A, ancestral origins, genotype-phenotype studies, and evidence for further genetic heterogeneity of periodic fevers. *American Journal of Human Genetics* **69**, 301–314. ISSN: 0002-9297 (Aug. 2001).
38. Gray, P. W., Barrett, K., Chantry, D., Turner, M. & Feldmann, M. Cloning of human tumor necrosis factor (TNF) receptor cDNA and expression of recombinant soluble TNF-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 7380–7384. ISSN: 0027-8424 (Oct. 1990).
39. Walter, R. & Stefan, O. in *Encyclopedia of Molecular Pharmacology* 2nd ed., 1505 (Springer, Nov. 2007). ISBN: 978-3-540-38918-7 978-3-540-38921-7.

40. Banner, D. W. *et al.* Crystal structure of the soluble human 55 kd TNF receptor-human TNF complex: Implications for TNF receptor activation. *Cell* **73**, 431–445. ISSN: 0092-8674 (May 7, 1993).
41. Segueni, N. *et al.* Innate myeloid cell TNFR1 mediates first line defence against primary Mycobacterium tuberculosis infection. *Scientific Reports* **6**. ISSN: 2045-2322. doi:10.1038/srep22454. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4773807/> (2019) (Mar. 2, 2016).
42. Naismith, J. H., Devine, T. Q., Brandhuber, B. J. & Sprang, S. R. Crystallographic Evidence for Dimerization of Unliganded Tumor Necrosis Factor Receptor. *Journal of Biological Chemistry* **270**, 13303–13307. ISSN: 0021-9258, 1083-351X (June 2, 1995).
43. Bender, L. M., Morgan, M. J., Thomas, L. R., Liu, Z.-G. & Thorburn, A. The adaptor protein TRADD activates distinct mechanisms of apoptosis from the nucleus and the cytoplasm. *Cell Death & Differentiation* **12**, 473. ISSN: 1476-5403 (May 2005).
44. Muppidi, J. R., Tschopp, J. & Siegel, R. M. Life And Death Decisions: Secondary Complexes and Lipid Rafts in TNF Receptor Family Signal Transduction. *Immunity* **21**, 461–465. ISSN: 1074-7613 (Oct. 1, 2004).
45. Vinay, K., Abul, K. & Jon, C. in *Robbins and Cotran Pathologic Basis of Disease, Professional Edition* 9th, 1464 (Elsevier, July 9, 2014). ISBN: 978-0-8153-4432-2.
46. Chen, G. & Goeddel, D. V. TNF-R1 Signaling: A Beautiful Pathway. *Science* **296**, 1634–1635. ISSN: 0036-8075, 1095-9203 (May 31, 2002).
47. Hengartner, M. O. The biochemistry of apoptosis. *Nature* **407**, 770. ISSN: 1476-4687 (Oct. 2000).
48. Aggarwal, B. B., Eessalu, T. E. & Hass, P. E. Characterization of receptors for human tumour necrosis factor and their regulation by -interferon. *Nature* **318**, 665. ISSN: 1476-4687 (Dec. 1985).
49. Hamosh, A. & McKusick, V. A. OMIM Entry - * 153440 - LYMPHOTOXIN-ALPHA; LTA <https://omim.org/entry/153440?search=lymphotoxin&highlight=lymphotoxin> (2019).
50. Kriegler, M., Perez, C., DeFay, K., Albert, I. & Lu, S. D. A novel form of TNF/cachectin is a cell surface cytotoxic transmembrane protein: Ramifications for the complex physiology of TNF. *Cell* **53**, 45–53. ISSN: 0092-8674 (Apr. 8, 1988).
51. Pieper, U. *et al.* modbase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research* **37**, D347–D354. ISSN: 0305-1048 (Database issue Jan. 2009).
52. Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS* **30**, S162–S173. ISSN: 1522-2683 (S1 2009).
53. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* **7**, 10480. ISSN: 2045-2322 (Sept. 5, 2017).
54. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350. ISSN: 1367-4803 (Feb. 1, 2011).
55. Bienert, S. *et al.* The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* **45**, D313–D319. ISSN: 0305-1048 (D1 Jan. 4, 2017).
56. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296–W303. ISSN: 0305-1048 (W1 July 2, 2018).
57. Poultney, C. S. *et al.* Rational Design of Temperature-Sensitive Alleles Using Computational Structure Prediction. *PLOS ONE* **6**, e23947. ISSN: 1932-6203 (Sept. 2, 2011).
58. Baugh, E. H. *et al.* SUPPLEMENTARY: Robust classification of protein variation using structural modelling and large-scale data integration. *Nucleic Acids Research* **44**, 2501–2513. ISSN: 0305-1048 (Apr. 7, 2016).

59. NCBI. *PSIBLAST* <http://www.biology.wustl.edu/gcg/psiblast.html> (2019).
60. NCBI. *PSSM Viewer* https://www.ncbi.nlm.nih.gov/Class/Structure/pssm/pssm_viewer.cgi (2019).
61. Wikipedia. in *Wikipedia* Page Version ID: 890694148 (Apr. 2, 2019). <https://en.wikipedia.org/w/index.php?title=BLAST&oldid=890694148> (2019).
62. Word, J. M. *et al.* Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms¹¹Edited by J. Thornton. *Journal of Molecular Biology* **285**, 1711–1733. ISSN: 0022-2836 (Jan. 29, 1999).
63. LAB, R. *Probe Software : Kinemage Website* <http://kinemage.biochem.duke.edu/software/probe.php> (2019).
64. Commons, R. *About — RosettaCommons* <https://www.rosettacommons.org/about> (2019).
65. Commons, R. *How to prepare structures for use in Rosetta* https://www.rosettacommons.org/docs/latest/rosetta_basics/preparation/preparing_structures (2019).
66. Commons, R. *Relax application* https://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/relax (2019).
67. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Science : A Publication of the Protein Society* **23**, 47–55. ISSN: 0961-8368 (Jan. 2014).
68. Tyka, M. D. *et al.* Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *Journal of Molecular Biology* **405**, 607–618. ISSN: 0022-2836 (Jan. 14, 2011).
69. Leaver-Fay, A. & Kellogg, E. *ddg-monomer application* https://www.rosettacommons.org/docs/latest/application_documentation/analysis/ddg-monomer (2019).
70. Jared, A.-B. *Score Commands* https://www.rosettacommons.org/docs/latest/application_documentation/analysis/score-commands (2019).
71. Betancourt, M. R. Efficient Monte Carlo trial moves for polypeptide simulations. *The Journal of Chemical Physics* **123**, 174905. ISSN: 0021-9606 (Oct. 31, 2005).
72. Smith, C. A. *Backrub application* https://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/backrub (2019).
73. Song, Y. *et al.* High-Resolution Comparative Modeling with RosettaCM. *Structure* **21**, 1735–1742. ISSN: 0969-2126 (Oct. 8, 2013).
74. Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960. ISSN: 1367-4803, 1460-2059 (Apr. 1, 2005).
75. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nature protocols* **7**, 1511–1522. ISSN: 1754-2189 (July 19, 2012).
76. Yang, Y., Faraggi, E., Zhao, H. & Zhou, Y. Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* **27**, 2076–2082. ISSN: 1367-4803 (Aug. 1, 2011).
77. Ovchinnikov, S. *et al.* Protein structure determination using metagenome sequence data. *Science* **355**, 294–298. ISSN: 0036-8075, 1095-9203 (Jan. 20, 2017).
78. LAB, Z. *LOMETS* <https://zhanglab.ccmb.med.umich.edu/LOMETS/help.html> (2019).
79. Wu, S. & Zhang, Y. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* **35**, 3375–3382. ISSN: 0305-1048 (May 2007).
80. CMBI. *HOPE* <http://www.cmbi.ru.nl/hope/> (2019).
81. CMBI. *HOPE about* <http://www.cmbi.ru.nl/hope/about/> (2019).

82. CMBI. *HOPE methods* <http://www.cmbi.ru.nl/hope/method/> (2019).
83. Modeller. *About MODELLER* <https://salilab.org/modeller/> (2019).
84. Eswar, N. *et al.* Comparative Protein Structure Modeling Using Modeller. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* **0 5**, Unit-5.6. ISSN: 1934-3396 (Oct. 2006).
85. Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779–815. ISSN: 0022-2836 (Dec. 5, 1993).
86. Fiser, A., Do, R. K. & Sali, A. Modeling of loops in protein structures. *Protein Science : A Publication of the Protein Society* **9**, 1753–1773. ISSN: 0961-8368 (Sept. 2000).
87. Schrödinger. *PyMOL* — pymol.org/2/ (2019).
88. Jeffrey, J. G., Sergey, L. & Team, P. *PyRosetta* <http://www.pyrosetta.org/> (2019).
89. ensembl. *Variant Effect Predictor - Homo sapiens - GRCh37 Archive browser 96* http://grch37.ensembl.org/Homo_sapiens/Tools/VEP (2019).
90. Plon, S. E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation* **29**, 1282–1291. ISSN: 1059-7794 (Nov. 2008).
91. gnomAD. *gnomAD* <https://gnomad.broadinstitute.org/> (2019).
92. Sarrauste de Menthière, C. *et al.* INFEVERS: the Registry for FMF and hereditary inflammatory disorders mutations. *Nucleic Acids Research* **31**, 282–285. ISSN: 0305-1048 (Jan. 1, 2003).
93. Naismith, J. H., Devine, T. Q., Kohno, T. & Sprang, S. R. Structures of the extracellular domain of the type I tumor necrosis factor receptor. *Structure* **4**, 1251–1262. ISSN: 0969-2126 (Nov. 15, 1996).
94. Burley, S. K. *et al.* RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Science* **27**, 316–330. ISSN: 0961-8368 (Jan. 1, 2018).
95. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204–D212. ISSN: 0305-1048 (Database issue Jan. 28, 2015).
96. Wickham, H. *Create Elegant Data Visualisations Using the Grammar of Graphics* <https://ggplot2.tidyverse.org/> (2019).
97. Dowle, m., Srinivasan, A., Gorecki, J. & Chirico, M. *R’s data.table package extends data.frame: Contribute to Rdatatable/data.table development by creating an account on GitHub* original-date: 2014-06-07T16:38:05Z. June 11, 2019. <https://github.com/Rdatatable/data.table> (2019).
98. Baugh, E. H. VIPUR: Variant Interpretation and Prediction Using Rosetta. doi:None. <https://osf.io/bd2h4/> (2019) (Sept. 15, 2015).
99. Sukits, S. F. *et al.* Solution structure of the tumor necrosis factor receptor-1 death domain. *Journal of Molecular Biology* **310**, 895–906. ISSN: 0022-2836 (July 20, 2001).
100. Liu, G.-H. *et al.* Lipin proteins form homo- and hetero-oligomers. *The Biochemical journal* **432**, 65–76. ISSN: 0264-6021 (Oct. 25, 2010).
101. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis*. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 98–104. ISSN: 0027-8424 (Feb. 1958).
102. Aksentijevich, I. *Infevers - Tabular list* <https://infevers.umai-montpellier.fr/web/search.php?n=2> (2019).
103. Commons, R. *Analyzing Results* <https://www.rosettacommons.org/docs/latest/getting-started/Analyzing-Results> (2019).

104. Li, S. C., Goto, N. K., Williams, K. A. & Deber, C. M. Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6676–6681. ISSN: 0027-8424 (June 25, 1996).
105. Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nature Chemical Biology* **3**, 29–35. ISSN: 1552-4469 (Jan. 2007).
106. Slundberg. *slundberg/shap: A unified approach to explain the output of any machine learning model.* <https://github.com/slundberg/shap> (2019).
107. Štrumbelj, E. & Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowl. Inf. Syst.* **41**, 647–665. ISSN: 0219-1377 (Dec. 2014).
108. Ribeiro, M. T., Singh, S. & Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv:1602.04938 [cs, stat]*. arXiv: 1602.04938. <http://arxiv.org/abs/1602.04938> (2019) (Feb. 16, 2016).
109. Shrikumar, A., Greenside, P. & Kundaje, A. *Learning Important Features Through Propagating Activation Differences* in International Conference on Machine Learning International Conference on Machine Learning (July 17, 2017), 3145–3153. <http://proceedings.mlr.press/v70/shrikumar17a.html> (2019).
110. Datta, A., Sen, S. & Zick, Y. Algorithmic Transparency via Quantitative Input Influence: 20.
111. Bach, S. *et al.* On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE* **10**. ISSN: 1932-6203. doi:10.1371/journal.pone.0130140. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498753/> (2019) (July 10, 2015).
112. Datadive. *Interpreting random forests — Diving into data* Datadive. <http://blog.datadive.net/interpreting-random-forests/> (2019).

Supplementary

Observed mutations table

Mutation	Classification
Ser4Phe	NA
Val6Met	NA
Pro7Thr	NA
Pro12Leu	NA
Glu14Lys	PATHOGENIC
Leu15Val	NA
Thr16Pro	PATHOGENIC
Thr16Ala	PATHOGENIC
Cys17Ser	PATHOGENIC
Val20Ala	Uncertain significance (VOUS)
Val21Ile	PATHOGENIC
Gly21Ala	NA
Val21Phe	PATHOGENIC
Val21Leu	PATHOGENIC
Arg24Trp	BENIGN
Gly26Ala	NA
Val27Asp	NA
Ile28Phe	NA
Arg28Trp	PATHOGENIC
Ala30Pro	PATHOGENIC
Ala30Gly	PATHOGENIC
Ala30Ser	PATHOGENIC
Ala30Thr	PATHOGENIC
Val31Gly	NA
Val33Leu	PATHOGENIC
Val33Met	PATHOGENIC
Arg34Thr	PATHOGENIC
Gly35Arg	NA
Ser38Pro	PATHOGENIC
Pro39Thr	PATHOGENIC
Asp41His	NA
Asp41Glu	NA
Thr43Ser	PATHOGENIC
Cys44Tyr	PATHOGENIC
Thr44Pro	PATHOGENIC
Thr44Ser	PATHOGENIC
Ser45Pro	PATHOGENIC
Ala48Thr	PATHOGENIC
Tyr49His	PATHOGENIC
Tyr49Cys	PATHOGENIC
Tyr49Asp	PATHOGENIC
Ala50Thr	POPULATION
Ile50Val	NA
His51Arg	PATHOGENIC
His51Tyr	PATHOGENIC
Ala51Gly	PATHOGENIC
Ala51Pro	PATHOGENIC
His51Gln	Likely pathogenic
Pro52Ala	PATHOGENIC
Asn54Asp	Uncertain significance (VOUS)
Ala54Thr	PATHOGENIC
Glu55Lys	PATHOGENIC
Ser56Thr	NA
Ser56Leu	NA
Ile57Ser	Uncertain significance (VOUS)
Cys58Ser	PATHOGENIC
Cys58Gly	PATHOGENIC
Cys58Phe	PATHOGENIC
Cys58Trp	Likely pathogenic
Cys58Tyr	PATHOGENIC
Cys58Arg	PATHOGENIC
Cys59Tyr	PATHOGENIC
Cys59Ser	NA
Cys59Arg	PATHOGENIC
Cys59Phe	PATHOGENIC
Cys62Tyr	PATHOGENIC
Cys62Gly	PATHOGENIC
Gly65Glu	PATHOGENIC
Pro66Arg	POPULATION
Ala66Thr	PATHOGENIC
Pro66Leu	BENIGN
Thr66Ile	Likely pathogenic
Tyr67Ser	PATHOGENIC
Tyr67Cys	PATHOGENIC
Pro68Leu	POPULATION
Leu68Phe	PATHOGENIC
Leu70Ser	POPULATION
Gly70Arg	PATHOGENIC
Thr70Met	PATHOGENIC
Asp71Glu	Likely pathogenic
Ser72Pro	PATHOGENIC
Cys72Gly	Pathogenic
Glu72Gly	POPULATION
Cys72Tyr	PATHOGENIC
Cys72Phe	PATHOGENIC
Cys72Arg	PATHOGENIC
Cys72Trp	PATHOGENIC
Cys72Ser	PATHOGENIC
Pro72Ser	POPULATION
Pro73Ser	NA
Cys73Arg	POPULATION
Cys74Arg	PATHOGENIC
Val74Gly	PATHOGENIC
Val74Met	PATHOGENIC

Val74Leu	PATHOGENIC
Cys74Tyr	PATHOGENIC
Cys75Arg	PATHOGENIC
Cys75Tyr	PATHOGENIC
Cys75Ser	PATHOGENIC
Thr75Ala	PATHOGENIC
Pro75Leu	NA
Pro75Arg	NA
Ala75Thr	POPULATION
Ala76Thr	PATHOGENIC
Glu76Asp	PATHOGENIC
Arg78Pro	PATHOGENIC
Thr79Met	PATHOGENIC
Ser79Pro	PATHOGENIC
Thr79Lys	Likely pathogenic
Phe80Ser	PATHOGENIC
Phe80Leu	PATHOGENIC
Thr80Ile	POPULATION
Phe80Val	PATHOGENIC
Cys81Ser	PATHOGENIC
Thr81Asn	PATHOGENIC
Val81Ala	POPULATION
Cys81Arg	PATHOGENIC
Cys81Trp	PATHOGENIC
Cys81Tyr	PATHOGENIC
Cys81Phe	PATHOGENIC
Arg82Gly	Uncertain significance (VOUS)
Cys82Trp	PATHOGENIC
Cys82Phe	PATHOGENIC
Cys82Tyr	PATHOGENIC
Gly83Asp	POPULATION
Cys84Arg	PATHOGENIC
Cys84Tyr	PATHOGENIC
Cys84Ser	PATHOGENIC
Glu85Asp	NA
Asn85Lys	PATHOGENIC
Asn85Ile	PATHOGENIC
His86Pro	PATHOGENIC
His86Tyr	PATHOGENIC
His86Leu	PATHOGENIC
Leu87Pro	PATHOGENIC
Leu87Phe	POPULATION
Lys87Glu	POPULATION
Gly87Ser	NA
Gln88Glu	POPULATION
Ser88Pro	PATHOGENIC
Phe89Ser	PATHOGENIC
Phe89Leu	PATHOGENIC
Phe89Val	PATHOGENIC
Cys90Ser	PATHOGENIC
Cys90Gly	PATHOGENIC
Cys90Tyr	PATHOGENIC
Ser90Ala	PATHOGENIC
Thr90Asn	PATHOGENIC
Thr90Ile	NA
Thr90Pro	NA
Arg90Trp	POPULATION
Ser90Pro	PATHOGENIC
Cys90Arg	PATHOGENIC
Tyr92His	POPULATION
Ser92Asn	POPULATION
Cys93Trp	PATHOGENIC
Cys93Ser	PATHOGENIC
Cys93Arg	PATHOGENIC
Asn94Lys	PATHOGENIC
Ser94Cys	PATHOGENIC
Ser94Gly	POPULATION
Asn94Ile	PATHOGENIC
His95Pro	PATHOGENIC
His95Leu	PATHOGENIC
Ala95Thr	POPULATION
Glu95Gly	POPULATION
His95Tyr	NA
Leu96Phe	POPULATION
Leu96Pro	PATHOGENIC
Cys96Tyr	PATHOGENIC
Arg97Gln	PATHOGENIC
Phe98Leu	PATHOGENIC
Phe98Cys	PATHOGENIC
His98Asn	NA
Phe98Ile	PATHOGENIC
Phe98Ser	PATHOGENIC
Cys99Arg	PATHOGENIC
Cys99Ser	PATHOGENIC
Cys99Tyr	PATHOGENIC
Cys99Gly	PATHOGENIC
Cys100Arg	PATHOGENIC
Ser101Asn	POPULATION
Pro102Ser	POPULATION
Asn102Asp	POPULATION
Cys102Arg	PATHOGENIC
Cys102Ser	PATHOGENIC
Cys102Trp	PATHOGENIC
Cys103Tyr	POPULATION
Ser103Cys	PATHOGENIC
Cys105Tyr	PATHOGENIC
Arg106Gln	PATHOGENIC
Ser106Pro	PATHOGENIC
Leu107Phe	POPULATION
Cys108Arg	PATHOGENIC
Cys108Tyr	PATHOGENIC
Thr109Ala	PATHOGENIC
Thr110Ile	POPULATION
Val111Leu	POPULATION

Gln111Lys	Uncertain significance (VOUS)
Arg112Pro	PATHOGENIC
Val112Leu	Likely pathogenic
His112Tyr	PATHOGENIC
Val112Met	NA
Leu113Phe	POPULATION
Thr114Ile	POPULATION
Ile114Ser	NA
Val115Ala	POPULATION
Ser115Pro	PATHOGENIC
Ser115Phe	NA
Cys116Tyr	PATHOGENIC
Cys116Phe	PATHOGENIC
Cys116Trp	PATHOGENIC
Cys117Tyr	PATHOGENIC
Cys117Arg	PATHOGENIC
Gly117Asp	POPULATION
Cys117Ser	Likely pathogenic
Cys118Tyr	PATHOGENIC
Thr118Ala	PATHOGENIC
Met118Val	POPULATION
Cys118Arg	PATHOGENIC
Val119Gly	NA
Val119Ala	NA
Arg121Trp	Likely pathogenic
Ser121Cys	POPULATION
Arg121Pro	NA
Arg121Gln	NA
Arg121Gly	NA
Asp122His	Likely pathogenic
Asp122Glu	Uncertain significance (VOUS)
Thr123Ile	NA
Val124Gly	NA
Val124Ala	NA
Val124Met	NA
Cys125Arg	Likely pathogenic
Cys125Trp	PATHOGENIC
Cys125Tyr	PATHOGENIC
Cys125Phe	PATHOGENIC
Gly126Asp	POPULATION
His126Arg	POPULATION
His126Tyr	POPULATION
Cys127Tyr	PATHOGENIC
Cys127Arg	PATHOGENIC
Phe129Leu	POPULATION
Asn130Lys	Likely pathogenic
Gln131Glu	NA
Arg133Trp	NA
Arg133Gln	NA
His134Pro	Uncertain significance (VOUS)
Tyr135His	NA
Glu135Val	POPULATION
Tyr135Cys	Likely pathogenic
Ser137Gly	NA
Glu138Gly	NA
Glu138Ala	Likely benign
Ser140Thr	POPULATION
Phe141Leu	PATHOGENIC
Phe141Ser	PATHOGENIC
Phe141Cys	PATHOGENIC
Phe141Ile	PATHOGENIC
Cys142Phe	POPULATION
Cys143Trp	Likely pathogenic
Cys143Arg	PATHOGENIC
Ser145Asn	POPULATION
Asn145Asp	NA
Asn145Ser	NA
Cys146Tyr	NA
Leu150Phe	POPULATION
Asn151Ser	NA
Gly152Ala	NA
Thr153Ile	POPULATION
Leu153Val	POPULATION
Val154Met	NA
Val154Leu	NA
His155Tyr	PATHOGENIC
Ile156Asn	PATHOGENIC
Leu156Phe	POPULATION
Val159Asp	PATHOGENIC
Arg160His	NA
Arg160Leu	NA
Arg160Ser	NA
Arg160Cys	NA
Lys161Arg	NA
Ser162Cys	NA
Ser162Ala	NA
Pro163Ser	NA
Asp164Asn	POPULATION
Glu164Lys	NA
Pro167Leu	NA
Pro167Ala	NA
Ser168Ala	NA
Pro169Thr	NA
His169Arg	NA
His169Tyr	NA
Pro169His	NA
His170Tyr	NA
Pro171Arg	NA
Phe172Leu	NA
Pro177Ser	NA
Glu178Val	NA
Glu178Lys	NA
Ala179Thr	NA
Ala179Ser	NA

Ala179Glu	NA
Ala179Val	NA
Leu182Val	NA
Ser183Thr	NA
Leu185Arg	NA
Cys185Phe	POPULATION
Pro186Ser	NA
Phe186Leu	BENIGN
Pro187His	NA
Pro187Leu	NA
Pro187Ala	NA
Ser188Asn	NA
Arg189Cys	NA
Glu190Lys	NA
Phe190Leu	NA
Phe190Val	NA
Gly191Asp	NA
Thr192Met	NA
Leu196Val	NA
Leu198Val	NA
Ile199Asn	PATHOGENIC
Ile199Thr	NA
Arg201Ser	NA
Arg201Cys	NA
Arg201Gly	NA
Arg201His	NA
Val202Asp	PATHOGENIC
His203Asn	NA
Thr205Asn	NA
Glu206Asp	NA
Asp207Asn	NA
Arg207Cys	NA
Arg207His	NA
His208Leu	NA
Phe209Tyr	NA
Gly214Arg	NA
Pro215Thr	NA
Trp215Cys	NA
Gly216Glu	NA
Pro217Leu	NA
Cys217Tyr	NA
Arg218Lys	NA
Arg219His	NA
Arg219Leu	NA
Arg219Cys	NA
Phe220Ser	NA
Phe220Leu	NA
Leu222Phe	NA
Phe222Ser	NA
Ser226Cys	NA
Ala226Thr	NA
Phe229Leu	NA
Ile230Val	NA
Gly231Val	NA
Met233Leu	NA
Arg235His	NA
Gln237Arg	NA
Arg238Gln	NA
Ser241Phe	NA
His242Arg	POPULATION
His242Tyr	POPULATION
Phe245Leu	POPULATION
Glu248Lys	NA
Glu248Asp	NA
Gly250Arg	NA
Leu251Phe	NA
Glu251Val	POPULATION
Pro253Thr	NA
Pro253Ala	NA
Glu253Lys	NA
Glu254Ala	NA
Glu255Gln	NA
Lys255Glu	NA
Gly256Glu	NA
Gly256Val	NA
Gly257Ala	NA
Gly257Arg	NA
Ala257Pro	NA
Ala257Val	NA
Gly259Arg	NA
Gly259Glu	NA
Leu259Ile	NA
Leu259Arg	NA
Gly260Val	NA
Thr262Ala	NA
Pro266Thr	NA
Pro266Ala	NA
Leu267Val	NA
Arg268Gln	NA
Trp269Arg	NA
Arg269Lys	BENIGN
Asn270Lys	NA
Asn270Asp	NA
Pro271Ser	NA
Pro271Ala	NA
Ser272Gly	NA
Phe273Leu	NA
Ser274Gly	NA
Gly274Arg	NA
Pro275Ser	NA
Gly277Glu	BENIGN
Thr280Ser	NA
Thr280Asn	NA
Pro281Arg	NA

Ser282Leu	NA
Arg284His	NA
Arg284Cys	NA
Arg284Leu	NA
Trp285Arg	NA
Phe285Leu	NA
Ser286Ile	NA
Pro287Ala	NA
Val288Leu	NA
Trp288Ser	NA
Val288Met	NA
Pro289Leu	NA
Ala289Thr	NA
Ser290Arg	NA
Phe293Leu	NA
Pro293Arg	NA
Trp295Cys	NA
Ser296Thr	NA
Arg298Gly	NA
Thr298Ala	NA
Pro301Ser	NA
Pro301His	NA
Gly302Ser	NA
Cys304Ser	NA
Pro305Arg	NA
Asn306Lys	NA
Pro310Leu	NA
Pro310Ser	NA
Arg311Cys	NA
Arg311His	NA
Arg312Lys	NA
Glu313Lys	NA
Glu313Gln	NA
Ala315Thr	NA
Pro317Ala	NA
Tyr318Cys	NA
Tyr318Phe	NA
Gly320Arg	NA
Gly320Glu	BENIGN
Asp322Glu	NA
Pro323Ser	NA
Ile324Val	NA
Ile324Asn	NA
Ala326Ser	NA
Leu329Phe	NA
Ala330Thr	NA
Ala330Val	NA
Pro333Ser	NA
Ile334Val	NA
Pro335His	NA
Pro335Arg	NA
Asn336Lys	NA
Asn336Asp	NA
Leu338Phe	NA
Lys340Arg	NA
Lys340Glu	NA
Glu342Asp	NA
His346Arg	NA
His346Gln	NA
Thr353Ala	NA
Thr353Pro	NA
Pro356Ser	NA
Pro356Ala	NA
Thr358Ala	NA
Glu364Val	NA
Pro368Ser	NA
Pro368Ala	NA
Ser381Asn	NA
Asp382His	NA
His383Asn	NA
Asp386Glu	NA
Leu388Arg	NA
Leu390Pro	NA
Arg394His	NA
Cys395Arg	NA
Cys395Trp	NA
Leu396Arg	NA
Glu398Lys	NA
Ser402Gly	NA
Ala405Gly	NA
Arg408Lys	NA
Arg409Trp	NA
Arg409Leu	NA
Pro412Leu	NA
Arg413Gly	NA
Arg413Gln	NA
Glu415Lys	NA
Ala416Gly	NA
Thr417Ala	NA
Glu419Gly	NA
Gly422Arg	NA
Arg423His	NA
Arg426His	NA
Gly432Val	NA
Glu435Gly	NA
Asp436Gly	NA
Glu438Lys	NA
Glu438Ala	NA
Glu438Gln	NA
Cys442Tyr	NA
Gly443Arg	NA
Gly443Val	NA
Pro448Leu	NA
Ser452Gly	NA

Leu453Pro | NA

Tables with applicable missense mutations to TNFRSF1A

Iteration number	Filename	Chain	Residue index in chain	New residue
34	ltnr3.TNFA	R	0	TYR
34	ltnr3.TNFA	T	0	TYR
34	ltnr3.TNFA	S	0	TYR
35	ltnr3.TNFA	R	0	PRO
35	ltnr3.TNFA	T	0	PRO
35	ltnr3.TNFA	S	0	PRO
36	ltnr3.TNFA	R	0	SER
36	ltnr3.TNFA	T	0	SER
36	ltnr3.TNFA	S	0	SER
37	ltnr3.TNFA	R	1	PRO
37	ltnr3.TNFA	T	1	PRO
37	ltnr3.TNFA	S	1	PRO
38	ltnr3.TNFA	R	4	THR
38	ltnr3.TNFA	T	4	THR
38	ltnr3.TNFA	S	4	THR
39	ltnr3.TNFA	R	5	HIS
39	ltnr3.TNFA	T	5	HIS
39	ltnr3.TNFA	S	5	HIS
40	ltnr3.TNFA	R	5	CYS
40	ltnr3.TNFA	T	5	CYS
40	ltnr3.TNFA	S	5	CYS
41	ltnr3.TNFA	R	5	ASP
41	ltnr3.TNFA	T	5	ASP
41	ltnr3.TNFA	S	5	ASP
42	ltnr3.TNFA	R	6	THR
42	ltnr3.TNFA	T	6	THR
42	ltnr3.TNFA	S	6	THR
43	ltnr3.TNFA	R	6	VAL
43	ltnr3.TNFA	T	6	VAL
43	ltnr3.TNFA	S	6	VAL
44	ltnr3.TNFA	R	7	ARG
44	ltnr3.TNFA	T	7	ARG
44	ltnr3.TNFA	S	7	ARG
45	ltnr3.TNFA	R	7	TYR
45	ltnr3.TNFA	T	7	TYR
45	ltnr3.TNFA	S	7	TYR
46	ltnr3.TNFA	R	7	GLY
46	ltnr3.TNFA	T	7	GLY
46	ltnr3.TNFA	S	7	GLY
47	ltnr3.TNFA	R	7	PRO
47	ltnr3.TNFA	T	7	PRO
47	ltnr3.TNFA	S	7	PRO
48	ltnr3.TNFA	R	7	GLN
48	ltnr3.TNFA	T	7	GLN
48	ltnr3.TNFA	S	7	GLN
49	ltnr3.TNFA	R	8	ALA
49	ltnr3.TNFA	T	8	ALA
49	ltnr3.TNFA	S	8	ALA
50	ltnr3.TNFA	R	10	ASP
50	ltnr3.TNFA	T	10	ASP
50	ltnr3.TNFA	S	10	ASP
51	ltnr3.TNFA	R	10	THR
51	ltnr3.TNFA	T	10	THR
51	ltnr3.TNFA	S	10	THR
52	ltnr3.TNFA	R	11	LYS
52	ltnr3.TNFA	T	11	LYS
52	ltnr3.TNFA	S	11	LYS
53	ltnr3.TNFA	R	12	THR
53	ltnr3.TNFA	T	12	THR
53	ltnr3.TNFA	S	12	THR
54	ltnr3.TNFA	R	12	LEU
54	ltnr3.TNFA	T	12	LEU
54	ltnr3.TNFA	S	12	LEU
55	ltnr3.TNFA	R	13	SER
55	ltnr3.TNFA	T	13	SER
55	ltnr3.TNFA	S	13	SER
56	ltnr3.TNFA	R	14	SER
56	ltnr3.TNFA	T	14	SER
56	ltnr3.TNFA	S	14	SER
57	ltnr3.TNFA	R	14	GLY
57	ltnr3.TNFA	T	14	GLY
57	ltnr3.TNFA	S	14	GLY
58	ltnr3.TNFA	R	14	PHE
58	ltnr3.TNFA	T	14	PHE
58	ltnr3.TNFA	S	14	PHE
59	ltnr3.TNFA	R	14	TRP
59	ltnr3.TNFA	T	14	TRP
59	ltnr3.TNFA	S	14	TRP
60	ltnr3.TNFA	R	14	TYR
60	ltnr3.TNFA	T	14	TYR
60	ltnr3.TNFA	S	14	TYR
61	ltnr3.TNFA	R	14	ARG
61	ltnr3.TNFA	T	14	ARG
61	ltnr3.TNFA	S	14	ARG
62	ltnr3.TNFA	R	15	TYR
62	ltnr3.TNFA	T	15	TYR
62	ltnr3.TNFA	S	15	TYR
63	ltnr3.TNFA	R	15	SER
63	ltnr3.TNFA	T	15	SER
63	ltnr3.TNFA	S	15	SER
64	ltnr3.TNFA	R	15	ARG
64	ltnr3.TNFA	T	15	ARG
64	ltnr3.TNFA	S	15	ARG
65	ltnr3.TNFA	R	15	PHE
65	ltnr3.TNFA	T	15	PHE
65	ltnr3.TNFA	S	15	PHE
66	ltnr3.TNFA	R	18	TYR
66	ltnr3.TNFA	T	18	TYR

66	ltnr3.TNFA	S	18	TYR
67	ltnr3.TNFA	R	18	GLY
67	ltnr3.TNFA	T	18	GLY
67	ltnr3.TNFA	S	18	GLY
68	ltnr3.TNFA	R	21	GLU
68	ltnr3.TNFA	T	21	GLU
68	ltnr3.TNFA	S	21	GLU
69	ltnr3.TNFA	R	22	ARG
69	ltnr3.TNFA	T	22	ARG
69	ltnr3.TNFA	S	22	ARG
70	ltnr3.TNFA	R	22	THR
70	ltnr3.TNFA	T	22	THR
70	ltnr3.TNFA	S	22	THR
71	ltnr3.TNFA	R	22	LEU
71	ltnr3.TNFA	T	22	LEU
71	ltnr3.TNFA	S	22	LEU
72	ltnr3.TNFA	R	22	ILE
72	ltnr3.TNFA	T	22	ILE
72	ltnr3.TNFA	S	22	ILE
73	ltnr3.TNFA	R	23	SER
73	ltnr3.TNFA	T	23	SER
73	ltnr3.TNFA	S	23	SER
74	ltnr3.TNFA	R	23	CYS
74	ltnr3.TNFA	T	23	CYS
74	ltnr3.TNFA	S	23	CYS
75	ltnr3.TNFA	R	24	LEU
75	ltnr3.TNFA	T	24	LEU
75	ltnr3.TNFA	S	24	LEU
76	ltnr3.TNFA	R	24	PHE
76	ltnr3.TNFA	T	24	PHE
76	ltnr3.TNFA	S	24	PHE
77	ltnr3.TNFA	R	26	SER
77	ltnr3.TNFA	T	26	SER
77	ltnr3.TNFA	S	26	SER
78	ltnr3.TNFA	R	26	ARG
78	ltnr3.TNFA	T	26	ARG
78	ltnr3.TNFA	S	26	ARG
79	ltnr3.TNFA	R	26	MET
79	ltnr3.TNFA	T	26	MET
79	ltnr3.TNFA	S	26	MET
80	ltnr3.TNFA	R	27	GLU
80	ltnr3.TNFA	T	27	GLU
80	ltnr3.TNFA	S	27	GLU
81	ltnr3.TNFA	R	28	PRO
81	ltnr3.TNFA	T	28	PRO
81	ltnr3.TNFA	S	28	PRO
82	ltnr3.TNFA	R	28	GLY
82	ltnr3.TNFA	T	28	GLY
82	ltnr3.TNFA	S	28	GLY
83	ltnr3.TNFA	R	28	GLY
83	ltnr3.TNFA	T	28	GLY
83	ltnr3.TNFA	S	28	GLY
84	ltnr3.TNFA	R	28	TYR
84	ltnr3.TNFA	T	28	TYR
84	ltnr3.TNFA	S	28	TYR
85	ltnr3.TNFA	R	28	PHE
85	ltnr3.TNFA	T	28	PHE
85	ltnr3.TNFA	S	28	PHE
86	ltnr3.TNFA	R	28	ARG
86	ltnr3.TNFA	T	28	ARG
86	ltnr3.TNFA	S	28	ARG
87	ltnr3.TNFA	R	28	TRP
87	ltnr3.TNFA	T	28	TRP
87	ltnr3.TNFA	S	28	TRP
88	ltnr3.TNFA	R	28	SER
88	ltnr3.TNFA	T	28	SER
88	ltnr3.TNFA	S	28	SER
89	ltnr3.TNFA	R	28	SER
89	ltnr3.TNFA	T	28	SER
89	ltnr3.TNFA	S	28	SER
90	ltnr3.TNFA	R	29	SER
90	ltnr3.TNFA	T	29	SER
90	ltnr3.TNFA	S	29	SER
91	ltnr3.TNFA	R	29	ARG
91	ltnr3.TNFA	T	29	ARG
91	ltnr3.TNFA	S	29	ARG
92	ltnr3.TNFA	R	30	ARG
92	ltnr3.TNFA	T	30	ARG
92	ltnr3.TNFA	S	30	ARG
93	ltnr3.TNFA	R	30	GLY
93	ltnr3.TNFA	T	30	GLY
93	ltnr3.TNFA	S	30	GLY
94	ltnr3.TNFA	R	30	MET
94	ltnr3.TNFA	T	30	MET
94	ltnr3.TNFA	S	30	MET
95	ltnr3.TNFA	R	30	LEU
95	ltnr3.TNFA	T	30	LEU
95	ltnr3.TNFA	S	30	LEU
96	ltnr3.TNFA	R	30	TYR
96	ltnr3.TNFA	T	30	TYR
96	ltnr3.TNFA	S	30	TYR
97	ltnr3.TNFA	R	31	ARG
97	ltnr3.TNFA	T	31	ARG
97	ltnr3.TNFA	S	31	ARG
98	ltnr3.TNFA	R	31	TYR
98	ltnr3.TNFA	T	31	TYR
98	ltnr3.TNFA	S	31	TYR
99	ltnr3.TNFA	R	31	SER
99	ltnr3.TNFA	T	31	SER
99	ltnr3.TNFA	S	31	SER
100	ltnr3.TNFA	R	31	ALA
100	ltnr3.TNFA	T	31	ALA
100	ltnr3.TNFA	S	31	ALA
101	ltnr3.TNFA	R	31	LEU

101	1tnr3.TNFA	T	31	LEU
101	1tnr3.TNFA	S	31	LEU
102	1tnr3.TNFA	R	31	ARG
102	1tnr3.TNFA	T	31	ARG
102	1tnr3.TNFA	S	31	ARG
103	1tnr3.TNFA	R	31	THR
103	1tnr3.TNFA	T	31	THR
103	1tnr3.TNFA	S	31	THR
104	1tnr3.TNFA	R	32	THR
104	1tnr3.TNFA	T	32	THR
104	1tnr3.TNFA	S	32	THR
105	1tnr3.TNFA	R	32	ASP
105	1tnr3.TNFA	T	32	ASP
105	1tnr3.TNFA	S	32	ASP
106	1tnr3.TNFA	R	34	PRO
106	1tnr3.TNFA	T	34	PRO
106	1tnr3.TNFA	S	34	PRO
107	1tnr3.TNFA	R	35	MET
107	1tnr3.TNFA	T	35	MET
107	1tnr3.TNFA	S	35	MET
108	1tnr3.TNFA	R	35	PRO
108	1tnr3.TNFA	T	35	PRO
108	1tnr3.TNFA	S	35	PRO
109	1tnr3.TNFA	R	35	LYS
109	1tnr3.TNFA	T	35	LYS
109	1tnr3.TNFA	S	35	LYS
110	1tnr3.TNFA	R	36	SER
110	1tnr3.TNFA	T	36	SER
110	1tnr3.TNFA	S	36	SER
111	1tnr3.TNFA	R	36	LEU
111	1tnr3.TNFA	T	36	LEU
111	1tnr3.TNFA	S	36	LEU
112	1tnr3.TNFA	R	36	ILE
112	1tnr3.TNFA	T	36	ILE
112	1tnr3.TNFA	S	36	ILE
113	1tnr3.TNFA	R	36	VAL
113	1tnr3.TNFA	T	36	VAL
113	1tnr3.TNFA	S	36	VAL
114	1tnr3.TNFA	R	37	SER
114	1tnr3.TNFA	T	37	SER
114	1tnr3.TNFA	S	37	SER
115	1tnr3.TNFA	R	37	ASN
115	1tnr3.TNFA	T	37	ASN
115	1tnr3.TNFA	S	37	ASN
116	1tnr3.TNFA	R	37	ALA
116	1tnr3.TNFA	T	37	ALA
116	1tnr3.TNFA	S	37	ALA
117	1tnr3.TNFA	R	37	ARG
117	1tnr3.TNFA	T	37	ARG
117	1tnr3.TNFA	S	37	ARG
118	1tnr3.TNFA	R	37	TRP
118	1tnr3.TNFA	T	37	TRP
118	1tnr3.TNFA	S	37	TRP
119	1tnr3.TNFA	R	37	TYR
119	1tnr3.TNFA	T	37	TYR
119	1tnr3.TNFA	S	37	TYR
120	1tnr3.TNFA	R	37	PHE
120	1tnr3.TNFA	T	37	PHE
120	1tnr3.TNFA	S	37	PHE
121	1tnr3.TNFA	R	38	GLY
121	1tnr3.TNFA	T	38	GLY
121	1tnr3.TNFA	S	38	GLY
122	1tnr3.TNFA	R	38	TRP
122	1tnr3.TNFA	T	38	TRP
122	1tnr3.TNFA	S	38	TRP
123	1tnr3.TNFA	R	38	PHE
123	1tnr3.TNFA	T	38	PHE
123	1tnr3.TNFA	S	38	PHE
124	1tnr3.TNFA	R	38	TYR
124	1tnr3.TNFA	T	38	TYR
124	1tnr3.TNFA	S	38	TYR
125	1tnr3.TNFA	R	39	ASP
125	1tnr3.TNFA	T	39	ASP
125	1tnr3.TNFA	S	39	ASP
126	1tnr3.TNFA	R	40	ARG
126	1tnr3.TNFA	T	40	ARG
126	1tnr3.TNFA	S	40	ARG
127	1tnr3.TNFA	R	40	TYR
127	1tnr3.TNFA	T	40	TYR
127	1tnr3.TNFA	S	40	TYR
128	1tnr3.TNFA	R	40	SER
128	1tnr3.TNFA	T	40	SER
128	1tnr3.TNFA	S	40	SER
129	1tnr3.TNFA	R	41	ASP
129	1tnr3.TNFA	T	41	ASP
129	1tnr3.TNFA	S	41	ASP
130	1tnr3.TNFA	R	41	LYS
130	1tnr3.TNFA	T	41	LYS
130	1tnr3.TNFA	S	41	LYS
131	1tnr3.TNFA	R	41	ILE
131	1tnr3.TNFA	T	41	ILE
131	1tnr3.TNFA	S	41	ILE
132	1tnr3.TNFA	R	42	PRO
132	1tnr3.TNFA	T	42	PRO
132	1tnr3.TNFA	S	42	PRO
133	1tnr3.TNFA	R	42	TYR
133	1tnr3.TNFA	T	42	TYR
133	1tnr3.TNFA	S	42	TYR
134	1tnr3.TNFA	R	42	LEU
134	1tnr3.TNFA	T	42	LEU
134	1tnr3.TNFA	S	42	LEU
135	1tnr3.TNFA	R	43	PRO
135	1tnr3.TNFA	T	43	PRO
135	1tnr3.TNFA	S	43	PRO

136	1tnr3.TNFA	R	43	PHE
136	1tnr3.TNFA	T	43	PHE
136	1tnr3.TNFA	S	43	PHE
137	1tnr3.TNFA	R	43	GLU
137	1tnr3.TNFA	T	43	GLU
137	1tnr3.TNFA	S	43	GLU
138	1tnr3.TNFA	R	43	SER
138	1tnr3.TNFA	T	43	SER
138	1tnr3.TNFA	S	43	SER
139	1tnr3.TNFA	R	44	GLU
139	1tnr3.TNFA	T	44	GLU
139	1tnr3.TNFA	S	44	GLU
140	1tnr3.TNFA	R	44	PRO
140	1tnr3.TNFA	T	44	PRO
140	1tnr3.TNFA	S	44	PRO
141	1tnr3.TNFA	R	45	SER
141	1tnr3.TNFA	T	45	SER
141	1tnr3.TNFA	S	45	SER
142	1tnr3.TNFA	R	45	LEU
142	1tnr3.TNFA	T	45	LEU
142	1tnr3.TNFA	S	45	LEU
143	1tnr3.TNFA	R	45	VAL
143	1tnr3.TNFA	T	45	VAL
143	1tnr3.TNFA	S	45	VAL
144	1tnr3.TNFA	R	46	SER
144	1tnr3.TNFA	T	46	SER
144	1tnr3.TNFA	S	46	SER
145	1tnr3.TNFA	R	46	GLY
145	1tnr3.TNFA	T	46	GLY
145	1tnr3.TNFA	S	46	GLY
146	1tnr3.TNFA	R	46	TYR
146	1tnr3.TNFA	T	46	TYR
146	1tnr3.TNFA	S	46	TYR
147	1tnr3.TNFA	R	46	ALA
147	1tnr3.TNFA	T	46	ALA
147	1tnr3.TNFA	S	46	ALA
148	1tnr3.TNFA	R	46	ASN
148	1tnr3.TNFA	T	46	ASN
148	1tnr3.TNFA	S	46	ASN
149	1tnr3.TNFA	R	46	ILE
149	1tnr3.TNFA	T	46	ILE
149	1tnr3.TNFA	S	46	ILE
150	1tnr3.TNFA	R	46	PRO
150	1tnr3.TNFA	T	46	PRO
150	1tnr3.TNFA	S	46	PRO
151	1tnr3.TNFA	R	46	TRP
151	1tnr3.TNFA	T	46	TRP
151	1tnr3.TNFA	S	46	TRP
152	1tnr3.TNFA	R	46	PRO
152	1tnr3.TNFA	T	46	PRO
152	1tnr3.TNFA	S	46	PRO
153	1tnr3.TNFA	R	46	ARG
153	1tnr3.TNFA	T	46	ARG
153	1tnr3.TNFA	S	46	ARG
154	1tnr3.TNFA	R	48	HIS
154	1tnr3.TNFA	T	48	HIS
154	1tnr3.TNFA	S	48	HIS
155	1tnr3.TNFA	R	48	ASN
155	1tnr3.TNFA	T	48	ASN
155	1tnr3.TNFA	S	48	ASN
156	1tnr3.TNFA	R	49	TRP
156	1tnr3.TNFA	T	49	TRP
156	1tnr3.TNFA	S	49	TRP
157	1tnr3.TNFA	R	49	SER
157	1tnr3.TNFA	T	49	SER
157	1tnr3.TNFA	S	49	SER
158	1tnr3.TNFA	R	49	ARG
158	1tnr3.TNFA	T	49	ARG
158	1tnr3.TNFA	S	49	ARG
159	1tnr3.TNFA	R	50	LYS
159	1tnr3.TNFA	T	50	LYS
159	1tnr3.TNFA	S	50	LYS
160	1tnr3.TNFA	R	50	CYS
160	1tnr3.TNFA	T	50	CYS
160	1tnr3.TNFA	S	50	CYS
161	1tnr3.TNFA	R	50	GLY
161	1tnr3.TNFA	T	50	GLY
161	1tnr3.TNFA	S	50	GLY
162	1tnr3.TNFA	R	50	ILE
162	1tnr3.TNFA	T	50	ILE
162	1tnr3.TNFA	S	50	ILE
163	1tnr3.TNFA	R	51	PRO
163	1tnr3.TNFA	T	51	PRO
163	1tnr3.TNFA	S	51	PRO
164	1tnr3.TNFA	R	51	LEU
164	1tnr3.TNFA	T	51	LEU
164	1tnr3.TNFA	S	51	LEU
165	1tnr3.TNFA	R	51	THR
165	1tnr3.TNFA	T	51	THR
165	1tnr3.TNFA	S	51	THR
166	1tnr3.TNFA	R	51	GLY
166	1tnr3.TNFA	T	51	GLY
166	1tnr3.TNFA	S	51	GLY
167	1tnr3.TNFA	R	51	TYR
167	1tnr3.TNFA	T	51	TYR
167	1tnr3.TNFA	S	51	TYR
168	1tnr3.TNFA	R	52	PHE
168	1tnr3.TNFA	T	52	PHE
168	1tnr3.TNFA	S	52	PHE
169	1tnr3.TNFA	R	52	PRO
169	1tnr3.TNFA	T	52	PRO
169	1tnr3.TNFA	S	52	PRO
170	1tnr3.TNFA	R	52	TYR
170	1tnr3.TNFA	T	52	TYR

170	ltnr3.TNFA	S	52	TYR
171	ltnr3.TNFA	R	53	GLN
171	ltnr3.TNFA	T	53	GLN
171	ltnr3.TNFA	S	53	GLN
172	ltnr3.TNFA	R	54	LEU
172	ltnr3.TNFA	T	54	LEU
172	ltnr3.TNFA	S	54	LEU
173	ltnr3.TNFA	R	54	CYS
173	ltnr3.TNFA	T	54	CYS
173	ltnr3.TNFA	S	54	CYS
174	ltnr3.TNFA	R	54	ASN
174	ltnr3.TNFA	T	54	ASN
174	ltnr3.TNFA	S	54	ASN
175	ltnr3.TNFA	R	54	ILE
175	ltnr3.TNFA	T	54	ILE
175	ltnr3.TNFA	S	54	ILE
176	ltnr3.TNFA	R	54	SER
176	ltnr3.TNFA	T	54	SER
176	ltnr3.TNFA	S	54	SER
177	ltnr3.TNFA	R	55	ARG
177	ltnr3.TNFA	T	55	ARG
177	ltnr3.TNFA	S	55	ARG
178	ltnr3.TNFA	R	55	SER
178	ltnr3.TNFA	T	55	SER
178	ltnr3.TNFA	S	55	SER
179	ltnr3.TNFA	R	55	TYR
179	ltnr3.TNFA	T	55	TYR
179	ltnr3.TNFA	S	55	TYR
180	ltnr3.TNFA	R	55	GLY
180	ltnr3.TNFA	T	55	GLY
180	ltnr3.TNFA	S	55	GLY
181	ltnr3.TNFA	R	56	ARG
181	ltnr3.TNFA	T	56	ARG
181	ltnr3.TNFA	S	56	ARG
182	ltnr3.TNFA	R	57	ASN
182	ltnr3.TNFA	T	57	ASN
182	ltnr3.TNFA	S	57	ASN
183	ltnr3.TNFA	R	58	SER
183	ltnr3.TNFA	T	58	SER
183	ltnr3.TNFA	S	58	SER
184	ltnr3.TNFA	R	58	ASP
184	ltnr3.TNFA	T	58	ASP
184	ltnr3.TNFA	S	58	ASP
185	ltnr3.TNFA	R	58	ARG
185	ltnr3.TNFA	T	58	ARG
185	ltnr3.TNFA	S	58	ARG
186	ltnr3.TNFA	R	58	SER
186	ltnr3.TNFA	T	58	SER
186	ltnr3.TNFA	S	58	SER
187	ltnr3.TNFA	R	58	TRP
187	ltnr3.TNFA	T	58	TRP
187	ltnr3.TNFA	S	58	TRP
188	ltnr3.TNFA	R	59	TYR
188	ltnr3.TNFA	T	59	TYR
188	ltnr3.TNFA	S	59	TYR
189	ltnr3.TNFA	R	59	CYS
189	ltnr3.TNFA	T	59	CYS
189	ltnr3.TNFA	S	59	CYS
190	ltnr3.TNFA	R	61	TYR
190	ltnr3.TNFA	T	61	TYR
190	ltnr3.TNFA	S	61	TYR
191	ltnr3.TNFA	R	62	GLN
191	ltnr3.TNFA	T	62	GLN
191	ltnr3.TNFA	S	62	GLN
192	ltnr3.TNFA	R	62	PRO
192	ltnr3.TNFA	T	62	PRO
192	ltnr3.TNFA	S	62	PRO
193	ltnr3.TNFA	R	63	PHE
193	ltnr3.TNFA	T	63	PHE
193	ltnr3.TNFA	S	63	PHE
194	ltnr3.TNFA	R	64	ARG
194	ltnr3.TNFA	T	64	ARG
194	ltnr3.TNFA	S	64	ARG
195	ltnr3.TNFA	R	64	TYR
195	ltnr3.TNFA	T	64	TYR
195	ltnr3.TNFA	S	64	TYR
196	ltnr3.TNFA	R	65	ALA
196	ltnr3.TNFA	T	65	ALA
196	ltnr3.TNFA	S	65	ALA
197	ltnr3.TNFA	R	66	ILE
197	ltnr3.TNFA	T	66	ILE
197	ltnr3.TNFA	S	66	ILE
198	ltnr3.TNFA	R	67	LEU
198	ltnr3.TNFA	T	67	LEU
198	ltnr3.TNFA	S	67	LEU
199	ltnr3.TNFA	R	67	LYS
199	ltnr3.TNFA	T	67	LYS
199	ltnr3.TNFA	S	67	LYS
200	ltnr3.TNFA	R	68	PRO
200	ltnr3.TNFA	T	68	PRO
200	ltnr3.TNFA	S	68	PRO
201	ltnr3.TNFA	R	68	LEU
201	ltnr3.TNFA	T	68	LEU
201	ltnr3.TNFA	S	68	LEU
202	ltnr3.TNFA	R	68	TYR
202	ltnr3.TNFA	T	68	TYR
202	ltnr3.TNFA	S	68	TYR
203	ltnr3.TNFA	R	68	MET
203	ltnr3.TNFA	T	68	MET
203	ltnr3.TNFA	S	68	MET
204	ltnr3.TNFA	R	69	PHE
204	ltnr3.TNFA	T	69	PHE
204	ltnr3.TNFA	S	69	PHE
205	ltnr3.TNFA	R	70	ILE

205	ltnr3.TNFA	T	70	ILE
205	ltnr3.TNFA	S	70	ILE
206	ltnr3.TNFA	R	70	SER
206	ltnr3.TNFA	T	70	SER
206	ltnr3.TNFA	S	70	SER
207	ltnr3.TNFA	R	71	ALA
207	ltnr3.TNFA	T	71	ALA
207	ltnr3.TNFA	S	71	ALA
208	ltnr3.TNFA	R	71	PRO
208	ltnr3.TNFA	T	71	PRO
208	ltnr3.TNFA	S	71	PRO
209	ltnr3.TNFA	R	71	PHE
209	ltnr3.TNFA	T	71	PHE
209	ltnr3.TNFA	S	71	PHE
210	ltnr3.TNFA	R	72	TYR
210	ltnr3.TNFA	T	72	TYR
210	ltnr3.TNFA	S	72	TYR
211	ltnr3.TNFA	R	72	PHE
211	ltnr3.TNFA	T	72	PHE
211	ltnr3.TNFA	S	72	PHE
212	ltnr3.TNFA	R	72	TRP
212	ltnr3.TNFA	T	72	TRP
212	ltnr3.TNFA	S	72	TRP
213	ltnr3.TNFA	R	73	TYR
213	ltnr3.TNFA	T	73	TYR
213	ltnr3.TNFA	S	73	TYR
214	ltnr3.TNFA	R	73	ARG
214	ltnr3.TNFA	T	73	ARG
214	ltnr3.TNFA	S	73	ARG
215	ltnr3.TNFA	R	73	ASP
215	ltnr3.TNFA	T	73	ASP
215	ltnr3.TNFA	S	73	ASP
216	ltnr3.TNFA	R	73	SER
216	ltnr3.TNFA	T	73	SER
216	ltnr3.TNFA	S	73	SER
217	ltnr3.TNFA	R	74	TYR
217	ltnr3.TNFA	T	74	TYR
217	ltnr3.TNFA	S	74	TYR
218	ltnr3.TNFA	R	74	ALA
218	ltnr3.TNFA	T	74	ALA
218	ltnr3.TNFA	S	74	ALA
219	ltnr3.TNFA	R	74	VAL
219	ltnr3.TNFA	T	74	VAL
219	ltnr3.TNFA	S	74	VAL
220	ltnr3.TNFA	R	74	ARG
220	ltnr3.TNFA	T	74	ARG
220	ltnr3.TNFA	S	74	ARG
221	ltnr3.TNFA	R	75	GLY
221	ltnr3.TNFA	T	75	GLY
221	ltnr3.TNFA	S	75	GLY
222	ltnr3.TNFA	R	75	ALA
222	ltnr3.TNFA	T	75	ALA
222	ltnr3.TNFA	S	75	ALA
223	ltnr3.TNFA	R	77	TRP
223	ltnr3.TNFA	T	77	TRP
223	ltnr3.TNFA	S	77	TRP
224	ltnr3.TNFA	R	77	CYS
224	ltnr3.TNFA	T	77	CYS
224	ltnr3.TNFA	S	77	CYS
225	ltnr3.TNFA	R	77	PRO
225	ltnr3.TNFA	T	77	PRO
225	ltnr3.TNFA	S	77	PRO
226	ltnr3.TNFA	R	77	GLN
226	ltnr3.TNFA	T	77	GLN
226	ltnr3.TNFA	S	77	GLN
227	ltnr3.TNFA	R	77	GLY
227	ltnr3.TNFA	T	77	GLY
227	ltnr3.TNFA	S	77	GLY
228	ltnr3.TNFA	R	78	HIS
228	ltnr3.TNFA	T	78	HIS
228	ltnr3.TNFA	S	78	HIS
229	ltnr3.TNFA	R	78	GLU
229	ltnr3.TNFA	T	78	GLU
229	ltnr3.TNFA	S	78	GLU
230	ltnr3.TNFA	R	79	ILE
230	ltnr3.TNFA	T	79	ILE
230	ltnr3.TNFA	S	79	ILE
231	ltnr3.TNFA	R	80	GLY
231	ltnr3.TNFA	T	80	GLY
231	ltnr3.TNFA	S	80	GLY
232	ltnr3.TNFA	R	80	ALA
232	ltnr3.TNFA	T	80	ALA
232	ltnr3.TNFA	S	80	ALA
233	ltnr3.TNFA	R	80	MET
233	ltnr3.TNFA	T	80	MET
233	ltnr3.TNFA	S	80	MET
234	ltnr3.TNFA	R	81	ARG
234	ltnr3.TNFA	T	81	ARG
234	ltnr3.TNFA	S	81	ARG
235	ltnr3.TNFA	R	81	TRP
235	ltnr3.TNFA	T	81	TRP
235	ltnr3.TNFA	S	81	TRP
236	ltnr3.TNFA	R	81	TYR
236	ltnr3.TNFA	T	81	TYR
236	ltnr3.TNFA	S	81	TYR
237	ltnr3.TNFA	R	81	PHE
237	ltnr3.TNFA	T	81	PHE
237	ltnr3.TNFA	S	81	PHE
238	ltnr3.TNFA	R	82	ASP
238	ltnr3.TNFA	T	82	ASP
238	ltnr3.TNFA	S	82	ASP
239	ltnr3.TNFA	R	82	ARG
239	ltnr3.TNFA	T	82	ARG
239	ltnr3.TNFA	S	82	ARG

240	ltnr3.TNFA	R	82	TYR
240	ltnr3.TNFA	T	82	TYR
240	ltnr3.TNFA	S	82	TYR
241	ltnr3.TNFA	R	83	TYR
241	ltnr3.TNFA	T	83	TYR
241	ltnr3.TNFA	S	83	TYR
242	ltnr3.TNFA	R	83	ARG
242	ltnr3.TNFA	T	83	ARG
242	ltnr3.TNFA	S	83	ARG
243	ltnr3.TNFA	R	85	LEU
243	ltnr3.TNFA	T	85	LEU
243	ltnr3.TNFA	S	85	LEU
244	ltnr3.TNFA	R	86	LYS
244	ltnr3.TNFA	T	86	LYS
244	ltnr3.TNFA	S	86	LYS
245	ltnr3.TNFA	R	87	GLU
245	ltnr3.TNFA	T	87	GLU
245	ltnr3.TNFA	S	87	GLU
246	ltnr3.TNFA	R	89	TRP
246	ltnr3.TNFA	T	89	TRP
246	ltnr3.TNFA	S	89	TRP
247	ltnr3.TNFA	R	89	GLN
247	ltnr3.TNFA	T	89	GLN
247	ltnr3.TNFA	S	89	GLN
248	ltnr3.TNFA	R	90	PRO
248	ltnr3.TNFA	T	90	PRO
248	ltnr3.TNFA	S	90	PRO
249	ltnr3.TNFA	R	91	HIS
249	ltnr3.TNFA	T	91	HIS
249	ltnr3.TNFA	S	91	HIS
250	ltnr3.TNFA	R	91	VAL
250	ltnr3.TNFA	T	91	VAL
250	ltnr3.TNFA	S	91	VAL
251	ltnr3.TNFA	R	91	CYS
251	ltnr3.TNFA	T	91	CYS
251	ltnr3.TNFA	S	91	CYS
252	ltnr3.TNFA	R	93	GLY
252	ltnr3.TNFA	T	93	GLY
252	ltnr3.TNFA	S	93	GLY
253	ltnr3.TNFA	R	94	GLY
253	ltnr3.TNFA	T	94	GLY
253	ltnr3.TNFA	S	94	GLY
254	ltnr3.TNFA	R	94	ALA
254	ltnr3.TNFA	T	94	ALA
254	ltnr3.TNFA	S	94	ALA
255	ltnr3.TNFA	R	96	THR
255	ltnr3.TNFA	T	96	THR
255	ltnr3.TNFA	S	96	THR
256	ltnr3.TNFA	R	97	LEU
256	ltnr3.TNFA	T	97	LEU
256	ltnr3.TNFA	S	97	LEU
257	ltnr3.TNFA	R	97	SER
257	ltnr3.TNFA	T	97	SER
257	ltnr3.TNFA	S	97	SER
258	ltnr3.TNFA	R	97	CYS
258	ltnr3.TNFA	T	97	CYS
258	ltnr3.TNFA	S	97	CYS
259	ltnr3.TNFA	R	97	ILE
259	ltnr3.TNFA	T	97	ILE
259	ltnr3.TNFA	S	97	ILE
260	ltnr3.TNFA	R	98	PHE
260	ltnr3.TNFA	T	98	PHE
260	ltnr3.TNFA	S	98	PHE
261	ltnr3.TNFA	R	99	TRP
261	ltnr3.TNFA	T	99	TRP
261	ltnr3.TNFA	S	99	TRP
262	ltnr3.TNFA	R	99	ARG
262	ltnr3.TNFA	T	99	ARG
262	ltnr3.TNFA	S	99	ARG
263	ltnr3.TNFA	R	101	ASN
263	ltnr3.TNFA	T	101	ASN
263	ltnr3.TNFA	S	101	ASN
264	ltnr3.TNFA	R	101	ASP
264	ltnr3.TNFA	T	101	ASP
264	ltnr3.TNFA	S	101	ASP
265	ltnr3.TNFA	R	101	SER
265	ltnr3.TNFA	T	101	SER
265	ltnr3.TNFA	S	101	SER
266	ltnr3.TNFA	R	102	TYR
266	ltnr3.TNFA	T	102	TYR
266	ltnr3.TNFA	S	102	TYR
267	ltnr3.TNFA	R	106	PHE
267	ltnr3.TNFA	T	106	PHE
267	ltnr3.TNFA	S	106	PHE
268	ltnr3.TNFA	R	107	SER
268	ltnr3.TNFA	T	107	SER
268	ltnr3.TNFA	S	107	SER
269	ltnr3.TNFA	R	108	ALA
269	ltnr3.TNFA	T	108	ALA
269	ltnr3.TNFA	S	108	ALA
270	ltnr3.TNFA	R	109	ILE
270	ltnr3.TNFA	T	109	ILE
270	ltnr3.TNFA	S	109	ILE
271	ltnr3.TNFA	R	109	VAL
271	ltnr3.TNFA	T	109	VAL
271	ltnr3.TNFA	S	109	VAL
272	ltnr3.TNFA	R	110	MET
272	ltnr3.TNFA	T	110	MET
272	ltnr3.TNFA	S	110	MET
273	ltnr3.TNFA	R	110	LEU
273	ltnr3.TNFA	T	110	LEU
273	ltnr3.TNFA	S	110	LEU
274	ltnr3.TNFA	R	111	TYR
274	ltnr3.TNFA	T	111	TYR

274	1tnr3.TNFA	S	111	TYR
275	1tnr3.TNFA	R	112	ASN
275	1tnr3.TNFA	T	112	ASN
275	1tnr3.TNFA	S	112	ASN
276	1tnr3.TNFA	R	112	PHE
276	1tnr3.TNFA	T	112	PHE
276	1tnr3.TNFA	S	112	PHE
277	1tnr3.TNFA	R	115	ASP
277	1tnr3.TNFA	T	115	ASP
277	1tnr3.TNFA	S	115	ASP
278	1tnr3.TNFA	R	116	HIS
278	1tnr3.TNFA	T	116	HIS
278	1tnr3.TNFA	S	116	HIS
279	1tnr3.TNFA	R	116	LEU
279	1tnr3.TNFA	T	116	LEU
279	1tnr3.TNFA	S	116	LEU
280	1tnr3.TNFA	R	116	SER
280	1tnr3.TNFA	T	116	SER
280	1tnr3.TNFA	S	116	SER
281	1tnr3.TNFA	R	116	CYS
281	1tnr3.TNFA	T	116	CYS
281	1tnr3.TNFA	S	116	CYS
282	1tnr3.TNFA	R	117	ARG
282	1tnr3.TNFA	T	117	ARG
282	1tnr3.TNFA	S	117	ARG
283	1tnr3.TNFA	R	118	CYS
283	1tnr3.TNFA	T	118	CYS
283	1tnr3.TNFA	S	118	CYS
284	1tnr3.TNFA	R	118	ALA
284	1tnr3.TNFA	T	118	ALA
284	1tnr3.TNFA	S	118	ALA
285	1tnr3.TNFA	R	119	SER
285	1tnr3.TNFA	T	119	SER
285	1tnr3.TNFA	S	119	SER
286	1tnr3.TNFA	R	120	ASN
286	1tnr3.TNFA	T	120	ASN
286	1tnr3.TNFA	S	120	ASN
287	1tnr3.TNFA	R	120	LYS
287	1tnr3.TNFA	T	120	LYS
287	1tnr3.TNFA	S	120	LYS
288	1tnr3.TNFA	R	123	LEU
288	1tnr3.TNFA	T	123	LEU
288	1tnr3.TNFA	S	123	LEU
289	1tnr3.TNFA	R	123	ALA
289	1tnr3.TNFA	T	123	ALA
289	1tnr3.TNFA	S	123	ALA
290	1tnr3.TNFA	R	124	ALA
290	1tnr3.TNFA	T	124	ALA
290	1tnr3.TNFA	S	124	ALA
291	1tnr3.TNFA	R	125	THR
291	1tnr3.TNFA	T	125	THR
291	1tnr3.TNFA	S	125	THR
292	1tnr3.TNFA	R	125	ARG
292	1tnr3.TNFA	T	125	ARG
292	1tnr3.TNFA	S	125	ARG
293	1tnr3.TNFA	R	125	TYR
293	1tnr3.TNFA	T	125	TYR
293	1tnr3.TNFA	S	125	TYR
294	1tnr3.TNFA	R	125	HIS
294	1tnr3.TNFA	T	125	HIS
294	1tnr3.TNFA	S	125	HIS
295	1tnr3.TNFA	R	126	TYR
295	1tnr3.TNFA	T	126	TYR
295	1tnr3.TNFA	S	126	TYR
296	1tnr3.TNFA	R	127	ARG
296	1tnr3.TNFA	T	127	ARG
296	1tnr3.TNFA	S	127	ARG
297	1tnr3.TNFA	R	128	LEU
297	1tnr3.TNFA	T	128	LEU
297	1tnr3.TNFA	S	128	LEU
298	1tnr3.TNFA	R	133	SER
298	1tnr3.TNFA	T	133	SER
298	1tnr3.TNFA	S	133	SER
299	1tnr3.TNFA	R	134	VAL
299	1tnr3.TNFA	T	134	VAL
299	1tnr3.TNFA	S	134	VAL
300	1tnr3.TNFA	R	134	LYS
300	1tnr3.TNFA	T	134	LYS
300	1tnr3.TNFA	S	134	LYS
301	1tnr3.TNFA	R	135	THR
301	1tnr3.TNFA	T	135	THR
301	1tnr3.TNFA	S	135	THR
302	1tnr3.TNFA	R	135	SER
302	1tnr3.TNFA	T	135	SER
302	1tnr3.TNFA	S	135	SER
303	1tnr3.TNFA	R	135	GLU
303	1tnr3.TNFA	T	135	GLU
303	1tnr3.TNFA	S	135	GLU
304	1tnr3.TNFA	R	135	VAL
304	1tnr3.TNFA	T	135	VAL
304	1tnr3.TNFA	S	135	VAL
305	1tnr3.TNFA	R	138	VAL
305	1tnr3.TNFA	T	138	VAL
305	1tnr3.TNFA	S	138	VAL

Iteration number	Filename	Chain	Residue index in chain	New residue
34	1tnr-3	R	0	YR
34	1tnr-3	T	0	YR
34	1tnr-3	S	0	YR
35	1tnr-3	R	0	RO
35	1tnr-3	T	0	RO
35	1tnr-3	S	0	RO
36	1tnr-3	R	0	ER
36	1tnr-3	T	0	ER
36	1tnr-3	S	0	ER
37	1tnr-3	R	1	RO
37	1tnr-3	T	1	RO
37	1tnr-3	S	1	RO
38	1tnr-3	R	4	HR
38	1tnr-3	T	4	HR
38	1tnr-3	S	4	HR
39	1tnr-3	R	5	IS
39	1tnr-3	T	5	IS
39	1tnr-3	S	5	IS
40	1tnr-3	R	5	YS
40	1tnr-3	T	5	YS
40	1tnr-3	S	5	YS
41	1tnr-3	R	5	SP
41	1tnr-3	T	5	SP
41	1tnr-3	S	5	SP
42	1tnr-3	R	6	HR
42	1tnr-3	T	6	HR
42	1tnr-3	S	6	HR
43	1tnr-3	R	6	AL
43	1tnr-3	T	6	AL
43	1tnr-3	S	6	AL
44	1tnr-3	R	7	RG
44	1tnr-3	T	7	RG
44	1tnr-3	S	7	RG
45	1tnr-3	R	7	YR
45	1tnr-3	T	7	YR
45	1tnr-3	S	7	YR
46	1tnr-3	R	7	LY
46	1tnr-3	T	7	LY
46	1tnr-3	S	7	LY
47	1tnr-3	R	7	RO
47	1tnr-3	T	7	RO
47	1tnr-3	S	7	RO
48	1tnr-3	R	7	LN
48	1tnr-3	T	7	LN
48	1tnr-3	S	7	LN
49	1tnr-3	R	8	LA
49	1tnr-3	T	8	LA
49	1tnr-3	S	8	LA
50	1tnr-3	R	10	ASP
50	1tnr-3	T	10	ASP
50	1tnr-3	S	10	ASP
51	1tnr-3	R	10	THR
51	1tnr-3	T	10	THR
51	1tnr-3	S	10	THR
52	1tnr-3	R	11	LYS
52	1tnr-3	T	11	LYS
52	1tnr-3	S	11	LYS
53	1tnr-3	R	12	THR
53	1tnr-3	T	12	THR
53	1tnr-3	S	12	THR
54	1tnr-3	R	12	LEU
54	1tnr-3	T	12	LEU
54	1tnr-3	S	12	LEU
55	1tnr-3	R	13	SER
55	1tnr-3	T	13	SER
55	1tnr-3	S	13	SER
56	1tnr-3	R	14	SER
56	1tnr-3	T	14	SER
56	1tnr-3	S	14	SER
57	1tnr-3	R	14	GLY
57	1tnr-3	T	14	GLY
57	1tnr-3	S	14	GLY
58	1tnr-3	R	14	PHE
58	1tnr-3	T	14	PHE
58	1tnr-3	S	14	PHE
59	1tnr-3	R	14	TRP
59	1tnr-3	T	14	TRP
59	1tnr-3	S	14	TRP
60	1tnr-3	R	14	TYR
60	1tnr-3	T	14	TYR
60	1tnr-3	S	14	TYR
61	1tnr-3	R	14	ARG
61	1tnr-3	T	14	ARG
61	1tnr-3	S	14	ARG
62	1tnr-3	R	15	TYR
62	1tnr-3	T	15	TYR
62	1tnr-3	S	15	TYR
63	1tnr-3	R	15	SER
63	1tnr-3	T	15	SER
63	1tnr-3	S	15	SER
64	1tnr-3	R	15	ARG
64	1tnr-3	T	15	ARG
64	1tnr-3	S	15	ARG
65	1tnr-3	R	15	PHE
65	1tnr-3	T	15	PHE
65	1tnr-3	S	15	PHE
66	1tnr-3	R	18	TYR
66	1tnr-3	T	18	TYR
66	1tnr-3	S	18	TYR
67	1tnr-3	R	18	GLY
67	1tnr-3	T	18	GLY
67	1tnr-3	S	18	GLY

68	ltnr-3	R	21	GLU
68	ltnr-3	T	21	GLU
68	ltnr-3	S	21	GLU
69	ltnr-3	R	22	ARG
69	ltnr-3	T	22	ARG
69	ltnr-3	S	22	ARG
70	ltnr-3	R	22	THR
70	ltnr-3	T	22	THR
70	ltnr-3	S	22	THR
71	ltnr-3	R	22	LEU
71	ltnr-3	T	22	LEU
71	ltnr-3	S	22	LEU
72	ltnr-3	R	22	ILE
72	ltnr-3	T	22	ILE
72	ltnr-3	S	22	ILE
73	ltnr-3	R	23	SER
73	ltnr-3	T	23	SER
73	ltnr-3	S	23	SER
74	ltnr-3	R	23	CYS
74	ltnr-3	T	23	CYS
74	ltnr-3	S	23	CYS
75	ltnr-3	R	24	LEU
75	ltnr-3	T	24	LEU
75	ltnr-3	S	24	LEU
76	ltnr-3	R	24	PHE
76	ltnr-3	T	24	PHE
76	ltnr-3	S	24	PHE
77	ltnr-3	R	26	SER
77	ltnr-3	T	26	SER
77	ltnr-3	S	26	SER
78	ltnr-3	R	26	ARG
78	ltnr-3	T	26	ARG
78	ltnr-3	S	26	MET
79	ltnr-3	R	26	MET
79	ltnr-3	T	26	MET
80	ltnr-3	R	27	GLU
80	ltnr-3	T	27	GLU
80	ltnr-3	S	27	GLU
81	ltnr-3	R	28	PRO
81	ltnr-3	T	28	PRO
81	ltnr-3	S	28	PRO
82	ltnr-3	R	28	GLY
82	ltnr-3	T	28	GLY
82	ltnr-3	S	28	GLY
83	ltnr-3	R	28	GLY
83	ltnr-3	T	28	GLY
83	ltnr-3	S	28	GLY
84	ltnr-3	R	28	TYR
84	ltnr-3	T	28	TYR
84	ltnr-3	S	28	TYR
85	ltnr-3	R	28	PHE
85	ltnr-3	T	28	PHE
85	ltnr-3	S	28	ARG
86	ltnr-3	R	28	ARG
86	ltnr-3	T	28	ARG
86	ltnr-3	S	28	ARG
87	ltnr-3	R	28	TRP
87	ltnr-3	T	28	TRP
87	ltnr-3	S	28	TRP
88	ltnr-3	R	28	SER
88	ltnr-3	T	28	SER
88	ltnr-3	S	28	SER
89	ltnr-3	R	28	SER
89	ltnr-3	T	28	SER
89	ltnr-3	S	28	SER
90	ltnr-3	R	29	SER
90	ltnr-3	T	29	SER
90	ltnr-3	S	29	SER
91	ltnr-3	R	29	ARG
91	ltnr-3	T	29	ARG
91	ltnr-3	S	29	ARG
92	ltnr-3	R	30	ARG
92	ltnr-3	T	30	ARG
92	ltnr-3	S	30	ARG
93	ltnr-3	R	30	GLY
93	ltnr-3	T	30	GLY
93	ltnr-3	S	30	GLY
94	ltnr-3	R	30	MET
94	ltnr-3	T	30	MET
94	ltnr-3	S	30	MET
95	ltnr-3	R	30	LEU
95	ltnr-3	T	30	LEU
95	ltnr-3	S	30	LEU
96	ltnr-3	R	30	TYR
96	ltnr-3	T	30	TYR
96	ltnr-3	S	30	TYR
97	ltnr-3	R	31	ARG
97	ltnr-3	T	31	ARG
97	ltnr-3	S	31	ARG
98	ltnr-3	R	31	TYR
98	ltnr-3	T	31	TYR
98	ltnr-3	S	31	TYR
99	ltnr-3	R	31	SER
99	ltnr-3	T	31	SER
99	ltnr-3	S	31	SER
100	ltnr-3	R	31	ALA
100	ltnr-3	T	31	ALA
100	ltnr-3	S	31	ALA
101	ltnr-3	R	31	LEU
101	ltnr-3	T	31	LEU
101	ltnr-3	S	31	LEU
102	ltnr-3	R	31	ARG
102	ltnr-3	T	31	ARG

102	ltnr-3	S	31	ARG
103	ltnr-3	R	31	THR
103	ltnr-3	T	31	THR
103	ltnr-3	S	31	THR
104	ltnr-3	R	32	THR
104	ltnr-3	T	32	THR
104	ltnr-3	S	32	THR
105	ltnr-3	R	32	ASP
105	ltnr-3	T	32	ASP
105	ltnr-3	S	32	ASP
106	ltnr-3	R	34	PRO
106	ltnr-3	T	34	PRO
106	ltnr-3	S	34	PRO
107	ltnr-3	R	35	MET
107	ltnr-3	T	35	MET
107	ltnr-3	S	35	MET
108	ltnr-3	R	35	PRO
108	ltnr-3	T	35	PRO
108	ltnr-3	S	35	PRO
109	ltnr-3	R	35	LYS
109	ltnr-3	T	35	LYS
109	ltnr-3	S	35	LYS
110	ltnr-3	R	36	SER
110	ltnr-3	T	36	SER
110	ltnr-3	S	36	SER
111	ltnr-3	R	36	LEU
111	ltnr-3	T	36	LEU
111	ltnr-3	S	36	LEU
112	ltnr-3	R	36	ILE
112	ltnr-3	T	36	ILE
112	ltnr-3	S	36	ILE
113	ltnr-3	R	36	VAL
113	ltnr-3	T	36	VAL
113	ltnr-3	S	36	VAL
114	ltnr-3	R	37	SER
114	ltnr-3	T	37	SER
114	ltnr-3	S	37	SER
115	ltnr-3	R	37	ASN
115	ltnr-3	T	37	ASN
115	ltnr-3	S	37	ASN
116	ltnr-3	R	37	ALA
116	ltnr-3	T	37	ALA
116	ltnr-3	S	37	ALA
117	ltnr-3	R	37	ARG
117	ltnr-3	T	37	ARG
117	ltnr-3	S	37	ARG
118	ltnr-3	R	37	TRP
118	ltnr-3	T	37	TRP
118	ltnr-3	S	37	TRP
119	ltnr-3	R	37	TYR
119	ltnr-3	T	37	TYR
119	ltnr-3	S	37	TYR
120	ltnr-3	R	37	PHE
120	ltnr-3	T	37	PHE
120	ltnr-3	S	37	PHE
121	ltnr-3	R	38	GLY
121	ltnr-3	T	38	GLY
121	ltnr-3	S	38	GLY
122	ltnr-3	R	38	TRP
122	ltnr-3	T	38	TRP
122	ltnr-3	S	38	TRP
123	ltnr-3	R	38	PHE
123	ltnr-3	T	38	PHE
123	ltnr-3	S	38	PHE
124	ltnr-3	R	38	TYR
124	ltnr-3	T	38	TYR
124	ltnr-3	S	38	TYR
125	ltnr-3	R	39	ASP
125	ltnr-3	T	39	ASP
125	ltnr-3	S	39	ASP
126	ltnr-3	R	40	ARG
126	ltnr-3	T	40	ARG
126	ltnr-3	S	40	ARG
127	ltnr-3	R	40	TYR
127	ltnr-3	T	40	TYR
127	ltnr-3	S	40	TYR
128	ltnr-3	R	40	SER
128	ltnr-3	T	40	SER
128	ltnr-3	S	40	SER
129	ltnr-3	R	41	ASP
129	ltnr-3	T	41	ASP
129	ltnr-3	S	41	ASP
130	ltnr-3	R	41	LYS
130	ltnr-3	T	41	LYS
130	ltnr-3	S	41	LYS
131	ltnr-3	R	41	ILE
131	ltnr-3	T	41	ILE
131	ltnr-3	S	41	ILE
132	ltnr-3	R	42	PRO
132	ltnr-3	T	42	PRO
132	ltnr-3	S	42	PRO
133	ltnr-3	R	42	TYR
133	ltnr-3	T	42	TYR
133	ltnr-3	S	42	TYR
134	ltnr-3	R	42	LEU
134	ltnr-3	T	42	LEU
134	ltnr-3	S	42	LEU
135	ltnr-3	R	43	PRO
135	ltnr-3	T	43	PRO
135	ltnr-3	S	43	PRO
136	ltnr-3	R	43	PHE
136	ltnr-3	T	43	PHE
136	ltnr-3	S	43	PHE
137	ltnr-3	R	43	GLU

137	ltnr-3	T	43	GLU
137	ltnr-3	S	43	GLU
138	ltnr-3	R	43	SER
138	ltnr-3	T	43	SER
138	ltnr-3	S	43	SER
139	ltnr-3	R	44	GLU
139	ltnr-3	T	44	GLU
139	ltnr-3	S	44	GLU
140	ltnr-3	R	44	PRO
140	ltnr-3	T	44	PRO
140	ltnr-3	S	44	PRO
141	ltnr-3	R	45	SER
141	ltnr-3	T	45	SER
141	ltnr-3	S	45	SER
142	ltnr-3	R	45	LEU
142	ltnr-3	T	45	LEU
142	ltnr-3	S	45	LEU
143	ltnr-3	R	45	VAL
143	ltnr-3	T	45	VAL
143	ltnr-3	S	45	VAL
144	ltnr-3	R	46	SER
144	ltnr-3	T	46	SER
144	ltnr-3	S	46	SER
145	ltnr-3	R	46	GLY
145	ltnr-3	T	46	GLY
145	ltnr-3	S	46	GLY
146	ltnr-3	R	46	TYR
146	ltnr-3	T	46	TYR
146	ltnr-3	S	46	TYR
147	ltnr-3	R	46	ALA
147	ltnr-3	T	46	ALA
147	ltnr-3	S	46	ALA
148	ltnr-3	R	46	ASN
148	ltnr-3	T	46	ASN
148	ltnr-3	S	46	ASN
149	ltnr-3	R	46	ILE
149	ltnr-3	T	46	ILE
149	ltnr-3	S	46	ILE
150	ltnr-3	R	46	PRO
150	ltnr-3	T	46	PRO
150	ltnr-3	S	46	PRO
151	ltnr-3	R	46	TRP
151	ltnr-3	T	46	TRP
151	ltnr-3	S	46	TRP
152	ltnr-3	R	46	PRO
152	ltnr-3	T	46	PRO
152	ltnr-3	S	46	PRO
153	ltnr-3	R	46	ARG
153	ltnr-3	T	46	ARG
153	ltnr-3	S	46	ARG
154	ltnr-3	R	48	HIS
154	ltnr-3	T	48	HIS
154	ltnr-3	S	48	HIS
155	ltnr-3	R	48	ASN
155	ltnr-3	T	48	ASN
155	ltnr-3	S	48	ASN
156	ltnr-3	R	49	TRP
156	ltnr-3	T	49	TRP
156	ltnr-3	S	49	TRP
157	ltnr-3	R	49	SER
157	ltnr-3	T	49	SER
157	ltnr-3	S	49	SER
158	ltnr-3	R	49	ARG
158	ltnr-3	T	49	ARG
158	ltnr-3	S	49	ARG
159	ltnr-3	R	50	LYS
159	ltnr-3	T	50	LYS
159	ltnr-3	S	50	LYS
160	ltnr-3	R	50	CYS
160	ltnr-3	T	50	CYS
160	ltnr-3	S	50	CYS
161	ltnr-3	R	50	GLY
161	ltnr-3	T	50	GLY
161	ltnr-3	S	50	GLY
162	ltnr-3	R	50	ILE
162	ltnr-3	T	50	ILE
162	ltnr-3	S	50	ILE
163	ltnr-3	R	51	PRO
163	ltnr-3	T	51	PRO
163	ltnr-3	S	51	PRO
164	ltnr-3	R	51	LEU
164	ltnr-3	T	51	LEU
164	ltnr-3	S	51	LEU
165	ltnr-3	R	51	THR
165	ltnr-3	T	51	THR
165	ltnr-3	S	51	THR
166	ltnr-3	R	51	GLY
166	ltnr-3	T	51	GLY
166	ltnr-3	S	51	GLY
167	ltnr-3	R	51	TYR
167	ltnr-3	T	51	TYR
167	ltnr-3	S	51	TYR
168	ltnr-3	R	52	PHE
168	ltnr-3	T	52	PHE
168	ltnr-3	S	52	PHE
169	ltnr-3	R	52	PRO
169	ltnr-3	T	52	PRO
169	ltnr-3	S	52	PRO
170	ltnr-3	R	52	TYR
170	ltnr-3	T	52	TYR
170	ltnr-3	S	52	TYR
171	ltnr-3	R	53	GLN
171	ltnr-3	T	53	GLN
171	ltnr-3	S	53	GLN

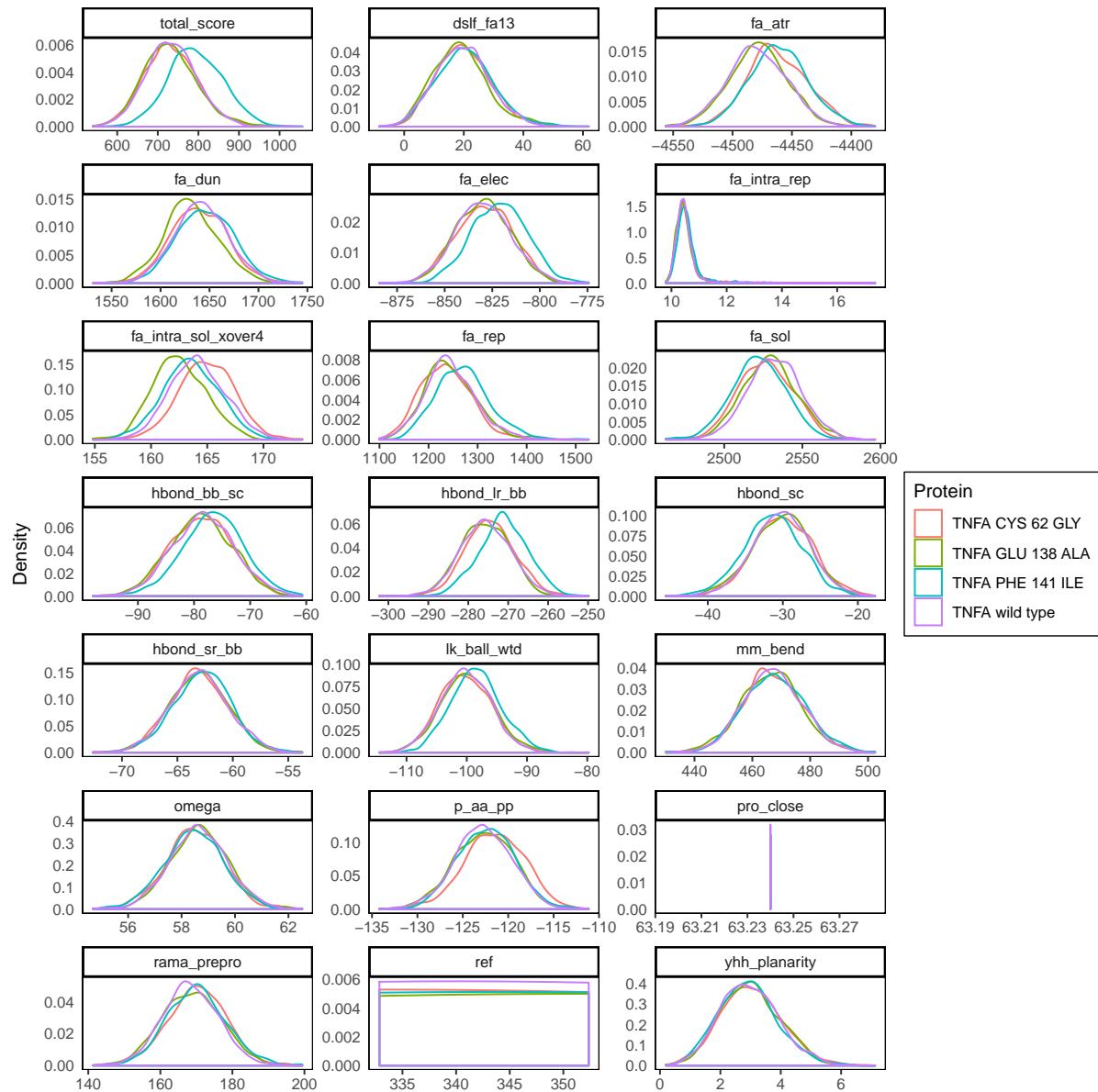
172	ltnr-3	R	54	LEU
172	ltnr-3	T	54	LEU
172	ltnr-3	S	54	CYS
173	ltnr-3	R	54	CYS
173	ltnr-3	T	54	CYS
173	ltnr-3	S	54	CYS
174	ltnr-3	R	54	ASN
174	ltnr-3	T	54	ASN
174	ltnr-3	S	54	ASN
175	ltnr-3	R	54	ILE
175	ltnr-3	T	54	ILE
175	ltnr-3	S	54	ILE
176	ltnr-3	R	54	SER
176	ltnr-3	T	54	SER
176	ltnr-3	S	54	SER
177	ltnr-3	R	55	ARG
177	ltnr-3	T	55	ARG
177	ltnr-3	S	55	ARG
178	ltnr-3	R	55	SER
178	ltnr-3	T	55	SER
178	ltnr-3	S	55	SER
179	ltnr-3	R	55	TYR
179	ltnr-3	T	55	TYR
179	ltnr-3	S	55	TYR
180	ltnr-3	R	55	GLY
180	ltnr-3	T	55	GLY
180	ltnr-3	S	55	GLY
181	ltnr-3	R	56	ARG
181	ltnr-3	T	56	ARG
181	ltnr-3	S	56	ARG
182	ltnr-3	R	57	ASN
182	ltnr-3	T	57	ASN
182	ltnr-3	S	57	ASN
183	ltnr-3	R	58	SER
183	ltnr-3	T	58	SER
183	ltnr-3	S	58	SER
184	ltnr-3	R	58	ASP
184	ltnr-3	T	58	ASP
184	ltnr-3	S	58	ASP
185	ltnr-3	R	58	ARG
185	ltnr-3	T	58	ARG
185	ltnr-3	S	58	ARG
186	ltnr-3	R	58	SER
186	ltnr-3	T	58	SER
186	ltnr-3	S	58	SER
187	ltnr-3	R	58	TRP
187	ltnr-3	T	58	TRP
187	ltnr-3	S	58	TRP
188	ltnr-3	R	59	TYR
188	ltnr-3	T	59	TYR
188	ltnr-3	S	59	TYR
189	ltnr-3	R	59	CYS
189	ltnr-3	T	59	CYS
189	ltnr-3	S	59	CYS
190	ltnr-3	R	61	TYR
190	ltnr-3	T	61	TYR
190	ltnr-3	S	61	TYR
191	ltnr-3	R	62	GLN
191	ltnr-3	T	62	GLN
191	ltnr-3	S	62	GLN
192	ltnr-3	R	62	PRO
192	ltnr-3	T	62	PRO
192	ltnr-3	S	62	PRO
193	ltnr-3	R	63	PHE
193	ltnr-3	T	63	PHE
193	ltnr-3	S	63	PHE
194	ltnr-3	R	64	ARG
194	ltnr-3	T	64	ARG
194	ltnr-3	S	64	ARG
195	ltnr-3	R	64	TYR
195	ltnr-3	T	64	TYR
195	ltnr-3	S	64	TYR
196	ltnr-3	R	65	ALA
196	ltnr-3	T	65	ALA
196	ltnr-3	S	65	ALA
197	ltnr-3	R	66	ILE
197	ltnr-3	T	66	ILE
197	ltnr-3	S	66	ILE
198	ltnr-3	R	67	LEU
198	ltnr-3	T	67	LEU
198	ltnr-3	S	67	LEU
199	ltnr-3	R	67	LYS
199	ltnr-3	T	67	LYS
199	ltnr-3	S	67	LYS
200	ltnr-3	R	68	PRO
200	ltnr-3	T	68	PRO
200	ltnr-3	S	68	PRO
201	ltnr-3	R	68	LEU
201	ltnr-3	T	68	LEU
201	ltnr-3	S	68	LEU
202	ltnr-3	R	68	TYR
202	ltnr-3	T	68	TYR
202	ltnr-3	S	68	TYR
203	ltnr-3	R	68	MET
203	ltnr-3	T	68	MET
203	ltnr-3	S	68	MET
204	ltnr-3	R	69	PHE
204	ltnr-3	T	69	PHE
204	ltnr-3	S	69	PHE
205	ltnr-3	R	70	ILE
205	ltnr-3	T	70	ILE
205	ltnr-3	S	70	ILE
206	ltnr-3	R	70	SER
206	ltnr-3	T	70	SER

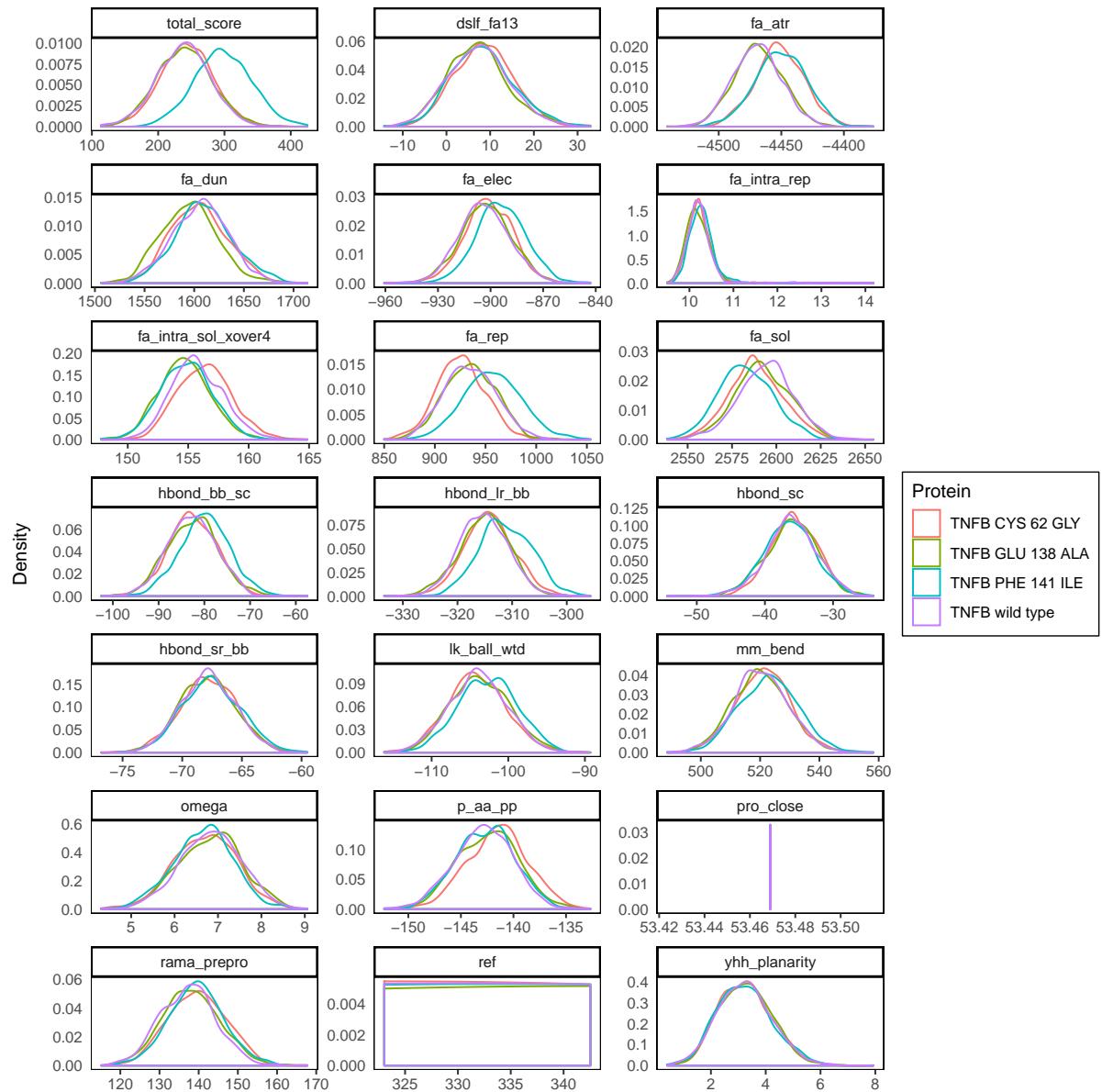
206	ltnr-3	S	70	SER
207	ltnr-3	R	71	ALA
207	ltnr-3	T	71	ALA
207	ltnr-3	S	71	ALA
208	ltnr-3	R	71	PRO
208	ltnr-3	T	71	PRO
208	ltnr-3	S	71	PRO
209	ltnr-3	R	71	PHE
209	ltnr-3	T	71	PHE
209	ltnr-3	S	71	PHE
210	ltnr-3	R	72	TYR
210	ltnr-3	T	72	TYR
210	ltnr-3	S	72	TYR
211	ltnr-3	R	72	PHE
211	ltnr-3	T	72	PHE
211	ltnr-3	S	72	PHE
212	ltnr-3	R	72	TRP
212	ltnr-3	T	72	TRP
212	ltnr-3	S	72	TRP
213	ltnr-3	R	73	TYR
213	ltnr-3	T	73	TYR
213	ltnr-3	S	73	TYR
214	ltnr-3	R	73	ARG
214	ltnr-3	T	73	ARG
214	ltnr-3	S	73	ARG
215	ltnr-3	R	73	ASP
215	ltnr-3	T	73	ASP
215	ltnr-3	S	73	ASP
216	ltnr-3	R	73	SER
216	ltnr-3	T	73	SER
216	ltnr-3	S	73	SER
217	ltnr-3	R	74	TYR
217	ltnr-3	T	74	TYR
217	ltnr-3	S	74	TYR
218	ltnr-3	R	74	ALA
218	ltnr-3	T	74	ALA
218	ltnr-3	S	74	ALA
219	ltnr-3	R	74	VAL
219	ltnr-3	T	74	VAL
219	ltnr-3	S	74	VAL
220	ltnr-3	R	74	ARG
220	ltnr-3	T	74	ARG
220	ltnr-3	S	74	ARG
221	ltnr-3	R	75	GLY
221	ltnr-3	T	75	GLY
221	ltnr-3	S	75	GLY
222	ltnr-3	R	75	ALA
222	ltnr-3	T	75	ALA
222	ltnr-3	S	75	ALA
223	ltnr-3	R	77	TRP
223	ltnr-3	T	77	TRP
223	ltnr-3	S	77	TRP
224	ltnr-3	R	77	CYS
224	ltnr-3	T	77	CYS
224	ltnr-3	S	77	CYS
225	ltnr-3	R	77	PRO
225	ltnr-3	T	77	PRO
225	ltnr-3	S	77	PRO
226	ltnr-3	R	77	GLN
226	ltnr-3	T	77	GLN
226	ltnr-3	S	77	GLN
227	ltnr-3	R	77	GLY
227	ltnr-3	T	77	GLY
227	ltnr-3	S	77	GLY
228	ltnr-3	R	78	HIS
228	ltnr-3	T	78	HIS
228	ltnr-3	S	78	HIS
229	ltnr-3	R	78	GLU
229	ltnr-3	T	78	GLU
229	ltnr-3	S	78	GLU
230	ltnr-3	R	79	ILE
230	ltnr-3	T	79	ILE
230	ltnr-3	S	79	ILE
231	ltnr-3	R	80	GLY
231	ltnr-3	T	80	GLY
231	ltnr-3	S	80	GLY
232	ltnr-3	R	80	ALA
232	ltnr-3	T	80	ALA
232	ltnr-3	S	80	ALA
233	ltnr-3	R	80	MET
233	ltnr-3	T	80	MET
233	ltnr-3	S	80	MET
234	ltnr-3	R	81	ARG
234	ltnr-3	T	81	ARG
234	ltnr-3	S	81	ARG
235	ltnr-3	R	81	TRP
235	ltnr-3	T	81	TRP
235	ltnr-3	S	81	TRP
236	ltnr-3	R	81	TYR
236	ltnr-3	T	81	TYR
236	ltnr-3	S	81	TYR
237	ltnr-3	R	81	PHE
237	ltnr-3	T	81	PHE
237	ltnr-3	S	81	PHE
238	ltnr-3	R	82	ASP
238	ltnr-3	T	82	ASP
238	ltnr-3	S	82	ASP
239	ltnr-3	R	82	ARG
239	ltnr-3	T	82	ARG
239	ltnr-3	S	82	ARG
240	ltnr-3	R	82	TYR
240	ltnr-3	T	82	TYR
240	ltnr-3	S	82	TYR
241	ltnr-3	R	83	TYR

241	ltnr-3	T	83	TYR
241	ltnr-3	S	83	TYR
242	ltnr-3	R	83	ARG
242	ltnr-3	T	83	ARG
242	ltnr-3	S	83	ARG
243	ltnr-3	R	85	LEU
243	ltnr-3	T	85	LEU
243	ltnr-3	S	85	LEU
244	ltnr-3	R	86	LYS
244	ltnr-3	T	86	LYS
244	ltnr-3	S	86	LYS
245	ltnr-3	R	87	GLU
245	ltnr-3	T	87	GLU
245	ltnr-3	S	87	GLU
246	ltnr-3	R	89	TRP
246	ltnr-3	T	89	TRP
246	ltnr-3	S	89	TRP
247	ltnr-3	R	89	GLN
247	ltnr-3	T	89	GLN
247	ltnr-3	S	89	GLN
248	ltnr-3	R	90	PRO
248	ltnr-3	T	90	PRO
248	ltnr-3	S	90	PRO
249	ltnr-3	R	91	HIS
249	ltnr-3	T	91	HIS
249	ltnr-3	S	91	HIS
250	ltnr-3	R	91	VAL
250	ltnr-3	T	91	VAL
250	ltnr-3	S	91	VAL
251	ltnr-3	R	91	CYS
251	ltnr-3	T	91	CYS
251	ltnr-3	S	91	CYS
252	ltnr-3	R	93	GLY
252	ltnr-3	T	93	GLY
252	ltnr-3	S	93	GLY
253	ltnr-3	R	94	GLY
253	ltnr-3	T	94	GLY
253	ltnr-3	S	94	GLY
254	ltnr-3	R	94	ALA
254	ltnr-3	T	94	ALA
254	ltnr-3	S	94	ALA
255	ltnr-3	R	96	THR
255	ltnr-3	T	96	THR
255	ltnr-3	S	96	THR
256	ltnr-3	R	97	LEU
256	ltnr-3	T	97	LEU
256	ltnr-3	S	97	LEU
257	ltnr-3	R	97	SER
257	ltnr-3	T	97	SER
257	ltnr-3	S	97	SER
258	ltnr-3	R	97	CYS
258	ltnr-3	T	97	CYS
258	ltnr-3	S	97	CYS
259	ltnr-3	R	97	ILE
259	ltnr-3	T	97	ILE
259	ltnr-3	S	97	ILE
260	ltnr-3	R	98	PHE
260	ltnr-3	T	98	PHE
260	ltnr-3	S	98	PHE
261	ltnr-3	R	99	TRP
261	ltnr-3	T	99	TRP
261	ltnr-3	S	99	TRP
262	ltnr-3	R	99	ARG
262	ltnr-3	T	99	ARG
262	ltnr-3	S	99	ARG
263	ltnr-3	R	101	ASN
263	ltnr-3	T	101	ASN
263	ltnr-3	S	101	ASN
264	ltnr-3	R	101	ASP
264	ltnr-3	T	101	ASP
264	ltnr-3	S	101	ASP
265	ltnr-3	R	101	SER
265	ltnr-3	T	101	SER
265	ltnr-3	S	101	SER
266	ltnr-3	R	102	TYR
266	ltnr-3	T	102	TYR
266	ltnr-3	S	102	TYR
267	ltnr-3	R	106	PHE
267	ltnr-3	T	106	PHE
267	ltnr-3	S	106	PHE
268	ltnr-3	R	107	SER
268	ltnr-3	T	107	SER
268	ltnr-3	S	107	SER
269	ltnr-3	R	108	ALA
269	ltnr-3	T	108	ALA
269	ltnr-3	S	108	ALA
270	ltnr-3	R	109	ILE
270	ltnr-3	T	109	ILE
270	ltnr-3	S	109	ILE
271	ltnr-3	R	109	VAL
271	ltnr-3	T	109	VAL
271	ltnr-3	S	109	VAL
272	ltnr-3	R	110	MET
272	ltnr-3	T	110	MET
272	ltnr-3	S	110	MET
273	ltnr-3	R	110	LEU
273	ltnr-3	T	110	LEU
273	ltnr-3	S	110	LEU
274	ltnr-3	R	111	TYR
274	ltnr-3	T	111	TYR
274	ltnr-3	S	111	TYR
275	ltnr-3	R	112	ASN
275	ltnr-3	T	112	ASN
275	ltnr-3	S	112	ASN

276	ltnr-3	R	112	PHE
276	ltnr-3	T	112	PHE
276	ltnr-3	S	112	PHE
277	ltnr-3	R	115	ASP
277	ltnr-3	T	115	ASP
277	ltnr-3	S	115	ASP
278	ltnr-3	R	116	HIS
278	ltnr-3	T	116	HIS
278	ltnr-3	S	116	HIS
279	ltnr-3	R	116	LEU
279	ltnr-3	T	116	LEU
279	ltnr-3	S	116	LEU
280	ltnr-3	R	116	SER
280	ltnr-3	T	116	SER
280	ltnr-3	S	116	SER
281	ltnr-3	R	116	CYS
281	ltnr-3	T	116	CYS
281	ltnr-3	S	116	CYS
282	ltnr-3	R	117	ARG
282	ltnr-3	T	117	ARG
282	ltnr-3	S	117	ARG
283	ltnr-3	R	118	CYS
283	ltnr-3	T	118	CYS
283	ltnr-3	S	118	CYS
284	ltnr-3	R	118	ALA
284	ltnr-3	T	118	ALA
284	ltnr-3	S	118	ALA
285	ltnr-3	R	119	SER
285	ltnr-3	T	119	SER
285	ltnr-3	S	119	SER
286	ltnr-3	R	120	ASN
286	ltnr-3	T	120	ASN
286	ltnr-3	S	120	ASN
287	ltnr-3	R	120	LYS
287	ltnr-3	T	120	LYS
287	ltnr-3	S	120	LYS
288	ltnr-3	R	123	LEU
288	ltnr-3	T	123	LEU
288	ltnr-3	S	123	LEU
289	ltnr-3	R	123	ALA
289	ltnr-3	T	123	ALA
289	ltnr-3	S	123	ALA
290	ltnr-3	R	124	ALA
290	ltnr-3	T	124	ALA
290	ltnr-3	S	124	ALA
291	ltnr-3	R	125	THR
291	ltnr-3	T	125	THR
291	ltnr-3	S	125	THR
292	ltnr-3	R	125	ARG
292	ltnr-3	T	125	ARG
292	ltnr-3	S	125	ARG
293	ltnr-3	R	125	TYR
293	ltnr-3	T	125	TYR
293	ltnr-3	S	125	TYR
294	ltnr-3	R	125	HIS
294	ltnr-3	T	125	HIS
294	ltnr-3	S	125	HIS
295	ltnr-3	R	126	TYR
295	ltnr-3	T	126	TYR
295	ltnr-3	S	126	TYR
296	ltnr-3	R	127	ARG
296	ltnr-3	T	127	ARG
296	ltnr-3	S	127	ARG
297	ltnr-3	R	128	LEU
297	ltnr-3	T	128	LEU
297	ltnr-3	S	128	LEU
298	ltnr-3	R	133	SER
298	ltnr-3	T	133	SER
298	ltnr-3	S	133	SER
299	ltnr-3	R	134	VAL
299	ltnr-3	T	134	VAL
299	ltnr-3	S	134	VAL
300	ltnr-3	R	134	LYS
300	ltnr-3	T	134	LYS
300	ltnr-3	S	134	LYS
301	ltnr-3	R	135	THR
301	ltnr-3	T	135	THR
301	ltnr-3	S	135	THR
302	ltnr-3	R	135	SER
302	ltnr-3	T	135	SER
302	ltnr-3	S	135	SER
303	ltnr-3	R	135	GLU
303	ltnr-3	T	135	GLU
303	ltnr-3	S	135	GLU
304	ltnr-3	R	135	VAL
304	ltnr-3	T	135	VAL
304	ltnr-3	S	135	VAL
305	ltnr-3	R	138	VAL
305	ltnr-3	T	138	VAL
305	ltnr-3	S	138	VAL

SPVAA Backrub density plots





HOPE reports

Method

The 3D-structure of your protein of interest is known. Information from this 3D-structure will be obtained using WHAT IF Web services, the UniProt database and the Reprof software.

The structural information was obtained from the analysis of PDB:

1EXT (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1EXT>)
Annotations about this protein were obtained from UniProt entry

p19438 (<http://www.uniprot.org/uniprot/p19438>)
See the method page for more information.

Amino Acids

You are interested in the mutation of a Cysteine into a Glycine at position 62.

The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.



Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

The mutant residue is smaller than the wild-type residue.

The wild-type residue is more hydrophobic than the mutant residue.

The report will evaluate the effect of the mutation on the following features: Contacts made by the mutated residue, structural domains in which the residue is located, modifications on this residue and known variants for this residue. A feature will only be shown when information is available. A short conclusion based on just the amino acid properties is shown always. In case a 3D-structure/model is available you will also find images and animations in the report.

Contacts

The wild-type residue is annotated in UniProt to be involved in a cysteine bridge, which is important for stability of the protein. Only cysteines can make these type of bonds, the mutation causes loss of this interaction and will have a severe effect on the 3D-structure of the protein.

Together with loss of the cysteine bond, the differences between the old and new residue can cause destabilization of the structure.

Structure

The mutation is located within a stretch of residues that is repeated in the protein, this repeat is named TNFR-Cys 1. The mutation into another residue might disturb this repeat and consequently any function this repeat might have.

The mutation introduces a glycine at this position. Glycines are very flexible and can disturb the required rigidity of the protein at this position.

Variants

This mutation matches a previously described variant, with the following description: Familial hibernian fever (FHF) [MIM:142680].

See the ExPASy site about this variant:

VAR_019303 (http://www.expasy.org/cgi-bin/variant_pages/get-sprot-variant.pl?VAR_019303)
The variant is annotated with severity: DISEASE

The mutation is located in a region with known splice variants, described as:

Familial hibernian fever (FHF) [MIM:142680] Familial hibernian fever (FHF) [MIM:142680]

Conservation

Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein.

The mutant and wild-type residue are not very similar. Based on this conservation information this mutation is probably damaging to the protein.

Domains

Hope Version 1.1.1

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Tnfr/Ngfr Cysteine-Rich Region IPR001368 (http://www.ebi.ac.uk/interpro/entry/IPR001368)	Protein Binding GO:0005515 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005515)	Binding GO:0005488 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005488) Molecular_Function GO:0003674 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0003674)
Tumor Necrosis Factor Receptor 1A, N-Terminal IPR033993 (http://www.ebi.ac.uk/interpro/entry/IPR033993)	None	None

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for binding. The mutation might disturb the interaction between these two domains and as such affect the function of the protein.

The mutated residue is located in a domain that is important for binding of other molecules. The mutated residue is in contact with residues in another domain. It is possible that the mutation disturbs these contacts.

Amino Acid Properties

The wild-type and mutant amino acids differ in size.

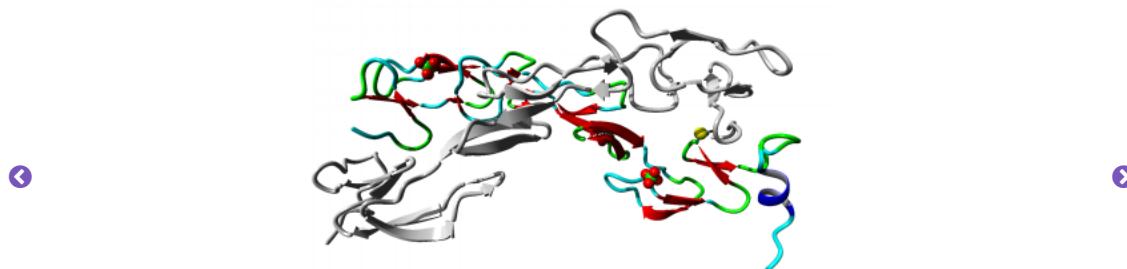
The mutant residue is smaller than the wild-type residue.

The mutation will cause an empty space in the core of the protein.

The hydrophobicity of the wild-type and mutant residue differs.

The mutation will cause loss of hydrophobic interactions in the core of the protein.

Images



Download (/hope/yasara/94214ee3-183d-441c-a1a1-c4a8e768cf26/16GLY_overview.png)

Overview of the protein in ribbon-presentation. The protein is coloured by element; α -helix=blue, β -strand = red, turn=green, 3/10 helix=yellow and random coil=cyan. Other molecules in the complex are coloured grey when present.

Citation



Please use the following citation when referencing the results in your report:

Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.

BMC Bioinformatics. 2010 Nov 8;11(1):548. DOI: 10.1186/1471-2105-11-548. (<http://dx.doi.org/10.1186/1471-2105-11-548>) PubMed: 21059217. (<http://www.ncbi.nlm.nih.gov/pubmed/21059217>)

Method

The 3D-structure of your protein of interest is known. Information from this 3D-structure will be obtained using WHAT IF Web services, the UniProt database and the Reprof software.

The structural information was obtained from the analysis of PDB:

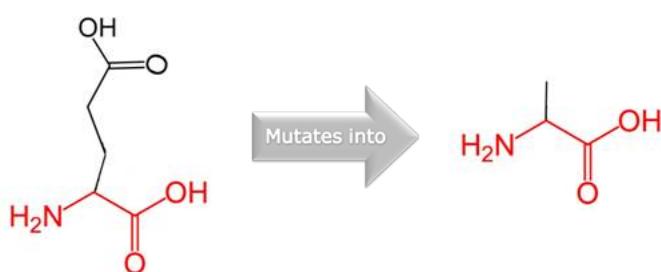
1EXT (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1EXT>)
Annotations about this protein were obtained from UniProt entry

p19438 (<http://www.uniprot.org/uniprot/p19438>)
See the method page for more information.

Amino Acids

You are interested in the mutation of a Glutamic Acid into a Alanine at position 138.

The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.



Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

The mutant residue is smaller than the wild-type residue.

The wild-type residue charge was NEGATIVE, the mutant residue charge is NEUTRAL.

The mutant residue is more hydrophobic than the wild-type residue.

The report will evaluate the effect of the mutation on the following features: Contacts made by the mutated residue, structural domains in which the residue is located, modifications on this residue and known variants for this residue. A feature will only be shown when information is available. A short conclusion based on just the amino acid properties is shown always. In case a 3D-structure/model is available you will also find images and animations in the report.

Contacts

The wild-type residue forms a salt bridge with:

- Proline at position 368
- Leucine at position 390

The difference in charge will disturb the ionic interaction made by the original, wild-type residue.

Structure

The mutation is located within a stretch of residues that is repeated in the protein, this repeat is named TNFR-Cys 3. The mutation into another residue might disturb this repeat and consequently any function this repeat might have.

Conservation

The wild-type residue occurs often at this position in the sequence, but other residues have also been observed here.

Your mutant residue is among the other residue types that have been observed at this position in homologous sequences. This means that this mutation can occur at this position and is probably not damaging to the protein.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
-----------------	--------------------	--------------------------

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Tumour Necrosis Factor Receptor 1A IPR020419 (http://www.ebi.ac.uk/interpro/entry/IPR020419)	Tumor Necrosis Factor-Activated Receptor Activity GO:0005031 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005031)	Molecular Transducer Activity GO:0060089 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0060089) Molecular_Function GO:0003674 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0003674)
Tnfr/Ngfr Cysteine-Rich Region IPR001368 (http://www.ebi.ac.uk/interpro/entry/IPR001368)	Protein Binding GO:0005515 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005515)	Binding GO:0005488 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005488) Molecular_Function GO:0003674 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0003674)
Tumor Necrosis Factor Receptor 1A, N-Terminal IPR033993 (http://www.ebi.ac.uk/interpro/entry/IPR033993)	None	None

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for binding. The mutation might disturb the interaction between these two domains and as such affect the function of the protein.

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is important for the activity of the protein. The mutation might affect this interaction and thereby disturb signal transfer from binding domain to the activity domain.

The mutated residue is located in a domain that is important for binding of other molecules. The mutated residue is in contact with residues in another domain. It is possible that the mutation disturbs these contacts.

Amino Acid Properties

There is a difference in charge between the wild-type and mutant amino acid.

The charge of the wild-type residue is lost by this mutation. This can cause loss of interactions with other molecules.

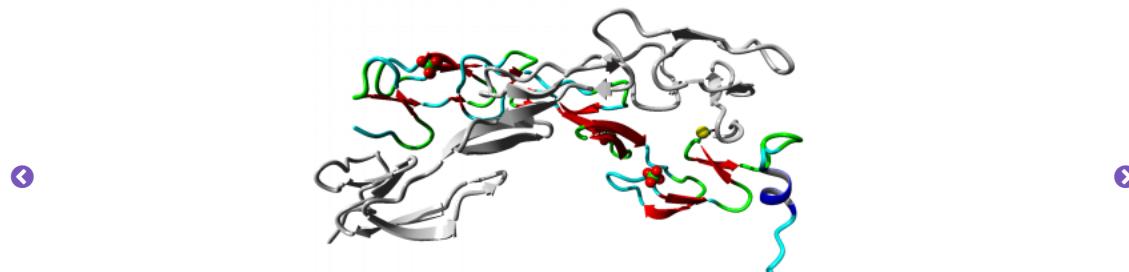
The wild-type and mutant amino acids differ in size.

The mutant residue is smaller than the wild-type residue.

This will cause a possible loss of external interactions.

The hydrophobicity of the wild-type and mutant residue differs.

Images



Download (/hope/yasara/5ecdbfa0-57b7-4570-b5d0-0def9401cf33/25ALA_overview.png/)

Overview of the protein in ribbon-presentation. The protein is coloured by element; α -helix=blue, β -strand = red, turn=green, 3/10 helix=yellow and random coil=cyan. Other molecules in the complex are coloured grey when present.

Citation

Please use the following citation when referencing the results in your report:

Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.

BMC Bioinformatics. 2010 Nov 8;11(1):548. DOI: 10.1186/1471-2105-11-548. (<http://dx.doi.org/10.1186/1471-2105-11-548>) PubMed: 21059217. (<http://www.ncbi.nlm.nih.gov/pubmed/21059217>)

Method

The 3D-structure of your protein of interest is known. Information from this 3D-structure will be obtained using WHAT IF Web services, the UniProt database and the Reprof software.

The structural information was obtained from the analysis of PDB:

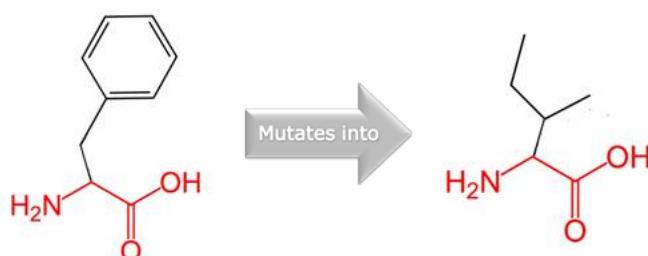
1EXT (<http://www.rcsb.org/pdb/explore/explore.do?structureId=1EXT>)
Annotations about this protein were obtained from UniProt entry

p19438 (<http://www.uniprot.org/uniprot/p19438>)
See the method page for more information.

Amino Acids

You are interested in the mutation of a Phenylalanine into a Isoleucine at position 141.

The figure below shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.



Each amino acid has its own specific size, charge, and hydrophobicity-value. The original wild-type residue and newly introduced mutant residue often differ in these properties.

The mutant residue is smaller than the wild-type residue.

The report will evaluate the effect of the mutation on the following features: Contacts made by the mutated residue, structural domains in which the residue is located, modifications on this residue and known variants for this residue. A feature will only be shown when information is available. A short conclusion based on just the amino acid properties is shown always. In case a 3D-structure/model is available you will also find images and animations in the report.

Structure

The mutation is located within a stretch of residues that is repeated in the protein, this repeat is named TNFR-Cys 3. The mutation into another residue might disturb this repeat and consequently any function this repeat might have.

Conservation

The wild-type residue is very conserved, but a few other residue types have been observed at this position too.

Your mutant residue was among the residues at this position observed in other sequences. This means that homologous proteins exist with the same residue type as your mutant at this position and this mutation is possibly not damaging to the protein.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Tumour Necrosis Factor Receptor 1A IPR020419 (http://www.ebi.ac.uk/interpro/entry/IPR020419)	Tumor Necrosis Factor-Activated Receptor Activity GO:0005031 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005031)	Molecular Transducer Activity GO:0060089 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0060089) Molecular_Function GO:0003674 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0003674)
Tnfr/Ngfr Cysteine-Rich Region IPR001368 (http://www.ebi.ac.uk/interpro/entry/IPR001368)	Protein Binding GO:0005515 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005515)	Binding GO:0005488 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0005488) Molecular_Function GO:0003674 (http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0003674)

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Tumor Necrosis Factor Receptor 1A, N-Terminal IPR033993 (http://www.ebi.ac.uk/interpro/entry/IPR033993)	None	None

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is also important for binding. The mutation might disturb the interaction between these two domains and as such affect the function of the protein.

The mutated residue is located in a domain that is important for binding of other molecules and in contact with residues in a domain that is important for the activity of the protein. The mutation might affect this interaction and thereby disturb signal transfer from binding domain to the activity domain.

The mutated residue is located in a domain that is important for binding of other molecules. The mutated residue is in contact with residues in another domain. It is possible that the mutation disturbs these contacts.

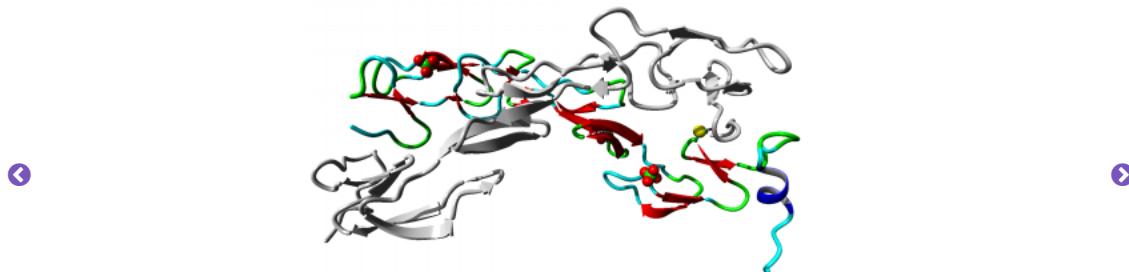
Amino Acid Properties

The wild-type and mutant amino acids differ in size.

The mutant residue is smaller than the wild-type residue.

The mutation will cause an empty space in the core of the protein.

Images



Overview of the protein in ribbon-presentation. The protein is coloured by element; α -helix=blue, β -strand = red, turn=green, 3/10 helix=yellow and random coil=cyan. Other molecules in the complex are coloured grey when present.

Citation

Please use the following citation when referencing the results in your report:

Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.
BMC Bioinformatics. 2010 Nov 8;11(1):548. DOI: 10.1186/1471-2105-11-548. (<http://dx.doi.org/10.1186/1471-2105-11-548>) PubMed: 21059217. (<http://www.ncbi.nlm.nih.gov/pubmed/21059217>)

R packages

Package	Version
ggplot2	3.1.1
data.table	1.12.2