Afilbercept

Taeyoon kim

May 23, 2022

CONTENTS

Introduction

The data in the report is intended to be a resource to guide the development and lead-optimization of a therapeutic protein. These data can be used in a preemptive fashion - for example, in the decision to substitute an exposed residue on the surface that may be prone to the kind of chemical modification that might affect the stability of the protein. They can also be used to assist the troubleshooting of problems that can arise in the course of an clinical development - for example if an therapeutic protein displays stability issues in storage, or unacceptably high levels of immunogenicity in early clinical trials.

There is always a great deal of risk involved in the development of any therapeutic molecule but experience has shown that the kind of data presented in this report is an invaluable tool for mitigating that risk - either by helping to identify potential problems before they occur, or by guiding the troubleshooting ofproblems that can occur during the antibody's development and lead-optimization.

History

> Aflibercept, sold under the brand names Eylea and Zaltrap, is a medication used to treat wet macular degeneration and metastatic colorectal cancer. It was developed by Regeneron Pharmaceuticals and is approved in the United States and the European Union. - wikipedia

Aflibercept is a recombinant fusion protein consisting of vascular endothelial growth factor (VEGF)-binding portions from the extracellular domains of human VEGF receptors 1 and 2, that are fused to the Fc portion of the human IgG1 immunoglobulin.

Regeneron commenced clinical testing of aflibercept in cancer in 2001. In 2003, Regeneron signed a major deal with Aventis to develop aflibercept in the field of cancer. In 2004 Regeneron started testing the compound, locally delivered, in proliferative eye diseases, and in 2006 Regeneron and Bayer signed an agreement to develop the eye indications.

Protein sequence

```
SDTGRPFVEMYSEIPEIIHMTEGRELVIPCRVTSPNITVTLKKFPLDTLIPDGKRIIWDSRKGFIISNATYKEIGLLTCE
ATVNGHLYKTNYLTHRQTNTIIDVVLSPSHGIELSVGEKLVLNCTARTELNVGIDFNWEYPSSKHQHKKLVNRDLKTQSG
SEMKKFLSTLTIDGVTRSDQGLYTCAASSGLMTKKNSTFVRVHEKDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISR
TPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT
ISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKS
RWQQGNVFSCSVMHEALHNHYTQKSLSLSPG
```

CHAPTER

ONE

PROTEIN PARAMETERS ANALYSIS

The program performs most of the same functions as the Expasy ProtParam tool.

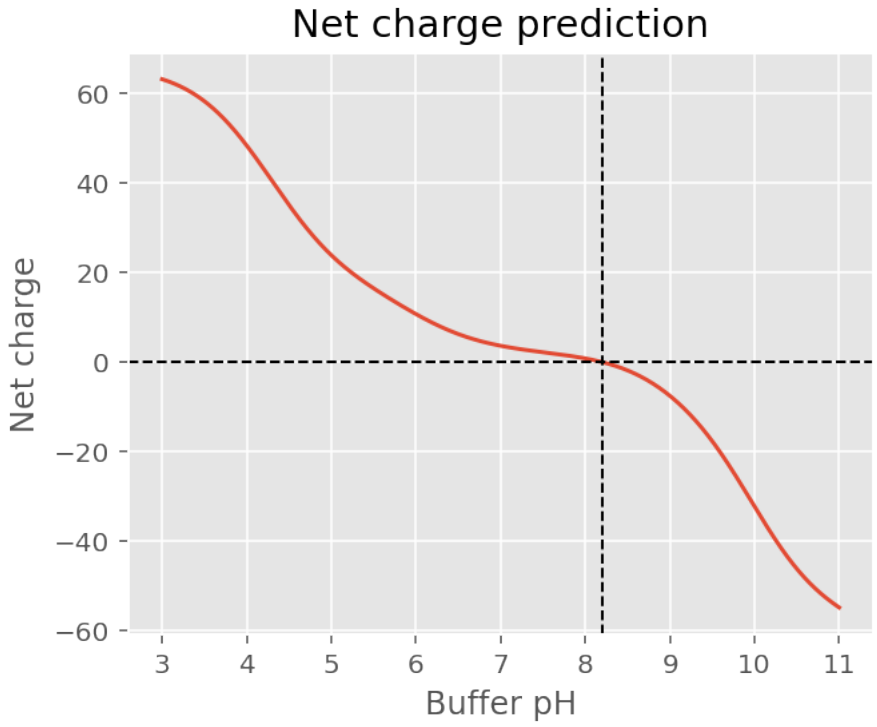
- Molecular weight
  - Amino acids are the building blocks that form polypeptides and ultimately proteins. Calculates the molecular weight of a protein.
- Chemical composition
  - A chemical formula is a way of presenting information about the chemical proportions of atoms that constitute a particular chemical compound or molecule, using chemical element symbols, numbers.
- Extinction coefficient
  - Extinction (or extinction coefficient) is defined as the ratio of maximum to minimum transmission of a beam of light that passes through a polarization optical train. extinction coefficient in units of M-1 cm-1, at 280 nm measured in water.
- Theoretical pI
  - The isoelectric point (pI, pH(I), IEP), is the pH at which a molecule carries no net electrical charge or is electrically neutral in the statistical mean. The pI value can affect the solubility of a molecule at a given pH. Such molecules have minimum solubility in water or salt solutions at the pH that corresponds to their pI and often precipitate out of solution. Biological amphoteric molecules such as proteins contain both acidic and basic functional groups.
- Aromaticity
  - Calculate the aromaticity according to Lobry, 1994. Calculates the aromaticity value of a protein according to Lobry, 1994. It is simply the relative frequency of Phe+Trp+Tyr.
- GRAVY
  - The GRAVY value is calculated by adding the hydropathy value for each residue and dividing by the length of the sequence (Kyte and Doolittle; 1982). A higher value is more hydrophobic. A lower value is more hydrophilic.
- Instability_index
  - Implementation of the method of Guruprasad et al. (1990, Protein Engineering, 4, 155-161). This method tests a protein for stability. Any value above 40 means the protein is unstable (=has a short half life).

```
# Name of target protein: ----------------------------Aflibercept
# Molecular weight(Dalton): ---------------------------------48,459
# Total number of amino acid: -------------------------------431
# Chemical formula: ------------------------C2159H3404N582O652S16
# Total number of atom in Aflibercept is ---------------------6813
```

(continues on next page)

(continued from previous page)

```
# Extinction coefficient(reduced): --------------------------55350
# Reduced Abs 0.1%(=1 g/L): ---------------------------------1.142
# Extinction coefficient(non-reduced): ----------------------55975
# Non-reduced Abs 0.1%(=1 g/L): -----------------------------1.155
# Theoretical pI: -------------------------------------------8.198
# Aromaticity: ----------------------------------------------7.89%
# The GRAVY value is ----------------------------------------0.460
  Aflibercept is more hydrophilic protein.
# The instability index: ------------------------------------39.222
  Aflibercept is seems stable.
```

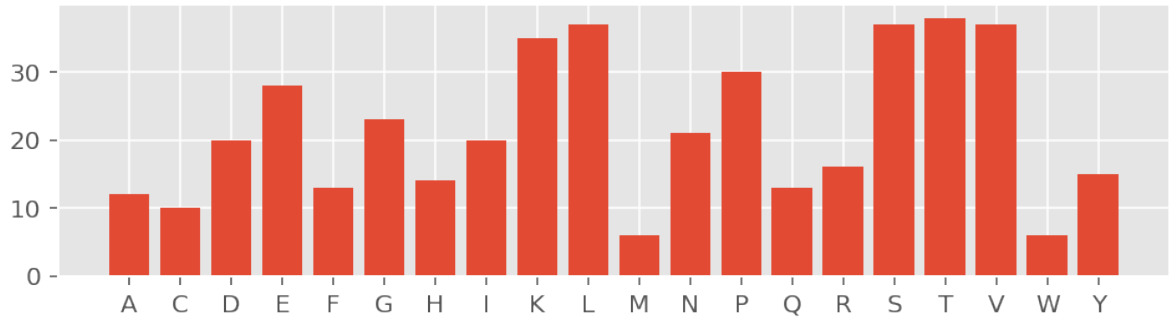


1.1 Amino acid composition

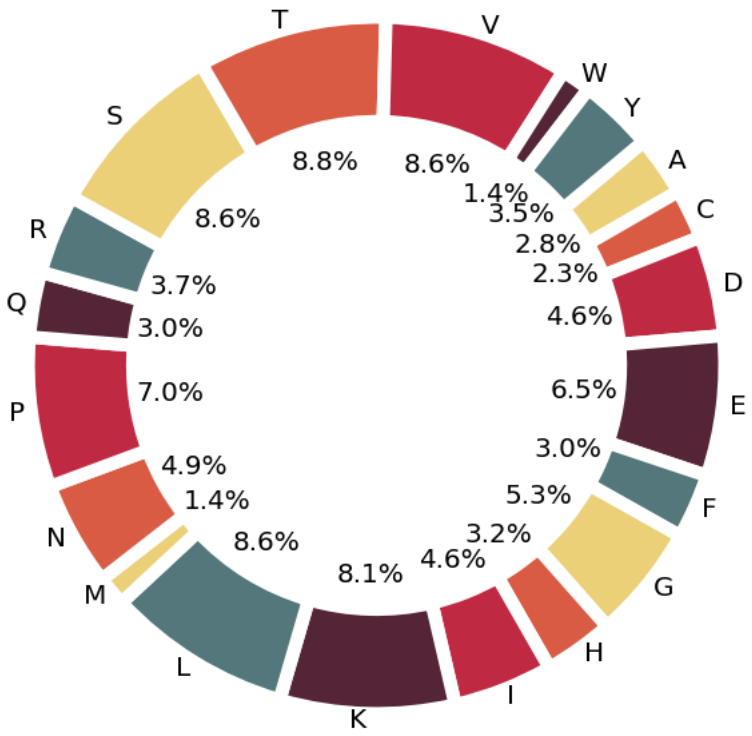
We can easily count the number of each type of amino acid.

- Number of each amino acids
  - Simply counts the number times an amino acid is repeated in the protein sequence.

```
Total number of positively charged residues(Arg + Lys):--------51
Total number of negatively charged residues(Asp + Glu):--------48
```



- Percent of amino acid contents



1.2 Potential sites of chemical modification

An initial scan of the protein sequences is presented based purely upon sequence. If a structural analysis was also requested, this section should be used in conjunction with the molecular surface analysis described in a subsequent section. Any of the sites listed below could be candidates for further consideration if the molecular surface analysis shows that they are significantly exposed on the surface of the protein, increasing their propensity for chemical modification. The canonical sequence analysis is also helpful here, since each of these sites can also be considered in the context of their frequency of occurrence within the canonical library of homologous sequences.

1.2.1 Potential deamidation positions

Asparagine (N) and glutamine (Q) residues are particularly prone to deamidation when they are followed in the sequence by amino acids with smaller side chains, that leave the intervening peptide group more exposed. Deamidation proceeds much more quickly if the susceptible amino acid is followed by a small, flexible residue such as glycine whose low steric hindrance leaves the peptide group open for attack.

- Search patterns: ASN/GLN-ALA/GLY/SER/THR

```
68-NA-69
84-NG-85
97-QT-98
99-NT-100
158-QS-159
180-QG-181
196-NS-197
271-NA-272
282-NS-283
300-NG-301
369-NG-370
404-QG-405
```

1.2.2 Potential o-linked glycosylation sites

The O-linked glycosylation of serine and threonine residues seems to be particularly sensitive to the presence of one or more proline residues in their vicinity in the sequence, particularly in the 2-1 and +3 positions.

- Search patterns: PRO-SER/THR

```
108-PS-109
141-PS-142
223-PS-224
338-PS-339
359-PS-360
```

- Search patterns: SER/THR-X-X-PRO

```
3-TGRP-6
12-SEIP-15
48-TLIP-51
210-TCPP-213
239-SRTP-242
335-TLPP-338
378-TTPP-381
427-SLSP-430
```

1.2.3 Potential n-linked glycosylation sites

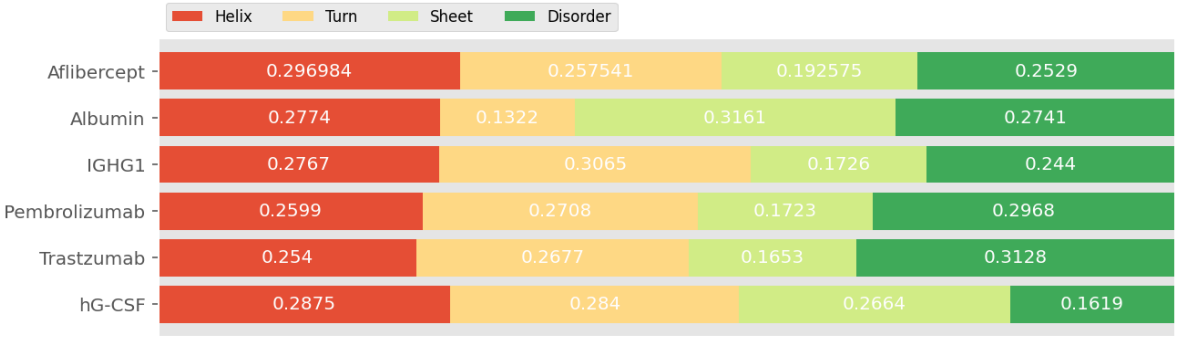
- Search patterns: ASN-X-SER/THR

```
36-NIT-38
68-NAT-70
123-NCT-125
196-NST-198
282-NST-284
```

1.3 Secondary structure fraction

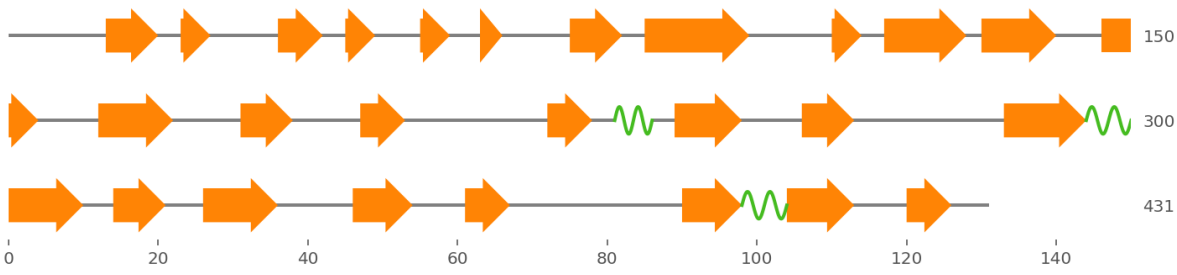
The fraction of amino acids that tend to be found in the three classical secondary structures. These are beta sheets, alpha helixes, and turns (where the residues change direction).

- Amino acids in helix: V, I, Y, F, W, L.
- Amino acids in turn: N, P, G, S.
- Amino acids in sheet: E, M, A, L.



1.4 Secondary structure prediction

Protein secondary structure prediction is one of the most important and challenging problems in bioinformatics. Here in, the P-SEA algorithm that to predict the secondary structures of proteins sequences based only on knowledge of their primary structure.



1.5 Structural analysis

1.5.1 Detection of disulfide bonds

This function detects disulfide bridges in protein structures. Then the detected disulfide bonds are visualized and added to the bonds attribute of the AtomArray. The employed criteria for disulfide bonds are quite simple in this case: the atoms of two cystein residues must be in a vicinity of Å and the dihedral angle of must be .

```
    A     352  CYS SG     S       -21.681    6.749   19.495
    A     410  CYS SG     S       -21.802    8.706   18.752
```

The found disulfide bonds are visualized with the help of Matplotlib: The amino acid sequence is written on the X-axis and the disulfide bonds are depicted by yellow semi-ellipses.

1.5.2 Calculation of protein diameter

This calculates the diameter of a protein defined as the maximum pairwise atom distance.

```
# Diameter of Aflibercept is: ----------------------145.705 Angstrong.
```

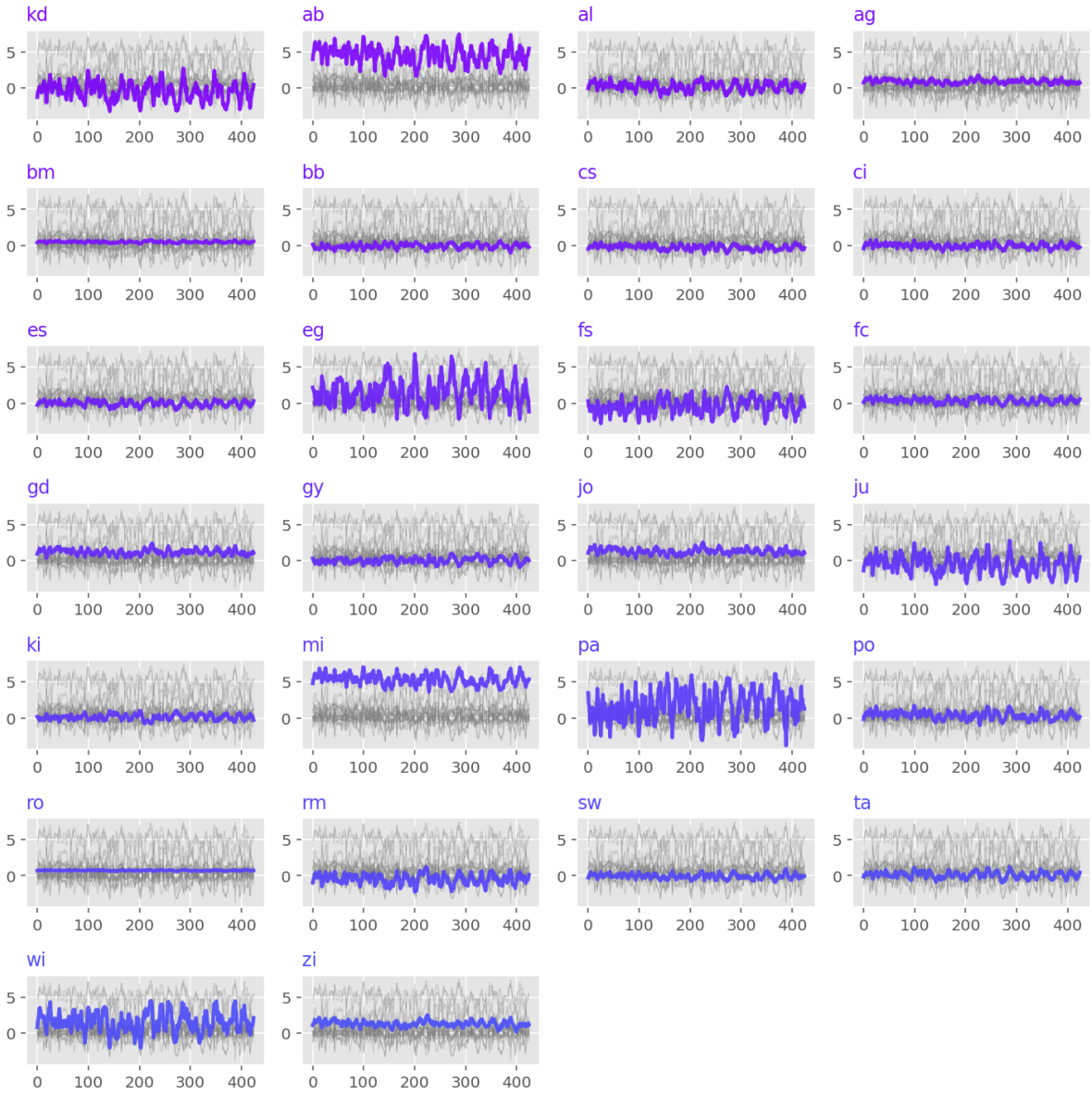
1.6 Protein Scales

Protein scales are a way of measuring certain attributes of residues over the length of the peptide sequence using a sliding window. Scales are comprised of values for each amino acid based on different physical and chemical properties, such as hydrophobicity, secondary structure tendencies, and surface accessibility. As opposed to some chain-level measures like overall molecule behavior, scales allow a more granular understanding of how smaller sections of the sequence will behave.

- kd → Kyte & Doolittle Index of Hydrophobicity
- Flex → Normalized average flexibility parameters (B-values)
- hw → Hopp & Wood Index of Hydrophilicity
- em → Emini Surface fractional probability (Surface Accessibility)

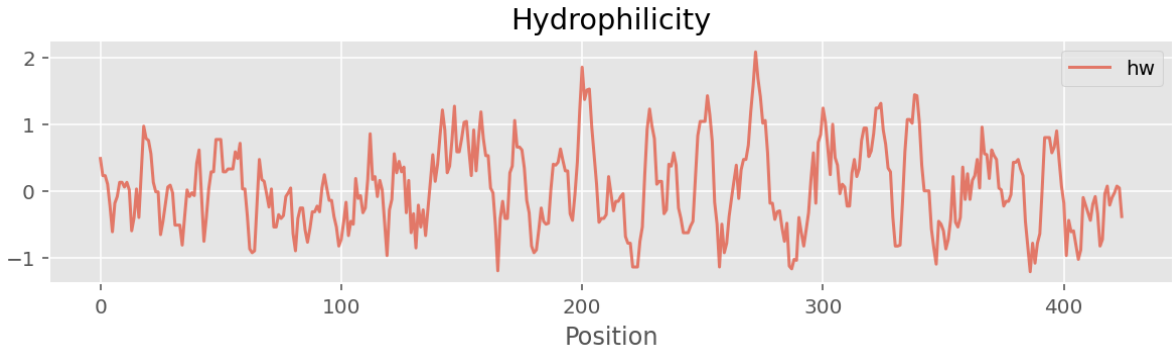
1.6.1 Hydrophobicity index

hydrophobicity is the physical property of a molecule that is seemingly repelled from a mass of water (known as a hydrophobe).



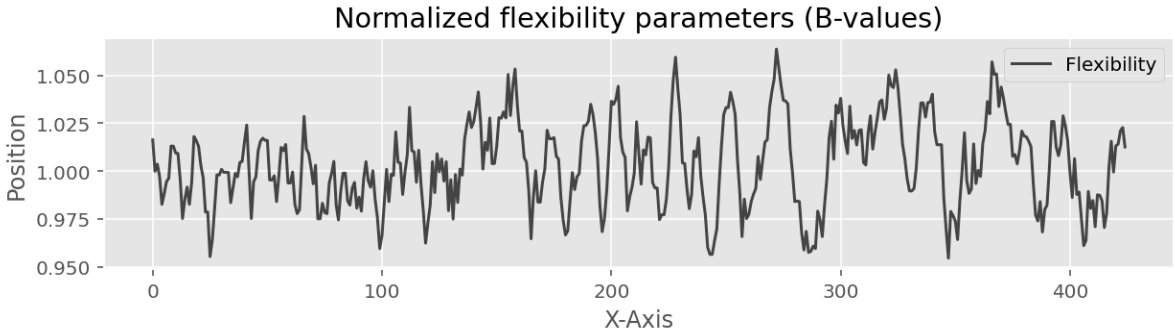
1.6.2 Hydrophilicity index

Hydrophilicity is the tendency of a molecule to be solvated by water.



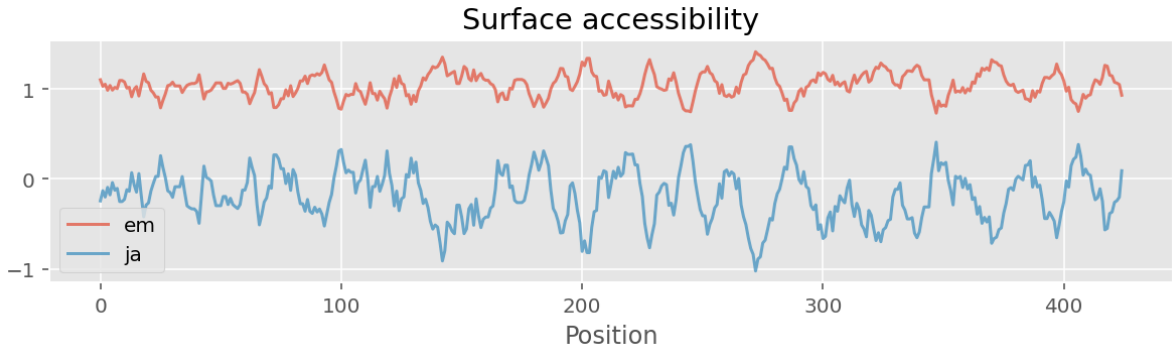
1.6.3 Flexibility index

Proteins are dynamic entities, and they possess an inherent flexibility that allows them to function through molecular interactions within the cell.



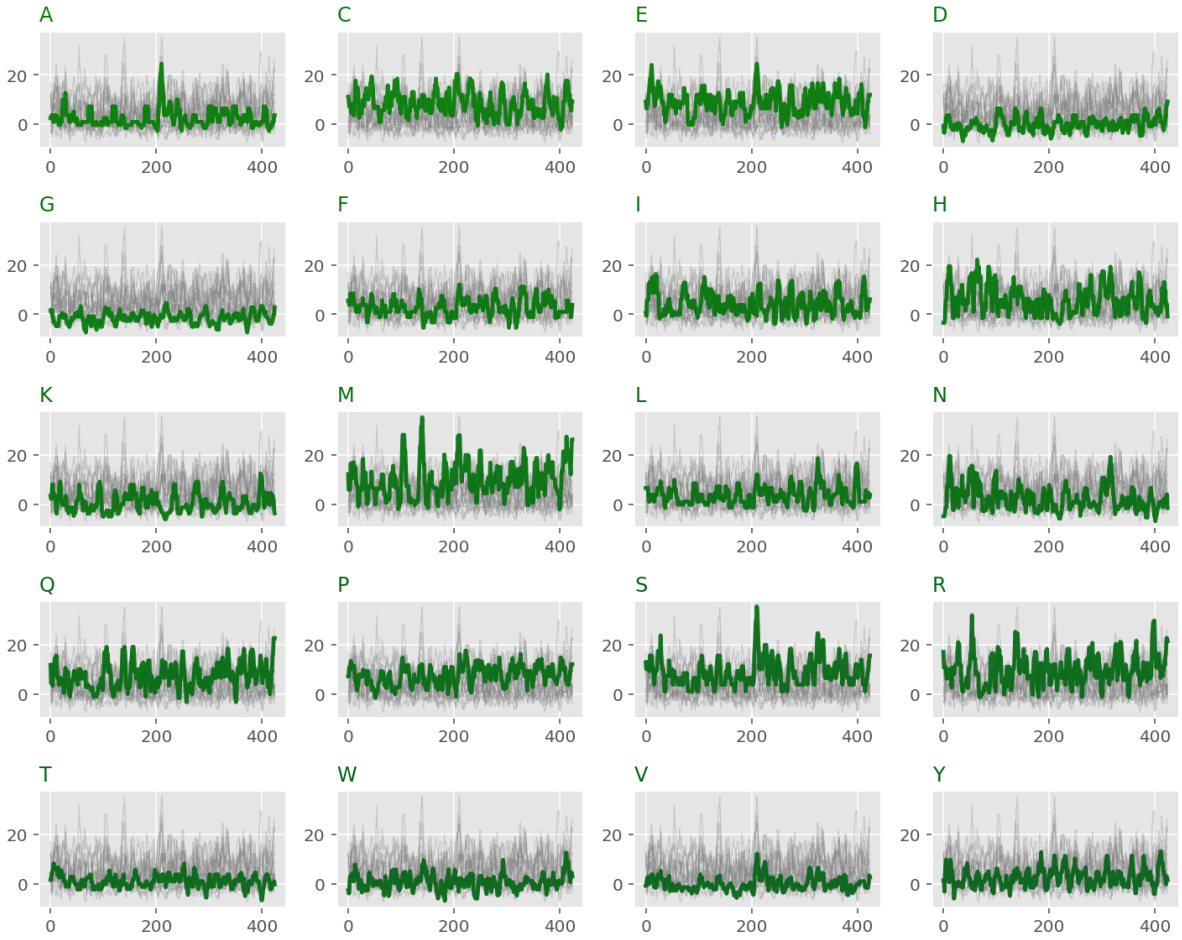
1.6.4 Surface accessibility

Data describing the solvent-accessible surface of a molecule is of great utility in the development of that molecule as a therapeutic, particularly in the case of antibodies. In the context of this report, the most obvious application of molecular surface data is in combination with the potential sites of chemical modification, described in the previous section. Proteins are known to undergo many different chemical modifications as a result of interactions with their aqueous environment. The probability and kinetic rate of such a modification is greatly enhanced by the degree of exposure of the potential modification site to the solvent environment. The solvent-accessible surface for each residue depends upon the degree of exposure of the residue on the surface, but also on the size of the residue side chain.



1.6.5 Instability index

The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed that there are certain dipeptides, the occurence of which is significantly different in the unstable proteins compared with those in the stable ones.



CHAPTER

TWO

IMMUNOGENICITY ANALYSIS

We uses the method of removing and/or reducing potential T-cell epitopes, as an approach to the management of the immunogenicity of biologics. The protein sequence is scanned in silico, for sequences that have a strong binding signature for a family of 50 MHC Class II receptors , whose alleles cover 96 – 98% of the human population. The presented histograms for each variable region sequence, show the average (for the n positively-testing MHC II alleles) of epitope strength at each position as a percentage for all epitopes above a threshold of 20%. At each position in the sequence, the number of alleles scoring above the threshold is shown above the histogram at that position. The epitopes of most concern for the antibody's immunogenicity are therefore those that have not just the highest average score per allele (as shown by the histogram), but which also score above the threshold across more alleles, since these epitopes are more likely to engender an immune response in a larger fraction of the patient population.

Experience using in silico algorithms of this kind in conjunction with laboratory immunogenicity assays has shown that epitopes below this threshold do not generally contribute significantly to the protein's immunogenicity. The number of alleles, the affected alleles and their individual scores are also listed in the detailed analyses below each histogram figure.

The raw immunogenicity score quoted is the total over all epitopes above the threshold for all affected alleles. The normalized immunogenicity score is this raw score divided by the sequence length, and represents epitope strength per unit sequence to enable comparisons of protein sequences of different lengths.

2.1 MHC class 1

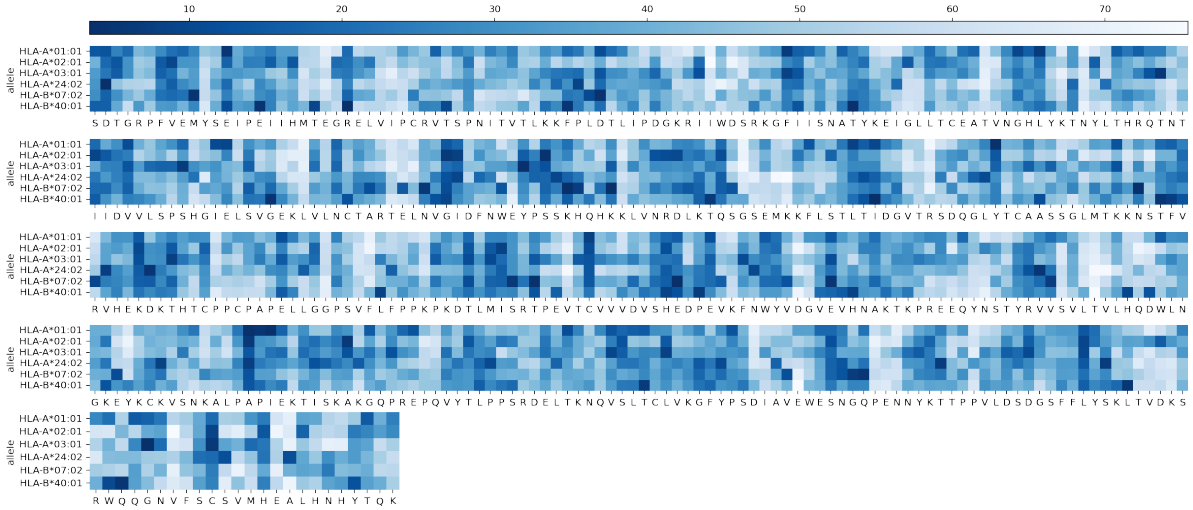
Class I major histocompatibility complex (MHC) molecules bind, and present to T cells, short peptides derived from intracellular processing of proteins. The peptide repertoire of a specific molecule is to a large extent determined by the molecular structure accommodating so-called main anchor positions of the presented peptide.

Their function is to display peptide fragments of proteins from within the cell to cytotoxic T cells; this will trigger an immediate response from the immune system against a particular non-self antigen displayed with the help of an MHC class I protein. Because MHC class I molecules present peptides derived from cytosolic proteins, the pathway of MHC class I presentation is often called cytosolic or endogenous pathway.[1]

- MHC class 1 superset
  - HLA-A01:01, HLA-A02:01, HLA-A03:01, HLA-A24:02, HLA-B07:02, HLA-B40:01

[1] Kimball's Biology Pages, Histocompatibility Molecules

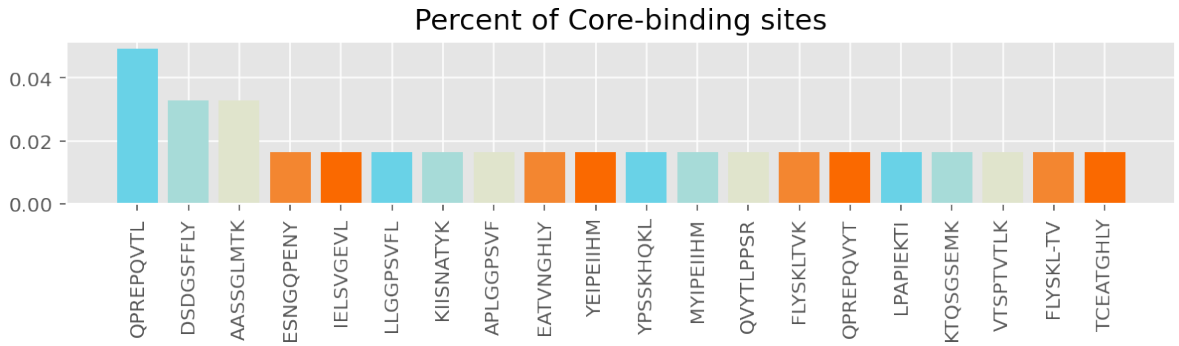
2.1.1 Predicts binding of peptides to MHC class1



2.1.2 Top10 strong binding peptide

	allele	peptide	Core	Rank
0	HLA-A*01:01	DSDGSFFLY	DSDGSFFLY	0.004
1	HLA-A*03:01	ALHNHYTQK	ALHNHYTQK	0.013
2	HLA-B*07:02	QPREPQVYTL	QPREPQVTL	0.013
3	HLA-A*01:01	LDSDGSFFLY	LSDGSFFLY	0.027
4	HLA-A*01:01	VLDSDGSFFLY	VSDGSFFLY	0.028
5	HLA-B*40:01	SEIPEIIHM	SEIPEIIHM	0.030
6	HLA-A*03:01	ATVNGHLYK	ATVNGHLYK	0.030
7	HLA-A*02:01	LMISRTPEV	LMISRTPEV	0.031
8	HLA-B*40:01	IELSVGEKL	IELSVGEKL	0.032
9	HLA-A*03:01	STFVRVHEK	STFVRVHEK	0.038

2.1.3 Frequency of binding peptide

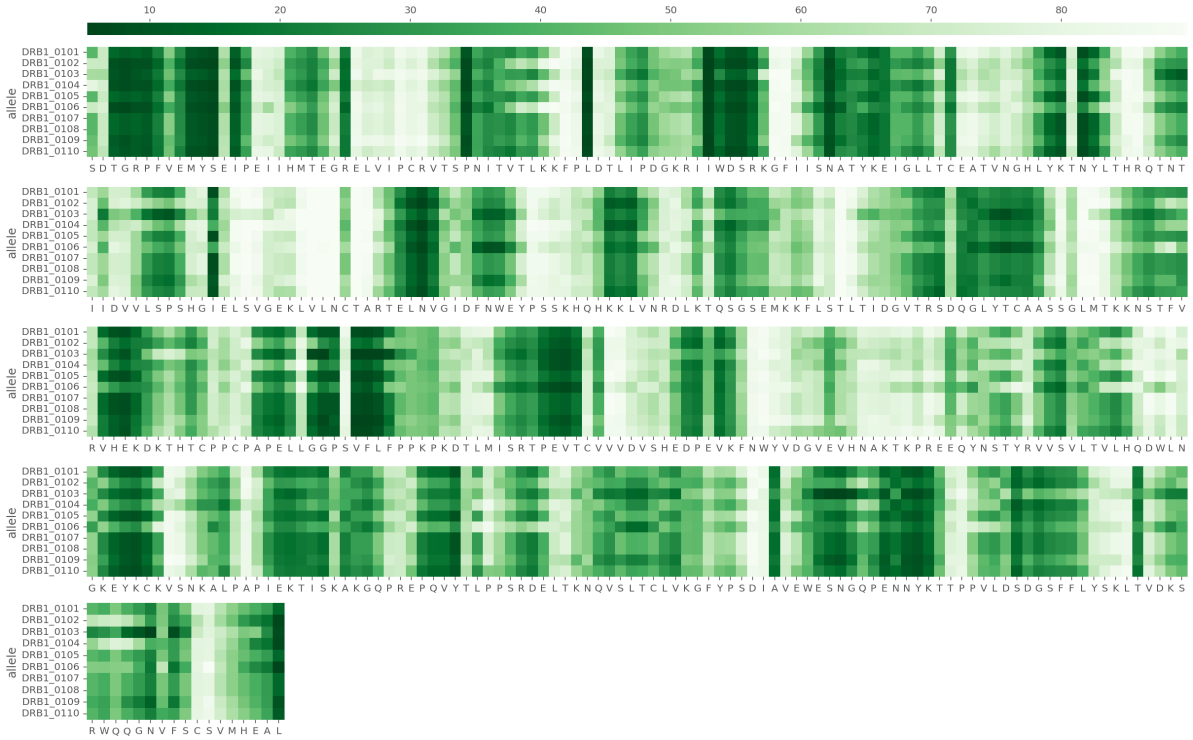


2.2 MHC class 2

MHC Class II molecules are a class of major histocompatibility complex (MHC) molecules normally found only on professional antigen-presenting cells such as dendritic cells, mononuclear phagocytes, some endothelial cells, thymic epithelial cells, and B cells. These cells are important in initiating immune responses.

- MHC class 2 allele superset
  - DRB1_0101,DRB1_0102,DRB1_0103,DRB1_0104,DRB1_0105,DRB1_0106,DRB1_0107,DRB1_0108,DRB1_0109,DRB

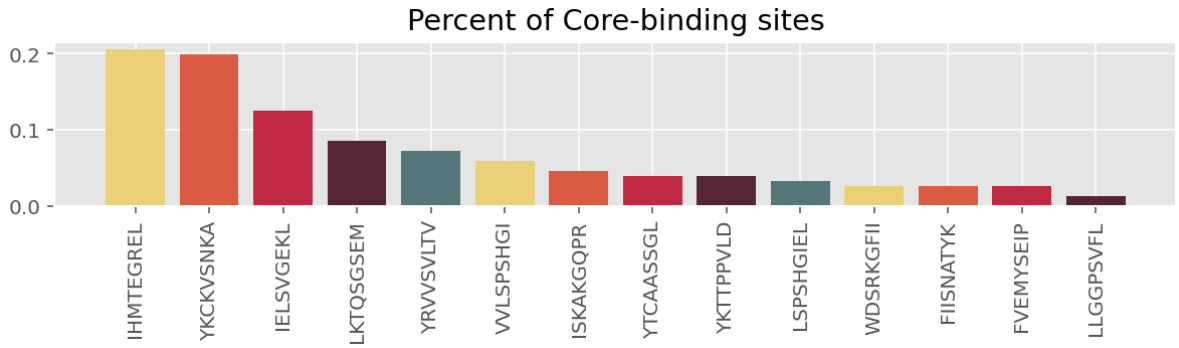
2.2.1 Predicts binding of peptides to MHC class2



2.2.2 Top10 binding peptide

	allele	peptide	Core	Rank
0	DRB1_0103	GKEYKCKVSNKALPA	YKCKVSNKA	0.06
1	DRB1_0103	NGKEYKCKVSNKALP	YKCKVSNKA	0.09
2	DRB1_0103	KEYKCKVSNKALPAP	YKCKVSNKA	0.25
3	DRB1_0103	LNGKEYKCKVSNKAL	YKCKVSNKA	0.33
4	DRB1_0106	PEIIHMTEGRELVIP	IHMTEGREL	0.36
5	DRB1_0109	GKEYKCKVSNKALPA	YKCKVSNKA	0.49
6	DRB1_0104	PEIIHMTEGRELVIP	IHMTEGREL	0.52
7	DRB1_0103	RIIWDSRKGFIISNA	WDSRKGFII	0.57
8	DRB1_0102	PEIIHMTEGRELVIP	IHMTEGREL	0.61
9	DRB1_0109	NGKEYKCKVSNKALP	YKCKVSNKA	0.74

2.2.3 Frequency of binding peptide



2.3 Appendix

nothing yet.