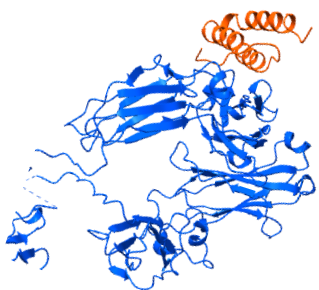




Host-Pathogen Interactions Prediction Using Computational Structural Biology Methods

Ben Shor
Instructor: Dr. Dina Schneidman
October 2021



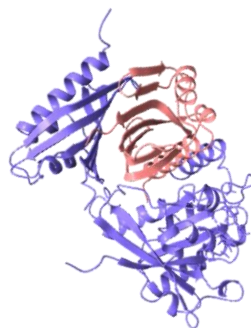
EffectiveDB

Prediction of bacterial protein secretion



HPIDB_{3.0}

Host-Pathogen
Interaction Database

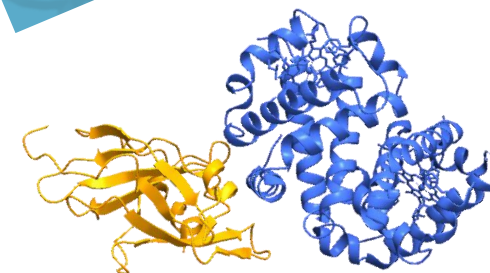


 Google DeepMind

THE HUMAN PROTEIN ATLAS 



Swiss Institute of
Bioinformatics

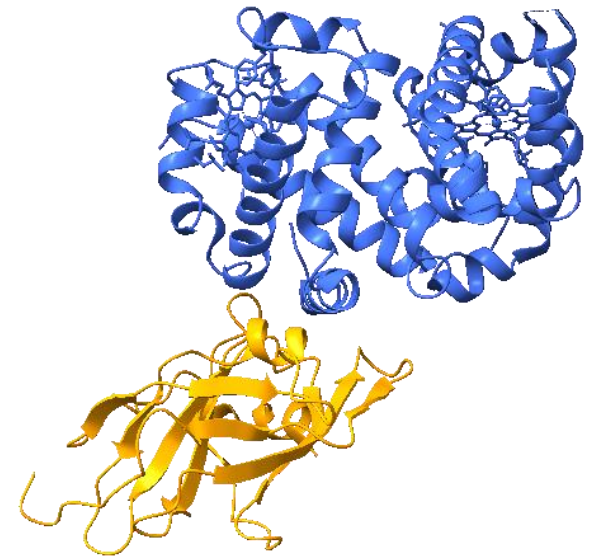


Host-Pathogen Interactions (HPIs)

Pathogen's proteins can **harm** their host (for example, human) by interacting with the host's proteins.

The protein's ability to interact depends on its **structure**.

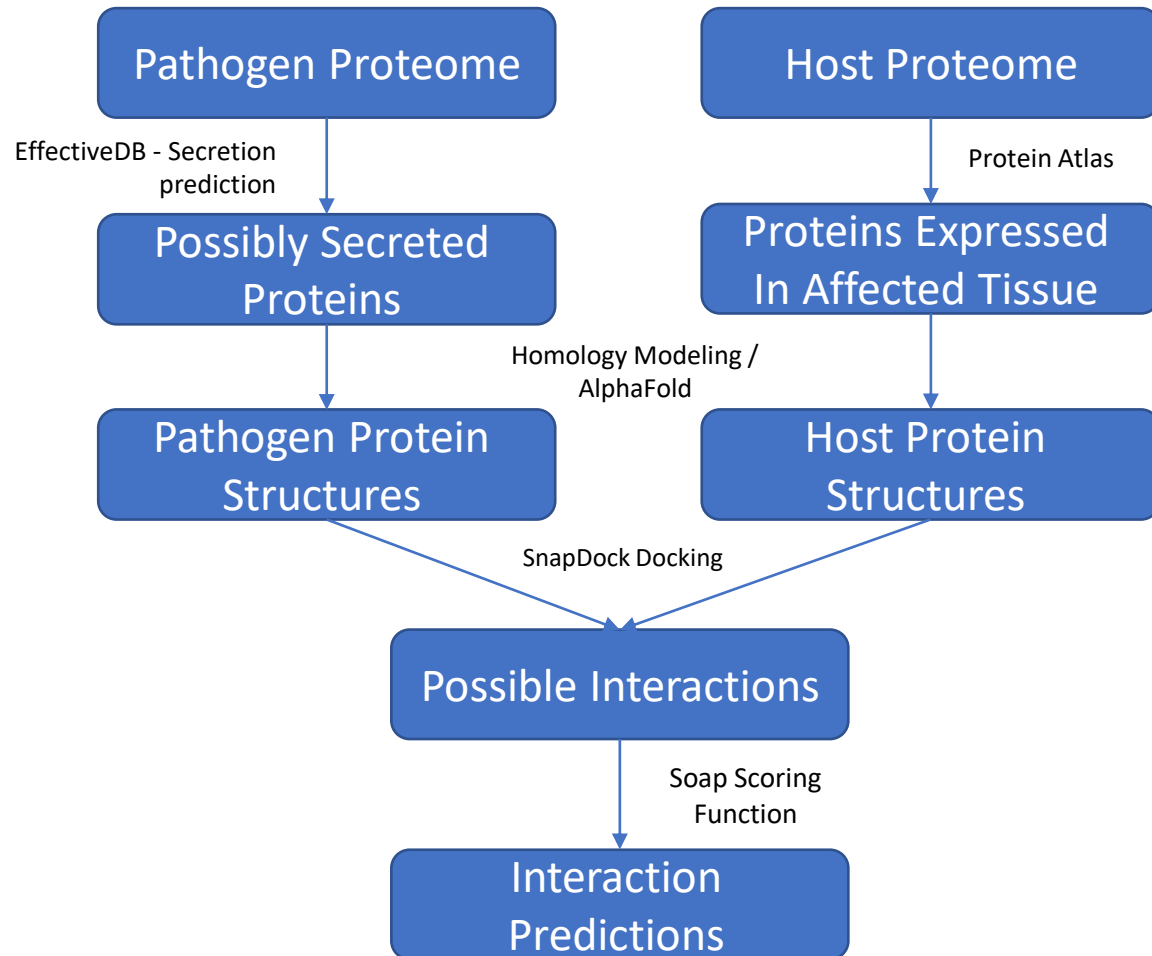
Identifying those host-pathogen interactions (HPIs) enables **better treatment**.



*Pathogen *S. aureus* protein
IsdH (orange) stealing an iron
atom from a human
metHaemoglobin protein (blue)*

PDB:3SZK

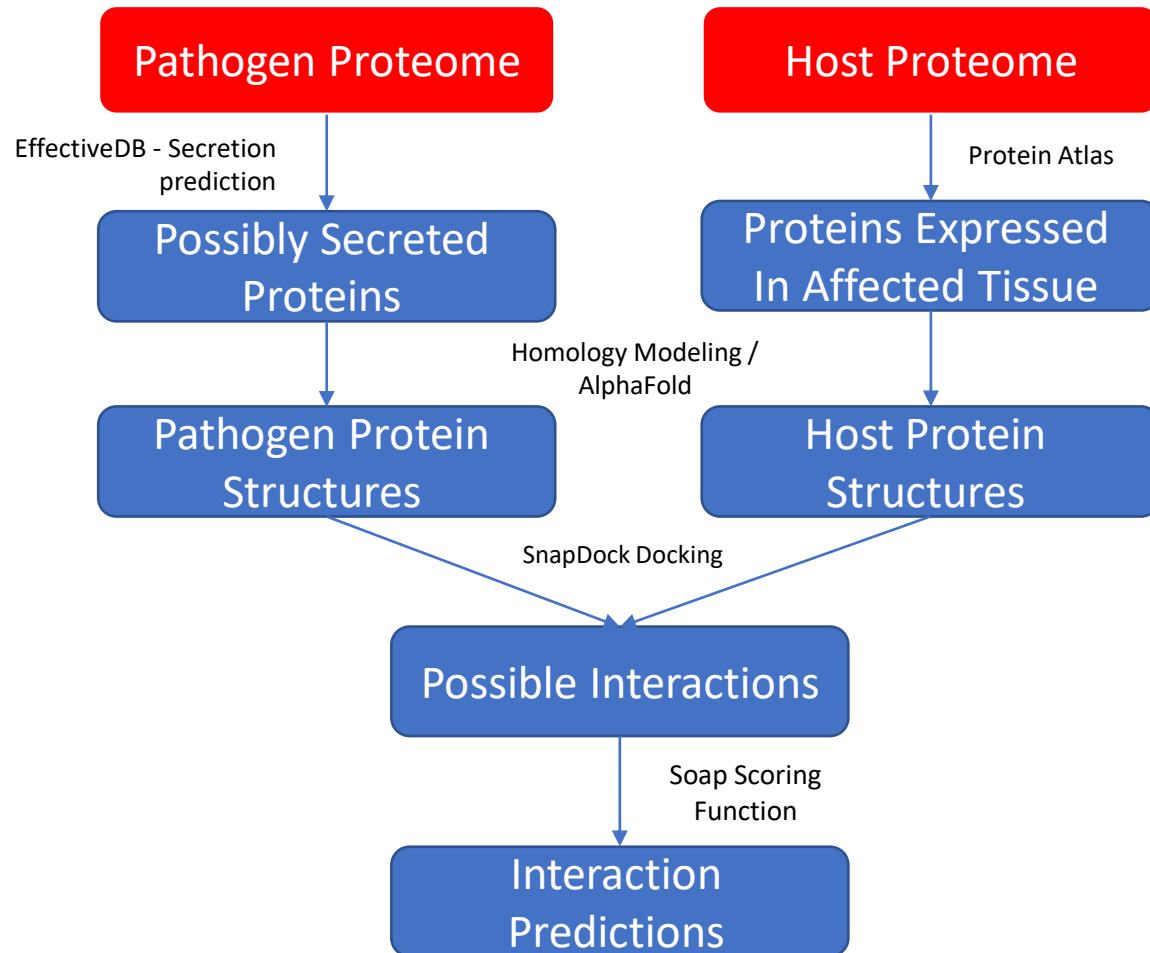
Predicting Host-Pathogen Interactions



We have created a **pipeline** that inputs the entire host and pathogens' proteomes, and outputs a list of predicted host-pathogen interactions.



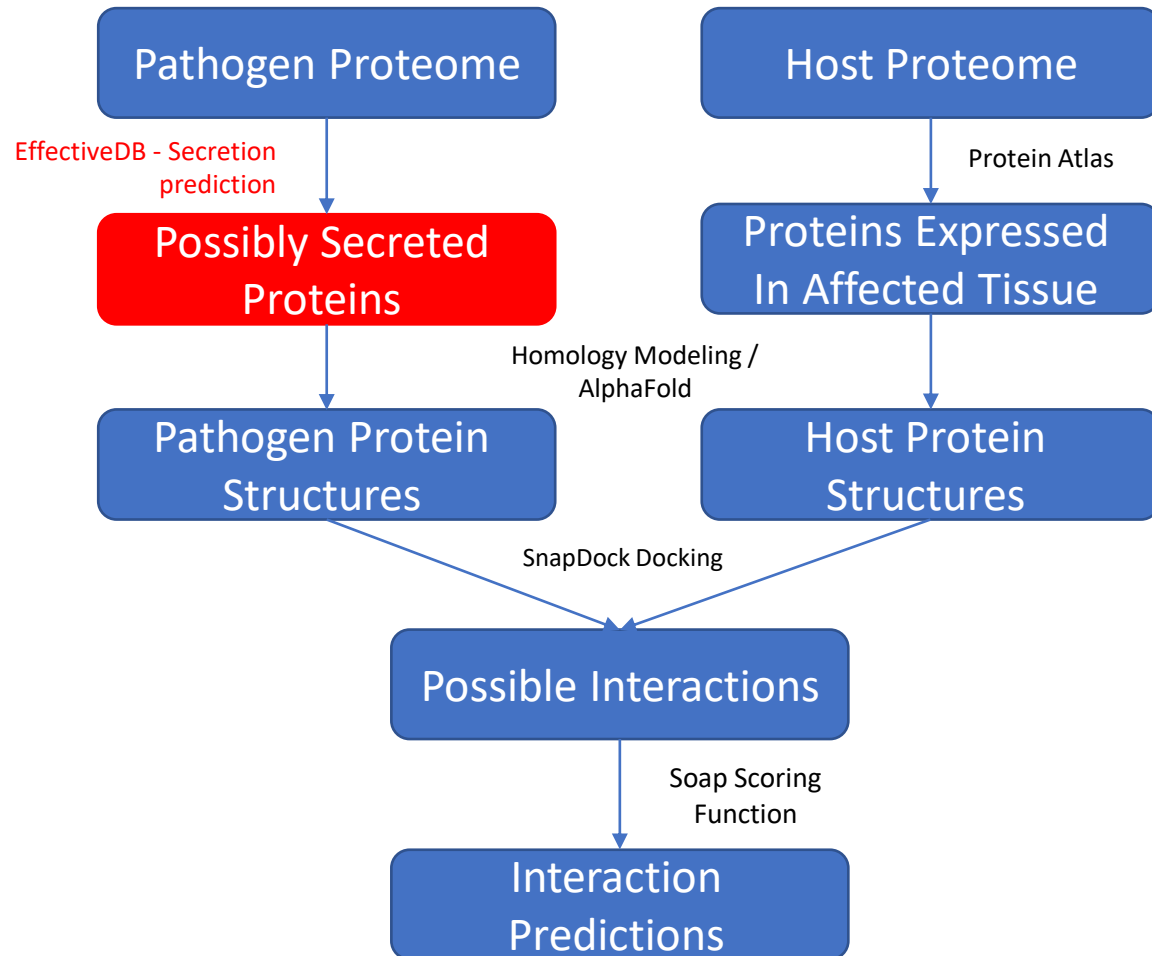
Acquiring Proteomes



Proteome is the complete set of proteins expressed by an organism.

Thanks to sequencing advancements, many organisms have freely available proteome in the UniProt database.

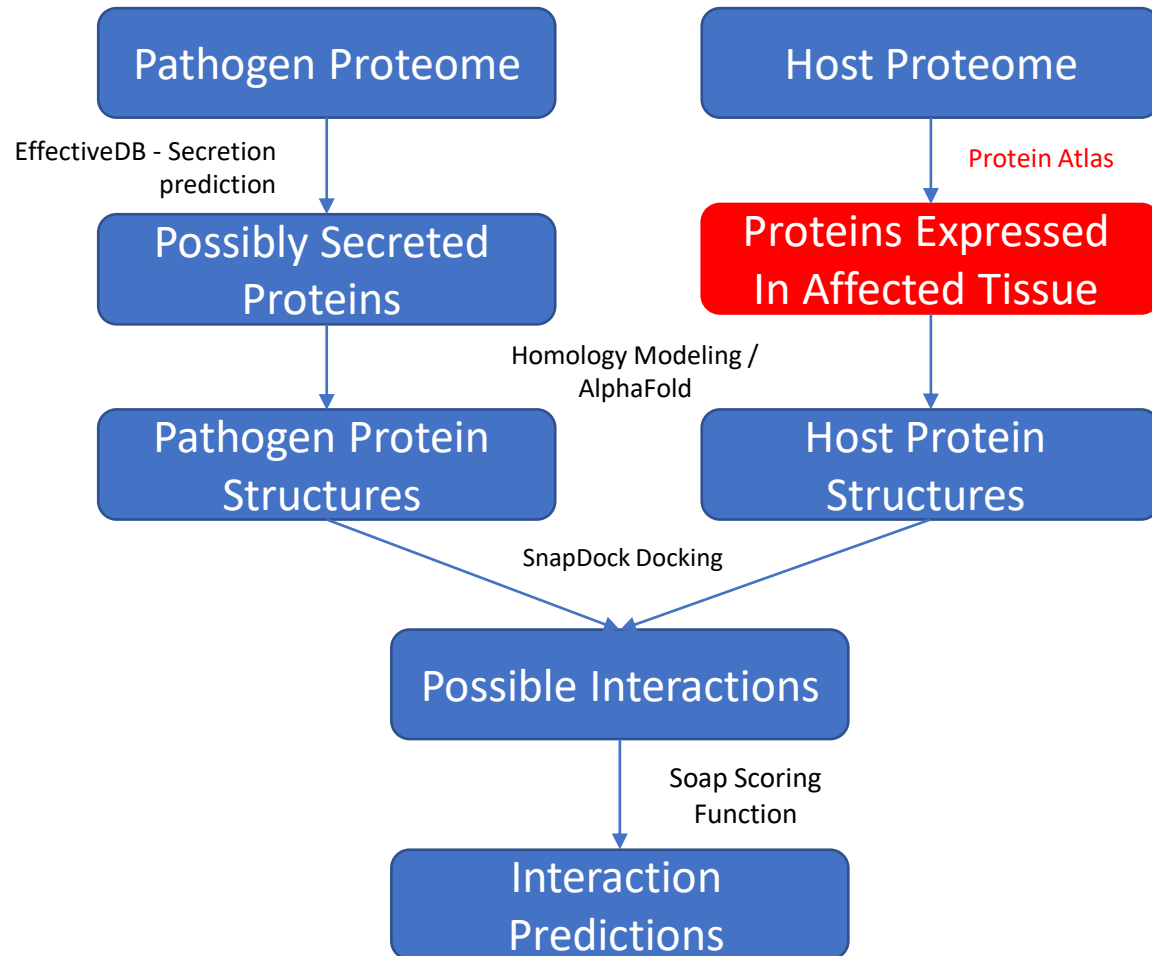
Predicting Secreted Proteins



Secreted protein is a protein that exits the pathogen cell, usually to be a part of HPIs.

We predicted secreted proteins from proteome using EffectiveDB prediction models.

Predicting Host-Pathogen Interactions

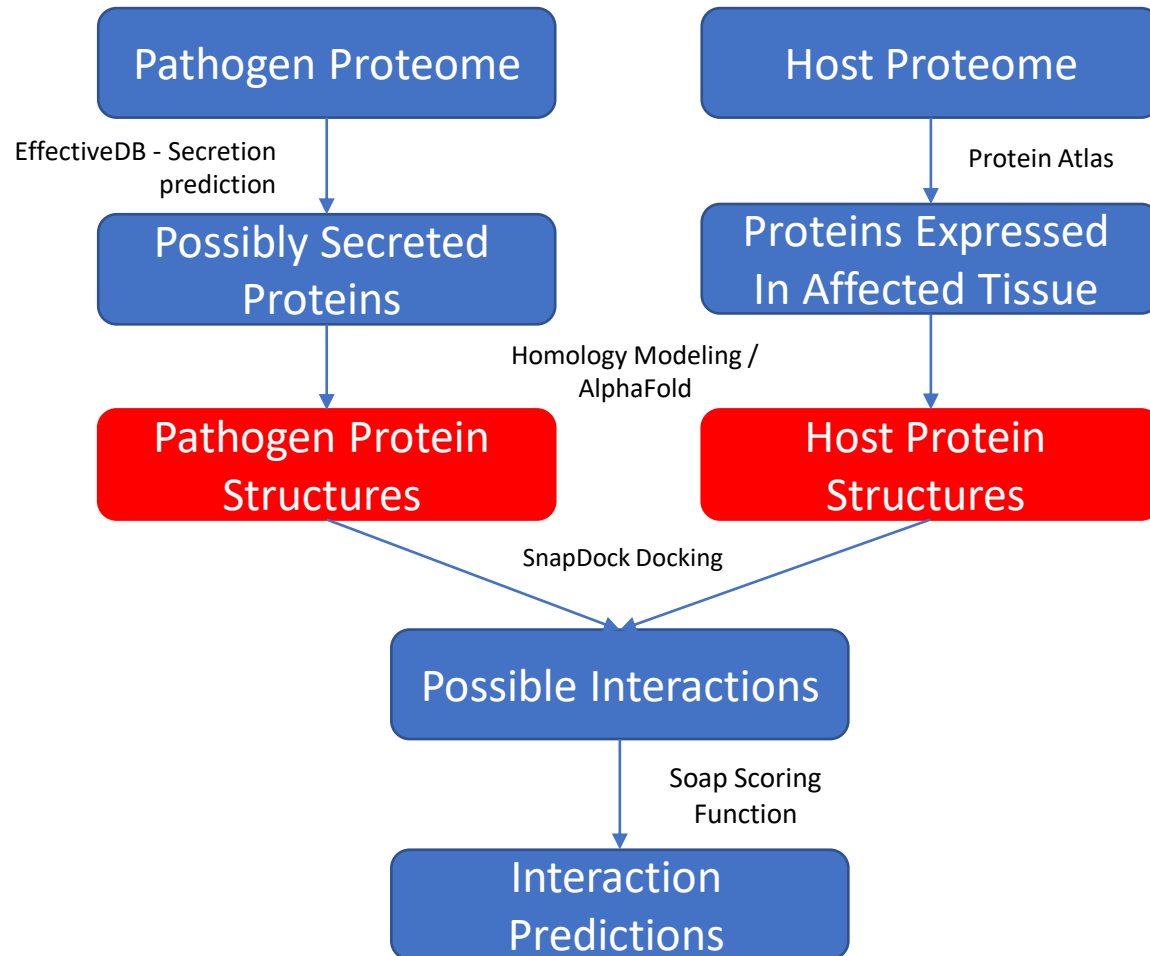


Protein Atlas Database enables querying proteins by tissue expression.

Only host proteins that are in the pathogen's affected tissues were used.

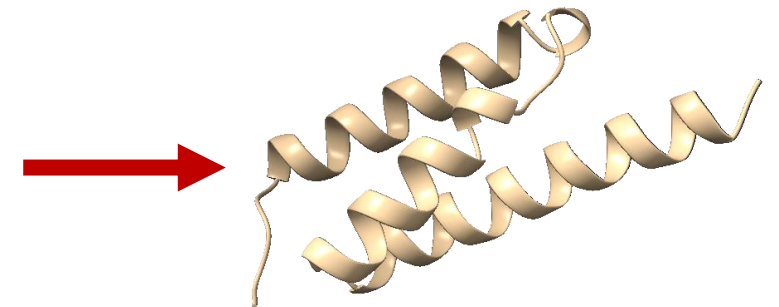
For example, a pathogen that causes pneumonia will only need to be checked against human proteins that are expressed in the lungs.

Predicting Structure

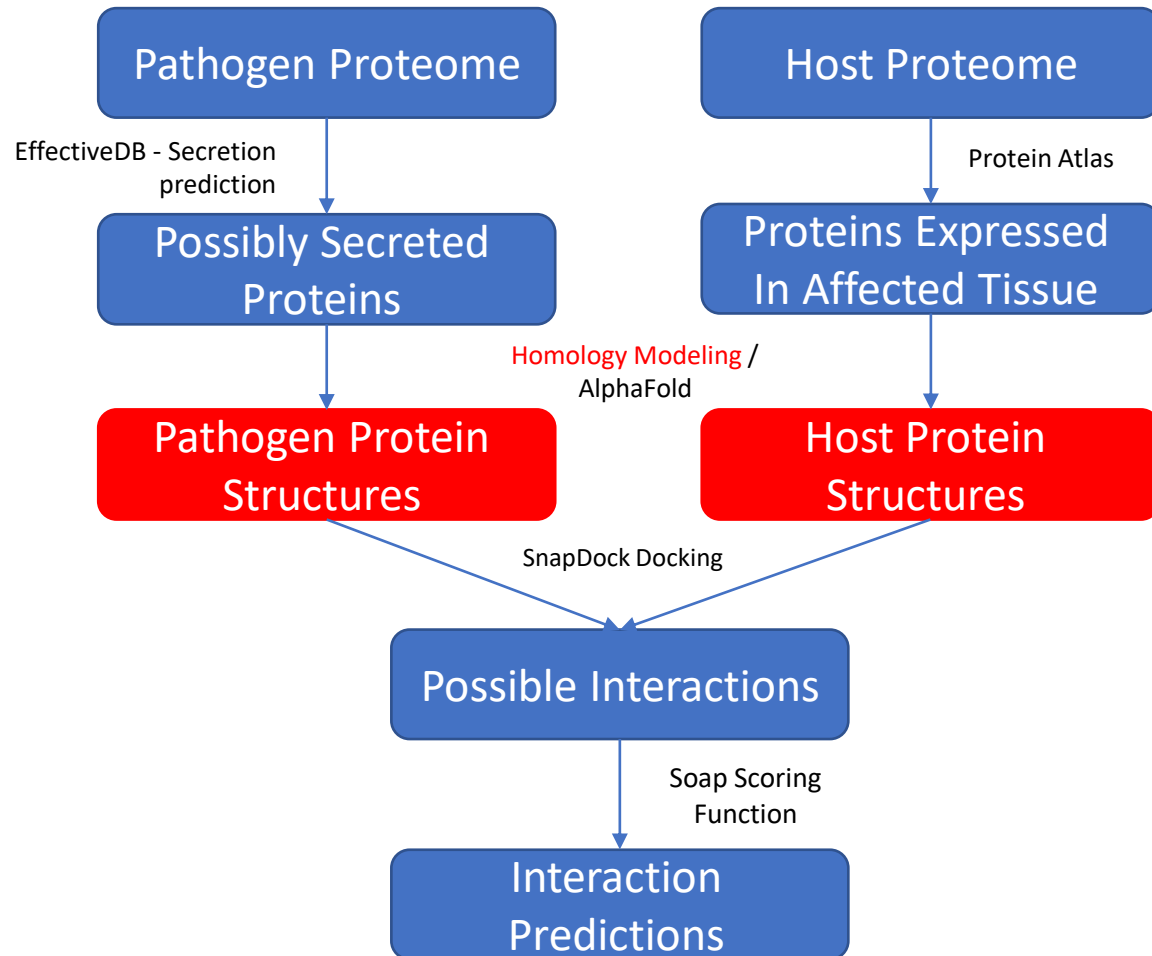


Protein Folding Prediction is the process of predicting the protein's 3D structure based on its sequence of amino acids.

MKNKLIAKSLLTLAAIGITTTTIA
STADASEGYGPREKKPVSIH
NIVEYNDGTFKYQSRPKFNST
PKYIKFKHDYNILEFNDGTFEY
GARPQFNKPAAKTDATIKKEQ
KLIQAQNLVREFEKTHTVSAHR
KAQKAVNLVSFEYKVKKMVLQ
ERIDNVLKQGLVK

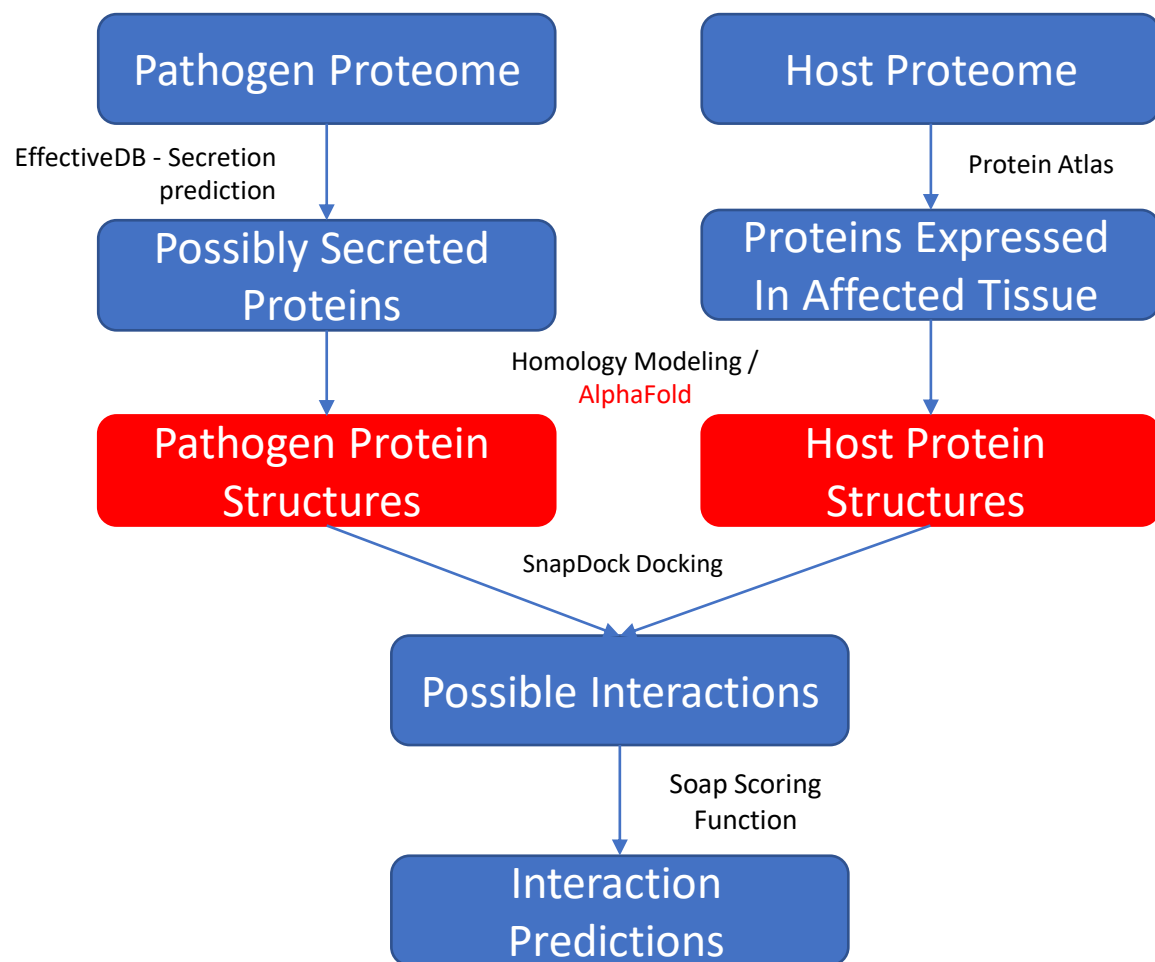


Predicting Structure – Swiss-Model



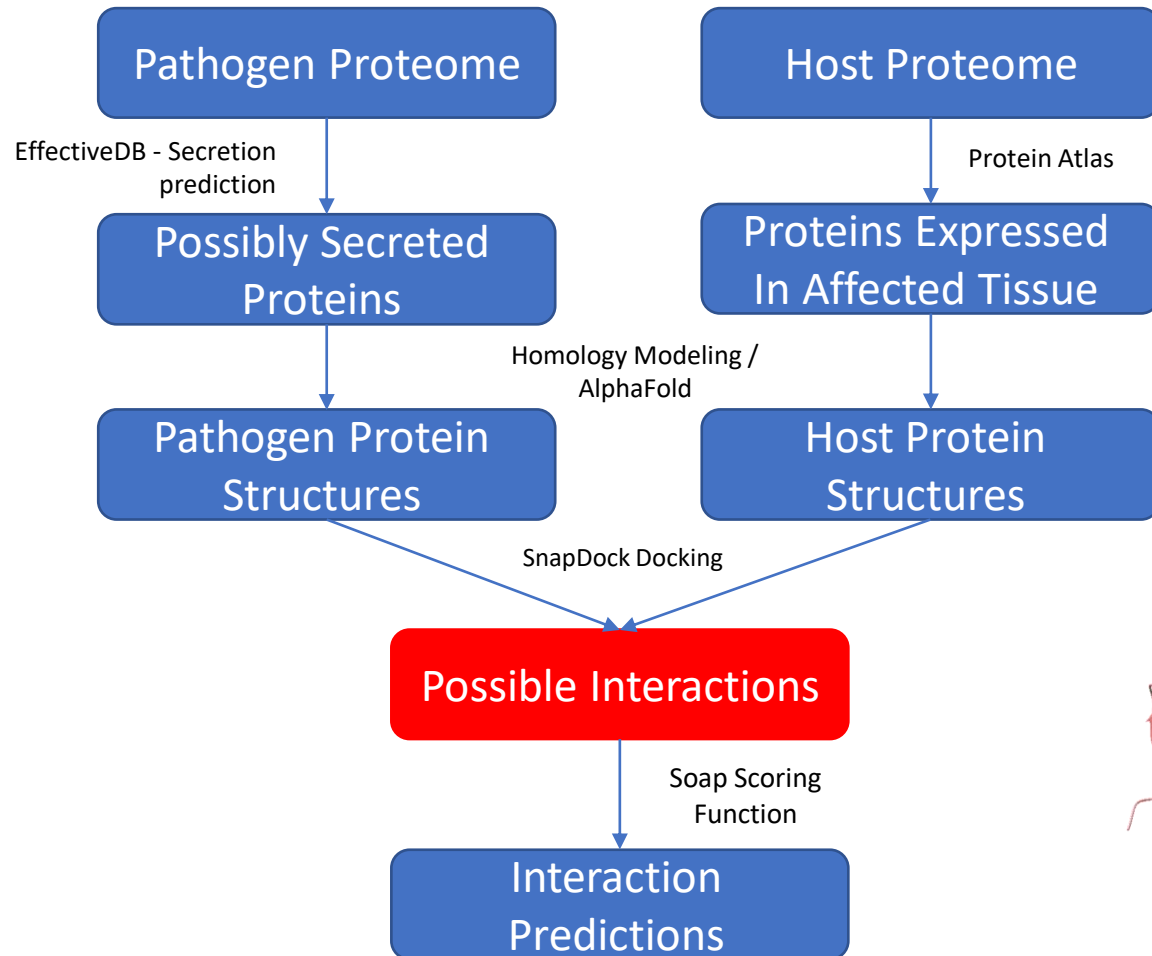
Swiss-Model Service uses **Homology Modeling** – Finding the structure of a target sequence by searching through all proteins with known structures for ones with a similar sequence, and using them as structure templates

Predicting Structure – AlphaFold2

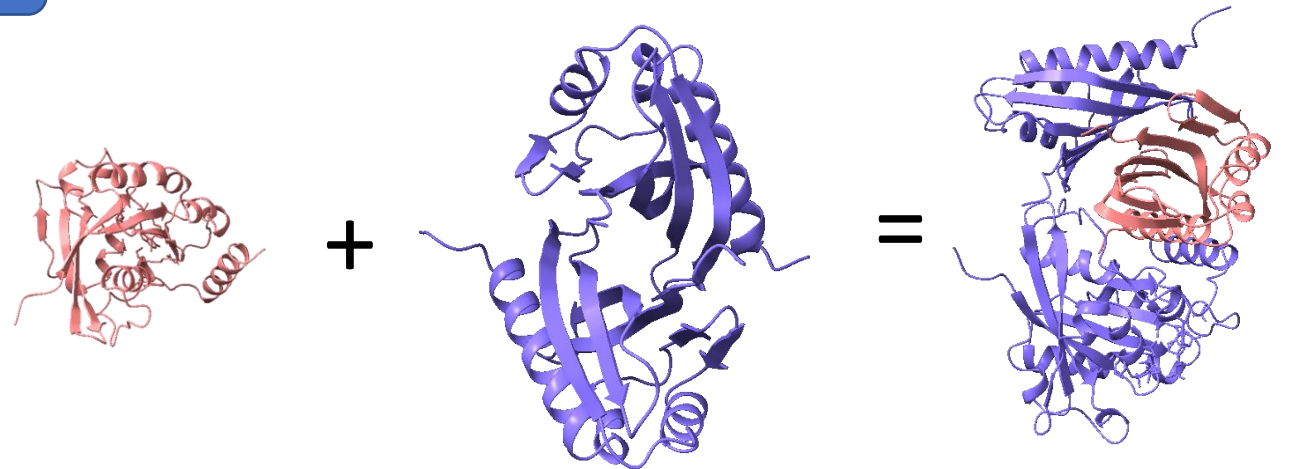


AlphaFold2 is a neural-network based program to predict protein folding. In 2020, it was able to accurately predict the structure of 90% of proteins with unknown structure, practically solving the protein folding problem.

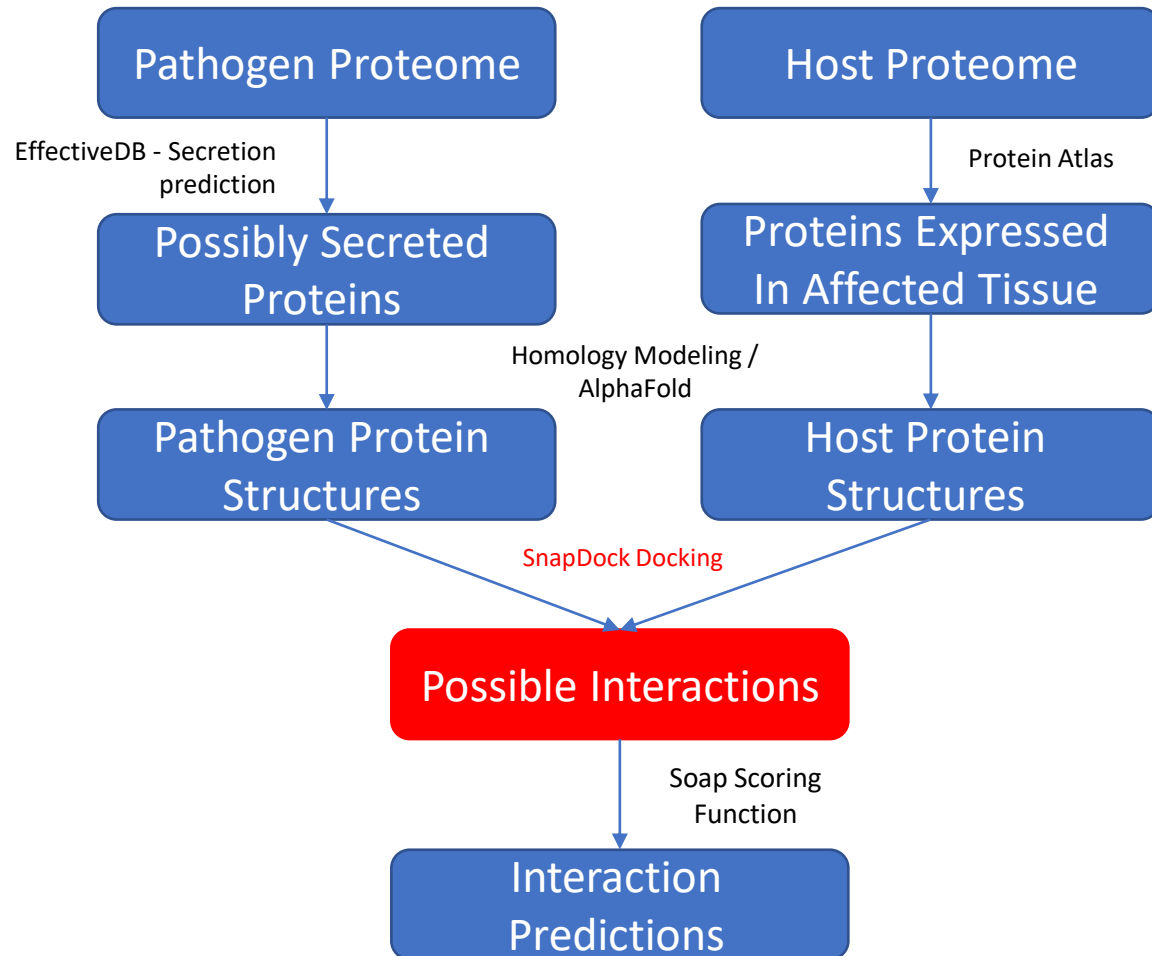
Protein Docking



Protein Docking is the prediction of interactions between two proteins' structures.

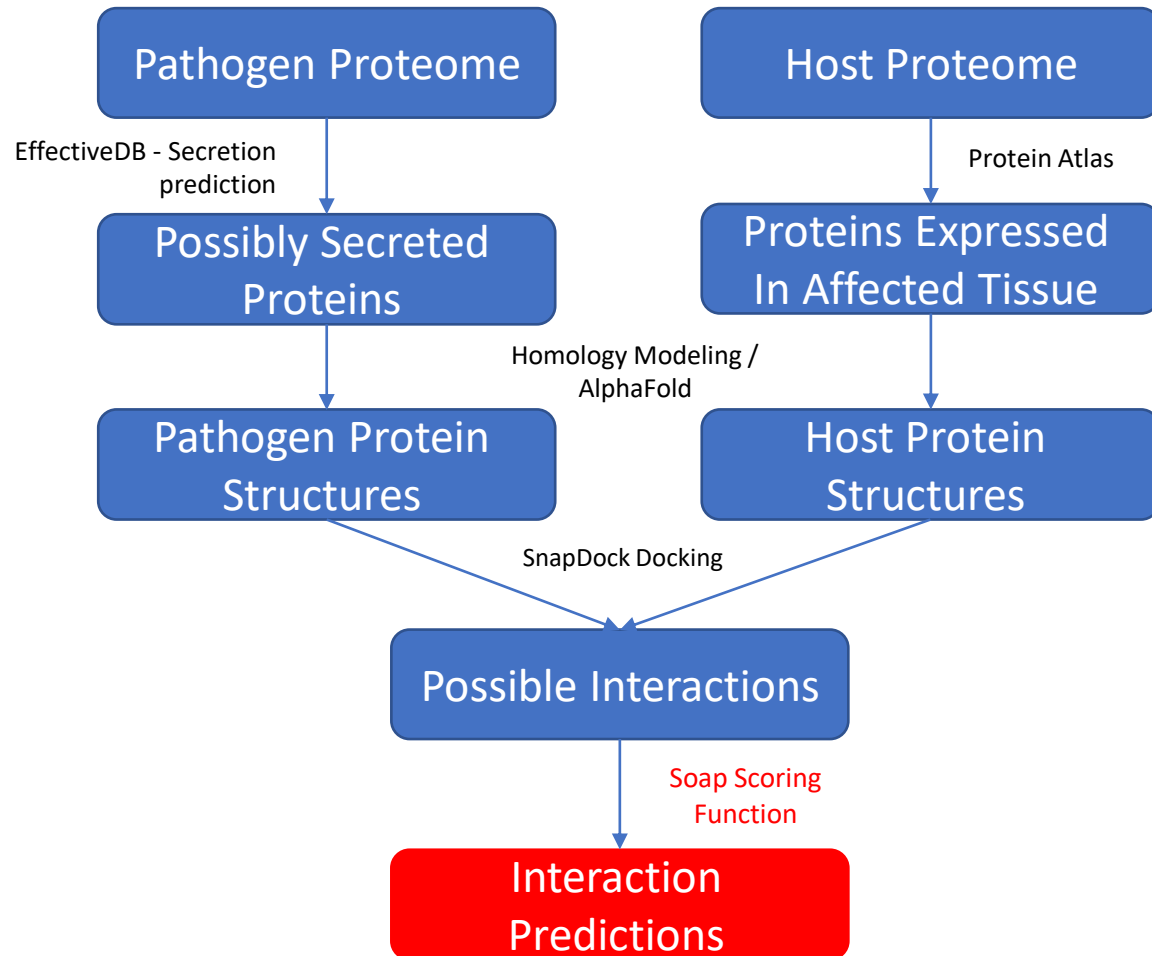


Protein Docking - SnapDock



SnapDock is one of the fastest docking algorithms. It is based on using a list of many protein interactions as templates and matching the target proteins to those templates.

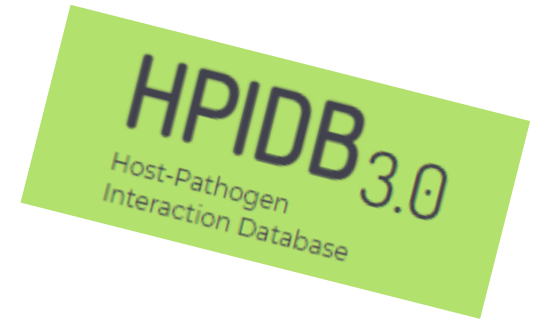
Docking Scoring – SOAP (statistically optimized atomic potentials)



SOAP is an algorithm used to score docked protein-protein structures based on how energetically favorable they are.

It is used as a final step to filter improbable interactions found by docking.

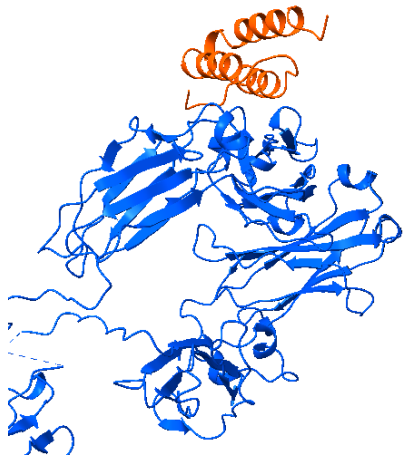
Validating Results



HPI-DB is a database of known HPis found in experiments.

Comparing our predicted interactions to a set of known pathogen's interactions validates our results.

Pipeline Demonstration – *S. aureus*



Example of predicted interaction, that was validated

Staphylococcus aureus is a common pathogen, known to cause several diseases in humans, such as pneumonia.

The described pipeline was used on *S. aureus* and human proteomes.

Over 90,000 possible interactions were tested, **500 predicted interactions were emitted.**

By validating results, it seems that predicted interactions are **13 times more likely** to be real than a regular possible interaction.

Discussion – Challenges

AlphaFold2 practically solves the structure modelling stage, which used to be one of the main challenges.

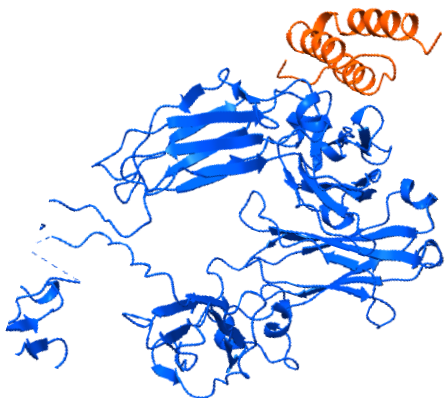
Predicting secreted proteins has low validation rates, and according to our tests can be improved.

The **main challenge** seems to be the **docking stage** – SnapDock has a false negative rate of about 65%, and a false positive rate which is difficult to measure .

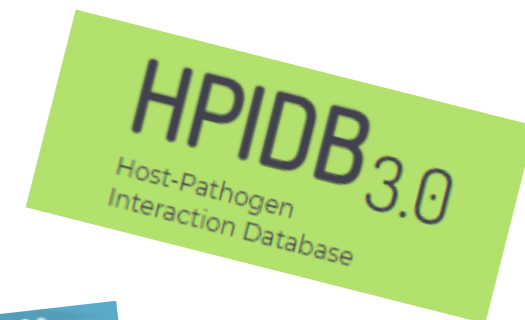
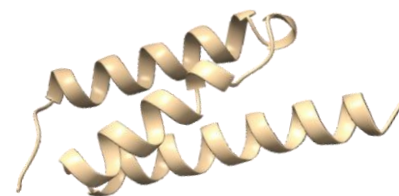
Discussion – Potential

Each stage described is **independent**, researched and advanced separately. Each method can be easily replaced with a new, improved one.

We can assume that with each advancement the ability of pipelines like this to predict HPIs will **improve greatly!**



MKNKLIAKSLLTLAAIGITTTIAS
TADASEGYGPREKKPVSINHNI
VEYNDGTFKYQSRPKFNSTPKY
IKFKHDYNILEFNDGTFEYGARP
QFNKPAAKTDATIKKEQKLIQAO
NLVREFEKTHTVSAHRKAQKAV
NLVSFEYKVKKMVLQERIDNVL
KQGLVK



Thank you for your attention!



THE HUMAN PROTEIN ATLAS

