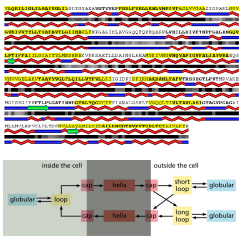
Transmembrane Protein Structure Prediction

Eight

Introduction to Bioinformatics

Sami Khuri
Computer Science
San José State University
June 2016

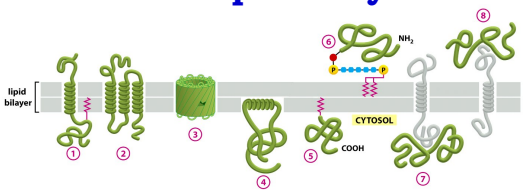
Transmembrane Protein Structure Prediction



- Lipid Bilayer
- Membrane Protein
- Bitopic
- Polytopic
- TMHMM
- HMMTOP
- PROFtmb

©2016 Sami Khuri

Membrane Proteins and Lipid Bilayer



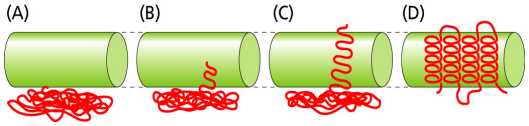
Most transmembrane proteins extend across the lipid bilayer as
1: a single alpha helix, 2: multiple alpha helices,
3: rolled-up beta sheets (beta barrel).

Figure 10-19 *Molecular Biology of the Cell* (© Garland Science 2008)

©2016 Sami Khuri

Types of Membrane Proteins

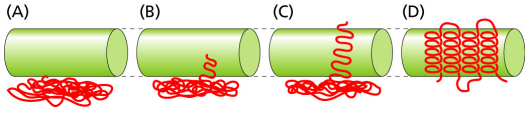
- Membrane proteins can be categorized by their degree of interaction with the membrane.



©2016 Sami Khuri

Membrane Proteins

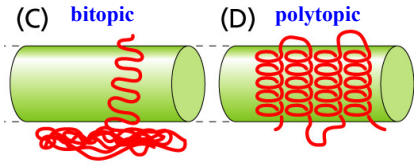
- Some are only anchored to one side of the membrane. See A and B.
  - These follow the general structural rules of proteins.



©2016 Sami Khuri

Transmembrane Proteins (I)

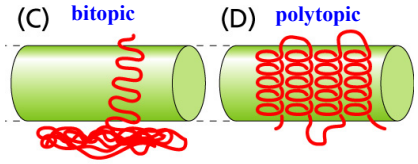
- Transmembrane or integral proteins have a part that is entirely embedded within the lipid bilayer.



©2016 Sami Khuri

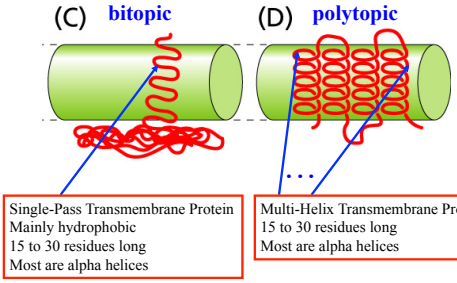
Transmembrane Proteins (II)

- Knowing a membrane protein's topology can be a significant step toward inferring both its structure and function.



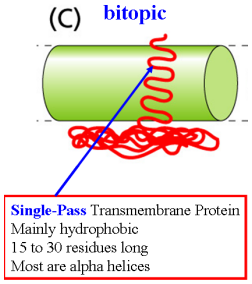
©2016 Sami Khuri

Transmembrane Proteins (III)



©2016 Sami Khuri

Single-Pass Transmembrane Proteins (I)



Hydrophobicity scales are used to assign values to individual residues. The values are converted into **hybrophobic profiles** by using a sliding window to average the values over a number of residues.

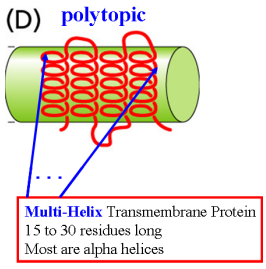
©2016 Sami Khuri

Single-Pass Transmembrane Proteins (II)

- There are many different **hydrophobicity scales**.
  - They produce different results.
  - Therefore, one has to use several transmembrane predictors.
- This method works pretty well for **single-pass transmembranes**.

©2016 Sami Khuri

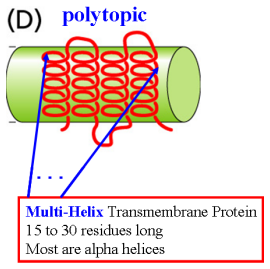
Multi-Helix Transmembrane Proteins (I)



Helices contain both **hydrophobic** and **charged residues**, forming a structural element that has a different character on each side – an **amphipathic helix**.

©2016 Sami Khuri

Multi-Helix Transmembrane Proteins (II)



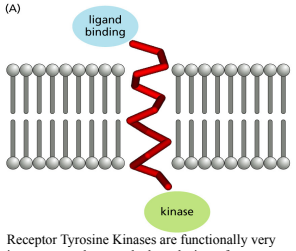
Use of hydrophobic profiles only will not suffice to guarantee good predictions.

The **hydrophobic moment** is used.

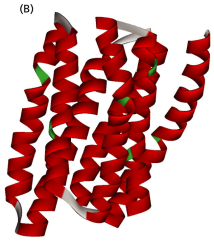
It measures the **hybrophobicity** of a peptide at different angles of rotation.

©2016 Sami Khuri

Functions of Membrane Proteins



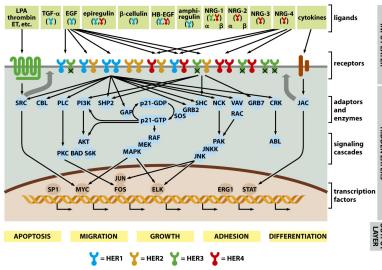
Receptor Tyrosine Kinases are functionally very important as they are the launch sites of many complex signal transduction pathways in the cell.



A seven-transmembrane spanning molecule.

©2016 Sami Khuri

The ErbB Signaling Network

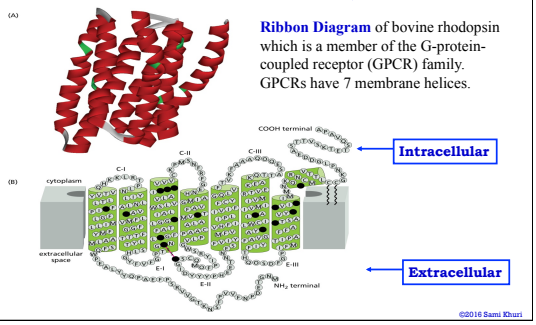


Normal cells receive growth-stimulatory signals from their surroundings. These signals are processed and integrated by complex circuits within the cell, which decide whether cell growth and division is appropriate or not.

Biology of Cancer by R. Weingberg

©2016 Sami Khuri

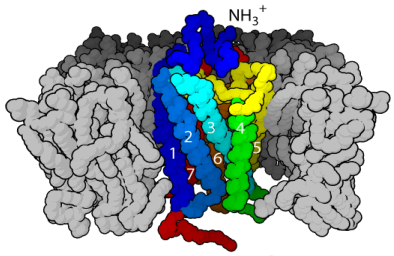
Bovine Rhodopsin



Ribbon Diagram of bovine rhodopsin which is a member of the G-protein-coupled receptor (GPCR) family. GPCRs have 7 membrane helices.

©2016 Sami Khuri

GPCR (I)



G-Protein-Coupled Receptors (GPCRs) share a conserved structure composed of seven transmembrane (TM) helices

en.wikipedia.org/wiki/G_protein-coupled_receptor

©2016 Sami Khuri

GPCR (II)

- **G-protein-coupled receptors** (**GPCRs**) constitute a large and diverse family of proteins whose primary function is to transduce extracellular stimuli into intracellular signals.
- **GPCRs** are among the largest and most diverse protein families in mammalian genomes.
- On the basis of homology with rhodopsin, **GPCRs** are predicted to contain seven membrane-spanning helices, an extracellular N-terminus and an intracellular C-terminus.
- This gives rise to **GPCRs** other names, the 7-TM receptors or the heptahelical receptors.

www.ibibiobase.com/projects/db-drd4/G_protein.htm

©2016 Sami Khuri

GPCR (III)

- **GPCRs** transduce extracellular stimuli to give intracellular signals through interaction of their intracellular domains with heterotrimeric G proteins.
- This class of membrane proteins can respond to a wide range of agonists, including photon, amines, hormones, neurotransmitters and proteins.
- Some agonists bind to the extracellular loops of the receptor, others may penetrate into the transmembrane region.

www.ibibiobase.com/projects/db-drd4/G_protein.htm

©2016 Sami Khuri

Cystic Fibrosis

- **Cystic Fibrosis** is an autosomal recessive disorder that affects the respiratory and digestive systems.
- **Cystic Fibrosis** is associated with mutations in the **CFTR** (**C**ystic **F**ibrosis **T**ransmembrane **R**egulator) gene.
- **Cystic Fibrosis** is fatal and treatment is limited to slowing the progress of the disease.

©2016 Sami Khuri

A vest designed to improve lung function for cystic fibrosis patients

©2016 Sami Khuri

CFTR Protein

- The **CFTR** protein is found in the membrane of epithelial cells.
- It forms a channel through which chloride ions (Cl^-) can pass.
- The channel can be opened or closed.
- The flow of ions is necessary for water to be released into secretions such as mucus in the lungs.

©2016 Sami Khuri

Function of CFTR Gene (I)

- The CFTR gene provides instructions for making a protein called the cystic fibrosis transmembrane conductance regulator. This protein functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes.
- The channel transports negatively charged particles called chloride ions into and out of cells.

http://ghr.nlm.nih.gov/gene=cftr

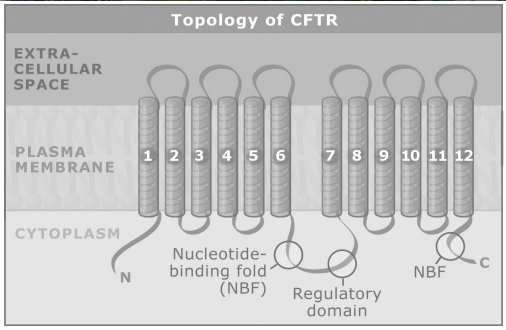
©2016 Sami Khuri

Function of CFTR Gene (II)

- The transport of chloride ions helps control the movement of water in tissues, which is necessary for the production of thin, freely flowing mucus.
- Mucus is a slippery substance that lubricates and protects the lining of the airways, digestive system, reproductive system, and other organs and tissues.

http://ghr.nlm.nih.gov/gene=cftr

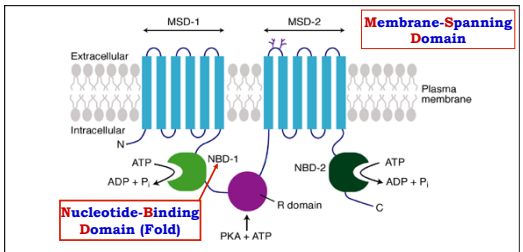
©2016 Sami Khuri



Schematic diagram showing the structure of the CFTR protein that regulates transport of chloride through cell membranes

©2016 Sami Khuri

CFTR



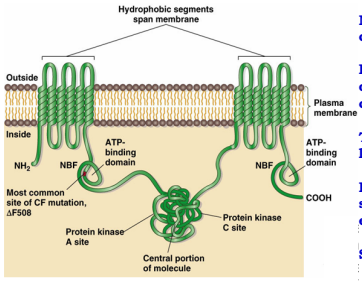
A model of the proposed structure of the cystic fibrosis transmembrane conductance regulator (CFTR)

Expert Reviews in Molecular Medicine ©2001 Cambridge University Press

journals.cambridge.org

©2016 Sami Khuri

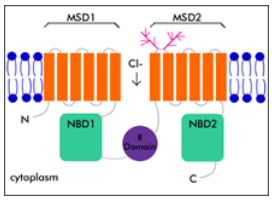
ΔF508 Mutation in CFTR



- **Missing or altered chloride transporter**
- **Insufficient secretion of water by epithelial cells**
- **Thick lung secretions lead to lung infections**
- **Defective pancreatic secretion of digestive enzymes**
- **Sterility in males**

©2016 Sami Khuri

CFTR Protein's Fate



CFTR Sequence:

Nucleotide	ATC	ATC	TTT	GGT	GTT
Amino Acid	Ile 506	Ile	Phe 508	Gly	Val 510

Deleted in ΔF508

ΔF508 CFTR Sequence:

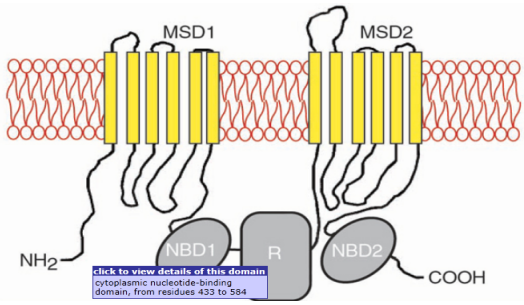
Nucleotide	ATC	ATT	GGT	GTT
Amino Acid	Ile 506	Ile	Gly	Val

With normal CFTR, once the protein is synthesized, it is transported to the endoplasmic reticulum (ER) and Golgi apparatus for additional processing before being integrated into the cell membrane. When a CFTR protein with the delta F508 mutation reaches the ER, the quality-control mechanism of this cellular component recognizes that the protein is folded incorrectly and marks the defective protein for degradation. As a result, delta F508 never reaches the cell membrane.

www.ornl.gov/sci/techresources/Human_Genome/posters/chromosome/cftr.shtml

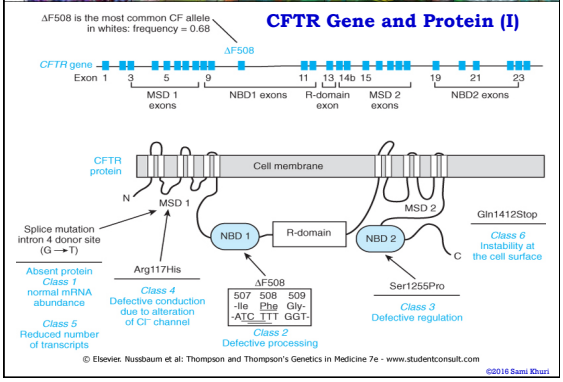
©2016 Sami Khuri

Location of the ΔF508



©2016 Sami Khuri

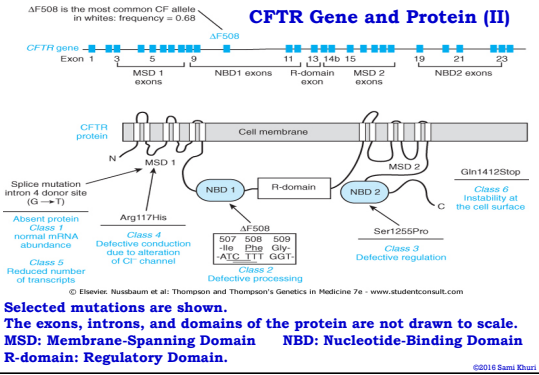
CFTR Gene and Protein (I)



© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

©2016 Sami Khuri

CFTR Gene and Protein (II)



© Elsevier. Nussbaum et al: Thompson and Thompson's Genetics in Medicine 7e - www.studentconsult.com

Selected mutations are shown.
The exons, introns, and domains of the protein are not drawn to scale.
MSD: Membrane-Spanning Domain NBD: Nucleotide-Binding Domain
R-domain: Regulatory Domain.

©2016 Sami Khuri

Cystic Fibrosis (I)

- More than 1,000 mutations in the **CFTR gene** have been identified in people with **cystic fibrosis**.
- Most of these mutations change single protein building blocks (amino acids) in the **CFTR protein** or delete a small amount of DNA from the **CFTR gene**.
- The most common mutation, called delta F508, is a deletion of one amino acid at position 508 in the **CFTR protein**.

http://ghr.nlm.nih.gov/gene=cftr

©2016 Sami Khuri

Cystic Fibrosis (II)

- The resulting abnormal channel breaks down shortly after it is made, so it never reaches the cell membrane to transport chloride ions.
- Disease-causing mutations in the **CFTR gene** alter the production, structure, or stability of the chloride channel.
- All of these changes prevent the channel from functioning properly, which impairs the transport of chloride ions and the movement of water into and out of cells.

http://ghr.nlm.nih.gov/gene=cftr

©2016 Sami Khuri

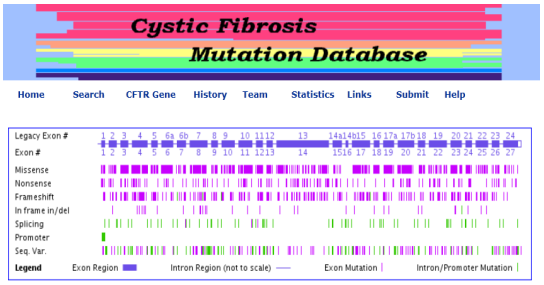
Cystic Fibrosis (III)

- As a result, cells that line the passageways of the lungs, pancreas, and other organs produce mucus that is abnormally thick and sticky.
- The abnormal mucus obstructs the airways and glands, leading to the characteristic signs and symptoms of cystic fibrosis.

http://ghr.nlm.nih.gov/gene=cftr

©2016 Sami Khuri

CF Mutation Database



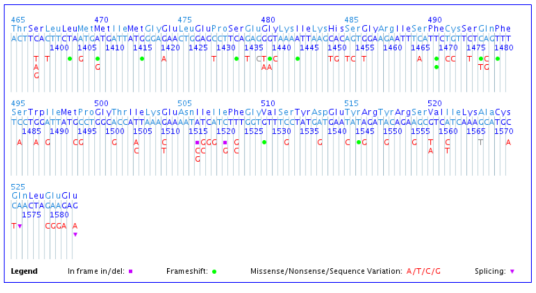
"Sequence Variation" is sometimes designated as "polymorphism", indicating that it is "non-disease causing". According to the general definition in human genetics, a "polymorphism" has to reach an allelic frequency of 1%. In addition, when a sequence variation is found in one single individual, it is not possible to determine if it is "non-disease causing"

©2016 Sami Khuri

Exon 11 of the CFTR Gene

Detailed View of exon 11

Get the summary of a mutation by putting your mouse over that mutation. Click to view the details of that mutation.



©2016 Sami Khuri

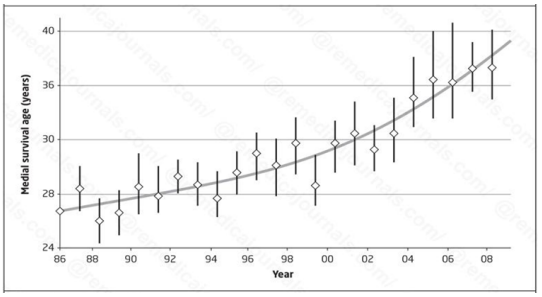
Cystic Fibrosis Carrier Frequency

Ashkenazi Jewish	1 in 24
Caucasian	1 in 25
Hispanic	1 in 46
African-American	1 in 65
Asian	1 in 94

http://progenity.com/cystic-fibrosis-cf

©2016 Sami Khuri

Median Survival Age



Cystic Fibrosis in the 21st Century by N Simmonds ©2016 Sami Khuri

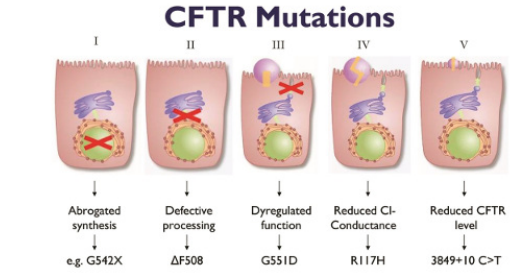
Functional Classification of CFTR Mutations

Class	Effect on CFTR	Examples of mutations
I	Defective protein production	G542X, R553X, W1282X, R1162X, 621-1G→T, 1717-1G→A
II	Defective protein processing	ΔF508, Δ1507, N1303K, S549N
III	Defective protein regulation	G551D, R560T
IV	Defective protein conductance	R117H, R334W, G85E, R347P
V	Reduced amounts of functioning CFTR protein	3849+10KbC→T, 2789+5G→A, A455E

It is important to know which allele(s) a CF patient carries because pharmaceutical companies are developing new drugs that target specific defects.
G542X: Defective Protein: Truncated
ΔF508: Defective Protein Processing: Folds incorrectly
G551D: Defective Protein Conductance: Unable to transport chloride
Mutation does not affect synthesis or localization [Kalydeco]

Cystic Fibrosis in the 21st Century by N. Simmonds ©2016 Sami Khuri

Five Classes of CF Mutations



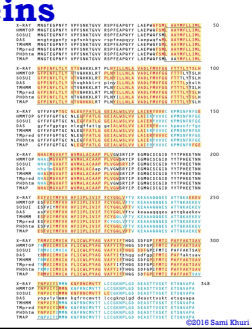
http://cysticfibrosis.org.uk/news/genotyping-pilot ©2016 Sami Khuri

Predicting Transmembrane Proteins

Predicting programs:

- HMMTOP
- SOSUI
- DAS
- TMHMM
- TMpred
- PHDhtm
- TMAP

are used to predict the structure of the bovine rhodopsin.



©2016 Sami Khuri

Predicting Programs (I)

- **HMMTOP** – Prediction of Transmembrane Helices and Topology of Proteins [www.enzim.hu/hmmtop]
- **SOSUI** – Classification and Secondary Structure Prediction of Membrane Proteins [bp.nuap.nagoya-u.ac.jp/sosui]
- **DAS** – Transmembrane Prediction Server [www.sbc.su.se/~miklos/DAS]
- **TMHMM** - Prediction of Transmembrane Helices in Proteins [www.cbs.dtu.dk/services/TMHMM]

©2016 Sami Khuri

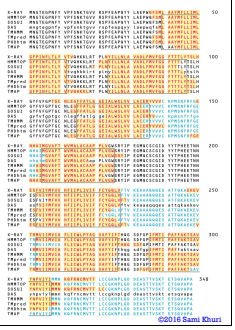
Predicting Programs (II)

- **Tmpred** - Prediction of Transmembrane Regions and Orientation [www.ch.embnet.org/software/TMPRED_form.html]
- **PHDhtm** – ProteinPredict [www.predictprotein.org]
- **TMAP** – Predict and plot transmembrane segments in protein sequences [emboss.bioinformatics.nl/cgi-bin/emboss/tmap]

©2016 Sami Khuri

Using X-Ray Crystallography

- The top row contains the results obtained from X-Ray crystallography.
- The transmembrane helices are highlighted in yellow.
- Extracellular loops are in black.
- Cytoplasmic loops are in blue.
- Boxed sequences are predicted to be transmembrane based on the consensus results of all prediction packages.



©2016 Sami Khuri

Part of the Alignment

```
X-RAY   GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE   150
HMMTOP  GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
SOSUI   GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
DAS     gyfvfgptgc nlegffatlg geIALWSLVV LAIERYVvvc kpmsnfrfge
TMHMM   GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
TMpred  GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
PHDhtm  GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
TMAP    GYFVFGPTGC NLEGFFATLG GEIALWSLVV LAIERYVVVC KPMSNFRFGE
```

- The third (out of 7) membrane (yellow residues) is shown in the box as predicted by the packages.
- Amino acids that are part of the extracellular loop are in black.
- Residues that are part of the cytoplasmic loop are in blue.

©2016 Sami Khuri

Transmembrane Predicting Methods

- There are many transmembrane predicting packages that are based on the following techniques:
  - Statistical Methods
    - Example: TMpred
  - Knowledge-based Methods
    - Example: SOSUI
  - Evolutionary-based Methods
    - Example: TMAP
  - Neural Networks
    - Example: PHDhtm

©2016 Sami Khuri

HMM for Transmembrane Protein Prediction (I)

- HMMs can incorporate:
  - Hydrophobicity
  - Charge bias
  - Helix length
  - Grammatical constraints

©2016 Sami Khuri

HMMs for Transmembrane Protein Prediction (II)

Simple Model:

- Define a set of **states**: each residue is then predicted to be in one of the states.
- Example:
  - A state for **inside loops**
  - A state for **outside loops**
  - A state for **transmembrane segments**

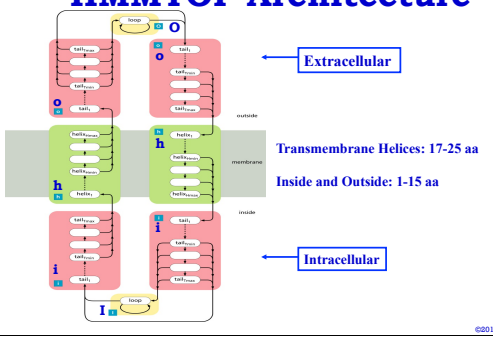
©2016 Sami Khuri

The Simple HMM

- Each state has an associated probability distribution over the 20 amino acids that describe the variability of each amino acid in the modeled region.
- States are connected to each other in a biological reasonable manner.
- The HMM is trained to have adequate emission and transition probabilities.

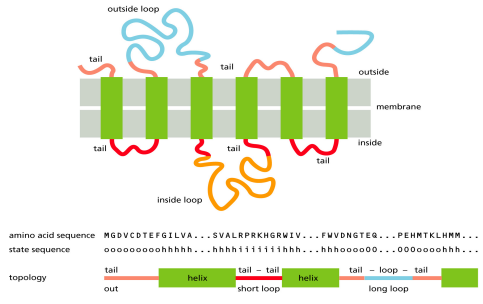
©2016 Sami Khuri

HMMTOP Architecture



©2016 Sami Khuri

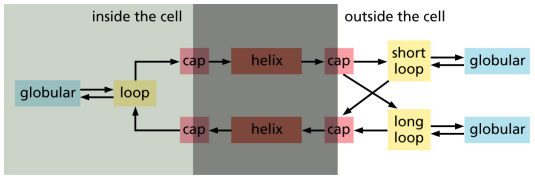
Five States of HMMTOP



©2016 Sami Khuri

TMHMM: Finite State Diagram

- TMHHM has seven states:



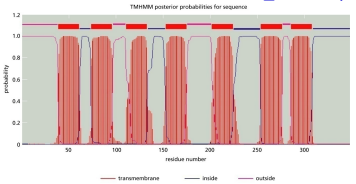
©2016 Sami Khuri

The TMHMM Package

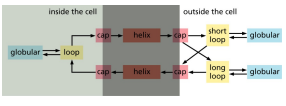
- TMHHM has seven states:
  - Core transmembrane helix
  - Helical cap
  - Helical tail
  - Loops on cytoplasmic end
  - Short loops outside the cell
  - Long loops inside the cell
  - Golbular-domain-like structures in the middle of each loop.

©2016 Sami Khuri

TMHMM: Output (I)



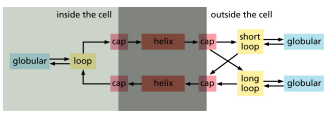
Sequence	outside	1	38
Sequence	TMhelix	39	61
Sequence	inside	62	73
Sequence	TMhelix	74	96
Sequence	outside	97	110
Sequence	TMhelix	111	133
Sequence	inside	134	152
Sequence	TMhelix	153	175
Sequence	outside	176	201
Sequence	TMhelix	202	224
Sequence	inside	225	253
Sequence	TMhelix	254	276
Sequence	outside	277	285
Sequence	TMhelix	286	308
Sequence	inside	309	348



©2016 Sami Khuri

TMHMM: Output (II)

Sequence	outside	1	38
Sequence	TMhelix	39	61
Sequence	inside	62	73
Sequence	TMhelix	74	96
Sequence	outside	97	110
Sequence	TMhelix	111	133
Sequence	inside	134	152
Sequence	TMhelix	153	175
Sequence	outside	176	201
Sequence	TMhelix	202	224
Sequence	inside	225	253
Sequence	TMhelix	254	276
Sequence	outside	277	285
Sequence	TMhelix	286	308
Sequence	inside	309	348



©2016 Sami Khuri

VKOR (I)

- Warfarin was approved for use as a medication in the early 1950s and has remained very popular.
- Warfarin is the most widely prescribed anticoagulant drug in North America.
- Warfarin decreases blood coagulation by inhibiting **vitamin K epoxide reductase** (**VKOR**).
- The gene encoding the catalytic subunit of VKOR was identified as an integral membrane protein.

©2016 Sami Khuri

VKOR (II)

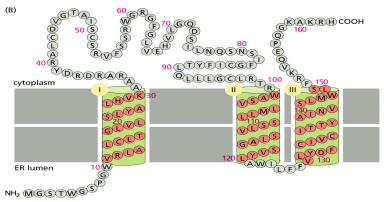
- These vitamin K-dependent proteins are important as coagulation factors, and are involved in bone metabolism and signal transduction.
- In order to understand structure-function relationship of these proteins, it is important to understand the membrane topology.
- Seven transmembrane prediction packages were used for that purpose.

©2016 Sami Khuri

(A)

programs	TM no.	C terminus	TM1	TM2	TM3	TM4
PHD	2	In		85–109 (1.0)		119–143 (0.87)
TMHMM 2.0	3	In	10–29 (1.0)		101–123 (0.90)	127–149 (0.93)
TopPred 2	3	In	9–29 (0.65)	78–98 (0.67)	109–129 (1.0)	
TMpred	3	In	9–29 (0.81)	75–97 (0.57)	101–129 (1.0)	
DAS	3		12–27	83–96	102–146	
SOSUI	3		11–31 (primary)	75–97 (secondary)		116–138 (primary)
MEMSAT	4	In	13–29 (0.76)	81–97 (0.60)	104–124 (1.0)	131–148 (0.68)

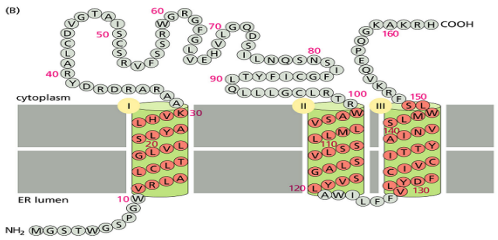
(B)



©2016 Sami Khuri

VKOR (III)

(B)



Experiments were performed and it was determined that VKOR has three transmembrane helices.

©2016 Sami Khuri

VKOR (IV)

programs	TM no.	C terminus	TM1	TM2	TM3	TM4
PHD	2	In		85–109 (1.0)		119–143 (0.87)
TMHMM 2.0	3	In	10–29 (1.0)		101–123 (0.90)	127–149 (0.93)
TopPred 2	3	In	9–29 (0.65)	78–98 (0.67)	109–129 (1.0)	
TMpred	3	In	9–29 (0.81)	75–97 (0.57)	101–129 (1.0)	
DAS	3		12–27	83–96	102–146	
SOSUI	3		11–31 (primary)	75–97 (secondary)		116–138 (primary)
MEMSAT	4	In	13–29 (0.76)	81–97 (0.60)	104–124 (1.0)	131–148 (0.68)

Two packages predicted the wrong number of helices.

Two packages predicted the wrong location of the C terminus

©2016 Sami Khuri

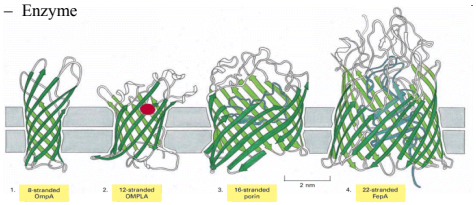
Other Types of Transmembranes

- Some transmembrane structures contain beta sheets instead of alpha helices.
- Tailored-made predictors were designed to detect them.
- Sometimes 2 or 3 alpha helices intertwine to form coiled-coil structures.
- Coiled-coil structures can be found in transmembrane as well as intracellular proteins.

©2016 Sami Khuri

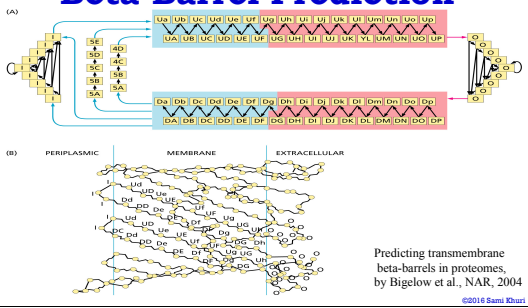
β -Barrels

- Located in mitochondria, chloroplasts, bacteria
- Functions include:
  - Transport channel
  - Receptor
  - Enzyme



©2012 Sami Khuri
©2016 Sami Khuri

PROFtmb for Beta-Barrel Prediction



Predicting transmembrane beta-barrels in proteomes, by Bigelow et al., NAR, 2004

©2016 Sami Khuri