SABBATICAL REPORT
Nancy Mozingo, Associate Professor of Biology
December 2009

PROJECT BACKGROUND

Septate junctions

Sea urchins are a widely used model organism for the study of embryology and gene regulation. One of the hallmarks of development in these animals is the blastula stage of development. The sea urchin blastula is characterized by a fluid-filled cavity called the blastocoel which is surrounded by an epithelial cell layer . Critical to the formation of the blastula is the development of intercellular junctions on adjacent epithelial cells. These junctions span the space between cells and are essential in maintaining functional and structural properties of the epithelium. One type of cell junction that forms early in development (perhaps as early as the 4-cell stage) is the septate junction. Septate junctions have a distinct, ladder-like appearance in which adjacent cells are connected by septa. Septate junctions function in the maintenance of cell polarity, cell-cell adhesion and establishment of a paracellular permeability barrier.

Septate junctions are only found in invertebrate animals, however, vertebrate animals possess tight junctions which appear to be the functional equivalent of invertebrate septate junctions . The similarity between invertebrate septate junctions and vertebrate tight junctions extends to the molecular level as well suggesting that proteins found in vertebrate tight junctions and invertebrate septate junctions share a common evolutionary history. To date, dozens of proteins associated with septate and tight junctions have been identified, but the sea urchin equivalents are not known.

The sea urchin genome has been sequenced and an analysis of the sea urchin genome was recently published. My project utilized the recently available sea urchin genome to identify and analyze a septate junction gene and perform preliminary phylogenetic comparisons of septate/tight-junction associated genes. To complete this project, I obtained training on use of comparative genomic tools using on-line tutorials.

PURPOSE AND GOALS

The purpose of my sabbatical leave was for the advancement of research and scholarship, instructional improvement and for faculty retraining. I accomplished the following goals:

Obtained training on bioinformatics by completing on-line tutorials.
Found a putative claudin protein in the sea urchin genome
Performed in-silico protein analysis of SP-claudin
Identified key regions for antibody production and contracted with Gallus Immunotech for production of an antibody
Produced preliminary multiple sequence alignment

RESULTS

Identification of putative sea urchin claudin (Sp Claudin)

I focused on trying to identify a sea urchin homologue to "claudin". Claudin is a protein that performs the paracellular barrier function in vertebrate tight junctions and an invertebrate homologue has recently been identified in *Drosophila* called sinuous. Initial BLAST searches did not reveal any statistically significant sequences so I turned to another search tool called Pattern Hit Initiated BLAST (PHI BLAST). PHI-BLAST uses a protein query sequence and a pattern contained in that sequence. (http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/Post_Blast.html) . A hypothetical septate junction protein was pulled out of the sea urchin genome using the Drosophila sinuous sequence and a pattern that I designed based on characteristics of sinuous and vertebrate claudins.

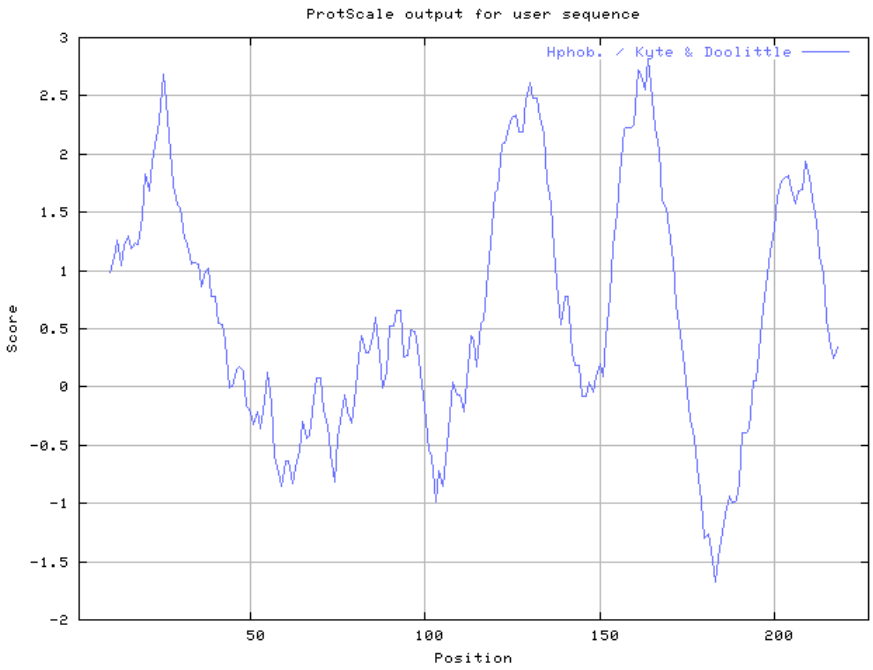
Amino Acid Sequence of putative Sp Claudin:
MALGCAAGSHISRSQLIVCIIGFVGFTLLALGAVSDYWVTYGITSAVSAGSSNGSS
PPPQAALLHREGLWRSCLQYIANNSTSVTGHMCFFGLSAPDSLSMSQGLNSQTR
YEVSFLIATWVLYGLGVVLSLIAVVMAIAALRHKKNQTLLRGVSAVFILAALLAF
LGLVIYAVRTSKFPNQWPNGDSPYSSSSLAWAYGISWVGLLLCFI
AGVGHLWVMRRYEDSMI

Protein sequence analysis

The Sp Claudin sequence was analyzed for structural and functional features. Each analysis that was run is listed below with a brief description of the tool from the website and a summary of the results.

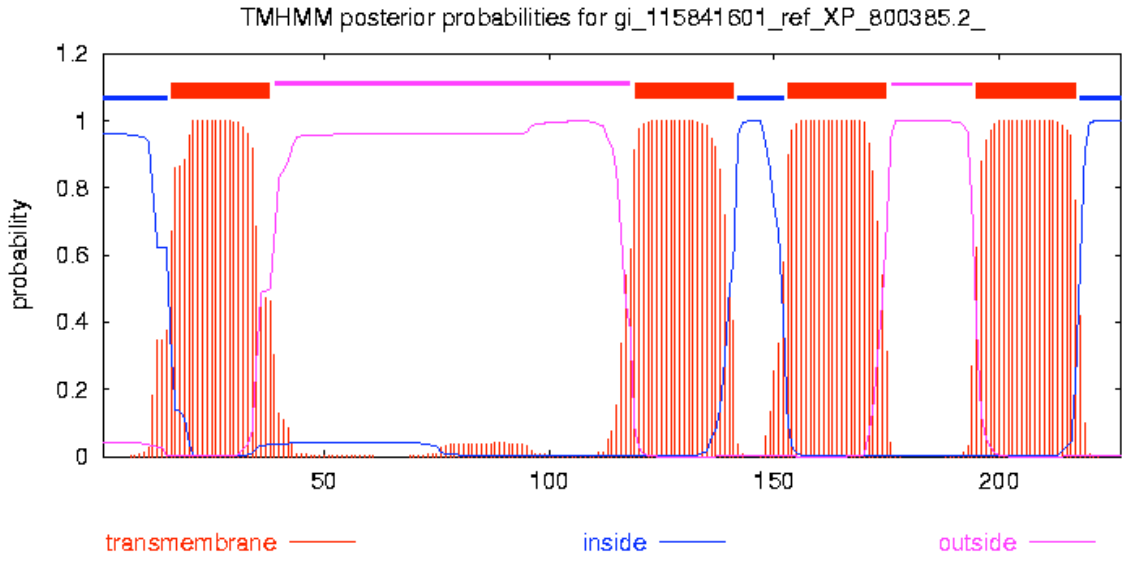
ProParam
ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein including the molecular weight, theoretical pI, amino acid composition, atomic composition, etc. http://www.expasy.ch/tools/protparam.html
This tool indicates that Sp claudin has 227 amino acids and a predicted molecular weight of 24.3K. It is classified as a stable protein (results shown in appendix).

ProtScale
ProtScale allows you to compute and represent the profile produced by any amino acid scale on a selected protein. The Kyte and Doolittle (J. Mol. Biol. 157:105-132, 1982) hydrophobicity scale was used (http://www.expasy.ch/tools/protscale.html)
The output from this analysis suggests that SP claudin has 4 hydrophobic domains (see peaks in figure below) suggestive of a membrane protein.



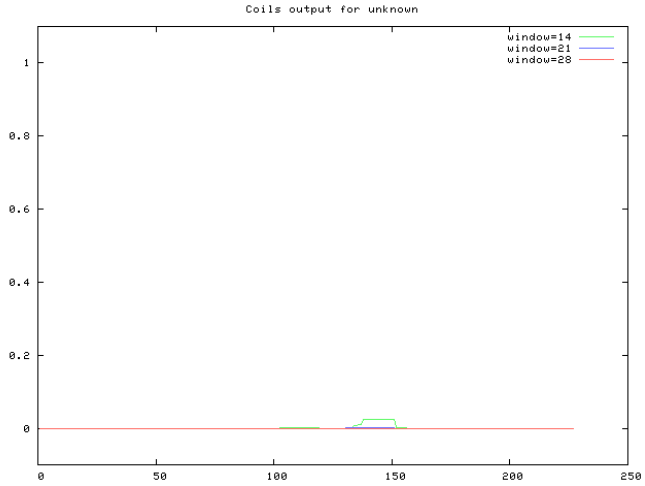
TMHMM
This server is for prediction of transmembrane helices in proteins.
http://www.cbs.dtu.dk/services/TMHMM/

The output of this tool shows that SP claudin has 4 transmembrane domains (shown in red below). This is significant because it is a hallmark of vertebrate claudins and *Drosophila* sinuous.



COILS
COILS is a program that compares a sequence to a database of known parallel two-stranded coiled-coils and derives a similarity score. By comparing this score to the distribution of scores in globular and coiled-coil proteins, the program then calculates the probability that the sequence will adopt a coiled-coil conformation.
http://www.ch.embnet.org/software/COILS_form.html

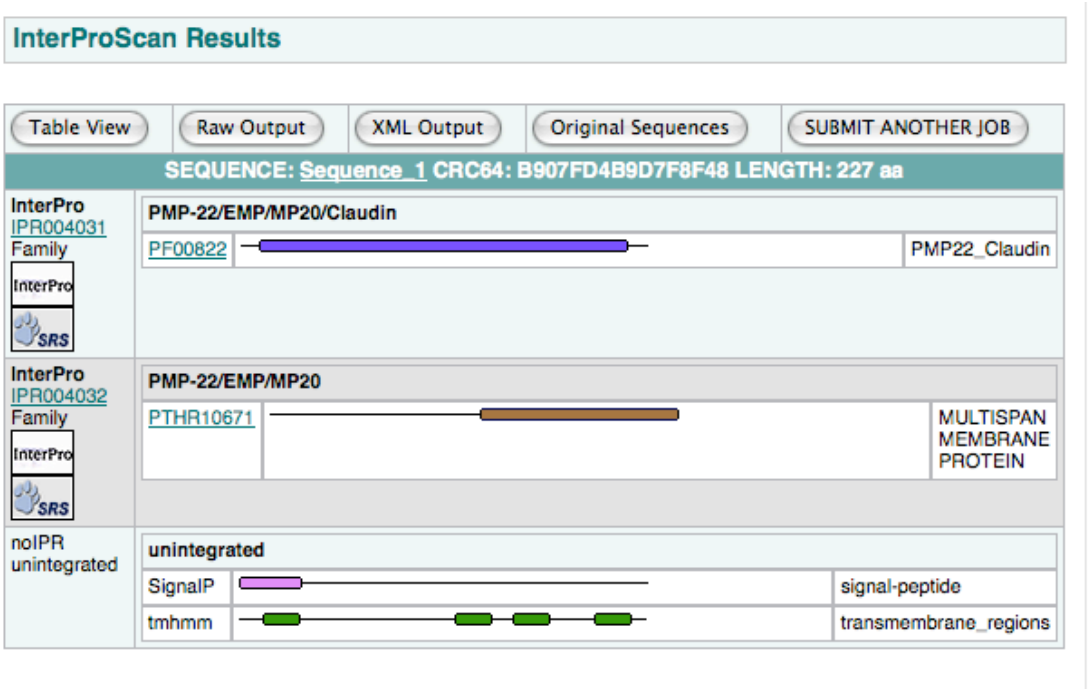
The output from this analysis shows that SP claudin does not possess coiled coil domains.



Scanprosite
The ScanProsite tool allows for scanning of protein sequence(s) for the occurrence of motif(s).
http://www.expasy.ch/tools/scanprosite/

The output from this tool (shown in appendix) reveals 4 potential N-glycosylation sites. One of the four sites is a predicted intracellular loop so is probably not valid. Seven possible N-myristoylation site are suggested, but myristolated proteins have been shown to be membrane associated or intracellular so this may be an artifact. This scan also revealed 1 possible Protein kinase C phosphorylation site and 2 possible casein kinase II phosphorylation sites but all of these site are wholly or partially extracellular and so are not valid.

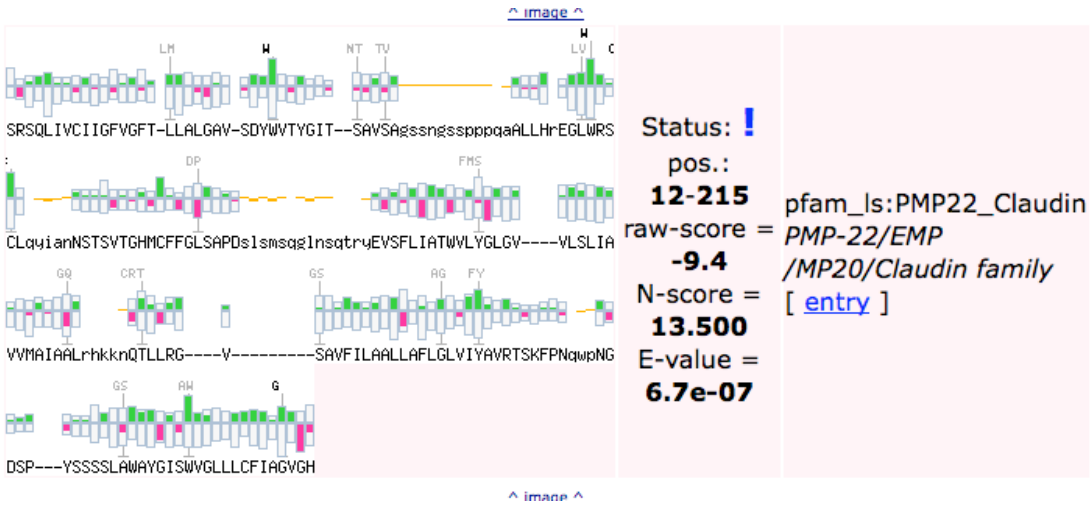
Interproscan
InterPro classifies sequences at superfamily, family and subfamily levels, predicting the occurrence of functional domains, repeats and important sites.
http://www.ebi.ac.uk/interpro/
The results of this analysis demonstrate that SP claudin is a member of the Claudin Family.



motif scan

This tool searches for all known motifs in a sequence.

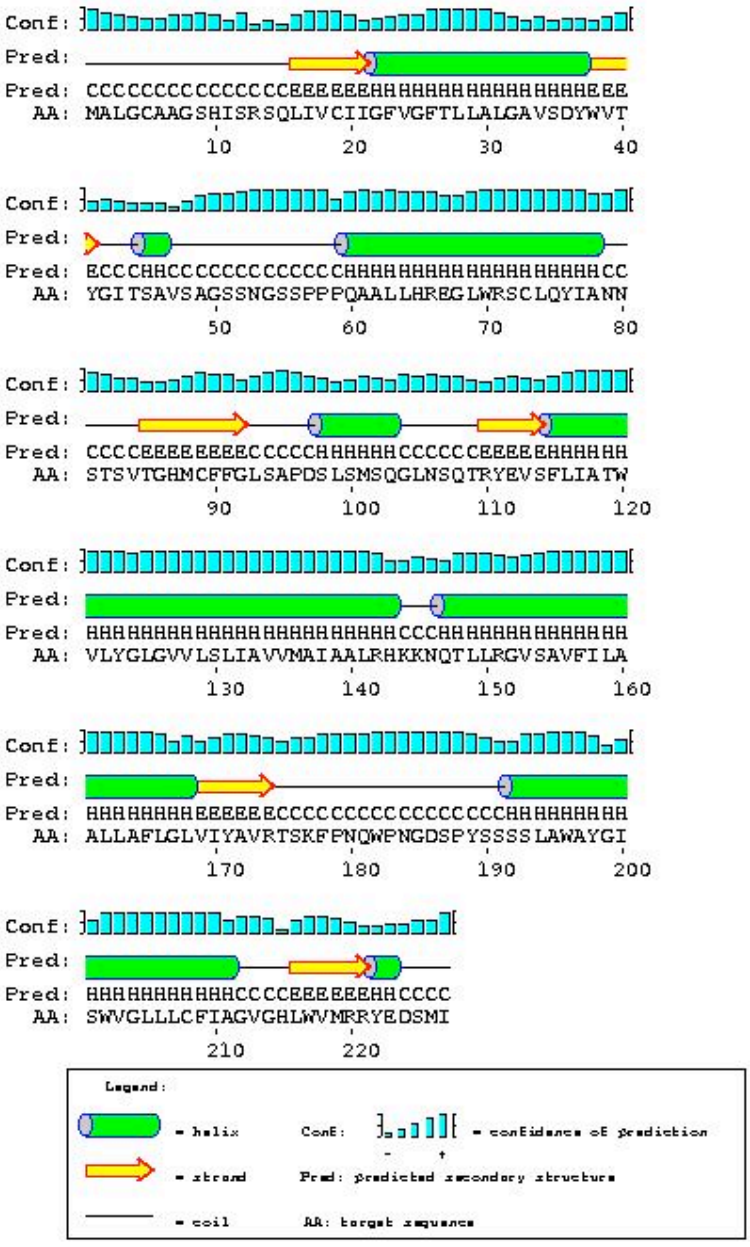
This analysis found a strong match with the PMP-22/EMP/MP20/Claudin family



PSIPRED

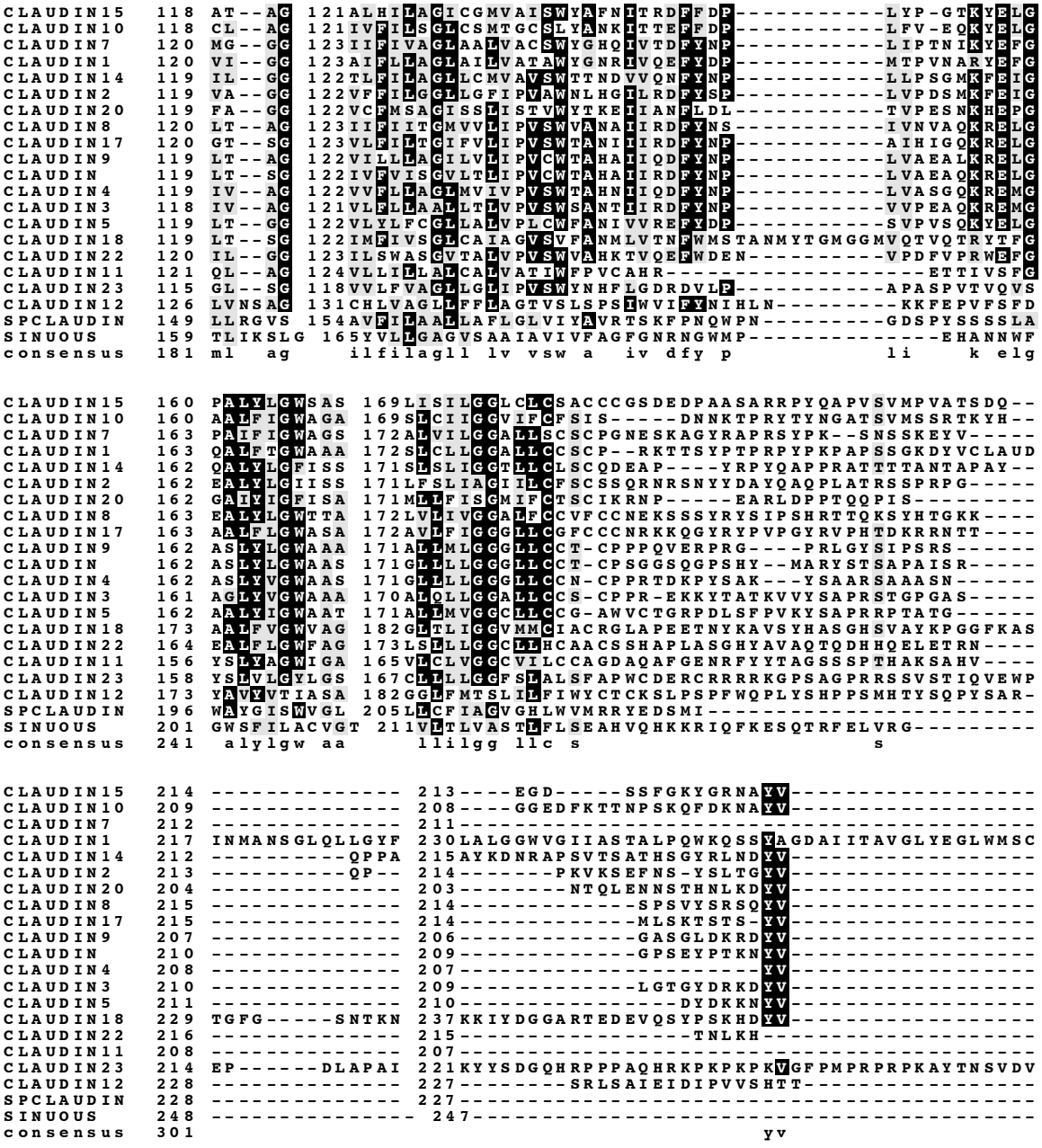
This tool predicts secondary structure from a primary sequence.

This analysis predicted 4 alpha helices corresponding to the transmembrane domains. It also predicts a helix in region of ~60-78 which is part of extracellular loop 1.



Antibody design
I was interested in obtaining an antibody that could be used for immunodetection of Sp claudin in sea urchin embryos and, in addition, I wanted an antibody that may be used to block function. Thus, I targeted the extracellular loops for antibody production. Based on these parameters, I contacted Gallus for sequence analysis for the most immunogenic regions. The sequence CPDSLSMSQGLNSQTRYE 96-112 was chosen which is in the first extracellular loop. This antibody has been produced and future work will utilize this antibody to study the function of Sp claudin in sea urchin embryos.

Multiple sequence alignment
To gauge the similarity between Sp claudin and members of the human claudin family a multiple sequence alignment was performed (shown on next page).



The analyses demonstrate that Sp claudin is a is a 227 amino acid protein with 4 transmembrane alpha helices. The protein has a theoretical molecular weight of 24.3K with 3 likely N-glycosylation sites. The signature motifs and the multiple sequence alignment provide evidence that Sp claudin is a member of the claudin superfamily.

APPENDICES

1) ProParam results

ProtParam

User-provided sequence:

```
        10         20         30         40         50         60
MALGCAAGSH ISRSQLIVCI IGFVGFTLLA LGAVSDYWVT YGITSAVSAG SSNGSSPFPQ

        70         80         90        100        110        120
AALLHREGLW RSCLQYIANN STSVTGHMCF PGLSAPDSLS MSQGLNSQTR YEVSFLIATW

       130        140        150        160        170        180
VLYGLGVVLS LIAVVMAIAA LSHKKNQTLL RGVSAVFILA ALLAFLGLVI YAVRTSKFPN

       190        200        210        220
QWPNGDSPYS SSSLAWAYGI SWVGLLLCFI AGVGHLWVMR RYEDSMI
```

References and documentation are available.

*** Please note the modified algorithm for extinction coefficient.

Number of amino acids: 227

Molecular weight: 24318.3

Theoretical pI: 9.08

Amino acid composition: CSV format

Ala (A)	24	10.6%
Arg (R)	9	4.0%
Asn (N)	7	3.1%
Asp (D)	4	1.8%
Cys (C)	5	2.2%
Gln (Q)	7	3.1%
Glu (E)	3	1.3%
Gly (G)	21	9.3%
His (H)	5	2.2%
Ile (I)	14	6.2%
Leu (L)	30	13.2%
Lys (K)	3	1.3%
Met (M)	6	2.6%
Phe (F)	9	4.0%
Pro (P)	7	3.1%
Ser (S)	29	12.8%
Thr (T)	9	4.0%
Trp (W)	7	3.1%
Tyr (Y)	9	4.0%

```
Val (V)  19       8.4%
Pyl (O)   0       0.0%
Sec (U)   0       0.0%

 (B)    0         0.0%
 (Z)    0         0.0%
 (X)    0         0.0%
```

Total number of negatively charged residues (Asp + Glu): 7
Total number of positively charged residues (Arg + Lys): 12

Atomic composition:

```
Carbon      C         1111
Hydrogen    H         1726
Nitrogen    N          288
Oxygen      O          303
Sulfur      S           11
```

Formula: $C_{1111}H_{1726}N_{288}O_{303}S_{11}$
Total number of atoms: 3439

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

```
Ext. coefficient    52160
Abs 0.1% (=1 g/l)   2.145, assuming ALL Cys residues appear as half cystines


Ext. coefficient    51910
Abs 0.1% (=1 g/l)   2.135, assuming NO Cys residues appear as half cystines
```

Estimated half-life:

The N-terminal of the sequence considered is M (Met).

```
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
                            >20 hours (yeast, in vivo).
                            >10 hours (Escherichia coli, in vivo).
```

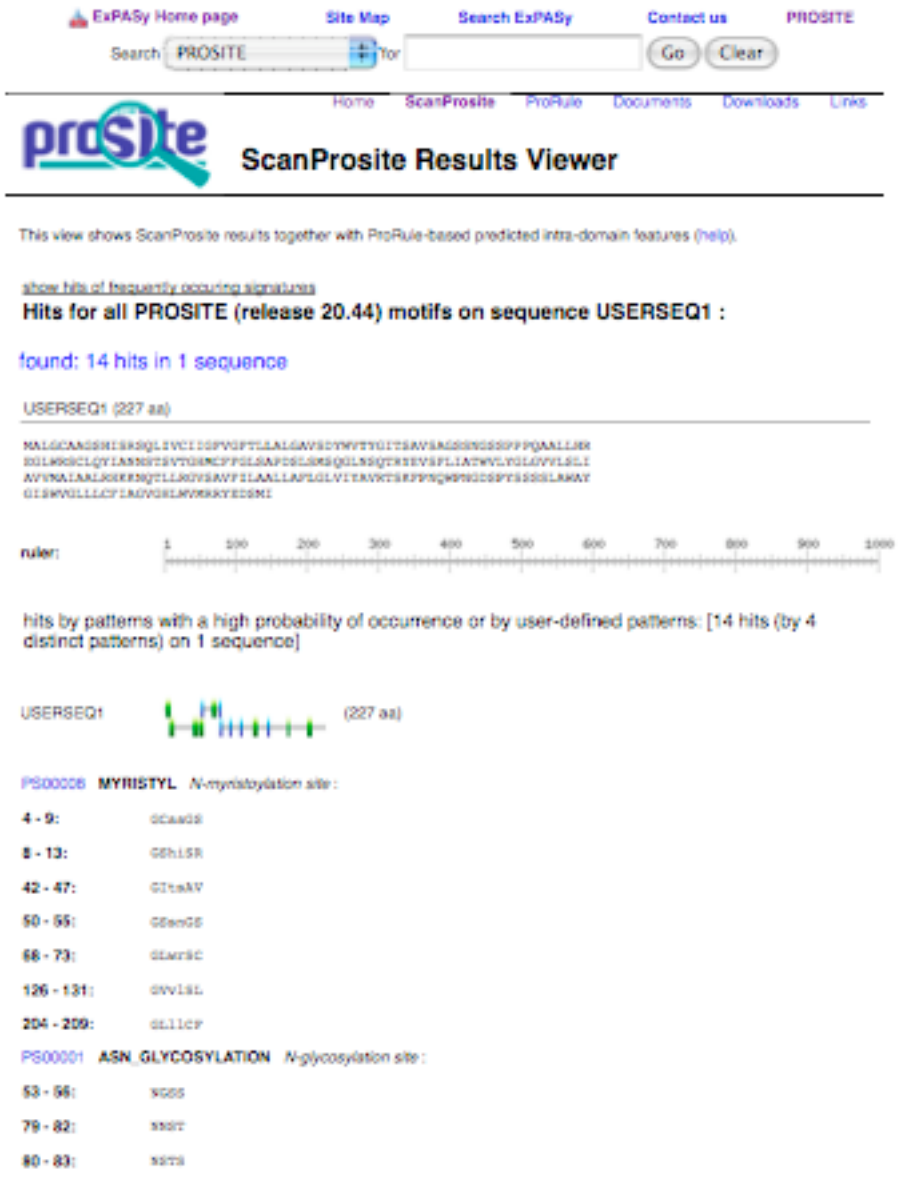
Instability index:

```
The instability index (II) is computed to be 38.91
This classifies the protein as stable.
```

Aliphatic index: 110.44

Grand average of hydropathicity (GRAVY): 0.618

2) ScanProsite results



146 - 149: sgtL

PS00006 **CK2_PHOSPHO_SITE** *Casein kinase II phosphorylation site :*

94 - 97: SapD

109 - 112: TtrY

PS00005 **PKC_PHOSPHO_SITE** *Protein kinase C phosphorylation site :*

175 - 177: TsK

Legend:

horizontal scaling: 0.6

do not show text labels:

do not show sites in hits:

do not show ranges in hits:

redisplay