

**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](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  
PPPQAALLHREGLWRSCQLQYIANNSTSVTGHMCFFGLSAPDSLMSQGLNSQTR  
YEVSFLIATWVLYGLGVVLSLIAVVMIAAALRHKNQTLLRGVSAVFILAALLAF  
LGLVIYAVRTSKFPNQWPNGDSPYSSSSLAWAYGISWVGLLLCFI  
AGVGHLWVMRRYEDSMI

### Protein sequence analysis

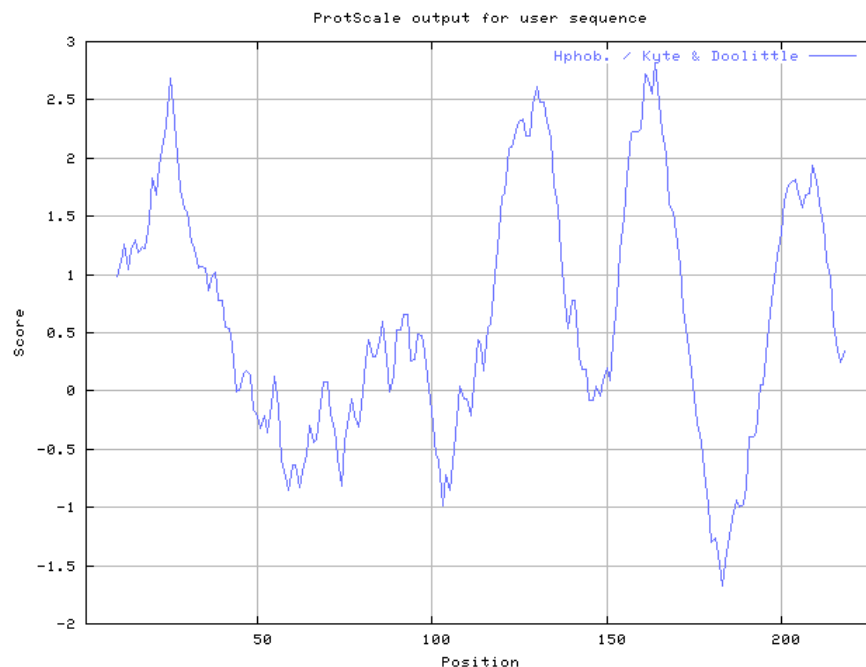
The Sp Claudin sequence was analyzed for structural and functional features. Each analysis that was run is listed below with a brief description of the tool from the website and a summary of the results.

#### *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.

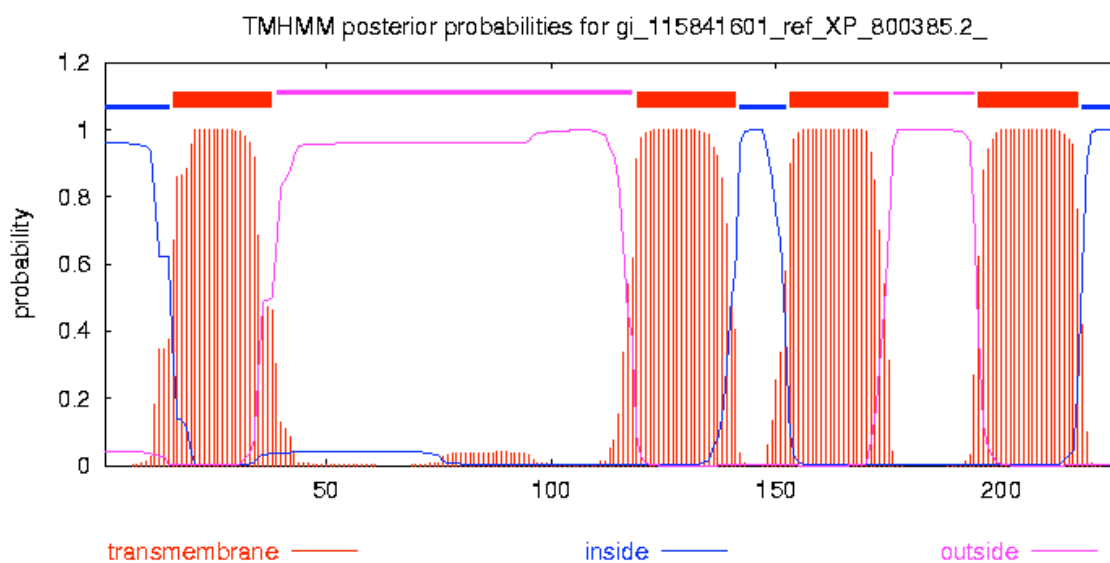


### TMHMM

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*.

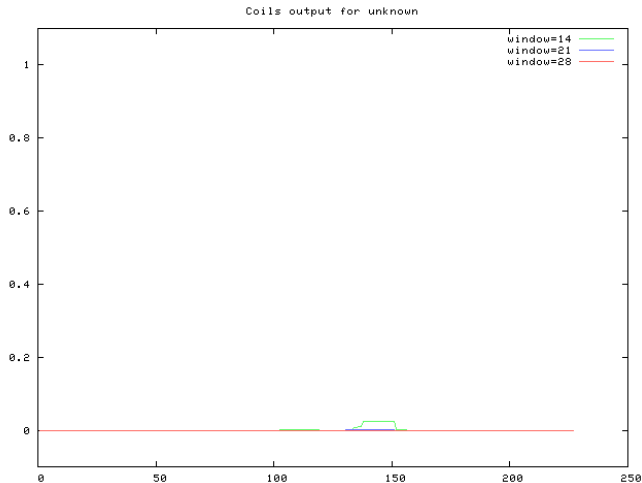


### *COILS*

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](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 sites are suggested, but myristoylated 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 sites 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.

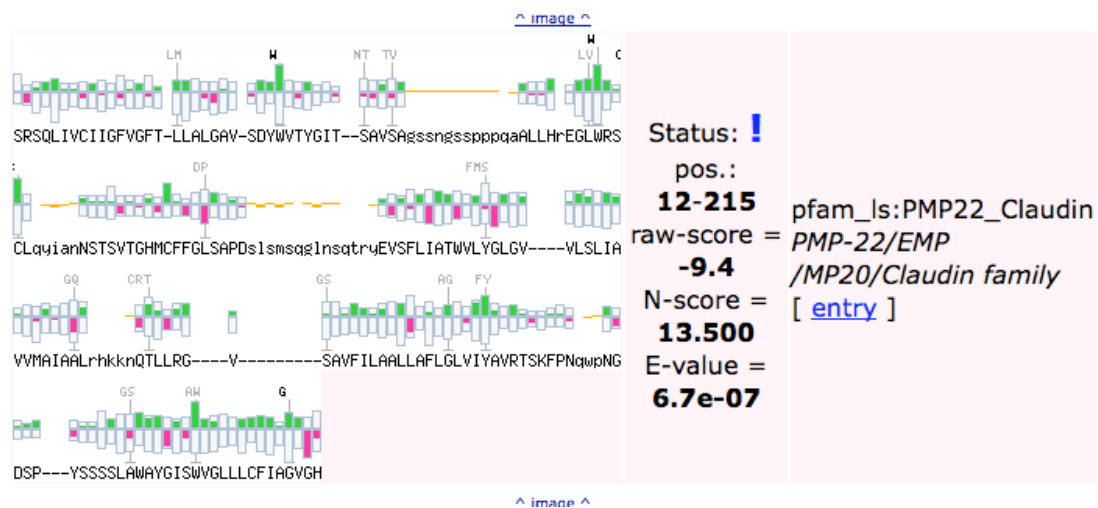
## InterProScan Results

Table View	Raw Output	XML Output	Original Sequences	SUBMIT ANOTHER JOB
SEQUENCE: Sequence_1 CRC64: B907FD4B9D7F8F48 LENGTH: 227 aa				
InterPro IPR004031 Family	PMP-22/EMP/MP20/Claudin			
InterPro SRS	PF00822  PMP22_Claudin			
InterPro IPR004032 Family	PMP-22/EMP/MP20			
InterPro SRS	PTHR10671  MULTISPAN MEMBRANE PROTEIN			
noIPR unintegrated	unintegrated			
SignalP	 signal-peptide			
tmhmm	 transmembrane_regions			

### motif scan

This tool searches for all known motifs in a sequence.

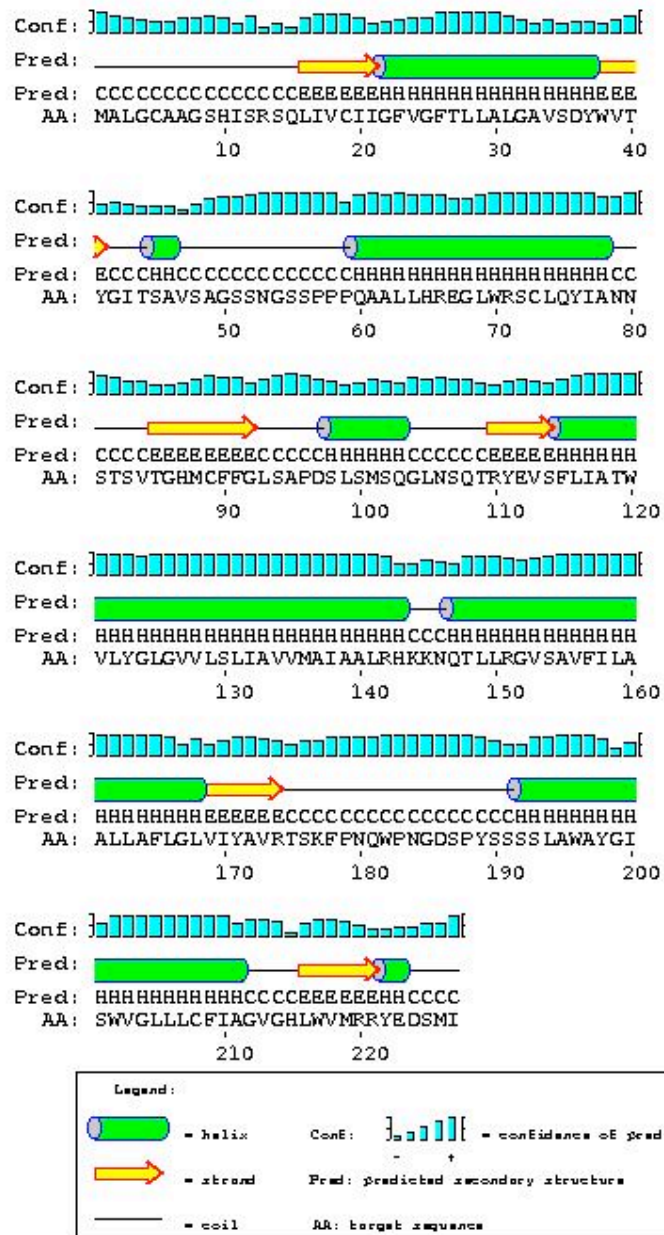
This analysis found a strong match with the PMP-22/EMP/MP20/Claudin family



## 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 CPDSLMSQGLNSQTRYE 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).

```

CLAUDIN15 118 AT--AG 121ALHI LAGICGMVAISWYAFNITTRDFFDP-----LYP-GTKYELG
CLAUDIN10 118 CL--AG 121IVFILSGLCSTMTGCSLYANKITTEFFDP-----LFV-EQKYELG
CLAUDIN7 120 MG--GG 123IIFIVAGLAAIVACSWYGHQIVTDFYNP-----LIFTNIIKYEFG
CLAUDIN1 120 VI--GG 123AIFLAGLAIIVATAWYGNRIVQEFYDP-----MTPVNNARYEFG
CLAUDIN14 119 IL--GG 122TLFILAGLTCMVAVSWTTNDVVQNFYNP-----LLPSGMKFEIG
CLAUDIN2 119 VA--GG 122VFFILGGLGFIPVAVNLHGLRDFYSP-----LVPDSMKFEIG
CLAUDIN20 119 FA--GG 122VCFMSAGISSTIVTWYTKETIANFLDL-----TPVESNKHHEPG
CLAUDIN8 120 LT--AG 123IIFIITGMVVLIPVSWVANAIIRDFYNS-----IVNVAQKRELG
CLAUDIN17 120 GT--SG 123VLFILGTGIFVLIIPVSWTANIIIRDFYNP-----AIHIGQKRELG
CLAUDIN9 119 LT--AG 122VILLLAGILVLIIPVGTAAHAIIQDFYNP-----LVAEALKRELG
CLAUDIN 119 LT--SG 122IVFVISGLTIIIPVGTAAHAIIRDFYNP-----LVAEALKRELG
CLAUDIN4 119 IV--AG 122VFFLAGLGMVLIIPVSWTANIIQDFYNP-----LVASGQKREMGE
CLAUDIN3 118 IV--AG 121VLFLLAALLTIPVSWSANTIIIRDFYNP-----VVEPAQKREMGE
CLAUDIN5 119 LT--GG 122VLYLFCGLLALVPLCWFANIVVREFYDP-----SVPVSQKYELG
CLAUDIN18 119 LT--SG 122IMFIVSGCATAGVSVFANMLVTNFWMTANMYTGMGMVQTVQTRYTFG
CLAUDIN22 120 IL--GG 123ILSWASGVTAVPVSVVAHKTQVEFWDEN-----VPDFVPRWEFG
CLAUDIN11 121 QL--AG 124VLLILLLALCALVATIMFPVCAHR-----ETTIVSFG
CLAUDIN23 115 GL--SG 118VVLFVAGLGLFIIVSMYNHFLGDRDVLFP-----APASPVTVQVS
CLAUDIN12 126 LVNSAG 131CHLVAGLFLFAGTSLSPSFWVIIFYNIHLN-----KKFEPVFSFD
SPCLAUDIN 149 LLRGVS 154AVFILAAFLAFLGLVIYAVRTSKFPNQWPN-----GDSPYSSSSSLA
SINUOUS 159 TLIKSLG 165YVLLGAGVSAIAIVIVFAGFGNRRNGWMP-----EHANNWF
consensus 181 ml ag ilfilagll lv vsw a iv dfy p li k elg

CLAUDIN15 160 PALYLGWSAS 169LISILGGLCLCSACCCGSDDEPAASARRPYQAPVSVMPVATSDQ--
CLAUDIN10 160 AALFIGWAGA 169SLCIIGGVIFGFSIS----DNNKTPRYTYNGATSVMSSTRKYH--
CLAUDIN7 163 PAIFIGWAGS 172ALVILGGALLSCSPGNEKAGYRAPRSYPK--SNSSKEYV----
CLAUDIN1 163 QALFTGWAAA 172SLCLIGGALLCSCCP--RKTTSYPTPRYPKPAPSSGKDYVCLAUD
CLAUDIN14 162 QALYLGFISS 171SLIGGTLCLCLSCQDEAP---YRPYQAPPRAATTTTANTAPAY--
CLAUDIN2 162 EALYLGFISS 171LFSLIAGIILCFSCSSQNRNSNYDAYQAQPLATRSPRPG----
CLAUDIN20 162 GAILYIGFISA 171MLFISGMIFCTSCIKRNP----EARLDPTTQQPIS-----
CLAUDIN8 163 EALYLGWTTA 172LVLIIGGALLFCVFCNEKSSSYRYSIPSHRTQKSYHTGKK----
CLAUDIN17 163 AALFLGWASA 172AVLFIGGGLLCGFCCNRKKQGYRYPVPGXRVPHDKRRNTT----
CLAUDIN9 162 ASLYLGWAAA 171ALLMIGGGLLCT-CPPPPQVERPRG---PRLGYSIPSR-----
CLAUDIN 162 ASLYLGWAA 171GLLLIGGGLLCT-CPSGSQGPSHY--MARYSTAPAI-----
CLAUDIN4 162 ASLYVGWAA 171GLLLIGGGLLCCN-CPPRTDKPYSAK---YSAARSAASN-----
CLAUDIN3 161 AGLYVGWAAA 170ALQLMGALLCCS-CPPR-EKKYTATKVVSAPRSTGPGAS-----
CLAUDIN5 162 AALYIGWAAT 171ALLMVGCCLLCG-AWVCTGRPDLSFPVKYSAPRPRATG-----
CLAUDIN18 173 AALFVGWVAG 182GLTLIGGVMMCIACRGLAPEETNYKAVSYHASGHSAVKPGGFKAS
CLAUDIN22 164 EALFLGWVAG 173LSLLIGGGLLHCAACSSHAPLASGHYAVAQTDHHELETRN----
CLAUDIN11 156 YSLYAGWIGA 165VLCVGGCVILCCAGDAQAFGENRFYTAGSSSPHTAKSAHV----
CLAUDIN23 158 YSLVLGYLGS 167CLLLGGFSLALSFAPWCDCRRRRKGPSAGPRSSSVSTIQVEWP
CLAUDIN12 173 YAVYVTIASA 182GGFLFMTSLILFIWYCTCKSLSPFWQPLYSHPSPMHTYSQPY SAR-
SPCLAUDIN 196 WYAGISWVGL 205LCLFIAVAGHLLWVMRRYEDSMI-----
SINUOUS 201 GWSFILACVGT 211VTLTVASTLFLSEAHVQHKKRIQFKESQTRFELVRG-----
consensus 241 alylgw aa lllilgg llc s s

CLAUDIN15 214 ----- 213-----EGD-----SSFGKYGRNA YV-----
CLAUDIN10 209 ----- 208-----GGEDFKTTNPSKQFDKNA YV-----
CLAUDIN7 212 ----- 211-----
CLAUDIN1 217 INMANSGLQLLG YF 230LALGQWVGIIASTALPQWKQSS YAGDAIITAVGLYEGLWMSC
CLAUDIN14 212 -----QPPA 215AYKDNRAPSVTSATHSGYRLND YV-----
CLAUDIN2 213 -----QP-- 214-----PKVKSEFNS-YSLTG YV-----
CLAUDIN20 204 ----- 203-----NTQLENNSTHNLKD YV-----
CLAUDIN8 215 ----- 214-----SPSVYSRSQ YV-----
CLAUDIN17 215 ----- 214-----MLSKTSTS-YV-----
CLAUDIN9 207 ----- 206-----GASGLDKRD YV-----
CLAUDIN 210 ----- 209-----GPSEYPTKN YV-----
CLAUDIN4 208 ----- 207----- YV-----
CLAUDIN3 210 ----- 209-----LGTGYDRKD YV-----
CLAUDIN5 211 ----- 210-----DYDKKN YV-----
CLAUDIN18 229 TGFG-----SNTKN 237KKIYDGGARTEDEVQSYPSKHD YV-----
CLAUDIN22 216 ----- 215-----TNLKH-----
CLAUDIN11 208 ----- 207-----
CLAUDIN23 214 EP-----DLAPAI 221KYYSQGHRRPPAQHRKPKPKK YGFPMPRPRPKAYTNSVDV
CLAUDIN12 228 ----- 227-----SRLSAIEIDI PVVSHTT-----
SPCLAUDIN 228 ----- 227-----
SINUOUS 248 ----- 247-----
consensus 301 yv

```

The analyses demonstrate that Sp claudin 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 ISRSGLIVCI IGFVGFLLA LGAVSDYNYT YGITSAVSAG SNGSSPPFQ
      70      80      90     100     110     120
AALLHREGWL RSCLQYIANN STSVTGHMCF FGLSAPDSLS MSQGLNSQTR YEVSFLIATW
     130     140     150     160     170     180
VLYGLGVVLS LIAVVMATAA LSEKKNQTL RGVSAVFILA ALLAPLGLVI YAVRTSKPPN
     190     200     210     220
QWPNGDSPYS SSSLAWAYGI SWVGLLLCFI AGVGHVWVR 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 (D)	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<sub>1111</sub>H<sub>1726</sub>N<sub>288</sub>O<sub>303</sub>S<sub>11</sub>

Total number of atoms: 3439

#### Extinction coefficients:

Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, 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

[ExPASy Home page](#)
[Site Map](#)
[Search ExPASy](#)
[Contact us](#)
[PROSITE](#)

Search

[Home](#)
[ScanProsite](#)
[ProRule](#)
[Documents](#)
[Downloads](#)
[Links](#)

# proSite

## ScanProsite Results Viewer

This view shows ScanProsite results together with ProRule-based predicted intra-domain features ([help](#)).

[show hits of frequently occurring signatures](#)

**Hits for all PROSITE (release 20.44) motifs on sequence USERSEQ1 :**

found: 14 hits in 1 sequence

USERSEQ1 (227 aa)

```

NALGCAANDSHIRSGQLIVCIISPPVQFTLLALGAVSDYVFTTQITSAVLSAGSSSSGPPQAALLER
EQLKRGCLQYIANNGTSVTDKNCFFSLSAFSLKMSGLNSGTREVSFLIATWLYQLAAPPFLSLT
AVPVALAALSGKRNQTLKQVSAVFLAALLAFLGLVIEANKTSKPPVQWQWSSPTSSGSLAKAY
GISWVULLLCPILADVGLAVKRYEDSMI
  
```

ruler:

hits by patterns with a high probability of occurrence or by user-defined patterns: [14 hits (by 4 distinct patterns) on 1 sequence]

USERSEQ1 (227 aa)

**PS00006 MYRISTYL** *N*-myristoylation site :

4 - 9:       GCAGDS  
 8 - 13:      GSHISR  
 42 - 47:     GItaAV  
 50 - 55:     GSeoSE  
 66 - 73:     GLmrSC  
 126 - 131:  GNVlSL  
 204 - 209:  GLlICP

**PS00001 ASN GLYCOSYLATION** *N*-glycosylation site :

53 - 56:      NGSS  
 79 - 82:      NSDT  
 80 - 83:      NSTS

146 - 149:      NQTL

PS00006    CK2\_PHOSPHO\_SITE    Casein kinase II phosphorylation site :

94 - 97:            IapB

109 - 112:        TtyS

PS00006    PKC\_PHOSPHO\_SITE    Protein kinase C phosphorylation site :

175 - 177:        Ttk

Legend:


 disulfide bridge     active site     other "ranges"     other sites

horizontal scaling:

do not show text labels: ☐

do not show sites in hits: ☐

do not show ranges in hits: ☐


 [ExPASy Home page](#)

[Site Map](#)

[Search ExPASy](#)

[Contact us](#)

[PROSITE](#)

Hosted by  SIB Switzerland    Minor sites: [Australia](#) [Brazil](#) [Canada](#) [China](#) [Korea](#)