

## Appendix 5

### 5.1 ProtParam (Gasteiger *et al.* 2005)

#### 5.1.1 NrtA Analysis

Number of amino acids: 507  
Molecular weight: 54925.1  
Theoretical pI: 8.82

##### Atomic composition:

Carbon (C): 2540  
Hydrogen (H): 3891  
Nitrogen (N): 635  
Oxygen (O): 675  
Sulphur (S): 25  
**Formula:** C<sub>2540</sub>H<sub>3891</sub>N<sub>635</sub>O<sub>675</sub>S<sub>25</sub>  
**Total number of atoms:** 7766

##### Extinction coefficients:

**Ext. coefficient** 74870  
Abs 0.1% (=1 g/l) 1.363, assuming all pairs of  
Cysteine residues form cystine  
**Ext. coefficient** 74370  
Abs 0.1% (=1 g/l) 1.354, assuming all Cysteine  
residues are reduced  
(Unit of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water)

##### 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 (*E.coli*, in vivo)

##### Instability index:

The instability index (II) is computed to be 35.57  
This classifies the protein as stable.

**Aliphatic index:** 96.63

**Grand average of hydropathicity (GRAVY):** 0.457

**Total number of negatively charged residues (N + E):** 30

**Total number of positively charged residues (R + K):** 36

##### Amino acid composition:

Ala (A)	52	10.3%
Arg (R)	19	3.7%
Asn (N)	19	3.7%
Asp (D)	19	3.7%
Cys (C)	8	1.6%
Gln (Q)	13	2.6%
Glu (E)	11	2.2%
Gly (G)	48	9.5%
His (H)	6	1.2%
Ile (I)	39	7.7%
Leu (L)	48	9.5%
Lys (K)	17	3.4%
Met (M)	17	3.4%
Phe (F)	41	8.1%
Pro (P)	31	6.1%
Ser (S)	36	7.1%
Thr (T)	26	5.1%
Trp (W)	10	2.0%
Tyr (Y)	13	2.6%
Val (V)	34	6.7%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

#### 5.1.2 NrtB Analysis

Number of amino acids: 497  
Molecular weight: 53731.9  
Theoretical pI: 9.12

##### Atomic composition:

Carbon (C): 2488  
Hydrogen (H): 3831  
Nitrogen (N): 633  
Oxygen (O): 653

##### Amino acid composition:

Ala (A)	50	10.1%
Arg (R)	20	4.0%
Asn (N)	20	4.0%
Asp (D)	18	3.6%
Cys (C)	7	1.4%
Gln (Q)	12	2.4%
Glu (E)	8	1.6%

Sulphur (S): 21

**Formula:** C<sub>2488</sub>H<sub>3831</sub>N<sub>633</sub>O<sub>653</sub>S<sub>21</sub>  
**Total number of atoms:** 7626

**Extinction coefficients:**  
**Ext. coefficient** 73255  
Abs 0.1% (=1 g/l) 1.363, assuming all pairs of Cysteine residues form cystines  
**Ext. coefficient** 72880  
Abs 0.1% (=1 g/l) 1.356, assuming all Cysteine residues are reduced  
(Unit of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water)

Gly (G)	45	9.1%
His (H)	10	2.0%
Ile (I)	40	8.0%
Leu (L)	45	9.1%
Lys (K)	14	2.8%
Met (M)	14	2.8%
Phe (F)	37	7.4%
Pro (P)	29	5.8%
Ser (S)	36	7.2%
Thr (T)	23	4.6%
Trp (W)	10	2.0%
Tyr (Y)	12	2.4%
Val (V)	47	9.5%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

**Estimated half-life:**  
The N-terminal of the sequence considered is M (Met).  
The estimated half-life is:

30 hours (mammalian reticulocytes, in vitro)	(B)	0	0.0%
>20 hours (yeast, in vivo)	(Z)	0	0.0%
>10 hours ( <i>E. coli</i> , in vivo).	(X)	0	0.0%

**Instability index:**  
The instability index (II) is computed to be 34.67  
This classifies the protein as stable.

**Aliphatic index:**  
104.19

**Grand average of hydropathicity (GRAVY):**  
0.548

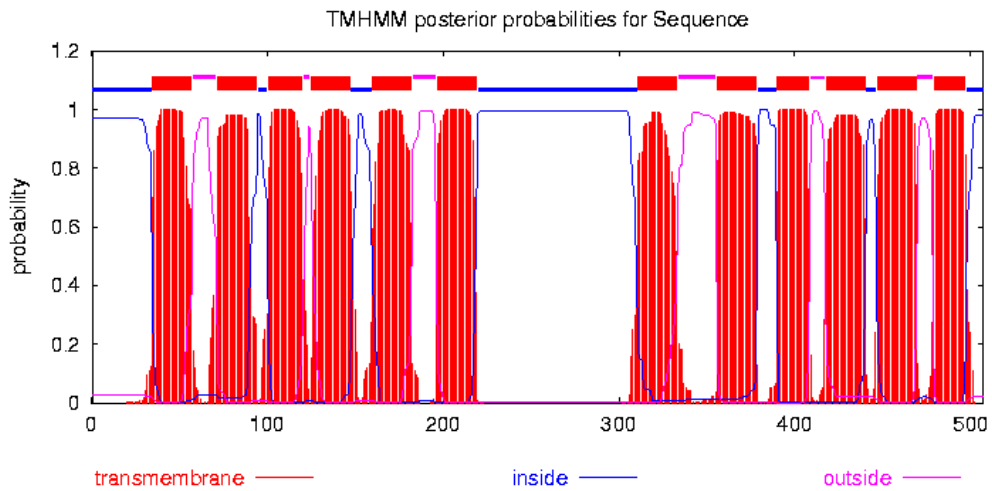
**Total number of negatively charged residues (N + E): 26**  
**Total number of positively charged residues (R+ K): 34**

## 5.2 TMHMM result Server v. 2.0

### 5.2.1 NrtA Analysis

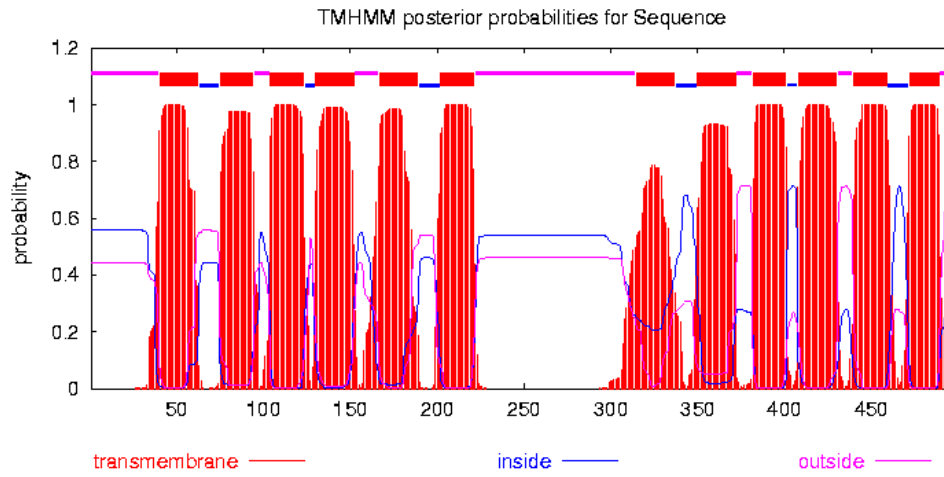
```
# Sequence Length: 507
# Sequence Number of predicted TMHs: 12
# Sequence Exp number of AAs in TMHs: 259.58402
# Sequence Exp number, first 60 AAs: 21.92685
# Sequence Total prob of N-in: 0.97219
# Sequence POSSIBLE N-term signal sequence
Sequence      TMHMM2.0      inside      1      34
Sequence      TMHMM2.0      TMhelix    35     57
Sequence      TMHMM2.0      outside    58     71
Sequence      TMHMM2.0      TMhelix    72     94
Sequence      TMHMM2.0      inside    95    100
Sequence      TMHMM2.0      TMhelix   101   120
Sequence      TMHMM2.0      outside   121   124
Sequence      TMHMM2.0      TMhelix   125   147
Sequence      TMHMM2.0      inside   148   159
Sequence      TMHMM2.0      TMhelix   160   182
Sequence      TMHMM2.0      outside   183   196
Sequence      TMHMM2.0      TMhelix   197   219
Sequence      TMHMM2.0      inside   220   310
Sequence      TMHMM2.0      TMhelix   311   333
Sequence      TMHMM2.0      outside   334   355
Sequence      TMHMM2.0      TMhelix   356   378
Sequence      TMHMM2.0      inside   379   389
Sequence      TMHMM2.0      TMhelix   390   408
```

Sequence	TMHMM2.0	outside	409	417
Sequence	TMHMM2.0	TMhelix	418	440
Sequence	TMHMM2.0	inside	441	446
Sequence	TMHMM2.0	TMhelix	447	469
Sequence	TMHMM2.0	outside	470	478
Sequence	TMHMM2.0	TMhelix	479	497
Sequence	TMHMM2.0	inside	498	507



## 5.2.2 NrtB Analysis

Sequence	Length:	497		
Sequence	Number of predicted TMHs:	12		
Sequence	Exp number of AAs in TMHs:	252.48148		
Sequence	Exp number, first 60 AAs:	21.63818		
Sequence	Total prob of N-in:	0.55805		
Sequence	POSSIBLE N-term signal sequence			
Sequence	TMHMM2.0	outside	1	39
Sequence	TMHMM2.0	TMhelix	40	62
Sequence	TMHMM2.0	inside	63	74
Sequence	TMHMM2.0	TMhelix	75	94
Sequence	TMHMM2.0	outside	95	103
Sequence	TMHMM2.0	TMhelix	104	123
Sequence	TMHMM2.0	inside	124	129
Sequence	TMHMM2.0	TMhelix	130	152
Sequence	TMHMM2.0	outside	153	166
Sequence	TMHMM2.0	TMhelix	167	189
Sequence	TMHMM2.0	inside	190	201
Sequence	TMHMM2.0	TMhelix	202	221
Sequence	TMHMM2.0	outside	222	314
Sequence	TMHMM2.0	TMhelix	315	337
Sequence	TMHMM2.0	inside	338	349
Sequence	TMHMM2.0	TMhelix	350	372
Sequence	TMHMM2.0	outside	373	381
Sequence	TMHMM2.0	TMhelix	382	401
Sequence	TMHMM2.0	inside	402	407
Sequence	TMHMM2.0	TMhelix	408	430
Sequence	TMHMM2.0	outside	431	439
Sequence	TMHMM2.0	TMhelix	440	459
Sequence	TMHMM2.0	inside	460	471
Sequence	TMHMM2.0	TMhelix	472	489
Sequence	TMHMM2.0	outside	490	497



## 5.3 ProtScale Tool

### 5.3.1 Summary of NrtA constructs analysed in this study using ProtScale tool

Construct	Bulkiness 0-20	Burried 0-20
<b>K19</b>	16	5
<b>A20</b>	16	5
<b>F36</b>	16.5	6
<b>W40</b>	16.6	5.5
<b>F43</b>	15.9	6.9
<b>F47</b>	17.7	6.6
<b>Y51</b>	17.3	5.7
<b>L84</b>	19	9.7
<b>L88</b>	17	8.4
<b>P113</b>	14	8.1
<b>R130</b>	17	8.2
<b>F140</b>	14.5	7.4
<b>F151</b>	14.5	5.6
<b>W197</b>	15.5	6.0
<b>R198</b>	16.6	5.7
<b>Y323</b>	16	7.8
<b>G328</b>	12.6	8.1
<b>F362</b>	16.4	7.9
<b>N364</b>	16.4	6.9
<b>C367</b>	15	7.3
<b>K388</b>	18	6.3
<b>F457</b>	11.7	9
<b>G448</b>	13.2	8
<b>G452</b>	11.5	10.1
<b>G455</b>	10.4	8.6
<b>G456</b>	11.7	9
<b>G458</b>	10.3	10.1
<b>N459</b>	10.3	9.7
<b>L460</b>	12.3	9.3
<b>G461</b>	14	8.6
<b>G462</b>	13.1	9.2
<b>R130QR198Q</b>	N/A	N/A

## Bulkiness

Using the scale the individual values for the 20 amino acids:

Ala: 11.500 Arg: 14.280 Asn: 12.820 Asp: 11.680 Cys: 13.460 Gln: 14.450  
Glu: 13.570 Gly: 3.400 His: 13.690 Ile: 21.400 Leu: 21.400 Lys: 15.710  
Met: 16.250 Phe: 19.800 Pro: 17.430 Ser: 9.470 Thr: 15.770 Trp: 21.670  
Tyr: 18.030 Val: 21.570

## % buried residues

Using the scale the individual values for the 20 amino acids:

Ala: 11.200 Arg: 0.500 Asn: 2.900 Asp: 2.900 Cys: 4.100 Gln: 1.600  
Glu: 1.800 Gly: 11.800 His: 2.000 Ile: 8.600 Leu: 11.700 Lys: 0.500  
Met: 1.900 Phe: 5.100 Pro: 2.700 Ser: 8.000 Thr: 4.900 Trp: 2.200  
Tyr: 2.600 Val: 12.900

---

## 5.4 Phosphorylation Sites

### 5.4.1 Kinase specific eukaryotic protein phosphorylation sites analysis (NrtA)

The NetPhosK 1.0 server (Blom *et al.* 2004) was used and the highest score obtained (0.88 PKC) was at position 353 (threonine)

### 5.4.2 NetPhos 2.0 Server (Blom *et al.* 1999)

This tool was used to predict phosphorylation sites of NrtA using the SwissProt FASTA sequence (P22152). There were 11 serine, 8 threonine and 1 tyrosine phosphorylation sites predicted for NrtA. Kinase specific phosphorylation predictions are available at: <http://www.cbs.dtu.dk/services/NetPhosK/>

#### Serine predictions

Name	Position	Context	Score	Pred
Sequence	10	LLVASPEVN	0.983	*S*
Sequence	39	VFFFSWFGF	0.050	.
Sequence	49	LAFLSWYAF	0.199	.
Sequence	68	DLDMSQTQI	0.655	*S*
Sequence	75	QIANSNIIA	0.151	.
Sequence	111	LLVGS IPTA	0.004	.
Sequence	122	GLVTSPQGL	0.110	.
Sequence	154	FFDKSIVGT	0.479	.
Sequence	161	GTANSLAAG	0.011	.
Sequence	184	AIFDSLIRD	0.027	.
Sequence	228	TGKWSERHI	0.178	.
Sequence	242	TQTASKGNI	0.614	*S*
Sequence	250	IVDLSSGAQ	0.341	.
Sequence	251	VDLSSGAQS	0.359	.
Sequence	255	SGAQSSRPS	0.967	*S*
Sequence	256	GAQSSRPSG	0.466	.
Sequence	259	SSRPSGPPS	0.978	*S*
Sequence	263	SGPPSIIAY	0.677	*S*
Sequence	285	LEPQSQAIG	0.016	.
Sequence	301	NAVASPSRK	0.988	*S*
Sequence	303	VASPSRKEA	0.997	*S*
Sequence	313	NVIFSLATM	0.006	.
Sequence	326	PYACSFSGSE	0.003	.
Sequence	329	CSFGSELAI	0.003	.
Sequence	335	LAIN SILGD	0.006	.
Sequence	393	KLLLSFLGV	0.015	.

Sequence	409	AMGFSDPKS	0.153	.
Sequence	413	SDPKSEATM	0.646	*S*
Sequence	430	FFLESCNGA	0.006	.
Sequence	437	GAI FSLVPH	0.005	.
Sequence	451	NGIVSGMVG	0.045	.
Sequence	472	IFRYSHHDY	0.996	*S*
Sequence	487	LGVISMAVF	0.011	.
Sequence	493	AVFISVSWV	0.045	.
Sequence	495	FISVSWVRP	0.303	.
Sequence	503	PVPKSQMRE	0.979	*S*

^

Threonine predictions

Name	Pos	Context	Score	Pred
v				
Sequence	22	RKALTIPVL	0.384	.
Sequence	31	NPFNTYGRV	0.564	*T*
Sequence	58	PLLTVTIR	0.073	.
Sequence	60	LLTVTIRDD	0.705	*T*
Sequence	70	DMSQTQIAN	0.230	.
Sequence	83	ALLATLLVR	0.010	.
Sequence	114	GSIPTAMAG	0.039	.
Sequence	121	AGLVTSPQG	0.023	.
Sequence	139	ILGGTFVPC	0.075	.
Sequence	148	QVWCTGFFD	0.011	.
Sequence	158	SIVGTANSL	0.048	.
Sequence	174	GGGITYFVM	0.043	.
Sequence	218	GMLFTCDDT	0.475	.
Sequence	222	TCDDTPTGK	0.562	*T*
Sequence	224	DDTPTGKWS	0.605	*T*
Sequence	238	MKEDTQTAS	0.368	.
Sequence	240	EDTQTASKG	0.226	.
Sequence	277	EKKGTETPL	0.211	.
Sequence	279	KGTEPLEP	0.948	*T*
Sequence	316	FSLATMAVA	0.033	.
Sequence	351	YMGQTQTGK	0.738	*T*
Sequence	353	GQTQTGKWA	0.913	*T*
Sequence	382	LYRKTNTPW	0.100	.
Sequence	384	RKTNTPWAK	0.813	*T*
Sequence	416	KSEATMFGL	0.266	.
Sequence	421	MFGLTAGLA	0.062	.

^

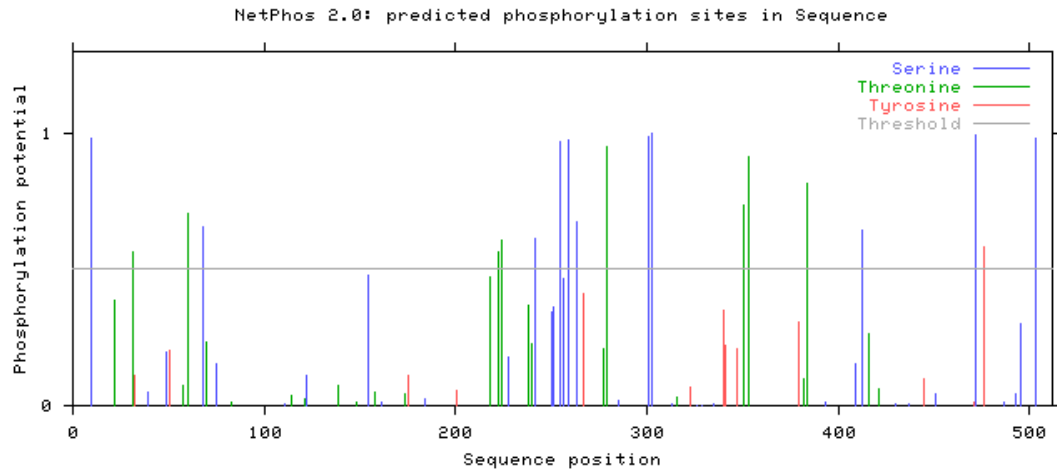
Tyrosine predictions

Name	Pos	Context	Score	Pred
v				
Sequence	32	PFNTYGRVF	0.109	.
Sequence	51	FLSWYAFPP	0.204	.
Sequence	175	GGITYFVMP	0.108	.
Sequence	201	WRVAYIVPF	0.054	.
Sequence	267	SIIAYAIPD	0.408	.
Sequence	323	VAVPYACSF	0.068	.
Sequence	340	ILGDYYDKN	0.348	.
Sequence	341	LGDYDKNF	0.221	.
Sequence	347	KNFPYMGQT	0.209	.
Sequence	379	ADFLYRKTN	0.305	.
Sequence	445	HVHPYANGI	0.099	.
Sequence	471	IIFRYSHHD	0.014	.
Sequence	476	SHHDYARGI	0.584	*Y*

^

Output score (value in the range 0.000-1.000); scores above the threshold of 0.500 are assigned as '\*S\*', '\*T\*', or '\*Y\*'; (.) indicates a residue predicted not

to be phosphorylated due to low score or below the threshold score also to indicate when the residue is not S, T, or Y.



## 5.5 NetSurfP (Gasteiger *et al.* 2005)

Protein Surface Accessibility and Secondary Structure Predictions

Construct	NrtA
K19	E
A20	B
F36	B
W40	B
F43	B
F47	B
Y51	B
L84	B
L88	B
P113	B
R130	B
F140	B
F151	B
W197	B
R198	B
Y323	B
G328	E
F362	E
N364	B
C367	E
K388	B
F457	B
G448	B
G452	B
G455	B
G456	B
G458	B
N459	B
L460	B
G461	B
G462	B

B for buried or E for Exposed - Threshold: 25% exposure