Functional analysis of the transcription factor Pax7 in mouse muscle stem cells
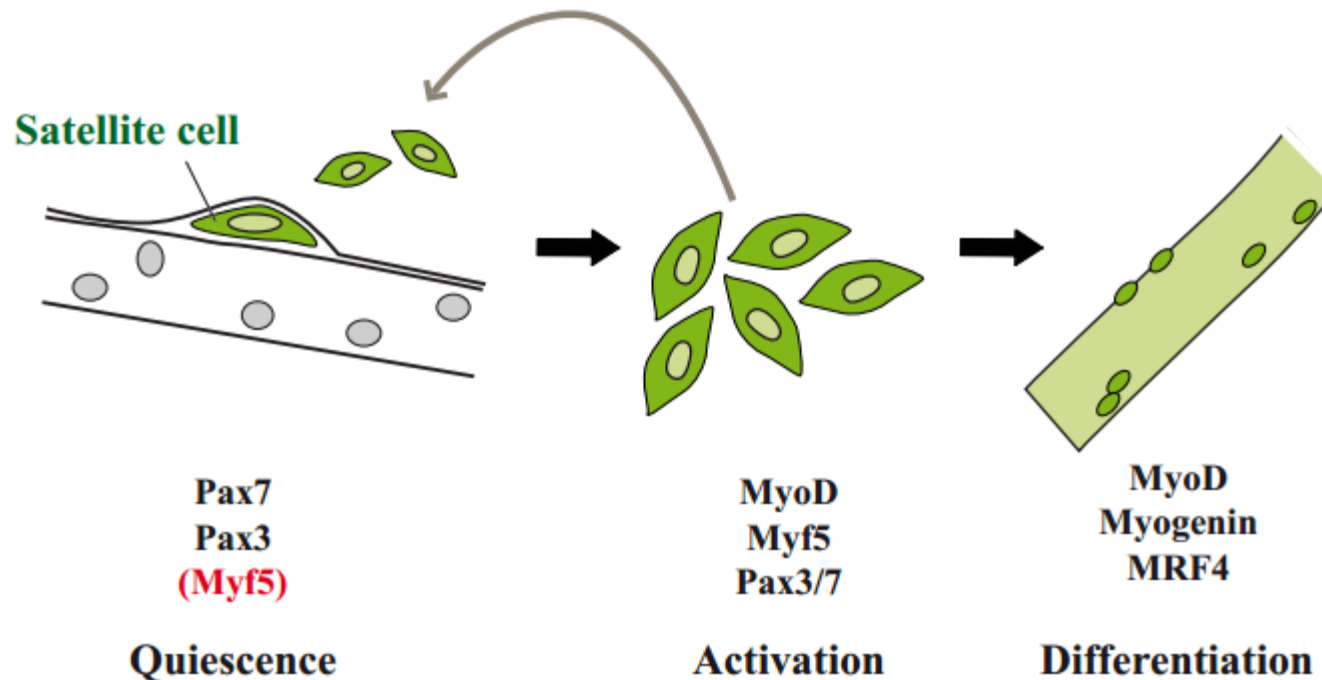
汇报小组：G3,G4
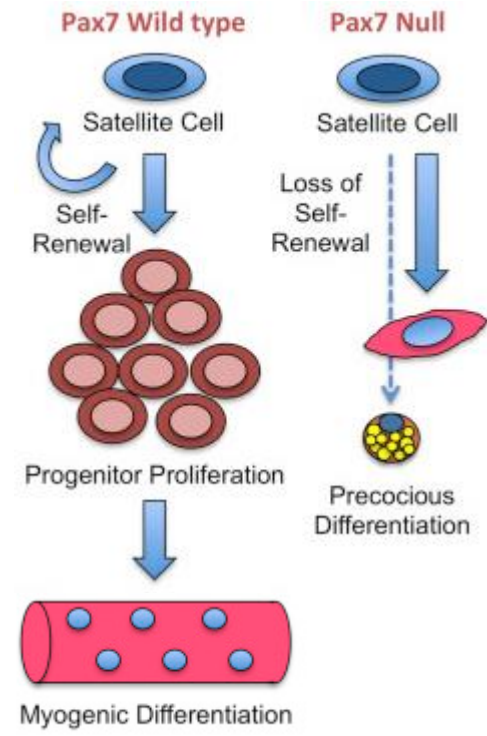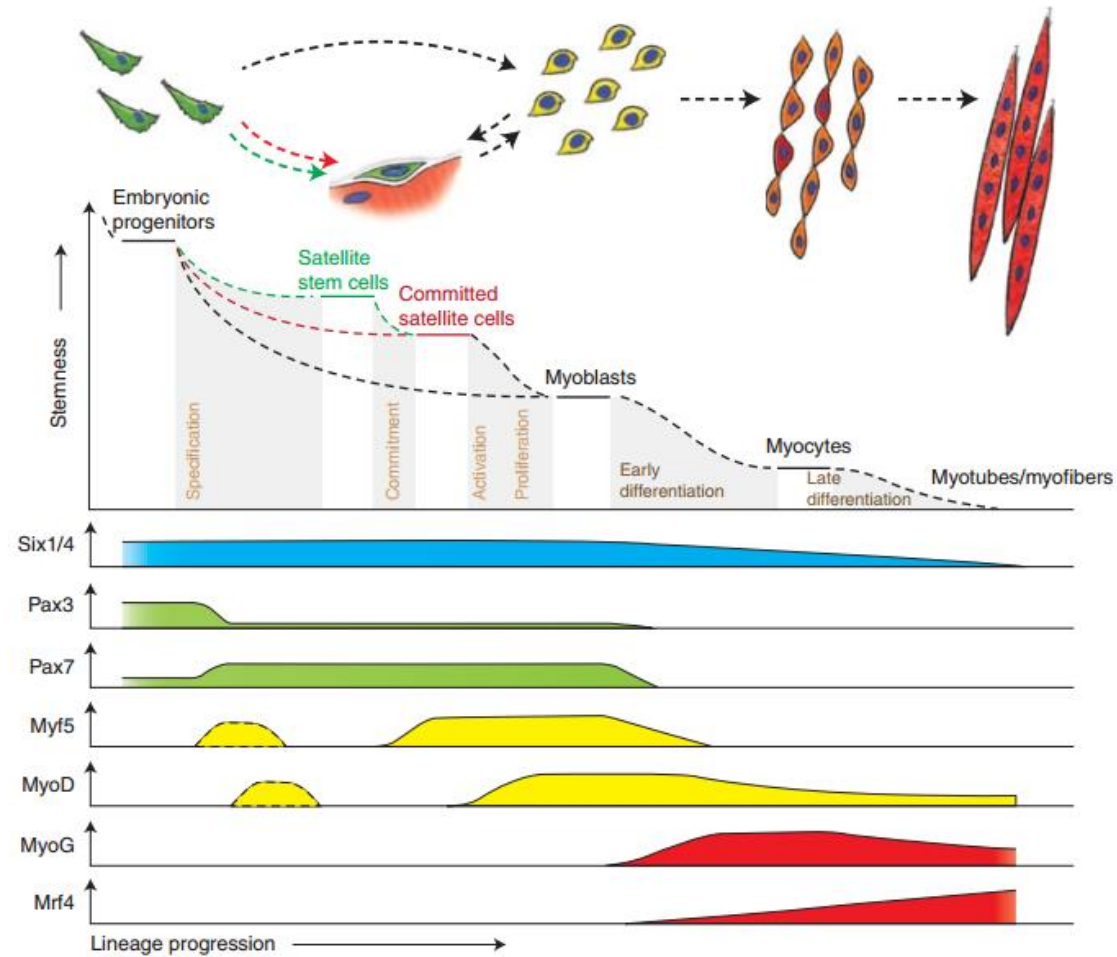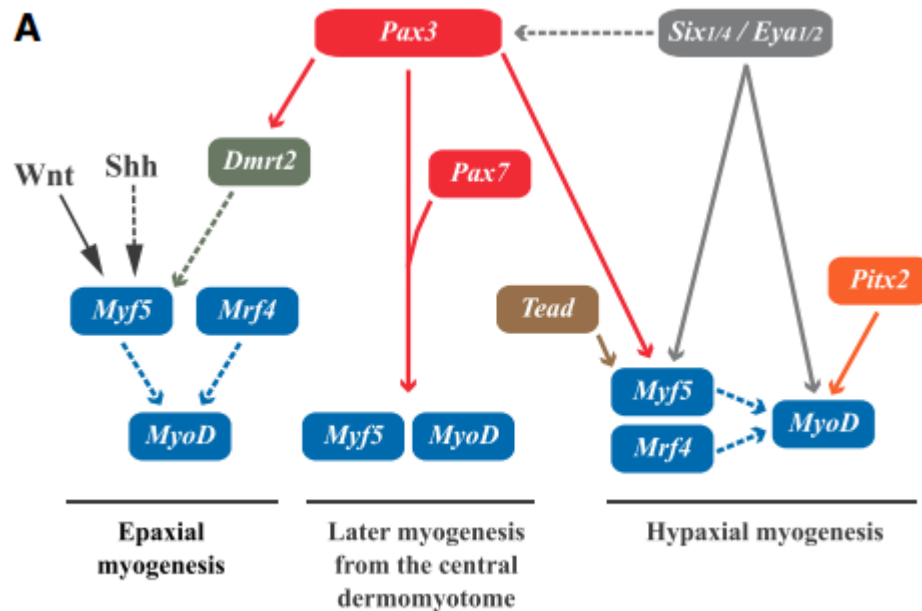
汇报人：G4-03

时间：20200104

Background

骨骼肌卫星细胞（Muscle satellite cells， MuSCs）是位于肌纤维质膜和基板之间的成体肌源性干细胞，正常生理条件下处于G0期静止状态，转录因子PAX7在MuSCs维持过程中发挥了重要作用。在外伤或病理状态下，PAX7+ MuSCs会从G0期激活进入G1期，表达肌肉生长因子MYOD，PAX7表达下调，诱导分化为肌原细胞，使肌肉组织得以修复。
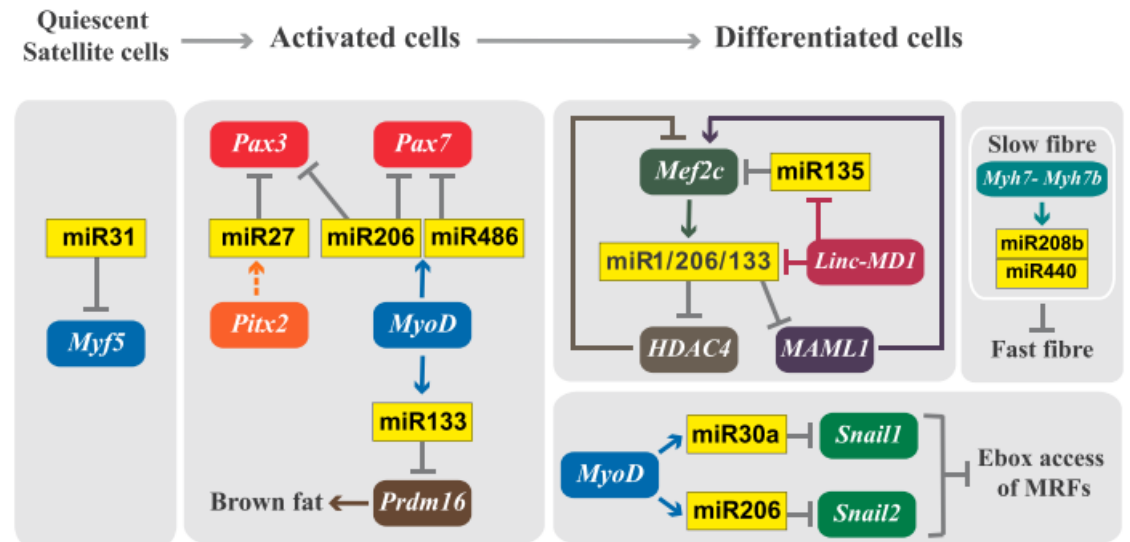
Pax7 is one of the muscle stem cell marker.

Role of Pax7 in Gene Regulatory networks

Gene Regulatory Networks
included Pax7 that Govern Myogenesis

MicroRNA Regulatory Networks in Satellite Cells,
when Cultured or during Regeneration of Adult Muscle after Injury

ChIP-seq

Cross-link protein to DNA

Shear DNA strands by sonicating

nucleus

cell lysate

Add bead-attached antibodies to immunoprecipitate target protein

precipitate

unlink protein; purify DNA

sequencing

map to genome

ATGCCTGGACCGTG

https://upload.wikimedia.org/wikipedia/commons/d/d8/Chromatin_immunoprecipitation_sequencing.svg

A

refGene   Myf5   Myf6

chr10   106900K   106910K   106920K   106930K   106940K

B

38.6%
1.2%
0.3%
3.6%
56.7%

- ☐ 5' UTR
- ■ exons
- ☐ 3' UTR
- ■ introns
- ■ intergenic

Pax7 remodels the chromatin landscape in skeletal muscle stem cells

MEME
Multiple Em for Motif Elicitation

Version 5.1.0

Data Submission Form

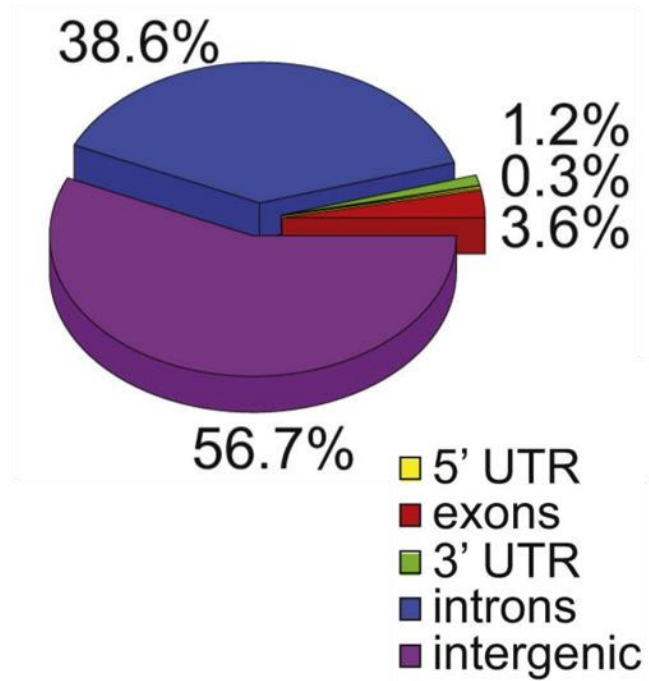Perform motif discovery on DNA, RNA, protein or custom alphabet datasets.

Select the motif discovery mode [?]

⦿ Classic mode   ◯ Discriminative mode   ◯ Differential Enrichment mode [NEW]

Select the sequence alphabet

Use sequences with a standard alphabet or specify a custom alphabet. [?]

⦿ DNA, RNA or Protein   ◯ Custom   选择文件 未选择任何文件

Input the primary sequences

Enter sequences in which you want to find motifs. [?]

[Upload sequences ▾]   选择文件 未选择任何文件 [?]

Select the site distribution

How do you expect motif sites to be distributed in sequences? [?]

[Zero or One Occurrence Per Sequence (zoops) ▴▾]

Select the number of motifs

How many motifs should MEME find? [?]

[3]

Input job details

(Optional) Enter your email address. [?]

[   ]

(Optional) Enter a job description. [?]

[   ]

▶ Advanced options

Note: if the combined form inputs exceed 80MB the job will be rejected.

[Start Search]   [Clear Input]
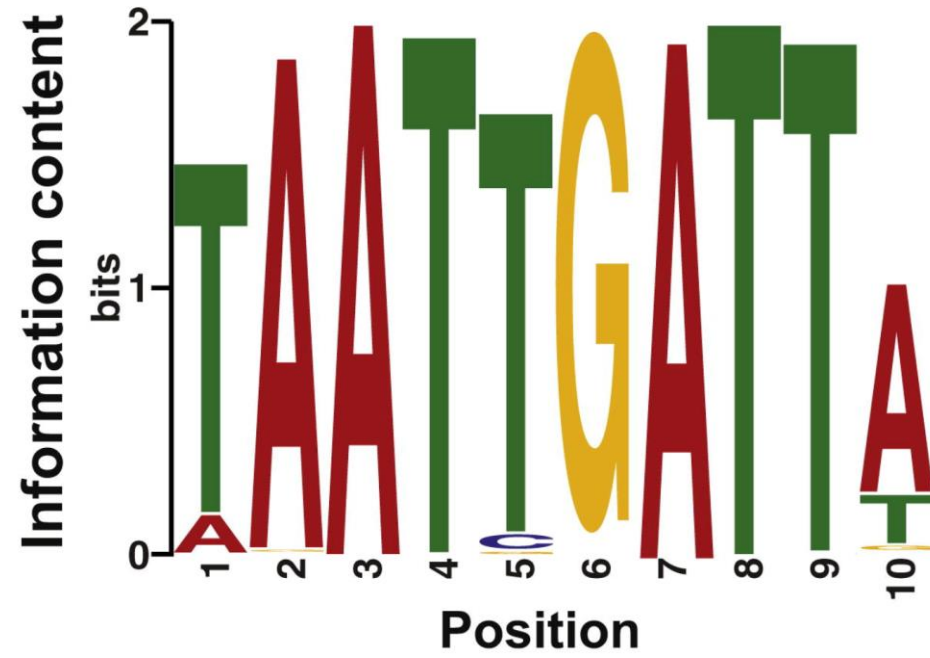
Version 5.1.0   Please send comments and questions to: meme-suite@uw.edu   Powered by Opal

Home   Documentation   Downloads   Authors   Citing

Pax7 binding DNA motif

Pax7基因的orf、cds序列、蛋白理化性质预测

一、找orf，并验证是否正确

- 首先在NCBI，nucleutide搜索Pax7，选择accession为NM_011039.2的小鼠DNA序列，用emboss-explorer网站的plotorf功能，得到如下结果：
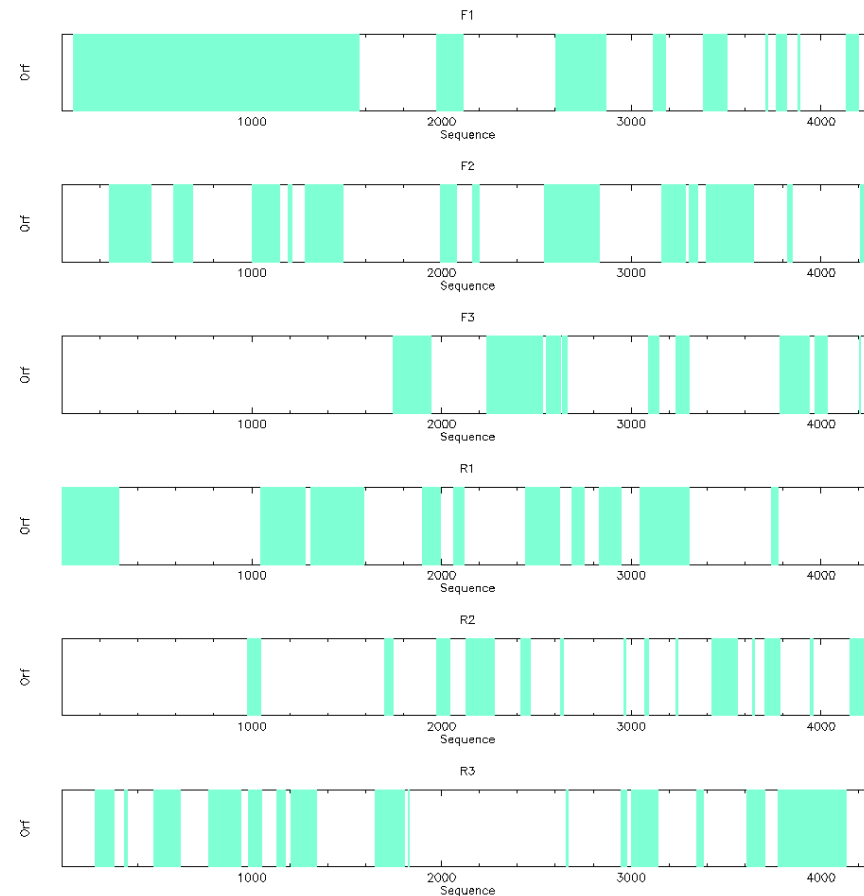
一、找orf，并验证是否正确

- 判断ORF在F1，长度在1000-2000之间。之后用emboss-explorer网站的getorf功能，经过上图判断，调整getorf参数，并设置output为起始密码子和终止密码子之间的核苷酸序列，得到orf;

- 接下来通过Sequence Manipulation Suite (SMS)网站中GenBank Feature Extractor功能，将该基因的GenBank作为输入，得到cds的核苷酸序列，二者通过EMBOSS Needle进行DNA的双序列比对，得到如下结果。二者序列完全一致（gap=3是因为预测的orf未包含终止密码子）

一、找orf，并验证是否正确

- 接下来通过Sequence Manipulation Suite (SMS)网站中GenBank Feature Extractor功能，将该基因的GenBank作为输入，得到cds的核苷酸序列，二者通过EMBOSS Needle进行DNA的双序列比对，得到如下结果。二者序列完全一致（gap=3是因为预测的orf未包含终止密码子）

```
#=======================================
#
# Aligned_sequences: 2
# 1: CDS
# 2: NM_011039.2_1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1512
# Identity:    1509/1512 (99.8%)
# Similarity:  1509/1512 (99.8%)
# Gaps:          3/1512 ( 0.2%)
# Score: 7545.0
#
#
#=======================================
```

```
CDS              1 atggcggcgctgcccggcgcgcggtccccaggatgatgagacccggcccggg       50
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
NM_011039.2_1    1 ATGGCGGCGCTGCCCGGCGCGCGGTCCCCAGGATGATGAGACCCGGCCCGGG       50

CDS             51 gcagaactacccgcgcaccggcttccccctggaagtgtccacccctcttg      100
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
NM_011039.2_1   51 GCAGAACTACCCGCGCACCGGCTTCCCCCTGGAAGTGTCCACCCCTCTTG      100

CDS            101 gccaaggccgggtcaatcagcttggtggggtcttcatcaacggtcgaccc      150
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
NM_011039.2_1  101 GCCAAGGCCGGGTCAATCAGCTTGGTGGGGTCTTCATCAACGGTCGACCC      150
```

二、翻译为蛋白并验证序列是否正确

- 将cds的DNA序列预测得到的氨基酸序列，和uniprot中该蛋白的序列利用EMBOSS Needle进行蛋白的双序列比对，发现二者完全一致，得分如下：

```
#=======================================
#
# Aligned_sequences: 2
# 1: PAX7_MOUSE
# 2: _1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 503
# Identity:     503/503 (100.0%)
# Similarity:   503/503 (100.0%)
# Gaps:           0/503 ( 0.0%)
# Score: 2651.0
#
#
#=======================================
```

三、理化性质

- 直接在ExPASy里利用登录号P47239进行compute parameters
  The sequence PAX7_MOUSE consists of 503 amino acids
  利用缺省值进行submit，得到如下结果：

  Number of amino acids: 503
  Molecular weight: 54938.72
  Theoretical pI: 9.10

Amino acid composition:

- Ala (A) 36   7.2%
- Arg (R) 36   7.2%
- Asn (N) 15   3.0%
- Asp (D) 24   4.8%
- Cys (C)   6   1.2%
- Gln (Q) 26   5.2%
- Glu (E) 27   5.4%
- Gly (G) 44   8.7%
- His (H) 15   3.0%
- Ile (I) 20   4.0%

- Leu (L) 36   7.2%
- Lys (K) 24   4.8%
- Met (M) 11   2.2%
- Phe (F) 12   2.4%
- Pro (P) 45   8.9%
- Ser (S) 52 10.3%
- Thr (T) 28   5.6%
- Trp (W) 3   0.6%
- Tyr (Y) 17   3.4%
- Val (V) 26   5.2%

Charged Residues

- Total number of negatively charged residues (Asp + Glu): 51
  Total number of positively charged residues (Arg + Lys): 60

- Atomic composition:

Carbon	C	2393
Hydrogen	H	3787
Nitrogen	N	709
Oxygen	O	744
Sulfur	S	17

- Formula: C2393H3787N709O744S17   Total number of atoms: 7650

Extinction coefficients

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

- Ext. coefficient   42205
  Abs 0.1% (=1 g/l)   0.768, assuming all pairs of Cys residues form cystines

- Ext. coefficient   41830
 Abs 0.1% (=1 g/l)   0.761, assuming all Cys residues are reduced

Estimated half-life

- 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).
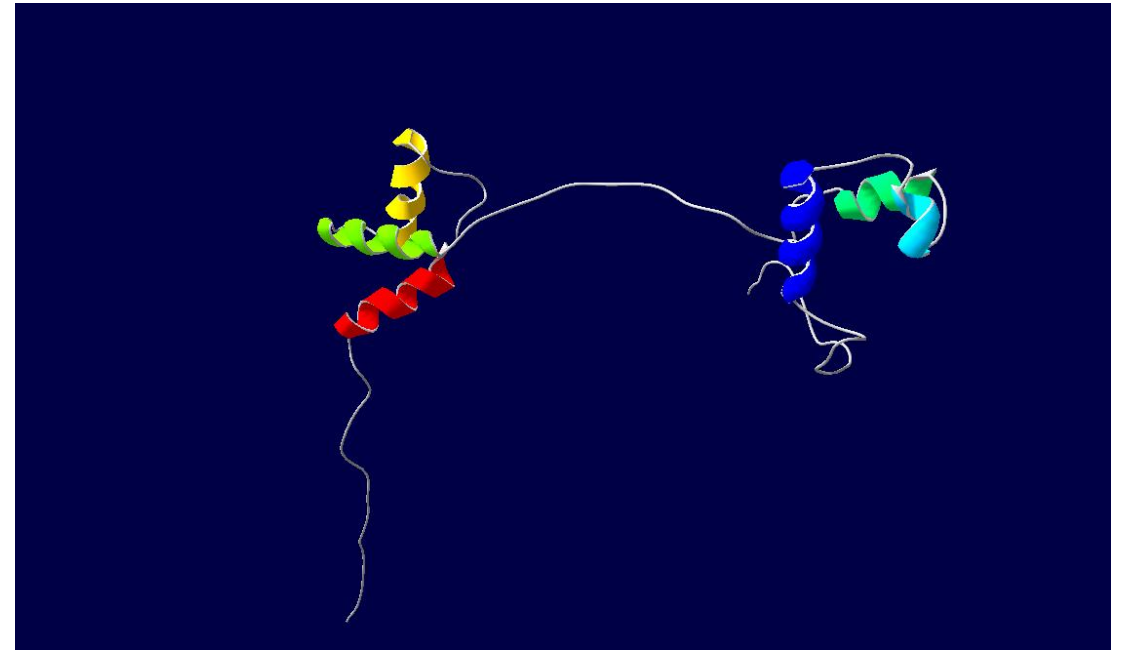 >10 hours (Escherichia coli, in vivo).

Instability index

- Instability index:
  The instability index (II) is computed to be 50.00
  This classifies the protein as unstable.

- Aliphatic index: 65.57

- Grand average of hydropathicity (GRAVY): -0.658
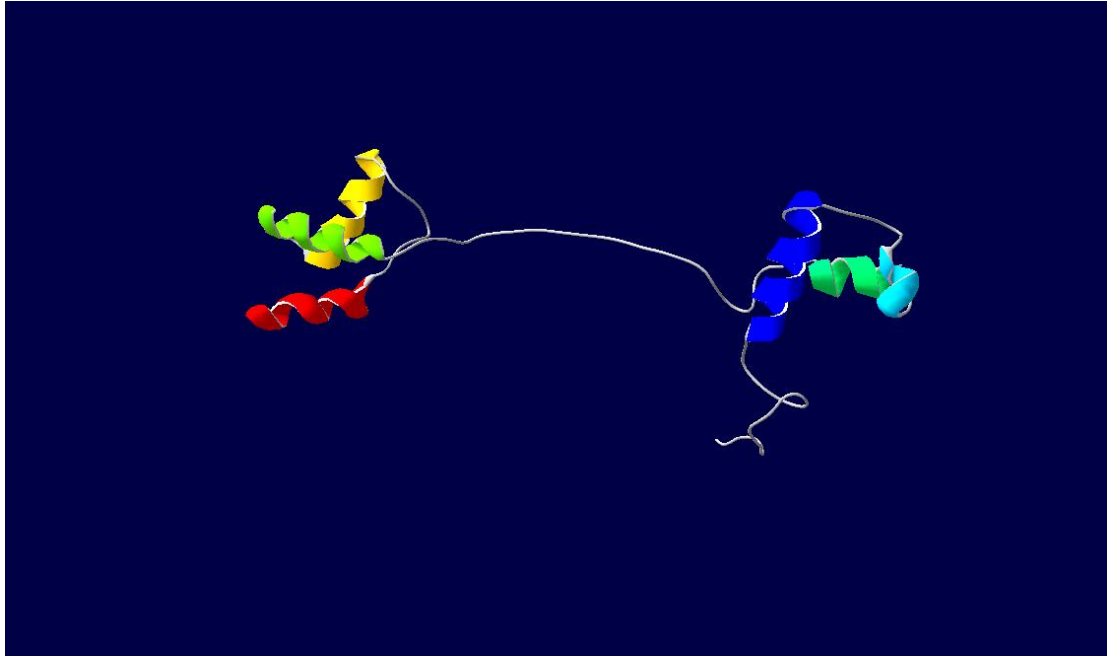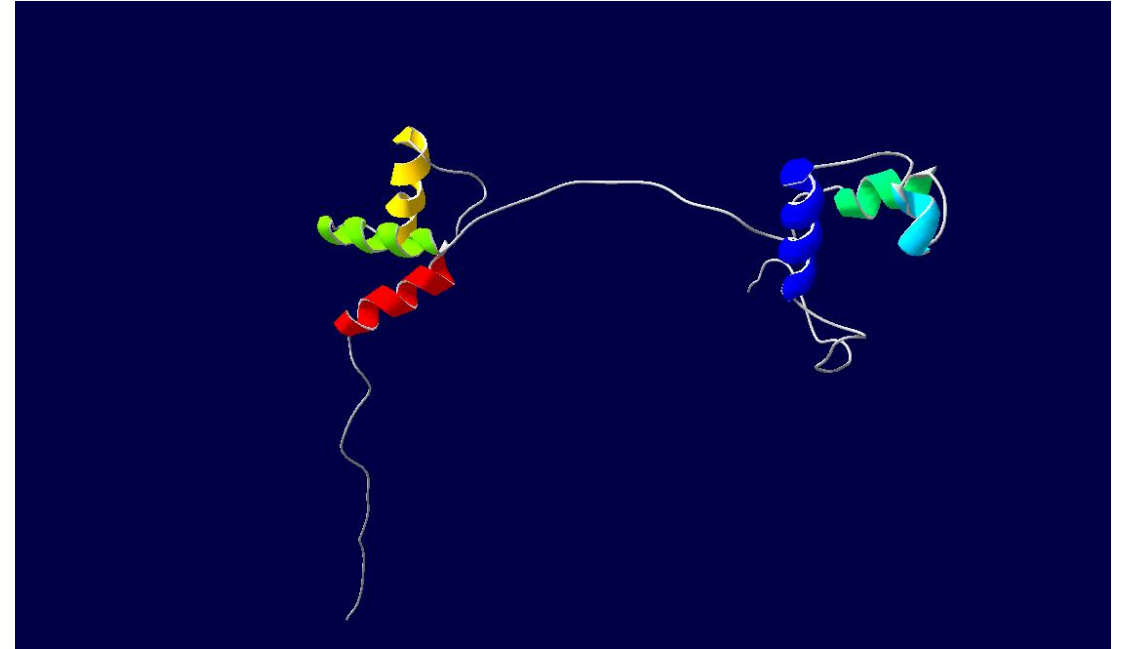
Protein modeling—phyre2
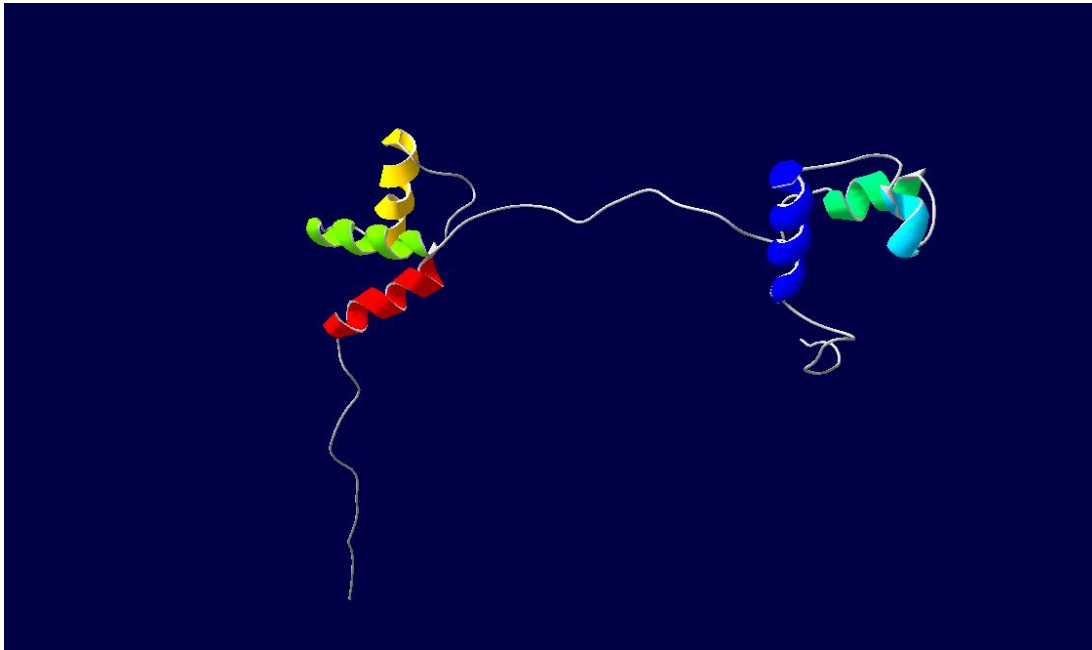
PAX7

top template—PAX8(human)

# Protein modeling—Swiss model

PAX7

template—pax8(human)
GMQE:0.17

蛋白名称、物种名称及对应的蛋白登录号

蛋白名称	物种名称	蛋白登录号（Entry name）
paired box 7	[Mus musculus (house mouse)]	P47239
paired box 7	[Homo sapiens (human)]	P23759
paired box 7	[Gallus gallus (chicken)]	O42349
paired box 7	[Rattus norvegicus (Norway rat)]	D3ZRA8
paired box 7	[Pan troglodytes (chimpanzee)]	H2PY64
paired box 7	[Pongo abelii (Sumatran orangutan)]	H2N8U5
paired box 7	[Chlorocebus sabaeus (green monkey)]	A0A0D9S8E8

Evolutionary relationships of taxa (Gene)
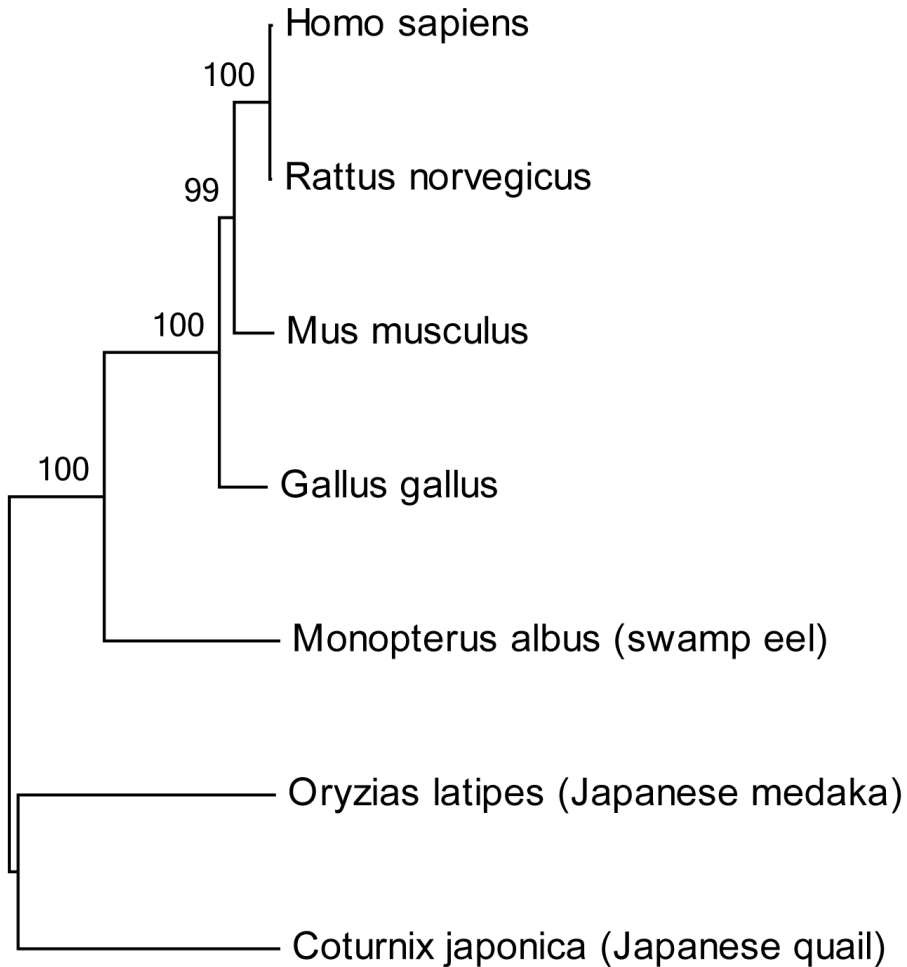
通过这7个基因的cds编码区序列，以Neighbor-Joining 法构建系统发生树，发现Homo sapiens（智人）和Rattus norvegicus（挪威大鼠）之间的亲缘关系最近。

G3和G4组期末汇报分工

1.小鼠Pax7蛋白的生物学背景介绍（王雅琦）

2.Pax7的组学研究背景--ChIP-seq数据分析（彭婷）

3.小鼠Pax7基因的orf预测和理化性质分析（魏琳、康美玲）

4.Pax7蛋白的结构预测（刘路路）

5.多物种Pax7基因cds序列和蛋白序列的进化分析（陈伟东）

6.PTT汇总修改和报告（钟新星）

谢谢