**Cell Biology Laboratory**

Module 11 – Project Practice 3: ER Transport

Table of Contents

[Lab Objectives\* 1](#_Toc54693147)

[ER Signal Sequences and Identifying a Protein 2](#_Toc54693148)

[**TASK 1 (Practice using 3of5) 6**](#_Toc54693149)

[**TASK 2 (Getting protein sequence to use 3of5 to validate cNLS Mapper) 7**](#_Toc54693150)

[**TASK 3 (Analysis of assigned protein – signal sequence) 9**](#_Toc54693151)

[**TASK 4 (Analysis of assigned protein – ER export signal) 9**](#_Toc54693152)

[**TASK 5 (Identify assigned protein) 10**](#_Toc54693153)

# Lab Objectives\*

**ER Signal Sequence**

*\* This laboratory module supplements topics being learned in Chapter 8 (endomembrane system) in the lecture portion of this course.*

1. Using appropriate syntax, create a text pattern algorithm to search for an ER specific signal sequence within the primary sequence of a protein
2. Given a protein’s sequence of amino acids, identify putative ER signal sequences
3. Use a protein’s sequence of amino acids to postulate the protein’s identity
4. Interpret an ER signal sequence analysis of the protein’s sequence of amino acids to support its postulated identify

# ER Signal Sequences and Identifying a Protein

**Purpose**

The purpose of this lab is to 1) identify putative ER signal sequences for import and export of a cellular protein, 2) use a protein’s sequence of amino acids to determine its putative identity and 2) practice skills needed for each student to complete their independent *in silico* project\*, namely generating and interpreting a hydropathy plot.

*\*Each student is assigned a unique, hypothetical protein code. During the semester, the student analyzes the code for a variety of classic cell protein properties. Prior to attempting each analysis independently, the instructor assigns a known protein for practice. Once a basic level of skill is attained, the student is responsible for developing a unique analysis on their own.*

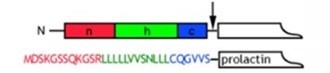
**Background**

Proteins secreted from the cell or located within the **endomembrane system** (including ER, peroxisome, Golgi, lysosome, transport vesicles and plasma membrane) are translated from a ribosome bound to the ER. For the ribosome to be ratcheted onto the cytosolic side of the ER, the **nascent proteins** are recognized by **signal recognition particles (SRP)**, which halts translation. The ribosome-nascent protein-SRP complex interacts with the **SRP-receptor** embedded in the ER membrane and is held to the ER as the ribosome is transferred to a **translocon** and the SRP is released. Once, SRP is released, translation resumes and the protein is translated into the ER lumen as either a soluble protein or an integral membrane protein (which requires that the protein has a stop transfer sequence).

**HOW PROTEINS ARE IMPORTED INTO THE ER**



SRP recognizes any nascent protein (protein in the process of being translated) that contains in its primary sequence an ER **signal sequence** (ER targeting signal).  Signal sequences have a **tripartite structure**, consisting of a hydrophobic core region (h-region) flanked by an n- and c-region.

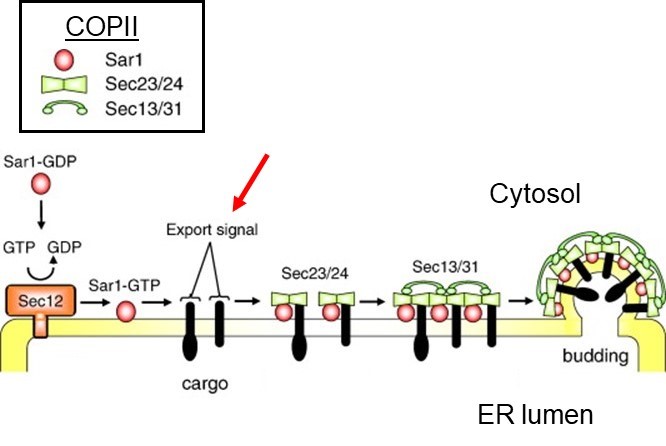


**Signal sequences** are typically located at the N-terminus of a protein and contain one or more positively charged (basic) amino acids [**lysine (K), arginine (R) or histidine (H)**]. followed by a *cluster* of 6-12 hydrophobic amino acids [**phenylalanine (F), leucine (L), alanine (A), tryptophan (W), valine (V), isoleucine (I), methionine (M)**] (such as in prolactin [a secreted hormone] RLLLLLVV).

**If the protein is secreted or resides in another part of the endomembrane system** (example, plasma membrane), the protein is packaged into a transport vesicle. Most of the proteins enter the vesicle by bulk flow. However, certain proteins have **ER export signals** or are ‘caught’ by ER membrane receptors that have ER export signals. These proteins contain both a signal sequence and an **ER-export signal**.

The ER export signal is used in the construction of a coat protein complex, called **COPII** (made up of Sar1, Sec23/24 and Sec13/31). Once the cytosolic coat is constructed, the vesicle buds from the ER and the proteins it contains (either soluble or integral membrane proteins) are transported to the Golgi (or sometimes the peroxisome).

**HOW PROTEINS ARE EXPORTED FROM THE ER**

****

**ER export signals** are typically located at the C-terminus of a protein and are diacidic containing one negatively charged amino acid [**aspartate (D) or glutamate (E)**], a spacer amino acid followed by one negatively charged amino acid [DE] (such as DXE).

Numerous web-based tools are available that identify targeting signal sequences in query protein sequences. For example, when we used the cNLS mapper, we were able to find mono- and bipartite nuclear localization signals. However, what if a novel tripartite nuclear localization signal (NLS) was discovered (doi: 10.1038/sj.cr.7290320), we wouldn’t be able to use cNLS to search for it.

To have a **user-friendly format** flexibility in searches is sacrifice, particularly when searching for novel signals or signals with length (such as ER signal sequence) and/or content (such as ER export signal) ambiguities. Therefore, **creating search-algorithms** based on published data gives a sequence analyst a whole new avenue for researching this protein of interest.

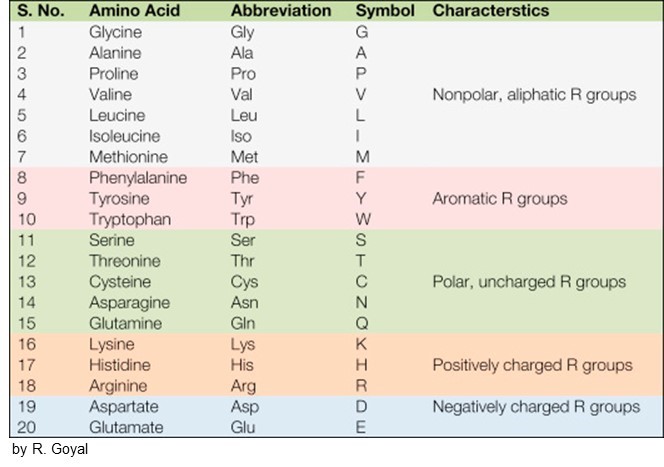
**Creating an Algorithm for Text Pattern Searches**



The “**3of5 website**” was developed by Dr. Seiler at the Division of Molecular Genome Analysis sponsored by the German Cancer Research Center (Deutsches Krebsforschungszentrum). This website allows the user to define their own search strings to predict variable sequences [Seiler *et al*., (2006), *BMC Bioinformatics* 7:14].

To use this web-based tool, we have to follow the rules of the program and put our text pattern search in the correct syntax. Here are a few of the (3of5) website’s commonly used **regular expressions** (RegEx), **a sequence of characters that define a search pattern.**

***In bioinformatics, amino acids are represented by their one letter code (see table)***



**No ambiguity (no inexactness) in content**

Enter the amino acid [it is what it is]

Example, to locate tryptophan search is W

**No ambiguity in length**

Enter the amino acid followed by the number in curly brackets

Example, to locate 9 tryptophans in a row, type W{9}

**Ambiguity (inexactness) in content**

To allow the presence of any particular amino acid at a given position, list all the possible amino acids surrounded by brackets.

Example, a positively charged amino acid in this position in the pattern, type [KHR]

To allow any amino acid at a given position, type a period

Example, glutamate-any amino acid-aspartate, type E.D

**Ambiguity in length**

To allow the repetition of an amino acid for a variable length enter the amino acid followed by range of numbers separated by a comma in curly brackets

Example, to locate 3 to 5 tryptophans in a row, type W{3,5}

**Combining RegEx**

Any combination of the regular expressions is allowed.

Example, to locate either lysine or arginine in each position in a stretch with a length between 1-3 positions (such as KKK or KRR or KRK, etc.), type [KR]{1,3}

Example, to allow any character in a stretch of 1-3 position, type .{1,3}

**Constraints on the Regular Expression** can be applied to limit or forbid an amino acid in a pattern. For example, any hydrophobic amino acid, except methionine, is allowed.

**Constraining Content**

If we want to forbid a particular amino acid from being in a particular position use a bracket and carat. Constraining content can be combined with ambiguity in length.

Example, allow any amino acid in one position except lysine or valine, type .[^KV]

Example, allow any amino acid in 1-3 positions except lysine, type .{1,3}[^K]

***Note: don’t forget the periods to indicate ‘any amino acid’***

**NofM Pattern Type**

The n-of-m pattern consists of two parentheses pairs with (nofm)(ACM). The first parenthesis pair contains information of the minimal number of a specified amino acid (n) and the total length of the pattern segment (m), while the second parenthesis pair determines the set of amino acids that are allowed (such as KRRxx, KKRxx, KKKxx, RKKxx, KxRxK, RRxxK, etc).

Example, at least 3 lysines and/or arginines in a stretch of 5 positions, type (3of5)(KR)

The n-of-m pattern can be combined with regular expressions to search for complex patterns.

Example, a bipartite nuclear localization signal searching for positively charged amino acids in two clusters separated by 10 of any amino acid, type [KR][KR].{10}(3of5)(KR)

*Practice:* Use these regular expressions to create an algorithm to search for an ER export signal. What is your text pattern search algorithm? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*Let’s start: ER export signals contain one negatively charged amino acid, a spacer amino acid followed by one negatively charged amino acid.*

Today, we are practicing with assigned proteins with a known identity. A protein’s sequence of amino acids can be identified using the US National Library of Medicine’s “**Basic Local Alignment Search Tool**” (BLAST) accessed through the “National Center for Biotechnology Information” (NCBI) sponsored by the US National Institute of Health (NIH). In short, we call it performing a BLAST search. This software aligns the protein of interest to all the proteins published in the database and provides the most likely match.

*These following tasks are conducted in silico so you need a computer with internet access.*

## TASK 1 (Practice using 3of5)

1. Open the (3of5) program:

[**https://3of5.dkfz.de/mga2/3of5/3of5.html**](https://3of5.dkfz.de/mga2/3of5/3of5.html)

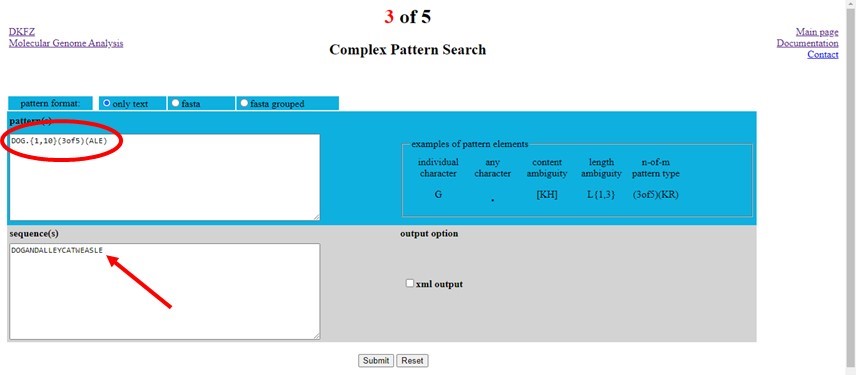
1. Into the pattern box, type the following pattern that looks for “DOG” followed by 1 to 10 of any character and subsequently 3 of any 5 characters being A, L or E.

**DOG.{1,10}(3of5)(ALE)**

*Note: don’t forget the period*

1. Based on this text pattern search, what is the smallest sequence length where this pattern would be possible? **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.**
2. Into the sequence(s) box, type the following

**DOGANDALLEYCATWEASLE**



1. Click “submit” and wait for Results page, which shows the examined patterns and the “matched in 1 sequences” list of results for position in the sequence, where matched, the length and pattern.
2. Results analysis. Look over the 6 matches listed on the results page.
   1. Start with DOG. Was the pattern found?
   2. Was your smallest sequence length prediction correct?
   3. In the last 5 characters is there at least 3 of any combination of A, L or E?
3. Answer question on the worksheet.

## TASK 2 (Getting protein sequence to use 3of5 to validate cNLS Mapper)

**Tasks 2, 3 and 4 completed with group.**

**If conducting through Zoom, join breakout room to begin**.

1. Open pubmed

[**https://www.ncbi.nlm.nih.gov/pmc/**](https://www.ncbi.nlm.nih.gov/pmc/)

1. Choose ‘protein’ from the pulldown menu and type the following into the search box to get the record for the regulator of chromosome condensation in homo sapiens. Hit Search. **NP\_001041659.1**



1. Once the file opens, click on FASTA



1. Highlight the protein sequence (starting with “M”) all the way to the end and copy.



1. Open the nuclear localization signal mapper program:

<http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi>

1. Paste the sequence into the box
2. Select a **cut-off score of 7**
3. Select **‘within terminal 60-amino-acid regions**
4. Click on **“Predict NLS”.**
5. **Note which position the bipartite sequence was found.**
6. Open the (3of5) program:

[**https://3of5.dkfz.de/mga2/3of5/3of5.html**](https://3of5.dkfz.de/mga2/3of5/3of5.html)

1. Into the pattern box, type the following pattern that looks for either K/R, followed by either K/R, followed by 10 of any character and then 3 of any 5 characters being K or R.

**[KR][KR].{10}(3of5)(KR)**

*Note: don’t forget the period*

1. Into the sequence(s) box, paste the **NP\_001041659.1** sequence from pubmed.
2. Click “submit” and wait for Results page.
3. Results analysis. Look over the 2 matches listed on the results page, particularly the position (pos) of the match.
4. Were the (3of5) results able to validate the cNLS mapper results? **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.**
5. Record the length of the (3of5) matches on the worksheet.

## TASK 3 (Analysis of assigned protein – signal sequence)

Using a (3of5) program and the amino acid sequence of the protein assigned to you, analyze the protein for ER signal sequence and ER export signal.

**Procedure for Identifying ER signal sequence**

*\* For ease (basically so you do not need to type in all of the letters of the protein sequence), the digital amino acid sequence for your assigned protein has been posted to Blackboard as a word document. Follow these steps:*

1. Open Blackboard to the folder containing the ER transport lab module.
2. Open the folder for ER transport assigned protein.
3. Open the word document with your assigned digital amino acid sequence.
4. Copy the digital amino acid sequence for assigned protein. Make sure to get the entire sequence from beginning to end.

EXAMPLE: MADVEKGKKIFIMKCSLCHTVEKGGKHKTGPNLHGLFGRKTGQ

1. Open the (3of5) program:

[**https://3of5.dkfz.de/mga2/3of5/3of5.html**](https://3of5.dkfz.de/mga2/3of5/3of5.html)

1. Paste the sequence into the bottom box labeled “sequence(s)”
2. Type into the top box labeled “pattern(s)” the following algorithm which searches for ER signal sequences:

[KHR][FLAWVIM]{6,12}

1. Click on “**Submit**” and Wait until the (3of5) result page is generated.
2. Record the start position of the signal sequence on worksheet.

***Make sure to also paste protein sequence into question#1 of on the worksheet.***

## TASK 4 (Analysis of assigned protein – ER export signal)

**Procedure for Identifying ER export signals**

1. Click on the ‘back’ arrow on the browser to return to the (3of5) homepage. The sequence should still be in box. If not, copy/paste it again.
2. Cut out the algorithm from the pattern(s) box and replace with **the algorithm you created for the ER export signal**
3. Click on “**Submit**” and Wait until the (3of5) result page is generated.
4. Record the start position of the signal sequence on worksheet.
5. **Each group chooses a spokesperson.** *(or ‘the person born closest to Des Moines”)*
6. **Group helps spokesperson develop a 15 second synopsis of results including:**
   1. **Based on results, how likely the protein is localized to the ER?**
   2. **Based on results, how likely the protein is exported from the ER?**
   3. **Confidence in algorithm created to search for ER export signal**
7. Exit breakout room and wait for everyone to enter main room. When called upon at random, spokespeople share synopsis about their protein
8. Progress to a BLAST search to support or contradict these results.

## TASK 5 (Identify assigned protein)

Use a BLAST search to determine the name of the assigned protein **and** research the location of the protein to assess whether the ER Transport determination makes sense for the determined protein.

**BLAST protein search**

1. Copy the same assigned digital amino acid sequence again from the word document posted to Blackboard.
2. Open the BLAST-protein program.

<http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome>

1. Copy the digital amino acid sequence for the assigned protein. Make sure to get the entire sequence from beginning to end.
2. Paste the sequence into the submission box in the BLAST program
3. Scroll to the middle of the program’s homepage to the “choose search set” area and type in the organism (optional) box the following: **9606**.

This is the taxonomic identification (taxid) number for *homo sapiens*.

1. Scroll to the bottom of the page and Hit the blue “BLAST” button
2. Wait while a list of putative proteins is generated. This may take a few minutes and the screen is automatically updated.
3. Record the description of the top putative protein on the worksheet.
4. Using your textbook, notes from the lecture and/or internet resources, research the assigned, now identified, protein.
5. Record on the worksheet the protein’s location and whether (3of5) ER signal sequence results support or contradict this location.

**DONE.**