

Making Molecular Tools to Determine the Effect of Amino Acid Supply on Protein Synthesis

Yutong Yang

Institute of Pharmacy and Biomedical Science, University of Strathclyde, Glasgow, UK

ABSTRACT

Background: The twenty amino acids making up proteins in animals are classified into three nutritional classes: EAAs, NEAAs and CEAAAs. Proteins with extreme amino acid ratios are closely related to growth, malnourishment, cancer and other health problems. Previous studies have shown that the amino acid composition of a protein can influence the protein expression level. The hypothesis that CEAA-rich proteins tend to reduce expression, especially in cells proliferating quickly is expected to be proven, which may provide new ideas for amino acid therapy for some diseases. **Lay Abstract:** The twenty amino acids making up proteins in animals are classified into three nutritional classes: EAAs, NEAAs and CEAAAs. Previous studies show that the amino acid composition of proteins can interact with amino acid availability to influence protein expression levels, independently of the amount of mRNA produced. By analysing the bioinformatic data of protein expression levels in human cells, protein expression was observed to be affected in the way that the more CEAAAs present in a protein, the lower its expression, which was particularly evident in cells proliferating quickly. In this project, plasmid constructs were created that express artificial proteins containing almost only one amino acid nutritional class with a V5 epitope at the terminal. These were transfected into A549 cells to test the hypothesis from data analysis in real-life cells that proteins with extreme proportions of CEAAAs tend to reduce expression and are most affected by proliferation compared with proteins with other proportions of amino acid class. The V5 tag is an epitope that can be detected by certain antibodies on Western blot or in cells. Through measuring the amount of certain A549 cell protein synthesis in different treatment groups, the results show that proteins with extreme proportions of CEAAAs tend to reduce in expression but are not the most affected by proliferation. However, proteins of what amino acid composition ratio are most affected by proliferation and the specific mechanism behind the effect need further study.

KEYWORDS

Cell Culture; Gateway® Technology; Microscopy; Immunofluorescence; Western Blot.

1. INTRODUCTION

1.1. The Way that Cells Make Protein

The process by which cells produce proteins can be summed up simply as the transformation of the order of bases in mRNA molecules into the order of amino acids in proteins or polypeptide chains. According to ‘Central Dogma’ [1], during transcription, information stored in DNA is copied into mRNA, In the next translation process, the ribosome synthesizes amino acid molecules into peptide chains according to the instructions of mRNA and completes the assembly of proteins [2]. Therefore, it is often assumed that the amount of protein is closely related only to the level of mRNA.

However, in fact, the quantity of each protein in a cell is only partially correlated with its gene transcription rate. The mRNA level of a typical gene and protein synthesis only shows a correlation coefficient of 0.4 to 0.6 [3,4]. Each protein's abundance in a cell is only partially related to the rate of gene transcription. mRNA sequence motifs, amino acyl-tRNA synthesis levels [5], elongation factor activity, protein susceptibility to degradation [6] and other factors also have independent effects on protein synthesis levels [7]. The research is aimed at exploring how the amino acid composition of proteins interacts with amino acid availability to affect protein expression.

1.2. Three Types of Amino Acids

People have the same twenty amino acids to make up all the proteins the body needs, and these amino acids are categorized into three nutritional groups. Actually, 500 million years ago, every creature had all the biochemical machinery to make each of the amino acids inside their bodies [8]. But as animals evolved, they destroyed the biochemical machinery for making nine of the amino acids under the pressure of natural selection [8,9], meaning they must get them through dietary supply. These nine amino acids are called essential amino acids (EAAs) [9]. There is another group called non-essential amino acids (NEAAs). Most people have all the amino acids in this group and never run out of them because the human body itself can produce them for synthesising proteins. There is also an intermediate group called conditionally essential amino acids (CEAAs), which are ones that people do have the machinery inside the body to make, but the machinery is not strong enough to meet the needs of the human body [10,11]. Under some stressful conditions, people do not have the capacity to make enough of CEAAs, so they also have to get those supplemented by diet [11]. Figure 1 gives an overview of the nutritional classification of amino acids.

Indispensable	Dispensable	Conditionally indispensable ^a
Histidine ^b	Alanine	Arginine
Isoleucine	Aspartic acid	Cysteine
Leucine	Asparagine	Glycine
Lysine	Glutamine acid	Glutamine
Methionine	Serine	Proline
Phenylalanine		Tyrosine
Threonine		
Tryptophan		
Valine		

^a Conditionally indispensable is defined as requiring a dietary source when endogenous synthesis cannot meet metabolic need.

^b Although histidine is considered indispensable, unlike the other eight indispensable amino acids, it does not fulfill the criteria used in this report of reducing protein deposition and inducing negative nitrogen balance promptly upon removal from the diet.

Figure 1. The classification of amino acids in the human diet (Figures taken from Laidlaw & Kopple, 1978). The amino acids are classified into indispensable (essential) ones, dispensable (non-essential) ones and conditionally indispensable (conditionally essential) ones. The corresponding types of amino acids are listed in the table.

1.3. The Ways that Amino Acid Composition Affects Protein Expression

By analysing the amino acid composition and protein expression levels in A549 cell lines from two different public expression datasets, previous research found that protein expression in A549 cell

lines was affected in two general rules: The first one is that proteins with relatively small proportions of EAAs were reduced in expression. The second one is that the more CEAAAs present in a protein, the lower its probable expression, which is particularly evident in cells that are proliferating quickly [12].

The evidence is shown below. In Figure 2.1, two expression datasets from the same lung cancer cell line, A549, exhibit notably different profiles. The protein expression of A549 (a) was not constrained while A549 (b) were described as actively proliferating. In both groups of A549 cells, higher CEAA compositions are generally associated with a modest reduction in A549 cell protein expression while higher EAA compositions are generally associated with an increase in expression. However, in A549 (b), there was a substantial increase in CEAA and EAA proportions (marked by green and red arrows) at lower expression levels, suggesting that proteins with greater proportions of those two amino acid classes were inefficiently translated. The differences between A549 (a) and (b) most likely reflect the effects of different proliferation rates on amino acid constraints. As a result, the inference about the interaction between proliferation rate, nutritional class and protein expression is made that if cells are dividing rapidly, it's hard to make enough proteins composed of CEAAAs because it is struggling to keep up with the demand [12].

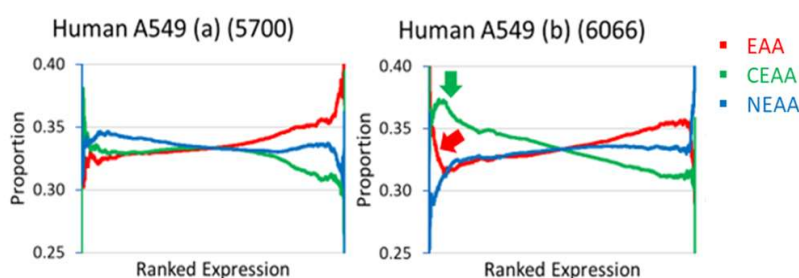


Figure 2.1 The relative proportions of the three nutritional amino acid classes change with protein expression levels in A549 cells with different proliferation rates (Figures taken from Thompson & Pickard, 2024, in press). The research methods refer to [13]. A smoothing procedure was applied to visualise trends in relative ranked amino acid class proportion when plotted against ranked protein expression level for A549 cells (data from PaxDB). Cumulative average proportions for each amino acid class were calculated from lowest-to-highest expressed proteins and, in parallel, highest-to-lowest expressed proteins. There was little evidence for constrained protein expression in A549(a) (data from [14]), while A549 (b) were described as actively proliferating (data from [15]). In the case of rapid proliferation, the expression of proteins with a high ratio of CEAA or EEA is significantly reduced (marked by green and red arrows respectively).

1.4. Practical Implications of the Research

As mentioned above, previous researches show that EAA or CEAA-rich proteins are not easy to make and tend to lose expression when EAAs or CEAAAs are being consumed [12]. So it may be assumed that amino acids making up the human body would have evolved to have an even amount of each amino acid class to avoid expression problems. However, this is not the case. The EAA and CEAA composition of all human proteins was plotted in the frequency histogram (Figure 2.2). It is shown that the evolution has placed many proteins at the outer edges of the composition range with extreme EAA or CEAA composition. The reason why this decision has a selective advantage is that these outlier proteins are the first to sense and drive biological responses to malnutrition to increase survival [12]. There is ample evidence that since animals lost the ability to make EAAs, they have evolved ways to make beneficial use of this seemingly problematical situation [16,17]. By making some

proteins purposefully susceptible to supply and demand issues of amino acids, evolution has allowed the reduced expression to signal 'hunger states' to the body: slowing metabolism, decreasing body size, and increasing the drive for more food [17].

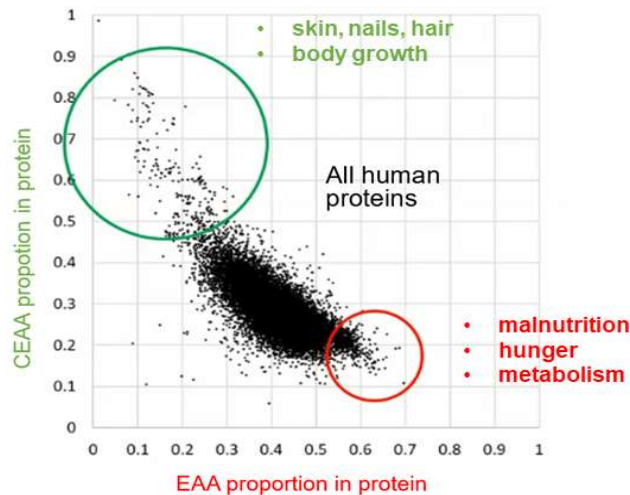


Figure 2.2 All human proteins are plotted according to fEAA against fCEAA (Figures taken from Thompson & Pickard, 2024, in press). The composition of 20,397 human proteins was plotted in the frequency histogram, fEAA against fCEAA. Proteins with a large proportion of CEAs circled in green are components of connective tissue, skin, nails hairs and their maturation enzymes involved in growth. Proteins with a large proportion of EAAs circled in red were observed in both leptin and the melanocortin receptor proteins, so they are closely connected with malnutrition, hunger and metabolism. Magnified outlier sectors of contained protein families or functionalities with extremely high fCEAA and fEAA are likely susceptible to reduced expression during nutritional insufficiency.

These findings could be applied to clinical treatments. For example, Child stunting is associated with low circulating EAAs and CEAs [18,19]. An inadequate amino acid supply will negatively impact the growth of cells and organisms [20]. In an amino acid malnourishment situation where CEEA-rich proteins are not easily produced, CEAs can be very useful as supplements to treat malnourishment and also to speed recovery from injury [21]. There are many established benefits to amino acid supplementation for patients undergoing wound healing or ulcer treatment [22,23].

Moreover, in the field of cancer, evidence shows that proteins linked to cancer have reduced levels of EAAs [12], potentially reflecting the amino acid 'starvation conditions' within the middle of tumours [24-26]. Protein, nucleotide, and fatty acid biosynthesis are crucial for the survival and malignant proliferation of cancer cells [27]. Cumulating research findings have demonstrated that amino acid restriction plays a role in cancer interventions, by attacking this weak point in its armour [28]. So dietary therapies are prevalent among cancer patients. However, there is lacking a certain rationale to clarify which strategy is optimal [28].

1.5. The Aim and Method of the Research

Our research is aimed at exploring how the amino acid composition of protein affects protein expression in the proliferating state. To measure the effect of different types of amino acids on protein

expression, the Gateway® Cloning Technology, which is a universal and efficient way to load DNA sequences into multiple vector systems [29], was used to create three kinds of pDEST40 plasmid constructs. Sequences encoding completely artificial proteins composed of almost only one amino acid class (EAA/CEAA/NEAA) were cloned into plasmids, with the V5 sequence at the terminal, which is an epitope that can be recognized by the certain antibody on Western blot or in cells, so that this V5 sequence can act as a reporter of levels of protein expression. The final structure of the pDEST40 plasmid is shown in Figure 3. The plasmids were then inserted into A549 cells (human lung epithelial cancer cell line). After selection, those successfully transfected cells were utilized as tools to help investigate the effect of proliferation on certain protein expressions. Proteins produced by each group were detected by immunofluorescence and quantified by Western blot for further analysis.

The hypothesis based on Thompson and Pickard’s paper [12] is expected to be proven from data in real-life cells that proteins with extreme proportions of CEAAAs tend to reduce expression and are most affected by proliferation compared with proteins with other proportions of amino acid class.

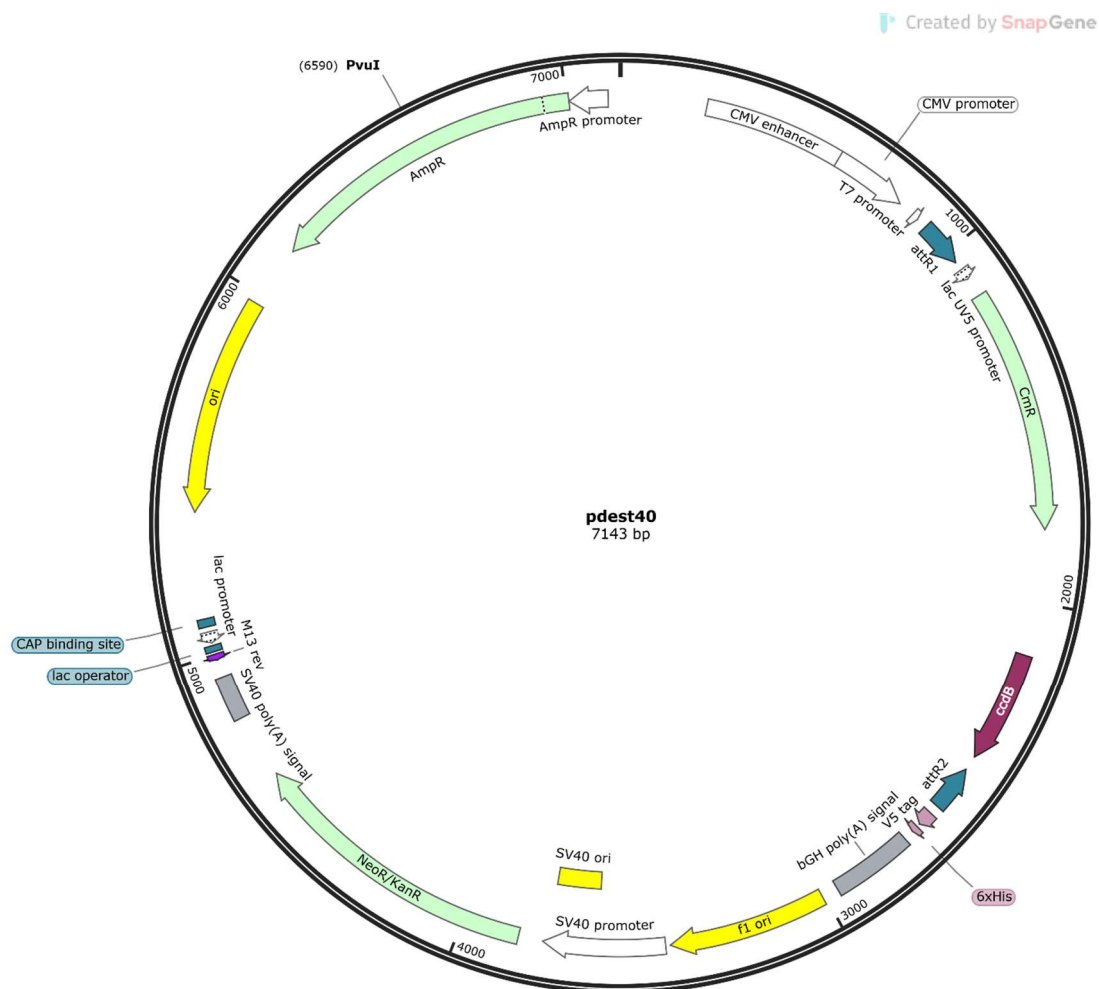


Figure 3. The structure of the pDEST40 plasmid. The figure was created by Snapgene. The pDEST40 plasmid is the commonly used Gateway® destination vector for expressing C-terminally V5- and 6xHis-tagged proteins in mammalian cells. The key components of the final vectors include CMV promoter, target gene between the attR1 and the attR2 sites, V5 tag and 6 Histidine. V5 sequence is an epitope which can be recognized by a certain antibody on Western blot or in cells.

2. MATERIALS AND METHODS

2.1. Gateway® Technology to Construct pDEST40 Plasmid

Gateway® Cloning Technology is a universal and efficient way to load DNA sequences into multiple vector systems, which is based on the site-specific recombination system used by *bacteriophage lambda* to integrate its DNA in the *E. coli* chromosome [29]. Gateway® Cloning Technology Kit (Invitrogen) was used according to the manufacturer's manual referring to [29] to get the *pDEST40* plasmid constructs with expressing target artificial proteins. The process contains BP reaction and LR reaction. The principle of the reactions is shown in Figure 4. PCR was used for the ligation of oligonucleotides into long stretches of sequence with *attB* ends. In the BP reaction, the following components were added to a 1.5 ml EP tube at RT and mixed: *attB*-PCR product, 10 µl *attB* expression clone, 2 µl pDONR™ vector, 4 µl 5XBP Clonase™ reaction buffer and TE buffer (volume to 16 µl). 4 µl BP Clonase™ enzyme was added to the components above, mixed well by vortexing briefly and incubated at 25°C for 1 h. 2 µl Proteinase K solution was added and the product was incubated at 37°C for 10 min. Competent *E. coli* was transformed and the antibiotic-resistant positive clones were selected. Then PCR was used to test the length of the cloned sequences and the longest ones were chosen for LR reaction. In LR Reaction, the following components were added to a 1.5 ml EP tube at RT and mixed: 10 µl entry clone, 2 µl destination vector, 4 µl 5XLR Clonase™ reaction buffer and TE buffer (volume to 16 µl). 4 µl BP Clonase™ enzyme was added to the components above. Then the same steps as the BP reaction were conducted to get the clones expressing LR products. The ISOLATE II Plasmid Mini Kit (Meridian Bioscience) was used to extract the plasmid DNA according to the manufacturer's manual.

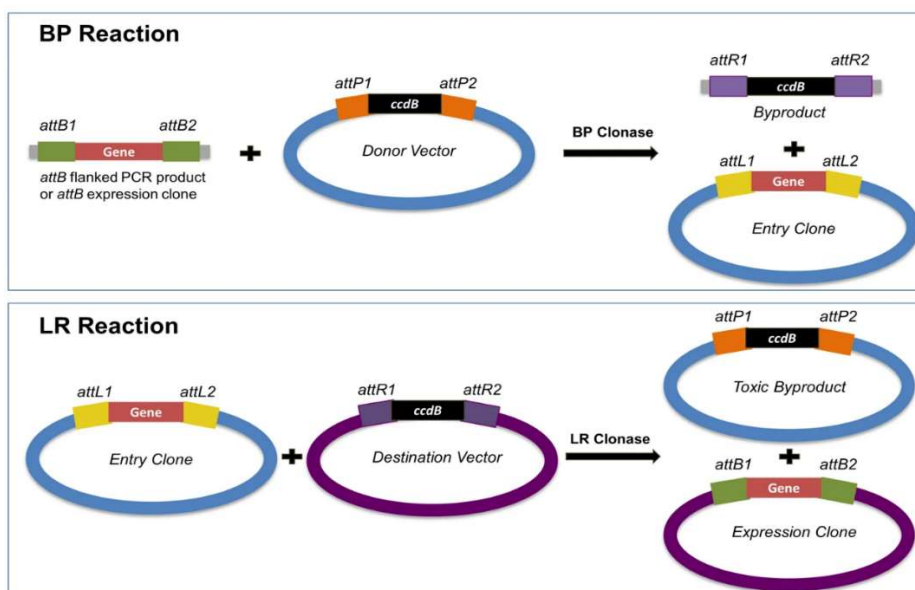


Figure 4. The Gateway system adopts phage integration into the BP and LR reactions (Figures adapted from <https://blog.addgene.org/plasmids-101-gateway-cloning>). The BP reaction takes place between the *attB* sites flanking the insert and the *attP* sites of the donor vector, generating the entry clone containing the DNA of interest flanked by *attL* sites. As a byproduct of the BP reaction, the *ccdB* gene is excised from the donor vector. The LR Reaction takes place between the *attL* sites of the generated entry clone and the *attR* sites of the destination vector. An expression clone with the DNA of interest flanked by *attB* sites is generated. As in the BP reaction, a DNA fragment containing the *ccdB* gene is excised from the destination vector. The BP reaction creates an *attL*-flanked entry clone. The LR reaction creates an expression clone with all of the components necessary for gene expression.

2.2. Transfection of Plasmid DNA into A549 Cells

Lipofectamine™ 2000 Transfection Reagent (Invitrogen) was used for the transfection of plasmid DNA into A549 cells according to the manufacturer's manual and neomycin (Thermofisher) was used for selection [30]. The cells were inoculated into 6-well plates at 37°C in the 5% CO₂ incubator for 48 h before transfection. DNA was diluted in 250 µl of Opti-MEM™ Reduced Serum Medium and 10 µl Lipofectamine 2000 was diluted in 250 µl medium, incubated for 5 min at RT. The diluted DNA was mixed with diluted Lipofectamine (total volume=500 µl) and incubated for 20 min at RT. 500µl of complexes was added to each well. The cells were incubated for 24h before testing for transgene expression. The medium was changed after 6 h. To get stable cell lines, the cells were passaged at a 1:10 dilution into a fresh medium 24 h after transfection. 28 µl neomycin was added to each well repeatedly the following day for selection and the medium was changed every 24 h.

2.3. Cell Culture and Treatments

The A549 cells (ECACC) were maintained in the medium DMEM/F12 (Thermofisher) with 10% FCS and 1% penicillin-streptomycin. Through all experiments, cell incubation was at 37°C in a 5% CO₂ atmosphere. Cells were incubated at 37°C with 5% CO₂ during experiments. Once the cells were about 80% confluent, they were detached using 0.05% Trypsin-EDTA and subcultured into clean T25cm³ culture flasks to preserve cell stock.

2.4. Experimental Design and Grouping of Proliferation Experiment

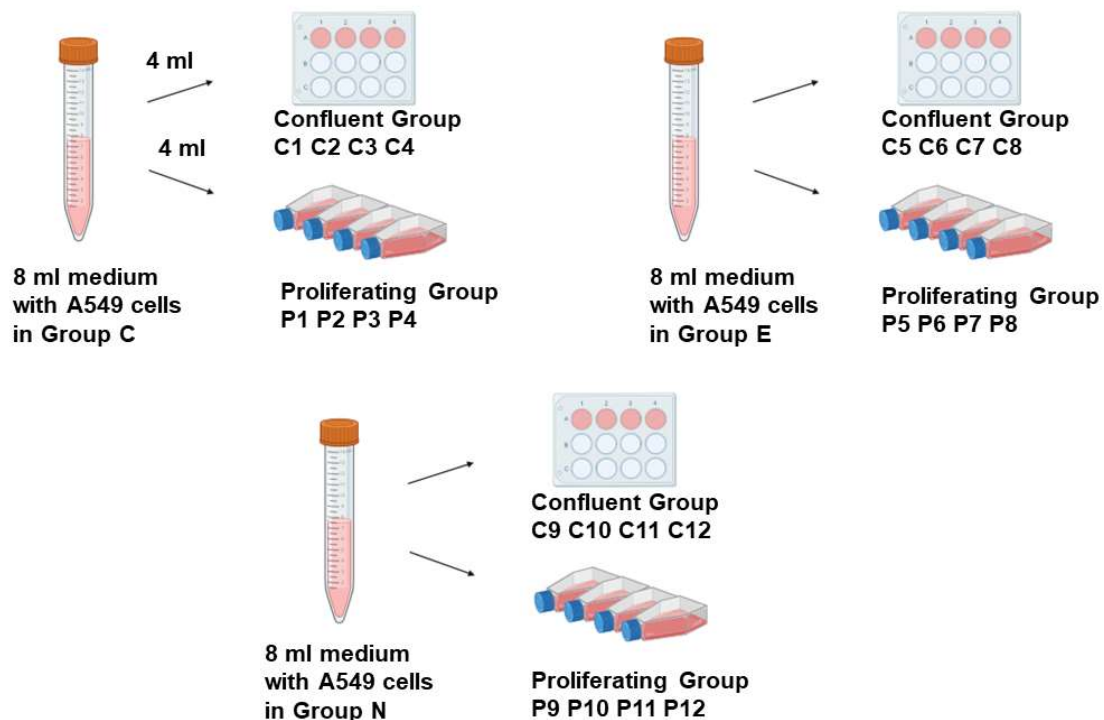


Figure 5. The grouping method of the three groups of A549 cells in the proliferation experiment. The cells in Group C, Group E and Group N are cells transfected with plasmids containing CEAA sequence, EAA sequence and NEAA sequence. The cells in each group were divided into four Confluent Groups and four Proliferating Groups respectively. The eight samples from Group C are numbered C1 to C4 and P1 to P4. The eight samples from Group E are numbered C5 to C8 and P5 to P8. The eight samples from Group N are numbered C9 to C12 and P9 to P12.

In the proliferation experiment, to make the cells take on two different growth states, half of the confluent cells were plated in a 12-well culture dish so they stayed confluent, while the other half were transferred to T25cm³ culture flasks so that they had more room to proliferate. The three successfully transfected cell lines incubated in a 6-well culture dish were trypsinised, and then 0.5 ml Trypsin and 0.5 ml fresh medium were added to each well. The cells were moved to 15 ml Falcon tubes for centrifugation (2000 rpm, 5 min). After pouring off the supernatant, fresh medium was added to 8 ml (the cell concentration is about 3X10⁵ cells/ ml) in the tube to break up the cell pellet and resuspend the cells. Then 4 ml medium containing cells were transferred to the 12-well culture dish as four Confluent Groups, 1 ml (about 3X10⁵ cells) in each well with 1 ml fresh medium supplemented in each well. The left 4 ml medium was transferred to four T25cm³ culture flasks as four Proliferating Groups, 1 ml in each flask with 6 ml fresh medium supplemented in each flask. The grouping of the cells is shown in Figure 5.

2.5. Immunofluorescent Staining to Detect V5 Tag

To determine whether the three transfected A549 cell lines can synthesise protein with the V5 tag successfully, immunofluorescent staining was used to visualise proteins with anti-V5 antibodies. Sterile 13 mm coverslips were put in a 6-well culture dish containing A549 cells suspension after selection, one in each well. The cells were washed with cold PBS on ice and then fixed to coverslips with 3.6% paraformaldehyde for 10 min on ice. After aspirating fixative and PBS wash, coverslips were incubated at -20°C for 20 min with methanol to enhance antibody permeability through the cell membrane, then washed in 2×PBS. To prevent unexpected binding, the cells were blocked in PBS with 0.1% Tween and 2% horse serum for 30 min. The anti-V5 primary antibody (Novex) and horseradish peroxidase-conjugated secondary antibodies (Pierce) were diluted with PBS before use according to the manufacturer's manual. The primary antibodies were added and specimens were incubated overnight at 6°C in the cold room. After incubation, PBS was used to wash the cells before incubation with the secondary antibody at RT for 1 h. PBS and dH₂O were used to wash the cells before staining.

Anti-V5 antibodies and nucleus were stained with FITC and DAPI respectively. Anti-fluorescence quench agent was dropped to the centre of the glass slides. The coverslips were placed on the glass slide and sealed with resin. Cell imaging was performed with laser scanning confocal microscopy.

2.6. Western Blot to Detect and Quantify Protein Synthesis

Western blot was used to detect the amount of protein each cell line produced. The Cellytic™ Cell Lysis Reagent (Thermofisher) was used for cell lysis according to the manufacturer's manual. After removing the medium in wells and flasks, the cells were washed twice with pre-cooled PBS on ice. The culture dishes and flasks were placed on ice for cracking for 30 min after adding the lysate. Then the cells were scraped off using clean scrapers. The cell fragments and lysate were moved into a 1.5 ml centrifuge tube and centrifuged at 12000 rpm for 5 min. After centrifugation, the supernatant containing protein was collected.

The loaded samples were quantified using Bradford protein Assay Kit (Thermofisher) and mixed with loading buffer in proportion of 1:1. Samples and 5µl EZ-Run™ Rec Protein Ladder (Thermofisher) were loaded onto the BIO-RAD Mini-PROTEAN TGX™ Precast Gels. The gel was resolved at 120 V for 45 min. Pierce™ Fast Western blot Kits, ECL™ Substrate (Themofisher) was used for blotting according to the manufacturer's manual. An electrotransfer stack was assembled consisting of two filter papers, ECL™ nitrocellulose membrane, the gel and pads, all pre-soaked in transfer buffer which was freshly prepared by dissolving 21.6 g glycine, 4.5 g Tris-Base and 300 ml methanol in 800 ml dH₂O. The stack was subjected to 4 mA for 40 min. The blot was washed in Fast Western Wash Buffer, then 10 ml Primary Antibody Working Dilution with anti-V5 primary antibody (1:1000) (one of the gels using anti-collagen IV antibody, Invitrogen) was added to incubate the blot overnight at

Figure 6 shows the sequences and sizes of the three LR products containing appropriate repeats of amino acids and with V5 tag and 6×Histidine at the terminal. The plasmid DNA containing the following three sequences was further used to be transfected into three groups of A549 cells respectively (named Group C, E and N) so that the three groups of cells could produce the corresponding three kinds of protein.

3.2. A549 Cells Successfully Expressing Artificial Protein with V5 Tag

3.2.1. Microscopy Images of Cells During Selection

Neomycin was used to select the A549 cells successfully transfected using Lipofectamine 2000. Neomycin resistance is mediated by the *neo* gene in the *pDEST40* plasmid. During selection, cells without the *pDEST40* plasmid died because they were unable to produce resistance products. In all the experimental groups, there were dead cells but some resistant colonies formed (Figure 7.1). Stable transfection can help implant the target gene into the cell chromosome, so the gene will not disappear with the cell passage [31]. During the experiment, these cells selected all had the gene expressing the artificial proteins inside.

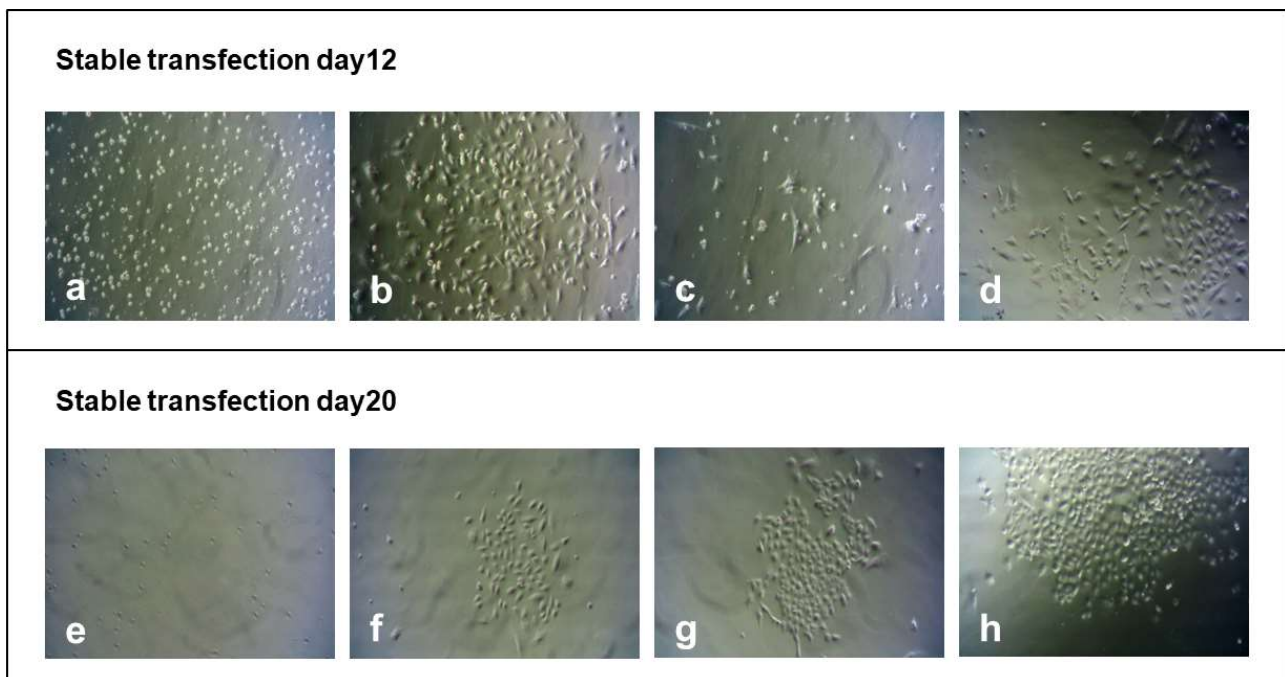


Figure 7.1 Microscopy images of the colonies in each group were captured during neomycin selection. On the 12th day after transfection, many dead cells can be seen in Control Group (a.), while in Group C (b.), Group E (c.) and Group N (d.), cell colonies formed, with dead cells floating. On the 20th day, it was a similar situation, dead cells were floating in the medium in Control Group (e.) and more colonies formed in Group C (f.), Group E (g.) and Group N (h.).

3.2.2. Immunofluorescence Composite Images

The nuclei were stained blue with DAPI and the anti-V5 antibody was stained green with FITC. In Figure 7.2, V5 reporters from all three groups were visualized by immunofluorescence, which indicates the target gene can be expressed in cells. So these three cell lines are successfully created and ready for further experiments. As can be seen in the images, the green fluorescence area was the largest in Group C (Figure 7.2b), which means the cells in Group C made a larger amount of artificial proteins than the other two groups. However, the reasons for this phenomenon are not clear.

It is worth noting that the green fluorescent proteins are scattered around the cell in dots. This is probably because the higher expressed proteins tend to clump together. It is also possible that the protein structure with the same repetitive amino acid units is predisposed to aggregate.

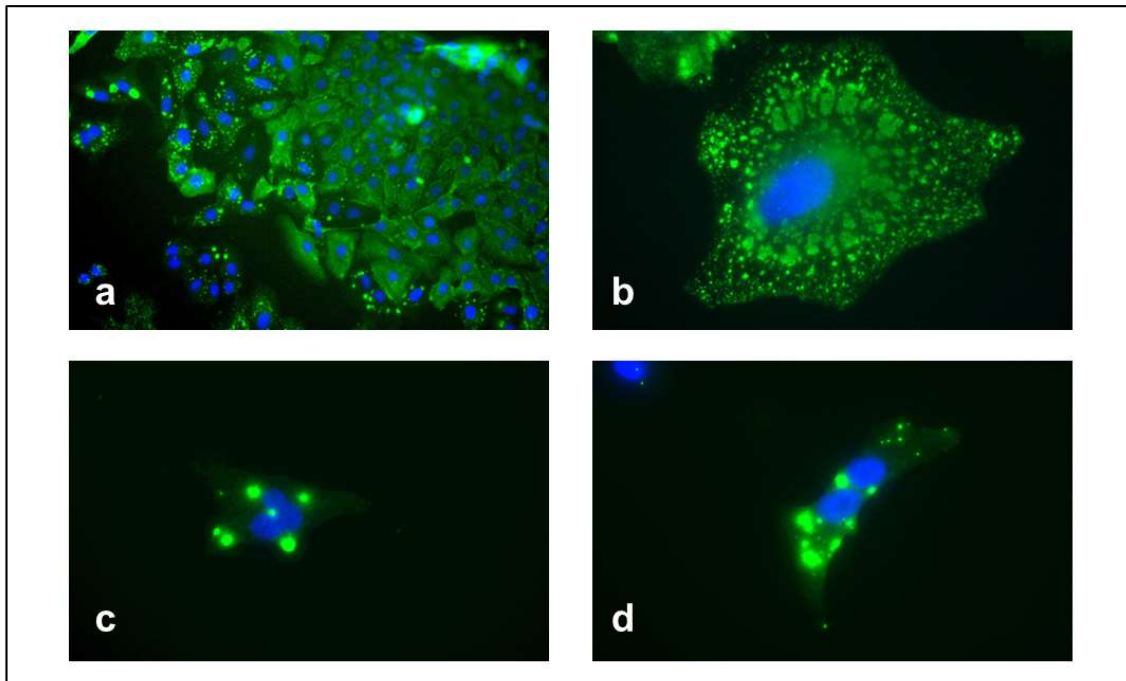


Figure 7.2 Composite images of cells in each group from immunofluorescence staining after selection using neomycin. The images were captured using laser scanning confocal microscopy. The nuclei were stained blue with DAPI and the anti-V5 tag antibody was stained green with FITC. (a.) Group C at low power. (b.) Group C at high power. (c.) Group E at high power. (d.) Group N at high power. In Group E and Group N, cells were not easy to find at low power.

3.2.3. Microscopy Images of Cells in Two Different Growth States

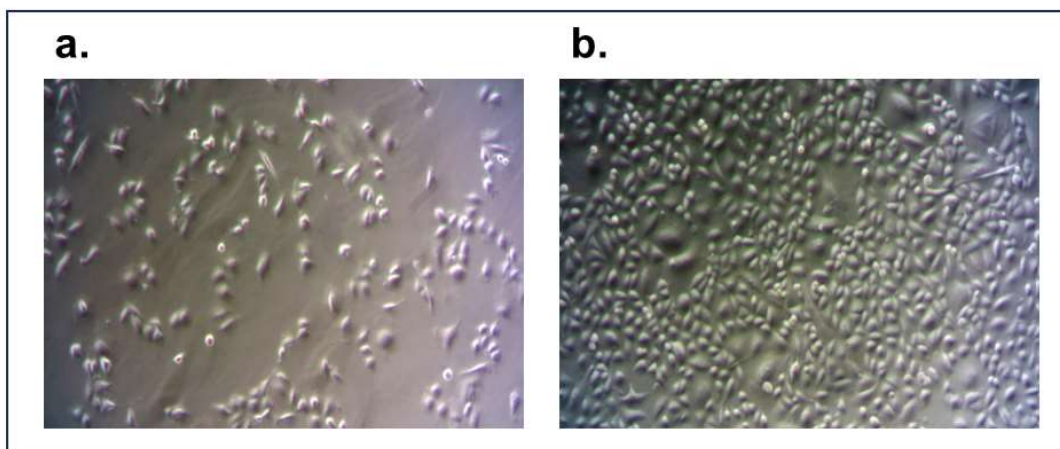


Figure 7.3 Microscopy images of the cells expressing Protein Cs in Proliferating Group and Confluent Group before the protein extraction in the proliferation experiment. The images were taken three days after setting the groups. (a.) Cells in Proliferating Group. (b.) Cells in Confluent Group.

1 ml medium (about 3×10^5 cells) was transferred respectively to the T25cm³ flask as Proliferating Group and each well of the 12-well culture dish as Confluent Group. As is seen from Figure 7.3, after three days of incubation by only changing the medium, the colonies in Confluent Group were already densely confluent while the cells in Proliferating Group still had a lot of room for growth, which provides the conditions needed for quantifying synthesized protein by Western blot and contrasting their differences.

3.3. Differences in the Expression of Artificial Proteins in Three Groups

3.3.1. Band Analysis of Western blot

Samples containing Protein C_S from the eight groups were loaded in turn as is shown in Figure 8.1. The location and size of Protein C_S and protein ladder polypeptide standards are indicated. Through contrast with the marker, the bottom band is determined as the band of the artificial Protein C_S that needs to be further analyzed. However, there are also bands of high molecular weight in the blot, presumably due to the aggregation of artificial proteins (referring to the immunofluorescence images shown in Figure 7.2). The same method was used in the experiments of the other two groups. Corresponding bands of artificial proteins were cropped for quantitative analysis.

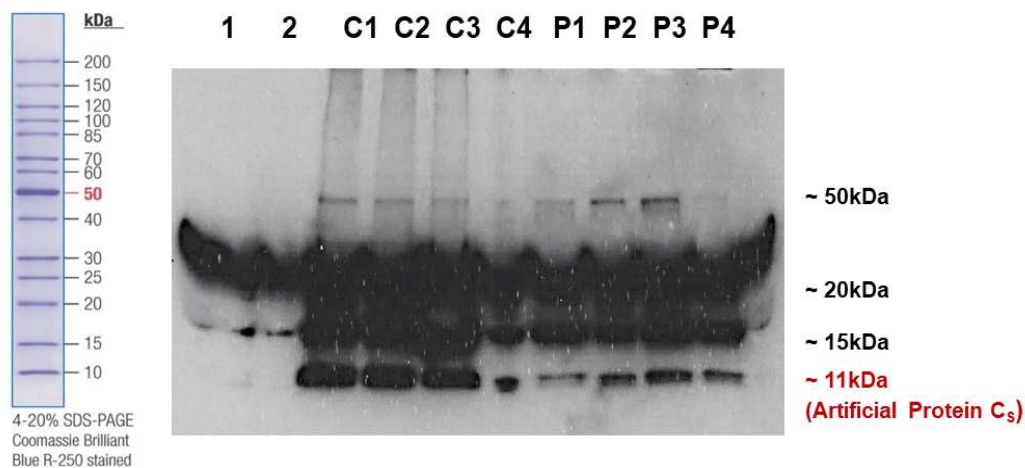


Figure 8.1 Blot analysis of Protein C_S in Confluent Group and Proliferating Group. There were some technical issues with this gel. The gel was resolved at 120 V for 45 min. The exposure time is three seconds. The blot was probed with anti-V5 antibody. After stripping anti-V5 antibodies, the blot was then probed with anti-beta actin antibodies. The membrane was later stained with Coomassie Blue dye to see the ladder. The location and size (kDa) of Protein C_S and protein ladder polypeptide standards are indicated. Lane 1 is loaded with the ladder and Lane 2 is blank. Eight groups of samples were loaded in turn in the remaining lanes as is shown above.

3.3.2. Band Intensity Quantification

Four groups of experiments were done to determine the effects of proliferation on the expression levels of four kinds of proteins (Protein C_S, Protein E_L, Protein N_S and collagen IV) in different proportions of amino acid class. Collagen IV is the natural synthetic protein made in the human body, which was chosen because it also has a relatively large proportion of CEAs (about 60% [32]) in its sequence. Beta-actin acts as the loading control in each experimental group. The bands of the Protein C_S, collagen IV and their beta-actin loading control are shown in Figure 8.2a and Figure 8.2b. It can

be seen that the band intensity of beta-actin in each blot was relatively consistent, indicating that proliferation did not have a great impact on the expression of all common proteins, so the experimental results are reliable to some extent. The relative band intensity of collagen IV produced by cells in Group C was measured and compared with that of Protein C_s in Figure 8.2c.

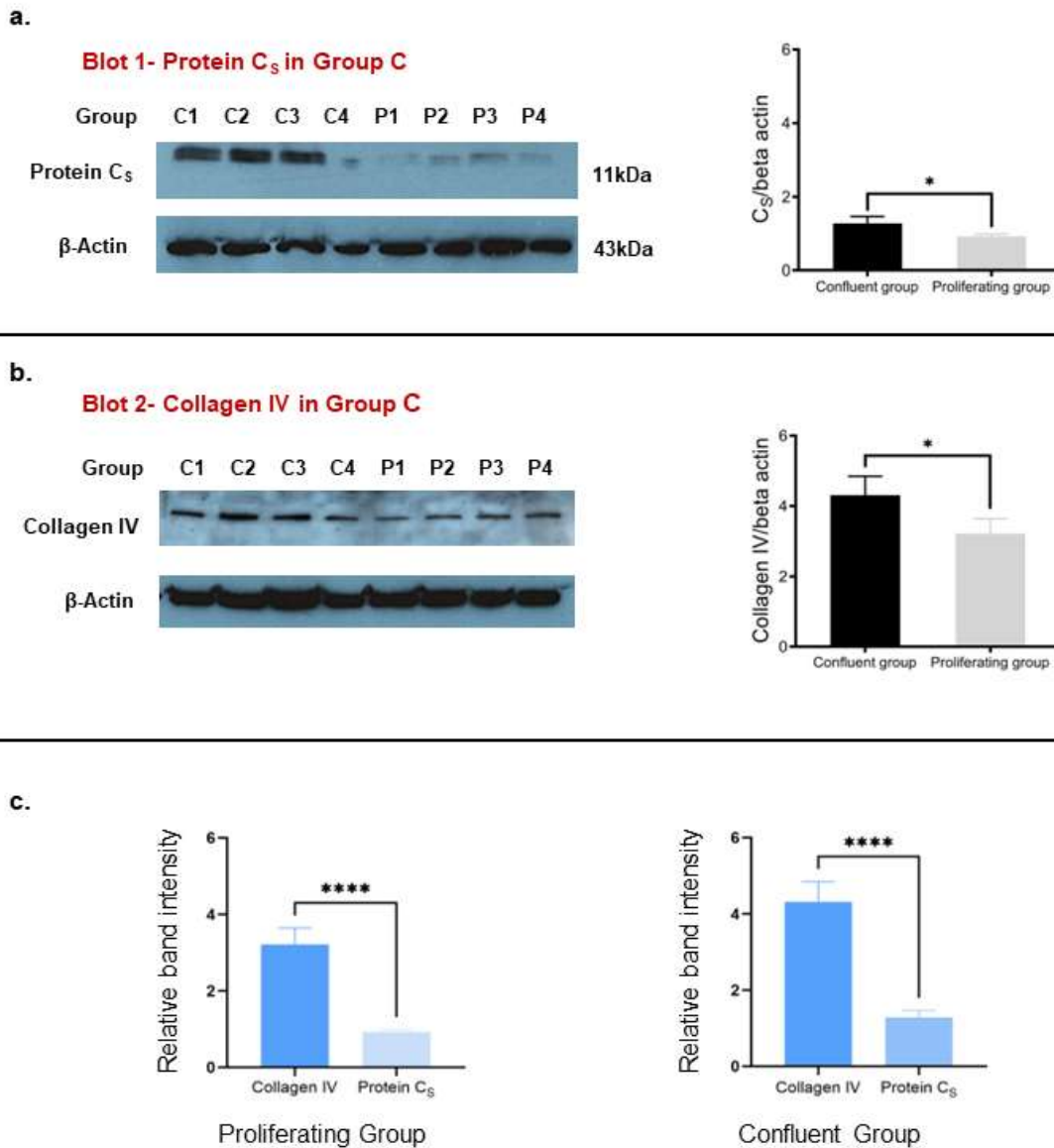


Figure 8.2 Comparison of the relative band intensity of Protein C_s and Collagen IV between Confluent Group and Proliferating Group. The obtained grayscale value of the bands (representing the band intensity) in Group C was normalised by dividing by the grayscale value of the corresponding loading control group. The average value of the four same treatment groups on each blot was taken for the bar chart, significance analysis and t-test using GraphPad Prism. (a.) Bands and relative band intensity of Protein C_s ($SD_{confluent1}=0.19$, $SD_{proliferating1}=0.06$). (b.) Bands and relative band intensity of collagen IV ($SD_{confluent2}=0.54$, $SD_{proliferating2}=0.42$). (c.) Comparison of the average band intensity of Protein C_s and collagen IV. The band of collagen IV (b.) was too weak to scan so it was photographed, which may lead to errors. The differences were considered statistically significant if $P<0.05$ (*). Figure 8.2c indicates that there is a very significant difference ($P<0.0001$) between the expression of Protein C_s and collagen IV.

As is shown in Figure 8.2a and Figure 8.2b, a statistically significant difference ($P < 0.05$ (*)) can be seen between Confluent Group and Proliferating Group in both two charts, which indicates that the expression of Protein C_S and collagen IV in Proliferating Group is decreased compared with Confluent Group. The proportion of CEAA in total amino acids of collagen IV (about 60%) is smaller than that of Protein C_S (about 95%). Through comparing the relative band intensity of Protein C_S with that of collagen IV (Figure 8.2c), collagen IV is found to be significantly more expressed ($P < 0.0001$ (****)) than Protein C_S in the same treatment group. So the hypothesis has been proven to some extent that the protein composed of a greater proportion of CEAA tends to be lower in its expression.

In order to further study the effect of cell proliferation on the expression of proteins with different proportions of amino acid types, the same experiment was done on cells from Group E and Group N. Surprisingly, in Figure 8.3, the expression of both Protein E_L and Protein N_S is significantly reduced ($P < 0.0001$ (****)) when proliferating compared with the confluent state, even more significant than the reduction of Protein C_S expression ($P < 0.05$ (*)). A comparison of the effects of proliferation on the expression of the four kinds of proteins was made in Figure 8.4. The results indicate that proliferation has a greater effect on Protein E_L, followed by Protein N_S, Protein C_S and collagen IV. Therefore, the hypothesis that proteins with extreme proportions of CEAA tend to reduce expression and are most affected by proliferation compared with proteins with other proportions of amino acid class cannot be proved completely in this experiment.

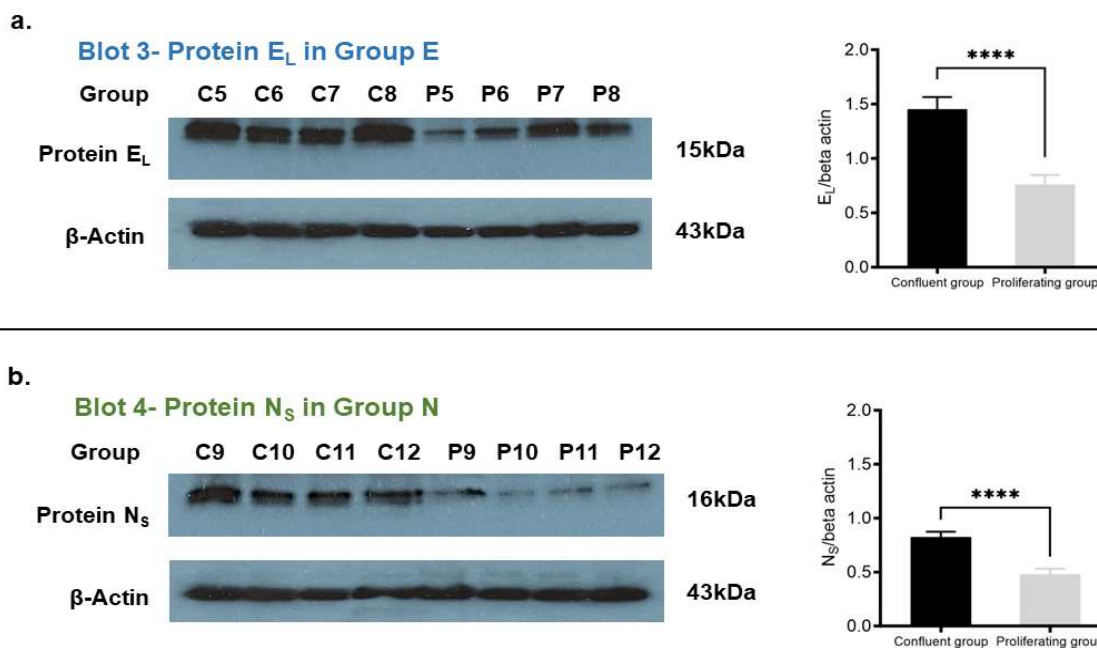


Figure 8.3 Comparison of relative band intensity of Protein EL and Protein N_S between Confluent Group and Proliferating Group. The obtained grayscale values of the bands (representing the band intensity) in Group E and Group N were normalised by dividing by the grayscale value of the corresponding loading control group. The average value of the four same treatment groups on each blot was taken for the bar chart, significance analysis and t-test using GraphPad Prism. (a.) Bands and relative band intensity of Protein EL ($SD_{confluent3}=0.11$, $SD_{proliferating3}=0.09$). (b.) Bands and relative band intensity of Protein N_S ($SD_{confluent4}=0.05$, $SD_{proliferating4}=0.05$), Both P values from the unpaired t-test are less than 0.0001 (**), indicating that there is a very significant difference between Confluent Group and Proliferating Group.

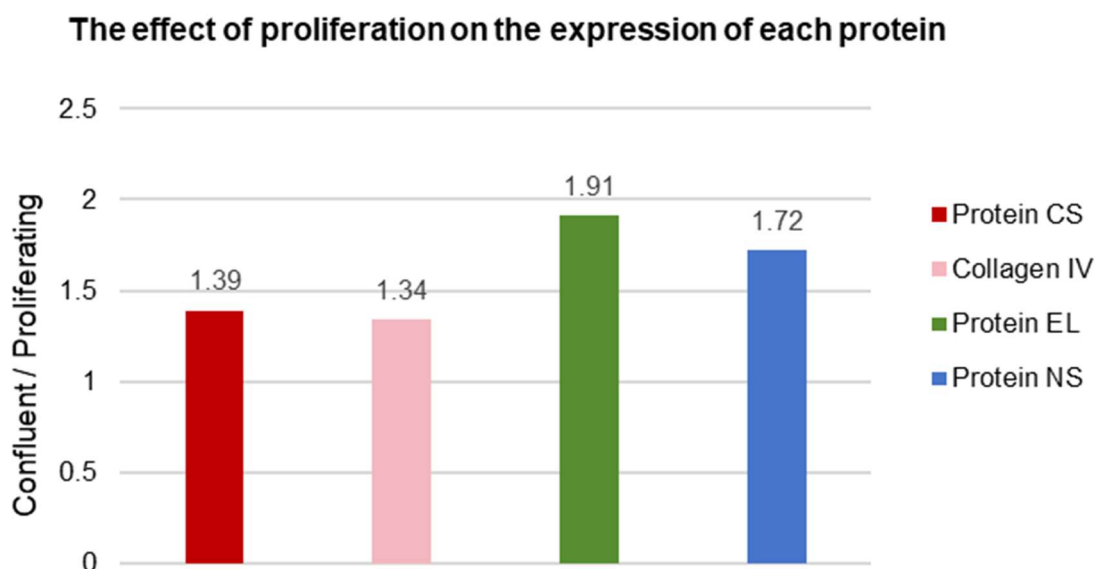


Figure 8.4 Comparison of the effect of proliferation on the expression of Protein Cs, collagen IV, Protein EL and Protein Ns. The bar chart was created using Excel. The value of the ordinate represents the average relative band intensity of each protein in Confluent Group divided by the average relative band intensity of each protein in Proliferating Group. The quotient was used to quantify the effect of proliferation on protein expression. The higher the value, the greater the influence of proliferation on the expression of the corresponding protein.

4. DISCUSSION

4.1. Results Analysis

According to the microscopy images above (Figure 7.1, Figure 7.2), A549 cells expressing artificial protein composed of extreme proportions of amino acids were successfully created within a relatively short period. Although the results show that proliferation has a greater effect on proteins with extreme EAA and NEAA ratios than on proteins with extreme CEAA ratios, which is contrary to our hypothesis, this method has proven to be reasonably straightforward and efficient.

In Figure 8.2a and 8.2b, a statistically significant difference ($P < 0.05$ *) can be seen between Confluent Group and Proliferating Group, indicating the expression of Protein Cs and collagen IV in Proliferating Group is decreased compared with Confluent Group. This shows that rapidly proliferating A549 cells reduce the expression of both Protein Cs and collagen IV compared with confluent A549 cells. The proportion of CEAA in total amino acids of collagen IV (about 60%) is smaller than that of Protein Cs (about 95%). When it comes to which of the above two proteins is more affected by proliferation, it can be seen in Figure 8.4 that the expression of Protein Cs is more affected by proliferation than collagen IV, which seems to be consistent with our hypothesis that proteins containing extremely more CEAA are more susceptible to reduced expression when affected by proliferation.

However, in Figure 8.3a and Figure 8.3b, a bigger statistically significant difference ($P < 0.0001$ (****)) compared with Figure 8.2a and 8.2b can be seen between Confluent Group and Proliferating Group, which means the expression of Protein EL and Protein NS reduced more and is more affected by proliferation. As can be seen from Figure 8.4, the expression of Protein EL is most affected by proliferation, followed by Protein NS, Protein Cs and collagen IV. However, according to our

hypothesis, it is expected that the expression of Protein C_S is most affected by proliferation, followed by collagen IV, Protein E_L or Protein N_S. Moreover, Protein N_S expression should not have changed significantly even with the rapid proliferation of cells. So the results do not exactly agree with the hypothesis.

In Figure 8.2c, by comparing the relative band intensity of Protein C_S with that of collagen IV, it can be seen that collagen IV is found to be significantly more expressed ($P < 0.0001$ ****) than Protein C_S in the same treatment group. The proportion of CEAA in total amino acids of collagen IV (about 60%) is smaller than that of Protein C_S (about 95%). So the hypothesis has been proven to some extent that the protein composed of a greater proportion of CEAA tends to be lower in its expression.

Based on the image and statistical analysis above, it seems that only half of the hypothesis has been confirmed through experiments that proteins with extreme proportions of CEAA tend to reduce expression. Rapid proliferation causes a reduction in the expression of proteins with extreme ratios of all three types of amino acids (Protein C_S, E_L and N_S) (Figure 8.2, Figure 8.3). However, it cannot be proven that proliferation had a greater effect on the expression of CEAA-rich proteins than other proteins, and the results are even contrary to our hypothesis.

The reason why our research results are inconsistent with the bioinformatic analysis results might be the particularity of artificial proteins, contingency of a single experiment with four replicates or inaccurate manipulation. In terms of the structure of the artificial proteins, Protein C_S, Protein E_L and Protein N_S are all composed of amino acid repeats with extreme proportions (over 95%) of only one amino acid class. In Figure 7.2, proteins with V5 tag are scattered around the cell in dots, which indicates that the structure of the artificial proteins has a certain influence on their expression. It can also be seen that cells in Group C made a larger amount of artificial proteins than the other two groups. These phenomena suggest that the repetitive structure of the artificial proteins has inherent specificity, which is different from proteins naturally synthesized by cells. Moreover, our hypothesis is based on the findings in Figure 2.1. However, the proteins in previous research based on the bioinformatic dataset do not contain an extreme proportion of amino acid class but are naturally produced in human tissues or cells with relatively moderate amino acid ratios (25%-40% each class). So extreme proportions are also a possible explanation for why the results differ from the hypothesis.

4.2. Limitations of the Research

This research has some limitations in terms of experimental methods and materials. Firstly, whether the repetitive sequences and the extreme amino acid composition of those artificial proteins themselves affect protein expression is unknown. The target proteins detected in our research are artificial proteins with extreme amino acid ratios instead of proteins naturally produced in human tissues or cells with relatively moderate amino acid ratios in the previous research [12]. Only three artificial proteins with extreme amino acid ratios are unable to provide the continuous and moderate data of amino acid ratios that bioinformatic datasets do. Additionally, the sequencing results for the three artificial proteins seem that they are quite close in size (11kDa, 15kDa, 16kDa), so it is difficult for them to separate on one gel and the amount of two proteins cannot be measured on one gel at the same time. Some variables cannot be fully controlled.

Moreover, after transfection, the cells in each well from the same group (Group C, E and N) were pooled together instead of isolated to individual colonies and then regrouped into Proliferating Group and Confluent Group. This means only one genetically identical cell line was dealt with, but each colony might have higher or lower levels of protein expression, so this pooled approach might have an effect on the results.

Finally, Western blot analysis was used for quantification in our research, which is mainly applied to qualitative and semi-quantitative proteins, so it cannot show the absolute protein amount, but only the relative change of protein abundance [33]. The band intensity is affected by factors such as protein

degradation, migration, and exposure time, so it is difficult to accurately compare the protein expressions.

4.3. Future Work

It would be beneficial to attempt repeats of the experiments using different cell lines and artificial proteins, for both the number and the object (A549 cell line) of experiments are single in this research, so the results may be accidental and a firm conclusion cannot be drawn. Further experiments could be designed where equal numbers of cells with different plasmids are mixed and cultured together and then the synthesised proteins were extracted for analysis.

To gain a more comprehensive insight into the effect of amino acid supply on protein synthesis, further studies could be designed, such as the effect of starvation, stress and drugs on the expression of proteins with extreme proportions of amino acid class. For example, whether starvation uses up EAAs is worth testing. Human starvation can be simulated by diluting the medium or reducing feeding, and environmental stress and drugs can be simulated by adding certain biochemical inhibitors. In theory, proliferation and starvation are expected to have less effect on proteins with a higher proportion of NEAAs than on proteins with a higher proportion of CEAAAs and EAAs respectively. The hypothesis is based on the fact shown in Figure 2.2 that CEAA-rich proteins are components of skin and nails and are involved in growth. EAA-rich proteins are the first to detect malnourishment and hunger. In an amino acid malnourishment situation, EAA-rich proteins won't be easily produced [12].

4.4. Conclusion

In our experiment, the molecular tools, which are those A549 cells expressing artificial protein composed of extreme proportions of one amino acid class, were successfully created. By analyzing the results integrated with the bioinformatic data, two facts have been confirmed through our research: (1) Proteins with extreme proportions of CEAAAs tend to reduce expression. (2) Rapid proliferation causes a reduction in the expression of certain proteins, probably with extreme amino acid proportions. However, what is the law and mechanism of the effect of proliferation on proteins with different amino acid ratios needs further study. It would be beneficial to attempt repeats of the experiments to avoid errors and further studies should be designed. Controlling the supply of certain amino acids and developing drugs targeting CEAA or EAA-rich proteins will provide a solution for cancer and problems in growth and development.

REFERENCES

- [1] Crick F. Central Dogma of Molecular Biology. *Nature* [Internet]. 1970 Aug;227(5258):561–3.
- [2] Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*. 2016;165(3):535-50.
- [3] Maier T, Guell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett*. 2009;583(24):3966-73.
- [4] Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13(4):227-32.
- [5] Kanaya S, Yamada Y, Kudo Y, Ikemura T. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and specific diversity of codon usage based on multivariate analysis. *Gene*. 1999;238(1):143-55.
- [6] Bulmer M. Coevolution of codon usage and transfer RNA abundance. *Nature*. 1987;325(6106):728-30.
- [7] Rattan SIS. Chapter 10 Protein synthesis and regulation in eukaryotes. *Cell Chemistry and Physiology: Part 1*. 1995;247–63.
- [8] Hernandez-Montes G, Diaz-Mejia JJ, Perez-Rueda E, Segovia L. The hidden universal distribution of amino acid biosynthetic networks: a genomic perspective on their origins and evolution. *Genome Biol*. 2008;9(6): R95.

- [9] Trolle J, McBee RM, Kaufman A, Pingley S, Berger H, German S, et al. Resurrecting essential amino acid biosynthesis in a mammalian cell.
- [10] Laidlaw SA, Kopple JD. 1987. Newer concepts of the indispensable amino acids. *Am J Clin Nutr* 46: 593–605.
- [11] Reeds PJ. Dispensable and indispensable amino acids for humans. *J Nutr.* 2000;130(7): 1835S40S.
- [12] Thompson R, Pickard B. The amino acid composition of a protein influences its expression. *bioRxiv* (Cold Spring Harbor Laboratory), 2024 Mar 29, in press.
- [13] Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell lines. *Proteomics.* 2015;15(18): 3163-8.
- [14] Balbin OA, Prensner JR, Sahu A, Yocum A, Shankar S, Malik R, et al. Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nat Commun.* 2013;4:2617.
- [15] Geiger T, Wehner A, Schaab C, Cox J, Mann M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics.* 2012;11(3): M111 014050.
- [16] Payne SH, Loomis WF. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryot Cell.* 2006;5(2):272-6.
- [17] Prentice AM. Starvation in humans: Evolutionary background and contemporary implications. *Mechanisms of Ageing and Development.* 2005 Sep;126(9):976-81.
- [18] Ghosh S, Suri D, Uauy R: Assessment of protein adequacy in developing countries: quality matters. *Br J Nutr* 2012;108(suppl 2):S77-S87.
- [19] Semba RD, Shardell M, Sakr Ashour FA, Moaddel R, Trehan I, Maleta KM, et al: Child stunting is associated with low circulating essential amino acids. *EBioMedicine* 2016;6:246-252.
- [20] Semba RD, Trehan I, Gonzalez-Freire M, Kraemer K, Moaddel R, Ordiz MI, et al: The potential role of essential amino acids and the mechanistic target of rapamycin complex 1 (mTORC1) pathway in the pathogenesis of child stunting. *Adv Nutr* 2016, in press.
- [21] Hendrickson NR, Davison J, Glass NA, Wilson ES, Miller A, Leary S, et al. Conditionally Essential Amino Acid Supplementation Reduces Postoperative Complications and Muscle Wasting After Fracture Fixation: A Randomized Controlled Trial. *J Bone Joint Surg Am.* 2022;104(9):759-66.
- [22] Barbul A, Lazarou SA, Efron DT, Wasserkrug HL, Efron G. Arginine enhances wound healing and lymphocyte immune responses in humans. *Surgery.* 1990;108(2):331-6; discussion 6-7.
- [23] Desneves KJ, Todorovic BE, Cassar A, Crowe TC. Treatment with supplementary arginine, vitamin C and zinc in patients with pressure ulcers: a randomised controlled trial. *Clin Nutr.* 2005;24(6):979-87.
- [24] Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer.* 2018;18(11):696- 705.
- [25] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-74.
- [26] Lieu EL, Nguyen T, Rhyne S, Kim J. Amino acids in cancer. *Exp Mol Med.* 2020;52(1):15-30.
- [27] Zhang H, Wang Y, Li J, Chen H, He X, Zhang H, et al. Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun.* 2018;9(1):4124.
- [28] Kang JS. Dietary restriction of amino acids for Cancer therapy. *Nutrition & Metabolism (Internet).* 2020 Mar 14 (cited 2021 Jul 18);17.
- [29] Katzen F. Gateway® recombinational cloning: a biological operating system. *Expert Opinion on Drug Discovery.* 2007 Apr;2(4):571-89.
- [30] Zhang A, Yan X, Li H, Gu Z, Zhang P, Zhang H, et al. TMEM16A protein attenuates lipopolysaccharide-mediated inflammatory response of human lung epithelial cell line A549. *Experimental Lung Research.* 2014 May 2;40(5):237-50.
- [31] Stepanenko AA, Heng HH. Transient and stable vector transfection: Pitfalls, off-target effects, artefacts. *Mutation Research/Reviews in Mutation Research.* 2017 Jul;773:91-103.
- [32] Zhou J, Jens Michael Hertz, Leinonen A, K Tryggvason. Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *Journal of Biological Chemistry.* 1992 Jun 1;267(18):12475-81.
- [33] Pillai-Kastoori L, Schutz-Geschwender AR, Harford JA. A systematic approach to quantitative Western blot analysis. *Analytical Biochemistry.* 2020 Mar;593(113608): 113608.