

Chapter 7

Allergen Gene Cloning

Xue-Ting Liu and Ailin Tao

Abstract For gene cloning of known allergens, primers are designed according to individual allergen genes and selective amplification by PCR, however, for cloning of an unknown or unstudied species, the design of the degenerate primers is crucial to its successful amplification by PCR. In order to achieve high expression efficiency of the desired allergen gene in a prokaryotic system or other protein expression system, it is necessary to optimize the allergen gene codon. After codon optimization of the allergen gene sequences, protecting bases and an affinity purification tag are added to the 3' and 5' end restriction loci and then sent out for sequence synthesis. The prokaryotic protein expression system is the most commonly used expression system and is also the most affordable. During the prokaryotic protein expression process, high yields of the target protein can be obtained by optimizing the host strain and induction conditions including inducing temperature, inducing agent, inducing time, and IPTG concentration. In the *Escherichia coli* system, most recombinant proteins are expressed in inclusion bodies. The denaturation and refolding of inclusion bodies are very important to obtain allergens that have biological activities. Finally, recombinant allergens can be purified using chromatography, isoelectric point precipitation, and salt fractionation.

Keywords Degenerate primer • Codon optimization • Prokaryotic expression • Inclusion body • Refolding • Purification of allergen

X.-T. Liu · A. Tao (✉)

Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, The State Key Clinical Specialty in Allergy, The State Key Laboratory of Respiratory Disease, The Second Affiliated Hospital of Guangzhou Medical University, 250# Changgang Road East, Guangzhou, China
e-mail: Aerobiologiao@163.com

X.-T. Liu
e-mail: jane102514@163.com

7.1 Introduction

Allergen gene cloning is the premise and foundation for the creation of recombinant allergens (Tao and He 2005). Primers are designed according to known individual allergen genes and selective amplification by PCR is the method most frequently used for allergen gene cloning. However, due to the differences between species DNA sequences and in those species less studied or with lower homologies, the PCR method of gene cloning is not ideal and is therefore limited in its application. Sequence homology exists in different species of allergens and lays a solid foundation for the design of common primers and subsequent gene clones. Cloning of an allergen for a species that has not yet been studied can be accomplished by designing degenerate primers (Linhart and Shamir 2005).

7.2 Cloning of Allergen Genes and Designing of Degenerate Primers

Due to differences in nucleotide sequences among species, the gene sequence cannot always be directly applied when cloning a homologous gene, especially in an unknown or unstudied species. In this case, the amino acid sequences that are conserved between the different species can be used. On the basis of sequence homology between a large number of allergens in a biological information database and by designing degenerate primers, we can perform selective PCR amplification on the cDNA pool using the Touchdown program. At the same time, the creation of degenerate primers can be further intensified and homologous gene allergens can be cloned by utilizing RACE technology to obtain full-length cDNA. The design of the degenerate primer is crucial to its successful amplification by PCR.

7.2.1 *The Search for Homologous Amino Acid Sequences*

The search begins with entering the common name of the gene to be studied into the NCBI database search text box and then searching the “Protein” database to obtain the amino acid sequence of the desired gene, which is not limited to any particular species. If the nucleotide sequence of the target gene is already known, another method can be employed, namely, the translation of the nucleotide sequence into an amino acid sequence that is then used to run a BLAST program in the NCBI database to obtain the homologous amino acid sequences.

7.2.2 Sequence Alignment to Determine Representative Sequences of a Homologous Gene

The Clustal W multiple sequence alignment program and the Mega3.0 software program can be used to bring the searched amino acid sequence into multiple sequence alignment (Chenna et al. 2003; Kumar et al. 2004). When multiple genes are involved, there may be more differences in the amino acid sequences and more clusters may be found; in such cases, a plurality of the representative sequences should be selected to carry out the next step.

7.2.3 Search for a Homologous Nucleotide Sequence

Use a representative nucleotide sequence to search for a homologous nucleotide sequence(s). This search process is the same as that using a homologous amino acid sequence and uses the BLAST program in the NCBI database to obtain a homologous nucleotide sequence (She et al. 2009).

7.2.4 Alignment of a Homologous Nucleotide Sequence and Identification of Conserved Regions

The Clustal W multiple sequence alignment procedures and the Mega3.0 software program, as well as other programs, can be used to align the obtained homologous nucleotide sequences and identify a relatively conserved nucleotide region to design the primers. If the conserved region is weak, primers can be created by designing degenerate bases at the changeable sites. When cloned by reverse transcription PCR, the desired target gene fragment can be obtained through combining universal primers with the upstream primer created from the conserved region. The full-length gene can be obtained through designing primers with specific sequences from gene fragments found using the 5'-RACE technique.

7.2.5 Degenerate Primer PCR

Although degenerate primers can be designed, they do not always work well, so it is very important to optimally adjust the reaction conditions. For example, try varying possible annealing temperatures or using the Gradient PCR or Touchdown PCR protocols, which should generally be successful in obtaining the desired target gene (Linhart and Shamir 2007; Souvenir et al. 2007).

(1) **Touchdown PCR**

The so-called annealing temperature is the temperature when half of the double-stranded DNA denatures. That is, at this temperature, half of the DNA may still exist in a double-stranded state. For a primer, the annealing temperature refers to the temperature at which the primer can still be combined with the template. Under these denaturing conditions, only the primer that is perfectly matched with the template can amplify, and, in spite of low amplification efficiency, the amount of primer is limited so as to avoid nonspecific amplification. This method is quite effective when cloning genes in a DNA or cDNA pool. Touchdown PCR is specifically designed to use this principle to enhance PCR amplification.

For the touchdown protocol, if the annealing temperature is normally T_m °C, then the annealing temperature should be set to $T_m + 5$ °C for a few cycles to specifically amplify the desired gene, with the main loop amplification finally set at T_m °C. Notably, the above-mentioned annealing temperatures are obtained using the OMEGA software. If you obtain annealing temperatures using a different software program, then the above-recommended annealing temperature adjustment may not be optimal.

Generally, the shorter the annealing time is, the stronger the specificity of the amplification. In order to enhance the specificity of amplification, the Touchdown program gradually reduces the annealing time so as to improve the specificity of primers. For example, the beginning annealing time is set to 45 s, then the annealing time is shortened by 1–2 s after every cycle, with the main cycling set for 30 s. This procedure is more efficient for the amplification of degenerate primers.

(2) **Gradient PCR**

When using degenerate primers for amplification, there can be uncertainty as to whether the primer is fully matched with the template. Under normal circumstances, the optimal annealing temperature of the reaction can be determined through trial and error. If using a gradient PCR system, you can accomplish this step all at once in one reaction plate. When setting the annealing temperature, select the “Gradient” program and, depending on the range of the various primer annealing temperatures to be tested, different rows (or columns) in the reaction plate can be programmed to achieve different annealing temperatures. The products of the different amplification reactions can be evaluated on an agarose gel. Based on these results, the optimum annealing temperature of the reaction can be determined, as well as whether the degenerate primers accept the original degenerate bases and whether the original gene primer sequences (using the RACE method) need to be confirmed after the gene is cloned.

(3) **RACE**

Rapid amplification of cDNA ends (RACE) is a PCR-based method for the rapid amplification of cDNA with a low number of transcripts at the 5' and 3' ends (Frohman 1993). Obtaining a complete cDNA sequence is essential for gene studies, protein expression, and gene function. Complete cDNA sequences can

be acquired by library screening and end cloning techniques. The classic RACE technology was developed by Frohman et al. (Frohman et al. 1988), mainly using RT-PCR technology and including unilateral PCR and anchored PCR in order to get the complete cDNA 5' and 3' terminal using a known cDNA sequence. Some improvements to the traditional RACE technology are mainly in primer design and improvements in RT-PCR technology are as follows: (a) the use of locking primers to construct the first-strand cDNA, in which in the oligo (dT) primers of the 3' end introduce two degenerate nucleotides (5'-Oligo(dT)₁₆₋₃₀MN-3', M=A/G/C; N=A/G/C/T), making the primer starting point at the poly(A) tail; (b) tailing at the 5' end with poly(C) rather than poly(A); (c) using the RNase H-Mononey Murine Leukemia Virus (MMLV) reverse transcriptase and thermophilic DNA polymerase, which can effectively reversely transcribe mRNA and amplify DNA at high temperatures (60–70 °C) to eliminate the influence of the mRNA secondary structure acting on reverse transcription due to the high CC content at the 5' end; and (d) the adoption of the hot-start PCR technique and the Touchdown PCR protocol to improve the specificity of the PCR reactions.

Compared with library screening, RACE has several advantages: (a) the methods are implemented using PCR technology, do not need the establishment of a cDNA library, and useful information can be obtained in a very short period of time; (b) the cost and experimental time are reduced; and (c) as long as the primer is designed correctly, a full-length coding region of the interested gene(s) can be obtained.

7.3 Codon Optimization

Whether the genes themselves match with the carrier and the host system is a very critical problem in gene protein expression. Codons are different in eukaryotic and prokaryotic cells, therefore, when we express eukaryotic genes in a prokaryotic system, some of the resulting eukaryotic genes may contain a rare codon for prokaryotic cells, resulting in low expression levels and low expression efficiency (Elena et al. 2014).

In order to achieve high expression efficiency of the desired allergen gene in a prokaryotic system or other protein expression system, it is necessary to optimize the allergen gene codon. This includes eliminating rare codons and using optimized codons adjusting GC content as well as the secondary structure of the RNA nucleotide sequence so that the formation of a complex stem-loop structure is avoided in order to have the most efficient mRNA translation.

This method was developed from many attempts by groups of graduate students in the Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, by the analysis of both positive and negative examples, and finalized by Professor Ailin Tao.

7.3.1 The Overall Principles of Codon Optimization

Codon optimization should be conducted according to the following principles:

- (1) The initiation codon AUG should not be wrapped in the stem-loop structure and terminator structures at the start site are prohibited. In order to improve efficiency, translation should be initiated while in a single-stranded state or on completely open linear structures.
- (2) Based on codon degeneracy, replace the rare codons in *Escherichia coli* with the optimal or suboptimal codons but strictly keep the target amino acid sequence unchanged.
- (3) Due to limitations on the amount of tRNAs, a single optimal codon cannot always be reused. Consider alternatively using the suboptimal codon instead.
- (4) Adjust the AT/GC content by alternative use of degenerate codons. GC and AT contents in the whole genome should be balanced and localized accumulation of GC or AT should be prevented. In order to avoid localized accumulation of CG or AT, the optimal codon(s) could be replaced by suboptimal codon(s).
- (5) The use of five or more consecutive identical nucleotides/bases should be prevented.

7.3.2 Specific Steps for Codon Optimization

The following steps are usually used for codon optimization:

- (1) Use the online software provided by the website (<http://www.kazusa.or.jp/codon/>) to initially optimize the codons of the allergen coding sequence.
- (2) Determine the rarely used codons. When the protein-expressing host is determined, codon usage is hence defined. For example, the codon usage of the commonly used prokaryote *E. coli* (strain k12) is shown in Table 7.1, which comes from the website (<http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/invitro-genetics/codon-usage.html>) and originally adapted from a book (Maloy et al. 1995), and can be utilized to make further codon optimization.
- (3) Optimize the RNA secondary structure using the DNASTar software. After initial codon optimization, the RNA secondary structure should be taken into consideration because the RNA secondary structure directly impedes protein translation; the more complex the RNA structure, the harder transcription and translation becomes. Therefore, it should be ensured that there has been no change to the amino acid sequence while the simplest RNA secondary structure is formed by applying abundant or sub-abundant RNA codons, when appropriate. The RNA secondary structure should not form complex winding or long hairpin structures. A large circular structure is optimal and the formation of six or more RNA stem-loop hydrogen bonds should be prevented. The RNA secondary structures of the rice allergen RA17, deduced before and after codon optimization, is exemplified in Fig. 7.1.

Table 7.1 Codon usage in *E. coli* K12 genes^a

	Codon	Amino acid ^b	% ^c	Ratio ^d	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio
U	UUU	Phe(F)	1.9	0.51	UCU	Ser(S)	1.1	0.19	UAU	Tyr(Y)	1.6	0.53
	UUC	Phe(F)	1.8	0.49	UCC	Ser(S)	1.0	0.17	UAC	Tyr(Y)	1.4	0.47
	UUA	Leu(L)	1.0	0.11	UCA	Ser(S)	0.7	0.12	UAA	STOP	0.2	0.62
C	UUG	Leu(L)	1.1	0.11	UCG	Ser(S)	0.8	0.13	UAG	STOP	0.03	0.09
	CUU	Leu(L)	1.0	0.10	CCU	Pro(P)	0.7	0.16	CAU	His(H)	1.2	0.52
	CUC	Leu(L)	0.9	0.10	CCC	Pro(P)	0.4	0.10	CAC	His(H)	1.1	0.48
A	CUA	Leu(L)	0.3	0.03	CCA	Pro(P)	0.8	0.20	CAA	Gln(Q)	1.3	0.31
	CUG	Leu(L)	5.2	0.55	CCG	Pro(P)	2.4	0.55	CAG	Gln(Q)	2.9	0.69
	AUU	Ile(I)	2.7	0.47	ACU	Thr(T)	1.2	0.21	AAU	Asn(N)	1.6	0.39
G	AUC	Ile(I)	2.7	0.46	ACC	Thr(T)	2.4	0.43	AAC	Asn(N)	2.6	0.61
	AUA	Ile(I)	0.4	0.07	ACA	Thr(T)	0.1	0.30	AAA	Lys(K)	3.8	0.76
	AUG	Met(M)	2.6	1.00	ACG	Thr(T)	1.3	0.23	AAG	Lys(K)	1.2	0.24
U	GUU	Val(V)	2.0	0.29	GCU	Ala(A)	1.8	0.19	GAU	Asp(D)	3.3	0.59
	GUC	Val(V)	1.4	0.20	GCC	Ala(A)	2.3	0.25	GAC	Asp(D)	2.3	0.41
	GUA	Val(V)	1.2	0.17	GCA	Ala(A)	2.1	0.22	GAA	Glu(E)	4.4	0.70
G	GUG	Val(V)	2.4	0.34	GCG	Ala(A)	3.2	0.34	GAG	Glu(E)	1.9	0.30
	C				A				G			
	U				A				G			

^aThe data shown in this table is from the Arabidopsis Research Companion on the World Wide Web ([//weeds/mgh.harvard.edu](http://weeds/mgh.harvard.edu)). Codon frequencies for many other bacteria can be found at <http://morgan.angis.su.oz.au/Angis/Tables.html>

^bThe letter in parenthesis represents the one letter code for the amino acid

^c% represents the average frequency this codon is used per 100 codons

^dRatio represents the abundance of that codon relative to all of the codons for that particular amino acid

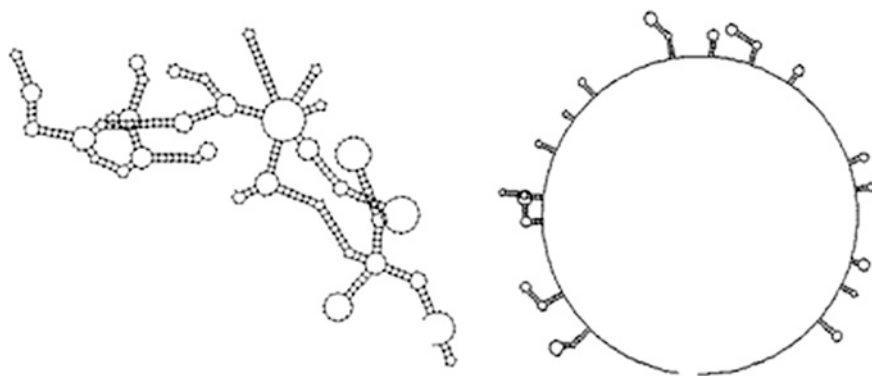


Fig. 7.1 Comparison of RNA secondary structures of the main rice allergen RA17 before and after codon optimization

7.3.3 Verification of Codon Optimization

In order to prevent an error in the optimization process, the optimized sequence should be validated. This can be accomplished using translation software (<http://expasy.org/tools/dna.html> etc.) to translate optimized nucleotide sequences into amino acid sequences and then see whether there is 100 % identity between the original target amino acid sequence and the optimized nucleotide sequence by applying a special BLAST program “bl2seq” (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=blast2seq), selecting “blastp” for amino acid sequence comparison. Gene synthesis should be considered only when there are no amino acid errors.

7.3.4 Synthesis of the Allergen Gene

After codon optimization of the allergen gene sequences, restriction sites should be added to the 3' and 5' ends of the target gene as well as affinity tags (such as Strep II, 6*His, etc.) in order to facilitate construction, digestion, and purification, after which the gene-coding sequence can be synthesized. There are some general considerations for choosing restriction sites:

- (1) The inside of target gene fragment should not contain selected restriction sites, otherwise, it will cleave off the target gene when it is identified as a positive recombinant with double enzyme digestion.
- (2) The multiple cloning sites in the vectors should contain the selected restriction sites to ensure that the direction of the target gene to be inserted into the

vector is correct (directional cloning). Otherwise, when changing the vector to express the allergen proteins, primers must be redesigned to introduce the new restriction sites.

- (3) As much as possible, choose frequently used restriction sites that have high efficiency and activity in commonly used buffers.

7.4 Prokaryotic Expression of Allergens

7.4.1 Vector Selection

A vector, by definition, is a medium carrying a substance. Cloning vectors or gene vectors can carry an endogenous gene of interest into the receptor cells, achieve asexual reproduction, or express a protein of interest. There are a great variety of vectors and a comprehensive analysis of them should be done when designing experiments. A literature search can be used to find the classic vectors, along with existing laboratory techniques and goals (such as eukaryotic or prokaryotic expression, inclusion body expression or soluble expression, resistance, etc.) that should then be taken into consideration when selecting the proper ones. Vector selection should be based on the purpose of gene construction and with consideration of the appropriate restriction sites. If the construction is for the expression of a specific gene, a suitable expression vector should be chosen.

- (1) According to their functions, vectors can be divided into cloning or expression vectors. Cloning vectors have a relaxed replicon, can help a foreign gene amplify in the host cell, and are mainly used for cloning and amplification of DNA fragments (gene). Expression vectors are capable of importing exogenous genes, transcribing them, and then translating them into proteins.
- (2) According to the type of receptor cell, vectors can be divided into prokaryotic vectors (such as plasmids in bacteria), eukaryotic vectors (such as plasmids in yeast), and shuttle vectors. Shuttle vectors are carrier molecules that can replicate and carry target genes back and forth between the two entities. Shuttle plasmids contain both prokaryotic and eukaryotic replicons to ensure amplification in two types of cells.
- (3) Attention should be centered on prokaryotic expression vectors. Select an appropriate promoter and the corresponding recipient strain with attention to reading frame dislocation when expressing eukaryotic proteins. Express the native protein or a fusion protein as a reference.
- (4) The restriction enzyme sites and the composing direction of the multiple cloning sites vary among different vectors; attention must be given to link the target gene with the vector without reading frame dislocation.

7.4.2 Plasmid Reconstruction

Bacterial strains containing the target plasmid are grown and the plasmid DNAs are extracted followed by a double restriction digestion to obtain the desired sticky gene fragments. In the meantime, the expression vector is double digested with the same restriction enzymes and the digested vector and the target gene are ligated by T4 DNA ligase, which uses ATP as energy source cofactor.

7.4.3 Allergen Protein Expression

After treatment with Ca^{2+} , original *E. coli* cells are made competent and able to uptake the constructed plasmid DNAs. The positive clones are then selected by restriction and colony PCR identification, or even by sequencing. The verified plasmids from the positive clones are then transformed into an expression host, such as *E. coli* Rosetta and alike, which contain several rare codons to improve protein expression.

During the expression process, high yields of the target protein can be obtained by optimizing the induction conditions, such as inducing temperature and time duration, and inducer concentration.

The protein of interest is inducibly expressed in the transformed host strain and is purified from the precipitate (inclusion bodies) and/or the supernatant (cytoplasm). In the *E. coli* system, many proteins are expressed in the form of inclusion bodies.

7.4.4 Formation of Inclusion Bodies

There is a synergistic effect of many factors that influence the formation of inclusion bodies (Strandberg and Enfors 1991). High expression rates and incorrect disulfide bond pairing may trigger the formation of inclusion bodies because the recombinant proteins cannot form the correct secondary bonds due to the lack of the cofactors. Culture conditions such as temperature, pH, medium composition, and other factors can also affect the formation of inclusion bodies. Moreover, in prokaryotic expression systems that lack a modification process for the newly expressed protein, inclusion bodies are easily formed (Ramon et al. 2014).

7.4.5 Denaturation of Inclusion Bodies

The denaturation of inclusion bodies usually includes the following processes:

(1) Cell lysis

To harvest inclusion bodies from genetically engineered bacteria, cell lysis can be accomplished either by mechanical methods, namely high-pressure homogenizer

and ultrasound or, when the sample is relatively small, by nonmechanical methods (Rodriguez-Carmona et al. 2010). The main nonmechanical methods include chemical disruption with sodium dodecyl sulfate (SDS) containing detergents, osmotic shock, freeze-thaw, enzyme lysis, etc. (Scotto-Lavino et al. 2006).

(2) Inclusion body washing

Crude inclusion bodies can be harvested by centrifugation of the cell lysate, which contains not only the recombinant protein but also impurities such as lipids, hybrid proteins, nucleic acids, lipopolysaccharides, etc. These impurities will seriously affect the refolding and yield of the target protein, as well as the biological activity of the protein and the downstream purification process. It is difficult to remove the membrane proteins and nucleic acids in the crude inclusion bodies, hence, some types of detergents, such as Triton X-100, NP40, or Tween 20, can be used for inclusion body washing. Sodium chloride is often added when washing the inclusion bodies because the washing power of detergents is enhanced with increased ionic strength. Usually, a high concentration of denaturant (e.g., 6 mol/L ~ 8 mol/L urea or guanidine hydrochloride, etc.) is used for dissolving and a low concentration of denaturant is used for washing. In addition, EDTA is also added into the wash solution; it chelates metal ions and prevents the target protein from degradation by any proteases.

(3) Protein dissolving

To obtain active protein, the first step is to dissolve the inclusion bodies using denaturing agents, which are generally either surfactants and/or chaotropic agents.

Surfactants: Surfactants, namely detergents, are an economical agent that can be used to dissolve the inclusion body protein. SDS and N-dodecyl sarcosine are the commonly used surfactants. The major advantage of surfactants is that the biological activity of the dissolved protein can be maintained and the aggregation of the dissolved protein is less than with the other solvents. The main drawback of surfactants is that they influence the subsequent protein refolding and purification. They strongly bind with proteins and are difficult to remove, thus, they impede ion exchange and hydrophobic interaction chromatography because of their dissociation function and the hydrophobic action of the ionic surfactants. Additionally, detergents can solubilize and activate the proteases in the membrane proteins, which may reduce the efficiency of the dissolving and refolding processes.

Chaotropic agents: The chaotropes guanidine hydrochloride and urea are commonly used during inclusion body protein treatment as a denaturant/solubilizing agent at a general concentration of 6–8 mol/L with the protein concentration being 1–10 mg/mL. Guanidine hydrochloride is a relatively expensive and stronger denaturant compared to urea. It can dissolve the inclusion bodies that urea cannot. In addition, the adventitious accumulation of cyanate in urea solutions during product manufacturing can cause unwanted carbamylation of proteins, leading to alterations in protein structure, stability, and function. These effects can be prevented by the use of an anionic buffer (such as glycineamide pH8, ethylenediamine pH7, citrate pH5, taurine pH9, etc.) and/or ion exchange

chromatography prior to urea preparation (Lin et al. 2004). The dissolving buffers can be held at ambient temperature for 14 days (Lin et al. 2004).

Extreme pH: This method dissolves certain specific inclusion body proteins. Organic acids, low in cost and high in efficiency, are generally used only for the dissolving of some inclusion bodies since some irreversible unwanted modifications or acid degradation may occur under the harsh conditions of extreme pH.

(4) Refolding of inclusion bodies

Heterologous proteins overexpressed in *Escherichia coli* often result in a significant proportion of the target protein accumulating in dense insoluble aggregates known as inclusion bodies, which are still unstructured even after being solubilized according to the above steps. The denatured unstructured proteins do not have biological activity, thus, it is necessary to refold these proteins to maximize the yield of soluble functional protein (Hwang et al. 2014).

An efficient refolding process transforms proteins from a denatured heat-unstable state to a thermodynamically stable state with a native structure and biological activity. Hence, the denaturing agents, along with fusion partners that have been specifically designed to accumulate in insoluble inclusion bodies, should first be removed. However, this process often leads to intermolecular misfolding and aggregation, and results in low refolding efficiency. The refolding efficiency actually depends on the balance between correct folding and reaggregation, the latter of which should be avoided.

Generally, when using urea or guanidine hydrochloride as the denaturants at a concentration of 6–8 mol/L, refolding begins when the concentration is reduced to 4–6 mol/L and ends after the concentration is gradually dropped to about 1–1.5 mol/L. There are many methods for the refolding of inclusion body proteins (Yamaguchi and Miyazaki 2014), therefore, it is necessary to comprehensively analyze the relevant literature and the appropriate refolding methods that are available, together with other information regarding the protein specificities such as disulfide bonds, the molecular isoelectric points, etc. (Burgess 2009) before commencing refolding. In any case, refolding efficiency is affected by several factors including the initial concentration of the refolding protein, the refolding buffer, pH, temperature, the interval time of refolding, the refolding sample methods, etc., all of which should be optimized accordingly (Porowinska et al. 2012; Su et al. 2011).

7.5 The Purification of Allergens

Similar to the usual methods, purification of recombinant proteins should take advantage of their physical and chemical properties, i.e., molecular size, shape, solubility, isoelectric point, hydrophilicity and affinity with other molecules, etc. There are many methods for protein purification. At present, the main purification methods include concentrated precipitation, chromatography, and electrophoresis.

Comprehensive analysis of the relevant literature should be done before designing experiments; common protein purification methods should be analyzed and summarized and then combined with the specific properties of the protein (isoelectric point, molecular weight, hydrophobic and hydrophilic properties, etc.) in order to choose the appropriate method of purification (Structural Genomics Consortium et al. 2008). Then, if problems are encountered, the purification scheme should be adjusted.

References

- Burgess RR. Refolding solubilized inclusion body proteins. *Methods Enzymol.* 2009;463:259–82.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res.* 2003;31(13):3497–500.
- Elena C, Ravasi P, Castelli ME, Peiru S, Menzella HG. Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Front Microbiol.* 2014;5:21.
- Frohman MA. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol.* 1993;218:340–56.
- Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A.* 1988;85(23):8998–9002.
- Hwang PM, Pan JS, Sykes BD. Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. *FEBS Lett.* 2014;588(2):247–52.
- Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 2004;5(2):150–63.
- Lin MF, Williams C, Murray MV, Conn G, Ropp PA. Ion chromatographic quantification of cyanate in urea solutions: estimation of the efficiency of cyanate scavengers for use in recombinant protein manufacturing. *J Chromatogr B Anal Technol Biomed Life Sci.* 2004;803(2):353–62.
- Linhart C, Shamir R. The degenerate primer design problem: theory and applications. *J Comput Biol.* 2005;12(4):431–56.
- Linhart C, Shamir R. Degenerate primer design: theoretical analysis and the HYDEN program. *Methods Mol Biol.* 2007;402:221–44.
- Maloy S, Stewart VJ, Taylor RK. Genetic analysis of pathogenic bacteria: a laboratory manual. New York: Cold Spring Harbor Laboratory Press; 1995.
- Porowinska D, Marszalek E, Wardecka P, Komoszynski M. In vitro renaturation of proteins from inclusion bodies. *Postepy Hig Med Dosw (Online).* 2012;66:322–9.
- Ramon A, Senorale-Pose M, Marin M. Inclusion bodies: not that bad. *Front Microbiol.* 2014;5:56.
- Rodríguez-Carmona EI, Cano-Garrido O, Seras-Franzoso J, Villaverde A, García-Fruitós E. Isolation of cell-free bacterial inclusion bodies. *Microb Cell Fact.* 2010;9:71.
- Scotto-Lavino E, Du G, Frohman MA. 3' end cDNA amplification using classic RACE. *Nat Protoc.* 2006;1(6):2742–5.
- She R, Chu JS, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* 2009;19(1):143–9.
- Souvenir R, Buhler J, Stormo G, Zhang W. An iterative method for selecting degenerate multiplex PCR primers. *Methods Mol Biol.* 2007;402:245–68.

- Strandberg L, Enfors SO. Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. Appl Environ Microbiol. 1991;57(6):1669–74.
- Structural Genomics Consortium, China Structural Genomics Consortiumm Northeast Structural Genomics Consortium, Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, et al. Protein production and purification. Nat Methods. 2008;5(2):135–46.
- Su Z, Lu D, Liu Z. Refolding of inclusion body proteins from *E. coli*. Methods Biochem Anal. 2011;54:319–38.
- Tao AL, He SH. Cloning, expression, and characterization of pollen allergens from *Humulus scandens* (Lour) Merr and *Ambrosia artemisiifolia* L. Acta Pharmacol Sin. 2005;26(10):1225–32.
- Yamaguchi H, Miyazaki M. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. Biomolecules. 2014;4(1):235–51.

Author's Biography



Dr. Xue-Ting Liu is a Docent and Junior Research Scientist at Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, the State Key Clinical Specialty in Allergy, the State Key Laboratory of Respiratory Disease, the Second Affiliated Hospital of Guangzhou Medical University. Email: jane102514@163.com.

Dr. Xue-Ting Liu received her Ph.D. from Jinan University majoring in Biochemistry and Molecular Biology. From 2005 to 2011, she worked on the antitumor effects of the protein drug sFGFR2 and how it exerts its antitumor activity. Since 2011, she has been working at the Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology. Her main research interests focus on the mechanisms responsible for chronic urticaria caused by Staphylococcal Enterotoxin A (SEA) and reducing the antigenicity of protein drugs, such as SEA. Dr. Liu has published several scientific papers within internationally well-recognized journals. She has participated in several projects for the Science Foundation of National and Guangdong Provincial and presided over a project from the National Natural Science Foundation.



Professor Ailin Tao is a Professor at Guangzhou Medical University, Director of Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, Principal Investigator of the State Key Laboratory of Respiratory Disease, Deputy Director of the State Key Clinical Specialty in Allergy of the Second Affiliated Hospital of Guangzhou Medical University, Member of the State Committee for Transgenic Safety Assessment, Standing Committee Member of Allergy Branch of Guangdong Medical Association, Member of Guangdong Provincial Committee for Transgenic Safety Assessment and Master Tutor.

Prof. Ailin TAO earned his doctorate degree from the State Key Laboratory of Crop Genetic Improvement of Huazhong Agricultural University in 2002, followed by a postdoctoral training at Postdoctoral Station of Basic Medicine in Shantou University Medical College, majoring in allergen proteins. His most recent research has been on allergy bioinformatics, allergy and clinical immunology and disease models such as allergic asthma, allergic rhinitis, infection and inflammation induced by allergy, inflammatory and protracted diseases caused by antigens or superantigens. He has gained experience in the field of allergology including the mechanisms of immune tolerance, allergy triggering factors and chronic inflammation pathways, and allergenicity evaluation and modification for food and drugs. He proposed some new concepts including “Representative Major Allergens”, “Allergenicity Attenuation” of immunotoxin and allergens, “broad-spectrum immunomodulator” as well as the theoretical hypothesis of “Balanced Stimulation by Whole Antigens”. Prof. TAO’s laboratory focuses on the diagnosis of allergic disease and the medical evaluation of food and drug allergenicity and its modification. Prof. TAO has now constructed a system for the prediction, quantitative assessment, and simultaneous modification of epitope allergenicity, which has been applied to more than 20 allergens, and he also developed a bioinformatics software program for allergen epitope prediction, SORTALLER (<http://sortaller.gzhmu.edu.cn>), which performed significantly better than other existing softwares, reaching a perfect balance of high specificity (98.4 %) and sensitivity (98.6 %) for discriminating allergenic proteins from several independent datasets of protein sequences of diverse sources. Furthermore, this program has a Matthews correlation coefficient as high as 0.970, a fast running speed, and can rapidly predict a set of amino acid sequences with a single click. The software has been frequently used by researchers from many institutions in China and over 30 countries worldwide, thus becoming the number one allergen epitope prediction software program. Prof. TAO has set up an allergen database ALLERGENIA (<http://ALLERGENIA.gzhmu.edu.cn>) that has several advantages over other databases such as a wide selection of nonredundant allergens, excellent stringency and accuracy, and friendly and usable analytical functions.