CHAPTER III
LITERATURE REVIEW

1. Biography of *Naja kaouthia* (Deraniyagala, 1960)

N. kaouthia (previously known as *Naja naja siamensis*) or Thai cobra snake is a reptile classified in the class: Reptilia; Subclass: Synptosuria; Order: Squamata; Suborder: Serpentes; Family: Elapidae, and Subfamily: Bungarinae.

2. Morphology and geographical distribution of *N. kaouthia* (www.saovabha.org/snake/thai.php).

N. kaouthia is a poisonous snake. It is a monocellate and non-spitting cobra. There are two groups of the poisonous snakes in the Elapidae family, *i.e.*, terrestrial and marine snakes. *N. kaouthia* is the terrestrial snake. It has a complete venom apparatus which contains real venom glands, highly toxic venom, and proteroglypha fangs. Its length is ~1,000 mm (head ~40 mm, body ~750 mm, and tail ~210 mm) and the color varies widely (**Figure 1**). *N. kaouthia* is oviparous and lays eggs per clutch. It preys on toads, snakes, birds, and small mammals, and is usually found in termite mounds, near human habitation, forests and foothills upto ~900 meters. *N. kaouthia* is also found in Myanmar, Cambodia, Loas PDR, Vietnam, and Malaysia

Figure 1 Morphology of *Naja kaouthia* or Thai cobra (from www.saovabha.org /snake/thai.php) Accessed on Feb 25, 2008

3. Epidemiology of snake bites in Thailand

In 2005, there were 7,370 of the reported survival cases of poisonous snake bites in Thailand (12 per 100,000 populations). There were five fatal cases in the year 2005 (mortality rate 0.01 per 100,000 populations) and the biting poisonous snakes could not be identified. When comparing with the years 2000-2004, there were 8,018-11,325 of the reported cases of poisonous snake bites (median = 9,071 cases). **Figure 2** shows the numbers of the reported cases of poison snake bites per 100,000 population during 1997-2006. For seasonal variation in the numbers of the reported cases in 2005, the highest number of the reported cases was found in May-October which are the rainy months in the Kingdom and the agricultural activities especially rice farming are active (**Figure 3**). Among the 7,370 of reported survival cases in the year of 2005, there were 6,424 cases which the biting poisonous snakes could not be identified (87%). Among the 946 cases which the biting poison snakes were identified (13%), they were 446 cases by Malayan pit viper, 352 cases by green pit viper, 114 cases by cobra, 9 cases by Russel's viper, 3 cases by King cobra, 3 cases by *Bungurus fasciatus*, 3 cases by marine snake, and 16 cases by the other poisonous snake(s) (**Figure 4**) (http://epid.moph.go.th).

In 2006, there were 8,299 of the reported survival cases of poisonous snake bites in Thailand (13 per 100,000 populations). Five of the reported cases in the year 2006 were dead (mortality rate 0.06 per 100,000 populations) and the biting poisonous snake was the cobra in one of the five fatal cases (biting snake of the other four cases could not be identified). For seasonal variation in numbers of the reported cases in 2006, the similar pattern of those of the 2005 was observed, *i.e.* the highest number of the reported cases was during May-October (**Figure 3**). Among the 8,299 of the reported survival cases, there were 7,140 cases which the biting poisonous snakes could not be identified (86%). Among the 1,159 cases which the biting poisonous snakes were identified (14%); there were 468 cases by Malayan pit viper, 427 cases by green pit viper, 189 cases by cobra, 11 case by Russel's viper, 1 case by King cobra, 6 cases by *Bungurus fasciatus*, 3 cases by sea snake, and 54 cases by the other poisonous snake(s) (**Figure 5**) (http://epid.moph.go.th).

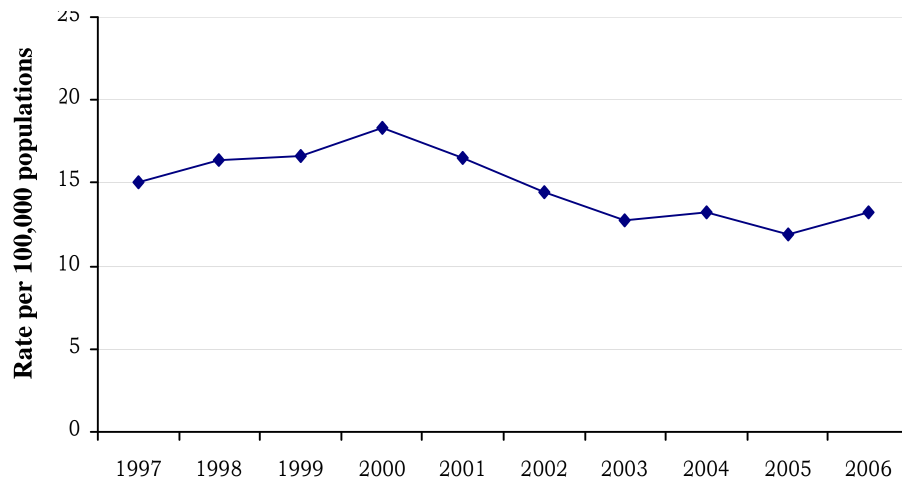


Figure 2 Reported cases of poisonous snake bites per 100,000 populations by year, Thailand, 1997-2006. (http://epid.moph.go.th, ISSN 0857-6521). Accessed on Feb 25, 2008

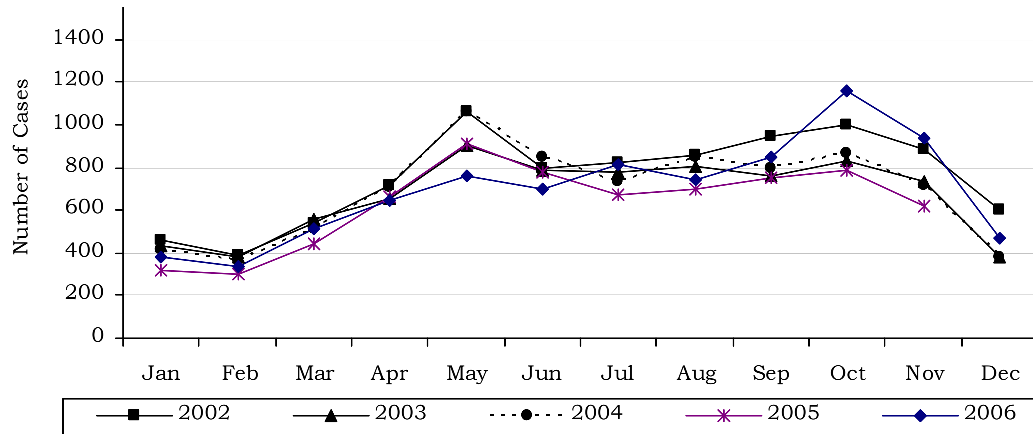


Figure 3 Reported cases of poisonous snake bites by months, Thailand, 2002-2006. (http://epid.moph.go.th, ISSN 0857-6521). Accessed on Feb 25, 2008

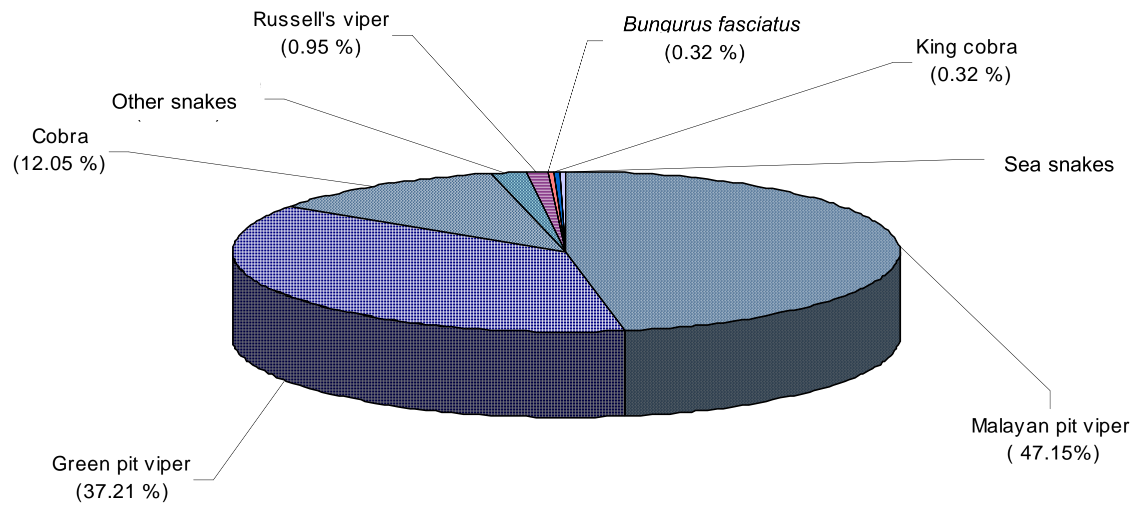


Figure 4 Proportion (%) of snake bites by snake types in Thailand, 2005. (http://epid.moph.go.th, ISSN 0857-6521). Accessed on Feb 25, 2008

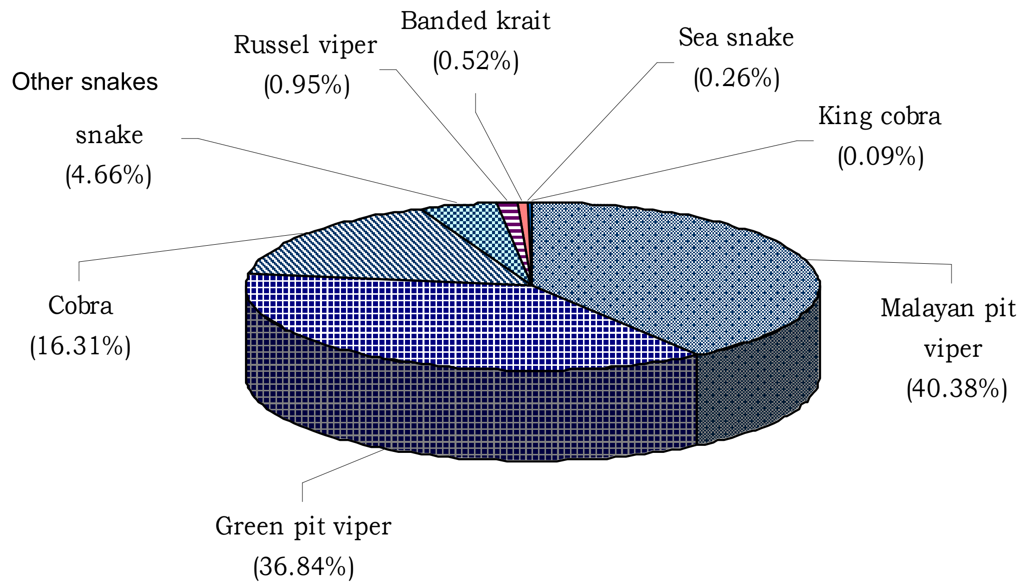


Figure 5 Proportion (%) of snake bites by snake types in Thailand, 2006. (http://epid.moph.go.th, ISSN 0857-6521). Accessed on Feb 25, 2008

4. Composition of *N. kaouthia* venom

N. kaouthia (Thai cobra) venom is composed of several protein and non-protein components produced by a special sero-mucous gland and inoculated by fangs which permit injection of venom under pressure into tissue of the preys.

Most of protein components have pharmacological properties which are either toxins or enzymes as described below:

4.1 Neurotoxins (Yang, 1996; Tsetlin, 1999)

Based on the pharmacological and biochemical properties, neurotoxins have been classified into 4 classes:

A. Pre-synaptic neurotoxins (Wernicke *et al.*, 1974)

Pre-synaptic neurotoxins or β-neurotoxins cause muscle contraction without stimulation of the nerve axon. The β-neurotoxin affects the pre-synaptic end of the nerve, and initiates the release of acetylcholine (Ach), and eventually stops the release of the neurotransmitter.

β-neurotoxin induces the influx of Ca^{2+} to the nerve terminal, which has acetylcholine containing-vesicles, and causes the exocytosis of the neurotransmitter which is made by the fusion of the vesicle and the nerve-terminal membranes. Based on the differences in the protein structure, various types of pre-synaptic neurotoxins have been identified and studied. Phospholipase activity is one common property of those pre-synaptic neurotoxins.

B. Post-synaptic neurotoxins (Monez, 1987; Ji-Fu *et al.*, 2003; Ji-Fu *et al.*, 2004)

Post-synaptic neurotoxins or α-neurotoxins are commonly found in the venom of elapids and hydrophids. *N. kaouthia* produces the highest content (approximately 25%) of the neurotoxins among all terrestrial snakes (Karlsson *et al.,* 1971).

α-neurotoxins are in a member of three-finger neurotoxin (**Figure 6**) and classified into long and short neurotoxins according to the length of the polypeptides and number of disulfide bonds. The short chain α-neurotoxin contains 60-63 amino acids with four disulfide bonds while the long chain α-neurotoxin contains 65-75

amino acids with five disulfide bonds. Tsetlin (1999) made an extensive review of the α-neurotoxins and other three-finger neurotoxins (please see **Figure 6**).

Three dimensional structure of the α-cobratoxin isolated from *N. naja siamensis* (presently, *N. kaouthia*) was firstly revealed by using X-ray diffraction (Walkinshaw, 1980). The short chain neurotoxin is a 71 amino acid polypeptide that folds into three major loops and one long tail at C-terminal. The molecule is hold by five disulfide bonds (**Figures 7 and 8**). The protruding long central loop (amino acids 21-40, named loop II) is flanked by two smaller loops (amino acids 4-13 and 44-55, named loops I and III, respectively). The long tail is located at C-terminal of the peptide chain (amino acids 63-71).

α-neurotoxin affects the neuromuscular junction at the post-synaptic site by combining with muscular type (heteropentamer, α_1, α_2, β, γ, and δ) and homopentameric neuronal (α7) nicotinic acetylcholine receptor (nAchR) on the muscle cell membrane (for more details, please see Tsetlin, 1999). When a normal nerve impulse (depolarization wave) passes through the axon and reaches the end of that axon, the calcium-ion concentration is increased and the neurotransmitter, Ach, is immediately released from the vesicles at the end of the nerve. Ach moves across the synaptic crevice and reaches the AchR on the muscle. The AchR is composed of five subunits, *i.e.*, α_1, α_2, β, γ, and δ. When two molecules of Ach attach to the α-subunits, the AchR changes configuration and becomes an open ion channel, permitting certain ions to pass through. By this mechanism, the depolarization reaches the muscle. Since there are two α-subunits, the stoichiometry of ligand-receptor interaction is two molecules of Ach per one AchR. Post-synaptic neurotoxin attaches to the same sites as the Ach; however, the AchR receptor fails to form a channel. Post-synaptic neurotoxins block post-synaptic neurotransmission, thus causing flaccid paralysis resulting in respiratory failure and possibly death of the envenomed victims (Monez, 1987; Ji-Fu *et al.*, 2003; Ji-Fu *et al.*, 2004).

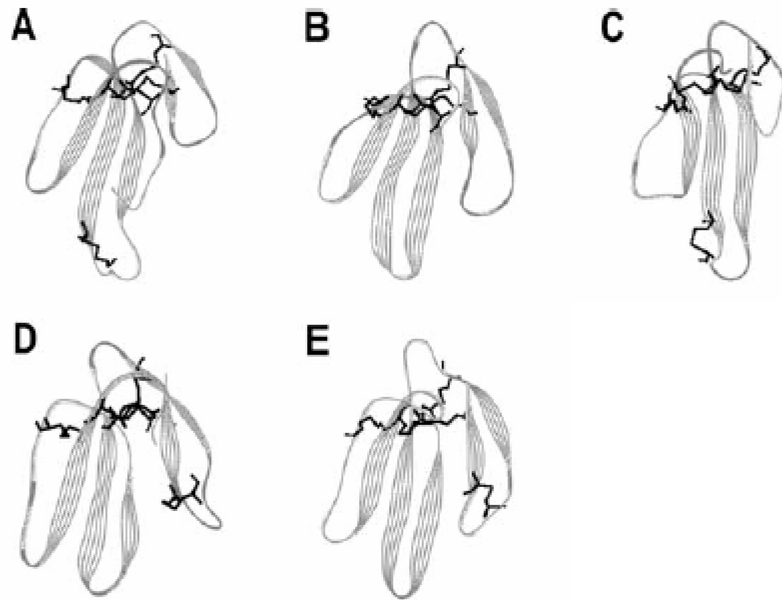


Figure 6 The three dimensional structure of five representatives of the three-finger toxins (Nirthanan *et al.*, 2003).

A, α-cobratoxin from *N. kaouthia*, accession no. 2CTX

B, Erabutoxin-a of *L. semifasciata*, accession no. 5EBX

C, k-bungarotoxin monomer of *B. multicinctus*, accession no. 1JGK

D, Bucandin of *B. candidus*, accession no. 1F94

E, Candoxin of *B. candidas*, accession no. 1JGK

Disulfide bonds are shown in black.

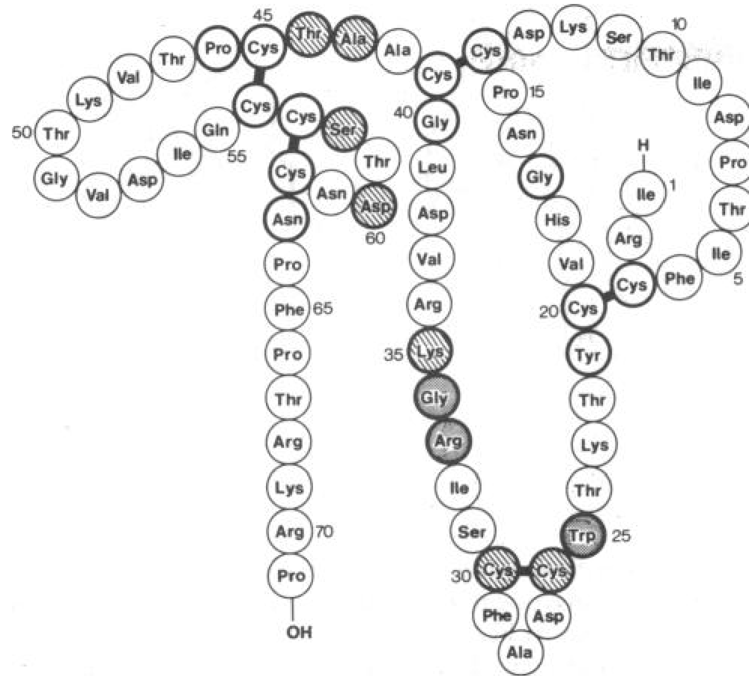


Figure 7 Two dimensional structure of long α-neurotoxin of *N. naja siamensis* (presently, *N. kaouthia*). Thick lines indicate invariant amino acid residue among the long, short, and cardiotoxin families, hatched circles are invariant amino acids in only long neurotoxin family. Strippled circles are invariant amino acids only in long and short toxin families but not in cardiotoxins (Walkinshaw *et al.*, 1980).

Amino acids 4-13, the first small loop or loop I

Amino acids 21-40, the second large loop or loop II

Amino acids 44-55, the third small loop or loop III

Amino acids 63-71, the C-terminal tail of the toxin

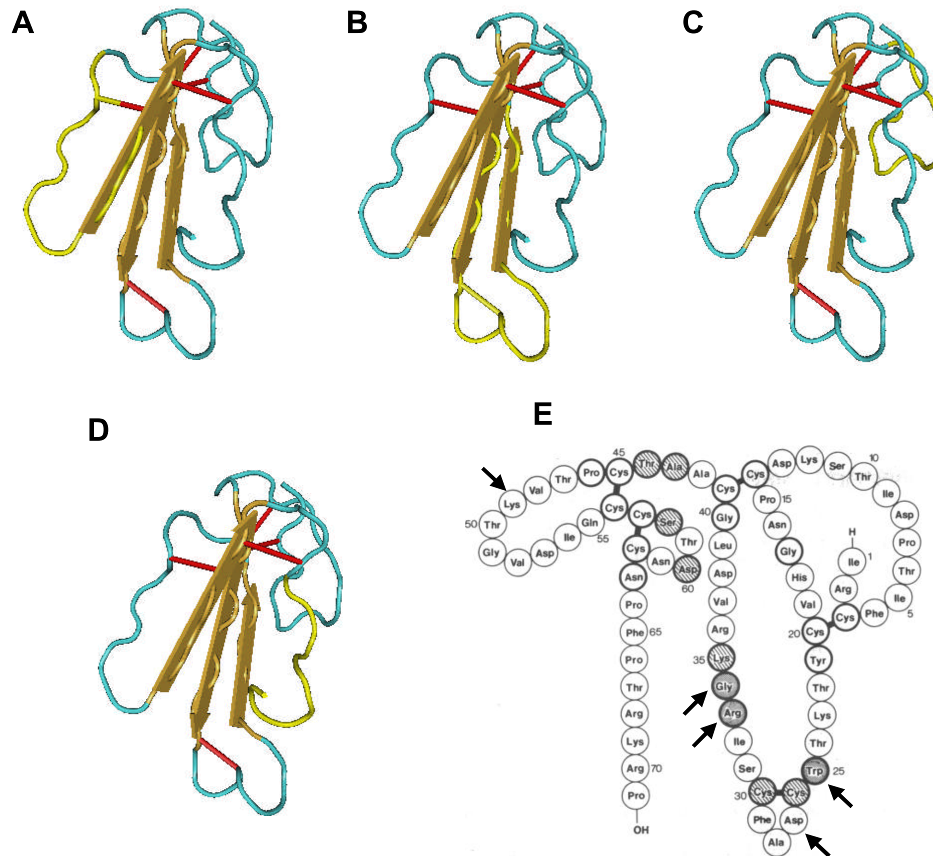


Figure 8 Three dimensional structure and secondary structure of long neurotoxin of *N. naja siamensis* (gi|229777|pdb|1CTX). The beta sheet structure of the protein is colored as dark yellow.

The blue color, long tube structure

The red color lines, disulfide bonds

A, Yellow ribbon indicates the small loop (amino acids 44-55)

B, Yellow ribbon indicates the large loop (amino acids 21-40)

C, Yellow ribbon indicates the small loop (amino acids 4-13)

D, Yellow ribbon indicates the tail of protein (amino acids 63-71)

E, Arrows indicate amino acid residues which play important role in acetylcholine receptor binding identified by Karlsson *et al.* (1979).

For molecular mechanism involved in ligand-receptor binding, Walkinshaw (1980) found from the chemical modification of amino acid and toxicity study that the invariant amino acids, *i.e.*, Trp-25, Asp-27, Arg-33, and Gly-34 in large central loop II (amino acid 21-40 in **Figures 7** and **8**), and Lys-49 in small loop III (amino acid 44-55 in **Figures 7** and **8**) are likely to be involved in acetylcholine receptor binding (for more details, please see Low, 1979 and Karlsson, 1979).

There are several studies that reported the important role of large loop II of long α-neurotoxin or α-cobratoxin (α-Cbtx) in receptor binding. From experiment based on molecular docking model of neuronal acetylcholine receptor (α7), Fruchart-Gaillard *et al*. (2002) reported that the tip of the large loop II plugs into the receptor between two α-subunits of α7 receptor (**Figure 9**). The critical amino acid residues that bind with high affinity to the α7 receptor are located at the large loop II: D27, F29, R33, K35, C26, C30, R36, and at C-terminal tail; F65.

Bourne *et al*. (2005) studied crystal structure of α-cobratoxin (or α-Cbtx) isolated from the venom of *N. naja siamensis* and the pentameric acetylcholine binding protein (AChBP) interaction. They found that the AChBP was a homolog of the extracellular domain of muscular type-nAChR which was isolated from a freshwater snail (*Lymnaea stagnalis*) (Brejc *et al.*, 2001). They found that the amino acid residues mostly located in loop II of toxin involved in receptor binding as shown in **Figure 10**.

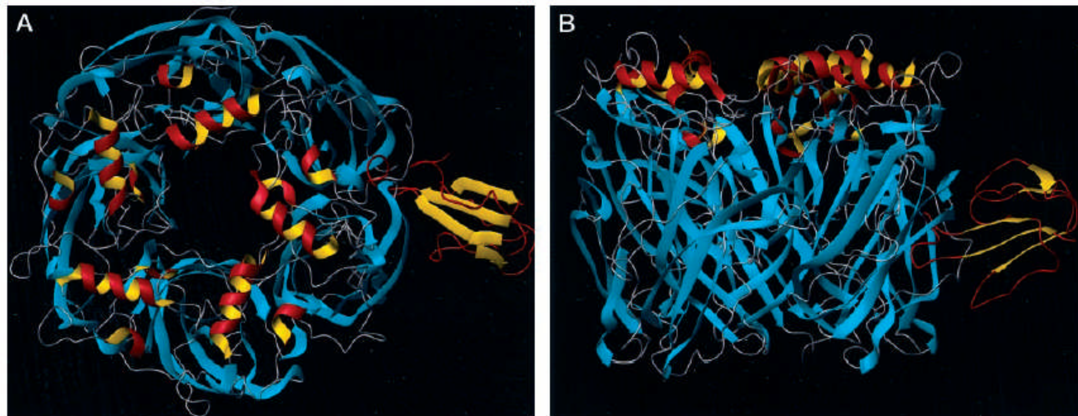


Figure 9 Three dimensional model of α7 receptor-α-cobratoxin (α-Cbtx) complex. The five subunits of α7 are indicated by blue β-sheet and the apical helix colored red and yellow. Beta-sheet and backbone of α-Cbtx are colored in yellow and red, respectively. (**A**) Top view and (**B**) side view are shown (Fruchart-Gaillard *et al*., 2002).

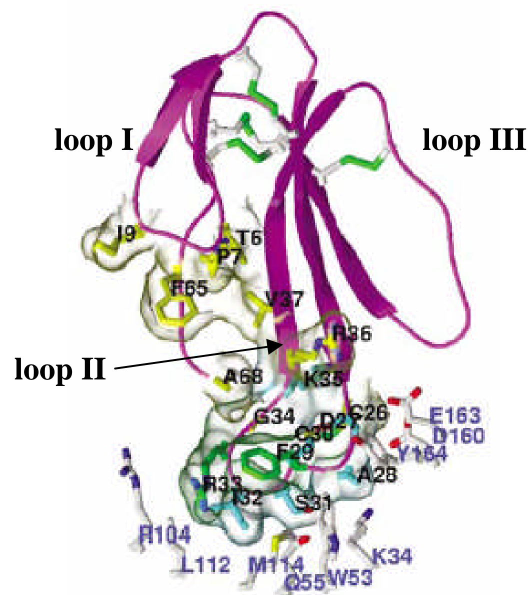


Figure 10 The three dimensional structure of Cbtx-AChBP complex. The amino acid residues in α-cobratoxin (α-Cbtx) labeled in black interacts with the binding-pocket of the AChBP subunit. The amino acid residues of AChBP are in white (blue labels) (Bourne *et al.*, 2005).

Another type of α-neurotoxin is a short chain α-neurotoxin. Meng *et al.* (2002) isolated a short chain α-neurotoxin from *N. kaouthia*. By Edman degradation technique, this short chain α-neurotoxin had 61 amino acids. The LD50 of this short chain α-neurotoxin was 0.08 μg/g of mouse. In receptor binding, the short chain α-neurotoxin binds specifically to the muscular type of nAChR. Mordvintsev *et al.* (2005) studied the interaction of the nAChR and the three short α-neurotoxins, *i.e.*, neurotoxin I of *N. mossambica mossambica* (NmmI), neurotoxin II of *N. oxiana* (NTII), and erabutoxin-a of *Laticauda semifasciata* (Erabu), by using computational molecular docking and dynamics model. All of the three short α-neurotoxins used the tip of their loops II to insert into the ligand-binding pocket between the α/γ or α/δ nAChR, while loops I and III contact with nAChR in surface touch manner (**Figure 11**). Amino acid residues involved in receptor binding also are shown in **Figure 12**.

C. **Anti-acetylcholine esterase neurotoxins** (Lee, 1972; Yang, 1974; Karlsson, 1979)

Anti-acetylcholine esterase neurotoxin is a single polypeptide chain with 57-60 amino acid residues that is cross-linked by three disulfide bonds. The toxin forms a specific complex with acetylcholine esterase and inhibits the enzyme. When acetylcholine esterase is not functioning, Ach (after binding to the AchR) cannot be hydrolyzed; consequently, normal nerve transmission is impaired. The toxin increases the amplitude and time course of the end-plate potential causing continuous excitation of the muscle. These toxins confer higher lethality when synergized with other facilitator which enhances the release of Ach.

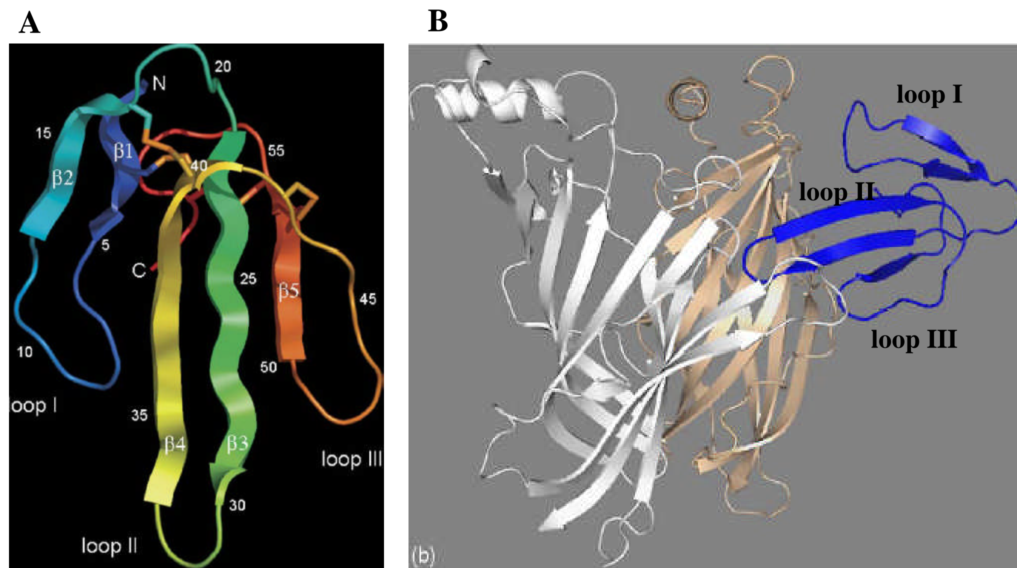
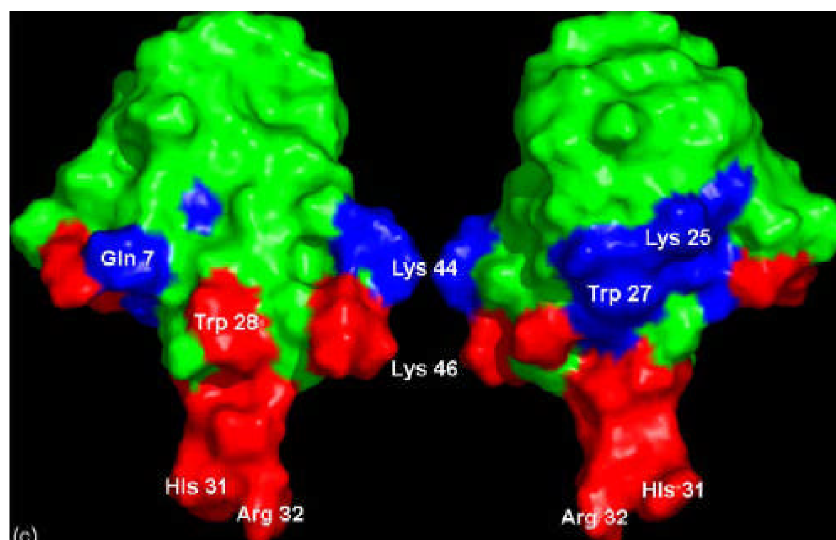


Figure 11 Three-dimensional structure of short α-neurotoxin (**A**) and the protein interaction between short α-neurotoxin (NTII) and α/γ-subunits of the muscular type-nicotinic acetylcholine receptor (nAChR) (**B**) of *Torpedo californica*. The ligand-binding site of α/γ interface is colored white and grey, respectively. The short α-neurotoxin (NTII of *N. oxiana*) is colored blue (Mordvintsev *et al.*, 2005).

A



B

```
              loop I                  loop II           loop III
NTII   1 LECHNQQSSQPPTTKTCS-GETNCYKKWWSDHRGTIIERGCGCPKVKPGVNLNCCRTDRCNN 62
Erabu  1 RICFNHQSSQPQTTKTCSPGESSCYNKQWSDFRGTIIERGCGCPTVKPGIKLSCCESEVCNN 62
NmmI   1 LECHNQQSSEPPTTTRCSGGETNCYKKRWRDHRGYRTERGCGCPTVKKGIELNCCTTDRCNN 62
```

Figure 12 (**A**) Amino acid residues involved in binding with α/γ interface of the muscular type-nicotinic acetylcholine receptor (nAChR). Three dimensional image of neurotoxin II of *N. oxiana* (NTII) is shown at concave side (left) and at convex side (right). In receptor binding, amino acid residues predicted by using the docking and molecular dynamics are colored red (**Trp28, His31, Arg32,** and **Lys46**) and predicted by only molecular dynamics are colored blue (**Gln7** and **Lys44**). (**B**) Multiple alignments of three short α-neurotoxins are shown. Cysteine residues are colored green. Loop I, II, and III are indicated by thick lines (Mordvintsev *et al.*, 2005).

NTII, neurotoxin II of *N. oxiana*

Erab, erabutoxin-a of *Laticauda semifasciata*

NmmI, neurotoxin I of *N. mossambica mossambica*

4.2 Cardiotoxins (Kumar *et al*., 1997)

Cardiotoxins are small molecular mass (6.5-7 kDa or approximately 60-62 amino acid residues), highly basic polypeptide cross-linked by four disulfide bridges (Kumar *et al*., 1997; Jang *et al*., 1997; Yu *et al*., 1994). They are β-sheet proteins with no helical structure, and are found only in the *Naja* species (the cobras) and the *Hemachatus* (the ringhals) (Jang *et al*., 1997). Cardiotoxin structure resembles the short-chain post-synaptic neurotoxins and exhibit a variety of biological activities on the different types of cells including hemolytic activity (hemolysis), cytotoxicity, depolarization of membranes of excitable cells, membrane fusion, selective killing of certain types of tumors cells, inhibition of protein kinase C activity, and muscle contraction (Kumar *et al*., 1996). Cardiotoxins isolated from venom of different poisonous snakes possess similar backbone folding (Kumar *et al*., 1997).

4.3 Acetylcholine esterase

Acetylcholine esterase which was purified from *N. naja oxiana* venom by affinity chromatography consists of a single polypeptide chain of molecular weight 67,000 ± 2,000 Da. At the high enzyme concentrations (> 0.2 mg/ml, > 1 μM) and ionic strength (0.1 M), it tends to form reversible higher-molecular-weight 7.1 S aggregates. The disc electrophoresis and isoelectric focusing experiments revealed that the acetylcholine esterase exists in one form with a common molecular weight but with different isoelectric points (pI) (Raba *et al*., 1979). There is no report about purification of acetylcholine esterase from *N. kaouthia* but one report before this study revealed that *N. kaouthia* venom had acetylcholine esterase activity which could be inhibited by plant polyphenols (Pithayanukul *et al*., 2005).

4.4 Phospholipase A_2

Phospholipase A_2 (PLA_2) is one of the major components of *N. kaouthia* venom (Doley and Mukherjee, 2003; Doley *et al*., 2004). PLA_2 is the lipolytic enzyme that specifically catalyzes the hydrolysis of fatty acid ester bonds at position 2 of 1,2-diacyl-*sn*-3-phosphoglycerides to produce free fatty acids and lysophospholipids (Dennis, 1983) (**Figure 13**). The *in vitro* assay of PLA_2 activity was described in article of Huang *et al*. in 1993. One unit of PLA_2 is designed as the amount of enzyme that could hydrolyze 1 μM phosphatidylcholine at 37°C per minute (Huang *et al*.,

1993). Doley and Mukherjee isolated PLA_2 which was deposited in the database as EC.3.1.1.4, and characterized functions such as anticoagulant (Doley and Mukherjee, 2003) and hydrolysis of erythrocytes and mitrochrondial membrane (Doley *et al.*, 2004). The direct toxic effects of snake's PLA_2 are the hydrolyzing phospholipids of cellular or subcellular membranes that can cause membrane dysfunction. Moreover, lysophospholipids and free fatty acids (**Figure 13**), which are products of phospholipid hydrolysis, can also cause membrane damage (Mukherjee *et al.*, 1997). Mukherjee isolated PLA_2 isoenzyme from *N. kaouthia* (Indian cobra) namely PLA_2A and PLA_2B, which had no lethal and no neurotoxic activity but had different molecular masses (13,619 and 13,303 Da, respectively) (Mukherjee, 2007). They had anticoagulant and cytotoxic functions but poor hemolytic activity. The PLA_2 isolated from various sources had different affinities for biomembranes which were composed of different phospholipid polar head groups and hydrophobic fatty acid tail chains (Grandbois *et al.*, 1998).

4.5 Cobra venom factor (Fritzinger *et al.*, 1994)

Cobra venom factor (CVF) is the complement-activating protein from cobra venom. CVF causes complement consumption in human and mammalian sera. It is a structural and functional analog of complement component C3. CVF functionally resembles C3b, one of the activated forms of C3. Like C3b, CVF binds factor B in human and mammalian sera, which is subsequently cleaved by factor D to form the biomolecular enzyme $\overline{\text{CVFBb}}$ and Ba $\overline{\text{CVFBb}}$ is a C3/C5 convertase that cleaves both complement components C3 and C5, then the downstream complement cascade is followed.

From deduced amino acid sequence, CVF is a three-chain 149,000-Da glycoprotein that structurally resembles the C3b degradation product C3c (which is the complement component that is unable to form a C3/C5 convertase). CVF are synthesized as single-chain pre-pro-proteins consists of a 22-amino acid signal sequence, a 627-amino acid α-chain, and a 989-amino acid precursor chain for CVF γ- and β-chains (Fritzinger *et al.*, 1994).

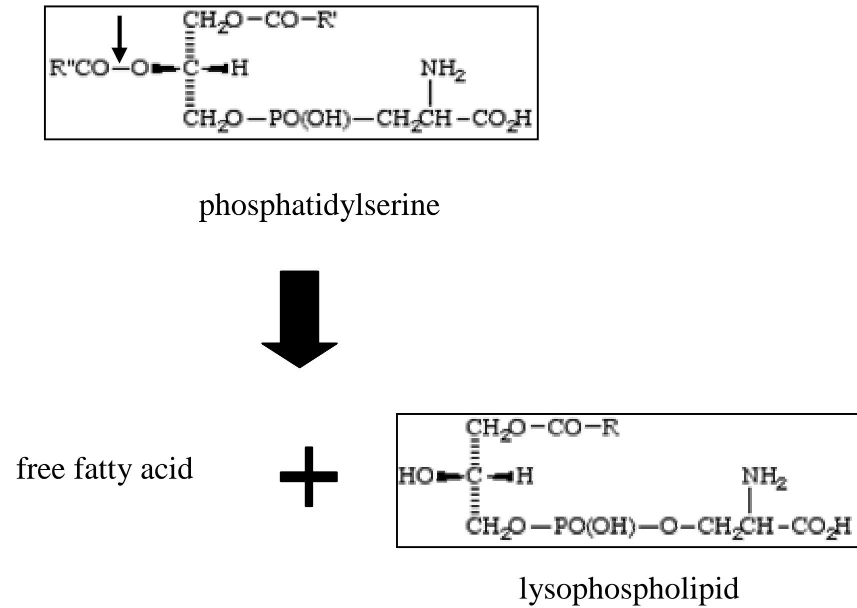


Figure 13 Mechanism of phospholipase A2 in hydrolysis of phospholipids (phosphatidylserine). Arrow indicates the site of ester bond. The products of hydrolysis are free fatty acid and lysophospholipid which are toxic to host cell.

4.6 Kaouthiagin (Ito *et al*., 2001)

Kaouthiagin is a metalloproteinase which specifically cleaves human von Willebrand factor (VWF). Metalloproteinases require divalent cation such as zinc or calcium ions for enzymatic activity or structural conformation. By direct amino acid sequencing, kaouthiagin is composed of 401 amino acid residues and one Asn-linked sugar chain. The sequence is highly similar to high molecular mass metalloproteinase of viperid and crotalid venoms comprised of metalloproteinase, disintegrin-like, and Cys-rich domains (Ito *et al*., 2001).

Kaouthiagin is an *o*-phenanthroine or EDTA sensitive metalloproteinase which binds to and cleaves human VWF (Hamako *et al*., 1998; Miura *et al*., 1999). VWF is a multimeric plasma protein essential for platelet adhesion to the damaged subendothelial matrixes to form a hemostatic plug (Fujimura and Titani, 1993). Kaouthiagin cleaves VWF at a peptide bond between Pro708 and Asp709, resulting in a loss of the platelet- and collagen-binding activities of the VWF by degrading the multimeric structure of VWF (Hamako *et al*., 1998).

5. Symptoms and signs of snake bites (reviewed by Warrell, 1999; WHO Guideline)

There are two types of symptoms and signs manifested by snake bitten victims. Symptoms and signs can occur locally only at the bitten part and the systemic manifestations that appear throughout the patient's body.

5.1 Local symptoms and signs at the bitten part

Early symptoms and signs are immediate pain due to mechanical penetration of the skin by the snake's fangs, increased local pain (burning, bursting, and throbbing) at the site of the bite, local swelling that gradually extends proximally up the bitten limb, and tender, and painful enlargement of the regional lymph node draining the site of the bite. The later symptoms and signs are inflammation (swelling, redness, and heat), local infection, abscess formation, and tissue necrosis (Khandelwal *et al*., 2007).

5.2 Generalized (systemic) symptoms and signs

General symptoms and signs found in some snake bites are nausea, vomiting, malaise, abdominal pain, weakness, drowsiness, and prostration. For the bites of elapidae snakes including Thai cobra, the neurological symptoms and signs are drowsiness, paraesthesiae, abnormalities of taste and smell, "heavy" eyelids, paralysis

of facial muscles and other muscles innervated by the cranical nerves, aphonia, difficulty in swallowing secretions, and respiratory and generalized flaccid paralysis (Khandelwal *et al.*, 2007).

The factors of the snake bite victims that can affect the degree of severity are: 1) age and size of the victim; 2) sensitivity of the victim to venom; 3) the location, depth, and number of bites; and 4) the degree and kind of first-aid treatment (Limithongkul, 1987; Nelson, 1989).

For the factors of snake that can affect the degree of severity are: 1) species and size; 2) amount of injected venom; 3) eke condition of the fangs and the venom gland; 4) pathogens present in snake's mouth; and 5) the extent of anger or fear that motivates the snake to strike (Limithongkul, 1987; Nelson, 1989).

6. Current methods for production of horse anti-snake venom

The first available antivenom was produced against cobra venom by Calmette in 1984. The production of anti-snake venom against almost venomous snake species is currently manufactured by immunizing horse because of the large blood volume available (Sunthornandh *et al.*, 1994). Horse whole blood is collected after a satisfactory level of specific antibodies is reached. The sequential steps of antivenom production are described below:

6.1 Animal immunization

6.1.1 Immunogen

The crude snake venom is used to immunize horse together with appropriate adjuvant (Sunthornandh *et al.*, 1992). The venom of the terrestrial elapid snakes, *i.e.*, *N. kaouthia*, *Opiophagus hannah*, and *Bungarus fascitus*, contain post-synaptic neurotoxin (MW~7-8 kDa) which is very toxic to horse and also confer low immunogenicity. Horse immunization with low dose of the venom of the terrestrial elapid snake usually results in low titer of neutralizing antibody in the horse blood.

In order to reduce the toxicity of the venom, toxiod or venoid preparation for using as immunogens in antivenom production were prepared. Aldehydes, *e.g.*, formaldehyde or glutaraldehyde, were used to treat toxins in the snake venom. However, the venom preparations still could not raise satisfactory level of the neutralizing antibody because the chemical modifications of the toxins by using the

aldehydes destroyed the active sites as well as the antigenic determinants of the toxins (Sunthornandh *et al*., 1992, 1994).

6.1.2 Adjuvants

Oil adjuvants, *i.e.*, Freund's complete and incomplete adjuvants have been used for antivenom production. Freund's adjuvant is a water-in-oil emulsion prepared with non-metabolized oil. The Freund's complete adjuvant contains killed *Mycobacterium tuberculosis* but there is no killed bacterium in the Freund's incomplete adjuvant. Main disadvantage of using the Freund's adjuvants in animal immunization is the formation of a persistent granuloma, inflammatory reaction, and sterile abscess. Granuloma in the tissue contains machrophages, lymphocytes, and plasma cells. New strategy for horse immunization against snake venom using Freund's adjuvants was developed by Chotwiwatthanakun *et al*. (2001) to reduce local inflammatory reaction in the horse tissue and to get high potency neutralizing antibody. This was done by injecting small volume (0.1-0.2 ml) of Freund's complete adjuvant emulsified-immunogen at multiple sites (10-20) instead of large volume (5-10 ml) of adjuvant emulsified-immunogen at one site. The other adjuvants, that have been used, included aluminium hydroxide and alginate (Kawamura, 1989).

6.2 Fractionation protocol

After a satisfactory anti-venom level was reached, the horse is bled and the anti-serum was collected. The conventional method for obtaining $F(ab')_2$ antibody fragments used by the majority of manufacturers, is the treatment of serum proteins with pepsin at acidic pH to remove the Fc fragment. The $F(ab')_2$ fragments are then purified by ammonium sulfate or sodium sulfate precipitation. The salt is then removed by dialysis or ultrafiltration (Reid, 2003). The $F(ab)'_2$ fragments can be further purified by ion-exchange column chromatography. Some laboratories produced antivenoms consisting of whole IgG molecules purified by ammonium sulfate precipitation of the serum protein (Nikalayenko, 2005). Some laboratories used caprylic acid precipitation of non-IgG proteins (dos Santos *et al.,* 1989), followed by ultrafiltration of the product.

6.3 Pasteurization and endotoxin exclusion

After the antibody is obtained, all microbial contaminants and endotoxin must be excluded. The pasteurization is performed for 10 hours at 60°C. Endotoxin should be excluded to less more than 0.5 unit/dose.

7. Immunoglobulin gene rearrangement and diversification

7.1 Basic structure of immunoglobulin (Goldsby *et al*., 2000)

Immunoglobulin or antibody is glycoprotein which consists of two identical light chains and two identical heavy chains (**Figure 14**). Each light chain is linked to a heavy chain by disulfide bonds (except for IgA1) and noncovalent bonds, *e.g*., Van der Waals force, salt linkage, hydrogen bonds, and hydrophobic bonds. Both heavy and light chain contains two main regions: variable (V) and constant (C) regions. Amino acid residues of V region of heavy chain (VH) and light chain (VL) are different among antibody molecules of different epitope specificities. The difference occurs within areas called "complementarity-determining regions (CDRs)". There are three CDRs, *i.e*., CDR1, CDR2, and CDR3 on each VH and VL (**Figure 14**). Three CDRs on V region of both heavy and light chains form antigen (epitope)-binding site. The constant region (C) of antibody is effector part with various biological functions such as opsonization, complement activation, and antibody-dependent cell-mediated cytotoxicity (ADCC).

7.2 Multigene organization of immunoglobulin genes (Goldsby *et al*., 2000)

The heavy chain and κ or λ light chain of immunoglobulin (Ig) are encoded by separate multigene families located on different chromosomes. Multigene families of heavy chain are located at human chromosome 22. Multigene families of κ chain are located at human chromosome 2 while the multigene families of λ chain are located at human chromosome 14.

In germ-line DNA, multigene families contain several protein-coding gene sequences called "**gene segments**" which are separated by non-coding DNA sequences. During B cell maturation, these gene segments are rearranged (deletion of the non-coding DNA sequences, introns) to form functional immunoglobulin gene.

The κ and λ light chains of Ig-coding DNA sequence contain V (variable), J (joining), and C (constant) gene segments. The genes encoding variable region of both light chains (Vκ or Vλ) contain V and J gene segments. The DNA sequences encoding heavy chain of Ig contain V (variable), D (diversity), J (joining), and C (constant) gene segments. The genes encoding variable region of heavy chain (VH) contain V, D, and J gene segments (**Figure 15A**).

7.3 Variable-region gene rearrangements

Variable-region gene rearrangements occur during B cell maturation in bone marrow. Heavy chain-variable region genes (*VH*) rearranged before light chain-variable region genes (*Vκ* or *Vλ*). After gene rearrangement, each B cell contains one functional *VH* for heavy chain and one functional *VL* for light chain and those mature, immunocompetent B cells are committed to produce antibody with an antigen (epitope)-binding site encoded by particular *VH* and *VL* gene sequences.

7.3.1 *Vκ* and *Vλ* rearrangement

For *Vκ* of human antibodies, any one of the $V_κ$ gene segments can be joined with any one of the functional $J_κ$ gene segments to generate rearranged $V_κJ_κ$ gene. During RNA splicing on primary RNA transcript, V-J gene combination is then joined with C gene segment to generate $V_κJ_κC_κ$ gene which encodes the kappa light chain polypeptide.

For *Vλ* of human antibodies, any of the functional $V_λ$ gene segments can combine with any of the four functional combined $J_λC_λ$ gene segments to generate rearranged $V_λJ_λC_λ$ gene which encodes the lambda light chain polypeptide.

7.3.2 *VH* rearrangement

To generate functional immunoglobulin heavy-chain gene, a D_H gene segment joins to a J_H gene segment to generate a D_HJ_H gene segment (**Figure 15B**). A V gene segment then joins to D_HJ_H gene segment to generate $V_HD_HJ_H$ gene segment (**Figure 15C**). During RNA splicing on primary RNA transcript, $V_HD_HJ_H$ gene segment combination is then joined with C_H gene segment to generate $V_HD_HJ_HC_H$ gene sequence (**Figure15D**) which encodes the heavy chain polypeptide.

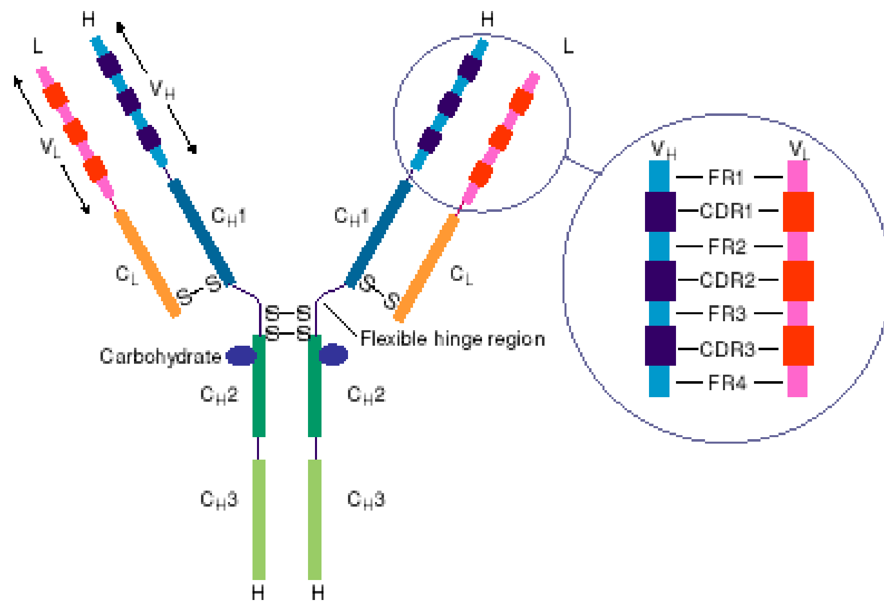


Figure 14 The protein structure of (human IgG_1) immunoglobulin molecules.

Source: journals.cambridge.org/fulltext_content (Accessed on Feb 25, 2008)

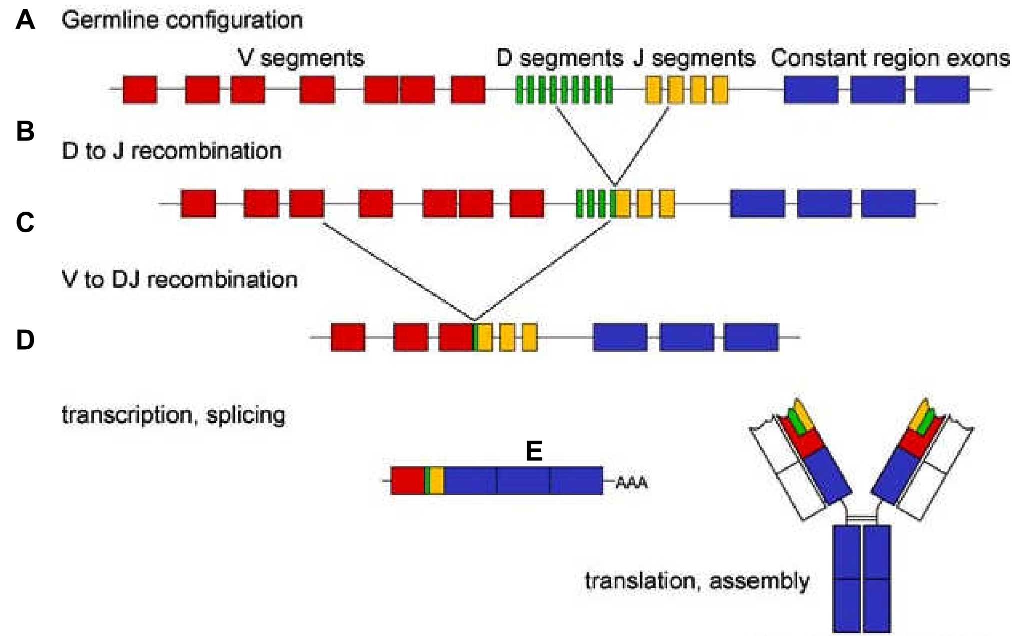


Figure 15 Multigene organization and V(D)J recombination of heavy chain of immunoglobulin gene segments (Janeway, 2001)

7.4 Mechanism of variable-region gene rearrangements (Lewis 1994; Shatz *et al*., 1992)

From DNA sequencing, it was found that there are unique recombination signal sequences (RSSs) which are located at 3' of each V gene segment, at 5' of J gene segment, and at both sides of each D gene segment. The function of RSSs is the signal for rearrangement of antibody gene segments: $V_\kappa J_\kappa$, $J_\lambda C_\lambda$, and $V_H D_H J_H$ gene recombination.

This recombination is catalyzed by enzymes called V(D)J recombinase which is designated RAG-1 and RAG-2. These two recombinases are only lymphoid-specific gene products (Oettinger *et al*., 1990). Sequential steps of antibody gene recombination are as follow:

1. RAG-1 and RAG-2 recognize RSSs at the 5'- and 3' of the gene exons to be rearranged (**Figure 16A**) and the synapsis of gene segments (**Figure 16B**).
2. Single-strand DNA cleavage by RAG-1 and RAG-2 at the junction of the signal sequences (RSSs) and gene segments
3. Hairpin formation in which the free 3'-OH group on the cut DNA strand attacks the phosphodiester bond linking the opposite strand to the signal sequence, and double-strand DNA break at signal sequence by RAG-1 and RAG-2 (**Figure 16B**)
4. Random cleavage of hairpin by endonuclease to generate sites for the addition of P-nucleotides, follow by the trimming of a few nucleotides from the gene segment by a single-strand endonuclease
5. Addition of N-region nucleotides at the end of the V, D, and J gene segments of heavy chain by terminal deoxynucleotidyl transferase (TdT)
6. Repair and ligation to join the gene segments, catalyzed by normal double-strand break repair (DSBR) enzymes

The recombination by these processes generates the coding joint between gene segments such as $V_\kappa J_\kappa$, $V_\lambda J_\lambda$, and $V_H D_H J_H$ coding joints.

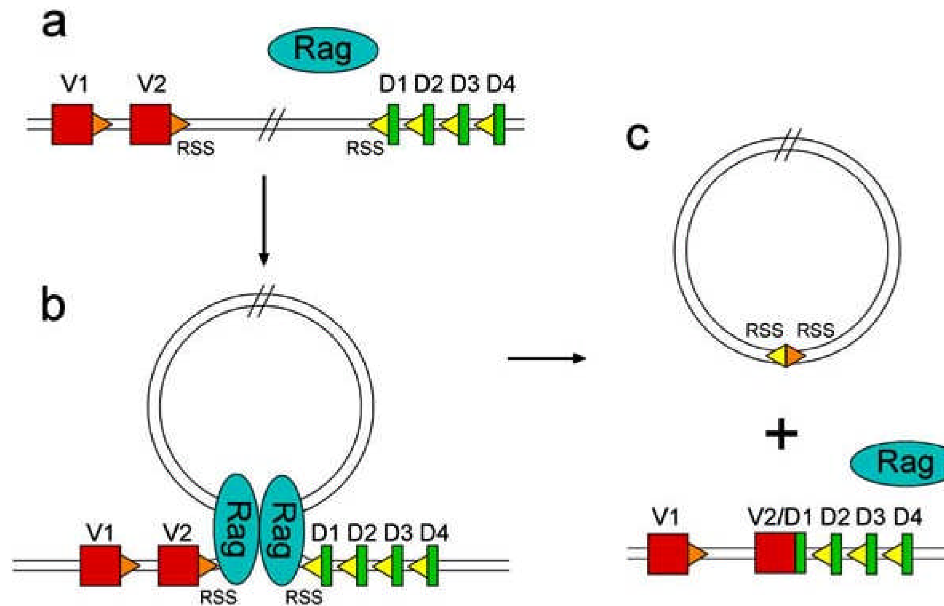


Figure 16 Molecular mechanism of immunoglobulin gene rearrangement (Janeway, 2001)

(**A**) RAG1 and RAG2 recognize the recombination signal sequences (RSSs).

(**B**) The RAG proteins bind to the RSSs of two gene segments (3' of V2 gene segment and 5' of D1 gene segment) and then two RSSs are in close contact to each other by deletion or inversion mechanism. RAG1 and RAG2 cleave the DNA at the junction site of gene segment and RSS.

(**C**) The double-stranded DNA breaks are repaired and re-ligated by DNA repair complexes. P-nucleotide and N-nucleotide addition occur which increase the repertoire of antibody diversity. At the end of process, V2 and D1 gene segments are linked as the coding joint. Two RSSs are also linked as a single joint and then are deleted.

7.5 Generation of antibody diversity

Currently, seven means of antibody diversification have been identified in humans: (1) Multiple germ-line gene segments, (2) combinatorial V-(D)-J joining, (3) junctional flexibility, (4) P-region nucleotide addition (P-addition), (5) N-region nucleotide addition (N-addition), (6) somatic hypermutation, and (7) combinatorial association of light and heavy chains (Goldsby *et al*., 2000).

A. Multiple germ-line gene segments

From DNA sequencing of the immunoglobulin loci of a single individual, human germ-line DNA of David Perry contains 51 V_H, 27 D, 6 J_H, 40 V_κ, 5J_κ, 30 V_λ, and 4 J_λ gene segments (Cook *et al*., 1995; Tomlinson *et al*., 1995; Williams *et al*., 1996).

[In the mouse, there appear to be about 85 V_κ gene segments and 134 V_H gene segments, 4 functional J_H, 4 functional J_κ, 3 functional J_λ, and an estimated 13 D_H gene segments, but only two V_λ gene segments (Goldsby *et al*., 2000)].

B. Combinatorial V-(D)-J joining

Random rearrangement of multiple germ-line gene segments contributes to antibody diversity. Diversity generated by the gene rearrangements in somatic cells can be calculated as shown in **Table 1**. In human, the ability of any of the 51 VH gene segments to combine with any of the 27 DH gene segments and any of the 6 JH gene segments allows a significant amount of heavy-chain gene diversity to be generated (51 x 27 x 6 = 8,262 possible combinations). Similarly, 40 V_κ gene segments randomly combining with 5 J_κ gene segments has potential of generating 200 possible combinations at the κ locus, while 30 V_λ and 4 J_λ gene segments allows up to 120 possible combinations at the human λ locus (Goldsby *et al*., 2000).

Table 1 Combinatorial antibody diversity in humans (Goldsby *et al*., 2000)

Multiple germ-line segments	Heavy chain	Light chain κ	Light chain λ
	Estimated number of segments in humans*		
V	51	40	30
D	27	0	0
J	6	5	4
Combinatorial V-D-J and V-J joining	51 x 27 x 6 = 8,262	40 x 5 = 200	30 x 4 = 120
Possible combinatorial association of heavy and light chains	8,262 x (200 +120) = 2.64 x 10^6		

* These numbers were investigated from only single subject, *i.e.*, David Perry. Only the functional gene segments have been listed. Some gene segments contain stop codons or are not able to rearrange.

C. Junctional flexibility

During recombination of the immunoglobulin gene segments, the joining of recombination signal sequences to form a coding joint is always precise but the joining of coding sequences to form a signal joint often is imprecise which increase the antibody diversity. Some of junctional flexibility may leads to non-productive rearrangements. Several productive combinations that encode different amino acids at each coding joint are fall within the CDR3 in immunoglobulin heavy-chain and light-chain DNA sequences (Lewis, 1994; Shatz *et al*., 1992).

D. P-region nucleotide addition (P-addition)

After random cleavage of hairpin by endonuclease, subsequent addition of complementary (palindromic or P) nucleotides (P-addition) to the trimmed hairpin by repair enzymes generates a palindromic sequence in the coding joint. Variation in the position at which the hairpin is cut, thus lead to the nucleotide sequence variation in the coding joint (Lewis, 1994; Shatz *et al*., 1992).

E. N-region nucleotide addition (N-addition)

Addition of N-region nucleotides (up to 15 amino acids) to both the D_H-J_H and V_H-D_HJ_H joints by terminal deoxynucleotidyl transferase (TdT) generates extra amino acids that are not encoded by the germ-line V, D, or J gene segments. N-addition is localized in CDR3 of heavy chain genes (Lewis, 1994; Shatz *et al*., 1992).

F. Somatic hypermutation (Tonegawa, 1983)

Somatic hypermutation occurs in B cells within germinal centers that form in the secondary lymphoid tissues within a week after the primary immunization. Target of somatic hypermutation is the rearranged VJ or VDJ gene segments. Somatic hypermutation occurs at a frequency 10^{-3} per base pair per generation. Most of the mutations are nucleotide substitution rather than deletion or insertion and this process generates antibodies with varying affinity for antigen. Following exposure to antigen, those B cells with higher-affinity receptors will be preferentially selected in the secondary lymphoid tissue for survival because of their greater ability to bind to the antigen on the icosomes of the follicular dendritic cells. This process is called "affinity maturation" and it occurrs within the germinal center.

G. Combinatorial association of light and heavy chains

Antigen-binding site of human and most animals is composed of the variable regions of both heavy and light chains. Combinational association of heavy and light chains can generate more antibody diversity as estimated in **Table 1**.

Antibody diversity could reach to 2.64 x 10^6 as estimated from combinatorial V-D-J and V-J joining and possible combinatorial association of heavy and light chains (**Table 1**). Moreover, the diversity can be contributed by other mechanisms such as junction flexibility, P-region nucleotide addition, N-region nucleotide addition, and some somatic hypermutation, the potential of diversity can exceed this estimation ($>10^6$). Thus, in theory a human may have an antibody repertoire of ~2.6 x 10^6 x 10^6. However, it has been estimated that during the life time approximately 10^7 of antibody diversity are produced (Goldsby *et al*., 2000).

8. Therapeutic antibody

In 1890, Emil von Behring and Shibasaburo Kitasato discovered that sera of the tetanus or diphtheria toxoid-immunized animals had ability to neutralize the toxin activity. The successful of serum therapy using antitoxin prepared from immunized animals resulting in the first Nobel Prize Award in 1901 to von Behring (cited by Roitt, 1998). Because of the heterologous nature of animal immune sera, the side effect *i.e.*, hypersensitivity in the form of anaphylaxis and/or serum sickness could be provoked. To overcome this problem, immunoglobulins were purified from the contaminated non-immunoglobulin serum proteins by using alcohol precipitation. The process is called “Cohn fractionation” (Cohn *et al.,* 1946).

In 1975, the production of murine monoclonal antibodies (MAbs) by hybridoma technology was first reported (Kohler and Milstein, 1975) and then in 1980, treatment of disease using MAbs had been studied in human. Because of the short serum half-life and the human anti-mouse antibody (HAMA) response (anti-mouse isotypes in the human recipient), the chimeric and humanized-MAbs, first reported in 1984 (Morrison *et al.,* 1984) and 1986 (Jones *et al.,* 1986), respectively, were developed (please see **Section 9.2**). The chimeric MAbs still can provoke HAMA in the recipients because they still contain mouse protein at variable regions of the chimeric antibodies. By using recombinant DNA technology, the therapeutic MAbs of murine origin was

humanized by grafting only the complementarity-determining regions (CDRs) of murine MAbs to human variable region frameworks (FR) of immunoglobulin molecules, so-called “antibody engineering” technology.

8.1 Murine monoclonal antibody production

In 1975, Georges Köhler and Cesar Milstein generated hybridoma which is a hybrid cell of an antibody-producing B cell and a myeloma cell (a cancerous plasma cell). Hybridoma has the immortal-growth properties of the myeloma cell and also can secrete the antibody as the parental B cell (Kohler and Milstein, 1975; Winter and Milstein, 1990).

8.1.1 Formation and selection of hybrid cell that secretes antibody

B-cell hybridomas are produced by using a cell fusogen such as polyethylene glycol to fuse myeloma cells with B cells from antigen-immunized animals, usually inbred mice (Chaicumpa *et al*., 1994). By using polyethylene glycol, a complex mixture of unfused myeloma and spleen cells (“parental cells”) as well as numbers of fused cells was generated. Among the fused cells, there are three different combinations; self fusions of B-cells, or of myeloma cells, and fusion of B cells and myeloma cells. To allow the survival and growth of only the B-cell x myeloma hybridomas, one method is to use myeloma cells that are unable to grow in a selective HAT medium (hypoxanthine, aminopterin, and thymidine medium) because they are deficient for one of the nucleotide salvage pathway. The parental myeloma cells cannot survive in HAT medium. The B-cell x myeloma cell hybrid can survive because the B cell provides the necessary enzyme for the salvage pathway which the myeloma cells lack. Although unfused B cells are able to survive in HAT medium, they cannot grow and survive *in vitro* for a long time and eventually die off in the prolonged cell culture (**Figure 17**).

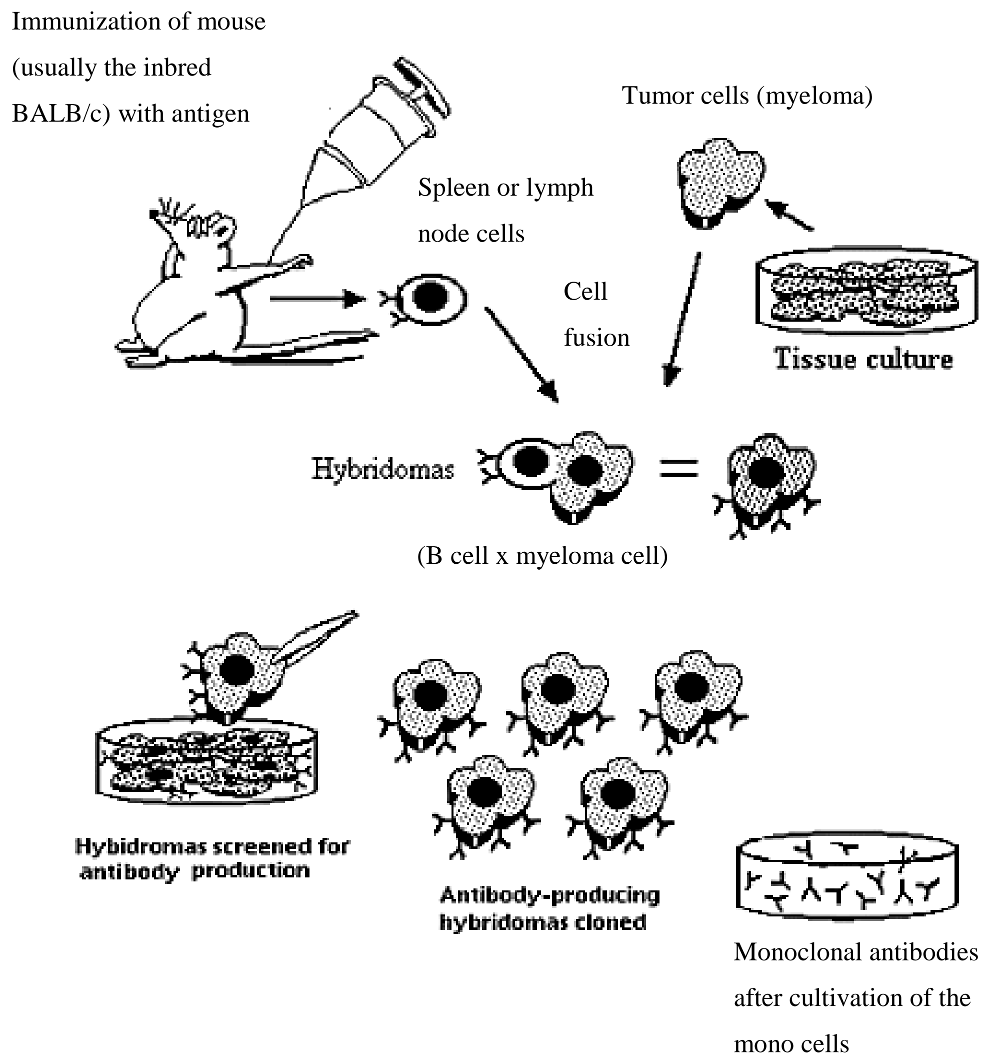


Figure 17 Sequential steps in mouse monoclonal antibody production. (adapted from www.accessexcellence.org/AB/GG/monoclonal.html) Accessed on Feb 25, 2008

8.1.2 Production of monoclonal antibodies

Antibody-secreting hybridomas are screened for the antigenic specificity by using several techniques such as enzyme linked-immunosorbent assay (ELISA), immunofluorescent antibody assay, etc. To generate truly monoclonal, hybridomas secreting antibody of an antigen specificity are re-cloned and are then propagated in tissue-culture flasks. By using *in vitro* culture technique, low concentrations (1-20 μg/ml) of antibody can be obtained. Much higher concentration (1-10 mg/ml) of antibody can be also obtained by propagating hybridoma in the peritoneal cavity of the pristane treated-histocompatible mice, where the monoclonal antibody is secreted into the ascitic fluid (Hay and Westwood, 2002).

8.2 Chimeric monoclonal antibody (Hayden, 1997)

The mouse MAbs are recognized as foreign when given to human and it can evoke anti-isotype antibody response so-called "human anti-mouse antibody" or HAMA (Losman, 1990). The induced HAMA can reduce the effectiveness of the MAb by forming immune complex with the MAb and being cleared from the bloodstream (Wong, 2000; Meeker *et al*., 1985; Pimm, 1985). In addition, circulating immune complexes of mouse MAb and human anti-mouse isotype antibodies can cause allergic reactions (Dillman *et al*., 1984; Meeker *et al*., 1985). In some cases, the deposition of immune complexes in organs such as the kidney can cause serious and even life-threatening reactions such as type III hypersensitivity or serum sickness (Losman, 1990). To avoid the undesirable reactions, using human monoclonal antibodies for clinical applications is desired. The human monoclonal antibodies have been developed but there are numerous technical problems as described in **Section 9.3**. Because of the difficulty of producing human monoclonal antibodies and the complications resulting from use of mouse monoclonal antibodies in humans, the recombinant DNA containing the promoter, leader, and variable-region sequences from a mouse antibody gene and the constant-region exons from a human antibody gene was constructed. The antibody encoded by such a recombinant gene segment is a mouse-human chimera, commonly known as a chimeric antibody (**Figure 18**). Its antigen binding specificity is derived from the mouse DNA; its isotype, which is determined by the constant regions are encoded by human genes. The chimeric

antibody has fewer mouse antigenic determinants (only in the immunoglobulin frameworks of the mouse VH and VL regions) and is far less immunogenic than the original mouse monoclonal antibodies when administered to human. Another advantage of all the humanized-chimeric antibodies is that they retain the biological effector functions of human antibody and are more likely to trigger human complement activation or Fc receptor binding (Opsonic activity).

The potential for the mouse variable region remaining in the chimeric antibody to induce a harmful antibody response in humans has encouraged further exploration of this approach. Only the sequences of the complementarily determining regions (CDRs) of mouse origin are grafted onto human antibody gene sequences encoding the immunoglobulin frameworks. The molecule is called humanized-antibody.

8.3 Human monoclonal antibody (Steinitz *et al.*, 1977)

Later, human monoclonal antibodies have been produced from Epstein-Barr virus-(EBV) transformed-B cells. This lymphotropic herpesvirus can converts normal human B cells into established lines which have immortal property. Memory B cells ($CD22^{+}$ IgM^{-}, IgD^{-}, IgA^{-}) isolated from peripheral blood mononuclear cells (PBMCs) of an individual recovering from infectious diseases or individual immunized with an immunogen/vaccine have been immortalized by EBV in the presence of a CpG oligodeoxynucleotide which are Toll-like receptor 4 (TLR4) ligands that increase transformation efficiency and irradiated allogeneic PBMCs (Steinitz *et al*., 1977). The EBV-immortalized B cells secreting antigen-specific antibody in culture supernatant were screened by using techniques as applied in mouse MAbs screening. The cells in the tissue cultures containing antigen-specific antibody are further cloned by limiting dilution technique (Chaicumpa *et al*., 1988). Human MAbs can be produced from the cloned B cells. Disadvantages of EBV-tranformed B cells are (i) difficulty in establishing antibody-secreting B cell line, (ii) possible contamination of EBV during antibody production, (iii) safety of the laboratory personnel, and (iv) ethical concern in using of human cells.

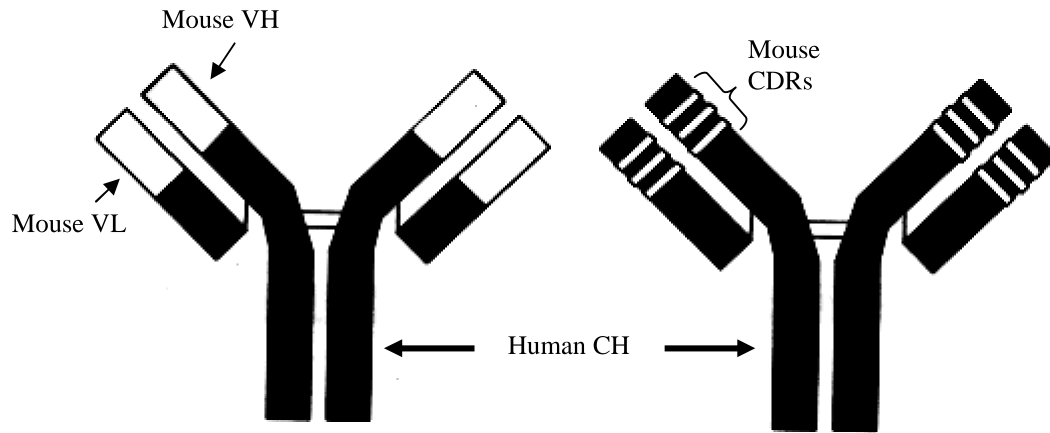


Figure 18 Chimeric antibody and CDR grafting of humanized-MAb

The chimeric antibody (at the left of the figure) contains the mouse heavy and light chain variable regions (white) which are attached to human constant regions (black). In humanized antibody (at the right of the figure), only CDRs of mouse MAb were grafted onto the frameworks and constant regions of the human antibody molecule.

(adapted from www.users.path.ox.ac.uk) Accessed on Feb 25, 2008

9. Phage display technology

In 1985, phage display technology was firstly introduced by G. Smith who used M13 filamentous bacteriophage to display *Eco*RI endonuclease on its surface. Phage display have been used to express proteins, including antibodies, or peptides on the surface of bacteriophages (or simply called phages) of *E. coli, e.g*., M13. Protein-coding DNA sequences have been inserted into the phage genome adjacent to phage coat protein-coding gene such that the proteins are expressed (displayed) on surface of a phage particle as a fusion protein with the phage coat protein, especially gene 3 protein (g3p) or protein 3 (pIII) of the M13 phage (Smith, 1985). By recombinant DNA technology, a large repertoire of antibody, estimated about 10^{11}-10^{12} diversity in human's body, can be fused to g3p and subsequently displayed on surfaces of phage particles, so-called "**human antibody phage display library**" (for extensive review, please see Azzazy and Highsmith, 2002). Antigens of interest can be used to select antigen-specific phages from the antibody repertoire *in vitro* in a process called "**bio-panning**".

9.1 Filamentous bacteriophage biology

The filamentous bacteriophages (genus *Inovirus*) are a group of bacterial viruses that contain a circular single-stranded DNA genome encased in a long coat protein as cylindrical shape. The Ff class of filamentous bacteriophages (f1, fd, and M13) have been extensively studied. These bacteriophages use the tip of the F conjugative pilus as a receptor and thus are specific for *Escherichia coli* bacteria containing the F plasmid. The DNA sequence of these three phages shows 98% identities (reviewed by Webster, 2001).

9.1.1 The phage genome

The genome of the Ff phages (fd and M13) encode 11 genes, *i.e.,* gene *I* to gene *XI* (**Table 2** and **Figure 19).** The phage genome is about 6,400 nucleotides in size. The gene products of gene *X* and gene *XI* (pX and pXI) are the result of a translational start at an internal methionine codon in genes II and I, respectively (Model and Russel, 1988; Guy-Caffey *et al*., 1992; Rapoza and Webster, 1995). The eleven genes are grouped into three groups according to their functions. The first group (genes *II*, *V*, and *X*) encodes the proteins required for the replication of the

phage genome. The second group (genes *VII*, *IX*, *VIII*, *III*, and *VI*) and the third group (genes *I*, *XI*, and *IV*) are involved in the host membrane-associated assembly of the bacteriophages. There is a short DNA sequence called the intergenic region (IG) that dose not encodes for protein. The IG (opened block, **Figure 19**) contains the sites of origin for viral (+) and complementary (-) DNA sequence synthesis. The 78-nucleotide sequence called the packaging signal (PS) is also found in the IG located near the end of gene *IV*.

9.1.2 Structure of Ff bacteriophage (extensively reviewed elsewhere by Webster, 2001)

The Ff filamentous bacteriophage is approximately 6.5 nm in diameter and 930 nm in length (**Figure 20**) (Webster, 2001). The mass of the particle is approximately 16.3 MDa, of which 87% is contributed by proteins. The Ff phages are consisted of a single-strand (ss) DNA enclosed in a cylindrical protein coat. Phage expresses about 2,700 copies of gene 8 protein (g8p or pVIII) as the major coat protein. There are minor coat proteins located at the two ends of each phage particle. One end of a phage particle contains three to five molecules each of gene III protein (g3p or pIII) and gene VI protein (g6p or pVI), accounting for 10-16 nm of phage length (**Figure 20**). Another end contains approximately five molecules each of the hydrophobic g7p (pVII) and the g9p (pIX). The packaging signal with 78-nucleotide hairpin structure is always located at this end of the particle containing the pVII and pIX (**Figure 20**).

The carboxyl terminal 10-13 residues of pVIII form the inside wall of the cylinder. This region contains 4 positively charged lysines that reside on one face of an amphiphilic helix. These positive charges interact with the sugar phosphate backbone of the DNA that is present in the particle with the base pointed inward (Greenwood *et al.*, 1991; Marvin *et al.*, 1994). The amino-terminal portion of pVIII is present on the outside of the particle. The residues connecting the amino and carboxyl regions of pVIII interact with the same region of other pVIII molecules to form the stable inner core of the protein cylinder. Most of this middle portion of pVIII spans the cytoplasmic membrane before being assembled into phage particles.

Table 2 Genes and gene products of the fl bacteriophage (Webster, 2001)

Gene	Function	No. of amino acids	Protein MW (Da)
II	DNA replication	410	46,137
X	DNA replication	111	12,672
V	Binding ssDNA	87	9,682
VIII	Major capsid protein	50	5,235
III	Minor capsid protein	406	42,522
VI	Minor capsid protein	112	12,342
VII	Minor capsid protein	33	3,599
IX	Minor capsid protein	32	3,650
I	Assembly	348	39,502
IV	Assembly	405	43,476
XI	Assembly	108	12,424

The number of amino acids and the molecular weight are for the mature proteins. The initiating methionine is included in proteins that do not contain an amino-terminal sequence.

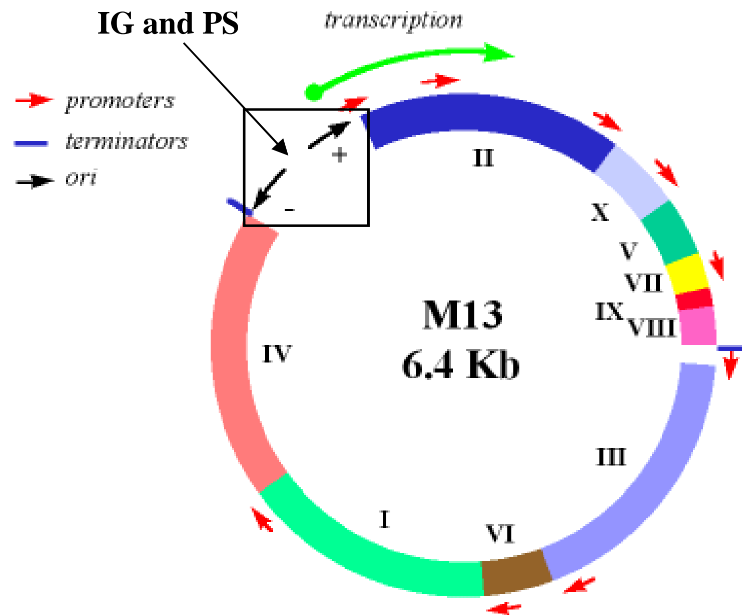


Figure 19 Gene organization of filamentous bacteriophage M13 genome.

Source: utminers.utep.edu/rwebb/html/m13_phage__intro

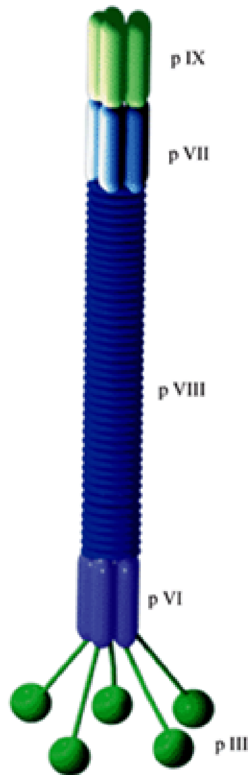


Figure 20 Schematic diagram of a filamentous bacteriophage M13. The phage consists of circular ssDNA surrounded by a major coat protein called pVIII whereas pIII, pVI, pVII, and pIX, at the tips of the phage, are the minor coat proteins.
Source: www.scielo.br/scielo.php?pid=S1415-4757200500

The pIII is made up of three domains which are designated as N1 or D1, N2 or D2, and CT or D3 and are separated by glycine-rich regions (Stengle *et al*, 1990). The N1 domain contains the amino-terminal 68 amino acids and is required during *E. coli* infection for the translocation of the DNA into the cytoplasm and the insertion of the coat proteins into the membrane. The N2 domain, 87-127 amino acid residues, is responsible for the F pilus binding (Deng *et al.*, 1999). The N1 and N2 domain are linked by intramolecular disulfide bonds and are exposed on the surface of the phage particle. Protease treatment of these two domains causes non-infectious phage (Gray *et al.*, 1981; Armstrong *et al.*, 1981).

The carboxyl terminal 150 residues make up the third domain of pIII called CT or D3 and are necessary for forming a stable phage particle (Crissman and Smith, 1984; Kremser and Rasched, 1994). It has been proposed that part or all of the CT domain, together with pVI, interact with pVIII to form one end of the particle (Rokonjac *et al.*, 1999). It has been proposed that the hydrophobic amino terminus of pVI is buried within the particle (Makowski, 1992).

9.1.3 Life cycle of filamentous bacteriophage (Webster, 2001)

The filamentous bacteriophages infect the F pilus of a male *E. coli* (*e.g.,* TG1 *E. coli*) by the attachment of the phage pIII to F pilus. Only the circular phage ssDNA is translocated into bacterial cells. The host DNA machinery converts ssDNA into the double-stranded plasmid like replicative form (RF). The RF then undergoes rolling circle replication to make ssDNA and also serves as a template for the pIII and pVIII expression. The coat proteins are all synthesized as integral membrane proteins that remain in the membrane until they are assembled around ssDNA. Assembly process occurs at specific sites in the bacterial envelope where the cytoplasmic and outer membranes of the bacterium are in close contact. During the assembly process, the viral ssDNA is extruded through the membrane-associated assembly site, where the phage ssDNA-binding proteins are removed and the coat proteins are packaged around the ssDNA. New complete phage progenies are extruded through the bacterial membrane without cell lysis into the bacterial culture medium. The bacterium can tolerate this process and continue to grow and divide with a generation time approximately 50% longer than that of the uninfected bacteria. There is a burst of about 1,000 phage particles produced in the first generation after infection, and then

the bacteria produce about 100-200 particles per generation resulting in phage titers of 10^{11} to 10^{12} particles per ml of the bacterial culture medium.

9.2 Human antibody phage display library

By using phage display technology, antibody-encoding DNA sequences are fused to phage coat protein-coding gene. Antibody molecules can be expressed or "displayed" on the surface of filamentous phage as a fusion protein with the phage coat protein.

The successful in the insertion of antibody gene into filamentous phage gene III protein (g3p) by Smith (Smith, 1985) and in the expression of functional antibody fragments in the periplasm of *E. coli* (Better, 1988; Skerra, 1988) are essential for the development of antibody phage display technology. The linkage between the phenotype (antibody-displaying phages) and the genotype (DNA sequences encoding antibody molecules) in the same virion permit the selection and amplification of antigen specific phage clones from pools of billions of phage clones in "bio-panning" process. Phage amplification can be simply done by infecting F-pili-expressing *E. coli*. The amino acid sequence of antigen-binding antibody can be revealed by DNA sequencing of gene insert in the phage genome.

9.2.1 Diversity of antigen binding sites

An million to billion number of antigen-binding site (paratope) generated by the N-terminal domains of the heavy and light chains is produced by immune system. Several mechanisms are responsible for the diversity of paratopes (Padlan, 1994): (i) a recombinatorial diversity: created by random selection of one variable (VH), one diversity (D) and one joining (JH) gene segment of heavy chain, or one variable (VL) and one joining (JL) gene segment of light chain from several gene segments in genomic DNA during gene rearrangement, to constitute the VH and VL domains, respectively; (ii) a junctional diversity: added by the imprecise joining mechanisms and by deletion or addition of random nucleotides at the borders of the recombining VH-D-JH gene segment; and (iii) a combinatorial diversity: generated by the assembly of the VH and VL domains. Other mechanisms include: (i) changing of the paratope architecture by adjusting the angle between the associated VH and VL domains; and (ii) a somatic hypermutation during B cell activation and proliferation

that the CDRs of the antibody can be improved. The antigenic affinity and specificity of antibody are determined by those combined mechanisms.

9.2.2 Antibody fragments for displaying on phage surface (Azzazy and Highsmith, 2002)

Antibody molecules contain discrete protein domains that can be separated by protease digestion or produced by recombinant DNA technology. Two antigen-binding fragments designated Fab or antibody fragment and Fv or variable fragment have been displayed on phage surface. The Fab is larger than Fv and consists of VH-CH and VL-CL segments linked by disulfide bonds. The smaller Fv is composed of only the VL and VH segments. By using recombinant DNA technology, the VH and VL-coding DNA sequences are linked via a flexible peptide linker-coding gene to generate single chain of DNA sequences which encode Fv called the single-chain variable fragment (scFv). The length of peptide linker is usually 5-15 aa linker and its sequence is four glycine and one serine (Gly4Ser3). This flexible linker allows the formation of antigen-binding site between VH and VL. The intact immunoglobulin molecule and several monomeric antibody fragments are presented in **Figure 21**.

9.2.3 Phage and phagemid cloning vector (Scott, 2001)

Filamentous bacteriophage genome especially M13 has been initially used to display protein molecules on phage surface as a fusion protein with phage coat protein (s) such as pIII and pVIII. Because phage genome contains only one copy of gene *III* and gene VIII, all pIII or pVIII molecules on phage surface display the two proteins. Foreign molecule-displaying pIII at N-terminal of protein can decrease infectivity of phage during phage propagation. The advantages of using phage genome are that no phage rescue step is needed (as contrast to the phagemids; please see below). Nevertheless, because of the large size of phage genome, the disadvantages of using phage genome are the difficulty in the construction of recombinant phage genome and the introduction of recombinant phage genome into *E. coli* host cell. Thus, the use of phagemids is more convenient.

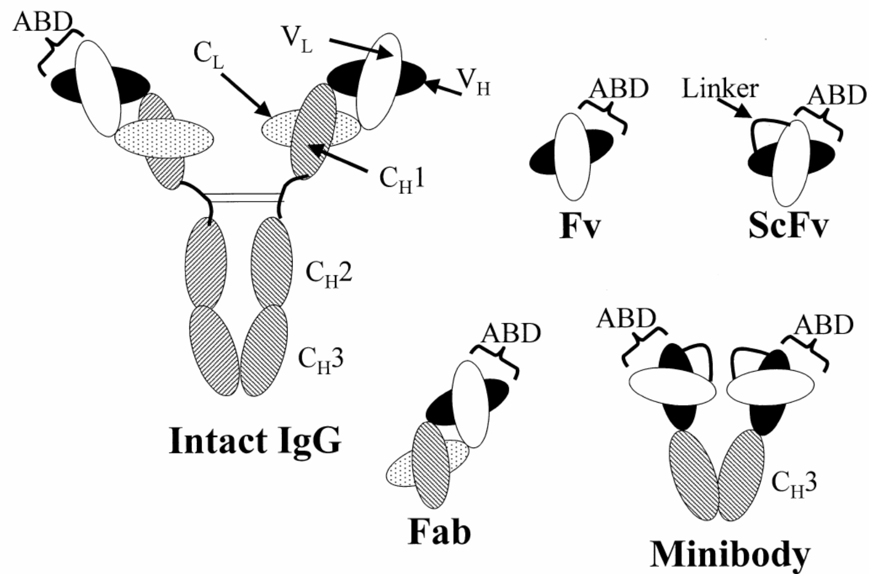


Figure 21 Structure of intact immunoglobulin G (IgG) molecule, a minibody, and monomeric antibody fragments: Fab (antibody fragment), Fv (variable fragment), and scFv (single chain variable fragment). Variable domains of heavy (VH) and light (VL) chains are represented by black and white ovals, respectively. Constant regions of heavy (CH1-3) and light (CL) chains are represented by shaded and dotted ovals, respectively. Antigen-binding domains (paratope) are shown as ABD.

Source: Azzazy and Highsmith, 2002

Phagemid is a hybrid of phage genome and plasmid vectors. Phagemid vector contains the origin of replication for the M13 phage and *E. coli*, gene *III*, cloning sites for antibody fragment-encoding genes, and an antibiotic-resistance gene (Mead and Kemper, 1988). In contrast to the M13 phage genomes, phagemid vector lacks other structural and non-structural protein-encoding genes required for generating a new progeny phage particle. However, phagemids can be packaged into phage coat protein as the wild-type M13 phage with the help of genetically modified phage strain that has a slightly defective origin of replication. This type of genetically modified phage is called "**helper phage**" such as M13KO7 or VCSM13. Helper phage supplies, *in trans*, all the structural proteins for generating a new progeny phage. This process is called "**phage rescue**". The new progeny phage particles incorporate either pIII derived from the helper phage or antibody fragment-pIII fusion protein derived from the recombinant phagemid. The proportion of antibody fragment-fused pIII: wild type pIII may range between 1:9 and 1:1,000 depending on the type of phagemid, growth conditions, the nature of the protein fused to pIII, and proteolytic cleavage of antibody-pIII fusions.

9.2.4 Construction of ScFv phage display library (Nissim *et al.,* 1994 and Vaughan *et al*., 1996)

The variable region of heavy (VH) and light (VL) chain-encoding DNA sequences are amplified from cDNA that are reverse transcribed from mRNA of B lymphocytes. There are several sources of B lymphocytes such as pheripheral blood, lymph node, and spleen. In the antibody gene amplification, because of the large variety of antibody genes, either degenerate or non-degenerate primers are designed to cover all antibody genes. The VH-encoding gene (*VH*) and VL-encoding gene (*VL*) amplicons are linked into a single DNA sequence called ScFv gene sequence (*scFv*) by using an oligonucleotide linker encoding short flexible peptide and dose not interfere the folding of ScFv molecule. The *scFv* are then inserted into phagemid vector. The recombinant *scFv*-phagemid vectors are introduced into competent *E. coli* by chemical transformation or electroporation. The recombinant phagemid-containing bacterial cells are then super-infected with helper phages to rescue the new progeny phages that display ScFv molecules as fusion protein to phage coat protein (**Figure 22**). The details of the steps in the construction of an immune ScFv library by using

phage display are shown in **Figure 23** (for more extensive review, please see Azzazy and Highsmith, 2002).

9.2.5 Types of ScFv phage display library

A. Non-immune library: "single pot libraries"

In non-immune library, VH and VL-encoded genes from the mRNA of B-cells of unimmunized human donors are isolated from peripheral blood lymphocytes, bone marrow, spleen cells, or from animal sources. Small-sized human single-pot libraries (3 x 10^7 antibody clones) had been constructed and used to isolate phage clones displaying antibodies to antigens, *e.g.,* bovine serum albumin and lysozyme, haptens (2-phenyloxazol-5-one), or "self" antigens, thyroglobulin, tumor necrosis factor (Nissim *et al.,* 1994). The affinity (*Ka)* of the ScFv displayed on the phages was similar to antibodies in primary immune response and could interact with particular antigen in Western blot analysis, ELISA, and FACS analysis. A large sized naïve library (1.4 x 10^{10} clones) was constructed by Vaughan *et al.* using lymphocytes from over 40 non-immunized human donors (Vaughan *et al*., 1996). The affinity of antibodies isolated from this library was in the low nanomolar range typically found in the secondary immune response. By using this kind of a single-pot library, human antibodies with high affinity could be generated without immunization. The single-pot naïve repertoires has several advantages include: (i) it can be used to isolate human antibodies to self, non-immunogenic, or toxic antigens; (ii) a single library can be used for all antigens; (iii) short time needed for antibody generation [2-4 rounds of selection (bio-panning) in two weeks]; and (iv) direct isolation of high affinity antibodies when the large repertoires are used. Nevertheless, naive libraries have disadvantages which are: (i) isolation of antibodies with low affinity from small sized libraries; (ii) long time consumption for large library construction, and (iii) content and quality of the library are influenced by the unequal expression of the V-genes repertoire, unknown history of the B-cell donors, and the diversity of antibody is limited in case of using the IgM mRNA repertoire.

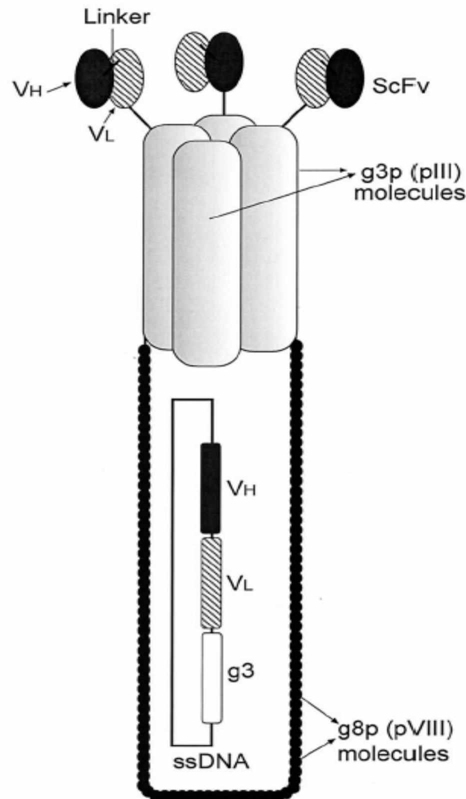


Figure 22 Schematic diagram of a ScFv-displaying filamentous bacteriophage M13. The phage consists of circular ssDNA surrounded by a major coat protein called pVIII. At the tip of phage, pIII is one of minor coat proteins. The scFv genes are ligated adjacent to gene *III* (g3) in the genome of phage. The ScFv is displayed as a fusion protein to pIII at the tip of phage but not all pIII are the ScFv-fused protein. The phage still has ability to infect bacterial host cell because there is/are still free pIII.
Source: Azzazy and Highsmith, 2002

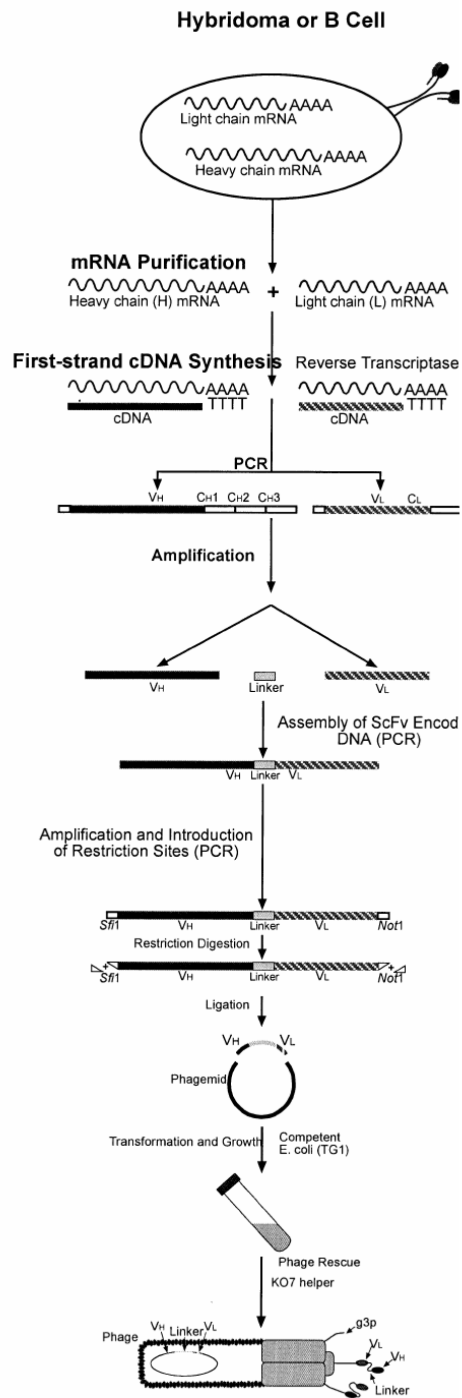


Figure 23 Schematic diagrams for constructing ScFv repertoire by using phage display technology (Azzazy and Highsmith, 2002).

B. Immune libraries

1) An immune library with increased antibody affinities can be constructed by several methods. VH and VL-encoded DNA sequences of B-lymphocytes obtained from immunized animals are used to construct the antibody phage library. High-affinity antibodies were generated from immunized mice (Andersen *et al*., 1996), chickens (Yamanaka *et al*., 1996), and rabbits (Lang *et al*., 1996).

2) Human immune repertoires can also be prepared using immunized transgenic mice (xenomice). Xenomice harbor human V-genes in the germline (instead of the native murine immune repertoire) (Yang *et al*., 1999). Immunization of xenomice with a hapten-carrier conjugate or a foreign antigen results in the production of human-like antibodies by their B-cells.

There are two main characteristics of immune library: (i) library is enriched in antigen-specific antibodies, and (ii) *in vivo* affinity maturation occurs in this library (Skerra and Pluckthun, 1988 and Clackson *et al*., 1991). From literature review, it was found that immune libraries were used to produce antibodies against carcino-embryonic antigen (CEA) (Chester *et al*., 1994), T-cell receptor-V alpha(s) (Popov *et al*., 1996) and major histocompatibility complex/peptide complexes (Andersen *et al*., 1996).

Disadvantages of immune libraries include:

(i) animal immunization process required long time

(ii) lack of immune response to self or toxic antigens

(iii) the unpredictability of the immune response to the antigen of interest

(iv) a new antibody library must be constructed for each antigen

(v) generating antibodies from immunized human is difficult because of the ethical concern.

Advantages of immune library include:

(i) High affinity antibodies could be generated from human after recovering from viral infections or cancer and antibodies to self-antigens could be generated from patients with autoimmune diseases

(ii) immune libraries was used for generating antibodies against minor or poorly immunogenic antigens by removal of irrelevant antibodies. Animals can be made tolerant to particular antigens and then were immunized with a mixture of the relevant antigen and the irrelevant ones. This approach was used to isolate human anti-melanoma antibodies from immune phage libraries of cancer patients who were immunized with autologous tumor cells (Cai and Garen, 1995).

C. Synthetic library

It is known that the three CDRs (CDR1, CDR2, and CDR3) in variable region (both VH and VL) of any antibodies determine the binding affinity and specificity of antibody. Among the six CDRs, CDR3 of the heavy chain (VH-CDR3) is the most important in antigen-binding site of all CDRs and is the most diverse site in term of composition and length of amino acid residues. It has been estimated that VH-CDR3 has a potential diversity of 10^{23} sequences (Chothia and Lesk, 1987). Oligonucleotide mutagenesis or PCR-based strategy has been used to randomly synthesize CDR3 of either VH or VL or both (Akamatsu *et al.*, 1993). Moreover, all CDRs loops in one V-gene segment can be diversified and were applied to construct synthetic library (Garrard and Henner, 1993).

9.2.6 Selection of antigen-specific antibodies from library: "bio-panning"

To select and enrich the antigen-specific phages from the library, the library (phage clones displaying antibodies) is incubated with an antigen of interest by using several methods (please see below). The earliest method is to use "immobilized antigen" as has been done in this thesis (please see **Section 8** of **Chapter IV**). Unbound phages are then removed by extensively washing in high stringency condition. Antigen-specific phages which bound specifically to the immobilized antigen are eluted and re-amplified by infecting in bacterial host cells (**Figure 24**). The re-amplified phage is used to bind with immobilized antigen in the next round. The repeated process is called "bio-panning".

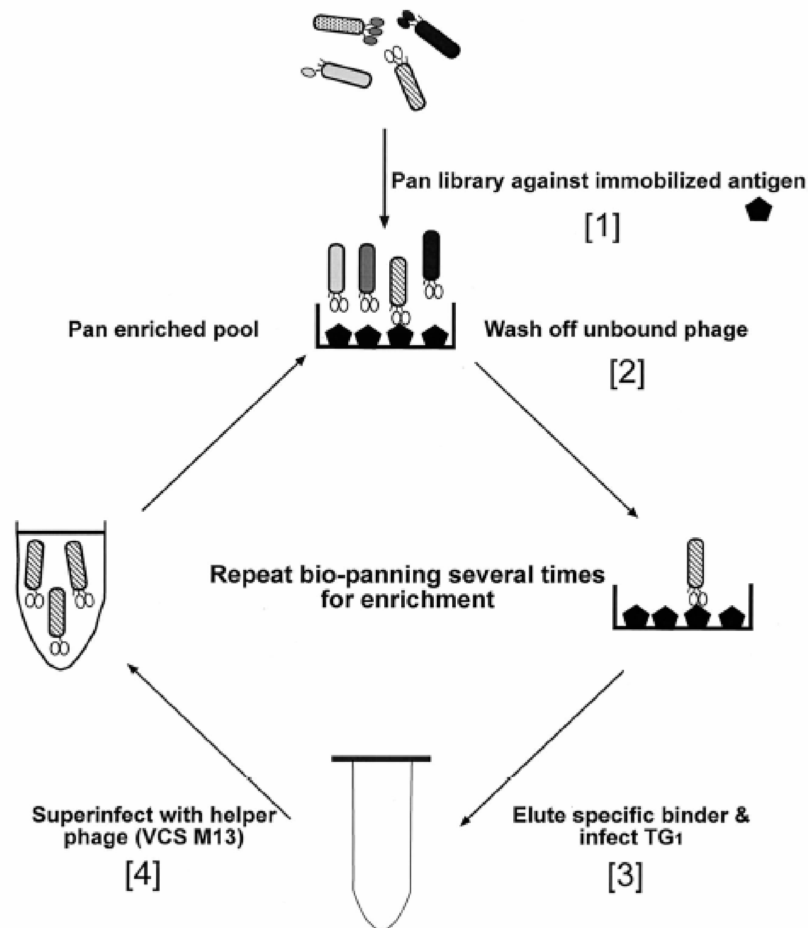


Figure 24 Schematic process of phage bio-panning against immobilized antigen (as used in this thesis). There are four steps in the bio-panning: (**1**) binding of antibody-displaying phages to the immobilized antigen on a solid phase; (**2**) washing to remove unbound phages; (**3**) disruption of antigen-antibody interaction and recovering antigen-specific phages by infecting TG1 *E. coli*; and (**4**) amplification of the antigen-specific phages by superinfection of the *E. coli* with helper phage. The pool of new progeny phage particles can be used in next round of bio-panning or they can be directly cloned and selected for the target-specific clones (one round of bio-panning)
Source: Azzazy and Highsmith, 2002

There are four types of bio-panning:

A. Bio-panning with immobilized antigens

An affinity column (Clack *et al.,* 1991; McCafferty *et al.,* 1990), immunotubes, ELISA plates (Marks *et al.,* 1991; Kang *et al.,* 1991), or BIAcore sensor chips (Malmborg *et al.,* 1996) with immobilized antigens were used in the phage bio-panning. The antibody library was applied and the antigen-specific phages were eluted from the immobilized antigen by using either acidic solutions (Kang *et al.,* 1991; Roberts *et al.,* 1992), basic solutions such as triethylamine (Marks *et al.,* 1991), by enzymatic cleavage of a protease sensitive site located between the antibody and pIII of the phage coat protein (Ward *et al.,* 1996), or by competition with excess antigen (Clackson *et al.,* 1991).

B. Bio-panning with antigens in solution

Antigen can be labeled with a ligand, *e.g.*, biotin. After incubating the phages with (biotin)-labeled-antigen in solution, the avidin or streptavidin-coated paramagnetic beads are then used to recover only bound phages from the solution. Bio-panning with antigens in solution was advantageous in selecting high-affinity antibody-displaying phages (Hawkins *et al.,* 1996).

C. Bio-panning with antigen on cells

The target molecules on cell surfaces, either monolayers of adherent cells or cells in suspension, can be used to select specific phages. To select only specific phages that bound to the target antigen on cell surface, the excess antigen-negative cells which are used as "absorber" to exclude other surface marker-specific phages, were incubated together with small number of antigen-positive cells (target cells). This process is called a negative and positive selection (de Kruif *et al.,* 1995). The target cells can be isolated from the absorber by labeling specific antigens (that are not found on surface of absorber) of target cells with antibodies, and subsequently sorting by FACS.

D. *In vivo* selection

This type of bio-panning is used for selecting phage clones displaying antibodies to tissue specific endothelial markers. Phage library is directly injected into an animal and then tissues are collected and examined for phage bound to tissue-

specific endothelial cell markers as was demonstrated for peptide phage library (Pasqualini and Ruoslahti, 1996).

9.2.7 Improvement of antibody affinity

The affinity of the antibody displayed on the selected phage clone can be improved by using several genetic engineering techniques, the affinity of human ScFv derived from antibody phage display library can be increased.

A. Multimer formation

ScFv is small molecule and can be formed as a multimer by using short linker of ~5 amino acid residues which lead to the association of two ScFv molecules called "diabody" with two paratopes, mimicking the F(ab)'2 fragment. The avidity of multimer ScFv molecule is higher than monomer ScFv (Dall'Acqua and Cater, 1998, Atwell *et al.,* 1999).

B. Site-directed mutagenesis

The nucleotides encoding one or more of the six CDRs are mutated. The ScFv displaying clones with the mutated amino acid residue(s) in CDRs are then screened for higher affinity clones (Balint and Larrick, 1993).

C. Chain shuffling

The VH gene of antigen-specific ScFv molecule is cloned into a repertoire of different VL chain-encoding DNA sequences. The resulting sub-ScFv library contains the phages displaying antigen specific VH chains and random VL chains. This sub-ScFv library is then panned with the same antigen to select phage clones which are displaying ScFv with higher affinity to the antigen of interest (Mark *et al.,* 1992).