# Directed evolution of novel properties starting from HisF of *Thermotoga maritima* as a structural scaffold

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultäten

der Georg-August-Universität zu Göttingen

vorgelegt von

**Zhenlian Ling**

aus Jiangsu, China PR

Göttingen 2005

D 7
Referent: Prof. Dr. H.-J. Fritz
Korreferent: PD Dr. A. Schwienhorst
Tag der mündlichen Prüfung: 17 Jan 2006

# Abstract

tHisF, imidazole glycerol phosphate synthase from hyperthermophilic organism *T. maritima*, with a $(\beta\alpha)_8$-barrel structure was used as scaffold for directed evolution of enzymes. The t*his*F gene libraries were constructed in phagemid vector pCANTAB 5E and in expression vector pKK223-3 by partially randomizing target positions in t*his*F gene. To evolve novel properties from the t*his*F library, a three-stage approach was adopted. In the first stage, tHisF variants that were able to complement eHisF activity in E. coli auxotroph strain UT860 were drawn out. 32 amino acids mutations were observed.A common mutation F86L was seen in 91 out of 98 tHisF variants. A proposed hypothesis was the residue Phe86 may be critical for the adaptation of hyperthermophilic enzyme tHisF to mesophilic environment. In the second stage, a tHisF mutant (termed HA03) with *N*'-[(5'-phosphoribosyl) formimino] -5-aminoimidazole-4- carboxamide ribonucleotide isomerase (HisA) activity was isolated from tHisF library by complementing an E. coli auxotroph strain Hfr G6. DNA sequencing of HA03 revealed that it carried ten amino acid exchanges. In the third stage, a transient state analog for target reaction (Aldol condensation) was designed and synthesized for future use in bio-panning to select desired biocatalyst candidates; meanwhile, in attempt to evolve aldolase from tHisF library by genetic selection, dihydrodipicolinate synthase (DHDPS, DapA) was targeted, a Δ*dap*A *E. coli* strain lacking DHDPS gene was constructed for this project, the biochemical assay system for DHDPS activity was established including the synthesis of an unstable substrate.

# Zusammenfassung

t*his*F, der Imidazol-Glyzerol-Phosphat-Synthase aus dem hyperthermophilen Organismus *T. maritima* mit der Struktur einer $(ßα)_8$-Barrels, wurde als Gerüst für eine direkte Evolution von Enzymen genutzt. Die t*his*F-Genbibliotheken wurden durch teilweises Randomisieren von Zielpositionen im t*his*F-Gen im Phagemid-Vektor pCANTAB 5E und im Expressionsvektor pKK223-3 konstruiert. Um neue Eigenschaften aus der t*his*F Bibliothek zu evolvieren, wurde ein dreistufiger Ansatz vorgenommen. Auf der ersten Stufe wurden tHisF-Varianten ausgewählt, die in der Lage waren e*His*F-Aktivität im auxotrophen *E. coli*-Stamm UT860 zu komplementieren. 32 Aminosäure-Austausche wurden beobachtet. Eine weit verbreitete Mutation, F86L, wurde in 91 von 98 Varianten von tHisF beobachtet. Eine mögliche Theorie war, dass der Rest Phe86 entscheidend für die Anpassung des hyperthermophilen Enzyms tHisF an die mesophile Umgebung ist. Im zweiten Schritt wurde eine tHisF Mutante (genannt HA03) mit *N*'-[(5'-phosphoribosyl) formimino] -5-aminoimidazol-4- carboxamidribonukleotid-isomerase (HisA)-Aktivität aus der HisF-Bibliothek durch Komplementation des auxotrophen *E. coli* Stammes Hfr G6 isoliert. Die Sequenzierung der DNA von HA03 ergab, dass hierbei zehn Aminosäure-Austausche vorlagen. Im dritten Schritt wurde für zukünftige Anwendungen ein Übergangszustands-Analogon für eine Zielreaktion (Aldo1-Kondensation) konstruiert und synthetisiert, um durch *bio-panning* gewünschte Kandidaten für Biokatalysatoren zu selektieren. In der Zwischenzeit wurde im Bestreben eine Aldolase aus der tHisF-Bibliothek durch genetische Selektion zu evolvieren, die Dihydrodipicolinat Synthase (DHDPS, DapA) ausgewählt. Ein Δ*dap*A *E. coli*-Stamm, bei dem das DHDPS-Gen deletiert ist, wurde für dieses Projekt konstruiert und das System eines biochemischen Assays für DHDPS-Aktivität etabliert samt der Synthese eines unstabilen Substrats.

# Content

# 1 Introduction

## 1.1 Chemical reaction, catalyst, enzyme and biocatalyst

Chemical reactions are underpinning all the chemical, biotechnological activities and biological functions. Thermodynamically, the reactants have to overcome the activation barrier to get activated, and subsequently form the products. The higher the activation energy the lower the reaction velocity, the difference of reaction speeds could be very huge.

Interestingly, reaction speed can be accelerated by catalysts. A catalyst is a substance that accelerates the rate of a chemical reaction while itself remains unchanged at the end of the reaction. Catalysis of a chemical reaction is achieved by reducing the activation energy for that reaction - the activation energy being the difference on free energy between the reactant(s) and the transition state for the reaction. This reduction in activation energy can be achieved either by stabilization (and hence reduction in free energy) of the transition state by the catalyst, or by the catalyst finding some other lower energy pathways for the reaction.

In bio-system, after millions of year's evolution organisms have enzymes that have become the most efficient catalytic substances to carry out all the reactions serving their life. Most of the enzymes are protein or protein complexes that catalyse different chemical reactions in organisms - enzymes are biological catalysts.

Comparing to conventional chemical catalysts, enzymes have a number of distinct advantages. First, enzymes have high efficiency - they are capable of catalysing reactions at rates well in excess of a million-fold faster than the uncatalyzed reaction, typical ratios of $k_{cat}/k_{uncat}$ being $10^6$-$10^{14}$, while chemical catalysts contribute to a reaction falling in $k_{cat}/k_{uncat}$ being $10^1$-$10^6$ (Bugg T 1997); Second, enzymes have high specificity - they are highly selective in the reactions that they catalyse. Since they bind their substrates via a series of selective enzyme-substrate binding interactions at a chiral active site, they are able to distinguish the most subtle changes in substrate structure, and are able to distinguish between regioisomers and between enantionmers. The high efficiency and high specificity are the hallmarks of enzymes, which are highly desired in synthetic reactions in organic chemistry.

The efficiency and specificity of enzyme are the results of natural mutation and selection. Each enzyme has evolved over millions of years to catalyse one particular reaction, which makes it an excellent catalyst for the reaction.

## 1.2 Enzyme Engineering

### 1.2.1 Background and achievement in enzyme engineering

Biological catalysts have become attractive for synthetic reactions because of their high efficiency and selectivity, their ability to produce relatively pure compounds compared with racemic mixtures, and their ability to produce regio-, chemo-, and stereo-specific compounds.

However, the variety of catalysts that nature offers is not enough to satisfy the needs for the production of new compounds, and most of them are only functional under biological conditions. To make use of the advantages of enzymes, there are still much works ahead.

Enzyme engineering is to generate enzymes toward desired enzymatic activities, enhance their selectivity, and stabilize enzymes towards the reaction conditions required for large-scale syntheses. Among these, the generation of stable enzymes with improved or novel catalytic activities is a fascinating topic of modern protein biochemistry. This goal is relevant not only for basic research, such as understanding the relationship between the structure and function of an enzyme, but also carries great potential for applications in biotechnology and biomedicine.

A number of enzymatic properties have been improved or altered by enzyme engineering. In these efforts, many successes have been reported, such as: improving enzyme catalytic efficiency (Muller BH et al., 2001; McCarthy JK et al., 2003); enhancing enzyme stability (Yano JK et al., 2003; Garrett JB et al., 2004); altering enzyme selectivity towards substrate (Wang XG et al., 2001; Joerger AC et al., 2003); and engineering to increase the yield to simplify manufacturing procedure, to lower the safety requirements, etc (Cherry JR and Fidantsef AL 2003).

Despite these achievements, there is still room for improvement remaining in this interesting and valuable field. Researchers are exploring by all means to the depth in the understanding of the enzyme mechanism and theoretic research in life science, and trying to broaden the application in biotechnology and related industries.

### 1.2.2 Approaches for enzyme engineering

As we know, almost all enzymes are protein or protein complex. Protein is the final product of genetic code, the advancement of DNA manipulation techniques has contributed much to the exploding research on enzyme engineering. An exciting development over the last few years is the application genetic engineering techniques to enzyme technology. Mutations in a gene that result in an altered amino acid sequence may directly affect its function through the structural perturbation of its folded state. This direct link between gene/protein sequence and protein

structure/function thus suggests a mechanism of evolution on the molecular level, in which mutation of a gene may lead to functional drift of the translated protein.

Two different strategies, rational design and directed evolution, have been applied to change enzymatic properties at wish or discover new proteins with catalytic activity, and design *de novo* structure with desired properties at wish is the utmost goal of protein engineering.

#### 1.2.2.1 Rational design

One of the basic enzyme engineering strategies is rational design, which is the planned redesign of the protein sequence by site-directed mutagenesis.

Rational design, or site-directed mutagenesis, depends on detailed structural and mechanistic information on the parent enzyme. The development of protein crystallography has opened many possibilities to obtain information about enzymes that could be exploited for rational design of desired properties. This strategy has been applied extensively with variable success, e.g. redesigning the substrate specificity of a large number of common classes of enzymes, such as oxidoreductases (dehydrogenases and reductases), hydrolases (acetylcholinesterases, lactamases,and proteases), transferases (aminotransferases,and glutathione-S-transferase), restriction enzymes, and enhancing the stability in different conditions, such as at high temperature, in extreme pH (acidic/basic) environment, in organic solvent (review: Cipolla L. 2004).

However, rational design is greatly hindered in practice by the complexity of protein function and still limited information of structures and functions of proteins. Generally, simple mutations (to be introduced by site-directed mutagenesis, for example) are not expected to have as drastic effects as altering an enzyme's substrate recognition pattern, as many amino acid residues in the enzyme (often not close to one another in the primary structure of the protein) affect the binding pocket of the substrate in the enzyme. Obviously, several amino acid residues may need to be altered simultaneously to achieve the goal of altering substrate specificities. However, as any amino acid residue may be altered into 19 other ones, the number of amino acid combinations that can be made if mutations are introduced at various residues simultaneously can become very large (Glick BR and Pasternak JJ 2003). For example, if five amino acid residues are altered simultaneously, there are 19-to-the-power-of-5 (that is ca $2.5 \times 10^6$) different combinations in which this can occur.

#### 1.2.2.2 Directed evolution

This strategy is directed evolution, which requires no structural information, because it simply mimics the natural evolution process: a combinatorial library of modified enzymes is created by random mutagenesis; and variants with the desired catalytic

properties are filtered by successive rounds of selection or screening. With advances in molecular biology, the possibility has arisen that an enzyme with the desired catalytic property can be isolated from a large pool of protein variants.

Recently directed evolution has been used successfully to modify the substrate, cofactor specificity of existing enzyme, to increase activity in organic solvents, to increase gene expression, increased enationselectivity, and increased thermostability (Jestin JL and Kaminski PA 2004).

Figure 1.1 is a schematic diagram of the processes of rational design and directed evolution.

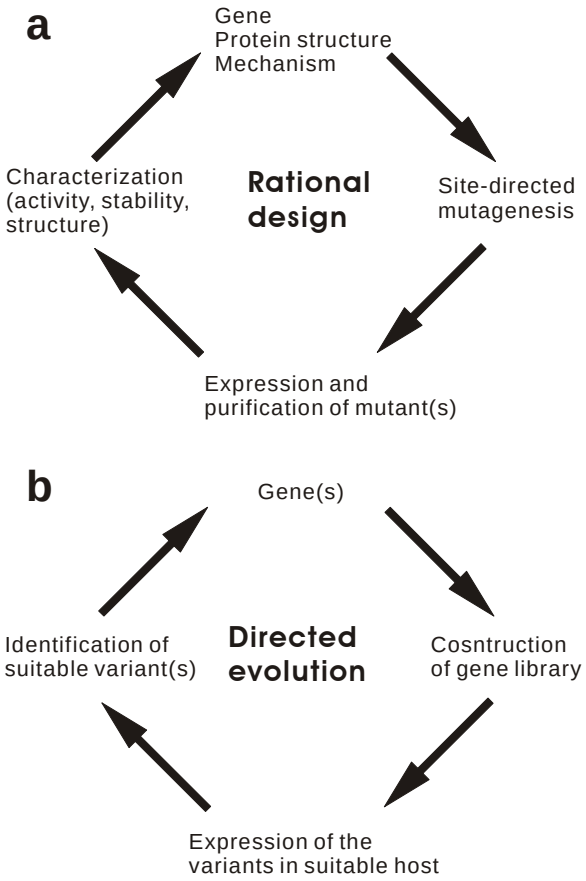



**Figure 1.1** Schematic diagram of the processes of rational design and directed evolution.

The methods of rational design and directed evolution have also been adopted in industry and therapeutic applications. It has been proven to be valuable for improving enzymes as well as for evolving new metabolic pathways, site-specifically engineered proteins (Wang et al., 2003), in genetics such as new selection markers (Hoseki et al., 1999) and in various industrial segments (Kirk et al., 2002). Therapeutic applications include an antiviral strategy by DNA polymerase engineering consisting of an

increase of the viral replication error rate thereby preventing viral replication etc (cf. review: Cipolla L. 2004; Valetti F and Gilardi G 2004).

#### 1.2.2.3 *de novo* design

In the attempts to discover new proteins with catalytic activities, chemists, biochemists and bioinformaticians also have achieved some success. Automated computational redesign of complete proteins, based on the "inverse folding approach" using statistics or first principles of protein folding and stability, or hybrids of these two, which aims to optimize the sequence of given structure, has led to the production of several redesigned proteins with interesting properties (review: Eijsink VG et al., 2004). More challenge work is computational *de novo* design proteins with desired activity. There are some successful examples now (recent review: Hecht MH et al., 2004).

### 1.2.3 Directed enzyme evolution

Despite continuous advances in our understanding of protein structure and function, however, there are many aspects of protein function we cannot predict. Many attempts to improve enzymes by rational design failed due to the still-limited understanding of enzyme structure/function relationships. Meanwhile, technical developments enormously increased the versatility and the success rate of directed evolution experiments. For these reasons, combinatorial strategies for protein engineering are appealing and were adopted in this project.

#### 1.2.3.1 General procedure of directed enzyme evolution

In the absence of detailed structural or mechanistic information, new functions can be engineered by introducing and recombining mutations, followed by subsequent testing of variants for the desired new function. Engineering the specificity and properties of enzymes within rapid time frames has become feasible with the advent of directed evolution.

All combinatorial optimization strategies require two fundamental components: a library, and a means of screening or selecting from that library. Much of the advantage of directed evolution of proteins lies in the fact that the coding information is held in a molecular medium which is straightforward to be amplified, read and manipulated, while the functional molecule, the protein, has a rich chemistry that provides a wide range of possible activities. This also has the advantage that (as long as a link between protein and nucleic acid is maintained) the identity of any selected protein can be directly determined by DNA sequencing. The methodology of choice in a directed evolution experiment is therefore to construct a library of variant genes, and screen or select from the protein products of these genes (Neylon C 2004).

### 1.2.3.2 Combinantorial construction of gene libraries

A range of methods are available for mutagenesis, and these can be used to introduce mutations at single sites, targeted regions within a gene or randomly throughout the entire gene. In addition, a number of different methods are available to allow recombination of point mutations or blocks of sequence space (Williams GJ et al., 2004).

Methods for the creation of protein-encoding DNA libraries may be divided into three categories. The first two categories encompass techniques that directly generate sequence diversity in the form of point mutations, insertions or deletions. These can further be divided into methods where changes are made at random along a whole gene and methods that involve randomization at specific positions within a gene sequence. The third category of techniques for library construction are those that do not directly create new sequence diversity but combine existing diversity in new ways (Neylon C 2004).

(i) Random methods introduce changes at positions throughout the gene sequence

Randomly targeted methods, encompasses most techniques in which the copying of a DNA sequence is deliberately disturbed. These methods, which include the use of physical and chemical mutagens, mutator strains and some forms of insertion and deletion mutagenesis as well as the various forms of error-prone PCR (epPCR), generate diversity at random positions within the DNA. For example, random mutagenesis by dNTP analogs (Zaccolo et al. 1996), the method bases on the incorporation of mutagenic dNTP analogs, such as 8-oxo-dGTP and dPTP, into an amplified DNA fragment by PCR. The mutagenic dNTPs are eliminated by a second PCR step in the presence of the four natural dNTPs only, resulting in a rate of mutagenesis of up to 19%. In error-prone PCR (Caldwell & Joyce 1992), mutations are introduced by adjusting PCR reaction conditions to induce an increased error-rate, these reaction conditions include: 1) using non-proofreading polymerase, such as Taq-pol; 2) using low annealing temperature to decrease product fidelity; 3) using low/unequal dNTP concentration; 4) using high concentration of $Mg^{2+}$ ion (up to 10mM); 5) using high cycle number e.g. 40, 60, or 80 cycles; 6) incorporation of $Mn^{2+}$ ion, typically 0.05-0.5mM, exceptionally 5mM; 7) re-PCR PCR-products or perform 2$^{nd}$ round of PCR after selection. The rate of mutagenesis achieved by error-prone PCR has been reported to be in the range of 0.6-2.0%.

(ii) Directed methods randomize a specific position or positions

The second category of methods targets a controlled level of randomization to specific positions within the DNA sequence (Ness JE et la., 2002; Zha DX et al., 2003) These methods involve the direct synthesis of mixtures of DNA molecules and are usually based on the incorporation of partially randomized synthetic DNA cassettes into genes

via PCR or direct cloning. The key to these methods is the introduction of diversity at specific positions within the synthetic DNA.

One of the methods to introduce random codon at specific positions within the synthetic DNA is oligonucleotide-directed mutagenesis. In standard oligonucleotide-directed mutagenesis schemes, a randomized DNA sequence is synthesized by sequentially coupling a mixture of the four nucleoside precursors to the growing oligonucleotide. In this way all 64 possible codon sequences (NNN, 4x4x4) are generated (including 41 redundant and 3 stop codons). This strategy can be improved by exploiting the third position redundancy of many codon assignments. By using all four nucleosides in the first two codon positions, but only G and C or A and T in the third position (NNG/C or NNA/T), the resulting mixture contains 32 triplets encoding all 20 amino acids, with 11 redundant and 1 stop codons. In this manner, the bias in favour of the amino acids encoded by multiple codon sequences is maintained, and the presence of a stop codon will produce truncated amino acid sequences upon translation. This event, which occurs with a frequency of $n/32$ where $n$ is the number of amino acids of the randomized sequence, considerably limits the complexity that can be achieved for long randomized peptide libraries. Another result of this form of codon bias is that it is difficult to insert codons for a subset of amino acids if this is desirable (Neuner P et al., 1998).

A number of solutions have been developed to tackle this problem. One of the solutions is to use pre-synthesized trinucleotide phosphoramidites as monomeric units in the synthesis of the oligonucleotides. This solves the problem of the codon bias by synthesizing the DNA one codon at a time. If it is desired to completely randomize one amino acid, then a mixture of 20 codons can be added. If a low level of mutagenesis is required then the mixture will be present at a lower concentration than the wild-type codon, and if a subset of amino acids is desired then this is easily accommodated by including only the desired codons. For example, the amino acid cysteine may become a problem in the protein engineering because of the possibility of disulfide bond formation and the other forms of oxidization of cysteine, so the codon for cysteine could be excluded from the gene sequence. With exclusion of one codon (here is the codon for cysteine) from incorporation of DNA molecules, it still results $3.23x10^{11}(19^9)$ possible combinations for a completely randomized sequence of only nine codons. However the size of libraries is usually limited by the efficiency of cloning and transformation and lies with approximately $10^8$-$10^9$ independent clones.

This method may easily achieve fine control over the presence of any set of residues at any given position, however the synthesis and efficient coupling of trinucleotide blocks are not straightforward. Early attempts to use triplets for the generation of protein mutants reported low coupling efficiency, as well as deletions in the final product. And it also showed a modest and uneven coupling efficiency of the triplets

and was found to generate a significant amount of single-base insertions (Neuner P et al., 1998).

(iii) DNA recombination by DNA shuffling/StEP/ITCHE

The third category of techniques for library construction are those that do not directly create new sequence diversity but combine existing diversity in other ways. These are the recombination techniques, such as DNA shuffling (Stemmer WP, 1994) and the staggered extension process (Zhao H et al., 1998) that take portions of existing sequences and mix them in novel combinations. These techniques make it possible to bring together advantageous mutations while removing deleterious mutations in a manner analogous to sexual recombination. To generate novel enzymes, the DNA-shuffling method of Stemmer has proven to be the most powerful. Also belonging to this category are methods such as iterative truncation for the construction of hybrid enzymes (ITCHE, Ostermeier M et al., 1999) that make it possible to construct hybrid proteins even when the genes have little or no sequence homology. While these techniques do not in principle produce new point mutations, they are generally dependent on a PCR reconstruction process that can be error prone, and new point mutations are usually produced as a by-product of these techniques (Neylon C 2004). DNA shuffling (Developed by Stemmer, 1994) generates libraries by random fragmentation of one gene or a pool of related genes, followed by the reassembly of the fragments in a self-priming PCR reaction. This method allows the recombination of sequences from different, related genes. Point mutations are also generated in the procedure of DNA shuffling. The overall rate of mutagenesis has been reported to be approximate 0.7%.

#### 1.2.3.3 Selection or screening for desired property

The sorting step is the most critical part of any evolution experiment: useful variants must be efficiently recognised and isolated from complex biocatalyst libraries.

Evolution requires a link between genotype (a nucleic acid that can be replicated encompassing a gene variant) and phenotype (a functional trait such as binding of a transition state analogue or catalytic activity). This linkage can be generated by physically coupling the gene to the encoded protein variant through protein display on phages, viruses, bacteria or eukaryotic cells or by using *in vitro* methods like ribosome display (Becker S et al., 2004). The searching for the target enzyme with the desired properties can be accomplished by genetic selection for a specific property, by phage display and selection for the binding property or direct screening all the candidates for desired property.

Screening, i.e. the one by one analysis of the catalytic properties of each member of the repertoire, allows typically between $10^3$ and $10^6$ distinct proteins to be considered (Lin and Cornish, 2002). Selections, which are different from screening, allow the simultaneous analysis of protein properties for sets of up to about $10^{8-10}$ different

proteins. Recently, new high-throughput screening/selection methods are emerging, e.g. cell surface display and sorting, compartment translation, *in vitro* ribosome or mRNA translation. See section 4.2.2.1 for detail information of methods for selection and screening.

## 1.3 (βα)8 barrel of protein and its catalytic activity

Structure determination of various proteins showed that in contrast to an infinite sequence space, protein tertiary structure is estimated to contain between 650 and 10,000 different folds (Chothia et al., 2003). This thesis work focuses on the most frequently encountered catalytic active protein fold -- $(\beta\alpha)_8$-barrel.

$(\beta\alpha)_8$-barrel is also called TIM-barrel, as it was first seen in triose phosphate isomerase (TIM). It consists of an eightfold repeat of (βα) units, such that 8 twisted β-strands are arranged in a closed parallel barrel-like structure, that forms the barrel's core surrounded by 8 α helices connected to the β-strands by loops of different length and form. (Figure 1.2)

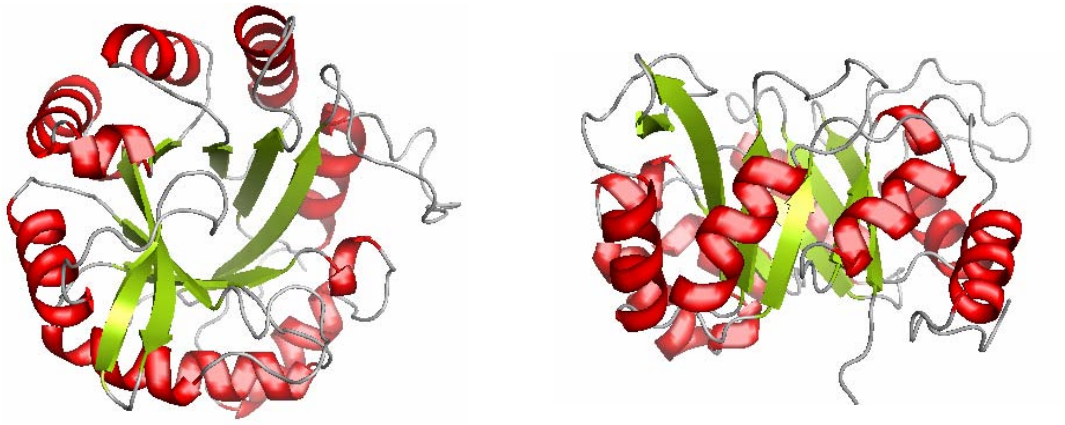


**Figure 1.2 Top and side view of $(\beta\alpha)_8$-barrel of tHisF**
Left: top-view of the active centre of tHisF; Right: side-view of tHisF. (Structural data from Lang D et al., 2000)

$(\beta\alpha)_8$-barrel is the most frequently encountered protein fold, and with only few exceptions, all members of this family are enzymes. It is the most common enzyme fold in the Protein Data Bank (PDB) database of known protein structures, accounting for at least 10% of all enzymes of known structure. At least 25 distinct enzyme families use this framework to generate the appropriate active site geometry (Schmidt DM, et al 2003). TIM barrel occurs in almost all the enzyme types which nominated by IECC (except the ligase) catalysing completely unrelated reactions (Nagano N et al 2002). The finding that the barrel of tHisF functions as an ammonia channel, broadens further functional diversity of the $(\beta\alpha)_8$-barrel fold (Douangamath A et al., 2002). The active site of $(\beta\alpha)_8$-barrel enzymes is always located at the C-terminal face of the central eight-stranded β-barrel, resulting in the spatial separation of elements

responsible for function and stability (Wierenga RK , 2001; Höcker B et al., 2001). The conformation of the loops located at the C-terminal end of the β-strands of the $(\beta\alpha)_8$-barrel is believed to determine the specificity to the substrate (Farber and Petsko, 1990; Lang et al., 2000).

Studies have shown that, although evolution occurs as a series of random mutations, stable folds such as the $(\beta\alpha)_8$ barrel and structural features of the active sites of enzymes are frequently reused in evolution and adapted for new catalytic purposes following the need of the substrate specificity, reaction mechanism, or active site architecture (Wise EL and Rayment I, 2004). The approximate partition of residues responsible for catalysis and binding in the $(\beta\alpha)_8$ fold offers the prospect to engineer new activities by changing catalysis while retaining binding, or changing binding while retaining catalysis (Gerlt JA and Babbitt PC 2001; Gerlt JA and Raushel FM 2003). A number of recent publications attested to the adaptability of $(\beta\alpha)_8$-barrels by directed evolution (Jurgens C et al., 2000; Gerlt JA et al., 2003; Joerger AC et al., 2003; Williams GJ et al., 2004).

The widespread occurrence of the TIM-barrel fold and the wide variety of biochemical reactions they involved, as well as their conserved secondary, tertiary and quaternary structure patterns have raised great interest in the structural, enzymological and evolutionary properties of this fold, make them ideal scaffolds which can be used to engineer novel catalytic activities.

### 1.4 Thermo stable protein with enzymatic activity

Microorganisms, which live at high temperatures such as 55-80°C, are called thermophile. Microorganisms that grow optimally above 80°C are called hyperthermophilic archaea or bacterial (e.g. *Thermotoga maritima*). To be able to live at those extreme temperatures their cell components display extreme resistance towards thermal degradation. Enzymes from extreme environments, sometimes called extremozymes, carry out their catalytic duties under conditions that are far beyond the boundaries within which more conventional biocatalysts operate. This characteristic is an attractive feature for use of these enzymes in biotechnology and industrial processes (Kaper T 2001).

Thermostable variants are also well suited for investigating the structure-function relationship of proteins. Heterologous expression of thermostable proteins in mesophilc hosts such as *Escherichi coli* allows their simple purification by heat precipitation of the thermolabile host cell proteins. The conformations of thermostable proteins are often very tolerant towards the exchange of amino acids, which allows to elucidate the functional role of individual amino acids by site-directed or random mutagenesis. Due to their high stability and high solubility, thermostable proteins

often readily crystallize, which makes possible the determination of their three-dimensional structures (Sterner R 2000).

## 1.5 tHisF: a thermalstable $(\beta\alpha)_8$-barrel-Protein

To evolve novel enzymes for non-biological use, a thermostable $(\beta\alpha)_8$ barrel fold enzyme—tHisF was chosen as the scaffold in this project.

HisF (EC 4.1.3.-) is the clyclase subunit of bacteria bienzyme Imidazole Glycerol Phosphate Synthase (ImGPS). It is one of the series of enzymes for the synthesis of histidine, and a key metabolic enzyme which links histidine and *de novo* purine biosynthesis.

HisF uses the nucleotide substrate N′-[(5′-phosphoribulosyl) formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) which is produced by another enzyme of the histidine biosynthesis pathway, HisA (catalyzing an isomerization reaction that precedes the ImGP synthase reaction) and nascent ammonia produced by HisH (amidotransferase subunit of bacteria bienzyme ImGPS) to catalyze the formation of the imidazole ring within histidine biosynthesis and additionally provides 5-aminoimidazole-4-carboxamide ribotide (AICAR) for use in the synthesis of purines (Fig. 1.3).

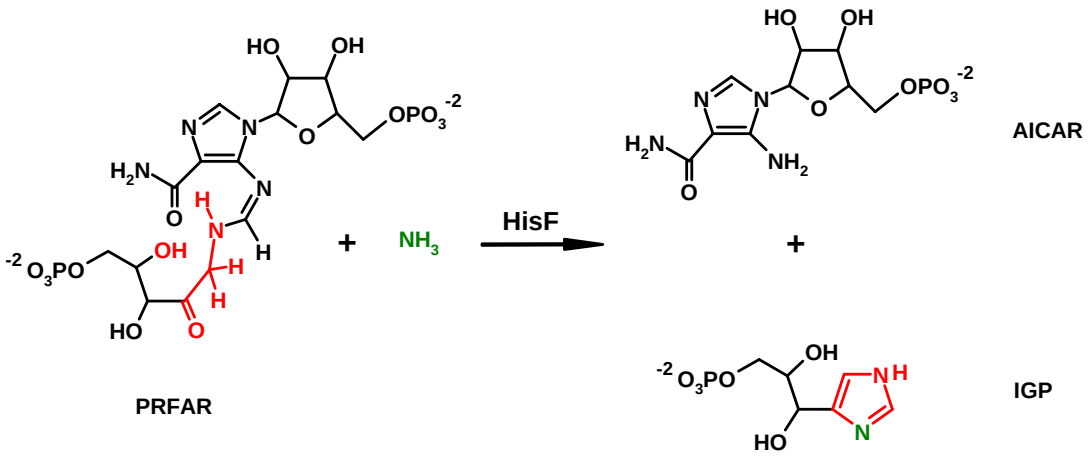



**Figure 1.3** The reaction catalyzed by HisF
the substrates PRFAR is supplied by HisA and $NH_3$ is supplied by HisH

tHisF is HisF protein from hyperthermophilic and phylogenetically early bacteria, *Thermotoga maritima*. It consists of 253 amino acids existing as a monomer, and has a molecular weight of 27,690 Da. In the secondary structure, tHisF contains 11 alpha-helices and 12 beta-strands. Sequence and secondary structure of tHisF are shown in Figure 1.4.

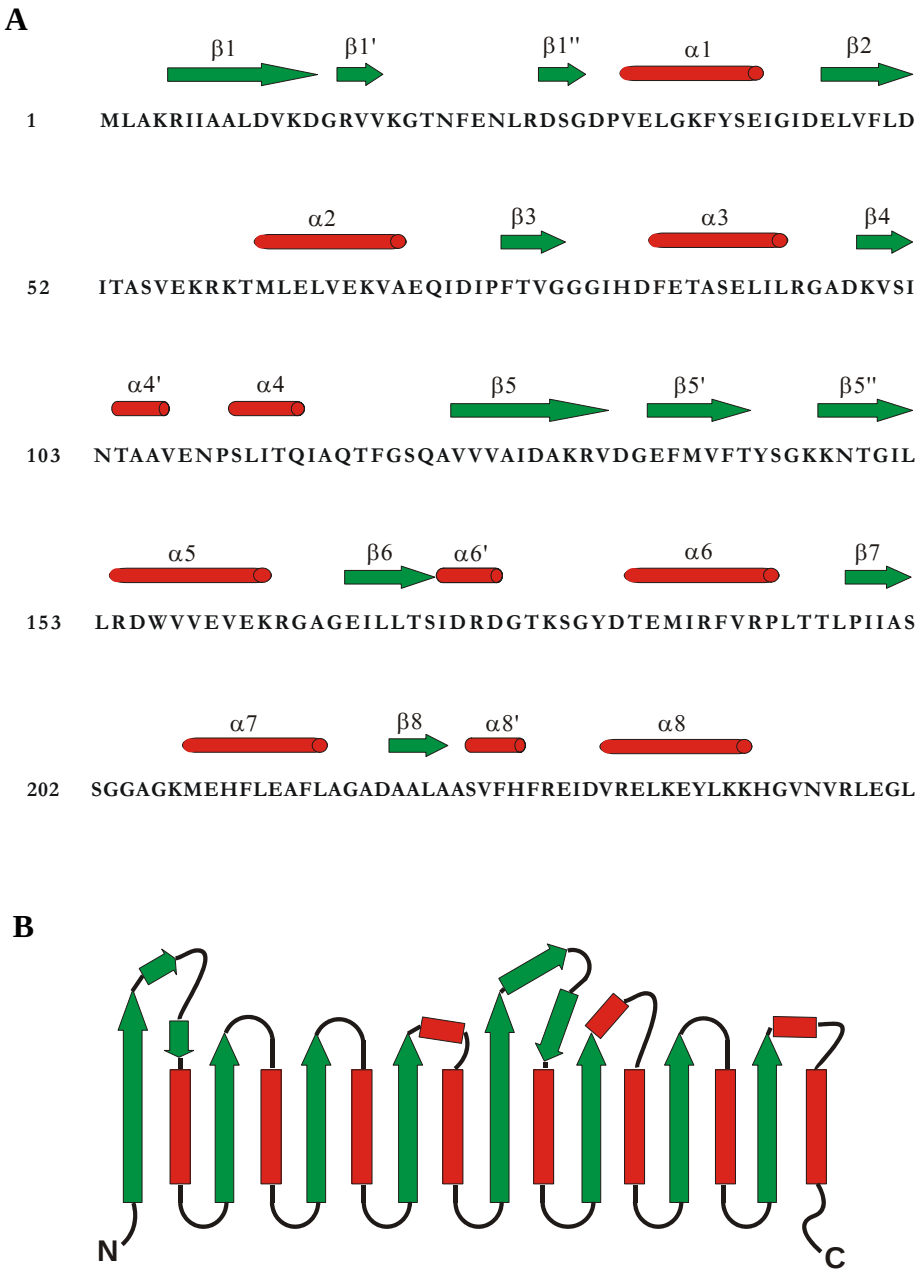



**Figure 1.4** Sequence and secondary structure of tHisF (Chain length: 253 amino acid residues) A, amino acid sequences of HisF from *T. maritima* (in single letter code); secondary structural elements are identified by green arrows (β strands) and red cylinders (α helices). The positions of secondary structure elements were taken from the X-ray structure of tHisF (Data from Lang et al., 2000). B, topography of secondary structural elements of tHisF, from left to right: β1/α1, β2/α2, β3/α3, β4/α41, β5/α5, β6/α6, β7/α7, β8/α8

X-ray analysis of HisF from *Thermotoga maritima* has revealed it as a $(\beta/\alpha)_8$-barrel structure (Lang D et al 2000, cf. Fig. 1.2). Biochemical evidence indicated that it is made up of two superimposable subdomains (HisF-N and HisF-C). HisF-N consists of the four N-terminal (β/α) units and HisF-C of the four C-terminal (β/α) units.

Separately, HisF-N and HisF-C are folded proteins, but are catalytically inactive. Upon co-expression *in vivo* or joint refolding *in vitro*, HisF-N and HisF-C assemble to the stoichiometric and catalytically fully active HisF-NC complex (Beismann-Driemeyer S and Sterner R 2001).

Crystal structure evidence (Douangmath A et al 2002) showed that tHisF bound with tHisH at the bottom of tHisF barrel (N-terminal face of eight β-strands of centre barrel), while the top of tHisF barrel was involved in PRFAR binding and is the catalytic center. Further more, computer modelling has suggested putative substrate binding sites located at N103, T104 for N-terminal half binding sites, and D176, G177, G203, A224, S225 for C-terminal half binding sites (Hennig M, personal communication).

The application of combinatorial strategies to protein engineering requires the construction of a library of variant proteins. The most straightforward method of constructing a library of variant proteins is to construct a library of nucleic acid molecules from which the protein library can be translated. For this purpose, the $(\beta\alpha)_8$-barrel fold enzyme - tHisF, a cyclase for the sixth step of histidine biosynthesis, from a hyperthermophile bacteria *Thomotoga mititima*, which consists of 253 amino acids, was used as the scaffold for enzyme engineering and the gene sequence information of tHisF was used in the combinatorial construction of gene library.

## 1.6 Aim of this project

The long-term aim of the project is to produce tHisF variants with novel catalytic properties other than those of the natural enzymes for non-biological use.

The aims of this PhD project are:

1. Using combinatorial method including self-priming PCR of controlled randomized synthetic oligonucleotides, DNA-shuffling and error prone PCR to (i) construct a phage library for potential use in panning against transient state analogs; (ii) construct a gene library coding for soluble tHisF variants for use in genetic selection.

2. Using the constructed library and the method of genetic complementation to investigate: the flexibility of residues involved in catalytic activity of enzyme and the protein-protein interaction; the relationship of the structure and stability of protein and the temperature adaptation of protein.

3. Using the gene library and the methods of directed evolution to search for novel catalytic activity from the $(\beta/\alpha)_8$-barrel.

# 2 Materials

## 2.1 Bacteria strains

BL21(DE3)pLysS

[*F - ompT hsdSB (rB - mB - ) gal dcm (DE3) pLysS (Cam* $^{R}$ *).*]

DH5α (Hanahan, 1983)

[$F^-$, *endA1*, *hsdR17* ($r_k^-m_k^-$), *supE44*, *thi1*, *recA1*, *gyrA* ($Nal^r$), *relA1*, Δ(*lacZYA-argF*)U169, Φ80*lacZ*ΔM15]

DH5α –JM1 (Home-made, 2004)

[$F^-$, *endA1*, *hsdR17* ($r_k^-m_k^-$), *supE44*, *thi1*, *recA1*, *gyrA* ($Nal^r$), *relA1*, *Δ(lacZYA-argF)U169, Φ80lacZΔM15 dapA::Kan*$^r$]

Hfr G6 (Matney et al 1964; Hofnung M et al., 1971; Thoma R et al 1998)

[*Hfr his thi Str* $^{s}$]

TG1 (Sambrook *et al.,* 1989)
[*supE thi-1 Δ*(*lac-proAB*) *Δ*(*mcrB-hsdSM*)*5* ($r_K^-m_K^-$) [F´ *traD36 proAB lacI*$^q$*ZΔM15*]

UTH 860 (also called SB3931) (Goldschmidt EP et al., 1970; Garrick-Silversmith L et al., 1970; Thoma R et al 1998)

[*araC14, glnV44(AS), galK2(Oc), LAM-, hisF860(stable), rpsL145(atrR), malT1 (LamR), xylA5, mtl-1*]

## 2.2 Bacterial phage

M13KO7 helper phage

A helper phage supplies wild-type version of coat proteins that is required for the successful re-infection of recombinant phage for amplification (Sambrook et al, 1989; Baek HJ et al 2002). M13KO7 is an M13 derivative which carries the mutation Met40Ile in gII, with the origin of replication from P15A and kanamycin resistance gene from Tn903 both inserted within the M13 origin of replication.

M13KO7 is able to replicate in the absence of phagemid DNA. In the presence of a phagemid bearing a wild-type M13 or f1 origin, single-stranded phagemid is packaged preferentially and secreted into the culture medium.This allows easy production of single-stranded phagemid DNA for mutagenesis or sequencing.

## 2.3 Buffers and solutions

Alkaline phosphatase buffer
100 mM NaCl, 5 mM MgCl2, 100 mM Tris/HCl, pH 9.5.

APS stock solution
10% Ammonium persultfate, stored at –20°C

BCIP (5-bromo-4-chloro 3-indolyl phosphate)
0.5 g BCIP dissolved in the 10ml dimethylformamide.

De-staining solution for PAGE
500 ml methanol, 100 ml acetic acid, add $H_2O$ till 1l.

DNase I dilution buffer(freshly prepared)
0.5M Tris pH7.5 + 100mM $MnCl_2$

dNTP stock solution
Mixture of 10 mM each dNTP (N=A,C,G, or T) in $H_2O$, stored at –20°C

E1-buffer (GENOMED)
50 mM Tris/HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNase.

E2-buffer (GENOMED)
0.2 M NaOH, 1 % (w/v) SDS.

E3-buffer (GENOMED)
3.2 M potassium acetate / acetic acid pH 5.5

E4-buffer (GENOMED)
600 mM NaCl, 100mM sodium acetate / acetic acid pH 5, 0.15% TritonX-100.

E5-buffer (GENOMED)
800 mM NaCl, 100 mM sodium acetate / acetic acid pH 5.

E6-buffer (GENOMED)
1250 mM NaCl, 100 mM Tris/HCl pH 8.5

EDTA stock solution 0.5M
18.6g EDTA(Ethylenediaminetetraacetic acid, disodium salt dehydrate) and 2.2g NaOH dissolved in H2O, adjust to pH8.0, add H2O till 100ml.

Ethidiumbromide stock solution
10 mg/ml ethidium bromide in $H_2O$

HEPES buffer
25 mM HEPES/KOH pH 7.6

IMAC buffer

25 mM HEPES/KOH pH 7.6, 0.5 M NaCl, 0--500 mM (25 mM, 50 mM, 75 mM, 80 mM, 90 mM, 100 mM, 150 mM, 200 mM, 300 mM, 500mM) imidazole.

6x Loading Dye Solution (MBI, Fermentas) for DNA analysis
0.2% bromophenol-blue, 0.2% xylene cyanol, 60 mM EDTA, 60% glycerol.

3% milk/PBST solution (block buffer)
PBST (wash buffer) containing 3% milk powder.

NBT (nitro blue tetrazolium)
0.5 g NBT dissolved in 10 ml of 70% Dimethyl formamide (DMF)

PBS buffer 10x
80 g NaCl, 2 g KCl, 14.4 g $Na_2HPO_4$, 2.4 g $KH_2PO_4$, pH 7.4, add $H_2O$ till 1l.

PBST buffer (Wash buffer)
100 ml PBS 10x buffer, 0.05% Tween 20, add $H_2O$ till 1l.

PEG/NaCl solution
20% polyethylene glycol 8000, 2.5M NaCl.

10x PfuTurbo® DNA polymerase reaction buffer (Stratagene)

100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris-Cl (pH 8.75), 20 mM $MgSO_4$
1% Triton® X-100, 1 mg/ml BSA.

Phage elution buffer
10 ml H2O + 140 µl Triethylamine
or 0.1M Glycine, pH2.2

Phenol: Chloroform (1:1)
1xV TE buffer-saturated phenol and 1xV $CHCl_3$.

5xProtein sample buffer

200 mM Tris/HCl pH 6.8, 20 % (v/v) glycerol, 3 %(w/v) SDS, 100 mM DTT, 0.05 % (w/v) bromophenol-blue.

Restriction buffer Blue (MBI, Fermentas)

50 mM Tris/HCl pH 7.5, 10 mM MgCl2.

Restriction buffer Green (MBI, Fermentas)

10 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl.

Restriction buffer Orange (MBI, Fermentas)

50 mM Tris/HCl pH 7.5, 10 mM MgCl, 100 mM NaCl.

Restriction buffer Red (MBI, Fermentas)

10 mM Tris/HCl pH 8.5, 10 mM $MgCl_2$, 100 mM KCl.

Restriction buffer Yellow (MBI, Fermentas)
33 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 66 mM KAc.

2x SDS-PAGE loading buffer (Laemmli loading dye)

160mM Tris/HCl pH6.8, 10% (v/v) glycerol, 4% (w/v) SDS,
0.021% (w/v) Bromphenolblue, 2% (v/v) ß-mercaptoethanol.

10x SDS-PAGE running buffer (Laemmli buffer, stock)

0.25 M Tris, 1.92 M glycine, 1 % (w/v) SDS, pH 8.3
36.3g Tris base, 144g Glycine, 10g SDS, dissolved in 1l $H_2O$
Dilute to 1x before use. Replace if the final pH is not within 0.1 pH units of pH 8.3.

SDS-PAGE resolving gel buffer
1.5 M Tris/HCl pH 8.8

SDS-PAGE stacking gel buffer
0.5 M Tris/HCl pH 6.8

Staining solution for PAGE
7.5% Coomasie Brilliant Blue R, 500 ml ethanol, 100 ml acetic acid, add $H_2O$ till 1l

T4 ligase buffer (MBI, Fermentas)
40 mM Tris/HCl pH 7.8, 10 mM MgCl2, 10 mM DTT, 0.5 mM ATP.

10x Taq DNA polymerase reaction buffer
50mM KCl, 15mM $MgCl_2$, 100mM Tris-HCl pH9.0, 1% Triton-100

TAE-buffer
0.89M Tris/HCl, 25mM Na2-EDTA, pH8.0

TBE-buffer
89 mM Tris, 89 mM boric acid, 2.5 mM EDTA

TE-buffer
10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0

TES-buffer 25x
250 mM Tris/HCl, 2.5mM EDTA, 2.5M NaCl, pH8.0

## 2.4 Chemicals and kits

### 2.4.1 Chemicals

| Chemicals | Company |
|---|---|
| Qualex Gold Agarose | AGS GmbH, Heidelberg |
| 30% Acrylamide, 2% Bisacrylamide solutions for A.L.F.-gel<br>Ampicillin (Na salt)<br>Isopropyl-β-D-thiogalactopyranosid (IPTG) | Applichem, Darmstadt |
| 5-Bromo-4-chloro-3-indolyl-phosphate-p-toluidine (BCIP)<br>1,4-Dithiothreitol (DTT)<br>Isopropyl-β-D-thio-Galactopyranoside (IPTG) | Biomol, Hamburg |
| Ampicillin (Na salt)<br>2'-Didesoxyribonucleoside-5'-triphosphate(dATP, dCTP, dGTP, dTTP) | Boehringer, Mannheim |
| Bacto Agar<br>Bacto Pepton<br>Bacto Trypton<br>Bacto Yeast Extract | Difco GmbH, Dreieich |
| Bromophenol blue<br>Coomassie Brilliant Blue G-250<br>Dimethylsulfoxide (DMSO)<br>Glycerol, 87%<br>β-Mercaptoethanol<br>Saccharose | Fluka Feinchemikalien GmbH, Neu-Ulm |
| Agarose electrophoresis grade | Gibco/Bethesda Research Laboratories (BRL) GmbH, Eggenstein |
| silica gel 60 F254 TLC- plates<br>All the other chemicals | Merck AG, Darmstadt |
| Chill out® liquid wax | MJ Research, Watertown, USA |
| Protogel ™ | National Diagnostics, New Jersey, USA |
| Acrylamid/Bisacrylamid (30% : 0.8%) mixture<br>Sequagel ™<br>Sequagel complete™ | National Diagnostics, Simerville, USA |

| | |
|---|---|
| Agar bacteriological grade<br>Trypton. | Oxoid, Hampshire |
| Chelating Sepharose Fast Flow | Pharmacia Biotech |
| Ammonium peroxodisulphate (APS)<br>N,N,N',N'-Tetramethylethylendiamine (TEMED) | Pharmacia/LKB GmbH, Uppsala |
| Ammonium acetate<br>Chloroform<br>Ethanol<br>Formamide<br>Hydrochloric acid 37%<br>Methanol<br>Potassium acetate<br>Sodium chloride | Riedel de Haen AG, Seelze |
| Boric acid<br>Chloramphenicol<br>Coomassie Brilliant Blue G250 and R250<br>Dimethyl sulfoxide (DMSO)<br>Ethidium bromide<br>Kanamycin<br>4-Nitrotetrazoliumchlorid blue (NBT)<br>Ponceau-S-ConcentrateTM<br>Tris-(hydroxymethyl)-aminomethane (Tris)<br>Triton X-100 (Octylphenol-polyethylenglycol)<br>Tween 20 (Polyoxyethylen-sorbitan-Mono-laureate)<br>Xylencyanol FF | Sigma Chemie GmbH, Deisenhofen |

### 2.4.2 Sample preparation Kits

| Kits | Company |
|---|---|
| Jetstar® Plasmid Midiprep Kit | Genomed, Bad Oyenhausen s.s. /Research Triangle Park |
| Quiagen® Plasmid Midipräp Kit | Quiagen, Hilden |
| NucleoTrap® and NucleoTrap®CR DNA-Purification Kit<br>NucleoSpin® Gel extraction Kit | Macherey-Nagel GmbH & Co, Düren |
| Thermo Sequinase fluorescent labelled primer cycle sequencing kit | Amersham, Braunschweig |

## 2.5 Enzymes and Antibodies

### 2.5.1 Enzymes

| Enzymes | Company |
|---|---|
| Ribonuclease A (RNase A)<br>DNase I (RNase Free) 10,000 units/ml | Boehringer,<br>Mannheim |
| Bovine serum albumin (BSA)<br>Calf Intestine Alkaline Phosphatase (CIAP)<br>T4 DNA-ligase,<br>various restriction endonucleases | MBI Fermentas,<br>Vilnius Litauen |
| various restriction endonucleases<br>Vent-Polymerase™ | NEB New England Biolabs,<br>Bad Schwalbach |
| *Pfu*$^{Turbo}$ ™ DNA Polymerase | Stratagene, California |
| Taq-DNA-Polymerase, λ-Exonuclease (homemade) | Abteilung molekulare Genetik und preparative Molekularbiologie, Universität Göttingen |

### 2.5.2 Antibody

| Antibodies | Company |
|---|---|
| Anti-mouse IgG-alkaline Phosphatase conjugate developed in goat | Sigma Deisenhofen |
| Anti-rabbit IgG-alkaline Phosphatase conjugate developed in goat | |
| polyclonal anti-HisF-antibody, rabbit serum | nanoTools, Teningen |

Antiserum solution:

1µl antibody or 20 µl anti-HisF-serum diluted into 10 ml PBS-Tween

### 2.5.3 Other products

RNase A stock solution (10mg/ml RNase A)

Dissolve 50 mg RNaseA in 5 ml 10 mM Tris-HCl (pH7.5), 15 mM NaCl, heat at 100°C for 10 min to inactivate DNase. Aliquot and store at –20°C. For short term use, store at 4°C. Do not freeze-thaw the stock.

Skinny milk powder
„Glücksklee“ (Nestlé Deutschland AG, Frankfurt a. M.)

## 2.6 Instruments and accessories

### 2.6.1 Instruments

| Instruments | Company |
|---|---|
| Branson Sonifier R-250 | Branson Sonic Power Company, USA |
| Biofuge pico and fresco | Heraeus GmbH, ZU |
| Centrikon® T1055 | KONTRON Instruments Ltd, UK |
| Centriprep® centrifugal devices | Amicon® , Beverly, US |
| MicroLitre bench centrifuge, cold rapid/k Microcentrifuge, Cold centrifuge Hettich –Laborzentrifuge Rotana/RPC, cold centrifuge | Hettich Zentrifugen, Tuttlingen |
| Sorvall RC-5C (Rotor type SS34, GSA) | Sorvall, Bad Nauheim |
| Constant Cell Disruption Systems | Constant Systems Ltd, Daventry, England |
| Cybertech CS-1 electrical ready picture camera | Cybertech GmbH, Berlin |
| Electrophoretische Transfer Apparature | Pharmacia/LKB, Freiburg |
| Eppendorf BioPhotometer | Schütt Labortechnik, Göttingen, Germany |
| Gel electrophoresis system for protein separation | ECPS 3000/150; Midget |
| Gel Imager UV-system | Gel Imager Intas, Göttingen, Germany |
| Gene Pulser®, E.coli Pulser TM Cuvette | BIO-RAD, California |
| Gene Pulser® and Pulse Controler | BIO-RAD, München |
| Horizontal electrophoresis chamber for DNA/RNA separation | Amersham Pharmacia Biotech |
| Ice machine | Ziegra, Stockport, England |
| Incubator | W. C. Heraeus GmbH, Hanau |
| Magnetic stirrer and Heater | IKA Labortechnik, Stauffen |

| | |
|---|---|
| MC410S Fine-balance | Sartorius, Göttingen |
| Metal block thermostat | Werkstatt Abt. mol. Genetik und präp., Uni. Goettingen |
| MGW Lauda RM6 (low temperature water bath) | DR.R. Wobser GMBH &Co KG, Lauda-Königshofen, Germany |
| LEFT-COR DNA Sequencer Model 4000L | LI-COR, Lincoln,USA |
| pH- Electrode U402-M5 | Ingold, Steinsbach/Taunus |
| pH-Meter-526 | Schütt Labortechnik, Göttingen, Germany |
| Speedvac Concentrator SVC 100H | Savant Instruments, Farmingdale, NY |
| Thermal cycler | MWG-Biotech, München |
| Thermocycler DNA Thermal-Cycler, PTC-150"MiniCyclerTM" | MJ Research, Watertown, USA |
| Thermoshaker | NFORS AG, Bottmingen |
| Two- ray spectrophotometer Uvikon 930 and 933 | Kontron Instruments GmbH, Eching |
| Ultrafree R-20 Nanopore-Water-Processor | MILLIPORE, Eschborn |
| Universal U 4800 P Fine-balance | Sartorius, Göttingen |
| UVT2035 UV-Light Box | Herolab, Wiesloch |
| Vertical electrophoresis system 2117-250 | NOVA-BLOT |
| Vortex Genie 2 | Bender & Hobein AG, Zürich, Swiss |
| Water bath | Gesellschaft fürLabortechnik mbH, Burgwedel |

### 2.6.2 Accessories

| Accessories | Company |
|---|---|
| Cylinders<br>Glass Pasteur pipettes | Brand GmbH &Co, Wertheim/Main |
| Dialyses tube 20/32 (16 mm) | SERVA, Heidelberg |
| Eppendorf reaction tubes | Eppendorf-Netheler-Hinz GmbH, Hamburg |
| Glass plate for SDS-Gel | Hoefer Scientific, SF, USA |
| Hamilton Microliter tips | Hamilton, Bansduz, Schweiz |
| LNGs 350-06 power source | Heinzinger, Germany |
| MaxiSorp™ NuncImmuno™ tube | NUNC, Wiesbaden |
| Minisart single use filter unit | Sartorius, Göttingen |
| 3 MM- filter paper (Whatman),<br>Nitrocellulose-Membranen | Schleicher & Schüll GmbH, Dassel |
| Parafilm | American National Can., Chicago,USA |
| Particles-free Kimwipes ™ | Kimbeley-Clark GmbH |
| PCR-Cups | Biozym, Hess.-Oldendor |
| Petri dishes<br>15 ml and 50 ml Falcon tubes | Greiner & Söhne GmbH & Co KG, Nürtingen |
| Pipette tips<br>Reaction tubes (1.5 ml, 2 ml)<br>50 ml tubes | Sarstedt, Nürnbrecht |
| Precision quartz cuvette | Hellma, Mühlheim/Baden |
| Scalpels | Bayha GmbH, Tuttlingen |
| Syringes | Terumo, Tokyo |
| Vivaspin centrifugal concentrators 20ml<br>(3,000MW,10,000MW, 30,000MW) | Vivascience AG, Hannover, Germany |

## 2.7 Media and antibiotics

### 2.7.1 Media

2 x YT medium (dYT) (Sambrook et al 1989)
16 g Bacto-trypton, 10 g yeast extract, 5 g NaCl, add 1l $H_2O$ and autoclave.

LB medium (Sambrook et al 1989)
10 g Bacto-trypton, 5 g yeast extract, 10 g NaCl, add 1l $H_2O$ and autoclave.

LB agar (Sambrook et al 1989)
10 g Bacto-trypton, 5 g yeast extract, 10 g NaCl, 15 g agar, add 1l $H_2O$ and autoclave.

M9-minimal medium
To 800ml $H_2O$ add: 64g $Na_2HPO_4$-$7H_2O$, 15g $KH_2PO_4$, 2.5g NaCl, 5.0g $NH_4Cl$, Stir until dissolved, Adjust to 1000ml with distilled $H_2O$. Sterilize by autoclaving.

MM (minimal medium) (Miller 1972)
MM contained 1x VB salts (Vogel and Bonner, 1956) supplemented with 0.4% glucose as carbon source, 5 µg of thiamine/ml, and 50 µg of amino acids/ml, when required.

SOC medium
20 g Bacto-trypton, 5 g yeast extract, 0.5 g NaCl, 20 ml 1M glucose, 10 ml 250 mM KCl, 5 ml 2M MgCl2, add $H_2O$ till 1l and autoclave (Sambrook et al 1989).

VB-medium 50 X (Vogel & Bonner 1956)
10g $MgSO_4$ .$7H_2O$, 100g citric acid. $H_2O$, 500g $K_2HPO_4$.anhydrous, and 175g $NaNH_4HPO_4$.4 $H_2O$ are dissolved successively in 670 ml $dH_2O$, add $H_2O$ to 1l. Store at room temperature; chloroform (ca 1ml) may be added as a preservative.

On 50-fold dilution with distilled water, the resulting 1x medium, designated E, has a pH of about 7.0; it is sterilized by autoclaving. Medium E is supplemented with dextrose (autoclaved separately) at a concentration of 5 g/liter.

### 2.7.2 Antibiotics

Ampicillin stock solution
100mg/ml, 1g ampicillin Na-salt dissolved in 10ml 70% ethanol (needn't further sterile), aliquot and stored at –20°C

Chloramphenicol stock solution
34 mg/ml, 340mg chloramphenicol dissolved in 10ml EtOH (needn't further sterile), aliquot and stored at –20°C

Kanamycin stock solution
75 mg/ml, 750mg kanamycin dissolved in 10ml $H_2O$, 0.2µm filter filtrated, aliquot and stored at –20°C

Streptomycin stock solution
10 mg/ml, 100mg kanamycin dissolved in 10ml $H_2O$, 0.2µm filter filtrated, aliquot and stored at –20°C

For the preparation of the selective media, sterilized antibiotic stock solution is added to the autoclaved medium. The stock solution's concentration and final concentrations are:

| **Antibiotic** | **Stock Conc.** | **Final Conc.** |
|---|---|---|
| Ampicillin | 100mg/ml | 100µg/ml |
| Chloramphenicol | 34 mg/ml | 34 µg/ml |
| Kanamycin | 75 mg/ml | 75 µg/ml |
| Streptomycin | 10 mg/ml | 10 µg/ml |

## 2.8 Molecular weight markers

### 2.8.1 DNA molecular size markers

The DNA size markers were mixed in 6x Loading Dye Solution (MBI, Fermentas) to the end concentration of 0.1µg/µl and stored at 4 °C.

GeneRuler™ 100 bp DNA Ladder, 11 bands:
1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 80 bp

GeneRuler™ 1 kb DNA Ladder, 10 bands:
10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 500 bp

Lambda DNA/Eco47I (AvaII) Marker, 36 fragments:
8126, 6555, 6442, 3676, 2606, 2555, 2134, 2005, 1951, 1611*, 1420, 1284, 985, 974, 894, 597, 590, 513, 511, 433, 398, 345, 310, 308, 272, 242, 215, 151, 88, 73, 67, 45, 42, 32, 29*, 23 bp

### 2.8.2 Protein molecular size markers

Mid-range marker (Promega)

| Protein Apparent molecular weight | [Dalton] |
|---|---|
| Phosphorylase b | 97,400 |
| Bovine serum albumin | 66,200 |
| Glutamate dehydrogenase | 55,000 |
| Ovalbumin | 42,000 |
| Aldolase | 40,000 |

| | |
|---|---|
| Carbonic anhydrase | 31,000 |
| Soyabean Trypsin Inhibitor | 21,500 |
| Lysozyme | 14,400 |

## 2.9 Oligodesoxyribonucleotides

### 2.9.1 Oligonucleotides for gene amplification

2'-deoxyribooligonucleotides were purchased from Purimex (Germany) and Metabion (Germany) in 'HPLC-purified' grade. Underlined nucleotides indicate the restriction sites, when incorporated. ATG is the start codon incorporated in the primers.

Primers for cloning *dapA* gene from E. coli chromosome into pET_B001 vector
DAPA_SEN(NcoI, 16GC/26mer) 5'AGA GGA TGG CCC ATG GTC ACG GGA AG-3'
DAPA_ANT(XhoI, 15GC/28mer) 5'TCT CCC TAA ACT CGA GCA AAC CGG CAT G-3'

Primers for cloning *dapB* gene from E. coli chromosome into pET_B001 vector
DAPB_SEN_C (NcoI, 11GC/26) 5'AGA GAA TAG CC ATG G TTG ATG CAA AC-3'
DAPB_ANT (XhoI, 10GC/31) 5'TAT TTT GTG GTT ACT CGA GAT TAT TGA GAT C-3'

Primers for cloning *his*F gene from pKK223-3 into pET_B001 vector
STHISF_S (NcoI, N, 11GC/26) 5'-GAA ACA GAA TCC ATG GTC GCT AAA AG-3'
STHISF_A (XhoI, C, 19GC/31) 5'-GGC TGC AGC TACTCG AGC AAA CCC TCC AGT C-3'

Primers for amplifying *his*F gene library from pCANTAB vector
sh-down (8GC/21) 5'-GTA TGT TGT GTG GAA TTG TGA-3'
sh-up(10GC/21) 5'-CAC GCC CTC ATA GTT AGC GTA-3'

Primers of final-PCR for cloning gene into pCANTAB5 E vector
Vb-01(Sfi I, 16GC/21) 5'-GTC AGC ATG CGG CCC AGC CGG-3'
Vb-38 (Bgl II, 11GC/23) 5'-CGA TAG GAT CAG ATC TCA AAC CC-3'

Primers for cloning *his*F gene into pKK223-3 vector
hisF-EcoR I (12GC/31) 5'-ATGC GAA TTC ATG CTC GCT AAA AGA ATA ATC-3'
hisF-Pst I (16GC/31) 5'-ATGC CTG CAG CTA CAA ACC CTC CAG TCT TAC-3'

### 2.9.2 Primers for colony screening PCR

For *dap*A gene in pKK223-3 vector

| | |
|---|---|
| dapA98 (13GC/20) | 5’-CAG CGG TAC TTC GGC GAT CG-3’ |
| pKK-rev(9GC9/20mer) | 5´-TTTTATCAGACCGCTTCTGC-3´ |

For *his*F gene in pCABTAB5 E vector

| | |
|---|---|
| Vb-03 (10GC/21) | 5’-GGC ACT AAC TTT GAG AAC CTG-3’ |
| sh-up (11GC/21) | 5’-CAC GCC CTC ATA GTT AGC GTA-3’ |

For *dap*A in pET 21d-B001 vector

| | |
|---|---|
| dapA98 (13GC/20) | 5’-CAG CGG TAC TTC GGC GAT CG-3’ |
| PETS1_UP(10GC/18) | 5’CAG CAG CCA ACT CAG CTT-3’ |

For *his*F gene in pET 21d vector

| | |
|---|---|
| Vb-03 (10GC/21) | 5’-GGC ACT AAC TTT GAG AAC CTG-3’ |
| PETS1_UP (10GC/18) | 5’CAG CAG CCA ACT CAG CTT-3’ |

For *his*F gene in pKK223-3 vector

| | |
|---|---|
| Vb-03 (10GC/21) | 5’-GGC ACT AAC TTT GAG AAC CTG-3’ |
| pKK-rev (9GC/20) | 5´-TTT TAT CAG ACC GCT TCT GC-3´ |

### 2.9.3 Oligonucleotides for sequencing

Primers for the vector pCANT AB 5E:

| | |
|---|---|
| cantupABI (8GC/20) | 5’-CAA CTT TCA ACA GTC TAT GC-3’ |
| cantdownABI (11GC/20) | 5’-TTA CGC CAA GCT TTG GAG CC-3’ |

Primers for the vector pET-21d

| | |
|---|---|
| PETS1_LO(9GC/18) | 5’ATA GGG GAA TTG TGA GCG-3’ |
| PETS1_UP(10GC/18) | 5’CAG CAG CCA ACT CAG CTT-3’ |

Primers for the vector pKK223-3

| | |
|---|---|
| pKK-for(9GC/20) | 5´-CGG TTC TGG CAA ATA TTC GT-3´ |
| pKK-rev(9GC/20) | 5´-TTT TAT CAG ACC GCT TCT GC-3´ |

### 2.9.4 Oligonucleotides for the synthesis of thisF-gene and thisF-gene libraries

st01-AgeI 5’-GTCAGCATGCGGCCCAGCCGGCCATGGCCGCTGAAACTACCGGTCTC-3’
st02 5’-Pho-GCTAAAAGAATAATCGCTGCTCTCGATGTGAAGGATGGTCGTGTGGTGAAG-3’
st03 5’-Pho-GGCACTAACTTTGAGAACCTGCGTGACAGCGGCGATCCT-3’
st04 5’-Pho-GTGGAACTGGGTAAATTCTACTCTGAGATTG-3’
st05 5’-Pho-GTATAGATGAACTCGTATTTCTGGATATCACTGCGTCTG-3’
st06 5’-Pho-TTGAGAAGAGGAAGACTATGCTGGAACTGGTTGAGAAGGTG-3’
st07 5’-Pho-GCCGAGCAGATTGATATTCCGTTCACTGTTGGTG-3’
st08 5’-Pho-GTGGTATCCATGACTTTGAGACCGCCTCTGAACTGATTCTG-3’
st09 5’-Pho-CGTGGTGCTGACAAGGTGTCTATTGCTGCTGCTGCTGTG-3’
st10 5’-Pho-GAAAATCCTTCTCTGATTACACAGATCGCTCAAACCTTCGG-3’
st11 5’-Pho-GAGTCAGGCTGTTGTTGTGGCGATAGATGCTAAGAGAGTG-3’
st12 5’-Pho-GATGGAGAGTTTATGGTATTCACCTACAGCGGTAAGAAGAACACGG-3’
st13 5’-Pho-GTATCCTGCTTAGAGACTGGGTGGTTGAAGTAGAGAAGAGA-3’
st14 5’-Pho-GGAGCAGGAGAGATTCTGCTCACCAGTATCGACAGAGACGGCACAAAATC GGGTTATGATACT-3’
st15 5’-Pho-GAGATGATTCGTTTCGTGCGTCCACTAACCACA-3’
st16 5’-Pho-CTTCCGATCATTGCTTCCGGTGGTGCGGGTAAGATGGAACAT-3’
st17 5’-Pho-TTCCTTGAGGCATTTCTGGCA-3’
st18 5’-Pho-GGTGCTGATGCTGCGCTTGCGGCGTCTGTCTTCCACTTTAGA-3’
st19 5’-Pho-GAGATTGATGTTAGAGAACTGAAAGAGTATCTGAAGAAGCAC-3’
st20 5’-Pho-GGAGTGAATGTAAGACTGGAGGGTTTGAGATCTGATCCTATCG-3’
st21-AgeI 5’-Pho-AGCGATTATTCTTTTAGCGAGACCGGTAGTTTCAGCGGCCATGGCCGGCTGG GCCGCATGCTGAC-3’
st22 5’-Pho-GGTTCTCAAAGTTAGTGCCCTTCACCACACGACCATCCTTCAC-3’
st23 5’-Pho-ATTTACCCAGTTCCACAGGATCGCCGCTGTCACGCA-3’
st24 5’-Pho-AAATACGAGTTCATCTATACCAATCTCAGAGTAGA-3’
st25 5’-Pho-TCCAGCATAGTCTTCCTCTTCTCAACAGACGCAGTGATATCCAG-3’
st26 5’-Pho-GAATATCAATCTGCTCGGCCACCTTCTCAACCAGT-3’
st27 5’-Pho-GCGGTCTCAAAGTCATGGATACCACCACCAACAGTGAACG-3’
st28 5’-Pho-GACACCTTGTCAGCACCACGCAGAATCAGTTCAGAG-3’
st29 5’-Pho-TGTAATCAGAGAAGGATTTTCCACAGCAGCAGCAGCAATA-3’
st30 5’-Pho-GCCACAACAACAGCCTGACTCCCGAAGGTTTGAGCGATCTG-3’
st31 5’-Pho-GTGAATACCATAAACTCTCCATCCACTCTCTTAGCATCTA TC-3’
st32 5’-Pho-TCTCTAAGCAGGATACCCGTGTTCTTCTTACCGCTGTAG-3’
st33 5’-Pho-AATCTCTCCTGCTCCTCTCTTCTCTACTTCAACCACCCAG-3’
st34 5’-Pho-GAAACGAATCATCTCAGTATCATAACCCGA-3’
st35 5’-Pho-AGCAATGATCGGAAGTGTGGTTAGTGGACGCAC-3’
st36 5’-Pho-CGCAGCATCAGCACCTGCCAGAAATGCCTCAAGGAAATGTTCCATCTTACC-3’
st37 5’-Pho-TACTCTTTCAGTTCTCTAACATCAATCTCTCTAAAGTGGAAGACAGACGCCGC
st38 5’-CGATAGGATCAGATCTCAAACCCTCCAGTCTTACATTCACTCCGTGCTTCTTCAGA-3’

Randomized ligonucleotides for the synthsis of hisF-gene-library

st02-mut 5’-Pho-GCTAAAAGAATAATCGCTXXXCTCXXXGTGAAGGATGGTCGTGTGGTGAAG-3’
st14-mut 5’-Pho-GGAGCAGGAGAGATTXXXCTCXXXAGTXXXGACAGAGACGGCACAXXXTCGGGTT ATGATACT-3’
st16-mut 5’-Pho-CTTCCGATCATTGCTXXXGGTGGTXXXGGTAAGATGGAACAT-3’
st18-mut 5’-Pho-GGTGCTGATGCTGCGXXXGCGGCGTCTGTCTTCCACTTT-3’

## 2.10 Vectors

### pCANTAB5 E (Pharmacia)

This is a versatile, bifunctional phagemid vector.

pCANTAB contains an M13 replication origin so, in the presence of helper phage, phagemid DNA is replicated in the normal M13 fashion.

Vector pCANTAB also carries the ColE1 replication origin and ampicillin resistance ($Amp^r$) gene for growth and selection in transformed E coli.

The interest genes can be cloned into the pCANTAB5 E between signal-sequence and gIIIp-gene. The gene III coding regions permits expression of the foreign proteins as N-terminal pIII fusions in *SupE* host strain like TG1. Thus, foreign protein can be displayed as a fusion with the gIIIp and the genetic information is packaged because of the packaging signal. This vector requires infection of the host cell with a helper virus to generate phage particles.

An amber codon is incorporated between the cloned gene and the start of gene 3 allowing the protein to be made as a soluble fragment in appropriate (non-suppressing) E.coli strain (e.g. WK6). It also includes peptide tag (E-tag), allowing detection of the soluble protein (Kay BK et al 1996) Schematic diagram: see Fig.2.1

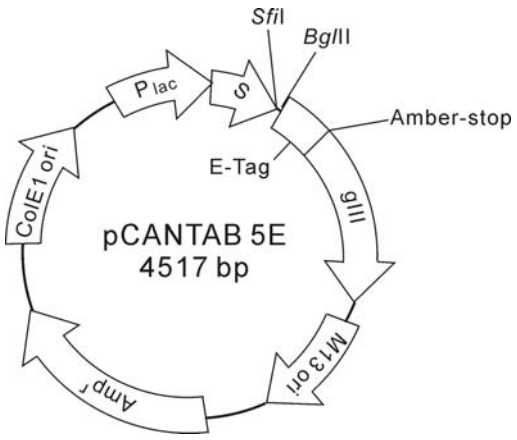



**Fig. 2.1:** Schematic representation of the phage display vector pCANTAB 5E. The restriction sites *Sfi*I and *Bgl*II are used to insert interest gene. Arrows indicate functional direction of genes.

$P_{lac}$: promotor/operator region
S: gIIIp signal sequence
E-tag: gene for the anti-E-tag-epitope
gIII: gene for the phage coading protein
M13 ori: replication origin of the filamental phagen M13;
$Amp^r$: β-Lactamase gene, mediate the resistence against ampicillin;
ColE1: replication origin

**pET_B_001**

pET_B_001 was modified from pET-21 d (+) (Novagen) by our laboratory. All the restriction cleavage sites were removed from the multiple clonal sites except NcoI and XhoI (Fig 2.2). No other modification.

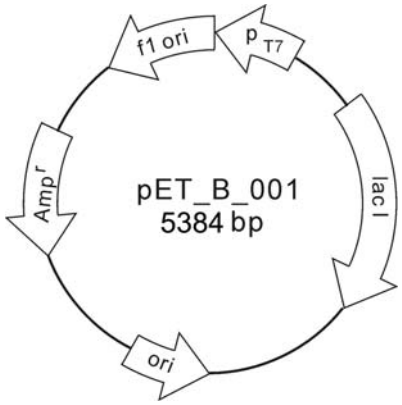



**Fig. 2.2 Map of pET_B_001**
$p_{T7}$: T7 promoter
lacI: lacI coding sequence
ori: pBR322 origin
$Amp^r$: β-Lactamase gene
f1 ori: f1 origin

**pKK223-3**(Amersham Pharmacia Biotech)

This *E. coli* expression vector contains a strong *tac* promoter for the multiple cloning site and strong *rrn*B ribosomal terminator downstream for control of protein expression (Amann E et al, 1983; Frost JW et al, 1984). *tac* promoter is inducible with 1-5mM IPTG, but even uninduced cells show a low level of expression.

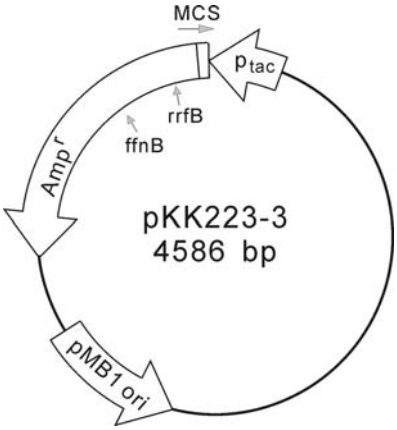



**Fig.2.3 Map of pKK223-3**
MCS:
HinIII
PstI
SaII
BamHI
SmaI
EcoRI

Ptac: tac promoter
MCS: multiple cloning site
Amp$^r$: **β-Lactamase gene**
pMB1 ori: pMB1 origin

## 2.11 Software and Databases

Chromas© Version 1.45 (32-bit) ©Mc Carthy

ClustalW http://www2.ebi.ac.uk/clustalw/
CorelDraw (Version 10, Corel Corporation)

DeLano, W.L. The PyMOL Molecular Graphics System (2002) http://www.pymol.org

ISIS Draw (Version 2.5, MDL Information System Inc.)

Photoshop (Version 6.0 Adobe)

SigmaPlot ® (Version 5.00, SPSS Inc.)

Swiss PdbViewer© Version 3.7 Glaxo Welcome Experimental Research,

Wisconsin Package Version 10.0 Genetic Computer Group (GCG), Madison, Wisconsin, USA

National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/

# 3 Methods

## 3.1 Microbiological methods

**Note**: All media and heat-stable solutions were sterilized by autoclaving for 20 min at 121 °C. The heat sensitive solutions were sterilized by filtering through a 0.2 µm membrane filter (Sartorius). Glassware was heated for 20 min at 160°C.

### 3.1.1 Culture and storage of *Escherichia coli* strains

*E. coli* strains were incubated overnight in 2xYT medium at 37°C with shaking at 150-200rpm. For long-time storage, the overnight culture was supplemented with glycerol to 20% and stored at -70°C.

The *E.coli* cells were also cultured on LB-Agar or 2xYT-Agar plates overnight at 37°C. Bacteria grew on rich media plates could be kept as working stock plates at 4°C for up to 3-4 weeks.

Bacteria, which served for helper-phage infection, were cultured on M9-Minimal plates. Bacteria grow much more slowly on minimal media than on rich media and do not survive prolonged periods of storage on minimal medium (even though at 4°C). At 1 week intervals, a fresh working stock plate should be prepared from the previous working stock plate.

For the selection for plasmids or phagemids with specific resistance, appropriate antibiotics were added into the medium.

### 3.1.2. Growth curve

The growth of E. coli cells includes several distinct phases: lag phase (0-2h after dilution), the logarithmic (log) phase (3–5 hours after dilution), stationary phase (~16 hours after dilution), eventually the culture enters the phase of decline as cells start to lyse, the number of viable bacteria falls, and DNA becomes partly degraded.

The growth curve of a bacterial culture can be monitored photometrically by reading the optical density at 600 nm. High $OD_{600}$ readings are calculated by diluting the sample in culture medium to enable photometric measurement in the linear range between 0.1–0.5 $OD_{600}$.

### 3.1.3 Preparation of log-phase E. coli cells

*E. coli* cells from glycerol stock or working stock plate were inoculated into 3ml 2xYT medium in sterile tube, cultured at 37°C on culture rotator overnight.

The overnight culture then was diluted into 2xYT in Erlenmeyer flasks at the ratio of 1:100, incubated at 37°C with shaking at 150-200rpm until the optical density at a wavelength of 600 nm ($OD_{600}$ ) reached the value of 0.4-0.6 (about 2-3 hours). This represented the log-phase *E. coli* culture.

For the preparation of electro-competent cells for gene libraries construction, SOB medium was used.

### 3.1.4 Preparation of CaCl2 treated competent cells (for the heat-shock transformation)

For single plasmid transformation, heat-shock of calcium-chloride treated competent cells with appropriate DNA is a convenient method, as it needs not special device, and multi transformations can be carried out in parallel.

1. Transfer one colony of appropriate strain from working stock plate or 10-20µl from glycerol stock into test tube with 3ml 2xYT medium and incubate it on culture rotator overnight at 37°C.
2. Dilute 0.5ml of the overnight culture into 50ml of fresh 2xYT medium (1:100). Incubate for 2 to 3 hours at 37°C with shaking at 150-200rpm until the culture reached an $OD_{600}$ of 0.45-0.55.
3. Transfer the culture into 50 ml Falcon-tube and centrifuge it at 2500g (4000rpm, Microcentrifuge Cold, Hettich) for 15 min at 4 $^{o}$C. Decant the supernatant, re-suspend the cell pellet in 25ml (1/2 volume) of sterile, ice-cold 100 mM $CaCl_2$. Incubate on ice for 5 min.
4. Centrifuge and decant as before, re-suspend in 10ml (1/5 volume) ice-cold 100 mM $CaCl_2$, incubate on ice for 5 min.
5. Centrifuge and pellet again, re-suspend the competent cells in 2ml of 100 mM $CaCl_2$, placed on ice.

   This competent cell can be used in 1-2 days in $CaCl_2$-solution at 4 $^{o}$C without much efficiency reduction. Or add in 15% sterile DMSO, mix well and aliquot to 200 µl, stored at -70 °C.
6. Transfer 200 µl competent cells to chilled eppendorf tube (e-tube) for the plasmid transformation.

### 3.1.5 Preparation of the electro-competent cells (for the electroporation)

For the preparation of electrical competent cells for library construction, SOB medium (without magnesium) was used.

1. One colony of appropriate strain from agar plate or 50µl from glycerol stock culture was transferred into test tube with 5ml SOB medium and incubated on rotator overnight at 37°C.
2. Dilute 2.5ml of the overnight culture into 250ml SOB medium (1:100) in 1l flask, incubate at 37°C with shaking at 150-200rpm until the $OD_{600}$ reaches 0.45-0.55 (about 2.5-3.5 hours).
3. Chill the culture at 0°C for 15 minutes, centrifuge it at 2500g, 4 °C for 15 minutes to pellet the cells.
4. Wash the pellet sequentially with 250ml, 125ml, and 60ml ice-cold sterile $H_2O$. Each wash step includes: re-suspend the cells; incubate cells on ice for 15 minutes; centrifuge to pellet the cells, and discard the supernatant.
5. The cells were finally resuspended in 1ml cold $H_2O$ (Usually cells can be resuspended in the $H_2O$ that remains in the centrifuge tube).

   Alternatively, wash and re-suspend the cells with 10% sterile glycerol-$H_2O$, these competent cells can be frozen in 0.2 ml aliquots and stored at -70 °C. However, for the preparation of gene libraries, the competent cells should be freshly prepared.

### 3.1.6 Transformation of the chemically competent cells (Heat-shock transformation)

For the heat-shock transformation, each of 10-500 ng DNA (ca. 10ng of supercoiled plasmid, or 500ng ligate-reaction) was added to 200µl aliquots of the $CaCl_2$ treated competent cell suspension.

The mixture of DNA and competent cell was incubated on ice for 30 min or longer (30-90min), followed by 3 min heat shock at 42 °C. Immediately add 1 ml of fresh 2xTY to each sample and recover the cells by incubating at 37 $^{o}$C for 60min (30-60 min depending on the resistance).

50-250 µl tranformants were plated out on the LB agar plates containing appropriate antibiotic to select for those transformed bacteria.

### 3.1.7 Transformation of the electro-competent E. coli cells (Electroporation) and construction of gene libraries (Dower et al 1988; Sharma RC et al 1996)

Electroporation is a high efficient transformation method. To achieve the **best** result, DNA and competent cell for electro-transformation must have a very low ionic strength to prevent arcing. The DNA may be purified by ethanol or iso-propanol precipitation, resuspend in sterile $H_2O$ to a concentration of 10-50 ng/µl of DNA.

Electro-transformation:

1. Mix 40 - 200 µl of the prepared electro-competent E. coli cell with 10-500ng salt free DNA solution in pre-chilled e-tube; transfer the mixture into a pre-chilled electroporation cuvette (gap width 2 mm), keep these materials on ice.
2. Set the BioRad E. coli pulser:
   Voltage at 2.5kV (note, for 1 mm gap cuvette, set the volt to 1.7 kV)
3. Place the electroporation cuvette (with the cell and DNA mixture) into the holder of cell-pulser; pulse the cells (the voltage display blinks), till the gene pulser beeps.
4. Add 1ml SOB into the transformed cells immediately after the gene pulser beeps; transfer the transformant to culture tube and incubate for 30-60 min (depending on the resistance) at 37°C to allow the expression of resistance gene.
5. Plate the cell suspensions onto appropriate plates:

   Take 10 µl for determination of independent clones (section 3.1.8), grow the rest of cells overnight on LB plates or in 2xYT medium containing appropriate antibiotics.

It may be stored at 4°C for use in a few days. Or wash the cells off plates with 2xYT medium and add glycerol to a final concentration of 20%. Store it at -70 °C until use.

### 3.1.8 Titer of the transformed or infected cells

1. Dilute the transformed or infected *E. coli* cells to ratios of 1:$10^3$, 1:$10^4$ and 1:$10^5$.
2. Plate 100 µl each of the dilutions onto LB-agar-antibiotic plates; incubate the plates overnight at 37°C.
3. Count the number of colonies on plates having between 30 and 300 colonies, calculate the cell populations according the colonies number and dilution ratio.

In the case of super-infection experiment, the number of colonies on LB-agar plates was counted to estimate the number of the infected cells. The number of the infected host cells reflects the number of phages because one phage could only infect one host cell.

### 3.1.9 Complementation assay

Complementation assay was carried out by transforming the plasmid libraries into specific auxotroph E. coli strain, plating the transformants on selecting plate without supplementary that the auxotroph E. coli strain normally needs. Incubate at 37°C till the colonies became visible. The cell without plasmid and the cell with plain plasmid served as negative control.

## 3.2 Molecular biology methods

### 3.2.1 DNA concentration Determination

One of the methods to estimate DNA concentration is agarose gel electrophoresis. The DNA sample was run in an agarose gel with two different volumes, paralleled to a DNA standard containing a known amount of DNA. The gel was visualized and photographed. The intensities of sample DNA bands were compared with standard DNA, and then the amount of DNA which had been loaded can be roughly estimated.

Another method to measure DNA concentration is carried out by a photometer. The DNA was diluted in $H_2O$. By determining the absorption of DNA at 260nm, the double strand DNA (dsDNA) concentration can be calculated using following formula:

$$C_{dsDNA}=OD_{260} \text{ x dilution folds x } 50\mu g/ml$$

Here, $OD_{260}$ 1 corresponds to 50µg dsDNA (or 33 µg single strand DNA, ssDNA). The concentration of oligos was calculated according this formula:

$$C\ (\mu M) = OD_{260}/(0.01 \text{ x } N)$$

where N is the number of bases.

The purity of the sample can be estimated from the ratio of $OD_{260}$ to $OD_{280}$. As proteins have maximum absorption at 280nm because of the existence of aromatic rings of Trp and Tyr. A protein-free DNA preparation usually has a relative coefficient of $OD_{260}$ to $OD_{280}$ above 1.8.

### 3.2.2 DNA molecular weight determination by electrophoresis

Analysis of DNA molecules was performed on the 0.8 - 1.2% agarose-1µg/ml ethidiumbromide-TAE-gel.

Usually 5 µl of DNA product was mixed with 1µl of 6x loading dye solution and loaded onto the agarose gel, paralleled with DNA molecular size marker as reference. The separation of the fragments was carried out at voltage of 60-120V for 25-60 min in 1x TAE running buffer. The gel was visualized with UV light at 302 nm and photographed.

### 3.2.3 DNA purification (Phenol-CHCl3 extraction) and DNA precipitation

DNA can be purified by Phenol-$CHCl_3$ extraction to remove proteins from DNA solutions.

1. Add 1x volume Phenol-$CHCl_3$ (considering the instability of phenol and impurities it may contain, phenol may be omitted, so only $CHCl_3$ is used) into DNA solutions, mix it thoroughly by vortex, centrifuge it at 13,000rpm (Biofuge pico, bench-top centrifuger, Heraeus) for 3-4 min.
2. Transfer the top aqueous phase (supernatant) to a new e-tube.
3. Extract a second time by repeating step 1 and 2.
4. Add 2.5xV 96% ethanol and 0.1xV 3M NaAc (or 7M $NH_4Ac$) into the supernatant, vortex briefly to mix all. Incubate it on ice for 10-15 min to precipitate the DNA.
5. Centrifuging at 12000 rpm, 4 °C for 15min to pellet the plasmid DNA. Remove the supernatant, wash the DNA pellet with 70 % ethanol to remove excess salt, dry at 37°C 5 min or at room temperature (rt.) for 10 min.
6. Re-suspend the DNA in appropriate volume of sterile $H_2O$.

### 3.2.4 Plasmid/phagemid DNA Mini-preparation

1. Inoculate 3 ml of 2xYT medium with single colony of *E. coli* cells harboring desired plasmid/phagemid, and incubate at 37°C overnight with shaking.
2. Transfer 1.5ml of the overnight culture to 1.5ml e-cup, harvest cells by centrifugation (13000 rpm, r.t., 0.5-1min) and remove supernatant as thoroughly as possible, resuspend the pellet in 150 µl E1-buffer (section 2.3).
3. Lyse the cells by adding in 150 µl E2-buffer, mix it gently, incubate at r.t. for 3-5 min; add in 150 µl E3-buffer to the lysate to neutralize the mixture and precipitate cell membrane particles, proteins and chromosomal DNA. Mix it carefully, and centrifuge this mixture at 13000 rpm, rt, 12 min.
4. Transfer the supernatant to a new 1.5ml e-cup, add in 1.5µl RNase A stock solution (10mg/ml), incubate at 37°C for 10-15 min.
5. Extract proteins with 1x volume Phenol-$CHCl_3$ solution (or $CHCl_3$) twice.
6. Precipitate DNA from the solution by adding in 2.5x volume 96% ethanol and 0.1x volume 3M NaAc, incubate on ice for 10-15 min.
7. Pellet the plasmid/phagemid DNA by centrifuging the mixture at 13000 rpm, 4 °C for 15min. Remove the supernatant, wash the DNA pellet with 70 % ethanol and re-suspend the DNA in 50µl $dH_2O$.

### 3.2.5 Plasmid/phagemid DNA Midi-preparation (anion exchange chromotography with Jet-Star kit)

1. Inoculate 50 ml 2 xYT medium with *E. coli* cells harboring the desired plasmid(s), incubate at 37 °C overnight with shaking.
2. Harvest cells by centrifuging at 2500g (4000 rpm, Hettich) r.t., 15 min. Discard the supernatant, resuspend the cell pellet in 4 ml E1-buffer.

3. Lyse the cells with 4 ml E2-buffer, incubate at r.t. for 5 min. Add in 4 ml E3-buffer to neutralize the lysate. Chromosomal DNA, proteins and membrane particles were removed by centrifuging at 11000g (13 000 rpm, Biofuge pico) r.t. for 15 min.
4. The supernatant was applied to a Midi JetStar-column (pre-equilibrated with 10 ml E4-buffer). The column was washed with 10 ml E5-buffer twice.
5. The plasmid/phagemid DNA were eluted with 5 ml E6-buffer.
   The eluted DNA was precipitated with 0.25 volume (1.2ml) of isopropanol, centrifuged at 13000rpm at 4°C for 30 min, the pellets were washed with 70 % ethanol. Re-suspend the DNA in 140 µl de-ion $H_2O$**.**

### 3.2.6 DNA digestion with Restriction Enzymes

Restriction digestion is the process of cutting DNA molecules with special enzymes called Restriction Endonucleases (or Restriction Enzymes, REs). These Restriction Enzymes recognize specific sequences in the DNA molecule (for example GATATC) wherever that sequence occurs in the DNA. The RE's activity is measured in Units where 1 Unit is the amount of enzyme needed to digest 1 µg of bacteriophage lambda DNA in 1 hour. REs are often shipped at 10 Units/µl.

Digestion of DNA was performed in its respective buffer and temperature recommended by the supplier. In general, about 0.5-5 µg double strands DNA (~5000bp) was digested with 2-20 units of a restriction endonuclease. When reaction conditions for two restriction enzymes are similar, two enzymes digestion can be carried out simultaneously. The amount of DNA and RE used depends on the task. A general example of components in RE digestion reaction is:

**Table 3.1** General composition of restriction digestion

| | |
|---|---|
| DNA | 0.5-5µg (ca 1-10 µl) |
| 10x RE Buffer | 2-20µl |
| Restrict endonuclease/s | 2-20 Units |
| H2O | xµl |
| --- | --- |
| Total volume | 20-200µl |

Mix the ingredients gently, incubate the mixture at the recommended temperature. The reaction time depends on the activity of REs, amount of RE and DNA used, normally, 1-1.5h. Some of the reaction's end can be judged by size or pattern of DNA bands on agarose-gel. To cut of the restrict sites of PCR products a slight excess of enzyme or a longer incubation time can be applied to ensure complete digestion. The digested DNA products then are separated and purified by agarose gel electrophoresis for further experiments (section 3.3.5).

### 3.2.7 Dephosphorylation of DNA 5'-termini with Calf Intestine Alkaline Phosphatase (CIAP)

After the restriction digestion, the DNA products can be treated with alkaline phosphatase to prevent self-ligation (Sambrook *et al.* 1989).

Dephosphorylate linear vector with alkaline phosphatase was performed under condition recommended by supplier. The enzyme was diluted with 1x reaction buffer, and the reaction was carried out according to the following procedure.

1. Prepare reaction mixture:

**Table 3.2** General composition of dephosphorylation

| | |
|---|---|
| DNA solution | 10-40µl (1-20 pmol of DNA 5'-termini) |
| 10x reaction buffer | 5µl |
| CIAP | 1 u /µl |
| Deionized water | xµl |
| --- | --- |
| Total volume | 50µl |

2. Incubate at 37°C for 30 minutes.
3. Stop reaction by heating at 85°C for 15 minutes or extract with phenol/chloroform and then precipitate with ethanol.

Dephosphorylation can also be performed by adding Calf Intestine Alkaline Phosphatase directly in mixture after DNA cleavage with restriction endonuclease. Use 0.05 units Calf Intestine Alkaline Phosphatase for dephosphorylation of 1 picomole DNA termini.

### 3.2.8 DNA purification from agarose gel with NucleoTrap kit

DNA purification from agarose gel was carried out according to the manual of NUCLEOTRAP extraction kit.

1. Excise the DNA fragment from agarose gel with a clean scalpel, transfer the gel slice into 2ml pre-weighed microcentrifuge tube (e-cup) and weigh the gel slice.
2. Add in three folds gel weight of N1 buffer and appropriate amount binding suspension (4µl /µg DNA, but at least 10 µl in one cup). Incubate the e-cup at 50°C and shake the e-cup in 2min interval till the gel dissolved completely.
3. Centrifuge the e-cup at 13 000rpm for 30 sec (on Biofuge freco/Heraeus) to pellet the binding suspension.
4. Wash the sediment with 500 µl N2 buffer twice and 500 µl N3 buffer one time, dry the sediment at rt for 5-10 min.
5. Elute the DNA from the sediment with elution buffer or deionized water by incubation at room temperature for 10-15min, followed by centrifugation to remove the binding particle. The eluted DNA was stored at -20 °C to avoid degradation.

### 3.2.9 DNA ligating reaction

DNA ligations were performed by incubating appropriately treated inserting DNA fragments, appropriately treated cloning vector and T4 DNA ligase in buffer.

A ligation mixture for single gene cloning usually included 20-50 ng vector DNA, five fold molar excess insert DNA, 1 x T4 ligase buffer and 2-10 units of the T4 DNA ligase in a final volume of 20 µl. Meanwhile, a parallel ligation in the absence of insert DNA was performed to determine the background clones arising from self-ligation of inefficiently de-phosphorylated or inefficiently purified linearized vector.

The components in the table below were mixed in a 1.5ml tube. The mixture was then incubated at 16°C overnight for ligating reaction.

**Table 3.3** General composition of the ligating reaction

| | |
|---|---|
| Cloning vector | 30ng |
| Insert DNA | 30ng |
| 10x Ligase Buffer | 1µl |
| ligase (5U/µl) | 1µl |
| H2O | xµl |
| --- | --- |
| Total volume | 10µl |

When the reaction is set for construction of gene libraries, more amount of vector and higher concentration of T4 ligase were used. The molar ratio of linear vector to insert fragments is ca 1: 3.

After the reaction, the mixture was then subjected to Phenol-$CHCl_3$ purification (section: 3.2.3) and the purified product was used for DNA transformation.

### 3.2.10 Polymerase chain reaction (PCR)

Polymerase Chain Reaction (PCR) is a repetitive bidirectional and exponential DNA synthesis via primer extension of a region of nucleic acid. Amplification of certain DNA template requires two oligonucleotide primers, four deoxynucleotide tripohosphates (dNTPs) and DNA polymerase to perform the synthesis.

#### 3.2.10.1 Composition of the PCR mixture

A. Template DNA

Usually the template DNA amount is in the range of 50pg-1ng for plasmid or phage DNA and 0.1-1µg for genomic DNA, for a total reaction mixture of 50µl.

B. Primers

PCR primers are usually 20-30 nucleotides in length. Longer primers provide sufficient specificity. The GC content should be 40-60%. More than three G or C nucleotides at the 3'-end of the primer should be avoided, as nonspecific priming may occur. The primer should not be self-complementary or complementary to any other primer in the reaction mixture, in order to avoid primer-dimer and hairpin formation. The melting temperature of flanking primers should not differ by more than 5°C, so the GC content and length must be chosen accordingly.

If the primer is shorter than 25 nucleotides, the approximate melting temperature (Tm) is calculated using the following formula:

$$T_m = 4\,(G + C) + 2\,(A + T)$$

G, C, A, T - number of respective nucleotides in the primer

If the primer is longer than 25 nucleotides, the interactions of adjacent bases, the influence of salt concentration etc. should be evaluated. The melting temperature may be calculated using the following formula:

$$T_m[°C] = 69.3 + 0.41\,(\%\ GC - 650/\text{length of nucleotide})$$

Optimal annealing temperature is generally 5°C lower than the melting temperature of the primer-template DNA duplex.

C. dNTPs

The concentration of each dNTP in the reaction mixture is usually 200µM. It is very important to have equal concentrations of each dNTP (dATP, dCTP, dGTP, dTTP), as inaccuracy in the concentration of even a single dNTP dramatically increases the misincorporation level. dNTPs concentrations of 100-250µM of each dNTP result in the optimal balance of product yield (greater at higher dNTP concentration) versus specificity.

### 3.2.10.2 Temperature Cycling

A. Initial Denaturation Step

The complete denaturation of the DNA template at the start of the PCR reaction is of key importance. Incomplete denaturation of DNA results in inefficient utilization of template in the first amplification cycle and a poor yield of PCR product. The initial denaturation should be performed over an interval of 1-3min at 95°C if the GC content is 50% or less, and this interval should be extended up to 10min for GC-rich templates or denaturation temperature may be increased up to 97°C.

B. Denaturation Step

Usually 0.5-2min denaturation at 94-95°C is sufficient, since the PCR product synthesized in the first amplification cycle is significantly shorter than the template DNA and is completely denatured under these conditions.

The temperature of the denaturizing step also depends on the DNA polymerase used: *Tlf* and *Tag* polymerases being less thermostable, *Pfu* and *Vent* polymerase - more thermostable. (Dieffenbach and Dveksler, 1995)

C. Primer Annealing Step

Usually the optimal annealing temperature is 5°C lower than the melting temperature of primer-template DNA duplex. Incubation for 0.5-2min is usually sufficient. If non-specific PCR products are obtained in addition to the expected product, the annealing temperature should be optimized by increasing it stepwise by 1-2°C.

D. Extending Step

Usually the extending step is performed at 70-75°C. Pfu DNA Polymerase exhibits lower than that of Taq DNA Polymerase extension rate (0.5kb/min), so 2min extension time is recommended for every 1 kb to be amplified.

E. Number of Cycles

The number of PCR cycles depends on the amount of template DNA in the reaction mix and on the expected yield of the PCR product. For most amplification reactions, 25-35 cycles are usually sufficient. In general, using the fewer cycles can ensure less amount of non-specific background product.

F. Final Extending Step

After the last cycle, the samples are usually incubated at 72°C for 5-10 min to fill-in the protruding ends of newly synthesized PCR products. (Skerra A, 1992)

There are a lot of variations of PCR method, differing in the type of DNA polymerase used, template as a starting material and desired products to be obtained, 3.2.11—3.2.14 are methods used in this work.

### 3.2.11 Amplification PCR

A general composition of PCR reaction used in this work was shown in Table 3.4; the amount of reaction was set up to 50- 100 µl. Annealing temperature is calculated by the formula in 3.2.10.1B. Table 3.5 is a general program for PCR cycling.

**Table 3.4** General composition of PCR reaction

| Reagent | End concentration |
|---|---|
| Template DNA | 50-100 ng |
| dNTPs | 200µM each |
| Primers | 10-100 pmol each |
| Reaction buffer | 1 x |
| DNA polymerase | 0.5 - 5 U |
| $Mg^{+2}$ (chloride or sulfate) | 1.5 mM |
| DMSO when needed | 2% |

Overlay the sample with a half volume of mineral oil or wax.

**Table 3.5** Cycling profile for a typical amplification PCR

| Step | Segment | Time | Temperature |
|---|---|---|---|
| 1 | Initiate denaturation | 1-2 min | 94-96°C |
| 2 | Denaturation | 30-45sec | 92-94°C |
| 3 | Annealing | 30sec -1min | 45-65°C |
| 4 | Extension | 1-2 min | 72°C |
| 5 | cycle step2-4 | 25-30 times | |
| 6 | Final extension | 2-5min | 72°C |

### 3.2.12 Colony screening PCR

Colonies with desired plasmid can be identified quickly by colony screening PCR.

Transfer single colony into 25µl 1x Taq-pol buffer in 200µl PCR tubes, incubate at 95°C for 10-15 min in thermo-cycler (PCR machine) for cell lysing. The mix was used as template DNA and combined with screening PCR solution (Table 3.6 lower part) for screening PCR reaction. Alternatively, isolate the plasmid DNA of interest colony by mini-preparation (section 3.2.4) and use the plasmid DNA as template.

Typically, the amplification was performed in a 50 µl volume with Taq-polymerase and a pair of primers specific for the insert gene and vector, respectively.

**Table 3.6** General composition of the Colony screening PCR

| | |
|---|---|
| Template solution: | |
| Cell lysate or plasmid in 1xTaq-pol buffer containing | 25µl |
| Screening PCR solution: | |
| $dH_2O$ | 19.5µl |
| 10x Taq-pol buffer (containing $MgCl_2$) | 2.5 µl |
| 10 mM dNTPs | 1 µl |
| 100 pmol/µl primer for insert | 0.1 µl |
| 100 pmol/µl primer for vector | 0.1µl |
| DMSO | 0.8µl |
| Taq polymerase | 1µl |
| Total volume | 50µl |

The reaction was carried out on Thermo-cycle with cycling program in Table 3.5 for 25-30 cycles. After reaction, load 5-10µl of each PCR product with 2µl DNA-loading-dye to agarose gel and run electrophoresis to examine the products by the band size of the desired products.

### 3.2.13 Error prone PCR (Cadwell & Joyce 1992)

The method introduces mutations in the gene of interest using a PCR reaction under conditions that induce an increased error-rate of the DNA-polymerase. The rate of mutagenesis achieved by error-prone PCR has been reported to be in the range of 0.6-2.0%. Proofreading enzymes such as Pfu exhibit error rates in the range from $10^{-6}$ to $10^{-7}$ whereas nonproofreading enzymes like Taq polymerase show error rates in the range from $10^{-4}$ to $10^{-5}$. This rate can be significantly enhanced by modifying the following parameters of a PCR reaction:

higher $Mg^{2+}$ concentration of up to 7 mM

partial substitution of $Mg^{2+}$ by $Mn^{2+}$

unbalanced dNTP concentrations

increased number of cycle

lower annealing temperature

### 3.2.14 DNA shuffling and self-priming PCR (Stemmer, 1994)

DNA shuffling generates libraries by random fragmentation of one gene or a pool of related genes, followed by reassembly of the fragments in a self-priming PCR reaction. This method allows the recombination of sequences from different, related genes. The overall rate of mutagenesis has been reported to be approx. 0.7%.

All the procedures in this experiment are carried out along the recommended conditions in the commercial products' manual, with a little modification.

A. DNase I digestion of interest genes and fragments selection

The pivotal step in DNA shuffling is digestion of the interest gene for production of fragments of appropriate size. Therefore, all the conditions need to be optimised to obtain fragments of the desired size within a convenient time frame.

Usually best results were achieved using fragments with mean sizes of 50-250 bp. Complete DNase I digest results in very short fragments and cannot be amplified by subsequent PCR.

To digest the interest gene after DNA purification, following steps were performed.

1. For a total volume of 50 µl, use 5 µl of 10x Digestion buffer and add 0.5-2 µg of each starting DNA
2. Add $dH_2O$ to a final volume of 50 µl
3. Add 0.1 u (1.0 µl) DNase I per µg of starting DNA
4. Depending on the desired fragment size incubate at room temperature for 2 min, 4 min, up to 10 min.
5. Stop the reaction by heat inactivation of the DNase I at 85°C for 10 min.
6. Determine the fragments size by agarose gel electrophoresis and decide the best digestion conditions.

An alternative way is to add DNaseI into PCR-cup after amplification PCR reaction:

1. Prepare 1x digestion buffer:

   0.5M Tris pH7.5 + 100mM $MnCl_2$

2. Dilute the DNaseI to 1U/µl with 1xdigestion buffer

3. Prepare a serial of digestion test by adding 0.5U, 1U, 2U, 4U, 8U DNase I into each tube of 100µl amplification PCR, incubate on thermo-cycler with the following programme:

   25°C 8-10min

   85°C 10min (to stop the reaction)

4. Determine the fragments size by agarose gel electrophoresis and decide the best digestion condition and isolate desired size fragments (section 3.2.1 & 3.2.8).

B. Assembling gene by self-priming (primerless) PCR

One method to assemble a larger number of oligonucleotide fragments to a gene is self-priming PCR (Stemmer et al, 1994). The homologous regions of the synthetic oligonucleotide, which serve as hybridization primers and template, are extended by the polymerase, so that in sequential cycles form the complete gene. Starting with the synthetic oligos or DNaseI digested fragments, the reaction conditions are described in Table 3.7

**Table 3.7** Compositions and conditions for self-priming PCR

| **Reaction components** | | **Reaction conditions** | | |
|---|---|---|---|---|
| Oligo-Mix (10 pmol/ Oligo) or DNA fragments ~100ng/µl | 10 µl | step | Temperature | time |
| Vent-Polymerase | 1µl (2 u/µl) | 1. | 95°C | 2 min |
| | | 2. | 95°C | 30 sec |
| dNTPs (10 mM) | 1 µl | 3. | 40-55°C | 30-45 sec |
| 10 x Vent-buffer | 2 µl | 4. | 72°C | 1-1.5 min |
| | | 5. | 72°C | 5-10 min |
| | | Step 2-4 were cycled 45 times | | |

### 3.2.15 Final PCR

Usually, to obtain enough amount and suitable size specific PCR product, following the self-priming PCR, a final PCR with a pair of internal primers was used to amplify the complete gene.

In nested PCR, two internal primers are used to amplify PCR products from preceding reaction (self-priming PCR, section 3.2.13). This step is performed in the event that previous yields are inadequate or extraneous bands are present. Nested amplification is used to increase fold of amplification and amplification specificity. Conditions for the nested PCR are described in Table 3.8.

The PCR product can be examined by agarose gel electrophoresis (section 3.2.1), and purified with NucleoTrap kit (3.2.8) if necessary.

**Table 3. 8** Compositions and conditions for nested PCR

| **Reaction components** | | **Reaction conditions** | | |
|---|---|---|---|---|
| *primerless-PCR*-product | 3 µl | step | Temperature | time |
| Primer forward (100 pmol/ µl) | 1 µl | 1. | 95°C | 2 min |
| | | 2. | 95°C | 30 sec |
| Primer reverse (100 pmol/ µl) | 1 µl | 3. | 45-60°C | 30-45 sec |
| Taq-Polymerase (homemade) | 1 µl | 4. | 72°C | 1 min |
| 10 x Taq-buffer | 5 µl | 5. | 72°C | 5-10 min |
| dNTPs (10 mM) | 1 µl | Step 2-4 were cycled 30 times | | |
| $H_2O$ | to 50 µl | | | |

### 3.2.16 DNA sequencing

DNA sequencing is based on the chain terminate method (Sanger et al, 1977; Ansorge et al, 1987). It is necessary to use PCR to make a huge number of copies of a gene, so that to have enough starting template.

At the beginning, four fluorescent labeling primers were extended, until it came to the installation of one of the four 2', 3'-dideoxynucleotide in the mixture of 2'-deoxynucleotide randomly, caused specific strand terminate. The fragments were then subjected to gel electrophoretic separation and detected on the basis of their fluorescence.

For the reaction composition, 1-2 µl of purified plasmid DNA (200-400 ng) was mixed with 0.5 µl of appropriate primer (5 pmol), 0.2 µl DMSO, 2 µl of premix (containing modified Taq DNA polymerase, buffer and nucleotides) and adjusted till 8 µl with sterile water. The sequencing reaction was performed according conditions described in Table 3.9

**Table 3. 9** Compositions and conditions for DNA cycle sequencing

| **Reaction components** | | **Reaction conditions** | | |
|---|---|---|---|---|
| DNA-template | 50 ng DNA/ kb Vector | step | Temperature | time |
| Primer (IDR800 labeled) | 1 µl | 1. | 94°C | 2 min |
| | | 2. | 94°C | 15 sec |
| A, C, G, and T-Mix | 2 µl | 3. | *annealing*°C | 20 sec |
| $H_2O$ | to 8 µl | 4. | 70°C | 30 sec |
| "*Chill-out 14*"-wax | 12 µl | 5. | 94°C | 15 sec |
| | | 6. | 70°C | 20 sec |
| | | Step 2-4 and 5-6 were cycled 20 times, respectively. | | |

After the reaction, 1 µl of stop-mix (85% formamide, 25 mM EDTA, pH 8.0, 50 mg/ml Dextran blue) was added to it, followed by incubation at 60 °C for 1h to denature DNA and reduce the sample volume. 0.5-1 µl of this product was loaded on the sequencing gel.

The sequencing gel consisted of following components: 18g urea, 5.3 ml 40% PAGE-Plus (Amresco), 6 ml 10x TBE buffer, 21.5 ml water. The components were well mixed, filtrated and degassed for 10-15 min. Polymerization was started by addition of 250 µl 10% APS and 25 µl TEMED. The gel was polymerized for 1-2h at room temperature.

Separation of the DNA fragments was performed on vertical LEFT-COR DNA Sequencer Model 4000L on gels of the size of 410 x 250 x 0.25 mm for short sequence less than 300 bases or 660 x 250 x 0.25 mm for longer sequence up to 500 bases. Run the gel with 1x running buffer in upper and bottom buffer reservoir, under these conditions: currency 21.0 mA, voltage 46 W, laser power 40 mW, temperature 45°C. Base ImagIR Image Analysis program was used to read the sequence.

However, most of the DNA sequencing in this work was done in Lab of Genome Analyse in Goettingen and Seqlab in Goettingen.

In this case, in stead of fluorescent-labeling primers, the fluorescence labeling 2', 3'-dideoxynucleotide was used as terminator in the thermal cycle reaction (Slatko, 1994), so that 3'-terminal fluorescence-marked single strand fragments were produced randomly during the strand synthesis. These fragments were then subjected to gel electrophoretic isolation and the final nucleotides were detected on the basis of their fluorescence characteristics. In this method, 100--200 ng plasmid DNA is used as template, the primer amount is 5 pmol. The received data were decrypted with the program Chromas (version 1.45).

## 3.3 Phage display and panning method

### 3.3.1 Preparation of M13KO7 helper phage stock

1. Inoculate a single, isolated plaque of helper phage M13KO7 from the working stock plate to 50 ml of 2xYT/Kam medium. Incubate at 37°C overnight (12-16 hr) with shaking at 200 rpm.
2. Transfer the culture to a sterile centrifuge tube. Centrifuge at 2500 x g for 30 min at 4°C (Hettich-Laborzentrifuge, 4000 rpm). Transfer the supernatant to a fresh tube and store at 4°C for up to 1 year.
3. Measure the titer of the M13KO7 stock by plaque formation on a TG1 lawn as described below (section 3.3.2)

### 3.3.2 Titering the M13KO7 helper phage

The helper phage titer may drop during prolonged storage. The helper phage stock should be titered before using and a fresh helper phage stock should be prepared when the titer drops lower than $10^9$pfu/ml.

1. Prepare ten-fold serial dilutions of the helper phage stock in 2xYT medium to $10^{-7}$, $10^{-8}$, and $10^{-9}$ dilution.
2. Transfer 200µl each of the $10^{-7}$, $10^{-8}$, $10^{-9}$ dilutions of the helper phage to sterile test tubes.
3. Add 200µl of prechilled, log-phase TG1 culture and 800µl 2xYT medium to each tube. Mix gently and incubate at room temperature for 5-10 min to allow infection (M13 phage adsorbs rapidly to TG1 cells).
4. Plate 200µl phage and cell mixture onto LB/Kam plates, incubate at 37°C overnight.

5. Count the number of the colonies on the plate having between 30 and 300 colonies.

   The low multiplicity of infection ensures that a host cell will be infected by only one phage. Thus, the titer (cfu/ml) of infected cells correlates to the titer (pfu/ml) of the library.

   pfu/ml = [(# plaques on the plate) x dilution factor]/ (0.2 x 0.2 ml)

7. Use Parafilm to seal one of the plates having distinct plaques and store the plate at 4°C. (This is the working stock plate for helper phage preparation.)

### 3.3.3 Infection of *E. coli* strain with M13 K07 helper-phage, rescue of the recombinant phage-display library and precipitation of M13 phages (Kay BK et al., 1996)

1. Inoculate 1.2ml out of 6 ml bacterial library into 50ml 2xYT medium containing 2% glucose and 100 µg/ml ampicillin, incubate at 37°C(about 1 h) with shaking.
2. As soon as the culture get cloudy (near OD600nm 0.3), super-infect this library culture (about $10^9$ clones) with 0.2 ml M13KO7 helper phage (ca $5x10^9$cfu/ml). Incubate at 37°C without shaking for 30 minutes to allow infection.
3. Remove the glucose by spinning the infected cells at 2500 rpm (Hettich-Laborzentrifuge), 15 min and discarding the supernatant;

   Resuspend the pellet in 400ml 2xYT/Amp/Kam medium, and incubate with shaking at 37°C overnight.
4. Spin the overnight culture at 4000 rpm (Hettich-Laborzentrifuge) for 30 minutes to pellet the cells. (The supernatant contains the recombinant phage.)
5. Supplement the supernatant with 1/5 volume of PEG/NaCl solution, kept on ice for at least 1 hour.
6. Centrifuge supernatant mixture at 11,000g (Sorval, SS34-rotor, 15 000 rpm), 4°C for 15 minutes to pellet phages.
7. Clarify the phages by re-suspending phage pellet in 20 ml TES-buffer, then centrifuging again in Sorval, SS34-rotor, 12 000rpm, 15 min at 4°C (the pellet consists of cells etc)
8. Precipitation again with 4 ml PEG/NaCl, mix well; incubation on ice 1h; then centrifuge: Sorval, SS34-rotor, 15 000 rpm, 30 min, 4°C.

   The phage pellet was resuspended in 1.5ml TES, stored at 4°C up to 2 weeks for proceeding panning; alternatively, add glycerol to 15-20% and store at -70°C.

### 3.3.4 Titering the phage displayed library

Library titer was determined after the phage particles were precipitated with PEG and resuspended in TES.

1. Prepare ten-fold serial dilutions of the phage display library in 2xYT medium to $10^{-6}$, $10^{-7}$, $10^{-8}$, and $10^{-9}$ dilution.
2. Transfer 100 µl of the $10^{-6}$, $10^{-7}$, $10^{-8}$, and $10^{-9}$ dilutions of the phage library to fresh tubes. 2xYT as (-)-control.

3. Add 100 µl of prechilled log-phase TG1 cells to each tube. Mix gently and incubate at room temperature for 5 min to allow infection; M13 phage adsorbs rapidly to TG1 cells.
4. Spread each phage/cell mixture on LB/Amp or 2xYT/amp plate. And incubate plates at 37°C overnight.
5. Count the number of the colonies on the plate having between 30 and 300 colonies.
6. Calculate the titer (cfu/ml) of the infected cells
   cfu/ml = [(# colonies on the plate)x dilution factor]/(0.5x 0.1ml)
   The low multiplicity of infection ensures that a host cell will be infected by only one phage. Thus, the titer (cfu/ml) of infected cells correlates to the titer (pfu/ml) of the library.

### 3.3.5 Panning, phage rescue and superinfection

A. Coating support surface with bait protein

Many different types of substances may be attached to a solid surface for use as bait ligands in a phage display experiment. The method used to attach the ligand to the surface depends on the nature of the ligand.

The matrix used in this work is polystyrene MaxiSorpTM NuncImmuno$^{TM}$ tube. The following is a protocol for coating antibody serum to the immnotube as bait.

1. Coat each NuncImmuno$^{TM}$ tube with 75 µl of anti-HisF serum (equivalent to 0.1-1 µg of protein). Incubate at 4°C overnight (16 hr).
2. Discard or save the coating solution and wash the tubes three times with PBST wash buffer.
3. Add 1.5ml of blocking buffer to each tube. Incubate at room temperature for 2 hr, or use 4°C incubation overnight (16 hr) if the bait protein is unstable at room temperature.
4. Discard blocking buffer and wash the tubes three times with wash buffer.
5. Dry the plate and wrap it in plastic wrap. Store at 4°C for up to 1-2weeks.

B. Panning protocol (modified from Nissim et al., 1994)

1. Add phage-displayed library 1.5 ml (~$10^{11-12}$cfu/ml) to the bait-coated tube. Incubate at room temperature with vibration for 30 min and then 1.5h without vibration. Or incubate at 4 °C for 4 h.
2. Remove the used library and save it for titering of unbound library.
3. Wash the tube surface 5 times with wash buffer (1-2ml each wash).

Note: use more washes (up to 20 times) for higher stringency binding conditions.

C. Eluting and rescuing trapped phages

1. Elute the panning tube with 0.5ml elution buffer, transfer the elution buffer into sterile 15ml tube; add 8.5 ml of pre-chilled, log-phase TG1 cells into the elution buffer in the 15ml tube; add 1 ml of pre-chilled, log-phase TG1 cells into the panning tube.
2. Incubate elution and cell mixture at 37°C for 30 min. Combine the infected TG1 cells, immediately place on ice or 4°C.

Note: Longer incubation at 37°C will allow the infected TG1 cells to duplicate and will make it difficult to use cell number to estimate percent of phage rescued.

This is the 1st **round enriched TG1 library**.

3. Dilute a small aliquot of the enriched TG1 library and plate it out to determine the titer and estimate recovery of trapped phage.
4. Store the enriched TG1 library at 4°C for up to 3 days, or add sterile glycerol to a final concentration of 15%, and store at –70°C for up to 1 yr.

   If the next round panning was planned within 3 days, keep 2.5ml of the enriched TG1 library at 4°C and store the rest at –70°C.

D. Preparation of enriched library for nest panning

1. Add the 2.5ml aliquot of enriched TG1 library to 50 ml of 2x YT/glucose medium. Incubate at 37°C for 1 hr with shaking at 250 rpm. Add ampicillin to a final concentration of 100 μg/ml.
2. Immediately add 5x $10^8$ pfu of M13KO7 to the cell culture for infection. Incubate at 37°C for 30min.
3. Repeat the step 3 –step 8 which are described in 3.3.3 for the preparation of phage-display library.

## 3.4 Protein biochemical methods

### 3.4.1 Protein concentration determination by UV absorption

Amino acid Tryptophan, Tyrosine, phenylalanin and disulphide bridges (Cysteine) absorb light within a range from 250 to 300 nm. On this basis, a formula for calculation of the molar extinction coefficient of proteins at 280nm was set up (Pace et al. 1995):

$$\varepsilon_{280}(M^{-1}cm^{-1}) = \sum Trp \cdot 5500 + \sum Tyr \cdot 1490 + \sum Cystine \cdot 125$$

According to the Lambert-Beer law, the protein concentration in a solution can be determined over the computed molar extinction coefficient and the measured absorption:

$$A = \varepsilon \cdot c \cdot d \quad \Leftrightarrow \quad c = \frac{A}{\varepsilon \cdot d}$$

A = Absorption
c = concentration (M)
d = cell lenth (cm)

### 3.4.2 IPTG induction for protein expression

Expression of the interest protein was carried out by using of pET_B001 Vector which contains a inducable T7 promoter, IPTG (Isopropyl-beta-D-thiogalactopyranoside) served as inductor.

The desired clones were cultured over night in 3 ml appropriate selective medium at 37°C. Dilute the cells in the rato of 1:100 with medium and shake at 37°C. When $OD_{600}$ reached 0.6-0.8, IPTG was added to a final concentration of 1 mM and the cells were further cultured at 30°C for 3 h, or alternatively, cultured at room temperature overnight for the protein expression. The culture was centrifuged at 4000rpm

(Hettich-Laborzentrifuge) for 30min at 4°C to harvest the cells and subjected to the next processing step.

### 3.4.3 Soluble protein fraction from host cell

1. The harvested cell pellets were washed with 20ml resuspend buffer (25mM HEPES, 0.5M NaCl, pH7.6), spin to pellet the cells and freeze at -70°C for 30min.
2. Thaw the cells at room temperature for 15-20min, resuspend the thawed cells in 20ml resuspend buffer.
3. Sonicate the cells on Branson Sonifier R-250 (Branson Sonic Power Company, USA) at the following conditions:

   output: 5;  cycle: 50%;  time: 2min*3
4. The cell mixture was then subject to cell disruptor (Constant Cell Disruption Systems, Constant Systems Ltd, Daventry, England), the pressure was set at 1.8Kpa.
5. Centrifuge to remove cell membrane and debris: 12000rpm, 4°C (Sorvall RC-5C Rotor type SS34, Bad Nauheim, Germany)

   The obtained soluble protein fraction then was subjected to purification, eg. IMAC purification (section 3.4.4)

### 3.4.4 Immobilized metal affinity chromatography (IMAC)

Histidine has a high affinity for bivalent transition metals e.g. $Ni^{2+}$. Proteins fusion with oligo-Histidine-Peptide can be purified by affinity chromatography with immobilized metal ions.

Pack 4 ml Chelating Sepharose$^{TM}$ Fast Flow (Amersham Pharmacia Biotech) into an empty column, charge the packed column with 2 column volumes 100 mM $NiCl_2$. Washed the column with five column volumes water to remove surplus nickel, and then equilibrate with three volumes starting buffer.

The protein solution was then loaded on the $Ni^{2+}$column, nonspecific bound molecules were removed by washing with three volumes starting buffer (25mM HEPES, 0.5M NaCl, pH7.6).

Elute the column with IMAC elution buffers, which were step-size incremental concentrations of imidazole in 0.5 M NaCl, 25 mM HEPES, pH 7.6. $His_6$-tag fusion protein then was displaced off from the column.

The eluate was analyzed on SDS-PAGE (section 3.4.5), appropriate fractions were pooled, desalted and concentrated 20-fold by ultrafiltration (Millipore Centriprep cartridge, molecular weight cut-off ca. 10000Da). The resulting solution can be diluted with an equal volumn of glycerol and add DTT to a final concentration of 1 mM, store at -20 °C.

### 3.4.5 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were analyzed over SDS Polyacrylamide gel electrophoresis (SDS PAGE) (Laemmli, 1970). By additive of the detergent sodium dodecyl sulfate (SDS), the proteins are denatured and the self-charges of the proteins are masked, so that the proteins are separated in electrophoresis according to their molecular weight.

Separation gels (resolving gel) consist of 10-15% (w/v) acrylamide/bis acrylamide, 187.5 mM of TrisHCl pH8.8 and 0.1% SDS. Collecting gels (stacking gel) consist of 5% acrylamide/bis acrylamide, 125 mM of TrisHCl pH6.8 and 0.1% SDS. The gel's dimension is 100 x 60 x 0.75 mm. The composition of the gels is shown in Table 3.10.

**Table 3.10** Solutions to prepare five pieces 10-15% discontinuous SDS-polyacrylamide gels

| | **Resolving gel (%)** | | | **Stacking gel (%)** |
|---|---|---|---|---|
| | 10% | 12.5% | 15% | 5% |
| Protogel (30% acrylamide) | 13.3ml | 16.7ml | 20ml | 2.7ml |
| $dH_2O$ | 18.7ml | 15.3ml | 12ml | 11.5ml |
| 1M Tris HCl pH8.8 | 7.5ml | | | 1.6ml 1.25M Tris pH6.8 |
| 10% SDS | 0.4ml | | | 0.16ml |
| Following components should be added into gel solution just before pouring the gel | | | | |
| 10% APS (fresh prepared) | 127.5 µl | | | 60 µl |
| TEMED | 20 µl | | | 10µl |
| Volume | ~40ml | | | ~16ml |

The gel casting includes two steps. First, casting the resolving gel, isopropanol was added onto the top to give a level surface. After polymerisation of resolving gel, the isopropanol was decanted and the gel was washed carefully. Then, cast the stacking gel on the top of resolving gel.

The proteins were loaded with 1/2 volume protein sample buffer (loading buffer) and denatured 5 min at 98 °C. The proteins were isolated with a constant current of 20 mA for one piece of gel or 40mA for two pieces of gel in SDS-PAGE running buffer.

After electrophoresis, the protein bands were visualised by dipping the gel into fresh "Coomassie Brilliant Blue" staining solution (section 2.3) for 2-3 h (if the staining solution is used one, dipping time should be longer, e.g. overnight). Alternatively, after the electrophoresis, proteins can be transferred from the gel onto the nitrocellulose membrane for the immunochemical detection.

### 3.4.6 Enzymatic activity assay

The enzyme activity was measured by coupling enzyme assay in this work. This assay system allows the accurate monitoring of initial rate data (Dobson R CJ et al., 2004). DHDPS activity was measured through the reduction of its catalytical product dihydrodipicolinate by NADH and DHDPR. The consumption of NADH is monitored spectrophotometrically at 340 nm. The initial velocity data were collected at a constant temperature of 20°C on a two-ray spectrophotometer Uvikon 930. All assays were performed in a volume of 100 µl buffer of 100 mM Hepes (pH 7.0 and 8.0).

Kinetic parameters were calculated by fitting the initial velocity data at different substrate concentrations to the Michaelis–Menten equation by the Lineweaver-Burke plot.

### 3.5 Bioinformatics methods

DNA sequencing chromatograms data was retrieved and read with Chromas© Version 1.45 on PC stations.

For the analysis and processing of DNA and protein sequence data, the GCG Wisconsin Package (Version 10.0) (Genetic Computer Group, Genetic Computer Group (GCG), Madison, Wisconsin, USA) was used.

For the diagram and modeling of protein structures, The PyMOL Molecular Graphics System (2002) (DeLano, W.L. http://www.pymol.org ) and Swiss PdbViewer© Version 3.7 (Glaxo Welcome Experimental Research) were used.

Multiple sequence alignments were generated by Vector NTI Suite 8 (InforMax, 2002). Alternatively, ClustalW program and ClustalX were used.

Sequence homologues searching was performed by BLAST algorithm (Basic Local Alignment Search Tool). Databases are from National Centre of Biotechnological Information (NCBI, http://www.ncbi.nlm.nih.gov/).

The other programs used in this work were listed in section 2.11.

# 4 Results and discussions

## 4.1 Combinatorial construction of thisF gene libraries

### 4.1.1 Aim and basis of the gene libraries' synthesis

The aim of the project is to produce tHisF variants with novel catalytic properties. For this purpose, gene libraries should be constructed first. These gene libraries could be displayed on phage surface for potential use in panning against transition state analogs (TSA), or they could be used to generate soluble tHisF variants in genetic complementation.

The construction of gene libraries requires nucleotide sequence of the scaffold. Following the introduction in section 1.5, the thermostable $(\beta\alpha)_8$ barrel structure enzyme—tHisF was chosen as the scaffold in this project, its nucleotide sequence and amino acid sequence are shown in Fig 4.1.1. In the design of the oligonucleotide fragments, some of the codons in t*his*F gene were changed for the host *E. coli* cell based on *E. coli* codon usage table.

```
    ATGCTCGCTAAAAGAATAATCGCGGCTCTCGATGTGAAAGACGGTCGTGTGGTGAAGGGA
1   ---------+---------+---------+---------+---------+---------+ 60
    TACGAGCGATTTTCTTATTAGCGCCGAGAGCTACACTTTCTGCCAGCACACCACTTCCCT
    M  L  A  K  R  I  I  A  A  L  D  V  K  D  G  R  V  V  K  G

    ACGAACTTCGAAAACCTCAGGGACAGCGGTGATCCTGTCGAACTGGGAAAGTTCTATTCC
61  ---------+---------+---------+---------+---------+---------+ 120
     TGCTTGAAGCTTTTGGAGTCCCTGTCGCCACTAGGACAGCTTGACCCTTTCAAGATAAGG
    T  N  F  E  N  L  R  D  S  G  D  P  V  E  L  G  K  F  Y  S

    GAAATTGGAATAGACGAACTCGTTTTTCTGGATATCACCGCGTCCGTTGAGAAGAGGAAA
121 ---------+---------+---------+---------+---------+---------+ 180
    CTTTAACCTTATCTGCTTGAGCAAAAAGACCTATAGTGGCGCAGGCAACTCTTCTCCTTT
    E  I  G  I  D  E  L  V  F  L  D  I  T  A  S  V  E  K  R  K

    ACCATGCTGGAACTGGTCGAAAAGGTGGCCGAGCAGATCGACATTCCGTTCACTGTTGGA
181 ---------+---------+---------+---------+---------+---------+ 240
    TGGTACGACCTTGACCAGCTTTTCCACCGGCTCGTCTAGCTGTAAGGCAAGTGACAACCT
    T  M  L  E  L  V  E  K  V  A  E  Q  I  D  I  P  F  T  V  G

    GGAGGTATCCACGACTTCGAAACGGCCTCGGAGCTCATTCTCCGTGGTGCGGACAAGGTG
241 ---------+---------+---------+---------+---------+---------+ 300
    CCTCCATAGGTGCTGAAGCTTTGCCGGAGCCTCGAGTAAGAGGCACCACGCCTGTTCCAC
    G  G  I  H  D  F  E  T  A  S  E  L  I  L  R  G  A  D  K  V

    AGCATAAACACGGCGGCTGTGGAGAATCCTTCTTTGATCACACAGATCGCTCAAACTTTT
301 ---------+---------+---------+---------+---------+---------+ 360
    TCGTATTTGTGCCGCCGACACCTCTTAGGAAGAAACTAGTGTGTCTAGCGAGTTTGAAAA
    S  I  N  T  A  A  V  E  N  P  S  L  I  T  Q  I  A  Q  T  F

    GGGAGTCAGGCCGTTGTCGTGGCGATAGATGCAAAAAGAGTGGATGGAGAGTTCATGGTC
361 ---------+---------+---------+---------+---------+---------+ 420
    CCCTCAGTCCGGCAACAGCACCGCTATCTACGTTTTTCTCACCTACCTCTCAAGTACCAG
    G  S  Q  A  V  V  V  A  I  D  A  K  R  V  D  G  E  F  M  V
```

```
    TTCACCTACTCCGGAAAGAAGAACACGGGCATACTTCTGAGAGACTGGGTGGTTGAAGTA
421 ---------+---------+---------+---------+---------+---------+ 480
    AAGTGGATGAGGCCTTTCTTCTTGTGCCCGTATGAAGACTCTCTGACCCACCAACTTCAT
    F  T  Y  S  G  K  K  N  T  G  I  L  L  R  D  W  V  V  E  V

    GAAAAGAGAGGAGCAGGAGAGATCCTGCTCACCAGTATCGACAGAGACGGCACAAAATCG
481 ---------+---------+---------+---------+---------+---------+ 540
    CTTTTCTCTCCTCGTCCTCTCTAGGACGAGTGGTCATAGCTGTCTCTGCCGTGTTTTAGC
    E  K  R  G  A  G  E  I  L  L  T  S  I  D  R  D  G  T  K  S

    GGTTACGATACGGAGATGATAAGGTTCGTGAGGCCACTAACCACACTTCCCATCATCGCT
541 ---------+---------+---------+---------+---------+---------+ 600
    CCAATGCTATGCCTCTACTATTCCAAGCACTCCGGTGATTGGTGTGAAGGGTAGTAGCGA
    G  Y  D  T  E  M  I  R  F  V  R  P  L  T  T  L  P  I  I  A

    TCCGGTGGTGCGGGAAAAATGGAACATTTCCTTGAAGCCTTCCTGGCAGGTGCCGACGCT
601 ---------+---------+---------+---------+---------+---------+ 660
    AGGCCACCACGCCCTTTTTACCTTGTAAAGGAACTTCGGAAGGACCGTCCACGGCTGCGA
    S  G  G  A  G  K  M  E  H  F  L  E  A  F  L  A  G  A  D  A

    GCCCTTGCGGCTTCTGTCTTTCACTTCAGAGAGATCGACGTGAGAGAACTGAAAGAGTAC
661 ---------+---------+---------+---------+---------+---------+ 720
    CGGGAACGCCGAAGACAGAAAGTGAAGTCTCTCTAGCTGCACTCTCTTGACTTTCTCATG
    A  L  A  A  S  V  F  H  F  R  E  I  D  V  R  E  L  K  E  Y

    CTCAAAAAACACGGAGTGAACGTGAGACTGGAGGGGTTG
721 ---------+---------+---------+--------- 759
    GAGTTTTTTGTGCCTCACTTGCACTCTGACCTCCCCAAC
    L  K  K  H  G  V  N  V  R  L  E  G  L
```

**Fig. 4.1.1** The nucleotide sequence and amino acid sequence of tHisF, a cysteine at the ninth position of the amino acid sequence was substituted by alanine (in bold) to avoid the potential disulfide bond formation and other potential oxidation of cysteine. The amino acid sequence is in single letter code.

### 4.1.2 The starting situation

As introduced in section 1.2.3.2, there are a range of methods available for construction of randomized gene libraries. In order to generate a repertoire at a possible and reasonable size, a controlled partially randomization strategy was used in this project. The basic design work had been done previously by S. Schiller (this lab) as skeletonized in the following sub-paragraphs 4.1.2.1 to 4.1.2.3.

#### 4.1.2.1 Choice of amino acid residues to be randomized

To select suitable amino acids for randomization to construct the gene library, a detail structure information which is derived from modeling the tHisF with its substrate PRFAR (N'-[(5'-phosphoribulosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide) (provided by M. Hennig, personal communication) based on the X-ray structure of tHisF (Lang et al., 2000) was used. Half of the circle of tHisF barrel (strand β6, β7, β8 and neighboring strand β1) was presumed to be catalytic residues rich region. Only this half circle was chosen for the first round of randomization.

To synthesize the t*his*F gene library, nine residues that were presumably involved in substrate binding and/or catalysis were selected for mutagenesis. All of them are

residues of tHisF within a radius of approximate 5 Å around the PRFAR. Two presumed phosphate binding residues in the amino terminal of the protein were removed and replaced with alanine codon (N103A, T104A) hoping to alter the substrate specificity for evolving novel property from this scaffold. The amino acids of the presumed phosphate binding sites in carboxyl terminal of protein (D176, G177, G203, A224, and S225) were kept intact. Amino acids whose side-chain pointed toward the inside of the barrel were excluded from the selection, since these residues could participate in the folding of the protein, hence are relevant to the structure stability of the protein. Only their neighboring residues were selected for randomized, they were: C/A9, D11, L169, T171, I173, K179, S201, A204 and L222. These residues were located at β1 strand (C/A9, D11), β6 strand (L169, T171), loop β6/α6 (I173, K179), β7 strand (S201), loop β7/α7 (A204) and β8 strand (L222), respectively (Figure 4.1.2).

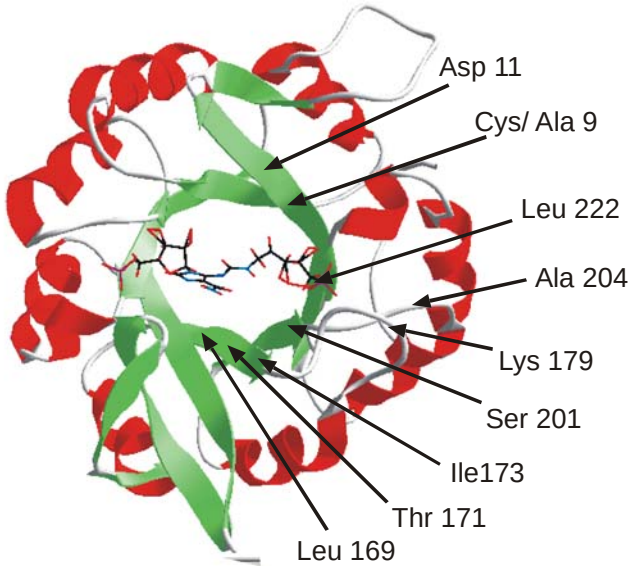



**Figure 4.1.2** tHisF protein (in ribbon) with modeled substrate 5'ProFAR (in sticks) (Hennig M, personal communication).
Black arrows point to those positions, which were selected for mutagenesis in the synthesis of controlled partially randomized t*his*F gene library. (Schiller S 2004)

### 4.1.2.2 Choice of mutation degree

Based on the empirical experience, the optimal number of exchanged amino acids which may achieve an ideal effect for enzyme engineering is between 4 and 6 in each protein molecule. The theoretical distribution of the mutant extent x (0 to n exchange) and its frequency $P_x$ (fraction in the total repertoire) can be calculated according to the binomial distribution:

$$P_x = \frac{n!}{(n-x)!\,x!} \bullet p^x \bullet (1-p)^{n-x}$$

n: number of total randomized positions

x: any integer from 0 to n

p: mutation rate at each randomized position

$P_x$: fraction of molecules having x exchanges

To generate t*his*F gene libraries, the number of total randomized positions (n) was set at nine. The binomial distribution was used to determine a suitable ratio of resident codon for the libraries. Fig 4.1.3 shows a comparison of the distribution of the different mutation degree, when p (ratio of non-resident codons used in trinucleotides mixture for codon exchanging at randomized position) is 0.5, 0.55 or 0.6, then the average codon exchange number x is 4.5, 5.0 and 5.5.

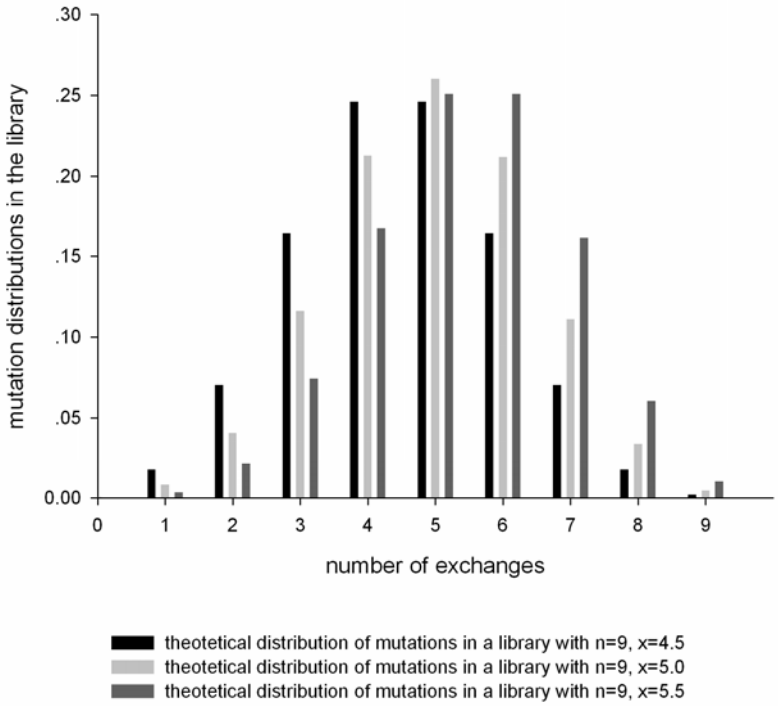



**Fig 4.1.3** Comparison of the distribution of the different mutant degree in libraries. If the number of total randomized positions (n) is 9, when the mutation rate (p) at randomized position is 0.5, 0.55 or 0.6, the average condon exchange number (x) is 4.5, 5.0 and 5.5.

For the synthesis of the partially randomized oligonucleotide fragments, defined mixtures of trinucleotides, which consist of the resident and non-resident codons were used. The codon coding for Cys was excluded from the trinucleotide mixture to avoid the potential oxidization of cysteine. If the resident and non-resident codons used in each randomized residue were in a molar ratio of 0.44 and 0.56, the exchange probability per position (p) would be 0.56, which would result in oligonucleotide products with the average of five exchanged codons in nine randomized positions. To simplify the sample handling in the synthesis, a mixture with equal mole of all 19 trinucleotides (including all the respective wild-type codons at any randomized position, the codon coding for Cys was changed to codon coding for Ala) was used at the molar ratio of 56%, this increased the molar ratios of all resident codon by 1/19 at each randomized position, and the molar ratio of non-resident codons were decreased accordingly, the actual molar ratio of resident and non-resident codons in each case were 0.47 and 0.53, and the theoretical average exchange number would be 4.77.

### 4.1.2.3 The set of synthetic oligonucleotides

To insert random mutations into the t*his*F gene, partial randomized oligonucleotides should be developed. The DNA sequence of tHisF was divided into 38 fragments (Fig. 4.1.4). The appropriate oligonucleotides possess an average length of 45.7 bases - the longest sequence measured 63 bases and the shortest 21 bases. All fragments were designed in such a way, that (1) they overlapped to each other at least 15 base pairs; (2) all oligonucleotides possessed 5'-phosphate group except fragment 01 and 38, so that it is possible to interconnect them after hybridizing the complementary regions by ligation; (3) with the help of the program *Primerfinder* (University of Texas Southwestern Medical Center) the sequences of the oligonucleotide were modified, so that formation of secondary structures such as dimer and/or hairpins, which in both cases are based on partially homologous sequence, were prevented.

As shown in Fig.4.1.4, four of the twenty oligonucleotides, which should form the coding strand of the t*his*F gene variants (upper line of the figure), carried the codons selected for randomization. The randomized codons were used only in the coding strand, since this reduced the expenditure of synthesis. Therefore these codon regions were single-strand, the corresponding regions in the eighteen oligonucleotides for the non-coding strand (the bottom line of the figure) were left uncovered. The detail of these oligonucleotides sequence is illustrated in Fig 4.1.5.

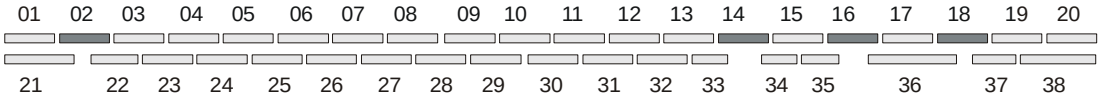



**Fig. 4.1.4** Assembly of *his*F-Gene from 38 synthetic oligonucleotides, the overlap region is at least 15 bp. In the dark fragments, randomized codons were incorporated in, the number of randomized codons contained in each oligo is in bracket following the name of each oligo: 02-mut (2), 14-mut (4), 16-mut (2), 18-mut (1). (ref. section 2.9.4)

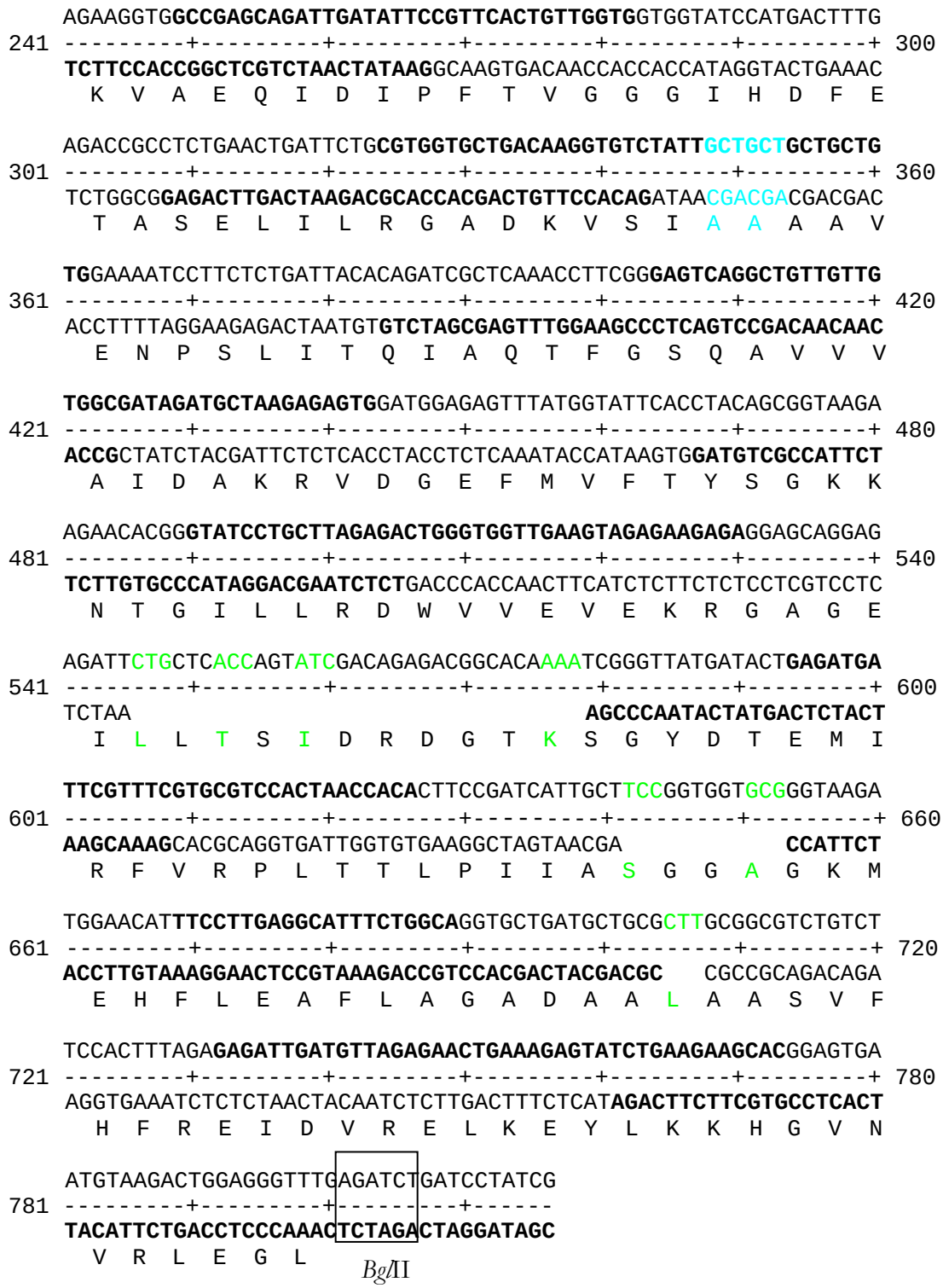
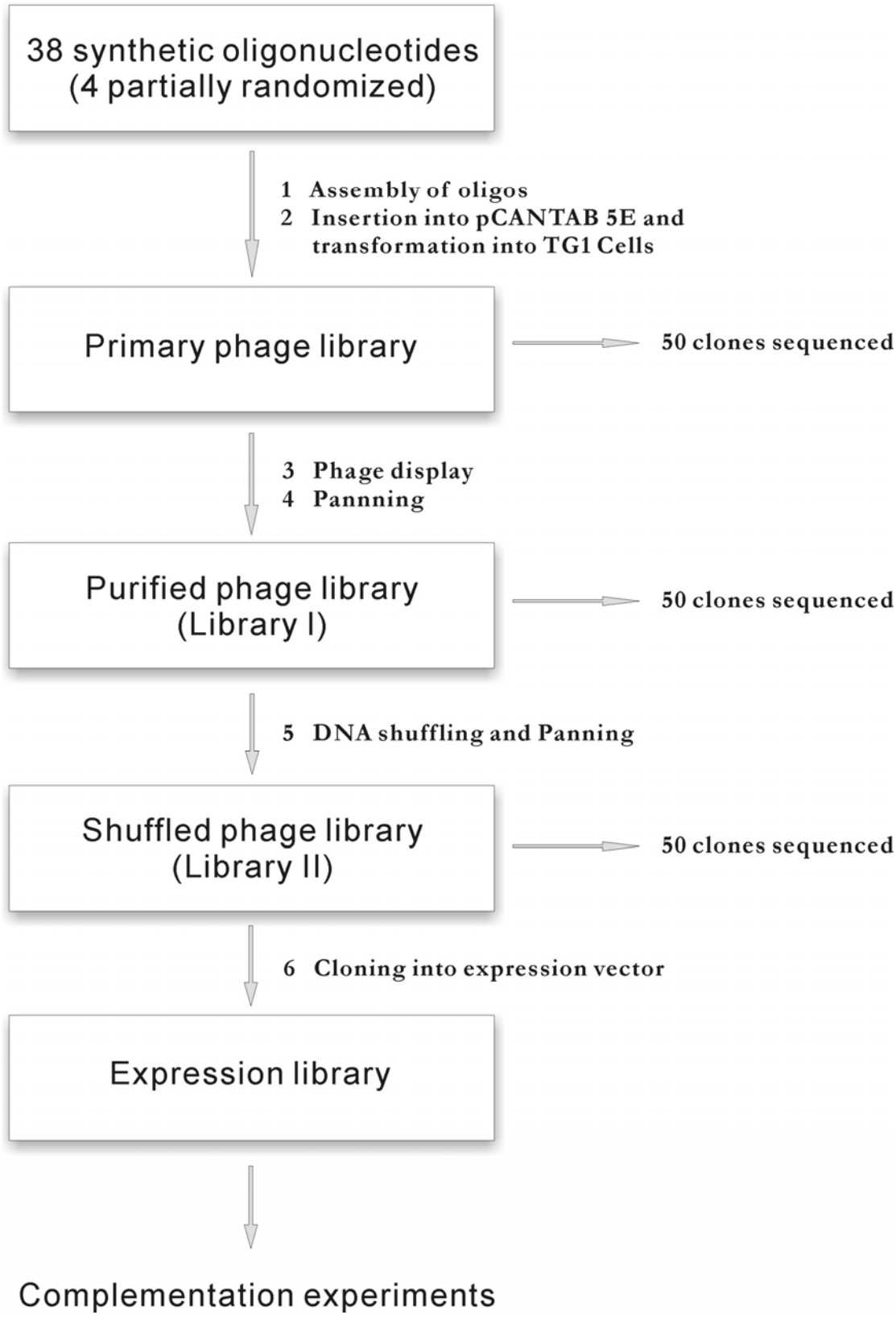



**Fig. 4.1.5** Nucleotide sequences of the synthetic oligonucleotides were built up from *this*F gene and the amino acid sequence derived from the $^{synth}$*this*F is in the single letter code. The oligonucleotides (see section 2.9.3) st01, st03, st05, st07, st09, st11, st13, st15, st17, st19, st22, st24, st26, st28, st30, st32, st34, st36, st38 are in bold. The codons of the amino terminal phosphate binding site N103 and T104 were exchanged to alanine codon (in cyan). The spacer (for successful display of tHisF on M13 phage, Schiller S 2004) between signal sequence and *this*F gene is in red, the randomized codons are highlight in green. The *Sfi*I and *Bgl*II restriction site incorporated in the fragmjents st01/st21 and st20/st38 are in rectangle boxes. L2 represents the second amino acid Leu in tHisF-sequence.

### 4.1.3 Road map of the t*his*F libraries construction

The procedures for the construction of t*his*F gene libraries are shown in scheme 4.1.1.




**Scheme 4.1.1** Construction of t*his*F gene library from synthetic oligonucleotide fragments

### 4.1.4 Construction of gene library from synthetic oligonucleotide fragments

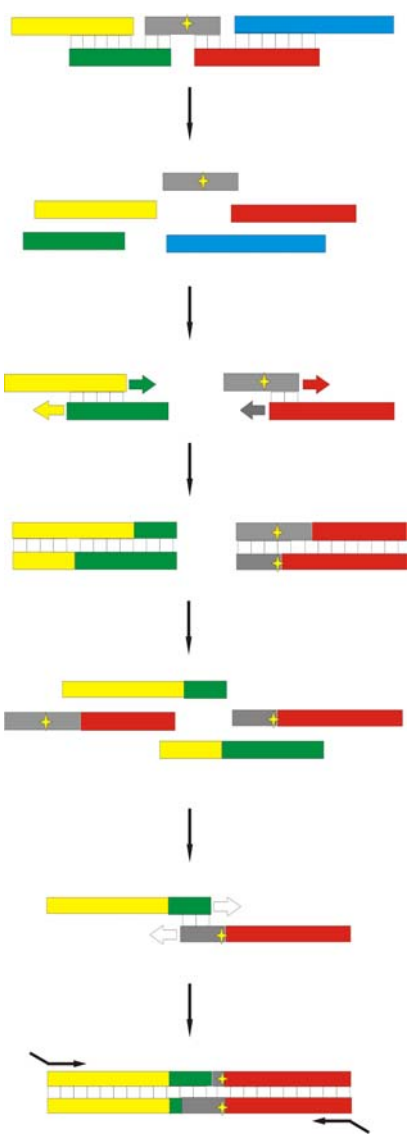
#### 4.1.4.1 Assembly strategy

The construction of gene libraries targeted a controlled level of randomization of nine specific positions within the t*his*F gene sequence. These positions were not randomized saturatedly, the respective wild-type codons were added as a portion in the tri-nucleotide mixtures. The key to this library construction method is the introduction of diversity at specific positions at codon level within the synthetic DNA fragments with pre-synthesized tri-nucleotides. The first step of the library construction is the direct synthesis of DNA molecules whose sequence design is guided by information of target gene (section 4.1.2.3), and then to incorporate the synthetic oligonucleotides into full genes via PCR.

To assemble the oligonucleotides into t*his*F gene library, several methods had been tried before (Schiller S 2004), e.g. (1) hybridize and ligate the overlapped oligonucleotide fragments to form full t*his*F genes; (2) assemble oligonucleotides by ligation on a scaffold of ssDNA (ligase chain reaction, LCR); (3) assemble t*his*F genes from PCR-fragments. However, there were always some problems in these methods, e.g. the first method (ligation of the overlapped oligonucleotides) and the second method (LCR) were only possible in theory and never succeeded in this lab; the third method is laborious and time and materials consuming.

The method which was chosen in this work is based on the self-priming PCR and shuffling PCR (Stemmer WP et al., 1994). This method relies on the assembly of oligonucleotides via self-priming PCR, controlled overlapping of these oligonucleotides leads to the full-length product with high recombination frequency.

As shown in Fig 4.1.6, the designed overlapping regions of the oligonucleotides served as primers, whereby the genes were constructed gradually over repeated cycles of separation, annealing and polymerase reactions. The complete genes then were amplified in primed final PCR with primers which carried restriction sites for cloning the assembled products into a suitable vector. Similar approaches had been reported in some publications, e.g. synthesis shuffling (Ness JE et al., 1999 & 2002); assembly of designed oligonucleotides (ADO, Zha D et al., 2001), etc.



1. Designed overlapping oligonucleotides (gray fragments carry random trinucleotides, indicated by a star)

2. Separating oligos by heating them to 90-95$^{o}$C

3. Annealing and polymerase chain reaction

4. First round PCR production

5. Separating again

6. Annealing and polymerase chain reaction again

7. Primed final PCR to amplify the chimerically assembled gene, the primers carry the restriction sites.

**Fig 4.1.6** Schematic diagram of gene synthesis by self-priming PCR using designed overlapping oligonucleotides

### 4.1.4.2 Assembling of oligonucleotides by self-priming PCR

(This paragraph comprises step1 of Scheme 4.1.1)

Using a mixture of 38 chemically synthesized oligonucleotide fragments that were designed according to *his*F gene from *Thermotoga maritima* (Fig 4.1.4 and Fig 4.1.5) and the length ranges between 21-63 bases (these oligonucleotides were ordered from Purimex (Goettingen, Germany) in 'HPLC-purified' grade), the complete genes were assembled by self-priming PCR (section 3.2.13). In order to introduce mutations into the t*his*F genes, the randomized oligos 02-mut, 14-mut, 16-mut and 18-mut (section 2.9.4) were added into the self-priming PCR reactions.

The self-priming PCR was carried out by using the mixture containing 50pmol each of the 38 oligonucleotides, 1µl 10mM dNTPs mixture with 2µl Vent polymerase (10u/1µl) in 20µl reaction volume. The PCR reaction programme consists of an initial

denaturation step for 2 min at 95$^{o}$C followed by 50 cycles of 95 $^{o}$C/30 sec, 43$^{o}$C/30 sec, 72 $^{o}$C/1 min and an additional incubation of 10 min at 72 $^{o}$C (ref. section 3.2.14).

A final PCR was applied to amplify the desired products using 1.5 µl assembled DNA products as templates. The primers for the amplification were oligonucleotide VB 01 containing a *Sfi*I site (in bold)(5'-GTCAGCATGC**GGCCCAGCCGG(CC)-**3') as a 5'-primer, and oligonucleotide VB 38 containing a *Bgl*II site (in bold) (5'-CGATAG GATC**AGATCT**CAAACCC-3') as a 3'-primer. 1µl 10 mM dNTPs mixture and 1µl Taq polymerase (homemade) were used in a 20µl reaction. The program for the final PCR is step 1, 95°C, 1 min; step 2, 95°C, 30 s; step 3, 50°C, 30 s; step 4, 72°C, 1 min; step 5, 72°C, 10 min; steps 2 to 4 were repeated 30 times (ref. section 3.2.15). The result of assembling of oligonucleotides and the final PCR product for t*his*F gene library are shown in Fig. 4.1.7.

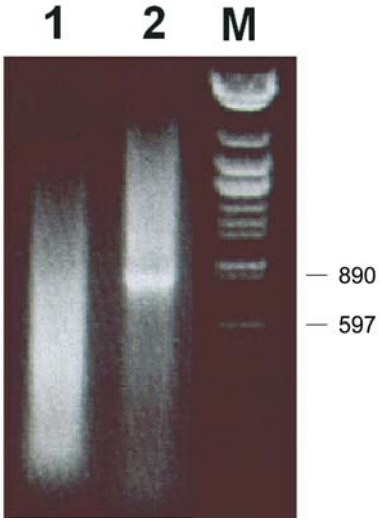



**Fig. 4.1.7** Assembling of oligonucleotides by self-priming PCR analyzed by agarose gel electrophoresis

1 Product of assembly PCR
2 Product of final PCR
M DNA molecular weight marker, Lambda DNA/Eco47I (AvaII)

### 4.1.5 Primary phage library

#### 4.1.5.1 Cloning t*his*F gene into phage display vector

(This paragraph comprises step2 of Scheme 4.1.1)

The PCR-amplified ensembles of t*his*F genes were digested with *Sfi*I and *Bgl*II, purified by agarose gel extraction and ligated into the corresponding *Sfi*I and *Bgl*II sites of phagemid vector pCANTAB 5E between its signal sequence and gIIIp gene to get the t*his*F gene phagemid library (Fig. 4.1.8, for protocols, see section 3.2.6-3.2.9).

After the construction of t*his*F gene library in phagemid vector, it was transformed into a host strain TG1(*SupE*) (section 3.1.7). An initial bacterial stock of the primary library was plated out on LB-Amp plates. The total independent clones of the t*his*F gene in pCANTAB 5E vector (primary library) are ca. 3.3x10$^{6}$, according to the titer plates (section 3.1.8).

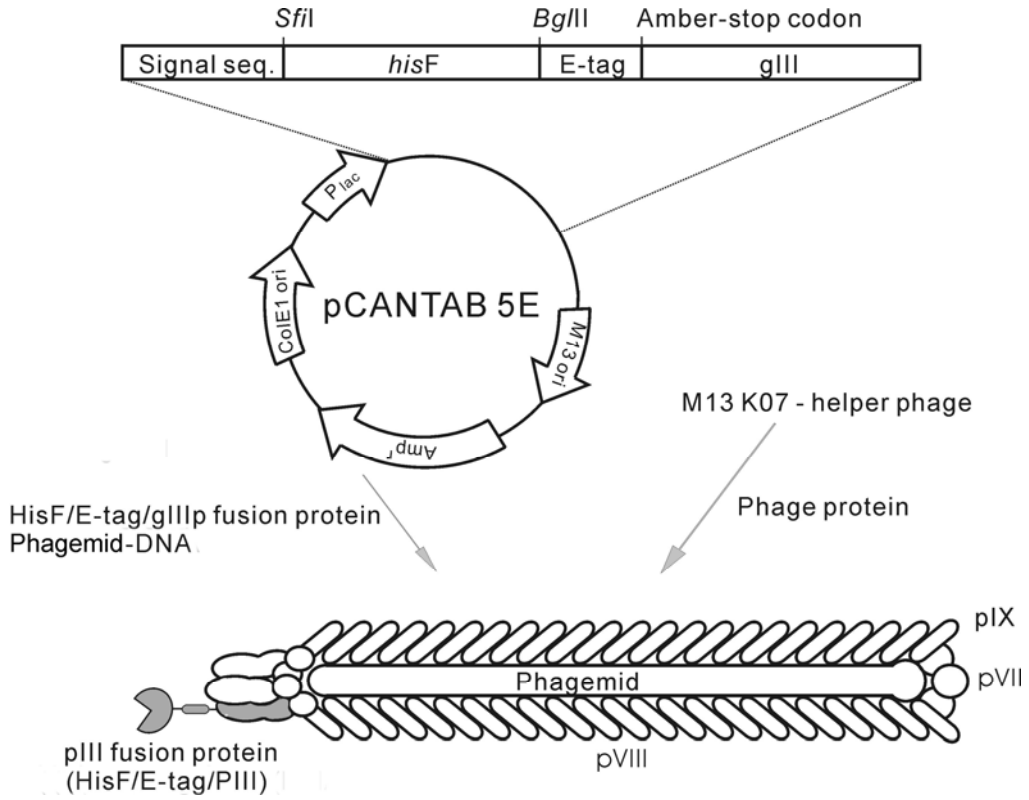



**Fig. 4.1.8** Schematic diagram of the construction of pCANTAB 5E-t*his*F phagemid library and M13 phage display

### 4.1.5.2 Colony screening PCR and characterization of primary library by DNA sequence analysis

To confirm the successful assembly of t*his*F gene and the successful ligation of the t*his*F gene into vector pCANTAB 5E, colony screening PCR (section 3.2.11) was performed to identify the clones containing the correct insertion. Two primers were used, one is specific for the inserted gene, and the other is specific for the vector (Fig.4.1.9). The two primers were: VB-03, 5'-GGC ACT AAC TTT GAG AAC CTG-3' for *his*F gene and sh-up, 5'-CAC GCC CTC ATA GTT AGC GTA-3' for pCANTAB 5E vector (section 2.9.2; protocol in section 3.2.11), the screening result revealed that ~60% clones contained plasmid pCANTAB 5E-t*his*F-variants.

To characterize the primary library, phagemid DNAs of independent clones from the primary library were prepared using the method of DNA-mini preparation (section 3.2.4), and subjected to DNA sequencing.

The sequence analysis (fifty clones) revealed a high ratio of base deletion and insertion in the DNAs from the primary library, only ~8% of sequenced DNAs from the primary library were complete t*his*F-variants without base deletion or insertion, this result revealed the disadvantage of the chemical synthetic method for the gene construction. To solve the problem of high ratio of base insertion and deletion in the primary library, additional purification step became necessary (section 4.1.5.4).

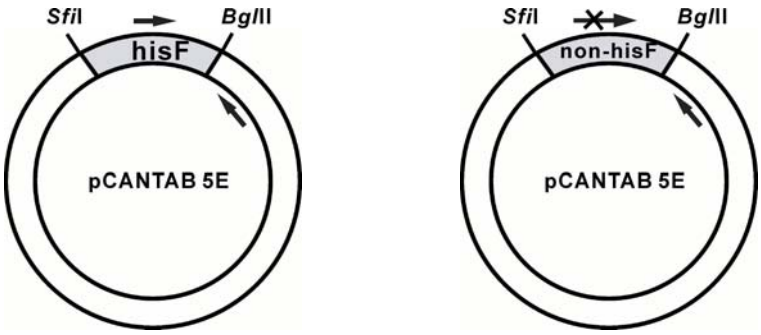



**Fig.4.1.9** Colony screening strategy for determination of the correct insertion of *his*F gene in a phagemid vector. A pair of primers was designed with one specific for *his*F gene and the other specific for the vector. The PCR product was checked by agarose gel electrophoresis.

### 4.1.5.3 Phage display

(This paragraph comprises step3 of Scheme 4.1.1)

In construction of primary t*his*F gene library with chemically synthesized oligos, a phagemid vector pCANTAB5 E was used (Fig. 4.1.8). pCANTAB 5E contains an M13 replication origin, so, in the presence of helper phage, phagemid DNA is replicated in the normal M13 fashion. pCANTAB also carries the ColE1 replication origin and ampicillin resistance gene ($Amp^r$) for growth and selection of transformed host cells. The gene III coding region permits expression of the foreign proteins as N-terminal pIII (gene III coat protein) fusions in host strain TG1(*Sup*E) and the genetic information is packaged thanks to the packaging signal (Fig. 4.1.8).

The initial bacteria stock library which was generated from transformed TG1 cells contained ca $3.3x10^6$ independent clones (section 4.1.5.1). After the infection of the host cells with the phagemid library under the help of a helper phage M13K07, phage particles were prepared as described with some modification, normally yield ca. $10^{11}$-$10^{12}$ pfu/ml (for protocol, see section 3.3.3). The foreign proteins were displayed as fusions with the minor coat proteins gIIIp and the genetic information was packaged following the packaging signal. This display allowed the enrichment of phage with selective probe.

### 4.1.5.4 Library weeding

(This paragraph comprises step4 of Scheme 4.1.1)

According to the result of colony screening PCR, ~92% phagemid DNAs in the primary t*his*F library contained base deletions or insertions (section 4.1.5.2), these base deletions or insertions would cause frame-shift and subsequently can not be

translated into tHisF variants. To exclude these abnormal genes and negative DNA products from the primary library, one immuno-depletion step was introduced to selectively enrich the t*his*F gene products. After displaying the primary phagemid library on the surface of M13 phages, panning the phage displayed library against anti-HisF antibodies would purify the library to those members that were actually viable tHisF derivatives. Fig. 4.1.10 shows the process of panning the phage displayed library against anti-HisF polyclonal antiserum.

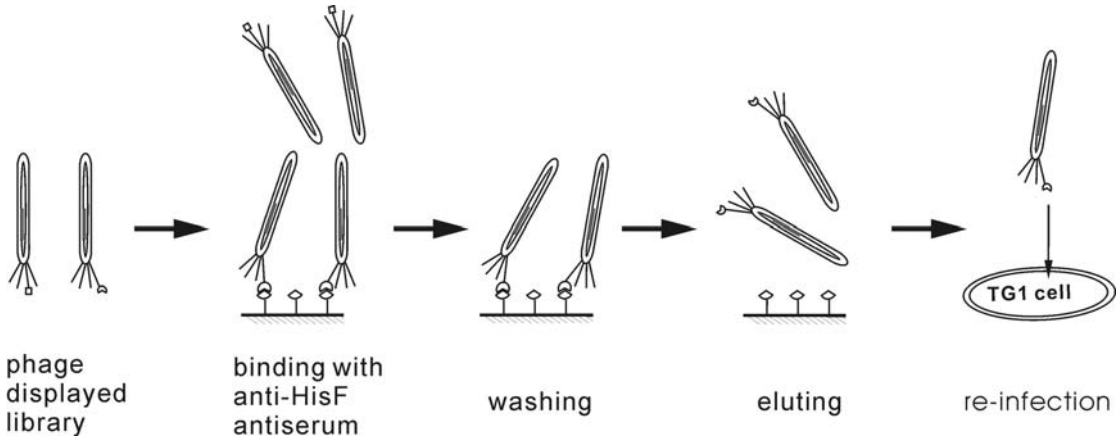



**Fig. 4.1.10** Process of panning the phage displayed library against anti-HisF polyclonal antiserum.

Panning procedure was carried out according to reference with some modification (section 3.3.5). In this method, anti-HisF antiserum (section 2.5.2) which was immobilized on MaxiSorp NuncImmuno$^{TM}$ tube surface was used to bind with the HisF variants on the M13 phage surface from displayed primary library while those non-HisF variants were washed away, and finally the HisF variants was enriched in the eluate after eluting them off the antibody with strong basic (or acidic) buffer. The enriched HisF-fusion phages were used to re-infection host cell TG1 to prepare the phagemid library.

After one round of panning against anti-HisF antiserum, the retained library was ca. $2.57x10^5$ independent phages according the titering (protocol in section 3.3.4). Phagemid DNAs were prepared from the purified library by Jetstar-kit (section 3.2.5), individual phagemid DNA of independent clones from the purified library were prepared with method of DNA Mini-preparation (section 3.2.4) and subjected to colony-screening PCR to identify the ratio of *his*F variants in the purified library and subsequently subjected to DNA sequencing.

Colony-screening PCR (same as section 4.1.5.2) result revealed that 95% of the tested clones from the purified library contain t*his*F variants. Fifty phagemid DNAs of independent clones from the purified library were sequenced, 40 out of the 50 phagemids are t*his*F-variants without any base deletion or insertion (one of them

contains an unexpected stop codon ´TAG´ which would produce truncated protein when the DNA was translated to protein), 6 out of the 50 phagemids contained one codon deletion (5 lie in randomized positions, 1 lies in the bulk region) but without frame shift, 4 phagemids have frame shift. This meant that about 90% (45 out of 50) clones from purified t*his*F phagemid library are t*his*F variants, while 10% clones contain DNA with base deletion or insertion.

This purified phagemid library was named **Lib.I** as will be called hereafter, and was ready for the use in next step.

### 4.1.6 Increasing the diversity of t*his*F gene library by DNA shuffling

(This section comprises step5 of Scheme 4.1.1)

DNA shuffling is a powerful technique for directed evolution of proteins *in vitro*. The major application of DNA shuffling in protein evolution is multiple gene shuffling (often referred to as molecular breeding). In this technique, homologous DNA sequences are digested and subsequently reassembled. The result is a library of chimeric genes containing additional point mutations (Fig. 4.1.11).

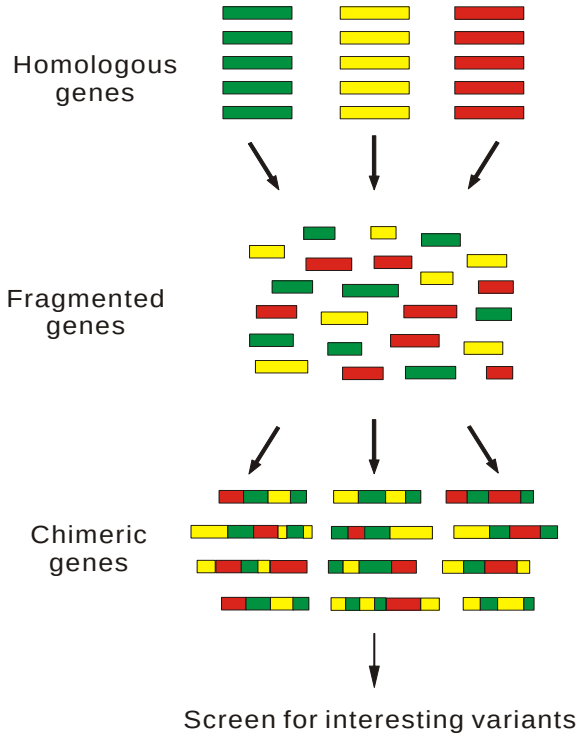



**Fig. 4.1.11** Schematic representation of DNA shuffling

Parents DNA were fragmentized with DNaseI digestion and selected for suitable size at range 50-250bp; the fragments then were separated under heating, and annealing to others, and extended by polymerase reaction; after cycles of separation, annealing and extension, full genes were constructed at chimeric way.

#### 4.1.6.1 Amplifying target genes

A population of t*his*F gene from purified gene library was prepared by amplification PCR (section 3.2.11). In the presence of 3.5 mM $MgCl_2$ with *Taq* DNA polymerase, using the phagemid DNA of purified library **(Lib.I**) as the template, and two oligonucleotides, one was located at the beginning of t*his*F gene in forward direction

(VB01, 5'-GTCAGCATGCGGCCCAGCCGG-3') and the other was located at 150bp downstream of the interest gene (sh-up, 5'-CACGCCCTCATAGTTAGCGTA) were used as the primers for the amplification, which could reduce the unspecific binding, thus avoid the unexpected by-product from amplification PCR. The following reaction conditions were applied: step 1, 95°C, 3 min; step 2, 95°C, 45 s; step 3, 55°C, 45 s; step 4, 72°C, 1min20s; step 5, 72°C, 10 min; steps 2 to 4 were repeated 30 times. The amplified DNA was purified by $CHCl_3$ extraction (section 3.2.3) or gel extraction (section 3.2.8).

#### 4.1.6.2 DNase I digestion and fragment reassembly

The purified amplification PCR product then were fragmentized by DNase I digestion (4 units DNaseI for DNA 2~4µg) in the presence of 4 mM $MnCl_2$ at 25°C for 10 min (see section 3.2.14A). The digested DNA was applied to agarose gel electrophoresis to check the size of fragments. Fragments with suitable size were purified by gel extraction (section 3.2.8) or $CHCl_3$ extraction (section 3.2.3), and used in reassembling PCR (self-priming PCR, section 3.2.14B) in the presence of 3.5 mM $MgCl_2$ with Taq DNA polymerase. The following program was used for reassembling PCR: step 1, 95°C, 2 min; step 2, 95°C, 30 s; step 3, 55°C, 40 s; step 4, 72°C, 1min20s; step 5, 72°C, 10 min; step 2 to 4 were repeated 45 times.

#### 4.1.6.3 Final PCR

The final PCR was performed with Taq DNA polymerase in the presence of 3.5 mM $MgCl_2$ , using the reassembled DNA products as template and oligonucleotide VB-01 5'-GTCAGCATG C**GGCCC AGC CGG(CC)-**3' with *Sfi*I site (in bold, ref. Fig. 4.1.1.2) as 5'-primer, and oligonucleotide VB-30 5'-CGATAG GATC**AGATCT**CA AAC CC -3' with *Bgl*II site (in bold) as 3'-primer. The program for the final PCR is step 1, 95°C, 2 min; step 2, 95°C, 30 s; step 3, 43°C, 30 s; step 4, 72°C, 1 min; step 5, 72°C, 10 min; steps 2 to 4 were repeated 30 times. The final PCR products were purified by chloroform extraction (section 3.2.3, Fig. 4.1.12).

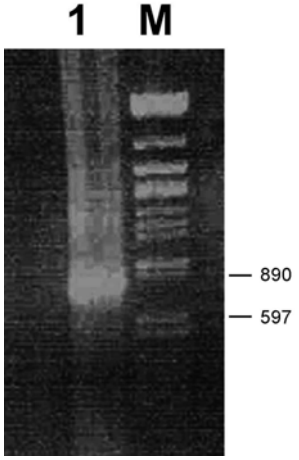



**Fig.4.1.12** Products of final PCR

1 Products of final PCR after DNA shuffling

M DNA molecular weight marker, Lambda DNA/Eco47I (AvaII)

#### 4.1.6.4 Cloning the re-shuffled t*his*F gene library into phagemid vector and constructing re-shuffled phage library

The purified final PCR products were digested with *Sfi*I and *Bgl*II (section 3.2.6) and purified by gel extraction (section 3.2.8) and then ligated into phagemid vector pCANTAB5 E in the corresponding *Sfi*I and *Bgl*II sites (section 2.10). The ligated pCANTAB5 E-t*his*F products were transformed into host cell TG1 (*SupE*) (section 3.1.7). The bacterial stock of the library was plated out on LB-Amp plates.

The construction of the second t*his*F gene library which was generated through DNA shuffling into pCANTAB 5E vector resulted in ca. $1.6x10^7$ clones, according to the titer plates.

#### 4.1.6.5 Colony screening PCR and purification of the re-shuffled thisF phagemid library

In the final PCR, there was a by-product of about 670bp which was very close to the band of desired t*his*F gene (780bp) on agarose gel. The incomplete separation of the by-product by agarose gel electrophoresis caused the non-t*his*F genes remaining in the re-shuffled gene library. To determine the quality of re-shuffled t*his*F gene library, colony-screening PCR (section 3.2.11) was performed again as in section 4.1.4.2 with two primes: VB-03 and sh-up. The screening result revealed that ~55% clones contained pCANTAB 5E-t*his*F-variants, which meant there were still considerable ratio of non- t*his*F products in the re-shuffled libraries. To purify the re-shuffled t*his*F gene library, the immuno-deletion method was adopted again (section 4.1.4.5) by using polyclonal anti-HisF antiserum to selectively bind with the phage displayed re-shuffled t*his*F gene library and washing away the non-*his*F gene products.

The clones retained by anti-HisF serum were ca. $6.58x10^6$ according to the titering plates. Colony-screening PCR result revealed that 90% of the tested clones contain t*his*F-variants. This purified re-shuffled t*his*F phagemid library was named **Lib.II** as will be called hereafter.

#### 4.1.6.6 DNA sequence analysis of re-shuffled phagemid library

Phagemid DNA of independent clones from **Lib.II** was prepared with the method of DNA-mini preparation (section 3.2.4), and then sequenced. The sequence analysis (fifty clones) revealed that 94% of these clones were t*his*F-variants (in detail, 82% sequences were complete t*his*F-like gene, 12% sequences contained codon deletion but without frame shift), only 6% clones contained base deletions or insertions which could cause frame-shift.

Phagemid DNAs of the purified re-shuffled t*his*F library (**Lib II)** were prepared with Jetstar-kit (see section 3.2.5) and ready for use in the following experiment.

### 4.1.7 Cloning the t*his*F gene library into expression vector pKK223-3

(This paragraph comprises step6 of Scheme 4.1.1)

To clone the t*his*F gene library into expression vector, the t*his*F genes were amplified from phagemid library **Lib.II** with two primers: hisF-EcoRI 5'-ATGC **GAA TTC** ATG CTC GCT AAA AGA ATA ATC-3' with *EcoR*I site (in bold) and hisF-PstI 5'-ATGC **CTG CAG** CTA CAA ACC CTC CAG TCT TAC-3' with *Pst*I site (in bold). The amplified products were digested with *EcoR* I and *Pst* I and were ligated into the *EcoR* I and *Pst* I sites in pKK223-3 expression vector. The pKK223-3 vector contains a strong tac promoter (section 2.10, Fig. 2.3) and an ampicillin resistant gene. This vector has a high base expression level without induction of isopropyl β-D-thiogalactoside (IPTG). The base expression level was shown to be enough for *in vivo* complementation in this work.

The size of the obtained t*his*F gene library in the pCANTAB vector is $6.58 \times 10^6$ independent clones. Colony screening PCR showed that 90% tested clones possessed *his*F variants, which represented the portion of positive clones in the total repertoire.

### 4.1.8 Characterization of the synthetic thisF gene library and statistic analysis

The sequences of 50 random picked clones from the first gene library (**Lib.I**) and 50 random picked clones from the second (re-shuffled) gene library (**Lib.II**) were used for the statistic analysis.

#### 4.1.8.1 Codon frequency as non-resident amino acid moieties

One hundred DNA sequences from the two libraries (**Lib.I** and **Lib.II**) were used for codon frequency analysis. All the codons coding for 19 amino acids (except cysteine which was not added into the random codon mixture) was all represented at the randomized positions of t*his*F gene. The average codon frequency in 100 sequences is 24.8 [471times/19positions], which was in accordance with the theoretical average frequency 25.1 for each amino acid [0.56*(1-1/19)*9*100/19], and the average relative frequency (using the highest frequency as 1) is 0.48 (Tab.4.1.1 and Fig.4.1.13).

However, there were some deviations away form the theoretical calculation – ATG which codes the amino acid methionine had the highest frequency (50 times out of the 100 sequences); CAC which codes the amino acid histidine had the lowest frequency (8 times out of 100 sequences). A possible explanation is that limited sample repertoire may lead to this bias - more sequences sample may reduce the statistic error. Another possible bias source may lie in the oligonucleotides synthesis – different tri-nucleotide block (codon) could have different coupling efficient in the synthetic procedure of coupling the oligonucleotide-triplet phosphoramidites (coupling blocks)

onto the growing oligonucleotide, which would alter molar ratios of tri-nucleotides (coding for the different amino acids).

**Table 4.1.1** Codon frequency as non-resident moieties in the two t*his*F libraries (**Lib.I** and **Lib.II**)

| Codon | Amino acid | Frequency in 100 sequences | Relative frequency |
|---|---|---|---|
| AAA | Lys | 34 | 0.68 |
| AAC | Asn | 37 | 0.74 |
| ACC | Thr | 24 | 0.48 |
| ATC | Ile | 15 | 0.3 |
| ATG | Met | 50 | 1 |
| CAC | His | 8 | 0.16 |
| CAG | Gln | 36 | 0.72 |
| CCG | Pro | 11 | 0.22 |
| CGT | Arg | 28 | 0.56 |
| CTG | Leu | 14 | 0.28 |
| GAA | Glu | 34 | 0.68 |
| GAC | Asp | 21 | 0.42 |
| GCT | Ala | 12 | 0.24 |
| GGT | Gly | 23 | 0.46 |
| GTT | Val | 38 | 0.76 |
| TAC | Tyr | 20 | 0.4 |
| TCT | Ser | 24 | 0.48 |
| TGG | Trp | 16 | 0.32 |
| TTC | Phe | 26 | 0.52 |
| | | Sum 471, average 24.8 | Average 0.48 |

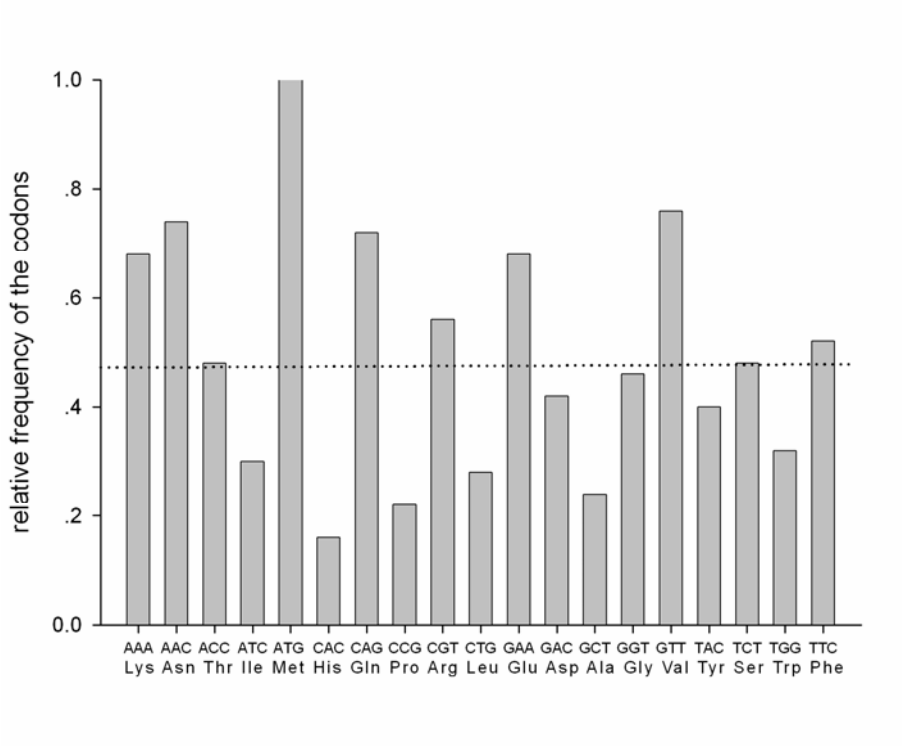



**Fig.4.1.13** Frequency of 19 amino acids as non-resident moieties observed in two t*his*F libraries (**Lib.I** and **Lib.II**). The dotted line represents the average frequency, 0.48.

### 4.1.8.2 Distribution of mutated codons in t*his*F gene libraries

A calculated binomial distribution for nine randomized positions with an average 4.77 exchange codons (section 4.1.2.2) and the experiment determined distribution of exchange are presented in Table 4.1.2 and Fig 4.1.14. The experiment determined distribution roughly matched the calculated binomial distribution.

**Table 4.1.2** Comparison of the experimentally determined distribution of codon mutation with the theoretical distribution of codon mutation of the repertoire with nine randomized positions and on the average 4.77 codon exchange (n=9, x =4.77)

| Total mutated codons in one clone | Binomial distribution | Observed distribution in 100 clones |
|---|---|---|
| 0 | 0.001 | 0 |
| 1 | 0.008 | 0 |
| 2 | 0.041 | 6 |
| 3 | 0.116 | 15 |
| 4 | 0.213 | 20 |
| 5 | 0.260 | 25 |
| 6 | 0.212 | 26 |
| 7 | 0.111 | 6 |
| 8 | 0.034 | 2 |
| 9 | 0.005 | 0 |

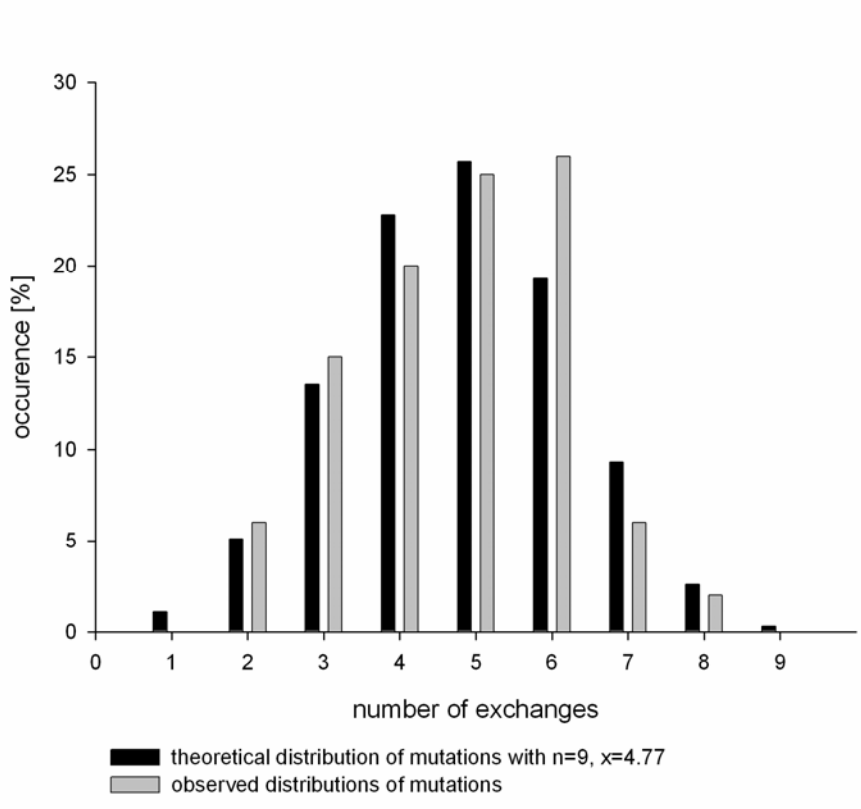



**Fig. 4.1.14** Comparison of the experimentally determined distribution of codon mutation with the theoretical distribution of codon mutation with nine randomized positions and on the average 4.77 codon exchange (μ =4.77)

### 4.1.8.3 Exchange ratio for each designed position and average exchange ratio

In the 100 sequenced clones, the average exchanged codons were 4.67(ref. Table 4.1.3). This result roughly matched the expectation of the design – the calculated exchange ratio was 4.77 [56%*(1-1/19)*9] according to the tri-nucleotides molar ratio used in the synthesis of oligonucleotides (section 4.1.2.2). In purified primary library Lib-I, the average exchanged codon of which was 4.36; in the re-shuffled library Lib-II, the exchange frequency was 4.99. Even though fluctuations could be seen in single random position and in the two libraries, the overall exchange frequency is consistent with design.

Average exchange ratio at single random position of **Lib.I** was 48.4%, and it was 55.4% for **Lib.II** – the mean of average exchange ratio at single random position of 100 clones was 51.9% (Table 4.1.3, Fig 4.1.15). This result also is consistent with the design of oligonucleotide where the molar ratio of non-resident codons was set at 53.1% (section 4.1.2.2).

**Table 4.1.3** Exchange frequency at individual random positions in *this*F-gene repertoire (100 clones)

| **Random positions** | C/A9 | D11 | L169 | T171 | I173 | K179 | S201 | A204 | L222 | Sum | Ave. ratio (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mutations in 50 clones from Lib-I** | 28 | 28 | 31 | 23 | 31 | 17 | 12 | 17 | 31 | 218 | 48.4 |
| **Mutations in 50 clones from Lib-II** | 26 | 25 | 38 | 35 | 18 | 29 | 23 | 25 | 30 | 249 | 55.4 |
| **Overall mutations in 100 clones** | 54 | 53 | 69 | 58 | 49 | 46 | 35 | 42 | 61 | 467 | 51.9 |

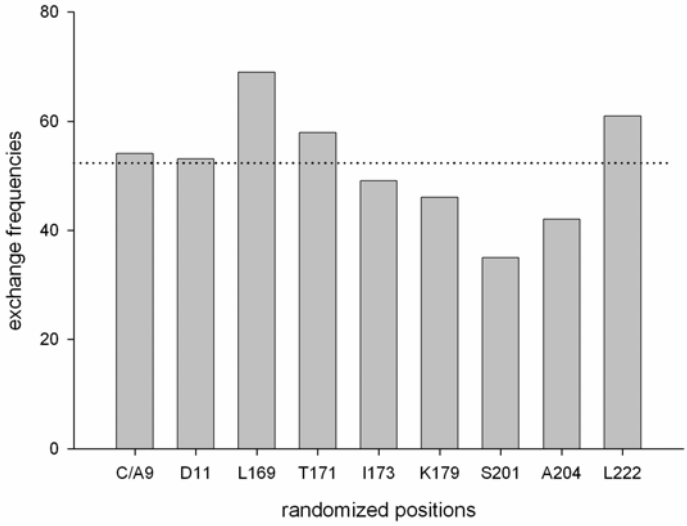



**Fig.4.1.15** Exchange frequency at individual random positions in 100 clones of *this*F-gene repertoire. The dotted line represents the average exchange frequency, 51.9%.

### 4.1.8.4 Extra point mutations

There were thirteen point mutations in the fifty clones from the primary library (**Lib.I**). Among these point mutations, eleven were missense mutations caused amino acid mutations; the other two were silent mutations. In addition, there were one codon deletion and one codon insertion in these fifty clones. There were one hundred and forty point mutations in the fifty clones from the second gene library (**Lib.II,** after DNA shuffling): eighty-five were missense mutations; fifty-five were silent mutations. There were three codon deletions in the fifty clones from **Lib.II**. The average point mutations per clone in **Lib.I** was 0.32 while it was 2.82 in **Lib.II**; in **Lib.I** 28% clones contained point mutations, in **Lib.II** it was increased to 94% (Table 4.1.4, Fig. 4.1.6).

The significant high ratio of point mutation in **Lib.II** obviously came from the procedure of DNA shuffling. In the DNA shuffling, some error prone parameters were employed, e.g. higher $Mg^{2+}$ concentration at 3.5mM, lower annealing temperature (43 $^{o}$C), and non-proofreading *Taq* polymerase all contributed to point mutations, and more PCR circling times (45 to 50 cycles) increased the mutations as well. So, the DNA shuffling increased the diversity of t*his*F gene library not only by recombination of the parent population but also by introducing of point mutations.

**Table 4.1.4** Summary of extra point mutations in the t*his*F gene libraries before (**Lib.I**, 50 clones) and after DNA shuffling (**Lib.II**, 50 clones)

| | | **Number of clones** | |
|---|---|---|---|
| | | Lib-I | Lib-II |
| **Number of extra point mutations in single clone** | 0 | 36 | 3 |
| | 1 | 11 | 11 |
| | 2 | 1 | 10 |
| | 3 | 0 | 6 |
| | 4 | 0 | 13 |
| | 5 | 0 | 3 |
| | 6 | 0 | 3 |
| | 7 | 0 | 1 |
| | 8 | 0 | 0 |
| **Clones containing codon deletion or insertion** | | 2 | 3* |
| **Sum of clones containing extra mutation/s** | | 14 | 47 |
| **Average point mutations in each clone** | | 0.32 | 2.82 |

* This number is included in the 47 clones containing extra mutations.

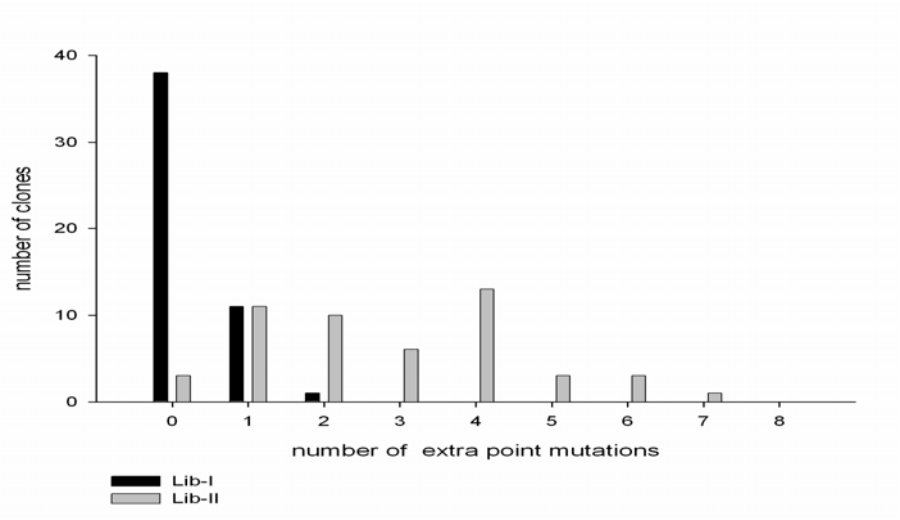



**Fig. 4.1.16** Distribution of extra point mutations in single clone in the t*his*F gene libraries before (**Lib.I**) and after DNA shuffling (**Lib.II**)

#### 4.1.8.5 Non-input codon

Besides the extra point mutations at non-randomized region, several codons which were not added in the trinucleotide mixture emerged at randomized positions in the sequenced clones, they were: AGA, CCT, GAT, GTC, TAG, TCC. This might be produced by point mutation at randomized positions in the PCR procedure, e.g. AGA could be produced by AAA with A to G mutation at the second base; CCT could be produced by TCT with T to C mutation at the first base, etc.

### 4.1.9 Summary of the t*his*F gene library construction

Based on x-ray structure of tHisF and computer modelling of its substrate (PRFAR) binding, nine residues neighbouring putative substrate binding sites were selected for controlled mutagenesis. The residues are: C/A9, D11, L169, T171, I173, K179, S201, A204, L222, their positions are strand β1 (C/A9, D11), strand β6 (L169, T171), loop β6/α6 (I173, K179), β7 (S201), loop β7/α7 (A204) and β8 (L222), respectively. In addition, two residues on the putative PRFAR binding site at N-terminal half of the protein were replaced with alanine codon (AAC/GCT, N103A; ACG/ GCT, T104A) hoping to alter the substrate specificity for evolving novel property from tHisF scaffold.

The construction of controlled partially randomized gene library was based on the incorporation of synthetic DNA oligonucleotides into genes via self-priming PCR. The key to this method is the introduction of diversity at specific positions at codon level within the synthetic DNA with pre-synthesized tri-nucleotides. In order to generate a repertoire at a possible and reasonable size, these positions were not

randomized saturatedly, the respective wild-type codons were added as a portion in the tri-nucleotide mixtures.

A gene library of t*his*F variants in phagemid vector pCANTAB 5E has been constructed. To avoid unnecessary screening or selection work in searching for novel properties from the library, an immuno-deleting method was successfully used to purify the t*his*F gene library, i.e. using the immobilized anti-HisF antiserum to bind HisF variants which were displayed on the M13 phage surface, those negative DNA product and abnormal DNA assembles which would cause frame-shift were deleted. DNA shuffling was used to increase the gene diversity.

A plasmid t*his*F gene library was constructed in an expression vector pKK223-3 for the production of soluble tHisF variants for the use of genetic complementation.

The t*his*F gene library has $5.58 \times 10^6$ independent clones. Ninety-four percent of the clones are t*his*F variants without nucleotide deletion or insertion according to DNA sequencing. The average exchanged codons in each clone are 4.67 and the average exchange ratio in single random position is 52%, both of which are consistent with the library design (Table 4.1.5).

**Table 4.1.5** Summary of the theoretical and experimentally determined mutation rates and the average number of exchanged codon

| Starting conditions for the library synthesis | |
|---|---|
| molar ratio of residential (wild type) codon in the trinucleotide mixture (q) | 0.47 |
| molar ratio of non-residential codon in the nucleotide mixture (p) | 0.53 |
| number of non-residential codons in the tri-nucleotide mixture | 18 |

| Theoretical calculated results for the library | |
|---|---|
| exchange probability in single random position | 0.53 |
| average number of exchanged codons in the repertoire (x) | 4.77 |

| **Experimentally observed results in the library** | |
|---|---|
| size of the library (independent clones) | $5.58 \times 10^6$ |
| exchange probability in single random position ($p_{exp}$) | 0.52 |
| average number of exchanged codons in the repertoire ($x_{exp}$) | 4.67 |

## 4.2 Search for novel properties from tHisF libraries

### 4.2.1 The methods for selection or screening

To find a proteinaceous catalyst for a given chemical reaction is the final objective of enzyme engineering. Searching libraries according to pre-set functional requirement constitutes a second pillar of evolutionary enzyme engineering, equal in importance to library construction.

There are two strategies for novel properties selection/screening. One is searching for the individual molecular species under the actual occurrence of the chemical reaction; the other is searching for the candidates for binding properties against transient state analogues (TSA) in the library and followed by further evaluation for the desired catalysis.

#### 4.2.1.1 Search for the individual molecular species under the actual occurrence of the chemical reaction

(i) Genetic complementation

Genetic selection takes advantage of a suitable system in which the presence of the catalytic activity provides a growth advantage to the bacteria or micro-organism possessing it. This approach includes a range of schemes for linking catalytic activity to a survival factor such as use of the substrate of interest as sole carbon source for growth, lack of a chemical compound within the culture medium e.g. amino-acids; resistance to molecules e.g. mutagens, antibiotics, metabolic analogues; resistance to altered environment, e.g. resistance to high temperatures, resistance to lowered temperatures; resistance to UV radiation etc (Jestin JL et al., 2004).

In genetic selection experiment, as many as $10^8$-$10^{10}$ variants can be assayed quickly, actually, transformation efficiency is the technique bottleneck in this method that limits the size of library. If a selection strategy turns out to be general enough so as to be applied to different chemical reactions, its use may not be more time-consuming than screening approach.

However, in these selections the mutant proteins are tested in a biological setting, which can complicate and limit the range of engineered properties and restrict the possible substrates and reactions. Many directed evolution experiments are usually impossible under genetic complementation.

(ii) Phage display and selection for catalytic activity

The system is based on the phage display (cf. section 4.2.1.2a). One method is designed for the *in vitro* evolution of protein catalysts in a biologically amplifiable system. Substrate is covalently and site specifically attached by a flexible tether to the pIII coat protein of a filamentous phage that also displays the catalyst. Intramolecular conversion of substrate to product provides a basis for selecting active catalysts from a library of mutants, either by release from or attachment to a solid support. An

analysis of factors influencing the selection efficiency of a model experiment with the enzyme staphylococcal nuclease showed that phage displaying staphylococcal nuclease can be enriched 100-fold in a single step from a library-like ensemble of phage displaying noncatalytic proteins (Pedersen H et al., 1998).

A similar selection strategy is phage display and catalytic elution. The metallo-enzyme was displayed on the filamentous phage fd and a selection process was designed to extract active phage-bound enzymes from libraries of mutants in three steps: 1. inactivation of active phage-bound enzymes by metal ion complexation e.g. by EDTA, 2. binding to substrate-coated magnetic beads, 3. release of phages capable of transforming the substrate into product upon zinc salt addition. The selection process was tested on model mixtures containing fd-betaLII plus either a dummy phage, a phage displaying an inactive mutant of the serine beta-lactamase TEM-1, or inactive and low-activity mutants of betaLII. The selection was then applied to extract active phage-bound enzymes from a library of mutants generated by mutagenic polymerase chain reaction (PCR). The activity of the library was shown to increase 60-fold after two rounds of selection (Ponsard I et al., 2001).

In another activity-based phage display and selection method, one substrate is linked to the phage-enzyme and the second substrate was labelled by a biotin group so that phage-enzymes catalysing the substrate to product conversion are specifically labelled and can be recovered on solid phase (Atwell and Wells, 1999; Xia G et al., 2002).

(iii) *In vitro* compartmentalisation (IVC)

In vitro compartmentalisation (IVC) allows billions of experiments to be performed in parallel by partitioning each experiment into a separate microscopic compartment. (Tawfik D & Griffiths AD, 1998; Cohen H et al., 2004). In this system, the linkage between genotype and phenotype is achieved by compartmentalizing single genes in the aqueous compartments of a water-in-oil emulsion. An *in vitro* transcription/ translation reaction mixture containing a library of genes encoding mutant enzymes is dispersed to form a water-in-oil (w/o) emulsion with typically one gene per aqueous microdroplet. The genes are transcribed and translated within their microdroplets. Proteins with enzymatic activity convert the substrate into product and the w/o emulsion is converted into a water-in-oil-in-water (w/o/w) emulsion. The microdroplets containing products are separated from non-product microdroplets using methods specific for the products. Genes that encode active enzymes are recovered and amplified using the polymerase chain reaction. And these genes can be re-compartmentalised for further rounds of selection.

(iv) Screening

When selection is not possible, the individual bacteria or microorganisms present in the library must be physically separated and individually assayed for catalysis. This can be done either in agar plates or in microtiter plates. Usually good selection may be

limited to searching $10^8$ protein variants; screening methods often can cover no more than $10^4$. This has caused the move of using sensitive detection methods for screening of large enzyme variant libraries adapted to high-throughput.

A range of new screening technologies have been developed in recent years (Ness JE et al., 1999; Schmidt-Dannert C et al., 2000; Raillard S et al., 2001; Glieder A et al., 2002, Jaeger KE and Eggert T 2002). One example is cell-surface display and high throughput screening. The promise of screening methods based on fluorescence-activated cell sorting (FACS) for directed enzyme evolution is being realized and significantly improved enzymes or desired function, including expression level, stability, ligand binding and catalysis (Daugherty PS et al., 2000). Enzyme libraries displayed on the surface of microbial cells or microbeads can be screened with fluorogenic substrates that provide a physical linkage of the reaction product to the corresponding enzyme. Libraries exceeding $10^9$ different variants can be quantitatively analyzed and screened by flow cytometry at a rate of 30 000 cells/ second (Becker S et al., 2004).

#### 4.2.1.2 Search for the individual molecular species based on binding properties

(i) Phage display and select candidates for binding property

This method combines *in vivo* expression in *E. coli*, M13 phage displayed proteins, and *in vitro* (solid-support) "biopanning" and is a general and robust *in vitro* selection method.

There are two kinds of vectors which can be used to generate phage-displayed protein libraries: phage vector and phagemid vector. Phage vector uses the genome of a phage e.g. M13 which was modified to allow insertion of interesting genes into phage's coating protein such as gIIIp or gVIIIp as a fusion protein. In a phagemid vector system, DNA of exogenous proteins is cloned into gene of phage's coating protein e.g. gIII or gVIII which is presented within phagemid vector, and the packaging of recombinant phagemid DNA and display of the fusions are provided by a helper phage such as M13KO7 (cf. Fig 4.1.8). The resulting phage particles display the fusion gIIIp on the lip of the filamentous phage, making them available for interaction with an immobilized bait ligand and ready for *in vitro* selection for a binding property or catalytic property (Baek H et al. 2002).

For the display of larger molecules, a phagemid vector system is more suitable and has more advantages over a phage vector system including the more efficient ligation-transformation that allows creating libraries of larger size and relatively easy genetic manipulation for introducing special features into a phagemid.

The origin of phage display and biopanning experiment demonstrated that phage containing foreign protein fusion could be enriched >1000-fold over wild-type phage with an immobilized polyclonal antibody (Smith GP 1985). Several rounds of panning enrich for the interacting clone(s), subsequent identification of positive clones can be

easily achieved after the amplification of phagemid DNA of selected clones. In attempts to select for improved catalysts, specialized selections using transition-state analog, active site ligands, reactive substrates, inhibitors, and reactive products have been used and had got significant success (Janda KD, et al 1997; Atwell S et al., 1999).

Phage display is a powerful *in vitro* selection technique, where more than $10^9$ protein or peptide variants can be subjected to selection for improved binding properties. For this reason, a phagemid vector pCANTAB 5E was chosen for the construction of the primary t*his*F gene library.

However, this selection is not directed for the catalytic ability but the binding property. Binding property doesn't necessarily lead to catalysis. So the selected results have to be evaluated for catalytic activity.

(ii) *In vitro* display

This technology includes ribosome display and mRNA display. They combine two important advantages for identifying and optimizing ligands by evolutionary strategies. First, by obviating the need to transform cells in order to generate and select libraries, they allow much higher library diversity. Second, by including PCR as an integral step in the procedure, they make PCR-based mutagenesis strategies convenient. The resulting iteration between diversification and selection allows true Darwinian protein evolution to occur *in vitro*.

In ribosome display, the translated protein remains connected to the ribosome and to its encoding mRNA; the resulting ternary complex is used for selection.

In mRNA display, mRNA is first translated and then covalently bonded to the protein it encodes, using an adaptor molecule. The covalent mRNA–protein adduct is purified from the ribosome and used for selection. Successful examples of high-affinity, specific target-binding molecules selected by *in vitro* display methods include peptides, antibodies, enzymes, and engineered scaffolds (Lipovsek D, Plueckthun A 2004).

### 4.2.2 Investigation of tHisF mutations under mesophilic environment by eHisF-complementation

#### 4.2.2.1 Aim of the investigation

The gene library described in section 4.1 is a potential source of catalysts accelerating a variety of chemical reactions. As introduced in section 4.2.1, to achieve the final goal of evolving novel catalytic activity from an existed protein scaffold, there are a range of methods available for choosing. Taking into consideration of the advantages and drawbacks of all the possible strategies for enzyme engineering, the most efficient and convenient method is genetic complementation provided the selection system is available.

Besides the t*his*F gene library, we have got several auxotrophic Escherichia coli stains with different gene mutations including $HisF^-$, $HisA^-$, and $DapA^-$ strains at hand. Using the t*his*F library to complement a *his*$F^-$ strain to a $HisF^+$ phenotype constitutes a simple control experiment for the practicality of the approach choice, based on the fact that: first, wild-type t*his*F is able to complement *his*$F^-$ in E. coli (Beismann-Driemeyer S et al., 2001); and second, theoretically, wild-type should be the most abundant single molecular species in the t*his*F library (Kansy E 2003, Schiller S 2004). As a consequence, the minimal outcome of the complementation experiment would be re-isolation of the wild-type sequence from the library, beyond this, a number of interesting insights into tHisF structure/function relationship may be gained, which may give answers to the following questions:

(i) Must all nine residues be conserved simultaneously in order to maintain HisF function?

(ii) How is the variability of the residues involved in the formation of the possible tHisF- eHisH heterodimer?

(iii) How is the variability with respect to conserved residues involved in coupling of glutaminase and synthase activities and the putative ammonia tunnel?

(iv) Which exchanges possibly increase HisF activity of tHisF in the environment of *E. coli*?

However, it should be kept in mind that two residues N103 and T104 were both changed to alanine while constructing the synthetic t*his*F library. So, the functional importance of the possible substrate binding residues N103 and T104 also can be evaluated in this complementation experiment.

#### 4.2.2.2 The eHisF$^-$ strain and eHisF complementation

An auxotrophic E. coli strain UTH860 (also called SB3931, Garrick-Silversmith L et al 1970) was used for the complementation assay. UHT860 is a mutant induced by UV irradiation (Goldsmidt EP et al 1970). Nucleotide G547 is deleted in its *his*F gene which causes frame shift of the gene. At protein level, this leads to amino acid replacement of G183A and consequently alters the amino acid sequence from 183-GVRNGYDLEQ to 183-ACVTVTTSRNN-Stop and the translation stops after ten new amino acids (R. Sterne, personal communication).

HisF$^-$ E. coli strain UTH860 was transformed with the pKK223-3-t*his*F gene library with a size of 5.58x10$^6$ independent clones (see section 4.1.6) by electroporation (section 3.1.7) and yielded 4.69x10$^7$ clones according to non-selective LB plates for tittering (section 3.1.8). The average representation fold of all the species in the library is 8.4. The tranformants were cultured overnight, washed and diluted with 1xM9 minimal medium (section 2.7), plated on minimal VB agar plates (section 2.7) containing 100 μg/ml ampicillin at 37°C for complementation assay. The pKK223-3 expression vector contains a strong tac promoter (section 2.10, Fig. 2.3) and an ampicillin resistant gene. This vector has a high base expression level. The base expression level was shown to be enough for complementation in this work without the induction of isopropyl β-D-thiogalactoside (IPTG).

Since there was no supplement histidine which was required by the UTH860 strain to survive on the selective VB plate, those clones which grew on the selective plate must be able to complement eHisF function. After incubation at 37°C for 20 h, numerous colonies grew on the main selective plates, countable colonies grew on the diluted selective plates. The ratio of colonies grew on selective VB plates and non-selective LB plates was 3.1x10$^{-3}$. One hundred independent colonies were randomly picked and plasmid DNAs of these eHisF$^-$ selected clones were prepared using standard DNA Mini preparation protocol (section 3.2.4).

#### 4.2.2.3 Retransformation, screening PCR and DNA sequencing

The selection ratio of 3.1x10$^{-3}$ is in the same range as 1.2x10$^{-3}$ of expected ratio for w.t. gene in the gene library (Kansy E, 2003; Schiller S 2004). Even though the genetic background can be expected to revert mutant strain to wild-type at fairly high frequency due to the frame shift nature of the *his*F mutation, their frequency can nevertheless be expected to be lower than 10$^{-3}$ so that background reversion should not be significant to the number of HisF$^+$ transformants. This was checked experimentally by choosing at random 50 HisF$^+$ transformants and using their plasmid DNA for re-transformation of UTH860, 100% tested samples showed complementing ability.

In order to identify if the eHisF$^-$ complementary plasmids contain t*his*F gene, twenty plasmids prepared from eHisF$^-$ selected colonies were subjected to screening PCR using oligonucleotides Vb-03 5'-GGC ACT AAC TTT GAG AAC CTG-3' which is specific for t*his*F gene and pKK-rev 5´-TTT TAT CAG ACC GCT TCT GC-3´ which is specific for pKK223-3 vector as primers. The result of screening PCR showed all the screened plasmids are pKK223-3-t*his*F variants (100% positive).

The one hundred plasmid DNAs prepared from eHisF$^-$ selected clones were submitted to sequencing, ninety-eight readable sequences were obtained.

### 4.2.2.4 Analysis of eHisF$^-$ complementing tHisF variants

The sequence information of ninety-eight successfully sequenced eHisF$^-$ complementing plasmids was summarized as following (Table 4.2.1.1).

(i) All the eHisF$^-$ selected plasmids DNA were pKK223-3-$^{syn}$t*his*F variants with or without point mutations. There was no mutation in the nine controlled randomized positions.

(ii) Only 6 out of 98 clones (clone H1, H15, H25, H29, H82, H95) were wild type synthetic t*his*F, which constituted 6.1% of the total sequenced clones; all the rest clones (92 out of 98, 93.9%) contained point mutations at non-randomized positions. The highest number of residue exchanges in single clones was three, the clones were: H19, H66, H76, H93 (Table 4.2.2.1).

**Tab. 4.2.2.1** Mutations in clones selected from eHisF$^-$ complementation assay

| Name of clone | Mutations | Number of clones | Name of clone | Mutations | Number of clones |
|---|---|---|---|---|---|
| H2 | F86L | 46 | H55 | F86L, T194A | 1 |
| H5 | F86L, V246A | 1 | H56 | F86L, T178A | 1 |
| H10 | F86L, N148S | 1 | H57- | F86L, s245 | 1 |
| H13 | N25D, F86L | 1 | H64- | s76, F86L | 1 |
| H14- | F86L, s152, s218, s238 | 1 | H65- | F86L, s216 | 1 |
| H16 | F86L, V234D | 1 | H66 | F86L, L94P, M186T | 1 |
| H17- | F86L, s217 | 1 | H67 | F86L, s130, K132E | 1 |
| H19 | E46D, F86L, T149A | 1 | H69- | s79, F86L | 1 |
| H20 | F86L, I151V | 1 | H72 | F86L, A104T | 1 |
| H23- | F86L, s248 | 1 | H73 | s45, s65, T149A | 1 |
| H24 | A3D,F86L | 1 | H74 | V56I, F86L | 1 |
| H25- | F86L, s111 | 1 | H76 | K60R, I75T, F86L | 1 |
| H27 | I75V, F86L | 1 | H77 | s53, D85E, F86L | 1 |
| H28 | L2R, F86L | 1 | H83- | L2R, F86L, s102 | 1 |
| H30 | F86L, E251D | 1 | H86 | F49S, F86L | 1 |
| H31 | F86L, s187, E208G | 1 | H90- | F86L, s229 | 1 |
| H32- | F86L, s239 | 1 | H91- | F86L, s192 | 1 |
| H33 | F86L, V125A | 2 | H93 | s30, F86L, V190E, F229S | 1 |
| H35 | F86L, E251A | 2 | H94 | F86L, Y240H | 1 |
| H38 | F86L, K99E | 1 | H97 | F86L, s91, L94P | 1 |
| H41 | F86L, T149A | 1 | H99 | E41G, F86L | 1 |
| H48 | E41V, F86L | 1 | H100- | s32, s53, F86L, s193 | 1 |
| H51 | F86L, M186V | 1 | | | |

'-' indicates clones contain silent mutation; 's' means silent mutation and is followed by the residue position.

(iii) Interestingly, almost all tHisF mutants (91 out of 92 mutants, except clone H73) contained a common nucleotide mutation T256C which turned codon TTT into CTT and caused a missense mutation F86L (Table 4.2.2.1 and Table 4.2.2.2a). 46 out of the 91 mutants contained additional point mutations, in which 33 mutants have missense mutation/s besides the F86L exchange (including 2 redundant clones); the other 13 mutants contained silent mutation/s besides F86L exchange (Table 4.2.2.1).

(iv) In terms of base mutations, there were totally 54 point mutations in the whole selected clones (excluding the repeated point mutations, Table 4.2.2.2a), 32 of which are missense mutations leading to 32 amino acid exchanges, 22 of which are silent mutations. The transition mutation A to G and T to C had the highest occurrence with 17 and 18 times in 54 point mutations, counted for 65% of total point mutations, their inverse transition mutations G to A and C to T counted up to 5 and 2 times in 54 point mutations, counted for 13% of total point mutations. All the rest base mutations including A to C, C/A, A to T, T to A, G to T, T to G, G to C transversions only count up to 12 times in 54 point mutations (22%) (Table 4.2.2.2b).

**Table 4.2.2.2a** Base changes, codon changes and amino acid changes in tHisF-variants selected from eHisF$^-$ complementation

| Base changes | Codon changes | Type of mutations | Amino acid changes |
|---|---|---|---|
| T5G | CTC/CGC | Missense | L2R |
| C8A | GCT/GAT | Missense | A3D |
| A73G | AAC/GAC | Missense | N25D |
| C90T | GGC/GGT | Silent | silent30 |
| T96C | CCT/CCC | Silent | silent32 |
| A122G | GAG/GGG | Missense | E41G |
| A122T | GAG/GTG | Missense | E41V |
| T135C | GAT/GAC | Silent | silent45 |
| A138G | GAA/GAG | Missense | E46D |
| T146C | TTT/TCT | Missense | F49S |
| T159C | ACT/ACC | Silent | silent53 |
| G166A | GTT/ATT | Missense | V56I |
| A179G | AAG/AGG | Missense | K60r |
| G195A | CTG/CTA | Silent | silent65 |
| T224C | ATT/ACT | Missense | I75T |
| G228T | CCG/CCT | Silent | silent76 |
| T237C | GTT/GTC | Silent | silent79 |
| C255A | GAC/GAA | Missense | D85E |
| T256C | TTT/CTT | Missense | F86L |
| A273G | GAA/GAG | Silent | silent91 |
| T281C | CTG/CCG | Missense | L94P |
| A295G | AAG/GAG | Missense | K99E |
| T306C | ATT/ATC | Silent | silent102 |
| G310A | GCT/ACT | Missense | A104T |
| C374T | GTT/GCT | Missense | V125A |
| T390C | GAT/GAC | Silent | silent130 |
| A394G | AAG/GAG | Missense | K132E |
| A443G | AAC/AGC | Missense | N148S |
| A445G | ACG/GCG | Missense | T149A |
| A451G | ATC/GTC | Missense | I151V |
| G456C | CTG/CTC | Silent | silent152 |
| A532G | ACA/GCA | Missense | T178A |

| | | | |
|---|---|---|---|
| A556G | ATG/GTG | Missense | M186V |
| T557C | ATG/ACG | Missense | M186T |
| T561C | ATT/ATC | Silent | silent187 |
| T569A | GTG/GAG | Missense | V190E |
| G576A | GTG/GTA | Silent | sllent192 |
| A579G | CTA/CTG | Silent | silent193 |
| A580G | ACC/GCC | Missense | T194A |
| A623G | GAA/GGA | Missense | E208G |
| A648T | GCA/GCT | Silent | silent216 |
| T651C | GGT/GGC | Silent | silent217 |
| T654C | GCT/GCC | Silent | silent218 |
| T686C | TTT/TCT | Missense | F229S |
| T687C | TTT/TTC | Silent | silent229 |
| T701A | GTT/GAT | Missense | V234D |
| A714G | AAA/AAG | Silent | silent238 |
| G717A | GAG/GAA | Silent | silent239 |
| T718C | TAT/CAT | Missense | Y240H |
| A735G | GGA/GGG | Silent | silent245 |
| T737C | GTG/GCG | Missense | V246A |
| A744C | GTA/GTC | Silent | silent248 |
| A752C | GAG/GCG | Missense | E251A |
| G753T | GAG/GAT | Missense | E251D |

**Table 4.2.2.2b** Occurrence of base mutations***** (derived from Table 4.2.2.2a)

| Base mutation | Number of occurrence | Probability | Type of mutations |
|---|---|---|---|
| A to G | 17 | 65% | Transition |
| T to C | 18 | | |
| G to A | 5 | 13% | |
| C to T | 2 | | |
| A to C | 2 | 22% | Transversion |
| C to A | 2 | | |
| A to T | 2 | | |
| T to A | 2 | | |
| G to T | 2 | | |
| T to G | 1 | | |
| G to C | 1 | | |
| C to G | 0 | | |

***** Coding strand of the gene

(v) In terms of amino acid mutations, there were 32 amino acids exchanges located at 29 positions in all the selected mutants, which constituted 11.5% of the whole 253 amino acid of tHisF variants (Table 4.2.2.1, Fig 4.2.2.1). In more detail, several amino acids had higher occurrence frequencies in the 32 selected variants: Ala was observed 6 times, Asp 5 times, Val 4 times, Glu 4 times, Ser 3 times and Thr 3 times; in terms of the substituted amino acids in tHisF variants: Glu was replaced 5 times, Val 5 times, Ile 3 times, Lys 3 times, Phe 3 times, Thr 3 times in the 32 mutants. There were 3 positions that were substituted by more than one amino acid in the selected mutants, they were E41G (clone H99) and E41V (clone H48), M186T (H66) and M186V (H51), E251A (H35, H40) and E251D (H30) (Table 4.2.2.1 & Table 4.2.2.3).

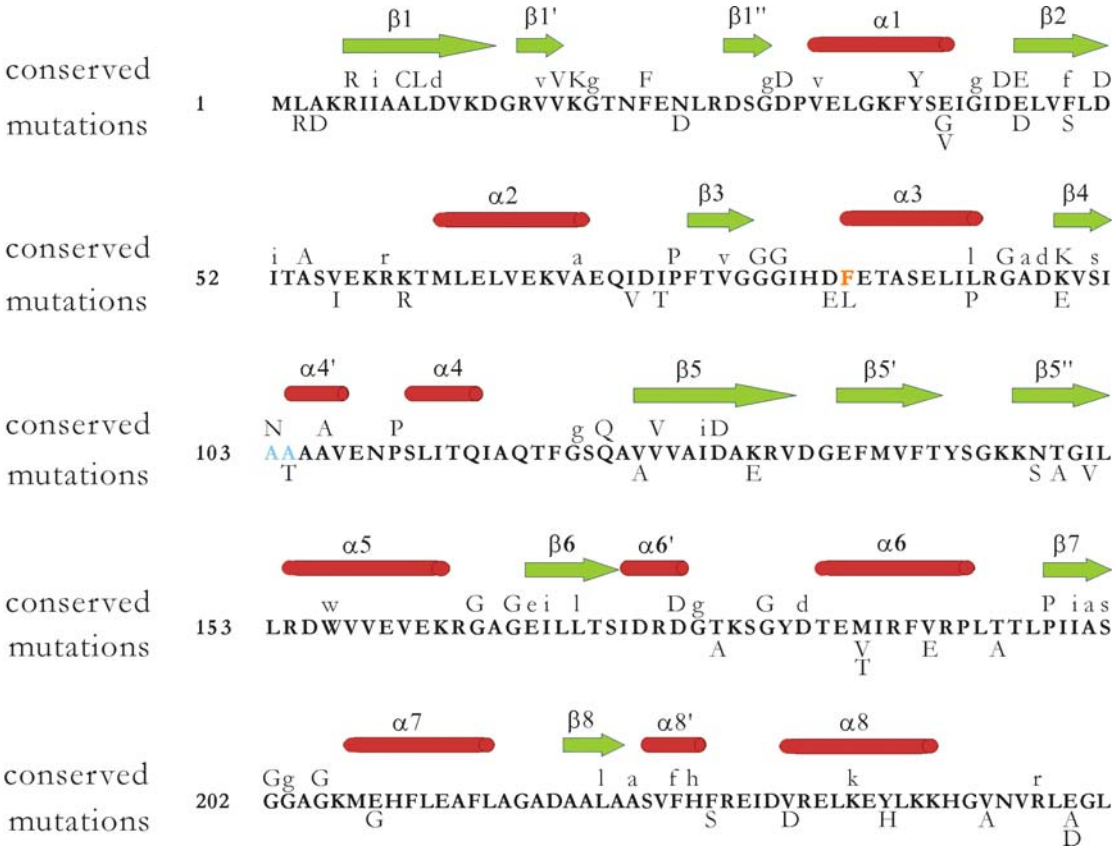



**Fig 4.2.2.1** Amino acid mutations in the secondary structure of tHisF. Green arrows represent β-strands, red cylinders represent α-helices. The conserved amino acid residues are listed above the corresponding positions, with the upper case letters indicate invariant residues and the lower case letters indicate more than 90% conservation in known HisF. All mutations are shown under the corresponding residues, E46, M186 and E251 have additional mutations. The common mutation position F86 is shown in orange, the designed mutation N103A and T104A are in cyan. (The structure data was based on Lang D et al, Science 2000)

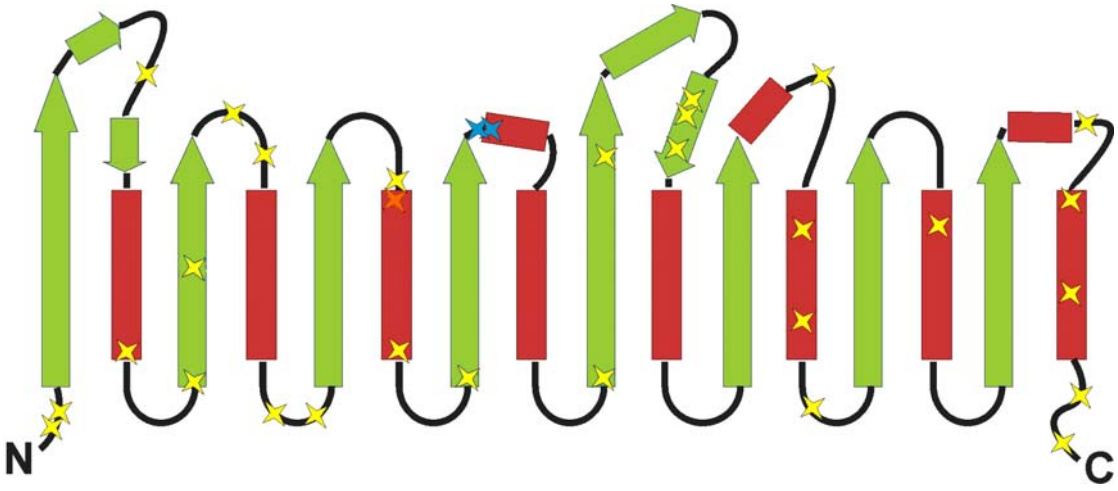


**Fig 4.2.2.2** Localization of the mutations in tHisF secondary structure elements. The stars in the figure indicate the mutated residues in that region; the orange star in α3 represents the common mutation F86L. The cyan stars are the mutations by design, N103A and T104A.Green arrows represent β-strands, red cylinders represent α-helices. From left to right:
β1/α1, β2/α2, β3/α3, β4/α41, β5/α5, β6/α6, β7/α7, β8/α8

**Table 4.2.2.3** Amino acids in tHisF variants and the corresponding substitutions in eHisF$^-$ selection. The sorting order is descending according to the number of exchanged amino acids. "x2" indicates two types of mutations.

| **Mutated amino acid in tHisF variants selected from eHisF$^-$ complementation** | | | | **Substitute amino acids for the mutated ones in tHisF selected from eHisF$^-$ complementation** | | |
|---|---|---|---|---|---|---|
| amino acid mutated | total number in $^{syn}$tHisF | mutated positions | number of mutated positions | amino acid emerged | position | number of substitute |
| V | 24 | V56<br>V125<br>V190<br>V234<br>V246 | 5 | A | 125A<br>149A<br>178A<br>194A<br>246A<br>251A | 6 |
| E | 22 | E41 x 2<br>E46<br>E208<br>E251x2 | 4 | D | 3D<br>25D<br>46D<br>234D<br>251D | 5 |
| F | 13 | F49<br>F86<br>F229 | 3 | V | 41V<br>73V<br>151V<br>186V | 4 |
| I | 20 | I73<br>I75<br>I151 | 3 | E | 85E<br>99E<br>132E<br>190E | 4 |
| K | 17 | K60<br>K99<br>K132 | 3 | S | 49S<br>148S<br>229S | 3 |
| T | 14 | T149<br>T178<br>T194 | 3 | T | 75T<br>104T<br>186T | 3 |
| A | 26 | A3<br>A104 | 2 | G | 41G<br>208G | 2 |
| L | 24 | L2<br>L94 | 2 | R | 2R<br>60R | 2 |
| N | 5 | N25<br>N148 | 2 | I | 56I | 1 |
| M | 5 | M186 x 2 | 1 | H | 240H | 1 |
| D | 17 | D85 | 1 | P | 94P | 1 |
| Y | 4 | Y240 | 1 | | | |

(vi) In terms of mutations in the secondary structures, there were mutations occurred in six α-helices (α1, α3, α4, α6, α7 and α8), three β-strands (β2, β4 and β5) (Fig. 4.2.2.2, Table 4.2.2.4), six loops at C-terminal face of the barrel and four loops at N-terminal face of the barrel ( Fig. 4.2.2.2, Table 4.2.2.5). Notably, the third piece of β5-strand β5" (when viewed in the tHisF-tHisH complex crystal structure, this region is a long loop in stead of a strand, Douangamath A et al 2002) was a hotspot of mutation, in which three neighboring residue N148, T149 and I151 were substituted

by 148S (H10), 149A (H19, H41, H73) and 151V (H20), respectively (Table 4.2.2.1, Fig 4.2.2.1and Fig 4.2.2.2).

**Table 4.2.2.4** Amino acid mutations localized in the α-helix and β-strand

| **Amino acid mutations localized in the α-helices and β-strands** | | | |
|---|---|---|---|
| α-helix | | β-strand | |
| α1 | E41G<br>E41V | β2 | E46D<br>F49S |
| α3 | F86L<br>L94P | β4 | K99E |
| α4 | T104A*<br>A104T | β5 | V125A<br>K132E |
| α6 | M186T<br>M186V<br>V190E | | |
| α7 | E208G | | |
| α8 | V234D<br>Y240H | | |

* designed substitutions

**Table 4.2.2.5** Amino acid mutations localized in the loop area

| **Amino acid mutations localized in the loop area** | |
|---|---|
| loops connecting β-strand and α-helix ( loops at the C-terminal face of the TIM barrel) | loops connecting α-helix and β-strand (loops at the N-terminal face of the TIM barrel) |
| N25D | L2R |
| V56I | A3D |
| K60R, | I73V |
| D85E | I75T |
| N103A* | T194A |
| N148S | E251A |
| T149A | E251D |
| I151V | |
| T178A | |
| F229S | |

* designed substitutions

(vii) The barrel of tHisF can be divided into two halves, N-terminal half and C-terminal half. The mutation frequencies found in both halves were similar. Equally, 15 mutated positions were founded in each half, and there were three residue positions which had two substitutes (Table 4.2.2.6, Fig 4.2.2.1).

**Table 4.2.2.6** Amino acid mutations localized in N- and C- terminal halves

| **Amino acid mutations localized in N- and C- terminal halves** | |
|---|---|
| N-terminal half (1-123) (18 different mutations) | C-terminal half (124-253) (17different mutations) |
| L2R | V125A |
| A3D | K132E |
| N25D | N148S |
| E41G | T149A |
| E41V | I151V |
| E46D | T178A |
| F49S | M186T |
| V56I | M186V |
| K60R | V190E |
| I73V | T194A |
| I75T | E208G |
| D85E | F229S |
| F86L | V234D |
| L94P | Y240H |
| K99E | V246A |
| N103A* | E251A |
| T104A*, A104T | E251D |

* designed substitutions

(viii) As to the putative substrate binding site (N103, T104) in the tHisF protein, the experiment result revealed that the designed mutation N103A and T104A all remained no change and still retained HisF activity with only one exception that in clone H72, a A104T reversion was observed (Table 4.2.2.1, Fig 4.2.2.1).

### 4.2.2.5 Implication of the mechanism of random mutagenesis in the t*his*F gene library

As described in section 4.2.2.4(ii), 93.9% selected clones contained point mutations at non-randomized positions. Where did these point mutations come from? Is there any specific mutational pattern in the random mutagenesis? To answer these questions, the first concern is the polymerase which was used in the construction of gene library.

During the construction of t*his*F gene library, *Taq* polymerase was used in all the PCR reactions, including assembling of nucleotide fragments, amplification of the assembled genes and DNA shuffling. *Taq* polymerase does not have proofreading function and is used dominantly in error-prone PCR to diversify the gene library. It would be interesting to check if *Taq* polymerase is the only source of the introduced mutations.

Wild type *Taq* pol I catalyzes frequent A to T and G to T transversions and relative high frequent T to C transition (Huang MM et al 1992). An investigation showed transversions count for 64% of base substitutions, transitions count for 36% of base substitutions by wild type *Taq* pol I (Suzuki M et al 2000; Table 4.2.2.7, right column).

However the point mutations observed in eHisF$^-$ selected mutants showed a significantly different mutational spectrum from reported error types introduced by *Taq* pol I (Table 4.2.2.7, three columns from left). Analysis of the base mutations in eHisF$^-$ selected t*his*F variants showed A to G and T to C transition mutations had the highest occurrence with 17 and 18 times in 54 observed point mutations (counted for 65% of point mutations, cf. Table 4.2.2.2b), G to A and C to T occurred 5 and 2 times in 54 point mutations (13%), which meant totally the transition mutations counted for 78% of point mutations. The transversions including all the rest base mutations A to C, C to A, A to T, T to A, G to T, T to G, G to C and C to G only occurred 12 times in 54 point mutations, which counted for 22% of point mutations.

**Table 4.2.2.7** Comparison of mutational types between observed mutations from eHisF$^-$ complementation selection and mutations preferred by *Taq* pol I*****.

| Type of mutations | Base mutation | Mutation probability Observed (%) | | Mutation probability by w.t. *Taq* pol I* (%) | |
|---|---|---|---|---|---|
| Transition | A to G | 31.5 | 65 | 1.6 | 25 |
| | T to C | 33.3 | | 23.0 | |
| | G to A | 9.3 | 13 | 1.6 | 11 |
| | C to T | 3.7 | | 9.8 | |
| Transversion | A to C | 3.7 | 22 | 0 | 64 |
| | C to A | 3.7 | | 0 | |
| | A to T | 3.7 | | 34.4 | |
| | T to A | 3.7 | | 4.9 | |
| | G to T | 3.7 | | 21.3 | |
| | T to G | 1.9 | | 1.6 | |
| | G to C | 1.9 | | 1.6 | |
| | C to G | 0 | | 0 | |

* Data from Suzuki M et al., 2000.

This observation indicated a complicate situation concerning the origin of base mutations in the synthetic t*his*F gene libraries. There may be more than one factors contributing to the random base mutations. While *Taq* pol I certainly contributed to the mutations by mismatching of dNTP in the extension of DNA, another possible source of random mutagenesis could be from incorrect incorporation of the nucleoside precursors during oligonucleotide synthesis. To understand totally this special mutational spectrum, more efforts are needed.

### 4.2.2.6 Importance of the residues near to active centre of HisF enzyme

The partially randomized t*his*F gene library (section 4.1) has nine randomized residue positions in its t*his*F genes: C/A9, D11, L169, T171, I173, K179, S201, A204 and L222. All of them are located in the vicinity of the substrate within a distance of 5Å, and their side chains orientation are toward the substrate or the inner side of the substrate interacting cleft. None of the nine residues involve in main chain contacts with the substrate the enzyme catalyze, and none of their side chains point to interior of the protein. As we knew, main chain contacts cannot be engineered by amino acid replacement, and side chains pointing into interior of protein should not be touched

because of their contribution to structure stability. Table 4.2.2.8 lists randomized residues in tHisF proteins (when the t*his*F genes were expressed into their proteins), the side chain orientation of these residues and their possible functions in tHisF were shown as well.

**Table 4.2.2.8** Randomized residues in tHisF library, the side chain orientation and their possible function.

| tHisF residue | Orientation of side chain | Possible function | Conserved residues or not |
|---|---|---|---|
| C/A9 | Toward cleft | Postulatedly involved in the ammonia tunneling | + |
| D11 | Toward cleft | Essential role in catalysis of HisF | + |
| L169 | Toward cleft | Postulatedly involved in the ammonia tunneling | - |
| T171 | Toward cleft | Difficult to assign a function without further evidence | - |
| I173 | Toward cleft | Difficult to assign a function without further evidence | - |
| K179 | Sideway | Difficult to assign a function without further evidence | - |
| S201 | Toward cleft | Postulatedly involved in the ammonia tunneling | + |
| A204 | Toward cleft | Difficult to assign a function without further evidence | - |
| L222 | Toward cleft | Postulatedly involved in the ammonia tunneling | + |

'+' means yes, '-' means no

How important are the nine randomized positions in HisF activity? Must all nine residues be required in order to maintain HisF function? The experiment results revealed that there was no mutation in the controlled randomized positions (cf. section4.2.2.4 (i)).

In these randomized positions, D11 had already been confirmed being essential for HisF catalytic activity although detail catalytic mechanism needs to be elucidated (Beismann-Driemeyer S et al., 2001); C9, L169, S201 and L222, are possibly involved in the ammonia tunnel across the β barrel of the HisF (Douangamath et al 2002, cf. section 4.2.1.5 (iii)). Four of the nine residues are conserved (C9, D11, S201, L222) in known HisFs (cf. Fig. 4.4.2.1, conserved residues in HisF sequence). So that no mutation was observed at D11, L169, S201 and L222 is reasonable. However, no substitution occurred at C/A9 position seems to be unexpected, although the complementation results is consistent with the *in vitro* experiment result that C9A substitution doesn't influence the HisF activity (Beismann- Driemeyer et al 2001).

The rest four randomized positions are not known of any role yet (T171, I173, K179 and A204). The interesting phenomenon is that no mutations in these sites were found. A possible explanation is that these residues may participate in some functions of HisF which are not understood yet, and library composition problem cannot be excluded, i.e. those clones with wild-type residues at their randomized positions were

over represented.

As introduced in chapter1, the active site of $(\beta\alpha)_8$-barrel enzymes is always located at the C-terminal face of the central eight-stranded β-barrel (top of the barrel). Besides the nine randomized positions, all observed substitutions that were located at the top of tHisF barrel (N25D, V56I, K60R, D85E, N148S, T149A, I151V, T178A, F229S, Table 4.2.2.5) did not significantly change the amino acid chemistry. None of these mutations belongs to the conserved amino acids. This observation is consistent with the common thought that the conserved residues at top of the TIM barrel are important for the enzyme activity, correspondingly, they are not replaceable.

Notably, the third segment of β5-strand, called β5" (this region is a long loop instead of a strand when viewed in the tHisF-tHisH complex crystal structure, Douangamath A et al 2002) was a hotspot for mutation, in which three neighboring residue N148, T149 and I151 were substituted by 148S (clone H10), 149A (clone H19, H41, H73) and 151V (clone H20), respectively (Table 4.2.2.1, Fig 4.2.2.1and Fig 4.2.2.2). This region showed little relevance to activity and enzyme stability. The high frequency of mutation at the β5"region may indicate that it is only a connective region or we can even presume that it is a redundant region in the protein. This finding showed that not all the loops at C-terminal face of tHisF are important for its catalytic activity.

### 4.2.2.7 Importance of the residues involved in formation of the possible tHisF-eHisH heterodimer

In bacteria, HisF forms a heterodimer with HisH, yielding a glutaminase-synthase complex. According to the X-ray structure of the heterodimeric bienzyme complex of tHisF-tHisH (Douangamath A et al 2002), the HisF cyclase subunit barrel has a rather flat surface at the N-terminal face, to which the HisH subunit is bound. Investigation of substitutions at the region which may involve in formation of the possible tHisF-eHisH heterodimer could give suggestion to which are the key residues in maintaining the heterodimer and whether they are replaceable.

The eHisF$^-$ complementation result revealed that 8 residues located at the region of bottom of barrel (N-terminal face of β-strands) were substituted by 10 amino acids: L2R, A3D, E41G and E41V, I73V, I75T, T194A, V246A, E251A and E251D (Table 4.2.2.4 and Table 4.2.2.5, Fig. 4.2.2.2).

Since the construction of gene library was based on *his*F gene from *Thermotoga maritima* and the complementation experiment was conducted in E coli HisF$^-$ strain, the possible bienzyme complex should be formed between tHisF variant and eHisH, which means tHisF interacted with eHisH to fulfill the histidine synthesis in E. coli. To understand the interaction of tHisF and eHisH, 3D structure of the complex would be ideal. Unfortunately, neither crystal structure of HisF nor HisH from E coli is available to date. Under such circumstance, it still makes sense by comparing the

sequence of tHisF and eHisF, tHisH and eHisH, and by making use of the structural information of tHsiF:tHisH.

An analysis of specific inter-subunit interactions between tHisF and tHisH identified ten protein- protein hydrogen bonds, which are distributed almost uniformly over the entire interface (Douangamath A et al 2002). The residues involved in the identified inter-subunit H-bonds are summarized in Table 4.2.2.10a and Table 4.2.2.10b. Table 4.2.2.10a shows the residues of tHisF involving in the inter-subunit H-bonds formation, their partner residues in tHisH, and a comparison to the corresponding residues in eHisF according to the sequence alignment of tHisF to eHisF. Table 4.2.2.10b shows the residues of tHisH involving in H-bonds formation, their partner residues in tHisF and a comparison to the corresponding residues in eHisH according to the sequence alignment of tHisH to eHisH.

**Table 4.2.2.10a** Residues of tHisF involving in inter-subunit H-bonds formation, its partner residues in tHisH and a comparison to the corresponding residues in eHisF according to the sequence alignment of tHisF to eHisF.

| **Residues in tHisF** | M1 | A3 | S40 | D45 | A70 | D74 | D98 | N247 |
|---|---|---|---|---|---|---|---|---|
| **Partner residues in tHisH*** | N124 +E157 | W123 | S183 | W123 | R18 | K181 S183 | K181 | Y136 |
| **Corresponding residues in eHisF** | M1 | A3 | A40 | D45 | A70 | D74 | D98 | E254 |

*** Data from Douangamath A et al., 2002.**

**Table 4.2.2.10b** Residues of tHisH involving in inter-subunit H-bonds formation, its partner residues in tHisF and a comparison to the corresponding residues in eHisH according to the sequence alignment of tHisH to eHisH.

| **Residues in tHisH** | R18 | W123 | N124 | Y136 | E157 | K181 | S183 |
|---|---|---|---|---|---|---|---|
| **Partner residues in tHisF*** | A70 | A3 D45 | M1 | N247 | M1 | D74 D98 | S40 D74 |
| **Corresponding residues in eHisH** | K17 | W119 | N120 | Y132 | N158 | R181 | G183 |

*** Data from Douangamath A et al., 2002.**

Table 4.2.2.10a shows that most of the residues involved in inter-subunit H-bond formation in tHisF-tHisH complex are conserved in tHisF and eHisF, except S40 (A40 in eHisF) and N247 (E254 in eHisF), while Table 4.2.2.10b shows that most of the residues of tHisH involved in inter-subunit H-bond formation in tHisF-tHisH complex are conserved in eHisH, except E157 (N158 in eHisH) and S183 (G183 in eHisH). This is a strong evidence supporting that the interaction between tHisF and tHisH can be used to reason the possible interaction between tHisF variant and eHisH.

The observed residue substitutions located at the region of bottom of HisF barrel, and a comparison to the corresponding residues in eHisF according to the sequence alignment of tHisF to eHisF were shown in Table 4.2.2.11(cf. Fig. 4.2.2.1 and Fig.4.2.2.2), most of these residues are conserved in tHisF and eHisF (except T194 and E251 of tHisF).

**Table 4.2.2.11** Amino acid substitutions located at bottom of tHisF barrel and a comparison to the corresponding residues in eHisF according to the sequence alignment of tHisF to eHisF.

| **Amino acid in tHisF** | L2 | A3 | E41 | I73 | I75 | T194 | V246 | E251 |
|---|---|---|---|---|---|---|---|---|
| **Substitution/s** | R | D | G/V | V | T | A | A | A/D |
| **Corresponding residues in eHisF** | L2 | A3 | E41 | I73 | I75 | C200 | V253 | C258 |

In these substituted residues, only the amino acid A3 involves in the inter-subunit H-bond formation with W123 of tHisH (Table 4.2.2.10a), it was replaced by a structurally and chemically different amino acid Asp. View of the existence of H-bond in this region shows that replacement of A3D retains the H-bond forming ability between the main chain amide of residue3 in tHisF and main chain carbonyl oxygen of W123 in tHisH, further more, a new H-bond is formed between the side chain amide of D3 in HisF and main chain carbonyl oxygen of W123 in tHisH (Fig. 4.2.2.3). That is to say A3D exchange enhanced the inter-subunit interaction by additional H-bond formation.

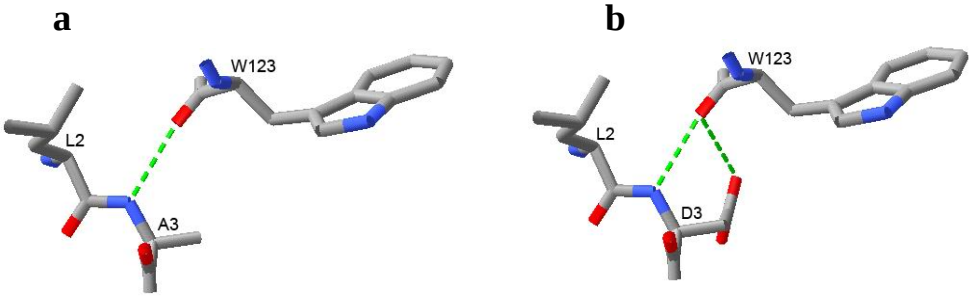



**Fig. 4.2.2.3** View of inter-subunit H-bond between residue3 of HisF and W23 of HisH (C atom, in grey; N atom, in blue; O atom, in red; H atom, not shown; H-bond, broken green line; modeled on the structure of tHisF-tHisH complex)

a. H-bond between A3 of tHisF and W123 of tHisH;

b. H-bond between D3 of tHisF and W123 of tHisH

The other residues that were substituted at that region, such as L2R, E41G/V, I73V, I75T, T194A, V246A and E251A/D do not involved in the inter-subunit H-bond formation, hence, have less contribution to the protein-protein interaction.

Referring to Table 4.2.1.10a, the fact is that almost all the amino acids which involved in the inter-subunit H-bond formation remained not changed (except A3D, where Asp can contribute more by increasing the number of inter-subunit H-bond). This observation indicates that amino acids involved in inter-subunit interaction are changeable only if the substitution can enhance the interaction, such as A3D.

### 4.2.2.8 Variability of the conserved residues involved in the putative ammonia tunneling

Since free ammonia is mainly present as nonreactive ammonium ion under physiological pH conditions, the nascent ammonia which is produced by HisH needs to be sequestered to avoid protonation in physiological pH conditions. The structure of the tHisF-tHisH heterodimer indicates that $(\beta\alpha)_8$-barrel of HisF is used as a channel for transferring ammonia from the HisH active site to the HisF active site (Douangamath A et al., 2002). The substitutions in tHisF variants which were located in the centre barrel of tHisF included: E46D, F49S, K99E, V125A, K132E (Table 4.2.2.4); in which E46, F49, K99 are conserved residues in known HisFs, and E46 and K99 were supposed to involve in the ammonia tunneling (see paragraph below).

There are already some investigations focused on the question of the ammonia tunneling. In one investigation, two conserved residues, R5 and E46 of HisF were shown to be crucial for the glutamine-dependent reaction of the bienzyme complex but not for the ammonia dependent reaction of the isolated cyclase subunit. The substitution of E46A, E46G impaired the glutamine hydrolysis reaction. This observation indicates that E46 is involved in the ammonia tunneling (Klem TJ et al., 2001). According to crystal structure of tHisF-tHisH complex, 17 residues were postulated involving in the ammonia tunneling: R5, I7, C9, E46, V47, T78, G86, K99, S101, V126, A128, E167, L169, I199, S201, A220, L222. There is a conserved salt bridge ring (cluster) consisting of four residues R5-E46-K99-E167 existed in molecule of HisF protein (Fig 4.2.2.4-A, B). These four residues are invariant within the known HisF sequences. Each of the two positively and two negatively charged side chains forms two salt bridges with its neighbors, resulting in a ring of salt bridges, which closes the putative ammonia tunnel. A hypothesis is this salt bridge served as a ‘gate’ on the interface of HisF-HisH complex, the nascent ammonia transferred through the interface upon the opening of this ‘gate’ (Douangamath A et al 2002).

The mutation E46D located at the position which was supposed to involve in the formation of salt bridge ring at N-terminal face of the central β strands barrel in HisF (Fig 4.2.2.4-C). This E46D substitution could be seen as a chemically conserved replacement which possibly did not change the formation of the salt bridge.

A very intriguing exchange was found at residue 99, the conserved basic residue Lys is substituted by acidic amino acid Glu. The K99 was supposed to involve in the interaction with another invariant residue E167 forming the conserved salt bridge in HisF. In a biochemical research, the basic residue K360 in yeast ImGPS (a functional counterpart of tHisF, K360 in yeast ImGPS corresponding to K99 in tHisF) was mutated to Ala and Arg (models of these mutations were shown in Fig 4.2.2.4-D and Fig 4.2.2.4-F). The experiment revealed that replacement of the lysine 99 with arginine or alanine decreases the efficiency of cyclase catalyzed reaction slightly, and resulted in a proposition that K360 (K99 in tHisF) plays a structural role in ImGPS

(Myers RS et al., 2003). However, molecular dynamics (MD) simulations showed ammonia can freely diffuse into the $(\beta\alpha)_8$ barrel of the synthase domain via a small side opening between E46, K99, P76, and D98 (cf. Fig.4.2.2.4-B). The simulations and kinetic analyses suggested lysine99 (and corresponding glutamate) controls the passage of ammonia into the barrel (Amaro RE et al., 2005).

Although structure information (Douangamath A et al 2002) suggested K99 involves in the formation of a conserved salt bridge, experimental result (Myers RS et al., 2003) and computer simulation (Amaro RE et al., 2005) suggested K99 plays a structure role and is a critical position in the bienzyme activity, however, the K99E substitute obviously changed the pattern of the possible salt bridge ring (Fig. 4.2.2.4-E). Combining with the biochemical result that replacement of the K99 with Arg or Ala only decreases the efficiency of cyclase catalyzed reaction slightly (Myers RS et al., 2003) and MD simulation showed the possible existence of another channel for ammonia tunnelling, the K99E substitution observed in eHisF$^-$ complementation reveals that the 'salt gate' did not necessarily exist, and the hypothetic mechanism of the nascent ammonia transportation through a 'salt gate' is less likely possible.

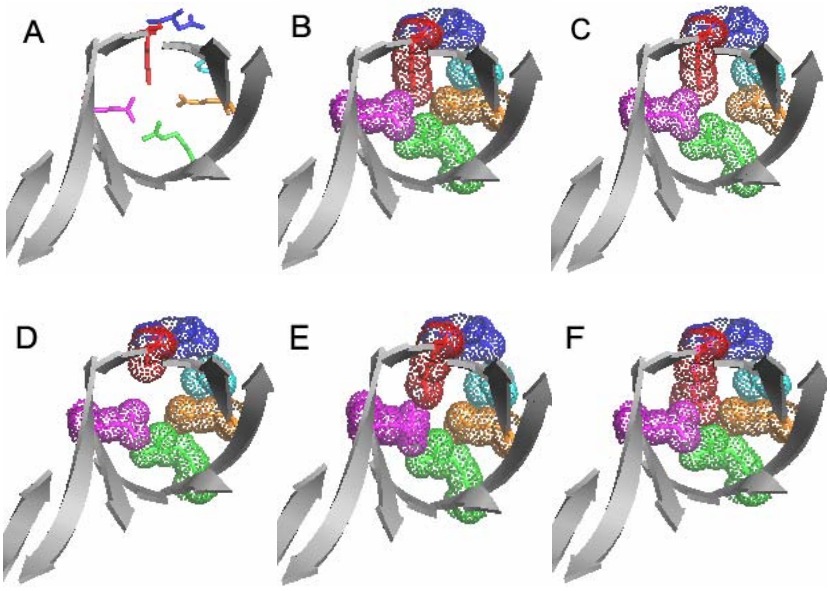



**Fig.4.2.2.4** Molecular model of a hypothetic gate of salt-ring at the N-terminal face of HisF barrel and mutations at this region

R5, green; E46, orange; P76, cyan; D98, Blue; K99, red; E167, magenta; β-strands of barrel, ribbons in gray (arrows point to C-termini)

A. Stick representation of the 'gate' region of w.t.tHisF

B. Dots representation of the 'gate' region of w.t.tHisF

C. Dots representation of mutation E46D at the 'gate' region

D. Dots representation of mutation K99A at the 'gate' region

E. Dots representation of mutation K99E at the 'gate' region

F. Dots representation of mutation K99R at the 'gate' region

As discussed in the first paragraph of this topic, the centre channel of the β-barrel of HisF is supposed to be used for transferring nascent ammonia (Douangamath A et al., 2002), what mutation occurred at this region other than the E46D and K99E? Another observed substitution at centre of β-barrel is a conserved Phe49 was replaced by a Ser. Although this Phe is conserved residue in known HisF protein, no specific function of Phe49 was identified yet. Phenylalanine and serine are structurally unrelated which means Phe49 isn't invariable for HisF activity. In view of H-bond forming situation (data not shown), main chain amide of residue49 and main chain carbonyl oxygen atom of T78 in tHisF forms a H-bond, the replacement of F49S does not influence this H-bond formation, in addition, the side chain of Ser forms another H-bond with main chain carbonyl oxygen atom of T78, so the F49S substitution increased the number of intra-H-bond in the centre of barrel, the mutant may benefit from increased structural stability.

The rest substitutions at the centre β-barrel include V125A, K132E. The substitution of V125A did not change the polarity and H-bond forming ability of the residue at this position. Mutation of K132E and previously discussed K99E suggests that either basic or acidic residue, strong polar residue in interior of central β-barrel of HisF can maintain HisF activity.

#### 4.2.2.9 Exchanges which is possibly related to conformational mobility of tHisF

Exploring the factors responsible for exhibiting extreme thermal stability of thermophilic proteins will contribute to the knowledge of mechanism of protein stability, as well as to design stable proteins. The experiment conducted here is the other way around, i.e. observing a hyperthermostable enzyme adapting itself to the mesophilic environment. The results may also provide clues to understand the mechanism of protein thermal stability.

Proteins are multifunctional: their amino acid sequences simultaneously determine folding, function and turnover. Evolution selected for compromises between rigidity (stability) and flexibility (folding/function/degradation), to the result that the free energy of stabilization of globular proteins in solution is the equivalent to only a few weak intermolecular interactions (Jaenicke R. 2000)**.** Recent developments revealed that the surface and surface electrostatics of a protein are more important for stability than previously thought. Very large stability differences in some cases are due to only one or very few point mutations (Sandgren M et al., 2003). It has also been shown that stability differences between homologous enzymes may be due to very few (out of many) of the naturally occurring sequence variations (Eijsink VG et al., 2004).

A remarkable observation in the eHisF$^-$ complementation was that a substitutions occurred at residue position 86, a Phe of tHisF was substituted by Leu in ~94% selected clones (section 4.2.1.4-2), where the nucleotide codon changed from TTT to CTT.

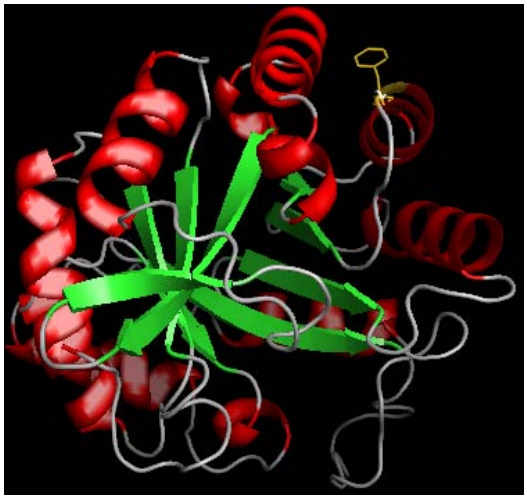


**Fig4.2.2.5** Mutation of F86 in tHisF.
The aromatic ring of Phe (shown in orange) is exposed to solvent.

Concerning the origin of the F86L change, if it is possible that F86L has at some stage been introduced by chance into that subset of the library that carries wild-type codons in all nine randomized positions - for instance when wild-type t*hisF* gene was 'crossed in' at the step of Stemmer-shuffling? It would then not be surprising that it was pulled out again when selecting for HisF$^-$ complementation. However, this objection is not supported by the fact that 94% selected clones contain the F86L change. If this F86L change was introduced in the PCR procedure of DNA shuffling and the change doesn't have any functional significance, then the gene without this change should have more chance to be presented in the selected clones as the wild-type gene is the template of the PCR reaction which produces the mutant products. This objection is also not supported by the fact that F86L comes associated with additional exchanges. Pairs of such additional exchanges are generally different in different clones making it unlikely (although it is not impossible) that they are of clonal origin (*i.e.* go back to the same variant that had only F86L and acquired the other mutations later).

Furthermore, in another unrelated experiment, when the t*his*F gene library was subjected to eHisA$^-$ complementation, a HisA$^-$ E coli strain selected clone HA03 also has the F86L exchange. The HA03 does not have nine wild-type codons at the randomized positions, so the presence of F86L is not coupled to the situation in eHisF$^-$ selection exclusively. The nature of the nucleotide exchange of HA03 is completely different, where in HA03 the codon changed from TTT to TTA (cf. section 4.2.3.3). This shows that this F86L has arisen independently from the one(s) found in the set of HisF$^-$ complementers.

All of these could either indeed mean that F86L has functional significance or be the result of a freak coincidence (not impossible but not very likely). Our hypothesis is the F86L exchange is of functional significance, and F86L contributes the same favorable characteristics to HisF function in one case and to HisA function in the

other. Since this can hardly be a contribution to two different catalytic activities, then it has to do with general structural properties, the most plausible assumption is the conformational mobility.

The substituted residue position 86 is located at the beginning of α3 helix in tHisF protein. The side chain of residue 86 in tHisF is exposed to solvent (Fig 4.2.2.5). In general, solvent exposed residues are not supposed to be key residues in the stabilization of proteins, or in other words, conformational mobility of proteins. But stabilization of proteins by substitution of solvent exposed residues had also been observed in several works.

One example is thermolysin-like protease from *Bacillus stearothermophilus* (TLP-ste), analysis of a series of mutations of a residue on the protein surface drew a conclusion that hydrophobic interactions between the phenylalanine ring and aliphatic parts of the surrounding (mostly polar) side chains were beneficial for stability (Imanaka T et al., 1986; Van den Burg et al., 1991, 1994). In another example, three neighbouring phenylalanines on the surface of CspA, a small cold shock protein which functions as single-stranded RNA and DNA binding protein, were mutated singly and in combination to leucine and to serine, experimental result revealed Phes on the protein surface are required not only for protein function but also for protein stability (Hillier BJ et al., 1998). In addition, three exposed Phe residues in CspB from *Bacillus subtilis* showed to be important for both binding to nucleic acids and conformational stability (Schindler T, 1998).

All these observations suggested that exposed aromatic residues especially Phe had special contributions to the protein stability, and mutation of the Phe to Leu, Ser cause the destabilizing of protein. The F86L mutation may reflect the reverse adaptation of a hyperthermostable protein to less stability demanding circumstance. However, it needs further experiment to test if the Phe86 is really important in determining the conformational flexibility, consequently, the thermal stabilities of tHisF protein. The test for this property was out of the scope of this thesis but is underway in the laboratory.

In this tHisF to eHisF adapting experiment, several Phe residues were found to be substituted, besides the F86L mutation, a conserved Phe49 (located on β2 strand, buried, clone H86) was replaced by a Ser; a similar substitution was F229S (located at C-terminal loop between β8 strand and α8 helix, clone H93). These results imply that under mesophilic environment where hyper-thermostability is no longer necessary in protein, serine at these positions is more favored.

In addition, by summarizing the substitutions in selected tHisF variants (Table 4.2.2.3), a fact was there were obviously higher occurrence of non-polar Ala in the selected clones, most of them came from the substitution of polar residue Threonine: T103A (by design), T104A (by design), T149A (clone H19, H41, H73), T178A (clone H56), T194A (clone H55), E251A (clone H35), and these changed residues all located

at different loops. Views of short range interaction of these substitutions (data not shown) reveal that the number of H-bonds formed by Ala is obviously less than the number of H-bonds formed by Thr or Glu, this implies that the tHisF protein losses part of short range interaction in adaptation to mesophilic environment, in other words, the hyperthermostable proteins possess more residues that can form stronger short range interactions and this may be one of the factor contributing to the thermostability.

The other notable exchange in selected tHisF variants is relatively high frequency of substitution of lysine: K60R (at C-terminal loop between β2 strand and α2 helix, clone H76), K99E (at N-terminal of β4 strand, clone H38), K132E (at β5 strand, clone H67). Lys was found to play an important role in thermophiles, as atoms that have a substantial positive charge make a significant contribution for cation-pi interactions so as to enhance protein stability from mesophilic to thermophilic (Gromiha MM et al., 2002). These may explain that several Lys residues were exchanged in selected tHisF variants under mesophilic condition.

However, not all observed mutations are apt to lower structural stability, one example is a conserved residue Leu94 located at α3 helix was found to be mutated to proline (L94P, clone H66 and H87). This position is the end of α3 helix according to the separate tHisF structure (Lang D et al 2000). Proline has a rigid five-membered ring which is normally seen at the turn of a protein chain and is the main contributor of entropy induced stability.

#### 4.2.2.10 Functional importance of residues N103 and T104

In the tHisF gene library, besides the nine randomized position, two point mutations, N103A and T104A were introduced in the construction of the t*hisF* gene library. N103 and T104 are presumed to be the substrate binding site of tHisF by computer modeling the tHisF crystal structure with its substrate PRFAR. This was in accordance with the structure evidence of yeast ImGPS in complex with substrate PRFAR, where G364 (loopβ4α4', corresponding to N103 in tHisF) and T365 (α4', corresponding to T104 in tHisF) interact with AICAR phosphate (Chaudhuri BN et al., 2003).

However the eHisF$^-$ complementation results showed that two consecutive mutations N103A and T104A did not influence the tHisF activity during genetic selection although there was an A104T inverse mutation occurred in the selected clone (clone H72, Table 4.2.1).

Sequence and structure alignment shows a high degree of similarity between the ImGPS from eukaryotic yeast and ImGPS from bacterial *T. maritima* (Amaro RE et al., 2005), so the information from structure of *S. cerevisiae* ImGPS in complex with cryo-trapped nucleotide substrate PRFAR can be used as reference for tHisF and its substrate. The interactions between *S. cerevisiae* ImGPS and its substrate as well as

the corresponding residues in tHisF were shown in Table 4.2.2.9 (data from Chaudhuri BN et al., 2003).

**Table 4.2.2.9** Residues of *S. cerevisiae* ImGPS interacting with its substrate and the corresponding residues in tHisF

| Interactions between ImGPS and its substrate PRFAR | | | |
|---|---|---|---|
| Residues in ImGPS from *S. cerevisiae* | Contacted atoms at binding site | Nature of contact | Corresponding residues in tHisF |
| G475 | Main chain amides | interact with the glycerol phosphate group | G177 |
| G501 | | | G203 |
| A523 | | | A224 |
| G524 | | | S225 |
| D245 | Side chain O or N | direct or water-bridged hydrogen bonds with glycerol hydroxyl groups | D11 |
| K258 | | | K19 |
| D474 | | | D176 |
| S500 | | | G202 |
| G332 | Main chain amides and side chain O of Thr | interact with AICAR phosphate | G82 |
| G364 | | | N103 |
| T365 | | | T104 |
| S402 | Side chain O | hydrogen bonds with ribose hydroxyl groups | A128 |
| D404 | | | D130 |

The complementation result of N103A and T104A can retain HisF activity indicates that those residues involved in substrate binding through main chain contact have less influence on enzyme activity.

### 4.2.2.11 Summary of eHisF$^-$ selection

After subjecting the pKK223-3-t*his*F library ($5.58 \times 10^6$ independent clones, partially randomized t*his*F gene library) to an auxotrophic E. coli strain UTH860 for eHisF complementation assay, substantial clones were observed to be able to complement eHisF activity, the selection ratio was $3.1 \times 10^{-3}$.

The eHisF complementation of selected t*his*F variants were confirmed by a second round complementation assay. Screening PCR of plasmids from twenty selected clones showed 100% of these plasmids are pKK223-3-t*his*F variants. Plasmid DNA of 98 clones were successfully sequenced, in which 92 out of 98 clones contained synthetic t*his*F gene with mutations, only 6 out of 98 plasmids were synthetic t*his*F gene without mutation.

Totally there were 54 point mutations in selected clones (excluding the repeated point mutations), 32 of which are missense mutations leading to 32 amino acid mutations, 22 of which are silent mutations. A to G and T to C base mutations had the highest occurrence, together with their inverse mutations, the transition mutations counted 78% of total point mutations. All the transversion base mutations count for 22% point mutations. This observation showed a significant difference from mutational spectrum

of *Taq* pol I and indicated a complicate situation concerning the origin of base mutations in the synthetic t*his*F gene libraries. To understand totally this special mutational spectrum, more efforts are needed.

The observed 32 amino acids mutations located at 29 residue positions, constituted 11.5% of the whole 253 amino acids of tHisF. Amino acid Ala, Asp, Val, Glu, Ser and Thr had higher appearance frequencies in the selected mutants; amino acid Glu, Val, Ile, Lys, Phe, and Thr had higher frequency to be replaced. There were 3 positions that were observed to be substituted by more than one amino acid, which were E41, M186 and E251. The highest number of residue exchanges in single clones was three.

In view of mutations in secondary structures, six α-helices, three β-strands, six loops at C-terminal face of the β-barrel and four loops at N-terminal face were involved in amino acid mutations. Notably, strand β5" (the third piece of β5) was a hotspot of mutation, in which three neighboring residue N148, T149 and I151 were substituted by Ser, Ala and Val, respectively.

No mutation was found at the nine randomized position of t*his*F gene. Totally, four conserved residues were found to be substituted: E46D, F49S, L94P and K99E, in which E46 and K99 are residues which were supposed to be involved in the formation of a hypothetic salt bridge ('salt gate') at the N-terminal face of β-barrel of the HisF consisting of four conserved residues R5-E46-K99-E167. The exchanges of E46D and K99E indicated that the 'salt gate' did not necessarily exist.

At the bottom of tHisF barrel, residue A3 which is involved in the HisF and HisH interaction was substituted by Asp, analysis of short range interaction shows A3D substitution enhanced the inter-subunit interaction by increasing the number of inter-subunit H-bond.

A remarkable observation from eHisF$^-$ complementation was 91 out of 98 selected tHisF variants contain a common mutation F86L. F86 located at α3 helix in tHisF structure and pointed outward from the barrel. A hypothesis is this residue may relate to the rigidity of hyperthermophilic enzyme tHisF, by mutating Phe to Leu, tHisF may adapt itself to mesophilic environment, as under mesophilic environment, the rigidity to resistant high temperature becomes not necessary but flexibility is more favored. To attest this assumption, further experiment e.g. fluorescence polarization which can provide further evidence to molecular flexibility is in need (unfortunately, this is out of the time table of this thesis work).

In addition, genetic selection result showed the designed consecutive mutations N103A and T104A don't affect HisF activity. This complementation result indicates that those residues involved in substrate binding through main chain contact have less influence on enzyme activity.

### 4.2.3 Directed evolution of HisA catalytic activity from tHisF library

In the first stage of this work, substantial t*his*F variants were drawn out of the controlled partially randomized t*his*F gene libraries in an eHisF$^-$ complementation assay, and plenty of information about the catalytic functional centre, protein-protein interaction, ammonia tunneling, structure stability and temperature adaptation of tHisF has been gained from those tHisF mutants (section 4.2.2).

To gradually elevate the demanding level of selection target, an E. coli *his*A deletion strain Hfr G6 (Matney et al 1964, Hofnung M, et al., 1971) was used for complementary assay to select t*his*F variants that are able to complement an eHisA$^-$ phenotype.

(The work in section 4.2.3.1 was carried out by Christiane Preiß under the help of Stina Schiller in this lab.)

#### 4.2.3.1 HisA$^-$ *E coli* strain and genetic selection for HisA activity from t*his*F library

An auxotrophic *Escherichia coli* strain Hfr G6 (Matney et al 1964, Hofnung M, et al., 1971) (section 2.1) whose *his*A gene was deleted from its chromosome and therefore cannot grow without histidine was used for complementation assay.

The plasmids library of pKK223-3-t*his*F ($5.58 \times 10^6$ independent clones, section 4.1.6) were transformed into the E. coli hisA deletion strain Hfr G6 by electroporation. The tranformants were plated on minimal VB agar plates (section 2.7) containing 100μg/ml ampicillin and incubated at 37°C for complementation selection, whereas the recipient of a plasmid containing wild-type t*his*F did not grew (negative control). The pKK223-3 expression vector contains a strong *tac* promoter (section 2.10, Fig. 2.3) and an ampicillin resistant gene. This vector has a high base expression level and the base expression level was shown to be enough for complementation in this work without the induction of isopropyl-D-thiogalactoside (IPTG, cf. section 4.2.2.2).

Since there was no supplement histidine which was required by the Hfr G6 strain to survive on the selective VB plate, clones which grew on the selective plate must be able to complement eHisA function. Colonies grew on the selective plates were picked and plasmid DNAs of these eHisF$^-$ selected clones were prepared using standard DNA Mini preparation protocol (section 3.2.4).

To confirm the eHisA$^-$ complementation selected variants, plasmids prepared from eHisF$^-$ selected clones were re-transformed into recipient eHisA$^-$ G6 cells, and plated on selective minimal VB plates without histidine supplement and incubated at 37°C,. The selected plasmids were proved to be able to complement eHisA function.

The plasmids DNA of twelve eHisA$^-$ selected clones were sequenced. One plasmid with the ability to complement eHisA$^-$ strain was tHisF-variant, and was termed HA03

(Tab.4.2.3.1). The other plasmids were integrated product of the expression vector and *his*A gene. This did not come to be too surprised, since that within a selection system with high selective pressure like complementation assay, the fused product would have more chance to be drawn out once the wanted gene (in this case, *his*A gene) from trace contaminant of chromosome DNA in the gene library had integrated with the vector containing the selective marker (in this case, ampicillin resistance).

### 4.2.3.2 Test of the HisF activity of HA03

In order to test whether the tHisF variant which can complement eHisA$^-$ strain still retain the original HisF activity, the plasmid of HA03 was transformed into an auxtrophic E coli strain UTH860 which contains an inactivated *his*F gene in its chromosome (section2.1, cf. section 4.2.2.2), and subjected to HisF$^-$ complementation assay. The previous experiment (section 4.2.2.4) had shown that wild-type synthetic t*his*F gene in pKK223-3 vector had the ability to complement eHisF$^-$ UTH860 cells.

The result showed that HA03 failed to complement eHisF$^-$ strain. This result indicated that those mutations introducing HisA activity to HA03 led to its lost at least part of its original HiaF activity and caused the failure of HA03 on eHisF$^-$ complementation.

### 4.2.3.3 Sequence result of eHisA$^-$ complementing gene

Totally, the DNA sequence of t*his*F variant HA03 carried ten residue exchanges, these mutations (and their location in secondary structure) were: C/A9R (located at β1 strand), F86L (α3 helix), N/A103T (loop4 at C-terminal face of β-barrel, between β4 strand and α4 helix), T104A (α4 helix, by design), T171I (β6 strand), I173R (α6 helix), G177D (loop between α6' and α6 helices), K179V (loop between α6' and α6 helices), S201Q (β7 strand), A204I (C-terminal face of β-barrel, loop between β7 strand and α7 helix). Additionally, there were four silent nucleotide mutations in the sequence of HA03, they were: A144G (s48; s, silent mutation; 48, number of the residue position), T351C (s117), T552C (s184), A579G (s193). All the exchanges are located in the C-terminal face of β-barrel, except the F86L. Table 4.2.3.1 shows nucleotides exchanges and amino acids exchanges in t*his*F variant HA03 (comparing with w.t.t*his*F); Fig. 4.2.3.1 shows the locations of amino acid exchanges in HA03 in secondary structure elements of tHisF; Fig. 4.2.3.2 shows the locations of amino acid exchanges in HA03 in 3D structure of tHisF.

**Table 4.2.3.1** Nucleotide exchanges and amino acid exchanges in clone HA03 (comparing with w.t.*this*F).

| mutated position | Randomized position in gene library | Wild type codon | Observed codon | A.A in $^{syn}$*this*F | Observed A.A. | A.A. exchange |
|---|---|---|---|---|---|---|
| **9** | + | GCT | CGT | Ala | Arg | + |
| **48** | - | GTA | GTG | Val | Val | - |
| **86** | - | TTT | TTA | Phe | Leu | + |
| **103** | - | AAC/GCT* | ACT | Gln/Ala | Thr | + |
| **104** | - | ACG/GCT* | GCT | Thr/Ala | Ala | + |
| **117** | - | GCT | GCC | Ala | Ala | - |
| **171** | + | ACC | ATC | Thr | Lle | + |
| **173** | + | ATC | CGG | Lle | Arg | + |
| **177** | - | GGC | GAC | Gly | Asp | + |
| **179** | + | AAA | GTT | Lys | Val | + |
| **184** | - | ACT | ACC | Thr | Thr | - |
| **193** | - | CTA | CTG | Leu | Leu | - |
| **201** | + | TCT | CAG | Ser | Gln | + |
| **204** | + | GCT | AAA | Ala | Lys | + |

'+' means yes, '-' means no; * Codon exchanges by design; A.A., amino acid
Color highlights: exchanged residues at random positions are in blue; exchanged residues at solvent exposed position is in orange; exchanged residues at presumed substrate binding positions are in cyan.

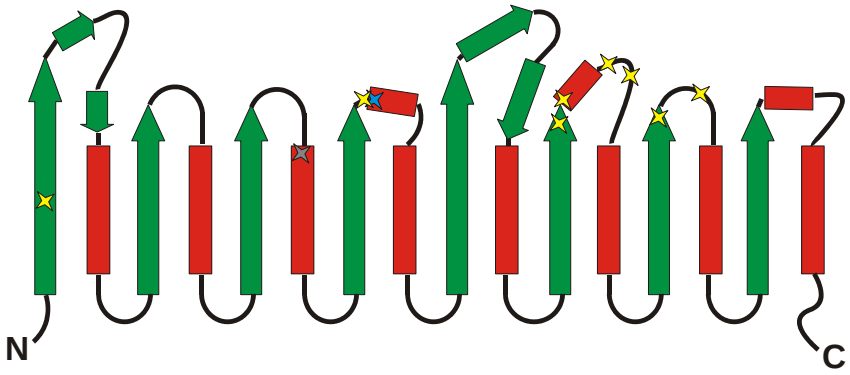



**Fig. 4.2.3.1** Representation of locations of amino acid exchanges in HA03 in secondary structure elements of tHisF

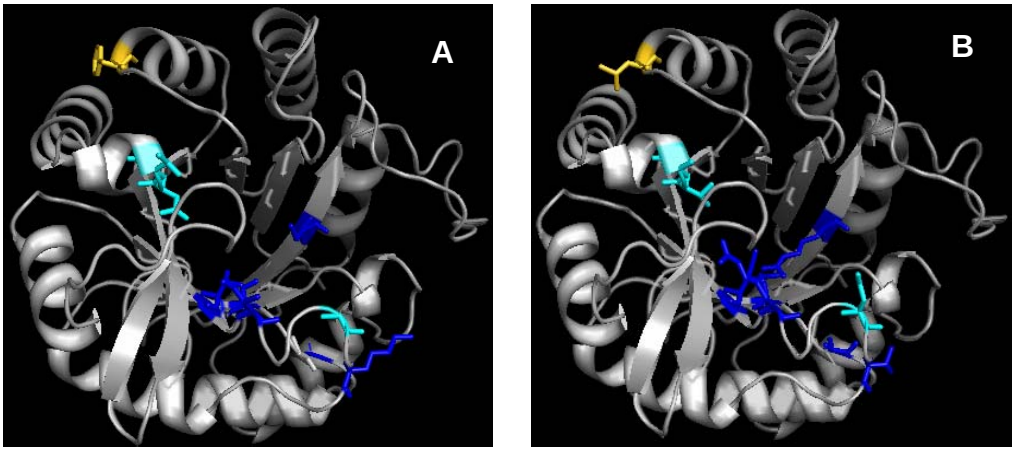



**Fig. 4.2.3.2** Representation of residue exchanges in HA03 in 3D structure of tHisF.
(A) Residue positions in $^{syn}$tHisF; (B) Observed residues in HA03. Color highlights: residues at randomized positions are in blue; residues 103, 104 and 177 are in cyan; residue 86 is in orange.

### 4.2.3.4 Analysis of selected result

The *his*A and *his*F genes are adjacent in most *his* operons responding the biosynthesis of amino acid histidine, the gene products HisA and HisF catalyze consecutive steps in histidine biosynthesis (Alifano P *et al*., 1996).

The enzyme *N*′-[(5′-phosphoribosyl) -formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase, HisA (EC 5.3.1.16), converts ProFAR into the aminoketoses *N*'-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR). HisF is part of a heterodimeric complex which is a glutamine amidotransferase, whereas the other subunit is HisH. HisF catalyzes the condensation of PRFAR and ammonia provided by HisH, through until now an unclear mechanism yielding 5-aminoimidazole- 4-carboxamide ribotide (AICAR) and imidazoleglycerol phosphate (ImGP) (Alifano P *et al*., 1996; Fig.4.2.3.3).

The substrate of HisA (ProFAR) and the substrate of HisF (PRFAR) share geometric similarity (cf. Fig 4.2.3.3). They are unusual metabolites with pseudo two-fold axis of symmetry. One half contains a ribose 5-phosphate moiety and the other contains another ribose 5-phosphate or an isomerized ribose 5-phosphate. The PRFAR is the intermediate metabolite between HisA and HisF. The X-ray structures of HisA and HisF show that both enzymes are folded as β/α barrels. Each half-barrel of these two enzymes contains a phosphate-binding motif that is imposed by the nature of their biphosphate substrates, ProFAR and PRFAR.

It was demonstrated that the substitution of Asp127 by valine in tHisA (tHisA-D127V) and the corresponding substitution of Asp130 by valine in tHisF (tHisF-D130V) generates phosphoribosylanthranilate isomerase (TrpF) activity, a related $(\beta\alpha)_8$-barrel enzyme participating in tryptophan biosynthesis (Jurgens C et al 2000; Leopoldseder

S et al., 2004). The ease with which TrpF activity can be established on both the tHisA and tHisF scaffolds highlights the functional plasticity of the $(\beta\alpha)_8$-barrel enzyme fold.

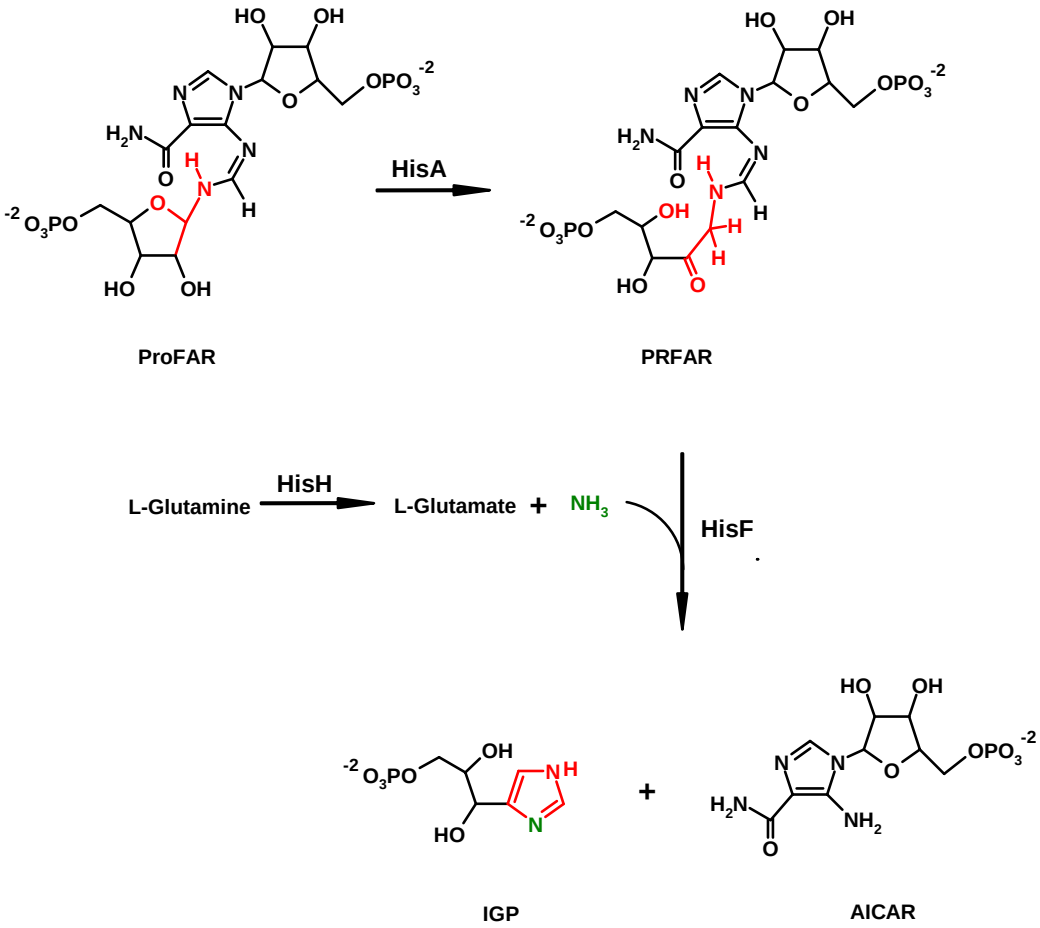



**Figure 4.2.3.3** Reaction catalysed by HisA and HisF, $NH_3$ is supplied by HisH

HisA and HisF share 24% sequence identity. Based on comparisons of gene and amino acid sequences, it was suggested that *his*A and *his*F have evolved via a series of gene duplication events (Fani et al., 1994). Although HisF showed detectable HisA activity in biochemical assay (Lang D et al., 2000), however, this intrinsic isomerase catalytic activity of HisF is too weak to complement eHisA$^-$ phenotype.

On one hand, tHisF may share a common ancestor with tHisA; on the other hand, the complementation was carried out in E coli Δ*his*A strain, so a sequence comparison of tHisF variant with eHisA and tHisA may give some useful information on which are the key exchanges enabling the tHisF variant, HA03, to acquire HisA activity. Table 4.2.3.2 showed residue comparison of $^{syn}$tHisF, HA03, eHisA and tHisA based on the structure based protein sequence alignment.

**Table 4.2.3.2** Residue comparison of tHisF, HA03, eHisA and tHisA (Data from structure based protein sequence alignment; Residue numbering is based on tHisF).

| | 9# | 11# | 86 | 103 | 104 | 169# | 171# | 173# | 177 | 179# | 201# | 204# | 222# |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **$^{syn}$tHisF (residue in w.t. tHisF)** | A (C) | D | F | A* (N) | A* (T) | L | T | I | G | K | S | A | L |
| **HA-03** | R | D | L | T | A | L | I | R | D | V | Q | K | L |
| **eHisA** | A | D | E | G | S | L | T | I | G | L | S | I | I |
| **tHisA** | A | D | L | S | S | V | T | I | G | L | A | I | I |

# Randomized positions in t*his*F gene library. * Exchanges by design. Color highlights: exchanged residues at random positions are in blue; exchange at solvent exposed position is in orange; exchanges at presumed substrate binding positions are in cyan.

As we know, the substrate of HisA and the substrate of HisF share geometric similarity (cf. Fig 4.2.3.3), hence, the information of HisF interacting with its substrate PRFAR could be used as a reference for the explanation of the gained function of HA03. The essential residues for tHisF activity such as D11, K19, D130 and D176 which involved in side chain contacting with its substrate (Table 4.2.3.3) remained not changed in HA03. Several exchanges in HA03 such as 103, 104 and 177 are involved in main chain contacting in the binding site of tHisF. This result is consistent with the fact that HisA and HisF sharing geometric similarity.

**Table 4.2.3.3** Residues of tHisF interact with its substrate (deduced from Table 4.2.2.9)

| **Interactions between HisF and its substrate PRFAR** | | |
|---|---|---|
| Residues in tHisF | Contacted atoms at binding site | Nature of contact |
| G177 | Main chain amides | interact with the glycerol phosphate group |
| G203 | | |
| A224 | | |
| S225 | | |
| D11 | Side chain O or N | direct or water-bridged hydrogen bonds with glycerol hydroxyl groups |
| K19 | | |
| D176 | | |
| G82 | Main chain amides and side chain O of Thr | interact with AICAR phosphate |
| N103 | | |
| T104 | | |
| D130 | Side chain O | hydrogen bonds with ribose hydroxyl groups |

An interesting observation was that a mutation F86L (codon exchange is TTT/TTA) in HA03 was also seen in the tHisF to eHisF adaptation experiment (section 4.2.2.4) but with different codon change (TTT/CTT) in 92.9% of t*his*F variants selected from

HisF$^-$ E coli strain UTH860. This residue is located at the beginning of α3 helix and is exposed to solvent (Fig. 4.2.3.2; cf. Fig 4.2.2.5). This observation gave further evidence that this phenylalanine may be related to the flexibility and structure stability of this protein, as it all occurred in the adaptation of hyperthermostable protein tHisF to mesophilic environment.

Comparison of HA03 with eHisA and tHisA (Table 4.2.3.2) suggested a postulation that the exchanges such as A103T and K179V may contribute more to the HisA activity acquiring than G177D and S201Q, while the exchanges A9R, T171I, I173R and A204R seem not to be necessary for the HisA activity acquiring. However, to identify the exchanges which are necessary for the HisA activity acquiring needs further experimental evidence.

### 4.2.3.5 Summary and perspective of directed evolution of HisA catalytic activity from tHisF library

Enzyme *N*'-[(5'-phosphoribosyl)formimino] -5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and imidazole glycerol phosphate synthase (HisF), which catalyze consecutive reactions in the biosynthesis of the amino acid histidine have a similar $(\beta\alpha)_8$-barrel structure. The substrate of HisA (ProFAR) and the substrate of HisF (PRFAR) share geometric similarity, which are unusual metabolites with pseudo two-fold axis of symmetry, one half contains a ribose 5-phosphate moiety and the other half contains another ribose 5-phosphate or an isomerized ribose 5-phosphate. The PRFAR is the intermediate metabolite between HisA and HisF.

The striking sequence and structural similarities have suggested that HisA and HisF may have evolved from one common ancestral enzyme. And biochemical assay revealed that tHisF did exhibit detectable HisA activity, however, this intrinsic isomerase catalytic activity of HisF is far below the need for HisA$^-$ complementation.

The gene library containing controlled partially randomized t*his*F gene in expression vector pKK223-3 was subjected to genetic complementation. An auxotrophic E coli strain Hfr G6 that lacks the *hisA* gene on its chromosome was used as recipient for eHisA$^-$ complementation assay. Plasmid containing synthetic wild-type t*hisF* was transformed into Hfr G6 recipient cells as negative control. A tHisF variant (termed HA03) that can efficiently complement *E coli his*A$^-$ strain was identified. Fresh recipient cells transformed with this plasmid confirmed the selective result.

DNA sequencing of HA03 revealed that the isolated t*hisF* variant carried the following ten amino acid exchanges: C/A9R, F86L, N/A103T, T104A, T171I, I173R, G177D, K179V, S201Q, A204I. Additionally, there were four silent nucleotide

mutations in the sequence of HA03, they were: A144G (s48), T351C (s117), T552C (s184), A579G (s193).

Complementing ability test of HA03 indicated that the introduced mutations in HA03 impaired its HisF activity, the acquired eHisA activity of HisF variant is at the expense of its original activity.

Perspective:

HA03 carried ten residue exchanges, a postulation is that less substitutions may be sufficient for the acquisition of HisA activity by tHisF variant. What is the minimal modification of the sequence which is sufficient to acquire HisA activity? This question may be addressed by producing each of the single amino acid substitution or double or triple substitutions (if necessary, systematical investigation on those ten residues maybe of interest) found in HA03 using site-directed mutagenesis, and further subject these single/double/triple mutation variants to complementation.

### 4.2.4 Groundwork and preliminary data in the effort toward creation of novel biocatalysts by directed evolution

#### 4.2.4.1 The targeted chemical reaction

This part is the third stage of this project. In biotechnology, it is tempting to alter thermo-stable enzymes' specificities towards different substrates, so they can be used as biocatalyst for organic synthesis.

The targeted reaction in this work is Aldol condensation, arguably, the most basic carbon-carbon bond forming reaction in chemistry and biology (Fig. 4.2.4.1).

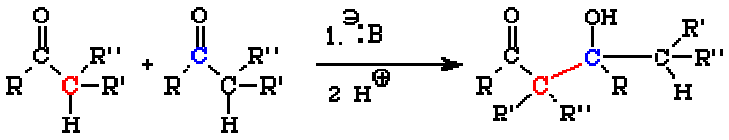


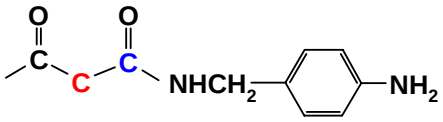
**Fig.4.2.4.1** General Features of the Aldol Reaction

#### 4.2.4.2 Search for biocatalysts for Aldol reaction via biopanning of phage displayed *this*F gene library against TSA

**(i) Design and synthesis of transition state analogue (TSA)**

The aldol reaction requires an aldehyde or ketone that contains at least one α-hydrogen. The α-carbon becomes nucleophilic when it is deprotonated by a base. The carbonyl carbon is electrophilic. Coulomb's Law brings these two oppositely charged species together to form a carbon-carbon (C-C) bond. Like other alcohols, β-hydroxyaldehydes and β-hydroxyketones are very prone to dehydration and produce α,β-unsaturated aldehydes or α,β -unsaturated ketons spontaneously.

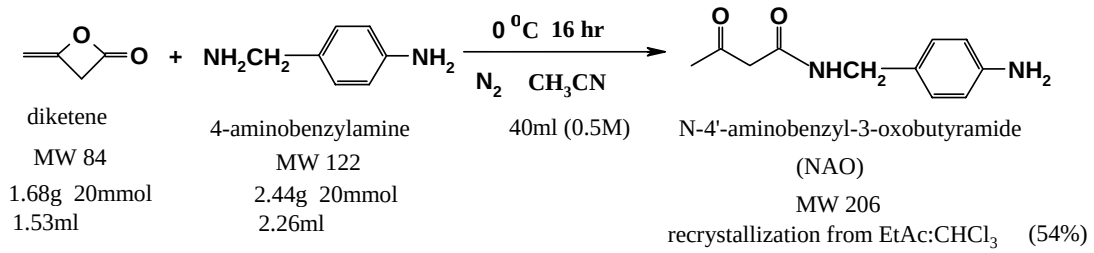
To construct a suitable structure as transition state analog (TSA), referring to Wagner and the co-worker's work (1995, Science), one compound N-4'-aminobenzyl-3-oxobutyramide (NAO) was designed to use the β-diketone structure to mimic the transition state of aldol reaction and use the aminobenzylamine group as linker for conjugating this molecule to a suitable carrier (Fig. 4.2.4.2)



N-4'-aminobenzyl-3-oxobutyramide
(NAO)

**Fig. 4.2.4.2** Molecular structure of transition state analog of Aldol reaction (NAO)
The β-diketone structure mimics the transition state of aldol reaction and the aminobenzylamine group serves as spacer for conjugating this molecule to a suitable carrier

Synthesis of the NAO molecule was carried out under the conditions described in Scheme 4.2.4.1. The product was confirmed by ESI-MS analysis (molecular weight: 207 ($M+H^1$), Fig. 4.2.4.3).




**Scheme 4.2.4.1** Synthesis of N-4'-aminobenzyl-3-oxobutyramide (NAO)

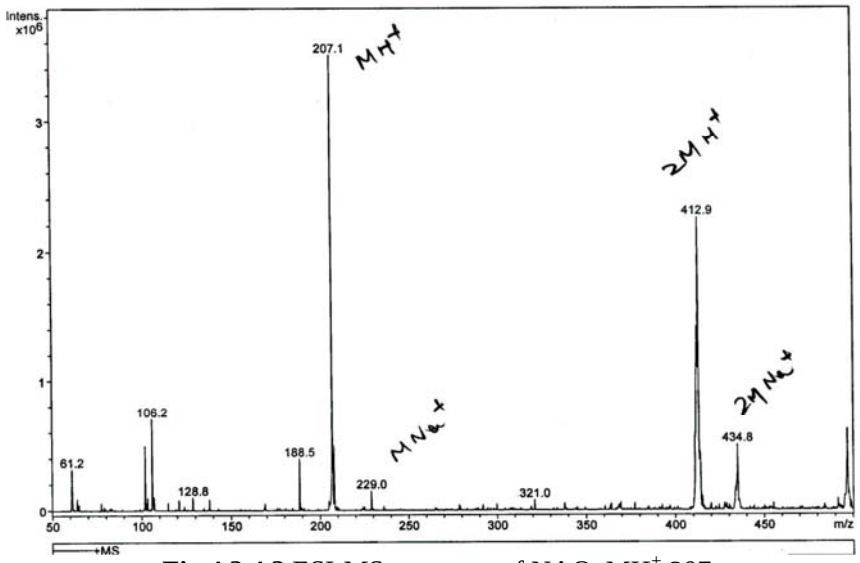


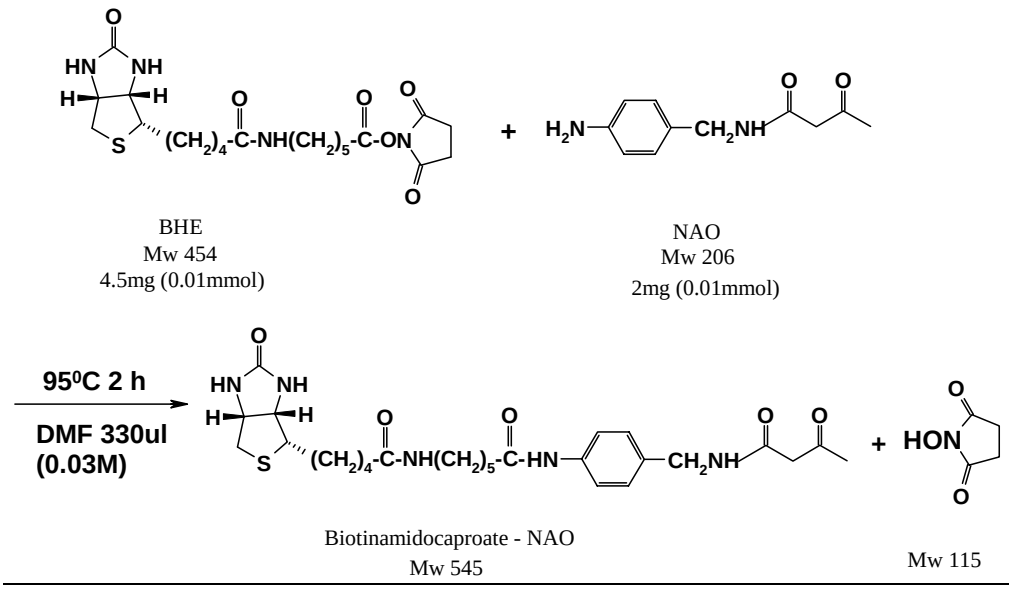

**Fig 4.2.4.3** ESI-MS spectrum of NAO, $MH^+$:207

## (ii) Immobilize the synthesized TSA

In order to use the TSA molecule as bait in bio-panning, the TSA needs to be properly connected to a suitable matrix.

(a) Conjugating TSA onto biotin derivative

Because biotin binds strongly and specifically with streptavidin, so steptavidin coated beads can be used for immobilization of biotin conjugated TSA. The conjugation of NAO to biotinamidocaproate N-hydroxysuccinimide ester (BHE) was carried out under conditions described in Scheme 4.2.4.2. ESI-MS analysis of the reaction mixture revealed the formation of the correct product (molecular weight: 546 ($M+H^1$), Fig 4.2.4.4).




**Scheme 4.2.4.2** Conjugating NAO to Biotinamidocaproate N-hydroxysuccinimide ester (BHE)

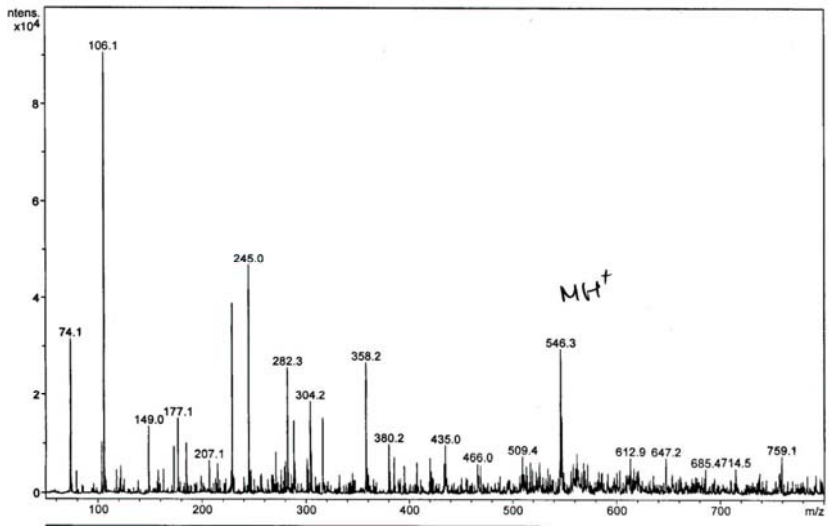
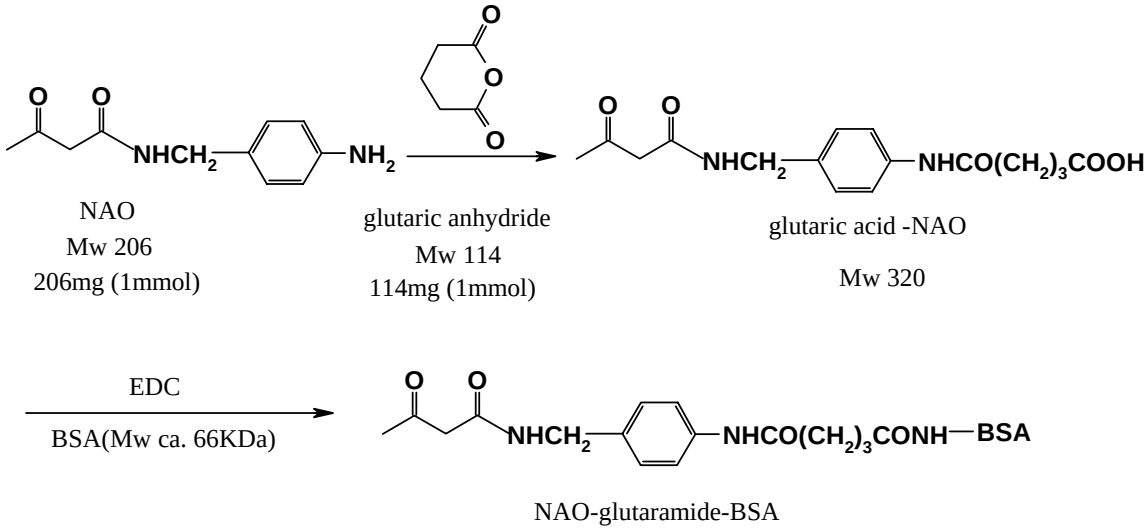



**Fig.4.2.4.4** ESI-MS analysis of the reaction mixture revealed the desired product: 546 ($MH^+$)

(b) Conjugating TSA onto Bovine serum albumin (BSA**)** (Wagner J et al., 1995 **)**

Another molecule, BSA, a protein which has a molecular weight ca. 66KDa (it is less expensive than biotin or biotin derivatives), was also considered as carrier for the immobilization of NAO. In this case, glutaric acid was used as a spacer between NAO and BSA. The reaction of NAO with glutaric anhydride was carried out under conditions described in the first step of Scheme 4.2.4.3. ESI-MS analysis of the reaction mixture revealed the formation of the correct product (molecular weight): 321 ($M+H^1$) (Fig. 4.2.4.5).




NAO, N-4'-aminobenzyl-3-oxobutyramide
EDC, 1-Ethyl-3-(3-dimethylaminopropyl) Carbodiimide
BSA, Bovine serum albumin

**Scheme 4.2.4.3** Synthesis pathway of conjugating NAO to BSA

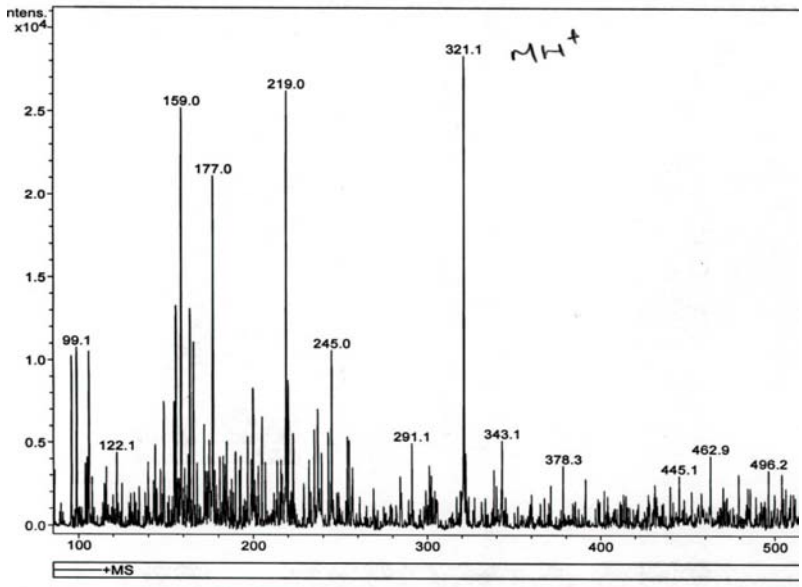



Fig. 4.2.4.5 **ESI-MS analysis of the reaction mixture revealed the presence of product: 321 ($MH^+$)**

These products could be used for bio-panning to select desired candidates for Aldol reaction's biocatalyst from phage displayed t*his*F gene library.

### 4.2.4.3 Directed evolution of DHDPS activity by genetic complementation

#### (i) DHDPS's biological function, DapA$^-$ strain and complementation assay

A number of enzymes catalyse the biologic Aldol condensation. Two mechanistic classes of aldolases enzymes have evolved, class I aldolases utilize the ε-amino group of a Lys in the active site to form a Schiff base with one of the substrates, which

activates the substrates as an Aldol donor. Class II aldolases are metalloenzymes that facilitate enolate formation by coordination to the substrate's carbonyl oxygen (Wagner J., et al 1995).

In order to evolve Aldolase catalytic activity from the controlled partially randomized t*his*F gene libraries, the auxotroph *E. coli* strain lacking dihydrodipicolinate synthase (DHDPS, *DapA*) activity was used to perform the genetic selection.

DHDPS (*DapA* ) catalyses the Aldol condensation of pyruvate and L-aspartate-semialdehyde (L-ASA), the first step in the biosynthesis of lysine via the diaminopimelate pathway (Fig. 4.2.4.6). The intermediate meso-diaminopimelic acid in this pathway not only serves as the precursor of lysine, but also involves in cell wall synthesis, subsequently the *E. coli* strain lacking DHDPS activity requires diaminopimelate to grow.

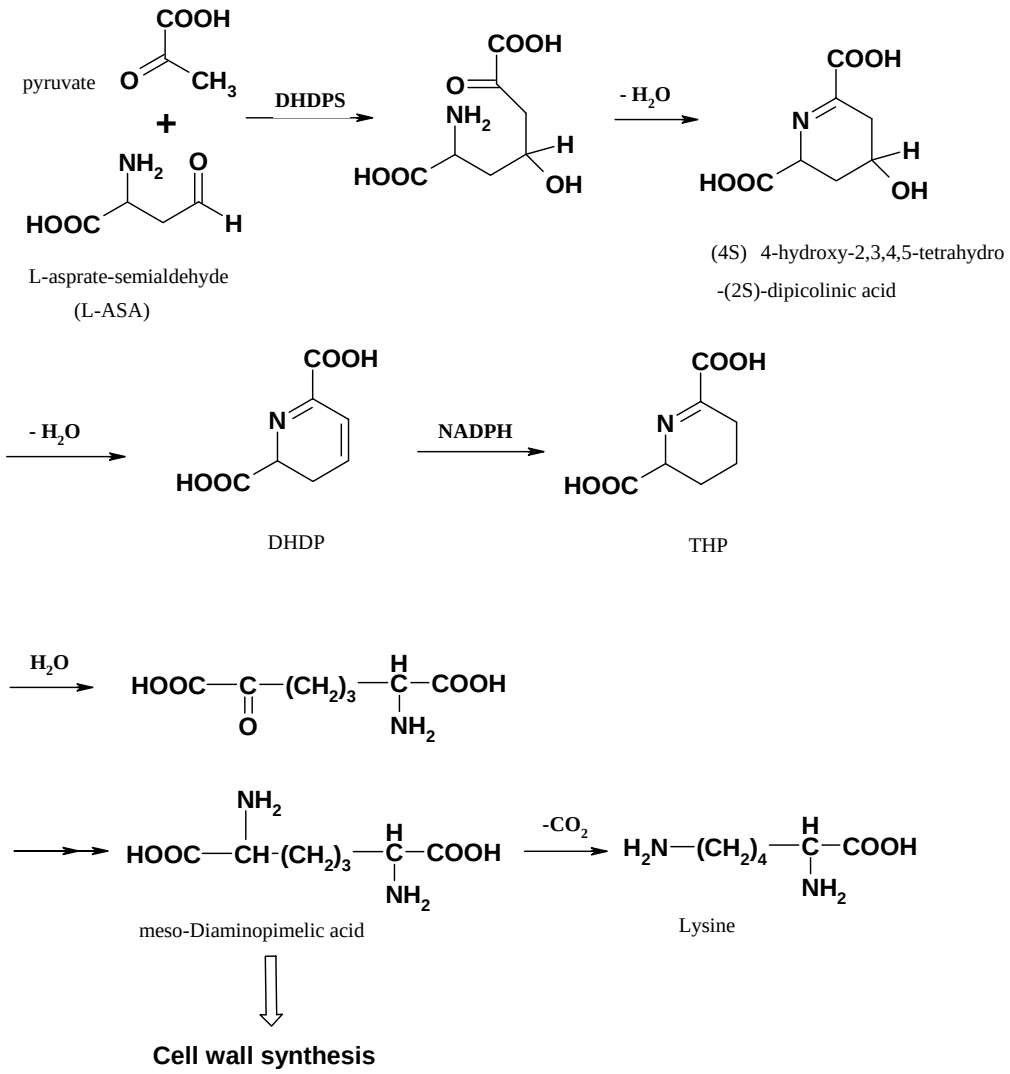



**Fig. 4.2.4.6** Lysine synthesis in *E.coli* via DHDP pathway

**(a) DapA mutation strain AT997 and complementation assay**

AT997 is an auxotroph *E. coli* strain with mutation on its *dap*A gene, consequently lacking DHDPS activity. Medium for this strain's maintenance is M9 minimal medium supplemented with 50 μg/ml DL-α,ε-diaminopimelate (DAP), 1 mM MgSO4, 1.7 mM thiamine hydrochloride, and 0.2% (wt/vol) glucose (Bukhari AI and Taylor AL 1971; Yeh P, et al 1988).

Functional complementation studies were performed by transforming plasmids library pKK223-3-t*his*F (section 4.1) into *E coli* strain AT997 and plating the transformants on M9 minimal plates (Difco agar) with the same supplements as for cell recovery plus 100 μg/ml ampicillin for plasmid selection but without DL- α,ε -diaminopimelate (DAP), incubated at 37°C.

The sequence of a plasmid from the best-grown clone showed it was an fused DNA of *dap*A gene and part of its adjacent *nlp*B gene of E. coli K12 chromosome in vector pKK223-3 (data not shown). The false positive result from fused DNA product of vector DNA and chromosome DNA was also seen in $HisA^-$ complementation, it is a big obstacle in this work.

Plasmids of clones from selective plates which were identified as t*his*F variants by colony screening PCR were subjected to second round transformation for confirming the selection. However, the quantitative characterization of the selected plasmids failed. This false positive result indicated that the spontaneous reverse of auxotroph AT997 had unneglectable impact on genetic selection experiment (the rate is as high as $10^{-5}$ ~$10^{-6}$, according to supplier's description and our test result, data not shown). For this reason, a *dap*A deletion strain derived from E. coli DH5α was constructed for this project (the work was carried out by Jessica Matter under the supervision of Prof. HJ Fritz).

**(b) ΔdapA *E. coli* strain DH5α-JM1 and complementation selection**

DH5α-JM1 was derived from DH5α by replacing the *dap*A gene in its chromosome with kanamycin resistance gene. Medium for strain maintenance is exactly the same as the medium for *dap*A mutation strain AT997. Experimental result showed the spontaneous reverse rate of Δ*dap*A strain DH5α-JM1 is lower than $10^{-8}$.

For complementation studies, plasmids library pKK223-3-t*his*F (section 4.1) and t*his*F library pre-selected by AT997 were transformed into DH5α-JM1cells by electroporation, separately. Transformed cells were plated on M9 minimal plates with the same supplements as for cell recovery plus 100μg/ml ampicillin for plasmid selection but without DL- α,ε -diaminopimelate (DAP) and incubated at 37°C. DH5α-JM1 harbouring plasmid pKK223-3-*dap*A was used as positive control (*dap*A gene was amplified from chromosome of E coli strain DH5α and cloned into

pKK223-3 vector). DH5α-JM1 harbouring pKK223-3-w.t.t*his*F was used as negative control.

Plasmids of individual clones from selective plates were isolated and subjected to colony screening PCR with three pairs of primers, two sets of primers specific for t*his*F (forward and reversed directions) and one set of primers for pKK223-3 vector to identify the selected genes. 8 out of 80 clones were identified that they contained plasmid pKK223-3-t*his*F. One of the eight plasmids didn't retain *dap*A$^-$ complementing ability in the second round transformation. Sequence of the 7 plasmids from selected clones revealed that three of them were *his*F-like DNAs and four of them were *dap*A and pKK223-3 fusions. The three *his*F-like DNAs were subjected to the next round transformation. After two round re-transformation, no plasmid was identified as t*his*F- variant.

This result indicated that those selected plasmids which were identified as t*his*F-like DNAs might co-existe with fused pKK223-3-*dap*A-*nlp*B. Chromosome DNA contamination was the cause of this type of false positive result. In the future, it is necessary to exclude chromosome DNA contamination thoroughly to avoid the false positive result.

**(ii) Establish the biochemical assay for DHDPS activity**

One of the groundwork for directed evolution of DHDPS activity is to establish the biochemical method for DHDPS activity assay.

The literature reports three methods for measuring kinetic data for DHDPS activity. Briefly, the easiest to perform is a discontinuous assay that monitors the formation of an unidentified adduct, after addition of o-aminobenzaldehyde, in the presence of acid. It is simple to set up and appears very sensitive to DHDPS activity, but shows an considerable lag phase and is therefore inappropriate for the measuring of initial rate data. The second assay directly monitors the change in absorbance at 270 nm when in the presence of imidazole buffer. This is thought to correspond to the spontaneous formation of dipicolinate from dihydrodipicolinate, but the details of the chemistry are unclear and the recording of initial rate data cannot be guaranteed. The third assay indirectly measures the reaction by monitoring the utilisation of NADPH at 340 nm by dihydrodipicolinate reductase (DHDPR), which is used as a coupling enzyme. This assay system allows the accurate monitoring of initial rate data and was used in this work (Joerger AC et al., 2003; Dobson R CJ et al., 2004). Protein DHDPS (DapA), DHDPR (DapB) and an unstable substrate L-aspartate- β -semialdehyde (L-ASA) are needed for DHDPA activity assay.

**(a) Expression and purification of DHDPS (DapA) and DHDPR (DapB)**

*dapA* gene and *dap*B gene were amplified from *E. coli* and cloned into the *Nco*I and *Xho* I restriction sites of vector pET_B_001 (a home-made derivative of pET21d, section 2.10) which can add C-terminal $His_6$-tag to the interest gene product. Primers for cloning *dapA* gene from E. coli chromosome into pET_B001 are DAPA_SEN with the restriction site NcoI and DAPA_ANT with the restriction site XhoI (section 2.9.1), a start codon ATG is incorporated in the primer DAPA_SEN. Primers for cloning *dapB* gene from E. coli chromosome into pET_B001 are DAPB_SEN_C with the restriction site NcoI and DAPB_ANT with the restriction site XhoI (section 2.9.1), a start codon ATG is incorporated in the primer DAPB_SEN_C.

The purified products of plasmid pET_B001-*dap*A and pET_B_001-*dap*B were transformed into E.coli DH5α to amplify these plasmid DNAs. The sequences of the two plasmids showed the successful cloning of the *dapA* gene and *dap*B gene into the expression vector pET_B_001. The expression of interested proteins was conducted in E.coli BL21(ΔE)plysS cells (protocol see section 3.4.2 and 3.4.3).

All the interested proteins with C-terminal $His_6$-tag were purified on Ni- Chelating Sepharose Fast Flow agarose column (Pharmacia Biotech) (section 3.4.5), the purity of products were estimated by SDS-PAGE gel (Fig.4.2.4.3). Protein concentrations were measured photometrically and calculated by using calculated extinction coefficients (section 3.4.1).

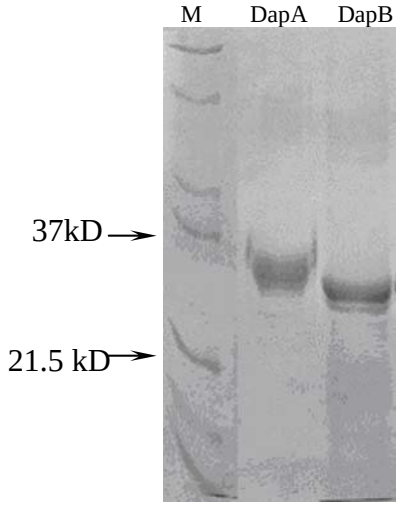
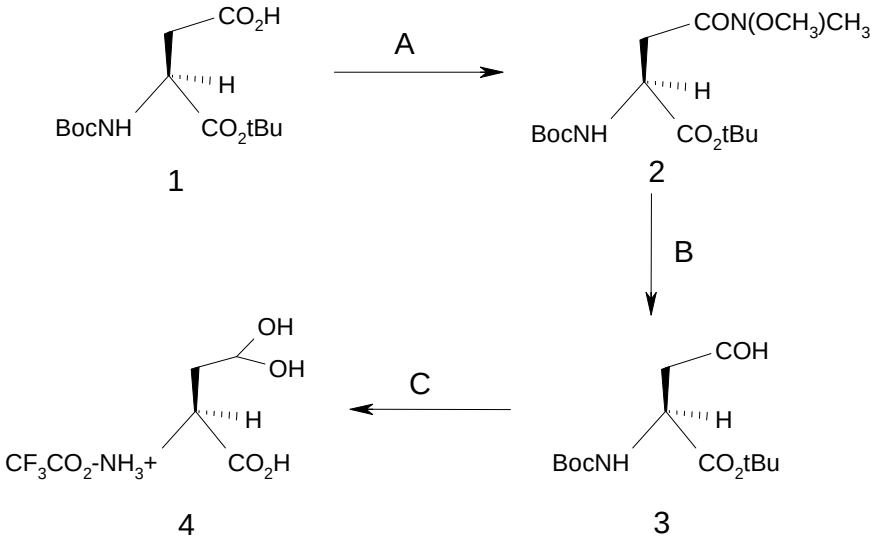



**Fig 4.2.4.3 SDS-PAGE of Ni-IMAC purified proteins**
(DapA 292aa, DapA-$His_6$ ~32.8kD; DapB 273aa, DapB--$His_6$ ~30.7kD)

**(b) Synthesis of L-ASA as one of the substrate for DHDPS activity assay**

One of the substrates for DHDPS activity assay, (*S*)-aspartate- β -semialdehyde (ASA), is unstable and not commercialized. It was synthesised through three steps shown in Scheme 4.2.4.4 (Roberts SJ et al., 2003; Tudor DW et al., 1993). The

intermediates and final product were identified by $^{1}$HNMR, $^{13}$CNMR (synthetic procedures and data were in appendices).




| | |
|---|---|
| Step A. | $Et_3N$, $BOP{\cdot}PF_6$, $CH_3ONHCH_3{\cdot}HCl$, $CH_2Cl_2$ rt (Mw1=289) |
| Step B. | DIBAL; THF; -78°C (Mw2=332) |
| Step C. | $CF_3COOH$; $CH_2Cl_2$, $N_2$, rt (Mw3=273) |
| Compound 4 | $CF_3COOH{\cdot}$ L-ASA$\cdot$ $H_2O$ $C_6H_{10}F_3NO_6$ (Mw4=133+136) |
| Boc | tert-butoxycarbonyl; |
| $BOP{\cdot}PF_6$ | (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafuoro-phosphate |
| tBu | tert Butyl; |
| DIBAL | diisobutyl aluminium hydride |
| $Et_3N$ | triethylamine |
| THF | tetrahydrofuran |

**Scheme 4.2.4.4** Synthesis of L-aspartate β-semialdehyde (ASA)

**(c) Coupling enzyme assay of DHDPS activity** (Dobson RC et al., 2004)

DHDPS activity can be measured through the reduction of its catalytical product dihydrodipicolinate by NADH and DHDPR, where DHDPR is used as a helper enzyme and does not interfere with the DHDPS activity. The consumption of NADH is monitored spectrophotometrically at 340 nm. The initial velocity data were collected at a constant temperature of 20°C on a two-ray spectrophotometer Uvikon 930. All assays were performed in a final volume of 100 µl buffer of 100 mM Hepes (pH 7.0 and 8.0), 0.15 mM NADH, 5µg of DHDPR, 1mM pyruvate and varied concentrations of the substrates L-ASA. Kinetic parameters were calculated by fitting the initial velocity data at different substrate concentrations to the Michaelis–Menten equation by the Lineweaver-Burke plot (Fig. 4.2.4.4a and b and Fig. 4.2.4.5a and b). The experimental result and comparisons to literature data were shown in Table 4.2.4.1.

**Table 4.2.4.1** Results of DHDPS activity assay and comparison to literature data

| | $k_{cat}$ per subunit, $s^{-1}$ | $K_{L\text{-}ASA}$, mM |
|---|---|---|
| DapA pH7.0 | 9.7±0.5 | 0.074±0.005 |
| DapA pH8.0 | 24±1.4 | 0.162±0.015 |
| Ref1.pH7.0 | 8.5± 0.5 | 0.075±0.006 |
| Ref1.pH8.0 | 51±3 | 0.172±0.012 |
| Ref 2 pH8.0 | - | 0.11 ± 0.01 |

1. All assays were performed in 100µL solution containing 100mM Hepes buffer (pH7.0 and pH8.0), 0.15mM NADH, 5µg DapB, at 20 $^{o}$C
2. $K_m$ $^{ASA}$ was measured at a fixed pyruvate concentration of 1mM.
3. $k_{cat}=V_{max}/E_{total}$
4. For deducing $K_m$ & $V_{max}$, the Lineweaver-Burke plot is used (Fig. 4.2.4.4b and Fig. 4.2.4.5b).
Ref1. Joerger AC et al., 2003
Ref2. Dobson RCJ et al., 2004

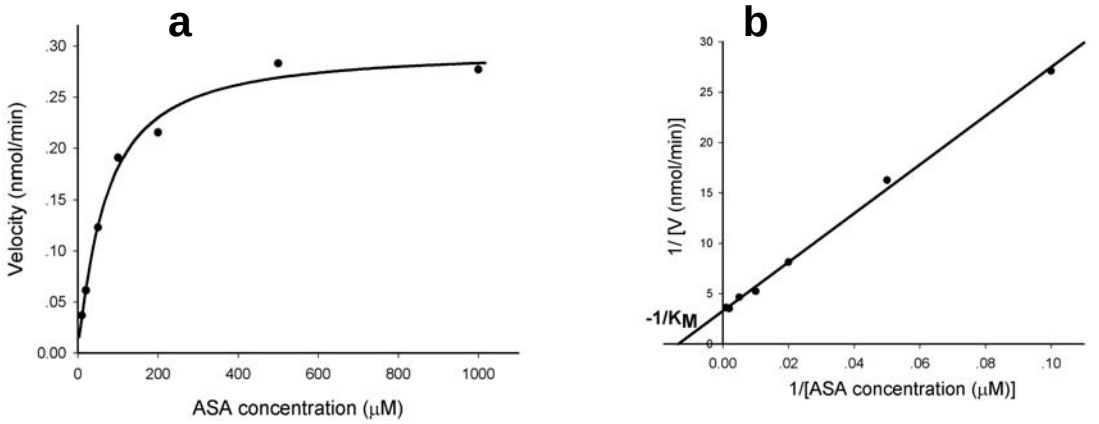



**Fig. 4.2.4.4** $K_M$ determination of DapA at pH 7.0.

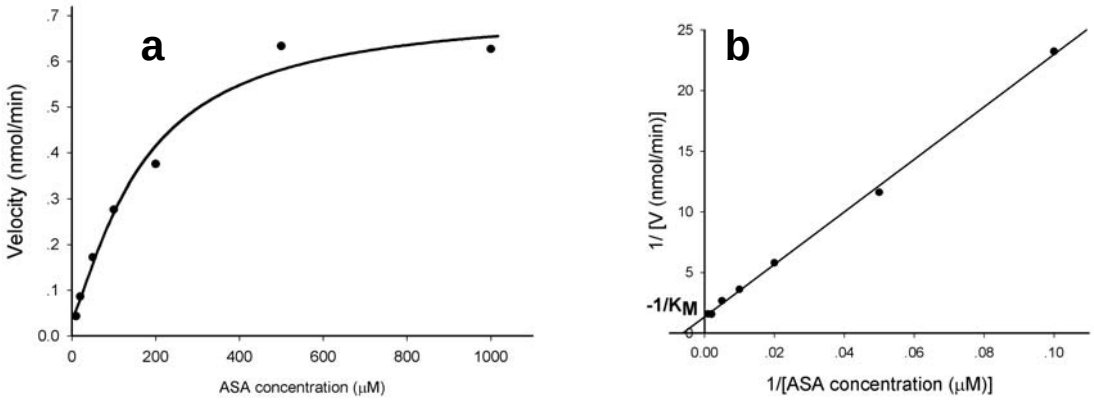



**Fig. 4.2.4.5** $K_M$ determination of DapA at pH 8.0.

#### 4.2.4.4 Summary and perspective of directed evolution of catalytic activity for Aldol reaction from tHisF library

To develop bio-catalysts for aldol reaction via biopanning of phage displayed tHisF library, a stable transition state analog (TSA) for aldol reaction was designed and synthesized.

In an alternative approach, the powerful selection method, genetic complementation, was used for directed evolving adolase activity from tHisF library. When an E coli *dap*A$^-$ mutation strain AT997 was employed for genetic selection, high frequent spontaneous reverse rate was observed. To solve this problem, a Δ*dap*A strain DH5α-JM1 lacking DHDPS gene in its chromosome was constructed in our lab. However, another false positive result became prominent which was due to the integration of chromosomal DNA and expression vector.

Biochemical assay for DHDPS activity had been established, including the synthesis of an unstable substrate in its salt form.

Perspective:

To tackle the false positive problem, chromosome DNA should be excluded thoroughly from gene library so to reach the final goal via genetic selection.

# 5 Summary

**Part one:** Construction of controlled partially randomized t*his*F gene libraries

A hyperthermostable enzyme with $(\beta\alpha)_8$ barrel, tHisF, was used as the scaffold for enzyme engineering. Synthetic oligonucleotides were directly assembled into t*his*F genes via self-priming PCR, and the gene diversity was introduced at specific positions on the codon level within the synthetic oligonucleotides with pre-synthesized tri-nucleotides. The controlled partially randomized t*his*F gene libraries were cloned into phagemid vector pCANTAB 5E and expression vector pKK223-3. The gene library was purified efficiently by immuno-depletion and the diversity of the gene library was increased by DNA shuffling.

The t*his*F gene library has $5.58x10^6$ independent clones. Both of the average exchanged codons in each clone and the average exchange ratio in single randomized position are consistent with the library design. The phage displayed t*his*F library can be used in bio-panning in search for proteins with desired property. Plasmid t*his*F library in expression vector pKK223-3 can produce soluble tHisF variants in recipient cells for genetic complementation.

**Part two:** Directed evolution of novel properties from t*his*F library

To evolve novel properties from the t*his*F gene library, a three-stage approach was adopted.

The first stage was to select tHisF variants that can complement HisF function in *E coli* strain. An auxotrophic $HisF^-$ *E. coli* strain UTH860 was used for $HisF^-$ complementation assay. Plasmids DNA of 98 selected clones were successfully sequenced. The statistics of nucleotide mutations in $eHisF^-$ selected DNAs showed a significant different mutational spectrum from *Taq* pol I and indicated a complicate situation concerning the origin of base mutations in the synthetic t*his*F gene libraries.

The 32 observed amino acids mutations located at 29 residue positions, constituted 11.5% of the whole 253 amino acids of tHisF. Strand β5" was a hotspot of mutation. No mutation was found at the nine randomized position of t*his*F gene.

At the bottom of β-barrel of tHisF, residue Ala3 which was seen participating in the inter-subunit interaction in tHisF-tHisH complex was substituted by Asp, analysis of short distance interaction showed A3D substitution enhanced the inter-subunit interaction by increasing the number of inter-subunit H-bond.

Four conserved residues were found to be substituted. The exchanges of E46D and K99E indicated that a hypothetic salt bridge ('salt gate') does not necessarily exist.

A remarkable observation from eHisF$^-$ complementation was 91 out of 98 selected tHisF variants contain a common mutation F86L. F86L mutation was also seen in HisA$^-$ E coli strain selected tHisF variant HA03 but with a different codon coding Leu. A hypothesis is this change may relate to the conformational mobility of hyperthermophilic enzyme tHisF. The test for this property was outside the scope of this thesis but is underway in the laboratory.

In addition, complementation result showed the designed consecutive mutations at tHisF substrate binding site N103A and T104A didn't affect HisF activity. This result indicates that those residues involved in substrate binding through main chain contact have less influence on enzyme activity.

The second stage of the work is selecting for HisA activity from the tHisF library. Enzyme *N*'-[(5'-phosphoribosyl)formimino] -5-aminoimidazole-4- carboxamide ribonucleotide isomerase (HisA) and imidazole glycerol phosphate synthase (HisF) catalyze consecutive reactions in the biosynthesis of the amino acids histidine, both have $(\beta\alpha)_8$-barrel structure. An auxotrophic *E coli* strain Hfr G6 lacking the *hisA* gene on its chromosome was used for HisA$^-$ complementation assay. A tHisF variant (termed HA03) that can complement HisA activity of *E coli* strain was identified. DNA sequence of HA03 revealed that it carried ten amino acid exchanges. Complementing ability test of HA03 indicated that the introduced mutations in HA03 impaired its HisF activity. The biochemical assay of HisA activity of HA03 gene product is also underway.

In the third stage, the groundwork had been paved for a target reaction of Aldol condensation, a stable transition state analog was designed and synthesized for future use in bio-panning to select desired biocatalyst candidates; meanwhile, in attempt to evolve aldolase from tHisF library by genetic selection, dihydrodipicolinate synthase (DHDPS, DapA) was targeted. A first attempt at using genetic complementation to select from the library a tHisF derivative with DHDPS activity failed due to high reversion rate of the *dapA* mutant strain originally employed; hence, a Δ*dap*A *E. coli* strain lacking DHDPS gene in its chromosome was constructed for this project, the biochemical assay system for DHDPS activity was established including the synthesis of an unstable substrate.

To tackle the false positive problem, chromosome DNA should be excluded thoroughly from gene library so to reach the final goal via genetic selection.

# 6 References

**Alifano P, Fani R, Lio P, Lazcano A, Bazzicalupo M, Carlomagno MS, Bruni CB (1996)** Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol Rev*. 60(1):44-69

**Altschul SF, Madden T, Scheaffer A, Zhang J, Zhang Z, Miller W and Lipman D (1997)** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nuclear Acids Res.*, 27(17):3389 - 3402

**Amann E, Brosius J, Ptashne M (1983)** Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in Escherichia coli. *Gene.* 25(2-3): 167-78

**Amaro RE, Myers RS, Davisson VJ, Luthey-Schulten ZA (2005)** Structural Elements in IGP Synthase Exclude Water to Optimize Ammonia Transfer. *Biophys J.* 89(1):75-87

**Ansorge W, Sproat B, Stegemann J, Schwager C, Zenke M (1987)** Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.*, 15(11):4593-602

**Arrizubieta MJ and Polaina J (2000)** Increased Thermal Resistance and Modification of the Catalytic Properties of a b-Glucosidase by Random Mutagenesis and *in Vitro* Recombination *The Journal of Biological Chemistry*, 275 (37):28843–28848

**Atwell S & Wells J (1999)** Selection for improved subtiligases by phage display. *Proc. Natl. Acad. Sci. USA*, 96(17):9497-9502

**Babbitt and Gerlt (1997)** Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities. *J Biol Chem* 272(49):30591-4

**Baek H, Suk KH, Kim YH, Cha S. (2002)** An improved helper phage system for efficient isolation of specific antibody molecules in phage display. *Nucleic Acids Res.* 30(5):e18

**Banfield MJ, Lott JS, Arcus VL, McCarthy AA, and BakerEN (2001)** Structure of HisF, a histidine biosynthetic protein from *Pyrobaculum aerophilum Acta Cryst.* 57(Pt 11):1518-25

**Baker K, Bleczinski C, Lin H, Salazar-Jimenez G, Sengupta D, Krane S, Cornish VW (2002)** Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A,* 99(26):16537-42

**Becker S , Schmoldt HU , Adams TM, Wilhelm S and Kolmar H (2004)** Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts
*Current Opinion in Biotechnoloy,* 15(4):323–329

**Beismann-Driemeyer S, Sterner R (2001)** Imidazole glycerol phosphate synthase from Thermotoga maritima. Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex. *J Biol Chem.*, 276(23):20387-96

**Bornscheuer UT, Altenbuchner J, Meyer HH (1999)** Directed evolution of an esterase: screening of enzyme libraries based on pH-indicators and a growth assay. *Bioorg. Med. Chem.* 7(10), 2169–2173

**Brakmann S, Grzeszik S (2001)** An error-prone T7 RNA polymerase mutant generated by directed evolution. *Chembiochemistry,* 2(3):212–219

**Bugg T (1997)** An Introduction to Enzyme and Coenzyme Chemistry, *Blackwell Science Ltd*.

**Bukhari AI & Taylor AL (1971)** Genetic Analysis of Diaminopimelic Acid- and Lysine-Requiring Mutants of *Escherichia coli, J. B*ac*ter.* 105(3):844-854

**Cadwell and Joyce (1992)** Randomization of genes by PCR mutagenesis. *PCR Meth. Appl.* **2**:28

**Canada KA, Iwashita S, Shimb H, and Wood TK (2002)** Directed Evolution of Toluene ortho-Monooxygenase for Enhanced 1-Naphthol Synthesis and Chlorinated Ethene Degradation, *J of Bacteriology*, 184(2):344–349

**Carl, PL (1970)** Escherichia coli mutants with temperature-sensitive synthesis of DNA. *Molecular and General Genetics,* 109(2):107-22

**Chaudhuri BN, Lange SC, Myers RS, Davisson VJ, Smith JL (2003)** Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerolphosphate synthase: crystal structures of a ternary complex and the free enzyme. *Biochemistry.* 42(23):7003-12

**Cherry JR and Fidantsef AL (2003)** Directed evolution of industrial enzymes: an update *Current Opinion in Biotechnology*, 14(4):438–443

**Chothia C, Gough J, Vogel C, Teichmann SA (2003)** Evolution of the protein repertoire. *Science*, 300(5626):1701-3

**Cipolla L (2004)** Combinatorial libraries of biocatalysts: application and screening, *Comb Chem High Throughput Screen*, 7(2):101-14

**Clackson T & Wells J (1994)** In vitro selection from protein and peptide libraries. *Trends Biotechnol*., 12(5):173-184

**Cohen H, Tawfik D and Griffiths A (2004)** Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalisation. *Protein Engineering, Design and Selection,* 17(1):3–11

**Daugherty PS, Iverson BL, Georgiou G (2000)** Flow cytometric screening of cell-based libraries. *J Immunol Methods,* 243(1-2):211-227

**Dobson R C, Gerrard JA and Pearce FG (2004)** Dihydrodipicolinate synthase is not inhibited by its substrate, (S)-aspartate β–semialdehyde, *Biochem. J.,* 377(Pt3):757–762

**Douangamath A, Walker M, Beismann-Driemeyer S, Vega-Fernandez MC, Sterner R, Wilmanns M (2002)** Structural Evidence for Ammonia Tunneling across the ($\beta\alpha$)8 Barrel of the Imidazole Glycerol Phosphate Synthase Bienzyme Complex, *Structure,* 10(2):185-193

**Dower WJ, Miller JF, Ragsdale CW (1988)** High efficiency transformation of E. coli by high voltage electroporation. *Nucleic Acids Res.*, 16(13):6127-45

**Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G (2004)** Rational engineering of enzyme stability. *Journal of Biotechnology,* 113(1-3): 105–120

**Fani R, Lio P, Chiarelli I, Bazzicalupo M (1994)** The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the hisA and hisF genes. *J Mol Evol.* 38(5):489-95

**Fani R, Tamburini E, Mori E, Lazcano A, Lio P, Barberio C, Casalone E, Cavalieri D, Perito B, Polsinelli M (1997)** Paralogous histidine biosynthetic genes: evolutionary analysis of the Saccharomyces cerevisiae HIS6 and HIS7 genes. *Gene.* 197(1-2):9-17

**Farber GK, Petsko GA (1990)** The evolution of alpha/beta barrel enzymes. *Trends Biochem Sci,* 15(6):228-34

**Frost JW, Bender JL, Kadonaga JT, Knowles JR (1984)** Dehydroquinate synthase from Escherichia coli: purification, cloning, and construction of overproducers of the enzyme. *Biochemistry.* 23(19):4470-5

**Garcia-Viloca M, Gao JL, Karplus M, Truhlar DG (2004)** How Enzymes Work: Analysis by Modern Rate Theory and Computer Simulations, *Science*, 303(5655): 186-195

**Garrett JB, Kretz KA, O'Donoghue E, Kerovuo J, Kim W, Barton NR, Hazlewood GP, Short JM, Robertson DE, Gray KA (2004)** Enhancing the thermal tolerance and gastric performance of a microbial phytase for use as a phosphate-mobilizing monogastric-feed supplement. *Appl Environ Microbiol.* 70(5):3041-6

**Garrick-Silversmith L, Hartman PE (1970)** Histidine-requiring mutants of Escherichia coli K12. *Genetics.,* 66(2):231-44

**Gerlt JA and Babbitt PC (2001)** Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.,* 70:209–246

**Gerlt JA and Raushel FM (2003)** Evolution of function in $(\beta\alpha)_8$8-barrel enzymes. *Curr. Opin. Chem. Biol*., **7(**2**):** 252–264

**Glieder A, Farinas ET, Arnold FH (2002)** Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat Biotechnol.*, 20(11):1135-1139

**Goldschmidt EP, Cater MS, Matney TS, Butler MA and Greene A (1970)** Genetic analysis of the histidine operon in *Escherichia coli* K12. *Genetics,* 66(2):219-29

**Greenwood J, Willis AE and Perham RN (1991)** Multiple display of foreign peptides on a filamentous bacteriophage. *J. Mol. Biol.*, 220(4):821–827

**Griffiths AD and Tawfik DS (2003)** Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* 22(1):24–35

**Griffiths JS, Cheriyan M, Corbell JB, Pocivavsek L, Fierke CA, Toone EJ (2004)** A bacterial selection for the directed evolution of pyruvate aldolases, *Bioorg Med Chem.* 12(15):4067-74

**Gromiha MM, Thomas S, Santhosh C (2002)** Role of cation-pi interactions to the stability of thermophilic proteins. *Prep Biochem Biotechnol*. 32(4):355-62(Abstract).

**Hanahan D (1983)** Studies on transformation of Escherichia coli with plasmids. *J Mol Biol.*, 166(4):557-80

**Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004)** De novo proteins from designed combinatorial libraries, *Protein Sci*., 13(7):1711-23

**Henn-Sax M, Hocker B, Wilmanns M, Sterner R (2001)** Divergent evolution of (beta/alpha)8-barrel enzymes. *Biol Chem*. 382(9):1315-20

**Hill DF and Petersen GB (1982)** Nucleotide sequence of bacteriophage f1 DNA. *J. Virol.*, 44(1):32–46

**Hillier BJ, Rodriguez HM, Gregoret LM (1998)** Coupling protein stability and protein function in Escherichia coli CspA. *Fold Des*. 3(2):87-93

**Hoeckker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R (2001)** Dissection of a (betaalpha)8-barrel enzyme into two folded halves**.** *Nature Structural Biology*, 8(1): 32-36

**Hocker B, Claren J, Sterner R (2004)** Mimicking enzyme evolution by generating new (betaalpha)8-barrels from (betaalpha)4-half-barrels. *Proc Natl Acad Sci U S A.*, 101(47):16448-53

**Hocker B, Jurgens C, Wilmanns M, Sterner R (2001)** Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold. *Curr Opin Biotechnol*., 12(4):376-81

**Hofnung M, Schwartz M, and Hatfield D (1971)** Complementation studies in maltose-A region of the Escherichia coli K12 genetic map. *J. Mol. Biol.*, 61**:**681–694

**Holm L, Sander C (1998)** Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.*, 26(1):316-9

**Hoseki J, Yano T, Koyama Y, Kuramitsu S, Kagamiyama H (1999)** Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermusthermophilus*. *J. Biochem.* (Tokyo) 126(5):951–956

**Huang MM, Arnheim N and Goodman MF (1992)** Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR, *Nucleic Acids Research*, 20(17):4567-4573

**Jaeger KE, Eggert T (2002)** Lipases for biotechnology. *Curr Opin Biotechnol*, 13(4):390-397

**Janda KD, Lo CH, Li T, Barbas CF 3rd, Wirsching P, Lerner RA (1994)** Direct selection for a catalytic mechanism from combinatorial antibody libraries *Proc. Natl. Acad. Sci. USA*, 91(7):2532-2536

**Janda KD, Lo LC, Lo CH, Sim MM, Wang R, Wong CH, Lerner RA (1997)** Chemical selection for catalysis in combinatorial antibody libraries. *Science*, 275(5302): 945-948

**Jaenicke R (2000)** Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc Natl Acad Sci U S A.* 97(7):2962-4

**Jaenicke R (2000)** Stability and stabilization of globular proteins in solution. *J Biotechnol.* 79(3):193-203

**Jestin JL, Kaminski PA(2004)** Directed enzyme evolution and selections for catalysis based on product formation *Journal of Biotechnology*, 113(1-3):85–103

**Joerger AC, Mayer S & Fersht AR (2003)** Mimicking natural evolution in vitro: An N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity *Proc Natl Acad Sci U S A.*, 100(10):5694-9

**Gerlt JA, Babbitt PC (2001)** Barrels in pieces? *Nature Structural Biology* 8(1):5-7

**Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M, Sterner R (2000)** Directed evolution of a (beta alpha)$_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A*, 97(18):9925-30

**Kang AS, Barbas CF, Janda KD, Benkovic SJ & Lemer RA (1991)** Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc. Natl. Acad. Sci. U.S.A.* 88(10):4363-4366

**Kaper T (2001)** Engineering of Glycosidases From Hyperthermophilic Archaea http://library.wur.nl/wda/abstracts/ab3058.html

**Kansy E (2003)** Kombinatorische Synthese einer Genbibliothek und Analyse ihrer statistischen Struktur. Dissertation, Universität Göttingen

**Kay B, Winter J & McCafferty J (eds.) (1996)** *Phaqe Display of Poplides and*

*Proteins: A Laboratory Manual* (Academic Press, San Diego, CA)

**Kirk O, Borchert TV, Fuglsang CC (2002)** Industrial enzyme applications. *Curr. Opin. Biotechnol.* 13(4):345–351

**Klem TJ, Chen Y and Davisson VJ (2001)** Subunit interactions and glutamine utilization by Escherichia coli imidazole glycerol phosphate synthase. *J. Bacteriol.* 183(3):989–996

**Klem TJ and Davisson VJ (1993).** Imidazole glycerol phosphate synthase: the glutamine amidotransferase in histidine biosynthesis. *Biochemistry,* 32(19):5177–5186.

**Knowles JR (1991)** Enzyme catalysis: not different, just better. *Nature,* 350(6314): 121-4

**Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003)** Design of a novel globular protein fold with atomic-level accuracy. *Science,* 302(5649):1364-8

**Kyhse-Andersen J (1984)** Electroblotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J Biochem Biophys Methods*., 10(3-4):203-9.

**Laber B, Gomis-Ruth FX, Romao MJ, Huber R (1992)** Escherichia coli dihydrodipicolinate synthase. Identification of the active site and crystallization *Biochem J.,* 288(Pt2):691-5

**Laemmli UK (1970)** Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature,* 227(52950):680-685

**Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns (2000)** Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion, *Science* 289(5484):1546-50

**Lebbink JH, Kaper T, Bron P, van der Oost J, de Vos WM (2000)** Improving low-temperature catalysis in the hyperthermostable Pyrococcus furiosus beta-glucosidase CelB by directed evolution. *Biochemistry.* 39(13):3656-65

**Leopoldseder S, Claren J, Jurgens C, Sterner R (2004)** Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. *J Mol Biol*. 337(4):871-9

**Light J & Lerner RA (1995)** Random mutagenesis of staphylococcal nuclease and phage display selection. *Bioorg. Med. Chem*., 3(7): 955-967

**Lin H, Cornish VW (2002)** Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed. Engl*., 41(23):4402–4425

**Lipovsek D, Pluckthun A (2004)** In-vitro protein evolution by ribosome display and mRNA display *Journal of Immunological Methods*, 290(1-2):51-67 (abstract)

**Lonn A, Gardonyi M, van Zyl W, Hahn-Hagerdal B, Otero RC (2002)** Cold adaptation of xylose isomerase from Thermus thermophilus through random PCR mutagenesis. Gene cloning and protein characterization. *Eur J Biochem*., 269(1):157-63

**Matney TS, Goldschmidt EP, Erwin NS, Scroggs RA (1964)** A preliminary map of genomic sites for F-attachment in Escherichia coli K12. *Biochem Biophys Res Commun*.17(3):278-81

**McCarthy JK, Uzelac A, Davis DF, and Eveleigh DE (2004)** Improved Catalytic Efficiency and Active Site Modification of 1,4-β-D Glucan Glucohydrolase A from *Thermotoga neapolitana* by Directed Evolution. *J Biol Chem.*, 279(12): 11495-502

**Merz A, Yee MC, Szadkowski H, Pappenberger G, Crameri A, Stemmer WP, Yanofsky C, Kirschner K (2000)** Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry.* 39(5):880-9

**Miller BG, Hassell AM, Wolfenden R, Milburn MV, Short SA (2000)** Anatomy of a proficient enzyme: the structure of orotidine 5'-monophosphate decarboxylase in the presence and absence of a potential transition state analog. *Proc Natl Acad Sci* USA. 97(5):2011-6

**Muller BH, Lamoure C, Le Du MH, Cattolico L, Lajeunesse E, Lemaitre F, Pearson A, Ducancel F, Menez A, Boulain JC (2001)** Improving Escherichia coli alkaline phosphatase efficacy by additional mutations inside and outside the catalytic pocket. *Chembiochem.* 2(7-8):517-23

**Murzin (1996)** Structural classification of proteins: new superfamilies. *Curr Opin Struct Biol.* 6(3):386-94

**Myers RS, Jensen J, Deras I, Smith J, and Davisson VJ (2003)** Substrate-Induced Changes in the Ammonia Channel for Imidazole Glycerol Phosphate Synthase. *Biochemistry*. 42(23):7013–7022

**Nagano N, Hutchinson EG and Thornton JM (1999)** Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. Protein Sci. 8(10):2072–2084

**Nagano N, Orengo CA, Thornton JM. (2002)** One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol.* 321(5):741-65

**Ness JE, Kim S, Gottman A, Pak R, Krebber A, Borchert TV (2002)** Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol*. 20(12):1251–1255

**Ness JE, Welch M, Giver L, Bueno M, Cherry JR, Borchert TV (1999)** DNA shuffling of subgenomic sequences of subtilisin. *Nat. Biotechnol.* 17((9): 893–896

**Neuner P, Cortese R, Monaci P** (1998) Codon-based mutagenesis using dimer-phosphoramidites. *Nucleic Acids Res.* 26(5):1223-7

**Neylon C (2004)** Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution, *Nucleic Acids Research*, 32(4):1448-1459

**Nissim A, Hoogenboom HR, Tomlinson IM, Flynn G, Midgley C, Lane D, Winter G (1994)** Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J.*, 13(3):692-8

**Ostermeier M, Shim JH and Benkovic SJ(1999)** A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.,* 17(12):1205-1209.

**Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995)** How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.,* 4(11):2411-23

**Pedersen H, Holder S, Sutherlin DP, Schwitter U, King DS, Schultz PG (1998)** A method for directed evolution and functional cloning of enzymes. *Proc Natl Acad Sci U S A.,* 95(18):10523-8

**Ponsard I, Galleni M, Soumillion P, Fastrez J (2001)** Selection of metalloenzymes by catalytic activity using phage display and catalytic elution. *Chembiochem.* 2(4):253-9

**Raillard S, Krebber A, Chen Y, Ness JE, Bermudez E, Trinidad R (2001)** Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chem. Biol*. 8(9):891–898

**Reetz MT(2002)** Lipases as practical biocatalysts. *Curr. Opin. Chem. Biol.* 6(2):145–150

**Roberts SJ, Morris JC, Dobson RCJ and Gerrard JA (2003)** The preparation of ( S)-aspartate semi-aldehyde appropriate for use in biochemical studies. *Bioorg. Med. Chem. Lett.* **13**, 265–267

**Sambrook, J, Fritsch, F & Maniatis, T (1989)** Molecular Cloning: a Laboratory Manual. 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

**Sandgren M, Gualfetti PJ, Shaw A, Gross LS, Saldajeno M, Day AG, Jones TA, Mitchinson C (2003)** Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability. *Protein Sci.*12(4):848-60

**Sanger F, Nicklen S, Coulson AR (1977)** DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.,* 74(12):5463-7.

**Schiller S (2004)** Evolutive in vitro-Adaption eines thermostabilen (βα)8-barrel-Proteins an die Katalyse einer abiotischen Reaktion. Dissertation, Universität Göttingen

**Schindler T, Perl D, Graumann P, Sieber V, Marahiel MA, Schmid FX (1998)** Surface-exposed phenylalanines in the RNP1/RNP2 motif stabilize the cold-shock protein CspB from Bacillus subtilis. *Proteins.* 30(4):401-6

**Schmidt DM, Mundorff EC, Dojka M, Bermudez E, Ness JE, Govindarajan S, Babbitt PC, Minshull J, Gerlt JA (2003)** Evolutionary potential of $(\beta\alpha)_8$-barrels: functional promiscuity produced by single substitutions in the enolase superfamily Single Substitutions in the Enolase Superfamily, *Biochemistr,y* 42(28):8387-93

**Scott JK and Smith GP (1990)** Searching for peptide ligands with an epitope library. *Science,* 249(4967):386–390

**Sidhu SS, Weiss GA and Wells JA (2000)** High copy display on phage for functional selections. *J. Mol. Biol.,* 296(2):487–495

**Sharma RC and Schimke RT (1996)** Preparation of Electro-competent E. coli Using Salt-free Growth Medium, *Biotechniques* 20: 42-44.

**Shedlarski JG, Gilvarg C (1970)** The pyruvate-aspartic semialdehyde condensing enzyme of Escherichia coli, *J Biol Chem.* 245(6):1362-73

**Slatko BE (1994)** Thermal cycle dideoxy DNA sequencing. *Methods Mol Biol.;* 31:35-45

**Smith GP (1985)** Filamentous fusion phage: Novel expression vectors that display cloned antigens on the surface of the virion. *Science*, 228(4705):1315–1317

**Stemmer WP (1994)** DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A,* 91(22):10747-51.

**Stemmer WP (1994)** Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370(6488):389-91.

**Stemmer W, Crameri A, Ha K, Brennan T, Heyneker H (1995)** Single-step assembly of a gene and entire plasmid from large numbers of oligodesoxyribonucleotides, *Gene,* 164(1):49-53

**Sterner R (2000)** Hetreologous expression, purification and crystallization of enzymes from hyperthermophiles. Courses and Workshops in 2000 EMBL

**Sterner R, Schmid FX (2004)** De novo design of an enzyme, *Science,* 304(5679): 1916-7

**Sterner R, Dahm A, Darimont B, Ivens A, Liebl W and Kirschner K (1995)** $(\beta\alpha)_8$-barrel proteins of tryptophan biosynthesis in the hyperthermophile Thermotoga maritima. *The EMBO Journal*, 14(18):4395-4402

**Stevenson JD, Lutz S, and Benkovic SJ (2001)** Retracing Enzyme Evolution in the $(\beta\alpha)_8$ -Barrel Scaffold *Angew. Chem. Int. Ed.*, 40(10):1854-1856

**Suzuki M, Yoshida S, Adman ET, Blank A and Loeb LA (2000)** *Thermus aquaticus* DNA Polymerase I Mutants with Altered Fidelity, *J Biol. Chem.*, 275(42):32728–32735

**Tabor S, Richardson CC (1985)** A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, *Proc Natl Acad Sci U S A* 82(4):1074-8

**Tatusova TA, Madden TL (1999)** Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*., 174(2):247-250

**Tawfik DS and Griffiths AD (1998)** Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 16(7): 652–656

**Taylor SV, Kast P and Hilvert D (2001**) Investigating and engineering enzymes by genetic selection. Angew. *Chem. Int. Ed. Engl.,* 40(18):3310-3335

**Thoma R, Schwander M, Liebl W, Kirschner K, Sterner R (1998)** A histidine gene cluster of the hyperthermophile Thermotoga maritima: sequence analysis and evolutionary significance. *Extremophiles*. 2(4): 379-89

**Towbin H, Staehelin T and Gordon J (1979)** Electrophoretic transfer of proteins from acrylamide gels to nitrocellulosesheets: Procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.*, 76(9):4350-4354

**Tudor DW, Lewis T and Robbins D J (1993)** Synthesis of the trifluoroacetate salt of aspartic acid *β*-semialdehyde, an intermediate in the biosynthesis of L-lysine, L-threonine, and L-methionine. *Synthesis* **11**, 1061–1062

**Valetti F and Gilardi G (2004)** Directed evolution of enzymes for product chemistry *Nat. Prod. Rep.*, 21(4):490-511

**Van den Burg B, Dijkstra BW, Vriend G, Van der Vinne B, Venema G, Eijsink VG (1994)** Protein stabilization by hydrophobic interactions at the surface. *Eur J Biochem.* 220(3):981-5 (Abstract)

**Van den Burg B, Enequist HG, Van der Haar ME, Eijsink VGH, Stulp BK, Venema G (1991)** A highly thermostable neutral protease from *Bacillus caldoelyticus :* cloning and expression of the gene in *Bacillus subtilis* and characterization of the gene product. *J. Bacteriol.* 173(13):4107–4115

**Vega MC, Lorentzen E, Linden A, Wilmanns M (2003)** Evolutionary markers in the (beta/alpha)8-barrel fold. *Curr Opin Chem Biol*. 7(6): 694-701

**Vogel HJ and Bonner DM (1956)** Acetylornithinase of Escherichia coli: partial purification and some properties. *J. Biol. Chem*., 218(1**)**:97-106

**Wagner J, Lerner RA & Barbas CF 3rd (1995)** Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes, *Science*, 270(5243):1797-1800

**Waldo GS (2003**) Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.*, 7(1):33-38.

**Wang L, Zhang Z, Brock A, Schultz PG (2003)** Addition of the keto functional group to the genetic code of *E. coli.*, *Proc. Natl.Acad. Sci. U.S.A.*, 100(1):56–61

**Wang XG, Britton KL, Stillman TJ, Rice DW and Engel PC (2001)** Conversion of a glutamate dehydrogenase into methionine/norleucine dehydrogenase by site-directed mutagenesis *Eur. J. Biochem*. 268(22):5791–5799

**Wernic D, DiMaio J, Adam J (1989)** Enantiospecific synthesis of L-alpha-aminosuberic acid. Synthetic applications in preparation of atrial natriuretic factor analogs. *J Org Chem* 54(17):4224-4228

**Wierenga RK (2001)** The TIM-barrel fold: a versatile framework for efficient enzymes *FEBS (Federation of European Biochemical Societies) Letters*, 492 (3):193-198

**Williams GJ, Nelson AS and Berry A (2004)** Directed evolution of enzymes for biocatalysis and the life sciences *Cell. Mol. Life Sci.*, 61(24):3034–3046

**Wilson DR, Finlay BB (1998)** Phage display: applications, innovations, and issues in phage and host biology *Can. J. Microbiol*., 44(4):313-329

**Wise EL, Rayment I (2004)** Understanding the Importance of Protein Structure to Nature's Routes for Divergent Evolution in TIM Barrel Enzymes. *Acc Chem Res*.;37(3):149-58

**Xia G, Chen L, Sera T, Fa M, Schultz PG, Romesberg FE (2002)** Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc Natl Acad Sci U S A.*, 99(10):6597-602

**Yano JK, Poulos TL (2003)** New understandings of thermostable and peizostable enzymes. *Curr Opin Biotechnol.* 14(4):360-5

**Yeh P, Sicard AM & Sinskey AJ (1988)** General organization of the genes specifically involved in the diaminopimelate-lysine biosynthetic pathway of *Corynebacterium glutamicum Mol. Gen. Genet*. 212:105-111

**Zaccolo M, Williams DM, Brown DM, Gherardi E (1996)** An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues.
J Mol Biol. 255(4):589-603

**Zha D, Eipper A and Reetz MT (2003)** Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution. *Chembiochem*, 4(1):34–39

**Zhao H, Giver L, Shao Z, Affholter J and Arnold F (1998)** Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol*., 16(3):258-261

# 7 Appendices

## 7.1 Abbreviations

| | |
|---|---|
| °C | Degree(s) Celsius |
| A | Angstrom |
| aa | amino acid |
| AICAR | 5-Aminoimidazole-4-carboxamide ribotide |
| Amp | Ampicillin |
| AP | Alkaline phosphatase |
| App. | Appendix |
| APS | Ammonium peroxydisulfate |
| BCIP | 5-Bromo-4-chloro-3-indolyl phosphate p-toluidine salt |
| *Bla* | Gene of ß-Lactamase |
| Boc | tert-butoxycarbonyl |
| BOP·PF6 | (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafuoro-phosphate |
| Bp | base pair |
| BSA | Bovine serum albumin |
| C | Cytosine |
| $CH_3COONH_4$ | Ammonium acetate |
| Cm | Chloramphenicol |
| cm | centimetre |
| DIBAL | diisobutyl aluminium hydride |
| DMF | Dimethyl formamide |
| DMSO | Dimethylsulfoxide |
| DNA | Deoxyribose nucleic acid |
| DNase I | Deoxyribonuclease I |
| dNTP | Deoxy-nucleoside-triphosphate |
| dsDNA | double stranded DNA |
| DTT | Dithiothreitol |
| $\varepsilon_x$ | molar extinction co-efficiency at x nm |
| EDTA | Ethylenediaminetetraacetat, disodium salt |
| $Et_3N$ | triethylamine |
| f1 | Replication's origin of Phage f1 |
| Fig. | Figure |
| G | Guanine; Gibbs energy or free enthalpy |
| GCG | Genetics computer group |
| GDB | Genome database |
| Gp | gene product |
| gIIIp | Gene product III; surface protein of Phage M13 |
| His | Histidine |
| HisF | Imidazole glycerol phosphate synthase |
| *his*F | Gene of imidazole glycerol phosphate synthase |

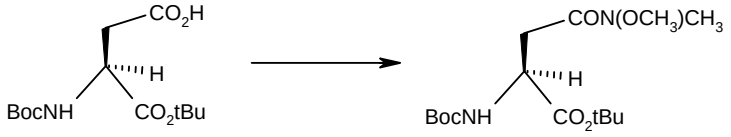
| | |
|---|---|
| *his*A | Gene of *N*'-[(5'-phosphoribosyl)formimino] -5-aminoimidazole-4- carboxamide ribonucleotide isomerase |
| IMAC | Immobilized metal ion chromatography |
| ImGP | Imidazole glycerol phosphate |
| IPTG | Isopropyl-beta-D-thiogalactopyranoside |
| kb | kilo base pair |
| kV | kilo volts |
| L | litre |
| LB | Luria and Bertani broth medium |
| LB-plate | LB agar Petri dish |
| M | Molar |
| mer | Oligomer |
| µg | microgram |
| mg | milligram |
| $MgCl_2$ | Magnesium chloride |
| min | minute(s) |
| µl | microlitre |
| ml | millilitre |
| µm | micrometer |
| NaAc | Sodium acetate |
| NaCl | Sodium chloride |
| NaOH | Sodium hydroxide |
| NBT | Nitro blue tetrazolium chloride |
| ng | nano gram |
| nmol | nano mole |
| $O.D._x$ | Optical density at x nm |
| ori | origin of replication |
| PAGE | Polyacrylamide-Gel electrophorus |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PEG | Polyethylene glycol |
| pfu | plaque forming units |
| 5'-PRFAR | N'-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide |
| 5'-ProFAR | N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide |
| RE | Restriction endonuclease |
| RNA | Ribonucleic acid |
| RnaseA | Ribonuclease A |
| rpm | Revolutions per minute |
| RT, rt | Room temperature |
| ssDNA | Single stranded DNA |

| | |
|---|---|
| SDS | Sodium dodecyl sulfate |
| sec | Second(s) |
| StEP | staggered extension process |
| T | Thymine |
| Ta | Annealing temperature |
| TAE | Tris-Acetate-EDTA buffer |
| Taq-p | *Thermus aquaticus* DNA polymerase |
| TBE | Tris-Borate- EDTA buffer |
| tBu | tert Butyl |
| TE | Tris-EDTA buffer |
| Te | Extension temperature |
| TEMED | N,N,N',N'-Tetramethylethylenediamin |
| TES | Tris-EDTA-NaCl buffer |
| THF | tetrahydrofuran |
| tHisF | Imidazoleglycerol phosphate synthase from *Thermotoga maritima* |
| t*his*F | Gene of imidazoleglycerol phosphate synthase from *Thermotoga maritima* |
| TIM-barrel | Triosephosphate isomerase barrel, $(\beta\alpha)_8$-barrel |
| Tm | Melting temperature |
| Tris | Tris(hydroxymethyl)aminomethane |
| TrpF | N-(5'-phosphoribosyl)anthranilate isomerase |
| u | unit |
| UV | Ultraviolet |
| V | Volt |
| vs. | versus |
| v/v | Volumen pro volumen |
| Vol. | Volume |
| w/v | Weight pro volume |

## 7.2 Synthesis of L-aspartic acid *β*-semialdehyde

The synthetic pathway is shown in Scheme 4.2.4.4.
All solvents were of commercially anhydrous grade. The reactions were carried out under a dry argon atmosphere and were monitored by thin-layer chromatography (TLC) using E.Merck 60F-254 precoated silica (0.25 mm) plates.

### 7.2.1 N-t-Boc-L-aspartic Acid 1-( *tert* -Butyl ester) N-Methoxy-N-methylamide *(Compound 2 )* (Wernic D et al., 1989)



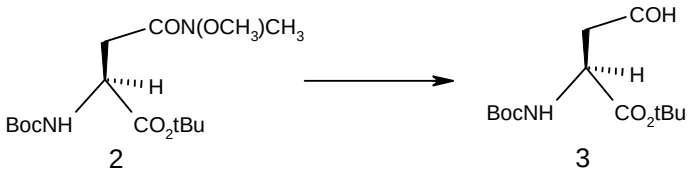

1 2

(Benzotriazol-1-yloxy)tris(dimethyl-amino)phosphonium hexafluoro- phosphate (BOP.PF6, 0.885 g, 1.9mmol) was added to a stirred solution of N-t-Boc-L-aspartic acid 1-(tert-butyl ester) (0.5g, 1.73mmol) and triethylamine (TEA, 0.27ml, 0.20g, 1.9 mmol) in dichloromethane ($CH_2Cl_2$, 17.5ml) at ambient temperature.

After 2min of stirring O,N-dimethylhydroxylamine hydrochloride (0.2 g, 2mmol) was added, followed by triethylamine (TEA, 0.27ml, 1.9mmol).

All solid materials were dissolved and the mixture were continued to stir for 2h at ambient temperature. The reaction mixture was then washed extensively with 1 M HCl (3 X 10 mL), $H_2O$ (3 x 10 mL), 1 M $NaHCO_3$, (2 X10 mL), and $H_2O$ (2 X 10 mL) and water phase was separated thoroughly. The organic phase was concentrated under reduced pressure, gave a pale yellow, oily product with colorless needle crystal (crude *Compound 2*, ~86%).
Dry the crude *Compound 2* with anhydrous $Na_2SO_4$ under reduced pressure, add in THF(dist., ~10ml) to dissolve the crude product, filtrate the solution to remove residue, vapor the THF, get pale yellow semi-solid. Repeat THF dissolve and filtrate procedure, dry the product in vacuum, gave purified *Compound 2* 0.3g (~44%).
$C_{l5}H_{28}N_2O_6$ Mw: 332.3 ESI-MS $MH^1$=333, $MNa^{23}$=355

### 7.2.2 *1-tert* -Butyl -2-(t-Boc-amino)-4-oxobutanoate *(Compound 3 )*




Under the protection of dry argon, a solution of diisobutylaluminum hydride (DIBAL) in hexane (1 M, 2.23 mL, 2.23 mmol) was added dropwise into a stirred solution of *Compound 2* (0,5g, 1.5 mmol) in anhydrous THF (14 mL) at -75 $^oC$. Upon addition the mixture continued to be stirred at -75 $^oC$ for 3 h. The reaction mixture was partitioned between 0.35 M $KHS0_4$ aqueous solution (15 mL) and ether (22.3 mL) and the aqueous layer was extracted with ether (3 X 7.5 mL). The combined ethereal solutions were washed with 1 M HCl (3 X 7.5 mL), 1 M NaHC0, (3 X 7.5 mL), and brine (3 X 7.5 mL) and dried with anhydrous $Na_2SO_4$. The dried organic phase was concentrated in vacuum gave the crude product *Compound 3* (78%) as a colorless oil, which solidified by standing at ambient temperature. Recrystallize *Compound 3* from hexane and analyze in with $^1$HNMR (Fig.A1), $^{13}$CNMR (Fig.A2.) and APT-NMR (Fig A3).

'H NMR ($CDC1_3$): 1.44 [s, 9 H, C(CH3)3], 1.45 [s,9 H, C(CH3)3], 2.98 (m, 2 H, CH2), 4.47 (m, 1 H, CH), 5.36 [d, 1 H, J(NHCH) = 6.65 Hz, NH], 9.74 (s, 1 H, CHO)

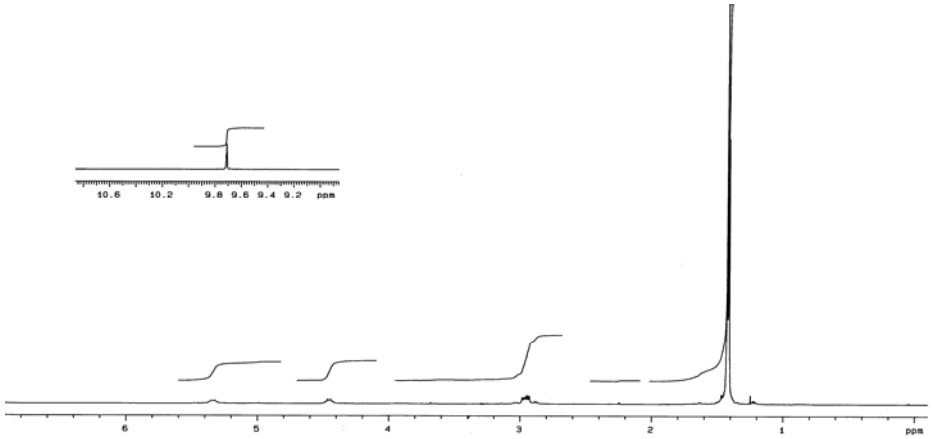


**Fig. A1** [1]HNMR of *1-t*-Butyl -2-(t-Boc-amino)-4-oxobutanoate *(Compound 3, in $CDCl_3$)*

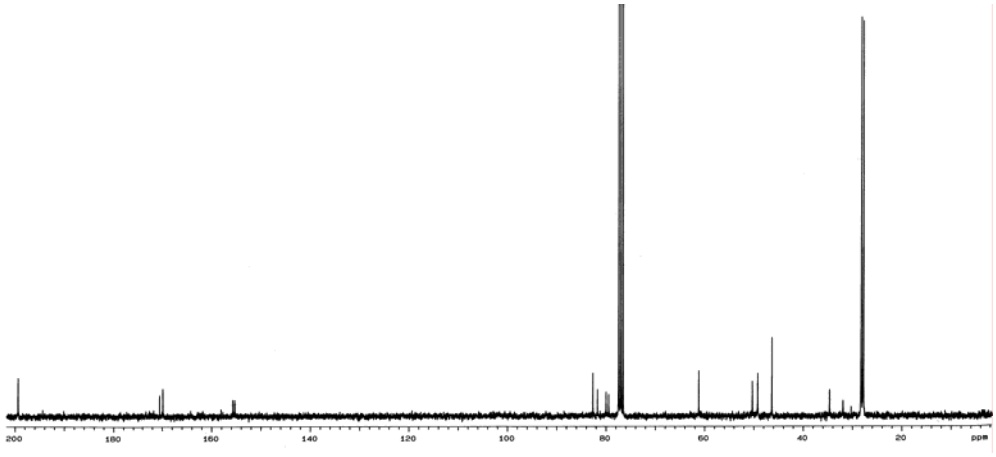


**Fig. A2** [13]CNMR of *1-t* -Butyl -2-(t-Boc-amino)-4-oxobutanoate *(Compound 3, in $CDCl_3$)*

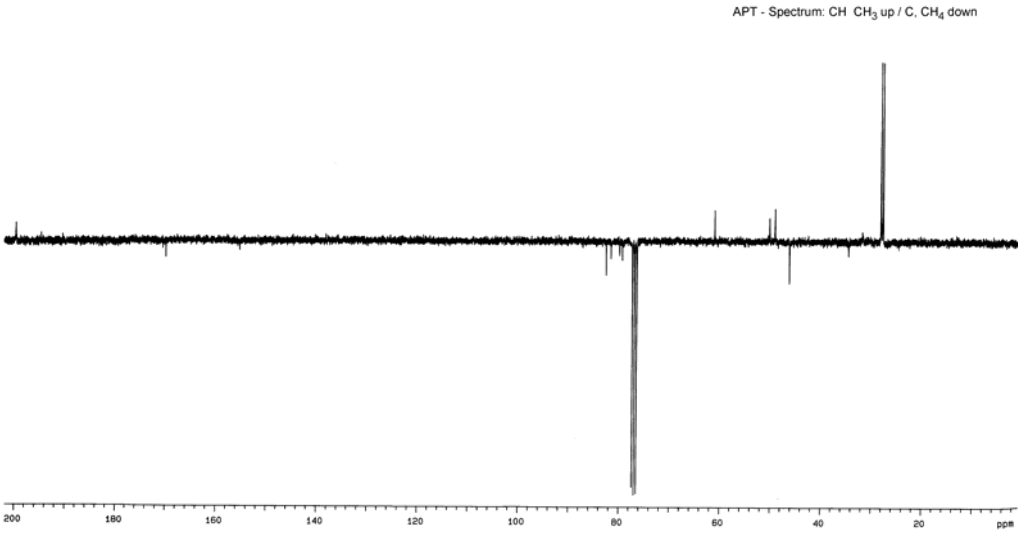
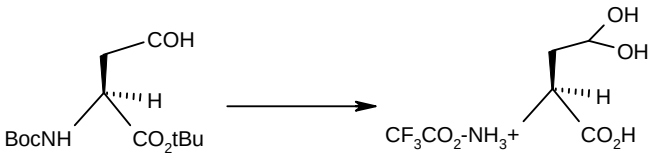



**Fig. A3** APT-NMR spectrumof *1-t*-Butyl -2-(t-Boc-amino)-4-oxobutanoate *(Compound 3, in $CDCl_3$))*

### 7.2.3 Trifluoroacetate salt of aspartic acid *β*-semialdehyde (*Compound 4*) (Tudor, D. W *et al., 1993 Synthesis*)




3 4

A solution of *Compound 3* (270mg, 1mmol) in trifluoroacetic acid (6mL) and $CH_2Cl_2$(6mL) was stirred under the protection of argon at rt for 2.5 h. The solvent was removed in vacuo to give an oily residue. This product was partitioned between water (30mL) and EtOAc (2X30mL). Removal of the water in vacuo gave a yellow solid, 150mg (60% yield). Its analysis is whoen in Fig.A4($^1$H NMR) and Fig.A5($^{13}$CNMR). $^1$H NMR (200 MHz, $D_2O$): 2,21(2H, m, 3-H2), 4,21(1H, dd,2-H), 5,38(1H, m, 4-H).

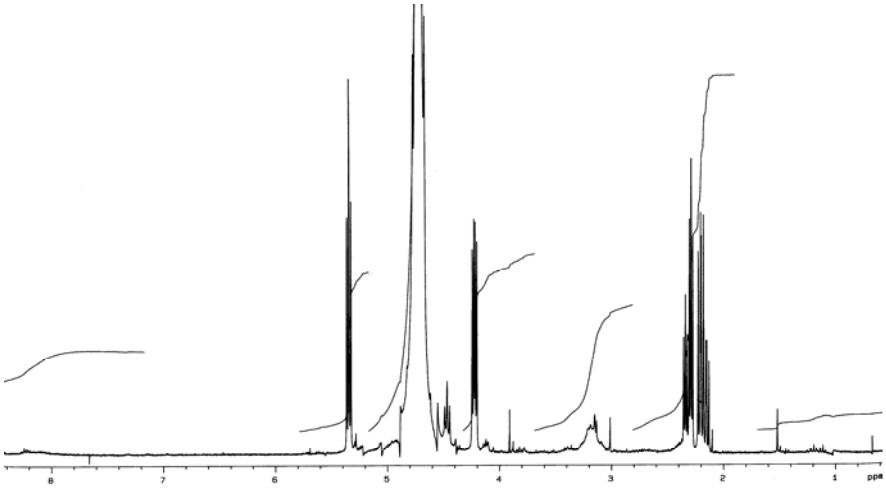


**Fig. A4** $^1$HNMR of trifluoroacetate salt of aspartic acid β-semialdehyde *(Compound 4,in $D_2O$)*

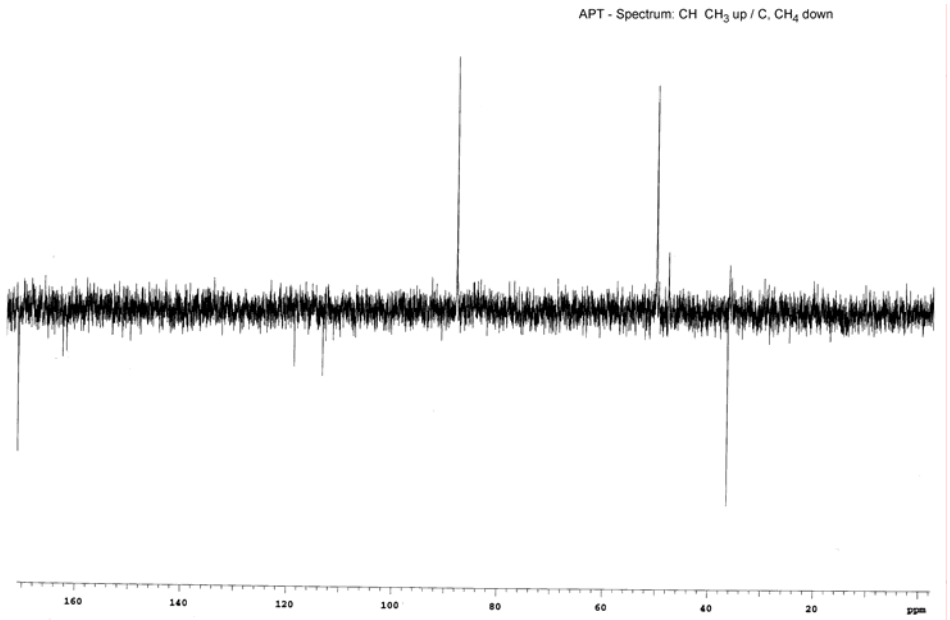



**Fig. A5** APT-NMR spectrum of trifluoroacetate salt of aspartic acid β-semialdehyde *(Compound 4, in* $D_2O$ *)*

## 7.3 CD-ROM with DNA sequences

# 8 Acknowledgement

I would like to express my sincere gratitude to Prof. Dr. Hans-Joachim Fritz, my supervisor, for accepting me as a PhD student in his research group, involving into the interesting field of enzyme engineering, for all his helps and supports, for his patience in all the procedure of my thesis work.

I am thankful to Dr. Andreas Schwienhorst for reading the manuscript and serving as co-referent of the thesis.

Many thanks to Dr. Susanne Behrens for kindly providing me vectors, sharing the working space and giving helpful advices, to Dr. Wilfried Kramer, Prof. Dr. Harald Kolmar for help in solving various problems.

My special thanks go to Prof. Dr. R. Sterner in the University of Rengensburg for kindly providing the E coli strains and helpful information for this work. I would like to thank Prof. Dr. U. Diederichsen in Institute for Organic and Biomolecular Chemistry, Uni. Goettingen for kindly providing me experiment space and material for chemical synthesis, thank his PhD students Nicole Hildebrandt and Angelina Weiß for their help in the experiments. Thanks to Dr M. Konrad in Max-Planck Institute for Biophysical Chemistry and his PhD student Sei Too for their help in enzyme activity assay.

Thanks to the staff members of the Genome Analysis Laboratory for their invaluable technique support.

I appreciate the tutorial work on molecular biology by Christine Härer, Stina Schille, Anke Schürer, Alexander Wentzel and Bettina Hucke. I am obligated to Christiane Preiß and Jessica Matter for their contribution to the project. Thanks to Torsten Adams, Andreas Christman, Gerrit Hennecke, Christian Hildmann, Eva Kansy, Lars Schomacher, Ulli Schmoldt, Julia Wittmann, Haiquan Xu and all the colleagues in the Department of Molecular Genetics and Preparative Molecular Biology for the friendly and scientific environment they have created, for giving hand whenever they were needed.

My appreciation goes to Olaf Waase and Patrick Regin for fixing of technical problems. Many thanks to Marita Kalck for her help in all the administrative procedure.

Thanks to all my friends for their numerous advices, helps and genuine friendship.

Finally, I would like to thank my family, my husband and my son, for their support and understanding. Without them I would never have got this far.

# Curriculum Vitae

## Zhenlian Ling

Birthday: 01 Oct. 1965
Place of birth: Jiangsu, PR China
Nationality: Chinese
Family status: Married

## Education

1972-1977 Lichengzhen Primary School, Jinhu, Jiangsu, China

1977-1983 Jinhu High School, Jiangsu, China

1983-1987 B.S. degree
Department of Chemistry, Wuhan University, China
Major: Organic chemistry

1987-1990 M.S. degree
Department of Chemistry, Wuhan University, China
Major: Synthetic organic chemistry
Supervisor: Prof. Chengtai Wu
Thesis: Mannich Reaction of Macrocyclic Imino-Compounds

2002- Ph.D student
Department of Molecular Genetics and Preparative Molecular Biology
School of Biology, University of Göttingen, Germany
Supervisor: Prof. Dr. Hans-Joachim Fritz
Thesis: Directed evolution of novel properties starting from HisF of Thermotoga maritima as a structural scaffold

## Working Experience

1990-1993 Teaching assistant
Nanjing University of Technology, Jiangsu, China

1993-2001 Lecturer
Nanjing University of Technology, Jiangsu, China