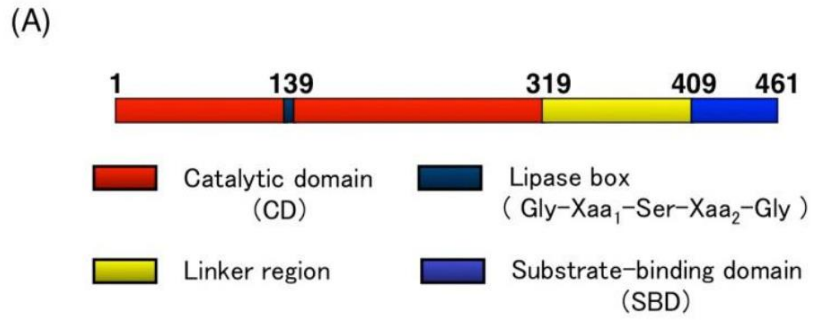
BIOCHEMICAL AND GENETIC PROPERTIES

BIOCHEMICAL AND GENETIC PROPERTIES OF PHB DEPOLYMERASES

A number of PHA depolymerases have been purified from diverse PHA-degrading microorganisms and characterized As described earlier, depending on the substrates and localization of PHA depolymerases, PHA depolymerases are grouped generally into four families: PHA depolymerases degrading the native intracellular granules (i-PHA_{MCL} depolymerases and i-PHA_{SCL} (i-PHB) depolymerases) and PHA depolymerases degrading the denatured extracellular PHA granules (e-PHA_{MCL} depolymerases and e-PHA_{SCL} (e-PHB) depolymerases). To date, the genes of about 30 PHA depolymerases with experimentally verified PHA depolymerase activity have been identified. On the basis of their sequence similarity, the PHA Depolymerase Engineering Database has been established as a tool for systematic analysis of PHA depolymerase family .

Among the PHA depolymerases, multi-domain e-PHB depolymerases have been extensively examined The multi-domain e-PHB depolymerases generally have a domain structure consisting of a catalytic domain (CD) at N-terminus, a substrate-binding domain (SBD) at C-terminus, and a linker region connecting the two domains, while e-PHB depolymerases from *Penicilliumfuniculosum*($PhaZ_{Pfu}$) and PhaZ7 from *Paucimonaslemoignei* ($PhaZ7_{Ple}$) have emerged as two exceptions (single-domain e-PHB depolymerases) Genetic analysis also shows that e-PHB depolymerases contain a lipase box pentapeptide [Gly-X_1-Ser-X_2-Gly] as an active residue, indicating that these enzymes are one of the serine hydrolases. As an example, the domain structure of e-PHB depolymerase from *Ralstoniapickettii* T1 ($PhaZ_{RpiT1}$) is illustrated in Figure 4(A). Such domain structure has been found in many biopolymer-degrading enzymes, such as cellulase, xylanase, and chitinase, which are capable of hydrolyzing water-insoluble polysaccharides The enzymatic degradation of PHB by the multi-domain e-PHB depolymerases is considered to proceed via a two-step reaction at the solid-liquid interface, as shown in Figure 4(B). The e-PHB depolymerase approaches and adheres to the PHB surface via SBD, followed by hydrolysis of the polymer chain by CD. Accordingly, it is considered that elucidation of the mechanisms of enzyme adsorption and enzymatic hydrolysis will contribute to the development of new PHB polymer materials with the desired environmental stability and biodegradability as well as the development of improved e-PHB depolymerases that can be used to effectively recycle PHB materials.



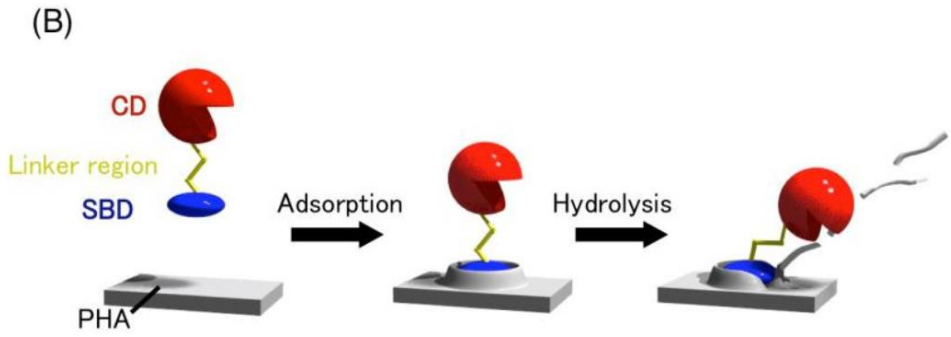


FIGURE 4.A) Domain structure of e-PHB depolymerase from *Ralstoniapickettii* T1 (Pha*ZRpiT1*). (B) Schematic illustration of the enzymatic degradation of PHA by e-PHB depolymerase.

From a biological viewpoint, the structure-function relationship of maluti-domain e-PHB depolymerases has been studied extensively, and several mutants were designed to analyze the function of each domain, in particular, SBD. Using a truncated multi-domain e-PHB depolymerases, Behrends et al., Nojiri and Saito, and our group revealed that the C-terminal domain is essential for PHB-specific binding Further, Nojiri and Saito genetically prepared many mutants of $PhaZ_{RpiT1}$ in various forms such as inversions, chimeras, and fusion to extra linker domains, and demonstrated that its SBD organization also influences the PHB degradation but not water-soluble substrates. Doi and co-workers prepared fusion proteins of SBDs of several PHB depolymerases with glutathione-S-transferase and demonstrated specific interactions based on molecular recognition between SBD and polyester surface.

EFFECTS OF CHEMICAL AND SOLID-STATE STRUCTURES AND SURFACE PROPERTIES OF PHAS ON ENZYMATIC DEGRADATION

Chemical structures of PHAs have influence on their enzymatic hydrolysis by multi-domain e-PHB depolymerases. Various types of PHAs including racemic PHA and 3HA oligomers PHAs with different main- and side-chain lengths (Kasuya et al., 1997) and random copolymers of (*R*)-3HB with various hydroxyalkanoate units have been synthesized to examine their enzymatic degradation by a variety of e-PHB depolymerases. For instance, Abe et al. proposed a schematic model of the

enzymatic cleavage of the PHA chain by $PhaZ_{RpiT1}$ (Figure 5), in which its active site can recognize at least three neighboring monomer units with a certain degree of difference in main-chain length Besides the chemical structure, the solid-state structure and surface properties of PHAs also influence the enzymatic hydrolysis. For example, the amorphous regions in PHA materials are preferentially hydrolyzed, followed by the hydrolysis of crystalline regions as a rate-limiting step in the enzymatic degradation process

Further, the enzymatic degradation rate of PHA materials decreases with increasing crystallinity, crystal size, and regularity of the chain packing state. In addition, Abe and co-workers demonstrated using proteinase K that the change in the surface properties of PLA film induced by end-capping with alkyl ester groups (carbon numbers 12 to 14) leads to a decrease in their enzymatic degradation rates

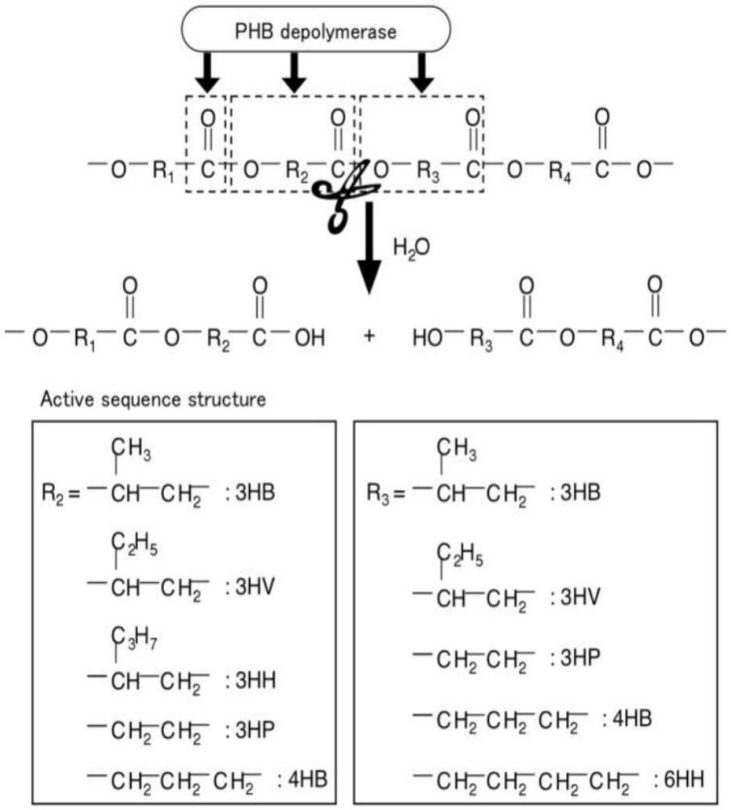


FIGURE 5.Schematic model of enzymatic cleavage of an ester bond in various sequences by PHB depolymerase.

To investigate the influence of the chemical structure or surface properties of polymer on enzymatic adsorption at nano-level sensitivity, several studies using quartz crystal microbalance (QCM) and atomic force microscopy (AFM) have been performed. Yamashita et al. investigated the

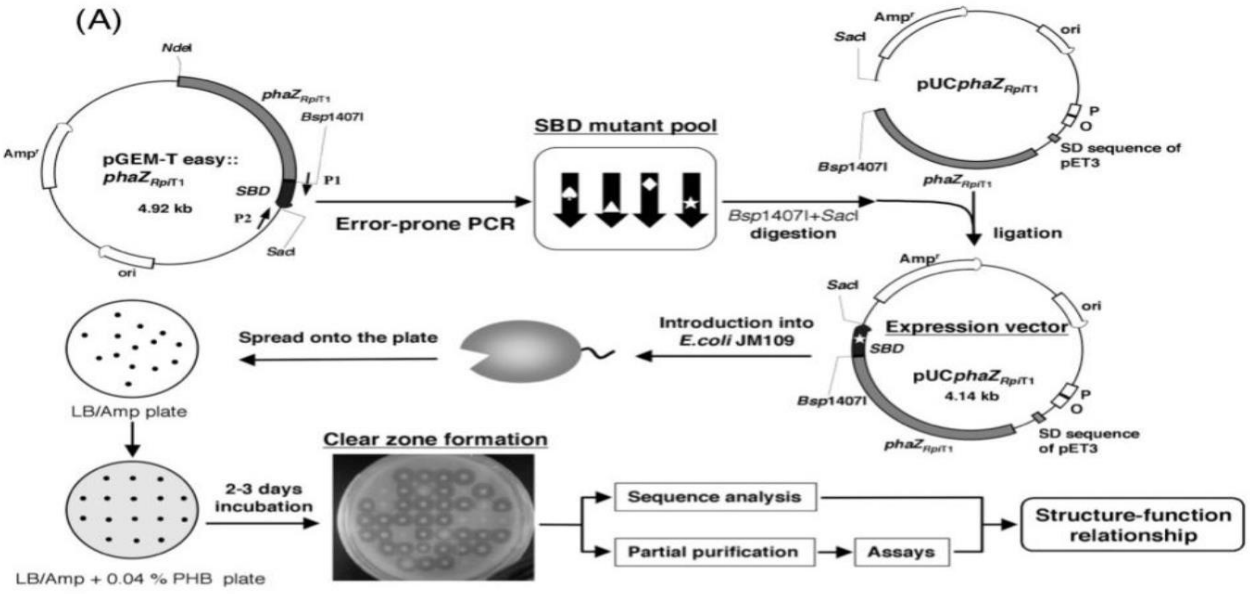
$PhaZ_{RpiT1}$adsorption to the film surface of several polymers including polyethylene, polystyrene and PHA using the QCM technique, and found that the enzyme showed adsorption specificity for PHA In addition, AFM analysis of $PhaZ_{RpiT1}$ on polyester surface has revealed that small ridges are formed around the enzyme molecule due to movement of some polyester chains at the adsorption area, suggesting that a strong chemical interaction exists between the enzyme and the polyester chains Furthermore, AFM analysis of interaction between PHB single crystal and a hydrolytic-activity-disrupted $PhaZ_{RpiT1}$ mutant has demonstrated that its SBD disturbs the molecular packing of PHB polymer chains, resulting in fragmentation of the PHB single crystal Taking these findings into consideration, the specific adsorption of PHB depolymerase to the PHB surface probably involves both the adsorption of the enzyme to the surface and the non-hydrolytic disruption of the substrate to promote PHB degradation. Recently, we have developed the AFM technique by using an AFM tip modified with SBD protein to evaluate the interaction between the SBD molecule and the PHB surface at the molecular level. Through this, it has been shown that the adsorption force of one SBD molecule to the PHB surface is approximately 100 pN

ANALYSIS OF POLYMER BINDING ABILITY OF E-PHB DEPOLYMERASE USING DIRECTED EVOLUTION TECHNIQUE

The structural aspects of an enzyme generally provide crucial information about the interaction between the enzyme and its ligand. Some researchers have reported the tertiary structures of polymer-degrading enzymes, such as glycoside hydrolases and single-domain e-PHB depolymerases, and proposed an interaction model between the enzymes and the polymer surfaces However, because of the paucity of information about the 3D structures of multi-domain e-PHB depolymerases, there are few insights into which and how amino acid residues in their SBD are involved in the enzyme adsorption to PHB surface.

Directed evolution is a useful and powerful tool to explore, manipulate, and optimize the properties of an enzyme as no information on the tertiary structure of the enzyme is required and new and unexpected beneficial mutations can be discovered Random mutagenesis via error-prone PCR (epPCR) and DNA recombination are widely used approaches to generate a large mutant pool and screen for the desired characteristics Using those approaches, many enzymes with improved substrate specificity, catalytic activity, thermostability, or solubility were obtained Further, analysis of the effects of mutations could also provide useful information for the improvement of enzyme function.

To improve e-PHB depolymerases, it is important to understand the mechanisms underlying its adsorption and hydrolysis, such as which and how amino acid residues participate in the catalytic process. To clarify this issue, we have investigated the interaction between $PhaZ_{RpiT1}$ and PHB surface by a combination of PCR random mutagenesis targeted to only SBD and an *in vivo* screening system as shown in Figure 6(A) In the analysis of recombinants showing low PHB-degrading activity, Ser410, Tyr412, Val415, Tyr428, Ser432, Leu441, Tyr443, Ser445, Ala448, Tyr455, and Val457 were replaced with other residues having hydropathy indices opposite to theirs at high frequency (Figure 6(B)). The results suggested that $PhaZ_{RpiT1}$ adsorbs to the PHB surface not only via the formation of hydrogen bonds between hydroxyl groups of Ser at these positions of the enzyme and carbonyl groups in the PHB polymer, but also via the hydrophobic interaction between hydrophobic residues at above-mentioned positions and methyl groups in the PHB polymer.



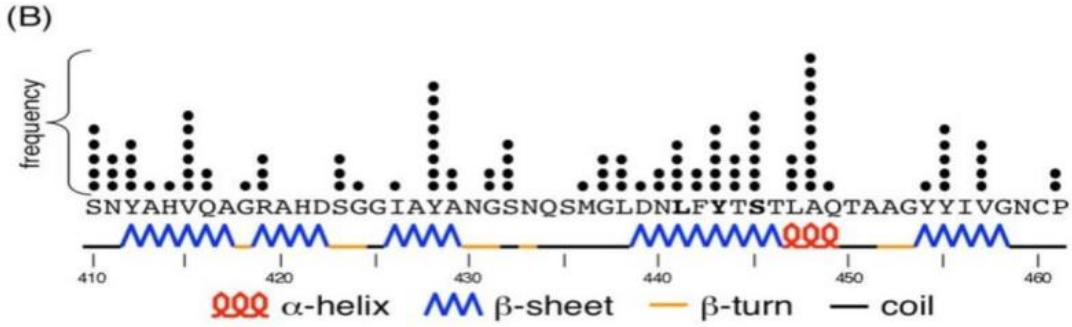


FIGURE 6.A) *In vivo* assay system for assessment of mutational effects of the substrate-binding domain of $PhaZ_{RpiT1}$ on PHB degradation. Schematic flow diagram of the system is illustrated. This system is composed of PCR-mediated random mutagenesis in the substrate-binding domain region of $PhaZ_{RpiT1}$ gene, preparation of mutant library, primary plate assay of PHB degradation (clear-zone formation), nucleotide sequencing and PHB degrading and adsorbing assays of partially purified mutant enzymes. (B) Positions and frequencies of PCR-mediated single mutations in the region coding for SBD of $PhaZ_{RpiT1}$, together with its predicted secondary structure.

Nevertheless, because only little knowledge was obtained on the biochemistry and kinetics of the purified mutant enzymes, the roles of these amino acids (Ser410, Tyr412, Val415, Tyr428, Ser432, Leu441, Tyr443, Ser445, Ala448, Tyr455, and Val457) and their contributions to the enzymatic activity remain poorly understood, resulting in little information to develop e-PHB depolymerases. Among these positions, Leu441, Tyr443, and Ser445 were predicted to form a β-sheet structure to orient in the same direction as shown in Figure 6(B). As polymer-degrading enzymes generally align their amino acid residues in a plane to interact with polymer surfaces, these three residues in $PhaZ_{RpiT1}$ may interact directly with the PHB surface. Since the hydropathy indices of such mutations as L441H (replacement of Leu441 with His), Y443H (replacement of Tyr443 with His), and S445C (replacement of Ser445 with Cys) dramatically changed among the mutations at these positions, their PHB-binding and -degrading properties were examined in detail . Functional analyses of the purified L441H, Y443H, and S445C enzymes indicated that these mutations had no influence on their structures and their ability to cleave the ester bond, while their PHB-degrading activity differed from that of the wild type. Kinetic analysis of PHB degradation by the mutants suggested that the hydrophobic residues at these positions are important for the enzyme adsorption to the PHB surface, and may more effectively disrupt the PHB surface to enhance the hydrolysis of PHB polymer chains than the wild-type enzyme. Further, surface plasmon resonance (SPR) analysis revealed that these substitutions mentioned above altered the association phase rather than the dissociation phase in the enzyme adsorption to the polymer surface.

Recently, Hisano et al. determined the crystal structure of $PhaZ_{Pfu}$ and proposed that hydrophobic residues, including Tyr, Leu, Ile, and Val, contribute to adsorption to the PHB surface, and that hydrophilic residues (Ser and Asn) located around the mouth of the enzyme crevice may also contribute to the affinity of the enzyme for PHB Jendrossek group determined $PhaZ7_{Ple}$ crystal structure and demonstrated that the enzyme was enriched in hydrophobic amino acids including eight tyrosine residues All tyrosine residues (Tyr103, Tyr105, Tyr172, Tyr173, Tyr189, Tyr190, Tyr203, and Tyr204), which are located at the surface of $PhaZ7_{Ple}$ but are far from the active site (Ser136), were changed to alanine or serine and the substitution effects were examined It turned out that mutation of Tyr105, Tyr189 or Tyr190 resulted in reduced PHB-degrading activity and in occurrence of a lag phase of the depolymerase reaction, indicating that these residues are possibly involved in the enzyme adsorption. Similar results have been obtained for the e-PHA_{MCL}depolymerase of *Pseudomonas fluorescens* GK13 by Jendrossek et al.

They reported that several hydrophobic amino acids (Leu15, Val20, Ile26, Phe50, Phe63, Tyr143 and Val198) were identified to be involved in interaction between the enzyme and poly(3-hydroxyoctanoate) substrate surface. This finding was supported with the recent study by Ihssen et al. (2009)

IMPROVEMENT IN SBD FUNCTION OF $PHAZ_{RPIT1}$

The above-mentioned findings imply that PHB binding ability of $PhaZ_{RpiT1}$ can be improved by substituting a hydrophilic residue with a hydrophobic one at the positions of 441, 443 and 445. Tyr at position 443 was targeted for substitution with a more highly hydrophobic amino acid residue because its hydrophobicity shows medium to high degree compared to those of general naturally occurring amino acid residues

Table 2 shows the hydrophobicity, the potential for β-sheet formation, and the volume of 20 common amino acid residues In this table, the properties of the original amino acid residue are colored blue and the desirable characteristics of the amino acid residues are colored orange, respectively. In the design of a mutant enzyme with an amino acid substitution at this position, the following factors were taken into consideration: (1) to achieve higher hydrophobicity than the original residue, (2) to retain the β-sheet structure, and (3) to change as little as possible the volume of the amino acid residue after the substitution. As a result, the substitution of Tyr443 with Phe (Y443F) was considered to be appropriate. Analysis of the purified Y443F enzyme indicated that the mutation had no influence on the structure and the

ester bond cleavage activity, while this mutant had higher PHB degradation activity than the wild type. Thus, this finding supports our previous assumption and indicates the importance of highly hydrophobic residues at these positions for PHB degradation.

amino acid residue	hydrophobicity[a]	Pb[b]	volume[c]
Ile	4.5	1.60h	100.1
Val	4.2	1.70h	83.9
Leu	3.8	1.30h	100.1
Phe	2.8	1.38h	113.9
Cys	2.5	1.19h	65.1
Met	1.9	1.05h	97.7
Ala	1.8	0.83i	53.2
Gly	-0.4	0.75b	36.1
Thr	-0.7	1.19h	69.7
Ser	-0.8	0.75b	53.4
Trp	-0.9	1.37h	136.7
Tyr(wild type)	-1.3	1.47h	116.2
Pro	-1.6	0.55b	73.6
His	-3.2	0.87i	91.9
Asn	-3.5	0.89i	70.6
Gln	-3.5	1.10h	86.3
Asp	-3.5	0.54b	66.7

Glu	-3.5	0.37b	83.0
Lys	-3.9	0.74b	101.1
Arg	-4.5	0.93i	104.1

TABLE 2. Hydrophobicity, potential for β-sheet formation, and volume of amino acid residues

CELL SURFACE DISPLAY SYSTEM FOR PROTEIN ENGINEERING OF $PHAZ_{RPIT1}$

Cell surface display is a valuable technique for the expression of peptides or proteins on the surface of bacteria and yeasts by fusion with the appropriate anchoring motifs Therefore, the cell surface display of functional and useful peptides and proteins, such as enzymes, receptors, and antigens, has become an increasingly used strategy in various applications, including whole-cell biocatalysts and bioabsorbents, live vaccine development, antibody production, and peptide library screening. In addition, this method is very useful for enzyme library screening because the displayed protein is accessible to the external environment and thus, is able to interact with substrates easily, allowing the screening of large libraries

A variety of surface anchoring motifs, including outer membrane proteins, lipoproteins, autotransporters, subunits of surface appendages, and S-layer proteins, have been employed to achieve the display systems We used the OprI anchoring motif for the functional display of $PhaZ_{RpiT1}$ on *Escherichia coli* cell surface The displayed enzyme retained its intrinsic characteristics, that is, hydrolytic activity for *p*-nitrophenyl butyrate (pNPC4) and the ability to adsorb to and degrade PHB, indicating that the engineered *E. coli* can be used in the form of a whole-cell biocatalyst by overcoming the uptake limitation of such substrates as insoluble PHB. These findings also indicate that the whole-cell catalyst is a promising and suitable tool to screen for mutant $PhaZ_{RpiT1}$ with enhanced catalytic activity.

PROTEIN ENGINEERING OF CD REGION OF $PHAZ_{RPIT1}$ USING CELL SURFACE DISPLAY SYSTEM

In contrast to SBD, there is little knowledge on the CD of $PhaZ_{RpiT1}$, such that which and how amino acid residues in the CD contribute to the enzymatic activity remain poorly understood, and this has resulted in the lack of information for the improvement of the CD function of $PhaZ_{RpiT1}$. The CD of $PhaZ_{RpiT1}$ was targeted for the directed evolution, employing random mutagenesis and DNA recombination to enhance its ester bond cleavage ability (Figure 7) The mutant genes generated from these reactions were expressed as surface-displayed enzymes, and the mutant enzymes were screened through a high-throughput system using pNPC4, a water-soluble substrate. As a result, clones displaying mutant enzymes with a 4- to 8-fold increase in pNPC4 hydrolysis activity were obtained in comparison with those displaying the wild type. This result was roughly consistent with the results of pNPC4 hydrolysis using purified enzymes with the unfused and undisplayed forms, concluding that the current screening system is feasible and effective for the search of improved enzymes.

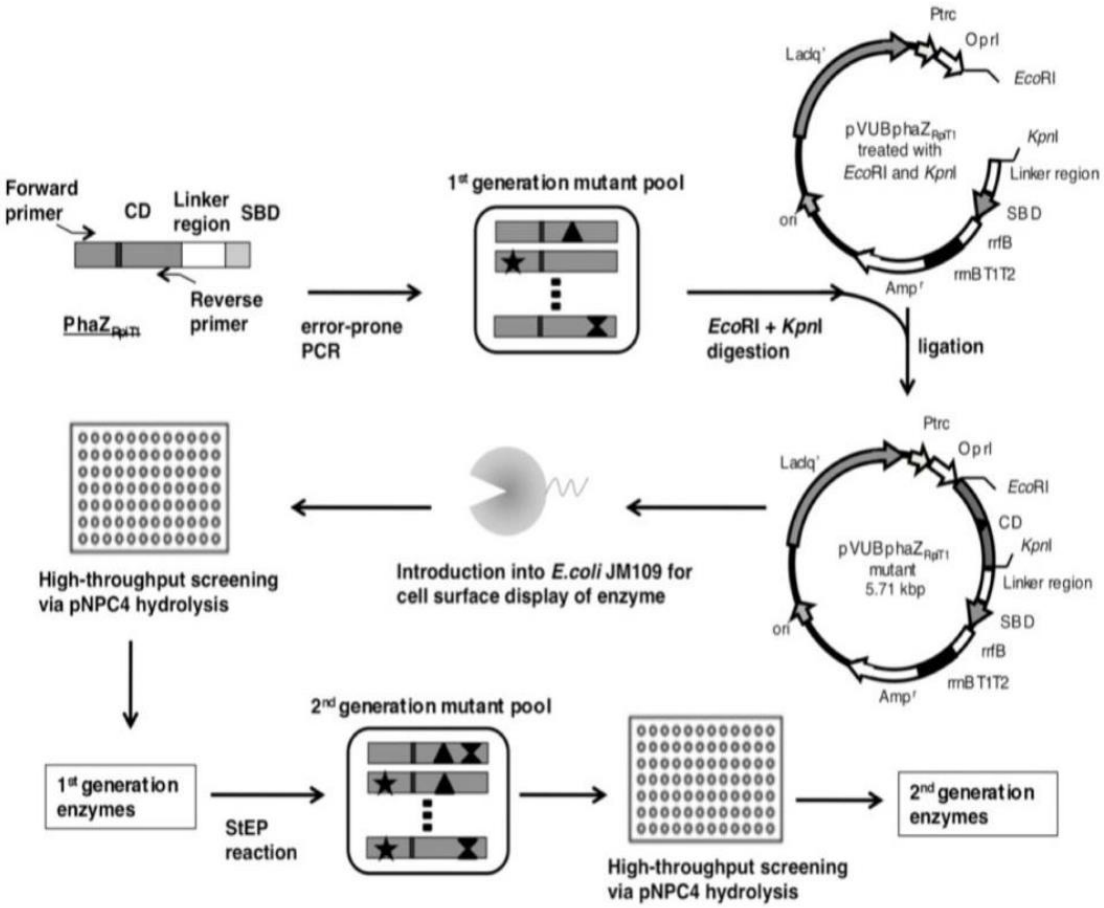


FIGURE 7.Directed evolution targeted at the catalytic domain (CD) of $PhaZ_{RpiT1}$ using the *in vivo* screening system in the cell surface display system. A schematic diagram of the mutational effects analysis is illustrated. This system consists of random mutagenesis by error-prone PCR in the CD of $PhaZR_{piT1}$, cell surface display of enzyme, high-throughput microplate screening via *p*-nitrophenyl butyrate (pNPC4) hydrolysis, staggered extension process (StEP), and nucleotide sequencing.

As the aliphatic part in pNPC4 is similar to the monomer unit in PHB polymer chain and pNPC4 is generally used as a model substrate, changes in pNPCn hydrolysis rates by the purified mutant enzymes as a function of the chain length of the aliphatic part in *p*-nitrophenyl esters (pNPCn, n=2 to 6) can provide the information regarding the substrate recognition of the enzyme. The results of pNPCn hydrolysis by the mutants demonstrated that the elevation on their pNPCn hydrolysis activity for each pNPCn substrate occurred. DNA sequencing showed that eight improved mutant enzymes contained N285D or N285Y mutations. As beneficial mutations are accumulated and deleterious mutations are simultaneously removed from the improved mutants through DNA recombination procedures the N285D and N285Y mutations found here are probably beneficial for pNPCn hydrolysis. Kinetic studies revealed that the increase in catalytic efficiency for pNPCn hydrolysis by the mutant enzymes is attributed to the high V_{max} values.

As opposed to pNPCn hydrolysis by the N285D and N285Y mutant enzymes, their PHB degradation rates were slower than that of the wild-type enzyme, indicating that these mutations are

unfavorable for PHB degradation. The kinetics of PHB degradation demonstrated that the N285D and N285Y mutations lowered the hydrolysis activity for the PHB polymer chain compared to the wild-type enzyme despite retention of the binding activity for the PHB polymer surface.

PROPOSED MODELS OF THE ACTIVE SITE IN E-PHB DEPOLYMERASES

The correct orientation of a PHB polymer chain to the active site is necessary to realize effective PHB degradation by e-PHB depolymerase. Hisano et al. have determined the crystal structures of $PhaZ_{Pfu}$-3HB trimer complex as well as $PhaZ_{Pfu}$ enzyme alone . In the $PhaZ_{Pfu}$-3HB trimer complex, 3HB trimer binds to the crevice with its carbonyl terminus towards the catalytic residues (Figure 8(A)). From the structural insight gained from $PhaZ_{Pfu}$, they proposed the mechanism of action of $PhaZ_{Pfu}$.Figure 8(B) shows the location of the catalytic residues and the interaction between PHB polymer chain and the residues in the subsite of the active site of $PhaZ_{Pfu}$. In their model, Ser39 participates in the nucleophilic attack of the carbonyl carbon of a PHB chain, resulting in the formation of a covalent acyl-enzyme intermediate followed by the hydrolysis by an activated water molecule. The nucleophilicity of the hydroxyl group of Ser39 is enhanced by the His155-Asp121 hydrogen bonding system.

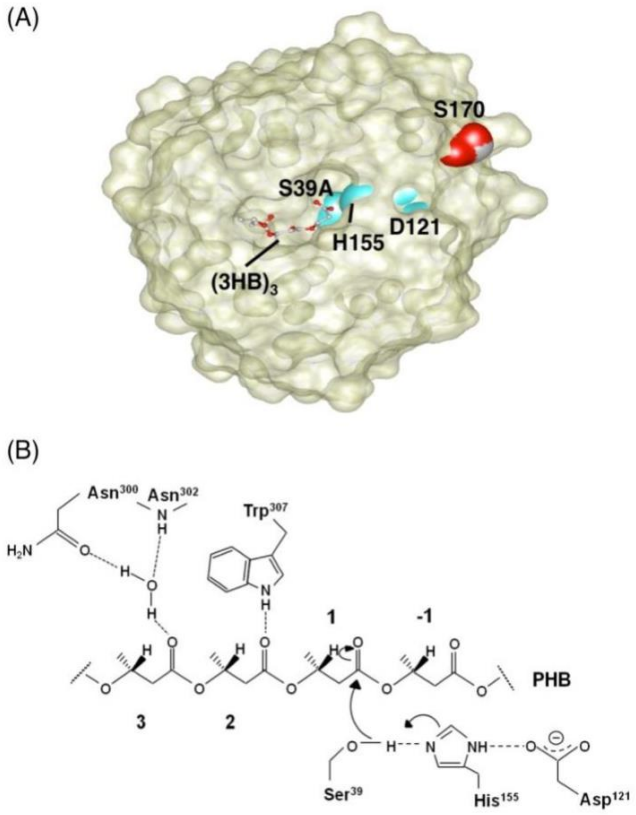


FIGURE 8.A) Molecular surface representation of $PhaZ_{Pfu}$. 3HB trimer in the crevice is shown as a ball and stick model. The positions of catalytic triad residues (S39A, D121, and H155) (cyan), as well as residue S170 (color-coded according to molecular species) are indicated. (B) Proposed model of the active site in $PhaZ_{Pfu}$ by Hisano et al. (Hisano et al. 2006).

For $PhaZ_{RpiT1}$, Bachmann and Seebach proposed that this enzyme has four subsites (2, 1, -1, and -2) in its active site, in which three of the subsites must be occupied by (*R*)-3-hydroxybutyrate (3HB) units for cleavage to occur at the center of the active site . Homology modeling of $PhaZ_{RpiT1}$ using the SWISS-MODEL program based on the crystal structure of $PhaZ_{Pfu}$ (PDB accession no. 2d81A) was performed to speculate the possible localization of Asn285 in the active site. Figure 9(A) shows the homology modeling structure of $PhaZ_{RpiT1}$, in which the modeled residue range was positioned from 124 to 294. The residue Asn285 (color-coded according to molecular species) of $PhaZ_{RpiT1}$ is located at the mouth of the crevice and also located immediately above His273, which corresponds to His155 in subsite -1 of $PhaZ_{Pfu}$. However, Asn285 was positioned as if to cover the subsite -1 and to inhibit the substrate access. Taking the homology modeling results and the aforementioned information on the cleavage mechanism into consideration, we propose a simple schematic model for $PhaZ_{RpiT1}$, as shown in Figure 9(B). In this model, Ser139 participates in the nucleophilic attack of the carbonyl carbon of a PHB chain, and its nucleophilicity is enhanced by the His273-Asp121 hydrogen bonding system. The residue Asn285 is positioned relatively close to His273 located in subsite -1 as if to cover the subsite. The location of Asn285 in the subsite probably leads to the regulation of the recognition of substrate molecules, such as pNPCn and PHB polymer chain, possibly indirectly via conformational change. A similar situation has been described in lipases and $PhaZ7_{Ple}$ where activation via conformational change is required to uncover the active site

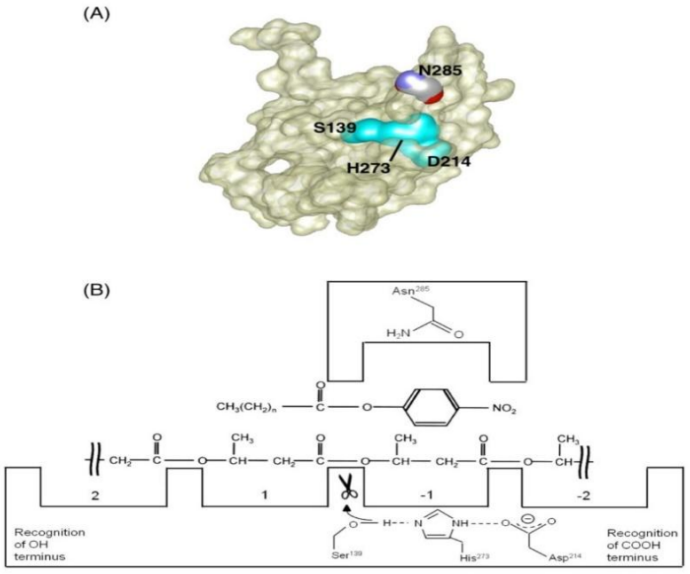


FIGURE 9.A) Molecular surface representation of the homology model of $PhaZ_{RpiT1}$. The positions of catalytic triad residues (S139, D214, and H273) (cyan), as well as residue N285 (color-coded according to molecular species) are indicated. (B) Newly proposed schematic model of the active site in the CD of $PhaZ_{RpiT1}$.

Bioengineering of Sequence - Repetitive Polypeptides

Protein - based materials, which correspond to polymers of tandemly repeated oligopeptide sequence motifs, have been the focus of significant research interest over the past two decades . The intellectual driving force for this process has come from two distinct directions: fi rst, from interest in the fundamental polymer science of architecturally uniform macromolecules; and second from interest in the structural biology of native, protein - based materials. From the viewpoint of fundamental polymer science, protein - based materials represent an approach to understand the effect of polymer architectural parameters (composition, sequence and molar mass) on macromolecular properties. Ribosomal protein synthesis ensures a uniformity of polymer microstructure that is impossible to achieve using the conventional synthetic methods employed for organic polymerization reactions. Thus, non - natural polypeptide sequences can been synthesized with near - absolute control of architectural parameters, and these biologically synthesized poly(α - amino acids) can be considered as model uniform polymers. These synthetic protein - based materials may provide insight into the fundamental aspects of polymer physical chemistry both in solution and in the solid - state, potentially enabling the creation of material constructs that display novel behavior . Furthermore, the observed control of polypeptide primary structure also implies the ability to defi ne a higher - order structure through the progression of protein structural hierarchy. Secondary and super - secondary elements, and the interactions between them, can be specified through the sequence identity although, as with more conventional targets of protein design, the currently limited ability of theoretical approaches to reliably defi ne the relationship between amino acid sequence and higher - order molecular and supramolecular structure is a significant constraint upon the design of novel polypeptide architectures. Nevertheless, genetic engineering methods have been employed to create artificial polypeptides of defined sequence that self - assemble into structurally definedsupramolecular aggregates, including lamellar crystallites surface - stabilizing coatings , smectic liquid crystalline mesophases , thermoresponsive nanoparticles and nanostructured hydrogels , on the basis of structural features programmed into the polypeptide sequences at the molecular level. These de novo - designed biomaterials provide an indication of the potential for biosynthesis to provide novel materials through the near - absolute control of macromolecular architecture.

The second factor that has motivated the investigation of protein - based materials lies in the desire to understand the chemical, biological and mechanical properties that underlie the native biological function of fi brousproteins . Natural evolutionary processes have afforded an array of structurally diverse protein - based materials that are produced within organisms as a natural consequence of their life cycle. These native protein - based materials usually display low complexity sequences that consist of tandem repeats of a fundamental oligopeptide motif that displays limited plasticity in amino acid sequence, and thus they bear

a nominal similarity to the repeat sequences of conventional organic polymers. The unique structural and functional properties of these native materials presumably arise as a consequence of their sequence specificity, which strongly influences the mode of self - assembly of the polymer chain into the supramolecular architectures that underlie their materials properties. Most notably, the materials properties of these native proteins often surpass the performance of synthetic materials within the relatively narrow compass of environmental conditions that defi ne these biological systems. Structural variants of these native proteins have been envisioned for technological applications as high - performance materials and, indeed, have provided the intellectual driving force for the development of conventional polymer science during the last century. Dragline silk fi bers from the spider Nephilaclavipes display a unique combination of high tensile and compressive strength that presumably originates in the segmented structure of the fi broin proteins that comprise the dragline fiber.

The biosynthesis of artificial protein - based polymers derived from sequence - repetitive polypeptides has developed in conjunction with the fundamental advances in recombinant DNA cloning and protein expression techniques over the past 25 years. Although this technology was not developed for the synthesis of protein - based materials, these techniques were soon applied to the synthesis of sequence - repetitive polypeptides based on the canonical repeats observed for native fi brous proteins such as elastin, collagen, keratin and silk. This approach met with mixed results in that, although significant knowledge was obtained with respect to cloning and expression of repetitive polypeptides, considerable challenges remained to be addressed, including the development of better methods to stabilize highly repetitive DNA sequences, to optimize recombinant protein yield, to promote appropriate post - translational modification, and to process the protein into a form that approximates that of the native state of the protein from which the sequence was originally derived.

Block Copolymers as Targets for Materials Design

Synthetic copolymers consisting of well - defined blocks of compositionally dissimilar monomers spontaneously self - assemble in the solid state into ordered domains of similar blocks. These hybrid materials have been extensively studied and often have unique, technologically significant properties in comparison to blends of the respective homopolymers. For example, copolymers comprising distinct blocks of different mechanical and chemical properties have been employed as polymer surfactants, pressure - sensitive adhesives, blend compatibilizers, thermoplastic elastomers, mineralization templates and lithographic resists. In contrast, block co - polypeptides have not garnered as much attention. However, the recent development of biosynthetic and chemosynthetic methods for the preparation of well - defined block copolymers of peptide sequences promises the potential for rapid advancements. These materials could be potentially interesting based on the diverse structures and functions observed for naturally occurring protein materials, in which the repetitive sequence pattern induces a regular secondary structure within the individual domains of the block co - polypeptide that has an important effect upon the supramolecular organization of the material. Segregation of the blocks into compositionally, structurally and spatially distinct domains occurs in analogy with synthetic block copolymers, affording ordered structures on the nanometer to micrometer size range. The sequence control and structural uniformity of these natural block copolymers is presumably responsible for their unique materials properties. The genetic engineering of synthetic

polypeptides enables the preparation of block copolymers composed of complex sequences in which the individual blocks may have different mechanical, chemical or biological properties. The utility of these protein materials depends on the ability to functionally emulate or enhance the materials properties of conventional polymer systems, while retaining the benefits of greater control over the sequence and microstructure that protein engineering affords for the construction of materials. This precise control of macromolecular architecture provides an opportunity for tailoring technologically significant materials properties for directed applications, for example, in biomedicine.

Amphiphilic Block Copolymers

One notable characteristic of technological significance for many conventional block copolymers is the property of amphiphilicity , that is, a difference in hydrophilic versus lipophilic (hydrophobic) character between the respective blocks. Amphiphilic block copolymers represent a special class of materials that are composed of compositionally defined blocks that have significantly different interaction affinities for aqueous solutions. These hybrid materials have attracted scientifi c interest due to their complex phase behavior in selective solvents, which parallels and complements that of small - molecule surfactant amphiphiles such as phospholipids. Amphiphilicdiblock (AB) and triblock (ABA or BAB) copolymers undergo selective segregation of the hydrophobic domain in aqueous solvents to form micellar structures in which the corona of the micelle is derived from the hydrophilic block (A) and the core of the micelle from the hydrophobic block (B). The identity and sequence of the individual block units within the polymer dictates the nature of the supramolecular assembly. In contrast to small - molecule surfactants, the phase behavior of amphiphilic block copolymers can be modified conveniently through manipulation of the macromolecular architecture – that is, the length, composition and sequence of the individual blocks. In addition, the hydrophilic/lipophilic balance can be adjusted systematically by variation of the relative lengths of the hydrophilic and hydrophobic blocks. These materials display several key features that may confer advantages over conventional surfactant amphiphiles in controlled delivery and release applications, namely, very low critical micelle concentrations, slow unimer – micelle exchange rates, high aggregate stabilities, and a controllable range of aggregate sizes and morphologies.

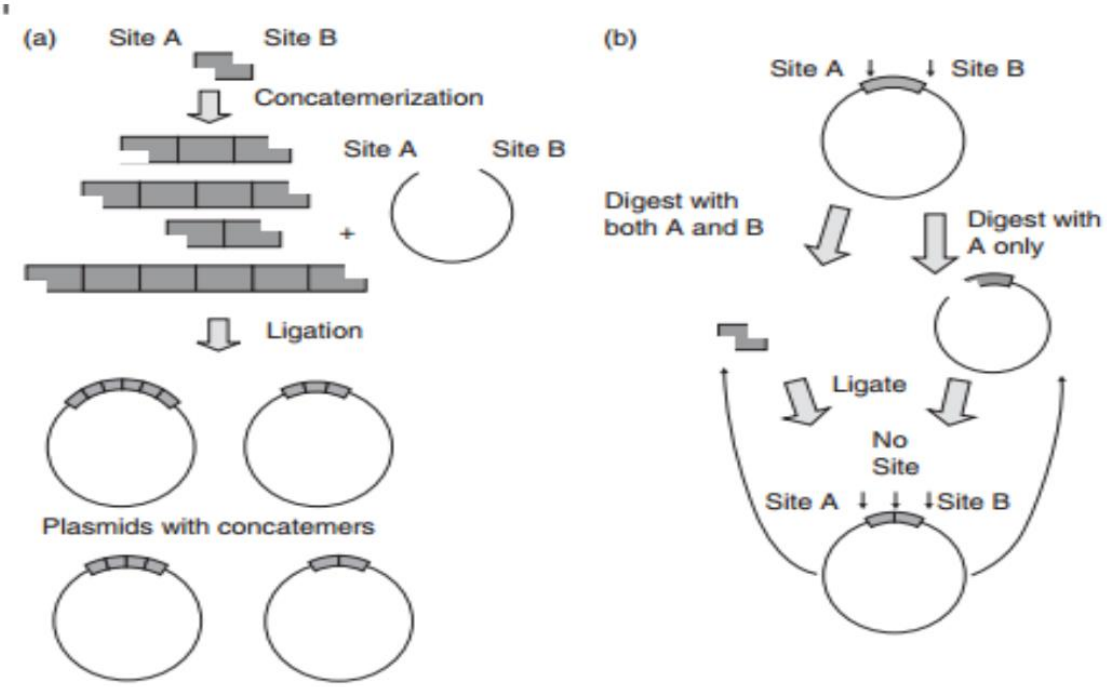
Elastin - Mimetic Block Copolymers

Elastin is a native protein - based material that is the primary structural component underlying the elastomeric mechanical response of compliant tissues in vertebrates and, therefore, has potential significance for human health as a medical biomaterial for the preparation of tissue - engineered analogues of native elastin - containing human systems. Moreover, elastin - mimetic polypeptides display a well - defined correlation between repeat sequence and macromolecular properties (vide infra), which enables the creation of a wide variety of synthetic elastin analogues with tailorable biophysical properties. The elastomeric domains of elastin comprise structurally similar oligopeptide motifs that are tandemly repeated in the native protein sequence. The local secondary structure and macromolecular thermodynamic and viscoelastic properties of the elastomeric domains can be emulated by synthetic polypeptides that are composed of a concatenated sequence of native oligopeptide motifs; the most common of which is the

pentapeptide (Val - Pro - Gly - Val - Gly). Polypeptides based on these pentameric repeat sequences undergo reversible, temperature - dependent, hydrophobic assembly from aqueous solution in analogy to the phase behavior of native tropoelastin, the soluble precursor of crosslinked elastin. This process results in a spontaneous phase separation of the polypeptide above a critical solution temperature, Tt , which is near ambient temperature in vitro. This inverse temperature transition coincides with a conformational rearrangement of the local secondary structure within the pentapeptide motifs.

Strategies for the Construction of Synthetic Genes Encoding Sequence - Repetitive Polypeptides

The synthesis of the protein - based materials based upon complex sequence repeats is best accomplished using the techniques of recombinant DNA (rDNA) technology and bacterial protein expression. The advantage of these methods lies in the ability to directly produce, with high fi delity, synthetic polypeptides of exact amino acid sequence and high molecular weight, as opposed to chemically synthesized oligopeptides, which are essentially limited to low degrees of polymerization (< 60 residues). With regards to the discussion herein, the term ‘ protein - based material ’ implies a sequence - repetitive polypeptide, or a multidomain protein consisting of one or more sequence - repetitive polypeptides, that is encoded within a synthetic DNA expression cassette. As materials properties generally scale to some degree with chain length, the biosynthesis of these protein polymers usually requires the construction and expression of large, synthetic genes containing multiple direct repeats of a ‘ monomeric ’ DNA sequence of approximately 50 to 150 base pairs in length. As automated DNA synthesis technology is currently limited to the production of oligodeoxynucleotides of lengths corresponding to about a hundred bases, sequences encoding medium to high - molecular - weight polypeptides cannot be obtained by direct synthesis of the entire gene. In addition, such repetitive DNA sequences may be unstable with respect to homologous recombination, and this may result in the structural instability of plasmid clones in vivo. Therefore, synthetic procedures for the cloning and expression of the repetitive genes may require special experimental considerations beyond conventional DNA manipulations (vide infra). Two main approaches have been described that are complementary in experimental methodology: (i**) DNA cassette concatemerization**and (ii) **recursive directional ligation.** Both strategies involve the chemical synthesis of the corresponding DNA sequence encoding the desired peptide repeat motif, enzymatically induced concatemerization, ligation of the concatemer into a plasmid vector, propagation in a bacterial host and, fi nally, expression of the repetitive polypeptides. However, the two strategies differ significantly in the method that is employed for generation of the concatemers and subsequent manipulation of the cloning vectors.



(i)DNA Cassette Concatemerization

This experimental protocol involves the construction of double - stranded oligonucleotide segments (DNA ‘ monomers ’) containing nonpalindromic, cohesive ends. Generation of the cohesive - ended DNA monomers is generally accomplished through the use of restriction endonucleases capable of recognizing and cleaving nonpalindromic sequences. The sizes of the oligopeptide repeats are usually chosen such that they could be conveniently encoded within single DNA cassettes of approximately 50 to 150 base pairs in length prior to concatemerization. Self - ligation of the DNA monomers proceeds in a head - to - tail fashion to generate a library of concatemers which differ in length by increments of the monomer. Preparative agarose gel electrophoresis is used to fractionate the concatemers according to the degree of concatemerization. Concatemers within the desired size range are extracted from the gel and used directly in subsequent cloning steps. A critical consideration for the successful application of this procedure is the efficacy of cloning and screening a population of concatemers to identify a construct of appropriate size. Unless concatemeric DNA cassettes corresponding to individual bands are excised from the gel, it is diffi cult to isolate and clone concatemers of determinate size using this approach. Usually, the sizes of individual concatemers are identified through screening a population of clones in parallel using either colony screening polymerase chain reaction(PCR) or restriction digestion of isolated plasmid - based constructs. Often, this process may require the screening of a large number of clones to identify a cassette of the desired size. DNA cassettes corresponding to very high degrees of concatemerization have been isolated using this procedure, although it is typically challenging to isolate a clone corresponding to a specifi c size. Although laborious, these protocols have been widely employed for the synthesis of artificial genes encoding sequence - repetitive polypeptides based on natural sequences, as well as artificial proteins having no natural parallel. However, difficulties have been reported in obtaining long concatemers as cloned inserts using this approach .Modifiedconcatemerization strategies have been described in which DNA adaptors have been appended to the termini to facilitate cloning into conventional plasmid - based vectors although these approaches do not necessarily address the problems associated with low yields of long concatemers.

Recursive Directional Ligation

In contrast to the DNA cassette concatemerization approach, recursive directional ligation permits the isolation of concatemers of determinate size through a controlled oligomerization process that is facilitated by the DNA manipulation experiments. Although several variations of the basic protocol have been described , the general procedure involves iterative directional insertions in a plasmid - based vector in which smaller concatemers are joined together recursively to form larger ones. The size of the DNA product that results from the cloning procedure corresponds to the sum of the initial DNA reactants. Thus, two DNA monomers can be joined together in a plasmid to form a dimeric construct. Two equivalents of the resulting dimeric construct can be joined to form a tetrameric construct, and so on. Repetitive application of this process, in which the products from a prior step are employed as the reactants in a successive step, can afford large concatemeric cassettes of determinate size. This procedure relies on the judicious choice of restriction sites at the termini of the DNA cassettes to facilitate the directional cloning process. As for DNA cassette concatemerization, restriction endonucleases that recognize and cleave nonpalindromic sites are very useful for the generation of cohesive - ended DNA fragments that are competent for selective ligation. Recursive directional ligation has the advantage that synthetic genes of determinate size and sequence can be obtained, although the process can be labor intensive for the assembly of large genes. For example, the assembly of a concatemer that encodes 32 (2 5) repeats of the basic sequence motif requires at least fi ve iterations of the directional cloning process in which the size of the concatemeric construct is doubled at the end of each step. Nevertheless, a significant number of sequence - repetitive polypeptides have been produced from synthetic genes assembled in this fashion. Recursive directional ligation is no doubt the technique of choice for the creation of synthetic genes of defined size and sequence. In theory, neither gene assembly strategy places any restriction on the size of the cloned DNA concatemers, although practical considerations (e.g. the efficiency of transformation of plasmid - based constructs and genetic instability of repetitive DNA sequences) may limit the effective size of cloned DNA inserts. It has been found that the ease of isolation of long DNA concatemers depends heavily on the identity of the DNA sequence. For many of the elastin - derived constructs, very large synthetic genes (≥ 8000 base pairs) can be obtained that encode sequence - repetitive polypeptides, whereas for other polypeptide sequences the isolation of long DNA concatemers becomes very diffi cult due to genetic instability leading to a recombinative loss of the majority of the coding sequence.

Application of protein folding to design new drug

Proteins are polymeric chains of amino acids that organisms and cells rely on for signaling, pathogen clearing, mobility, catalysis, recognition, shape, ordering, and stability. The precise ordering of the amino acids in a protein sequence determines how the protein folds into a 3D structure, and thus its biological function. As our knowledge of the connection between sequence, structure, and function has advanced, interest has grown in designing proteins on a sequence level to produce novel folds and function. Brute-force experimental approaches to resolving protein structures and designing protein sequences for new functions remain time consuming and expensive, and add little to our understanding of the physical principles required for both problems. Protein structure prediction aims to determine accurately

the full 3D structure of a protein given only its amino acid sequence. Structure prediction is challenging if only low homology templates exist. De novo protein design is the inverse problem given a rigid or flexible backbone structure, one aims to determine a sequence that will fold into that structure. Different sequences can fold into the same structure, so there is degeneracy in the protein design space. The existence and accuracy of protein structures as templates for protein design can have a significant impact on potential success. For this reason, the ability to produce viable protein templates through protein structure prediction is important for protein design, and for advancement in biotechnology and drug discovery. Figure 1 schematically shows the roadmap and key challenges in protein structure prediction and de novo protein design. The past few years have shown impressive applications of computational structure prediction and design to biotechnology, spanning peptide or antibody therapeutics, novel biocatalysts, and self-assembling nanomaterials.

State-of-the-art advances and challenges in protein structure prediction and refinement

The consistent determination of structure from sequence is one of the greatest unsolved problems in nature and has recently passed the 50-year milestoneAccurately predicting the 3D structure of a protein involves a series of steps performed on a sequence of amino acids: secondary structure prediction (identifying whether local segments are helical, beta-strand, or loop), structural alignment to candidate template structures, conformational sampling, and selection (Figure 2A and Box 1). A predicted structure may then undergo refinement, in an attempt to improve the accuracy of that structure. Historically, most refinement methods degrade rather than improve the accuracy of the predicted structure, making protein structure refinement a substantial unsolved problem in its own right.

The Critical Assessment of Techniques for Protein Structure Prediction (CASP) occurs biennially and recently completed its tenth experiment. For the CASP competition, prediction targets are categorized into two groups depending on the availability of structural templates: (i) template-based modeling, in cases for which templates are available; and (ii) free modeling, in cases for which templates are not available.

Successful protein designs with biotechnological applications

Design of proteins and peptides for therapeutic applications:

Over 200 peptides, proteins, or antibody therapeutics have been marketed as of 2010. Computational approaches have recently been applied to design new proteins and peptides for therapeutic applications. Elucidation of the sequences, structures, and interaction patterns of several disease-related proteins have allowed for the application of computational approaches for peptide therapeutic design .
Craik et al. predict that by 2020 we will see more prevalence of peptides as drugs, while outlining the challenges to meeting that outcome. Here, we review timely applications by target.

Cancer. Generally, therapeutic proteins/peptides can: (i) interfere with signal transduction cascades; (ii) arrest the cell cycle through modulation of cyclin-dependent kinase activity; or (iii) directly induce apoptosis by modulation of the proteins controlling apoptosis Cysteine-rich intestinal protein 1 (CRIP1) is an early biomarker for breast cancer. Hao et al. used phage display to identify peptide sequences that bound to

CRIP1. Subsequently, they computationally redesigned the scaffold sequence to optimize the binding free energy to increase its affinity for CRIP1, finding experimentally that it improved the IC50 27.5x over the phage-displayed sequence.

HIV. A computational method using side-chain grafting and to transplant a continuous structural epitope, 4E10, into scaffold proteins for conformational stabilization and immune presentation was developed. . The method produces epitope-containing designs that bind stronger to monoclonal antibody (mAb) 4E10 than 4E10 alone, and inhibits neutralization by HIV+ sera. Floudas and coworkers designed HIV-1 entry inhibitors starting from the structure of the C14linkmid peptide in complex with the hydrophobic core of gp41. C14linkmid is a crosslinked peptide derived from the C-terminal heptad repeat gp41. A global optimization-based sequence selection was performed with a distance-dependent force- field originally developed for protein folding to select candidate sequences from the vast combinatorial space. These sequences were reranked using fold-specificity calculations, which sample conformations near the template structure with substitutions dictated by the newly designed sequences. It aims to determine how favorably a new sequence folds into the fold of the design template. A subset of top-ranked sequences identified in the fold-specificity stage was evaluated using approximate bindingaffinity calculations, which approximate the binding equilibrium constant. The best design had an IC50 between 29 and 253 mM for different HIV-1 donors and mutants. This de novo design approach was made into an interactive web interface, Protein WISDOM.

Alzheimer's disease. Eisenberg performed computationally guided design to predict and experimentally validate peptide inhibitors of fibril formation by the t protein associated with Alzheimer's disease, as well as an amyloid promoting the sexual transmission of HIV. The designs bind to the end of the steric zipper and inhibit elongation. Focusing on the t protein inhibitor methodology, for a rotameric, fixed-backbone sequence optimization, they inverted the chirality of the design target to enable use of the Rosetta suite of tools. They designed L-amino acid sequences that favorably interact with a fixed-atom D version of the scaffold. Subsequently, the scaffold was reverted to its native L- form, and D-amino-acid-containing peptides were used as inhibitors experimentally. The designed D-peptides were then verified for shape complementarity, noting that D-Leu2 of the peptide was designed to clash with the target VQIVYK on the opposite sheet, and upon alanine substitution, inhibitory activity ceased. Introducing a tight-binding interface and clashes destroying the ability of a cascade of amyloid-forming sequences to propagate is effective for inhibition. Pande and coworkers, guided by observations made in simulations of Ab42, designed a noncanonical and D-amino-acid-containing peptide that organizes Ab42 into stable oligomers.

Self-assembling proteins/peptides

Controlling ordered (i.e., crystals) or disordered (i.e., hydrogels) self-assembly of proteins is a critical test of our understanding of both structure and interactions, having applications in biologically inspired materials. Lanci et al. computationally designed a protein crystal starting from an idealized homotrimeric parallel coiled-coil template and redesigned the interfaces. They utilized strictly physics-based energy functions to discriminate favorable interfaces. Stranges et al. took the solvent-exposed b strands of two monomeric proteins and redesigned them to form an intermolecular b sheet symmetric homodimer with near atomic-level accuracy. . This design demonstrated the creation of unique stabilizing interactions at an

interface. King et al. designed symmetric self-assembling complexes to atomic level accuracy. They performed symmetric docking of subunits followed by redesign at the interfaces to design cage-like nanomaterials with tetrahedral or octahedral point group symmetry. The designed structures were confirmed experimentally by crystallography and electron microscopy to high agreement. The control over such selfassembling can be used to design advanced functional materials and molecular machines.

QUESTIONS TO PRACTICE:

1. What are the applications of lipases/esterases.
2. Discuss the various strategies implemented to alter the subtsrate specificity and steroselectivity of lipases/esterases
3. What is co-factor engineering?
4. What is metabolic engineering?
5. Enumerate on the tools and processes involved in co-factor engineering for metabolic engineering
6. What is gene therapy? Discuss the role of cofactor engineering for treating disorders
7. Give a detailed account on strategies employed to manipulate polyketides and non ribosomal peptides.
8. Discuss in detail the different types of PHA synthase enzymes.
9. Write a detailed account on structure based engineering of e-PHB depolymerases
10. Write a detailed account on structure based engineering of PHA synthase enzymes
11. What are the Strategies for the Construction of Synthetic Genes Encoding Sequence
12. Enumerate on Successful protein designs with biotechnological applications

References

1. Visualization of Enzymatic Degradation of Poly[(R)-3-hydroxybutyrate] Single Crystals by an Extracellular PHB Depolymerase by Tadahisa Iwata, Yoshiharu Doi, Ken-ichi Kasuya, and Yoshio Inoue (1997)

2. Degradation of poly(3-hydroxyoctanoic acid) [P(3HO)] by bacteria: purification and properties of a P(3HO) depolymerase from Pseudomonas fluorescens GK13. by A Schirmer, D Jendrossek, H G Schlegel

3. Use of Extracellular Medium Chain Length Polyhydroxyalkanoate Depolymerase for Targeted Binding of Proteins to Artificial Poly[(3-hydroxyoctanoate)-co-(3-hydroxyhexanoate)] Granules by Julian Ihssen, David Magnani, Linda Thöny-Meyer, and Qun Ren

4. The Crystal Structure of Polyhydroxybutyrate Depolymerase from Penicillium funiculosum Provides Insights into the Recognition and Degradation of Biopolyesters paper (2006) by - Tamao Hisano, Ken-ichi Kasuya, Yoko Tezuka, Nariaki Ishii, Teruyuki Kobayashi, Mari Shiraki, Emin Oroudjev, Helen Hansma, Tadahisa Iwata, Yoshiharu Doi, Terumi Saito, Kunio Miki

5. Investigation of the Enzymatic Cleavage of Diastereomeric Oligo(3-hydroxybutanoates) Containing Two to Eight HB Units. A Model for the Stereoselectivity of PHB Depolymerase from Alcaligenes faecalis T1 by Beat M. Bachmann and Dieter Seebach

6. The Future of Peptide-based Drugs by David J. Craik, David P. Fairlie, Spiros Liras and

David Price

7. Two-Mechanism Model for the Interaction of Etoposide Quinone with Topoisomerase - Elizabeth G. Gibson, McKenzie M. King, Susan L. Mercer, and Joseph E. Deweese