약학박사학위논문

Structure determination of Rv0569 from *Mycobacterium tuberculosis* and Lead compound development of HsPDF from *Homo sapiens*

결핵균 유래 Rv0569 단백질의 구조 규명 및

사람 유래 단백질 HsPDF의 선도물질 개발

2014년 6월

서울대학교 대학원

제약학과 물리약학전공

김원제

약학박사학위논문

Structure determination of Rv0569 from *Mycobacterium tuberculosis* and Lead compound development of HsPDF from *Homo sapiens*

결핵균 유래 Rv0569 단백질의 구조 규명 및

사람 유래 단백질 HsPDF의 선도물질 개발

2014년 6월

서울대학교 대학원

제약학과 물리약학전공

김원제

약학박사학위논문

Structure determination of Rv0569 from *Mycobacterium tuberculosis* and Lead compound development of HsPDF from *Homo sapiens*

결핵균 유래 Rv0569 단백질의 구조 규명 및

사람 유래 단백질 HsPDF의 선도물질 개발

2014년 6월

서울대학교 대학원

제약학과 물리약학전공

김원제

Structure determination of Rv0569 from *Mycobacterium tuberculosis* and Lead compound development of HsPDF from *Homo sapiens*

결핵균 유래 Rv0569 단백질의 구조 규명 및
사람 유래 단백질 HsPDF의 선도물질 개발

지도교수 이 봉 진

이 논문을 약학박사학위논문으로 제출함.
2014년 6월

서울대학교 대학원
제약학과 물리약학전공
김 원 제

김원제의 박사학위논문을 인준함.
2014년 6월

위 원 장 _____ (인)

부 위 원 장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

ABSTRACT of the Dissertation

Structure determination of Rv0569 from *Mycobacterium tuberculosis* and lead compound development of HsPDF from *Homo sapiens*

Won-Je Kim

Dissertation director : Professor Bong-Jin Lee

Structure-based drug design (SBDD) is one of the rational drug development method and can be divided largely into two parts. First, the structure of the target protein is revealed by using Nuclear Magnetic Resonance (NMR) or X-ray crystallography. Second, lead compounds are generated through binding site research. The structure of Rv0569 protein from *Mycobacterium tuberculosis* was

calculated by NMR, achieving the early part of SBDD. After a serial process of cloning, massive expression, and purification of Rv0569, circular dichroism (CD) experiments was performed to predict the secondary structure and identify the basic physicochemical properties of the protein. The three-dimensional structure was calculated through NMR experiments. The bioinformatic research, structure calculation, and previous papers suggest that Rv0569 could be involved in hypoxia signal transduction. The latter part of SBDD was accomplished with peptide deformylase (HsPDF) derived from *Homo sapiens*. The lead compounds for HsPDF inhibitors drug were developed using library of bacterial peptide deformylase inhibitors. Forty-three species of the bacterial peptide deformylase inhibitors were selected by docking simulation. The species of top ten scores out of 43 were selected and their antitumor activities were tested. NMR binding experiments were conducted with selected four compounds showing the excellent antitumor activities to identify the HsPDF protein–ligand binding. Consequently, two species of bacteria-derived inhibitors were confirmed to bind human-derived protein, and were chosen as a lead.

……

Keywords : Structure–based drug design (SBDD), Tuberculosis, Rv0569, Nuclear Magnetic Resonance (NMR), hypoxic signal transduction, docking simulation, peptide deformylase inhibitors

Student Number : 2004-22292

Contents

Chapter 2. Lead compound development of HsPDF from *Homo sapiens*

List of Tables

List of Figures

Figure 10. In phase (blue peaks) and anti phase (red peaks) RDC spectrum of Rv0569.

Figure 11. RDC back-calculation using REDCAT. The value of Q factor is 0.056 and the value of RMSD is 0.394.

Figure 12. CD spectra in the various pH and salt conditions.

Figure 13. Solution NMR structure of *M. tb* Rv0569. The superimposition with ordered residues of the 20 representative structures is shown in red and flexible residues are shown in blue (A, B). The lowest clash scored structure from the Molprobity validation was represented with ribbon colored in orange (α-helix), magenta (sheet), and white (loop, terminus), respectively (C, D). There is very short α-helix except α-helix (α1) but is not found consistently through 20 structures of ensemble.

Figure 14. Sequence alignment of Rv0569 with Rv2302 (PDB code 2A7Y) from *M. tb*. Sequence alignment was performed with ClustalW and JalView and the conserved residues in both sequences are colored in boxes.

Figure 15. Structure alignment of Rv0569 with Rv2302 (PDB code 2A7Y) from *M. tb*. The Match maker module in UCSF Chimera was used to align structure with Needleman-Wunsch algorithm (A, B). Aligned proteins are shown with ribbon presentation method, colored in yellow (Rv2302) and green (Rv0569), respectively (A, B). And the mismatched regions in the aligned proteins are marked with blue dotted circles. The electrostatic surface potential is colored in blue (positively charged residues) and red (negative charged residues) with an energy scale of ±3 in

units of kcal/ (mole*e) for Rv0569 (C, D) and Rv2302 (E, F). All protein sequences and structures were obtained from the PDB database.

Figure 16. Hypothetical model of Rv0569 function. The electrostatic potential of protein (A) is shown as same scheme with that of Figure 15C and 15D. The residues on the largest positively charged area are represented as blue spheres including Trp8, Val10, Ile11, Lys12, Gly13, Ala14, Thr15, Ile16, Gln18, Pro19, Asp20, Arg22, and Val57. The color scheme of ribbon representation is same with that of Figure 13C and 13D.

Figure 17. Peptide deformylase in bacterial cell (A) and mammalian cell (B).

Figure 18. HsPDF crystal structure. This figure came from Journal of Molecular Biology 2009.

Figure 19. HsPDF crystal structure with actinonin. This figure came from Journal of Molecular Biology 2009.

Figure 20. Development of human PDF inhibitor from bacterial PDF inhibitor library.

Figure 21. The sequences of original and codon-optimized HsPDF gene (amino acid residues 40–243). The residues 40–62 of HsPDF (underlined) were not used for the expression of HsPDF protein, and the residues 63–243 of HsPDF (bold character) were expressed in the context of Trx-fusion protein. The rare codons of *E. coli* are marked with a gray box.

Figure 22. Molecular docking structures of HsPDF-PMT complexes. All PMT compounds in the binding pocket of HsPDF are presented with cyan line-stick

model including PMT-172 (A), PMT-173 (B), PMT-199 (C), and PMT-201 (D), respectively. Electrostatic surface potential of HsPDF was colored in blue (positively charged region) and red (negatively charged region) with 10 scaled energy unit. Hydrogen bonds of PMT-199 and PMT-201 were shown in magenta line.

Figure 23. The GC content of original (thin line) and codon-optimized (thick line) HsPDF genes. The GC content was calculated using the window of nine base pairs. The positions of rare codons in the original sequence were indicated by a vertical bar on the top of the figure.

Figure 24. Estimated stabilities of the overall secondary structure. The possible stem structures of the original HsDF (orHsPDF) gene (A) and the codon-optimized HsPDF (coHsPDF) gene (B) were estimated. The predicted stem regions of which energy is lower than -15 kcal/mol are indicated.

Figure 25. Expression and purification of HsPDF (63–243) with the N-terminal Trx-fusion protein. The expression of the Trx-tagged codon optimizing HsPDF (B) was greatly increased compared to that of Trx-tagged original HsPDF (A).

Figure 26. CPMG-T2 NMR binding assay. Antinonin (A), PMT-172 (B), PMT-173 (C), PMT-199 (D), and PMT-201 (E). The black spectrum was from 1D reference zgesgp pulse and the red spectrum was from the CPMG-T2 filter zgesgp pulse. The dashed blue circle means the differences of intensities with the CPMG-T2 filter experiment.

List of Abbreviations

AA Amino Acid

CD Circular Dichroism

CPMG Carr-Purcell-Meiboom-Gill pulse sequence

DNA DeoxyriboNucleic Acid

DSS 2,2-Dimethyl-2-Silapentane-5-Sulfonic acid

DTT DiThioThreitol

E. coli *Escherichia coli*

GC Guanine-Cytosine

GOLD Genetic Optimization for Ligand Docking

HsPDF Homo sapiens Peptide DeFormylase

HSQC Heteronuclear Single Quantum Coherence

IMAC Immobilized Metal ion Affinity Chromatography

IPTG IsoPropyl β-d-ThioGalactopyranoside

M. tb *Mycobacterium tuberculosis*

NMR Nuclear Magnetic Resonance

NOE Nuclear Overhauser Enhancement

OD Optical Density

PMT ProMediTec

RDC Residue Dipolar Coupling

RMSD Root Mean Square Deviation

SBDD Structure Based Drug Design

TB Tuberculosis

TCEP *Tris* (2-CarboxyEthyl) Phosphine

TEV Tobacco Etch Virus

WHO World Health Organization

General introduction

Structure-based drug design

Structure-based drug design (SBDD) is the technique of designing a new active substance based on three-dimentional structure of the target protein, the protein–ligand complex structure, the information of protein–ligand interaction and the computer modeling [1]. From search on lead compounds for a drug candidate to optimize compounds, SBDD uses the three-dimentional structure of protein in order to perform the drug development rationally (Figure 1). After obtaining the structure of the target protein, lead compounds are made from X-ray crystallographic screening method or structure-based virtual screening method. Because the massive compound libraries are simulated to dock efficiently to the target protein using computer calculation, SBDD reduces costs and time-consuming of the drug development to the comparing conventional high-throughput screening (HTS) method [2]. Major pharmaceutical companies have already adopted this method, and developed more than 10 kinds of new drugs (Table 1).

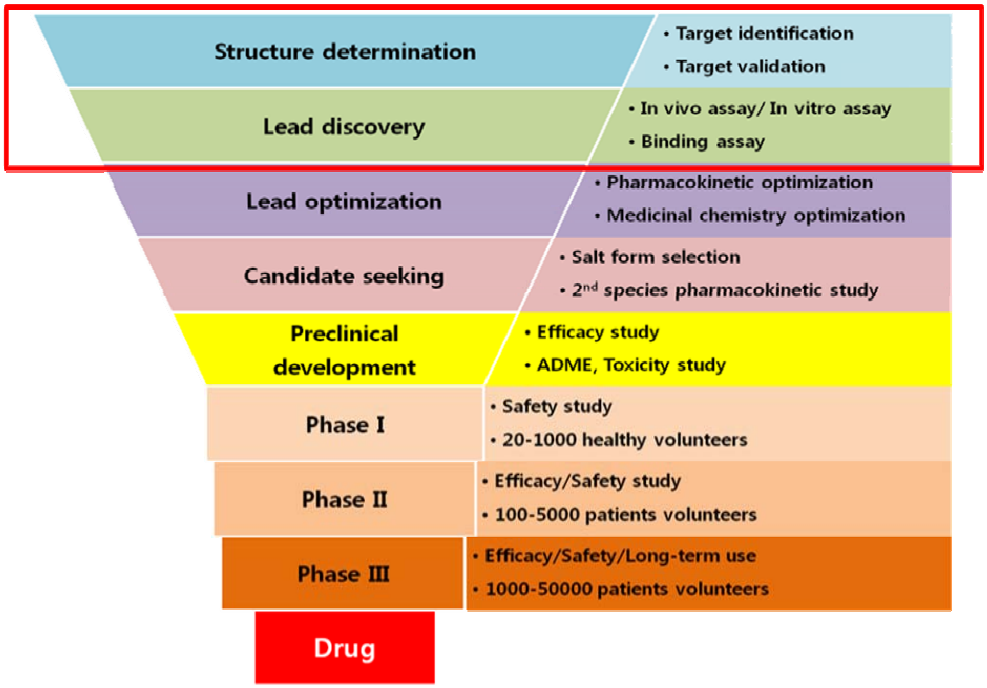


Figure 1. General scheme of drug development. Structure-based drug design (Red box) can reduce cost and time-consuming of early steps of the drug development [3].

Table 1. New drugs development using SBDD method.

Trade name	Popular name	Year	Company	Target protein	Target disease
Agenerase	Amprenavir	1999	Vertex/ GSK	HIV protease	AIDS
Aluviran	Lopinavir	2000	Abbott	HIV protease	AIDS
Relenza [4]	Zanamivir	1999	GSK	Neuraminidase	Influenza
Tamiflu [4]	Oseltamivir	1999	Gilead/ Roche	Neuraminidase	Influenza
Gleevec [5]	Imatinib	2001	Norvatis	BCR-ABL kinase	Chronic myeloid Leukemia
Tasigna [6]	Nilotinib	2006	Norvatis	BCR-ABL kinase	Chronic myeloid Leukemia
Tarceva	Erlotinib	2004	Genentech	EGFR	Non-small cell lung c cancer
Exanta	ximelagatran	2004 in Europe	AstraZeneca	Thrombin	Thromboembolic disease
Viagra [7]	Sildenafil	1998	Pfizer	PDE-5	Erectile dysfunction
Celsentri, Selzentry [8]	Maraviroc	2007	Pfizer	CCR5 antagonist	AIDS
Prezista [9]	Darunavir	2006	Tibotec	HIV protease	Multidrug-resistant HIV infection
Pradaxa [10]	Dabigatran etexilate	2008 in Europe	Boehringer Irigelheim	Thrombin	Venous Thromboembolism
Viracept [11]	Nelfinavir	1997	Agouron, Pfizer	HIV protease	AIDS
Reyataz, Zrivada [12]	Atazanavir sulphate	2003	Bristol-Myers Squibb, Novartis	HIV protease	AIDS

SBDD is composed of several steps, which can be divided largely into two part : structure determination and lead development. In this study, the Chapter 1 consists of the structure determination, the first part of SBDD, of Rv0569 protein sample from *Mycobacterium tuberculosis* and the Chapter 2 is about lead development, the second part of SBDD, of HsPDF protein sample from *Homo sapiens* (Figure 2).

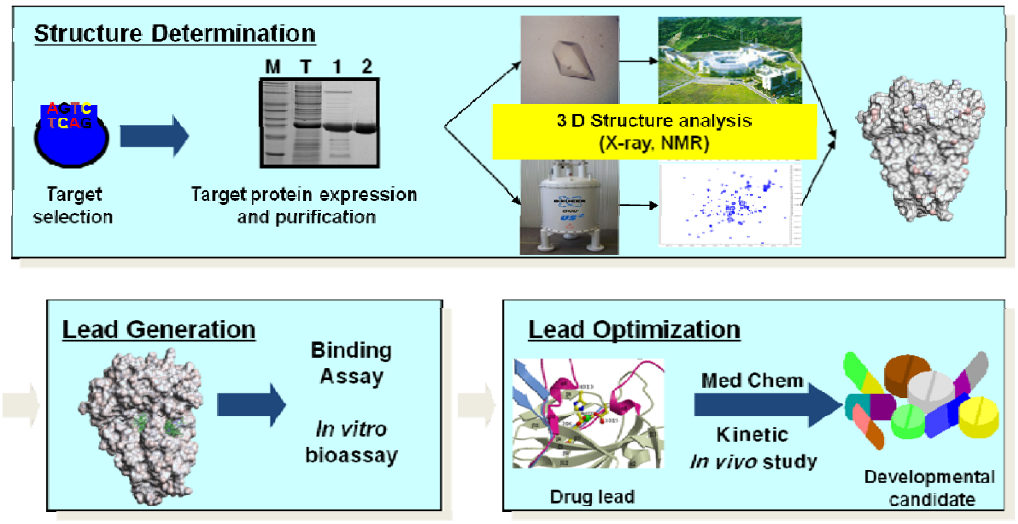


Figure 2. General scheme of SBDD.

Chapter 1. Structure Determination of Rv0569, Potent Hypoxic Signal Transduction Protein, from *Mycobacterium tuberculosis*

1.1 Abstract

The latent infection is the unique characteristic of *Mycobacterium tuberculosis* to overcome human immune response for survival. The *M. tb* develops adaptation to extreme stress conditions to increase the viability, thus acquires drug resistance easier than any other bacteria and maintains a long-term infection status without any symptoms. Rv0569 is a conserved hypothetical protein that overexpresses under dormant state induced by hypoxia, starvation, and medication. To study function and structure in detail, we have determined the solution structure of Rv0569 by NMR. The NOE and RDC restraints were used to calculate the structure, which was further refined with AMBER. Rv0569 is composed of five antiparallel β-sheets and one α-helix. Rv0569 shows structural similarity with its homolog Rv2302, yet there is a big difference in the orientation of the C-terminal α-helix between Rv0569 and Rv2302. According to the previous studies, Rv0569 might comprise a hypoxia induced operon with the Rv0570 which is located 29bp downstream of the Rv0569 and Rv0570 plays an important role in the latent

infection. From our structure and bioinformatics research, we suggest that Rv0569 contributes signaling transduction in hypoxic condition by binding with DNA for upregulation of Rv0570 or supports Rv0570 to bind ATP during dormancy of tuberculosis.

1.2 Introduction

Tuberculosis is a disease caused by *M. tb* mainly impacting host lung. One third of the world population are infected by *M. tb* and an annual morality rate has been reached 1.6 million people (Figure 3) [13, 14]. Because *M. tb* easily develops drug resistance against one drug, cocktail method has been generally used for the treatment of tuberculosis [15]. Because of the lack of nutrition and antibiotics, the infection rate is still high in the under-developing countries [13]. Even in advanced countries and developing countries, infection rate is also high. The causes of tuberculosis are weakened immune system due to stress [16], nosocomial infection [17], nutritional imbalance obtained from harsh diet [18], and so on. As tuberculosis is a worldwide issue, researchers have conducted many experiments in order to prevent the emergence of tuberculosis [19]. The strain mainly used for structural studies is *M. tb* H37Rv, which contains 3,989 genes [20, 21]. During past two decades, the structures of many proteins have been studied for new drug development.

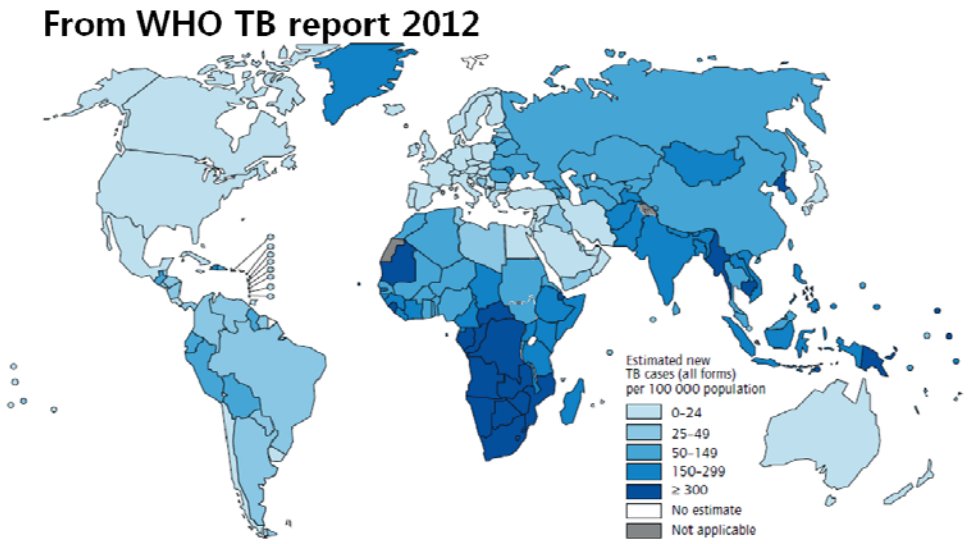


Figure 3. Estimated new TB cases per 100,000 population from WHO TB report [13]

Rv0569 is a conserved hypothetical protein composed of 88 amino acids. Its molecular weight is about 9,500 Da and its theoretical pI is 5.7. According to previous studies, Rv0569 is upregulated by DosR gene, which establishes and maintains the dormant stage of *M. tb*, in hypoxia [22-28]. Rv0569 and Rv0570 (NrdZ) located 29bp downstream of the Rv0569 together form a possible operon in the hypoxic condition [29]. When *M. tb* enters the latent state (Figure 4), Rv0570 converts RNA to DNA for DNA repairing system to survive [29]. Rv0570 is composed of the three domains. The first part is the ATP cone domain, the second part is the dimerization domain, and the third is the ribonucleotide reductase. From the result of BLAST search, Rv0569 is highly surmised to be a signal transduction protein or DNA-binding protein. BLAST search for the C-terminal α-helical region

of Rv0569 shows that the C-terminal α-helix has similarity in sequence with the part of ATP-dependent DNA helicase and nucleotide-binding proteins. Other paper reported that Rv0569 is overexpressed in the resistant strain of *M. tb* against streptomycin, the primary therapeutic [30]. Consequently, it is speculated that when *M. tb* encounters stress condition such as hypoxia or antibiotic environment, Rv0569 is overexpressed. Then, the overexpressed Rv0569 might participate in signal transduction to upregulate Rv0570 by binding promoter region or complement an ATP-related function of Rv0570. However, the mechanism of signal transduction and ATP-related function of Rv0569 and Rv0570 are still unclear.

In this study, the solution structure of Rv0569 was obtained and compared with the structure of Rv2302 [31], that has highly similar sequence identity to Rv0569. Here, we present the structure of Rv0569 and the structural differences between Rv0569 and its structural homolog Rv2302 and suggest the possible function of Rv0569.

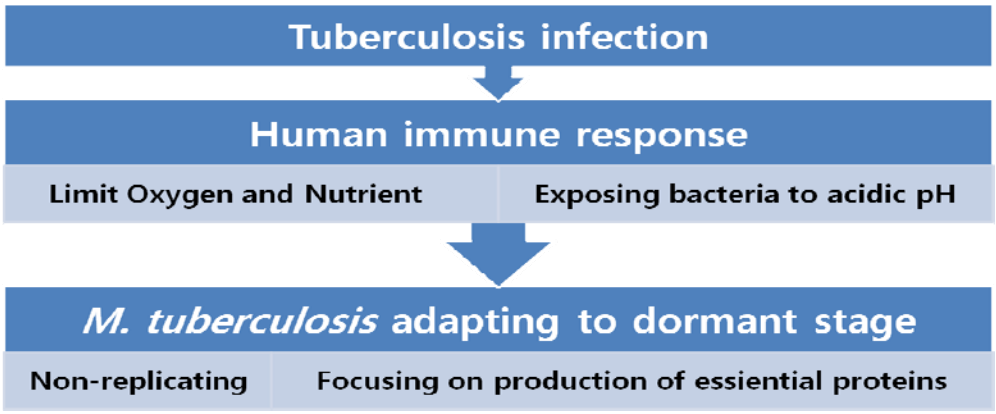


Figure 4. Latent infection. To overcome tuberculosis infection, human immune system limits oxygen and nutrient and exposes bacteria to acidic pH. Then *M. tb* reacts the response of human immune response by adapting to dormant state which *M. tb* focuses on the production of essential proteins to survive and stays in non-replicating state due to limit of resources.

1.3 Experimental procedures

1.3.1 Protein preparation

Rv0569 was cloned from genomic DNA of *M. tuberculosis* H37Rv (ATCC 27294) by PCR, and inserted into the pET-28a (+) vector (Novagen Inc., USA) between the restriction enzyme sites EcoRI (New England Biolab Inc., USA) and XhoI (New England Biolab Inc., USA). For the cleavage of N-terminal histidine tag, TEV [32, 33] recognized sequence was inserted in the sense primer following EcoRI restriction site. The uniformly ^{15}N- or $^{15}N/^{13}C$-labeled Rv0569 was expressed in BL21 (DE3) (Novagen Inc., USA) host cells grown in M9 minimal media using ^{15}N NH_4Cl and ^{13}C-glucose as a stable isotope source at 37 °C. Protein expression was induced by adding 0.4 mM IPTG when the OD_{600} reached 0.6 and the cells were grown at 20°C for next 16 hours. The bacterial cells were lysed and centrifuged to remove the insoluble debris. The supernatant was applied to IMAC column (GE healthcare Inc., USA) and the eluted with 100 mM imidazole. To remove imidazole and salt after IMAC column (GE Healthcare Inc., USA), eluted sample was dialysed three times with optimal buffer for TEV cleavage. For eliminating N-terminal histidine tag, TEV cleavage was carried out under the following condition [20 °C 160 rpm incubation, protein/TEV= 3 ratio]. After cleavage, the reaction mixture was loaded onto IMAC (GE healthcare Inc., USA) column, and our protein was identified in flow through. Size exclusive column XK16/60 (GE healthcare Inc., USA) was used for further purification and

exchanging buffer to the optimal NMR measurement. The molecular weight of the purified protein was confirmed by MALDI-TOF mass spectrometry (VoyagerDE Pro Biospectrometry Workstation, Applied Biosystems Inc., USA). The NMR sample was prepared at a concentration of approximately 2mM in 90% H_2O/ 10% D_2O containing 40mM HEPES (pH 7.0), 1mM DTT, 1mM NaN_3.

1.3.2 Circular Dichroism measurement

Circular dichroism (CD) spectra were collected with protein concentration 30 μM using 10 mm path length cuvette cell on a J-715 spectropolarimeter (Jasco Inc., Japan). The machine was equipped with a Peltier temperature control system (Model PTC-348WI). Samples were scanned three times for the wavelength ranges from 190 to 240 nm at the rate of 100 nm/min with a bandwidth of 1 nm and a response time of 2 s. The temperature wavelength scanning was carried out in the temperature range of 20–80 °C at molar ellipticity of 222 nm at the rate of 1 °C/min. To analyze physical and thermodynamic properties of protein, various solutions were used.

1.3.3 NMR measurement

All NMR spectra were acquired at 303K on Bruker AVANCE AMX 500, Bruker AVANCE DRX 600 with cryoprobe and Varian 900MHz with a cryogenic probe. The NMR data were processed with NMRPipe [34] and analyzed using NMRViewJ [35] . Chemical shifts for all atoms were calibrated with internal DSS (2,2-dimethyl-2-silapentane-5-sulfonate) as previously reported [36]. Backbone and side-chain assignments were carried out with 2D and 3D experiments, ^{1}H-^{15}N HSQC [37, 38] , HNCACB, CBCA(CO)NH [39], HNCO, HNCACO, HBHA(CO)NH, ^{15}N-edited TOCSY-HSQC, HCCH-TOCSY [40] , and aromatic ^{13}C-edited NOESY-HSQC. Distance restraints NOE were obtained from ^{15}N-edited NOESY-HSQC [41, 42], ^{13}C-edited NOESY-HSQC [41, 42], and aromatic ^{13}C-edited NOESY-HSQC spectra. Proton-deuterium exchange was investigated by collecting a series of ^{1}H-^{15}N HSQC spectra recorded 0.5, 2, 4, 16, 48 hours after lyophilizing the NMR sample and redissolving in 99.9% D_2O. The ^{1}H-^{15}N steady-state heteronuclear NOE experiment [43] was performed to analyze dynamic properties of the protein. ^{1}H-^{15}N residual dipolar coupling values of protein were obtained by using Pf1 phage (ALAS biotech Inc., USA) at the concentration of 12 mg/ml [44]

1.3.4 Structure calculation and validation

The structures were calculated by the program package CYANA [45] and refined using AMBER [46-48]. Distance restraints were derived from inter-proton nuclear Overhauser effect (NOE). Dihedral angles (ϕ, Ψ) were determined from backbone chemical shifts predicted by TALOS [48, 49]. The initial structures were calculated with the automatic assignment module of the CYANA program. The 60 lowest energy structures were selected as the models for refinement with RDC restraints. Among ten thousand structures were calculated by CYANA, the 60 lowest energy structures were used as the initial structures and refined using AMBER in implicit solvent mode. Finally, the 20 lowest energy structures were selected to represent the Rv0569 protein. The final 20 structures were analyzed by using the program packages Molprobity [50, 51] and PSVS validation package [52]. The RDC back-calculations were conducted using REDCAT for the validation of RDC experiments [53]. The DNA binding site of Rv0569 was predicted by Castp [54] and Patch Finder Plus [55]. The programs MOLMOL [56], PyMOL (www.pymol.org), and UCSF Chimera [57] were used to visualize the 20 energy-minimized conformers. The APBS program [58] and Coulombic surface coloring method with UCSF Chimera were used to calculate the electrostatic surface potentials of the protein. The electrostatic potential maps were calculated by numerically solving the Poisson-Boltzmann equation based on molecular mechanics.

1.4 Result

1.4.1 Protein preparation

Rv0569 was highly expressed in BL21(DE3) host cells in soluble form and its expression was confirmed with SDS-PAGE (Figure 5). For reducing additive terminal residues of the fusion protein, we designed the sense primer containing TEV recognition site to cleave N-terminal six histidine tag [32]. Although TEV cleavage is usally conducted at the concentration ratio (protein/TEV = 50 ratio), Rv0569 was completely cleaved by the excess amount of TEV (protein/TEV = 3 ratio) at 20°C in the incubator (160 rpm). After TEV cleavage, the molecular weight of the protein was measured with MALDI-TOF mass spectroscopy in NCIRF (Figure 6). As a result of purification, we obtained 40 mg of the protein from 2 liters of M9 culture.

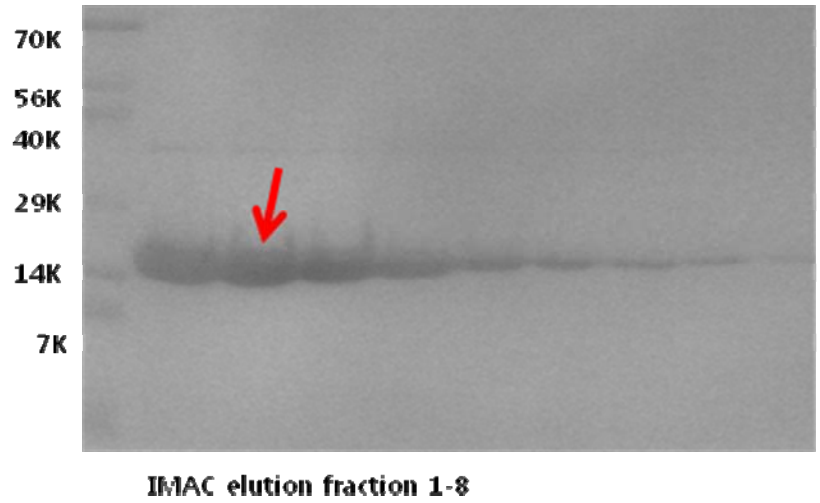


Figure 5. SDS-PAGE of IMAC elution of Rv0569 with TEV recognition site.

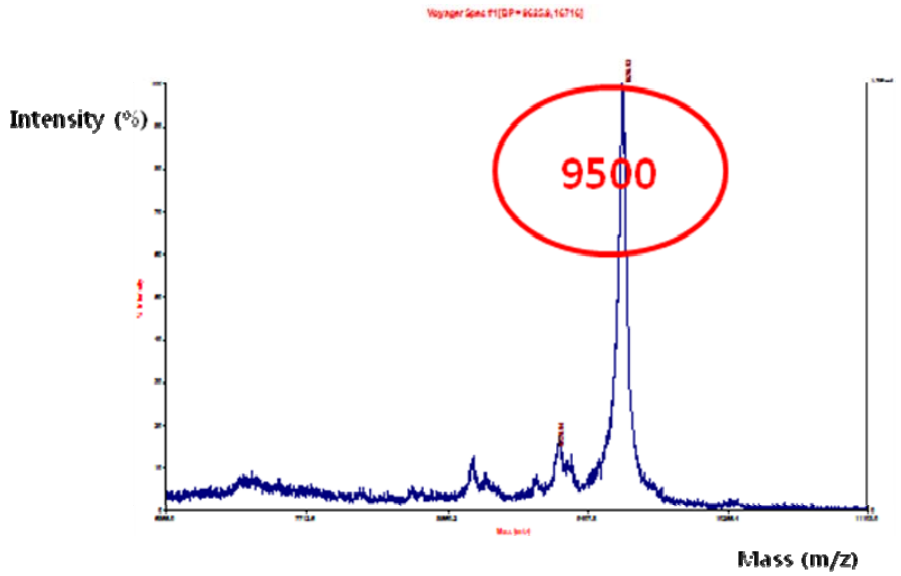


Figure 6. MALDI–TOF mass spectroscopy measurement of Rv0569 after TEV cleavage.

1.4.2 NMR spectrum analysis

Ninety-eight percent (%) of the resonance assignment of all ^{1}HN, ^{15}N, ^{13}CO, $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and 81.2% of all side-chain was accomplished with 2D-, 3D- NMR spectrum analysis. ^{1}H-^{15}N HSQC spectrum of ^{15}N uniformly labeled sample showed very good signal dispersion and each amide ^{1}H and ^{15}N spin correlation was assigned in Figure 7 without five proline residues. Unfortunately, the peak of the residue 31 serine was not detected and it seemed to be related with labile property of solvent exposed residues [59]. The peaks from a flexible region tend to appear at a ^{1}H resonance frequency lower than 8.6 ppm. All peaks from residues 73 to the C-terminus appear at a ^{1}H resonance frequency lower than 8.6 ppm which represented in red (Figure 7).

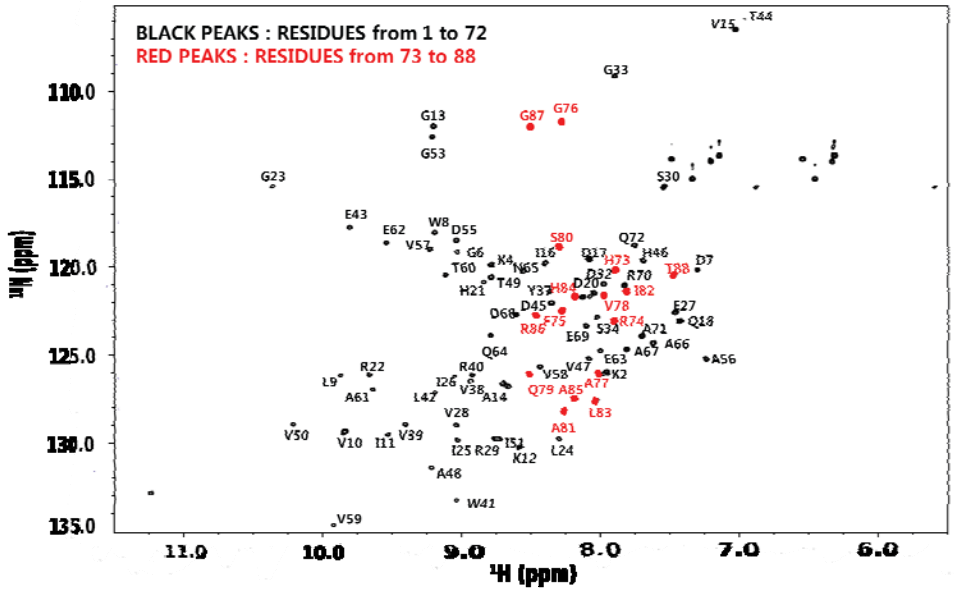


Figure 7. 2D HSQC spectra of *M. tb* Rv0569. The black peaks represent residues from 1 to 72, the red peaks represent residues from 73 to 88. All peaks from residues 73 to the C-terminus appear at a ^{1}H resonance frequency lower than 8.6 ppm.

In addition, there is no intramolecular NOE from residue 73 to residue 88 in ^{15}N-edited NOESY-HSQC and ^{13}C-edited NOESY-HSQC. The fitted data of heteronuclear NOE demonstrated that the residues from 73 to 88 are very flexible comparing other part of protein and this result explained why this region is lack of intramolecular NOE (Figure 8).

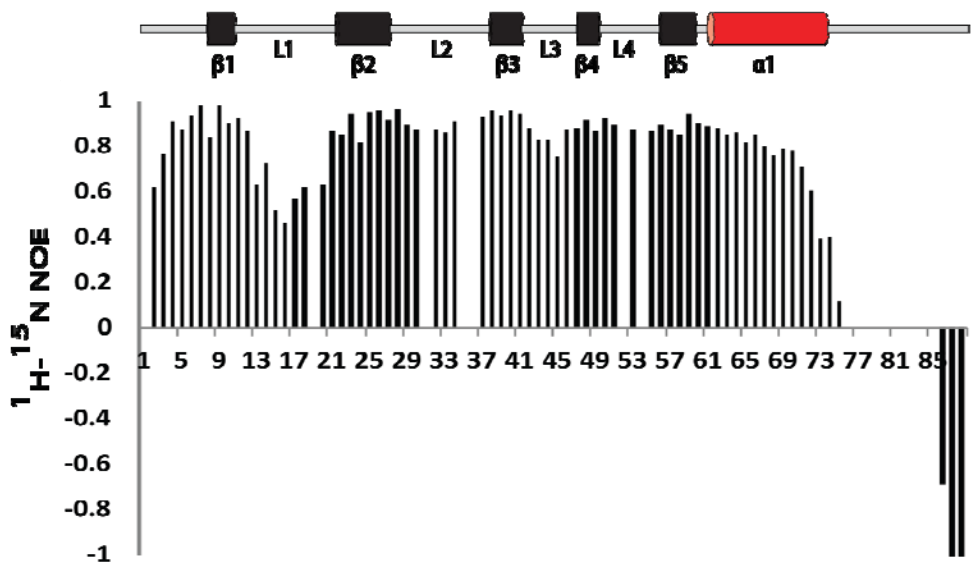


Figure 8. **^{1}H - ^{15}N heteronuclear NOE data and secondary structure.**

The hydrogen bond information was revealed by the H-D exchange experiment. The ^{1}H-^{15}N HSQC spectra were obtained in 0.5 hour, 2 hour, 4 hour, 16 hour, and 48 hour after redissolving ^{15}N-labeled NMR sample into D_2O. After 16 hours, 12 residual amide protons were remained (residue number 6, 7, 8, 9, 10, 23, 24, 25, 26, 38, 39, and 40), and all the residues are located throughout β-strands (β1, β2, β3, and β5) except for the amide proton of residue number 7. From NOE data analysis, those 12 residual amide protons can make hydrogen bonds with carbonyl oxygen 25, 4, 59, 23, 57, 9, 40, 7, 38, 27, 48, and 24, respectively (Figure 9) [59].

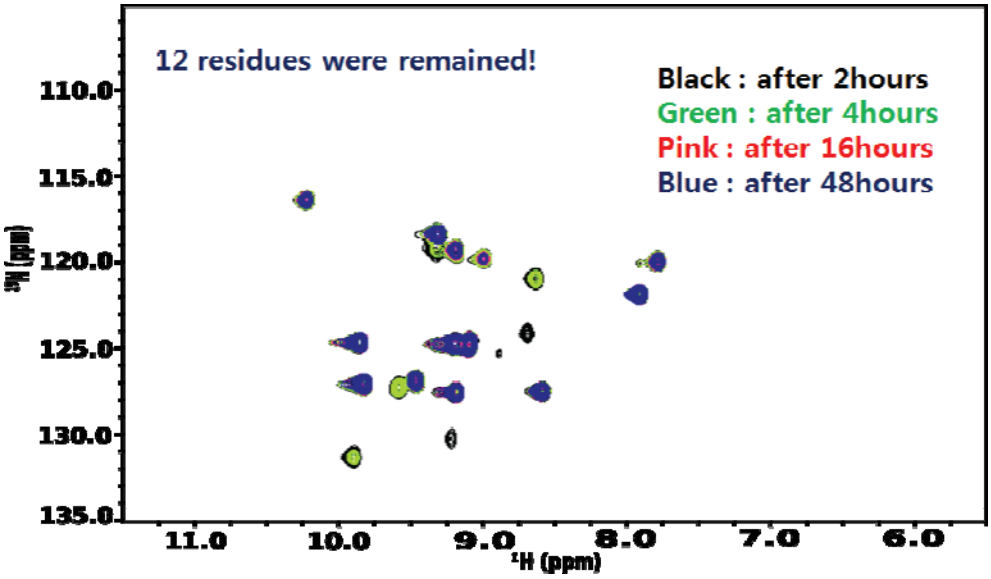


Figure 9. Hydrogen-deuterium exchange measurement. Black, Green, Pink and Blue peaks represent HSQC spectrum redissolving after 2 hours, 4 hours, 16 hours and 48 hours, respectively.

To obtain the orientation of the protein for the structure calculation, the residual dipolar coupling experiment was conducted on Pf1 macrophage [44]. About 5Hz~15Hz of ^{15}N-^{1}H dipolar couplings were detected at the 12 mg/ml of the Pf1 in the NMR sample (Figure 10) [44]. The RDC data was correlated with the heteronuclear NOE data in the point of rigidity of secondary structure.

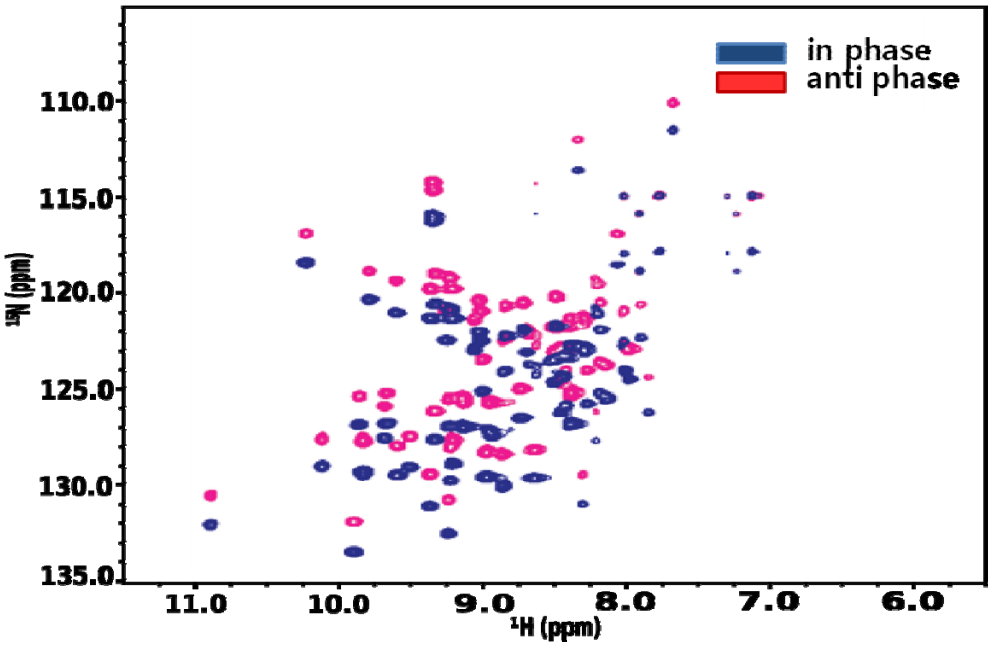


Figure 10. In phase (blue peaks) and anti phase (red peaks) RDC spectra of Rv0569.

1.4.3 Structure calculation and validation

The structure statistics including the NMR restraints for calculation are summarized in Table 2. From the results of size exclusion chromatography and the lack of intermolecular NOE in NOE experiments show that Rv0569 is a monomer. The 60 lowest energy structures were selected from the 10000 structures that were calculated and refined using NOE and RDC parameters by CYANA. The RMSD of the backbone atoms of the 20 superimposed lowest energy structures (ordered residues) after AMBER refinement was 0.9 Å even containing four loops.

Table 2. Structural statistics of Rv0569.

Parameters	Values
Total Distance Restraints	1076
Intraresidue (i,i)	260
Sequential (i, i+1)	344
Medium Range (2<=|i-j|<=4)	154
Long Range (|i-j| > 4)	318
Total Dihedral Angle Restraints	114
PHI	57
PSI	57
Total RDC Restraints	75
H-N (pf1 phage)	75
Ensemble statistics (20 structures)	
Secondary Structure Elements	

alpha helices:	61A-74R
beta strands:	47V-49T, 38V-41W, 22R-27E, 8W-11I, 56A-59V
Energies	
mean restraint energy (kcal mol^{-1})	19.82
mean AMBER energy (kcal mol^{-1})	-3196.88
RMSD deviation	
All residues (1-88)	
backbone atoms	5.7 Å
heavy atoms	6.1 Å
Ordered residues (4-12, 14-29, 36-52, 54-70)	
backbone atoms	0.9 Å
heavy atoms	1.3 Å
Close Contacts and Deviations from Ideal Geometry	
Number of close contacts (within 1.6 Å for H atoms, 2.2 Å for heavy atoms)	0
RMS deviation for bond angles	2.2 °
RMS deviation for bond lengths	0.014 Å
MolProbity Analysis	
All atom clash score[a]	1.27
Ramachandran favored (%)	90.61
Ramachandran outliners (%)	2.68
Bad rotamers (%)	5.53

[a]Clash score is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.

After conducting RDC back-calculation using REDCAT, we can conclude that our observed RDC values are well correlated with the calculated RDC values [53]. The

Q factor (0.0563467) and RMSD (0.394649) resulted from the RDC back-calculation are significantly low (Figure 11 and Table 3).

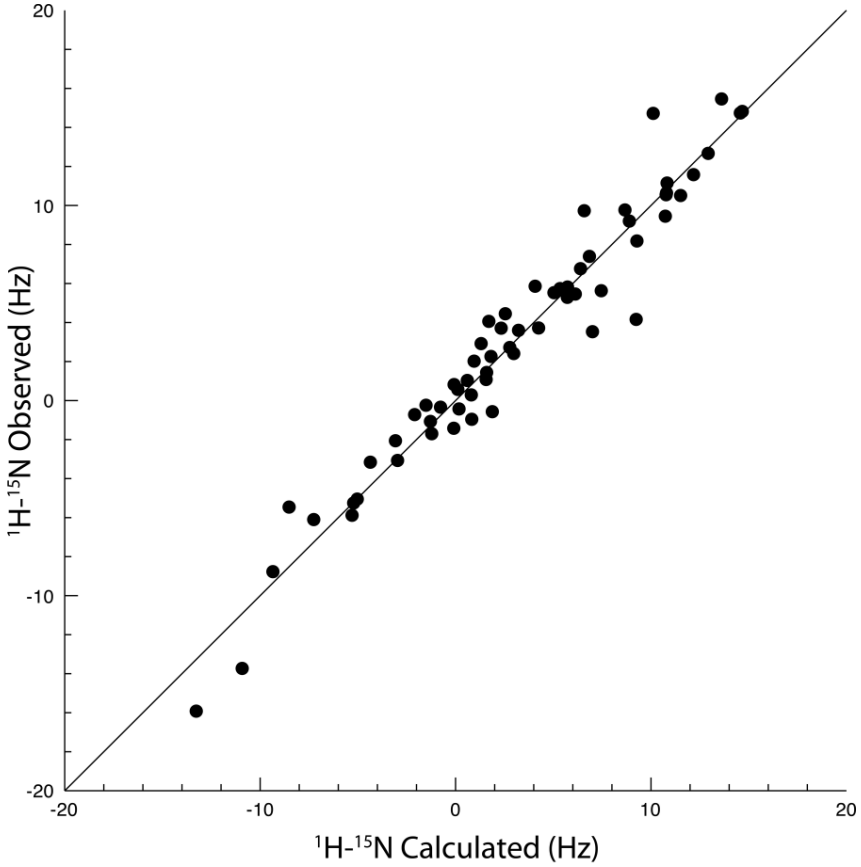


Figure 11. RDC back-calculation using REDCAT. The value of Q factor is 0.056 and the value of RMSD is 0.394.

Table 3. RDC table.

Residue Number	Observed RDC (Hz)	Calculated RDC (Hz)	Residue Number	Observed RDC (Hz)	Calculated RDC (Hz)
1	n.d.*	n.c.**	45	-0.96	0.828086
2	6.76	6.39443	46	-5.25	-5.21493
3	n.d.	8.09243	47	3.6	3.22194
4	8.18	9.2784	48	7.39	6.84805
5	-5.46	-8.51897	49	2.72	2.77421
6	2.26	1.81181	50	-8.77	-9.34725
7	14.74	14.5712	51	n.d.	-13.6552
8	3.71	2.33336	52	n.d.	n.c.
9	-1.42	-0.09196	53	3.72	4.24865
10	9.73	6.58317	54	n.d.	n.c.
11	2.41	2.97539	55	5.63	7.45441
12	5.82	5.72753	56	15.46	13.6107
13	10.51	11.5155	57	1.03	0.599727
14	n.d.	10.076	58	0.57	0.114473
15	-5.05	-5.03109	59	2.02	0.947892
16	-6.1	-7.25812	60	-0.34	-0.76953
17	1.08	1.5635	61	11.15	10.8225
18	4.16	9.2404	62	9.45	10.7329
19	n.d.	n.c.	63	14.82	14.6709
20	14.72	10.1118	64	9.77	8.66678
21	5.29	5.71943	65	5.46	6.12501
22	4.06	1.70159	66	12.67	12.9252
23	-0.43	0.175288	67	11.58	12.1766
24	0.29	0.804403	68	5.74	5.34622
25	9.2	8.89275	69	3.53	7.00509
26	-5.88	-5.29743	70	n.d.	14.0484
27	n.d.	-14.0832	71	5.86	4.07103
28	-15.92	-13.2702	72	0.82	-0.07567
29	n.d.	-10.0989	73	10.64	10.7932
30	4.45	2.54489	74	n.d.	2.2017
31	n.d.	10.5477	75	-1.07	-1.2866
32	-1.7	-1.21956	76	-0.72	-2.09404
33	1.44	1.58895	77	n.d.	-0.843796
34	-2.06	-3.0769	78	n.d.	4.09893
35	n.d.	n.c.	79	n.d.	13.9207
36	n.d.	n.c.	80	n.d.	3.94881
37	-13.73	-10.9192	81	n.d.	13.2282
38	n.d.	-13.4801	82	n.d.	-7.93752
39	2.92	1.30729	83	n.d.	-5.72349
40	n.d.	3.27256	84	n.d.	-12.7652
41	10.55	10.7733	85	n.d.	4.21482
42	-3.07	-2.96853	86	n.d.	3.25877
43	-0.57	1.87927	87	-0.24	-1.51285

44	-3.16	-4.36037	88	5.53	5.03465

n.d.* means not detected.
n.c.** means not calculated.

Rv0569 possesses five β-strands and one α-helix. This result is well matched with the CD data (Figure 12), concluding that the secondary structure of Rv0569 mostly consists of β-sheets. Each β1 - β2, β2 - β3, β3 - β4, and β5 - β1 forms antiparallel β-sheets (Figure 13). The five β-strands correspond to residues 8 - 11, 22 - 27, 38 - 41, 47 - 49, and 56 - 59, respectively and the α-helix corresponds to residues 61-74. The hydrogen bonds in antiparallel β-strands make the structure more rigid and the heteronuclear NOE data supports the rigidity of these regions (Figure 8). The orientation of α-helix is outward, not facing the globular β-strands. The lack of intermolecular NOE from NOE experiment confirms that the C-terminal region (residue 74-88) is unstructured. Also, the results of heteronuclear NOE experiment for monitoring the dynamics of the protein support that this C-terminal region is very flexible. The loop1 between β1 and β2 is more flexible than loop2, loop3, and loop4 because these loop2, loop4 have proline residues and loop2 and loop3 have aromatic residues which have more interresidue NOE.

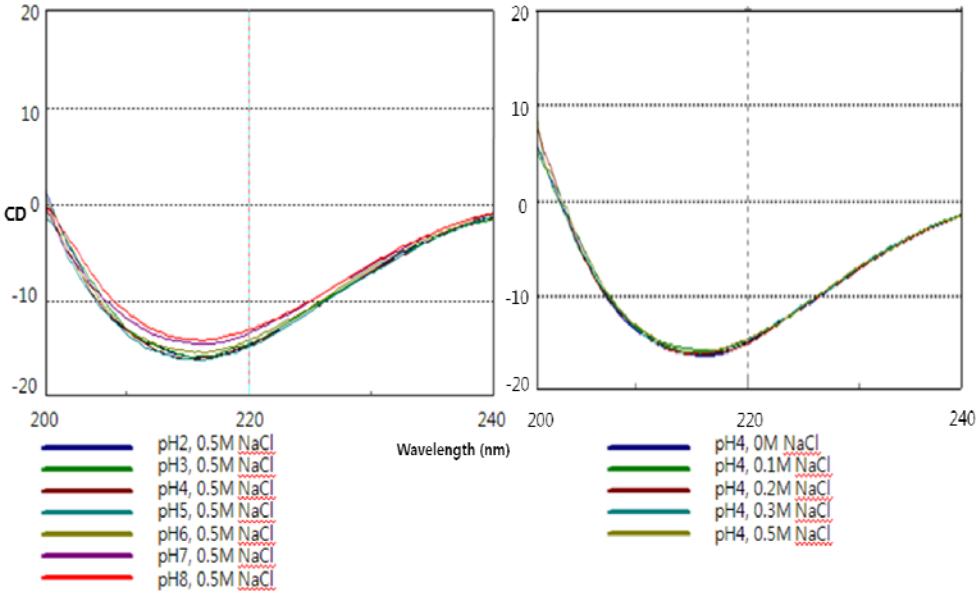


Figure 12. CD spectra in the various pH and salt conditions.

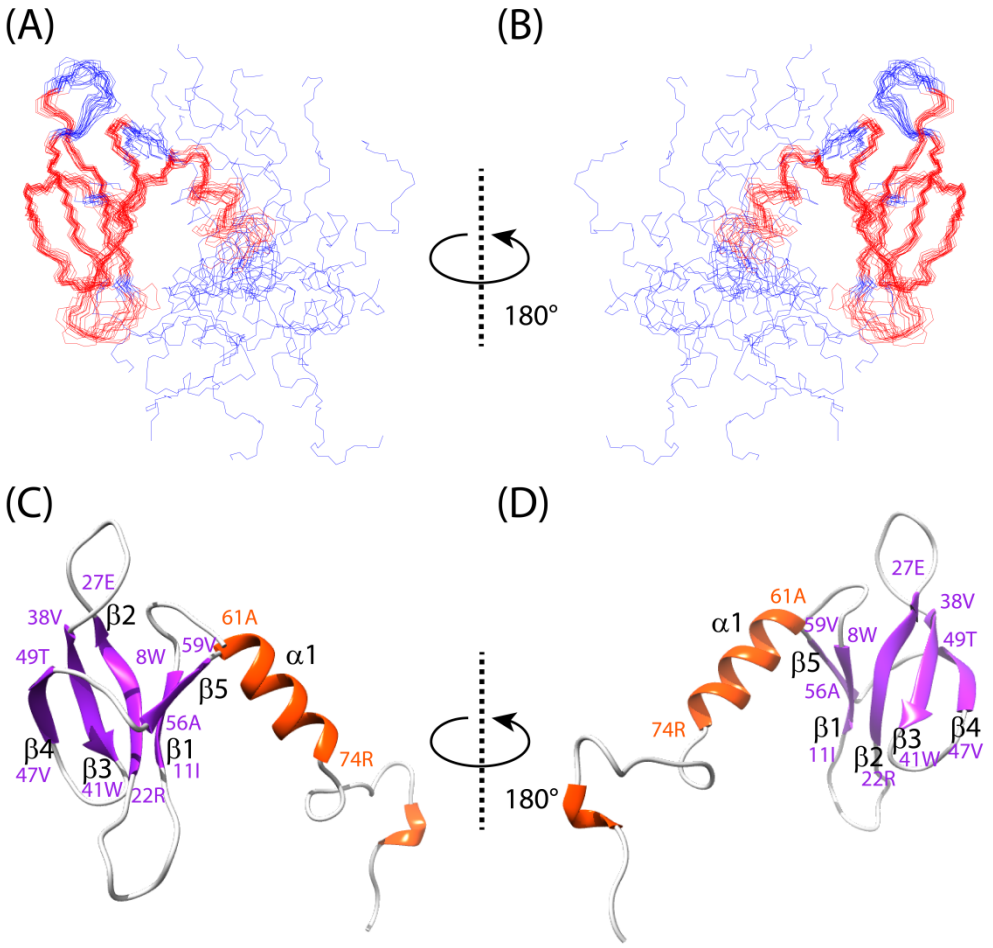


Figure 13. Solution NMR structure of *M. tb* Rv0569. The superimposition with ordered residues of the 20 representative structures is shown in red and flexible residues are shown in blue (A, B). The lowest clash scored structure from the Molprobity validation was represented with ribbon colored in orange (α-helix), magenta (sheet), and white (loop, terminus), respectively (C, D). There is very short α-helix except α-helix (α1) but is not found consistently through 20 structures of ensemble.

1.5 Discussion

TEV cleavage was performed for eliminating N-terminal histidine tag of the fusion protein, and the condition of restriction was rather harsh than the usual condition. In the previously reported papers, TEV cleavage was conducted at a ratio of 50:1 (protein: TEV) without incubation. However, Rv0569 was restricted at the ratio of protein (3), TEV (1) with 160 rpm incubation. This was quite unusual, and we could know possible reasons after structure calculation. To increase the efficiency of the TEV cleavage, the recognition site should be well exposed [60]. However, the N-terminal region of Rv0569 is quite rigid as shown in heteronuclear NOE data (Figure 8), and surrounded by loop 2 and β5. This steric hindrance interrupts the TEV recognition, reducing the efficiency of digestion

Comparing Rv0569 and Rv2302 of which structure was solved by solution NMR in 2006 [31], residues 1-64 showed high sequence identity as shown in Figure 14 and this region is composed of the five β-sheets and the partial α-helix (Figure 15A and 15B). The residues 64-88 of Rv0569 did not have high sequence identity with Rv2302. Through the overlaid structures of both proteins, Rv2302 has a similar topology βββββα with that of Rv0569. However, the α-helix of Rv2302 is close with globular β-sheets enough to observe NOE between α-helix and β1, β2, β3, and β4, respectively in the ^{15}N-edited or ^{15}N/ ^{13}C-edited NOESY spectra. It looks like that the “arm-shaped” α-helix embraces the “body -shaped” bundle of β-sheets (Figure 15A and 15B). On the other hand, the α-helix of Rv0569 does not embrace

the β-sheets and this was confirmed with the lack of NOE in the NMR spectra. Especially, the orientation of α-helix is outward against the rigid β-sheets. Base on the sequence analysis of Rv0569, residues 1-64 have high sequence identity with Rv2302. However, in the view of structure, residue 1-59 showed high structural identity with Rv2302 and the direction of α-helix is totally different.

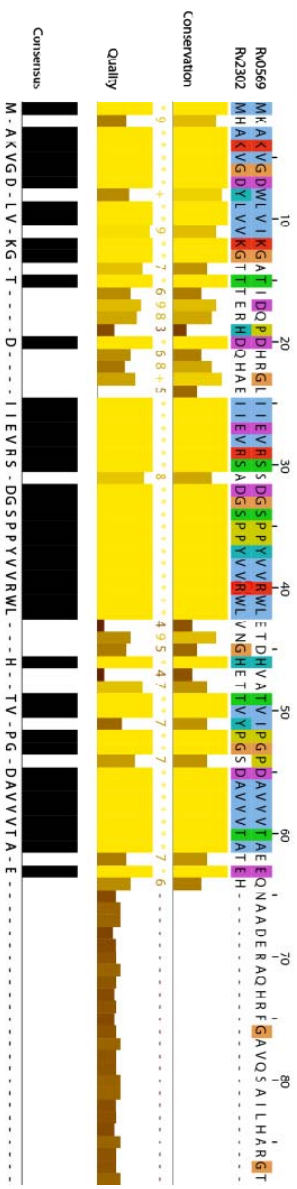


Figure 14. Sequence alignment of Rv0569 with Rv2302 (PDB code 2A7Y) from *M. tb*. Sequence alignment was performed with ClustalW [61, 62] and JalView [63, 64] and the conserved residues in both sequences are colored in boxes.

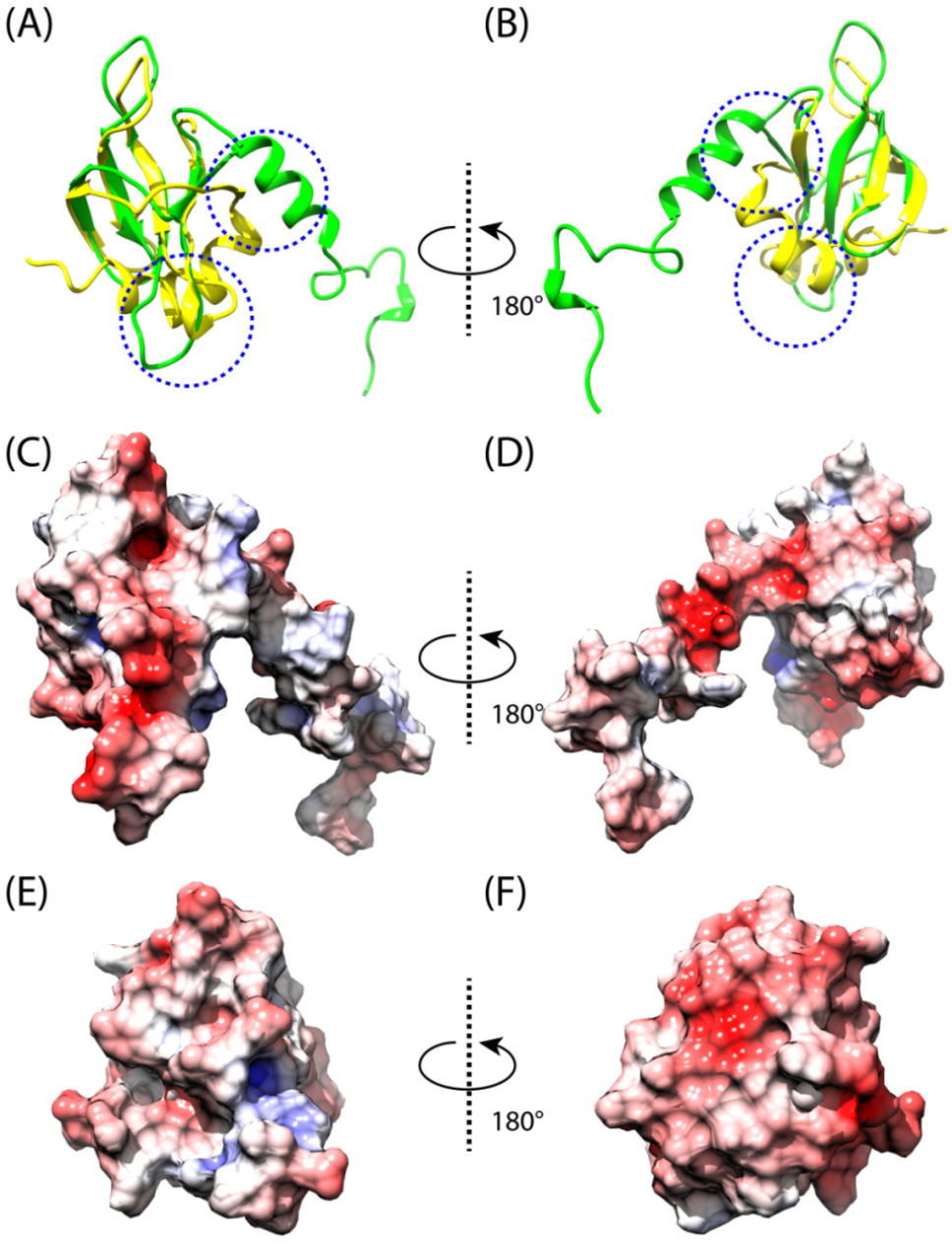


Figure 15. Structure alignment of Rv0569 with Rv2302 (PDB code 2A7Y) from *M. tb*.

The Match maker module in UCSF Chimera was used to align structure with

Needleman-Wunsch algorithm (A, B). Aligned proteins are shown with ribbon presentation method, colored in yellow (Rv2302) and green (Rv0569), respectively (A, B). And the mismatched regions in the aligned proteins are marked with blue dotted circles. The electrostatic surface potential is colored in blue (positively charged residues) and red (negative charged residues) with an energy scale of ±3 in units of kcal/ (mole*e) for Rv0569 (C, D) and Rv2302 (E, F). All protein sequences and structures were obtained from the PDB database.

A search for structure homologues using DALI server [65] with in Protein Data Bank [66] showed the same result by inputting PDB file with or without unstructured C-terminal region. Rv0569 has structural similarity with spindlin-1 (PDB code:2ns2, Z-score 4.2, RMSD 13.6, identity 12%) [67], restriction endonuclease HPY99I (PDB code:3gox, Z-score 4.1, RMSD 6.6, identity 15%) [68], PHD finger protein 19 (PDB code: 4bd3, Z-score 4.0, RMSD 3.5, identity 17%) [69], transcription antitermination protein NusG (PDB code: 1npp, Z-score 3.9, RMSD 2.1 identity 13%) [70], the hypothetical protein RV2302 (PDB code: 2a7y, Z-score 3.8, RMSD 4.0, identity 61%) [31], DNA-directed RNA polymerase SUBUNIT E (PDB code: 3p8b, Z-score 3.7, RMSD 2.6 identity 12%) [71] and so on. Unfortunately, the six kinds of proteins with highest Z-score do not have a consistent function. However spindlin-1 and PHD finger protein 19 are associated with the DNA damage response and the epigenetic control of the gene expression by interacting with histones, and the protein with the second highest Z-score, the

restriction endonuclease HPY99I recognizes a specific DNA fragment for digestion. The transcription antitermination protein, NusG, is a gene regulator by binding DNA or RNA. The DNA-directed RNA polymerase subunit E is also related with the nucleic acids. It is remarkable that these proteins which have similar folding with Rv0569 including the orientation of the α-helix show higher Z-score than Rv2302 which has the highest sequence identity with Rv0569. Therefore, Rv0569 might have a function related to DNA or RNA based on the result of DALI server search.

According to the previous research, Rv0569 is highly expressed in stress conditions such as hypoxia, starvation and toxic environments [22, 23, 26, 27]. Another research reports that the dormant program of *M. tb* was interrupted when Rv0569 gene was deleted [72]. Rv0570, the right next gene of Rv0569, is also upregulated in hypoxia and thought to form a hypoxia-induced operon with Rv0569 [29]. Rv0570 is an ATP-dependent ribonucleotide reductase that preserves DNA by converting useless RNA in resource-limit condition of dormancy [25]. It is obvious that Rv0569 is very essential for surviving of M. tb in the latent infection state and the function of Rv0569 is closely related to Rv0570. The BLAST search implies that Rv0569 has homology with DNA-binding proteins or signal transduction proteins (Table 4). Rv2302, which has similar topology of β-strands with that of Rv0569, shows almost the same BLAST result as the result from Rv0569. However, from BLAST search result with the C-terminal region (residues 60 - 88) showing different orientation of α-helix, Rv0569 has high sequence identity with

ribonucleotide reductase beta subunit or ATP-dependent helicase from heterogeneity. Rv0570 is annotated as ribonucleotide reductase, and its funtion involves with ATP during non-replicating period of bacteria such as the dormant stage when maintaining low level of ATP is critical for surviving [25, 29]. The hypoxia-induced dormancy response is established and maintained by regulator, DosR together with ATP-related proteins such as USP-domain containing Rv2623 and CBS-domain containing Rv2626c [23].

Table 4. Result of sequence similarity research.

Rv0569 sequence	Matched sequence	Identities (%)
C-terminal region with α-helix: DERAQHRFGAVQSAILT	ribonucleotide-diphosphate reductase subunit beta [*Saccharomonospora azurea* SZMC 14600]: QYAAD-RAQRRLGAIES----ARGT Sequence ID: ref|ZP_13032212.1|	58 (14/24)
C-terminal region with α-helix: DERAQHRFGAVQSAILT	ATP-dependent DNA helicase RecG [*Geobacter* sp. M18]: DE—QHRFGVVQRALL Sequence ID: ref|YP_004198467.1|	69 (11/16)
N-terminal region with β-strands MKAKVGDWLVIKGATIDQPDHRGLIIEVRSSDGSPPYVVRWLETDHVATVIPGPDA	DNA-binding protein [*Streptomyces acidiscabies* 84-104]: MRARIGDILVVESPTTGVTRRDGEIVGLRHEDGTPPYDVRWSETHEVTLVFPGPDA Sequence ID: ref|ZP_10457630.1|	48 (27/56)

N-terminal region with β-strands MKAKVGDWLVIKGATI DQPDHRGLIIEVRSSDG SPPYVVRWLETDHVAT VIPGPDAVV	signal-transduction protein [*Nocardioides* sp. JS614] MHATKGDHLVIHGTHVGEPLRDGEIV EVRGPDGAPPYVVRWADNGHESLTY PGPDAEV Sequence ID: ref|YP_921516.1|	52 (30/58)

The results from DALI database, structure calculation, previous papers and BLAST search suggested that Rv0569 could be involved in hypoxia signal transduction. It might function in upregulation of Rv0570 by binding promoter region of Rv0570 or in complementation of ATP-related operation of Rv0570 in dormancy stage of *M. tb*. Although, Rv0569 and Rv2302 have similar sequence and structure, Rv2302 is not related with Rv0570. It seems that the orientation of the α-helix strongly contributes to the functional differences of Rv0569 and Rv2302. The result from DNA binding prediction shows that there is enough space between β-sheets and α-helix in Rv0569 to bind DNA but Rv2302 has too limited space between β-sheets and α-helix to bind DNA due to the position of embraced α-helix around β-sheets. Especially, the positively charged area in Rv0569 could increase the possibility of DNA binding and stability of DNA-protein complex (Figure 16). Also, it suggested that the long flexible C-terminal α-helix of Rv0569 could bind small molecules including ATP and participate in a functional process of DNA-protein complex through binding or transferring small molecules.

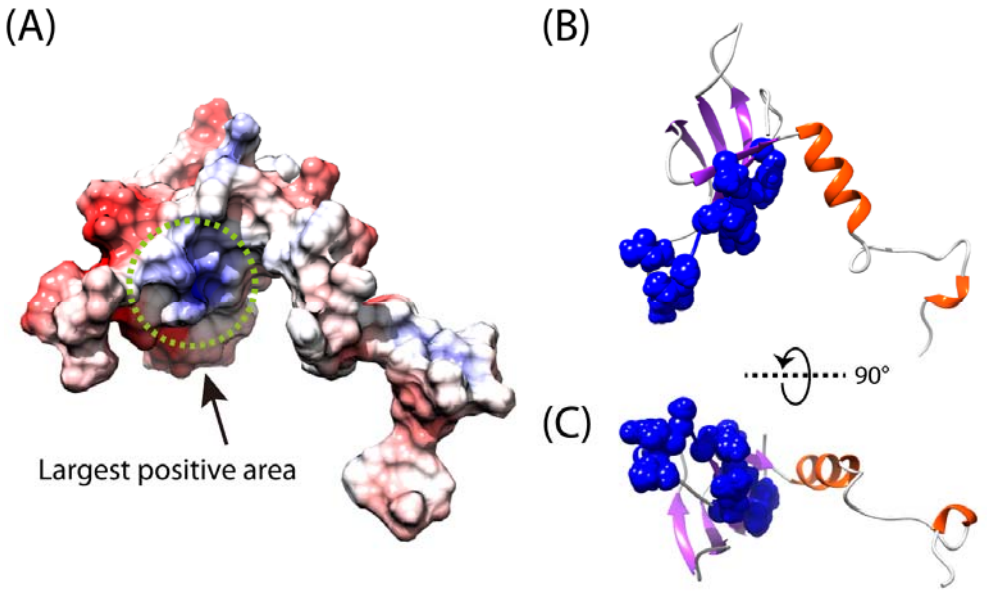


Figure 16. Hypothetical model of Rv0569 function. The electrostatic potential of protein (A) is shown as same scheme with that of Figure 15C and 15D. The residues on the largest positively charged area are represented as blue spheres including Trp8, Val10, Ile11, Lys12, Gly13, Ala14, Thr15, Ile16, Gln18, Pro19, Asp20, Arg22, and Val57. The color scheme of ribbon representation is same with that of Figure 13C and 13D.

1.6 Conclusion

To overcome tuberculosis infection, human immune system limits oxygen and nutrients, and exposes bacteria to acidic condition [25]. Then *M. tb* reacts the response of human immune response by adapting to dormant state which *M. tb* focuses on the production of essential proteins to survive and stays in non-replicating state due to limit of resources [23, 25]. The dormancy is the key of understanding latent infection of *M. tb* which increases a risk of death by phenomena such as infecting without symptoms or drug resistance. The dormancy is a big theme of recent tuberculosis research. As Rv0569 was known to be essential for dormancy program [72], we determined the three-dimensional structure of Rv0569 in order to discover its function. Rv0569 possesses βββββα topology, and the orientation of α-helix is different from that of Rv2302, a homolog of the Rv0569. However, Rv2302 is not related to Rv0570 which make a possible operon with Rv0569 under hypoxic condition and Rv0570 is very essential for survival during latent infection [29, 31]. The reason of functional differences between Rv0569 and Rv2302 is much likely to be the different orientation of α-helix at C-terminus. Based on our study, we conclude that Rv0569 could bind to promoter DNA of Rv0570 for upregulation of Rv0570 or support ATP related function of Rv0570 in dormancy stage of *M. tb.*

Chapter 2. Lead compound development of HsPDF from *Homo sapiens*

2.1 Abstract

Potent inhibitors of human peptide deformylase (HsPDF) were screened using known PMT analog inhibitors of bacterial peptide deformylase. Forty-three species of PMT analogs that are non-peptidyl bacterial PDF inhibitors like actinonin were selected using virtual screening GOLD. Ten species out of 43 that could bind to HsPDF were selected and their antitumor activities were tested. Among them, 4 species (PMT-172, PMT-173, PMT-199, and PMT-201) showed outstanding antitumor effects in the MTT assay. HsPDF-PMT binding was confirmed by a ^{1}H-CPMG-T_2 filter NMR experiment, leading to a significant change in peak intensity for PMT-172 and PMT-199. These results suggest that PMT analogs could possibly interact with HsPDF and be a novel anticancer drug candidate.

2.2 Introduction

The peptide deformylase (PDF) is a unique feature of prokaryotes and believed not to exist in eukaryotes up to even only a few years ago [73]. PDF is essential for ribosomal translation by deformylating the initiating N-formylmethionine in prokaryotes. Since the initiating methionine of eukaryotes is not N-formylated, PDF inhibitors were considered as a possible antibacterial drug target (Figure 17). Many researchers have studied PDF and PDF inhibitors to overcome the drug resistance problem. Currently, there are 280 papers on the structure and function of bacterial PDF, and 103 papers on bacterial PDF inhibitors. Among them, LBM415 [74] and BB-83698 [75] are already under clinical study as novel antibiotics candidates.

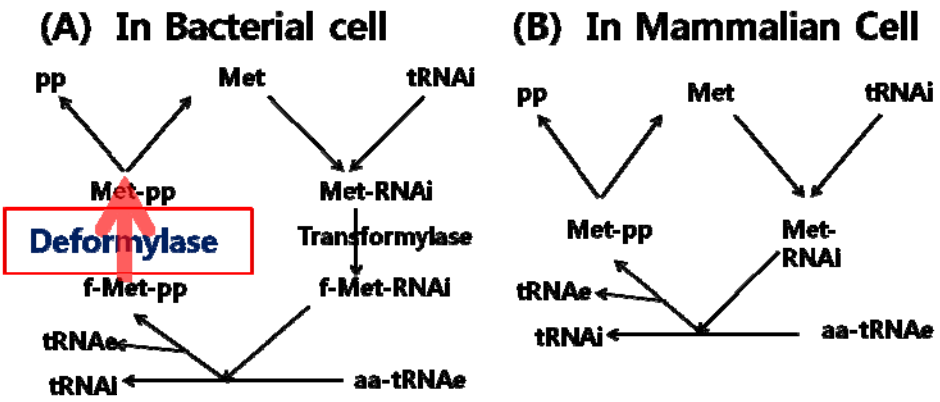


Figure 17. Peptide deformylase in bacteria cell (A) and mammalian cell (B) [76].

However, PDF also exists in the mitochondria of eukaryotes including humans, and is crucial for cell growth according to a recent paper [77, 78]. The first eukaryotic PDF was reported to be in the mitochondria and plastids of Arabidopsis thaliana [77]. A bioinformatics study found many PDF-like genes in parasites, plants, and mammals. In 2004, Lee and his co-workers reported that non-peptidyl bacterial PDF inhibitor actinonin suppresses the proliferation of cancer cells more than normal cells through a tumor-specific mitochondrial membrane depolarization and ATP depletion mechanism by interacting with HsPDF [79]. It also could remove the N-formyl capping of model human mitochondrial peptides in prokaryotes. However, the function of HsPDF *in vivo* was unclear. The crystal structure of HsPDF was determined by Sindy Escobar-Alvarez et al. and the mechanism of action was proposed (Figure 18) [80]. It is known that HsPDF processes formylated peptides derived from the mitochondrial DNA-encoded proteins by a Co^{2+} dependent manner. The bacterial PDF and HsPDF have both similarities and differences in their structure. The complex structure of HsPDF and actinonin (Figure 19), a naturally occurring inhibitor derived from the Streptomyces species, shows an overall structural similarity to the complex structure of bacterial PDF and actinonin, but they have structural differences at the C-terminus.

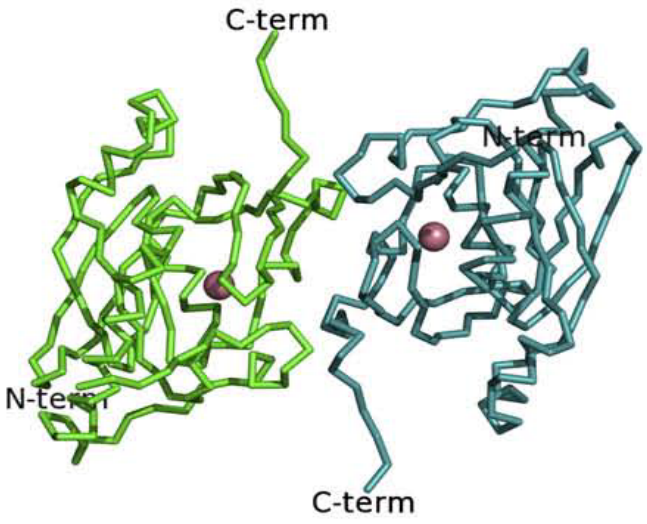


Figure 18. HsPDF crystal structure. This figure came from Journal of Molecular Biology 2009 [80].

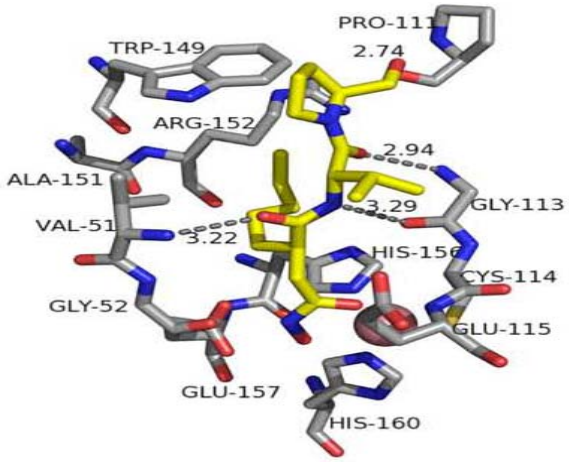


Figure 19. HsPDF crystal structure with actinonin. This figure came from Journal of Molecular Biology 2009 [80].

In the previous study, we made novel bacterial PDF inhibitors, PMT analogs, which was named after ProMediTech Ltd [9]. In the progress of developing PMT analogs, we discovered about 600 kinds of compounds in our library. PMT analogs have a potent antibacterial activity against *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* [81, 82]. These bacteria were known to cause respiratory tract-associated infections. Structural modification of PMT analogs makes these compounds more effective especially against Staphylococcus aureus [83]. PMT analogs have a peptidomimetic structure like actinonin, but have stronger activities and better pharmacokinetic properties than actinonin.

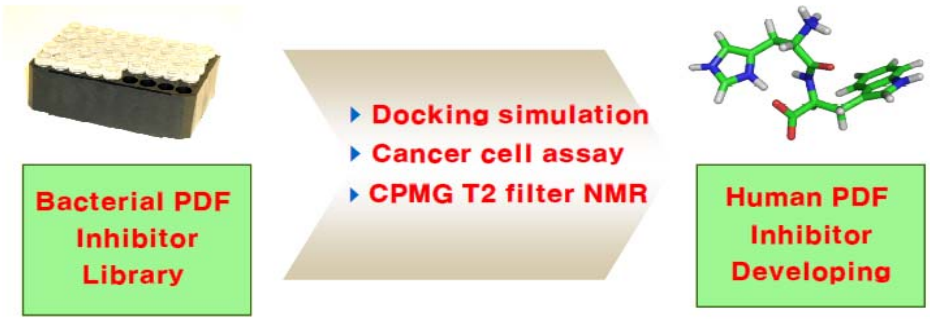


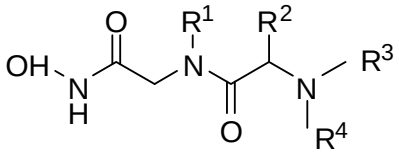
Figure 20. Development of human PDF inhibitor from bacterial PDF inhibitor library.

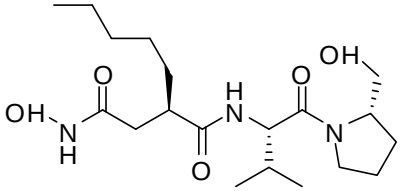
Based on our previous studies, we predicted that PMT analogs could bind to HsPDF and inhibit the proliferation of cancer cells like actinonin (Figure 20). In order to prove our hypothesis, we carried out virtual screening of PMT analogs with GOLD software [84] and DOCK6.5 from UCSF [85-88] and selected 43 compounds from our library. The MTT assay [89] was performed to examine the antitumor activity for 10 highly scored compounds out of the 43 compounds. Among the 10 compounds, 4 analogs, PMT-172, PMT-173, PMT-199, and PMT-201, showed similar or better activity compared to actinonin.

We prepared a HsPDF construct that has an enhanced expression level by codon optimization in *E.coli* [90]. Then, ^{1}H-CPMG-T2 filter NMR experiments [91] were performed with purified protein, and it was confirmed that PMT-172 and PMT-199 could bind directly to HsPDF. The docking models between PMT-172, PMT-199 and HsPDF from the X-ray crystal structure (PDB 3G5K) with DOCK6.5 were obtained. It was identified that PMT-172 and PMT-199 have an N-alkyl amide and sulfonyl amide group, respectively, at the functional group R^3 and R^4 (Table 5 and Supplemental data Table S1).

Understanding the structural characteristics of PMT-172 and PMT-199 might suggest an idea to design novel anticancer drug candidates that have better physical properties and binding affinities via modification of the functional groups.

Table 5. Compound list. (A) The main structure of PMT analogs. Each R^1, R^2, R^3, R^4 means functional group of PMT analogs. The detail structures of compounds were contained in supplemental Table 1. (B) Actinonin structure.





(A) PMT analogs main chain structure (B) Actinonin structure

	R^1	R^2	R^3	R^4
PMT- 131	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$
PMT- 132	$(CH_2)_3CH_3$	H	(C_5H_9)	$CO(C_6H_5)$
PMT- 137	$(CH_2)_3CH_3$	H	(C_6H_{11})	$COCH(CH_3)_2$
PMT- 151	$(CH_2)_3CH_3$	H	(C_3H_5)	$COCH(CH_3)_2$
PMT- 152	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$COCH(CH_3)_2$
PMT- 153	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$COCH(CH_3)_2$
PMT- 163	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-pOCH_3
PMT- 164	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-mOCH_3
PMT- 165	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$COCH_2(C_6H_5)$
PMT- 166	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_{10}H_7)$
PMT- 172	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$COCH_2CH(CH_3)_2$
PMT- 173	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$COCH_2C(CH_3)_3$
PMT- 174	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-*p*CH_3
PMT- 175	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-*m*CH_3
PMT- 176	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-*p*Cl
PMT- 177	$(CH_2)_3CH_3$	H	$CH(CH_3)_2$	$CO(C_6H_5)$-*m*Cl

PMT- 188	$(CH_2)_3CH_3$	$CH_2(C_6H_5)$	H	$SO_2(C_6H_5)$
PMT- 189	$(CH_2)_3CH_3$	$CH_2(C_6H_5)$	H	$SO_2(C_6H_5)$-*p*CH_3
PMT- 190	$(CH_2)_3CH_3$	$CH_2(C_6H_5)$	H	$SO_2(C_6H_5)$-*p*Cl
PMT- 191	$(CH_2)_3CH_3$	$CH_2(C_6H_5)$	H	$SO_2(C_6H_5)$-*p*OCH_3
PMT- 199	$(CH_2)_3CH_3$	$CH_2CH(CH_3)_2$	H	$SO_2(C_6H_5)$
PMT- 200	$(CH_2)_3CH_3$	*S*-$CH_2CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*CH_3
PMT- 201	$(CH_2)_3CH_3$	*R*-$CH_2CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*Cl
PMT- 202	$(CH_2)_3CH_3$	$CH_2CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*OCH_3
PMT- 203	$(CH_2)_3CH_3$	$CH_2CH(CH_3)_2$	H	$SO_2CH_2(C_6H_5)$
PMT- 247	$(CH_2)_3CH_3$	*S*-CH_3	H	$SO_2(C_6H_5)$-*p*Cl
PMT- 248	$(CH_2)_3CH_3$	*S*-CH_3	H	$SO_2(C_6H_5)$-*p*OCH_3
PMT- 249	$(CH_2)_3CH_3$	$CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*Cl
PMT- 250	$(CH_2)_3CH_3$	$CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*OCH_3
PMT- 251	$(CH_2)_3CH_3$	$CH_2(C_6H_5)$	H	$SO_2CH_2(C_6H_5)$
PMT- 272	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*OCH_3
PMT- 324	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	SO_2CH_3
PMT- 325	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2CH_2CH_3$
PMT- 326	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2(CH_2)_2CH_3$
PMT- 327	$(CH_2)_3CH_3$	$C(CH_3)_3$	H	$SO_2(C_6H_5)$
PMT- 328	$(CH_2)_3CH_3$	$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*CH_3
PMT- 331	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2CH_2(C_6H_5)$
PMT- 338	$(CH_2)_3CH_3$	*R*-$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*Cl
PMT- 339	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*$C(CH_3)_3$
PMT- 340	$(CH_2)_3CH_3$	*S*-$C(CH_3)_3$	H	$SO_2(C_{10}H_7)$
PMT- 557	$CH_2(C_5H_9)$	$CH(CH_3)_2$	H	$SO_2(C_6H_5)$-*p*CH_3
PMT- 567	$CH_2(C_3H_5)$	$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*CH_3
PMT- 572	$CH_2(C_4H_7)$	$C(CH_3)_3$	H	$SO_2(C_6H_5)$-*p*CH_3

2.3 Experimental procedures

2.3.1 Virtual screening

The 43 species of PMT analogs were docked into the binding pocket of HsPDF by the program GOLD [84]. The X-ray structure of HsPDF (PDB code 3G5K) was used to define the binding site for the molecular docking studies. PMT analogs were prepared in the form of a 3D structure with the MOE software [92] that immediately converts the 2D chemical structure into a 3D structure. Energy minimization was conducted for both protein and ligand with the CHARMM force field [93]. The standard set parameters were used in all the calculations. Default cutoff values of 2.5 Å for the hydrogen bonds and 4.0 Å for van der Waals (vdW) were used. To increase speed for the calculation, the genetic algorithm (GA) docking was stopped when the top three solutions were within 1.5 Å RMSD of each other. All other values were set to the default values. The scoring function GOLD fitness score, implemented in GOLD, was used to rank the tested PMT analogs.

To increase an accuracy of the protein-PMT docking analysis, another docking program, UCSF DOCK6.5 [85-88], was used with a flexible algorithm. Before the refined docking experiment, complete energy minimization was performed for the PMT analogs and HsPDF in Amber force field [46, 47]. Energy minimizations were performed using a distance-dependent dielectric constant and a cutoff of 12 Å

for nonbonded interactions through 5,000 steps in an explicit water solvent. The HsPDF structure was prepared by removing all non-protein atoms, and non-standard amino acids. Binding site residues were identified as residues with at least one atom within 5 Å of any heavy atoms of the original ligand from the initial model. Protein-derived spheres were calculated using the SPHGEN (part of the UCSF DOCK6.5 distribution) [85]. The final docking structures were generated using an anchor and grown by allowing the flexibility of the ligand method. The docking structures of the HsPDF-PMT analog complexes were ranked by DOCK scores consisting of van der Waals, Poisson-Boltzmann electrostatic, and ligand desolvation penalty terms. The APBS program [58] was used to calculate the electrostatic surface potentials of the HsPDF model. Electrostatic potential maps were calculated by numerically solving the Poisson-Boltzmann equation based on molecular mechanics. Molecule preparation and trajectory analysis were done using UCSF Chimera [57]. All the calculations were performed on an 8-node Intel Zeon CPU (192 threads) cluster system.

2.3.2 MTT assay

The cell lines were obtained from ATCC^{TM}. Various human and mouse cancer cell lines were used to measure antitumor activity. The 8 cancer cell lines used in this study were DU145 (Human prostate carcinoma (brain metastasis)), Daudi (Human B lymphoblast Burkitt lymphoma), HeLa (Human cervical adenocarcinoma), HL60 (Human acute promyelocytic leukemia), MCF-7 (breast cancer), MDA-MB (Human breast adenocarcinoma), PC3 (Human prostate carcinoma (bone metastasis)), and WI-38 (Human fibroblast (normal, lung tissue)) from a human, and one cancer cell line, NIH-3T3 (Mouse fibroblast (normal, embryo)) from a mouse. Cell lines were stored as aliquots in liquid nitrogen. To remove the DMSO and maintain the normal cell condition, after complete thawing within 2 min, cell lines were subcultured 3 times in a 25 ml T-flask. Cell lines were incubated under constant humidity at 37°C and 5% CO_2. Culture media and other additives were used as recommended by ATCC^{TM}.

Anticancer activity was measured as follows; an aliquot of 100 µl of cells (10000 cells/ml) was placed in a 96-well plate. 100 µl of PMT analogs were added to the plate at various concentrations made from serial dilutions with 4 mM stocks. After 72 hours of incubation, 40 µl of MTS solution (Promega Inc.) were added to each well of the plate. Cells were incubated for an additional 3 hours and the absorbances at 570 nm were measured. The IC_{50} was calculated with the PRISM software (GraphPad Inc.). Before adding the MTS solution, the growth status and

the number of cells were checked with an optical microscope for reliability. These results were confirmed with SRB assay (data not shown)

2.3.3 Protein preparation

The original HsPDF (orHsPDF) gene was obtained from KUGI (Korean UniGene Information) at the Korean BioInformation Center. The genetic codons of HsPDF were optimized using the web-based program, Optimizer [94], not only to eliminate the rare codons of HsPDF gene, but also to reduce its high GC content. Possible secondary structures of the codon-optimized HsPDF (coHsPDF) gene were evaluated using the web program (http://www.genebee.msu.su/services/rna2_reduced.html), and further modification of the genetic codes was performed in order to reduce the stability of possible local secondary structures (higher than -15.0 kcal/mol). The codon-optimized hsPDF (coHsPDF) gene (residues E40–D243) was synthesized by EpochBioLabs, USA, and the sequences of the orHsPDF and coHsPDF genes are shown in Figure 21.

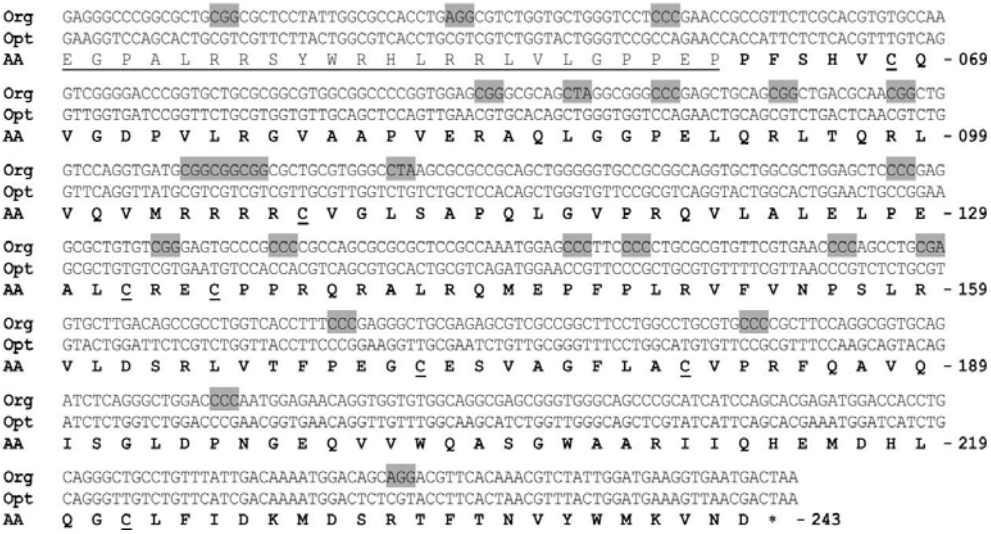


Figure 21. The sequences of original and codon-optimized HsPDF gene (amino acid residues 40–243). The residues 40–62 of HsPDF (underlined) were not used for the expression of HsPDF protein, and the residues 63–243 of HsPDF (bold character) were expressed in the context of Trx-fusion protein. The rare codons of E. *coli* are marked with a gray box.

The orHsPDF and coHsPDF gene (amino acid residues 63–243) were subcloned into the modified pET32a vector (NcoI/XhoI), in which the original enterokinase protease site was replaced by a tobacco etch virus (TEV) protease site [32]. For subcloning into the pET32a vector, an additional Gly residue was incorporated to match the coding frame during subcloning using NcoI. The orHsPDF gene has extremely high GC content (80–90%) in its first half, and thus its PCR amplification was performed in the presence of 5% DMSO.

The orHsPDF and coHsPDF was expressed in the BL21(DE3) strain (Novagen Inc., USA), by growing *E. coli* in 1 L of LB media containing 100 mg/ml Amp. Cells were grown at 37℃ until the OD_{600} reached 0.5–0.6. Protein expression was induced by adding 1.0 mM IPTG and the cells were grown at 16°C for 20 hours in the presence of 50 μM of $CoCl_2$. Cells were harvested and lysed by sonication on ice. Ni Sepharose™ high performance affinity column (GE Healthcare Inc., USA) was used for protein purification. Proteins were eluted with 100 mM imidazole. For the NMR experiment, purified coHsPDF was stored in deuterated buffer [pH 6.0 10 mM MES (d13), 10 mM Bis-tris (d18), 100 mM NaCl, 1 mM TCEP]. Buffer exchange was performed with a PD-10 desalting column (GE Healthcare Inc., USA).

2.3.4 NMR experiment

All NMR measurements were carried out 32 scans at 303K. Actinonin was used as a reference for the NMR binding assay. 10 uM of purified protein and 5 mM of compounds were used for NMR experiments. The proton NMR spectra of the PMT analogs were measured with a 1D reference pulse and CPMG-T2 filter pulse [91]. When compared to the spectrum of the 1D reference, the spectrum acquired by 50 ms of the CPMG-T2 filter pulse showed reduced signals of the compounds that interact with the purified HsPDF. The 0.222 mM of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used as internal reference to estimate the solubility of the PMT analogs. Since it was difficult to obtain highly concentrated protein due to its low solubility, all NMR measurements were carried out with a 900 MHz machine with a cryogenic probe at the Korea Basic Science Institute (KBSI) for high sensitivity.

2.3.5 Chemistry

The synthesis of PMT compounds used in this study is described in the supplemental data (S2).

2.4 Result and Discussion

2.4.1 Initial screen

Screening millions of compounds and synthesizing multiple number of compounds for initial hits are costly and time consuming in the initial stage of drug discovery [95]. Using a previously known template as a starting point and synthesizing derivatives using structural information might be a good strategy [96]. In our study, virtual screening was performed with the GOLD and DOCK6.5 software to select compounds that could bind to HsPDF. Fourty-three species of the PMT analogs, as bacterial PDF inhibitor were obtained from our compounds library [81-83].

Initially, the GOLD scoring fitness function [97] was used to predict the inhibitor-binding site and affinity. The ten PMT analogs with high scores were chosen by the GOLD scoring fitness function to select candidates for HsPDF inhibitors (Table 6). The selected PMT analogs had significant GOLD fitness scores over 40 [98-100].

Table 6. Docking scores of HsPDF-PMT models. The Fitness of GOLD scores is made up of four components including external H-bond (hydrogen bond energy), external vdw (van der Waals energy), internal vdw, and internal torsion of GOLD [84]. Grid score and Grid VDW (van der Waals energy) are based on the non-bonded terms of the molecular mechanic force field of DOCK6.5 [85-88].

Compound No.	GOLD Scores			Dock Scores		
	Fitness	External H bond	Internal VDW Weighted	Grid Score	Grid VDW	No. of H bond
164	37.1364	24.6773	-34.0868	-42.0775	-38.3242	3
165	62.2395	20.7618	-10.4896	-41.5432	-41.4142	0
172	60.8347	25.7452	-7.6311	-39.8787	-39.4525	0
173	61.1030	19.5151	-10.3763	-37.4734	-35.9974	0
177	60.6741	23.5356	-8.5385	-44.1043	-40.6854	1
199	43.2651	6.0401	-10.4704	-49.0813	-47.0280	2
200	54.7528	12.9037	-2.9260	-41.4464	-40.6003	1
201	41.7651	0.1267	-3.5470	-45.1018	-41.9975	1
202	60.3487	25.8973	-12.9692	-43.1834	-40.7711	1
203	64.8043	25.6610	-14.1875	-42.7174	-41.1963	1

2.4.2 MTT assay

The anticancer activity of the PMT analogs selected from GOLD was verified using actinonin as a control. Various cell lines including DU145 (Human prostate carcinoma (brain metastasis)), Daudi (Human B lymphoblast Burkitt lymphoma), HeLa (Human cervical adenocarcinoma), HL60 (Human acute promyelocytic leukemia), MCF-7 (breast cancer), MDA-MB (Human breast adenocarcinoma), PC3 (Human prostate carcinoma (bone metastasis)), WI-38 (Human fibroblast (normal, lung tissue)), and NIH-3T3 (Mouse fibroblast (normal, embryo)) were used to measure the IC_{50} with the MTT assay [89]. PMT-172, PMT-173, PMT-199, and PMT-201 had a similar or superior activity for all the cell lines compared to actinonin. However, it has not been shown whether the compounds with antitumor activity bind to HsPDF.

It was shown from the MTT assay that four species of the analogs, PMT-172, PMT-173, PMT-199, and PMT-201 had outstanding anticancer activity (Table 7). PMT-199 and PMT-201 had a rather lower score in the GOLD score fitness system compared to the rest of the selected compounds. While the docking simulation is a powerful technique for screening millions of compounds to find leads, it sometimes shows a poor correlation between the results of the scoring function and the results of the *in vitro* assay.

PMT-173 with a high score in the GOLD score fitness system (61.103) had good activity. However, PMT-199 and PMT-201 with a score of about 40 for the GOLD

score fitness (43.2651 and 41.7651, respectively, Table 6) showed excellent performance in the cell lines like the other selected compounds. In recent studies, target compounds with over a 40 score in the GOLD score fitness algorithm were frequently chosen as the lead [101] [19] .

Table 7. *in vitro* MTT assay.

	MTT Assay IC50 (μM)								
Compounds	**Du145**	**Daudi**	**HELA**	**HL60**	**MCF-7**	**MDA MB**	**PC-3**	**WI-38**	**NIH-373**
Actinonin	79.67	202.9	35.51	178.7	106.8	185.8	180.2	207.7	40.18
PMT-164	234.1	N.D	183.3	N.D	223.5	0.1301	219.7	191.3	192.9
PMT-165	29878	418.2	126.2	297.1	351.3	5711	233.7	147.8	N.D
PMT-172	162.5	185	130.6	146.3	192.1	194.7	76.7	37.41	N.D
PMT-173	159.1	176.5	134.4	144.1	78.16	173.6	112	31.37	N.D
PMT-177	173	201.7	165.2	195.7	111.3	197.9	93.86	31.63	N.D
PMT-199	84.79	487.7	97.64	190.6	79.93	171.4	108.4	23.81	N.D
PMT-200	129.25	N.D	N.D	N.D	32.77	41.66	N.D	68.36	N.D
PMT-201	30.91	N.D	N.D	N.D	29.09	31.44	N.D	47.55	N.D
PMT-202	43.06	N.D	N.D	N.D	45.11	99.19	N.D	44.06	N.D
PMT-203	49.77	N.D	N.D	N.D	48.29	70.45	N.D	57.86	N.D

(N.D = not detected)

2.4.3 Various docking simulation

Since docking simulation varies the results depending on the algorithm, using multiple programs rather than one program could be more accurate in analyzing the result. The same PMT analogs with a good docking score, each from GOLD and DOCK6.5, were commonly derived, but some PMT analogs were different. This phenomenon explains that PMT analogs with a bad docking score from a single algorithm showed good activity in the MTT assay. In the case of docking with PMT-172 and PMT-173, these compounds showed high scores (60.8347 and 61.1030, respectively, Table 6) of docking in the GOLD methods but not so high scores of docking in the DOCK6.5 algorithm (-39.8787 and -37.4734, respectively, in Table 6) with no hydrogen bonds (Figure 22 (A) and (B)). From the MTT assay results, PMT-172 and PMT-173 had good anticancer activity (Table 7) suggesting that the docking scores from the GOLD method may suitable for selecting these compounds. However, there were one or two hydrogen bonds in the docking structures of the PMT-199 and PMT-201 compounds with the DOCK6.5 algorithm, and the existence of hydrogen bonds could explain the good scores with the DOCK method (-49.0813 and -41.4464, respectively, Table 6) and the anticancer activity from the MTT assay (Table 7). Thus, in the case of docking with PMT-199 and PMT-201, the DOCK6.5 algorithm may more appropriate for selecting these compounds.

The Grid score and the Grid VDW of actinonin were -44.2244 and -40.0520,

respectively. The number of the hydrogen bond of actinonin is two. The lowest and highest value of Grid score of PMT analogs were -49.0813 and -37.4734 and the lowest and highest value of PMT Grid VDW of PMT analogs were -47.0280 and -35.9974. The docking score of actinonin is somewhere in the ranges of PMT compounds.

2.4.4 Structure of selected compounds

All of the PMT analogs have the same structure in the main chain. Like actinonin, the PMT analogs are hydroxamic acid with a peptidomimetic structure (Table 5 and Supplementary data). However, there are some differences between actinonin and the PMT analogs. Actinonin has a carbon at the position 3' of the main chain with a pentyl functional group, whereas the PMT analogs have nitrogen with a butyl- or cycloalkyl- group. In addition, atoms in the 5' and 6' positions of the main chain between actinonin and the PMT analogs are different. The PMT analogs have broadly four types of functional groups shown in Table 5. The ten compounds that were selected with GOLD and DOCK6.5 have a characteristic butyl group at the R^1 position. It is surmised that this R^1 group may play an essential role in binding to HsPDF. The other positions, R^2, R^3, and R^4 have various functional groups.

2.4.5 Molecular modeling

In the binding of the ligand to the protein, hydrogen bonding is a crucial factor due to its strong binding force rather than other interactions in making the typical salt bridge. In the docking structure with PMT-199, Val49 and Trp147 residues in the active site could make 2 hydrogen bonds with a compound (PMT-199-O3···H-Val49, PMT-199-O4···$H^{\varepsilon 1}$-Trp147, Figure 22 (C)) and Pro109 residue could be involved in making a hydrogen bond with PMT-201 (PMT-201-H6···$O^{\varepsilon 2}$-Glu113, Figure 22 (D)). Accordingly, it is believed that taking into consideration the number of hydrogen bonds and multiple methods of docking may lead to more accurate results of docking with the protein-PMT compound.

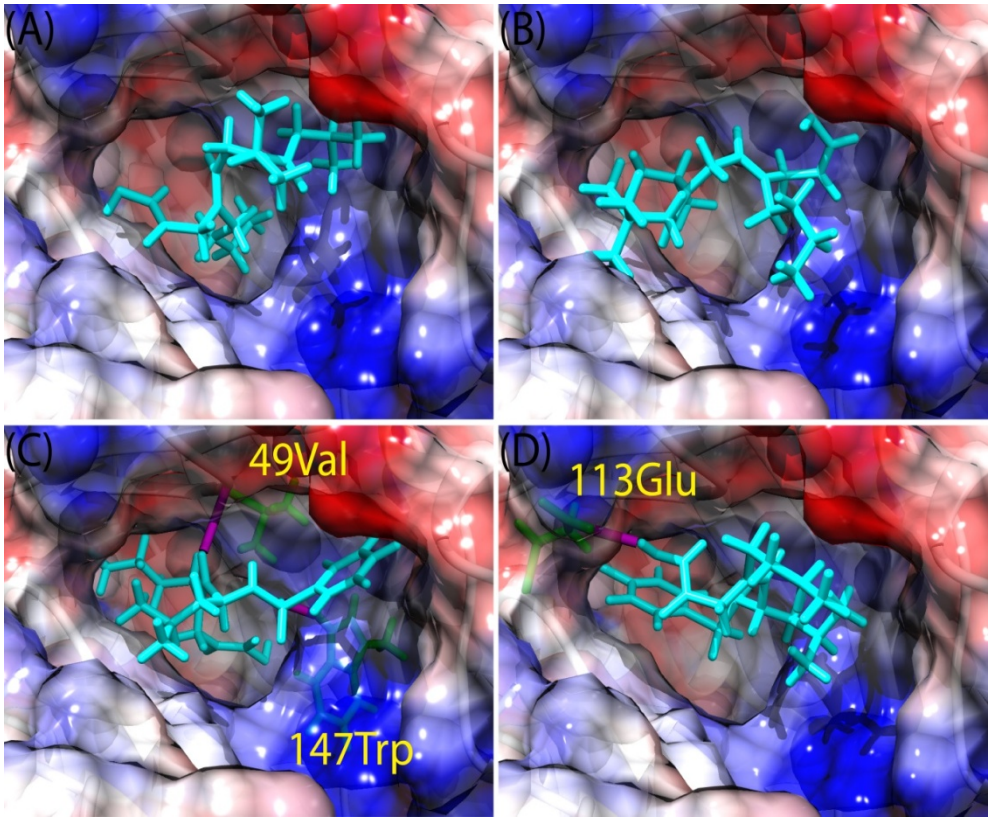


Figure 22. Molecular docking structures of HsPDF-PMT complexes. All PMT compounds in the binding pocket of HsPDF are presented with cyan line-stick model including PMT-172 (A), PMT-173 (B), PMT-199 (C), and PMT-201 (D), respectively. Electrostatic surface potential of HsPDF was colored in blue (positively charged region) and red (negatively charged region) with 10 scaled energy unit. Hydrogen bonds of PMT-199 and PMT-201 were shown in magenta line.

2.4.6 Codon optimization and protein purification

Drug development generally requires a large amount of target protein. Moreover, the NMR screening of drug candidates often requires ^{13}C- and ^{15}N-labeled protein, and higher protein expression has great advantages for obtaining isotope-labeled protein. We found that IPTG induction performed at a lower temperature results in increased expression of soluble Trx-HsPDF proteins. An alternative method to increase the amount of soluble HsPDF protein is to increase the amount of protein expression. The orHsPDF gene has extremely high GC content in its first half of sequence (Figure 23). However, the optimal GC content for gene expression in *E. coli* is within 30–70% [102].

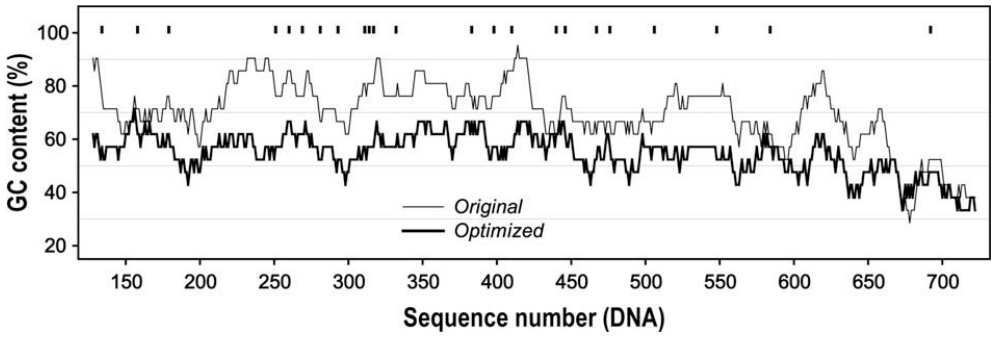


Figure 23. The GC content of original (thin line) and codon-optimized (thick line) HsPDF genes. The GC content was calculated using the window of nine base pairs. The positions of rare codons in the original sequence are indicated by a vertical bar on the top of the figure.

It was recently reported that the elimination of rare codons does not increase the expression level of the Rv3297 gene (*Mycobacterium tuberculosis*) in *E. coli*, and that the presence of a stable secondary structure in the translation initiation region (TIR) inhibits expression [103]. Although a stable secondary structure could also be present in the TIR of the orHsPDF gene, the presence of the N-terminal tagged protein, such as Trx-tag, could reduce possible presence of stable secondary structure in the TIR. The presence of N-terminal tagged proteins that are expressed well alone likely prevents the formation of stable secondary structures in the TIR. We found that the orHsPDF gene alone contains many stable stem structures, which are likely due to the high GC content (Figure 24). We assumed that these stable stem structures might lower the expression of HsPDF proteins, as well as cause expression of the truncated forms and Trx alone in the purified fraction of Trx-orHsPDF. Therefore, we performed codon optimization of the HsPDF gene and reduced the GC content in order to eliminate these stem structures as well as the presence of rare codons (Figure 21 and 23). The whole HsPDF gene, except for the mitochondrial targeting sequence (amino acids residues 40–243), was optimized first, and then part of the coHsPDF gene (residues 63–243) was subcloned into the Trx-tagged expression vector (Trx-coHsPDF). Overall free energy of the secondary structures of the codon-optimized HsPDF (coHsPDF) gene was greatly reduced compared to that of the orHsPDF gene (-94.8 vs. -185.5 kcal/mol), and the stable stem structure was almost eliminated in the coHsPDF gene (Figure 24). Indeed, Trx-tagged coHsPDF (63–243) were expressed well

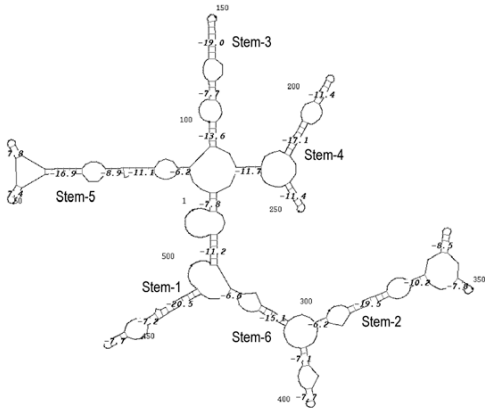
following IPTG induction in BL21(DE3) *E. coli* cells. The expression of Trx-coHsPDF protein was greatly increased compared to that of Trx-orHsPDF protein (Figure 25A and Figure 25B). The amount of purified Trx-coHsPDF was more than 14 mg per 1-L LB culture.

A

orhPDF (63-243)

-185.5 kcal/mol

Stem-1 (-20.5)	Stem-2 (-19.5)
438-GGCAGCCC-445	306-CCGCCTGG-313
484-CCGTCGGG-477	379-GGCGGACC-372
Stem-3 (-19.0)	Stem-4 (-17.1)
132-GCGGCGC-138	187-GCGCTGG-193
160-CGCCGCG-154	237-CGCGACC-231
Stem-5 (-16.9)	Stem-6 (-15.1)
39-GCTGCGC-45	287-GCCTGC-292
76-CGACGCG-70	428-CGGACG-423



B

cohPDF (63-243)

-94.8 kcal/mol

No stem structure
more stable than -15 kcal/mol

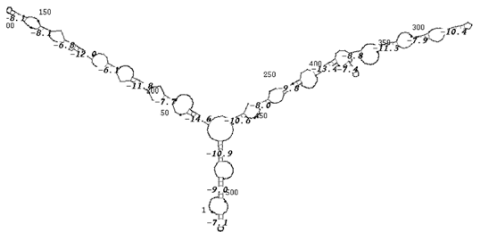


Figure 24. Estimated stabilities of the overall secondary structure. The possible stem structures of the original HsPDF (orHsPDF) gene (A) and the codon-optimized HsPDF (coHsPDF) gene (B) were estimated. The predicted stem regions of which energy is lower than -15 kcal/mol are indicated.

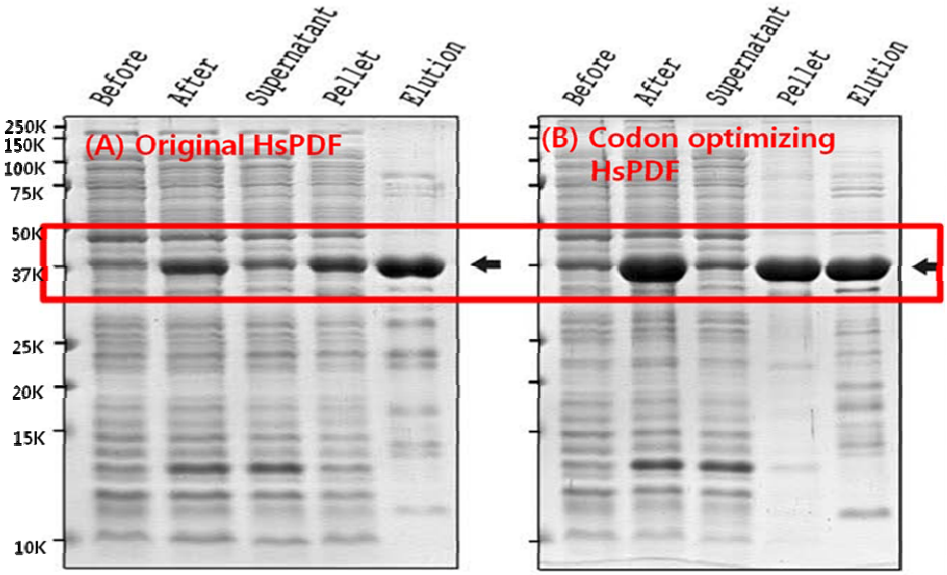


Figure 25. Expression and purification of HsPDF (63–243) with the N-terminal Trx-fusion protein. The expression of the Trx-tagged codon optimizing HsPDF (B) was greatly increased compared to that of Trx-tagged original HsPDF (A).

2.4.7 NMR binding experiment

CPMG-T2 filter NMR experiment was used for the binding test of actinonin which has a micro-molar binding constant to HsPDF. Actinonin, which was used as a control, was verified to bind to HsPDF by CPMG-T2 filter pulse (Figure 26). The intensity of the two pseudo-amide protons of actinonin was reduced by a decrease in the T2-relaxation time induced by the CPMG-T2 filter pulse compared to the 1D reference pulse.

In the process of NMR sample preparation, there were some aggregation after dissolving compounds to NMR sample buffer due to the insolubility of compounds. We could not get exact concentration of compounds in NMR sample, DSS (0.222mM) was added to each samples as an internal reference to indirectly measure the concentration of the compounds. The concentration of actinonin was estimated to be about 1 mg/ml compared to the peak intensity of the DSS. By applying the same method, PMT-172, PMT-173, PMT-199, and PMT-201 were estimated to be about 0.82 mg/ml, 0.03 mg/ml, 0.11 mg/ml, and 0.02 mg/ml, respectively (data now shown).

Thus, CPMG-T2 filter NMR binding experiments were carried out and the binding pattern of the compounds were characterized. PMT-172 had a similar binding constant to actinonin. PMT-173 and PMT-201 could not be confirmed to interact with HsPDF because they showed very low signal sensitivity due to their poor solubility. In the case of PMT-199, it seemed to bind with HsPDF.

The interaction of the PMT analogs was confirmed by observing the peaks of the compounds bound to Trx-HsPDF. When compounds bind to high molecular weight proteins, the T2-relaxation time is decreased due to an increase in the apparent molecular weight of the compounds. Comparing the spectrum of the 1D reference pulse zgesgp to one of the CPMG-T2 filter pulse zgesgp [91], the peak intensity of the bound protons from the compounds was decreased in the CPMG-T2 filter NMR spectra. Since the exchange rate between the bound and free PMT analogs was low, the amplification effect from the decrease in the T2-relaxation time induced by protein binding might be small. Therefore, complementary experiments such as enzyme assays are required for an accurate analysis about the nature of the interaction. Further modification of the functional group is needed even if PMT-199 binds to the HsPDF strongly since solubility of compounds is an important factor in drug development.

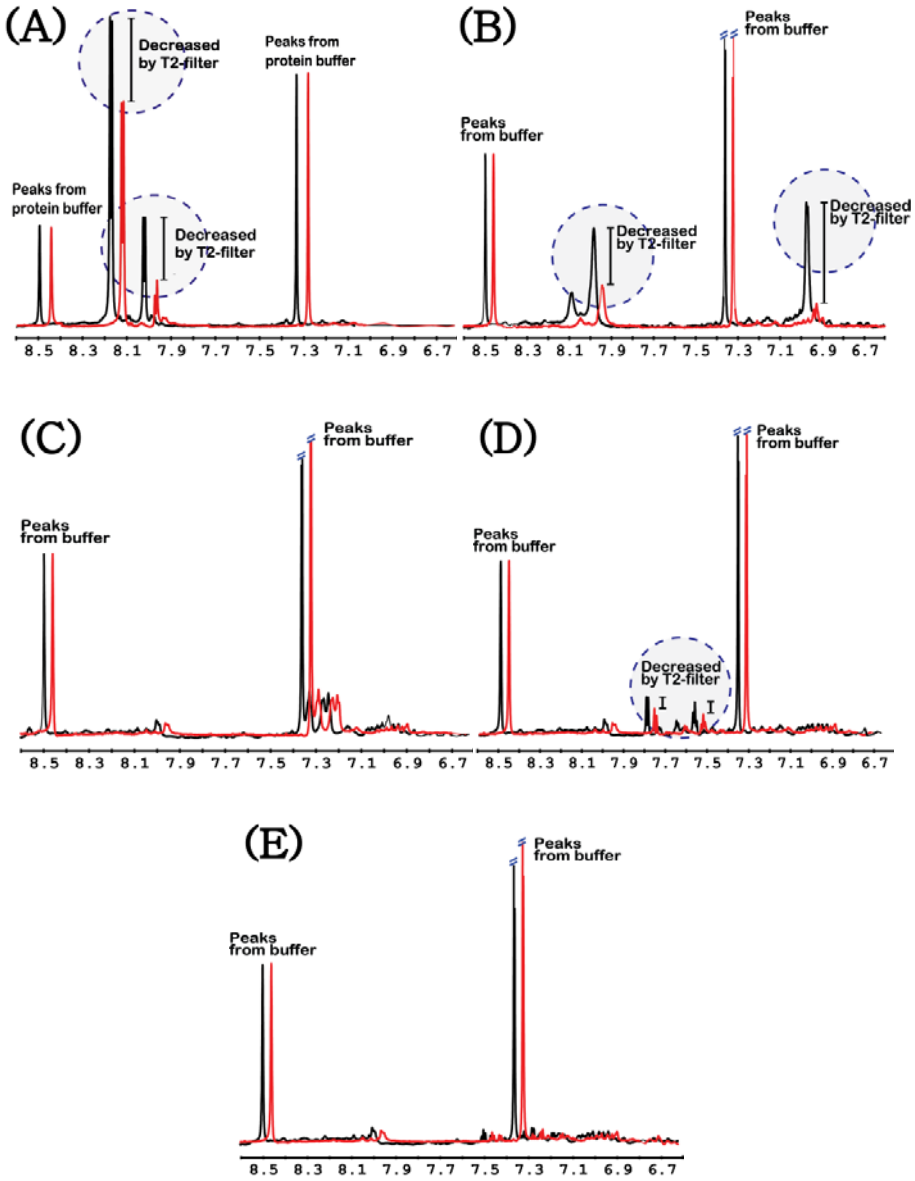


Figure 26. CPMG-T2 NMR binding assay. Antinonin (A), PMT-172 (B), PMT-173 (C), PMT-199 (D), and PMT-201 (E). The black spectrum was from 1D reference zgesgp pulse and the red spectrum was from the CPMG-T2 filter zgesgp pulse. The dashed blue circle means differences of intensities with the CPMG-T2 filter experiment.

2.5 Conclusion

The inhibitors of HsPDF were screened using PMT analogs, bacterial PDF inhibitors made from previous study. Various docking programs make screening more accurate and reliable. The antitumor effect of the selected compounds were checked with the MTT assay and four species (PMT-172, PMT-173, PMT-199, and PMT-201) of PMT analogs had excellent activities. NMR binding experiment showed that two species of them directly interact with HsPDF and these analogs could be helpful in developing novel anticancer drugs.

Summary

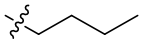
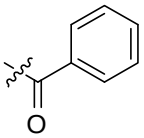
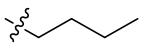
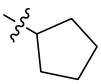
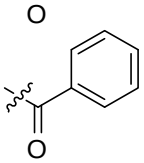
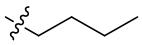
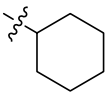
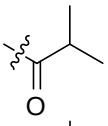
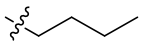
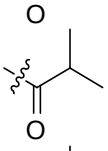
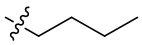
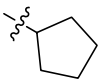
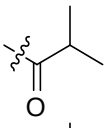
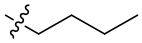
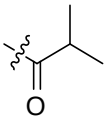
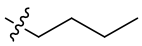
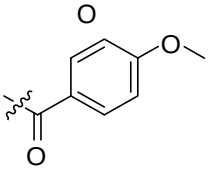
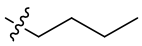
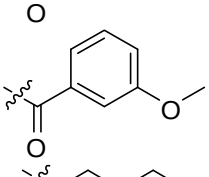
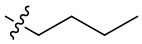
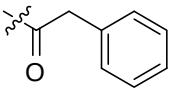
In this study, the general scheme of structure-based drug design (SBDD) was divided roughly into two parts and two part of process were accomplished with two different sample. The first part was successfully conducted with Rv0569 protein sample from *Mycobacterium tuberculosis.*

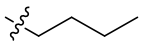
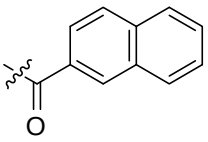
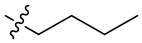
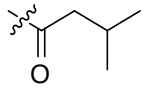
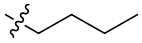
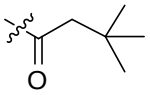
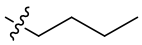
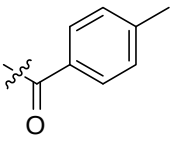
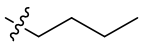
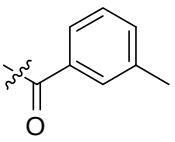
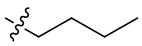
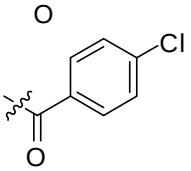
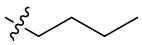
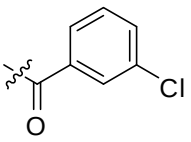
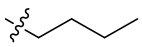
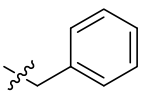
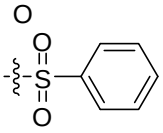
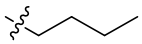
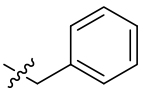
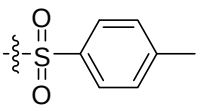
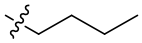
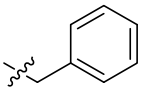
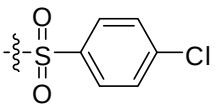
The structure of Rv0569 was determined with high quality using NMR, and based on structure, previous studies, DALI server and BLAST search, we suggest that Rv0569 is a potent hypoxic signal transduction protein. The second part was carried out with HsPDF protein sample from human. We found the lead compound of Human PDF inhibitors from screening the bacterial PDF inhibitors library. Further studies are required for both projects, and these studies are expected to help to develop anti-tuberculosis drug and anti-cancer drug, respectively.

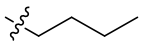
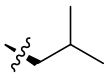
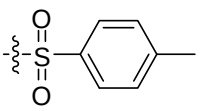
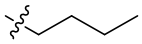
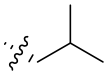
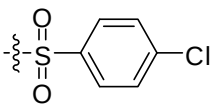
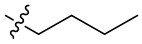
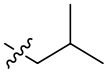
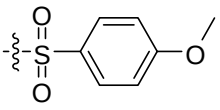
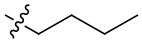
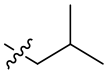
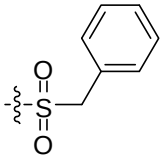
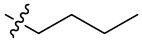
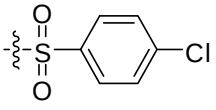
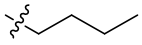
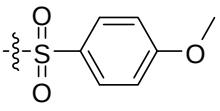
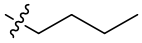
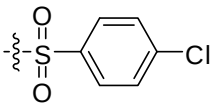
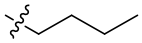
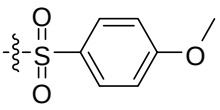
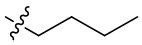
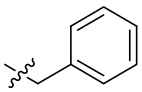
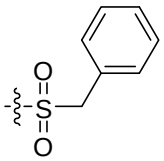
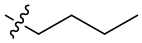
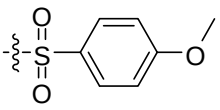
Supplemental data

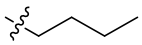
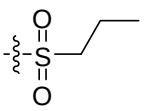
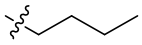
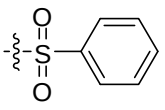
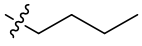
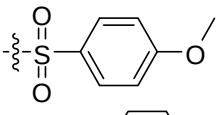
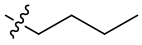
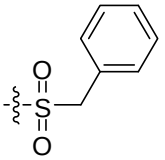
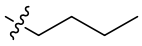
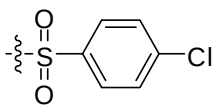
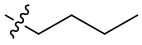
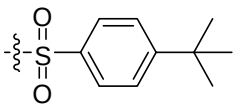
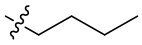
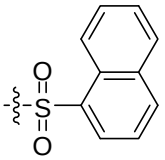
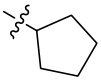
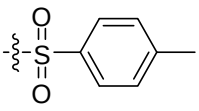
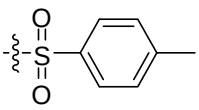
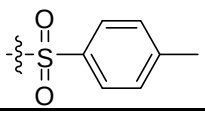
S1. PMT Compound structures

Table S1. List of compound structures.

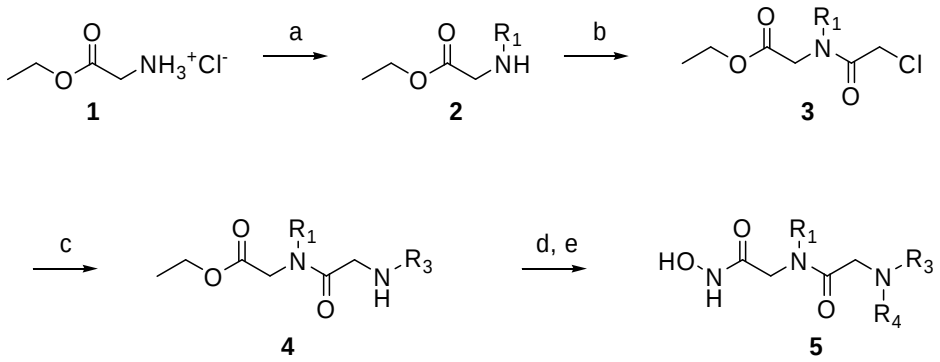
	R^1	R^2	R^3	R^4
PMT-131				
PMT-132				
PMT-137				
PMT-151				
PMT-152				
PMT-153				
PMT-163				
PMT-164				
PMT-165				

PMT-166				
PMT-172				
PMT-173				
PMT-174				
PMT-175				
PMT-176				
PMT-177				
PMT-188				
PMT-189				
PMT-190				
PMT-191				
PMT-199				

PMT-200				
PMT-201				
PMT-202				
PMT-203				
PMT-247				
PMT-248				
PMT-249				
PMT-250				
PMT-251				
PMT-272				
PMT-324				
PMT-325				

PMT-326				
PMT-327				
PMT-328				
PMT-331				
PMT-338				
PMT-339				
PMT-340				
PMT-557				
PMT-567				
PMT-572				

S2. Chemistry of PMT compounds



Scheme 1. Reagents and conditions: a) R_1OTs, $NaHCO_3$, acetonitrile, 60 °C, 18 h; b) chloroacetyl chloride, Et_3N, dichloromethane, rt, 1 h; c) R_3NH_2, Et_3N, acetonitrile, reflux, 18 h; d) R_4COCl, triethylamine, dichloromethane; e) hydroxylamine, MeOH, rt, 30 min.

Compound 2

An acetonitrile (15 V/Wt) suspension of glycine ethyl ester hydrochloride (5 eq.), alkyl (R^1) tosylate (1 eq.), and sodium bicarbonate (10 eq.) was heated at 60 °C and then stirred for 18 h. The precipitate was removed by filtration of the reaction suspension and the filtrate was concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 1) to give **2.**

Compound 3

Chloroacetyl chloride (1.5 eq.) and triethylamine (1.5 eq.) were added to a dichloromethane (10 V/Wt) solution of compound **2** (1 eq.). The reaction mixture

was stirred at room temperature for 1 h and washed with distilled water and a solution of 1 N aq. HCl. The organic layer was dried on anhydrous sodium sulfate and concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 5) to give **3.**

Compound 4

Alkyl (R^3) amine (1.5 eq.) and triethylamine (1.5 eq.) were added to a solution of compound **3** (1 eq.) in acetonitrile (10 V/Wt). The reaction mixture was refluxed for 18 h and concentrated under vacuum. The resulting residue was dissolved in EtOAc and washed with distilled water. The organic layer was dried on anhydrous sodium sulfate and concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 3) to give **4.**

Compound 5

Triethylamine (1.2 eq.) and acyl (R^4) chloride (1 eq.) were added to a dichloromethane (10 V/Wt) solution of compound **4** (1 eq.). The reaction mixture was stirred at room temperature for 18 h and then washed with water. The organic layer was dried on anhydrous sodium sulfate and concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 3) and then the purified ethyl ester was treated with NH_2OH solution in methanol to give **5**.

The following PMT compounds were synthesized by **scheme 1.**

Analytical data

PMT-131

^{1}H-NMR ($CDCl_3$, 300 MHz) : 83-1.01 (m, 3H), 1.11-1.67 (m, 8H), 3.51 (m, 2H), 3.97 (m, 2H), 4.16 (m, 3H), 7.42 (s, 5H).

PMT-132

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.98 (m, 3H), 1.25-1.62 (m, 10H), 1.88 (m, 2H), 3.52 (m, 2H), 3.97 (s, 2H), 4.17 (m, 3H), 7.42 (m, 5H).

PMT-137

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.9 (m, 3H), 1.14 (d, J=6.6Hz, 6H), 1.30 (m, 6H), 1.60 (m, 3H), 1.88 (m, 4H), 2.83 m, 1H), 3.44 (m, 4H), 3.82 (s, 2H), 4.08 (s, 2H).

PMT-151

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.88-0.97 (m, 7H), 1.14 (d, 6H), 1.25-1.60 (m, 4H), 2.99 (m, 1H), 3.39 (m, 3H), 4.05 (m, 4H).

PMT-152

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.96 (m, 3H), 1.15 (d, 6H), 1.25-2.00 (m, 12H), 2.92 (m, 1H), 3.48 (m, 2H), 3.77 (s, 2H), 4.11 (s, 2H), 4.39 (m, 2H).

PMT-153

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.96 (m, 3H), 1.15 (d, 6H), 1.24-1.61 (m, 10H), 2.87 (m, 1H), 3.49 (m, 2H), 3.76 (s, 2H), 4.11 (s, 2H),4.27 (m, 2H).

PMT-163

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.97 (m, 3H), 1.21-1.61 (m, 10H), 2.50 (m, 2H),

3.83 (s, 3H), 3.95-4.20 (m, 5H), 6.92 (d, 2H), 7.49 (d, 2H).

PMT-164

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.94 (m, 3H), 1.16-1.64 (m, 10H), 3.45 (m, 2H), 3.81 (s, 3H), 4.02-4.16 (m, 5H), 6.96-7.33 (m, 4H).

PMT-165

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.94 (m, 3H), 1.08 (m, 6H), 1.26-1.60 (m, 4H), 3.44 (m, 2H), 3.80-4.19 (m, 7H), 7.23- 7.35 (m, 5H).

PMT-166

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.87-1.10 (m, 9H), 1.26-1.68 (m, 4H), 3.56-4.32 (m, 7H), 7.40-8.10 (m, 7H).

PMT-172

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.93 (m, 9H), 1.21-1.62 (m, 10H), 2.16-2.29 (m, 3H), 3.46 (m, 2H), 3.79-4.28 (m, 5H).

PMT-173

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.98 (m, 3H), 1.06 (s, 9H), 1.21 (m, 6H), 1.33-1.61 (m, 4H), 2.31 (s, 2H), 3.43 (m, 2H), 3.77 (s, 2H), 4.12 (s, 2H), 4.28 (m, 1H).

PMT-174

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.96 (m, 3H), 1.17 (m, 6H), 1.23-1.64 (m, 4H), 2.37 (s, 3H), 3.49 (m, 2H), 3.97-4.16 (m, 5H), 7.19-7.35 (m, 4H).

PMT-175

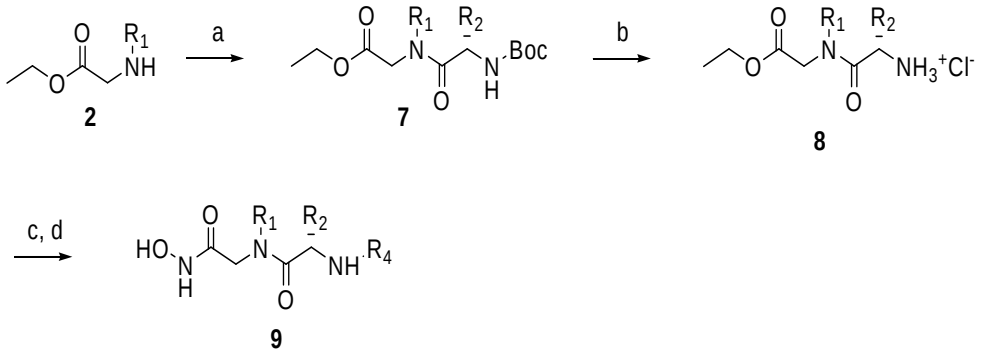
^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.95 (m, 3H), 1.17 (m, 6H), 1.26-1.65 (m, 4H), 2.37 (s, 3H), 3.47 (m, 2H), 3.98-4.12 (m, 5H), 7.19-7.31 (m, 4H).

PMT-176

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.97 (m, 3H), 1.17 (m, 6H), 1.25-1.63(m, 4H), 3.46 (m, 2H), 3.97-4.13 (m, 5H), 7.38 (m, 4H).

PMT-177

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.97 (m, 3H), 1.17 (m, 6H), 1.23-1.65 (m, 4H), 3.47 (m, 2H), 4.03-4.22 (m, 5H), 7.32-7.40 (m, 4H).



Scheme 2. Reagents and conditions: a) *N*-Boc-L-amino acid (introduction of R_2), HOBt, EDC, diethylamine, dichloromethane, rt, 1 h; b) 4 M HCl in 1,4-dioxane, rt, 18 h; c) R_4Cl, triethylamine, dichloromethane; e) hydroxylamine, MeOH, rt, 30 min.

Compound 7

1-Hydroxybenzotriazole (1.5 eq.), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (1.5 eq.), *N*,*N*-diisopropylethylamine (2.5 eq.), and *N*-Boc-L-amino acid (R^2) (1 eq.) were added to a dichloromethane (10 V/Wt) solution of compound **2** (1 eq.). The reaction mixture was stirred at room temperature for 18 h and washed with distilled water and a solution of 1 N aq. HCl. The organic layer was dried on anhydrous sodium sulfate and concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 5) to give **7.**

Compound 8

A 1,4-dioxane solution of 4 N HCl was added to compound **7** and the reaction mixture was stirred at room temperature for 18 h and concentrated under vacuum.

The resultant concentrate was used in the next step without further purification.

Compound 9

Triethylamine (2.2 eq.) and sulfonyl (R^4) chloride (1 eq.) were added to a dichloromethane (10 V/Wt) solution of compound **8** (1 eq.). The reaction mixture was stirred at room temperature for 18 h and then washed with water. The organic layer was dried on anhydrous sodium sulfate and concentrated under vacuum. The resultant concentrate was purified by silica gel chromatography (EtOAc : n-Hex = 1 : 3) and then the purified ethyl ester was treated with NH_2OH solution in methanol to give **9.**

The following PMT compounds were synthesized by **scheme 2.**

Analytical data

PMT-188

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.79 (m, 3H), 1.06 (m, 2H), 1.25 (m, 2H), 2.82 - 3.07 (m, 4H), 3.64 (t, J=14.2 Hz, 1H), 3.89 (d, J=14.8Hz, 1H), 4.32 (m, 1H), 7.06 (m, 2H), 7.25 (m, 3H), 7.48 (m, 3H), 7.77 (m, 2H).

PMT-189

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.79 (m, 3H), 1.06 (m, 2H), 1.25 (m, 2H), 2.37 (s, 3H), 2.82 - 3.07 (m, 4H), 3.64 (t, J=14.2 Hz, 1H), 3.89 (d, J=14.8Hz, 1H), 4.32 (m, 1H), 7.01 (m, 2H), 7.24 (m, 5H), 7.64 (m, 2H).

PMT-190

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.81 (m, 3H), 1.12 (m, 2H), 1.28 (m, 2H), 2.82 - 3.07 (m, 4H), 3.64 (t, J=14.2 Hz, 1H), 3.89 (d, J=14.8Hz, 1H), 4.32 (m, 1H), 7.05 (m, 2H), 7.26 (m, 3H), 7.38 (m, 2H), 7.64 (m, 2H).

PMT-191

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.79 (m, 3H), 1.06 (m, 2H), 1.25 (m, 2H), 2.82 - 3.09 (m, 4H), 3.64 (d, J=14.2 Hz, 1H), 3.84 (s, 3H), 3.94 (d, J=14.8Hz, 1H), 4.27 (m, 1H), 6.89 (m, 2H), 7.08 (m, 2H), 7.25 (m, 3H), 7.69 (m, 2H).

PMT-199

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.76-0.92 (m, 9H), 1.19-1.28 (m, 4H), 1.42-1.52 (m, 3H), 3.19-3.26 (m, 2H), 3.78(s, 1.5H), 4.07-4.16 (m, 1.5H), 6.66 (bs, 0.7H), 7.49-7.59 (m, 3H), 7.56-7.88 (d 2H), 9.69(bs, 0.5H).

PMT-200

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.76-0.92 (m, 9H), 1.19-1.28 (m, 4H), 1.42-1.52 (m, 3H), 2.40 (s, 3H), 3.19-3.24 (m, 2H), 3.79-4.13 (m, 3H), 6.55 (bs, 0.7H), 7.29 (d, 2H, J=7.7Hz), 7.74 (d, 2H, J=7.7Hz) 9.67 (bs, 0.6H).

PMT-201

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.78-0.93 (m, 9H), 1.20-1.28 (m, 4H), 1.47-1.50 (m, 3H), 3.23-3.32 (m, 2H), 3.81(s, 1.5H), 4.08-4.16 (m, 1.5H), 6.76 (bs, 0.7H), 7.47 (d, 2H, J=8.0Hz), 7.81 (d, 2H, J=8.0Hz) , 9.75 (bs, 0.6H).

PMT-202

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.80-0.95 (m, 9H), 1.21-1.28 (m, 4H), 1.46-1.51 (m,

3H), 3.20-3.25 (m, 2H), 3.63 (d, 1H, J=15.2Hz), 3.87 (s, 3H), 4.01-4.09 (m, 2H), 6.01 (d, 0.8H, J=9.6Hz), 6.97 (d, 2H, J=8.7Hz), 7.879(d, 2H, J=8.7Hz) , 9.41(bs, 0.7H).

PMT-203

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.81-0.93 (m, 9H), 1.23-1.28 (m, 4H), 1.46-1.50 (m, 3H), 3.15-3.22 (m, 2H), 3.83-4.33 (m, 5H), 6.14 (bs, 0.8H), 7.32-7.45 (m, 5H), 9.82 (bs, 0.6H).

PMT-247

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.90 (m, 3H), 1.25 (m, 5H), 1.45 (m, 2H), 3.30 (m, 2H), 3.80-4.23 (m, 3H), 7.47 (d, 2H), 7.81 (d, 2H).

PMT-248

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.89 (m, 3H), 1.24 (m, 5H), 1.45 (m, 2H), 3.27 (m, 2H), 3.75-4.19 (m, 6H), 6.97(d, 2H), 7.79 (d, 2H).

PMT-249

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.843-0.99 (m, 9H), 1.17-1.90 (m, 5H), 3.25 (m, 2H), 3.70-3.91 (m, 3H), 6.17 (m, 1H), 7.48 (d, 2H), 7.79 (d, 2H).

PMT-250

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.84-1.01 (m, 9H), 1.18-1.88 (m, 5H), 3.22 (m, 2H), 3.60 (m, 1H), 3.87 (s, 4H), 4.05 (m, 1H), 76.97 (d, 2H), 7.77 (d, 2H).

PMT-251

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.85 (m, 3H), 1.02-1.26 (m, 4H), 2.85 (m, 4H), 3.62-4.18 (m, 5H), 7.05-7.34 (m, 10H).

PMT-272

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.94 (m, 12H), 1.24 (m, 4H), 3.32 (m, 2H), 3.90 (s, 6H), 6.99 (m, 2H), 7.79 (m, 2H).

PMT-324

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.99 (m, 12H), 1.32 (m, 2H), 1.66 (m, 2H), 2.96 (s, 3H), 3.42 (m, 1H), 3.62 (m, 1H), 3.82 (m, 1H), 4.15 (m, 2H), 5.96 (m, 1H), 9.76 (br, 1H).

PMT-325

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.99 (m, 12H), 1.32 (m, 5H), 1.66 (m, 2H), 3.05 (m, 2H), 3.42 (m, 1H), 3.62 (m, 1H), 3.82 (m, 1H), 4.15 (m, 2H), 5.55 (m, 1H), 9.76 (br, 1H).

PMT-326

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.99 (m, 15H), 1.32 (m, 2H), 1.66 (m, 2H), 1.88 (m, 2H), 2.96 (s, 2H), 3.42 (m, 1H), 3.62 (m, 1H), 3.82 (m, 1H), 4.15 (m, 2H), 5.50 (m, 1H), 9.76 (br, 1H).

PMT-327

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.99 (m, 12H), 1.25 (m, 4H), 3.26 (m, 2H), 3.94 (m, 3H), 6.41 (m, 1H), 7.54 (m, 3H), 7.86 (m, 2H), 9.76 (br, 1H).

PMT-328

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.99 (m, 12H), 1.25 (m, 4H), 2.41 (s, 3H), 3.23 (m, 2H), 3.94 (m, 3H), 6.27 (m, 1H), 7.26 (m, 2H), 7.74 (m, 2H), 9.76 (br, 1H).

PMT-331

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.95 (m, 12H), 1.28 (m, 4H), 3.23 (m, 1H), 3.41 (m, 1H), 3.76 (m, 1H), 4.09 (m, 2H), 4.30 (m, 2H), 7.35 (m, 5H).

PMT-338

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.92 (m, 12H), 1.23-1.64 (m, 4H), 3.24-3.37 (m, 2H), 3.76-3.95 (m, 3H), 6.17 (m, 1H), 7.47-7.81 (m, 4H), 9.67 (br, 1H).

PMT-339

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.86-0.96 (m, 12H), 1.21-1.41 (m, 13H), 3.17-3.26 (m, 2H), 3.67-3.91 (m, 3H), 5.95 (m, 1H), 7.51-7.80 (m, 4H), 9.64 (br, 1H).

PMT-340

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.82-0.95 (m, 12H), 1.24-1.53 (m, 4H), 3.22 (m, 2H), 3.55-3.99 (m, 3H), 5.89 (m, 1H), 7.51-8.71 (m, 7H), 9.47 (br, 1H).

PMT-557

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.91 (m, 9H), 1.46-1.86 (m, 4H), 2.34 (m, 3H), 3.55 (m, 1H), 3.61 (m, 1H), 4.09 (m, 2H), 7.40-7.86 (m, 4H), 8.05 (br, 1H).

PMT-567

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.86 (m, 4H), 0.94 (m, 12H), 2.32 (m, 1H), 2.34 (m, 3H), 3.54 (m, 1H), 4.09 (m, 2H), 7.40-7.74 (m, 4H), 8.05 (br, 1H).

PMT-572

^{1}H-NMR ($CDCl_3$, 300 MHz) : 0.94 (m, 12H), 2.01-2.23 (m, 6H), 2.34 (m, 3H), 3.54 (m, 1H), 4.06 (m, 1H), 4.09 (m, 2H), 7.40-7.74(m, 4H), 8.05 (br, 1H).

References

1 Kitchen DB, Decornez H, Furr JR, et al. Docking and scoring in virtual screening for drug discovery: Methods and applications. Nat Rev Drug Discov, 2004, 3: 935-949

2 Bleicher KH, Bohm H-J, Muller K, et al. Hit and lead generation: Beyond high-throughput screening. Nat Rev Drug Discov, 2003, 2: 369-378

3 Kramer JA, Sagartz JE, Morris DL. The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. Nat Rev Drug Discov, 2007, 6: 636-649

4 Russell RJ, Haire LF, Stevens DJ, et al. The structure of h5n1 avian influenza neuraminidase suggests new opportunities for drug design. Nature, 2006, 443: 45-49

5 Capdeville R, Buchdunger E, Zimmermann J, et al. Glivec (sti571, imatinib), a rationally developed, targeted anticancer drug. Nat Rev Drug Discov, 2002, 1: 493-502

6 Weisberg E, Manley P, Mestan J, et al. Amn107 (nilotinib): A novel and selective inhibitor of bcr-abl. Br J Cancer, 2006, 94: 1765-1769

7 Morales A, Gingell C, Collins M, et al. Clinical safety of oral sildenafil citrate (viagra) in the treatment of erectile dysfunction. International journal of impotence research, 1998, 10: 69-73; discussion 73-64

8 Dorr P, Westby M, Dobbs S, et al. Maraviroc (uk-427,857), a potent,

orally bioavailable, and selective small-molecule inhibitor of chemokine receptor ccr5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. Antimicrobial Agents and Chemotherapy, 2005, 49: 4721-4732

9 Ghosh AK, Dawson ZL, Mitsuya H. Darunavir, a conceptually new hiv-1 protease inhibitor for the treatment of drug-resistant hiv. Bioorganic & Medicinal Chemistry, 2007, 15: 7576-7580

10 Eriksson BI, Smith H, Yasothan U, et al. Dabigatran etexilate. Nat Rev Drug Discov, 2008, 7: 557-558

11 Kaldor SW, Kalish VJ, Davies JF, et al. Viracept (nelfinavir mesylate, ag1343): A potent, orally bioavailable inhibitor of hiv-1 protease. Journal of Medicinal Chemistry, 1997, 40: 3979-3985

12 Raja A, Lebbos J, Kirkpatrick P. Atazanavir sulphate. Nat Rev Drug Discov, 2003, 2: 857-858

13 Organization WH, 978 92 4 156450 2. World Health Organization, 2012

14 Programme GT. Global tuberculosis control: Who report. Global Tuberculosis Programme, World Health Organization, 2010

15 Dye C, Williams BG, Espinal MA, et al. Erasing the world's slow stain: Strategies to beat multidrug-resistant tuberculosis. Science, 2002, 295: 2042-2046

16 Glaser R, Rabin B, Chesney M, et al. Stress-induced immunomodulation: Implications for infectious diseases? JAMA, 1999, 281: 2268-2270

17 Beck-Sague C, Dooley SW, Hutton MD, et al. Hospital outbreak of multidrug-resistant mycobacterium tuberculosis infections. Factors in transmission

to staff and hiv-infected patients. JAMA, 1992, 268: 1280-1286

18 Chandra RK. Nutrition, immunity and infection: From basic knowledge of dietary manipulation of immune responses to practical application of ameliorating suffering and improving survival. Proc Natl Acad Sci U S A, 1996, 93: 14304-14307

19 Xie Q, Tang Y, Li W, et al. Investigation of the binding mode of (−)-meptazinol and bis-meptazinol derivatives on acetylcholinesterase using a molecular docking method. J Mol Model, 2006, 12: 390-397

20 Camus JC, Pryor MJ, Medigue C, et al. Re-annotation of the genome sequence of mycobacterium tuberculosis h37rv. Microbiology, 2002, 148: 2967-2973

21 Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of mycobacterium tuberculosis from the complete genome sequence. Nature, 1998, 393: 537-544

22 Park HD, Guinn KM, Harrell MI, et al. Rv3133c/dosr is a transcription factor that mediates the hypoxic response of mycobacterium tuberculosis. Mol Microbiol, 2003, 48: 833-843

23 Boon C, Dick T. How mycobacterium tuberculosis goes to sleep: The dormancy survival regulator dosr a decade later. Future Microbiol, 2012, 7: 513-518

24 Trauner A, Lougheed KE, Bennett MH, et al. The dormancy regulator dosr controls ribosome stability in hypoxic mycobacteria. J Biol Chem, 2012, 287:

24053-24063

25 Rustad TR, Sherrid AM, Minch KJ, et al. Hypoxia: A window into mycobacterium tuberculosis latency. Cell Microbiol, 2009, 11: 1151-1159

26 Rosenkrands I, Slayden RA, Crawford J, et al. Hypoxic response of mycobacterium tuberculosis studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. Journal of bacteriology, 2002, 184: 3485-3491

27 Klinkenberg LG, Sutherland LA, Bishai WR, et al. Metronidazole lacks activity against mycobacterium tuberculosis in an in vivo hypoxic granuloma model of latency. J Infect Dis, 2008, 198: 275-283

28 Sherman DR, Voskuil M, Schnappinger D, et al. Regulation of the mycobacterium tuberculosis hypoxic response gene encoding α-crystallin. Proceedings of the National Academy of Sciences, 2001, 98: 7534-7539

29 Dawes SS, Warner DF, Tsenova L, et al. Ribonucleotide reduction in mycobacterium tuberculosis: Function and expression of genes encoding class ib and class ii ribonucleotide reductases. Infect Immun, 2003, 71: 6124-6131

30 Sharma P, Kumar B, Gupta Y, et al. Proteomic analysis of streptomycin resistant and sensitive clinical isolates of mycobacterium tuberculosis. Proteome science, 2010, 8: 59

31 Buchko GW, Kim CY, Terwilliger TC, et al. Solution structure of the conserved hypothetical protein rv2302 from mycobacterium tuberculosis. Journal of bacteriology, 2006, 188: 5993-6001

32 Kapust RB, Tozser J, Fox JD, et al. Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng, 2001, 14: 993-1000

33 Kapust RB, Waugh DS. Controlled intracellular processing of fusion proteins by tev protease. Protein Expr Purif, 2000, 19: 312-318

34 Delaglio F, Grzesiek S, Vuister GW, et al. Nmrpipe: A multidimensional spectral processing system based on unix pipes. J Biomol NMR, 1995, 6: 277-293

35 Johnson BA, Blevins RA. Nmr view: A computer program for the visualization and analysis of nmr data. J Biomol NMR, 1994, 4: 603-614

36 Wishart DS, Bigam CG, Yao J, et al. 1h, 13c and 15n chemical shift referencing in biomolecular nmr. J Biomol NMR, 1995, 6: 135-140

37 Kay L, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. Journal of the American Chemical Society, 1992, 114: 10663-10665

38 Schleucher J, Schwendinger M, Sattler M, et al. A general enhancement scheme in heteronuclear multidimensional nmr employing pulsed field gradients. Journal of biomolecular NMR, 1994, 4: 301-306

39 Muhandiram D, Kay LE. Gradient-enhanced triple-resonance three-dimensional nmr experiments with improved sensitivity. Journal of Magnetic Resonance, Series B, 1994, 103: 203-216

40 Kay L. A gradient-enhanced hcch-tocsy experiment for recording side-chain 1h and 13c correlations in h2o samples of proteins. J Magn Reson, 1993,

101: 333-337

41 Pascal S, Muhandiram D, Yamazaki T, et al. Simultaneous acquisition of 15n-and 13c-edited noe spectra of proteins dissolved in h2o. Journal of magnetic resonance Series B, 1994, 103: 197-201

42 Zhang O, Kay LE, Olivier JP, et al. Backbone 1h and 15n resonance assignments of the n-terminal sh3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient nmr techniques. Journal of biomolecular NMR, 1994, 4: 845-858

43 Farrow NA, Muhandiram R, Singer AU, et al. Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by 15n nmr relaxation. Biochemistry, 1994, 33: 5984-6003

44 Zweckstetter M, Bax A. Characterization of molecular alignment in aqueous suspensions of pf1 bacteriophage. J Biomol NMR, 2001, 20: 365-377

45 Guntert P. Automated nmr structure calculation with cyana. Methods Mol Biol, 2004, 278: 353-378

46 Case DA, Cheatham TE, Darden T, et al. The amber biomolecular simulation programs. J Comput Chem, 2005, 26: 1668-1688

47 Pearlman DA, Case DA, Caldwell JW, et al. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications, 1995, 91: 1-41

48 Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR, 1999, 13: 289-302

49 Shen Y, Delaglio F, Cornilescu G, et al. Talos+: A hybrid method for predicting protein backbone torsion angles from nmr chemical shifts. J Biomol NMR, 2009, 44: 213-223

50 Davis IW, Leaver-Fay A, Chen VB, et al. Molprobity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res, 2007, 35: W375-383

51 Davis IW, Murray LW, Richardson JS, et al. Molprobity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res, 2004, 32: W615-619

52 Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins, 2007, 66: 778-795

53 Valafar H, Prestegard JH. Redcat: A residual dipolar coupling analysis tool. Journal of Magnetic Resonance, 2004, 167: 228-241

54 Dundas J, Ouyang Z, Tseng J, et al. Castp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic acids research, 2006, 34: W116-118

55 Shazman S, Celniker G, Haber O, et al. Patch finder plus (pfplus): A web server for extracting and displaying positive electrostatic patches on protein surfaces. Nucleic acids research, 2007, 35: W526-530

56 Koradi R, Billeter M, Wuthrich K. Molmol: A program for display and analysis of macromolecular structures. J Mol Graph, 1996, 14: 51-55, 29-32

57 Pettersen EF, Goddard TD, Huang CC, et al. Ucsf chimera--a visualization system for exploratory research and analysis. J Comput Chem, 2004, 25: 1605-1612

58 Baker NA, Sept D, Joseph S, et al. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98: 10037-10041

59 Creighton TE. Proteins: Structures and molecular properties. WH Freeman, 1992

60 Phan J, Zdanov A, Evdokimov AG, et al. Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem, 2002, 277: 50564-50572

61 Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using clustalw and clustalx. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al], 2002, Chapter 2: Unit 2 3

62 Fukami-Kobayashi K, Saito N. [how to make good use of clustalw]. Tanpakushitsu kakusan koso Protein, nucleic acid, enzyme, 2002, 47: 1237-1239

63 Clamp M, Cuff J, Searle SM, et al. The jalview java alignment editor. Bioinformatics, 2004, 20: 426-427

64 Waterhouse AM, Procter JB, Martin DM, et al. Jalview version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics, 2009, 25: 1189-1191

65 Holm L, Sander C. Touring protein fold space with dali/fssp. Nucleic acids research, 1998, 26: 316-319

66 Berman HM, Westbrook J, Feng Z, et al. The protein data bank. Nucleic acids research, 2000, 28: 235-242

67 Zhao Q, Qin L, Jiang F, et al. Structure of human spindlin1 tandem tudor-like domains for cell cycle regulation. Journal of Biological Chemistry, 2007, 282: 647-656

68 Sokolowska M, Czapinska H, Bochtler M. Crystal structure of the ββα-me type ii restriction endonuclease hpy99i with target DNA. Nucleic acids research, 2009, 37: 3799-3810

69 Ballaré C, Lange M, Lapinaite A, et al. Phf19 links methylated lys36 of histone h3 to regulation of polycomb activity. Nature structural & molecular biology, 2012, 19: 1257-1265

70 Steiner T, Kaiser JT, Marinkoviç S, et al. Crystal structures of transcription factor nusg in light of its nucleic acid-and protein-binding activities. The EMBO journal, 2002, 21: 4641-4653

71 Klein BJ, Bose D, Baker KJ, et al. Rna polymerase and transcription elongation factor spt4/5 complex structure. Proceedings of the National Academy of Sciences, 2011, 108: 546-550

72 Magombedze G, Mulder N. Understanding tb latency using computational and dynamic modelling procedures. Infection, Genetics and Evolution, 2012,

73 Yuan Z, Trias J, White RJ. Deformylase as a novel antibacterial target.

Drug Discov Today, 2001, 6: 954-961

74 Watters AA, Jones RN, Leeds JA, et al. Antimicrobial activity of a novel peptide deformylase inhibitor, lbm415, tested against respiratory tract and cutaneous infection pathogens: A global surveillance report (2003-2004). J Antimicrob Chemother, 2006, 57: 914-923

75 Lofland D, Difuntorum S, Waller A, et al. In vitro antibacterial activity of the peptide deformylase inhibitor bb-83698. J Antimicrob Chemother, 2004, 53: 664-668

76 Yuan Z, Trias J, White RJ. Deformylase as a novel antibacterial target. Drug Discovery Today, 2001, 6: 954-961

77 Meinnel T. Peptide deformylase of eukaryotic protists: A target for new antiparasitic agents? Parasitol Today, 2000, 16: 165-168

78 Dirk LM, Williams MA, Houtz RL. Eukaryotic peptide deformylases. Nuclear-encoded and chloroplast-targeted enzymes in arabidopsis. Plant Physiol, 2001, 127: 97-107

79 Lee MD, She Y, Soskis MJ, et al. Human mitochondrial peptide deformylase, a new anticancer target of actinonin-based antibiotics. J Clin Invest, 2004, 114: 1107-1116

80 Escobar-Alvarez S, Goldgur Y, Yang G, et al. Structure and activity of human mitochondrial peptide deformylase, a novel cancer target. J Mol Biol, 2009, 387: 1211-1228

81 Lee BJ, Choi SK, Lee KW, et al. Patent pct/kr2004/000502. 2004-.

82 Lee SK, Choi KH, Lee SJ, et al. Peptide deformylase inhibitors with retro-amide scaffold: Synthesis and structure-activity relationships. Bioorg Med Chem Lett, 2010, 20: 4317-4319

83 Lee SK, Choi KH, Lee SJ, et al. Peptide deformylase inhibitors with non-peptide scaffold: Synthesis and structure-activity relationships. Bioorg Med Chem Lett, 2011, 21: 133-136

84 Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol, 1995, 245: 43-53

85 Kuntz ID, Blaney JM, Oatley SJ, et al. A geometric approach to macromolecule-ligand interactions. Journal of molecular biology, 1982, 161: 269-288

86 Ewing TJ, Makino S, Skillman AG, et al. Dock 4.0: Search strategies for automated molecular docking of flexible molecule databases. Journal of computer-aided molecular design, 2001, 15: 411-428

87 Moustakas DT, Lang PT, Pegg S, et al. Development and validation of a modular, extensible docking program: Dock 5. Journal of computer-aided molecular design, 2006, 20: 601-619

88 Lang PT, Brozell SR, Mukherjee S, et al. Dock 6: Combining techniques to model rna-small molecule complexes. RNA, 2009, 15: 1219-1230

89 Pieters R, Huismans DR, Leyva A, et al. Adaptation of the rapid automated tetrazolium dye based (mtt) assay for chemosensitivity testing in childhood leukemia. Cancer Lett, 1988, 41: 323-332

90 Han JH, Choi YS, Kim WJ, et al. Codon optimization enhances protein expression of human peptide deformylase in e. Coli. Protein Expr Purif, 2010, 70: 224-230

91 Zartler ER, Yan J, Mo H, et al. Id nmr methods in ligand-receptor interactions. Curr Top Med Chem, 2003, 3: 25-37

92 Nayeem A, Sitkoff D, Krystek S, Jr. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. Protein Sci, 2006, 15: 808-824

93 Brooks BR, Brooks CL, 3rd, Mackerell AD, Jr., et al. Charmm: The biomolecular simulation program. J Comput Chem, 2009, 30: 1545-1614

94 Puigbò P, Guzmán E, Romeu A, et al. Optimizer: A web server for optimizing the codon usage of DNA sequences. Nucleic acids research, 2007, 35: W126-W131

95 DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: New estimates of drug development costs. J Health Econ, 2003, 22: 151-185

96 Kuntz ID. Structure-based strategies for drug design and discovery. Science, 1992, 257: 1078-1082

97 Annamala MK, Inampudi KK, Guruprasad L. Docking of phosphonate and trehalose analog inhibitors into m. Tuberculosis mycolyltransferase ag85c: Comparison of the two scoring fitness functions goldscore and chemscore, in the gold software. Bioinformation, 2007, 1: 339-350

98 John S, Thangapandian S, Sakkiah S, et al. Discovery of potential

pancreatic cholesterol esterase inhibitors using pharmacophore modelling, virtual screening, and optimization studies. J Enzyme Inhib Med Chem, 2011, 26: 535-545

99 Lu IL, Huang CF, Peng YH, et al. Structure-based drug design of a novel family of ppargamma partial agonists: Virtual screening, x-ray crystallography, and in vitro/in vivo biological activities. J Med Chem, 2006, 49: 2703-2712

100 Aparna V, Rambabu G, Panigrahi SK, et al. Virtual screening of 4-anilinoquinazoline analogues as egfr kinase inhibitors: Importance of hydrogen bonds in the evaluation of poses and scoring functions. J Chem Inf Model, 2005, 45: 725-738

101 Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. Journal of Medicinal Chemistry, 2000, 43: 4759-4767

102 Redwan E-RM. The optimal gene sequence for optimal protein expression in escherichia coli: Principle requirements. Arab J Biotechnol, 2006, 9: 493-510

103 Guo Y, Wallace SS, Bandaru V. A novel bicistronic vector for overexpressing *mycobacterium tuberculosis* proteins in *escherichia coli*. Protein expression and purification, 2009, 65: 230-237

국문초록

구조 기반 약물 설계는 합리적인 약물 개발 방법의 하나로 목적 단백질의 구조를 밝히고 이후 결합자리 연구를 통해 선도물질을 도출한다. 결핵균 유래의 단백질 Rv0569 의 구조 연구를 통하여 구조기반 약물 설계의 초반 부분을 수행하였다. Rv0569 단백질의 클로닝, 대량발현, 단백질 정제의 과정을 거쳐 원편광이색성 (Circular Dichrosim) 실험을 통해 단백질의 2차구조를 예측하고 기본적인 물성을 알아 본 뒤 핵자기공명 (Nuclear Magnetic Resonance) 실험을 통하여 단백질의 구조를 계산하였다. 계산된 구조와 구조 기반 기능 예측과 기존연구 분석을 통해 Rv0569 단백질이 결핵균의 잠복기 감염시 저산소 신호 전달에 관여하는 기능을 가짐을 제안하였다. 구조기반 약물 설계의 후반부분을 사람 유래 펩타이드 디포밀라제 단백질의 저해제 선도물질 도출연구를 통해 수행하였다. 43개의 박테리아 유래 저해제 라이브러리에서 결합 모의 실험 (Docking Simulation)을 통해 10 종의 화합물을 선택하고 이의 항암능을 실험하였다. 항암능 실험결과에서 우수한 결과를 보인 4종의 화합물과

핵자기 공명법 실험을 통하여 사람유래의 펩타이드 디포밀라제 단백질과 결합 실험을 수행하였다. 4종 중 2종의 박테리아 유래의 저해제가 사람 유래의 단백질과 결합함을 확인하였고 이 2종의 화합물을 선도물질로 도출하였다.

.....

주요어 : 구조 기반 약물 설계, 결핵, Rv0569, 핵자기공명, 저산소 신호 전달, 펩타이드 디포밀라제, 결합 모의 실험, 항암능, 선도물질

학번 : 2004-22292