Duplicated DNA Damage-Inducible Genes *DDI2/DDI3* in Budding Yeast *Saccharomyces cerevisiae* Encodes a Functional Cyanamide Hydratase

A Thesis Submitted to the College of Graduate Studies and Research in Partial Fulfillment of the Requirement for the Degree of Doctor of Philosophy
in the Department of Microbiology and Immunology
University of Saskatchewan
Saskatoon

By

Jia Li, M.Sc

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material on this thesis in whole or part should be addressed to:

Head of the Department of Microbiology and Immunology

Health Science Building, 107 Wiggins Road

University of Saskatchewan

Saskatoon SK S7N 5E5 Canada

ABSTRACT

In a yeast genome-wide microarray analysis, two *Saccharomyces cerevisiae* open reading frames *YFL061w* and *YNL335w* displayed the highest fold levels of induction after methyl methanesulfonate (MMS) treatment, a typical DNA damaging agent. Hence the genes were named as DNA damage inducible genes 2 and 3 (*DDI2* and *DDI3*). *DDI2* and *DDI3* share identical DNA sequences in their coding regions and promoters (except for one nucleotide). Their deduced protein product Ddi2/3 is 36% identical to an identified cyanamide hydratase (EC 4.2.1.69) from a soil fungus *Myrothecium verrucaria*. The enzyme catalyzes cyanamide hydration to urea. Hence Ddi2/3 likely has the same enzymatic activity. To study the encoded protein, the *DDI2/3* open reading frame was cloned into a bacterial expression system, and the resulting recombinant protein was purified for biochemical analysis. The cyanamide hydratase activity of Ddi2/3 was measured by using an enzyme-based urea detection assay. The kinetic study showed that Ddi2/3 has a K_M of 17 mM for cyanamide, and k_{cat} is 9.8 ± 0.1 S^{-1}. The cyanamide hydratase activity of endogenous Ddi2/3 was also detected by using yeast whole cell extracts (WCEs).

The cyanamide hydratase domain belongs to the Histidine–Aspartate (HD) domain superfamily. Double mutations in conserved His88 and Asp89 residues abolished the enzymatic activity, suggesting that the conserved HD residues contribute to the cyanamide hydratase activity of Ddi2/3.

To identify whether *DDI2/3* can be induced by cyanamide, the *DDI2/3* promoter was fused to *lacZ* as a reporter gene, and the β-galactosidase assay demonstrated that *DDI2/3-lacZ* can be activated by both MMS and cyanamide, and the optimal fold

induction by cyanamide was even higher than by MMS. Interestingly, the basal level expression of *DDI2/3-lacZ* was extremely low, and yeast WCEs from untreated wild-type cells do not have measurable cyanamide hydratase activity.

To investigate the physiological function of *DDI2/3*, we constructed yeast *ddi2/3Δ* single and double null mutant strains by sequential targeted one-step gene disruption via homologous recombination. Disruption of *DDI2/3* genes enhanced cellular sensitivity to both MMS and cyanamide by up to 1,000-fold. Furthermore, WCEs from the *ddi2/3*Δ single mutant had reduced cyanamide hydratase activity, while that from the *ddi2Δ ddi3*Δ double mutant had no detectable cyanamide hydratase activity. Since the induction of *DDI2/3-lacZ* by cyanamide and MMS was not compromised in the *ddi2/3*Δ strains, the *DDI2/3* gene is not required for its transcriptional regulation.

Since no structural study with a cyanamide hydratase domain has been reported, we undertook crystallographic studies on Ddi2/3. After protein crystallization, X-ray diffraction, and data interpretation, Ddi2/3 was revealed as a zinc-binding protein and composed of 14 helices A zinc anion is coordinated to three conserved HD residues, H55, H88 and D89, and a water molecule, forming a tetrahedral geometry. To further investigate the active site, several amino acid substitutions were made, and enzymatic assays revealed that H137, Q138, T157, and N161 are required for the activity.

HD domain represents a superfamily of metal-dependent phosphohydrolases and characterized members are involved in nucleotide metabolism and signal transduction. However, CAH has been reported as an exception. Hence, the study on Ddi2/3 will enrich our knowledge of the variation of HD domain proteins, and has identified a unique subfamily within the HD domain family.

ACKNOWLEDGEMENTS

I would first like to express my sincere thanks to my supervisor Dr. Wei Xiao for providing me with a great research opportunity.

I would also like to thank the members of my supervisory committee, Dr. Stanley Moore, Dr. Linda Chelico, Dr. Hugh Goldie, and Dr. Yu Luo, for their selfless supports, helpful insights and useful suggestions for the progression of my research. My gratitude is extended to all the members of Dr. Wei Xiao's laboratory: Michelle Hanna, Parker Anderson, Landon Pastushok, Lindsay Ball, Hanna Dworaczek, Rui Wen, Micheal Biss, Susan Blackwell, Amanda Lambrecht and Zhaojia Wu, for their help and providing a friendly and productive environment. Yunhua Jia from Dr.Moore's lab helped me a lot in doing protein purification and crystallization. Yuqing Feng from Dr.Chelico's lab helped me using the HPLC facility in protein purification and gel filtration. I would also like to thank the professors, support staff and graduate students in the department.

Financial supports to my project are from the Natural Sciences and Engineering Research Council (NSERC), College of Medicine of University of Saskatchewan, the Proteomics Research in Interactions and Structure of Macromolecules (PRISM) of University os Saskatchewan, and the Canadian Institute for Health Research (CIHR) Training in Health Research Using Synchrotron Techniques (THRUST). Finally, I would thank my parents and my husband who give me unconditional love, understanding, help and encouragement all the time, which is necessary to ensure my graduation.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4-NQO	4-Nitroquinoline-N-oxide
aa	amino acid
Ala, A	alanine
Amp	ampicillin
Arg, R	arginine
Asn, N	asparagine
Asp, D	aspartate
BER	base excision repair
bp	base pair
CA	carbonic anhydrase
CAH	cyanamide hydratase
co-IP	co-immunoprecipitation
COMW	cut-off molecular weight
Cys, C	cysteine
ddH2O	double distilled water
DDI	DNA damage-inducible
DMSO	Dimethu=yl sulfoxide
dNDP	deoxynucleoside diphosphate
dNTP	deoxynucleoside triphosphate
dNTPase	deoxynucleoside triphosphase
dsDNA	double stranded DNA
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMS	ethyl methanesulfonate
EtBr	ethidium bromide
Gln, Q	glutamine
Glu, E	glutamate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His, H	histidine
HPLC	High-performance liquid chromatography
HR	homologous recombination
HU	hydroxyurea
IEC	ion-exchange chromatography

Ile, I	isoleucine
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
Kb	kilo base pair
kDa	kilo Dalton
LB	Luria-Bertani broth
Leu, L	leucine
Lys, K	lysine
Met, M	methionine
MES	2-(N-morpholino)ethanesulfonic acid
MI	*myo*-inositol
MIOX	*myo*-inositol oxygenase
MMS	methyl methanesulfonate
MNNG	N-methyl-N'-nitro-N-nitrosoguanidine
*M*r	relative molecular mass
MW	molecular weight
NDP	nucleoside diphosphate
NTP	nucleoside triphosphate
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pI	isoelectric point
PMSF	phenylmethylsulfonyl fluoride
PSI	pound per square inches
RNR	ribonucleotide reductase
RVP domain	retroviral-like proteinase domain
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SCB	sodium carbonate-bicarbonate buffer
SD medium	synthetic dextrose medium
SDS	sodium dodecyl sulfate
Ser, S	serine
ssDNA	single-stranded DNA or salmon sperm DNA
TFIIH	transcription factor II human
Thr, T	threonine
Trp, W	tryptophan
UAS	upstream activating sequence
Ub	ubiquitin
UBA	ubiquitin-associated domain
UBL	ubiquitin-like domain
Ura	uracil
UV	ultraviolet radiation
Val, V	valine

WCE	whole cell extract
Wt	wild type
YPD medium	yeast extract-peptone-dextrose medium
β-Gal	β-galactosidase

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1. Cyanamide

1.1.1. Distribution of Cyanamide in Nature and the Natural Synthesis of Cyanamide

Cyanamide, H_2N-C≡N, (*M*r 42.04) was obtained in 1838 by passing gaseous ammonia over cyanogen chloride. Presently, cyanamide is manufactured from calcium cyanamide (CaNCN) by continuous carbonation in an aqueous medium (May, 1967). Cyanamide is a weak acid that is highly soluble in water. It is entirely soluble at 43℃ and has a minimum solubility at -15°C. It is highly soluble in polar organic solvents including alcohols, esters, nitriles, and amides, but less soluble in nonpolar solvents like saturated or aromatic hydrocarbons. Cyanamide exists in two tautomeric forms: nitrile form (H_2N-C≡N) and carbodiimide form (HN=C=NH) that cannot be isolated in pure form. Spectroscopic evidence and reactivity clearly show that the nitrile form predominates by far (Güthner and Mertschenk, 2000).

This chemical was likely present in abundance due to the environment conditions in prebiological time (Schimpl *et al*., 1965). Moreover, it has been proposed that cyanamide is involved in chemical evolution (Steinman *et al*., 1964). It is hypothesized that, at an early stage in the origin of life, an informational polymer must have arisen by purely chemical means. According to the 'RNA world' hypothesis, this polymer was RNA, and it is hypothesized that 'activated' ribonucleotides can polymerize to form RNA (Powner *et al*., 2009). However, it is far from obvious how such ribonucleotides could

have formed from their constituent parts (ribose and nucleobases). Reported studies showed that activated pyrimidine ribonucleotides could be formed by using materials including cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate. Moreover, the conditions of the synthesis were consistent with potential early-Earth geochemical models (Powner *et al*., 2009).

Cyanamide has long been recognized as a synthetic compound, but recently Japanese researchers found cyanamide to be a natural product from several leguminous plants. Kamo and his colleagues firstly isolated cyanamide from *Vicia villosa* as a plant growth inhibitor (Kamo *et al*., 2003). Subsequently, they extracted cyanamide from potassium [^{15}N] nitrate treated plants and analyzed ^{15}N/ (^{14}N + ^{15}N) ratios of cyanamide. The isotopic ratio ^{15}N/ (^{14}N + ^{15}N) of the cyanamide was calculated to be 0.143, and the isotopic ratio the cyanamide extracted from *V. villosa* grown in a natural nitrogen environment was 0.0065. Thus, it demonstrated that cyanamide can be biosynthesized *de novo* by *V. villosa* utilizing nitrogen from the environment (Kamo *et al*., 2006a). Crude extracts from hairy vetch leaves and stems containing 1.3 ppm of cyanamide inhibited radicle growth of lettuce by 40% (Kamo *et al*., 2006b). After investigating the distribution of natural cyanamide in *Vicia* species including *V. villosa* subsp. *varia, V. cracca*, and *V. amoena* during their pre-flowering and flowering seasons, Kamo and his colleagues confirmed that *V. cracca* is superior to *V. villosa* subsp. *varia* in accumulating natural cyanamide and that *V. amoena* is unable to biosynthesize this compound under the laboratory conditions examined. The presence of cyanamide in the leaves of *V. villosa* subsp. *varia* seedlings were also clarified. In a screen for plants producing cyanamide, only *Robinia pseudo-acacia* was found to contain cyanamide among 452 species of

higher plants. After investigating 553 species to date, researchers found that only three species, *V villosa* subsp. *varia, V. cracca* and *R. pseudo-acacia* can biosynthesize cyanamide (Kamo *et al*., 2008). The same group continued investigating the material and synthesis mechanism of cyanamide by *V. villosa* subsp. *Varia*. The incorporation of ^{15}N from [^{15}N] nitrate and [^{15}N] ammonium into cyanamide using shoots of *V. villosa* subsp. *varia* was demonstrated. [$^{15}N_2$] cyanamide was applied to shoots of *V. villosa* subsp. *varia* to monitor the turnover of cyanamide de novo; [$^{15}N_2$] cyanamide was detected in the leaves within four hours, and it was present without detectable degradation for more than four days. In contrast, maximum incorporation of ^{15}N into cyanamide molecules was observed after four days of feeding the shoots with ^{15}N-labeled inorganic ions and l-[amide-^{15}N]-glutamine, indicating that these nitrogenous compounds are distant precursors of cyanamide (Kamo *et al*., 2009). To investigate the carbon source of synthesized cyanamide, the ^{13}C labels of [^{13}C]carbon dioxide and D-[$^{13}C_6$] glucose were incorporated into cyanamide (NH_2CN) when they were administered to *V. villosa* subsp. *varia* shoots, indicating that the carbon atom of cyanamide is derived from any of the carbohydrates present upstream of pyruvate during the metabolism (Kamo *et al*., 2010). These researchers noticed that the distribution of L-canavanine in the plant kingdom appeared to overlay that of cyanamide, and they noticed that the guanidino group structure in L-canavanine contained the cyanamide skeleton. The latest research progress of Kamo group suggested that cyanamide is biosynthesized from L-canavanine. Their experiments showed that the addition of L-[guanidineimino-$^{15}N_2$] canavanine to young *Vicia villosa* seedlings resulted in the significant incorporation of ^{15}N-label into cyanamide, verifying its presumed biosynthetic pathway (Kamo *et al*., 2015).

Cyanamide (H_2N-C≡N) can be regarded as the amide of a cyanic acid (HO-CN), or as the nitrile of carbamic acid (H_2N-COOH). The bifunctional molecule is capable of reacting as a nucleophile at the amino group or as an electrophile at the cyano group. Many reactions with polyfunctional compounds involve both reactive sites, thus forming heterocyclic structures. One of the most important reactions of cyanamide is dimerization to cyanoguanidine (dicyanamide). Unless kept cool and stabilized by weak acids, crystalline cyanamide may polymerize violently. In aqueous solution, cyanamide has the highest stability between pH 4.0 and pH 4.5; polymerization becomes significant when the pH value is above 5, with maximum reactivity between pH 8 and pH 10. Cyanamide reacts with water; the hydrolysis (hydration) of cyanamide to urea is catalyzed by acids. Cyanamide also reacts with amino acids; it adds to amino acids to form carboxyalkylguanidines (Young and Luce, 1967). The conversion of glycine to glycocyamine was initially reported in 1952 (Güthner and Mertschenk, 2000).

1.1.2. Cyanamide Is Widely Used In Agriculture and the Pharmaceutical Industry

Cyanamide is the aqueous formation of calcium cyanide [$Ca(CN_2)$], which is used as nitrogen fertilizer in agriculture. Cyanamide can be naturally degraded to urea ($(NH_2)_2CO$), a functional ingredient in nitrogen fertilizer, via hydrating water molecules in the air. Urea will be further broken down to ammonia (NH_3) and carbon dioxide (CO_2). Thus, cyanamide is an environmentally “clean” fertilizer. For other applications in agriculture, cyanamide is found as a key raw material in herbicides, such as Amitrol and Hexazinone (Fuchs. and Wommak, 1974), fungicides, and pesticides. And it is also used to synthesize the pyrimidine or 1,3,5-triazine subunit for a board range of sulfonylurea

herbicides like bensulfuron methyl (Güthner and Mertschenk, 2000; Levitt and Petersen, 1983; Sauers, 1983). It is reported that cyanamide inhibits the growth of tomato root by affecting cell division, disturbing phytohormone balance and expansin gene expression (Soltys *et al*., 2012). It is also used as plant growth regulator and bud breaking agent (Dormex®) used in mild winter to improve the bud break of grapes and other fruit crops (Nee and Fuchigami, 1992; Shulman *et al*., 1983). In the pharmaceutical industry, cyanamide is the key raw material for the anti-ulcer drug Cimetidine (Durant *et al*., 1975). The majority of antiviral and anticancer drugs with pyrimidine or purine structures are based on cyanamide and guanidine chemistry (Güthner and Mertschenk, 2000). Moreover, cyanamide is also reported as an important compound in the synthesis of guanidine derivatives (Katla *et al*., 2013).

1.1.3. Cyanamide Is an Inhibitor to Aldehyde Dehydrogenase (ALDH) and Carbonic Anhydrase (CA)

Cyanamide is taken as a medication to treat alcoholic patients because its metabolized intermediate can inhibit aldehyde dehydrogenase (ALDH) and thus interferes with the oxidation of ethanol to acetate (DeMaster *et al*., 1984). DeMaster and his colleagues later discovered that the mitochondrial enzyme catalase mediated the oxidation of cyanamide. The catalysis results in the formation of nitroxyl (HNO) and cyanide (HCN); the generated HNO interacts with and inhibits the active-site cysteine thiolate in ALDH. (DeMaster *et al*., 1998). Besides ALDH, HNO also inhibits other thiol-containing enzymes such as yeast transcription factor Ace1, Cathepsin P, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and mitochondrial respiratory

complex I and II (Paolocci *et al*., 2007). Cyanamide is only clinically approved HNO donor to date for its minimal toxicity with no carcinogenic effect (Irvine *et al*., 2008).

Cyanamide is also identified as an inhibitor of carbonic anhydrase II (CAII), but by a different mechanism. Carbonic anhydrases (CAs) are zinc metalloenzymes vital for living cells because they catalyze the conversion of CO_2 and carbonic acid to maintain the physiological pH for cellular events, and to help transport carbon dioxide out of tissues (McCall *et al*., 2000). Cyanamide has a linear triatomic molecule, like CO_2. Therefore, it mimics the CO_2 binding to the active sites of CAII (Briganti *et al*., 1999), though later studies revealed that the two substrates have different binding modes at the active site (Guerri *et al*., 2000). After hydration, cyanamide is converted to urea; however,

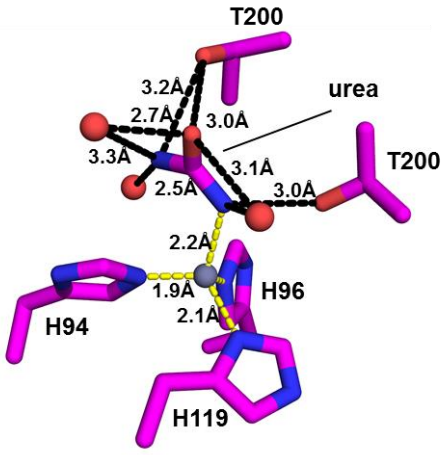


Figure 1–1. Structural model of urea bound to human CAII (hCAII) based on the crystallographic study.

The deprotonated ureate nitrogen is coordinated to the catalytic Zinc ion, a total of 8 hydrogen bonds fix the molecule within the active site. PDB ID: 1BV3. Zinc ion (gray) and water molecules (red) are shown as spheres. Metal coordination is shown as yellow dashes; hydrogen bond is represented as black dashes. The side chains of residues are colored by elements: oxygen, red, nitrogen, blue.

it forms eight hydrogen bonds with the zinc and the surrounding threonine residues as shown in Figure 1–1. Therefore, the ureate cannot be released from the active site, which results in the "death" of the enzyme. Cyanamide is the first reported suicide inhibitor of CAs (Alterio *et al*., 2012).

Cyanamide has a mild toxicity with an acute oral LD50 (rat) at 125 mg/kg. When applied dermally to rabbits, the LD50 was found to be 848 mg/kg. Cyanamide solid or 50% aqueous solution has distinct irritating properties to the skin and eyes and is regarded as sensitizing in the guinea pig by using the Magnusson–Kligman method (Magnusson and Kligman, 1969). Moreover, cyanamide did not display any mutagenic effects in several assays such as Ames test (Güthner and Mertschenk, 2000).

1.1.4. Degradation of Cyanamide by Microorganisms

Since cyanamide is produced by limited species of plants as a chemical defense to fungi and weeds, it is not surprising that some microorganisms, especially those living in soil, have developed a cyanamide biodegradation pathway for their survival.

Biodegradation of cyanamide in microorganisms is via hydrolysis by cyanamide hydratase (EC 4.2.1.69), which was initially identified from soil fungus *M. verrucaria* (*Mv*Cah) that specifically catalyzes the formation of urea from cyanamide but not with other organic nitriles. The Cah was recognized as a zinc binding protein, and it formed a homo-hexamer as a functional enzyme. Moreover, the protein expression was induced by cyanamide. Identified K_M of Cah was 27 mM (Maiergreiner *et al*., 1991).

Cyanamide is toxic to plant cells, and plants apparently do not contain a homologous *CAH* gene, according to a search using Basic Local Alignment Search Tool

(BLAST) (Altschul *et al*., 1990). Due to the detoxification capability of Cah, the gene *CAH* was used as a selectable marker for transformation of *Arabidopsis*, potato, rice and tomato as well as wheat (Weeks, 2001; Weeks *et al*., 2000). Zhang *et al*. reported using the *CAH* gene as a selective marker in transgenic soybean. After transformation, the transgenic plants expressed *CAH* mRNA, and a cyanamide hydratase activity was present in extracts of leaves and embryogenic tissue cultures when measured by a colorimetric-based enzymatic assay. The presence of Cah protein was identified by enzyme-linked immunosorbent assay (ELISA), and the whole plant displayed cyanamide resistance. However, the transgenic plants grew and set seeds normally, indicating that the Cah enzyme activity did not affect soybean metabolism (Zhang *et al*., 2005).

1.1.5. Biodegradation of Nitriles

Cyanamide is a nitrile, a class of compounds distributed widely in the environment, and most of them are toxic to higher eukaryotes. Naturally existing nitriles are primarily synthesized by plants as hormones required for growth, or as a chemical defense against microorganisms. It is not surprising that organisms have developed nitrile biodegradation systems includes hydrolysis, oxidation, and reduction pathways.

The enzymes perform the carbon-nitrogen (-CN) bond hydrolysis and are commonly grouped into a nitrilase superfamily. Nitrilases are ubiquitously present in all species of plant, animal, and fungi, as well as many prokaryotes, with more than 150 known members classified into 13 branches based on structure-sequence analysis (Barglow *et al*., 2008). However, only the first branch of the nitrilases family (EC 3.5.5.1), found exclusively in bacteria, directly hydrolyzes nitriles (R-CN) to the

corresponding acids (R-COOH) and ammonia (NH_3), the others of the nitrilases superfamily are performing as amidases (Pace and Brenner, 2001). The members of nitrilases superfamily are nitrile-inducible enzymes composed of one or two types of subunits of different size and numbers, but they all share a common CN hydrolase domain. All nitrilases appear to share a common catalytic mechanism that involves a conserved catalytic triad (Glu-Lys-Cys) including a cysteine nucleophile, a glutamate base, and a lysine residue, as shown in Figure 1–2. Mutation of any of these three residues results in an inactive enzyme (Brenner, 2002). Identified nitrilases do not contain a metal co-factor (Banerjee *et al*., 2002).

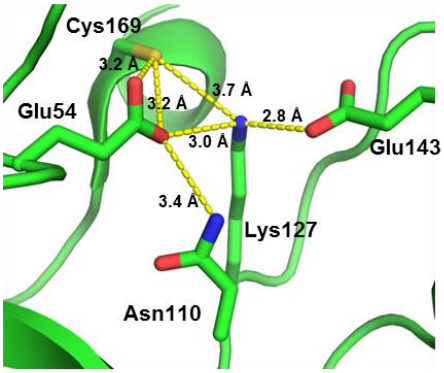


Figure 1–2. Structure model of the active site of NitFhit.

The active site of NitFhit from *Caenorhabditis elegans* (PDB ID: 1EMS). NitFhit protein consists of the branch 10 of nitrilases superfamily. All nitrilases proteins contain a catalytic triad of residues (Glu-Lys-Cys) which is essential for the function of its active site. The side chains of residues were colored by elements: oxygen: blue. Reproduced from (Brenner, 2002) by using Pymol (Schrodinger, 2015).

The catalytic properties were investigated by several groups; a proposed catalyzing model (Figure 1–3) is that the nitrile carbon bears a positive charge and is subjected to nucleophilic attack, by the -SH group of cysteine at the active site in nitrilases. The intermediate is further hydrolyzed to the corresponding ketone while releasing NH_3 as a by-product. The acyl-enzyme is then hydrolyzed by the addition of H_2O, and finally liberated the carbonhydroxylic acid along with the regenerated enzyme (Thimann and Mahadevan, 1964).

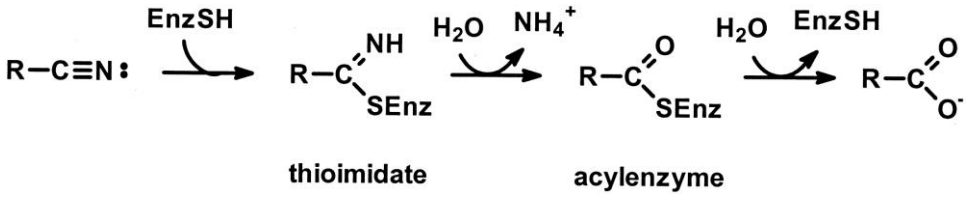


Figure 1–3. The proposed mechanism of nitrilases-mediated bio-catalysis.

Nitrile carbon was subjected to nucleophilic attack, by an SH group on the nitrilases. The resulting intermediate was then hydrolyzed to the corresponding ketone while releasing an NH_3 molecule. Acyl-enzyme was then hydrolyzed by a second H_2O molecule, and finally liberated the carboxylic acid along with the regenerated enzyme. Adopted from (Pace and Brenner, 2001) .

In nature, nitrilase enzymes are widely expressed in plants, animals, fungi and certain prokaryotes (Gong *et al*., 2012; Heinemann *et al*., 2003; Petříčková *et al*., 2011; Vorwerk *et al*., 2001). The majority of the studied nitrilases were obtained from bacteria, fungi and plants, by doing screens with media containing nitriles as nitrogen sources (Martinkova *et al*., 2008; O'Reilly and Turner, 2003), or through direct cloning and expression. Based on the substrate specificity, nitrilases from different organisms are classified into three broad categories, namely aromatic nitrilases, aliphatic nitrilases and arylacetonitrilase (Kobayashi and Shimizu, 1994). Although the physiological role of nitrilases is still unclear, several diverse roles were reported. Hsu and Camper discovered that nitrilases from *Fusarium solani*, a filamentous fungus, was able to break down the herbicides 3,5-dibromo-4-hydroxybenzonitrile (bromoxynil) and 3,5-diiodo-4-hydroxybenzonitrile (ioxynil) (Hsu and Camper, 1979). Besides detoxification, one strain of *Fusarium solani* was found to be able to use benzonitrile as the sole source of carbon and nitrogen (Harper, 1977). In plants, the nitrilases were reported being involved in the synthesis of indole acetic acid (Bartel and Fink, 1994; Bartling *et al*., 1992). Thimann and Mahadevan firstly reported in 1964 that the nitrilases hydrolyzing indole-3-acetonitile (IAN) to indole-3-acetic acid (IAA), also known as auxin, one of plant hormones/plant growth substances (Thimann and Mahadevan, 1964). Recent studies on plant nitrilases verified their cyanide detoxification function (Piotrowski *et al*., 2001). Over 3,000 plant species are cyanogenic, indicating that their hydroxynitrile lyase catalyzed the production of a cyanide group in the form of HCN, and the formation of corresponding aldehyde or ketone. When a plant is attacked, it releases HCN as a self-defense mechanism (Kassim and Rumbold, 2014). Cyanide is highly toxic; therefore, cyanide detoxification is

necessary for plant growth. Nitrilases were also reported participating in the degradation of glucosinolates (Bestwick *et al*., 1993), a typical secondary metabolite of almost all plants of *Brassicales*, such as mustard, cabbage, and horseradish (Agerbirk and Olsen, 2012). A family of nitrilases discovered in these glucosinolate-containing plants displayed rather broad substrate specificities and were able to hydrolyze nitriles that come from the degradation of glucosinolates (Janowitz *et al*., 2009). In industry, nitrilases can be used as catalysts to convert nitriles directly into corresponding acids and ammonia. The enzymes display high specificity and frequent enantioselectivity, and that makes them excellent biocatalysts for the production of fine chemicals and pharmaceutical intermediates (Thuku *et al*., 2009). The nitrilases are also used in the herbicide degradation and the detoxification of industrial waste (Banerjee *et al*., 2002). These enzymes function in aqueous solutions at moderate temperatures and pH, minimizing the costs of chemical processes and the negative impact of industry on the environment (Brady *et al*., 2006; Singh *et al*., 2006).

Nitrile hydratase (NHase, EC 4.2.1.84) family is another enzyme family which directly hydrolyze the CN bond. NHase catalyzes the formation of amides ($R\text{-}CONH_2$) from the nitriles (R-CN), following the further hydrolysis by amidase to the corresponding acids (R-COOH) and ammonia (NH_3). Nitriles are firstly converted to corresponding amides and then further hydrolyzed to corresponding acids by amidase. NHases are composed of equal molar of two subunits (named α and β subunits). NHases are metalloenzymes containing either cobalt or iron; hence the NHase are classified into two groups: ferrite NHase (Fe-NHase) and cobalt NHase (Co-NHase).

Ferrite NHases have a non-heme iron atom at the catalytic center and are characterized by their grayish green colors. Structural studies on Fe-NHase from *Rhodococcus* sp. N771 has shown that the iron (III) center is located at the interface of two subunits. As illustrated in Figure 1–4A, the sulfur atoms from three cysteine residues in the α subunit (αCys109, αCys112, and αCys114) and two amide nitrogen atoms of the main chain in αSer113 and αCys114 form ligands with iron (III) center on five vertices of an octahedron. The sixth position is occupied by an endogenous NO molecule.

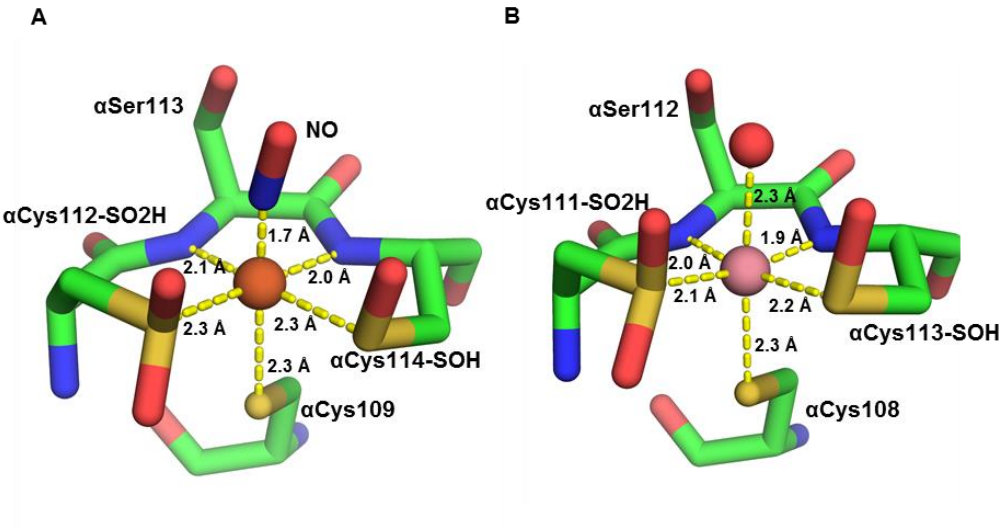


Figure 1–4. The structure of the active centers of Fe-type NHase in the inactive state and the active centers of Co-type NHase.

A. The structure of the active centers of Fe-type NHase in the inactive state (PDB ID: 2AHJ) from *Rhodococcus* sp. N771. ***B.*** The structure of the active centers of Co-type NHase (PDB ID: 3VYH) from *Pseudonocardia thermophilla.* The Fe^{3+} ion (orange), Co^{3+} ion (salmon), and water (red) are represented as spheres. The coordination of metal ions is presented as yellow dashes. The side chains of residues are colored by element, oxygen, red; nitrogen, blue; carbon; green.

The presence of the NO molecules blocks the access to solvent and prevents substrate-binding. Thus, the NHase is in an inactive state. After exposure to light, light stimulation causes the release of NO the molecule, and thus the activities of NHases are restored (Endo *et al*., 2001; Mascharak, 2002).

Cobalt NHases have a non-corrinoid cobalt atom at the catalytic center. The coordination of cobalt center is almost the same with that in ferrite NHase (Figure 1–4B). Studies one of the NHase from *P. thermophila* shows that ligands to the cobalt atoms include three sulfur atoms of the cysteine residues (αCys108, αCys111, and αCys113) and two amide nitrogen atoms (αSer112 and αCys113), forming an octahedral coordination with a water hydrogen atom (Miyanaga *et al*., 2001).

Although the catalyzing mechanism of NHase remains unclear, a possible reaction mechanism for NHase catalysis might be: (1) the metal-bond hydroxide ion acting as a nucleophile attacking the nitrile carbon atom when it approaches, or, (2) a water molecule is activated by metal-bond hydroxide ion, and it then attacks nitrile carbon. The intermediate finally tautomerizes to form an amide (Banerjee *et al*., 2002). The mechanism of NHase is becoming clear from multiple recent studies indicating that the cysteine-derived sulfenato ligand of the active site metal acts as the nucleophile in initially attacking the nitrile (Figure 1–5). The latest study on a cobalt-binding nitrile hydrates (Co-NHase) from *Streptomyces rimosus* further supported this proposal. Nelp and his colleagues used ^{18}O-labeled water under single turnover conditions, combined with high-resolution native protein mass spectrometry, to observe the hydrolysis reaction. Their study showed that that the incorporation of labeled oxygen into both product and

protein is turnover-dependent, and that only a single oxygen is exchanged into the protein even after multiple reactions (Nelp *et al*., 2016).

The biological function of NHases is unknown so far. However, studies showed that the presence of NHase enables the particular organism to utilize aliphatic, aromatic and heteroaromatic nitriles as the sole nitrogen source under laboratory conditions (Blakey *et al*., 1995; Layh *et al*., 1994). Due to their ability to hydrolyze cyano groups with selectively and efficiently, NHases are commonly used in the biotechnological industry, e.g. for the synthesis of the essential chemicals acrylamide (30,000 tons/year (Nagasawa and Yamada, 1995)) and nicotinamide (> 3500 tons/year (Shaw *et al*., 2003)). Also, their enzymatic activities are used to remove toxic nitriles, such as nitrile herbicides, during waste water treatment (Narayanasamy *et al*., 1990).

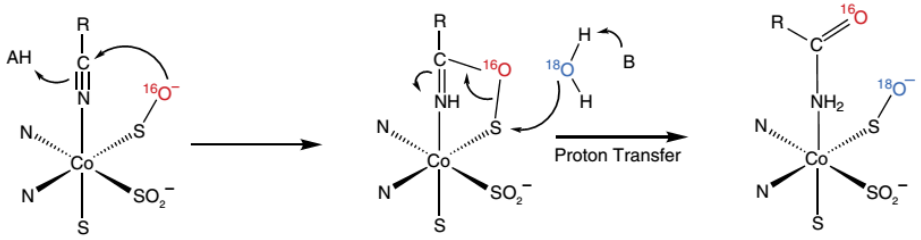


Figure 1–5. The proposed mechanism of Co-NHase.

Supported mechanism of Co-NHase in which the sulfenate oxygen acts as the primary nucleophile to attack the nitrile carbon. Bonds from enzyme to modified cysteine and amidate ligands are omitted for clarity. Adopted from (Nelp *et al*., 2016).

The majority of identified NHases were discovered from species belonging to the phyla *Proteobacteria*, *Actinobacteria*, *Cyanobacteria* and *Firmicutes*, which are living in soil (DiGeronimo and Antoine, 1976), coastal marine sediments (Langdahl *et al*., 1996), and deep sea sediments (Brandao and Bull, 2003; Layh *et al*., 1994) to geothermal environments (Pereira *et al*., 1998; Yamaki *et al*., 1997). Recently, Foerstner *et al*. described the discovery of the first eukaryotic NHase by using a large-scale screen in public sequence databases and metagenomics datasets (Foerstner *et al*., 2008). Unlike bacteria NHase, the two usually separated NHase subunits fused into one protein of *Monosiga brevicollis*, which is a recently sequenced unicellular model organism from choanoflagellate. Phylogenetic analyses and genomic context imply a possible ancient horizontal gene transfer (HGT) from proteobacteria (Foerstner *et al*., 2008). Then a second eukaryotic NHase was discovered in the stramenopile *Aureococcus anophagefferens*, and only the α subunit has been reported in *A. anophagefferens* NHase (Gobler *et al*., 2011).

Most recent studies revealed that nitrile hydratases exist in at least five eukaryotic supergroups: opisthokonts, amoebozoa, archaeplastids, CCTH (haptophytes and relatives) and SAR (comprising the stramenopiles, rhizarians, and alveolates). The majority of eukaryotic NHase are encoded by α and β subunit fusion genes. However, NHases containing only one subunit have been identified. An individual α subunit is found in a dinoflagellate, and an individual β subunit is found in a haptophyte (Marron *et al*., 2012).

1.2. Discovery and Identification Of the HD Domain

CAH was identified containing a cyanamide hydratase domain which belongs to the HD domain superfamily due to the presence of conserved histidine (H) and aspartate (D) residues in the motif (Aravind and Koonin, 1998). The most conserved histidine and aspartate residues in HD domain display a characteristic …H…HD….D… pattern. The doublet HD residues are the signature of an HD domain. HD domains superfamily represents a variety of enzymes, the majority of them exhibiting metal-dependent phosphohydrolase activity (Aravind and Koonin, 1998).

Structural studies show that the conserved His and Asp residues coordinate one or more metals which are required for enzymatic activity (Beloglazova *et al*., 2011; Brownt *et al*., 2006; Singh *et al*., 2015; White *et al*., 2013; Zhang *et al*., 2010). By sequence conservation and motif architectures, HD domain proteins were classified into 17 clusters (Aravind and Koonin, 1998). Well-characterized members in the HD superfamily include bacterial dGTPase, the guanosine polyphosphate hydrolase RelA/SpoT family, and the cyclic-NMP phosphodiesterases (PDEs). Studies of HD proteins indicate that many exhibit a common function of divalent metal-dependent phosphohydrolase. Characterized members are involved in nucleic acid metabolism and cellular signal transduction (Aravind and Koonin, 1998; Kondo *et al*., 2004; Kondo *et al*., 2007; Yakunin *et al*., 2004)

1.3. Structurally Characterized HD Domain Proteins

1.3.1. HD Domain Deoxyribonucleoside Triphosphate (dNTP) Triphosphatases (dNTPase)

The dNTPase enzyme specifically hydrolyzes deoxyribonucleoside triphosphates (dNTPs) to deoxynucleoside and triphosphate (PPP_i). Based on substrate specificity, the triphosphohydrolases can be divided into two broad groups. One group includes dGTPase from *E. coli* and PA1124 from *Pseudomnas aeruginosa*, both of which have specificity for dGTP. The other group includes TT1383 from *Thermus thermophilus*, EF1143 from *Enterococccus faecali*s, and PA3043 from *P. aeruginosa*, which have the ability to hydrolyze various dNTPs (Kondo *et al*., 2004; Mega *et al*., 2009; Seto *et al*., 1988; Vorontsov *et al*., 2011).

Deoxyguanosine triphosphate triphosphohydrolase (dGTPase) (EC 3.1.5.1) was firstly reported from *E. coli* in 1958 (Kornberg *et al*., 1958). The enzyme encoded by *dgt* gene was purified based on its high affinity for single-stranded DNA (ssDNA). The protein consists of 504 amino acids, a search of sequence databases shows that genes encoding homologs of dGTPase widely exist among almost all eubacteria. The enzyme cleaves dGTP into deoxyguanosine and tripolyphosphate (PPP_i) (Beauchamp and Richardson, 1988). Of the four dNTPs, the enzyme is highly specific for dGTP hydrolysis with a K_M of approximately 2 μM. The enzyme can hydrolyze neither dGDP nor dGMP. GTP is the second best substrate of the enzyme, with a K_M of about 110 μM. In the standard assay, the rate of GTP hydrolysis is only 3% of that of dGTP hydrolysis (Beauchamp and Richardson, 1988). The Mg^{2+} is required for optimal dGTPase activity, the presence of Mn^{2+} and Co^{2+} also allows the dGTP hydrolysis proceed at 30% and 17%

of the maximum rate (Huber *et al*., 1988; Seto *et al*., 1988). Deepa and his colleagues obtained the crystal structure of *E. coli* dGTPase (Dgt) bound to DNA molecules. The structure model revealed that the enzyme existed as a hexamer, containing two DNA molecules. Conserved histidine and aspartate residues, H69, H117, and D268, coordinate a Ni^{2+} ion, in a model of dGTPase without DNA (Singh *et al*., 2015). It is noticed that the aspartate residue (D118) of the HD doublet does not interact with the metal ion.

Studies on other bacterial dNTPase (TT1383, PA1124, PA3043, and EF1143) have revealed that the enzymes were present as a hexamer, except EF1143 existing as a tetramer (Kondo *et al*., 2007; Mega *et al*., 2009). The study on a complex of EF1143 with its substrate (dATP bound at the active site) and its activator (dGTP bound at the regulatory site) indicated an allosteric mechanism in which a dNTP bound at regulator site promoted the binding of a substrate dNTP at the catalyzing site (Vorontsov *et al*., 2011).

Sterile-Alpha Motif (SAM) domain and HD domain-containing protein 1 (SAMHD1) is a dGTP-activated deoxynucleoside triphosphate (dNTP) triphosphohydrolase (dNTPase) found in humans. It is proposed that SAMHD1 is expressed in dendritic cells and it restricts the replication of human immunodeficiency virus typ1 (HIV-1) by hydrolyzing the majority of cellular dNTPs to inhibit the reverse transcription and viral complementary DNA (cDNA) synthesis (Goldstone *et al*., 2011). The protein has 626 amino acids and contains two domains as mentioned above. Residues 45 to 110 consist of the SAM domain, and a region from residues 164 to 319 is characterized as an HD domain (White *et al*., 2013).

The human SAMHD1 protein has dNTP triphosphatase activity, specifically dGTP-stimulated dNTP triphosphohydrolase activity. Its binding affinity to single-stranded DNA (ssDNA) and ssRNA has been observed (Goncalves *et al*., 2012; Seamon *et al*., 2015). The nuclease activity of SAMHD1 against ssRNA has been reported as in an exogenous phosphate-dependent manner, indicating the protein catalyze the cleavage of RNA via a phosphorolysis mechanism instead of a hydrolytic mechanism (Ryoo *et al*., 2016). And the nuclease activity.is associated with its HD domain (Beloglazova *et al*., 2013). DNA-binding is required for the oligomerization of the protein which is essential for its dNTPase activity (Tüngler *et al*., 2013; Yan *et al*., 2013). It hydrolyzes nucleotide triphosphates (NTPs) to triphosphate and a nucleoside (Goldstone *et al*., 2011). Besides being responsible for SAMHD1's phosphohydrolase activity, the C-terminus HD domain is confirmed to be essential for protein to oligomerize (Goldstone *et al*., 2011). The presence of the HD domain also contributes to the inhibition of a broad range of retroviruses (White *et al*., 2013). The crystal structure of the enzyme revealed that a zinc ion is coordinated by two conserved histidine residues and two conserved aspartate residues in the HD motif (Zhu *et al*., 2013).

1.3.2. Clustered Regularly Interspaced Short Palindromic Repeat (CRISPER)-Associated Protein 3 (Cas3)

Clustered regularly interspaced short palindromic repeat (CRISPER) loci are composed of short DNA repeat sequences separated by stretches of variable spacer sequence that are derived from viral and plasmid DNA. CRISPER loci are located near CRISPER-associated (Cas) gene that works with RNA transcribed from CRISPER loci,

mediating the resistance pathway. Although Cas gene clusters are diverse in number, sequence, and organization, CRISPER/Cas systems are grouped into three types. Cas3 is the defining gene in type I system, the most popular presence in living organisms. A typical Cas3 protein is composed of an N-terminal HD domain and a C-terminal superfamily two helicase domain (Beloglazova *et al*., 2011). The HD domain is identified having endonuclease activity on ssDNA. The C-terminal helicase domain is responsible for the unwinding of dsDNA and DNA-RNA duplex. Structural and biochemical studies of the HD domain of *T. thermophillus* Cas3 reveal that the protein functions as a metal-dependent ssDNA nuclease, and a Ni^{2+} ion is bound at the active site of the protein (Mulepati and Bailey, 2011). *Methanocaldococcus jannaschii* expresses two Cas3 proteins: a Cas3 HD domain protein MJ0384, and a Cas3 helicase MJ0383. Cas3 HD protein MJ0384 is found to have both endonuclease and exonuclease (3'-5') activities on ssDNAs and RNAs. The structural study on Cas3 protein MJ0384 revealed the active site containing two metal ions. Evidence suggests that the Cas3 HD nucleases MJ0384 works together with the Cas3 helicase protein (MJ0383) for complete degradations of invasive DNAs (Beloglazova *et al*., 2011).

1.3.2.1. HD-GYP Domain Cyclic-Di-GMP (c-di-GMP) Phosphodiesterases (PDEs)

Bis-(3′-5′)-cyclic-di-GMP (c-di-GMP) is a near ubiquitous bacterial signaling molecule involved in the regulation of a variety of processes including virulence, motility, and biofilm formation (Römling *et al*., 2013). C-di-GMP is not observed being used by archaea, and only been found in eukaryotes in *Dictyostelium* (Chen and Schaap, 2012). In structure, c-di-GMP comprises two GMP molecules connected by 5′ to 3′ phosphodiester

bonds. It is synthesized from GTP by diguanylate cyclase (DGCs) (EC 2.7.7.65) containing GGDEF domain. It is hydrolysed to linear form 5′-pGpG, by c-di-GMP phosphodiesterases (PDEs) (EC 3.1.4.52) containing either EAL or HD-GYP domain (Hengge, 2009). However, HD-GYP PDEs further hydrolyze 5′-pGpG to two GMP molecules (Stelitano *et al*., 2013). The HD-GYP domain is named after the conserved residues in the domain, the consensus sequence of GYP motif is HHExxDGxGYP (Ryan and Dow, 2010). The HD-GYP motif, which constitutes a subset of HD superfamily, makes up more than 30% of all predicted c-di-GMP PDEs. In some phyla such as *Thermotogae* and *Spirochaetes*, HD-GYP PDEs are the predominant c-di-GMP PDEs (Römling *et al*., 2013).

Andrew and his colleagues obtained the first crystal structure of HD-GYP domain protein Bd1817 from predatory bacterium *Bdellovibrio bacteriovorus* (Lovering *et al*., 2011). Though purified Bd1817 was inactive, and it missed the conserved tyrosine (Tyr, or Y) residue in the GYP motif, the crystal structure was still instructive for showing the architecture of the conserved domain. The structural model revealed that the domain was mainly comprised of helices, and the active site of HD-GYP domain contained a non-heme di-iron site coordinated by histidine, aspartate and glutamate residues (Lovering *et al*., 2011).

The second structural study of HD-GYP protein was performed with *Pm*GH, which was a functional c-di-GMP specialized PDE from *Persephonella marina*. The structural model revealed that *Pm*GH presents as a dimer, and each monomer contained three irons forming a triangle. The mutagenesis of the eight metal-interacting residues

either abolished or significantly reduced the enzymatic activity, indicating that all three metal ions are necessary for the catalysis (Bellini *et al*., 2014).

The crystal structure of a third HD-GYP protein, PA4781 from *P. aeruginosa* showed occupancy by two nickel ions, which was different from other identified HD-GYP proteins. Moreover, PA4781 was reported to bind various divalent metals with similar affinities and was recognized as a pGpG-preferring PDE (Rinaldo *et al*., 2015).

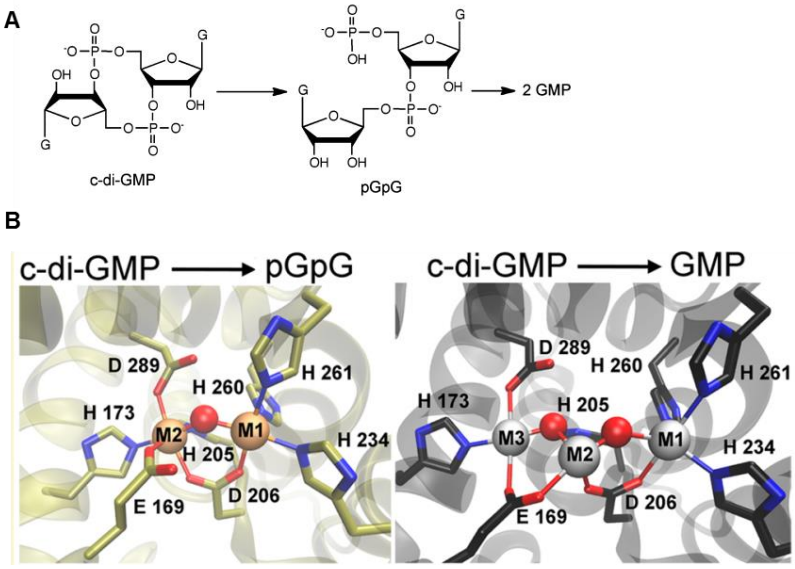


Figure 1–6. Scheme of c-di-GMP PDE activity and metal occupancy of TM0186 HD-GYP domain.

A.Scheme of c-di-GMP PDE activity. ***B.*** Metal occupancy of TM0186 HD-GYP domain. The left panel shows di-metal site in the TM0186 HD-GYP domain structural model. Right panel illustrates tri-metal site in the TM0186 HD-GYP domain structural model. Metal coordination is represented as red lines. The side chains of residues are colored by element, oxygen, red; nitrogen, blue. The metal (gray and tan) and the water molecules (red) are represented as spheres. Adopted from (Miner and Kurtz, 2016).

The most recent crystal structure of an HD-GYP protein was of c-di-GMP PDE TM0186 from *Thermotoga maritima*. The structural study showed that TM0186 can accommodate both di- and tri-metal active sites (Miner and Kurtz, 2016). Combined enzymatic assays proved that the presence of di-iron at the active site was necessary and sufficient to hydrolyze c-di-GMP to pGpG, but the further conversion of pGpG to GMP required the tri-iron active site (Figure 1–6). The study also showed that a glutamate residue conserved in a subset of HD-GYPs is needed for the formation of the tri-metal and can also serve a labile ligand to the bimetal site. In all these HD-GYP proteins, the residues of the HD motif provide a carboxylate bridge between two metal ions. However, the conserved GYP residues have not been found having direct interactions with the metals in any HD-GYP protein (Miner and Kurtz, 2016).

1.3.3. HD Domain Oxygenase

A special subgroup of the HD family is the di-iron oxygenases. The first identified member of this class was *myo*-insitol (MI) oxygenase (MIOX; EC 1.13.99.1),

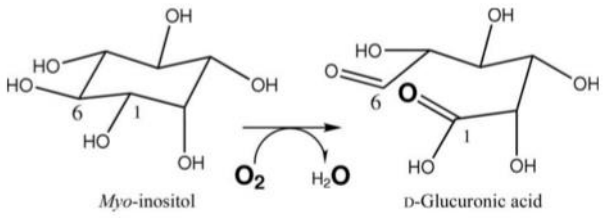


Figure 1–7. The reaction catalyzed by *myo*-inositol oxygenase (MIOX).

MIOX is an enzyme containing a non-heme di-iron center that oxidizes *myo*-inositol to glucuronic acid Adopted from (Brown *et al*., 2006).

which catalyzes the oxidative conversion of MI to D-glucuronic acid (Brownt *et al*., 2006). As shown in Figure 1–7, MIOX breaks the bond between C6 and C1 of MI to form a glucuronic acid (Brown *et al*., 2006; Moskala *et al*., 1981).

Diabetes mellitus (diabetes) is a chronic disease characterized by hyperglycemia caused by the defective action and or secretion of insulin. Evidence shows that both type-1 and type-2 diabetes are connected with altered inositol metabolism, especially of *myo*-inositol (MI) and its epimeric form, D-chiro-inositol (DCI) (Asplin *et al*., 1993; Kawa *et al*., 2003; Ostlund Jr *et al*., 1993). The MIOX is a key regulator of inositol levels by catalyzing the first committed step in the glucuronate–xylulose pathway (Charalampous, 1959; Moskala *et al*., 1981), the only known pathway for MI catabolism (Hankes *et al*., 1969). Moreover, MIOX also acts on DCI and potentially mediates its catabolism as well (Arner *et al*., 2001). MIOX is a 33 kDa iron-containing protein exclusively expressed in kidney, and its amino acid sequence remains highly conserved among species (Arner *et al*., 2006).

Brown and his colleagues presented crystal structure of mouse MIOX, in complex with MI (Brownt *et al*., 2006). The 2.0-Å resolution structural model revealed that the MIOX existed as monomeric, single-domain protein, with a mostly helical fold. Among the total 9 helices, five helices formed the structural core. Six histidine and aspartate residues in the HD motif provided 6 ligands for the di-iron center, in which the two iron atoms are bridged by a putative hydroxide ion and one of the ligands provided by an aspartate residue. Also, the MI substrate is found to be bound to one of the iron atoms. Different from other di-iron oxygenases, MIOX has a mixed-divalent di-iron pair Fe^{2+}/Fe^{3+} involved in catalysis (Xing *et al*., 2006c). The proposed catalytic mechanism is

that Fe^{III} firstly coordinates with MI, serving as a Lewis acid to activate the substrate, then the other Fe^{2+} activates dioxygen to produce a superoxide-Fe^{3+}/Fe^{3+} intermediate, which initiates the reaction by abstracting a hydrogen atom from C1 of MI (Xing *et al*., 2006a; Xing *et al*., 2006b).

MIOX is the first identified member of a special subgroup in HD domain superfamily, named as mixed-valent di-iron oxygenases (MVDO). Different from identified HD domain phosphohydrolases, this subgroup was identified as di-iron oxygenase. The second representative of this class is PhnZ discovered in marine bacteria that catalyzes the oxidative cleavage of a carbon-phosphorus bond (C-P bond) of 2-amino-1-hydroxyethylphoaphonate, resulting in the formation of glycine and phosphate (Wörsdörfer *et al*., 2013).

1.4. Structural Features of the HD Domain

1.4.1. The HD Domain Is αn-Helix Bundle

Published crystal structures of HD domain proteins in RCSB protein data bank (RCSB PDB, www.rcsb.org.) (Berman *et al*., 2000) have been listed below (Table 1-1). They are mainly divided into two classes respective to their enzymatic functions. The majority of HD domain proteins belong to the phosphohydrolase family; the rest belong to oxygenases. Secondary structure analysis reveals that all these proteins display a helical structure; helical region covers 40% to 60% of the lengths of the proteins, while β-sheet region covers less than 10% of the proteins. These metalloenzymes prefer recruiting

Table 1-1. Summary of published structures of HD domain proteins

Protein	Function	PDB ID	Metal cofactor	Percentage of helical region	Percentage of beta-sheet region	Ref.
Phosphohydrolases						
CRISPR-Cas3 (HD motif)	nuclease	3SKD 3SK9	Ni^{2+}	55% (12 helices; 148 residues)	5% (6 strands; 14 residues)	(Mulepati and Bailey, 2011)
EF1143	dNTPase	3IRH 2O6I	Ca^{2+}	52% (33 helices; 253 residues)	10% (15 strands; 51 residues)	(Vorontsov *et al*., 2011)
SAMHD1 (HD motif)	dNTPase	3U1N	Zn^{2+}	43% (27 helices; 229 residues)	8% (16 strands; 45 residues)	(Goldstone *et al*., 2011)
dGTPase	dGTPase	4XDS 4X9E	Ni^{2+} Mg^{2+}	61% (25 helices; 313 residues)	1% (6 strands; 9 residues)	(Singh *et al*., 2015)
YfbR	5'-deoxy-ribonucleotidase	2PAQ 2PAR 2PAU	Co^{2+}	59% (12 helices; 120 residues)	0% (2 strands; 2 residues)	(Zimmerman *et al*., 2008)
YpgQ	nucleotide pyrophosphohydrolase	5DQV 5DQW 5IHY	Ni^{2+}	63% (10 helices; 134 residues)	0 % (0 residue)	(Jeon *et al*., 2016)
PA4781(residue 151-368)	c-di-GMP phosphodiesterase	4R8Z	di-Ni^{2+}	64% (14 helices; 140 residues)	1% (4 strands; 4 residues)	(Rinaldo *et al*., 2015)
*Pm*GH	c-di-GMP phosphodiesterase	4MCW 4MDZ 4ME4	tri-iron	57% (21 helices; 211 residues)	11% (10 strands; 43 residues)	(Bellini *et al*., 2014)
PgpH (HD domain)	c-di-AMP phosphodiesterase	4S1C 4S1B	di-iron	65% (14 helices; 146 residues)	1% (3 strands; 3 residues)	(Huynh *et al*., 2015)
Bd1817	inactive c-di-GMP phosphodiesterase	3TM8 3TMB 3TMC 3TMD	di-iron	56% (18 helices; 184 residues)	7% beta sheet (8 strands; 25 residues)	(Lovering *et al*., 2011)

RelSeq (bifunctional catalytic fragment)	(p)ppGpp synthesis and hydrolysis	1VJ7	Mn^{2+}	51% (19 helices; 202 residues)	7% (8 strands; 29 residues)	(Hogg *et al*., 2004)
Oxygenases						
PhnZ	oxygenases	4N6W 4N71	di-iron	54% (14 helices; 125 residues)	0% (2 strands; 2 residues)	(Wörsdörfer *et al*., 2013)
MIOX	*myo*-inositol oxygenase	3BXD 2IBN	di-iron	47% (16 helices; 137 residues)	2% (4 strands; 6 residues)	(Thorsell *et al*., 2008)

iron as metal cofactors, appearing in the form of non-heme di-iron or tri-iron of mixed oxidation state. Second most frequently recruited metal ion is nickel, present forms can be either single nickel anion or di-nickel. Other metal cofactors include zinc, calcium, cobalt, and manganese. Almost all the coordinated metals are in the form of a divalent cation, except iron, which can be present as either Fe^{2+} or Fe^{3+} in the protein.

1.4.2. Conserved HD Residues in the HD Motif Are Coordinating Metals

Crystallographic studies on functional HD domain proteins reveal that the conserved HD residues bind to divalent metal ions, Fe^{2+}/Fe^{3+}, Mg^{2+}, Ni^{2+}, Co^{2+} and Zn^{2+}, and substrates interact with the surrounding amino acids (Brown *et al*., 2006; Kondo *et al*., 2007). Mutations in the conserved HD residues drastically affect the enzymatic activity (Zimmerman *et al*., 2008). It indicates that the metals are likely involved in the

Table 1-2. Summary of metal ion coordination residues in published HD domain protein structures.

Protein	PDB ID	Metal cofactor	Metal-coordination residues
Phosphohydrolases			
CRISPR-Cas3 (HD motif)	3SKD 3SK9	Ni^{2+}	H24, **H69**, **D70**, and D205
EF1143	3IRH 2O6I	Ca^{2+}	H66, **H110**, **D111**, and D183
SAMHD1 (HD motif)	3U1N	Zn^{2+}	H67, **H206**, **D207**, and D311.
dGTPase	4XDS 4X9E	Ni^{2+} Mg^{2+}	H69, **H117**[a], and D268
YfbR	2PAR 2PAQ 2PAU	Co^{2+}	H33, **H68**, **D69**, and D137
YpgQ	5DQV 5DQW 5IHY	Ni^{2+}	H29, **H58**[a] , and D124
PA4781 (residue 151-368)	4R8Z	di-Ni^{2+}	Ni1: H180, **H220**, and **D221** Ni2:**D221**, H249, H281, and H282
*Pm*GH	4MCW 4MDZ 4ME4	tri-iron	Fe1: E185, H189, and D305 Fe2: E185 and **D222**[a], Fe3: **D222**[a], H250, H276, and H277
PgpH (HD domain)	4S1C 4S1B	di-iron	Fe1: H514, **H543**, **D544** and D648 Fe2: **D544**, H580, H604 and H605
Bd1817	3TM8 3TMB 3TMC 3TMD	di-iron	Fe1: H150, **H183**, **D184**, and N265 Fe2: **D184,** H212, H237, and E238
RelSeq (bifunctional catalytic fragment)	1VJ7	Mn^{2+}	H53, **H77**[a], and D144
Oxygenases			
PhnZ	4N6W 4N71	di-iron	Fe1: **D59**, H80, and H104 Fe2: H34, **H58**, **D59**, and D161
MIOX	3BXD 2IBN	di-iron	Fe1: H98, **H123**, **D124**, and D253 Fe2: **D124**, H194 and H220

[a]The histidine (aspartate) residue belongs to the HD doublet pair, but the other aspartate (histidine) residue is not interacting with the metal ion.

Information on metal interactions with residues was retrieved from PDB. The HD doublet residues are shown in bold.

catalytic process and/or substrate binding. A summary of residues interacting with a metal cofactor in HD domain proteins is presented in Table 1-2.

From the table, it is observed that all residues coordinating metal ions are histidine and aspartate residues, except in protein *Pm*GH and Bd1817, which additionally recruit a glutamate residues and an asparagine residue in iron coordination. In addition, in HD proteins those recruit only one metal ion, such as SAMHD1 (Zn^{2+} coordinated), CRISPER-Cas3 (Ni^{2+} coordinated), EF1143 (Ca^{2+} coordinated), YfbR (Co^{2+} coordinated), the metal-binding residues can be shown as H...HD…D. However, exceptions are observed in the proteins YpgQ (Ni^{2+} coordinated), RelSeq (Mn^{2+} coordinated) and dGTPase (Ni^{2+} coordinated), residues binding to metal can be represented as H…H…D, the aspartate residue of doublet HD is seen in the sequence, but it does not interact with the metal ion.

Figure 1–8 is displaying the superposition of structural models of HD motifs of HD domain proteins YpgQ (PDB ID: 5DQV), SAMHD1 (PDB ID: 3U1N), and EF1143 (PDB ID: 3IRH), to present the conserved architecture of the HD domain core. The above three proteins contain only one metal ion and they all belong to phosphohydrolases. The overlaid model indicates that the core region of HD motif is composed of 5 α-helices, named as αA to αE. The conserved HD residues that are interacting with metal ions are located at αA, αB and αE.

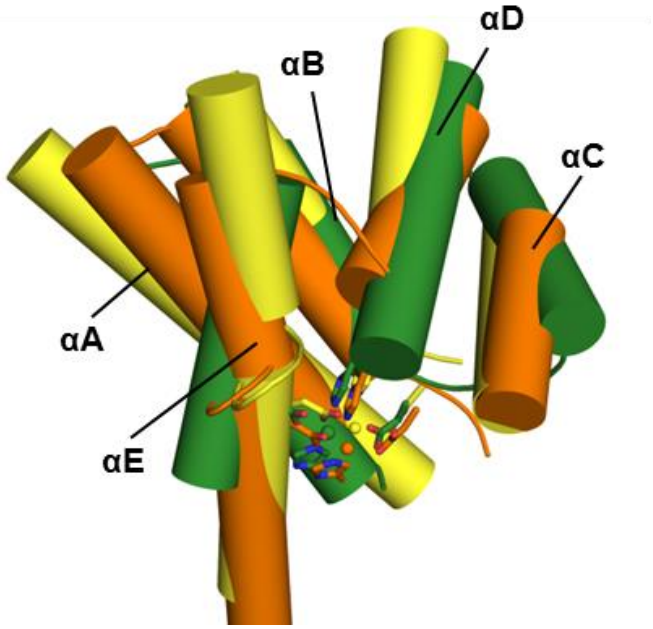


Figure 1–8. Superposition of HD motifs of HD domain proteins YpgQ, SAMHD1 and EF1143 illustrating the conserved helices architecture.

Superposition of HD motif of HD proteins YpgQ (green, PDB ID: 5DQV, residue 25-130), SAMHD1 (yellow, PDB ID: 3U1N, residue 160-320), and EF1143 (orange, PDB ID: 3IRH, residue 59-197), indicates that the core of HD motif consists five α-helices, αA-αE. Metal cofactors of each protein are represented as spheres. Side chains of HD residues that interact with metal ions are colored by element, oxygen: red; nitrogen; blue. The figure was drawn with Pymol.

Figure 1–9A is showing the structural model of HD motif in protein YpgQ (PDB ID: 5DQV, residue 25-130), illustrating the conserved helices (αA to αE) in the domain. YpgQ displays nucleotide pyrophosphohydrolase activity, it hydrolyzes (deoxy)ribonucleoside triphosphate [(d)NTP] to (deoxy)ribonucleoside monophosphate and pyrophosphate using the HD domain. The Ni^{2+} ion is coordinated by the three conserved HD residues, which can be presented as H…H…D. The first histidine residue is located at the starting of the αA, the HD doublet residues (H58D59) are located at the end of the αB, but only the histidine interacts with the metal ion. The two helices (αA and αB) are roughly paralleled, following are another two helices (αC and αD) that lie above the αB. The other aspartate residue bound to metal ion is located at the last α-helix (αE). Figure 1–9B is showing the structure model of HD motif of YfbR (PDB ID: 2PAR, residue 30-137), which also coordinates one metal ion and displays phosphohydrolase activity.YfbR is a 5′-nucleotidase (5′-NT) against to 2′-deoxyribonucleotide-5′-monophosphates (dNMPs) and does not discriminate among particular nucleotide bases (Zimmerman *et al*., 2008). HD motif of the YfbR contains only four conserved helices instead of five.. But the HD motif in YfbR is composed of four helices instead of five helices in YpgQ. The three helices (αA, αB, and αE) are also found in other HD domain proteins, it also contains an α-helix (αC′) which is in a different orientation of αC. It is noticed that the αD is missing in HD motif of YfbR (Figure 1–9B). However, the YfbR shares the same architecture of metal center, the Co^{2+} is bound to conserved HD residues, which are located at αA, αB and αE, respectively

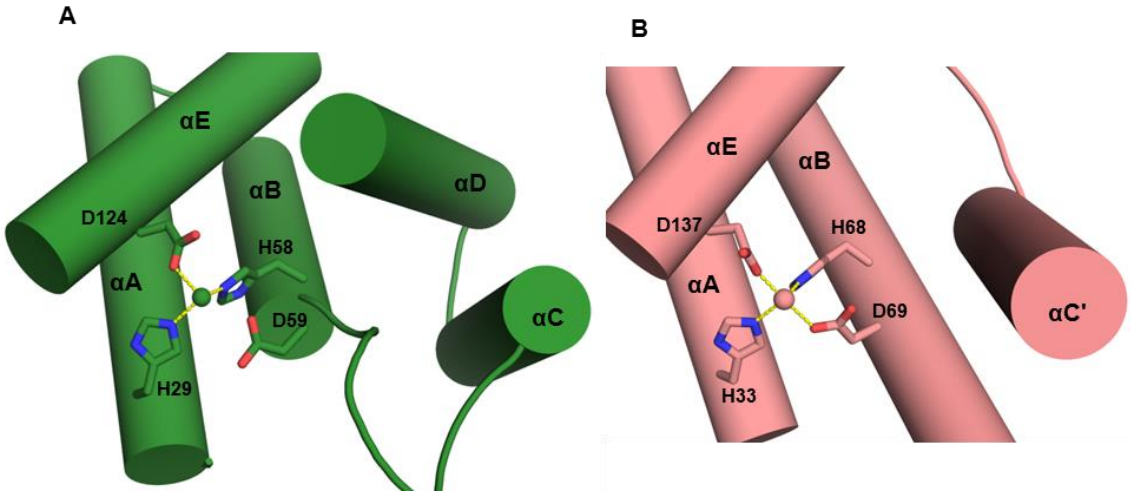


Figure 1–9. Structural models of HD motifs of YpgQ and YfbR.

A. The structural model of HD motif of YpgQ (PDB ID: 5DQV, residue 25-130). HD residues coordinating Ni^{2+} are located at αA, αB and αE, respective. Ni^{2+} ion is represented as a green sphere. Metal coordination is shown as yellow dashes. ***B***. The structural model of HD motif of YfbR (PDB ID:2PAR, residue 30-137). HD residues coordinating Co^{2+} are located at αA, αB and αE, respective. Co^{2+} ion is represented as a salmon sphere. Metal coordination is shown as yellow dashes. The side chains of HD residues that interact with the metal ion are colored by element, oxygen: red; nitrogen; blue. The figure was drawn with Pymol.

In proteins holding a di-metal center, such as PA4781 (di-nickel coordinated), PhnZ (di-iron coordinated), PgpH (di-iron coordinated), and MIOX (di-iron coordinated), one of the two ions is coordinated by H…HD…D residues. And the aspartate residue in the HD doubl*et al*ways bridges the other metal ion with another two histidine residues, in oxygenases PhnZ, MIOX; or with another three histidine residues, in phosphodiesterases PA4781 and PgpH. Superposition of HD motif of HD domain proteins that contain multiple metal ions indicates the same conserved helical architecture, as shown in Figure 1–10. Overlaid model of YpgQ (Ni^{2+} coordinated, PDB ID: 5DQV), PA4781 (di-Ni^{2+}

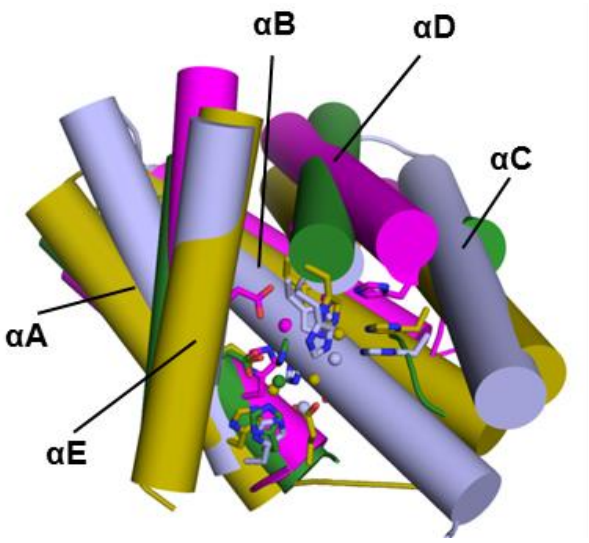


Figure 1–10. Superposition of HD motifs of HD domain proteins that contain multiple metal ions to illustrate the conserved helices architecture.

Superposition of HD motif of YpgQ (green, PDB ID: 5DQV, residue 25-130), PA4781 (light blue, PDB ID: 4R8Z, residue 180-310), *Pm*GH (olive, PDB ID: 4MCW, residue 182-319), and PhnZ (magenta, PDB ID: 4NCW, residue 25-110) indicates that the core of HD motif consists five α-helices, αA-αE. Metal cofactors of each protein are represented as spheres. Side chains of HD residues that interact with metal ions are colored by element, oxygen: red; nitrogen; blue. The figure was drawn with Pymol.

coordinated, PDB ID: 4R8Z), *Pm*GH (tri-iron coordinated, PDB ID: 4MCW), and PhnZ (di-iron coordinated, PDB ID: 4NCW) shows the overlapped five α-helices, αA-αE. Compared to the superposed model showing in Figure 1–8, this overlaid model of proteins containing multiple metals has elongated αB and αC.

Figure 1–11 is showing the structural models of two HD proteins which contain multiple metal ions, PhnZ (PDB ID: 4N6W, residue 25-106,149-166) and *Pm*GH (PDB ID: 4MCW, residue 182-319), respectively. Both the two proteins display the conserved five helices in the motifs. PhnZ contains a di-iron center; residues holding one of the metals (M1) are located at αA, αB and αE (Figure 1–11A), same as that in YpgQ shown in Figure 1–9A. Residues that hold the other metal ion (M2) are located at αC and αD. Similarities are seen in the model of HD motif in *Pm*GH (PDB ID: 4MCW, residue 182-319), HD residues bound to the third metal ion (M3) are located at αC and αD (Figure 1–11B), respectively. PmGH is the only one which is characterized containing three metal ions, the histidine residue of the doublet HD residues is not interacting with the metal ions. In addition, one glutamate residue is recruited in iron coordination, and it also performs as a bridge two metal ions (Fe1 and Fe2) as well as the aspartate residue in doublet HD sites (bridging Fe2 and Fe3). Besides the aspartate residue, Fe3 is coordinated by another three histidine residues including two adjacent histidine residue (H276 and H277), this pattern was also observed in proteins which have a di-metal center, such as PA4781 (di-Ni^{2+} coordinated) and PgpH (di-iron coordinated). It is noticed that the histidine in HD doublet of protein *Pm*GH is not interacting with any of the tri-iron ions.

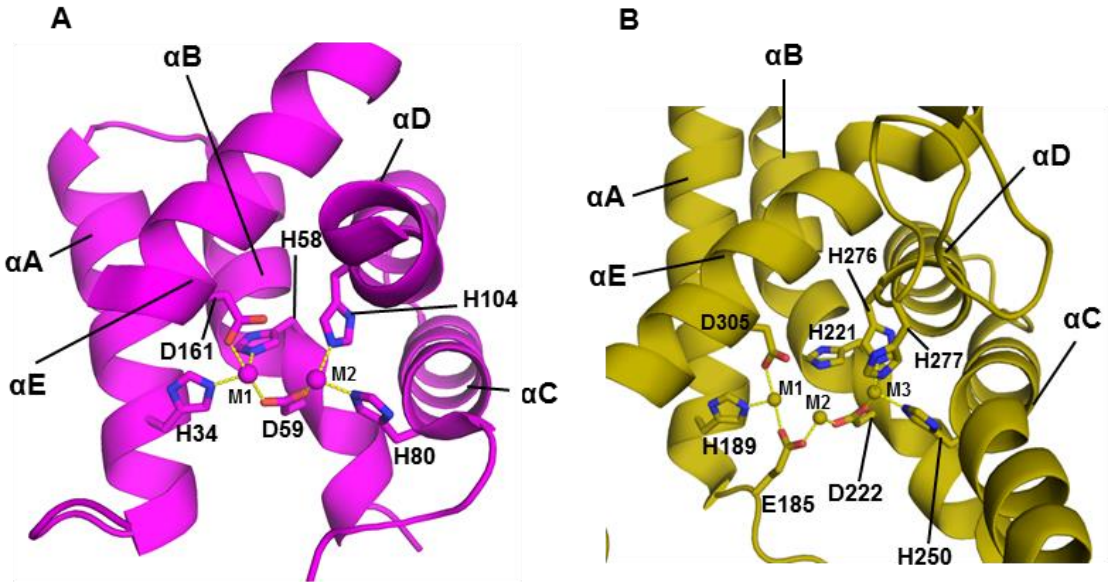


Figure 1–11. The structural models of HD motifs of PhnZ and *Pm*GH.

A. The structural model of HD motif of PhnZ (PDB ID: 4N6W, residue 25-106,149-166). Di-iron ions are represented as spheres and labeled as M1, M2. Metal coordination is shown as yellow dashes. ***B***. The structural model of HD motif of *Pm*GH (PDB ID:4MCW, residue 182-319). Tri-iron metal ions are represented as spheres and labeled as M1, M2 and M3. Metal coordination is shown as yellow dashes. The side chains of HD residues that interact with the metal ion are colored by element, oxygen: red; nitrogen; blue. The figure was drawn with Pymol.

1.5. Discovery of *DDI2* and *DDI3* Genes

DDI2 and *DDI3* are two genes identified in *S. cerevisiae* through a genome-wide microarray analysis of gene transcription levels in response to MMS, which is a typical DNA-methylating agent. In this microarray analysis, two open reading frames (ORFs) *YNL335W* and *YFL061W* displayed the highest fold induction (> 100-fold) among the entire transcripts after 0.1% MMS treatment (Fu, 2008). Due to their properties, the ORFs were named DNA-damage inducible genes 2 and 3 (*DDI2* and *DDI3*) respectively , as *DDI1* had been previously reported. Ddi1 was reported being involved in numerous interactions with the ubiquitin system (White *et al*., 2011), and is known to be regulated by the DNA-damage checkpoint (Zhu and Xiao, 1998, 2001).

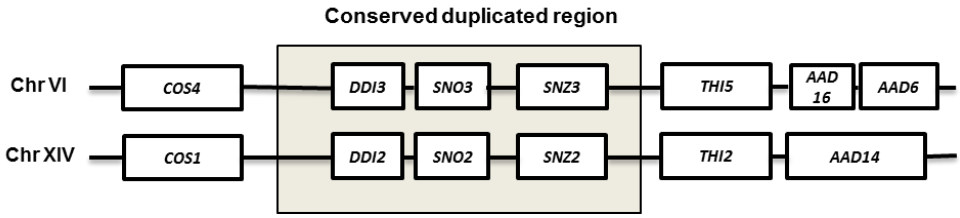


Figure 1–12.The diagram shows *DDI2* and *DDI3* genes are located within a highly conserved duplicated region.

DDI2 and *DDI3* and their flanking 20-kb regions are thought to derive from gene duplication, in which the boxed regions are highly conserved in DNA sequence.

Interestingly, these two ORFs are identical and are located in a board region of gene duplication, including their promoter regions (SGD, www.yeastgenome.org, Stanford University). (Figure 1–12). Besides *DDI2/3*, two downstream genes, *SNO2/3* and *SNZ2/3*, are also found in this duplicated region. However, Sno2 has two amino acids differences (M94 and E105) from Sno3 (L94 and A105). Functions of both Sno2 and Sno3 remains unknown, but their expression is known being induced before the diauxic shift, and in the absence of thiamin (Padilla *et al*., 1998; Rodríguez-Navarro *et al*., 2002). Snz2 has one amino acid (I266) difference from Snz3 (V266). *SNZ2* and *SNZ3* belong to stationary phase-induced gene family; their proteins are expressed in the presence of galactose. The transcription of *SNZ2(3)* is induced prior to the diauxic shift, and also in the absence of thiamin, through a Thi2p-dependent pathway. It is noticed that *SNZ2(3)* forms a co-regulated gene pair with *SNO2(3)* (Braun *et al*., 1996; Padilla *et al*., 1998; Rodríguez-Navarro *et al*., 2002).

1.6. DNA Damage-Inducible Genes in Budding Yeast *S. cerevisiae*

DNA-damage inducible genes are those whose transcriptional levels can be induced by DNA damage. The budding yeast *S. cerevisiae* is one of the simplest unicellular eukaryotic organisms, but it shares most complex internal cell structures with plants and animals. *S. cerevisiae* shares many of the technical advantages that permit rapid progress in the molecular genetics of prokaryotes, such as rapid growth, dispersed cells, the ease of replica plating and mutant isolation, a well-defined genetic system, a highly versatile DNA transformation system, various selective markers and powerful genetic manipulations. Therefore, *S. cerevisiae* is one of the model organisms used in the

study of DNA repair pathways in eukaryotes. A large number of genes involved in yeast DNA repair and tolerance pathways are identified as DNA-damaging inducible genes (Friedberg, 2006; Fu *et al*., 2008).

The *PHR1* gene encodes a DNA photolyase involved in photoreactivation, which directly removes pyrimidine dimers in the presence of visible lights (Schild *et al*., 1984). The transcriptional level of *PHR1* was reported to be induced by UV irradiation (Sebastian and Sancar, 1991), which mainly cause pyrimidine dimers and UV-mimetic agent 4-Nitroquinoline 1-oxide (4NQO). It is, also induced by methyl methanesulfonate (MMS) and methylnitronitrosoguanidine (MNNG), which are DNA alkylating agents; and also by bleomycin and cisplatinum, which halt DNA synthesis (Sancar, 2000; Sebastian *et al*., 1990).

The *MAG1* gene encodes a DNA glycosylase that specifically recognizes and cleaves 3-methyladnine produced by DNA alkylating damage, and initiates the base excision repair (BER) pathway (Chen *et al*., 1989). The transcription of *MAG1* is induced by a variety of DNA-damaging agents including DNA-alkylating agents, such as MMS and MNNG, oxidative agents. The gene is also induced by agents causing bulky adducts such as 4-nitroquinoline oxide (4NQO) and UV radiation (Chen and Samson, 1991), and also by hydroxyurea (HU), which results in depletion of deoxyribonucleotide (dNTPs) pools and promotes stalls at the replication forks (Liu and Xiao, 1997).

DNA damage-inducible gene 1 (*DDI1*) is located immediately upstream of the *MAG1* gene (Liu and Xiao, 1997) encoding a 3-methyladenine DNA glycosylase required for the base excision repair (NER) pathway (Chen *et al*., 1989). An interesting discovery is that the upstream activating sequence that regulates the transcription of *MAG1*

(UAS_{MAG1}) lies in protein coding region of *DDI1* (Liu *et al*., 1997), and *DDI1* was named because this gene was found being inducible by DNA-damaging agents and co-regulated with *MAG1* (Liu and Xiao, 1997). (Liu and Xiao, 1997; Zhu and Xiao, 2001).Subsequent studies reveal that Ddi1 contains ubiquitin-associated (UBA), ubiquitin-like (UBL), and retroviral-like proteinase (RVP) domains (Clarke *et al*., 2001). The function of Ddi1 remains unclear; studies reveals that the protein belongs to a family of shuttle proteins which are targeting polyubiquitinated substrates for proteasomal degradation, and it is confirmed that Ddi1 can bind to ubiquitin via its UBL domain (Nowicka *et al*., 2015). The presence of RVP domain is required for the repression of protein secretion, and it is proposed that Ddi1 functions as an aspartic protease *in vivo*. (White *et al*., 2011).

BER is a DNA repair process to remove mainly modified bases due to alkylation and oxidation. The damaged base is directly recognized by corresponding glycosylases, following steps including base excision, DNA synthesis, and ligation. Nucleotide excision repair (NER) is a more complex process to remove many DNA helix-distorting lesions such as UV or UV-mimetic agent-induced damage and chemical adducts. In *S. cerevisiae*, NER is mediated by Rad1, Rad2, Rad4, Rad7, Rad10, Rad14, Rad16, Rad23, Met18, the transcription factor TFIIH, and the heterotrimeric complex RPA (Rfa1, Rfa2, Rfa3). Together these proteins bind DNA lesions and make incisions on both sides of the damaged DNA, which releases a fragment of 25-30 bp (Hoeijmakers, 1993; Prakash *et al*., 1993). It is identified that expression of Rad2, which functions an endonuclease making the incision on single-strand NDA, is induced in response to UV in alpha-factor-arrested or stationary phase cells (Siede and Friedberg, 1992; Siede *et al*., 1989). Other NER genes including *RAD7* (Jones *et al*., 1990), *RAD16* (Bang *et al*., 1995) and *RAD23*

(Madura and Prakash, 1990) were reported to be induced at the transcriptional level after UV treatment.

The homologous recombination repair pathway is responsible for repairing DNA double-strand breaks; genes involved in this pathway belong to the *RAD52* epistasis group, including *RAD50, RAD51, RAD52, RAD54, RAD55, RAD56, RAD57, RAD58/MRE11, RAD59* and *XRS2* .(van den Bosch *et al*., 2002). Among these genes, the transcription of *RAD51* was reported to be induced by low-dose X-ray treatment (Basile *et al*., 1992). Moreover, a study on the expression of *RAD54* showed that it was inducible in response to DNA-damaging agents including X-rays, UV light and MMS (Basile *et al*., 1992).

Microarray chip was frequently used in genome-wide screening for DNA-damaging inducible genes as a high throughput study method. A DNA microarray chip is a high-density array of DNA spots that contain addressable complementary sequences to many or most genes in a genome. The chip is hybridized with fluorescently tagged nucleic acids, representative of expressed transcripts. Imaging and computational analysis are used to monitor changes in transcriptional levels of thousands of genes simultaneously (Gasch *et al*., 2001; Jelinsky and Samson, 1999; Lee *et al*., 2007).

Microarray studies provide an opportunity for thorough study of the DNA damage-induced transcription in budding yeast. By using a DNA microarray, Jelinsky and Samson reported that 325 genes of 6,218 ORFs (5.2%) increase more than 4-fold at the transcript level after exposing yeast cells to the typical alkylating agent MMS at 0.1% for 1 hour (h) (Jelinsky and Samson, 1999). These genes can be classified into several functional groups: DNA repair, cell cycle, signal transduction, cell wall biogenesis,

membrane transport, detoxification, and protein degradation. Obviously, besides genes involved in DNA repair and cell cycle regulation, genes participating in many other cellular processes are also regulated in response to DNA damage. The same group also studied the genomic transcriptional spectrum of yeast cells treated with different MMS doses and time courses (Jelinsky *et al*., 2000). More than 1,000 genes were found to be up-regulated in response to MMS treatment. Interestingly, many of the genes induced by MMS treatment were also induced in response to the arrest of cells in the stationary phase. Thus, there appears to be an overlap of responsive genes under two different stressful conditions, MMS exposure, and stationary growth arrest. This implies the existence of a general stress response pathway in budding yeast. Studies revealed that the DNA damage-inducible genes could be induced by a spectrum of DNA damaging agents rather than a particular kind of chemical. Examples include *PHR1* and *MAG1* as previously referred. Another well-studied DNA damage-inducible gene that responds to a variety of DNA-damaging agents is *RNR3*. *RNR3* encodes one of the large subunits of ribonucleotide reductase (RNR) that catalyzes the conversion of the nucleotides to deoxynucleotides, which is a rate-limiting step in de novo deoxyribonucleotides biosynthesis (An *et al*., 2006; Elledge and Davis, 1990; Yao *et al*., 2003). RNR plays an essential role in DNA replication and repair as it maintains the concentrations and balance among the individual dNTPs (dATP, dGTP, dCTP and dTTP) (Kumar *et al*., 2010). Functional RNR is a tetrameric protein complex containing two large subunit and two small subunits. The main isoform of large subunit is encoded by *RNR1*, the other isoform is encoded by *RNR3*, and two small subunits are encoded by *RNR2* and *RNR4* (Basso *et al*., 2008; Elledge and Davis, 1990). The homodimer of Rnr1 contains the

regulatory and catalytic sites, and heterodimer consisting of Rnr2 and Rnr4 houses the essential diferric-tyrosyl radical cofactor (Xu *et al*., 2006). The contribution of Rnr3 to ribonucleotide reduction is not clear. Though Rnr3 is not expressed in normal growth condition, it is strongly induced by DNA damage (Domkin *et al*., 2002). Using *lacZ* gene as a reporter gene, Endo-Ichikawa found that the *RNR3* promoter could be activated by a variety of DNA damaging agents including UV, alkylating agents MMS and MNNG, oxidation agent H_2O_2, hydroxyurea (HU), and γ-ray irradiation (Endo-Ichikawa *et al*., 1995). Due to this feature, the *RNR3* promoter-*lacZ* fusion system was developed as a tool to assess genotoxicity (Jia and Xiao, 2003; Jia *et al*., 2002).

Based on these observations, it seems likely that DNA damage induction is regulated by a global damage response pathway rather than a DNA lesion-specific pathway in budding yeast. The model of transduction pathway of DNA damage-inducible genes includes signals, sensors, transducers and effectors (Zhou and Elledge, 2000).

Regulation of Rnr3 expression is the best-known example of a eukaryotic transcriptional response to DNA damage. As previously mentioned, the transcriptional level of *RNR3* gene is very low under normal conditions. However, when treating yeast cells with DNA-damaging agents such as UV, MMS and 4-NQO, the transcriptional level of *RNR3* can be induced 100 to 500 fold (Elledge and Davis, 1990). In order to discover the mechanism of induction, Zhou studied a series of mutant yeast strains that have constitutive expression of *RNR3* (*crt* mutants) (Zhou and Elledge, 1992). A series of genes were identified as suppressors of Rnr3 expression. These negative regulators of Rnr3 expression can be divided into two groups, the first group includes indirect regulators that come from endogenous DNA damage or a state of metabolic stress such as

nucleotide depletion which induces the upregulation of Rnr3 expression. The other group includes direct regulators involved in the entire signal transduction pathway, transcription factors are also included. The *CRT1*, *TUP1* (*CRT4*) and *SSN6* (*CRT8*) genes encode direct negative regulators binding to the *RNR3* promoter (Figure 1–13). A second study was carried out by the same research group to identify mutations that disrupt DNA damage induction on *RNR3*. These genes were designated *DUN* for DNA-damage uninducible (Zhou and Elledge, 1993). Genetic analysis of *crt1Δ*, *tup1Δ* and *ssn6Δ* showed that these mutations were epistatic to the *dun1Δ* mutation, providing a strong genetic verification that *CRT1*, *TUP1* and *SSN6* function downstream of *DUN1* (Huang *et al*., 1998). The same group identified that Mec1 and Rad53 which are upstream kinases activating Dun1, are necessary for the DNA damage induction of *RNR3* (Huang *et al*., 1998) (Figure 1–13).

To sum up, in response to DNA damage or replication blocks, the Rad53 protein kinase is activated via a Mec1-dependent pathway, and activated Rad53 further phosphorylates the protein kinase Dun1. The Mec1-Rad53-Dun1 kinase cascade culminates in the phosphorylation of Crt1. Crt1 is a DNA-binding protein that binds to a 13-bp consensus sequence termed the X-box. Multiple X-box related sequences of different strength can be found in the promoter region of *RNR3*, *RNR2* and *RNR4*. Crt1 is able to recruit the general repressor Tup1-Ssn6 complex to suppress the transcription of *RNR* genes. The Crt1 is phosphorylated by Dun1 with the present of DNA damage or replication blocks, and loses the ability to bind to the X-box, leading to transcriptional induction of *RNR* genes (Figure 1–13.).

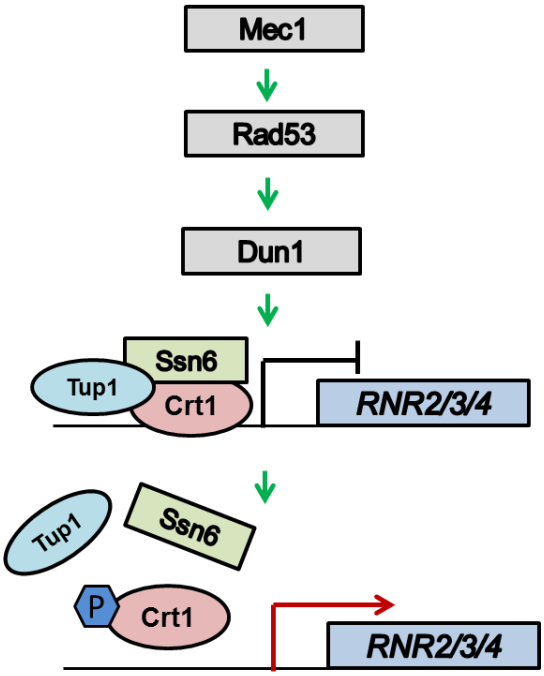
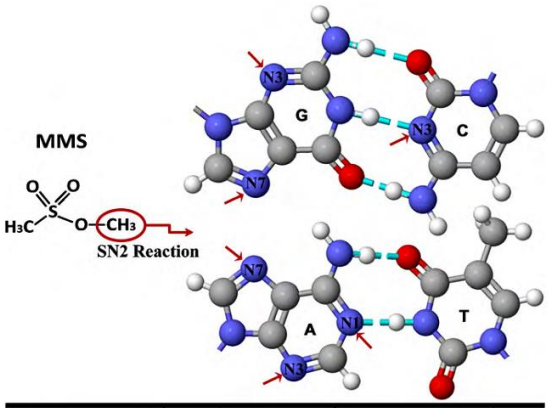


Figure 1–13. Mec1-Rad53-Dun1–dependent regulation pathway of transcription of *S. cerevisiae RNR* genes.

Dun1 kinase relieves the inhibition of transcription of RNR genes by targeting suppressor proteins binding upstream of RNR genes. Reproduced from (Sanvisens *et al*., 2013).

1.6.1. DNA Damage Caused by MMS Is Mainly Repaired by the Base Excision Repair (BER) Pathway in Yeast

MMS is a typical DNA alkylating agent. It mainly methylates guanine (G) and adenine (A), producing N7-methylguanine (7MeG) and N3-methyladenine (3MeA), as shown in Figure 1–14. While 7MeG does not block replication or cause increased mutations, 3MeA is a malignant lesion and affects base pairing during DNA replication. Therefore, N3-methyladenine needs to be removed otherwise it will cause genome damage and eventually cell death.



		C	G		A			Sum
		N^3	N^3	N^7	N^1	N^3	N^7	
dsDNA	*In vitro*	<1	0.6	85	3.8	10	1.8	101
	In cells	<1	1.3	83	>1	8.7	2.4	96
ssDNA	*In vitro*	10	1	68	18	1.4	3.8	102
	In cells	7		74	10	2.1		93

Figure 1–14. Frequency (%) of MMS-induced base lesions.

The reactive methyl group of MMS can attack cellular targets including both dsDNA or ssDNA by nucleophilic substitution through the SN_2 mechanism. The six nitrogens (pointed by arrows) with a double bond are the primary DNA targets of methylation by MMS and are labeled and highlighted with arrows. Atoms are colored by elements (nitrogen: blue; carbon: gray; oxygen: red; hydrogen: white) and are presented at their 25% van der Waals radii and bonds are shown in 0.1Å. Adopted from (Yang *et al.*, 2010).

DNA repair is the cellular response to DNA damage that alters nucleotide sequence and DNA structure to its native state (Friedberg, 2006). The major pathways include base damage reversal, base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), recombination repair and post-replication repair (PRR). These DNA repair pathways are highly conserved in all organisms from bacteria to human. Defects in DNA repair pathways will cause regional or global alteration of genetic information, leading to the accumulation of mutations and ultimately cell death. For mammals, DNA repair deficiency may lead to cancer and other diseases.

In general, BER repairs DNA base damage that causes relatively minor disturbances to the DNA helical structure. Such damage includes deamination, oxidation, and alkylation. Methylated bases resulting from MMS treatment are predominantly repaired by the base excision repair (BER) pathway. As shown in Figure 1–15, firstly, damaged bases are recognized by DNA glycosylases. Substrate-specific DNA glycosylases remove the damaged base from the sugar and phosphate backbone, resulting in the formations of apurinic/apyrimidinic (AP) sites. In *S. cerevisiae*, 3MeA sites are recognized and cleaved by Mag1 (3-methyl-adenine DNA glycosylase) (Chen *et al*., 1989; Xiao *et al*., 2001).. Other genes encoding glycosylases include *NTG1, NTG2, UNG1* and *OGG1* (Boiteux and Jinks-Robertson, 2013). Ntg1 acts as a DNA N-glycosylase that also exhibits an AP lyase activity. *NTG2* is a paralog of *NTG1*, arising from genome duplication. *UGN1* encodes a uracil-DNA glycosylase that recognizes the deamination of cytosine. *OGG1* encodes a glycosylase specifically excises 7, 8-dihydro-8-oxoguanine caused by oxidation. Starting from AP sites, DNA backbone can be cleaved by an AP endonuclease (Apn1 and Apn2 in yeast), resulting in the formation of a 3′-hydroxyl and a

5′-abasic sugar phosphate, deoxyribose phosphate (3′-dRP and 5′-dRP). This cleavage results in either a short gap (1 mer) or long single-strand gap (2-10 mer) at the DNA backbone, and thus named short-patch and long-patch BER, respectively. Alternatively, the AP sites can be cleaved by AP lyases associated with a subset of DNA glycosylases, as Ntg1 and Ntg2 in yeast. AP lyases catalyze the formation of a 5′-phosphate and 3′-fragmented deoxyribose. In yeast cells, the resulting gap can be filled in by DNA polymerase ε (Pol ε), and the remained DNA strand nick is sealed by DNA ligase (Cdc9).

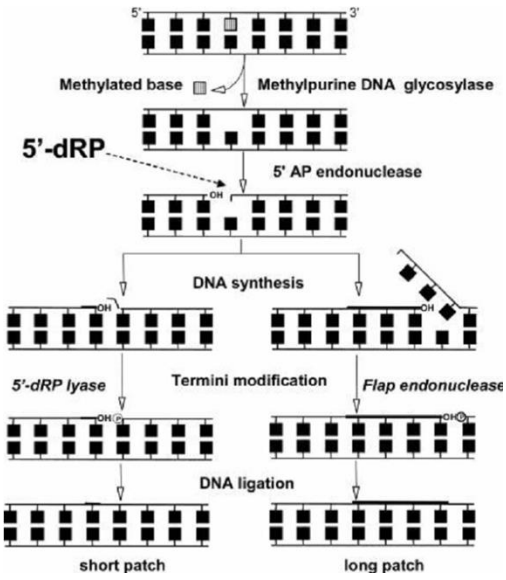


Figure 1–15. A general schematic of BER after methyl base damage.

BER repairs damaged DNA throughout the cell cycle. The repair process is initiated by DNA glycosylases which remove specific damaged bases. The resulting AP sites are then cleaved by an AP endonuclease, following by either short patch or long patch DNA synthesis. Adopted from (Wyatt and Pittman 2006).

It is important to notice that BER involves a series of enzymatic steps, and the intermediates themselves could be various forms of DNA damage. An AP site is a form of DNA damage and will give rise to form DNA strand breaks due to its inherent instability. As shown in Figure 1–16, the actions of DNA glycosylases and AP endonuclease creates breaks at DNA strands (single-strand breaks, SSBs) and two closely located SSBs on opposite strands are possible to form double strand breaks (DSBs). Alternatively, AP sites have the ability to block DNA replication and stalled replication forks can give rise to DSBs. Unrepaired strand breaks could cause loss of genetic information and/or cell death while mis-repaired strand breaks could lead to mutations.

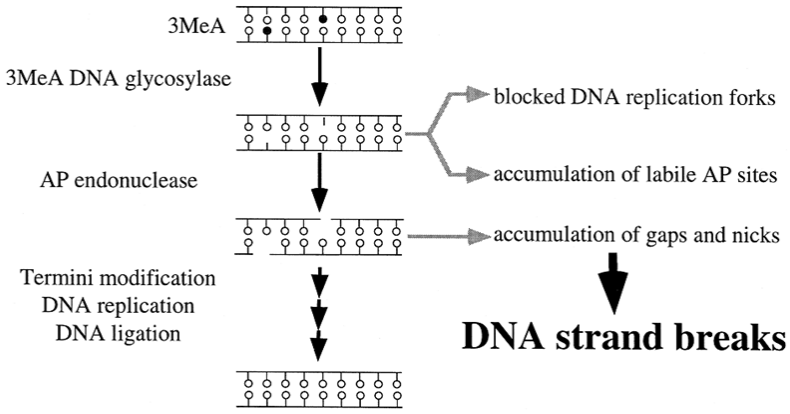


Figure 1–16. Illustration of BER intermediates giving rise to DNA strand breaks (DSBs).

Methylated DNA and AP sites accumulate lead to increased levels of DSBs. Adopted from (Memisoglu and Samson 2000).

Studies on model organisms reveal that strand breaks are mainly repaired by homologous recombination (HR) pathway and non-homologous ending joint (NHEJ) pathway (Figure 1–17).

HR is an error-free repair pathway to repair DSBs. When DNA replication is halted at a strand break, HR process is recruited to bypass the break by utilizing the complimentary chain as a template for DNA replication. The genes involved in this pathway were named *RAD52* group genes, due to their mutants were specifically sensitive to ionizing radiation. The genes involved in HR include *RAD50, RAD51*, *RAD52, RAD54, RAD55, RAD57, RAD59, MRE11*, and *XRS2*. Though *RAD51* initiates the HR process by binding to the breaking end of DNA to form filamentous tail searching for the homologous sequence, *RAD52* plays the central role in HR. Rad52 binds to single-strand DNA (ssDNA) as oligomers and accelerates the annealing of the complementary DNA strands. Other proteins involved in the annealing process include Rad55, Rad57, and Rad59. The strand with the break is introduced to anneal to the sister chromatin at the complementary region and continues replication. The formed four-strand DNA complex is named as a Holiday Junction (HJ). HJ cleavage could be performed at either direction and thus leading to an exchange of DNA, the transaction of DNA material could happen once or twice, based on the direction of HJ cleavage.

Non-homologous end joining (NHEJ) pathway is to rejoin DNA double-strand breaks directly; this process is considered as an error-prone repair because the direct rejoining would cause genetic information missing and chromosomal rearrangement. Studies have revealed that yeast processes a canonical NHEJ pathway dependent on Ku proteins and DNA ligase IV (Daley *et al*., 2005). Yeast Ku protein is a heterodimer

consisted of Ku70 and Ku80. Soon after DSB formation, Ku and a complex of Mre11, Rad50, and Xrs2 (MRX), then DNA ligase IV is recruited. After pairing, the two ends will be ligated.

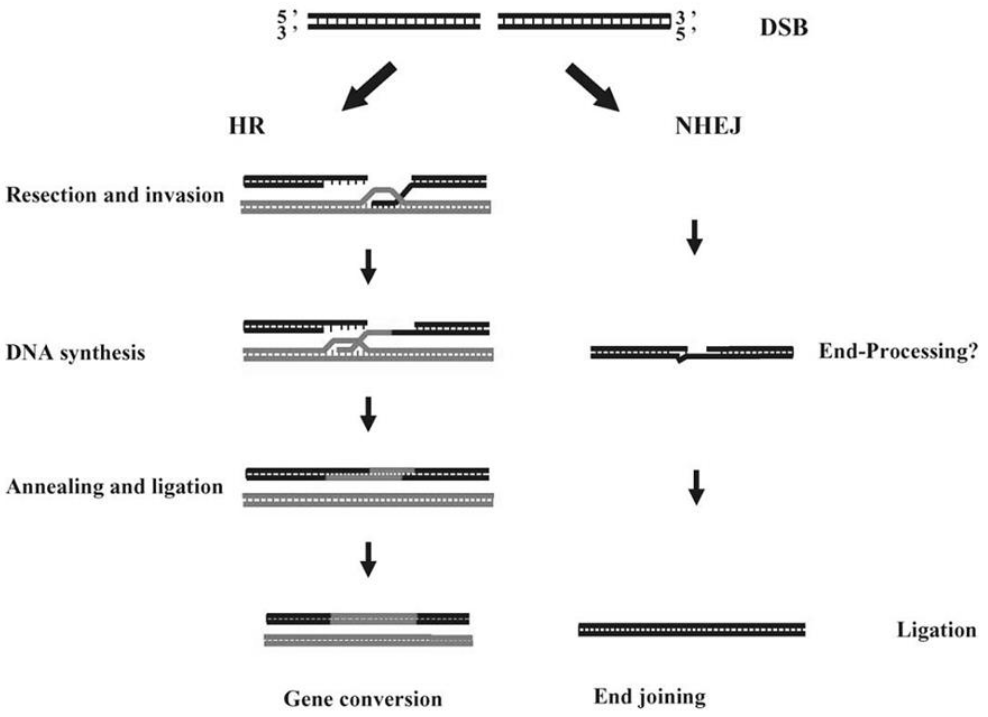


Figure 1–17. Scheme of homologous recombination (HR) and non-homologous ending jointing (NHJE) pathways repair double-strand breaks (DSBs) in yeast.

In cells that divide through mitosis, HR and NHJE repair DNA DSBs caused by ionizing radiation or DNA-damaging chemicals. Unrepaired DSBs can cause chromosomes rearrangement in somatic cells, which could lead to cancer. Unlike HR pathway, NHEJ process does not require sequence homology at DSBs region. Adopted from (Aylon and Kupiec 2004).

1.7. The Objective of This Study

The primary purpose of this study is to identify the function of Ddi2 and Ddi3 proteins, through genetic and biochemistry approaches.

The biochemical study focuses on identification of any function of the protein, such as enzymatic activity, any ability to bind to other proteins or DNAs. In this study, *DDI2/3* encoding regions would be subcloned and overexpressed in *E.coli*. Prokaryotically expressed Ddi2/3 protein would be collected and applied in an enzymatic assay, to determine if the protein has any enzymatic activity. To investigate the function of the endogenous Ddi2/3 protein, *ddi2/3Δ* and *ddi2Δ ddi3Δ* yeast strains would be created and studied. Compared to wildtype cells, any changes of phenotypes due to the lack of the genes would be observed and providing clues of the functions *in vivo*.

Due to its name, *DDI2/3* was proposed playing a role in DNA repair pathways to recover from damage caused by MMS. To investigate in which DNA repair pathway *DDI2/3* is participating, a series of triple mutants of yeast strain would be created, besides the disruptions of *DDI2* and *DDI3*, one of representative DNA repair genes is also disrupted in these strains. The triple mutants would be studied in sensitivity assays with exposure to MMS and other typical DNA damaging agents. Any enhanced sensitivities due to the three genes disruptions would help to determine which pathway requires *DDI2/3 in vivo*.

All these studies would be helpful to identify the physiological role of *DDI2/3* in yeast. A side-by-side study focusing on transcription regulation of *DDI2/3* is also launched. These studies may provide information that helps us to understand why this region is duplicated in the yeast genome.

CHAPTER 2 MATERIALS AND METHODS

2.1. Cloning Of *DDI2/3* Gene into Bacteria Overexpression System

To clone *DDI2/3* into a GST-tagged expression vector, the *DDI2/3* ORF was amplified from the yeast genome using primers YFL061w-2 (5′-GCC*GAATTC*ATGTCACAGTACGGATTT -3′) and YFL061w-3 (5′-GCC*GCGGCCGC*CCTCATTGAAACTTACCT -3′), and inserted into bacteria vector pGEX-6P-1 (GE Healthcare.) at the *Eco*RI and *Not*I sites (enzyme sites are shown in italic). This construct was made by Yu Fu (Fu, 2008).

To clone *DDI2/3* into a His_6-tagged expression vector, the *DDI2/3* ORF was amplified from the yeast genome using primers YFL061w-9 (5′-GGC*CCATGG*GCATGTCACAGTACGGA-3′) and YFL061w-10 (5′-GGC*GCGGCCGC*GTTATACCCAAATGTATT-3′). The PCR program includes pre-denaturation at 94°C for 3 minutes (min), and then followed by 30 cycles of 30 seconds (s) at 94°C, 30 s at 56°C and then 45 s at 72°C. At last, the reaction was incubated for 10 min at 72°C. PCR products were digested with *Nco*I and *Not*I and inserted into plasmid pET28a (EMD Millipore) at the corresponding multiple cloning sites. Restriction enzyme sites are shown as italic in primers.

Constructed pGEX-DDI2/3 and pET-DDI2/3 plasmids were first transformed into *Escherichia coli* strain DH10B for preservation. After confirming by DNA sequencing, correct plasmids were transformed into *E. coli* strain BL21(DE3) for overexpression.

2.2. Construction of Double Mutations on Ddi2/3: Ddi2/3-H88A; D89A and Ddi2/3-H137A; D139A by Using Mega-Primer Method

Site-specific mutations were created on Ddi2/3 in the above pGEX-DDI2/3 plasmid by a mega-primer approach (Ke and Madison, 1997) as previously described (Pastushok *et al*., 2005). The site-specific mutagenesis was introduced by doing PCR with two common primers pGEX-5′ and pGEX-3′, and two mutation-specific primers Ddi2/3-H88D89AA-P (5′-CTGTTGTTGCAAT**AG**CA**GC**AAGTAAGCAGGTG-3′) and Ddi2/3-H137D139AA-P (5′-GGGCTGCAGAATGAGCCCCAAGGTGGTAATGTAGCCAGTCCCAGTCAAA**G**CCTGG**GC**ACGAATGATGGC-3′). Mutated sites are in bold. The entire ORFs of the resulting plasmids were sequenced to confirm the mutant constructs. The mutated *DDI2/3* genes were expressed and proteins purified like the methods for wild type Ddi2/3.

Figure 2–1 shows the scheme of construction of double mutations by using megaprimer method. Briefly, to construct the site-specific mutation H88A; D89A, plasmid pGEX-DDI2/3 was used as the template for PCR (Figure 2–1A), and mega-primers were obtained from first round PCR using primers pGEX-5′ and a primer containing the H88D89AA mutation sites (Figure 2–1B). The PCR program includes pre-denaturation at 94°C for 3 min, followed by 30 cycles of 30 s at 94°C, 30 s at 58°C and 30 s at 72°C, and then ended with incubation for 10 min at 72°C. The amplified mega-primers were purified by doing agarose electrophoresis. In the second round of PCR, the template plasmid DNA was cut to linearize DNA by using restriction enzyme *Eco*RI (Figure 2–1C). Purified megaprimers and pGEX3′ were used to amplify the rest part of

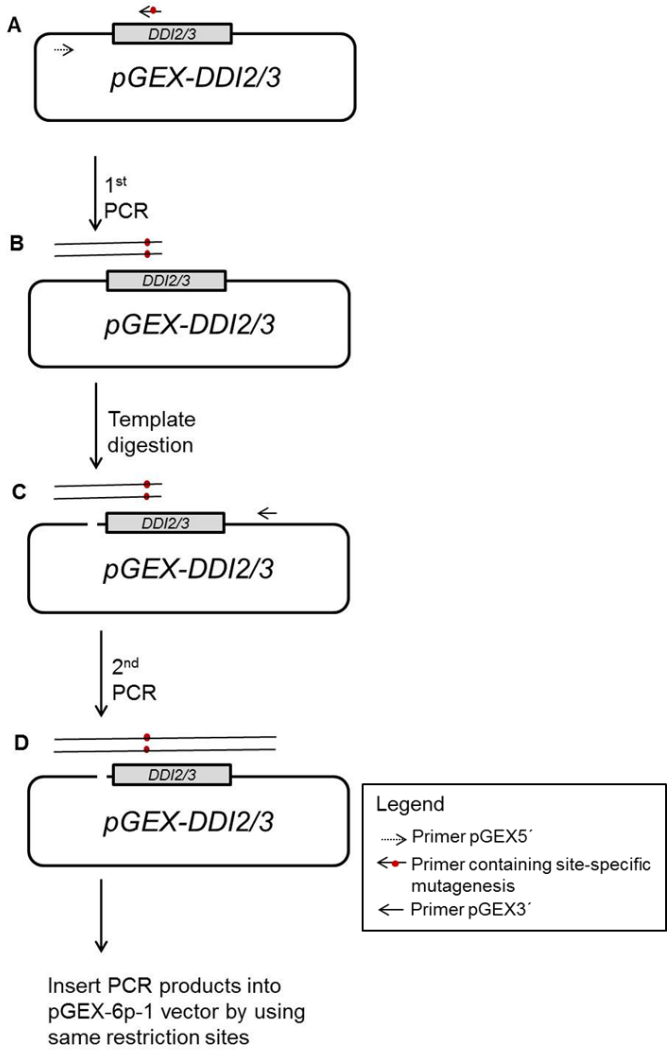


Figure 2–1. Scheme of Ddi2/3 double mutant construction by using mega-primers containing site-specific mutations.

pGEX-DDI2/3 was used as the template for PCR (panel ***A***), and mega-primers were obtained from first round PCR using primers pGEX-5´ and a primer containing the H88A; D89A mutation sites (panel ***B***). The amplified mega-primers were purified by doing agarose electrophoresis. In the second round of PCR, the template plasmid DNA was digested to linearize.(panel ***C***). Purified megaprimers and pGEX3´ were used to amplify the rest part of the *DDI2/3* coding region including restriction sites (penal ***D.***). Products obtained from the second round of PCR were inserted into the pGEX-6p-1 vector by using same restriction sites.

the *DDI2/3* coding region including restriction sites. The PCR program includes pre-denaturation at 94°C for 3 min, and then followed by 30 cycles of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C, and then ended with incubation for 10 min at 72°C. Products obtained from the second round of PCR were inserted into the pGEX-6P-1 vector by using EcoRI and NotI sites as well.

Similarly, to construct site-specific mutagenesis H137A; D139A, plasmid pGEX-DDI2/3 was used as template in PCR, and mega-primers were obtained from first round PCR using mega-primer containing H137A and D139A mutagenesis sites and pGEX3′. Then the template plasmid DNA was digested to linearize DNA by using restriction enzyme *Eco*RI and *Pst*I, while the latter site was mutated in mega-primers due to the site-specific mutagenesis. The second round of PCR was performed using pGEX5′ and mega-primers to amplify the rest part of the *DDI2/3* coding region including restriction sites. Products obtained from the second round of PCR were inserted into the pGEX-6P-1 vector by using *Eco*RI and *Not*I sites as well.

2.3. Purification of Prokaryotically Expressed Ddi2/3

2.3.1. Overexpression and Purification of Recombinant Ddi2/3 from pGEX Expression System

The recombinant plasmid pGEX-DDI2/3 was transformed into *E. coli* strain BL21(DE3) for recombinant protein overexpression. The transformed cells were cultured in LB containing 50 µg/ml ampicillin (LB+ Amp) at 37°C to A_{600nm} = 0.6 before starting induction with 0.1 mM IPTG and continued for 16 hours (h) at room temperature (23°C).

The cells were harvested by centrifugation at 10,000 × *g* for 10 min at 4°C. Five milliliter (ml) lysis buffer (50 mM Tris-HCl [pH 8.0], containing 150 mM NaCl and 1 mM EDTA) was used to suspend per gram pellet, then added lysozyme (1 mM), DTT (4 mM), DNaseI (100 μg/ml) and $ZnCl_2$ (6 μM) to the suspension. The cell lysate was collected after passing through a cell disrupter (TS Series Benchtop, Constant System Ltd) at 35k PSI, phenylmethanesulfonyl fluoride (PMSF) solution was added to cell lysate to 1 mM, and then the cell debris were removed by spinning at 15,000 × rpm (rotor JA-17 from Beckman Coulter) at 4°C for 60 min. GST-fused Ddi2/3 in the supernatant fraction was collected by using glutathione Sepharose™ 4B resin (GE Healthcare) and eluted by adding an excess amount of reduced glutathione (10 mM). After purification, the GST tag was cleaved by PreScission protease (GE Healthcare) in a cleavage buffer (50 mM Tris-HCl [pH 8.0], containing 150 mM NaCl, 1 mM EDTA and 4 mM DTT) at 4°C for 16 h. The GST tag and PreScission protease, which is also GST-tagged but cannot be cleaved, was removed by running the digested product over glutathione Sepharose™ 4B resin. Collected Ddi2/3 was dialyzed in 50 mM HEPES-Na [pH 7.5], containing 100 mM NaCl. During the purification process, protein samples were taken after each step and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Western blot analysis was also performed to identify the GST fusion, in which a goat anti-GST (GE Healthcare) primary antibody (1:20,000 diluted) and bovine anti-goat IgG-HRP (Santa Cruz Biotechnology) secondary antibody (1:10,000 diluted) were used.

2.3.2. Overexpression and Purification of Recombinant Ddi2/3 from pET Expression System

The plasmid pET-DDI2/3 was transformed into *E. coli* strain BL21(DE3) for recombinant protein overexpression. The transformed cells were cultured in an LB medium containing 50 μg/ml kanamycin (LB+ Kan) at 37°C to A_{600nm} = 0.6 before starting induction by 0.5 mM IPTG at 16°C for over 16 h. The cells were harvested by centrifugation at 10,000 × *g* at 4°C for 10 min. Five ml of lysis buffer (50 mM Tris-HCl [pH8.0], containing 150 mM NaCl and 1 mM EDTA) was used to suspend each gram of pellet; lysozyme was added to 1 mM for cell lysis, followed by additions of 4 mM DTT, 100 μg/ml DNaseI and 6 μM $ZnCl_2$. The cell lysate was collected after passing through a cell disrupter (TS Series Benchtop from Constant System Ltd) at 35k PSI, phenylmethanesulfonyl fluoride (PMSF) solution was added to cell lysate to 1 mM, and then the cell debris was removed by spinning at 15,000 × rpm (rotor JA-17 from Beckman Coulter) at 4ºC for 60 min. Cell lysate through cell disrupter was added to 5 volume of His_6-tag binding buffer (50 mM Tris-HCl [pH 8.0], containing 500 mM NaCl and 20 mM imidazole) to dilute EDTA. Then His_6- tagged Ddi2/3 in the cell lysate was collected by using Ni Sepherose™ 6 Fast Flow resin (GE healthcare) and eluted by adding an excess amount of imidazole (500 mM). Imidazole was removed by dialyzing protein in storage buffer (50 mM Tris-HCl [pH 8.0], containing 150 mM NaCl). Protein purity was assessed by SDS-PAGE.

2.4. Enzymatic Assay of Cyanamide Hydratase

2.4.1. Urease Based Enzymatic Assay of Cyanamide Hydratase

Purified prokaryotically expressed Ddi2/3 was incubated with various concentrations of cyanamide at room temperature (23ºC), and the urea formation was monitored by adding excessive urease to hydrolyze urea to ammonia. The produced ammonia was then quantified using an ammonia assay kit purchased from Sigma (Cat. AA0100). The kit provides α-ketoglutarat acid (KGA), reduced nicotinamide adenine dinucleotide phosphate (NADPH) and L-glutamate dehydrogenase (LDH). Incorporation of ammonia is coupled with NADPH consumption in the formation of glutamate, and its molarity is equal to the consumption of NADPH, which has an absorbance peak at 340 nm. The reaction mixture for measuring kinetic parameters contains 0.62 μM urease (Sigma), 0.45 μM L-GDH, 3.4 mM KGA, and 0.23 mM NADPH. For each assay, 100 μl of Ddi2/3 solution, containing 0.027 μM (0.7 μg) protein was incubated with the above reaction mixture and cyanamide solution was added last to start the reaction. Each reaction was monitored at different time intervals for at least 20 min to calculate the initial velocity. When cyanamide concentrations were below 5 mM, the cyanamide hydration reactions were continuously monitored for 15 min. When cyanamide concentrations were above 5 mM, real-time monitoring of reactions for 15 min would be impossible since cyanamide increases the background in the ammonia assay and the A_{340nm} soon drops to undetectable levels. In these situations, cyanamide hydration reactions and ammonia assays were separated. Briefly, the cyanamide hydration by Ddi2/3 was started in a 1.5 ml Eppendorf tube containing 0.027 μM of Ddi2/3 and 0.62 μM of urease, with cyanamide added last to initiate the reaction, making the total volume

to 1 ml. 50-200 µl of the reaction mixture was withdrawn every 7-10 min for the ammonia assay and the total amount of ammonia produced in the original reaction was determined using the ammonia assay kit. To account for possible background ammonia production, a negative control reaction lacking recombinant Ddi2/3 was performed. All assays were repeated at least three times to calculate the mean and the standard deviations. The purified GST protein was used as a negative control in the enzymatic assay because it is readily available in the lab and its molecular weight is comparable to that of Ddi2/3.

2.4.2. Colorimetric Assay-Based Enzymatic Assay of Cyanamide Hydratase from Yeast Whole Cell Extracts (WCEs)

Cyanamide was added to 0.5 ml yeast WCEs to a final concentration of 2 mM and the solution was incubated at 30ºC. The cyanamide concentration was monitored by using a colorimetric assay as described by Weeks (Weeks *et al*., 2000). Briefly, 100 µl reaction mixture was added to 500 µl PBS, followed by adding 400 µl of 0.1 M sodium carbonate-bicarbonate buffer (SCB) [pH 10.4], and 200 µl of 4% sodium pentacyanoammine-ferroate (II) (TCI chemicals) solution as the color reagent. After reaction in the dark for 10 min, A_{530nm} was measured to determine the remaining cyanamide concentration. The decrease of A_{530nm} after incubation indicates the consumption of cyanamide in solution, reflecting the activity of cyanamide hydratase in WCEs.

2.4.3. Colorimetric Cyanamide Hydratase Assay with Putative Inhibitors

Purified His-tagged Ddi2/3 was adjusted to 0.5 mg/ml (0.019 mM) and incubated with 10 mM of a variety of chemicals at room temperature for 30 min and then was

incubated with 5 mM of cyanamide. The reaction was continued at room temperature for 2 h and the remaining cyanamide was measured by using the colorimetric assay described above. The decrease of A_{530nm} after incubation indicates the cyanamide hydratase activity of Ddi2/3 in the presence of the chemicals.

2.5. DMSO Enhanced Yeast Transformation

Wild type yeast cells were cultured overnight in 2 ml YPD broth at 30°C with shaking. On the second day, 10 ml fresh YPD broth was added to the overnight culture and cells were allowed to grow for another 5 h with shaking. Then 1.5 ml yeast cells were transferred to an Eppendorf tube and centrifuged at 6,000 × *g* for 2 min at room temperature. The pellet was washed with 0.8 ml total of LiOAc solution (10 mM Tris-HCl [pH 8.0], containing 1 mM EDTA and 0.1 M LiOAc,) once and resuspended in 0.1 ml of LiOAc solution; competent yeast cells are ready to use. Salmon sperm DNA (ssDNA) solution (10 mg/ml) was boiled for 5 min and immediately put on the ice. 10 µl ssDNA and 5 µl plasmid DNA (50 ng/µl) were added to competent yeast cells. The contents were mixed gently and then left on the bench for 5 min. Then 280 µl of polyethylene glycol (PEG) 4,000 solution (50% [w/v] PEG 4,000 dissolved in the LiOAc solution) was added, the contents were mixed by inverting 4-6 times, and then the tube was placed at 30°C for 45 min without shaking. After incubation, 39 µl dimethyl sulfoxide (DMSO) was added to give an approximately 10% (vol/vol) DMSO solution in the tube. The contents were mixed by inversion then heat shocked at 42°C for 5 min. Cells were spun at 6,000 × *g* for 2 min, washed in 1 ml ddH_2O once, and then suspended in 0.2 ml ddH_2O. All transformed cells were plated on selective SD medium

supplemented with the necessary bases and amino acids, and incubated for 3 days at 30°C. Individual colonies were streaked on a fresh prepared selective plate for future use.

2.6. β-Galactosidase (β-Gal) Assay

The promoter region of *DDI2/3* was amplified by PCR from genomic DNA by primers YFL061w-1 (5′-GGAAAATCCAAGCTTTCAAG-3′) and YFL061w-3 (5′-GCC*GCGGCCGC*CCTCATTGAAACTTACCT-3′); the restriction enzyme sequence used for cloning is italicized. The PCR product (containing -718 to +678 of *DDI2/3*) was then cloned as a *Hin*dIII-*Pst*I (-711 to +457) fragment into YEp365R (Myers *et al*., 1986) to form YEpDDI2-lacZ. The resulting plasmid YEpDDI2-lacZ was transformed into yeast cells though DMSO enhanced transformation method as described above and transformants were maintained in an SD-Ura medium. Yeast cells containing the plasmid were cultured in SD-Ura medium making sure the cells maintained the plasmids. The β-gal assay was performed as described previously (Jia *et al*., 2002). Firstly, yeast cells containing YEpDDI2-lacZ plasmid was inoculated in 2 ml SD-Ura medium and growing at 30ºC for overnight. On the next day, 0.5 ml overnight culture was subcultured in 2.5 ml SD-Ura medium and growing at 30ºC for another 2 h. Two sets of cells were set up in the subculture for the following treatment. The test chemical was added to one set of cells at the desired concentration, and all cells were incubated at 30°C for another 4 h. After incubation, cell density was determined by taking 1 ml of cell cultures to measure A_{600nm}. The remaining 2 ml of cells were collected and suspended in 1 ml of Buffer Z (60 mM Na_2HPO_4, 40 mM NaH_2PO_4 [pH 7.0], 10 mM KCl, 1 mM $MgSO_4$, 40 mM β-mercaptoethanol). To break cells, 50 µl 0.1% sodium dodecyl sulfate (SDS) and

chloroform were added to the cell suspension, followed by vortexing at top speed for 10 s. The β-gal enzyme reaction was started by adding 200 µl 4 mg/ml orthonitrophenyl-β-D-galactosidase (ONPG) solution. The reaction continued 20 min at 30ºC, in a shaker. Then 500 µl 1 M Na_2CO_3 solution was added to stop the reaction. The test tube was spinning at 3,500 × *g* at 4ºC for 10 min to remove cell debris. After the spin, 1 ml of supernatant was transferred into a cuvette to determine the absorbance at 420 nm. A_{420nm} The β-gal activity was determined using the following equation:

$$1 \text{ Miller Unit} = \frac{1000 \times A_{420nm}}{\text{reaction time (min)} \times \text{volume of reaction (ml)} \times A_{600nm}}$$

The β-gal assay was performed with several independent yeast transformants from the same transformation to avoid internal inconsistency. Results from various transformants/treatments presented for comparison (e.g., treated *vs.* untreated with chemicals) were always from the same experiment to minimize inter-experimental variation.

2.7. Construction of *ddi2/3Δ* and *ddi2Δ ddi3Δ* Yeast Strains

To construct *ddi2/3Δ* disruption cassettes, the YFL061w-1/YFL061w-3 PCR product was cleaved by *Pvu*II-*Not*I and the 1.17-kb fragment containing a 0.5-kb promoter and the entire *DDI2/3* ORF were cloned into *Sma*I-*Not*I of the pBlueScript vector. A 0.54-kb *Hpa*I-*Bam*HI fragment (-55 to +487) was then deleted from the resulting pBS-DDI2 plasmid and replaced by a *Bam*HI linker, which was used to clone either 1.16-kb *HIS3* from plasmid YDp-H or 1.6-kb *LEU2* from plasmid YDp-L (Berben

et al., 1991). The resulting *ddi2/3*Δ*::HIS3* cassette was released by *Xba*I-*Xho*I digestion and the *ddi2/3*Δ*::LEU2* disruption cassette by *Bgl*II-*Xho*I digestion before yeast transformation. To achieve high-efficiency and specificity of target gene disruption, the disruption cassettes were purified from 1% agarose gel after electrophoresis. A single copy of *DDI2/3* gene was disrupted by either the *ddi2*Δ*::HIS3* or *ddi2*Δ*::LEU2* cassette by a one-step gene disruption method (Rothstein, 1983). Briefly, either of two disruption cassette was transformed into yeast cells and during homologous recombination, one copy of the *DDI2/3* genes was replaced by the selective marker gene and, therefore, could survive on selective medium, which lacks either histidine or leucine. Then the transformants were used for the second round transformation in which the other selective marker was used to disrupt the second copy of *DDI2/3* genes. The double disruption lines were screened for their ability to grow on the SD agar plates lacking both histidine and leucine, and then further confirmed by genomic PCR with primers YFL061w-1 and YFL061w-3 flanking the deleted region.

2.8. Sensitivity Assay with *ddi2/3Δ* and *ddi2Δ ddi3Δ* Yeast Strains

Wild type yeast strain and confirmed *ddi2/3Δ* mutant strains were cultured overnight in 2 ml YPD broth at 30ºC with shaking. Sterilized ddH_2O was used in adjusting cell density and making a series of tenfold dilution of yeast cultures. All these dilutions were equally spotted on YPD agar plate and plates containing testing chemicals. After the liquid had been absorbed, the plates were incubated at 30ºC for three days before taking photographs. The relative growth on control and testing plates represents the level of cellular resistance to the testing chemical.

A gradient plate assay is also used to measure sensitivities of yeast mutants to DNA damaging agents. Thirty ml of molten YPD agar were mixed with an appropriate concentration of a specific DNA damaging agent to form the bottom layer. The gradient was created by pouring the medium into a tiled square petri dish. After brief solidification for one hour, the petri dish was returned to a flat position and 30 ml of the molten YPD agar was poured to form the top layer. A 100 µl of yeast culture was taken from an overnight culture, mixed with 0.4 ml sterile water and 0.5 ml of molten YPD agar, and then immediately imprinted onto freshly made gradient plates using a microscope slide. Gradient plates were incubated at 30℃ for 2 days before taking. The growth reflected the resistance of yeast cells to the DNA-damaging agent.

2.9. Preparation of Yeast Whole Cell Extracts (WCEs)

Wild type yeast strain BY4741, *ddi2/3Δ* and *ddi2Δ ddi3Δ* mutant strains were inoculated in 5 ml YPD broth and incubated overnight at 30ºC with shaking. On the next day, the overnight culture was subcultured into 50 ml fresh YPD medium and cells were

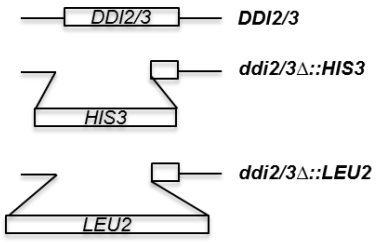


Figure 2–2. Scheme of *DDI2* disruptions through homologous recombination.

The two terminal sequences at two termini of the disruption cassette are identical to the target gene. Through homologous recombination, the target gene will be replaced by the selective marker gene, and thus, mutants can survive on selective medium.

continuously cultured until A_{600nm} reached 0.4. The yeast cultures were treated with 25 mM cyanamide for 2 h at 30ºC and cells were collected by centrifugation and then washed with 5 ml yeast lysis buffer (20 mM Tris-HCl [pH 8.0], containing 150 mM NaCl and 0.5 mM EDTA). The pellet was weighed, and each gram pellet was suspended in 1 ml lysis buffer. PMSF solution was added to 1 mM to the cell suspension. After adding 0.3 g acid-treated glass beads, yeast cells were lysed by vortexing 30 s followed by leaving on ice 30 s; after 8 rounds of the vortex, glass beads and cell debris were removed by centrifugation and the supernatant was centrifuged at 15,000 × *g* at 4ºC for 10 min. The clear supernatant was collected as WCEs, and the protein concentration was measured by the Bradford assay using the Bio-Rad protein assay reagent and then adjusted to 1 mg/ml in PBS.

2.10. Purification and Concentration of Recombinant Protein for Crystallization

Two liters of *E.coli* BL21(DE3)-pGEX-DDI2/3 culture was used in Ddi2/3 purification as previously described. Purified Ddi2/3 was further purified by passing through prepacked RESOURCE Q (GE Healthcare) ion exchange column to separate Ddi2/3 and residual GST tag. Briefly, purified Ddi2/3 collected from second GST affinity purification was dialyzed in 200 volume of buffer A/low-salt buffer (20 mM Tris-HCl [pH 8.0], containing 20 mM NaCl) at 4ºC for overnight. Then the protein was loaded to SourceQ ion-exchange column (GE Healthcare) by using the Duo Flow chromatography system (Bio-Rad) at a speed of 0.2 ml/min. Ddi2/3 was eluted by gradient salt elution containing 0% to 25% buffer B/high-salt buffer (20 mM Tris-HCl [pH 8.0], containing

2.0 M NaCl). Elution was collected 1 ml per tube, and elution fractions were identified on SDS-PAGE to determine the purity of Ddi2/3. Fractions containing pure Ddi2/3 were collected and concentrated to 10 mg/ml by using UltraSpin (EMD Millipore) concentration tubes with molecular weight cut-off (MWCO) of 10 kDa. Concentrated Ddi2/3 was dialyzed in storage buffer (20 mM bis-Tris propane [pH 7.0], containing 150 mM NaCl and 0.01 mM $ZnCl_2$) and identified in SDS-PAGE to check the purity; purified Ddi2/3 with over 90% purity was used in crystallization grid screens.

2.11. Protein Crystallization and Cryoprotection

Crystals were grown by the hanging-drop method at room temperature (23℃) using 0.8 µl to 2.0 µl of protein solution (8-11 mg/ml) mixed with equal volume of reservoir solution.

Grid screening of protein crystallization was performed on 24-well crystallization plates (VDX plate with sealant from HAMPTON research) with crystallization buffer kit, including Wizard I and Wizard II crystallization screen series (Rigaku), and $AmSO_4$ suite (QIAGEN). Crystallization trials were set using hanging drops method. 0.8 µl to 2 µl of protein solution (8-11 mg/ml) was mixed with equal volume of reservoir buffer on a clean plastic cover slide (HAMPTON research). Then the cover slide was quickly flipped and sealed on the top of the well. The plates were sitting at room temperature for three days to a week before reviewing under a microscope. Cryoprotection was attained by the sequential addition of increments of mother solution supplemented with 22%-26% (vol/vol) glycerol, followed by the subsequent flash frozen in liquid nitrogen. Substrate/inhibitor soaking was performed as equilibrium crystals in cryo-buffers

containing 0.2 M-0.35 M corresponding chemicals for few seconds, followed by the subsequent flash cooling in liquid nitrogen.

2.12. Data Collection and Structure model building

X-ray diffraction data on Ddi2/3 protein crystals were collected using beamlines 08B1-1 and 08ID-1 of the Canadian Light Source (CLS). The distance from crystal to the detector was adjusted to 320 mm, exposure time was set from 1 s to 4 s to avoid radiation damage to protein crystals; one frame was taken for every 0.2°. A typical dataset included 100-120 degrees of data.

Diffraction datasets were processed using HKL2000 (Otwinowski and Minor, 1997). Manipulations with the resulting diffraction data were performed using the CCP4 suite of programs (Winn *et al*., 2011) .and PHENIX to yield an interpretable map. Structure building and refinement were performed using PHENIX (Adams *et al*., 2010) and COOT (Emsley and Cowtan, 2004). Structural figures were prepared using PyMol (Schrodinger, 2015)

2.13. Site-Specific Mutagenesis

In order to study the involvements of amino acids surrounding the zinc binding site of substrate binding and likely participating the reaction, five site-specific mutations were applied to surrounding residues. Total five mutants, H137N, Q138E, Q138A, T157V, and N161A, were made by using QuickChange site-specific mutagenesis kit (Stratagene). Site-specific mutations were introduced by doing PCR using primers that

contain mutated sites. Five pairs of primers corresponding to the site-specific mutations were designed and synthesized (Table 2-1).

As shown in Figure 2–3, firstly, plasmid pET-DDI2/3 or pGEX-DDI2/3 was isolated and used as template in PCR. Primers containing mutated sequences introduced site-specific mutations to PCR products. After PCR, the template DNA was completely digested by *Dpn*I. The PCR product was directly transformed into *E. coli* cells, where it formed plasmid. Resulting plasmids containing site-specific mutations were verified by DNA sequencing. Recombinant Ddi2/3 proteins containing site-specific amino acid alterations were produced and purified as previously described. The cyanamide hydratase activities of Ddi2/3 mutants were tested in enzymatic assays.

Table 2-1. Primers used for site-specific mutagenesis.

Primers	Sequence (5′ – 3′)
Ddi2-H137N-F	CTGAGGCCATCATTCGT**A**ACCAGGATTTGACTGGG
Ddi2-H137N-R	CCCAGTCAAATCCTGGT**T**ACGAATGATGGCCTCAG
Ddi2-Q138A-F	GAGGCCATCATT CGTCAC**GC**GGATTTGACTGGG
Ddi2-Q138A-R	CCCAGTCAAATCC**GC**GTGACGAATGATGGCCTC
Ddi2-Q138E-F	GGCCATCATTCGTCAC**G**AGGATTTGACTGGG
Ddi2-Q138E-R	CCCAGTCAAATCCTCGTGACGAATGATGGCC
Ddi2-T157V-F	CATTCTGCAGATTGCT**GT**TACGCTTGACAATGTCGGATCC
Ddi2-T157V-R	GGATCCGACATTGTCAAGCGTA**AC**AGCAATCTGCAGAATG
Ddi2-N161A-F	GCTACTACGCTTGAC**GC**TGTCGGATCCAATACCGATC
Ddi2-N161A-R	GATCGGTATTGGATCCGACA**GC**GTCAAGCGTAGTAGC

Mutagenic sites were shown as bold.

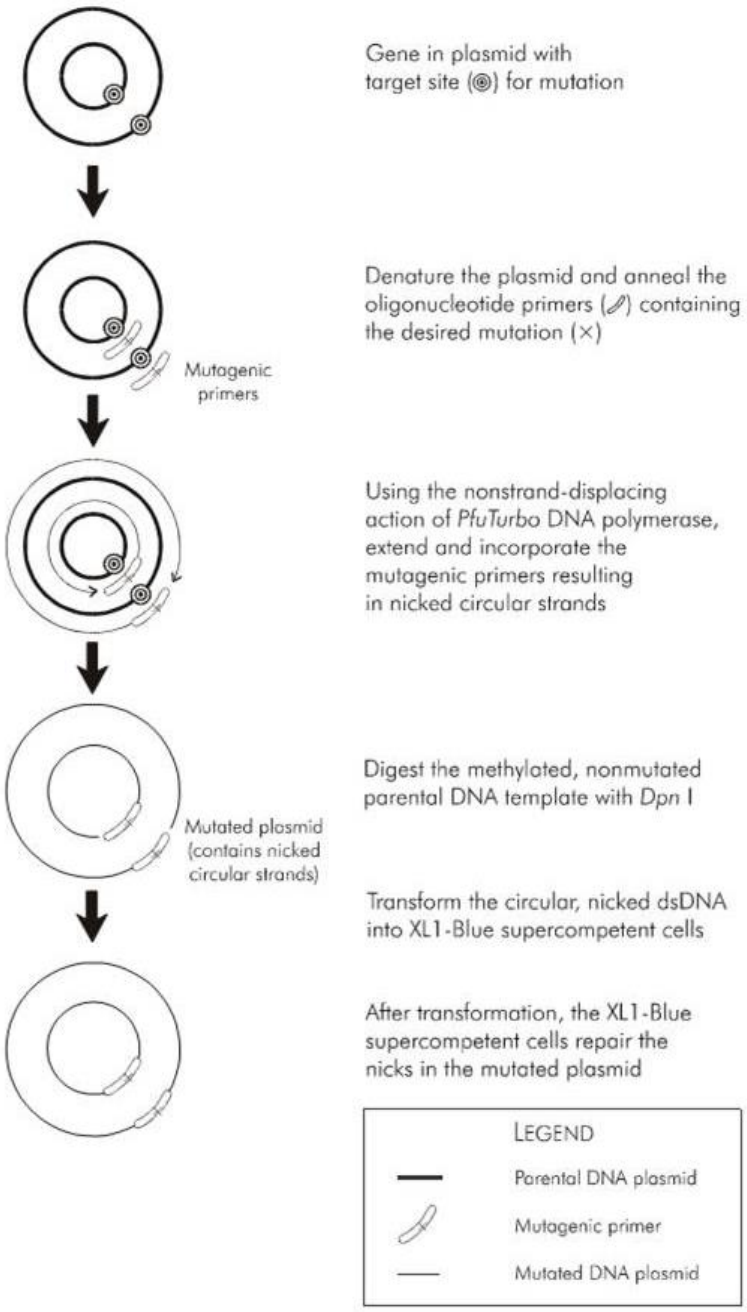


Figure 2–3. Overview of constructions of site-specific mutations by using QuickChange™ Site-Directed Mutagenesis kit.

pET-DDI2/3 or pGEX-DDI2/3 was used as template in PCR. Primers containing mutated sequences introduced site-specific mutations to PCR products. After PCR, the template

DNA was completely digested by *Dpn*I and thus removed. The PCR product was directly transformed into *E. coli* cells, where it formed plasmid. Adopted from the user manual.

CHAPTER 3 *DDI2* AND *DDI3* ENCODE A CYANAMIDE HYDRATASE

Most of the results in this chapter have been published in the article below:

Li, J., Biss, M., Fu, Y., Xu, X., Moore, S.A., and Xiao, W. (2015). Two duplicated genes *DDI2* and *DDI3* in budding yeast encode a cyanamide hydratase and are induced by cyanamide. *J Biol Chem* **290**, 12664-12675.

JL, MSA and WX designed the experiment; JL performed most experiments, conducted data collection and interpretation, and drafted the article; BM performed β-gal assays with pYEDDI2/3-lacZ transformants in response to MMS and cyanamide; FY performed Southern blot and β-gal assays with pYEDDI2/3-lacZ transformants to a variety of DNA-damaging agents; XX constructed *Chr6Δ*, *Chr14Δ*, *Chr6Δ ddi2Δ* and *Chr14Δ ddi3Δ* mutant strains used in this study. XW is the corresponding author. Results obtained by other authors included in this chapter are indicated in the figure legends.

3.1. Abstract

To study the functions of two identical DNA damage-inducible genes in *Saccharomyces cerevisiae*, *DDI2* and *DDI3*, the gene was cloned into bacterial expression vectors, and the recombinant proteins were purified and studied. Since Ddi2/3 is characterized as an HD domain protein and it shows limited homology to a fungal cyanamide hydratase that converts cyanamide to urea. Prokaryotically expressed Ddi2/3 protein was tested in a cyanamide hydratase assay. From the assay, a K_M of the recombinant Ddi2/3 to cyanamide was determined as 17.3 ± 0.049 mM. Studies with site-specific mutations at the HD domain indicated that the cyanamide hydratase activity of

Ddi2/3 requires the HD domain. Unlike most other DNA damage-inducible genes, *DDI2/3* is the only one induced by a particular set of alkylating agents and surprisingly is massively induced by cyanamide. To characterize the biological function of *DDI2/3*, both genes were deleted, and the resulting double mutant showed enhanced sensitivity to growth inhibition by cyanamide, suggesting that the *DDI2/3* genes protect host cells from cyanamide toxicity. Despite the physiological relevance of the cyanamide induction, *DDI2/3* is not involved in its transcriptional regulation, since the cyanamide induction of a reporter gene driven by the *DDI2/3* promoter is not compromised in the *ddi2Δ ddi3Δ* double mutant. The significance of cyanamide hydratase and its induced expression is discussed.

3.2. Introduction

3.2.1. MMS-Inducible Genes *DDI2* and *DDI3* Are Duplicated Genes

DDI2 and *DDI3* are two genes identified in *S. cerevisiae* through a genome-wide microarray analysis of transcription levels in response to methyl methanesulfonate (MMS), which is a typical DNA-methylating agent. In this microarray analysis, two open reading frames (ORFs) *YNL335W* and *YFL061W* displayed the highest induction (> 100-fold) among the whole genome after 0.1% MMS treatment (Figure 3–1) (Fu, 2008). Therefore, they were named DNA-damage inducible genes 2 and 3 (*DDI2* and *DDI3*), respectively, as *DDI1* was reported as co-regulated with *MAG1* (Liu *et al*., 1997; Liu and Xiao, 1997), involved in DNA-damage checkpoint (Clarke *et al*., 2001) and required for repression of protein secretion (White *et al*., 2011).

DNA sequencing reveals that *DDI2* and *DDI3* are duplicated genes located on different chromosomes (Figure 3–1B), with identical ORF sequences and only one nucleotide difference in their promoter (up to 1 kb) regions. Therefore, the genes are referred as *DDI2/3* in the following studies. Since all of the well-documented DNA damage-inducible genes in budding yeast seem to respond to a broad spectrum of DNA-damaging agents (Fu *et al*., 2008), for example, *RNR3* encodes the large subunit of ribonucleotide reductase (Rnr3) in yeast; its induction responds to all typical DNA

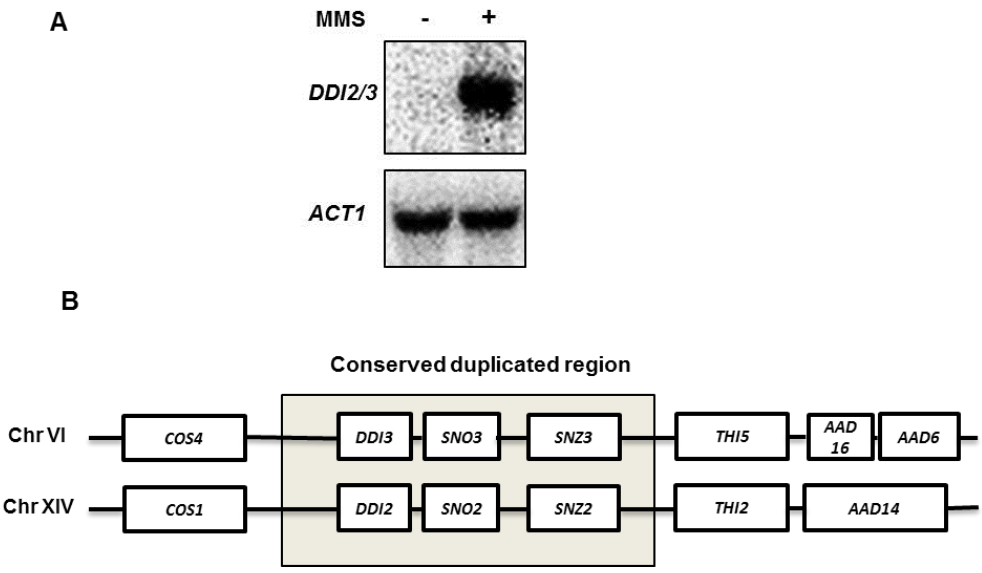


Figure 3–1. Results of the Northern blot of *DDI2/3* after MMS treatment and the diagram showing the duplication region containing *DDI2/3* genes.

A.. Northern blot of *DDI2/3* after MMS treatment, *ACT1* was used as a reference. The assay was performed by Fu, Y as described in (Fu, 2008; Zhu and Xiao, 1998). ***B.*** *DDI2* and *DDI3* genes are located within a highly conserved duplicated region. *DDI2* and *DDI3* and their flanking 20-kb regions are thought to derive from gene duplication, in which the boxed regions are highly conserved in DNA sequence.

damaging agents. Therefore, *RNR3* is used as a reporter to monitor genotoxicity (Jia *et al*., 2002). To examine the induction of *DDI2/3* to representative DNA-damaging agents, the *lacZ* gene was cloned behind the *DDI2/3* promoter and used as a reporter gene. The expression level of LacZ (β-galactosidase) after treatment is monitored by doing the β-gal assay. The relative fold induction after treatment represents the activation of the *DDI2/3* promotor.

Yu Fu (Fu, 2008) tested representative DNA damaging agents in the β-gal assay by using plasmid pYEDDI2/3-lacZ (Figure 3–2). Results of β-gal assay show that, unlike other characterized budding yeast DNA damage-inducible genes that are often induced by a variety of DNA-damaging agents regardless of whether they are involved in the repair of that type of DNA damage, the transcriptional level of *DDI2/3* was only highly induced by SN_2-type alkylating agents including MMS and dimethyl sulfate (DMS), up to 350-fold and 150-fold, respectively (Figure 3–2A-B). These compounds alkylate predominantly at nitrogens rather than oxygens in DNA bases. The SN_1-type alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), which efficiently alkylates both nitrogens and oxygens, weakly induce the expression of *DDI2/3-lacZ* (Figure 3–2C), probably due to its overlapping activities with SN_2-type alkylating agents. Other DNA-damaging agents, such as ethyl methanesulfonate (EMS), γ-ray and hydroxyurea (HU), only mildly induced the expression of *DDI2/3-lacZ* by no more than eightfold (Figure 3–2D-F). Furthermore, UV irradiation, a well-known DNA-damaging agent, did not induce *DDI2/3-lacZ* expression at all (Figure 3–2G). The above observations indicate that

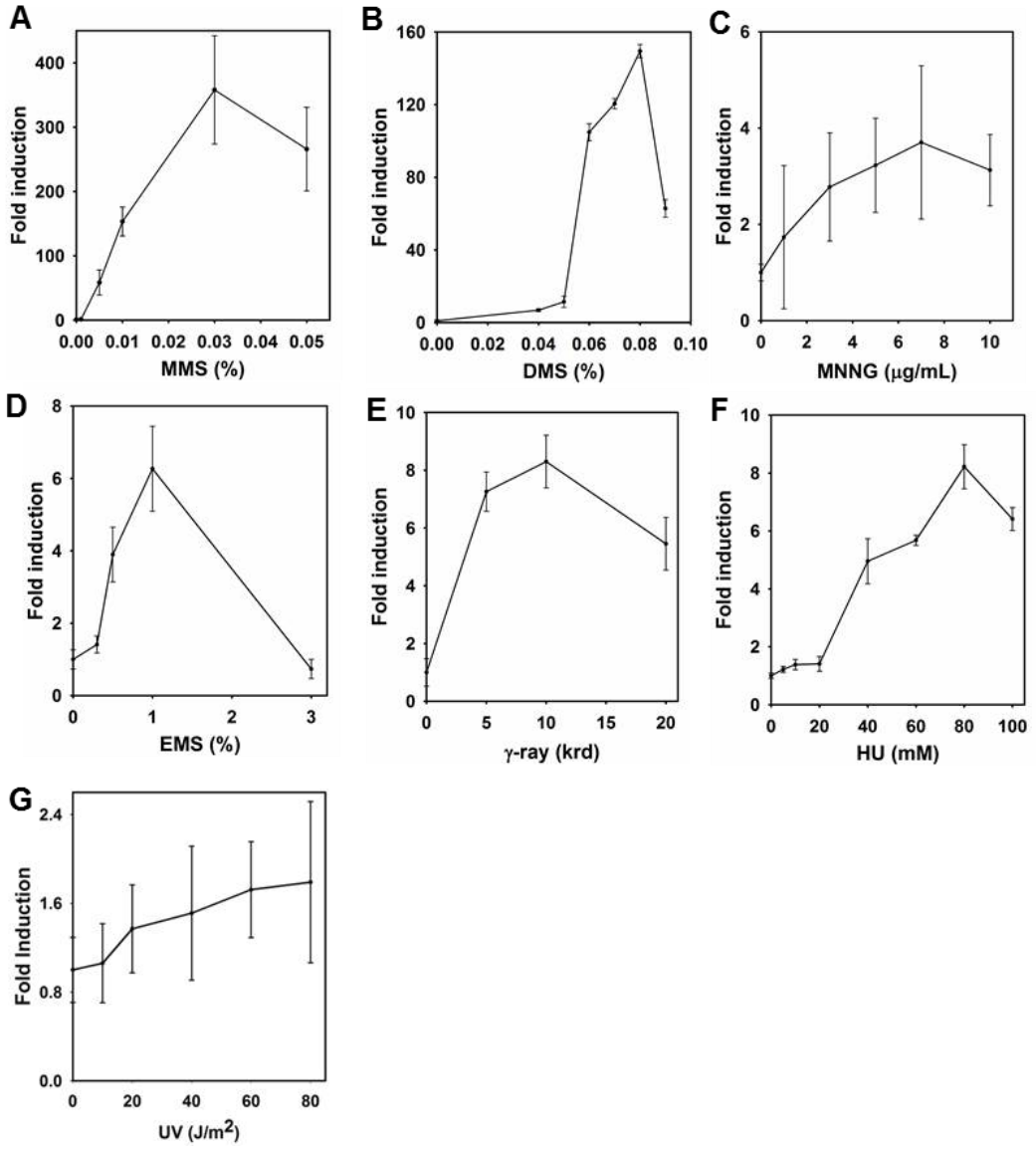


Figure 3–2. The induction of *DDI2-lacZ* by different DNA-damaging agents.

β-gal assays were performed by Fu, Y., the results after various DNA-damaging agents are shown as induction relative to untreated. All data represent the average of at least three experiments with standard deviations. Adopted from (Fu 2008).

DDI2/3 genes are not typical DNA damage inducible genes and that they may be involved in a cellular metabolic response related to SN_2-type alkylation stress. It also suggests that its induction is regulated by a unique mechanism and that the gene may function differently than most other DNA damage-inducible genes.

3.2.2. *DDI2/3* Encodes A Peptide Containing Cyanamide Hydratase Conserved Domain

DDI2/3 encodes a polypeptide (Ddi2/3) of 225 amino acids having a characteristically conserved domain named cyanamide hydratase domain. This domain is

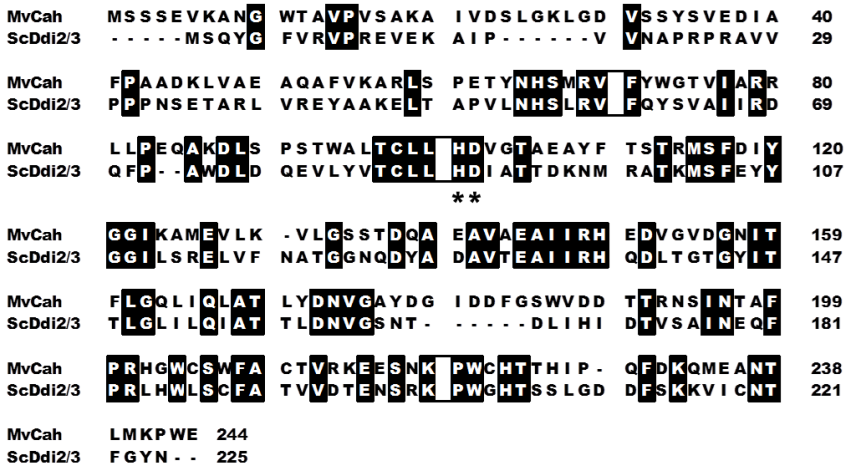


Figure 3–3. Protein sequence alignment of yeast Ddi2/3 and identified Cah from *Myrothecium verrucaria*.

Sequence alignment was performed by ClustalW (Larkin *et al*., 2007). Identical residues are shaded, and about 36% of Ddi2/3 protein is identical to Cah. The HD doublet residues which are the signature of HD domain are pointed out by asterisks.

related to an identified cyanamide hydratase (EC 4.2.1.69), Cah, from a soil fungus *Myrothecium verrucaria* (Maiergreiner *et al*., 1991). Protein sequence alignment of Ddi2/3 and MvCah indicates that Ddi2/3 is a homolog of MvCah.

3.2.3. Hypothetic functions of *DDI2/3*

Due to the sequence similarity to identified Cah, Ddi2/3 is hypothesized to possess cyanamide hydratase activity as well. Since the transcription levels of *DDI2/3* genes are induced by MMS, the genes are hypothesized to be involved in DNA repair to MMS damage.

3.3. Results

3.3.1. Prokaryotically Expressed Ddi2/3 Exhibits Cyanamide Hydratase Activity

To date, there is only one reported characterization of cyanamide hydratase (Cah) isolated from *M. verrucaria* cell extract, with a measured K_M of 27 mM. Cah was identified as a zinc metalloenzyme, and it formed homo-hexamer as a functional enzyme (Maiergreiner *et al*., 1991). Yeast Ddi2/3 has been characterized containing the cyanamide hydratase domain, and its protein sequence shows homology to Cah (Figure 3–3). Therefore it is hypothesized that Ddi2/3 could exhibit same enzymatic activity to Cah.

To test our hypothesis, *DDI2/3* ORF was cloned into bacterial expression vector pGEX6p-1 (GE Healthcare) and the recombinant protein was expressed as a GST fusion. Following overexpression, purification and PreScission protease (GE Healthcare) cleavage to remove the GST tag, the resulting recombinant Ddi2/3 contains eight

additional amino acid residues (Gly-Pro-Leu-Gly-Ser-Pro-Asn-Ser) at the N-terminus, with a calculated molecular weight of 25.8 kDa. The cleaved Ddi2/3 protein was purified to apparent homogeneity as identified by SDS-PAGE, and the corresponding western blot shows that it is free of detectable GST contamination (Figure 3–4A).

For His_6-tagged recombinant Ddi2/3 expressed from pET28 expression system, it contains two additional amino acids (Met-Gly) at the N-terminus and 11 additional amino

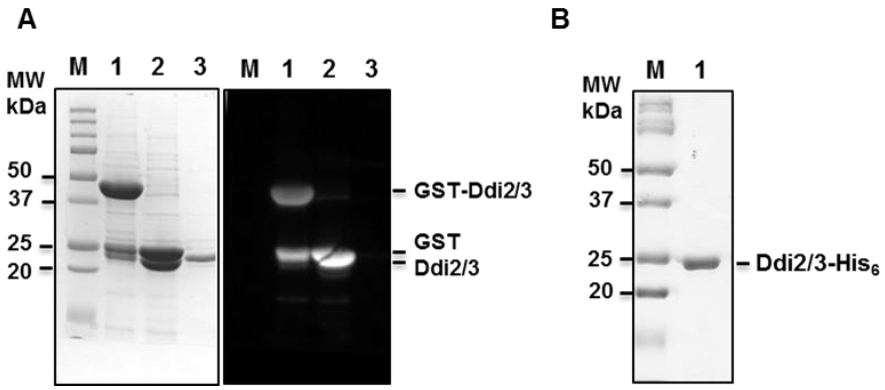


Figure 3–4. SDS-PAGE images showing purification of the recombinant Ddi2/3 protein.

A. SDS-PAGE gel image and the corresponding anti-GST western blot to demonstrate the purification of the recombinant Ddi2/3 protein obtained from pGEX expression system. *Lane M,* Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1*, purified GST-Ddi2/3; *lane 2,* GST-Ddi2/3 incubated with PreScission Protease for 16 hours; *lane 3,* recombinant Ddi2/3 after removal of the GST tag, which was used in the enzymatic assay. In the anti-GST western blot, the primary antibody was goat anti-GST, the secondary antibody was bovine anti-goat IgG. ***B.*** SDS-PAGE gel image to show the purification of the recombinant Ddi2/3 protein from *pET*28 expression system. *Lane M*, Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1*, purified His_6-tagged Ddi2/3.

acids (Ala-Ala-Ala-Leu-Glu-His_6) at the C-terminus. The purified Ddi2/3-His_6 protein was analyzed by SDS-PAGE and its molecular weight is agreeable to its theoretical molecular weight of 26.6 kDa (Figure 3–4B).

Maiergreiner *et al.* first isolated and identified cyanamide hydratase from the fungus *M. verrucaria* in 1991. The initial Cah enzymatic activity was determined by measuring both the consumption of cyanamide after incubation with Cah for 15-60 min, followed by a colorimetric assay at 530 nm (Steller *et al*., 1965), and the urea production. Since the decreased cyanamide and produced urea were of equal moles, the cyanamide hydratase activity of Cah was confirmed. The enzymatic activity is presented as a reaction rate, which needs to be measured with substrate concentration as high as five times K_M to guarantee that the enzyme is saturated, and the reaction rate approaches V_{max}

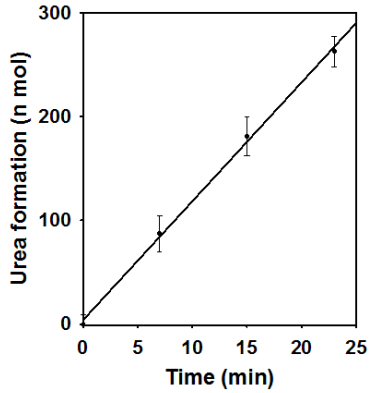


Figure 3–5. Time course monitoring of urea formation by recombinant Ddi2/3.

Protein applied was 0.7 μg (27 nM), cyanamide concentration was 44.8 mM, linear regression R^2=0.998, calculated reaction rate was 11.4 n mol*min^{-1}.

to ensure an accurate K_M of the enzyme. Since the established K_M of Cah is 27 mM, and Ddi2/3 is expected to have similar properties, the reported colorimetric assay is not likely to provide accurate data on enzymatic reaction rate with high cyanamide concentrations.

Here a novel enzymatic assay to cyanamide hydratase was developed based on the determination of the product urea concentration. In principle, excessive urease in the cyanamide hydratase reaction allows immediate hydrolysis of the produced urea to ammonia, which can be quantified using a commercial ammonia assay kit. The kit was based on glutamate synthesis reaction, utilizing ammonia, alpha-keto glutarate (KGA) and glutamate dehydrogenase (GDH), monitoring the consumption of partner molecule NADPH (Mondzac *et al*., 1965; Olson and Anfinsen, 1953). By combining the two reactions, the real-time monitoring of urea formation can be archived. However, cyanamide produces a background in this assay, which limits the reaction to 15 min and less than 5 mM cyanamide. When more cyanamide was used in the reaction, 1/10 volume of the reaction was applied to the ammonia assay kit. The time courses of urea formation can be plotted, as shown in Figure 3–5.

Initial velocities of the cyanamide hydration reaction to various concentrations of the substrate were measured, and an initial velocity versus substrate concentration curve for Ddi2/3 was plotted using cyanamide concentrations ranging from 0.3 mM to 88 mM for 0.027 μM (1 μg) Ddi2/3. As shown in Figure 3–6A, the kinetic curve was fitted by the program SigmaPlot12 with the Michaelis-Menten equation, in which the R^2 of regression fitting reached 0.9997 and Ddi2/3 fits a one-site saturation model for cyanamide. In summary, the recombinant Ddi2/3 exhibits a K_M = 17.3 ± 0.5 mM, V_{max}= 15.9 ± 0.2 n mol* $min^{-1,}$ and k_{cat} = 9.8 ± 0.1 n mol * s^{-1} * n mol^{-1} enzyme.

To rule out the possibility that the N-terminal extra amino acid residues in the recombinant Ddi2/3 influence its activity, purified His_6-tagged recombinant Ddi2/3 was also expressed from pET28 expression system and tested in the Cah assay, with similar activities obtained (Figure 3–6B). 1 μg of His-tagged Ddi2/3 was applied in the kinetic assay as described above. The kinetic curve was fitted by the program SigmaPlot12 with the Michaelis-Menten equation, the R^2 of regression fitting reached 0.9962 and Ddi2/3

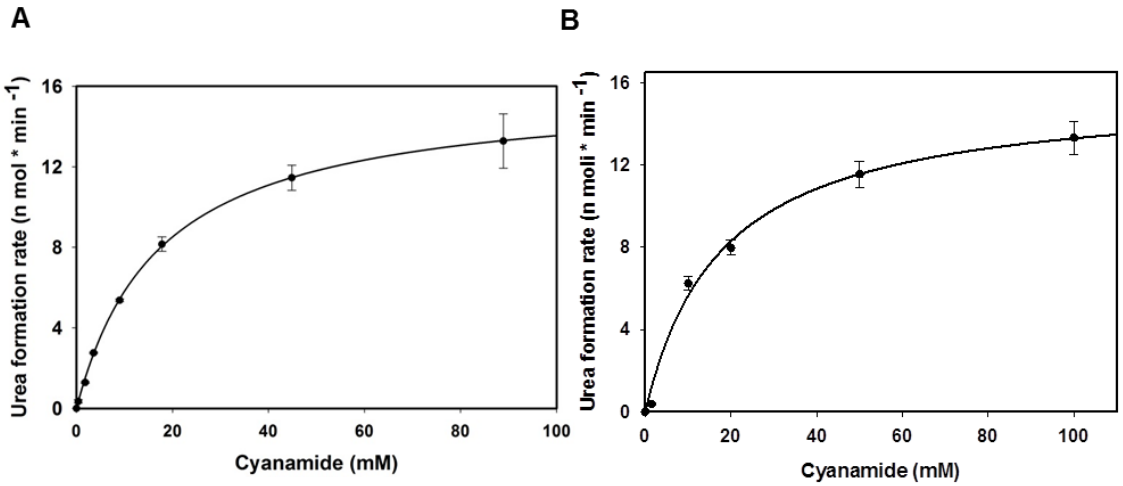


Figure 3–6. Kinetic studies of prokaryotically expressed Ddi2/3.

A. Michaels-Menten curve of Ddi2/3 obtained from pGEX expression system to cyanamide. 0.7 μg (27 nM) Ddi2 was applied in the assay. The urea formation rates to 0.3 mM-88 mM cyanamide were measured. ***B***. Michaels-Menten curve of Ddi2/3 obtained from pET28 expression system to cyanamide. 1.0 μg (37.6 nM) Ddi2 was applied in the assay. The urea formation rates to 5 mM-100 mM cyanamide were measured. Non-linear regression was determined by the SigmaPlot12 program with the Michaels-Menten equation.

fits a one-site saturation model for cyanamide. In summary, the recombinant His-tagged Ddi2/3 exhibits a K_M = 16.5 ± 2.0 mM, V_{max}= 15.4 ± 0.6 n mol* $min^{-1,}$ and k_{cat} = 6.8 ± 0.3 n mol * s^{-1} * n mol^{-1} enzyme.

3.3.2. Screen for Inhibitors to Ddi2/3's Enzymatic Activity

Maiergreiner *et al.* reported that Cah displayed high specificity on cyanamide. Other compounds closely related chemically to cyanamide, such as formylcyanamide (HCO-NH-CN), acetylcyanamide (H_3C-CO-NH-CN), acetonitrile (H_3C-CN), cyanate (OCN-), cyanide (CN-), dicyandiamide (H_2N-CNH-NH-CN) cyanourea (H_2N-CO-CNH-CN), formamide (H_2N-COH), urea (H_2N-CO-NH_2), or creatine, were incubated with Cah, but the chemicals were not hydrated to any detectable extent.

The inhibitions of Cah by a variety of chemicals were also studied. Maiergreiner *et al.* reported that the enzymatic activity is sensitive to chelating agents such as EDTA or o-phenanthroline, pointing to a functional role of zinc in the protein. Compounds containing structural elements of cyanamide, such as urea, dicyandiamide, azide (N-NN^-), hydroxylamine (H_2N-OH), thiocyanate (SCN^-), or cyanate (OCN^-) the activity. The most potent inhibitor among them is cyanide (Maiergreiner *et al*., 1991).

Based on his report, a series of chemicals were selected and tested in cyanamide hydratase assay to determine whether they are inhibitors to Ddi2/3. These chemicals include metal chelates, such as ethylenediaminetetraacetic acid (EDTA), sodium azide; chemicals with structural similarities to cyanamide, such as cyanide salt, cyanate salt, thiocyanate salt, acetonitrile, and acetohydroxamic acid were also tested.

The purified Ddi2/3 protein was adjusted to 0.5 mg/ml (0.019 mM), and incubated with 10 mM of possible inhibitors, including EDTA, cyanide (NaCN), urea, hydroxyurea, azide (NaN_3) cyanate (KOCN), thiocyanate (NaSCN), acetonitrile (CH_3CN) and acetohydroxamic acid (AHA), at room temperature for 30 min. The above mixtures were then incubated with 5 mM cyanamide for another 2 h and the remaining cyanamide was monitored by the colorimetric assay. PBS buffer was used as a negative control;

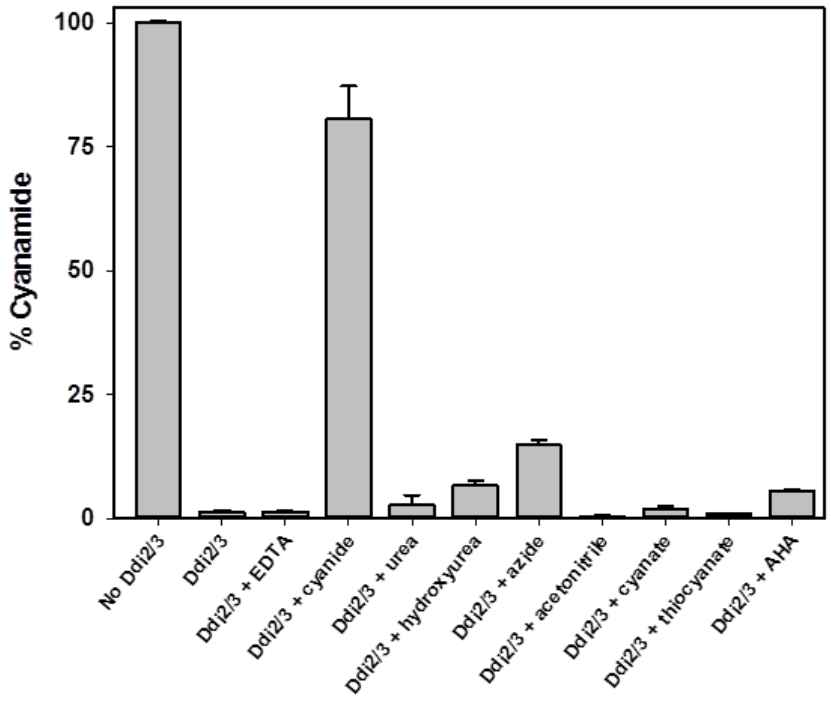


Figure 3–7. Cyanamide hydratase assay of Ddi2/3 in the presence of inhibitors.

His-tagged Ddi2/3 was adjusted to 0.5 mg/ml (0.019 mM) and incubated with 10 mM each of chemicals for 30 min before incubation with 5 mM cyanamide. Remained cyanamide was monitored by a colorimetric assay as previously described (Section 2.4.2.). Three independent assays were performed to calculate the mean and standard deviations.

A_{530nm} of the negative control was taken as 100%. Consumed cyanamide reflected the enzymatic activity in the presence of each chemical. As shown in Figure 3–7, cyanide shows the strongest inhibition of Ddi2/3's enzymatic activity. In addition to cyanide, hydroxyurea, azide and AHA also display moderate inhibitory effects to Ddi2/3. Urea and cyanate have little impact on the Ddi2/3's activity, the rest of chemicals do not show any inhibition of Ddi2/3's activity.

3.3.3. The HD Domain Is Required for Ddi2/3's Enzymatic Activity

HD domains are found in a superfamily of enzymes with metal-dependent phosphohydrolase activity (Aravind and Koonin, 1998). The HD domain is named due to the most conserved HD residues in the region, and they have displayed a characteristic …H…HD….D… pattern. The doublet HD residues are the signature of HD domain. The conserved HD residues are have been identified as metal-binding ligands, and substrates interact with the surrounding amino acids, from structural studies of several HD family members (Brown *et al*., 2006; Kondo *et al*., 2007). Mutations in the conserved HD residues drastically affect the enzymatic activity (Zimmerman *et al*., 2008). Since the cyanamide hydratase domain identified in Ddi2/3 belongs to the HD domain superfamily, it is wondered whether the HD domain contributes to the cyanamide hydratase activity. In NCBI conserved domain database (Marchler-Bauer *et al*., 2015), cyanamide hydratase domain family has 9 entries including Cah. Among the rest 8 proteins, only Ddi2/3 has been identified. It is noticed that all the 9 proteins are from fungi, especially from *Ascomycota* phylum. The sequence alignment of all 9 putative cyanamide hydratase domain containing proteins (Figure 3–8) reveals 7 conserved HD

residues including Ddi2/3-H55, H88, D89, H137, D139 D160, and H205. The doublet HD pair, H88 and D89, is the signature of HD domain. If these residues were critical to the Ddi2/3 enzymatic activity, site-specific mutations of these residues would compromise the enzymatic activity.

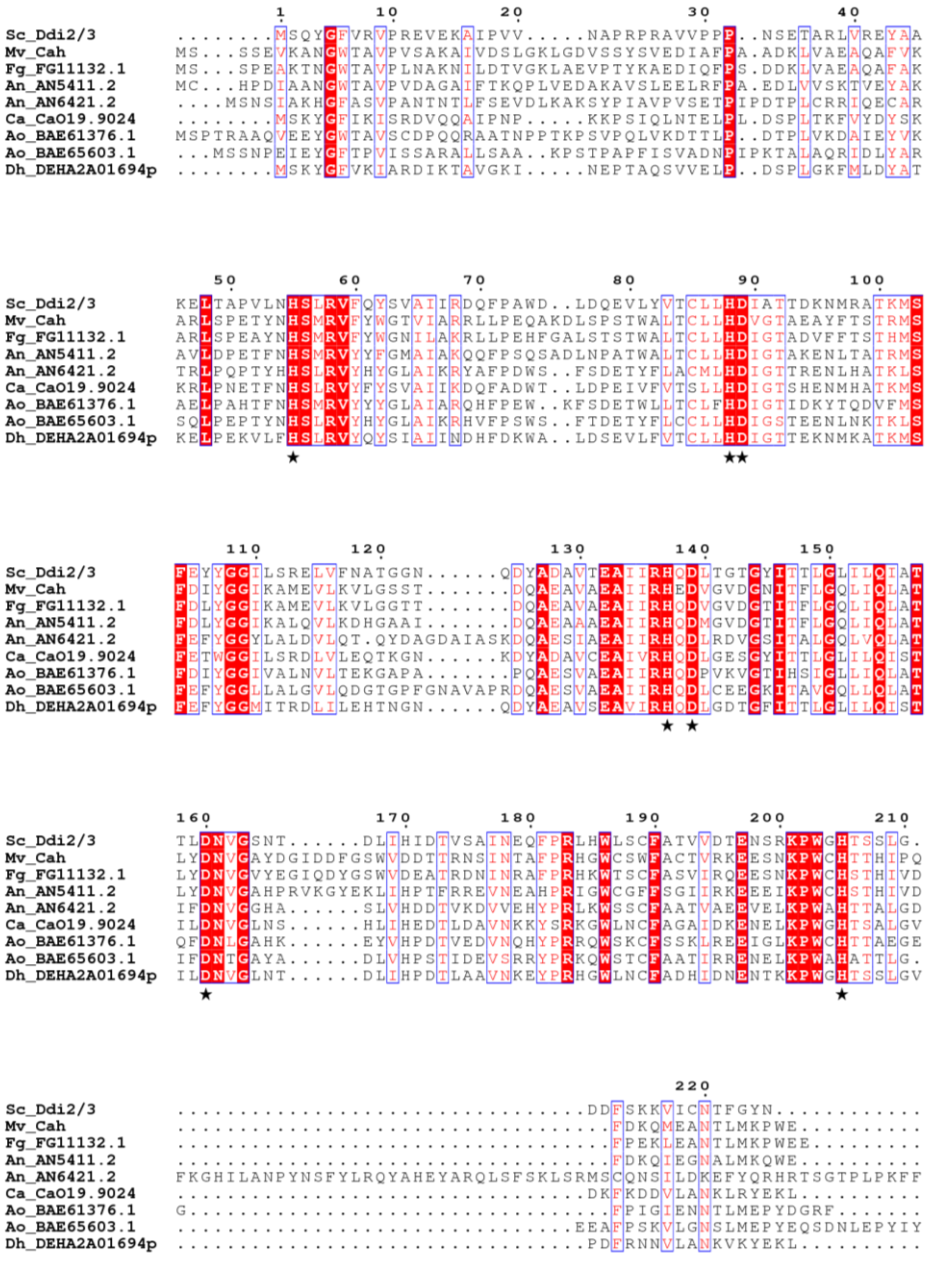


Figure 3–8.Sequence alignment of cyanamide hydratase domain containing proteins.

Amino acid sequences of proteins were retrieved from the NCBI conserved domain database (Marchler-Bauer *et al*., 2015). Sequence alignment was processed by ClustalW (Larkin *et al*., 2007) and presented by ESpript 3.0 (Robert and Gouet, 2014), globe score was set as 0.8. GenBank names, the source organism abbreviation are given in the left-hand column. Sc, *S.cerevisiae*; Mv, *M. verrucaria*; Fg, *Fusarium graminearum*; An, *Aspergillus nidulans*; Ca, *Candida albicans*; Ao, *Aspergillus oryzae*; Dh, *Debaryomyces hansenii.* Identified residues are colored in red. Conserved HD residues are pointed out by stars.

At last, two such Ddi2/3 double mutants, Ddi2/3-H88A; D89A and Ddi2/3-H137A; D139A, in which both HD residues were mutated to alanine, were created, and their enzymatic activities were tested. During recombinant protein preparation, it was noticed that the H137A; D139A mutation on Ddi2/3 severely affected protein yield and solubility, while the Ddi2/3-H88A; D89A protein was readily expressed and purified (Figure 3–9A). As shown in Figure 3–9B, both HD mutant proteins completely lost their

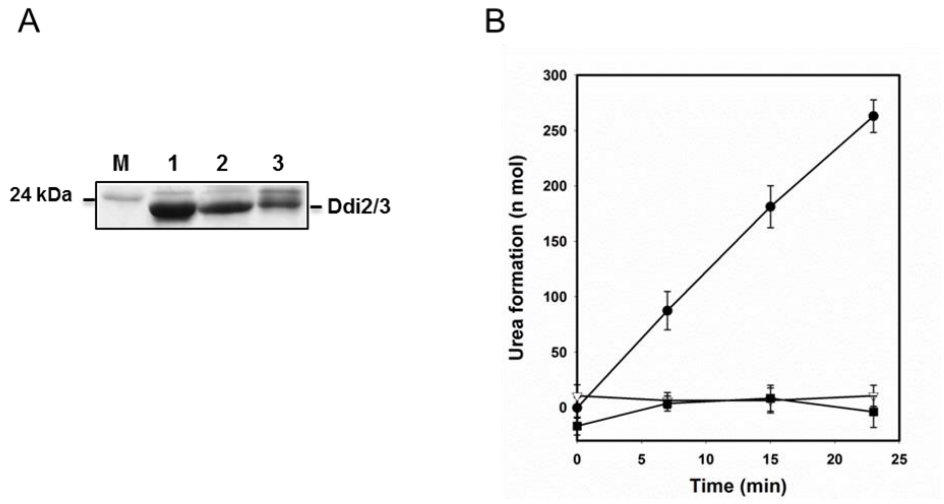


Figure 3–9. SDS-PAGE gel showing purification of Ddi2/3-H88A; D89A and Ddi2/3-H137A; D139A, and the enzymatic activities of the above Ddi2/3 mutants.

A. SDS-PAGE gel to demonstrate the purification of the recombinant Ddi2/3 proteins and its mutants obtained from pGEX expression system. *Lane M,* Precision Plus Protein™ Unstained Standards (from Bio-Rad); *lane 1,* purified wild type Ddi2/3; *lane 2,* purified Ddi2/3-H88A; D89A; *lane 3,* purified Ddi2/3-H137A; D138A. ***B.*** Time courses of urea formation of Ddi2/3 and its site-specific mutants. Each reaction contained 44.8 mM cyanamide and 0.7 µg wild type Ddi2/3 or 10 µg mutant Ddi2/3. (●) wild type Ddi2/3; (▽) Ddi2/3-H88A; D89A and (■) Ddi2/3-H137A; D139A. Note that Ddi2/3-H137A; D139A was not purified to the same extent as WT and Ddi2-H88D89AA proteins.

cyanamide hydratase activity, indicating that the HD domain is essential for the above enzymatic activity. It also suggests that the required metal ions can be incorporated into prokaryotically expressed Ddi2/3.

3.3.4. *DDI2/3* is Massively Induced by Cyanamide at the Transcriptional Level

CAH was reported as a cyanamide inducible gene (Stransky and Amberger, 1973). Since *DDI2/3* encodes a cyanamide hydratase that is a homolog of CAH, it is wondered whether *DDI2/3* may also be induced by cyanamide. To test this hypothesis, the *lacZ* gene encoding β-galactosidase was fused to *DDI2/3* promoter as a reporter gene. The amount of expressed β-galactosidase after induction was monitored by a β-gal assay; the fold induction of β-galactosidase after treatment represented the response of *DDI2/3* promoter to the induction signal. Results of the β-gal assay were as shown in Figure 3–10A, indicating that as low as 0.01% MMS can induce *DDI2/3* over 150-fold, and treatment with 0.03% MMS achieves the highest induction of over 350-fold. Under the same experimental conditions, cyanamide can lead to even higher induction than MMS. As shown in Figure 3–10B, 10 mM cyanamide reached induction of over 600-fold. Hence, *DDI2/3* is identified as a cyanamide-inducible gene. Moreover, compared to MMS, cyanamide induction can archive higher induction fold.

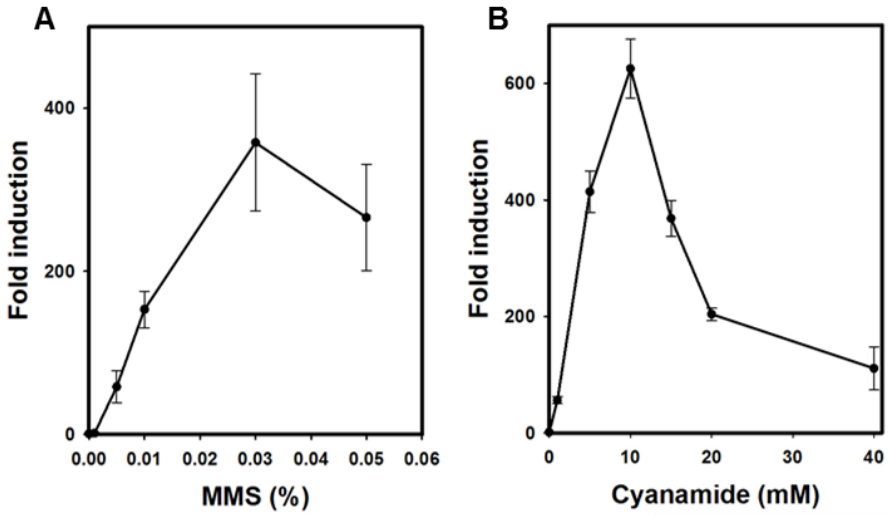


Figure 3–10. β-gal assay with *DDI2/3-lacZ* to MMS and cyanamide.

A*.** Induction of *DDI2/3-lacZ* by MMS. ***B. Induction of *DDI2/3-lacZ* by cyanamide. Experimental conditions are as described previously. β-gal assays were performed by Biss, M., the results after induction are shown as fold induction relative to untreated. All data represent the average of at least three experiments with standard deviations. Adopted from (Li *et al*., 2015).

3.3.5. *DDI2/3* Confers Resistance to Both Cyanamide and MMS

To characterize the biological functions of *DDI2/3*, two disruption cassettes with different selectable markers were created (Figure 3–11A) and applied in a sequential gene deletion procedure. Disruptions of *DDI2/3* genes were confirmed by genome PCR. As shown in Figure 3–11B, the *ddi2/3*Δ*::HIS3* and *ddi2/3*Δ*::LEU2* single mutants contained a wild type copy and a copy of a selectable marker gene used in disruption, although the PCR method cannot distinguish which chromosomal copy was disrupted. Nevertheless, in the *ddi2/3*Δ*::HIS3 ddi2/3*Δ*::LEU2* double mutant, both *DDI2* and *DDI3* genes were disrupted, and the strain no longer contained a wild type allele. The growth of the *ddi2/3*Δ single and double mutants was tested by a serial dilution assay in the presence of MMS or cyanamide. As shown in Figure 3–11C, the *ddi2*Δ *ddi3*Δ double mutant displayed increased sensitivity to cyanamide compared with the parental wild type strain, suggesting that *DDI2/3* protects cells from cyanamide toxicity. It was surprising that the *ddi2/3*Δ single mutant displayed a level of sensitivity similar to the double mutant, suggesting that although *DDI2* and *DDI3* are identical genes, they are not simply redundant in function. Otherwise, the double mutant would show enhanced sensitivity than that of single mutant. Similarly, both single and double mutants displayed comparable levels of sensitivity to MMS.

However, the *ddi2/3Δ* single mutant strain created in this way cannot identify which gene of the two copies was disrupted Two possibilities are proposed to explain the observation; the first hypothesis is that one copy of *DDI2/3* is not transcribed, as a silenced or dysfunctional gene. The second hypothesis is that the toxicity of the present

chemical is not strong enough to let double mutants display enhanced sensitivity to single mutants.

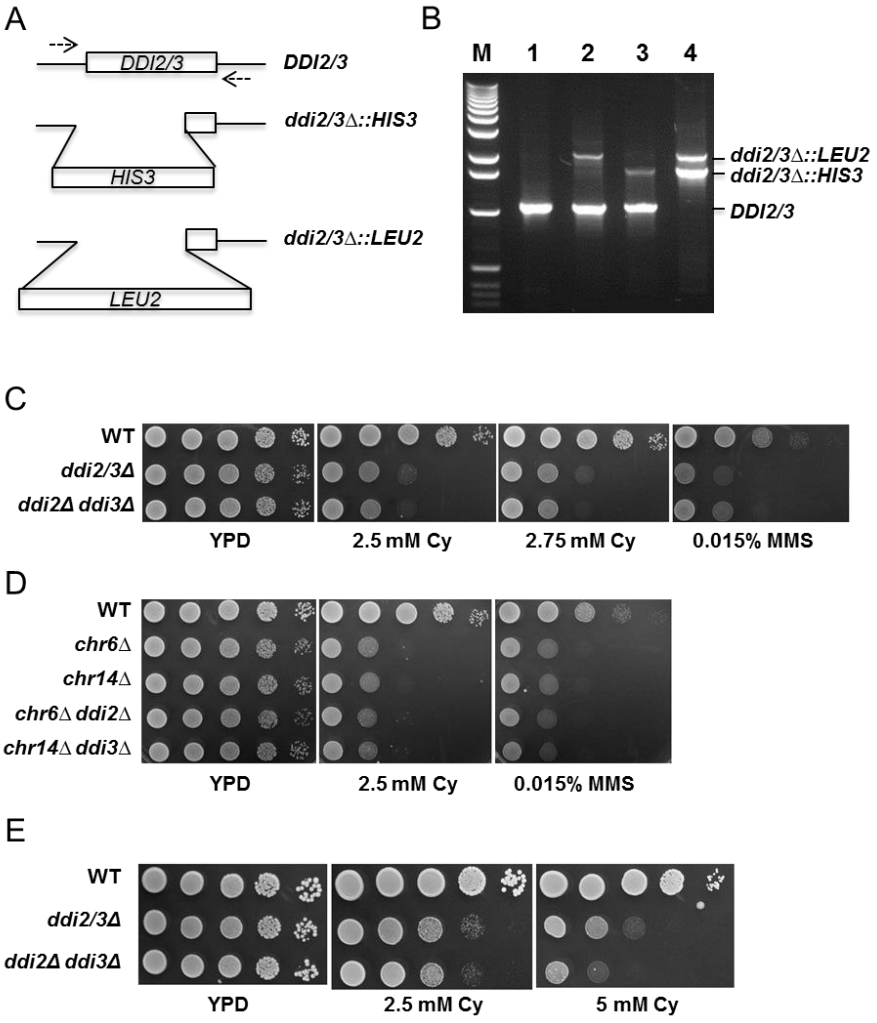


Figure 3–11. Sensitivity assays of yeast *ddi2/3Δ* and *ddi2Δ ddi3Δ* mutants.

A. Scheme of disruptions of *DDI2* and *DDI3* gene in yeast by utilizing homologous recombination. Arrows represent the primers binding sites in PCR. ***B***. Genomic PCR confirms the disruptions of *DDI2/3* genes. ***C, D*** and ***E.*** Sensitivity assays with wild type yeast, *ddi2/3Δ* mutant and *ddi2Δ ddi3Δ* mutant to cyanamide and MMS. Notice that the pictures of panel ***C*** and ***D*** were taken after two days incubation while panel ***E*** shows the picture after three days incubation. Adopted from (Li *et al*., 2015).

To further distinguish the role of individual *DDI2* and *DDI3* gene, chromosomal deletion strains were created by Xin Xu, by targeting each of the two repeats, followed by deleting the single remaining *DDI2* or *DDI3* gene (Li *et al*., 2015). As seen in Figure 3–11D, deletion of either single chromosome region alone is sufficient to confer cyanamide or MMS sensitivity regardless of further *DDI2* or *DDI3* deletion, which effectively rules out the first possibility. To test the second possibility, we thought that perhaps the agent concentration is the most influential parameter. Indeed, at higher cyanamide concentrations, the double mutant is apparently more sensitive than the single isogenic mutant (Figure 3–11E). When treated with 5 mM cyanamide, *ddi2Δ ddi3Δ* double mutant shows enhanced sensitivity than single mutant by about 10-fold.

3.3.6. Endogenous Ddi2/3 Exhibits Cyanamide Hydratase Activity

The study of recombinant Ddi2/3 has shown that it exhibits cyanamide hydratase activity. To further confirm that *DDI2/3* genes encode cyanamide hydratase, endogenous Ddi2/3 isolated from yeast is required for the enzymatic assay. An ideal approach would be to purify endogenous Ddi2/3 from yeast cells through chromatography for the enzymatic analysis. An alternative approach is to tag the endogenous Ddi2/3 through genetic manipulations followed by affinity purification and enzymatic analysis.

The first trial was to culture wild type yeast and the *ddi2Δ ddi3Δ* double mutant in media containing cyanamide, and remained cyanamide in media was monitored by utilizing the colorimetric assay as previously described. In this colorimetric assay, the A530nm is proportional to the cyanamide concentration up to 5 mM, as shown by a standard curve in Figure 3–12A. As the result shown in Figure 3–12B, after 36 hours

incubation, wild type cells used 80% cyanamide in the medium, while about 90% cyanamide remained in the *ddi2Δ ddi3Δ* double mutant, indicating that the consumption of cyanamide in the culture medium is primarily due to the activity of *DDI2/3* gene products.

The above result shows that wild type yeast cells consume more cyanamide in the medium than the *ddi2Δ ddi3Δ* double mutant cells, but the assay is time-consuming and it is an indirect measurement. To develop a more rapid and direct assay, the yeast whole cell extract (WCE) was prepared and tested in a cyanamide hydratase assay based on a colorimetric assay as previously described (Zhang *et al*., 2005). Yeast WCEs were adjusted to 1 mg/ml and was incubated with 5 mM cyanamide solution, and the colorimetric assay was used to monitor the reduction of cyanamide as quantitative measurement of the cyanamide hydratase activity. As shown in Figure 3–12C, the cyanamide concentration in the reaction mixture remained stable over 24 h in the absence of WCE, and the addition of WCE from wild type cells without the prior cyanamide treatment did not alter the cyanamide concentration. The addition of WCE from cyanamide-induced wild type cells resulted in the reduction of cyanamide concentration by 1/3 in 5 h and more than fivefold in 24 h, indicating that the cyanamide hydratase activity in yeast cells is inducible by cyanamide. In contrast, the WCE from the cyanamide-treated *ddi2/3*Δ single mutant cells showed moderate activity to use cyanamide, and WCE from cyanamide-treated *ddi2*Δ *ddi3*Δ double mutant cells was unable to reduce cyanamide concentration even over 24 h. The above observations indicate that *DDI2/3* is solely responsible for the cyanamide metabolism *in vivo* and most likely encodes cyanamide hydratase.

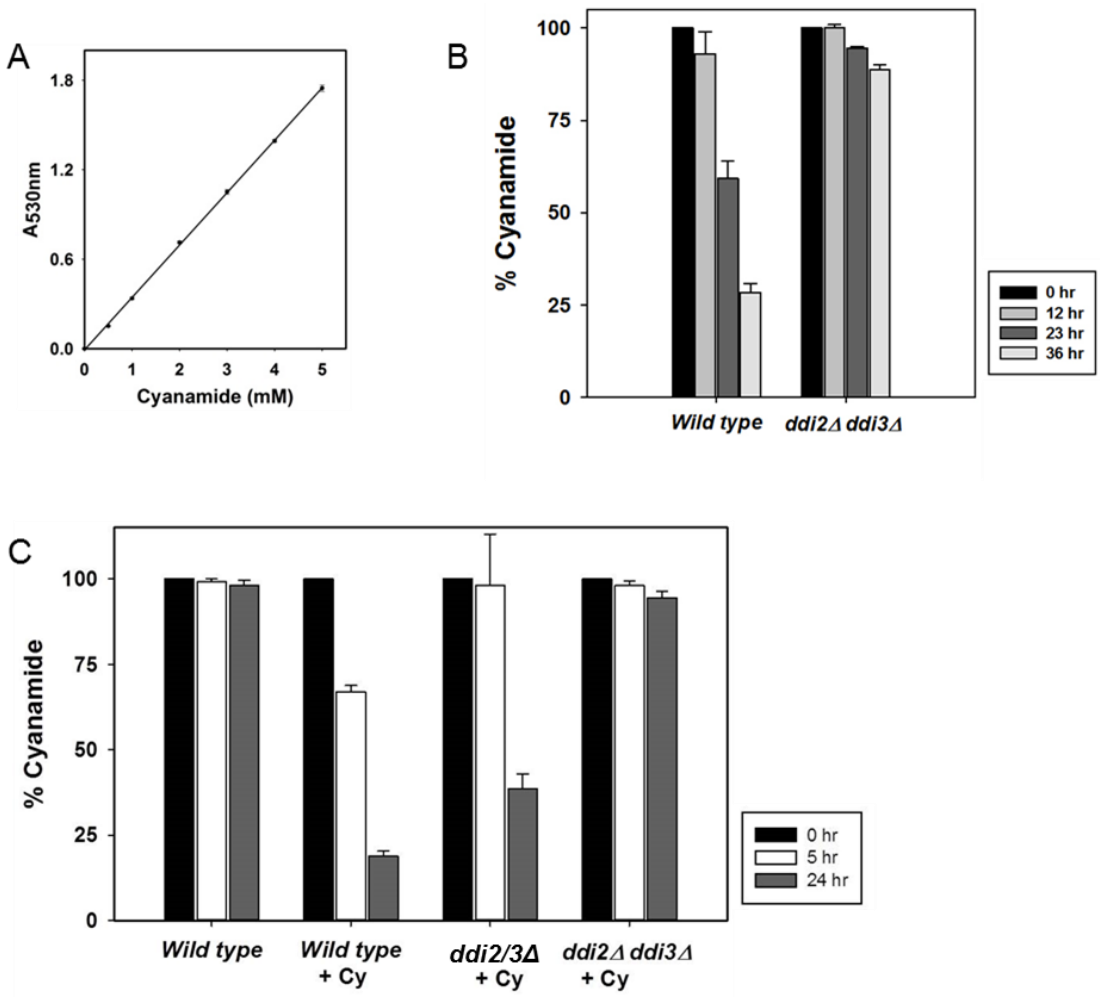


Figure 3–12. Cyanamide hydratase assays with yeast WCEs.

A. A standard curve of cyanamide in a colorimetric assay. ***B***. Cyanamide hydratase assays with yeast cultures. **C**. Cyanamide hydratase assays with yeast WCEs. WCEs were adjusted to 1 mg/ml and incubated with 5 mM cyanamide. Colorimetric assays were processed as previously described in Section 2.4.2.

3.3.7. Ddi2/3 Is Not Involved in Its Transcriptional Regulation Pathway

The fact that *DDI2/3* is highly induced by cyanamide, and the gene itself encodes a functional cyanamide hydratase, raises a possibility that *DDI2/3* may serve as a sensor in its induction pathway. To test whether Ddi2/3 are involved in the regulatory circuit, β-gal assays were performed by using YEpDDI2-lacZ with the treatment of both cyanamide and MMS in the *ddi2/3Δ* single mutant and *ddi2*Δ *ddi3*Δ double mutant. If the Ddi2/3 is required in transcription regulation, the disruptions of *DDI2/3* would compromise the induction by MMS and cyanamide. Especially in the *ddi2Δ ddi3Δ* double mutant, since no Ddi2/3 is present, the induction would be abolished. However, as seen in Figure 3–13,

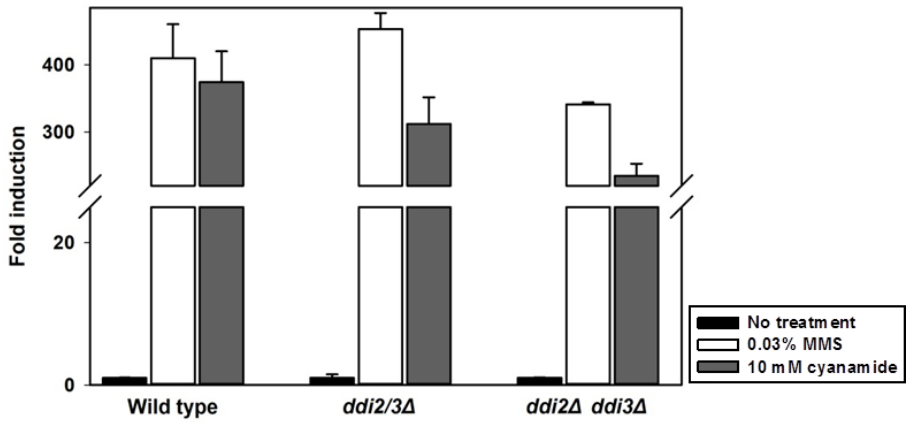


Figure 3–13. β-gal assays to MMS and cyanamide induction in *ddi2/3Δ* and *ddi2Δ ddi3Δ* mutants by using YEpDDI2-lacZ.

Deletion of both *DDI2* and *DDI3* genes does not affect YEpDDI2-lacZ induction by cyanamide or MMS. β-gal assays were performed by Biss, M., the results after induction are shown as fold induction relative to untreated. All data represent the average of at least three experiments with standard deviations. Experimental conditions are as described in section 2.6.

deletion of both *DDI2* and *DDI3* genes does not compromise the induction fold of MMS and cyanamide at all.

Figure 3–13 is presenting that both cyanamide and MMS activates the transcription of *DDI2/3*, and archive high fold induction of gene expression, in all *ddi2/3Δ* mutant strains. The disruptions of *DDI2/3* only has little effect to lower the induction fold, but still archives over 200 fold. While the induction of MMS is even less affected by the gene disruptions. This observation indicates that Ddi2/3 is not involved in the transcriptional regulation of its own gene.

3.3.8. *DDI2/3* Gene May Not Be Involved In Known DNA Repair Pathways

To investigate which DNA repair pathway Ddi2/3 may play a role, *MAG1*, *RAD1* and *YAP1* genes were disrupted from the *ddi2Δ ddi3Δ* double mutant to make three triple mutants strains: *ddi2Δ ddi3Δ mag1Δ, ddi2Δ ddi3Δ rad1Δ,* and *ddi2Δ ddi3Δ yap1Δ*.

MAG1 encodes a 3-methyl-adenine DNA glycosylase, which is involved in protecting DNA against alkylation damage. Mag1 initiates a BER pathway by removing damaged bases to create AP sites that are subsequently repaired (Xiao *et al*., 1994). The *mag1Δ* strain displays enhanced sensitivity to alkylating agents, such MMS (Chen *et al*., 1990; Chen and Samson, 1991; Prakash and Prakash, 1977). *RAD1* encodes a subunit of a single-stranded DNA endonuclease playing a role in the NER pathway and double-strand break repair (Prakash and Prakash, 2000; Symington, 2002). The *rad1Δ* mutant is sensitive to a variety of DNA-damaging agents including X-rays and UV radiation, and DNA synthesis inhibitors (Higgins *et al*., 1983). *YAP1* encodes a basic leucine zipper (bZIP) transcription factor required for oxidative stress tolerance (Moye-Rowley *et al*.,

1989). The *yap1Δ* mutant shows, enhanced sensitivity to oxidizing and reducing agents, carcinogens, antifungals, alkylating agents, metals, DNA metabolism inhibitors, protein and sterol synthesis inhibitors (Schnell *et al*., 1992; Wemmie *et al*., 1994). Disruption of *YAP1* increases the sensitivity of a yeast genotoxicity testing system to a broad range of DNA damaging agents (Zhang *et al*., 2011).

Constructed triple mutants strains: *ddi2Δ ddi3Δ mag1Δ, ddi2Δ ddi3Δ rad1Δ,* and *ddi2Δ ddi3Δ yap1Δ* were applied in gradient plates assays to test their sensitivities to DNA damaging agents, as shown in Figure 3–14. In panel A, the plate contained a gradient amount of MMS up to 0.03%, the length of growth reflected the resistance of cells to MMS. Here shows that, compared to wildtype, the *ddi2Δ ddi3Δ* double mutant did not show enhanced sensitivity, properly because the concentration of MMS was low, while the *mag1Δ* mutant was far more sensitive to MMS, since the growth of the mutant was much weaker than wildtype. If *DDI2/3* is involved DNA repair to MMS damage, the *ddi2Δ ddi3Δ mag1Δ* triple mutant should display enhanced sensitivity than the *mag1Δ* single mutant. However, the *ddi2Δ ddi3Δ mag1Δ* triple mutant did not perform as expected, in fact, it displayed almost same sensitivity as *mag1Δ* single mutant (Figure 3–14A), indicating that *ddi2/3* had no synergistic effect to *mag1*. This observation implies that *DDI2/3* genes are not involved in DNA repair to MMS-induced damage.

4-Nitroquinoline 1-oxide (4-NQO) is a quinoline derivative and a tumorigenic compound. 4-NQO mimics the biological effects of UV light on various organisms (Ikenaga *et al*., 1975). Both 4-NQO and its metabolite 4-hydroxyaminoquinoline 1-oxide (4-HAQO) bind covalently to cellular macromolecules, such as nucleic acids and proteins (Tada, 1975). 4-NQO also damages DNA through the production of reactive oxygen

species thought to arise from the enzymatic reduction of its nitro group (Arima *et al*., 2006). In response to damage from 4-NQO, cells attempt to repair and initiate a transcriptional response to detoxify the cell from 4NQO and its metabolites (Fry *et al*., 2005). DNA lesions caused by 4-NQO is usually corrected by the NER pathway. Since *RAD1* is required for NER, inactivation of the gene makes cells very sensitive to 4-NQO, as shown in Figure 3–14B. Moreover, the *yap1Δ* mutant also shows increased sensitivity to 4-NQO (Figure 3–14C) because the chemical functions as an oxidation agent as well. Meanwhile, *ddi2Δ ddi3Δ* double mutant displayed same resistance as wild type.

However, both *ddi2Δ ddi3Δ rad1Δ* and *ddi2Δ ddi3Δ yap1Δ* triple mutants did not show enhanced sensitivities compared to relative either *rad1Δ* or *yap1Δ* single mutant, indicating that *ddi2/3* genes are not synergistic to either *rad1* or *yap1*.

Although we can rule out a possibility that the above triple mutants or other uncharacterized triple mutants display an enhanced sensitivity to a specific DNA damaging agent, the above observations collectively imply that *DDI2/3* is unlikely to play a direct role in DNA repair.

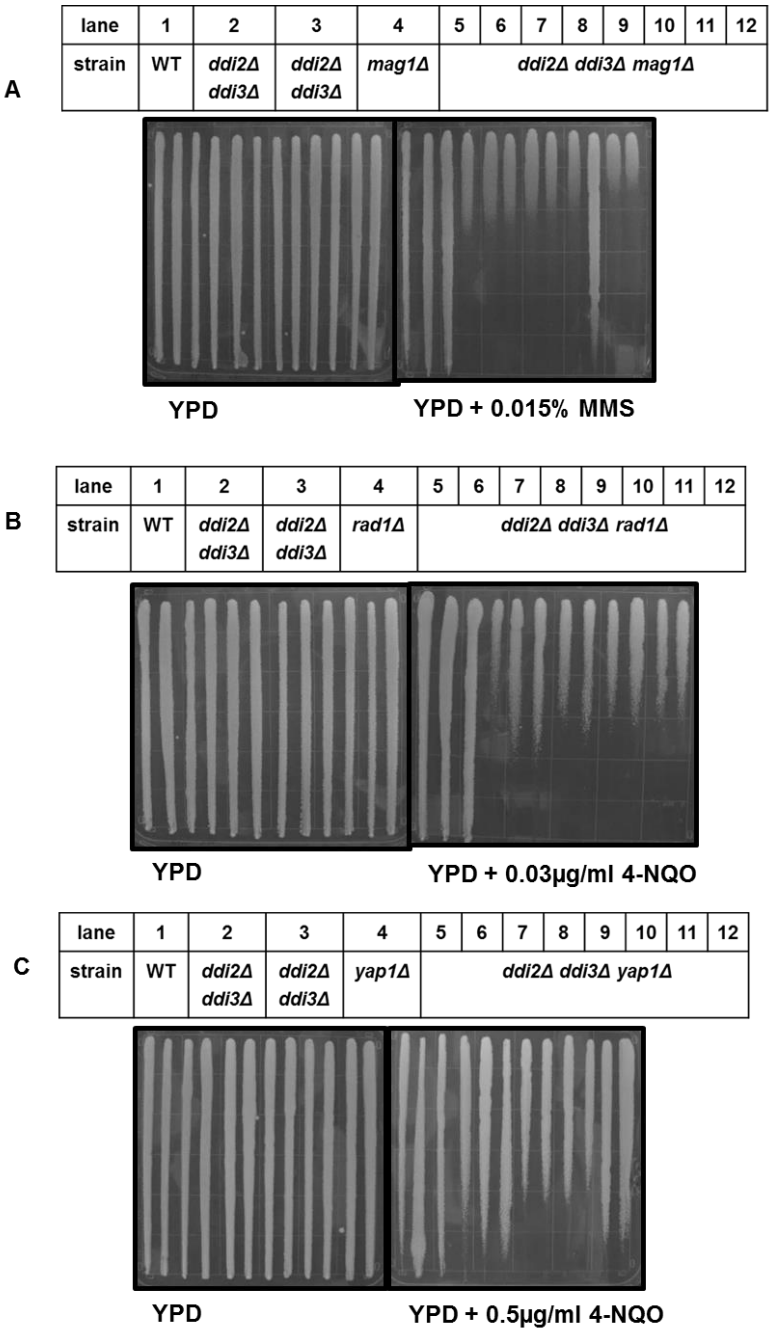


Figure 3–14. Gradient plate assays of triple mutants yeast cells to MMS and 4-NQO.

The addition of disruptions of *DDI2* and *DDI3* did not make cell display enhanced sensitivity than a *mag1Δ*, *rad1Δ* and *yap1Δ* single mutant. Two *ddi2Δ ddi3Δ* double mutants and seven triple mutant strains were applied on plates to avoid systematic errors.

3.4. Summary and Discussion

3.4.1. *DDI2/3* Encodes a Functional Cyanamide Hydratase

In this study, biochemical activity and biological functions of two duplicated genes, *DDI2* and *DDI3*, from budding yeast *S. cerevisiae* were investigated. Prokaryotically expressed recombinant Ddi2/3 protein was identified to have cyanamide hydratase activity that can catalyze the urea formation from cyanamide. Enzymatic assay with yeast WCEs also confirmed cyanamide hydratase activity of endogenous Ddi2/3. Similar to its homolog Cah found in *M. verrucaria*, Ddi2/3 was also identified as a cyanamide inducible gene. Disruptions of Ddi2/3 genes made cells more sensitive to both MMS and cyanamide.

Cyanamide hydratase (EC 4.2.1.69, or urea hydrolase) is a poorly studied enzyme and to date, there is only one report on its enzymatic activity based on proteins extracted from the fungus *M. verrucaria* (Maiergreiner *et al*., 1991). Our study reports for the first time real-time monitoring of a recombinant cyanamide hydratase reaction and the determined K_M for Ddi2/3 (17 mM) is comparable to that of the reported native *M. verrucaria* Cah (27 mM). This relatively low substrate affinity indicates that either cyanamide is not a physiological substrate, or that Ddi2/3 and Cah are not effective enzymes to hydrolyze cyanamide into urea.

Besides the enzymatic assay, several pieces of evidence are consistent with Ddi2/3 as a *bona fide* cyanamide hydratase. Firstly, since cyanamide is synthesized by limited plants as vetch species, the presence of such an enzyme may be vital to the survival of soil bacteria and fungi. Secondly, the expression of *DDI2/3* is highly induced by cyanamide, which may explain the poor efficiency of the enzyme for the reaction

($k_{cat}/K_M = 5.67 * 10^2\ M^{-1}s^{-1}$). Thirdly, it is shown that deletion of *DDI2* and *DDI3* genes sensitize cells to as low as 2 mM cyanamide in the medium, suggesting that the genes protect cells from cyanamide toxicity. Fourthly, yeast cells lacking *DDI2/3* genes or wild type cells without prior cyanamide induction were unable to metabolize cyanamide, whereas wild type cells treated with cyanamide can metabolize cyanamide, indicating a cyanamide-inducible hydratase activity *in vivo*. Finally, microorganisms with putative cyanamide hydratase activity may be able to utilize cyanamide as a carbon and/or nitrogen source. To this end, it is of great interest to notice that the budding yeast *DUR1* and *DUR2* genes encode two ureases that effectively convert urea to ammonia and carbon dioxide (CO_2) (Olson and Anfinsen, 1953).

3.4.2. Cyanamide Hydratase Domain Is a Subfamily of HD Domain Family and HD Domain Is Responsible For Its Activity

The cyanamide hydratase domain belongs to HD domain superfamily due to the most conserved histidine and aspartate residues in the region. Though HD domain was reported as representing a superfamily of metal-dependent phosphohydrolases existed from bacteria to human (Aravind and Koonin, 1998), members that displaying other enzymatic activities were also discovered, such as oxygenase MIOX, and hydratase Cah. Experimental data showed that the mutations on the doublet HD residues of Ddi2/3 make the protein lose the enzymatic activity. The observation confirms that the conserved HD residues are critical for the enzymatic activity.

Neither Cah nor Ddi2/3 has been reported having any phosphohydrolase activity, but we cannot rule out the possibility. A general phosphatase substrate p-nitrophenyl

phosphate (p-NPP) was incubated with Ddi2/3 to test the possible phosphatase activity, while Calf-intestinal alkaline phosphatase (CIP) was used as positive control. But even after overnight incubation, Ddi2/3 incubation did not break down any p-NPP (data not shown), this indicates that the protein does not have any phosphatase activity.

3.4.3. *DDI2/3* Gene Is Unlikely Involved In DNA Repair

As its name indicated, the transcriptional level of *DDI2/3* is highly induced by a typical DNA-damaging agent MMS; thus, it was hypothesized that the gene is involved in DNA repair to MMS damage. Indeed, the disruption of *DDI2/3* makes yeast cells more sensitive to MMS. But when combining *ddi2Δ ddi3Δ* with inactivation of a representative DNA damage repair gene, *MAG1*, *RAD1* and *YAP1*, resulting triple mutants did not display enhanced sensitivities than the corresponding *mag1Δ, rad1Δ*, and *yap1Δ* single mutant cells to DNA damage caused by MMS and 4-NQO, indicating that *ddi2/3* is not synergistic with either of the three DNA repair defects. In other words, *DDI2/3* is not involved in either BER (represented by *MAG1*), NER (represented by *RAD1*) pathway or even a broad sense of DNA damage (represented by *YAP1*). Therefore, following studies on *DDI2/3* were focused on its cyanamide hydratase activity instead of DNA repair ability.

3.4.4. Proteins Containing a Cyanamide Hydratase Domain Are Evolutionarily Related

In the NCBI conserved domain database (Marchler-Bauer *et al*., 2015), to date only nine proteins including Cah and Ddi2/3 are identified containing cyanamide

hydratase domain. All these protein are from fungi, especially from *Ascomycota* phylum. Among them, only Cah has been demonstrated to have a cyanamide hydratase activity. A phylogenetic tree was constructed based on the sequences of these nine proteins. As shown in Figure 3–15, the homology among all nine proteins indicates that they are evolved from one ancestor. It is likely that all nine proteins have cyanamide hydratase activity. Since limited plants synthesis cyanamide as a chemical defense to soil fungi, it's reasonable for the fungi bearing a cyanamide degradation pathway to survive.

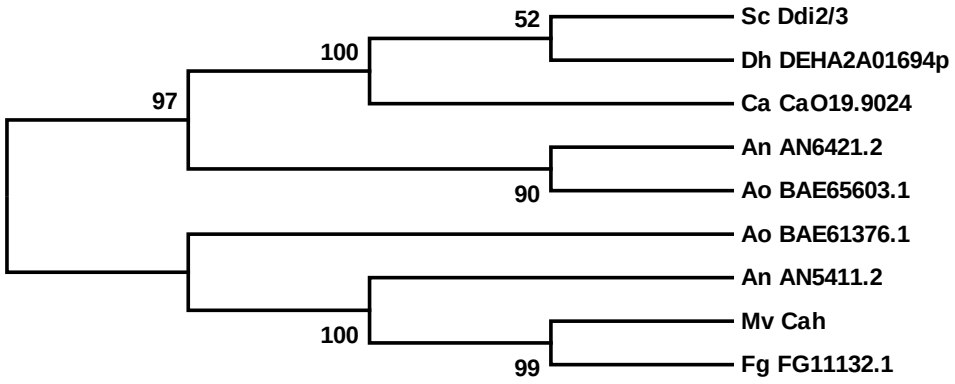


Figure 3–15. A phylogenetic tree illustrating the homology among available cyanamide hydratase domain containing proteins.

Amino acid sequences of proteins were retrieved from the NCBI conserved domain database. GenBank names, the source organism abbreviation are given in the left-hand column. Sc, *S.cerevisiae*; Mv, *M. verrucaria*; Fg, *F. graminearum*; An, *A. nidulans*; Ca, *C. albicans*; Ao, *A. oryzae*; Dh, *D. hansenii*. MEGA7 (Kumar *et al*., 2016) was used to predict the phylogenetic tree based on the maximum likelihood. The phylogeny was tested by using a bootstrap method, and the bootstrap replications were set as 1,000.

3.4.5. The Induction Pathway of *DDI2/3* Is Unique Among Known DNA Damage-Inducible Genes

In the β-gal assay with plasmid YEpDDI2-lacZ, both cyanamide and MMS strongly induced *DDI2/3* gene expression, indicating that the transcription of *DDI2/3* is tightly controlled. Indeed, its basal transcriptional level is barely detectable by northern hybridization. Since the basal level of LacZ in the β-gal assay is very low, any induction of transcriptional level of *DDI2/3* gene would lead to significant high induction fold. In other words, the *DDI2/3* promoter is sensitive to transcriptional induction. However, so far, only two chemicals, MMS, and cyanamide were reported can highly induce *DDI2/3* transcription, indicating that its transcription is under tightly regulated, and it is activated to certain inducers. This specific response is different from known DNA damaging inducible genes, whose transcriptions can be activated be a series chemical that causes the same kind of DNA damage. Therefore, the transcription regulation pathway of *DDI2/3* is unique.

So what is in common between MMS ($CH_3SO_2OCH_3$) and cyanamide (NH_2CN) that can activate *DDI2/3* transcription? In fact, the two chemicals do not have a similar structure or same functional group. To their toxicities of the two chemicals, MMS performs as a methylating agent mainly causes DNA damage. It also methylates proteins including histones affecting a broad range of cellular events. Whereas, cyanamide was reported as an inhibitor of aldehyde dehydrogenase and carbonic anhydrase, therefore, it affects ethanol metabolism. However, it has not been reported having DNA damaging effect, even if it has undiscovered DNA-damaging ability, it will not be a methylating agent as MMS because cyanamide does not have a methyl group at all. Also, MMS does

not appear to be a small molecular inhibitor of Ddi2/3 enzymatic activity (data not shown), indicating MMS is not a substrate of Ddi2/3. Therefore, the two has no common in function compared to cyanamide. Two distantly related chemicals can activate *DDI2/3* transcription; the regulation pathway would contain a sensor that can detect both chemicals to start transcription process.

The experimental data confirms that disruptions of *DDI2/3* genes did not compromise the activation of *DDI2/3* transcription by MMS and cyanamide, in other words, Ddi2/3 does not perform as a "sensor" in the transcription regulation pathway.

3.4.6. Other Identical Genes at Upstream and Downstream Of *DDI2/3* Are Involved In Vitamin B_{12} Synthesis

The *Saccharomyces* genome database reveals that the haploid laboratory strain contains up to 30% genomic duplications (Coissac *et al*., 1997). However, the chromosome VI and XIV regions where *DDI2* and *DDI3* reside are the only areas (except the rDNA cluster) in which the nucleotide sequences are highly conserved with three genes and their flanking regions nearly identical in sequence.

Though *SNO2* and *SNZ2* genes are also present in the duplicated region, however, their paralogs *SNO3* and *SNZ3* are not encoding identical proteins. Among these three pairs of genes, *SNO2/3* and *SNZ2/3* are found to be regulated by thiamine (Padilla *et al*., 1998; Rodríguez-Navarro *et al*., 2002), and their adjacent *THI5* and *THI2* genes are related to thiamine (vitamin B1) synthesis (Wightman and Meacock, 2003), suggesting that the duplicated regions contain a gene cluster associated with thiamine metabolism. Thiamine is a crucial physiological molecule since it serves as a cofactor (in the form of

thiamine diphosphate) for several enzymes involved primarily in carbohydrate catabolism. Thiamine is composed of pyrimidine and thiazole rings linked by a methylene bridge, while cyanamide can be utilized in pyrimidine synthesis (Hulme *et al*., 2008). Hence, it is plausible to speculate that cyanamide could be used by Ddi2/3 in pyrimidine rings and further in thiamine synthesis. Future studies are needed to reveal whether and how Ddi2/3 is involved in thiamine metabolism.

CHAPTER 4 CRYSTALLOGRAPHIC STUDY OF Ddi2/3

4.1. Abstract

Ddi2/3 encoded by *DDI2/3* genes in yeast *S. cerevisiae* was identified as a cyanamide hydratase that catalyzes cyanamide conversion to urea. Ddi2/3 belongs to HD domain superfamily due to the presence of conserved histidine and aspartate residues. Site-specific mutations with conserved HD residues result in the loss of enzymatic activity of Ddi2/3, indicating that the HD domain is responsible for the enzyme function. Cyanamide hydratase represents a special subclass of HD domain proteins that has not been well characterized. Crystals of recombinant Ddi2/3 were obtained and its structural model was constructed based on X-ray diffraction data at a resolution of 2.6 Å. The structural model shows that Ddi2/3 is mainly composed of helices. Native Ddi2/3 forms dimers and each monomer coordinates a zinc ion. Zinc is observed coordinating with conserved HD residues H55, H88, D89, and a water/hydroxyl group, forming tetrahedral geometry. Site-specific mutations of residues surrounding the metal, H137, Q138, T157 and N161, resulted in the loss of enzymatic activity. This observation indicates that theses residues are involved in the metal binding in catalysis. Substrate binding was observed by soaking crystals of Ddi2/3 mutants, Ddi2/3-H137N and Ddi2/3-T157V with cyanamide at 0.35M in cryo-buffer before performing X-ray diffraction.

4.2. Introduction

DDI2 and *DDI3* are duplicated genes in yeast *S.cerevisiae*; the proteins encoded by the genes were identified as having cyanamide hydratase activity that catalyzes cyanamide conversion to urea. Ddi2/3 is characterized containing a cyanamide hydratase domain, which belongs to the HD domain superfamily due to the presence of conserved histidine and aspartate residues. Site-specific mutations with conserved HD residues result in the loss of enzymatic activity of Ddi2/3, indicating that the HD domain is responsible for the enzyme function. Most identified HD domain proteins are characterized as being metal-dependent phosphohydrolases; hence cyanamide hydratase represents a special subgroup of HD domain proteins. Little was known about this family of proteins since its discovery in 1991 (Maiergreiner *et al*., 1991). Ddi2/3 from yeast is a second identified cyanamide hydratase, and proteins that were characterized containing cyanamide hydratase domain are all from fungi, in particular the *Ascomycota* phylum. Homologs of Ddi2/3 were also found in *Candida albicans*, *Aspergillus nidulans*, and *Aspergillus oryzae*. More interestingly, two genes of *DDI2/3* homologs were found in both *Aspergillus* species. Although proteins that are predicted to display cyanamide hydratase activity were found in bacteria, no biochemical studies were reported.

The hydration of cyanamide (H_2N-CN) to urea (H_2N-CO-NH_2) is similar the reactions carried out by nitrile hydratase (NHase), which catalyze the hydration of nitriles (R-CN) to the corresponding amides (R-CO-NH_3). Although the two categories of enzymes catalyze the same reactions, they have different distributions in organisms and are not evolutionally related. Most of the identified NHases are from bacteria, only a few have been discovered in eukaryotes. In addition, NHases and cyanamide hydratase are

reported as having very high specificities to their own substrates; there is no report indicating that NHases can hydrate cyanamides.

Cyanamide hydratase represents a special subclass of the HD domain protein and has not been extensively studied. Previously studies suggested that the HD domain is responsible for cyanamide hydratase activity of Ddi2/3. Here a crystallographic study of Ddi2/3 was conducted to reveal the structure of the protein and the working mechanism of the enzyme.

This study will help to characterize this class of HD domain proteins and enrich our knowledge of HD domain superfamily.

4.3. Results

4.3.1. Purification of Recombinant Ddi2/3 for Crystallization

Yunhua Jia from Dr.Moore's lab first harvested recombinant Ddi2/3 and successfully obtained crystals of Ddi2/3. Recombinant Ddi2/3 was expressed in the form of a GST fusion protein; after purification, the GST tag was cleaved by PreScission protease (GE healthcare) and removed by passing through a glutathione Sepharose column (GE healthcare). However, residual GST was still present after the affinity purification. To achieve high purity (greater than 95%) of Ddi2/3 for protein crystallization, a further purification was required to remove the remaining GST.

After removing the N-terminal GST tag, the recombinant Ddi2/3 protein has a theoretical molecular weight of 25.9 kDa with pI (isoelectric point) of 5.49, while the GST tag from vector pGEX-6P-1 after cleavage has a theoretical molecular weight of 26.4 kDa and its pI is 5.73. The MWs of two proteins are too close to be separated by

size-exclusive gel filtration. Ion exchange chromatography (IEC) is the only one practical method to separate the two. Samples were dialyzed in low salt buffer (20 mM Tris-HCl [pH 8.0], 20 mM NaCl, 4 mM DTT), and then were loaded on a RESOURCE Q column (GE Healthcare). The protein was eluted by using a gradient salt buffer containing up to 25% high salt buffer (20 mM Tris-HCl, [pH 8.0], 2 M NaCl, 4 mM DTT) (Figure 4–1A). Two peaks were observed during elution, after identification using SDS-PAGE, it showed that Ddi2/3 was firstly eluted, followed by elution of GST.

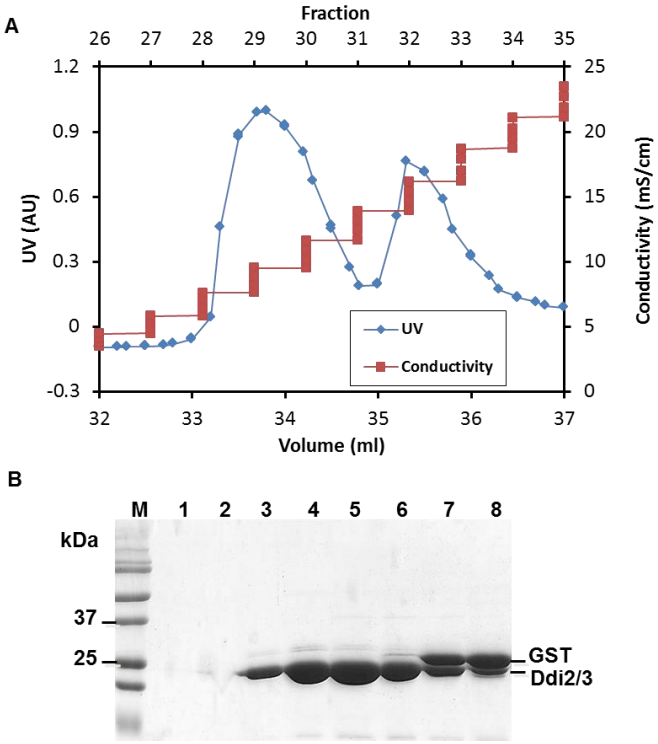


Figure 4–1. IEC elution profile showing purification of Ddi2/3

A. Elution profile of IEC; ***B***. SDS-PAGE gel image to show the separation of Ddi2/3 and GST in IEC. *Lane M*, Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1-lane 8*, elution fractions #26 - #33. Elution fractions that contained pure Ddi2/3 were collected and concentrated.

Elution fractions were evaluated on an SDS-PAGE for protein purity, as shown in Figure 4–1B. Fractions that contain pure Ddi2/3 were collected and concentrated to 10 mg/ml by using UltraSpin concentration tube with cut-off molecular weight (COMW) of 10 kDa (EMD Millipore). The concentrated protein was dialyzed in storage buffer (20 mM bis-Tris propane [pH 7.0], containing 50 mM NaCl and 0.01 mM $ZnCl_2$) and the purity was checked by SDS-PAGE.

4.3.2. Crystallization of recombinant Ddi2/3

Protein crystallization conditions were screened by using crystallization kit, including Wizard I, Wizard II (Rigaku) and $AmSO_4$ kit (QIAGEN), 0.8 µl of Ddi2/3-H88A; D89A (20 mg/ml) or Ddi2/3 (10 mg/ml) was mixed with equal volume of reservoir buffer and setting as hanging drops. The 24-well crystallization plate was sitting at room temperature for three days before reviewing under a microscope.

Crystals of the double mutant Ddi2/3-H88A; D89A were observed in 0.1 M Tris-HCl [pH 7.0], containing 1.0 M potassium/sodium (K/Na) tartrate and 0.2 M $LiSO_4$. The oval shaped crystals grew to the size of 0.3 × 0.8 × 0.3 mm after one-week incubation (Figure 4–2B). However, the crystals displayed poor diffraction properties; therefore no useful diffraction data were obtained.

Crystals of Ddi2/3 were not observed in any conditions from the kit. Yunhua Jia from Dr. Moore's lab has grown crystals of Ddi2/3 when using MES [2-(N-morpholino) ethane sulfonic acid] buffers as reservoir buffer with the presence of ammonia sulfate. Following her suggestion, a series of MES buffers in the range of pH 5.0 to pH 7.0, HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffers from pH 6.5 to pH

8.0 and Tris-HCl buffers from pH 6.5 to pH 8.0 were tested, with the addition of 0.9 M to 1.3 M ammonium sulfate and 0.2 M.

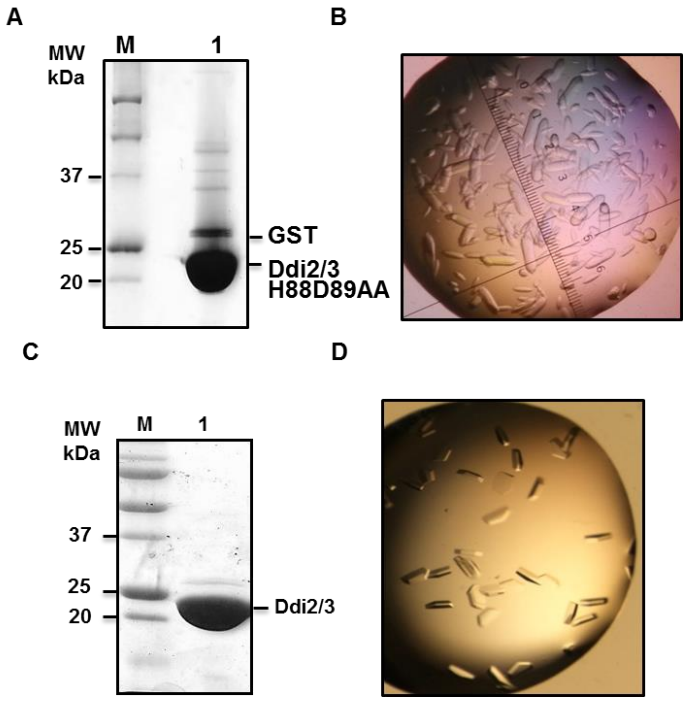


Figure 4–2. Crystallization of Ddi2/3 and Ddi2/3-H88A; D89A

A. SDS-PAGE gel image to show the purity of Ddi2/3-H88A; D89A used for crystallization. *Lane M,* Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1,* concentrated Ddi2/3-H88A; D89A for crystallization. ***B.*** Crystals of Ddi2/3-H88A; D89A, having a size of 0.3 × 0.8 × 0.3 mm. ***C.*** SDS-PAGE gel image showing the purity of wild type Ddi2/3 used for crystallization. *Lane M,* Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1,* concentrated Ddi2/3 for crystallization. ***D.*** Crystals of wild type Ddi2/3, having a size of 0.20 × 0.50 × 0.20 mm.

Ddi2/3 protein was adjusted to 10 mg/ml, and 0.8 µl protein was mixed with equal amount of reservoir buffer in setting crystallization using the hanging drop method. From crystallization trials, premium crystals of Ddi2/3 crystals were observed in 0.1 M MES buffer (pH 5.5), containing 1.2 M ammonium sulfate and 0.2 M arginine. With optimum conditions of crystallization, 2 µl protein solution was used in setting up hanging drops with 0.1 M MES buffers from pH 5.25 – pH 6.00, containing 1.1 M - 1.25 M ammonium sulfate, and 0.2 M arginine. Protein crystals of Ddi2/3 were growing at room temperature for a month, having a size of 0.20 × 0.50 × 0.20 mm (Figure 4–2D).

4.3.3. Ddi2/3 Is a Zinc-Metalloenzyme

Protein crystals of Ddi2/3 were brought to Canadian Light Source (CLS) for X-ray diffraction, by using beamline 08B1-1 and 08ID-1. The presence of zinc in Ddi2/3 crystals was determined by X-ray fluorescence scattering experiment, which was performed by Dr.Moore.

Diffraction data process and initial model building were carried out by Dr.Moore. The structure of Ddi2/3 was solved to 3.2-Å using single wavelength anomalous dispersion (SAD) method on Ddi2/3 crystals. Briefly, three datasets were collected using wavelength at zinc peak (λ = 1.2821 Å), each dataset consists of 500 frames (with the rotation of 0.2° for each frame). Parameters of collected dataset are shown in Table 4-1. Crystals belong to the space group *P*321, a=b=264.45 Å, c=119.15 Å. Each dataset was scaled and merged using HKL2000 program (Otwinowski and Minor, 1997), keeping anomalous pairs separate. Following density modification was processed using PHENIX, as the solvent content of Ddi2/3 crystals were 72%, extreme solvent flattering was carried

out in PHENIX (Terwilliger, 2004). Total 9 heavy atoms were located in the electron density map, indicating that the asymmetric unit contained 9 Ddi2/3 molecules, assigned as chain A to chain I. The resulted map was interpreted and the core 6 HD helices were built into 8 of the 9 subunits. Further work on model construction and refinement yielded a refined structure of 9 subunits. Following solvent refinement was carried out by Jia Li.

Table 4-1. Parameters of X-ray crystallographic data collection and model refinement of Ddi2/3

	Ddi2/3 with hydroxyurea
Data statistics[a]	
Space group	*P*321
a, b, c, (Å)	264.45, 264.45, 119.15
α, β, γ, (°)	90, 90, 120
Resolution (Å) (high)	50-2.6 (2.64-2.60)
Wavelength (Å)	1.2821
Temperature (K)	100
R_{sym} (%) [b]	10.4 (64.3)
I/σ (I)	18.8 (2.3)
No. of reflections	1,111,859 (146,237)
Redundancy	7.7 (7.7)
Completeness (%)	99.0 (97.9)
Refinement statistics	
Resolution (Å) (high)	39.7-2.60 (2.63-2.60)
Completeness (%)	98.7 (92.0)
Numbers of independent molecules	9
R_{work} [c] (%) (total/high)	18.6 (27.0)
R_{free} (%) (total/high)	23.2 (31.2)
Root mean square deviation, bonds/angles (Å/°)	0.015 /1.55
Average B factor	59.6
No. of protein atoms	31865
No. of water molecules	233
Residues in most favored region (%)	92.29
Residues in generously allowed regions (%)	6.04

[a] Each data set was obtained from a single crystal.

[b.] $R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I_i(hkl)\rangle| \ / \ \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and $\langle I_i(hkl)\rangle$ is the average intensity of reflection *hkl*.

[c.] $R = \sum||F_o| - |F_c|| \ / \ \sum|F_o|$

This dataset was collected with Ddi2/3 crystal soaked in 0.2 M hydroxyurea, however, after data processing, no significant electron density peak other than water molecule was observed at each of the zinc site. Therefore, the diffraction data was used as wild type Ddi2/3 refinement. The structure model was refined to an R_{work} of 18.6% and an R_{free} of 23.2%. In total, 98.33% of residues are in Ramachandran allowed region. The tertiary structure of Ddi2/3, a compact helical architecture, is illustrated in Figure 4–3. It composed of a total of 15 helices, including 11 α helices and four 3^{10} helices, and two short β-strands. The Zinc ion is coordinated by conserved HD resides, H55, H88, and D89, which are located at α3 and α4, respectively.

Table 4-2. B factors of structural model of Ddi2/3 after refinement

	Average B factor (backbone/total)	B factor of zinc
chain A	48.2 / 53.7	64.7
chain B	46.7 / 52.0	69.8
chain C	39.0 / 44.2	71.8
chain D	42.6 / 48.0	64.9
chain E	39.6 / 44.3	72.3
chain F	47.2 / 52.3	65.5
chain G	57.1 / 63.2	92.4
chain H	97.1 / 105.1	131.2
chain I	116.7 / 126.1	183.0

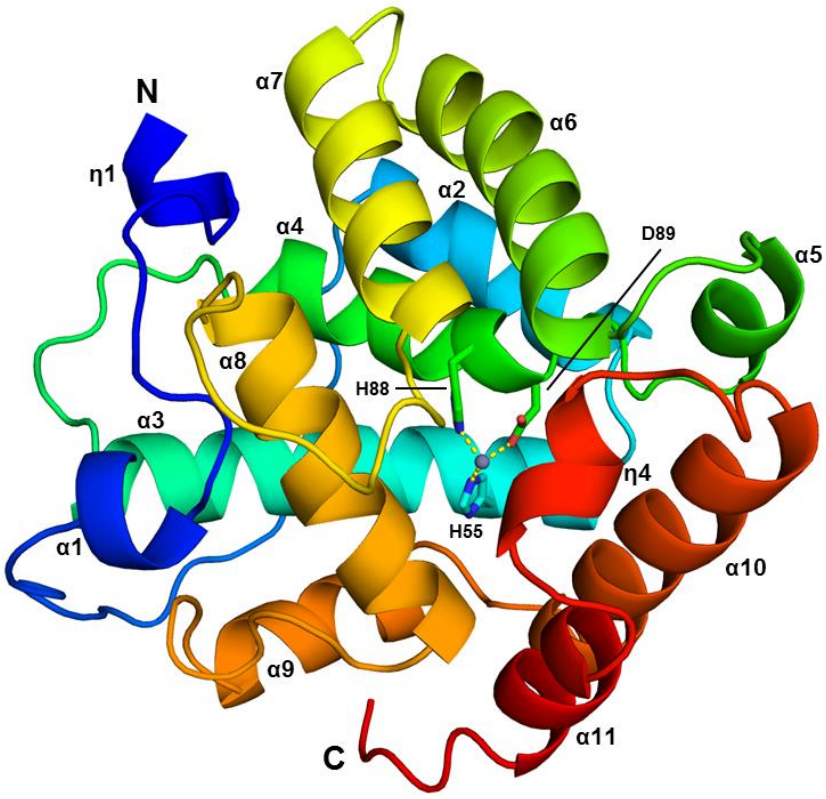


Figure 4–3. Ribbon trace of Ddi2/3 monomer (residue 1-225).

The first four residues at N-terminus (residue -3 to 0) after removal of GST tag was observed with electron densities, but the residues were removed in the model. Total 11 α-helix (α1-α11) and four 3^{10}-helix (η1–η4) were observed. Coordination of zinc ion with H55, H88, and D89 is represented as yellow dashes. Side chains of the residues are shown as sticks and colored by elements, oxygen; red; nitrogen: blue. The zinc ion is represented as a gray sphere.

Table 4-2. presents the B factors of each chain and the metal ion in the crystallographic structural model. Chain C, chain D and chain E have lower B factors than other chains. Therefore, they are displaying more convinced structural model. Zinc ion was observed in each of the 9 molecules of the asymmetry unit cell. However, the metals from chain G, chain H and chain I have higher B-factors as 92.4, 131.2 and 183.0. The refined model of the zinc coordination and the possible active sites are shown in Figure 4–4. The zinc ion is also coordinated by a water (hydroxyl) molecule with roughly equal distance (2.1-2.2 Å), forming a tetrahedral geometry with the other zinc ligands. Surrounding residues F104, H137, Q138, T157 and N161 form a pocket which is likely to be the active sites. Multiple water molecules were observed in this pocket (Figure 4–4), increasing the likelihood of being the active site of this hydratase. Figure 4–4A is presenting the refined model of the metal site in chain E: total four water molecules were observed, assigned as W1 - W4. W1 molecule is involved in metal binding; the rest water molecules form hydrogen bonds with H137, Q138, T157 and N161, respectively. The same architecture was observed in chain A, chain B, chain C and chain F.

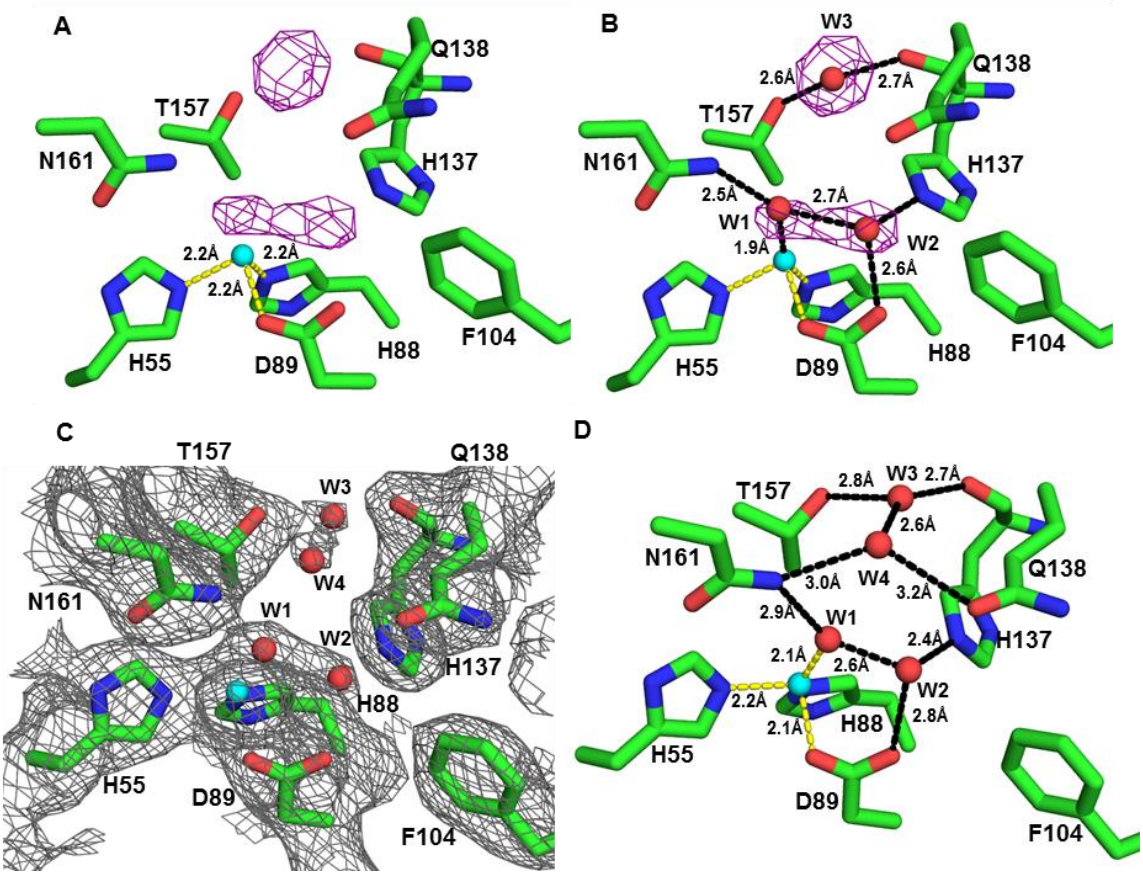
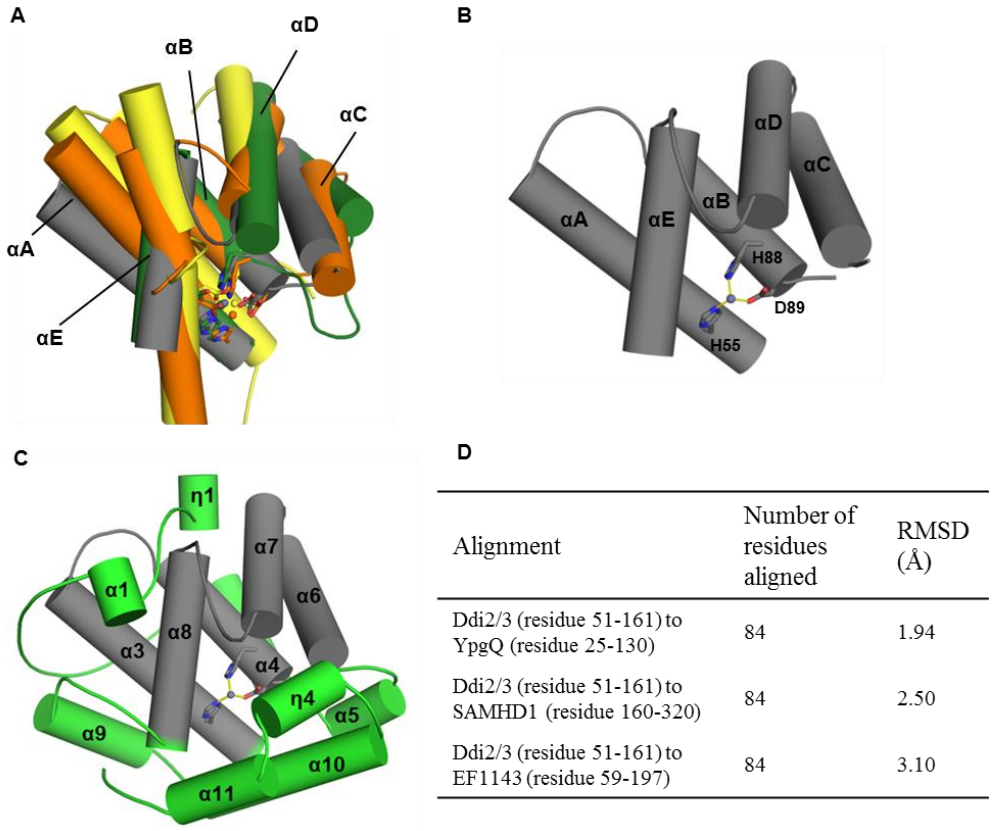


Figure 4–4. The omit map and the coordinates presenting the metal binding site and possible active site of Ddi2/3.

A. The Fo-Fc map (purple) showing the electron density near the metal site (chain E), the contour level was 3.50 for the Fo-Fc map. ***B.*** The Fo-Fc map (purple) showing the electron density near the metal site (chain E), modeling with solvent molecules (red). The contour level was 3.50 for the Fo-Fc map. ***C.*** The 2Fo-Fc map (gray) showing the electron density after refinement near the metal site (chain E), the contour level was 1.0 for the 2Fo-Fc map. ***D.*** The coordinates showing the structural model at the metal site (chain E). Side chains of residues are colored by element, oxygen; red; carbon: green; nitrogen: blue. Coordination bond to metal ion is represented as yellow dashes; hydrogen bonds are represented as black dashes. The zinc ion (cyan) and water molecules (red) are represented as spheres.

4.3.4. Ddi2/3 Bears Conserved Helical Structure of HD Domain

A total of nine Ddi2/3 molecules were present in asymmetric unit; the model of chain A was used in protein superposition to several typical HD domain proteins.



Alignment	Number of residues aligned	RMSD (Å)
Ddi2/3 (residue 51-161) to YpgQ (residue 25-130)	84	1.94
Ddi2/3 (residue 51-161) to SAMHD1 (residue 160-320)	84	2.50
Ddi2/3 (residue 51-161) to EF1143 (residue 59-197)	84	3.10

Figure 4–5. Superposition of HD motif of Ddi2/3 to core region of HD motifs of showing the conserved helical architecture

A. Superposition of HD motif of Ddi2/3(gray, residue 51-161 of chain A) to HD motives of YpgQ (green, PDB ID: 5DQV, residue 25-130), SAMHD1 (yellow, PDB ID: 3U1N, residue 160-320), and EF1143 (orange, PDB ID: 3IRH, residue 59-197), indicating that the Ddi2/3 overlaps the core region of HD motif, containing five α-helices, αA-αE. Metal cofactors of each protein were represented as spheres. ***B.*** Models of HD motif of Ddi2/3 (residue 51-161, chain A). ***C.*** Ribbon trace of Ddi2/3 showing the location of the HD motif (gray, residue 51-161). ***D.*** A table listing the parameters of structures superposition. Side chains of HD residues that interact with metal ions were colored by element, oxygen: red; nitrogen; blue.

Residues 51-161 of Ddi2/3 was identified as the HD motif, since the protein only coordinates one metal ion, the protein was overlaid to HD motifs of other HD domain proteins containing only one metal ion (Figure 4–5A). HD motifs used in superposition are from proteins YpgQ (Ni^{2+} coordinated protein, PDB ID: 5DQV), SAMHD1 (Zn^{2+} coordinated protein, PDB ID: 3UIN) and EF1143 (Ca^{2+} coordinated protein, PDB ID: 3IRH). The overlaid models (Figure 4–5A) showed that the structure of HD motif of Ddi2/3 overlaps the five helices (αA-αE) of the HD domain core region. Figure 4–5B is presenting the structural model of HD motif in Ddi2/3 (residue 51-161, chain A) consisting the conserved five helices. It is observed that the HD residues in the zinc coordination, H55 and H88D89 doublet are located at αA and αB, respectively. It is consistent with the distribution pattern of HD conserved residue in the other shown HD domain proteins. The observed structure similarity indicates that Ddi2/3 contains a typical HD motif.

Because YpgQ is a well-studied representative HD domain phosphohydrolase, structure model of Ddi2/3 monomer (shown in Figure 4–6A) was also used to superpose to YpgQ to identify any structure differences. YpgQ from *Bacillus subtilis* was characterized as a pyrophosphohydrolase that hydrolyzes (deoxy)ribonucleoside triphosphate (dNTP) to (deoxy)ribonucleoside monophosphate and pyrophosphate, and its activity requires the presence of Mn^{2+}, while the enzyme displayed compromised activity with the presence of Fe^{2+} Cu^{2+} or Ca^{2+} (Jeon *et al*., 2016). The HD domain is responsible for its pyrophosphohydrolase activity. Crystal structure model revealed that the protein contains a helical structure and assembles into a dimeric architecture. Each

monomer accommodates a Ni^{2+} ion, which was determined by the X-ray fluorescence scattering experiment (Jeon *et al*., 2014).

Ribbon trace model of the Ddi2/3 monomer is shown in Figure 4–6A, zinc ion is coordinated by H55, H88 and H89. H55 is located in α3, H88 and D89 doublet is located at α4. Figure 4–6B shows the model of YpgQ, which displays a helical structure. Total 10 helices are observed in the YpgQ model (PDB ID: 5DQV), up to 63% region of protein, 134 residues were composed of helices. Nickel ion is coordinated by H29′, H58′, and D124′. H29 is located at α2′, H58′ and D59′ doublet is located at α3′ (however, the aspartate residue is not interacting with the metal ion), D124′ is located at α6′. Figure 4–6C is presenting the superposed model of Ddi2/3 monomer (gray) and YpgQ monomer (violet). Overlaid model shows that the α3 and α4 of Ddi2/3, overlapping to α2′, and α3′ of YpgQ, respectively. Besides, α7 and α8 of Ddi2/3 overlap to α5′ and α6′. However, helices at C-terminal of Ddi2/3, α10, η4, and α11, forming a "lid", is more close to metal ion than α7′, α8′, and α9′ in YgpQ. The compact architecture of Ddi2/3 limits the opening of the active site thus only small molecules can fit in. Therefore YpgQ is able to work on relatively large molecules (dNTPs) whereas Ddi2/3 can only hydrate a smaller molecule (cyanamide). Figure 4–6D shows the superposition of metal binding sites of Ddi2/3 and YpgQ. H55, H88 and D89 of Ddi2/3 overlap to the H29′, H58′ and D59′ of YpgQ. However, the corresponding residue to the second aspartate residue in Ddi2/3 is threonine (T157), the side chain of threonine residue is shorter than aspartate residue, and contains a hydroxyl group instead of a carboxylate group; therefore it is unable to coordinate the metal ion.

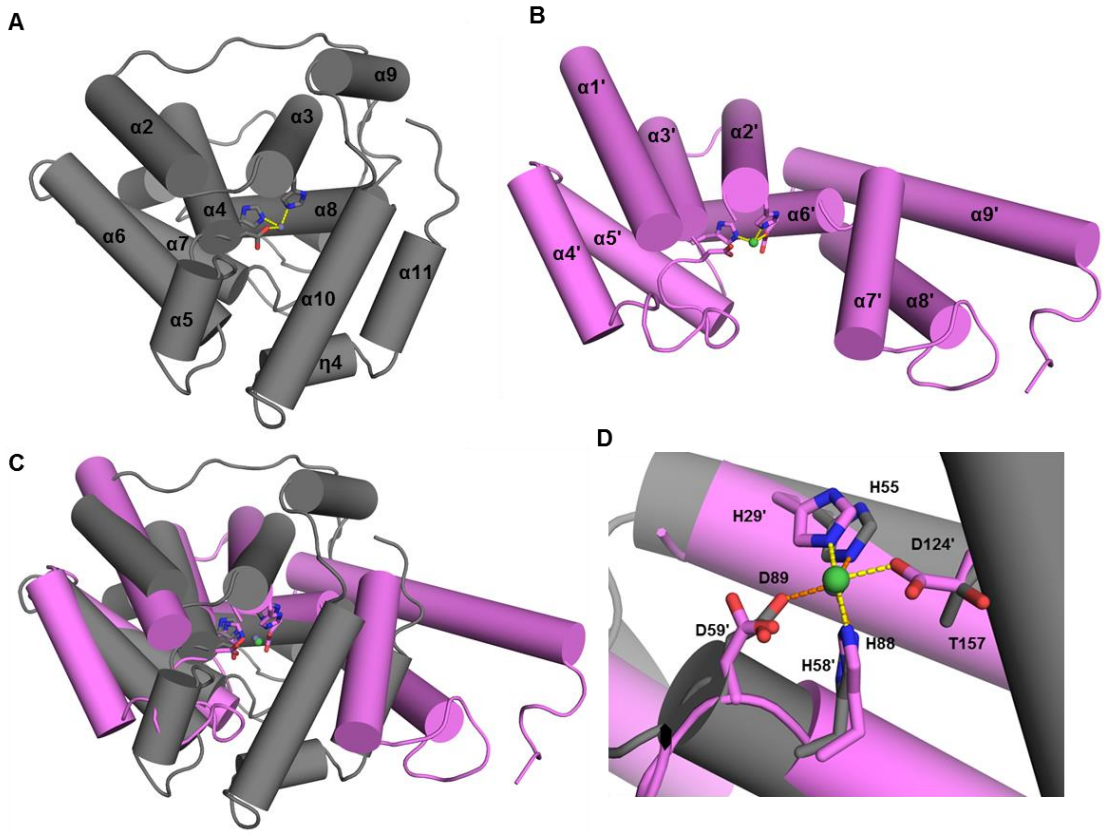


Figure 4–6. Superposition of a model of Ddi2/3 to the model of YpgQ

A. Ribbon trace of Ddi2/3 monomer (chain A). The coordination of zinc by H55, H88 and D89 are represented as yellow dashes. ***B.*** Ribbon trace of YpgQ (PDB ID: 5DQV, chain A) monomer. Nickel ion is represented as a green sphere, the coordination of nickel ion H29′, H58′, D59′ and D124′ are represented as yellow dashes. ***C.*** Superposition of Ddi2/3 (gray) to YpgQ (violet). 84 residues of the HD domains of each proteins were aligned with RMSD of 1.94 Å ***D.*** Superposition of the metal binding site of Ddi2/3 (gray) and YpgQ (violet). Nickel (green) and zinc (gray) are represented as spheres. The coordination of nickel in YpgQ is represented as yellow dashes, the coordination of zinc in Ddi2/3 is represented as orange dashes.. The side chains of residues are colored by element: nitrogen: blue; oxygen: red.

In addition, the structure model of Ddi2/3 was also superposed onto HD domain protein which contains multiple metal ions to illustrate the differences at the metal binding sites. Protein PgpH from *Listeria monocytogenes* was identified as a cyclic di-3',5'- adenosine monophosphate (c-di-AMP) phosphodiesterase. The c-di-AMP is an essential and ubiquitously presented second messenger in bacterial signaling. Biochemical and structural studies revealed that the HD motif of PgpH was responsible for hydrolyzing c-di-AMP to 5'-pApA. A crystal structure model of the HD motif of PgpH (residue 497-708, PDB ID: 4S1C) revealed that the HD motif contains a di-iron metal center which directly binds to c-di-cAMP molecules. After processing the anomalous diffraction data, the two metal ions in the active site are most likely to be Fe^{3+} (Huynh *et al*., 2015). Fe1 is interacting with H514′, H543′, D544′ and D648′, while Fe2 is coordinated by D544′, H580′, H604′ and H605′.

Figure 4–7 illustrates the superposition of Ddi2/3 to the HD motif of PgpH. The overlap of helices of HD motif is observed, indicating the two maintain the conserved tertiary structure. The overlaid model also illustrate that, two histidine residues and one aspartate residue coordinating Fe1 in PgpH (H514′, H543′, D544′) overlaps to HD residues coordinating zinc in Ddi2/3 (H55, H88, and D89). However, the second aspartate residue interacting to Fe1 (D648´) in PgpH has changed to threonine (T157) in Ddi2/3; the side chain is short and contains a hydroxyl group instead of a carboxylate group, therefore it is unable to coordinate the metal ion. to bind to the zinc ion in the protein. In addition, residues coordinating Fe2 in PgpH (D544′, H580′, H604′ and H605′) are corresponding to D89, F104, H137, and Q138, in Ddi2/3. The changes of histidine

residues to phenylalanine and glutamine residues in Ddi2/3 prevent protein recruiting a second metal cofactor.

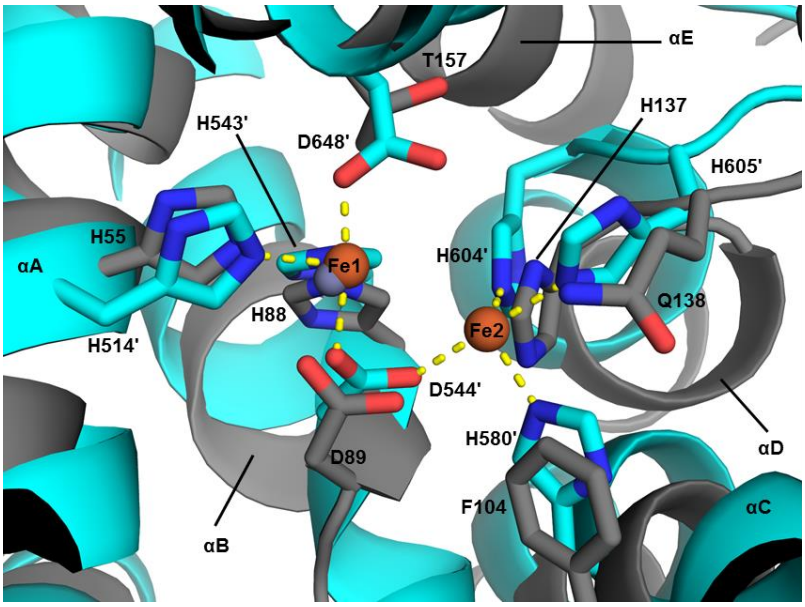


Figure 4–7. Superposition of Ddi2/3 to PgpH showing the metal binding sites.

Superposition of model of Ddi2/3 monomer (gray) and model of HD motif of PgpH (cyan, PDB ID: 4S1C, residue 507-662). In this superposition, 82 residues f HD domains of each proteins were aligned with RMSD of 1.82 Å The Zinc (gray) and di-iron molecules (orange) are represented as spheres. Metal coordination with residues in 4S1C model is represented as yellow dashes. The side chains of residues are colored by element: oxygen: red; nitrogen: blue.

4.3.5. Ddi2/3 Has Different Tertiary Structure to That of Typical Zinc-Metalloenzymes

Carbonic anhydrase II (CAII) is one of fourteen forms of human α carbonic anhydrases (CAs). Carbonic anhydrase (EC 4.2.1.1) catalyzes the reversible hydration of carbon dioxide. CAII is a well-characterized metalloenzyme, containing one zinc ion per polypeptide chain, whose main physiological function is to catalyze the interconversion between CO_2 and the bicarbonate ion (Briganti *et al*., 1999). As shown in Figure 4–8A,

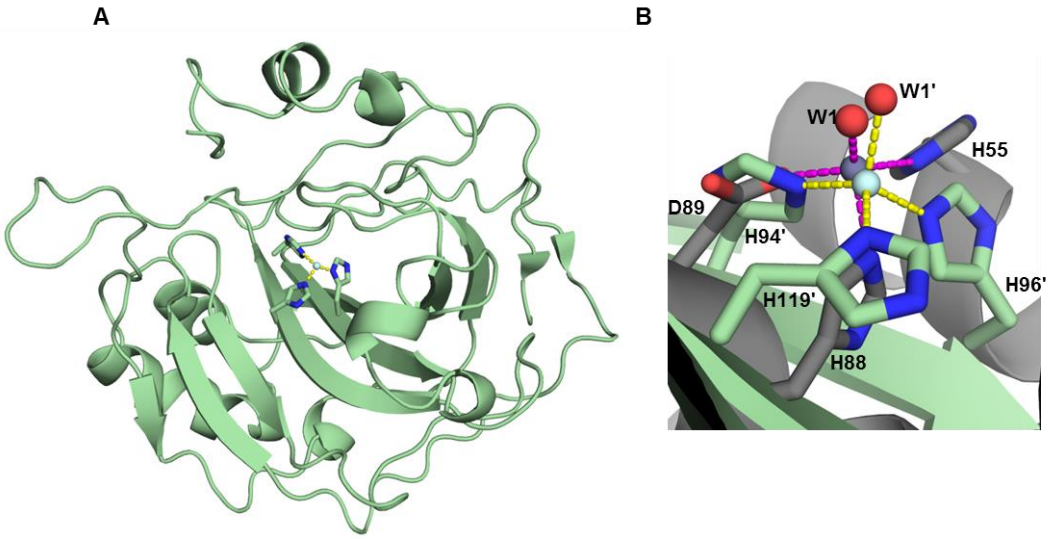


Figure 4–8. Superposition of model of Ddi2/3 to the model of human carbonic anhydrase II (CAII)

A. Ribbon trace of human CAII (PDB ID: 1CA2), showing the protein is mainly composed of β-sheet and loops. Zinc ion in 1CA2 model is represented as a palecyan sphere and its coordination by H94′, H96′ and H119′ are represented as yellow dashes. ***B.*** Superposition of model of Ddi2/3 (gray) and model of CAII (palecyan, PDB ID: 1CA2). Zinc ion in the Ddi2/3 model is represented as a gray sphere; its coordination is represented by magenta dashes. Zinc ion in CAII model is represented as a palecyan sphere, and its coordination is represented as yellow dashes. The side chains of residues are colored by element: oxygen: red; nitrogen, blue.

CAII (PDB ID: 4S1C) is mainly composed of loops and β-sheets. The zinc ion is coordinated by three histidine residue (H94′, H96′ and H119′) and a water molecule forming tetrahedral geometry. Zinc in Ddi2/3 has roughly same coordination geometry to that of CAII.

When Ddi2/3 was overlaid to the CAII, as shown in Figure 4–8B, it was found that conserved HD residues in Ddi2/3 did not overlap to the histidine residues in CAII. Although Ddi2/3 and CAII share same zinc coordination geometry, the secondary structure composition of CAII is entirely different from Ddi2/3. Although both enzymes catalyze the hydration reaction, and CAII even works on cyanamide, Ddi2/3 and CAII have totally differently tertiary structure, indicating the two do not share any homology.

4.3.6. Native Ddi2/3 Exists As A Dimer

Nine molecules of Ddi2/3 were designated chain A to I. Dimerization of Ddi2/3 molecules were observed, total 4 pairs of dimerization were formed among the 9 chains, chain B:chain C, chain D:chain E, chain F:chain G, chain H:chain I. Chain A was forming dimer with the chains from an another asymmetric unit. Structural models of chain B and C were shown in Figure 4–9. In the dimer of chain B: chain C, chain B is rotated 180° around Z axis to overlap chain C. Interfaces of dimers were composed of α7 and η4, which located between α10 and α11. Residues located at the interface and involved in dimerization are listed in Table 4-3. It is observed that the dimerization is maintained by hydrogen bonds between two molecules, including directly hydrogen bonds between two residues or via a water molecule.

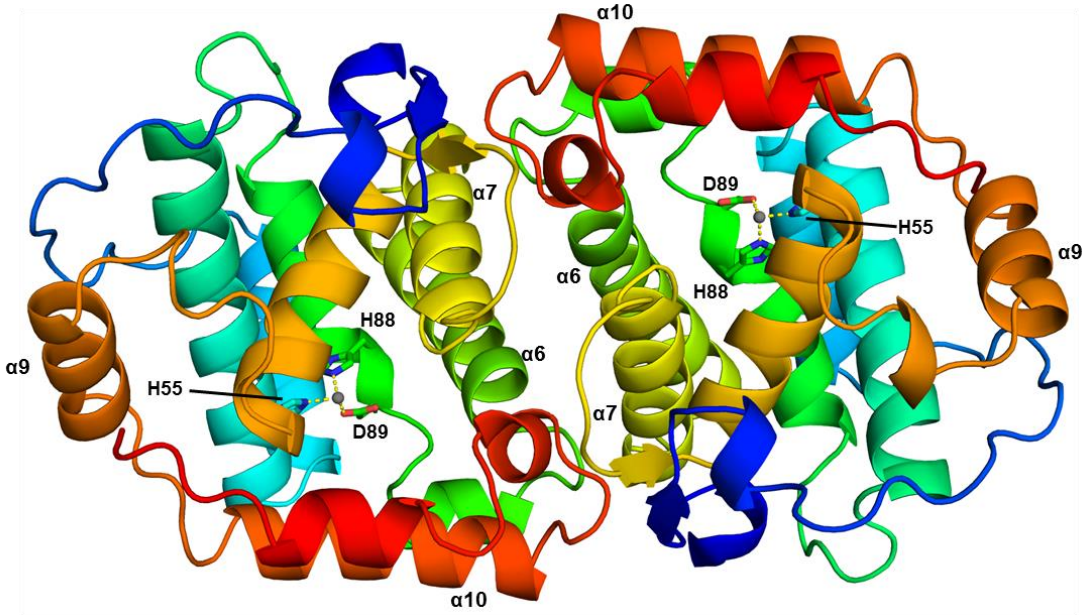


Figure 4–9. Ribbon trace of Ddi2/3 dimer.

The side chains of the residues are shown as sticks and colored by elements. The zinc ion is represented as a gray sphere. Coordination of zinc ion to H55, H88 and D89 are represented as yellow dashes.

Table 4-3. List of residues at interface of chain B: chain C that forms hydrogen bonds in dimerization

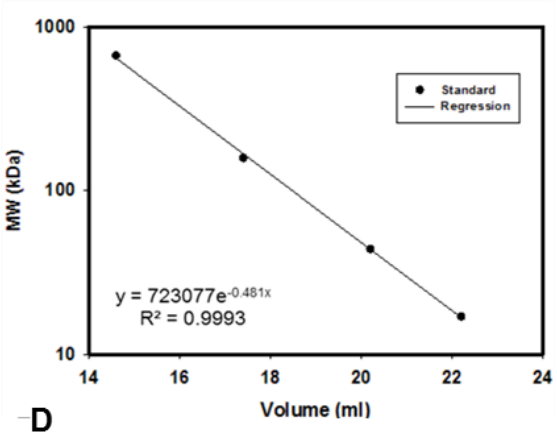
Residue in molecule B (location/atom)	Residue in molecule C (location/atom)	Interactions
Thr100 (loop between α5 and α6/OG1)	Asp128 (α7/OD2)	H bond
Lys101 (loop between α5 and α6/NZ)	Asp125 and Asp128 (α7/OD1)	H bond
Ser103 (loop between α5 and α6/OG)	Glu132 (α7/OE1, OE2)	H bond
Arg 113 (α6/NH2)	Thr 100 (loop between α5 and α6/OG1)	H bonds via HOH
Try145(β2, between α7 and α8/O)	Ser207 (η4, between α10 and α11/N)	H bond
Asp125 (α7/OD1)	Lys101 (loop between α5 and α6/NZ)	H bond
Asp128 (α7/OD1, OD2)	Thr100 (loop between α5 and α6/OG)	H bond via HOH
Glu132 (α7/OE1)	Ser103 (loop between α5 and α6/OG)	H bonds
Asp139 (loop between α7 and β2/OD2)	Trp106 (α6/OH)	H bonds via HOH
Trp203 (loop between α10 and η4/O)	Arg136 (α7/NH2)	H bonds via HOH
Ser207(η4, between α10 and α11/OG)	Tyr145 (β2, between α7 and α8/O)	H bond

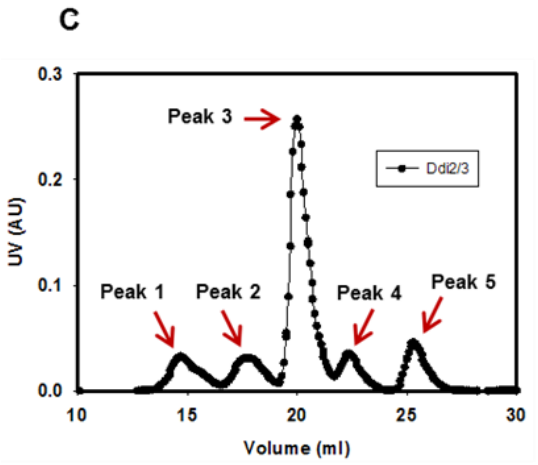
Size exclusion chromatography was used to confirm the oligomerization of Ddi2/3, 500 µg purified protein (the same protein used for crystallization) was applied to Superdex G-200 Increase (GE Healthcare) column to separate possible oligomers of protein. A gel filtration chromatography standard (Bio-Rad) was used as a reference to plot standard curve, as shown in Figure 4–10B. The elution profile of Ddi2/3 is presented in Figure 4–10C, in which the primary peak (peak 3) contains eluates having a calculated molecular weight of 48 kDa, consistent with the size of a Ddi2/3 dimer. This confirms

that native Ddi2/3 forms dimers in crystallization buffer (20 mM bis-Tris propane [pH 7.0], containing 150 mM NaCl, and 0.01 mM $ZnCl_2$).

A

Components in standard	MW (kDa)	Elution volume (ml)
Thyroglobulin (bovine)	670	14.6
γ-globulin (bovine)	158	17.4
Ovalbumin (chicken)	44	20.2
Myoglobin (horse)	17	22.2
Vitamin B_{12}	1.35	25.2





D

	Elution volume (ml)	MW (kDa)
peak 1	14.7	615.3
peak 2	17.7	145.4
peak 3	20	48.1
peak 4	22.3	15.9
peak 5	25.3	3.8

Figure 4–10. Size exclusion filtration chromatography showing the presence of a dimer of native recombinant Ddi2/3 protein

A. A table showing components of gel filtration chromatography standards (Bio-Rad). ***B***. A standard curve of gel filtration chromatography by using Superdex G-200 Increase column. ***C.*** Elution profile of Ddi2/3, total 5 peaks were observed. ***D.*** Calculated molecular weight for eluates in each peak.

4.3.7. Residues around Metal Ion Are Involved In Forming Active Sites

To investigate the substrate binding geometry in Ddi2/3, diffraction data from crystals soaked in a cryo-buffers containing cyanamide, cyanide, hydroxyurea (HU), and acetohydroxamic acid (AHA) were collected. Briefly, crystals of Ddi2/3 were shortly soaked with the substrate and/or inhibitors, including 0.1 M to 0.2 M cyanamide, 0.2 M hydroxyurea, 0.2 M urea, 0.1 M to 0.2 M sodium cyanide, and 0.2 M AHA, and X-ray diffraction data was collected. The diffraction data showed that the soakings did not affect diffraction properties of Ddi2/3 crystals. However, collected diffraction data did not reveal a peak at metal ion site in electron density map.

Structural studies on HD domain proteins show that the substrate is usually bound with residues close to the metal sites (Beloglazova *et al*., 2011; Brownt *et al*., 2006; Mashhadi *et al*., 2009; Van Staalduinen *et al*., 2014). In the structural model of Ddi2/3, H137, Q138, T157, and N161 formed a pocket in which Zinc ion is at center (Figure 4–4). In addition, these residues are conserved in all proteins that have been identified to contain the cyanamide hydratase domain (Figure 4–11), indicating that their conservation is necessary to maintain proteins function or structural stability.

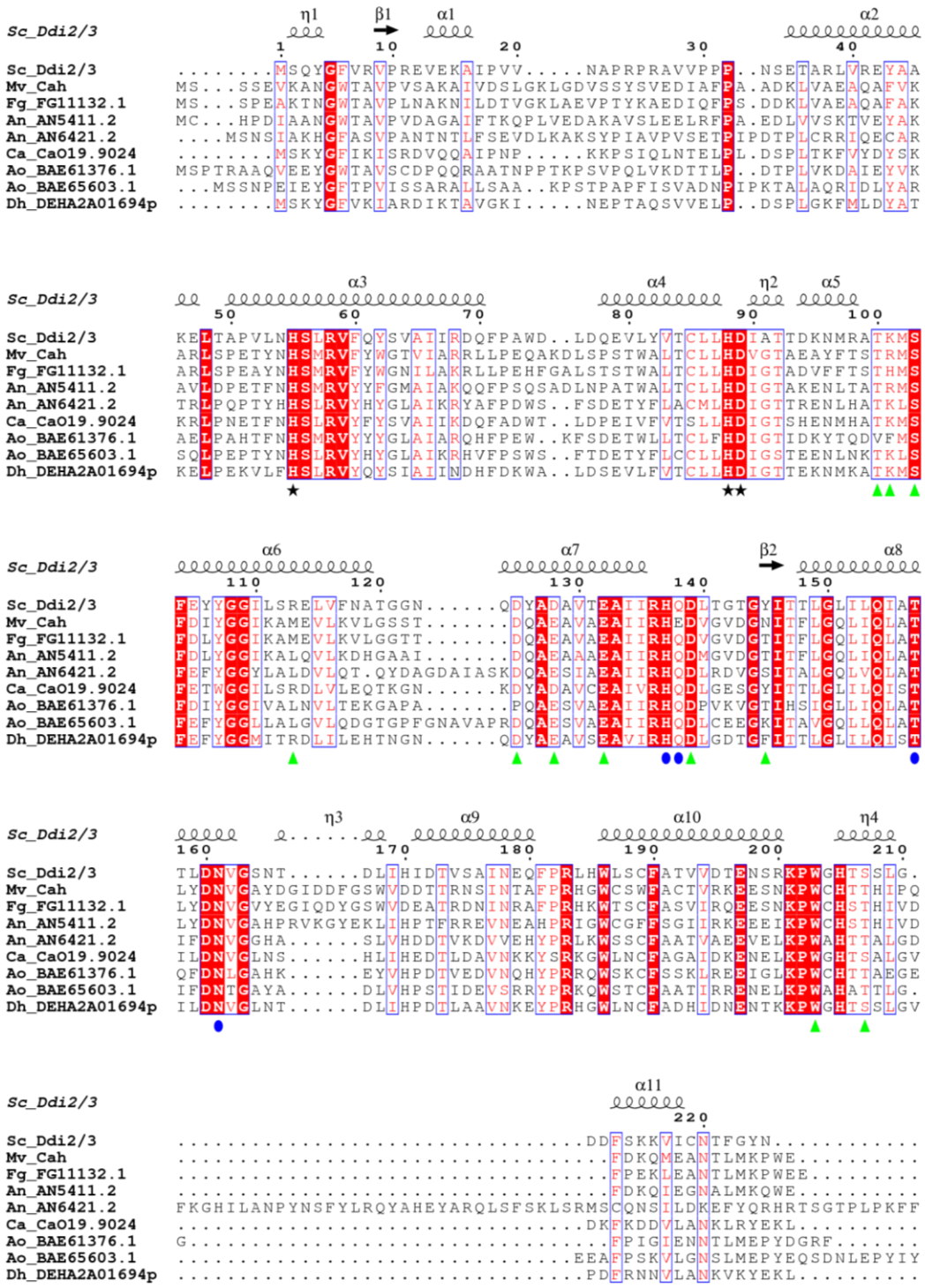


Figure 4–11. Sequence alignment of proteins containing cyanamide hydratase domain

Amino acid sequences of proteins were retrieved from the NCBI conserved domain database. Sequence alignment was processed by ClusterW, the secondary structure was illustrated by ESPript 3.0, based on crystal structure model. In secondary structure

illustration, α represents α-helix, β represents β-sheet, η represents 3^{10}-helix. HD residues coordinating zinc ion are pointed out by blue asters. The source organism abbreviations are given in the left-hand column. Sc, *S.cerevisiae*; Mv, *M. verrucaria*; Fg, *Fusarium graminearum*; An, *Aspergillus nidulans*; Ca, *Candida albicans*; Ao, *Aspergillus oryzae*; Dh, *Debaryomyces hansenii*. Identified residues are colored in red. Conserved HD residues are pointed out by stars, the Site-specific mutagenesis sites are indicated by green triangles. Residues involved in dimerization are pointed out by solid blue circles.

Taking vector pET-DDI2/3 as a template, total five site-specific mutations, H137N, Q138E, Q138A, T157V, and N161A, were made by using the QuickChange site-specific mutagenesis kit. Mutants were verified by DNA sequencing. Ddi2/3 mutants were harvested (Figure 4–12A) and tested in an enzymatic assay (Figure 4–12B). All proteins were adjusted to 0.1 mg/ml and incubated with 5 mM cyanamide for two hours

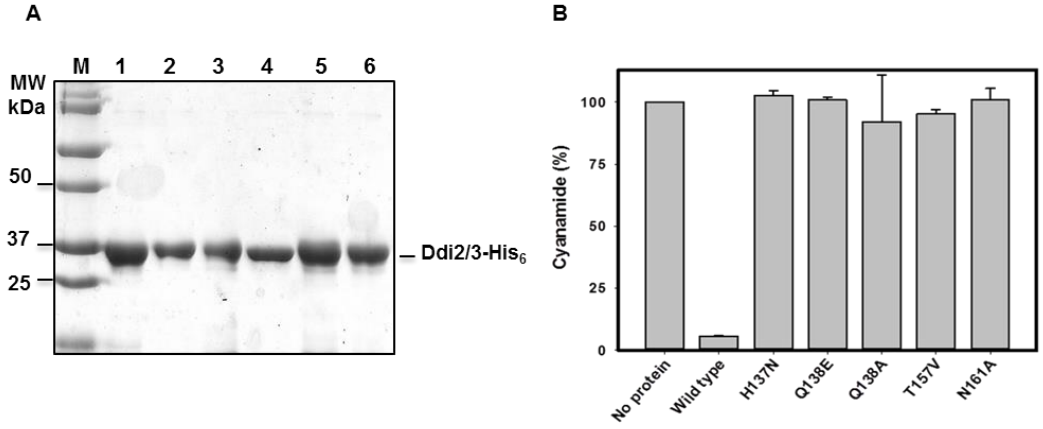


Figure 4–12. Preparation of Ddi2/3 mutants and the corresponding enzymatic assays

A. SDS-PAGE image to show purified His-tagged Ddi2/3 and its mutants. Lane M, Precision Plus Protein™ Unstained Standards (from Bio-Rad); *lane 1*, wild type Ddi2/3; *lane 2*, Ddi2/3-H137N; *lane 3*, Ddi2/3-Q138E*; lane 4*, Ddi2/3-Q138A; *lane5*, Ddi2/3-T157V; lane 6, Ddi2/3-N161A. ***B***. Cyanamide hydratase assays with His_6-tagged Ddi2/3 and its mutant derivatives.

at room temperature; remained cyanamide was measured by the colorimetric assay as previously described. The ability to consume cyanamide indicates the enzymatic activity of the mutant protein. As shown in Figure 4–12B, mutations of H137N, Q138E and N161A resulted in an almost entirely loss of the cyanamide hydratase activity of Ddi2/3, however, T17V mutant presented a little enzymatic activity. Q138A displayed varies in three independent assays, could remain a little activity. However, in initial velocity assays, all mutants failed to produce any detectable urea in 20 min (data not shown). This observation indicates that H137, Q138, T157, and N161 are all required for the catalysis.

4.3.8. Crystallization of Ddi2/3 Mutants

Substrate soaking with crystals of wild typeDdi2/3 did not give any information on the substrate binding mode; therefore, the inactive Ddi2/3 mutants were soaked with the substrate, and it was expected that the inactive Ddi2/3 mutants could not hydrate the substrate and thus would “trap” the substrate at the active site.

First, His_6-tagged Ddi2/3 mutants were prepared for crystallization due to its ease of purification. Purified His-tagged Ddi2/3 mutants, H137N and N161A, from affinity purification were further purified by doing ion-exchange chromatography (IEC) and dialyzed in storage buffer (20 mM bis-Tris propane [pH 7.0], containing 50 mM NaCl and 0.01 mM $ZnCl_2$). The proteins were concentrated to 10 mg/ml and then were subjected to crystallization trials with commercial sparse matrix crystallization kit (Hampton Research). Crystallization trials were performed by robots (Gryphon from Art Robot Instrument) in sitting drops method with commercial reservoir buffer (drop size of 0.3 ul + 0.3 ul) with a 96-3 shallow well Intelli-plate (HAMPTON research). The

crystallization trials were performed by Yunhua Jia from Dr.Moore's lab, she has tried a total of 384 conditions with His-tagged Ddi2/3, and its two mutants H137N and N161A. However, no crystals were observed with any of these conditions.

Therefore, Ddi2/3 mutants obtained from the pGEX system were used for

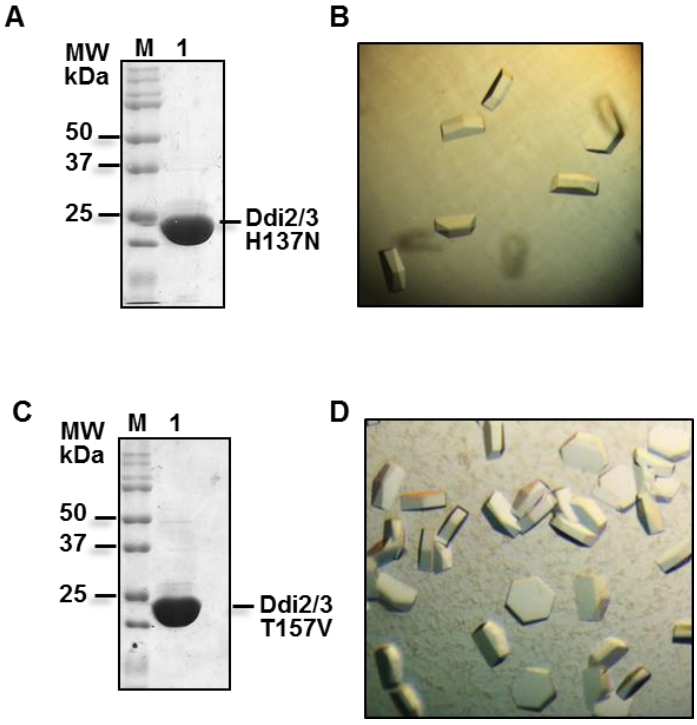


Figure 4–13. Crystals of Ddi2/3-H137N and Ddi2/3-T157V

A. SDS-PAGE gel image to show the purity of Ddi2/3-H137N used for crystallization. *Lane M,* Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1,* concentrated Ddi2/3-T157V for crystallization. ***B.*** Crystals of Ddi2/3-H137N, having a size of 0.3 × 0.5 × 0.2 mm. ***C.*** SDS-PAGE gel image to show the purity of wild type Ddi2/3 used for crystallization. *Lane M,* Precision Plus Protein™ Unstained Standards (Bio-Rad); *lane 1,* concentrated Ddi2/3 for crystallization. ***D.*** Crystals of wild type Ddi2/3-T157V, having a size of 0.5 × 0.5 × 0.2 mm.

crystallization. As previously described, GST fusion Ddi2/3 mutants were collected by doing affinity purification, after GST cleavage and removal; Ddi2/3 mutants were further purified by doing IEC, and were concentrated to 10 mg/ml for crystallization trials (Figure 4–13A and C). Conditions used in crystallization trials are the same as that of wild type Ddi2/3. Three Ddi2/3 mutants were used in crystallization trials, at last, Ddi2/3-H137N and Ddi2/3-T157V were successful to produce useful crystals (Figure 4–13B and D), but Ddi2/3-N161A was only able to grow tiny crystals.

Crystals of Ddi2/3-H137N were observed in 0.1 M MES (pH 6.0 – pH 6.5), containing 1.25 M - 1.40 M $(NH_4)_2SO_4$ and 0.2 M arginine. Obtained crystals were able to grow to the size of. 0.3 × 0.5 × 0.2 mm. Crystals of Ddi2/3-T157V were observed in 0.1 M MES (pH 5.5 – pH 6.2), containing 1.25 M - 1.40 M $(NH_4)_2SO_4$ and 0.2 M arginine. Obtained crystals were able to grow to the size of. 0.5 × 0.5 × 0.2 mm.

4.3.9. Substrate Binding Model of Cyanamide to Ddi2/3

Both crystals of Ddi2/3-H137N and Ddi2/3-T157V were shortly soaked in cryobuffer containing 0.2 M -0.35 M cyanamide and 0.2 M -0.35 M cyanide, respectively, before being applied to X-ray diffraction, by using beamline 08B1-1 or 08ID-1 at the CLS for X-ray diffraction.

The structure model of Ddi2/3-H137N at 3.00-Å resolution was solved, and structure model of Ddi2/3-T157V at a 2.90-Å resolution of was solved. However, resolved model of a complex of Ddi2/3 mutants and cyanide did not show significant results. Structure model of the complex of Ddi2/3-T157V with cyanamide was obtained and refined to 2.90-Å resolution. Table 4-4 is presenting the parameters of collected X-

ray diffraction dataset and of model refinement.

Table 4-4. Parameters of X-ray crystallographic data collection and refinement of Ddi2/3 mutants.

Complex	Ddi2/3-H137N	Ddi2/3-T157V	Ddi2/3-T157V with cyanamide
Data statistics[a]			
Space group	*P*321	*P*321	*P*321
a, b, c, (Å)	264.64, 264.64, 119.30	263.61, 263.61, 119.03	263.43, 263.43, 119.21
α, β, γ, (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å) (high)	50-3.00 (3.05-3.00)	50-2.90 (2.95-2.90)	50-2.90 (2.95-2.90)
Wavelength (Å)	1.0246	1.0246	1.0246
Temperature (K)	100	100	100
R_{sym} (%)[b]	9.2 (56.8)	10.5 (69.5)	11.4 (63.0)
I/σ (I)	20.9 (2.7)	17.6 (2.9)	16.1 (2.2)
No. of reflections (total/unique)	614,102 / 95,972	538,150 / 105,200	671,645 / 104,906
Redundancy	6.6 (6.6)	5.3 (5.2)	6.6 (6.5)
Completeness (%) [c] (total/high)[b]	96.5 (97.8)	97.0 (98.8)	96.8 (98.4)
Refinement statistics			
Resolution (Å)	39.77-3.00 (3.07-3.00)	39.32-2.90 (2.96-2.90)	29.97-2.90 (2.96-2.90)
Completeness (%)	96.5 (97.0)	97.0 (99.0)	96.7 (97.0)
Numbers of independent molecules	9	9	9
R_{work} (%) [c]	18.7 (27.1)	18.7 (23.2)	18.3 (24.7)
R_{free}(%)	23.4 (34.5)	22.8 (27.5)	22.8 (30.2)
Root mean square deviation, bonds/angles (Å/°)	0.016/1.46	0.015/1.38	0.016/1.38
Average B factor	77.5	52.7	65.8
No. of protein atoms	31,761	31,725	31,725
No. of water molecules	325	408	490
Residues in most favored region (%)	92.73	93.72	93.62
Residues in generously allowed regions (%)	5.79	5.11	5.30

[a] Each data set was obtained from a single crystal.

[b.] $R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I_i(hkl)\rangle| \; / \; \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation of reflection *hkl* and $\langle I_i(hkl)\rangle$ is the average intensity of reflection *hkl*.

[c.] $R = \sum\big||F_{obs}| - |F_{calc}|\big| \; / \; \sum|F_{obs}|$

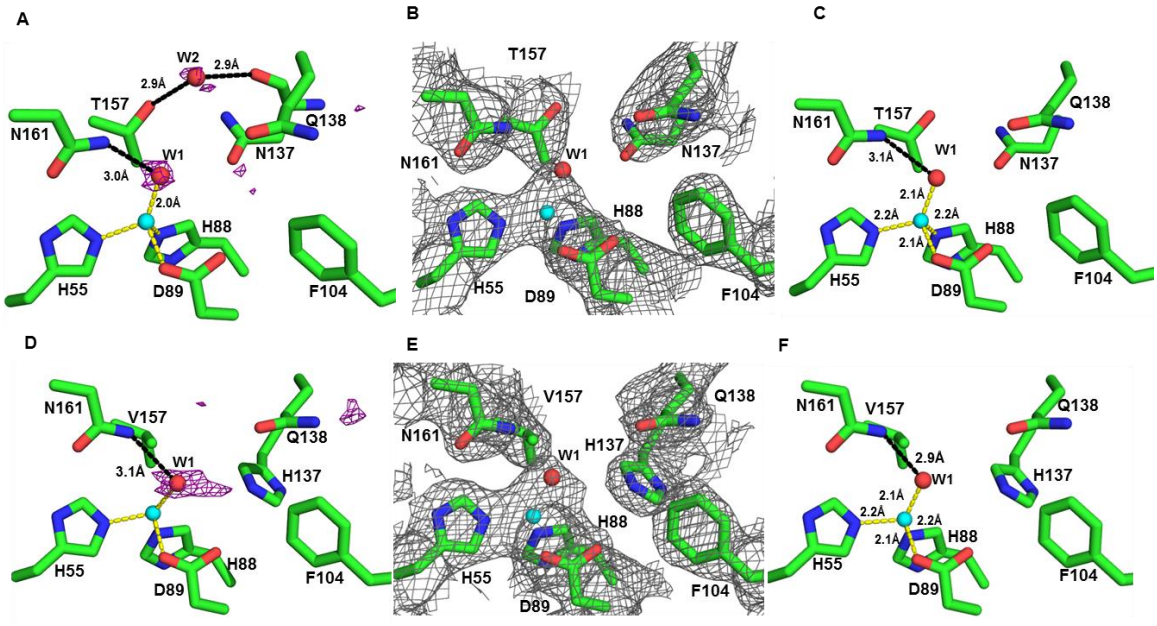


Figure 4–14. The 2Fo-Fc map and the coordinates presenting the metal binding site of Ddi2/3-H137N and Ddi2/3-T157V mutants.

A. The Fo-Fc map (purple) showing the electron density near the metal site in the structure model of Ddi2/3-H137N (chain D), modeling with solvent molecules (red). The contour level was 3.50 for the Fo-Fc map. ***B.*** The 2Fo-Fc map (gray) showing the electron density near the metal site in the structure model of Ddi2/3-H137N (chain D), the contour level was 1.0 for the 2Fo-Fc map. ***C.*** The coordinates showing the zinc coordination and water molecules at the metal site of Ddi2/3-H137N (chain D). ***D.*** The Fo-Fc map (purple) showing the electron density near the metal site in the structure model of Ddi2/3-T157V (chain D), modeling with solvent molecules (red). The contour level was 3.50 for the Fo-Fc map. ***E.*** The 2Fo-Fc map (gray) showing the electron density near the metal site in the structure model of Ddi2/3-T157V (chain D), the contour level was 1.0 for the 2Fo-Fc map. ***F.*** The coordinates are showing the zinc coordination and water molecules at the metal site of Ddi2/3-T157V (chain D). The side chains of residues are colored by element, oxygen; red; carbon: green; nitrogen: blue. Coordination bond to metal ion was represented as yellow dashes; hydrogen bonds are represented as black dashes. The Zinc (cyan) and water molecules (red) are represented as spheres.

The refined model of zinc coordination and the possible active sites in Ddi2/3-H137N and T157V mutants (of chain D) are shown in Figure 4–14. It was observed that the mode of zinc metal coordination remains in both mutants. These diffraction data were obtained in March, 2016; initial refinement has located most of the solvent in 7 of the 9 subunits. Further refinement will be carried out to confirm these solvent sites.

The complex of Ddi2/3-T157V and cyanamide revealed a piece of useful information on the cyanamide binding mode. Figure 4–15 shows the refinement of the model of a complex of Ddi2/3-T157V and cyanamide (chain A). Panel A displays the Fo-Fc map showing extra density at metal center indicating the presence of atoms, indicating that a molecule binding at zinc was pointing towards H137. Panel B shows the 2Fo-Fc map after refinement with added solvent, displaying the fittings of water molecules. However, there was excessive electron density at the zinc suggesting the presence of a larger molecule. Panel C shows the refined 2Fo-Fc map after fitting a cyanamide molecule at zinc. It was observed that the cyanamide molecule fits better to the electron density than a water molecule.

After refinement, the most representative model (chain A) showed that the stick-like cyanamide molecule directly binds to zinc ion via N2 (from NC- group) atom. The Zn-N2 bond has a distance of 2.0 Å, roughly equal to that of interactions by HD residues. N2 atom also formed a hydrogen bond with N161, having a length of 3.0 Å. While the ammonia group formed hydrogen bonds with H137 and Q138 at roughly equal lengths of 2.7-2.8 Å, it also formed a weak hydrogen bond with D89 at distance of 3.3 Å (Figure 4–15). This observation also confirms that the pocket containing zinc site is the active site

of the enzyme. The electron density of cyanamide molecule was observed in 7 of the 9 molecules.

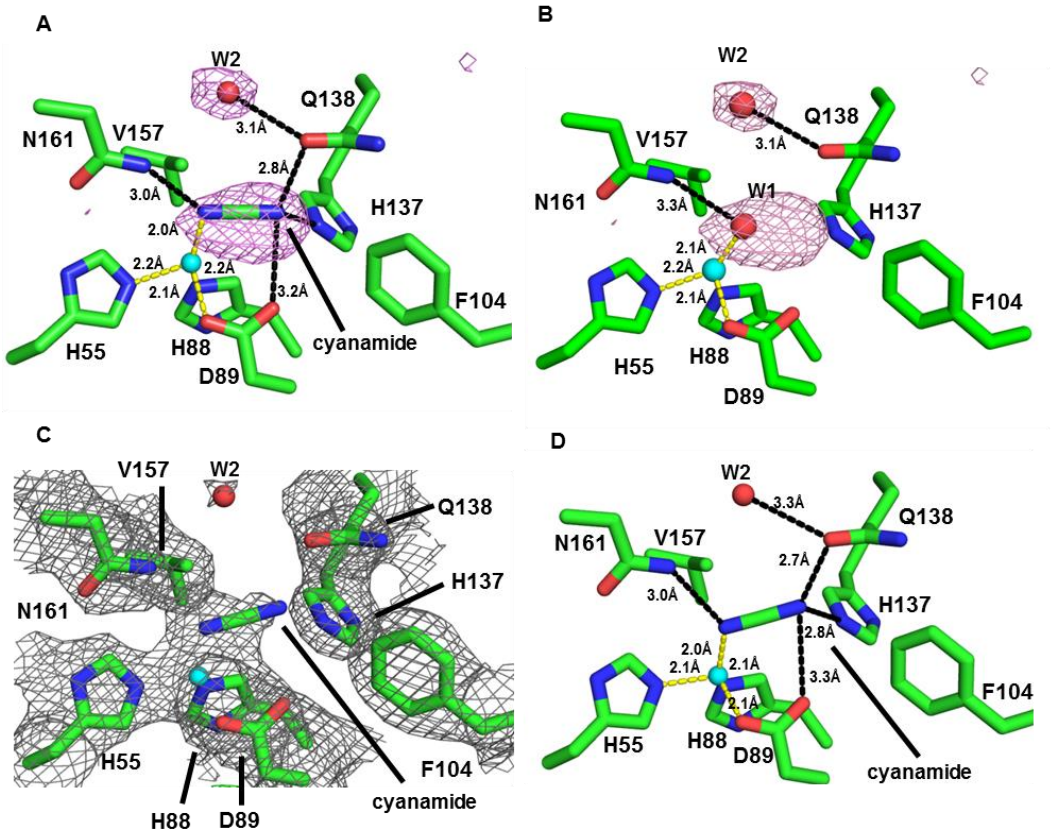


Figure 4–15. Model refinement of Ddi2/3-T157V soaking with 0.35 M cyanamide

A. The Fo-Fc map of electron density near zinc site, modeling with the cyanamide molecule (chain A). The contour level was 3.50 for the Fo-Fc map (purple). ***B.*** The Fo-Fc map of electron density near the zinc site, modeling with the water molecule. The contour level was 3.50 for the Fo-Fc map (purple). ***C.*** The 2Fo-Fc map is showing the electron density after refinement with cyanamide molecule at the zinc site (chain A). The contour level was 1.00 for the 2Fo-Fc map (gray). ***D.*** The corresponding coordinate is showing the binding mode of cyanamide. Zinc ion was represented as a cyan sphere, water molecules were represented as red spheres. The side chains of residues were colored by element: oxygen: red; carbon: green; nitrogen: blue. Coordination of the zinc was represented as yellow dashes; hydrogen bonds were represented as black dashes. The side chains of residues were colored by elements: oxygen: red; carbon: green; nitrogen: blue.

After refinement of available chains in the asymmetric unit (chain A - chain G. the electron densities of chain H and chain I are too poor to refine), it is observed that, the binding mode of cyanamide is relatively stable, though the molecule was not fully covered in electron density map. Figure 4–16 is presenting the models of cyanamide binding at the active sites of Ddi2/3.

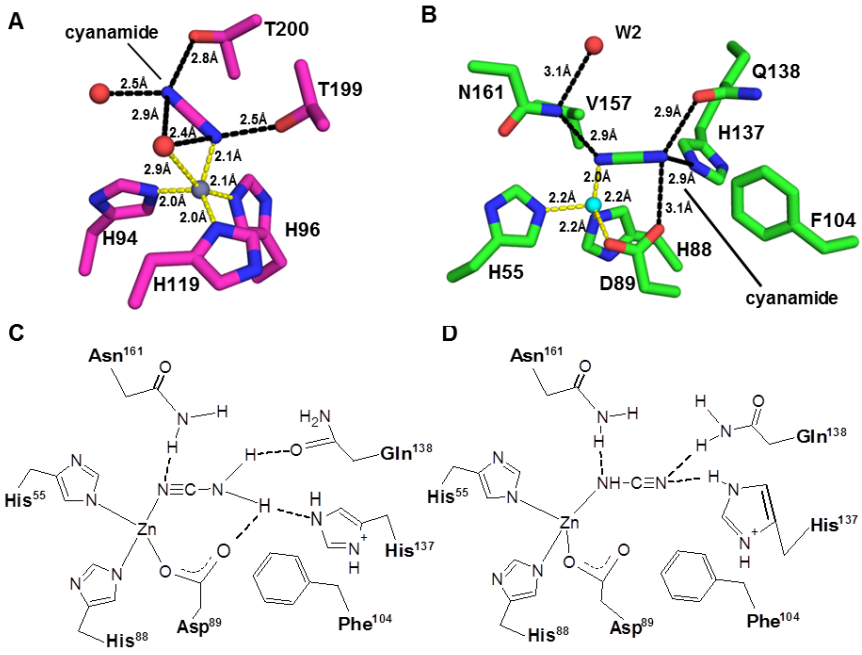


Figure 4–16. Models of cyanamide binding at the active site of Ddi2/3-T157V mutant and in CAII.

A. Cyanamide binding model in CAII (magenta, PDB ID: 1F2W). Zinc ion (gray) and water molecules (red) are represented as spheres. ***B.*** Cyanamide binding model in Ddi2/3-T157V mutant (chain A). Zinc ion (cyan) and water molecules (red) represented as spheres. Metal coordination is shown as yellow dashes; hydrogen bond is represented as black dashes. The side chains of residues are colored by elements: oxygen, red, nitrogen, blue. ***C.*** and ***D.***, the models of cyanamide binding to the metal ion via either cyanide group or ammonia group. The hydrogen bond is represented as dashes.

Figure 4–16A displays the model of cyanamide binding to the zinc ion in CAII (PDB ID: 1F2W). The cyanamide molecule binds to the zinc via N2 of the cyanide group, and the whole molecule forms hydrogen bonds with T199, T200 and other two water molecules. Though cyanamide is a molecule of linear structure, the structural model (PDB ID: 1F2W) has a relatively high resolution of 1.90 Å (Guerri *et al*., 2000) and thus it is able to distinguish the orientation of the cyanamide molecule binding to the zinc ion in CAII. Therefore, in the modeling of cyanamide binding to the zinc ion in Ddi2/3-T157V, cyanamide was proposed binding to the zinc via the N2 of the cyanide group as well (Figure 4–16B and C). In this model, N2 of the cyanide group shares the lone pair with the zinc ion, but the covalent bond does not line up to the molecule of cyanamide, therefore the nitrogen atom is still possible to serve as a hydrogen bond acceptor. Thus the nitrogen of the cyanide group also forms a hydrogen bond with N161. The ammonia group in the other side of the stick-like molecule performs as hydrogen bond donor in the formations of hydrogen bonds with Q138 and D89. H137 carries positive charge since the crystallization solution is of pH 6.0, which is lower than the pKa of the histidine (pKa=6.5). The nitrogen atom of ammonia group has a lone pair and thus serves as hydrogen bond acceptor in the formation of a hydrogen bond with H137.

However, the current structural model of Ddi2/3-T157V and cyanamide has a relatively low resolution of 2.90 Å, we are not able to determine the orientation of cyanamide binding to the zinc ion in this condition. Therefore, the possibility of cyanamide binding to the zinc ion via the ammonia group can not be ruled out. In the model shown in Figure 4–16D, the ammonia group interacts with the zinc ion. The nitrogen atom of the ammonia group still has a lone pair and thus serves as the hydrogen

bond acceptor in the formation of hydrogen bond with N161. While the nitrogen atom of the cyanide group also has a lone pair but can perform as a hydrogen bond acceptor, thus it forms hydrogen bonds with both imidazole ring of H137 and the ammonia group of Q138, since the side chain of Q138 can be flipped by180°. But the nitrogen atom of the cyanide group would not form a hydrogen bond with D89.

Both binding models of cyanamide in Ddi2/3-T157 indicate that carbon would carry more positive charges and would be the target of a nucleophile to initial reaction. And at least one water molecule (W2 in Figure 4–15D) was present at active site was present in each Ddi2/3 monomer, forming a hydrogen bond to N161. This water could be recruited to produce attacking hydroxide in hydration reaction; therefore its position in the active site is crucial to reveal the catalytic mechanism, further refinement of cyanamide binding model is needed to identify the proper position of a water molecule. The accurate location of cyanamide at the active site also needs further model refinement to determine the interactions with the residues at active sites. All these information are required to discover the reaction mechanism of Ddi2/3.

4.4. Summary and Discussion

4.4.1. Ddi2/3 Contains a Typical HD Motif

Superposition of HD motif of Ddi2/3 to other identified HD domain proteins revealed that Ddi2/3 displays the conserved helical structure of HD domain composed of 5 α-helices. Metal-coordinating residues in HD domain proteins that contain single metal ion can be presented as H…HD…D, and they are located at αA, αB, and αE. In Ddi2/3, HD residues coordinating the zinc ion are shown as H…HD, and their location at αA and

αB overlap to the HD residues from other identified HD domain. Thus the structural similarity between Ddi2/3 and identified HD domain protein reveals that Ddi2/3 contains a typical HD motif in the region.

4.4.2. Ddi2/3 Has Unique Metal Coordination Geometry As Compared To Other HD Domain Proteins

Although Ddi2/3 contains a typical HD motif, it is unique to other HD domain proteins since only three HD residues (H...HD) are holding the zinc ion, instead of H…HD…D pattern in other identified HD domain proteins. The second aspartate residue has been mutated to threonine (T157) in Ddi2/3; the threonine residue has a short side chain and has not been reported to have the metal coordinating ability. Therefore, the mutation of aspartate to threonine in Ddi2/3 makes the zinc ion have unique coordination geometry to other HD domain proteins. HD domain protein SAMHD1 (PDB ID: 3U1N) is also a zinc metalloenzyme, in which the zinc ion is coordinated by the side chains of H167, H206, D207, D311 and a phosphate oxygen, forming a 5-ligand coordination. While in Ddi2/3, the zinc ion is coordinated by H55, H88, D89 and water (hydroxyl) forming a four ligand tetrahedral geometry.

4.4.3. The Mode of Zinc Coordination Reveals That Ddi2/3 Is a Typical Zinc Binding Protein

Zinc is the second most abundant transition element in living organisms after iron. About 10% of total human proteins are predicted as having the ability to binding the zinc *in vivo.* Well-studied examples of the zinc binding proteins include the carbonic

anhydrase I and II (CAs), alcohol dehydrogenase and thermolysin. Statistical studies with all zinc-binding proteins in the protein data bank (PDB) have shown that zinc proteins traditionally adopt a coordination number of 4 or 5 in their metal coordination sphere. Histidine residues are the most common amino acid ligands, existing in 48% of the ligands to zinc, followed by cysteine, which presents in 27% of the ligands to zinc. Aspartate and glutamate are the other main residues interacting with the zinc. These four residues are responsible for 96% of zinc coordination (Tamames *et al*., 2007).

Our structural study on Ddi2/3 reveals that the zinc ion is coordinated by three amino acid residues, His55, His88, Asp89, and a water (hydroxyl) molecule; total four ligands have roughly same length and forming an almost regular tetrahedron, which is typical coordination geometry of the zinc proteins. It indicates that Ddi2/3 is a typical zinc binding protein, and the opening site (hydroxyl bound) of coordination sphere indicates that the zinc ion in Ddi2/3 is likely the catalytic zinc directly participating in the reaction, interacting with the substrates (McCall *et al*., 2000).

4.4.4. The Binding Mode Of Cyanamide At Zinc Site In Ddi2/3 Is Similar To That In Carbonic Anhydrase II (CAII).

The refined model of cyanamide binding at the active site of Ddi2/3-T157V mutant is showing that, the substrate molecule directly binds to the zinc ion, via the N2 atom of the cyanide group of cyanamide. The rest of the molecule forms the hydrogen bonds with residues at active site to stabilize the position (Figure 4–17A) and to expose the carbon atom which will be attacked in the presumed reaction. This binding mode is the same as that of cyanamide binding in CAII (Figure 4–17B), since CAII is weakly

active on cyanamide and the reaction mechanism is well studied, the similarity of substrate binding would be helpful to unveil the catalytic mechanism of Ddi2/3.

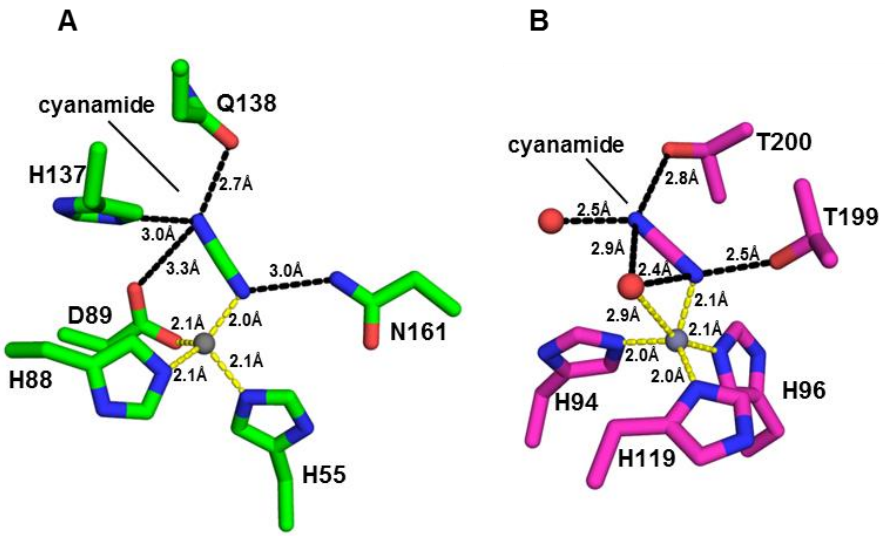


Figure 4–17. Models of cyanamide binding at the active sites of the Ddi2/3-T157V and in CAII.

A. Cyanamide binding model in Ddi2/3-T157V mutant (chain A, green). ***B.*** Cyanamide binding model in CAII (magenta, PDB ID: 1F2W). Zinc ion (gray) and water molecules (red) in both proteins are represented as spheres. Metal coordination is shown as yellow dashes; hydrogen bond is represented as black dashes. The side chains of residues are colored by elements: oxygen, red, nitrogen, blue.

CHAPTER 5 DISCUSSION AND FUTURE DIRECTIONS

5.1. Proposed Reaction Mechanism of Ddi2/3

As described in section 4.4.4. , cyanamide binding mode in Ddi2/3-T157V is the same as that in CAII, indicating the hydration mechanism could be similar to two enzymes. An intermediate of the reaction of cyanamide to urea catalyzed by hCAII has been determined by cryo-crystallographic techniques (Guerri *et al*., 2000). In this observed hCAII-cyanamide-water ternary complex (PDB ID: 2CBA), cyanamide is bound to the catalytic zinc ion, a water molecule approached to zinc ion, together with the rest three histidine residues forming a five-coordinate adduct. It is hypothesized that the water molecule/hydroxide ion is about to perform a nucleophilic attack on the zinc-activated cyanamide substrate in the catalyzed reaction. The structural evidence is consistent with the kinetic data, and indicates that a different mechanism was carried out compared to that of the physiologic carbon dioxide (CO_2) hydration reaction. In the reaction of CO2 hydration by CAII, CO_2 molecule was brought close to the zinc ion, allowing the electron-rich hydroxide produced by zinc ion to attack the carbon dioxide, forming bicarbonate, and then it is displaced by a water molecule (Lindskog, 1997).

As shown in Figure 4–16, cyanamide is possible to bind to zinc ion via either cyanide group or ammonia group. Both the two binding models of cyanamide to the zinc suggest that carbon is polar positive and would be the target of a nucleophile to initial reaction.

If the cyanamide molecule directly binds to zinc ion via N2 from cyanide group, the ammonia group would form hydrogen bonds to Q138 and D89, helping to position the cyanamide molecule and carbon is exposed as the nucleophilic attacking site. It is proposed that the binding of the nitrogen atom of the cyanide group to the zinc ion cause the polarization of carbon atom in cyanamide, making it become more polar positive, and thus it will be easily attacked by an approaching water molecule. Besides, the H137 carries positive charge under this condition, if a water molecule gets close enough to the charged H137, it would be activated to become a hydroxide to attack the carbon to initial the reaction.

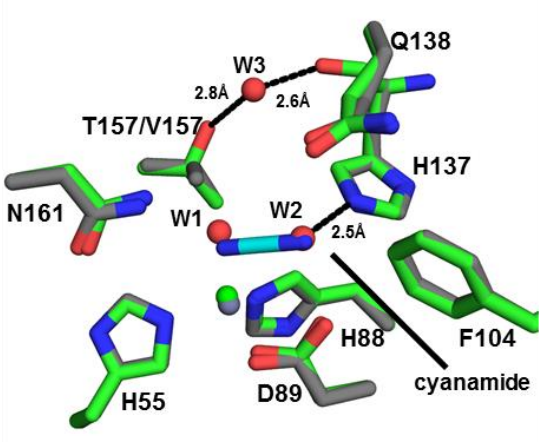


Figure 5–1. Overlaid models of the active sites of wild type Ddi2/3 and that of a complex of Ddi2/3-T157V and cyanamide.

Coordinates showing the overlaid models of active sites of wild type Ddi2/3 (chain D, green) and complex of Ddi2/3-T157V and cyanamide. (chain D, gray). Zinc ions (gray and green) and water (red) are represented as spheres. Hydrogen bonds are represented as black dashes. The side chains of residues are colored by element, oxygen, red; nitrogen, blue.

In the other hand, if the cyanamide binds zinc via ammonia group, the nitrogen atom of the cyanide group has a lone pair and performs as a donor of hydrogen bonds. The hydrogen bonds between the nitrogen atom and both H137 and Q138 makes the carbon more polar positive, and therefore becomes the target of the nucleophile.

However, the source of nucleophilic hydroxide cannot be determined; the solvent molecules observed in the current model are less likely to be the source of the hydroxides. Figure 5–1 is showing the overlaid models of wild type Ddi2/3 and Ddi2/3-T157V with cyanamide. After checking the overlaid models in 7 chains of the 9 chains, the closest water molecule in active sites of wild type Ddi2/3 is 3.4Å from the carbon of the cyanamide. The distance indicates the water molecule is less likely to be the source of the nucleophilic hydroxide. Further model refinement with solvent molecules, and future studies are needed to determine the details of the catalytic mechanism of Ddi2/3.

5.2. Ddi2/3 Represents a Unique Subfamily of HD Domain Protein

Although Ddi2/3 shares structural similarity to HD domain proteins, the difference of metal binding residues make the protein unique to identified HD domain proteins. Only three HD residues (H…HD) instead of four residues (H…HD…D) are binding to a Zinc ion in Ddi2/3. In order to identify the conserved residues of the cyanamide hydratase domain, a number of proteins that were proposed as cyanamide hydratase were retrieved from NCBI database and then used in a multiple sequence alignment.

A search for predicted cyanamide hydratase in NCBI protein database shows that the hits are distributed to bacteria (450 hits) and fungi (235 hits) (updated in Jun., 2016). Predicted Cah proteins in bacteria are enriched in *Actinobacteria* (398/450 hits), a group

of Gram-positive bacteria with high guanine and cytosine content in their DNA. Most of the predicted Cah in fungi, including Sc_Ddi2/3 and Mv_Cah, are from the *Ascomycetes* (212/235 hits) branches. No hits were found in plants and archaea, though 1 hit was found in animal, the predicted cyanamide hydratase from *Marmota marmota marmot* was identical to Sc_Ddi2/3, indicating that it was likely to be a contamination during genome sequencing.

As shown in Figure 5–2, proteins in group 1 are predicted cyanamide hydratase from fungi; candidates in group 2 are hypothesized cyanamide hydratases from bacteria. Sequence alignment of these proteins reveals that the HD residues coordinating metal ions (H…HD) are conserved in all 18 protein candidates, indicating that the metal coordination geometry is conserved and the metal cofactor is critical to protein structure and/or enzymatic activity.

In addition, conserved residues include $F104_{Ddi2/3}$ and $H137_{Ddi2/3}$ that form a pocket at the metal ion, suggesting that they are important to support the active site and could be involved in catalysis. However, $Q138_{Ddi2/3}$ is conserved in all candidates from fungi, but it is variable as either leucine or threonine in all candidates from bacteria. Leucine residue has a hydrophobic side chain; therefore it will not perform any interaction with solvent or substrate (cyanamide) during the reaction (if the protein has cyanamide hydrates activity). The threonine residue has a weakly nucleophilic side chain and can function differently than glutamine that is an amide. $T157_{Ddi2/3}$ is substituted to either valine or alanine in all candidates from bacteria, both of which have small hydrophobic side chains; therefore it will not have the same function as $T157_{Ddi2/3}$. $N161_{Ddi2/3}$ is changed to valine in all candidates from bacteria. As valine has a hydrophobic side chain, it cannot form a

hydrogen bond with solvent or substrate (cyanamide) as $N161_{Ddi2/3}$. And the hydrogen bond between $N161_{Ddi2/3}$ is critical in reaction cyanamide hydration to urea as described in the previous section, therefore this mutation in bacteria cyanamide hydratase make the protein unlikely to perform the enzymatic activities.

Identified conserved residues among all protein candidates (Figure 5–2) residues shaded in red in also include $R58_{Ddi2/3}$, $D160_{Ddi2/3}$, $G163_{Ddi2/3}$ and $P202_{Ddi2/3}$. $D160_{Ddi2/3}$ forms a hydrogen bond with $H55_{Ddi2/3}$; thus its conservation is likely helpful to maintain and stabilize the metal coordination. $R58_{Ddi2/3}$ is involved in an α-helix and its side chain faces the opposite direction of metal ion sites, indicating that its long and charged side chain may be necessary to maintain the tertiary structure of protein. The $G164_{Ddi2/3}$ and $P202_{Ddi2/3}$ are at the surface of the tertiary structure of protein. $P202_{Ddi2/3}$ is involved in forming hydrogen bonds with R8 in another Ddi2/3 molecule to maintain dimers, while $G164_{Ddi2/3}$ is located at a loop between α-helix and a 3^{10} helix; thus the residue could be essential to maintain proper folding of the protein.

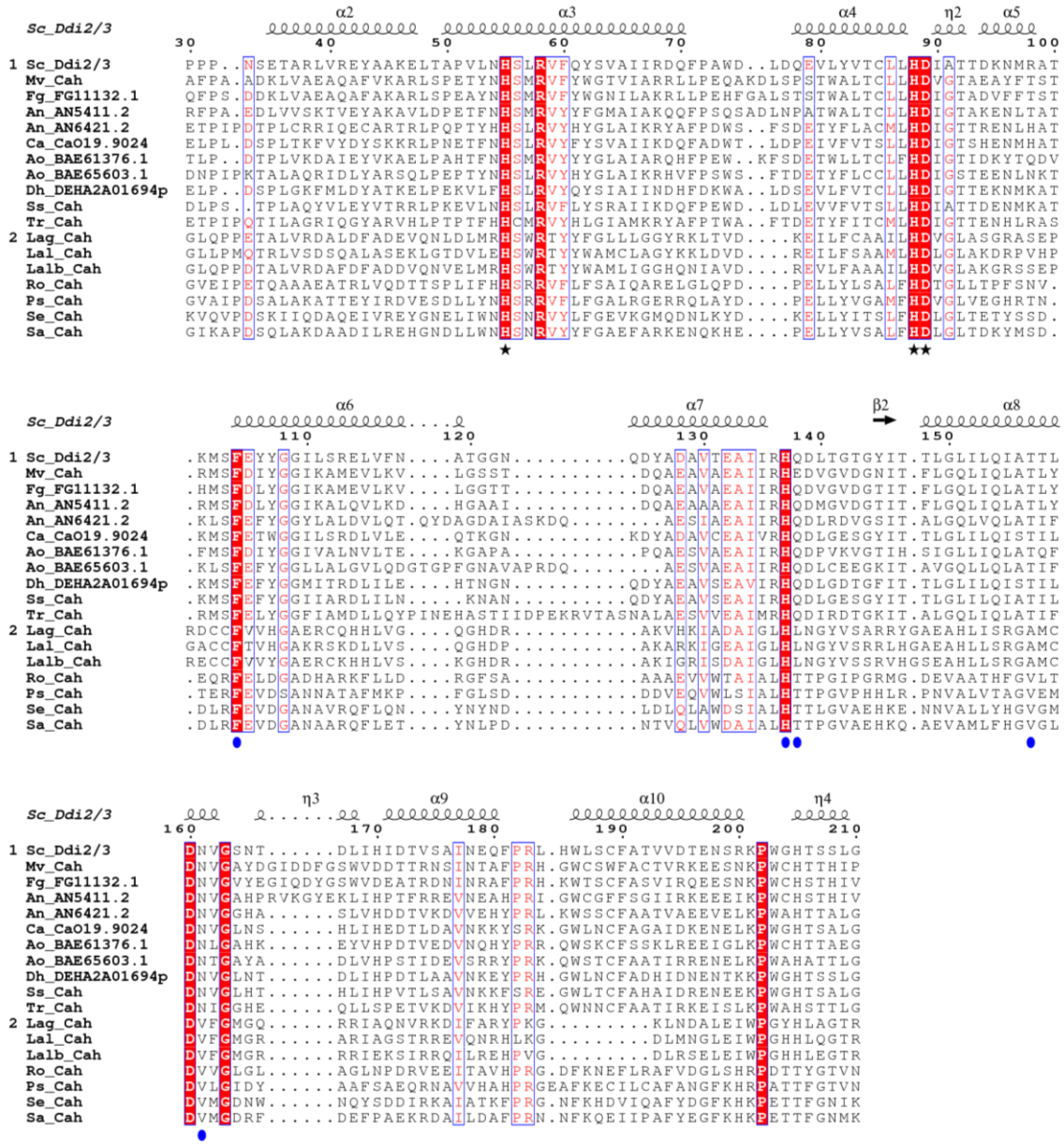


Figure 5–2. Sequence alignment of proteins containing cyanamide hydratase domain and predicted cyanamide hydratase in bacteria.

Amino acid sequences of proteins were retrieved from the NCBI protein database. The source organism abbreviations are given in the left-hand column. In Group 1: Sc, *S.cerevisiae*; Mv, *M. verrucaria*; Fg, *F. graminearum*; An, *A. nidulans*; Ca, *C. albicans*; Ao, *A. oryzae*; Dh, *D. hansenii*; Ss, *S. stipitis*; Tr, *T rubrum*. In group 2: Lag, *Labrenzia aggregate*; Lal, *Labrenzia alexandrii*; Lalb, *Labrenzia alba*; Ro, *Rhodococcus opacus*; Ps, *Pseudomonas sp. TAA207*; Se, *Streptococcus equi subsp. equi*; Sa, *Salinicoccus albus*. Amino acid sequences of proteins were retrieved from the NCBI conserved domain

database. Sequence alignment was processed by ClustalW and illustrated by using ESPript 3.0, globe score was set as 0.8. Consensus residues are shaded in red. Only regions containing identical residues are displayed. HD residues coordinating zinc ion are pointed out by asterisks. $H137_{Ddi2/3}$, $Q138_{Ddi2/3}$, $T157_{Ddi2/3}$ and $N161_{Ddi2/3}$ are pointed out by blue solid circles.

5.3. Future Research Directions

5.3.1. Physiological Roles of Ddi2/3

Why are the *DDI2/3* genes kept in the yeast genome? Why two copies of *DDI2/3* are present in the genome? It is believed that a genome keeps these genes for a purpose, most likely for the survival of the species under conditions of chemical or oxidation stress.

Our studies demonstrate that *DDI2/3* encode a functional cyanamide hydratase that specifically catalyzes cyanamide hydration. The transcription of *DDI2/3* is tightly regulated and activated by cyanamide. The observations suggest that the presence of *DDI2/3* is to break down cyanamide to nutrient. However, cyanamide is synthesized by limited plants, and it is not commonly present in nature. In addition, the chemical has only mild toxicity to yeast, and the *ddi2Δ ddi3Δ* mutant yeast strain can survive in 5 mM cyanamide. Therefore it looks like a waste for the genome keeping two copies in the genome. It is possibly inherited from an ancestor soil fungus to survive from cyanamide-synthesizing plants. However, only *S. cerevisiae* contains two identical copies of the gene. *A. oryzae* and *A nidulans* were identified to contain two copies of hypothesized cyanamide hydratase genes, but their sequences diverge from each other during evolution.

Therefore it is hypothesized that Ddi2/3 has a second substrate that is more important in cell physiology; thus the yeast keeps two identical copies of *DDI2/3*. Further research would be to study the possible substrate of Ddi2/3. Constructed *ddi2Δ ddi3Δ* can

be exposed to a variety of conditions to observe any phenotypic change compared to wild type. This would be helpful to identify the physiological role(s) of *DDI2/3*.

5.3.2. Transcription Induction Mechanism of *DDI2/3* by Cyanamide and MMS

As previously described, the transcriptional level of *DDI2/3* is restricted, as it is almost undetectable under normal condition, but after treatment with cyanamide or MMS, the transcriptional level of *DDI2/3* could be induced by over hundred-fold. This indicates that the transcription of *DDI2/3* is under tight regulation. Unlike other identified DNA damage inducible genes, the transcriptional level of *DDI2/3* is only induced by SN_2-type alkylating agents and cyanamide. It indicates that the transcription regulation pathway is different from that of known DNA damage-inducible genes.

MMS is known as a typical DNA methylating agent. It methylates nucleotides on DNA and free (d)NTPs, while cyanamide has not been identified to have any ability to damage DNA, and the two chemicals have no structural similarity or share any functional group. Therefore, it is wondered what is common between the two chemicals that induce the transcription level of *DDI2/3*.

The first hypothesis was that excessive cyanamide could disturb (d)NTPs pool in the cell. It is reported that in prebiotic age, cyanamide was present in a rich amount in nature due to the environmental conditions (Schimpl *et al*., 1965; Steinman *et al*., 1964). Moreover, cyanamide could be used in pyrimidine synthesis under the prebiotic condition. Therefore, it is possible that excessive cyanamide disturbs the nucleotide synthesis in vivo and thus imbalance the (d)NTPs pool, which activates the transcription of *DDI2/3* to break down cyanamide in cells. MMS could trigger the transcription of *DDI2/3* in the

same way. MMS is reported to prefer methylating adenine and guanidine, which also cause an imbalance of (d)NTPs pool and thus induce the expression of *DDI2/3*. However, the Ddi2/3 protein has not been found to have any activity on MMS; indicating that induction of *DDI2/3* could not help to reduce damage caused by MMS. Furthermore, the imbalance of (d)NTPs pool would activate a series of genes involved in (d)NTPs synthesis as a direct feedback, especially *RNR* genes which play critical roles in (d)NTPs synthesis. To repair DNA damage, cells needs extra (d)NTPs and thus usually cause disturbance of (d)NTPs pool in cells, which may explain why *RNR3*, a DNA damage-inducible gene, responds to a wide range of DNA damaging agents, including MMS. However, cyanamide does not induce the expression of *RNR3* at all, indicating that cyanamide is unlikely to damage (d)NTPs or DNA molecules.

Another hypothesis is based on gene expression regulation by epigenetics. It is reported that MMS methylates histone, which could affect protein expression through epigenetic regulation. MMS treatment induced the phosphorylation of serine 1 (S1) of histone H4 in a casein kinase II (CK2)-dependent manner. It was found that this phosphorylation event was associated with MMS- or phleomycin-induced DSBs, but not with UV-induced DNA damage. It was proposed that histone H4 S1 was one of the physiological substrates of CK2 in regulating a DNA-damage response upon DSBs (Cheung *et al*., 2005).

Cyanamide is used as a dormancy-break agent on plants. It was found that the treatment could up-regulate GDBRPK, a transcript for an SNF-like protein kinase in grapevine. Since SNF and SNF-like protein kinases are known as sensors of stress signals, it was hypothesized that GDBRPK is involved in the regulation of stress response after

cyanamide treatment (Or *et al*., 2000). A recent study on inhibition of cyanamide to the growth of root tips of maize (*Zea mays*) revealed that the cell cycle activity, endoreduplication intensity, and modifications of cyclins CycA2, CycD2, and histone H3 gene expression were all affected by only 3 mM cyanamide treatment. Furthermore, prolonged cyanamide treatment induced oxidative stress in root cells, accompanied with an increased concentration of auxin and stimulated ethylene emission. It was proposed that inhibition of root growth by cyanamide may be a consequence of stress-induced morphogenic responses (Soltys *et al*., 2014).

These studies indicate that both MMS and cyanamide affects histone modification and in turn activate the corresponding stress responses. It is reasonable to propose that the activation of transcription of *DDI2/3* by both MMS and cyanamide is through a histone-dependent pathway.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D *66*, 213-221.

Agerbirk, N., and Olsen, C.E. (2012). Glucosinolate structures in evolution. Phytochemistry *77*, 16-45.

Alterio, V., Di Fiore, A., D'Ambrosio, K., Supuran, C.T., and De Simone, G. (2012). Multiple binding modes of inhibitors to carbonic anhydrases: How to design specific drugs targeting 15 different isoforms? Chem Rev *112*, 4421-4468.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

An, X., Zhang, Z., Yang, K., and Huang, M. (2006). Cotransport of the heterodimeric small subunit of the Saccharomyces cerevisiae ribonucleotide reductase between the nucleus and the cytoplasm. Genetics *173*, 63-73.

Aravind, L., and Koonin, E.V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci *23*, 469-472.

Arima, Y., Nishigori, C., Takeuchi, T., Oka, S., Morimoto, K., Utani, A., and Miyachi, Y. (2006). 4-Nitroquinoline 1-oxide forms 8-hydroxydeoxyguanosine in human fibroblasts through reactive oxygen species. Toxicol Sci *91*, 382-392.

Arner, R.J., Prabhu, K.S., Krishnan, V., Johnson, M.C., and Reddy, C.C. (2006). Expression of myo-inositol oxygenase in tissues susceptible to diabetic complications. Biochem Biophys Res Commun *339*, 816-820.

Arner, R.J., Prabhu, K.S., Thompson, J.T., Hildenbrandt, G.R., Liken, A.D., and Reddy, C.C. (2001). myo-Inositol oxygenase: molecular cloning and expression of a unique enzyme that oxidizes myo-inositol and D-chiro-inositol. Biochem J *360*, 313-320.

Asplin, I., Galasko, G., and Larner, J. (1993). chiro-Inositol deficiency and insulin resistance: a comparison of the chiro-inositol- and the myo-inositol-containing insulin mediators isolated from urine, hemodialysate, and muscle of control and type II diabetic subjects. Proc Natl Acad Sci USA *90*, 5924-5928.

Banerjee, A., Sharma, R., and Banerjee, U.C. (2002). The nitrile-degrading enzymes: current status and future prospects. Appl Microbiol Biotechnol *60*, 33-44.

Bang, D.D., Timmermans, V., Verhage, R., Zeeman, A.M., van de Putte, P., and Brouwer, J. (1995). Regulation of the Saccharomyces cerevisiae DNA repair gene RAD16. Nucleic Acids Res *23*, 1679-1685.

Barglow, K.T., Saikatendu, K.S., Bracey, M.H., Huey, R., Morris, G.M., Olson, A.J., Stevens, R.C., and Cravatt, B.F. (2008). Functional proteomic and structural insights into molecular recognition in the nitrilase family enzymes. Biochemistry *47*, 13514-13523.

Bartel, B., and Fink, G.R. (1994). Differential regulation of an auxin-producing nitrilase gene family in Arabidopsis thaliana. Proc Natl Acad Sci USA *91*, 6649-6653.

Bartling, D., Seedorf, M., MithÖFer, A., and Weiler, E.W. (1992). Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-acetonitrile to the plant hormone, indole-3-acetic acid. Eur J Biochem *205*, 417-424.

Basile, G., Aker, M., and Mortimer, R.K. (1992). Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene RAD51. Mol Cell Biol *12*, 3235-3246.

Basso, T.S., Pungartnik, C., and Brendel, M. (2008). Low productivity of ribonucleotide reductase in Saccharomyces cerevisiae increases sensitivity to stannous chloride. Genet Mol Res *7*, 1-6.

Beauchamp, B.B., and Richardson, C.C. (1988). A unique deoxyguanosine triphosphatase is responsible for the optA1 phenotype of Escherichia coli. Proc Natl Acad Sci USA *85*, 2563-2567.

Bellini, D., Caly, D.L., McCarthy, Y., Bumann, M., An, S.Q., Dow, J.M., Ryan, R.P., and Walsh, M.A. (2014). Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. Mol Microbiol *91*, 26-38.

Beloglazova, N., Flick, R., Tchigvintsev, A., Brown, G., Popovic, A., Nocek, B., and Yakunin, A.F. (2013). Nuclease activity of the human SAMHD1 protein implicated in the Aicardi-Goutieres syndrome and HIV-1 restriction. J Biol Chem *288*, 8101-8110.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A.F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. EMBO J *30*, 4616-4627.

Berben, G., Dumont, J., Gilliquet, V., Bolle, P.A., and Hilger, F. (1991). The YDp plasmids: a uniform set of vectors bearing versatile gene disruption cassettes for Saccharomyces cerevisiae. Yeast (Chichester, England) *7*, 475-477.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res *28*, 235-242.

Bestwick, L.A., Grønning, L.M., James, D.C., Bones, A., and Rossiter, J.T. (1993). Purification and characterization of a nitrilase from Brassica napus. Physiol Plant *89*, 811-816.

Blakey, A.J., Colby, J., Williams, E., and O'Reilly, C. (1995). Regio- and stereo-specific nitrile Hydrolysis by the nitrile hydratase from Rhodococcus AJ270. FEMS Microbiol Lett *129*, 57-61.

Boiteux, S., and Jinks-Robertson, S. (2013). DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae. Genetics *193*, 1025-1064.

Brady, D., Dube, N., and Petersen, R. (2006). Green chemistry: highly selective biocatalytic hydrolysis of nitrile compounds. S Afr J Sci *102*, 339-344.

Brandao, P.F., and Bull, A.T. (2003). Nitrile hydrolysing activities of deep-sea and terrestrial mycolate actinomycetes. Antonie van Leeuwenhoek *84*, 89-98.

Braun, E.L., Fuge, E.K., Padilla, P.A., and Werner-Washburne, M. (1996). A stationary-phase gene in Saccharomyces cerevisiae is a member of a novel, highly conserved gene family. J Bacteriol *178*, 6865-6872.

Brenner, C. (2002). Catalysis in the nitrilase superfamily. Curr Opin Struct Biol *12*, 775-782.

Briganti, F., Mangani, S., Scozzafava, A., Vernaglione, G., and Supuran, C.T. (1999). Carbonic anhydrase catalyzes cyanamide hydration to urea: is it mimicking the physiological reaction? JBIC *4*, 528-536.

Brown, P.M., Caradoc-Davies, T.T., Dickson, J.M., Cooper, G.J.S., Loomes, K.M., and Baker, E.N. (2006). Purification, crystallization and preliminary crystallographic analysis of mouse myo-inositol oxygenase. Acta Crystallogr F *62*, 811-813.

Brownt, P.M., Caradoc-Davies, T.T., Dickson, J.M.J., Cooper, G.J.S., Loomes, K.M., and Baker, E.N. (2006). Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism. Proc Natl Acad Sci USA *103*, 15032-15037.

Charalampous, F.C. (1959). Biochemical studies on inositol. V. Purification and properties of the enzyme that cleaves inositol to D-glucuronic acid. J Biol Chem *234*, 220-227.

Chen, J., Derfler, B., Maskati, A., and Samson, L. (1989). Cloning a eukaryotic DNA glycosylase repair gene by the suppression of a DNA repair defect in Escherichia coli. Proc Natl Acad Sci USA *86*, 7961-7965.

Chen, J., Derfler, B., and Samson, L. (1990). Saccharomyces cerevisiae 3-methyladenine DNA glycosylase has homology to the AlkA glycosylase of E. coli and is induced in response to DNA alkylation damage. EMBO J *9*, 4569-4575.

Chen, J., and Samson, L. (1991). Induction of S.cerevisiae MAG 3-methyladenine DNA glycosylase transcript levels in response to DNA damage. Nucleic Acids Res *19*, 6427-6432.

Chen, Z.H., and Schaap, P. (2012). The prokaryote messenger c-di-GMP triggers stalk cell differentiation in Dictyostelium. Nature *488*, 680-683.

Cheung, W.L., Turner, F.B., Krishnamoorthy, T., Wolner, B., Ahn, S.H., Foley, M., Dorsey, J.A., Peterson, C.L., Berger, S.L., and Allis, C.D. (2005). Phosphorylation of histone H4 serine 1 during DNA damage requires casein kinase II in S. cerevisiae. Curr Biol *15*, 656-660.

Clarke, D.J., Mondesert, G., Segal, M., Bertolaet, B.L., Jensen, S., Wolff, M., Henze, M., and Reed, S.I. (2001). Dosage suppressors of pds1 implicate ubiquitin-associated domains in checkpoint control. Mol Cell Biol *21*, 1997-2007.

Coissac, E., Maillier, E., and Netter, P. (1997). A comparative study of duplications in bacteria and eukaryotes: the importance of telomeres. Mol Biol Evol *14*, 1062-1074.

Daley, J.M., Palmbos, P.L., Wu, D., and Wilson, T.E. (2005). Nonhomologous end joining in yeast. Annu Rev Genet *39*, 431-451.

DeMaster, E.G., Redfern, B., and Nagasawa, H.T. (1998). Mechanisms of inhibition of aldehyde dehydrogenase by nitroxyl, the active metabolite of the alcohol deterrent agent cyanamide. Biochem Pharmacol *55*, 2007-2015.

DeMaster, E.G., Shirota, F.N., and Nagasawa, H.T. (1984). The metabolic activation of cyanamide to an inhibitor of aldehyde dehydrogenase is catalyzed by catalase. Biochem Biophys Res Commun *122*, 358-365.

DiGeronimo, M.J., and Antoine, A.D. (1976). Metabolism of acetonitrile and propionitrile by Nocardia rhodochrous LL100-21. Appl Environ Microbiol *31*, 900-906.

Domkin, V., Thelander, L., and Chabes, A. (2002). Yeast DNA damage-inducible Rnr3 has a very low catalytic activity strongly stimulated after the formation of a cross-talking Rnr1/Rnr3 complex. J Biol Chem *277*, 18574-18578.

Durant, G.J., Emmett, J.C., and Ganellin, C.R. (1975). Certain n-cyanoguanidines (Smith Kline French Lab).

Elledge, S.J., and Davis, R.W. (1990). Two genes differentially regulated in the cell cycle and by DNA-damaging agents encode alternative regulatory subunits of ribonucleotide reductase. Genes Dev *4*, 740-751.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Endo-Ichikawa, Y., Kohno, H., Tokunaga, R., and Taketani, S. (1995). Induction in the gene RNR3 in Saccharomyces cerevisiae upon exposure to different agents related to carcinogenesis. Biochem Pharmacol *50*, 1695-1699.

Endo, I., Nojiri, M., Tsujimura, M., Nakasako, M., Nagashima, S., Yohda, M., and Odaka, M. (2001). Fe-type nitrile hydratase. J Inorg Biochem *83*, 247-253.

Foerstner, K.U., Doerks, T., Muller, J., Raes, J., and Bork, P. (2008). A nitrile hydratase in the eukaryote Monosiga brevicollis. PloS one *3*, e3976.

Friedberg, E.C. (2006). DNA repair and mutagenesis (Washington, D.C.: ASM Press).

Fry, R.C., Begley, T.J., and Samson, L.D. (2005). Genome-wide responses to DNA-damaging agents. Annu Rev Microbiol *59*, 357-377.

Fu, Y. (2008). The regulatory network controlling DNA damage responses in Saccharomyces cerevisiae. In Microbiology and Immunology (University of Saskatchewan).

Fu, Y., Pastushok, L., and Xiao, W. (2008). DNA damage-induced gene expression in Saccharomyces cerevisiae. FEMS Microbiol Rev *32*, 908-926.

Fuchs., J.J., and Wommak, J.B. (1974). Process for preparing herbicidal triazines (Du Pont).

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J., and Brown, P.O. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. Mol Biol Cell *12*, 2987-3003.

Gobler, C.J., Berry, D.L., Dyhrman, S.T., Wilhelm, S.W., Salamov, A., Lobanov, A.V., Zhang, Y., Collier, J.L., Wurch, L.L., Kustka, A.B., *et al.* (2011). Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. Proc Natl Acad Sci USA *108*, 4352-4357.

Goldstone, D.C., Ennis-Adeniran, V., Hedden, J.J., Groom, H.C.T., Rice, G.I., Christodoulou, E., Walker, P.A., Kelly, G., Haire, L.F., Yap, M.W., *et al.* (2011). HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. Nature *480*, 379-382.

Goncalves, A., Karayel, E., Rice, G.I., Bennett, K.L., Crow, Y.J., Superti-Furga, G., and Burckstummer, T. (2012). SAMHD1 is a nucleic-acid binding protein that is mislocalized due to aicardi-goutieres syndrome-associated mutations. Hum Mutat *33*, 1116-1122.

Gong, J.-S., Lu, Z.-M., Li, H., Shi, J.-S., Zhou, Z.-M., and Xu, Z.-H. (2012). Nitrilases in nitrile biocatalysis: recent progress and forthcoming research. Microb Cell Fact *11*, 1-18.

Guerri, A., Briganti, F., Scozzafava, A., Supuran, C.T., and Mangani, S. (2000). Mechanism of cyanamide hydration catalyzed by Carbonic Anhydrase II suggested by cryogenic X-ray diffraction. Biochemistry *39*, 12391-12397.

Güthner, T., and Mertschenk, B. (2000). Cyanamides. In Ullmann's Encyclopedia of Industrial Chemistry (Wiley-VCH Verlag GmbH & Co. KGaA).

Hankes, L.V., Politzer, W.M., Touster, O., and Anderson, L. (1969). myo-Inositol catabolism in human pentosurics: the predominant role of the glucuronate-xylulose-pentose phosphate pathway. Ann NY Acad Sci *165*, 564-576.

Harper, D.B. (1977). Fungal degradation of aromatic nitriles. Enzymology of C-N cleavage by Fusarium solani. The Biochemical journal *167*, 685-692.

Heinemann, U., Engels, D., Bürger, S., Kiziak, C., Mattes, R., and Stolz, A. (2003). Cloning of a nitrilase gene from the cyanobacterium Synechocystis sp. strain PCC6803 and heterologous expression and characterization of the encoded protein. Appl Microbiol *69*, 4359-4366.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol *7*, 263-273.

Higgins, D.R., Prakash, S., Reynolds, P., and Prakash, L. (1983). Molecular cloning and characterization of the RAD1 gene of Saccharomyces cerevisiae. Gene *26*, 119-126.

Hoeijmakers, J.H. (1993). Nucleotide excision repair I: from E. coli to yeast. Trends Genet *9*, 173-177.

Hogg, T., Mechold, U., Malke, H., Cashel, M., and Hilgenfeld, R. (2004). Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response [corrected]. Cell *117*, 57-68.

Hsu, J.C., and Camper, N.D. (1979). Degradation of ioxynil by a soil fungus, Fusarium solani. Soil Biol Biochem *11*, 19-22.

Huang, M., Zhou, Z., and Elledge, S.J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. Cell *94*, 595-605.

Huber, H.E., Beauchamp, B.B., and Richardson, C.C. (1988). Escherichia coli dGTP triphosphohydrolase is inhibited by gene 1.2 protein of bacteriophage T7. J Biol Chem *263*, 13549-13556.

Hulme, R., Zamora, O.D.P., Mota, E.J., Pastén, M.A., Contreras-Rojas, R., Miranda, R., Valencia-Hernández, I., Correa-Basurto, J., Trujillo-Ferrara, J., and Delgado, F. (2008). Cyanamide: a convenient building block to synthesize 4-aryl-2-cyanoimino-3,4-dihydro-1H-pyrimidine systems via a multicomponent reaction. Tetrahedron *64*, 3372-3380.

Huynh, T.N., Luo, S., Pensinger, D., Sauer, J.D., Tong, L., and Woodward, J.J. (2015). An HD-domain phosphodiesterase mediates cooperative hydrolysis of c-di-AMP to affect bacterial growth and virulence. Proc Natl Acad Sci USA *112*, E747-E756.

Ikenaga, M., Ichikawa-Ryo, H., and Kondo, S. (1975). The major cause of inactivation and mutation by 4-nitroquinoline 1-oixde in Escherichia coli: excisable 4NQO-purine adducts. J Mol Biol *92*, 341-356.

Irvine, J.C., Ritchie, R.H., Favaloro, J.L., Andrews, K.L., Widdop, R.E., and Kemp-Harper, B.K. (2008). Nitroxyl (HNO): the Cinderella of the nitric oxide story. Trends Pharmacol Sci *29*, 601-608.

Janowitz, T., Trompetter, I., and Piotrowski, M. (2009). Evolution of nitrilases in glucosinolate-containing plants. Phytochemistry *70*, 1680-1686.

Jelinsky, S.A., Estep, P., Church, G.M., and Samson, L.D. (2000). Regulatory networks revealed by transcriptional profiling of damaged Saccharomyces cerevisiae cells: Rpn4 links base excision repair with proteasomes. Mol Cell Biol *20*, 8157-8167.

Jelinsky, S.A., and Samson, L.D. (1999). Global response of Saccharomyces cerevisiae to an alkylating agent. Proc Natl Acad Sci USA *96*, 1486-1491.

Jeon, Y.J., Park, S.C., Song, W.S., Kim, O.H., Oh, B.C., and Yoon, S.I. (2016). Structural and biochemical characterization of bacterial YpgQ protein reveals a metal-dependent nucleotide pyrophosphohydrolase. J Struct Biol *195*, 113-122.

Jeon, Y.J., Song, W.S., and Yoon, S.I. (2014). Purification, crystallization and preliminary X-ray analysis of a putative nucleotide phosphohydrolase, YpgQ, from Bacillus subtilis. Acta Crystallogr F *70*, 984-986.

Jia, X., and Xiao, W. (2003). Compromised DNA repair enhances sensitivity of the yeast RNR3-lacZ genotoxicity testing system. Toxicol Sci *75*, 82-88.

Jia, X.M., Zhu, Y., and Xiao, W. (2002). A stable and sensitive genotoxic testing system based on DNA damage induced gene expression in Saccharomyces cerevisiae. Mutat Res-Gen Tox En *519*, 83-92.

Jones, J.S., Prakash, L., and Prakash, S. (1990). Regulated expression of the Saccharomyces cerevisiae DNA repair gene RAD7 in response to DNA damage and during sporulation. Nucleic Acids Res *18*, 3281-3285.

Kamo, T., Endo, M., Sato, M., Kasahara, R., Yamaya, H., Hiradate, S., Fujii, Y., Hirai, N., and Hirota, M. (2008). Limited distribution of natural cyanamide in higher plants: Occurrence in Vicia villosa subsp varia, V. cracca, and Robinia pseudo-acacia. Phytochemistry *69*, 1166-1172.

Kamo, T., Hiradate, S., and Fujii, Y. (2003). First isolation of natural cyanamide as a possible allelochemical from hairy vetch Vicia villosa. J Chem Ecol *29*, 275-283.

Kamo, T., Kasahara, R., Abe, S., Hirota, M., Sugano, M., Yamaya, H., Hiradate, S., and Fujii, Y. (2010). Carbon sources of natural cyanamide in Vicia villosa subsp. varia. Nat Prod Res *24*, 1637-1642.

Kamo, T., Kato, K., Abe, S., Hirota, M., Yamaya, H., Hiradate, S., and Fujii, Y. (2009). Biosynthetic origin of the nitrogen atom in cyanamide in Vicia villosa subsp varia. Soil Sci Plant Nutr *55*, 235-242.

Kamo, T., Kato, K., Hiradate, S., Nakajima, E., Fujii, Y., and Hirota, M. (2006a). Evidence of cyanamide production in hairy vetch Vicia villosa. Nat Prod Res *20*, 429-433.

Kamo, T., Sakurai, S., Yamanashi, T., and Todoroki, Y. (2015). Cyanamide is biosynthesized from l-canavanine in plants. Sci Rep *5*, 10527.

Kamo, T., Sato, M., Kato, K., Hiradate, S., Nakajima, E., Fujii, Y., and Hirota, M. (2006b). Quantification of cyanamide contents in herbaceous plants. Biosci Biotechnol Biochem *70*, 2310-2312.

Kassim, M.A., and Rumbold, K. (2014). HCN production and hydroxynitrile lyase: a natural activity in plants and a renewed biotechnological interest. Biotechnol Lett *36*, 223-228.

Katla, V., Syed, R., Kuruva, C., Kuntrapakam, H., and Chamarthi, N. (2013). Synthesis of Novel Phosphorylated Guanidine Derivatives from Cyanamide and Their Anti-inflammatory Activity. Chem Pharm Bull *61*, 25-32.

Kawa, J.M., Przybylski, R., and Taylor, C.G. (2003). Urinary chiro-inositol and myo-inositol excretion is elevated in the diabetic db/db mouse and streptozotocin diabetic rat. Exp Biol Med (Maywood) *228*, 907-914.

Ke, S.H., and Madison, E.L. (1997). Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method. Nucleic Acids Res *25*, 3371-3372.

Kobayashi, M., and Shimizu, S. (1994). Versatile nitrilases: nitrile-hydrolysing enzymes. FEMS Microbiol Lett *120*, 217-223.

Kondo, N., Kuramitsu, S., and Masui, R. (2004). Biochemical characterization of TT1383 from Thermus thermophilus identifies a novel dNTP triphosphohydrolase activity stimulated by dATP and dTTP. J Biochem *136*, 221-231.

Kondo, N., Nakagawa, N., Ebihara, A., Chen, L., Liu, Z.-J., Wang, B.-C., Yokoyama, S., Kuramitsu, S., and Masui, R. (2007). Structure of dNTP-inducible dNTP triphosphohydrolase: insight into broad specificity for dNTPs and triphosphohydrolase-type hydrolysis. Acta Crystallogr D Biol Crystallogr *63*, 230-239.

Kornberg, S.R., Lehman, I.R., Bessman, M.J., Simms, E.S., and Kornberg, A. (1958). Enzymatic cleavage of deoxyguanosine triphosphate to deoxyguanosine and tripolyphophate. J Biol Chem *233*, 159-162.

Kumar, D., Viberg, J., Nilsson, A.K., and Chabes, A. (2010). Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint. Nucleic Acids Res *38*, 3975-3983.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol *33*, 1870-1874.

Langdahl, B.R., Bisp, P., and Ingvorsen, K. (1996). Nitrile hydrolysis by Rhodococcus erythropolis BL1, an acetonitrile-tolerant strain isolated from a marine sediment. Microbiology *142*, 145-154.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Layh, N., Stolz, A., Bohme, J., Effenberger, F., and Knackmuss, H.J. (1994). Enantioselective hydrolysis of racemic naproxen nitrile and naproxen amide to S-naproxen by new bacterial isolates. J Biotechnol *33*, 175-182.

Lee, M.W., Kim, B.J., Choi, H.K., Ryu, M.J., Kim, S.B., Kang, K.M., Cho, E.J., Youn, H.D., Huh, W.K., and Kim, S.T. (2007). Global protein expression profiling of budding yeast in response to DNA damage. Yeast (Chichester, England) *24*, 145-154.

Levitt, G., and Petersen, W.C. (1983). Herbicidal ureas and isoureas (E. I. Du Pont De Nemours And Company).

Li, J., Biss, M., Fu, Y., Xu, X., Moore, S.A., and Xiao, W. (2015). Two duplicated genes DDI2 and DDI3 in budding yeast encode a cyanamide hydratase and are induced by cyanamide. J Biol Chem *290*, 12664-12675.

Lindskog, S. (1997). Structure and mechanism of carbonic anhydrase. Pharmacol Ther *74*, 1-20.

Liu, Y., Dai, H., and Xiao, W. (1997). UAS(MAG1), a yeast cis-acting element that regulates the expression of MAG1, is located within the protein coding region of DDI1. Mol Gen Genet *255*, 533-542.

Liu, Y.L., and Xiao, W. (1997). Bidirectional regulation of two DNA-damage-inducible genes, MAG1 and DDI1, from Saccharomyces cerevisiae. Mol Microbiol *23*, 777-789.

Lovering, A.L., Capeness, M.J., Lambert, C., Hobley, L., and Sockett, R.E. (2011). The structure of an unconventional HD-GYP protein from bdellovibrio reveals the roles of conserved residues in this class of Cyclic-di-GMP phosphodiesterases. mBio *2*, e00163.

Madura, K., and Prakash, S. (1990). Transcript levels of the Saccharomyes cerevisiae DNA repair gene RAD23 increase in response to UV light and in meiosis but remain constant in the mitotic cell cycle. Nucleic Acids Res *18*, 4737-4742.

Magnusson, B., and Kligman, A.M. (1969). The identification of contact allergens by animal assay. The guinea pig maximization test. J Invest Dermatol *52*, 268-276.

Maiergreiner, U., Obermaierskrobranek, B., Estermaier, L., Kammerloher, W., Freund, C., Wulfing, C., Burkert, U., Matern, D., Breuer, M., Eulitz, M., *et al.* (1991). Isolation and properties of a nitrile hydratase from the soil fungus Myrothecium-verrucaria that is highly specific for the fertilizer cyanamide and cloning of its gene. Proc Natl Acad Sci USA *88*, 4260-4264.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., *et al.* (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res *43*, D222-226.

Marron, A.O., Akam, M., and Walker, G. (2012). Nitrile hydratase genes are present in multiple eukaryotic supergroups. PloS one *7*, e32867.

Martinkova, L., Vejvoda, V., and Kren, V. (2008). Selection and screening for enzymes of nitrile metabolism. J Biotechnol *133*, 318-326.

Mascharak, P.K. (2002). Structural and functional models of nitrile hydratase. Coord Chem Rev *225*, 201-214.

Mashhadi, Z., Xu, H., and White, R.H. (2009). An Fe2+-dependent cyclic phosphodiesterase catalyzes the hydrolysis of 7,8-dihydro-D-neopterin 2′,3′-cyclic phosphate in methanopterin biosynthesis. Biochemistry *48*, 9384-9392.

May, D.R. (1967). Method for manufacture of cyanamide (American Cyanamid Co).

McCall, K.A., Huang, C.-c., and Fierke, C.A. (2000). Function and mechanism of zinc metalloenzymes. J Nutr *130*, 1437S-1446S.

Mega, R., Kondo, N., Nakagawa, N., Kuramitsu, S., and Masui, R. (2009). Two dNTP triphosphohydrolases from Pseudomonas aeruginosa possess diverse substrate specificities. FEBS J *276*, 3211-3221.

Miner, K.D., and Kurtz, D.M., Jr. (2016). Active site metal occupancy and Cyclic Di-GMP phosphodiesterase activity of Thermotoga maritima HD-GYP. Biochemistry *55*, 970-979.

Miyanaga, A., Fushinobu, S., Ito, K., and Wakagi, T. (2001). Crystal structure of cobalt-containing nitrile hydratase. Biochem Biophys Res Commun *288*, 1169-1174.

Mondzac, A., Ehrlich, G.E., and Seegmill.Je (1965). An enzymatic determination of ammonia in biological fluids. J Lab Clin Med *66*, 526-&.

Moskala, R., Reddy, C.C., Minard, R.D., and Hamilton, G.A. (1981). An oxygen-18 tracer investigation of the mechanism of myo-inositol oxygenase. Biochem Biophys Res Commun *99*, 107-113.

Moye-Rowley, W.S., Harshman, K.D., and Parker, C.S. (1989). Yeast YAP1 encodes a novel form of the jun family of transcriptional activator proteins. Genes Dev *3*, 283-292.

Mulepati, S., and Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). J Biol Chem *286*, 31896-31903.

Myers, A.M., Tzagoloff, A., Kinney, D.M., and Lusty, C.J. (1986). Yeast shuttle and integrative vectors with multiple cloning sites suitable for construction of lacZ fusions. Gene *45*, 299-310.

Nagasawa, T., and Yamada, H. (1995). Interrelations of chemistry and biotechnology-VI. Microbial production of commodity chemicals (technical report). Pure Appl Chem *67*, 1241-1256.

Narayanasamy, K., Shukla, S., and Parekh, L.J. (1990). Utilization of acrylonitrile by bacteria isolated from petrochemical waste waters. Indian J Exp Biol *28*, 968-971.

Nee, C.C., and Fuchigami, L.H. (1992). Overcoming rest at different growth stages with hydrogen cyanamide. Sci Hortic *50*, 107-113.

Nelp, M.T., Song, Y., Wysocki, V.H., and Bandarian, V. (2016). A protein-derived oxygen is the source of the amide oxygen of nitrile hydratases. J Biol Chem *291*, 7822-7829.

Nowicka, U., Zhang, D., Walker, O., Krutauz, D., Castaneda, C.A., Chaturvedi, A., Chen, T.Y., Reis, N., Glickman, M.H., and Fushman, D. (2015). DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin. Structure *23*, 542-557.

O'Reilly, C., and Turner, P.D. (2003). The nitrilase family of CN hydrolysing enzymes - a comparative study. J Appl Microbiol *95*, 1161-1174.

Olson, J.A., and Anfinsen, C.B. (1953). Kinetic and equilibrium studies on crystalline L-glutamic acid dehydrogenase. J Biol Chem *202*, 841-856.

Or, E., Vilozny, I., Eyal, Y., and Ogrodovitch, A. (2000). The transduction of the signal for grape bud dormancy breaking induced by hydrogen cyanamide may involve the SNF-like protein kinase GDBRPK. Plant Mol Biol *43*, 483-494.

Ostlund Jr, R.E., McGill, J.B., Herskowitz, I., Kipnis, D.M., Santiago, J.V., and Sherman, W.R. (1993). D-chiro-Inositol metabolism in diabetes mellitus. Proc Natl Acad Sci USA *90*, 9988-9992.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology (Academic Press), pp. 307-326.

Pace, H.C., and Brenner, C. (2001). The nitrilase superfamily: classification, structure and function. Genome Biol *2*, REVIEWS0001.

Padilla, P.A., Fuge, E.K., Crawford, M.E., Errett, A., and Werner-Washburne, M. (1998). The highly conserved, coregulated SNO and SNZ gene families in Saccharomyces cerevisiae respond to nutrient limitation. J Bacteriol *180*, 5718-5726.

Paolocci, N., Jackson, M.I., Lopez, B.E., Miranda, K., Tocchetti, C.G., Wink, D.A., Hobbs, A.J., and Fukuto, J.M. (2007). The pharmacology of nitroxyl (HNO) and its therapeutic potential: not just the Janus face of NO. Pharmacol Ther *113*, 442-458.

Pastushok, L., Moraes, T.F., Ellison, M.J., and Xiao, W. (2005). A single Mms2 "key" residue insertion into a Ubc13 pocket determines the interface specificity of a human Lys63 ubiquitin conjugation complex. J Biol Chem *280*, 17891-17900.

Pereira, R.A., Graham, D., Rainey, F.A., and Cowan, D.A. (1998). A novel thermostable nitrile hydratase. Extremophiles : life under extreme conditions *2*, 347-357.

Petříčková, A., Veselá, A.B., Kaplan, O., Kubáč, D., Uhnáková, B., Malandra, A., Felsberg, J., Rinágelová, A., Weyrauch, P., Křen, V., *et al.* (2011). Purification and characterization of heterologously expressed nitrilases from filamentous fungi. Appl Microbiol Biotechnol *93*, 1553-1561.

Piotrowski, M., Schönfelder, S., and Weiler, E.W. (2001). The Arabidopsis thaliana isogene NIT4 and its orthologs in tobacco encode β-cyano-l-alanine hydratase/nitrilase. J Biol Chem *276*, 2616-2621.

Powner, M.W., Gerland, B., and Sutherland, J.D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature *459*, 239-242.

Prakash, L., and Prakash, S. (1977). Isolation and characterization of MMS-sensitive mutants of Saccharomyces cerevisiae. Genetics *86*, 33-55.

Prakash, S., and Prakash, L. (2000). Nucleotide excision repair in yeast. Mutat Res-Fund Mol M *451*, 13-24.

Prakash, S., Sung, P., and Prakash, L. (1993). DNA repair genes and proteins of Saccharomyces cerevisiae. Annu Rev Genet *27*, 33-70.

Rinaldo, S., Paiardini, A., Stelitano, V., Brunotti, P., Cervoni, L., Fernicola, S., Protano, C., Vitali, M., Cutruzzolà, F., and Giardina, G. (2015). Structural basis of functional diversification of the HD-GYP domain revealed by the Pseudomonas aeruginosa PA4781 protein, which displays an unselective bimetallic binding site. J Bacteriol *197*, 1525-1535.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-W324.

Rodríguez-Navarro, S., Llorente, B., Rodríguez-Manzaneque, M.T., Ramne, A., Uber, G., Marchesan, D., Dujon, B., Herrero, E., Sunnerhagen, P., and Pérez-Ortín, J.E. (2002). Functional analysis of yeast gene families involved in metabolism of vitamins B1 and B6. Yeast (Chichester, England) *19*, 1261-1276.

Römling, U., Galperin, M.Y., and Gomelsky, M. (2013). Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev *77*, 1-52.

Rothstein, R.J. (1983). One-step gene disruption in yeast. Methods Enzymol *101*, 202-211.

Ryan, R.P., and Dow, J.M. (2010). Intermolecular interactions between HD-GYP and GGDEF domain proteins mediate virulence-related signal transduction in Xanthomonas campestris. Virulence *1*, 404-408.

Ryoo, J., Hwang, S.Y., Choi, J., Oh, C., and Ahn, K. (2016). SAMHD1, the Aicardi-Goutieres syndrome gene and retroviral restriction factor, is a phosphorolytic ribonuclease rather than a hydrolytic ribonuclease. Biochem Biophys Res Commun *477*, 977-981.

Sancar, G.B. (2000). Enzymatic photoreactivation: 50 years and counting. Mutat Res *451*, 25-37.

Sanvisens, N., de Llanos, R., and Puig, S. (2013). Function and regulation of yeast ribonucleotide reductase: cell cycle, genotoxic stress, and iron bioavailability. Biomed J *36*, 51-58.

Sauers, R.F. (1983). Herbicidal sulfonamides (E. I. Du Pont De Nemours & Company).

Schild, D., Johnston, J., Chang, C., and Mortimer, R.K. (1984). Cloning and mapping of Saccharomyces cerevisiae photoreactivation gene PHR1. Mol Cell Biol *4*, 1864-1870.

Schimpl, A., Lemmon, R.M., and Calvin, M. (1965). Cyanamide formation under primitive earth conditions. Science *147*, 149-150.

Schnell, N., Krems, B., and Entian, K.D. (1992). The PAR1 (YAP1/SNQ3) gene of Saccharomyces cerevisiae,a c-jun Homologue, is involved in oxygen metabolism. Curr Genet *21*, 269-273.

Schrodinger, LLC (2015). The PyMOL Molecular Graphics System, Version 1.8.

Seamon, K.J., Sun, Z., Shlyakhtenko, L.S., Lyubchenko, Y.L., and Stivers, J.T. (2015). SAMHD1 is a single-stranded nucleic acid binding protein with no active site-associated nuclease activity. Nucleic Acids Res *43*, 6486-6499.

Sebastian, J., Kraus, B., and Sancar, G.B. (1990). Expression of the yeast PHR1 gene is induced by DNA-damaging agents. Mol Cell Biol *10*, 4630-4637.

Sebastian, J., and Sancar, G.B. (1991). A damage-responsive DNA binding protein regulates transcription of the yeast DNA repair gene PHR1. Proc Natl Acad Sci USA *88*, 11251-11255.

Seto, D., Bhatnagar, S.K., and Bessman, M.J. (1988). The purification and properties of deoxyguanosine triphosphate triphosphohydrolase from Escherichia coli. J Biol Chem *263*, 1494-1499.

Shaw, N.M., Robins, K.T., and Kiener, A. (2003). Lonza: 20 Years of Biotransformations. Adv Synth Catal *345*, 425-435.

Shulman, Y., Nir, G., Fanberstein, L., and Lavee, S. (1983). The effect of cyanamide on the release from dormancy of grapevine buds. Sci Hortic *19*, 97-104.

Siede, W., and Friedberg, E.C. (1992). Regulation of the yeast RAD2 gene: DNA damage-dependent induction correlates with protein binding to regulatory sequences and their deletion influences survival. Mol Gen Genet *232*, 247-256.

Siede, W., Robinson, G.W., Kalainov, D., Malley, T., and Friedberg, E.C. (1989). Regulation of the RAD2 gene of Saccharomyces cerevisiae. Mol Microbiol *3*, 1697-1707.

Singh, D., Gawel, D., Itsko, M., Hochkoeppler, A., Krahn, J.M., London, R.E., and Schaaper, R.M. (2015). Structure of Escherichia coli dGTP triphosphohydrolase: a hexameric enzyme with DNA effector molecules. J Biol Chem *290*, 10418-10429.

Singh, R., Sharma, R., Tewari, N., and Rawat, D.S. (2006). Nitrilase and its application as a 'green' catalyst. Chem Biodivers *3*, 1279-1287.

Soltys, D., Rudzińska-Langwald, A., Gniazdowska, A., Wiśniewska, A., and Bogatek, R. (2012). Inhibition of tomato (Solanum lycopersicum L.) root growth by cyanamide is due to altered cell division, phytohormone balance and expansin gene expression. Planta *236*, 1629-1638.

Soltys, D., Rudzińska-Langwald, A., Kurek, W., Szajko, K., Sliwinska, E., Bogatek, R., and Gniazdowska, A. (2014). Phytotoxic cyanamide affects maize (Zea mays) root growth and root tip function: from structure to gene expression. J Plant Physiol *171*, 565-575.

Steinman, G., Lemmon, R.M., and Calvin, M. (1964). Cyanamide - possible key compound in chemical evolution. Proc Natl Acad Sci USA *52*, 27-30.

Stelitano, V., Giardina, G., Paiardini, A., Castiglione, N., Cutruzzolà, F., and Rinaldo, S. (2013). C-di-GMP hydrolysis by Pseudomonas aeruginosa HD-GYP phosphodiesterases: analysis of the reaction mechanism and novel roles for pGpG. PloS one *8*, e74920.

Steller, W.A., Frederic.Ib, and Morgan, P.W. (1965). Defoliant residues, determination of cyanamide residues on ginned cottonseed. J Agr Food Chem *13*, 329-330.

Stransky, H., and Amberger, A. (1973). Isolierung und eigenschaften einer cyanamid-hydratase (E.C.-Gruppe 4. 2.1.) aus Myrothecium verrucaria alb. u. schw. Zeitschrift Fur Pflanzenphysiologie *70*, 74-87.

Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev *66*, 630-670.

Tada, M. (1975). Seryl-tRNA synthetase and activation of the carcinogen 4-nitroguinoline 1-oxide. Nature *255*, 510-512.

Tamames, B., Sousa, S.F., Tamames, J., Fernandes, P.A., and Ramos, M.J. (2007). Analysis of zinc-ligand bond lengths in metalloproteins: trends and patterns. Proteins *69*, 466-475.

Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat *11*, 49-52.

Thimann, K.V., and Mahadevan, S. (1964). Nitrilase. I. Occurrence, preparation, and general properties of the enzyme. Arch Biochem Biophys *105*, 133-141.

Thorsell, A.-G., Persson, C., Voevodskaya, N., Busam, R.D., Hammarström, M., Gräslund, S., Gräslund, A., and Hallberg, B.M. (2008). Structural and Biophysical Characterization of Human myo-Inositol Oxygenase. J Biol Chem *283*, 15209-15216.

Thuku, R.N., Brady, D., Benedik, M.J., and Sewell, B.T. (2009). Microbial nitrilases: Versatile, spiral forming, industrial enzymes. J Appl Microbiol *106*, 703-727.

Tüngler, V., Staroske, W., Kind, B., Dobrick, M., Kretschmer, S., Schmidt, F., Krug, C., Lorenz, M., Chara, O., Schwille, P., *et al.* (2013). Single-stranded nucleic acids promote SAMHD1 complex formation. J Mol Med *91*, 759-770.

van den Bosch, M., Lohman, P.H.M., and Pastink, A. (2002). DNA double-strand break repair by homologous recombination. Biol Chem *383*, 873-892.

Van Staalduinen, L.M., McSorley, F.R., Schiessl, K., Séguin, J., Wyatt, P.B., Hammerschmidt, F., Zechel, D.L., and Jia, Z. (2014). Crystal structure of PhnZ in complex with substrate reveals a di-iron oxygenase mechanism for catabolism of organophosphonates. Proc Natl Acad Sci USA *111*, 5171-5176.

Vorontsov, I.I., Minasov, G., Kiryukhina, O., Brunzelle, J.S., Shuvalova, L., and Anderson, W.F. (2011). Characterization of the deoxynucleotide triphosphate triphosphohydrolase (dNTPase) activity of the EF1143 protein from Enterococcus faecalis and crystal structure of the activator-substrate complex. J Biol Chem *286*, 33158-33166.

Vorwerk, S., Biernacki, S., Hillebrand, H., Janzik, I., Müller, A., Weiler, E.W., and Piotrowski, M. (2001). Enzymatic characterization of the recombinant Arabidopsis thaliana nitrilase subfamily encoded by the NIT2/NIT1/NIT3-gene cluster. Planta *212*, 508-516.

Weeks, J.T. (2001). Transformation of wheat with the cyanamide hydratase gene (The United States Of America As Represented By The Secretary Of Agriculture).

Weeks, J.T., Koshiyama, K.Y., Maier-Greiner, U., Sch a ¨ effner, T., and Anderson, O.D. (2000). Wheat transformation using cyanamide as a new selective agent. Crop Sci *40*, 1749-1754.

Wemmie, J.A., Wu, A.L., Harshman, K.D., Parker, C.S., and Moye-Rowley, W.S. (1994). Transcriptional activation mediated by the yeast AP-1 protein is required for normal cadmium tolerance. J Biol Chem *269*, 14690-14697.

White, R., Dickinson, J., Semple, C., Powell, D., and Berry, C. (2011). The retroviral proteinase active site and the N-terminus of Ddi1 are required for repression of protein secretion. FEBS Lett *585*, 139-142.

White, T.E., Brandariz-Nuñez, A., Carlos Valle-Casuso, J., Amie, S., Nguyen, L., Kim, B., Brojatsch, J., and Diaz-Griffero, F. (2013). Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. Virology *436*, 81-90.

Wightman, R., and Meacock, P.A. (2003). The THI5 gene family of Saccharomyces cerevisiae: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. Microbiology *149*, 1447-1460.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallographica Section D *67*, 235-242.

Wörsdörfer, B., Lingaraju, M., Yennawar, N.H., Boal, A.K., Krebs, C., Bollinger Jr, J.M., and Pandelia, M.E. (2013). Organophosphonate-degrading PhnZ reveals an emerging family of HD domain mixed-valent diiron oxygenases. Proc Natl Acad Sci USA *110*, 18874-18879.

Xiao, W., Chow, B.L., Hanna, M., and Doetsch, P.W. (2001). Deletion of the MAG1 DNA glycosylase gene suppresses alkylation-induced killing and mutagenesis in yeast cells lacking AP endonucleases. Mutat Res-DNA Repair *487*, 137-147.

Xiao, W., Penugonde, V., and Rank, G.H. (1994). The MAG1, 3-methyladenine DNA glycosylase gene is closely linked to the SPT15 TATA-Binding TFIID gene on chromosome V-R In Saccharomyces cerevisiae. Yeast (Chichester, England) *10*, 687-691.

Xing, G., Barr, E.W., Diao, Y., Hoffart, L.M., Prabhu, K.S., Arner, R.J., Reddy, C.C., Krebs, C., and Bollinger Jr, J.M. (2006a). Oxygen activation by a mixed-valent, diiron(II/III) cluster in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. Biochemistry *45*, 5402-5412.

Xing, G., Diao, Y., Hoffart, L.M., Barr, E.W., Prabhu, K.S., Arner, R.J., Reddy, C.C., Krebs, C., and Bollinger Jr, J.M. (2006b). Evidence for C-H cleavage by an iron-superoxide complex in the glycol cleavage reaction catalyzed by myo-inositol oxygenase. Proc Natl Acad Sci USA *103*, 6130-6135.

Xing, G., Hoffart, L.M., Diao, Y., Prabhu, K.S., Arner, R.J., Reddy, C.C., Krebs, C., and Bollinger Jr, J.M. (2006c). A coupled dinuclear iron cluster that is perturbed by substrate binding in myo-inositol oxygenase. Biochemistry *45*, 5393-5401.

Xu, H., Faber, C., Uchiki, T., Racca, J., and Dealwis, C. (2006). Structures of eukaryotic ribonucleotide reductase I define gemcitabine diphosphate binding and subunit assembly. Proc Natl Acad Sci USA *103*, 4028-4033.

Yakunin, A.F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C.H., and Edwards, A.M. (2004). The HD domain of the Escherichia coli tRNA nucleotidyltransferase has 2 ',3 '-cyclic phosphodiesterase, 2 '-nucleotidase, and phosphatase activities. J Biol Chem *279*, 36819-36827.

Yamaki, T., Oikawa, T., Ito, K., and Nakamura, T. (1997). Cloning and sequencing of a nitrile hydratase gene from Pseudonocardia thermophila JCM3095. J Ferment Bioeng *83*, 474-477.

Yan, J., Kaur, S., DeLucia, M., Hao, C., Mehrens, J., Wang, C., Golczak, M., Palczewski, K., Gronenborn, A.M., Ahn, J., *et al.* (2013). Tetramerization of SAMHD1 is required for biological activity and inhibition of HIV infection. J Biol Chem *288*, 10406-10417.

Yang, Y., Gordenin, D.A., and Resnick, M.A. (2010). A single-strand specific lesion drives MMS-induced hyper-mutability at a double-strand break in yeast. DNA Repair *9*, 914-921.

Yao, R., Zhang, Z., An, X., Bucci, B., Perlstein, D.L., Stubbe, J., and Huang, M. (2003). Subcellular localization of yeast ribonucleotide reductase regulated by the DNA replication and damage checkpoint pathways. Proc Natl Acad Sci USA *100*, 6628-6633.

Young, H.H., and Luce, S.B. (1967). Reaction products of protein with cyanamide under acidic conditions (Google Patents).

Zhang, M., Zhang, C., Li, J., Hanna, M., Zhang, X., Dai, H., and Xiao, W. (2011). Inactivation of YAP1 enhances sensitivity of the yeast RNR3-lacZ genotoxicity testing system to a broad range of DNA-damaging agents. Toxicol Sci *120*, 310-321.

Zhang, X., Zhong, W., and Widholm, J. (2005). Expression of a fungal cyanamide hydratase in transgenic soybean detoxifies cyanamide in tissue culture and in planta to provide cyanamide resistance. J Plant Physiol *162*, 1064-1073.

Zhang, Y., Pohlmann, E.L., Serate, J., Conrad, M.C., and Roberts, G.P. (2010). Mutagenesis and functional characterization of the four domains of GlnD, a bifunctional nitrogen sensor protein. J Bacteriol *192*, 2711-2721.

Zhou, B.B., and Elledge, S.J. (2000). The DNA damage response: putting checkpoints in perspective. Nature *408*, 433-439.

Zhou, Z., and Elledge, S.J. (1992). Isolation of crt mutants constitutive for transcription of the DNA damage inducible gene RNR3 in Saccharomyces cerevisiae. Genetics *131*, 851-866.

Zhou, Z., and Elledge, S.J. (1993). DUN1 encodes a protein kinase that controls the DNA damage response in yeast. Cell *75*, 1119-1127.

Zhu, C., Gao, W., Zhao, K., Qin, X., Zhang, Y., Peng, X., Zhang, L., Dong, Y., Zhang, W., Li, P., *et al.* (2013). Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. Nat Commun *4*, 2722.

Zhu, Y., and Xiao, W. (1998). Differential regulation of two closely clustered yeast genes, MAG1 and DDI1, by cell-cycle checkpoints. Nucleic Acids Res *26*, 5402-5408.

Zhu, Y., and Xiao, W. (2001). Two alternative cell cycle checkpoint pathways differentially control DNA damage-dependent induction of MAG1 and DDI1 expression in yeast. Mol Genet Genomics *266*, 436-444.

Zimmerman, M.D., Proudfoot, M., Yakunin, A., and Minor, W. (2008). Structural insight into the mechanism of substrate specificity and catalytic activity of an HD-domain phosphohydrolase: The 5'-deoxyribonucleotidase YfbR from Escherichia coli. J Mol Biol *378*, 215-226.