Fall 12-7-2012

Structure Based Study of CA SPASE-3 and D-Arginine Dehydrogenase

Guoxing Fu
Georgia State University

STRUCTURE BASED STUDY OF CASPASE-3 AND D-ARGININE DEHYDROGENASE

by

GUOXING FU

Under the Direction of Irene T. Weber

ABSTRACT

Caspases are important players in programmed cell death. Normal activities of caspases are critical for the cell life cycle and dysfunction of caspases may lead to the development of cancer and neurodegenerative diseases. They have become a popular target for drug design against abnormal cell death. In this study, the recognition of P5 position in substrates by caspase-3, -6 and -7 has been investigated by kinetics, modeling and crystallography. Crystal structures of caspase-3 and -7 in complexes with substrate analog inhibitor Ac-LDESD-CHO have been determined at resolutions of 1.61 and 2.45 Å, respectively, while a model of caspase-6/LDESD is constructed. Enzymatic study and structural analysis have revealed that Caspase-3 and -6 recognize P5 in pentapeptides, while caspase-7 lacks P5-binding residues.

D-arginine dehydrogenase catalyzes the flavin-dependent oxidative deamination of D-amino acids to the corresponding imino acids and ammonia. The X-ray crystal structures of DADH and its complexes with several imino acids were determined at 1.03-1.30 Å resolution. The DADH crystal structure comprises a product-free conformation and a product-bound conformation. A flexible loop near the active site forms an “active site lid” and may play an essential role in substrate recognition. The DADH Glu87

forms an ionic interaction with the side chain of iminoarginine, suggesting its importance for DADH preference of positively charged D-amino acids. Comparison of the kinetic data of DADH activity on different D-amino acids and the crystal structures demonstrated that this enzyme is characterized by relatively broad substrate specificity, being able to oxidize positively charged and large hydrophobic D-amino acids bound within a flask-like cavity.

Understanding biology at the system level has gained much more attention recently due to the rapid development in genome sequencing and high-throughput measurements. Current simulation methods include deterministic method and stochastic method. Both have their own advantages and disadvantages. Our group has developed a deterministic-stochastic crossover algorithm for simulating biochemical networks. Simulation studies have been performed on biological systems like auto-regulatory gene network and glycolysis system. The new system retains the high efficiency of deterministic method while still reflects the random fluctuations at lower concentration.

INDEX WORDS: Apoptosis, Caspases, D-arginine dehydrogenase, Conformational change, Substrate specificity, Systems biology, Stochastic simulation, Biochemical networks

STRUCTURE BASED STUDY OF CASPASE-3 AND D-ARGININE DEHYDROGENASE

by

GUOXING FU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2012

STRUCTURE BASED STUDY OF CASPASE-3 AND D-ARGININE DEHYDROGENASE

by

GUOXING FU

Committee Chair: Irene T. Weber

Committee: Robert W. Harrison

Chung-Dar Lu

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2012

ACKNOWLEDGEMENTS

The person I would like to thank the most is my dear advisor Dr. Irene Weber, who brought me into the fantastic world of structural biology. This dissertation would not have been possible without her guidance, advice, and encouragement during the past a few years. I have learned how to read, write, and perform scientific research from her; but most importantly, she has taught me how to think and solve problems as a scientist. Besides her fully support in the research, she also cares about the students' personal life and encourages the students to develop their own interests. Not only a great mentor, she is also a great friend to the students.

I would like to acknowledge my Committee members: Dr. Robert Harrison and Dr. Chung-Dar Lu. Dr. Harrison is my advisor in computational biology. Whenever I get stuck in the experiment, he can always inspire me with brilliant ideas and constructive advices. Without his guidance and help, I will not be able to finish my bioinformatics project succefully. Dr. Lu has given me a lot of suggestions and guidance in molecular cell biology. I am truly thankful to him making his lab resources always available to me.

I give special thanks to Dr. Johnson Agniswamy, Dr. Bin Fang and Yuanfang Wang who taught me protein purification and crystallography. Thanks also go to Dr. Tracy Tie, Dr. Ping Liu, Dr. Fengling Liu, Brian Shen, and Ying Zhang who gave me sincere advices and help for research and presentations. I would also like to thank all the faculties, staff and colleagues in biology department.

I thank the staff at the SER-CAT beamline at Argonne National Laboratory for their assistance during X-ray data collection. I give special thanks to Dr. Albert Fu who helped me a lot for solving the structure of D-arginine dehydrogenase. My study was supported in part by the Georgia Research Alliance, the Georgia Cancer Coalition, the National Institute of Health grants and Molecular Basis of Disease Program at Georgia State University.

Finally, I would love to give my whole-heart thankfulness to my father Desheng Fu, my mother Yulan Hou, and my wife Hongmei Zhang. Hongmei is not only my greatest life-partner; she is always supportive and has walked through every step in my Phd program with me. Without their unconditional love and support, I would never be able to make this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Å	angstrom
ALA	alanine
Asp	aspartic acid
Asn	asparagines
Ac	acetyl
Cα	alpha carbon
cDNA	complementary deoxyribonucleic acid
CHO	aldehyde
CYS	cysteine
C-terminal	carboxyl terminal
Da	dalton
DAAD	D-amino acid dehydrogenase
DAAO	D-amino acid oxidase
DADH	D-arginine dehydrogenase
DDT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
ESI	electrospray ionization
FAD	flavin adenine dinucleotide
Glu	glutamic acid
Gln	glutamine
Gly	glycine
hDAAO	human D-amino acid oxidase
His	histidine
Ile	isolucine

L	liter
LAAO	L-amino acid oxidase
LADH	L-arginine dehydrogenase
LB	Luria-Bertani
Leu	leucine
Lys	lysine
kDa	kilodaltons
kLeu	4-methyl-2-oxopentanoic acid
kVal	3-methyl-2-oxobutyric acid
MAD	multiplewavelength anomalous diffraction
Met	methionine
mM	millimolar
MW	molecular weight
NMR	nuclear magnetic resonance
N-terminal	amino terminal
Pro	proline
PAGE	polyacrylamide gel electrophoresis
pDAAO	porcine D-amino acid oxidase
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PMS	phenazine methosulfate
pNA	p-nitroanilide
Phe	phenylalanine
RMSD	root mean square deviation
SAD	singlewavelength anomalous diffraction

SDS	sodium dodecyl sulphate
Ser	serine
Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine

1 INTRODUCTION

1.1 Substrate Specificity of Executioner Caspases

1.1.1 Apoptosis

The programmed cell death, also known as apoptosis, is a critical event in the life cycle of all living organisms [1]. It is a highly regulated process important in normal physiological activities as well as in many diseases [2]. Apoptosis will lead to the cell morphological and biochemical changes and eventually cell death. Changes including blebbing, membrane detachment, nuclear fragmentation, and chromatin condensation are normally seen during the development of apoptosis. The cell fragments produced by apoptosis are quickly degraded and removed before further damages occur [3]. Regular apoptosis is a non-harmful activity and is required in a healthy organism. For example, during the development of a human embryo, apoptosis occurs to the tissue cells between the fingers and triggers the differentiation of fingers and toes. Other factors such as virus infection, toxic materials, and radiation can also induce apoptosis.

The occurrence of apoptosis can be triggered by a diverse range of signals, which could be intracellular (intrinsic pathway) or extracellular (extrinsic pathway). The execution of apoptosis is shown to be carried out by the caspase family, which will be further discussed in the following sections. The intrinsic pathway is activated in response to cell stresses, such as DNA damage and loss of cell survival factors. The Bcl-2 protein family plays important roles in the regulation of the mitochondria-initiated intrinsic apoptosis [4]. More than twenty members of the Bcl-2 family have been identified. The Bcl-2 family members are grouped into three subfamilies according to their sequence and function similarity [5]. The anti-apoptosis subfamily includes Bcl-2 and Bcl-xL in mammals and CED-9 in nematodes. Bcl-2 and Bcl-xL prevent the mitochondrial proteins from being released. CED-9 inhibits apoptosis through preventing CED-4 from activating CED-3 [4]. The pro-apoptotic Bcl-2 members are grouped into multi-domain pro-apoptotic subfamily including Bax and Bak, and BH3-only pro-apoptotic subfamily [6]. Bax

and Bak exist as cytosolic monomers in healthy cells and will oligomerize upon receiving the death signal. It is suggested that these oligomers will form pores on the outer mitochondrial membrane, thus allowing the release of mitochondrial proteins like cytochrome c and SMAC/DIABLO into the cytosol [7]. Cytochrome c binds the adaptor apoptotic protease activating factor-1 (Apaf-1) to facilitate the formation of apoptosome, which recruits and activates caspase-9. Caspase-9 then activates the downstream executioner caspases to carry out the apoptosis.

The extrinsic apoptosis pathway is initiated by extracellular signals, such as hormones, toxins, nitric oxide, growth factors and cytokines [4]. Extracellular molecules such as Apo2L/TRAIL and CD95L/FasL are also known as pro-apoptotic ligands [8]. These pro-apoptotic ligands bind to the death receptors on the cell surface. The death receptors are characterized by extracellular cysteine-rich domains (CRDs) and belong to the tumor necrosis factor (TNF) family of proteins [9]. Members of the well-studied death receptors include TNFR1 (p55 or CD120a), Fas (Apo1 or CD95), and TRAIL (R1/R2) [4]. A death domain (DD) is found in the intracellular domain of all death receptors. The death domain is suggested to be responsible for the binding of adapter proteins such as Fas-associated death domain (FADD) during the activation of initiator caspases. The activation of caspase-8 is a result of a cascading oligomerization. The activated death ligands exist as homotrimers, thus leading to the oligomerization of death receptors upon binding. Adaptor proteins and other effector proteins are then recruited by the death receptors. The adaptor proteins like FADD contain death effecter domains (DED) that interact with the DED of procaspase-8 and trigger the oligomerization of procaspase-8 [10]. Therefore, the multi-component death-inducing signaling complex (DISC) is formed at the plasma membrane. Procaspase-8 molecules are activated by cleaving each other when they are brought together, while only weak proteolytic activity is detected for single monomers of procaspase-8 [10]. The active caspase-8 then starts to process downstream executioner procaspases, including procaspase-3, -6 and -7.

Abnormal inhibition of apoptosis results in uncontrolled cell proliferation, which leads to cancer and autoimmune disease [11]. In contrast, excessive cell death has been identified in a number of neurodegenerative disorders such as Alzheimer's disease [12]. Current treatments for cancer include surgery,

radiation, chemotherapy, hormone therapy, biological therapy, and targeted therapy [13]. Most of these approaches kill cancer cells by inducing apoptosis. However, cancer cells often develop resistance to these therapies. Many cancer therapies introduce chemical or physical damages to DNA, thus increasing the chances of developing mutated cancer cells with resistance [6]. Therefore, knowledge of how apoptotic pathways are regulated in cancer is critical for understanding the development of the disease and designing effective and reliable anti-cancer treatments. Novel therapies that are both safer and more effective could be developed through directly activating or suppressing apoptosis by regulating the caspase activity.

1.1.2 Caspase Family

Caspases are cysteinyl aspartate proteinases that play important roles in the regulation of cell life cycle [14]. The first member of caspase family was initially identified as interleukin-1β-converting enzyme (ICE, now named caspase-1), which is involved in the maturation of IL1β [15]. Caspase-2 was found as the product of the *Caenorhabditis elegans* cell death gene ced-3 in 1993 [16]. Many caspase members have been identified in different organisms, including eleven from human, ten from mouse, four from chicken and four from zebrafish [17]. Caspases are classified into three groups according to their functions and structures. Some of the caspase members like caspase-1 do not have a role in apoptosis execution, whereas some caspases are functional in both apoptotic and non-apoptotic signaling. Caspase-1, -4, -5, -11 and -12 are involved in inflammation. Caspase-2, -8, -9 and -10 are initiators while caspase-3, -6 and -7 are executioners of apoptosis (**Figure 1.1**) [6, 18].

In healthy cells, caspases normally exist as inactive precursors called procaspases or caspase zymogens. Procaspases are shown to have limited catalytic activity and they are under critical regulation by various regulatory molecules [19]. They are composed of three domains: a large subunit (~20kDa, p20), a small subunit (~10kDa, p10) and an N-terminal prodomain with various lengths (**Figure 1.2a**). The long prodomain contains different protein-protein interaction domains that play important roles in caspase activation, such as the caspase recruitment domain (CARD) in caspase-2 and -9, and death effector domains (DEDs) as in caspase-8 and -10. These are known as initiators and their prodomains recruit the

procaspases to the death-inducing signaling complex (DISC), while oligomerization leads to their auto-catalytic activation, known as "proximity-induced" activation [18, 20]. The executioner caspases, caspase-3, -6, and -7, do not have a long prodomain and could not perform self-activation. Therefore, the downstream caspases are activated through cleavage by active initiator caspases. Procaspases are activated through removal of the prodomain and cleavage of the linker region between the p20 and p10. One molecule of p20 and one molecule of p10 will form an active heterodimer. Two heterodimers will form a heterotetramer as the mature caspase under physiological conditions (**Figure 1.2a**). The 3D structure of caspase-3 heterotetramer is shown in Figure 1.2b. It consists of two large subunits and two small subunits which are indicated by different colors in the figure.

In mammals, the apoptotic caspases have been classified as initiator caspases (caspases-2, -8, -9, and -10) and effector caspases (caspases-3, -6, and -7) based on their structures and functions. The activation of initiator caspases results from cell death signaling through either the extrinsic (death receptor) pathway or the intrinsic (mitochondrial) pathway (**Figure 1.3**). Caspase-8 is an essential component of the extrinsic cell death pathways initiated by the TNF family members [19]. The extrinsic pathway initiates when the death ligand binds to the cell surface receptor, such as Fas/CD95/Apo-1, which is a member of the TNF family of apoptosis-inducing receptors. Binding of Fas and procaspase-8 to the adaptor molecule FADD leads to the formation of the death-inducing signaling complex. The aggregated procaspase-8 molecules then undergo auto-activation and activate the downstream effector caspases [21-22]. It is suggested that caspase-8 plays non-redundant roles in normal embryonic development and apoptosis mediated by the TNFR family members, while it is not critically required in other apoptotic pathways [19]. Furthermore, caspase-8 is also shown to be involved in macrophage differentiation and in the formation of yolk sac vasculature of endothelial cells [23]. Caspase-9 is regarded as the canonical caspase in the intrinsic mitochondrial pathway [18]. When the intrinsic pathway is initiated, cytochome c from the inner-mitochondrial membrane space is released into the cytosol. Binding of cytochrome c and dATP to the adaptor molecule, Apaf-1, forms a large multimeric complex called apoptosome. Procaspase-9 is then recruited to the apoptosome through the interaction of the caspase CARD domains. Procaspase-9 is acti-

vated by autocatalysis and then activates the downstream effector caspases [24]. Caspase-10 contains two DEDs like caspase-8. Caspase-2 contains one CARD domain similar to caspase-9. However, their functions in apoptosis still remain unclear [19]. It is suggested that human caspase-10 may have a function overlapping with caspase-8 in Fas ligand-mediated cell death pathway [25]. Caspase-2 was demonstrated to be required for mitochondrial outer membrane permeabilization and the release of apoptogenic factors in response to DNA damaging agents [26].

Caspase-3, -6 and -7 are the downstream effector caspases. Caspase-3 is the major executioner that cleaves the majority of the cellular substrates in apoptotic cells while caspase-6 and -7 are involved to a lesser extent [27]. Activation of caspase-3 is carried out by caspase-8 and -9 but not by caspase-2 [18]. Capase-7 is quite similar to caspase-3 and they share similar substrate specificity [18-19]. Reduced cell death and increased cell number in the central nervous system (CNS) have been observed in caspase-3 mutant mice [28]. DNA fragmentation defects are also detected in the cells derived from these animals. Caspase-3 knockout animals in C57BL/6 background showed certain viability with reduced fertility, indicating that caspase-3 is redundant for most developmental cell death [29]. If only caspase-7 was knocked out, the animals showed normal development and their cells were still sensitive to a variety of apoptotic agents [29]. However, when both caspase-3 and caspase-7 are knocked out, the animals will die rapidly after birth, suggesting that these two caspases are critical for the apoptosis execution. Caspases-3 and -7 share high sequence identity of 56% and they preferentially recognize DEVD in their substrate sequence. Caspase-6 shares 37% and 41% sequence identity with caspase-7 and -3, respectively, and it shows structural homology to caspase-3 and -7. However, it has a different substrate specificity showing preference for peptide sequence VEHD rather than DEVD preferred by caspase-3 and -7 [30]. The function of caspase-6 remains unclear. The caspase-6 knock-out mice developed normally and the cells were still sensitive to various apoptotic agents, suggesting caspase-6 is not essential for apoptosis [19]. These studies suggest that any of the effector caspases may be important for certain apoptosis pathways while others may act as redundant or compensatory executioners.

1.1.3 Caspase Substrate Specificity

Caspases are major executioners of apoptosis. Substrates of caspases have been identified in many biological pathways, such as proteins involved in framework of the cytoplasm and nucleus, transcription regulatory, and DNA replication and repair [31]. The analysis of binding preferences in caspase active site or the cleavage site sequences on their natural substrates are two different ways for investigating caspase substrates [32]. More than 280 protein substrates of caspase have been identified in cells and the discovery has been considerably accelerated by the systematic proteome analysis of apoptotic cells [31].

Most of the caspases recognize certain tetrapeptide sequences in their substrates (**Figure 1.4a**) [32]. This was initially investigated by Thornberry et al. [30] using positional scanning synthetic peptide-based libraries and Talanian et al. [33] using sets of individual peptide substrates and later refined by others. The positions on the N-terminal side of the scissile bond are called P1, P2, P3 and P4 with the corresponding binding sites called S1-S4, while the positions on the C-terminal side are called P1' and P2'. The active site groove of caspase-3 with DEVD binding in the active site is shown in **Figure 1.4b** (PDB: 2H5I) [34]. The substrate P1-P4 residues bind into the corresponding S1-S4 pockets in the active site. A requirement for aspartic acid is highly restricted at the P1 position while the residues at other positions may vary with different preferences. The substrate specificities of caspases are summarized in **Figure 1.4c**. Previous studies have showed that the substrate preference is W/YExD (x refers to any amino acid) for caspase-1, 4, 5, 14, I/LExD for caspase-8, 9, 10, DExD for caspase-3, 7, VExD for caspase-6, and DExD for caspase-2 [32]. Furthermore, the residue at the P1' position was also shown to be important for the caspase substrate specificity. Small and uncharged residues like Gly, Ala, Thr, Ser and Asn at P1' form good substrates, while bulky or charged residues like Glu, Asp, Lys, Arg, Trp are poorly tolerated [35]. The substrate recognition was extended to P5 for caspase-2 [36]. Similarly, it is reported that caspase-3 prefers pentapeptides with hydrophobic amino acids at P5 over tetrapeptides [34, 37].

Knowledge of caspase substrate specificity has been verified and refined by many investigators and has led to various novel methods to study the activity of caspases. Peptide-based small substrate re-

porters have been developed to measure the activity of caspases [38]. These reporters normally contain a caspase preferred tetrapeptide sequence with a C-terminal fluorogenic reporter, which is released to produce a fluorescent signal upon cleavage. This method has greatly facilitated the characterization of caspase substrate specificity. However, the over-simplified traditional understanding of caspase substrate specificity does not work well for prediction of caspase protein substrates [32]. Due to the simplicity of caspase preferred tetrapeptide sequence, many proteins may possess one or more similar sequences. However, this is far from sufficient requirement for the protein to be a natural caspase substrate. First of all, the cleavage site needs to be properly exposed to the caspase active site, while most of the time these potential cleavage sites are deeply embedded inside the protein 3-dimensional structures. This brings out another concept called exosite, the enzyme regions that are distant from the active site but involved in substrate binding. Some studies have showed additional interactions between caspases and their substrates. However, the mechanism still remains unclear. The crystal structure of viral caspase inhibitor p35 with caspase-8 revealed a p35 loop region interacts with a surface area of caspase-8 distant from the active site, suggesting a possible exosite interaction [39]. Moreover, the cleavage site sequences on caspase natural substrates are also different from canonical sequences. It is suggested that about 45% of reported caspase-3 natural substrates contain cleavage site sequences different from classical substrate sequences [31]. Therefore, the presence of a preferred cleavage site sequence on a protein does not always guarantee it is a real caspase cleavage site *in vivo*.

As caspases play essential roles in the regulation of apoptosis, their substrate specificities have been intensively studied by many research groups over a decade. The previous research results helped the discovery of various natural caspase substrates. Determination of these canonical substrate sequences helped us to better understand the caspase substrate specificity. However, many aspects of the caspase substrate specificities still remain elusive. We are studying the substrate binding of caspases by applying crystallography, kinetic study and computational modeling in order to provide valuable information for future studies of caspase substrate specificity. The recognition of P5 position in substrates by caspase-3, -6 and -7 has been investigated in this study [37]. In caspase-3, the enzyme residues Ser209, Phe250 and

Phe252 interact with the P5 residue in substrate. It showed 50% increased hydrolysis of LDEVD relative to DEVD in enzymatic experiment. A model of caspase-6 indicated its preference for polar P5 in QDEVD likely due to interactions with polar Lys265 and hydrophobic Phe263 in loop-4. Unlike the other two executioner caspases, Caspase-7 exhibited no preference for P5 residue in agreement with the absence of P5 interactions in the caspase-7/LDESD crystal structure. These findings will provide valuable information for better understanding the different roles of the executioner caspases. The differences identified in P5 recognition are particularly important for caspase-3 and -7, which showed very similar preferences for tetrapeptide sequence in substrate. Such knowledge can be applied in the design of selective inhibitors for caspases, which can be potentially used in the treatment of diseases caused by increased activities of caspases.

1.2 Structural Analysis of D-Arginine Dehydrogenase

1.2.1 Flavoenzymes

Flavoenzymes play important roles in a large variety of biological processes. They typically contain a molecule of flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) as a cofactor. In most flavoenzymes, the cofactor is non-covalently bound to the apoprotein, with a few exceptions that the cofactor forms covalent bond with a Histidine, Cysteine or Tyrosine of the protein [40-41]. The cofactor FAD is composed of one molecule of adenosine monophosphate (AMP) and one flavin mononucleotide (FMN), which are linked together by a pyrophosphate bond (**Figure 1.5a**). The FMN moiety contains an isoalloxazine ring, which is the catalytic center of FAD, while the other groups mainly contribute to the binding of the cofactors to the protein. The isoalloxazine ring contains a xylene group, which is hydrophobic and forms better interaction with hydrophobic protein residues, and a pyrimidine ring, which is hydrophilic and is easier to form hydrogen bonds [42]. The flavin can transfer one or two elections from the substrate to another substrate or an electron acceptor. Therefore, three different redox states for flavin have been identified: oxidized, semiquinoid, and fully reduced (**Figure 1.5b**). Furthermore, the catalytic activities of the flavin are also affected by the surrounding protein residues. For example, protein positive

charges have been identified in flavoprotein oxidases such as glucose oxidase and choline oxidase. The positive charges are involved in the electrostatically stabilization of the oxygen activation in flavoprotein oxidases [43]. Other protein structures affecting the flavin's oxidation-reduction potential are also reported, such as the Ala113 in L-galactono-γ-lactone dehydrogenase, which is suggested to block the oxygen access to FAD. Mutation of this Ala to a Gly has increased the reactivity of the reduced flavin with oxygen by about 400-fold [44]. A similar structure has been observed for D-arginine dehydrogenase Ala46, which will be discussed in details in the following chapters [45].

Flavoenzymes are involved in a wide range of biological processes. Their functions are not only limited to electron-transfer reactions, but extended to energy production, light sensing and emission, neural development and protein folding (**Figure 1.6**) [41, 46]. The cofactor is reduced by the enzyme substrate and then oxidized by an electron acceptor. The flavoenzymes are characterized into different groups based on the type of their electron acceptors. Those convert oxygen to hydrogen peroxide are named flavoprotein oxidases, such as L-amino acid oxidase and D-amino acid oxidase [47-48]. Those employ other electron acceptors are called flavoprotein dehydrogenases, such as Cytokinin dehydrogenase and L-proline dehydrogenase [49-50]. Another group of flavoenzymes are known as monooxygenases, which introduce a single oxygen atom into the substrate and reduce the other oxygen atom to water [41]. A transient C4a-hydroperoxide intermediate is formed during the activation of molecular oxygen in the monooxygenases. Several members of this family have been identified in the past few years, including Baeyer-Villiger monooxygenases, flavin-containing monooxygenases and N-hydroxylating enzymes [51]. A great variety of substrate specificity has been observed for flavoenzymes. Some have very narrow substrate specificity, while others have a broader selection of susbtrates. For example, glycine oxidase is able to react with various amines and D-amino acids in thiamine biosynthesis [52]. On the contrary, L-tryptophan oxidase, which is involved in the biosynthesis of the antitumor compound rebeccamycin, strictly prefers 7-chloro-L-tryptophan [53].

Flavoenzymes have been the subject of many studies for more than 50 years [54]. Hundreds of members have been identified in prokaryotic and eukaryotic organisms, for which about 1-3% of their

genes are suggested to encode flavin-containing proteins [46, 54]. The enzymological and structural properties of many flavoenzymes have been intensively studied and characterized, while certain aspects of the reaction mechanism still remain unclear. Although they are identified in different organisms and involved in various biological processes, similarities have been found in sequences and structures of flavoenzymes. So far, four different FAD-binding folds have been defined for flavoenzymes [55]. They are represented by glutathione reductase (GR), ferredoxin reductase (FR), *p*-cresol methylhydroxylase (PCMH), and pyruvate oxidase (PO), respectively (**Figure 1.7**). Unique FAD-binding patterns have been identified to distinguish different flavoenzyme families. However, sequence and structure studies of flavoenzymes have revealed some conserved sequence motifs and similar 3-D structures that are essential for cofactor binding. One of the most well-known folds was first identified in NADH-binding dehydrogenases by Rossmann, and is named Rossmann fold [56]. This structure is later found to be conserved in the GR flavoenzyme family. The Rossmann fold is composed of a central five-stranded parallel β-sheet with α-helices on one side and a three-stranded antiparallel β-sheet on the other side (**Figure 1.7a**). A conserved sequence motif has been identified in the GR family members, xhxhGxGxxGxxxhxxh(x)8hxhE(D), in which x stands for any residue and h represents a hydrophobic residue [57-58]. A sequence-structure analysis of 32 nonredundant FAD-containing enzymes has revealed many structural and sequence properties that are conserved among the flavoenzymes [55]. For example, all the members of the FR family contain a cylindrical β-domain in their FAD-binding fold, while the FAD-binding domain of PCMH members all consists of two α+β subdomains (**Figure 1.7b, c**) [55, 59]. The PO family contains a double Rossmann fold similar to the FAD fold in GR family; despite the fact the two families share very little similarity in their sequences (**Figure 1.7d**) [56]. Further, the various FAD-binding domains allows FAD bind to the enzyme in different conformations, such as the elongated or bent butterfly conformation. In the elongated conformation, the adenine ring is located far from the isoalloxazine ring as seen in the structures of D-Amino acid oxidase [60] and glutathione reductase (FAD in Figure 1.7a) [61]. In the bent conformation, the adenine and isoalloxazine ring are very close to each other as observed in flavodoxin reductase and other members of the FR family (FAD in Figure 1.7b) [62]. In conclusion, the sequence and

structure analysis on the conserved motifs, cofactor binding, and cofactor conformation of flavoenzymes will improve our understanding of the reaction mechanisms and provide useful information for recognizing new flavin-containing proteins from the sequence.

1.2.2 D-Amino Acids

The standard amino acids except glycine can exist as either L- or D- optical isomers. An example of L- and D- forms of alanine is shown in **Figure 1.8**. The L-amino acids are utilized as building blocks for proteins synthesized by all living organisms. D-amino acids are important for bacteria as fundamental elements of the bacterial cell wall peptidoglycan layer, but their catabolism mechanism is not well understood [63]. Several D-amino acids are reported to be involved in disassembly of bacterial biofilms [64] and growth phase-dependent cell wall remodeling [65]. Moreover, specific D-amino acids have been discovered in many physiological processes in vertebrates, including humans [66]. Dunlop et al. identified D-aspartate in the brain and other tissues of mammals and suggested that it may play a role in regulating the development of these organs [67]. D-serine was identified at significant levels in rat brain, at about a third of the level of L-serine [68]. Further, D-serine in the rat brain is distributed in close association with N-methyl-D-aspartate (NMDA) and it may act as an endogenous agonist of the NMDA receptor in mammalian brains [68].

The catabolism of D-amino acids has not been intensively studied when compared to those of L-amino acids. D-amino acids could be metabolized either directly or after conversion into the L-enantiomers. Conversion of L- and D-amino acids in living organisms is commonly catalyzed by racemases. Various amino-acid racemases have been identified in bacteria, archaea, and eucaryotes including mammals. These racemases are classified into two groups: pyridoxal 5'-phosphate (PLP) -dependent and -independent enzymes [66]. D-serine racemases and D-aspartate racemases are intensively studied in mammals due to their involvement in cell aging and neural signaling [69-70]. In bacteria, D-amino acids are deaminated by flavin-dependent oxidases or dehydrogenase, allowing the bacteria to grow using D-amino acids as the nutrient and energy source [71-73].

Many flavoenzymes have been identified in the catabolism pathway of D-amino acids. D-amino acid oxidase has been the most extensively studied member since its first discovery in 1930s [48]. This enzyme catalyzes the dehydrogenation of D-amino acids into their corresponding imino acids. The imino acids are converted to keto acids and ammonium through a non-enzymatical reaction, while the reduced enzyme cofactor FAD is oxidized by molecule oxygen. The structure of DAAO is composed of two domains, a FAD-binding domain and a substrate-binding domain (or interface domain) (**Figure 1.9a**) [74]. The cofactor FAD adopts the elongated conformation with the flavin ring located at the interface between the two domains. The enzyme active site is identified near the *re* face of the cofactor as shown in the structures of DAAO with ligands [74-75].

DAAO is widely distributed in nature, from microorganisms to mammalian tissues, such as brain, kidney, and liver. The studies on DAAO used to be difficult due to the low content in tissues and low stability of this enzyme. However, investigation in DAAO has increased dramatically since mid 1990s because of the importance of its physiological roles and development of genetic engineering which allows producing the enzyme in large quantity [76-77]. DAAO is able to catalyze the deamination of many D-amino acids. Therefore, it is involved in many physiological processes. In microorganisms, DAAO can provide D-amino acids as a source of carbon, nitrogen, and energy [78]. In mammalian cells, DAAO plays important roles in many processes such as neutral signaling and hormone secretion by regulating level of D-amino acids. One of the most important functions of DAAO is its regulation of D-serine, which is known to be an endogenous agonist of NMDA-receptors in mammalian brains [68]. Increased DAAO activity will lead to decreased level of D-serine, hence lowering the activity of NMDA-receptors. It is suggested that this regulation is related to the development of schizophrenia [76]. Moreover, DAAO is involved in the process of hormone secretion due to its regulation of D-Aspartate [77]. DAAO also regulates the levels of D-proline and D-alanine. The level of D-oxyproline has been suggested to be related to aging [79]. D-alanine is found to be important for intracellular osmotic pressure and increased level of D-alanine has been reported in the gray matter of Alzheimer's patient [80-81].

One of the major characteristics of DAAO is its high specificity to D-amino acids, while it is almost inactive to the corresponding L-amino acids. The deamination of L-amino acids is catalyzed by another flavin-dependent oxidoreductase, L-Amino acid oxidase (LAAO), which accepts most of the 20 L-amino acids as substrates but does not react with D-amino acids [47]. The structures of LAAO from *Rhodococcus opacus* (roLAAO) with different ligands have been determined (**Figure 1.9b**) [47]. The overall topology of roLAAO is very similar to that of DAAO from pig kidney, although they share very low sequence identity (15.3%). The two enzymes are both composed of a substrate-binding domain and a FAD-binding domain except that LAAO contains an extra helical domain (colored yellow in Figure 1.9b), which is involved in the dimerisation of the enzyme [47]. How these enzymes precisely recognize the stereospecificity in their substrates has become an interesting topic. Detailed comparison of the structures of DAAO and LAAO has revealed a mirror-symmetrical relationship through the plane perpendicular to the isoalloxazine ring between their active sites, which will be responsible for their different enantiomeric selectivity [45]. Meanwhile, similar composition and arrangement of the active site residues have been observed between the D- and L- amino acid oxidases, which may explain the similarities in their functions. Moreover, a different mode of mirror-symmetrical relationship for opposite stereospecificity has been observed between DAAO and flavocytochrome b2, which oxidizes L-lactate to pyruvate [82]. The active sites of the two enzymes are mirrored through the plane of isoalloxazine ring. The ligand binds to the *re* side of FAD in pDAAO, while it binds to the *si* side of FAD in flavocytochrome b2 [74].

1.2.3 D-Arginine Dehydrogenase

Endogenous D-amino acids are normally converted to L-amino acids by racemases [66]. In 2009, Li et. al reported a novel 2-component amino acid racemase involved in D-to-L inversion in *Pseudomonas aeruginosa*, which is an opportunistic human pathogen [83]. A *dauBAR* operon was found to be highly induced by D-arginine from the DNA microarray analysis. Mutagenesis study showed that mutations at *dauA* or *dauB* prevented the bacteria from using D-arginine as sole carbon source. Further studies have identified two coupled dehydrogenases: D-arginine dehydrogenase, which is encoded by *dauA* and cata-

lyzes the conversion of D-arginine into 2-ketoarginine and ammonia, and L-arginine dehydrogenase, which is encoded by *dauA* and converts 2-ketoarginine to L-arginine (**Figure 1.10**) [83]. Enzymatic studies on 19 D-amino acids have revealed that DADH has broad substrate specificity, with D-arginine and D-lysine as the two best substrates [45, 84].

In order to understand the reaction mechanism and substrate specificity of DADH, the crystal structure of DADH was determined at the atomic resolution of 1.06 Å, while the structures of DADH crystallized in the presence of iminoarginine, iminohistidine, iminolysine, iminomethionine, iminophenylalanine and iminoproline were also determined at the resolutions of 1.03 Å to 1.30 Å. Well-defined electron densities for the non-covalently bound FAD and imino intermediates of the reaction allowed detailed analysis of the enzyme active site. Alternative conformations have been observed for a loop region near the DADH active site, suggesting its involvement in substrate binding and product release. Iminoarginine and iminohistidine bind to the active site in distinct modes, which is in agreement with detailed kinetic analysis on substrate specificity reported previously [84]. The structural characteristics described here will provide insights into substrate recognition and the catalytic reaction mechanism of DADH.

DADH contains one molecule of FAD as a cofactor and is composed of a substrate-binding domain and a FAD-binding domain, while the active site is located at the interface of the two domains (**Figure 1.11a**). Sequence of this enzyme showed that it shares very low sequence identity (<20%) with other FAD-containing enzymes such as L-proline dehydrogenase (18.5% [50]), DAAO (17.2%, [74]), and LAAO (16.4%, [47]). However, the three-dimensional structure of DADH highly resembles those of other flavoenzymes (see detailed comparison in chapter 3). The FAD-binding domain of DADH adopts the rossmann fold that is predominant in the GR flavoenzyme family as discussed in above. It consists of a central six-stranded β-sheet surrounded by five α-helixes on one side and a three-stranded anti-parallel β-sheet with two α-helixes on the other side (**Figure 1.11b**). The substrate binding domain is composed of an eight-stranded β-sheet and two short antiparallel β-strands forming a sandwich surrounded by four α-helices.

Examination of the electron density map demonstrated that the reaction product imino acids were captured in the active sites of the determined crystal structures. Further comparison of structures of the ligand-free DADH and its complexes with imino acids has revealed that the protein active site undergoes major conformational change upon binding of the ligand. It possesses a ligand-free conformation in the free enzyme but shows a product-bound conformation upon binding of the ligand. Residues 50-56 were designated as an active site lid controlling the substrate accessibility to the active site. This active site lid structure has been reported in many flavoenzymes, such as D-amino acid oxidase (residues 216-228 [48]), pyranose 2-oxidase (residues 454-457 [85]), and choline oxidase (residues 64-95 [86]) (**Figure 1.12**). The loop composed of residues 216-228 in DAAO have been proposed to act as a lid that is able to switch from closed and open conformation to control substrate binding and product release [74]. This hypothesis is supported by the experiment that proteolysis of the peptide bond between loop residues 221-222 is prevented by inhibitor binding, suggesting a conformational alternation upon binding of a ligand [87]. The closed conformation of DAAO active site lid has been observed in several structures of DAAO in complexes of different ligands [74-75]. However, the open conformation has never been captured in crystal structure based on our knowledge. The ligand-free and product-bound conformations observed in the DADH helped confirm the hypothesis about this active site lid. Not only it is involved in ligand binding and releasing, it may also play a role in the substrate specificity of DADH [45].

DADH is characterized by broad substrate specificity, being able to oxidize basic and hydrophobic D-amino acids of various sizes, but not reacting with acidic D-amino acids [45, 84]. D-amino acids are converted enzymatically to imino acids, which then dissociate from the enzyme and are hydrolyzed to keto acids in a nonenzymatic reaction. Steady-state kinetic studies with D-arginine or D-histidine as substrate have established a ping-pong bi-bi kinetic mechanism for this enzyme [88]. A product release experiment with D-arginine indicated that the release of imino product is partially rate limiting for the overall turnover of the enzyme, which is in agreement with the observation that imino acids were trapped in the crystal structure. DADH has been suggested to be a true dehydrogenase due to its lack of oxygen reactivity [88]. Structural analysis suggested that the access of O_2 to the flavin C4a atom is blocked by the

Ala46 and the absence of positive charge proximal to the flavin C4a and N1-C2 atoms is also responsible for its poor oxygen reactivity. A mutation of DAAO Gly52, which is located at the corresponding position of DADH Ala46, to Val leads to the loss of DAAO reactivity with oxygen due to the steric hindrance blocking the oxygen access [89]. Furthermore, substitution of the structurally equivalent Ala113 with Gly in L-galactono-γ-lactone dehydrogenase increased the reactivity of the reduced flavin with oxygen by about 400-fold [44].

The structural studies on DADH presented here may facilitate knowledge of the mechanism of catabolism of D-amino acids *in vivo*. The conformational change at the active site lid has been clearly shown in the atomic resolution structure of DADH, suggesting a role in controlling the substrate binding and product releasing. This active site lid has been proposed in other enzymes such as DAAO but with no direct evidence. The ionic interaction of iminoarginine with Glu87 in DADH is suggested to be critical for the enzyme specificity for basic D-amino acids. This is supported by the kinetic study showing that D-arginine and D-lysine are the two best substrates and the k_{cat}/K_{Arg} value is about 1000-fold higher than the k_{cat}/K_{His} value. Moreover, the structural analysis of DADH in complexes with hydrophobic imino acids revealed the hydrophobic residues responsible for the binding of bulky hydrophobic substrates. Further comparison of DADH with other functional related enzymes like DAAO and L-amino acids oxidase has revealed some key components and structural arrangement that are responsible for their specificity toward different enantiomers and different amino acids [45]. Detailed structural comparison of DADH with D-amino acid oxidase and D-aspartate oxidase indicated that although they share active site similarities in terms of recognition of D-amino acids, different arrangement and composition of the key structural features determine very distinct substrate specificities. Overall, the high-resolution structures of DADH will provide useful information for future studies of similar flavin-dependent enzymes.

1.3 A Deterministic-Stochastic Crossover Algorithm

1.3.1 Systems Biology

Systems biology is an interdisciplinary science of experimental and computational research studying biological functions and mechanisms [90]. Rapid development in molecular biology especially for genome sequencing and high-throughput measurements has greatly facilitated the advancement of systems biology [91]. Meanwhile, significant advances in computational software and hardware allowed the creation of models and investigation by *in silico* simulation of complex and still realistic biological systems. Systems biology studies focus on the understanding of the whole system's structure and dynamics, while the traditional molecular biology concentrates more on the properties and functions of individual genes and proteins. Although many issues remain unsolved, systems biology studies are able to provide in-depth insights and reasonable predictions for biological networks such as the cell cycle and metabolic analysis.

Biological systems, such as cells, are considered as "complex systems", in which large numbers of various elements selectively interact with each other to produce complex behaviors. Functions of complex systems in the traditional mechanical concept are mainly determined by properties of the networks they form rather than properties of each element. However, functions in biological systems relay on a combination of the network and the specific elements involved. The protein p53 acts as a tumor suppressor when involved in a network of transcription factors. Its functions are strictly regulated by modifications like phosphorylation, dephosphorylation and proteolytic degradation, which are properties that reflect the complexity of the element itself. Neither p53 nor the network functions as a tumor suppressor when taken apart [90, 92]. Molecular biology has uncovered numerous biological facts like protein properties and genome sequences in the past a few decades. However, with the dramatically growth of information, it alone is not sufficient for interpreting the complex biological systems. Therefore, a system-level understanding of biology is greatly needed and a combination of experimental and computational approaches is expected to achieve this goal. An example of the integration of systems-biology research

with drug discovery and treatment optimization is illustrated in **Figure 1.13**. Systems biology can employ computational modeling, systems analysis, and development of experimental techniques in to the process of drug discovery and treatment optimization. Therefore, it is able to significantly reduce the cost and increase the success of drug discovery [90].

Knowledge discovery and simulation-based analysis are two different branches in computational biology. Knowledge discovery is often known as data-mining. Hypotheses are often derived from extracting the patterns from huge quantities of experimental data. Areas for which knowledge discovery has been widely used include predicting protein structure from sequence and investigating the gene regulatory networks [91, 93]. Simulation-based analysis is used to test hypotheses with *in silicon* experiments, which is then tested by *in vitro* and *in vivo* studies. Computer simulations of biological systems have been intensively developed and can now replicate important biological behavior. The phages are employed as the initial study model due to their simplicity. Several groups have studied the lambda phage 'switch' [94-95]. A successful combination of simulation and laboratory measurements has been performed on the T-7 phage [96-97]. However, eukaryote organisms especially mammalian cells, are much more complex systems and are more difficult to study by computational simulation [98]. Many efforts have been dedicated in studies of cellar processes like gene expression, cellular signaling and metabolism pathways.

1.3.2 Deterministic, Stochastic and Crossover Methods

Two major techniques used in system simulation are deterministic methods and stochastic methods. Both techniques have their own advantages and disadvantages. For a long time the most common approach was deterministic method, in which the systems are assumed to be continuous or deterministic. Their kinetics can be described by ordinary differential equations (ODEs), which are the reaction rate equations derived from the classical theory of mass action kinetics. The deterministic method works well for large-quantity biological networks and can reasonably represent the overall dynamics of the system. Since the deterministic method assumes that time evolution of a chemically reacting system is both continuous and deterministic, some behaviors of the systems associated to the discreteness of molecular

quantities and random fluctuations in systems containing species at lower concentration, are neglected. Since the molecular population of reacting species can only be changed by discrete integer amounts, the chemical reaction is not a continuous process. Moreover, it is impossible to predict whether a reaction will occur without knowing the precise positions and velocities of the reacting species, suggesting that the chemical reaction is not deterministic either. The accuracy of deterministic method is greatly compromised, especially when dealing with ecological systems, microscopic biological systems, and nonlinear systems driven to conditions of chemical instability [99].

In intracellular molecule kinetics, stochasticity is significant because some molecule species exist in very low numbers. Gene regulation can serve as a good example because only one gene is involved in most of gene expressions. It has been reported that the transcription level of a single gene in *Escherichia coli* is very low by measuring mRNA levels in individual living cells [100]. Typical numbers of some protein molecules in a cell are less than a few hundreds. Therefore, fluctuations in these biological systems are significant at low concentrations. The stochastic methods were developed to address this problem based on the knowledge that a reaction occurs when two or more reacting species collide with each other. In stochastic models, the reacting species are represented as discrete number of molecules, allowing the dynamics of the whole system to be updated discretely rather than continuously. The stochastic simulation has been widely used since its first description in 1977 [99, 101]. Efforts in accelerating simulations are desired because stochastic simulation is computationally costly and it requires a significant amount of computer time when dealing with large models. However, accelerated simulation usually comes at the price of approximation and thus loss of exactness. Another disadvantage of stochastic method is that it can only take into account elementary reactions rather than more complex reactions such as Michaelis-Menten type reactions. These complex reactions have to be broken into elementary reactions before they can be simulated using the stochastic method. However, this is not easily achievable due to the difficulties of experimental determination of the rate constants for those elementary reactions. Therefore, the stochastic method is hindered from being widely used for complex systems.

Many studies have been dedicated to improving the computational efficiency of stochastic method, while some groups have developed hybrid algorithms to facilitate the simulation by partitioning the system into “fast” and “slow” reactions [101-103]. Our group has developed a novel deterministic-stochastic crossover method in order to improve the computational accuracy and efficiency of simulation for biochemical systems [104]. This crossover method works by introducing a certain degree of randomness into a deterministic simulation. First of all, each reaction in the system is treated as deterministic and the law of mass action can be applied to obtain ODEs. For a certain species A, its rate of changes can be described as the following formula:

$$dA/dt = \Sigma f(k_i, [R]_i) - \Sigma f(k_j, [R]_j)$$

Here “i” stands for the reactions that generate A and “j” indicates reactions that consume A. ‘k’ is the reaction rate constant and [R] is the concentration of the reactants for that reaction. In other words, the net change (ΔA) of species A in a certain time dt is denoted by the difference of generated A (A_g) and consumed A (A_c).

$$\Delta A = A_g - A_c$$

Therefore, the concentration of A at time t+1 (A_{t+1}) can be described as

$$A_{t+1} = A_t + \Delta A$$

However, since the reaction is treated as continuous, the calculated net change ΔA will also be a continuous value, a real number. This is contrary to the fact that the number of any species should be discrete integers. Therefore, in this crossover method, a Bernoulli trial is conducted to determine the integer value of the net change. The fractional part of the real number from the deterministic method will be used as the success probability for Bernoulli trial. For example, the success probability will be 0.45 if the ΔA from deterministic step is 10.45. A true outcome will round up ΔA to 11, while a false outcome rounds it down to 10, ensuring the net change for each time step is always an integer. In this way, certain degree of randomness has been introduced into the deterministic solution, allowing the stochastic effect to be integrated into this crossover method.

The crossover method has been tested with systems such as auto-regulatory gene network and dimerization pathway [104]. It not only retains the efficiency of deterministic methods, but also reflects the fluctuations generally only captured by stochastic methods. However, similar to the pure stochastic methods, the previous version of crossover method was only able to handle elementary reaction. A standard Michaelis-Menten reaction, shown as following, will need to be broken into three elementary reactions: association, in which the substrate (S) binds to the enzyme (E) to form a complex (ES); dissociation, in which the complex dissociates back into the substrate and enzyme; conversion, in which the enzyme converts the substrate into the product (P) with the turnover number k_{cat}. This process, however, will introduce more difficulties to the system representation and add more computational cost to the simulation.

$$E + S \rightleftharpoons ES \overset{k_{cat}}{\rightarrow} E + P$$

In this study, we have improved crossover method and allowed it to directly model Michaelis-Menten type reactions. The two types of reactions will be marked differently in the system model and calculated by different functions. In the crossover algorithm, the rate of product formation of this type of reaction is given by the following equation, in which [E] is the enzyme concentration and [S] is the substrate concentration. K_m is the substrate concentration at which the reaction rate is at half-maximum, and is an inverse measure of the substrate's affinity for the enzyme.

$$\frac{d(P)}{dt} = k_{cat}[E]\frac{[S]}{K_m + [S]}$$

One of the major issues of decomposing Michaelis-Menten type reactions into elementary reactions is that the kinetic parameters for the elementary reactions are often not available. The ability of this crossover algorithm of directly modeling Michaelis-Menten type reactions allows it to simulate more complex systems.

1.3.3 Model Representation of the Crossover System

Various representations of biological systems have been developed to fulfill different purposes. For example, graphical representations are useful for both visualization and analysis, while matrix representations are more convenient for mathematical and computational analysis. However, users often need to combine results from different simulation or analysis tools in one project. Most of the time, the models and output need to be re-constructed in order for them to be used across different platforms. This process could be very time-consuming and error-prone. A commonly accepted software environment is greatly in need to allow individual research groups to easily exchange their models and reuse the code. Several undergoing project including Systems Biology Markup Language (SBML http://www.sbml.org/), CellML (http://www.cellml.org/) and the Systems Biology Workbench are examples of efforts that aim to form a standard and open software platform for modeling and analysis [105-106]. The SBML developed by Hucka et al. is a powerful presentation for biological networks such as gene regulation, cell signaling pathways and metabolic pathways [107]. The CellML is a XML based standard for describing mathematical models, especially the models relevant to biology processes [108]. The Systems Biology Workbench is a software framework that allows different applications to communicate with each other in an efficient manner [109]. Such efforts have significantly facilitated the development of new generation of databases concerned with biological pathways by employing better formatted models.

The SBML representation of biological networks consists of lists of functions, units, compartments, species, parameters, rules, reactions, and events [107]. In order to standardize and improve the readability of our models, we have developed a similar but much simpler representation using eXtensible Markup Language (XML). We will focus on compartments, species, parameters and reactions, as these are fairly sufficient for describing most simple discrete stochastic models. The system design in this crossover algorithm is demonstrated in **Figure 1.14**. The model will consist of one global system, with the parameters like initial time, time step, total time, and report interval defined at the system level. Since complex biological systems are often composed of multiple components or subsystems with close interactions, we have introduced the idea of multiple compartments into this crossover system. Each model may

contain one or more compartments. Each compartment will have its own species and reactions. The name, initial number and chemostat status (constant concentration for a species) will be assigned to each species, while every reaction will have its own kinetic parameters, input species and output species. For biological systems that are composed of more than one compartment, there might be transportable species between different compartments, which means the behavior of one compartment will affect that of others. This behavior is closer to reality since most of the biological systems are complex networks and one system is often affected by its environment. In the meantime, an event feature was also developed to represent specific activities during the simulation.

Here we took the auto-regulatory gene network, which is typically a low concentration biochemical network whose detailed dynamics are usually captured only by stochastic algorithms, as an example to illustrate the difference of model representation in SBML (**Figure 1.15a**) and crossover algorithm (**Figure 1.15b**). SBML has been now widely used for electronic representation of models in systems biology studies. Many simulation tools have provided various levels of support for SBML. However, although it can represent comprehensive information for the model and is well-formatted for computers to parse and generate, its verbosity makes it quite inconvenient for humans to read and write [110]. From the comparison in Figure 1.15, it can be seen that the representation in crossover algorithm is much shorter and easier to understand than that of SBML when presenting the same amount of information. Therefore, the simplicity of this crossover algorithm would make it easy for researchers to understand and apply across different applications for systems biology.

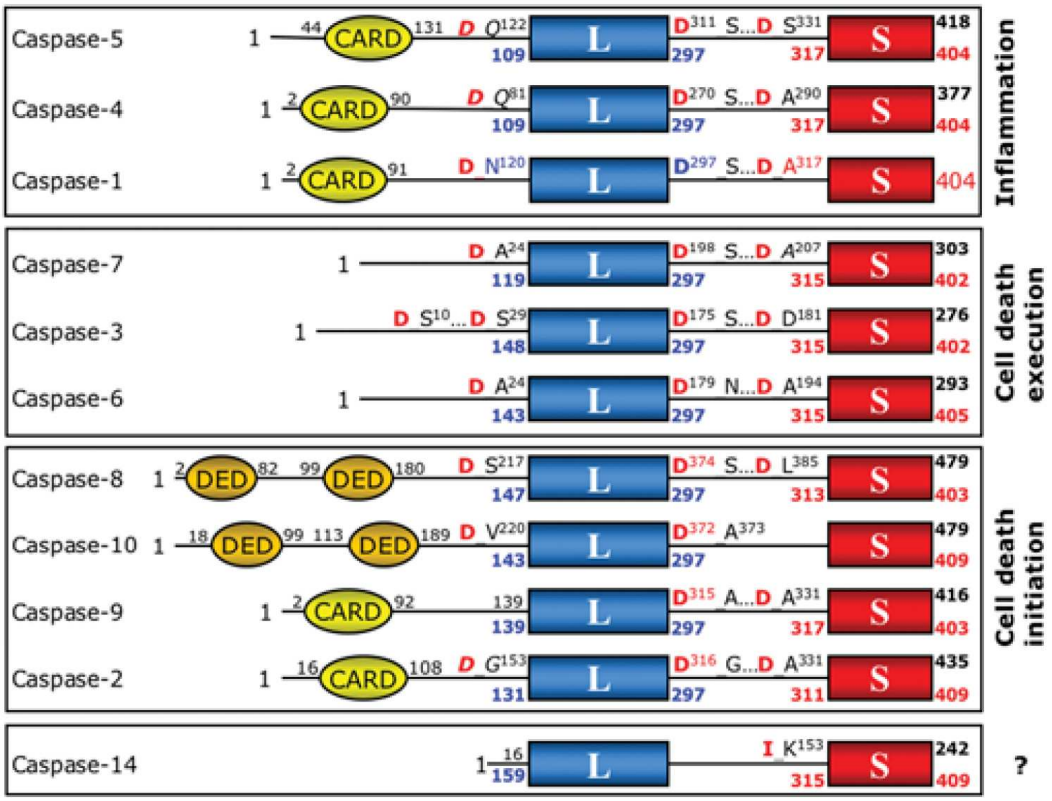


Figure 1.1 Domain organizations of human caspases. Different domains are indicated by colored blocks. (Fuentes-Prior et al., 2004)

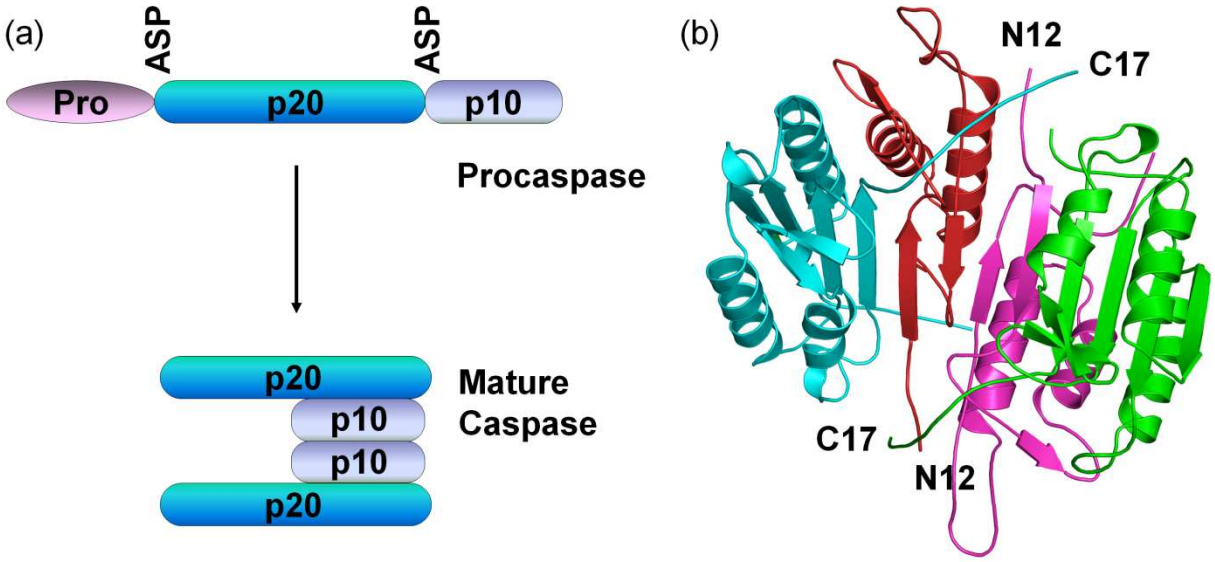


Figure 1.2 Activation and structure of caspase. (a) Scheme of procaspase activation. Caspases are synthesized as inactive procaspases which contain one prodomain, one large (p20) and one small (p10) subunit. Cleavage at specific aspartate (Asp) sites results in an active heterotetramer consisting of two large and two small subunits. (b) The 3D structure of caspase-3 heterotetramer. It consists of two large subunits (cyan and green) and two small subunits (red and magenta). N and C termini of the small and large subunits are indicated.

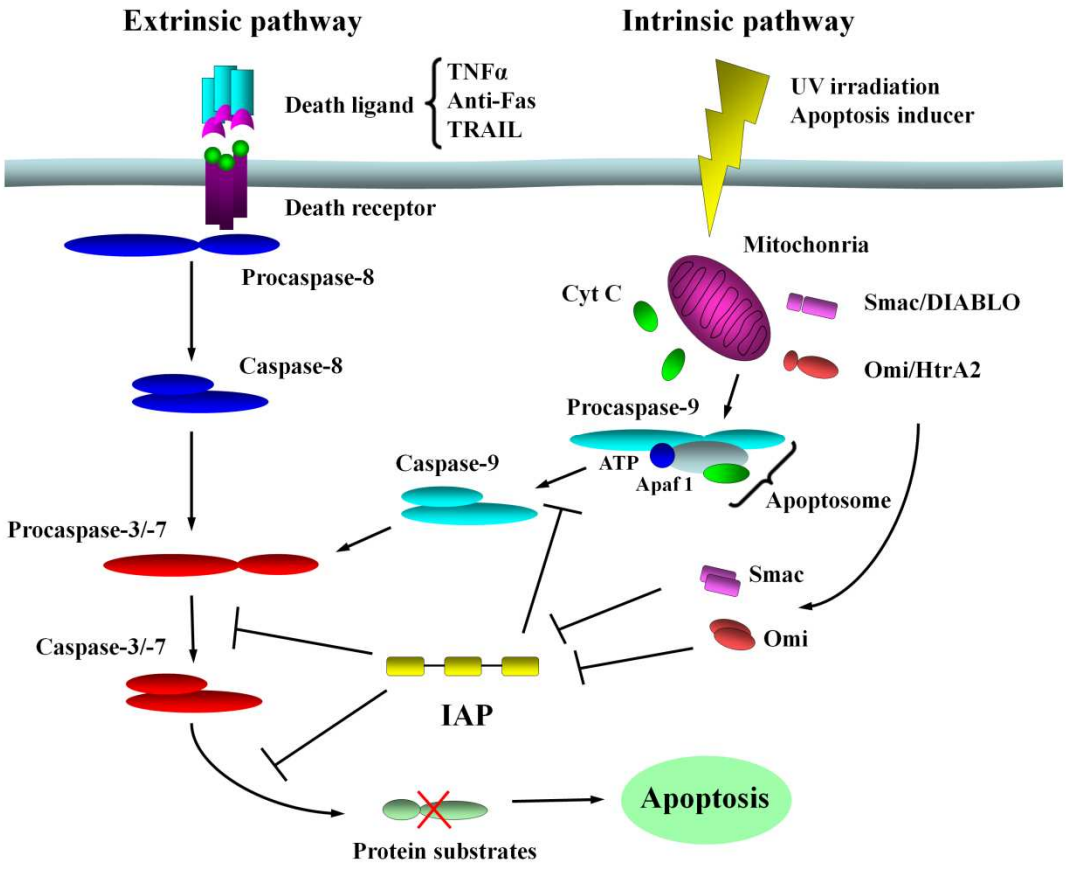
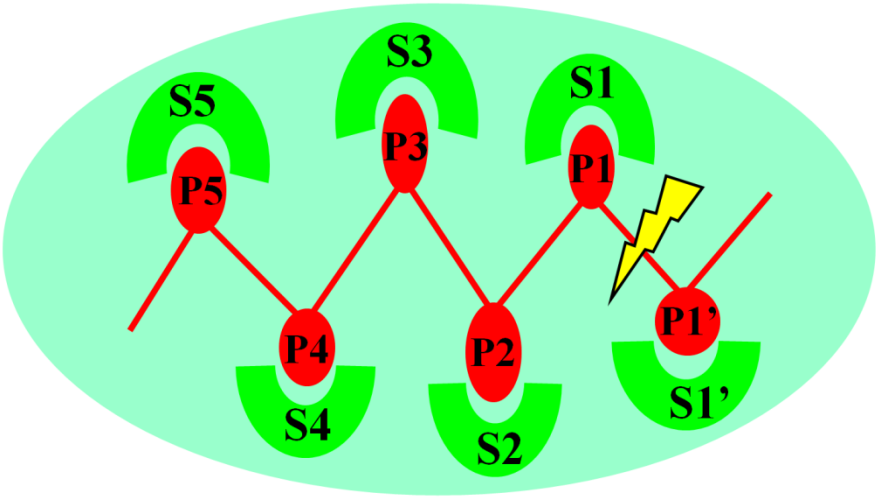
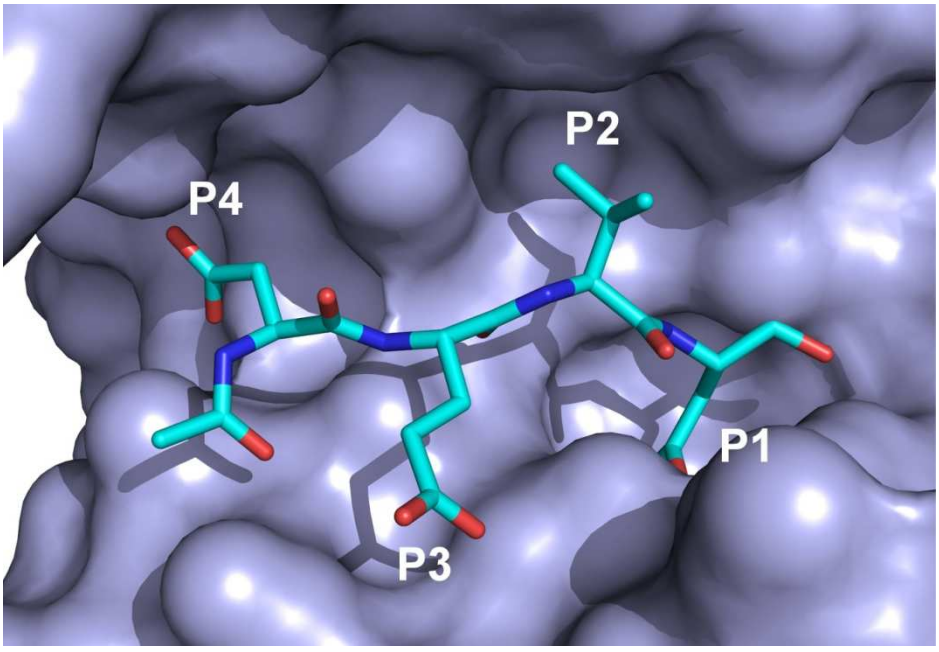


Figure 1.3 Extrinsic and intrinsic signaling pathways of apoptosis.

(a)



(b)



(c)

		Preferred P4-P1
Inflammatory caspases	caspases 1,4,5	WEHD
Initiators of apoptosis	caspase 2 caspases 8,9,10	DEHD LEXD
Executioners of apoptosis	caspase 6 caspase 3,7	VEHD DEVD

Figure 1.4 Substract specificities of caspases. (a) Schematic presentation of caspase substrate binding site. The substrate P1'-P5 are colored red. The active site S1'-S5 are colored green. The cleavage site on the substrate is indicated by yellow arrow. (b) Crystal structure of 2H5I shows the binding mode of DEVD in caspase-3. Surface representation is shown for the caspase-3 active groove. Substrate P1-P4 residues are labeled. (c) Specificities of caspases for positions P4-P1 in substrates. X refers to any amino acid.

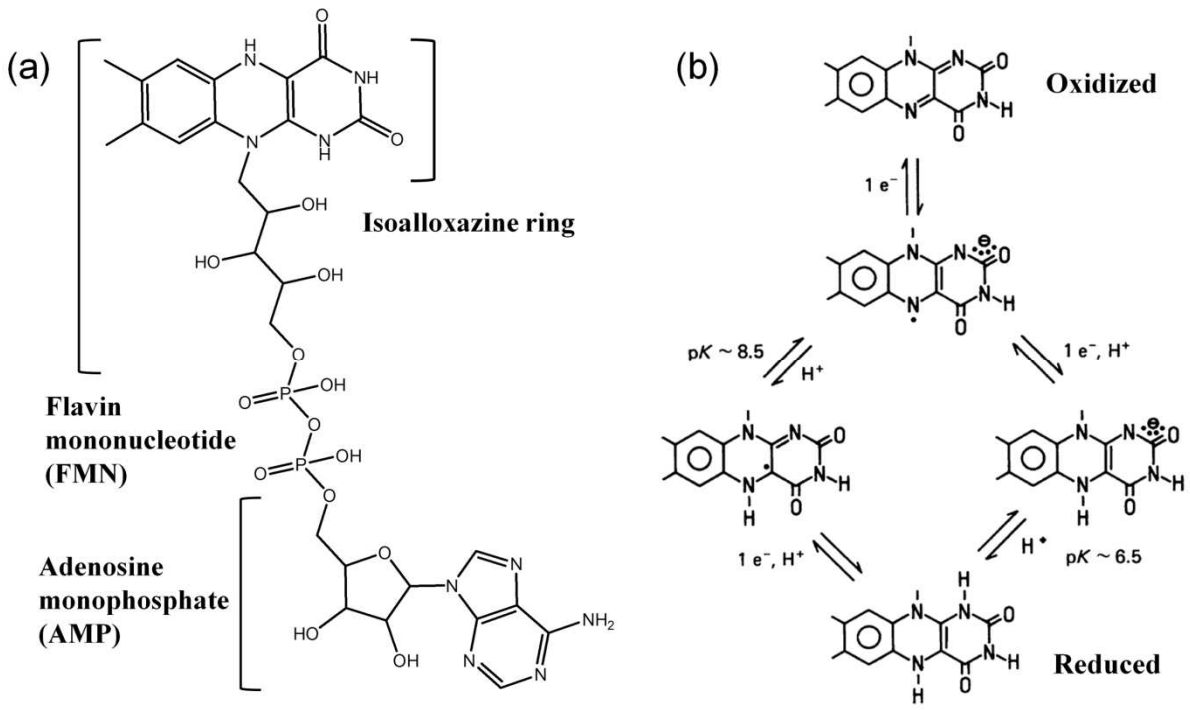


Figure 1.5 Strucutre of of FAD cofactor. (a) Schematic presentation of FAD cofactor. The division between the component parts FMN and AMP composing the FAD is shown. (b) Redox states of flavin. Modified from Ghisla et al., 1986.

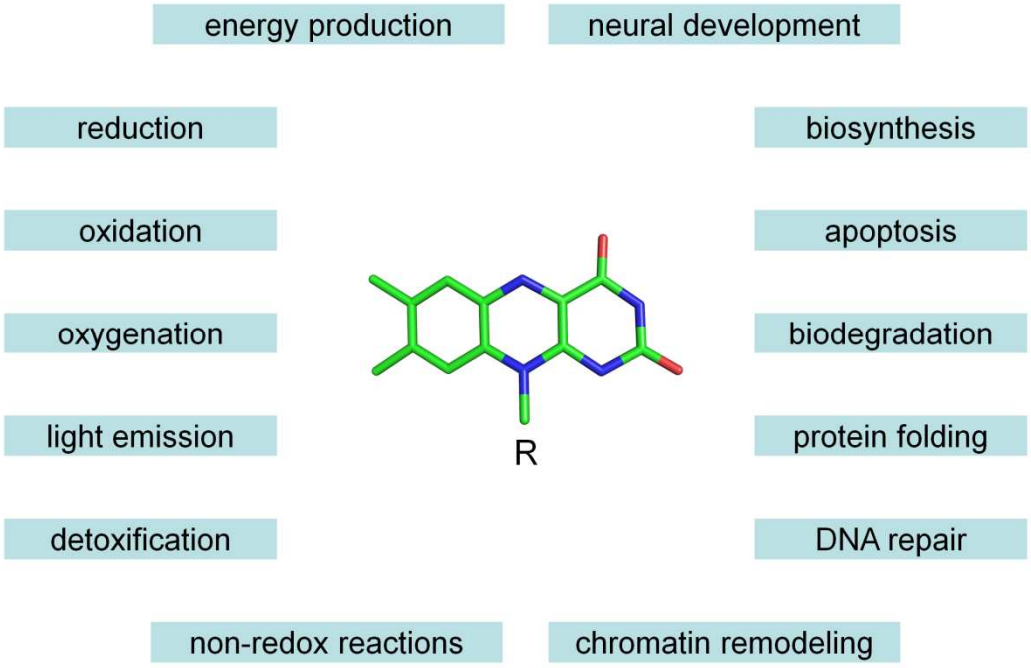


Figure 1.6 Biological functions of flavoenzymes.

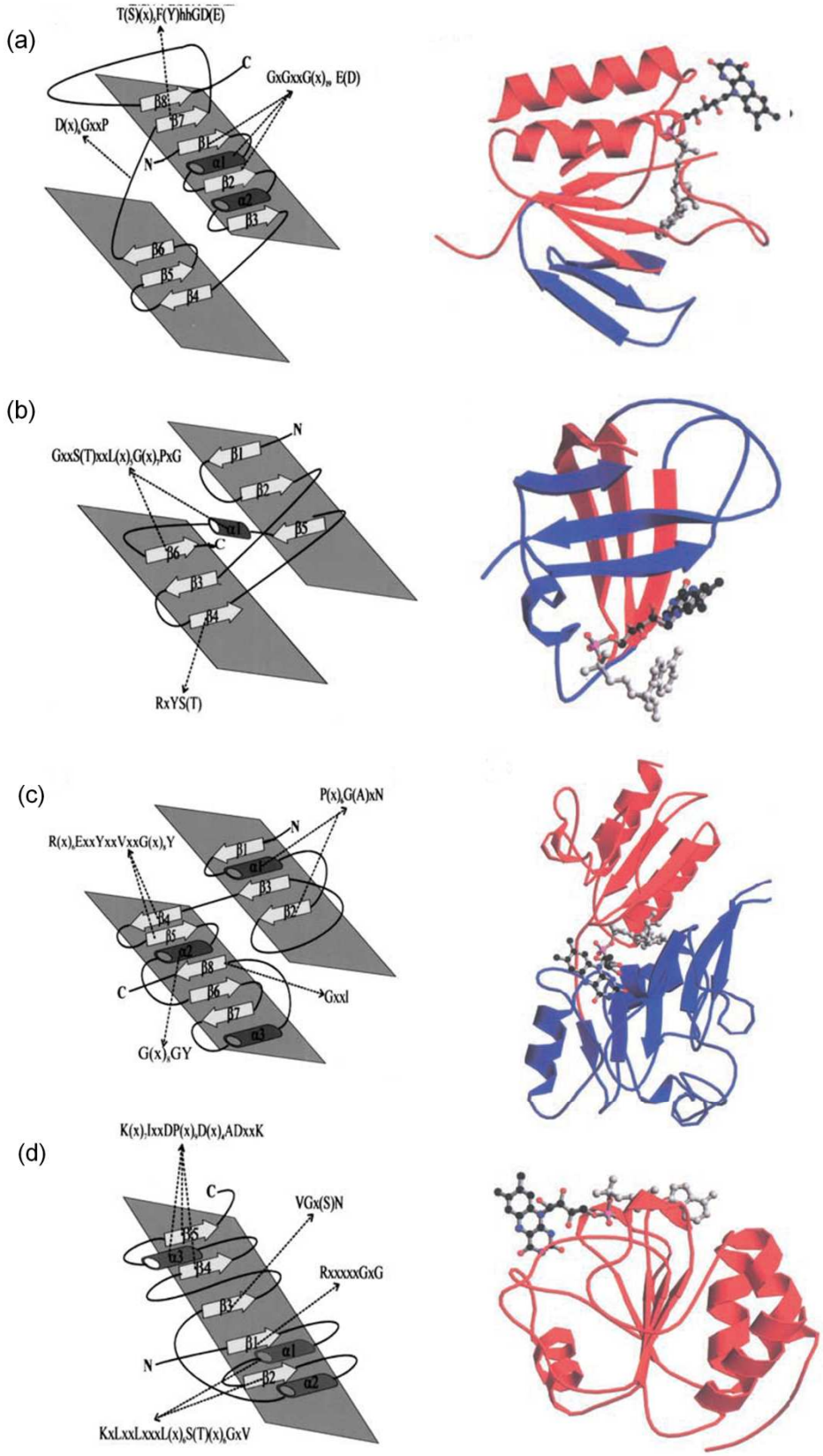

(a)
T(S)(x)2F(Y)hhGD(E)
C
GxGxxG(x)m E(D)
β8
β7
D(x)4GxxP
β1
N
α1
β2
α2
β3
β6
β5
β4
(b)
GxxS(T)xxL(x)3G(x)2PxG
N
β1
β2
α1
β5
β6
C
β3
β4
RxYS(T)
(c)
P(x)3G(A)xN
R(x)4ExxYxxVxxG(x)8Y
N
β1
α1
β3
β2
β4
β5
α2
β8
C
β6
Gxxl
β7
α3
G(x)4GY
(d)
K(x)3IxxDP(x)4D(x)4ADxxK
C
VGx(S)N
β5
α3
β4
RxxxxxGxG
β3
β1
α1
N
β2
α2
KxLxxLxxxL(x)6S(T)(x)4GxV

Figure 1.7 Topological diagram (left panel) and ribbon representation (right panel) of FAD-binding domain of the four FAD-family folds. (a) Rossmann fold adopted by the glutathione reductase (GR) family members. The crossover connection composed of a three-stranded antiparallelβ-sheet is colored blue. The red shading indicates the central five-stranded parallel β-sheet surrounded by two α-helices. The FAD cofactor adopts an elongated conformation. (b) Ferredoxin reductase (FR) family fold adopting a cylindricalβ-domain organized into two orthogonal sheets. The two antiparallel three-stranded β-sheets are shown in red and blue. The FAD in its bent conformation is shown with the isoalloxazine ring (black circles) pointing toward the FAD-binding domain. (c) The p-cresol methylhydroxylase (PCMH) family fold consists of two α+β subdomains. The FAD molecule adopts an elongated conformation and is located in between the two subdomains with the adenine ring pointing toward them. (d) The pyruvate oxidase (PO) family fold is similar to the double Rossmann fold found in dehydrogenases. The cofactor adopts an elongated conformation and lies perpendicular to the β-strands. Modified from Dym et al., 2001.

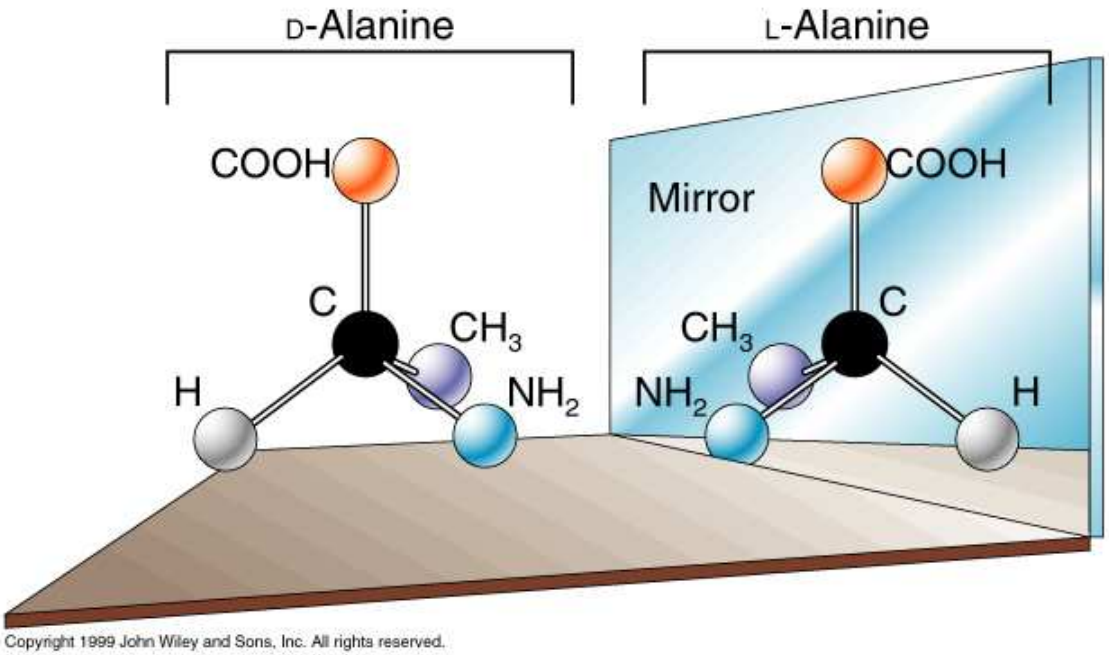


Figure 1.8 Optical isomers of D-alanine exist as mirror images of each other. (John Wiley and Sons. Inc. 1999)

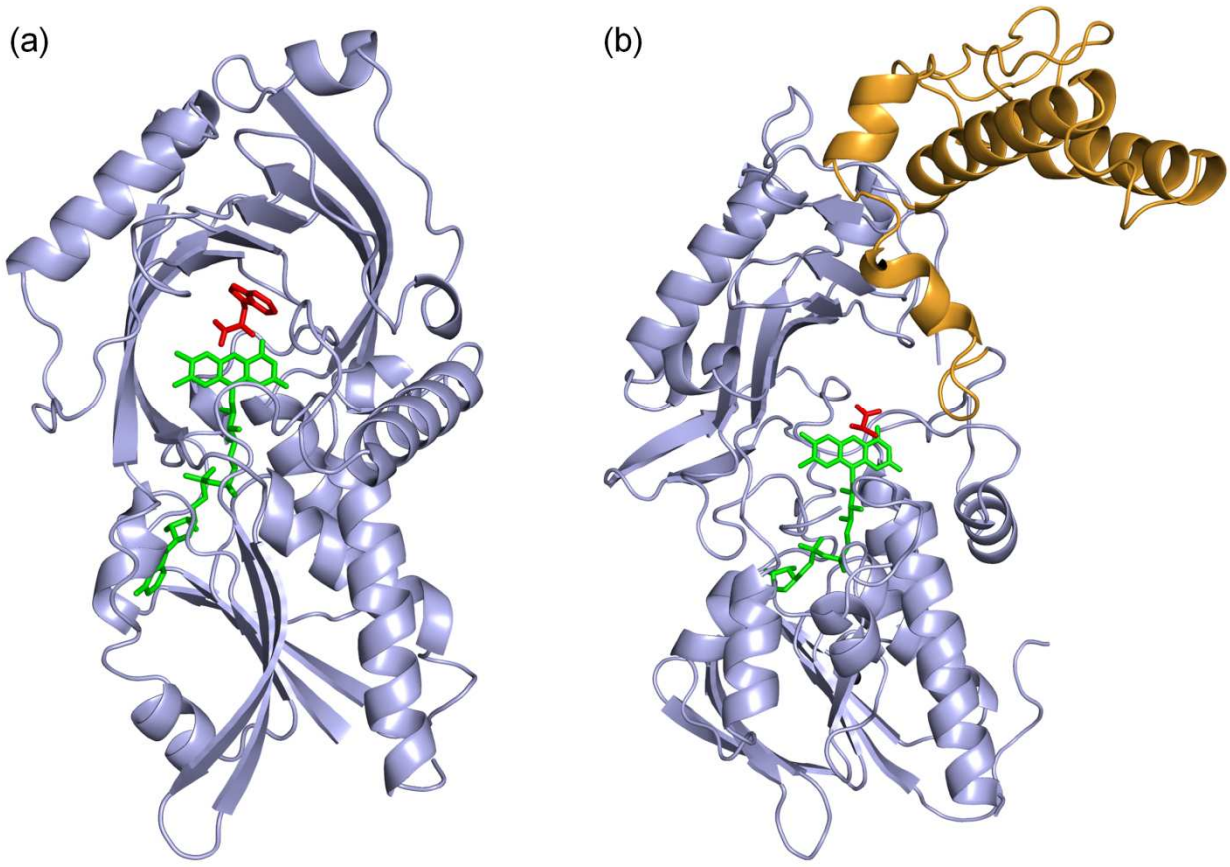


Figure 1.9 Structures of DAAO and LAAO. Cartoon representations of (a) DAAO/ iminotryptophan (PDB entry: 1DDO) and (b) LAAO/ L-alanine (PDB entry: 2JB1). Both enzymes are composed of a FAD-binding domain and a substrate-binding domain, while LAAO contains an extra helical domain (yellow color). Ligands and cofactor FAD are shown as sticks and colored red and green, respectively.

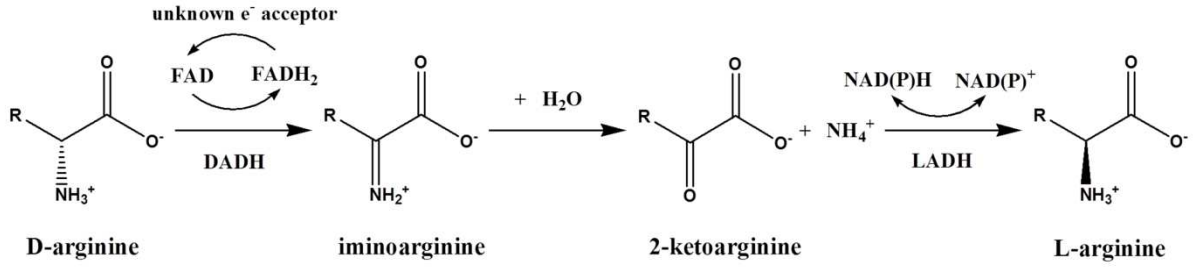


Figure 1.10 Reaction scheme for conversion of D-amino acids to L-amino acids by DADH and LADH.

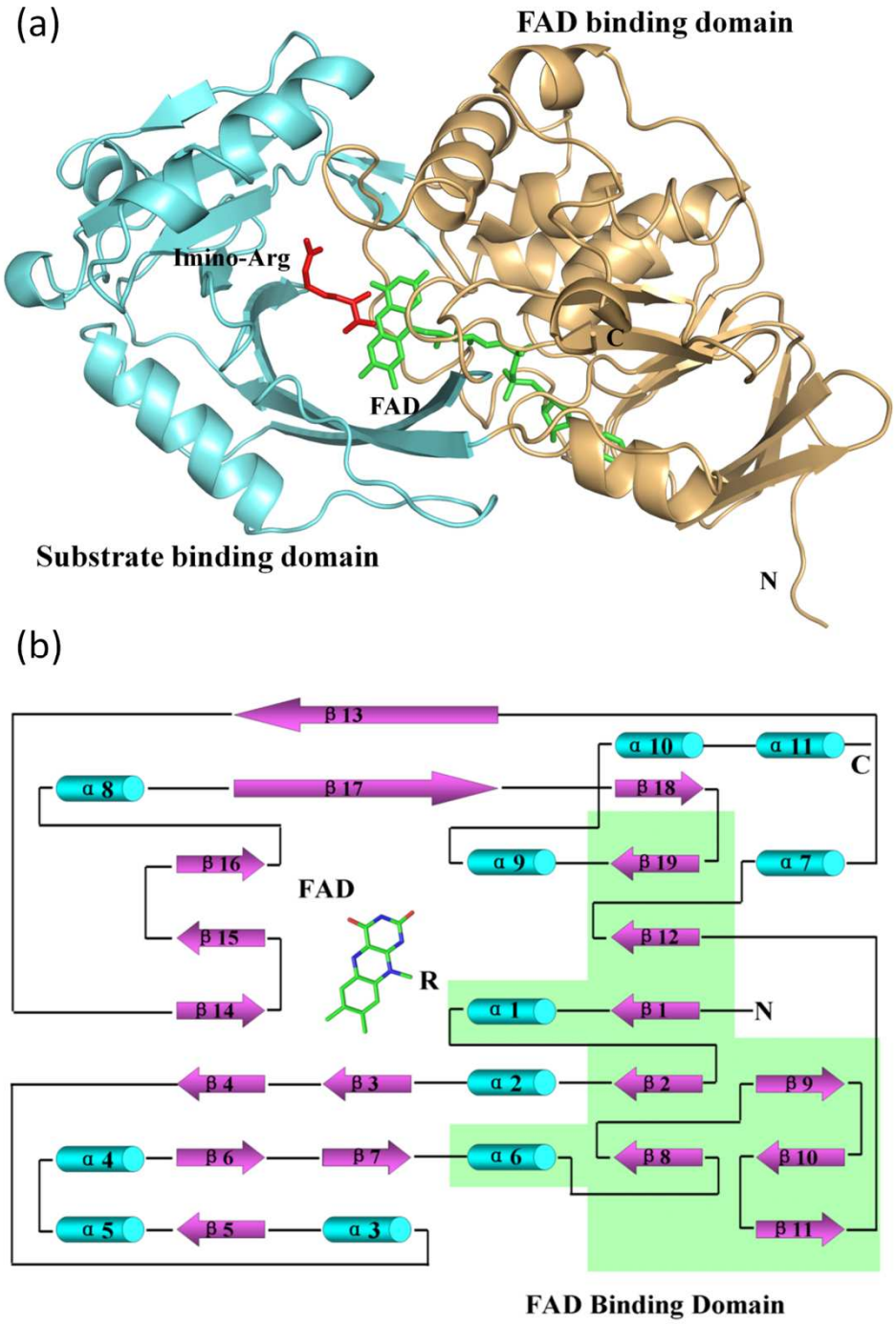


Figure 1.11 Structure of DADH. (a) Overall structure of DADH shown in cartoon representation. The enzyme is composed of FAD-binding domain (orange) and substrate-binding domain (cyan). Iminoarginine and cofactor FAD are shown as sticks and colored green and red, respectively. (b) Secondary structure topology for DADH. α-helices are shown as cylinders whereas β-strands are indicated by arrows.

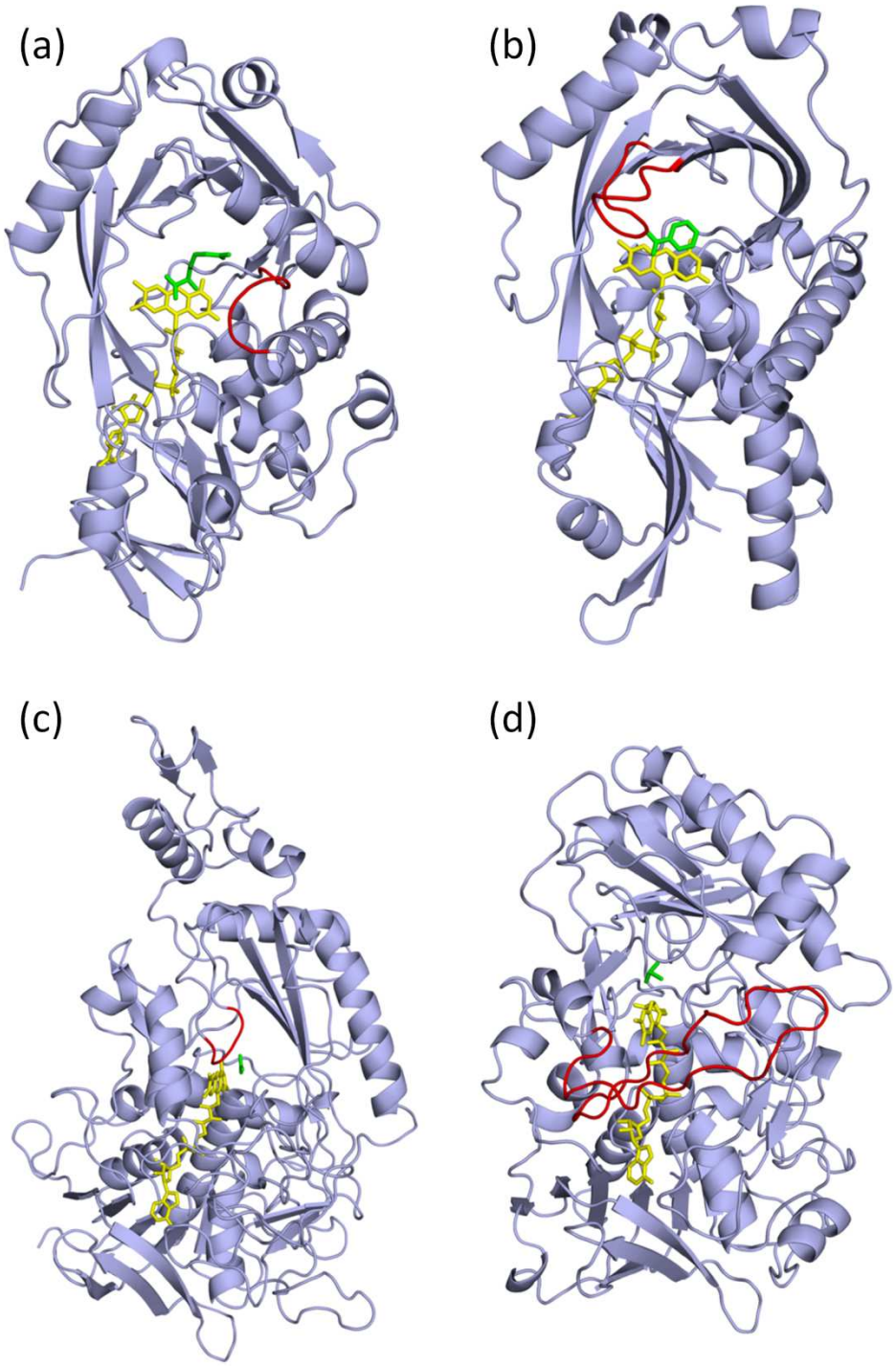


Figure 1.12 Active site lid in different enzymes. Active site lid structures (red color) in (a) D-arginine dehydrogenase (PDB: 3NYE), (b) D-amino acid oxidase (PDB: 1DDO), (c) pyranose 2-oxidase (PDB: 1TT0), and (d) choline oxidase (PDB: 2JBV). Active site lids are colored red. Ligands and cofactor FAD are shown as sticks and colored green and yellow, respectively.

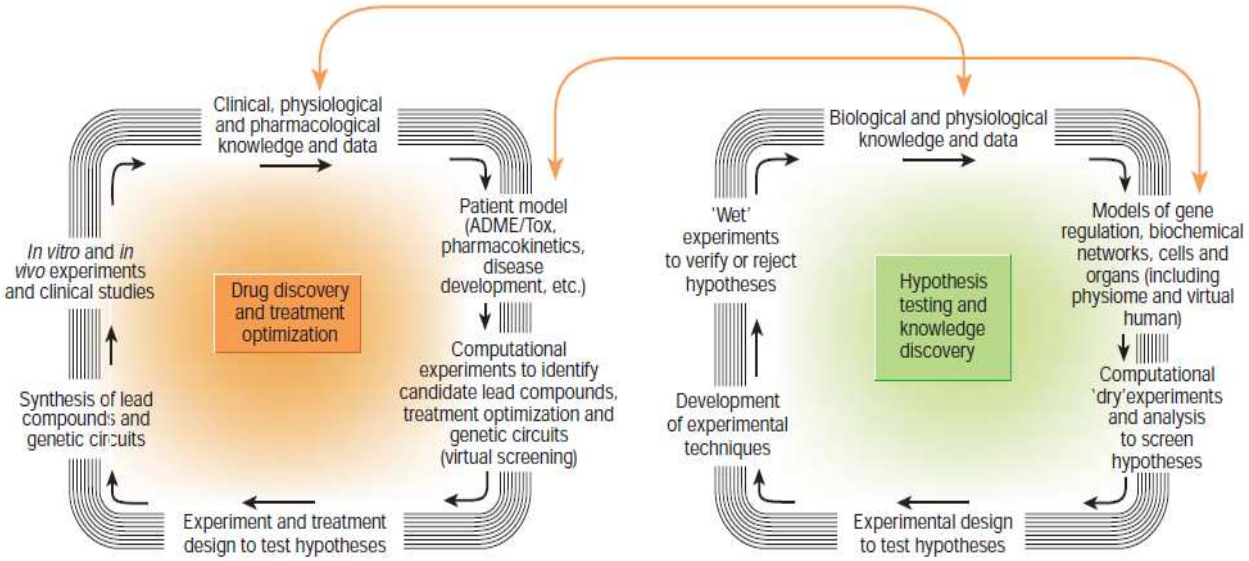


Figure 1.13 Linkage of a basic systems-biology research cycle with drug discovery and treatment cycles. Systems biology is an integrated process of computational modelling, system analysis, technology development for experiments, and quantitative experiments.

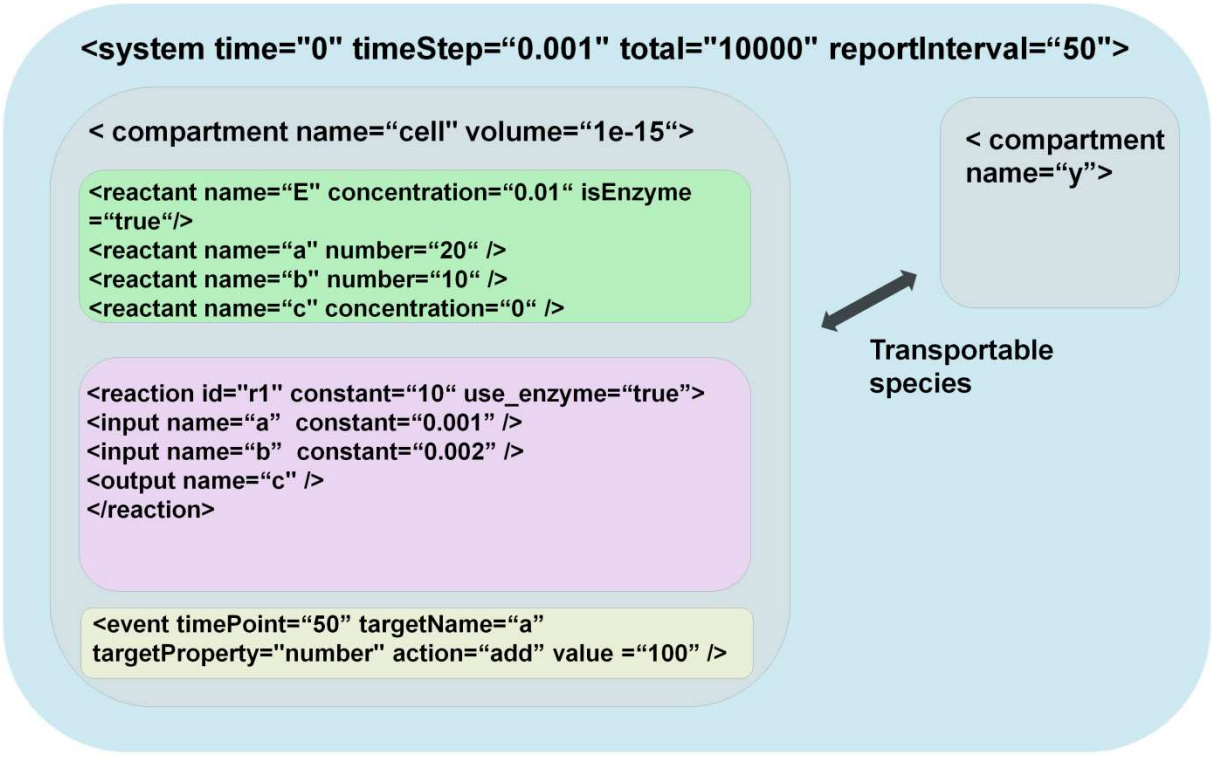


Figure 1.14 System design of the crossover algorithm. The model will consist of one global system, with the parameters such as initial time, time step, and others defined at the system level. Each system may contain one or more compartments with their own species and reactions. Some species may be transportable between different compartments.

(a)

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level3/version1/core" level="3" version="1">
<model id="AutoRegulatoryNetwork" name="Auto-regulatory network" >

<listOfCompartments>
<compartment id="Cell"/>
</listOfCompartments>

<listOfSpecies>
<species id="Gene" compartment="Cell" initialAmount="10"
hasOnlySubstanceUnits="true"/>
<species id="Rna" compartment="Cell" initialAmount="0"
hasOnlySubstanceUnits="true"/>
…
</listOfSpecies>

<listOfReactions>
<reaction id="Transcription" reversible="false">
<listOfReactants>
<speciesReference species="Gene" stoichiometry="1"/>
</listOfReactants>
<listOfProducts>
<speciesReference species="Gene" stoichiometry="1"/>
<speciesReference species="Rna" stoichiometry="1"/>
</listOfProducts>
<kineticLaw>
<math xmlns="http://www.w3.org/1998/Math/MathML">
 <apply>
 <times/>
 <ci> k2 </ci>
 <ci> Gene </ci>
</apply>
 </math>
<listOfLocalParameters>
<localParameter id="k2" value="0.01"/>
</listOfLocalParameters>
</kineticLaw>
</reaction>
…
</listOfReactions>

</model>
</sbml>
```

(b)

```
<system time="0" timestep="0.001" total="100000" reportInterval ="500">
<partition name="cell" volume="1e-15">

<reactant name="gene" number="10" chemostat="false"/>
<reactant name="rna" concentration="0" chemostat="false"/>
…

<reaction id="r1" constant="0.01">
<input name="gene" />
<output name="gene" />
<output name="rna" />
</reaction>
…

</partition>
</system>
```

Figure 1.15 System representations of auto-regulatory gene network shown in SBML. (a) and crossover algorithm (b). The same amount of information from the model is shown in both systems.

2 STRUCTURAL BASIS FOR EXECUTIONER CASPASE RECOGNITION OF P5 POSITION IN SUBSTRATES

(Published: Fu G, Chumanevich AA, Agniswamy J, Fang B, Harrison RW, Weber IT. *Apoptosis* 2008;13: 1291-1302.)

2.1 Abstract

The executioner caspases-3, -6 and -7 induce apoptosis by cleavage of many known and unknown protein substrates. Many studies have identified tetrapeptide (P1-P4) recognition motifs for caspases. Studies of caspase-3 recognition of hydrophobic P5 in pentapeptides have been extended to caspase-6 and -7 and the binding of peptides with polar or non-polar P5 residues has been investigated by kinetics, modeling and crystallography. Caspase-3 and -6 recognize P5 in pentapeptides as shown by enzyme activity data and interactions observed in the 1.6 Å resolution crystal structure of caspase-3 with the aldehyde inhibitor Ac-LDESD-Cho and conserved interactions predicted in a model for caspase-6. Caspase-3 preferred the hydrophobic P5 in LDEVD with 50% increased activity relative to the tetrapeptide DEVD consistent with observed hydrophobic contacts of P5 Leu with loop-4 residues Phe250 and Phe252, while caspase-6 showed a preference for polar P5 in QDEVD likely due to interactions with polar Lys265 and hydrophobic Phe263. Caspase-7 shows the highest activity on the DEVD tetrapeptide substrate in agreement with the absence of interactions with P5 in the 2.45 Å resolution crystal structure of caspase-7/LDESD. Also, these studies have identified a critical P5-anchoring residue in loop-3; the hydroxyl side chain and amide of Ser or Thr in caspase-2, -3 and -6 form hydrogen bonds with the main chain of P5 in the bound pentapeptide, in agreement with increased activity on the pentapeptide substrates. In contrast, caspase-7 has Pro at the equivalent loop-3 position, which cannot form hydrogen bonds with the P5 main chain atoms, consistent with the relative activity. Moreover, the initiator caspase-8, which lacks a loop-4 and has Pro in the loop-3, had only 20% activity on the tested pentapeptides relative to DEVD. This analysis of substrate recognition will help identify the specific protein substrates and signaling pathways of caspases.

2.2 Introduction

Caspase family members induce apoptosis by cleavage of diverse protein substrates, although many of the signaling pathways and potential substrates are poorly characterized. Altered caspase-mediated apoptosis is associated with many diseases. Increased caspase activity and apoptosis are observed in stroke and neurodegenerative diseases like Alzheimer's, Parkinson's and Huntington's disease [111-114]. Moreover, apoptosis has been suggested to be responsible for the death of cardiomyocytes during acute myocardial infarction, as well as for the progressive loss of surviving cells in failing hearts [115-116]. In contrast, reduced caspase activity is associated with cancer, autoimmune diseases, and viral infections [29, 117-119]. Pan-caspase inhibitors are in clinical trials for acute myocardial infarction and liver transplants [120-121]. Therefore, knowledge of the substrate specificity of caspases is critical for design of selective therapeutic agents to control cell death.

The caspase family consists of cysteine proteases that cleave the peptide bond after an aspartic acid in their substrates. They are classified as inflammatory and apoptotic caspases based on their function and prodomain structure. The apoptotic caspases are further divided into initiators and effectors of apoptosis. The initiator caspases-2, -8, -9, -10 and -12 activate the effector, or executioner, caspases -3, -6, and -7 by cleaving the procaspase to form the active enzyme [10, 122]. The active caspases-3,-6, and -7 hydrolyze many protein substrates in the signaling pathways leading to apoptosis [31]. The crystal structures have been solved for the apoptotic caspase-2, -3, -7, -8, -9, in their complexes with protein inhibitors, peptide analogs and non-peptide inhibitors [36, 123-124]. The enzymatically active caspase structure is a heterotetramer of two large and two small subunits formed by cleavage of a procaspase dimer. Peptide substrates or inhibitors bind in two similar active sites formed between the large and small subunits of the two heterodimers.

The substrate specificity of apoptotic caspases has been defined by studies of peptides. Most studies have analyzed the preference for tetrapeptides of P4-P1. Caspase-2, -3 and -7 prefer the tetrapeptide DEXD, while caspase-6, -8 and -9 recognize L/VEXD [10, 30]. The two effector caspases-3 and -7 are very similar; they share 56% sequence identity and preferentially recognize DEVD. Recently, the sub-

strate recognition sequence was extended to P5 for caspase-2 and -3. The P5 position of substrates was determined to be essential for recognition by caspase-2 [36]. Similarly, caspase-3 was shown to preferentially recognize pentapeptides with hydrophobic amino acids at P5 over tetrapeptides based on structural and kinetic data [34]. Caspase-6 shares 37 and 41% sequence identity with caspase-7 and -3, respectively. However, caspase-6 differs from the other two caspases by preferring the tetrapeptide sequence of LEHD over DEVD in substrates. The possibility of recognizing the P5 residue in substrates has not been explored for caspase-6.

We report structural and kinetic analysis of caspase recognition of the P5 position in peptide substrates. The specificity of the executioner caspases was investigated for polar or hydrophobic P5 residues using the colorimetric substrates Ac-QDEVD-pNA and Ac-LDEVD-pNA, and tetrapeptide Ac-DEVD-pNA as a control, where Ac is the acetyl group and pNA is *p*-nitroanilide. Crystal structures were solved of caspase-3 and -7 in their complexes with the pentapeptide Ac-LDESD-CHO, and molecular models were constructed for the caspase-6 complex. The relative activity on the substrates agrees with the caspase residues forming the S5 site and their interactions with the P5 position.

2.3 Experimental Procedures

2.3.1 Materials

Restriction enzymes were purchased from both New England BioLabs, Inc. and Stratagene. PCR SuperMix was purchased from Stratagene. All media were obtained from Difco or Gibco BRL. SDS–PAGE standards were purchased from Pharmacia. Other chemicals were obtained from Sigma.

2.3.2 Vector Construction for Caspase-6

Human caspase-6 (***EC 3.4.22.59*)** was engineered as a truncated, prodomain-less, construct *MetSerPhe*^{25}*-Asn*^{293} (similar to [32]) of the caspase-6 isoform alpha. The *pOTB7* vector (Invitrogen), containing full-length cDNA of caspase-6 (NCBI database ID: NM_001226, NCBI accession BC000305) was used as a template and amplified by PCR using the primers: 5’-

GAATTCATGTCGTTCTATAAAAGAGAA-3’ and 3’– CTCGAGATTAGATTTTGGAAAGAAATG-5’ to introduce *EcoR1* and *XhoI* restriction sites, respectively. The PCR transcript was subcloned into the *TOPO XL* vector (Invitrogen) through the blunt ends. The amplified DNA was excised with *EcoRI*/*XhoI* endonucleases and the caspase-6 construct was recloned into the bacterial expression vector *pET23b* as an *EcoRI*-to-*XhoI* fragment by ligation reaction and expressed in *Escherichia coli BL21DE3* (Stratagene Inc) cells with a C-terminal $(His)_6$ tag. The sequence of the construct was confirmed by the dideoxynucleotide chain-termination method using primers to T7 promoter and terminator.

2.3.3 Expression and Purification of Caspase-6

The expression construct was transformed into *E. coli BL21DE3* cells (Stratagene Inc) according to the manufacture’s protocol and the cells were grown on an LB/ampicillin plate overnight. Next day, one colony was transferred into 10 ml LB medium containing ampicillin (100 mg/ml) and the culture was grown overnight at 37°C with shaking. The 1000 ml of LB medium containing ampicillin was inoculated with the overnight culture and incubated at 37°C with shaking until OD_{600} = 0.6 – 0.8 (about 3-4 h) followed by induction with 2mM isopropyl- , D-thiogalactopyranoside (final concentration) for four hours at 30°C. Cells were collected by centrifugation at 5,000*g* for 20 min, resuspended in 40 ml working buffer (50 mM Tris–HCl, pH 7.4, 250 mM NaCl) and lysed by ultra-sonication without either protease inhibitors or lysozyme. The cellular debris was removed by centrifugation (20,000*g*, 25 min). The supernatant was examined for the presence of the caspase-6 by SDS–PAGE and stored at -20°C. The protein was purified from the soluble cell fraction using Ni-NTA chromatography (GE-Healthcare) according to the manufacturer’s standard protocol. The protein fractions eluted in 0.35-0.5 M imidazole in working buffer were collected, dialyzed against working buffer and concentrated to approximately 3 mg/ml using Centricon-10 ultrafiltration (Amicon Inc). During the concentration step the caspase-6 processed (cleaved) itself into two subunits to become an active protease. The final purification step was done on Sephacryl S-100HR column (Pharmacia). The purity of caspase-6 was above 95% as estimated from Coomassie blue stained 10-20% (w/v) polyacrylamide gels, which were done in denaturing, reducing conditions. The purified

protein was equilibrated with 100 mM HEPES, pH 7.25, containing 100 mM NaCl, 0.1% CHAPS, 10% sucrose, using Centricon-5 ultrafiltration, and stored at -20°C. The final yield of catalytically active caspase-6 was about 1.5 mg per liter of bacterial culture.

2.3.4 Expression and Purification of Caspase-3, -7 and -8

Recombinant human caspase-3, -7 and -8 were expressed in *E. coli* and purified as described previously [34, 125].

2.3.5 Determination of Activity

Measurements of the caspase activity were performed with the colorimetric assay as described previously [34]. Caspase-3, -7, and -8 were preincubated in assay buffer (50 mM HEPES, 100 mM NaCl, 0.1% CHAPS, 10% glycerol, 1 mM EDTA and 10 mM DTT, pH7.5) at room temperature for 5 mins prior to the addition of substrate at different concentrations. The assay buffer for caspase-6 was 50 mM HEPES, pH 7.25, containing 100 mM NaCl, 10% sucrose, 0.1% CHAPS. The *p*-nitroanilide released by the substrate hydrolysis was measured at a wavelength of 405 nm using a Polarstar Optima microplate reader (BMG Labtechnologies, NC). All assays were performed in triplicate and the mean values were plotted. Kinetic constants were calculated by direct fits of the data, obtained at less than 20% substrate proteolysis, to the Michaelis–Menten equation using non-linear regression analysis in SigmaPlot 9.0 (SPSS Inc).

2.3.6 Crystallization, X-Ray Data Collection and Refinement of Caspase-3 and -7 with LDESD

The crystals of caspase-7 complexed with inhibitor Ac-LDESD-CHO were grown by hanging drop vapor diffusion method at room temperature using 1:15 ratio of the enzyme to inhibitor. The crystals were obtained with a well solution of 0.1M sodium citrate buffer (pH 5.0-5.5) and 2.1 M sodium formate. Caspase-3 was incubated at room temperature with Ac-LDESD-CHO at 10-fold molar excess. Crystallization was performed by the hanging-drop vapor-diffusion method: 1 μl of protein solution (4

mg/ml) was mixed with an equal volume of the reservoir solution (100 mM sodium citrate, 5% glycerol, 10 mM dithiothreitol, 14–16% (w/v) PEG 6000, pH 7.0). Crystals grew at room temperature within 24 h. The crystals were soaked in the mother liquor with 22% (caspase-7/LDESD) and 20% (caspase-3/LDESD) glycerol as cryoprotectant for ~1 min and immediately frozen in liquid nitrogen. Diffraction data were collected at 100°K on beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory. The data were integrated and scaled with HKL2000[126].

The crystal structures were solved by molecular replacement using PHASER [127] in the CCP4i suite of programs [128] using the starting models of caspase-7 complexed with Ac-DEVD-CHO (PDB code 1F1J) [123] and caspase-3 complexed with Ac-VDVAD-CHO (PDB code 2H65). The structures were subjected to several rounds of refinement in CNS or SHELX97 [129] for caspase-7 and -3, respectively. The molecular graphics programs O 9.0 [130] and Coot 0.33 [131] were used in model rebuilding. The inhibitor was fitted into unambiguous electron density. The solvent molecules were inserted at stereochemically reasonable positions based on the peak height of the $2F_o$-F_c and F_o-F_c electron density maps, hydrogen bond interactions and interatomic distances. The geometry of the refined structures was validated according to the Ramachandran plot criteria of [132]. Sequence alignment was performed by using the ClustalW server [133] and the figure was generated with GeneDoc (http://www.cris.com/2Ketchup/genedoc.shtml). Molecular figures were prepared with Molscript, Raster3D [134] and PyMol (http://www.pymol.org).

The crystal structures have been deposited in the Protein Data Bank with accession codes 3EDQ for caspase-3/LDESD, and 3EDR for caspase-7/LDESD.

2.3.7 Construction of Molecular Model of Caspase-6

The model for caspase-6 in complex with the peptide aldehyde Ac-LDESD-CHO was constructed by homology modeling from the new high resolution crystal structure of caspase-3/LDESD. Sequence alignment for caspase-6 was performed with a profile-profile method using full dynamics programming

and the Kullback entropy [135] with frequency profiles obtained from Psi-BLAST [136]. The program AMMP [137-138] was used with the latest version of the potential set, which was reparameterized to minimize RMS deviation when applied to high resolution crystal structures (parameter set atoms.tuna). All hydrogen atoms were included in this potential. The adducts of the catalytic cysteine 163 with P1 Asp, and the N-acetyl Leu (P5) residue was explicitly modeled with charges derived with the method of moments approximation [139]. Side chains were built using the analytic atom building routines in AMMP [140] followed by annealing over local dominating sets, minimization and conjugate gradients.

2.4 Results and Discussion

2.4.1 Predictions for Caspase Recognition of P5 in Pentapeptides

The P5 specificity of caspase-6 was predicted from comparative analysis of the sequence and structures of caspases. Molecular models were generated for human caspase-6 in the absence of diffraction quality crystals. Alignment of the amino acid sequences (**Figure 2.1**) suggested that caspase-6 shared the P5 binding loop observed in the known crystal structures of caspase-3 [34, 141] and caspase-2 with pentapeptides [36]. The hydrophobic P5 residue interacts with Phe250 and Phe252 in caspase-3 and with Tyr273 and Pro275 at the structurally equivalent residues of caspase-2. Caspase-6, which has Phe263 and Lys265 at the equivalent positions, was predicted to recognize both hydrophobic and polar P5 residues in substrates. In contrast, caspase-7 has the polar residues Gln276 and Asp278 in the corresponding positions of S5, and was predicted to recognize polar P5 residues. Caspase-8 was used as a control since there was no evidence for an S5 binding pocket in the sequence alignment or structure. Therefore, caspase-8 was expected to recognize only tetrapeptides, although it is possible that pentapeptides can bind in a different manner to that observed for caspase-2 and -3. These predictions were evaluated by solving the crystal structures of caspase-3 and -7 with the peptide aldehyde inhibitor Ac-LDESD-CHO, and determining the activity of caspase-3, -6, -7 and -8 for the pentapeptide substrates Ac-LDEVD-pNA and Ac-QDEVD-pNA, containing hydrophobic or polar P5 residues.

2.4.2 Kinetic Data for P5-Containing Substrates

The substrate hydrolysis was analyzed for the effector caspases-3, -6, and -7 in comparison to the initiator caspase-8. The effect of P5 in substrates was investigated using the pentapeptides Ac-LDEVD-pNA containing a hydrophobic P5 residue and Ac-QDEVD-pNA with a polar P5 residue. The tetrapeptide substrate Ac-DEVD-pNA, which is hydrolyzed efficiently by all four caspases, was used as a reference. The kinetic parameters for activity of caspase-3, -6, -7 and -8 are listed in **Table 2.1** and the relative activities are compared in **Figure 2.2**. Caspase-3 showed increased relative activity of 146% for the pentapeptide with P5 Leu, in agreement with the previous study [34], while the relative activity was lower (85%) for the peptide with the polar P5 Gln, due primarily to changes in K_m. Caspase-6 showed increased activity for both pentapeptides; a small increase to 120% for LDEVD and a substantial increase to 150% for QDEVD, due to correlated changes in both k_{cat} and K_m values relative to the activity on DEVD. LEHD, however, is a better tetrapeptide substrate than DEVD for caspase-6 with 3-fold higher k_{cat}/K_m (0.19 $min^{-1}\mu M^{-1}$ and 0.058 $min^{-1}\mu M^{-1}$, respectively), in agreement with the known tetrapeptide specificity [30]. Caspase-7 showed decreases in relative activity to 56% and 67% for the substrates with P5 Leu and Gln, respectively, with changes in both k_{cat} and K_m. Substantial changes were observed for caspase-8 where the relative activity dropped sharply to about 20% for both pentapeptides compared to DEVD with 3 to 4-fold increased values of K_m.

2.4.3 Crystal Structures of Caspase-3/ LDESD and Caspase-7/ LDESD

Caspase-3 and caspase-7 were crystallized in their complexes with the peptide aldehyde inhibitor Ac-LDESD-CHO. The caspase-3 complex was crystallized in the orthorhombic space group of $P2_12_12_1$ while caspase-7 complex was crystallized in the trigonal space group of $P3_221$. The structures were refined to the resolutions of 1.61 Å and 2.45 Å, respectively. The data collection and refinement statistics are listed in **Table 2.2**. The mature caspase-3 contains two heterodimers $(p17/p12)_2$, which are derived from processing of procaspase-3 (**Figure 2.3**). The residues 35-174 and 34-174 are visible in the electron density for the two p17 subunits, respectively, and 186-278 in the two p12 subunits in the heterotetramer.

The overall structure of caspase-3/LDESD is very similar to the previously reported structures of caspase-3 with the canonical tetrapeptide DEVD and pentapeptide VDVAD [34], with rmsd values of 0.38 Å and 0.17 Å for Cα atoms, respectively. The asymmetric unit of caspase-7/LDESD contains a complete catalytic unit of two p20-p10 heterodimers. Residues 58-196 and 52-196 are visible in the two p20 subunits, respectively. Residues 211-303 are visible in both p10 subunits in the heterotetramer. The refined overall structure of the caspase-7/LDESD compares very well with the previously reported structure of caspase-7 in complex with the canonical tetrapeptide DEVD [123]. The entire catalytic domain with the two heterodimers can be superposed with those of caspase-7/DEVD with rmsd values of 0.27 for 460 topologically equivalent Cα atoms.

2.4.4 Interactions of Caspase-3 with LDESD

Most of the residues in the active site of caspase-3/LDESD are in similar position and orientation when compared to those of the caspase-3/VDVAD complex [34]. The reported alternate conformations of His121 were also observed in current structure, supporting the proposed Cys-His catalytic dyad hydrolysis mechanism [142]. All the residues in the peptide substrate are clearly visible in the electron density as shown in **Figure 2.4**. The pentapeptide of caspase-3/LDESD can be superimposed with those of caspase-2/LDESD and caspase-3/VDVAD structures with RMS deviations of 0.33 (0.17) Å and 0.09 (0.18) Å for the 5 topologically equivalent Cα atoms, respectively. (The values in the parenthesis show the RMS deviations of the inhibitor in the second catalytic site.) The interactions between caspase-3 and the peptide analog Ac-LDESD-CHO are summarized in **Figure 2.5a**. Compared to the caspase-3/VDVAD structure, caspase-3/LDESD has substitutions at three positions, P2 (Ala/Ser), P3 (Val/Glu) and P5 (Val/Leu). The interactions at the P2 position are similar in both structures. Both Ala and Ser have van der Waals interactions with caspase residues Tyr204 and Trp206. However, the main chain amide, carbonyl oxygen and side chain oxygen of P2 Ser form hydrogen bond interactions with three water molecules since the P2 side chain is exposed to solvent. The side chain of P3 Val in VDVAD forms van der Waals interactions with caspase-3 atoms. In contrast, the negatively charged side chain of P3 Glu in LDESD forms a strong

ionic interaction with the side chain of Arg207 and their main chains are connected by two hydrogen bond interactions between the carbonyl oxygen atoms and amide groups.

At the P5 position, the main chain amide of Leu forms a hydrogen bond interaction with the side chain hydroxyl of Ser209. The carbonyl oxygen of Leu forms hydrogen bond interactions with both the side chain hydroxyl and main chain amide of Ser209. The hydrophobic Leu side chain is accommodated in the hydrophobic pocket formed by the side chains of Phe250 and Phe252, which is similar to the P5 interactions in the caspase-3/VDVAD structure. The CD1 of P5 Leu forms van der Waals interactions with the aromatic side chain of Phe252. In addition, CD2 of P5 forms van der Waals interactions with the side chains of both Phe250 and Phe252. The acetyl group of Ac-LDESD-CHO protrudes into the solvent region. It has been suggested that the P5 residue is unnecessary for the efficient cleavage of caspase-3 and -7 substrates [33]. However, the current investigation confirms that caspase-3 possesses a hydrophobic S5 pocket, which plays an important role in the substrate recognition and cleavage.

2.4.5 Predicted Interactions of Caspase-6 with LDESD

The modeled structure of caspase-6/LDESD showed very similar overall structure and interactions with the pentapeptide as observed in the crystal structure of caspase-3/LDESD. These two caspases are predicted to have conserved hydrogen bond interactions with the pentapeptide, with the exception of the interactions with P5 (**Figure 2.5b**). The caspase-6 has different S5 amino acids in loop-3 and -4, forming similar interactions with P5 Leu to those in caspase-3/LDESD. The main chain carbonyl oxygen and amide of P5 are anchored by hydrogen bonds with the amide and side chain hydroxyl of Thr222 in caspase-6 rather than the structurally equivalent Ser209 in caspase-3. The side chain of P5 Leu forms hydrophobic interactions with loop-4 residues Phe263 and Lys265 in caspase-6, which correspond to Phe250 and Phe252 in caspase-3.

2.4.6 Interactions of Caspase-7 with LDESD

The pentapeptide inhibitor in the caspase-7/LDESD complex adopts an extended conformation as seen for the canonical tetrapeptide DEVD. The P5 Leu in the pentapeptide buries an additional ~ 208 $Å^2$

of surface area when compared to DEVD in the caspase-7/DEVD complex. The pentapeptide of caspase-7/LDESD can be superposed with those of caspase-2/LDESD, caspase-3/LDESD and caspase-3/VDVAD structures with RMS deviation of 0.54(0.48) Å, 0.22(0.35) Å and 0.31(0.43) Å for the 5 topologically equivalent Cα atoms, respectively. (The values within the parenthesis are those of the inhibitor in the second catalytic site.) The orientation and position of the P1-P4 residues of caspase-7/LDESD are very similar to those in the caspase-7/DEVD complex. The only difference is the Val/Ser substitution at the P2 position, and both side chains form similar hydrophobic interactions with Tyr230 of caspase-7 (Fig. 5c). The P5 Leu side chain of LDESD interacts with the residues of the loop-4 in the active site cleft. However, the main chain amide and carbonyl oxygen of P5 Leu do not form hydrogen bonds with caspase-7, unlike the interactions observed in caspase-3/LDESD (**Figure 2.5a**). The CD1 of P5 Leu has van der Waals interactions with the side chain of Asp278. The P5 CD2 forms van der Waals interactions with Cα atom of Ser277, and the main chain carbonyl groups of both Gln276 and Ser277. The hydrogen bond interaction between the main chain carbonyl of Gln276 and the P4 amide in the caspase-7/DEVD structure is conserved in the caspase-7/LDESD complex, while the acetyl group of Ac-LDESD-CHO is exposed to the solvent.

2.4.7 Comparison of Caspase/LDESD Structures

The P5 interactions were compared for the pentapeptide complexes with caspase-2, -3, -6 and -7 (**Figure 2.6**). The P5 Leu side chain forms similar hydrophobic interactions with the side chains of equivalent residues Phe250 and Phe252 in caspase-3, Tyr273 and Pro275 in caspase-2, Phe263 and Lys265 in caspase-6. However, in caspase-7 the P5 Leu side chain forms van der Waals interactions with the main chain atoms of residues 276-277 and the side chain of Asp278. The main chain of P5 Leu is anchored by the loop-3 residue Thr243 in caspase-2, Ser209 in caspase-3, and Thr222 in caspase-6. In caspase-3, the P5 amino acid is secured by the interaction of its main chain amide with side chain hydroxyl of Ser209, and by the two hydrogen bonds formed by its carbonyl with main chain amide and side chain hydroxyl of Ser209. Similar interactions are predicted for Thr222 in the caspase-6 complex. In caspase-2, the main

chain amide of Thr243 forms a hydrogen bond interaction with the carbonyl of P5 Leu and the side chain hydroxyl group of Thr243 interacts with the backbone amide of P5 [36]. Therefore, the hydroxyl of Ser or Thr of loop-3 in caspase-2, -3 and -6 forms stabilizing interactions with the main chain of the P5 residue. In contrast, Pro235 is the corresponding loop-3 residue of caspase-7, which cannot form the hydrogen bonds that anchor P5 in the other caspases. When caspase-2/LDESD, caspase-3/LDESD and caspase-7/LDESD structures are superposed, the residues P1-P4 of the peptides have almost identical position and orientation (**Figure 2.7**). However, the Cα of the P5 residue of caspase7/LDESD is ~2 Å further away from Pro235 in loop-3 when compared to the corresponding positions in the pentapeptide complexes of caspase-2 and caspase-3. This change is due to loss of the critical hydrogen bonds in the caspase-7/LDESD complex. Instead, the P5 Leu side chain forms closer interactions with the loop-4 residues Ser277 and Asp278. Therefore, unlike caspase-2, -3 and -6, caspase-7 is not structurally suited to accommodate the hydrophobic P5 residue in peptides.

Interactions between P5 residues and S5 subsite were shown to produce conformational changes at loop-1 and loop-4 regions of caspase-3 when compared with the complex with tetrapeptides [34]. Structures of caspase-3 with tetrapeptide DEVD, pentapeptide LDESD and VDVAD were superimposed, and the distance between the Cα atoms of Gly60 (in loop-1) and Asp253 (in loop-4) was measured to show the conformational differences (**Figure 2.8**). The two loops have moved closer by about 3 Å in the structures of caspase-3/LDESD and caspase-3/VDVAD compared with that of caspase-3/DEVD. This conformational change partly closes the active pocket formed by the four loops and enables the P5 side chain to form favorable van der Waals interactions with caspase-3 residues. In the caspase-3/VDVAD structure, only one of the P5 Val CD atoms has hydrophobic interactions within the S5 pocket formed by Phe250 and Phe252 [34]. In caspase-3/LDESD structure, however, both CD atoms of the P5 Leu have van der Waals contacts with Phe250 and Phe252 (**Figure 2.6a** and **Figure 2.8a**). This observation confirms that loop-1 and loop-4 regions of caspase-3 are flexible and accommodate a variety of residues by

an induced-fit mechanism. However, it is not possible to determine from these static crystal structures whether the observed conformational differences result from entropic or enthalpic contributions.

Similar analyses were performed for the structures of caspase-7 in complex with tetrapeptide and pentapeptide (**Figure 2.8b**). The separation of loop-1 and -4 is similar in caspase-7/DEVD and caspase-7/LDESD (19.9 and 20.3 Å, respectively), indicating that no conformational change was induced in agreement with the kinetic data. The existence of a conformational change in caspase-6 upon binding of P5 residue cannot be verified in the absence of crystal structures. However, its preference for a P5 residue in kinetic study and structural similarity with caspase-3 suggested that caspase-6 also had a flexible S5 binding site formed by the four loops.

2.4.8 Correlation of Kinetic and Structural Data

Knowledge of the molecular structures of caspases has been critical for understanding their recognition of substrates and inhibitors. The kinetic data can be interpreted in light of the caspase residues in the loops of the S5 pocket interacting with the P5 residue. The caspase-3/LDESD structure showed hydrophobic interactions between the side chains of P5 Leu and Phe250 and Phe252 in the S5 pocket, similar to those reported for P5 Val in caspase-3/VDVAD and consistent with the increased activity (150%) for pentapeptides with hydrophobic P5 Leu compared to tetrapeptides of the same P1-P4 sequence (**Figure 2.2** and [34]). However, the activity was similar for the tetrapeptide DEVD and the corresponding pentapeptide QDEVD, suggesting little preference for polar P5 residues. The putative S5 pocket of caspase-6 shares similarities with those of caspase-3 and -2 since Phe and Lys are present at equivalent positions in loop-4, and Thr is at the P5 anchoring position in loop-3 (**Figure 2.6b**). Moreover, the Phe263 in S5 can form hydrophobic contacts with P5 Leu, while the Lys265 may interact with polar P5 residues like Gln. Hence, caspase-6 was predicted to form favorable interactions with both hydrophobic and polar P5 residues in agreement with the relative activity data, although there is a preference for polar P5 (**Figure 2.2**).

Caspase-7 has the polar residues Gln and Asp at structurally equivalent positions in the S5 pocket (**Figure. 2.6c**) and was predicted to prefer polar residues at P5. However, caspase-7 showed lower activity for P5 Gln and P5 Leu in pentapeptides relative to DEVD suggesting no preference for polar P5 residues (**Figure. 2.2**). This specificity difference of caspase-7 compared to the caspase-2, -3 and -6 correlated with the observed interactions of P5 with the residue at the loop-3 anchoring position. Structural analysis suggests the loop-3 residue Thr243/Ser209 in caspase-2, -3 and -6 is crucial for the recognition and anchoring of P5 residues. In caspase-7, the substitution of an imino acid, proline, at the structurally equivalent position in the loop-3 prevents the formation of the hydrogen bond with the carbonyl of P5, while the nonpolar side chain of proline abolishes the other hydrogen bond with the amide of P5 that anchors this residue in the S5 binding pocket (**Figure. 2.7**). The presence of Pro rather than Ser or Thr explains the lower activity of caspase-7 with P5-containing substrates.

Caspase-8 lacks the long surface loop-4 that forms the S5 pocket in the executioner caspases, while the P5 anchoring residue in loop-3 is proline as in caspase-7 (**Figure. 2.1**). Therefore, caspase-8 strongly prefers the tetrapeptide substrate DEVD compared to the two tested pentapeptides. Comparison of the primary structures of other human caspases reveals that inflammatory caspase-1 and apoptotic initiator caspase-9 also have proline as the S5 residue in loop-3 and hence it is likely that these caspases have no preference for the P5 residue in their substrates.

Our structural and kinetic analysis of the preference of executioner caspases for the P5 position of substrates will help identify the specific protein substrates and apoptotic pathways activated by each caspase. The P1 to P4 substrate recognition sequences of the executioner caspases are similar and hard to differentiate, especially between caspase-3 and -7. The P5 residue will contribute less than the P1 to P4 residues to the overall affinity for caspases and therefore, the differences in the relative kcat/Km are also small. Our studies have shown that the hydrophobicity of the S5 subsite in caspase-3 and -6 is consistent with the relative kcat/Km for cleavage of peptides with hydrophobic or polar P5 residues, while caspase-7 does not strongly interact with P5 of peptides. These findings are valuable in understanding the presence of redundant executioner caspases. Importantly, differences were identified in P5 recognition of caspase-3

and -7, which otherwise recognize very similar substrates, and these differences were correlated with the residue present in the newly defined loop-3 anchoring position and the caspase conformation with pentapeptides rather than tetrapeptides. Such differences in substrate recognition can be exploited in the rational design of selective inhibitors, which has proved a challenge, especially for caspase-3 and -7 [32, 143]. Selective pharmacological inhibitors of specific caspases are desirable to reduce cell death for treatment of diseases characterized by increased apoptosis, such as stroke, heart disease and neurodegenerative diseases.

2.5 Acknowledgements

G.F. and B_F. were supported in part by the Georgia State University Research Program Enhancement award. B_F. was supported by the Georgia State University Molecular Basis of Disease Fellowship. I.T.W. and R.W_H. are Georgia Cancer Coalition Distinguished Cancer Scholars. This research was supported in part by the Georgia Research Alliance, the Georgia Cancer Coalition and the National Institutes of Health award GM065762. We thank the staff at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory, for assistance during X-ray data collection. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

Table 2.1 Kinetic parameters of caspases for peptide substrates

	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m ($min^{-1}\mu M^{-1}$)	Relative k_{cat}/K_m (%)
Caspase-6				
DEVD	28±3	490±60	0.058±0.01	100
QDEVD	92±10	1070±110	0.086±0.01	148
LDEVD	30±3	440±40	0.068±0.01	118
Caspase-3				
DEVD	55±3	44±2	1.3±0.1	100
QDEVD	57±3	54±3	1.1±0.1	85
LDEVD	53±3	28±1	1.9±0.1	146
Caspase-7				
DEVD	77±4	44±2	1.8±0.1	100
QDEVD	63±3	51±3	1.2±0.1	67
LDEVD	58±3	61±3	1.0±0.1	56
Caspase-8				
DEVD	14±1	34±3	0.40±0.02	100
QDEVD	11±1	151±8	0.07±0.01	18
LDEVD	9±1	107±5	0.08±0.01	20

Table 2.2 Crystallographic data collection and refinement statistics

	Caspase3/ Ac-LDESD-CHO	**Caspase7/ Ac-LDESD-CHO**
Space group	$P2_12_12_1$	$P3_221$
a (Å)	67.37	88.64
b (Å)	93.56	88.64
c (Å)	97.62	187.71
β (°)	90	120
Resolution range	50-1.61	50-2.45
Total observations	358,844	175,449
Unique reflections	72,895	28,712
Completeness	90.1 (63.9)[a]	90.3 (57.3)[a]
$\langle I/\sigma(I)\rangle$	15.3 (2.4)	18.5 (2.7)
R_{sym} (%)[b]	7.5 (41.6)	9.5 (38.9)
Refinement statistics		
Resolution range	10-1.61	10-2.45
R_{cryst} (%)[c]	17.4	20.1
R_{free} (%)[d]	22.3	24.8
Mean B-factor (Å^2)	32.4	64.2
Number of atoms		
Protein	3803	3777
Inhibitor	84	84
Water	276	82
r.m.s. deviations		
Bond length (Å)	0.009	0.006
Angles	0.029(Å)[e]	1.35(°)[f]

[a]Values in parentheses are given for the highest resolution shell

[b]$R_{sym} = \Sigma_{hkl}|I_{hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}I_{hkl}$.

[c]$R = \Sigma|F_{obs}-F_{cal}|/\Sigma F_{obs}$.

[d]$R_{free} = \Sigma_{test}(|F_{obs}|-|F_{cal}|)^2/\Sigma_{test}|F_{obs}|^2$.

[e] The angle rmsd in SHELX97 is indicated by distance in Å.

[f] The angle rmsd in CNS is indicated by angle in degree.

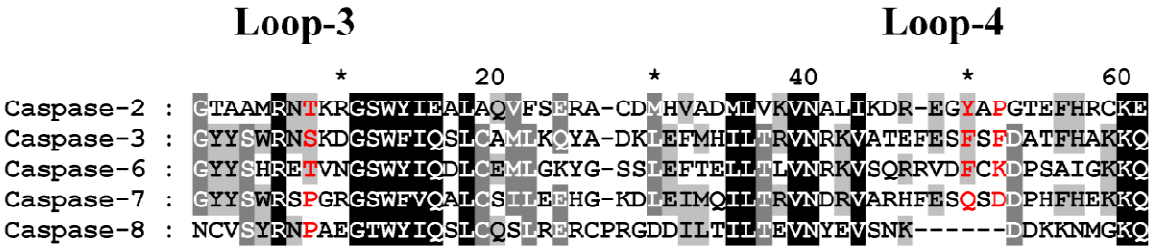


Figure 2.1 Sequence homology among five caspase family members. Only the regions of loop-3 and loop-4 in small subunits are compared. The residues involved in P5 substrate recognition and binding are colored in red.

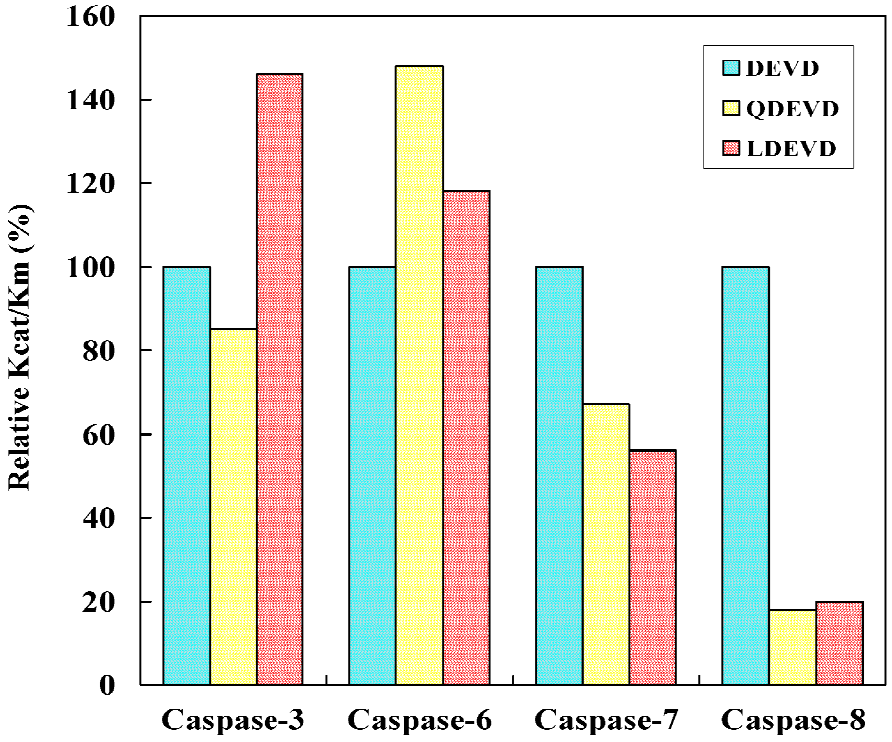


Figure 2.2 Relative activity on substrates. The relative k_{cat}/K_m values are shown for Ac-DEVD-CHO, Ac-QDEVD-CHO and Ac-LDEVD-CHO.

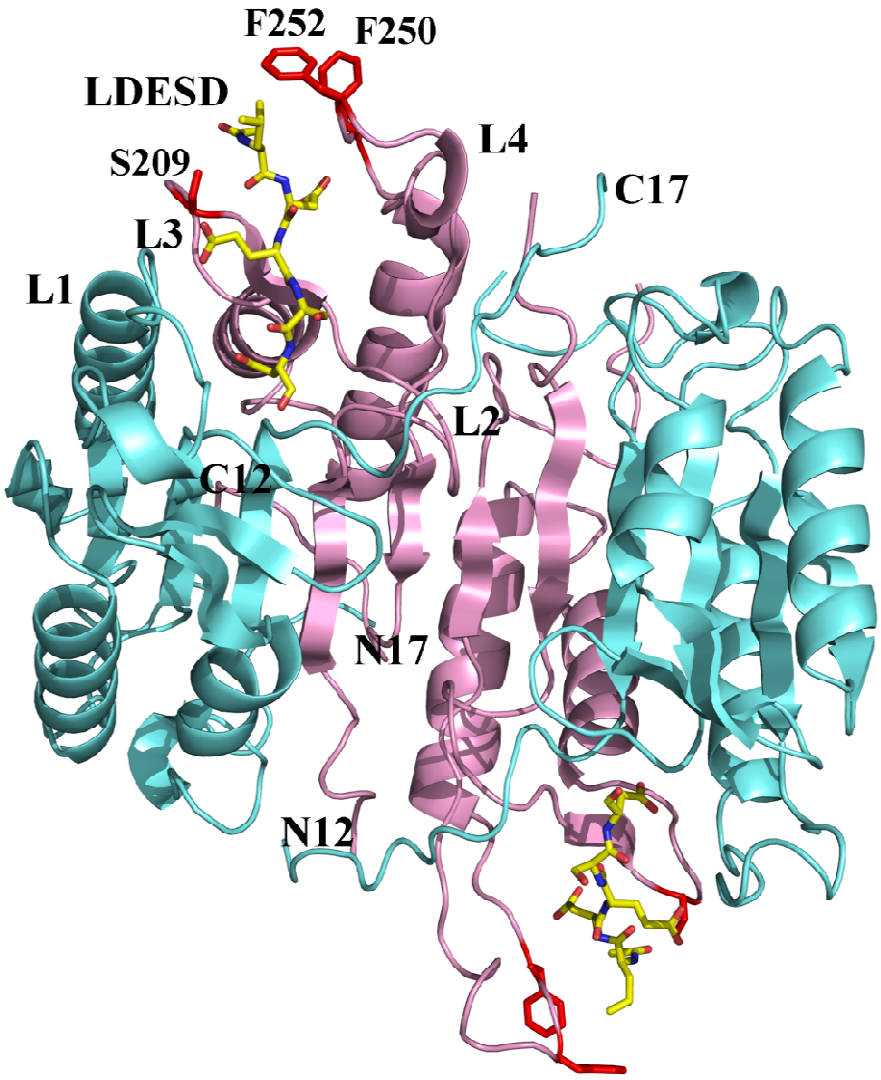


Figure 2.3 Overall structure of caspase-3/LDESD. Two heterodimers $(p17/p12)_2$ of caspase-3/LDESD are shown in a ribbon representation with the large and small subunits colored cyan and pink, respectively. The inhibitor Ac-LDESD-CHO is colored by element type. Side chains of the caspase-3 residues that interact with P5 Leu are colored in red. The N and C termini are indicated for the 12 kDa and 17 kDa chains. L1 to L4 indicate loops 1 to 4 that form the substrate binding site.

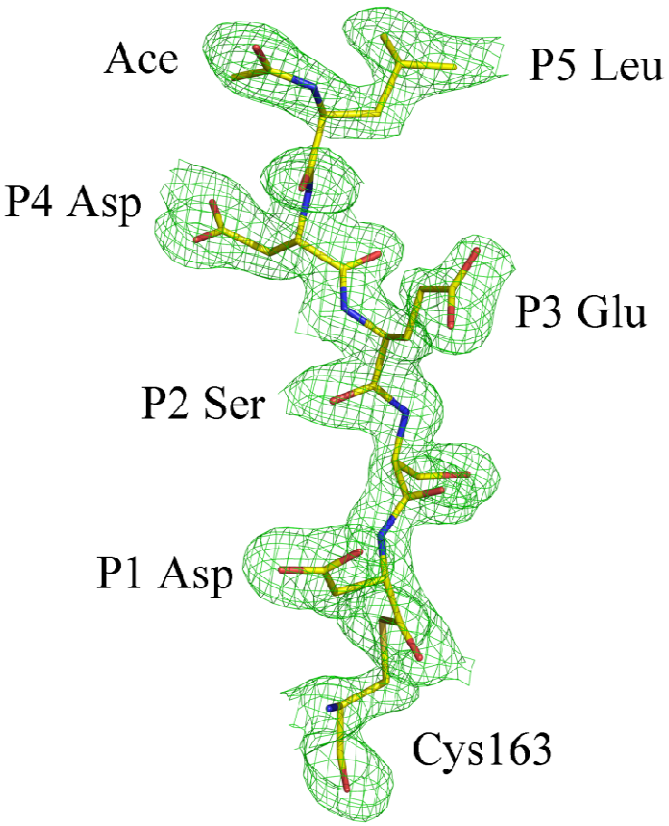


Figure 2.4 Structure of peptide analog inhibitor Ac-LDESD-CHO. 2Fo – Fc electron density map contoured at a level of 1.6 σ for Ac-LDESD-CHO in the caspase-3/LDESD complex. The active site Cys163 of caspase-3 forms a hemithioacetal bond with the aldehyde group of the inhibitor.

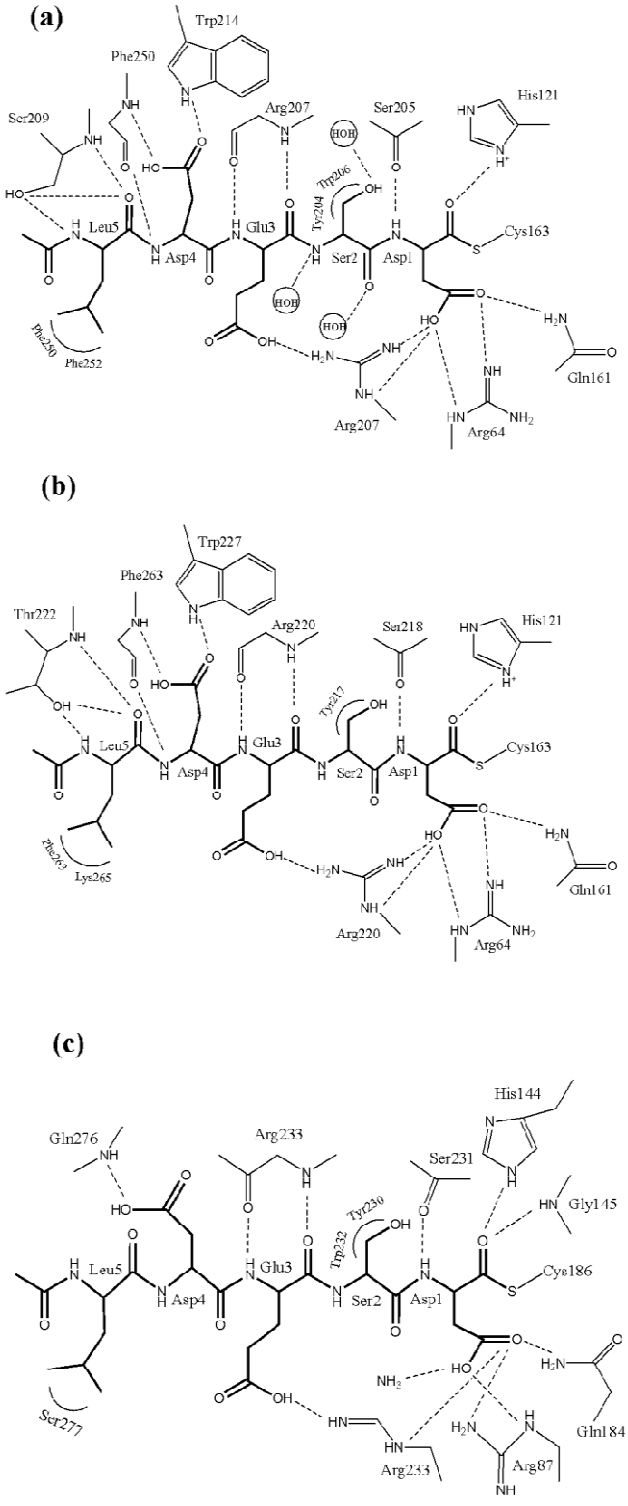


Figure 2.5 Interactions of caspase-3, 6 and 7 with Ac-LDESD-CHO. (a) Caspase-3/LDESD. (b) Caspase-6/LDESD. (c) Caspase-7/LDESD. Thicker lines represent the peptide analog inhibitor. The inhibitor is covalently bonded to the catalytic cysteine. Dashed lines represent hydrogen bonds and ion pairs, while curved lines indicate van der Waals interactions.

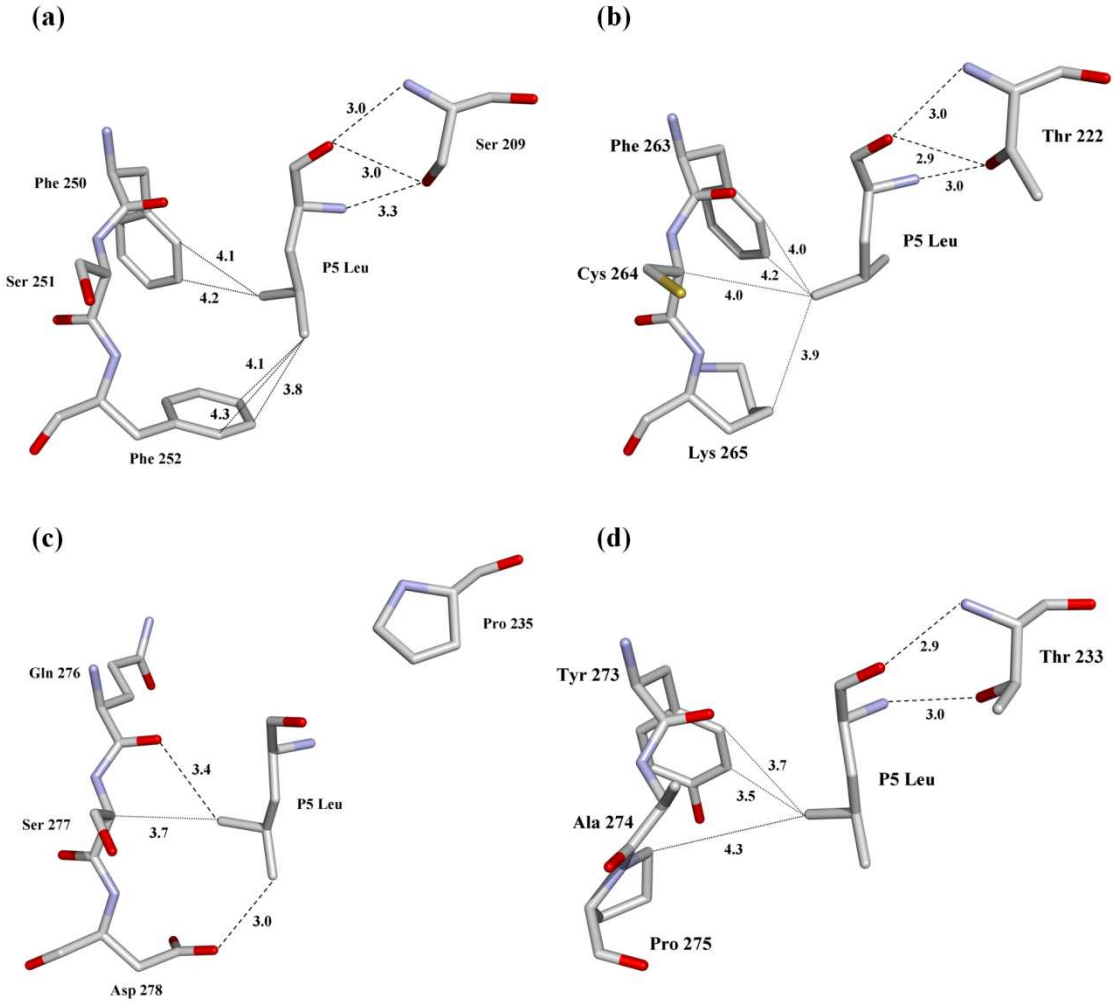


Figure 2.6 Interactions of P5 Leu in S5 subsite of caspases. (a) caspase-3/LDESD. (b) caspase-6/LDESD. (c) caspase-7/LDESD. (d) caspase-2/LDESD. The atoms are colored by element type. Hydrogen bond interactions are represented by broken lines and the van der Waals contacts are shown in dotted lines.

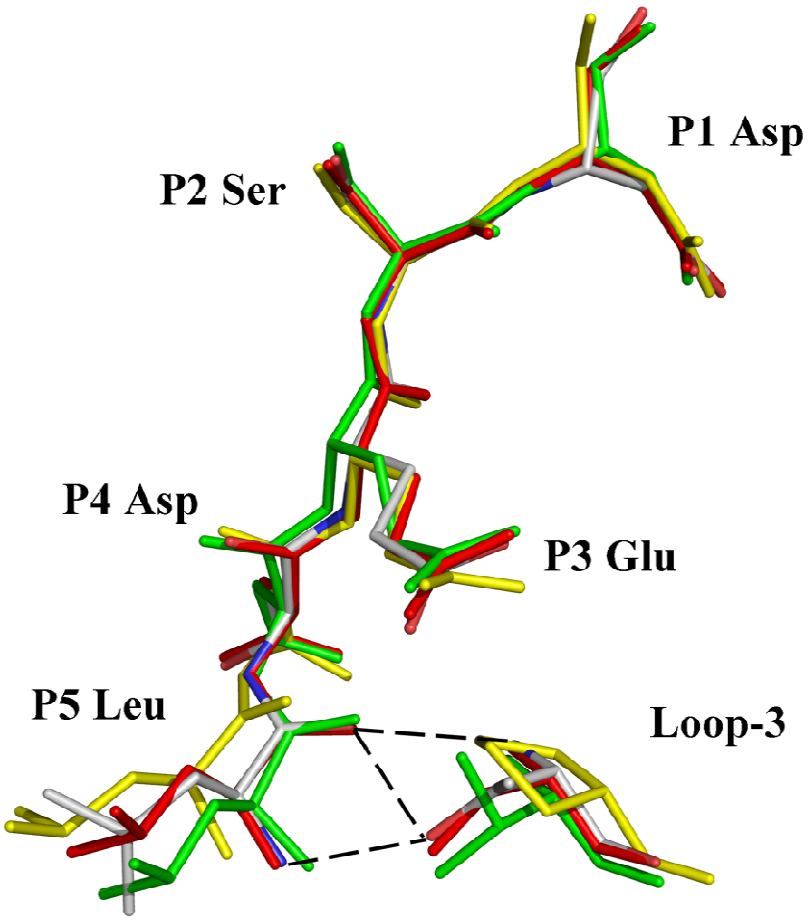


Figure 2.7 Superposition of pentapeptide substrate analogs in the caspase active site. The inhibitor and the critical loop-3 residue that fixes the P5 position in structures of caspase-2, -3, - 6 and -7 complexes with LDESD are shown in green, red, element type and yellow, respectively. Hydrogen bonds between P5 and loop-3 residue (in caspase-2, -3 and -6) are indicated by broken lines. The substitution of Pro in the loop-3 of caspase-7 results in loss of crucial hydrogen bonds that fixes the P5 residue.

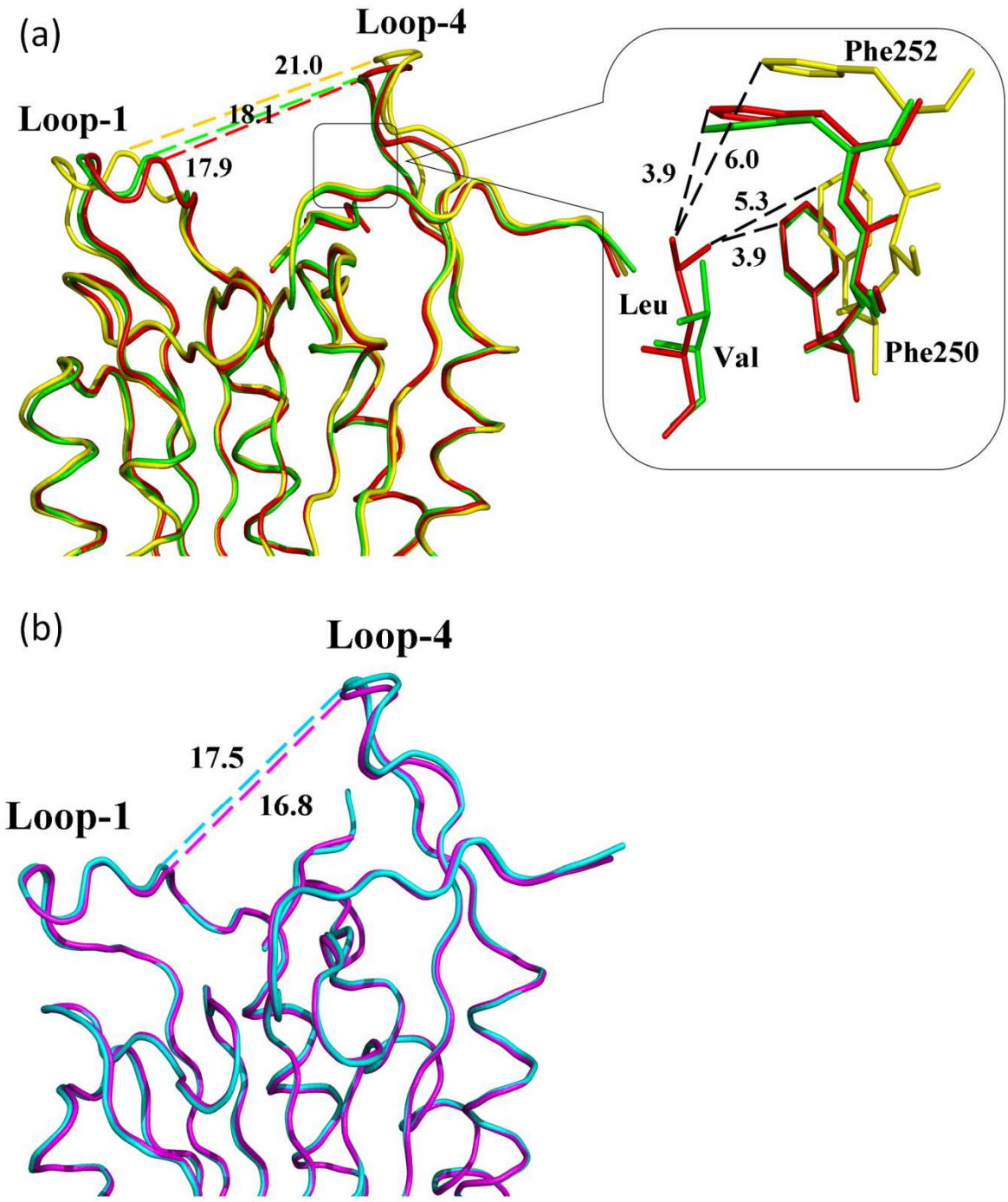


Figure 2.8 Comparison of caspase comlexes with pentapeptides and tetrapeptides. (a) Superposition of Cα backbone of caspase-3 with LDESD (red), VDVAD (green) and DEVD (yellow). Distances between atoms are shown by broken lines. Conformational change is indicated by the altered distances between loop-1 and loop-4. The S5 binding site (boxed region) is shown in detail, where Phe250 and Phe252 form hydrophobic interactions with the P5 residues. (b) Superposition of Cα backbone of caspase-7 with LDESD (cyan) and DEVD (magenta). No significant structural change is observed in the two complexes of caspase-7.

3 CASPASE-3 BINDS DIVERSE P4 RESIDUES AS REVEALED BY CRYSTALLOGRAPHY AND STRUCTURAL MODELING

(Published: Fang B, Fu G, Agniswamy J, Harrison RW, Weber IT. *Apoptosis.* 2009; 14(5):741-752.)

3.1 Abstract

Caspase-3 recognition of various P4 residues in its numerous protein substrates was investigated by crystallography, kinetics, and calculations on model complexes. Asp is the most frequent P4 residue in peptide substrates, although a wide variety of P4 residues are found in the cellular proteins cleaved by caspase-3. The binding of peptidic inhibitors with hydrophobic P4 residues, or no P4 residue, is illustrated by crystal structures of caspase-3 complexes with Ac-IEPD-Cho, Ac-WEHD-Cho, Ac-YVAD-Cho, and Boc- D(OMe)-Fmk at resolutions of 1.9–2.6 Å. The P4 residues formed favorable hydrophobic interactions in two separate hydrophobic regions of the binding site. The side chains of P4 Ile and Tyr form hydrophobic interactions with caspase-3 residues Trp206 and Trp214 within a non-polar pocket of the S4 subsite, while P4 Trp interacts with Phe250 and Phe252 that can also form the S5 subsite. These interactions of hydrophobic P4 residues are distinct from those for polar P4 Asp, which indicates the adaptability of caspase-3 for binding diverse P4 residues. The predicted trends in peptide binding from molecular models had high correlation with experimental values for peptide inhibitors. Analysis of structural models for the binding of 20 different amino acids at P4 in the aldehyde peptide Ac-XEVD-Cho suggested that the majority of hydrophilic P4 residues interact with Phe250, while hydrophobic residues interact with Trp206, Phe250, and Trp214. Overall, the S4 pocket of caspase-3 exhibits flexible adaptation for different residues and the new structures and models, especially for hydrophobic P4 residues, will be helpful for the design of caspase-3 based drugs.

3.2 Introduction

Caspases are key effectors in the processes of apoptosis and inflammation. The family of cysteine aspartyl proteases comprises 14 mammalian caspases that are divided into three groups according to their

roles in the apoptotic and inflammatory pathways. The first group consists of inflammatory caspase-1, 4, 5, and 11 [144]. They play essential roles in cytokine maturation and inflammatory responses. The second and the third groups are involved in the apoptotic process. Caspase-2, 8, 9, 10, and 12 are called initiator caspases because they are located on the upstream of the signaling pathway. They initiate the apoptotic cascade upon receiving interior or exterior death signals. Caspase-3, 6, 7, and 14 are downstream caspases that are activated by initiator caspases during apoptosis. The mature proteins are called executioner caspases because they selectively hydrolyze cellular proteins in the pathways leading to cell death. Several hundred different proteins have been found to be caspase substrates, as reviewed in [31-32].

Caspase activity is highly regulated in the cell. Some natural caspase regulators have been identified, such as caspase recruitment domains (CARD) and inhibitor of apoptosis proteins (IAPs), however, the entire regulatory machinery is very complicated and not yet fully understood. Dysregulation of caspase activity is associated with many severe human diseases. For example, neuronal crush injury, stroke, and neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's diseases are associated with increased activities of caspases [111-113]. On the other hand, suppression of caspase activity is implicated in cancer, autoimmune diseases and viral infections [29, 119, 145]. The development of caspase inhibitors as potential drugs has attracted great attention. Several non-peptide caspase inhibitors are currently in clinical trials, such as IND6556 and VX-740 [146-147].

Caspase recognition of substrates has been investigated in order to identify the caspase-mediated pathways leading to cell death as well as to facilitate drug development. Early studies on peptides showed that caspases recognize at least four residues from P1 to P4, and they exclusively require aspartic acid at P1 in their substrates. From P2 to P4, different caspases have distinct preferences. Caspase-1, 4, and 5 were shown to prefer the tetrapeptide sequence WEHD. Caspase-2, 3, and 7 prefer DEXD, whereas, caspase-6, 8, and 9 prefer (L/V)EXD [10, 148]. However, a series of recent studies have questioned these preferences. For example, the application of substrate-phage approach implied that DLVD was hydrolyzed up to 170% faster than canonical caspase-3 substrate DEVD [149]. A computer based virtual screen found a potent caspase-3 inhibitor Ac-DNLD-Cho with comparable inhibition to Ac-DEVD-Cho. Simi-

larly, Ac-DFPD-Cho was found to be the best tetrapeptide inhibitor for caspase-7 [150]. Moreover, a non-polar P5 residue was shown to facilitate substrate recognition of caspase-2 and 3, while polar P5 is preferred by caspase-6 [34, 151]. Considering the number of native caspase substrates identified with diverse non-canonical cleavage site sequences [152], it is apparent that the substrate selectivity of caspases is far from well understood.

Caspase-3 is one of the critical executioner caspases. In our previous study, the S2, S3 and S5 pockets of caspase-3 were analyzed using aldehyde peptidyl inhibitors Ac-DEVD-Cho, Ac-DMQD-Cho, and Ac-VDVAD-Cho. Analysis of the P5 interaction in S5 was extended to the Ac-LDESD-Ac complex [153]. The complexes with non-canonical P4 residues in the S4 pocket of caspase-3 have not been analyzed, due to the assumption that P4 Asp was critical for substrate binding [35]. Although P4 Asp is common, many other amino acids can be found at the P4 position in the known protein substrates of caspase-3, including hydrophobic residues [31]. Evidently, P4 Asp is not essential for substrate recognition and hydrolysis. In addition, the analysis of non-polar residues in the S4 pocket is important for inhibitor and drug development, because hydrophobic moieties provide better cell permeability in drugs. Therefore, the binding of P4 residues was studied with the dual aims of helping to identify natural caspase-3 substrates and to design new inhibitors and drugs. The plasticity of the S4 pocket was studied in four caspase-3 crystal structures with inhibitors Ac-IEPD-Cho (IEPD), Ac-WEHD-Cho (WEHD), Ac-YVAD-Cho (YVAD), and Boc-D(OMe)-Fmk (BocD). This crystallographic analysis was combined with energy calculations on structural models of P4 residues as well as enzyme inhibition data, and suggested that S4 pocket of caspase-3 can accommodate a variety of residues with different affinities.

3.3 Materials and Methods

3.3.1 Plasmids and Recombinant Proteins

The cloned full length human caspase-3 cDNA was expressed in *E.coli* BL21(DE3). The over-expressed protein was purified using nickel affinity chromatography, ion exchange chromatography, and

gel-filtration chromatography as previously described [34]. Protein was concentrated to 4mg/ml and stored at -80°C. The purity was determined to be over 99% by SDS-PAGE.

3.3.2 Enzyme Kinetic Assays

The catalytic activities of three caspase-3 mutants were determined using the colorimetric caspase-3 substrate Ac-DEVD-pNA (Biomol, PA), where Ac is the acetyl group and pNA is *p*-nitroanilide. Mutant caspase-3 was preincubated in reaction buffer (50 mM HEPES, 100 mM NaCl, 0.1% CHAPS, 10% glycerol, 1mM EDTA and 10 mM dithiothreitol, pH 7.5) at room temperature for 5 mins prior to the addition of substrate at various concentrations. *p*-nitroanilide released by enzyme cleavage was measured at a wavelength of 405 nm using a Polarstar Optima microplate reader (BMG Labtechnologies, NC). SigmaPlot 9.0 (SPSS Inc. IL) was used to obtain the K_m and V_{max} values by fitting reaction velocities as described [154].

The inhibition constants of reversible aldehyde caspase-3 inhibitors Ac-IEPD-Cho, Ac-WEHD-Cho, and Ac-YVAD-Cho were measured using the same protocol as previously described [34]. Briefly, caspase-3 was preincubated with its inhibitors in reaction buffer at room temperature for 30 mins. Then, substrate was added and reaction velocity was calculated according to substrate cleavage. The inhibition constants of each inhibitor were determined by a dose-response curve described by the equation: $K_i = (IC_{50}-0.5[E])/(1+[S]/K_m)$, where [E], [S] and IC_{50}, respectively, correspond to active enzyme concentration, substrate concentration, and the inhibitor concentration needed to suppress half enzyme activity [155]. The inhibition profile of irreversible inhibitor BocD was obtained using the same method and 30 mins preincubation. Time dependence of the initial velocity was measured using 300nM of caspase-3 and a constant amount (2μM) of BocD after preincubation for times from 0 to 60 mins.

3.3.3 Crystallographic analysis

The inhibitors were dissolved in dimethylsulfoxide. Caspase-3 was incubated at room temperature with the inhibitor at 10 to 20-fold molar excess. Crystallization was performed by the hanging-drop vapor diffusion method. Specifically, 1 μl of protein solution (4 mg/ml) was mixed with an equal volume

of mother liquid (100 mM sodium citrate, 5% glycerol, 10 mM dithiothreitol, and 14-18% PEG6000, pH 6.5). Crystals grew within 24 hrs at room temperature. The crystals were frozen in liquid nitrogen with 20% glycerol as a cryoprotectant. X-ray diffraction data were collected on the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory.

The diffraction data were processed with HKL2000 [156]. The structures were solved by molecular replacement with the program AmoRe [157]. The structure of caspase-3/DMQD (PDB code 2H5J) was used as the initial model for all the structures in this study. All the structures were refined using CNS [158]. The molecular graphics program O 8.0 [159] was used to display the electron density map and to refit structures. Water molecules and alternate conformations of caspase-3 residues were modeled when observed in the election density maps. Structural figures were made by Weblab viewer pro (Accelrys Inc., MA) and images of electron density map were obtained using Molscript [160-161]. The crystal structures have been deposited in the RCSB Protein Data Bank with accession codes 3GJT for caspase-3/IEPD, 3GJQ for caspase-3/WEHD, 3GJS for caspase-3/YVAD, and 3GJR for caspase-3/BocD.

3.3.4 Molecular Modeling

Models of caspase-3 complexes with inhibitors of the form Ac-XEVD-Cho, where X is one of the 20 amino acids, were built using the program AMMP[162]. The models were generated from the crystal structure of caspase-3/DEVD [34], which comprises the DEVD inhibitor bound to one large and one small subunit of a caspase-3 heterodimer. The other P4 residues were introduced into the inhibitor using Coot [131]. The hydrogen atoms were built with the sp4 potential set using 15 cycles of the analytic model building algorithm in AMMP [140, 163], followed by conjugate gradients minimization of the non-bonded and geometrical terms. Finally, each caspase-3/inhibitor complex was optimized by 1000 steps of conjugate gradients energy minimization. The total non-bonded interaction energy was calculated for each caspase-peptide model. Structural models were examined with the molecular graphics program RasMol [164] on Linux PCs. The experimental binding energy was calculated as $\Delta G = RT\ln Ki$ from the seven measured inhibition constants for correlation with the calculated interaction energies. The regres-

sion line from this correlation was used to predict Ki values for the peptides with other P4 residues from the peptide-caspase interaction energies calculated by the modeling program.

3.4 Results

3.4.1 Analysis of P4 Residues in Known Caspase-3 Substrates

The frequency of occurrence of different P4 amino acids was analyzed for 182 natural substrates of caspase-3 listed in [31]. Almost all amino acids can be found at least once at the P4 position as shown by the relative occurrence (**Figure 3.1**). Aspartic acid was the most frequent residue in this position appearing in 55% of the substrate cleavage sites. The other cleavage sites had sixteen different residues at the P4 position, comprising 29% polar and 16% nonpolar residues. Glu and Ser are ranked second and third in frequency after Asp, appearing in 10% and 9 % of cleavage sites, respectively. Hydrophobic residues appear in the relative frequency of Val > Ala > Leu > Met > Phe > Ile/Cys. None of the 182 proteins had Arg, Lys, or Gln at P4, so these positively charged or polar amino acids are likely to be rare in caspase-3 cleavage sites. Therefore, in normal physiological conditions, the S4 pocket of caspase-3 can accommodate a variety of residues including the preferred negatively charged Asp, small polar and a variety of hydrophobic side chains.

3.4.2 Inhibition Constants of Caspase-3 Inhibitors

Seven substrate analog reversible inhibitors of caspase-3 are commercially available. They were divided into two groups based on their P4 residues. DEVD, DQMD, and DMQD are caspase-3/7 inhibitors with the standard motif of DxxD; while IEPD, ESMD, WEHD, and YVAD possessing different P4 residues are optimal inhibitors of initiator caspase-8 and inflammatory caspases. The kinetic study was performed using the canonical caspase-3 substrate Ac-DEVD-pNA with the Km value determined as 62.5±2.5 μM. The inhibition constants of the seven inhibitors against caspase-3 are listed in **Table 2**. The first group of inhibitors with the standard P4 Asp exhibited potent inhibition with Ki values from 1.2 nM for DEVD to 12 nM for DMQD, whereas the second group of inhibitors had Ki values approximately a

thousand times weaker. In the second group, IEPD was the most potent with the Ki value of 1.2 μM, while YVAD was the weakest with Ki of 10 μM. The inhibition values are comparable with those of previous studies with fewer inhibitors showing Ki values of 0.2-1nM for DEVD, 195 nM for IETD, 1.9 μM for WEHD and 10-12 μM for YVAD [33, 125, 165].

BocD is a pan-caspase irreversible inhibitor widely used for in vivo studies. Its inhibitory profile was compared with IEPD as shown in **Fig. 3.2a**. At low inhibitor concentration, the enzyme activity was insensitive to the inhibitor. However, a substantial decrease of enzyme activity was observed when the inhibitor concentration was raised to 1 μM. The sigmoidal dependence of inhibition upon inhibitor concentration suggests that BocD has a cooperative or two-state binding as described for another irreversible inhibitor in [166]. The two-states may reflect different binding modes of BocD in the same site, or binding to the two active sites in the caspase heterotetramer. The inhibition increased when the BocD was incubated longer with the caspase-3 before the activity was assayed (**Figure 3.2b**). The time dependency suggests that BocD acts as a slow binding inhibitor.

3.4.3 Overall Structures of Four Caspase-3 Complexes

Caspase-3 was crystallized in complex with three tetrapeptide aldehyde inhibitors, IEPD, WEHD, and YVAD, as well as one modified monopeptide inhibitor BocD. The crystallographic data and refinement statistics are summarized in **Table 3.1**. The proteins were crystallized in two space groups: primitive orthorhombic space group $P2_12_12_1$ and primitive monoclinic space group $P2_1$. The four structures were refined to the resolutions of 1.9-2.6Å with R factors of 21.4-24.4%. All the four complexes consist of two p17/p12 heterodimers in their asymmetric units. The main chain conformations of caspase-3 in the four complexes are essentially identical and they closely resemble our previously reported structure of caspase-3\DEVD (2H5I) with overall rmsd of 0.41-0.49 Å for Cα atoms.

The three tetrapeptide inhibitors were bound in extended conformations in the S1-S4 substrate binding sites of caspase-3 (**Figure 3.3a**). The *2Fo-Fc* electron density map of inhibitor YVAD is shown in **Figure 3.3b**. A thiohemiacetal bond was formed between the aldehyde group (-CHO) of the inhibitor

and the mercapto group (-SH) of the catalytic site Cys163. The main chain atoms of all three inhibitors shared similar positions and their Cα atoms superimposed very well with average rmsd of 0.27 Å for four Cα atoms. On the other hand, significant conformational variations were observed when side chain atoms were compared. Compared with the canonical inhibitor DEVD, the three weaker inhibitors differ mainly at P2 and P4 positions, and the structural differences will be explained in the later sections.

Caspase-3 was crystallized with the inhibitor BocD in order to study the unliganded S4 pocket. BocD was observed in a unique mode of binding. The *Fo-Fc* map in the substrate binding site clearly indicated that this inhibitor only occupied the S1 and S2 pockets of caspase-3 (**Figure 3.3c**). A zwitterionic intermediate was formed between the fluoromethyl ketone (-FMK) on the inhibitor and the mercapto group (-SH) of Cys163. Although it has been claimed that this inhibition could be reversed in certain conditions, such as high DTT (1,4-dithio-threitol) [167], this class of inhibitors is conventionally considered to be irreversible caspase inhibitors in physiological conditions.

3.4.4 Inhibitor Interactions in the S1-S3 Subsites

The peptidic inhibitors formed a series of hydrogen bond and ionic interactions in the caspase active site cavity, as shown in **Figure 3.4**. The interactions in each subsite are described separately. Caspases are known to have a stringent requirement for aspartic acid at the P1 position of their substrates. In agreement, the P1 Asp of the three tetrapeptide inhibitors bound in the S1 pocket in very similar conformations. The major interactions in S1 include both ionic bonds and hydrogen bonds formed between P1 Asp and Arg64, Arg207, Gln161, and Ser205. Although P1 Asp is critical for tight binding of peptides, its negative charge is undesirable in potential drugs. The inhibitor BocD shows one way to improve cell permeability by methylation of one side chain oxygen atom of the P1 Asp. The methyl-Asp occupied the S1 pocket of caspase-3 in a comparable conformation to the P1 Asp of the canonical inhibitor DEVD (**Figure 3.4a**) except that the methyl group extended into the pocket towards Gln161. The neutralized P1 group has hydrogen bond interactions instead of ionic interactions with the positively charged Arg64 and

Arg207 (**Figure 3.4d**). From the protein perspective, the conformations of S1 residues were conserved in all four complexes.

The S2 subsite is formed by three hydrophobic residues, Tyr204, Trp206, and Phe256. Its plasticity was explored using the residues Pro and His, as well as the hydrophobic moiety Boc (tert-butoxycarbonyl) in the new structures. In the structure of caspase-3/IEPD, the P2 Pro neatly fitted into the S2 subsite. Although its side chain is smaller than Val in the canonical DEVD, three S2 residues shifted their side chains towards the P2 Pro reducing the size of the S2 pocket. Consequently, proline has favorable van der Waals interactions with these three S2 residues (**Figure 3.5b**). Compared with proline, histidine was not expected to be an attractive P2 residue due to its bulky side chain. However, in the structure of caspase-3/WEHD, the P2 histidine was accommodated in the S2 pocket with its aromatic ring pointing into the solvent (**Figure 3.5c**). All the side chain carbon atoms formed van der Waals interactions with the three hydrophobic S2 residues. Compared with P2 valine of canonical DEVD, the Cα atom of histidine moved a little further from the S2 residues, which was likely due to the larger size of the histidine side chain. In the caspase-3/BocD complex, the hydrophobic Boc group occupies the equivalent P2 position. Unexpectedly, the electron density map clearly showed that the Boc group was directed out of the S2 subsite, (**Figure 3.3c; 3.5d**). Although the terminal carbon atoms of the Boc group had interactions with Tyr204, the interactions with Trp206 and Phe256 were lost. This analysis suggests that the S2 subsite of caspase-3 prefers small hydrophobic moieties like Val or Pro, and the Boc group is too bulky for optimal fit.

The primary interactions in the S3 pocket are the main chain hydrogen bonds formed between the P3 residue and Arg207. They are conserved independent of the type of amino acid at P3. In these structures of caspase-3 with IEPD, WEHD, and YVAD, Glu and Val formed similar main chain interactions in the S3 pocket as described in an earlier study [34]. Therefore, P3 Val can be accommodated in the S3 subsite, although earlier binding studies showed that caspase-3 preferred hydrophilic residues, with the highest affinity for P3 Glu in peptides [148]. The side chain of P3 Glu formed hydrogen bond interactions with the side chain hydroxyls of Ser65 and Ser209 in caspase-3/WEHD, however, the P3 Glu in IEPD

only retained the interaction with Ser65 while the side chain of Ser209 rotated to interact with its neighbor Lys210. A water mediated hydrogen bond was formed between P3 Glu and Arg207 in both structures. In contrast, the side chain of P3 Glu in the canonical inhibitor DEVD formed a direct hydrogen bond with Arg207 but had no hydrogen bonds with Ser65 and Ser209 (**Figure 3.5e**). These differences imply that the interactions of the P3 side chain depend on the type of amino acid at positions P2 and P4, however, further studies are necessary to address this question.

3.4.5 S4 Subsite

In our previously reported structure of caspase-3/DEVD, the P4 Asp interacted with the main chain atoms of Phe250 and Asn208 through its side chain oxygen atoms and main chain amide group. Those polar interactions tightly anchored the P4 Asp in the S4 subsite. In contrast, the three hydrophobic P4 residues in the new structures showed no direct hydrogen bonds with S4 residues, except for the hydrogen bond between the amide of P4 Ile in IEPD and the carbonyl oxygen of Phe250 (**Figure 3.4a**). In caspase-3/YVAD a water mediated hydrogen bond interaction was found between the side chain hydroxyl of P4 Tyr and the main chain atoms of Gln248. Despite the paucity of polar interactions, the three P4 residues formed favorable hydrophobic interactions in two separate regions of the S4 pocket. In the structure of caspase-3/IEPD, the side chain of Ile was rotated slightly into the S4 subsite compared with the conformation of P4 Asp in caspase-3/DEVD, and fitted in a non-polar pocket formed by the aromatic side chains of Trp206 and Trp214 (**Figure 3.5f**). Similarly favorable van der Waals interactions between P4 Ile and two Trp residues were also observed in the complex of caspase-7/IEPD [168]. Also, the S4 subsite of initiator caspase-8 is formed by two aromatic residues, Trp420 and Tyr412, consistent with the preference for Ile at P4 position. In the complex of caspase-3/YVAD, the P4 Tyr lay in the S4 pocket with its side chain hydroxyl directed out of the substrate binding groove. The P4 Tyr formed favorable van der Waals bonds with Trp214 and Trp206 (**Figure 3.5g**). The terminal acetyl oxygen of YVAD formed hydrogen bonds with the main chain and side chain of Ser209, whereas, the acetyl group of IEPD formed a

water-mediated interaction with the side chain of Ser209, which also contribute to the binding affinity (**Figure 3.6**).

In the complex of caspase-3/WEHD, the P4 Trp binds in the S4 subsite in a unique manner. The entire side chain of Trp and the acetyl group were rotated by about 90° (**Figure 3.5h**). Consequently, the aromatic ring of P4 Trp interacted with the hydrophobic side chains of Phe250 and Phe252 that were proposed to form the S5 pocket of caspase-3 [34]. Because of the rotation, the acetyl oxygen formed a hydrogen bond with the main chain amide of Phe250 instead of the interaction with Ser209 seen in the other complexes (**Figure 3.4b**). Moreover, the interaction of P4 Trp with Phe250 and 252 in the S5 subsite appeared to trigger a closure of about 0.9 Å of the loops 1 and 4, similar to the conformational change and induced fit mechanism proposed in our previous studies of caspase-3 with P5-containing pentapeptides [34].

Although BocD does not possess a P4 residue, a glycerol molecule was found in the S4 subsite in the crystal structure of caspase-3/BocD. The glycerol molecule was positioned to form two hydrogen bonds between glycerol oxygen O1 and O2 and the side chain nitrogen of Trp214 as well as the amide group of Phe250, respectively (**Figure 3.5i**). The binding of glycerol reflects the hydrophilic nature of the S4 subsite. In the absence of P4-containing ligand, S4 is highly accessible for water and other polar solvent molecules.

Overall, the structures showed that the different P4 residues can bind with only small conformational changes in the caspase residues, similar to the observations for caspase-7 [168]. The divergent binding modes of the various P4 residues imply that, although polar residues bind favorably in the S4 subsite, non-polar residues can be accommodated without triggering substantial conformational change. Two distinct hydrophobic pockets provide interactions with hydrophobic P4 in the S4 region. Trp206 and Trp214 form an inner subsite, which is suitable for smaller non-polar side chains while Phe250 and Phe252 form an outer pocket, which is suitable for larger hydrophobic P4 side chains like Trp or P5 residues like Val or Leu. The binding modes of P4 Trp and P4 Tyr in YVAD and WEDH, respectively, illus-

trate the two distinct pockets (**Figure 3.6**). This structural analysis explains why some natural substrates of caspase-3 have hydrophobic P4 resides, such as Ile and Leu in CREB and DCC, respectively [152].

3.4.6 Correlation of Structural Interactions with Inhibition

The three substrate analog inhibitors share the same P1 Asp residue. They exhibit comparable conformations for P1 and have similar interactions with S1 residues. In the next subsite, all three different P2 residues exhibited favorable van der Waals interactions with S2 residues. Therefore, the variations of P3 and P4 are the primary factors influencing inhibitory potency. The three tetrapeptide inhibitors showed weak inhibition compared with canonical inhibitor DEVD (**Table 3.2**). IEPD is the strongest among three with a K_i value of thousand times higher than for DEVD. The weakest inhibitor YVAD is ten-fold weaker than IEPD. Structural analysis suggests that Ile is the most favorable of the three P4 residues because it preserves a main chain hydrogen bond interaction with Phe250. Tyr is the next most favorable since its side chain has a water mediated hydrogen bond with Gln248. The Trp is ranked last because of two reasons. First, it has no polar interactions with S4 residues unlike the other two residues. Second, the binding to the S5 pocket requires rotation of its entire bulky side chain. At the P3 position, Glu is clearly more favorable than Val in the S3 subsite. The positively charged Arg207 and hydrophilic Ser65 firmly anchored the negatively charged Glu in the S3 pocket. On the contrary, the Val side chain has no favorable interaction with S3 residues. Since hydrogen bond interactions are stronger than van der Waals interactions, the variation of P3 is the primary reason for the inhibitory differences of the three inhibitors. Overall, the substitution of P4 Asp with hydrophobic residues resulted in dramatic drop in inhibitory potency. IEPD was the best because P4 Ile has a better fit in S4 than other two and P3 Glu is favorable. YVAD showed the weakest inhibition mainly because of the mismatch of P3 Val in the S3 subsite. Although, WEHD had the least favorable P4 residue, the favorable P3 Glu interaction still made a stronger inhibitor than YVAD. Moreover, the side chain of P3 Glu in WEHD formed a hydrogen bond with Ser209 and a more direct ionic interaction with Arg207 relative to P3 Glu in IEPD. This analysis illustrates the effects

of combining sequence variations at the P4 and P3 positions, which is expected to alter the efficiency of hydrolysis of proteins with different sequences at their cleavage sites.

From the viewpoint of numbers of favorable interactions with caspase-3, BocD should exhibit the lowest binding affinity as reflected in our inhibition assay. The methylated P1 Asp is uncharged and less favorable than the negatively charged Asp for binding the positively charged S1 pocket. The consequence is that a higher concentration of the inhibitor needs to be accumulated in order to effectively block the active site of the enzyme. Hence, the poor performance of BocD observed in the low concentration range. In the higher concentration range, Boc-D-Fmk was a stronger inhibitor than WEHD and YVAD because it can irreversibly inactivate caspase-3.

3.4.7 Predicted Binding of Diverse P4 Residues

The analysis of mode of binding and inhibition of P4 residues was extended to all 20 amino acids by molecular modeling. 20 residues at the P4 position in peptide inhibitor XEVD were modeled in their complexes with caspase-3. The program AMMP [162] was used for modeling with the crystal structure of caspase-3/DEVD (2H5I) comprising one inhibitor bound to one large and one small caspase subunit as the initial template. The structural 49 interactions and the interaction energies were predicted for the peptides in the molecular models of caspase-3/XEVD. Importantly, the predicted conformations of the P4 Ile, Trp and Tyr in the inhibitor XEVD closely resemble those observed in our crystal structures **(Figure. 3.7**).

The accuracy of the energy calculation was examined using the seven commercial caspase tetrapeptide inhibitors with Ki values determined in our caspase-3 activity assay (**Table 3.2**). To our knowledge, this is the most complete set of aldehyde tetrapeptide inhibitors of caspase-3 at present. As shown in the **Figure 3.7a**, the observed free energies (ΔG) derived from experimental Ki values for the seven inhibitors showed excellent correlation (R of 0.83) with the calculated interaction energies for the models. The calculated interaction energy is an estimate of ΔU, which is assumed to dominate the changes in free energy. In these calculations on a single configuration the effects of entropy and energy of solv-

ation are assumed to be small or similar for the series of inhibitors. Therefore, the estimate for ΔU can be used to estimate trends in ΔG for different peptides. The high correlation indicates that our structural models and energy calculations are reliable.

In order to estimate relative binding affinities for all 20 P4 residues in the peptide XEVD, structural models were also made with Asp, Ile, Trp, and Tyr in the P4 position using the same procedure. The predicted Ki values of inhibitors XEVD, which were derived from mapping the calculated interaction energies onto the regression line in **Figure 3.8a**, are shown in **Figure 3.8b**. The P4 Asp showed the highest affinity in the prediction, in agreement with experimental results. The inhibitor with P4 Glu was predicted as the second strongest one. Arg and Asn were predicted to be the next best at P4 in this tetrapeptide, although Lys and His were among the weaker inhibitors. Among the hydrophobic P4 residues, aliphatic residues were predicted to show better binding than aromatic residues. These results are consistent with the previous experimental analysis of caspase-3 substrate preference [148]. The predicted trends in binding affinity showed general agreement with the residue ranking in our analysis of natural substrates (**Figure 3.1**), although deviations appeared for some residues such as Arg and Asn. The predictions for relative inhibition from the energy calculations assume that changes in entropy and solvation energy are negligible. Also, the binding affinity depends on the entire substrate and the models used tetrapeptides with optimal residues at P1–P3 instead of the full-length protein substrates with a variety of residues at these positions. Therefore, experimental data for proteins with different cleavage site sequences are required in order to more fully understand the substrate specificity of caspase-3.

The model complexes were analyzed for the binding mode of the different P4 residues (**Figure 3.9**). Most of the sixteen residues showed similar conformations with interactions depending on the type of side chain at P4. The eight polar P4 residues Arg, Lys, Glu, Asn, Gln, His, Ser and Thr were predicted to form at least one side hydrogen bond interaction with caspase residues. Most of them interacted with the main chain atoms of Phe250 (**Figure 3.9**). The P4 Arg and Lys were predicted to form hydrogen bonds with the main chain of Glu248 and the side chain of Asn208, and P4 His formed a hydrogen bond with Ser209. On the other hand, Pro and six hydrophobic residues Ala, Cys, Val, Leu, Met, Phe interacted

with S4 subsite through van der Waals interactions with distance cutoff of 3.8-4.2 Å. All of them interacted with Trp206, P4 Leu, Met and Phe interacted with Phe250, and P4 Leu also interacted with Trp214. Glycine is not shown in the figure since it does not have a side chain. Most acetate groups on the inhibitors formed hydrogen bond interactions with Ser209, as observed for crystal structures of caspase-3/IEPD and YVAD (**Figure 3.4**). An exception was observed in the model with P4 Phe, where the acetate interacted with Asn208 and Trp214. This change was because the side chain of P4 Phe had rotated towards Phe250 and Phe252 similar to the rotated conformation seen for P4 Trp in the crystal structure of caspase-3/WEHD (**Figure 3.6**).

Independent of the predicted binding affinities, most P4 residues were accommodated without greatly changing the conformation of the S4 pocket. His and Phe were only two P4 residues that showed considerable conformational change when bound to the enzyme, which is consistent with their higher estimated Ki compared with others. Our results imply that the S4 subsite is capable of accommodating various P4 residues. Hydrophilic P4 is most likely to interact with F250 and N208, whereas hydrophobic P4 is most likely to interact with W206 and F250. According to the calculated interaction energy, most substitutions of P4 resulted in a hundred fold decrease in the binding potency compared with inhibitor DEVD. The binding affinity, however, will also depend on the residues present at other positions, such as P3. Further analysis will be needed to explore the effect of variations at other positions in the cleavage sites. Moreover, a substrate protein with a significantly lower binding affinity can still effectively bind to caspase-3 and be hydrolyzed in physiological conditions. That explains why a large number of non-canonical substrate sequences have been found.

3.5 Discussion

Since the original work of Thornberry and Talanian [148, 169], the idea of consensus recognition sequences was established as the basis of the current substrate and inhibitor research of caspases. DxxD was considered the classic substrate recognition motif for caspase-3 and -7 from studies of short peptides. This information, however, is misleading in searching for natural caspase substrates because a substrate

protein does not have to bind with the highest affinity. Numerous natural substrates of caspases have been discovered by *in vitro* kinetic studies, *in vivo* cleavage studies, and mutagenesis studies. Among 59 natural caspase-3 substrates listed in Backes' study, 27 substrates have different P4 residues rather than Asp at their cleavage sites [152]. In fact, 10 out of those 27 substrates possess hydrophobic P4 residues. We have extended the statistics on natural substrates of caspase-3 and found that 45% of 182 reported natural caspase-3 substrates possess residues distinct from Asp, and over a third of them are non-polar residues. It challenges our old knowledge about substrate recognition of caspase-3 suggested by earlier *in vitro* kinetic studies. Crystal structures of caspase-substrate complexes can help us to better understand this problem by providing direct evidence of whether a putative substrate can bind and how it binds.

Our structures with three substrate analog inhibitors have demonstrated the binding modes of non classical residues from P2 to P4 positions. The hydrophobic S2 subsite of caspase-3 was suggested to preferentially recognize aliphatic residues. Nevertheless, some natural caspase-3 substrates have aromatic residues at P2, such as PAPD in CAMK4 (calcium/calmodulin dependent protein kinase IV) and DRHD in NFKBIA (nuclear factor of Kappa light peptide gene enhancer in B-cells inhibitor, alpha). In our structures, the aromatic side chain of His and hydrophobic Pro were both accommodated in the S2 subsite. This observation is consistent with a previous screening study where proline was found to be of the best P2 residues [149]. Furthermore, the pyrrolidine ring has been successfully utilized at P2 position in the development of caspase-3/7 inhibitors, such as the isatin sulfonamide inhibitors [170]. Clearly, the S2 subsite of caspase-3 can accommodate both aliphatic residues and aromatic residues. At the P3 position, our study confirmed previous studies that various residues can bind, among which Glu is the best.

At the P4 position, the three groups of caspases exhibit distinct preferences. Initiator caspases prefer Leu/Ile/Val, inflammatory caspases prefer Trp, while executioner caspases demonstrate a strong preference for Asp at P4 position and any substitution resulted in thousand-fold difference in the binding affinity. This characterization was challenged by several previous studies where caspase-8 was shown to tolerate P4 Asp [171] and caspase-7 was shown to accommodate hydrophobic P4 residues such as Trp [168]. The cleavage sites of many caspase-3 natural substrates, such as YVPD in CDC2L1 (cell division

cycle 2 like 1 protein) and ILND in CREB (cAMP response element-binding protein) underlined the natural capability of caspase-3 to bind diverse P4 residues *in vivo*. Several polar residues have showed tight interactions with S4 in our predictions, including Asp, Glu, Asn and Ser. Although hydrophobic residues showed weaker interactions in the S4, most of them can be accommodated in favorable conformations in our models as well as in the crystal structures. Our previous studies revealed that the S4 subsite of caspase-7 has dual functionality with a hydrophilic area and a hydrophobic area [168]. In the current study, the same feature was also found in the S4 subsite of caspase-3. The side chain of P4 Ile in IEPD and P4 Tyr in YVAD interacted with a hydrophobic pocket formed by Trp206 and Trp214 deep inside the S4 subsite. In contrast, the side chain of P4 Trp in WEHD extended out of the S4 subsite and made contact with the reported hydrophobic S5 residues [34]. The physiological function of hydrophobic S5 was reflected in the crystal structure of XIAP bound with caspase-3 [172], where Ile149 and Ile153 of XIAP interacted with S5 residues. In the absence of structural information for natural caspase substrates, the functions of Trp206 and Trp214 in the S4 subsite were not understood before. Nevertheless, their capability of interacting with non-polar residues was revealed in this study. Van der Waals interactions between hydrophobic P4 residues and Trp206 were also conserved in our structural models. This analysis supports the existence of the inner non-polar pocket of the S4 subsite. The discovery of dual functionality of S4 subsite as well as the binding modes of diverse P4 residues provides the structural basis for binding of non-polar residues in S4 subsite. Small aliphatic non-polar residues such as Ala, Val, and Leu are more likely to bind to the inner pocket of S4. Aromatic residues may need more conformational adjustments. This explains why aliphatic non-polar residues appear more frequently than aromatic ones at P4 in the caspase cleavage sites. Of course, it is likely that caspase-3 recognition of non-polar P4 residues will depend on other conditions, such as early stage binding between enzyme and its targets. In such way, non-polar interactions at P4 can become more favorable when the cleavage site is buried in the active site groove and becomes less exposed to the solvent. This may explain why many natural substrates with non-polar P4 can be cleaved efficiently in physiological conditions. Although the real mechanism may be

more complex, it is clear that some non-polar P4 residues such as Ala, Val, and Leu, must be considered when searching for potential caspase-3 substrates.

Last but not least, the information on binding of non-polar P4 residues to caspase-3 can help the development of new inhibitors and drugs. Current drug design for caspse-3 is mostly limited to the P1, P2 and P3 positions. Introduction of a hydrophobic P4 moiety could not only facilitate the binding, but also improve the cell permeability. Compound IDN-6556 is one such example [173]. Although the structure of its complex with caspase is not available, the aromatic moiety at its N terminus is likely to bind in the S4 of caspase-3 [174]. This oxamyl peptide compound showed inhibition of all caspases and is currently under phase II clinical trials in patients with liver transplants [121]. The results from this study should be considered in the future development of therapeutic compounds. Material and Methods.

3.6 Acknowledgements

B.F. was supported by the Georgia State University Molecular Basis of Disease Fellowship. G.F. was supported in part by the Georgia State University Research Program Enhancement award. I.T.W. and R.W.H. are Georgia Cancer Coalition Distinguished Cancer Scholars. This research was supported in part by the Georgia Research Alliance and the Georgia Cancer Coalition. We thank the staff at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory, for assistance during X-ray data collection. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

Table 3.1 Crystallographic data collection and refinement statistics.

	Caspase3/ Ac-YVAD-Cho	Caspase3/ Ac-IEPD-Cho	Caspase3/ Ac-WEHD-Cho	Caspase3/ Ac-D(Me)-Fmk
Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
a (Å)	50.4	68.2	69.1	50.4
b (Å)	70.3	88.3	88.2	69.7
c (Å)	93.4	96.9	96.6	93.4
β(°)	102.4	90	90	102
Resolution range	50-1.9	50-2.2	50-2.6	50-2.2
Completeness[a]	85.7(52.0)	99.7(99.3)	99.7(99.7)	95.1(75.0)
$\langle I/\sigma(I)\rangle$	13.18(3.0)	21.3(10.7)	12.3(7.0)	11.9(2.8)
R_{sym} (%)[b]	7.3(21.4)	7.6(17.9)	11.2(33.1)	11(30.5)
Refinement statistics				
Resolution range	10-1.9	10-2.2	10-2.6	10-2.2
R_{cryst} (%)[c]	21.6	24.4	23.6	21.4
R_{free} (%)[d]	24.8	28.8	29.0	25.3
Average B factor (Å^2)	25.3	25.5	31.2	20.7
No. protein atoms	3908	3786	3776	3774
No. water atoms	299	134	49	181
Bond length rmsd (Å)	0.007	0.007	0.007	0.007
Angles rmsd (°)	1.3	1.3	1.3	1.3

[a]Values in parentheses are given for the highest resolution shell. [b]$R_{sym} = \Sigma_{hkl}|I_{hkl} - \langle I_{hkl}\rangle|/\Sigma_{hkl}I_{hkl}$. [c]$R = \Sigma|F_{obs}-F_{cal}|/\Sigma F_{obs}$. [d]$R_{free} = \Sigma_{test}(|F_{obs}|-|F_{cal}|)^2/\Sigma_{test}|F_{obs}|^2$.

Table 3.2 Inhibition constants

Inhibitor	K_i (nM)
Ac-DEVD-Cho	1.2 ± 0.05
Ac-DQMD-Cho	11.0 ± 0.5
Ac-DMQD-Cho	12.0 ± 0.5
Ac-IEPD-Cho	170 ± 7
Ac-ESMD-Cho	1.2×10^3 ± 48
Ac-WEHD-Cho	4.7×10^3 ± 190
Ac-YVAD-Cho	10.2×10^3 ± 410

Caspase-3 was preincubated with its inhibitors in reaction buffer (50 mM HEPES, 100 mM NaCl, 0.1% CHAPS, 10% glycerol, 1mM EDTA and 10 mM dithiothreitol, pH 7.5) at room temperature for 15 mins. Substrate was then added and reaction product *p*-nitroanilide was measured at a wavelength of 405 nm. The inhibition constants were determined by using the equation: Ki = (IC50-0.5[E])/(1+[S]/Km), where [E], [S] and IC50, respectively, correspond to active enzyme concentration, substrate concentration, and the inhibitor concentration needed to suppress half enzyme activity. The enzyme concentration was determined by active site titration.

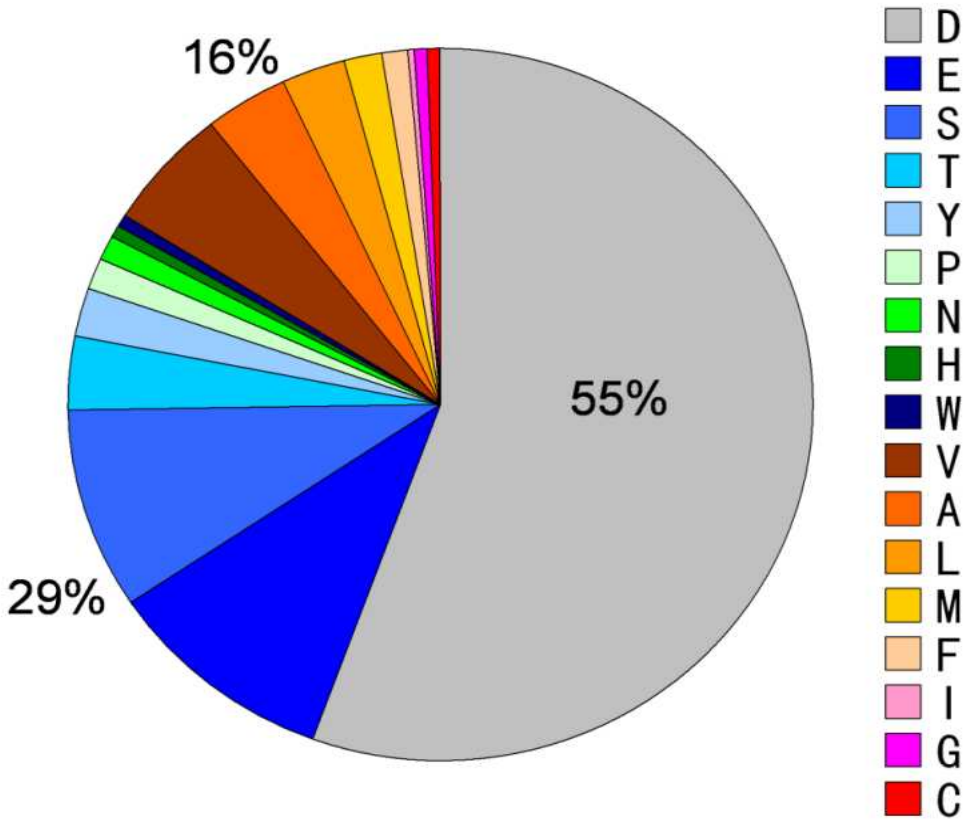


Figure 3.1 The occurrence of different amino acids at P4 in the cleavage sites of natural substrates of caspase-3. Different residues are represented in different colors.

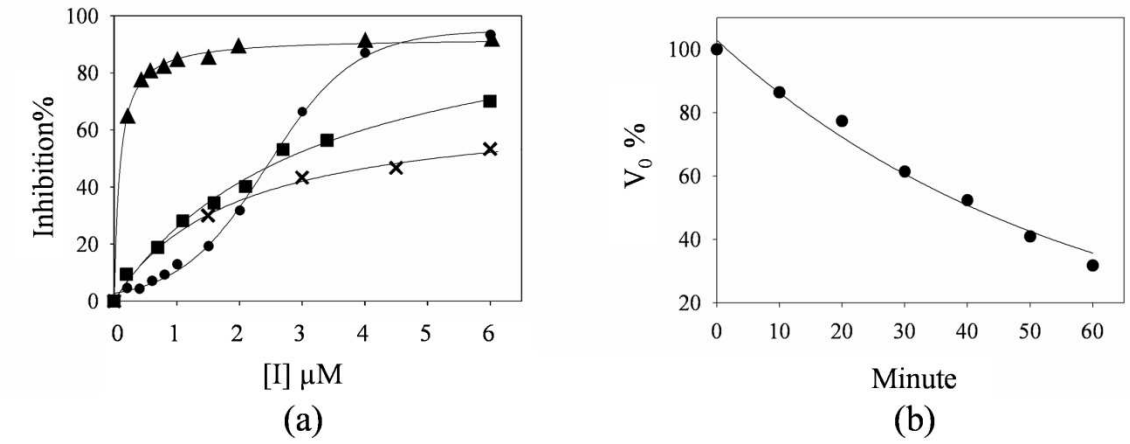


Figure 3.2 Inhibition characterizations of caspase-3 inhibitors. (a) Inhibition profiles of caspase-3 inhibitors IEPD (▲) and BocD (●). The fractional inhibition is calculated as V/V0×100%, where V and V0 refer to velocities with and without inhibitor respectively. Caspase-3 was preincubated with inhibitors in reaction buffer (50 mM HEPES, 100 mM NaCl, 0.1% CHAPS, 10% glycerol, 1mM EDTA and 10 mM dithiothreitol, pH 7.5) at room temperature for 30 mins. Inhibition profile of IEPD was fitted using hyperbola equation: Inh=Inhmax[I]/(Kd+[I]), while inhibition profile of IEPD was fitted using sigmoidal equation: Inh=Inhmin+(Inhmax- Inhmin)/(1+10^(logEC50-[I])). Inh, [I], and EC50 refer to fractional inhibition, inhibitor concentration, and the inhibitor concentration provides the halfway inhibition. The sigmoidal inhibition profile of BocD suggests a cooperative binding mechanism. BocD may bind to an allosteric site on caspase-3, which in turn facilitates the binding of BocD in the active site of the protein. (b) Time dependence of the initial velocity was measured using 300 nM of caspase-3 and 2μM of BocD after preincubation for 0 to 60 mins. Because time dependence inhibition profiles of irreversible inhibitors typically follow exponential decay, data in the graph was fitted into the exponential decay curve defined by equation: V=Vmin+a*exp(-bt), where V and t refer to initial reaction velocity and time, respectively.

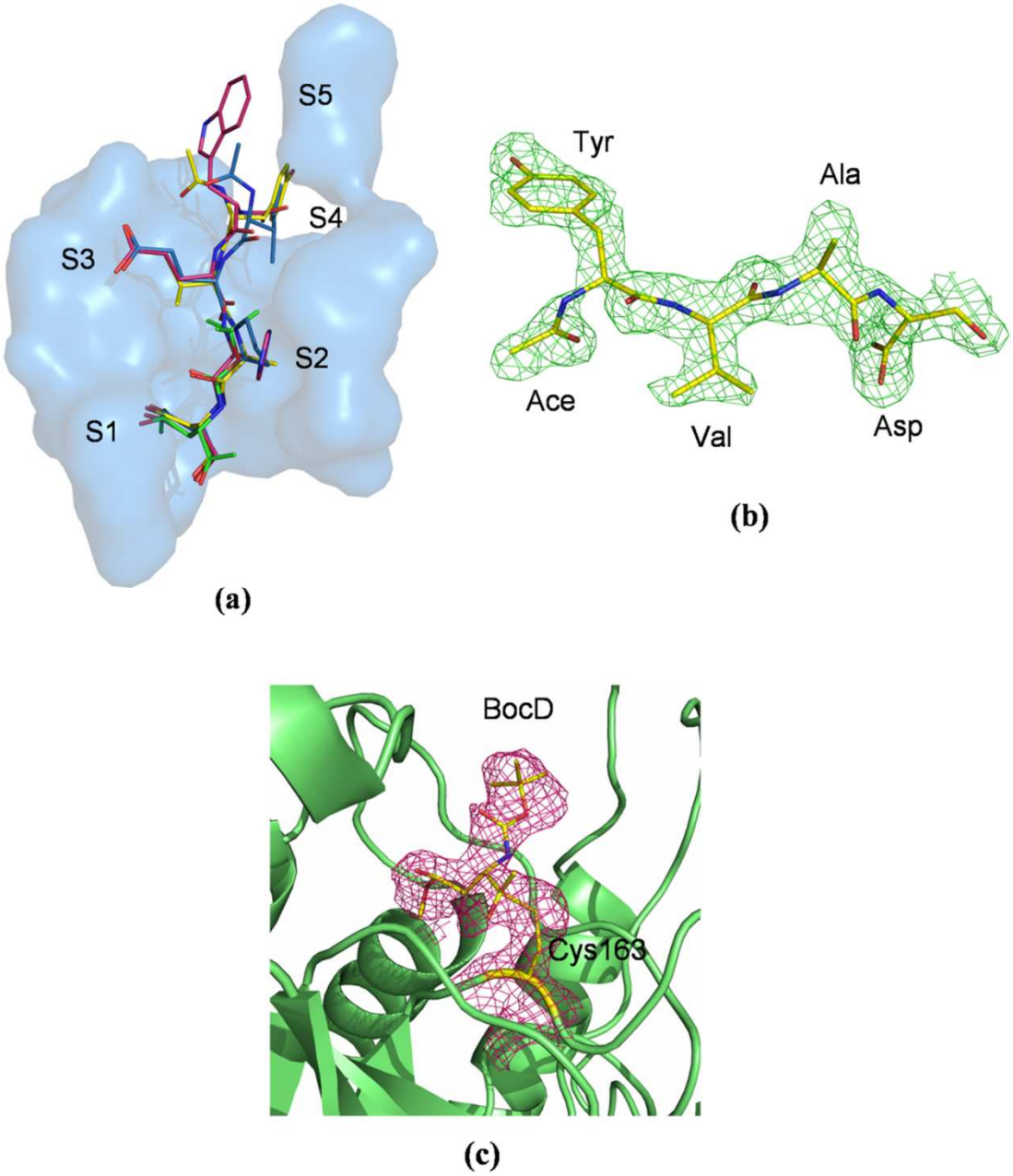


Figure 3.3 Binding conformations of inhibitors. (a) Superposition of four inhibitors in the substrate binding site of caspase-3. Inhibitors IEPD (blue), WEHD (magenta), YVAD (yellow), and BocD (green) are shown in stick representation. The inding surface of S1-S4 subsites is shown in pale blue. (b) 2Fo-Fc electron density map of inhibitor YVAD contoured at a level of 1.4σ. (c) Fo-Fc electron density map of inhibitor BocD in the complex of caspase-3/BocD. The map was contoured at a level of 1.4σ. The inhibitor covalently binds to active site Cys163. The backbone of loops forming the substrate binding site is illustrated as ribbons.

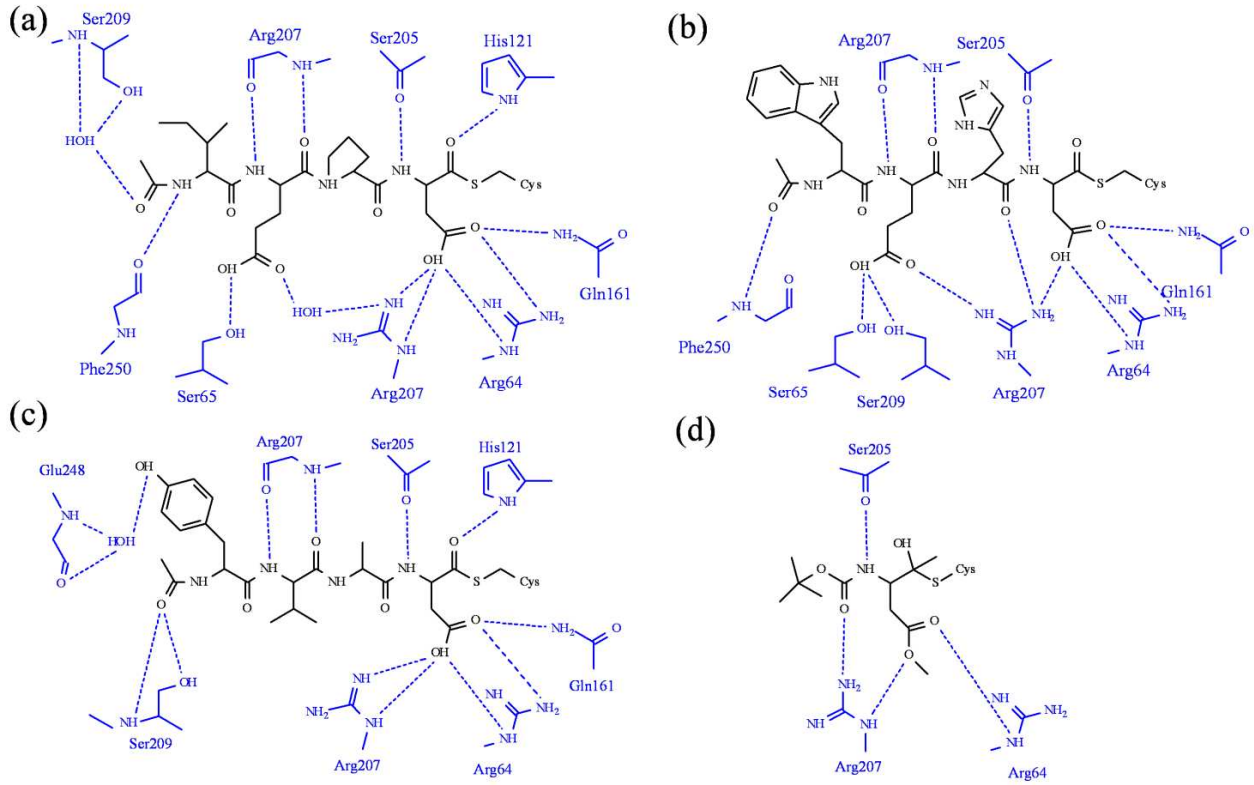


Figure 3.4 Schematic representation of hydrogen bond and ionic interactions between caspase-3 and the inhibitors. (a) IEPD, (b) WEHD, (c) YVAD, and (d) BocD. The protein residues are in blue and inhibitors are in black. Interactions are indicated in dashed lines.

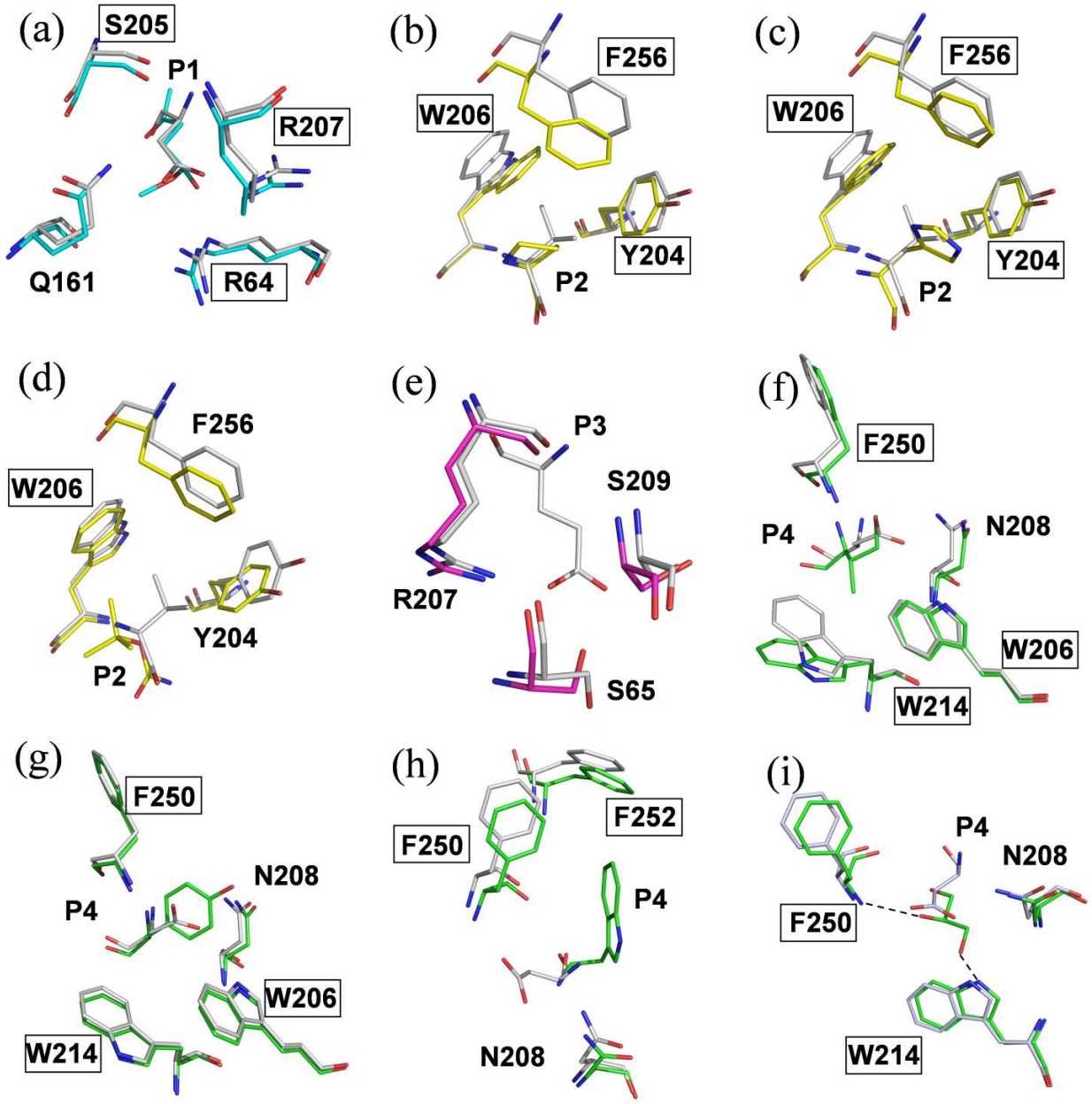


Figure 3.5 Comparison between subsites in four new complexes (color) and published structure of caspase-3/DEVD (grey) (2H5I). (a) The methylated P1 Asp of BocD and S1 residues. (b)-(d) P2 residues Pro, His, and Boc moiety, respectively are shown in S2. (e) The empty S3 subsite of the BocD complex. (f)-(h) The binding of P4 Ile, Tyr, and Trp, respectively, in S4. (i) The unbound S4 has glycerol in the BocD complex. Hydrogen bonds are represented by dashed lines. The boxed caspase-3 residues interact with inhibitors.

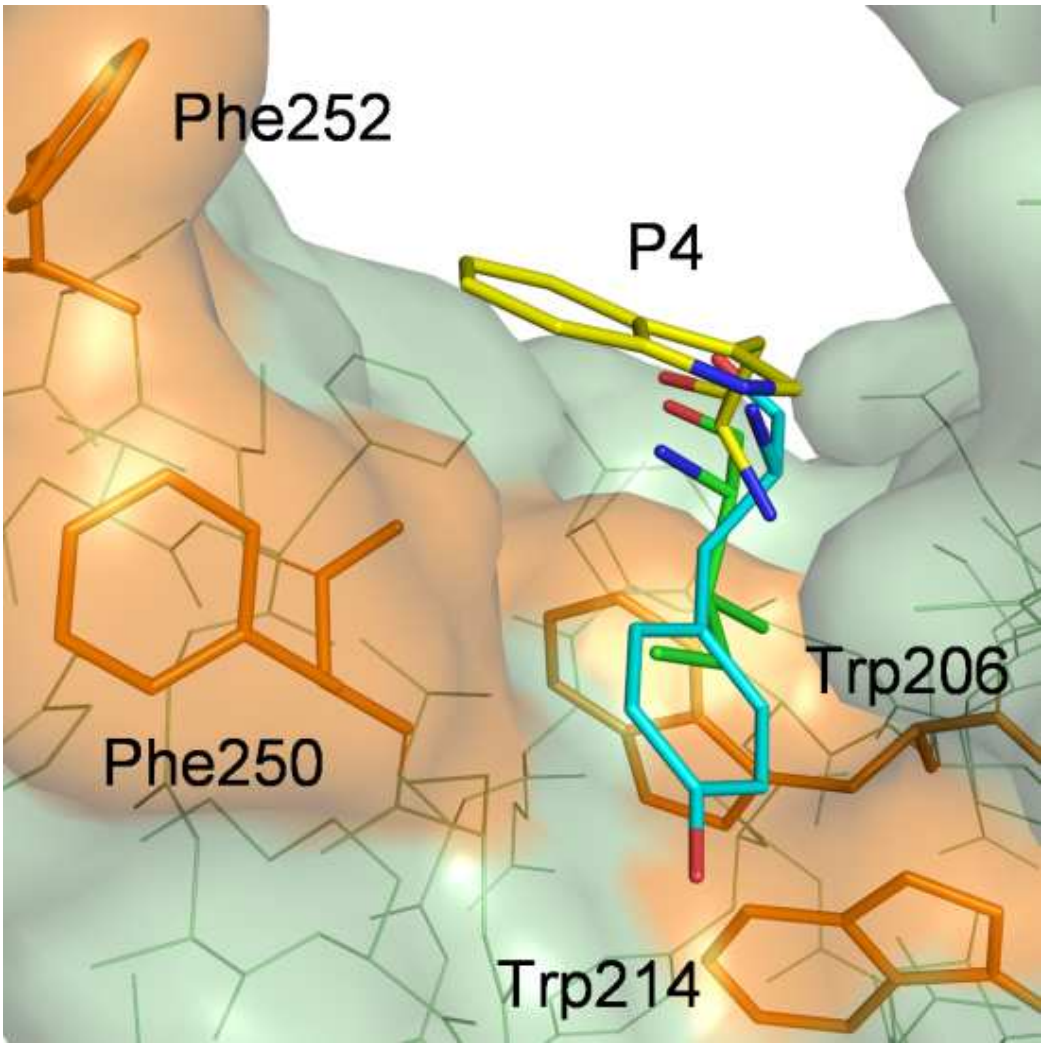


Figure 3.6 P4 binding site on caspase-3. P4 residues Ile (green), Tyr (cyan), and Trp (yellow) can bind in either the S5 pocket formed by Phe250 and 252 (orange) or another non-polar pocket formed by Trp206 and 214 (orange).

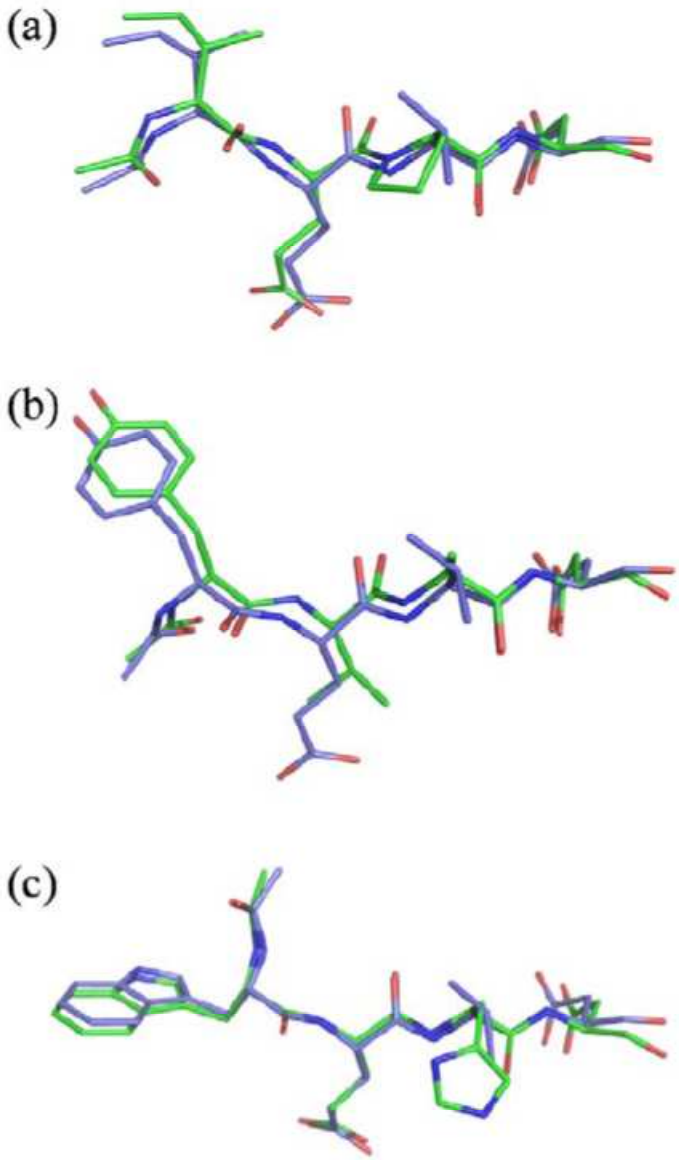


Figure 3.7 Superposition of inhibitors in the crystal structures (green) and structural models (purple). (a) IEPD/IEVD (b) YVAD/YEVD (c) WEHD/WEVD

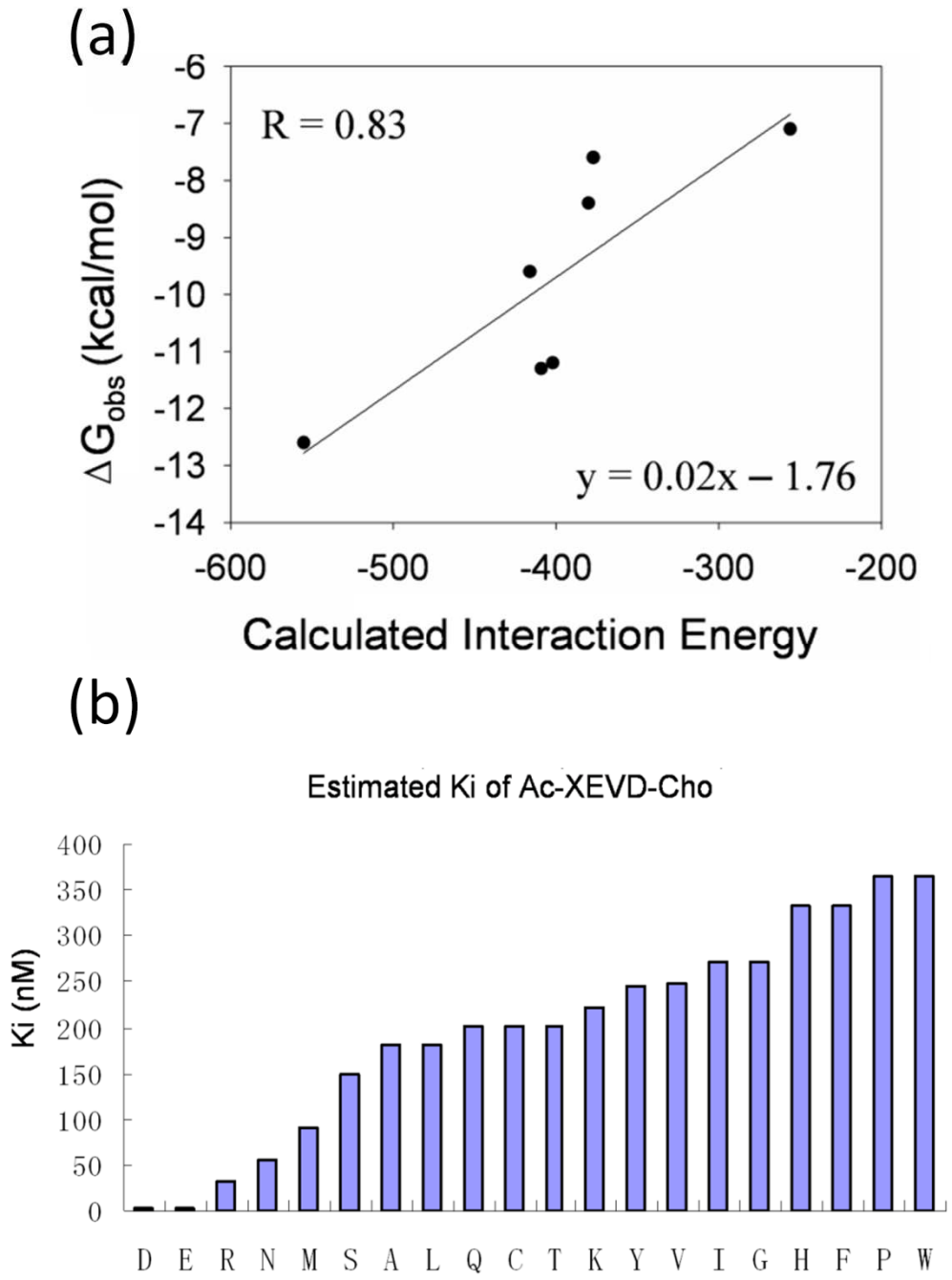


Figure 3.8 Predicted binding affinities for diverse P4 residues. (a) The correlation between experimental binding energies and calculated interaction energies. (b) The occurrence of different P4 residues found in caspase-3 natural substrates normalized to occurrence of P4 Asp. (c) Predicted Ki values of inhibitors Ac-XEVD-Cho.

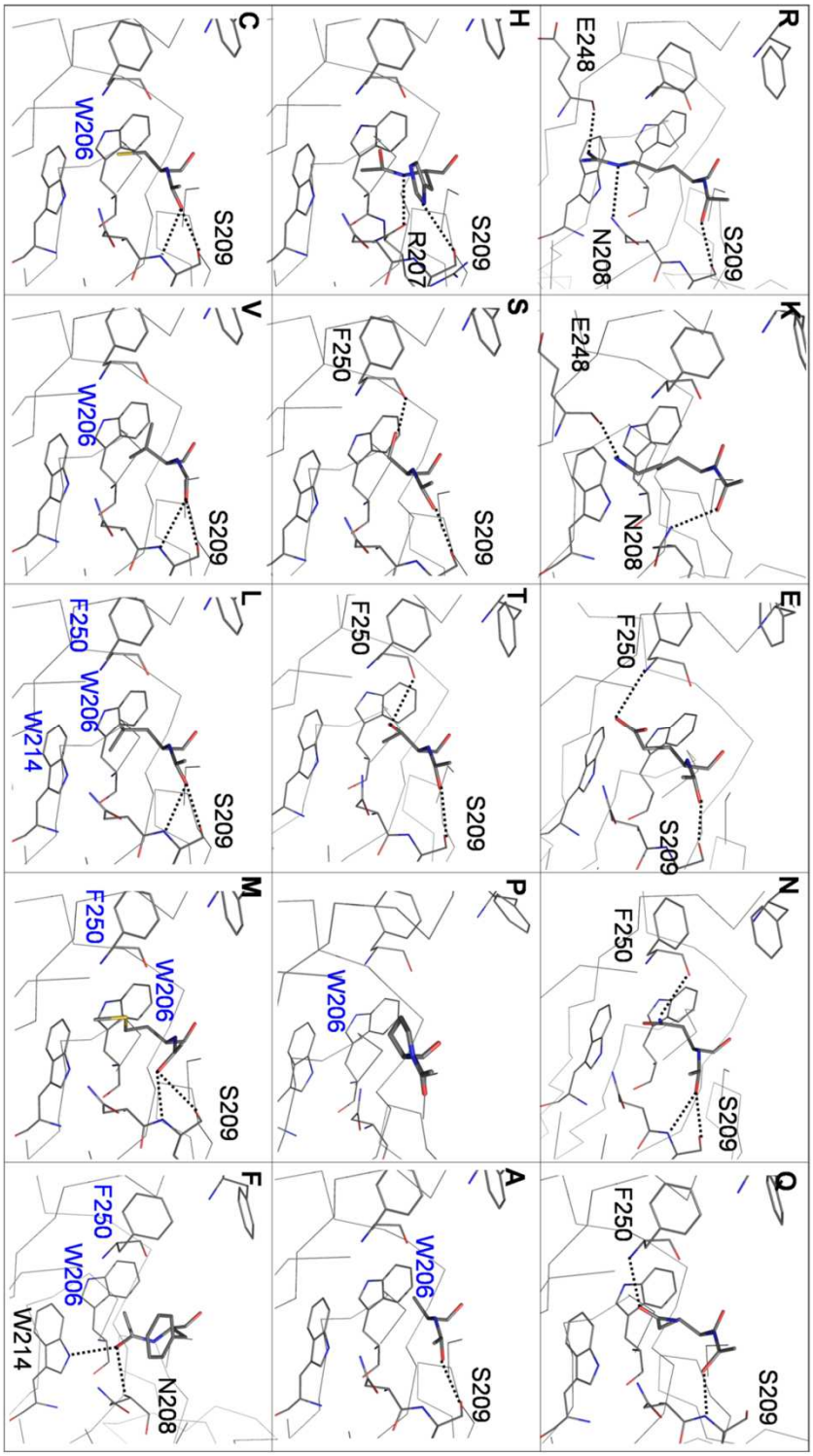


Figure 3.9 Predicted binding conformations of fifteen different P4 residues in caspase-3/XEVD. The P4 residue is located in the center in the stick representation, while caspase residues are shown as lines. The letter in the left top corner indicates the residue name. Hydrogen bond interactions are indicated by dash lines. Caspase residues involved in polar interactions are labeled in black while those involved in van der Waals interactions are labeled in blue.

4 CONFORMATIONAL CHANGES AND SUBSTRATE RECOGNITION IN PSEUDOMONAS AERUGINOSA D-ARGININE DEHYDROGENASE

(Published: Fu, G., Yuan, H., Li, C., Lu, C., Gadda, G. and Weber, I., (2010) *Biochemisty* 49: 8535-8545.)

4.1 Abstract

DADH catalyzes the flavin-dependent oxidative deamination of D-amino acids to the corresponding α-keto acids and ammonia. Here we report the first X-ray crystal structures of DADH at 1.06 Å resolution, and its complexes with iminoarginine ($DADH_{red}$/iminoarginine) and iminohistidine ($DADH_{red}$/iminohistidine) at 1.30 Å resolution. The DADH crystal structure comprises an unliganded conformation and a product-bound conformation, which is almost identical to the $DADH_{red}$/iminohistidine crystal structure. The active site of DADH was partially occupied with iminoarginine product (30% occupancy) that interacts with Tyr53 in the minor conformation of a surface loop. This flexible loop forms an "active site lid", similar to those seen in other enzymes, and may play an essential role in substrate recognition. The guanidinium side chain of iminoarginine forms a hydrogen bond interaction with the hydroxyl of Thr50 and an ionic interaction with Glu87. In the structure of DADH in complex with iminohistidine, two alternate conformations were observed for iminohistidine where the imidazole groups formed hydrogen bond interactions with the side chains of His48 and Thr50 and either Glu87 or Gln336. The different interactions and very distinct binding modes observed for iminoarginine and iminohistidine are consistent with the 1000-fold difference in k_{cat}/K_m values for D-histidine and D-arginine. Comparison of the kinetic data for the activity of DADH on different D-amino acids and the crystal structures in complex with iminohistidine and iminoarginine establishes that this enzymes characterized by a relatively broad substrate specificity, being able to oxidize positively charged and large hydrophobic D-amino acids bound within a flask-like cavity.

4.2 Introduction

All the standard amino acids except glycine can exist as either L- or D- optical isomers. L-amino acids are predominant in proteins synthesized by all living organisms. D-amino acids are important for bacteria as fundamental elements of the bacterial cell wall peptidoglycan layer [63]. Several D-amino acids, D-leucine, D-methionine, D-tyrosine and D-tryptophan, were recently reported to regulate disassembly of bacterial biofilms [64]. Also, specific D-amino acids have been discovered in many physiological processes in vertebrates, including humans [66]. Dunlop et al. identified D-aspartate in the brain and other tissues of mammals and suggested it may play a role in regulating the development of these organs [67]. D-serine was identified at significant levels in rat brain, at about a third of the level of L-serine [68]. Moreover, D-serine in the rat brain is distributed in close association with N-methyl-D-aspartate (NMDA) and it may act as an endogenous agonist of the NMDA receptor in mammalian brains [68].

Conversion of L- and D- amino acids in living organisms is commonly catalyzed by racemases. Various aminoacid racemases have been identified in bacteria, archaea, and eucaryotes including mammals. These racemases are classified into two groups: pyridoxal 5'-phosphate (PLP) -dependent and -independent enzymes [66]. In mammals, D-serine racemases and D-aspartate racemases are the most intensively studied enzymes due to their involvement in cell aging and neural signaling [69-70]. In bacteria, D-amino acids are deaminated by flavin-dependent dehydrogenases [73], allowing the bacteria to grow using D-amino acids as the sole carbon, nitrogen and energy source, in a concentration-sensitive manner, since some D-amino acid analogs have toxic effects on bacterial growth [71-72].

Pseudomonas aeruginosa, an opportunistic human pathogen, can grow with D-arginine as the sole source of carbon and nitrogen [175]. D-to-L arginine racemization by two coupled dehydrogenases serves as prerequisite of D-arginine utilization through L-arginine catabolic pathways [83, 176]. One enzyme is an FAD-dependent catabolic DADH and the other is an NAD(P)H-dependent anabolic LADH. DADH catalyzes the conversion of D-arginine into 2-ketoarginine and ammonia, and LADH converts 2-ketoarginine to L-arginine. In order to understand the reaction mechanism and substrate specificity of DADH, the crystal structure of DADH was determined at the atomic resolution of 1.06 Å, while the struc-

tures of DADH crystallized in the presence of D-arginine and D-histidine were both determined at 1.30 Å resolution. Well-defined electron density for the non-covalently bound FAD and imino intermediate of the reaction allowed detailed analysis of the enzyme active site. A loop region with alternative conformations was identified in the DADH structure and is involved in substrate binding. Very distinct binding modes were observed for iminoarginine and iminohistidine, in agreement with detailed kinetic analysis on substrate specificity reported previously [84] and in this paper. The structural characteristics described here provide insights into substrate recognition and the catalytic reaction mechanism of DADH.

4.3 Experimental Procedures

4.3.1 Materials

D-amino acids and L-amino acids were from Alfa Aesar and Sigma-Aldrich (St. Louis, MO). pms and PMSF were obtained from Sigma-Aldrich (St. Louis, MO). All of the other reagents were of the highest purity commercially available.

4.3.2 Expression and Purification of DADH and SeMet-DADH

Hexa-histidine-tagged DADH was expressed in *Escherichia coli* TOP10 and purified as described previously [83]. No FAD was added during expression and purification. The selenomethionine (SeMet) DADH protein was prepared following a protocol with slight modifications as described in a previous study [177]. 10-20 mM DTT was incorporated in the buffer throughout the purification steps in order to avoid oxidation of selenium. Mass spectrometry confirmed that about 88% of the eight methionines were replaced by Se-Met.

4.3.3 Crystallization and X-ray Data Collection

Purified DADH and SeMet-DADH were concentrated to 6 mg/mL and 3 mg/mL, respectively, in 50 mM Tris at pH 7.5. Crystals were grown by the hanging drop vapor diffusion method using a well solution of 0.1M 2-(N-morpholino)-ethanesulfonic acid (MES) pH 6.5-7.0, 5% glycerol, and 6-10% (w/v) PEG6000. Crystals can grow to a size about 0.1-0.2 mm^3 within two weeks. DADH was co-crystallized

with D-arginine or D-histidine under similar conditions using a 1:10 molar ratio of the enzyme (~0.15mM) to substrate. The crystals were soaked in the reservoir solution with 20% glycerol as cryoprotectant for ~1 min and frozen immediately in liquid nitrogen. X-ray data were collected at 100 K on beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory. Single-wavelength anomalous diffraction (SAD) data for SeMet-DADH were collected at the wavelength of 0.97182 Å (high energy remote) on beamline 22-BM of SER-CAT at Argonne National Laboratory.

4.3.4 Structure Determination and Model Refinement

The X-ray data were integrated and scaled using HKL2000 [178]. Program SGXPRO was used to perform phasing and automatic tracing with scaled but unmerged SeMet-DADH data [179]. With this approach, 331 of the 381 residues (375 amino acids residues of the enzyme and the N-terminal hexahistidine tag) (~87%) were successfully built. This model was then used for automated model building by ARP/WARP [180], in which 96% of the structure was fit. The structure of SeMet-DADH was then employed to solve the native DADH data set (1.06 Å) by molecular replacement using PHASER [127] in the CCP4i suite of programs [128]. Notably, all the FAD atoms were distinctly visible in the electron density maps. FAD was refined with 100% occupancy. The structure of DADH was used to solve the structures of DADH co-crystallized with D-arginine or D-histidine by molecular replacement. Crystal structures were refined with SHELX97 [129]. Manual adjustments and rebuilding were performed using the molecular graphics program COOT [131]. The structures of DADH co-crystallized with D-arginine and D-histidine showed difference density for iminoarginine and iminohistidine, respectively, bound at the enzyme active site, as observed for D-amino acid oxidase crystallized in the presence of D-tryptophan [75]. The D-amino acids are converted enzymatically to imino acids (R-C=NH), which then dissociate from the enzyme and are hydrolyzed to the final keto acids (R-C=O) in a non-enzymatic reaction. The iminoarginine was refined with 100% occupancy, while two alternate conformations were visible for the iminohistidine that refined to relative occupancies of 60% and 40%. Further analysis showed that the

structure of DADH, which was crystallized without added D-amino acids, contained extra density for a low occupancy iminoarginine adjacent to a disordered loop showing two alternate conformations. This structure was refined with two conformers corresponding to 70% occupancy of a ligand-free, open conformation DADH, and 30% occupancy of iminoarginine bound to a closed conformation of DADH. Higher peaks in the 2Fo-Fc electron density (>5σ) were observed for the two main-chain oxygen atoms of iminoarginine, while smaller density peaks (~1.3σ) were visible for other atoms. Therefore, in the ligand-free conformation, two molecules of water were refined at the locations of the carboxylate oxygen atoms of iminoarginine. Solvent molecules were inserted at stereochemically reasonable positions based on the peak height of the 2Fo–Fc and Fo–Fc electron density maps, hydrogen bond interactions and interatomic distances. Anisotropic B-factors were refined for all the structures. Hydrogen atom positions were included in the last stage of refinement using all data.

Sequence similarity searches were performed using BLAST [181]. Protein structures were superimposed on Cα atoms by using the secondary structure matching (SSM) module in COOT [182]. Figures of the structures were generated with PYMOL (http://www.pymol.org).

4.3.5 Enzyme Assay

The determination of the apparent steady state kinetic parameters was carried out by measuring initial rates of reaction with an oxygen electrode at varying concentrations of the D-amino acid substrate and a fixed concentration of 1 mM for PMS as electron acceptor. The reaction mixture (1 mL) was first equilibrated with organic substrate and PMS at the desired concentrations before the reaction was started by the addition of DADH. Enzyme assays were conducted in 20 mM Tris-Cl, pH 8.7, 25℃. All reaction rates are expressed per molar concentration of enzyme-bound flavin[1].

4.4 Results

4.4.1 Overall Structure of DADH

The DADH was crystallized without the addition of FAD or D-amino acid, and the structure was solved in the orthorhombic space group $P2_12_12_1$with one molecule per asymmetric unit using single anomalous dispersion (SAD) phasing and automated tracing. The structure was refined to an R-factor of 13.2% at 1.06 Å resolution. Structures were obtained also for DADH crystallized with substrates D-arginine or D-histidine under similar conditions. The structures were solved in the same space group $P2_12_12_1$ by using molecular replacement with the DADH structure as a template, and were both determined at the resolution of 1.30 Å and refined to R-factors of 13.4% and 12.8%, respectively. The crystallographic data and refinement statistics are summarized in **Table 4.1**. In the three structures, all 375 DADH residues and N-terminal hexa-histidine tag were defined clearly in the electron density map. Further analysis of the N-terminus of the structure revealed that the His-tag residues form several direct or water-mediated polar interactions with residues from symmetry related molecules, which stabilize the flexible terminus. Overall, the three structures are almost identical with pairwise RMSD values of 0.10-0.11 Å for 381 Cα atoms. The protein folds into two domains: an FAD-binding domain and a substrate-binding domain (**Figure 4.1**). The FAD-binding domain includes residues 1-82, 147-217 and 309-375. It consists of a central six-stranded β-sheet surrounded by five α-helices on one side and a three-stranded anti-parallel β-sheet with two α-helices on the other side. The substrate-binding domain is formed by residues 83-142 and 218-297, and consists of an eight-stranded β-sheet and two short antiparallel β-strands forming a sandwich surrounded by four α-helices. The substrate binding site was identified near the *re* face of the FAD 7,8-dimethyl-isoalloxazine ring at the interface of the two domains.

4.4.2 FAD-Binding Site

DADH contains a non–covalently bound FAD cofactor as identified in the X-ray crystal structures. All non-hydrogen atoms of the FAD were clearly visible in the electron density map (**Figure 4.2A**). The flavin adopts an elongated conformation with the adenine ring buried inside the FAD-binding domain

and the isoalloxazine ring located at the interface of the two domains. The adenine portion of the cofactor points toward the flavin binding domain and the 7, 8-dimethyl-isoalloxazine ring is directed towards the interface of the two domains, similar to the FADs observed in other flavin-dependent enzymes, such as L-proline dehydrogenase [50], D-amino acid oxidase [74] and L-amino acid oxidase [47, 183]. As illustrated in **Figure 4.2B**, the ribose ring moiety has hydrogen bond interactions with the side chain of Glu32, two water molecules and the main chain amide of Arg33. Numerous hydrogen bonds are present between the phosphate groups and the peptidyl atoms of Ala14, Thr42, Gly331 and 5 water molecules. The central part of the conserved glycine-rich sequence GXGXXG in the N-terminal region of the polypeptide is part of an α-helix (residues 13-25) that points toward the phosphate group and is assumed to contribute to the stabilization of the two negative charges of the diphosphate[55]. The 2'-OH and 4'-OH of the ribityl moiety of the cofactor establish hydrogen bond interactions with hydroxyl groups of Ser41 and Thr42. The 3'-OH group interacts with peptidyl atoms of Gly331 and Gly334 and one water molecule.

The 7,8-dimethyl-isoalloxazine is sharply bent between the two planes containing the benzene and pyrimidine moieties, defining a 15^{o} angle along the N5-N10 axis. This conformation is in agreement with several other crystallographic structures of flavin-dependent enzymes in which the flavin is in the reduced state [184]. The pyrimidine portion of the 7,8-dimethyl-isoalloxazine ring interacts through several hydrogen bonds with backbone atoms of the protein (i.e. C2=O atom with N atom of Ile335, Gln336, N3-H atom with the O atom of His48 and a water molecule, and C4=O atom with the N atom of His48). The benzene portion of the 7,8-dimethyl-isoalloxazine ring is excluded from solvent and interacts with residues Arg44, Ser45, Ala46, Arg222, Tyr249, Gly303 and Arg305.

4.4.3 Interactions of DADH with Iminoarginine

In order to investigate the structural basis for the substrate recognition of DADH, the enzyme was crystallized with D-arginine that was converted to iminoarginine. Iminoarginine binds with the plane formed by its Cα atom, carboxylate group and imino group approximately parallel to the *re* face of the flavin, while the side chain points away from the FAD (**Figure 4.3A**). The α-carbon of the iminoarginine

is 3.3 Å away from the FAD N5 atom, which is compatible with direct involvement of these atoms in the arginine oxidation catalyzed by DADH. The iminoarginine imino group is pointed out of the plane of the carboxylate atoms and can form a 2.9 Å long hydrogen bond interaction with the FAD O4 atom (**Figure 4.3B**). Several DADH residues form extensive polar interactions with the two main-chain carboxyl oxygen atoms of iminoarginine that anchor the ligand in the active site. One of the iminoarginine carboxyl oxygen forms hydrogen bonds with the side-chain hydroxyl of Tyr53, the guanidinium side chain of Arg305, and the carbonyl oxygen of Gly332. The other ligand carboxyl oxygen forms hydrogen bonds with the side-chain nitrogen of Arg222 and side-chain hydroxyl of Tyr249. Two water molecules are located near the ligand imino group, and form a hydrogen bond network extending to the imidazole side chain of His48. The side chain of iminoarginine forms hydrophobic interactions with the Val242 side chain, and a hydrogen bond interaction with the hydroxyl of Thr50. A strong ionic interaction (2.5 Å distance between pairs of O and N atoms) was observed between the guanidinium group of iminoarginine and the carboxylate group of Glu87. Hence, Glu87 may play an important role in the high specificity of DADH for D-arginine, in agreement with the kinetic analysis on substrate specificity of DADH reported previously [84] and performed in this study (*vide infra*).

4.4.4 Interactions of DADH with Iminohistidine

The structure of DADH was also determined for crystals grown in the presence of D-histidine, thereby yielding a $DADH_{red}$/iminohistidine complex. The iminohistidine binds to the DADH active site in two discrete conformations with clear density for the overlapping Cα atoms and carboxylate groups and weaker density for the alternate positions of the side chains (**Figure 4.3C**). The two conformations are like overlapping mirror images. In fact, there might be two additional conformations that are indistinguishable in this structure due to the potential 180º rotation of the imidazole ring. The imidazole orientation showing more hydrogen bond interactions with DADH is illustrated in **Figure 4.3D**. The iminohistidine atoms of conformation B (40% occupancy) all lie nearly parallel to the isoalloxazine ring of FAD. The main-chain atoms of conformation A (60% occupancy) lie on the same plane, whereas its

imidazole side chain is rotated by about 30º. DADH residues Tyr53, Arg222, Tyr249, Arg305, and Gly332 form conserved polar interactions with the ligand main-chain oxygen atoms in both conformations (**Figure 4.3D**). The imino group of the ligand forms a polar interaction with Gly332 in conformation A and with the hydroxyl of Tyr249 in conformation B. In both conformations the side chain of iminohistidine forms hydrogen bonds with the side chain of His48 and the hydroxyl of Thr50. The imidazole group also interacts with the side chain of Glu87 and the O4 of FAD in conformation A, and the side chain of Gln336 in conformation B. When the imidazole is rotated by 180º the interactions with His48, Thr50 and Gln336 are retained for conformation B. The interactions of conformation A with Glu87 and FAD are lost, however, suggesting they are not critical since they appear in only one of four possible conformations of iminohistidine.

4.4.5 Conformational Flexibility of the Active Site

No substrate or other ligand was added to the protein solution during crystallization of the DADH. However, the solved ligand-free structure contained weaker density at the substrate-binding site that was fit by 30% occupancy of the iminoarginine reaction product. It is possible that the ligand was trapped during the bacterial expression of the protein, as observed for other enzymes, for example, a tetrahedral reaction intermediate was discovered in the crystal structure of bacterial Est30 [177]. The iminoarginine lies adjacent to residues 50-56 of loop L1 with two alternate conformations that were refined with 0.7/0.3 relative occupancy (**Figure 4.4A**). This loop had a single well-defined conformation in the electron density for structures of DADH with iminoarginine and iminohistidine that corresponds to the higher occupancy conformation in the ligand-free structure. Therefore, the crystal structure of DADH contains both ligand-free and product-bound conformations, which suggests that a conformational change occurs in the adjacent loop regions upon binding of the substrate. Comparison of the two conformations of DADH revealed distinct changes in two regions of loop L1: the residues 50-56 adjacent to the imino product and the residues 45-47 near the flavin ring of FAD.

One of the major conformational changes between the ligand-free and the product-bound conformations in the DADH structure was observed at the substrate binding site for residues 50-56 of loop L1 region (**Figure 4.4B**). Four loops (L1, residues 33-56; L2, 244-248; L3, 261-276; L4, 329-336) form a flask-like substrate binding pocket. The binding pocket has a small entrance but expands at the bottom near FAD. The iminoarginine product binds to DADH in a similar manner as observed in the structures of pDAAO/iminotryptophan [75] and hDAAO/iminoserine [185]. In the ligand-free conformation, the side chain of Tyr53 points away from the active site and forms a hydrogen bond with Thr137, whereas in the structure of the product-bound enzyme, the aromatic ring of Tyr53 moves into the active site and forms hydrogen bonds with the carboxyl oxygen of the imino acid and also the side chains of Glu246 and Arg305. Other residues also form different interactions in the two conformations of this loop. In the ligand-free conformation, the side chain of Thr50 forms hydrogen bonds with Asp39 and Ala52, while Gly54 and Thr55 both form polar interactions with Arg59. In the product-bound conformation, Thr50 interacts with Glu87 and Gln336, and its main-chain peptidyl oxygen forms a hydrogen bond with the main-chain amide of Tyr53. The hydroxyl side chain of Thr55 forms a hydrogen bond with Asp312. The interactions of Tyr53 and Thr50, in particular, stabilize the ligand bound in the active site, suggesting that this flexible loop L1 region may act as a lid controlling substrate accessibility to the active site. The shape and flexibility of the active site enables DADH to control the entrance of substrates and accommodate a variety of substrates.

The second region showing significant differences in the ligand-free and product-bound conformations comprised residues 45-47 at the N-terminal end of loop L1, which is located at the *si* face of the flavin ring (**Figure 4.4C**). This region possesses two conformations in the DADH structure and shows a single well-defined conformation in the structures of DADH with high occupancy products. In the ligand-free conformation, the Ser45 hydroxyl group forms a hydrogen bond with the flavin N5 atom, and the side chain of Ala46 is pointed away from FAD. In the product-bound conformation, this region is flipped about 90º with the side chain of Ser45 pointed away from FAD and the Ala46 side chain approaching closer to the FAD. This structural change causes the loss of the hydrogen bond interaction between Ser45

and FAD and, instead, a new hydrogen bond is formed between the main-chain amide of Ala46 and the N5 of FAD. A similar interaction was reported between the Ala49 of pDAAO and its FAD N5 atom [74-75]. Due to the conformational changes of residues 50-56 and 45-47 in loop L1 region, the active site of DADH shrinks by about 3.8 Å when the product is bound. The distances between the Cα atoms of Ala46 and Tyr53 are 17.5 Å and 13.7 Å in the ligand-free conformation and product-bound conformation, respectively (**Figure 4.4B**). Loop L2 also has a slight conformational change and moves about 0.7 Å closer to loop L1 based on the distances between the Cα atoms of loop L1 Tyr53 and loop L2 Asp245.

4.4.6 Substrate Specificity

The steady state kinetic parameters were determined by measuring initial rates of reaction with various D-amino acids in 20 mM Tris-Cl, pH 8.7 at 25 °C. The concentration of the electron acceptor PMS was kept fixed at 1 mM PMS, ensuring saturation of the enzyme[2]. As illustrated in **Table 4.2** the enzyme showed substrate preferences for amino acids with positively charged side chains, of which D-arginine appears to be the best substrate. D-lysine, D-tyrosine and D-methionine were also good substrates. In contrast, D-glutamate, D-aspartate, L-arginine and glycine were not detectably oxidized by the enzyme. Cysteine could not be assayed due to its non-enzymatic reaction with PMS in the enzyme reaction mixture.

4.4.7 DADH Recognition of Iminoarginine and Iminohistidine

Comparison of the structures of DADH in complex with iminoarginine and iminohistidine reveals the molecular basis for the relatively broad substrate specificity of this enzyme. The internal cavity of the substrate-binding site has a triangular cross-section with a narrow entrance at the top, as shown in **Figure 4.5** (DADH residues Tyr53, Gly332 and Gln336 were removed to view the internal site). The bottom of the cavity extends about 14.9 Å from loop L1 to loop L3 and 14.1 Å from loop L2 to loop L4. The depth of this substrate-binding pocket is about 10 Å from the Cα atom of Gly54 in loop L1 to the FAD O2 atom. The main chain of iminohistidine is located almost on the same plane and parallel to the isoalloxazine ring of FAD. The main-chain atoms of iminoarginine are very close to those of iminohistidine, and form

similar polar interactions with the DADH residues Tyr53, Arg222, Tyr249, Arg305, and Gly332. Since the binding pocket interactions with the main chain of the substrate are conserved, DADH has a broad substrate specificity for a variety of D-amino acids, as shown in **Table 4.2**, which is also supported by the kinetic studies reported previously [84]. The different side chain interactions with DADH are presumed to be responsible for the specificity differences. The longer D-arginine side chain fits well in the large substrate-binding pocket of DADH, while the smaller D-histidine cannot fill the cavity, which is consistent with the differences in catalytic activity for the two substrates.

4.4.8 Structural Comparison with Related Proteins

Determination of the crystal structure of DADH allows a detailed comparison with other mechanistically related flavin-dependent oxidoreductases. Analysis of the sequences and structures showed that DADH shares low sequence identity (<20%) but similar three-dimensional structures with L-Proline dehydrogenase β subunit (18.5% sequence identity, RMSD 2.2 Å for 342 Cα atoms, [50]), heterotetrameric sarcosine oxidase (18.4% sequence identity, RMSD 2.2 Å for 339 Cα atoms, [186]), porcine D-amino acid oxidase pDAAO (17.2% sequence identity, RMSD 2.4 Å for 270 Cα atoms, [74]) and L-amino acid oxidase LAAO (16.4% sequence identity, RMSD 3.0 Å for 245 Cα atoms, [47]). For a detailed comparison of the overall structure and active site geometry, we have chosen the structures of the enzyme complexes with dicarboxylate ligands: pDAAO from pig kidney in complex with iminotryptophan (PDB entry 1DDO, [75]) and LAAO from *Rhodococcus opacus* in complex with L-alanine (PDB entry 2JB1, [47]).

pDAAO and LAAO both are homodimeric enzymes formed by two subunits interacting through their helical domain or substrate binding domain, respectively [47, 74, 183]. Structural superimposition indicates that the overall structure of DADH resembles the monomer structures of pDAAO and LAAO (**Figure 4.6**). DADH and LAAO share the same topology for the FAD binding domain in which a central β-sheet is surrounded by α-helixes on one side and a three-stranded anti-parallel β-sheet on the other side. However, in the FAD-binding domain of pDAAO, an α-helix is observed instead of the three-stranded anti-parallel β-sheet located in DADH (**Figure 4.6A**). All three enzymes contain a large β-pleated sheet in

the substrate binding domain, while the orientation of this β-sheet in LAAO is significantly different from those in DADH and pDAAO. Two extra β-strands are observed for LAAO at the Gly57-Cys70 region (**Figure 4.6B**). Furthermore, the helical domain involved in dimerization of LAAO is absent from the structures of DADH and pDAAO.

The substrate binding sites of these three enzymes are all located near the *re* face of the FAD isoalloxazine ring at the interface of the FAD-binding and substrate-binding domains. Despite the opposite orientation of L-alanine bound to LAAO compared to the binding mode of ligands in the DADH and pDAAO complexes, similar protein-ligand interactions are retained among the three enzymes. The specific arrangement of active site residues is suggested to be responsible for the strict enantioselectivity of each enzyme [47, 74-75, 183]. In the pDAAO/iminotryptophan complex, the carboxylate group of the ligand forms a salt bridge with the guanidinium side chain of Arg283 (**Figure 4.7A**). Similar interactions are formed between L-alanine and Arg84 in the structure of LAAO/L-alanine (**Figure 4.7B**). However, in the $DADH_{red}$/iminoarginine complex, this ionic interaction is replaced by interactions between the ligand and two arginines: Arg222 and Arg305. The carboxylate oxygen atoms of the ligands also form polar interactions with Tyr53, Tyr249 and Gly332 of DADH and Tyr224, Tyr228 and Gly313 of pDAAO, respectively. L-alanine forms similar interactions with the polar residues Gln228 and Tyr371 of LAAO. The side chains of these ligands form distinct interactions with enzyme residues. In DADH iminoarginine forms strong ionic interactions with Glu87 and a hydrogen bond with Thr50. In contrast, in pDAAO the indole side chain of iminotryptophan is surrounded by several hydrophobic residues (Ala49, Leu51, Ile215, Ile230, and Tyr228). The short side chain of L-alanine interacts with several hydrophobic residues Trp426, Ala466 and Trp467 of LAAO. These similarities and differences in interactions are presumed to be important determinants of their different substrate specificities.

4.5 Discussion

4.5.1 An "Active Site Lid" Controls Substrate Accessibility

A loop covering the active site in the DADH structures shows conformational change with substrate binding. A similar loop region described as "active site lid" was reported in the structure of pDAAO (residues 216-228, [74]). It is hypothesized that this lid switches between closed and open conformations to allow substrate binding and product release [74]. Importantly, this atomic resolution DADH crystal structure demonstrates two loop conformations corresponding to the ligand-bound and ligand-free forms. Once the substrate binds to the enzyme, it is sandwiched between the FAD isoalloxazine and the active site lid. After the enzymatic reaction is completed, this loop region is expected to switch back to the open conformation allowing the product to exit the cavity. A similar "loop-and-lid" structure has been assigned in some of the glucose-methanol-choline (GMC) family members, including glucose oxidase (residues 76-97, [187]), cholesterol oxidase (residues 95-109, [188]), pyranose 2-oxidase (residues 452-456, (*34*)), and the flavoprotein domain of cellobiose dehydrogenase (residues 289-299 [189]). Besides its role as a gate in opening and closing the enzyme active site, the active site lid may be important in determining the substrate specificity of DADH. The conformational change of the lid, especially for Tyr53, may allow DADH to accommodate bulky residues like D-phenylalanine or D-tryptophan. Similar interactions have been observed in pDAAO (Tyr224, [75]) and flavocytochrome b_2 (Tyr254, [82]). Furthermore, closing of the lid shields the active site and the FAD from solvent. An increase of the active site hydrophobicity caused by loop closure may facilitate the hydride transfer step leading to substrate oxidation in pDAAO [75]. Similar phenomena have been described in several NAD(P)H-dependent dehydrogenases [190]. Further insight into the functionality of this active site lid may be obtained by protein engineering studies on DADH.

4.5.2 Substrate Specificity

DADH is characterized by a broad substrate specificity, being able to oxidize basic and hydrophobic D-amino acids of various size [84]. This conclusion is supported by comparison of the structures

of DADH in complex with iminoarginine and iminohistidine, and kinetic analysis. Iminoarginine forms a strong ionic interaction with the side chain of Glu87 and a hydrogen bond with Thr50, whereas the side chain of iminohistidine extends in a different direction and forms hydrogen bonds with His48, Thr50 and Gln336. Steady state kinetics data summarized in **Table 4.2** are also consistent with this conclusion. Indeed, the large k_{cat}/K_{Arg} value in the range of 10^6 $M^{-1}s^{-1}$ is 1,000-fold higher than the k_{cat}/K_{His} value. Kinetic data for other D-amino acids further indicate that the negatively charged side chain of Glu87 is the major determinant for the specificity of DADH for positively charged substrates. Strong electrostatic interactions formed between the ligand and Glu87 allow D-arginine and D-lysine to be the two most effective substrates for DADH. A notable example of similar case can be found from the well defined substrate specificity of trypsin, in which the enzyme residue Asp189 is responsible for its narrow selection for positively charged arginine and lysine [191-192]. In addition, the hydrophobic walls (Tyr53, Met240, Val 242 and Tyr249) of the DADH active site pocket create a favorable environment for the long aliphatic and unbranched parts of the basic D-arginine and D-lysine. Furthermore, D-tyrosine, D-methionine and D-phenylalanine may form favorable van der Waals contacts with the hydrophobic walls of the active site and are also good substrates for DADH. In contrast, DADH shows low or undetectable activities toward other D-amino acids, especially the negatively charged D-glutamate and D-aspartate.

4.5.3 The Ser/Ala Switch in the FAD Binding Site

The Ser45-Ala47 region also has two conformations corresponding to the ligand-free and product-bound conformations in the DADH structure, while this region has a well-defined single conformation in the structures of $DADH_{red}$/iminoarginine and $DADH_{red}$/iminohistidine. Comparison of the two conformations indicates that, upon substrate binding, the side chain of Ala46 rotates from the ligand-free conformation to approach closer to the isoalloxazine of FAD in the product-bound conformation, while maintaining a hydrogen bond with the N5 of the flavin. A hydrogen bond (O_{Ser45}—$N5_{FAD}$) in the ligand-free conformation is replaced by another polar interaction (N_{Ala46}—$N5_{FAD}$) in the product-bound conformation (**Figure 4.4C**). Furthermore, interactions of the FAD N5 atom with a residue structurally equivalent to

Ala46 of DADH are conserved among different FAD-dependent enzymes, such as pDAAO (Ala49, [74]), yeast D-amino acids oxidase (Gly52, [193]), L-proline dehydrogenase (Gly48, [50]), flavocytochrome b2 (Ala198, [82]), and L-galactono-γ-lactone dehydrogenase (Ala113, [44]). Human DAAO shares the same sequence ($V_{47}AAGL_{51}$) with pDAAO near the *si* face of FAD ring. However, this region has a conformational shift in hDAAO, resulting in the loss of a hydrogen bond between Ala49 and the FAD N5 atom. This conformational change in hDAAO has been suggested to be responsible for its low binding affinity for FAD as well as the slower rate of flavin reduction [185]. Furthermore, the structurally equivalent Ala113 in L-galactono-γ-lactone dehydrogenase has been reported to play a substantial role in the enzyme activity with oxygen. Site-directed replacement of Ala113 with glycine increased the reactivity of the reduced flavin with oxygen by about 400-fold [44]. Therefore, the functional significance of this Ser/Ala switch in DADH remains unclear; it might be involved in structural stability and also the enzymatic properties of this protein. Further studies are required to elucidate its function.

4.5.4 Two Conformations of D-Histidine in the Complex with DADH

The iminohistidine binds to the DADH active site in two discrete conformations, while iminoarginine shows a single conformation. The main chain atoms of the two iminoacids are bound in similar positions as shown in **Figures 4.3** and **4.5**, however, the side chains extend in different directions. The iminoarginine side chain lies almost perpendicular to the isoalloxazine ring of FAD, whereas the side chain of iminohistidine is more parallel to the flavin. The length and charge of the ligand side chain may both contribute to the different binding modes of the two imino acids. The long iminoarginine side chain extends to form strong electrostatic interactions with Glu87 and a hydrogen bond with Thr50. Instead, the flexibility of the shorter iminohistidine results in conserved interactions with His48, Thr50, while alternate conformations interact with Glu87 or Gln336. The substrate carboxylate group is essential for binding of D-amino acids to the DADH active site. Therefore, these atoms maintain the same positions between different conformations of iminohistidine, while the side chains have more freedom to move. Simi-

larly, the ligand side chains display two discrete conformations in the structures of monomeric sarcosine oxidase in complex with methylselenoacetate and methyltelluroacetate [194].

4.5.5 Structural Comparison with Related Proteins

Structural comparison of DADH with pDAAO and LAAO revealed that the active site residues of DADH are more similar to those of pDAAO (**Figure 4.7**), in agreement with their specificity for D-amino acids rather than L-amino acids [84]. The spatial arrangement of active residues is critically related to the enzyme enantiomeric selectivity. A mirror-symmetrical relationship of active sites has been described among enzymes with opposite stereospecificities and two different modes have been reported. One is observed between pDAAO and flavocytochrome b_2, in which the active sites of the two enzymes are mirrored through the plane of isoalloxazine (on the *re* side in pDAAO, on the *si* side of flavocytochrome b_2) [74]. Another situation is that the substrate binding sites are mirrored through the plane perpendicular to the isoalloxazine ring, which is observed in the comparison between pDAAO and LAAO [183]. The active site of DADH is highly similar to that of pDAAO, and mirror-symmetrically related to that of LAAO (**Figure 4.7**). This observation is in agreement with the kinetic study described previously [84] and in this paper showing that DADH can oxidize a variety of D-amino acids but not L-arginine. A similar active site arrangement has been observed in other flavoenzymes like glycolate oxidase [195], L-glutamate oxidase [196] and L-aspartate oxidase [197], suggesting these enzymes employ a common mechanism to control enantioselectivity.

In addition to their specificity toward different enantiomers of amino acids, DADH, pDAAO and LAAO display other distinct features. For example, the ligands are held in the active site by a salt bridge formed with an arginine in pDAAO and LAAO, while two active site arginines are involved in DADH. The active sites of DADH and pDAAO are covered by a loop termed the “active site lid”. This feature was not reported in the structure of LAAO, and a funnel-shaped entrance for substrate binding was proposed instead [198]. In addition to the polar interactions between the ligands and the enzyme active site residues, a hydrophobic environment has been observed in all three enzymes. In fact, all the enzymes in-

teract with a wide range of hydrophobic amino acids [75, 84, 199]. Overall, the different composition and arrangement of active site residues determine differences in substrate specificity, while critical interactions for catalysis are conserved among these mechanistically related enzymes.

4.6 Conclusions

Comparison of the crystal structures of DADH and its complexes with iminoarginine and iminohistidine has highlighted important structural differences that rationalize the catalytic activities and substrate specificity of the enzyme. The imino products of D-arginine and D-histidine bind to the active site in very distinct side-chain conformations. Glu87 forms strong electrostatic interactions with iminoarginine, which is likely responsible for the high selectivity of DADH for positively charged residues like D-arginine and D-lysine. A loop region has been designated as an active site lid controlling the substrate accessibility to the active site, similar to those reported in other flavin-dependent enzymes. Structural comparison of DADH with other related flavin-dependent enzymes reveals that the spatial arrangement of active site residues is essential for the differences in enzyme enantioselectivity, while some specific interactions needed for catalysis are conserved among the enzymes. Overall, the high-resolution structures for DADH described in this study will provide new guidelines for future studies of similar flavin-dependent enzymes.

4.7 Acknowledgements

We thank Dr. Siming Wang and Yanyi Chen for the mass spectrometry data, Dr. Johnson Agniswamy and Dr. Yuan-Fang Wang for providing help with refinement and valuable discussions. We are especially grateful for the assistance of Dr. Zheng-Qing Fu and the staff at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory, for assistance during X-ray data collection. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Table 4.1 Crystallographic Data Collection and Refinement Statistics

	SeMet DADH	DADH	DADH/ D-Arg	DADH/ D-His
Data Collection Statistics				
Redox state of flavin		Mixture	Reduced	Reduced
Wavelength (Å)	0.97182 [a]	0.99999	1.00000	1.00000
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	62.65	62.18	62.17	62.10
b (Å)	78.21	78.08	78.43	78.15
c (Å)	89.47	89.72	89.95	89.59
Resolution range (Å)	50-1.50	50-1.06	50-1.30	50-1.30
Total observations	1016,247	1258,212	600,949	627,618
Unique reflections	70270	190,545	98,021	97,110
Completeness	99.7 (98.0)[b]	96.2 (69.1)	90.6(53.9)	90.4(54.7)
$<I/\sigma(I)>$	27.6(16.6)	26.5(2.3)	18.5(2.7)	31.8(3.2)
R_{sym} (%)	7.6(18.1)	6.2(41.4)	8.2(37.3)	4.9(30.6)
Figure of merit	0.72			
f'	-4.2234			
f"	3.7853			
Refinement Statistics				
Resolution range (Å)		10-1.06	10-1.30	10-1.30
R_{work} (%)		13.2	13.4	12.8
R_{free} (%)		15.9	16.6	16.3
Mean B-factor ($Å^2$)		16.0	15.9	17.5
Protein		14.4	14.6	16.1
FAD		7.3	6.5	8.3
Ligand		29.5	37.0	16.3
Water		27.9	29.7	31.8
Number of atoms				
Protein		3102	2961	2971
FAD		53	53	53
Ligand		12	12	22
Water		467	358	367
r.m.s. deviations				
Bond length (Å)		0.017	0.013	0.012
Angle[c]		0.035	0.031	0.030

[a]The SeMet DADH data were collected at the wavelength of high energy remote.
[b]Values in parentheses are given for the highest resolution shell.
[c]The angle RMSD in SHELX97 is indicated by distance in Å.

Table 4.2 Steady State Kinetics of DADH[a]

substrates	k_{cat}/K_m, $M^{-1}s^{-1}$	k_{cat}, s^{-1}	K_m, mM
D-arginine	$(3.4 \pm 0.3) \times 10^6$	204 ± 3	0.06 ± 0.01
D-lysine	$(5.3 \pm 0.2) \times 10^5$	141 ± 3	0.26 ± 0.01
D-tyrosine	27,600±3,800	23±1	0.8 ± 0.1
D-methionine	14,800±600	154±3	10 ±1
D-phenylalanine	6,900±300	75±3	11±1
D-histidine	3,140 ± 30	35 ± 1	11 ± 1
D-leucine	515±60	6.4± 0.3	12± 1
D-proline	420 ± 10	-[b]	-
D-tryptophan	245 ± 3	-	-
D-isoleucine	195 ± 3	-	-
D-valine	47 ± 1	-	-
D-alanine	41 ± 1	-	-
D-glutamine	186 ± 3	-	-
D-asparagine	16 ± 1	-	-
D-serine	3.8 ± 0.1	-	-
D-threonine	0.75 ± 0.01	-	-
D-glutamate	-	-	-
D-asparate	-	-	-
L-arginine	-	-	-
Glycine	-	-	-
D-cysteine	nd[c]	nd	nd

[a]Enzyme activity was measured at varying concentrations of substrateand 1 mM PMS in 20 mM Tris-Cl, pH8.7, at 25 °C. [b]Cannot be saturated with the substrate, thereby k_{cat} and K_m values are not reported.[c] Not determined. PMS was reduced by cysteine.

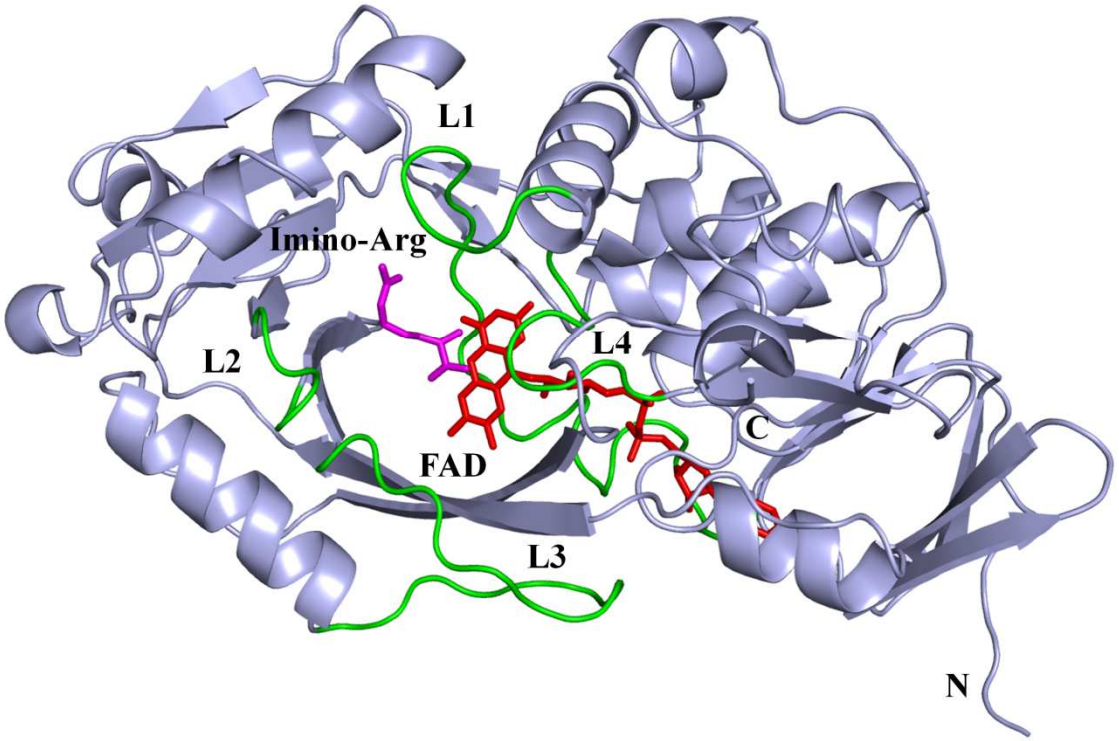


Figure 4.1 Overall structure of DADH with iminoarginine. The DADH structure is shown in cartoon representation. Iminoarginine and cofactor FAD are shown as sticks and colored magenta and red, respectively. Loops that contribute to the active site are colored green.

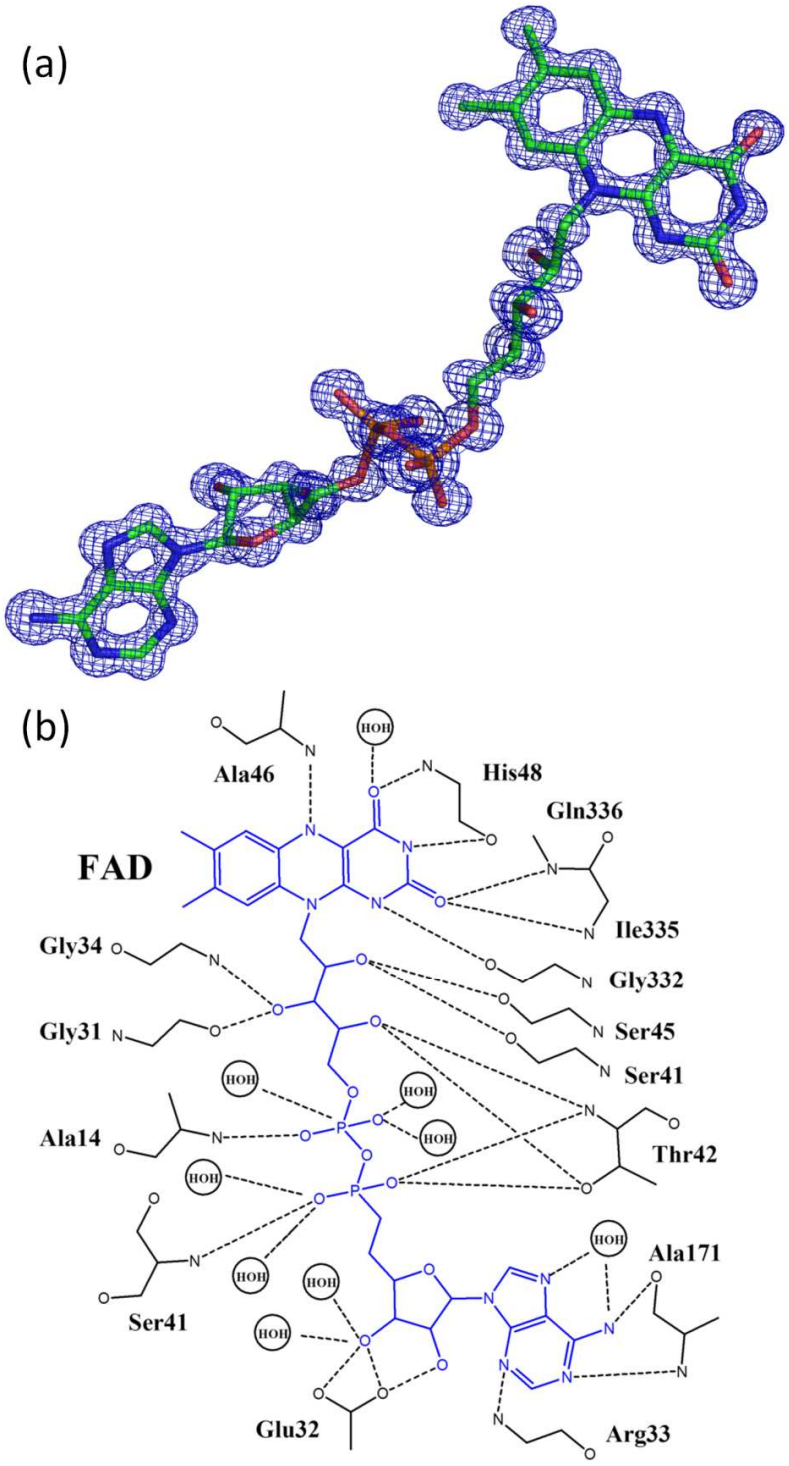


Figure 4.2 Cofactor FAD in DADH. (A) The *2Fo-Fc* electron density map of FAD contoured at 2σ illustrates the high quality of the 1.06 Å resolution structure. FAD (colored by element type) adopts an elongated conformation. (B) Schematic diagram of the flavin-apoprotein interactions in the product-bound DADH conformation. Hydrogen bonds are indicated by dashed lines. The flavin cofactor is colored blue. For clarity, H atoms are not shown.

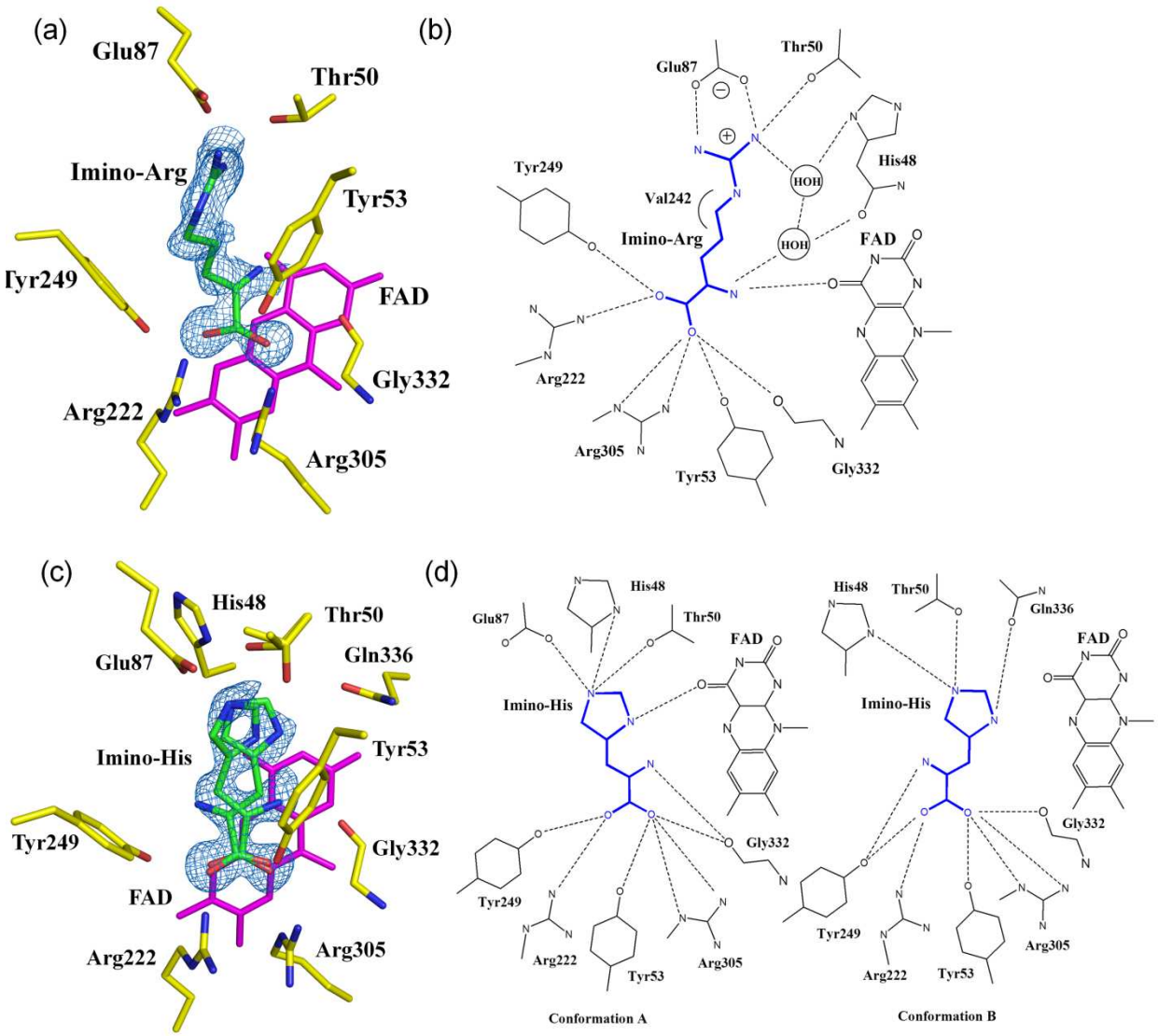
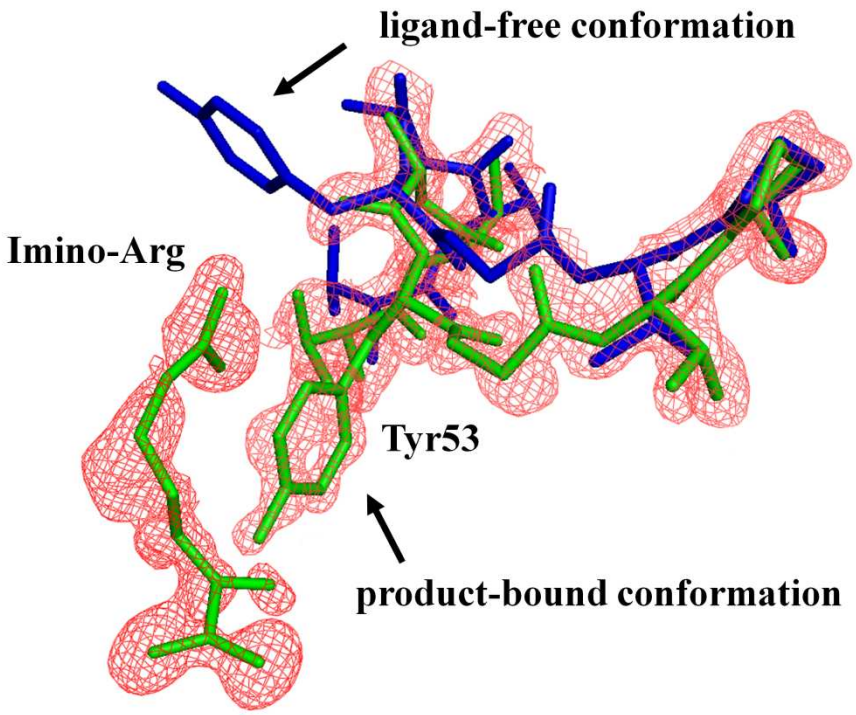
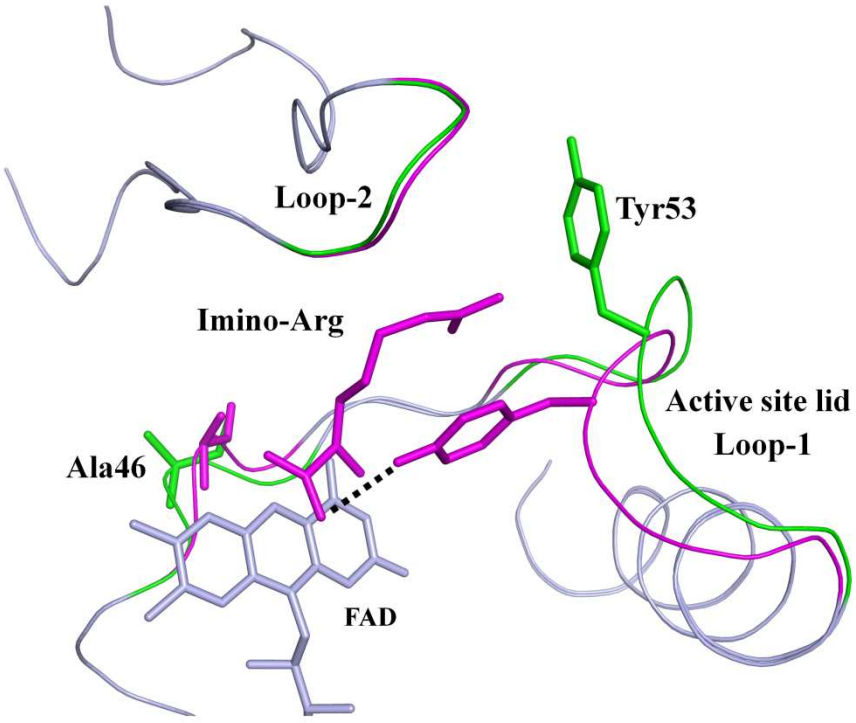


Figure 4.3 DADH interactions with iminoarginine (A, B) and iminohistidine (C, D). Carbons are colored yellow for DADH active site residues and green for the imino acids. FAD is represented by its isoalloxazine ring in magenta. Omit map of ligand is indicated as blue mesh and contoured at 3σ. Dashed lines represent hydrogen bonds and ionic interactions.

(a)


ligand-free conformation
Imino-Arg
Tyr53
product-bound conformation

(b)


Loop-2
Tyr53
Imino-Arg
Active site lid
Loop-1
Ala46
FAD

(c)

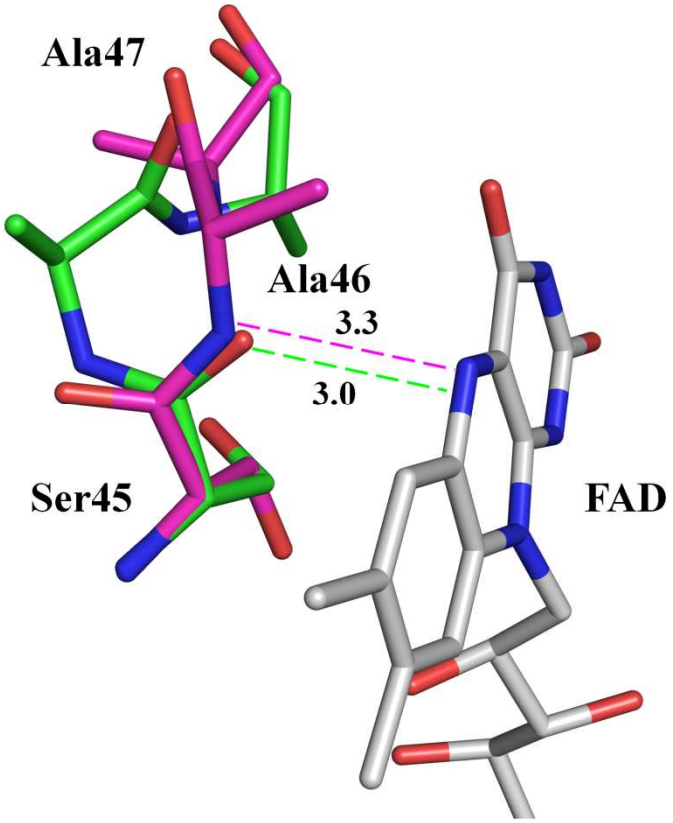


Figure 4.4 Conformational flexibility of DADH structure. (A) The DADH structure comprises two conformers: a 30% occupied population with iminoarginine product, and 70% one without ligand. The iminoarginine and its adjacent loop region with low occupancy (product-bound conformation) are shown as green sticks. The omit map of this region (red mesh) is contoured at 2.5σ. The corresponding loop region with high occupancy (ligand-free conformation) is shown as blue sticks. (B) Comparison of ligand-free conformation (green) and product-bound conformation (magenta) at loop L1 and L2 regions in DADH structure. Ala46, Tyr53 and iminoarginine are represented as sticks. A hydrogen bond (black dotted line) is formed between Tyr53 and iminoarginine in the product-bound conformation. (C) Conformational change of residues Ser45-Ala46-Ala47 at the *si* face of flavin ring. Ser45 in the ligand-free conformation (green carbon atoms) forms a 3.0 Å hydrogen bond with the FAD N5 atom (gray carbon atoms). Ala46 forms a 3.3 Å long hydrogen bond with the N5 atom of FAD in the product-bound conformation (magenta carbon atoms).

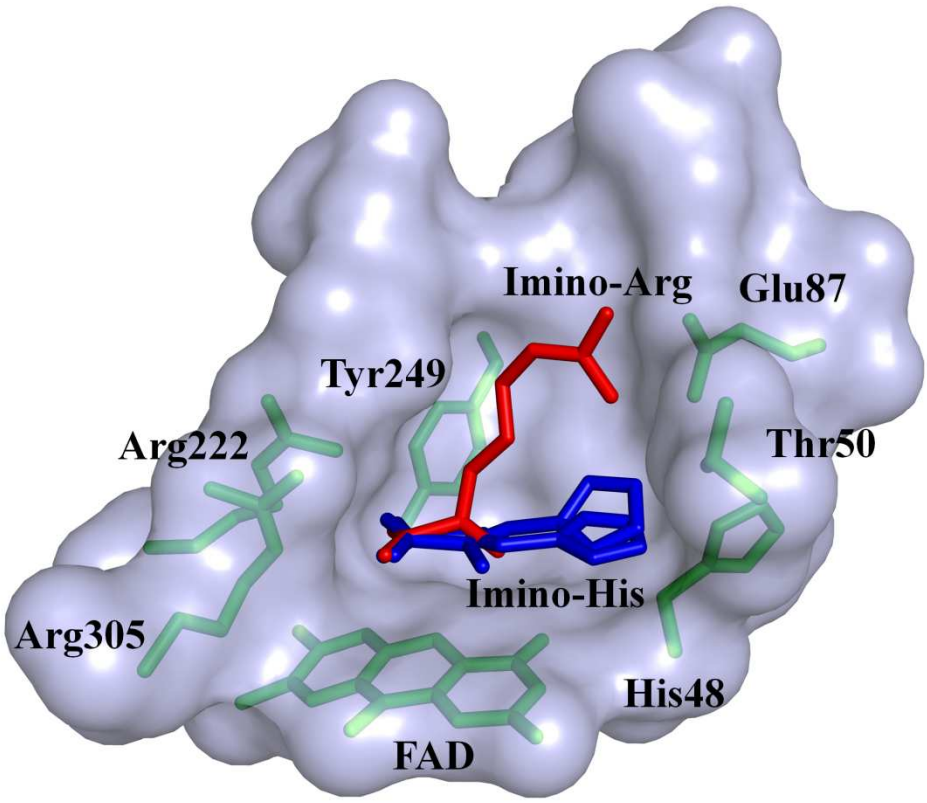


Figure 4.5 Comparison of DADH structures in complex with iminoarginine and iminohistidine. DADH active site residues (green sticks) line the interior of the substrate binding pocket (Thr53, Gly332 and Gln336 were omitted for clarity). Iminoarginine (red) and iminohistidine (blue) bind to the active site in very distinct conformations. FAD is represented by its isoalloxazine ring in green.

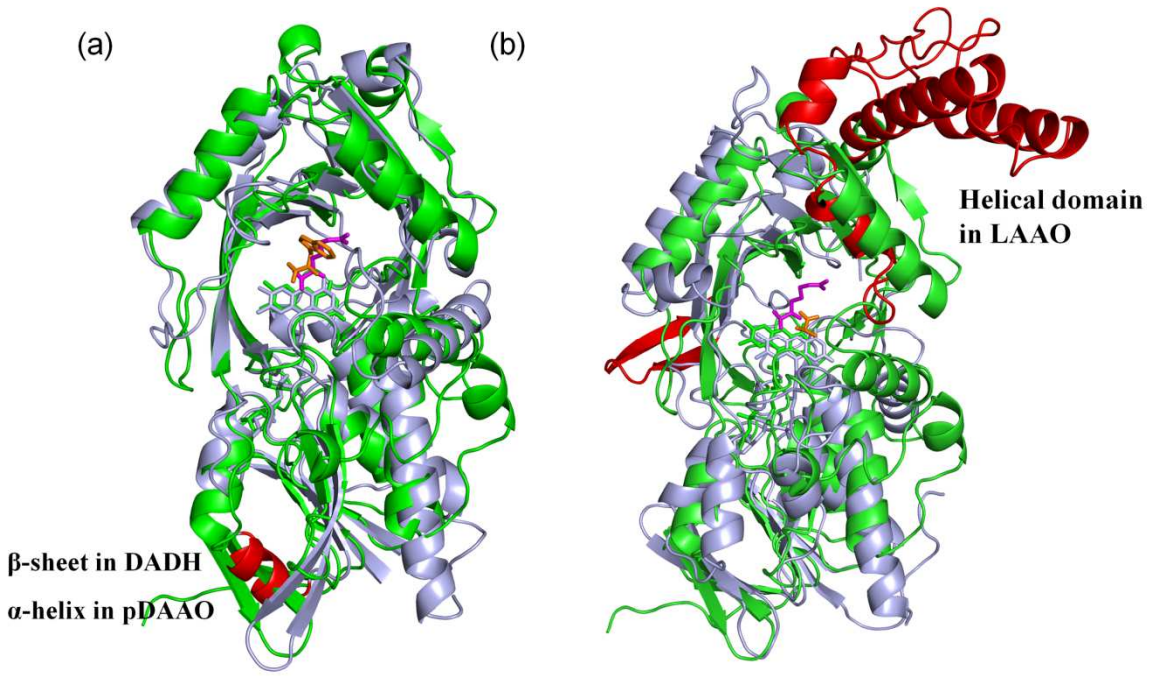


Figure 4.6 Structural comparison of DADH with pDAAO (A) and LAAO (B). DADH is shown in green cartoon representation while pDAAO and LAAO are colored in light blue with red regions indicating their major structural differences compared to DADH. FAD is represented by sticks in green and light blue corresponding to the compared structures. The ligands are: iminoarginine in DADH (magenta), iminotryptophan in pDAAO (orange) and L-alanine in LAAO (orange).

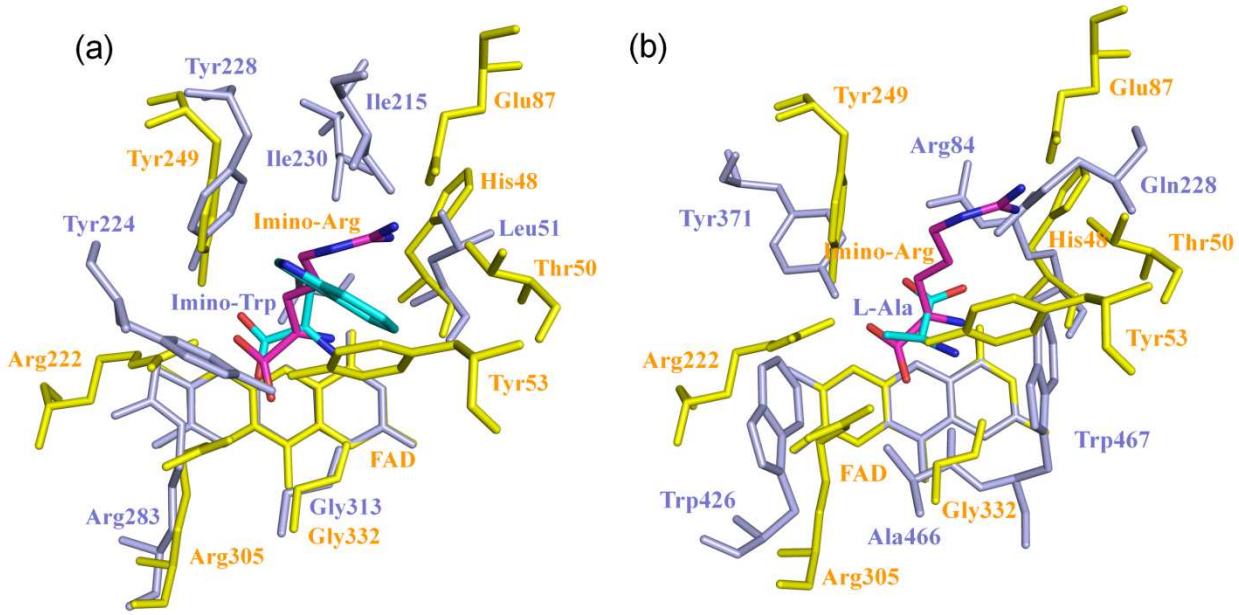


Figure 4.7 Comparison of the active sites of DADH, pDAAO and LAAO. The active site of DADH (yellow) is superimposed on that of pDAAO (a, lightblue) and LAAO (b, lightblue) along the FAD isoalloxazine ring. Iminoarginine in DADH (magenta carbon atoms) and iminotryptophan in pDAAO (cyan carbon atoms) and L-alanine in LAAO (cyan) are shown as sticks. FAD is represented by its isoalloxazine ring.

5 ATOMIC RESOLUTION STRUCTURE OF AN N(5) FLAVIN ADDUCT IN D-ARGININE DEHYDROGENASE

(Published: Fu, G., Yuan H., Wang S., Gadda, G. and Weber, I. (2011). *Biochemistry* 50(29): 6292-6294.)

5.1 Abstract

D-Arginine dehydrogenase (DADH) catalyzes the flavin-dependent oxidative deamination of D-arginine and other D-amino acids to the corresponding imino acids. The 1.07 Å atomic resolution structure of DADH crystallized with D-leucine unexpectedly revealed a covalent flavin N(5) adduct, instead of the expected iminoleucine product in the active site. This acyl adduct has been successfully reproduced by photoreduction of DADH in the presence of 4-methyl-2-oxopentanoic acid (ketoleucine). The iminoleucine may be released readily due to weak interactions in the binding site, in contrast to iminoarginine, converted to ketoleucine, which reacts with activated FAD to form the covalently linked acyl adduct.

5.2 Introduction

D-arginine dehydrogenase was first identified as the catabolic enzyme of a novel 2-enzyme-coupled system for D-to-L inversion of D-arginine [83]. This 2-component racemase allows *Pseudomonas aeruginosa*, which is a common opportunistic human pathogen, to grow with D-arginine as the sole source of carbon and nitrogen [83-84]. DADH noncovalently binds FAD as a prosthetic group and catalyzes the oxidative deamination of many D-amino acids to their corresponding imino acids with concomitant reduction of FAD. The imino acid is released into the solvent and nonenzymatically hydrolyzed to the corresponding α-keto acid and ammonia. Iminoarginine is then converted to L-arginine by the anabolic enzyme of the system, an NAD(P)H-dependent L-arginine dehydrogenase (LADH). Our group recently reported the crystal structure of DADH and the complex of DADH with iminoarginine or iminohistidine

at or close to atomic resolution (1.06-1.30 Å) [45]. Further investigation of this enzyme by performing steady-state kinetic studies has established a ping-pong bi-bi kinetic mechanism [88]. A loop region of DADH (residues 50-56) was designated as an active site lid controlling the substrate accessibility to the active site, similar to those reported in other flavin-dependent enzymes. The DADH Glu87 forms strong ionic polar interactions with iminoarginine and may play a crucial role for the strict preference of DADH for basic D-amino acids.

5.3 Results and Discussion

In order to further investigate the substrate specificity of DADH, we have performed crystallographic study on DADH in complex with a hydrophobic substrate, D-leucine. However, an acyl adduct rather than the imino products observed in previous DADH structures has been identified in this structure, which forms a covalent bond between the decarboxylated keto product Cα atom and the FAD N(5) atom. It has been reported that α-keto acids with small hydrophobic side chains are competitive inhibitors of DAAO [200]. Furthermore, these compounds are able to react with flavins and form a covalent bond between their Cα atoms and the flavin N(5) atom in the presence of light [75, 201]. The structural properties of this reaction byproduct have been thoroughly characterized by analyzing the atomic-resolution structure and its formation is studied by photoreduction of oxidized DADH in the presence of kLeu.

DADH was crystallized with the hydrophobic substrate D-leucine during our investigations into the substrate specificity of this enzyme. The structure was determined in the orthorhombic space group P212121 with one molecule per asymmetric unit. It was solved by molecular replacement using the structure of DADH/iminoarginine as template (PDB entry: 3YNE [45]), and refined to the atomic resolution of 1.07 Å and R-factor of 13.1%. The crystallographic data and refinement statistics are summarized in **Table 5.S1**. The overall structure closely resembles the structure of DADH/iminoarginine with RMSD value of 0.14 Å for 381 Cα atoms. Unlike the imino acid products observed in the other DADH structure [45], however, an acyl adduct was identified at the active site of this structure (**Figure 5.1A**). The adduct is assumed to arise from a nucleophilic substitution in which the N(5) atom of the reduced flavin formed in

the X-ray beam reacts with the Cα of ketoleucine, yielding 4-methyl-2-pentanone-FAD and CO2 (Scheme S1). A similar structure has been described in the complex of D-amino acid oxidase (DAAO) with 3-methyl-2-oxobutyric acid (kVal) (PDB entry: 1DAO, 3.20 Å, [75]). However, DAAO was deliberately soaked with kVal to form the complex in the previous study, while the acyl adduct in this report is an un-expected byproduct of co-crystallizing DADH with D-leucine. This result is consistent with previous studies showing that α-keto acids bearing small hydrophobic side chains are competitive inhibitors of DAAO and can form FAD N(5) adducts in the presence of light [75, 200-201]. Formation of this acyl ad-duct results in an additional bending of the flavin over that seen in the DADH/iminoarginine structure, characterized by a 20° angle between the pyrimidine and benzene moieties.

The decarboxylated kLeu is covalently linked to the N(5) atom of FAD. The oxygen atom of acyl group forms hydrogen bonds with the Arg222 and Tyr249 side chains (**Figure 5.1B**). Its aliphatic side chain points away from the FAD isoalloxazine ring and interacts with several hydrophobic residues in-cluding Tyr53, Met240 and Tyr249. Interestingly, a water molecule is found to mediate a hydrogen bond network between the acyl oxygen and the enzyme residues Tyr53, Arg305 and Gly332. The polar interac-tion between this water molecule and Tyr53 appears to be important for holding the active site lid in its closed conformation and no significant conformational change is observed for this loop region when compared to the structures of DADH with other D-amino acids. A water molecule is also found at the structurally corresponding position of DAAO/kVal structure, which forms polar interactions with DAAO Gln53, Tyr224 and Gly313, but not with the acyl oxygen of kVal [75]. Tyr224 is located on the active site lid of DAAO and is functionally equivalent to Tyr53 in DADH. Therefore, this water molecule may have an important role in maintaining the active site conformation when the carboxylate group is absent from the ligand.

The formation of the acyl N(5) adduct in solution was confirmed by photoreduction of DADH and 4-methyl-2-oxopentanoic acid followed by confirmation of the adduct structure by ESI mass spec-trometry. The DADH in a solution of 10 mM potassium phosphate in the presence of 50 mM EDTA and 43 mM 4-methyl-2-oxopentanoic acid was slowly reduced, as shown in **Figure 5.2A.** The bleaching of

the yellow color and the decreased absorbance at 450 nm indicate photoreduction of the flavin. The DADH was separated on a Sephadex G-25 PD-10 column and the flavin adduct was extracted with 50% acetonitrile. The spectrum of the flavin adduct (inset of **Figure 5.2A**) shows an absorbance peak in the ~420 nm region, which is consistent with formation of a reduced flavin-N(5) adduct with an oxo-compound [75, 201]. In the presence of light, the 4-methyl-2-oxopentanoic acid was decarboxylated and reacted with the flavin, yielding a covalently linked adduct between the Cα of the acid and N(5) atom of flavin, as shown in **Figure 5.2B**.

The structure of the isolated flavin was analyzed by negative ESI mass spectrometry. As shown in **Figure 5.2C**, the electrospray ion source produced a doubly charged flavin complex due to the adenosine diphosphate group of FAD. Based on the mass measurement and **Figure 5.2B**, the peak with m/z of 434.6 corresponding to a molecular weight of 871.6 indicates the addition of $C_5H_{10}O$ to FAD. Indeed, the calculated flavin adduct with doubly charged m/z is 434.8 and the difference of 0.2 mass units corresponds to a mass measurement accuracy of 0.046%. Thus, the ESI-MS measurement has confirmed the process of N(5) adduct formation.

The active site geometry and interactions of DADH with kLeu were compared to those in the iminoarginine complex. The active sites of these two structures were superimposed along the FAD isoalloxazine ring (**Figure 5.3**). The kLeu Cα atom forms a covalent bond of 1.4 Å with the FAD N(5) atom. This covalent link between kLeu and FAD results in a slight shift of Tyr53, Arg222 and Tyr249 towards the FAD isoalloxazine ring. In the structure of DADH/iminoarginine, the main chain of the ligand is slightly shifted so that its Cα is about 3.4 Å away from FAD N(5). The main chain of iminoarginine is held in position by the polar interactions of its two carboxylate oxygen atoms with the residues Tyr53, Arg222, Tyr249, Arg305 and Gly332 of DADH. kLeu forms fewer polar interactions; its acyl oxygen atom forms direct hydrogen bonds with the side chains of Arg222 and Tyr249, and a water molecule mediates a hydrogen bond network with Tyr53, Arg305 and Gly332 (**Figure 5.1B**). The charged side chain of iminoarginine forms strong polar interactions with Glu87 and Thr50, while the long aliphatic part of its side chain forms hydrophobic interactions with Tyr53, Met240, Val242, and Tyr249.

In contrast, the short hydrophobic side chain of kLeu can only form hydrophobic interactions with Tyr53, Met240, and Tyr249 and has lost contact with Val242, which displays two alternate conformations in the structure. Also, the water mediated interaction between the iminoarginine and His48 is absent in the DADH/kLeu structure. Overall, the weaker interactions of kLeu are in agreement with the kinetic data showing that the k_{cat}/K_{Arg} value is about 6600-fold higher than the k_{cat}/K_{Leu} value [45, 84].

5.4 Conclusion

In conclusion, our crystal structure of DADH with an FAD N(5) adduct solved at the atomic resolution of 1.07 Å is possibly the highest resolution for such an acyl adduct. Furthermore, photoreduction of DADH in the presence of kLeu successfully reproduced this acyl adduct, which apparently requires either intense synchrotron X-radiation or activation by light. D-leucine is a much poorer substrate of DADH by almost 4 orders of magnitude compared to the best substrate D-arginine. Comparison of the DADH complexes suggests that, in contrast to iminoarginine, the relatively small side chain of iminoleucine will form weaker interactions within the binding site. Therefore, iminoleucine may be partially dissociated from the enzyme and become hydrolyzed to ketoleucine, which then decarboxylates and reacts with photoactivated FAD to form the covalently linked acyl adduct.

5.5 Acknowledgements

We thank Dr. Johnson Agniswamy and Yuan-Fang Wang for valuable discussions. We are grateful for the assistance of the staff at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

5.6 Experimental procedures (Supporting Information)

5.6.1 Materials

4-methyl-2-oxopentanoic acid was purchased from VWR. All other reagents were of the highest purity commercially available.

5.6.2 Expression and Purification of DADH

The *Escherichia coli* strain TOP10 bearing hexa-histidine-tagged DADH was kindly provided by Dr. Chung-Dar Lu, Georgia State University. Protein expression and purification were as described previously [83]. No FAD was added during expression and purification.

5.6.3 Crystallization and X-ray Data Collection

Purified DADH was concentrated to 6 mg/mL, in 50 mM Tris at pH 7.5. DADH was co-crystallized with D-Leucine by the hanging drop vapor diffusion method at 4 ˚C using a well solution of 0.1M 2-(N-morpholino)-ethanesulfonic acid (MES) pH 6.5-7.0, 5% glycerol, and 6-10% (w/v) PEG6000. A 1:10 molar ratio of the enzyme (~0.15mM) to substrate was used. Crystals can grow to a size of 0.1-0.2 mm^3 within two weeks. The crystals were soaked in the reservoir solution with 20% glycerol as cryoprotectant for ~1 min and frozen immediately in liquid nitrogen. X-ray data were collected at 100 K on beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory.

5.6.4 Structure Determination and Model Refinement

The X-ray data were integrated and scaled using HKL2000 [178]. The structure of DADH/iminoarginine (PDB: 3YNE [45]) was employed to solve the new DADH structure by molecular replacement using PHASER [127] in the CCP4i suite of programs [128]. Surprisingly, the electron density of the ligand was quite different from those observed in other DADH structures. Further analysis suggested that a covalent bond is formed between Cα atom of ketoleucine and the flavin N(5) atom, forming a FAD N(5) adduct. A small amount of extra density was observed near the ligand acyl group, suggesting that iminoleucine may also exist at a very low occupancy in the crystal structure. Crystal structures were refined with SHELX97 [129]. Manual adjustments and rebuilding were performed using the molecular graphics program COOT [131]. Solvent molecules were inserted at stereochemically reasonable positions based on the peak height of the *2Fo–Fc* and *Fo–Fc* electron density maps, hydrogen bond interactions and interatomic distances. Anisotropic B-factors were refined for all the structures. Hydrogen atom posi-

tions were included in the last stage of refinement using all data. The crystallographic statistics are shown in **Table 5.S1**. Protein structures were superimposed on Cα atoms by using the secondary structure matching (SSM) module in COOT [182]. Figures of the structures were generated with PYMOL (http://www.pymol.org).

5.6.5 Anaerobic Light Irradiation

DADH was expressed and purified as previously described in [88]. Photoreduction studies were performed as described by Massey [202]. DADH was subjected to gel filtration with Sephadex G-25 column (PD-10, Amersham-Pharmacia Biotech) using 10 mM potassium phosphate and 50 mM EDTA, pH 8.0. The enzyme solution then mixed with final concentration of 43 mM 4-methyl-2-oxopentanoic acid, 0.2 µM FAD, 2 µM methyl viologen. Solutions contained in the anaerobic spectrophotometer cell were made anaerobic by repeated cycles of degassing and flushing with argon. The anaerobic cuvette was then placed in a water bath at 4 ^{o}C. The light irradiation was carried out by 100W tungsten light and spectra were recorded at intervals using an Agilent Technologies Hewlett-Packed 8453A diode array spectrophotometer. Complete photoreduction was attained with 710 min irradiation. The enzyme mixture and excess reactant were separated through Sephadex G-25 PD-10 column equilibrated with H_2O. Fractions containing DADH were collected and pooled together. The flavin adduct was extracted by denaturing the enzyme with 50% acetonitrile. After centrifugation to remove denatured protein, the supernatant containing the extracted flavin was used for mass spectrometry analyzing.

5.6.6 ESI TOF mass spectrometry studies

The flavin adduct sample was analyzed by mass spectrometry on an ESI at the Mass Spectrometry Laboratory of the Georgia State University, Atlanta. The mass spectrometry analysis was performed on a Waters Q-TOF micro mass spectrometer equipped with electrospray ionization source (ESI) in negative ion mode (Waters Corporate, Milford, MA). The mass spectrometer was calibrated with sodium formate and the sample was analyzed with Leucine Enkephalin as internal standard (*m/*z=554.2615 Da). The mass range was from 100-1000 Da. The sample was introduced into the ion source through direct

infusion at 5 μl/min flow rate. The instrument parameters were as follows: capillary voltage of 3000 V, cone voltage of 30 V, desolvation temperature of 100 ^{o}C and source temperature of 70 ^{o}C. Nitrogen was used as nebulizing and drying gas. Data were analyzed using MassLynx 4.1 software.

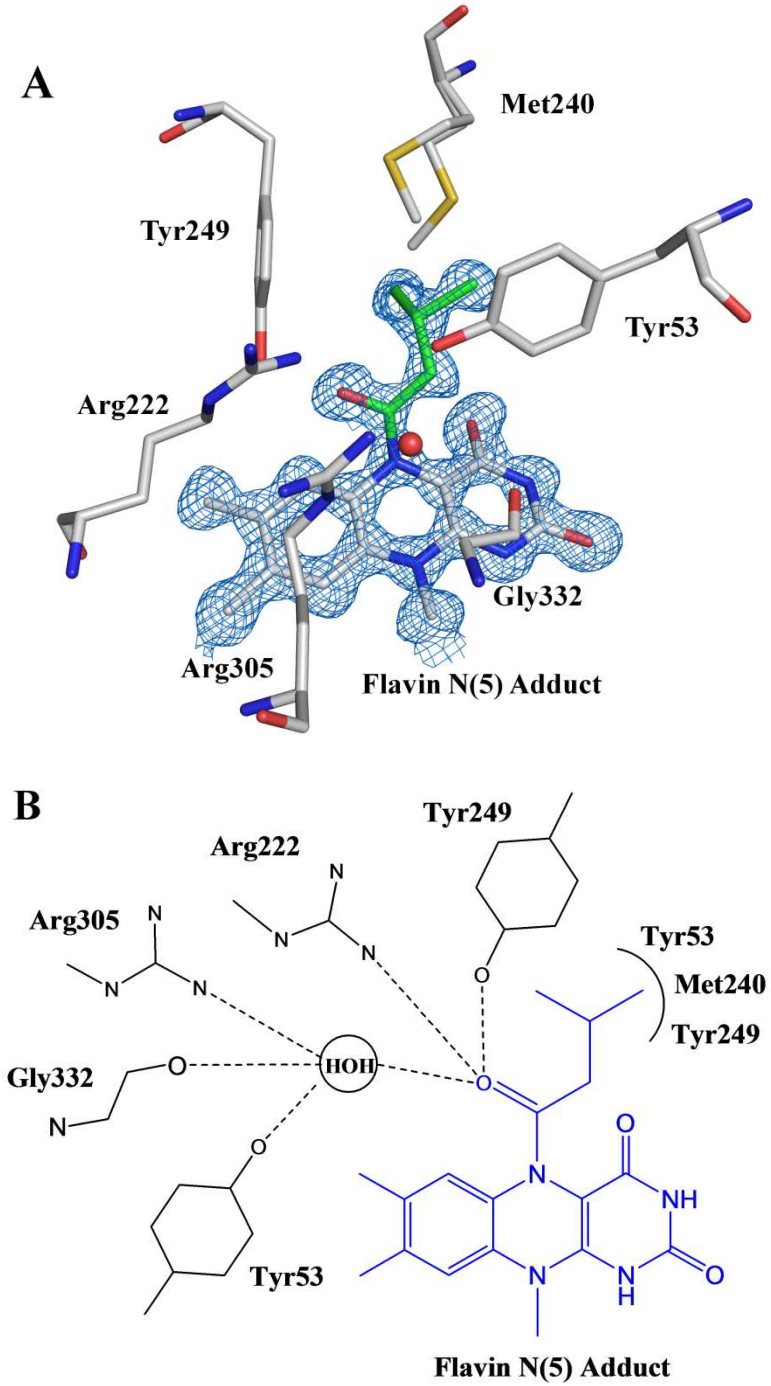


Figure 5.1 Interaction between flavin N(5) adduct and DADH. (A) Active site of the DADH/kLeu structure. Carbons are colored gray for DADH active site residues and green for kLeu part of the acyl adduct. Water molecule is shown as red sphere. *Fo-Fc* omit map of the flavin N(5) adduct is indicated as blue mesh and contoured at 3σ. (B) Schematic diagram of the interactions between DADH and the flavin N(5) adduct (blue color). Hydrogen bonds are indicated by dashed lines, while curved lines indicate van der Waals interactions.

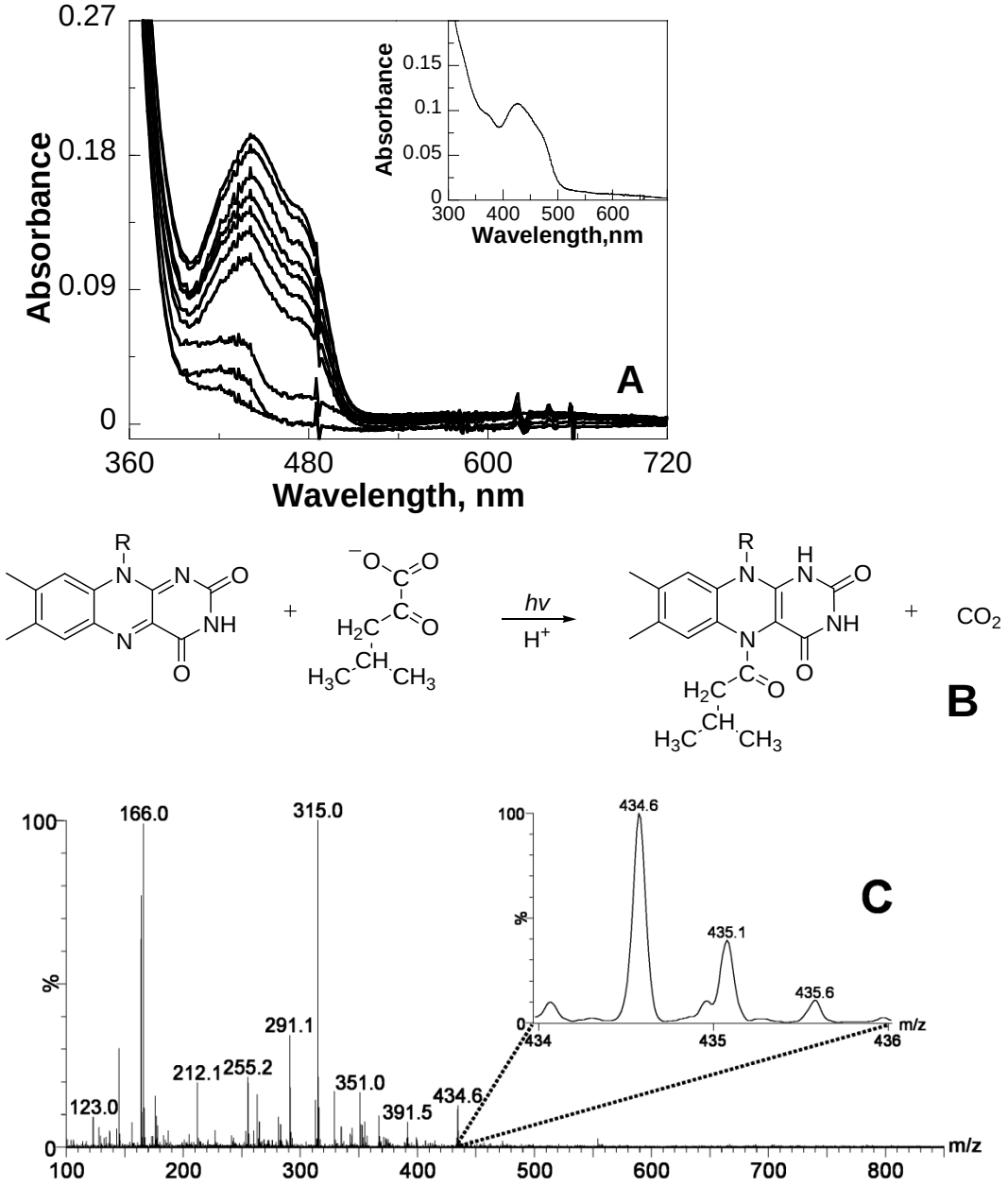


Figure 5.2 Anaerobic light irradiation of DADH. Panel A shows the effect of light on DADH under anaerobic conditions in the presence of 43 mM 4-methyl-2-oxopentanoic acid, 0.2 µM FAD, 2 µM methyl viologen. Curves for top to bottom represent 0 min, 70 min, 165 min, 280 min, 370 min, 460 min, 525 min, 585 min, 620 min, and 710 min. Further irradiation to 830 min and 955 min did not yield spectral changes. The inset shows the extracted flavin adduct in 50% acetonitrile. The spectrum of the flavin adduct is estimated by multiplying by a factor of 3. Panel B: Formation of 4-methyl-2-pentanone-FAD through reductive decarboxylation of FAD and 4-methyl-2-oxopentanoic acid by photoirradiation. Panel C: Negative ion ESI mass spectrum of the isolated flavin adduct. The peak at m/z 434.6 corresponds to the $(M\text{-}2H)^{2-}$ ion of the 4-methyl-2-pentanone-FAD extracted from DADH. Inset, expansion of the $(M\text{-}2H)^{2-}$ ion region 434-437 of the mass spectrum.

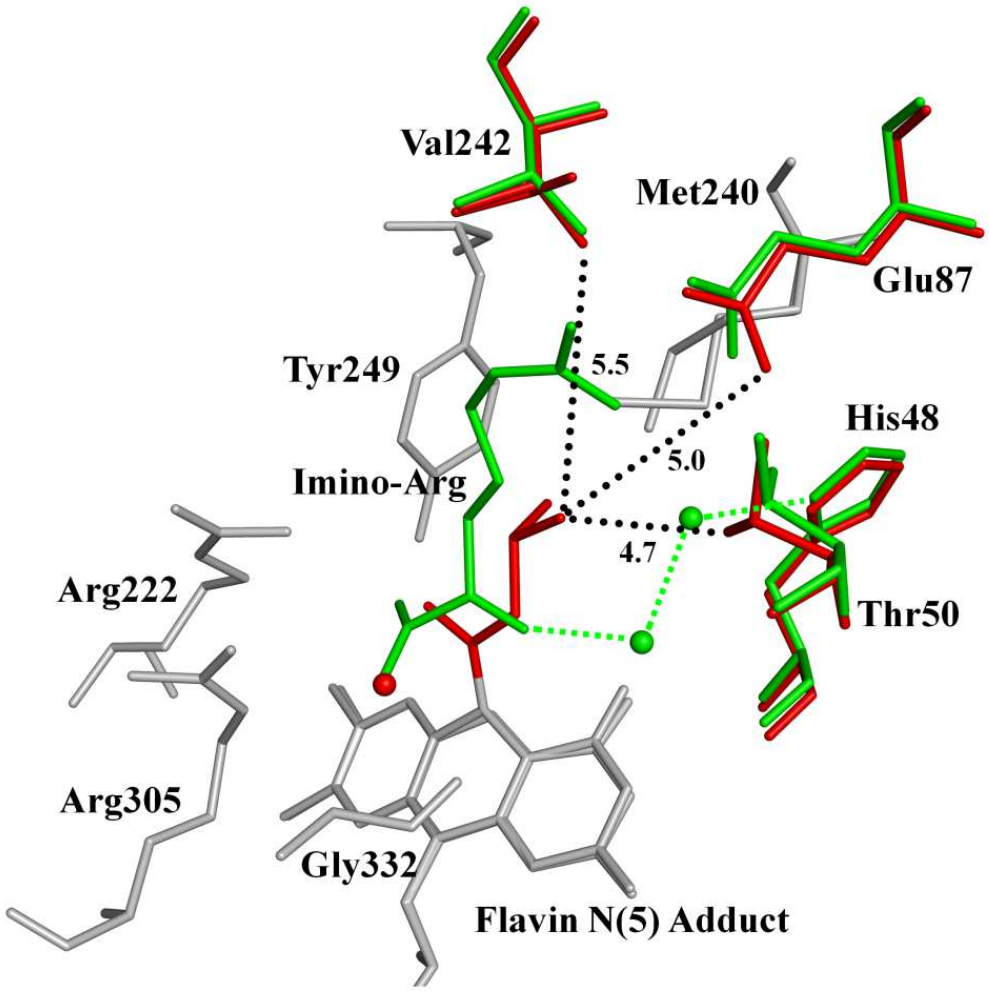
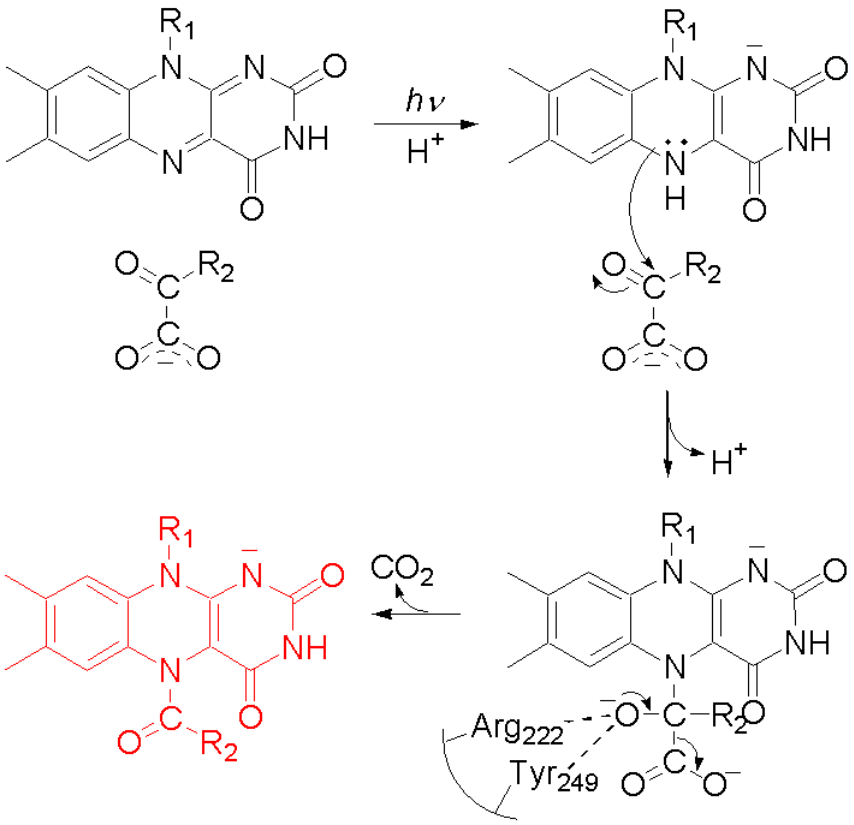


Figure 5.3 Comparison of the active sites of DADH structures in complex with iminoarginine and kLeu. Residues from DADH/iminoarginine forming similar interactions in both structures are shown as gray sticks. The ligands, water and the residues showing differences in interactions are colored red for kLeu adduct and green for iminoarginine. Water molecules are indicated by spheres. kLeu has lost interactions with several DADH residues as shown by interatomic distances (dotted lines labeled in Å). The water mediated interaction of iminoarginine with His48 (green dashed lines) is absent in the DADH/kLeu structure. Also, kLeu forms a water-mediated hydrogen bond with Tyr53, unlike the direct interaction of iminoarginine (Tyr53 was omitted for clarity).

Table 5.1 Crystallographic Data Collection and Refinement Statistics (S1)

	DADH/D-Leu
Data Collection Statistics	
Redox state of flavin	Reduced
Wavelength (Å)	0.80000
Space group	$P2_12_12_1$
a (Å)	62.07
b (Å)	77.93
c (Å)	89.31
Resolution range (Å)	50-1.07
Total observations	2,025,781
Unique reflections	187,536
Completeness	98.6(80.8) [a]
$\langle I/\sigma(I) \rangle$	32.9(2.3)
R_{sym} (%)	6.7(43.8)
Refinement Statistics	
Resolution range (Å)	10-1.07
R_{work} (%)	13.1
R_{free} (%)	15.4
Mean B-factor ($Å^2$)	
Protein	12.7
FAD	5.9
Ligand	14.8
Water	23.3
Number of atoms	
Protein	3010
FAD	53
Ligand	6
Water	363
r.m.s. deviations	
Bond length (Å)	0.017
Angle [b]	0.034

[a] Values in parentheses are given for the highest resolution shell.
[b] The angle RMSD in SHELX97 is indicated by distance in Å.



Scheme 5.S1. Proposed mechanism for the formation of 4-methyl-2-pentanone-FAD from reduced DADH and ketoleucine. Red defines the 4-methyl-2-pentanone-FAD observed in the active site of DADH. R_1 is adenosine diphosphate ribose; R_2 is $–CH_2CH(CH_3)_2$.

6 PROBING THE SUBSTRATE SPECIFICITY OF D-ARGININE DEHYDROGENASE BY CRYSTALLOGRAPHY AND COMPUTATIONAL MODELING

Guoxing Fu, Robert W. Harrison and Irene T. Weber

6.1 Abstract

D-Arginine dehydrogenase (DADH) catalyzes the flavin-dependent oxidative deamination of D-amino acids to the corresponding imino acids. In order to investigate the molecular basis for the high activity of DADH on basic and hydrophobic D-amino acids crystal structures of DADH complexes with reaction products iminolysine, iminomethionine, iminophenylalanine and iminoproline were solved at atomic resolution (1.03-1.09 Å). The structure of DADH/iminolysine confirms that DADH residue Glu87 forms ionic interactions with positively charged D-amino acids. The hydrophobic side chains of iminophenylalanine and iminomethionine form hydrophobic contacts with DADH residues Tyr53, Met240, Val242 and Tyr249 lining the binding pocket. The structure of DADH/iminoproline shows fewer ligand contacts within the binding pocket and alternate conformations for residues 46-48 and 50-52 of the active site lid. Molecular mechanics calculations on DADH with 16 imino acids gave excellent agreement with measured kinetic data with a correlation coefficient of 0.84 after estimating the relative populations of charged and neutral states for the side chain of the trapped product (the correlation coefficient is 0.77 without this estimate). Finally, comparison of DADH with D-amino acid oxidase and D-aspartate oxidase gave new insights into key structural features contributing to the distinct substrate specificity of these enzymes.

6.2 Introduction

D-amino acids and their metabolizing enzymes have important, but poorly understood, roles in living organisms compared to those of L-amino acids, which are the predominant form in biologically

synthesized proteins. In bacteria, D-amino acids are fundamental elements of the cell wall peptidoglycan layer [63], while D-leucine, D-methionine, D-tyrosine, and D-tryptophan were reported to regulate disassembly of bacterial biofilms [64]. In eukaryotes, specific D-amino acids such as D-aspartate and D-serine have critical roles in regulating several physiological processes [66]. Catabolism of D-amino acids is commonly catalyzed by D-amino acid oxidases or dehydrogenases, which have been identified in bacteria, archaea, and eukaryotes. D-amino acid oxidase (DAAO) has been the focus of numerous studies and a paradigm for flavin-dependent oxidases since its identification in 1935 [48]. In bacteria, D-amino acids are deaminated by flavin-dependent dehydrogenases [73], which allows the bacteria to grow using D-amino acids as a nutrient [71-72].

D-arginine dehydrogenase was identified as the catabolic enzyme of a novel 2-enzyme-coupled system for D-to-L inversion of D-arginine. This 2-component racemase allows *Pseudomonas aeruginosa*, which is a common opportunistic human pathogen, to grow with D-arginine as the sole source of carbon and nitrogen [175]. DADH noncovalently binds FAD as a prosthetic group and catalyzes the oxidative deamination of many D-amino acids to their corresponding imino acids with concomitant reduction of FAD. The imino acid is then released into the solvent and nonenzymatically hydrolyzed to the corresponding α-keto acid and ammonia. Iminoarginine is then converted to L-arginine by the coupled anabolic enzyme, an NAD(P)H-dependent L-arginine dehydrogenase (LADH).

Our group recently reported the crystal structure of DADH determined at the atomic resolution of 1.06 Å, while the structures of DADH with iminoarginine and iminohistidine reaction products were determined at 1.30 Å resolution [45]. Further investigation of this enzyme by steady-state kinetic studies has established a ping-pong bi-bi kinetic mechanism [88]. A loop region of DADH (residues 50-56) was designated as an active site lid controlling the substrate accessibility to the active site, similar to those reported in other flavin-dependent enzymes [75, 188]. Moreover, the residues Ser45-Ala46 were observed to switch conformation between the ligand-free and product-bound structures of DADH. Distinct side-chain conformations were observed for the imino products of D-arginine and D-histidine bound in the active site. Glu87 of DADH formed ionic interactions with iminoarginine, which is likely responsible for the

high activity of DADH for positively charged residues. The different interactions and binding modes observed for iminoarginine and iminohistidine are consistent with the 1000-fold difference in k_{cat}/K_m values for D-arginine and D-histidine.

DADH is characterized by broad substrate specificity, being able to oxidize basic and hydrophobic D-amino acids of various sizes, but not reacting with acidic D-amino acids [45, 84, 203-204]. Similarly, DAAO acts on many neutral and basic D-amino acids but not acidic D-amino acids. The oxidative deamination of acidic D-amino acids, such as D-aspartate and D-glutamate, is specifically catalyzed by D-aspartate oxidase (DDO), which is widely distributed in a variety of eukaryotes, but has not been identified in bacteria or plants yet [205]. In this study, we co-crystallized DADH with basic and hydrophobic D-amino acids: D-lysine, D-methionine, D-phenylalanine and D-proline. The atomic resolution crystal structures facilitated detailed analysis of the enzyme-substrate interaction. Additionally, we performed molecular mechanics energy calculations on DADH with 19 imino acids to further explore the molecular basis for the substrate specificity of DADH. Finally, structural and functional comparison of DADH with DDO and DAAO has given insights into the key residues responsible for the recognition of different D-amino acids.

6.3 Experimental Procedures

6.3.1 Expression and Purification of DADH

The *Escherichia coli* strain TOP10 bearing hexa-histidine-tagged DADH was kindly provided by Dr. Chung-Dar Lu, Georgia State University. Protein expression and purification were as described previously [83].

6.3.2 Crystallization and X-ray Data Collection

Purified DADH was concentrated to 6 mg/mL in 50 mM Tris at pH 7.5. DADH was co-crystallized with selected D-amino acids by the hanging drop vapor diffusion method at 4 ˚C using a 1:10 molar ratio of the enzyme (~0.15mM) to substrate. The well solution was 0.1 M 2-(N-morpholino)-

ethanesulfonic acid (MES) pH 6.5-7.0, 5% glycerol, and 6-10% (w/v) PEG6000. Crystals can grow to a size of 0.1-0.2 mm^3 within two weeks. The crystals were soaked in the reservoir solution with 20% glycerol as cryoprotectant for ~1 min and frozen immediately in liquid nitrogen. X-ray data were collected at 100 K and a wavelength of 0.85 Å on beamline 22-ID of the Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source, Argonne National Laboratory.

6.3.3 Structure Determination and Model Refinement

The X-ray data were integrated and scaled using HKL2000 [178]. The crystal structure of DADH/iminoarginine (Protein Data Bank (PDB) entry: 3YNE) was employed to solve the structures of DADH in complexes with iminolysine, iminomethionine, iminophenylalanine and iminoproline by molecular replacement using PHASER [127] in the CCP4i suite of programs [128]. Notably, all the FAD atoms were distinctly visible in the electron density maps. Crystal structures were refined with SHELX97 [129]. Manual adjustments and rebuilding were performed using the molecular graphics program COOT [131]. The active site lid and the Ala46-Ser48 region of the DADH/iminoproline complex were refined with two alternate conformations where the major (~70% occupancy) and minor conformations (~30% occupancy) corresponded to the product-bound and ligand-free conformations in the first reported DADH structure (PDB entry: 3NYC [45]). Solvent molecules were inserted at stereochemically reasonable positions based on the peak height of the 2*Fo–Fc* and *Fo–Fc* electron density maps, hydrogen bond interactions and interatomic distances. Anisotropic B-factors were refined for all the structures. Hydrogen atom positions were included in the last stage of refinement using all data. Protein structures were superimposed on Cα atoms by using the secondary structure matching (SSM) module in COOT [182]. Figures of the structures were generated with PYMOL (http://www.pymol.org).

6.3.4 Molecular Modeling of DADH

Models of DADH complexes with different imino acids were built using the program AMMP [29]. The models were generated from the crystal structure of DADH/iminoarginine. The other imino acids were generated from the main-chain atoms of iminoarginine. The template and restraints for FAD

were generated from the crystal structure. The hydrogen atoms were built with the tuna potential set [153] using 15 cycles of the analytic model building algorithm in AMMP [137, 206], followed by conjugate gradients minimization of the non-bonded and geometrical terms. The charges for each imino acid were generated by using the method of moments atomic charge calculation implemented in AMMP before energy minimization [163]. Finally, each DADH/imino acid complex was optimized by 1,000 steps of conjugate gradients energy minimization. The total non-bonded interaction energy was calculated for each protein-ligand model. The experimental binding energy was calculated as $\Delta G = -RT \ln (k_{cat}/K_m)$ from the 16 measured kinetic constants [45]. Only the data from 16 imino acids with experimentally determined kinetic parameters (omitting iminocysteine, iminoglutamate and iminoaspartate) were used to calculate the correlation with calculated energies.

6.3.5 Molecular modeling of DDO

Due to the lack of an experimental structure of DDO, a theoretical model of the mouse DDO structure was constructed for comparison with DADH. Sequence similarity searches with available crystal structures were performed by BLAST [181] and the AMMP structure prediction server (http://bmcc3.cs.gsu.edu) [137]. The structure of porcine DAAO showed the highest sequence identity (40.3%) with DDO. Therefore, the crystal structure of porcine DAAO (PDB entry: 1DDO [75]) was submitted to the AMMP structure prediction server to construct the initial model of DDO. FAD was then incorporated into the model and hydrogen atoms were generated by using the AMMP program. The enzymatic reaction product, iminoaspartate, was modeled in the active site to allow a comprehensive comparison with the structures of DADH.

6.4 Results and Discussion

6.4.1 Overall structure of DADH with four different imino acids

DADH was crystallized with three hydrophobic D-amino acids, D-methionine, D-phenylalanine and D-proline, as well as the basic D-amino acid D-lysine, in order to investigate the substrate specificity

of this enzyme. All structures were solved in the orthorhombic space group $P2_12_12_1$ with one molecule per asymmetric unit and were refined to atomic resolutions of 1.03-1.09 Å and R-factors of 16.2-17.8 %. The crystallographic data and refinement statistics are summarized in **Table 1**. All 375 DADH residues were defined clearly in the electron density map, as well as the N-terminal hexa-histidine tag, which forms several direct or water-mediated polar interactions with residues from symmetry related molecules. The main chain conformations of DADH in the four complexes closely resemble our previously reported structure of DADH/iminoarginine with pairwise RMSD values of 0.09-0.11 Å for 381 Cα atoms. The cofactor FAD adopts an elongated conformation with the adenine ring buried inside the FAD-binding domain and the isoalloxazine ring located at the interface with the substrate-binding domain (**Figure 1**), as observed in the reported structures [45].

6.4.2 Interactions of DADH with iminolysine

Based on the previously determined structure of DADH with iminoarginine, we proposed that Glu87 was critical for recognition of positively charged D-amino acids as substrates. Therefore, DADH was co-crystallized with D-lysine to confirm the role of Glu87. As expected, the bound iminolysine forms similar polar interactions with Glu87 as observed for iminoarginine. Iminolysine binds with the plane formed by its Cα atom, carboxylate group and imino group approximately parallel to the *re* face of the flavin, while the side chain points away from the FAD (**Figure 2A**). A slight difference was observed near the α-carbon of the iminolysine, which is 3.0 Å away from the FAD N5 atom. This distance is about 0.4 Å longer in the structure of DADH/iminoarginine (**Figure 2B**), possibly due to the shorter interactions of iminoarginine with Glu87 relative to those of iminolysine.

Several DADH residues form extensive polar interactions with the two main-chain carboxyl oxygen atoms of iminolysine and anchor the ligand in the active site. One of the iminolysine carboxyl oxygen atoms forms hydrogen bonds with the side-chain hydroxyl of Tyr53, the guanidinium side chain of Arg305, and the carbonyl oxygen of Gly332. The other ligand carboxyl oxygen forms hydrogen bonds with the side-chain nitrogen of Arg222 and side-chain hydroxyl of Tyr249. The imino group of

iminolysine remains on the same plane as the carboxylate atoms, unlike the out of plane imino group of iminoarginine, resulting in the loss of the polar interaction with the FAD O4 atom; instead, a hydrogen bond is formed with the carbonyl oxygen of Gly332 (**Figure 2A**). The two water molecules that mediate the hydrogen-bond network between the ligand imino group and the imidazole side chain of His48 are retained in both structures. The side chain of iminolysine is recognized by two hydrogen bonds (2.7 Å and 3.0 Å long) with the carboxylate group of Glu87, which is consistent with the high activity of DADH for positively charged D-amino acids [45, 84, 88].

6.4.3 Interactions of DADH with hydrophobic imino acids.

In order to understand its substrate specificity for hydrophobic residues, structures were also determined for DADH crystallized with D-methionine, D-phenylalanine and D-proline, which have 260~10,000-fold lower k_{cat}/K_m values when compared to that of D-arginine. The hydrophobic walls (Tyr53, Met240, Val242, and Tyr249) of the DADH active site create a favorable environment for hydrophobic D-amino acids with larger side chains. Iminomethionine and iminophenylalanine form extensive van der Waals interactions with the hydrophobic residues in the binding site (**Figure 3A, 3B**). Interestingly, the benzene ring of iminophenylalanine forms parallel-displaced π-stacking with Tyr53 and T-shaped π-stacking with Tyr249, which are favorable arrangements for aromatic-aromatic interactions (π…π and C-H…π) [207-208]. Iminoproline has the smallest side chain of the three hydrophobic imino acids and shows fewer hydrophobic contacts with DADH (**Figure 3C**). The diminished interactions of iminoproline are in excellent agreement with the substrate specificity of DADH, since D-proline is a poorer substrate than the larger hydrophobic D-amino acids. Detailed examination of the DADH/iminoproline structure revealed conformational disorder around the enzyme active site. Residues Thr50-Ala52 in the active site lid show a minor alternative conformation (~30% occupancy) (**Figure 4A, red mesh**) similar to the ligand-free conformation described previously [45]. However, Tyr53, which forms a hydrogen bond with the ligand, has a single conformation as seen in the product-bound complex without visible disorder (**Figure 4A, green mesh**). On the *si* face of the flavin ring, Ala47, rather than

Ala46 in the previous DADH structure, displays an alternate conformation directed away from the flavin. This structural disorder may arise from the small side chain and weak affinity of D-proline.

Superimposition of the active sites of DADH illustrates the flexible conformations adopted by Thr50, Glu87, Met240 and Val242 in complexes with basic and hydrophobic imino acids (**Figure 4B**). In contrast, DADH residues interacting with the product main-chain atoms display almost identical conformations. More variation was seen for the side chains of Thr50, Glu87 and Val242. Glu87 mainly interacts with iminoarginine and iminolysine, although slight structural shifts are seen in other structures. Thr50 has the largest variation since it forms polar interactions with iminoarginine and iminolysine and has hydrophobic contacts with iminophenylalanine. However, Thr50 does not interact with iminomethionine, which has moved toward Val242. The smaller iminoproline does not form hydrophobic contacts with the above enzyme residues. The water-mediated hydrogen bond network depicted in **Figure 2B** is conserved in all these structures, except for the complex with iminophenylalanine, in which only one water molecule is visible, probably due to exclusion of the other water by the bulky phenylalanine side chain.

6.4.4 Predicted selectivity of DADH for different D-amino acids

Molecular modeling calculations were employed to analyze the substrate selectivity of DADH for all 19 D-amino acids. The observation that imino acids are trapped in the crystal structures, and a product release experiment with D-arginine, suggested that the release of imino product is the rate-limiting step of the reaction [45, 88]. Therefore, the calculation on product complexes is likely to reflect the reaction kinetics. The comparison of model and crystal structure of DADH with iminoarginine showed great similarity with slight differences appearing in the ligand interaction with Glu87 as shown in **Figure S1**. The predicted conformations of the imino acids and their interactions with protein residues closely resemble those observed in the crystal structures with only minor deviations, except for iminohistidine (**Figure S2**). Iminohistidine in the crystal structure showed several alternate conformations and lacked the two water molecules present in the binding sites for other imino acids. These two water molecules were included in the models and may prevent iminohistidine forming the conformation(s) observed in the crystal structure.

The molecular mechanics interaction energies were calculated for the 19 imino acid products as listed in Table S1. The experimentally determined kinetic parameters are available for 16 D-amino acid substrates [45, 84]. The kinetic data excluded D-cysteine because its -SH group causes reduction of the phenazine methosulfate in the enzyme assay, while D-glutamate and D-aspartate are not substrates of DADH. Changes in the free energy of the reaction can be related to the measured kinetic parameters by using the equation $\Delta G = -RT \ln(k_{cat}/K_m)$ [209]. The predicted trends in binding affinity showed high agreement with the ranking in our kinetic analysis of natural substrates [45, 84, 88]. The calculated interaction energies gave a respectable correlation coefficient of 0.77 with the experimentally derived binding energies (**Figure 5A**).

The charged state was analyzed further for iminoarginine and iminolysince since the calculated interaction energies appeared to be overestimated relative to the other imino acids. There is little direct information about aqueous ionization constants for imino acids since they are unstable in water and quickly hydrolyze into keto acids and ammonium ions [210]. Iminoarginine and iminolysine are titratable ligands so that the experimental binding data will reflect the population weighted average binding energy of both charged and neutral ligands. The initial calculations assumed fully charged states (+1), since the side chain pKa values for lysine (10.53) and arginine (12.48) are considerably higher than the pH of the kinetic experiment (8.7). The pKa of the side chain may be elevated due to the presence of the α–amino group on the amino acid, and it is possible that the pKa is lower in the α–imino acid. Both phenylguanidine and n-acetyl guanidine have lower pKa values than arginine, which is consistent with a shift of the pKa in an imino acid [211]. Therefore, it is necessary to consider the possibility of a mixture of charged and uncharged molecules for iminoarginine and iminolysine. Calculations were performed for the neutral species as well as the charged forms. The correlation coefficient between calculated and experimental binding energies dropped from 0.77 for fully charged imino acids to 0.67 for neutral forms (**Figure 5A**), which is similar to the correlation coefficient of 0.66 when the basic imino acids are excluded. Linear combinations of the interaction energies for the neutral and fully charged side chains were used to estimate the relative populations of the two ionization states. The correlation between experimental data

and interaction energies calculated with mixed populations of ionized species was, as expected, a smooth curve with a single maximum (**Figure 5B**). The combination of 17% charged and 83% neutral gave the highest correlation coefficient of 0.84 (**Figure 5A**), which suggests that DADH stabilizes binding of the neutral intermediate rather than the fully charged state.

Overall, the observed free energies (ΔG) derived from experimental k_{cat}/K_m values showed excellent correlation with the calculated interaction energies for the models. This agreement verifies the accuracy of the calculations on DADH with the imino acid products. The calculations confirm that D-arginine and D-lysine are the best two substrates for DADH, while D-aspartic acid and D-glutamic acid are the worst.

6.4.5 Key structural features determining substrate specificity

Our structural and kinetic studies have characterized the substrate specificity of DADH. This enzyme prefers positively charged polar residues and also interacts well with several hydrophobic residues. Structural analysis suggested that Glu87 is responsible for its strong preference for D-arginine and D-lysine. Similar key structural components determining the substrate specificity have been observed in different enzymes from many organisms. DAAO catalyzes oxidative deamination of D-amino acids and is widespread in nature, from microorganisms to mammals. DAAO shows broad substrate specificity for neutral and basic D-amino acids but not acidic D-amino acids, with D-phenylalanine determined as its best substrate [212]. On the other hand, the oxidative deamination of acidic D-amino acids, D-aspartate and D-glutamate, is specifically catalyzed by D-aspartate oxidase [205]. To gain insights into how these enzymes recognize different D-amino acids, detailed sequence and structural analyses were performed on DADH, DAAO and DDO.

DADH shares rather low sequence identities with DAAO (17.2%) and DDO (16.3%), while DAAO and DDO are more similar (40.3% identity). Despite their low sequence similarity, DADH and DAAO display similar overall structures (RMSD of 2.4 Å for 270 Cα atoms). Katane et al. have investigated the substrate specificity of mouse DDO [212-213]. Their kinetic and mutagenesis studies have re-

vealed that Arg216 is critical for its restrict preference for acidic D-amino acids like D-glutamate and D-aspartate [212]. Unfortunately, the experimental structure of mouse DDO is not available. Hence, we constructed a theoretical model of mouse DDO in complex with iminoaspartate by applying the DAAO structure as the template to the AMMP structure prediction server to perform structural comparison with the other two enzymes.

Comparison of the active site composition and geometry revealed significant similarities among the three enzymes and some key components that may be responsible for their distinct substrate specificities. In the DADH/iminoarginine complex (PDB entry: 3NYE), two arginines Arg222 and Arg305 form ionic interactions with the ligand carboxylate (**Figure 6A**), while the ligand carboxylate group forms a salt bridge with the guanidinium side chain of Arg283 in the pDAAO/iminotryptophan complex (PDB entry: 1DDO) (**Figure 6B**). Arg278 is predicted to be at the same location in DDO and may serve a similar function in binding to iminoaspartate (**Figure 6C**). The main chain carboxylate oxygen atoms of the ligands also form polar interactions with Tyr53, Tyr249 and Gly332 of DADH, and Tyr224, Tyr228 and Gly313 of DAAO. In the modeled structure of DDO, Tyr223, Arg237 and Ser308 have almost equivalent locations with Tyr53, Tyr249 and Gly332 of DADH and may form similar interactions with the ligand main-chain atoms. A previous study on mouse DDO suggested that Arg237 is possibly involved in substrate recognition but may not be essential for its enzymatic activity against acidic D-amino acids [212].

The three enzymes have very different residues around the ligand side chain, which may be mainly responsible for their distinct substrate specificities. In DADH, iminoarginine forms strong polar interactions with Glu87. However, a glycine residue is found at the structurally equivalent position in both DAAO (Gly94) and DDO (Gly94) structures. The glycines are remote (~12 Å) from the ligand Cα in both enzymes and likely have little influence on the substrate specificity. It was suggested that Ile215 in DAAO is important for its preference for hydrophobic D-amino acids [212, 214], while Arg216 at the same position in DDO may play a crucial role in its high selectivity for acidic D-amino acids [212]. The detailed structural comparison of these enzymes indicated that although they share active site similarities

in terms of recognition of D-amino acids, different arrangement and composition of the key structural features provide very distinct substrate specificities.

6.5 Conclusions

The crystal structures of DADH with different imino acids and the computational modeling analysis have highlighted important structural differences that rationalize the catalytic activities and substrate specificity of the enzyme. The structure of DADH with iminolysine further confirms the crucial role of Glu87 in the strong preference for basic D-amino acids. Moreover, the structures of DADH in complexes with hydrophobic imino acids revealed the hydrophobic residues responsible for the binding of bulky hydrophobic substrates. Our predicted structural models closely resemble the crystal structures and the predicted DADH-imino acid interaction energies showed excellent agreement with the kinetic data for D-amino acid substrates. This computational analysis provides a powerful approach for investigating protein-ligand interactions. Structural comparison of DADH with related enzymes suggests a few key structural components may be critical for the distinct substrate specificity of enzymes. Overall, the new high-resolution structures and predictive molecular mechanics calculations will guide future studies of DADH and other flavin-dependent enzymes.

6.6 Acknowledgements

G. F. was supported by the Georgia State University Molecular Basis of Disease Fellowship. The authors are thankful to Dr. Chung-Dar Lu for providing the dauA gene encoding D-arginine dehydrogenase. We thank Dr. Johnson Agniswamy and Yuan-Fang Wang for valuable discussions. Data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) beamline 22ID at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

Table 6.1 Crystallographic Data Collection and Refinement Statistics

	DADH/ D-Lys	DADH/ D-Met	DADH/ D-Phe	DADH/ D-Pro
Data Collection Statistics				
Wavelength (Å)	0.85	0.85	0.85	0.85
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	62.14	62.11	62.20	62.10
b (Å)	78.39	78.44	78.40	78.21
c (Å)	89.91	90.05	89.92	89.85
Resolution range (Å)	50-1.09	50-1.05	50-1.05	50-1.03
Total observations	1080,291	1294,535	1296,946	1434,616
Unique reflections	182,430	203,230	203,913	212,660
Completeness	99.6 (97.1) [a]	99.7(99.0)	99.9(99.4)	98.7(76.8)
$\langle I/\sigma(I)\rangle$	24.3(2.2)	19.7(2.2)	26.5(2.2)	24.5(2.1)
R_{sym} (%)	6.7(52.5)	8.8(64.4)	6.2(60.9)	7.2(51.0)
Refinement Statistics				
Resolution range (Å)	10-1.09	10-1.05	10-1.05	10-1.03
R_{work} (%)	14.9	15.1	13.8	14.1
R_{free} (%)	17.8	17.6	16.1	16.2
Mean B-factor ($Å^2$)				
Protein	13.2	12.7	13.7	12.5
FAD	6.1	5.4	6.1	5.6
Ligand	24.5	28.0	29.3	14.8
Water	27.9	19.6	26.2	23.4
Number of atoms				
Protein	2993	2967	2994	2968
FAD	53	53	53	53
Ligand	10	9	12	8
Water	389	232	363	318
r.m.s. deviations				
Bond length (Å)	0.016	0.017	0.017	0.016
Angle[b]	0.034	0.035	0.035	0.035

[a]Values in parentheses are given for the highest resolution shell.
[b]The angle rmsd in SHELX97 is indicated by distance in Å.

Table 6.2 Calculated interaction energies of DADH with different imino acids (S1)

substrates	k_{cat}/K_m, $M^{-1}s^{-1}$	Predicted binding energy (kcal/mol)
D-arginine	$(3.4 \pm 0.3) \times 10^6$	-235.34
D-lysine	$(5.3 \pm 0.2) \times 10^5$	-256.34
D-tyrosine	27,600±3,800	-134.9
D-methionine	14,800±600	-132.2
D-phenylalanine	6,900±300	-140.5
D-histidine	3,140 ± 30	-77.5
D-leucine	515±60	-152.4
D-proline	420 ± 10	-112.5
D-tryptophan	245 ± 3	-116.8
D-isoleucine	195 ± 3	-135
D-glutamine	186 ± 3	-106.8
D-valine	47 ± 1	-109.7
D-alanine	41 ± 1	-92.6
D-asparagine	16 ± 1	-74.5
D-serine	3.8 ± 0.1	-77
D-threonine	0.75 ± 0.01	-78.3
D-cysteine	nd[a]	-82.4
D-glutamic acid	-[b]	351.2
D-aspartic acid	-	381

[a]Not determined. PMS was reduced by cysteine. [b]Cannot be saturated with the substrate, thereby k_{cat} and K_m values are not reported.

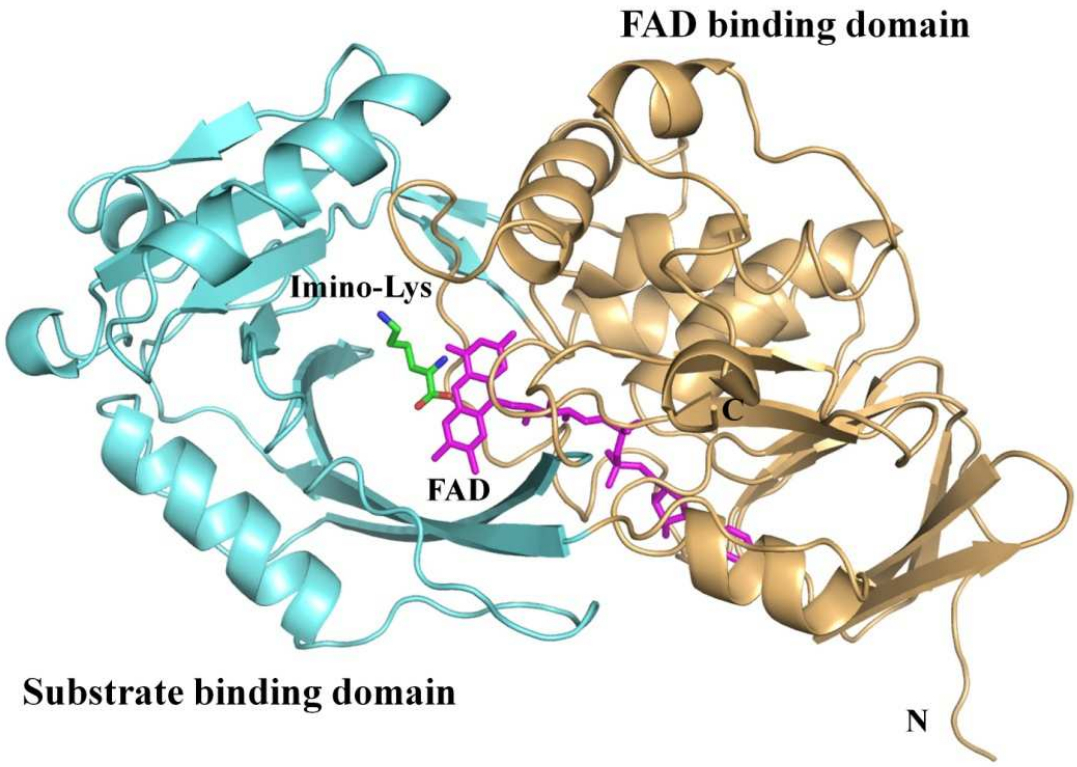


Figure 6.1 Overall structure of DADH. The DADH structure is shown in cartoon representation. The enzyme is composed of an FAD-binding domain (orange) and a substrate-binding domain (cyan). Iminolysine and cofactor FAD are shown as sticks and colored green and magenta, respectively.

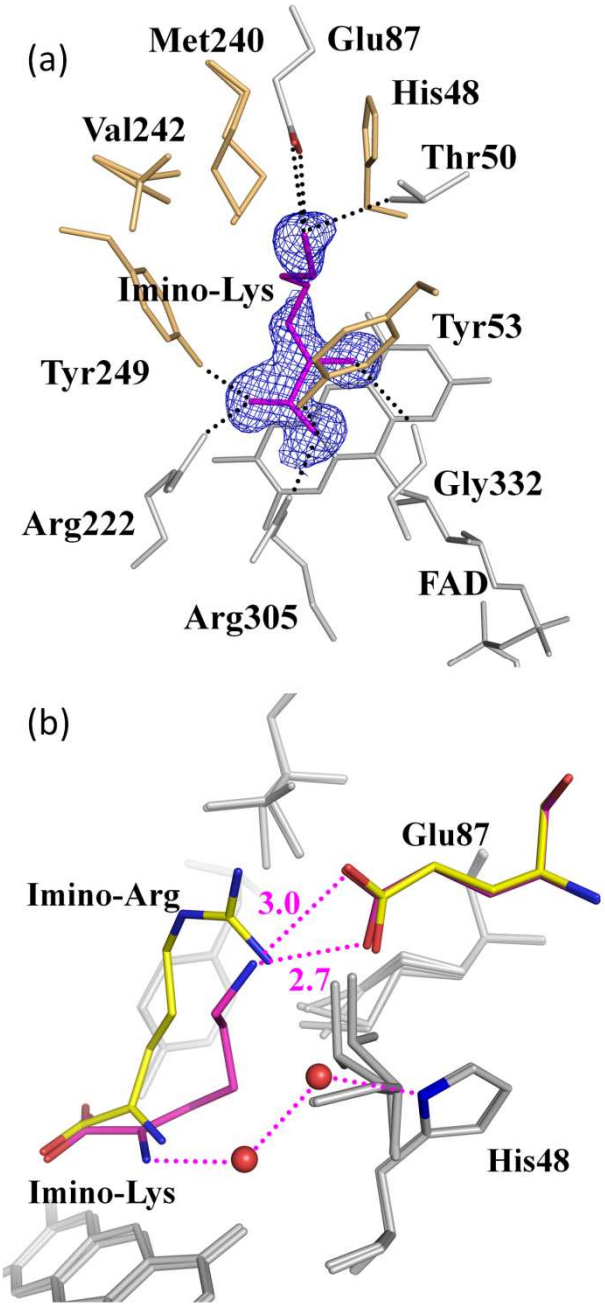


Figure 6.2 : Crystal structure of DADH in complex with iminolysine. (A) Interactions between DADH and iminolysine (magenta). DADH residues form hydrophobic interactions with ligand are colored light orange. Fo-Fc omit map of the iminolysine is indicated as blue mesh and contoured at 3σ. (B) Structural comparison of DADH/iminoarginine (yellow) and DADH/iminolysine (magenta). The active sites residues from two structures are superimposed. Hydrogen bonds are represented by dashed lines.

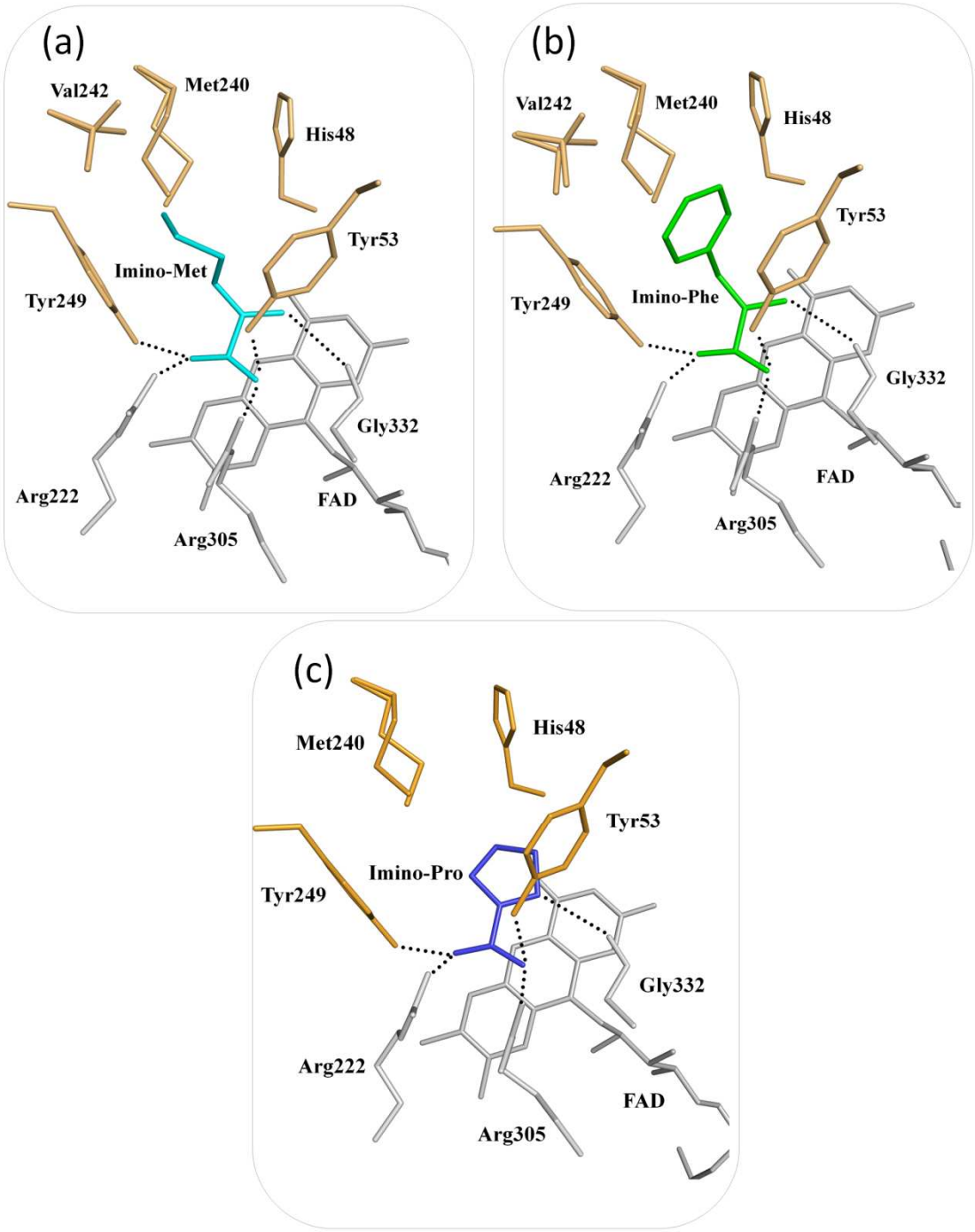


Figure 6.3 Interactions between DADH and (A) iminomethionine (cyan), (B) iminophenylalanine (green), and (C) imnoproline (blue). DADH residues form hydrophobic interactions with ligand are colored light orange. Hydrogen bonds are shown as dashed lines.

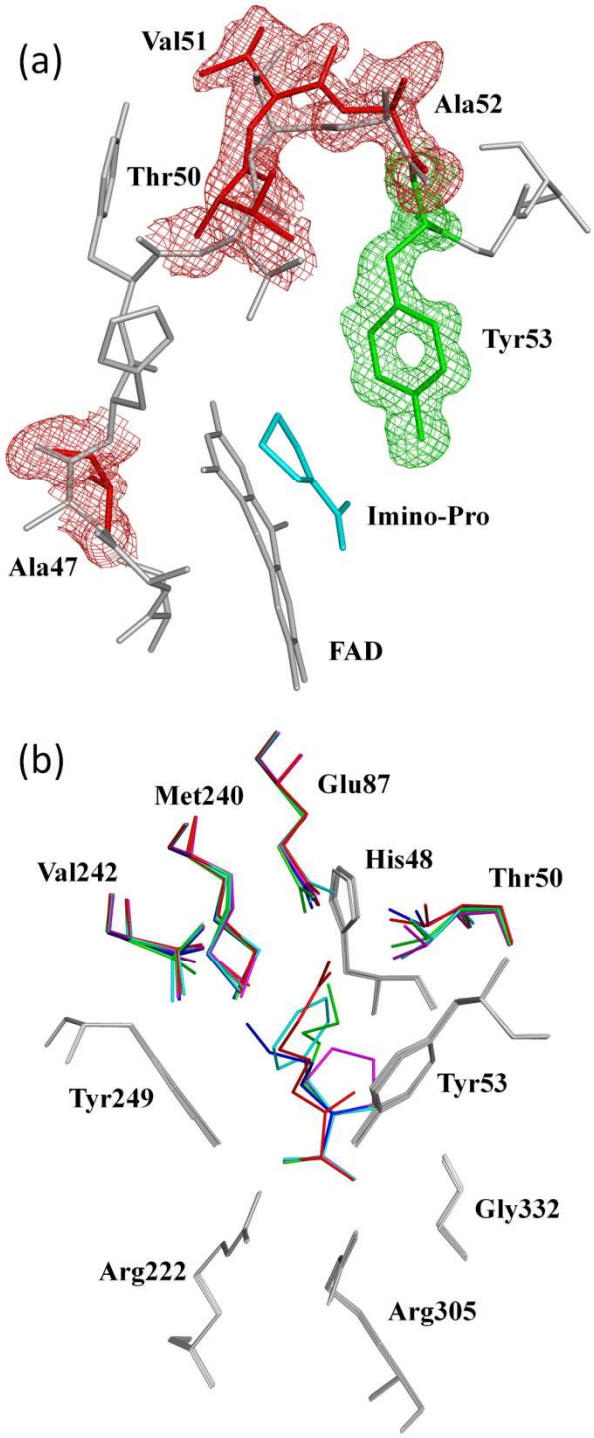


Figure 6.4 Crystal structures of DADH with imino acids. (A) Conformational disorder in the structure of DADH/iminoproline. Omit map of the alternate conformation of Ala47 and Thr50-Ala52 are shown as red mesh. Tyr53 is determined in the product-bound conformation and it's omit map is shown as green mesh. (B) Superimposition of the active sites of DADH in complexes with iminoarginine (red), iminolysine (green), iminomethionine (blue), iminophenylalanine (cyan), and iminoproline (magenta). DADH residues forming similar interactions with ligands are colored gray, while residues showing different conformations are colored corresponding to the ligand color.

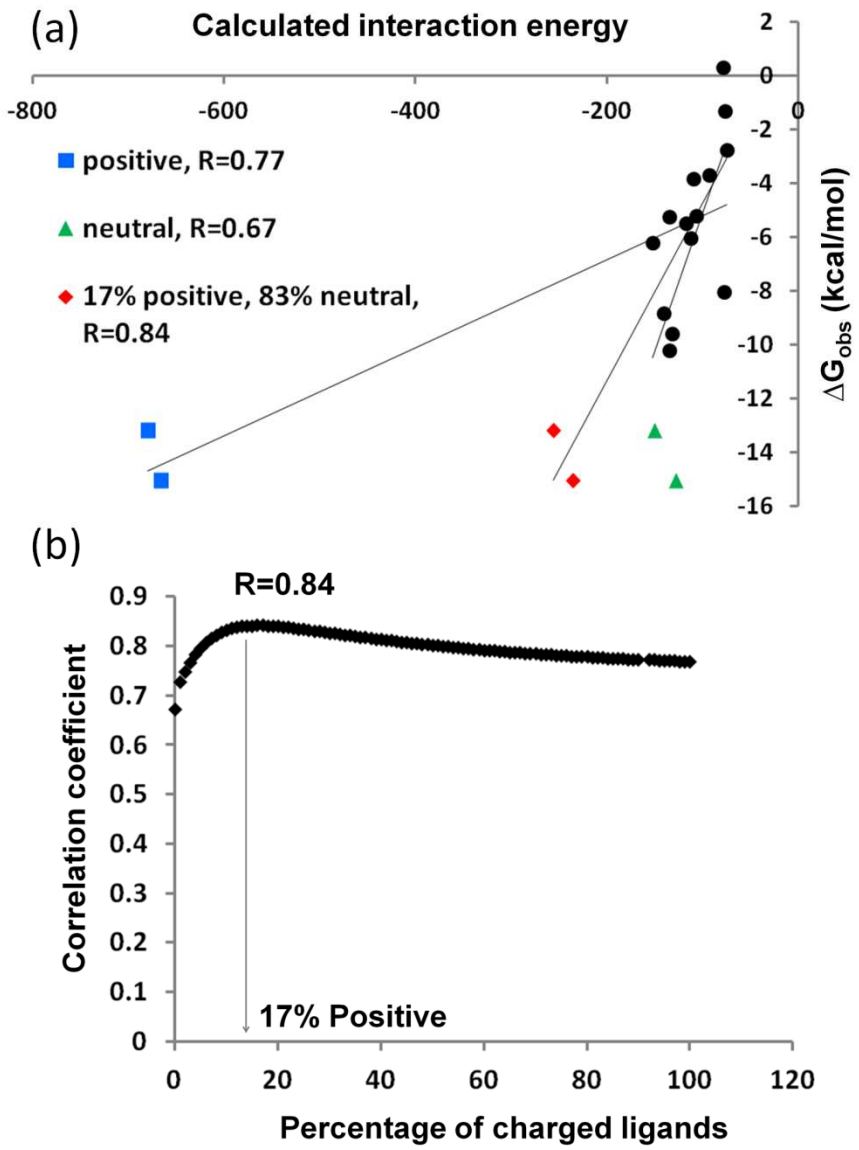


Figure 6.5 Computational modeling of DADH with different D-amino acids. (A) Correlation coefficients calculated for linear combinations ranging from neutral to fully charged side chains of basic imino acids. The maximum coefficient of 0.84 was obtained when 17% of the basic imino acids are charged. (B) The correlation between experimental binding energies and calculated interaction energies. Correlation has been calculated in three ionization states of iminoarginine and iminolysine: positive (blue), neutral (green), and the combination of 17% positive and 83% neutral states (red).

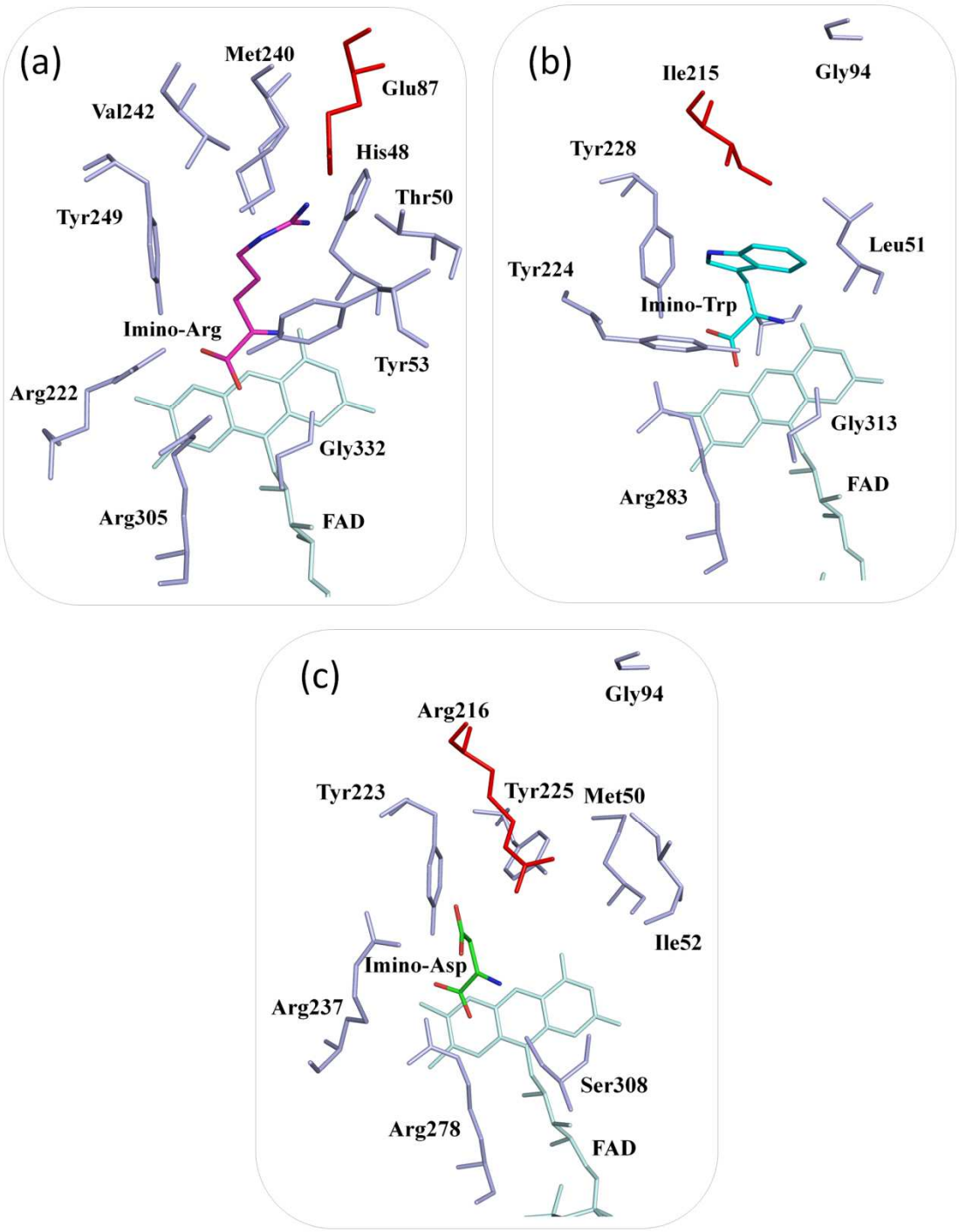


Figure 6.6 Comparison of DADH with other flavin-dependent enzymes. Comparison of the active sites of (A) DADH (yellow, iminoarginine in magenta) with (B) pDAAO (iminotryptophan in cyan), and (C) DDO (iminoaspartate in green). Glu87 in DADH, Ile215 in pDAAO and Arg216 in DDO are suggested to be important for the distinct specificities of the enzymes.

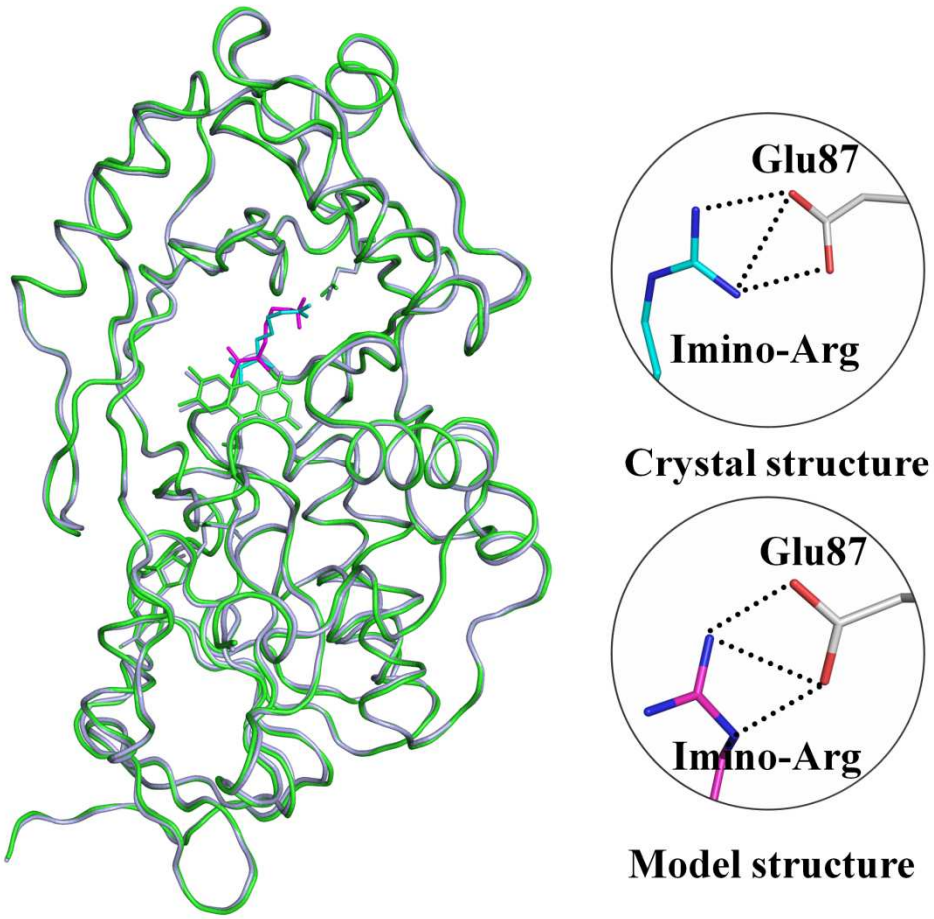


Figure 6.7 Comparison of predicted structure (green) and crystal structure (grey) of DADH in complex with iminoarginine. (S1) Inset: Iminoarginine (carbons in cyan) forms ionic interactions with Glu87 in the crystal structure, with slight differences appearing in the modeled structure (iminoarginine carbons in magenta).

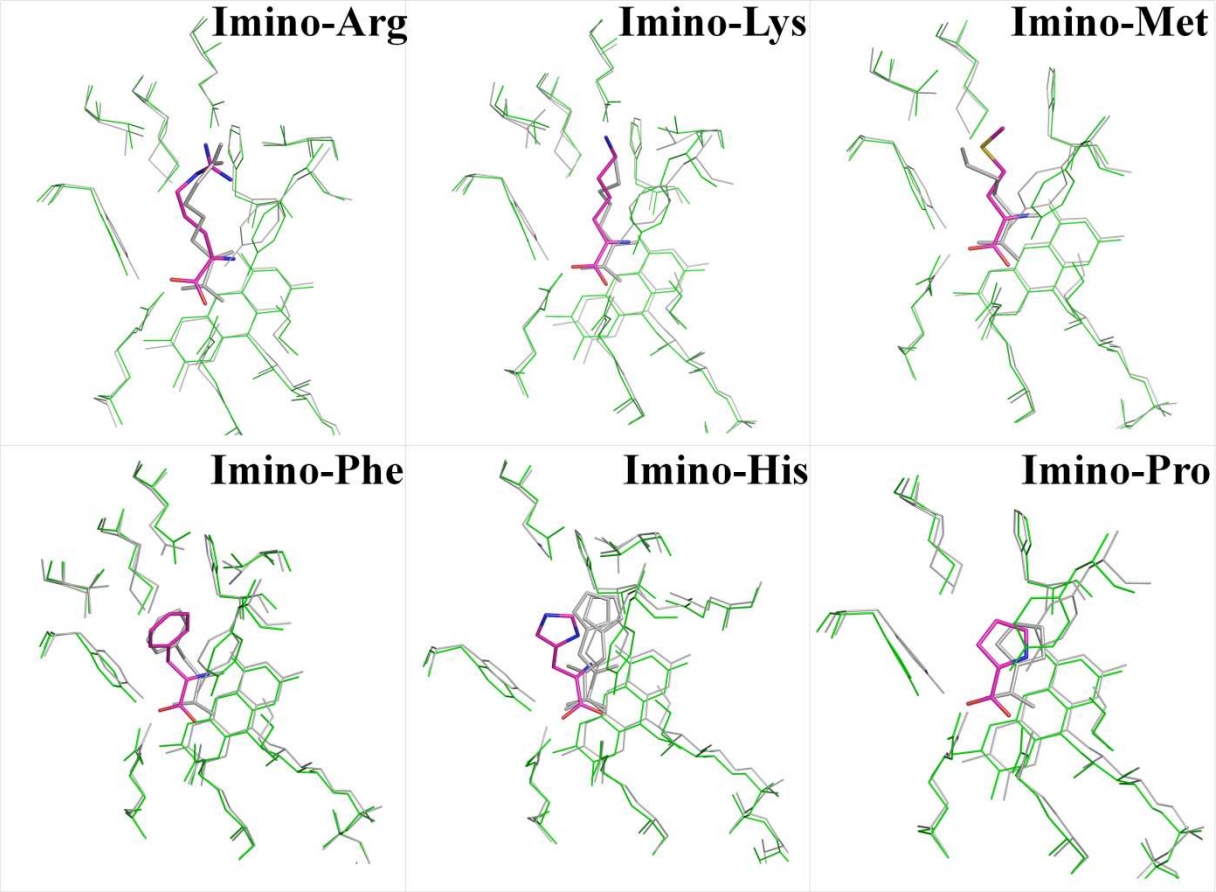


Figure 6.8 Comparison of the active sites of predicted structures (green) and crystal structures (grey) of DADH in complexes with imino acids. (S2) Imino acids from predicted structures are colored magenta, while the ligands from crystal structures are shown as grey sticks.

7 A DETERMINISTIC-STOCHASTIC CROSSOVER ALGORITHM FOR SIMULATING BIOCHEMICAL NETWORKS

Guoxing Fu, Amit Sabnis, Robert W. Harrison

7.1 Abstract

Understanding biology at the system level has gained much more interest recently due to the rapid development in genome sequencing and high-throughput measurements. Mathematical descriptions of biological systems are normally formalized using two different approaches. The deterministic method is very efficient in predicting the overall behavior of the system but ignores the inherent fluctuations and correlations at lower concentration. The stochastic method, on the other hand, captures the intrinsic randomness but is often mathematically intractable and computationally expensive. Our group has developed a deterministic-stochastic crossover algorithm for simulating biological networks. The biological models are formalized into a standard format using the eXtensible Markup Language (XML). Simulation studies have been performed on biological systems such as auto-regulatory gene network and glycolysis system. The new algorithm retains the high efficiency of deterministic method while still reflects the random fluctuations at lower concentration. The ability of revealing the stochastic property with high efficiency makes this algorithm very useful for researches and applications based on systems biology.

Categories and Subject Descriptors

Simulation and modeling, model validation and analysis, simulation output analysis

General Terms

Algorithms, standardization, verification

Keywords

Systems biology, stochastic simulation, deterministic, biochemical networks

7.2 Introduction

Systems biology is an interdisciplinary science of studying biological functions and mechanisms [90]. Rapid development in molecular biology, especially in genome sequencing and high-throughput measurements, has made the collection of comprehensive information of biological systems possible, which in turn greatly facilitated the advancement of systems biology [91]. Meanwhile, significant development in computational software and hardware allowed the creation and investigation of complex and still realistic biological systems. Systems biology studies focus on the understanding of whole system's structure and dynamics, while the traditional molecular biology concentrates more on the properties and functions of genes and proteins. Although many issues remain unsolved, systems biology studies are able to provide in-depth insights and reasonable predictions for biological networks such as the cell cycle and metabolic analysis.

Molecular biology has uncovered numerous biological facts like protein properties and genome sequences in the past a few decades. However, with the dramatically growth of information, it alone is far from sufficient for investigating the complex biological systems. Therefore, a system-level understanding of biology is greatly in need and a combination of experimental and computational approaches is expected to achieve this goal. Computer simulations of biological systems have been intensively studied and can now replicate important biological behavior. Successful computational studies have been performed on simple systems such as the lambda phage 'switch' [94-95] and the T-7 phage [96-97]. However, complex systems such as eukaryote organisms, especially mammalian cells, are much more challenging to be studied by computer simulation [98]. Many efforts have been dedicated into studies of cellar processes like gene expression, cellular signaling and metabolism pathways.

Two major techniques used in system simulation are deterministic methods and stochastic methods. The deterministic methods work well for systems with large molecular populations and can reasonably represent the overall dynamics of the systems. However, it fails to capture the inherent fluctuations and correlations of the species at lower concentration. The stochastic methods are able to capture the fluctuations at lower concentration, but are very computationally inefficient. Therefore, novel algorithms are needed to improve the accuracy of deterministic methods and efficiency of stochastic methods. Many studies have been dedicated to improving the computational efficiency of stochastic method, while some groups have developed hybrid algorithms to facilitate the simulation [101-103]. Our group has developed a novel deterministic-stochastic crossover method in order to improve the computational accuracy and efficiency of simulation for biochemical systems [104]. The crossover method retains the efficiency of deterministic methods and still reflects the fluctuations generally only captured by stochastic methods.

In this study, we have extended the functionality and adaptability of the deterministic-stochastic crossover method. First of all, we have developed the model representation for biological systems in XML, which makes the input and output of data easier. This feature also allows transformation of our models between different software platforms feasible. Moreover, we have incorporated compartment and event into our model design, making the models more similar to the real biological or experimental conditions. Most importantly, the new algorithm is able to simulate Michaelis-Menten type reactions directly, which have to be broken into elementary reactions for simulation in stochastic methods and the previous version of deterministic-stochastic crossover method.

7.3 The Corssover Algorithm

7.3.1 System Design and Model Representation

Biological systems can be represented in various formats to fulfill different purposes. For example, graphical representations are useful both for visualization and analysis, while matrix representations are more convenient for mathematical and computational analysis. However, different representations often make it difficult to transfer models between different software platforms. A commonly accepted

software environment would allow individual research groups to easily exchange their models. Several undergoing project including Systems Biology Markup Language (SBML http://www.sbml.org/), CellML (http://www.cellml.org/) and the Systems Biology Workbench are examples of efforts that aim to form a standard and open software platform for modeling and analysis [105-106]. Such efforts have significantly facilitated the development of new generation of databases concerned with biological pathways by employing machine-executable models rather than human-readable forms only.

The SBML developed by Hucka et al. is a powerful presentation of biological networks [107]. It consists of lists of functions, units, compartments, species, parameters, rules, reactions, and events. While it is a good format for computers to parse and generate, its verbosity makes it quite inconvenient for humans to read and write [110]. In order to standardize and improve the readability of our models, we have developed a similar but much simpler representation using eXtensible Markup Language. We will concentrate on compartments, species, parameters and reactions, as these are fairly sufficient for describing most simple discrete stochastic models. The system design is demonstrated in **Figure 7.1**. The model is composed of one global system, with the parameters such as initial time, time step, total time, and report interval defined at the system level.

Most of the computational biology studies are focused on single simple system. However, we may need to work with multiple systems or single system containing multiple compartments depending on the nature of the study. Therefore, we have introduced the idea of multiple compartments into this crossover metgod. Each compartment will contain its own species and reactions. The name, initial number and chemostat status will be assigned to each species, while every reaction will have its own kinetic parameters, lists of input species and output species. For biological systems that are composed of more than one compartment, there might be transportable species between different compartments, which means behavior of one compartment will affect that of others. This is more close to the reality since most of the biological systems are complex networks and one certain system is often affected by its environment. In the meantime, an event feature is also developed to represent specific activities during the simu-

lation, which are very common in biological processes and designed experiments at certain time points or when some conditions are met.

7.3.2 Simulation of Michaelis-Menten type Reactions

The fundamental of this deterministic-stochastic crossover method has been elucidated in the previous report [104]. Deterministic functions are used to compute a continuous updating value (a real number) for each reacting species. A Bernoulli trial was then conducted to obtain an integer number for each species using the fractional part of the real number as the probability of success. This method has shown to be successful for simple systems such as auto-regulatory gene network and dimerization pathway [104]. However, the previous crossover method was only able to handle elementary reactions similar to stochastic methods. A standard Michaelis-Menten reaction, shown as following, will need to be broken into three elementary reactions: binding, dissociation and conversion, which will increase the complexity and computational time of the simulation. Here E stands for enzyme; S is substrate; ES is the complex of enzyme and substrate; P indicates product; kcat is the turnover number.

$$E + S \rightleftharpoons ES \overset{k_{cat}}{\rightarrow} E + P$$

One of the major progresses of the current crossover method is its capability of directly modeling Michaelis-Menten type reactions. The two types of reactions will be marked differently in the system model and calculated by different functions. In the crossover algorithm, the rate of product formation of this type of reaction is given by the following equation, in which [E] is the enzyme concentration and [S] is the substrate concentration. K_m is the substrate concentration at which the reaction rate is at half-maximum, and is an inverse measure of the substrate's affinity for the enzyme.

$$\frac{d(P)}{dt} = k_{cat}[E]\frac{[S]}{K_m + [S]}$$

7.4 Results and Discussion

7.4.1 System Validation by Auto-Regulatory Gene Network

Since the crossover algorithm is newly implemented in Java with a few features like XML mode representation added, we first validated it by testing the auto-regulatory gene network, which has been described in the previous work [104]. The auto-regulatory gene network is typically a low concentration biochemical network whose detailed dynamics are usually captured only by stochastic algorithms [7]. According to the concepts of our system design, an example of system representation in XML for auto-regulatory gene network is shown as following:

```
<system time="0" timestep="0.001" total="100000" reportInterval ="500">
<partition name="cell" volume="1e-15">

<reactant name="gene" number="10" chemostat="false"/>
<reactant name="rna" concentration="0" chemostat="false"/>
…

<reaction id="r1" constant="0.01">
<input name="gene" />
<output name="gene" />
<output name="rna" />
</reaction>
…

</partition>
</system>
```

The system is initialed with 10 molecules of "Gene" and none of the other species. Reaction volume was set to 1E-15 liters. The Δt for each time step is set to 0.001s. The results from crossover method are compared with the results from stochastic method and deterministic method (**Figure 7.2A-E**). The stochastic method was simulated by using the software Dizzy [215], while the deterministic results were obtained by using XPPAUT [http://www.math.pitt.edu/~bard/xpp/xpp.html].

The simulation solution from the crossover method is in reasonable agreement with the results from stochastic method and deterministic method. Similarities between the results from crossover method and the stochastic method indicated that the crossover method is able to reflect those fluctuations featured

in stochastic simulation but missed in deterministic simulation. The stoichiometric consistency of the crossover method is demonstrated by analyzing the species Gene and Gene.Dimer. Gene cannot be produced or degraded in this system, thus the sum of Gene and Gene.Dimer should remain constant throughout the whole process of simulation, which is 10 molecules as shown in **Figure 7.2F**.

7.4.2 Testing of Compartments by Michaelis-Menten Model

The Michaelis-Menten model has been widely used in biochemistry for studying the kinetics of many enzymes. The substrate (S) can bind to enzyme (E) forming complex (ES), which then dissociates to either enzyme and substrate or enzyme and product (P) (**Table 7.1**). Here we employed this model to test the multiple-compartment feature of the crossover algorithm, due to its simplicity and well-understood dynamics. An artificial two-compartment model was defined for this system, with S and E in one compartment, ES and P in another (**Figure 7.3A**). Complex ES generated in compartment I will be transported to compartment II, while S and E produced by ES dissociation in compartment II are output into compartment I. The transportation reactions will be defined in each compartment with the transferring rate being experimentally determined or artificially chosen as in this test.

The simulation results from deterministic method (XPPAUT), stochastic method (Dizzy), and crossover method are compared in **Figure 7.3**. It is clear that two-compartment model of crossover method has similar dynamic behavior when compared with results from the other two methods. The conservation law holds true for this system that the sum of [E] and [ES] is constant. Taken together, this multiple-compartment feature works properly and can be applied to more complex biological systems.

7.4.3 Glycolysis System

One of the major difficulties of simulating complex systems is searching for the right combination of parameters. Traditional optimization methods often simulate simple systems by integrating the kinetic equations with a set of parameters and initial conditions. However, this approach does not always work for complex systems if many kinetic parameters are unknown. Here we have chosen the glycolysis system in Saccharomyces cerevisiae as the studying subject of this crossover algorithm [216]. Glycolysis

is the metabolic pathway that converts glucose into pyruvate and has been thoroughly studied in classical biochemistry, thus making it a good subject for computational studies.

There are 38 species and 36 reactions involved in this system (Table 7.2). The system is composed of two compartments, extracellular and intracellular. Extracellular compartment includes 11 species and 12 reactions, while 27 species and 24 reactions are defined in the intracellular compartment. The chemstat statuses of extracellular glucose flow and extracellular cyanide (CN^-) flow are set to true in order to provide constant supply for extracellular glucose and cyanide. Species like glucose, glycerol, acetaldehyde and ethanol are transportable between the two compartments. Unlike the auto-regulatory gene network, which is composed of elementary reactions only, this glycolysis contains both elementary reactions and Michaelis-Menten type reactions. Reactions such as transportation or degradation of some species are simplified into elementary reactions, with the reaction rate mainly determined by the reaction constant and the concentration of the reactants. Reactions such as conversion of glucose into glucose 6-phosphate catalyzed by hexokinase obey the Michaelis-Menten kinetics, with the reaction rate determined by reaction rate, reactant concentration, enzyme concentration and affinities of reactant to the enzyme.

The system is initialed following the conditions from the model described in previous study [216]. The time step is set to 0.001s with a total time of 100 seconds simulated. The results from the crossover method for some of the species are shown in **Figure 7.4**. The concentration of ATP and ADP decreased while that of AMP increased. After about 10 seconds, they reached an equilibrium state and did not show significant changes. Again, since the system is closed to ATP, the sum of ATP, ADP and AMP is constant throughout the simulation. Meanwhile, the crossover method is able to capture the random fluctuations at lower concentration for species such as 1, 3-bisphosphoglycerate and phosphoenol pyruvate. A fluctuation is also observed for pyruvate even its concentration is higher than other species. We tried to compare the simulation of this glycolysis system by crossover method, stochastic method and deterministic method. However, there are some limitations with XPPAUT and Dizzy when applied with such kind of complex system. XPPAUT and Dizzy can only take elementary reactions, thus cannot handle Michaelis-Menten type reactions driectly as discussed previously. They cannot simulate models with mul-

tiple compartments. Therefore, this two-compartment model has to be converted to one-compartment model to be tested by other programs. Moreover, constant concentration for certain species is not supported in XPPAUT and Dizzy. Due to these limitations, we were not able to obtain comparable simulation results from these two programs for this glycolysis system.

7.4.4 Testing Events in Glycolysis System

The event feature is then tested in this glycolysis system. For each event, we need to define when the event will happen. What is the target of this event? For example, the number, concentration or chemostat status of the reactant can be altered. Here we defined two events in the intracellular compartment of glycolysis system in order to test whether this event function is working properly. At 20 second, 3000 molecules of ATP were added into the system. At 50 second, the number of NADH was adjusted to 1000, no matter what the actual value was at that moment. The results from the system with and without events are compared in Figure 7.5. At 20 second, the number of ATP has been increased by 3000. Since ATP, ADP and AMP are convertible to each other and the system is closed to them, the numbers of AMP and ADP are also increased accordingly. Similar situation has been observed for NAD^+ and NADH. The number of NAD^+ was increased when the number of NADH was adjusted to 1000 at 50 second. These changes can be directly observed since these species are the subjects of the testing events (**Figure 5A-B**). However, indirect changes were also seen for other species in the system (**Figure 5C-F**). For example, a decrease in number of dihydroxyacetone phosphate was observed at about 50s. Most likely it is mainly affected by the adding of NADH at 50s since it reacts with NADH to form glycerol and NAD^+. The number of 1, 3-bisphosphoglycerate has been increased during 20-60 second. This could possibly be explained by the increased conversion of glyceraldehydes 3-phosphate and phosphoenol pyruvate to 1, 3-bisphosphoglycerate due to the increase of ATP and NAD^+. Almost in the same time period, the number of phosphoenol pyruvate has dropped, probably because its conversion to 1, 3-bisphosphoglycerate and pyruvate. Further, the overall simulation result for pyruvate is different from the data without the event. This could be the result of combined impact from multiple reactions. Overall, the test of this event system

was successful. It allows this crossover algorithm to simulate some of the specific activities in biological systems.

7.4.5 Run-time Analysis

In order to test the efficiency of this crossover algorithm, we have performed different types of run-time analysis. The tests were performed on a desktop computer with an Intel I5-750 (2.66G) processor and 8G DDR3 memory. Each test is performed 10 times and the average value is calculated. Firstly, the performance of crossover algorithm and stochastic methods (Dizzy) on auto-regulatory gene network has been compared over different simulation time (**Table 7.3**). At a first glance, the crossover algorithm is slower than Dizzy when the total simulation time is short, such as 100s and 1000s. However, as we increased the total simulation time, the advantage of crossover algorithm became more significant and eventually it surpassed the stochastic methods. The crossover algorithm is 10.2% faster than dizzy at 10,000s and the difference increased to 18.5% at 100,000s, suggesting that the crossover algorithm is more efficient than the stochastic methods in Dizzy for long time simulation.

Next we compared the crossover algorithm with Dizzy over different systems (Table 7.4). The Michaelis-Menten model contains only 3 reactions, thus is considered as a small system here. The auto-regulatory gene network is composed of 8 reactions and has a relatively medium size. The glycolysis system has 36 reactions and is a more complex system. The first two systems contain only elementary reactions while the glycolysis system contains both elementary reactions and Michaelis-Menten type reactions. In order to eliminate the affect of data loading on the performance, we have chosen a simulation time of 10,000s. The crossover algorithm is 9.1% faster than Dizzy for Michaelis-Menten model, and 10.2% faster for the auto-regulatory gene network. It is able to finish a 10,000s simulation of the glycolysis system in 32.37s. A simulation of 100s of the glycolysis system by this crossover algorithm takes only 0.361s. On the other hand, Dizzy is not able to handle such a complex system as discussed in the previous sessions. Therefore, the crossover algorithm is slower than Dizzy for shorter simulation time, but is faster

for longer simulation time, when small and simple systems are studied. Moreover, the crossover algorithm can simulate big and more complex systems which cannot be handled by stochastic methods.

7.5 Conclusions

In this study we have extended the functions of this deterministic-stochastic crossover method, making it more competitive for the simulation of biochemical networks in both accuracy and efficiency. A lot of efforts have been made to overcome the drawbacks of deterministic and stochastic methods. Some hybrid algorithms employing both deterministic and stochastic techniques have been developed to increase the computational efficiency at a small cost of accuracy [102-103, 217]. However, one of the major disadvantages of these hybrid systems is that they require partitioning the system into subsets of fast and slow reactions. The fast reactions will be treated as a continuous process and simulated by deterministic method, while the slow reactions are considered as discrete events and simulated by stochastic algorithm. The major advantage of this crossover method over these hybrid algorithms is it does not require partition of the modeled system. It is not actually a combination of deterministic and stochastic methods, but a deterministic method with the randomness introduced by Bernoulli trial. At higher concentration, the fluctuation induced by Bernoulli trial is neglectable; however, the effect of Bernoulli trial will become significant at lower concentration, thus allowing the stochastic effects to be captured. Such kind of “smooth” transition from deterministic (high concentration) to stochastic (low concentration) has not been achieved by other methods based on our knowledge. In fact, this is more similar to the condition that a fast reaction behaves continues at high concentration may become a slow reaction at lower concentration due to some physiological impacts, which is very common in biological systems and under some designated experimental conditions.

Another notable feature of this crossover algorithm is its capability of simulating more complex Michaelis-Menten type reactions, instead of being restricted to simple elementary reactions. Most of other simulation programs are not able to handle complex reactions without breaking them down into elementary reactions, which will introduce more difficulties to the system representation and add more computa-

tional cost. This crossover algorithm made the simulation of Michaelis-Menten type reactions feasible. Further, the added features of compartment and event made the simulation of biological systems become more realistic. One of the benefits of using multiple- compartment system is the potential ability of applying parallel computing in this system. Simulation of different compartments simultaneous will greatly enhance the efficiency of this crossover method and will be a future direction of our study. Run-time analysis has indicated that this crossover algorithm is more efficient than stochastic methods for long simulation time and more complex systems. Overall, development of this crossover algorithm would be very useful for researches and applications based on systems biology.

7.6 Acknowledgment

G. F. was supported by the Georgia State University Molecular Basis of Disease Fellowship. The authors are thankful to the Georgia Research Alliance and the Georgia Cancer Coalition for supporting this work (Robert W. Harrison is a Georgia Cancer Coalition distinguished cancer scholar).

7.7 Appendix: Biochemical Abbreviations

ADH: alcohol dehydrogenase

ALD: aldolase

GAPDH: glyceraldehyde 3-phosphate dehydrogenase

PDC: pyruvate decarboxylase

HK: hexokinase

PFK: phosphofructokinase-1

PGI: phosphoglucoisomerase

TIM: triosephosphate isomerase

ACA: intracellular acetaldehyde

ACA_{ex}: extracellular acetaldehyde

ADP: adenosine 5’-diphosphate

AMP: adenosine 5’-monophosphate

ATP: adenosine 5’-triphosphate

DHAP: dihydroxyacetone phosphate

BPG: 1, 3-bisphosphoglycerate

CN^-_{ex}: extracellular CN^-

CN^-_{exf}: extracellular CN^- flow

EtOH: intracellular ethanol

$EtOH_{ex}$: extracellular ethanol

F6P: fructose 6-phosphate

FBP: fructose 1, 6-bisphosphate

G6P: glucose 6-phosphate

GAP: glyceraldehyde 3-phosphate

Glc: intracellular glucose

Glc_{ex}: extracellular glucose

Glc_{exf}: extracellular glucose flow

Glyc: intracellular glycerol

$Glyc_{ex}$: extracellular glycerol

NAD^+: nicotinamide adenine dinucleotide (oxidized form)

NADH: nicotinamide adenine dinucleotide (reduced form)

PEP: phosphoenol pyruvate

Pyr: pyruvate

Table 7.1 Details of the Michaelis-Menten model

Description	**Reaction**	**Constant**
binding	E + S → ES	1.0E-2
dissociation	ES → E + S	1.0E-4
conversion	ES → E + P	1.0E-1

Table 7.2 Details of the glycolysys model (See appendix for abbreviations)

Reaction ID	Reaction	Reaction constant	Enzyme	Parameter
r1	$Glc_{exf} \rightarrow Glc_{ex}$	8.0E-7		
r2	$Glc_{ex} \rightarrow Glc_{exf}$	8.0E-7		
r3	$Glc_{ex} \rightarrow Glc$	1.6916E-2		
r4	$Glc \rightarrow Glc_{ex}$	1.6916E-2		
r5	$Glc + ATP \rightarrow ADP + G6P$	8.6257E-1	HK	$K_{Glc} = 0.01$, $K_{ATP} = 0.1$
r6	$G6P \rightarrow F6P$	8.2673E-3	PGI	$K_{G6P} = 0.8$, $K_{F6P} = 0.15$
r7	$F6P \rightarrow G6P$	8.2673E-3	PGI	
r8	$F6P + ATP \rightarrow ADP + FBP$	7.5721E-1	PFK	$K_{F6P} = 0.021$, $K_{ATP} = 0.15$
r9	$FBP \rightarrow GAP + DHAP$	3.6797E-2	ALD	$K_{FBP} = 0.3$
r10	$GAP + DHAP \rightarrow FBP$	1.8399E-1	ALD	$K_{GAP} = 4.0$, $K_{DHAP} = 2.0$
r11	$DHAP \rightarrow GAP$	1.9394E-3	TIM	$K_{DHAP} = 1.23$
r12	$GAP \rightarrow DHAP$	1.9394E-3	TIM	$K_{GAP} = 1.27$
r13	$GAP + NAD^{+} \rightarrow BPG + NADH$	1.3897E-3	GAPDH	$K_{GAP} = 0.6$, $K_{NAD^{+}} = 0.1$
r14	$BPG + NADH \rightarrow GAP + NAD^{+}$	1.3897E-3	GAPDH	$K_{BPG} = 0.01$, $K_{NADH} = 0.06$
r15	$BPG + ADP \rightarrow PEP + ATP$	7.3977		
r16	$PEP + ATP \rightarrow BPG + ADP$	2.5477		
r17	$PEP + ADP \rightarrow Pyr + ATP$	5.7182E-5		
r18	$Pyr \rightarrow ACA$	8.8554E-1	PDC	$K_{Pyr} = 0.3$
r19	$ACA + NADH \rightarrow EtOH + NAD^{+}$	1.4967E-3	ADH	$K_{ACA} = 0.71$, $K_{NADH} = 0.1$
r20	$EtOH \rightarrow EtOH_{ex}$	2.7866E-1		
r21	$EtOH_{ex} \rightarrow EtOH$	2.7866E-1		
r22	$EtOH_{ex} \rightarrow empty$	8.0E-7		
r23	$DHAP + NADH \rightarrow Glyc + NAD^{+}$	1.3579E-5		
r24	$Glyc \rightarrow Glyc_{ex}$	3.1666E-2		
r25	$Glyc_{ex} \rightarrow Glyc$	3.1666E-2		
r26	$Glyc_{ex} \rightarrow empty$	8.0E-4		
r27	$ACA \rightarrow ACA_{ex}$	4.1166E-1		
r28	$ACA_{ex} \rightarrow ACA$	4.1166E-1		
r29	$ACA_{ex} \rightarrow empty$	8.0E-1		
r30	$ACA_{ex} + CN^{-}_{ex} \rightarrow empty$	4.7304E-2		
r31	$CN^{-}_{ex} \rightarrow CN^{-}_{exf}$	8.0E-4		
r32	$CN^{-}_{exf} \rightarrow CN^{-}_{ex}$	8.0E-4		
r33	$G6P + ATP \rightarrow ADP$	3.7655E-5		
r34	$ATP \rightarrow ADP$	5.346E-1		
r35	$ATP + AMP \rightarrow ADP + ADP$	7.215E-3		
r36	$ADP + ADP \rightarrow ATP + AMP$	2.2222E-3		

Table 7.3 Run-time analysis of auto-regulatory gene network over different simulation time

Time (s)	Crossover (s)	Dizzy (s)	Difference
100	0.077	0.042	-83.3%
1000	0.670	0.481	-39.3%
10,000	6.670	7.434	10.2%
100,000	67.514	82.844	18.5%

Table 7.4 Run-time analysis of different systems

Systems (# of reactions)	**Crossover (s)**	**Dizzy (s)**	**Difference**
Michaelis-Menten (3)	0.019	0.021	9.1%
auto-regulatory gene network (8)	6.670	7.434	10.2%
glycolysis (36)	32.37	N/A	N/A

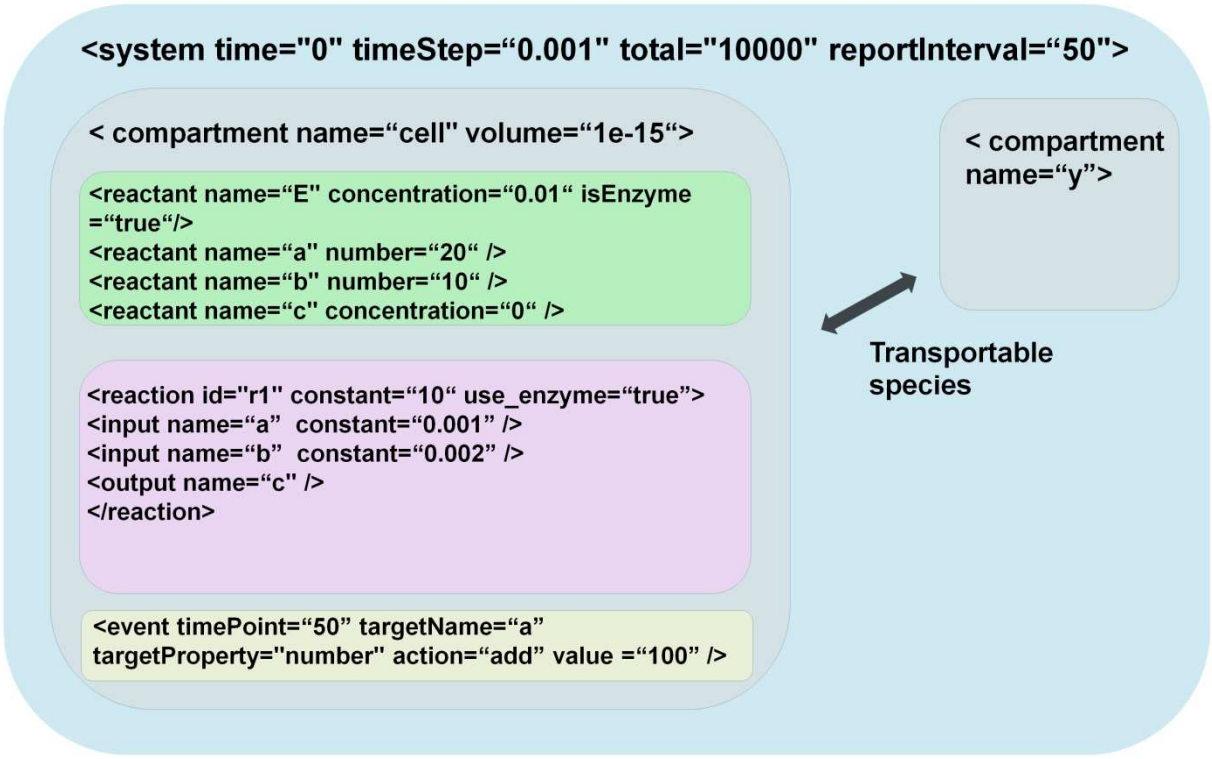


Figure 7.1 System design of the crossover algorithm

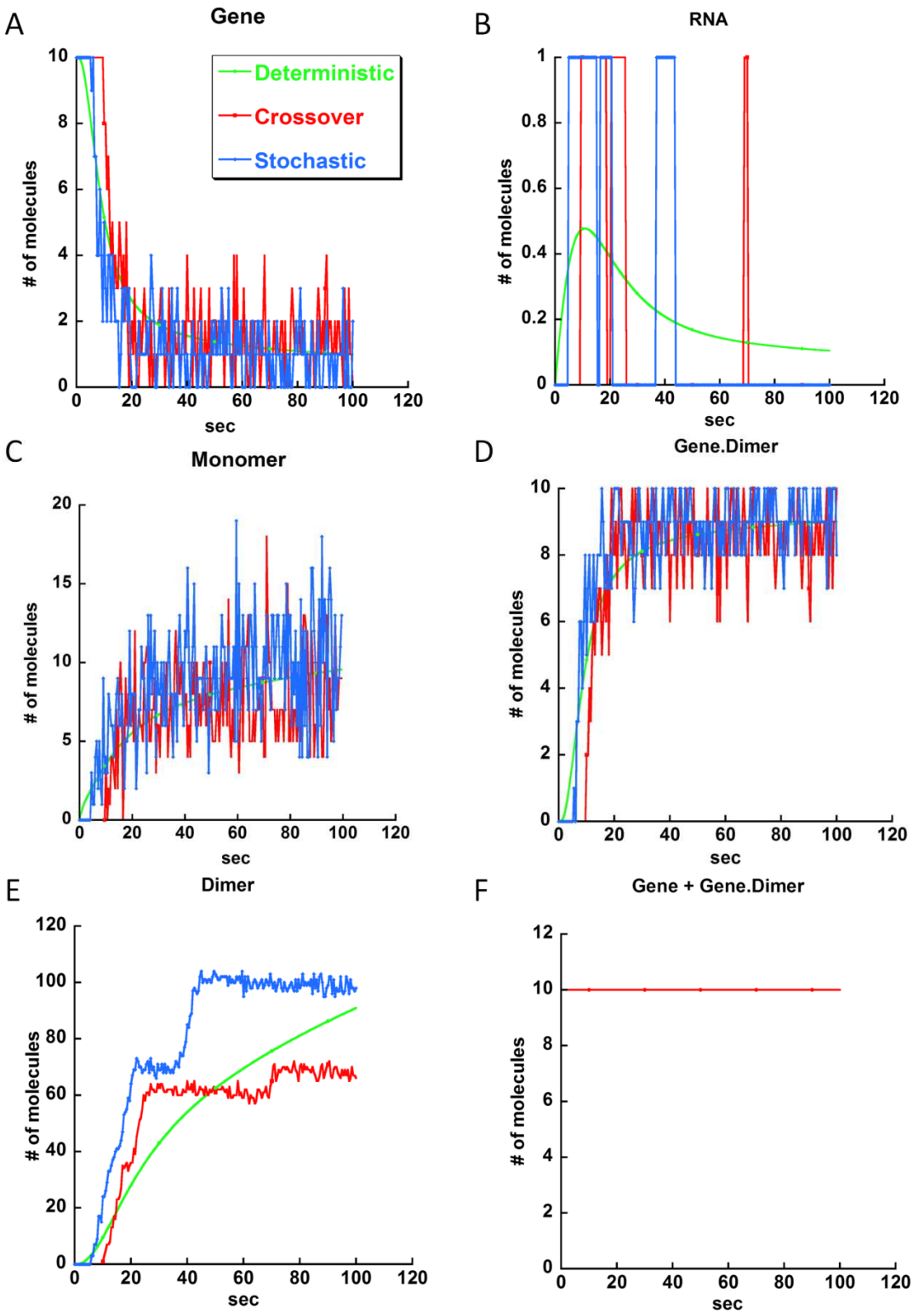


Figure 7.2 Comparison of simulation results for auto-regulatory gene network

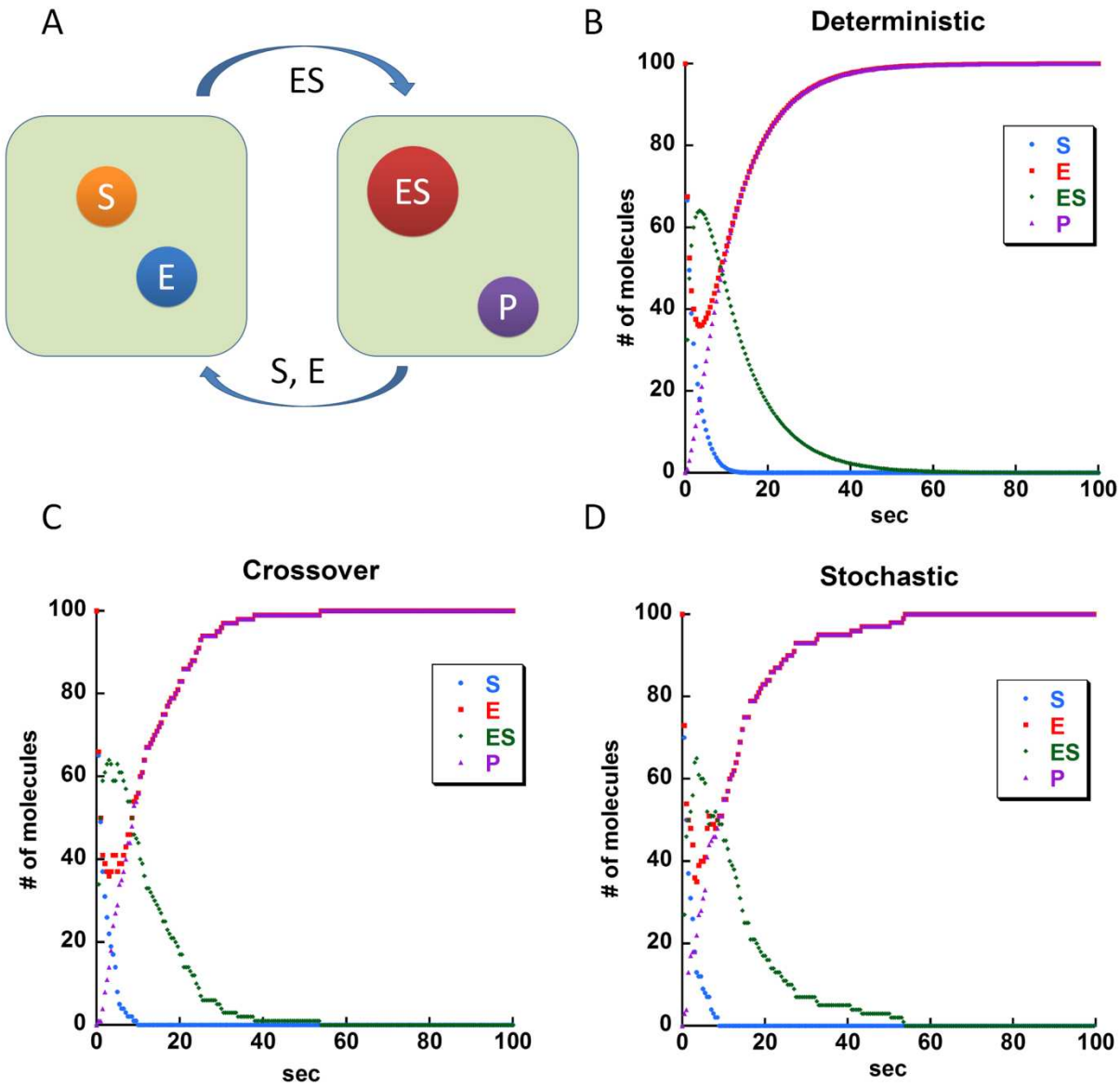


Figure 7.3 Testing of multiple-compartment system

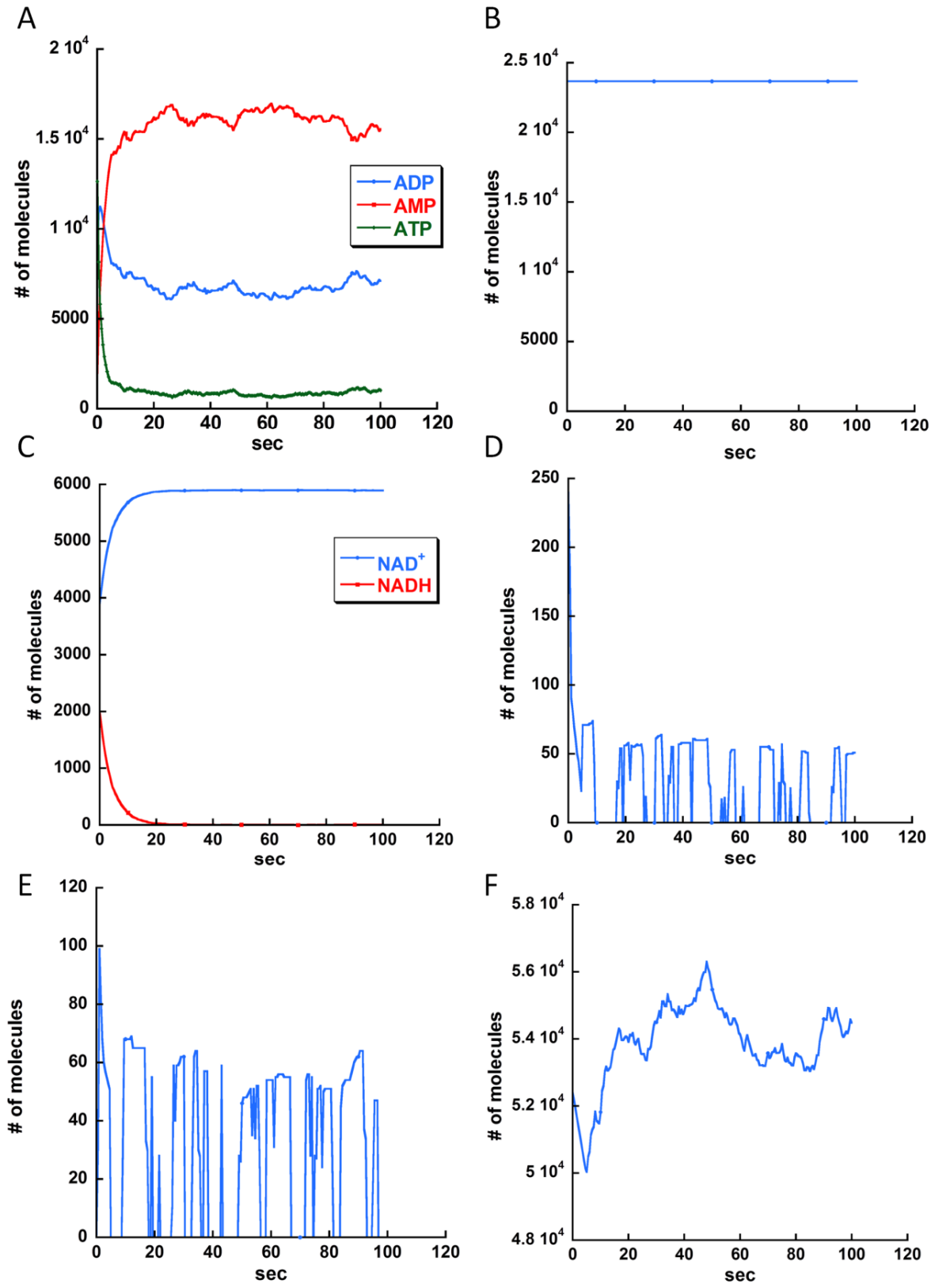


Figure 7.4 Simulation results of glycolysis system. (A) ATP, ADP, AMP; (B) Sum of ATP, ADP and AMP; (C) NAD^+, NADH; (D) Phosphoenol pyruvate; (E) 1, 3-Bisphosglycerate; (F) Pyruvate

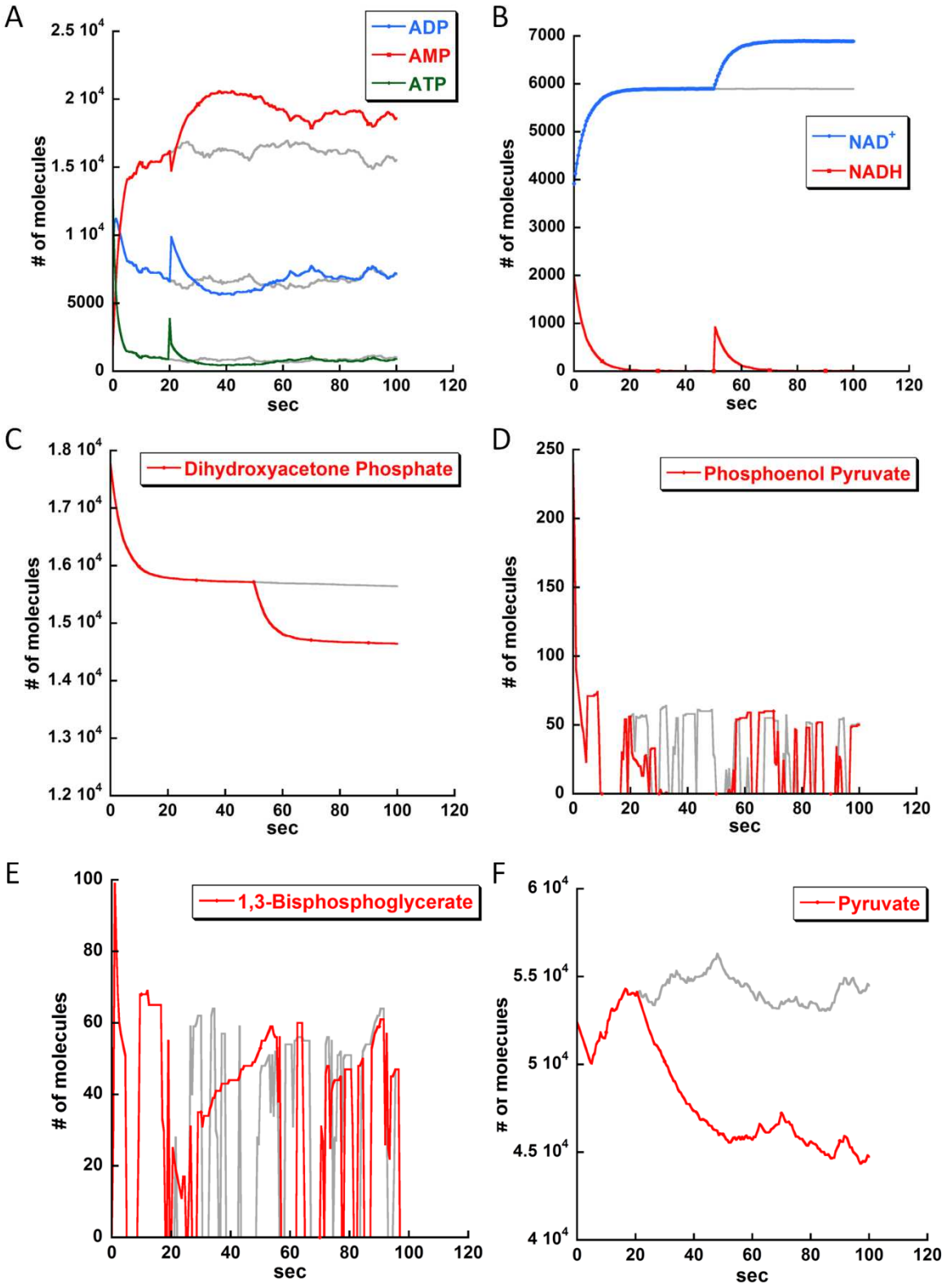


Figure 7.5 Testing events in glycolysis system. Results without events are shown in gray.

8 OVERALL SUMMARY

In this study, we have performed crystallographic study and structural modeling to investigate the structural basis of P5 recognition of executioner caspases. The S5 subsite of caspase-3 was identified by our group and further studied in caspase-6 and -7 [34, 37]. Caspase-3 and -6 recognize P5 in pentapeptides as shown by enzyme activity data and interactions observed in the crystal structure of caspase-3/LDESD and in a model for caspase-6. In caspase-3 the P5 main-chain was anchored by interactions with Ser209 in loop-3 and the P5 Leu side-chain interacted with Phe250 and Phe252 in loop-4 consistent with 50% increased hydrolysis of LDEVD relative to DEVD [37]. Binding of P5 residue into caspase-3 active site leads to conformational changes in the loop-4 via an induced fit mechanism [34, 37]. Caspase-6 formed similar interactions and showed a preference for polar P5 in QDEVD likely due to interactions with polar Lys265 and hydrophobic Phe263 in loop-4. Caspase-7 exhibited no preference for P5 residue in agreement with the absence of P5 interactions in the caspase-7/LDESD crystal structure. Our structural and kinetic analysis of the preference of executioner caspases for the P5 position of substrates will help identify the specific protein substrates and apoptotic pathways activated by each caspase. These findings are valuable in understanding the presence of redundant executioner caspases. Importantly, differences were identified in P5 recognition of caspase-3 and -7, which otherwise recognize very similar substrates, and these differences were correlated with the residue present in the newly defined loop-3 anchoring position and the caspase conformation with pentapeptides rather than tetrapeptides. Such differences in substrate recognition can be exploited in the rational design of selective inhibitors, which has proved a challenge, especially for caspase-3 and -7. Selective pharmacological inhibitors of specific caspases are desirable to reduce cell death for treatment of diseases characterized by increased apoptosis, such as stroke, heart disease and neurodegenerative diseases.

The first crystal structure of D-arginine dehydrogenase was determined at 1.06 Å resolution and its substrate specificity has been thoroughly investigated by crystallography and kinetic study [45, 88]. DADH catalyzes the conversion of D-arginine into 2-ketoarginine and ammonia. However, it is character-

ized by broad substrate specificity, being able to oxidize basic and hydrophobic D-amino acids of various sizes, but not reacting with acidic D-amino acids [45, 84]. A ping-pong bi-bi kinetic mechanism has been established by steady-state kinetic studies with D-arginine or D-histidine as substrate [88]. Comparison of the structures of DADH in complexes with iminoarginine and iminohistidine indicated very distinct binding modes between the two ligands, which is in agreement with detailed kinetic analysis on substrate specificity reported previously [45, 84]. Structural analysis indicates that the negatively charged side chain of Glu87 may be responsible for the high selectivity of DADH for positively charged substrates like D-arginine and D-lysine. Other active site residues, such as Tyr53, Tyr249, Met240 and Val242, form hydrophobic walls for specificity of the long aliphatic and unbranched substrate. Further comparison of structures of the ligand-free DADH and its complexes with imino acids has revealed that the protein active site undergoes major conformational upon binding of the ligand. Residues 50-56 were designated as an active site lid controlling the substrate accessibility to the active site, similar to those reported in other flavin-dependent enzymes [74]. It possesses a ligand-free conformation in the free enzyme but shows a product-bound conformation upon binding of the ligand. Comparison of DADH with other functional related enzymes like DAAO and L-amino acids oxidase has revealed some key components and structural arrangement that are responsible for their specificity toward different enantiomers and different amino acids [45]. Overall, the high-resolution structures for DADH will provide useful information for future studies of similar flavin-dependent enzymes [218].

Besides the crystallographic study, we have performed computational study on a novel crossover system for simulating biochemical networks. We have extended the functions of this deterministic-stochastic crossover method, making it a more competitive system for the simulation of biochemical networks while retaining accuracy and efficiency. The crossover method employs Bernoulli trial to introduce randomness into the simulation. Testing of this method on several systems like auto-regulatory gene network indicated that it not only retains the efficiency of deterministic methods, but also reflects the fluctuations generally only captured by stochastic methods [104]. When compared with other hybrid methods, the major advantage of our crossover method over these hybrid algorithms is it does not require partition-

ing of the modeled system. It is not actually a combination of deterministic and stochastic methods, but a deterministic method with the randomness introduced by Bernoulli trial. We have also introduced XML representation into this crossover system. The standard format of model representation makes it very convenient for researchers to read, write, and modify the system. Further, it would be easier to transfer the system among different software platforms. Another notable feature of this crossover system is its capability of simulating more complex Michaelis-Menten type reactions, instead of being restricted to simple elementary reactions. Overall, development of this crossover system would be very useful for researches and applications based on systems biology.

REFERENCES

1. Jacobson, M.D., M. Weil, and M.C. Raff, *Programmed cell death in animal development.* Cell, 1997. **88**(3): p. 347-54.
2. Rathmell, J.C. and C.B. Thompson, *THE CENTRAL EFFECTORS OF CELL DEATH IN THE IMMUNE SYSTEM.* Annual Review of Immunology, 1999. **17**(1): p. 781-828.
3. Alberts, K., et al., *"Chapter 18 Apoptosis: Programmed Cell Death Eliminates Unwanted Cells." Molecular Biology of the Cell* 5th ed. 2008: Garland Science.
4. Yan, N. and Y. Shi, *MECHANISMS OF APOPTOSIS THROUGH STRUCTURAL BIOLOGY.* Annual Review of Cell and Developmental Biology, 2005. **21**(1): p. 35-56.
5. Zimmermann, K.C., C. Bonzon, and D.R. Green, *The machinery of programmed cell death.* Pharmacol Ther, 2001. **92**(1): p. 57-70.
6. Hunter, A., E. LaCasse, and R. Korneluk, *The inhibitors of apoptosis (IAPs) as cancer targets.* Apoptosis, 2007. **12**(9): p. 1543-1568.
7. Cory, S., D.C.S. Huang, and J.M. Adams, *The Bcl-2 family: roles in cell survival and oncogenesis.* Oncogene, 2003. **22**(53): p. 8590-8607.
8. Zhang, N., et al., *The role of apoptosis in the development and function of T lymphocytes.* Cell Res, 2005. **15**(10): p. 749-769.
9. Ashkenazi, A. and V.M. Dixit, *Death Receptors: Signaling and Modulation.* Science, 1998. **281**(5381): p. 1305-1308.
10. Lavrik, I.N., A. Golks, and P.H. Krammer, *Caspases: pharmacological manipulation of cell death.* J Clin Invest, 2005. **115**(10): p. 2665-72.
11. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
12. Thompson, C.B., *Apoptosis in the pathogenesis and treatment of disease.* Science, 1995. **267**(5203): p. 1456-62.
13. Schmitt, C.A., *Senescence, apoptosis and therapy--cutting the lifelines of cancer.* Nat Rev Cancer, 2003. **3**(4): p. 286-95.
14. Hengartner, M.O., *The biochemistry of apoptosis.* Nature, 2000. **407**(6805): p. 770-6.
15. Thornberry, N.A., et al., *A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes.* Nature, 1992. **356**(6372): p. 768-74.
16. Yuan, J., et al., *The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme.* Cell, 1993. **75**(4): p. 641-52.
17. Lamkanfi, M., et al., *Alice in caspase land. A phylogenetic analysis of caspases from worm to man.* Cell Death Differ, 2002. **9**(4): p. 358-61.
18. Fuentes-Prior, P. and G.S. Salvesen, *The protein structures that shape caspase activity, specificity, activation and inhibition.* Biochem. J., 2004. **384**(2): p. 201-232.
19. Kumar, S., *Caspase function in programmed cell death.* Cell Death Differ, 2007. **14**(1): p. 32-43.
20. Degterev, A., M. Boyce, and J. Yuan, *A decade of caspases.* Oncogene, 2003. **22**(53): p. 8543-67.
21. Nicholson, D.W., *Apoptosis. Baiting death inhibitors.* Nature, 2001. **410**(6824): p. 33-4.
22. Goyal, L., *Cell death inhibition: keeping caspases in check.* Cell, 2001. **104**(6): p. 805-8.
23. Kang, T.-B., et al., *Caspase-8 Serves Both Apoptotic and Nonapoptotic Roles.* The Journal of Immunology, 2004. **173**(5): p. 2976-2984.
24. Bratton, S.B., et al., *Recruitment, activation and retention of caspases-9 and -3 by Apaf-1 apoptosome and associated XIAP complexes.* EMBO J, 2001. **20**(5): p. 998-1009.
25. Kischkel, F.C., et al., *Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8.* J Biol Chem, 2001. **276**(49): p. 46639-46.
26. Zhivotovsky, B. and S. Orrenius, *Caspase-2 function in response to DNA damage.* Biochem Biophys Res Commun, 2005. **331**(3): p. 859-67.
27. Porter, A.G. and R.U. Janicke, *Emerging roles of caspase-3 in apoptosis.* Cell Death Differ, 1999. **6**(2): p. 99-104.

28. Milhas, D., et al., *Caspase-10 triggers Bid cleavage and caspase cascade activation in FasL-induced apoptosis.* J Biol Chem, 2005. **280**(20): p. 19836-42.
29. Lakhani, S.A., et al., *Caspases 3 and 7: key mediators of mitochondrial events of apoptosis.* Science, 2006. **311**(5762): p. 847-51.
30. Thornberry, N.A., et al., *A Combinatorial Approach Defines Specificities of Members of the Caspase Family and Granzyme B. FUNCTIONAL RELATIONSHIPS ESTABLISHED FOR KEY MEDIATORS OF APOPTOSIS.* J. Biol. Chem., 1997. **272**(29): p. 17907-17911.
31. Fischer, U., R.U. Janicke, and K. Schulze-Osthoff, *Many cuts to ruin: a comprehensive update of caspase substrates.* Cell Death Differ, 2003. **10**(1): p. 76-100.
32. Timmer, J.C. and G.S. Salvesen, *Caspase substrates.* Cell Death Differ, 2007. **14**(1): p. 66-72.
33. Talanian, R.V., et al., *Substrate Specificities of Caspase Family Proteases.* J. Biol. Chem., 1997. **272**(15): p. 9677-9682.
34. Fang, B., et al., *Structural and kinetic analysis of caspase-3 reveals role for s5 binding site in substrate recognition.* J Mol Biol, 2006. **360**(3): p. 654-66.
35. Stennicke, H.R., et al., *Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8.* Biochem J, 2000. **350 Pt 2**: p. 563-8.
36. Schweizer, A., C. Briand, and M.G. Grutter, *Crystal Structure of Caspase-2, Apical Initiator of the Intrinsic Apoptotic Pathway.* J. Biol. Chem., 2003. **278**(43): p. 42441-42447.
37. Fu, G., et al., *Structural basis for executioner caspase recognition of P5 position in substrates.* Apoptosis, 2008.
38. Huang, X., S. Lee, and X. Chen, *Design of "smart" probes for optical imaging of apoptosis.* Am J Nucl Med Mol Imaging, 2011. **1**(1): p. 3-17.
39. Fisher, A.J., et al., *Crystal structure of baculovirus P35: role of a novel reactive site loop in apoptotic caspase inhibition.* EMBO J, 1999. **18**(8): p. 2031-9.
40. Hefti, M.H., J. Vervoort, and W.J. van Berkel, *Deflavination and reconstitution of flavoproteins.* Eur J Biochem, 2003. **270**(21): p. 4227-42.
41. Joosten, V. and W.J.H. van Berkel, *Flavoenzymes.* Current Opinion in Chemical Biology, 2007. **11**(2): p. 195-202.
42. Ghisla, S. and V. Massey, *New flavins for old: artificial flavins as active site probes of flavoproteins.* Biochem J, 1986. **239**(1): p. 1-12.
43. Gadda, G., *Oxygen activation in flavoprotein oxidases: the importance of being positive.* Biochemistry, 2012. **51**(13): p. 2662-9.
44. Leferink, N.G., et al., *Identification of a gatekeeper residue that prevents dehydrogenases from acting as oxidases.* J Biol Chem, 2009. **284**(7): p. 4392-7.
45. Fu, G., et al., *Conformational changes and substrate recognition in Pseudomonas aeruginosa D-arginine dehydrogenase.* Biochemistry, 2010. **49**(39): p. 8535-45.
46. Mattevi, A., *To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes.* Trends Biochem Sci, 2006. **31**(5): p. 276-83.
47. Faust, A., et al., *The Structure of a Bacterial l-Amino Acid Oxidase from Rhodococcus opacus Gives New Evidence for the Hydride Mechanism for Dehydrogenation.* Journal of Molecular Biology, 2007. **367**(1): p. 234-248.
48. Krebs, H.A., *Metabolism of amino-acids: Deamination of amino-acids.* Biochem J, 1935. **29**(7): p. 1620-44.
49. Frebortova, J., et al., *Catalytic reaction of cytokinin dehydrogenase: preference for quinones as electron acceptors.* Biochem J, 2004. **380**(Pt 1): p. 121-30.
50. Tsuge, H., et al., *Crystal structure of a novel FAD-, FMN-, and ATP-containing L-proline dehydrogenase complex from Pyrococcus horikoshii.* J Biol Chem, 2005. **280**(35): p. 31045-9.
51. van Berkel, W.J.H., N.M. Kamerbeek, and M.W. Fraaije, *Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts.* Journal of Biotechnology, 2006. **124**(4): p. 670-689.
52. Martinez-Martinez, I., et al., *Maximization of production of his-tagged glycine oxidase and its M261 mutant proteins.* Biotechnol Prog, 2006. **22**(3): p. 647-52.

53. Nishizawa, T., C.C. Aldrich, and D.H. Sherman, *Molecular analysis of the rebeccamycin L-amino acid oxidase from Lechevalieria aerocolonigenes ATCC 39243.* J Bacteriol, 2005. **187**(6): p. 2084-92.
54. Nishino, T., et al., *Flavins and Flavoproteins.* Proceedings of the 15th International Symposium on Flavins and Flavoproteins, 2005.
55. Dym, O. and D. Eisenberg, *Sequence-structure analysis of FAD-containing proteins.* Protein Sci, 2001. **10**(9): p. 1712-28.
56. Rossmann, M.G., D. Moras, and K.W. Olsen, *Chemical and biological evolution of a nucleotide-binding protein.* Nature, 1974. **250**(5463): p. 194-199.
57. Schulz, G.E., R.H. Schirmer, and E.F. Pai, *FAD-binding site of glutathione reductase.* Journal of Molecular Biology, 1982. **160**(2): p. 287-308.
58. Schulz, G.E., *Binding of nucleotides by proteins: Current opinion in structural biology 1992, 2: 61…-67.* Current Opinion in Structural Biology, 1992. **2**(1): p. 61-67.
59. Dym, O., et al., *The crystal structure of D-lactate dehydrogenase, a peripheral membrane respiratory enzyme.* Proc Natl Acad Sci U S A, 2000. **97**(17): p. 9413-8.
60. Mizutani, H., et al., *Three-dimensional structure of porcine kidney D-amino acid oxidase at 3.0 A resolution.* J Biochem, 1996. **120**(1): p. 14-7.
61. Karplus, P.A. and G.E. Schulz, *Refined structure of glutathione reductase at 1.54 A resolution.* J Mol Biol, 1987. **195**(3): p. 701-29.
62. Ingelman, M., V. Bianchi, and H. Eklund, *The three-dimensional structure of flavodoxin reductase from Escherichia coli at 1.7 A resolution.* J Mol Biol, 1997. **268**(1): p. 147-57.
63. Pollegioni, L., et al., *Physiological functions of D-amino acid oxidases: from yeast to humans.* Cell Mol Life Sci, 2007. **64**(11): p. 1373-94.
64. Kolodkin-Gal, I., et al., *D-Amino Acids Trigger Biofilm Disassembly.* Science, 2010. **328**(5978): p. 627-629.
65. Lam, H., et al., *D-amino acids govern stationary phase cell wall remodeling in bacteria.* Science, 2009. **325**(5947): p. 1552-5.
66. Yoshimura, T. and N. Esak, *Amino acid racemases: functions and mechanisms.* J Biosci Bioeng, 2003. **96**(2): p. 103-9.
67. Dunlop, D.S., et al., *The presence of free D-aspartic acid in rodents and man.* Biochem Biophys Res Commun, 1986. **141**(1): p. 27-32.
68. Hashimoto, A., et al., *Endogenous D-serine in rat brain: N-methyl-D-aspartate receptor-related distribution and aging.* J Neurochem, 1993. **60**(2): p. 783-6.
69. Snyder, S.H. and P.M. Kim, *D-amino acids as putative neurotransmitters: focus on D-serine.* Neurochem Res, 2000. **25**(5): p. 553-60.
70. Wolosker, H., *NMDA receptor regulation by D-serine: new findings and perspectives.* Mol Neurobiol, 2007. **36**(2): p. 152-64.
71. Raunio, R.P. and W.T. Jenkins, *D-alanine oxidase form Escherichia coli: localization and induction by L-alanine.* J Bacteriol, 1973. **115**(2): p. 560-6.
72. David, H.L., et al., *Methionine as methyl-group donor in the synthesis of Mycobacterium avium envelope lipids, and its inhibition by DL-ethionine, D-norleucine and DL-norleucine.* Acta Leprol, 1989. **7 Suppl 1**: p. 77-80.
73. Olsiewski, P.J., G.J. Kaczorowski, and C. Walsh, *Purification and properties of D-amino acid dehydrogenase, an inducible membrane-bound iron-sulfur flavoenzyme from Escherichia coli B.* J Biol Chem, 1980. **255**(10): p. 4487-94.
74. Mattevi, A., et al., *Crystal structure of D-amino acid oxidase: a case of active site mirror-image convergent evolution with flavocytochrome b2.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7496-501.
75. Todone, F., et al., *Active site plasticity in D-amino acid oxidase: a crystallographic analysis.* Biochemistry, 1997. **36**(19): p. 5853-60.

76. Verrall, L., et al., *The neurobiology of D-amino acid oxidase and its involvement in schizophrenia.* Mol Psychiatry, 2010. **15**(2): p. 122-37.
77. Khoronenkova, S.V. and V.I. Tishkov, *D-amino acid oxidase: physiological role and applications.* Biochemistry (Mosc), 2008. **73**(13): p. 1511-8.
78. LaRue, T.A. and J.F. Spencer, *The utilization of D-amino acids by yeasts.* Can J Microbiol, 1967. **13**(7): p. 777-88.
79. Ohtani, S., *Estimation of age from dentin by utilizing the racemization of aspartic acid: influence of pH.* Forensic Sci Int, 1995. **75**(2-3): p. 181-7.
80. Abe, H., et al., *Physiological function and metabolism of free D-alanine in aquatic animals.* Biol Pharm Bull, 2005. **28**(9): p. 1571-7.
81. Fisher, G.H., et al., *Free D-aspartate and D-alanine in normal and Alzheimer brain.* Brain Res Bull, 1991. **26**(6): p. 983-5.
82. Xia, Z.X. and F.S. Mathews, *Molecular structure of flavocytochrome b2 at 2.4 A resolution.* J Mol Biol, 1990. **212**(4): p. 837-63.
83. Li, C. and C.D. Lu, *Arginine racemization by coupled catabolic and anabolic dehydrogenases.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 906-11.
84. Li, C., X. Yao, and C.D. Lu, *Regulation of the dauBAR operon and characterization of D-amino acid dehydrogenase DauA in arginine and lysine catabolism of Pseudomonas aeruginosa PAO1*. Microbiology, 2010. **156**(Pt 1): p. 60-71.
85. Hallberg, B.M., et al., *Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase.* J Mol Biol, 2004. **341**(3): p. 781-96.
86. Quaye, O., et al., *Role of Glu312 in binding and positioning of the substrate for the hydride transfer reaction in choline oxidase.* Biochemistry, 2008. **47**(1): p. 243-56.
87. Vanoni, M.A., et al., *Limited Proteolysis and X-ray Crystallography Reveal the Origin of Substrate Specificity and of the Rate-Limiting Product Release during Oxidation of d-Amino Acids Catalyzed by Mammalian d-Amino Acid Oxidase†,‡.* Biochemistry, 1997. **36**(19): p. 5624-5632.
88. Yuan, H., et al., *Steady-state kinetic mechanism and reductive half-reaction of D-arginine dehydrogenase from Pseudomonas aeruginosa.* Biochemistry, 2010. **49**(44): p. 9542-50.
89. Saam, J., et al., *O2 reactivity of flavoproteins: dynamic access of dioxygen to the active site and role of a H+ relay system in D-amino acid oxidase.* J Biol Chem, 2010. **285**(32): p. 24439-46.
90. Kitano, H., *Computational systems biology.* Nature, 2002. **420**(6912): p. 206-10.
91. Kitano, H., *Systems biology: a brief overview.* Science, 2002. **295**(5560): p. 1662-4.
92. May, P. and E. May, *Twenty years of p53 research: structural and functional aspects of the p53 protein.* Oncogene, 1999. **18**(53): p. 7621-36.
93. Baldi, P. and S. Brunak, *Bioinformatics: the machine learning approach*. 2001: MIT Press. 400.
94. Shea, M.A. and G.K. Ackers, *The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation.* Journal of Molecular Biology, 1985. **181**(2): p. 211-230.
95. Arkin, A., J. Ross, and H.H. McAdams, *Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells.* Genetics, 1998. **149**(4): p. 1633-1648.
96. Endy, D., D. Kong, and J. Yin, *Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7.* Biotechnology and Bioengineering, 1997. **55**(2): p. 375-389.
97. You, L., P.F. Suthers, and J. Yin, *Effects of Escherichia coli Physiology on Growth of Phage T7 In Vivo and In Silico.* Journal of Bacteriology, 2002. **184**(7): p. 1888-1894.
98. Asthagiri, A.R. and D.A. Lauffenburger, *BIOENGINEERING MODELS OF CELL SIGNALING.* Annual Review of Biomedical Engineering, 2000. **2**(1): p. 31-53.
99. Gillespie, D.T., *Exact stochastic simulation of coupled chemical reactions.* The Journal of Physical Chemistry, 1977. **81**(25): p. 2340-2361.
100. Golding, I., et al., *Real-Time Kinetics of Gene Activity in Individual Bacteria.* Cell, 2005. **123**(6): p. 1025-1036.

101. Gillespie, F.T., *Approximate accelerated stochastic simulation of chemically reacting systems.* J Chem Phys, 2001. **115**: p. 1716-1733.
102. Kiehl, T.R., R.M. Mattheyses, and M.K. Simmons, *Hybrid simulation of cellular behavior.* Bioinformatics, 2004. **20**(3): p. 316-322.
103. Salis, H. and Y. Kaznessis, *Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions.* The Journal of Chemical Physics, 2005. **122**(5): p. 054103.
104. Sabnis, A. and R.W. Harrison, *A continuous-time, discrete-state method for simulating the dynamics of biochemical systems.* IEEE/ACM Trans Comput Biol Bioinform, 2011. **8**(2): p. 335-41.
105. Kitano, H., *Standards for modeling.* Nat Biotechnol, 2002. **20**(4): p. 337.
106. Hucka, M., et al., *The ERATO Systems Biology Workbench: enabling interaction and exchange between software tools for computational biology.* Pac Symp Biocomput, 2002: p. 450-61.
107. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, 2003. **19**(4): p. 524-531.
108. Lloyd, C.M., M.D. Halstead, and P.F. Nielsen, *CellML: its future, present and past.* Prog Biophys Mol Biol, 2004. **85**(2-3): p. 433-50.
109. Sauro, H.M., et al., *Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration.* OMICS, 2003. **7**(4): p. 355-72.
110. Wilkinson, D.J., *Stochastic modelling for systems biology*. Chapman & Hall/CRC mathematical and computational biology series. 2006, Boca Raton: Taylor & Francis. 254 p.
111. Tacconi, S., et al., *Increased caspase activation in peripheral blood mononuclear cells of patients with Alzheimer's disease.* Exp Neurol, 2004. **190**(1): p. 254-62.
112. Hartmann, A., et al., *Caspase-8 is an effector in apoptotic death of dopaminergic neurons in Parkinson's disease, but pathway inhibition results in neuronal necrosis.* J Neurosci, 2001. **21**(7): p. 2247-55.
113. Hermel, E., et al., *Specific caspase interactions and amplification are involved in selective neuronal vulnerability in Huntington's disease.* Cell Death Differ, 2004. **11**(4): p. 424-38.
114. LeBlanc, A.C., *The role of apoptotic pathways in Alzheimer's disease neurodegeneration and cell death.* Curr Alzheimer Res, 2005. **2**(4): p. 389-402.
115. Zidar, N., et al., *Caspases in myocardial infarction.* Adv Clin Chem, 2007. **44**: p. 1-33.
116. Takemura, G. and H. Fujiwara, *Morphological aspects of apoptosis in heart diseases.* J Cell Mol Med, 2006. **10**(1): p. 56-75.
117. Kim, H.S., et al., *Inactivating mutations of caspase-8 gene in colorectal carcinomas.* Gastroenterology, 2003. **125**(3): p. 708-15.
118. Soung, Y.H., et al., *CASPASE-8 gene is inactivated by somatic mutations in gastric carcinomas.* Cancer Res, 2005. **65**(3): p. 815-21.
119. Volkmann, X., et al., *Caspase activation is required for antiviral treatment response in chronic hepatitis C virus infection.* Hepatology, 2006. **43**(6): p. 1311-6.
120. Fischer, U. and K. Schulze-Osthoff, *Apoptosis-based therapies and drug targets.* Cell Death Differ, 2005. **12 Suppl 1**: p. 942-61.
121. Baskin-Bey, E.S., et al., *Clinical Trial of the Pan-Caspase Inhibitor, IDN-6556, in Human Liver Preservation Injury.* Am J Transplant, 2007. **7**(1): p. 218-25.
122. Fuentes-Prior, P. and G.S. Salvesen, *The protein structures that shape caspase activity, specificity, activation and inhibition.* Biochem J, 2004. **384**(Pt 2): p. 201-32.
123. Wei, Y., et al., *The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity.* Chem Biol, 2000. **7**(6): p. 423-32.
124. Shiozaki, E.N., et al., *Mechanism of XIAP-mediated inhibition of caspase-9.* Mol Cell, 2003. **11**(2): p. 519-27.
125. Garcia-Calvo, M., et al., *Purification and catalytic properties of human caspase family members.* Cell Death Differ, 1999. **6**(4): p. 362-9.

126. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*, in *Methods in Enzymology*, Charles W. Carter, Jr., Editor. 1997, Academic Press. p. 307-326.
127. McCoy, A.J., et al., *Likelihood-enhanced fast translation functions.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 4): p. 458-64.
128. Potterton, E., et al., *A graphical user interface to the CCP4 program suite.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 7): p. 1131-7.
129. Sheldrick, G.M. and T.R. Schneider, *SHELXL: high-resolution refinement.* Methods Enzymol, 1997. **277**: p. 319-43.
130. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
131. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
132. Lovell, S.C., et al., *Structure validation by Calpha geometry: phi,psi and Cbeta deviation.* Proteins, 2003. **50**(3): p. 437-50.
133. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
134. Merritt, E.A. and M.E. Murphy, *Raster3D Version 2.0. A program for photorealistic molecular graphics.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 6): p. 869-73.
135. Solomon, K., *Information theory and statistics*. 1959: Wiley, New York.
136. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.
137. Harrison, R.W., *Stiffness and energy conservation in molecular dynamics: An improved integrator*. Journal of Computational Chemistry, 1993. **14**(9): p. 1112-1122.
138. Harrison, R.W., D. Chatterjee, and I.T. Weber, *Analysis of six protein structures predicted by comparative modeling techniques.* Proteins, 1995. **23**(4): p. 463-71.
139. Peter, B., et al., *Modification of parameters of the charge equilibrium scheme to achieve better correlation with experimental dipole moments.* J Mol Model, 1999(5): p. 143–152.
140. Harrison, R.W., *A self-assembling neural network for modeling polymers.* Journal of Mathematical Chemistry, 1999. **26**(1/3): p. 125-137.
141. Agniswamy, J., B. Fang, and I.T. Weber, *Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis.* FEBS Journal, 2007. **274**(18): p. 4752-4765.
142. Wilson, K.P., et al., *Structure and mechanism of interleukin-1 beta converting enzyme.* Nature, 1994. **370**(6487): p. 270-5.
143. Onteniente, B., *Natural and synthetic inhibitors of caspases: targets for novel drugs.* Curr Drug Targets CNS Neurol Disord, 2004. **3**(4): p. 333-40.
144. Martinon, F., et al., *Activation of a pro-apoptotic amplification loop through inhibition of NF-kappaB-dependent survival signals by caspase-mediated inactivation of RIP.* FEBS Lett, 2000. **468**(2-3): p. 134-6.
145. Vucic, D., et al., *ML-IAP, a novel inhibitor of apoptosis that is preferentially expressed in human melanomas.* Curr Biol, 2000. **10**(21): p. 1359-66.
146. O'Brien, T. and D. Lee, *Prospects for caspase inhibitors.* Mini Rev Med Chem, 2004. **4**(2): p. 153-65.
147. Callus, B.A. and D.L. Vaux, *Caspase inhibitors: viral, cellular and chemical.* Cell Death Differ, 2007. **14**(1): p. 73-8.
148. Thornberry, N.A., et al., *A combinatorial approach defines specificities of members of the caspase family and granzyme B. Functional relationships established for key mediators of apoptosis.* J Biol Chem, 1997. **272**(29): p. 17907-11.
149. Lien, S., et al., *A substrate-phage approach for investigating caspase specificity.* Protein J, 2004. **23**(6): p. 413-25.

150. Yoshimori, A., et al., *Structural and functional definition of the specificity of a novel caspase-3 inhibitor, Ac-DNLD-CHO.* BMC Pharmacol, 2007. **7**: p. 8.
151. Schweizer, A., C. Briand, and M.G. Grutter, *Crystal structure of caspase-2, apical initiator of the intrinsic apoptotic pathway.* J Biol Chem, 2003. **278**(43): p. 42441-7.
152. Backes, C., et al., *GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W208-13.
153. Fu, G., et al., *Structural basis for executioner caspase recognition of P5 position in substrates.* Apoptosis, 2008. **13**(11): p. 1291-302.
154. Howard, A.D., et al., *IL-1-converting enzyme requires aspartic acid residues for processing of the IL-1 beta precursor at two distinct sites and does not cleave 31-kDa IL-1 alpha.* J Immunol, 1991. **147**(9): p. 2964-9.
155. Maibaum, J. and D.H. Rich, *Inhibition of porcine pepsin by two substrate analogues containing statine. The effect of histidine at the P2 subsite on the inhibition of aspartic proteinases.* J Med Chem, 1988. **31**(3): p. 625-9.
156. Otwinowski, Z.M., W., *Processing of X-ray diffraction data in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-326.
157. Navaza, J., *AMoRe: an automated package for molecular replacemnet.* Acta Crystallog. sect. D, 1994. **50**: p. 157-163.
158. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54 (Pt 5)**: p. 905-21.
159. Jones, T.A., et al., *Improved methods for building protein models in electron density maps and the location of errors in these models.* Acta Crystallogr A, 1991. **47 (Pt 2)**: p. 110-9.
160. Esnouf, R.M., *An extensively modified version of MolScript that includes greatly enhanced coloring capabilities.* J Mol Graph Model, 1997. **15**(2): p. 132-4, 112-3.
161. Esnouf, R.M., *Further additions to MolScript version 1.4, including reading and contouring of electron-density maps.* Acta Crystallogr D Biol Crystallogr, 1999. **55 (Pt 4)**: p. 938-40.
162. Harrison, R.W., *Stiffness and energy conservation in the molecular dynamics: an improved integrator.* J. Comp. Chem. , 1993. **14**: p. 1112-1122.
163. Bagossi, P., et al., *Improved Parameters for Generating Partial Charges: Correlation with Observed Dipole Moments.* Journal of Molecular Modeling, 1999. **5**(9): p. 143-152.
164. Sayle, R.A. and E.J. Milner-White, *RASMOL: biomolecular graphics for all.* Trends Biochem Sci, 1995. **20**(9): p. 374.
165. Nicholson, D.W., et al., *Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis.* Nature, 1995. **376**(6535): p. 37-43.
166. Qiu, L., et al., *Irreversibly inhibitory kinetics of 3,5-dihydroxyphenyl decanoate on mushroom (Agaricus bisporus) tyrosinase.* Bioorg Med Chem, 2005. **13**(22): p. 6206-11.
167. Ganesan, R., et al., *Extended substrate recognition in caspase-3 revealed by high resolution X-ray structure analysis.* J Mol Biol, 2006. **359**(5): p. 1378-88.
168. Agniswamy, J., B. Fang, and I.T. Weber, *Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis.* FEBS J, 2007. **274**(18): p. 4752-65.
169. Talanian, R.V., et al., *Substrate specificities of caspase family proteases.* J Biol Chem, 1997. **272**(15): p. 9677-82.
170. Lee, D., et al., *Potent and selective nonpeptide inhibitors of caspases 3 and 7 inhibit apoptosis and maintain cell functionality.* J Biol Chem, 2000. **275**(21): p. 16007-14.
171. Blanchard, H., et al., *Caspase-8 specificity probed at subsite S(4): crystal structure of the caspase-8-Z-DEVD-cho complex.* J Mol Biol, 2000. **302**(1): p. 9-16.
172. Riedl, S.J., et al., *Structural basis for the inhibition of caspase-3 by XIAP.* Cell, 2001. **104**(5): p. 791-800.
173. Pockros, P.J., et al., *Oral IDN-6556, an antiapoptotic caspase inhibitor, may lower aminotransferase activity in patients with chronic hepatitis C.* Hepatology, 2007. **46**(2): p. 324-9.

174. Irene T. Weber, B.F.a.J.A., *Caspases: Structure-Guided Design of Drugs to Control Cell Death.* Mini-Reviews in Medicinal Chemistry, 2008. **8**(11): p. 1154-1162.
175. Haas, D., et al., *Arginine degradation in Pseudomonas aeruginosa mutants blocked in two arginine catabolic pathways.* Mol Gen Genet, 1984. **193**(3): p. 437-44.
176. Lu, C.D., *Pathways and regulation of bacterial arginine metabolism and perspectives for obtaining arginine overproducing strains.* Appl Microbiol Biotechnol, 2006. **70**(3): p. 261-72.
177. Liu, P., et al., *Covalent reaction intermediate revealed in crystal structure of the Geobacillus stearothermophilus carboxylesterase Est30.* J Mol Biol, 2004. **342**(2): p. 551-61.
178. Otwinowski, Z., and Minor, W., *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol 1997. **267**: p. 307-326.
179. Fu, Z.Q., J. Rose, and B.C. Wang, *SGXPro: a parallel workflow engine enabling optimization of program performance and automation of structure determination.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 7): p. 951-9.
180. Lamzin, V.S. and K.S. Wilson, *Automated refinement of protein models.* Acta Crystallogr D Biol Crystallogr, 1993. **49**(Pt 1): p. 129-47.
181. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
182. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
183. Pawelek, P.D., et al., *The structure of L-amino acid oxidase reveals the substrate trajectory into an enantiomerically conserved active site.* EMBO J, 2000. **19**(16): p. 4204-4215.
184. Dixon, D.A., et al., *Conformations and electronic structures of oxidized and reduced isoalloxazine.* Biochemistry, 1979. **18**(26): p. 5770-5.
185. Kawazoe, T., et al., *Crystal structure of human D-amino acid oxidase: context-dependent variability of the backbone conformation of the VAAGL hydrophobic stretch located at the si-face of the flavin ring.* Protein Sci, 2006. **15**(12): p. 2708-17.
186. Chen, Z.W., et al., *Heterotetrameric sarcosine oxidase: structure of a diflavin metalloenzyme at 1.85 A resolution.* J Mol Biol, 2006. **360**(5): p. 1000-18.
187. Wohlfahrt, G., et al., *1.8 and 1.9 A resolution structures of the Penicillium amagasakiense and Aspergillus niger glucose oxidases as a basis for modelling substrate complexes.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 5): p. 969-77.
188. Li, J., et al., *Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases.* Biochemistry, 1993. **32**(43): p. 11507-15.
189. Hallberg, B.M., et al., *Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase.* J Mol Biol, 2002. **315**(3): p. 421-34.
190. Stoll, V.S., M.S. Kimber, and E.F. Pai, *Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase.* Structure, 1996. **4**(4): p. 437-47.
191. Lu, W., et al., *Binding of amino acid side-chains to S1 cavities of serine proteinases.* Journal of Molecular Biology, 1997. **266**(2): p. 441-461.
192. Brandsdal, B.O., J. Aqvist, and A.O. Smalas, *Computational analysis of binding of P1 variants to trypsin.* Protein Sci, 2001. **10**(8): p. 1584-95.
193. Umhau, S., et al., *The x-ray structure of D-amino acid oxidase at very high resolution identifies the chemical mechanism of flavin-dependent substrate dehydrogenation.* Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12463-8.
194. Wagner, M.A., et al., *Monomeric sarcosine oxidase: 1. Flavin reactivity and active site binding determinants.* Biochemistry, 2000. **39**(30): p. 8813-24.
195. Lindqvist, Y., *Refined structure of spinach glycolate oxidase at 2 A resolution.* J Mol Biol, 1989. **209**(1): p. 151-66.

196. Arima, J., et al., *Structural characterization of L-glutamate oxidase from Streptomyces sp. X-119-6.* FEBS J, 2009. **276**(14): p. 3894-903.
197. Bossi, R.T., et al., *Structure of FAD-bound L-aspartate oxidase: insight into substrate specificity and catalysis.* Biochemistry, 2002. **41**(9): p. 3018-24.
198. Moustafa, I.M., et al., *Crystal structure of LAAO from Calloselasma rhodostoma with an L-phenylalanine substrate: insights into structure and mechanism.* J Mol Biol, 2006. **364**(5): p. 991-1002.
199. Massey, V. and B. Curti, *On the reaction mechanism of Crotalus adamanteus L-amino acid oxidase.* J Biol Chem, 1967. **242**(6): p. 1259-64.
200. Dixon, M. and K. Kleppe, *d-amino acid oxidase I. Dissociation and recombination of the holoenzyme.* Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis, 1965. **96**(3): p. 357-367.
201. Ghisla, S., V. Massey, and Y.S. Choong, *Covalent adducts of lactate oxidase. Photochemical formation and structure identification.* J Biol Chem, 1979. **254**(21): p. 10662-9.
202. Massey, V., M. Stankovich, and P. Hemmerich, *Light-mediated reduction of flavoproteins with flavins as catalysts.* Biochemistry, 1978. **17**(1): p. 1-8.
203. Fu, G., et al., *Atomic-resolution structure of an N5 flavin adduct in D-arginine dehydrogenase.* Biochemistry, 2011. **50**(29): p. 6292-4.
204. Weber, I.T., et al., *Chapter 3 - Reaction Intermediates Discovered in Crystal Structures of Enzymes*, in *Advances in Protein Chemistry and Structural Biology*, C. Christo and K.-C. Tatyana, Editors. 2012, Academic Press. p. 57-86.
205. Katane, M. and H. Homma, *D-aspartate oxidase: the sole catabolic enzyme acting on free D-aspartate in mammals.* Chem Biodivers, 2010. **7**(6): p. 1435-49.
206. Harrison, R., *A self-assembling neural network for modeling polymers.* Journal of Mathematical Chemistry, 1999. **26**(1): p. 125-137.
207. McGaughey, G.B., M. Gagne, and A.K. Rappe, *pi-Stacking interactions. Alive and well in proteins.* J Biol Chem, 1998. **273**(25): p. 15458-63.
208. Kovalevsky, A.Y., et al., *Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114.* J Mol Biol, 2006. **363**(1): p. 161-73.
209. Weber, I.T. and R.W. Harrison, *Molecular mechanics calculations on protein-ligand complexes.* Perspectives in Drug Discovery and Design, 1998. **9-11**(0): p. 115-127.
210. Hafner, E.W. and D. Wellner, *Reactivity of the imino acids formed in the amino acid oxidase reaction.* Biochemistry, 1979. **18**(3): p. 411-417.
211. Albert, A., R. Goldacre, and J. Phillips, *The strength of heterocyclic bases.* J. chem. Soc., 1948: p. 2240 - 2249.
212. Katane, M., et al., *Role of the active site residues arginine-216 and arginine-237 in the substrate specificity of mammalian D-aspartate oxidase.* Amino Acids, 2011. **40**(2): p. 467-76.
213. Katane, M., et al., *Molecular cloning of a cDNA encoding mouse D-aspartate oxidase and functional characterization of its recombinant proteins by site-directed mutagenesis.* Amino Acids, 2007. **32**(1): p. 69-78.
214. Setoyama, C., et al., *Engineering the substrate specificity of porcine kidney D-amino acid oxidase by mutagenesis of the "active-site lid".* J Biochem, 2006. **139**(5): p. 873-9.
215. Ramsey, S., D. Orrell, and H. Bolouri, *Dizzy: stochastic simulation of large-scale genetic regulatory networks.* J Bioinform Comput Biol, 2005. **3**(2): p. 415-36.
216. Hynne, F., S. Dano, and P.G. Sorensen, *Full-scale model of glycolysis in Saccharomyces cerevisiae.* Biophys Chem, 2001. **94**(1-2): p. 121-63.
217. Pahle, J., *Biochemical simulations: stochastic, approximate stochastic and hybrid approaches.* Briefings in Bioinformatics, 2009. **10**(1): p. 53-64.
218. Fang, B., et al., *Caspase-3 binds diverse P4 residues in peptides as revealed by crystallography and structural modeling.* Apoptosis, 2009. **14**(5): p. 741-52.

APPENDICES

Appendix A: List of crystal structures

Protein	Ligand	PDB code	Resolution (Å)
Caspases-3	Ac-LDESD-Cho	3EDQ	1.61
Caspases-3	Ac-YVAD-Cho	3GJS	1.90
DADH		3NYC	1.06
DADH	Imino-arginine	3NYE	1.30
DADH	imino-histidine	3NYF	1.30
DADH	Ketoleucine N5 adduct	3SM8	1.07
DADH	imino-lysine	4H1A	1.09
DADH	imino-methionine	4H1B	1.05
DADH	imino-phenylalanine	4H1C	1.05
DADH	imino-proline	4H1F	1.03

Appendix B: List of publications

1. Fu G, Chumanevich AA, Agniswamy J, Fang B, Harrison RW, Weber IT. Structural basis for executioner caspase recognition of P5 position in substrates. *Apoptosis 2008*

2. Fang B, Fu G, Agniswamy J, Harrison RW, Weber IT. Caspase-3 binds diverse P4 residues in peptides as revealed by crystallography and structural modeling. *Apoptosis*. 2009

3. Fu, G, Yuan, H., Li, C., Lu, C. D., Gadda, G., Weber, I. T Conformational Changes and Substrate Recognition in Pseudomonas aeruginosa d-Arginine Dehydrogenase. *Biochemistry.* 2010

4. Yuan, H., Fu, G, Phillip T. B, Weber, I. T, Gadda, G., Steady-state kinetic mechanism and reductive half-reaction of D-arginine dehydrogenase from Pseudomonas aeruginosa. *Biochemistry*, 2010.

5. Fu G, Yuan H, Wang S, Gadda G, Weber IT Atomic-resolution structure of an N5 flavin adduct in D-arginine dehydrogenase. *Biochemistry*, 2011.

6. Irene T. Weber, Johnson Agniswamy, Guoxing Fu, Chen-Hsiang Shen, Robert W. Harrison. (2012) Reaction Intermediates Discovered in Crystal Structures of Enzymes. In “Structural and Mechanistic Enzymology: Bringing Computations and Experiments Together”, Eds., Christo Christov and

Tatyana Karabencheva, Advances in Protein Chemistry and Structural Biology, Elsevier, Inc. APCSB-Vol. 87, ISBN: 9780123983121

Appendix C: Mutagenesis study of DADH

Crystallographic study and kinetic studies have showed that DADH prefers positively charged D-amino acids and our analysis suggested that DADH Glu87 may play an important role in its substrate specificity. In order to test this hypothesis, we have performed mutagenesis study followed by crystallographic study on this enzyme. The idea of this study is to mutate the negatively charged Glu87 into positively charged arginine or hydrophobic phe, in an attempt to see decreased activity towards positively charged D-amino acids but increased activity towards negatively charged or hydrophobic D-amino acids. In addition, we also made a mutation of Tyr53 into phe. Tyr53 is located on the "active site lid" of DADH. It undergoes a major conformational change and forms a polar interaction with the ligand with substrate binding. This residue may be important for the active site conformational change during the substrate binding and product releasing. A mutation of Y53F will eliminate the polar interaction between Tyr53 and the ligand. We would like to investigate how this will affect the activity of DADH.

The standard site-directed mutagenesis protocol did not work well for introducing mutations E87R, E87F, and Y53F into the wild type DADH cDNA, respectively, due to the high GC content near these regions. Asymmetric PCR was then applied to construct the mutations. Taking mutation E87R as an example, the sequence between 5' NcoI site and DADH residue 87 were amplified with primers 5'-GGAGGAATTAACCATGGGTCATCATC (NcoI cleavage site underlined) and 5'-GACCACCATACGCGGACGC (E87R mutation site underlined). The sequence between DADH residue 87 and 3' HindIII site were amplified with primers 5'- GCGTCCGCGTATGGTGGTC (E87R mutation site underlined) and 5'- GGTAAGCTTCA GGGGGACAGGCG (HindIII cleavage site underlined). The corresponding primers for E87F are 5'-GACCACCATGAACGGACGC and 5'-GCGTCCGTTCATGGTGGTC, and the primers for Y53F are 5'-GTGCCGAAGGCCACCGTGTAG and 5'- CTACACGGTGGCCTTCGGCAC. The two pieces of PCR products were then purified by gel extrac-

tion kit and used as megaprimers for each other in the second PCR to generate the full length DADH sequence. Primers containing 5' NcoI site and 3' HindIII site were used to amplify the whole DADH sequence with desired mutation in the third PCR. The final PCR transcript was excised by NcoI and HindIII endonucleases and is subcloned into the expression vector pBAD-His-6 used in previous DADH studies [83]. The resulting plasmid was used to transform *E. coli* Top10, and transformants were selected on LB plates with ampicillin. The sequences of the mutations were confirmed by nucleotide sequencings with pBAD forward primer. The expression of these mutants followed the same as described in previous studies [45, 83].