

12-2009

Structural Basis of Caspase-3 Substrate Specificity Revealed by Crystallography, Enzyme Kinetics, and Computational Modeling

Bin Fang
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/biology_diss

 Part of the [Biology Commons](#)

Recommended Citation

Fang, Bin, "Structural Basis of Caspase-3 Substrate Specificity Revealed by Crystallography, Enzyme Kinetics, and Computational Modeling." Dissertation, Georgia State University, 2009.
https://scholarworks.gsu.edu/biology_diss/69

This Dissertation is brought to you for free and open access by the Department of Biology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Biology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

STRUCTURAL BASIS OF CASPASE-3 SUBSTRATE SPECIFICITY REVEALED BY
CRYSTALLOGRAPHY, ENZYME KINETICS, AND COMPUTATIONAL MODELING

by

BIN FANG

Under the Direction of Irene T. Weber

ABSTRACT

Caspase-3 is a cysteine protease that hydrolyzes diverse intracellular proteins during programmed cell death (known as apoptosis). It has been a popular target for drug design against abnormal cell death for more than a decade. No approved caspase based drug, however, is available so far. Therefore, structural insights about the substrate recognition of caspase-3 are needed for the future development of caspase-3 based inhibitors and drugs. In this study, crystal structures of recombinant caspase-3 in complex with seven substrate analog inhibitors, including acetyl (Ac)-DEVD-aldehyde (Cho), Ac-DMQD-Cho, Ac-IEPD-Cho, Ac-YVAD-Cho, Ac-WEHD-Cho, Ac-VDVAD-Cho, and tert-butoxycarbonyl (Boc)-D-fluoromethylketone (Fmk), have been analyzed in combination with enzyme kinetic data and computational models.

Seven crystal structures were determined at resolutions of 1.7-2.6Å. The binding conformation of each inhibitor residue at P1-P4 position was analyzed. The negative P1 aspartic acid side chain is exclusively required by the positive S1 pocket of caspase-3. Small hydrophobic P2 residues are preferred by the nonpolar S2 pocket formed by Y204, W206, and F256.

Although hydrophilic residues at P3 position tend to fit better, hydrophobic residues also can be accommodated by the plastic S3 pocket. Two substrate binding sites were found in the S4 pocket, one formed by main chain atoms of F250 and side chain atoms of N208 and the other formed by aromatic side chains of W206 and W214. These two binding sites are responsible for the binding of hydrophilic and hydrophobic P4 residues, respectively. Furthermore, the S5 subsite of caspase-3 formed by side chains of F250 and F252 was discovered. It stabilizes hydrophobic P5 residues on the substrates by an induced fit mechanism.

Computational studies were performed to help improve prediction of protein structures and protein-ligand interactions. Based on the Morse's function, a novel potential function with only three adjustable parameters per residue pair was developed, which will significantly increase the efficiency of protein structure prediction and molecular mechanics. Altogether, our studies have provided valuable information for the future caspase-3 based drug development.

INDEX WORDS: Enzyme catalysis, Cysteine protease, Protein recognition, Apoptosis, Induced fit

STRUCTURAL BASIS OF CASPASE-3 SUBSTRATE SPECIFICITY REVEALED BY
CRYSTALLOGRAPHY, ENZYME KINETICS, AND COMPUTATIONAL MODELING

by

BIN FANG

A Dissertation Submitted in Partial Fulfillment of Requirements for the Degree of
Doctor of Philosophy
in the College of Arts and Sciences
Georgia State University

2009

Copyright by
Bin Fang
2009

STRUCTURAL BASIS OF CASPASE-3 SUBSTRATE SPECIFICITY REVEALED BY
CRYSTALLOGRAPHY, ENZYME KINETICS, AND COMPUTATIONAL MODELING

by

BIN FANG

Committee Chair: Irene T. Weber

Committee: Robert W. Harrison
Giovanni Gadda

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2009

ACKNOWLEDGEMENTS

I would love to give my thanks first to my dear advisor Dr. Irene T. Weber. She is the person who led me into the fantastic area of structural biology. She is the person who gave me guidance, advice, and encouragement throughout my entire graduate study at Georgia State University. She is the person who not only educated me about reading, writing, and experimental skills, but also trained me to think as a scientist. She is the person who not only cared about students' studies but also respected their interests and needs. She is the person to whom I have to say: Thank you so much!

My sincere thanks also go to my committee member Dr. Robert Harrison. He advised my studies in computational biology. His brilliant ideas always gave me inspirations. I also want to thank another committee member Dr. Giovanni Gadda who gave me a lot of valuable advices on my study, writing, and presentation.

I give special thanks to Dr. Peter Boross who taught me protein purification and enzyme kinetics; Dr. Tracy Tie, Dr. Ping Liu, and Yuanfang Wang who taught me crystallography; and Dr. Johnson Agniswamy who gave me in-depth suggestions for my caspases studies. Thanks also go to Andrey Kovalevsky, Xianfeng Chen, Alexander Chumanevich, Fengling Liu, Hao Wang, Patra Volarath, Guioxing Fu, Brian Shen, Ying Zhang, Xi Xia Yu, and Ting Chiu who gave me sincere advice and help for research and presentations.

I also thank the staffs at the SER-CAT (ID-22 beamline) at the Advanced Photon Source, Argonne National Laboratory for their assistance during X-ray data collection. 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 wife Wei Li, my mother Wei Feng, and my father Shiyu Fang. Thank you for being my all-time support!

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
LIST OF ABBREVIATIONS.....	xiii
1. GENERAL INTRODUCTION.....	1
1.1. Caspase Enzyme.....	1
1.2. Caspases and Apoptosis.....	2
1.3. Caspase Substrate Specificity.....	6
1.4. Caspase Inhibitors.....	9
1.5. Protein Crystallography and Caspase Structures.....	16
1.6. Protein Structure Prediction and Potential Functions.....	21
1.7. Objectives.....	24
2. CASPASE-3 BINDS DIVERSE P4 RESIDUES IN PEPTIDES AS REVEALED BY CRYSTALLOGRAPHY AND STRUCTURAL MODELING.....	35
2.1. Introduction.....	35
2.2. Materials and Methods.....	37
2.2.1. Plasmids and Recombinant Proteins.....	37
2.2.2. Enzyme Kinetic Assays.....	38
2.2.3. Crystallographic Analysis.....	38
2.2.4. Molecular Modeling.....	39

2.3. Results.....	40
2.3.1. Analysis of P4 Residues in Known Caspase-3 Substrates.....	40
2.3.2. Inhibition Constants of Caspase-3 Inhibitors.....	40
2.3.3. Overall Structures of Four Caspase-3 Complexes	41
2.3.4. Inhibitor Interactions in the S1-S3 Subsites.....	43
2.3.5. S4 Subsite.....	45
2.3.6. Correlation of Structural Interactions with Inhibition	47
2.3.7. Predicted Binding Of Diverse P4 Residues	48
2.4. Discussion.....	51
3. STRUCTURAL AND KINETIC ANALYSIS OF CASPASE-3 REVEALS ROLE FOR S5 BINDING POCKET IN SUBSTRATE RECOGNITION.....	66
3.1. Introduction.....	66
3.2. Materials and Methods.....	68
3.2.1. Protein Expression and Purification.....	68
3.2.2. Enzyme Kinetic Assays	69
3.2.3. Crystallographic Analysis.....	70
3.2.4. Protein Data Bank Accession Codes.....	71
3.3. Results and Discussion	71
3.3.1. Crystal Structures.....	71
3.3.2. Caspase-3 Interactions with Peptide Analogs.....	73
3.3.3. Conformational Change when Caspase-3 Binds the P5-Containing Peptide	75

3.3.4.	Enzyme Kinetics and Relative Inhibition	75
3.3.5.	Roles of S2 and S3 in Substrate Recognition and Caspase-3 Activity	76
3.3.6.	Role of S5 Pocket in Caspase Recognition of Substrates	78
3.4.	Conclusions.....	80
4.	COMPACT, DIFFERENTIABLE, KNOWLEDGE-BASED POTENTIAL FUNCTIONS FOR EVALUATING PROTEIN MODEL QUALITY	93
4.1.	Introduction.....	93
4.2.	Materials and Methods.....	95
4.2.1.	Calculation of Total Potential	95
4.2.2.	Parameter Determination Using Genetic Algorithm.....	96
4.2.3.	Training Data and Decoy Sets	97
4.2.4.	Performance Measurement	98
4.3.	Results and Discussion	98
4.3.1.	Parameters Determined for 210 Potential Functions	98
4.3.2.	Performance on 70 Standard Multiple Decoy Sets.....	99
4.3.3.	Performance on 9 Docking Decoy Sets	102
4.3.4.	Conclusions and Future Work	102
5.	OVERALL SUMMARY	111
	LITERATURE CITED	113
	APPENDICES	126

LIST OF TABLES

Table 2.1. Crystallographic data collection and refinement statistics	55
Table 2.2. Inhibition constants.....	56
Table 3.1. Crystallographic Data Collection and Refinement Statistics.....	81
Table 3.2. Polar interactions of caspase-3 with peptide analogs	82
Table 3.3: Inhibition constants.....	83
Table 3.4: Kinetic parameters of caspase-3 substrates	84
Table 4.1. 70 multiple decoy sets for testing.....	104
Table 4.4. The success rates and average Z-scores of different SCM potentials.....	107
Table 4.5. The success rate of other energy models	108
Table 4.6. The Ranking of native structures for 9 docking decoy sets.....	109

LIST OF FIGURES

Figure 1.1. Grouping of 14 mammalian caspase family members	25
Figure 1.2. Extrinsic and intrinsic signaling pathway of apoptosis	26
Figure 1.3. Schematic presentation of caspase substrate binding site.	27
Figure 1.4. Peptidomimetic inhibitors.	28
Figure 1.5. Isatin sulfonamide analog inhibitors.	29
Figure 1.6. Quinoline derivative inhibitors.....	30
Figure 1.7. Structures of VX-765 and IDN-6556	31
Figure 1.8. Crystal packing and the Bragg's law.....	32
Figure 1.9. Hanging drop vapor diffusion system and crystal samples.....	33
Figure 1.10. The list of caspase structures currently deposited in PDB.	34
Figure 2.1. The occurrence of different amino acids at P4 in the cleavage sites of natural substrates of caspase-3.....	57
Figure 2.2. Inhibition characterization of caspase-3 inhibitors.....	58
Figure 2.3. Binding conformations of inhibitors.	59
Figure 2.4. Schematic representation of hydrogen bond and ionic interactions between caspase-3 and the inhibitors.....	60
Figure 2.5. Comparison between subsites in four new complexes (color) and published structure of caspase-3/DEVD (grey) (2H5I).....	61
Figure 2.6. P4 binding site on caspase-3.....	62
Figure 2.7. Superposition of inhibitors in the crystal structures (green) and structural models (purple).....	63
Figure 2.8. Predicted binding affinities for diverse P4 residues.	64

Figure 2.9. Predicted binding conformations of fifteen different P4 residues in caspase-3/XEVD.	65
Figure 3.1. Overall structure of caspase-3/DMQD.....	85
Figure 3.2. His121 has different side-chain conformations in the two p17/p12 heterodimers of caspase-3/DMQD.....	86
Figure 3.3. Structure of peptide analog inhibitors. (.....	87
Figure 3.4. Schematic representation of the caspase-3 interactions with inhibitors.....	88
Figure 3.5 The superimposed complexes of caspase-3/DEVD and caspase-3/VDVAD.....	89
Figure 3.6. Structural comparison of caspase-3 and caspase-2.	90
Figure 3.7. Sequence homology among six caspase family members.....	91
Figure 3.8 Different compositions of active site groove.....	92
Figure 4.1. Accuracy of different potential functions.....	110

LIST OF ABBREVIATIONS

Å	angstrom
ALA	alanine
Asp	aspartic acid
Asn	asparagines
Ac	acetyl
Boc	tert-butoxycarbonyl
Arg	arginine
C α	alpha carbon
cDNA	complementary deoxyribonucleic acid
CHO	aldehyde
CYS	cysteine
C-terminal	carboxyl terminal
Da	dalton
DDT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
Fmk	fluoromethyl ketone
Glu	glutamic acid
Gln	glutamine
Gly	glycine
His	histidine
Ile	isoleucine

L	liter
LB	Luria-Bertani
Leu	leucine
Lys	lysine
kDa	kilodaltons
MAD	multiplexwavelength anomalous diffraction
Met	methionine
mM	millimolar
MW	molecular weight
NMR	nuclear magnetic resonance
N-terminal	amino terminal
OMe	methylated oxygen
Pro	proline
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pNA	p-nitroanilide
Phe	phenylalanine
RMSD	root mean square deviation
SDS	sodium dodecyl sulphate
Ser	serine

Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine

1. GENERAL INTRODUCTION

1.1. Caspase Enzyme

Caspases are cysteinyl aspartate proteases that specifically cleave their substrate proteins after an aspartic acid. Since the discovery of the first caspase family member caspase-1 (initially known as interleukin-1 β -converting enzyme, ICE) in 1992 (Cerretti, Kozlosky et al. 1992; Thornberry, Bull et al. 1992), numerous caspase family members have been found in various vertebrate and invertebrate species, including 11 in humans, 10 in mice, 4 in birds, 4 in zebra fish, 8 in amphibians, 7 in insects, and 3 in nematodes (Lamkanfi, Declercq et al. 2002). Caspases are widely involved in differentiation, cell death, and inflammation processes. Although many caspases function in nonapoptotic pathways, the evolutionarily conserved role of caspases is to execute programmed cell death.

Caspase zymogens (known as procaspases) consist of three domains. A big subunit with a molecular weight around 20 kDa (p20) and a small subunit around 10 kDa (p10) are connected by a short linker region. At the N-termini, different caspases have their prodomains varying in length (figure 1.1). Some of them have big prodomains, such as caspase-2, 8, and 9. These caspases usually interact with death-inducing signaling complex (DISC) through their prodomains during the activation process. Other caspases, such as caspase-3, 6, and 7, have short prodomains, and their activation normally depends on cleavage by other caspases. The full length sequences of human caspases share homologies from 20% to 77%. Caspase-4 and 8 share the lowest homology while caspase-4 and 5 share the highest (Cohen 1997).

Procaspases have limited catalytic activities. During maturation, the prodomain is removed and the linker region between p20 and p10 is cleaved. As a result, two subunits will be assembled into a heterodimer p10/p20 which contains one catalytic groove responsible for

substrate recognition and hydrolysis. X-ray crystallography data suggested that caspases normally form heterotetramers $(p10/p20)_2$ under physiological conditions, although it is not required for the catalytic activity.

Based on their physiological functions in cells, caspases are divided into inflammatory caspases and apoptotic caspases. As illustrated in the Figure 1.1, the group I caspases, including caspase-1, 4, 5, 11, 12, 13, 14, belong to inflammatory caspases. Apoptotic caspases fall into two subgroups based on the position and function in the apoptotic signaling pathways. Caspase-2, 8, 9, 10 are initiator caspases (group II) and caspase-3, 6, 7 are executioner caspases (group III). The detailed functions of apoptotic caspases are introduced in the following section.

1.2. Caspases and Apoptosis

The programmed cell death, also named apoptosis, is a critical event in the cell life cycle. It is characterized by a series of highly ordered cell morphological and biochemical changes, including blebbing, loss of membrane asymmetry and attachment, cell shrinkage, nuclear fragmentation, chromatin condensation, and chromosomal DNA fragmentation. In the end, the cell will be lysed and debris will be degraded. In general, apoptosis occurs when a cell is injured beyond repair, infected by virus, or under stressful conditions such as heat or radiation. On the other hand, it is also required for tissue development, such as the differentiation of fingers and toes. The regular apoptosis is a highly controlled process and is not harmful to the organisms, which makes it different from necrosis, another type of cell death.

The apoptotic process is controlled by diverse cell signals initiated both extracellularly and intracellularly. Caspases are the major component of both pathways. According to the position in the signaling pathway, human apoptotic caspases are divided into two groups. Caspase-2, 8, 9, and 10 are initiator caspases because they are located on the upstream of the

signaling cascade. Caspase-3, 6, and 7 are executioner caspases which are located on the bottom of the cascade. Almost all healthy cells contain several caspases. These caspases, however, are in their inactive zymogen form, named procaspases. All the procaspases consist of an N-terminal domain and a catalytic domain formed by a big subunit p20 and a small subunit p10. The primary structures of catalytic domains share high homology for all apoptotic caspases while their N-terminal domains vary a lot. The initiator caspases usually contain large N-terminal domains, such as the caspase recruitment domain (CARD) in caspase-2, -9, and a pair of death effector domains (DEDs) in caspase-8 and -10. These domains consist of specific protein-protein interaction motifs that play crucial roles in the activation process. Upon the stimulation of cell death signal, the N-terminal domains of initiator caspases will mediate the recruitment of the procaspases to specific death signaling complexes. The procaspase molecules will subsequently undergo self activation by 'proximity-induced' mechanism (Degterev, Boyce et al. 2003; Fuentes-Prior and Salvesen 2004). In contrast, the executioner caspases only have short N-terminal domains and thus lack the capability of self activation *in vivo*. Their activation depends on the cleavage of initiator caspases.

The apoptosis pathway initiated by extracellular signals is called the extrinsic pathway (Fig 1.2) (Zhang, Hartig et al. 2005). It starts from specific molecules outside of the cells, known as pro-apoptotic ligands. These ligands include Apo2L/TRAIL and CD95L/FasL. The pro-apoptotic ligand binds to their receptors on the cell surface. The receptor, known as death receptor, is characterized by extracellular cysteine-rich domains (CRDs) and belongs to the tumor necrosis factor (TNF) family of proteins (Lavrik, Golks et al. 2005). A number of death receptors have been discovered, and some of them have been well studied, such as TNFR1 (p55 or CD120a), Fas (Apo1 or CD95), and TRAIL (R1/R2) (Yan and Shi 2005). All the death

receptor proteins have a death domain (DD) in their intracellular domains. They are responsible for binding adapter proteins during the caspase activation. The death ligands form homotrimers when they are activated, which induces the oligomerization of death receptor upon binding. The death domains of death receptor will subsequently bind to adaptor proteins such as FADD and TRADD. These adapter proteins also have death domains and they interact with the death receptor through homotypic interactions (Cohen 1997). In addition, homotypic interactions are also formed between the death effector domains (DED) of the adaptor protein and the pro-domain of procaspase-8, which leads to the oligomerization of procaspase-8 (Lavrik, Golks et al. 2005). Then, the death-inducing signaling complex (DISC) is formed. When the local enzyme concentration of procaspase-8 is increased in the DISC, procaspase-8 molecules begin to cleave each other, although a single procaspase-8 only has weak proteolytic activity (Lavrik, Golks et al. 2005). As a result, the active initiator caspase-8 molecules are released. In type I cells, once caspase-8 is activated, it will start to process downstream executioner procaspases, including procaspase-3, 6 and 7. After that, the active executioner caspases will cleave their substrates and the apoptosis begins (Kumar 2007).

In the type II cells, the apoptosis signal coming from extrinsic death ligands is usually not strong enough to trigger the entire apoptotic cascade. As a result, it requires another loop to amplify the signal (Fig 1.2) (Zhang, Hartig et al. 2005). This loop is mitochondria-dependent signal pathway which will be introduced in the next paragraph. The bridge between these two pathways is built by a Bcl-2 family protein Bid. Following the activation, Bid will be cleaved by active caspase-8. The truncated protein tBid in turn translocates to the mitochondria and starts the apoptosis with other two Bcl-2 family members Bax and Bak (Yan and Shi 2005).

Another type of apoptosis pathway is called the intrinsic pathway (Fig 1.2). This is usually triggered by the death signals from inside the cells in response to diverse cell stresses, such as DNA damage, a defective cell cycle, detachment from the extracellular matrix, hypoxia, and loss of cell survival factors. The intrinsic apoptosis initiates from mitochondria and it is regulated by Bcl-2 protein family. The Bcl-2 protein family is characterized by the conserved Bcl-2 homology domain (BH) (Yan and Shi 2005). Over twenty Bcl-2 family members fall into two categories, the pro-apoptotic Bcl-2 and the anti-apoptotic Bcl-2 (Kumar 2007). The anti-apoptotic Bcl-2 proteins, such as Bcl-2 and Bcl-xL in mammals, reduce the permeability of mitochondrial membrane and thus suppress the release of mitochondrial proteins. During activation, the pro-apoptotic Bcl-2 proteins, such as Bax and Bak, will undergo homo-oligomerization. It is thought that these oligomers form pores which facilitate the release of mitochondrial proteins, although the conclusive evidence is lacking (Yan and Shi 2005). With the help of Bax and Bak, mitochondria proteins including cytochrome c and SMAC/DIABLO are released into the cytosol. The SMAC/DIABLO can promote apoptosis by directly interacting with inhibitor of apoptosis proteins (IAPs). The interaction interrupts the binding between IAPs and caspases and thus activates caspases. Meanwhile, Cytochrome c binds the adaptor apoptotic protease activating factor-1 (Apaf-1) to form a big multiprotein complex known as apoptosome. The apoptosome is a key component of the intrinsic pathway because procaspase-9 is recruited and activated from within the complex. Similar to the extrinsic pathway, the initiator caspase-9 will activate downstream executioner caspases upon maturation.

The other two initiator caspases are caspase-2 and caspase-10. The activation of caspase-2 is similar to that of caspase-9. A big multiprotein complex is formed resembling the Apaf-1/caspase-9 complex. The pro-domain of caspase-10 contains two DEDs, and its activation

depends on the oligomerization process similar to caspase-8. The physiological functions of caspase-2 and 10 in apoptosis, however, remain a matter of considerable debate (Kumar 2007).

Caspase-3 is a major executioner caspase that cleaves the majority of cellular substrates in apoptotic cells. It can be activated by caspase-8 and 9 but not caspase-2 (Kumar 2007). Previous study in caspase-3 mutant mice in genetically mixed 129/SvJ_C57BL/6 background showed that mutant mice die prenatally due to reduced cell death in the central nervous system (CNS). However, in pure C57BL/6 background, caspase-3 null animals are viable but show reduced fertility. This comparison indicates that caspase-3 is redundant for most developmental cell death (Lakhani, Masud et al. 2006). In addition, caspase-7 null animal also showed normal development. In C57BL/6 background, Casp3/Casp7 double knockout (DKO) mice die rapidly after birth. Cells from DKO mice showed resistance to extrinsic apoptosis but not intrinsic apoptosis, although the activation of intrinsic apoptosis was delayed. It suggests that caspase-3 and 7 may amplify the signal of intrinsic apoptosis. Caspase-3, 6, and 7 share a similar structure, but caspase-6 shows different substrate specificity from caspase-3 and 7. The function of caspase-6 in apoptosis is still unclear. The mutation and knockout results suggest that one executioner caspase may be most important in a particular cell death pathway while others are either redundant or compensate. And their activities may be context dependent (Kumar 2007).

1.3. Caspase Substrate Specificity

Understanding the substrate specificities of caspases can help us to predict natural caspase substrates and understand apoptotic signaling pathways. More importantly, it can provide guidance for the design of inhibitors and drugs. This work can be done in two different ways, either by investigating binding preferences in caspase active site grooves or by analyzing the cleavage site sequences on their natural substrates. Obviously the first way is easier at the

early stage of investigation as finding natural substrates of caspases is relatively difficult. In 1997, two different groups, Thornberry and Talanian, established the basic understanding of caspase substrate preference by using synthetic peptide substrates (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997). In the following years, the result was refined by a number of studies, which led to a widely accepted understanding of caspase substrate specificity. As illustrated in Figure 1.3, most caspases recognize tetrapeptide sequences. From the N-terminus to the C-terminus, positions of each residue on the substrate are named P4, P3, P2, and P1. The corresponding binding sites on the protein are named S4-S1. The enzyme can hydrolyze the peptide bond after the P1 Asp. From P4 to P1, caspase-1,4,5,14 prefer W/YExD (x refers to any amino acid); caspase-8,9,10 prefer I/LExD; caspase-3,7 prefer DExD; caspase-6 prefers VExD; and caspase-2 prefers the pentapeptide substrate V/LDExD (Timmer and Salvesen 2007). Further studies suggested that the P1' residue (the residue after P1) is also important for substrate specificity. For example, Gly, Ala, Thr, Ser and Asn are preferred at P1' by caspase-3, while bulky or charged residues were poorly tolerated (Stennicke, Renatus et al. 2000).

The traditional understanding of caspase substrate specificity has been verified by many *in vitro* kinetic studies, however, it is vast oversimplified and dangerous to use for the prediction of protein substrates (Timmer and Salvesen 2007). The presence of the preferred cleavage site sequence is not sufficient for a protein substrate to be cut. The cleavage site must be properly presented to the enzyme. The interactions between enzymes and larger substrates are usually mediated by exosites distant from the catalytic site for many other regulatory proteases. Although additional interactions between caspase and substrate have been suggested, the mechanism still remains unrevealed. The crystal structure of p35 interacting with caspase-8 at a surface area distant from the active site may suggest a possible exosite (Fisher, Cruz et al. 1999).

In addition to exosites, the cleavage site sequences on caspase natural substrates also diverge from canonical sequences. For example, about 45% of reported caspase-3 natural substrates contain noncanonical cleavage site sequences (Fischer, Janicke et al. 2003). Thus, the preferred cleavage site sequence on a protein is neither sufficient nor necessary to be a real cleavage site *in vivo*. As the functions of caspases are better understood, more and more natural substrates of caspases are being discovered. As discussed in a previous review (Earnshaw, Martins et al. 1999), natural caspase substrates in apoptotic cells contain a wide range of proteins, including but not limited to cytoskeleton proteins in cytoplasm, structural proteins in the nucleus, proteins involved in DNA metabolism and repair, proteins involved in cell cycle regulation and proliferation, proteins involved in signal transduction pathways, proteins related to human genetic diseases, and proteins in the apoptotic regulation pathway. The identification of a real substrate target is a challenging process with experiments from *in vitro* kinetic assay using recombinant caspases to the test of cleavage site mutant transgenically introduced into whole animals. Indeed only a small portion of the reported events are found to be significant *in vivo*. Therefore, we must be aware that the published results have different levels of reliability and they should not be considered equally for the analysis of caspase substrates.

The studies of caspase substrate specificity have been progressing for more than a decade. They began from investigating the catalytic rates of different substrates. The preliminary understanding helped the discovery of various natural caspase substrates, which in turn provides valuable information for refining our understanding of caspase substrate specificity. In the second round of investigation, we are studying the substrate binding from the structural point of view. We hope that our study will provide valuable information to improve knowledge of caspase substrate specificity.

1.4. Caspase Inhibitors

Caspases are involved in apoptosis and inflammatory responses. They have been hot targets of drug design for many years because dysregulation of caspase activity leads to many severe diseases. Suppression of caspase activity is usually observed in cancer patients while increased caspase activity is associated with many neuronal degenerative diseases and autoimmune diseases, such as Parkinson's disease (Blandini, Sinforiani et al. 2006), Alzheimer's disease (Blandini, Sinforiani et al. 2006), Huntington's disease (Sanchez Mejia and Friedlander 2001), stroke (Schulz, Weller et al. 1999), and sepsis (Hotchkiss and Nicholson 2006). Although caspase is a very attractive drug target and many inhibitors have been developed, no compound has become an approved drug against caspase so far. Therefore, developing better caspase based inhibitors is necessary for future drug development.

Current caspase inhibitors can be divided into three groups, natural inhibitor proteins, substrate analog inhibitors, and non-peptide compounds. The most well known cellular caspase inhibitors are the inhibitors of apoptosis protein (IAP) family. They are characterized by their consensus N-terminal baculoviral inhibitory repeat (BIR) domains. XIAP is the most studied family member. It has three BIR domains among which BIR2 is shown to inhibit caspase-3 and 7 while BIR3 can inhibit caspase-9. The functions of other IAPs, however, still remain unclear (Eckelman, Salvesen et al. 2006).

Substrate analog inhibitors are usually peptides with ketone or aldehyde warheads. These peptide inhibitors are not attractive for therapeutic development due to lack of specificity, inhibiting all caspases and other cysteine proteases, or poor cell penetration and metabolic stability. The mode of binding of peptide inhibitors has been studied in crystal structures, and many structure based designs have aimed to reduce the peptidyl features (Weber, Fang et al.

2008). This strategy has led to the identification of several effective substitutions for P2-P4 residues, as summarized in an earlier review (O'Brien and Lee 2004). Recently, a number of new peptidomimetic caspase inhibitors were reported. One example is MX1153 (compound **1** in Fig. 1.4) (Wang, Guan et al. 2005). Although this inhibitor has fluoromethylketone as a warhead, it showed >5000 fold IC_{50} for caspase-3 over several other cysteine proteases and serine proteases, and reduced apoptosis in a mouse model. Another potent caspase-3 inhibitor M867 was synthesized recently (Han, Giroux et al. 2005) (compound **2** in Fig 1.4). The P1 Asp was retained in this compound, while the P2-P4 backbone was replaced by an amino pyrazinone template. The hydrophilic furazanmethylamino showed the best binding affinity at P4, and at P2, the hydrophobic *t*-Bu displayed the optimal whole cell activity, consistent with the substrate specificity of caspase-3. At the P1' position, an N-methyl-N-hexyl group dramatically improved cell permeability. Overall, M867 was highly effective in the *in vivo* anti-apoptosis tests (IC_{50} 20-1200 nM). Interestingly, M867 was shown to be about 60-fold more effective against recombinant caspase-3 versus caspase-7, and may facilitate the development of inhibitors selective for caspase-3 rather than caspase-7, which has been a difficult issue for many years. Other novel compounds are the aza-peptide inhibitors (Ekici, Li et al. 2006; Ganesan, Jelakovic et al. 2006). These inhibitors carry P2-P4 peptide residues, but the P1 Asp is modified to aza-Asp (C_{α} was replaced by N). Ketone and aldehyde were substituted with Michael acceptors as warheads in these compounds, resulting in better selectivity against caspases over clan CA and other clan CD proteases, such as legumain, clostripain, and gingipain K (Ekici, Li et al. 2006). The interactions between these aza-Asp inhibitors and caspases are similar to those of conventional tetrapeptide inhibitors at the P1-P4 positions, although the extension at P1' allows the exploration of the S1' site of caspases. Crystal structures showed that the S1' pocket of

caspase-3 is surrounded by four loops forming an internal space of 900 Å³ (Ekici, Li et al. 2006). Thr166 and Tyr204 are located at one side of this pocket and Phe128 and Met61 are on the other side (Fig. 1.4B). Analysis of kinetic data and crystal structures suggested that a bulky hydrophobic group can penetrate deeply into the S1' pocket of caspase-3 forming favorable interactions. Consequently, this inhibitor had enhanced binding affinities compared with compounds with small P1' groups. In contrast, caspase-8 has a relatively small S1' site created by Leu254, Ile257 and Tyr324 (Fig. 1.4B), and prefers small non-polar groups such as ethyl. Enzyme inhibition assays suggested that caspase-7 has the same S1' selectivity as caspase-3, perhaps due to their highly homologous sequences; esters were the best candidates at P1' for caspase-2; ethyl was the optimal P1' group for caspase-9 and 10; while there was no distinct S1' selectivity observed for caspase-6 (Ekici, Li et al. 2006).

Parallel to the modification of peptidic inhibitors, potential caspase inhibitors have been discovered by searching in the available chemical libraries. 5-nitroisatin was first discovered as a nonpeptide caspase-3 inhibitor by using a highthroughput screen in the SmithKline Beecham compound collection (Lee, Long et al. 2000). The modification of this compound led to the development of a series of potent isatin analog inhibitors of caspase-3/7, including compound **3** in Fig. 1.5A with appK_i values of 1.2 nM and 6 nM for caspase-3 and caspase-7, respectively (Lee, Long et al. 2001). Kinetic data and crystallographic analysis on compound **5** in the complex with caspase-3 indicated that a reversible covalent, tetrahedral adduct was formed between the isatin carbonyl and the active site cysteine (Lee, Long et al. 2000). Unlike most peptidomimetic inhibitors, the isatin inhibitors lack a P1 Asp and the caspase S1 pocket is thus empty (Lee, Long et al. 2000) Fig. 1.5C. The pyrrolidine rings interact with caspase-3/7 primarily in the hydrophobic S2 pocket formed by three aromatic residues: tyrosine, tryptophan,

and phenylalanine (Fig. 1.5D). Because caspases differ in their S2 pockets, this exclusive binding profile of these inhibitors results in 1000-fold higher selectivity for caspase-3/7 versus many other caspases (1, 2, 4, 6, and 8). This unique feature enables them to specifically inhibit caspase-3 and 7, which has been unachievable by peptidomimetic inhibitors to date. Surprisingly, a recent biochemical and biophysical study showed that compound **3** abolished caspase-3 activity upon binding to only one active site of the homodimer (Aulabaugh, Kapoor et al. 2007). This interesting phenomenon, however, was not observed in the previous crystal structure (Lee, Long et al. 2000). Despite the ambiguities in the inhibition mechanism of isatin sulfonamide inhibitors, their effectiveness in reducing apoptosis has been clearly demonstrated (Lee, Long et al. 2000; Chapman, Magee et al. 2002). N1-substituted 5-pyrrolidinylsulfonyl isatins have been shown to inhibit caspase processing in apoptotic endothelial cells (Kopka, Faust et al. 2006). For instance, the compound (s)-(+)-5- [1-(2-methoxymethylpyrrolidine) sulfonyl]isatin (MMPSI) reduced myocardial ischemic injury in an isolated rabbit heart model (Chapman, Magee et al. 2002). The isatin sulfonamide compounds, which are potent and selective reversible nonpeptide caspase inhibitors, have become good templates for structure-based drug design. Recently, modifications at three different regions of an isatin sulfonamide molecule were evaluated for caspase-3/7 inhibition (Fig. 1.5A) (Chu, Zhang et al. 2005). In region I, neither substitution of the para-position nor replacement with a pyridine ring significantly changed the inhibitory potency. Nevertheless, a 20 fold lower potency was observed when the phenoxyethyl moiety in this region was removed, indicating the hydrophobic group in region I possibly binds in the non-polar S1' site of caspase-3/7. In region II, the replacement of the pyrrolidine ring with an azetidine ring did not improve the potency. In region III, the substitution of the benzene ring with a pyridine ring increased the potency by 3-4 times against caspase-3/7, possibly due to the

introduction of a hydrophilic interaction between the phenoxymethyl moiety and the S3 binding pocket of caspase-3/7 (Chu, Zhang et al. 2005). An earlier study claimed that the *in vivo* applications of isatin sulfonamide inhibitors may be limited because of the highly reactive nature of their ketone carbonyl groups toward nucleophiles (Lee, Long et al. 2000). A possible solution for this issue has been achieved recently by a new class of isatin sulfonamide analog compounds called isatin Michael acceptors (IMAs) (Chu, Rothfuss et al. 2007). They contain Michael addition acceptors as their warheads, which can be attacked by the thiol nucleophile of cysteine (Fig. 1.5B). The binding affinities of certain IMAs, such as compound **4** in Fig. 1.5B, have reached the nanomolar range against caspase-3/7. Interestingly, all IMAs showed 10-fold higher potency for caspase-6 relative to their isatin sulfonamide analogs (Chu, Rothfuss et al. 2007). This potency may provide a valuable hint for the design of caspase-6 inhibitors, which has lagged far behind the development of caspase-3/7 inhibitors.

1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]quinoline (Fig. 1.6B) was recently discovered as a small molecule inhibitor of caspase-3/7. It represents a novel scaffold for non-peptide inhibitors of executioner caspases. A number of derivatives varying at the R1, R2 and R3 positions were synthesized and evaluated in caspase-3 *in vitro* inhibition assays (Kravchenko, Kysil et al. 2005). Compound **6** with the combination of a morpholinesulfonyl moiety at position R1 and 1,3,5-trimethyl-1*H*-pyrazol-4-yl group at R2 was the lead compound among those tested when R3 was fixed to -CH₃ (Fig. 1.6A) (Kravchenko, Kysil et al. 2005). It showed an IC₅₀ value of 4 nM against caspase-3, comparable to the commercial tetrapeptide inhibitor Ac-DEVD-CHO (IC₅₀=3.1 nM under the same conditions). The best R3 moiety was examined by fixing R1 and R2 groups. Methyl group and phenyl group displayed the best IC₅₀ values of 23nM and 27nM, respectively (Kravchenko, Kuzovkova et al. 2005). Overall, the best combination of R1-R3 is

shown in compound **8** in Fig. 1.6A, however, its activity remains to be determined. The inhibitory mechanism of this class of inhibitors has been suggested to arise from nucleophilic attack of the caspase catalytic cysteine on the 'phthalimide' carbonyls (Kravchenko, Kysil et al. 2005) (Fig. 1.6B). However, the mode of binding of these inhibitors has not been confirmed due to the absence of structural evidence. In addition, recent kinetic studies indicated that these compounds were noncompetitive reversible inhibitors for caspase-3 (Kravchenko, Kuzovkova et al. 2005), implying that they do not bind in the active site groove of caspase-3 or interact with the catalytic cysteine. Possibly these compounds bind in another region of caspase-3, such as the allosteric binding site. Further insights into the mode of binding of these inhibitors could lead to the discovery of new inhibitory mechanisms. Another subset of this class of inhibitors is the isoquinoline-1,3,4-trione (compound **9**) derivatives (Chen, Zhang et al. 2006). The best compound (compound **10**) showed a comparable potency against caspase-3 to that of the peptidic inhibitor Ac-DEVD-CHO. Although, most of these compounds demonstrated better activities against caspase-3/7 over other caspases, the differences were not sufficient to ensure selectivity. However, some isoquinoline-1,3,4-trione derivatives specifically inactivated caspase-1 (Ma, Zhang et al. 2007). Importantly, this class of compounds showed 50-1000 fold higher specificities against caspases relative to other cysteine and serine proteases (Chen, Zhang et al. 2006). The compounds were validated in PC21 cells and primary neuronal cells, where they effectively attenuated apoptosis induced by the amyloid-beta protein that is released by caspase-3 in the progression of Alzheimer's disease (Zhang, Zhang et al. 2006). Enzyme kinetic studies indicated that these compounds performed as noncompetitive inhibitors (Chen, Zhang et al. 2006) and they bound to caspases irreversibly in a slow-binding manner (Zhang, Zhang et al. 2006). The binding mode of these compounds is unknown. Therefore, the future development of these

promising quinoline derivatives would benefit greatly from structural analysis of the caspase-inhibitor complexes.

Despite substantial efforts, relatively few compounds targeting caspases are currently in clinical trials. The caspase inhibitors VX-765 (Fig. 1.7 compound **11**), VX-740, and MX-1013 were reviewed previously (Callus and Vaux 2007). Caspase-1 inhibitor VX-756 is in Phase II clinical trials for the treatment of inflammatory diseases (Fischer and Schulze-Osthoff 2005). VX-799, a small molecule pan caspase inhibitor developed by Vertex/Serono for septic organ failure, is in Phase I clinical trials. The pan caspase inhibitor IDN-6734 developed for the treatment of acute myocardial infarction is also in Phase I clinical trials (Fischer and Schulze-Osthoff 2005). IDN-6556 (Fig. 1.7 compound **12**) is an oxamyl dipeptide pancaspase inhibitor (O'Brien and Lee 2004) with promising results in Phase I clinical trials (Callus and Vaux 2007) and in Phase II clinical trials in patients with liver transplants (Baskin-Bey, Washburn et al. 2007). Also, IDN-6556 is promising as a potential drug for chronic hepatitis C (Pockros, Schiff et al. 2007).

Information from crystal structures of caspase complexes has proven valuable in the design of more selective and potent caspase inhibitors such as the aza-peptides with Michael acceptors, and the non-peptide isatin derivatives, and for the design of SMAC mimics to activate caspases. In parallel, structure-based screening of chemical libraries has identified novel compounds to control caspase activity like the isatin derivative inhibitors and the activator embelin. Overall, the development of pharmacological agents to control caspase mediated cell death has greatly benefited from structural studies, and several compounds are in clinical trials or preclinical development for treatment of various diseases.

1.5. Protein Crystallography and Caspase Structures

Determining protein 3D structure can help the understanding of protein functions and thus is a critical component of proteomics. Two methods are currently widely used for determining protein structures, X-ray protein crystallography and nuclear magnetic resonance (NMR). The basics of X-ray crystallography were established by a number of scientists in the early twentieth century, but this technique was not used on investigating protein structures until 1950s'. By first solving the structure of sperm whale myoglobin using X-ray crystallography, Max Perutz and Sir John Cowdery Kendrew were awarded the Nobel Prize in Chemistry in 1962. In the following half century, protein crystallography technique has been significantly improved in respect to both accuracy and efficiency. Today protein crystallography has become an almost routine technique in biological studies. Compared with NMR, crystallography has no limitation on the size of macromolecules, thus it is becoming more and more popular. According to the protein data bank (www.pdb.org), 52025 structures of biological macromolecules have been determined by X-ray crystallography by Sep-29-2009.

Protein crystal is a solid material where protein molecules are arranged in an orderly repeating pattern extending in all three spatial dimensions. When an X-ray beam strikes a crystal, it will be diffracted into many specific directions. By analyzing the diffraction pattern, a crystallographer can determine the positions of protein atoms in the crystal. The basis of protein crystallography is the Bragg's law ($n\lambda = 2 d \sin\theta$) whose principle is shown in the Figure 1.8b. Crystal is formed by highly ordered repeats of unit cells (Figure 1.8a). We can consider the atoms 1 and 1' in the figure 1.8b as equivalent atoms in two repeated unit cells. When an X-ray beam strikes on these two atoms, there will be two reflections (a and b) generated by the reflection planes (m and m'). Because X-rays are a type of electromagnetic wave, the amplitudes

of a and b will be amplified if they are in the same phase when they reach the detector. In contrast, their amplitudes will be canceled out if they are in opposite phases. According to Bragg's law, if the difference ($2l$ in the figure 1.8b) between the pathlengths of a and b equals $n\lambda$ (n is any integer, λ refers to the wave length) then a and b will have the same phase when they reach the detector. This means we can find the right angle θ by simply rotating the crystal because λ and d are constants for a particular experiment. Since every equivalent atom l in each unit cell will generate a reflection, once the X-rays strike the crystal from the right angle θ , the amplitudes of all these reflections will be summed up on the detector and thus form one bright spot. On the other hand, reflections generated by nonequivalent atoms will not form a spot because their amplitudes will be canceled out due to different phases. The intensity of a spot is proportional to the square of the amplitude of the diffracted wave. The intensities together with the phase angles can be transformed by Fourier Transformation into 3D visualizations of electron surfaces of atoms, called electron density map. As a result, the position of each atom can be determined.

Protein crystallography, although useful, is a challenging technique because several bottlenecks limit its success rate. For example, growing diffraction quality crystals, including previous protein expression and purification, finding appropriate cryoprotectant, and phasing are all very challenging tasks. The entire process of protein crystallography includes the following steps: crystallization, data collection, phasing, modeling, and refinement.

Crystallization is the first and the most difficult step. Normally, proteins can form into crystals in very particular conditions depending on protein concentration, pH, salt, temperature, buffer, precipitant, and additives. Crystallization conditions of different proteins normally vary a lot from one to another. Even single residue mutation or replacement of a small ligand can

dramatically change the crystallization condition. Therefore a standard crystallization protocol does not exist, and one successful crystallization usually comes out of hundreds even thousands of failures. Crystallization techniques fall into several categories, vapor diffusion, batch, and dialysis. The vapor diffusion method is the most widely used method. Taking hanging drop method as an example, the principle of vapor diffusion is illustrated in Figure 1.9a. The sealed system contains reservoir solution in a small well covered by a glass coverslip. The reservoir solution consists of buffer, precipitant, and sometimes additives. A small amount (1-2 μ l) of protein solution is mixed with equal amount of reservoir solution and suspended as a droplet underneath the coverslip. Because the precipitant concentration in the droplet is lower than that in the reservoir solution, water in the droplet will slowly evaporate and migrate into the reservoir until the system equilibrates. As a result, protein concentration in the droplet will gradually increase. If the final concentrations of protein and precipitant are optimal, protein crystals will be formed (Rhodes 2000). A picture of protein crystals is shown in the Figure 1.9b.

The second step is data collection. Because low temperature can increase the stability of molecules and thus enhance diffraction quality, X-ray diffraction data are currently collected with the help of liquid nitrogen (boiling point -196°C). Therefore, the protein crystal needs to be mounted on to a fine glass capillary loop with the reservoir liquid and frozen in the liquid nitrogen before the diffraction experiment. The freezing step requires the protection of cryoprotectant, which is some chemical agent that prevents the formation of both crack and ice in the protein crystal during the flash freezing in liquid nitrogen. Regular cryoprotectants include glycerol, small molecular weight polyethylene glycol, sugar, and some oils. Because the type and concentration of cryoprotectant and the appropriate procedure for freezing crystals vary from crystal to crystal, the cryoprotection is also a challenging step. During the data collection, the

crystal is placed in front of an X-ray detector. Following the Bragg's law, an X-ray beam will be diffracted into many particular directions after striking the crystal and thus form spots on the detector. The detector will record the diffraction pattern as one diffraction frame. The crystal is rotating during the data collection. In order to determine the 3D structure of the protein, a large number of frames must be taken at different angles. Because long-term X-ray radiation can damage the alignment of crystal cells, the aim of data collection is to accomplish the highest resolution and completeness within the shortest time.

To solve the protein structure, both diffraction density and phase are required. However, the phase cannot be measured directly in the experiment and it needs to be deduced indirectly. Actually, the phasing problem is the second serious bottleneck of crystallography. There are two basic approaches to solve the phasing problem: one is to perturb the structure and diffraction, and the other is to guess the phases. Two methods are widely used for the first approach: multiple isomorphous replacement and multiple wavelength anomalous dispersion. The isomorphous replacement is a classic phasing method. It uses a crystal nearly identical to the one being studied, except that there are several atoms replaced or added. Usually heavy atoms are co-crystallized or soaked into the crystal where they can bind tightly with the protein. These heavy atoms can perturb the diffractions and their positions are likely to be deduced. From positions of heavy atoms, phases of other protein atoms can be deduced. Because replacing or adding atoms can disturb structures of nearby protein atoms, the isomorphism is never perfect. The multiple wavelength anomalous dispersion (MAD) then becomes a substitution. In this method, only one crystal is studied. It contains atoms called anomalous scatterers, such as selenium. The degree to which the anomalous scatterers perturb the diffraction pattern can be changed by using different X-ray wavelengths. In this way, the position of anomalous scatterers can be deduced. This

method is more accurate than the isomorphous replacement because all the experiments are done on one crystal. In contrast to deduction, phase can be solved more efficiently by guessing sometimes, especially when the protein structure has been solved in another crystal form, or the protein sequence is similar to another protein (identity > 25%) with known structure. In this case, the known structure can be used as a template to guess the phase of the current crystal by using rotation and translation functions (Blow 2002).

When the phase problem is solved, the diffraction data can be transformed into electron density map by using Fourier transform (Sherwood 1976):

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

According to Fourier transform, the electron density function $\rho(x, y, z)$ can be determined by the volume of the unit cell V , structure factor F_{hkl} , and the phase. The structure factor is calculated by the following equation:

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

In the equation, the parameter f_j refers to scattering factor of a particular atom j .

Based on the electron density map, a protein structural model can be built. Usually the main chain atoms of the protein are first fitted into the electron density map. Side chain atoms are then added one by one according the protein sequence. Building structural model is much more straightforward for molecular replacement because most atoms would have been positioned after the phasing. Normally, high resolution data can provide more detailed and confident information to the protein conformations. Roughly speaking, the overall shape of a protein can be seen with the data of resolution 5.5 Å; main chain structures can be determined with

resolution of 3.5 Å; side chain conformations can be determined with resolution of 2.5 Å; and positions of single atoms can be determined with resolution of 1.5 Å or better. Since this step is relatively straightforward, some software has made it automatic, such as ArpWarp (Lamzin and Wilson 1993) and Coot (Emsley and Cowtan 2004).

The last and most tedious step is the model refinement. In this step, the conformation of each individual residue needs to be examined. If the model structure does not fit the electron density perfectly, then it needs to be adjusted as close to the density map as possible. Subsequently, new electron density map will be calculated based on adjusted model. This process needs to be repeated many times until the discrepancy R between the X-ray structure factors calculated for the model structure $|F_{cal}|$ and the observed intensities $|F_{obs}|$ drops to a

reasonable level ($R = \frac{\sum_h ||F_{obs}| - |F_{cal}||}{\sum_h |F_{obs}|} < 0.25$). Usually a protein has one to several hundreds of

residues, therefore the refinement process can last days to weeks.

3D structures of six caspases family members have been determined using X-ray crystallography, including caspase-1, 2, 3, 7, 8, and 9 (Yan and Shi 2005). The structures of other caspases still remain unclear. All the identified mature caspase structures share similar overall structure. In most caspase structures, a heterotetramer is formed by two caspase molecules, each of which contains a big subunit p20 and a small subunit p10. The active site cysteine is surrounded by loops, forming the substrate binding pocket. The structures of six caspase family members are shown in the Figure 1.10.

1.6. Protein Structure Prediction and Potential Functions

The number of proteins discovered in different species is growing very fast because of the development of new techniques for investigating biological mechanisms. The determination of protein structures, however, lags behind due to the efficiency limit of current experimental

techniques for structural determination. Fortunately, it is believed that proteins always fold into their 3D structures by following certain physical and chemical rules. Roughly speaking, a protein molecule always folds into the most stable conformation in a certain environment, which means the internal energy of a protein molecule is the lowest in its native conformation. Therefore, the protein structure can be predicted by simply comparing the internal energies of all possible conformations. With the help of modern computers, protein structure prediction, or structural modeling, has made great contributions to the understanding of protein structures and functions.

The internal energy of a protein molecule cannot be directly measured in the experiment. It is thus approximated by mathematical functions. One type of potential functions is derived from physical laws and they are called empirical potentials. Normally, the molecular system is represented in atomic level for molecular mechanics and electronic level for quantum mechanics. In the molecular mechanics, atoms are typically described as points with certain charges and masses. Their positions are represented by coordinates in Cartesian space (Erkoc 2001). The interactions between atoms include bonded interactions and nonbonded interactions. The bonded interaction is usually represented with a quadratic expansion around the equilibrium bond distance. Similar functions are used to represent other aspects of covalent geometry such as bond angles, planarity, chirality and torsional (around bond) rotations. The nonbonded interactions, including van der Waals interactions (attractive and repulsive force between molecules) and electrostatic interactions, are normally described by the Lennard-Jones potential and the Coulomb's law, respectively. Total potential energy of a protein molecule is determined by the summation of the bonded and non-bonded energies (Erkoc 2001). This type of potential function has large complexity and moderate accuracy.

Another type of potential energy function is named knowledge-based potentials (or statistical potentials). It differs from the empirical potentials in developing potential functions using statistical approaches but not physical rules. Based on the assumption that a protein molecule always folds into its native structure with the lowest potential energy in native conditions, computers can be trained to develop statistical potential functions to discriminate the native structure (experimentally determined protein structure) from structural models having different conformations (known as decoys). These potential functions, therefore, can be applied for evaluating the qualities of predicted protein structural models. More specifically, lower potential energy reflects better quality for a structural model, or vice versa. Many knowledge-based potentials have been developed. In some methods, a single point is used to represent the position of a residue. Different methods use different residue representation, such as Ca atom, C β atom, or side chain center of mass (SCM) (Zhang, Liu et al. 2004). The entire potential energy of a protein is calculated as the summation of the energy between each residue pair. These methods are thus residue level potentials, such as RAPDF-SCM (Samudrala and Moult 1998), KBP-SCM (Lu and Skolnick 2001), and DFIRE-SCM (Zhang, Liu et al. 2004). In order to achieve better accuracy, many methods calculate the potential energy based on each atom pair. These methods are named atomic level potentials, such as DFIRE-A (Zhou and Zhou 2002), and Atomic KBP (Lu and Skolnick 2001).

The knowledge-based potentials are developed by using machine learning techniques, where training data has large impact on the accuracy of the potential functions. The training data normally consists of decoy sets. Each decoy set contains one native structure of a protein and a certain amount (from less than one hundred to thousands) of distinct structural models of the same protein. Decoy structures are generated by using different algorithms. Commonly used

decoy sets include lattics_ssfit (Xia, Huang et al. 2000), lmds (Kearar and Levitt 2003), and Rosetta (Simons, Bonneau et al. 1999), etc. Hidden features of the native protein structure can be extracted by machine learning methods, such as genetic algorithm which is a widely used machine learning technique for searching exact or approximate solutions to optimization problems (Banzhaf, Nordin et al. 1998). The extracted hidden feature can be consequently integrated into potential functions to discriminate the native structure and closely similar structural models from other models. Although traditional knowledge-based potentials have better computational efficiency and cover hidden features of protein structures, the potential functions themselves are less physically meaningful and cannot be used for molecular mechanics. Therefore, this method still needs to be improved.

1.7. Objectives

Although a tremendous amount of effort has been put into the development of caspase-based drugs for over a decade, no caspase-based drug is yet available. Major challenges for drug development include cell permeability, metabolic stability, and target specificity. Since a large number of caspase inhibitors have been developed (Weber, Fang et al. 2008), a potential drug is likely to be obtained by modifying these caspase inhibitors. Structural information of substrate/inhibitor binding in caspases can provide valuable guidance in the design of the next generation of drug candidates. Therefore, this study aims to help the future caspase-3 based drug design from two aspects. First, the structural basis of caspase-3 substrate binding at P1-P5 will be investigated by analyzing crystal structures of caspase-3 bound with its substrate analog inhibitors. Second, novel knowledge-based potential functions will be developed in order to facilitate the prediction of protein structures and protein-ligand structures.

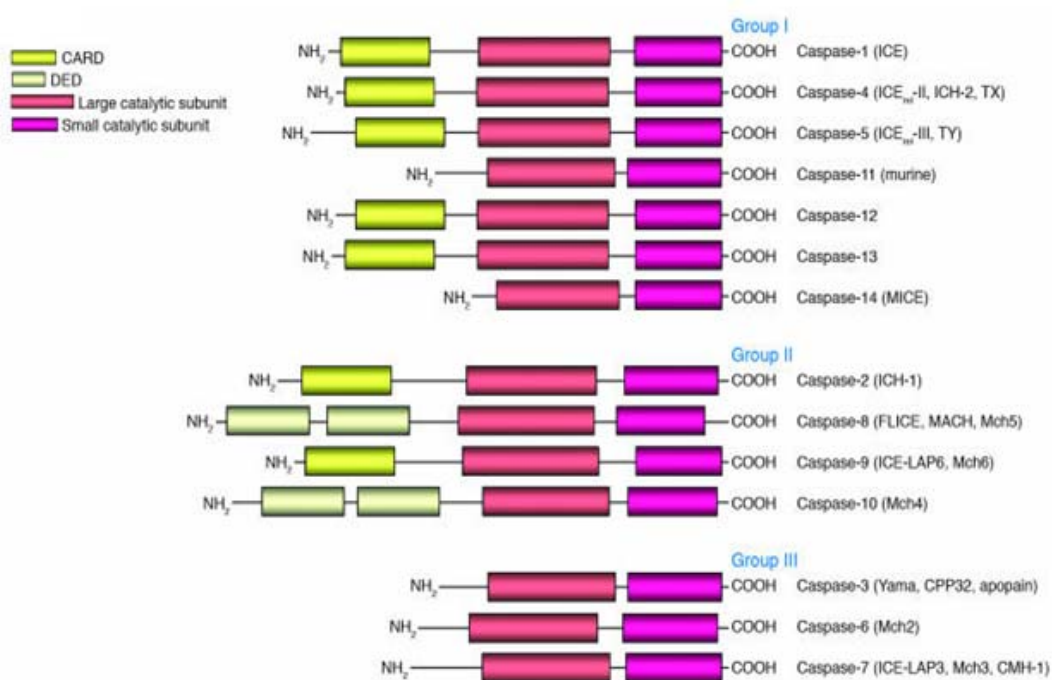


Figure 1.1. Grouping of 14 mammalian caspase family members (Lavrik, Golks et al. 2005). Colorful blocks refer to different domains.

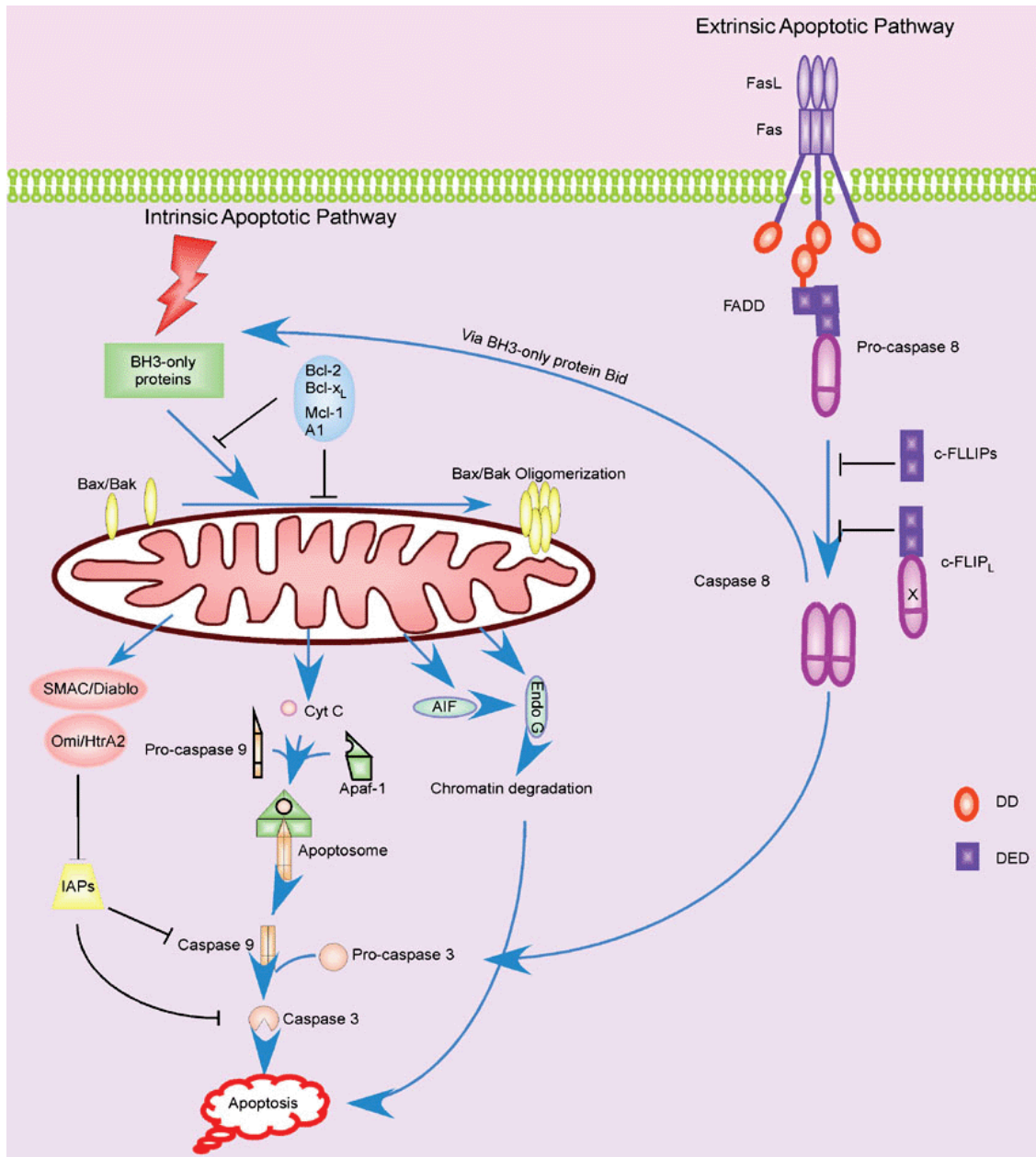
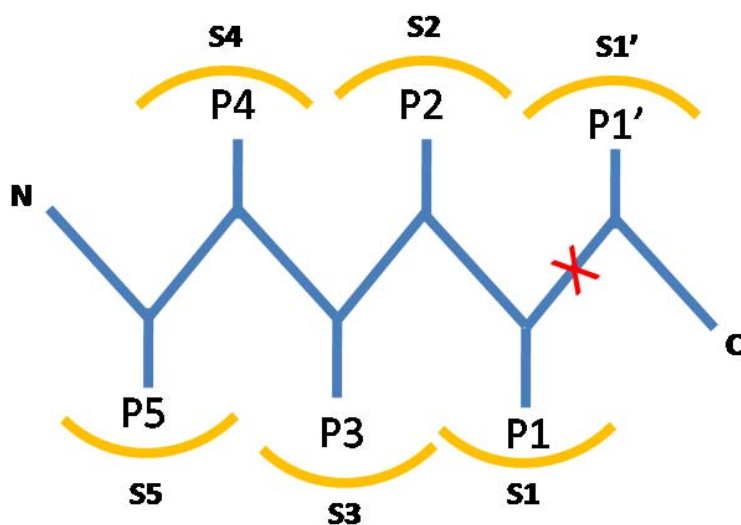


Figure 1.2. Extrinsic and intrinsic signaling pathway of apoptosis (Zhang, Hartig et al. 2005).



Caspases	P5	P4	P3	P2	P1
1,4,5,14		W/Y	E	X	D
8,9,10		I/L	E	X	D
3,7		D	E	X	D
6		V	E	X	D
2	V/L	D	E	X	D

Figure 1.3. Schematic presentation of caspase substrate binding site. The substrate specificities of caspases are shown in the Table where X refers to any amino acid.

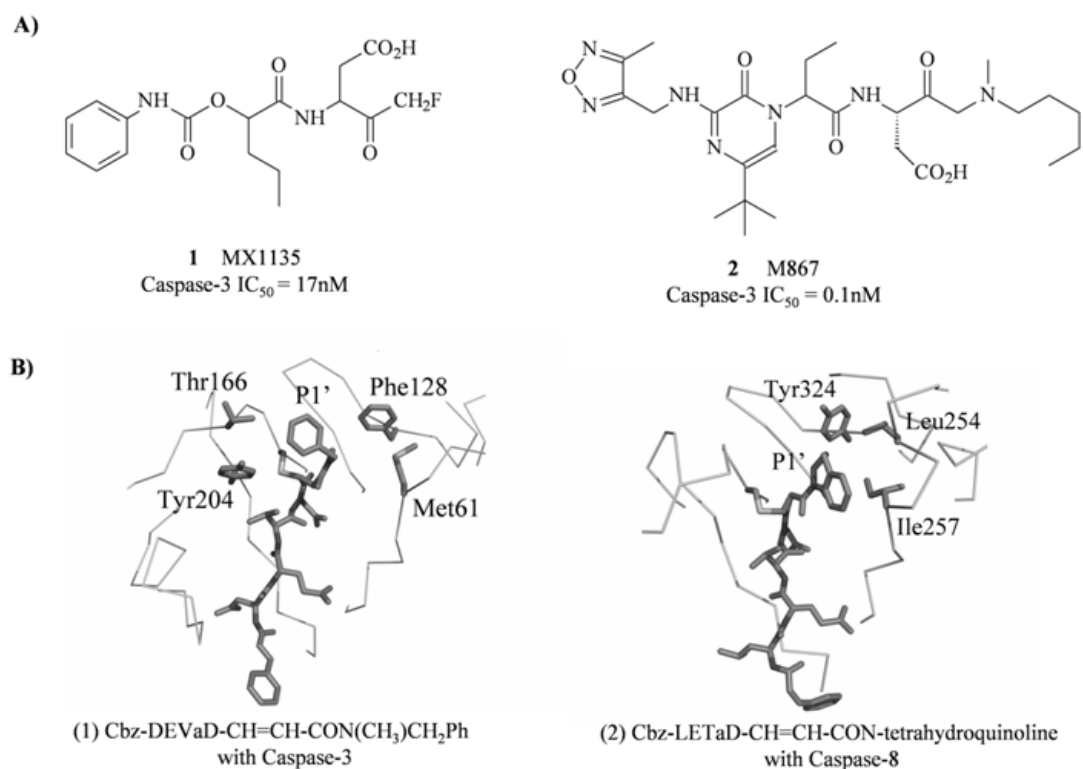


Figure 1.4. Peptidomimetic inhibitors. **A.** MX1135 and M867 are two examples of peptidomimetic inhibitors of caspase-3. **B.** The binding of aza-peptide inhibitors in caspase-3 (PDB code 2C2M) and caspase-8 (PDB code 2C2Z). Side chains of caspase S1' residues are shown in stick format and labeled with residue names. The aza-peptide Michael acceptor inhibitors are shown in stick format and their P1' groups are labeled.

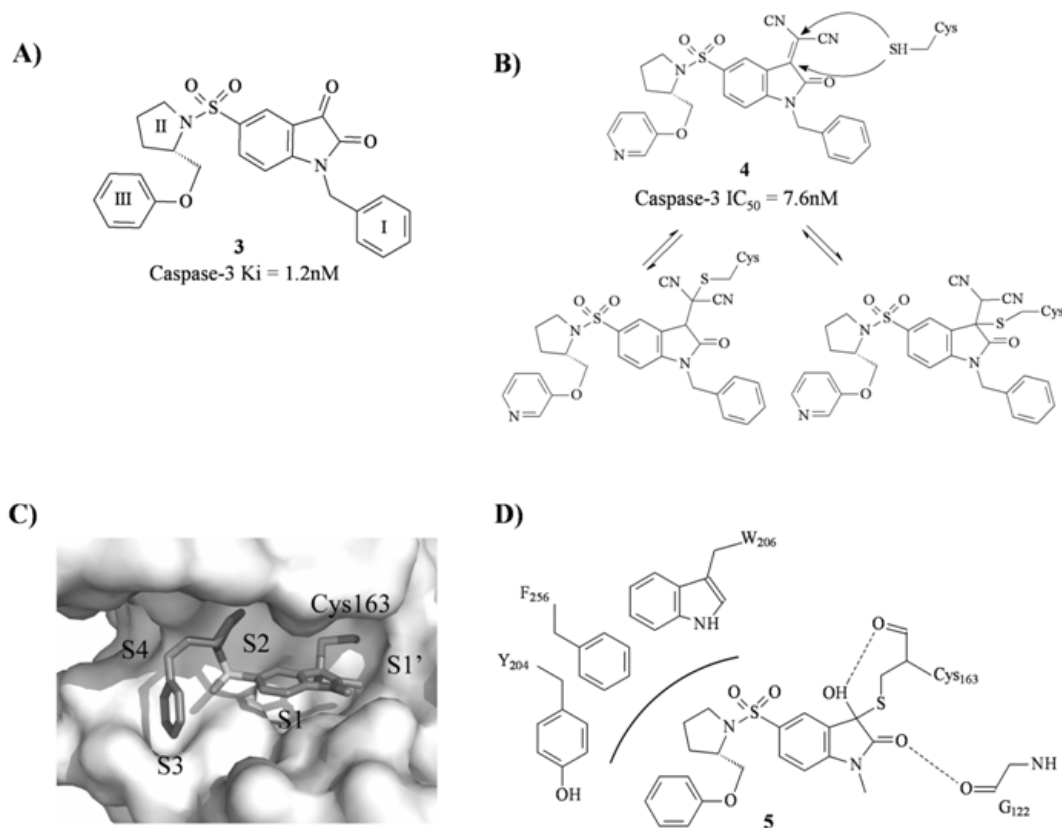


Figure 1.5. Isatin sulfonamide analog inhibitors. **A.** Isatin sulfonamide analog inhibitors of caspase-3/7. I, II, and III indicate three groups modified in later studies (12,49). **B.** IMA inhibitor, and proposed reaction mechanism with caspases. **C.** Crystal structure of 1GFW shows the binding mode of compound **5** in caspase-3. Caspase-3 active groove is in a surface representation. S1'~S4 binding pockets are labeled. **D.** Schematic representation of interactions between inhibitor **5** and caspase-3. Hydrogen interactions are shown in dashed lines. Caspase-3 S2 residues having non-polar interactions with inhibitor **5** are labeled.

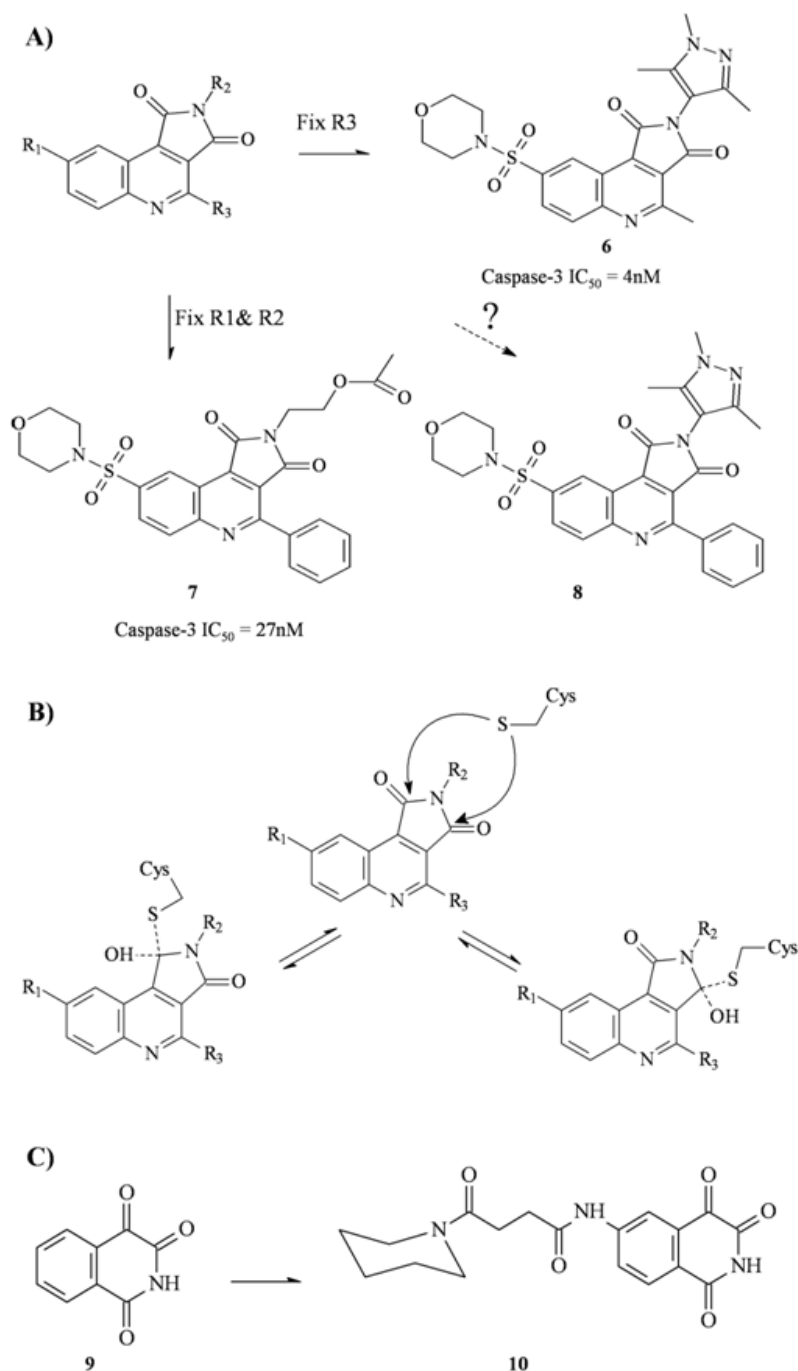


Figure 1.6. Quinoline derivative inhibitors. **A.** Modification of a general quinoline derivative inhibitor led to the discovery of two potent caspase-3 inhibitors, compounds **6** and **7**. Compound **8** suggests a putative better inhibitor. **B.** Putative inhibitory mechanism of quinoline inhibitors against caspases. **C.** Compound **10** is an example of isoquinoline inhibitor derived from compound **9**.

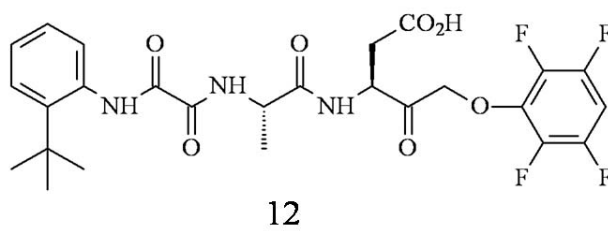
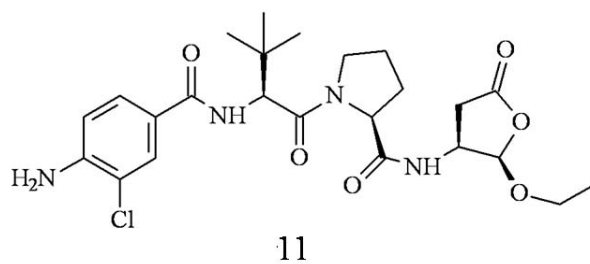


Figure 1.7. Structures of VX-765 and IDN-6556. Structure of VX-765 (11) and IDN-6556 (12) in clinical trials.

From <http://www.structmed.cimr.cam.ac.uk>

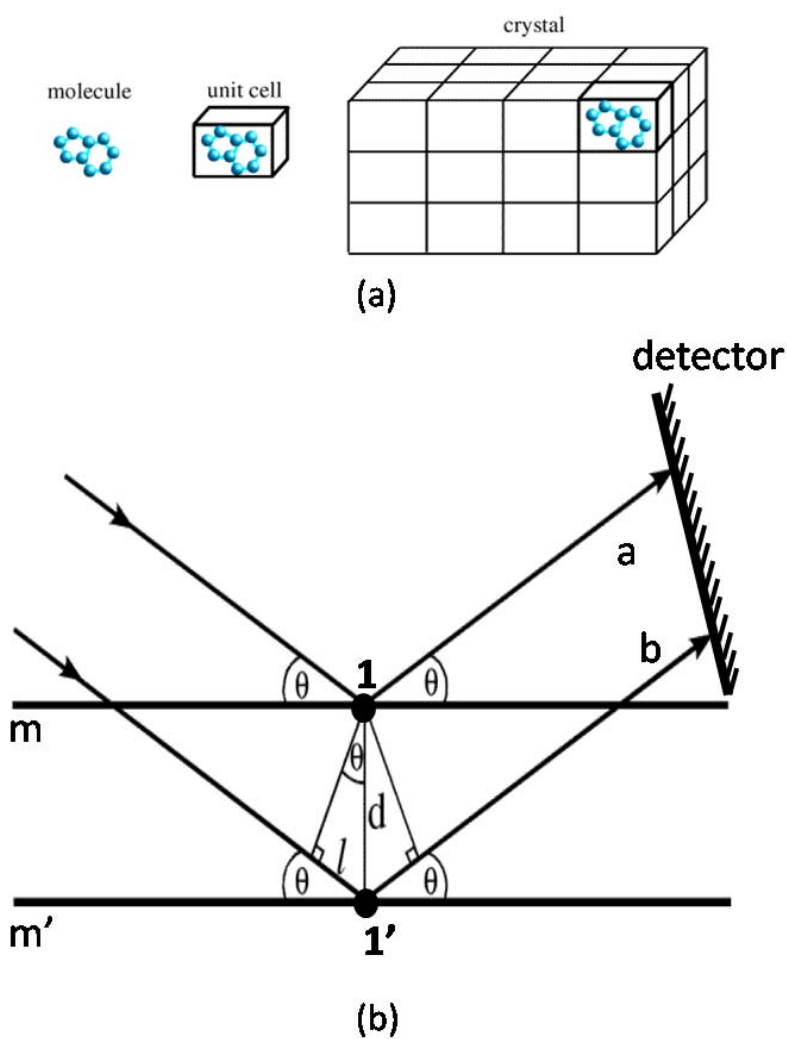
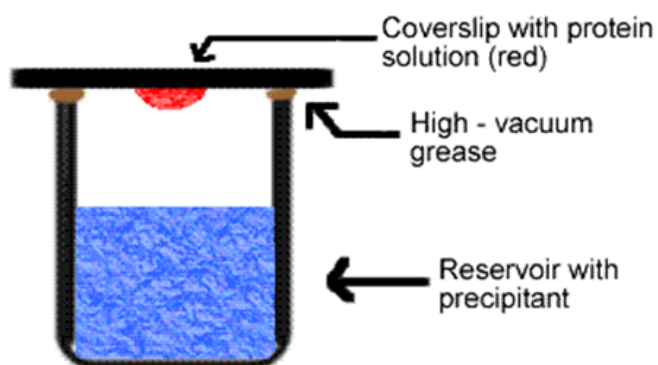
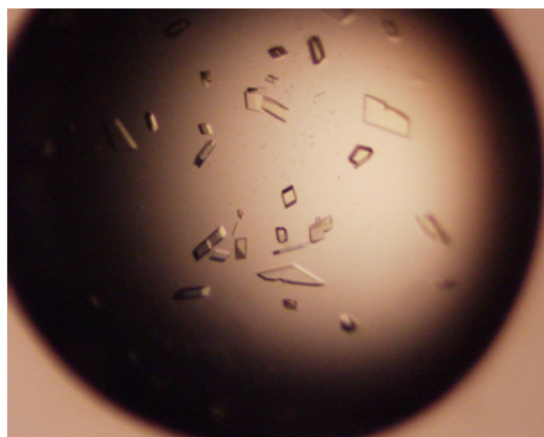


Figure 1.8. Crystal packing and the Bragg's law. (a) Packing of protein crystals. (b) Schematic presentation of the Bragg's law.



From <http://www.bio.davidson.edu/courses/MolBio/>

(a)



(b)

Figure 1.9. Hanging drop vapor diffusion system and crystal samples. (a) Hanging drop vapor diffusion system. (b) Protein crystals of caspase-3/DMQD complex under microscope.

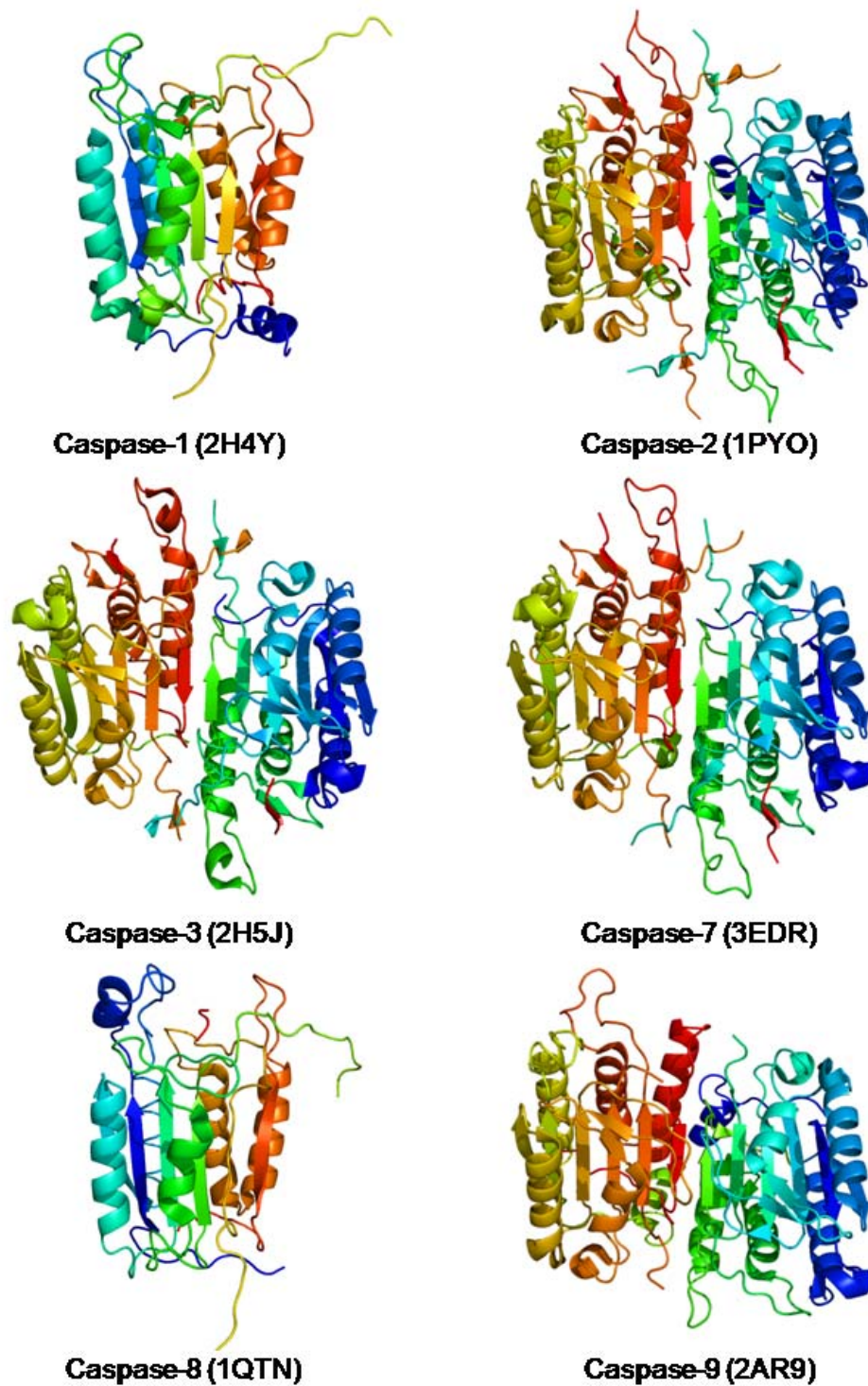


Figure 1.10. The list of caspase structures currently deposited in PDB. The structures are in ribbon representation and rainbow coloring.

2. CASPASE-3 BINDS DIVERSE P4 RESIDUES IN PEPTIDES AS REVEALED BY CRYSTALLOGRAPHY AND STRUCTURAL MODELING

2.1. 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 structure and role in the apoptotic and inflammatory pathways. The first group consists of inflammatory caspase-1, 4, 5, and 11 (Martinon, Holler et al. 2000). 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 identified as caspase substrates, as reviewed in (Fischer, Janicke et al. 2003; Timmer and Salvesen 2007).

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, heart failure and neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's diseases are associated with increased activities of caspases (Hartmann, Troadec et al. 2001; Hermel, Gafni et al. 2004; Tacconi, Perri et al. 2004). On the other hand, caspase activity is suppressed in cancer, autoimmune diseases and viral

infections (Vucic, Stennicke et al. 2000; Lakhani, Masud et al. 2006; Volkmann, Cornberg et al. 2006). 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 (O'Brien and Lee 2004; Callus and Vaux 2007).

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 (Thornberry, Rano et al. 1997; Lavrik, Golks et al. 2005). 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 (Lien, Pastor et al. 2004). A computer based virtual screen found a potent caspase-3 inhibitor Ac-DNLD-Cho with comparable inhibition to Ac-DEVD-Cho. Similarly, Ac-DFPD-Cho was found to be the best tetrapeptide inhibitor for caspase-7 (Yoshimori, Sakai et al. 2007). 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 (Schweizer, Briand et al. 2003; Fang, Boross et al. 2006). Since many caspase substrates are observed with diverse non-canonical sequences at their cleavage sites (Backes, Kuentzer et al. 2005), it is apparent that the substrate selectivity of caspases is still poorly defined.

Caspase-3 is one of the critical executioner caspases. Previously, we analyzed the S2, S3 and S5 pockets of caspase-3 using aldehyde peptidyl inhibitors Ac-DEVD-Cho, Ac-DMQD-Cho,

and Ac-VDVAD-Cho (17). Analysis of the P5 interaction in S5 was extended to the Ac-LDESD-Ac complex (Fu, Chumanevich et al. 2008). 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 (Stennicke, Renatus et al. 2000). Although P4 Asp is common, many other amino acids, including hydrophobic residues, can be found at the P4 position in the known protein substrates of caspase-3 (Fischer, Janicke et al. 2003). 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 development of inhibitors into drugs, 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. Previous structural studies have focused on peptide with P4 Asp. 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 molecular mechanics energy calculations on structural models of P4 residues in relation to enzyme inhibition data, and suggested that S4 pocket of caspase-3 can accommodate a variety of residues with different affinities.

2.2. Materials and Methods

2.2.1. 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 (Fang, Boross et al. 2006). Protein was concentrated to 4 mg/ml and stored at -80 °C. The purity was determined to be over 99% by SDS-PAGE.

2.2.2. Enzyme Kinetic Assays

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 (Fang, Boross et al. 2006). Briefly, 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. Then, substrate was added and reaction velocity was calculated from *p*-nitroanilide released by enzyme cleavage as measured at a wavelength of 405 nm using a Polarstar Optima microplate reader (BMG Labtechnologies, NC). 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 (Maibaum and Rich 1988). The enzyme concentration was determined by active site titration against Ac-DEVD-Cho using method described previously (Donepudi, Mac Sweeney et al. 2003). 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 300 nM of caspase-3 and a constant amount (2 μ M) of BocD after preincubation for times from 0 to 60 mins.

2.2.3. 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 μ l 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 (Otwinowski 1997). The structures were solved by molecular replacement with the program AmoRe (Navaza 1994). 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 (Brunger, Adams et al. 1998). The molecular graphics program O 8.0 (Jones, Zou et al. 1991) 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 electron density maps. Structural figures were made by Weblab viewer pro (Accelrys Inc., MA) and images of electron density map were obtained using Molscrip (Esnouf 1997; Esnouf 1999). 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.

2.2.4. 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 (Harrison 1993). The models were generated from the crystal structure of caspase-3/DEVD (Fang, Boross et al. 2006), 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 (Emsley and Cowtan 2004). The hydrogen atoms were built with the sp4 potential set using 15 cycles of the analytic model building algorithm in AMMP (Bagossi, Tozser et al. 1999; Harrison 1999), 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 (Sayle and Milner-White 1995) on Linux PCs. The experimental binding energy was calculated as $\Delta G = RT \ln K_i$ from the seven measured inhibition constants for correlation with the calculated interaction energies. The regression line from this correlation was used to predict K_i values for the peptides with other P4 residues from the peptide-caspase interaction energies calculated by the modeling program.

2.3. Results

2.3.1. 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 (Fischer, Janicke et al. 2003). Almost all amino acids can be found at least once at the P4 position as shown by the relative occurrence (Fig. 2.1). 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.

2.3.2. 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 K_m value determined as $62.5 \pm 2.5 \mu\text{M}$. 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 K_i values from 1.2 nM for DEVD to 12 nM for DMQD, whereas the second group of inhibitors had K_i values approximately a thousand times weaker. In the second group, IEPD was the most potent with the K_i value of 1.2 μM , while YVAD was the weakest with K_i of 10 μM . The inhibition values are comparable with those of previous studies with fewer inhibitors showing K_i values of 0.2-1nM for DEVD, 195 nM for IETD, 1.9 μM for WEHD and 10-12 μM for YVAD (Nicholson, Ali et al. 1995; Talanian, Quinlan et al. 1997; Garcia-Calvo, Peterson et al. 1998).

BocD is a pan-caspase irreversible inhibitor widely used for *in vivo* studies. Its inhibitory profile was compared with IEPD as shown in Fig. 2.2a. 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 (Qiu, Chen et al. 2005). 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 (Fig 2.2b). The time dependency suggests that BocD acts as a slow binding inhibitor.

2.3.3. 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, in order to study the

binding of hydrophobic P4 residues. The crystallographic data and refinement statistics are summarized in Table 2.1. The proteins 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 1.9-2.6 Å 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 $C\alpha$ atoms. Overall, the side chain atoms share similar conformations since the average all atom rmsd for the four structures is 0.7 Å, and the average for residues in the active site cavity is almost identical at 0.8 Å.

The three tetrapeptide inhibitors were bound in extended conformations in the S1-S4 substrate binding sites of caspase-3 (Fig. 2.3a). The $2Fo-Fc$ electron density map of inhibitor YVAD is shown in Figure 2.3b. 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 $C\alpha$ atoms superimposed very well with average rmsd of 0.27 Å for four $C\alpha$ 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 (Fig. 2.3c). 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) (Ganesan, Mittl et al. 2006), this class of inhibitors is conventionally considered to be irreversible caspase inhibitors in physiological conditions.

2.3.4. 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 2.4. 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 illustrates 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 (Fig. 2.4a) except that the methyl group extended into the pocket towards Gln161. The neutralized P1 group formed single hydrogen bond interaction, instead of ionic interaction, with the positively charged Arg64 and Arg207. The interaction with Gln161 was not observed in this structure (Fig. 2.4d). 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 (Fig. 2.5b). Compared with proline, histidine was not expected to be an attractive P2 residue due to its bulkier side chain. In the structure of caspase-3/WEHD, however, the P2 histidine was accommodated in the S2 pocket with its aromatic ring pointing into the solvent (Fig. 2.5c). All the side chain carbon atoms formed van der Waals interactions with the three hydrophobic S2 residues. Compared with P2 Val of canonical DEVD, the C α atom of His had moved a little further from the S2 residues, which was likely due to the larger size of the His 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, (Fig. 2.3c; 2.5d). 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 (Fang, Boross et al. 2006). 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 (Thornberry, Rano et al. 1997). 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 (Fig. 2.5e). 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.

2.3.5. 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 (Fig. 2.4a). 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 (Fig. 2.5f). Similar favorable van der Waals interactions between P4 Ile and two Trp residues were observed in the complex of caspase-7/IEPD (Agniswamy, Fang et al. 2007). 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 (Fig. 2.5g). 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, and these interactions also contribute to the binding affinity (Fig. 2.6).

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° (Fig. 2.5h). 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 (Fang, Boross et al. 2006). 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 (Fig. 2.4b). Moreover, the interaction of P4 Trp with Phe250 and Phe252 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 (Fang, Boross et al. 2006).

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 (Fig. 2.5i). The binding of glycerol reflects the polar 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 small conformational changes in the caspase-3 residues, similar to the observations for caspase-7 (Agniswamy, Fang et al. 2007). 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 changes. 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, alternatively, P5 residues like Val or Leu. The binding modes of P4 Trp and P4 Tyr in YVAD and WEDH, respectively, illustrate the two distinct pockets (Fig. 2.6). This structural analysis explains why some natural substrates of caspase-3 have hydrophobic P4 residues, such as Ile and Leu in CREB and DCC, respectively (Backes, Kuentzer et al. 2005).

2.3.6. 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 2.2). 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. Thus, 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 dramatically reduced 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. WEHD had the least favorable P4 residue, however, the favorable P3 Glu interaction still made a stronger inhibitor than YVAD. Moreover, the side chain of P3 Glu in WEHD formed one more hydrogen bond with Ser209 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 influence 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 was observed in the low concentration range. In the higher concentration range, BocD was a stronger inhibitor than WEHD and YVAD because it can irreversibly inactivate caspase-3.

2.3.7. 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 (Harrison 1993) 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

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 (Fig. 2.7).

The accuracy of the energy calculation was evaluated using the seven commercial caspase tetrapeptide inhibitors with K_i values determined in our caspase-3 activity assay (Table 2.2). To our knowledge, this is the most complete set of aldehyde tetrapeptide inhibitors of caspase-3 at present. As shown in the Fig 2.7a, the observed free energies (ΔG) derived from experimental K_i 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 the internal energy 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 solvation are assumed to be small or similar for the series of inhibitors. Therefore, the estimate for ΔU can be used to estimate trends in binding free energy ΔG for different peptides. The high correlation indicates that these 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 K_i values of inhibitors XEVD, which were derived from mapping the calculated interaction energies onto the regression line in Fig 2.8a, are shown in Fig 2.8b. The P4 Asp showed the highest affinity in the prediction, in agreement with experimental results. The inhibitor with P4 Glu was predicted to be 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 (Thornberry, Rano et al. 1997). The predicted trends in binding affinity showed general agreement with the residue ranking in our analysis of natural substrates (Fig 2.1), 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 to 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 (Fig. 2.9). 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 chain hydrogen bond interaction with caspase residues. Most of them interacted with the main chain atoms of Phe250 (Fig. 2.9). 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, the hydrophobic residues Pro, Ala, Cys, Val, Leu, Met and Phe interacted with S4 subsite through van der Waals interactions with distance cutoff of 3.8-4.2 Å. All of these residues interacted with Trp206. In addition, 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 acetyl group on the inhibitors formed hydrogen bond interactions with Ser209, as observed for crystal structures of caspase-3/IEPD and YVAD (Fig. 2.4). An exception was observed in the model with P4 Phe, where the acetyl interacted with Asn208 and Trp214 instead. 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 (Fig. 2.6).

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 K_i compared with others. Our results imply that the S4 subsite is capable of accommodating various P4 residues. Hydrophilic P4 residues are predicted to interact with F250 and N208, whereas hydrophobic P4 residues are predicted 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. This analysis explains why a large number of non-canonical substrate sequences have been found.

2.4. Discussion

Since the original work of Thornberry and Talanian (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997), 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 cellular 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. For example, in one recent study, 27 out of 59 natural caspase-3 substrates do not have Asp at P4 in their cleavage sites (Backes, Kuentzer et al. 2005). In fact, 10 of those 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 other than Asp at P4, and over a third of them are non-polar residues. It challenges the older understanding of substrate recognition of caspase-3 from earlier *in vitro* kinetic studies on peptides. Crystal structures of caspase-substrate complexes can shed light on 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) (Backes, Kuentzer et al. 2005). 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 one of the best P2 residues (Lien, Pastor et al. 2004). Furthermore, the pyrrolidine ring has been successfully utilized at P2 in the development of caspase-3/7 inhibitors, such as the isatin sulfonamide inhibitors (Lee, Long et al. 2000). Clearly, the S2 subsite of caspase-3 can accommodate both aliphatic residues and aromatic residues. At the P3 position, our analysis confirmed previous studies that various residues can bind in S3, 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 (Blanchard, Donepudi et al. 2000) and caspase-7 was shown to accommodate hydrophobic P4 residues such as Trp (Agniswamy, Fang et al. 2007). 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) (Backes, Kuentzer et al. 2005) underlined the natural ability 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 subsite, 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 (Agniswamy, Fang et al. 2007). In the current study, the same feature was 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 (Fang, Boross et al. 2006). The physiological function of hydrophobic S5 is reflected in the crystal structure of XIAP bound with caspase-3 (Riedl, Renatus et al. 2001), where Ile149 and Ile153 of XIAP interacted with S5 residues. In the absence of structural information for natural caspase substrates, the roles of Trp206 and Trp214 in the S4 subsite were not understood before. Nevertheless, their ability to interact 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 the S4 subsite and 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 better accommodated than aromatic residues in the inner pocket of S4, which may explain the relative frequency in the caspase cleavage sites. However, the P4 preferences will likely differ in the context of a full-length protein substrate rather than a short peptide. Protein substrates may form additional interactions at exosites, as described for caspase-7 (Agniswamy, Fang et al. 2007). Although the mechanism is not fully understood, 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 facilitate the development of new inhibitors and drugs. Current drug design for caspase-3 is mostly limited to the P1, P2 and P3 positions. Introduction of a hydrophobic P4 moiety would not only facilitate the binding, but also improve the cell permeability. Compound IDN-6556 is one such example (Pockros, Schiff et al. 2007). 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 (Irene T. Weber 2008). This oxamyl peptide compound showed inhibition of all caspases and is currently under phase II clinical trials in patients with liver transplants (Baskin-Bey, Washburn et al. 2007). The results from this study should be considered in the future development of therapeutic compounds.

Table 2.1. 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 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 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)
<I/σ(I)>	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

^aValues in parentheses are given for the highest resolution shell. ^b $R_{\text{sym}} = \frac{\sum_{\text{hkl}} |I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle|}{\sum_{\text{hkl}} I_{\text{hkl}}}$. ^c $R = \frac{\sum |F_{\text{obs}} - F_{\text{cal}}|}{\sum F_{\text{obs}}}$. ^d $R_{\text{free}} = \frac{\sum_{\text{test}} (|F_{\text{obs}}| - |F_{\text{cal}}|)^2}{\sum_{\text{test}} |F_{\text{obs}}|^2}$.

Table 2.2. 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	1200 ± 48
Ac-WEHD-Cho	4700 ± 190
Ac-YVAD-Cho	10200 ± 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: $K_i = (IC_{50} - 0.5[E]) / (1 + [S]/K_m)$, where [E], [S] and IC_{50} , respectively, correspond to active enzyme concentration, substrate concentration, and the inhibitor concentration needed to suppress half enzyme activity. The enzyme concentration was determined by active site titration.

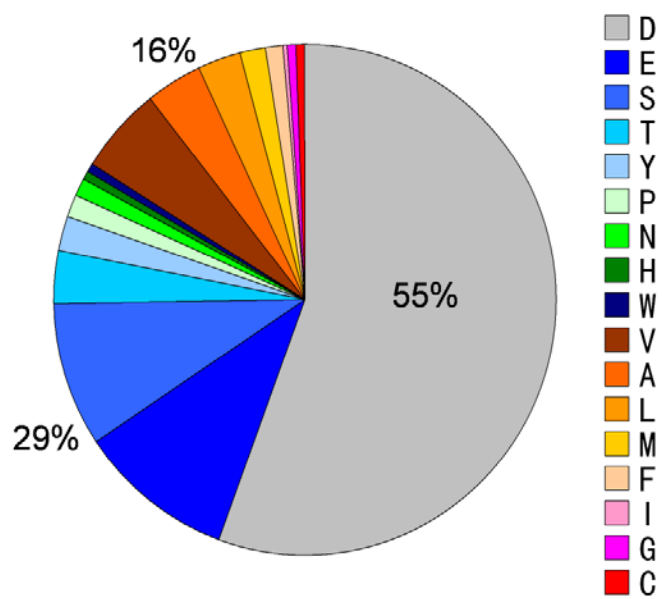


Figure 2.1. The occurrence of different amino acids at P4 in the cleavage sites of natural substrates of caspase-3. Different residues are represented in different colors.

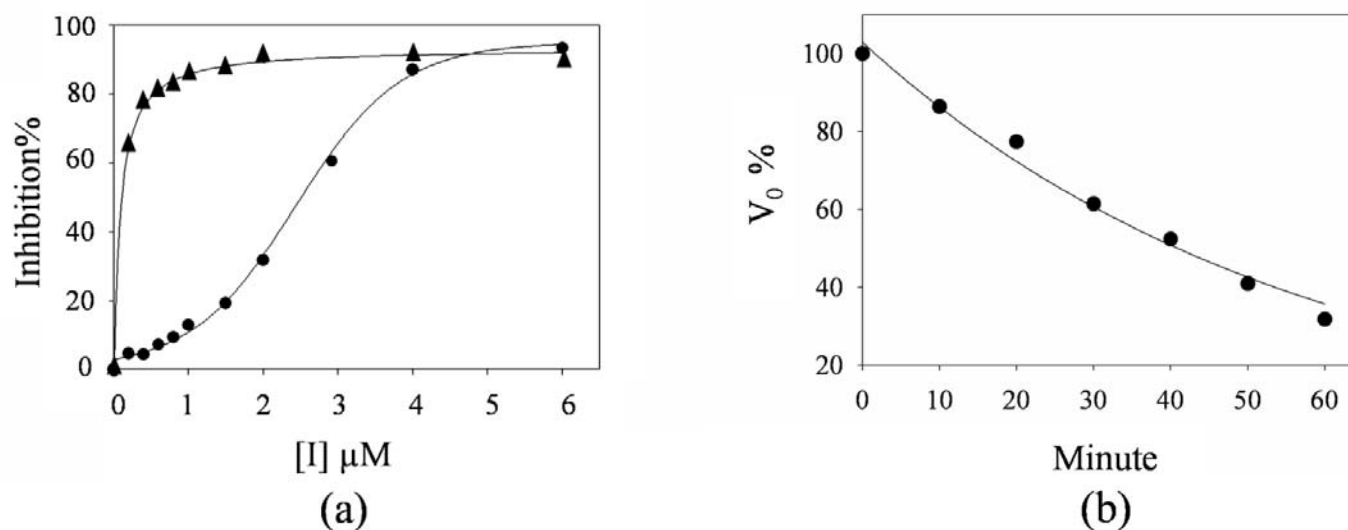


Figure 2.2. Inhibition characterization of caspase-3 inhibitors. (a) Inhibition profiles of caspase-3 inhibitors IEPD (▲) and BocD (●). The fractional inhibition is calculated as $V/V_0 \times 100\%$, where V and V_0 refer to velocities with and without inhibitor respectively. Caspase-3 was preincubated 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: $Inh = Inh_{max}[I]/(K_d + [I])$, while inhibition profile of BocD was fitted using sigmoidal equation: $Inh = Inh_{min} + (Inh_{max} - Inh_{min}) / (1 + 10^{(\log EC_{50} - [I])})$. Inh , $[I]$, and EC_{50} 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 2 μM 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: $V = V_{min} + a \cdot \exp(-bt)$, where V and t refer to initial reaction velocity and time, respectively.

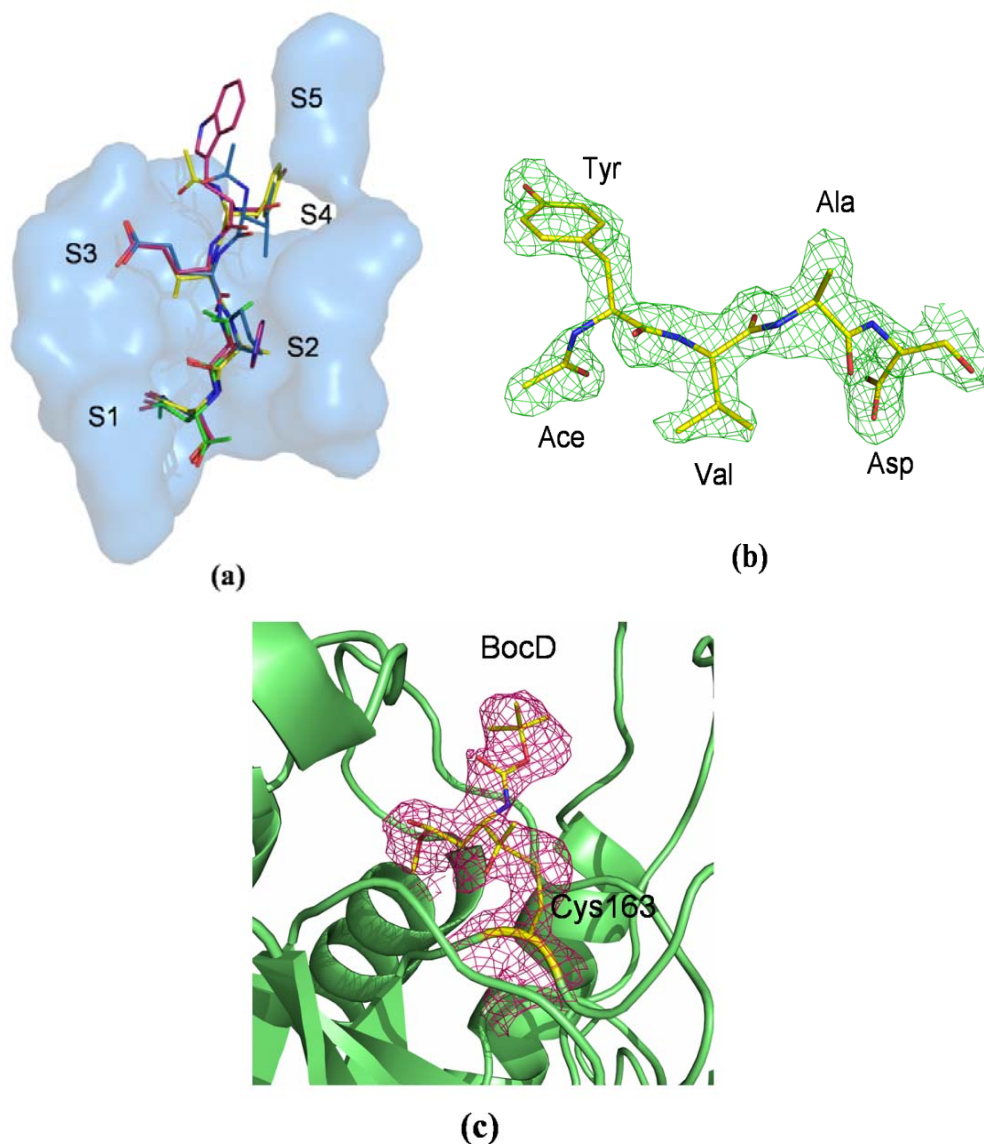


Figure 2.3. Binding conformations of inhibitors. (a) Superposition of four inhibitors in the substrate binding site of caspase-3. Inhibitors IEPD (blue), WEHD (magenta), YVAD (yellow), and BocD (green) are shown in stick representation. The binding surface of S1-S4 subsites is shown in pale blue. (b) $2F_o - F_c$ electron density map of inhibitor YVAD contoured at a level of 1.4σ . (c) $F_o - F_c$ 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.

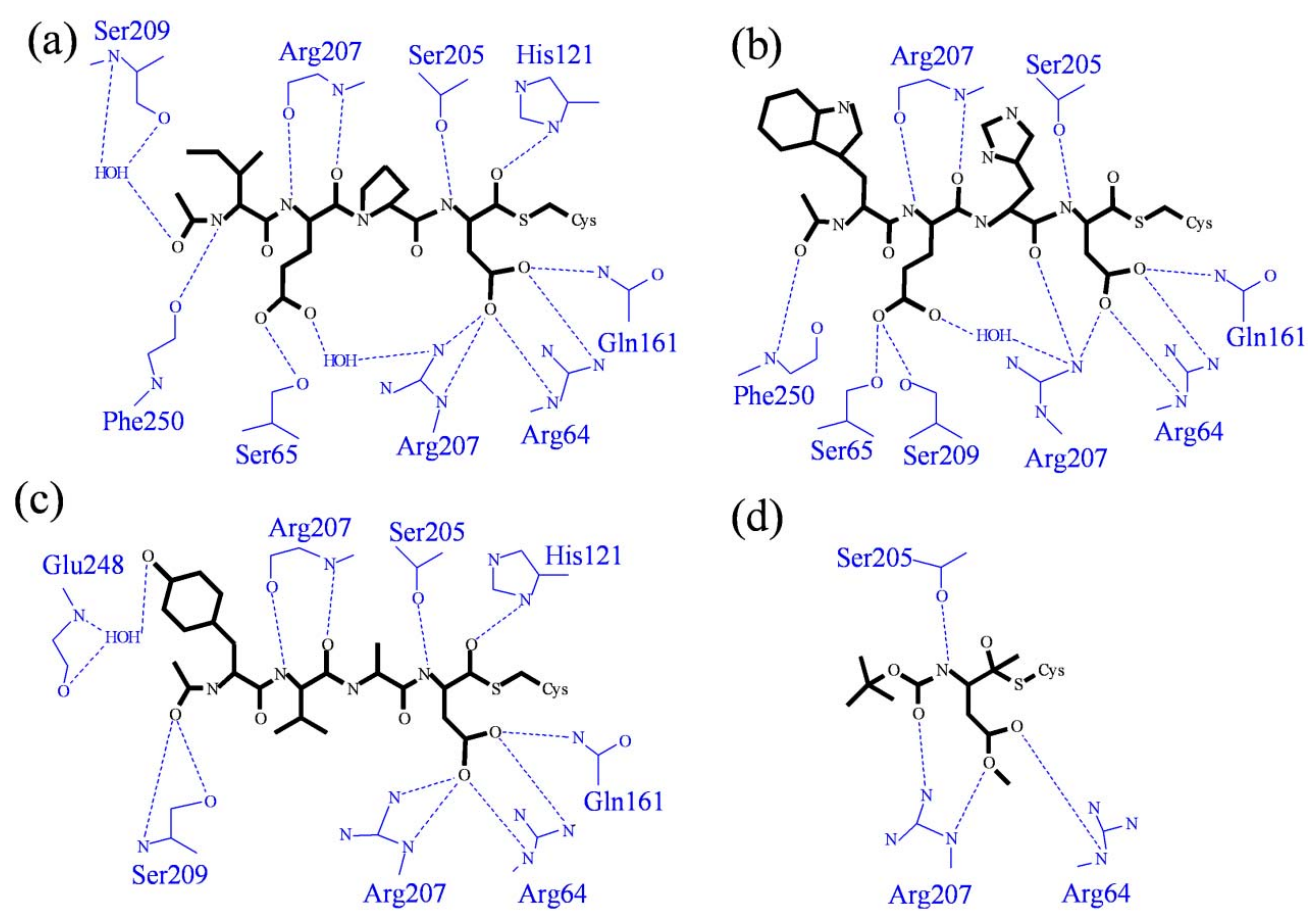


Figure 2.4. Schematic representation of hydrogen bond and ionic interactions between caspase-3 and the inhibitors. (a) IEPD, (b) WEHD, (c) YVAD, and (d) BocD. The protein residues are in blue and inhibitors are in black. Interactions are indicated in dash lines.

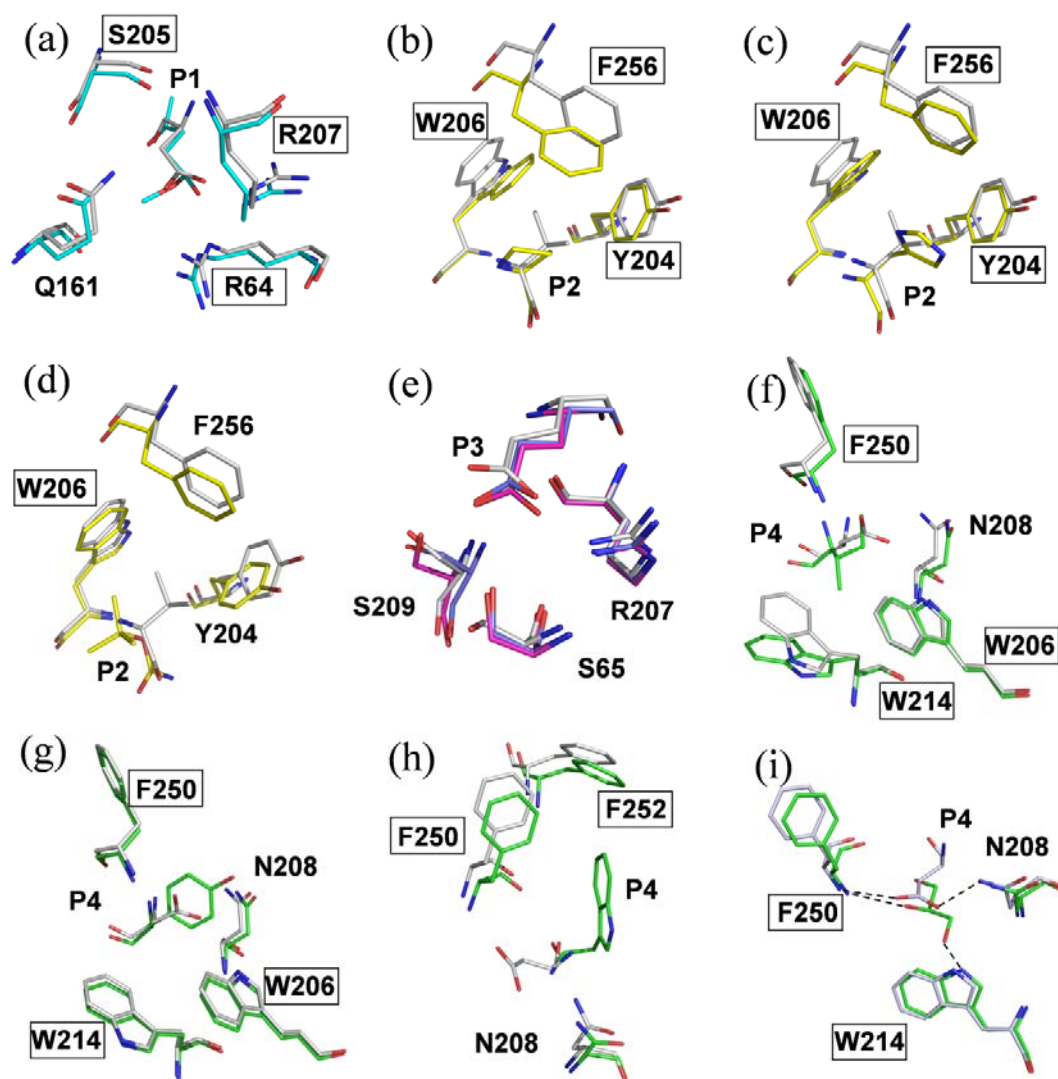


Figure 2.5. Comparison between subsites in four new complexes (color) and published structure of caspase-3/DEVD (grey) (2H5I). (a) The methylated P1 Asp of BocD and S1 residues. (b)-(d) P2 residues Pro, His, and Boc moiety, respectively are shown in S2 subsite. (e) P3 Glu in structure caspase-3/IEPD (purple) and caspase-3/WEHD (magenta). (f)-(h) The binding of P4 Ile, Tyr, and Trp, respectively, in S4 subsite. (i) The unbound S4 has glycerol in the BocD complex. Hydrogen bonds are represented by dashed lines. The boxed caspase-3 residues form hydrogen bonds or hydrophobic interactions with inhibitors.

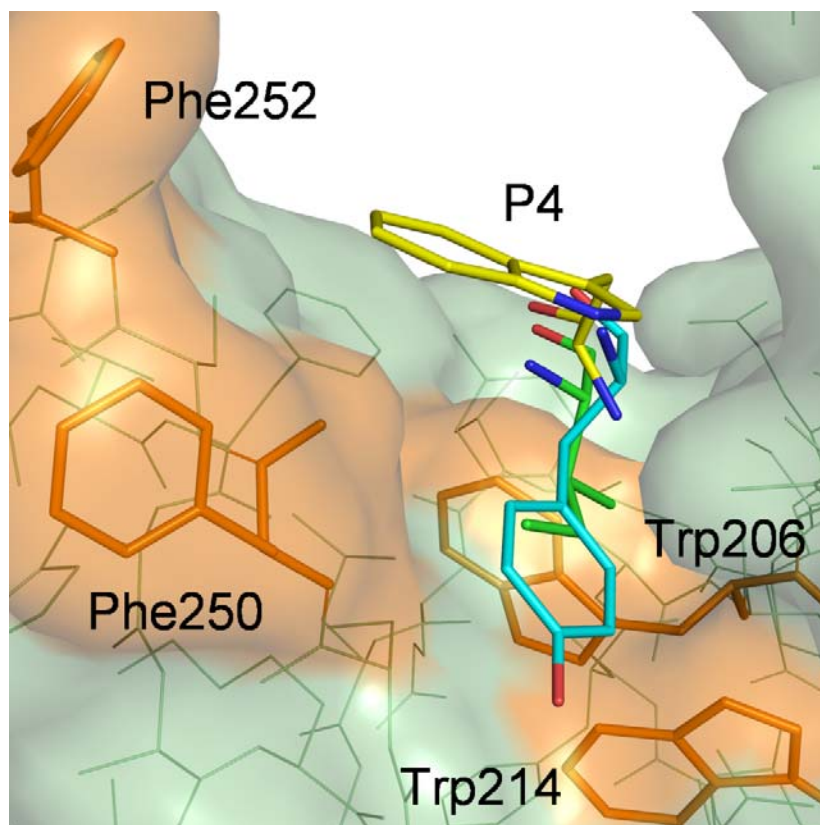


Figure 2.6. P4 binding site on caspase-3. P4 residues Ile (green), Tyr (cyan), and Trp (yellow) can bind in either the S5 pocket formed by Phe250 and 252 (orange) or another hydrophobic pocket formed by Trp206 and 214 (orange).

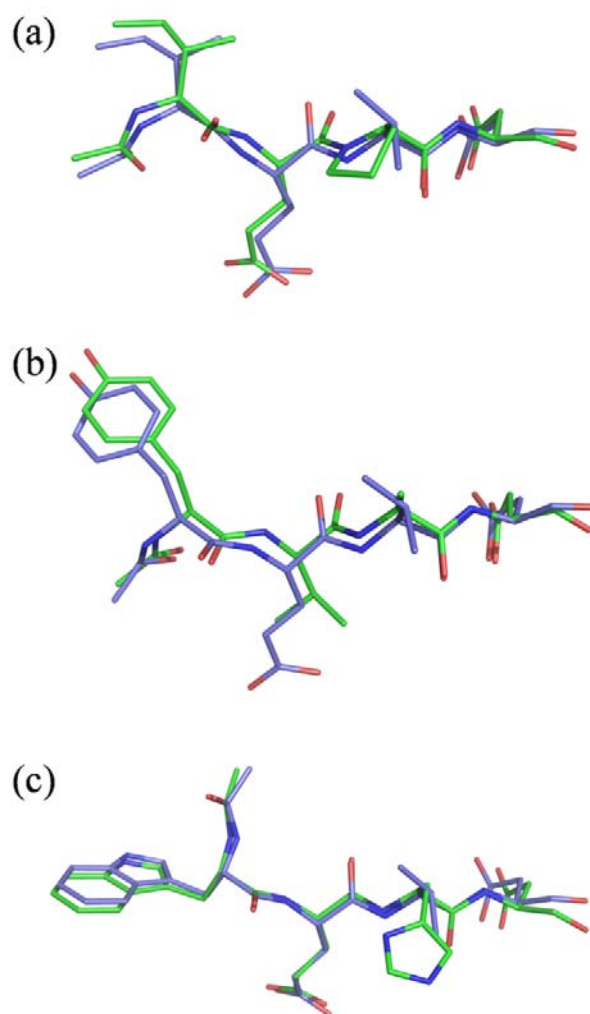


Figure 2.7. Superposition of inhibitors in the crystal structures (green) and structural models (purple). (a) IEPD/IEVD (b) YVAD/YEVD (c) WEHD/WEVD

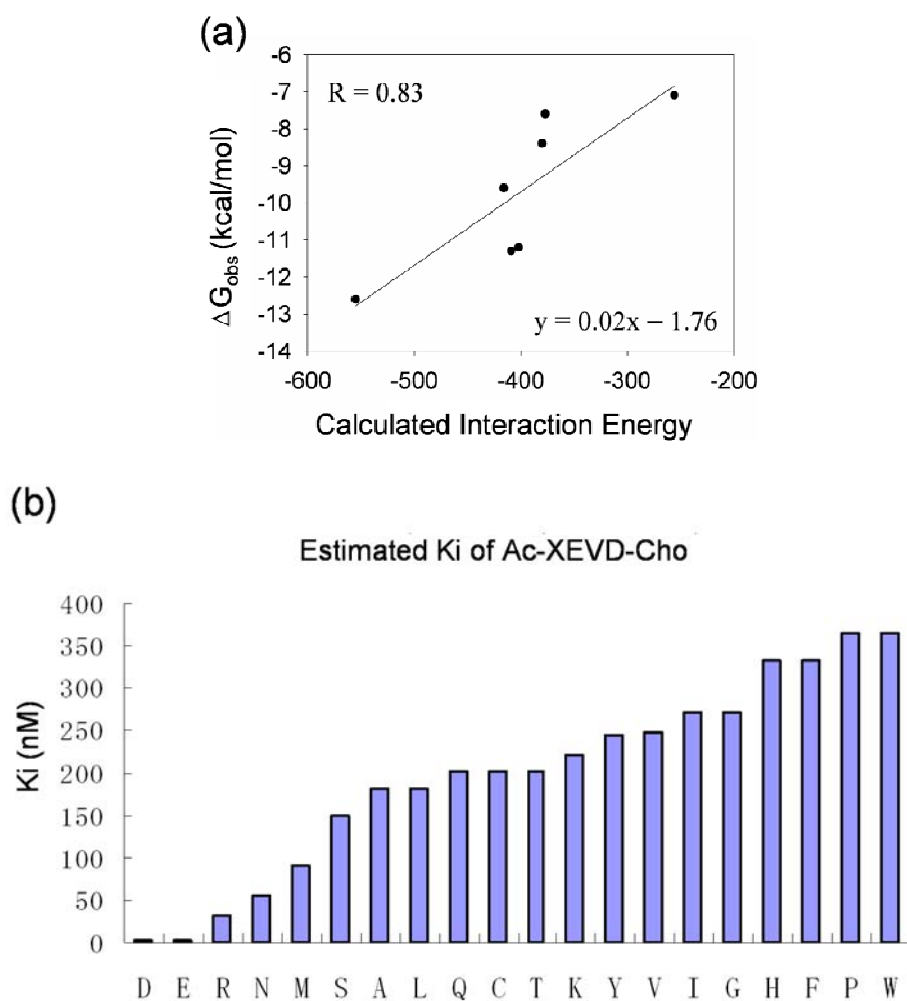


Figure 2.8. Predicted binding affinities for diverse P4 residues. (a) The correlation between experimental binding energies and calculated interaction energies. (b) Predicted K_i values of inhibitors Ac-XEVD-Cho.

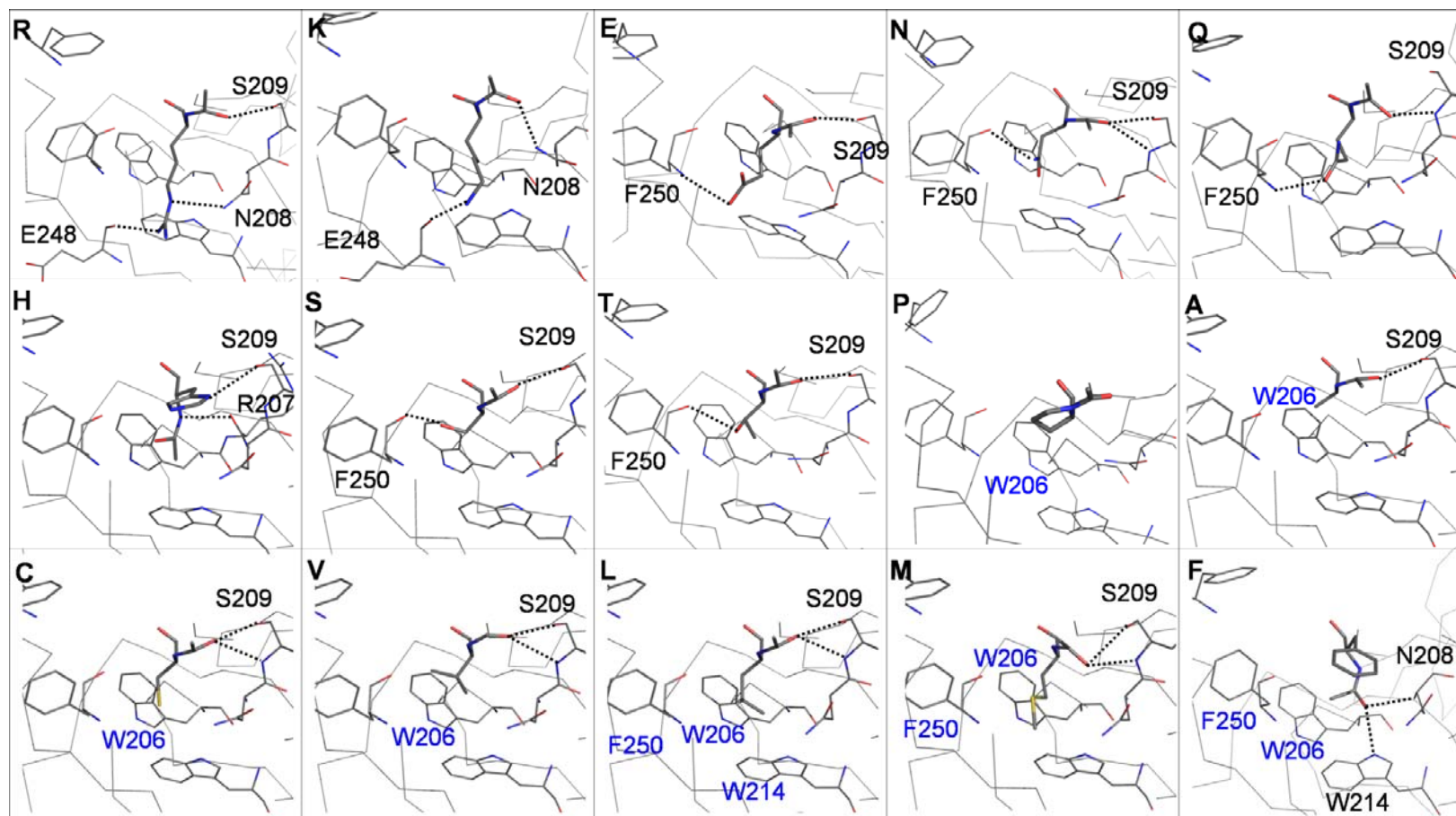


Figure 2.9. Predicted binding conformations of fifteen different P4 residues in caspase-3/XEVD. The P4 residue is located in the center in the stick representation, while caspase residues are shown as lines. The letter in the top left corner is the name of the P4 amino acid. Hydrogen bond interactions are indicated by dashed lines. Caspase residues involved in polar interactions are labeled in black, while those involved in van der Waals interactions are labeled in blue.

3. STRUCTURAL AND KINETIC ANALYSIS OF CASPASE-3 REVEALS ROLE FOR S5 BINDING POCKET IN SUBSTRATE RECOGNITION

3.1. Introduction

Caspases promote apoptosis by proteolytic cleavage of a number of downstream protein substrates (Alnemri, Livingston et al. 1996). Caspase activity is associated with a variety of diseases including neurodegenerative disorders, ischemic injury and cancers (Thompson 1995). Caspase-3 mediated apoptosis has a major role in neurodegenerative diseases (Charriaut-Marlangue 2004). Caspase-3 is activated in spinal cord injury and Alzheimer's disease, where it can cleave the amyloid-beta precursor protein and influence apoptosis of neurons (Kim, Pettingell et al. 1997; Gervais, Xu et al. 1999). Several peptidomimetic and nonpeptide caspase-3 inhibitors have been found to inhibit apoptosis (Lee, Long et al. 2000; Micale, Vairagoundar et al. 2004). Inhibitors of caspase-3 were shown to prevent neuronal loss in mouse models (Takuma, Tomiyama et al. 2004). On the other hand, the deficiency or suppression of caspases is a factor in the development of cancer and autoimmune diseases (Thompson 1995; Soengas, Alarcon et al. 1999), and tumor specific gene therapy based on caspase-3 has been explored (Yang, Cao et al. 2004). Moreover, abnormal heart development was found in knockout mice without caspase-3 and -7 (Lakhani, Masud et al. 2006). Consequently, caspase-3 is required for normal development and control cell death in many diseases (Philchenkov 2004).

Information on the molecular structure and substrate specificity of caspase-3 is valuable for understanding the development of disease and for the design of new therapies. The crystal structures are known for the catalytic domain of caspase-3, both unliganded and in complex with different peptidic, non-peptidic, or protein inhibitors (Fuentes-Prior and Salvesen 2004; Salvesen and Abrams 2004). The catalytic domain comprises a small (12 kD) and large (17 kD) subunit

arising from cleavage of the procaspase-3. The peptide substrates or inhibitors consisting of residues P4-P1', where the scissile peptide bond is between P1 and P1', are bound in pockets S4-S1' formed by the caspase. Generally, peptidic inhibitors contain aldehydes (Cho) or fluororomethylketones (fmk) that form a covalent link with the catalytic Cys 163, as shown in the caspase-3 complexes with peptide inhibitors Ac-DVAD-fmk (Mittl, Di Marco et al. 1997) and Ac-DEVD-Cho (Rotonda, Nicholson et al. 1996) (where Ac is the acetyl group).

All caspases have a stringent specificity for Asp at P1 in the substrate. Caspase-3 and -7 are effector caspases that recognize the canonical peptide sequence of DEVD and are involved in similar signaling pathways. In caspase-3 the P1 Asp is bound in a deep basic pocket formed by the conserved residues Arg 64, Arg 207 and Gln 161, and a peptide with Glu at P1 instead of Asp was hydrolyzed at a 20,000-fold lower k_{cat}/K_m (Stennicke, Renatus et al. 2000). Caspase-3 was shown to preferentially cleave the peptide bond after the optimal sequence of DE(V/I)D in a combinatorial peptide library (Thornberry, Rano et al. 1997). Other studies using peptides with substitutions at different positions suggested the preferred peptide sequence of DE(M/L)D-(S/G) (Talanian, Quinlan et al. 1997; Stennicke, Renatus et al. 2000). However, peptide library searches with a substrate-phage selected DLVD with 170% faster hydrolysis than the canonical DEVD peptide (Lien, Pastor et al. 2004). These results suggest that the substrate preferences for P2 and P3 are not fully understood. Although most caspase family members have been shown to recognize a tetrapeptide of P4-P1, the P5 position was determined to be essential for the substrate selectivity of caspase-2 (Talanian, Quinlan et al. 1997; Schweizer, Briand et al. 2003). Few studies, however, have been performed to investigate the P5 position in other caspases. Clearly, the substrate or inhibitor specificities of caspases are not fully understood.

In order to improve our knowledge of the molecular basis for substrate specificity the crystal structures were determined of complexes of recombinant human caspase-3 with the peptide analogs Ac-DEVD-Cho, Ac-DMQD-Cho and Ac-VDVAD-Cho. The complex with Ac-DEVD was obtained at the significantly higher resolution of 1.70 Å compared to 2.5 Å for the previously reported structure (Rotonda, Nicholson et al. 1996). These complexes explore the specificity for positions P2 and P3 in the peptide substrates, and examine the effect of adding the P5 residue. The analysis of these new structures with peptide analog inhibitors will help in the design of new caspase-3 inhibitors as potential therapeutic agents for neurodegenerative diseases.

3.2. Materials and Methods

3.2.1. Protein Expression and Purification

The cloned full length human caspase-3 cDNA was expressed as previously described (Stennicke and Salvesen 1997). Cells were harvested and resuspended in lysis buffer (20 mM Tris-HCl, 5 mM imidazole, 25 mM NaCl, 5 mM dithiothreitol, 0.1 mg/ml lysozyme and 0.1% Triton X-100, pH 7.5). Cell crude extract, obtained by sonication and centrifugation, was filtered with 0.2µm filters and loaded onto the nickel affinity column (HisTrap™ HP, Amersham, NJ). Caspase-3 was eluted by a gradient of 20-1000 mM imidazole. Imidazole in the protein solution was subsequently removed by dialysis against 25 mM Tris-HCl, 20 mM NaCl, 10mM dithiothreitol, pH 7.5. The partially purified caspase-3 was then loaded onto the anion exchange column (High Q Cartridge, Bio-Rad, CA) and eluted by 20-1000 mM NaCl gradient. Salt concentration was reduced to 20mM by buffer exchange using ultrafiltration. Further purification was performed using a gel-filtration column (Superdex 75, Amersham, NJ) with the buffer 20 mM Tris-HCl, 20 mM NaCl, 10 mM dithiothreitol, pH 7.5. Finally, caspase-3 was concentrated to 4mg/ml and stored at -80°C. The purity was determined to be over 99% by SDS-PAGE.

3.2.2. Enzyme Kinetic Assays

Caspase-3 activity was determined using the colorimetric caspase-3 substrate Ac-DEVD-pNA (Biomol, PA), where Ac is the acetyl group and pNA is *p*-nitroanilide. 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 (Howard, Kostura et al. 1991). The catalytic constants k_{cat} of three caspase-3 substrates: Ac-DVAD-pNA (GenScript, NJ), Ac-VDVAD-pNA (Axxora, CA) and Ac-LDVAD-pNA (GenScript, NJ) were determined by using the equation $k_{cat}=V_{max}/[E]$, where $[E]$ values were measured by active site titration during K_i determination as described below.

Caspase-3 substrate analog inhibitors Ac-DEVD-Cho (Biomol, PA), Ac-DMQD-Cho (Calbiochem, CA), and Ac-VDVAD-Cho (Axxora, CA) bind to caspase-3 through thioether bonds between the aldehyde (–CHO) group of the inhibitor and the mercapto (–SH) group of Cys 163 on the protein. According to the vendor instructions, the binding of these inhibitors is reversible although it is strong. Therefore, they were treated as reversible tight binding inhibitors. For the measurement of inhibition constant K_i , caspase-3 was preincubated with its substrate analog 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 (Maibaum and Rich 1988).

3.2.3. Crystallographic Analysis

The inhibitors were dissolved in dimethylsulfoxide. Caspase-3 was incubated at room temperature with the inhibitor at 20-fold molar excess. Crystallization was performed by the hanging-drop vapor diffusion method. 1 μ l of protein solution (4 mg/ml) was mixed with an equal volume of 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 15% 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 (Otwinowski 1997). The structures were solved by molecular replacement with the program AmoRe (Navaza 1994). The caspase-3 structure (1NME) (Erlanson, Lam et al. 2003) in the space group I222 was used as the initial model for caspase-3/DEVD, while the structure 1CP3 (Mittl, Di Marco et al. 1997) was used as the initial model for the two complexes in the P2₁ space group. The structures of caspase-3/DEVD and caspase-3/DMQD were refined using the program SHELX97 (Sheldrick 1997) and caspase-3/VDVAD was refined using CNS (Brunger, Adams et al. 1998) due to the lower resolution. The molecular graphics program O 8.0 (Jones, Zou et al. 1991) was used to display the electron density map and to refit structures. The initial inhibitor structures were generated by energy minimization using AMMP (Harrison 1993). Water molecules and alternate conformations of caspase-3 residues were modeled where observed in the electron density maps. Structural figures were made by WebLab viewer pro (Accelrys Inc., MA) and images of electron density map were obtained using Molscript (Esnouf 1997; Esnouf 1999).

3.2.4. Protein Data Bank Accession Codes

The crystal structures have been deposited in the RCSB Protein Data Bank with accession codes 2H5I for caspase-3/DEVD, 2H5J for caspase-3/DMQD, and 2H65 for caspase-3/VDVAD.

3.3. Results and Discussion

3.3.1. Crystal Structures

The crystal structures of caspase-3 were determined in the complexes with the three substrate analog inhibitors Ac-DEVD-Cho, Ac-DMQD-Cho and Ac-VDVAD-Cho. The crystallographic statistics are summarized in Table 3.1. The crystal structures were refined to the resolutions of 1.7-2.3 Å and R-factors of 19.4-22.7. The structures of caspase-3/DMQD and caspase-3/VDVAD have not been reported previously, while our structure of caspase-3/DEVD was determined at a significantly higher resolution of 1.7 Å compared to 2.5 Å for the reported structure of 1PAU (Rotonda, Nicholson et al. 1996). The mature caspase-3 consists of the p17 and p12 subunits derived from processing of the procaspase-3. The crystal structure of caspase-3/DEVD has the p17/p12 heterodimer consisting of residues 29-174 and 185-277 in the asymmetric unit of the *I*222 space group. Residues 175 in p17 and 176-184 in p12 were not visible due to the weak electron density in the terminal regions. The complexes of caspase-3/DMQD and caspase-3/VDVAD were crystallized in the *P*2₁ space group with two heterodimers (p17/p12)₂ (residues 34-174 and 186-277 for each heterodimer) in the asymmetric unit (Fig 3.1). The terminal residues (29-33 and 175 in p17; 176-185 in p12) were not visible in the electron density map.

The three caspase-3 structures are closely similar overall, with RMS deviations of 0.35-0.60 Å for C α atoms although they were obtained in two different space groups. The major differences, with RMS deviations on C α atoms of greater than 1.0 Å, were in the loop1 (residues

56-61) and loop4 (residues 251-254). In addition, the residues 29-33 at the N-terminus of the p17 subunit have visible electron density only in the structure of caspase-3/DEVD in the space group *I*222. The p17 N-terminal region extends out of the p12/p17 heterodimer and interacts with the C-terminal region of the p12 subunit of a symmetry related molecule. Hydrogen bond interactions are formed between two symmetry related molecules: the main chain amide and carbonyl oxygen of His277 interact with the carbonyl oxygen of Ile31 and the amide of Leu33, respectively. In the structures of caspase-3/DMQD and caspase-3/VDVAD in space group *P*21, no ordered density was observed for residues 29-33 and the p17 N-terminal region has no contacts with nearby symmetry related molecules. Self processing is unlikely since the peptide sequence of residues 30-33 GISL is not a site for self-cleavage of caspase-3 (Stennicke and Salvesen 1997). Therefore, the conformation of p17 N-terminus is more flexible in the structure in the space group *P*21.

Most of the residues in the active site had very similar positions in all three complexes. The exception was the side chain of His 121, which is located beside the catalytic Cys 163 at the bottom of the active site groove. In the structure of caspase-3/DEVD, the imidazole group of His 121 is directed away from the catalytic Cys 163 and into the substrate binding groove. In the other two structures, the side chain of His 121 has different conformations in the two heterodimers: one resembles that in caspase-3/DEVD, and the other is rotated so that the imidazole group is closer to Cys 163 (Fig 3.2). This conformational change was not observed in the previous caspase-3 structures (Rotonda, Nicholson et al. 1996; Mittl, Di Marco et al. 1997); and it suggests that the side chain of His121 can approach Cys 163 to participate in the catalytic reaction, and then move away after the cleavage has occurred to facilitate release of the reaction products.

3.3.2. Caspase-3 Interactions with Peptide Analogs

All the amino acids in the peptide analogs were clearly visible in the three structures. The electron density map for the inhibitor in the caspase-3/DEVD structure is shown in Figure 3.3a. The substrate analogs bind with thioether bonds between the aldehyde group (–CHO) and the mercapto group (–SH) of Cys 163 of the caspase-3. The three inhibitors bind in the β -strand conformation with almost identical overall conformation and atomic positions (Figure 3.3b). Their C α atoms (P1-P4) have small RMS deviations from 0.09Å to 0.3Å. The side chain atoms of aspartates at P1 and P4 also have very similar positions. This rigid conformation allows caspase-3 residues to interact with the β -strand inhibitor tightly in the active site groove. The P2 and P3 amino acid side chains of the inhibitor share very similar positions for their equivalent atoms. The differences in the size and type of amino acid at P2 and P3, as well as the presence of P5 in one inhibitor, however, are expected to lead to the potency differences of the three inhibitors.

Caspase-3 showed very similar interactions with the main chain atoms of P1-P4 and the Asp side chains at P1 and P4 of the analogs in all three structures (Fig 3.4 and Tab 3.2). In the structure of caspase-3/DEVD, the side chain of Asp at P1 has ionic interactions with Arg 64 and Arg 207 and a hydrogen bond interaction with Gln 161 in the S1 pocket (Fig 3.4a). The main chain amide of P1 Asp interacts with the main chain carbonyl oxygen of Ser 205. At the P2 position, the hydrophobic side chain of Val has van der Waals interactions with caspase residues Tyr 204, Trp 206 and Phe 256. The negatively charged Glu at P3 interacts closely with Arg 207 forming an ionic interaction between their side chains and two hydrogen bond interactions between their main chain carbonyl oxygen and amide groups. In addition, P3 Glu shows a water-mediated interaction with Thr 62. The carboxylate side chain oxygens of P4 Asp form hydrogen bond interactions with the side chain of Asn 208 and main chain amide of Phe 250. A water-

mediated interaction is formed between main chain amide of P4 Asp and the main chain carbonyl oxygen of Phe 250. At the N-terminus of the inhibitor, the acetyl protecting group interacts with the main chain amide of Ser 209.

At P2 and P3, the protein-inhibitor interactions of caspase-3/DMQD (Fig 3.4b) differ from those in the complex of caspase-3/DEVD. The polar side chain of P2 Gln is directed out of the hydrophobic S2 pocket. At the P3 position, the long hydrophobic side chain of methionine lies in the S3 pocket, but forms no specific interactions. Fortunately, the hydrogen bonds between the main chain atoms of P3 and Arg 207 are conserved. Similar interactions are present in both two inhibitor binding sites of this (p17/p12)₂ structure.

The longer peptide analog Ac-VDVAD-Cho binds to caspase-3 in a similar manner to the other two inhibitors at positions P1-P4 (Fig 3.4c). At the P2 position, the side chain of Ala forms van der Waals interactions with hydrophobic residues in the S2 pocket. However, the hydrophobic side chain of P3 Val has no contacts with nearby caspase-3 atoms, which is similar to P3 Met in Ac-DMQD-Cho. At the P4 position, in addition to the hydrogen bond interactions observed in the caspase-3/DEVD structure, another hydrogen bond is formed between the main chain carbonyl oxygen of P4 Asp and main chain amide of Phe 250 (Tab 3.2).

The peptide analog Ac-VDVAD-Cho contains P5 Val, unlike the other two analogs. The P5 main chain amide is positioned to interact with the side chain hydroxyl group of Ser 209, and its carbonyl oxygen interacts with both the side chain hydroxyl group and main chain amide of Ser 209 (Fig 3.4c). These three hydrogen bonds stabilize the main chain of P5 Val and help to position its hydrophobic side chain in a hydrophobic cleft formed by the side chains of Phe 252 and Phe 250. The acetyl group, on the other hand, has moved out of the active site groove of caspase-3 into the solvent. Therefore, caspase-3 forms specificity pockets for recognition of

substrate residues from P5 to P1. The hydrophobic S5 pocket in caspase-3 has not been described previously.

3.3.3. Conformational Change when Caspase-3 Binds the P5-Containing Peptide

The caspase-3 residues forming S5 in the structure of caspase-3/VDVAD undergo a conformational change relative to the complexes with tetrapeptides (Fig 3.5). The S5 pocket is formed by residues from loop4, which has moved towards loop1 by up to 2.9 Å for C α atoms compared to the positions in the complexes with tetrapeptides. This conformational change partly closes the entire active site groove of caspase-3 and enables P5 Val to form good van der Waals contacts with Phe 250 and 252. This more closed conformation of the substrate-binding groove has not been reported for other caspase-3 structures. In contrast to previous suggestions that the active site groove of caspase-3 is rigid (Ni, Li et al. 2003), our results indicate that the loop4 region of caspase-3 near S5 is flexible, and it forms a mouth-like binding groove together with loops 1, 2 and 3. This “mouth” can open to different extents to accommodate a variety of substrates by an induced fit mechanism. These structures suggest that the loop1 and loop4 regions are flexible in physiological conditions and this flexibility contributes to the binding of substrate. Other caspases may also have flexible binding sites formed by their four loops. This observation provides valuable insight into the dynamic mechanism of caspase recognition and binding of substrates.

3.3.4. Enzyme Kinetics and Relative Inhibition

Three substrate analogs Ac-DEVD-Cho, Ac-DMQD-Cho and Ac-VDVAD-Cho that differ in the P2, P3, and presence of the P5 position, were assayed for inhibition of the caspase-3 activity (Tab 3.3). Overall, the three analogs showed similar potency as reversible tight-binding inhibitors. Ac-DEVD-Cho, with the canonical sequence, was the strongest inhibitor with a K_i value of 0.2 nM. Ac-VDVAD-Cho was five-fold weaker, and Ac-DMQD-Cho was the weakest

with a K_i value of 1.9 nM. The K_i values were comparable to those reported in previous studies (Mittl, Di Marco et al. 1997; Talanian, Quinlan et al. 1997).

Three colorimetric peptides Ac-DVAD-pNA, Ac-VDVAD-pNA and Ac-LDVAD-pNA were designed based on the crystal structures to evaluate the importance of the P5 position for specificity of caspase-3. These peptides were shown to be substrates of caspase-3 and the kinetic parameters are listed in Table 4. The major differences were observed in the K_m values, while the k_{cat} values showed little variation, which suggested that the P5 residue affected the binding affinity. Caspase-3 showed the highest catalytic efficiency (k_{cat}/K_m) and lowest K_m for Ac-LDVAD-pNA, about 40% higher than the values for Ac-DVAD-pNA. Similarly, the k_{cat}/K_m for Ac-VDVAD-pNA was 20% higher than for Ac-DVAD-pNA. These results demonstrated that the hydrophobic P5 residue has an important contribution to the recognition and hydrolysis of substrates by caspase-3.

3.3.5. Roles of S2 and S3 in Substrate Recognition and Caspase-3 Activity

The three crystal structures of caspase-3 with peptide analogs that vary in the amino acids at P2, P3 and the presence of P5 have demonstrated the molecular basis for substrate recognition. Caspase-3 interactions with the three analogs are conserved for the P1 and P4 positions and differ for P2, P3 and P5. Kinetic studies have indicated that the caspase-3 substrate preferences at P1 and P4 are almost absolute, since only aspartic acid is acceptable at these two positions and any substitution resulted in a dramatic decrease in the binding affinity (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997). In agreement with these studies, all the protein-inhibitor interactions in S1 and S4 are strong and conserved in all three crystal structures (Tab 3.2). Unlike P1 and P4, the substrate selectivity of caspase-3 can vary for P2 and P3 positions. Theoretically, hydrophobic amino acids are preferred at P2, and hydrophilic amino acids are preferred at P3.

This is confirmed by the 10-fold difference in K_i values for Ac-DEVD-Cho and Ac-DMQD-Cho (Tab 3.3).

The S2 pocket has an important role in both substrate recognition and regulation of caspase-3 activity. Previous studies suggested that various hydrophobic amino acid residues at P2 bound with high affinity, while polar amino acids were weakly bound (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997). At the P2 position, the main chain atoms cannot form hydrogen bonds with caspase-3. The affinity for P2 binding is therefore largely dependent on the side chain atoms. Hydrophilic residues will be unfavorable in the hydrophobic S2 pocket. Previous studies showed that a pyrrolidine ring bound in S2 can increase the inhibitory potency of 1-Methyl-5-nitroisatin by 30-fold (Lee, Long et al. 2000). In our kinetic assay, Ac-DMQD-Cho showed twofold weaker inhibition of caspase-3 than did Ac-VDVAD-Cho (Table 3.3). Similarly, the catalytic efficiency for hydrolysis of Ac-DMQD-pNA was 54% of the value for Ac-DVAD-pNA and 17% of the value for Ac-DEVD-pNA (Tab 3.4). The structural explanation is that the binding of polar P2 Gln in hydrophobic S2 is unfavorable and decreases the binding affinity of the inhibitor, although the P5 position will also have an effect. Nevertheless, the CB and CG atoms in the long side chain of P2 Gln form favorable van der Waals interactions with residues in the S2 pocket. Therefore, the inhibition is not greatly decreased. The S2 pocket appears to have a role in regulation of caspase-3 activity. The crystal structure of caspase-3 in complex with the inhibitor of apoptosis protein XIAP showed that the side chain of Tyr 204 was rotated into the S2 pocket and blocked the binding of substrate (Riedl, Renatus et al. 2001). Also, the side chain of Tyr 204 filled the S2 pocket in the structure of an unliganded caspase-3 (Ni, Li et al. 2003). These results suggest that the S2 pocket has an important role in both substrate recognition and regulation of caspase activity.

The P3 position can tolerate a wide range of amino acids. Although negatively charged or polar amino acids have better binding affinity, some non-polar amino acids such as valine and alanine can also bind in the S3 pocket quite stably (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997). Analysis of our structures showed that the hydrogen bond interactions formed between the main chain atoms of P3 residue and Arg 207 are conserved in all three structures (Tab 3.2). These interactions stabilize the P3 residue in the correct location, independent of the type of side chain. Hence, the P3 residue can vary without significantly altering the binding affinity.

3.3.6. Role of S5 Pocket in Caspase Recognition of Substrates

The structure of caspase-3/VDVAD shows that the side chain of P5 Val lies in a hydrophobic cleft formed by two aromatic residues in the loop4 region, Phe 250 and Phe 252 (Fig 3.6b). Moreover, the caspase-3 loops forming the binding site had a conformational change relative to the two complexes with tetrapeptides, suggesting that the P5 residue of a pentapeptide bound by an induced fit mechanism. The importance of a hydrophobic P5 residue for binding and hydrolysis of caspase-3 substrates was confirmed by our kinetic assays showing k_{cat}/K_m values in the order Ac-LDVAD-pNA > Ac-VDVAD-pNA > Ac-DVAD-pNA (Tab 3.4). Because the P5 residue is located at one end of the major active site groove of caspase-3, polar side-chains are free to rotate into the solvent and away from the hydrophobic S5 site, which might explain why little apparent substrate preference at P5 has been observed previously for caspase-3. Nevertheless, hydrophobic residues at P5 enhance the binding affinities and specificities of the tested caspase-3 substrates. In contrast, the addition of the hydrophobic P5 residues had the opposite effect on hydrolysis by caspase-7, and the k_{cat}/K_m values were in the reverse order of Ac-DVAD-pNA > Ac-VDVAD-pNA > Ac-LDVAD-pNA (Tab 3.4). This discovery of the S5 recognition site in caspase-3 will be helpful for the future design of novel inhibitors.

Previous work on caspase-2 has revealed that the addition of a hydrophobic P5 residue can increase the V_{\max}/K_m value of its substrate (Talanian, Quinlan et al. 1997), underscoring the critical role of the S5 pocket, and consistent with the observed interactions in the caspase-2/LDESD crystal structure (Schweizer, Briand et al. 2003). The main chain conformations of the two bound pentapeptides are very similar in the structures of caspase-2/LDESD and caspase-3/VDVAD (Fig 3.6a). Also, the hydrogen bond and van der Waals interactions of P5 in the S5 binding pocket are similar in caspase-3 and caspase-2 (Fig 3.6b and c). The main chain amide and carbonyl oxygen of the P5 residue form hydrogen bond interactions with Ser209 of caspase-3 or with Thr233 of caspase-2. The hydrophobic P5 side chain has van der Waals contacts with structurally equivalent hydrophobic side chains from loop4 in both caspase-2 and -3. In caspase-2, the P5 Leu interacts with Tyr273 and Pro275, while in caspase-3 P5 Val interacts with Phe250 and Phe252.

No apparent substrate selectivity at P5 has been reported for other caspases. Therefore, the potential S5 binding site formed by residues of loop4 was examined for other caspases of known structure. The sequences and structures of six human caspases (caspase-1(Wilson, Black et al. 1994), 2(Schweizer, Briand et al. 2003), 3, 7(Chai, Wu et al. 2001), 8(Watt, Koeplinger et al. 1999), and 9(Chao, Shiozaki et al. 2005)) are compared in Figure 3.7 and 3.8. These six caspases have diverse sequences and lengths for the loop4 regions (Fig 3.7), suggesting that they will vary in their substrate specificity. The loop4 regions of caspase-2, -3 and -7 have similar lengths and conformations. Caspase-2 and -3 have two aromatic residues forming the S5 pocket; while the polar Gln and Asp were located at the equivalent positions of caspase-7. We therefore predict that either caspase-7 forms a hydrophilic S5 pocket, or else it has no preference for the type of residue at the P5 position. These differences in substrates recognition of caspase-3 and -7

are likely to result in different cellular effects, including activation of distinct signaling pathways. The other three caspases, -1, -8 and -9, have shorter loop4 regions and may not be able to form an S5 binding pocket (Fig 3.8). Therefore, these caspase proteins are likely to differ in their recognition of the P5 position of their substrates.

3.4. Conclusions

Our crystal structures of caspase-3 with three different peptide inhibitors have revealed the molecular basis for substrate preferences at P2, P3, and suggested the preference for hydrophobic side chains at P5. The importance of the hydrophobic P5 residue was confirmed by studies of caspase-3 activity on substrates with different P5 residues. The newly defined hydrophobic S5 pocket of caspase-3 is similar to the S5 pocket in caspase-2, but polar residues are found in equivalent positions of caspase-7, suggesting that these caspases will differ in their substrate selectivity at P5. Moreover, caspase-1, 8 and 9 are unlikely to have similar S5 binding pockets on the basis of their structures. These discoveries will be valuable for the future design of novel inhibitors that are more specific for caspase-3. The distinct preferences observed for the P5 residue in caspase substrates will help define the particular cellular signaling pathways associated with each caspase protein.

Table 3.1. Crystallographic Data Collection and Refinement Statistics

Structure	Caspase-3/DEVD	Caspase-3/DMQD	Caspase-3/VDVAD
Inhibitor	Ac-DEVD-Cho	Ac-DMQD-Cho	Ac-VDVAD-Cho
Space group	<i>I</i> 222	<i>P</i> 2 ₁	<i>P</i> 2 ₁
A (Å)	69.9	50.2	50.4
B (Å)	86.1	69.3	69.7
C (Å)	98.0	94.1	93.4
β	90°	102°	101°
Resolution (Å)	50-1.7	50-2.0	50-2.3
Data range for refinement (Å)	10-1.7	10-2.0	10-2.3
Completeness overall (last shell) (%)	85.9 (79.0)	94.7 (66.9)	97.6 (89.4)
R merge overall (last shell) (%)	12.5 (36.4)	7.1 (24.5)	8.8 (40.2)
I/sigma overall (last shell)	14.0 (3.3)	9.7 (3.0)	12.0 (2.0)
R _{work} (%)	19.5	20.3	22.7
R _{free} (%)	23.7	27.2	25.4
Number of waters	240	593	86
RMS deviation (Å)			
Bonds	0.011	0.015	0.007
Angle	0.03 ^a	0.025 ^a	1.3 ^b
Average B-factor (Å)			
Main chain	14.5	21.2	27.9
Side chain	23.5	27.8	32.3
Inhibitor	18.4	24.1	36.4

a. In SHELX97, the angle RMS deviation is indicated by distance in Å; *b.* Structure refined with CNS where the angle RMS deviation is indicated by angle in degrees.

Table 3.2. Polar interactions of caspase-3 with peptide analogs

	Analog atoms	Protein atoms	Distance (Å)		
			Ac-DEVD-Cho	Ac-DMQD-Cho Chain E/F	Ac-VDVAD-Cho Chain E/F
P1	Asp N	Ser205 O	3.0	2.9/2.9	2.8/2.8
	Asp OD1	Arg64 NE	2.9	2.7/2.8	2.6/2.7
	Asp OD1	Arg207 NE	2.9	2.8/2.8	3.2/3.4
	Asp OD1	Arg207 NH1	3.1	3.0/2.8	2.8/3.1
	Asp OD2	Arg64 NH1	2.9	2.9/2.9	2.9/2.6
	Asp OD2	Gln161 NH2	2.9	2.9/2.9	3.2/3.1
P2					
P3	E/M/V N	Arg207 O	2.8	2.6/2.8	2.8/2.8
	E/M/V O	Arg207 N	2.9	2.8/2.8	2.8/3.0
	E OE1	Arg207 NH2	2.9		
P4	Asp OD1	Asn208 ND2	3.1	2.9/2.9	2.8/2.9
	Asp OD2	Phe250 N	3.1	3.0/3.0	3.1/3.2
	Asp N	Phe250 O			3.0/3.2
P5	Val O	Ser209 N			3.0/2.8
	Val O	Ser209 OG			2.9/2.6
	Val N	Ser209 OG			3.0/2.8
	Ace O	Ser209 N	2.9	2.7/2.9	

Table 3.3: Inhibition constants

Inhibitor	K_i (nM)
Ac-DEVD-Cho	1.3 ± 0.1
Ac-VDVAD-Cho	6.5 ± 0.4
Ac-DMQD-Cho	12.4 ± 0.7

Caspase-3 activity was assayed using the substrate Ac-DEVD-pNA and the K_m value was $67 \pm 4 \mu\text{M}$. Ac indicates the acetyl group and Cho represents the aldehyde group. 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: $K_i = (IC_{50} - 0.5[E]) / (1 + [S]/K_m)$, where [E], [S] and IC_{50} , respectively, correspond to active enzyme concentration, substrate concentration, and the inhibitor concentration needed to suppress half enzyme activity. The enzyme concentration was determined by active site titration.

Table 3.4: Kinetic parameters of caspase-3 substrates

	k_{cat} (min^{-1})		K_m (μM)		k_{cat}/K_m ($\text{mM}^{-1}\text{min}^{-1}$)	
	C3	C7	C3	C7	C3	C7
DEVD	53.1±3.2	NA	67.1±4.3	NA	790.4±47.4	NA
DMQD	211.8±12.7	NA	1600.3±96.0	NA	132.4±7.9	NA
DVAD	54.2±1.8	56.8±1.1	222.3±7.3	219.3±4.4	243.8±8.0	259.0±5.2
VDVAD	48.4±1.6	75.6±2.3	164.7±5.4	314.9±9.4	293.9±9.7	240.1±7.2
LDVAD	50.6±1.5	66.1±1.3	147.2±4.4	323.9±6.5	343.8±10.3	204.1±4.1

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 substrates at various concentrations. *p*-nitroanilide released by enzyme cleavage was measured at a wavelength of 405 nm using a Polarstar Optima microplate reader. K_m and V_{max} values were determined by fitting the Michaelis-Menten plot: $V=V_{\text{max}}[S]/(K_m+[S])$ where [S] refers to substrate concentration. The catalytic constants k_{cat} of caspase-3 substrates were determined by the equation $k_{\text{cat}}=V_{\text{max}}/[E]$, where active enzyme concentration [E] was measured by active site titration.

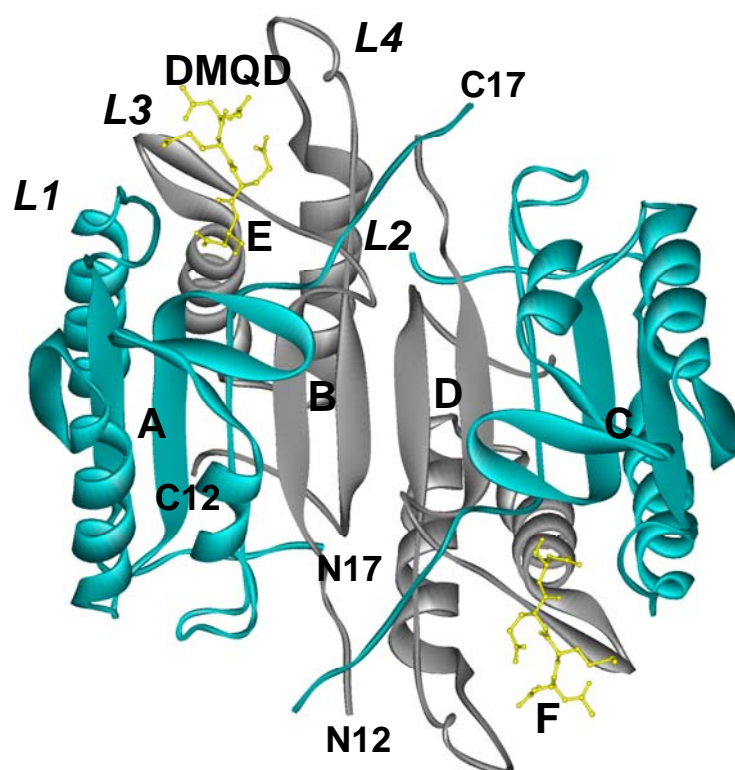


Figure 3.1. Overall structure of caspase-3/DMQD. Two heterodimers (p17/p12)₂ of caspase-3/DMQD are shown in a ribbon representation with the large and small subunits colored blue and grey, respectively. The inhibitor Ac-DMQD-Cho is shown in a yellow ball and stick representation. The four polypeptide chains are labeled A to D and the two peptide analog inhibitors are labeled E and F. The N and C-termini are indicated for the 12 and 17 kD chains. L1 to L4 refer to Loops 1 to 4.

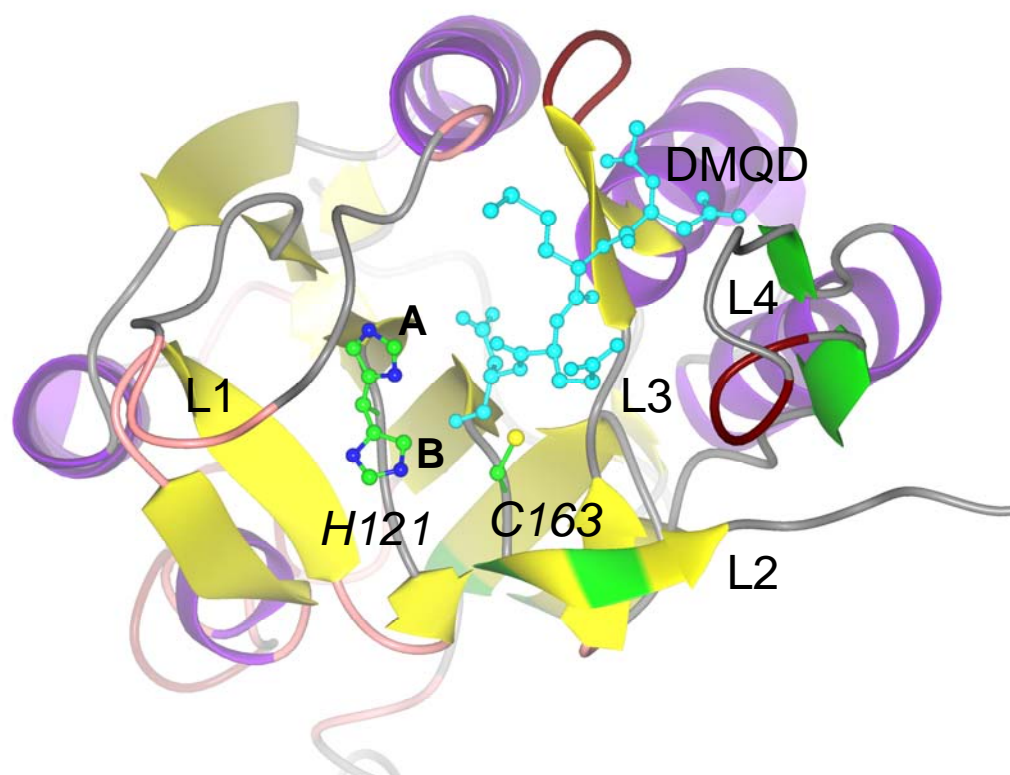


Figure 3.2. His121 has different side-chain conformations in the two p17/p12 heterodimers of caspase-3/DMQD. The A conformation of the His 121 side-chain lies in the active site groove, while conformation B is closer to the catalytic Cys163. The active site groove is formed by residues from the four loops (L1–L4). The peptide analog Ac-DMQD-Cho is shown in cyan.

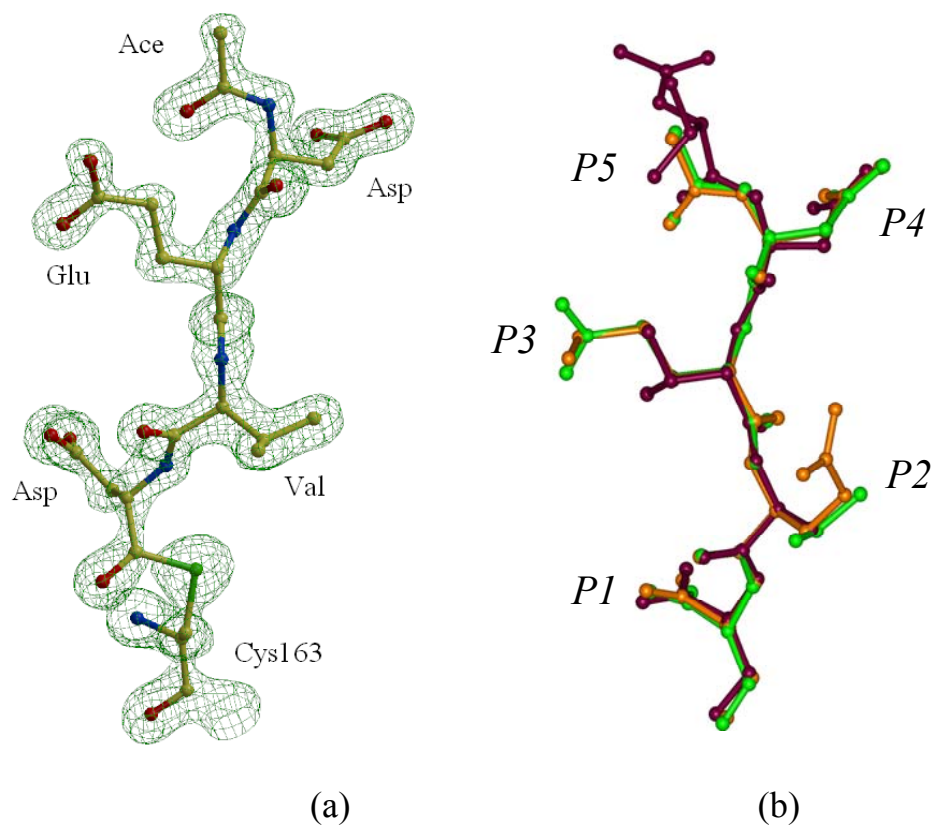


Figure 3.3. Structure of peptide analog inhibitors. (a) $2F_o - F_c$ electron density map of Ac-DEVD-Cho in the caspase-3/DEVD complex. The map was contoured at a level of 1.8σ . The active site Cys163 of caspase-3, located at the bottom, forms a hemithioacetal bond with the aldehyde group of the inhibitor. Inhibitor P1–P4 residues and the acetyl group are labeled. (b) Superposition of the three peptide analogs Ac-DEVD-Cho (green), Ac-DMQD-Cho (orange), and Ac-VDVAD-Cho (purple). The equivalent atoms on both the main chain and side chain have similar positions from P1 to P4.

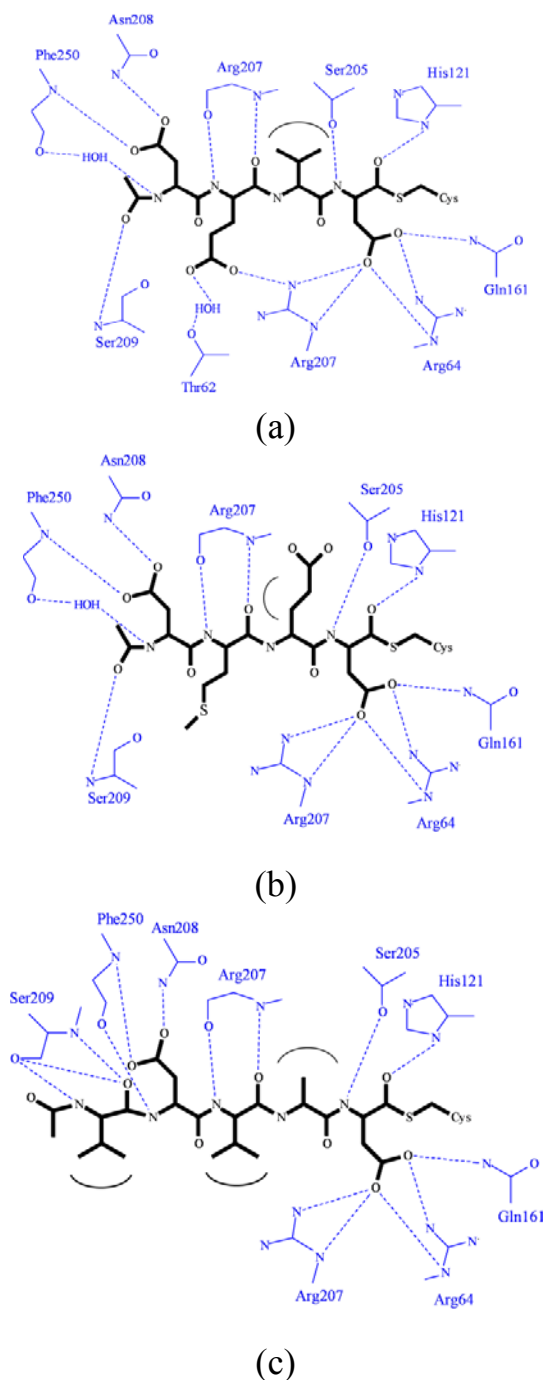
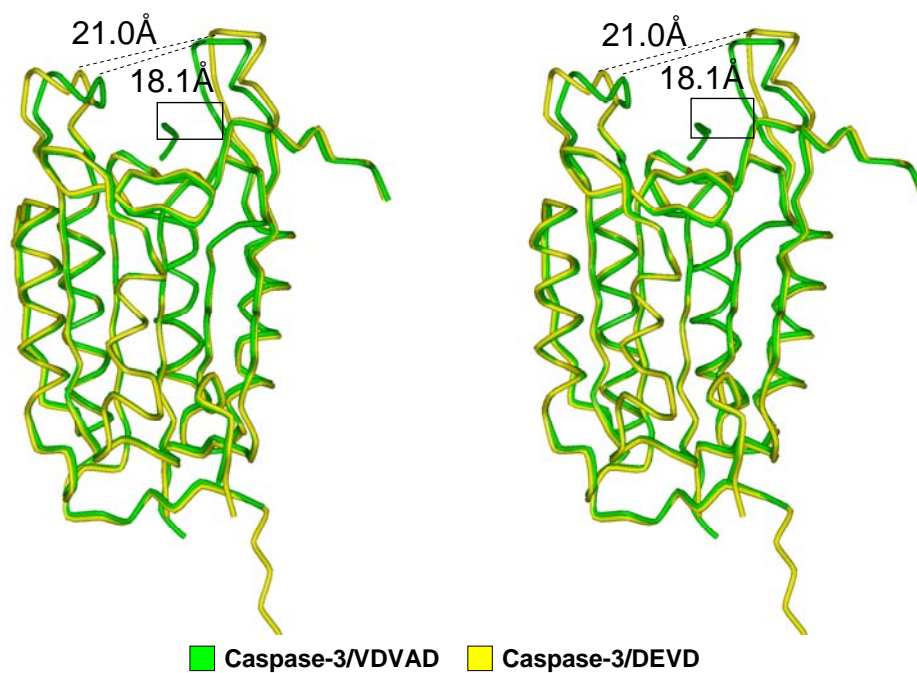
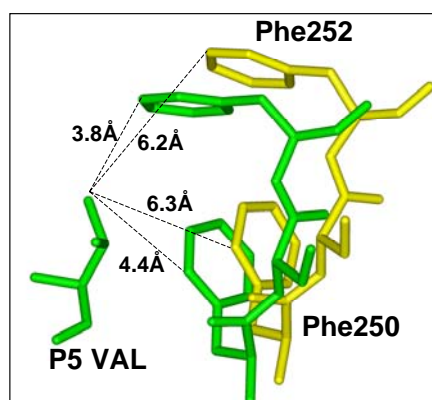


Figure 3.4. Schematic representation of the caspase-3 interactions with inhibitors. (a) caspase-3/DEVD, (b) caspase-3/DMQD, and (c) caspase-3/VDVAD. The substrate analog inhibitors are indicated by thicker lines. Inhibitors are covalently linked to the active site Cys163 via a thioether bond. Hydrogen bonds and salt bridges are indicated by dashed lines. Van der Waals interactions are shown as thick curves.



(a)



(b)

Figure 3.5 The superimposed complexes of caspase-3/DEVD and caspase-3/VDVAD. (a) Stereoview of C α backbone of caspase-3/DEVD (yellow) and caspase-3/VDVAD (green). The complex with P5 residue has a narrower active site groove, as reflected by the indicated distances. The boxed region is shown in detail in (b) where the S5 residues of Phe 250 and 252 on the loop4 form van der Waals contacts with the P5 Val of the substrate analog. Distances between atoms are shown by dashed lines.

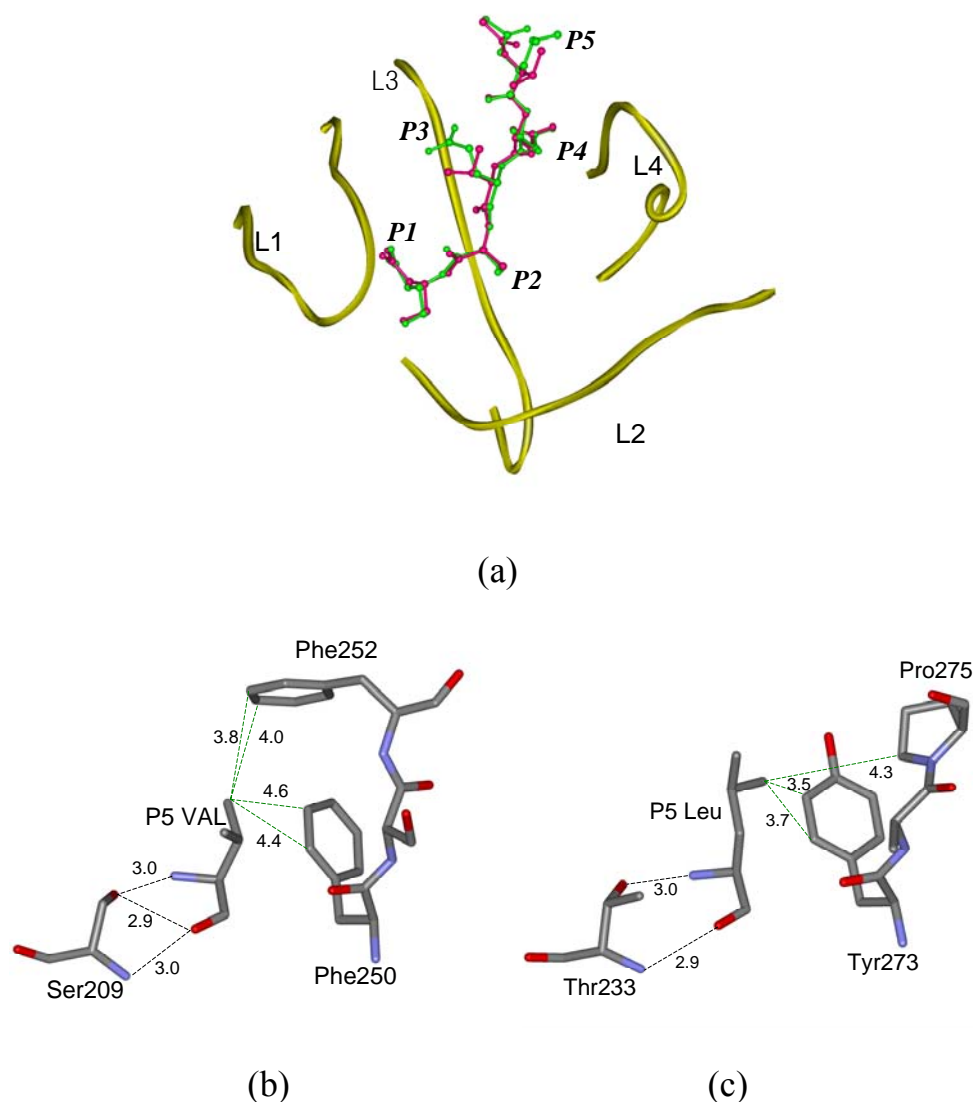


Figure 3.6. Structural comparison of caspase-3 and caspase-2. (a) Comparison of interactions of substrate analogs VDVAD (red) and LDES (green) with caspase-3 and -2. The substrate binding groove is formed by four loops labeled L1 to L4. The inhibitor amino acids from P1 to P5 are labeled. Protein-inhibitor interactions are compared in the S5 pocket of (b) caspase-3 and (c) caspase-2. The P5 residue of the inhibitor was located in the middle. Its interactions are indicated by dashed lines. Hydrogen bond interactions and hydrophobic interactions are indicated by black and green dashed lines respectively. Distances between atoms are shown in angstroms.

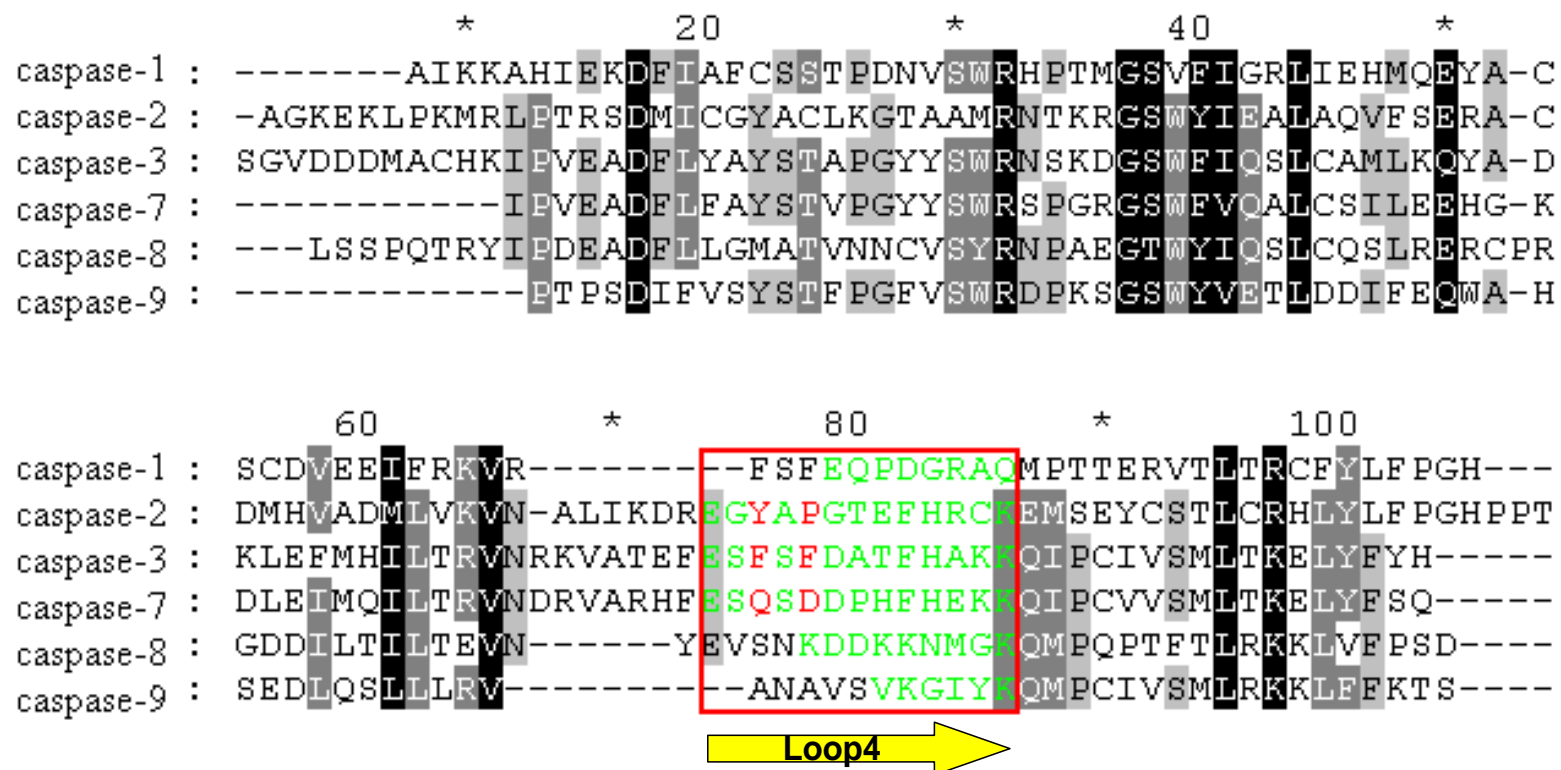


Figure 3.7. Sequence homology among six caspase family members. Only the regions of the small subunits including loop3 and loop4 are compared. The loop4 regions from sequence alignment are shown in the red box. The loop4 regions defined from the crystal structures, are indicated by the green letters. The S5 residues of caspase-2 and -3, as well as the residues at equivalent positions of caspase-7, are colored red.

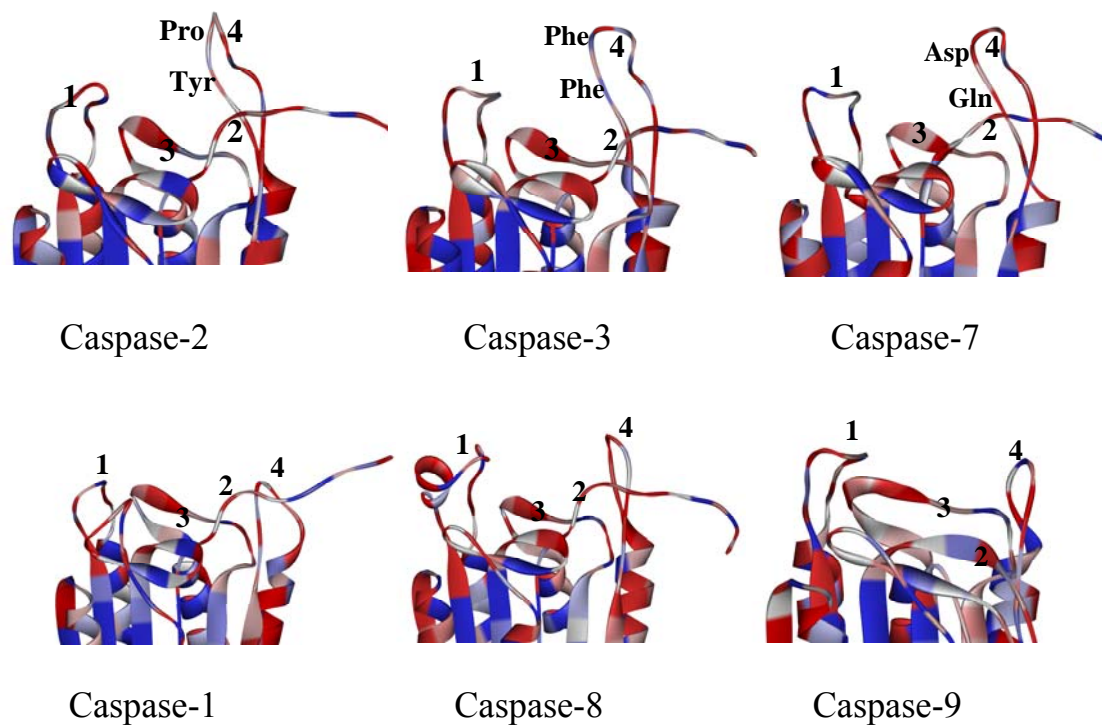


Figure 3.8 Different compositions of active site groove. The structures of the active site grooves of six human caspase family members are shown. The numbers 1 to 4 indicate the loops 1-4. Hydrophobic protein residues are colored blue, and hydrophilic ones are colored red. The loop4 regions are longer in caspase-2, 3 and 7. The S5 residues on the loop4 of caspase-2, 3 as well as the equivalent residues of caspase-7 are labeled.

4. COMPACT, DIFFERENTIABLE, KNOWLEDGE-BASED POTENTIAL FUNCTIONS FOR EVALUATING PROTEIN MODEL QUALITY

4.1. Introduction

An effective and efficient potential function is required for protein structure prediction. There are currently two major categories of potential functions: physical-based potentials and knowledge-based potentials. The physics-based potentials are the summation of non-bonded electrostatic and van der Waals interaction energies. Numerous factors have to be considered for this method, such as bond length, angle, dihedral angle, entropy, and solvent effects (Brooks and Bruccol 1983; Jorgensen, Maxwell et al. 1996; Lazaridis and Karplus 1999). However, this physics-based potential is becoming less and less popular for protein structural prediction due to expensive computation. In contrast, the knowledge-based potentials provide a more efficient way of assessing potentials and thus have become widely used (Zhou, Zhou et al. 2006). The typical knowledge-based potential is built on the mean potentials from the distribution of pairwise distances of experimentally determined protein structures (Zhou, Zhou et al. 2006). Boltzman's law is used to calculate the interaction energy between a particular particle pair (Sippl 1990). And the entire potential is considered as the summation of all pairwise potentials. Usually one residue is represented by one interacting point, such as C_{α} , C_{β} , and side chain center of mass (SCM) (Zhang, Liu et al. 2004; Fogolari, Pieri et al. 2007). This type of potential is called residue-level potential. All atoms or residue specific heavy atoms were also used for representing residues in many successful models (Zhou and Zhou 2002; Makino and Itoh 2008; Mirzaie, Eslahchi et al. 2009). Their potentials are called atomic-level potentials.

Although current knowledge-based potential functions achieved good performance in protein structure prediction (Zhou, Zhou et al. 2006), molecular mechanics is still dominated by

physical-based potentials. This is because of the common limitation of current knowledge-based potentials. The potential curve of the knowledge-based function is built from a series of discrete points defined by distance-potential pairs. Each point represents the values of a certain distance bin and is independently determined by the statistical analysis of the entire protein structure. It means the relation between potential energy and pairwise distance is not directly described by a mathematical function. This makes the calculation of molecular mechanics difficult because the force exerted on a particle is determined by the gradient of the potential function (Alder and Wainwright 1959) and the gradients cannot easily be obtained for potential functions consisting of discrete points.

In order to overcome the limitations of current knowledge-based potentials, here we designed a novel method of constructing knowledge-based potentials. The potential energies of a pair of residues of a certain type of amino acid, such as nonpolar-nonpolar pair, or charged-charged pair, should follow a certain pattern although there must be some variation between different pairs of residues. Therefore, a small number of potential patterns can be used to represent the potential energies of all the residue pairs. Based on this assumption, if we can find a polynomial expression to describe each potential pattern, the potentials of all the residue pairs can be defined by a set of polynomial expressions. Then the total potential can be easily calculated. The atomic-level potentials can also be obtained using the same strategy, but it will be more complex since more potential patterns might be needed. The advantage of our proposed method is that we do not need to consider how the potential is derived from either physical laws or statistical analysis, the only thing we need is a number of mathematical expressions determined by regression, which can reflect the natural properties of protein potentials.

Since polynomial expressions need more computational power to be determined, we started from using simple, physical equations to approximate potential energies. Physical

potential functions like Gaussian: $E = ae^{-\frac{(r-b)^2}{2c^2}}$ (Laplace 1812), Lennard Jones 6/12

potentials: $E = 4a \left[\left(\frac{b}{r} \right)^{12} - \left(\frac{b}{r} \right)^6 \right]$ (Lennard-Jones 1924), and Morse potentials:

$E = D \left(1 - e^{-a(r-r_k)} \right)^2$ (Morse 1929) have the desirable feature of converging to a constant at

large distances, which is not true, in general, with polynomial expansions. Therefore they are well suited to use in expansions of inter-residue and inter-atomic potentials. In this study, the Morse potential function showed the best outcome among a number of models we have tested. Although different types of residue pairs have distinct potential patterns, we simplify this problem by using the Morse potential curve to approximate the potentials of all residue pairs. Thus the potential values in this study are considered as “pseudo” or “relative” potentials. Unless specified, protein residues are represented by the side chain center (SCM) of mass in this study. Since our potentials are defined by many fewer parameters than other knowledge-based potentials, we name the entire set of 210 potential functions as the Compact-SCM.

4.2. Materials and Methods

4.2.1. Calculation of Total Potential

In our study, the total potential of mean force for a protein was approximated as the sum of pairwise potentials of all the residues. The potential of each different residue pair was defined by one potential function. Therefore, 210 functions were determined for the total potential of a protein. The Morse potential function was used to approximate the potential of mean force $u(i, j, r)$ between two residues i and j at the distance r :

$$u(i,j,r)=D_e(1-e^{-a(r-r_e)})^2$$

r_e refers to the equilibrium distance between two residues, a and D_e control the width and depth of the potential curve respectively. The ranges of three parameters were set to: $1 \leq D_e \leq 6$, $0.3 \leq a \leq 3.0$, and $0 < r_e < 14$. The basic step sizes were 1, 0.1, and 0.1 for tuning D_e , a and r_e , respectively. For each residue pair ij , one triple (D_e , a , and r_e) was determined for its unique potential function. Thus, a total number of 630 parameters (210 triples) were determined for 210 potential functions. Finally, the total potential of mean force u for a given protein was calculated as:

$$u(i,j,\infty) = \frac{1}{2} \sum_{i,j} u(i,j,r_{ij}) - \infty$$

Since the potential energy of a well formed protein molecule is negative, the value at infinity (∞) was subtracted to ensure negative values of the potential energy. In this study, the position of each residue was represented using three different approaches: C_α atom, C_β atom, and side chain center of mass (SCM).

4.2.2. Parameter Determination Using Genetic Algorithm

Genetic algorithm is particularly effective for tuning large number of variables (Kaufman 1998). In this study, a vector containing 630 parameters with designed initial values was considered as an individual chromosome. Each population was formed by 30 distinct chromosomes. The elitism selection (Mashohor, Evans et al. 2005) was applied in reproduction. Combination rate and mutation rate were 90% and 1%, respectively. The selection pressure was the rank of the native fold in the decoy sets. The parameters were tuned for over 300 parallel cycles on the Octans cluster at Georgia State University. In each cycle, 4 to 6 decoy sets were used simultaneously. A set of 630 parameters were generated only if the potentials of all the

native structures in tested decoy sets were correctly ranked to the 1st. The final parameter set was produced in the integration step, where all sets of produced parameters were analyzed and the most frequent value for each parameter was extracted as the final value. The averaged values were also tested in our study, but the result was worse than the most frequently appeared values.

4.2.3. Training Data and Decoy Sets

The training data contains decoys from CASP6 and 4state_reduced (Park and Levitt 1996). CASP6 decoys were downloaded from <http://predictioncenter.org/casp6>. Among those decoy sets, only X-ray structures of monomer proteins without gaps were selected. Specifically, 18 decoy sets (t0199, t0200, t0201, t0262, t0263, t0264, t0267, t0268, t0269, t0271, t0273, t0274, t0275, t0277, t0279, t0280, t0281), each containing 60 decoy structures on average, were selected as part of our training data. Because most proteins in CASP6 are relatively big (100-300 residues), seven 4state_reduced decoy sets from <http://dd.compbio.washington.edu/download.shtml> (1ctf, 1r69, 1sn3, 2cro, 3icb, 4pti, 4rxn), were added to our training data set to minimize the bias towards the protein size. Each native structure in this decoy sets has 50-70 residues and each decoy set contains about 660 decoy structures on average. Training decoys were preselected by removing structures with gaps, missing atoms, incomplete or longer sequences. The testing decoy sets are listed in the Table 1. All the decoy sets were used for the tests of C_α and C_β representations while two groups of decoys were used on the examination of SCM representation. One group contains all the 70 multiple decoys and the other group only contains the filtered decoys. Decoys marked with * in the Table 1 were discarded during the filtering as missing side chain atom was found in the PDB file of the native structure. The docking decoy sets were downloaded from http://www.sbg.bio.ic.ac.uk/docking/all_decoys.html.

4.2.4. Performance Measurement

Z-score is the difference from the mean divided by the standard deviation. It places values of the potential in a “standard measure” and reflects the degree to which a value at an individual is distinct from the values at the rest of the data points. It reflects the bias towards the native structure (calculation follows the standard definition (Zhang, Liu et al. 2004)). CC refers to correlation coefficient between C_{α} rmsd and calculated potential energy. FE represents fraction enrichment which means the percentage of the top 10% best scoring structures that are also found in the top 10% lowest rmsd structures.

4.3. Results and Discussion

4.3.1. Parameters Determined for 210 Potential Functions

Ideally, different residue pairs will have distinct potential trajectories, especially for different types of residue pairs, such as nonpolar-nonpolar pairs and charged-charged pairs. We simplified this complication by approximating the potentials of all residue pairs by one type of trajectory. However, we set a wide value range for each adjustable variable in the potential function, allowing the potential curves to be different from one to another. We tested four types of potential curves defined by single Gaussian's function (Laplace 1812), double Gaussian function, Lenard-Jones function (Lennard-Jones 1924), and Morse function (Morse 1929), respectively. As shown in Figure 4.1, the Morse potential curve showed the best accuracy in our test. The genetic algorithm successfully generated 630 parameters for 210 potential functions. Although there are 210 functions, only a narrow range of values were selected for each variable after training. All 210 functions have the same D_e value of 5.0. The majority functions have their a values in the range of 0.5 ± 0.2 (Table 4.2) although several exceptions exist. In contrast, a variety of values from 4.3 to 8.1 were selected for the parameter r_e as indicated in the Table 4.3. Since the r_e defines the equilibrium distance between two residues, it suggests that our method

tends to use the equilibrium distance as the most important feature to distinguish potentials of different residue pairs. The opening width of the potential curve, which is determined by a , also plays some roles for the discrimination. The well depth of the potential curve, however did not contribute to the discrimination. By analyzing single potential function, correlation between the shape of the curve and the residue type was not observed. This is not surprising as our approximated potentials were designed to function as an intact system.

4.3.2. Performance on 70 Standard Multiple Decoy Sets

In previous studies, it was suggested that a residue can be represented by a single interaction point instead of all the atoms. The potential function using side chain center of mass (SCM) showed the most comparable result to the atomic level potentials (Zhang, Liu et al. 2004). In order to find out the most efficient residue representation for our model, three most widely used approaches, C_α , C_β , and SCM, were tested. The success rates on 70 decoy sets increased significantly from 30%, 46% to 70% as the interaction center moved from C_α , C_β , to SCM. We subsequently compared the performance of the Compact-SCM with three previously published models using the SCM representation. The performance data of KBP, RAPDF, and DFIRE-SCM was obtained from the previous study of Zhang (Zhang, Liu et al. 2004). The detailed comparison was shown in the Table 4.4.

The *lattice_ssfit* decoy set was generated by tetrahedral lattice model with all-atom ENCAD energy function (Xia, Huang et al. 2000). The original package has 2000 decoys for each of eight targets. The performances of the Compact-SCM were quite different on filtered and non-filtered decoys. 5 targets were correctly ranked out of 8 non-filtered decoy sets while all three filtered decoy sets were correctly ranked. Our average Z-scores were lower than other three methods. The *lmds* was generated by a local minimization method with a reduced ENCAD function (Kearar and Levitt 2003). The best success rate of three previous methods on this decoy

was 3/10. Our method achieved comparable performance for both filtered and non-filtered decoys. The *fisa* and *fisa_casp3* decoy sets were generated by simulated annealing method with Bayesian scoring function (Simons, Kooperberg et al. 1997; Simons, Bonneau et al. 1999). Our method successfully rated all the seven targets while all other three approaches missed at least one target. Our average Z-score for *fisa* decoys was comparable to other methods whereas it is somehow lower for the *fisa_casp3* decoys. Among 10 filtered targets of CASP4 decoys, our method discriminated 7 native structures. The accuracy is a little better for the non-filtered decoys. The success rate and average Z-score were both lower than other three methods. In contrast, our method achieved 79% of success for 24 Rosetta targets and only DFIRE-SCM got the similar accuracy. All together, the Compact-SCM achieved an average accuracy of 70% for the entire 70 non-filtered decoy sets, which is close to the DFIRE and better than other two methods. However, if only the filtered decoys are considered, the Compact-SCM achieved an accuracy of 73% which is better than all other three methods. Considering that the performance of knowledge-based potential relies on the training data, it is necessary to exclude low quality structures from the training set. Similarly, the non-filtered testing set can possibly generate false result due to the ignoring of missing side chain atoms or even residues. Therefore, our result for the filtered decoy sets is more reliable than other three methods. In addition, we discarded model structures having gaps or sequence differences from the native structure by filtering, thus some decoys far from the native conformation were excluded from our testing data. This could result in the decrease on the average Z-score for the filtered decoys.

Many knowledge-based potential functions were developed in recent years, some of which use different residue representations other than the SCM. We therefore compared the performance of Compact-SCM with some other methods. The data of comparison was obtained

from (Makino and Itoh 2008) for DFMAC and (Mirzaie, Eslahchi et al. 2009) for all other methods. To exclude the possible training bias of different studies, only commonly tested decoys were compared. The number of correctly ranked targets for each decoy set is shown in the Table 4.5. Our method correctly rated 13 targets out of 19 decoy sets, which is significantly higher than the Rosetta and three ModPipe methods. However, DFIRE-A, DOPE, PC2CA, and Force model all got 15-17 correct rankings. This result, however, is not surprising because those methods use more complicated residue representations. DFIRE-A, DOPE, and Force model use residue specific heavy atom representation, and the PC2CA use both C_{α} and SCM information. The DFMAC is a recently reported method using only main chain atoms. Although its success rate was 6/8 in our comparison, its tested decoy set was significantly less and our method got the same success rate if only the same decoy sets were counted.

CC and FE are other two parameters used in assessing the ability to discriminate the native fold from structural models. In this study, the average CC and FE for all 70 multiple decoy sets were 30% and 20%, respectively. The best average CC was 61% on *Casp4* decoy set. And our results on *lattice_ssfite*, *fisa*, *fisa_casp3* and *lmds* decoy sets were similar to the DFMAC (Makino and Itoh 2008). Clearly, the correlation between rmsd and potential score is not significant for our method. Nevertheless, to achieve a significant correlation coefficient for a large number of decoy sets is still a very challenging problem and most studies can only obtain good results on some particular decoy sets (Makino and Itoh 2008) (Zhou and Zhou 2002).

Overall, the Compact-SCM has demonstrated one of the best performances among methods using the SCM representation. Although the performance is moderate compared with some atomic-level potentials, it is likely due to the limit of the SCM representation. Moreover, our method uses significantly fewer adjustable parameters than those used in other methods

(about 90% fewer than other residue level potentials (Table 4.4) and over 99% fewer than atomic level potentials (Table 4.5)). It suggests that our potential function captures the hidden features of potential energy better than other methods and thus is more statistically significant. Last but not least, although improving the correlation between rmsd and function score is a challenging task, future development will benefit since the construction method of our potentials is much simpler and more flexible than other studies.

4.3.3. Performance on 9 Docking Decoy Sets

In the study of Zhang (Zhang, Liu et al. 2004), the performance of DFIRE-SCM was tested by 21 docking decoys sets. To understand Compact-SCM's capability of discriminating docking decoys, we performed the same test using 9 decoy sets available at http://www.sbg.bio.ic.ac.uk/docking/all_decoys.html. Four decoy sets contain dimer interfaces and five contain trimer interfaces. The performance of Compact-SCM was compared with the LLS (all atom version) (Lu, Lu et al. 2003) and the DFIRE-SCM (data from (Zhang, Liu et al. 2004)). The ranking of native structure for each decoy set is indicated in the Table 4.6. For the dimer decoys, our method successfully rated 3 out of 4 native structures, which is equivalent to the LLS and a bit lower than the DFIRE-SCM. For the trimer decoys, our method correctly rated 4 native structures out of 5. DFIRE-SCM got the same accuracy while LLS got only 1 hit. Overall, the performance of Compact-SCM was much better than the LLS and comparable to the DFIRE-SCM. This result indicated that the Compact-SCM has considerable capability of discriminating docking decoys. Since the Compact-SCM was not trained with any interface decoys, the performance has a good chance to be enhanced in the future.

4.3.4. Conclusions and Future Work

We introduced a new method of generating knowledge-based potentials that can be used for both structural prediction and molecular mechanics. The potential of mean force for each

residue pair was defined by one Morse function determined by training against 25 sets of multiple decoys. The total potential of mean for a protein was calculated as the summation of all residue pairs. The capability of the Compact-SCM on discriminating native structure from structural models was comparable to some of the best models using residue-level potentials. What's more, each potential function was described by only three parameters (D_e , a , and r_e), which is much less (<10%) than other potentials. And derivatives of each potential functions on the distance can be easily calculated. These advantages make the Compact-SCM favorable for both structure prediction and molecular mechanics.

The development of our potential functions is still in progress. We are testing more mathematical models to approximate the potential patterns. The Morse function will be further refined by integrating auxiliary expressions. Meanwhile, we are trying to distinguish different types of residue pairs using different mathematical expressions. And we will eventually establish a set of atomic-level potential functions. With the further development, we believe the performance of our potential functions will be considerably enhanced. This new type of knowledge-based potentials will help the future studies of structural bioinformatics.

Table 4.1. 70 multiple decoy sets for testing.

Source	Average decoys	PDB ID
lattice_ssfit ^a	1998	1beo, 1ctf, 1pgb, 1dtk-A*, 1fca*, 1nkl*, 1trl-A*, 4icb*
lmdb ^b	445	1b0n-B, 1bba, 1ctf, 1dtk, 1fc2, 1igd, 1shf-A, 2cro, 2ovo, 4pti*
fisa ^c	500	1fc2, 1hdd-C, 2cro, 4icb
fisa_casp3 ^d	1498	1bg8-A, 1bl0, 1jwe
CASP4 ^e	37	t0092(1im8), t0098(1fc3), t0100(1qjv), t0103(1ga6), t0104(1fl9), t0108(1j83), t0115(1fwk), t0117(1j90), t0123(1exs), t0125(1ghk), t0086(1fw9)*, t0087(1i74)*, t0096(1e2x)*, t0106(1ijx)*, t0107(1i8u)*, t0111(1e9i)*, t0112(1e3j)*, t0113(1e3w)*, t0118(1fzr)*, t0121(1g29)*, t0126(1f35)*
Rosetta ^d	846	1aa2, 1acf, 1ail, 1bdo, 1csp, 1ctf, 1erv, 1gvp, 1kte, 1mbd, 1msi, 1pal, 1pdo, 1r69, 1ris, 1tul, 1utg, 1vls, 1who, 2acy, 2fha, 4fgf, 2gdm, 5icb

^a(Xia, Huang et al. 2000), ^b(Keasar and Levitt 2003), ^c(Simons, Kooperberg et al. 1997), ^d(Simons, Bonneau et al. 1999), ^e(Feig and Brooks 2002)

*decoys not tested for SCM approach due to the missing side chain atoms in the native structure.

Table 4.4. The success rates and average Z-scores of different SCM potentials.

Source	RAPDF-SCM ^a	KBP-SCM ^a	DFIRE-SCM ^a	Compact-SCM ^b	Compact-SCM ^c
lattice_ssfit	6/8 (3.21)	6/8 (5.11)	8/8 (6.19)	5/8 (1.17)	3/3 (1.52)
lmds	2/10 (1.78)	4/10 (2.59)	3/10 (2.56)	3/10 (1.98)	3/9 (2.04)
fisa	1/4 (2.51)	3/4 (3.99)	3/4 (4.70)	4/4 (3.96)	4/4 (3.96)
fisa_casp3	2/3 (3.70)	3/3 (4.96)	3/3 (6.05)	3/3 (1.49)	3/3 (1.49)
CASP4	19/23 (2.74)	17/23 (3.83)	19/23 (3.15)	15/21 (2.49)	7/10 (1.01)
Rosetta	27/41 (3.55)	29/41 (4.16)	33/41 (4.90)	19/24 (2.54)	19/24 (2.38)
Summary	64.0% (2.92)	69.7% (4.11)	71.9% (4.59)	70.0% (2.27)	73.6% (2.06)
Parameters	7560	5880	8400	630	630

^a (Zhang, Liu et al. 2004)

^b Tested with all decoys

^c Tested with filtered decoys

Numbers in parentheses are average Z-scores

Table 4.5. The success rate of other energy models

Source	DFIRE-A ^a	Rosetta ^a	ModPipe-Pair ^a	ModPipe-Surf ^a	ModPipe-Comb ^a	DOPE ^a	PC2CA ^a	Force model ^a	DFMAC ^b	Compact-SCM
lattice_ssfit	3/3	3/3	3/3	2/3	3/3	3/3	3/3	3/3	2/2	3/3
lmds	6/9	2/9	3/9	2/9	3/9	6/9	6/9	9/9	2/3	3/9
fisa	3/4	1/4	0/4	1/4	0/4	3/4	4/4	4/4	1/1	4/4
fisa_casp3	3/3	0/3	2/3	0/3	2/3	3/3	3/3	1/2	1/2	3/3
Correct	15	6	8	5	8	15	16	17	6	13
Parameters					>200,000					630

^a(Mirzaie, Eslahchi et al. 2009)

^b(Makino and Itoh 2008)

Table 4.6. The Ranking of native structures for 9 docking decoy sets.

PDB ID	1ugh ^a	2sic ^a	1cgl ^a	1dfj ^a	1ahw ^b	1bvk ^b	1dqj ^b	1mlc ^b	1wej ^b
LLS ^c	1	1	1	4	3	4	4	3	1
DFIRE-SCM ^d	1	1	1	1	1	1	1	1	2
Compact-SCM	1	61	1	1	1	1	1	1	55

^a Dimers^b Trimers^c (Lu, Lu et al. 2003)^d (Zhang, Liu et al. 2004)

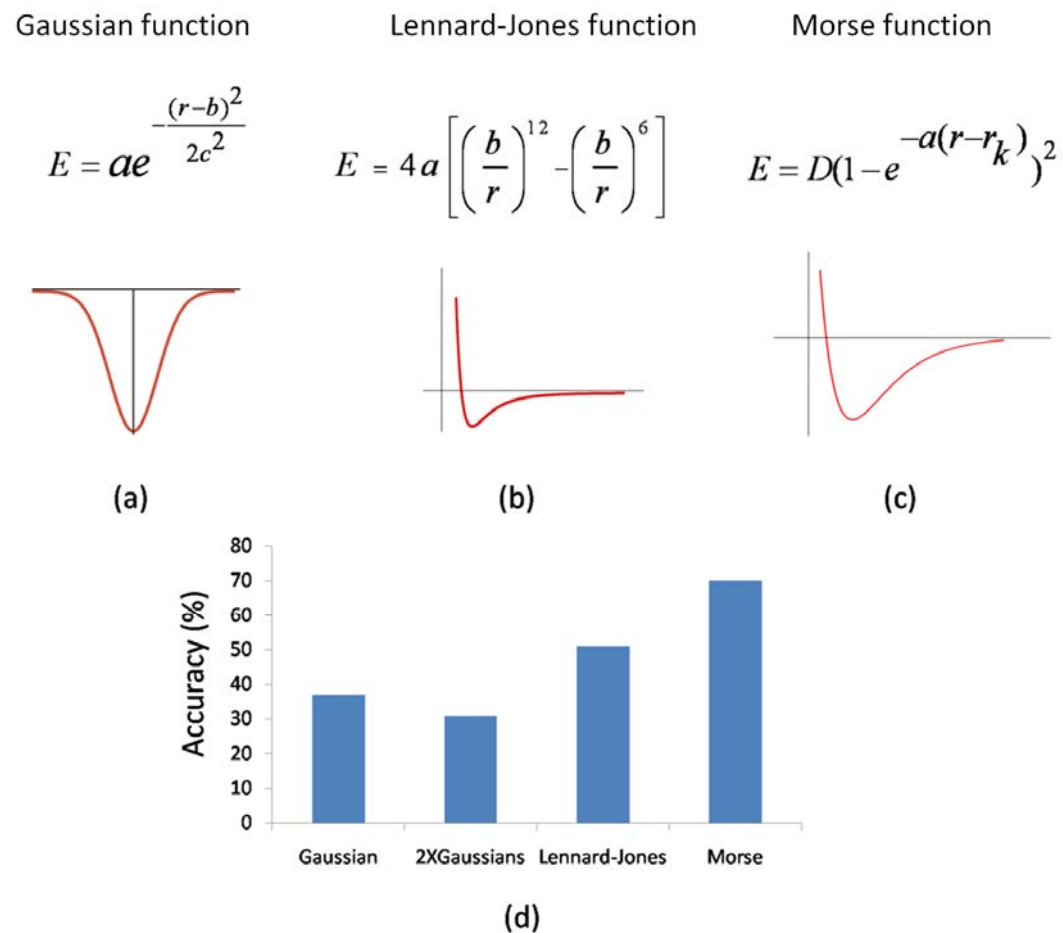


Figure 4.1. Accuracy of different potential functions. The mathematical expressions and curves were shown for (a) Gaussian's function, (b) Lennard-Jones function, and (c) Morse function. (d) In the test of 70 standard multiple decoy sets, four different functions obtained accuracies of 37%, 31%, 50%, and 70%, respectively.

5. OVERALL SUMMARY

In this study, the structural basis of caspase-3 substrate specificity was investigated by crystallographic study and structural modeling. Subsites from S1 to S5 in the caspase-3 substrate binding groove showed different characteristics for substrate recognition. Specifically, the S1 pocket which is highly positive and stringent for P1 Asp, is formed by Arg64, Arg207, and Gln161. Our observation is consistent with the previous conclusion that the S1 pocket exclusively recognizes P1 Asp (Talanian, Quinlan et al. 1997; Thornberry, Rano et al. 1997). The S2 pocket is surrounded by three aromatic residues, Tyr204, Trp206, and Phe256. This pocket prefers hydrophobic residues with small side chains, and its size can be adjusted upon the binding of different P2 residues. The S3 pocket, mainly formed by Arg207, is open at the protein surface and prefers hydrophilic P3 residues. The S4 pocket can accommodate both polar and nonpolar residues. Crystal structures showed that Trp206 and Trp214 in the S4 pocket interacted with P4 Ile or Tyr. Computational modeling suggested that most nonpolar P4 residues, except Trp, can interact with either Trp206 or Trp214 (Fang, Fu et al. 2009). The S5 pocket, formed by Phe250 and Phe252, was newly identified by our study. It can interact with P4 Trp and P5 residues with hydrophobic side chains. Binding of P5-containing peptides leads to conformational changes in the L4 loop via an induced fit mechanism (Fang, Boross et al. 2006). The comparison of seven caspase-3 crystal structures indicated that most (S2-S5) of the caspase-3 substrate binding groove is flexible, and only the S1 pocket is stringent in recognition of amino acids. The S1-S5 binding sites are interdependent, as different substrate residues at one position also showed impact on the binding of residues at other positions (Fang, Boross et al. 2006).

Besides the crystallographic analysis, a computational study was performed to develop novel knowledge-based potential functions for evaluating the quality of protein structural models.

Our new potential energy function Compact-SCM was established using Morse's potential functions. It achieved a comparable accuracy to the best previous method in the same category. More importantly, the number of adjustable parameters used by Compact-SCM is 90% fewer than other residue-level potentials and over 99% fewer than atomic-level potentials. It suggests that our method is more statistically significant and thus better captures the hidden features of protein potential energy. In contrast to traditional knowledge-based potentials, our potential energy functions are defined by differentiable mathematical expressions, which are more physically meaningful and can be directly applied for molecular mechanics. Thus the Compact-SCM is a new type of potential function that combines the advantages of both physics-based potentials and knowledge-based potentials.

Overall, my research will be helpful for the future design of caspase-based drugs in the following two aspects. First, the analysis of the substrate binding sites in caspase-3 and conformations of different residues in the inhibitors provides direct guidance for the design of the next generation of drug candidates. Second, our new potential functions can be used in the prediction of protein structures and protein-ligand structures, which is a widely used technique for screening potential protein inhibitors.

LITERATURE CITED

- Agniswamy, J., B. Fang, et al. (2007). "Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis." FEBS J **274**(18): 4752-65.
- Alder, B. J. and T. E. Wainwright (1959). "Studies in Molecular Dynamics. I. General Method." J. Chem. Phys. **31**(2).
- Alnemri, E. S., D. J. Livingston, et al. (1996). "Human ICE/CED-3 protease nomenclature." Cell **87**(2): 171.
- Aulabaugh, A., B. Kapoor, et al. (2007). "Biochemical and biophysical characterization of inhibitor binding to caspase-3 reveals induced asymmetry." Biochemistry **46**(33): 9462-71.
- Backes, C., J. Kuentzer, et al. (2005). "GraBCas: a bioinformatics tool for score-based prediction of Caspase- and Granzyme B-cleavage sites in protein sequences." Nucleic Acids Res **33**(Web Server issue): W208-13.
- Bagossi, P., J. Tozser, et al. (1999). "Modification of parameters of the charge equilibrium scheme to achieve better correlation with experimental dipole moments." J Mol Model **5**: 143-152.
- Banzhaf, w., P. Nordin, et al. (1998). Genetic Programming - An Introduction. M. Kaufmann. San Francisco.
- Baskin-Bey, E. S., K. Washburn, et al. (2007). "Clinical Trial of the Pan-Caspase Inhibitor, IDN-6556, in Human Liver Preservation Injury." Am J Transplant **7**(1): 218-25.

- Blanchard, H., M. Donepudi, et al. (2000). "Caspase-8 specificity probed at subsite S(4): crystal structure of the caspase-8-Z-DEVD-cho complex." J Mol Biol **302**(1): 9-16.
- Blandini, F., E. Sinforiani, et al. (2006). "Peripheral proteasome and caspase activity in Parkinson disease and Alzheimer disease." Neurology **66**(4): 529-34.
- Blow, D. (2002). "Outline of crystallography for biologists." New York, Oxford University Press.
- Brooks, B. R. and R. E. Bruccol (1983). "CHARMM: a program for macromolecular energy, minimization and dynamics calculations." Journal of Computational Chemistry **4**(2): 30.
- Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.
- Callus, B. A. and D. L. Vaux (2007). "Caspase inhibitors: viral, cellular and chemical." Cell Death Differ **14**(1): 73-8.
- Cerretti, D. P., C. J. Kozlosky, et al. (1992). "Molecular cloning of the interleukin-1 beta converting enzyme." Science **256**(5053): 97-100.
- Chai, J., Q. Wu, et al. (2001). "Crystal structure of a procaspase-7 zymogen: mechanisms of activation and substrate binding." Cell **107**(3): 399-407.
- Chao, Y., E. N. Shiozaki, et al. (2005). "Engineering a dimeric caspase-9: a re-evaluation of the induced proximity model for caspase activation." PLoS Biol **3**(6): e183.
- Chapman, J. G., W. P. Magee, et al. (2002). "A novel nonpeptidic caspase-3/7 inhibitor, (S)-(+)-5-[1-(2-methoxymethylpyrrolidinyl)sulfonyl]isatin reduces myocardial ischemic injury." Eur J Pharmacol **456**(1-3): 59-68.
- Charriaut-Marlangue, C. (2004). "Apoptosis: a target for neuroprotection." Therapie **59**(2): 185-90.

- Chen, Y. H., Y. H. Zhang, et al. (2006). "Design, synthesis, and biological evaluation of isoquinoline-1,3,4-trione derivatives as potent caspase-3 inhibitors." J Med Chem **49**(5): 1613-23.
- Chu, W., J. Rothfuss, et al. (2007). "Isatin sulfonamide analogs containing a Michael addition acceptor: a new class of caspase 3/7 inhibitors." J Med Chem **50**(15): 3751-5.
- Chu, W., J. Zhang, et al. (2005). "N-benzylisatin sulfonamide analogues as potent caspase-3 inhibitors: synthesis, in vitro activity, and molecular modeling studies." J Med Chem **48**(24): 7637-47.
- Cohen, G. M. (1997). "Caspases: the executioners of apoptosis." Biochem J **326** (Pt 1): 1-16.
- Degterev, A., M. Boyce, et al. (2003). "A decade of caspases." Oncogene **22**(53): 8543-67.
- Donepudi, M., A. Mac Sweeney, et al. (2003). "Insights into the regulatory mechanism for caspase-8 activation." Mol Cell **11**(2): 543-9.
- Earnshaw, W. C., L. M. Martins, et al. (1999). "Mammalian caspases: structure, activation, substrates, and functions during apoptosis." Annu Rev Biochem **68**: 383-424.
- Eckelman, B. P., G. S. Salvesen, et al. (2006). "Human inhibitor of apoptosis proteins: why XIAP is the black sheep of the family." EMBO Rep **7**(10): 988-94.
- Ekici, O. D., Z. Z. Li, et al. (2006). "Design, synthesis, and evaluation of aza-peptide Michael acceptors as selective and potent inhibitors of caspases-2, -3, -6, -7, -8, -9, and -10." J Med Chem **49**(19): 5728-49.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Erkoc, S. (2001). "Empirical potential energy functions used in the simulations of materials properties." annual reviews of computational physics IX **9**: 103.

- Erlanson, D. A., J. W. Lam, et al. (2003). "In situ assembly of enzyme inhibitors using extended tethering." Nat Biotechnol **21**(3): 308-14.
- Esnouf, R. M. (1997). "An extensively modified version of MolScript that includes greatly enhanced coloring capabilities." J Mol Graph Model **15**(2): 132-4, 112-3.
- Esnouf, R. M. (1999). "Further additions to MolScript version 1.4, including reading and contouring of electron-density maps." Acta Crystallogr D Biol Crystallogr **55 (Pt 4)**: 938-40.
- Fang, B., P. I. Boross, et al. (2006). "Structural and kinetic analysis of caspase-3 reveals role for s5 binding site in substrate recognition." J Mol Biol **360**(3): 654-66.
- Fang, B., G. Fu, et al. (2009). "Caspase-3 binds diverse P4 residues in peptides as revealed by crystallography and structural modeling." Apoptosis **14**(5): 741-52.
- Feig, M. and C. L. Brooks, 3rd (2002). "Evaluating CASP4 predictions with physical energy functions." Proteins **49**(2): 232-45.
- Fischer, U., R. U. Janicke, et al. (2003). "Many cuts to ruin: a comprehensive update of caspase substrates." Cell Death Differ **10**(1): 76-100.
- Fischer, U. and K. Schulze-Osthoff (2005). "Apoptosis-based therapies and drug targets." Cell Death Differ **12 Suppl 1**: 942-61.
- Fisher, A. J., W. Cruz, et al. (1999). "Crystal structure of baculovirus P35: role of a novel reactive site loop in apoptotic caspase inhibition." EMBO J **18**(8): 2031-9.
- Fogolari, F., L. Pieri, et al. (2007). "Scoring predictive models using a reduced representation of proteins: model and energy definition." BMC Struct Biol **7**: 15.
- Fu, G., A. A. Chumanevich, et al. (2008). "Structural basis for executioner caspase recognition of P5 position in substrates." Apoptosis **13**(11): 1291-302.

- Fuentes-Prior, P. and G. S. Salvesen (2004). "The protein structures that shape caspase activity, specificity, activation and inhibition." Biochem J **384**(Pt 2): 201-32.
- Ganesan, R., S. Jelakovic, et al. (2006). "Exploring the S4 and S1 prime subsite specificities in caspase-3 with aza-peptide epoxide inhibitors." Biochemistry **45**(30): 9059-67.
- Ganesan, R., P. R. Mittl, et al. (2006). "Extended substrate recognition in caspase-3 revealed by high resolution X-ray structure analysis." J Mol Biol **359**(5): 1378-88.
- Garcia-Calvo, M., E. P. Peterson, et al. (1998). "Inhibition of human caspases by peptide-based and macromolecular inhibitors." J Biol Chem **273**(49): 32608-13.
- Gervais, F. G., D. Xu, et al. (1999). "Involvement of caspases in proteolytic cleavage of Alzheimer's amyloid-beta precursor protein and amyloidogenic A beta peptide formation." Cell **97**(3): 395-406.
- Han, Y., A. Giroux, et al. (2005). "Novel pyrazinone mono-amides as potent and reversible caspase-3 inhibitors." Bioorg Med Chem Lett **15**(4): 1173-80.
- Harrison, R. W. (1993). "Stiffness and energy conservation in the molecular dynamics: an improved integrator." J. Comp. Chem. **14**: 1112-1122.
- Harrison, R. W. (1999). "A Self-Assembling Neural Network for Modeling Polymers." J Math Chem **26**: 125-137.
- Hartmann, A., J. D. Troadec, et al. (2001). "Caspase-8 is an effector in apoptotic death of dopaminergic neurons in Parkinson's disease, but pathway inhibition results in neuronal necrosis." J Neurosci **21**(7): 2247-55.
- Hermel, E., J. Gafni, et al. (2004). "Specific caspase interactions and amplification are involved in selective neuronal vulnerability in Huntington's disease." Cell Death Differ **11**(4): 424-38.

- Hotchkiss, R. S. and D. W. Nicholson (2006). "Apoptosis and caspases regulate death and inflammation in sepsis." Nat Rev Immunol **6**(11): 813-22.
- Howard, A. D., M. J. Kostura, et al. (1991). "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 **147**(9): 2964-9.
- Irene T. Weber, B. F. a. J. A. (2008). "Caspases: Structure-Guided Design of Drugs to Control Cell Death." Mini-Reviews in Medicinal Chemistry **8**(11): 1154-1162.
- Jones, T. A., J. Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A **47 (Pt 2)**: 110-9.
- Jorgensen, W. L., D. S. Maxwell, et al. (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." J. Am. Chem. Soc. **118**(45): 11.
- Kaufman, M. (1998). Genetic Programming - An Introduction.
- Keasar, C. and M. Levitt (2003). "A novel approach to decoy set generation: designing a physical energy function having local minima with native structure characteristics." J Mol Biol **329**(1): 159-74.
- Kim, T. W., W. H. Pettingell, et al. (1997). "Alternative cleavage of Alzheimer-associated presenilins during apoptosis by a caspase-3 family protease." Science **277**(5324): 373-6.
- Kopka, K., A. Faust, et al. (2006). "5-pyrrolidinylsulfonyl isatins as a potential tool for the molecular imaging of caspases in apoptosis." J Med Chem **49**(23): 6704-15.

- Kravchenko, D. V., Y. A. Kuzovkova, et al. (2005). "Synthesis and structure-activity relationship of 4-substituted 2-(2-acetyloxyethyl)-8-(morpholine-4-sulfonyl)pyrrolo[3,4-c]quinoline-1,3-diones as potent caspase-3 inhibitors." J Med Chem **48**(11): 3680-3.
- Kravchenko, D. V., V. M. Kysil, et al. (2005). "Synthesis and caspase-3 inhibitory activity of 8-sulfonyl-1,3-dioxo-2,3-dihydro-1H-pyrrolo[3,4-c]quinolines." Farmaco **60**(10): 804-9.
- Kravchenko, D. V., V. V. Kysil, et al. (2005). "1,3-Dioxo-4-methyl-2,3-dihydro-1H-pyrrolo[3,4-c]quinolines as potent caspase-3 inhibitors." Bioorg Med Chem Lett **15**(7): 1841-5.
- Kumar, S. (2007). "Caspase function in programmed cell death." Cell Death Differ **14**(1): 32-43.
- Lakhani, S. A., A. Masud, et al. (2006). "Caspases 3 and 7: key mediators of mitochondrial events of apoptosis." Science **311**(5762): 847-51.
- Lamkanfi, M., W. Declercq, et al. (2002). "Alice in caspase land. A phylogenetic analysis of caspases from worm to man." Cell Death Differ **9**(4): 358-61.
- Lamzin, V. S. and K. S. Wilson (1993). "Automated refinement of protein models." Acta Crystallogr D Biol Crystallogr **49**(Pt 1): 129-47.
- Laplace, P.-S. (1812). "Analytical Theory of Probabilities."
- Lavrik, I. N., A. Golks, et al. (2005). "Caspases: pharmacological manipulation of cell death." J Clin Invest **115**(10): 2665-72.
- Lazaridis, T. and M. Karplus (1999). "Effective energy function for proteins in solution." Proteins **35**(2): 133-52.
- Lee, D., S. A. Long, et al. (2000). "Potent and selective nonpeptide inhibitors of caspases 3 and 7 inhibit apoptosis and maintain cell functionality." J Biol Chem **275**(21): 16007-14.
- Lee, D., S. A. Long, et al. (2001). "Potent and selective nonpeptide inhibitors of caspases 3 and 7." J Med Chem **44**(12): 2015-26.

- Lennard-Jones, J. E. (1924). Proc. Roy. Soc. A **106**.
- Lien, S., R. Pastor, et al. (2004). "A substrate-phage approach for investigating caspase specificity." Protein J **23**(6): 413-25.
- Lu, H., L. Lu, et al. (2003). "Development of unified statistical potentials describing protein-protein interactions." Biophys J **84**(3): 1895-901.
- Lu, H. and J. Skolnick (2001). "A distance-dependent atomic knowledge-based potential for improved protein structure selection." Proteins **44**(3): 223-32.
- Ma, X. Q., H. J. Zhang, et al. (2007). "Novel irreversible caspase-1 inhibitor attenuates the maturation of intracellular interleukin-1beta." Biochem Cell Biol **85**(1): 56-65.
- Maibaum, J. and D. H. Rich (1988). "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 **31**(3): 625-9.
- Makino, Y. and N. Itoh (2008). "A knowledge-based structure-discriminating function that requires only main-chain atom coordinates." BMC Struct Biol **8**: 46.
- Martinon, F., N. Holler, et al. (2000). "Activation of a pro-apoptotic amplification loop through inhibition of NF-kappaB-dependent survival signals by caspase-mediated inactivation of RIP." FEBS Lett **468**(2-3): 134-6.
- Mashohor, S., J. R. Evans, et al. (2005). "Elitist selection schemes for genetic algorithm based printed circuit board inspection system." Evolutionary Computation **2**: 4.
- Micale, N., R. Vairagoundar, et al. (2004). "Design and synthesis of a potent and selective peptidomimetic inhibitor of caspase-3." J Med Chem **47**(26): 6455-8.
- Mirzaie, M., C. Eslahchi, et al. (2009). "A distance-dependent atomic knowledge-based potential and force for discrimination of native structures from decoys." Proteins.

- Mittl, P. R., S. Di Marco, et al. (1997). "Structure of recombinant human CPP32 in complex with the tetrapeptide acetyl-Asp-Val-Ala-Asp fluoromethyl ketone." J Biol Chem **272**(10): 6539-47.
- Morse, P. M. (1929). "Diatomic molecules according to the wave mechanics. II. Vibrational levels." Phys. Rev. **34**.
- Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Crystallog. sect. D **50**: 157-163.
- Ni, C. Z., C. Li, et al. (2003). "Conformational restrictions in the active site of unliganded human caspase-3." J Mol Recognit **16**(3): 121-4.
- Nicholson, D. W., A. Ali, et al. (1995). "Identification and inhibition of the ICE/CED-3 protease necessary for mammalian apoptosis." Nature **376**(6535): 37-43.
- O'Brien, T. and D. Lee (2004). "Prospects for caspase inhibitors." Mini Rev Med Chem **4**(2): 153-65.
- Otwinowski, Z. M., W. (1997). "Processing of X-ray diffraction data in oscillation mode." Methods Enzymol **276**: 307-326.
- Park, B. and M. Levitt (1996). "Energy functions that discriminate X-ray and near native folds from well-constructed decoys." J Mol Biol **258**(2): 367-92.
- Philchenkov, A. (2004). "Caspases: potential targets for regulating cell death." J Cell Mol Med **8**(4): 432-44.
- Pockros, P. J., E. R. Schiff, et al. (2007). "Oral IDN-6556, an antiapoptotic caspase inhibitor, may lower aminotransferase activity in patients with chronic hepatitis C." Hepatology **46**(2): 324-9.

- Qiu, L., Q. X. Chen, et al. (2005). "Irreversibly inhibitory kinetics of 3,5-dihydroxyphenyl decanoate on mushroom (*Agaricus bisporus*) tyrosinase." Bioorg Med Chem **13**(22): 6206-11.
- Rhodes, G. (2000). "Crystallography made crystal clear : a guide for users of macromolecular models." Academic.
- Riedl, S. J., M. Renatus, et al. (2001). "Structural basis for the inhibition of caspase-3 by XIAP." Cell **104**(5): 791-800.
- Rotonda, J., D. W. Nicholson, et al. (1996). "The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis." Nat Struct Biol **3**(7): 619-25.
- Salvesen, G. S. and J. M. Abrams (2004). "Caspase activation - stepping on the gas or releasing the brakes? Lessons from humans and flies." Oncogene **23**(16): 2774-84.
- Samudrala, R. and J. Moult (1998). "An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction." J Mol Biol **275**(5): 895-916.
- Sanchez Mejia, R. O. and R. M. Friedlander (2001). "Caspases in Huntington's disease." Neuroscientist **7**(6): 480-9.
- Sayle, R. A. and E. J. Milner-White (1995). "RASMOL: biomolecular graphics for all." Trends Biochem Sci **20**(9): 374.
- Schulz, J. B., M. Weller, et al. (1999). "Caspases as treatment targets in stroke and neurodegenerative diseases." Ann Neurol **45**(4): 421-9.
- Schweizer, A., C. Briand, et al. (2003). "Crystal structure of caspase-2, apical initiator of the intrinsic apoptotic pathway." J Biol Chem **278**(43): 42441-7.
- Sheldrick, R. W. S., T. R. (1997). "High resolution refinement." Methods Enzymol **277**: 319-343.
- Sherwood, D. (1976). "Crystals, X-rays and Proteins."

- Simons, K. T., R. Bonneau, et al. (1999). "Ab initio protein structure prediction of CASP III targets using ROSETTA." Proteins Suppl **3**: 171-6.
- Simons, K. T., C. Kooperberg, et al. (1997). "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions." J Mol Biol **268**(1): 209-25.
- Sippl, M. J. (1990). "Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins." J Mol Biol **213**(4): 859-83.
- Soengas, M. S., R. M. Alarcon, et al. (1999). "Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition." Science **284**(5411): 156-9.
- Stennicke, H. R., M. Renatus, et al. (2000). "Internally quenched fluorescent peptide substrates disclose the subsite preferences of human caspases 1, 3, 6, 7 and 8." Biochem J **350 Pt 2**: 563-8.
- Stennicke, H. R. and G. S. Salvesen (1997). "Biochemical characteristics of caspases-3, -6, -7, and -8." J Biol Chem **272**(41): 25719-23.
- Tacconi, S., R. Perri, et al. (2004). "Increased caspase activation in peripheral blood mononuclear cells of patients with Alzheimer's disease." Exp Neurol **190**(1): 254-62.
- Takuma, H., T. Tomiyama, et al. (2004). "Amyloid beta peptide-induced cerebral neuronal loss is mediated by caspase-3 in vivo." J Neuropathol Exp Neurol **63**(3): 255-61.
- Talanian, R. V., C. Quinlan, et al. (1997). "Substrate specificities of caspase family proteases." J Biol Chem **272**(15): 9677-82.
- Thompson, C. B. (1995). "Apoptosis in the pathogenesis and treatment of disease." Science **267**(5203): 1456-62.

- Thornberry, N. A., H. G. Bull, et al. (1992). "A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes." *Nature* **356**(6372): 768-74.
- Thornberry, N. A., T. A. Rano, et al. (1997). "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* **272**(29): 17907-11.
- Timmer, J. C. and G. S. Salvesen (2007). "Caspase substrates." *Cell Death Differ* **14**(1): 66-72.
- Volkman, X., M. Cornberg, et al. (2006). "Caspase activation is required for antiviral treatment response in chronic hepatitis C virus infection." *Hepatology* **43**(6): 1311-6.
- Vucic, D., H. R. Stennicke, et al. (2000). "ML-IAP, a novel inhibitor of apoptosis that is preferentially expressed in human melanomas." *Curr Biol* **10**(21): 1359-66.
- Wang, Y., L. Guan, et al. (2005). "Dipeptidyl aspartyl fluoromethylketones as potent caspase inhibitors: peptidomimetic replacement of the P2 alpha-amino acid by a alpha-hydroxy acid." *Bioorg Med Chem Lett* **15**(5): 1379-83.
- Watt, W., K. A. Koeplinger, et al. (1999). "The atomic-resolution structure of human caspase-8, a key activator of apoptosis." *Structure* **7**(9): 1135-43.
- Weber, I. T., B. Fang, et al. (2008). "Caspases: structure-guided design of drugs to control cell death." *Mini Rev Med Chem* **8**(11): 1154-62.
- Wilson, K. P., J. A. Black, et al. (1994). "Structure and mechanism of interleukin-1 beta converting enzyme." *Nature* **370**(6487): 270-5.
- Xia, Y., E. S. Huang, et al. (2000). "Ab initio construction of protein tertiary structures using a hierarchical approach." *J Mol Biol* **300**(1): 171-85.
- Yan, N. and Y. Shi (2005). "Mechanisms of apoptosis through structural biology." *Annu Rev Cell Dev Biol* **21**: 35-56.

- Yang, L., Z. Cao, et al. (2004). "Tumor-specific gene expression using the survivin promoter is further increased by hypoxia." Gene Ther **11**(15): 1215-23.
- Yoshimori, A., J. Sakai, et al. (2007). "Structural and functional definition of the specificity of a novel caspase-3 inhibitor, Ac-DNLD-CHO." BMC Pharmacol **7**: 8.
- Zhang, C., S. Liu, et al. (2004). "An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state." Protein Sci **13**(2): 400-11.
- Zhang, N., H. Hartig, et al. (2005). "The role of apoptosis in the development and function of T lymphocytes." Cell Res **15**(10): 749-69.
- Zhang, Y. H., H. J. Zhang, et al. (2006). "Isoquinoline-1,3,4-trione and its derivatives attenuate beta-amyloid-induced apoptosis of neuronal cells." FEBS J **273**(21): 4842-52.
- Zhou, H. and Y. Zhou (2002). "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction." Protein Sci **11**(11): 2714-26.
- Zhou, Y., H. Zhou, et al. (2006). "What is a desirable statistical energy function for proteins and how can it be obtained?" Cell Biochem Biophys **46**(2): 165-74.

APPENDICES

I. List of All Caspase Complexes Studied and Publications

Protein	Inhibitor	PDB code	Purification	Crystallization	Solve	Enzyme Kinetics
Caspase-3	Ac-DEVD-Cho	2H5I	#	#	#	#
Caspase-3	Ac-DMQD-Cho	2H5J	#	#	#	#
Caspase-3	Ac-VDVAD-Cho	2H65	#	#	#	#
Caspase-3	Ac-IEPD-Cho	3GJT	#	#	#	#
Caspase-3	Ac-WEHD-Cho	3GJQ	#	#	#	#
Caspase-3	Ac-YVAD-Cho	3GJS	*	*	*	#
Caspase-3	Boc-D(OMe)-Fmk	3GJR	#	#	#	#
Caspase-3	Ac-LDESD-Cho	3EDQ	*	*	*	#
Caspase-7	Ac-DMQD-Cho	2QL5	*			#
Caspase-7	Ac-DQMD-Cho	2QL9	*			#
Caspase-7	Ac-DNLD-Cho	2QLF	*			#
Caspase-7	Ac-IEPD-Cho	2QL7	*			#
Caspase-7	Ac-ESMD-Cho	2QLB	*			#
Caspase-7	Ac-WEHD-Cho	2QLJ	*			#
Caspase-7	Ac-YVAD-Cho	3IBC	*			#
Caspase-7	Ac-LDESD-Cho	3EDR	*			#

full contribution *partial contribution N/A not published yet

Related publications are listed below.

1. Agniswamy J, Fang B, Weber IT. Catalytic restrictions in the active site of human caspase-7 revealed by unliganded and inhibited structures. *Apoptosis*. 2009 14(10):1135-44.
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 May 14(5):741-752.
3. Weber IT, Fang B, Agniswamy J. Caspases: structure-guided design of drugs to control cell death. *Mini Rev Med Chem* 2008;8: 1154-1162.
4. 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;13: 1291-1302.
5. Agniswamy J, Fang B, Weber IT. Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis. *FEBS J*. 2007 Aug 14
6. Fang B, Boross PI, Tozser J, Weber IT. Structural and kinetic analysis of caspase-3 reveals role for S5 binding site in substrate recognition. *J Mol Biol*. 2006 Jul 14;360(3):654-66.

II. Mutagenesis Study in the S4 Binding Site of Caspase-3

Enzyme kinetic studies have showed that caspase-3 prefers the aspartic acid residue at P4 position on its substrate and any substitution dramatically decrease substrate binding affinity. Our molecular modeling study in chapter three, however, suggested that the glutamic acid can bind in the S4 pocket and form hydrogen bonds with F250 and S209. Since Glu and Asp only differ from one side chain carbon atom, what make caspase-3 stringently prefer Asp at P4? This question still remains to be a mystery. We observed in several caspase-3 structures that the P4 aspartic acid form water mediated hydrogen bond interactions with some protein residues nearby the S4 subsite, including D221, W214, and Q217. In order to examine the effect of these indirect

interactions to the P4 binding, mutagenesis and enzyme kinetic studies were performed. In addition, crystallization of caspase-3 mutant was also performed.

Three caspase-3 mutants, D221A, W214A, and Q217A were made following the standard site-directed mutagenesis protocol using PCR. The following protein expression, purification, and kinetic analysis were performed using the same protocol described in previous chapters.

Three caspase-3 mutants were successfully constructed and expressed. However, the yield of purified enzyme is low. The enzyme kinetic study indicated that all three caspase-3 mutants have very low catalytic activities on the canonical substrate Ac-DEVD-pNA. Since the exclusion of one water mediated hydrogen bond should not substantially decrease the catalytic activity, we suspect that the mutated residues must be important for the correct folding of the substrate binding pocket. The active site conformation of caspase-3 might be changed by the mutation leading to the dropped activity. This explains the low yield of protein purification as caspase-3 needs to be processed to the active form by its own proteolytic activity. However, the crystallization was unsuccessful, thus structural evidence of our hypothesis is still missing.