THE PROTEIN SOCIETY

Phage display and structural studies reveal plasticity in substrate specificity of caspase-3a from zebrafish

Matthew B. Tucker,[1] Sarah H. MacKenzie,[1] Joseph J. Maciag,[1] Hayley Dirscherl Ackerman,[2] Paul Swartz,[1] Jeffrey A. Yoder,[2] Paul T. Hamilton,[3] and A. Clay Clark[4]*

[1]Department of Molecular and Structural Biochemistry, NC State University, Raleigh, North Carolina 27608
[2]Department of Molecular Biomedical Sciences, NC State University, Raleigh, North Carolina 27608
[3]Department of Plant and Microbial Biology, NC State University, Raleigh, North Carolina 27608
[4]Department of Biology, University of Texas at Arlington, Arlington, Texas 76019

Received 1 May 2016; Revised 2 August 2016; Accepted 25 August 2016
DOI: 10.1002/pro.3032
Published online 31 August 2016 proteinscience.org

Abstract: The regulation of caspase-3 enzyme activity is a vital process in cell fate decisions leading to cell differentiation and tissue development or to apoptosis. The zebrafish, *Danio rerio*, has become an increasingly popular animal model to study several human diseases because of their transparent embryos, short reproductive cycles, and ease of drug administration. While apoptosis is an evolutionarily conserved process in metazoans, little is known about caspases from zebrafish, particularly regarding substrate specificity and allosteric regulation compared to the human caspases. We cloned zebrafish caspase-3a (*casp3a*) and examined substrate specificity of the recombinant protein, Casp3a, compared to human caspase-3 (CASP3) by utilizing M13 bacteriophage substrate libraries that incorporated either random amino acids at P5-P1' or aspartate fixed at P1. The results show a preference for the tetrapeptide sequence DNLD for both enzymes, but the P4 position of zebrafish Casp3a also accommodates valine equally well. We determined the structure of zebrafish Casp3a to 2.28Å resolution by X-ray crystallography, and when combined with molecular dynamics simulations, the results suggest that a limited number of amino acid substitutions near the active site result in plasticity of the S4 sub-site by increasing flexibility of one active site loop and by affecting hydrogen-bonding with substrate. The data show that zebrafish Casp3a exhibits a broader substrate portfolio, suggesting overlap with the functions of caspase-6 in zebrafish development.

Keywords: phage display substrate library; substrate recognition; Danio rerio; apoptosis

Abbreviations: Ac-DEVD-AFC, acetyl-Asp-Glu-Val-Asp-7-amino-4-trifluoromethylcoumarin; Ac-DEVD-AMC, acetyl-Asp-Glu-Val-Asp-7-amino-4-methylcoumarin; Ac-DEVD-CMK, acetyl-Asp-Glu-Val-Asp-chloromethyl ketone; MD molecular dynamics; protomer, large and small subunit obtained by processing a monomer of procaspase-3; WT, wild-type

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH (National Institutes of Health); Grant number: GM065970; Grant sponsor: NIH Biotechnology Traineeship; Grant number: T32 GM008776; Grant sponsor: U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences; Grant number: W-31-109-ENG-38.

*Correspondence to: A. Clay Clark, Department of Biology, Box 19498, 501 S. Nedderman Drive, 337 Life Science Building, University of Texas at Arlington, Arlington, TX 76019. E-mail: clay.clark@uta.edu

Introduction

Apoptosis is a process that ensures normal cell turnover in multicellular organisms, and a number of diseases arise due to the improper regulation of apoptosis. Cancer, for instance, results from insufficient apoptosis, and targeting cell death pathways represents a major effort in developing new cancer therapies.[1–3] In contrast, excess apoptosis is problematic in neurodegenerative diseases.[4,5] Understanding how cells regulate caspase activity will be important for developing therapeutic strategies that return caspase activity to normal levels required for cell development and cell death. Caspase-3 is the primary executioner in apoptosis, but cells also use caspase-3 enzyme activity for a variety of physiological reactions (adaptive responses) when activity is held at levels lower than that required for apoptosis.[1] Adaptive responses such as cell differentiation[6] and cell development,[7] particularly of the eye lens[8] and inner ear,[9] rely on the ability to fine-tune caspase activity. Caspase mechanisms in apoptosis have been studied for more than two decades, and the signaling pathways that result in caspase activation are well-known. In contrast, mechanisms used by cells to fine-tune caspase activity for adaptive responses are largely unknown.

Danio rerio, more commonly known as zebrafish, are small tropical fish that have become an increasingly popular animal model to study various human diseases.[10,11] Zebrafish embryos offer visual transparency, short reproductive cycles, ease of drug administration,[10,12] and lower costs compared to mouse and rat models.[13] In addition, most organ systems are fully developed by 5 days post-fertilization (dpf).[14] Although the zebrafish model is widely used to investigate signaling pathways, tissue, and organ development, apoptosis, and disease development in vertebrates,[12] the zebrafish caspases are largely uncharacterized.

Caspases are cysteinyl aspartate-specific proteases that recognize aspartate at the P1 position of the substrate (using the nomenclature of Schechter and Berger),[15] and the specificity among the various caspases is determined primarily by the amino acid in the P4 position.[1] In human caspase-3 (CASP3), the preferred amino acid sequence is DEVD, and more generally, CASP3 recognizes sequences with a DXXD motif.[16,17] Human caspases cleave at least four hundred cellular proteins[18] and although techniques developed to identify CASP3 substrates, such as N-terminomics,[19,20] should also apply to zebrafish embryos, comparative studies in zebrafish have not been performed. To date, only a limited number of caspase substrates have been identified in zebrafish, although current evidence demonstrates overlap with homologues from human cells.[12]

While apoptosis is an evolutionarily conserved process, caspase activation and regulation is not always identical among species.[21] Zebrafish share at least seven caspases with humans, including orthologues of the three executioner caspases: caspase-3, caspase-6, and caspase-7.[22,23] Zebrafish also encode Bcl-2 family members, inhibitor of apoptosis proteins (IAP), death receptors and ligands, and apoptosis-related kinases.[24,25] Aside from genomics data[26]; however, the regulation of caspases in development or apoptosis in zebrafish has not been studied extensively.

Two orthologs of human CASP3 are encoded in the zebrafish genome, *caspase-3a* (*casp3a*) and *caspase-3b* (*casp3b*), where the encoded enzymes share 61.8% and 56% sequence identity, respectively, with human CASP3 [Fig. 1(A)]. Zebrafish *casp3a* and *casp3b* are predicted to be paralogues that arose from a common progenitor sequence when a teleost-specific whole genome duplication event occurred ~350 million years ago (Supporting Information Fig. S1).[27] The functional roles of the two zebrafish caspase-3 enzymes have not been determined, but in the Japanese rice fish, medaka (*Oryzias latipes*), the expression pattern of the two genes is distinctly different in early embryo development, where one caspase-3 gene is expressed in the tailbud while the other is expressed in the head.[28,29] In medaka, it appears that apoptosis is dependent on both caspase-3 paralogues. Yamashita et al. initially characterized *casp3a* (initially named *casp3*) from zebrafish and demonstrated that *casp3a* transcripts are present in early embryos as a maternal factor, while the gene is expressed throughout the body after gastrulation by zygotic expression.[30] The role of zebrafish *casp3b* (previously named *casp3l*), if any, in development is currently unknown, but *casp3a* was shown to induce apoptosis when transfected in cultured fish cells,[30] zebrafish embryos,[30] *Xenopus,*[31] rat,[32] and HEK 293T cells.[33] The zebrafish Casp3a enzyme was examined against a limited set of synthetic tetrapeptide substrates and showed highest activity with Ac-DEVD-MCA.[30]

In the studies presented here, we further examined the substrate specificity of zebrafish Casp3a compared to human CASP3 using a substrate-phage display library,[34,35] and we report the first X-ray crystal structure of Casp3a. The data show that both human CASP3 and zebrafish Casp3a exhibit a preference for the tetrapeptide sequence DNLD, but Casp3a also accommodates valine at the P4 position. Structural and molecular dynamics studies of Casp3a suggest increased flexibility due to changes in several amino acids near the active site compared to human CASP3. Overall, the data further establish substrate-phage techniques as an alternative method to positional scanning synthetic combinatorial peptide libraries (PS-SCL) for determining enzyme specificity. The results show that zebrafish Casp3a may have functional overlap with Casp6 due to relaxed substrate specificity in the P4 site. Despite nearly 400 million years of evolution between zebrafish and human, the caspase 3 proteins share similar traits in

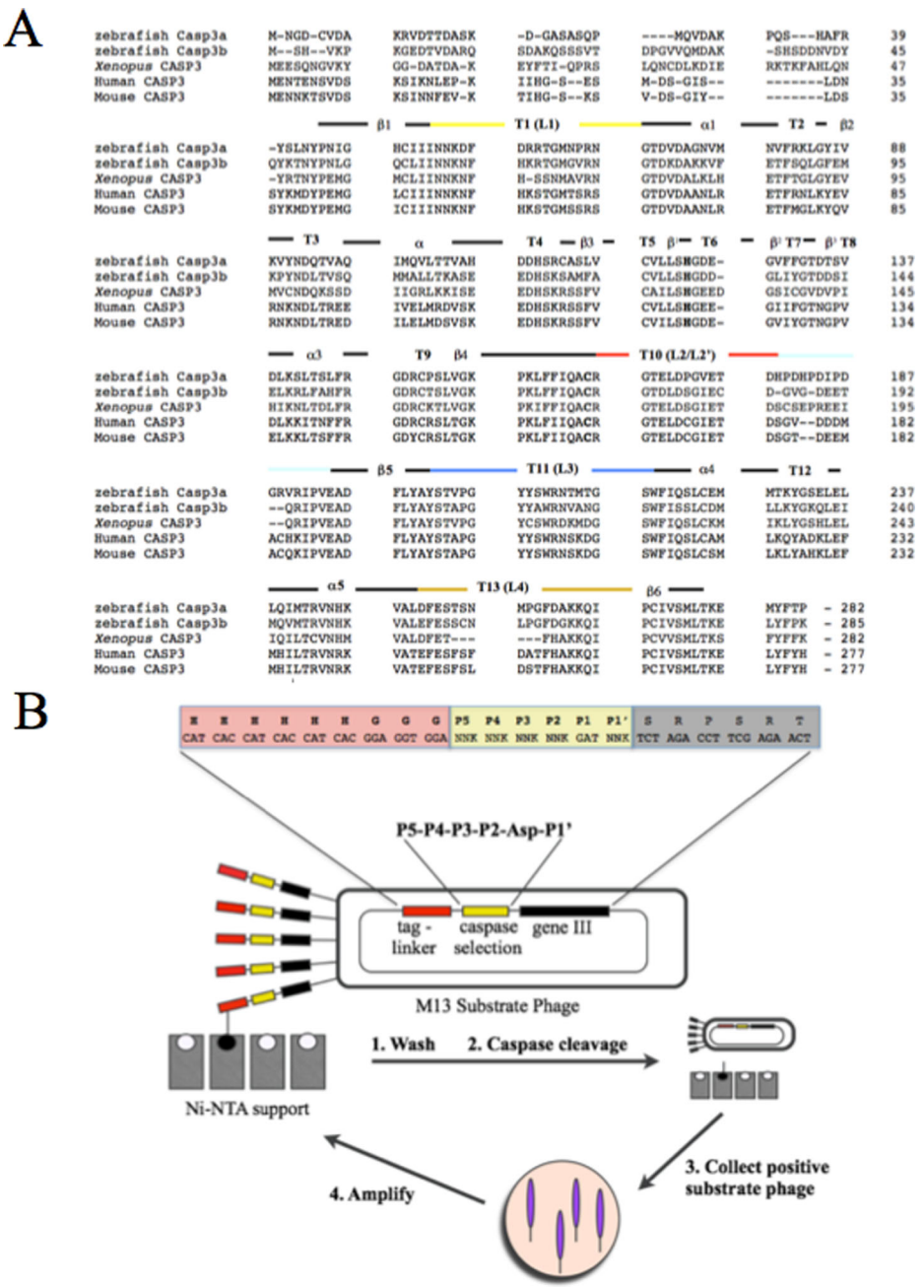

Figure 1. Zebrafish Casp3a used in substrate-phage display methods. A: Sequence alignment of caspase-3 proteins. All sequences are from the Universal Protein Resource Knowledgebase (UniProtKB).[60] The following UniProt accession codes were used for the caspase-3 alignment: zebrafish Casp3b, Q0PKX2; *Xenopus*, P55866; human, P42574; mouse, P70677. The sequence for zebrafish zCasp3a was determined as described in the text (GenBank KX084794). Regions of secondary structure are shown above the sequence, and the colors are the same as those in Figure 3A-B. The catalytic histidine and cysteine residues are shown in bold. B: Schematic of substrate-phage display method for determining caspase substrate specificity using the X_4DX library.

substrate specificity and predicted allosteric sites. Differences in specificity may result from only a few changes in amino acids near the active site.

Results and Discussion

Determining caspase-3 substrate specificity from phage-substrate display libraries

A *casp3a* transcript was cloned from a single zebrafish of the AB line, and the encoded protein differs from the published sequence[30] at three positions: 57 (N->D), 180 (T->P), and 190 (E->V) [Fig. 1(A)]. Aspartate 57 is in active site loop 1 (L1), and the crystal structure of Casp3a (described below) shows that the amino acid overlays closely to N54 of human CASP3. In contrast, residues 180 and 190 are located in the intersubunit linker (IL) and are not observed in the crystal structure. The zebrafish Casp3a zymogen is cleaved at D178 during activation, and there is no evidence that the substitutions

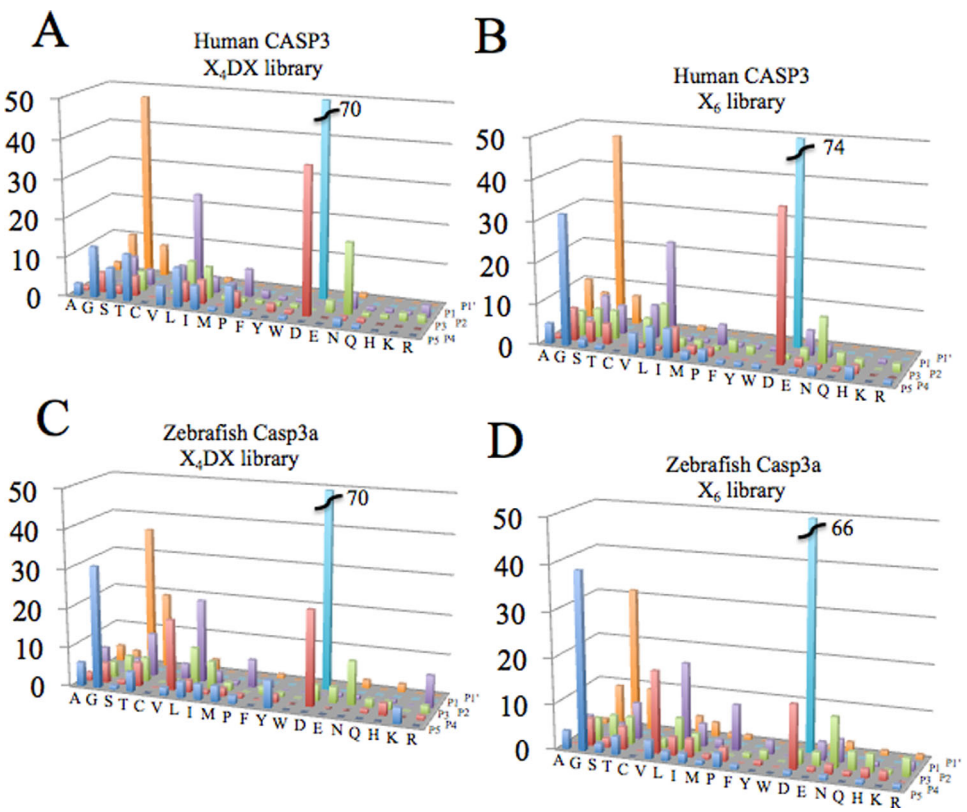

Figure 2. Results of substrate-phage display to determine substrate specificity for human and zebrafish caspase-3. Two libraries were used for human CASP3 (A-B) or zebrafish Casp3a (C-D) in which aspartate was fixed at the P1 position, called the X_4DX library, or six random amino acids were examined, called the X_6 library. Values on the *y*-axis represent number of phage with the representative sequence, while amino acids are labeled on the *x*-axis. The six positions - P5 (blue), P4 (red), P3 (green), P2 (purple), P1 (cyan) and P1' (orange) - are shown on the *z*-axis.

in the IL affect activation or enzyme activity compared to the previously published sequence. We expressed and purified recombinant zebrafish Casp3a with a C-terminal histidine tag and examined enzyme activity against the tetrapeptide substrate, Ac-DEVD-AFC. The data show a specificity constant (k_{cat}/K_M) of 3.4×10^5 M^{-1} s^{-1}, with k_{cat} of 7.5 ± 0.14 s^{-1} and K_M of 22 ± 0.9 μM (Supporting Information Fig. S2). Compared to human CASP3,[36–38] the activity of zebrafish Casp3a in cleaving the peptide substrate is approximately three times higher due to a ~10-fold higher k_{cat}, while the K_M is also higher than that for human CASP3.

In order to examine substrate specificity for zebrafish Casp3a, we utilized a substrate-phage display library [Fig. 1(B)]. In this technique, the M13 bacteriophage displays an N-terminal six-histidine tag, a bridge consisting of three glycines, and either control sequences or P5-P4-P3-P2-P1-P1' substrate amino acids followed by the remainder of the pIII sequence [Fig. 1(B)]. The control sequences were (P4-P1') DEVDG, DEVEG, and DEVAG, and the random libraries expressed were XXXXDX (X_4DX) and XXXXXX (X_6), where the former (X_4DX) contained an aspartate residue fixed at P1 and the latter (X_6) contained six random amino acids at P5-P1'. Control studies of elution with imidazole showed bacteriophage plaque forming units (PFU) of 5-8x10^{10} on the nickel-sepharose (Ni-NTA) beads, while washing with PBS demonstrated a low removal of bacteriophage, 3-10x10^4 PFU (Supporting Information Table S2). The imidazole treatment represents the upper limit of phage bound to the Ni-NTA column, while the PBS treatment represents the lower limit of phage release from the beads. Both CASP3 and Casp3a showed a preference for a P1 aspartate in the control libraries, by 1-2 orders of magnitude, compared to glutamate or alanine. We note that sequencing phage cleaved in the DEVEG and DEVAG control libraries showed that the enzyme selected the GGG linker and P4 aspartate, thus cleaving the GGGD sequence in the GGGDEVEG or GGGDEVAG control libraries. Both enzymes displayed high activity in the X_4DX and X_6 libraries (Table S2). Finally, inhibiting human CASP3 and zebrafish Casp3a with Ac-DEVD-CMK resulted in basal levels of phage removal (~10^4 PFU, Table S3), demonstrating that release of phage in the control libraries was due to proteolytic cleavage by the caspase-3 enzyme.

Substrate sequences were determined after 3-5 rounds of selection, and the data showed no

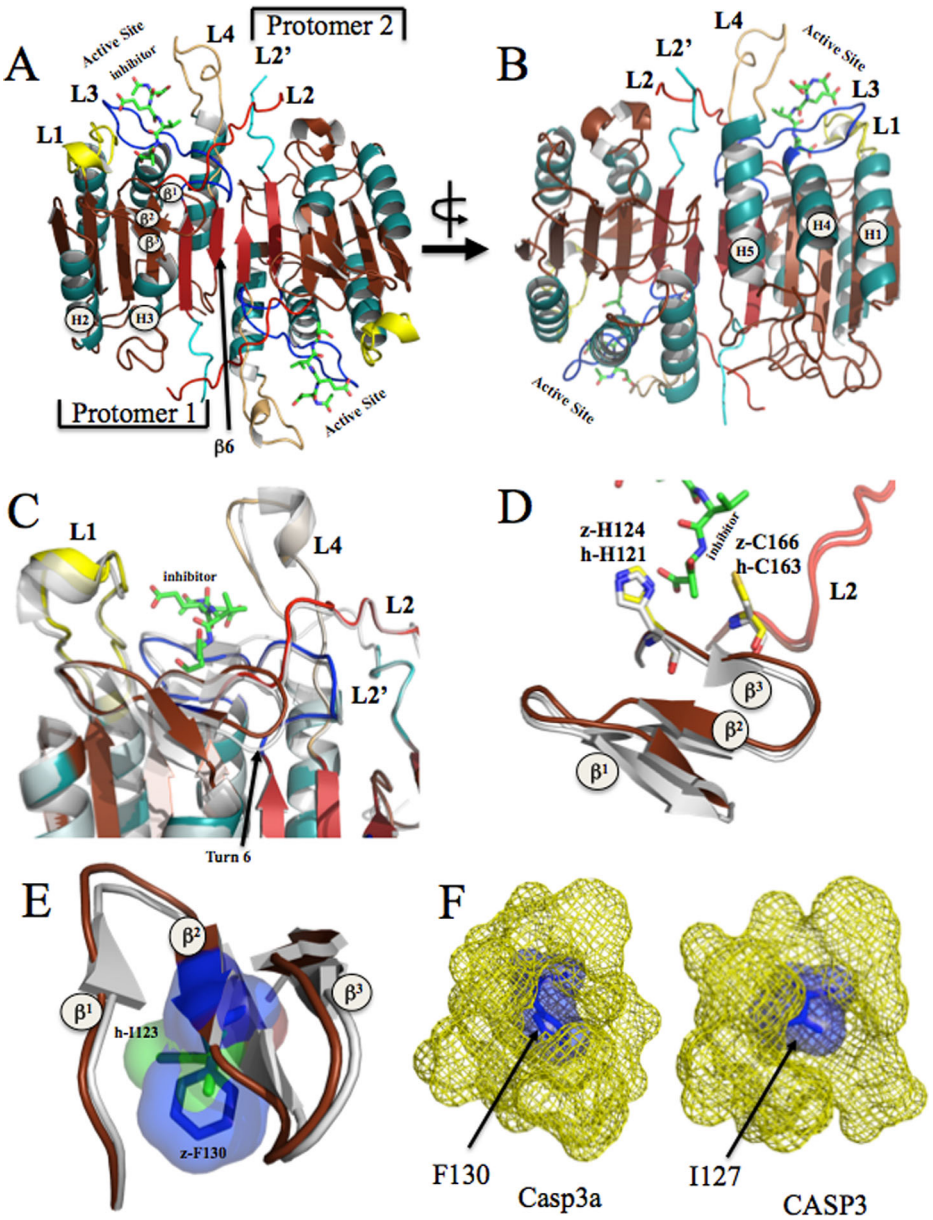

Figure 3. Comparison of human and zebrafish caspase-3 structures. A,B: Structure of zebrafish Casp3a determined by X-ray crystallography. The active site loops (L1-L4, L2') and regions of secondary structure are labeled for one protomer of the dimer. C: Comparison of active sites for human CASP3 (gray) and zebrafish Casp3a (color). D: The short surface β-strand (β^1-β^3) of zebrafish Casp3a (color) is displaced compared to that of human CASP3 (gray), although the catalytic histidine and cysteine (shown as sticks) overlay very closely. E: The displacement of β^2 may be caused by the amino acid substitution F130 in zebrafish Casp3a (blue) compared to I123 in human CASP3 (green). As in panels C-D, human CASP3 is shown in gray. F: Surface representation of F130 (zebrafish) or I127 (human) caspase 3, shown as blue spheres. Atoms within 5 Å are shown as yellow mesh.

difference between rounds 4 and 5, so the results for those rounds were combined. For both enzymes, there was little difference in selection from the X_4DX or X_6 libraries. In the case of human CASP3 [Fig. 2(A,B)], the data from both libraries showed a preference for small amino acids at the P5 position, particularly glycine. There was a clear preference for aspartate at the P4 position and a slight preference for asparagine at the P3 position, although valine, leucine, and isoleucine were also observed. There was a preference for leucine in the P2 position, and the random library (X_6) showed a preference for aspartate at P1. Finally, either glycine or serine was preferred at P1', so the consensus sequence for CASP3 from the data in Figure 2(A,B) is (P5-P1') GDNLD(G/S).

An analysis of the data for zebrafish Casp3a [Fig. 2(C,D)] showed a similar consensus sequence except that both valine and aspartate were selected in the P4 position. Aspartate was slightly preferred over valine in the X_4DX library [Fig. 2(C)] while the converse was true in the X_6 library [Fig. 2(D)]. In

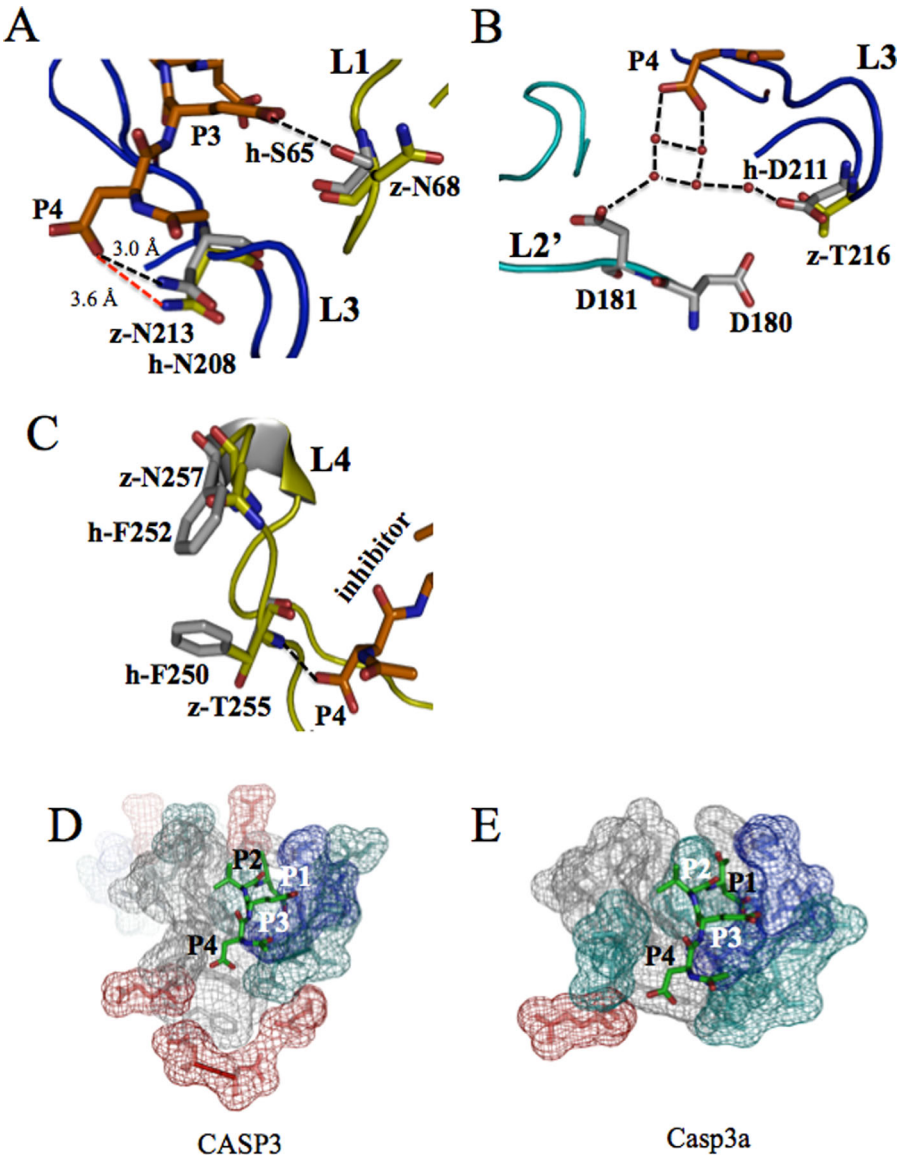

Figure 4. Differences in active site residues of zebrafish and human caspase-3. A: The substitution of N68 in zebrafish Casp3a results in fewer hydrogen bonds to the P3 aspartate, while changes in active site loop 3 (L3) result in weaker hydrogen bonds to the P4 aspartate. B: The substitution of T216 in zebrafish Casp3a removes a water-mediated hydrogen-bonding network between L3, L2' and the P4 aspartate residue. C: Two substitutions in active site loop 4 (L4) (T255 and N257) of Casp3a compared to CASP3. For panels A-C, amino acids found in human CASP3 are shown in grey. D,E: Amino acids (shown in mesh) within 5 Å of the P4-P1 residues (sticks) demonstrate changes in polarity of the active site for human CASP3 (panel D) versus zebrafish (panel E) Casp3a. Positively charged amino acids are shown in blue, polar amino acids are shown in cyan, negatively charged amino acids are shown in red, and hydrophobic amino acids are shown in gray.

addition, there was a clear preference for serine over glycine at the P1' position. So, for Casp3a, the consensus sequence determined from the substrate-phage display method is (P5-P1') G(V/D)NLDS.

In order to further confirm the broader specificity for Casp3a, we examined enzyme activity against the caspase-6 tetrapeptide substrate, Ac-VEID-AMC (Supporting Information Fig. S2). The data show that Casp3a is active against the caspase-6 substrate, whereas CASP3 did not cleave the caspase-6 substrate. We determined a specificity constant (k_{cat}/K_M) for Casp3a against Ac-VEID-AMC of 4.8×10^4 $M^{-1}sec^{-1}$, with k_{cat} of 2.0 ± 0.2 sec^{-1} and K_M of 42 ± 3 μM (Supporting Information Fig. S2). The data are very similar to those determined previously for the cleavage of Ac-VEID-AMC by human caspase-6,[39] where Hardy and colleagues report k_{cat} of 0.92 s^{-1} and K_M of 45 μM ($k_{cat}/K_M = 2.0\text{x}10^4$ $M^{-1}sec^{-1}$).

Crystal structure of zebrafish Casp3a reveals plasticity in P4 site

We determined the X-ray crystal structure of zebrafish Casp3a to 2.28Å resolution (Table S4), and the

data [Fig. 3(A)] show a very close alignment with human CASP3 (PDB ID 2J30),[40] with RMSD of 0.46Å. Changes in two regions of the protein near the active site, however, may affect either allosteric regulation or enzyme specificity. While all caspase structures are similar overall,[1] a comparison of Casp3a with human caspases-6 (PDB ID 3OD5)[41] and -7 (PDB ID 1F1J)[42] show RMSD of 0.83Å and 0.47Å, respectively. This is again consistent with Casp3a being structurally similar to CASP3 and the close homologue, caspase-7.

The short surface β-sheet, β^1-β^3, of human CASP3 is important for positioning the catalytic histidine for proton transfer to the catalytic cysteine.[1] This region of the protein contains Turn 6, which in CASP3 is important for stabilizing active site loop 2 (L2) through electrostatic interactions in the dimer interface. In addition, the loop undergoes large conformational changes upon activation of the zymogen.[43] In caspase-9, an insertion in Turn 6 results in a longer loop that occludes the central cavity,[44] which is a common allosteric site in caspases.[45,46] Notably, the short surface β-strand undergoes a strand-to-helix transition in caspase-6, which inhibits the enzyme by increasing the distance between the catalytic groups.[47] In addition, the surface β-strand connects to helix 3, near the dimer interface [Fig. 3(A,B)]. Rotation of the helix toward the central cavity of the dimer interface inactivates the enzyme by distorting the active site groups. Movements of helix 3 are part of a mechanism that stabilizes a high-energy inactive state in the native ensemble of human CASP3.[37,48]

While the catalytic groups in zebrafish Casp3a align closely to those in human CASP3, the short surface β-sheet, particularly β^2 and β^3, is misaligned compared to CASP3 [Fig. 3(C,D)]. Changes in the β-strand appear to be due to the substitution of phenylalanine in Casp3a (F130) from isoleucine in CASP3 (I123). The larger amino acid in Casp3a results in improved hydrophobic contacts in the β-strand [Fig. 3(E,F)]. We note that F130 is unique to Casp3a, where isoleucine is observed in other homologues [Fig. 1(A)]. At present, it is not known if helix 3 in zebrafish Casp3a undergoes similar conformational transitions as observed in human CASP3, that is, the static structure does not provide information on changes to the native ensemble compared to CASP3, but the improved hydrophobic contacts in Casp3a may prevent active site distortions as observed in CASP3.

The second area of difference between zebrafish Casp3a and human CASP3 is in the S3 and S4 sites (Fig. 4). In CASP3, the side-chain of the P3 glutamate forms two hydrogen bonds in the S3 site, one each with R207 and S65. In Casp3a, the P3 glutamate forms only one hydrogen bond, to R212 in the S3 site. This is due to the substitution of serine to asparagine in Casp3a [S65 -> N68, Fig. 4(A)]. In this case, the side-chain of N68 is rotated toward solvent and is not positioned to hydrogen bond to the P3 aspartate. We note that although R212 is conserved in caspases [Fig. 1(A)], N68 is found in both zebrafish and *Xenopus* caspase-3 enzymes, but serine is observed in human and mouse caspase-3 [Fig. 1(A)]. In addition, the position of active site loop 3 (L3) in zebrafish Casp3a varies in the S4 site such that the asparagine at the beginning of the turn no longer forms a hydrogen bond with the P4 aspartate, even though the amino acid is conserved (N208 in CASP3 *versus* N213 in Casp3a) [Fig. 4(A)]. This is likely due to a substitution at the end of the turn, D211->T216 in Casp3a. In CASP3, D211 forms part of a hydrogen bonding network between water molecules and the P4 aspartate. The network also includes the N-terminus of loop 2' from the second protomer [Fig. 4(B)], which is generated upon cleavage of the zymogen at D175. The water-mediated hydrogen-bonding network stabilizes the active site loops, L2' and L3. In Casp3a, the substitution of T216 results in only one hydrogen bond to the P4 aspartate side-chain, where one observes a hydrogen bond between the amide nitrogen of T255 in active site loop 4 (L4) and the P4 aspartate [Fig. 4(C)]. As a result, the region of L2' between H179 and R190 [see Fig. 1(A)] is disordered in zebrafish Casp3a.

Finally, two substitutions in active site loop 4 (L4) affect the position of the loop. In human CASP3, F250 and F252 form a hydrophobic cluster in the loop, and replacement by threonine (T255) or asparagine (N257), respectively, in zebrafish Casp3a results in re-positioning of the loop relative to the substrate-binding pocket [Fig. 4(C)] as well as a more polar loop [Fig. 4(D,E)]. The L4 active site loop is a major factor in determining specificity of the P4 amino acid, and with the notable exception of CASP3 from *Xenopus* [Fig. 1(A)], the loop is longest in caspase 3 relative to other caspases.[1] In addition, F250 and F252 in CASP3 were shown to form a hydrophobic cleft for binding valine in the P5 position of a substrate, (P5-P1) VDVAD.[49] We note that in our phage-substrate display studies (Fig. 2), Casp3a displays a lower preference for V/L/I at the P5 position compared to CASP3, which is consistent with the substitutions of the phenylalanines in L4.

Molecular dynamics simulations show increased flexibility in the active site of zebrafish Casp3a

In order to further examine differences in zebrafish Casp3a compared to human CASP3, we performed molecular dynamics (MD) simulations for 50 ns. Compared to our previous data for CASP3,[37] the MD simulations show that the fluctuations in Casp3a are generally within 1-2.5 Å of those observed for CASP3. The data for both protomers in the dimer are reported as ΔRMSF ($RMSF_{(Casp3a)}$ –

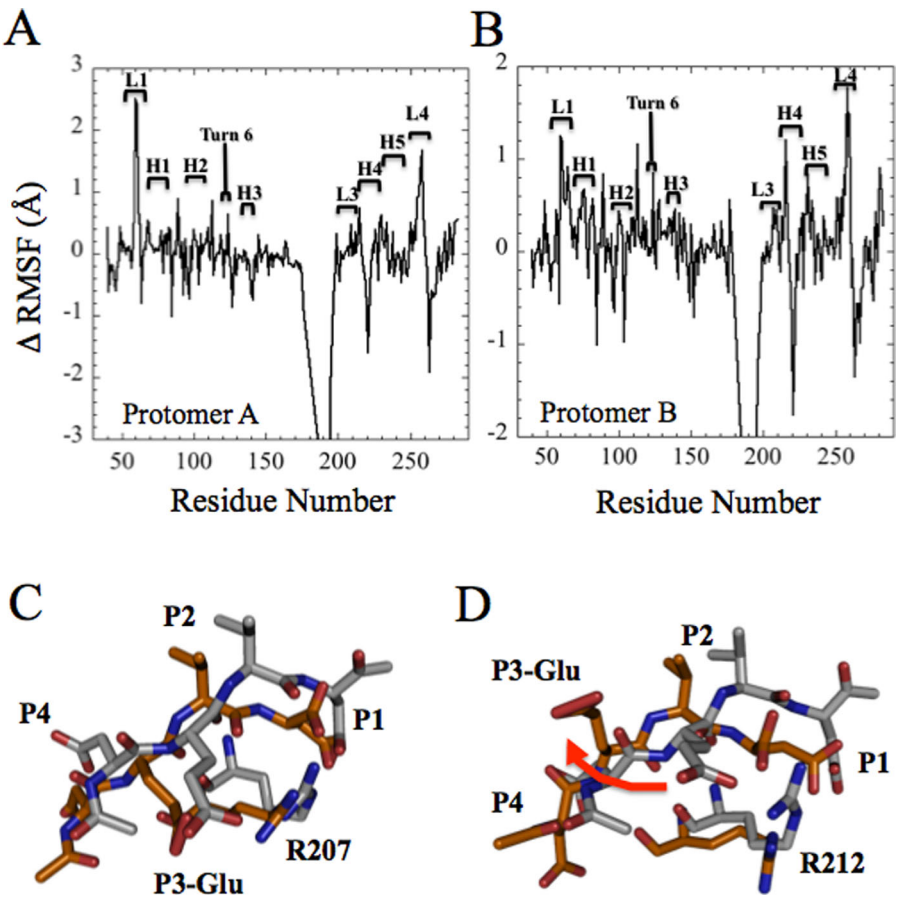

Figure 5. Molecular dynamics simulations of zebrafish Casp3a. Root mean square fluctuations (RMSF) were determined for zebrafish Casp3a from 50 ns MD simulations. Active site loops and surface helices are labeled. Panels A and B represent the two protomers in the dimer. Data are shown as the change in RMSF, calculated as $\Delta RMSF = (RMSF_{(Casp3a)} - RMSF_{(CASP3)})$. Fluctuations in P4-P1 residues for human (Panel C) or zebrafish (Panel D) caspase 3. The starting position, determined from the X-ray crystal structure is shown in gray versus the position at 28.9 ns of the simulation (orange).

$RMSF_{(CASP3)}$), so values greater than zero show larger fluctuations in Casp3a compared to those of CASP3, while values less than zero reflect larger fluctuations in CASP3 (Fig. 5). Overall, the data show that active site loops L1 and L4 are more flexible in Casp3a. Unlike allosteric mutants of CASP3,[37,48] there appears to be little rotation in helix 3 of Casp3a, suggesting that if the high-energy inactive state of the native ensemble is present in Casp3a, then the state is selected by different mechanisms than occurs in the human enzyme.

In the active site of human CASP3, there are few fluctuations of the P1-P4 residues of the inhibitor [Fig. 5(C)], likely due to the extensive hydrogen bonding with amino acids in the S1, S2, and S4 sites. In contrast, the P3 glutamate of zebrafish Casp3a is quite dynamic and is observed to rotate out of the S3 site to become exposed to solvent [Fig. 5(D)]. These movements are consistent with fewer hydrogen bonds in the S3 site, as noted above (Fig. 4). In CASP3, Y204 on active site loop 3 (L3) forms part of the hydrophobic S2 binding pocket. In the unliganded enzyme, Y204 occupies the S2 binding pocket and must rotate out of the pocket prior to ligand binding [Fig. 6(A,B)]. Two tryptophans in the active site, W206 and W214, partially occupy the S4 binding pocket, but minor movements in these two residues open the pocket for ligand binding. The three residues are conserved in Casp3a [Fig. 1(A)], and Y209 behaves similarly to the comparable residue (Y204) in CASP3, that is, it occupies the S2 binding pocket in the unliganded enzyme [Fig. 6(C,D)]. In contrast to W206 in CASP3, however, W211 in Casp3a is highly dynamic and occupies the S4 binding pocket in the unliganded enzyme. Thus, in zebrafish Casp3a, both Y209 and W211 must rotate out of the S2 and S4 binding pockets, respectively, prior to ligand binding. Together, the molecular dynamics data show increased fluctuations in three active site loops, L1, L3, and L4, and in the P3 glutamate of Casp3a compared to the human enzyme. The amino acid substitutions described above (Figs. 3 and 4) change the hydrogen-bonding networks in the active site as well as the hydrophobicity of L4, likely resulting in the increased fluctuations observed in zebrafish Casp3a.

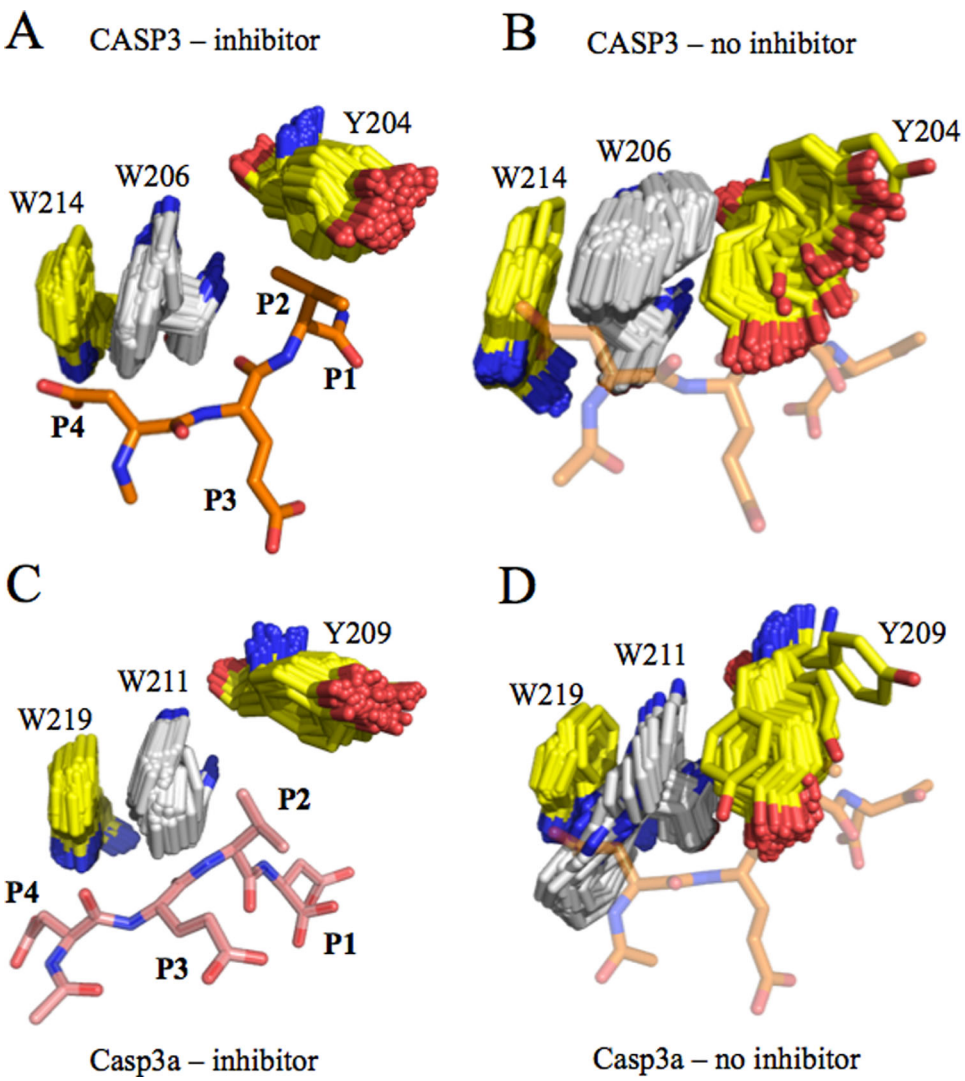

Figure 6. Fluctuations in the caspase 3 active site. For human CASP3 (Panels A-B) or zebrafish Casp3a (Panels C-D), fluctuations for three active site residues are shown as 200 frames (at 250 ps intervals) of the 50 ns simulation. In the absence of inhibitor (Panels B and D), Y204 (Panel B) or Y209 (Panel D) occupy the S2 site, while in zebrafish Casp3a (Panel D), W211 occupies the S4 site. The position of the inhibitor is shown as sticks in Panels A and C and as semi-transparent in Panels B and D for reference.

Conclusions

We have shown that zebrafish Casp3a is structurally very similar to human CASP3, with an RMSD of 0.46Å. Several substitutions in or near the active site affect active site dynamics and result in fewer and/or weaker hydrogen bonds to the P3 and P4 amino acids of the substrate, likely resulting in plasticity of the P4 site. From phage-substrate display methods, we observe that both Casp3a and CASP3 demonstrate substrate specificity of DNLD and that Casp3a accommodates valine or aspartate equally well at the P4 position of the substrate. Similar to data described by Liu and colleagues,[12] Casp3a may overlap functionally with caspase-6 in zebrafish, and they also observed cleavage of the caspase-6 substrate, VEID, by Casp3a.

The structural studies reported here demonstrate that zebrafish Casp3a is very similar to human CASP3 with the noted exceptions in substrate binding. The common allosteric site in the central cavity of the dimer interface, described in human caspases,[45,46] appears to be conserved in Casp3a. A known phosphorylation site in human CASP3, S150, is also conserved in Casp3a [Fig. 1(A)]. We note, however, that a unique allosteric site described in human caspases-6 (CASP6)[50] and caspase-7 (CASP7)[51] differs between CASP3 and Casp3a. In human CASP6, the site evolved for zinc binding, while the site appears to have evolved for peptide binding in human CASP7. In human CASP3, several amino acid substitutions prevent either zinc or peptide binding.[1] A comparison of the amino acids in the site suggest that the site in zebrafish Casp3a is more similar to that of human CASP7 than CASP3 or CASP6, although to date no peptides have been observed to bind at this site.

The zebrafish model offers a number of advantages over mouse and rat models in studies of disease development in vertebrates. The similarities between zebrafish Casp3a and human CASP3 shown here suggest that zebrafish Casp3a may respond to various drugs in a similar fashion to human CASP3. It will be of interest to compare the structures and activities of zebrafish Casp3a and Casp3b to determine if subfunctionalization has occurred for these paralogues. Finally, these data further validates the use of zebrafish caspases in developmental and apoptosis assays.

Materials and Methods

Cloning, expression, and purification. Total RNA was isolated (Trizol, Invitrogen) from a single adult zebrafish from the AB line and reverse transcribed to generate cDNA using SuperScript III Reverse transcriptase (Invitrogen). An initial full-length *casp3a* amplicon was generated from this cDNA by PCR using Titanium Taq DNA polymerase (Clontech) with a forward primer (CTCGTTAAGCGGTTGGAGATG) and a reverse primer (CAAAGTTTCCCTGGTGGTTTTA) that incorporated the translational start and stop codons, respectively. This amplicon was generated using a "touch-down" PCR strategy during which the annealing temperature was lowered from 65°C to 55°C by 0.5°C per cycle for 20 cycles and then an additional 20 cycles were completed with a 55°C annealing temperature. This amplicon was then used as the template to reamplify *casp3a* in a second PCR reaction using Platinum Pfx DNA polymerase (Invitrogen) with nested primers that introduced restriction sites to the ends of amplicon. The forward nested primer introduced a 5' *Nde*I restriction site (GACTCATATGAACGGAGACTGTGTGGAC) and the reverse nested primer replaced the *casp3a* stop codon with a *Xho*I restriction site (GATCCTCGAGAGGAGTGAAGTACATCTC). This amplicon was generated using a PCR strategy using 30 cycles and an annealing temperature of 55°C. The final amplicon was digested with *Nde*I and *Xho*I and ligated into the *Nde*I and *Xho*I sites of pET-21b (EMD Millipore) resulting in a vector encoding Casp3a with a carboxyl-terminal His-Tag. This Casp3a sequence differs from the reference sequence (NP_571952) by three residues at positions 57 (N->D), 180 (T->P), and 190 (E->V) and has been deposited in the GenBank database (accession KX084794). *Escherichia coli* BL21(DE3)pLysS cells were transformed with the plasmid, and Casp3a protein was expressed and purified as previously described for human CASP3.[36,38,52]

Enzyme activity assays

Initial velocity of substrate cleavage was measured at 25°C in a buffer of 150 m*M* Tris-HCl, pH 7.5, 50 m*M* NaCl, 0.1% CHAPS, 1% sucrose and 10 mM DTT in the presence of varying concentrations of Ac-DEVD-AFC or Ac-VEID-AMC substrates, as described previously for human CASP3[38,52] and caspase-6.[39] The total reaction volume was 200 μL and the final enzyme concentration was 10 n*M*. Following the addition of substrate, the samples were excited at 400 nm and emission was monitored at 505 nm for 60 s. The steady-state parameters, K_M and k_{cat}, were determined from plots of initial velocity versus substrate concentration.

Phage display substrate libraries and selection

Caspase substrate libraries were constructed with an amino-terminal hexa-histidine sequence used to bind the phage to a Ni-NTA resin followed by a randomized caspase substrate sequence and the carboxyl-terminal domain of M13 gene III. Each phage particle displays a unique caspase substrate sequence. Phage displaying this library of fusion proteins were bound to the Ni-NTA resin, washed to remove unbound phage and then treated with caspase. Phage containing a good caspase cleavage site were released from the resin whereas phage with sequences that are resistant to cleavage by caspase remain bound to the resin.

The oligonucleotide sequences shown in Supporting Information Table S1 were used to generate three control sequences and two random libraries by annealing the oligonucleotide to a modified single-stranded M13 SAM33 DNA vector for introduction of sequences to the pIII gene, introducing the vector into *E. coli* SS320 cells by electroporation, and growing cells in 2XYT media.[53] After electroporation, 1 mL of SOC media was added to the mixture of electroporated cells, and the solution was transferred to a culture containing 30 mL of SOC media at 37°C. To determine the diversity of the phage display library, a sample was removed from the culture after 25 minutes of incubation, used in a 10-fold dilution series, and each dilution plated on a lawn of *E. coli* cells to determine the phage titer. The titer represents the number of cells originally transformed with phage DNA and therefore the diversity of peptide sequences represented in the phage library. The remaining culture was added to 1 L of 2XYT media and placed in a shaking incubator at 37°C overnight. After overnight growth, the culture was centrifuged at 8,000 rpm for 20 min to remove cells. The phage in the supernatant were recovered by precipitation: addition of 0.25 volumes of a solution of 30% PEG-8000, 1.5*M* NaCl, followed by incubation on ice for 1 h and then centrifugation at 10,000 rpm for 20 min. The supernatant was discarded and the phage pellet was resuspended in PBS. The phage titer was determined, and the phage library was stored at 4°C. For long-term

storage, glycerol was added to 20% final concentration and the phage library was stored at −80°C.

For phage selection, Ni-NTA sepharose beads (100 μL; HisPur™, Thermo Scientific) were centrifuged at 14,000 rpm for 1 min, the supernatant was removed, and the pelleted resin was resuspended in 1 mL blocking buffer (1% BSA, 1X PBS, 1*M* NaCl, and 0.1% Tween-80) in a 1.5 mL microcentrifuge tube. After incubation for 1 h on a rotator at room temperature, the resin was centrifuged at 14,000 rpm for 1 min, the supernatant was removed, and the pelleted resin was resuspended in 250 μL of blocking buffer. A sample of the phage library (250 μΛ; $>10^9$ pfu) was added, and the resin plus phage were mixed at room temperature on a rotator for 45 min. The resin was pelleted by centrifugation, the supernatant removed, and the pelleted resin was washed 10 times with 1 mL of wash buffer (1X PBS, 1*M* NaCl, and 0.1% Tween-80) followed by centrifugation and removal of the wash buffer each time. The resin was then washed 3 times with 1 mL of 1X PBS. Nonspecifically bound and unbound phage were removed with the 13 wash steps described above. After the final wash step, caspase-3 (500 μL of a 500 n*M* solution) was added to the tube, and the reaction was carried out at room temperature for 4 h. Following incubation with the protease, the solution was centrifuged (1 min at 14,000 rpm), and the supernatant containing the released phage particles was collected. An aliquot of the phage-containing supernatant (10 μL) was removed and used to determine the phage titer. The remainder of the released phage were added to 3 mL of 2XYT media that also contained 60 μL of an overnight culture of *E. coli* ER2738 cells, and the solution was incubated at 37°C with shaking for 4 − 5 h. After growth, the cells were removed by centrifugation and the amplified phage in the supernatant were ready for the next round of selection. To start the round of selection, the amplified phage (250 μL; $>10^9$ pfu) were added to Ni-NTA sepharose resin as described above. To determine the number of phage that were released from the resin without protease treatment, one sample was left untreated and incubated in PBS for 4 hours at room temperature. Additionally, a sample was incubated with 500 μL of 250 mM imidazole to release all phage that were bound to the Ni-NTA resin.

To determine the phage titer, phage-containing samples were mixed with 100 μL of an overnight culture of *E. coli* ER2738 cells, added to 3 mL of molten 2xYT top agar at 50°C that contained 2% X-gal and 2% IPTG and then plated onto 2xYT agar plate. After the top agar solidified, the plates were incubated overnight at 37°C. To characterize individual phage, plaques were picked, added to 3 mL of 2xYT broth with 30 μL of an overnight culture of *E. coli* ER2738 cells and grown for 4 hours with shaking at 37°C. The cells were removed by centrifugation, and the phage-containing supernatant was used in PCR amplification of the DNA encoding the putative protease substrate sequence. The PCR amplification was performed using 25 μL of 2x Taq Master mix (Econ-Taq Plus, Lucigen), 18 μL of H_2O, 5 μL of phage-containing supernatant, 1 μL of gene III primer (TTTTTTTTGGAGATTTTCAACGTG) and 1 μL of -96 primer (CCCTCATAGTTAGCGTAACG) for 30 cycles at 95°C for 30 s, 50°C for 45 s, and 72°C for 60 s. The PCR products were purified using a PCR purification kit (QiaQuick, Qiagen). DNA sequence of the PCR products was determined using the gene III primer (Eton Biosciences). From the DNA sequence, the peptide sequence of the caspase substrate site was deduced.

Crystallization and data collection

Zebrafish Casp3a was crystallized as described previously for human CASP3.[37,48,54] Briefly, Casp3a was dialyzed in a buffer of 10 m*M* Tris-HCl, pH 8.5, and 1 m*M* DTT. The protein was concentrated to 10 mg/mL and inhibitor, Ac-DEVD-CMK (reconstituted in DMSO) was added at a 5:1 (w/w) inhibitor/protein ratio. Crystals were obtained at 18°C by the hanging drop vapor diffusion method using 4 μL drops that contained equal volumes of protein and reservoir solutions over a 0.5 mL solution of 100 m*M* sodium citrate, pH 5.1, 17.5% PEG 6000, 10 m*M* DTT, and 3 m*M* NaN_3. Crystals appeared within 14 days and were briefly immersed in cryogenic solution containing 10% MPD (2-methylpentane-2,4-diol) and 90% reservoir solution. Data sets were collected at 100 K at the SER-CAT synchrotron beamline (Advance Photon Source, Argonne National Laboratory, Argonne, IL). The Casp3a crystallized with the symmetry of the orthorhombic space group $P2_12_12_1$ and was phased with a previously published human CASP3 structure (PDB entry 2J30). The inhibitor and all water molecules were removed from the initial model and all B-factors for protein atoms were set to 20 A^3. Inhibitor and water molecules were added in subsequent rounds of refinement performed with COOT[55] and Phenix[56] and were positioned based on 2Fo -Fc and Fo -Fc electron density maps contoured at the 1σ and 3σ levels respectively. A summary of the data collection and refinement statistics is shown in Table S4.

Molecular dynamics simulations

Molecular dynamics simulations were performed as described previously[37] with GROMACS 4.5,[57] using the Amber99 force field[58] and the TIP3P water model.[59] All simulations started with the structure obtained from X-ray crystallography, as described above, and the P1 position of the inhibitor was substituted with aspartate. For simulations in the absence of inhibitor, Ac-DEVD-CMK inhibitor was

removed from the structure file. As described previously for human CASP3,[37] the proteins were solvated in a periodic box of 62 Å x 48 Å x 66 Å, with approximately 13,500 water molecules. Sodium or chloride ions were added as required to neutralize the charge on the system. The system was first minimized using steepest descent, and then the waters were relaxed during a 20 ps MD simulation with positional restraints on the protein. Simulations of 50 ns were then run for each protein (with inhibitor present or with inhibitor absent) under constant pressure and temperature (300 K). A time step of 2 fs was used, and coordinates were saved every 5 ps. In each simulation, the protein was equilibrated within 500 ps.

Accession number

Protein data bank accession number for zebrafish caspase-3a: PDB ID 5JFT.

References

1. Clark AC (2016) Caspase allostery and conformational selection. Chem Rev 111:6666–6706.
2. Butt AJ (2008) Review of: tamoxifen and TRAIL synergistically induce apoptosis in breast cancer cells. Breast Cancer Online 11:e4.
3. Kirkin V, Joos S, Zörnig M (2004) The role of Bcl-2 family members in tumorigenesis. Biochim Biophys Acta 1644:229–249.
4. Wellington CL, Hayden MR (2000) Caspases and neurodegeneration: on the cutting edge of new therapeutic approaches. Clin Genet 57:1–10.
5. Marks N, Berg MJ (1999) Recent advances on neuronal caspases in development and neurodegeneration. Neurochem Int 35:195–220.
6. Schwerk C, Schulze-Osthoff K (2003) Non-apoptotic functions of caspases in cellular proliferation and differentiation. Biochem Pharmacol 66:1453–1458.
7. Crawford ED, Wells JA (2011) Caspase substrates and cellular remodeling. Ann Rev Biochem 80:1055–1087.
8. Zandy A, Lakhani S, Zheng T, Flavell RA, Bassnett S (2005) Role of the executioner caspases during lens development. J Biol Chem 280:30263–30272.
9. Makishima T, Hochman L, Armstrong P, Rosenberger E, Ridley R, Woo M, Perachio A, Wood S (2011) Inner ear dysfunction in caspase-3 deficient mice. BMC Neurosci 12:1–9.
10. Parng C, Seng WL, Semino C, McGrath P (2004) Zebrafish: a preclinical model for drug screening. Assay Drug Dev Technol 1:41–48.
11. Ignatius MS, Langenau DM (2011) Fluorescent imaging of cancer in zebrafish. Methods Cell Biol 105:437–459.
12. Alexander Valencia C, Bailey C, Liu R (2007) Novel zebrafish caspase-3 substrates. Biochem Biophys Res Commun 361:311–316.
13. Lieschke GJ, Currie PD (2007) Animal models of human disease: zebrafish swim into view. Nat Rev Genet 8:353–367.
14. Langenau DM, Ferrando AA, Traver D, Kutok JL, Hezel J-PD, Kanki JP, Zon LI, Look AT, Trede NS (2004) In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. Proc Natl Acad Sci USA 101:7369–7374.
15. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27:157–162.
16. Thornberry NA, Rano TA, Peterson EP, Rasper DM, Timkey T, Garcia-Calvo M, Houtzager VM, Nordstrom PA, Roy S, Vaillancourt JP, 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 272:17907–17911.
17. Stennicke HR, Salvesen GS (1999) Catalytic properties of the caspases. Cell Death Differ 6:1054–1059.
18. Lüthi AU, Martin SJ (2007) The CASBAH: a searchable database of caspase substrates. Cell Death Differ 14:641–650.
19. Mahrus S, Trinidad JC, Barkan DT, Sali A, Burlingame AL, Wells JA (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. Cell 134:866–876.
20. Wejda M, Impens F, Takahashi N, Van Damme P, Gevaert K, Vandenabeele P (2012) Degradomics reveals that cleavage specificity profiles of caspase-2 and effector caspases are alike. J Biol Chem 287:33983–33995.
21. Crawford ED, Seaman JE, Barber AE, David DC, Babbitt PC, Burlingame AL, Wells JA (2012) Conservation of caspase substrates across metazoans suggests hierarchical importance of signaling pathways over specific targets and cleavage site motifs in apoptosis. Cell Death Differ 19:2040–2048.
22. Chowdhury I, Tharakan B, Bhat GK (2008) Caspases—an update. Comp Biochem Physiol B 151:10–27.
23. Eimon PM (2014) Studying apoptosis in the zebrafish. Methods Enzymol 544:395–431.
24. Inohara N, Nuñez G (2000) Genes with homology to mammalian apoptosis regulators identified in zebrafish. Cell Death Differ 7:509–510.
25. Zon LI, Peterson RT (2005) In vivo drug discovery in the zebrafish. Nat Rev Drug Discov 4:35–44.
26. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Al E (2013) The zebrafish reference genome and its relationship to the human genome. Nature 496:498–503.
27. Glasauer SMK, Neuhauss SCF (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics 289:1045–1060.
28. Sakamaki K, Satou Y (2009) Caspases: evolutionary aspects of their functions in vertebrates. J Fish Biol 74:727–753.
29. Iijima N, Yokoyama T (2007) Apoptosis in the medaka embryo in the early developmental stage. Acta Histochem Cytochem 40:1–7.
30. Yabu T, Kishi S, Okazaki T, Yamashita M (2001) Characterization of zebrafish caspase-3 and induction of apoptosis through ceramide generation in fish fathead minnow tailbud cells and zebrafish embryo. Biochem J 360:39–47.
31. Nakajima K, Takahashi A, Yaoita Y (2000) Structure, expression, and function of the Xenopus laevis caspase family. J Biol Chem 275:10484–10491.
32. Van De Craen M, Vandenabeele P, Declercq W, Van Den Brande I, Van Loo G, Molemans F, Schotte P, Van Criekinge W, Beyaert R, Fiers W (1997) Characterization of seven murine caspase family members. FEBS Lett 403:61–69.
33. Dorstyn L, Read SH, Quinn LM, Richardson H, Kumar S (1999) DECAY, a novel Drosophila caspase related to mammalian caspase-3 and caspase-7. J Biol Chem 274:30778–30783.

34. Matthews DJ, Wells JA (1993) Substrate phage: selection of protease substrates by monovalent phage display. Science 260:1113–1117.
35. Lien S, Pastor R, Sutherlin D, Lowman HB (2004) A substrate-phage approach for investigating caspase specificity. Protein J 23:413–425.
36. Bose K, Pop C, Feeney B, Clark AC (2003) An uncleavable procaspase-3 mutant has a lower catalytic efficiency but an active site similar to that of mature caspase-3. Biochemistry 42:12298–12310.
37. Walters J, Schipper JL, Swartz P, Mattos C, Clark AC (2012) Allosteric modulation of caspase-3 through mutagenesis. Biosci Rep 32:401–411.
38. Feeney B, Pop C, Tripathy A, Clark AC (2004) Ionic interactions near loop L4 are important for maintaining the active site environment and the dimer stability of (pro)caspase-3. Biochem J 384:515–525.
39. Vaidya S, Velázquez-Delgado EM, Abbruzzese G, Hardy JA (2011) Substrate-induced conformational changes occur in all cleaved forms of caspase-6. J Mol Biol 406:75–91.
40. Feeney B, Pop C, Swartz P, Mattos C, Clark AC (2006) Role of loop bundle hydrogen bonds in the maturation and activity of (pro)caspase-3. Biochemistry 45:13249–13263.
41. Wang X-J, Cao Q, Liu X, Wang K-T, Mi W, Zhang Y, Li L-F, LeBlanc AC, Su X-D (2010) Crystal structures of human caspase 6 reveal a new mechanism for intramolecular cleavage self-activation. EMBO Rep 11:841–847.
42. Wei Y, Fox T, Chambers SP, Sintchak J, Coll JT, Golec JMC, Swenson L, Wilson KP, Charifson PS (2000) The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. Chem Bio 7:423–432.
43. Thomsen ND, Koerber JT, Wells JA (2013) Structural snapshots reveal distinct mechanisms of procaspase-3 and -7 activation. Proc Natl Acad Sci USA 110:8477–8482.
44. Renatus M, Stennicke HR, Scott FL, Liddington RC, Salvesen GS (2001) Dimer formation drives the activation of the cell death protease caspase 9. Proc Natl Acad Sci USA 98:14250–14255.
45. Hardy JA, Lam J, Nguyen JT, O'Brien T, Wells JA (2004) Discovery of an allosteric site in the caspases. Proc Natl Acad Sci USA 101:12461–12466.
46. Scheer JM, Romanowski MJ, Wells JA (2006) A common allosteric site and mechanism in caspases. Proc Natl Acad Sci USA 103:7595–7600.
47. Vaidya S, Hardy JA (2011) Caspase-6 latent state stability relies on helical propensity. Biochemistry 50: 3282–3287.
48. Cade C, Swartz P, MacKenzie SH, Clark AC (2014) Modifying caspase-3 activity by altering allosteric networks. Biochemistry 53:7582–7595.
49. Fang B, Boross PI, Tozser J, Weber IT (2006) Structural and kinetic analysis of caspase-3 reveals role for S5 binding site in substrate recognition. J Mol Biol 360: 654–666.
50. Velazquez-Delgado EM, Hardy JA (2012) Zinc-mediated allosteric inhibition of caspase-6. J Biol Chem 287: 36000–36011.
51. Agniswamy J, Fang B, Weber I (2007) Plasticity of S2–S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis. febs J 274:4752–4765.
52. MacKenzie SH, Schipper JL, England EJ, Thomas IIIME, Blackburn K, Swartz P, Clark AC (2013) Lengthening the intersubunit linker of procaspase 3 leads to constitutive activation. Biochemistry 52:6219–6321.
53. Scholle MD, Kehoe JW, Kay BK (2005) Efficient construction of a large collection of phage-displayed combinatorial peptide libraries. Comb Chem High Throughput Screen 8:545–551.
54. Walters J, Pop C, Scott FL, Drag M, Swartz P, Mattos C, Salvesen GS, Clark AC (2009) A consitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. Biochem J 424:335–345.
55. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Cryst D60:2126–2132.
56. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
57. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854.
58. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21:1049–1074.
59. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935.
60. Consortium U (2008) The Universal Protein Resource. Nucleid Acid Res 36:D190–D195.