Structural and Functional Analysis of Caspase Active Sites

David Chéreau, Lalitha Kodandapani, Kevin J. Tomaselli, Alfred P. Spada, and Joe C. Wu*

Idun Pharmaceuticals, Inc., 9380 Judicial Drive, San Diego, California 92121

Received September 23, 2002; Revised Manuscript Received February 20, 2003

ABSTRACT: Amino acid sequences of caspases 1, 3, 7, and 8 were aligned with their published three-dimensional (3D) structures. The resultant alignment was used as a template to compare the primary structures of caspases 2, 4−6, and 9−11 to build 3D homology models. The structural models were subsequently refined and validated using structure−activity relationship data obtained from an array of substrate-like inhibitors. All caspases were shown to have identical S1 and catalytic dyad architecture but diverse S2−S4 structures. S2 pockets of these 11 caspases can be briefly categorized into two groups: Csp3, -6, and -7 as one and Csp1, -2, -4, -5, -8, -9, -10, and -11 as the other. S2 pockets of Csp3, -6, and -7 are smaller than those of the other eight caspases, and are limited to binding small P2 residues such as Ala and Val. At the S3 site, the presence of a conserved Arg in all caspases suggests that Glu would be a universally preferred P3 residue. Csp8 and Csp9 have an additional Arg in this pocket that can further enhance the binding of a P3 Glu, whereas Csp2 has a Glu adjacent to the conserved Arg. As such, Csp2 is the only caspase that can accommodate both positively and negatively charged P3. At S4, Csp1, -4, -5, and -11 are closely related with respect to their structures and binder preferences; all have a large hydrophobic pocket and prefer large hydrophobic residues such as Trp. S4 of Csp2, -3, and -7 represents an opposite group with a conformation that is highly specific in binding an Asp. The S4 structures of Csp6, -8, -9, and -10 appear to be hybrids of the two extremes, and have little specificity for any P4. Information revealed from this work provides a guide for designing potent caspase inhibitors with desirable specificity.

Apoptosis is one of the most fundamental biological processes that controls, from fetal development to the homeostasis and immune system, regulation of almost all complex life forms. Although a detailed mechanism of how apoptosis is triggered and carried out in cells has not yet been completely elucidated, genetic and biochemical studies have demonstrated that a proteolytic cascade controlled by the caspase (Csp) family is central to this process (for reviews, see refs *1* and *2*). Caspases are a family of intracellular cysteine proteases that cleave their substrates very specifically at an aspartic acid (P1) residue. To date, 13 mammalian caspases have been described (*1*, *2*). Inspection of their primary sequences indicates that all caspases share an extremely high degree of homology, not only within the caspases from the same species but also across various species. Although all caspases recognize Asp as the P1 cleavage site and utilize the same mechanism for the catalysis, the substrate specificity of each individual caspase is distinct and is determined by the P2−P4 residues of their substrates. Distinction between caspases also extends to the prodomains at N-termini. The prodomains might function as anchors for locating caspases to their respective action sites, whereas the P1−P4 substrate specificity defines their distinct proteolytic roles in the cells that undergo apoptosis. In healthy cells, caspases exist as inactive zymogens (procaspases), but are activated via cleavages at specific aspartic acid residues when cells are triggered to undergo apoptosis. Processing at these aspartic residues results in liberation of the mature large and small subunits from the precursor and ultimate formation of active tetrameric caspases composed of two pairs of large and small subunits. Once sufficient caspase molecules are activated, the caspase cascade is initiated. This subsequently leads to a sequence of cellular morphological changes that are recognized as "apoptosis". As such, caspases present an attractive set of targets for the development of therapeutic compounds for controlling apoptosis.

Many broad-spectrum caspase irreversible inhibitors have demonstrated efficacy *in vivo* in several acute disease models (*3*−*6*). However, specific caspase inhibitors that offer desirable therapeutic benefit for chronic applications remain a challenge, mostly due to a lack of knowledge of which caspase to inhibit in a given cell type, and how to inhibit it selectively. To design a legitimate strategy for the synthesis of caspase inhibitors with desirable selectivity for any given caspase, comprehensive structural information on every caspase, particularly with respect to their substrate binding sites, is essential. Several caspase crystal structures with peptidyl inhibitors bound at substrate binding sites have been made available over the past eight years, including Csp1, Csp3, and Csp7−Csp9 (*7*−*14*). These structures have delineated the location of the substrate pocket and specific intermolecular interactions required for inhibitor binding to specific caspases. However, more than two-thirds of caspases still have no three-dimensional (3D) structural information,

* To whom correspondence should be addressed: Idun Pharmaceuticals, Inc., 9380 Judicial Dr., San Diego, CA 92121. Phone: (858) 646-8101. Fax: (858) 623-2765. E-mail: jwu@idun.com.

10.1021/bi020593l CCC: $25.00
Published on Web 03/21/2003

and it is unlikely that they will appear soon enough to aid drug design. Given the high degree of the primary and secondary structural resemblance among caspases and the guidance from the known 3D structures of the five caspase members, we may construct structural models for the remaining caspases. Nevertheless, computer-simulated models are often speculative because of uncertainties involved in calculating local energy minima and because of inaccurate alignments of primary sequences. Such shortcomings could be minimized if independent methods that allow examination of proposed models can be incorporated. In this report, a panel of substrate-like tetrapeptidyl inhibitors were utilized experimentally to probe the active site structures of eight caspases proposed by computer-aided simulations.

MATERIALS AND METHODS

Recombinant Caspases. Human Csp3, -6, -7, -8, and -10 and murine Csp11 were expressed in *Escherichia coli* BL21-DE3 cells harboring pET-21b plasmids that contained integrated caspase DNA according to published procedures (*15*, *16*). Each of these caspases was purified to near homogeneity using a Ni-chelating column (Amersham Biosciences). Murine Csp1 was expressed using *E. coli* BL21-DE3 pLys S cells harboring the pET3ap30mICEFLAG plasmid as described previously (*17*). Human Csp1, -4, and -5 were purchased from Biovision (Mountain View, CA).

Prodomain-truncated P30 Csp2 and P30 Csp9 were also expressed in BL21-DE3 cells using plasmids containing the prodomain-truncated versions of the human Csp9 DNA (residues 140−416) and human Csp2 DNA (residues 153−435). The truncation sites were based on the protein secondary structure predictions using the program published by Rost and Sander (*18*, *19*). Each of these DNA fragments was inserted into *Nde*I and *Xho*I sites of the pET-21b expression plasmid (Novagen, Madison, WI). In each of these constructs, the C-terminus of the protein is fused to a polyhistidine tag. A culture of BL21-DE3 cells containing the plasmid was grown at 37 °C in induction medium [20 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 0.5 M M9 salts, 0.4% glucose, 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, and 0.1 mg/mL ampicillin (pH 7.4)]. When the culture density reached an A_{600} of 1.0, isopropyl β-D-thiogalactopyranoside (IPTG, Sigma, St. Louis, MO) was added to a concentration of 1.0 mM and the culture was shaken at 25 °C for 4 h. The induced cells were harvested by centrifugation and resuspended in cold binding buffer [20 mM Tris-HCl, 5 mM imidazole, and 500 mM NaCl (pH 8.0)] containing 0.1 mg/mL lysozyme and 0.1% Triton X-100. Cells were frozen on dry ice and stored at −80 °C. After being thawing, the cells were sonicated to reduce sample viscosity and centrifuged at 30000*g* for 40 min at 4 °C. The lysate was filtered through a 0.8 μm syringe filter, and caspases were purified using a 10 mL Ni^{2+}-charged HiTrap chelating column.

Caspase Inhibitors and Substrates. Fluorogenic caspase substrates tetrapeptidyl aminomethylcoumarin were purchased from Anaspec, Inc. Peptidic aldehydes Ac-DEVDald, Ac-ZEKDald, Ac-ZAVDald, Ac-ZKVDald, and Ac-ZEVDald were prepared following a general procedure reported by Graybill et al. (*20*). Ac-YVADald, Ac-YVKDald, Ac-LEHDald, and Ac-WEHDald were from Peptides International, Inc. (Louisville, KY). Boc-AEVDald was purchased from Bachem, Inc. (Torrance, CA), and Ac-LEVDald was from Biosource International, Inc. (Camarillo, CA).

Computational Modeling. The GenBank database accession numbers of caspases used in this work are P29466 for *Homo sapiens* Csp1, P42575 for Csp2, P42574 for Csp3, P49662 for Csp4, P51878 for Csp5, P55212 for Csp6, P55210 for Csp7, Q14790 for Csp8, P55211 for Csp9, BAB32553 for Csp10A, P29452 for *Mus musculus* Csp1, and P70343 for *M. musculus* Csp11. Caspase-1, -3, -7, and -8 amino acid sequences were aligned manually on the basis of their 3D crystal structures. This structure-based sequence alignment was then used as a template for the subsequent matching up of all other caspases. After several refinements that emphasized the conserved secondary structures and at the active site conserved regions, the obtained alignment was shown as in Figure 2. The refined alignment was then uploaded to run the MOE Homology Model program (Chemical Computing Group, Inc.). Given the assigned alignment of amino acids and coordinates from the crystal structures of Csp1, -3, -7, and -8 with inhibitors designated to the active site, the simulation program was instructed to give the initial solutions in which WEHDald in Csp4, -5, and -11, DEVDald in Csp2 and -6, and IETDald in Csp9 and -10 were assigned to the active sites. These are the ligands cocrystallized in the four reported crystal structures (*7*−*13*) shown in Figure 1. After a preliminary model was obtained for a given caspase, the original ligand was replaced with its optimal inhibitor shown in Figure 3 according to the literature or our own data. Using the Engh−Huber algorithm, the models were iteratively energy minimized and further refined to agree with our inhibitor data for the final models. The structure obtained with this procedure represents one of the possible active site conformational states of the corresponding caspase with a potent ligand bound.

Caspase Assay and Inhibition. Caspase assays were performed in a 150 μL buffer solution containing 25 mM Hepes, 1 mM EDTA, 0.1% CHAPS, 100 g/L sucrose, and 8 mM DTT at pH 6.0 for mouse Csp11, at pH 6.5 for Csp2 and -9, and at pH 7.5 for Csp1, -3, -6, -7, -8, and -10. Csp4 and Csp5 were assayed in PBS buffer [10 mM Na_2HPO_4, 1 mM KH_2PO_4, 137 mM NaCl, and 2.7 mM KCl (pH 7.4)]. Assays were run with 3 nM hCsp1, 4 nM Csp2, 0.15 nM Csp3, 5 nM Csp4, 5 nM Csp5, 2.5 nM Csp6, 2.5 nM Csp7, 2.5 nM Csp8, 1 nM Csp9, 5 nM Csp10, 8 nM Csp11, or 0.25 nM mCsp1. Caspase concentrations were based on the active site titration. Tetrapeptide aminomethylcoumarin (AMC) was used as the substrate: 10, 50, 50, and 50 μM WEHDamc in mCsp1, hCsp1, Csp4, and Csp5 assays, respectively; 10 μM VDVADamc in the Csp2 assay; 10 μM DEVDamc in Csp3 and Csp7 assays; 10 μM IETDamc in Csp6, -8, and -10 assays; 50 μM IETDamc in the Csp11 assay; and 50 μM LEHDamc in the Csp9 assay. Under these conditions, 0.2−0.5 nmol of substrates was hydrolyzed per hour at 30 °C as followed by the increase in fluorescence (excitation at 360 nm and emission at 460 nm) resulting from formation of the product AMC. All assays were run in duplicate Cytoplate wells using the CytoFluor plate reader (PerSeptive Biosystems). For inhibition studies, 50 μL of inhibitors was mixed with 50 μL of substrate solutions prior to the addition of 50 μL of caspases. Inhibitor IC_{50} values

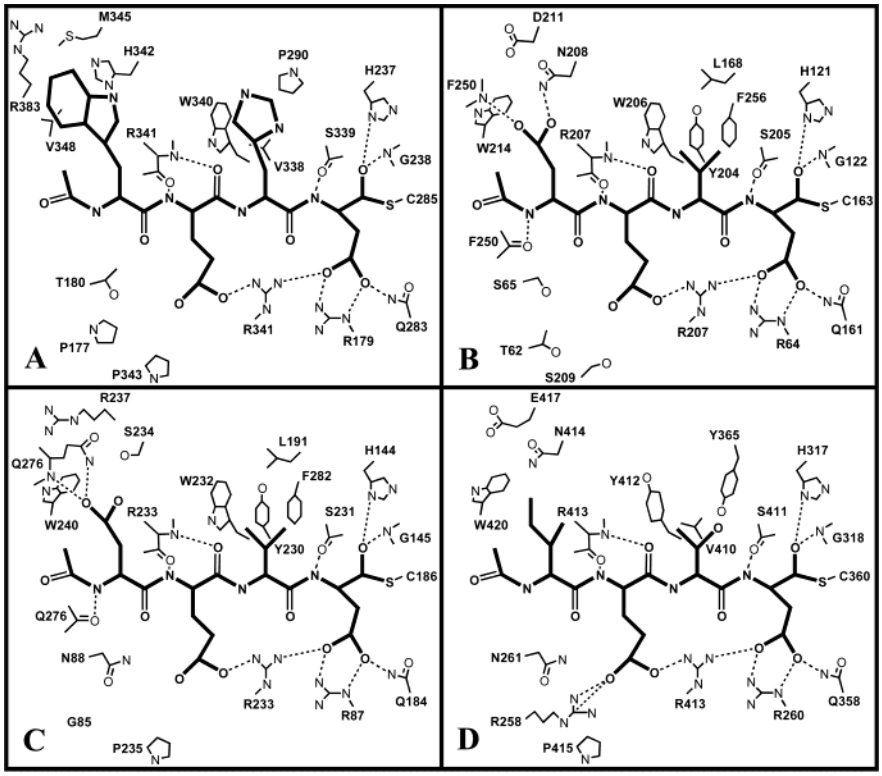

FIGURE 1: Schematic representation of Csp1 (A), Csp3 (B), Csp7 (C), and Csp8 (D) active site structures. Structures are shown with the inhibitors Ac-WEHD, Ac-DEVD, Ac-DEVD, and Ac-IETD covalently bound to the respective caspase. These projections are derived from the published crystallographic coordinates (*8*, *9*, *12*, *13*). The PDB entry codes for Csp1, -3, -7, and -8 are 1IBC, 1PAU, 1F1J, and 1QTN, respectively.

were calculated using the equation

$$\Delta F/\Delta t = (\Delta F/\Delta t)_0/(1 + [\mathrm{I}]/\mathrm{IC}_{50})$$

based on the inhibition data obtained throughout the inhibitor concentration range from 0.3 nM to 300 μM, where $\Delta F/\Delta t$ is the observed averaged rate of fluorescence change at inhibitor concentration [I] and $(\Delta F/\Delta t)_0$ is the rate of fluorescence change for the uninhibited enzyme. K_i values were calculated according to the equation

$$K_i = \mathrm{IC}_{50}/(1 + [\mathrm{S}]/K_M)$$

where K_M values were determined by measuring caspase activity throughout the substrate range of 1−1000 μM, and subsequently calculated using the equation

$$\Delta F/\Delta t = (\Delta F/\Delta t)_{max}/(1 + K_M/[\mathrm{S}])$$

where [S] denotes the substrate concentration and $(\Delta F/\Delta t)_{max}$ is the maximum catalysis rate when [S] approaches infinity.

RESULTS

Csp1, -3, -7, and -8 Three-Dimensional Structure Overlay and Function-Based Sequence Alignment. In an attempt to construct a template for aligning all caspase primary structures in a manner related to their substrate recognition, we have analyzed the reported coordinates of the four solved caspase structures, i.e., Csp1, -3, -7, and -8 (*7*−*13*), and aligned their primary sequences correlating to their active site 3D structures. Figure 1 illustrates the key residues in the active sites of these four caspases extracted from the reported coordinates (*8*, *9*, *12*, *13*), and Figure 2A shows the primary sequence alignment derived from the 3D overlay. Notably, the S1 pockets of all four caspases are almost identical, consistent with their common specificity of Asp at P1. In each of four caspases, there are two Arg residues and a Gln, e.g., Arg179, Arg341, and Gln283 in Csp1 (Figure 1A), in this pocket that interact with the carboxyl side chain of the P1 Asp via ionic interaction and hydrogen bonding, respectively. A Ser, e.g., Ser339 of Csp1, that forms a hydrogen bond between its backbone carbonyl oxygen and the P1 backbone nitrogen is also conserved among all four caspases. The essential groups involved in the catalysis are identical among all four caspases as well. A His, e.g., His237 of Csp1, and a Gly backbone nitrogen, e.g., Gly238 of Csp1, anchor the carbonyl group of the P1 Asp at an appropriate position that allows the active site Cys thiol to form a tetrahedral configuration with the carbonyl center of the bound substrate, a key intermediate during catalysis.

In contrast to the highly conserved S1 structure, S2−S4 pockets are considerably different among the four caspases. The S2 pockets of Csp3 and Csp7 resemble each other to a high degree, but are only slightly homologous with that of Csp1 and Csp8. The S2 pockets of Csp3 and Csp7 consist of three key hydrophobic side chains, i.e., Tyr204, Trp206, and Phe256 of Csp3 and Tyr230, Trp232, and Phe282 of Csp7 (Figure 1B,C), and only allow relatively small room for small aliphatic P2 side chains such as Ala, Val, or Thr. In Csp1 and Csp8, instead of a Tyr as in Csp3 at position 204 and in Csp7 at position 230, the same position is occupied with a smaller residue, Val (Val338 in Csp1 and Val410 in Csp8). This substitution along with other minor changes in the region (Figure 1) results in a larger S2 pocket

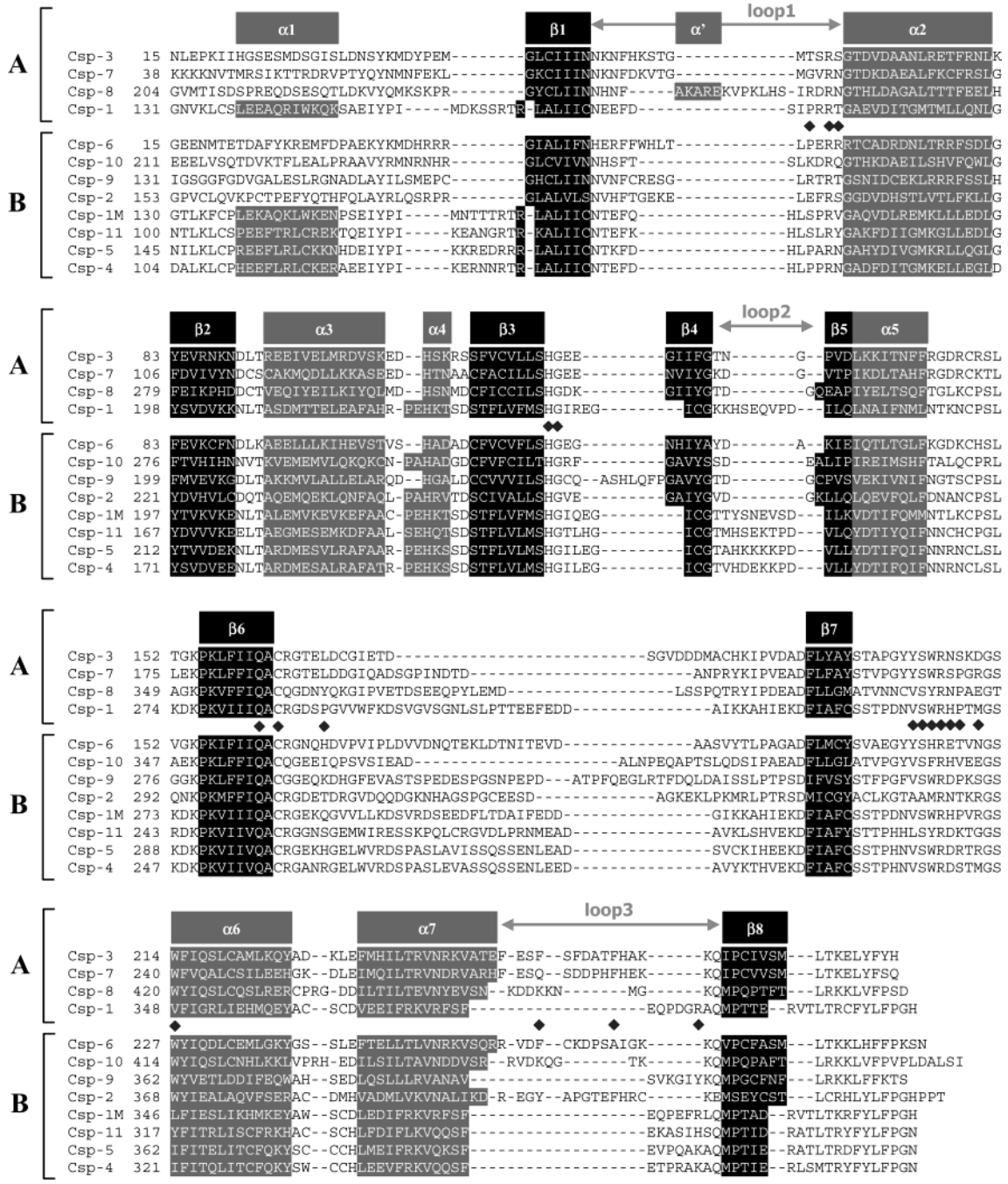

FIGURE 2: SAR-related alignment of caspase primary sequences. (A) Primary sequence alignment of Csp1, -3, -7, and -8 based on superimposed 3D structures. Superimposed residues are vertically aligned. Residues that do not have spatial counterparts in the other caspases are aligned with dashed lines. Amino acids with side chains in the active sites are indicated (black diamonds). Amino acids that give α- and *β*-secondary structures are shaded as indicated. (B) Primary sequence alignment of the remaining caspases that are without crystal structures. The alignment was carried out using the A panels as the template as described in the text. Csp9 crystal structural data were unavailable at the time this work was conducted.

of Csp1 and Csp8, which could accommodate larger P2 side chains.

The S3 pockets of the four caspases appear to be highly homologous. The key residue is a conserved Arg, i.e., Arg341 in Csp1, Arg207 in Csp3, Arg233 in Csp7, and Arg413 in Csp8. This conserved Arg provides an ionic interaction with the bound P3 Glu. It is also the one that forms a backbone−backbone interaction with the bound P3 residue. Such an arrangement allows a favorable binding of Glu at P3 for all four caspases. Notably, there is a second Arg (Arg258) in the Csp8 S3 pocket that provides an additional ionic interaction with the P3 Glu. The same spatial positions in Csp1, Csp3, and Csp7 are occupied by Pro177, Thr62, and Gly85, respectively. The additional Arg in the Csp8 S3 pocket has been predicted to contribute to a tighter binding of a Glu at P3 for Csp8 than for Csp1, -3, and -7 (*11*).

Significant structural divergence at S4 pockets among the four caspases has been observed. The Csp1 S4 pocket is considerably larger and more hydrophobic than those of the other three caspases. A Trp at P4 was shown to fit in the pocket in Csp1 quite favorably, not only in its size but also having multiple interactions with residues in the S4 pocket, including Arg383, His342, Trp340, Met345, and Val348. Although the Csp1 S4 pocket is best filled with a large

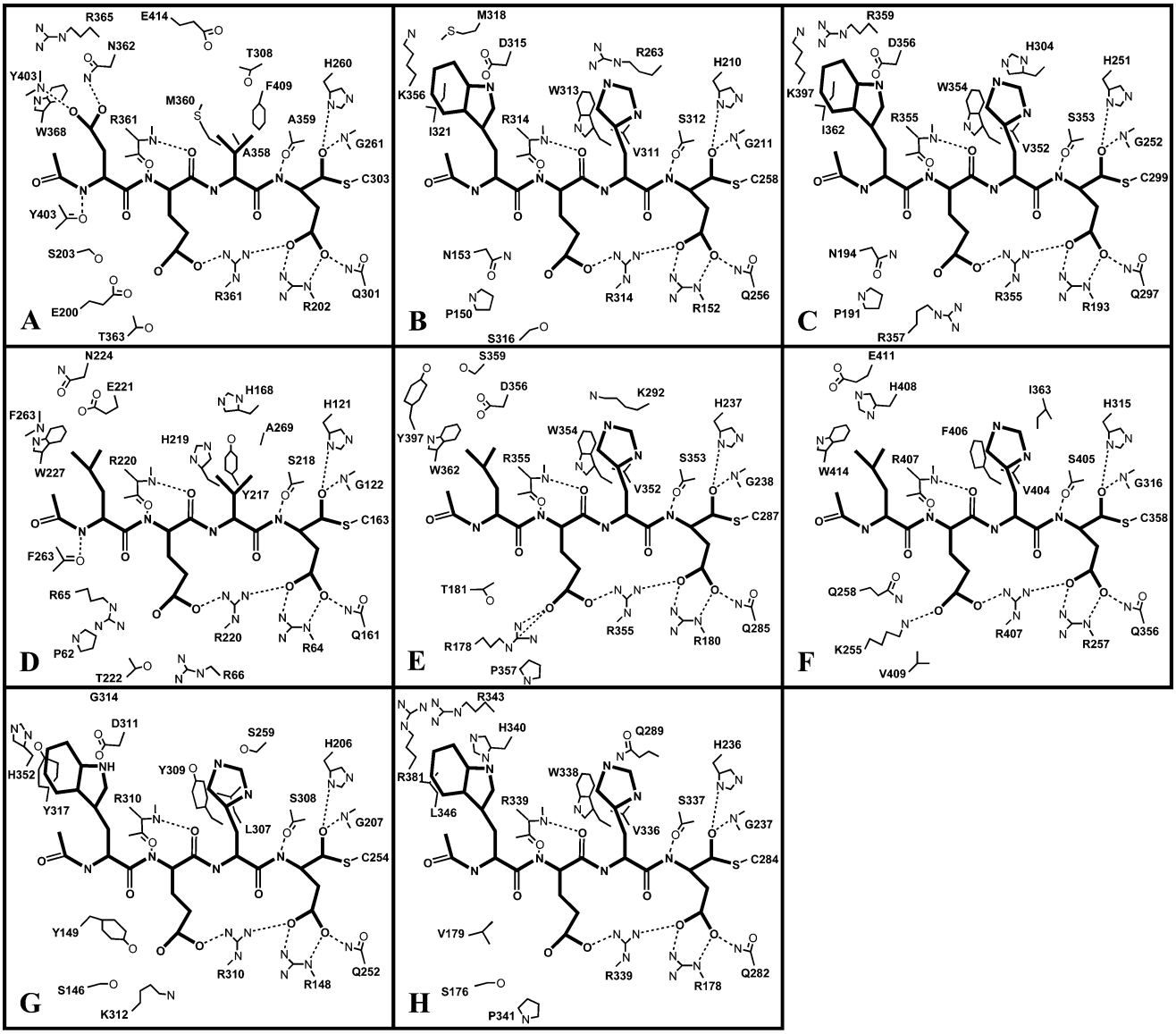

FIGURE 3: Schematic presentation of modeled human Csp2 (A), Csp4 (B), Csp5 (C), Csp6 (D), Csp9 (E), and Csp10 (F) and mouse Csp11 (G) and Csp1 (H) active sites with Ac-DEVD (human Csp2), Ac-WEHD (human Csp4 and -5 and mouse Csp11 and -1), Ac-LEVD (human Csp6), and Ac-LEHD (human Csp9 and -10) covalently bound to the active site Cys. 3D structural models were obtained on the basis of the alignment in Figure 2B and simulated with the MOE homology program.

hydrophobic P4, it is possible to insert a charged residue such as Asp, due to the presence of Arg383 (*13*); however, the affinity would be significantly reduced. The S4 pockets of Csp3 and Csp7 are significantly smaller and more hydrophilic than that of Csp1, whereas the Csp8 S4 pocket appears to be a hybrid of the two extremes; it may accommodate both hydrophilic and hydrophobic ligands with less affinity in either case as in the other three caspases.

Similar structural comparisons have also been extended outside the substrate binding areas. A high degree of conservation in α- and *β*-secondary structures, but not in the loop structures, was observed. Using active site amino acids that are at the same space position in their 3D structure as the anchors, and fine-tuning with conserved secondary structures, we obtained the point-to-point alignment of Csp1, -3, -7, and -8 primary sequences as illustrated in Figure 2A. As such, this provides a function-relevant primary sequence alignment.

Using the alignment of these four caspases as a template, we have extended the comparison to the primary structures of the other caspases that do not have crystallographic data. Figure 2B illustrates such an alignment excluding the prodomain regions. Not surprisingly, a high degree of homology was observed; most of the key residues used in constructing the template are also identified in the rest of caspases, and almost all of the secondary structures are conserved.

Structural Modeling of Csp2, -4, -5, -6, -9, -10, and -11 Active Sites. With the high degree of homology in the function-based sequence alignment shown in Figure 2B, and with the known 3D structures of four caspase members, one can build a 3D active site structural model with reasonable authenticity for a given caspase. This can be started by constraining the key conserved residues, including the ones at the S1 pocket and at the catalytic motifs, where their counterparts in the four known crystal structures are. This can then be followed by adding the remaining residues that would provide essential interactions with substrate guided by the four known crystal structures. As such, a preliminary framework of the substrate binding site is created. The gaps

are filled in by using information provided from the primary sequence alignment shown in Figure 2B. After several minimizations and refinements with the MOE homology program, simulated active site structures of Csp2, -4, -5, -6, -9, -10, and -11 were constructed. The local two-dimensional projections are shown in Figure 3. S1 pockets and the catalytic Cys-His dyad of all caspases appear to be indistinguishable as expected. At the S2 pockets, the participating amino acid residues are quite diverse. However, residue 338 in Csp1 and its equivalents in other caspases seem to be restricted to either Val or Tyr with the exception of Csp2 and Csp11, i.e., Val338 (Csp1), Ala358 (Csp2), Val311 (Csp4), Val352 (Csp5), Val410 (Csp8), Val352 (Csp9), Val404 (Csp10), Tyr204 (Csp3), Tyr217 (Csp6), Tyr230 (Csp7), and Leu307 (Csp11). Since the amino acid at this Csp1 338 position is important in determining the size of an appropriate P2 residue as suggested by the crystal structures of Csp1, -3, -7, and -8, S2 pockets of Csp6, like those of Csp3 and -7, might also be restricted to only small β-branched P2 amino acids such as Val or Thr. On the other hand, Csp2, -4, -5, -9, -10, and -11, like Csp1 and -8, might be more tolerant of larger residues at P2.

Although Csp1, -2, -4, -5, -8, -9, and -10 all contain a Val or an Ala (Csp2) at the position equivalent to Csp1 Val338, and allow bigger S2 pockets of these caspases compared to those of Csp3, -6, and -7, they are individually distinct because of the other amino acids in the pockets. Among these seven caspases, Csp2 has the largest S2 with a negatively charged Glu414 in this pocket that is not seen in the other six caspases where the same position is exclusively a Gln with no other negatively charged residue residing observed near this position. Collectively, these would suggest that Csp2 is able to accommodate a large positively charged P2 that is not expected for the other six caspases. This prediction has been examined, and the results are shown in the next section.

In our reconstituted models, the S3 pockets of Csp2, -4, -5, -6, -9, -10, and -11 are all solvent-exposed and each has a Glu-friendly groove in the pocket as in Csp1, -3, -7, and -8 shown by the crystallographic data. As demonstrated in Figures 1 and 2, there is an Arg that is conserved in Csp1, -3, -7, and -8, e.g., Csp1 Arg341. This Arg interacts with both P1 Asp and P3 Glu of the binder. The same conserved Arg is also present in the remaining seven caspases we modeled. This suggests that Glu could be a universally preferred residue at P3 for these seven caspases as for the other four. Besides this conserved Arg, there are additional positively charged residues in the S3 pocket of Csp5, Csp6, and Csp8−Csp11, i.e., Csp5 Arg357, Csp6 Arg65 and Arg66, Csp8 Arg258, Csp9 Arg178, Csp10 Lys255, and Csp11 Lys312. Among these, Csp9 Arg178 is at the same position as Csp8 Arg258 and, like it, may be able to interact with the carboxyl side chain of a bound P3 Glu. However, in the cases of Csp5, Csp6, Csp10, and Csp11, it is not clear whether and how the additional Arg or Lys would interact with P3 Glu. Data from inhibitor binding experiments would inform us.

Unlike other caspases that have one or multiple positively charged Arg or Lys residues at their S3 pocket, Csp2 is the only human caspase that has a negatively charged residue (Glu200) at the location corresponding to Csp1 Pro177. Given that Csp2 also has the conserved Arg361 to interact with the P3 side chain of a bound ligand, the presence of Glu200 might not be sufficient to completely repel the binding of a P3 Glu, but would be sufficient to reduce the binding affinity significantly. On the other hand, unlike most caspases that cannot accommodate binders with positively charged residues such as Lys at the P3 pocket, Csp2 is likely to bind them reasonably well. In summary, all caspases seem to prefer a Glu for binding to their S3 sites though the degree may vary significantly. Csp2 could be the only caspase that is capable of binding a positively charged P3. These predictions can be examined by inhibitor binding.

While the crystallographic analyses show diverse structures of S4 among Csp1, -3, -7, and -8, our reconstituted models of the remaining caspases suggest that their S4 pockets resemble one or the other of the four known S4 structures relatively well. The S4 pockets of Csp4, -5, and -11 contain hydrophobic side chains and closely resemble that of Csp1, though there are differences that allow distinguishing one from the others. His342 in Csp1 is substituted with an Asp in Csp4, -5, and -11, i.e., Asp315 (Csp4), Asp356 (Csp5), and Asp311 (Csp11), whereas Val348 in Csp1 is substituted with an Ile or Tyr in the latter three enzymes, i.e., Ile321 (Csp4), Ile362 (Csp5), and Tyr317 (Csp11) (Figure 2B). These differences would result in an S4 pocket for Csp4, -5, and -11 slightly smaller and less hydrophobic than that of Csp1. A P4 Trp that fits very well to Csp1 S4 might not be as optimal for Csp4, -5, and -11 S4. The same Csp1 Val348 equivalent position in the remaining caspases is strictly a Trp, i.e., Csp2 (Trp368), Csp3 (Trp214), Csp6 (Trp227), Csp7 (Trp240), Csp8 (Trp420), Csp9 (Trp362), and Csp10 (Trp414). With a Trp at this position, these caspases would not be able to accommodate another sizable side chain from the P4 pocket of a binder such as Trp.

As observed from crystal structures of Csp3 and Csp7, an Asp fits very well to their S4. The binding is mainly attributed to the interaction of the Asp carboxyl side chain with Csp3 Asn208 and Phe250, and with Csp7 Gln276. As demonstrated in Figure 3A, the Csp2 S4 structure based on the model appears to be quite similar to that of Csp3 and Csp7; all essential groups for interacting with a P4 Asp are present. As such, Csp2 may be able to bind an Asp at P4 as favorably as Csp3 or -7.

Csp6 has a loop 3 structure similar to that of Csp3 and -7, and might be able to interact with the carboxylic group of a P4 Asp via the Phe263 backbone nitrogen at the base of the loop. However, the critical Asn for interacting with the P4 Asp as in Csp3 (Asn208) is missing in Csp6, and is replaced with a Glu. This would make Asp a less favorable P4 residue for Csp6 than for Csp3. Similarly, this Asn is also missing in Csp9 and -10. As such, these caspases might not bind ligands with a P4 Asp as tightly as Csp2, -3, and -7. In addition, Csp6 and Csp9 have a negatively charged residue (Glu221 and Asp356, respectively) at the position corresponding to Csp1 His342 in their S4 pockets. This could further make a P4 Asp particularly unfavorable for these two caspases. Csp4, -5, and -11 also have a negatively charged residue at the same location, and hence are not expected to bind a P4 Asp in a favorable manner.

As shown in Figure 1D, the Csp8 S4 pocket is comprised of both hydrophilic and hydrophobic residues, and therefore may be able to accommodate a variety of residues at P4. Csp10 S4 has structure similar to that of Csp8 S4 except

Table 1: K_i Values and K_i Ratios of Various Substrate-like Inhibitors against Caspases

	K_i											
inhibitor	Csp1	Csp2	Csp3	Csp4	Csp5	Csp6	Csp7	Csp8	Csp9	Csp10	mCsp11	mCsp1
AEVDald	54.0	160000	42.0	1280	480	48.7	814	8.4	1.6	762	615	51.3
DEVDald	79.7	884	2.3	1402	602	41.6	37.5	6.8	1.7	59.5	337	82.4
LEHDald	41.1	2580	229	343	37.4	115	15600	1.8	1.2	14.9	795	10.1
LEVDald	41.1	2850	30.3	328	38.8	28.8	1460	2.4	1.1	25.1	450	9.2
WEHDald	2.0	62700	1050	83.8	11.2	1370	45700	16.2	17.4	492	488	1.3
YVADald	2.3	130000	10600	1020	102	230000	120000	293	28.6	697	885	1.5
YVKDald	20.6	23100	70500	15400	796	530000	600000	2090	1300	15300	22600	7.0
ZAVDald	272	530000	1900	40200	3120	39900	11200	2490	901	45500	3820	355
ZEKDald	118	85.8	42700	86100	1850	2080	101	12200	62.1	3620	135	146
ZEVDald	7.2	13100	23.5	503	72.8	204	91.4	5.2	1.8	279	54.8	3.8
ZKVDald	1350	726	86600	160000	8830	16700	1240	61800	2450	630000	67500	1610

	K_i ratio											
	Csp1	Csp2	Csp3	Csp4	Csp5	Csp6	Csp7	Csp8	Csp9	Csp10	mCsp11	mCsp1
LEHD/LEVD	1.0	0.9	7.6	1.0	1.0	4.0	10.7	0.8	1.1	0.6	1.8	1.1
YVAD/YVKD	0.11	5.6	0.15	0.07	0.13	0.43	0.20	0.14	0.02	0.05	0.04	0.21
ZEVD/ZEKD	0.06	152	0.001	0.006	0.04	0.10	0.91	0.001	0.03	0.08	0.41	0.03
ZAVD/ZEVD	38	40	81	80	43	195	123	478	501	163	70	93
ZKVD/ZEVD	189	0.06	3680	318	121	82	13.6	11900	1360	2260	1230	424
LEHD/WEHD	20.6	0.04	0.22	4.1	3.3	0.08	0.34	0.11	0.07	0.03	1.6	7.8
LEVD/DEVD	0.52	3.2	13.2	0.23	0.06	0.69	39.0	0.35	0.65	0.42	1.3	0.11
AEVD/DEVD	0.68	181	18.4	0.91	0.80	1.2	21.7	1.2	0.94	12.8	1.8	0.62

that Asn414 in Csp8 is His408 in Csp10. These two caspases should share similar P4 preferences.

In summary, Csp4, -5, and -11 are closely related to Csp1 with respect to their S4 structures. All of their S4 pockets are relatively sizable and hydrophobic. Csp2, -3, and -7 belong to a different category where a hydrophobic P4 is not preferred and an Asp is expected to be most favorable. Csp6, -8, -9, and -10 are between the two extremes, and are expected to accommodate both hydrophobic and hydrophilic P4 residues with less significant binding affinity in either case compared to the other two groups of S4.

Active Site Structures and Inhibitor Binding. Given the appropriate alignment of sequences, the four known crystal structures of caspases provide a highly instructive guide with which we can construct homology models for the remaining caspases. However, such models are not highly precise; too many variables remain even after several refinements with MOE simulations. Approaches to verifying the proposed structures are required. With this objective, peptidic inhibitors were designed for testing the simulated structures, and the results are summarized in Table 1.

As described in the previous section, our simulated models predict that S2 pockets of Csp1, -2, -4, -5, -8, -9, -10, and -11 tend to be more flexible than those of Csp3, -6, and -7 in accommodating P2 of different sizes. LEHDald and LEVDald were chosen for demonstrating this difference. As shown in Table 1B, $K_{i-LEHDald}/K_{i-LEVDald}$ ratios for Csp1, -2, -4, -5, -8, -9, -10, and -11 are close to unity, whereas the same K_i ratios for Csp3, -6, and -7 are significantly higher, consistent with what the models suggest. Another distinct feature among S2 pockets of various caspases shown in Figures 2 and 3 is the presence of E414 at a distant corner in the Csp2 S2 pocket, while such a negatively charged residue is absent in all other caspases. This would imply that a residue with a long, flexible, and positively charged side chain at P2 could be particularly favored by Csp2. This prediction can be tested by comparing the binding affinities of YVADald, YVKDald, zEVDald, and zEKDald bound to Csp2 and to other caspases. As demonstrated in Table 1, all caspases but Csp2 favor Val or Ala over Lys at P2 as indicated by the $K_{i-YVADald}/K_{i-YVKDald}$ and $K_{i-zEVDald}/K_{i-zEKDald}$ ratios. The ratio of $K_{i-YVADald}$ to $K_{i-YVKDald}$ for Csp2 is 5.6, and ranges from 0.02 to 0.43 for the remaining caspases. More dramatically, the $K_{i-zEVDald}/K_{i-zEKDald}$ ratio for Csp2 is 152, whereas it ranges from 0.001 to 0.91 for all other caspases. Accordingly, one could expect that an inhibitor with a Lys at the P2 position along with appropriate P3 and P4 residues may give a specificity index of 40−38000 for Csp2 over the remaining caspases. It should be noted that when two inhibitors are analyzed for their binding to a caspase, the effect of their difference in solvation energy should not be overlooked. However, as the difference in solvation energy remains relatively constant for the same pair of inhibitors regardless the caspase against which they are tested, its effect on the binding is at least partially offset when the K_i ratio of the same pair of inhibitors is compared from one caspase to another.

Of all proposed S3 pockets of the various caspases shown in Figure 3, each predicts a favorable binding of Glu at P3 because of a conserved Arg that bridges P1 Asp and P3 Glu, e.g., Csp1 Arg341. In Csp8 and Csp9, there is an additional Arg (Csp8 Arg258 and Csp9 Arg178) present in this pocket that is expected to offer additional binding energy when bound to a ligand with a P3 Glu. Comparison of the binding of zAVDald and zEVDald to all caspases is shown in Table 1. It is evident that a Glu at P3 is greatly favored by every caspase as the ratios of $K_{i-zAVDald}$ to $K_{i-zEVDald}$ are all at least 38. Among all the caspases, the ratios for Csp8 and Csp9 are significantly higher than for others, i.e., 478 and 501, respectively, consistent with the expected extra binding energy from the additional Arg of Csp8 and Csp9. Csp5, -6, -10, and -11 also have an additional Arg or Lys in the neighborhood. Solely on the basis of the modeling, it is difficult to determine whether and how these additional positively charged residues would interact with the P3 Glu. However, $K_{i-zAVDald}/K_{i-zEVDald}$ ratios suggest that Csp6 Arg65

or Arg66, and Csp10 Lys255, might help in binding a P3 Glu, though less significantly than the additional Arg in Csp8 and Csp9. On the other hand, Csp5 Arg357 and Csp11 Lys312 seem to make little contribution as their $K_{i-zAVDald}/K_{i-zEVDald}$ ratios are comparable to that for caspases without an additional Arg or Lys in the area.

Another important feature exhibited by the models is the presence of a negatively charged Glu200 that coexists with the conserved Arg in the Csp2 S3 pocket but not in the other caspases. Although this Glu200 might not be sufficient to completely repel the binding of a ligand with Glu at its P3 because of the nearby Arg361, one may expect that the contribution of this P3 Glu in binding to Csp2 would be significantly smaller than that for binding to other caspases. The $K_{i-zAVDald}/K_{i-zEVDald}$ ratio for Csp2 is among the lowest in Table 1. In addition, when zEVDald P3 Glu is replaced with an oppositely charged Lys, the potency of the inhibitor is diminished dramatically in all cases except Csp2, where the binding affinity was actually improved ~17-fold. The ratio of $K_{i-zKVDald}$ to $K_{i-zEVDald}$ for Csp2 is 0.06, whereas it ranges from 13.6 to 11 900 for the remaining caspases.

There are several predictions one could make on the basis of the simulated 3D structures of S4 pockets for caspases. (1) Although the S4 pockets of Csp4, -5, and -11 resemble the Csp1 S4 pocket quite closely, the replacements of Csp1 Val348 with an Ile in Csp4 (Ile321) and Csp5 (Ile362) and a Tyr in Csp11 (Tyr317) may inhibit their ability to accommodate a Trp at P4 compared to Csp1. (2) Since the same Csp1 Val348 equivalent in Csp2, -3, -6, -7, -8, -9, and -10 is strictly a Trp, the latter caspases could be incapable of binding ligands with another sizable Trp at P4. (3) Like Csp3 and Csp7, Csp2 would greatly prefer a charged residue such as Asp at P4. Such a preference for Asp should not be evident for the remaining caspases.

K_i ratios for LEHDald to WEHDald were taken to examine predictions 1 and 2. This ratio is 20.6 for Csp1 (Table 1) and is between 1.6 and 4.1 for Csp4, Csp5, and Csp11, but it ranges from 0.03 to 0.34 for the remaining caspases. Also demonstrated in the same table is the fact that Csp2 is the only caspase besides Csp3 and Csp7 that prefers Asp over Leu or Ala at P4, though to a lesser extent than Csp3 and Csp7. The ratio of $K_{i-LEVDald}$ to $K_{i-DEVDald}$ is less than 1 for most caspases, but is 13.2, 39, and 3.2 for Csp3, Csp7, and Csp2, respectively. A similar trend was also observed in comparing the K_i values of AEVDald and DEVDald.

Caspase Active Site Divergence across Species. On the basis of the same alignment strategy and modeling procedure, one may compare the active site structure of a given caspase from different species. A simulated 3D structural model for mouse Csp1 was constructed and compared with that of human Csp1 (Figures 3 and 4A). In general, the active site structures of human and mouse Csp1 are highly conserved, though some differences in their primary sequences are observed. For instance, Pro290 (S2), Asp288 (S2), Thr180 (S3), Pro177 (S3), Val348 (S4), and Met345 (S4) of human Csp1 are replaced with Gln, Glu, Val, Ser, Leu, and Arg, respectively, in the mouse enzyme. Judging from our protein models, these substitutions do not result in major structural changes and should not give significant specificity variance. Various K_i ratios listed in Table 1 that consistently differ from 2- to 3-fold between the two caspases support this conclusion.

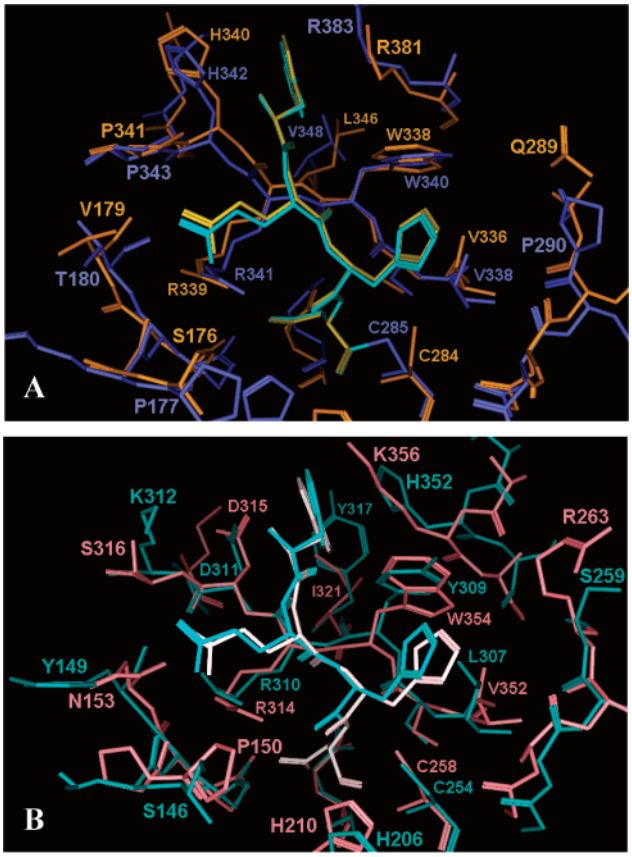

FIGURE 4: Superposition of the active site regions of human and mouse caspases. (A) Human (blue) and mouse (yellow) Csp1. (B) Human Csp4 (pink) and mouse Csp11 (blue). All four structures are with WEHD covalently bound to the active site Cys. The human Csp1 structure was constructed using published coordinates (1IBC), whereas human Csp4 and mouse Csp1 and Csp11 are homology models built as described in the text.

We have also compared the active site structural models of a mouse-specific caspase, Csp11, and its suggested human counterpart, Csp4. The overall active site architectures of these two caspases from different species are not significantly different except at S2 (Figure 4B). There is an Arg (Arg263) located at a distant corner of the Csp4 S2 pocket, whereas the equivalent position in Csp11 is Ser259. With this difference, a positively charged P2 residue such as Lys is expected to be unfavorable for Csp4, but make no impact on Csp11. Shown in Table 1, Csp4 and Csp11 share almost identical binding specificity throughout the four S subsites, except that Csp4 prefers zEVDald 167-fold over zEKDald, in contrast to Csp11 that shows no significant preference between the two inhibitors (~2-fold difference in K_i).

DISCUSSION

With recent improvements, homology modeling is being frequently utilized in contemporary structural biology. However, the accuracy and usefulness of such models depend not only on the degree of homology between the two proteins but also on how the two primary structures are aligned. A structure of Csp9 was proposed on the basis of a homology model (*21*). However, since the alignment of the primary sequence used to build the model was slightly off, the resultant 3D structure disagreed significantly with the subsequently published crystallographic result (*14*). Alternatively, we have aligned caspase primary sequences for

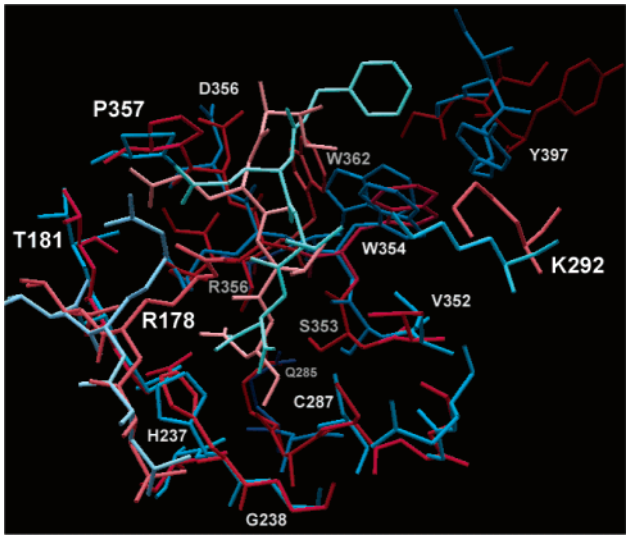

FIGURE 5: Comparison of modeled and crystallographic Csp9 active site structures. The active site of modeled Csp9 (red) complexed with IETD (pink) is superimposed with that obtained from a crystallographic study reported by Renatus et al. (*14*; PDB entry 1JXQ) presented in blue. ZEVD cocrystallized in the structure is shown in light blue.

model building on the basis of the key functional groups in the known 3D structures, and supplemented with the information from conserved secondary structures among caspases. As there are typically multiple possible low-energy conformers for a given protein, an independent method that allows verification and fine-tuning of the preliminary structural models needs to be in place. For this, various tetrapeptide binders were designed to validate the modeling. This approach appears to be useful as our Csp9 homology structure was constructed independent of crystallographic data, and was found to agree well at the active site with the crystal coordinates that later became available to us. Comparison of the two active site structures is shown in Figure 5. Note that all key side chains in the S1−S3 sites in both structures superimpose well. However, orientations of Trp362 and Tyr397 in the S4 pocket are different in the model and in the crystal structure determined at 2.8−3.5 Å resolution. The indole side chain of Trp362 in the model tilts about 30° from that in the crystal structure. This might be due to different ligands bound at the active site of the model and of the crystal, i.e., IETD and zEVD, respectively. Tyr397 from loop 3 that is located at the edge of the S4 pocket in the crystal structure is far away in our model, indicating some limitation of the modeling. Since this Tyr does not participate directly in the binding of substrates, the variable position of this Tyr in the model does not necessarily imply different binder specificities.

The active site structural models presented in this report offer a useful guide that may help in designing potent and selective binders for caspases. Although the accuracy of these models needs to be examined by crystallographic studies, these models are nonetheless of great utility as they agree closely with our SAR data and are usefully predictive for designing new inhibitors with desirable potency and selectivity among the caspases. Besides the application to drug design, accurate homology models also allow close characterization of caspases based on their substrate specificities. The most commonly used method for categorizing caspases is based on the primary sequence analysis that led to initiators, mediators, executioners, etc. Such a categorization may be of some use in some areas, but it may also be misleading if the functional and structural analyses of enzymes are overlooked, especially before a comprehensive cascade map is constructed. In a similar vein, comparison of caspases across species where solving structures of every single caspase is almost impossible can also benefit from such structural models, insofar as they are accurate.

Using a combinatorial substrate library, Thornberry and co-workers have proposed optimal amino acids at each of the P2−P4 sites for various caspases (*22*−*25*). While the combinatorial substrate library offers an efficient tool for identifying good substrates, it cannot be definitive in identifying the best combinations of P2−P4, and sometimes could be misleading. The substrate library method measures the overall activity of 400 substrates that are made of a defined amino acid at one of three P sites and with the combination of 20 × 20 amino acids at the other two P sites. Comparing the activities given by the 20 sublibraries that contain 400 individual substrates each, one could determine the best residue(s) for a specific P position. Repeating the same procedure throughout all three P sites may give the best combination of P2−P4 as the best substrate sequence. This method would lead to identifying the best combination of three P residues for the best binder if the contribution of each P residue to the binding is independent of the other two P residues. However, in many cases, the significance of a specific P residue in contributing to the binding depends strongly on the identity of the other adjacent P residues. For instance, a P2 residue X might fit the S2 site best when the P3 and P4 residues are Y and Z, respectively, but when the P3 and P4 residues are anything but Y and Z, respectively, the same residue X could fit the S2 poorly. On the other hand, another P2 residue X′ might be a less-than-perfect residue to fit to the S2, but it is much less sensitive to the residues at the adjacent P3 and P4 positions. As such, the sum activity from 400 substrates that all have X′ at P2 may exceed the sum activity from the sublibrary that has X at the P2 position of its 400 substrates, though among those the P2X−P3Y−P4Z sequence offers the best binding. Accordingly, amino acid X′ can be determined to be a better P2 residue than amino acid X on the basis of the combinatorial substrate library screening, whereas the truly best sequence (X−Y−Z) is overlooked. Examples related to this hypothetical scenario are readily available. According to the combinatorial peptide library results, Val is one of the least favorable residues at the P2 pocket for Csp1, Csp2, Csp4, Csp5, and Csp9 (*22*, *25*), whereas His is the best residue at the P2 pocket, prevailing over the other amino acids by at least 10-fold in binding. But in a comparison of the binding of the individual peptides LEHDald and LEVDald in each case, Val is as effective as His for P2 (Table 1). A more prominent example is demonstrated by the binding of zEKDald and zKVDald to Csp2. Lys was not predicted by the peptide library screening to be a favorable amino acid at P2 or P3 for binding to Csp2. However, zEKDald is the best tetrapeptide, 11-fold better than DEVDald, in our hands for binding to Csp2, and zKVDald binds better than zEVDald by 17-fold to this caspase, contrary to what was predicted by the peptide library screening where Glu was far superior than Lys for P3.

ACKNOWLEDGMENT

We thank Mr. Chung-Jeng Lai and Mr. Robert Smidt for purifying caspases 3, 6−8, and 10, Dr. Donald Karanewsky and Dr. Vittorio Rasetti for syntheses of Ac-DEVDald, Ac-ZEKDald, Ac-ZAVDald, Ac-ZKVDald, and Ac-ZEVDald. We also thank Professor Jean-Louis Romette of the Centre National de la Recherche Scientifique, France, for his critical reading of the manuscript.

REFERENCES

1. Thornberry, N. A., and Lazebnik, Y. (1998) *Science 281*, 1312−1316.
2. Earnshaw, W. C., Martins, L. M., and Kaufmann, S. H. (1999) *Annu. Rev. Biochem. 68*, 383−424.
3. Endres, M., Namura, S., Shimizu-Sasamata, M., Waeber, C., Zhang, L., Gomez-Isla, T., Hyman, B. T., and Moskowitz, M. A. (1998) *J. Cereb. Blood Flow Metab. 18*, 238−247.
4. Holly, T. A., Drincic, A., Byun, Y., Nakamura, S., Harris, K., Klocke, F. J., and Cryns, V. L. (1999) *J. Mol. Cell. Cardiol. 31*, 1709−1715.
5. Wiessner, C., Sauer, D., Alaimo, D., and Allegrini, P. R. (2000) *Cell. Mol. Biol. 46*, 53−62.
6. Rabuffetti, M., Sciorati, C., Tarozzo, G., Clementi, E., Manfredi, A. A., and Beltramo, M. (2000) *J. Neurosci. 20*, 4398−4404.
7. Wilson, K. P., Black, J. A., Thomson, J. A., Kim, E. E., Griffith, J. P., Navia, M. A., Murcko, M. A., Chambers, S. P., Aldape, R. A., Raybuck, S. A., et al. (1994) *Nature 370*, 270−275.
8. Rotonda, J., Nicholson, D. W., Fazil, K. M., Gallant, M., Gareau, Y., Labelle, M., Peterson, E. P., Rasper, D. M., Ruel, R., Vaillancourt, J. P., Thornberry, N. A., and Becker, J. W. (1996) *Nat. Struct. Biol. 3*, 619−625.
9. Rano, T. A., Timkey, T., Peterson, E. P., Rotonda, J., Nicholson, D. W., Becker, J. W., Chapman, K. T., and Thornberry, N. A. (1997) *Chem. Biol. 4*, 149−155.
10. Mittl, P. R., Di Marco, S., Krebs, J. F., Bai, X., Karanewsky, D. S., Priestle, J. P., Tomaselli, K. J., and Grutter, M. G. (1997) *J. Biol. Chem. 272*, 6539−6547.
11. Blanchard, H., Kodandapani, L., Mittl, P. R., Marco, S. D., Krebs, J. F., Wu, J. C., Tomaselli, K. J., and Grutter, M. G. (1999) *Struct. Folding Des. 7*, 1125−1133.
12. Watt, W., Koeplinger, K. A., Mildner, A. M., Heinrikson, R. L., Tomasselli, A. G., and Watenpaugh, K. D. (1999) *Struct. Folding Des. 7*, 1135−1143.
13. Wei, Y., Fox, T., Chambers, S. P., Sintchak, J., Coll, J. T., Golec, J. M., Swenson, L., Wilson, K. P., and Charifson, P. S. (2000) *Chem. Biol. 7*, 423−432.
14. Renatus, M., Stennicke, H. R., Scott, F. L., Liddington, R. C., and Salvesen, G. S. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 14250−14255.
15. Fernandes-Alnemri, T., Litwack, G., and Alnemri, E. S. (1995) *Cancer Res. 55*, 2737−2742.
16. Fernandes-Alnemri, T., Armstrong, R. C., Krebs, J., Srinivasula, S. M., Wang, L., Bullrich, F., Fritz, L. C., Trapani, J. A., Tomaselli, K. J., Litwack, G., and Alnemri, E. S. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7464−7469.
17. Krebs, J. F., Srinivasan, A., Wong, A. M., Tomaselli, K. J., Fritz, L. C., and Wu, J. C. (2000) *Biochemistry 39*, 16056−16063.
18. Rost, B., and Sander, C. (1993) *J. Mol. Biol. 232*, 584−599.
19. Rost, B., and Sander, C. (1994) *Proteins 20*, 216−226.
20. Graybill, T. L., Dolle, R. E., Helaszek, C. T., Miller, R. E., and Ator, M. A. (1994) *Int. J. Pept. Protein Res. 44*, 173−182.
21. Chou, K. C., Tomasselli, A. G., and Heinrikson, R. L. (2000) *FEBS Lett. 470*, 249−256.
22. Thornberry, N. A., Rano, T. A., Peterson, E. P., Rasper, D. M., Timkey, T., Garcia-Calvo, M., Houtzager, V. M., Nordstrom, P. A., Roy, S., Vaillancourt, J. P., Chapman, K. T., and Nicholson, D. W. (1997) *J. Biol. Chem. 272*, 17907−17911.
23. Talanian, R. V., Quinlan, C., Trautz, S., Hackett, M. C., Mankovich, J. A., Banach, D., Ghayur, T., Brady, K. D., and Wong, W. W. (1997) *J. Biol. Chem. 272*, 9677−9682.
24. Lee, D., Adams, J. L., Brandt, M., DeWolf, W. E., Jr., Keller, P. M., and Levy, M. A. (1999) *Bioorg. Med. Chem. Lett. 9*, 1667−1672.
25. Kang, S. J., Wang, S., Hara, H., Peterson, E. P., Namura, S., Amin-Hanjani, S., Huang, Z., Srinivasan, A., Tomaselli, K. J., Thornberry, N. A., Moskowitz, M. A., and Yuan, J. (2000) *J. Cell. Biol. 149*, 613−622.

BI020593L