Structure-Based Thermodynamic Analysis of Caspases Reveals Key Residues for Dimerization and Activity†

Stefano Piana, Marialore Sulpizi, and Ursula Rothlisberger*

Institute of Molecular and Biological Chemistry, Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland

Received January 9, 2003; Revised Manuscript Received May 28, 2003

ABSTRACT: Cysteine-dependent aspartic proteases (caspases) are a family of enzymes which play a crucial role in apoptosis. Caspases accumulate in eukaryotic cells in the form of low-activity proenzyme precursors. Proteolytic cleavage of specific sites triggers conformational changes that lead to full activation and thus to the initiation of the apoptotic cascade. Several experimental observations suggest that dimerization is crucial for activity and regulation, but the underlying molecular mechanisms have not yet been completely resolved. In this work, we have used a structure-based thermodynamic analysis method [Edgcomb, S. P., and Murphy, K. P. (2000) *Curr. Opin. Biotechnol. 11*, 62−66] to calculate the free energy of association and folding for all the caspases and procaspases whose structures are known at present. In all cases, analysis of the single-residue contributions to the dimerization energy shows that 30−50% of the dimer stability originates from the highly specific interaction of 12−14 residues located at the N- and C-termini of the large and small subunits, respectively. Moreover, our calculations indicate that these residues are also critical for stabilizing the conformation of the active site loops, which in turn is crucial for the binding of substrates and inhibitors. Thus, our results help to rationalize the relation between dimerization and activity in this important class of enzymes and can be used as a starting point for an active manipulation of the monomer−dimer equilibrium for preparatory and regulatory purposes.

Caspases are a family of highly homologous enzymes that play a central role in the process of apoptosis (*1*). They accumulate in the cell as scantily active proenzymes (zymogens) (*2*), which reach the fully functional form upon proteolytic cleavage at specific sites (*3*). Activation takes place in the form of a cascade in which effector caspases (caspases 3, 6, and 7) are cleaved by initiator caspases (caspases 2 and 8−10) (*4*). In contrast with effector zymogens, initiator procaspases possess a higher residual proteolytic activity; they are thus able to self-activate under proper conditions and to instigate the apoptosis pathway (*5*). The regulation of caspase activity is a crucial issue in the life cycle of all multicellular organisms as irregular apoptosis leads to a variety of disorders such as fetal malformation and neurodegeneration (*4*) and is almost invariably associated with the development of neoplasia (*6*).

The crystal structures of all caspases determined to date (caspases 1, 3, and 7−9) reveal that the mature enzymes are tetramers composed of homodimers of heterodimers (Figure 1a,b) (*7*, *8*). Each heterodimer is composed of a large (p20) and a small subunit (p10). For the sake of simplicity, we will refer to the heterodimer as the *monomer* caspase and to the homodimer of heterodimers as the *dimer* caspase (*9*). Moreover, all throughout this paper, the caspase structures with a bound inhibitor are called *active caspases* to distinguish them from the unbound caspases (*free caspases*) and the *procaspases*.

A full molecular understanding of the activation and regulation of caspases is still lacking. However, some crucial facts have started to emerge. Caspases are synthesized as single-polypeptide chains containing both p20 and p10. Maturation of procaspases requires proteolytic cleavage of the *linker* (*10*), a highly flexible region connecting the C-terminus of the large subunit (p20) with the N-terminus of the small subunit (p10) of the heterodimer (T2 and T3, Figure 1a). Several experimental findings also indicate that dimer formation (*4*, *5*, *11*−*13*) is an essential requirement for procaspase activation (*11*, *14*) and caspase activity (*14*). Indirectly, this idea is further supported by the observation that the dimer−monomer equilibrium can be shifted toward dimer formation by adding a caspase inhibitor (*8*, *15*). However, effector procaspases 3, 6, and 7, which are characterized by a very low residual activity, are dimers even at low concentrations (*16*−*19*). Thus, dimerization, although apparently required for caspase activity, appears not to be sufficient to activate effector procaspases.

Very recently, the first two crystal structures of a procaspase, procaspase 7, have been published (*17*, *18*) and provide a major step toward the understanding of the regulation mechanism of caspase activity via dimerization and linker cleavage. The tertiary structure of procaspase 7 closely resembles that of the active enzyme with the notable exceptions that one of the two linkers is inserted between the two subunits and the folding of most of the loops that constitute the active site is different. To a lesser extent, a similar picture is also observed in the crystal structure of

† S.P. gratefully acknowledges a fellowship from the Roche Research Foundation.

* To whom correspondence should be addressed: Institute of Molecular and Biological Chemistry, Federal Institute of Technology (EPFL), 1015 Lausanne, Switzerland. Phone: +41-21-6930321. Fax: +41-21-6930320. E-mail: ursula.roethlisberger@epfl.ch.

10.1021/bi034032l CCC: $25.00
Published on Web 07/04/2003

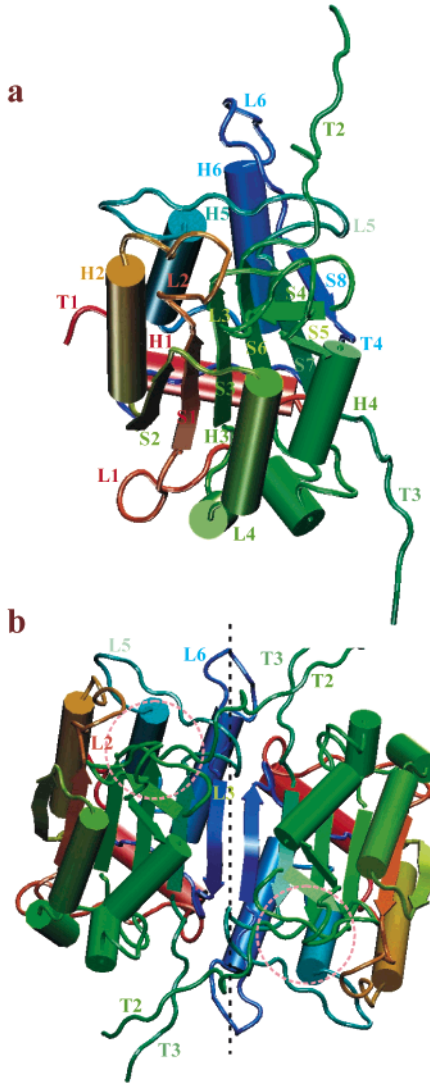


FIGURE 1: Caspase 1 tertiary structure. (a) Monomer caspase 1 (*27*). The coloring of the secondary structure goes from red to blue upon going from the N- to the C-terminus. Helices are numbered from H1 to H6, loops from L1 to L5, β-strands from S1 to S8, and the tails (C- and N-termini) from T1 to T4. (b) Caspase 1 dimer. The symmetry axis is represented by a dashed line. The positions of the active sites are indicated with pink circles.

free caspase 7 (*18*), where N-terminal T3 is folded into the intersubunit groove and active site loops L5 and L6 are partially unfolded. The differences in the conformations of the active site loops provide a first rationale for the observed low activity of the zymogen (*20*). On the basis of this structural data, the authors conclude that the interaction of the T2 C-terminus and the T3 N-terminus should be crucial in stabilizing the conformation of the active site loops, thus pointing to the requirement of dimerization and linker cleavage for the activity of caspases (*20*).

To investigate the molecular origin of these observations and to provide a detailed picture of the interaction energies at the dimer interface, we have calculated the free energy of dimerization and folding (*21*−*24*) of all the caspases whose structures have been determined to date (caspases 1, 3, and 7−9). In the special case of caspase 7, structures of the procaspase, free caspase, and active caspase are available, enabling a detailed comparison of the structure and energetics of the active site loops in different forms. Anticipating our results, we find that dimerization is crucial for the stability of the active conformation. In all caspases that have been studied, a large part (30−50%) of the dimerization energy originates from the interaction of C-terminal T2 of the first monomer with N-terminal T3 of the second monomer (Figure 1b). These tails are located in the proximity of the L5 and L6 active site loops whose conformation appears to be critical for activity. Indeed, our structure-based analysis of the folding energetics of the active site loops in caspase 7 indicates this interaction as a key element. Furthermore, because of geometric constraints imposed by the (intact) linker, this structural feature is absent in the corresponding procaspase. As a consequence, the conformation of the active site loops is only favorable for catalysis in the dimeric mature caspase as it is stabilized by the interaction of the p20 C-terminus of one monomer with the p10 N-terminus of the other. These findings strongly confirm and generalize the mechanism for the regulation of caspase activity through dimerization and linker cleavage that emerged from analysis of the crystal structures of procaspase 7 (*17*). Most importantly, the structure-based thermodynamic analysis presented here allows us to pinpoint the hot spots that are most relevant for stabilizing the active enzyme conformation and to characterize the nature of the molecular interactions of the involved residues. Thus, our calculations identify new potential target regions for the rational design of specific noncovalent caspase activation inhibitors acting at the level of proenzyme maturation.

MATERIALS AND METHODS

The starting structures for all the calculations were taken from the Protein Data Bank (*25*). The following caspase and procaspase structures were considered: 1PAU (*7*), 1CP3 (*8*), 1I3O (*26*), 1BMQ (*27*), 1IBC (*28*), 1F1J (*29*), 1I51 (*30*), 1I4O (*31*), 1KMC (*32*), 1QDU (*33*), 1QTN (*34*), 1JXQ (*11*), 1K86 (*18*), 1K88 (*18*), and 1GQF (*17*). Residues are numbered according to their relative position in procaspase 1. According to this numbering, Cys285 and His237 form the catalytic dyad. Residues of the second subunit are marked with a prime where necessary (e.g., Cys285′ and His237′). Crystallographic water molecules and the inhibitor molecules were removed where present. Hydrogen atoms were added with the XLEAP module of Amber6 (*35*). All histidine residues were considered ϵ-protonated, consistent with the H-bond pattern of the protein environment. All Asp and Glu residues were considered deprotonated, whereas Arg and Lys were considered protonated. All the other residues were taken to be neutral. The 34 linker residues that were not resolved in the GQF crystal structure (*17*) were considered disordered in the native state (*36*). Prior to the calculations of the dimerization free energies, dimer and monomer caspases underwent 100 steps of steepest descent followed by 4900 steps of conjugate gradient geometry optimization using the Amber94 force field (*37*). A generalized Born (GB) model as implemented in Amber6 was used to implicitly account for the solvation energy contribution (*38*). A cutoff of 15 Å was used for the nonbonded electrostatic and van der Waals interactions.

Dimerization (ΔG_{dim}) and folding energies [ΔG_F(M) and ΔG_F(D)] at 25 °C were calculated according to the methodology developed by Freire et al. (*21*). This approach has been shown to provide an accurate description of the thermodynamics of protein−protein interactions and folding

Table 1: Calculated Folding and Dimerization Free Energies (kcal/mol) and Dimer Interface Areas (Å^2)[a]

caspase	PDB entry	$\Delta SASA_{dim}$	ΔG_{dim}[b]	$\Delta G_F(D)$[b]	$\Delta G_F(M)$[b]
1	1BMQ	5144	−57	−61	−4
	1IBC	5302	−58	−61	−3
3	1PAU	4260	−44	−42	2
	1CP3	4302	−45	−49	−4
	1I3O	5043	−51	−31	20
7	1F1J	4473	−47	−39	8
	1I51	4590	−50	−34	16
	1I4O	4669	−50	−30	20
	1KMC	5034	−60	−40	20
8	1QDU	5366	−55	−39	16
	1QTN	5038	−47	−33	14
9	1JXQ	4434	−45	−43	2
procaspase 7	1K88	3302	−34	−39	−5
	1GQF	3887	−40	−33	7
free caspase 7	1K86	3468	−35	−30	5

[a] Calculated folding free energies for the dimer [$\Delta G_F(D)$] and isolated monomers [$\Delta G_F(M)$]. $\Delta SASA_{dim}$ is the subunit−subunit interface area. ΔG_{dim} is the calculated dimerization energy. [b] Calculated according to ref *21*.

for a large number of systems (*21*, *23*, *39*−*47*). It allows prediction of free energies of protein folding or binding free energies at a reference temperature by explicitly calculating the entropy ΔS, enthalpy ΔH, and heat capacity ΔC_p changes. These terms are expressed as linear functions of the changes in polar (ASA_{POL}), apolar (ASA_{AP}), and hydroxyl group (ASA_{OH}) accessible surface areas: $\Delta C_p = \Delta C_{p,ap} + \Delta C_{p,pol} = a_c(T) \times \Delta ASA_{ap} + b_c(T) \times \Delta ASA_{pol} + c_c(T) \times \Delta ASA_{OH}$, where $a_c(T) = 0.45 + 2.63 \times 10^{-4} \times (T - 298.15) - 4.2 \times 10^{-5} \times (T - 298.15)^2$ cal K^{-1} mol^{-1} Å^{-1}, $b_c(T) = -0.26 + 2.85 \times 10^{-4} \times (T - 298.15) + 4.31 \times 10^{-5} \times (T - 298.15)^2$ cal K^{-1} mol^{-1} Å^{-1}, and $c_c = 0.17$ cal K^{-1} mol^{-1} Å^{-1}. The bulk of the enthalpy change is also expressed in terms of the ΔASA. At the reference temperature of 333.15 K, it can be written as $\Delta H_{gen}(333) = -8.44\Delta ASA_{ap} + 31.4\Delta ASA_{pol}$. The enthalpy at any other temperature can be calculated as $\Delta H_{gen}(T) = \Delta H_{gen}(333) + \Delta C_p(T - 333.15)$. There are two main contributions to the entropy term: a solvation and a conformational entropy change. The solvation entropy $\Delta S_{solv}(T)$ is temperature-dependent and is defined in terms of the heat capacity as follows: $\Delta S_{solv}(T) = \Delta S_{solv,ap}(T) + \Delta S_{solv,pol}(T) = \Delta C_{p,ap} \times \ln(T/T^*_{S,ap}) + C_{p,pol} \times \ln(T/T^*_{S,pol})$. The conformational entropies are evaluated by considering the following three contributions for each amino acid: $\Delta S_{bu\text{-}ex}$, associated with the transfer of a buried side chain to the surface of the protein; $\Delta S_{ex\text{-}u}$, the side chain entropy change associated with the backbone folding, and ΔS_{bb}, the backbone entropy change associated with backbone folding. These contributions have been estimated for each amino acid and are reported in ref *48*. More details of the method can be found in refs *22*−*24*, *47*, and *48*. The accessible surface area was defined as the surface accessible to a spherical probe with a 1.4 Å radius (*49*). Reference ASA values for the unfolded state were calculated for each residue X as the solvent accessible surface area in the Gly-X-Gly tripeptide. Details of the method can be found in refs *22*−*24*, *47*, and *48*.

To validate our implementation of the method presented in ref *21*, we performed calculations on HIV-1 protease, the β-trypsin−pancreatic trypsin inhibitor complex, and the *Streptomyces* subtilisin inhibitor (SSI) for which both structural (PDB entries 1HHP, 2PTC, and 3SSI) (*50*−*52*) and thermodynamic (*36*, *53*−*56*) data are available. The calculated dimerization energies for the three systems (−39, −21, and −21 kcal/mol, respectively) are in good agreement with the values derived from experiments (−40.6, −18.1, and −16 kcal/mol, respectively) (*36*, *55*, *56*). Furthermore, thermodynamic parameters calculated for the unfolding of HIV-1 protease at 25 °C (*36*) ($\Delta G = 12$ kcal/mol, $\Delta S = -90$ cal K^{-1} mol^{-1}, $\Delta H = -9.4$ kcal/mol, and $\Delta C_p = 3.4$ kcal K^{-1} mol^{-1}) are in very good agreement with the experimental values ($\Delta G = 14.6$ kcal/mol, $\Delta S = -79.15$ cal K^{-1} mol^{-1}, $\Delta H = -9.0$ kcal/mol, and $\Delta C_p = 3.2$ kcal K^{-1} mol^{-1}) (*39*). Reasonable agreement is also found for the calculated thermodynamic parameters for thermal unfolding of SSI at 86 °C [$\Delta S = 857$ cal K^{-1} mol^{-1}, $\Delta H = 226$ kcal/mol, and $\Delta C_p = 2.8$ kcal K^{-1} mol^{-1} (calculated) compared to $\Delta S = 603$ cal K^{-1} mol^{-1}, $\Delta H = 216$ kcal/mol, and $\Delta C_p = 2.1$ kcal K^{-1} mol^{-1} (experimental)]. The large dimerization energies calculated for HIV-1 protease and, to a lesser extent, for SSI are consistent with the experimental evidence which shows that dimerization is a folding event (*36*, *54*). Indeed, in HIV-1 protease and SSI, folding and dimerization proceed through a single step, as separate folding of the isolated monomers is not possible. Thus, dimerization appears to be a folding event that is crucial for stabilizing the final conformation (*36*, *54*). On the contrary, in the β-trypsin−pancreatic trypsin inhibitor complex, the isolated molecules of trypsin and its inhibitor are stable. The structures of the isolated molecules have been determined (PDB entries 2PTN and 1BPI) (*57*, *58*) and differ little from those in the complex (*52*) (root-mean-square deviation of the backbone atoms of <1 Å). Thus, in this case, dimerization induces only very limited folding. As a consequence, the calculated association free energy directly reflects the dissociation constant.

Validation calculations were also performed using the method developed by Honig and co-workers (*59*) or by simply considering the energy differences between the dimer and monomer as given by the Amber force field as a measure of the binding enthalpy. All the methods employed gave results in fair agreement. The results and details of these calculations are presented in the Supporting Information.

Active site folding free energies were calculated according to the methodology developed by Freire et al. (*21*) and outlined above. The starting structure for active caspase 7 was PDB entry 1I51 (*30*); those for free caspase 7 and procaspase 7 were PDB entries 1K86 (*18*) and 1GQF (*17*),

respectively. To make the three sequences identical, some residues were modified. In particular, the first Thr of the p20 subunit present in I51 and K86 but not in GQF was eliminated; the catalytic Cys285 and Asp268 were mutated into Ala. In procaspase 7, the linker residues that were not present in the other two structures were eliminated.

MD Simulations of Monomer and Dimer Caspase 3. The structural model of the active form of caspase 3 was built starting from 1PAU (*7*). Caspase 3 was immersed in an 81.0 Å × 65.3 Å × 74.2 Å box containing 10673 water molecules, corresponding to a density of 1.04 g/cm^3. The classical MD simulations were carried out using the Amber6 suite of programs (*35*). The force field parameters were those of the Amber94 force field (*37*). For water, we used the TIP3P model. Electrostatic interactions were calculated with the Ewald particle mesh method (*60*−*63*) with a mesh of 81 × 72 × 75 points. A cubic interpolation between the points was used. Constant temperature and pressure conditions were achieved by coupling the system to a Berendsen's thermostat and barostat (*64*), with relaxation times of 2.0 and 3.0 ps, respectively. Bonds involving hydrogen atoms were constrained to their equilibrium bond length with the SHAKE algorithm (*65*). The time step was 1.5 fs.

The caspase 3 dimer structure was relaxed with 700 steps of energy minimization. Subsequently, the system was equilibrated for 150 ps at 300 K and 1 atm. Finally, 4.0 ns of MD simulation trajectory was collected for analysis. The structural model for monomer caspase 3 was built starting from the structure of the caspase dimer obtained after 1.17 ns of MD simulation. For this purpose, one of the caspase 3 subunits was removed. The system was then equilibrated at 300 K and 1 atm with 300 ps of NPT MD simulation. Finally, 3.5 ns was collected for analysis.

RESULTS AND DISCUSSION

Dimerization and Folding Free Energies. Dimerization and folding free energies, ΔG_F and ΔG_{dim}, respectively (*21*), were calculated for 15 different caspase structures (Table 1). Where available, several different crystallographic data sets were considered for each member of the caspase family. In particular, two structures were considered for caspase 1 [PDB entries 1BMQ (*27*) and 1IBC (*28*)] and caspase 8 [PDB entries 1QDU (*33*) and 1QTN (*34*)], three for caspase 3 [PDB entries 1PAU (*7*), 1CP3 (*8*), and 1I3O (*26*)], and four for caspase 7 [PDB entries 1F1J (*29*), 1I51 (*30*), 1I4O (*31*), and 1KMC (*32*)]. For caspase 9, only one structure is available [PDB entry 1JXQ (*11*)]. Dimerization and folding energies were also calculated for free caspase 7 [PDB entry 1K86 (*18*)] and the two available structures of procaspase 7 [PDB entries 1K88 (*18*) and 1GQF (*17*)]. Comparison of the results obtained for different structures of identical members of the caspase family indicates that the calculations are robust with respect to small variations in the crystal structures induced by crystal packing or inhibitor binding. Total folding and dimerization free energies vary at most by 10 and ∼5 kcal/mol, respectively. Most importantly, the relative accuracy of the single-residue contributions is extremely high. The standard deviation that can be calculated for each residue with respect to the average residue value is as small as 0.07 kcal/mol (Table 1).

A number of factors can introduce possible sources of errors in addition to variations in the crystallographic data. These include the approximation of removing any inhibitor molecule from the calculations, the large number of poorly resolved residues present in several crystal structures, and the relatively large size of the proteins that were investigated (∼500 amino acids). However, we would like to stress that this work is aimed primarily at a characterization of the relative differences and similarities among the various caspases as well as at a qualitative identification of the residues that are most relevant for activation and enzymatic activity. Clearly, providing values for folding and dimerization energies with chemical accuracy is beyond the scope of this work.

As an indication of the relative accuracy of the dimerization energy calculations, we notice that folding energies ΔG_F(D) calculated for the very similar caspase dimers 3 and 7 range between −30 and −50 kcal/mol (Table 1). Differences in the folding energies for specific members can originate from conformational changes induced by the binding of the different inhibitors, as well as from different crystallization conditions. The former effect is clearly apparent in the folding energies ΔG_F(M) calculated for the isolated monomers. ΔG_F(M) values of free caspase 7 or caspases bound to small covalent peptidomimetic inhibitors (1K86, 1PAU, 1CP3, and 1F1J) are much smaller [ΔG_F(M) = −5 to 8 kcal/mol) (Table 1) than those calculated for caspase 3 and 7 bound to the bulky inhibitor Xiap [1I3O, 1I51, 1I4O, and 1KMC; ΔG_F(M) = 16−20 kcal/mol] (Table 1). This result suggests that Xiap binding induces some destabilizing conformational changes that are partially compensated by a decrease in the dimerization free energy [ΔG_F(D) ranging from −50 to −60 kcal/mol in Xiap-bound caspases compared to −35 to −47 kcal/mol for other inhibitors] (Table 1). Despite the changes in the absolute value of the dimerization energy, the single-residue contributions to the dimerization are surprisingly well conserved not only between the two dimers of the same crystal structure or different crystal structures of the same caspase but very often also between crystal structures of different members of the caspase family (Figure 2).

There are a number of other noteworthy trends in the folding and dimerization energies of the caspase family. Caspase 1 appears to be characterized by the largest folding stability [ΔG_F(D) = −61 kcal/mol as compared to a typical value of −30 to −43 kcal/mol calculated for the other members] (Table 1). This caspase features an extended dimer interface characterized by a number of salt bridges that are not present in the other structures.

The dimerization energies ΔG_F(D) calculated for caspase 8 (−33 to −39 kcal/mol) are within the range calculated for caspases 3, 7, and 9 (Table 1). However, for this enzyme, the folding of the isolated monomers [ΔG_F(M) = 14−16 kcal/mol] appears to be less favorable than that of other caspases bound to peptidomimetic inhibitors [ΔG_F(M) = −4 to 8 kcal/mol) (Table 1). A possible rationale could be that the binding of peptidomimetic inhibitors induces larger rearrangements in the structure of free caspase 8 than in other members of the family.

Finally, the folding free energies calculated for procaspase 7 [ΔG_F(D) = −33 to −40 kcal/mol] are in reasonable agreement with the value of −25.8 kcal/mol measured for the very similar procaspase 3 (*66*). It is worth noting that dimerization energies calculated for procaspase and free

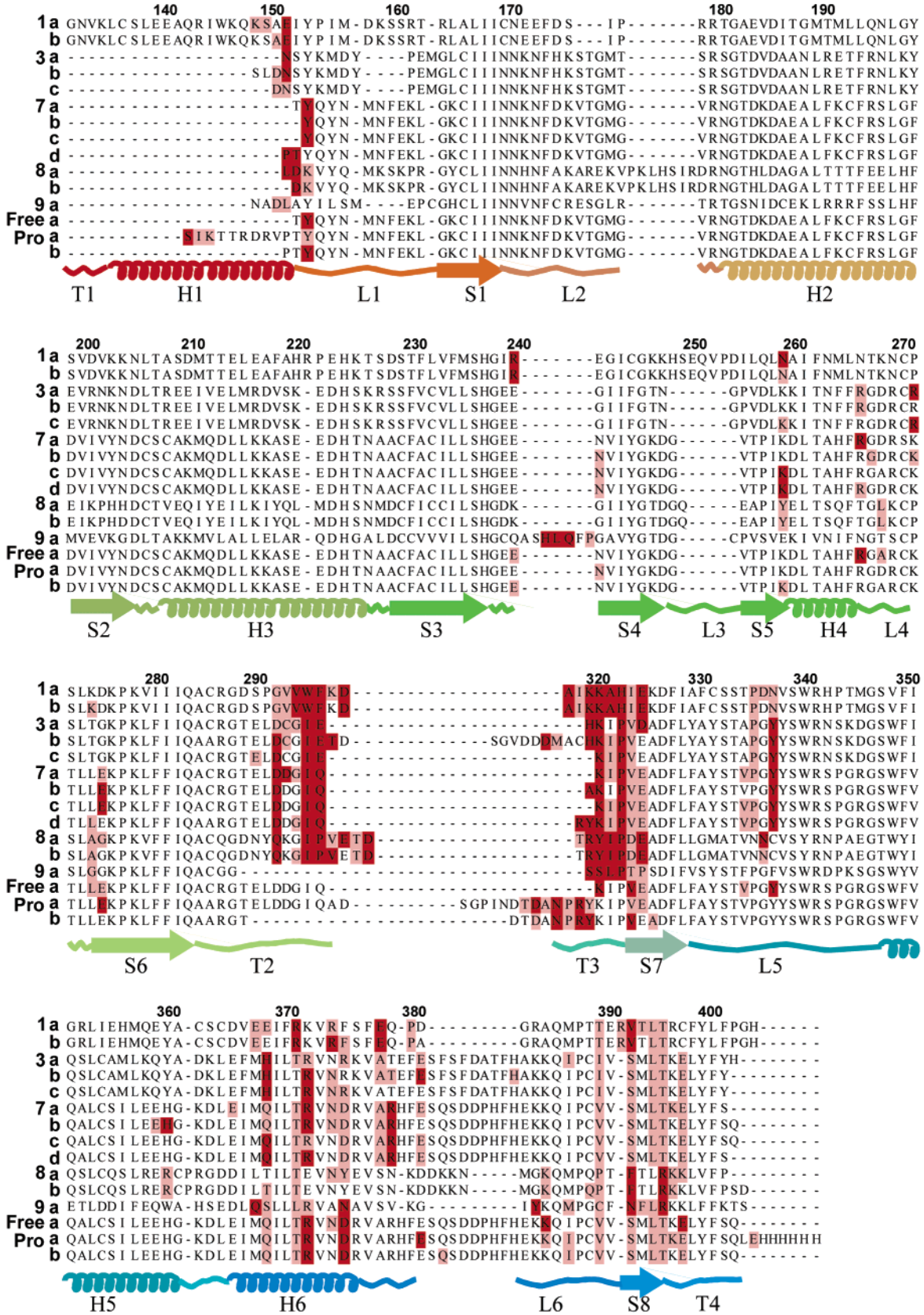


FIGURE 2: Single-residue contributions to the dimerization energy. Each residue in the picture is colored according to its contribution to the dimerization free energy (*21*): blue for greater than 0.5 kcal/mol, white for 0.5 to −0.3 kcal/mol, pink for −0.3 to −0.7 kcal/mol, and red for less than −0.7 kcal/mol. Only one subunit for each of the 15 structures that were studied is presented for the sake of clarity as the other subunits exhibit very similar patterns: (1a) 1IBC, (1b) 1BMQ, (3a) 1CP3, (3b) 1I3O, (3c) 1PAU, (7a) 1F1J, (7b) 1I51, (7c) 1I4O, (7d) 1KMC, (8a) 1QDU, (8b) 1QTN, (9a) 1JXQ, (Free a) 1K86, (Pro a) 1K88, and (Pro b) 1GQF. At the bottom of the graph, the tertiary structure of caspase 1 is indicated as reference; the coloring and labeling of the secondary structural motifs follow the scheme of Figure 1.

caspase 7 are smaller than those calculated for the active enzymes (ΔG_{dim} of −35 to −40 kcal/mol compared to values of −47 to −60 kcal/mol for active caspase 7).

The calculated folding energies indicate that in all cases, separation of the dimer subunits produces marginally unstable isolated monomers [ΔG_F(M) ranges from −5 to 20 kcal/mol] (Table 1). This observation suggests that the actual structure of the isolated monomers might be somehow different with respect to the active conformation observed in the dimer. Thus, dimerization provides a substantial contribution to the stabilization of the dimer conformation. These findings are consistent with the experimental evidence that dimerization is a folding event that substantially contributes to the conformational stability of procaspase 3 (*66*).

Single-Residue Decomposition of the Dimerization Energy. To identify possible hot spots for dimerization, the ΔG_F(D) values were decomposed in terms of single-residue contributions (Figure 2). A first surprising result coming from this calculation is that all caspases show very similar patterns. It turns out that in all the inhibitor-bound caspase structures more than 90% of the total contributions to the dimerization free energy originate from only three contact regions, namely, S8, H6, and T2−T3. The first corresponds to the residues in the core of the interface that form a buried antiparallel β-sheet that connects the two p10 chains of each monomer. The second region corresponds to the contact between the two long H6 α-helices of the p10 subunit. The third region encompasses five to nine of the C-terminal residues of the p20 chain of one monomer (T2) and four to eight N-terminal residues of the p10 chain of the other (T3) (Figure 1b). Remarkably, this region contributes 30−50% to the total dimerization energy in all the active caspases. The T2 and T3 tails are connected by a dense hydrogen bond network. They are also anchored to the protein surface and the selectivity loop of the nearby active site by a number of salt bridges and/or hydrophobic interactions (Figure 3). The corresponding sequences are poorly conserved among different caspases (*67*) with the exception of position 322 where either a proline or a histidine is found throughout and position 295 that is always occupied by a bulky hydrophobic residue. In caspases 1, 4, 5, and 13 where His is present in position 322, a Trp is found at position 295, whereas in other caspases, either Ile, Phe, or Val is present. The two conserved residues (295 and 322) provide major contributions to the dimerization free energy. His322, for instance, is the energetically most important residue for the dimerization free energy in caspase 1, a result that correlates well with the experimental finding that mutations in this position completely suppress caspase 1 activity (*68*).

Residue 295 is buried in a hydrophobic pocket strongly interacting with the hydrophobic side chains of residues 321 and 323 (Figure 3) and, to a lesser extent, residues 274 and 275. In the case of caspase 1, the C-terminal conformation is slightly different and Val321 is buried in the same cavity. Indeed, in all caspases where Trp is found in position 322, Val or Leu is found in position 321, thus supporting the idea that either one of the two binding modes is conserved throughout.

Our calculations can also offer a first rationale for some surprising results from mutation experiments on caspase 1. The dimer interface of active caspase 1 is characterized by the presence of a number of salt bridges. Two of the involved residues (Glu390 and Arg391) were mutated to Arg and Glu, respectively (*14*). As expected, the R391E mutant was not able to dimerize, but unexpectedly, the E390R mutant exhibited wild-type behavior (*14*). Our calculations clearly demonstrate that the salt bridge formed by Arg391 with Glu324 is energetically more important [$\Delta G_F(D)_{391} = -2$

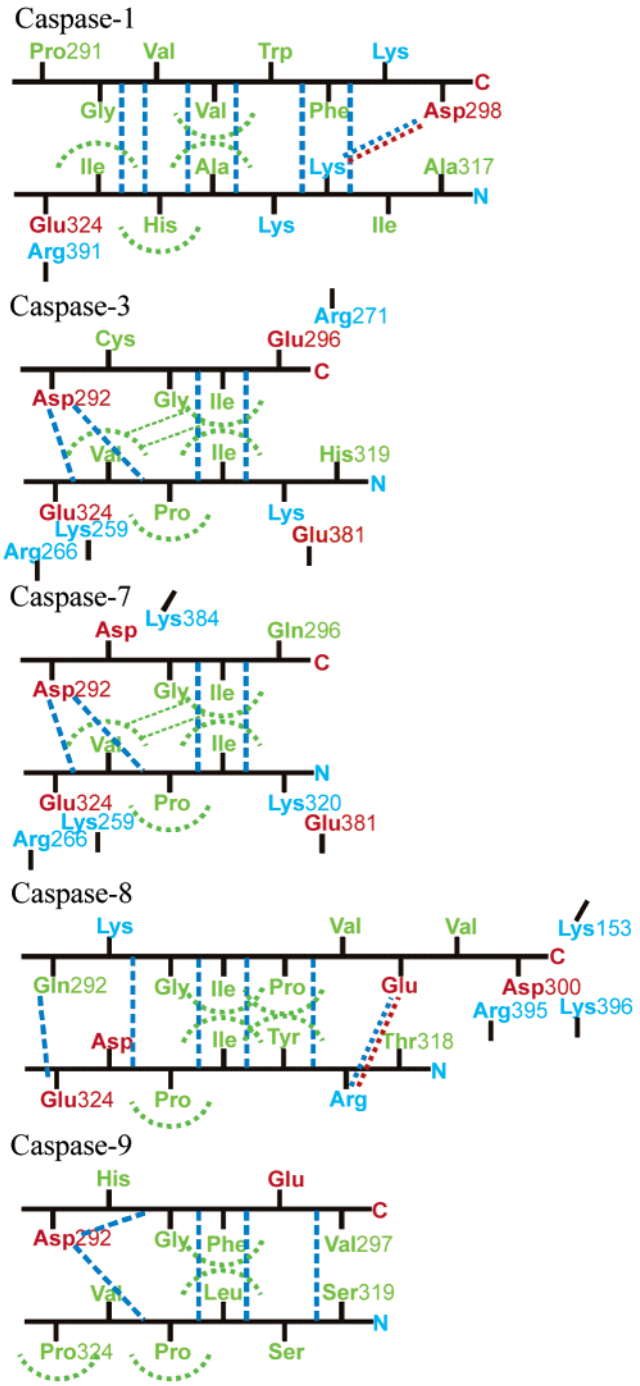


FIGURE 3: Interaction of the p20 C-terminus with the p10 N-terminus. Hydrogen bonds are represented by blue dashed lines; hydrophobic interactions are represented by green dashed lines, and salt bridges are represented by blue and red dashed lines. Hydrophobic pockets are depicted as green semicircles. Residues are colored red (negatively charged), blue (positively charged), and green (neutral). Despite the differences in sequence among various caspases, the hydrophobic interactions between the C- and N-terminus appear to be conserved in all the caspases whose structures have been determined.

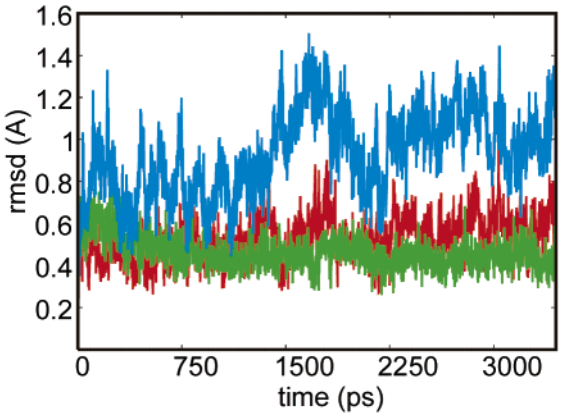
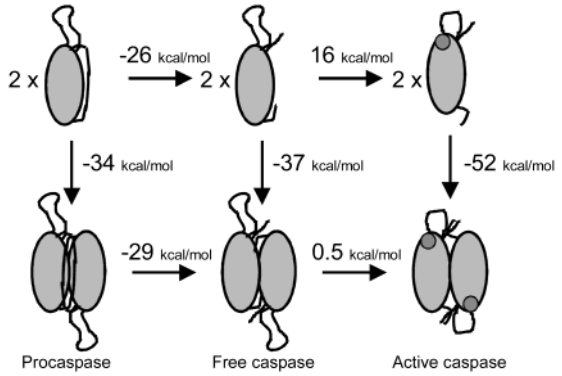


FIGURE 4: Flexibility of the active site loops in monomer and dimer caspase 3. Comparison of the root-mean-square deviation (angstroms) calculated for the selectivity loop (L8 in Figure 1, residues 380−385) as a function of time (picoseconds) during the MD simulation of the caspase 3 monomer (blue) and dimer (red and green).

Scheme 1



kcal/mol] for the dimerization free energy than the one formed by Glu390 [$\Delta G_F(D)_{390} = -0.3$ kcal/mol].

A residue decomposition of the dimerization free energy also enables an energetic analysis of the effects of the conformational changes that have been observed upon activation. In free caspase 7, the T3 N-terminus is folded inside the intersubunit groove and is not in contact with the T2 C-terminus (*18*). In the corresponding procaspase, the formation of this contact is prevented by the presence of the linker connecting the p10 and p20 subunits. As a consequence, the residues from the T3 N-terminus still give some contribution to the dimerization energy (Figure 2), while residues from T3 are mostly disordered and do not contribute significantly. This missing interaction mostly accounts for the difference between the dimerization energies calculated for active caspase 7, free caspase 7, and procaspase 7.

Relative Folding Stability of the Monomer and Dimer in Procaspase and Free and Active Caspase 7. Crystal structures of free caspase 7 and procaspase 7 reveal that, although the overall folding is very similar, the active site loops display a different conformation with respect to the structures of active caspase 7. This indicates that substantial conformational changes in the active site loops are required for the binding of substrates or inhibitors. We have calculated the free energy required for this conformational change in the cases of dimer and monomer caspase 7, in the pro, free, and active forms (Scheme 1). A crystal structure of a monomeric caspase has not been resolved yet. For this reason, the starting structures for the monomers were taken from the crystal structure of the dimer enzymes and their geometries were optimized locally within a molecular mechanics approach. Within this approximation, it turns out that in the *monomer* ~8 kcal/mol is required to change the conformation of the active site loops from the free to the active form, whereas the corresponding forms of the *dimers* are almost isoenergetic (Scheme 1). This further confirms that the structural stability of the active form is strongly correlated with dimerization. In particular, our calculations predict that dimerization stabilizes the active fold with respect to the free caspase. Analogous findings emerge from an analysis of the dynamical properties of the active site loops obtained from MD simulations of the highly homologous caspase 3. Three nanoseconds of NTP molecular dynamics simulations was performed on monomer and dimer free caspase 3. During the time scale that was investigated, the overall tertiary structure is maintained in both simulations. However, the monomer appears to be more flexible than the dimer, as indicated by the larger average rmsd with respect to the starting structure (1.5 and 0.9 Å for monomer and dimer simulation, respectively). The conformation of loop L6 in the two MD simulations is of particular interest. This loop is part of the active site and is crucial for substrate recognition. Figure 3 shows the rmsd calculated for the backbone atoms of L6 during the MD simulations of the monomer and dimer. In the simulation of the monomer, the L6 loop exhibits large fluctuations and deviates from the active enzyme structure by as much as 1.2 Å, compared to 0.5−0.6 Å in the case of the dimer. We conclude that both the folding free energy calculations and the MD simulations indicate that the active site loop conformation is less stable in the monomer than in the dimer. As a consequence, substrate binding is predicted to be energetically more favorable in the dimeric form. This observation provides a possible rationale for the requirement of dimerization for caspase activity.

In the case of dimeric procaspase 7, folding to the active conformation appears to be energetically favorable although prevented by geometric constraints imposed by the linker (*17*, *18*). Removal of this constraint by proteolytic cleavage of the linker is thus expected to be followed by a refolding of the active site loops toward the free and active conformations. Refolding of monomer procaspase 7 to the free form is also energetically favorable. A simple molecular modeling approach indicates that, in the monomer, the linker is long enough to be reconnected without inducing strain in the structure. In summary, these observations suggest that the folding of the active site loops in monomeric procaspase 7 and free caspase 7 should be similar. Some support to this hypothesis is provided from the calculated folding free energies. The estimated $\Delta G_F(D)$ for the formation of dimeric procaspase 7 from two free caspase 7-like monomers (approximately −8 kcal/mol, Scheme 1) is comparable to the value of −10.5 kcal/mol associated with dimerization for the very similar procaspase 3 (*66*).

The relative folding free energies calculated for the active caspase 7 monomer and dimer strongly suggest that the stability of the active site folding is correlated with dimerization. To identify the specific residues, which are most relevant for linking dimerization with active site stability, we have subtracted the folding free energies calculated for

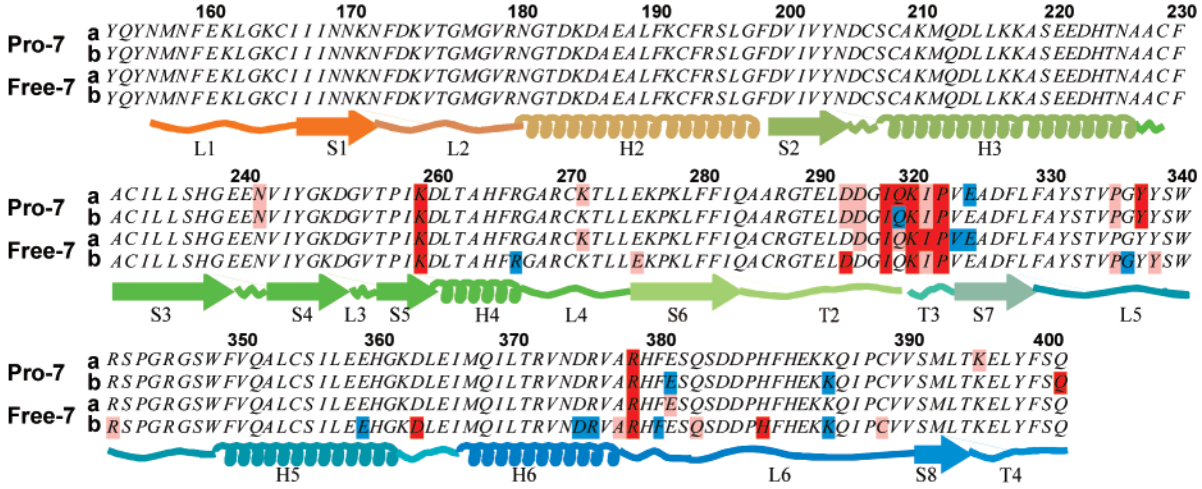


FIGURE 5: Single-residue contributions to the folding stability in the caspase 7 dimer compared to the caspase 7 monomer. The difference in the folding free energies between monomer and dimer forms of active caspase 7 is plotted for each residue. Coloring of residues and secondary structural motifs follows the scheme of Figure 2. (Pro-7) Folding of active caspase 7 from procaspase 7: (a) first subunit and (b) second subunit. (Free-7) Folding of active caspase 7 from free caspase 7: (a) first subunit and (b) second subunit.

monomer caspase 7 from the corresponding values of the dimer. If dimerization and folding are two unrelated events, the difference between the two quantities (as well as its single-residue decomposition) is zero. Scheme 1 shows that the difference between the two free energies is −15.5 kcal/mol for the folding of active caspase 7 from its free form and −18.5 kcal/mol with respect to the proenzyme. The single-residue contributions to the difference are reported in Figure 5 and show that the residues that contribute to the stabilization of the active conformation in the dimer with respect to the monomer are very much the same independent of whether active caspase 7 is folded starting from procaspase 7 or it is folded from free caspase 7. Three hot spots appear to be particularly relevant in this respect: Lys259, Arg379, and the T2−T3 contact region. Lys259 forms a salt bridge with Glu324 that is part of the T3 terminus (Figure 3), and Arg379 is involved in an intersubunit salt bridge with Glu275. However, most of the contributions to the folding free energy of the active site loops through dimerization originate from residues located in the T2−T3 contact region, namely, Asp292, Ile295, Lys320, Ile321, and Pro322. Asp292, located in the T2 segment, is connected to the backbone of T3 by a network of hydrogen bonds (Figure 3); Ile295 and Ile323, located in T2 and T3, respectively, form a hydrophobic pocket (Figure 3). This interaction appears to be of particular interest as it is conserved in all caspase structures. The charged side chain of Lys320 forms hydrogen bonds with the backbone of Glu382 and Asp293. Finally, Pro322 is folded in a hydrophobic pocket formed by active site loop L6 (the so-called *selectivity loop*) residues Ala378 and Ile386 and the nonpolar part of the side chain of Lys384.

In summary, our calculations indicate that dimerization and active site folding are correlated events. The residues that are most relevant for this correlation are located in the T2−T3 contact region and are connected by hydrogen bonds and hydrophobic interactions to active site loop L6.

A consistent picture emerges from this study linking dimerization and enzymatic activity. Dimerization is a folding event that is important for stabilizing the conformation of the active site loops. In particular, interaction of the T2 C-terminus of one monomer with the T3 N-terminus of the other is crucial both for the structural stability of the dimer and for establishing an active site loop conformation appropriate for the binding of substrates. Our calculations provide further confirmation of the activation mechanism proposed on the basis of the X-ray diffraction data (*17*, *18*). Moreover, they also allow identification of the specific molecular interactions which are most important for dimerization and active site stability, namely, the interaction of residue 322 with active site loop L6 and the interaction of residues 294 and 295 with the hydrophobic pocket formed by residues 321 and 323. The hydrophobic and electrostatic interactions that stabilize the conformations of the T3 N-terminus and the T2 C-terminus in the active caspase define a conserved binding pocket located in the intersubunit groove. This binding pocket could potentially be exploited as a new target for molecules that stabilize or destabilize the active conformation of active site loops L5 and L6. The spatial arrangement and physical nature of the T2−T3 interactions differ in the various caspases; thus, it might be possible to design molecules that are selective for specific members of the family.

SUPPORTING INFORMATION AVAILABLE

Comparison of different methods for calculating the free energy of association. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Earnshaw, W. C., Martins, L. M., and Kaufmann, S. H. (1999) *Annu. Rev. Biochem. 68*, 383−424.
2. Ramage, P., Cheneval, D., Chvei, M., Graff, P., Hemmig, R., Heng, R., Kocher, H. P., Mackenzie, A., Memmert, K., and Revesz, L. (1995) *J. Biol. Chem. 270*, 9378−9383.
3. Yamin, T. T., Ayala, J. M., and Miller, D. K. (1996) *J. Biol. Chem. 271*, 13273−13282.
4. Chang, H. Y., and Yang, X. (2000) *Microbiol. Mol. Biol. Rev. 64*, 821−846.

5. Muzio, M., Stockwell, B. R., Stennicke, H. R., Salvesen, G. S., and Dixit, V. M. (1998) *J. Biol. Chem. 273*, 2926−2930.
6. Kaufmann, S. H., and Gores, G. J. (2000) *BioEssays 22*, 1007−1017.
7. 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.
8. 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.
9. Walker, N. P., Talanian, R. V., Brady, K. D., Dang, L. C., Bump, N. J., Ferenz, C. R., Franklin, S., Ghayur, T., Hackett, M. C., and Hammill, L. D. (1994) *Cell 78*, 343−352.
10. Mosley, B., Urdal, D. L., Prickett, K. S., Larsen, A., Cosman, D., Conlon, P. J., Gillis, S., and Dower, S. K. (1987) *J. Biol. Chem. 262*, 2941−2944.
11. 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.
12. Colussi, P. A., Harvey, N. L., Shearwin-Whyatt, L. M., and Kumar, S. (1998) *J. Biol. Chem. 273*, 26566−26570.
13. MacCorkle, R. A., Freeman, K. W., and Spencer, D. M. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 3655−3660.
14. Gu, Y., Wu, J., Faucheu, C., Lalanne, J. L., Diu, A., Livingston, D. J., and Su, M. S. (1995) *EMBO J. 14*, 1923−1931.
15. Talanian, R. V., Dang, L. C., Ferenz, C. R., Hackett, M. C., Mankovich, J. A., Welch, J. P., Wong, W. W., and Brady, K. D. (1996) *J. Biol. Chem. 271*, 21853−21858.
16. Kang, B. H., Ko, E., Kwon, O. K., and Choi, K. Y. (2002) *Biochem. J. 364*, 629−634.
17. Riedl, S. J., Fuentes-Prior, P., Renatus, M., Kairies, N., Krapp, S., Huber, R., Salvesen, G. S., and Bode, W. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 14790−14795.
18. Chai, J., Wu, Q., Shiozaki, E., Srinivasula, S. M., Alnemri, E. S., and Shi, Y. (2001) *Cell 107*, 399−407.
19. Pop, C., Chen, Y. R., Smith, B., Bose, K., Bobay, B., Tripathy, A., Franzen, S., and Clark, A. C. (2001) *Biochemistry 40*, 14224−14235.
20. Salvesen, G. S., and Dixit, V. M. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 10964−10967.
21. Luque, I., Mayorga, O. L., and Freire, E. (1996) *Biochemistry 35*, 13681−13688.
22. Hilser, V. J., Gomez, J., and Freire, E. (1996) *Proteins 26*, 123−133.
23. Xie, D., and Freire, E. (1994) *Proteins 19*, 291−301.
24. Xie, D., and Freire, E. (1994) *J. Mol. Biol. 242*, 62−80.
25. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G. L., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res. 28*, 235−242.
26. Riedl, S. J., Renatus, M., Schwarzenbacher, R., Zhou, Q., Sun, C., Fesik, S. W., Liddington, R. C., and Salvesen, G. S. (2001) *Cell 104*, 791−800.
27. Okamoto, Y., Anan, H., Nakai, E., Morihira, K., Yonetoku, Y., Kurihara, H., Sakashita, H., Terai, Y., Takeuchi, M., Shibanuma, T., and Isomura, Y. (1999) *Chem. Pharm. Bull. 47*, 11−21.
28. 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.
29. 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.
30. Chai, J., Shiozaki, E., Srinivasula, S. M., Wu, Q., Dataa, P., Alnemri, E. S., and Shi, Y. (2001) *Cell 104*, 769−780.
31. Huang, Y., Park, Y. C., Rich, R. L., Segal, D., Myszka, D. G., and Wu, H. (2001) *Cell 104*, 781−790.
32. Salvesen, G. S., and Duckett, C. S. (2002) *Nat. Rev. Mol. Cell Biol. 3*, 401−410.
33. 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.
34. 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.
35. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M. F., Ferguson, D. M., Radmer, R. J., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) Amber, version 5.0, University of California, San Francisco.
36. Todd, J. M., Semo, N., and Freire, E. (1998) *J. Mol. Biol. 283*, 475−488.
37. Cornell, W. D., Cieplack, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
38. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) *J. Am. Chem. Soc. 112*, 6127−6129.
39. Todd, J. M., Semo, N., and Freire, E. (1998) *J. Mol. Biol. 283*, 475−488.
40. Luque, I., Todd, J. M., Gomez, J., Semo, N., and Freire, E. (1998) *Biochemistry 37*, 5791−5797.
41. Todd, J. M., and Freire, E. (1999) *Proteins: Struct., Funct., Genet. 36*, 147−156.
42. Murphy, K. P., Xie, D., Garcia, K. C., Amzel, L. M., and Freire, E. (1993) *Proteins 15*, 113−120.
43. Burrows, S. D., Doyle, M. L., Murphy, K. P., Franklin, S. G., White, J. R., Brooks, I., McNulty, D. E., Scott, M. O., Knutson, J. R., et al. (1994) *Biochemistry 33*, 12741−12745.
44. Baker, B. M., and Murphy, K. P. (1997) *J. Mol. Biol. 268*, 557−569.
45. Baker, B. M., and Murphy, K. P. (1998) *Methods Enzymol. 295*, 294−315.
46. Murphy, K. P. (1999) *Med. Res. Rev. 19*, 333−339.
47. Edgcomb, S. P., and Murphy, K. P. (2000) *Curr. Opin. Biotechnol. 11*, 62−66.
48. Lee, K. H., Xie, D., Freire, E., and Amzel, L. M. (1994) *Proteins 20*, 68−84.
49. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
50. Miller, M., Jaskolski, M., Rao, J. K. M., Mohana, J. K., Leis, J., and Wlodawer, A. (1989) *Nature 337*, 576−579.
51. Mitsui, Y., Satow, Y., Watanabe, Y., and Iitaka, Y. (1979) *J. Mol. Biol. 131*, 697−724.
52. Huber, R., Bode, W., Kukla, D., Kohl, U., and Ryan, C. A. (1975) *Biophys. Struct. Mech. 1*, 189−201.
53. Komiyama, T., Miwa, M., Yatabe, T., and Ikeda, H. (1984) *J. Biochem. 95*, 1569−1575.
54. Takahashi, K., and Sturtevant, J. M. (1981) *Biochemistry 20*, 6185−6190.
55. Akasaka, K., Fujii, S., Hayashi, F., Rokushika, S., and Hatano, H. (2002) *Biochem. Int. 5*, 637−642.
56. Vincent, J. P., and Lazdunski, M. (1972) *Biochemistry 11*, 2967−2977.
57. Fehlhammer, H., and Bode, W. (1975) *J. Mol. Biol. 98*, 683−692.
58. Bode, W., and Schwager, P. (1975) *J. Mol. Biol. 98*, 693−717.
59. Norel, R., Sheinerman, F., Petrey, D., and Honig, B. (2001) *Protein Sci. 10*, 2147−2161.
60. Darden, T. A., and York, D. (1993) *J. Chem. Phys. 98*, 10089−10094.
61. Essman, U., Perera, L., Berkowitz, M. L., Darden, T. A., Lee, H., and Pedersen, L. G. (1995) *J. Chem. Phys. 103*, 8577−8593.
62. Weerasinghe, S., Smith, P. E., Mohan, V., Cheng, Y.-K., and Pettitt, B. M. (1995) *J. Am. Chem. Soc. 117*, 2147−2158.
63. Duan, Y., and Kollman, P. A. (1999) *Science 282*, 740−744.
64. Berendsen, H. J. C., Postma, J. P. M., Van Gusteren, W. F., DiNola, A., and Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684−3690.
65. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.
66. Bose, K., and Clark, A. C. (2001) *Biochemistry 40*, 14236−14242.
67. Earnshaw, W. C., Martins, L. M., and Kaufmann, S. H. (1999) *Annu. Rev. Biochem. 68*, 383−424.
68. 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., and Raybuck, S. A. (1994) *Nature 370*, 270−275.

BI034032L