Use of Papain as a Model for the Structure-Based Design of Cathepsin K Inhibitors: Crystal Structures of Two Papain−Inhibitor Complexes Demonstrate Binding to S′-Subsites

Judith M. LaLonde,*,† Baoguang Zhao,† Ward W. Smith,† Cheryl A. Janson,‡ Renee L. DesJarlais,§ Thaddeus A. Tomaszek,|| Thomas J. Carr,⊥ Scott K. Thompson,⊥ Hye-Ja Oh,⊥ Dennis S. Yamashita,⊥ Daniel F. Veber,⊥ and Sherin S. Abdel-Meguid*,†

Departments of Structural Biology, Protein Biochemistry, Physical and Structural Chemistry, Molecular Recognition, and Medicinal Chemistry, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406

Received April 23, 1998

Papain has been used as a surrogate enzyme in a drug design effort to obtain potent and selective inhibitors of cathepsin K, a new member of the papain superfamily of cysteine proteases that is selectively and highly expressed in osteoclasts and is implicated in bone resorption. Here we report the crystal structures of two papain−inhibitor complexes and the rational design of novel cathepsin K inhibitors. Unlike previously known crystal structures of papain−inhibitor complexes, our papain structures show ligand binding extending deep within the S′-subsites. The two inhibitor complexes, carbobenzyloxyleucinyl-leucinyl-leucinal and carbobenzyloxy-L-leucinyl-L-leucinyl methoxymethyl ketone, were refined to 2.2- and 2.5-Å resolution with *R*-factors of 0.190 and 0.217, respectively. The S′-subsite interactions with the inhibitors are dominated by an aromatic−aromatic stacking and an oxygen−aromatic ring edge interaction. The knowledge of S′-subsite interactions led to a design strategy for an inhibitor spanning both subsites and yielded a novel, symmetric inhibitor selective for cathepsin K. Simultaneous exploitation of both S- and S′-sites provides a general strategy for the design of cysteine protease inhibitors having high specificity to their target enzymes.

Introduction

Recent successes in the rational design of potent, selective HIV-1 protease inhibitors and the subsequent demonstration that they are highly effective drugs in the treatment of AIDS have validated the important role of rational design in the drug discovery process. Knowledge of the three-dimensional structure of HIV protease, particularly with bound ligands, was critical for the rapid progress in the design of HIV-1 protease inhibitors. Equally important was the vast knowledge gained earlier from other aspartyl proteases such as renin and pepsin. Here, we will describe the use of papain as a model for the structure-based design of cathepsin K inhibitors that are potential drugs for the treatment of osteoporosis. The use of papain as a drug design target proved instrumental in obtaining inhibitors for cathepsin K at a time when large quantities and structural information of cathepsin K were unavailable.

Papain is a plant cysteine protease isolated from papaya latex. It is a 212-amino acid polypeptide with well-known biological and structural properties.[1] Its catalytic mechanism involves an active site triad, consisting of Cys25, His159, and Asn175, that resides in a cleft between two domains. The reactive thiol of Cys25 hydrolyzes peptide bonds by formation of a covalent acyl enzyme intermediate with the carbonyl carbon of the scissile bond.[2]

Papain shares sequence and structural homology with other mammalian cysteine proteases. Several of these cysteine proteases are implicated in diseases caused by abnormal levels of proteolytic activity. For example, cathepsin K is selectively expressed in human osteoclasts[3] and is a target for treatment of diseases involving excessive bone loss such as osteoporosis. Cathepsin B is involved in intracellular protein degradation and is implicated in muscular dystrophy[4] and metastasis.[5] The high degree of sequence homology between papain and cathepsin K (46% sequence identity) suggests that papain can serve as a suitable model for designing disease-specific cysteine protease inhibitors. Moreover, in the presence of structure−activity relationship (SAR) data on cathepsin K, papain can serve as a model for elucidating components important in selectivity.

Thus far, a number of inhibitors for papain or papain family members have been identified and the three-dimensional structures of their complexes with papain have been elucidated. These include the peptide aldehyde[6] (leupeptin), the chloromethyl ketone,[7] and the epoxide[8,9] inhibitors of papain. In these structures, the inhibitors were found to bind to Cys25 through a reactive group, with the inhibitors interacting exclusively with the "nonprime" side of the active site (S-subsites). Two other papain−inhibitor structures, succinyl-*p*-nitroanilide[10] and N^{α}-*p*-tolylsulfonyl-L-lysine chloromethyl ketone,[11] show limited penetration into the S′-subsite (extending only into the S1′-subsite). Also, a crystal structure of cathepsin B with an irreversible epoxide inhibitor bound was recently published.[12] The inhibitor binds within the S′-side of the active site, but

* To whom reprint requests should be addressed.
† Department of Structural Biology.
‡ Department of Protein Biochemistry.
§ Department of Physical and Structural Chemistry.
|| Department of Molecular Recognition.
⊥ Department of Medicinal Chemistry.

10.1021/jm980249f CCC: $15.00
Published on Web 10/21/1998

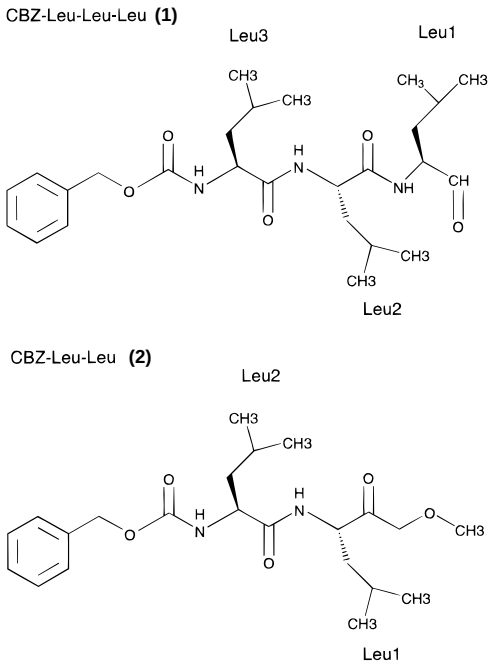

Figure 1. Depiction of the aldehyde **1** and methoxymethyl ketone **2** inhibitors.

cathepsin B has a significant insertion on the S′-side compared to the other papain family members including cathepsin K. Therefore, the structure was of little use in designing cathepsin K inhibitors. The previously known inhibitor−papain complexes have been used to extrapolate the interactions of the S-portion of the peptide substrate complex. However, they only reveal limited information on inhibitor specificity for the S′-subsite. Here, we report the structures of two papain−inhibitor complexes where the inhibitors bind extensively into the S′-subsites. The inhibitors carbobenzyloxy-leucinyl-leucinyl-leucinal (abbreviated aldehyde **1**) and carbobenzyloxy-L-leucinyl-L-leucinyl methoxymethyl ketone (abbreviated methoxymethyl ketone **2**), depicted in Figure 1, serve as a starting point for investigating both S- and S′-subsite selectivity for cysteine proteases.

Results and Discussion

Overall Structure of the Papain−Inhibitor Complexes. The structures of papain in the two papain−inhibitor complexes are virtually identical to the structure of native papain. The rms differences between native papain and the aldehyde **1** and methoxymethyl ketone **2** complexes for main-chain atoms and side-chain atoms are 0.14 and 2.11 Å and 0.30 and 1.57 Å, respectively. Most of the differences observed in side-chain positions occur on the surface of the protein and involve atoms with high thermal motion. For the methoxymethyl ketone **2** complex, differences with the native structure occur at Gly180 and Trp181, where the peptide bond is rotated to allow the carbonyl group to interact with the inhibitor Cbz ring (see below).

The structural differences between the two inhibitor complexes are also small. The rms differences between protein atoms in the aldehyde **1** and methoxymethyl ketone **2** complexes with papain are 0.28 and 1.45 Å for main-chain and side-chain atoms, respectively. Some differences in main-chain positions occur as a result of inhibitor binding to the S′-subsites. Aldehyde **1** is a larger inhibitor and makes more contacts with residues comprising the active site cavity. Small differences in peptide bond conformation occur in order to avoid bad contacts or form closer contacts with the inhibitor. For example, in the aldehyde **1** complex, the carbonyl of Ser21 is rotated away relative to its position in methoxymethyl ketone **2** to avoid a bad contact with the second leucyl group of the inhibitor (see Figures 5 and 6). Main-chain differences at Gly65 and Asp158 occur in the aldehyde **1** complex to enhance van der Waals contacts with the inhibitor. These are minor changes and do not affect the overall structure of the enzyme. Most of the differences in side-chain positions for the two enzyme−inhibitor complexes occur in regions with high thermal motion and are not a consequence of inhibitor binding.

Furthermore, the papain structure in the aldehyde **1** and methoxymethyl ketone **2** complexes is nearly identical to its structure for the leupeptin−papain complex,[6] a S-subsite inhibitor−papain complex. The rms differences between protein atoms in the aldehyde **1** and methoxymethyl ketone **2** complexes with the leupeptin-papain complex for main-chain and side-chain atoms are 0.36 and 1.46 Å and 0.41 and 1.41 Å, respectively. Only small differences occur in the conformation of main-chain and side-chain atoms that form the S- and S′-subsites between papain, papain−S-subsite inhibitor complex, and papain−S′-subsite inhibitor complex. Therefore, no change in protein conformation is observed upon inhibitor binding to either the S- or S′-side of the active site.

Papain−Inhibitor Interactions. The electron densities of the bound inhibitors are shown in Figure 2. In both cases the density of the inhibitors is easily distinguished from side-chain density in the binding cleft. Methoxymethyl ketone **2** binds noncovalently to papain ($k_{obs}/I = 26\ M^{-1}\ s^{-1}$), while it appears that aldehyde **1** binds covalently ($k_{obs}/I = 430\ 000\ M^{-1}\ s^{-1}$).

The substrate binding pocket of papain lies in a cleft between two domains. Domain 1 consists of residues 1−11 and 113−207, while domain 2 consists of residues 12−112 and 208−212. The binding pocket of papain has been defined by Berger and Schechter[13] in terms of seven subsites designated S4 to S1 and S1′ to S3′. Each subsite accommodates one amino acid of the peptide substrate. The four S-subsites (S4 to S1) are located on the acyl side of the scissile bond of the bound substrate, while the three S′-subsites (S1′ to S3′) are located on the amino side. The primary specificity of papain is for binding hydrophobic residues such as Phe in the S2-subsite.[1] Both inhibitors bind in the groove between papain's two domains, along the S′-subsites (Figure 3). This is in contrast to binding along the S-subsites by other papain inhibitors for which crystal structures are known, including the chloromethyl ketone,[7] epoxysuccinyl,[8,9] and leupeptin[6] peptide inhibitors. The peptide groups of aldehyde **1** and methoxymethyl ketone **2** bind in the opposite direction to that postulated for a peptide substrate.[7] In addition, the binding of the inhibitors in the cleft is such that one edge of the benzene ring of the Cbz group is entirely buried and the other edge is exposed to solvent.

Aldehyde 1 Binding Interactions. The amino acid residues that form contacts with the inhibitor are shown

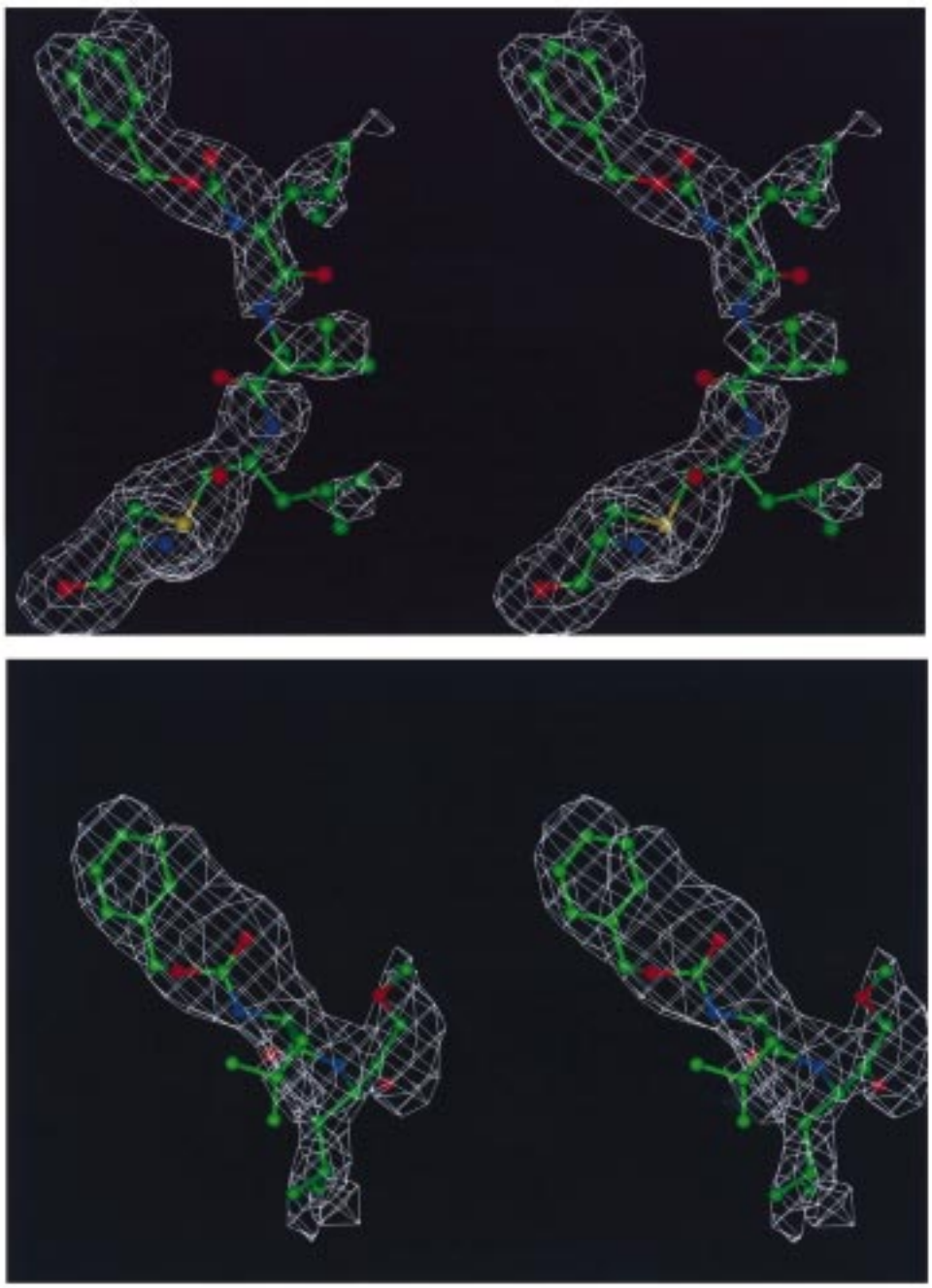

Figure 2. Omit electron density maps around the aldehyde **1** (top) and methoxymethyl ketone **2** (bottom) inhibitors. The maps were calculated with coefficients $|F_o - F_c|$ omitting the inhibitor positions from phasing.

in Figure 4. On the basis of the reactivity of the aldehyde group toward thiols and the observation of the "omit" electron density maps, we presume that the aldehyde **1** is covalently bound to Cys25 as a hemimercaptal. This covalent bond occurs at the analogous position to the P1 carbonyl carbon in a peptide substrate. The inhibitor contacts (less than 4.0 Å) 11 protein residues: Gln19, Gly20, Ser21, Gly23, Asn64, Gly65, Gln142, Asp158, His159, Trp177, and Trp181 and 2 water molecules (Wat412 and Wat480). These interactions include hydrophobic, electrostatic, and hydrogen bonding, but most interesting are the stacking interactions between the aromatic ring of the Cbz group and Trp177 and Trp181. Figure 5 shows interactions between the inhibitor and the protein in the active site.

Aromatic stacking interactions in proteins have been defined by Burley and Petsko as pairs of aromatic residues with phenyl ring centroid separations between 4.5 and 7.0 Å and dihedral angles between aromatic planes within 30−90°.[14] These separations and angles between the inhibitor's phenyl ring and Trp177 and Trp181 are 5.7 Å and 33° and 7.0 Å and 50°, respectively. On the basis of the Burley and Petsko definition,[14] the tryptophan−Cbz stacking interactions can be characterized as weak. The stacking interaction with Trp181 appears stronger (closer to perpendicular) than that for Trp177. The position of the phenyl ring is further stabilized by a favorable electrostatic interaction with the carbonyl oxygen of Gln142 (3.3 Å).[15] The carbonyl oxygen faces the phenyl ring so that the partial negative charge on the carbonyl oxygen is directed toward the partial positive charge at the edge of the

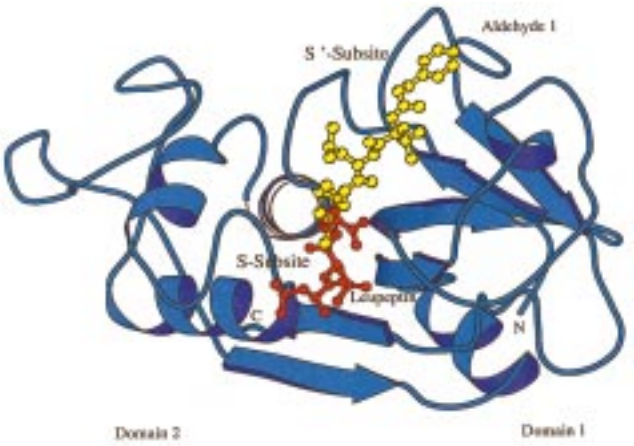

Figure 3. Folding diagram of papain showing the inhibitor bound in the cleft between two domains. The ball-and-stick model of aldehyde **1** bound in the S′-subsite is shown in yellow. For comparison, leupeptin bound in the S-subsite is drawn in red. Figure was prepared using MOLSCRIPT.[23]

phenyl ring. A second favorable oxygen−aromatic ring edge interaction is formed between Wat480 and the Cbz ring (3.5 Å). Hydrophobic contacts are also observed between the CH_2 group of the Cbz and Trp177.

There are four carbonyl oxygens in the inhibitor: the first is the hemimercaptal oxygen, the second and third are peptidyl amide oxygens, and the fourth is a carbamate carbonyl oxygen. Only the first peptidyl amide oxygen is hydrogen-bonded to the protein (Figures 5 and 6); it interacts with the imidazole nitrogen of His159 and Wat412. The hemimercaptal oxygen forms two hydrogen bonds: one to the amide group of Gln19 (a hydrogen bond between the amide nitrogen of Gln19 and the carbonyl oxygen of the substrate is postulated to stabilize the tetrahedral intermediate *by forming the so-called oxyanion hole*[10]) and a second to the backbone amide nitrogen from Cys25.

The three Leu side chains (Leu1, Leu2, Leu3) of the inhibitor are partially buried forming a number of van der Waals contacts in their respective subsites. Leu1 forms van der Waals contacts with Asn64, Gly65, and Asp158. This side chain is bent, allowing it to fall into the S1-subsite rather than extend into the S1′-subsite. Thus, it mimics a P1 rather than a P1′ residue. Leu2 is actually the P1′ side chain; it makes van der Waals contacts with Gly20, Ser21, and Cys22 in the S1′-subsite. Leu3 forms van der Waals contacts with the amide group of Gln142 in the S2′-subsite.

Methoxymethyl Ketone 2 Binding Interactions. Methoxymethyl ketone **2** makes contacts with seven residues (Gln19, Gly23, Gly65, Gln142, His159, Trp177, Trp181) as well as with a water molecule (Wat511) along the S′-side, Figure 4. As observed with aldehyde **1**, the aromatic ring of the Cbz in this case is also stacked with Trp177 and Trp181, but the interactions are weaker. The centroid distances between phenyl groups and dihedral angles are 8.0 Å and 24° and 7.9 Å and 59° for Trp177 and Trp181, respectively. The distances are greater than those observed with aldehyde **1**, but the angles are approximately the same. Again better stacking (closer to perpendicular) is observed with the Cbz to Trp181 than to Trp177. The position of the aromatic ring is stabilized by an electrostatic interaction with the main-chain carbonyl oxygen of Gly180 (3.0 Å).

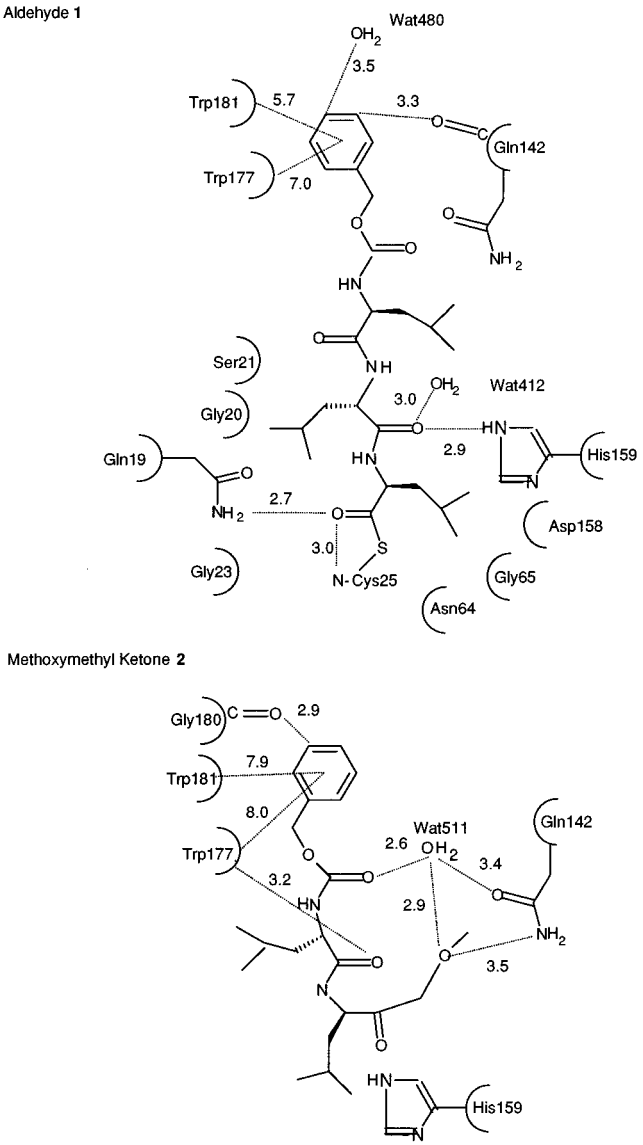

Figure 4. Diagram of residues that form contacts with the inhibitors aldhehyde **1** (top) and ketone **2** (bottom). Distances (in Å) are shown as dashed lines.

There are two peptidyl amide oxygens and one carbamate carbonyl oxygen in the inhibitor. The second peptidyl amide oxygen forms a hydrogen bond to the aromatic ring of Trp177. The carbamate carbonyl oxygen hydrogen bonds to Wat511. The carboxymethyl oxygen is also involved in hydrogen bonding. It forms a weak hydrogen bond to the amido nitrogen on Gln142 and a second hydrogen bond to Wat511. In addition, Wat511 forms a hydrogen bond to the amido group of Gln142. Therefore, Wat511 stabilizes the bent conformation of the inhibitor by hydrogen bonding to the carbamate carbonyl oxygen near the Cbz group, the carboxymethyl oxygen, and a protein side chain.

Methoxymethyl ketone **2** binds noncovalently. As a result it does not bind in the S1-subsite, and the first Leu side chain, Leu1, does not make the same set of contacts as P1′ Leu2 of aldehyde **1** (Figure 6); instead it forms a hydrophobic contact with His159. The inhibitor is forced into a bent conformation by contacts formed with Gln142 of the S2′-subsite as noted above. Leu2 forms hydrophobic contacts with Trp177 in the

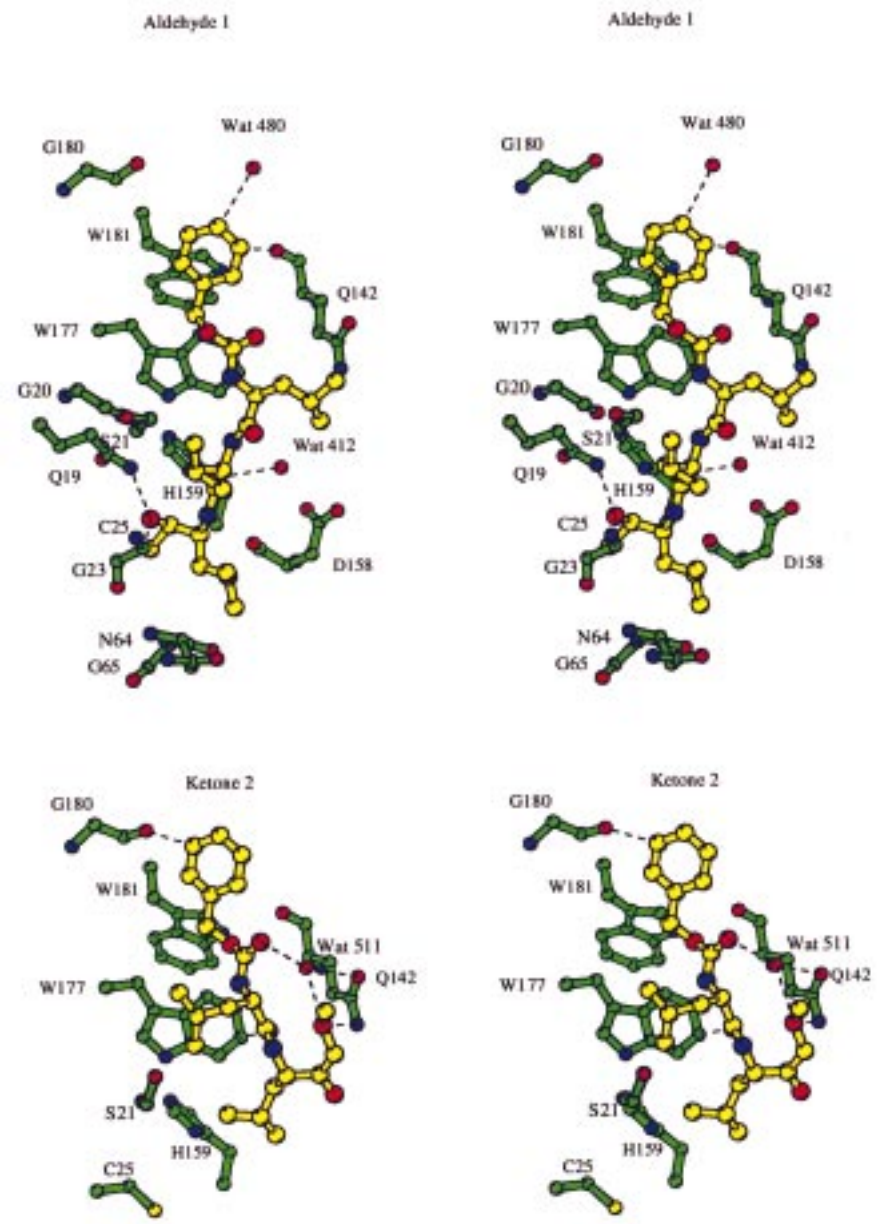

Figure 5. Stereodiagram of aldehyde **1** (top) and methoxymethyl ketone **2** (bottom) bound in the S′-subsites of papain. Residues forming interactions with the inhibitor are drawn in green. The inhibitors are drawn in yellow. Hydrogen bonds and oxygen−aromatic ring edge interactions are shown with dashed lines. Figure was prepared using MOLSCRIPT.[23]

S3′-subsite. The Cbz moiety forms contacts with Trp177 and Trp181 in the S3′-subsite as does aldehyde **1**.

Comparison of Binding Modes for the Two Inhibitors. Only minor differences in protein side-chain conformations are observed between the two inhibitor complexes, and no major difference occurs in the S′-subsite upon inhibitor binding, Figure 6. As described above the conformations of both inhibitors are stabilized by aromatic−aromatic interactions, specific hydrogen bonds to peptidyl amide oxygens, and hydrophobic contacts from leucine side chains. Antiparallel β-sheet type hydrogen bonding between the inhibitor and the enzyme is not observed with these two inhibitors. Moreover, each inhibitor exhibits a different hydrogen-bonding pattern of the peptidyl amide oxygens. A hydrogen bond is formed between the His159 indole ring and the first peptidyl amide oxygen for aldehyde **1** and between the Trp177 aromatic ring and the second peptidyl amide oxygen for ketone **2**. There is the common use of water molecules for inhibitor binding in the cleft. Both inhibitors use a water molecule for hydrogen bonding, but to a different carbonyl of the inhibitor. These are the first peptidyl oxygen for aldehyde **1** and the carbamate carbonyl oxygen adjacent to the Cbz group in ketone **2**.

Water molecules are observed in the active site of most of the other reported papain−inhibitor complexes as well as in native papain. Water 511 observed in the ketone **2** complex is not observed in other papain structures including the 1.6-Å structure of papain.[16] Both of the waters associated with aldehyde **1** are observed in other papain−inhibitor complexes. Therefore, no clear pattern of hydration can be associated with inhibitor binding in the S′-subsite.

In comparing the position of the two inhibitors, the benzyl ring of methoxymethyl ketone **2** is translated 2.6

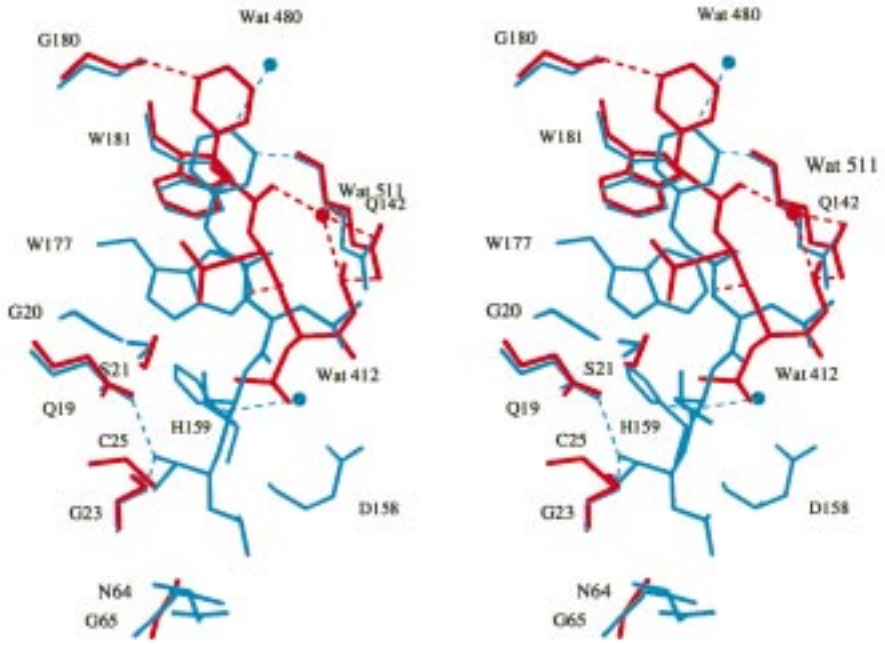

Figure 6. Superposition of aldehyde **1** (blue) and methoxymethyl ketone **2** (red) in the active site. Hydrogen bonds and oxygen−aromatic ring edge interactions are shown with blue and red dashed lines for aldehyde **1** and methoxymethyl ketone **2**, respectively. Figure was prepared using MOLSCRIPT.[23]

Å farther away from the catalytic cysteine than the benzyl ring of aldehyde **1** (Figure 6). Despite this translation, both Cbz moieties are stacked with Trp177 and Trp181. Cbz groups of both inhibitors form an oxygen−aromatic ring edge interaction. However, two different main-chain carbonyls are used, one from Gly180 for methoxymethyl ketone **2** and the other from Gln142 for aldehyde **1**. The carbonyl oxygens from Gly180 and Gln142 approach the inhibitors from opposite sides of their aromatic rings (Figure 6). The differences in the position and interactions of aldehyde **1** and ketone **2** despite the similarity in the active site regions indicate that the binding mode is determined by chemical functionality of the inhibitor. The fact that aldehyde **1** binds covalently whereas ketone **2** binds noncovalently may be the sole reason that the two inhibitors have different positions in the S′-subsite.

Subsite Specificity. Binding of aldehyde **1** and methoxymethyl ketone **2** inhibitors deep within the S′-subsites of papain is a novel observation. The presence of a Cbz moiety and its aromatic interactions with tryptophans and carbonyl oxygens are important components of binding specificity. However, the presence of the Cbz group alone is not sufficient to explain the specificity for the S′-subsite. Two other papain−inhibitor structures also contain the Cbz group. These are the chloromethyl ketone inhibitors Cbz-Phe-Ala-CK and Cbz-Gly-Phe-Gly-CK.[7] These chloromethyl ketone inhibitors bind covalently, along the S-subsites with their Cbz groups forming interactions with Tyr61 and Tyr67. The Phe side chain of Cbz-Phe-Ala-CK and Cbz-Gly-Phe-Gly-CK sits in the S2 pocket where primary substrate specificity is determined.

The S-subsite inhibitor Cbz-Gly-Phe-Gly-CK is similar to aldehyde **1** in that both bind covalently, contain three peptide residues, and are approximately the same length (15.5 Å). However, aldehyde **1** is bound to Cys25 as a hemimercaptal with the covalent bond between the S atom of Cys25 and the first carbonyl group of the inhibitor, whereas the inhibitor Cbz-Gly-Phe-Gly-CK contains an extra methylene group between the first carbonyl of the inhibitor and the S atom of Cys25. Although crystallized under different conditions, differences in the crystal do not account for the subsite preference, since none of the inhibitors or their active site residues are involved in crystal contacts. Furthermore, there are no major differences in side-chain conformation in the residues comprising the S- and S′-subsites between the chloromethyl ketone inhibitor−papain complexes and the aldehyde **1**−papain and ketone **2**−papain complexes. Factors that may explain subsite preference would include the presence of a CH_2 spacer in Cbz-Gly-Phe-Gly-CK and the difference in side-chain composition. Examination of the Cbz-Gly-Phe-Gly-CK−papain complex shows that the extra CH_2 group forms interactions with H159. In aldehyde **1**, H159 forms a hydrogen bond to the first peptidyl amide oxygen instead, Figure 5. In Cbz-Gly-Phe-Gly-CK the CH_2 group prevents H159 from interacting with the first peptidyl amide oxygen. Instead, its first peptidyl amide oxygen hydrogen bonds to Gly66 in the S-subsite. In regards to side-chain composition, a model of aldehyde **1** bound in the S-subsite shows that the larger Leu side chain restricts the position of the Cbz group. Thus, aldehyde **1** bound in the S-subsite would have diminished interactions with Tyr61 and Tyr67 in comparison to Cbz-Gly-Phe-Gly-CK. Although aldehyde **1** appears to fit in the S-subsite, Cbz-Gly-Phe-Gly-CK is not accommodated as easily in the S′-subsite because of the large Phe residue. Therefore, in the absence of a Phe side chain that would bind to the P2 pocket, a Cbz-containing inhibitor may bind to either subsite. The explanation for S′-subsite preference may be a combination of the factors discussed above as well as the aromatic interactions formed between the Cbz and tryptophan side chains and the main-chain carbonyls.

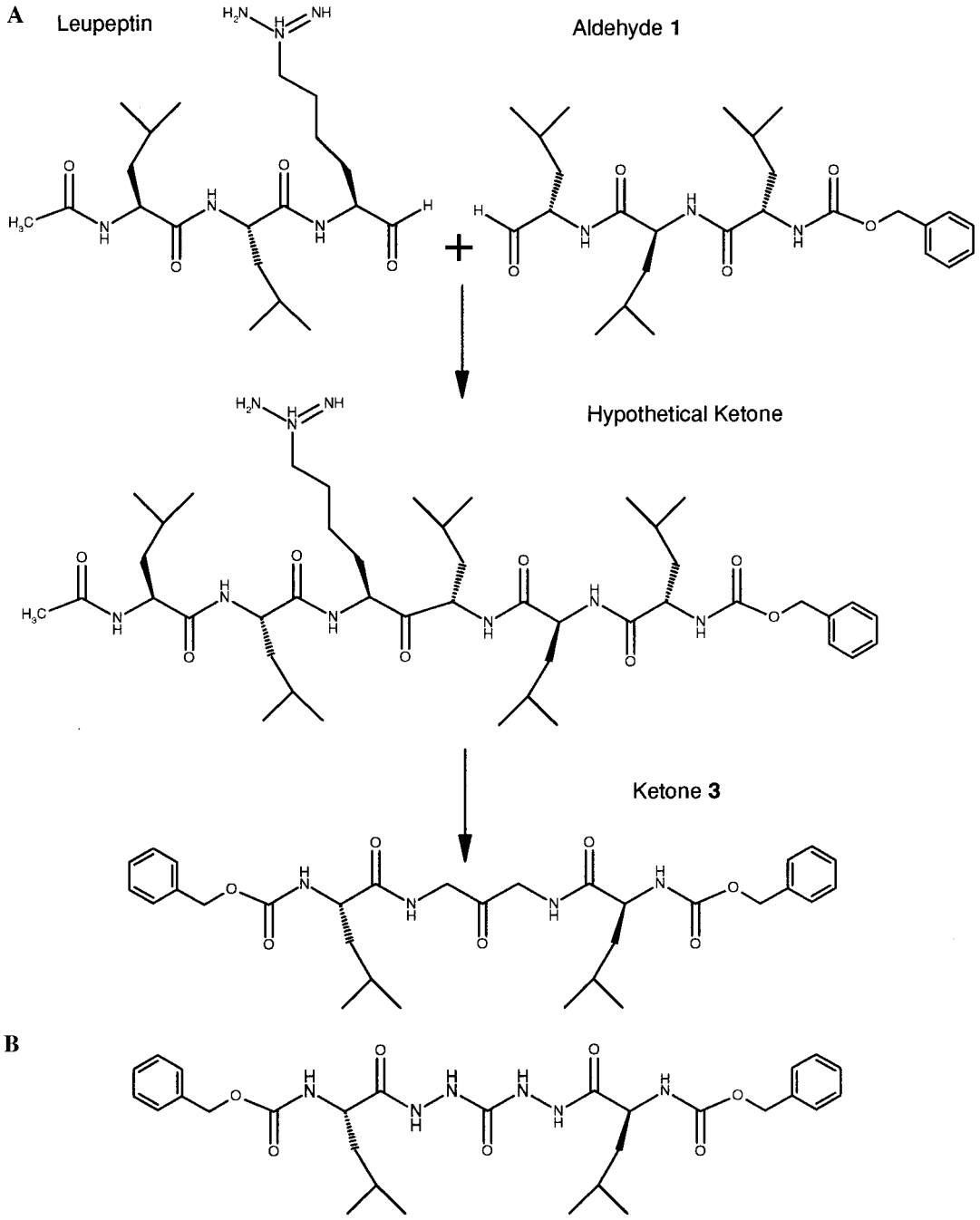

Figure 7. (A) Design of the diacylaminomethyl ketone from leupeptin and aldehyde **1**. (B) Compound **3**, diacylaminomethyl ketone, and compound **4**, diacyl carbohydrazide.

Design of Cathepsin K Inhibitors. Prior to the structure determination of cathepsin K,[17,18] papain was chosen as a surrogate for the structure-based design of cathepsin K inhibitors. Papain serves as a suitable model since it has 46% amino acid sequence identity to cathepsin K. Papain was also selected because it is commercially available in large quantities and its structure is known. To ensure that the novel binding mode observed in the crystal structure of the aldehyde **1**-papain complex was not an artifact of crystallization, crystals of the leupeptin−papain complex were produced under identical conditions. The structure of the leupeptin−papain complex (data not shown) was nearly identical to its previously reported structure.[6] Furthermore, it was isomorphous with the aldehyde **1**-papain complex except that leupeptin was bound on the S-side of the active site as observed in the earlier reported structure.[6]

By overlaying the structures containing leupeptin in the S-subsite and aldehyde **1** in the S′-subsite, a hypothetical 1,3-bis(acylaminomethyl) ketone was designed that spanned both subsites. This resulting inhibitor was simplified by (1) removing the two side chains on either side of the ketone, (2) shortening the structure by one amino acid on both sides, and (3) targeting a symmetrical structure, ketone **3**, by replacing the acetyl group found in leupeptin with a Cbz group (Figure 7A).[19] Modeling suggested that even though the aromatic−aromatic interaction with Trp188 (Trp181 in papain) and the Cbz group would be eliminated by

shortening the inhibitor structure, it appeared that a better interaction with Trp184 (Trp177 in papain) could compensate. Furthermore, the high potency of the dipeptide aldehyde, Cbz-Leu-Leu-H, against cathepsin K supported this hypothesis.[19]

Ketone **3** was found to be a potent inhibitor of cathepsin K. A subsequent crystallographic structure determination of ketone **3** bound to cathepsin K[19] revealed that (1) the covalent inhibitor spans both the S- and S′-subsites, (2) one of the leucine side chains fills the hydrophobic S2 pocket, (3) aromatic−aromatic interactions are formed between the two Cbz groups and Trp184 (face-to-face, centroid distance and angle of 5.5 Å and 23°) and Tyr67 (face-to-edge, centroid distance and angle of 5.5 Å and 83°), and (4) aromatic ring edge interactions are observed between the Cbz groups and the amide carbonyl oxygens of Asn18 (3.02 Å) and weakly with Leu160 (3.12 Å).

Interestingly, the potency of inhibition by ketone **3** is much weaker against papain ($K_{i,app} > 10\ \mu M$) than against cathepsin K ($K_{i,app}$ of 22 nM).[19] One difference in the structures of cathepsin K and papain is near the S-subsite where cathepsin K contains a 2-amino acid residue insertion in the loop from residues 151 to 161. Nonetheless, overlays of the papain and cathepsin K−ketone **3** crystal structures show no obvious steric hindrance between ketone **3** and papain side chains. The difference in potency of ketone **3** in cathepsin K versus papain can also be rationalized from the two observed binding modes of Cbz groups on the S-side of cathepsin K. In the ketone **3**−cathepsin K crystal structure, the Cbz group on the S-side is oriented between Tyr67 and Leu160; however, in the crystal structure of cathepsin K and $(Cbz\text{-}LeuNHNH)_2CO$ (**4**), an isosteric compound of the diacyl carbohydrazide class of inhibitors[20] (Figure 7B), the Cbz binds between Tyr 67 and Asp 61, the S3-subsite. Modeling indicates that ketone **3** is not able to bind to papain as in the $(Cbz\text{-}LeuNHNH)_2CO$ (**4**)−cathepsin K complex,[20] because a steric clash would occur between the Cbz group and Tyr61 which is replaced by the smaller Asp61 in cathepsin K. In the cocrystal of cathepsin K−ketone **3**, this S3-subsite is occupied by Lys44 from an adjacent molecule in the crystal lattice. The two binding modes are observed in different crystal forms for the two complexes. Therefore, understanding the large difference in inhibitor potency of ketone **3** in papain versus cathepsin K may be aided by additional structural studies of ketone **3** bound to papain.

Conclusions

Binding of the aldehyde (**1**) and methoxymethyl ketone (**2**) inhibitors deep within the S′-subsite is unique for papain, although it has been reported for inhibitor binding to cathepsin B. The directionality of binding appears to be influenced by aromatic−aromatic stacking and oxygen−aromatic ring edge interactions. This observation has inspired a drug design strategy using papain as a surrogate for cathepsin K. With the knowledge of S′-subsite interactions in papain, numerous potent, selective cathepsin K inhibitors were designed that span both sides of the active site.[19,20] Ultimately, the papain structures combined with information from structure−activity studies on peptide aldehyde inhibitors and molecular modeling provided pivotal insights for the design of novel, selective inhibitors of cathepsin K. The use of papain as a surrogate for cathepsin K illustrates the utility and limitations of employing surrogate enzymes in the design of protease inhibitors.

Experimental Section

Inhibitors. Aldehyde **1** was prepared as previoiusly reported:[24] ^{1}H NMR ($CDCl_3$) δ 9.43 (s, 1H), 7.28 (m, 5H), 6.68 (d, 1H), 6.47 (d, 1H), 5.12 (d, 1H), 5.02 (s, 2H), 4.39 (m, 2H), 4.08 (m, 1H), 1.70−1.32 (m, 9H), 0.81 (m, 18H). Anal. ($C_{26}H_{41}N_3O_5 \cdot 0.5H_2O$) C, H, N.

Ketone **2** was synthesized by following the method of Jones et al.[25,26] and is described below:

Cbz-Leu-Leu α-Bromomethyl Ketone. Isobutyl chloroformate (1.37 mL, 10.6 mmol) was added to a solution of Cbz-Leu-Leu-OH (4.0 g, 10.6 mmol; Bachem) and *N*-methylmorpholine (1.16 mL, 10.6 mmol) in THF (20 mL) at −40 °C, and the reaction was stirred for 15 min. The reaction mixture was then filtered; then a solution of diazomethane (from 5.9 g of 1-methyl-3-nitro-1-nitrosoguanidine and 18 mL of 40% aqueous KOH in 150 mL of diethyl ether) was added slowly. The reaction was sealed and was maintained at 0 °C in a refrigerator overnight. The reaction mixture was then treated with 30% HBr in AcOH (7 mL) and was stirred for 5 min. The reaction mixture was diluted with EtOAc (50 mL) and then was extracted with 15% aqueous citric acid, then saturated sodium bicarbonate (with CO_2 evolution), and then brine. The combined organics were dried with anhydrous sodium sulfate, filtered, and concentrated in vacuo to give the title compound as a white solid (3.2 g, 67%): ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.4−7.3 (m, 5H), 6.74 (d, J = 7.6 Hz, 1H), 5.34 (d, J = 7.6 Hz, 1H), 5.10 (s, 2H), 4.80−4.75 (m, 1H), 4.22 (brs, 1H), 4.05 (AB, J_{AB} = 16.0 Hz, $\Delta\delta_{AB}$ = 0.045, 2H), 1.7−1.6 (m, 4H), 1.55−1.47 (m, 2H), 0.95 (d, J = 6.8 Hz, 3H), 0.91 (3d, J = 6.4 Hz, 9H); IR (thin film) 1716, 1700, 1670 cm^{-1}. Anal. ($C_{21}H_{31}BrN_2O_4$) C, H, N, Br.

Cbz-Leu-Leu α-Hydroxymethyl Ketone. Benzoylformic acid (1.1 g, 7.1 mmol) was added to a solution of Cbz-Leu-Leu α-bromomethyl ketone (2.7 g, 5.9 mmol) in DMF (20 mL) at room temperature. Then solid potassium fluoride (0.52 g, 8.9 mmol) was added, and the reaction mixture was stirred overnight. The reaction mixture was then diluted with EtOAc (100 mL), and was poured into water (50 mL). The combined organics were extracted with brine, dried with magnesium sulfate, filtered, concentrated in vacuo, and then dissolved in THF (200 mL). A solution of potassium bicarbonate (200 mL, 200 mmol, 1 M) was added, and the reaction mixture was stirred vigorously overnight. The reaction mixture was diluted with EtOAc (200 mL) and then was extracted with water and then brine. The combined organics were dried with anhydrous magnesium sulfate, filtered, and concentrated in vacuo to give the title compound as a white solid (3.2 g, 100%): ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.35−7.31 (m, 5H), 6.74 (d, J = 7.6 Hz, 1H), 5.35 (d, J = 7.6 Hz, 1H), 5.10 (s, 2H), 4.7−4.6 (m, 1H), 4.36 (s, 2H), 4.25−4.18 (m, 1H), 3.2 (brs, 1H), 1.7−1.43 (m, 6H), 0.93 (d, J = 6.0 Hz, 3H), 0.91 (2d, J = 6.0 Hz, 6H), 0.90 (d, J = 6.0 Hz, 3H); IR (thin film) 3307, 1716, 1699, 1652 cm^{-1}.

Cbz-Leu-Leu α-Methoxymethyl Ketone. Methyl iodide (0.32 mL, 5.1 mmol) was added to a solution of Cbz-Leu-Leu α-hydroxymethyl ketone (0.4 g, 1.0 mmol) and silver(I) oxide (0.46 g, 2.0 mmol) in methylene chloride (30 mL), and the reaction mixture was refluxed for 7 h. The reaction was incomplete; therefore, additional methyl iodide (1.0 mL, 15.5 mmol) was added, and the reaction mixture was stirred at room temperature over 2 days. The reaction mixture was filtered, concentrated in vacuo, and then chromatographed (silica gel, 25% EtOAc−hexanes) to give the title compound as a white solid (0.17 g, 42%): ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.35−7.24 (m, 5H), 6.35−6.33 (d, 1H, J = 7.74 Hz), 5.06 (s, 2H), 4.81−4.75 (m, 1H), 4.16 (brs, 1H), 4.12 (AB, J_{AB} = 10.3 Hz, $\Delta\delta_{AB}$ = 0.097, 2H), 3.40 (s, 3H), 1.60−1.30 (m, 6H), 0.92

Table 1. Data Collection Statistics

	aldehyde **1**−papain	ketone **2**−papain
cell (Å)	$a = 100.54$	$a = 101.00$
	$b = 50.74$	$b = 50.99$
	$c = 62.28$	$c = 62.48$
no. of crystals	6	2
resolution limit	2.2	2.5 Å
no. of observations	62560	41190
no. of reflections	14458	11412
$R_{sym}(I)$	8.78	13.7
no. refls, $I/\sigma I > 2$	12955	9152
redundancy	4.3	3.6
% complete	95	91

Table 2. Final Refinement Statistics

	aldehyde **1**−papain	ketone **2**−papain
resolution Å	10−2.2	8−2.5
initial *R*-free/*R*-factor	0.284/0.291	0.316/0.311
final *R*-free/*R*-factor	0.240/0.190	0.306/0.217
no. of macrocycles	10	9
no. of reflections	12800	8679
no. of atoms	1772	1727
no. of solvent	88	44
average overall *B*-factor (Å^2)	22.7	26.4
inhibitor *B*-factor (Å^2)	43.3	48.7
inhibitor occupancy	0.65	0.85
rms deviations		
bonds (Å)	0.010	0.011
angles (deg)	1.529	1.595
dihedrals (deg)	24.743	24.859
impropers (deg)	1.387	1.441

Table 3. Disordered Side Chains

aldehyde **1**	methoxymethyl ketone **2**
R59, N77, R98, R111, K139, K145, K156	R59, Q73, N77, R98, K145

(d, 9H, $J = 6.2$ Hz), 0.90 (d, 3H, $J = 6.47$ Hz); IR (KBr) 3301, 1731, 1694, 1646 cm^{-1}. Anal. ($C_{22}H_{34}N_2O_5$) C, H, N.

Crystallization. Papain from papaya latex was purchased from ICN Biomedicals (#100924) as a suspension. Prior to crystallization the suspension buffer was removed by centrifugation. A concentrated solution (approximately 35 mg/mL) of papain was prepared in water. Papain was crystallized by hanging drop vapor diffusion method with 0.1 M Tris HCl, pH 8.5, 0.5 M trisodium citrate, and 20% PEG 600. Crystals of the enzyme−inhibitor complexes were prepared by either cocrystallization (methoxymethyl ketone **2**) or by soaking the crystal (aldehyde **1**). Solutions of inhibitors were prepared in DMSO to give approximately 0.03 M. For cocrystallization of methoxymethyl ketone **2**, a 5-fold molar excess of the inhibitor was added to the enzyme solution and allowed to stand overnight before setting up crystals. The aldehyde **1** complex crystals were formed by soaking a native crystal with a 5-fold molar excess of the inhibitor in the crystallization droplet.

X-ray Methods and Refinement. The native enzyme and inhibitor complexes crystallized in space group *C*2 with one molecule in the asymmetric unit. Data were collected on a Siemens area detector. The cell dimensions and data collection statistics are listed in Table 1.

The structure of the native papain was solved first by molecular replacement using the program XPLOR,[21] using the previously solved structure of papain (code 1PIP)[8] as a starting model. The cross-rotation function was calculated using data from 10 to 4 Å with a radius of integration of 38 Å resulting in the highest peak at 3.8σ. The translation search was computed using data from 10 to 3.5 Å and gave a peak at 32σ. The resulting model yielded an *R*-factor of 0.33. Several cycles of refinement of the native papain were conducted using rigid body, simulated annealing, and positional and temperature factor refinement as implemented in XPLOR.[21] The final native model consisted of 1694 atoms, including 64 water molecules, with an *R*-factor of 0.218 for data from 10 to 2.5 Å and an average temperature factor of 24.2 Å^2. The phases from native papain were used in calculating difference Fourier maps (coefficients $|2F_o - F_c|$ and $|F_o - F_c|$) into whose density the inhibitor atoms were built. The molecular graphics program CHAIN[22] was used for model building and map inspection in subsequent cycles of simulated annealing and positional and temperature factor refinement of the papain−inhibitor complexes. Refinement statistics for the papain−inhibitor complexes are shown in Table 2.

In the final model, several of the side-chain positions have weak electron density and are disordered beyond either the C-β or C-γ atom. These disordered side chains are listed in Table 3 and have their occupancies set to 0 in the coordinate file. The coordinates for the aldehyde **1**−papain and methoxymethyl ketone **2**−papain complexes have been submitted to the PDB and given codes 1BP4 and 1BQI, respectively.

References

(1) Baker, E. N.; Drenth, J. The Thiol Proteases: Structure and Mechanism. In *Biological Macromolecules and Assemblies, Vol. 3. Active Sites of Enzymes;* Jurnak, F. A., McPherson, A., Eds.; John Wiley and Sons: New York, 1987; pp 313−368.

(2) Storer, A. C.; Menard, R. Catalytic Mechanism in Papain Family of Cysteine Peptidases. *Methods Enzymol.* **1994**, *244*, 487−500.

(3) Drake, F. H.; Dodds, R.; James I.; Connor J.; Debouck, C.; Richardson, S.; Lee, E.; Rieman, D.; Barthlow, R.; Hastings, G.; Gowen, M. Cathepsin K, But Not Cathepsins B, L, or S, Is Abundantly Expressed in Human Osteoclasts. *J. Biol. Chem.* **1996**, *271*, 12511−12516.

(4) Katunuma, N.; Kominami, E. Abnormal Expression of Lysosomal Cysteine Proteases in Muscle Wasting Diseases. *Rev. Physiol. Biochem. Pharmacol.* **1987**, *108*, 1−20.

(5) Sloane, B. F. Cathepsin B and Cystatins: Evidence for a Role in Cancer Progression. *Semin. Cancer Biol.* **1990**, *1*, 137−152.

(6) Schroeder, E.; Phillips, C.; Garman, E.; Harlos, K.; Crawford, C. X-ray Crystallographic Structure of a Papain-Leupeptin Complex. *FEBS Lett.* **1993**, *315*, 38−42.

(7) Drenth, J.; Kalk, K. H.; Swen, H. M. Binding of Chloromethyl Ketone Substrate Analogues to Crystalline Papain. *Biochemistry* **1976**, *15*, 3731−3738.

(8) Varughese, K. I.; Ahmed, F. R.; Carey, P. R.; Hasnain, S.; Huber, P.; Storer, A. C. Crystal Structure of a Papain-E−64-Complex. *Biochemistry* **1989**, *28*, 1220−1332.

(9) Yamamoto, D.; Matsumoto, K.; Ohishi, H.; Ishida, T.; Inoue, M.; Kitamura, K.; Mizuno, H. Refined X-ray Structure of Papain E−64-c Complex at 2.1 Å Resolution. *J. Biol. Chem.* **1991**, *266*, 14771−14777.

(10) Yamamoto, A.; Tomoo, K.; Doi, M.; Ohishi, H.; Inoue, M.; Ishida, T.; Yamamoto, D.; Tsuboi, S.; Okamoto, H.; Okada, Y. Crystal Structure of Papain-Succinyl-Gln-Val-Val-Ala-Ala-*p*-Nitroanilide complex at 1.7 Å resolution: Noncovalent Binding Mode of a Common Sequence of Endogenous Thiol Protease Inhibitors. *Biochemistry* **1992**, *31*, 11305−11309.

(11) Drenth, J.; Jansonius, J. N.; Koekoek, R.; Swen, H. M.; Wolthers, B. G. Structure of Papain. *Nature* **1968**, *218*, 929−932.

(12) Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Huber, R.; Turk, V. Crystal Stucture of Cathepsin B Inhibited with CA030 at 2.1 Angstroms Resolution: A Basis for the Design of Specific Epoxysuccinyl Inhibitors *Biochemistry* **1995**, *34*, 4791−4797.

(13) Berger, A.; Schechter, I. Mapping the Active Site of Papain With Aid of Peptide Substrates and Inhibitors. *Philos. Trans. R. Soc. London* **1970**, *B257*, 249−264.

(14) Burley, S. K.; Petsko, G. A. Aromatic−Aromatic Interaction: A Mechanism of Protein Structure Stablization. *Science* **1985**, *229*, 23−28.

(15) Deber, C. M.; Joshua, H. Side Chain Interactions in Aromatic Dipeptides. *Biopolymers* **1972**, *11*, 2493−2503.

(16) Pickersgill, R. W.; Harris, G. W.; Garman, E. Structure of Monoclinic Papain at 1.60 Å Resolution. *Acta Crystallogr., Sect. B* **1992**, *48*, 59−67.

(17) Zhao, B.; Janson, C. A.; Amegadzie, B. Y.; D'Alessio, K.; Griffin, C.; Hanning, C. R.; Jones, C.; Kurdyla, J.; McQueney, M.; Qiu, X.; Smith, W. W.; Abdel-Meguid, S. S. Crystal Structure of Human Osteoclast Cathepsin K Complex With E−64. *Nature Struct. Biol.* **1997**, *4*, 109−111.

(18) McGrath, M. E.; Klaus, J. L.; Barnes, M. G.; Bromme, D. Crystal Structure of Human Cathepsin K Complexed with A Potent Inhibitor. *Nature Struct. Biol.* **1997**, *4*, 105−109.

(19) Yamashita, D. S.; Smith, W. W.; Zhao, B.; Janson, C. A.; Tomaszek, T. A.; Bossard, M. J.; Levy, M. A.; Marquis, R. W.; Oh, H.-J.; Ru, Y.; Carr, T. J.; Thompson, S. K.; Ijames, C. F.; Carr, S. A.; McQueney, M.; D'Alessio, K. J.; Amegadzie, B. Y.; Hanning, C. R.; Abdel-Meguid, S. S.; DesJarlais, R. L.; Gleason, J. G.; Veber, D. F. Structure And Design of Potent and Selective Cathepsin K Inhibitors *J. Am. Chem. Soc.* **1997**, *119*, 11351−11352.

(20) Thompson, S. K.; Halbert, S. M.; Bossard, M. J.; Tomaszek, T. A.; Levy, M. A.; Meek, T. D.; Zhao, B.; Smith, W. W.; Janson, C. A.; D′Alessio, K. J.; McQueney, M.; Abdel-Meguid, S. S.; DesJarlais, R. L.; Briand, J.; Sarkar, S. K.; Huddleston, M.; Ijames, C.; Carr, S.; Garnes, K. T.; Shu, A.; Heys, J. R.; Bradbeer, J.; Zembryki, D.; Lee-Rykaczewski, L.; James, I. E.; Lark, M.; Drake, F. H.; Gowen, M.; Gleason, J. G.; Veber, D. F. Design of Potent and Selective Human Cathepsin K Inhibitors Which Span the Active Site. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14249−14254.

(21) Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic R Factor Refinement by Molecular Dynamics. *Science* **1987**, *235*, 458−460.

(22) Sack, J. S. CHAIN − A Crystallographic Modeling Program. *J. Mol. Graph.* **1988**, *6*, 224−225.

(23) Kraulis, P. J. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *Appl. Crystallogr.* **1991**, *24*, 946−950.

(24) Votta, B.; Levy, M. A.; Badger, A.; Bradbeer, J.; Dodds, R. A.; James, I. E.; Thompson, S.; Bossard, M. J.; Carr, T.; Connor, J. R.; Tomaszek, T. A.; Szewczuk, L.; Drake, F. H.; Veber, D. F.; Gowen, M. Peptide Aldehyde Inhibitors of Cathepsin K Inhibit Bone Resorption Both In Vitro and In Vivo. *J. Bone Miner. Res.* **1997**, *12*, 1396−1406.

(25) Jones, D. M.; Atrash, B.; Teger-Nilsson, A. C.; Gyzander, E.; Deinum, J.; Szelke, M. Design and synthesis of thrombin inhibitors. *Lett. Pept. Sci.* **1995**, *2*, 147−154.

(26) Jones, D. M.; Atrash, B.; Teger-Nilsson, A. C.; Gyzander, E.; Deinum, J.; Szelke, M. Inhibitors of thrombin containing arginine ketone derivatives. In *Pept. 1994, Proc. Eur. Pept. Symp.*, 23rd; Maia, H. L. S., Ed.; ESCOM: Leiden, 1995; pp 899−900.

JM980249F