ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/7686230

Inhibition of a Cathepsin L-Like Cysteine Protease by a Chimeric Propeptide-Derived Inhibitor †

ARTICLE *in* BIOCHEMISTRY · SEPTEMBER 2005

Impact Factor: 3.02 · DOI: 10.1021/bi047590o · Source: PubMed

CITATIONS
10

READS
18

6 AUTHORS, INCLUDING:

Shafinaz F Chowdhury
Lady Davis Institute for Medical Research
30 PUBLICATIONS **626** CITATIONS
SEE PROFILE

Fabien Lecaille
University of Tours
56 PUBLICATIONS **1,455** CITATIONS
SEE PROFILE

Enrico O. Purisima
McGill University
75 PUBLICATIONS **2,178** CITATIONS
SEE PROFILE

Gilles Lalmanach
University of Tours
86 PUBLICATIONS **1,776** CITATIONS
SEE PROFILE

Available from: Gilles Lalmanach
Retrieved on: 28 January 2016

Inhibition of a Cathepsin L-Like Cysteine Protease by a Chimeric Propeptide-Derived Inhibitor†

Emmanuel Godat,[‡,§] Shafinaz Chowdhury,[§,∥] Fabien Lecaille,[‡] Maya Belghazi,[⊥] Enrico O. Purisima,[∥] and Gilles Lalmanach*[,‡]

INSERM, U618, Tours F-37000, France, Université François Rabelais, Tours F-37000, France, IFR 135, Tours F-37000, France, Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec, Canada, and INRA de Tours, Service de Spectrométrie de Masse pour la Protéomique, Nouzilly, France

Received November 15, 2004; Revised Manuscript Received June 6, 2005

ABSTRACT: Like other papain-related cathepsins, congopain from *Trypanosoma congolense* is synthesized as a zymogen. We have previously identified a proregion-derived peptide (Pcp27), acting as a weak and reversible inhibitor of congopain. Pcp27 contains a 5-mer YHNGA motif, which is essential for selectivity in the inhibition of its mature form [Lalmanach, G., Lecaille, F., Chagas, J. R., Authié, E., Scharfstein, J., Juliano, M. A., and Gauthier, F. (1998) *J. Biol. Chem. 273*, 25112−25116]. In the work presented here, a homology model of procongopain was generated and subsequently used to model a chimeric 50-mer peptide (called H3-Pcp27) corresponding to the covalent linkage of an unrelated peptide (H3 helix from *Antennapedia*) to Pcp27. Molecular simulations suggested that H3-Pcp27 (pI = 9.99) maintains an N-terminal helical conformation, and establishes more complementary electrostatic interactions (E_{coul} = −25.77 kcal/mol) than 16N-Pcp27, the 34-mer Pcp27 sequence plus the 16 native residues upstream from the proregion (E_{coul} = 0.20 kcal/mol), with the acid catalytic domain (pI = 5.2) of the mature enzyme. In silico results correlated with the significant improvement of congopain inhibition by H3-Pcp27 (K_i = 24 nM), compared to 16N-Pcp27 (K_i = 1 μM). In addition, virtual alanine scanning of H3 and 16N identified the residues contributing most to binding affinity. Both peptides did not inhibit human cathepsins B and L. In conclusion, these data support the notion that the positively charged H3 helix favors binding, without modifying the selectivity of Pcp27 for congopain.

Cysteine cathepsins of clan CA family C1 are mostly found in protozoa, plants, and mammals, and represent the most studied family of cysteine proteases (CPs)[1] (*1*). Their catalytic domains share very similar sequences (MEROPS protease database, http://merops.sanger.ac.uk) and three-dimensional (3D) structures (*2*, *3*). CPs are involved in a variety of biological processes, including housekeeping tasks within the endosomal/lysosomal system (intracellular protein degradation), specific functions in MHC-II antigen presentation or in prohormone processing, and pathological processes such as Alzheimer's disease, allergic reactions, tumor invasion, osteoporosis, or rheumatoid arthritis (see for review ref *4*). They are synthesized as zymogens, which are converted to the mature form by proteolytic cleavage and the release of the proregion. These propeptides take part in the proper folding, intracellular trafficking, or secretion of the mature protease (*5*). In addition, propeptides bind and inhibit competitively their mature enzyme (*6*, *7*). Analysis of the 3D structures of CP precursors has provided evidence that the prosegments inhibit their parent enzymes in a similar fashion, except cathepsin X, though they differ slightly in length and sequence (*8*−*14*). The prosegment contacts the enzyme in two main areas: one is a prosegment binding loop (PBL) involving mainly aromatic side chains in hydrophobic interactions, and the other is along the substrate binding cleft, but in an opposite and nonproductive orientation compared to that of the substrate (see for review ref *15*). Experiments carried out with recombinant truncated proregions of cathepsin L have demonstrated that the two N-terminal α-helices (α-1p and α-2p) are essential for propeptide binding and contribute greatly to stability (*9*, *16*). In addition,

† E.G. holds a doctoral fellowship from MENRT (Ministére de l'Education Nationale, de la Recherche et de la Technologie, France). This work was supported by l'ARC (Association pour la Recherche contre le Cancer), VML (Vaincre les Maladies Lysosomales), and the EU (5th Framework Programme, Inco-Dev, ICA4-CT-2000-30035).

* To whom correspondence should be addressed: Université François Rabelais, Faculté de Médecine, INSERM U618 Protéases et Vectorisation Pulmonaires, Equipe Protéases et Pathologies Pulmonaires, 10 Boulevard Tonnellé, 37032 Tours cedex, France. Phone: +33 2 47 36 61 51. Fax: +33 2 47 36 60 46. E-mail: gilles.lalmanach@univ-tours.fr.

‡ INSERM U618, Université François Rabelais, and IFR 135.

§ These authors contributed equally to this work.

∥ National Research Council of Canada.

⊥ INRA de Tours.

[1] Abbreviations: AMC, 7-amino-4-methylcoumarin hydrochloride; Boc, *tert*-butoxycarbonyl; CD, circular dichroism; CP, cysteine protease; DTT, DL-dithiothreitol; E-64, L-3-carboxy-*trans*-2,3-epoxypropionyl-leucylamido-(4-guanido)butane; ESI-Q-TOF, ionization quadrupole time-of-flight; Fmoc, *N*-(9-fluorenyl)methoxycarbonyl; HATU, *o*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MHC-II, major histocompatibility complex class II; PBL, prosegment binding loop; PDB, Protein Data Bank; rink amide MBHA resin, 4-(2′,4′-dimethoxyphenyl-Fmoc-aminomethyl-phenoxyacetamido-norleucyl)-4-methylbenzhydrylamine resin; SCR, structurally conserved region; SIE, solvated interaction energy; SVR, structurally variable region; TFA, trifluoroacetic acid; TFE, trifluoroethanol; Z, benzyloxycarbonyl.

10.1021/bi047590o CCC: $30.25 © 2005 American Chemical Society
Published on Web 07/12/2005

a short helix (α-3p) and aromatic residues (Phe63p and Phe56p) of procathepsin L interact with residues of the hydrophobic cluster of the protein binding loop (*9*). Although the C-terminus proregion does not contribute significantly to the overall binding energy toward the cognate enzyme, it provides compensating charges to those on the contacting surface of the enzyme and may, therefore, contribute to the selectivity of the recognition process (*15*). However discrepancies and deviating data have been reported concerning the relative importance of propeptide regions for specificity and selectivity (*5*).

Congopain, the major CP from *Trypanosoma congolense*, the causative agent of African trypanosomosis in livestock, presents immunosuppressive effects and may act as a pathogenic factor in the host−parasite relationships (*17*). Like cruzipain from *Trypanosoma cruzi*, congopain is structurally and functionally related to mammalian cysteine cathepsins (see for review refs *18* and *19*). The sequence of its catalytic domain is 68% identical with that of cruzain and more than 65% identical with that of cathepsin K (*20*, *21*). Congopain has a C-terminal extension of unknown structure and function that is linked to the catalytic domain by a polyproline stretch (*22*). During activation of the zymogen, removal of the proregion [105 residues (*23*)], which contains the ER(F/W)N-(I/V)N-like motif which is specific to cathepsin L-like enzymes (*24*), most probably occurs via an autocatalytic mechanism (*25*). Additionally, we have identified a proregion-derived peptide (Asn59p−Leu92p) located after the putative N-terminal α-2p helix, which acts as a selective inhibitor of congopain and cruzipain. This peptide (called Pcp27) contains a 5-mer YHNGA motif reported to be essential for selectivity in the inhibition of the mature enzyme by its propeptide (*26*).

The 16-mer peptide (amino acids 43−58), derived from the third helix (H3) homeodomain of the *Drosophila Antennapedia* transcription factor, is also called penetratin (*27*). This peptide may carry hydrophilic cargoes (such as oligonucleotides or proteins) through the biological membranes, and has been used with success in the intracellular delivery of an epoxysuccinyl-based cathepsin B inhibitor (*28*) and a calpastatin-derived calpain inhibitor (*29*). Nevertheless, in our work, the rationale for selecting the H3 domain from *Antennapedia* does not depend on its cell penetrating properties but on its ability to adopt a helical secondary structure (*30*). Thus, substitution of the native α-2p helix by H3 represents an appropriate approach to evaluating the specific role of this structural motif in binding of the propeptide to its parent enzyme. Taking together the critical contribution of α-helices for propeptide binding and molecular studies of the 105-residue procongopain propeptide, we synthesized two 50-mer peptides corresponding to the native 50-mer proregion-derived peptide (16N-Pcp27) and the chimera of penetratin and Pcp27 (H3-Pcp27). Comparison of their inhibitory properties supports the notion that affinity dramatically depends on electrostatic interactions between the catalytic domain of congopain and the positively charged residues of the H3 helix.

EXPERIMENTAL PROCEDURES

L-3-Carboxy-*trans*-2,3-epoxypropionyl-leucylamido-(4-guanidino)butane (E-64) and DL-dithiothreitol (DTT) were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France). Ac-Phe-Arg-AMC and Z-Phe-Arg-AMC were from Bachem (Weil am Rhein, Germany). All other reagents were analytical grade.

Structure Prediction and Model Building. The model of procongopain was built in two stages. First, a model of the mature enzyme was constructed. Then, the proregion was constructed and added. The sequence of mature congopain is significantly homologous with a number of other cysteine proteases, with the highest degree of sequence identity (68%) shared with cruzain. Therefore, the cruzain crystal structure (PDB entry laim) was used as a template to model the catalytic domain of mature congopain. The homology modeling was carried out using COMPOSER and PROTEIN LOOPS (SYBYL 6.6, Tripos Inc., St. Louis, MO). Only one structurally conserved region (SCR) was needed to build an initial model of the catalytic domain of congopain without any insertion or deletion.

No crystal structure of the proenzyme of cruzain is available. Hence, we had to rely on proenzyme crystal structures of more distant homologues. The sequence of the proregion of procathespin L is 25.5% identical with that of the proregion of procongopain. Despite the relatively low level of sequence identity, a reasonable alignment could be made. In particular, the sequence spanning the active site cleft could be readily identified. The crystal structure of procathepsin L (PDB entry Ics8) was therefore used as a template to model the proregion of procongopain using COMPOSER. Four structurally conserved regions (SCRs) were built on the basis of the procathepsin L template. Separating the four SCRs are three structurally variable regions (SVRs). Two of the three SVRs were constructed using PROTEIN LOOPS. The third SVR is basically a one-residue insertion of a threonine residue that forms part of the active site spanning sequence. This insertion was generated manually. The N- and C-terminal extensions for which there is no corresponding template structure were assigned sterically allowed, random conformations. The proregion model thus assembled was then docked into the modeled structure of mature congopain using procathepsin L as a guide. The two chains were then joined by forming a peptide bond between the N-terminal residue of mature congopain and the C-terminal residue of the proregion of the procongopain.

Structural refinement of the complex was carried out using conjugate-gradient energy minimization with the AMBER force field (*31*) to a gradient of 0.05 kcal mol^{-1} $Å^{-1}$. Arg, Lys, and His side chains were given a +1 charge. Asp, Glu, and the catalytic Cys side chains had a −1 charge. A distance-dependent ($\epsilon = 4r$) dielectric function and an 8 Å nonbonded cutoff was applied. Sequential energy minimization runs were carried out to allow a gradual relaxation of the structure, starting from SVRs and constructed non-homologous loops, then including the side chain of the SCRs, and finally energy minimization of the entire structure. The quality of the model was assessed with PROCHECK (*32*).

The H3-Pcp27 Chimeric Peptide. The procongopain model was used as a template for modeling the complex of mature congopain with the chimeric peptide H3-Pcp27. The inhibitory segment Pcp27, i.e., residues Asn59p−Leu92p of the propeptide of procongopain, is a 34-amino acid fragment that includes the substrate binding cleft-occupying segment

YHNGA (*26*). The 16-residue H3 at the N-terminus of the chimeric peptide is derived from the H3 helix of *Antennapedia* (*27*). The bound conformation of the Pcp27 end was taken directly from the procongopain model. The H3 end was constructed on the basis of the native 16-residue N-terminal segment preceding Pcp27 in the procongopain model. Energy refinement of the complex was carried out as described above. In addition to the complex of mature congopain with H3-Pcp27, we also constructed models of the complex of congopain with Pcp27 and with 16N-Pcp27.

Binding Energy Calculations. Predicted binding free energies of the inhibitory peptides were calculated using a solvated interaction energy (SIE) approach. The SIE is calculated as

$$\mathrm{SIE} = E_{\mathrm{coul}} + \beta E_{\mathrm{vdw}} + \Delta G_{\mathrm{rf}} + \Delta G_{\mathrm{cav}} \qquad (1)$$

where E_{coul} and E_{vdw} are the intermolecular Coulomb and van der Waals interaction energies, respectively. No cutoffs were used in the Coulomb and van der Waals energy calculations. An internal dielectric constant of 20 was used for E_{coul}; the coefficient β of 0.0361 simulated the loss of solute−solvent van der Waals interactions upon binding. ΔG_{rf} is the difference in reaction field energy between the complex and the free state. The reaction field energy, ΔG_{rf}, is calculated with a continuum model using a boundary element solution to the Poisson equation using BRI BEM (*33*, *34*) with internal and external dielectric constants of 20 and 78.5, respectively. ΔG_{cav} is the cavity cost of complex formation and is taken to be proportional to the change in molecular surface area upon binding.

$$\Delta G_{\mathrm{cav}} = \alpha \Delta \mathrm{MSA} \qquad (2)$$

The coefficients (i.e., cavity term coefficient, $a = 0.011$ kcal mol^{-1} Å^2) were obtained by calibrating the functional form in eqs 1 and 2 on an extensive database of protein−ligand complexes (Naim et al., manuscript in preparation).

In addition, we carried out virtual alanine scanning of H3 and the corresponding native sequence (16 N) to identify side chain contributions to binding, according to the method of Huo et al. (*35*). This was performed by mutating each residue in turn to Ala and calculating the change in the binding affinity (eq 1) of the mutated H3-Pcp27 and 16N-Pcp27 for mature congopain.

Enzymes. Congopain (CP-2 type) was purified as previously reported (*36*). The recombinant catalytic domain of cruzipain, i.e., cruzain, was a kind gift from J. McKerrow (University of California, San Francisco, CA). Human cathepsins B and L were supplied by Calbiochem (VWR international, Pessac, France). The activation buffers for enzyme assays were as follows: 0.1 M acetate buffer for pH 5.5 containing 6 mM DTT and 2 mM EDTA for congopain and cruzipain, and 0.1 M phosphate buffer for pH 6.0 containing 1 mM EDTA and 2 mM DTT for cathepsins B and L. Enzymes were incubated in their activation buffer for 5 min at 37 °C prior to kinetic measurements (Kontron SFM 25 spectrofluorimeter). Their active sites were titrated with E-64, using Z-Phe-Arg-AMC as a substrate (excitation wavelength of 350 nm and emission wavelength of 460 nm).

Peptide Synthesis. L-Amino acids were purchased from NeoMPS (Strasbourg, France), Novabiochem (VWR international), and Advanced Chemtech (Cambridge, U.K.). The AYHNGAA peptide, corresponding to the 5-mer substrate binding site occupying residues flanked by two alanyl residues, was synthesized by the Fmoc chemistry procedure (ABI 431A synthesizer, Applied Biosystems, Warrington, U.K.), using a Rink Amide MBHA resin (Novabiochem) (*26*). The Pcp27 peptide was prepared according to the same protocol. The H3 peptide (RQIKIWFQNRRMKWKK) was synthesized on an automated solid phase peptide synthesizer (The Pioneer, Applied Biosystems), using a Fmoc-Lys(Boc)-Novasyn TGT resin (Novabiochem), and HATU as the activator (double coupling), and analyzed by analytical RP-HPLC (Brownlee C18 OD 300 column), using a 35 min linear (0 to 60%) gradient of acetonitrile in 0.1% trifluoroacetic acid. The H3 peptide was further purified by semi-preparative reverse phase chromatography (Vydac C18 218TPS1 column), using the elution conditions indicated above.

Synthesis of the chimeric 50-mer peptide H3-Pcp27 was performed as summarized in Figure 1. The condensation step between H3 and Pcp27 was controlled by PAGE analysis (*37*) and analyzed by RP-HPLC on a Brownlee BU 300 column, using a 30 min linear (0 to 90%) gradient of acetonitrile in 0.1% TFA at a flow rate of 0.3 mL/min (ChromQuest Chromatography Workstation, ThermoFinnigan, les Ulis, France). H3-Pcp27 was purified by semi-preparative RP-HPLC, using a Pep RPC HR 5/5 column (Amersham Pharmacia Biotech) [30 min linear (0 to 90%) gradient of acetonitrile in 0.1% TFA]. A similar procedure was used for the synthesis of 16N-Pcp27. Molecular weights of H3, Pcp27, 16N-Pcp27, and H3-Pcp27 were checked by MALDI-TOF (Bruker biflex III, Wissembourg, France) or ESI-Q-TOF mass spectroscopy (Micromass, Birmingham, U.K.).

Circular Dichroism Spectroscopy. CD spectra were measured at room temperature (22 °C) over 280−190 nm, using a Jasco J-810 150 S spectropolarimeter. The path length of the cell was 1 mm, and data from repeated scans were averaged. Experiments were carried out in 0.1 M phosphate buffer (pH 5.5) or 0.1 M phosphate buffer (pH 5.5) with 50% TFE (v/v).

Kinetics. The activities of trypanosomal and mammalian enzymes were measured with Ac-Phe-Arg-AMC and with Z-Phe-Arg-AMC, respectively. Congopain (3.3 nM) was activated for 2 min at 37 °C in its assay buffer prior to incubation with peptides AYHNGAA, Pcp27, 16N-Pcp27, and H3-Pcp27 (0−200 μM) for 5 min. The residual enzymatic activity was recorded following addition of substrate (1−10 μM). The K_{i} values were calculated by plotting $1/v$ against [I] (triplicate experiments). The K_{i} values toward cruzain (1 nM) were determined under the same experimental conditions. The capacities of these peptides to inhibit cathepsin B (4 nM) and cathepsin L (4 nM) were measured using Z-Phe-Arg-AMC (10 μM) as a substrate.

Peptide stability upon incubation with proteases was determined as previously described (*20*). Briefly peptides (60 μM) and congopain (3.3 nM) were mixed and incubated in the activation buffer at 37 °C for 0−5 h. After addition of E-64 (10 μM) to inactivate the enzyme, the sample was analyzed by RP-HPLC (C18 OD 300 Brownlee column, linear 0 to 60% gradient of acetonitrile in 0.1% TFA), by using Spectacle (ThermoQuest). The same experiment was

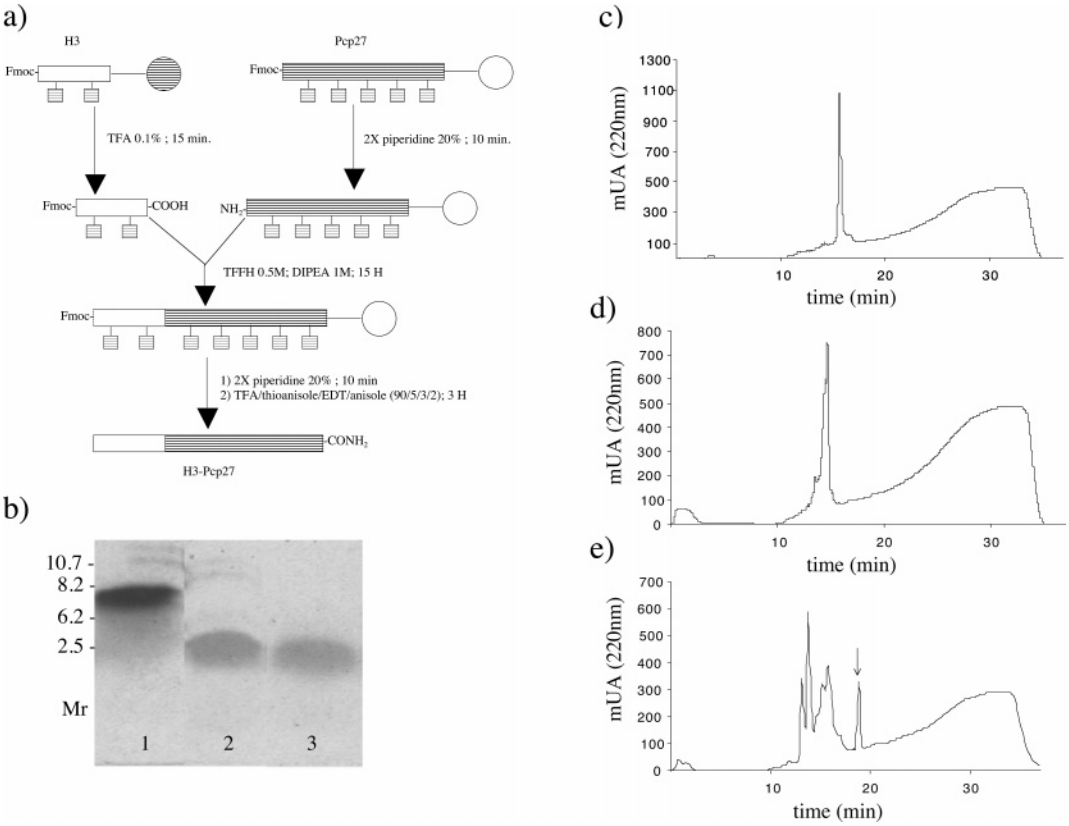

FIGURE 1: Synthesis of H3-Pcp27. (a) Schematic procedure for the convergent synthesis of the chimeric 50-mer peptidyl amide H3-Pcp27. After separate syntheses of peptides H3 and Pcp27, the two peptidyl moieties were conjugated by a condensation step, as described in Experimental Procedures: (●) acid-sensitive resin and (○) MBHA resin. (b) The ligation between H3 and Pcp27 was controlled by PAGE analysis, according to Schagger and von Jagow G (*37*): lane 1, H3-Pcp27; lane 2, H3; and lane 3, Pcp27. Crude synthetic peptides were checked by analytical RP-HPLC, (c) H3, (d) Pcp27, and (e) H3-Pcp27 (indicated with an arrow), and analyzed by mass spectroscopy before purification.

repeated for cruzain and cathepsin L, and proteolysis products were fractionated by reverse phase chromatography, as described above.

RESULTS AND DISCUSSION

Homology Modeling of Procongopain. In a previous study using overlapping 15-mer peptides that cover the prosequence of congopain, Lalmanach and co-workers (*26*) have identified peptides having the common 5-mer motif Tyr81p-His-Asn-Gly-Ala85p (74p−78p, procathepsin L numbering) that inhibited selectively congopain and cruzipain. Although the prosequences of trypanosomal cysteine proteases share a low degree of homology with their mammalian and plant homologues, the YHNGA moiety is in the same position, relative to the C-terminus of the procongopain sequence, as the residues that contact the substrate binding cleft of cathepsin B, cathepsin L, and caricain. Moreover, this 5-mer motif contains a Gly residue, a steric requirement at the S1 subsite for a binding mode opposite to that of a substrate (*38*), supporting the notion that the prosegment of trypanosomal cysteine proteases shares a common autoinhibitory mechanism as the mammalian procathepsins and that this 5-mer is the active site-spanning peptide sequence.

These similarities prompted us to model the three-dimensional structure of procongopain. The secondary structure of the proregion of congopain as predicted using PSI-PRED (*39*) and PHD (*40*) revealed a high degree of similarity between the secondary structures of procathepsin

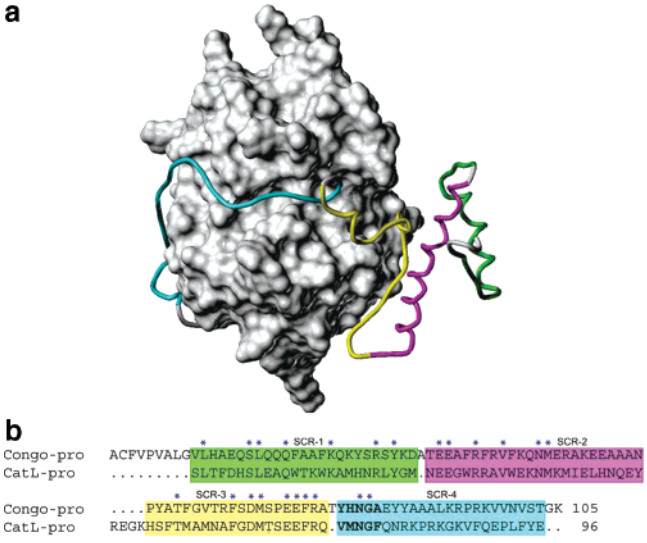

FIGURE 2: Modeled structure of procongopain. (a) The catalytic domain of procongopain is displayed as a Connolly surface. The modeled proregion is depicted as a tube. (b) Alignment of the primary sequences of the proregions of congopain (105 residues) and of human cathepsin L (96 residues). Structurally conserved regions (SCRs) are enclosed in colored boxes. Identical residues are labeled with asterisks. Substrate binding site-occupying residues are highlighted in bold. The color of the segments corresponds with the SCRs used during model generation.

L and procongopain proregions (Figure 2), facilitating the alignment for homology modeling. The catalytic domain of the procongopain model is very similar to the cruzipain

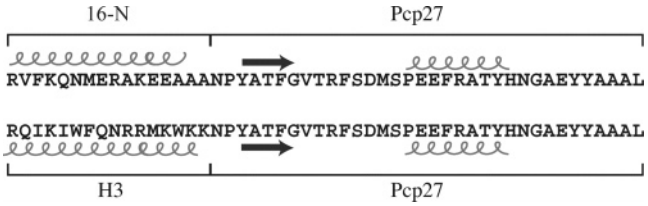

FIGURE 3: Secondary structures of H3-Pcp27 and 16N-Pcp27. Secondary structures of H3-Pcp27 and its corresponding native peptide, 16N-Pcp27, were predicted by PSI-PRED (*39*) and PHD (*40*), as reported in Experimental Procedures. Helices are shown as coils; sheets are shown as arrows.

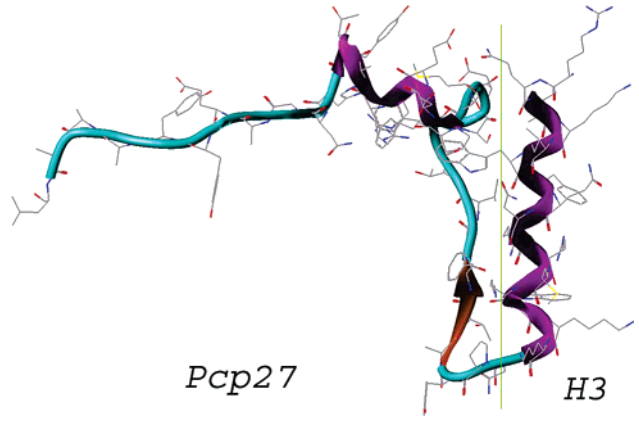

FIGURE 4: Molecular model of H3-Pcp27. Schematic ribbon diagram of the overall fold of the 16-mer H3 (derived from the homeodomain of the *Drosophila Antennapedia* transcription factor) covalently linked to the inhibitory procongopain-derived peptide Pcp27 (*26*). Pcp27 displayed one β-strand (orange) and one short α-helix preceding the substrate binding cleft (magenta). The H3 fragment displayed one extended α-helix (magenta), which forms part of the loop connecting H3 and Pcp27.

crystal structure as expected (68% identical) and is composed of one chain of 214 amino acid residues folded into two domains. The α-helical structures are predominant in one domain and extensive β-sheets in the other. Catalytic triad residues Cys25, His162, and Asn182 are positioned in a cleft between the two domains, which is in agreement with other papain-like enzymes (*1*−*3*, *15*, *20*). The propeptide chain of 105 amino acids displays similar secondary structural features to 96-mer procathepsin L (*9*), namely, a short α-helix (Gln15p−Lys27p), a longer helix (Arg30p−Ala58p) corresponding to the α-2p procathepsin L, a loop followed by a short β-strand (Ala62p−Phe64p), a loop followed by a two-turn α-helix (Pro74p−Tyr81p) related to the α3p procathepsin L, and an extended C-terminal fragment occupying the substrate binding cleft which ends in a short β-strand (Arg94p and Pro95p) (Figure 2).

Modeling of H3-Pcp27. If the α-2p-related helix upstream of Pcp27 (residues Asn59p−Leu92p) contributes to binding in a nonspecific manner, replacement of these native N-terminal residues with an unrelated helix will not abrogate selectivity. To validate this hypothesis, the chimeric peptide H3-Pcp27 was designed. The 16-mer H3 maintains its helicoidal structure in the chimeric peptide, as suggested by secondary structure prediction (Figure 3), and H3-Pcp27 adopts a structural pattern identical to its corresponding 50-mer native peptide 16N-Pcp27. Congopain-bound H3-Pcp27 was further modeled as described in Experimental Procedures (Figure 4). Energy refinement of the complex suggests that H3 is accommodated well by congopain. Binding free

Table 1: Modeled Binding Free Energies of Pcp27, H3-Pcp27, and 16N-Pcp27 against Mature Congopain[a]

peptide	βE_{vdw}	E_{coul}	ΔG_{rf}	ΔG_{cav}	ΔG_{calc}
H3-Pcp27	−7.29	−25.77	32.30	−17.34	−18.10
Pcp27	−6.84	6.38	6.91	−15.91	−9.46
16N-Pcp27	−6.88	0.20	11.96	−15.97	−10.69

[a] All quantities are in kilocalories per mole.

energies were calculated for Pcp27, 16N-Pcp27, and H3-Pcp27, using the modeled complexes. The results are summarized in Table 1. H3-Pcp27 was predicted to bind more strongly (ΔG_{calc} = −18.10 kcal/mol) than Pcp27 to congopain (ΔG_{calc} = −9.46 kcal/mol). Interestingly, calculations also predict that H3-Pcp27 binds better than 16N-Pcp27 (ΔG_{calc} = −10.69 kcal/mol). Indeed, free 16N-Pcp27 (pI = 5.69) makes favorable electrostatic interactions with solvent, which are lost when 16N-Pcp27 binds with the enzyme. This is not fully compensated by the Coulomb interaction energy, which is unfavorable in the bound state because of repulsive charge interactions with the mature domain of congopain (pI = 5.2). On the other hand, while H3-Pcp27 presents also an unfavorable reaction field binding energy (ΔG_{rf} = 32.30 kcal/mol), this is mostly compensated by favorable Coulomb interactions (E_{coul} = −25.77 kcal/mol) between H3-Pcp27 (pI = 9.99) and congopain. In fact, compared to the basic residues present in the native 16N procongopain sequence (Arg43p, Lys46p, Arg51p, and Lys53p), H3 carries an overall higher positive charge (Arg43p, Lys46p, Arg52p, Arg53p, Lys55p, Lys57p, and Lys58p), enhancing electrostatic interactions with the negatively charged catalytic domain of congopain. Furthermore, H3-Pcp27 also exhibits better intermolecular van der Waals interaction (E_{vdw}) and nonpolar solvation energy (ΔG_{cav}) compared to 16N-Pcp27 (Table 1). To obtain insights into the individual contribution of H3 residues to the affinity of H3-Pcp27, virtual alanine scanning was carried out by calculating relative binding free energies of both H3-Pcp27 and 16N-Pcp27 to mature congopain, according to ref *35*. $\Delta\Delta G_{binding}$ data from the computational alanine scanning are reported in Figure 5. A total of nine (Arg43p, Lys46p, Trp48p, Arg52p, Arg53p, Lys55p, Trp56p, Lys57p, and Lys58p) of 16 residues in H3 favor binding to mature congopain (Figure 5a). On the other hand, only five residues (Arg43p, Lys46p, Asn48p, Arg51p, and Lys53p) favor binding in the native sequence (16N) (Figure 5b), and are partly counterbalanced by the presence of three unfavorable negatively charged glutamic acid residues (50p, 54p, and 55p). It has to be noticed that Arg43p, Lys46p, Arg51p, and Lys53p (from the procongopain sequence) are substitued with Ala43p, Glu46p, Ile52p, and Glu53p, respectively, in the corresponding mer-procathepsin L moiety (see ref *9*), and this supports the notion that this overall negative charge induces repulsive charge interactions with the mature domain of congopain.

Kinetics. Because of the extensive length of the chimeric peptide (50 residues), we have adopted a strategy of convergent solid phase peptide synthesis (as summarized in Figure 1), consisting of the separate synthesis of H3 and Pcp27 followed by a condensation step. The ligation between H3 and Pcp27 was checked by SDS−PAGE (*36*) prior to mass analysis. Results are as follows: H3, calcd MW 2468 Da, found MW 2469 Da; Pcp27, calcd MW 3849 Da, found MW 3848 Da; and H3-Pcp 27, calcd MW 6300 Da, found

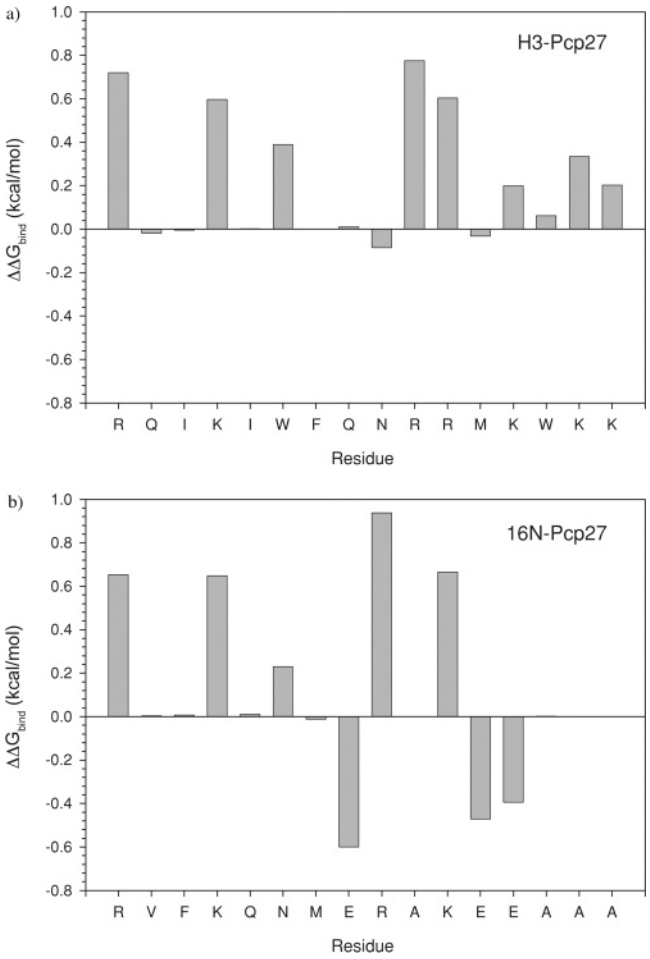

FIGURE 5: Alanine scanning of the binding contribution of each residue in the H3 and 16N segments (residues 43p−58p) to mature congopain: (a) H3-Pcp27 and (b) 16N-Pcp27. Computational alanine scanning was performed according to the method of ref *35*. Positive values indicate a favorable binding contribution from the original side chain.

Table 2: Inhibition Constants for Inhibition of Congopain by *T. congolense* Proregion-Related Peptides and by H3-Pcp27

	K_i (nM)				
enzyme	AYHNGAA	Pcp27[a]	16N-Pcp27	H3-Pcp27	affinity gain
congopain	225000	2000	1000	24	9375
cathepsin L	nd[b]	nsi[c]	hydr[d]	hydr[d]	–
cathepsin B	nd[b]	nsi[c]	nsi[c]	nsi[c]	–

[a] From Lalmanach et al., 1998 (*26*). [b] Not determined. [c] No significant inhibition. [d] Hydrolysis of the peptide.

MW 6302 Da. Likewise, results for the native proregion-derived peptide 16N-Pcp 27 are as follows: calcd MW 5704 Da, found MW 5704 Da. Peptides were further assayed for their ability to inhibit congopain. The 7-mer peptide AYHNGAA, corresponding to the active site-interacting YHNGA sequence flanked by two alanyl residues (*26*), inhibited congopain very poorly (Table 2). Pcp27 inhibited congopain with a significantly lower K_i value ($K_i = 2\ \mu$M), due to an increased level of stabilization of the congopain−Pcp27 complex, which could be related to the presence of a putative short helix (corresponding to the α-3p helix from procathepsin L) (*26*) and aromatic residues (Phe63p and Phe56p) interacting with residues of the hydrophobic cluster of the protein binding loop (PBL) as reported for cathepsin L (*9*). The N-terminal addition of 16 residues did not significantly improve the binding of 16N-Pcp27 to congopain (K_i in the micromolar range) despite the presence of residues partly corresponding to the α-2p helix. This is in good agreement with molecular simulations because of unfavorable electrostatic interactions between 16N-Pcp27 (pI = 5.69) and the catalytic domain of congopain (pI = 5.2) (Table 1). In addition, selectivity is not affected by N-terminal elongation given that 16N-Pcp27 did not inhibit human cathepsins L and B (Table 2).

Conversely, the presence of the H3 helix dramatically increased the affinity of Pcp27 for congopain, leading to a potent reversible inhibitor ($K_i = 24$ nM). In close agreement with this experimental finding, H3-Pcp27 exhibits a basic pI of 9.99 which compares to the theoretical pI of the 105-mer proregion of congopain (pI = 9.5), supporting the notion that electrostatic interactions play a critical role in the binding affinity. This K_i value compares to data reported for recombinant truncated propeptides of cathepsin L (*16*), and confirms that the N-terminal addition of Lys/Arg-rich H3 stabilizes the inhibitory complex between H3-Pcp27 and congopain, by establishing tight and extensive interactions with the acidic mature form of congopain. Moreover, H3-Pcp27, as well as Pcp27 and 16N-Pcp27, did not inhibit human cathepsins L and B, indicating that H3 greatly improves the overall energy of binding of Pcp27 to congopain, without modifying its selectivity toward its parent enzyme. This supports the importance of having a helical secondary structure element with complementary charges, thus providing a supplementary anchoring site for binding, and confirms in silico alanine scanning data. Analysis of the Pcp27 peptide by circular dichroism has suggested that it adopts a partial α-helical conformation (*26*). Despite no increase in secondary structure observed for unbound H3-Pcp27 by CD (data not shown), this is in agreement with other studies indicating that conformational properties of the H3 helix of the homeodomain of *Antennapedia* and of its derivatives mostly depend of its environment (*41*). Nevertheless, taken together with the common inhibition mechanism of the catalytic domain by its corresponding propeptide (see for review refs *5* and *15*), kinetics data show that H3 may adopt the required scaffold upon interactions between H3-Pcp27 and congopain. Interestingly, cruzain (theoretical pI = 5.1), the recombinant catalytic domain of cruzipain which lacks the C-terminal extension (*42*), was inhibited as well as the full-length CP2-type congopain by Pcp27 ($K_i = 2\ \mu$M) and H3-Pcp27 ($K_i = 36$ nM). Therefore, this sustains the hypothesis that the C-terminal domain of trypanosomal CPs does not interfere with the binding of the propeptide to the active site, as previously suggested (*18*).

In the vicinity of the active site cleft, prosegment residues are positioned in a reverse direction to that of a substrate, conferring resistance to proteolysis (*15*). Pcp27 and H3-Pcp27 retained their full inhibitory capacity after incubation for 5 h, suggesting their orientation is similar to that of the whole prosegment and they obey the general reverse binding mode. However, under these conditions, some limited cleavage, corresponding to a minor, substrate-like binding mode, occurred after the Phe78p-Arg79p pair, which is one of the preferred pairs of residues accommodated at S2 and S1 subsites in papain-like cysteine proteases. In contrast, PB8

(60 μM), a procathepsin B-derived inhibitor (*43*), did not inhibit congopain (3.3 nM) and is completely hydrolyzed in less than 30 min, as shown by reverse phase HPLC.

In brief, a chimeric peptide consisting of Pcp27, a native fragment of procongopain, covalently attached to the unrelated helical peptide H3 from *Antennapedia*, was predicted not only to form a complex with its parent enzyme but also to bind better than the native 50-mer corresponding sequence (16N-Pcp27). Kinetic studies confirmed that the chimera indeed bound cathepsin L-like congopain in the nanomolar range, supporting the notion that affinity primarily depends on the presence of an anchoring positively charged helix, but not in a strict sequence-specific manner. On the other hand, the chimera was rapidly hydrolyzed by human cathepsin L, presumably because specific residues of the active site-spanning fragment (YHNGA) imposed selectivity toward congopain from *T. congolense*.

ACKNOWLEDGMENT

We thank Drs. Alain Boulangé and Edith Authié (ILRI, Nairobi, Kenya) for supplying us with congopain, Dr. James McKerrow (University of California, San Francisco, CA) for his generous gift of recombinant cruzipain, and Dr. Pradip Nandi (Centre INRA de Tours) for circular dichroism facilities.

REFERENCES

1. Turk, V., Turk, B., and Turk, D. (2001) Lysosomal cysteine proteases: Facts and opportunities, *EMBO J. 20*, 4629−4633.
2. Turk, B., Turk, V., and Turk, D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors, *Biol. Chem. 378*, 141−150.
3. McGrath, M. E. (1999) The lysosomal cysteine proteases, *Annu. Rev. Biophys. Biomol. Struct. 28*, 181−204.
4. Lecaille, F., Kaleta, J., and Bromme, D. (2002) Human and parasitic papain-like cysteine proteases: Their role in physiology and pathology and recent developments in inhibitor design, *Chem. Rev. 102*, 4459−4488.
5. Wiederanders, B., Kaulmann, G., and Schilling, K. (2003) Functions of propeptide parts in cysteine proteases, *Curr. Protein Pept. Sci. 4*, 309−326.
6. Fox, T., de Miguel, E., Mort, J. S., and Storer, A. C. (1992) Potent slow-binding inhibition of cathepsin B by its propeptide, *Biochemistry 31*, 12571−12576.
7. Mach, L., Mort, J. S., and Glossl, J. (1994) Noncovalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme, *J. Biol. Chem. 269*, 13036−13040.
8. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) Structure of rat procathepsin B: Model for inhibition of cysteine protease activity by the proregion, *Structure 4*, 405−416.
9. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment, *EMBO J. 15*, 5492−5503.
10. Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, J. A. (1996) The prosequence of procaricain forms an α-helical domain that prevents access to the substrate-binding cleft, *Structure 4*, 1193−1203.
11. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide, *FEBS Lett. 384*, 211−214.
12. LaLonde, J. M., Zhao, B., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M., and Smith, W. W. (1999) The crystal structure of human procathepsin K, *Biochemistry 38*, 862−869.
13. Sivaraman, J., Lalumiere, M., Menard, R., and Cygler, M. (1999) Crystal structure of wild-type human procathepsin K, *Protein Sci. 8*, 283−290.
14. Sivaraman, J., Nagler, D. K., Zhang, R., Menard, R., and Cygler, M. (2000) Crystal structure of human procathepsin X: A cysteine protease with the proregion covalently linked to the active site cysteine, *J. Mol. Biol. 295*, 939−951.
15. Groves, M. R., Coulombe, R., Jenkins, J., and Cygler, M. (1998) Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes, *Proteins 32*, 504−514.
16. Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J. S., and Menard, R. (1996) Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases, *Biochemistry 35*, 8149−8157.
17. Authie, E. (1994) Trypanosomiasis and trypanotolerance in cattle: A role for congopain? *Parasitol. Today 10*, 360−364.
18. Lalmanach, G., Boulange, A., Serveau, C., Lecaille, F., Scharfstein, J., Gauthier, F., and Authie, E. (2002) Congopain from *Trypanosoma congolense*: Drug target and vaccine candidate, *Biol. Chem. 383*, 739−749.
19. Sajid, M., and McKerrow, J. H. (2002) Cysteine proteases of parasitic organisms, *Mol. Biochem. Parasitol. 120*, 1−21.
20. Lecaille, F., Authie, E., Moreau, T., Serveau, C., Gauthier, F., and Lalmanach, G. (2001) Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. Probing the S2 pocket with phenylalanine-derived amino acids, *Eur. J. Biochem. 268*, 2733−2741.
21. Lecaille, F., Weidauer, E., Juliano, M. A., Bromme, D., and Lalmanach, G. (2003) Probing cathepsin K activity with a selective substrate spanning its active site, *Biochem. J. 375*, 307−312.
22. Chagas, J. R., Authie, E., Serveau, C., Lalmanach, G., Juliano, L., and Gauthier, F. (1997) A comparison of the enzymatic properties of the major cysteine proteinases from *Trypanosoma congolense* and *Trypanosoma cruzi*, *Mol. Biochem. Parasitol. 88*, 85−94.
23. Boulange, A., Serveau, C., Brillard, M., Minet, C., Gauthier, F., Diallo, A., Lalmanach, G., and Authie, E. (2001) Functional expression of the catalytic domains of two cysteine proteinases from *Trypanosoma congolense*, *Int. J. Parasitol. 31*, 1435−1440.
24. Karrer, K. M., Peiffer, S. L., and DiTomas, M. E. (1993) Two distinct gene subfamilies within the family of cysteine protease genes, *Proc. Natl. Acad. Sci. U.S.A. 90*, 3063−3067.
25. Serveau, C., Boulange, A., Lecaille, F., Gauthier, F., Authie, E., and Lalmanach, G. (2003) Procongopain from *Trypanosoma congolense* is processed at basic pH: An unusual feature among cathepsin L-like cysteine proteases, *Biol. Chem. 384*, 921−927.
26. Lalmanach, G., Lecaille, F., Chagas, J. R., Authie, E., Scharfstein, J., Juliano, M. A., and Gauthier, F. (1998) Inhibition of trypanosomal cysteine proteinases by their propeptides, *J. Biol. Chem. 273*, 25112−25116.
27. Derossi, D., Chassaing, G., and Prochiantz, A. (1998) Trojan peptides: The penetratin system for intracellular delivery, *Trends Cell Biol. 8*, 84−87.
28. Schaschke, N., Deluca, D., Assfalg-Machleidt, I., Hohneke, C., Sommerhoff, C. P., and Machleidt, W. (2002) Epoxysuccinyl peptide-derived cathepsin B inhibitors: Modulating membrane permeability by conjugation with the C-terminal heptapeptide segment of penetratin, *Biol. Chem. 383*, 849−852.
29. Gil-Parrado, S., Assfalg-Machleidt, I., Fiorino, F., Deluca, D., Pfeiler, D., Schaschke, N., Moroder, L., and Machleidt, W. (2003) Calpastatin exon 1B-derived peptide, a selective inhibitor of calpain: Enhancing cell permeability by conjugation with penetratin, *Biol. Chem. 384*, 395−402.
30. Lindberg, M., and Graslund, A. (2001) The position of the cell penetrating peptide penetratin in SDS micelles determined by NMR, *FEBS Lett. 497*, 39−44.
31. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
32. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
33. Purisima, E. O., and Nilar, S. H. (1995) A simple yet accurate boundary element method for continuum dielectric calculations, *J. Comput. Chem. 16*, 681−689.

34. Purisima, E. O. (1998) Fast summation boundary element methods for calculating solvation free energies of macromolecules, *J. Comput. Chem. 19*, 1494−1505.
35. Huo, S., Massova, I., and Kollman, P. A. (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex, *J. Comput. Chem. 23*, 15−27.
36. Authie, E., Muteti, D. K., Mbawa, Z. R., Lonsdale-Eccles, J. D., Webster, P., and Wells, C. W. (1992) Identification of a 33-kilodalton immunodominant antigen of *Trypanosoma congolense* as a cysteine protease, *Mol. Biochem. Parasitol. 56*, 103−116.
37. Schagger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, *Anal. Biochem. 166*, 368−379.
38. Chowdhury, S. F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cygler, M., Sulea, T., and Purisima, E. O. (2002) Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides, *J. Med. Chem. 45*, 5321−5329.
39. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol. 292*, 95−202.
40. Rost, B., and Sander, C. (1994) Combining evolutionary information and neural networks to predict secondary structure, *Proteins: Struct., Funct., Genet. 19*, 55−72.
41. Lindberg, M., Biverstahl, H., Graslund, A., and Maler, L. (2003) Structure and positioning comparison of two variants of penetratin in two different membrane mimicking systems by NMR, *Eur. J. Biochem. 270*, 3055−3063.
42. Gillmor, S. A., Craik, C. S., and Fletterick, R. J. (1997) Structural determinants of specificity in the cysteine protease cruzain, *Protein Sci. 6*, 1603−1611.
43. Chagas, J. R., Ferrer-Di Martino, M., Gauthier, F., and Lalmanach, G. (1996) Inhibition of cathepsin B by its propeptide: Use of overlapping peptides to identify a critical segment, *FEBS Lett. 392*, 233−236.

BI047590O