J|A|C|S
ARTICLES
Published on Web 01/18/2003

# Synthesis and NMR Structure of P41icf, a Potent Inhibitor of Human Cathepsin L

Cristina Chiva,[‡] Philippe Barthe,[‡] Anna Codina,[‡] Margarida Gairí,[§] Franck Molina,[||] Claude Granier,[||] Martine Pugnière,[||] Tatsuya Inui,[⊥] Hideki Nishio,[⊥] Yuji Nishiuchi,[⊥] Terutoshi Kimura,[⊥] Shumpei Sakakibara,[⊥] Fernando Albericio,[†,‡] and Ernest Giralt*[,†,‡]

*Contribution from the Institut de Recerca Biomèdica de Barcelona, Parc Científic de Barcelona, Josep Samitier 1, E-08028 Barcelona, Spain, Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain, Unitat de RMN d'Alt Camp, SCT, Universitat de Barcelona, Josep Samitier 1, E-08028 Barcelona, Spain, Faculté de Pharmacie IBPh, CNRS UMR 5094, 15 Av. Charles Flahault, 34093 Montpellier Cedex 5, France, and Peptide Institute Inc., 4-1-2 Ina, Minoh-shi, Osaka 562-8686, Japan*

Received June 3, 2002; E-mail: egiralt@qo.ub.es

***Abstract:*** The total synthesis and structural characterization of the MHCII-associated p41 invariant chain fragment (P41icf) is described. P41icf plays a crucial role in the maturation of MHC class II molecules and antigen processing, acting as a highly selective cathepsin L inhibitor. P41icf synthesis was achieved using a combined solid-phase/solution approach. The entire molecule (65 residues, 7246 Da unprotected) was assembled in solution from fully protected peptides in the size range of 10 residues. After deprotection, oxidative folding in carefully adjusted experimental conditions led to the completely folded and functional P41icf with a disulfide pairing identical to that of native P41icf. CD, NMR, and surface plasmon resonance (SPR) were used for the structural and functional characterization of synthetic P41icf. CD thermal denaturation showed clear cooperative behavior. Tight cathepsin L binding was demonstrated by SPR. $^1$H NMR spectroscopy at 800 MHz of unlabeled P41icf was used to solve the three-dimensional structure of the molecule. P41icf behaves as a well-folded protein domain with a topology very close to the crystallographic cathepsin L-bound form.

## Introduction

Proteolysis is strictly regulated in all living organisms. Proof of this precise regulation in humans comes from the discovery by Bevec et al.[1] of a small protein domain which acts as a common "cornerstone" in two very important but previously unrelated biological processes, antigen presentation to the immune system and cathepsin L inhibition. This domain is known as the MHCII-associated p41 invariant chain fragment (P41icf). P41icf is a member of the family of thyroglobulin-like domains.[2,3] In 1992, Peterson and Miller reported its crucial function of enhancing the antigen presentation activity of MHCII molecules,[4,5] and four years later, Bevec et al. reported that P41icf was a strong cathepsin L inhibitor.[1] Recently, the crystal structure of the P41icf bound to cathepsin L has been reported.[6] Human cathepsin L is involved in a number of disease processes, such as bone resorption,[7] pathogenesis of rheumatoid arthritis,[8] and tumor invasion and metastasis[9] and is recognized as a key target for therapeutic intervention.[10−12]

Despite its moderate size (see amino acid sequence in Figure 1A), all previous attempts to obtain P41icf either by chemical synthesis or via recombinant DNA were unsuccessful.[13] P41icf was only accessible through isolation from human kidneys as a complex with cathepsin L. This lack of availability has left several questions unanswered: (i) Does P41icf behave as an autonomous protein domain, and does it fold with the native disulfide pairing in the absence of the rest of the P41 invariant chain? (ii) Does P41icf have a well-defined 3D structure in the absence of cathepsin L? And, (iii) is the eventual P41icf 3D structure similar to that of cathepsin L-bound, or on the contrary, is there a large amount of induced fit during complexation?

[†] Parc Científic de Barcelona.
[‡] Departament de Química Orgànica, Universitat de Barcelona.
[§] Unitat de RMN d'Alt Camp, SCT, Universitat de Barcelona.
[||] Faculté de Pharmacie IBPh.
[⊥] Peptide Institute Inc.

(1) Bevec, T.; Stoka, V.; Pungercic, G.; Dolenc, I.; Turk, V. *J. Exp. Med.* **1996**, *183*, 1331.
(2) Molina, F.; Pau, B.; Granier, C. *FEBS Lett.* **1996**, *391*, 229.
(3) Molina, F.; Bouanani, M.; Pau, B.; Granier, C. *Eur. J. Biochem* **1996**, *240*, 125.
(4) Peterson, M.; Miller, J. *Nature* **1992**, *357*, 596.
(5) Cresswell, P. *Cell* **1996**, *84*, 505.
(6) Guncar, G.; Pungercic, G.; Klemencic, I.; Turk, V.; Turk, D. *EMBO J.* **1999**, *18*, 793.
(7) Millest, A. J.; Breen, S. A.; Loveday, B. E.; Clarkson, P. N.; Simpson, C. A.; Waterton, J. C.; Johnstone, D. *Bone* **1997**, *20*, 465.
(8) Esser, R. E.; Angelo, R. A.; Murphey, M. D.; Watts, L. M.; Thornburg, L. P.; Palmer, J. T.; Talhouk, J.; Smith, R. E. *Arthritis Rheum.* **1994**, *37*, 236.
(9) Yagel, S.; Warner, A. H.; Nellans, H. N.; Lala, P. K.; Waghorne, C.; Denhardt, D. T. *Cancer Res.* **1989**, *49*, 3553.
(10) Brinker, A.; Weber, E.; Stoll, D.; Voigt, J.; Müller, A.; Sewald, N.; Jung, G.; Wiesmüller, K. H.; Bohley, P. *Eur. J. Biochem.* **2000**, *267*, 5085.
(11) McGrath, M. E. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 181.
(12) Turk, B.; Turk, D.; Turk, V. *Biochim. Biophys. Acta* **2000**, *1477*, 98.
(13) Chiva, C. PhD Thesis, University of Barcelona, Barcelona, Spain, 2001.


**10.1021/ja0207908 CCC: $25.00** 

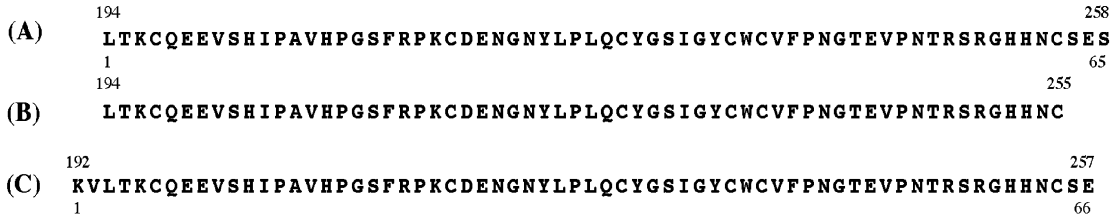


***Figure 1.*** (A) Amino acid sequence of P41icf as found in the human kidney[6] (above, p41 numbering; below, P41icf numbering); (B) amino acid sequence of P41icf, identified as a thyroglobulin-like domain[5] (p41 numbering); (C) amino acid sequence of elongated P41icf (above, p41 numbering; below, P41icf numbering).

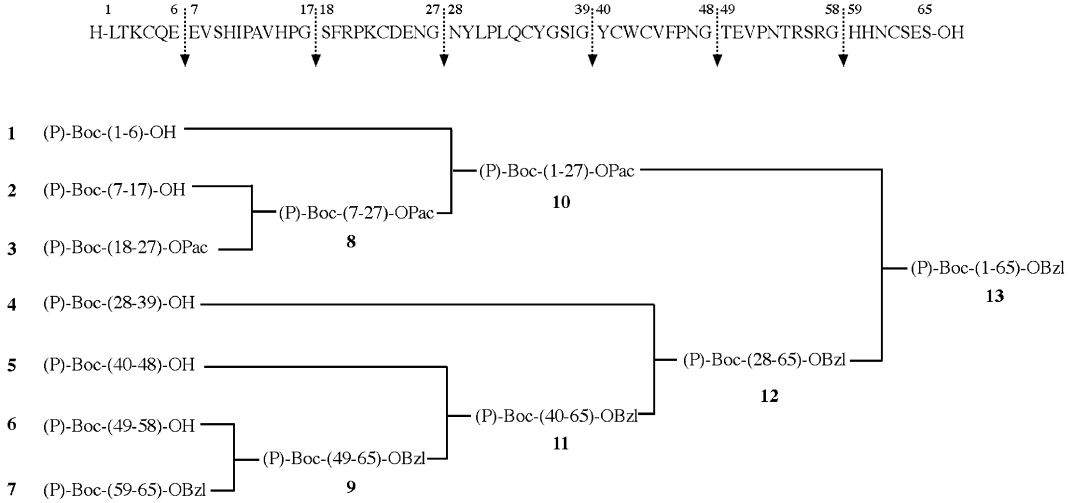


***Figure 2.*** Coupling strategy for the synthesis of protected P41icf.

Here, we report the first total synthesis of P41icf using a convergent approach based on solution coupling of protected peptide segments obtained by solid-phase synthesis[14,15] and the determination of the three-dimensional solution structure of the free molecule by $^1$H NMR spectroscopy at 800 MHz. This process allowed us to provide appropriate answers to these questions.

## Results

**Peptide Synthesis.** The P41icf sequence identified as a thyroglobulin-like domain[3] is shown in Figure 1B. Our initial synthetic attempts were focused on this sequence elongated with four residues, two at the C-terminus and two at the N-terminus, from an entire p41 (Figure 1C). Several attempts at stepwise synthesis, using either Boc or Fmoc for temporary protection of the $N^{\alpha}$-amino group, were carried out (data not shown). The low yield of the stepwise strategy (<1%) and the significant amount of sample required for the NMR study made this approach inadequate.

Next, a convergent solid-phase strategy was attempted. In the retrosynthetic analysis, the 66-residue peptide was divided into three protected peptides, corresponding to the 1−14, 15−29, and 30−41 segments of the protein, in addition to the peptide-resin H-(42−66)-AM-PEG-PS. The protected fragments were obtained in high yield and purity using the Fmoc/$^t$Bu strategy of solid-phase peptide synthesis. A highly acidic labile resin Cl−ClTrt−PS[16] with a limited incorporation of the first amino acid[17] was used for the preparation of the protected peptides. Several coupling conditions were tested for the incorporation of the protected peptide 30−41 onto the peptide-resin H-(42−66)-AM-PEG-PS. Conditions assayed involved the use of various coupling reagents, including aza compounds, solvents disrupting peptide chains aggregation, and high-temperature couplings. However, good conditions for coupling the protected peptide to the resin could not be found.

Finally, a convergent strategy using a combined solid-phase and solution approach[18] was successfully applied to the synthesis of P41icf as found in the human kidney (Figure 1A). The final protected peptide of 65 amino acids was obtained in solution by assembling seven protected peptides (Figure 2). This approach requires having Gly or Pro, which cannot racemize, as C-terminal residues in all fragments to be coupled. All fragments of P41icf have a Gly residue at the C-terminal position with the exception of **1.** In this case, it was preferred to choose a shorter fragment with Glu in the C-terminal than a longer one with either Gly or Pro in this position. The C-terminus of **3** was temporarily protected with the phenacyl (Pac) ester and the C-terminus of **7** was permanently protected with the benzyl (Bzl) ester. In this approach, a Boc/Bzl protection scheme and the base labile handle *N*-[9-(hydroxymethyl)-2-fluorenyl] succinamic acid[19] (HMFS) were used for the synthesis of protected

(14) Lloyd-Williams, P.; Albericio, F.; Giralt, E. *Chemical Approaches to the Synthesis of Peptides and Proteins*. CRC Press: Boca Raton, 1997.
(15) Nishiuchi, Y.; Inui, T,; Nishio, H.; Bódi, J.; Kimura, T.; Tsuji, F. I.; Sakakibara, S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13549.
(16) (a) Barlos, K.; Gatos, D.; Kallitsis, J.; Papaphotiu, G.; Sotiriu, P.; Wenqing, Y.; Schäfer, W. *Tetrahedron Lett.* **1989**, *30*, 3943. (b) Barlos, K.; Gatos, D.; Kapolos, S.; Papaphotiu, G.; Schäfer, W.; Wenqing, Y. *Tetrahedron Lett.* **1989**, *30*, 3947. (c) Barlos, K.; Chatzi, O.; Gatos, D.; Stavropoulos, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 513.
(17) Chiva, C.; Vilaseca, M.; Giralt, E.; Albericio, F. *J. Pept. Sci.* **1999**, *5*, 131.
(18) Nishiuchi, Y.; Nishio, H.; Inui, T.; Bódi, J.; Kimura, T. *J. Pept. Sci.* **2000**, *6*, 84.

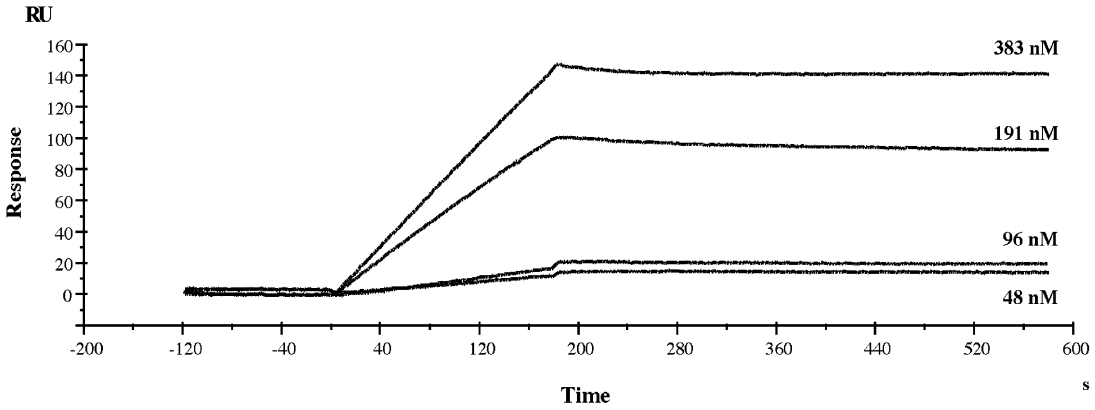


***Figure 3.*** Sensogram for the interaction between immobilized P41icf and cathepsin L at four different concentrations.

peptides. The side-chains protecting groups for Tyr and Trp were Pen[20] and Hoc,[21] respectively, which are completely stable in the basic conditions (20% morpholine in DMF) necessary to detach the peptide from the resin; other acid-labile/basic-stable groups were Tos for Arg, cHxl for Asp and Glu, Bom for His, ClZ for Lys, and Bzl for Ser and Thr. All these protecting groups were removed with anhydrous HF. Thiol groups of the six Cys residues were protected with the Acm group, which is stable to HF and can be removed by $I_2$ or $Hg^{2+}$. For side-chain protection of Asn and Gln, the Xan group was used, which prevents dehydration to the corresponding nitrile during the coupling step and is eliminated by TFA during the Boc removal. Each protected peptide was purified by recrystallization, reprecipitation, or silica gel chromatography using appropriate solvents. Homogeneity was checked by amino acid analysis, TLC, MS (ESI), and RP-HPLC. All fragments showed purity higher than 95%.

Fragment coupling reactions were carried out according to the scheme shown in Figure 2, using EDC as coupling reagent in the presence of HOBt or HOOBt and DMF or NMP as solvent. This scheme allowed us to obtain 2.4 g of the fully protected peptide **13**, which represents an 11% yield based on the nominal loading of the initial resin used for the preparation of each fragment. The degree of purity of this peptide, around 65% by RP-HPLC, is extremely good considering its size.

The final protected peptide thus obtained was treated with HF/anisole (9:1) in the presence of 20 equiv of HCl·H−Cys−OH·$H_2O$ to suppress the side reaction associated with the $N^{im}$-benzyloxymethyl group.[22] The crude product was purified by RP-HPLC, and the (6Acm)-peptide thus obtained was treated with $Hg(OAc)_2$ in 50% aqueous AcOH at room temperature for 1 h to remove the Acm groups.[23] The Hg ions were removed by adding $\beta$-mercaptoethanol, followed by gel filtration on Sephadex G-25.

Several conditions of oxidative folding of the reduced peptide were tried. Best results were obtained when the reaction was done in 0.1 M $AcONH_4$ buffer (pH 7.8) containing 1mM EDTA and 0.5 M Gu·HCl at a peptide concentration of $7 \times 10^{-6}$ M in the presence of reduced and oxidized glutathione (GSH/GSSG) for 3 days at room temperature. Monomeric oxidized P41icf was isolated by RP-HPLC. Analysis of the product by analytical RP-HPLC using gradient conditions showed a single peak, but, when analyzed in isocratic conditions, it showed two peaks with a ratio of 5:1. RP-HPLC purification in the same isocratic conditions at 40 °C allowed the separation of the two disulfide isomers. When the preparative chromatography was carried out at room temperature, the two products coeluted.

Disulfide bridges in every isomer were identified after digestion with chymotrypsin. The proteolytic crude corresponding to the major isomer was analyzed by RP-HPLC coupled to MS (ESI), the peaks obtained were isolated, and their amino acid sequences were determined by Edman sequencing. Three disulfide bonds connecting $Cys^4$-$Cys^{23}$, $Cys^{34}$-$Cys^{41}$ and $Cys^{43}$-$Cys^{62}$ were identified. This pattern of disulfide bonds coincides with that reported for the P41icf bound to cathepsin L.[6] A similar analysis by RP-HPLC coupled to MS (ESI) of the proteolytic crude of the minor isomer showed different connectivities of the three disulfide bonds: $Cys^4$-$Cys^{23}$, $Cys^{34}$-$Cys^{43}$, and $Cys^{41}$-$Cys^{62}$.

**Binding Assay.** The affinity of synthetic P41icf to cathepsin L was assayed by surface plasmon resonance. The synthetic peptide was immobilized on the sensor chip, and solutions of different concentrations of human cathepsin L were passed through it (Figure 3). The formation of a high-affinity complex characterized by an equilibrium dissociation constant ($K_D$) of $4.9 \times 10^{-9}$ M was detected. Furthermore, when the same experiment was done using cathepsin B instead of cathepsin L, no binding was detected, which demonstrated the specificity of the interaction between cathepsin L and immobilized P41icf.

**CD Spectroscopy.** The circular dichroism (CD) spectra of synthetic P41icf showed a strong positive band at 226 nm, which could not be assigned to any of the most common secondary structures in peptides or proteins and was attributed to an aromatic side-chain contribution to the far-ultraviolet CD spectra.[24] This phenomenon has been described for a number of proteins where the $\alpha$-helix content is low and for all-$\beta$ proteins.[25]

The thermal unfolding curves of the synthetic peptide measured by CD in the far and near UV showed a sigmoid shape

(19) (a) Rabanal, F.; Giralt, E.; Albericio, F. *Tetrahedron Lett.* **1992**, *33*, 1775. (b) Rabanal, F.; Giralt, E.; Albericio, F. *Tetrahedron* **1995**, *51*, 1449.
(20) Bódi, J.; Nishiuchi, Y.; Nishio, H.; Inui, T.; Kimura, T. *Tetrahedron Lett.* **1998**, *39*, 7117.
(21) Nishiuchi, Y.; Nishio, H.; Inui, T.; Kimura, T.; Sakakibara, S. *Tetrahedron Lett.* **1996**, *37*, 7529.
(22) Kumagaye, K. Y.; Inui, T.; Nakajima, K.; Kimura, T.; Sakakibara, S. *Pept. Res.* **1991**, *4*, 84.
(23) Nishio, H.; Kimura, T.; Sakakibara, S. *Tetrahedron Lett.* **1994**, *35*, 1239.
(24) Woody, R. W. *Biopolymers* **1978**, *17*, 1451.
(25) Woody R. W. In *Circular dichroism. Principles and Applications;* Nakanishi, K., Berova, N., Woody, R. W., Eds.; VCH Publishers Inc.: New York, USA, 1994; p 473.

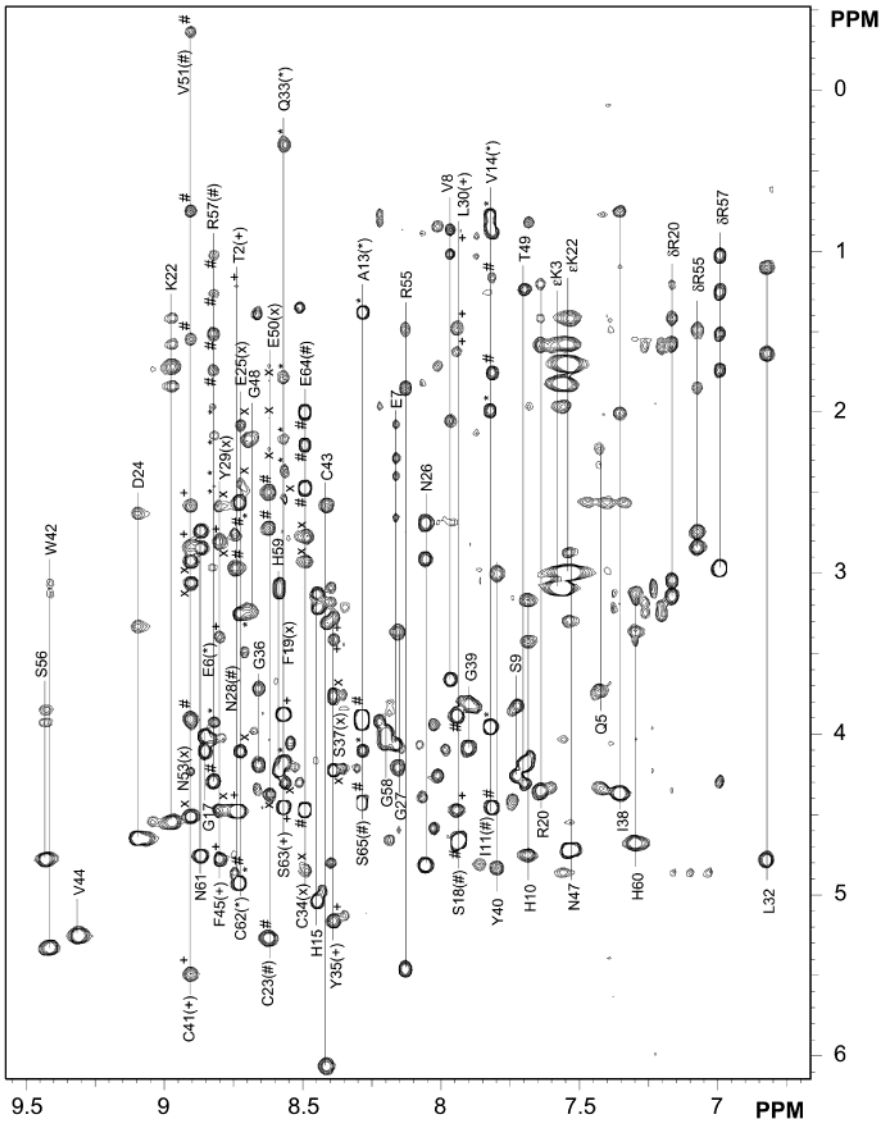


***Figure 4.*** Portion of a TOCSY spectrum of P41icf (1 mM, 15%$D_2O/H_2O$, pH 5.7) recorded at 800 MHz and 288 K.

where the transition between the two states occurs in a narrow range of temperature. An experiment of Gu·HCl-induced unfolding showed a cooperativity profile during denaturation. These CD spectra are provided as Supporting Information.

**NMR Experiments.** Nearly complete sequence-specific $^1$H resonance assignments for P41icf were obtained from 2D homonuclear NMR spectroscopy, using a combination of DQF-COSY, TOCSY, and NOESY experiments.[26] 2D spectra were recorded at 5, 15, and 25 °C, which helped to solve residual overlap problems in the assignment of the proton resonances. A portion of a TOCSY spectrum at 15 °C is shown in Figure 4. The degree of $^1$H resonance dispersion in the 2D homonuclear spectra at 800 MHz was sufficient to enable the identification of a continuous stretch of sequential NOEs over the whole sequence, which allowed us to accomplish sequential assignment. Therefore, heteronuclear NMR experiments, which would have required $^{15}$N/$^{13}$C labeling of the protein, were unnecessary. The presence of strong sequential $d_{\alpha\delta(i,i+1)}$ NOEs in the six Xxx-Pro dipeptides unambiguously established the trans geometry of the peptide bonds involved.

The presence of a large stretch of medium-range NOE contacts ($i$, $\leq i+4$) in the N-terminal residues suggested the presence of an α-helix. This was supported by the inspection of the αH chemical shifts, with $\delta$ values shifted upfield from random-coil values,[27] and by the low $^3J_{NH\alpha}$ coupling constant values indicating an α-helix encompassing residues 3−10. Chemical shift index (CSI) values indicated the presence of two $\beta$-strands spanning residues 32−35 and 40−44 supported by large $^3J_{NH\alpha}$ coupling constant values and strong $d_{\alpha N(i,i+1)}$ NOE contacts. The pattern of numerous backbone−backbone NOE contacts connecting different peptide segments (Tyr35-Pro31, Tyr40-Val44, and Arg57-Thr54) revealed the presence of an antiparallel $\beta$-sheet, consisting of three strands.

**Structure Calculation and Refinement.** A total of 981 conformational constraints (distance and dihedral) were used in the distance−geometry structure calculations (see Experimental Section). Disulfide bonds were treated as additional distance restraints. Of the final 600 structures, 30 had a final target function smaller than 0.70 Å$^2$ and satisfied all experimental constraints with no NOE distance constraint violation

(26) Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley and Sons*:* New York, 1986.

(27) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67.

***Table 1.*** Experimental Constraints and Refinement Statistics of the 30 Conformers Representing the Solution Structure of P41icf before and after Molecular Mechanics Energy Refinement

| | | |
|---|---|---|
| distance constraints | | |
| intraresidue | 255 | |
| sequential | 257 | |
| medium range | 141 | |
| long range | 272 | |
| disulfide bonds | 3 | |
| constraints per residue | 24.5 | |
| dihedral constraints | | |
| PHI (Φ) | 28 | |
| CHI1 ($\chi_1$) | 28 | |
| parameter | DYANA | DYANA + AMBER |
| target function (Å$^2$) | 0.63 ± 0.05 | |
| upper limit violations | | |
| number > 0.2 Å | 0 | 0 |
| sum of violations (Å) | 3.40 ± 0.35 | 1.33 ± 0.26 |
| maximum violations (Å) | 0.18 ± 0.01 | 0.18 ± 0.07 |
| dihedral angle violations | | |
| number > 5° | 0 | 0 |
| sum of violations (°) | 0.07 ± 0.05 | 6.79 ± 6.60 |
| maximum violations (°) | 0.02 ± 0.01 | 4.22 ± 0.82 |
| van der Waals violations | | |
| number > 0.2 Å | 0 | |
| sum of violations (Å) | 2.20 ± 0.27 | |
| maximum violations (Å) | 0.18 ± 0.01 | |
| AMBER energies (kcal mol$^{-1}$) | | |
| bond energy | | 30.6 ± 0.8 |
| valence angle energy | | 184.6 ± 3.7 |
| van der Waals energy | | −429.7 ± 9.0 |
| electrostatic energy | | −1655.2 ± 17.9 |
| constraint enrgy | | 8.9 ± 1.8 |
| total nonbonding energy | | −1160.8 ± 12.6 |
| total energy | | −644.7 ± 11.6 |
| PROCHECK values (%) for residues | | |
| in most favored regions | 64.4 | 78.5 |
| in additional allowed regions | 32.8 | 20.7 |
| in generously allowed regions | 2.7 | 0.8 |
| in disallowed regions | 0.1 | 0.0 |
| RMSD values (Å) | | |
| residues | | |
| 1−65 (BA[a]/HA[b]) | 1.43 ± 0.32/ 1.78 ± 0.28 | 1.51 ± 0.41/ 1.91 ± 0.34 |
| 1−62 (BA[a]/HA[b]) | 1.09 ± 0.30/ 1.41 ± 0.25 | 1.07 ± 0.34/ 1.53 ± 0.28 |

[a] Backbone atoms. [b] All heavy atoms.

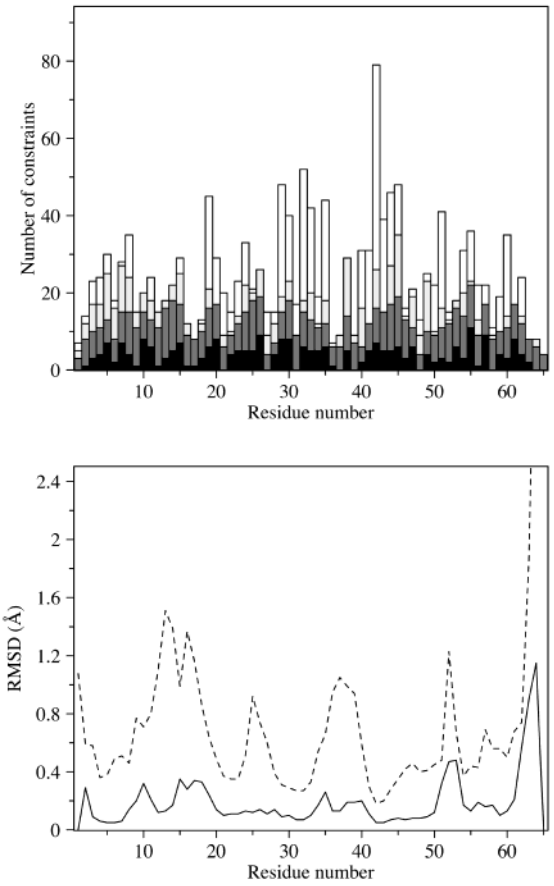


***Figure 5.*** (A) Number of distance constraints derived from observed NOE cross peaks used in the structure calculation of P41icf. Intraresidue (black), sequential (dark gray), medium-range (light gray), and long-range (white) NOE constraints combined for each residue are shown. (B) Average RMSDs of coordinates for backbone atoms (N, Cα, C') (broken line) superimposed for each residue in 30 structures and average RMSDs for the backbone superposition of all tripeptide segments along the sequence (solid line). The RMSD values for the tripeptide segments are plotted at the position of the central residue.

exceeding 0.22 Å and no dihedral angle violation exceeding 0.03°. Only those 30 structures with a mean target function equal to 0.63 ± 0.05 Å$^2$ and with mean maximum distance and dihedral violations lower than 0.2 Å and 5°, respectively, were chosen for further refinement. The surveys of the structural statistics and the residual violations of experimental constraints for these 30 conformers are shown in Table 1. When the C-terminal residues (63−65) were discarded, for which no long-range NOEs were available, the low average pairwise RMSD (root-mean-square deviation) value for all heavy atoms (1.41 ± 0.25 Å) indicates that all conformers converge to the same overall fold. Thus, they were considered representative of the conformational space consistent with the data. These 30 DYANA conformers with low target function value were further refined by simulated annealing and restrained energy minimization using the AMBER force field.[28] As expected, the refinement resulted in a substantial improvement in van der Waals energy and of the electrostatic contribution to the energy. The decrease in nonbonding energy was achieved without a significant increase in the residual violations of the experimental constraints, as compared with the case of DYANA conformers (Table 1). This fact, in conjunction with the high number of NOEs per residue (Figure 5A), supports the reliability of the calculated structures. Perhaps the most important result of the refinement was the marked improvement in the geometry of most hydrogen bonds as a result of the electrostatic term in the force field. Altogether, the refined structures provide a better estimate of the conformational space, consistent with the experimental constraints, than those obtained from DYANA alone. The quality of the 30 refined structures was checked with a Ramachandran plot of $\phi$ and $\psi$ torsion angles (see Supporting Information) using the PROCHECK[29] program. The protein contains 51 "meaningful" residues (i.e., nonglycine, nonproline, and nonterminal residues) for the Ramachandran plot. Of all the 30 refined structures, 78.5% of residues fall into the most favorable region of the Ramachandran plot, 20.7% are in the additionally allowed regions, 0.8% in generously allowed regions (essentially the last two residues, Ser63 and Glu64), and no residue falls in the disallowed region.

(28) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.

(29) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283.

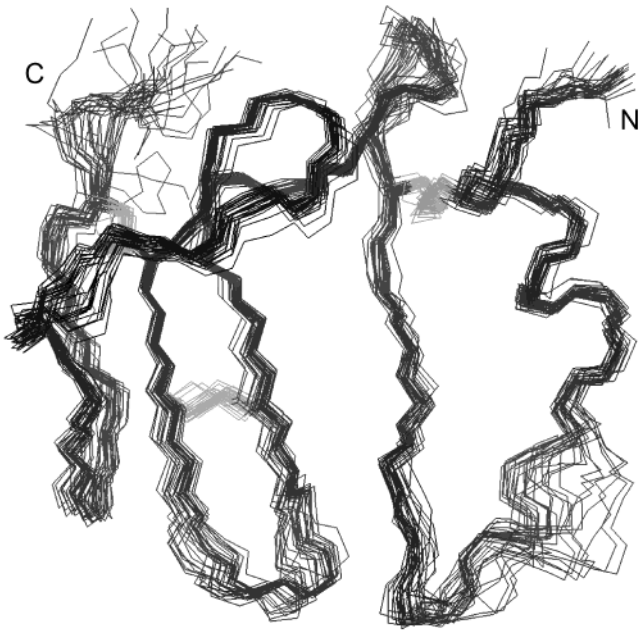


***Figure 6.*** Superposition of the 30 refined structures of P41icf over the backbone heavy atoms for residues 1−62. Only the backbone is shown, except for the three disulfide bonds (4−23, 34−41, and 43−62) (shaded gray).

The P41icf solution structure consists of two subdomains (Figure 6). The N-terminal subdomain (residues 1 to 28) is characterized by a short α-helix and a $\beta$-strand, stabilized by a well-defined disulfide bond Cys4-Cys23, which adopts a type-3 conformation, according to Srinivasan et al.[30] The C-terminal subdomain (residues 29−65) consists of several $\beta$-strands stabilized by two disulfide bonds (Cys34-Cys41 and Cys43-Cys62). The P41icf solution structure described previously is quite similar to the crystal structure of the P41icf−cathepsin L complex reported by Guncar et al.[6] The differences between the apo and the holo form are discussed later.

## Discussion

Although peptide synthesis has evolved enormously during the last years, efforts have been put mainly in the synthesis of complex but small peptides, such as those presenting cyclic topologies, labile components (e.g., glyco-, phospho-, nucleopeptides), and/or non natural amino acids.[14] Thus, synthesis of peptides of more than 50−60 residues containing several Cys constitutes still nowadays a real challenge, and therefore, there is not a method of choice for their preparation.

A solid-phase stepwise strategy has been widely used for the synthesis of large peptides and small proteins. Nevertheless, some amino acid sequences are difficult to synthesize, owing to the aggregation of peptide chains during elongation. The physicochemical basis of the aggregation phenomena of these "difficult sequences" is not well understood and is still under discussion. In the case of P41icf, the application of a stepwise synthetic strategy led to a complex mixture, from which the target molecule was isolated in low yield. A convergent solid-phase synthesis was also attempted, but the strategy was unsuccessful because of problems during the coupling of protected peptide fragments, again associated with the aggregation of the growing peptide chains. Several solutions to the aggregation problem have been proposed, such as the use of Hmb[31] as a protecting group for the amide bond or the use of pseudoprolines[32] as secondary structure disrupters. However, these strategies are far from being a general solution to the aggregation problem.

Alternatively, the chemical ligation approach,[33] which has proven to be an effective method for the synthesis of some of these peptides, presents some sequence requirements and difficulties in the synthesis and handling of the thioester peptide fragments. Furthermore, classical solution methods are tedious and very time-consuming.[34] Thus, a combined solid-phase/solution approach can be an attractive alternative for the synthesis of these peptides[18] because it combines the advantages of both solid-phase and solution strategies, that is, the speed and efficiency in the solid-phase preparation of the protected segments and higher yields in the coupling of these fragments in solution. This method has been successfully applied by the Osaka group to the synthesis of the 238-residue green fluorescent protein (GFP),[15] the longest protein chemically synthesized to date. It consists of the solution assembly of the protected peptides previously obtained by solid-phase synthesis using the HMFS handle.[19]

The combined solid-phase/solution approach was successfully applied to the synthesis of P41icf. The coupling of the fragments proceeded smoothly using stoichiometric amounts of the corresponding protected peptides. The fully protected peptide obtained had very good purity and an acceptable yield.

The best folding conditions for P41icf were the use of the redox pair GSH/GSSG at a peptide concentration of $7 \times 10^{-6}$ M and 0.5 M Gu·HCl. The use of a denaturing reagent during folding reactions is recommended in order to control aggregation.[35] The addition of a redox buffer enables physiological conditions to be mimicked; the reaction proceeds via a thiol−disulfide exchange mechanism. Thus, even if non-native disulfide bridges were kinetically favored, they would equilibrate to the thermodynamically most favored arrangement, which usually corresponds to the native structure. In the folding experiment, two disulfide isomers in a 5:1 ratio were found. Preparative RP-HPLC under isocratic conditions at high temperature (40 °C) was necessary to achieve the separation of the two isomers. At high temperature, the efficiency of the stationary phase is magnified, mobile phase viscosity decreases, and the equilibrium of the interaction between the sample and the stationary phase is achieved sooner. Moreover, at 40 °C, the minor isomer, a priori less stable to thermal denaturation, could undergo a conformational change, modifying its interaction with the stationary phase and therefore its retention time. This would result in a better separation of the two isomers.

The major isomer corresponds to the native structure, and the minor isomer derives from the scrambling of the disulfide

(30) Srinivasan, N.; Sowdhamini, R.; Ramakrishnan, C.; Balaram, P. *Int. J. Pept. Protein Res.* **1990**, *36*, 147.
(31) Johnson, T.; Quibell, M.; Sheppard, R. C. *J. Pept. Sci.* **1995**, *1*, 11.
(32) Wöhr, T.; Wahl, F.; Nefzi, A.; Rohwedder, B.; Sato, T.; Sun, X.; Mutter, M. *J. Am. Chem. Soc.* **1996**, *118*, 9218.
(33) Dawson, P. *Synthesis of Peptides and Peptidomimetics*; Goodman, M., Felix, A., Moroder, L., Toniolo, C., Eds.; Houben−Weyl. Methods of Organic Chemistry, Vol. E22a; Thieme: Stuttgart, Germany and New York, 2002; pp 627−641.
(34) Okada, Y. *Synthesis of Peptides and Peptidomimetics*; Goodman, M., Felix, A., Moroder, L., Toniolo, C., Eds.; Houben−Weyl. Methods of Organic Chemistry, Vol. E22a; Thieme: Stuttgart, Germany and New York, 2002; pp 591−626.
(35) Wetlaufer, D. B.; Xie, Y. *Protein Sci.* **1995**, *4*, 1535.

bridges in the C-terminal subdomain, which gives rise to the new pairing: $Cys^{34}$-$Cys^{43}$ and $Cys^{41}$-$Cys^{62}$. In the P41icf NMR structure, the distances between sulfur atoms of these pairs of Cys are 9.6 and 8.6 Å, respectively. In the minor isomer structure, a rearrangement of the $\beta$-sheet would probably be necessary, to shorten the distance between the sulfur atoms involved in the new pattern of disulfide bonds.

The affinity assays with the synthetic peptide showed high affinity between immobilized P41icf and cathepsin L, demonstrating the functional folding of the molecule. The differences between our calculated $K_D$ and the reported[1] $K_D$ ($10^{-9}$ vs $10^{-12}$) could be due to different experimental conditions for the two assays. The main difference is that P41icf is immobilized on a solid surface in our experiment which might distort in some way the activity of the molecule. There is also a glycosylation point in the natural molecule that is not present in the synthetic peptide. Furthermore, the two experiments were performed in different pH conditions.[36] No binding was detected when the assay was performed using cathepsin B, another lysosomal cysteine protease that presents high homology with cathepsin L. This demonstrated the high specificity of the interaction between P41icf and cathepsin L.

Thermal and chemically induced denaturation experiments followed by CD showed the cooperative character of the interactions stabilizing the tertiary structure of P41icf. Thermal denaturation showed that the structure is stable up to 35 °C and, consequently, the NMR structure calculated from the 15 °C data represents the stable structure of P41icf. Chemically induced denaturation showed that the structure is stable up to 2.5 M Gu·HCl, clearly above the concentration of Gu·HCl (0.5 M) in the folding conditions.

The NMR structure of P41icf reveals a well-folded domain, structured in solution, in the absence of the rest of the invariant chain. Analysis of the local backbone displacements calculated from the 30 energy-refined conformers (Figure 5B) indicates that the P41icf structure is well defined from residues 2−61 (average pairwise local RMS deviation: 0.16 Å ± 0.10 Å) but has certain less well-defined regions: the N- and C-terminus of the helix, the three turns, involving residues 13−17, 36−39, and 57−62, the dipeptide 52−53, and the C-terminal part of the protein.

The first turn (residues 13−17) remains ill-defined in the NMR structure. This is related to the scarcity of long-range NOEs involving these residues, because of spectral overlap or of real flexibility of this part of the molecule. Despite the acquisition of the spectra at very high field (800 MHz) and at different temperatures, we could not avoid ambiguities in a few NOEs, which hampered their use as structural restraints in the structure calculation. In addition, in the TOCSY spectra, some weak extra peaks, which can be assigned to three consecutive spin systems, Ala13-Val14-His15, were observed. The presence of this set of minor signals was attributed to the existence of one or more minor conformations. Heterogeneity of this turn could result from cis−trans isomerization of the flanking

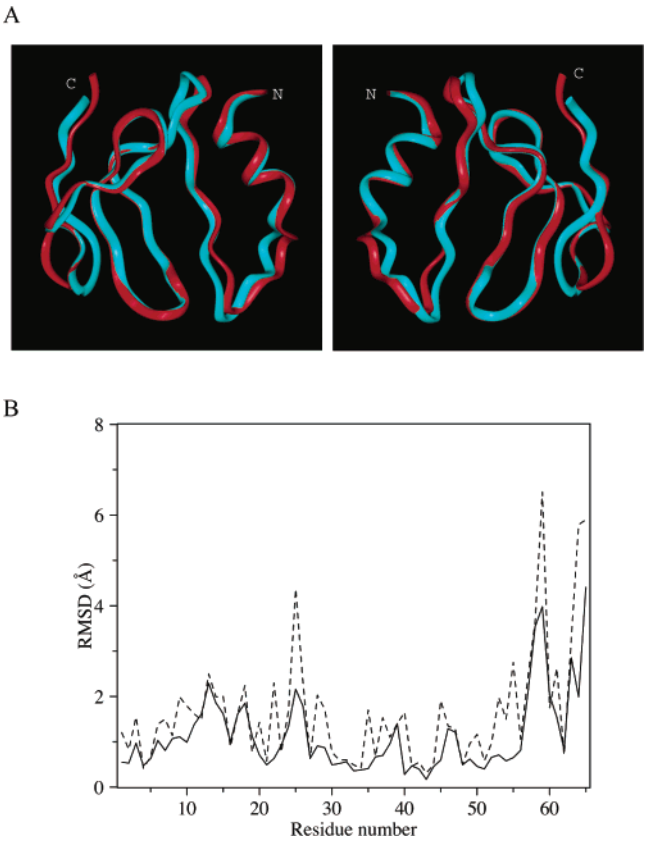


***Figure 7.*** (A) Superposition of the ribbon plots of the average NMR structure of free P41icf in solution (cyan) and the X-ray structure (PDB code 1icf) of the protein complexed with cathepsin L (red). The two views are related by a 90° rotation about the vertical axis. (B) RMSDs of coordinates between the average NMR structure and the X-ray structure superposed over residues 1−62, for backbone atoms (solid line) and for side-chain heavy atoms (broken line)

residues, Pro12 and Pro16, although the characteristic $d_{\alpha\alpha(i,i+1)}$ NOE of a cis conformation in a dipeptide Xxx-Pro[26] was not observed.

Superposition of the backbone heavy atoms (residues 1−62) of the last 30 refined structures highlights the less well-defined regions in the structure (Figure 5B): as well as the mentioned turn (residues 13−17), two other loops (around residues 25 and 47, respectively) show different orientations in the 30 conformers.

When superimposing the average NMR structure and the X-ray structure described by Guncar et al.,[6] we observed a high general similarity between the folding of the free protein in solution and the structure of the protein complexed with cathepsin L (Figure 7A). This is supported by the low RMSDs of coordinates of both structures over residues 1−62 (1.25 Å for the backbone, 1.80 Å for all heavy atoms). The main differences are found on turns (residues 13−17, 36−39, and 57) involved in the interaction with the enzyme[6] and in exposed loops which appear flexible in the NMR structures (Figure 7B). Natural P41icf is glycosylated on the Asn47,[6] whereas synthetic P41icf is not. The backbone RMSD for residues 46−47 in the superimposed average NMR and X-ray structures (1.25 ± 0.06 Å) is smaller than the backbone RMSD for the loop involving residues 24−26 (1.76 ± 0.42 Å), suggesting that the *N*-acetyl-D-glycosamine on Asn47 does not induce any structural modification.

(36) Cathepsin L activity is very pH sensitive (Mason, R. W.; Green, G. G.; Barrett, A. J. *Biochem. J.* **1985**, *226*, 233−241.). Bevec and co-workers, in their experiment, performed the assay under more basic conditions than we did. Under these conditions, the cathepsin could be less active and, as they followed the inactivation of the functional activity of cathepsin L using a subtract (and not the direct binding of the two molecules), a loss in the optimal activity of the enzyme should result in a greater dissociation constant.

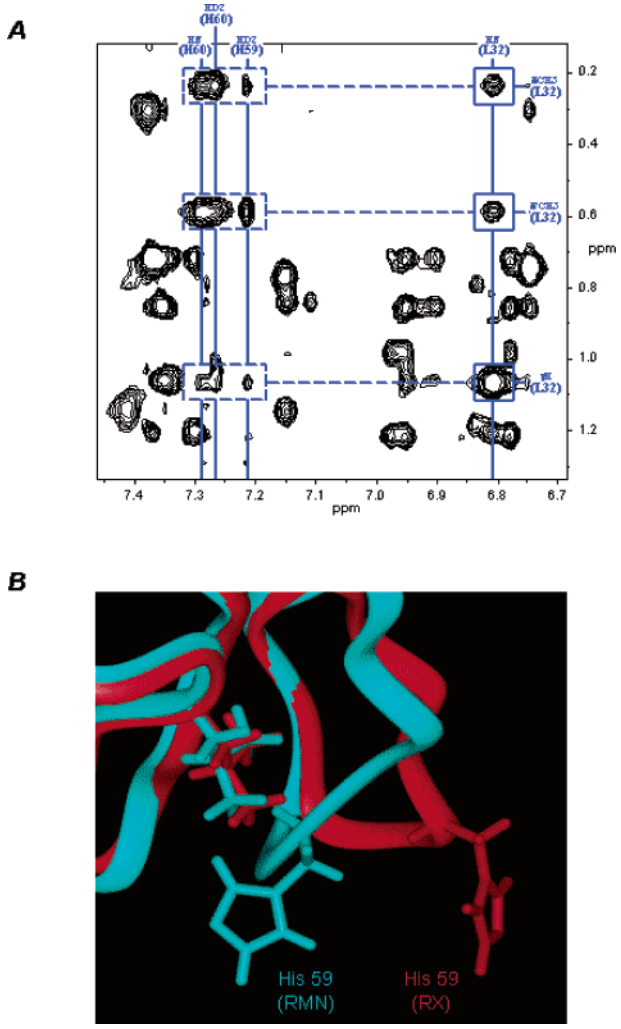


***Figure 8.*** (A) Region of a NOESY spectrum showing NOEs connecting side-chain protons of His59, His60, and Leu32. (B) Detail of the superposition of the average NMR structure (blue) and the X-ray structure (red) of P41icf, showing a significant difference between the His59 side-chain orientation in the two structures.

The largest backbone RMSD (3.98 Å) is found in the third turn, around His59, which is involved in the interaction with cathepsin L. Figure 8B shows a detail of the side-chain orientation of this residue in both structures. The NMR structure reveals the proximity of the side-chain protons of Leu32 with those of both His59 and His60, a consequence of the proton NOEs involving these residues (Figure 8A). This suggests that the loop is "closed" and His59 protects disulfide bond 34−41 (Figure 9A). On the other hand, in the crystal structure, this loop is "opened" by Glu141 from cathepsin L, which replaces the His59 position (Figure 9B). This observation reveals the induced fitting necessary for the interaction between P41icf and cathepsin L. His59 is highly conserved in all P41 sequences but not in the family of Tg-type-1 modules. This suggests that His59 could be responsible for the specificity of P41icf for cathepsin L.

## Experimental Section

**Equipment, Materials and General Methods.** Boc amino acids and other reagents for peptide synthesis were obtained from the Peptide Institute Inc. (Osaka, Japan). TLC was performed on Merck (Darmstadt, Germany) Kieselgel 60F-254 precoated plates. The compounds were viewed with UV light (254 nm) and/or ninhydrin reagent using different solvent systems. HPLC analysis was performed on a Shimadzu liquid chromatograph model LC-10A or LC-8A (Kyoto, Japan). Amino acid sequence analysis was carried out on an ABI 470 automated Edman degradation sequencer. Amino acid analysis was carried out on a JEOL analyzer model JLC-300 (Tokyo, Japan), after hydrolysis of the peptide with 6 M HCl in the presence or absence of 4% thioglycolic acid at 110 °C for 22 h. Molecular weights were measured with an MS (ESI) (HP 1100 LC/MSD, Palo Alto, CA). CD measurements were carried out on a Jasco J700 spectropolarimeter. Binding assays were performed on a BIACORE 2000. NMR spectra were recorded on a Bruker Avance DRX-800 and on a Bruker Avance DRX-600 operating at proton frequencies of 800 and 600 MHz, respectively.

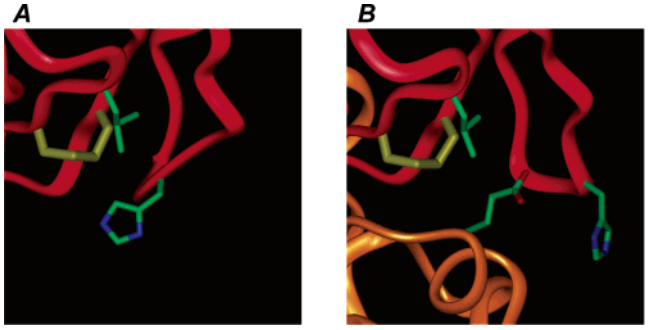


***Figure 9.*** (A) Ribbon plot of a region of the average NMR structure of free P41icf in solution, showing the side chains of Leu32, Cys34-Cys41, and His59. (B) Ribbon plot of a portion corresponding to the X-ray structure of P41icf (red) complexed with cathepsin L (orange). Side chains of Leu32, Cys34-Cys41, and His59 from the p41 fragment and Glu141 from the cathepsin L are shown.

**General Procedure for the Solid-Phase Synthesis of Protected Peptides on an HMFS Resin.** Each protected peptide was obtained using a semiautomatic peptide synthesizer (model ACT-90, Advanced ChemTech, Louisville, KY) on the respective Boc amino acid−HMFS resin (2 mmol). Peptides were synthesized using the "in situ" neutralization method.[37] One synthesis cycle consists of the following steps: (a) DCM (6 × 1 min); (b) TFA−DCM (8:2) (1 × 5 min + 1 × 20 min); (c) DCM (4 × 1 min); (d) NMP (6 × 1 min); (e) addition of Boc amino acid/HBTU/HOBt/DIEA (4:4:4:6 equiv) in NMP (45 min); (f) NMP (6 × 1 min). When the Boc group was removed from the Boc-His(Bom) peptide resin, the deprotection step was extended (1 × 5 min + 1 × 30 min), whereas the Boc group of the Boc-Asn peptide resin and Boc-Gln peptide resin was removed by adding 1% triisopropylsilane to the reaction mixture to quench the Xan cation generated during the deprotection reaction. Coupling reactions were monitored by the Kaiser ninhydrin test, and the coupling step was repeated when couplings were not completed.

**General Procedure for the Detachment of Protected Segments from the Resin.** The peptide resins were placed in a 200 mL cylindrical vessel with a sintered glass bottom and were treated with 80 mL of morpholine−DMF (2:8) (2 × 30 min). After filtration, DMF solution was evaporated to dryness and the resultant residue was triturated with 1 N HCl, washed successively in water, and dried. Each segment was purified by recrystallization, reprecipitation, or silica gel chromatography using appropriate solvents. Homogeneity of each fragment was analyzed using HPLC: the minimum required purity was 95%. The protected peptides were also analyzed by TLC, amino acid analysis, and MS (ESI).

**General Procedure for the Removal of the Boc Group in Solution.** The protected peptides were dissolved in TFA at −10 °C and then allowed to react for 1 h at room temperature. After removal of the TFA under vacuum, the residue was triturated with 4.6 N HCl

(37) Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, *40*, 180.

solution in dioxane to convert the TFA salt to the HCl salt, and the product was precipitated with ether and dried over NaOH under vacuum.

**Synthesis of the Fully Protected Peptide 13.** Protected Boc-(59−64)-OH (1.75 g, 1.34 mmol) was dissolved in DMF (17.5 mL), and pTsOH·H-Ser(Bzl)-OBzl (0.736 g, 1.61 mmol) and HOOBt (0.284 g, 1.74 mmol) were added. The mixture was cooled to −10 °C. EDC (0.295 mL, 1.61 mmol) was then added and allowed to react for 6 h at room temperature. The reaction mixture was precipitated by adding an excess of saturated brine at 0 °C. The precipitates were collected by filtration and then washed with water and ether. The product was purified by reprecipitation from MeCN; the yield was 1.51 g (71%). Amino acid analysis: Asx 1.00 (1), Ser 1.82 (2), Glu 1.04 (1), Cys n.d. (1), His 2.00 (2). MS (ESI) $[M+H^+]^+$: 1577.0 Da (1577.81 calculated).

The Boc group of protected peptide **7** (1.46 g, 0.93 mmol) was removed. The HCl salt of **7** and fragment **6** (1.55 g, 0.83 mmol) were dissolved in DMF−NMP (6:1) (12 mL), and HOBt (0.123 g, 0.91 mmol) and EDC (0.166 mL, 0.91 mmol) were added at −10 °C. The mixture was allowed to react for 4 h at room temperature, and it was precipitated by adding an excess of chilled water saturated with NaCl. The precipitates were collected by filtration and then washed with water and ether. Product **9** was purified by reprecipitation from hot MeCN; the yield was 1.79 g (64%). Amino acid analysis: Asx 2.00 (2), Thr 1.91 (2), Ser 2.71 (3), Glx 2.01 (2), Pro 1.02 (1), Cys n.d. (1), Gly 1.01 (1), Val 1.03 (1), His 1.99 (2), Arg 2.04 (2). MS (ESI) $[M+H^+]^+$: 3337.0 Da (3336.87 calculated).

The Boc group of protected peptide **9** (1.75 g, 0.52 mmol) was removed. The HCl salt of **9** (1.19 g, 0.36 mmol) and protected peptide **5** (0.57 g, 0.37 mmol) were dissolved in DMF (12 mL), and HOBt (0.065 g, 0.48 mmol) and EDC (0.087 mL, 0.48 mmol) were added at −10 °C. The mixture was allowed to react for 18 h at room temperature, and it was precipitated by adding an excess of chilled water. The precipitates were collected by filtration and then washed with water and ether. The resulting product **11** was purified by reprecipitation from MeCN; the yield was 1.56 g (86%). Amino acid analysis: Asx 2.92 (3), Thr 1.84 (2), Ser 2.62 (3), Glx 2.02 (2), Pro 2.01 (2), Cys n.d. (3), Gly 1.98 (2), Val 2.05 (2), Tyr 0.96 (1), Phe 1.00(1), His 1.92 (2), Trp n.d. (1), Arg 1.96 (2). MS (ESI) $[M+Na^+]^+$: 4766.0 Da (4766.56 calculated).

The Boc group of protected peptide **11** (1.69 g, 0.36 mmol) was removed. The HCl salt of **11** (1.27 g, 0.27 mmol) and protected peptide **4** (0.55 g, 0.29 mmol) were dissolved in NMP (30 mL), and HOBt (0.041 g, 0.30 mmol) and EDC (0.055 mL, 0.30 mmol) were added at −10 °C. The mixture was allowed to react for 10 h at room temperature and was precipitated by adding an excess of chilled water saturated with NaCl. The precipitates were collected by filtration and then washed with water and ether. Product **12** was purified by reprecipitation from MeCN; the yield was 1.78 g (91%). Amino acid analysis: Asx 3.93 (4), Thr 1.86 (2), Ser 3.57 (4), Glx 3.04 (3), Pro 3.03 (3), Cys n.d. (4), Gly 4.01 (4), Val 2.04 (2), Ile 0.97 (1), Leu 2.08 (2), Tyr 2.88 (3), Phe 1.00(1), His 1.94 (2), Trp n.d. (1), Arg 1.97 (2). MS (ESI) $[M+H^+]^+$(derivate de-Boc, de-Xan): 6255.3 Da (6255.39 calculated).

$Cs_2CO_3$ (0.22 g, 0.685 mmol) was added to the stirred solution of protected Boc-(18−27)-OH (2.6 g, 1.37 mmol) in DMF (20 mL). The solution was cooled to −10 °C, and phenacyl bromide (0.32 g, 1.59 mmol) was added. The mixture was stirred for 5 h at room temperature. The reaction mixture was precipitated by adding an excess of chilled water. The precipitates were collected by filtration and then washed with water and ether. Product **3** was purified by being triturated with $CHCl_3$ and MeOH, washed in ether, and dried; the yield was 2.28 g (83%). amino acid analysis: Asx 2.00 (2), Ser 0.91 (1), Glx 1.04 (1), Pro 0.96 (1), Gly 1.00 (1), Cys n.d. (1), Phe 0.97 (1), Lys 0.99 (1), Arg 0.97 (1). MS (ESI) $[M+H^+]^+$: 2018.8 Da (2019.74 calculated).

The Boc group of protected peptide **3** (2.23 g, 1.10 mmol) was removed. The HCl salt of **3** (1.90 g, 0.97 mmol) and protected peptide **2** (1.72 g, 1.04 mmol) were dissolved in DMF (19 mL), and HOBt (0.147 g, 1.09 mmol) and EDC (0.199 mL, 1.09 mmol) were added at −10 °C. The mixture was allowed to react for 20 h at room temperature and was precipitated by adding an excess of chilled water. The precipitates were collected by filtration and then washed with water and ether. The resulting product **8** was purified by reprecipitation from $CHCl_3$−TFE (3:1); the yield was 1.75 g (49%). Amino acid analysis: Asx 1.99 (2), Ser 1.90 (2), Glx 2.09 (2), Pro 3.08 (3), Cys n.d. (1), Gly 2.05 (2), Ala 1.04 (1), Val 1.99 (2), Ile 0.97 (1), Phe 1.00 (1), His 1.94 (2), Lys 1.01 (1), Arg 1.00 (1). MS (ESI) $[M+H^+]^+$: 3555.8 Da (3556.55 calculated).

The Boc group of protected peptide **8** (1.73 g, 0.49 mmol) was removed. The HCl salt of **8** (1.61 g, 0.46 mmol) and protected peptide **1** (0.63 g, 0.51 mmol) were dissolved in DMF−NMP (6:1) (21 mL), and HOBt (0.088 g, 0.54 mmol) and EDC (0.098 mL, 0.54 mmol) were added at −10 °C. The mixture was allowed to react for 21 h at room temperature and was precipitated by adding an excess of chilled water. The precipitates were collected by filtration and then washed with water and ether. Product **10** was purified by reprecipitation from MeCN; the yield was 1.44 g (66%). amino acid analysis: Asx 2.01 (2), Thr 1.01 (1), Ser 1.87 (2), Glx 4.33 (4), Pro 3.10 (3), Cys n.d. (2), Gly 2.07 (2), Ala 1.07 (1), Val 2.03 (2), Ile 0.96 (1), Leu 1.06 (1), Phe 1.00 (1), His 1.96 (2), Lys 2.01 (2), Arg 1.02 (1). MS (ESI) $[M+H^+]^+$: 4671.0 Da (4671.30 calculated).

Protected peptide **10** (1.40 g, 0.30 mmol) was dissolved in AcOH (50 mL). Zn dust (5.82 g, 89.1 mmol) was added, and the whole mixture was stirred vigorously under argon atmosphere at 40 °C for 3 h. The Zn dust was removed by filtration, and the filtrate was concentrated to a residue, which was triturated with HCl 0.5 N, and washed with water and ether. The product was purified by reprecipitation from $CHCl_3$−TFE (3:1); the yield of Boc-(1−27)-OH was 1.21 g (89%). Protected peptide **12** (1.75 g, 0.27 mmol) was treated with TFA−$^i$Pr$_3$SiH−$H_2O$ (95:2.5:2.5) (18 mL), and the product was converted to HCl salt, as described previously. The HCl salt (1.50 g, 0.24 mmol) and protected Boc-(1−27)-OH (1.64 g, 0.25 mmol) were dissolved in NMP (20 mL), and HOBt (0.038 g, 0.28 mmol) and EDC (0.051 mL, 0.32 mmol) were added at −10 °C. The mixture was allowed to react for 8 h at room temperature and was precipitated by adding an excess of chilled water saturated with NaCl. The precipitates were collected by filtration and then washed with water and ether. Product **13** was purified by reprecipitation from MeCN; the yield was 2.46 g (86%). Amino acid analysis: Asx 5.75 (6), Thr 2.79 (3), Ser 5.02 (6), Glx 7.07 (7), Pro 5.98 (6), Cys n.d. (6), Gly 5.76 (6), Ala 0.98 (1), Val 3.88 (4), Ile 1.86 (2), Leu 3.00 (3), Tyr 2.81 (3), Phe 1.94 (2), His 3.69 (4), Lys 2.37 (2), Trp n.d. (1), Arg 2.97 (3). MS (ESI) $[M+H^+]^+$: 10789.0 Da (10790.56 calculated).

**Synthesis of P41icf.** The fully protected peptide **13** (1.3 g, 0.12 mmol) was treated by HF (45 mL) in the presence of anisole (5 mL) and HCl·H-Cys-OH·$H_2O$ (0.42 g, 2.41 mmol) at −5 °C for 1 h. The HF was removed under vacuum, and the residue was triturated with ether and collected by filtration. The crude product was dissolved in 10% AcOH (30 mL) and was purified by preparative RP-HPLC on a YMC ODS-A $C_{18}$ column (250 mm × 30 mm) using a linear gradient (17.6−37.6% MeCN in 0.1% TFA/$H_2O$ for 80 min) and high temperature (40 °C) to obtain 388 mg (42%) of purified 6-Cys(Acm)-containing peptide.

To a 50% AcOH aqueous solution (40 mL) of the 6-Cys(Acm)-containing peptide (380 mg, 0.050 mmol) $Hg(OAc)_2$ (103 mg, 0.327 mmol) was added. The mixture was stirred for 1 h at room temperature. $\beta$-Mercaptoethanol (0.45 mL) was added to the solution, which was stirred for an additional hour. The solution was then applied to a Sephadex G-25 column and eluted with 10% AcOH. The main fraction was collected and lyophilized to obtain 340 mg (95%) of 6-SH peptide.

The 6-SH peptide (170 mg, 0.023 mmol) was dissolved in 3.4 L of 0.1 M $AcONH_4$ buffer (pH 7.8) containing 0.5 M Gu·HCl, 1mM EDTA, reduced glutathione, and oxidized glutathione. The ratio of peptide and redox reagents was 1:100:10. The oxidative folding was carried out

under gentle stirring in a stream of Ar for 3 days at room temperature. The mixture was acidified to pH 2 by adding TFA, and the folded peptide was purified by preparative RP-HPLC on a YMC ODS-A $C_{18}$ column (250 mm × 30 mm) using a linear gradient (17.1−37.1% MeCN in 0.1% TFA/$H_2O$ for 80 min). The fraction containing the folded peptide was further purified using the same column, with isocratic conditions (25% MeCN in 0.1% TFA/$H_2O$ for 80 min) at 40 °C, to separate the two disulfide isomers. The aggregated fraction was dissolved in 0.1 M tris·HCl buffer (pH 8.2) containing 7.3 M Gu·HCl and 0.5% EDTA, and an excess of DTT was added. The mixture was stirred at 40 °C for 90 min. The mixture was acidified to pH 2 by adding TFA, and the reduced peptide was purified by preparative RP-HPLC on a YMC ODS-A $C_{18}$ column (250 mm × 30 mm) using a linear gradient (19.7−39.7% MeCN in 0.1% TFA/$H_2O$ for 80 min). The fractions containing the reduced peptide were collected and lyophilized and were then submitted to a new folding reaction under the same conditions described previously. The two lots of folded peptide were joined to yield 21 mg (12%) of folded peptide. Amino acid analysis: Asx 5.82 (6), Thr 2.83 (3), Ser 5.33 (6), Glx 7.15 (7), Pro 5.83 (6), Cys n.d. (6), Gly 5.79 (6), Ala 0.99 (1), Val 3.82 (4), Ile 1.84 (2), Leu 3.00 (3), Tyr 2.80 (3), Phe 1.94 (2), His 3.77 (4), Lys 1.85 (2), Trp n.d. (1), Arg 2.89 (3). MS (ESI) $[M+H^+]^+$: 7247.3 Da (7247.09 calculated).

**Disulfide Structure Determination.** The major isomer of the folded peptide (0.1 mg, 13.8 nmol) obtained above was digested with α-chymotrypsin (5 μL of a solution 1 mg/mL) in 0.1 M tris·HCl buffer (pH 6.5) for 8 h at 37 °C. The digestion was stopped by addition of TFA until pH 2. The mixture was analyzed by RP-HPLC on a YMC-Pak ODS $C_{18}$ column (150 mm × 4.6 mm) using a linear gradient (10−60% MeCN in 0.1% TFA/$H_2O$ for 40 min), and the four peaks found were collected and characterized by sequence analysis (Edman degradation) and MS (ESI). The minor isomer was digested in the same way as described previously. Major isomer MS (ESI): first peak [$Gly^{36}$-$Tyr^{40}$] $[M+H^+]$ 496.3 Da (496.20 calculated); second peak [$Cys^{43}$-$Arg^{55}$, $Ser^{56}$-$Ser^{55}$] $[M+H^+]$ 2545.3 Da (2545.10 calculated); third peak [$Gln^{33}$-$Tyr^{35}$, $Cys^{41}$-$Trp^{42}$] $[M+H^+]$ 718.3 Da (717.97 calculated); fourth peak [$Leu^{1}$-$Phe^{19}$, $Arg^{20}$-$Leu^{32}$] $[M+H^+]$ 3596.8 Da (3596.75 calculated). Minor isomer MS (ESI): first peak [$Cys^{41}$-$Trp^{42}$, $Ser^{56}$-$Ser^{65}$] $[M+H^+]$ 1419.1 Da (1419.48 calculated); second peak [$Gly^{36}$-$Tyr^{40}$] $[M+H^+]$ 496.3 Da (496.20 calculated); third peak [$Gln^{33}$-$Tyr^{35}$, $Cys^{43}$-$Arg^{55}$] $[M+H^+]$ 1844.7 Da (1845.04 calculated); fourth peak [$Leu^{1}$-$Phe^{19}$, $Arg^{20}$-$Leu^{32}$] $[M+H^+]$ 3596.8 Da (3596.75 calculated).

**Binding to Cathepsin L.** The binding assay of synthetic P41icf with human cathepsin L was performed on a BIACORE 2000 apparatus. Synthetic P41icf was immobilized onto a flow cell of a CM5 sensor chip (Biacore AB), and solutions of different concentration of human cathepsin L dissolved in 10 mM acetate buffer (pH 5), 150 mM NaCl, were injected over the chip at a flow rate of 10 μL/min for 3 min. The dissociation step (6 min) was performed in running buffer at the same flow rate. The sensorgrams were corrected by subtracting a control flow cell and analyzed by the software BIAevaluation 3.1.

**Circular Dichroism Spectroscopy.** CD measurements were carried out at 0.2 nm intervals over the wavelength range from 190 to 260 nm, a spectral bandwidth of 1 nm, and a time constant of 4 s (scan speed 10 nm/min). The CD spectra represented an average of three accumulations and were corrected by subtracting the buffer baseline. CD spectra were recorded at a peptide concentration of 25 μM using a 0.1 cm path length quartz cell. The samples were prepared by dissolving a lyophilized peptide previously quantified by amino acid analysis in 5 mM phosphate buffer (pH 7.2). Data are represented in molar ellipticities per residue ($[\theta]_M$ deg $cm^2$ $dmol^{-1}$).

**NMR Measurements.** NMR experiments were performed with 500 μL of 1 mM samples containing 20 mM phosphate buffer (pH 5.7) and 0.01 mM $NaN_3$ dissolved either in 85% $H_2O$ and 15% $D_2O$ or in 100% $D_2O$. Dioxane was used as internal standard. NMRPipe[38] and NMRview[39] programs were used for data processing and spectral analysis, respectively. Two-dimensional homonuclear TOCSY (70ms, 85ms), DQF-COSY, and NOESY (150ms, 200ms) experiments at different temperatures (5 °C, 15 °C, and 25 °C) were used for the spin system and sequential assignment. $^3J_{NH-C\alpha H}$ and $^3J_{C\alpha H-C\beta H}$ coupling constants were measured in the DQF-COSY spectra. We used the method previously described by Ludvingsen et al.[40] to compensate for the overestimation of the $^3J_{NH-H\alpha}$ due to the broad line width of the antiphase multiplets.

**Structure Calculations.** Proton−proton distance restraints were obtained from NOEs recorded in 2D NOESY experiments. NOEs were translated into upper-limit constraints, according to their intensity by using the following qualitative criterion: strong, medium, and weak were set to distances lower than 2.8, 3.6, and 4.8 Å, respectively. Angular restraints ($\phi$ dihedral angles) were obtained from the coupling constants $^3J_{NH-H\alpha}$, using two cutoffs. For coupling constants greater than 8 Hz, the dihedral angle was constrained to $-120° \pm 35°$; and, for constants smaller than 6 Hz, the angle was constrained to $60° \pm 30°$. In addition, $\chi_1$ angles were obtained from the analysis of the $^3J_{H\alpha-H\beta}$ coupling constants and intraresidue NOEs.[41] A total of 925 NOEs and 56 angular restraints were used to initiate the calculation of 600 randomized conformers using the torsion angle dynamics protocol of DYANA.[42] The 30 conformers with the lowest final target value ($<0.7$ Å$^2$) with neither distance nor dihedral violation were energy minimized with the AMBER 5 program[43] using the 1995 all-atom force field. The atomic coordinates for the 30 best conformers described in this paper have been deposited in the protein data bank (accession number 1L3H).

**Acknowledgment.** We acknowledge support from the NMR Facility of the Serveis Cientificotècnics (Universitat de Barcelona) where NMR experiments were carried out. P.B. gratefully acknowledges a postdoctoral fellowship from the French "Institut National de la Santé et de la Recherche Médicale" (INSERM). This work was supported by MCYT (Bio 2002−2301) and Generalitat de Catalunya (CeRBa and SGR).

**Supporting Information Available:** RP-HPLC profile of the protected peptide **13**; scheme used for identification of disulfide bridges of P41icf; far-UV CD and near-UV CD spectra of synthetic P41icf at different temperatures; far-UV CD spectra of synthetic P41icf at different Gu·HCl concentrations; $^1$H NMR spectra of P41icf at different temperatures; residue contact map based on observed NOEs for P41icf; NOE connectivities (short and medium range) along the P41icf sequence; pattern of characteristic NOEs defining the three-strand antiparallel $\beta$-sheet; $^1$Hα conformational chemical shifts and chemical shift index for the p41icf fragment (at 298 K); Ramachandran ($\phi$,$\psi$) plot for the 30 NMR structures of P41icf; assignment of $^1$H resonances, temperature coefficients, and $^3J_{NH\alpha}$ coupling constant values of P41icf. This material is available free of charge via the Internet at http://pubs.acs.org.

**Note Added after ASAP:** The version published on the Web 1/18/2003 contained an error in an author name. The final version published 2/5/2003 and the print version are correct.

JA0207908

(38) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(39) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603.
(40) Ludvingsen, S.; Andersen, K. V.; Poulsen, F. M. *J. Mol. Biol.* **1991**, *217*, 731.
(41) Hyberts, S. G.; Märki, W.; Wagner, G. *Eur. J. Biochem.* **1987**, *164*, 625.
(42) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283.
(43) 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.; Ferguson, D. M.; Radner, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 5*; University of California, San Francisco, 1997.