NMR Solution Structure of the C-Terminal Fragment 255−316 of Thermolysin: A Dimer Formed by Subunits Having the Native Structure[†,‡]

Manuel Rico,*[,§] M. Angeles Jiménez,[§] Carlos González,[§] Vincenzo De Filippis,[∥] and Angelo Fontana[∥]

Instituto de Estructura de la Materia, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain

Received August 2, 1994; Revised Manuscript Received October 6, 1994[⊗]

ABSTRACT: The solution structure of the C-terminal fragment 255−316 of thermolysin has been determined by two-dimensional proton NMR methods. For this disulfide-free fragment there was a previous proposal according to which it would fold into a stable helical structure forming a dimer at concentrations above 0.06 mM. A complete assignment of the proton NMR resonances of the backbone and amino acid side chains of the fragment was first performed using standard sequential assignment methods. On the basis of 729 distance constraints derived from unambiguously assigned nuclear Overhauser effect (NOE) proton connectivities, the three-dimensional structure of a monomeric unit was then determined by using distance geometry and restrained molecular dynamic methods. The globular structure of fragment 255−316 of thermolysin in solution, composed of three helices, is largely coincident with that of the corresponding region in the crystallographic structure of intact thermolysin [Holmes, M. A., & Matthews, B. W. (1982) *J. Mol. Biol. 160,* 623−639]. This fact allowed identification as intersubunit of up to 52 NOE cross correlations, which were used to dock the two subunits into a symmetric dimer structure. The obtained dimeric structure served as the starting structure in a final restrained molecular dynamic calculation subjected to a total of 1562 distance constraints. In the resulting dimeric structure, the interface between the two subunits, of a marked hydrophobic character, coincides topologically with the one between the 255−316 fragment and the rest of the protein in the intact enzyme. The present work decisively shows that the thermolysin fragment 255−316 can attain a stable and nativelike structure independently of the rest of the polypeptide chain. Considering that the thermolysin molecule is constituted of two structural domains of equal size (residues 1−157 and 158−316), the results of this study show that autonomously folding units can be substantially smaller than entire domains.

The detailed mechanism by which proteins acquire their specific, biologically active, three-dimensional structure is one of the central unsolved problems of modern molecular biology (Creighton, 1992). At present, the protein folding problem is the subject of numerous studies using a variety of physicochemical and biological approaches and techniques. It is generally accepted now that the folding of a polypeptide chain does not occur by a random search of all possible conformations, but proceeds via folding intermediates, which thus direct the folding through a limited number of pathways and consequently lead to the final folded species within a reasonably short time (Levinthal, 1968; Kim & Baldwin, 1982, 1990). However, since the folding process is highly cooperative, the isolation and detailed structural characterization of folding intermediates has become a major problem (Creighton, 1990; Kim & Baldwin, 1990). Nevertheless, in recent studies measurements of hydrogen−deuterium exchange by nuclear magnetic resonance (NMR)[1] spectroscopy have provided considerable insights into the nature and populations of transient species on the folding pathway of a protein (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Bycroft et al., 1990; Miranker et al., 1991; Radford et al., 1992; Dobson et al., 1994). Current models of protein folding mechanisms imply that fluctuating elements of secondary structure are formed at the earliest stages of folding, followed by the onset of long-range interactions leading to the final globular state (Dill, 1985; Kim & Baldwin, 1990).

Considering that the early stages of protein folding appear to be characterized by the absence of tertiary interactions, the use of peptide models seems a logical approach for studying the initiation of protein folding, since the cooperative interactions that determine the rapid protein folding are necessarily absent (Wright et al., 1988). As an example,

[†] This work was supported by a special grant from the Commission of the European Communities, Biotechnology Action Program (Contract BAP-0249), and by grants from the Dirección General de Investigación Científica y Técnica (Spain) (Project PB90-0120) and the National Council of Research, Biotechnology and Bioinstrumentation Project (Italy).

[‡] Atomic coordinates for the eight converged dimeric structures have been deposited in the Brookhaven Protein Data Bank under filename 1TRL.

* Author to whom correspondence should be addressed (fax: 34-1-5642431).

[§] Consejo Superior de Investigaciones Científicas.

[∥] CRIBI Biotechnology Centre, University of Padua, Via Trieste 75, 35121 Padua, Italy.

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1994.

[1] Abbreviations: CD, circular dichroism; COSY, scalar coupling correlated spectroscopy; 2D, two-dimensional; dαα (*i, j*) NOE connectivity between the $C_\alpha H$ proton on residue *i* and the $C_\alpha H$ proton on residue *j*; dαN (*i, j*) NOE connectivity between the $C_\alpha H$ proton on residue *i* and the NH proton on residue *j*; dβN(*i, j*) NOE connectivity between the $C_\beta H$ proton on residue *i* and the NH proton on residue *j*; dNN(*i, j*), NOE connectivity between the NH proton on residue *i* and the NH proton on residue *j*; DSC, differential scanning calorimetry; HPLC, high-performance liquid chromatography; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; RMD, restrained molecular dynamics; RMSD, root mean square deviation; TOCSY, total correlation spectroscopy; TSP, sodium 3-(trimethylsilyl)[(2,2,3,3-$^2H_4$]propionate.

0006-2960/94/0433-14834$04.50/0 © 1994 American Chemical Society

we have examined in a previous paper (Jiménez et al., 1993) the propensity to form nonrandom structures on the part of the thermolysin peptides 258−276, 279−298, and 299−316 (see below) and found that these fragments form significant populations of helices spanning the same residues that are helical in the intact protein. Whereas this kind of study is now common practice, studies on longer peptides, which are potentially useful as models for conformations with long-range interactions and thus for folding steps subsequent to the initial formation of secondary structure elements, are rare [Oas & Kim, 1988; for a recent study, see Shin et al. (1993)]. On the other hand, there is clear evidence from X-ray analysis that relatively large globular proteins are assemblies of compactly folded substructures usually called domains. It is also possible to describe globular proteins in terms of a "hierarchic" architecture ranging from elements of secondary structure (helices, strands) to subdomains (supersecondary structures of folding units), domains, and whole protein molecules (Crippen, 1978). It is conceivable to suggest that specific segments of an unfolded polypeptide chain first refold to individual domains, which then associate and interact to give the final tertiary structure of the globular protein, much the same as do subunits in oligomeric proteins. The major implication of this model of protein folding by a mechanism of modular assembly is that isolated protein fragments corresponding to domains in the intact protein are expected to fold autonomously, thus resembling in their properties a small globular protein. Whereas independent folding of large protein domains has been demostrated in a number of cases (Wetlaufer, 1981), there are not many examples of well-defined and fixed three-dimensional folding on the part of protein subdomains, or fragments formed by a few elements of secondary structure.

In recent years we have addressed the question of the existence and properties of protein domains and subdomains by examining the folding and stability characteristics of fragments of the thermolysin molecule, a well-characterized metalloprotease isolated from *Bacillus thermoproteolyticus* (Titani et al., 1972; Holmes & Matthews, 1982). This protein is a single polypeptide chain of 316 amino acid residues without thiol or disulfide groups and containing a functional zinc ion and four calcium ions. The three-dimensional structure of thermolysin at 1.6-Å resolution establishes that this protein is constituted of two structural domains of equal size (residues 1−157 and 158−316), with the active site located at the interface between them (Holmes & Matthews, 1982). In previous work, we have examined the conformational properties of a number of fragments of thermolysin by using spectroscopic and immunochemical techniques. On this basis, it was concluded that fragments 121−316 and 206−316 are able to refold in water into a helical structure of nativelike characteristics. The smallest fragment able to fold into a remarkably stable, nativelike structure as deduced from CD estimates and immunochemical data was the disulfide-free fragment 255−316 (Dalzoppo et al., 1985; Vita et al., 1989). Nevertheless, a definite confirmation of the extent, the degree of analogy with the intact protein, and the dynamic behavior of the adopted structure was lacking. As is well-known, NMR spectroscopy, in contrast to CD, provides complete structural and dynamic information, once resonances have been assigned, at all individual sites throughout the molecule. Then, we decided to examine fragment 255−316 of thermolysin by NMR, and according to that study, we present in this paper the ^{1}H-NMR sequence-specific resonance assignment for this fragment, together with a description of its three-dimensional structure calculated by distance geometry and restrained molecular dynamics (RMD) techniques on the basis of distance constraints derived from observed NOE cross-correlations. At the concentrations used in this NMR study (~1 mM) this fragment folds into a symmetric dimer formed by two subunits having the native structure present in the intact enzyme. Where the subunit structure is concerned, the results of the present study definitely demonstrate that autonomously folding units can be substantially smaller than entire structural domains seen in the crystal structure of relatively large proteins (Wetlaufer, 1981). Moreover, this study represents an advance with respect to the numerous conformational studies conducted on short peptides, since thermolysin fragment 255−316 assumes in water a specific and stable tertiary structure most typical of a globular protein. An additional result of special significance is the determination of the topology of the symmetric dimer formed by the two monomeric subunits. As determined previously by far-UV CD measurements, dimerization does not lead to changes in fragment conformation. Simply, the highly hydrophobic surface created by the folding in each subunit is mutually protected through dimer formation.

MATERIALS AND METHODS

Sample Preparation. Fragment 255−316 of thermolysin was prepared by limited proteolysis with subtilisin of cyanogen bromide fragment 206−316 of thermolysin, as previously described (Dalzoppo et al., 1985). The sample of fragment 255−316 used for NMR studies was at least 95% pure by analytical reverse-phase HPLC.

NMR Spectroscopy. NMR samples were prepared by dissolving the peptide in either H_2O/D_2O, 9:1, or D_2O at a concentration of ca. 1 mM and a pH of 4.8 (direct pH meter reading without correction for isotope effects). Sodium 3-(trimethylsilyl)[2,2,3,3-$^2H_4$] propionate (TSP) was used as an internal reference. NMR experiments were performed on a Bruker AMX-600 spectrometer. All the two-dimensional spectra were acquired in the phase-sensitive mode using the time-proportional phase incrementation technique (Redfield & Kuntz, 1975; Aue et al., 1976; States et al., 1982; Marion & Wüthrich, 1983) with presaturation of the water signal. COSY (Wagner, 1983; Rance, 1987) and NOESY (Jeener et al., 1979; Macura & Ernst, 1980) spectra, using mixing times of 50, 100, and 200 ms in the latter, were recorded using standard phase-cycling sequences. TOCSY (Bax & Davis, 1985; Rance, 1987) spectra were performed using the standard MLEV17 spinlock sequence and 75 ms mixing time (Levitt & Freeman, 1979; Braunschweiler & Ernst, 1983). The size of the acquisition data matrix was 2048 × 512 words in f_2 and f_1, respectively, and prior to Fourier transformation the 2D data matrix was multiplied by a phase-shifted square sine bell window function in both dimensions and zero-filled to 4096 × 1024 words. The phase shift was optimized for every spectrum. Qualitative measurements of amide hydrogen exchange were made by freshly dissolving the sample in D_2O and observing the intensity of peptide NH resonances after a 48-h period. Slowly exchanging protons were those still present after that period.

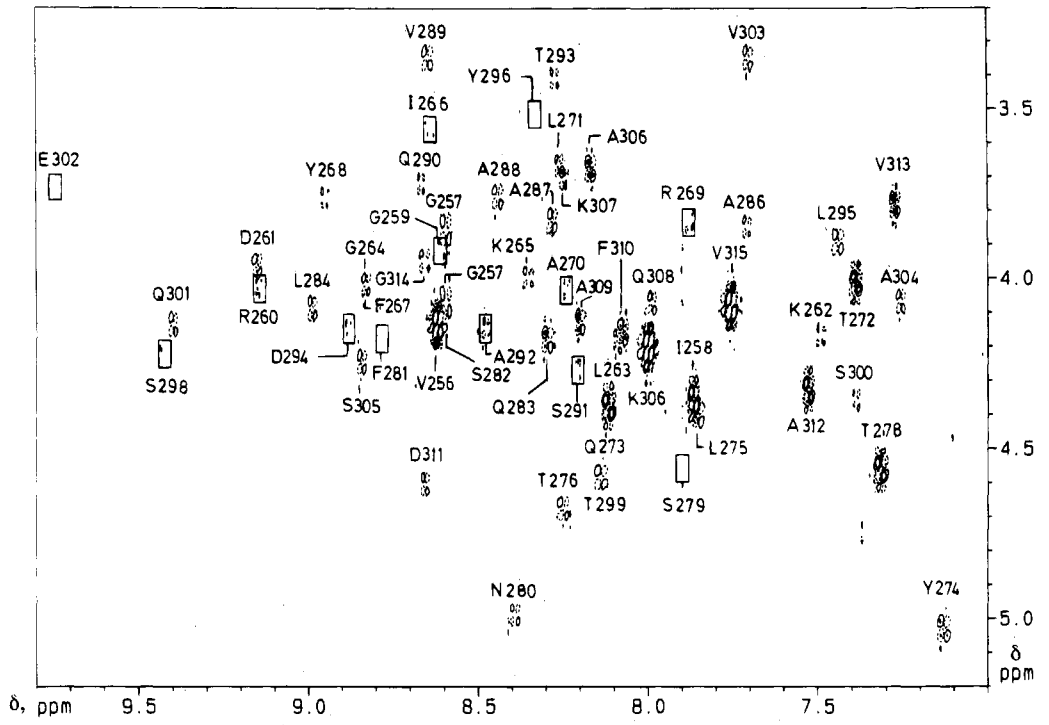

FIGURE 1: Fingerprint region of the 600-MHz phase-sensitive COSY spectrum of fragment 255−316 of thermolysin (1 mM; H_2O/D_2O, 9:1; pH 4.8; 25 °C). The boxes indicate cross peaks clearly observed at a lower contour level.

Structure Calculation. Distance constraints were obtained from nuclear Overhauser effect spectroscopy (NOESY) data acquired with mixing times of 50 and 100 ms. The intensities of the cross peaks were qualitatively divided into three different categories and translated into upper distance limit constraints according to the following criterion: up to 3 Å, 4 Å, and 5 Å for strong, medium, and weak NOE intensities, respectively. In cases of conflict, priority was given to data from the shortest mixing time spectrum. Stereospecific assignments of CβH2 and CβH3 side-chain protons were performed on the basis of the intensities of sequential and intraresidue NOE cross correlations together with $^3J_{\alpha\beta}$ coupling values. The usual corrections for pseudoatoms (Wüthrich, 1986) were added when no stereospecific assignments could be made. A pseudoatom correction was also added to the distance limit (Wüthrich, 1986) when a methyl group was involved in the constraints.

Preliminary structures of the monomer were calculated by distance geometry as implemented in the program DIANA (Gunter et al., 1991) and were further refined by using restrained molecular dynamics (RMD) with the GROMOS package (Van Gunsteren & Berendsen, 1987). In each of the RMD runs, a set of eight structures was generated by following an annealing strategy. The starting structure was first energy minimized and then heated to 1000 K over 10 ps. At this temperature, 40 ps of RMD was carried out and eight structures were extracted from the trajectory, one every 5 ps, and then submitted to a cooling strategy consisting of 5 ps at 750 K, 5 ps at 500 K, and 10 ps at 300 K. The last 5 ps of the trajectory was averaged, and the resulting structures were finally energy minimized. By using this protocol, which is ordinarily employed in our lab for structure calculations (Rico et al., 1991; Santoro et al., 1993), reasonably disordered structures are obtained in regions poorly defined by the distance constraints, thus demonstrating that the 40 ps of high-temperature molecular dynamics is a sufficiently long period for an appropriate sampling of the conformational space.

Docking of the two subunits for the initial dimer structure as well as all the molecular graphics manipulations was carried out with the program INSIGHT II, version 2.2.0 (Biosym Technologies, San Diego, 1993).

RESULTS

Assignment of the ^{1}H-NMR Spectra. The ^{1}H-NMR assignment of thermolysin 255−316 fragment was achieved by means of the standard sequential assignment strategy (Wüthrich, 1986). The first stage, involving the identification of systems of spin-coupled resonances belonging to a particular residue, was performed by means of COSY and TOCSY experiments in H_2O and D_2O. Illustrations of the COSY spectrum in H_2O (fingerprint region) and the TOCSY spectrum in H_2O (NH amide region) are given in Figures 1 and 2, respectively, where a number of assignments have been marked. The eight Ala, the five Thr, the two Ile, and the seven Val residues were identified without difficulty by their characteristic spin systems. The N-terminal Val 255 was unambiguously determined also, because of its lack of an NH amide resonance. Further complete spin systems were found for two Leu, four Glx, nine AMX, and one Lys. Scalarly correlated NH, CαH, and in some cases CβH resonances were identified for 17 additional residues. Some of these spin systems were shown to be complete in the next stage (the sequential assignment) like those belonging to Gly 314, with the two degenerate CαH protons, as well as Ser 279 and Tyr 296, both with degenerate CβH resonances. The spin systems of two of the five Leu residues were completely assigned, whereas for the three remaining ones the two methyl CδH$_3$ and the CγH resonances could be identified. The correspondence of these to their backbone protons was determined along with the sequential assignment process. The unique Pro 277 was also identified at this stage, and its

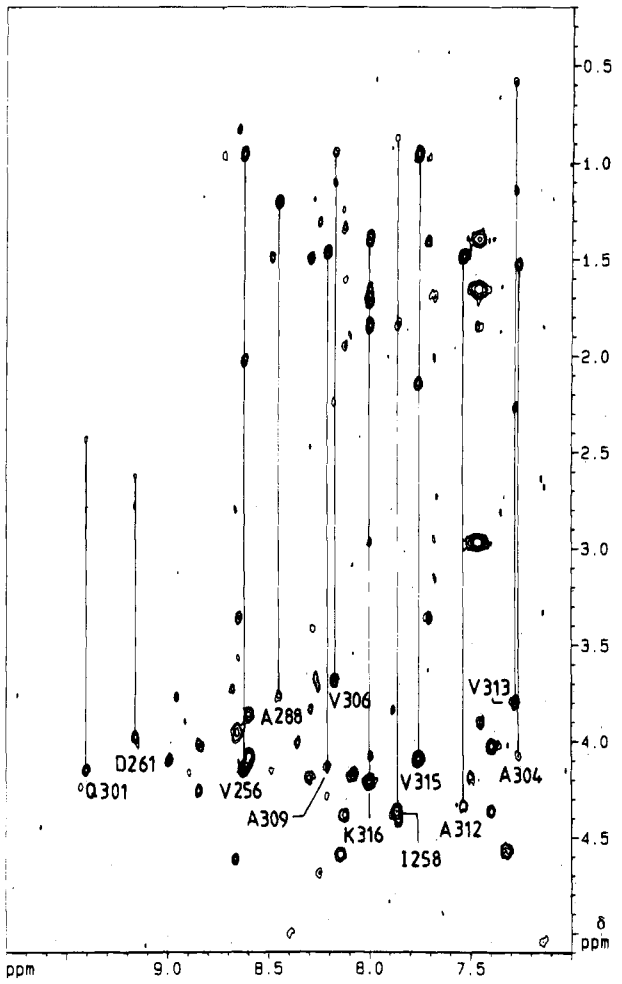

FIGURE 2: Region of the TOCSY spectrum of fragment 255−316 of thermolysin (1 mM; H_2O/D_2O, 9:1; pH 4.8; 25 °C). Some relayed connectivities are marked by continuous lines, and the labels are at the positions of the direct NH−CαH cross peaks.

characteristic spin system was confirmed in the NOESY experiments. On the whole, only three spin systems, which were identified in later steps, remained unassigned at this stage.

The aromatic ring protons belonging to the three Phe and the two Tyr residues were identified easily by a joint analysis of COSY and TOCSY experiments. Although the *ortho* and *meta* proton resonances of Phe 281 coincide accidentally, distinction of its *para*−*meta* proton cross-peak pattern from that of a Tyr residue was straightforward from the signal intensities in the 1D spectra obtained in D_2O solution. The *para* proton signal integrates for one proton, and the signal for the *ortho* and *meta* protons corresponds to 4 protons, what is not compatible with a Tyr ring. The ring proton signals of a remaining Tyr residue were missed on our initial analysis. Afterwards, it could be found that protons on this ring give rise to very broad signals. In fact, broad and weak *ortho*−*meta* cross peaks were observed in the TOCSY spectrum. Ring protons could be correlated to the corresponding backbone spin systems through the observed *ortho*-$C_\beta H$, *ortho*-$C_{\beta'}H$, and/or *ortho*-$C_\alpha H$ NOE connectivities. In a similar way, assignments of the Asn and Gln amide side-chain proton resonances to particular residues could be achieved on the basis of observed side-chain amide−$C_\beta HH'$ or amide−$C_\gamma HH'$ NOE connectivities in the case of Asn or Gln residues, respectively.

The second stage involves the assignment of the spin systems identified in the previous stage to specific residues in the protein sequence. The sequence-specific assignment is achieved by correlating one amino acid spin system with those of its neighbors, by means of sequential NN, αN, and/or βN NOE connectivities by using mainly cross peaks observed in the NOESY spectrum with the largest mixing time. For most residue pairs the two sequential backbone NOE connectivities were observed. In addition, many assignments were confirmed by sequential side chain NH NOE connectivities. As can be seen in Figure 3, the only point where no sequential NOE connectivity could be found was between Arg 260 and Asp 261 due to coincidence of their NH δ-values. Sequential $C_\alpha H-C_\delta H$ and $C_\alpha H-C_{\delta'}H$ NOE correlations were observed between Thr 276 and Pro 277, indicating that the corresponding peptide bond is in a *trans* configuration. The sequential assignment of segment 301−315 performed on the basis of NN NOE correlations and those of segments 271−274 and 299−305 through αN NOEs are illustrated in Figure 4 .

The hydroxylic proton resonances belonging to Ser 279 and Ser 291 were identified. They do not appear to be coupled, or rather their couplings to other protons are sufficiently small so that no COSY or TOCSY cross-peaks are observed for them. In D_2O, they are readily substituted by deuterium, but in H_2O exchange is sufficiently slow so that a large number of NOE connectivities can be observed for those resonances. Assignment to specific Ser residues was accomplished after an initial structure had been calculated, on the basis of finding OH protons fulfilling all the NOE correlations observed for them. In Table 1, the chemical shifts (δ-values in ppm from TSP) of all proton resonances of the thermolysin 255−316 fragment are given.

Stereospecific assignments of Val and Leu methyl groups, $C_\beta H-C_\beta H'$ protons, and Pro $C_\delta H-C_\delta H'$ protons were performed by a combined analysis of NOE intensities and $^3J_{\alpha\beta}$ coupling constants and a further check of the compatibility of the assignments with the preliminary calculated structures. Only those stereospecific assignments showing no ambiguity at all were introduced in the following steps of the structure calculations. These were the methyl groups of Val 303, 306, 313, and 315; the methyl groups of Leu 263, 271, 275, and 284; the $C_\beta H2$ and $C_\beta H3$ protons of Leu 263, Tyr 268, Tyr 274, Leu 275, and Gln 301; and the $C_\delta H2-C_\delta H3$ protons of Pro 277.

Secondary Structure Analysis. Secondary structure was readily delineated from a qualitative interpretation of ^1H chemical shifts, together with an analysis of observed NOE patterns and NH-exchange data. Figure 3 summarizes the medium-range NOE connectivities observed for thermolysin 255−316 fragment in addition to the sequential NN, αN, and βN NOE correlations used for the sequence-specific assignment, whose relative intensities have been marked. Three long stretches of $NN_{i,i+2}$, $\alpha N_{i,i+3}$, $\alpha N_{i,i+4}$, and $\alpha\beta_{i,i+3}$ NOE connectivities, all of them characteristic of helices, are clearly distinguished. Figure 5 shows the spectral region where the $\alpha\beta_{i,i+3}$ NOE cross peaks are found, with those belonging to the C-terminal helix marked. Hydrogen-exchange data for the NH amide protons, shown also in Figure 3, provide additional support for the location and extension of the three helices. Residues within the regions defined as helical, with the exception of a few residues at the N-terminus of each helix, present slow to medium NH-

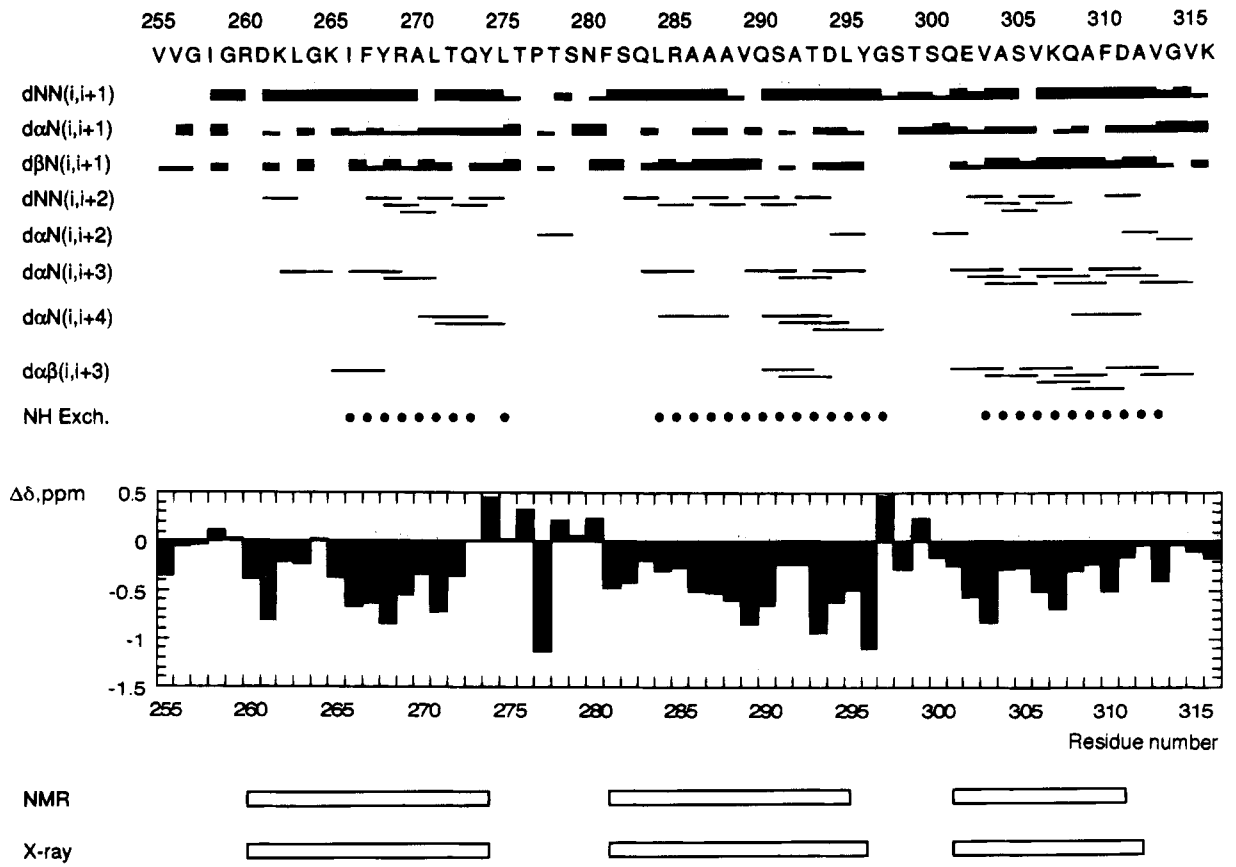

FIGURE 3: Sequence of fragment 255−316 of thermolysin together with a summary of observed sequential and medium-range NOEs involving NH, CαH, and Cβ H protons, amide-exchange data, and CαH conformational shifts (Δδ). The NOEs are classified as strong, medium, and weak by the thickness of the line. NH amide protons slowly-exchanging with solvent deuterons are marked by filled circles. The CαH conformational shifts were evaluated as the difference between the observed δ and that accepted as the randon coil value for each amino acid residue (Bundi & Wüthrich, 1979). At the bottom, the α-helices of fragment 255−316 of thermolysin in solution and those found in the crystal for the intact thermolysin are compared.

exchange rates. All this information indicates the presence of three helical regions spanning residues 260−274, 281−296, and 301−315, which is further confirmed by the upfield shift (relative to random coil values) of most of the $C_\alpha H$ protons within these stretches. A few exceptions can be observed regarding both chemical shifts and exchange rates which can have an origin different from the formation of secondary structure. In the case of chemical shifts, for instance, the values used are not corrected for aromatic ring current effects which can affect some of the conformational chemical shifts illustrated in Figure 3. In summary, the thermolysin 255−316 fragment in aqueous solution contains three helices which very closely match the helical segments of crystalline thermolysin (see Figure 3).

Tertiary Structure Determination. Complete assignment of cross peaks in crowded NOESY spectra is strongly facilitated by comparing all the potential assignments with interprotonic contacts predicted in a model structure (Kraulis, 1989). This model can be a crystallographic structure, when there is experimental evidence of close similarity with the structure in solution, or a set of preliminary NMR structures arising from unambiguous NOESY cross peaks.

In the case of symmetric dimers, the situation is even more complicated because of the additional ambiguity of inter- and intrasubunit cross peaks. Although new methods to deal with this problem have been developed very recently (Nilges, 1993), favorable cases can be treated by using traditional procedures, and intra- and intersubunit cross peaks can be sorted out during the assignment−structure calculation process (Kay et al., 1991). Crystallographic structures, when they are available, can greatly facilitate this assignment process (Breg et al., 1990).

In our case, initial structure calculations of the isolated thermolysin 255−316 fragment were carried out simultaneously with the analysis of NOESY spectra. Structures were calculated with the program DIANA (Gunter et al., 1991). When a distance constraint was strongly violated in all the resulting structures and an alternative assignment for the corresponding cross peak in the NOESY spectra could not be found, the cross peak was considered as intersubunit. After several cycles of assignment and structure calculation, 52 intersubunit cross peaks were identified. Some of the observed intersubunit NOE cross peaks are marked in the NOESY regions shown in Figures 5 and 6. The resulting structures were very similar to the crystallographic one except in some poorly defined regions where not many distance constraints were found. Thus, they each consist of a bundle of three helices, with the same extension and location in the sequence as found in the intact enzyme, showing moreover a very close relative orientation. This similarity supported the use of the X-ray structure as the starting point for docking the two subunits. The docking was carried out manually by using the program INSIGHT. In the resulting dimer, the symmetry was maintained and intersubunit distance constraints were roughly accomplished. This structure was submitted to a RMD refinement following the protocol

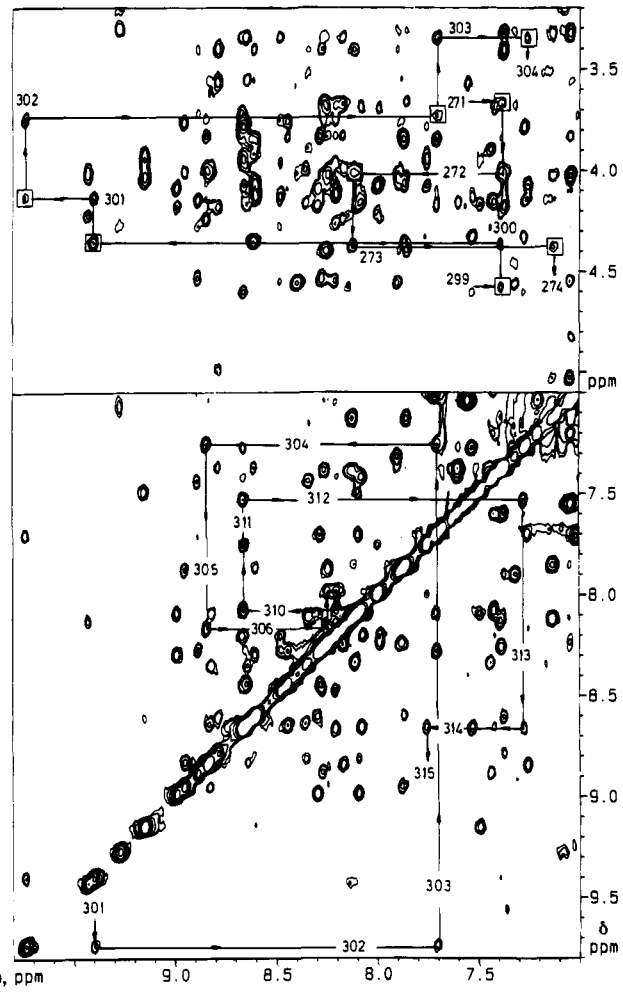

FIGURE 4: Portion of the NOESY spectrum of fragment 255−316 of thermolysin (1 mM; H_2O/D_2O, 9:1; pH 4.8; 25 °C; mixing time, 100 ms). Sequential connectivities $d_{\alpha N}(i, i + 1)$ leading to the assignment of the segments 271−274 and 299−305 are illustrated in the top panel. The assignment pathway through the sequential connectivities $d_{NN}(i, i + 1)$ is shown for segment 301−315 in the bottom panel.

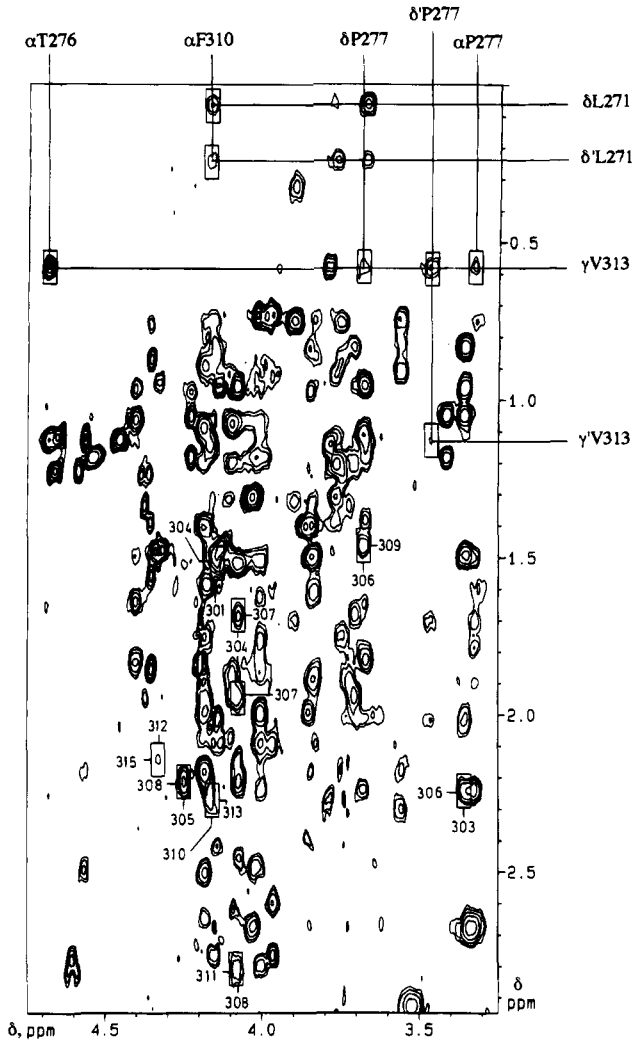

FIGURE 5: Selected region of the NOESY spectrum of fragment 255−316 of thermolysin (1 mM; D_2O; pH 4.8; 25 °C; mixing time, 100 ms). The nonsequential $d_{\alpha\beta}(i, i + 3)$ NOE connectivities belonging to the C-terminal helix are labeled. Observed intersubunit NOE cross correlations are marked outside the vertical and horizontal scales.

described in Materials and Methods. The 52 constraints classified as intersubunit turned out to be consistent with the final dimer structure since no significant distance constraint violations were observed, proving definitely that the NMR spectra are compatible with a symmetric dimer conformation.

A final assignment cycle was carried out, and some intrasubunit constraints were eliminated since they turned out to be ambiguous. The NOEs from which they derive were classified as intrasubunit in earlier stages of the process because no constraint violation was observed, but they could also correspond to intersubunit cross correlations. After removal of all ambiguities, the final set consisted of 729 intra- and 52 intersubunit distance constraints, making an overall set of 1562 constraints for the whole dimer. A set of eight dimer structures were calculated with this final set of constraints by using a high-temperature restrained molecular dynamics protocol without imposing any explicit symmetry conditions. Except for the initial six residues and the last two, the final dimer structure is remarkably well-defined, with a backbone RMSD of 0.85 Å, and it conserves the initial symmetry in the well-defined regions. Figure 7 shows a stereoview with the superposition of the eight final structures for one of the subunits. The dimer topology is illustrated in Figure 8 in a ribbon representation. Figure 9 shows the root mean square deviations of the backbone torsional angles of subunits 1 and 2 in the eight converged dimeric structures, and Figure 10 shows the backbone torsional differences between the average structures of both subunits, as well as the differences between a global average solution conformation and the crystal structure.

In figures 9 and 10, it can be seen that the largest values for the torsional root mean square deviations corresponding to the eight structures of the monomer as well as those for differences between subunits are located at the same main-chain regions, roughly coincident with terminal segments and turns or loops between helices. In addition, and more interestingly, the differences between subunits are very small, with the largest ones placed at those same regions. Thus, they are reflecting local chain mobility and not a real assymmetry in the dimeric structure. An analysis of the hydrophobic properties of the subunit surface indicates that the most hydrophobic area is located at the dimer interface.

Figure 11 is a summary of the NOE information employed in the structure determination (■) as well as the resulting contacts between residues (□) and hydrogen bonds (○) in the final dimeric structures. Thus, contact regions are readily

Table 1: ^{1}H Chemical Shifts of Thermolysin 255−316 Fragment at 22 °C, pH 4.8 (δ, ppm from TSP)

residue	NH	$C_\alpha H$	$C_\beta H$	others
Val 255		3.84	2.18	$C_\gamma H_3$ 0.97, 0.97
Val 256	8.62	4.14	2.02	$C_\gamma H_3$ 0.96, 0.96
Gly 257	8.59	4.07, 3.85		
Ile 258	7.87	4.35	1.85	$C_\gamma H$ 1.56, 1.38; $C_\gamma H_3$ 0.86; $C_\gamma H_3$ 0.75
Gly 259	8.61	4.08, 3.90		
Arg 260	9.15	4.00	2.08, 1.85	$C_\gamma H$ 1.62, 1.62; $C_\gamma H$ 3.10, 2.79; $N_\epsilon H$ 7.35
Asp 261	9.15	3.95	2.76, 2.61	
Lys 262	7.49	4.16	1.95, 1.95	$C_\gamma H$ 1.42, 1.38; $C_\gamma H$ 1.58, 1.58; $C_\epsilon H$ 3.00, 3.00
Leu 263	8.10	4.16	1.99, 1.76	$C_\gamma H$ 2.00; $C_\delta H_3$ 1.12, 1.08
Gly 264	8.82	4.00, 4.00		
Lys 265	8.35	4.00	2.23, 1.99	$C_\gamma H$ 1.76, 1.49; $C_\gamma H$ 1.76, 1.76; $C_\epsilon H$ 2.98, 2.98
Ile 266	8.64	3.56	2.18	$C_\gamma H$ 2.29, 2.29; $C_\gamma H_3$ 0.92; $C_\gamma H_3$ 0.88
Phe 267	8.84	4.02	2.66, 2.48	$C_\delta H$ 6.79, 6.79; $C_\epsilon H$ 7.04, 7.04; $C_\zeta H$ 7.21
Tyr 268	8.95	3.76	3.01, 2.98	$C_\gamma H$ 6.89, 6.89; $C_\epsilon H$ 6.74, 6.74
Arg 269	7.88	3.84	2.08, 2.08	$C_\gamma H$ 1.61, 1.61; $C_\gamma H$ 3.30, 3.20; $N_\epsilon H$ 9.26; $N_\zeta H$ 7.10, 7.10, 6.24, 6.24
Ala 270	8.24	4.02	1.30	
Leu 271	8.26	3.66	1.36, 1.36	$C_\gamma H$ 1.15; $C_\gamma H_3$ 0.24, 0.06
Thr 272	7.38	4.01	3.97	$C_\gamma H_3$ 0.72
Gln 273	8.12	4.37	1.32, 1.22	$C_\gamma H$ 1.94, 1.62; $N_\epsilon H$ 7.10, 6.63
Tyr 274	7.12	5.04	3.33, 2.67	$C_\delta H$ 7.04, 7.04; $C_\epsilon H$ 6.62, 6.62
Leu 275	7.85	4.40	1.82, 1.63	$C_\gamma H$ 1.82; $C_\delta H_3$ 1.05, 0.95
Thr 276	8.25	4.68	4.45	$C_\gamma H_3$ 1.12
Pro 277		3.32	2.25, 1.69	$C_\gamma H$ 2.02, 1.78; $C_\delta H$ 3.68, 3.46
Thr 278	7.32	4.56	4.66	$C_\gamma H_3$ 1.12
Ser 279	7.90	4.56	4.05, 4.05	$O_\gamma H$ 5.90
Asn 280	8.39	4.99	3.57, 3.11	$N_\delta H$ 7.55, 7.04
Phe 281	8.78	4.18	3.41, 3.03	$C_\delta H$ 7.38, 7.38; $C_\epsilon H$ 7.38, 7.38; $C_\zeta H$ 7.60
Ser 282	8.61	4.08	3.84, 3.84	
Gln 283	8.30	4.18	2.51, 2.18	$C_\gamma H$ 2.64, 2.64; $N_\epsilon H$ 7.57, 6.74
Leu 284	8.99	4.09	1.92, 1.86	$C_\gamma H$ 2.06; $C_\delta H_3$ 1.19, 1.07
Arg 285	8.09	4.12	1.58, 0.84	$C_\gamma H$ 1.20, 0.64; $C_\delta H$ 3.14, 3.07; $N_\epsilon H$ 7.67
Ala 286	7.71	3.85	1.39	
Ala 287	8.28	3.83	1.49	
Ala 288	8.44	3.75	1.19	
Val 289	8.64	3.34	1.98	$C_\gamma H_3$ 0.82, 0.82
Gln 290	8.67	3.72	1.33, 1.33	$C_\gamma H$ 1.88, 1.82; $N_\epsilon H$ 7.71, 7.02
Ser 291	8.21	4.27	4.10, 3.96	$O_\gamma H$ 4.55
Ala 292	8.48	4.12	1.49	
Thr 293	8.27	3.41	4.54	$C_\gamma H_3$ 1.18
Asp 294	8.89	4.15	3.10, 2.76	
Leu 295	7.44	3.89	1.32, 0.32	$C_\gamma H$ 1.69; $C_\delta H_3$ 0.75, 0.75
Tyr 296	8.33	3.51	2.91, 2.91	$C_\delta H$ 7.61, 7.61; $C_\epsilon H$ 7.16, 7.16
Gly 297	8.11	4.85, 4.03		
Ser 298	9.43	4.22	4.75, 4.02	
Thr 299	8.15	4.58	4.65	$C_\gamma H_3$ 1.22
Ser 300	7.38	4.35	4.05, 3.91	
Gln 301	9.40	4.13	2.08, 2.00	$C_\gamma H$ 2.42, 2.42; $N_\epsilon H$ 7.63, 6.95
Glu 302	9.74	3.73	2.15, 1.50	$C_\gamma H$ 1.75, 1.75
Val 303	7.70	3.35	2.03	$C_\gamma H_3$ 1.04, 0.96
Ala 304	7.25	4.07	1.52	
Ser 305	8.84	4.24	3.99, 3.83	
Val 306	8.16	3.67	2.23	$C_\gamma H_3$ 1.09, 0.94
Lys 307	8.25	3.69	1.94, 1.94	$C_\gamma H$ 1.68, 1.68; $C_\delta H$ 1.68, 1.68; $C_\epsilon H$ 2.94, 2.72; $N_\zeta H$ 7.67
Gln 308	7.99	4.07	2.21, 2.21	$C_\gamma H$ 2.61, 2.44; $N_\epsilon H$ 7.48, 6.86
Ala 309	8.20	4.12	1.46	
Phe 310	8.08	4.15	3.08, 3.02	$C_\delta H$ 7.42, 7.42; $C_\epsilon H$ 6.82, 6.82; $C_\zeta H$ 6.98
Asp 311	8.66	4.61	2.80, 2.80	
Ala 312	7.53	4.33	1.48	
Val 313	7.27	3.78	2.26	$C_\gamma H_3$ 1.13, 0.58
Gly 314	8.66	3.94, 3.94		
Val 315	7.76	4.08	2.13	$C_\gamma H_3$ 0.96, 0.94
Lys 316	8.00	4.21	1.85, 1.72	$C_\gamma H$ 1.39, 1.39; $C_\delta H$ 1.66, 1.66; $C_\epsilon H$ 2.97, 2.97; $N_\zeta H$ 7.46

appreciated within a given subunit (up-diagonal, helix I−helix II) as well as between subunits (down-diagonal, mainly helix I−helix III).

Hydrogen Bonds. An analysis of atomic positions in the eight converged structures provides information about the hydrogen bonds present in those structures. In this regard, one hydrogen bond is considered to be formed if the H distance to the acceptor atom is less than 2.5 Å and the O−H−N angle is greater than 125°. Regular α-helices are characterized by repeated hydrogen bonds between the CO of residue i and the NH of residue i+4. As expected, three stretches of such i,i+4 hydrogen bonds are found in the regions 262−274, 280−297, and 300−312 in all resulting structures. Also, backbone torsional angles ϕ and ψ consistent with standard α-helices are obtained for residues 260−274, 281−295, and 301−311, thus confirming the qualitative description of elements of secondary structure made at the time of the assignment.

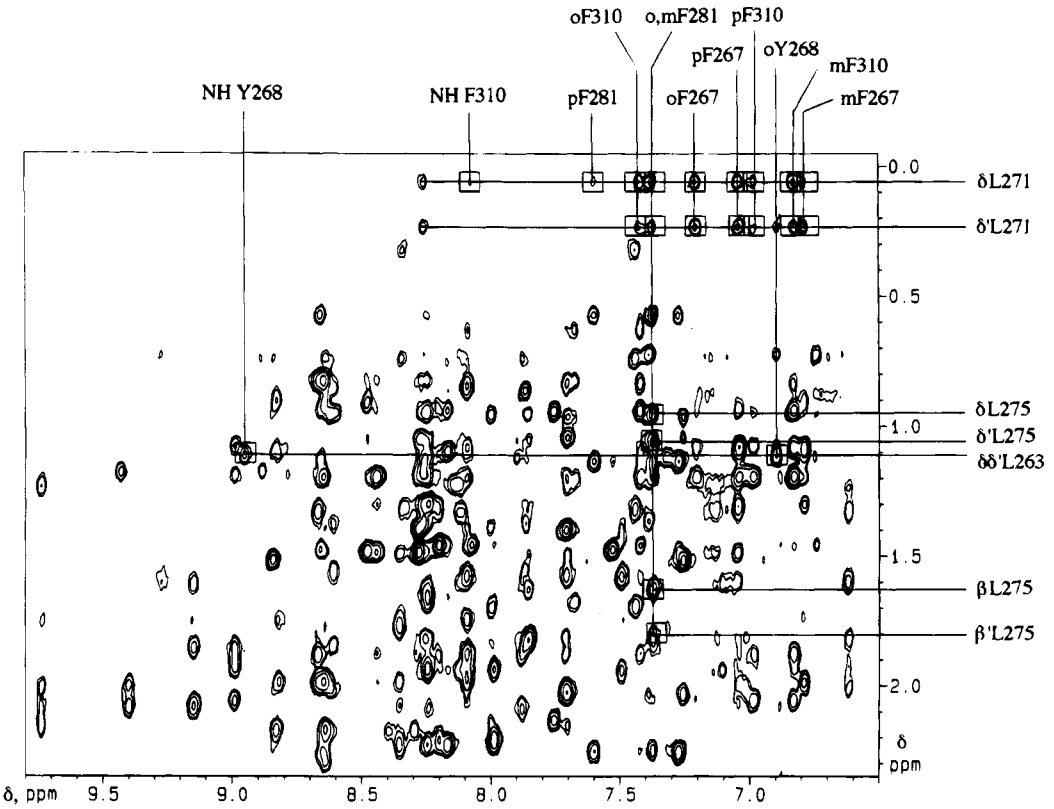

FIGURE 6: Part of the NOESY spectrum of fragment 255−316 of thermolysin (1 mM; H_2O/D_2O, 9:1; pH 4.8; 25 °C; mixing time, 100 ms) showing some intersubunit NOE connectivities.

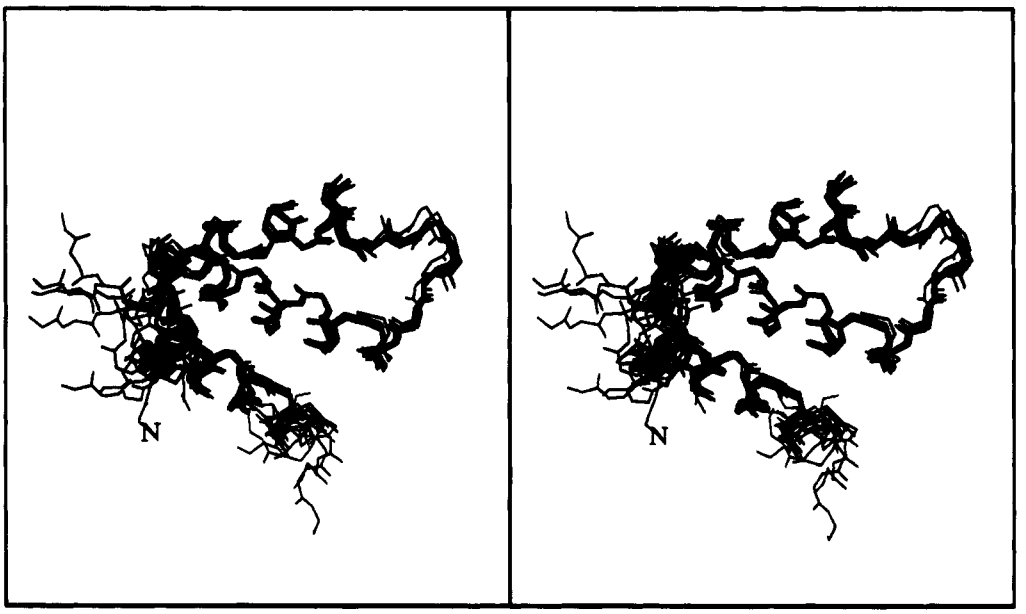

FIGURE 7: Stereoview of the superposition of the eight final structures for one of the subunits.

In the loop joining helices I and II, well-populated hydrogen bonds (those present in at least 50% of the computed structures) were found: Thr 276 N*H*−Ser 279 *O*H, Ser 279 O*H*−Thr 276 C*O*, and Ser 279 N*H*−Pro 277 C*O*, the first of these appearing also in the crystal structure of this segment in the intact enzyme. In the loop linking helices II and III, the hydrogen bonds present in the solution structure are the following: Ser 300 N*H*−Ser 298 C*O*, Glu 302 N*H*−Ser 300 *O*H, and Ser 300 O*H*−Tyr 296 C*O*; the first two are also found in the crystal structure. The hydrogen bond between Gln 283 N*H* and Asn 280 *Oδ* acts as a typical N-cap of the second helix and the Ser 291 O*H*−Ala 287 C*O* and Thr 293 O*H*−Val 289 C*O* hydrogen bonds contribute to a further stabilization of helix II. The fact that the hydroxylic protons of Ser 279 and Ser 291 are involved in hydrogen bonds is consistent with their relatively slow exchange rates observed in the spectra in light water. Helix I is also stabilized by the hydrogen bond between Arg 269 NH_ϵ and Tyr 274 *O*H. The salt bridge observed in the crystal between one $C_\eta H$ in the guanidinium group of Arg 285 and the carboxylate group of Asp 311 is found in solution in only one of the eight computed structures.

Two intersubunit hydrogen bonds are present in the calculated dimeric structures: Phe 281 N*H*−Ser 279′ C*O*,

FIGURE 8: Dimeric structure of thermolysin fragment 255−316 in a ribbon representation constructed with the MOLSCRIPT program (Kraulis, 1991).

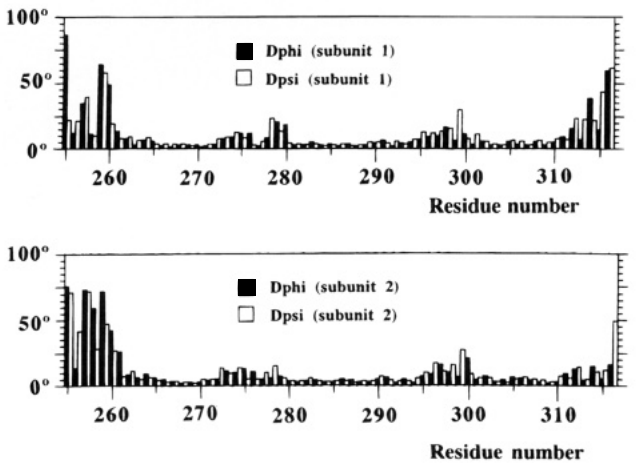

FIGURE 9: Root mean square deviations of the backbone torsional angles ϕ and ψ of the eight final structures of subunits 1 (top) and 2 (bottom).

which appears in all structures, and Gly 314 N*H*−Thr 272′ C*O*, present only in half of them.

Side-Chain Conformational Flexibility. An analysis of side-chain conformational flexibility was carried out on the basis of the values found for side-chain torsion angles in the computed structures. A side-chain torsion angle was considered as well-defined when its root mean square deviation between the values in the computed structures (eight for each subunit) was less than ±30°. In terms of the order parameter, which is 0 for a totally random dihedral angle and 1 for a completely fixed conformation, a well-defined conformation will have a value in the range 0.87−1.00. In Figure 12, we illustrate the order parameters for side-chain torsions, as a function of the sequence, together with the calculated relative accessible surface areas. Note that differences between accessible surface areas in the monomer and the dimer are an index of the residue being located at the interface between the two subunits.

If Gly, Ala, and Pro residues are not considered, we are left with 48 residues with undefined values for at least one side-chain torsion angle, 20 of which show a markedly random conformational behavior. With the exception of three valines located at the N- and C-termini, all other residues in this group are polar and their accessible surface areas fall in the range 29−95%, which therefore locates them

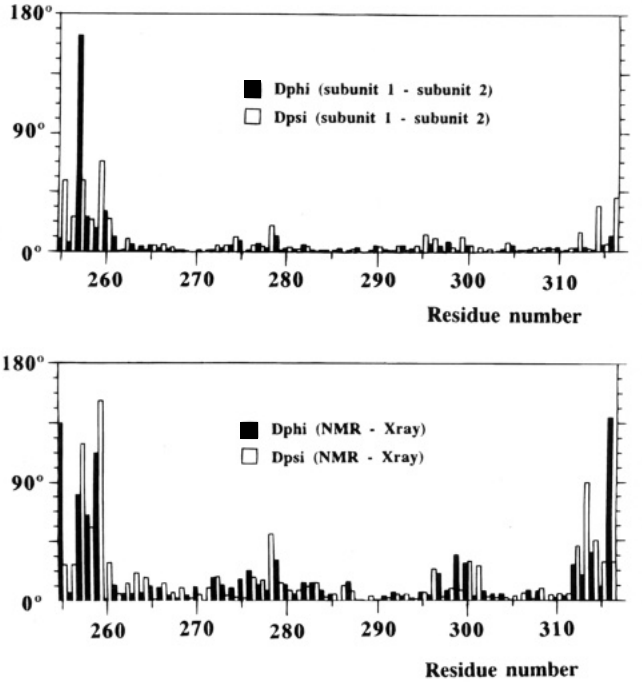

FIGURE 10: Top panel: Absolute values of the differences between the averaged torsional backbone angles ϕ and ψ in subunits 1 and 2. Bottom panel: Absolute values of the differences between a global average solution structure and the structure of the segment in the crystal state of intact thermolysin.

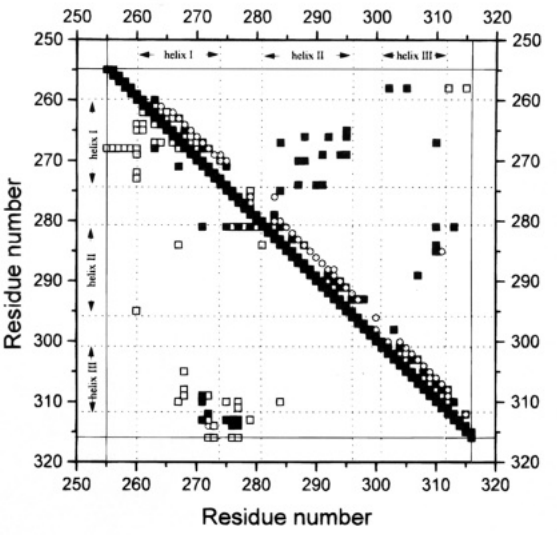

FIGURE 11: Contact diagonal plot. Up-diagonal region: intrasubunit correlations. Down-diagonal region: intersubunit correlations. Open circles: hydrogen bonds detected in at least four of the eight calculated structures. Open squares: distances less than 5 Å in one of the eight final structures between at least one proton in each of the correlated residues. Filled squares: observed interproton NOE cross correlations.

at the external surface of the dimer. Exceptions are Glu 302 and Ser 305, which, being located apparently at the interior of the dimer (accessible surface areas of only 1 and 5%, respectively), show a dynamic conformational averaging of their side chains. In this regard, it is to be noted that the accessible surface areas have been computed on an average structure, in which the N- and C-termini appear as rigid. Glu 302 and Ser 305 are in contact with the N-terminus of their own subunit, and very likely their solvent accessibility would be much larger had we considered the individual computed structures.

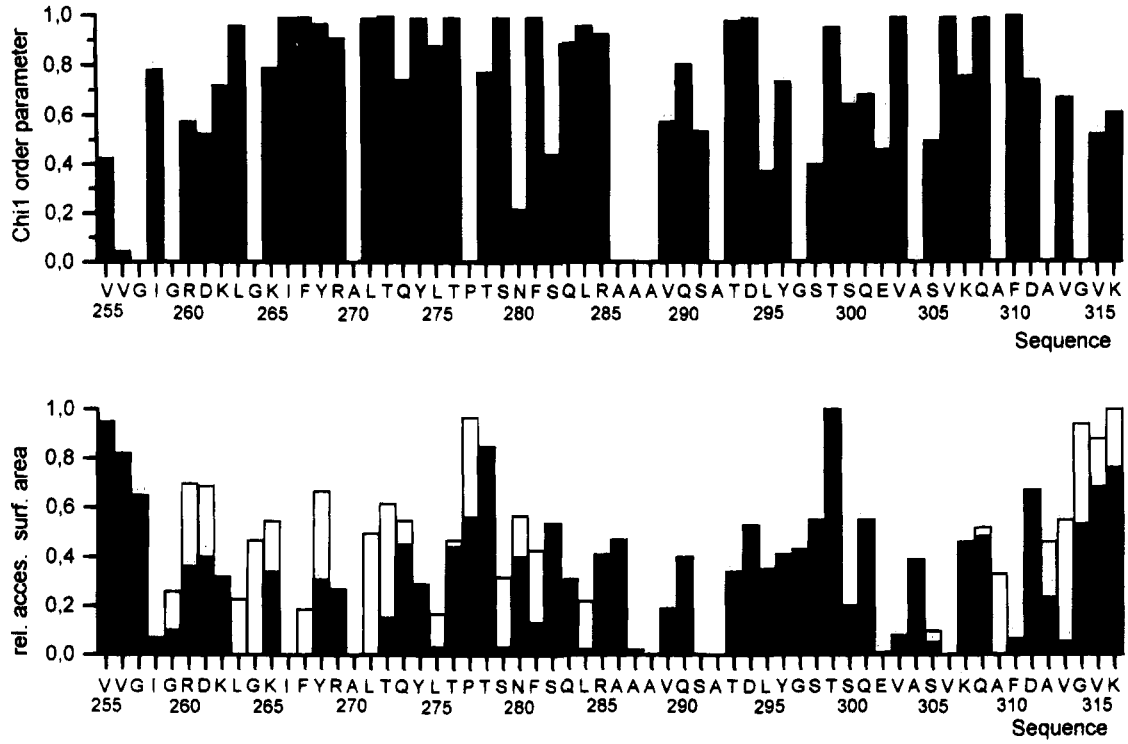

FIGURE 12: Top panel: Order parameters (Hyberts et al., 1992) around the torsion angle χ_1 along the fragment sequence. Bottom panel: Relative accessible surface areas (Richards, 1977) of the different residues along the sequence. Height of the filled bars: data for the dimer. Height of the open bars data for the monomer (when different).

A relatively large number of side chains, 28 out of 48, show some conformational constraints. Of 12 residues with a long side chain, five show a well-defined χ_1 (Lys 262, Leu 263, Lys 265, Gln 290, and Lys 307) and seven display a single conformation for the entire side chain (Ile 258, Ile 266, Arg 269, Leu 271, Leu 275, Leu 284, and Leu 295). Residues Ile 266, Arg 269, and Leu 295 are located at the interface of helices I and II within a given subunit; Leu 284 is located between helices II and III; and Ile 258, Leu 271, and Leu 275 are located at the interface between the two monomeric subunits. All of them, with the exception of Arg 269 and Leu 295, are solvent inaccessible, so they are located at the interior of the globular structure. Arg 269 is hydrogen bonded through its NH_ϵ to Tyr 274 OH, and this may be the reason for its preferred side-chain conformation. On the other hand, Leu 295 is solvent exposed (accessible surface area, 31%) and nevertheless shows a fixed conformation for its side chain. Some local steric hindrance arising from its location in the loop joining helices II and III must be at the origin of the observed facts.

Five aromatic residues (Tyr 268, 274, and 296 and Phe 287 and 310), seven residues with hydroxylic groups (Thr 272, 276, 293, and 299 and Ser 279, 291, and 300), and four valines (Val 289, 303, 306, and 313) also show well-defined side-chain conformations. In all cases, with the exception of Tyr 296 and Thr 299, the corresponding residues are either at the interior of the globular structure of the dimer or hydrogen bonded through their side chains. Tyr 296 and Thr 299, both solvent exposed and located at the loop joining helices II and III, may adopt fixed conformations of their side chains so as to optimize the solvation of their corresponding hydroxylic groups.

In summary, there are a relatively large number of conformationally restricted side chains both at the interior of each subunit and at the interface between the two subunits forming the dimer. Well-defined side-chain conformations in solvent-exposed residues are either involved in hydrogen bonds through a group in their side chains or conformationally restricted because of their location in turns or short loops.

DISCUSSION

Description and Quality of the Structure. Fragment 255−316 of thermolysin adopts a dimeric structure in aqueous solution at the concentration used in this NMR study, as is evident from the three-dimensional structure determined here. Each subunit of the dimer consists of three helices which span residues 260−274, 281−295, and 301−311 and two non-ideal β-turns covering residues 275−279 and 296−300 which respectively link helices I and II and helices II and III. The six N-terminal residues and the two C-terminal ones do not show any kind of ordered structure and are considerably more mobile than the rest of the chain. Excluding these disordered residues, the backbone structure is well-defined, having an average root mean square deviation among the 16 converged structures of 0.85 Å. When side chains are taken into account, the root mean square deviation for all residues increases to 3.02 Å. This relatively large value has its origin in the residues with mobile side chains, located mainly at the solvent-exposed surface. It is, on the other hand, a positive index of an efficient exploration of the conformational space. Energy values and distance constraint violations for the final dimeric structure are given in Table 2. The distance constraints are relatively well distributed along the fragment chain with the exception of the terminal segments. Many of the constraints involve pairs of protons far away in the sequence, consistent with a compact globular structure. The number of detected intersubunit NOE restraints, given in a table as supplementary material, is remarkably large, thus allowing a very precise and unambiguous docking of the two subunits. It should be noted

Table 2: Structural Statistics

No. of NOE Restraints per Subunit		distance (Å)		
type[a]	total	<3.5	<4.5	≥4.5
intra	291	69	135	87
SEQ	153	54	54	45
MBB	78	1	24	53
LBB	13	2	4	7
LGN	194	19	39	136
all intrasubunit	729			
intersubunit	52			

Residual Constraint Violations	
range (Å)	av no. of distance constraint violations
0.00−0.25	49.25
0.25−0.50	17.25
0.50−0.75	3.12
0.75−1.00	0.50
>1.00	0.00
max. violation (Å)	0.74
av sum of violations (Å)	15.20
rms deviation from ideality	
bond lengths (Å)	0.04
bond angles (deg)	1.72
impropers (deg)	2.57
total energy (kJ mol^{-1})	
av	−5433
range	−4520 to −6200
Lennard−Jones energy (kJ mol^{-1})	
av	−4575
range	−4395 to −4990

rms Deviations (Å)	NMR	X-ray
backbone atoms		
all residues	2.29	2.63
residues 261−314	0.85	0.78.69
all atoms	3.02	2.63
if rel acces. surf. area < 25% (56 residues)	1.26	1.69

[a] Abbreviations: intra, intraresidue; SEQ, sequential backbone or backbone−β; MBB, medium-range backbone−backbone or backbone−β; LBB, long-range backbone−backbone or backbone−β; LNG, all other interresidue NOEs.

that some of the intersubunit NOE constraints involve pairs of residues which are far away both in the sequence and in the three-dimensional structure of the corresponding subunit.

Comparison with the Structure of the Segment in the Crystal State of Native Thermolysin. The structure adopted by fragment 255−316 of thermolysin in each subunit of the dimer is extremely similar to the structure that this segment shows in the intact native thermolysin in the crystalline state. This was recognized in the early stages of the process as discussed above. The global shape, the overall main-chain fold, and even the side-chain position of many residues with a well-defined side-chain conformation in the solution structure match closely those found in the crystal structure. The average root mean square deviation of backbone atom positions in the single crystal state with respect to those in the average solution subunit structure is 0.78 Å for the structurally well-defined residues, while that for the eight solution structures is 0.85 Å. The location and extension of the three helices in the fragment in solution and in the crystal state are also markedly coincident (see Figure 3). In Figure 13 (bottom), the structure of the fragment has been superimposed on that of the same segment in the intact protein and compared with that adopted in the dimer (Figure 13, top). The close analogy between the two structures is apparent.

Analogies and differences between hydrogen bonds in the solution and crystal structures have been described in the Results. In general, hydrogen bonds in the helical segments are closely coincident. Differences appear in loop regions and in the N- and C-termini, where a greater number of hydrogen bonds, partially populated, are observed in the solution structures, thus reflecting the existence of different conformational states in those regions.

Dimeric Structure. Previous studies by sedimentation analysis in the analytical ultracentrifuge (Vita et al., 1989) had shown that fragment 255−316 of thermolysin forms a dimer at concentrations higher than 0.06 mM. Circular dichroism studies showed, however, that dimerization does not lead to alterations of the secondary structure of the fragment, thus signifying that the association process is not critical for dictating the folding of the monomer. Very recently, the energetics of the thermal unfolding of the fragment has been characterized by high-sensitivity differential scanning calorimetry (Conejero-Lara et al., 1994). It was found that the process is consistent with a two-state unfolding with no intermediate states, according to the scheme F2⇔2U, where F and U stand for the folded and unfolded state. The thermodynamic parameters of the process are characteristic of a compact globular structure for the dimer, and the energetics of the interactions at the contact surface between the two monomers appear to be significantly high.

The symmetrical dimeric structure of the 255−316 fragment of thermolysin, determined in this work, can be analyzed in full detail on the basis of Figures 8, 11, and 14. As shown pictorially in Figure 8 and in more detail in Figure 11, the subunit regions interacting with each other, so as to form the dimer, are mainly the following: (a) the loop between helices I and II in one subunit with the beginning of the second helix as well as with the C-end of helix III in the other subunit, (b) the C-terminal end of helix I with the C-terminal segment in the opposite subunit, and (c) the first residues in helix I with residues in the middle and end of the corresponding helix in the opposite subunit. The interface between the two subunits is the region with the most marked hydrophobic character of all the subunit surface. Not unexpectedly, this interface topologically coincides with the one which separates the 255−316 subdomain from the rest of the protein in the intact thermolysin (see Figure 13).

Figure 14 gathers all hydrophobic interactions, potentially significant from an energetic point of view, present at the interface between the two subunits, resulting from the calculated solution structures (note that the interface orientation is that corresponding to Figure 8). Symmetric hydrogen bonds between the amide proton of Phe 281 and the peptide carbonyl oxygen of Ser 279 in the opposite subunit have been marked in Figure 14. These hydrogen bonds were present in all eight calculated structures for the dimer. Another symmetric hydrogen bond (not shown), that between the NH of Gly 314 and the CO of Thr 272, present only in half of the converged structures, may also contribute to the dimer stability. Nevertheless, the largest energetic contribution must have a hydrophobic origin. Two symmetric pockets, each containing not fewer than nine hydrophobic side chains, can be readily appreciated in Figure 14. In particular, the side chain of Leu 271 is surrounded in an

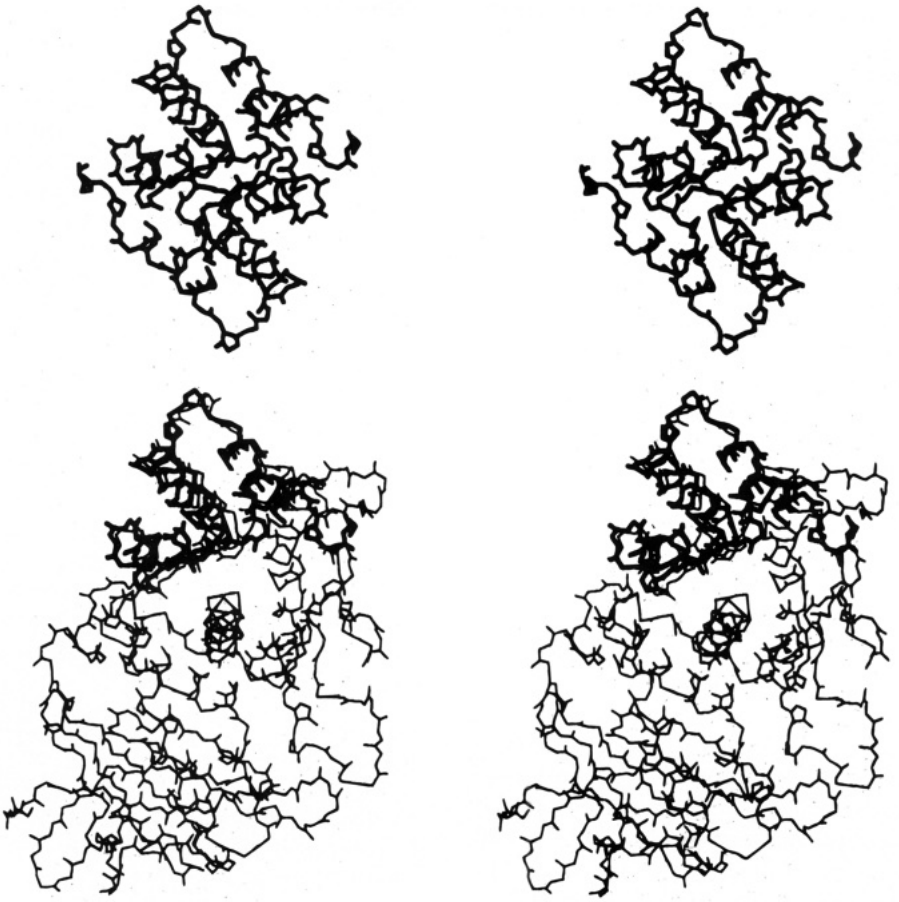

FIGURE 13: Stereoview of the dimer backbone structure (top) and superposition of one subunit (thick line) upon the structure of intact thermolysin (thin line) (bottom). Note the different orientation in relation to Figure 8.

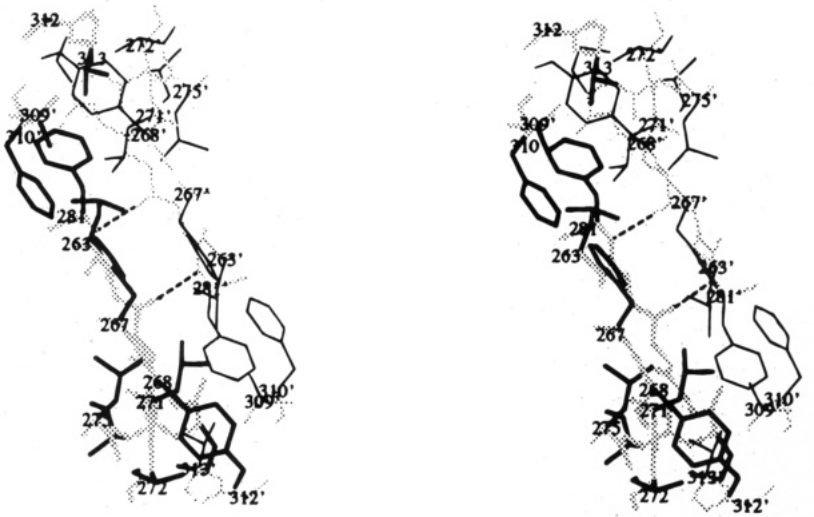

FIGURE 14: Stereoview of the dimer interface showing intersubunit hydrogen bonds and pockets of hydrophobic side chains. Dotted and continuous lines correspond respectively to backbone and side-chain bonds. Lines in the two subunits are distinguished by different thicknesses. Note that the orientation shown corresponds to that of Figure 8.

almost spherical arrangement by the side chains from residues Phe 267, Tyr 268, and Leu 275 belonging to the same subunit as well as by those from Leu 263, Phe 267, Phe 281, Phe 310, and Val 313 in the opposite subunit.

All residues present in the dimer interface have a well-defined side-chain conformation (root mean square deviations for χ_1 and χ_2 in the eight converged structures were lower than 30°, and in most cases they were lower than 20°). However, the conformations of some of them differ clearly from those adopted by the same residues in the intact protein. Thus Leu 263 and Leu 275 rearrange one of their side-chain torsion angles (χ_2 and χ_1, respectively), whereas Leu 271 changes both of them. The change of χ_1 in Phe 281 has the largest effect. As may be seen in Figure 14, the ring is folded toward the subunit to which it belongs. In the intact protein, however, the ring is oriented orthogonally to the interface of the segment and directed toward the rest of the protein, where it fills an otherwise empty volume. These facts confirm that steric packing at the interior of a protein is not like a rigid matching of two complementary interfaces, but

rather one interface may be coupled to two different ones by the rearrangement of a small number of side chains. The resulting interdigitation of side chains belonging either to the same subunit or to opposite subunits brings to mind the perfect organization of side chains in globular proteins. All these facts are in agreement with the conclusions about the dimerization energetics obtained by DSC (Conejero-Lara et al., 1994).

Implications for Protein Structure and Folding. The striking result of this study is that the folding of thermolysin fragment 255−316 in its dimeric state is essentially identical to that of the corresponding chain region in the X-ray structure of thermolysin (Holmes & Matthews, 1982). These structural similarities encompass many details of both secondary and tertiary structure. The dimeric state of the fragment appears to be stabilized largely by hydrophobic interactions, but it was earlier shown, on the basis of CD measurements of the monomer−dimer equilibrium (Vita et al., 1989), that fragment dimerization is not critical for dictating the folding and stability of the fragment. Thus, the present work put on a firm basis a previous proposal of a stable, nativelike structure of the 62-residue fragment 255−316, based on global estimates of structure derived from CD spectroscopy and immunochemical data (Dalzoppo et al., 1985; Vita et al., 1985, 1989; Fontana, 1990). Considering that this fragment was produced by proteolysis of the longer fragment 206−316, it was earlier inferred that fragment 255−316 could acquire a folded and rigid structure not amenable to further digestion (Dalzoppo et al., 1985). Since the independent folding of fragment 255−316 was also predicted on the basis of surface area measurements of the X-ray structure of thermolysin (Rashin, 1984), our results appear to support the validity of hierarchical analyses of protein structure and folding (Crippen, 1978; Rose, 1979; Zehfus, 1987; Wodak & Janin, 1981; Rashin, 1981). This work also firmly substantiates the proposal made first by Wetlaufer (1973) that fragments corresponding to protein domains could fold autonomously and, more important, that this property is confined to relatively short sequences corresponding to subdomains, i.e., folding units composed of a few elements of secondary structure which are definitely smaller than protein domains. Thus our results add weight to the idea that different parts of protein molecules can form nativelike structure independently (Kim & Baldwin, 1990).

The results of this study show that fragment 255−316 acquires in aqueous solution a remarkably well-defined tertiary structure, in agreement with its noteworthy thermal stability ($T_m \sim 70$ °C) (Conejero-Lara et al., 1994), which is quite striking for a rather short polypeptide chain devoid of disulfide bridges. Considering the size of fragment 255−316 and its helical secondary structure and well-defined side-chain packing, it is of interest to comment on the results of the present study in relation to those obtained by utilizing similar fragment systems. Recent attempts at the *de novo* design of helical proteins, such as the four-helix bundle, have produced molecules that show a folded and native structure, but lack unique side-chain packing (De Grado et al., 1989; Handel et al., 1993) and thus possess characteristics related to the molten globule state of proteins (Ptitsyn, 1992). Also the isolated helical domain of α-lactalbumin, produced by recombinant methods, forms a molten globule with the same overall tertiary fold as that found in the intact protein, but without specific side-chain packing interactions (Peng & Kim, 1994). Thermolysin fragment 255−316 in its dimeric state instead behaves as a small globular protein with a well-defined and fixed three-dimensional fold.

The concept that globular proteins are composed of domains and subdomains is a central feature of our understanding of protein structure. This modular architecture is generally observed with protein chains of at least 100 amino acid residues (Zehfus, 1987) and, in particular, in many much larger extracellular proteins involved in cell adhesion, clotting, and signaling (Campbell & Downing, 1994). It is obvious that a successful strategy for an understanding of both protein structure and function could be to dissect the proteins into their constituent modular units and to study these modules (domains and subdomains) in isolation. The results of this study show that individual modules in proteins can be identified by predictive methods and produced by protein fragmentation and that their structural properties can be studied in detail. Thus the complex problems of protein structure and folding can be simplified by analyzing proteins in parts. This approach of protein dissection is likely to provide useful advances in our understanding of the properties and functions of proteins (Peng & Kim, 1994; Kippen et al., 1994).

SUPPLEMENTARY MATERIAL AVAILABLE

One table listing observed intersubunit NOEs with an indication of the evaluated upper limit distance constraint (1 page). Ordering information is given on any current masthead page.

REFERENCES

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229−2246.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.

Breg, J. N., van Ophensden, J. H. J., Burgering, J. H. J., Boelens, M. J. M., & Kaptein, R. (1990) *Nature 346*, 586−589.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488−490.

Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285−297.

Campbell, I. D., & Downing, A. K. (1994) *Trends Biotechnol. 12*, 168−172.

Conejero-Lara, F., De Filippis, V., Fontana, A., & Mateo, P. (1994) *FEBS Lett. 344*, 154−156.

Creighton, T. E. (1990) *Biochem. J. 270*, 1−15.

Creighton, T. E., Ed. (1992) *Protein Folding*, W. H. Freeman & Co., New York.

Crippen, G. M. (1978) *J. Mol. Biol. 126*, 315−332.

Dalzoppo, D., Vita, C., & Fontana, A. (1985) *J. Mol. Biol. 182*, 331−340.

De Grado, W., Wassermann, Z., & Lear, J. (1989) *Science 243*, 622−628.

Dill, K. A. (1985) *Biochemistry 24*, 1501−1509.

Dobson, C. M., Evans, P. A., & Radford, S. E. (1994) *Trends Biochem. Sci. 19*, 31−37.

Fontana, A. (1990) in *Peptides: Chemistry, Structure and Biology* (Rivier, J. E., & Marschall, G., Eds.) pp 557−565, ESCOM, Leiden.

Günter, P., Braun, W., & Wüthrich, K. J. (1991) *J. Mol. Biol. 217*, 517−530.

Handel, T. M., Williams, S. A., & De Grado, W. F. (1993) *Science 261*, 879−885.

Holmes, M. A., & Matthews, B. W. (1982) *J. Mol. Biol. 160*, 623−639.

Hyberts, S. G., Golberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. A. (1979) *J. Chem. Phys. 71*, 4546−4553.
Jiménez, M. A., Bruix, M., González, C., Blanco, F. J., Nieto, J. L., Herranz, J., & Rico, M. (1993) *Eur. J. Biochem. 211*, 569−581.
Kay, L. E., Forman-Kay, J. D., McCubbin, W. D., & Kay, C. M. (1991) *Biochemistry 30*, 4323−4333.
Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459−489.
Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.
Kippen, A. D., Sancho, J., & Fersht, A. R. (1994) *Biochemistry 33*, 3778−3786.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
Levinthal, C. (1968) *J. Chim. Phys. Phys. Chim. Biol. 65*, 44−45.
Levitt, M., & Freeman, R. (1979) *J. Magn. Reson. 33*, 473−476.
Macura, A., & Ernst, R. R. (1980) *Mol. Phys. 41*, 95−117.
Marion, D., & Wüthrich, K. (1983) *J. Magn. Reson. 79*, 352−356.
Miranker, A., Radford, S. E., Karplus, M., & Dobson, C. M. (1991) *Nature 349*, 633−636.
Nilges, M. (1993) *Proteins 17*, 297−309.
Oas, T. G., & Kim, P. S. (1988) *Nature 336*, 42−48.
Peng, Z.-Y., & Kim, P. S. (1994) *Biochemistry 33*, 2136−2141.
Ptitsyn, O. B. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 243−300, W. H. Freeman & Co., New York.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302−307.
Rance, M. (1987) *J. Magn. Reson. 74*, 557−564.
Rashin, A. A. (1981) *Nature 291*, 85−86.
Rashin, A. A. (1984) *Biochemistry 23*, 5518−5519.
Redfield, A. G., & Kuntz, S. D. (1975) *J. Magn. Reson. 19*, 250−254.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Rico, M., Santoro, J., González, C., Bruix, M., Neira, J. L., Nieto, J. L., & Herranz, J. (1991) *J. Biomol. NMR 1*, 283−298.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.
Rose, G. D. (1979) *J. Mol. Biol. 134*, 447−470.
Santoro, J., González, C., Bruix, M., Neira, J. L., Nieto, J. L., Herranz, J., & Rico, M. (1993) *J. Mol. Biol. 229*, 722−734.
Shin, H.-C., Merutka, G., Waltho, J. P., Tennaut, L. L., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 6356−6364.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Titani, K., Hermodson, M. A., Ericsson, L. H., Walsh, K. A., & Neurah, H. (1972) *Nature 238*, 35−37.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694−699.
Van Gunsteren, W. F., & Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands.
Vita, C., Fontana, A., & Chaiken, I. M. (1985) *Eur. J. Biochem. 151*, 191−196.
Vita, C., Fontana, A., & Jaenicke, R. (1989) *Eur. J. Biochem. 183*, 513−518.
Wagner, G. (1983) *J. Magn. Reson. 55*, 151−156.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697−701.
Wetlaufer, D. B. (1981) *Adv. Protein Chem. 34*, 61−62.
Wodak, S. J., & Janin, J. (1981) *Biochemistry 20*, 6544−6552.
Wright, P. E., Dyson, H. J., & Lerner, R. A. (1988) *Biochemistry 27*, 7167−7175.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley & Sons, New York.
Zehfus, M. H. (1987) *Proteins: Struct., Funct., Genet. 2*, 90−110.