Structural Investigation of the HIV-1 Envelope Glycoprotein gp160 Cleavage Site

Romina Oliva,[a] Marilisa Leone,[b] Lucia Falcigno,[a, b] Gabriella D'Auria,[a, b] Monica Dettin,[c] Claudia Scarinci,[c] Carlo Di Bello,[c] and Livio Paolillo*[a, b]

Abstract: The selective proteolytic activation of the HIV-1 envelope glycoprotein gp160 by furin and other precursor convertases (PCs) occurs at the carboxyl side of the sequence Arg508-Glu-Lys-Arg511 (site 1), in spite of the presence of another consensus sequence: Lys500-Ala-Lys-Arg503 (site 2). We report on the solution structural analysis of a 19-residue synthetic peptide, p498, which spans the two gp160-processing sites 1 and 2, and is properly digested by furin at site 1. A molecular model is obtained for p498, by means of molecular dynamics simulations, from NMR data collected in trifluoroethanol/water. The peptide N-terminal side presents a 9-residue helical segment, enclosing the processing site 2; the C-terminal segment can be described as a loop exposing the processing site 1. A hypothesis for the docking of p498 onto the catalytic domain of human furin, modeled by homology and fitting previous site-directed mutagenesis studies, is also presented. p498 site 1 is shown to have easy access to the furin catalytic site, unlike the nonphysiological site 2. Finally, on the basis of available data, we suggest a possible structural motif required for the gp160–PCs recognition.

Keywords: conformation analysis · HIV · molecular modeling · proteases · protein structures

Introduction

The HIV envelope glycoprotein gp160 is synthesized as an inactive precursor. It is then cleaved, in a late Golgi compartment,[1] to give the noncovalently associated glycoproteins gp120 and gp41. The cleavage of the HIV envelope glycoprotein gp160 is a prerequisite for the virus infectivity. Indeed, it is gp120 that mediates the interaction of the viral particle with CD4 receptor–coreceptor complexes, while gp41 carries out the fusion of viral and cellular lipid membranes.[2]

The endoprotease candidates for the intracellular processing of the HIV envelope glycoprotein are the precursor convertases (PCs), newly discovered mammalian subtilisin-like serine proteases. Furin was the first PC enzyme shown to cleave gp160 intracellularly into gp120 + gp41.[3] Later the involvement in this process of other mammalian PCs (PC5/6/7)[4] and, most recently, of other protease families[5] has been proposed.

Analogously to prohormones, viral glycoproteins have multibasic cleavage sites, and in particular consensus sequences of the type Lys/Arg-Xaa-Lys/Arg-Arg. From activity studies it is known that the proteolytic cleavage of the HIV-1 glycoprotein gp160 precursor occurs at the carboxyl side of the sequence Arg508-Glu-Lys-Arg511 (site 1).[6] The basic residues within the cleavage recognition sequence are crucial in determining the efficiency of the proteolytic processing, as confirmed by site-directed mutagenesis studies.[7] Many data indicate, however, that the amino acid sequence of site 1 alone is not sufficient to allow the correct processing of HIV-1 gp160. Strikingly, gp160 cleavage occurs selectively at site 1, in spite of the presence of a consensus motif next to site 1, corresponding to the Lys500-Ala-Lys-Arg503 sequence (site 2). It has been shown that the basic amino acids of site 2 are also important for the proteolytic processing at site 1; indeed, multiple replacements of the site 2 basic residues drastically reduce the gp160 cleavage.[8]

To date, no structural information is available for a large sequence around the gp160 cleavage sites (about 50 residues).[9] Brakch et al.[10] showed that two peptides, each spanning one only of the two gp160 consensus sequences

[a] Prof. L. Paolillo, Dr. R. Oliva, Dr. L. Falcigno, Prof. G. D'Auria
Dipartimento di Chimica, Università di Napoli "Federico II"
Complesso Universitario Monte S. Angelo, via Cintia
80126 Napoli (Italy)
Fax: (+39) 081-674090
E-mail: paolillo@chemistry.unina.it

[b] Prof. L. Paolillo, Dr. M. Leone, Dr. L. Falcigno, Prof. G. D'Auria
Centro di Studio di Biocristallografia del C.N.R.
Università di Napoli "Federico II", Via Mezzocannone 4
80134 Napoli (Italy)

[c] Dr. M. Dettin, Dr. C. Scarinci, Prof. C. Di Bello
Dipartimento di Processi Chimici dell'Ingegneria
Università di Padova, Via Marzolo 9
35131 Padova (Italy)

Supporting information for this article is available on the WWW under http://wiley-vch.de/home/chemistry/ or from the author (NMR structure determination statistics, proton–proton distances from NOESY experiments and RMD calculations, $^3J_{NH,\alpha CH}$ coupling constants).

  0947-6539/02/0806-1467 $ 17.50+.50/0

(site 1 and site 2), were both cleaved by furin and other PCs. Very interestingly, however, the 19-residue peptide p498, spanning the junction between gp120 and gp41, and containing sites 1 and 2 (Figure 1), is recognized and properly digested by furin at the site 1.[10,11] These data confirm that contributions of surrounding amino acids are required to promote the correct gp160-like processing, in addition to the multibasic consensus sequence. Furthermore they suggest that such contributions, as specific secondary structure motifs, can be reproduced in a medium-sized peptide fragment like p498. Therefore the activity findings for the p498 peptide have prompted the study of other synthetic peptides, mimicking precursor processing sites of HIV-1 gp160, HIV-2 gp140, and SIV gp140.[12]

To clarify the structural features that allow the correct gp160-like processing to be performed by furin on the p498 peptide, here we report on the structural analysis of the peptide, carried out in solution by means of NMR techniques. Then we present a hypothesis for the docking of the substrate p498 in the catalytic domain of the processing enzyme furin, built by homology and matching previous homology[13] and site-directed mutagenesis studies.[14] Finally, on the basis of our experimental results and of gp160 secondary-structure predictions, we suggest a hypothesis concerning the secondary structure motif required for the molecular recognition between gp160 and the PCs processing enzymes.

Table 1. Proton chemical shifts (δ) of p498 in TFE/H_2O (90:10 v/v) at 298 K with TSP as internal reference.

Residue	HN	Hα	Hβ	Hγ	Others
Pro1	–	4.52	2.15, 2.54	2.11	Hδ 3.50, 3.40
Thr2	8.28	4.52	4.45	1.32	
Lys3	8.27	4.19	1.86	1.48, 1.61	Hδ 1.78 Hε 3.12
Ala4	7.85	4.18	1.46		
Lys5	7.65	4.09	1.92	1.52	Hδ 1.75 Hε 3.03
Arg6	7.91	4.04	1.95	1.72	Hδ 3.21 NHε 7.17
Arg7	7.82	4.14	2.02	1.75, 1.86	Hδ 3.23 NHε 7.20
Val8	7.74	3.72	2.30	1.01, 1.11	
Val9	8.04	3.75	2.20	1.01, 1.11	
Gln10	8.14	4.04	2.21, 2.28	2.45, 2.60	NHε 6.32, 7.06
Arg11	7.93	4.10	2.09	1.75, 1.88	Hδ 3.22 NHε 7.11
Glu12	8.38	4.13	2.31	2.50, 2.69	
Lys13	8.39	4.13	2.00	1.52	Hδ 1.71 Hε 2.97
Arg14	7.88	4.24	1.98	1.78, 1.83	Hδ 3.24 NHε 7.06
Ala15	7.99	4.26	1.54		
Val16	7.71	4.14	2.26	1.04	
Gly17	7.85	3.90, 4.05			
Ile18	7.63	4.16	1.95	1.24, 1.58	CH3γ 0.97 Hδ 0.92
Gly19	7.93	3.92			CONH 6.51, 7.16

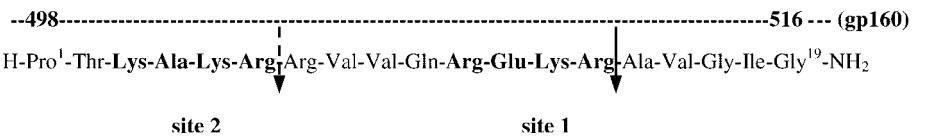

Figure 1. Amino acid sequence of the p498 synthetic peptide; p498 spans the gp160 sequence from Pro498 to Gly516, containing the two processing sites 1 and 2 (in bold). Solid and broken arrows are used to indicate the cleavage loci for the recognized processing site 1 and the secondary site 2, respectively.

Figure 2. Summary of the most relevant NOE effects in TFE/H_2O (90/10).

Results

p498 Conformational analysis: In aqueous solution containing 2,2,2-trifluoroethanol (TFE), p498 CD spectra present three bands (one intense and positive at 190 nm and two negative at 205 nm and 222 nm) characteristic of right-handed α-helical conformations (data not shown).

The NMR analysis was performed in a TFE/H_2O (90/10) solution. The complete sequential assignment was achieved according to the standard procedure proposed by Wüthrich.[39] Proton chemical shifts are listed in Table 1.

The NOE pattern, a series of strong NH_i-NH_{i+1} NOE connectivities, together with α_i-NH_{i+3} NOEs and the $\alpha_i-\beta_{i+3}$ contact between Arg7 and Gln10, points to a helical conformation in the *N*-terminal/middle region of the peptide (Figure 2). The helix hypothesis is further supported by the proton chemical shift deviations from the random coil values (Figure 3). In the segment from Lys3 to Arg14, the diagram of the αCH deviations vs. residue number does indeed show a trend which is typical of helical secondary structures.[15] A set of 126 experimental constraints from NOE data (67 intra-residual, 35 sequential and 24 medium-range, including 17 constraints characteristic of helical structure) was determined for structure calculations. These constraints have been

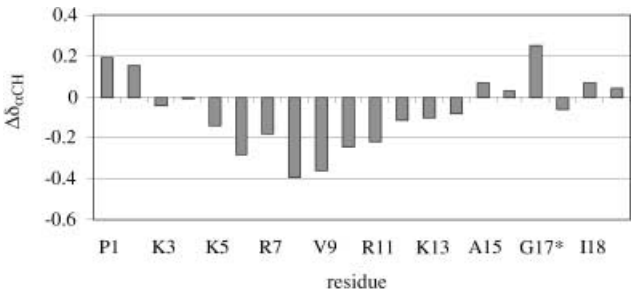

Figure 3. Chemical shift deviations from the random coil values for αCH protons.[15] For Gly17* both the αCH and α'CH chemical shift deviations are reported.

0947-6539/02/0806-1468 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 6

imposed as upper bounds on interproton distances, for sampling the conformational space allowed to the peptide, by means of torsion angle dynamics (DYANA program[16]).

A first set of 50 conformers was calculated with the DYANA program. All of these conformers agree well with experimental constraints: they do not violate 93 out of the 126 constraints, while the best 34 DYANA structures show no more than three violations larger than 0.2 Å from the experimental restraints. 50 additional structures were then calculated with the redundant dihedral angle constraints (REDAC) strategy, in order to improve the convergence.[17] The REDAC structures are indeed more compatible with experimental data. All of these structures never violate 117 out of the total 126 experimental restraints by more than 0.2 Å. All the calculated DYANA structures exhibit a 4–6-residue helical segment in the region from Arg7 to Glu12, while their N- and C-terminal segments appear rather flexible.

A set of eight out of the 50 structures calculated with the REDAC strategy was chosen, by means of a RMSD cluster analysis,[18] as representative of the conformational space allowed to the peptide, together with the best one obtained with the direct strategy (Figure 4a). Such structures were then

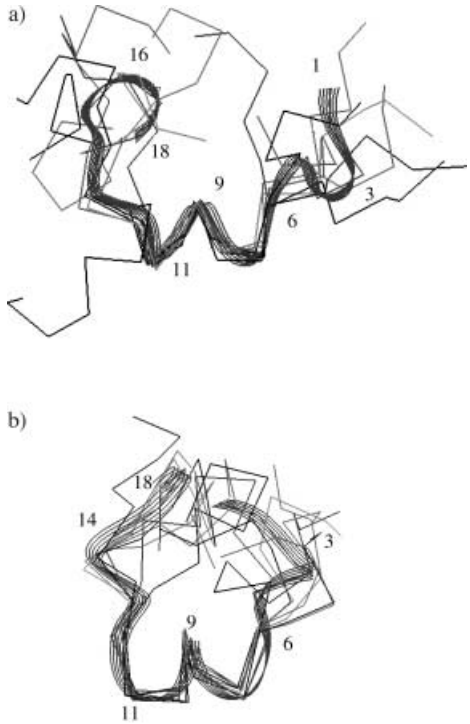

Figure 4. The best trace superposition of the Lys5–Lys13 segment in nine selected DYANA structures, before (a) and after (b) the RMD simulations. The same gray scale is used for corresponding structures. The best REDAC structure is represented by a ribbon of lines.

subjected to restrained molecular dynamics (RMD) simulations while the experimental restraints were applied. After RMD simulations (Figure 4b), the helical segment of the starting conformers is preserved. Eight out of nine calculated models show a regular α-helix extending from Val8 to Arg11. This helix is stabilized by the backbone hydrogen bonds of the $CO_i - NH_{i+3}$ and $CO_i - NH_{i+4}$ type and by those formed by the positively charged side chain of Lys13 with the C-terminal carbonyl oxygens (helix capping). A 3_{10}-helix, from Lys3 to Lys5, is present in seven out of nine models. However, the residues Arg6 and Arg7, flanking the two helical segments, are less well defined. They are either enclosed within the helical segments, or assume angles representative of noncanonical turns. The N-terminal tail of the peptide is less defined; in fact a smaller number of constraints is present for this region.

The C-terminal segment from Glu12 to Gly19, containing the processing site, does not show regular secondary structure elements and can be considered as an Ω-loop. However, after RMD simulations, its flexibility appears reduced. It assumes a quite defined shape (see Figure 4b), with bending causing the exposure of the cleavage point residues (Arg14 and Ala15), and, in five out of the nine calculated structures, residues Gly17 and Ile18 exhibit dihedral angles falling in the α_R region.

The best RMD molecular model with respect to the fit with experimental data and the Ramachandran plot is shown in Figure 5. It has a 9-residue helical structure extending from Lys3 to Arg11 (3_{10}-helix for Lys3-Lys5 and α-helix for Arg6-Lys11) and a loop in the Glu12-Gly19 sequence, exposing the processing site 1. This model can be considered quite representative of the whole set of structures obtained after RMD, and it has been used for study of the peptide docking onto the catalytic domain of the furin enzyme.

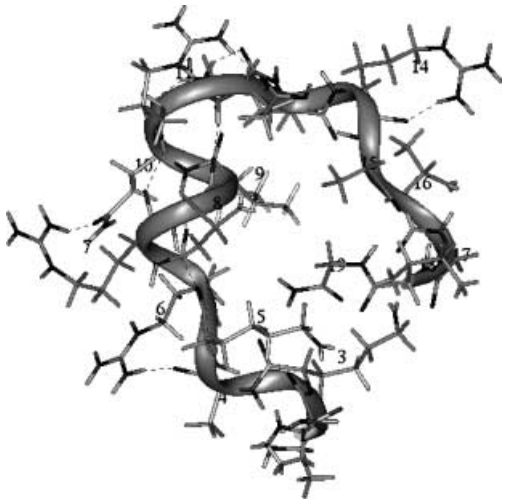

Figure 5. Ribbon and stick representation of the best RMD average molecular model for p498. Hydrogen bonds are shown by broken lines.

Model of the furin catalytic domain complexed to a model substrate containing the consensus sequence Arg-Glu-Lys-Arg: A three-dimensional model of the catalytic domain of human furin in a complex with a model substrate containing site 1 was built by homology on the basis of the crystal structure of thermitase complexed to the inhibitor eglin-c.[19] The sequence alignment obtained by Siezen et al.[20] and verified by site-directed mutagenesis studies[14] was employed. The side chains of the eglin-c residues 40–47 were substituted to resemble appropriate P6–P2′ residues of gp160, namely, Val-Gln-Arg-Glu-Lys-Arg ↓ Ala-Val. (Nomenclature refers to Schechter et al.:[21] residues before the cleavage point are

numbered consecutively, starting from the C-terminal one, as P1, P2 etc.; residues after the cleavage point, starting from the N-terminal one, as P1′, P2′, etc.) Predicting enzyme – substrate interactions for furin does indeed have a reasonable basis in the observation that many other subtilases appear to retain essentially the same mode of substrate/inhibitor binding.[13]

The complex is shown in Figure 6. Furin exhibits, as expected, all the secondary structure elements, α-helices and β-sheets, characteristic of the subtilase family. Eight substrate

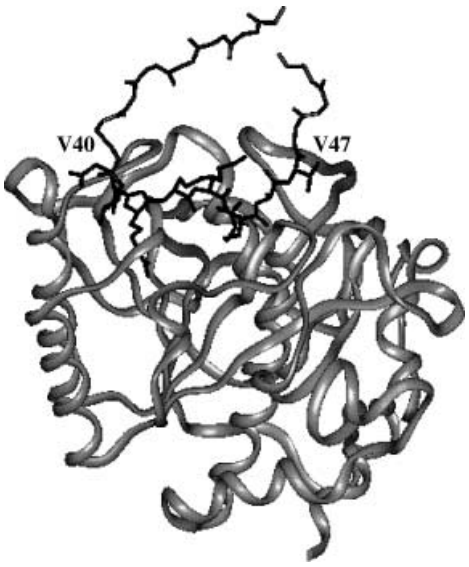

Figure 6. Homology model of furin catalytic domain (light gray ribbon) complexed with a model substrate containing site 1. The model substrate was obtained by modifying the inhibitor eglin-c to present the gp160-like sequence Val40-Gln-**Arg-Glu-Lys-Arg** ↓ Ala-Val47 in the P6 – P2′ positions. A 19-residue segment around the cleavage site of the model substrate, modified eglin, is shown as dark gray sticks. The N- and C-terminal residues of the modified sequence, shown in black with side chains, are labeled.

residues around the multibasic sequence (P6 – P2′) are accommodated in an enzyme-binding region that may be described as a surface channel (see Siezen et al.[13]). The furin catalytic residues (Ser225, His71, Asp38 and Asn163) are located at the bottom, together with several surface-exposed acidic residues. Large contributions to the binding of the model substrate (modified eglin) to the enzyme are apparently made by specific electrostatic interactions between each of the basic amino acids of the multibasic substrate sequence with the acidic residues belonging to the enzyme catalytic region. In particular, we observed interactions of the substrate Arg45 and Lys44 with furin Asp199 and Asp47, respectively, and of Arg42 with Glu129. The specific interactions we observed are fully consistent with site-directed mutagenesis studies performed on furin by Creemers et al.[14]

Attempts were made to dock p498 to furin by using the best RMD molecular model for the peptide conformation. Suitable peptide docking was obtained by inserting the peptide site 1 in the furin catalytic region. In particular, the backbone of four residues around the p498 physiological cleavage point (Glu12 – Ala15) was superimposed on the backbone of the corresponding residues (Glu43 – Ala46) in the homology-modeled substrate (Figure 7). In such an arrangement, the entire peptide is indeed nicely accommodated in the furin-binding region without any modification required in its backbone. In particular, the C-terminal side of site 1 (Arg11-Glu-Lys-Arg14), on which the proteolytic activation occurs, makes a very good match with the furin catalytic site (Ser225, His71, Asp38 and Asn163), with the Arg14 carboxy carbon atom sitting at a catalytic distance (3.0 Å) from the hydroxy oxygen atom of Ser225.

Analogous attempts were then made to dock the peptide by inserting the site 2 (Lys3-Ala-Lys-Arg6) into the furin catalytic site. In that case, the results were very different: in fact, no access seems possible for the p498 site 2 to furin catalytic residues in the conformation observed for the peptide.

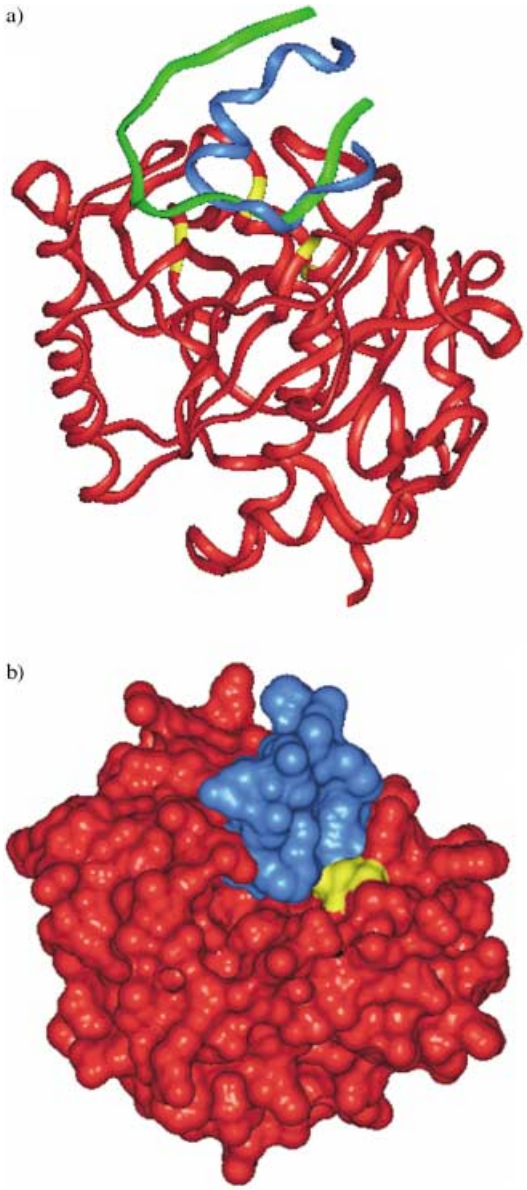

Figure 7. Docking of the p498 best RMD model in furin catalytic domain. The backbone of p498 Glu12-Ala15 residues around the physiological processing site has been superimposed on the corresponding one of modified eglin (Glu43-Ala46). p498, modified eglin, and furin are shown in blue, green, and red, respectively. Furin catalytic residues (Ser225, His71, Asp38, and Asn163) are shown in yellow. a) Ribbon representation. b) Connolly surface representation of only p498 and furin catalytic domain.

0947-6539/02/0806-1470 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 6

Discussion

The HIV-1 envelope glycoprotein gp160 is intracellularly processed into the gp120 and gp41 glycoproteins by the newly discovered mammalian subtilisin-like PC enzymes such as furin and PC1. PC enzymes recognize the sequence Arg-Xaa-Lys/Arg-Arg. However, the proteolytic cleavage of the HIV-1 glycoprotein precursor gp160 occurs selectively at the carboxyl side of the sequence Arg508-Glu-Lys-Arg511 (site 1), in spite of the presence of another area of potential enzyme sensitivity next to site 1, with the sequence Lys500-Ala-Lys-Arg503 (site 2). These data suggest that other factors, as specific secondary structure motifs, are required in addition to the consensus sequence Arg-Xaa-Lys/Arg-Arg for the molecular recognition.

We have reported here on the structural analysis in solution of a 19-residue synthetic peptide, p498, that contains both the gp160 processing sites 1 and 2, and that is properly digested by furin at the physiological processing site of gp160: site 1.[10,11] The functional data thus suggest that the structural motifs participating in the recognition of gp160 by furin are significantly preserved in the p498 peptide. Therefore such a gp160 fragment has been assumed to be a suitable model for investigating the structural requirements for the processing of the gp160 glycoprotein.

NMR and CD techniques were employed to analyze the peptide structural features. CD spectra showed the peptide propensity to adopt an α-helical structure in the presence of TFE. NMR analysis performed in the same medium, TFE/water, confirms and completes these results. TFE is well known to favor solvent-shielded amide conformations, thus promoting ordered structures. We chose to employ such a cosolvent, since it is also known from literature that the local conformation of native proteins is usually better reproduced by the conformation assumed by corresponding fragments in solutions of TFE/water[22–24] than in pure water.[25]

The molecular model obtained from NMR data by means of DYANA[16] and RMD simulations (Figure 5) shows a helical segment at the N-terminal side, from Lys3 to Arg 11, with the processing site 2 enclosed in it. The C-terminal side, from Lys13 to Gly 19, is disordered and can be described as a loop that presents a bend exposing the processing site 1.

Such findings (namely, site 1 located in a loop, next to a helix, and secondary processing site located in the helix), perfectly agree with current hypotheses about secondary structure of multibasic cleavage sites in other precursors. Indeed, several studies suggest that dibasic sites of prohormones could be present in bending regions, like turns[26] or loops,[27] because these can better adapt to the active sites of the processing enzymes. Recently a loop was proposed as crucial for the interaction of a gp140 fragment of the SIV virus, containing the consensus motif Arg-Xaa-Lys-Arg, with the processing enzyme furin.[12] Furthermore, several furin-substrate precursors (the toxin proaerolysin, diphtheria toxin, the hemagglutinin precursor of influenza virus) have been shown to be cleaved within a flexible loop.[28–30]

In addition, secondary structure predictions and experimental studies on prohormones support the occurrence of ordered secondary structures, such as α-helices, flanking the physiological sites.[31–34] Such ordered structures could play a key role in regulating the exposition and accessibility of the cleavage sites. This is the case for the furin substrate toxin proaerolysin, whose cleavage site is exposed to the surface at the end of a helical segment.[28] In addition, analogously to our findings about p498 site 2, secondary processing sites of prohormones have been hypothesized to be preferentially enclosed in more rigid structures (i.e., helices), which are less accessible to the processing enzymes.[35]

Accessibility therefore seems to be the main factor for the preferential recognition of physiological sites. In order to test the accessibility of the p498 sites 1 and 2 to the enzyme, we performed docking experiments on the p498 NMR molecular model in the catalytic domain of human furin, modeled by homology. The docking attempts indeed showed that, unlike site 2, the p498 site 1 does not need any drastic rearrangement of the peptide backbone to fit the furin binding channel (Figure 7). These findings may explain the exhibited preference of furin in cleaving p498 at site 1. On the basis of the predicted interactions for furin with the homology-modeled substrate (Figure 6), it can, however, be hypothesized that, when approaching the binding site, the peptide arranges itself to optimize both the van der Waals and the electrostatic energetic interactions. In particular, electrostatic interactions between the basic residues of the consensus sequence and the acidic residues located in the furin binding site are expected to promote the recognition, on analogy with the observations from the enzyme/modified-eglin model.

To date the secondary structure motifs required for the molecular recognition between HIV-1 gp160 and PC processing enzymes have not been clearly assessed. As the structure of the gp120/gp41 junction remains to be elucidated, current hypotheses can only rely on conformational studies of model peptides and protein structure predictions. A loop has already been suggested as a common structural feature for physiological cleavage sites in retroviral envelope glycoproteins.[5,12] In the peptide p498 we found an N-terminal helix, enclosing the secondary processing site 2, which could play a key role in regulating the exposition and accessibility of the physiological cleavage site, although it cannot be excluded that, on recognition, the helix partially unfolds in order to optimize energetic interactions between the enzyme and the substrate. In addition, gp160 secondary structure predictions, obtained from the PHDsec server,[36] locate with a very high reliability index an α-helix at the *C*-terminal side of the loop, starting from Ala517.

Altogether, these data point to a helix–loop–helix structural motif. To date at least one natural substrate of furin, diphtheria toxin (DT), whose crystal structure is known,[37] has been shown to be indeed cleaved by the enzyme on a loop, flanked by two α-helices.[29]

Experimental Section

Peptide substrate: The peptide synthesis, purification, and characterization have been described elsewhere.[10]

CD analysis: CD spectra were recorded at room temperature on a Jasco Model J710 automatic recording circular dichrograph. A cylindrical fused

quartz cell (0.1 cm path length) was used. A weighed quantity of p498 was dissolved in TFE/H_2O solution. Peptide concentration (1.6×10^{-5} M) was determined by quantitative amino acid analysis in the TFE/H_2O 90:10 (*v/v*) solution.

NMR analysis: NMR experiments were carried out on a Varian Unity 400 spectrometer equipped with a SPARC station 330, located at the Centro di Studio di Biocristallografia C.N.R., University of Naples "Federico II". NMR characterization was performed in TFE/H_2O 90:10 (v/v) at 298 K. The sample was prepared by dissolving the peptide (≈6.7 mg) in [D_3]TFE (0.75 mL, 99% isotopic purity, Aldrich) and H_2O (0.075 mL). Chemical shifts were referred to internal 2,2′,3,3′-[D_4]3-(trimethylsilyl) sodium propionate (TSP). Two-dimensional (2D) experiments, such as total correlation spectroscopy (TOCSY),[38] nuclear Overhauser effect spectroscopy (NOESY),[39] rotating frame Overhauser effect spectroscopy (ROESY),[40] double quantum-filtered correlated spectroscopy (DQFCOSY)[41] were recorded by the phase-sensitive States–Haberkorn method. The data file generally consisted of 512 and 2048 (4096 for DQFCOSY) data points in the $\omega 1$ and $\omega 2$ dimensions, respectively. TOCSY results were acquired with a 70 ms mixing time, NOESY experiments with an 80, 150, 250, 300, 350 ms mixing time, and ROESY experiments with a 100 ms mixing time, using a continuous spin-lock. Off-resonance effects caused by the low power spin-lock field were compensated by means of two 90° hard pulses before and after the spin-lock period.[42] Free induction decays (FIDs) were multiplied in both dimensions with shifted sine-bell weighting functions, and data points were zero-filled to 1 K in $\omega 1$ prior to Fourier transformation. The water resonance was suppressed by low-power irradiation during the relaxation delay and, for NOESY, during the mixing time. Temperature coefficients of amide protons were measured from one-dimensional (1D) spectra and from TOCSY spectra, acquired with 4096 data points, in the 298–310 K temperature range. NOESY spectra were used for NOE analysis; NOE intensities were evaluated by integration of cross-peaks by the appropriate VARIAN software and then converted into interproton distances using the $1/r^6$ relationship for rigid molecules.[39] Geminal $\gamma - \gamma'_{CH_2}$ protons of Glu12 and Gln10 were chosen as the reference with a distance of 1.78 Å. The vicinal $^3J_{NH,\alpha CH}$ coupling constants were measured from DQFCOSY spectra.

Following Wüthrich,[39] amino acid spin systems were identified by comparison of TOCSY and DQF-COSY, while sequential assignment was obtained by the analysis of NOESY. In order to solve residual ambiguities, additional TOCSY and NOESY spectra were recorded on a 600 MHz spectrometer by courtesy of Pharmacia Upjohn, Nerviano (MI, Italy).

Computational analysis: Conformational analysis was performed on a Silicon Graphics INDIGO2 Workstation. Starting models were calculated by the DYANA (dynamics algorithm for NMR applications) program (version 1.2) and they were then refined by restrained molecular dynamics (RMD), according to a standard procedure.[43]

DYANA calculations: Torsion-angle dynamics calculations were carried out by the DYANA program.[16] The library program was modified for the N- and C-terminal residues. A total of 100 three-dimensional structures were obtained with the interproton distances evaluated from NOEs (raised by 30%) as upper limits, without the use of stereospecific assignments. 50 conformers were calculated with the standard parameters of the DYANA program. To improve convergence, the redundant dihedral angle constraints (REDAC) strategy[17] was also employed and 50 more structures were calculated by carrying on five REDAC cycles. Dihedral angle constraints were created with an ang-cut for the target function (TF) equal to 0.8 Å^2 in the first step, 0.6 in the second one and 0.4 in the third one. In the fourth step the structures were calculated using the constraints previously established. In the final step no other dihedral angle constraints were created and the structures were minimized at the highest level (L_{19}) using all the experimental restraints.

To obtain a set of structures representative of the conformational space allowed to the peptide, the best 23 conformers from the REDAC strategy were subjected to a RMSD cluster analysis.[18] Eight structures representative of different conformational families resulted, and they were used as starting models for RMD calculations, together with the best one calculated with the direct strategy.

RMD calculations: The INSIGHT/DISCOVER (Biosym Technologies, San Diego, CA, USA) program with the Amber Force Field[44,45] was employed for molecular dynamics; a steepest descendent algorithm was applied for the initial stages of the refinement and a Quasi Newton–Raphson algorithm for the final ones.[46] The RMD simulations were performed in vacuo at 300 K with a distance-dependent dielectric constant ($\varepsilon = r$)[47,48] and a time step of 0.5 fs. The motion equation algorithm was the leapfrog.[49] RMD simulations were carried out for 50 ps in the equilibration phase and for 160 ps without velocity rescaling; in the meantime, the temperature was kept constant at 300 K. Interproton distances evaluated from NOEs were inserted as restraints with a tolerance of 10% during the simulations.[50] Pseudoatoms were used instead of protons nonstereospecifically assigned.[51] The ϕ angle of Thr2, exhibiting a $^3J_{NH,\alpha CH}$ value of 9.9, was kept constant to $-120°$ with a force constant of 200 Kcal rad^{-2} mol^{-1}. Coordinates and velocities for the system were dumped into a disk every 1000 steps. Data recorded during the last 50 ps of the simulations were used for the statistical analysis. The molecular graphics program MOLMOL[52] was employed to calculate secondary structure and Ramachandran maps of the average models.

Modeling of the catalytic domain of furin and docking of the peptide: A three-dimensional model of the catalytic domain of human furin was built by homology, using WHAT IF software,[53] starting from the crystal structure coordinates of thermitase complexed with eglin-c (ref. [19], Brookhaven Data Bank code: 2TEC). The sequence identity between thermitase and human furin is 29%; the multiple sequence alignment previously reported by Siezen et al. for the entire subtilase family[20] was employed. The insertions were not modeled. The eglin backbone was not altered, while the side chains of the eglin residues 40–47 were substituted to resemble appropriate P6–P2′[21] residues of gp160, that is Val-Gln-Arg-Glu-Lys-Arg↓Ala-Val. The position-specific rotamer method was used to model the side chain conformations.[54] The entire complex was then regularized by energy minimization performed by the INSIGHT/DISCOVER program (100 cycles of conjugate–gradients energy minimization). The structure of the p498 peptide obtained by NMR data and MD simulations was docked onto the binding site of furin by superimposing the backbone of the Glu12-Ala15 (P3–P1′) segment onto the backbone of the modified eglin Glu43-Ala46 residues.

Acknowledgements

We are grateful to Dr. Marco Tatò of Pharmacia Upjohn (Nerviano, MI, Italy) for recording the 600 MHz NMR spectra.

[1] B. S. Stein, E. G. Engleman, *J. Biol. Chem.* **1990**, *265*, 2640–2649.
[2] J. N. Weber, R. A. Weiss, *Sci. Am.* **1988**, *259*, 100–109.
[3] S. Hallenberger, V. Bosh, H. Angliker, E. Shaw, H. D. Klenk, W. Garten, *Nature* **1992**, *360*, 358–361.
[4] E. Decroly, M. Vandenbranden, J. M. Ruysschaert, J. Cogniaux, G. S. Jacob, S. C. Howard, G. Marshall, A. Kombelli, A. Basak, F. Jean, C. Lazure, S. Benjannet, M. Chrétien, N. G. Seidah, *J. Biol. Chem.* **1994**, *269*, 12240–12247.
[5] M. Moulard, E. Decroly, *Biochim. Biophys. Acta* **2000**, *1469*, 121–132, and references therein.
[6] J. M. McCune, L. B. Rabin, M. B. Feinberg, M. Lieberman, J. C. Kosek, G. R. Reyes, I. L. Weissman, *Cell* **1988**, *53*, 55–67.
[7] a) E. O. Freed, D. J. Myers, R. Risser, *J. Virol.* **1989**, *63*, 4670–4675; b) H. G. Guo, F. D. Veronese, E. Tschachler, R. Pal, V. S. Kalyanaraman, R. C. Gallo, M. S. Reitz, Jr., *Virology* **1990**, *174*, 217–224.
[8] V. Bosch, M. Pawlita, *J. Virol.* **1990**, *64*, 2337–2344.
[9] G. Turner, M. F. Summers, *J. Mol. Biol.* **1999**, *285*, 1–32.
[10] N. Brakch, M. Dettin, C. Scarinci, N. G. Seidah, C. Di Bello, *Biochem. Biophys. Res. Commun.* **1995**, *213*, 356–361.
[11] E. Decroly, S. Wouters, C. Di Bello, C. Lazure, J. M. Ruysschaert, N. G. Seidah, *J. Biol. Chem.* **1996**, *271*, 30442–30450.
[12] M. Moulard, L. Challoin, S. Canarelli, K. Mabrouk, H. Darbon, *Biochemistry* **1998**, *37*, 4510–4517.
[13] R. J. Siezen, J. W. Creemers, W. J. Van De Ven, *Eur. J. Biochem*. **1994**, *222*, 255–266, and references therein.
[14] J. W. Creemers, R. J. Siezen, A. J. M. Roebroek, T. A. Y. Ayoubi, D. Huylebroeck, W. J. M. Van De Ven, *J. Biol. Chem.* **1993**, *268*, 21826–21834.

  0947-6539/02/0806-1472 $ 17.50+.50/0 *Chem. Eur. J.* **2002**, *8*, No. 6

[15] D. S. Wishart, B. D. Sykes, F. M. Richards, *J. Mol. Biol.* **1991**, *222*, 311–333.
[16] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283–298.
[17] P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1991**, *1*, 447–456.
[18] A. D. McLachlan, *J. Mol. Biol.* **1979**, *128*, 49–79.
[19] P. Gros, C. Betzel, Z. Dauter, K. S. Wilson, W. G. Hol, *J. Mol. Biol.* **1989**, *210*, 347–367.
[20] R. J. Siezen, W. M. de Vos, B. W. Dijkstra, *Prot. Eng.* **1991**, *4*, 719–737.
[21] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157–162.
[22] S. Segawa, T. Fukuno, K. Fujiwara, Y. Noda, *Biopolymers* **1991**, *31*, 497–509.
[23] M. A. Jimenez, M. Bruix, C. Gonzales, F. J. Blanco, J. L. Nieto, J. Herranz, M. Rico, *Eur. J. Biochem*. **1993**, *211*, 569–581.
[24] M. T. Reymond, S. Huo, B. Duggan, P. E. Wright, H. J. Dyson, *Biochemistry* **1997**, *36*, 5234–5244.
[25] D. E. Callihan, T. M. Logan, *J. Mol. Biol.* **1999**, *285*, 2161–2175.
[26] M. Rholam, P. Nicolas, P. Cohen, *FEBS Lett.* **1986**, *207*, 1–6.
[27] E. Bek, R. Berry, *Biochemistry* **1990**, *29*, 178–183.
[28] L. Abrami, M. Fivaz, E. Decroly, N. G. Seidah, F. Jean, G. Thomas, S. H. Leppla, J. T. Buckley, F. G. van der Goot, *J. Biol. Chem*. **1998**, *273*, 32656–32661.
[29] M. Tsuneoka, K. Nakayama, K. Hatsuzawa, M. Komada, N. Kitamura, E. Mekada, *J. Biol. Chem.* **1993**, *268*, 26461–26465.
[30] J. Chen, K. H. Lee, D. A. Steinhauer, D. J. Stevens, J. J Skehel, D. C. Wiley, *Cell* **1998**, *95*, 409–417.
[31] L. Paolillo, M. Simonetti, N. Brakch, G. D'Auria, M. Saviano, M. Dettin, M. Rholam, A. Scatturin, C. Di Bello, P. Cohen, *EMBO J.* **1992**, *11*, 2399–2405.
[32] C. Di Bello, M. Simonetti, M. Dettin, L. Paolillo, L. Falcigno, G. D'Auria, M. Saviano, A. Scatturin, G. Vertuani, P. Cohen, *J. Pept. Sci.* **1995**, *1*, 251–265.
[33] L. Falcigno, L. Paolillo, G. D'Auria, M. Saviano, M. Simonetti, C. Di Bello, *Biopolymers* **1996**, *39*, 837–848.
[34] M. Simonetti, L. Falcigno, L. Paolillo, C. Di Bello, *Biopolymers* **1997**, *41*, 461–479.
[35] N. Brakch, M. Rholam, H. Bousetta, P. Cohen, *Biochemistry* **1993**, *32*, 4925–4930.
[36] B. Rost, C. Sander, *J. Mol. Biol.* **1993**, *232*, 584–599.
[37] S. Choe, M. J. Bennett, G. Fujii, P. M. G. Curmi, K. A. Kantardjieff, R. J. Collier, D. Eisenberg, *Nature* **1992**, *357*, 216–222.
[38] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *65*, 355–360.
[39] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[40] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *63*, 207–213.
[41] U. Piantini, O. W. Sørensen, R. R. Ernst, *J. Am. Chem. Soc.* **1982**, *104*, 6800–6801.
[42] C. Griesinger, E. E. Ernst, *J. Magn. Res.* **1987**, *75*, 261–271.
[43] J. M. Crippen, *Distance Geometry and Conformational Calculations*, Chemometric Research Studies Series 1, Wiley, New York, **1981**.
[44] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
[45] W. D. Cornell, P. Cieplak, C. I. Bayly, K. F. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, J. M. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[46] C. L. Brooks III, M. Karplus, B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics,* in *Advances in Chemical Physics 71*, Wiley, New York, **1988**.
[47] J. A. McCammon, P. G. Wolynes, M. Karplus, *Biochemistry* **1979**, *18*, 927–942.
[48] G. Melacini, Q. Zhu, M. Goodman, *Biochemistry* **1997**, *36*, 1233–1241.
[49] R. W. Hockney in *Methods in Computational Physics*, *Vol. 9 (Plasma Physics)*, Academic Press, New York, London, **1970**, pp. 136–211.
[50] W. F. van Gunsteren, H. J. C. Berendsen, *Angew. Chem.* **1990**, *101*, 1020–1051; *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992–1023.
[51] K. Wüthrich, M. Billeter, W. Brown, *J. Mol. Biol.* **1983**, *169*, 949–961.
[52] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51–55.
[53] G. Vriend, *J. Mol. Graphics* **1990**, *8*, 52–56.
[54] G. Chinea, G. Padron, R. W. W. Hooft, C. Sander, G. Vriend, *Proteins* **1995**, *23*, 415–421.

Received: October 8, 2001 [F3600]