# Structural Investigation of the HIV-1 Envelope Glycoprotein gp160 Cleavage Site, 2: Relevance of an N-Terminal Helix**

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

*Proteolytic activation of the HIV-1 envelope glycoprotein gp160 is selectively performed by the proprotein convertase furin at the C-terminus of the sequence R508 – E – K – R511 (site 1), in spite of the presence of another consensus sequence, Lys500 – Ala – Lys – Arg503 (site 2). On the basis of the solution structural analysis of the synthetic peptide p498, spanning the gp160 sequence Pro498 – Gly516, we previously suggested a possible role of an N-terminal helix in regulating the exposure and accessibility of the gp160 physiological cleavage site, enclosed in a loop. Here we report on the activity and conformation of the 23-residue peptide h-REKR, designed to exhibit a large N-terminal helix, followed by the gp160 native sequence, Arg508 – Gly516. h-REKR is digested by furin with high efficiency, comparable to the full native p498. Circular dichroism analyses, in mixtures from pure water to 98% trifluoroethanol, outline a significant content of helical structure in the peptide conformation. The molecular model obtained from NMR data collected in trifluoroethanol/water, by means of DYANA and AMBER simulations, indeed has helical structure on a large N-terminal segment. Such a long helix does not seem to affect the loop conformation of the C-terminal site 1-containing sequence, which exhibits the same proton chemical shifts already observed for the full native p498.*

**KEYWORDS:**

conformation analysis · gp160 · HIV · molecular modeling · NMR spectroscopy

## Introduction

Proteolytic processing of the HIV envelope glycoprotein gp160 is a prerequisite for the virus infectivity. HIV gp160 is indeed synthesized as an inactive precursor in a late Golgi compartment,[1] and is then cleaved to give the noncovalently associated glycoproteins gp120 and gp41. Both gp120 and gp41 participate into the virus infectivity process: gp120 mediates the interaction of the viral particle with CD4 receptor – coreceptor complexes; gp41 carries out the fusion of viral and cellular lipid membranes.[2]

Proprotein convertases (PCs) are newly discovered mammalian subtilases that are candidates for the intracellular processing of the HIV envelope glycoprotein gp160. Furin was the first PC enzyme shown to cleave gp160 intracellularly into gp120/gp41[3] and is to date the best candidate for the gp160 proteolytic processing.[4]

By analogy with prohormones, viral glycoproteins present multibasic cleavage sites and, in particular, consensus sequences of the type Lys/Arg – Xaa – Lys/Arg – Arg.[5] The proteolytic cleavage of the HIV-1 glycoprotein gp160 precursor occurs for more than 85% at the C-terminus of the sequence Arg508 – Glu – Lys – Arg511 (site 1),[6] although a second putative cleavage site (Lys500 – Ala – Lys – Arg503, site 2) is present eight residues upstream. Basic residues within the site 1 recognition sequence are necessary for the proteolytic processing, as confirmed by site-directed mutagenesis studies.[7] However, the role of the nonphysiological site is still hypothetical.

The fact that proteolytic cleavage of the HIV-1 glycoprotein precursor selectively occurs at site 1 suggests that for the molecular recognition other factors, such as specific secondary structural motifs, are required in addition to the consensus sequence Lys/Arg – Xaa – Lys/Arg – Arg. Structural features required for the molecular recognition between HIV-1 gp160 and proteolytic enzymes have not yet been clearly assessed. As the structure of the gp120/gp41 junction remains to be elucidated, current hypotheses can only rely on protein structure predictions and conformational studies of model peptides.

[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)*
*E-mail: paolillo@chemistry.unina.it*

[b] *Prof. G. D'Auria*
*Istituto di Biostrutture e Bioimmagini del C.N.R.*
*Università di Napoli "Federico II"*
*Via Mezzocannone 4, 80134 Napoli (Italy)*

[c] *Dr. M. Dettin, Dr. C. Scarinci, Dr. A. Pasquato, Prof. C. D. Bello*
*Dipartimento di Processi Chimici dell'Ingegneria*
*Università di Padova*
*Via Marzolo 9, 35131 Padova (Italy)*
*Fax: (+39) 081-674090*

[**] *Part 1: ref. [10]*

 **DOI: 10.1002/cbic.200200541**  

The 19-residue peptide p498, spanning the junction between gp120 and gp41 and containing sites 1 and 2, was shown by Brakch et al.[8] to be recognized and properly digested by furin at site 1,[8, 9] suggesting that contributions of specific secondary structure motifs from surrounding amino acids can be reproduced even in a middle-sized peptide fragment like p498. The p498 structural analysis by NMR techniques that we recently reported,[10] outlines the presence of a C-terminal loop exposing the physiological site 1, and a helix at the N-terminal side enclosing site 2.

A loop had previously been suggested as a common structural feature for physiological cleavage sites in retroviral envelope glycoproteins.[4, 11] We hypothesized that the N-terminal helix enclosing the secondary processing site 2 could play a key role in regulating the exposure and accessibility of the physiological cleavage site 1 enclosed in a loop. Moreover, our evaluation of the secondary structure propensity around the cleavage site of 37 precursors processed by furin (reported by Nakayama[5]), has indeed shown that the sequence at the N-terminus of the signal loci has a significant $\alpha$ helical conformation propensity in which the helix is the only predicted ordered structure. In particular, by applying the hierarchical neural network (HNN) prediction method[12] to 14-residue segments preceding the consensus sequence, the presence of helical structure is predicted for 30 of the 37 analyzed sequences and an $\alpha$ helix content above the 50% level is estimated for 11 of them.

In order to investigate the possible role of specific secondary structure motifs flanking the gp160 cleavage site, we designed and synthesized a series of peptide analogues exhibiting: 1) at the C-terminal, the native gp160 sequence Arg508 – Gly516 containing site 1, and 2) at the N-terminal, model sequences known to assume specific secondary structure motifs.

Here we report the synthesis, activity, and conformation of h-REKR, a 23-residue peptide exhibiting a large N-terminal helical segment,[13] followed by the gp160 native sequence Arg508 – Gly516 (Figure 1). The h-REKR helical sequence Glu1 – Asn14 corresponds to the Glu14 – Asn27 sequence in granulocyte-macrophage colony-stimulating factor (GM-CSF), enclosed in the N-terminal Trp13 – Leu28 $\alpha$ helical segment (helix A). Such a sequence was previously shown to assume a helical structure even as a fragment in trifluoroethanol (TFE)/water mixtures (a partial onset of $\alpha$ helical structure is present in pure water, and 56% or more helical content is observed at saturation above 30% TFE).[14]

It is noteworthy that although h-REKR does not contain the secondary processing site 2, it is nevertheless a very good substrate for furin. The h-REKR conformational analysis, carried out in solution by NMR techniques, is presented and related to the exhibited activity. Special effort is devoted to refining the native C-terminal Arg15 – Gly23 sequence conformation, which seems to be preserved in p498 and h-REKR analogues.

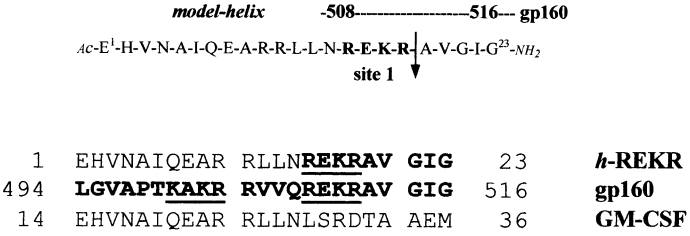


***Figure 1.*** *Amino acid sequence of the h-REKR synthetic peptide. h-REKR C-terminal sequence (Arg15 – Gly23) spans the gp160 sequence from Arg508 to Gly516 containing the site 1; the N-terminal model-helix sequence (Glu1 – Asn14) corresponds instead to the Glu14-Asn27 sequence from GM-CSF helix A. An arrow is used to indicate the cleavage loci for the recognized processing site 1. The alignment with the corresponding segments of the gp160 (ENV HV1 LW) and GM-CSF (pdbCode: 2GMF[13]) native sequences, bold and normal respectively, is also reported. gp160 processing sites 1 and 2 are underlined.*

## Results

### Cleavage of synthetic peptides by furin

p498 and h-REKR were incubated with furin for 1 or 3 h at 37 °C. With peptide p489, data clearly show that furin acts specifically at the potential cleavage site Arg – Glu – Lys – Arg (site 1); the resulting conversion percentages are 75 and 100, respectively (Table 1). The substitution of the N-terminal native gp160 sequence with an $\alpha$ helix sequence creates a good substrate for furin. In fact the conversion of h-REKR is nearly complete (98% after 3 h incubation). Therefore activity studies show that h-REKR is digested by the enzyme furin with an efficiency comparable to the native p498.

***Table 1.*** *Conversion percentages of p498 and h-REKR after 1 and 3 h incubation with human furin at 37 °C.*

| Substrate | Incubation time [h] | Digestion [%] |
|---|---|---|
| p498 | 1 | 75 |
| | 3 | 100 |
| h-REKR | 1 | 41 |
| | 3 | 98 |

### h-REKR conformational analysis

Circular dichroism (CD) investigations show that the peptide folds into an $\alpha$ helical conformation, characterized by a positive band at 190 nm and two negative bands at 207 and 220 nm, in the presence of increasing amounts of TFE (Figure 2). The helical content is 64% at saturation conditions, as estimated according to the method of Greenfield and Fasman.[15] The maximum helix content is reached above 30% TFE, but significant helix content (48%) is present at 20% TFE.

The NMR analysis was performed in a TFE/$H_2O$ (90:10 v/v) solution, to allow a direct comparison with available data for p498. The complete sequential assignment was achieved according to the standard procedure proposed by Wüthrich.[16] Proton chemical shifts are presented in Table 2. All the proton chemical shifts observed for the native Arg15 – Gly23 sequence, except for the Arg15, Lys17, and Arg18 amide protons, do not differ by more than 0.04 ppm from the observed values for the corresponding Arg11 – Gly19 sequence in the p498 analogue. The $\alpha$CH chemical shift deviations from the random coil values[17] versus residue number (Figure 3) are all negative for the Val3 –

  

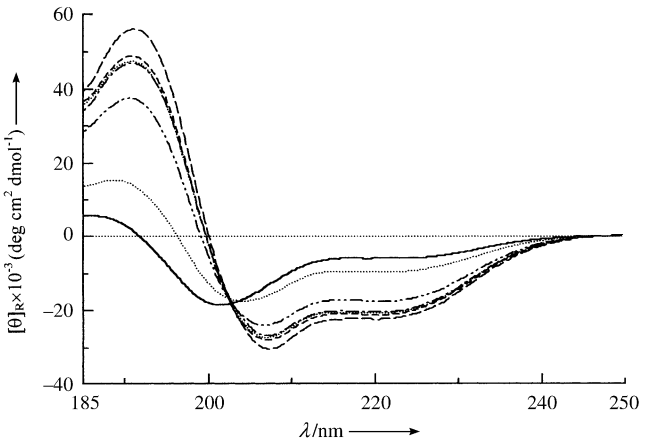


***Figure 2.*** *CD spectra of $4 \times 10^{-5}$ M h-REKR in water/TFE at increasing percentages of TFE (v/v): —— 0%; •••• 10%; –••– 20%; •••• 30%; –•– 40%; - - - - 50%; – – 98%.*

***Table 2.*** *Proton chemical shifts [ppm] of h-REKR at 298 K in TFE/$H_2O$ (90:10 v/v) with TSP as internal reference.*

| Residue | HN | Hα | Hβ | Hγ | Others |
|---|---|---|---|---|---|
| Glu1 | 7.74 | 4.26 | 2.02 | 2.46 | Ac 2.06 |
| His2 | 8.35 | 4.70 | 3.37, 3.27 | | 2H 8.39, 4H 7.25 |
| Val3 | 7.65 | 3.97 | 2.17 | 1.04, 1.01 | |
| Asn4 | 8.16 | 4.61 | 2.88 | | NHγ 7.29, 6.35 |
| Ala5 | 7.88 | 4.16 | 1.55 | | |
| Ile6 | 7.64 | 3.86 | 2.03 | 1.72, 1.25 | $CH_3$γ 0.96, Hδ 0.91 |
| Gln7 | 8.08 | 4.12 | 2.25, 2.15 | 2.57, 2.48 | NHδ 6.83, 6.27 |
| Glu8 | 8.20 | 4.21 | 2.25, 2.20 | 2.57 | |
| Ala9 | 8.26 | 4.13 | 1.60 | | |
| Arg10 | 8.23 | 4.01 | 2.02, | 1.98, 1.71 | Hδ 3.19, NHε 6.93 |
| Arg11 | 7.99 | 4.08 | 2.11 | 2.00, 1.69 | Hδ 3.23, NHε 7.09 |
| Leu12 | 8.32 | 4.14 | 1.95, 1.73 | 1.84 | Hδ 0.94 |
| Leu13 | 8.65 | 4.18 | 1.91, 1.63 | 1.85 | Hδ 0.94 |
| Asn14 | 8.08 | 4.48 | 3.03, 2.80 | | NHγ 7.59, 6.41 |
| Arg15 | 8.06 | 4.12 | 2.10 | 1.88, 1.76 | Hδ 3.24, NHε 7.12 |
| Glu16 | 8.39 | 4.14 | 2.28 | 2.67, 2.50 | |
| Lys17 | 8.19 | 4.11 | 2.03 | 1.69, 1.53 | Hδ 1.75, Hε 3.00 |
| Arg18 | 7.77 | 4.24 | 1.95 | 1.84, 1.74 | Hδ 3.23, NHε 7.04 |
| Ala19 | 7.98 | 4.27 | 1.53 | | |
| Val20 | 7.71 | 4.14 | 2.24 | 1.06, 1.00 | |
| Gly21 | 7.85 | 4.07, 3.90 | | | |
| Ile22 | 7.62 | 4.16 | 1.94 | 1.58, 1.24 | $CH_3$γ 0.96, Hδ 0.91 |
| Gly23 | 7.94 | 3.96, 3.87 | | | CONH 7.20, 6.53 |

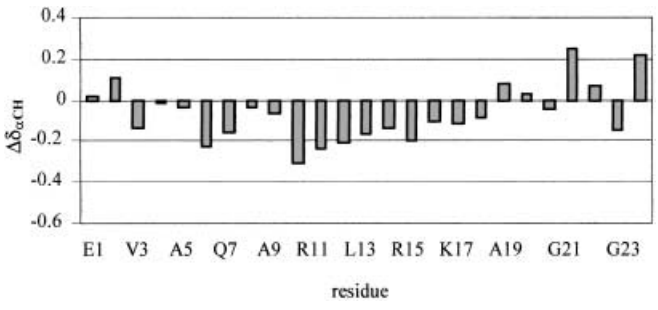


***Figure 3.*** *Chemical shift deviations from the random coil values for αCH protons.[17] For Gly21 and Gly23 both the αCH and α′CH chemical shift deviations are reported.*

Arg18segment, which points to a helical structure or turn.[18] Moreover a complete NOE pattern that is diagnostic of a helical conformation, that is, strong $NH_i - NH_{i+1}$, together with $NH_i - NH_{i+2}$, $\alpha_i - NH_{i+3}$, $\alpha_i - \beta_{i+3}$, and $\alpha_i - NH_{i+4}$ NOE "contacts",[16] is observed for the N-terminal/middle region of the peptide (from the N-terminus to Asn14); $\alpha_i - NH_{i+3}$ "contacts" are also observed in the C-terminal region (Figure 4). A set of 211 experimental constraints from NOE data (95 intraresidual, 75 sequential, and 41 medium-range, including 23 constraints characteristic of a helical structure) was determined for structure calculations. These constraints were imposed as upper bounds on inter-proton distances for sampling the conformational space allowed to the peptide by means of torsion angle dynamics (DYANA program).[19]

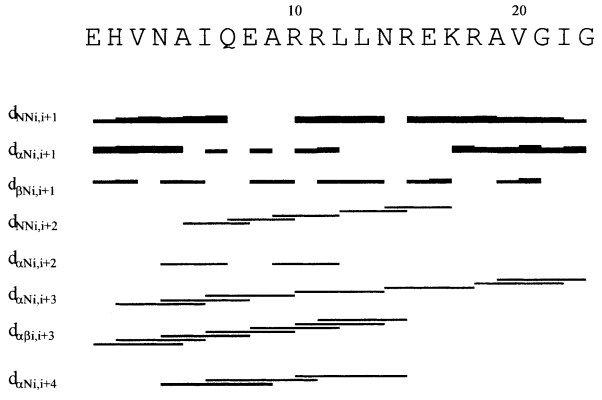


***Figure 4.*** *Summary of the most relevant NOE effects in TFE/$H_2O$ (90:10 v/v).*

A first set of 100 conformers was calculated with the DYANA program, and 100 additional structures were calculated by using the redundant dihedral angle constraints (REDAC) strategy[20] to improve the convergence. The REDAC structures are indeed more compatible with experimental data. All of these structures never exhibit violations larger than 0.2 Å for 200 out of the overall 211 experimental restraints. The 30 DYANA structures with the lowest values of target function (average value: $1.70 \pm 0.06$ Å$^2$; mean global backbone root-mean-square difference (rmsd) on the best 30 structures: $2.15 \pm 0.79$ Å) were then subjected to a restrained energy minimization by using the SANDER module of the AMBER 6.0 package.[21] The best 10 structures, in terms of the fitting with experimental restraints, were selected among those with a residual restraint energy lower than $-260$ kcal mol$^{-1}$ to represent the h-REKR solution structure (Figure 5a, Table 3). The whole N-terminal fragment, from Glu1 to Lys17, is well defined (average rmsd on backbone: 0.40 Å). A canonical helix extends from Val3 to Leu13 (average rmsd on backbone: 0.22 Å). An axial γ turn is further present around Arg15, which is enclosed in a larger bending involving five residues (Leu13 – Lys17), stabilized by a hydrogen bond between the Lys17 amide proton and the Leu13 carbonyl oxygen atom, and classified as a type II-$\alpha_{RS}$ turn.[22]

The C-terminal side, from Arg18 to Gly23, is less ordered (average rmsd on backbone: 1.2 Å), and may be described as a loop in which the processing site 1 is exposed. Nevertheless, a β turn mostly of the I′ type involving residues Ala19 – Ile22 is

  

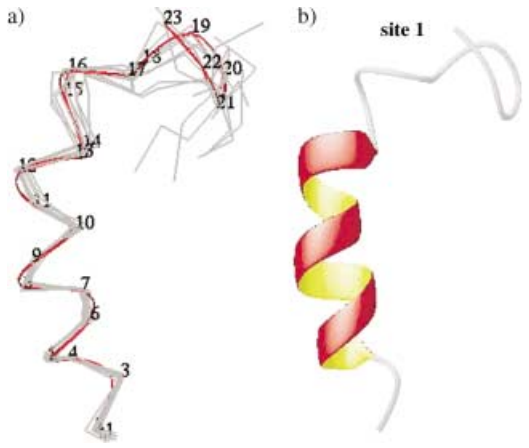


***Figure 5.*** *a) The best backbone superimposition of the Val3 – Leu13 segment in the best 10 DYANA structures, after the AMBER refinement (gray), together with the mean structure (red). b) The mean AMBER structure generated with MOLMOL[23] is represented by a cartoon ribbon.*

***Table 3.*** *Structural statistics.*

| Parameter | Value |
|---|---|
| Quantity | 10 conformers[a] |
| Residual distance constraint violations [Å] | |
| $0.2 < d \leq 0.3$ | $6.6 \pm 1.8$ |
| $0.3 < d \leq 0.4$ | $4.8 \pm 0.8$ |
| $0.4 < d = 0.5$ | $1.3 \pm 1.6$ |
| Maximum violation | $0.42 \pm 0.04$ |
| AMBER energies [kcal mol$^{-1}$] | |
| Distance constraint | $21.26 \pm 1.72$ |
| Van der Waals | $-81.61 \pm 3.35$ |
| Total | $-267.51 \pm 7.27$ |
| rmsd to the averaged coordinates[b] [Å] | |
| Backbone (1 – 23) | $2.23 \pm 0.77$ |
| All heavy (1 – 23) | $2.63 \pm 0.60$ |
| Backbone (3 – 13) | $0.32 \pm 0.14$ |
| All heavy (3 – 13) | $1.48 \pm 0.38$ |

[a] Average value for the 10 energy-minimized conformers. [b] Average coordinates of the best 10 energy-minimized conformers after superposition for the best fit of the atoms of residues indicated in parentheses.

observed in all 10 structures. The C-terminal amide group (on Gly23) bends to make a hydrogen bond several residues upstream with, alternatively, the carbonyl oxygen atom Glu16, Lys17, or Ile22.

The mean AMBER structure generated by MOLMOL[23] is shown in Figure 5b. It is well representative of the entire conformational family, exhibiting a helical structure from His2 to Leu13, an $\alpha$ turn involving residues Leu13 – Lys17 enclosing an equatorial $\gamma$ turn around Arg15, a $\beta$ turn of type I′ on Ala19 – Ile22, and a hydrogen bond connecting the C-terminal amide with the Glu16 carbonyl.

#### h-REKR docking onto furin catalytic region

h-REKR was docked to furin to investigate the accessibility of its cleavage site to the enzyme catalytic region by using the mean AMBER molecular model as the representative peptide conformation. A three-dimensional model was employed for the catalytic domain of human furin, which was built by homology on the basis of the crystal structure of thermitase complexed to eglin-c,[24] and is described elsewhere.[10] In the homology-modeled complex, the side chains of the eglin-c residues 40 – 47 were substituted to resemble appropriate P6 – P2′ residues of gp160 (according to the nomenclature of Schechter and Berger[25]), namely, Val – Gln – Arg – Glu – Lys – Arg ↓ Ala – Val, thus mimicking a model substrate. We remind the reader that the enzyme binding region was described as a surface channel which was able to accommodate eight substrate residues around the multibasic sequence, from P6 to P2′[10] (see also ref. [26]). The furin catalytic residues (Ser225, His71, Asp38, and Asn163) are located at the bottom of the channel, together with several surface-exposed acidic residues, making specific electrostatic interactions with the basic amino acids of the multibasic substrate sequence. In particular, in our previous paper[10] we hypothesized interactions of the substrate P1 – P2 residues, Arg45 and Lys44, with furin Asp199 and Asp47 (in the subsites S1 and S2, respectively) and of Arg42 at P4 with Glu129 (in the subsite S4), which are fully consistent with site-directed mutagenesis studies performed on furin by Creemers et al.[27]

A h-REKR rigid docking was performed by taking advantage of the presence of a model substrate (modified eglin-c) in the previously modeled furin catalytic domain. By analogy to p498,[10] a suitable peptide docking onto the furin catalytic region was obtained by superimposing the backbone of the P3 – P1′ residues, around the h-REKR cleavage point (Glu16 – Lys17 – Arg18 – Ala19) on the backbone of the corresponding residues (Glu43 – Lys44 – Arg45 – Ala46), in the modified eglin-c. The result is shown in Figure 6. Residues at the cleavage site present a good fit with the furin catalytic residues (Ser225, His71, Asp38, and Asn163). Indeed, in such an arrangement, no modification in the h-REKR backbone is required to nicely accommodate the entire peptide into the furin binding region. Furthermore, the Arg18 carbonyl carbon atom, on which the proteolytic activation occurs, sits at a catalytic distance (3.0 Å) from the hydroxy oxygen atom of Ser225, while its side chain nicely fits the enzyme subsite S1, with residue Asp199 exposed at the bottom (Figure 6b).

## Discussion

Processing of HIV-1 envelope glycoprotein gp160 is a key event in the virus infectivity. Clarification of the structural requirements for the gp160/PCs recognition is a crucial step in obtaining properly designed inhibitors. To date, no structure is available for HIV-1 gp160. We recently reported a structural analysis of a fragment spanning the Pro498 – Gly516 sequence (p498). There we found a loop exposing the physiological cleavage site 1, and a helix at the N-terminal side. Also, on the basis of the peptide p498 docking onto the furin catalytic region, as modeled by homology, we suggested a role for the N-terminal helix in regulating the exposure and accessibility of the loop enclosing the physiological cleavage site.

  

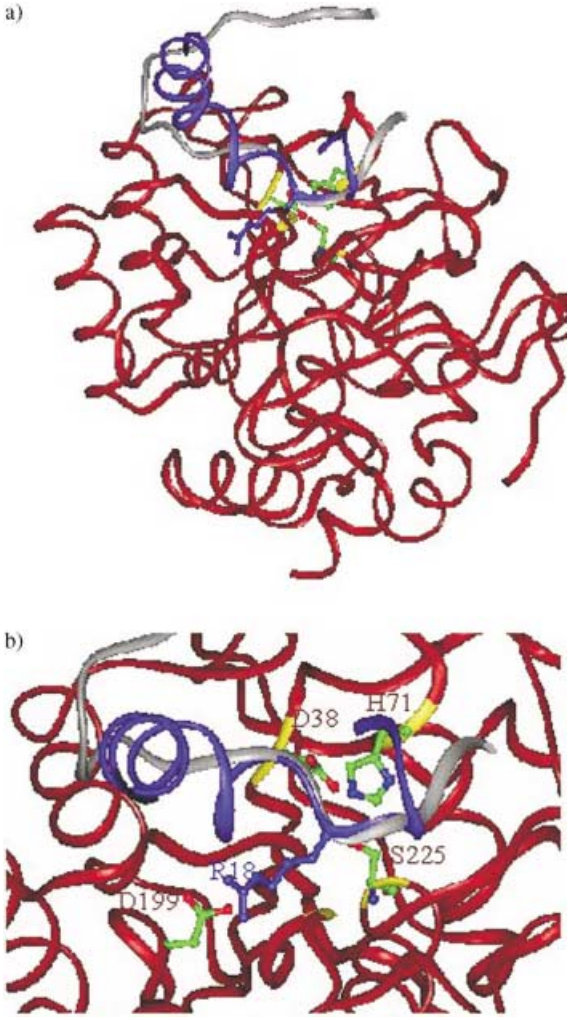


***Figure 6.*** *Docking of the h-REKR mean AMBER structure onto the furin catalytic domain. The backbone of h-REKR Glu16 – Ala19 residues, around the processing site, has been superimposed to the corresponding one of the model inhibitor (Glu43 – Ala46). a) Ribbon representation: h-REKR, a segment of the inhibitor and furin are shown in violet, gray, and red, respectively. Furin catalytic residues (Ser225, His71, Asp38, and Asn163) are shown in yellow. Side chains of the furin catalytic triad (Ser225, His71, and Asp38) and of the h-REKR Arg18 residue at P1, undergoing the proteolytic attack, are shown as ball and sticks. b) Expansion of the binding site in a slightly different orientation, the Asp199 side chain of the furin subsite S1 is also shown. All the side chains shown are labeled.*

Here, we report the structure and activity of the 23-residue peptide h-REKR, which belongs to a series of analogues we designed and synthesized to investigate the role of specific secondary structures flanking the gp160 cleavage site. The key feature of this model peptide is the introduction of a large non-native helical sequence at the N-terminal side of site 1, instead of the site 2-containing sequence. We showed that the peptide's capability to be processed by furin is high and not significantly affected by such modification, as compared to the full native analogue p498. This result is not trivial, as a very low cleavage efficiency was observed for other analogues of the series (work in progress). It must be noted, however, that the five residues preceding the cleavage site 1, Arg10 – Arg – Leu – Leu – Asn14 from the GM-CSF helix A, strictly resemble the corresponding gp160 sequence Arg503 – Arg – Val – Val – Gln507 (see also Figure 1).

NMR and CD techniques were employed to analyze the peptide structural features in TFE/water. TFE is well known to promote ordered structures by favoring solvent-shielded amide conformations, and it has been shown that the local conformation of native proteins is usually better reproduced by the corresponding fragments in TFE/water solutions rather than in pure water.[28–31] An NMR study showed that h-REKR is strictly respondent to the design expectations. It is indeed organized in a stable N-terminal helix, extending from Lys3 to Leu13 (note the complete pattern of NOE effects is diagnostic of helical structure, Figure 4), followed by a loop on the native C-terminal sequence (Figure 5). The helix is one turn longer than in p498, while the loop conformation seems to be preserved, considering that even the proton chemical shifts are identical with those already observed in the corresponding sequence of p498. As h-REKR is processed with an efficiency comparable to p498 and much higher relative to other analogues bearing different N-terminal model conformation, "helix + loop" seems to be a proper secondary structure motif for the digestion by furin.

Modeling studies revealed that the main determinants for the proteolytic sites susceptibility are exposure, flexibility, and ability to unfold locally and adapt to the enzyme active site.[32] We performed attempts to dock peptides to furin to investigate the exposure/accessibility of their cleavage sites to the enzyme catalytic region. The p498[10] and h-REKR (Figure 6) docking indeed confirms that a "helix + loop" conformation allows the peptides to have an easy access to the furin catalytic region. These results support the hypothesis that in gp160 the processing site may be enclosed in a loop, which is preceded by a helix helping in correctly orienting it.

It is however expected that upon recognition, peptides arrange themselves to optimize the energetic interactions with the enzyme, especially electrostatic interactions of the consensus basic residues with the acidic residues in the furin substrate channel. Note that according to the available models for the furin binding mode,[10, 26] the h-REKR P6 – P2′ sequence, corresponding to the substrate region candidate to the binding, is strictly similar to that of gp160/p498 (Figure 1). The N-terminal helix instead stops at residue P6 (L13), that is, before the binding region, and seems thus to play the important structural role of exposing the functional loop to the proteolysis enzymes.

The fact that the loop conformation of the C-terminal sequence is not affected by the presence of a non-native helix at the N-terminus then raises the question of the intrinsic propensity of an amino acidic sequence around the cleavage site 1 to adopt a well-defined conformation. It is a highly hydrophilic sequence, but no strong propensity to a specific secondary motif could be found by available prediction methods.[12] The experimental conformation of such a loop has been quite clearly assessed here. In particular, a type I′ β turn was observed on residues Ala19 – Ile22 next to the cleavage site.

One more interesting observation concerns site 2, whose role, like other secondary processing sites, remains speculative. We have already proposed that it could be blocked in a rigid structure and could just play a structural role in directing the

  

processing,[10] without close participation in the recognition with furin. It has been shown that multiple replacements of the site 2 basic residues drastically reduce the gp160 cleavage. Indeed activity and structure results about h-REKR show not only that site 2 residues are not necessary for the processing at site 1, but also that their absence does not reduce the cleavage efficiency, provided that they are substituted by helix-promoting residues. Recently the role of a short helix-promoting sequence at the N-terminus of the multibasic cleavage site in governing the proteolytic processing was also emphasized for pro-somatostatin.[33]

Structural studies are in progress on other HIV-1 gp160 cleavage site analogues to further investigate the structural factors that play a role in gp160-proteolytic enzyme recognition.

## Experimental Section

**Peptide synthesis**: h-REKR was synthesized by solid-phase peptide synthesis with an automated peptide synthesizer Model 431A (Applied Biosystems, Forster City, CA, USA) by standard Fmoc chemistry. The Rink Amide MBHA Resin (0.49 mmol $g^{-1}$) was employed. Amino acids were incorporated by using 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate/1-hydroxybenzotriazole(HBTU/HOBt) reagents. The following side chain protecting groups used were: Thr – *t*Bu; Lys – *tert*-butyloxycarbonyl (Boc); Asn-, Gln-, and His – tripheylmethyl (Trt); Arg – 2,2,5,7,8-pentamethyl-chroman-6-sulfonyl (Pmc); Glu- and Asp – O*t*Bu. Double couplings were introduced in the sequence 1 – 13 and for Ala19 and Gly23 insertions. At the end of the synthesis, after deprotection of the 9-fluorenylmethyloxycarbonyl (Fmoc) group of the N-terminal residue, the h-REKR peptide was acetylated by using 10% acetic anhydride and 5% DIEA/NMP solution for 1 h. The protected peptide – resin was treated with $H_2O$ (0.5 mL), ethanedithiol (EDT) (0.25 mL), thioanisole (0.5 mL), phenol (0.75 g), and trifluoroacetic acid (TFA) (10 mL) for 90 min to cleave the peptide from the solid support and to remove the side chain protecting groups. The crude peptides were purified by reversed-phase high-performance liquid chromatography (HPLC) with a semi-preparative column (Deltapak $C_{18}$, Waters, 15 μm, 100 Å, 7.8 × 300 mm). The purified h-REKR was analyzed by using the following conditions: column, Vydac $C_{18}$ (5 μm, 300 Å, 4.6 × 250 mm); eluant A: 0.05% TFA in water; eluant B: 0.05% TFA in $CH_3CN$; gradient, 25 – 35% B over 20 min; flow rate: 1 mL $min^{-1}$; detector: 214 nm. The retention time of h-REKR was 16.38 min. The integration of chromatographic pattern gave a 93% purity grade. The homogeneity of the products was investigated by capillary electrophoresis analysis by using an Applied Biosystems instrument Model 270A. The identity of the product was confirmed by MALDI spectrometric analysis (calcd. for h-REKR: 2671; found: 2667).

**Enzyme assay**: Synthetic peptides were dissolved in $H_2O$ at 1 M concentration. The solution (15 μL) of each peptide underwent proteolysis in a reaction volume (85 μL) containing purified furin (10 μL, activity 153.5 pmol AMC $\mu L^{-1}$ enzyme $h^{-1}$) in Tris acetate (50 μL, 100 mM, pH 7), $CaCl_2$ (2 μL, 100 mM), and $H_2O$ (23 μL). The reaction mixture was incubated at 37 °C for 1 or 3 h. Products were identified by HPLC purification followed by MALDI mass spectrometry analysis. The substrate conversion was determined by integration of the peak area of uncleaved substrate and comparison with peak area of furin-untreated sample.

**CD analysis**: Circular dichroism spectra were obtained on a Jasco J-710 automatic recorder spectropolarimeter. All measurements were performed at room temperature in quartz cells of 0.1-cm path length. Spectra were recorded with a 2.0-nm bandwidth, time constants of 2 s, and a scan speed of 5 nm $min^{-1}$; two scans were collected to improve the signal-to-noise ratio and the solvent baseline was recorded and subtracted from the spectra of the samples. All CD spectra were smoothed by using the Jasco FT noise reduction software, and are reported in terms of ellipticity units per mole of peptide residue ($[\Theta]_R$). Peptide concentration was determined by amino acid analysis. The helix content was estimated by the amplitude of the CD band at 220 nm according to the method of Greenfield and Fasman.[15]

**NMR analysis**: NMR experiments were carried out on a Varian Unity – Inova 600 MHz spectrometer, equipped with a Sun Station Ultra5, located at the Istituto di Biostrutture e Bioimmagini C.N.R., University of Naples Federico II. Spectra were also acquired on an INCA (Consorzio Interuniversitario Chimica per L'Ambiente) Varian Inova 500 MHz. NMR characterization was performed in TFE/$H_2O$ 90:10 (v/v) at 298 K. The sample was prepared by dissolving the peptide (approximately 5.1 mg) in [$D_3$]TFE (0.75 mL, 99% isotopic purity, Aldrich) and $H_2O$ (0.075 mL). Chemical shifts were referenced to internal sodium 3-(trimethylsilyl)[2,2′,3,3′-$d_4$] propionate (TSP). Two-dimensional experiments, such as total correlation spectroscopy (TOCSY),[34] nuclear Overhauser effect spectroscopy (NOESY),[35] rotating frame Overhauser effect spectroscopy (ROESY),[36] and double quantum-filtered correlated spectroscopy (DQFCOSY)[37] were recorded by the phase-sensitive States – Haberkorn method. The data file generally consisted of 512 and 2048 data points (4096 for DQFCOSY) in the $\omega 1$ and $\omega 2$ dimensions, respectively. TOCSY experiments were acquired with a 70-ms mixing time, and the water resonance was suppressed by the watergate sequence.[38] NOESY experiments were acquired with a 100-, 200-, and 300-ms mixing time, and ROESY experiments with a 100-ms mixing time, by using a continuous spin-lock. Off-resonance effects, because of the low-power spin-lock field, were compensated by means of two 90° hard pulses before and after the spin-lock period.[39] The water resonance was suppressed by low-power irradiation during the relaxation delay and, for NOESY, during the mixing time. 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. Temperature coefficients of amide protons were measured from one-dimensional spectra and from TOCSY spectra, acquired with 4-K data points in the 298 – 310 K temperature range. NOE analysis was achieved by means of NOESY spectra. NOE intensities were evaluated by integration of cross-peaks by using the appropriate Varian software and then converted into inter-proton distances by using the $r^{-6}$ relationship for rigid molecules.[16] *Geminal* $\beta - \beta'_{CH2}$ protons of Asn14 were chosen as reference with a distance of 1.78 Å. In accordance with Wüthrich's method,[16] identification of amino acid spin systems was performed by comparison of TOCSY and DQF-COSY data, while sequential assignment was obtained by the analysis of NOESY spectra.

**Computational analysis**

**DYANA calculations**: Torsion angle dynamics calculations were carried out by using the DYANA program.[19] The library program was modified for the N- and C-terminal residues. A total of 200 three-dimensional structures were obtained by using inter-proton distances evaluated from NOEs (raised of 20%) as upper limits, without using stereospecific assignments. One hundred conformers were calculated by using the standard parameters of the DYANA program. To improve convergence, the redundant dihedral angle constraints (REDAC) strategy[20] was also employed and 100 more structures

  

were calculated by carrying out five REDAC cycles. Dihedral angle constraints were created with an angle cutoff for the target function (TF) equal to 0.8 Å$^2$ in the first step, 0.6 Å$^2$ in the second, and 0.4 Å$^2$ in the third. In the fourth step the structures were calculated by 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 ($L_{23}$) with all the experimental restraints.

**Energy minimization**: The 30 DYANA structures with the lowest values of target function were subjected to restrained energy minimization by the SANDER module of the AMBER 6.0 package.[21] The 1991 version of the force field was used[40] with a distance-dependent dielectric constant $\varepsilon = r_{ij}$. The charge of the ionizable groups was reduced to 20% of its full value to reduce possible artifacts because of the in vacuo simulations. A distance cutoff of 12 Å was used in the evaluation of nonbonded interactions. Distance restraints were applied as a flat well with parabolic penalty within 0.5 Å outside the upper bound; a linear function beyond 0.5 Å with a force constant of 16 kcal mol$^{-1}$ Å$^{-2}$ was used. The restrained energy minimization was carried out with a total of 2000 steps of conjugate gradients minimization, after 200 of the steepest descents, for each conformer. The molecular graphics program MOLMOL was employed to perform the structural statistics analysis.

*We are grateful to Dr. Orlando Crescenzi, University Federico II of Naples, for helpful discussions about the structure calculations and advice in the use of AMBER. We also thank Dr. Nabil G. Seidah, Clinical Research Institute of Montreal (IRCM), Canada, for providing furin.*

[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] M. Moulard, E. Decroly, *Biochim. Biophys. Acta* **2000**, *1469*, 121 – 132.
[5] For a sequence alignment of multibasic cleavage sites in precursor proteins, see: K. Nakayama, *Biochem. J.* **1997**, *327*, 625 – 635.
[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] N. Brakch, M. Dettin, C. Scarinci, N. G. Seidah, C. Di Bello, *Biochem. Biophys. Res. Commun.* **1995**, *213*, 356 – 361.
[9] E. Decroly, S. Wouters, C. Di Bello, C. Lazure, J. M. Ruysschaert, N. G. Seidah, *J. Biol. Chem.* **1996**, *271*, 30442 – 30450.
[10] R. Oliva, M. Leone, L. Falcigno, G. D'Auria, M. Dettin, C. Scarinci, C. Di Bello, L. Paolillo, *Chem. Eur. J.* **2002**, *8*, 1467 – 1473.
[11] M. Moulard, L. Challoin, S. Canarelli, K. Mabrouk, H. Darbon, *Biochemistry* **1998**, *37*, 4510 – 4517.
[12] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, *Trends Biochem. Sci.* **2000**, *25*, 147 – 150.
[13] D. A. Rozwarski, K. Diederichs, R. Hecht, T. Boone, P. A. Karplus, *Proteins* **1996**, *26*, 304 – 313.
[14] S. Fiori, S. Mammi, E. Peggion, P. Rovero, S. Pegoraro, R. P. Revoltella, *J. Pept. Sci.* **1997**, *3*, 336 – 346.
[15] N. Greenfield, G. D. Fasman, *Biochemistry* **1969**, *8*, 4108 – 4116.
[16] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[17] D. S. Wishart, B. D. Sykes, F. M. Richards, *J. Mol. Biol.* **1991**, *222*, 311 – 333.
[18] K. Osapay, D. A. Case, *J. Biomol. NMR* **1994**, *4*, 215 – 230.
[19] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283 – 298.
[20] P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1991**, *1*, 447 – 456.
[21] a) D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, P. A. Kollman, AMBER 6, University of California, San Francisco, **1996**; b) D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. deBolt, D. Ferguson, G. Seibel, P. A. Kollman, *Comp. Phys. Commun.* **1995**, *91*, 1 – 41.
[22] V. Pavone, G. Gaeta, A. Lombardi, F. Nastri, O. Maglio, C. Isernia, M. Saviano, Biopolymers **1996**, *38*, 705 – 721.
[23] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51 – 55.
[24] P. Gros, C. Betzel, Z. Dauter, K. S. Wilson, W. G. Hol, *J. Mol. Biol.* **1989**, *210*, 347 – 367.
[25] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157 – 162.
[26] R. J. Siezen, J. W. Creemers, W. J. Van De Ven, *Eur. J. Biochem.* **1994**, *222*, 255 – 266.
[27] 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.
[28] S. Segawa, T. Fukuno, K. Fujiwara, Y. Noda, *Biopolymers* **1991**, *31*, 497 – 509.
[29] M. A. Jimenez, M. Bruix, C. Gonzales, F. J. Blanco, J. L. Nieto, J. Herranz, M. Rico, *Eur. J. Biochem.* **1993**, *211*, 569 – 581.
[30] M. T. Reymond, S. Huo, B. Duggan, P. E. Wright, H. J. Dyson, *Biochemistry* **1997**, *36*, 5234 – 5244.
[31] D. E. Callihan, T. M. Logan, *J. Mol. Biol.* **1999**, *285*, 2161 – 2175.
[32] S. J. Hubbard, R. J. Beynon, J. M. Thornton, *Protein Eng.* **1998**, *11*, 349 – 359.
[33] N. Brakch, N. Lazar, M. Panchal, F. Allemandou, G. Boileau, P. Cohen, M. Rholam, *Biochemistry* **2002**, *41*, 1630 – 1639.
[34] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *65*, 355 – 360.
[35] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *63*, 207 – 213.
[36] A. Kumar, G. Wagner, R. R. Ernst, K. Wüthrich, *J. Am. Chem. Soc.* **1981**, *103*, 3654 – 3658.
[37] U. Piantini, O. W. Sørensen, R. R. Ernst, *J. Am. Chem. Soc.* **1982**, *104*, 6800 – 6801.
[38] M.Piotto, V. Saunder, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661 – 665.
[39] C. Griesinger, E. E. Ernst, *J. Magn. Res.* **1987**, *75*, 261 – 271.
[40] S. J. Weiner, P. A. Kollmann, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **1986**, *7*, 230 – 238.

Received: December 16, 2002;
Revised version: April 23, 2003 [F541]

  