Structural Investigation of the HIV-1 Envelope Glycoprotein gp160 Cleavage Site 3: Role of Site-Specific Mutations**

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

Proteolytic processing of HIV gp160 to produce gp120 and gp41 is performed by PC enzymes. This process is a prerequisite for the virus infectivity, since both gp120 and gp41 participate in the virus HIV-1 entry mechanism. The structure of the gp120/gp41 junction remains to be elucidated, and the structural features required for molecular recognition between HIV-1 gp160 and proteolytic enzymes have not been clarified. Furin is the best PC candidate for the gp160 proteolytic processing known to date.

In previous studies on model peptides, we have shown the relevance of an N-terminal helix for the proper recognition of the gp160 processing site by furin. Here we analyze the effect of point mutations in peptides lacking a regular N-terminal helix. To this end, we present the structure–activity characterization of three peptide analogues of the HIV gp160 processing site that all present mutations in proline at positions P3 and/or P2′, while sharing the same N-terminal sequence, containing helix-breaking D*-amino acids. Conformational analysis of the peptides was carried out in solution by NMR techniques, and furin's efficiency in cleaving them was measured. Structural findings are presented and discussed in relation to the different exhibited activity.*

Introduction

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

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

Analogously to prohormones, viral glycoproteins present multibasic cleavage sites and, more specifically, consensus sequences of the Lys/Arg-Xaa-Lys/Arg-Arg type. The proteolytic cleavage of the HIV-1 glycoprotein gp160 precursor occurs at the carboxyl side of the Arg508-Glu-Lys-Arg511 sequence (site 1) in over 85% of cases,[5] although a second putative cleavage site—Lys500-Ala-Lys-Arg503 (site 2)—is present eight residues upstream. Site-directed mutagenesis studies[6] showed that basic residues within the site 1 recognition sequence are necessary for the proteolytic processing. The role of the non-physiological site is still hypothetical, however.

While the first PC structure—for mouse furin, which is highly homologous (sequence identity 98%) with the human form[7]—has recently been reported, the structural features required for 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 rely on protein structure predictions and conformational studies of model peptides.

We have already reported on the structural analysis of a fragment spanning the Pro498–Gly516 sequence: p498.[8] The 19-residue peptide p498, spanning the junction between gp120 and gp41, and containing sites 1 and 2, was shown by Brakch et al.[9] to be recognized and duly digested by furin at site 1;[9,10] this suggests that contributions of specific secondary structure motifs from surrounding amino acids can be reproduced even in a middle-sized peptide. In our study we found a loop exposing the physiological cleavage site 1, and a helix at the N-terminal side.

More recently we reported on the structure and activity of the 23-residue peptide h-REKR, the key features of which were the introduction of a large helical sequence replacing that of site 2 and located at the N terminus.[11] We showed that the propensity of the peptide to be processed by furin is high and

[a] *Dr. L. Falcigno, Dr. R. Oliva, Prof. G. D'Auria, Dr. M. Maletta, Prof. L. Paolillo*
Dipartimento di Chimica, Università di Napoli "Federico II"
Complesso Universitario Monte S. Angelo, via Cintia, 80126 Napoli (Italy)
Fax: (+39) 081-674287
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 6, 80134 Napoli (Italy)

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

[**] *For Part 2, see: R. Oliva, L. Falcigno, G. D'Auria, M. Dettin, C. Scarinci, C. Di Bello, L. Paolillo,* ChemBioChem ***2003***, 4, *727–733.*

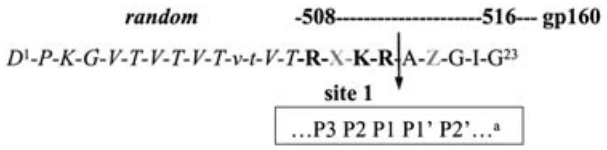
Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

 DOI: 10.1002/cbic.200400181

not significantly affected by such modification, in relation to the full-native analogue p498.

"Helix + loop" thus seems to be a suitable secondary structure motif for digestion by furin. These results supported the hypothesis that the processing site in gp160 may be enclosed in a loop, preceded by a helix helping in orienting it correctly.

Here we have undertaken a structure–activity study in a series of analogues designed to bear helix-breaking amino acids at the N terminus of the processing site (Scheme 1).



Scheme 1. *Amino acid sequence of the analogues **1**–**4**. X and Z symbols, reported in gray, indicate Proline mutations (see Table 1). An arrow is used to indicate the cleavage locus for the recognized processing site 1. [a] Schechter and Berger nomenclature.*[12]

These analogues (**1**–**4**, Table 1) are 23-mer peptides containing the native gp160 sequence Arg508–Gly516, containing site 1, at the C terminus, and, at the N terminus, model sequences indicated for simplicity as "random" (*r*-). In analogues **2**–**4**, mutations in prolines are also made at positions P3 and/or P2′ (according to the nomenclature of Schechter and Berger)[12].

Table 1. *Names and sequences of analogues with digestion experiment results; p498 reproduces the native gp160 sequence.*[8] *The non-native N-terminal residues are given in italics and indicated by r- (random). The processing site residues are reported in bold. Mutations in prolines are shown in gray.*

	Name	Sequence	Substrate processing by Furin[a]
	p498 (498–516 gp160)	P¹T**KAKRR**VVQ**REKR**↓AVGIG¹⁹	100%
1	*r*-**REKR**[14]	*D¹PKGVTVTVTvtVT***REKR**↓AVGIG²³	21%
2	*r*-**RPKR**	*D¹PKGVTVTVTvtVT***RPKR**↓AVGIG²³	100%
3	*r*-**REKR-Pro20**	*D¹PKGVTVTVTvtVT***REKR**↓APGIG²³	29%
4	*r*-**RPKR-Pro20**	*D¹PKGVTVTVTvtVT***RPKR**↓APGIG²³	77%

[a] Digestion percentages after 3 h incubation with human furin at 37 °C.

The random N-terminal sequence of analogues **1**–**4** contains two D-residues (val11 and thr12) and corresponds to the Asp5–Thr18 sequence of the peptide D15 in a series of analogues designed and characterized by Krause and colleagues.[13] Krause and co-workers showed, by circular dichroism (CD) analysis, that the double Val-Thr D-amino acid substitution dramatically decreases the amount of helical structure and that no propensity for ordered secondary structure appears to exist.

In our analogues, mutations were also made at positions P3 and P2′ in order to gain insight in the role of these substrate positions. From modeling studies, protein engineering, and comparative analysis of native substrates, it is indeed known that furin is highly specific towards positions P1, P2, and P4, whereas the role of the P3 and P2′ positions is still not clear. Proline residues were inserted in such positions, to introduce structural constraints.

Digestion experiments with furin, performed on our peptides showed that they all are recognized by furin, but with significantly different efficiencies.

The introduction of helix-breaking amino acids into the N terminus in analogue **1** (with the full-native C-terminal sequence) dramatically reduces the cleavage efficiency. Mutations with prolines (analogues **2**–**4**) seem, instead, to affect the activity significantly. In particular, the proline mutation at P3 seems to enhance the cleavage efficiency.

The previously reported[14] NMR conformational analysis of analogue **1** (namely *r*-REKR) failed to demonstrate a well defined secondary structure, in particular for the N-terminal region, where spectral overlapping prevents unambiguous assignment of structurally diagnostic NOEs.

Here we present the structure–activity analysis of three other analogues of the series incorporating proline substitutions (analogues **2**–**4**). In particular, these analogues **2**–**4** share the same N-terminal (*r*-) sequence with the following mutations: analogue **2** (*r*-RPKR) presents the E509P mutation at P3, analogue **3** (*r*-REKR-Pro20) presents the V513P mutation at P2′, and analogue **4** (*r*-RPKR-Pro20) presents both the E509P and V513P mutations at P3 and P2′.

Conformation analysis, carried out by NMR techniques and by molecular modeling, is presented and related to the different exhibited activities.

Results

Cleavage of synthetic peptides by furin

Digestion experiments with furin were performed. All peptides were incubated with furin for 1 h and for 3 h at 37 °C. Products were identified by high-performance liquid chromatography (HPLC) purification followed by a MALDI mass spectroscopic analysis. The degree of substrate conversion was determined by integrating the peak area of the uncleaved substrate and comparing it with the peak area of the furin-untreated sample.

The digestion results are summarized below (see also Table 1).

The substrate *r*-REKR (analogue **1**), containing an N-terminal segment designed to adopt a random conformation, reaches 21% conversion after 3 h incubation and only 9% digestion after 1 h. The mutations introduced in the *r*-REKR sequence (analogues **2**–**4**) also gave different and interesting results:

1) Analogue **2**, with Glu16→Pro substitution (P3 position), is extensively digested by furin after 1 h (91%) and it is the best synthetic substrate among the peptide analogues under investigation.
2) Analogue **3**, with the modification Val20→Pro at the P2′ position, is processed with comparable efficiency to analogue **1**, with no proline mutation.

3) Analogue **4**, with both point mutations described above, shows 40% conversion over 1 h and 77% conversion after 3 h incubation.

Analogue 2 (*r*-RPKR): conformational analysis

NMR data: NMR analysis was performed in a TFE/H_2O (90:10, *v/v*) solution. The complete sequential assignment was achieved by the standard procedure proposed by Wüthrich.[15] Proton chemical shifts are presented as Supporting Information (Table S1).

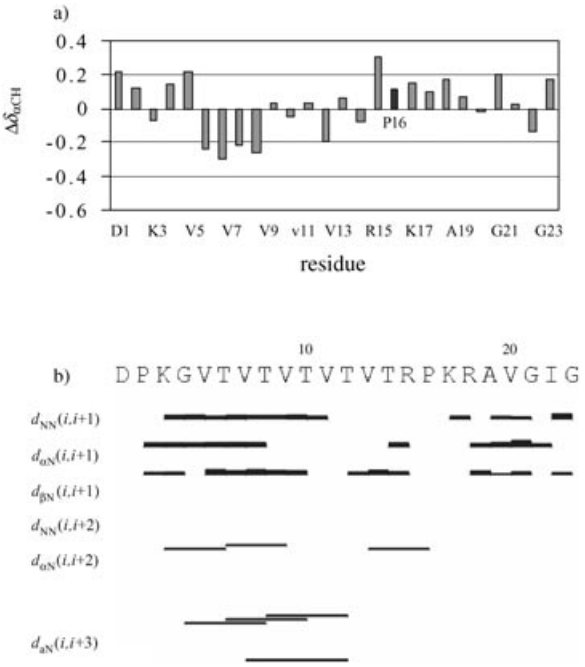


Figure 1. *a) Chemical shift deviations from the random-coil values for αCH protons[16] of analogue **2**. For Gly4 and Gly21 both the αCH and the α′CH chemical shift deviations are reported. b) Summary of analogue **2**'s most relevant NOE effects in TFE/H_2O (90:10). Val11 and Thr12 are D-amino acids.*

Deviations of the αCH chemical shifts from random-coil values are reported in Figure 1a. The negative deviations observed for the Val5–Thr8 segment point to a helical or turn structure in this region.[16] In the C-terminal segment Arg15–Val20, positive deviations are observed instead, pointing to an extended local conformation, probably due to the presence of the Pro16. The NOE pattern (Figure 1b) shows a few medium-range contacts (three α_i–β_{i+3} [3–6, 5–8, 7—10] and one α_i–NH_{i+4} [6—10]), diagnostic of a helical conformation, in the N-terminal region.

In the C-terminal segment, starting from Pro16, a pattern of stronger α_i–NH_{i+1} contacts, relative to the corresponding NH_i–NH_{i+1}, is observed, which is diagnostic of an extended conformation. A couple of α_i–NH_{i+2} contacts, pointing to the presence of turns, are also observed in the N-terminal tail. A set of 262 experimental constraints from NOE data (130 intraresidual, 103 sequential, and 29 medium-range) was determined for structure calculations. These constraints were imposed as upper bounds on interproton distances, in order to sample the conformational space, by torsion angle dynamics (DYANA program).[17]

Structure calculations: A first set of 100 conformers was calculated with the DYANA program. One hundred additional structures were then calculated by the redundant dihedral angle constraints (REDAC) strategy, in order to improve the convergence.[18] The REDAC structures are indeed more compatible with the experimentally obtained data (for details see the Experimental Section). The 30 DYANA structures with the lowest target function values were then subjected to restrained

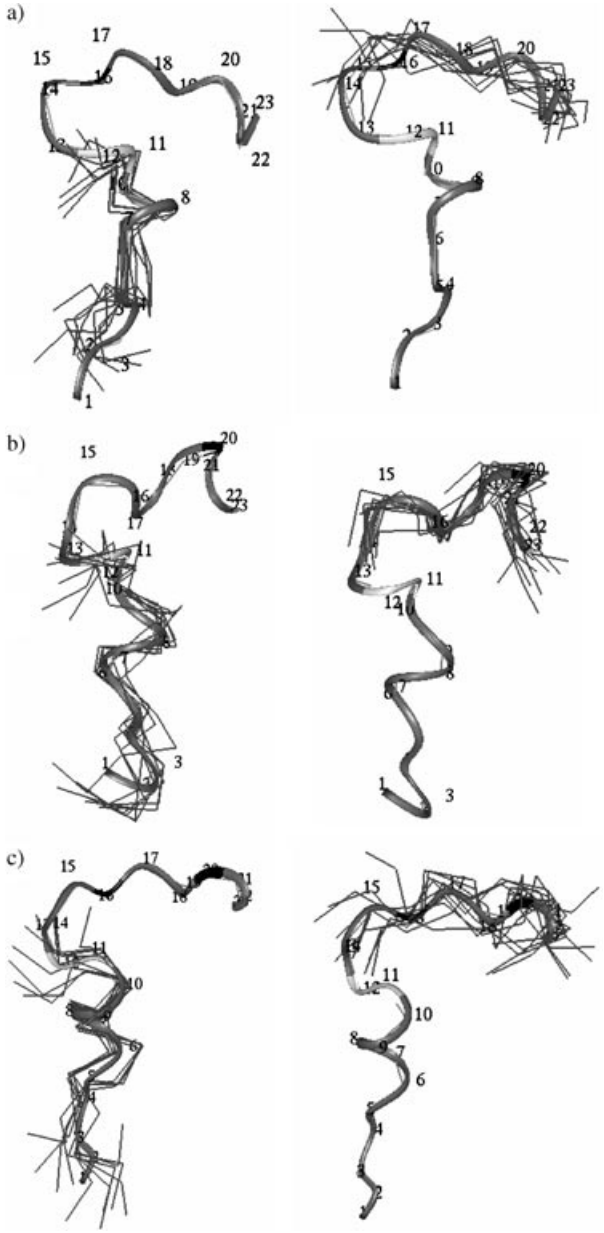


Figure 2. *The best backbone superposition of the 2–12 (left) and 14–22 (right) segments in the ten selected AMBER structures: a) analogue **2**, b) analogue **3**, and c) analogue **4**. The mean structures are shown as gray ribbons. Proline mutations are shown in black and D-amino acids in white.*

energy minimization by use of the SANDER module of the AMBER 6.0 package.[19,20] The best 10 structures in terms of the agreement with experimental restraints, out of those with a residual restraint energy lower than -138 kcal mol^{-1}, were selected to represent the analogue **2** solution structure (Figure 2a). The whole N-terminal fragment, from Pro2 to D-Thr12, assumes a well defined conformation (average root mean square deviation (rmsd) on backbone 1.55 ± 0.56 Å calculated from the best 10 structures).

In particular, the residues from Thr6 to Thr10 are involved in an α-helix (mean global rmsd on backbone 0.69 ± 0.46 Å) in all ten structures, interrupted by the pair of adjacent D-amino acids, DVal11–DThr12. A type I β-turn is further present on residues Asp1–Gly4 in five out of the ten selected structures.

The C-terminal side, from Arg15 to Ile22, appears globally quite extended although it is less defined than the N-terminal one (mean global rmsd on backbone of Arg15–Ile22 1.81 ± 0.50 Å). On the C-terminal residues Val20–Gly23, a non-canonical four-residue bending is observed.

The region DThr12–Arg15 instead presents large variability.

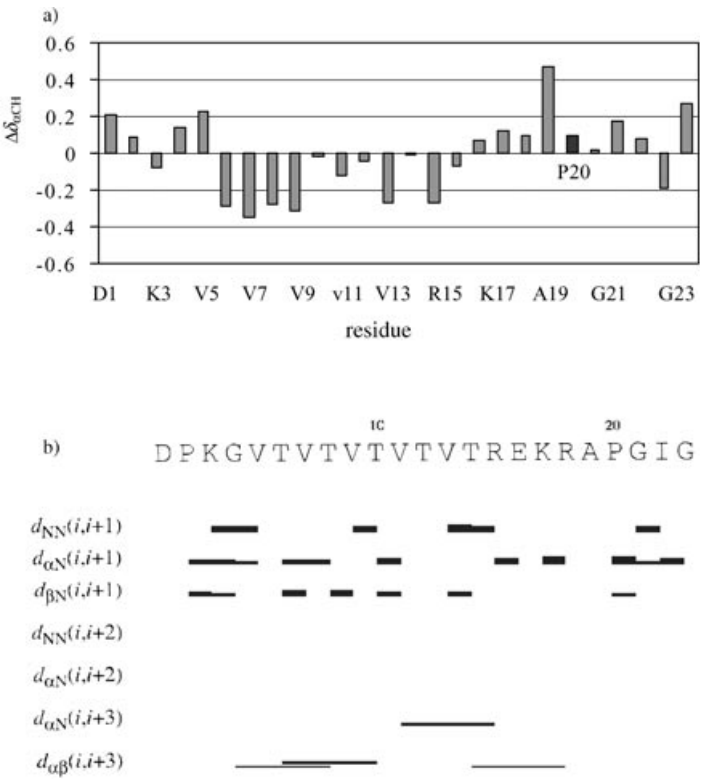


Figure 3. *a) Chemical shift deviations from the random-coil values for the αCH protons[16] of analogue **3**. For Gly4 and Gly21 both the αCH and the α'CH chemical shift deviations are reported. b) Summary of analogue **3**'s most relevant NOE effects in TFE/H_2O (90:10). Val11 and Thr12 are D-aminoacids.*

Analogue 3 (*r*-REKR-Pro20): conformational analysis

NMR data: Analogue **3** has a large N-terminal random sequence,[13] followed by the gp160 native sequence Arg508–Gly516 presenting the mutation V513P (Table 1).

NMR analysis in a TFE/H_2O (90:10, *v*/*v*) solution allowed a complete sequential assignment. Proton chemical shifts are presented as Supporting Information (Table S2).

At the N terminus, the chemical shift deviations reported in Figure 3a point to a helical or turn structure in the Val5–Thr8 segment.[16] In the C-terminal Glu16–Val20 segment, weak positive deviations are observed instead, pointing to an extended local conformation, very probably due to the presence of the Pro20. The NOE pattern (Figure 3b) shows a few medium-range contacts diagnostic of a helical conformation, in the Thr4–DVal10 region, two α_i–β_{i+3} contacts (4–7, 6—9) among them. Several medium-range contacts are also present in the 11–17 region.

A set of 233 NOE experimental constraints (128 intraresidual, 78 sequential, and 27 medium-range) was used for structure calculations. These constraints were imposed as upper bounds on interproton distances in order to sample the allowed conformational space by torsion angle dynamics (DYANA program).[17]

Structure calculations: The molecular model of analogue **3** was calculated with the procedure already outlined for analogue **2** (see also the Experimental Section). The ten best structures among those with a residual restraint energy lower than -143 kcal mol^{-1} were considered to represent the analogue **3** solution structure. These structures belong to one conformational family (mean global backbone rmsd on residues 2–22: 3.44 ± 1.39 Å). The mean global backbone rms deviations in the 2–14 and 15–22 regions are 1.92 ± 0.85 Å and 1.38 ± 0.88 Å, respectively. Figure 2b shows the mean AMBER structure representative of the entire conformational family, superimposed on all the selected structures. It presents a quite unusual N-terminal conformation, also including the residues D-Val11 and D-Thr12: one turn of an α-helix spans the Thr6–Val9 segment, followed by two consecutive β-turns, one on residues D-Val11–Thr14 and one on Thr14–Lys17, enclosing a γ-turn around Arg15. At the C terminus, the expected β-turn around Pro20–Gly21 is not found; instead, a large non-canonical bending involving residues Lys17–Gly21 is observed, with the presence of Glu16–Arg18 and Ala19–Gly21 main-chain hydrogen bonds. In this large bending, the residues from Lys17 to Pro20 assume a quite extended conformation. (Pro20 forms a distorted equatorial γ-turn in half of the structures). A fairly tight network of main-chain hydrogen bonds involves those residues located around the cleavage site 1, in agreement with the low-temperature coefficients for the Thr14–Arg18 amide protons (data not shown).

Analogue 4 (*r*-RPKR-Pro20): conformational analysis

NMR data: Analogue **4** has the previously mentioned large N-terminal random sequence,[13] followed by the gp160 native sequence Arg508–Gly516 presenting the double mutation E509P and V513P (Table 1).

Proton chemical shifts are reported as Supporting Information (Table S3).

The αCH chemical shift deviations for each residue number are reported in Figure 4a. Negative deviations for the Val5–

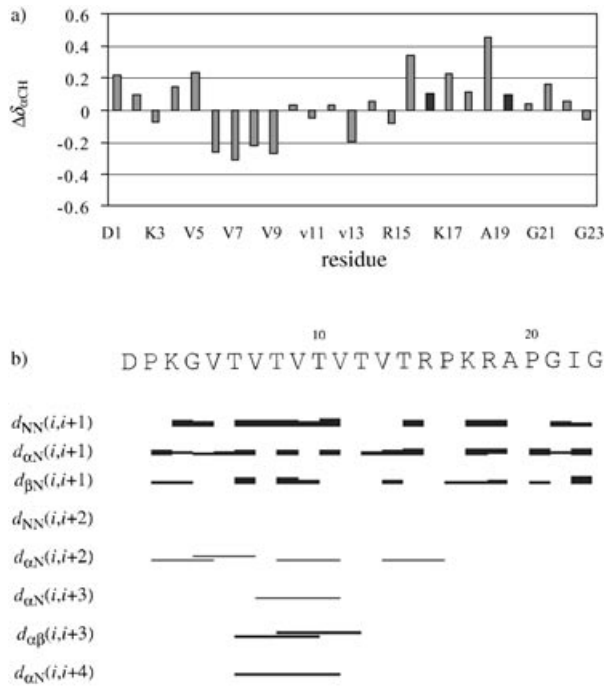


Figure 4. *a) Chemical shift deviations from random-coil values for the αCH protons[16] of analogue **4**. For Gly4 and Gly21 both the αCH and the α′CH chemical shift deviations are reported. b) Summary of analogue **4**'s most relevant NOE effects in TFE/H_2O (90:10). Val11 and Thr12 are D-amino acids.*

Thr8 segment point to a helical or turn structure in this region.[16] In the C-terminal segment Arg15–Gly21, positive deviations are observed instead, pointing to an extended local conformation, very probably due to the Pro16 and Pro20 substitutions.

On the basis of the chemical shift comparison, analogue **4** would thus be expected to exhibit an N-terminal conformation closely resembling that of analogue **2**, from Asp1 to Pro16, and a C-terminal conformation very similar to that of analogue **3** in the Arg18–Gly23 segment. The Arg15–Arg18 segment would then be expected, on the basis of the αCH chemical shift deviations, to be found in an extended conformation.

The NOE pattern (Figure 4b) shows a few medium-range contacts diagnostic of a helical conformation in the Thr6–DThr11 region (among them one α_i–N_{i+3} [7–10], two α_i–β_{i+3} [6–9, 8–11], and one α_i–N_{i+4} [6–10]).

A set of 300 experimental constraints from NOE data (148 intraresidual, 121 sequential, and 31 medium-range) was used for structure calculations.

Structure calculations: In analogy with the previously described cases, the ten best structures (among those with a residual restraint energy lower than -100 kcal mol^{-1}) were selected to represent the analogue **4** solution structure. They belong to one conformational family (mean global backbone rmsd on residues 2–22: 3.52 ± 0.88 Å). The mean global backbone rmsds are 2.44 ± 1.06 Å and 2.19 ± 0.56 Å in the peptide regions 2–14 and 15–22, respectively. The mean AMBER structure representative of the entire conformational family is shown in Figure 2c, superimposed on the selected structures. One α-helix turn is present in the Thr6–Val9 segment, while the two D-residues D-Val11 and D-Thr12 contribute to a folded structure through a kink characterized by H-bonds formed by D-Val11 with both Val7 and Thr8 and dihedral angles typical of a type I β-turn around residues Thr10–DVal11. Two γ-turns—axial and equatorial, respectively—around Val13 and Thr14 invert the direction of the chain. At the C terminus, the expected β-turn around Pro20–Gly21 is not found, whereas a fairly extended conformation is observed for the Arg15–Ile22 segment.

Discussion

The results obtained from the furin digestion experiments demonstrate the importance of the N-terminal conformation and of the P3 and P2′ residues in determining the efficiency of the enzyme. These results can be summarized as follows:

1) When no mutation occurs in or close to the cleavage site, but different sequences are exhibited at the N terminus (see analogues p498,[8] h-REKR,[11] *r*-REKR,[14]) the efficiency of cleavage seems to be driven by degree of order of the N terminus.
2) In the case of mutations (analogues **2**–**4**), proline mutation at the P3 position seems to enhance the cleavage efficiency, while proline mutation at P2′ seems not to affect it (or to reduce it if a proline is already present at P3).

D-Residues

All studied analogues contain a pair of adjacent D-amino acids: D-Val11 and D-Thr12.

Several studies have shown that the replacement of amino acids by their enantiomers induces destabilization of ordered secondary structures.[13,21,22] The substitution of amino acids by their enantiomers is thus advantageous for structure–activity studies, to probe the relationship between conformational domains and bioactivity, for example, as it affects only the structure without changing properties such as side-chain hydrophobicity, functionality, or charge distribution.

The incorporation of an adjacent pair of D-amino acids was shown to enhance the effect of structure disturbance. In particular, Krause and colleagues analyzed, by CD, NMR and HPLC methods, the helix-destabilizing abilities of the 19 standard proteinogenic counterpart D-amino acids, when inserted as pairs in a 16-mer model amphipathic α-helix.[21] They showed that all D-amino acids have a helix-destabilizing effect, which is strongly dependent on their side-chains and not related to the structure propensity of the corresponding L-amino acids. Accordingly, the D-isomers of bulky and β-branched amino acids, particularly D-Val and D-Thr, were shown to be the most effective in destabilizing the amphipathic helix. From the NOE pattern, Krause and colleagues suggested that D-residues might induce turn-like structures in the helix, causing kinks.[21]

Our NMR study confirms the propensity of the Asp1–Thr14 segment to adopt a helical conformation in the presence of

TFE, as well as partial helix disruption (by induction of a kink) due to the insertion of the D-amino acid pair.

From our structure elucidations it appears that the last L-residue before the double D-insertion (Thr10) and the first D-residue of the pair (val11) exhibit dihedral angles in unusual regions of the Ramachandran map, while the dihedral angles of the second D-residue (thr12) and of the following L-residue (Val13) are consistent with an α_R region (data not shown). Such an unusual conformation for Thr10–val11 backbones seems to stabilize helix capping between their backbone amide protons and the carbonyls of residues three or four positions upstream, inserted in the helix turn. This can be seen in Figure 5, which shows a superposition of the N-terminal regions in the three analogues' mean structures. (The rmsd for the backbone superposition of Pro2–thr12 ranges from 1.2 to 2.2 Å). These structural features help to explain the helix-disrupting ability of two consecutive D-residues inserted into a stretch of L-amino acids.

Figure 5. *A superposition of the mean N-terminal region structures of analogues **2** (purple), **3** (green), and **4** (blue): ribbon representation. The backbone H-bonds between 7–10 and 7–11 residues are shown in green.*

Furin–substrate binding mode: The crystal structure of mouse furin, highly homologous to the human form (sequence identity 98%), has been recently determined at a 2.6 Å resolution.[7] A tetrapeptide analogue inhibitor (decanoyl-Arg-Val-Lys-Arg-chloromethylketone) is bound in its catalytic site, occupying the substrate P1–P4 positions (according to Schechter and Berger's nomenclature[12]).

As already hypothesized on the basis of homology models for the furin catalytic domain (Siezen and co-workers[23,24] and our own laboratories[8,11]), the furin binding region can be described as a narrow channel, able to accommodate several substrate residues around the multibasic sequence (see also Siezen et al.[23]). The furin catalytic triad (Ser368–His194–Asp153) is located at the bottom of the channel, together with several surface-exposed acidic residues, making specific electrostatic interactions with the basic amino acids at positions P1, P2, and P4.

As in the case of the substrate positions currently under investigation, the inhibitor Val side chain at P3 in the mouse furin structure extends into the bulk solvent, pointing outside the binding site, in agreement with the lack of specificity observed at this position. A proline can be easily accommodated at P3, as shown in Figure 6a in the form of the furin-inhibitor structure, bearing a mutation in Pro at P3, as in peptide analogues **2** and **4**.

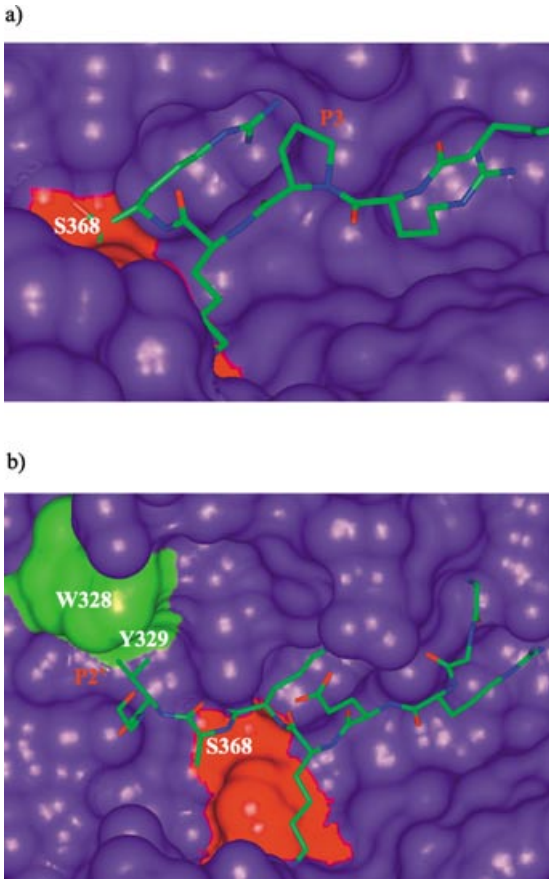


Figure 6. *Furin catalytic region shown as a purple Connolly surface, with catalytic triad residues (Ser368–His194–Asp153) in red. a) The structure of mouse furin-tetrapeptide inhibitor, with a proline mutation modeled at position P3. b) The model of human furin with bound eglin c inhibitor, modified to mimic the gp160 sequence around the processing site: V-Q-R-E-K-R↓A-V. The Val residue at P2′ interacts with the hydrophobic W328–Y329 residues (in green) in the enzyme binding region.*

A model of the human furin–eglin c inhibitor complex was built in order to model the P2′ position, absent in the mouse furin structure. The inhibitor sequence was modified in order to mimic the gp160 processing site: namely, Val-Gln-Arg-Glu-Lys-Arg↓Ala-Val.[8] This model shows that residues W328 and Y329 are located in a very favorable position for interaction with the hydrophobic Val residue at P2′, thus defining the preference for a medium-sized hydrophobic residue at such a position (Figure 6b).

Conclusion

In this paper we discuss the activity and conformational properties of three peptide analogues related to the HIV-1 gp160 processing site. These peptides were designed to exhibit a common N-terminal random sequence and proline mutations at positions P3 and/or P2′, in order to elucidate the relative importance of these structural features in the gp160 processing.

A proline at P3 has been shown to increase the rate of molecular recognition in our unstructured peptides (analogue **2** vs. **1**) while proline mutation at P2′ does not seem to affect

the cleavage efficiency (analogue **3** vs. **1**), but reduces it if a proline is already present at position P3, as in analogue **2** (analogue **4** vs. **2**).

It is known that proline residues are present at position P3 in the canonical protein inhibitor BPTI[25] and in some furin substrates, such as human pro-factor IX and rat pro-endopeptidase.[26] However, P3 does not seem to be a conserved position in furin substrates (although hydrophilic, particularly basic, residues are favored). Moreover, from the current furin model the side chains of residues located at P3 indeed do not point toward the furin catalytic region and, in addition, the furin subsite S3 does not appear to be a distinct site.

On the contrary, no proline substitution at P2′ was observed for any furin substrate analyzed, where a large predominance of hydrophobic residues was observed instead (specifically, a Val is found in 12 out of the 37 total analyzed cases; see Nakayama[26]). In our model for the furin catalytic domain, a hydrophobic interaction between the substrate residue Val47, at P2′, and the furin sub-site residue Y222 (4 Å away) is indeed observed.

According to Bode and Huber,[25] the binding loops containing the proteolytic site for Ser-protease substrates exhibit a quite characteristic extended main-chain conformation, from P3 to P3′, when bound to proteases. Such a conformation exposes the side chains flanking the scissile bond (except for P3) in such a way that they protrude from the supporting scaffold toward the enzyme's active site. When free, however, substrates' proteolytic sites are usually quite different in structure; they thus undergo a conformational change in order to facilitate the binding into the proteinase active site.[27] Globally, the structural effect of prolines in the mutants under study appears to be to induce an extended conformation in a segment of four–six adjacent residues. Residues preceding prolines show, as expected,[28] a marked preference for the β-region. This is apparent from the αCH chemical shift deviation diagrams, in which residues preceding prolines always exhibit strong positive deviations from random-coil values, indicative of extended conformations.

The high efficiency of the cleavage of analogue **2**, comparable to that of the p498 native sequence, thus seems to be governed at first by the C-terminal conformation, which could play an important role in promoting recognition. The presence of proline at P3 indeed introduces constraints on the dihedral angles of the adjacent residues P4–P2′ toward the expected values for recognition, according to the Bode and Huber model.[25] Therefore, the proline mutation at P3 should entropically favor the adoption of the correct orientation by the substrate.

In analogue **3**, the observed activity is probably due to a balance between the entropic gain due to the Pro residue at position P2′ and the loss of a specific interaction required in such position. Indeed the presence of the proline at P2′ induces the correct dihedral angles for the P2–P2′ residues, and particularly for the Ala residue at P1′, although less efficiently than at P3. Therefore, according to the Bode and Huber model,[25] the proline mutation at P2′ could be structurally favored. At the same time, the proline mutation at P2′ is enthalpically unfavorable, as hydrophobic interactions between the Val residue at P2′ and the furin subsite S2′ are lost.

Analogue **4** exhibits an intermediate activity. It shares with the most active analogue, **2**, the conformation of a large segment (Asp1–Arg18) enclosing the site 1 processing residues. At the same time, the presence of a proline at the P2′ position with the subsequent loss of a hydrophobic interaction with S2′ may be invoked to explain the lower efficiency of cleavage in relation to analogue **2**. However, the severe structural constraint due to the presence of two prolines, separated by only three residues, could also play a role.

Experimental Section

Synthesis: All the analogues, reported in Table 1, were synthesized by solid-phase peptide synthesis on an automated peptide synthesizer Model 431 A (Applied Biosystems, Forster City, CA, USA) by standard Fmoc chemistry. Rink Amide MBHA Resin (0.49 mmol g^{-1}) was employed as solid support. Amino acids were incorporated by the use of 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate/1-hydroxybenzotriazole (HBTU/HOBt) reagents. The following side chain protecting groups were used:

- Thr: *t*Bu,
- Lys: *tert*-butyloxycarbonyl (Boc),
- Arg: 2,2,5,7,8-pentamethyl-chroman-6-sulfonyl (Pmc), and
- Asp and Glu: O*t*Bu.

Double couplings were introduced in the sequence 1–13 and for Gly23 insertion for analogue **4** synthesis; in the sequence 1–10 and for Gly23 insertion for analogue **2** synthesis and for 1–10 sequence and for Ala19 and Gly23 insertions in the analogue **3** synthesis.

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) over 90 min to cleave the peptide from the solid support and to deblock the side chains. The crude peptides were purified by reversed-phase high-performance liquid chromatography (HPLC) on a semi-preparative column (Delta-pak C_{18}, Waters, 15 μm, 100 Å, 7,8×300 mm). HPLC analysis of the purified analogue **4** (conditions: column, Vydac C_{18}; eluent A, 0.05% TFA in water, eluent B, 0.05% TFA in CH_3CN; gradient, 19–27% B over 16 min; flow rate, 1 mL min^{-1}; detector, 214 nm) gave a 95% purity grade peak at 7.3 min. The analysis of purified analogues **2** and **3** was carried out on a C_{18} Luna column (Phenomenex, 5 μm, 100 Å, 4,6×250 mm) under the following conditions: eluent A, TFA in water (0.05%), eluent B, TFA in CH_3CN (0.05%); gradient, 18–26% B over 16 min; flow rate, 1 mL min^{-1}; detector, 214 nm. Elution was reached at 10.3 min for analogue **2** and at 7.6 min for analogue **3**. Integration of the chromatographic readouts gave 98% purity for analogue **2** and 99% purity for analogue **3**. The homogeneity of the products was also confirmed by capillary electrophoresis analysis on an Applied Biosystems instrument Model 270 A. The identities of the products were confirmed by MALDI spectrometric analysis (analogue **2**: theor. value: 2350 Da, exp. value 2349 Da; analogue **3**: theor. value: 2380 Da, exp. value 2383 Da; analogue **4**: theor. value: 2348 Da, exp. value 2349 Da).

Enzyme assay: Synthetic peptides were dissolved in H_2O at 1 mM concentration. The solution (15 μL) of each peptide was subjected to proteolysis in a reaction volume (85 μL) containing purified human furin (10 μL, activity 153.5 pmoles AMC μL^{-1} enzyme×hour)

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 h and for 3 h. Products were identified by HPLC purification followed by MALDI mass spectroscopic analysis. The degree of substrate conversion was determined by integration of the peak area of uncleaved substrate and comparison with peak area of a furin-untreated sample.

NMR analysis: NMR characterization was performed in TFE/H_2O 90:10 (*v/v*) at 298 K. The samples were prepared by dissolving each peptide (about 5.0 mg) in [D_3]TFE (0.75 mL, 99% isotopic purity, Aldrich) and H_2O (0.075 mL). Chemical shifts were referred to internal sodium [D_4]-2,2′,3,3′-(trimethylsilyl)propionate (TSP).

For all the analogues, NMR experiments were carried out on a Varian Unity-Inova 600 MHz spectrometer, fitted with a Sun Station Ultra5, located at the Istituto di Biostrutture e Bioimmagini C.N.R. (IBB), University of Naples "Federico II". Spectra were also acquired on an INCA (Consorzio Interuniversitario Chimica per L'Ambiente) Varian Inova 500 MHz, located at the Centro Interdipartimentale di Metodologie Chimico-Fisiche (CIMCF), University of Naples "Federico II". Two-dimensional (2D) experiments, such as total correlation spectroscopy (TOCSY),[29] nuclear Overhauser effect spectroscopy (NOESY),[30] rotating frame Overhauser effect spectroscopy (ROESY),[31] and double quantum-filtered correlated spectroscopy (DQFCOSY)[32] 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 experiments were acquired with a 70 ms mixing time, and the water resonance was suppressed by use of the WATERGATE sequence.[33] NOESY experiments were acquired with 100, 200, and 300 ms mixing times; ROESY experiments were all acquired with a 100 ms mixing time, with use of a continuous spin-lock. Off-resonance effects, due to the low power spin-lock field, were compensated by means of two 90° hard pulses before and after the spin-lock period.[34] 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. According to Wüthrich,[15] identification of amino acid spin systems was performed by comparison of TOCSY and DQFCOSY, while sequential assignment was achieved by analysis of NOESY spectra. Temperature coefficients of amide protons were measured from one-dimensional (1D) 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 in the 200 ms NOESY spectra, by use of the appropriate VARIAN software and then converted into interproton distances by use of the $1/r^6$ relationship for rigid molecules.[15] Geminal δ–$\delta'CH_2$ protons of Pro20 for analogue **3** and C-terminal NH_2 protons for analogues **2** and **4**, all with a distance of 1.78 Å, were chosen as references.

Computational methods: Torsion angle dynamics calculations were carried out with the DYANA program.[17] The library program was modified for the N- and C-terminal residues. A total of 200 three-dimensional structures were obtained from interproton distances evaluated from NOEs (raised by 20–30%) as upper limits, without use of stereospecific assignments. 100 conformers were calculated with the standard parameters of the DYANA program. All of these conformers showed good agreement with experimental constraints (lowest target function value is 0.57 $Å^2$ for analogue **2**, 0.39 $Å^2$ for analogue **3**, and 0.43 $Å^2$ for analogue **4**; the mean global backbone rmsd is 5.84 ± 1.40 Å for analogue **2**, 4.66 ± 1.01 Å for analogue **3**, and 5.04 ± 1.32 Å for analogue **4**). To improve convergence, the redundant dihedral angle constraints (REDAC)[18] strategy was also employed, and 100 more structures were calculated by carrying out five REDAC cycles for the analogues. 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 $Å^2$ in the second, and 0.4 $Å^2$ in the third. In the fourth step, the structures were calculated with the constraints previously established. In the final step, no other dihedral angle constraints were created and the structures were minimized at the highest level by use of all the experimental restraints.

The REDAC structures are indeed more compatible with the experimentally obtained data. Each of these structures shows few violations greater than 0.2 Å from the experimentally derived restraints (2 vs. 262 for analogue **2**, 4 vs. 233 for analogue **3**, 5 vs. 300 for analogue **4**). For all the analogues, the 30 DYANA structures with the lowest values of target function (average value $= 0.68 \pm 0.07$ $Å^2$ for analogue **2**, 0.27 ± 0.07 $Å^2$ for analogue **3**, and 0.77 ± 0.17 $Å^2$ for analogue **4**; mean global backbone rmsd $= 5.86 \pm 1.68$ Å for analogue **2**, 3.83 ± 1.21 Å for analogue **3**, and 4.91 ± 1.42 Å for analogue **4**) were subjected to restrained energy minimization by use of the SANDER module of the AMBER 6.0 package.[19,20] The 1991 version of the force field was used,[35] with a distance-dependent dielectric constant $\varepsilon(r)$. The charge due to the ionizable groups was reduced to 20% of its full value, in order to reduce possible artifacts due to the in vacuo simulations. A distance cutoff of 12 Å was used in the evaluation of non-bonded 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 conjugated gradient minimization, after 200 of steepest descent, for each analogue. The best 10 structures in terms of the fitting with experimentally derived restraints were selected from those with a residual restraint energy lower than: -138 kcal mol^{-1} for analogue **2**, -143 kcal mol^{-1} for analogue **3**, and -100 kcal mol^{-1} for analogue **4** to represent the peptides' solution structures. The molecular graphics program MOLMOL[36] was employed to perform the structural statistics analysis.

Modeling of the human furin–eglin c complex: A three-dimensional model of the catalytic domain of human furin was built by homology, with use of the WHAT IF software,[37] starting from the crystal structure coordinates of mouse furin (Brookhaven Data Bank code: 1P8J).[7] Note that the sequence identity between mouse and human furin is strikingly high—98%—and that none of the few observed mutations falls into functionally significant regions, thus implying virtually identical catalytic regions.

The eglin c coordinates were taken from a thermitase–eglin c complex structure (Brookhaven Data Bank code: 2TEC).[38]

The relative orientation between eglin c and human furin was settled by superimposition of the catalytic triad in the eglin-bound thermitase onto the catalytic triad of human furin. The eglin backbone was not altered, while the side chains of the eglin residues 40–47 were substituted to resemble appropriate P6–P2′ residues of gp160: that is, Val-Gln-Arg-Glu-Lys-Arg↓Ala-Val. A simple rearrangement of a few inhibitor side chains was required in order to fit the modified eglin c inhibitor into the furin binding site. The entire complex was then regularized by energy minimization performed with the INSIGHT/DISCOVER program (100 cycles of conjugate gradients energy minimization).

Keywords: conformation analysis · D-amino acids · gp160 · HIV · NMR spectroscopy

[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] 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.
[6] 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.
[7] S. Henrich, A. Cameron, G. P. Bourenkov, R. Kiersauer, R. Huber, I. Lindberg, W. Bode, M. E. Than, *Nature* **2003**, *10*, 520 – 526.
[8] 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.
[9] N. Brakch, M. Dettin, C. Scarinci, N. G. Seidah, C. Di Bello, *Biochem. Biophys. Res. Commun.* **1995**, *213*, 356 – 361.
[10] E. Decroly, S. Wouters, C. Di Bello, C. Lazure, J. M. Ruysschaert, N. G. Seidah, *J. Biol. Chem.* **1996**, *271*, 30442 – 30450.
[11] R. Oliva, L. Falcigno, G. D'Auria, M. Dettin, C. Scarinci, A. Pasquato, C. Di Bello, L. Paolillo, *ChemBioChem* **2003**, *4*, 727 – 733.
[12] I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157 – 162.
[13] E. Krause, M. Beyermann, H. Fabian, M. Dathe, S. Rothermund, M. Bienert, *Int. J. Peptide Protein Res.* **1996**, *48*, 559 – 568.
[14] R. Oliva, L. Falcigno, G. D'Auria, M. Vacatello, M. Dettin, R. Gambaretto, C. Di Bello, L. Paolillo in *Peptides 2002: Proceedings of the 27th European Peptide Symposium* (Eds.: E. Benedetti, C. Pedone), Edizioni Ziino, Naples, pp. 834 – 835.
[15] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
[16] D. S. Wishart, B. D. Sykes, F. M. Richards, *J. Mol. Biol.* **1991**, *222*, 311 – 333.
[17] P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283 – 298.
[18] P. Güntert, K. Wüthrich, *J. Biomol. NMR* **1991**, *1*, 447 – 456.
[19] AMBER 6, 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, **1999**, University of California, San Francisco.
[20] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. de Bolt, D. Ferguson, G. Seibel, P. A. Kollman, *Comp. Phys. Commun.* **1995**, *91*, 1 – 41.
[21] E. Krause, M. Bienert, P. Schmieder, H. Wenschuh, *J. Am. Chem. Soc.* **2000**, *122*, 4865 – 4870.
[22] K. Janek, S. Rothemund, K. Gast, M. Beyermann, J. Zipper, H. Fabian, M. Bienert, E. Krause, *Biochemistry* **2001**, *40*, 5457 – 5463.
[23] R. J. Siezen, J. W. Creemers, W. J. Van De Ven, *Eur. J. Biochem.* **1994**, *222*, 255 – 266.
[24] 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.
[25] W. Bode, R. Huber, *Eur. J. Biochem.* **1992**, *204*, 433 – 451.
[26] K. Nakayama, *Biochem. J.* **1997**, *327*, 625 – 635.
[27] S. J. Hubbard, R. J. Beynon, J. M. Thornton, *Protein Eng.* **1998**, *11*, 349 – 359.
[28] M. W. MacArthur, J. M. Thornton, *J. Mol. Biol.* **1991**, *218*, 397 – 412.
[29] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *65*, 355 – 360.
[30] A. Kumar, G. Wagner, R. R. Ernst, K. Wüthrich, *J. Am. Chem. Soc.* **1981**, *103*, 3654 – 3658.
[31] A. Bax, D. G. Davis, *J. Magn. Res.* **1985**, *63*, 207 – 213.
[32] U. Piantini, O. W. Sørensen, R. R. Ernst, *J. Am. Chem. Soc.* **1982**, *104*, 6800 – 6801.
[33] M. Piotto, V. Saunder, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661 – 665.
[34] C. Griesinger, E. E. Ernst, *J. Magn. Res.* **1987**, *75*, 261 – 271.
[35] S. J. Weiner, P. A. Kollmann, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **1986**, *7*, 230 – 238.
[36] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51 – 55.
[37] G. Vriend, *J. Mol. Graphics* **1990**, *8*, 52 – 56.
[38] P. Gros, C. Betzel, Z. Dauter, K. S. Wilson, W. G. Hol, *J. Mol. Biol.* **1989**, *210*, 347 – 367.

Received: June 1, 2004
Early View Article
Published online on November 3, 2004

