Retroviral Envelope Glycoprotein Processing: Structural Investigation of the Cleavage Site

Maxime Moulard,*,‡ Laurent Challoin,§ Stéphane Canarelli,|| Kamel Mabrouk,|| and Hervé Darbon§

C.I.M.L., Marseille, France, Laboratoire AFMB, IBSM, Marseille, France, and Laboratoire de Biochimie IFR Jean Roche, CNRS UMR6560, Marseille, France

Received October 27, 1997; Revised Manuscript Received January 22, 1998

ABSTRACT: Proteolytic activation of retroviral envelope glycoprotein precursors occurs at the carboxyl side of a consensus motif consisting of the amino acid sequence (Arg/Lys)-Xaa-(Arg/Lys)-Arg. Synthetic peptides spanning the processing sites of HIV-1/2 and SIV glycoprotein precursors were examined for their ability to be cleaved by the subtilisin-like endoproteases kexin and furin. To determine the potential role of secondary structure on proteolytic activation, we examined the secondary structure of synthetic peptides by circular dichroism and NMR spectroscopy. The results indicate that (i) the peptides were correctly cleaved by kexin and furin and therefore could be used as specific substrates for the purification and characterization of the lymphocyte endoprotease(s) responsible for proteolytic processing of precursors; (ii) the regions surrounding the cleavage sites could be characterized by their flexibility in aqueous solutions. However, a loop has been shown to be a determinant for the specificity of the interaction between the enzyme and its substrate as determined by molecular modeling. Furthermore, we determine and propose a possible structure of the cleavage site which fits to the active site of the modeled furin.

Viral envelope glycoproteins are synthesized as precursors which are cleaved by a cellular endoprotease in a late Golgi compartment (*1, 2*), probably by a kexin-like protease (*3*−*5*). The mature envelope glycoproteins are generated from an inactive precursor by selective and limited endoproteolysis which occurs at the carboxyl side of a consensus motif consisting of (Arg/Lys)-Xaa-(Arg/Lys)-Arg (*6*). Except for the presence of this sequence, proteolytic processing sites in precursors of viral glycoproteins exhibit only minor similarities at the amino acid sequence, suggesting that other parameters are likely to be involved in the viral glycoprotein precursor processing.

The HIV-1 glycoprotein precursor (gp160) undergoes a proteolytic cleavage at Arg508-Glu-Lys-Arg511 (site 1) (*7*). The cleavage generates the external glycoprotein gp120 associated noncovalently to the transmembrane glycoprotein gp41. The gp120 is responsible for the interaction of the viral particle with CD4 receptor−coreceptor complexes, and gp41 is required for the fusion of viral and cellular lipid membranes. Site-directed mutagenesis suggested that basic residues within the cleavage recognition sequence are crucial in determining the efficiency of the cleavage (*8, 9*). For example, replacement of Arg511 with Ser completely abolishes cleavage leading to the production of noninfectious viral particles (*10*). In addition to Arg511, Arg508, and Lys510 of site 1, there is another area of potential enzyme sensitivity corresponding to Lys500-Ala-Lys-Arg-Arg504 (site 2). Basic amino acids of site 2 are also shown to be important for proteolytic processing at site 1 (*10*). Replacement of each of the four basic residues of site 2 by neutral amino acid abolished the cleavage, suggesting that, in addition to the sequence Arg508-Glu-Lys-Arg511, a contribution of surrounding amino acids is required. The newly discovered kexin-like enzyme family is considered to be a candidate for the intracellular processing of the HIV envelope glycoprotein (*11, 12*). Among the members of this family, furin and PC7 have been shown to cleave correctly the HIV-1 gp160 (*13*) as well as the HIV-2 gp140 (*11*) in vitro using the recombinant vaccinia viruses for expression of each of the proteins. In the case of HIV-1 gp160, of the seven dibasic sites which constitute potential substrates for proprotein convertases, cleavage occurs selectively at site 1, generating biologically active glycoproteins. Taken together, these data strongly suggest that the amino acid sequence of site 1 alone is not sufficient to allow for correct processing of gp160.

To investigate the importance of secondary structure in determining the processing of such precursors, we have addressed several questions concerning the role of these structures in the catalytic process. Indeed, we have made use of a series of synthetic peptides mimicking precursor processing sites of gp160 HIV-1, gp140 HIV-2, and gp140 SIV to measure their reactivity for purified kexin from *Saccharomyces cerevisiae* and human recombinant soluble furin. The cleavability and the cleavage selectivity for each peptide is presented. Secondary structures of the synthetic peptides have also been investigated by CD and NMR. Experimental data from NMR demonstrates the peptide p511 (gp140 SIV) is likely to adopt turn structures. Therefore,

* Correspondence to: Centre d'Immunologie CNRS-INSERM de Marseille-Luminy, Parc Scientifique de Luminy, Case 906, F-13288 Marseille Cedex 9, France. Phone: (33) 491 26 94 94. Fax: (33) 491 26 94 30. E-mail: moulard@ciml.univ-mrs.fr.

‡ C.I.M.L.

§ Laboratoire AFMB.

|| Laboratoire de Biochimie.

S0006-2960(97)02662-7 CCC: $15.00
Published on Web 03/10/1998

Table 1: Amino Acid Sequence of Synthetic Peptides Used in This Study

Peptide p510 (HIV−1) Val510-Val-Gln-**Arg-Glu-Lys-Arg**-Ala-Val-Gly-Ile-Gly521
Peptides Gp4 and Gp1 (HIV−2) (Gp4) Val487-Glu-Ile-Thr-Pro-Ile-Gly-Glu-Ala-Pro-Thr-**Lys-Glu-Lys-Arg**-Tyr-Ser-Ser-Ala-His-Gly507 (Gp1) Tyr502-Ser-Ser-Ala-His-Gly507
Peptide p511 (SIV) Pro511-Thr-Asp-Val-**Lys-Arg**-Tyr-Thr-Thr-Gly-Gly-Thr-Ser-**Arg-Asn-Lys-Arg**-Gly-Val-Phe530

we attempted to demonstrate that the obtained NMR structure of p511 is relevant to a potential structure in gp160 since p511 fits into the molecular model of the active site of furin.

MATERIALS AND METHODS

Enzymes. Recombinant soluble kexin was purified from media of kexin secreting cells by concentration and elution on Q-Sepharose followed by Mono-Q chromatography (*14*). The enzyme was 90 μg/mL and 1000 units/μL corresponding to 1.1×10^7 units/mg (1 unit can cleave 1 pmole boc-Gln-Arg-Arg per minute under our standart conditions) and was from R. Fuller (University of Michigan Medical Center, Ann Arbor, MI). Recombinant soluble kexin was purified as published previously (*15*) and was from G. Thomas (Vollum Institute, Portland, OR). The concentration of the enzyme was 160 ng/μL.

Peptide Substrates. Synthetic peptides which mimic the processing cleavage sites of the envelope glycoproteins used in this study are listed in Table 1. The substrate peptide Gp4 (HIV-2 ROD) and its N-terminal peptide Gp1 were synthesized by solid-phase procedures following the general method of Merrifield (*16*). Stepwise elongation of these peptides was carried out on an automated peptide synthesizer (Applied Biosystems Inc., model 430 A) utilizing NMP-HOBt (*N*-methylpyrrolidone-2 hydroxybenzotriazole) as adapted by Applied Biosystems. After anhydrous hydrogen fluoride cleavage, the crude peptides were highly purified (>99%) by C18 reversed-phase medium-pressure liquid chromatography. Homogeneous fractions were pooled, lyophilized, and characterized by analytical HPLC and by amino acid analysis after acid hydrolysis for 24 h. The substrate peptides p510 (HIV-1) and p511 (SIV mac251) where obtained from Neosystem (Strasbourg, France).

HPLC and Amino Acid Analysis. Peptides were fractionated by HPLC using a 5-mm RP C18 endcapped column (5 × 250 mm) as described elsewhere (*17*). Amino acid analysis was performed with a Beckman model 6300 apparatus.

Protein Substrates. Recombinant gp160 (rgp160)[1] was derived from the *env* gene of HIV-1 IIIB expressed in baculovirus-infected insect cells. rgp160 including the two potential cleavage sites (L_{494}GVAPTKAKRRVVQREKR-AVGIGALF) was supplied in PBS (AGMED, Bedford, MA). Before enzymatic proteolysis, rgp160 was extensively dialyzed against a buffer composed of 20 mM Tris-HCl, pH 7.4, and 2 mM $CaCl_2$.

Enzyme Assay. Synthetic peptides (1 mg/mL) were dissolved in H_2O. Ten nanomoles of each peptide underwent proteolysis in a 0.1 mL reaction volume containing 2 μL of kexin or 5 μL of the purified furin (0.08 μg) in 20 mM Tris-HCl buffer, pH 7.4, containing 2 mM $CaCl_2$ and 0.5% (v/v) Triton X-100. The reaction mixture was incubated at room temperature for 0−24 h. Products were identified by HPLC purification followed by amino acid analysis.

Western Blotting. Transfer of the proteins from the SDS polyacrylamide gel (8%) to nitrocellulose membranes was conducted for 3 h at 200 mA. Immunodetection of cleaved fragments was performed with a 10 μg/mL dilution of a monoclonal antibody directed against gp120 (mAb 5.5) or gp41 (mAb 41.A) obtained from F. Traincard (Hybridolab, Paris, France). Membranes were then incubated with peroxidase-conjugated goat anti-mouse IgG and developed with chemiluminescence reagents from Amersham following suppliers instructions.

CD Measurements. CD spectra were acquired on a Jobin-Yvon Mark IV spectrophotometer (Longjumeau, France). The spectra were recorded at 25 °C by using 0.5 mm path length cell, with a 2 s time constant and a scan rate of 0.5 nm/s. The instrument was calibrated with (+)-10-camphosulfonic acid. The peptide concentration in the solutions, based on the absorption spectra, was from 0.5 to 1 mg/mL. The spectra were cumulated 5-fold in water or in water-TFE (2,2,2-trifluoroethanol) and automatically averaged.

NMR. Proton NMR spectra were routinely recorded at 25 °C on a Bruker AMX400 spectrometer in the phase sensitive mode by time-proportional phase incrementation of the first pulse [TPPI] (*18*). A double-quantum-filtered 2-D correlation spectrum (DQF-COSY) (*19*), a total correlation spectrum (TOCSY) was acquired with a mixing time of 100 ms using TPPI, and a phase-sensitive 2-D nuclear Overhauser effect (NOE) spectrum (NOESY) (*20*) with mixing times of 300 and 400 ms were acquired. The solvent OH resonance was suppressed by low-power irradiation during the relaxation delay and, for NOESY, during the mixing time. After the two-dimensional data matrixes were zero-filled to 2K in both dimensions (600, 400, and 800 experiments with 1K data points were recorded for, respectively, COSY, TOCSY, and NOESY spectra) and multiplied by a shifted sine-bell window in both dimensions, two-dimensional spectra were Fourier transformed and baseline corrected with FTTOOL software (Eccles, personal communication) running on an IRIS 4D-380 Silicon Graphics computer. Identification of the amino acid spin systems and sequential assignment were achieved using the well-known general strategy described by Wüthrich (*21*) with the help of EASY software (*22*) and gave rise to a nearly complete proton assignment. Identification of spin systems was obtained by analysis and comparison of 2D-DQF-COSY and TOCSY recorded in water. NOE intensities, used as input for the structure calculations were obtained from a NOESY spectrum recorded with 100 ms mixing time on the fully protonated sample, and the peaks integrated by the peak-integration routine of the EASY software (*22*), running on an IPC SUN station. The intra-residual and sequential NOE's were partitioned into three categories (<2.7, <3.5, and <5.0 Å, respectively) using a calibration curve obtained

[1] Abbreviations: DQF-COSY, double quantum filtered correlation spectroscopy; HIV, human immunodeficiency virus; NOESY, nuclear Overhauser effect spectroscopy; rgp, recombinant glycoprotein; SIV, simian immunodeficiency virus; TFE, 2,2,2-trifluoroethanol; TOCSY, total correlation spectroscopy.

after averaging the volume of all resolved peaks used for the calibration. The medium range and long-range NOEs were then converted into <4 or <5 Å distance restraints, depending on whether they occur between two backbone protons. Because of the lack of stereospecific assignments, pseudo-atom corrections were added when required (*21*). Lower distance restraints were systematically set to 1.8 Å.

Structure Calculations. Distance geometry calculations were performed with the DIANA package (*23*), without the use of stereospecific assignments. A total of 1000 DIANA structures with randomly generated starting conformations were initiated. The first round of calculation included only intra-residual and sequential restraints. The 100 best solutions as judged from the residual constraint violations were then calculated using distances extracted from medium- and long-range NOEs. Among these 100 structures, the best 25 structures were kept (as far as the NMR data are concerned) for graphic analysis using the TURBO-FRODO software (*24*), running on an IRIS-380 VGX Silicon Graphics computer.

Modeling of the Catalytic Domain of Furin. The coordinates of the crystal structure of the thermitase in complex with eglin were used (*ref 25*; Brookhaven Protein Data Bank code, 1TEC). Molecular modeling was performed using the graphic software TURBO-FRODO (*24*) running on a Silicon Graphics workstation. The backbone conformation for furin was modeled starting from appropriate segments of thermitase as described by Siezen et al. (*26*).

Modeling of Enzyme−Substrate Interaction. The obtained structure of the SIV peptide (p511) determined according to the NMR data was docked manually into the binding site of furin. The docking was then refined by energy minimization to remove bad van der Waals contacts using Powell minimizer in Xplor software (Bringer, AT, 1992, Xplor v3.1 Manual, Yale University, New Haven, CT).

Electrostatic Calculations. The dipole moments together with the electrostatic calculations and analysis of p511 and of a model of furin obtained from thermitase structure by amino acid replacement for all ionisable residues were done using the GRASP software (*27*) running on Silicon Graphics workstations. The potential maps were calculated with a simplified Poisson−Boltzmann solver (*28*) on the basis of an AMBER-derived parameter file.

RESULTS

Kexin and Furin Cleave HIV-1 Recombinant gp160. In vitro digestions of recombinant soluble gp160 by kexin or soluble furin were performed for 4 h at 37 °C. Products were separated by SDS−PAGE, transferred to a nitrocellulose membrane, and immunodetected with a monoclonal antiboby directed against the external glycoprotein gp120. Results in Figure 1 show that the precursor gp160 is completly processed to mature gp120 by recombinant soluble furin (lane 1) and partially by kexin (lane 2). The specificity of the cleavage of gp160 by kexin has been described in a previous report (*5*). These results are in agreement with previous papers (*12, 29*). However, taking into account the fact that gp160 contains seven dibasic doublets which represent potential cleavage sites of kexin and furin, the limited and restricted cleavage of gp160 by these enzymes supports the hypothesis that the proteolytic processing of gp160 at the authentic cleavage site may require specific conformational information in addition to the presence of the consensus basic motif. To assess the possible involvement of secondary structure in the selectivity and efficiency of the cleavage, we synthesized peptides covering processing sites of viral envelope precursors and analyzed the relation between their secondary structures and their cleavability by these endoproteases.

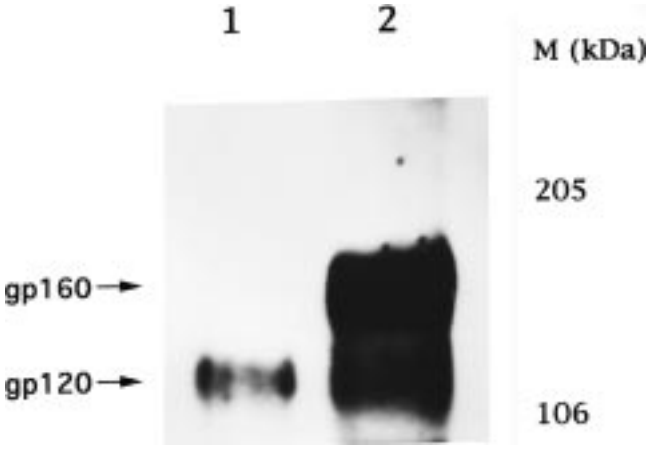

FIGURE 1: Cleavage of wild-type gp160 by kexin and furin. A total of 300 ng of rgp160 was cleaved by 2 μL of the soluble kexin or 5 μL of purified furin for 16 h at 37 °C. Products were analyzed by SDS−PAGE (8%) and Western blot using an anti-gp120 monoclonal antibody. Lanes 1 and 2 represent furin and kexin treatment of rgp160, respectively.

Table 2: Amino Acid Composition of Products Generated by Cleavage of HIV-1 Synthetic Peptide p510 with Recombinant Soluble Furin[a]

	peak 1	peak 2
Glu + Gln	3.83 (2)	
Gly		1.97 (2)
Ala		1.10 (1)
Val	3.51 (2)	0.97 (1)
Ile		0.95 (1)
Lys	1.99 (1)	
Arg	3.71 (2)	
total	(7)	(5)

[a] p510 was incubated with recombinant soluble furin for 4 h at 37 °C. Products were fractionated by HPLC and analyzed by amino acid analysis. Two peaks were generated: peak 1 (14.83 min) and peak 2 (15.60 min).

Cleavage of Synthetic Peptides by Kexin and Furin. Peptide p510 (Table 1) that includes the cleavage site of HIV-1 gp160 (site 1) was incubated with soluble furin for 4 h at 37 °C. The products were characterized by HPLC and amino acid analysis. p510 is converted by furin to the N-terminal fragment Val-Val-Gln-Arg-Glu-Lys-Arg eluting at 14.83 min (peak 1) and the C-terminal fragment Ala-Val-Gly-Ile-Gly at 15.60 min (peak 2) as shown by amino acid analysis (Table 2). Similarly, Gp4 was converted by kexin to material eluting at 12.92 min (peak 1) and 22.02 min (peak 2). Peaks 1 and 2 corresponded to the C-terminal and N-terminal fragments, respectively, since they were characterized by amino acid analysis (data not shown). Results of this analysis show clearly that furin and kexin act specifically at the potential cleavage site Arg/Lys-Glu-Lys-Arg. Similarly, p510 is cleaved by soluble kexin and Gp4 is processed correctly by the soluble human furin (data not shown).

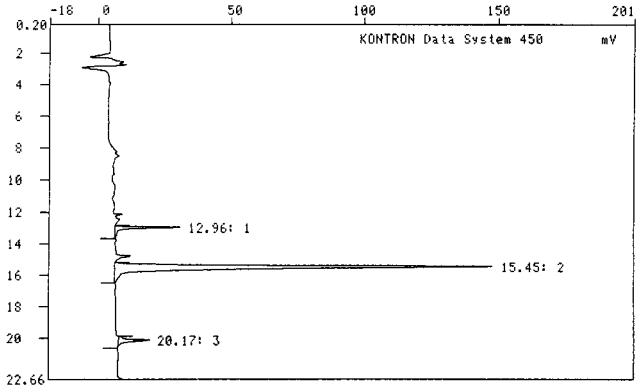

FIGURE 2: HPLC of the reaction products from kexin digestion of peptide p511. Chromatography conditions are described in Materials and Methods. p511 was incubated with kexin for 4 h. Products corresponding to peaks 1 (12.96 min), 2 (15.45 min) and 3 (20.17 min) were subjected to amino acid analysis.

Table 3: Amino Acid Composition of Products Generated by Cleavage of SIV Synthetic Peptide p511 with Recombinant Soluble Kexin

	peak 1	peak 2	peak 3
Asp + Asn	1.02 (1)	2.00 (2)	
Thr	2.64 (3)	3.68 (4)	
Ser	1.00 (1)	0.82 (1)	
Gly	2.02 (2)	2.14 (2)	0.91 (1)
Val		0.96 (1)	0.86 (1)
Tyr	0.92 (1)	1.07 (1)	
Phe			1.09 (1)
Lys	0.98 (1)	1.86 (2)	
Arg	2.00 (2)	3.00 (3)	
Pro		1.06 (1)	
total	(11)	(17)	(3)

[a] The products formed from peptide p511 incubated with kexin for 4 h at room temperature were analyzed by amino acid analysis. Peaks 1 (12.96 min), 2 (15.45 min) and 3 (20.17 min) were from Figure 3.

Interestingly, p511 (Table 1) encompasses the cleavage site of gp140 SIV and contains in addition to the natural processing site (Arg524-Asn-Lys-Arg527), a dibasic doublet Lys515-Arg516 which is also a potential substrate for kexin. While p511 is cleaved to completion by soluble furin only after Arg527, we showed that recombinant soluble kexin cleaved p511 not only after Arg527 but to a lesser extent after Arg516. In both cases, the substrate is 100% converted after Arg527. Incubation of p511 with either furin (data not shown) or kexin leads to a complete processing of the substrate (Figure 2). Three products were identified by HPLC purification and amino acid analysis (Table 3). Products with retention times of 12.96 min (peak 1), 15.45 min (peak 2), and 20.17 min (peak 3) were identified as follow: Tyr517-Thr-Thr-Gly-Gly-Thr-Ser-Arg-Asn-Lys-Arg527 (double cleavage), Pro511-Thr-Asp-Val-Lys-Arg-Tyr-Thr-Thr-Gly-Gly-Thr-Ser-Arg-Asn-Lys-Arg527 (N-terminal), and Gly528-Val-Phe530 (C-terminal). As expected, kexin is able to cleave the substrate not only after Arg527 (the expected site of hydrolysis), but also, although with a somewhat reduced efficiency, after Arg516. Efficiency of the cleavage was evaluated on the basis of molar ratios of arginine and aspartic acid residues quantified by amino acid analysis. This result shows that cleavage at the authentic site is 6−7-fold higher than cleavage after the first site (Lys515-Arg516). Furthermore, if p511 is exposed to kexin for more than 24 h, even after repeated addition of enzyme, cleavage at this site still does not reach completion. This final result suggests that the amino acid primary structure alone is not sufficient for recognition and cleavage by kexin. Our data are consistent with the previous report of Decroly et al. (*12*).

FIGURE 3: Summary of sequential and medium-range NOEs. Sequential NOEs are indicated with boxes whose sizes is proportional to the upper distance bound calculated from NOE intensity.

Conformational Analysis of the Precursor Processing Sites. In aqueous solution, the main feature of the CD spectra of p510, p511, and Gp4 is an intense negative band around 200 nm (data not shown), typical of a predominance of unordered conformer population (*30*). At 55 and 100% TFE, the CD spectra show a modified shape indicating that all of peptides are in a folded conformation and tend to adopt an α-helix conformation characterized by a typical double minimum at 208−222 nm and a positive band at 194 nm (*31*). However, the conformation of the peptides was determined under solution conditions in which TFE is considered to enhance ordered structures (*32*). On the basis of the CD spectra alone we cannot exclude that these peptides have the potential to form other folded conformations.

Although we were unable to generate exploitable results from NMR spectroscopy of p510 and Gp4, p511 was shown to be of interest with respect to the determination of its tridimensional structure. The sequence assignment of resonances was performed using conventional procedures (*21*). Cross-peaks between NH and CαH were identified by examination of the DQF-COSY spectrum (data not shown). Spin systems were identified on the basis of DQF-COSY, and the TOCSY spectrum was used to correlate these side-chain spin systems with the NH-CαH cross-peaks.

The sequential assignment was achieved by virtue of $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ connectivities (Figure 3). We have been able to sequentially connect the N-terminal sequence up to Ser523 residue and the C-terminal sequence from Gly528 to the C-terminus. The resonance assignments obtained are listed in Table 4. Due to the gap in the sequential assignment (no sequential connectivities could be identified in the stretch between Ser523 and Gly528), the assignment of resonance belonging to the Arg524 and Arg527 is ambiguous.

The analysis of the secondary structure of p511 was based on the data summarized in Figure 4. As can be seen from this figure, the pattern of d_{NN}, $d_{\alpha N}$, $d_{\alpha N}$ (i, $i + 2$), and $d_{\alpha N}$ (i, $i + 4$) strongly suggests a turn with Lys515 and Arg516 as central residues, connecting extended conformations of peptide 511−514 and 517−523. Due to the lack of constraints involving residues 524−527, it is not possible to propose a conformation for this sequence. The superposition of the N-terminal fragment, i.e., sequence 511−523, as well as the superposition of the C-terminal fragment, i.e., sequence 527−530, is fairly good and reveals, respectively,

Table 4: ^{1}H Chemical Shifts for p511 in Water Taking Water Resonance as a Reference

residue	HN	Hα	Hβ	others
Pro511		4.49	2.26	CγH_2 1.79
Thr512	9.05	4.43	4.17	CγH_3 0.79
Asp513	9.00	4.86	2.75, 2.63	
Val514	8.44	4.00	1.84	CγH_3 0.51
Lys515	8.60	4.34	1.59, 1.44	CγH_2 1.03, CδH_2 1.33 CϵH_2 2.82, NH_2 8.00
Arg516	8.39	4.21	1.37	CγH_2 1.11, CδH_2 2.97 NH_2 7.51
Tyr517	8.41	4.66	2.88, 3.00	
Thr518	8.39	4.46	4.33	CγH_3 0.83
Thr519	8.44	4.42	4.35	CγH_3 0.85
Gly520	8.64	3.90		
Gly521	8.81	3.98		
Thr522	8.51	4.45	4.37	CγH_3 0.83
Ser523	8.71	4.49	3.85, 3.92	
Arg524	8.71	4.33	1.46, 1.59	CγH_2 1.30, CδH_2 3.05 NH_2 7.57
Asn525	9.38	4.93	2.57, 2.76	
Lys526	8.61	4.26	1.48	CγH_2 1.06, CδH_3 1.06 CϵH_2 2.85, NH_2 8.61
Arg527	8.68	4.39	1.47, 1.59	CγH_2 1.32, CδH_2 3.07
Gly528	8.66	3.91		
Val529	8.09	4.08	1.68	CδH_3 0.36
Phe530	8.27	4.73	2.86, 3.07	

a turn structure centered on residues 515−516 and an extended conformation (Figure 4). The turn structure is confirmed by the existence of NOEs between Val514-CαH and the side chain of Tyr517 as well as between Arg516-NH and the side chain of Val514. However, as shown in Figure 5, where the best structure in terms of energy is depicted, the second dibasic sequence (Arg524-Asn-Lys-Arg527) also has a strong tendency to be organized as a turn.

Substrate Binding to the Engineered Catalytic Domain of Furin. We have modelized the complex between the obtained structure for p511 and furin taking the three-dimensional structure as template. The residues have been modified according to the primary sequence comparison. The backbone conformation for the furin catalytic domain was modeled starting from appropriate segments of thermitase as described by Siezen et al. (*26*). The residues of the active site: Ser109, Trp112, Val115, Gly139, and Asn170 from thermitase were exchanged, respectively, into Glu123, Asp126, Glu129, Asp152, and Asp199 in furin. The hydrophobic pocket from thermitase Ile119, Pro172, Val147, Leu143, and Ala116 were exchanged, respectively, into Leu133, Tyr201, Phe1667, Ala163, and Ala130. The coordinates of the crystal structure of thermitase complexed with eglin was used (*ref 25*; Brookhaven Protein Data Bank code, 1TEC). Molecular modeling was performed as described in the Materials and Methods and the resulting model is depicted in Figure 6. The predicted substrate-binding region in furin can be described as a crevice capable of accommodating at least seven residues of the polypeptide substrate. The substrate-binding site of furin appeared to be covered by Glu and Asp, which probably contribute to substrate specificity and could be involved in ionic interactions. The NMR structure of p511 was first docked manually in the binding site of furin, then docking was further refined by energy minimization. The resulting complex is shown on Figure 6. The synthetic peptide p511 is stabilized into the active site toward electrostatic interactions between Lys526 of p511 and Asp159, Glu129, Asp46 of furin, Arg527 of the substrate, and Asp199 of the furin. Electrostatic interactions between Arg524 of p511 and Asp121 and Glu123 of furin are present outside the active site and reinforce the interaction between furin and the peptide to be cleaved. Further hydrophobic interactions were found between Val529 of p511 and the hydrophobic pocket of furin made of Leu133, Tyr201, Phe1667, Ala163, and Ala130.

Dipole Moments. Both furin and p511 possess an anisotropic charge repartition. This anisotropy can be represented by a dipole moment. We have previously made a hypothesis attempting to describe the docking of a ligand to its receptor: according to this hypothesis, the ligand orientates in the electic field of the receptor according to its dipole orientation. This privileged orientation helps the ligand to present the correct surface for interaction with the receptor (*33*−*35*). The calculated dipole moment of furin in one hand and the p511 docked in the active site of furin in the other hand is actually collinear to the main axis of furin's active site. Thus, the peptide clearly enters the active site in a way such that the dibasic segment (Lys526-Arg527) fits correctly in the bottom of the crevice (data not shown).

DISCUSSION

A large number of mammalian prohormone convertases have been described in the last 10 years since the isolation and characterization of kexin (*36*). The biological functions of these endoproteases remain speculative even if PC1/3 and PC2 seem to be responsible for the proteolytic processing of protein precursors in the regulated secretory pathway. PACE4 (*37, 38*), PC6 (*39*), and furin are thought to be engaged in the constitutive pathway of secretion (*40*).

Proteolytic processing of viral envelope glycoprotein precursors was shown to be determinant in the life cycle of retroviruses (*7, 8, 41*). Furin and PC6 and PC7 both activate the precursors of many viral glycoproteins (*11, 42*−*45*). Cleavage of precursors by subtilisin-like endoproteases occurs at the carboxyl side of a consensus sequence consisting in (Arg/Lys)-Xaa-(Arg/Lys)-Arg. Several reports suggested that the primary sequence itself is not sufficient to determine a correct cleavage of precursors (*12, 46*). Conserved cysteine residues and glycosylation sites in HIV-1 gp160 are essential for proteolytic processing (*47, 48*) and may contribute to formation of a suitable conformation for cleavage. In this study, we have shown that cleavage of gp160 in vitro by recombinant soluble furin or kexin yielded mature gp120 and gp41 without additional cleavage in the V3-loop in our standardized conditions. A cleavage in the V3-loop is expected by increasing the furin concentration as demonstrated by Decroly et al. (*12*). Such cleavage of gp120 in the hypervariable V3-loop (Arg313-Ile-Glu-Arg316) by furin has been reported by our group (*11*) and others (*49*) when glycoprotein and endoprotease were co-expressed by recombinant vaccinia viruses. This cleavage could be expected because furin activity is expressed with the minimal sequence Arg-Xaa-Xaa-Arg (*15*). Here again, we show that selectivity of furin or kexin activation could be directed by secondary structure not excluding other factors which remain to be elucidated.

To evaluate the possible role of secondary structure in this processing step, we synthesized peptides mimicking cleavage sites of envelope glycoprotein precursors of HIV-1, HIV-2,

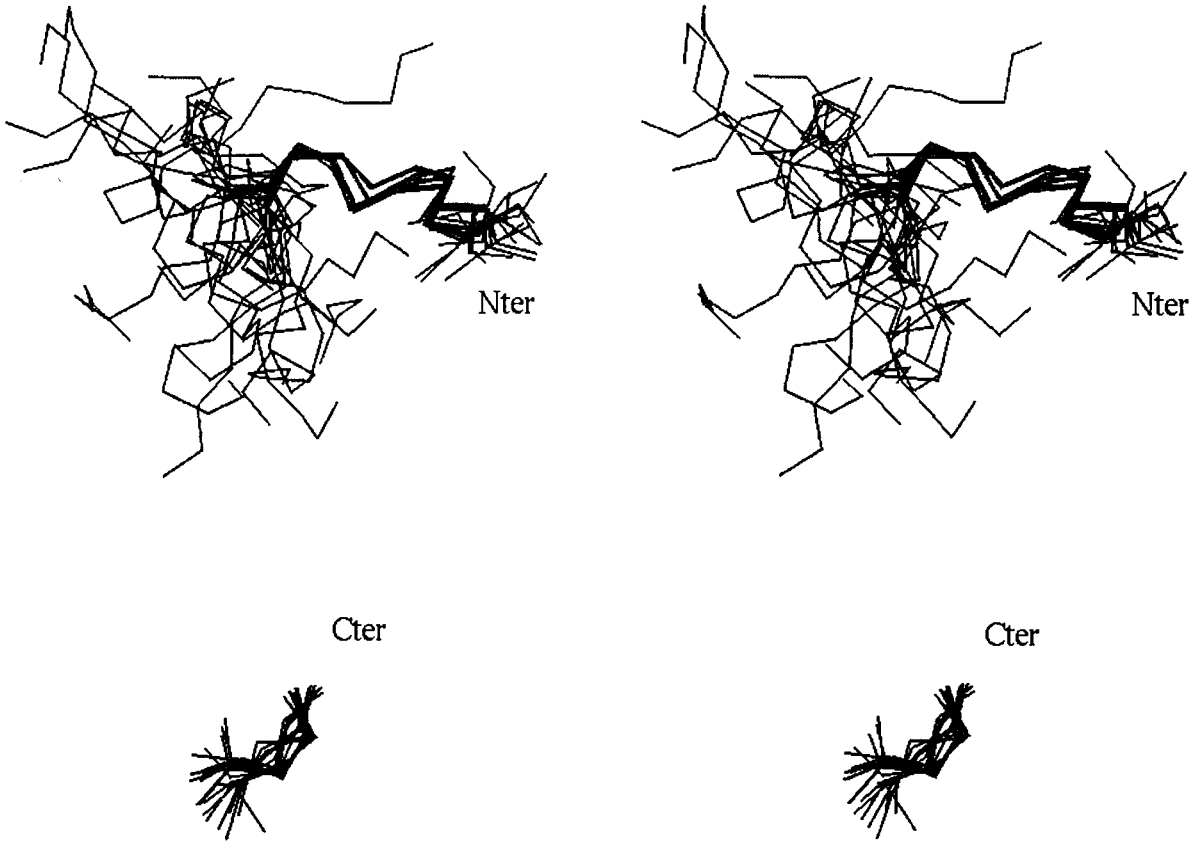

FIGURE 4: Stereoview of the best molecular p511 structures. Only the Cα atoms are displayed. The stereoview of the best molecular structures were superimposed for best fit on backbone 1−13 heavy atoms (top) and on backbone 16−20 heavy atoms (bottom). In that view, only the 16−20 sequence is shown for clarity.

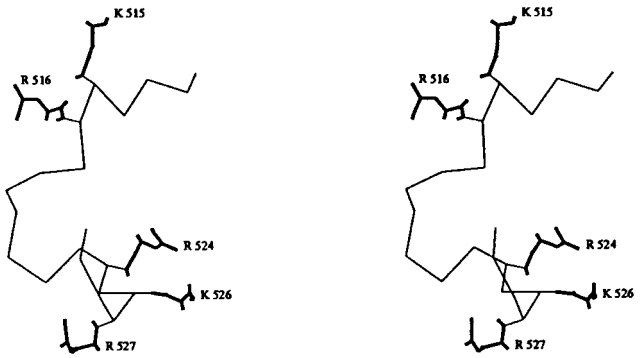

FIGURE 5: Stereoview of the lowest energy structure of p511. Only Cα atoms are displayed showing the possible conformation of the dibasic sequences.

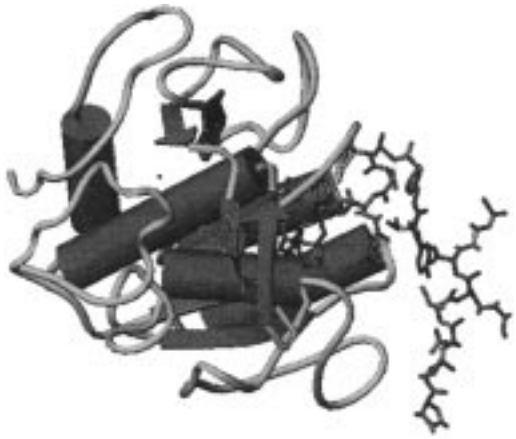

FIGURE 6: Model of the substrate-binding region of furin in complex with p511 substrate. The catalytic domain of furin was modeled starting from the crystal structure of thermitase. Furin is drawn according to the secondary structure elements. α-Helix and *β*-sheet were represented as cylinders and arrows, respectively. The docked p511 is drawn in stick.

and SIV. These peptides were subjected to proteolysis by kexin and furin and analyzed for their secondary structures by CD and NMR. Synthetic peptides (Gp4, p510, and p511) were correctly cleaved by kexin and furin. We demonstrated that kexin shows a preference for multibasic site over dibasic site of the SIV peptide (p511), suggesting either a role for the Asp at P4 as a major determinant in the specificity or, alternatively, an importance for the secondary structures which direct the enzyme specificity or both simultaneously. A similar approach was reported by Decroly et al. (*12*). The authors clearly showed that in vitro furin digestion of a model synthetic peptide spanning the two potential cleavage sites of HIV-1 gp160 generates two major products corresponding to cleavage at the physiological cleavage site of gp160. These experimental results raise the question of the role of secondary structures in the enzyme−subtrate recognition/ activity for kexin and furin. Interestingly, our results are in good agreement with a previous report (*43*) which showed that the dibasic motif Lys-Arg of a mutant F protein of human parainfluenza virus type III was very poorly cleaved by kexin in comparison with the wild-type precursor which contains the consensus sequence Arg-Thr-Lys-Arg. Using an ex vivo coexpression system in cell lines coinfected with vaccinia recombinants, we previously showed that both PC7 and furin were able to cleave, at the right position and without degradative products, HIV-1 and HIV-2 glycoprotein precursors, implying either that a redundant enzymatic system

exists in the cell, or more likely, that the information for correct glycoprotein processing is mainly encoded by the viral protein structure.

The CD spectra of synthetic peptides described in this study exhibit a random coil structure in water and have a propensity to adopt α-helical structures in the presence of TFE. However, TFE is an apolar helix-promoting solute which is used to stimulate the water-limited microenvironment surrounding a molecule. To determine the aqueous conformation of the cleavage site, NMR was used for resolution of p511. As far as the conformation of the two dibasic sequences is concerned, it is evident that the first (Lys515-Arg516) is organized in a turn conformation. The second one could not be determined because of the lack of NMR constraints which may be due to a higher mobility of the 524−527 sequence. Taken together, these data provide direct evidence for the presence of secondary structure close to, or including, the cleavage site of such precursors. In the case of the gp140 SIV, we should expect that destroying the turn around the Lys515-Arg516 pair of basic amino acids would reduce the cleavability of the precursor.

Little is known about how these enzymes select cleavage sites. In the case of prohormones, it has been postulated that amino acid flanking dibasic cleavage sites and secondary structures may direct endoprotease to the appropriate sites (*50*−*52*). Studies conducted with short peptides reproducing cleavage sites of prosomatostatin and of prooxytocin/neurophysin have shown that the sequences flanking the basic amino acid doublets might participate in the recognition of the cleavage site by providing accessible peptide segments constituted by *β*-turns (*50*) or Ω-loops (*53, 54*) and may indeed play a role in the processing (*55, 56*). From these studies, it was assumed that such structures, which are flexible and mobile, favor both the accessibility and the segmental adaptability when compared with other structures such as α-helices or *β*-sheets. Interestingly, the proteolytic cleavage of diphtheria toxin (DT) occurs on the COOH-terminal side of the sequence Arg190-Val-Arg-Arg, which have been determined to be a suitable subtrate for furin (57). The tertiary structure of DT was determined at 2.0 Å (58), and the cleavage site was reported to be included in a highly flexible loop. These data support our finding. Dibasic sites enclosed in α-helix or *β*-sheet structures appeared to be poorer subtrates for cellular enzymes. However, a number of cleavage sites analyzed in prohormones were predicted to be associated with α-helices, and in other cases no assignment of structure was possible.

Understanding viral glycoprotein precursor processing will require crystallographic structure or NMR information, but to date, no such data are available for these regions. Because it appears that a variety of structural motifs are capable of directing appropriate processing, it would be interesting to determine secondary structures of recognition motifs that might correlate with different classes of precursors processing enzymes (*26*).

ACKNOWLEDGMENT

We are grateful to G. Thomas (Vollum Institute, Oregon Health Sciences University, OR) for the gift of the recombinant soluble furin and R. Fuller (University of Michigan Medical Center, Ann Arbor, MI) for provinding us with recombinant purified kexin. We are indebted to Q. Sattentau for its suggestions. E. Bahraoui is thanked for its initial interest. We wish to thank E. Blanc for dipole moment calculations and S. Meunier for recording NMR spectra at the "Centre Régional de R.M.N., Marseille", headed by Professor A. Thevand.

REFERENCES

1. Stein, B. S., and Engleman, E. G. (1990) *J. Biol. Chem. 265*, 2640−2649.
2. Willey, R. L., Bonifacio, J. S., Potts, B. J., Martin, M. A., and Klausner, R. D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9580−9584.
3. Hallenberger, S., Bosh, V., Angliker, H., Shaw, E., Klenk, H. D., and Garten, W. (1992) *Nature 360*, 358−361.
4. Decroly, E., Vandenbranden, M., Ruysschaert, J.-M., Cogniaux, J., Jacob, S. J., Howard, S. C., Marshall, G., Kompelli, A., Basak, A., Jean, F., Lazure, C., Benjannet, S., Chrétien, M., Day, R., and Seidah, N. G. (1994) *J. Biol. Chem. 269*, 12240−12247.
5. Moulard, M., Achstetter, T., Montagnier, L., Kiény, M. P., and Bahraoui, E. (1994) *Eur. J. Biochem. 225*, 565−572.
6. Barr, P. J. (1991) *Cell 66*, 1−3.
7. McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R., and Weissman, I. L. (1988) *Cell 53*, 55−67.
8. Freed, E. O., Myers, D. J., and Risser, R. (1989) *J. Virol. 63*, 4670−4675.
9. Guo, H.-G., Veronese, F. D., Tschachler, E., Pal, R., Kalyanaraman, V. S., Gallo, R. C., and Reitz, M. S., Jr. (1990) *Virology 174*, 217−224.
10. Bosh, V., and Pawlita, M. (1990) *J. Virol. 64*, 2337−2344.
11. Hallenberger, S., Moulard, M., Sordel M, Klenk, H. D., and Garten, W. (1997) *J. Virol. 71*, 1036−1045.
12. Decroly, E., Wouters, S., Di Bello, C., Lazure, C., Ruysschaert, J.-M., and Seidah, N. G. (1996) *J. Biol. Chem. 271*, 30442−130450.
13. Decroly, E., Benjannet, S., Savaria, D., and Seidah, N. G. (1997) *FEBS Lett. 405*, 68−72.
14. Fuller, R. S., Brake, A., and Thorner, J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1434−1438.
15. Molloy, S. S., Bresnahan, P., Leppla, S. H., Klimpel, K. R., and Thomas, G. (1992) *J. Biol. Chem. 267*, 16396−16402.
16. Merrifield, R. B. (1986) *Science 232*, 341−347.
17. Moulard, M., Achstetter, T., Ikehara, Y., and Bahraoui, E. (1994) *Biochimie 76*, 251−256.
18. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
19. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800−6801.
20. Kumar, A., Ernst, R. R., and Wüthrich, K. (1981) *Biochem. Biophys. Res. Commun. 95*, 1−6.
21. Wüthrich, K. (1986) in *NMR of Protein and Nucleic Acid*, John Wiley & Sons, New York.
22. Eccles, C., Güntert, P., Billeter, M., and Wüthrich, K. (1991) *J. Biomolecular NMR 1*, 111−130.
23. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.
24. Roussel, A., and Cambillau, C. (1989) in *Silicon Graphics Geometry Partner Directory* (Silicon Graphics, Ed.) pp 77−78, Silicon Graphics, Mountain View, CA.
25. Gros, P., Betzel, C., Dauter, Z., Wilson, K. S., and Hol, W. G. (1989) *Mol. Biol. 210*, 347−367.
26. Siezen, R. J., Creemers, J. W. M., and Van de Ven, W. J. M (1994) *Eur. J. Biochem. 222*, 225−266.
27. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins 11*, 281−296.
28. Nicholls, A., and Honig, B (1991) *J. Comput. Chem. 12*, 435−445.
29. Franzusoff, A., Volpe, A. M., Josse, D., Pichuantes, S., and Wolf J. R. (1995) *J. Biol. Chem. 270*, 31541−31549.
30. Woody R. (1985) in *Peptides, Polypeptides and Proteins* (Blout, E. R., Bovey, F. A., Lotan, N., & Goodman, M., Eds.) pp 338−360, Wiley, New York.

31. Bandekar, J., Evans, D. J., Krimm, S., Leach, J., Lee, S., Mcquie, J. R., Minasian, E., Memethy, G., Pottle, M. S., Sheraga, H. A., Stimson, E. R., and Woody, R. W. (1982) *Int. J. Pep. Protein Res. 19*, 187−205.
32. Urry, D. W. (1972) *Biochim. Biophys. Acta 265*, 115−168.
33. Fremont, V., Blanc, E., Crest, M., Eauclaire, M. F., Gola, M., Darbon, H., and Van Rietschoten, J. (1997) *J. Pept. Sci.* (in press).
34. Blanc, E., Lecomte, C., Van Rietschoten, J., Sabatier, J. M., and Darbon, H. (1997) *Proteins 29*, 359−369.
35. Blanc, E., Sabatier, J. M., Karrhat, R., El Ayeb, M., Van Rietschoten, J., and Darbon, H. (1997) *Proteins 29*, 321−333.
36. Van de Ven, W. J. M., Roebroek, A. J. M., and Van Duijnhoven, H. L. P. (1993) *Crit. Rev. Oncogen. 4*, 115−136.
37. Kiefer, M. C., Tucker, J. E., Joh, R., Landsberg, K. E., Saltman, D., and Barr, P. J. (1991) *DNA Cell Biol. 10*, 757−769.
38. Rehemtulla, A., Barr, P. J., Rhodes, C. J., and Kaufman, R. J. (1993) *Biochemistry 32*, 11586−11590.
39. Lusson, J., Vieau, D., Hamelin, J., Day, R., Barale, J. C., Chrétien, M., and Seidah, N. G. (1993) *Proc. Natl. Acad. Sci U.S.A. 90*, 6691−6695.
40. Steiner, D. F., Smeekens, S. P., Ohagi, S., and Chan, J. S. (1992) *J. Biol. Chem. 267*, 23435−23438.
41. Moulard, M., Montagnier, L., and Bahraoui, E. (1994c) *FEBS Lett. 338*, 281−284.
42. Miranda, L., Wolf J., Pichuantes, S., Duke, R., and Franzusoff, A. (1996) *Proc. Natl. Acad. Sci U.S.A. 93*, 7695−7700.
43. Ortmann, D., Ohuchi, M., Angliker, H., Shaw, E., Garten, W., and Klenk, H. D. (1994) *J. Virol. 68*, 2772−2776.
44. Horimoto, T., Nakayama, K., Smeekens, S. P., and Kawaoka, Y. (1994) *J. Virol. 68*, 6074−6078.
45. Gotoh, B., Ohnishi, Y., Inocencio, N., Esaki, E., Nakayama, K., Barr, P. J., Thomas, G., and Nagai, Y. (1992) *J. Virol. 66*, 6391−6397.
46. Brakch, N., Dettin, M, Scarinci, C, Seidah, N. G., and Di Bello, C. (1995) *Biochem. Biophys. Res. Commun. 213*, 356−361.
47. Dedera, D., Gu, R., and Ratner, L. (1992) *J. Virol. 66*, 1207−1209.
48. Bolmstedt, A., Hemming, A., Flodby, P., Berntsson, P., Travis, B., Lin, J. P. C., Ledbetter, J., Tsu, T., Wigzell, H., Hu, S. L., and Olofsson, S. (1991) *J. Gen. Virol. 72*, 1269−1277.
49. Morikawa, Y., Barsov, E., & Jones, I. (1993) *J. Virol. 67*, 3601−3604.
50. Rholam, M., Nicolas, P., and Cohen, P. (1986) *FEBS Lett. 207*, 1−6.
51. Rholam, M., Brakch, N., Germain, D., Thomas, D. Y., Fahy, C., Boussetta, H., Boileau, G., and Cohen, P. (1995) *Eur. J. Biochem. 227*, 707−714.
52. Brennan, S. O., and Nakayama, K. (1994) *FEBS Lett. 347*, 80−84.
53. Bek, E., and Berry, R. (1990) *Biochemistry 29*, 178−183.
54. Rayne, R. C., and O'Shea, M. (1993) *Eur. J. Biochem. 217*, 905−911.
55. Paolillo, L., Simonetti, M., Brakch, N., D'Auria, G., Saviano, M., Dettin, M., Rholam, M., Scatturin, A., Di Bello, C., and Cohen, P. (1992) *EMBO J. 11*, 2399−2405.
56. Brakch, N., Rholam, M., Bousetta, H., and Cohen, P. (1993) *Biochemistry 32*, 4925−4930.
57. Tsuneoka, M., Nakayama, K., Hatsuzawa, K., Komada, M., Kitamura, N., and Nekada, E. (1993) *J. Biol. Chem. 268*, 26461−26465.
58. Choe, S., Bennett, M. J., Fujii, G., Curmi, P. M., Kantardjieff, K. A., Collier, R. J., and Eisenberg, D. (1992) *Nature 357*, 216−222.

BI972662F