Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 7310–7315, June 1998
Biochemistry

A model for the structure of the P domains in the subtilisin-like prohormone convertases

GREGORY M. LIPKIND, AN ZHOU, AND DONALD F. STEINER*

Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637

Contributed by Donald F. Steiner, April 20, 1998

ABSTRACT **The proprotein convertases are a family of at least seven calcium-dependent endoproteases that process a wide variety of precursor proteins in the secretory pathway. All members of this family possess an N-terminal proregion, a subtilisin-like catalytic module, and an additional downstream well-conserved region of ≈150 amino acid residues, the P domain, which is not found in any other subtilase. The pro and catalytic domains cannot be expressed in the absence of the P domains; their thermodynamic instability may be attributable to the presence of large numbers of negatively charged Glu and Asp side chains in the substrate binding region for recognition of multibasic residue cleavage sites. Based on secondary structure predictions, we here propose that the P domains consist of 8-stranded β-barrels with well-organized inner hydrophobic cores, and therefore are independently folded components of the proprotein convertases. We hypothesize further that the P domains are integrated through strong hydrophobic interactions with the catalytic domains, conferring structural stability and regulating the properties and activity of the convertases. A molecular model of these interdomain interactions is proposed in this report.**

Most prohormones and neuroendocrine peptide precursors, including proopiomelanocortin, proinsulin, and proglucagon, are processed by specific cellular enzymes—prohormone convertases—through the selective endoproteolytic cleavage at dibasic sites, usually Lys-Arg ↓ and Arg-Arg ↓ (1–4). Over the last decade a family of precursor processing endoproteases has been discovered, which includes the yeast prohormone processing enzyme kex2 or kexin (5, 6), the mammalian endoproteinases furin (subtilisin-like proprotein convertase 1; SPC1) (7), PACE4 (SPC4) (8), and the prohormone convertases PC2 (SPC2) (9, 10), PC1/PC3 (SPC3) (11, 12), PC4 (SPC5) (13), PC5/PC6 (SPC6) (14), and PC7/PC8/LPC (SPC7) (15–17). The core specificity of all these enzymes—paired basic amino acids—is further enhanced by the presence of additional basic residues at the P4 and further upstream sites (1, 3).

When kexin was cloned and sequenced (5), it was found to contain a catalytic domain homologous to that of the bacterial subtilisins. Amino acid sequence alignments of all the mammalian convertases and kexin have shown a very high degree of similarity in their catalytic domains (18, 19). The most significant difference between the subtilisins and the catalytic domains of the SPCs is the large increase in the number of negatively charged residues (Glu and Asp) in the substrate binding region relative to the subtilisins, a feature that probably contributes to the great selectivity of this family of enzymes for substrates containing multiple basic residues (1, 20).

Although these proteases all possess catalytic domains similar to the bacterial subtilisins, no tertiary structural data from x-ray analysis are available at this time. However, molecular modeling of the catalytic domains of furin (19) and SPC3/SPC1 (21), based on the known spatial structures of subtilisin and thermitase, has indicated that they are closely related in their three-dimensional structures to the subtilisins. The SPCs and kexin share common structural features, including an N-terminal signal peptide followed by a propeptide of 80–90 residues terminating with the canonical cleavage motif for (auto) activation, R-X-K/R-R, a catalytic domain of ≈285 residues (having ≈25–30% homology to the subtilisins), and a well-conserved P, or Homo B, domain of ≈150 residues (1, 3). The presence of the P domains is a distinctive characteristic of all known members of the proprotein convertase family. In subtilisin and other subtilisin-like proteases, such additional downstream sequences are not found. The P domain appears to be necessary to both fold and maintain the subtilisin-like active catalytic module and to regulate its specialized features of calcium and more acidic pH dependence (22). However, the structural basis for this regulatory role (i.e., the folding motif of the P domain and its spatial integration with the catalytic domain) is unclear.

Despite the homology of the SPC catalytic domains to the bacterial subtilisins, we have not found any corresponding homologous protein for the P regions in the database. In this report we propose a possible protein structure (or structural motif) for the P domain based on predictions of essential elements of its secondary structure, and from these and related considerations we have gained insights into the probable structural and functional integration of the P domains within the spatial organization of the convertases.

The P Domain Has a β-Structural Motif

An amino acid sequence alignment of the P domains (18) of representative members of the SPC family—mammalian SPC2–SPC7, furin, and kexin—is given in Fig. 1. The residues are numbered according to those of mouse SPC3. The region of greatest similarity among the P domains is restricted to residues 454–594. On this basis, the P domain of SPC3 can be considered to contain 141 residues. Mutagenesis studies also indicate that C-terminal deletions proximal to Thr-594 in SPC3 give rise to nonfunctional enzymes (22).

As a first step in considering the problem of the spatial organization of the P domains, it is important to distinguish possible elements of secondary structure, α-helices and β-strands, and also irregular sequences or loops that might connect regular folded regions. Assessment of the structural protein group (23) restricts the choice of meaningful template structures for modeling of the conserved core of the P domains. On the whole, this strategy represents a knowledge-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SPC, subtilisin-like proprotein convertase.

*To whom reprint requests should be addressed at: Howard Hughes Medical Institute, University of Chicago, 5841 South Maryland Avenue, MC 1028, Room N-216, Chicago, IL 60637. e-mail: dfsteine@midway.uchicago.edu.

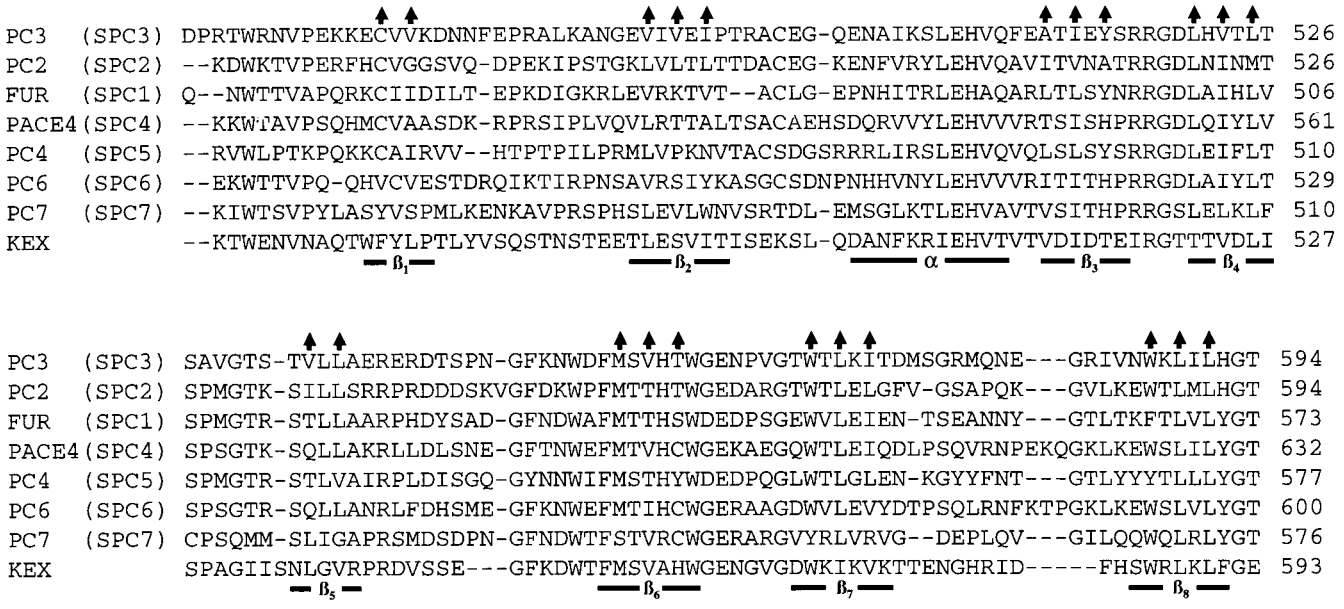

FIG. 1. Amino acid sequence alignment of the P domains of SPC2–SPC7, furin, and kexin (15). The predicted eight β-strands and a region of amphipathic α-helix are underlined. Vertical arrows indicate those β-strand side chains that form the inner hydrophobic core of the P domain.

based approach to protein model building (24–26). Here we have followed the predictive schemes as proposed by Chou and Fasman (27, 28) and Lim (29, 30). The first scheme is based on statistical analyses of protein primary sequences and classifies the amino acids as favoring, breaking, or being indifferent to each type of conformation. The second scheme takes into account the main structural rules of folded proteins such as compactness of form and the presence of a tightly packed hydrophobic core and is based on stereochemical considerations of interactions of hydrophobic and hydrophilic residues in α-helices and β-strands.

In assessing the secondary structure of the P domains, it proved useful to distinguish probable nonregular sequences in the protein. Such loop regions usually are located on the external surfaces of globular proteins. Five amino acid residues are present most frequently in such nonregular loops and reverse turns, Gly and Pro, and residues with short hydrophilic side chains, Ser, Asn, and Asp (31, 32). Furthermore, as a rule, insertions and deletions between homologous sequences are often located within external nonregular loops (31, 33). Therefore, regions in the amino acid sequences of the P domains with a high proportion (more than half) of such "structure disrupting" residues, and which also tend to have deletions or insertions, are likely to be unfolded. These regions were identified by the presence of sequences meeting these criteria in at least one of all homologous sequences and are as follows: residues 454–457, 470–483, 491–493, 517–520, 527–530, 540–550, 560–565, and 572–583. Thus, about 40% of the residues in the P domains participate in the formation of nonregular surface loops, a proportion that is characteristic of many globular proteins (34). We therefore infer that the P domains of the SPCs most likely also have this type of structure.

Putative elements of secondary structure were first calculated for each considered amino acid sequence of SPC2–SPC7, furin, and kexin on the basis of the Chou and Fasman method (27, 28) by using the most recent values for the conformational preferences P_α, P_β, and P_t (35). Each prediction was then checked in terms of its conformity with the rules proposed by Lim (29, 30), which, in some instances, significantly restricted the boundaries of possible α-helices and β-strands. For example, the sequence AERERD (residues 537–542 in SPC3) is predicted by Chou–Fasman rules to have a high α-helical potential, whereas according to Lim (30), this sequence is considered antihelical as it does not form hydrophobic pairs and triplets; accordingly it was excluded from the predicted secondary structure.

Sequences of the P domains having a high probability of forming elements of secondary structure and which are also compatible with the primary structures of SPC2–SPC7, furin, and kexin are underlined in Fig. 1. Except for one long segment, residues 498–508, all other identified segments are rather short (4–7 amino acid residues), consistent with their predicted preference for forming β-strands (the length of β-strands in globular proteins usually does not exceed 3–8 residues; ref. 34). This conclusion is also supported by the significant content in these regions of amino acid residues with bulky hydrophobic side chains (Val, Ile, Leu, Tyr, Trp) that tend to occur more frequently in β-sheets (32, 35, 36).

The most characteristic peculiarity of the β-strands of globular proteins is the alternation of hydrophobic residues with polar or charged hydrophilic residues. Such an alternation is a basic rule for prediction of β-strands in proteins (29). Because of the extended conformation of the peptide chain in β-strands, alternating hydrophobic and hydrophilic residues are segregated into separate hydrophobic and hydrophilic surfaces on either side of the β-strand backbone. Thus β-strands can simultaneously participate in the formation of the hydrophobic core and the external surface of globular proteins. Such alternating sequences are unlikely to be α-helical though, because hydrophobic residues at positions i, $i + 2$, and $i + 4$ would be located on the opposite sides of an α-helix, and such α-helices would have neither hydrophobic nor hydrophilic surfaces.

All eight of the β-strands we have identified in the P domains of the SPCs are in excellent agreement with the alternation rule, with several notable exceptions where hydrophobic residues are found on both sides, suggesting a possible site of interdomain interaction (see below). Importantly, the alternating hydrophobic residues, which are marked by arrows in Fig. 1, are present in the same relative positions in all eight homologous sequences. A good example of such a sequence is residues 566–571 (TWTLKI in SPC3 and DWKIKV in kexin in Fig. 1).

The formation of an α-helix is predicted in the region 498–508 (Fig. 1). This sequence has a high content of charged residues and a high α-helical potential according to the

Chou–Fasman predictive scheme (27, 28). The pattern of hydrophobics at every third residue is also strongly suggestive of α-helical secondary structure for this sequence. Moreover, the N-terminal ends of these helices contain predominantly Glu, Asp, and Pro residues, which tend to occur at the N-terminal ends of α-helices (32) and in the case of kexin also "a capping box" motif: Ser/Thr-Xaa-Xaa-Glu/Asp (37). Most likely, this α-helix is on the surface. However, its hydrophobic side screens the inner hydrophobic core on the bottom of the β-barrels of the P domains.

The presence of eight short β-strands having mainly hydrophobic side chains along one side suggested to us that the P domains may belong to the β-structure-based group of globular proteins (23, 38). The nonpolar sides of these strands could thus form a hydrophobic core, whereas the polar sides could be oriented toward the surface where they would participate in the formation of the hydrophilic outer shell of the domains.

The predicted participation of eight short β-strands in the structure indicates a high probability for the formation of a β-barrel-like structure. An antiparallel β-barrel consists of a closed β-structure, which is made up of just eight strands (38) that have the characteristic alternating hydrophobic and polar side chains. Because both of these conditions are met by the predicted β-strands of the P domains, it is likely that they form a β-barrel structure with an inner hydrophobic core and outer hydrophilic surface and with the polar residues of the loops, located at the top and bottom of the β-barrels, also participating in the formation of its hydrophilic surface.

The number of observed topologies of β-barrel structures in real proteins is small (38). The three most frequently occurring groups belong to the up-and-down barrels, the Greek key motif barrels, and the jelly roll barrels. However, taking into account requirements for docking of the P domains with the catalytic domains of SPCs (see the next section), we propose that the β-barrel of the P domains belongs to the type having a Greek key motif. Accordingly, it consists of two β-sheets, formed by strands 2, 1, 4, 7 and 3, 8, 5, 6, as denoted in the amino acid sequences of the P domains (Fig. 1), and arranged as shown in Fig. 2 to form a closed β-barrel structure. For this construction we used the x-ray structure of another small protein, γ-crystallin (39) as a structural template for the relative arrangement of the β-strands, because γ-crystallin forms a β-sandwich with the Greek key motif. In terms of unified topologies, β-sandwiches also belong to the family of β-barrel proteins (40).

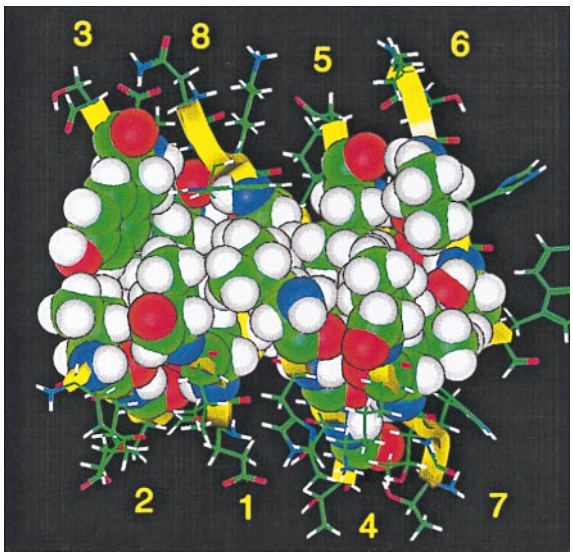

FIG. 2. Proposed β-barrel (β-sandwich) of the P domain of SPC3 with its central hydrophobic core shown by space-filling models. In this representation, the view is from above with the two β-sheets, each made up of four β-strands, arranged in the Greek key motif (numbers correspond to β-strand segments shown in Fig. 1).

Purposeful packing and optimization of the potential energy of interactions between the side chains of β-strands in such an initial arrangement [by using INSIGHT and DISCOVER graphical environment (Biosym Technologies, San Diego)] led to the structure, shown in Fig. 2, where the hydrophobic side chains of the eight β-strands of the P domains form a β-barrel with a densely packed hydrophobic core. The form of this hydrophobic core is structurally closely similar to that of α-crystallin and is protected by external hydrophilic loops. The α-helix (residues 498–508) also can be included in one of these loops, and thus one function of this single α-helix may be to protect the hydrophobic core of the central β-barrel. In the construction of the P domain, the side chain of Cys at the N-terminal end of first β-strand lies near the Cys residue of the loop between β-strands 2 and 3, enabling the formation of a disulfide bond.

Finally, we can conclude on the basis of this model that the P domain possesses, in itself, all the characteristic features of the structural organization of typical globular proteins. Therefore, the P domains form independently folded compact domains in the subtilisin-like prohormone convertases that are juxtaposed with the domains of the catalytic units.

The Catalytic and P Domains Interact Through Hydrophobic Patches

The catalytic domains form a core structure in the SPCs that is structurally related to the subtilisins (41). The structurally conserved regions (42) of the catalytic domains of the SPCs consist of about 200 amino acid residues and include the secondary structural elements (α-helices and β-strands) and also the active site groove of the subtilisins (43). The main differences between the primary structures of subtilisins and catalytic domains of the SPCs consist, as a rule, of numerous additional insertions in the latter, which are located in the external connecting loops between helices and strands. Accordingly, these regions are considered as variable regions (42). Therefore, in considering possible interactions between the catalytic and P domains in the SPCs, we will pay attention only to the structurally conserved framework of the subtilisins, taking into account corresponding amino acid substitutions in the SPCs and the conserved β-barrel core of the P domains.

Following our conclusion that P domains form discrete structural units, we propose that both catalytic and P domains coexist as independently folded domains in the structures of SPCs. We further posit that nonbonded interactions between the catalytic and P domains provide additional structural stabilization of the catalytic domains, which alone appear to be thermodynamically unstable. The reason for this instability may be that the catalytic domains of the prohormone convertases have a large increase in the number of Glu and Asp residues, whereas subtilisin contains practically equal numbers of negatively and positively charged side chains. The SPCs contain 22 conserved Asp and Glu residues, whereas conserved basic residues number only five (18, 41). Moreover, nearly all of the excess of negatively charged side chains are located asymmetrically in the modeled spatial structures of the catalytic domains, around and inside the active sites, resulting in the electrostatic destabilization of the catalytic domains within the SPCs.

That the P domains are required for the stabilization of SPC structure was shown in truncation experiments. Deletion of the P domains in SPC2 (44), SPC3 (22), and kexin (45) led to unstable constructs. Most likely these truncated forms were misfolded because they were not able to traverse the secretory pathway.

We can postulate that there must be a region on the surface of the catalytic domains of the SPCs that contains predomi-

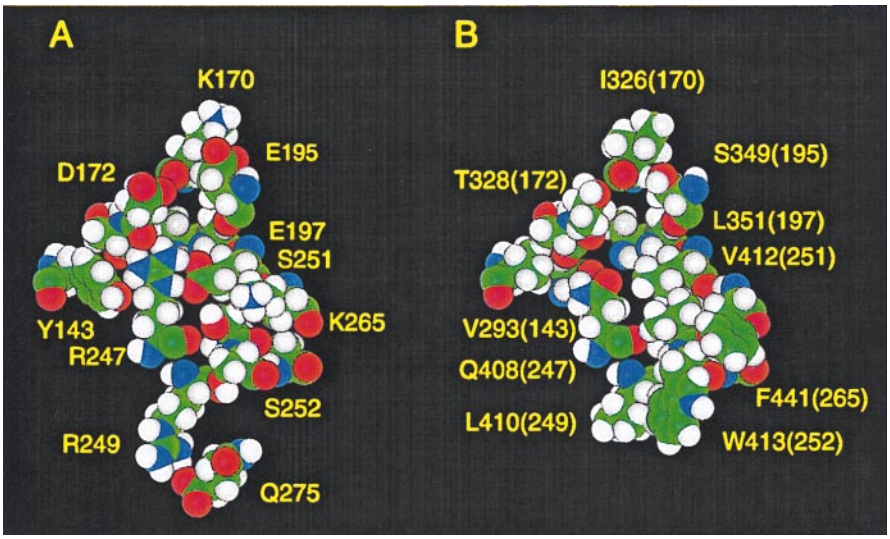

FIG. 3. Topography of a portion of the hydrophilic surface of subtilisin Carlsberg (*A*), which is transformed into a hydrophobic patch on the surface of the catalytic domain of the SPCs, as exemplified here by SPC3 (*B*). Note that Q275 in *A* is the C-terminal residue of subtilisin and its carboxylate forms a salt bridge with R249, whereas in the SPCs the peptide chain is not interrupted.

nantly hydrophobic residues and constitutes a site for interaction with the P domains, whereas the corresponding residues on the surface of the globular subtilisin molecule are hydrophilic. Indeed, analysis of protein–protein interfaces distinguishes the significant contribution of steric contacts between hydrophobic amino acid residues (46–48). Moreover, surface hydrophobicity can be used to identify regions of a protein surface most likely to interact with a binding ligand (47). Such hydrophobic interactions play an especially important role in the stabilization of long-lived protein homocomplexes (49). In this sense, the situation we are considering in the SPCs is closest to this latter case.

Comparative analysis of the surface residues of the structurally conserved framework of subtilisin Carlsberg (43) and SPC3 revealed a region of subtilisin where charged amino acid residues are substituted by hydrophobic residues in SPC3 and related convertases (Fig. 3). The acidic residues Asp-172, Glu-195, and Glu-197 of subtilisin (Fig. 3) are substituted by Val, Ser, and Leu residues, respectively in the case of PC3, whereas the basic residues Lys-170, Arg-247, Arg-249, and Lys-265 are changed to Ile, Leu, Val, and Phe, respectively. Therefore, all former salt bridges are substituted by contacts of the hydrophobic residues in PC3 (Fig. 3), which at the same time are very highly conserved in all the SPCs (41). For example, Glu-197 and Arg-249 of subtilisin are substituted by Leu residues, not only in PC3, but also in PC2, furin, kexin, and other members of the SPC family (including some conserved substitutions by Ile). From the foregoing consideration of substitutions of charged amino acid residues in the subtilisins by bulky hydrophobic residues in the SPCs (by using a published alignment of the amino acid sequences of the subtilases; ref. 41) we have identified a single, quite large hydrophobic patch unique to the surfaces of the SPCs. Importantly, the surface segment bearing this hydrophobic patch in its center is not located in the region that has an excess of conserved Asp and Glu residues in the SPCs. These are practically all concentrated within the span of residues 122–326 in SPC3 (or 1–170 in the subtilisin numbering system) and surround the catalytic groove on the opposite surface of the catalytic domain. Normally, such extensive hydrophobic surface patches are rare in globular proteins. It therefore seems reasonable to propose that this conserved surface hydrophobic region on the SPC catalytic domains interacts with the P domains, and this interaction in turn stabilizes the catalytic domains.

Similar asymmetry of surface hydrophobicity was also found in our proposed β-barrel model of the P domains. The external surface of this β-barrel (Fig. 2), formed by β-sheet 2, 1, 4, 7, in accordance with the numbering of its component β-strands, contains a large number of negatively and positively charged residues and, therefore, this β-sheet also probably lies on the

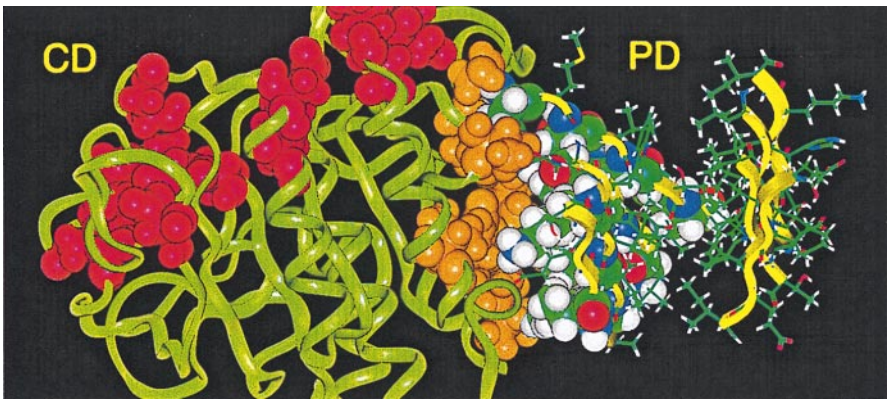

FIG. 4. Conserved framework of the catalytic (CD) and P domains of SPC3 showing proposed dense packing between the hydrophobic surfaces of the catalytic and P domains. The P domain (PD) is shown here in a side view with β-strands oriented vertically. External hydrophobic residues of β-strands 8, 5, and 6 form a hydrophobic patch on the P domain that interacts with the catalytic domain. Surface hydrophobic residues (see Fig. 3*B*) and conserved Glu and Asp residues of the catalytic groove of the SPCs are shown by orange and red space-filling images, respectively.

external surface of the SPCs. On the other hand, the external residues of β-strands 5 and 6 of β-sheet 3, 8, 5, 6 are mainly hydrophobic (Fig. 1) and therefore could participate in interdomain interactions with the hydrophobic patch on the surface of the catalytic domains. Analogously, in α-crystallin nonpolar side chains of the 5th, 6th, and 8th β-strands of the second β-barrel in the C-terminal end domain participate in van der Waals contacts with the N-end domain (39), thus providing a model for the relative arrangement and interface docking of the catalytic and P domains. Then Ile-590 and the hydrophobic part of the side chain of Lys-588 of β-strand 8 of the P domain were packed with the hydrophobic patch of the catalytic domain, similarly to the interface arrangement of Val-170 and Arg-168 of β-strand 8 of the second domain of α-crystallin (39). The possible spatial stacking between the hydrophobic surfaces of the catalytic domain and the P domain is illustrated in Fig. 4. Putative interdomain hydrophobic contacts are shown in this figure by space-filling models. Thus, molecular modeling does not exclude the formation of a region of dense side-chain packing between these two domains in the prohormone convertases.

In the proposed arrangement, the surface hydrophobic contacts of the catalytic domains screen residues of β-strand 8 of the P domains, especially Ile-590 and His-592, which neighbor the 5th strand in the 3, 8, 5, 6 β-sheet. Accordingly, the residues near the C-terminal ends of the P domains (Gly-593 and Thr-594 in SPC3; Fig. 1) could also participate in immediate van der Waals contacts inside this interdomain hydrophobic core, and this interaction could determine the important functional role of these C-terminal residues in conferring structural stability on the SPCs (22, 44).

This proposed docking scheme for the P domains would not be expected to interfere with the formation of inhibitor protein complexes. Molecular modeling reveals that interactions of furin with such bulky peptide inhibitors as turkey ovomucoid (50) and α_1-antitrypsin Portland (51) would not extend over the surface region which is favorable for interaction with the P domains (data not shown). This finding is also the case for the association of SPC2 with the neuroendocrine protein 7B2. Site-directed mutagenesis of Tyr-194 in proPC2 to Asp (as in SPC3) blocked both its binding and activation (44). However, this locus for binding of 7B2 is located strictly on the side of the catalytic domain that lies opposite to the P domain contact site.

In conclusion, we have herein proposed a model for the spatial organization of the prohormone convertases in which these consist of two independent structural domains (catalytic and P). The catalytic domains have subtilisin-like structures, whereas the P domains form structures of the antiparallel β-barrel or β-sandwich type. Both the catalytic and P domains possess extensive hydrophobic surfaces, and interdomain hydrophobic interactions most likely provide additional thermodynamic stability to maintain the SPCs in their catalytically active conformation.

We thank Rosie Ricks for her expert assistance in the preparation of this manuscript and Sean Martin for the alignment of amino acid sequences of SPC2–SPC7, furin, and kexin. We also thank Tobin Sosnick and Carl Correll for critically reading the manuscript. This study was supported by National Institutes of Health Grant DK-13914 and in part by the Howard Hughes Medical Institute.

1. Steiner, D. F., Smeekens, S. P., Ohagi, S. & Chan, S. J. (1992) *J. Biol. Chem.* **267,** 23435–23438.
2. Steiner, D. F., Bell, G. I., Tager, H. S. & Rubenstein, A. H. (1995) in *Endocrinology*, ed. DeGroot, L. (Saunders, Philadelphia), pp. 1296–1328.
3. Rouille, Y., Duguay, S., Lund, K., Furuta, M., Gong, Q., Lipkind, G., Oliva, A., Jr, Chan, S. J. & Steiner, D. F. (1995) *Front. Neuroendocrinol.* **16,** 322–361.
4. Steiner, D. F., Rouille, Y., Gong, Q., Martin, S. Carroll, R. & Chan, S. J. (1996) *Diabetes Metab.* **22,** 94–104.
5. Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S. & Matsuo, M. (1988) *Biochem. Biophys. Res. Commun.* **156,** 246–254.
6. Fuller, R. S., Brake, A. & Thorner, J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1434–1438.
7. Van de Ven, W. J. M., Voorberg, J., Fontijn, R., Pannekoek, H., Van den Ouweland, A. M. W., Van Duijnhoven, H. L. P., Roebroek, A. J. M. & Siezen, R. J. (1990) *Mol. Biol. Rep.* **14,** 265–275.
8. Kiefer, M. C., Tucker, J. E., Joh, R., Landsberg, K. E., Saltman, D. & Barr, P. J. (1991) *DNA Cell Biol.* **10,** 757–769.
9. Smeekens, S. P. & Steiner, D. F. (1990) *J. Biol. Chem.* **265,** 2997–3000.
10. Seidah, N. G., Mattei, M. G., Gaspar, L., Benjannet, S., Mbikay, M. & Chretien, M. (1991) *Genomics* **11,** 103–107.
11. Smeekens, S. P., Avruch, A. S., LaMendola, J., Chan, S. J. & Steiner, D. F. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 340–344.
12. Seidah, N. G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M. G., Lazure, C., Mbikay, M. & Chretien, M. (1991) *Mol. Endocrinol.* **5,** 111–122.
13. Nakayama, K., Kim, W.-S., Torii, S., Hosaka, M., Nakagawa, T., Ikemizu, J., Baba, T. & Murakami, K. (1992) *J. Biol. Chem.* **267,** 5897–5900.
14. Nakagawa, T., Hosaka, M., Torii, S., Watanabe, T., Murakami, K. & Nakayama, K. (1993) *J. Biochem. (Tokyo)* **113,** 132–135.
15. Seidah, N. G., Hamelin, J., Mamarbachi, M., Dong, W., Tardos, H., Mbikay, M., Chretien, M. & Day, R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3388–3393.
16. Bruzzaniti, A., Goodge, K., Jay, P., Taviaux, S. A., Lam, M. H. D., Berta, P., Martin, T. J., Moseley, J. M. & Gillespie, M. T. (1996) *Biochem. J.* **314,** 727–731.
17. Meerabux, J., Yaspo, M-L., Roebroek, A. J., Van den Ven, M. J. M., Lister, T. A. & Young, B. D. (1996) *Cancer Res.* **56,** 448–454.
18. Smeekens, S. P., Chan, S. J. & Steiner, D. F. (1992) in *Progress in Brain Research*, eds. Joosse, J., Buijs, R. & Tilders, F. (Elsevier, Amsterdam), Vol. 92, pp. 235–246.
19. Siezen, R. J., de Vos, W. M., Leunissen, J. A. M. & Dijkstra, B. W. (1991) *Protein Eng.* **4,** 719–737.
20. Creemers, J. W. M., Siezen, R. J ., Roebroek, A. J. M., Ayoubi, T. A. Y., Huylebroeck, D. & Van de Ven, W. J. M. (1993) *J. Biol. Chem.* **268,** 21826–21834.
21. Lipkind, G. M., Gong, Q. & Steiner, D. F. (1995) *J. Biol. Chem.* **270,** 13277–13284.
22. Zhou, A., Martin, S., Lipkind, G., LaMendola, J. & Steiner, D. F. (1998) *J. Biol. Chem.* **273,** 11107–11114.
23. Levitt, M. & Chothia, C. (1976) *Nature (London)* **261,** 552–558.
24. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. (1987) *Nature (London)* **326,** 347–352.
25. Blundell, T. L. & Johnson, M. S. (1993) *Protein Sci.* **2,** 877–883.
26. Bajorath, J., Stenkamp, R. & Aruffo, A. (1993) *Protein Sci.* **2,** 1798–1810.
27. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
28. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
29. Lim, V. I. (1974) *J. Mol. Biol.* **88,** 857–872.
30. Lim, V. I. (1974) *J. Mol. Biol.* **88,** 873–894.
31. Benner, S. A., Cohen, M. A. & Gerloff, D. (1993) *J. Mol. Biol.* **229,** 295–305.
32. Creighton, T. E. (1993) *Proteins: Structural and Molecular Properties* (Freeman, New York).
33. Zvelebil, M. J., Barton, G. J. & Taylor, W. R. (1987) *J. Mol. Biol.* **195,** 957–961.
34. Shulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York).
35. Williams, R. W., Chang, A., Juretic, D. & Loughran, S. (1987) *Biochim. Biophys. Acta* **916,** 200–204.
36. Rooman, M. J. & Wodak, S. J. (1991) *Proteins Struct. Funct. Genet.* **9,** 69–78.
37. Harper, E. T. & Rose, G. D. (1993) *Biochemistry* **32,** 7605–7609.
38. Branden, C. & Tooze, J. (1991) *Introduction to Protein Structure* (Garland, New York).
39. Wistow, G., Turnell, B., Summers, L., Slingsby, C., Moss, D., Miller, L., Lindley, P. & Blundell, T. L. (1983) *J. Mol. Biol.* **170,** 175–202.
40. Richardson, J. (1981) *Adv. Protein Chem.* **34,** 167–339.
41. Siezen, R. J. & Leunissen, J. A. M. (1997) *Protein Sci.* **6,** 501–523.
42. Greer, J. (1990) *Proteins Struct. Funct. Genet.* **7,** 317–334.

43. McPhalen, C. A. & James, M. N. G. (1988) *Biochemistry* **27,** 6582–6598.
44. Zhu, X., Muller, L., Mains, R. E. & Lindberg, I. (1998) *J. Biol. Chem.* **273,** 1158–1164.
45. Gluschankof, P. & Fuller, R. S. (1994) *EMBO J.* **13,** 2280–2288.
46. Vakser, I. & Aflalo, C. (1994) *Proteins Struct. Funct. Genet.* **20,** 320–329.
47. Young, L., Jernigan, R. L. & Covell, D. G. (1994) *Protein Sci.* **3,** 717–729.
48. Tsai, C.-J., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1997) *Protein Sci.* **6,** 53–64.
49. Jones, S. & Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13–20.
50. Lu, W., Zhang, W., Molloy, S. S., Thomas, G., Ryan, K., Chiang, Y., Anderson, S. & Laskowski Jr, M. (1993) *J. Biol. Chem.* **268,** 14583–14585.
51. Jean, F., Stella, K., Thomas, L., Liu, G., Xiang, Y., Reason, A. J. & Thomas, G. (1998) *Proc. Natl. Acad. Sci. USA,* in press.