

The self-inhibited structure of full-length PCSK9 at 1.9 Å reveals structural homology with resistin within the C-terminal domain

Eric N. Hampton, Mark W. Knuth, Jun Li, Jennifer L. Harris, Scott A. Lesley, and Glen Spraggon*

Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121

Edited by James A. Wells, University of California, San Francisco, CA, and approved July 25, 2007 (received for review April 12, 2007)

Mutations in proprotein convertase subtilisin/kexin type 9 (PCSK9) are strongly associated with levels of low-density lipoprotein cholesterol in the blood plasma and, thereby, occurrence or resistance to atherosclerosis and coronary heart disease. Despite this importance, relatively little is known about the biology of PCSK9. Here, the crystal structure of a full-length construct of PCSK9 solved to 1.9-Å resolution is presented. The structure contains a fully folded C-terminal cysteine-rich domain (CRD), showing a distinct structural similarity to the resistin homotrimer, a small cytokine associated with obesity and diabetes. This structural relationship between the CRD of PCSK9 and the resistin family is not observed in primary sequence comparisons and strongly suggests a distant evolutionary link between the two molecules. This three-dimensional homology provides insight into the function of PCSK9 at the molecular level and will help to dissect the link between PCSK9 and CHD.

hypercholesterolemia | low-density lipoprotein receptor | proprotein convertase | x-ray crystallography | adipocytokine

PCSK9 (also known as neural apoptosis-regulated convertase, NARC-1) is a 692-residue extracellular protein expressed primarily in the kidneys, liver and intestines (1) representing the 9th member of the secretory subtilase family. Various genetic observations subsequently mapped PCSK9 as the third gene (along with LDLR and APOB) to cause autosomal dominant hypercholesterolemia (ADH). These studies suggested that gain-of-function mutations increase plasma levels of LDL-c (2–6), whereas nonsense or missense (loss-of-function) mutations, which interfere with folding or secretion of PCSK9, lead to a reduction of plasma levels of LDL-c and an 88% decrease in the risk of coronary heart disease (CHD) (5). In mice, adenoviral overexpression of PCSK9 results in increased plasma LDL-c level in normal mice but not in LDLR-deficient mice (7). Deletion of PCSK9 causes an increase in level of LDLR protein but not mRNA (8). These findings lead to a hypothesis that PCSK9 exerts its role in cholesterol metabolism through post-translational down-regulation of LDLR, the receptor responsible for clearing LDL-c from plasma. In support of this hypothesis, a recent mouse parabiosis study revealed a nearly complete loss of liver LDLR proteins in wild-type recipients of human PCSK9 parabiosed from a transgenic littermate (9).

Like other secretory subtilisin-like serine proteases, PCSK9 is synthesized in the endoplasmic reticulum (ER) as a precursor that undergoes autocatalytic processing, intracellular trafficking, and, finally, release into the extracellular medium with the capability of decreasing the cellular LDLR level and LDL-c uptake (10). Despite the rapid progress in understanding the genetics and biology of PCSK9, little is known of its structure, substrate specificity, or mode of action of PCSK9 in the regulation of LDLR. The full-length, mature sequence consists of three domains, [the first 30 residues of the N terminus contain signal sequence (Fig. 1A)]; the first two correspond to an inhibitory prodomain (amino acids 31–152) and a serine protease domain (amino acids 153–452) of the proteinase K sub-

family member of the subtilisin-like serine proteases (1, 11), respectively. The overall fold of the catalytic domain has been well characterized because of the early x-ray crystallography work done on bacterial subtilisins (11), followed by structures of the catalytic domain in complex with its prodomain (12). Recently, representative structures from eukaryotes have also been solved. These structures, Furin and Kex2 (13, 14) reveal the structure of the ≈ 150 -residue, C-terminal, P (processing)-domain believed to be essential for folding and regulation of the activated protease (15). The equivalent domain in PCSK9 is longer, 239 residues in length (amino acids 453–692), rich in cysteine and histidine residues, and structurally and functionally uncharacterized but thought to play an analogous role to the P-domains.

Here, the high-resolution full-length structure of PCSK9, at 1.9-Å resolution is presented. The structure not only allows mapping of the gain- and loss-of-function mutations to three-dimensional coordinates but also structurally characterizes the C-terminal domain a unique domain of the proprotein convertases showing structural homology to another plasma protein resistin.

Results

Overall Structure. The core of the structure conforms to a standard subtilisin-like serine protease domain, an α/β protein consisting of a seven-stranded parallel β -strand sandwiches between sets of helices, inhibited by its prodomain (12). A C-terminal, accessory cysteine-rich domain (CRD) is situated next to the catalytic domain forming a cloverleaf of subdomains displaying a clear pseudothreefold axis (Fig. 1A). The CRD is loosely attached to the catalytic domain (Fig. 1A) occluding a surface area of 1,450 Å² at the interface between the protease domain. Despite the relatively weak binding between the domains and the presence of a flexible linker between them, it seems likely that this is the physiological conformation of the molecule as indicated by the rigid body refinement of a structure at lower resolution (≈ 4.0 Å) crystallized at a different pH 6.5 and space group, as opposed to pH 10.5 for the high-resolution structure, which exhibited the same conformation (data not shown). Both chains were well defined in the electron density

Author contributions: E.N.H., M.W.K., J.L., J.L.H., S.A.L., and G.S. designed research; E.N.H., M.W.K., and G.S. performed research; E.N.H., J.L., and G.S. analyzed data; and J.L., J.L.H., S.A.L., and G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: ADH, autosomal dominant hypercholesterolemia; CHD, coronary heart disease; LDL, low-density lipoprotein; NARC-1, neural apoptosis-regulated convertase 1; PCSK9, prohormone convertase subtilisin kexin-9; CRD, cysteine-rich domain.

Data deposition: The structure reported in this paper has been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2Q7W).

*To whom correspondence should be addressed. E-mail: gspraggo@gnf.org.

This article contains supporting information online at www.pnas.org/cgi/content/full/0703402104/DC1.

© 2007 by The National Academy of Sciences of the USA

stability within the resistin fold, because equivalent positions in the PCSK9 resistin-like CRD are characterized by regions of high atomic displacement parameters or disorder. Given the shapes of the two molecules, the most obvious candidate region for interaction is the head domain containing the nine loops connecting the β -sheets surrounding the deep cleft at the center of the pseudothreefold axis (Fig. 3D). The topological arrangement of β -strands between the two molecules is identical, but the loops connecting the β -strands to the head of the trimer in resistin of the molecule are conformationally divergent, suggesting an alternative binding partner (Fig. 3A and C).

Naturally Occurring Gain- and Loss-of-Function Mutations. A number of genetic studies of naturally occurring variants of PCSK9 have been discovered (2–6, 20, 28). If the general model is that PCSK9 is responsible for plasma LDLR levels and, thus, plasma LDL-c levels, then loss-of-function mutations can be linked with non-sense mutations or those mutation related to processing/folding/active-site abnormalities within PCSK9, which thus increase levels of LDLR (Fig. 2). One example in the resistin-like CRD is the nonsense mutation C679X that occurs in the penultimate disulfide bond, suggesting a disruption in the folding pattern of the CRD and thus incorrect processing or folding of the molecule. Another loss-of-function mutation, R46L (20), has been reported to produce a 15% reduction in plasma LDL-c levels for heterozygous individuals. This mutation is located within the N terminus of the prodomain, a region that shows no ordered electron density (amino acids 31–60). The flexibility makes it unclear whether the loss of function is a result of folding or secretion problems. It is possible that the structure becomes ordered in response to a binding partner or a different physiochemical environment. In such circumstances, the loss of function mutation could influence this binding or ordering of the sequence.

Gain-of-function mutations are likely more diverse in phenotypic action, the general hypothesis being that the mutations increase the binding affinity for substrates, thus providing a more active molecule. Gain-of-function mutations are distributed around the surface (Fig. 2): one region occurs at the negatively charged lip of the active site where D374Y presumably influences binding affinity for substrate; one mutation in the prodomain S127R may increase binding or enhance the susceptibility of the domain to cleavage by proteases such as furin; and regions at both the front and back ends of the first subdomain of the resistin-like CRD. Those in the resistin-like CRD are interesting and validate the importance of the resistin-like CRD in function. The two mutations R469W and F515L occur in the putative binding region of the resistin-like CRD at the head of the domain within the first monomer of the heterotrimer (SD1 in Figs. 2 and 3).

Discussion

Despite the obvious biomedical interest, the function of PCSK9 has been enigmatic. The structure of PCSK9 provides valuable insight into its specificity and function and the structural relationship to other cellular proteins involved in similar interactions. The existence of a unique C-terminal domain in PCSK9 relative to other proprotein convertases shows novelty, whereas its homology to resistin, not easily discernable at the primary sequence level, exhibits conservation. This observation indicates that the functions of the proprotein convertases, like their functionally homologous counterparts, the trypsin like serine proteases, arose by exons shuffling and gene duplication [see Pathy (29) for review]. The molecular function of resistin has also been mysterious; the gene has been associated with metabolic diseases [for review, see Stepan (27)], but, as yet, no molecular receptor has been identified. With the relationship between the two molecules, it is tempting to speculate that the two share or compete for a common receptor, LDLR. However,

it is more likely, given the low sequence identity, structural divergence of putative binding sites (Fig. 3B), and shape changes due to the nature of the hetero- versus homotrimer structure, that the two molecules have evolved to recognize different binding partners. It is tempting to suggest that the CRD provides a means to localize the protease to the substrate molecule [LDLR (8, 30, 31)] or a receptor that localizes function, allowing the PCSK9 protease domain to facilitate degradation or removal of LDLR. An alternative hypothesis, and one consistent with current knowledge, would be the absence of proteolytic activity altogether, whereby binding of PCSK9 to LDLR (or other receptor) provides a signal to lower LDLR levels. A recent publication by Nassoury *et al.* (32) has provided evidence that the CRD binds to the cell surface via LDLR. A nonproteolytic mechanism has been shown to exist in proteases of the trypsin family such as hepatocyte growth factor (HGF) (33), whereby the protease has lost catalytic activity but functions as a growth factor receptor. An analogous situation may occur in PCSK9, with the exception that positive selection for the maintenance of the active site is apparent, because the protein must autocleave itself to enable the correct folding and secretion of the protein. This connection between structure and disease may be important to dissecting the complex pathway of PCSK9 in CHD.

While the present manuscript is in review, two structures of PCSK9 were published (34, 35). The orthorhombic crystal form detailed in these two articles is identical to the ones presented in this article, with the exception that growth conditions for the two crystals differed widely, with pHs of 8.4 and 7.6 for the Cunningham (34) and Piper (35) structures, respectively. The Piper article (35) also details a cubic crystal form grown at pH 4.6. This, combined with the fact that our structure crystallized at pH 10.5, strongly suggests that the orientation of the domains relative to one another is independent of pH and suggests that recent studies of differential binding affinities of LDLR at different pHs (36) are due to conformational changes in LDLR or other physiochemical changes. In addition, the structures do not contain a Ca^{+2} at the conserved calcium-binding site. Because calcium is not a major determinant of PCSK9 activity (19), this suggests that PCSK9 can accommodate a Ca^{+2} but has little effect on the structure. Both articles mistakenly identify the CRD as a new fold rather than a heterotrimeric arrangement of a resistin-like fold.

Materials and Methods

Expression and Purification. The full-length DNA sequence encoding PCSK9 was amplified from a full-length construct in the baculovirus expression vector pFastBac1 by PCR. The construct included a fused hexahistidine C-terminal affinity tag, and the honey bee melittin secretion signal substituted for the native signal peptide. The forward primer (5'-CCCAAGCTTGCCGC-CACCATGAAATTCTTAGTC AAC-3') incorporates a 5' HindIII restriction site and Kozak sequence. The reverse primer (5'-GC TCTAGATCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGAG-3') includes a 3' XbaI restriction site and introduces a stop codon. The PCR product was digested with HindIII and XbaI, and cloned into these respective sites of the pRS5a mammalian expression vector. PCSK9 was expressed in the Freestyle 293 expression system (Invitrogen, Carlsbad, CA). One liter of cells was transfected according to the manufacturer's protocol. Forty-eight hours after transfection, cultures were harvested by centrifugation at $400 \times g$. The supernatant was loaded on to a 4-ml NiNTA column (Qiagen, Valencia, CA) equilibrated with 50 mM Tris·HCl (pH 7.4), 300 mM NaCl, 1 mM CaCl_2 , and 2 mM 2-mercaptoethanol and washed with 20 ml of equilibration buffer plus 20 mM imidazole. Protein was eluted with 50 mM Tris·HCl (pH 7.4), 250 mM imidazole, 300 mM NaCl, 1 mM CaCl, and 2 mM 2-mercaptoethanol and analyzed by SDS/PAGE. The eluted PCSK9 was dialyzed against 50 mM

Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM CaCl₂, and 2 mM 2-mercaptoethanol before concentration to 10 mg/ml.

Crystallization, Data Collection, and Structure Solution. Crystals were grown by sitting-drop vapor diffusion and formed within 3 weeks in crystallization conditions containing 20% PEG 6000, 0.2 M sodium chloride at 20°C in 0.1 M CAPS buffer at pH 10.5. Data were collected at the Advanced Light Source beamline 5.0.3 to a maximum resolution of 1.9 Å. Data were processed with the HKL2000 package (37), and initial phases were derived by using Phaser (38) and two ensembles of a prodomain structure (PDB ID codes 1SCJ and 1SPB) and one containing five representatives from the

subtilisin-like serine protease family (PDB ID codes 1CNM, 1SH7, 2B6N, 1P8J, and 1R64). Iterative building and refinement was carried out with ARP/WARP, Coot and, Refmac5 (39–42). The *R*-factor and free *R*-factor for all data up to 1.9 Å converged at 19.0% and 24%, respectively (SI Table 2).

We thank A. Kreuzsch, C. Lee, M. Didonato, and P. Chamberlain for help with data collection; J. Trauger, E. Nigoghossian, C. Tumanut, K.-F. Suen, and J.-A. Gavign for help with protein expression; D. Yowe, B. Kreider, J. Levell, S. Izumo, C. Hernandez, and G. Liao for useful discussion; and P. Schultz for continued support. Data were collected at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, CA) on beamline 5.0.3; we thank all of the staff for their help.

1. Seidah NG, Benjannet S, Wickham L, Marcinkiewicz J, Jasmin SB, Stifani S, Basak A, Prat A, Chretien M (2003) *Proc Natl Acad Sci USA* 100:928–933.
2. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, Cruaud C, Benjannet S, Wickham L, Erlich D, et al. (2003) *Nat Genet* 34:154–156.
3. Leren TP (2004) *Clin Genet* 65:419–422.
4. Allard D, Amsellem S, Abifadel M, Trillard M, Devillers M, Luc G, Krempf M, Reznik Y, Girardet JP, Fredenrich A, et al. (2005) *Hum Mutat* 26:497.
5. Cohen JC, Boerwinkle E, Mosley TH, Jr, Hobbs HH (2006) *N Engl J Med* 354, 1264–1272.
6. Berge KE, Ose L, Leren TP (2006) *Arterioscler Thromb Vasc Biol* 26:1094–1100.
7. Maxwell KN, Breslow JL (2004) *Proc Natl Acad Sci USA* 101:7100–7105.
8. Rashid S, Curtis DE, Garuti R, Anderson NN, Bashmakov Y, Ho YK, Hammer RE, Moon YA, Horton JD (2005) *Proc Natl Acad Sci USA* 102:5374–5379.
9. Lagace TA, Curtis DE, Garuti R, McNutt MC, Park SW, Prather HB, Anderson NN, Ho YK, Hammer RE, Horton JD (2006) *J Clin Invest* 116:2995–3005.
10. Cameron J, Holla OL, Ranheim T, Kulseth MA, Berge KE, Leren TP (2006) *Hum Mol Genet* 15:1551–1558.
11. Wright CS, Alden RA, Kraut J (1969) *Nature* 221:235–242.
12. Jain SC, Shinde U, Li Y, Inouye M, Berman HM (1998) *J Mol Biol* 284:137–144.
13. Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber R, Lindberg I, Bode W, Than ME (2003) *Nat Struct Mol Biol* 10:520–526.
14. Holyoak T, Wilson MA, Fenn TD, Kettner CA, Petsko GA, Fuller RS, Ringe D (2003) *Biochemistry* 42:6709–6718.
15. Zhou A, Martin S, Lipkind G, LaMendola J, Steiner DF (1998) *J Biol Chem* 273:11107–11114.
16. Seidah NG, Chretien M (1999) *Brain Res* 848:45–62.
17. Schechter I, Berger A (1967) *Biochem Biophys Res Commun* 27:157–162.
18. Naureckiene S, Maa L, Sreekumarb K, Purandarea U, Loa CF, Huangc Y, Chiangd LW, Grenierd JM, Ozenbergera BA, Jacobsena JS, et al. (2003) *Arch Biochem Biophys* 420:55–67.
19. Benjannet S, Rhainds D, Essalmani R, Mayne J, Wickham L, Jin W, Asselin MC, Hamelin J, Varret M, Allard D, et al. (2004) *J Biol Chem* 279:48865–48875.
20. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, Hobbs HH (2006) *Am J Hum Genet* 78:410–422.
21. Banerjee RR, Rangwala SM, Shapiro JS, Rich AS, Rhoades B, Qi Y, Wang J, Rajala MW, Poci A, Scherer PE, et al. (2004) *Science* 303:1195–1198.
22. Patel SD, Rajala MW, Rossetti L, Scherer PE, Shapiro L (2004) *Science* 304:1154–1158.
23. Rangwala SM, Rich AS, Rhoades B, Shapiro JS, Obici S, Rossetti L, Lazar MA (2004) *Diabetes* 53:1937–1941.
24. Holcomb IN, Kabakoff RC, Chan B, Baker TW, Gurney A, Henzel W, Nelson C, Lowman HB, Wright BD, Skelton NJ, et al. (2000) *EMBO J* 19:4046–4055.
25. Ohue M, Kato K, Nomura A, Toyosawa K, Furutani Y, Kimura S, Kadowaki T, Kim KH (2001) *Diabetes* 50:113–122.
26. Steppan CM, Bailey ST, Bhat S, Brown EJ, Banerjee RR, Wright CM, Patel HR, Ahima RS, Lazar MA (2001) *Nature* 409:307–312.
27. Steppan CM, Lazar MA (2004) *J Int Med* 255:439–447.
28. Jirholt P, Adiels M, Borén J (2004) *Arterioscler Thromb Vasc Biol* 24:1334–1336.
29. Patthy L (2003) *Genetica* 118:217–231.
30. Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin NV, Horton JD, Cohen JC, Hobbs HH (2006) *Am J Hum Genet* 79:514–523.
31. Park SW, Moon YA, Horton JD (2004) *J Biol Chem* 279:50630–50638.
32. Nassoury N, Blasiole DA, Olter AT, Benjannet S, Hamelin J, Poupon V, McPherson PS, Attie AD, Prat A, Seidah NG (2007) *Traffic* 8:718–732.
33. Donate LE, Gherardi E, Srinivasan N, Sowdhamini R, Aparicio S, Blundell TL (1994) *Protein Sci* 3:2378–2394.
34. Cunningham D, Danley DE, Geoghegan KF, Griffor MC, Hawkins JL, Subashi TA, Varghese AH, Ammirati MJ, Culp JS, Hoth LR, et al. (2007) *Nat Struct Mol Biol* 14:413–419.
35. Piper DE, Jackson S, Liu Q, Romanow WG, Shetterly S, Thibault ST, Shan B, Walker NP (2007) *Structure (London)* 15:545–552.
36. Fisher TS, Lo Surdo P, Pandit S, Mattu M, Santoro JC, Wisniewski D, Cummings RT, Calzetta A, Cubbon RM, Fischer PA, et al. (2007) *J Biol Chem* 282:20502–20512.
37. Otwinowski Z, Minor W (1997) *Methods Enzymol* 276:307–326.
38. Storoni LC, McCoy AJ, Read RJ (2004) *Acta Crystallogr D* 60:432–438.
39. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS (2001) *Acta Crystallogr D* 57:1445–1450.
40. Emsley P, Cowtan K (2004) *Acta Crystallogr D* 60:2126–2132.
41. Murshudov GN, Vagin AA, Dodson EJ (1997) *Acta Crystallogr D* 53:240–255.
42. Collaborative Computing Project Number4 (1994) *Acta Crystallogr D* 50:760–763.
43. Delano W (2002) *PYMOL* (DeLano Scientific, San Carlos, CA).
44. Russell RB, Barton GJ (1992) *Proteins Struct Funct Genet* 14:309–323.
45. Barton GJ (1993) *Protein Eng* 6:37–40.