addition of DNA did not change the structure of ΔN4-NColE7-C* in solution. This is in line with the crystal structures of NColE7 and the nuclease domain of ColE9 bound to dsDNA [16, 29, 31, 32, 57]. There was, however, a clear difference between the calculated and experimental spectra upon addition of DNA to ΔN25-NColE7, suggesting that the dsDNA binding induces a slight conformational change in the more flexible mutant, probably by the stabilization of the wild-type structure.

Gel mobility shift assays

To compare the DNA-binding ability of the truncated mutants, a gel mobility shift experiment was conducted, in which increasing amounts of proteins were added to a 0.874 μM solution of an approximately 400 bp dsDNA sample (Fig. 4). For ΔN25-NColE7, approximately ten times greater protein concentration than for ΔN4-NColE7-C* was applied to achieve a substantial gel mobility shift, in agreement with its weaker DNA binding.

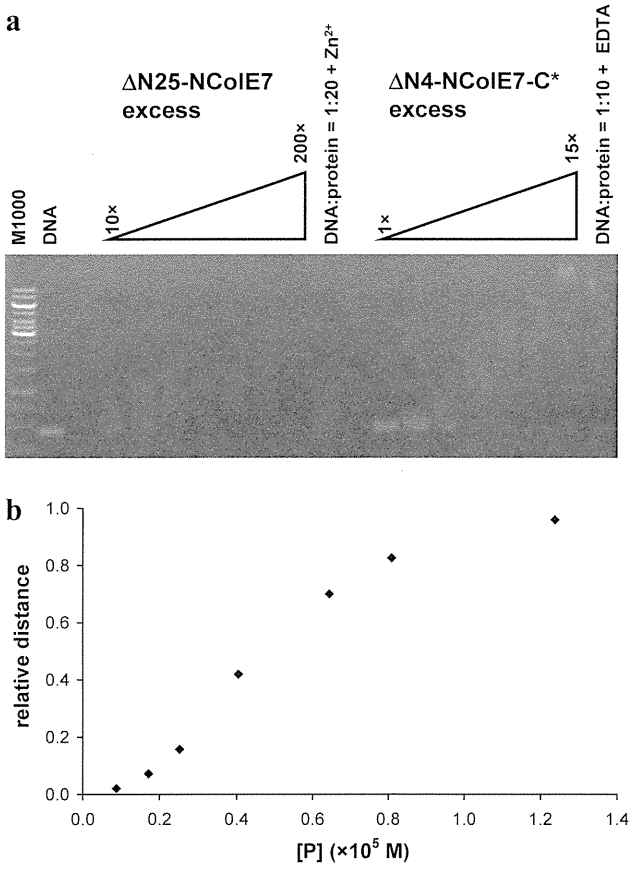

Fig. 4 **a** Gel mobility shift assay for studying the DNA-binding ability of ΔN25-NColE7 (*left*) and ΔN4-NColE7-C* (*right*). The first lane from the left contains a 1,000 bp marker DNA; the second lane contains a 0.874 μM double-stranded DNA sample. In the following lanes equal amounts of the same DNA sample incubated for 1 h with increasing amounts of mutant proteins in a constant volume of 10 μl were loaded. The excess of the proteins was tenfold, 20-fold, 30-fold, 50-fold, 80-fold, 100-fold and 200-fold for ΔN25-NColE7 and onefold, twofold, threefold, fivefold, eightfold, tenfold and 20-fold for ΔN4-NColE7-C*. **b** The relative gel mobility shift of the ΔN4-NColE7-C* mutant (i.e. the normalized distance of the shifted band and the band of the unbound double-stranded DNA; the saturation distance was taken as 1.0) versus the equilibrium protein concentration, [P]. At the inflection point [P] = K_D

Addition of Zn^{2+} ions to the ΔN25-NColE7 solution containing 20-fold excess of protein (see Fig. 4a) did not result in a change in the position of the band. Two reasons may account for this: (1) the protein had already bound metal ions, however, this would be in contrast with our previous results, or (2) the binding of Zn^{2+} ions is not necessary for DNA binding—similarly as for NColE7. In agreement with this latter observation, an excess of EDTA added to the ΔN4-NColE7-C*–DNA system did not cause any change in the position of the DNA band (Fig. 4a).

For ΔN4-NColE7-C* an apparent stability constant was estimated on the basis of the gel mobility shift assay. Simplifying conditions were introduced assuming 1:1 DNA binding site (10 bp DNA)—protein complex (P-DNA) formation and 100 % complex formation at the saturation of the curve. In Fig. 4b the relative gel mobility shift versus equilibrium protein concentration ([P]) is plotted. The latter was estimated as [P] = c_P − [P-DNA], where c_P is the total concentration of the protein, and [P-DNA] is the equilibrium concentration of the protein–DNA complex, which is proportional with the relative distance of the shifted band from the unbound DNA on the gel. K_D = ([P] × [DNA])/[P-DNA] at the inflection point, where 50 % of the DNA binding sites are occupied by the protein, i.e. [DNA] = [P-DNA], and thus [P] = K_D, where K_D is the apparent dissociation constant related to the formation of protein–DNA complexes at each binding site. By the above considerations, K_D was estimated to be approximately 5.0 μM (pK_D ~ 5.3) for a ΔN4-NColE7-C*–DNA binding site complex.

Molecular dynamics calculations

Structural changes of NColE7 and the mutant proteins (it should be noted that in the calculations the native sequences were applied without any tags) in explicit SPC/E water were tracked by 22 ns molecular dynamics calculations. Figure 5a describes the change in the root mean square deviation (RMSD) of backbone atoms in the molecule with respect to the reference structure at the 500th picosecond during the simulation. According to the RMSD diagram for the NColE7 (446–576) simulation, a relatively stable structure is formed after 2.5 ns of solvation, causing a 0.2 nm difference, as compared with the start of the simulation. The structure slightly changes until 10 ns, and then it fluctuates around 0.2 nm.

The structure of the ΔN4-NColE7 (450–576) mutant behaves similarly to the wild-type NColE7, but the ΔN25-

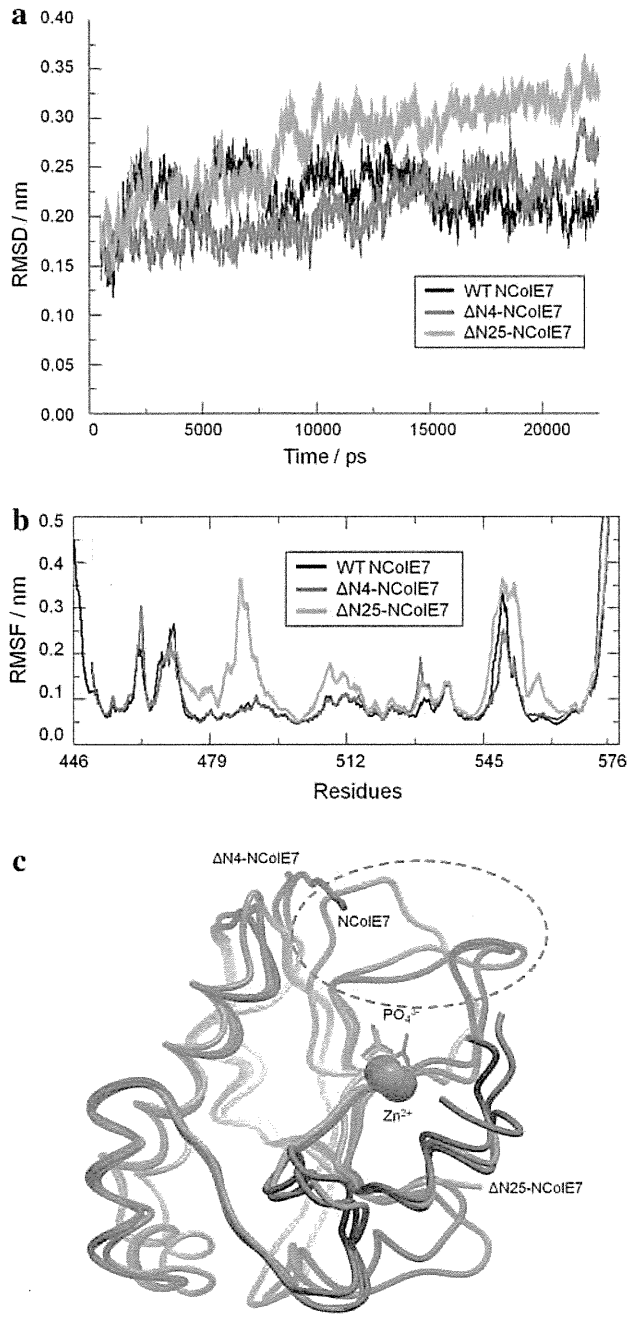

Fig. 5 **a** The root mean square deviation (*RMSD*) versus time as a result of 22 ns molecular dynamics calculations in explicit SPC/E water. **b** The average motion of backbone atoms during the molecular dynamics simulation in proteins. **c** The average structure of the proteins in the 20th to 25th nanosecond range of the simulations. The loop between the two β-strands of the HNH motif is highlighted by a *circle*. *RMSF* root mean square fluctuation, *WT* wild type

NColE7 (471–576) mutant goes through more serious changes, reaching an RMSD of 0.35 nm by the end of simulation. The RMSD for both mutants increased with time during the whole simulation, in contrast to that of wild-type NColE7. This means that shortening the protein caused remarkable changes in protein dynamics even if only four residues were cut at the N-terminus.

Figure 5b shows the average motion of each backbone atoms (root mean square fluctuation) during the whole simulation. The HNH motif is at the C-terminal part of the protein, and as such is at the right side of the diagram. The intense peak at about the 550th residue corresponds to the loop that joins the two β-sheets of the HNH motif. It is a functionally important part of the protein: the conserved residue N560 is located here, and is responsible for orienting the general base H545.

A significant difference can be observed between NColE7 and the shortened mutants. Two regions of ΔN25-NColE7, i.e. amino acid residues 485–487 and 511–515, show increased motion. These parts of NColE7 are loops leaning approximately parallel to the original N-terminal part that is missing from ΔN25-NColE7. The α-helices in the neighbourhood remained unchanged. The region including residues 530–535 also forms a loop at the N-terminal end of ΔN25-NColE7, and it interacts with the helix of the HNH motif. Interestingly, these residues also show an increased root mean square fluctuation in the case of ΔN4-NColE7, which suggests that the deletion of the last four residues at the N-terminus has an influence on the dynamics of the middle part of the protein. Residues 547–560 form the loop between the β-strands of the HNH motif. Changes in the dynamics of this loop can strongly influence the function of the protein as mentioned above.

Figure 5c shows the average structure of each mutants taken from the 20th to 25th nanosecond region of the simulation. The most obvious effect of shortening the N-terminus is the change in the orientation of the loop between the two β-strands of the HNH motif (highlighted by a circle in Fig. 5c). In case of ΔN25-NColE7, the missing N-terminal loop caused the two neighbouring loops to approach one another. Therefore, the loop in the HNH motif lost its original orientation. A smaller but not negligible movement can also be seen in ΔN4-NColE7: the HNH loop is also shifted in this case. As mentioned above, there are catalytically important residues here. The change of the average orientation and flexibility of the loop between β-strands of the HNH motif could be a reason for the decreased nuclease activity of the shortened proteins. The N-terminal loop can be considered as a structural spacer between the HNH loop and the DNA binding loop of the protein. It is also worth mentioning, that the N-terminal loop remained in an unchanged position in ΔN4-NColE7 even without the positively charged amino acids being deleted in this mutant. This has also been observed for the ΔN4-NColE7-C* mutant [43].

In a previous study of NColE7 [32], it was supposed that NColE7 can bind the DNA substrate in two different manners: coordinating also a water molecule or without it.

That is, the Zn^{2+} ion may have a temporary fifth coordination site that can provide a general acid (assisting in the protonation of the leaving group) in the form of an induced coordinated water molecule. During the 25 ns simulations, no such structure was found: there was no water molecule near the metal ion. However, a change was detected in the solvent distribution around the Zn^{2+} ion in the mutants. The shorter the protein, the looser the structure, which allows more water molecules to get near to or in the active centre (Fig. S6).

Semiempirical quantum chemical calculations

PM6/MOZYME/conductor-like screening model semiempirical quantum chemical calculations were performed to further investigate the fine changes in the active centre of the protein. Figure 6 shows the active centre in the optimized structures of NColE7, ΔN4-NColE7 and ΔN25-NColE7. The proteins were superimposed with PyMOL [58] using the whole length of the corresponding sequences. The RMSD of the full-length backbone relative to the initial structure of the calculations (PDB ID 1M08) was 1.010 nm for NColE7 (127 atoms fitted), 0.604 nm for ΔN4-NColE7 (109 atoms fitted) and 0.893 for ΔN25-NColE7 (100 atoms fitted). Aligning with the optimized structure of the NColE7 resulted in an RMSD of 0.527 nm (110 atoms) for ΔN4-NColE7 and 0.824 (89 atoms) for ΔN25-NColE7. The active centre of the ΔN4-NColE7 mutant is similar to that of the wild-type enzyme, as the orientation of the histidine side chains is almost identical (Fig. 6b). However, the small differences in the structures lead to different orientations of the phosphate and Zn^{2+} ions. This is even more obvious in the case of ΔN25-NColE7. The changed geometry around Zn^{2+} and phosphate ions could be a reason for the decreased metal- and DNA-binding ability of the ΔN25-NColE7 mutant. This again points to the fact that the removal of the N-terminal part has a significant effect on the structure of the C-terminal active centre.

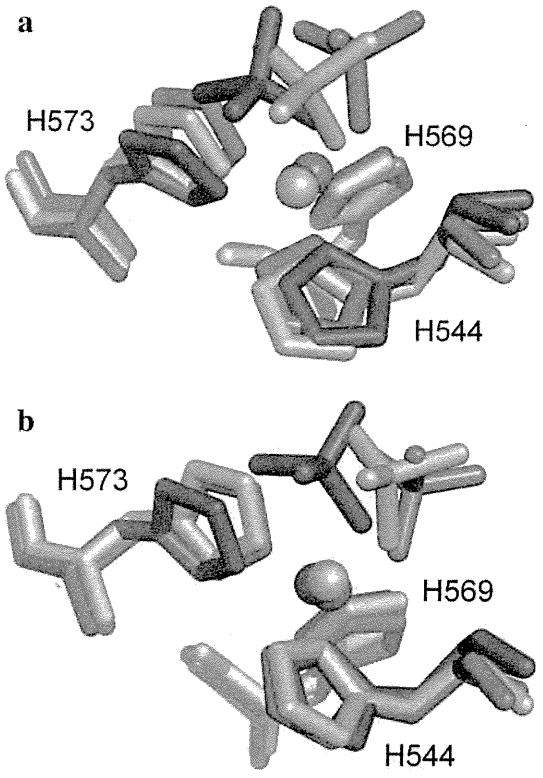

Fig. 6 Superposition of the metal ion binding residues in the optimized structures of NColE7 (*blue*), ΔN4-NColE7 (*yellow*) and ΔN25-NColE7 (*red*) with phosphate and zinc ions. The backbone atoms of the proteins were aligned with **a** the 1M08 structure [6] and **b** the optimized structure of NColE7

Conclusions

The necessity of the arginine residue at the N-terminus for hydrolytic activity of NColE7 poses the possibility of positive allosteric control in this protein. Mass spectrometry, SRCD and fluorescence spectroscopy and agarose gel mobility shift assays provided information on the effect of the removal of N-terminal sequences on the Zn^{2+} ion and DNA binding in ΔN4-NColE7-C* and ΔN25-NColE7 mutants. The longer protein bound both Zn^{2+} ion and DNA more strongly than its shorter counterpart owing to the structural stabilization effect of the N-terminal amino acids. The C-terminal mutation in ΔN4-NColE7-C* might affect these properties, but our results here and in [43] strongly suggest that the C-terminal flanking sequence does not participate in the metal ion or DNA binding. Molecular dynamics and semiempirical quantum chemical calculations performed in parallel showed that the absence of the N-terminal sequences resulted in significantly increased movement of the backbone atoms in regions of possible interactions with the N-terminal loop: residues 485–487, 511–515 and 570–571 for ΔN25-NColE7, and residues 467–468, 530–535 and 570–571 for ΔN4-NColE7. The distortion of the active centre predicted by semiempirical quantum chemical calculations could also be the reason for weak Zn^{2+} binding of ΔN25-NColE7. These results lead to the conclusion that the N-terminal loop plays an important role in the positioning of the arginine residue for the control of the DNAse activity. The question arose whether this could be a common feature among the HNH family of endonucleases. Since the amino acid sequences of the bacterial colicins and pyocins display high similarity, the arginine is frequently found in a position similar to that in NColE7. Also in the available crystal structures of different members of the HNH family, e.g. Sm endonuclease, Vvn

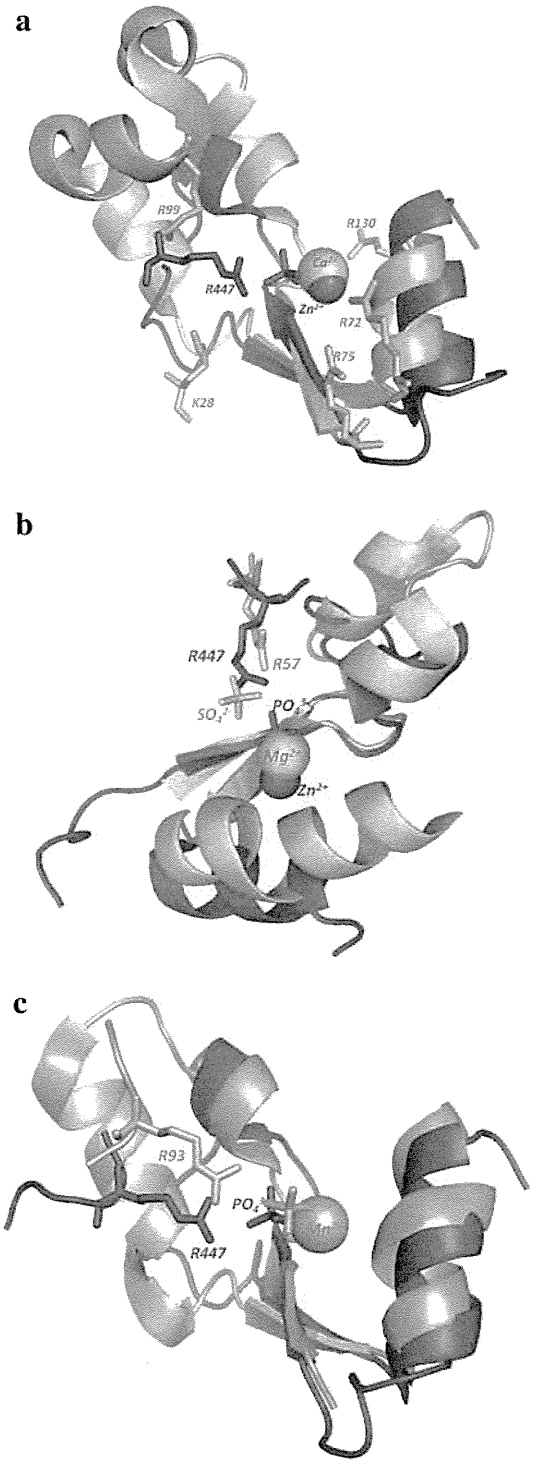

Fig. 7 The alignment of the HNH motifs of NColE7 (*red*, PDB ID 1MZ8) and selected proteins belonging to the HNH superfamily. **a** Vvn endonuclease (PDB ID 1OUP). **b** Sm endonuclease (PDB ID 1G8T). **c** Nuclease *A* (PDB ID 1ZM8)

endonuclease or nuclease A, we found arginine side chains in the spatial vicinity of the catalytic centre (Fig. 7). The answer thus seems to be positive. However, since the arginines are not always situated at the N-termini of the proteins, it is difficult to identify them without knowing the 3D structure. Therefore, a detailed bioinformatic study is required.

Acknowledgments This work has received support through the Hungarian Science Foundation (OTKA-NKTH CK80850), TÁMOP-4.2.1/B-09/1/KONV-2010-0005 and TÁMOP-4.2.2/B-10/1-2010-0012. The computational resources at High Performance Computing of the University of Szeged and financial support from the European Union Research Infrastructure Action FP7 (Integrated Activity on Synchrotron and Free Electron Laser Science, contract no. FP7/2007-2013; no. 226716) are also gratefully acknowledged. B.G. thanks the Japan Society for the Promotion of Science, and I.N.J-S. and A.C. thank the Hungarian Scholarship Board for the fellowships provided.

References

1. Chak K-F, Kuo W-S, Lu F-M, James R (1991) J Gen Microbiol 137:91–100
2. Lin Y-H, Liao C-C, Liang P-H, Yuan HS, Chak K-F (2004) Biochem Biophys Res Commun 318:81–87
3. Liao C-C, Hsia K-C, Liu Y-W, Leng P-H, Yuan HS, Chak K-F (2001) Biochem Biophys Res Commun 284:556–562
4. Cheng Y-S, Shi Z, Doudeva LG, Yang W-Z, Chak K-F, Yuan HS (2006) J Mol Biol 356:22–31
5. Sui M-J, Tsai L-C, Hsia K-C, Doudeva LG, Ku W-Y, Han GW, Yuan HS (2002) Protein Sci 11:2947–2957
6. Cheng Y-S, Hsia K-C, Doudeva LG, Chak K-F, Yuan HS (2002) J Mol Biol 324:227–236
7. Chak K-F, Safo MK, Ku W-Y, Hsieh S-Y, Yuan HS (1996) Proc Natl Acad Sci USA 93:6437–6442
8. Hsieh S-Y, Ko T-P, Tseng M-Y, Ku W-Y, Chak K-F, Yuan HS (1997) EMBO J 16:1444–1454
9. Dennis CA, Videler H, Paupit RA, Wallis R, James R, Moore GR, Kleanthous C (1998) Biochem J 333:183–191
10. Ko T-P, Liao C-C, Ku W-Y, Chak K-F, Yuan HS (1999) Structure 7:91–102
11. Kleanthous C, Walker D (2001) Trends Biochem Sci 26:624–631
12. Kolade OO, Carr SB, Kühlmann UC, Pommer A, Kleanthous C, Bouchcinsky CA, Hemmings AM (2002) Biochimie 84:439–446
13. Orlowski J, Bujnicki JM (2008) Nucleic Acids Res 36:3552–3569
14. Eastberg JH, Eklund J, Monnat R, Stoddard BL (2007) Biochemistry 46:7215–7225
15. Mehta P, Katta K, Krishnaswamy S (2004) Protein Sci 13:295–300
16. Hsia K-C, Chak K-F, Liang P-H, Cheng Y-S, Ku W-Y, Yuan HS (2004) Structure 12:205–214
17. Michel-Briand Y, Baysse C (2002) Biochimie 84:499–510
18. Shen BW, Landthaler M, Shub DA, Stoddard BL (2004) J Mol Biol 342:43–56
19. Ghosh M, Meiss G, Pingoud A, London RE, Pedersen LC (2005) J Biol Chem 280:27990–27997
20. Kriukiene E, Lubiene J, Lagunavicius A, Lubys A (2005) Biochim Biophys Acta 1751:194–204
21. Saravanan M, Bujnicki JM, Cymerman IA, Rao DN, Nagaraja V (2004) Nucleic Acids Res 32:6129–6135
22. Saravanan M, Vasu K, Ghosh S, Nagaraja V (2007) J Biol Chem 282:32320–32326
23. Cymerman IA, Obarska A, Skowronek KJ, Lubys A, Bujnicki MJM (2006) Proteins 65:867–876
24. Jakubauskas A, Giedriene J, Bujnicki JM, Janulaitis A (2007) J Mol Biol 370:157–169

25. Sokolowska M, Czapinska H, Bochtler M (2009) Nucleic Acids Res 37:3799–3810
26. Veluchamy A, Mary S, Acharya V, Mehta P, Deva T, Krishnaswamy S (2009) Bioinformation 6:80–83
27. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, Bateman A (2008) Nucleic Acids Res 36:D281–D288
28. Huang H, Yuan HS (2007) J Mol Biol 368:812–821
29. Wang Y-T, Yang W-J, Li C-L, Doudeva LG, Yuan HS (2007) Nucleic Acids Res 35:584–594
30. Papadakos G, Wojdyla JA, Kleanthous C (2012) Q Rev Biophys 45:57–103
31. Mate MJ, Kleanthous C (2004) J Biol Chem 279:34763–34769
32. Doudeva LG, Huang H, Hsia K-C, Shi Z, Li C-L, Shen Y, Cheng C-L, Yuan HS (2006) Protein Sci 15:269–280
33. Pommer AJ, Kuhlmann UC, Cooper A, Hemmings AM, Moore GR, James R, Kleanthous C (1999) J Biol Chem 274:27153–27160
34. Hannan JP, Whittaker SBM, Hemmings AM, James R, Kleanthous C, Moore GR (2000) J Inorg Biochem 79:365–370
35. Keeble AH, Hemmings AM, James R, Moore GR, Kleanthous C (2002) Biochemistry 41:10234–10244
36. van den Bremer ETJ, Jiskoot W, James R, Moore GR, Kleanthous C, Heck AJR, Maier CS (2002) Protein Sci 11:1738–1752
37. Hannan JP, Whittaker SB, Davy SL, Kuhlmann UC, Pommer AJ, Hemmings AM, James R, Kleanthous C, Moore GR (1999) Protein Sci 8:1711–1713
38. van den Bremer ETJ, Keeble AH, Visser AJWG, van Hoek A, Kleanthous C, Heck AJR, Jiskoot W (2004) Biochemistry 43:4347–4355
39. Ku W-Y, Liu Y-W, Hsu Y-C, Liao C-C, Liang P-H, Yuan HS, Chak K-F (2002) Nucleic Acids Res 30:1670–1678
40. Shi Z, Chak K-F, Yuan HS (2005) J Biol Chem 280:24663–24668
41. Li C-L, Hor L-I, Chang Z-F, Tsai L-C, Yang W-Z, Yuan HS (2003) EMBO J 22:4014–4025
42. Gyurcsik B, Czene A (2011) Future Med Chem 3:1935–1966
43. Tóth E, Czene A, Gyurcsik B, Otten H, Poulsen J-CN, Larsen S, Christensen HEM, Nagata K (2013) Acta Crystallogr Sect D
44. Limao-Vieira P, Giuliani A, Delwiche J, Parafita R, Mota R, Duflot D, Flament JP, Drage E, Cahillane P, Mason NJ, Hoffmann SV, Hubin-Franskin MJ (2006) Chem Phys 324:339–349
45. Fahrni CJ, O'Halloran TV (1999) J Am Chem Soc 121:11448–11458
46. Berendsen HJC, van der Spoel D, van Drunen R (1995) Comput Phys Commun 91:43–56
47. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) J Chem Theory Comput 4:435–447
48. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) J Comput Chem 25(13):1656–1676
49. Olsson MHM, Søndergaard CR, Rotkowski M, Jensen JH (2011) J Chem Theory Comput 7:525–537
50. Stewart JJP (2008) MOPAC2009. Stewart Computational Chemistry, Colorado Springs. http://openmopac.net
51. Stewart JJP (2007) J Mol Model 13:1173–1213
52. Stewart JJP (2009) J Mol Model 15:765–805
53. Stewart JJP (1996) Int J Quantum Chem 58:133–146
54. Klamt A, Schüümann G (1993) J Chem Soc Perkin Trans 2 799–805
55. Anthony LC, Suzuki H, Filutowicz M (2004) J Microbiol Methods 58:243–250
56. Levin KB, Dym O, Albeck S, Magdassi S, Keeble AH, Kleanthous C, Tawfik DS (2009) Nat Struct Mol Biol 16:1049–1055
57. Wang Y-T (2009) Wright JD, Doudeva LG, Jhang H-C, Lim C, Yuan HS. J Am Chem Soc 131:17345–17353
58. DeLano WL (2006) PyMOL version 0.99rc6. DeLano Scientific, San Carlos
59. Jecklin MC, Schauer S, Dumelin CE, Zenobi R (2009) J Mol Recognit 22:319–329
60. Miles AJ, Wallace BA (2006) Chem Soc Rev 35:39–51

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of the deacetylase domain of *Escherichia coli* PgaB, an enzyme required for biofilm formation: a circularly permuted member of the carbohydrate esterase 4 family

Takashi Nishiyama, Hiroki Noguchi, Hisashi Yoshida, Sam-Yong Park and Jeremy R. H. Tame*

Protein Design Laboratory, Graduate School of Nanobioscience, Yokohama City University, Suehiro 1-7-29, Yokohama, Kanagawa 230-0045, Japan

Correspondence e-mail: jtame@tsurumi.yokohama-cu.ac.jp

Received 5 July 2012
Accepted 8 October 2012

PDB Reference: N-terminal domain of PgaB, 3vus

Bacterial biofilm formation is an extremely widespread phenomenon involving the secretion of a protective exopolysaccharide matrix which helps the bacteria to attach to surfaces and to overcome a variety of stresses in different environments. This matrix may also include proteins, lipids, DNA and metal ions. Its composition depends on the bacterial species and growth conditions, but one of the most widely found components is polymeric β-1,6-*N*-acetyl-D-glucosamine (PGA). Several studies have suggested that PGA is an essential component of biofilm and it is produced by numerous bacteria, including *Escherichia coli*, *Staphylococcus epidermis*, *Yersinia pestis*, *Bordetella* spp. and *Actinobacillus* spp. In *E. coli*, PGA production and export are dependent on four genes that form a single operon, *pgaABCD*, which appears to have been transferred between various species. Biofilms themselves are recognized as environments in which such horizontal gene transfer may occur. The *pga* operon of *E. coli*, which is even found in innocuous laboratory strains, is highly homologous to that from the plague bacterium *Yersinia pestis*, and biofilm is believed to play an important role in the transmission of *Yersinia*. The crystal structure of the N-terminal domain of PgaB, which has deacetylase activity, is described and compared with models of other deacetylases.

1. Introduction

Bacteria may adopt either a free-living 'planktonic' or a sessile lifestyle. Sessile bacterial growth of a structured community of cells attached to a surface is known as a biofilm (Hall-Stoodley *et al.*, 2004). Biofilm formation is a general feature of microorganisms, and has been of intense interest for decades owing to its medical relevance; metabolically quiescent bacteria within biofilms are hidden from the immune system and are difficult to treat with antibiotics, and biofilms frequently form on catheters or other surgical implants (Costerton *et al.*, 1999). Polymeric β-1,6-*N*-acetyl-D-glucosamine (PGA) is a principal component of the exopolysaccharide matrix of many bacteria. Without the *pgaABCD* operon, *Escherichia coli* is unable to form biofilms (Wang *et al.*, 2004), and cleavage of PGA breaks up biofilms (Itoh *et al.*, 2005). The enzymes encoded by the operon produce a linear homopolymer of β-1,6-linked *N*-acetylglucosamine. This polymer was first described in studies of *Staphylococcus epidermis* and was referred to as polysaccharide intracellular adhesin (PIA; Mack *et al.*, 1996). As well as being involved in bacterial attachment to biofilm (Agladze *et al.*, 2005), PGA can play important roles in

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

host–microbe interactions. It has been implicated in the colonization and virulence of both Gram-positive and Gram-negative bacteria (Cerca *et al.*, 2007; Vuong, Kocianova *et al.*, 2004; Vuong, Voyich *et al.*, 2004). Biofilm formation by *E. coli*, in particular the role of the *pgaABCD* operon, has been studied in detail by the group of Romeo (Itoh *et al.*, 2008; Wang *et al.*, 2004). Transcription of the *pgaABCD* operon is repressed by CsrA (carbon storage regulatory A), an RNA-binding protein that binds to the untranslated leader of target mRNA (Wang *et al.*, 2005). *pgaABCD* transcription requires NhaR (a LysR-family DNA-binding protein), which switches on PGA production in response to high pH and also to high concentrations of salt (Goller *et al.*, 2006).

PgaC and PgaD are inner membrane proteins, with the latter having no sequence similarity to any known structure; PgaC is a member of the GT-2 glycosyltransferase family. They consume UDP-GlcNAc from the cytoplasm and release PGA into the periplasm. PgaA and PgaB are not directly involved in PGA synthesis, but are required for its export (Itoh *et al.*, 2008). PgaA is believed to form a β-barrel in the outer membrane through which the polymer passes to exit the cell; a homologous structure from *Pseudomonas aeruginosa* has recently been crystallized and shown to form an 18-stranded β-barrel (Whitney *et al.*, 2011). PgaB is anchored to the outer membrane by attachment of lipid; it has a classical signal sequence followed by a cysteine (Cys21), which is the lipid-attachment site. Sequence analysis of PgaB suggests that the protein has an N-terminal deacetylase domain, a member of carbohydrate esterase family 4 (CE4) in the CAZy classification scheme (http://www.cazy.org; Cantarel *et al.*, 2009). Other members of this family act on carbohydrate polymers such as xylan or chitin, but only PgaB has been reported to be involved in PGA synthesis or export. The C-terminal domain of PgaB has an unknown function, but may bind to PgaA or to PGA itself. The deacetylase activity of PgaB appears to be necessary for PGA export, presumably by the exposure of amine groups, which acquire a positive charge (Itoh *et al.*, 2008). NMR analysis of PGA obtained by overexpression of the *pgaABCD* operon showed that the polymer has a high molecular weight (about 400 kDa) but contains fewer than 3% deacetylated residues (Wang *et al.*, 2004). PGA from other species may show a much higher proportion; that from *S. epidermis*, for example, shows 15–20% deacetylation, although the proportion may vary according to the growth conditions (Vuong, Kocianova *et al.*, 2004). Published studies therefore suggest that the deacetylase activity of PgaB is essential for biofilm formation but that it is only weakly active. The growing polymer chain presumably passes close to PgaB in order to be acted upon by it, yet fewer than one residue in 20 undergoes modification. The proportion of residues which must be deacetylated in order to promote secretion appears to be low, but the possibility exists that inhibition of this weak enzyme activity could be an effective route to prevent biofilm formation by *E. coli*. Biofilm formation by the plague bacterium *Yersinia pestis* involves an operon (*hmsHFRS*) that is highly homologous to *E. coli pgaABCD* and appears to play a significant role in bacterial transmission (Bobrov *et al.*, 2008; Darby, 2008). We have therefore solved the crystal structure of the catalytic domain of PgaB (PGABN), the first CE4 model from this organism, and compared it with known structures.

2. Materials and methods

2.1. Cloning

pgaB was amplified by PCR from *E. coli* JM109 chromosomal DNA and cloned into a modified pET28 expression vector using *Bam*HI and *Xho*I restriction sites. The primer sequences were CGGGATCCGCCCAGTCAAGAACATC-ATTTATACCG and CCGAGCCTCGAGTTACTGGAG-GTTTTCGTCATAAAC. The PCR product was digested with *Bam*HI and *Xho*I at 310 K for 2 h before purification using a QIAquick PCR Purification Kit (Qiagen). The purified PCR product was ligated into the cut vector using T4 DNA ligase (Wako) at room temperature for 1 h. The ligation mixture was used to transform *E. coli* DH5α, and pET28b-pgaBN was prepared from cultures using QIAprep (Qiagen). This construct directs expression of residues Gln24–Gln330 of PgaB with a hexahistidine tag at the N-terminus that is cleavable with TEV protease.

2.2. Expression and purification

pET28b-pgaBN was transformed into *E. coli* BL21 (DE3) and cells were grown at 310 K with shaking in 3 l LB medium containing kanamycin (50 μg ml^{-1}). When the OD_{600} of the culture reached 0.5–0.6, PGABN expression was induced by adding IPTG to a final concentration of 0.5 m*M* and growth was continued overnight at 288 K. The cells were collected by centrifugation at 3000*g* at 277 K for 30 min. The pellet was suspended in 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl and was then lysed by sonication on ice. The lysate was centrifuged at 38 000*g* and 277 K for 30 min. The supernatant solution was loaded onto a 10 ml nickel Sepharose column (GE Healthcare) equilibrated with 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl, 10 m*M* imidazole; after washing, it was eluted with 50 m*M* Tris–HCl pH 8.0, 250 m*M* imidazole. The major protein fractions were collected and digested with TEV protease overnight at 277 K during dialysis into 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl. The protease:PgaB ratio was 1:50. The protein was reloaded onto the washed nickel Sepharose column and eluted with 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl. The pooled fractions containing PGABN were dialyzed into 50 m*M* Tris–HCl pH 8.0, 0.1 *M* NaCl before concentration to 30 mg ml^{-1} using Amicon centrifugal filter units (Millipore).

2.3. Crystallization and structure determination

Crystallization experiments were performed at 293 K using the hanging-drop vapour-diffusion method. Crystals grew in 16%(*w*/*v*) PEG 3350, 0.18 *M* sodium acetate, 0.1 *M* bis-Tris–HCl pH 6.5. Data were collected on beamline 17A of the Photon Factory, Tsukuba. The highest resolution data were collected from a crystal which had been soaked briefly in 1 m*M* mercury chloride but which appeared to be native in phasing trials. The data were used in the final refinement and

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the outer shell.

Data set	Native	Hg, remote	Hg, inflection	Hg, peak	Pt
Space group	$P2_1$	$P2_1$			$P2_1$
Wavelength (Å)	1.00000	0.99321	1.00938	0.99957	1.00000
Unit-cell parameters (Å, °)	a = 39.6, b = 53.1, c = 144.2, β = 95.2	a = 39.0, b = 51.9, c = 144.7, β = 95.5			a = 39.3, b = 53.0, c = 144.4, β = 95.5
Resolution range (Å)	50.0–1.65 (1.68–1.65)	50.0–2.50 (2.54–2.50)			50.0–2.30 (2.34–2.30)
Reflections (measured/unique)	337723/63364	87449/19886	88848/19847	86805/19838	106033/25579
Completeness (%)	96.7/83.7	98.0/96.6	98.1/97.1	98.0/97.3	96.0/86.2
$R_{\rm merge}$† (%)	4.9/37.5	4.9/19.1	4.4/13.3	5.4/12.4	7.5/39.8
Multiplicity	4.9	4.4	4.5	4.4	4.2
$\langle I/\sigma(I)\rangle$	49.0	44.3	47.9	48.2	34.9
Refinement statistics					
Resolution range (Å)	25.0–1.65				
R factor/free R factor	0.205/0.257				
R.m.s.d. bond lengths (Å)	0.027				
R.m.s.d. bond angles (°)	2.29				
No. of water molecules	175				
Average B factor (protein/water) (Å^2)	33.4/35.2				
Ramachandran plot, residues in (%)					
Most favourable regions	87.6				
Additional allowed regions	11.1				
Generously allowed regions	0.9				
Disallowed regions	0.4				

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation, $\langle I(hkl)\rangle$ in the mean value for that reflection and the summations are over all equivalents.

revealed a bound Hg atom with low occupancy on refinement. A total of 250 images of 1° oscillation were collected for each data set. Data processing and scaling were carried out with *HKL*-2000 and *SCALEPACK* (Otwinowski & Minor, 1997). The space group was found to be $P2_1$, with two molecules in the asymmetric unit. Data statistics are given in Table 1. Multiple-wavelength data were collected to 2.5 Å resolution on the same beamline using a crystal soaked in 2 m*M* mercury chloride for 14 h. A single data set to 2.3 Å resolution was also collected using a crystal soaked in 2 m*M* K_2PtCl_4 for 24 h. The native data set and the Pt-soak data set were collected using incident radiation of 1.000 Å wavelength. Phases were calculated using *PHENIX* (Adams *et al.*, 2010). Model building was carried out with *ARP/wARP* (Langer *et al.*, 2008; Morris *et al.*, 2003) and *Coot* (Emsley *et al.*, 2010; Emsley & Cowtan, 2004). Refinement was carried out with *REFMAC* (Murshudov *et al.*, 2011) and the *CCP*4 suite (Winn *et al.*, 2011). Noncrystallographic symmetry restraints and TLS group refinement were not applied. The Ramachandran plot showed several residues in unusual positions, but the agreement between the two copies of the molecule was very good. Slight main-chain disorder at residues 66 and 100 led to several pairs of equivalent residues in chains *A* and *B* lying at different positions in the Ramachandran plot. Two Ramachandran outliers were the active-site residues His55 and Asp115. Isotropic temperature factors were refined with default restraints. Figures were prepared with *PyMOL* (DeLano, 2002). Water molecules were checked manually. A single Hg atom was modelled into the structure with an occupancy of 20%. Density around the Zn atom of chain *B* showed less than full occupancy of the acetate, and a partially ordered water molecule was also modelled at this site. The final model and structure factors have been deposited in the Protein Data Bank as entry 3vus.

3. Results

3.1. Overall structure

Initial sequence analysis of the *pgaB* gene from *E. coli* suggested that the deacetylase activity resided in an N-terminal domain (PGABN) of roughly 300 residues in length. A threading search of known structures using the *Wurst* server (Torda *et al.*, 2004) suggested a strong match to aldolase from *Trypanosoma brucei* (PDB entry 1f2j), which has 14% sequence identity to the PgaB N-terminus (Chudzik *et al.*, 2000). This TIM β-barrel structure has no enzymatic activity in common with PgaB, but has eight $\alpha\beta$ structure repeats, whereas the CE4 enzymes have seven. PGABN shows sequence similarity to PdaA, an *N*-deacetylase from *Bacillus subtilis*, which is a CE4 member of known structure (Blair & van Aalten, 2004). On the basis of these results, a construct was made to express PgaB from Gln24 to Gln330 by PCR from *E. coli* genomic DNA, omitting the signal peptide and associated cysteine residue. This construct yielded 25 mg purified protein (with the N-terminal histidine tag removed) per litre of culture. Analytical ultracentrifugation showed the protein to exist as a monomer in solution (data not shown).

Crystals were grown that diffracted to almost 1.6 Å resolution and the structure (shown in Fig. 1*a*) was solved using two heavy-atom derivatives and multiple-wavelength anomalous measurements. Two copies of the molecule were found in the asymmetric unit. The ordered residues visible in the electron-density map began at Pro43 and ended at Val308 or

Gln309, so that roughly 20 disordered residues at each end of the polypeptide were not modelled. EXAFS analysis of a native crystal clearly showed the presence of zinc, a metal not purposely added to the protein at any stage of preparation or purification. In each copy of the molecule a single metal ion was readily located and modelled (Fig. 1*b*). A search for related models in the PDB using *DALI* (Holm & Rosenström, 2010) yielded the best match as SlCE4 (PDB entry 2cc0), an acetylxylan esterase from *Streptomyces lividans* involved in plant cell-wall degradation which has a structure similar to PdaA (Taylor *et al.*, 2006). Both enzymes are members of Pfam family PF01522. Overlaying the structures with *SSM* (Krissinel & Henrick, 2004) matched 128 residues of PdaA with PGABN with 18% sequence identity, giving an C^{α} r.m.s.d. of 2.3 Å. Similar results were obtained matching PGABN to SlCE4, giving an alignment of 138 residues (19.6% identity) and 2.4 Å r.m.s.d. Straightforward sequence searches against the PDB gave much poorer alignments, with the best fits being about 30% identical sequences of roughly 50 residues in length. Although most of these hits were to CE4 enzymes, they also included a subunit of the yeast ribosome and clearly were not all biologically relevant. Examination of the models quickly showed that PGABN is circularly permuted with respect to known CE4 members and that the last β-strand of their TIM-like barrel corresponds to the first strand of PGABN. Simple sequence alignments of PGABN with other members of the CE4 family therefore suggested quite different matches to those found from comparisons of the actual models and did not match conserved active-site residues. A sequence alignment between SlCE4 and PGABN is shown in Fig. 2(*a*), demonstrating the shift of one strand of the central barrel from the N-terminus of PGABN to the C-terminus of other CE4 enzymes. The result of fitting PGABN and SlCE4 by *SSM* is shown in Fig. 2(*b*).

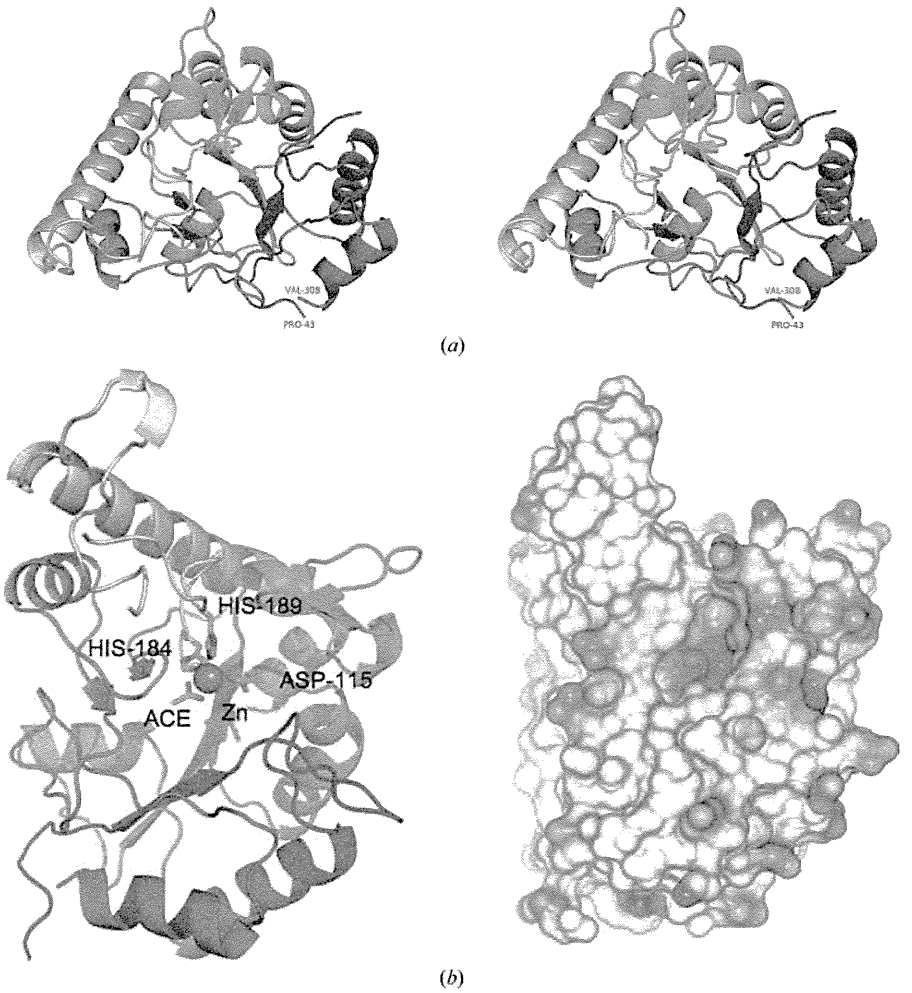

Figure 1
(*a*) A stereoview of the C^{α} trace of the PgaB N-terminal domain (PGABN), looking into the central barrel. The trace is coloured from blue (Pro43) to red (Val308), with α-helices shown as coils and β-strands as arrows. The figure was drawn using *PyMOL* (DeLano, 2002). Secondary structure was determined automatically. (*b*) The C^{α} trace of the PGABN monomer, coloured as in (*a*), is shown on the left, with the active-site zinc ion shown as a grey sphere. Residues coordinating the metal are shown as sticks, with O atoms coloured red and N atoms blue. The right-hand panel shows a surface representation of the monomer in the same orientation coloured by electrostatic potential. The zinc ion and acetate ligand were omitted from the calculation of the potential and are shown as ball-and-stick models over the protein surface. The strong negative charge of the active site is apparent. The remainder of the protein surface shows no distinct pattern of charge.

3.2. Active site

The active site of the CE4 family is found at the centre of the barrel-like β-sheet and involves a metal ion in some, but not all, cases. PdaA notably has no metal in the active site on purification and does not bind zinc ions even when these are added; cadmium ions do bind, but only at very high concentrations (Blair & van Aalten, 2004). Other members of the CE4 family have been reported to require cobalt for activity or to prefer cobalt to zinc (Taylor *et al.*, 2006). Nevertheless, two active-site residues (histidine and aspartic acid) were identified from a comparison of the PdaA and PgaB sequences; mutation of either residue in PgaB to alanine (D115A and H184A) blocked activity (Itoh *et al.*, 2008). PGABN shares the common coordination pattern of a zinc ion bound by two histidines and an aspartic acid residue (Fig. 3). The initial crystallization screen and subsequent optimization showed that the inclusion of acetate in the buffer greatly improved the crystal quality, and an acetate ion can be found coordinated to the metal ion of each monomer in the asymmetric unit. However, in one of these sites the occupancy seems to be less than 1.0, and a water molecule with partial occupancy was also modelled close to the zinc ion. Well diffracting

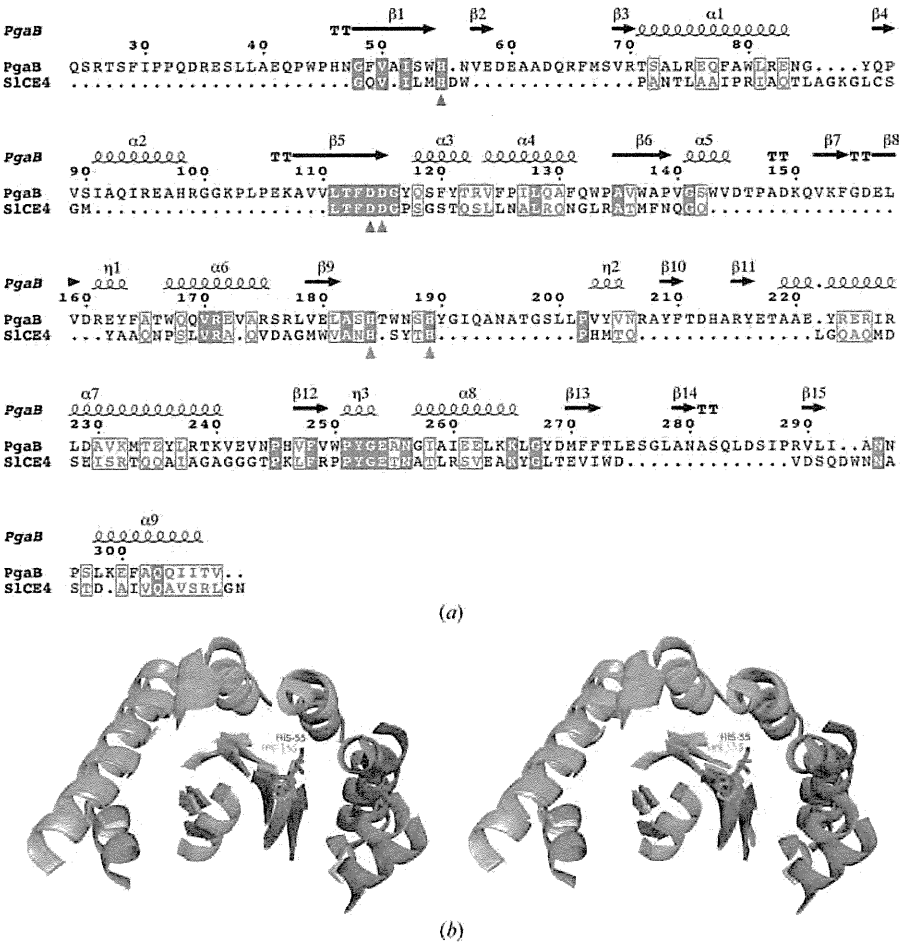

Figure 2
(*a*) A sequence alignment based on the crystal structures of PGABN and SlCE4 (PDB entry 2cc0). Conserved residues are shown in white on a red background. Red triangles show residues that coordinate the active-site zinc ion and blue triangles show nearby residues involved in catalysis. The residue numbering refers to PGABN. The first β-strand of PGABN (residues 48–55) matches the last strand of SlCE4, the sequence of which has been shifted from the C-terminus to the N-terminus in this alignment. His55 of PGABN matches His155 of SlCE4. The figure was produced by *ESPript* (Gouet *et al.*, 1999). (*b*) Overlay of the C^{α} traces of the PgaB N-domain and SlCE4 (PDB entry 2cc0), showing the helices and strands of both structures, but not coil. Alignment of the models was carried out with *SSM* (Krissinel & Henrick, 2004). Each structure is coloured from blue to red (N-terminus to C-terminus) as in (*a*), but circular permutation of the sequences leads to a marked difference in colouring, despite the close structural similarity. His55 of PGABN lies close to the active site, overlapping His155 of SlCE4, which is a conserved histidine in the CE4 family.

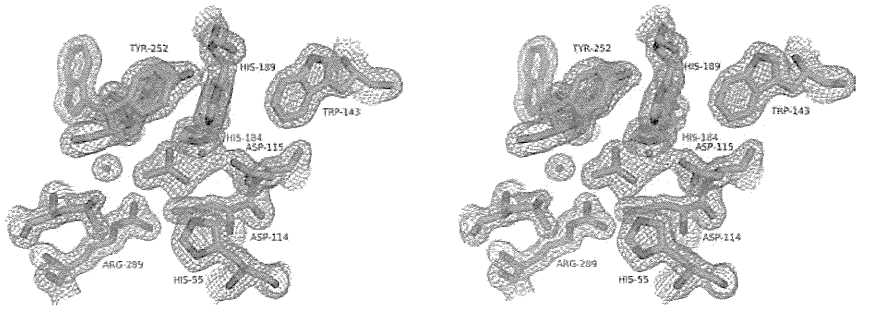

Figure 3
A stereo figure of the $2mF_o - DF_c$ electron-density map for the refined model covering the active site. Electron density is shown at a level of 1σ. The zinc ion is shown as a grey sphere and water molecules are shown as red spheres. The acetate ion coordinating the zinc ion is clearly visible in the density map and is found to hydrogen-bond to the conserved His55. The conserved LT*X*DDG motif of the NodB domain includes Asp114 and Asp115.

crystals could not be grown in the absence of acetate, suggesting that this ligand stabilized the protein. Attempts to remove acetate by soaking the crystals in acetate-free buffer limited the diffraction resolution.

Comparing PGABN with PdaA, it can be seen that many features of the active site are preserved. 128 C^{α} atoms from the core of the structures show roughly 17% sequence identity, and the zinc-binding residues His184 and His189 of PGABN closely overlap with His124 and His128 of PdaA. Loss of metal binding by PdaA is explained by the fact that it has no third coordinating residue; Asp115 of PGABN is replaced by an asparagine residue (Asn74) that points away from the active site (Fig. 4*a*). Other conserved residues of the active site of PdaA such as Phe98, Trp187 and Leu220 are not found in PGABN, the fold of which is quite different from PdaA in the region of the latter two residues. The active-site residue Asp73 of PdaA is preserved as Asp114 in PGABN, and in both enzymes this aspartic acid forms a salt bridge with an arginine side chain. This arginine residue (Arg163 in PdaA and Arg289 in PGABN) is found on a different β-strand of the barrel; this does not reflect the altered order of the secondary-structure elements but is a consequence of a quite different spatial geometry. Another conserved histidine residue (His55) lies close to the active site at the end of the first β-strand in PGABN, but its equivalent in PdaA (His222) lies at the end of the last β-strand of the barrel. In PdaA this histidine forms a salt bridge (His222–Asp193), but the aspartate has no equivalent in PGABN and His55 is only stabilized by a water molecule. However, His55 does hydrogen-bond to the acetate placed in density near the zinc ion (Fig. 3) and must contact the substrate in the active site. The ion pairs in this region of PdaA have been proposed to play a role in stabilizing charges during catalysis (Blair & van Aalten, 2004), but the details of any such mechanism presumably differ in members of the CE4 family with or without active-site metal ions.

Overlaying SlCE4 with PGABN gives a similar pattern to that with PdaA (Fig. 4*b*). Although SlCE4 is metal-dependent and PdaA has no metal centre, these two enzymes are not circularly permuted relative to one another, and the Asp73–Arg163 pair of PdaA is exactly mirrored by Asp12–Arg100 in SlCE4. His55 of PGABN is preserved as His155 in SlCE4. Thus, although the three enzymes share a common overall three-dimensional structure, each has an active site with characteristics not found in the others. PGABN and SlCE4 presumably share a common mechanism and the altered residues around the active site reflect the substrate preference.

3.3. Substrate binding

CE4 enzymes have a highly distorted barrel that lacks one of the $\alpha\beta$ repeats of regular TIM barrels, which creates a groove into which the extended polymer substrates of these enzymes can fit. PdaA has several positively charged residues lining this pocket (such as Lys34, Arg35 and Arg166) that are proposed to be involved in binding to the negatively charged peptidoglycan substrate. PGABN has no equivalent residues and acts upon an uncharged substrate; its enzyme activity is also presumably strongly affected by the insertion of two loops blocking one end of the substrate groove, apparently a feature unique to PGABN among the CE4 enzymes (Fig. 5). Residues Thr146–Phe164 face residues Tyr190–Pro202 on the opposite side of the groove. Neither loop appears to be stably fixed in position and residues 195–200 are not modelled in one copy of the molecule, but the presence of these loops may help to control access to the active site. A further loop (Arg207–Thr218) is more distant from the active site but may impede a long polymeric substrate from easy access to it. Overall, the structured domain of PGABN is rather longer than SlCE4 owing to these and other loops, which decorate its surface. Not all of these loops appear to be mobile; the surface loop from Ser276 to Pro288 appears to be fixed in place by hydrogen bonds formed between the side chain of Gln283 and the main-chain atoms of neighbouring residues.

Repeated attempts to grow well diffracting crystals of PGABN without acetate failed. The reason why acetate, the product of the reaction, strongly promotes crystallization is unclear from the structure, but the enzyme appears to be product-inhibited. As with previous structures from this family, we were

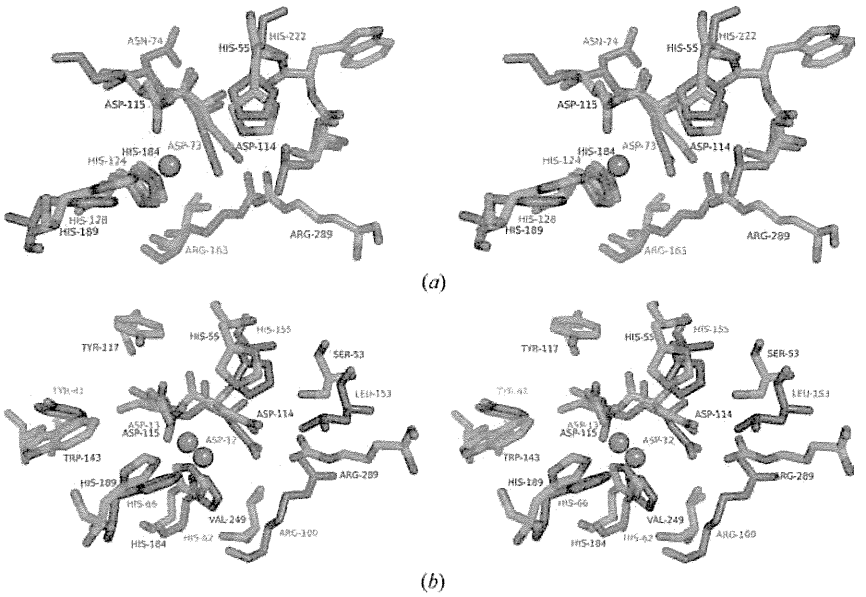

Figure 4
(*a*) A stereo overlay of PGABN (green) and PdaA (light blue) showing the active site. Labels in black indicate the residues in PGABN. O atoms are coloured red and N atoms blue. The models were fitted by overlapping 128 C^{α} atoms of the conserved core residues. The zinc ion of PGABN is shown as a grey sphere. PdaA (PDB entry 1w17) does not bind metal ions owing to the loss of the coordinating aspartate residue. Instead, Asn74 points away from the active site. Arg163 forms a salt bridge to Asp73, but the equivalent arginine in PGABN is Arg289, which sits on a different β-strand to Arg163 of PdaA. (*b*) Stereo overlay of the active sites of PGABN (green) and SlCE4 (brown). Both structures have a bound zinc ion with very similar coordination geometry. Other residues around the active site show low sequence conservation between the two structures, and Tyr117 of PGABN has no counterpart in SlCE4.

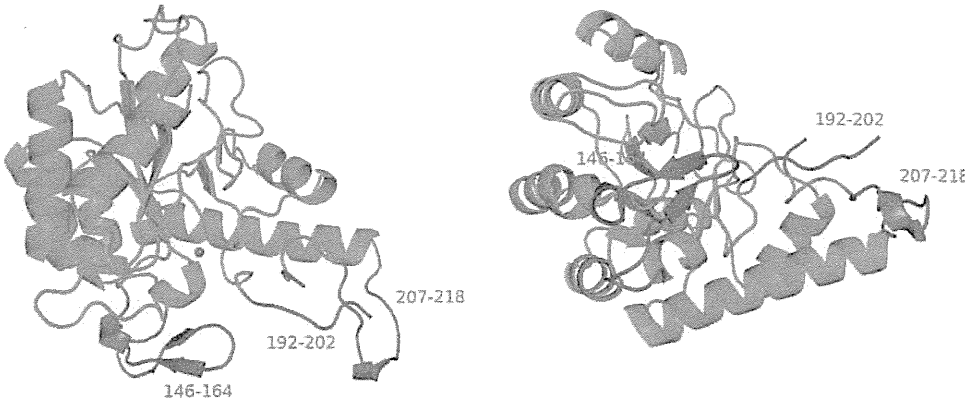

Figure 5
Two views of the C^{α} trace of PGABN rotated 90° relative to one another about a horizontal axis. The zinc ion is shown as a grey sphere. Surface loops near the active site are shown in blue. These regions are presumably flexible in solution and have no counterpart in SlCE4 or PdaA. They may bind to PgaA in the periplasm. Residues 195–200 are not visible in the electron-density map for the copy of the molecule shown.

therefore obliged to attempt to fit substrate analogues into the binding site by docking (Blair *et al.*, 2005). No clearly preferred binding sites were found which can be confidently predicted to bind single glucosamine residues, but the polymeric substrate may bind by weak interactions at several adjacent sites, just as a protease may require a minimum length of peptide in order to cut a single peptide bond. *In vivo*, it is also possible that the substrate is held in proximity to PgaB by interaction with other proteins in the synthetic machinery.

4. Discussion

A number of structures from the CE4 family have been solved by X-ray crystallography: the first of these was solved by a structural genomics consortium (PDB entry 1ny1; Northeast Structural Genomics Consortium, unpublished work), but no accompanying report has been published to date. The first enzyme from this family to be studied in detail was NodB, which removes the acetyl group from a GlcNAc residue in one step of the synthesis of Nod factors, which are molecular signals involved in the symbiosis of legumes and nitrifying bacteria (John *et al.*, 1993). A NodB homology domain was identified in this sequence (Kafetzopoulos *et al.*, 1993) and has been shown to be present in a variety of esterases with different substrates (Caufrier *et al.*, 2003). This conserved region begins roughly 20 residues upstream of the highly conserved LT*X*DDG motif, where *X* is F or Y (Fig. 2*a*). However, accurate sequence alignment of PGABN without the X-ray model was made difficult by the unique inserts that form loops near the substrate-binding site, and sequence analysis alone was unable to show that PGABN is a circularly permuted member of the CE4 family. The NodB domain forms part of a conserved tertiary structure but with associated secondary-structure elements attached to the N-terminus instead of the C-terminus, an arrangement not previously described. PGABN maintains commonly found features at the active site, including coordination to a catalytic zinc ion and nearby aspartic acid and histidine residues. The observed weak activity of the enzyme presumably arises from obstruction of the polymeric substrate by the unique surface loops of PGABN and/or a low intrinsic affinity of the catalytic domain for PGA. The crystallization of a fragment of PgaB has recently been reported. Little and coworkers attempted to crystallize full-length PgaB and identified a crystallizable fragment by proteolysis (Little, Whitney *et al.*, 2012). Our approach was to express both the N-terminal domain and the full-length protein directly, but so far only the N-terminal domain has yielded crystals.

Circular permutation of proteins has been known for some time and has become a route to artificial proteins of enhanced stability (Yu & Lutz, 2011). Hydrolases form one of the largest groups of naturally occurring circular permutants, although many types of protein are found to have such variants (Lo *et al.*, 2009). PGABN is clearly an offshoot of the CE4 family, as shown by the close similarity of the CE4 structures, and the fact that the N- and C-termini of PGABN are close in space, roughly 10 Å apart, a distance which can easily be bridged by a few amino-acid residues. It seems unlikely that a permuted variant of PGABN with the connectivity of the parent family would have markedly different properties from the wild-type protein. Circular permutation has also been noted to play a role in the evolution of calcium-dependent carbohydrate-binding modules involved in xylan recognition (Montanier *et al.*, 2010).

The very high level of conservation shown between PGABN and the equivalent domain of HmsF from *Yersinia* is unequivocal evidence of gene transfer between species. Codon-usage analysis implies that the gene has been introduced into *E. coli*. Whereas *Yersinia* apparently requires biofilm for its preferred method of infection, blocking the gut of fleas to cause them to expel the bacteria into the bloodstream of an animal, exopolysaccharide clearly plays a very different role in a bacterium found in the soil and the gut of higher animals. The gene used in the work described in this paper involved an attenuated laboratory strain of *E. coli*. The *hms* operon of *Yersinia* is named for haemin storage and is found within the *pgm* (pigmentation) locus, a 102 kb long region associated with colouration and iron uptake. The *hmsHFRS* operon was originally shown to be required for the haemin-storage phenotype, but subsequently it was found that an extra gene *hmsT* was also required which lies far from the *hmsHFRS* operon (outside the *pgm* locus) on the genome of *Y. pestis* (Jones *et al.*, 1999). This extra gene is not required for biofilm formation and is not found in *Y. enterolitica*. *E. coli* strain MG1655 has homologues for all five genes (*ycdSRQPT*), but does not show a *hms*$^+$ phenotype. Only *ycdQ* and *ycdP* from *E. coli* MG1655 (equivalent to *pgaC* and *pgaD* in *E. coli* K-12) complement mutations in *hmsR* and *hmsS* in *Y. pestis* (Jones *et al.*, 1999).

Biofilm research is an extremely active and growing field, driven by both medical and biotechnological goals, and the mucoid phenotype of *P. aeruginosa* is a much-studied model (Franklin *et al.*, 2011). This organism causes chronic lung infections in cystic fibrosis patients and produces an exopolysaccharide layer of alginate, a linear polymer of 1,4-linked β-D-mannouronic acid and its C5 epimer α-L-glucuronic acid. The structures of the alginate-export proteins clearly resemble those involved in PGA export in *E. coli* (Keiski *et al.*, 2010; Whitney *et al.*, 2011), raising the possibility that a similar strategy may be employed to tackle biofilm production by both microorganisms. After this paper was reviewed, the structure of PgaB was reported independently by the Howell group (Little, Poloczek *et al.*, 2012).

The authors are extremely grateful to the staff of the Photon Factory for the provision of beamtime in the aftermath of the tsunami disaster of 2011.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Agladze, K., Wang, X. & Romeo, T. (2005). *J. Bacteriol.* **187**, 8237–8246.

Blair, D. E., Schüttelkopf, A. W., MacRae, J. I. & van Aalten, D. M. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 15429–15434.
Blair, D. E. & van Aalten, D. M. (2004). *FEBS Lett.* **570**, 13–19.
Bobrov, A. G., Kirillina, O., Forman, S., Mack, D. & Perry, R. D. (2008). *Environ. Microbiol.* **10**, 1419–1432.
Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
Caufrier, F., Martinou, A., Dupont, C. & Bouriotis, V. (2003). *Carbohydr. Res.* **338**, 687–692.
Cerca, N., Maira-Litrán, T., Jefferson, K. K., Grout, M., Goldmann, D. A. & Pier, G. B. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 7528–7533.
Chudzik, D. M., Michels, P. A., de Walque, S. & Hol, W. G. J. (2000). *J. Mol. Biol.* **300**, 697–707.
Costerton, J. W., Stewart, P. S. & Greenberg, E. P. (1999). *Science*, **284**, 1318–1322.
Darby, C. (2008). *Trends Microbiol.* **16**, 158–164.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Franklin, M. J., Nivens, D. E., Weadge, J. T. & Howell, P. L. (2011). *Front. Microbiol.* **2**, 167.
Goller, C., Wang, X., Itoh, Y. & Romeo, T. (2006). *J. Bacteriol.* **188**, 8022–8032.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. (2004). *Nature Rev. Microbiol.* **2**, 95–108.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T. J., Preston, J. F. III & Romeo, T. (2008). *J. Bacteriol.* **190**, 3670–3680.
Itoh, Y., Wang, X., Hinnebusch, B. J., Preston, J. F. III & Romeo, T. (2005). *J. Bacteriol.* **187**, 382–387.
John, M., Röhrig, H., Schmidt, J., Wieneke, U. & Schell, J. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 625–629.
Jones, H. A., Lillard, J. W. & Perry, R. D. (1999). *Microbiology*, **145**, 2117–2128.
Kafetzopoulos, D., Thireos, G., Vournakis, J. N. & Bouriotis, V. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 8005–8008.
Keiski, C. L., Harwich, M., Jain, S., Neculai, A. M., Yip, P., Robinson, H., Whitney, J. C., Riley, L., Burrows, L. L., Ohman, D. E. & Howell, P. L. (2010). *Structure*, **18**, 265–273.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Little, D. J., Poloczek, J., Whitney, J. C., Robinson, H., Nitz, M. & Howell, P. L. (2012). *J. Biol. Chem.* **287**, 31126–31137.
Little, D. J., Whitney, J. C., Robinson, H., Yip, P., Nitz, M. & Howell, P. L. (2012). *Acta Cryst.* F**68**, 842–845.
Lo, W.-C., Lee, C.-C., Lee, C.-Y. & Lyu, P.-C. (2009). *Nucleic Acids Res.* **37**, D328–D332.
Mack, D., Fischer, W., Krokotsch, A., Leopold, K., Hartmann, R., Egge, H. & Laufs, R. (1996). *J. Bacteriol.* **178**, 175–183.
Montanier, C., Flint, J. E., Bolam, D. N., Xie, H., Liu, Z., Rogowski, A., Weiner, D. P., Ratnaparkhe, S., Nurizzo, D., Roberts, S. M., Turkenburg, J. P., Davies, G. J. & Gilbert, H. J. (2010). *J. Biol. Chem.* **285**, 31742–31754.
Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). *Methods Enzymol.* **374**, 229–244.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S., Prates, J. A., Puchart, V., Ferreira, L. M., Fontes, C. M., Biely, P. & Davies, G. J. (2006). *J. Biol. Chem.* **281**, 10968–10975.
Torda, A. E., Procter, J. B. & Huber, T. (2004). *Nucleic Acids Res.* **32**, W532–W535.
Vuong, C., Kocianova, S., Voyich, J. M., Yao, Y., Fischer, E. R., DeLeo, F. R. & Otto, M. (2004). *J. Biol. Chem.* **279**, 54881–54886.
Vuong, C., Voyich, J. M., Fischer, E. R., Braughton, K. R., Whitney, A. R., DeLeo, F. R. & Otto, M. (2004). *Cell. Microbiol.* **6**, 269–275.
Wang, X., Dubey, A. K., Suzuki, K., Baker, C. S., Babitzke, P. & Romeo, T. (2005). *Mol. Microbiol.* **56**, 1648–1663.
Wang, X., Preston, J. F. III & Romeo, T. (2004). *J. Bacteriol.* **186**, 2724–2734.
Whitney, J. C., Hay, I. D., Li, C., Eckford, P. D., Robinson, H., Amaya, M. F., Wood, L. F., Ohman, D. E., Bear, C. E., Rehm, B. H. & Howell, P. L. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 13083–13088.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yu, Y. & Lutz, S. (2011). *Trends Biotechnol.* **29**, 18–25.

PLOS ONE

VCP Is an Integral Component of a Novel Feedback Mechanism that Controls Intracellular Localization of Catalase and H_2O_2 Levels

Katsuhiro Murakami[1], Yuzuru Ichinohe[1], Masaaki Koike[1], Norio Sasaoka[1], Shun-ichiro Iemura[2], Tohru Natsume[2], Akira Kakizuka[1]*

1 Laboratory of Functional Biology, Kyoto University Graduate School of Biostudies, Sakyo-ku, Kyoto, Japan, **2** National Institutes of Advanced Industrial Science and Technology, Biological Information Research Center (JBIRC), Kohtoh-ku, Tokyo, Japan

Abstract

Catalase is a key antioxidant enzyme that catalyzes the decomposition of hydrogen peroxide (H_2O_2) to water and oxygen, and it appears to shuttle between the cytoplasm and peroxisome via unknown mechanisms. Valosin-containing protein (VCP) belongs to the AAA class of ATPases and is involved in diverse cellular functions, e.g. cell cycle and protein degradation, etc. Here we show that VCP and PEX19, a protein essential for peroxisome biogenesis, interact with each other. Knockdown of either VCP or PEX19 resulted in a predominantly cytoplasmic redistribution of catalase, and loss of VCP ATPase activity also increased its cytoplasmic redistribution. Moreover, VCP knockdown decreased intracellular ROS levels in normal and H_2O_2-treated cells, and an oxidation-resistant VCP impaired the ROS-induced cytoplasmic redistribution of catalase. These observations reveal a novel feedback mechanism, in which VCP can sense H_2O_2 levels, and regulates them by controlling the localization of catalase.

Citation: Murakami K, Ichinohe Y, Koike M, Sasaoka N, Iemura S-i, et al. (2013) VCP Is an Integral Component of a Novel Feedback Mechanism that Controls Intracellular Localization of Catalase and H_2O_2 Levels. PLoS ONE 8(2): e56012. doi:10.1371/journal.pone.0056012

Editor: Junji Yodoi, Institute for Virus Research, Laboratory of Infection and Prevention, Japan

Received October 2, 2012; **Accepted** January 4, 2013; **Published** February 14, 2013

Copyright: © 2013 Murakami et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported in part by research grants from Solution Oriented Research for Science and Technology (SORST) of Japan Science and Technology Agency (JST), the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and the Ministry of Health, Labour, and Welfare of Japan. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kakizuka@lif.kyoto-u.ac.jp

Introduction

Reactive oxygen species (ROS), e.g. superoxide radicals, hydrogen peroxide, etc., are natural byproducts of the aerobic metabolism of foods, and they have been shown to play important roles in several physiological functions, e.g. transcriptional regulation, mitogen signaling, integrin signaling, Wnt signaling, etc. (see refs in [1–3]). On the other hand, ROS are also produced by UV and X-ray exposure or inflamation, and excess ROS can damage cellular functions by oxidizing proteins, lipids, and DNA, leading to cell aging as well as cancer (see refs in [4]). Among ROS, the superoxide radical is enzymatically converted by superoxide dismutases (SODs) to hydrogen peroxide (H_2O_2), which, in turn, is converted by catalase or peroxidases to H_2O and O_2. Mammalian cells typically possess three SODs, several peroxidases, and one catalase. Among these ROS-scavenging enzymes, only catalase resides in peroxisomes. In certain conditions, such as aging, catalase also resides in the cytoplasm [5], which is believed to be due to its weak peroxisome-targeting signal (PTS). Two types of PTS, PTS1 and PTS2, are known [6]. Typically, PTS1 consists of three sequential amino acids, SKL, and it is present in peroxisome-localized proteins such as peroxisomal Acyl-CoA thioesterase, PTE1. Catalase has a unique PTS1, consisting of four sequential amino acids, KANL. Both PTS1s are recognized by PEX5 (Peroxisome biogenesis factor 5); however, PEX5 binds to SKL more strongly than to KANL [7], and thus it is believed that PEX5 can transfer SKL-containing proteins more effectively than catalase to peroxisomes. In aged cells, cellular levels of ROS increase, and it is thought that such ROS may weaken PEX5 functions, with transport of catalase to peroxisomes being preferentially compromised, as opposed to transport of SKL-possessing proteins [8]. However, no clear evidence has been provided supporting this speculation.

VCP belongs to the AAA class of ATPases and has been shown to function in many cellular events, including ERAD (endoplasmic reticulum-associated degradation), cell cycle control, membrane fusion, maintenance of Golgi apparatus, protein aggregate formation and clearance, etc. (see refs in [9]). VCP has also been shown to play important roles in several human neurodegenerative disorders [10–12]. We have shown that VCP is modified post-translationally at 60 amino acids, at least, including 18 serines, 14 threonines, 6 tyrosines, and 22 lysines [13]. To investigate the role of post-transcriptional modifications of VCP, we created several modification-mimic forms of VCP, and characterized them [12–14]. These analyses have revealed novel VCP functions and have led us to speculate that VCP may have unidentified functions. In this study, we report a novel VCP function in regulation of intracellular H_2O_2 levels via the control of catalase localization.

Materials and Methods

Antibodies

The following antibodies were purchased: anti-actin (Chemicon), anti-catalase (Calbiochem), anti-PTE1 (ACOT8) (Santa Cruz), anti-FLAG M2 (Sigma), anti-PMP70 (Zymed), anti-HA (Santa Cruz), and anti-PEX19 (BD PharMingen). The affinity-purified rabbit polyclonal anti-VCP antibody was described previously [15].

Plasmids and siRNAs

The cDNAs for PEX5, PEX19, catalase, and PTS2 signal sequences of ACAA1 (acetyl-CoA acyltransferase 1) were amplified by RT–PCR from total RNA isolated from HeLa cells, and their sequences were confirmed. The VCP cDNAs (wtVCP, VCP[K251A], VCP[K524A] [16]) or PEX5 cDNA was subcloned into pmCherry vector (Clontech). The PEX19 cDNA was subcloned into pCMV-HA vector (Clontech).

The targeting sequences of siRNAs for VCP and PEX5 mRNAs were as follows:

VCP(nc), 5′-CGGGAGAGGCGCGCGCCAT-3′;
VCP(286), 5′-GGTTAATTGTTGATGAAGCCATCAA-3′;
PEX5(192), 5′-CAAGCCTTTGGGAGTAGCTTCTGAA-3′;
PEX5(955), 5′-GACCTTACGTCAGCTACCTATGATA-3′.
Control, 5′-CGGACGCGTCAGGAGCCGGTT-3′.

The siRNAs for PEX19 were purchased from Invitrogen (Stealth Select RNAi, HSS108913 and HSS108914, respectively).

Cell Culture and Cell Lines

HeLa cells and HEK293A cells were grown at 37°C in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. HeLa cell lines stably expressing organelle-targeted GFPs were created by transfection of organelle-targeted GFP expression vectors, and selected in the presence of 2.5 μg/ml of puromycin (Invivogen). The HEK293A cell line stably expressing GFP-catalase, was also created by similar methods.

Transfection and Immunostaining

Plasmid transfection was carried out using Lipofectamine plus (Invitrogen), and siRNA transfection was carried out using Oligofectamine (Invitrogen) according to the manufacturer's protocol. In co-transfection experiments, cells were transfected with siRNA and plasmid using Lipofectamine 2000 (Invitrogen).

Cells were fixed with 4% formaldehyde for 10 min at room temperature. Fixed cells were permeabilized with 0.5% Triton X-100 in PBS for 10 min at room temperature and blocked with blocking buffer (0.1% bovine serum albumin and 0.1% skim milk in PBS) for 1 h. Cells were then incubated 1 h at room temperature with primary antibodies. Subsequently, cells were treated with Alexa Fluor 488-conjugated secondary antibodies (Invitrogen). To detect PMP70, fixed cells were permeabilized with 25 μg/ml digitonin in PBS for 5 min at room temperature, and cells were processed for immunostaining as describe above.

Subcellular Fractionation and Immunoprecipitation

Cells were fractionated into cytosol, membrane/organelle, and nucleus, using a Subcellular Proteome extraction kit (Calbiochem), according to the manufacturer's protocol. Immunoprecipitation assays were performed as described previously [16]. Briefly, samples were lysed on ice and debris was removed by centrifugation for 30 min at 15,000×g at 4°C. The supernatant was mixed with an anti-HA or anti-FLAG antibody and rotated at 4°C overnight after addition of protein G-Sepharose beads (Amersham Biosciences). After washing of beads, bound proteins were analyzed by Western blot.

Intracellular ROS Detection

Cells were washed twice with HBSS and incubated with 5 μM CM-H_2DCFDA, a ROS-detection reagent (Invitrogen), in HBSS at 37°C for 30 min. Subsequently, cells were washed twice with HBSS and incubated with growth medium at 37°C for 30 min with or without H_2O_2. Then cells were analyzed by FACScan flow cytometer (BD Biosciences) or LSM510 confocal microscopy (Carl Zeiss).

Statistical Analysis

Each experiment was conducted at least three times with consistent results. The gel or blot representative of each experiment is presented in this study. The statistical significance was analyzed using Student's *t* test.

Results

Involvement of VCP in Intracellular Localization of Catalase

In order to visualize organelle in live cells, we generated several HeLa cell sublines in which GFP was expressed as a fusion protein with a peroxisomal (PTS1 or PTS2)-, nuclear (NLS)-, ER (KDEL)-, or mitochondrial (mito)-targeting signal. Among these, we observed clear mislocalization of GFP-PTS1 (namely, GFP-SKL and GFP-KANL) into the cytoplasm when the cells were treated with VCP siRNAs but not a control siRNA. VCP siRNAs perturbed GFP-KANL localization much more severely than GFP-SKL localization (**Fig. 1A and B**). By contrast, VCP siRNAs did not induce clear mislocalization of PTS2-GFP, mito-GFP, GFP-ER, or GFP-NLS (Fig. S1). We observed similar mislocalization of GFP-KANL by expressing ATPase-negative or dominant-negative VCP mutants, e.g. VCP[K251A] and VCP[K524A] [16] (Fig. S2). In addition, treating cells with DBeQ, a VCP inhibitor [17], also induced cytoplasmic localization of GFP-catalase (**Fig. 1C**). These results suggest that the ATPase activity of VCP is necessary for proper localization of catalase.

We then examined the effects of VCP knockdown on intracellular localization of endogenous PTE1 or catalase. In more than 50% of cells treated with VCP siRNAs, endogenous catalase was diffusely observed in the cytoplasm. In contrast, PTE1 localization was not apparently affected by VCP knockdown (**Fig. 1D**). These results were confirmed by cell fractionation experiments. VCP siRNA treatments increased the amounts of catalase but not PTE1 in the cytoplasmic fraction (**Fig. 1E**). Mislocalization of catalase as well as GFP-KANL decreased in cells treated with VCP siRNA together with cycloheximide (Fig. S3), supporting the idea that newly synthesized catalase is transported into peroxisomes with the help of VCP.

Interaction between VCP and PEX19

In order to obtain insights for molecular mechanisms underlying VCP-mediated regulation of catalase localization, we searched for VCP-interacting proteins using an immunoprecipitation method followed by a very sensitive MS/MS analysis [18], and identified PEX19 as a potential VCP-interacting protein. Indeed, we could observe a physical association between VCP and PEX19 via immunoprecipitation and western blotting (**Fig. 2**). This interaction appeared very weak, suggesting the possibility that yet-unknown VCP modification may enhance this interaction. This possibility remained to be clarified.

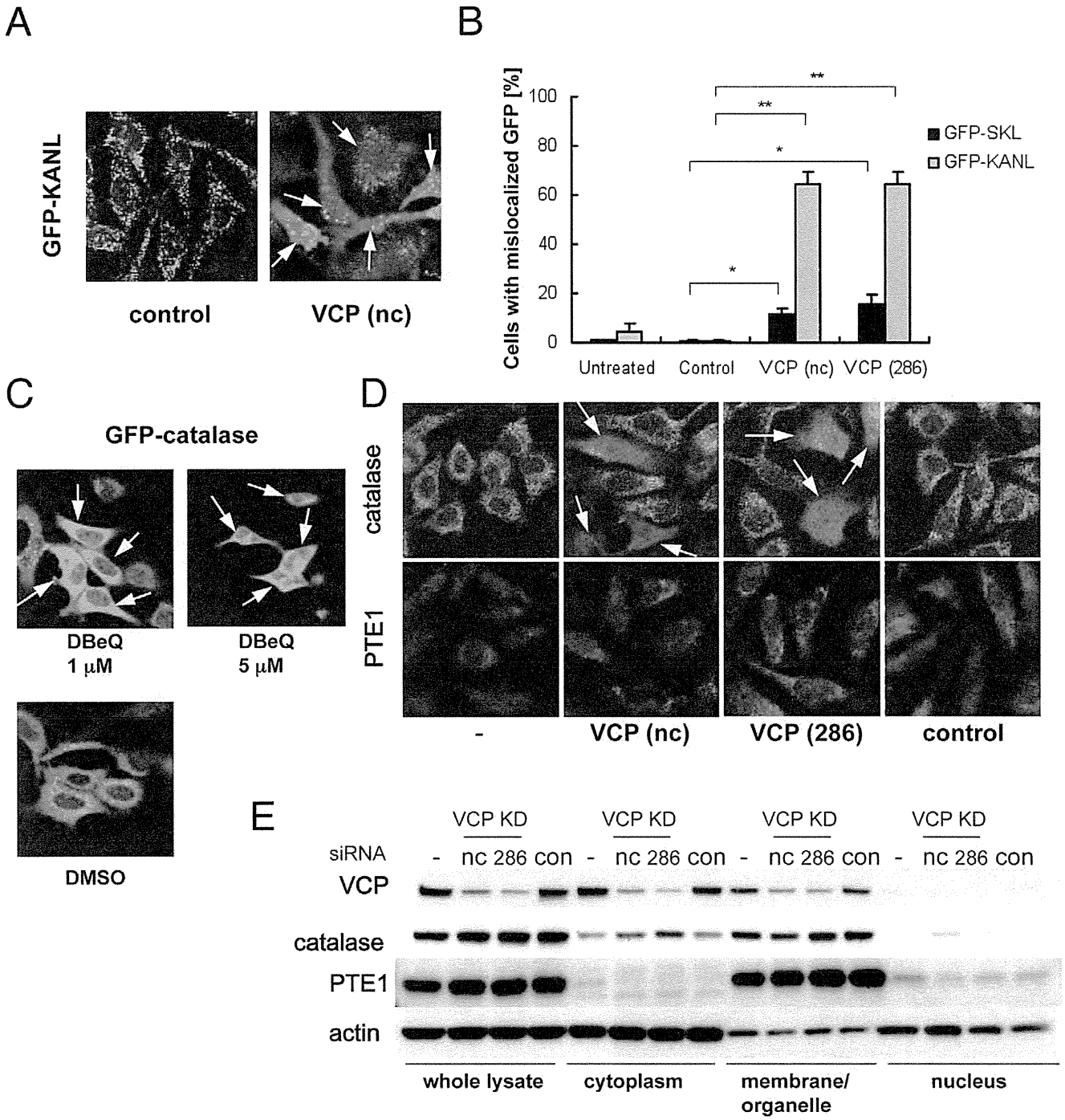

Figure 1. VCP siRNAs induce cytoplasmic localization of catalase. (A) Cytochemical analysis of intracellular localization of GFP-KANL. HeLa cells continuously expressing GFP-KANL were treated with control siRNA (control) or VCP siRNAs (nc and 286). Seventy-two hours later, GFP images were analyzed by confocal microscopy. Arrows indicate cells with cytoplasmic localization of GFP-KANL. (B) Quantification of cytochemical analysis on GFP-KANL in (A) and on GFP-SKL. More than 200 cells were examined in each sample, and the fraction (%) of cells with diffuse GFP signals in the cytoplasm were scored. $^{**}p<0.01$, $^{*}p<0.05$. (C) Fluorescence microscopy analysis of intracellular localization of GFP-catalase. HEK293A cells continuously expressing GFP-catalase were treated with 1 μM or 5 μM DBeQ, a VCP inhibitor [17], or DMSO for 24 hours, and then GFP signals were detected. Arrows indicate cells with cytoplasmic localization of GFP-catalase. (D) Immunocytochemical analysis of intracellular localization of catalase and PTE1. HeLa cells were treated without (−) or with control siRNA (control), or VCP siRNAs (nc and 286). Seventy-two hours later, catalase and PTE1 were detected with anti-catalase and anti-PTE1 antibodies, respectively. Arrows indicate cells with cytoplasmic localization of catalase. (E) Western blot analyses of protein levels of VCP, catalase, and PTE1 in different cell compartments. HeLa cells were treated without (−) or with control siRNA (control) or VCP siRNAs (nc and 286). Seventy-two hours later, cells were fractionated as described in Methods. Fractionated samples equivalent to 7.5 μg total protein of whole cell lysates were separated by SDS-PAGE and analyzed by western blotting using specific antibodies. Actin served as a loading control.

doi:10.1371/journal.pone.0056012.g001

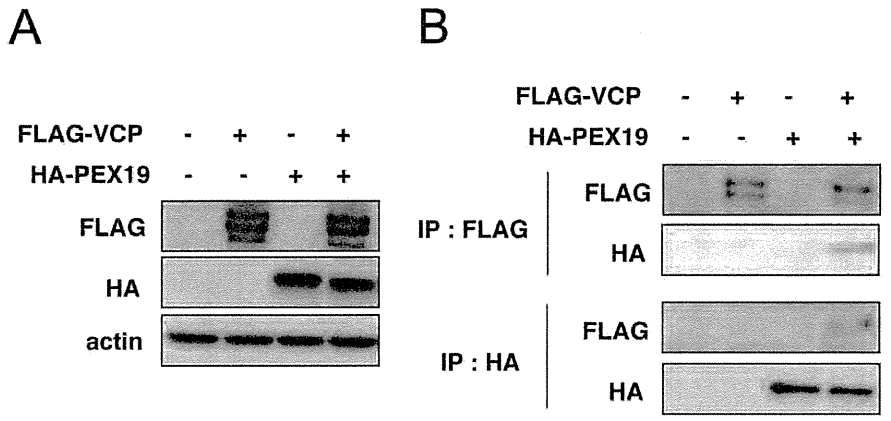

Figure 2. Immunoprecipitation assays to detect physical interactions between VCP and PEX19. (A) HEK293A cells were transfected with expression vectors for FLAG-VCP and HA-PEX19. Twenty-four hours later, cells were harvested and the cell lysates were analyzed by western blots with antibodies indicated in the panels. Actin served as a loading control. (B) The immunoprecipitation was performed on the cell lysates in (A) with an anti-FLAG or anti-HA antibody. The precipitates were analyzed by western blots with antibodies indicated in the panels. See details in **Materials and methods.**
doi:10.1371/journal.pone.0056012.g002

Involvement of PEX19 in Intracellular Localization of Catalase

Given that VCP could potentially make a complex with PEX19 and that VCP knockdown apparently affected the transport of catalase into peroxisomes, PEX19 knockdown could also affect the intracellular localization of catalase. Indeed, PEX19 knockdowns produced virtually identical distributions of intracellular catalase as were observed in VCP knockdowns (**Fig. 3A and B**). Moreover, in PEX19 knockdown cells, PTE1 localization was not apparently affected (**Fig. 3A and B**). PEX19 is reportedly involved in the transport of membrane proteins, such as PMP70 (peroxisome membrane protein 70), to peroxisomes. However, we could not detect any clear mislocalization of endogenous PMP70 in VCP-depleted HeLa cells (Fig. S4).

Consistent with previous reports, PEX5 knockdown induced mislocalization of both catalase and PTE1 (Fig. S5). It is notable that over-expression of PEX5 and VCP could not rectify the mislocalization of GFP-KANL in VCP and PEX5 knockdown cells, respectively (Fig. S6). These results indicate that VCP/PEX19 complexes are required for PEX5 to transport catalase, but not other typical PTS1- or PTS2-possessing proteins, to peroxisomes.

VCP Activity, Catalase Localization, and ROS Levels

We next examined the possibility that VCP-depleted cells have a greater capacity to scavenge H_2O_2 as compared with non-treated cells, due to the presence of catalase in the cytoplasm. This was indeed the case. Basal ROS levels were reduced in cells treated with VCP siRNAs compared to those treated with control siRNAs (**Fig. 4A**). Reduction of ROS levels was more pronounced when cells were treated with H_2O_2 (**Fig. 4A and B**). We have previously shown that the ATPase activity of VCP was inactivated

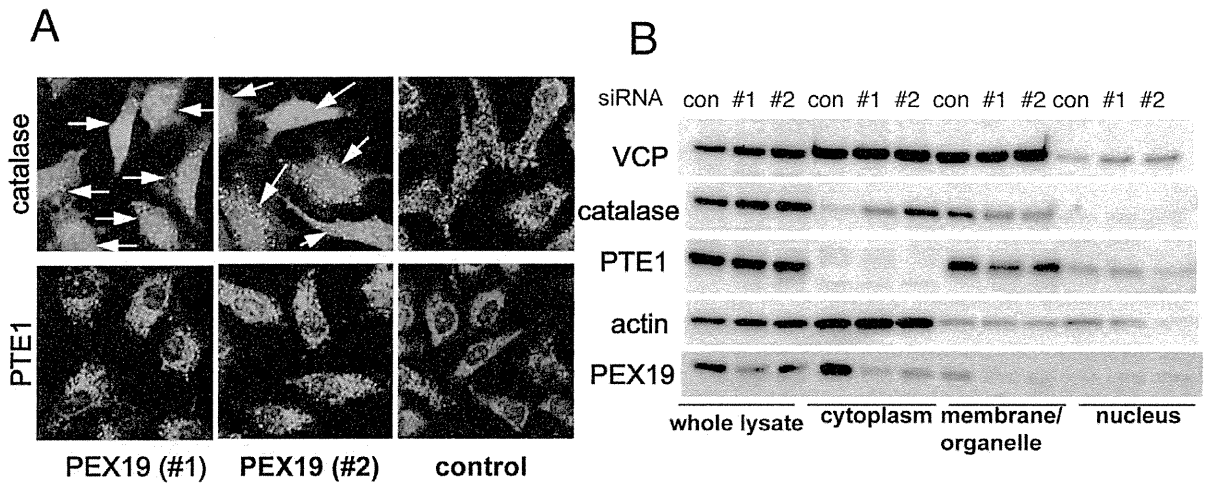

Figure 3. PEX19 siRNAs also induce cytoplasmic localization of catalase. (A) Immunocytochemical analysis of intracellular localization of catalase and PTE1. HeLa cells were treated with control siRNA (control) or PEX19 siRNAs (#1; HSS108913 and #2; HSS108914). Seventy-two hours later, catalase and PTE1 were detected as in (Fig. 1D). Arrows indicate cells with cytoplasmic localization of catalase. (B) Western blot analyses of protein levels of VCP, catalase, PTE1, and PEX19 in different cell compartments. HeLa cells were treated with control siRNA (control) or PEX19 siRNAs (#1; HSS108913 and #2; HSS108914). Seventy-two hours later, cells were analyzed as in (Fig. 1E).
doi:10.1371/journal.pone.0056012.g003

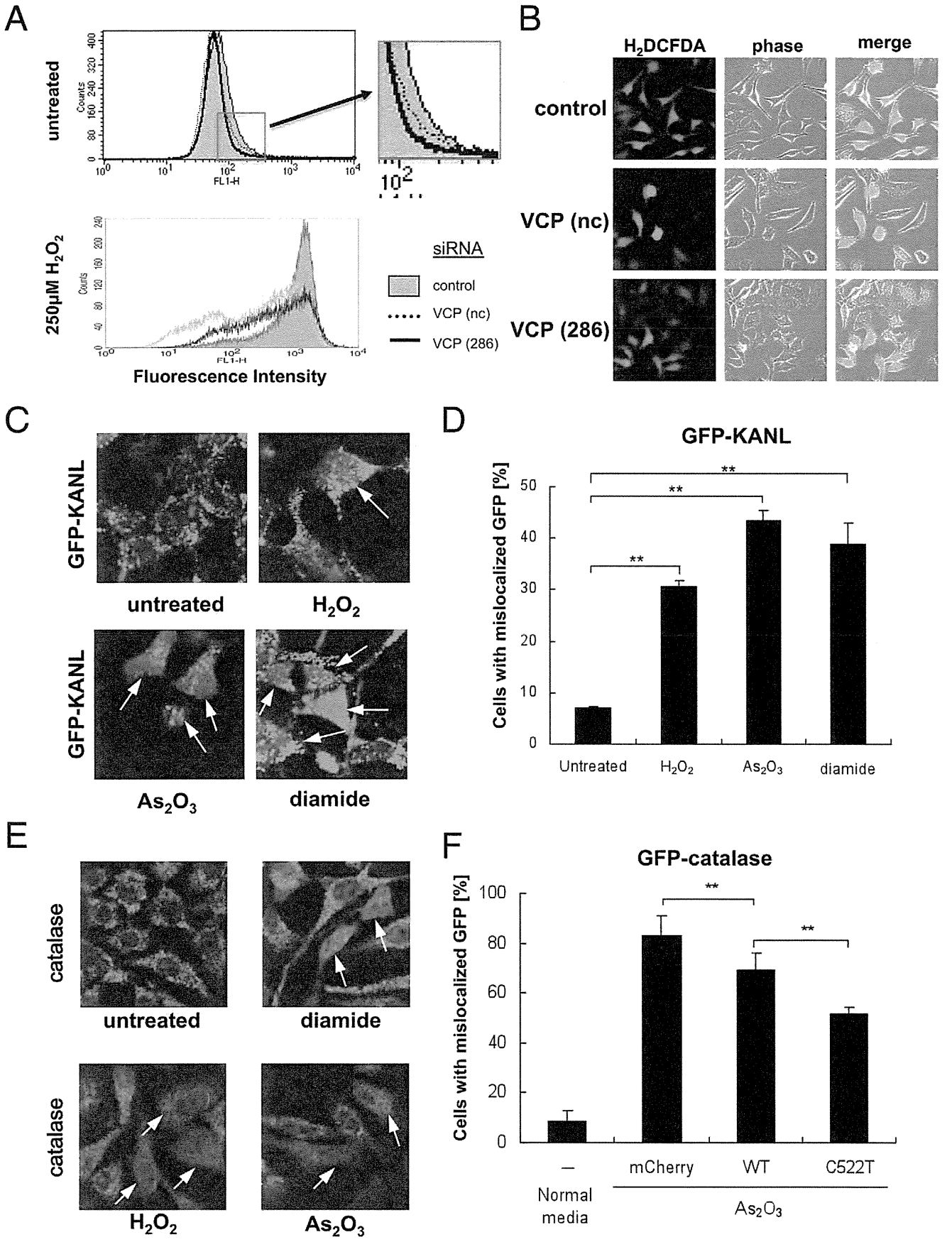
A
untreated
Counts
FL1-H
250μM H_2O_2
siRNA
control
VCP (nc)
VCP (286)
Fluorescence Intensity
B
H_2DCFDA
phase
merge
control
VCP (nc)
VCP (286)
C
GFP-KANL
untreated
H_2O_2
As_2O_3
diamide
D
GFP-KANL
Cells with mislocalized GFP [%]
Untreated
H_2O_2
As_2O_3
diamide
E
catalase
untreated
diamide
H_2O_2
As_2O_3
F
GFP-catalase
Cells with mislocalized GFP [%]
Normal media
mCherry
WT
C522T
As_2O_3

Figure 4. ROS levels and intracellular localization of GFP-KANL and catalase in cells with VCP depletion or overexpression. (A) ROS levels in cells with VCP depletion. HeLa cells were treated with control (control) or VCP (nc and 286) siRNAs. Seventy-two hours later, 5 μM H_2DCFDA was added to the media for 30 min. Then, cells were treated with or without 250 μM H_2O_2 at 37°C for additional 30 min, and analyzed by FACS (see details in **Materials and methods**). (B) Cytochemical analysis of HeLa cells treated with 250 μM H_2O_2. H_2DCFDA fluorescence was analyzed by confocal microscopy as in (A). (C) Cytochemical analysis of intracellular localization of GFP-KANL after treatment of ROS-producing agents. HEK293A cells continuously expressing GFP-KANL were treated with 250 μM H_2O_2, 20 nM As_2O_3, or 250 μM diamide. Twenty-four hours later, GFP images were analyzed by confocal microscopy. Arrows indicate cells with cytoplasmic localization of GFP-KANL. (D) Quantification of cytochemical analysis in (C). More than 200 cells were examined in each sample, and the fraction (%) of cells with diffuse GFP signals in the cytoplasm were scored. $^{**}p<0.01$. (E) Immunocytochemical analysis of intracellular localization of catalase after treatment with ROS-producing agents. HeLa cells were treated with 500 μM H_2O_2, 20 nM As_2O_3, or 250 μM diamide. Twenty-four hours later, catalase was detected with an anti-catalase antibody. Arrows indicate cells with cytoplasmic localization of catalase. (F) Overexpression of VCP[C522T] weakened cytoplasmic localization of catalase by ROS more significantly than that of wild-type VCP. HEK293A cells continuously expressing GFP-catalase were transfected with VCP[C522T]-mCherry (C522T) or wild-type VCP-mCherry (wtVCP), and treated with 20 nM As_2O_3 for 24 hours. More than 200 mCherry-positive cells were examined in each sample, and the fraction (%) of cells with diffuse GFP signals in the cytoplasm were scored. $^{**}p<0.01$.
doi:10.1371/journal.pone.0056012.g004

by oxidation of Cys522 by ROS, such as H_2O_2 [14]. The observation that ATPase activity of VCP is necessary for proper catalase localization to peroxisomes raised the possibility that ROS treatments would also induce redistribution of catalase. We next examined this possibility, and confirmed that all tested ROS-inducing agents (such as H_2O_2, As_2O_3, and diamide) induced cytoplasmic localization of GFP-KANL as well as catalase (**Fig. 4C–E**). We then examined whether VCP[C522T], a VCP mutant with the ROS-sensitive cysteine to threonine substitution [14], had protective effects on redistribution of catalase in cells treated with ROS. Indeed, overexpression of VCP[C522T] significantly inhibited cytoplasmic localization of catalase in cells treated with As_2O_3, compared to overexpressed wild-type VCP (**Fig. 4F**).

Discussion

The results presented in this study, taken together, point to the existence of a novel feedback mechanism: when H_2O_2 levels increase, VCP ATPase is inactivated by Cys522 oxidation, which in turn keeps catalase in the cytoplasm, leading to reduced H_2O_2 levels. After H_2O_2 levels are reduced, glutathione as well as thioredoxine levels would recover, which would then restore VCP ATPase activity, leading to catalase transport into peroxisomes. This VCP-mediated system has the great merit of specifically changing the localization of catalase without affecting the localization of other peroxisome proteins.

In *S. cerevisiae*, Cys522 is not conserved in Cdc48p, a VCP homologue [14]. In what appears to be an evolutionary alternative design, *S. cerevisiae* possesses two catalases, one of which resides in peroxisomes and the other in the cytoplasm [19,20]. *C. elegans* also possesses two catalases, one in peroxisomes and the other in the cytoplasm [21]. These lines of evidence strongly indicate that for living organisms catalase is needed in both peroxisomes and the cytoplasm. In mammals, a certain level of ROS, namely H_2O_2, is utilized in several physiological conditions, and, therefore the continuous presence of catalase in the cytoplasm might not be favorable. On the other hand, when mammalian cells meet conditions with a large amount of H_2O_2 in the cytoplasm, catalase would more effectively degrade and reduce H_2O_2 by accumulating in the cytoplasm. Thus, mammals have developed an integrated system to utilize one catalase rather than to have two differently localized catalases.

Supporting Information

Figure S1 Fluorescence microscopy analysis of intracellular localization of PTS2-GFP, mito-GFP, GFP-ER, and GFP-NLS. (A) Schematic drawings of GFP-fused proteins. (B) HeLa cells were treated with control siRNA (control) or VCP siRNAs (nc and 286). Seventy-two hours later, GFP signals were detected.
(TIFF)

Figure S2 Fluorescence microscopy analysis of intracellular localization of GFP-KANL in the presence of ATPase activity-defective mutant VCPs. (A) HEK293A cells continuously expressing GFP-KANL were transfected with an expression vectors for mCherry or VCP (wtVCP, VCP[K251A] [15], or VCP[K524A] [15])-mCherry. Forty-eight hours later, GFP signals were detected. (B) Quantification of fluorescence microscopy of GFP-KANL in (A). More than 200 mCherry-positive cells were examined in each sample, and the fraction (%) of cells with diffuse GFP signals in the cytoplasm were scored. $^{**}p<0.01$, $^{*}p<0.05$.
(TIFF)

Figure S3 Immunocytochemical and fluorescence microscopy analyses of intracellular localization of catalase and GFP-KANL. (A) HeLa cells were treated with control siRNA (control) or VCP siRNA (286) for 72 hours, and treated with or without cyclohexamide (CHX) (5 μg/ml) for additional 24 hours. Then catalase was detected with anti-catalase antibody. (B) HeLa cells continuously expressing GFP-KANL were treated with VCP siRNA (286). Cells were treated with or without 5 μg/ml of CHX from 48 (24 h) or 24 (48 h) to 72 hours after siRNA treatment. Then, GFP signals were detected.
(TIFF)

Figure S4 Immunocytochemical analysis of intracellular localization of PMP70. HeLa cells were treated without (−) or with control siRNA (control), or VCP siRNAs (nc and 286). Seventy-two hours later, PMP70 was detected with an anti-PMP70 antibody. Note that VCP protein levels decreased by VCP siRNA treatments, as shown in Fig. 1E
(TIFF)

Figure S5 Immunocytochemical analysis of intracellular localization of catalase and PTE1. HeLa cells were treated with control siRNA (control) or PEX5 siRNAs (192 and 955). Seventy-two hours later, catalase and PTE1 were detected with anti-catalase and anti-PTE1 antibodies, respectively.
(TIFF)

Figure S6 Fluorescence microscopy analysis of intracellular localization of GFP-KANL. (A) HeLa cells continuously expressing GFP-KANL were treated with control siRNA (control), VCP siRNA (nc), or PEX5 siRNA (192) for 48 hours, and then transfected with an expression vector for mCherry, VCP-mCherry, or mCherry-PEX5. Twenty-four hours later, GFP signals (green) and mCherry signals (red) were examined. (B) Quantification of fluorescence microscopy of GFP-KANL in (A).

More than 120 mCherry-positive cells were examined in each sample, and the fraction (%) of cells with diffuse GFP signals in the cytoplasm were scored. n.s., not significant.
(TIFF)

Acknowledgments

We thank Professor James A. Hejna (Kyoto University) for critical reading of the manuscript.

References

1. Vurusaner B, Poli G, Basaga H (2012) Tumor suppressor genes and ROS: complex networks of interactions. Free Radic Biol Med 52: 7–18.
2. Funato Y, Miki H (2010) Redox regulation of Wnt signalling via nucleoredoxin. Free Radic Res 44: 379–388.
3. Hu CT, Wu JR, Cheng CC, Wang S, Wang HT, et al. (2011) Reactive oxygen species-mediated PKC and integrin signaling promotes tumor progression of human hepatoma HepG2. Clin Exp Metastasis 28: 851–863.
4. Valko M, Leibfritz D, Moncol J, Cronin MT, Mazur M, et al. (2007) Free radicals and antioxidants in normal physiological functions and human disease. Int J Biochem Cell Biol 39: 44–84.
5. Legakis JE, Koepke JI, Jedeszko C, Barlaskar F, Terlecky LJ, et al. (2002) Peroxisome senescence in human fibroblasts. Mol Biol Cell 13: 4243–4255.
6. Heiland I, Erdmann R (2005) Biogenesis of peroxisomes. Topogenesis of the peroxisomal membrane and matrix proteins. FEBS J 272: 2362–2372.
7. Maynard EL, Gatto GJ Jr, Berg JM (2004) Pex5p binding affinities for canonical and noncanonical PTS1 peptides. Proteins 55: 856–861.
8. Terlecky SR, Koepke JI, Walton PA (2006) Peroxisomes and aging. Biochim Biophys Acta 1763: 1749–1754.
9. Manno A, Noguchi M, Fukushi J, Motohashi Y, Kakizuka A (2010) Enhanced ATPase activities as a primary defect of mutant valosin-containing proteins that cause inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia. Genes Cells 15: 911–922.
10. Kakizuka A (2008) Roles of VCP in human neurodegenerative disorders. Biochem Soc Trans 36: 105–108.
11. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, et al. (2010) Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 68: 857–864.
12. Koike M, Fukushi J, Ichinohe Y, Higashimae N, Fujishiro M, et al. (2010) Valosin-containing protein (VCP) in novel feedback machinery between abnormal protein accumulation and transcriptional suppression. J Biol Chem 285: 21736–21749.
13. Mori-Konya C, Kato N, Maeda R, Yasuda K, Higashimae N, et al. (2009) p97/valosin-containing protein (VCP) is highly modulated by phosphorylation and acetylation. Genes Cells 14: 483–497.
14. Noguchi M, Takata T, Kimura Y, Manno A, Murakami K, et al. (2005) ATPase activity of p97/valosin-containing protein is regulated by oxidative modification of the evolutionally conserved cysteine 522 residue in Walker A motif. J Biol Chem 280: 41332–41341.
15. Hirabayashi M, Inoue K, Tanaka K, Nakadate K, Ohsawa Y, et al. (2001) VCP/p97 in abnormal protein aggregates, cytoplasmic vacuoles, and cell death, phenotypes relevant to neurodegeneration. Cell Death Differ 8: 977–984.
16. Kobayashi T, Tanaka K, Inoue K, Kakizuka A (2002) Functional ATPase activity of p97/valosin-containing protein (VCP) is required for the quality control of endoplasmic reticulum in neuronally differentiated mammalian PC12 cells. J Biol Chem 277: 47358–47365.
17. Chou TF, Brown SJ, Minond D, Nordin BE, Li K, et al. (2011) Reversible inhibitor of p97, DBeQ, impairs both ubiquitin-dependent and autophagic protein clearance pathways. Proc Natl Acad Sci U S A 108: 4834–4839.
18. Natsume T, Yamauchi Y, Nakayama H, Shinkawa T, Yanagida M, et al. (2002) A direct nanoflow liquid chromatography-tandem mass spectrometry system for interaction proteomics. Anal Chem 74: 4725–4733.
19. Cohen G, Rapatz W, Ruis H (1988) Sequence of the Saccharomyces cerevisiae CTA1 gene and amino acid sequence of catalase A derived from it. Eur J Biochem 176: 159–163.
20. Hartig A, Ruis H (1986) Nucleotide sequence of the Saccharomyces cerevisiae CTT1 gene and deduced amino-acid sequence of yeast catalase T. Eur J Biochem 160: 487–490.
21. Togo SH, Maebuchi M, Yokota S, Bun-Ya M, Kawahara A, et al. (2000) Immunological detection of alkaline-diaminobenzidine-negativeperoxisomes of the nematode Caenorhabditis elegans purification and unique pH optima of peroxisomal catalase. Eur J Biochem 267: 1307–1312.

Author Contributions

Conceived and designed the experiments: KM AK. Performed the experiments: KM YI MK NS S-II. Analyzed the data: KM TN AK. Contributed reagents/materials/analysis tools: NS. Wrote the paper: KM AK.