Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structure of *Debaryomyces castellii* CBS 2923 phytase

M. Ragon,[a] F. Hoh,[b]* A. Aumelas,[b] L. Chiche,[b] G. Moulin[a] and H. Boze[a]

[a]UMR IR2B, Equipe Génie Microbiologique et Enzymatique, ENSAM–INRA, 2 Place Viala, 34060 Montpellier CEDEX 01, France, and [b]CNRS UMR 5048, INSERM U554, Université Montpellier 1 and 2, Centre de Biochimie Structurale, 29 Rue de Navacelles, 34090 Montpellier, France

Correspondence e-mail: f.hoh@cbs.cnrs.fr

Received 5 January 2009
Accepted 9 March 2009

PDB Reference: phytase, 2gfi, r2gfisf.

Phytate (*myo*-inositol hexakisphosphate) is the primary storage form of phosphate in seeds and legumes (Reddy *et al.*, 1982). Phytases are phosphatases that hydrolyze phytate to less phosphorylated *myo*-inositol derivatives and inorganic phosphate. The crystal structure of phytase from *Debaryomyces castellii* has been determined at 2.3 Å resolution. The crystals belonged to space group $P6_522$, with unit-cell parameters a = 121.65, c = 332.24 Å. The structure was solved by molecular replacement and refined to a final R factor of 15.7% (R_{free} = 20.9%). The final model consists of a dimer (with two monomers of 458 residues), five NAG molecules and 628 water molecules.

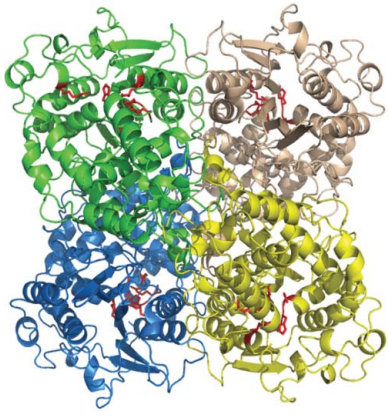
1. Introduction

Phytases (*myo*-inositol hexakisphosphate 3-phosphohydrolases and 6-phosphohydrolases; EC 3.1.3.8 and EC 3.1.3.26) catalyze the release of phosphate from phytic acid (*myo*-inositol hexakisphosphate), the major phosphorus-storage form in plants, including most cereals and legumes. Phytic acid acts as an antinutrient owing to its chelation of various metals and binding of proteins and therefore diminishes the bioavailability of proteins and nutritionally important minerals.

Phytases are produced by a wide range of organisms: plants, animals and especially microorganisms (Pandey *et al.*, 2001; Nakamura *et al.*, 2000). The number of phytases described has increased over the last decade (Konietzny & Greiner, 2002). Most phytases belong to the histidine acid phosphatases (HAPs; Mitchell *et al.*, 1997; Oh *et al.*, 2004). The catalytic histidine is part of the amino-acid sequence motif RHG*X*R*X*P which is characteristic of HAPs. However, not all of the enzymes are structurally similar. Many pathways for the hydrolysis of phytate by phytate-degrading enzymes have been described (Konietzny & Greiner, 2002). Moreover, their biochemical properties (specific activities, broad substrate specificity, broad pH stability and thermostability) are different.

Debaryomyces castellii CBS 2923 phytase (PhytDc) is a glycosylated protein thermostable up to 339 K that hydrolyses the six phytate-bound phosphates (Ragon *et al.*, 2008; Fig. 1). Its activity is observed at pH values between 3 and 7. The primary sequence of PhytDc has 36% identity to that of *Aspergillus niger* acid phosphatase (PAAn), 25% to that of *A. niger* phytase A (PhytAn) and 23% to that of *A. fumigatus* phytase A (PhytAf; Fig. 2).

In this paper, we report the first yeast phytase structure from *D. castellii* CBS 2923 and make a brief comparison with other histidine acid phosphatase structures.



2. Experimental

2.1. Protein expression and purification

D. castellii CBS 2963 cells were grown at 301 K in synthetic MSA-B medium containing 10 g l^{-1} glucose, 0.4 g l^{-1} sodium phytate, 3 g l^{-1} ammonium sulfate, 7.5 mg l^{-1} $MnSO_4.H_2O$, 0.5 g l^{-1} KCl, 0.1 g l^{-1}

$CaCl_2.2H_2O$, 0.5 g l^{-1} $MgSO_4.7H_2O$, 500 µg l^{-1} H_3BO_4, 40 µg l^{-1} $CuSO_4.5H_2O$, 100 µg l^{-1} KI, 200 µg l^{-1} $Na_2MoO_4.2H_2O$, 400 µg l^{-1} $ZnSO_4.7H_2O$, 200 µg l^{-1} $FeCl_3.6H_2O$, 2 mg l^{-1} calcium pantothenate, 2 mg l^{-1} thiamine, 2 mg l^{-1} *myo*-inositol, 2 mg l^{-1} pyridoxin, 500 µg l^{-1} nicotinic acid and 20 µg l^{-1} biotine. Cultures were performed in an Applikon fermentor (The Netherlands) with a useful capacity of 1.5 l. The pH, which was measured using an Ingold probe, was regulated at pH 4 by automatic addition of 2 *M* NaOH or 1 *N* H_2SO_4.

After 24 h growth, the culture supernatant was filtered (0.22 µm cutoff) and concentrated using a tangential flow ultrafiltration system (Filtron, 10 kDa cutoff). Prior to purification, the concentrated supernatant fluid was equilibrated with 2 *M* ammonium sulfate (2–16 h, 277 K) and centrifuged (12 000*g*, 30 min; 277 K). The phytase was purified by hydrophobic interaction chromatography (HiPrep 16/10 Phenyl FF column, internal diameter 16 mm, length 100 mm, Pharmacia) at 293 K. The column had been equilibrated with 50 m*M* Tris–HCl buffer pH 6.1, 2 *M* ammonium sulfate. 1–5 ml concentrated supernatant fluid was loaded onto the column. Unbound proteins were removed by washing with equilibration buffer. Bound proteins were then eluted using a linear gradient of ammonium sulfate (2–1.7 *M* in 50 m*M* Tris–HCl buffer pH 6.1). The phytase was eluted with 1.7 *M* ammonium sulfate in the same buffer. The other bound proteins were then eluted using a linear gradient of ammonium sulfate (1.7–0 *M* in the same buffer). The flow rate was 5 ml min^{-1}. The absorbance was measured at 280 nm. Fractions containing phytase were pooled, desalted by ultrafiltration using a PM-10 membrane (Amicon) and concentrated to 3.7 mg ml^{-1}. The purified protein (3.14 mg) was deglycosylated for 7 h with 8000 units of endo-*β*-*N*-acetylglycosaminidase H (Endo H, New England Biolabs Inc., Beverly, Massachusetts, USA) in 25 m*M* sodium acetate pH 5.5 at 310 K. The protein was dialyzed with a 10 kDa cutoff against 5 m*M* sodium acetate pH 5.5 and concentrated to 11.2 mg ml^{-1} by ultrafiltration. Protein purity was checked using SDS–PAGE.

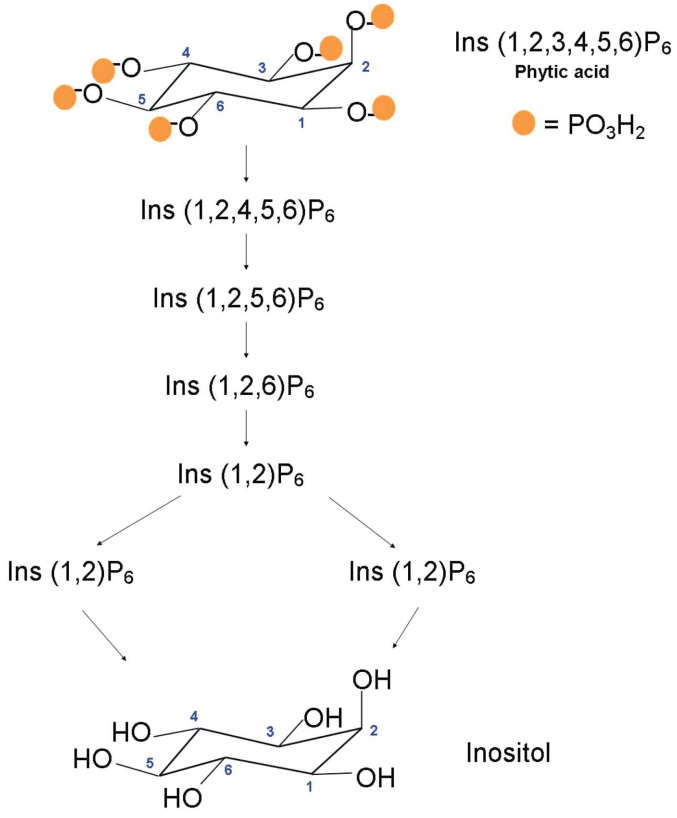


Figure 1
Hydrolysis sequence of phytic acid by PhytDc (Ragon *et al.*, 2008).

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell (2.42–2.3 Å).

Data-collection statistics	
Space group	$P6_522$
Wavelength used (Å)	0.933
Unit-cell parameters (Å)	a = 121.65, c = 332.24
Resolution range (Å)	65.2–2.3
Total No. of observations	320802 (21949)
Total No. unique observations	59592 (7925)
Mean $(I)/\sigma(I)$	11.2 (3.5)
Completeness (%)	94.4 (84.3)
R_{merge} (%)	11.0 (21.9)
Refinement statictics	
Asymmetric unit content	Dimer
R factor (%)	15.7 (16.1)
R_{free} factor (%)	20.9 (23.9)
No. of reflections in working set	59592
No. of reflections in test set	3186
Protein atoms	7195 (916 residues)
Chain *A*	4–461
Chain *B*	4–461
Other atoms	70 (5 NAG)
Water atoms	628
Mean temperature factor (Å^2)	26.4
Matthews coefficient (Å^3 Da^{-1})	3.49
R.m.s.d. bond lenths (Å)	0.016
R.m.s.d. bond angles (°)	1.512
Ramachandran plot, residues in	
Most favoured region (%)	97.6
Additionally allowed region (%)	2.4

2.2. Crystallization, data collection and processing

Crystallization was carried out by the sitting-drop technique using the Classics, PEG and MPD suites (Qiagen) and low-profile microplates (Greiner). 0.5 µl protein solution was mixed with an equal volume of reservoir solution. Several conditions yielded crystals with good appearance and of sufficient size. Some of these grew directly in cryoprotectant medium. We obtained well diffracting crystals (2.0 Å) using 0.2 *M* $CaCl_2$, 0.1 *M* sodium acetate pH 4.6, 15% MPD. Further optimizations did not yield better diffracting crystals. Diffraction data were collected at 100 K at the ESRF in Grenoble (beamline ID14-1). Data were indexed and integrated using *MOSFLM* (Leslie, 1992) and scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994).

The crystals belonged to space group $P6_522$, with unit-cell parameters a = 121.65, c = 332.24 Å. According to Matthews coefficient calculations, the asymmetric unit should consist of two molecules, with a V_M of 3.46 Å^3 Da^{-1} and a solvent content of 63%. A summary of data collection is given in Table 1.

2.3. Structure determination and refinement

The structure of *D. castellii* CBS 2923 phytase was solved by molecular replacement using the program *Phaser* (McCoy *et al.*, 2005). The *A. niger* acid phosphatase (PAAn; PDB code 1qfx; Kostrewa *et al.*, 1999) structure was chosen as the search model because of its relatively high sequence identity with PhytDc (36%). The initial model was built using a combination of automatic building and refinement with *ARP*/*wARP* (Perrakis *et al.*, 1999). Final manual building/refinement steps were carried out with *Coot* (Emsley & Cowtan, 2004) and maximum-likelihood refinement was carried out

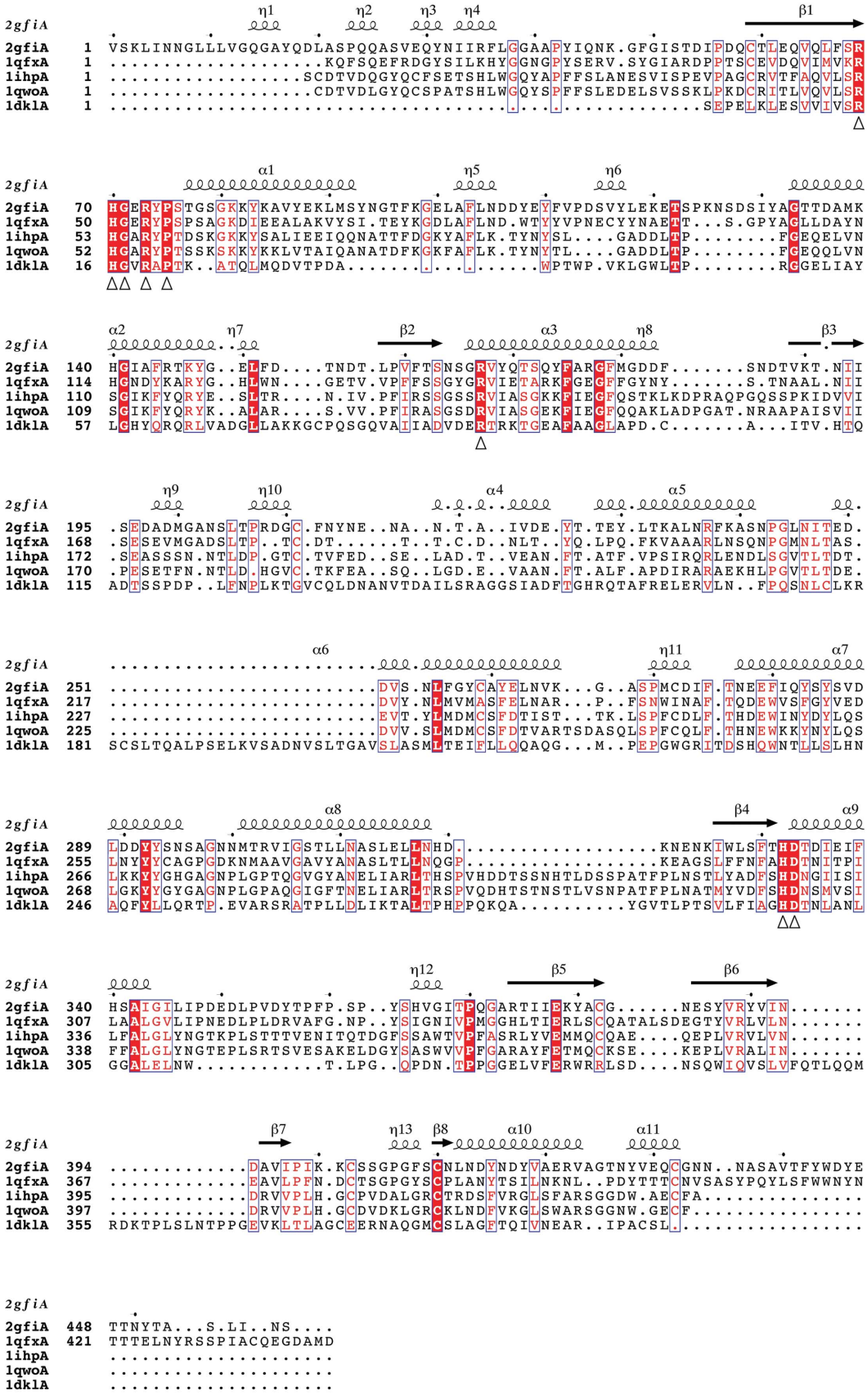


Figure 2
Structure-based sequence alignment of HAPs (Table 2). The secondary structure of PhyDc is shown above its sequence. (η, α and β represent 3_{10}-helix, α-helix and β-strand, respectively). Conserved residues are displayed in white on a red background. Residues with a high average similarity score are displayed in red in a blue frame. Active-site residues are indicated by a triangles. The alignment was performed using the programs *TM-align* (Zhang, 2008) and *ESPript* (Gouet *et al.*, 2003).

Table 2
Comparison of PhytDc with the other HAP phytases.

HAP-family member (PDB code, No. of residues)	Abbreviation used in this article	Sequence identity *versus* PhyDc (%)	R.m.s.d. *versus* PhyDc (Å)	Oligomeric state	Activity spectrum	Reference
D. castelli phytase (2gfi, 458)	PhyDc			Tetramer	Very large	Ragon *et al.* (2008)
A. niger acid phosphatase (1qfx, 441)	PAAn	36	2.8 overall, 1.3 over 314 C^{α}	Tetramer	Very large but not phytic acid	Kostrewa *et al.* (1999)
A. niger phytase (1ihp, 434)	PhytAn	25	3.2 overall, 1.6 over 327 C^{α}	Monomer	Specific for phytic acid	Kostrewa *et al.* (1997)
A. fumigatus phytase (1qwo, 435)	PhytAf	23	5.1 overall, 1.5 over 314 C^{α}	Monomer	Very large	Xiang *et al.* (2004)
E. coli phytase (1dkl, complexed)		5	Not applicable	Monomer	Highly specific for phytic acid	Lim *et al.* (2000)

using the program *REFMAC* (Murshudov *et al.*, 1997). The stereochemical quality of the final model was checked using *PROCHECK* (Laskowski *et al.*, 1993) and an updated Ramachandran plot (Lovell *et al.*, 2003). The final refinement statistics are summarized in Table 1.

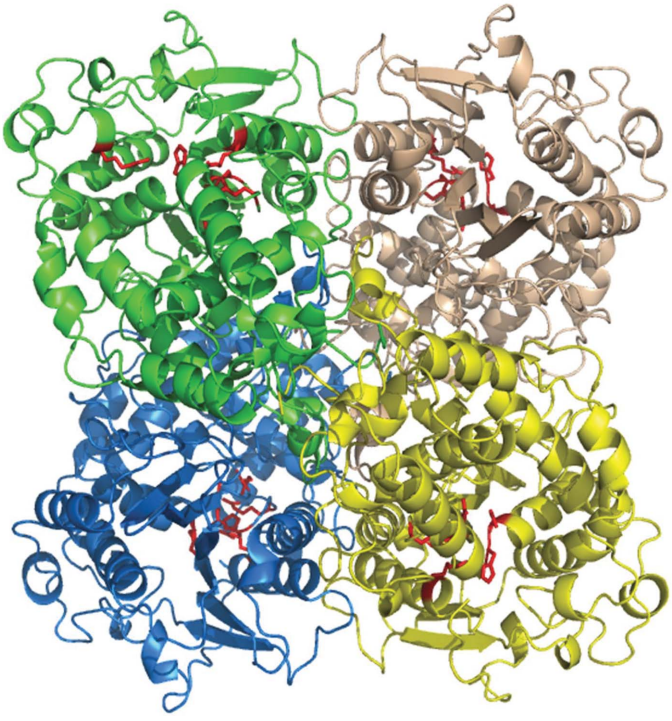


Figure 3
Tetramer of PhytDc with the residues of the active site shown in red. Figs. 3–6 were generated using *PyMOL* (DeLano, 2002).

3. Results and discussion

3.1. Overall structure of PhytDc

The final model comprises two monomers of PhyDc per asymmetric unit, consisting of residues 4–461. A total of five NAG molecules are visible and all are linked to an Asn. In the asymmetric unit, PhyDc presents a dimer with a contact area of 2550 $Å^2$ per monomer (13% of the monomer surface). Crystallographic symmetry generates the tetrameric biological molecule (Fig. 3). The tetramerization contact area per monomer is about 2400 Å (corresponding to 12% of the monomer surface). As expected from the sequence identity, the structures of PhytDc and PhytAn are very similar at the monomer level, as shown in Fig. 4 (overall main-chain r.m.s.d. of 1.3 Å; Table 2). Briefly, the structure can be divided into two parts: a large α/β-domain with a six-stranded β-sheet and a small α-domain. As in PAAn, the four active sites of the tetramer are solvent-exposed and are accessible to substrate. The two monomers completely superimpose with an overall root-mean-square deviation of 0.2 Å for main chains (0.6 Å for all atoms). Some minor differences can be seen in the NAG composition (two for monomer *A* and three for monomer *B*) and in the number of disulfide bonds (three in monomer *A* and four in monomer *B*).

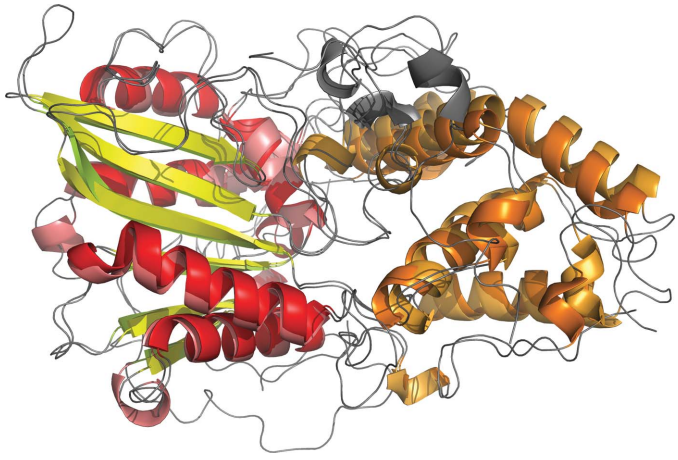


Figure 4
Overall superposition of the C^{α} trace of the *D. castellii* phytase monomer (PDB code 2gfi) with that of the *A. niger* acid phosphatase monomer (PDB code 1qfx). α-Domains are shown in orange and bright orange and α/β-domains in red/yellow and salmon/lemon.

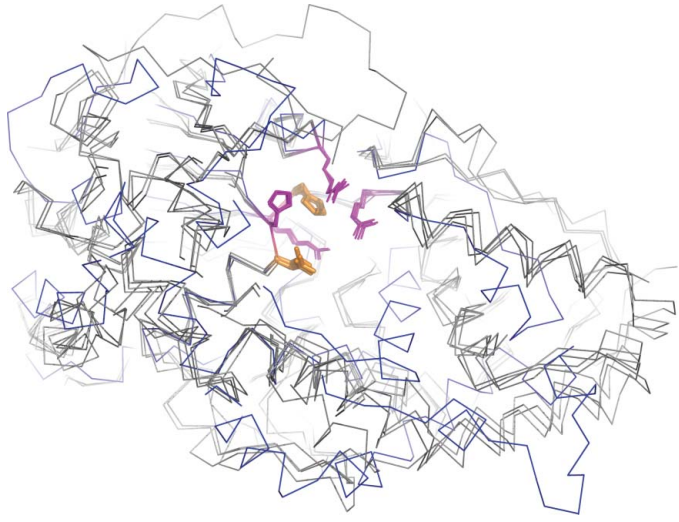
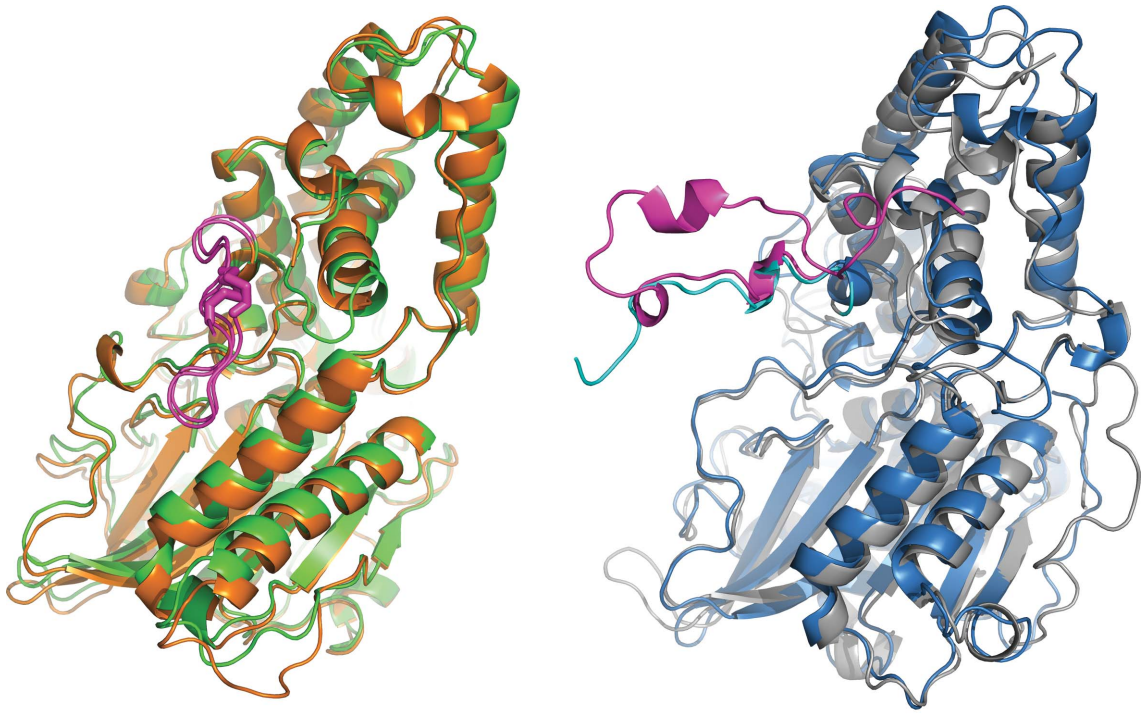


Figure 5
C^{α} superposition of all HAP-family phytases (PhyDc, PAAN, PhytAn and PhytAf are shown in grey and phytase from *E. coli* in blue) with active-site residues in purple (Arg72, Arg76, Arg170 and His 335) and catalytic residues shown as orange sticks (His73 and Asp336 in PhytDc numbering).



(*a*) (*b*)

Figure 6
(*a*) Superposition of PhytAf (green) and PhytAn (orange) with their N-terminal parts coloured pink. (*b*) Superposition of PhytDc (blue) and PAAn (grey) with their N-terminal parts coloured pink and cyan, respectively.

3.2. PhytDc *versus* others phytases

Several structures of phytase are available. Most of them are in the apo form, while two are of complexes with either persulfated phytic acid (a nonhydrolysable substrate) or phytic acid (the natural substrate) using an inactive mutant.

According to *SCOP* (Andreeva *et al.*, 2004), phytases are classified into three structural families. One is the thermostable phytase family with a six four-stranded β-sheet motif (β-propeller) and only contains the structure of the apo form of *Bacillus amyloliquefaciens* phytase (PDB code 1h6l; Shin *et al.*, 2001). Another family, the *myo*-inositol hexaphosphate phosphohydrolase family (with a core formed by a parallel β-sheet of four strands), only contains the structure from *Selenomonas ruminantium* (PDB code 1u24; Chu *et al.*, 2004). Finally, the histidine acid phosphatase family (or HAP family; one α-domain and one α/β-domain) contains phytases from various species and various organisms (such as *A. niger*, *A. fumigatus* and *Escherichia coli*). The PhytDc structure is very similar to that of PAAn and contains the sequence motif RHG*X*R*X*P which is characteristic of the HAP family. Hence, PhytDc will certainly be classified into this HAP family. All HAP-family phytase members, together with some of their properties, are listed in Table 2.

Superposition of all HAP-family phytases shows that all active-site residues are absolutely superposable and that with the exception of that from *E. coli* all HAP-family phytases present very similar structures (Table 2; Fig. 5). In this major subfamily, two phytases are monomeric and two are tetrameric (Table 2). Superposition of the monomers from monomeric or tetrameric phytases shows limited but significant differences. The PhytDc and PAAn N-terminal regions are extended (Fig. 6) and are strongly involved in dimeric and tetrameric contacts (data not shown). The corresponding parts of PhytAn and PhytAf are attached to the core of the protein.

At present, two structures of complexed forms of phytases are available. One is from the HAP family (*E. coli*; PDB codes 1dkq and 1dkp; an inactive mutant complexed with phytic acid). The second is from *S. ruminantium* (PDB code 1u24) complexed with persulfated phytate (a nonhydrolysable phytic acid analogue). The latter belongs to a different *SCOP* family and therefore has a structure that is completely different from those of the HAP family. As mentioned previously, *E. coli* phytase differs slightly from the other members of the HAP family. Thus, to date, no structure of a complexed phytase from the major subfamily of the HAP family has been resolved. Therefore, it will be interesting to obtain the structure of complexed PhyDc. Moreover, this structure will help to explain the substrate-specificity of PhyDc. Indeed, of the HAP-family phytases, only PhyDc is able to cleave all six phosphate groups of phytic acid.

This work was supported by ADISSEO. The financial support from ADISSEO and ANRT (Agence Nationale de la Recherche Technologique) to MR in the form of a doctoral grant is gratefully acknowledged. We thank the beamline ID14-1 team of the ESRF for technical support during data collection (especially Dr Chloe Zubieta) and the European Union for support of the work at the ESRF.

References

Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J. P., Chothia, C. & Murzin, A. G. (2004). *Nucleic Acids Res.* **32**, D226–D229.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Chu, H. M., Guo, R. T., Lin, T. W., Chou, C. C., Shr, H. L., Lai, H. L., Tang, T. Y., Cheng, K. J., Selinger, B. L. & Wang, A. H. (2004). *Structure*, **12**, 2015–2024.

DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* http://www.pymol.org.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Konietzny, U. & Greiner, R. (2002). *Int. J. Food Sci. Technol.* **37**, 791–812.
Kostrewa, D., Gruninger-Leitch, F., D'Arcy, A., Broger, C., Mitchell, D. & Van Loon, A. P. (1997). *Nature Struct. Biol.* **3**, 185–190.
Kostrewa, D., Wyss, M., D'Arcy, A. & Van Loon, A. P. (1999). *J. Mol. Biol.* **288**, 965–974.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lim, D., Golovan, S., Forsberg, C. W. & Jia, Z. (2000). *Nature Struct. Biol.* **7**, 108–113.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* D**61**, 458–464.
Mitchell, D. B., Vogel, K., Weimann, B. J., Pasamontes, L. & Van Loon, A. P. (1997). *Microbiology*, **143**, 245–252.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nakamura, Y., Fukuhara, H. & Sano, K. (2000). *Biosci. Biotechnol. Biochem.* **64**, 841–844.
Oh, B. C., Choi, W. C., Park, S., Kim, Y. O. & Oh, T. K. (2004). *Appl. Microbiol. Biotechnol.* **63**, 362–372.
Pandey, A., Szakacs, G., Soccol, C. R., Rodriguez-Leon, J. A. & Soccol, V. Y. (2001). *Bioresour. Technol.* **77**, 203–214.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ragon, M., Aumelas, A., Chemardin, P., Galvez, S., Moulin, G. & Boze, H. (2008). *Appl. Microbiol. Biotechnol.* **78**, 47–53.
Reddy, N. R., Sathe, S. K. & Salunkhe, D. K. (1982). *Adv. Food Res.* **28**, 1–92.
Shin, S., Ha, N. C., Oh, B. C., Oh, T. K. & Oh, B.-H. (2001). *Structure*, **9**, 851–858.
Xiang, T., Liu, Q., Deacon, A. M., Koshy, M., Kriksunov, I. A., Lei, X. G., Hao, Q. & Thiel, D. J. (2004). *J. Mol. Biol.* **339**, 437–445.
Zhang, Y. (2008). *Curr. Opin. Struct. Biol.* **18**, 342–348.