# Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the F1 antigen Caf1M–Caf1 chaperone–subunit pre-assembly complex from *Yersinia pestis*

**Anton Zavialov, Jenny Berglund
and Stefan D. Knight***

Department of Molecular Biology, Uppsala Biomedical Center, Swedish University of Agricultural Sciences, Box 590, SE-753 24 Uppsala, Sweden

Correspondence e-mail: stefan@xray.bmc.uu.se

The F1 capsular antigen of the plague-causing pathogen *Yersinia pestis* is assembled from monomeric Caf1 subunits *via* the Caf1M/Caf1A chaperone/usher system. *Y. pestis* Caf1M–Caf1 chaperone–subunit complex was purified from the periplasm of *Escherichia coli* cells overexpressing Caf1M and Caf1 and was crystallized in PEG 4000 solution using hanging-drop vapour diffusion. The crystals diffract to a minimum Bragg spacing of 1.8 Å and belong to space group $P2_1$, with unit-cell parameters $a = 36.0$, $b = 69.2$, $c = 69.1$ Å, $\beta = 93.0°$. SeMet-labelled Caf1M–Caf1 complexes were purified and crystallized under the same conditions. The SeMet crystals were identical to the native crystals and diffracted to 1.9 Å. Heavy-atom derivative crystals were prepared by soaking in 10 m$M$ $K_2PtCl_4$, giving two Pt sites per complex. The experimental electron-density map was obtained by a combination of MAD and MIR methods using both Se- and Pt-derivative crystals.

## 1. Introduction

The success of a bacterial pathogen depends critically on its ability to manipulate its surface structures. Gram-negative bacteria from the *Enterobacteriaceae* family assemble a rich arsenal of virulence organelles on their surfaces using the chaperone/usher pathway (Knight *et al.*, 2000; Sauer, Barnhart *et al.*, 2000; Sauer, Knight *et al.*, 2000; Soto & Hultgren, 1999). The morphology of these organelles varies considerably from thick and rigid pili to large amorphous capsules. Not surprisingly, they play a role in diverse events in microbial diseases, such as bacterial colonization of specific tissues by mediating binding to receptors on host cells or spreading of bacteria in blood by conferring anti-adhesive and anti-phagocytic properties (Du *et al.*, 2002; Soto & Hultgren, 1999).

*Yersinia pestis* is the causative agent of plague and is considered to be the most dangerous pathogen of the *Enterobacteriaceae* family (Perry & Fetherston, 1997). Following infection of a mammalian host, *Y. pestis* assembles a massive amorphous capsule, F1 antigen (F1-Ag), on its surface (Baker *et al.*, 1952). This capsule is comprised of a polymeric form of the 15.5 kDa Caf1 protein subunit (Galyov *et al.*, 1990). It has been known for a long time that F1-Ag efficiently inhibits phagocytosis (Cavanaugh & Randall, 1959; Du *et al.*, 2002; Williams *et al.*, 1972), although it is its paradoxical property to induce a protective immune response in humans that has attracted most scientific attention (Baker *et al.*, 1952). As a result, efficient vaccines against bubonic plague (Meyer, 1970; Meyer *et al.*, 1974) as well as its serodiagnostic assays (Poland & Barnes, 1979) have been created based on F1-Ag.

F1-Ag has been studied as a prototype of non-pilus aggregative organelles assembled *via* FGL periplasmic chaperone/usher pathways (Chapman *et al.*, 1999; Hung *et al.*, 1996; MacIntyre *et al.*, 2001; Zav'yalov *et al.*, 1997; Zav'yalov, Zav'yalova *et al.*, 1995; Zavialov *et al.*, 2002). In contrast to the rigid pili assembled *via* FGS chaperone/usher pathways, these non-pilus structures are built of only one or sometimes two types of self-polymerizing subunits (Hung *et al.*, 1996; Zav'yalov, Zav'yalova *et al.*, 1995). Perhaps the most striking difference between pili and non-pilus organelles is that the latter do not contain a specialized receptor-binding adhesin domain as is normally present at the tips of pili (Soto & Hultgren, 1999). Despite this, many of the non-pilus organelles have been shown to possess adhesive properties, suggesting that their subunits play a dual role as both structure-forming subunits and receptor-binding adhesins.

Two specialized proteins are required for assembly of the F1-Ag: a 26.3 kDa periplasmic chaperone (Caf1M) and a 90.4 kDa outer membrane usher (Caf1A) (Karlyshev *et al.*, 1994). As Caf1 subunit emerges into the periplasm, Caf1M binds it, assists in its folding and delivers it to Caf1A, where subunits are

# crystallization papers

assembled into linear polymers and secreted to the cell surface (Zavialov *et al.*, 2002).

Caf1M is the prototype of the FGL subfamily of periplasmic chaperones assembling non-pilus organelles (Hung *et al.*, 1996; Zav'yalov, Zav'yalova *et al.*, 1995). The amino-acid sequences of FGL periplasmic chaperones share about 35% identity. Compared with the FGS subfamily of periplasmic chaperones involved in assembly of pili, FGL chaperones have large additional sequences at their N-termini and between the F1 and G1 $\beta$-strands (Chapman *et al.*, 1999; Zav'yalov, Zav'yalova *et al.*, 1995). In addition, they contain an invariant pair of Cys residues forming a disulfide bridge between the F1 and G1 $\beta$-strands (Zav'yalov *et al.*, 1997).

Structures of three FGS chaperones, PapD, FimC and SfaE, all revealed the same general fold: two immunoglobulin (Ig)-like domains joined at approximately right angles with a large cleft in between (Holmgren & Brändén, 1989; Knight *et al.*, 2002; Pellecchia *et al.*, 1998). Structures of the FimC–FimH chaperone–adhesin (Choudhury *et al.*, 1999) and the PapD–PapK chaperone–adapter (Sauer *et al.*, 1999) pre-assembly complexes from the type 1 pilus and P pilus, respectively, showed that pilin subunits have incomplete Ig-like folds. The absence of a final (seventh) $\beta$-strand of the fold creates a deep hydrophobic cleft on the surface of pilin subunits, rendering them dependent on the periplasmic chaperone for their stability and proper folding. The chaperone binds to the pilins by donating its G1 $\beta$-strand to complete the pilin Ig-like fold in a process called donor-strand complementation (Choudhury *et al.*, 1999; Sauer *et al.*, 1999). Assembly of subunits is thought to proceed by a donor-strand-exchange mechanism in which the chaperone G1 donor-strand interaction is replaced by a similar interaction between subunits (Choudhury *et al.*, 1999; Sauer *et al.*, 1999).

No structure of a FGL chaperone or a non-pilus subunit has been determined and the structural basis for the striking difference in morphology between pilus and non-pilus organelles is not known. In this paper, we present the overexpression, purification and crystallization of the FGL pre-assembly chaperone–subunit complex Caf1M–Caf1, along with X-ray diffraction data. The structure of the Caf1M–Caf1 complex will shed light on important questions about the biogenesis, ultrastructure and immunogenic properties of the F1 capsule in particular and of non-pilus organelles in general. The structure may also help clarify how surface organelles with very different architectures are achieved using similar but distinct classes of periplasmic chaperones.

## 2. Materials and methods

### 2.1. Expression system and genetic modification of Caf1

A six-histidine tag was genetically introduced in Caf1 by substituting Ser-His$_6$ for the sequence between Thr4 and Val14 of mature Caf1 using inverse PCR with Pfu DNA polymerase (Stratagene, USA), oligonucleotides 5'-tat**actagt**catcaccatcac-GTTGAACCAGCCCGCATCA-3' and 5'-tat**actAGT**TAAATCTGCCGCATTAGCA-GTTG-3' containing a SpeI restriction site (bold) (MedProbe, Norway), and pFM1 (Zav'yalov, Chernovskaya *et al.*, 1995) as a template. The PCR product was purified from gel, digested with SpeI and ligated with T4 DNA ligase (Promega, USA) to produce pFM-1-6H. Host strain *E. coli* B834 (DE3) strain (Novagene, USA), a methionine-requiring auxotroph, was transformed with pFM-1-6H. For expression of native complexes, cells were grown to an $A_{600nm}$ of 0.6 in LB medium (Maniatis *et al.*, 1998) containing 80 µg ml$^{-1}$ ampicillin. Protein expression was induced with 0.5 m*M* IPTG for 1.5 h. To express SeMet-labelled protein, cells were grown to an $A_{600nm}$ of 0.5 in NMM medium (Budisa *et al.*, 1995) supplemented with 80 µg ml$^{-1}$ ampicillin and 0.6 m*M* L-SeMet (Fluka, Switzerland), followed by induction of proteins with 0.5 m*M* IPTG for 4.5 h. Periplasmic extracts were obtained by osmotic shock as described in MacIntyre *et al.* (2001).

### 2.2. Purification of Caf1M–Caf1 complex

Periplasmic extracts were dialyzed against 50 m*M* phosphate buffer pH 7.8, 300 m*M* NaCl, 10 m*M* imidazole containing Complete protease inhibitors (Roche, France) overnight and loaded onto a HiTrap Chelating HP 5 ml column (Pharmacia, Sweden). Complexes were eluted with a sharp 10–500 m*M* imidazole gradient. Fractions containing complexes were pooled, dialysed against 50 m*M* CH$_3$COONa pH 5.25 and loaded onto a Mono S HR 8 ml column. Complexes were eluted with a 0–500 m*M* NaCl gradient (at about 300 m*M* NaCl), desalted with an Econo-Pac 10 DG (BioRad, USA) column and concentrated using Vivaspin 20 ml (Vivascience, UK) to 15–20 mg ml$^{-1}$. Normally, about 10 mg of pure native and 4 mg of SeMet-labelled complex was isolated from cells grown in 1 l of LB and NMM, respectively.

### 2.3. Analytical procedures

Samples of purified complexes were analysed by SDS and native PAGE on 12.5% polyacrylamide gels and by isoelectric focusing (IEF) on pH 3–9 gradient gels in a Pharmacia Phast system (Pharmacia, Sweden), by gel filtration using a Superose 12 HR column and by dynamic light scattering using a DynaPro-801 TC light-scattering apparatus (Protein Solutions, UK). The concentration was routinely determined by measuring the absorption at 280 nm on a Hewlett Packard 845A spectrophotometer (Hewlett Packard, USA), using an extinction coefficient of $\varepsilon_{280nm}^{0.01\%} = 1$, or by a Coomassie blue binding assay (BioRad, USA). The SeMet complex was analysed for SeMet incorporation by comparative analysis of the amino-acid composition of the native and SeMet complexes (Shepard & Huber, 1969) at the Analytical Service Centre (Department of Biochemistry, University of Uppsala). No methionine band was visible in the SeMet sample, indicating full SeMet incorporation.

### 2.4. Crystallization

Crystallization was performed by the hanging-drop vapour-diffusion method at 293 K. In all experiments the crystallization drops, containing 2.5 µl protein solution (15–20 mg ml$^{-1}$) and 2.5 µl precipitant solution, were equilibrated against 1 ml precipitant solution. The initial crystallization conditions were found using Hampton Crystal Screen (Hampton Research, USA). Crystalline-like precipitates were found in solutions containing 20–30% PEG 4000 and PEG 8000. Single crystals suitable for X-ray analysis were obtained by finer screening around these conditions.

### 2.5. X-ray diffraction study

Diffraction data were collected under liquid-nitrogen cryoconditions at 100 K. To avoid damage on freezing, crystals were soaked for 10–30 s in a cryoprotection solution prepared by mixing one part precipitant solution with one part 25% PEG 400. Crystals were flash-cooled by rapidly moving them into the cold nitrogen stream or by dipping them in liquid nitrogen. Native X-ray diffraction data were collected on beamline ID14-1, ESRF, France. SeMet MAD experiments were performed at beamlines ID14-4 and ID29 at the ESRF; Pt data were collected both at the ESRF

(beamline ID14-2) and at Max-lab beamline 7-11, Lund. All ESRF data were collected using a ADSC Q4 CCD detector; Lund data were collected on a MAR Research 165 mm CCD detector. Data were processed, integrated and scaled using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Heavy-atom positions were determined using *RSPS* (Collaborative Computational Project, Number 4, 1994; Knight, 2000). Heavy-atom parameters were refined and initial phases calculated using *SHARP* (de La Fortelle & Bricogne, 1997).

**Table 1**
Statistics of X-ray data collection.

Values in parentheses indicate the statistics for the last resolution shell.

| Data set | Se, peak | Se, inflection | Native | Pt derivative, ESRF | Pt derivative, Lund |
|---|---|---|---|---|---|
| Wavelength (Å) | 0.9798 | 0.9801 | 0.933 | 0.933 | 1.076 |
| Resolution (Å) | 34–2.7 | 36–2.7 | 20–1.8 | 49–2.2 | 20–2.5 |
| No. of unique reflections | 17046 (928) | 17458 (941) | 34122 (2587) | 17430 (1724) | 12329 (1709) |
| Multiplicity | 3.2 | 3.4 | 3.3 | 3.7 | 3.8 |
| Completeness (%) | 96.7 (84.6) | 97.3 (85.4) | 90.5 (69.5) | 99.9 (99.6) | 99.5 (99.5) |
| $R_{merge}$† (%) | 4.6 (34.8) | 4.8 (35.2) | 7.7 (79.1) | 8.2 (36) | 5.9 (16.5) |
| $I/\sigma(I)$ | 11 (2.1) | 9.9 (1.9) | 20.9 (1.6) | 18.6 (4.2) | 10.2 (3.9) |

† $R_{merge} = \sum_h \sum_i |I(h)_i - \langle I \rangle| / \sum_h \sum_i I(h)_i$, where $I(h)$ is the intensity of a reflection $h$, $\sum_h$ is the sum over all reflections and $\sum_i$ is the sum over $i$ measurements of reflection $h$.

## 3. Results and discussion

Overexpression of Caf1M and Caf1 in *E. coli* results in accumulation of Caf1M–(Caf1)$_n$ complexes in the periplasm (Zavialov et al., 2002). However, because of spontaneous F1 assembly such as polymerization of Caf1, periplasmic complexes differ in the amount of Caf1 molecules incorporated into the complex. We previously showed that the polymerization of Caf1 can be abolished by deletion of a short N-terminal sequence without disruption of the chaperone–subunit interactions. Since this 12-amino-acid sequence was found to be flexible and apparently is not necessary for the general fold of Caf1 (Zavialov et al., 2002), we replaced it by a six-histidine tag to block polymerization and in order to be able to use metal-chelate chromatography for purifica-
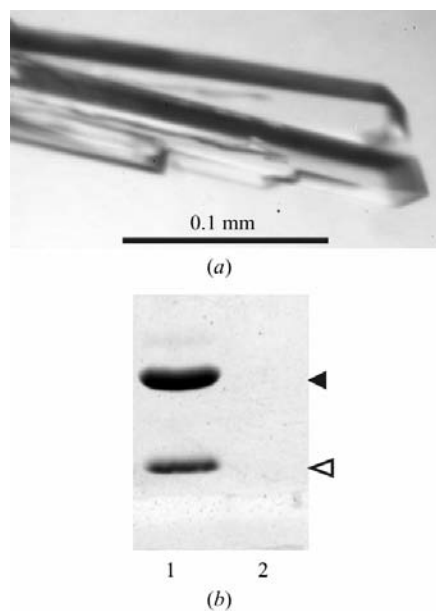
tion. Coexpression of this modified Caf1 with Caf1M resulted in high-level accumulation of stable binary Caf1M–Caf1 complex in the periplasm.

Pure complex was isolated by Ni-chelate chromatography of periplasmic extracts. However, dynamic light scattering and gel filtration detected a small amount of high-molecular-weight aggregates (presumably of Caf1) in these preparations. Therefore, cation-exchange chromatography at pH 5.25 followed by desalting gel filtration was used for further purification. Complexes appeared to be very stable at pH 5.25 and no aggregation was evident even after freezing and thawing of the samples. Protein was concentrated to 15–20 mg ml$^{-1}$ and stored frozen at 193 K for several months.

Crystallization screening and optimization produced single crystals in drops with 19–21% PEG 4000 in 0.1 *M* Tris–HCl pH 7.5. SeMet crystals could be obtained under almost identical conditions. Generally, crystals of three different shapes were formed: thick plates, thin needles and bundles of thick needles (Fig. 1*a*). The presence of both Caf1M and Caf1 in the crystals was confirmed by SDS–PAGE on carefully washed crystals. The resulting gel (Fig. 1*b*) shows two bands corresponding to Caf1M and Caf1 (lane 1), whereas there is no protein in the final wash solution (lane 2). Thick plates diffracted to a minimum Bragg spacing of 1.5 Å on the ID14-4 beamline at the ESRF. However, these crystals showed high mosaicity. The thin needles diffracted well, but they were quickly damaged by radiation. Crystals suitable for X-ray analysis were obtained by gentle separation of thick needles from bundles. These crystals diffract to a minimum Bragg spacing of 1.8–1.9 Å and belong to space group $P2_1$, with unit-cell parameters $a = 36.0$, $b = 69.2$, $c = 69.1$ Å, $\beta = 93.0°$. The asymmetric unit contains one molecule of Caf1M–Caf1 complex, with a solvent content of 30%. Native data to 1.8 Å Bragg spacing were

collected at ESRF beamline ID14-4 from a single crystal (Table 1).

SeMet-labelled crystals obtained in the same way as those of the wild type belonged to the same space group, with essentially the same unit-cell parameters, and diffracted to about 1.9 Å. SeMet crystals underwent a considerable loss in diffraction quality after one month of storage. Presumably, this was because of oxidation of Se atoms, with the formation of acidic $-SeO_2^-$ groups that destabilize the protein fold. A three-wavelength MAD data set was collected from a fresh single crystal on beamline ID14-4 at the ESRF, France (Table 1). Based on the absorption spectrum for the Se $K$ edge, two energies were chosen: at the peak ($\lambda = 0.9798$ Å) and the inflection point ($\lambda = 0.9801$ Å). Data to 2.7 Å were obtained at each of these energies. A third energy was selected at $\lambda = 0.9399$ Å as a remote point. The crystal showed clear signs of radiation damage and the remote wavelength was not included in further calculations. The positions of all four Se sites were determined from anomalous difference Patterson maps using *RSPS* (Collaborative Computational Project, Number 4, 1994; Knight, 2000). The heavy-atom parameters were refined and the experimental phases calculated using *SHARP* (La Fortelle & Bricogne, 1997). However, we were unable to interpret the resulting map even after application of density-modification techniques (Cowtan & Zhang, 1999). The failure to produce an interpretable map was presumably a consequence of the low phasing power afforded by only four Se atoms in a protein of 42 kDa and the relatively weak power of density modification at moderate resolution, low solvent content and in the absence of non-crystallographic symmetry.

In order to improve phasing, a Pt-derivative data set was collected at Max-lab (Lund) at $\lambda = 1.076$ Å, very close to the $L_{III}$ absorption edge of platinum (Table 1). Another platinum data set was collected at



**Figure 1**
(*a*) A photograph of monoclinic crystals of the Caf1M–Caf1 complex. (*b*) SDS–PAGE gel of a washed crystal (lane 1) and the final wash solution (lane 2). Arrowheads indicate the positions of Caf1M (filled) and Caf1 (open).

# crystallization papers

ID14-1, ESRF at $\lambda = 0.933$ Å, close to the Pt $L_{II}$ edge (Table 1). Both these data sets showed strong anomalous signals and two Pt positions could be found using *RSPS*. Combining SeMet MAD data with the two Pt data sets in *SHARP* gave a final good experimental map with a mean figure of merit of 0.434 for the outer 2.06–2.00 Å resolution shell after application of density modification. Modelling of the Caf1M–Caf1 complex in this map is now in progress.

## References

Baker, E. E., Sommer, H., Foster, L. E., Meyer, E. & Meyer, K. F. (1952). *J. Immunol.* **68**, 131–140.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). *Eur. J. Biochem.* **230**, 788–796.

Cavanaugh, D. C. & Randall, R. (1959). *J. Immunol.* **83**, 348–363.

Chapman, D. A. G., Zavialov, A. V., Chernovskaya, T. V., Karlyshev, A. V., Zav'yalova, G. A., Vasiliev, A. M., Dudich, I. V., Abramov, V. M., Zav'yalov, V. P. & MacIntyre, S. (1999). *J. Bacteriol.* **181**, 2422–2429.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. L. & Knight, S. D. (1999). *Science*, **285**, 1061–1066.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Du, Y., Rosqvist, R. & Forsberg, A. (2002). *Infect. Immun.* **70**, 1453–1460.

Galyov, E. E., Smirnov, O. Y., Karlishev, A. V., Volkovoy, K. I., Denesyuk, A. I., Nazimov, I. V., Robtsov, K. S., Abramov, V. M., Dalvadyanz, S. M. & Zav'yalov, V. P. (1990). *FEBS Lett.* **277**, 230–232.

Holmgren, A. & Brändén, C.-I. (1989). *Nature (London)*, **342**, 248–251.

Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. & Hultgren, S. J. (1996). *EMBO J.* **15**, 3792–3805.

Karlyshev, A. V., Galyov, E. E., Smirnov, O. Y., Abramov, V. M. & Zav'yalov, V. P. (1994). *Biological Membranes: Structure, Biogenesis and Dynamics*, edited by J. A. F. Op den Kamp, Vol. 11, pp. 321–330. Berlin: Springer–Verlag.

Knight, S. D. (2000). *Acta Cryst.* D**56**, 42–47.

Knight, S. D., Berglund, J. & Choudhury, D. (2000). *Curr. Opin. Chem. Biol.* **4**, 653–660.

Knight, S. D., Choudhury, D., Hultgren, S., Pinkner, J., Stojanoff, V. & Thompson, A. (2002). *Acta Cryst.* D**58**, 1016–1022.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

MacIntyre, S., Zyrianova, I. M., Chernovskaya, T. V., Leonard, M., Rudenko, E. G., Zav'yalov, V. P. & Chapman, D. A. G. (2001). *Mol. Microbiol.* **39**, 12–25.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1998). *Molecular Cloning: A Laborotory Manual.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Meyer, K. F. (1970). *Bull. W. H. O.* **42**, 653–666.

Meyer, K. F., Hightower, J. A. & McCrumb, F. R. (1974). *J. Infect. Dis.* **129**, S41–S45.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, pp. 307–326.

Pellecchia, M., Guntert, P., Glockshuber, R. & Wuthrich, K. (1998). *Nature Struct. Biol.* **5**, 885–890.

Perry, R. D. & Fetherston, J. D. (1997). *Clin. Microbiol. Rev.* **10**, 35–66.

Poland, J. D. & Barnes, A. M. (1979). *Plague.* Boca Raton, FL, USA: CRC Press.

Sauer, F. G., Barnhart, M., Choudhury, D., Knight, S. D., Waksman, G. & Hultgren, S. J. (2000). *Curr. Opin. Struct. Biol.* **10**, 548–556.

Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). *Science*, **285**, 1058–1061.

Sauer, F. G., Knight, S. D., Waksman, G. & Hultgren, S. J. (2000). *Semin. Cell. Dev. Biol.* **11**, 27–34.

Shepard, L. & Huber, R. E. (1969). *Can. J. Biochem.* **47**, 877–891.

Soto, G. E. & Hultgren, G. S. (1999). *J. Bacteriol.* **181**, 1059–1071.

Williams, R. C. Jr, Gewurz, H. & Quie, P. G. (1972). *J. Infect. Dis.* **126**, 235–241.

Zavialov, A. V., Kersley, J., Korpela, T., Zav'yalov, V. P., MacIntyre, S. & Knight, S. D. (2002). *Mol. Microbiol.* **45**, 983–995.

Zav'yalov, V. P., Chernovskaya, T. V., Chapman, D. A. G., Karlyshev, A. V., MacIntyre, S., Zavialov, A. V., Vasiliev, A. M., Denesyuk, A., Zav'yalova, G. A., Dudich, I. V., Korpela, T. & Abramov, V. M. (1997). *Biochem. J.* **324**, 571–578.

Zav'yalov, V. P., Chernovskaya, T. V., Navolotskaya, E. V., Karlyshev, A. V., MacIntyre, S., Vasiliev, A. M. & Abramov, V. M. (1995). *FEBS Lett.* **371**, 65–68.

Zav'yalov, V. P., Zav'yalova, G. A., Denesyuk, A. I. & Korpela, T. (1995). *FEMS Immunol. Med. Microbiol.* **11**, 19–24.