Acanthamoeba polyphaga mimivirus NDK: preliminary crystallographic analysis of the first viral nucleoside diphosphate kinase

Sandra Jeudy,[a] Bruno Coutard,[b] Régine Lebrun[c] and Chantal Abergel[a]*

[a]Information Génomique et Structurale, CNRS UPR 2589, 31 Chemin Joseph Aiguier, 13402 Marseille CEDEX 20, France, [b]Architecture et Fonction des Macromolecules Biologiques, CNRS UMR 6098, 31 Chemin Joseph Aiguier, 13402 Marseille CEDEX 20, France, and [c]IBSM, 31 Chemin Joseph Aiguier, 13402 Marseille CEDEX 20, France

Correspondence e-mail: chantal.abergel@igs.cnrs-mrs.fr

The complete sequence of the largest known double-stranded DNA virus, *Acanthamoeba polyphaga* mimivirus, has recently been determined [Raoult *et al.* (2004), *Science*, **306**, 1344–1350] and revealed numerous genes not expected to be found in a virus. A comprehensive structural and functional study of these gene products was initiated [Abergel *et al.* (2005), *Acta Cryst.* F**61**, 212–215] both to better understand their role in the virus physiology and to obtain some clues to the origin of DNA viruses. Here, the preliminary crystallographic analysis of the viral nucleoside diphosphate kinase protein is reported. The crystal belongs to the cubic space group $P2_13$, with unit-cell parameter 99.425 Å. The self-rotation function confirms that there are two monomers per asymmetric unit related by a twofold non-crystallographic axis and that the unit cell thus contains four biological entities.

1. Introduction

With a 400 nm particle size and a genome of 1.2 Mbp, the recently discovered *Acanthamoeba polyphaga* mimivirus (La Scola *et al.*, 2003) challenges the established frontier between viruses and parasitic cellular organisms of the same size, such as *Mycoplasma*.

In addition to its unusual genome size, *A. polyphaga* mimivirus has an unusually complex genome (Raoult *et al.*, 2004). It presents numerous genes that are unexpected in a virus, such as genes encoding protein-translation components, including four amino-acyl tRNA synthetases, various translation-initiation, elongation and termination factors, as well as proteins implicated in DNA repair, protein folding or proteins involved in new metabolic pathways. Among the proteins never identified before in a viral genome, *A. polyphaga* mimivirus includes a nucleoside diphosphate kinase (NDK). NDKs are required for the synthesis of nucleoside triphosphates (NTP) other than ATP (Parks & Agarwal, 1973; EC 2.7.4.6). They are non-specific enzymes active on both purine and pyrimidine, ribonucleotides or deoxyribonucleotides and can provide NTPs or dNTPs for nucleic acid synthesis, CTP for lipid synthesis, UTP for polysaccharide synthesis and GTP for protein elongation, signal transduction and microtubule polymerization.

Because these features that distinguish *A. polyphaga* mimivirus from other nucleocytoplasmic large DNA viruses (NCLDVs) may help in understanding its origin and lineage, we initiated a comprehensive structural and functional study of the unique *A. polyphaga* mimivirus genes. The sequence analysis of the *A. polyphaga* mimivirus nucleoside diphosphate kinase (NDK) highlights a specific feature of the viral protein located at the substrate-binding site. Enzymatic studies of this enzyme together with its structural analysis should provide some insights into the viral NDK specificity and its functional role in the context of the amoeba infection. In this work, we report the cloning, expression and crystallization of the 137-amino-acid NDK protein (Genbank: MIMI_R418).

2. Results and discussion

2.1. Expression of the NDK gene product

The gene encoding the *A. polyphaga* mimivirus nucleoside diphosphate kinase was amplified from mimivirus genomic DNA and
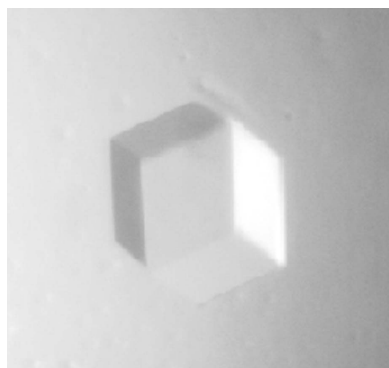
directional cloning was performed using the Gateway system (Invitrogen) as previously described (Abergel *et al.*, 2003). The PCR product was inserted by homologous recombination in the pDIGS02 expression plasmid in phase with an N-terminal His$_6$ tag under the control of a T7 promoter. The pDIGS02 is an in-house vector engineered to selectively coexpress the GroEL–GroES chaperone complex with the gene of interest by tetracyclin induction in order to help the protein folding. After transformation into DH10B cells (Invitrogen), the purified plasmids were used for the overexpression of the recombinant proteins using a specific expression-screening protocol. Incomplete factorial experimental design (as implemented by the *SAmBa* software; Audic *et al.*, 1997; http://igs-server.cnrs-mrs.fr/samba/) was used to define a set of 12 conditions corresponding to the combination of three variables.

(i) Three temperatures: 315, 310 and 298 K.

(ii) Four *Escherichia coli* expression strains: BL21(DE3)pLysS, Rosetta(DE3)pLysS, Origami(DE3) (Novagen) and C41(DE3) (Avidis).

(iii) Three tetracycline concentrations to vary the level of GroEL–GroES chaperon coexpression: 0, 10 and 50 µg l^{-1}.

The best results were obtained when transforming *E. coli* Rosetta(DE3)pLysS at 298 K. Since the decrease in temperature was visibly a positive factor for soluble expression, we optimized our results by further decreasing the growth temperature. Initial growth was performed at 310 K in 2YT media containing ampicillin and chloramphenicol. After induction with 0.5 m*M* IPTG when A_{600} reached 0.6–0.8, the temperature was decreased to 290 K. Chaperon coexpression had a neutral effect on the protein solubility and was not used in the final protocol. The pellet was resuspended in 50 m*M* sodium phosphate, 300 m*M* NaCl buffer pH 9.0 (buffer *A*) containing 0.1% Triton X-100 and 5% glycerol and total proteins were extracted by sonication.

2.2. Purification

Purification of the recombinant protein was performed using a standard protocol set for all our structural genomics targets (Abergel *et al.*, 2003). The recombinant NDK corresponds to the native protein where the N-terminal methionine is replaced by an extended His tag inherent to the use of the Gateway system (21-residue tag: SYYHHHHHHLESTSLYKKAGL). A microdialysis experiment was performed and allowed us to identify the best buffer as 10 m*M* CHES pH 9.0 (Abergel *et al.*, 2003). The purified protein was characterized by mass spectroscopy and by N-terminal Edman sequencing.

Isoelectric focusing using pH 3–10 gradient pre-cast gels (Novex) revealed a band around pI 6.7. Dynamic light scattering using the Dynapro system (Protein Solutions) indicated that the protein solution was monodisperse. Aliquots of the purified protein were stored at various temperatures (293, 277 and 193 K) to monitor the protein stability over time. After two weeks, 90% of the NDK recombinant protein has been transformed into a truncated form. N-terminal sequencing revealed an N-terminal cleavage affecting the first 15 residues of the tag. After two months of storage at 297 K, the NDK was fully degraded.

2.3. Crystallization

The NDK was concentrated to 17 mg ml^{-1} in 10 m*M* CHES pH 9.0 using a centrifugal filter device (Ultrafree Biomax 5K, Millipore, Bedford, MA, USA). The NDK recombinant protein was initially tested at 293 K against 480 different conditions corresponding to commercially available solution sets (Crystal Screens from Hampton Research, Wizard Screens from Emerald Biostructures) and conditions designed in-house using *SAmBA* software (Audic *et al.*, 1997). The screening for crystallization conditions was performed on 3 × 96-well crystallization plates (Greiner) loaded by an eight-needle dispensing robot (Tecan, WS 100/8 workstation modified for our needs), using one 1 µl sitting drop per condition (Abergel *et al.*, 2003).

A few small crystals appeared after several weeks in 0.1 *M* MES pH 6.5 and 1.44 and 1.6 *M* (NH$_4$)$_2$SO$_4$, respectively. We tried to refine these conditions at 293 K by hanging-drop vapour diffusion using 24-well culture plates (Greiner). Each hanging drop was prepared by mixing 0.5 µl 17 mg ml^{-1} NDK with 0.5 µl reservoir solution. The hanging drop on the cover glass was vapour-equilibrated against 1 ml reservoir solution with increasing (NH$_4$)$_2$SO$_4$ concentrations from 1 to 2.4 *M* in each well of the tissue-culture plate. These conditions failed to produce crystals after two weeks of observation and we thus decided to test if we could increase the size of the crystals obtained in 96-well crystallization plates using the macroseeding procedure and thus transferred one of them to the corresponding 24-well culture-plate conditions [1.6 *M* (NH$_4$)$_2$SO$_4$] after equilibration. This crystal was then tested for diffraction (Fig. 1).

2.4. Data collection and processing

The crystal was picked up in a Hampton Research 0.05 × 0.05 mm loop, flash-frozen to 105 K in a cold nitrogen-gas stream and subjected to X-ray diffraction. The data set was collected on a MAR
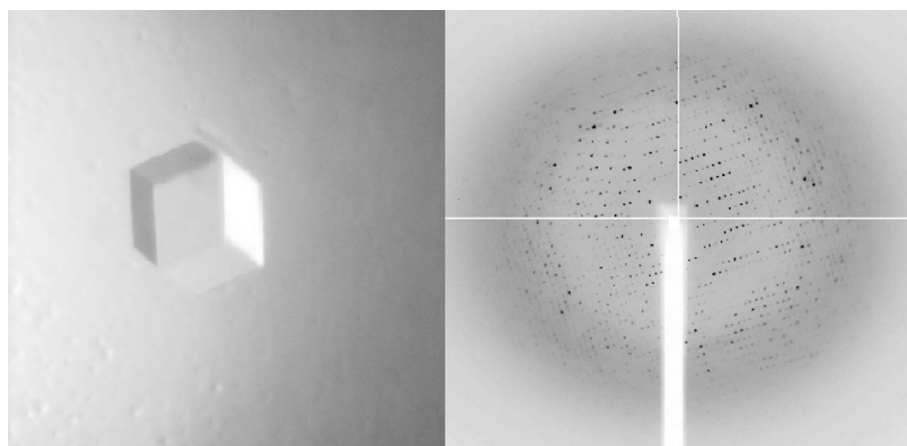

Figure 1
Picture and diffraction image of the NDK crystal used for data collection (ID29, ESRF, Grenoble).

Table 1
X-ray diffraction data.

Values in parentheses are for the outer resolution shell.

Wavelength (Å)	0.97563
Resolution (Å)	70.36–2.55 (2.64–2.55)
Reflections	40079 (1660)
Unique reflections	10356 (977)
Completeness (%)	95.9 (94.1)
$I/\sigma(I)$	9.7 (2.7)
Multiplicity	3.9 (1.7)
R_{sym}† (%)	5.9 (25.6)

† $R_{sym} = \sum_h \left||F_{obs}(h)| - |F_{calc}(h)|\right| / \sum_h |F_{obs}(h)|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes for the reflection with Miller indices $h = (hkl)$.

Table 2
Preliminary molecular-replacement statistics.

	Correlation	R factor
1k44	51.7	47.1
1ndl	32.5	53.1
CaspR model	49.4	47.4

CCD camera at the European Synchrotron Radiation facility (ID29 beamline) at a wavelength of 0.97563 Å. Data collection was performed with an oscillation angle of 1° and a crystal-to-detector distance of 305 mm; a total of 90 images were collected. Given the crystal size, radiation damage rapidly became obvious. Since the crystal belongs to the cubic $P2_13$ space group, with unit-cell parameter 99.425 Å, we only used 40° of the collected data. *MOSFLM* and *SCALA* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) were used for processing, scaling and data reduction of the data set.

To determine the sequence of the NDK in the crystal, another crystal from the same condition was analysed by mass spectroscopy,

revealing a total weight of 18 142 Da corresponding to the recombinant protein with the tag (theoretical MW 18 146 Da). The packing density for two monomers of NDK (36 284 Da) in the asymmetric unit of the crystal is 2.26 Å^3 Da^{-1}, a low but reasonable value for globular proteins, indicating an approximate solvent content of 45.5% (Matthews, 1968). We used the *AMoRe* (Navaza, 2001) self-rotation function to confirm the presence of two molecules in the asymmetric unit and identified a peak corresponding to twofold non-crystallographic symmetry (correlation: 33%). The crystal diffracted to 2.55 Å and statistics are presented in Table 1.

The sequence of the *A. polyphaga* mimivirus NDK shows 44% identity to the *Mycobacterium tuberculosis* NDK structure (PDB code 1k44; Chen *et al.*, 2002) and 38% to the *Drosophila melanogaster* structure (PDB code 1ndl; Chiadmi *et al.*, 1993) (Fig. 2). The structure will be solved by molecular replacement using the CaspR webserver (Claude *et al.*, 2004; http://igs-server.cnrs-mrs.fr/Caspr/index.cgi) and these structures as references to generate the *A. polyphaga* mimivirus NDK models. Preliminary results are presented in Table 2.

In NDKs, a highly conserved loop, the Kpn loop (Fig. 2), is involved in substrate binding and the oligomeric state of the enzyme (Janin *et al.*, 2000). Interestingly, the multiple alignment reveals an intriguing feature of the *A. polyphaga* mimivirus NDK sequence, where this loop appears shorter. To address this oddity, the activity of the purified recombinant enzyme has been verified (data not shown) and we are currently assaying its specificity. The structural analysis of the *A. polyphaga* mimivirus NDK should thus provide a better understanding of its molecular function and therefore of its functional role in the context of *A. polyphaga* infection.
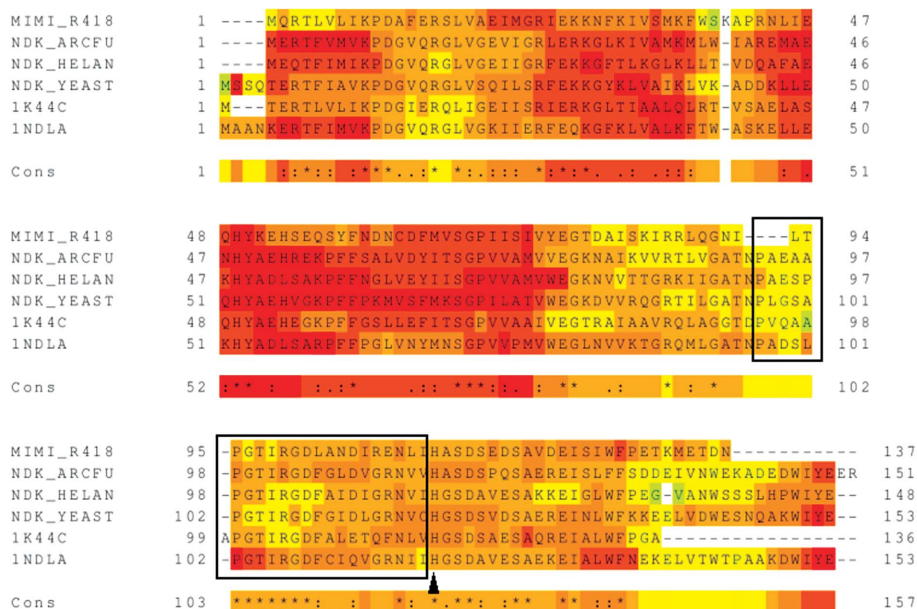
Figure 2
Multiple alignment of the *A. polyphaga* mimivirus NDK sequence (MIMI_R418) with structural homologues and related eukaryotic sequences. 1K44C and 1NDLA correspond to the PDB structures 1k44 chain *C* and 1ndl chain *A*, respectively. Swiss-Prot accession names are given for the homologous NDK sequences used in the alignment (NDK_ARCFU, *Archeoglobulus fulgidus*; NDK_HELAN, the sunflower *Helianthus annuus*; NDK_YEAST, *Saccharomyces cerevisiae*). The black box corresponds to the Kpn loop known to be involved in the interaction with the substrate and oligomerization state. The histidine corresponding to NDK's prosphophohistidine intermediate is marked by a black triangle. The multiple alignment combining structural and sequence information was generated using *3DCoffee* software (http://igs-server.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi; Poirot *et al.*, 2004).

crystallization communications

References

Abergel, C., Chenivesse, S., Byrne, D., Suhre, K., Arondel, V. & Claverie, J. M. (2005). *Acta Cryst.* F**61**, 212–215.

Abergel, C., Coutard, B., Byrne, D., Chenivesse, S., Claude, J. B., Deregnaucourt, C., Fricaux, T., Gianesini-Boutreux, C., Jeudy, S., Lebrun, R., Maza, C., Notredame, C., Poirot, O., Suhre, K., Varagnol, M. & Claverie, J. M. (2003). *J. Struct. Funct. Genomics*, **4**, 141–157.

Audic, S., Lopez, F., Claverie, J. M., Poirot, O. & Abergel, C. (1997). *Proteins*, **29**, 252–257.

Chen, Y., Morera, S., Mocan, J., Lascu, I. & Janin, J. (2002). *Proteins*, **47**, 556–557.

Chiadmi, M., Morera, S., Lascu, I., Dumas, C., Le Bras, G., Veron, M. & Janin, J. (1993). *Structure*, **1**, 283–293.

Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M. & Abergel, C. (2004). *Nucleic Acids Res.* **32**, W606–W609.

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

Janin, J., Dumas, C., Morera, S., Xu, Y., Meyer, P., Chiadmi, M. & Cherfils, J. (2000). *J. Bioenerg. Biomembr.* **32**, 215–225.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J. M. & Raoult, D. (2003). *Science*, **299**, 2033.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Parks, R. E. Jr & Agarwal, R. P. (1973). *The Enzymes*, 3rd ed., edited by P. D. Boyer, Vol. 8, pp. 307–334. New York: Academic Press.

Poirot, O., Suhre, K., Abergel, C., O'Toole, E. & Notredame, C. (2004). *Nucleic Acids Res.* **32**, W37–W40.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. & Claverie, J. M. (2004). *Science*, **306**, 1344–1350.