

PROTEIN STRUCTURE REPORT

Crystal structure of AFV1-102, a protein from the acidianus filamentous virus 1

Jenny Keller,¹ Nicolas Leulliot,¹ Bruno Collinet,¹ Valerie Campanacci,² Christian Cambillau,² David Pranghisvilli,³ and Herman van Tilbeurgh^{1*}

¹Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, IFR115, UMR8619-CNRS, 91405 Orsay, France

²Architecture et Fonction des Macromolécules Biologiques, CNRS and Universités d'Aix-Marseille I & II, UMR 6098, Case 932, 163 avenue de Luminy, 13288 Marseille cedex 9, France

³Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrémophiles, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Received 6 November 2008; Revised 18 November 2008; Accepted 24 November 2008

DOI: 10.1002/pro.79

Published online 10 February 2009 proteinscience.org

Abstract: Viruses infecting hyperthermophilic archaea have intriguing morphologies and genomic properties. The vast majority of their genes do not have homologs other than in other hyperthermophilic viruses, and the biology of these viruses is poorly understood. As part of a structural genomics project on the proteins of these viruses, we present here the structure of a 102 amino acid protein from acidianus filamentous virus 1 (AFV1-102). The structure shows that it is made of two identical motifs that have poor sequence similarity. Although no function can be proposed from structural analysis, tight binding of the gateway tag peptide in a groove between the two motifs suggests AFV1-102 is involved in protein protein interactions.

Keywords: archaeal virus; function; structure; fold evolution; gateway tag

Introduction

Double strand DNA viruses from hyperthermophilic Archaea show exceptionally diverse and unusual morphotypes. Moreover, analysis of their genomes revealed that they are unrelated to other known viruses [reviewed in refs. 1–4]. Based on their morphological and genomic properties, they have been classified in seven novel viral families: rod-shaped *Rudiviridae*,⁵ filamentous *Lipothrixviridae*,^{6–9} spindle-shaped *Fuselloviridae*,^{10–12} droplet-shaped *Gutta-*

viridae,¹³ spherical *Globuloviridae*,^{14,15} bottle-shaped *Ampullaviridae*,¹⁶ and two-tailed *Bicaudaviridae*.¹⁷ Three more hyperthermophilic archaeal viruses have still not been classified: the icosahedral STIV,¹⁸ spindle-shaped STSV1¹⁹ and PSV.²⁰ The genomes of most isolated archaeal viruses have been sequenced showing that sequence similarities between genes of the different archaeal viral families are generally limited and most predicted genes yield good matches only with those of other members of the same family.⁴ Despite this wealth of genomic information, the biological knowledge of the biology of these viruses is still very limited and the function of the majority of the proteins encoded by their genomes is unknown. In cases where no functional information can be retrieved from sequence analysis, 3D-structure determination may offer an alternative and provide testable hypothesis on

Grant sponsor: Agence Nationale de la Recherche; Grant number: NT05-2_41674.

*Correspondence to: Herman van Tilbeurgh, Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, IFR115, UMR8619-CNRS, 91405 Orsay, France. E-mail: herman.van-tilbeurgh@u-psud.fr

Table I. Data Collection and Refinement Statistics

	SeMet	Native
Space group	P6 ₂ 22	P6 ₂ 22
Unit cell a,b,c (Å)	62.36, 62.36, 111.27	62.89, 62.89, 110.84
Resolution (Å)	38.75–2.5 (2.64–2.5)	55.38–1.95 (2.06–1.95)
Total number of reflexions	51,952 (7457)	347,114 (54,335)
Number of unique reflexions	4860 (666)	10,006 (1431)
Multiplicity	10.7 (11.2)	34.7 (38)
Rmerge	0.11 (0.65)	0.11 (0.59)
I/σI	5.4 (1.1)	4.4 (0.8)
Overall completeness	100 (100)	99.5 (100)
R/R free		21.3/27.4
rmsd bond (Å)		0.011
rmsd angle (°)		1.31
Ramachandran plot (%)		
Favored		99.1
Allowed		0.9

biochemical function. Our decision to embark upon a structural genomics project on proteins from *Lipothrixviridae* family viruses has two motivations: (1) the discovery of new protein folds (many of the protein sequences are orphans and for the vast majority no reliable structural model can be proposed) and (2) obtaining functional information through 3D-structure determination of the proteins. Here, we present the 3D crystal structure of a small protein (102 amino acids) encoded by acidianus filamentous virus 1 (AFV1-102), which has no sequence homologues and hence no predictable structure and function. We found that AFV1-102 consists of a repeat of a four-stranded β -sheet motif packed against a α helix, together forming a $\beta\alpha\beta$ sandwich. The unexpected presence of the 19 N-terminal residues belonging to the linker between the protein and the 6 \times histidine purification tag in a groove enclosed by these two motifs suggests that AFV1-102 is involved in protein protein interactions.

Results and Discussion

Overall structure of AFV1-102

AFV1-102 crystallized in space group P6₂22 and the structure was solved at 1.95 Å resolution. The statistics on data collection and refinement are provided in Table I. Gel filtration chromatography showed that AFV1-102 is a monomer in solution (results not shown). The asymmetric unit contains a single copy of the AFV1-102 protein. AFV1-102 contains two four-stranded β sheets and three α helices forming an overall $\beta\alpha\beta$ sandwich fold [Fig. 1(A,B)]. No structural analogs could be identified using DALI or the SSM server of EBI (<http://www.ebi.ac.uk/msd-srv/ssm/>). The $\beta\alpha\beta$ sandwich can also be described as an assembly of two motifs with quasi identical structures that are related by pseudo twofold symmetry [Fig. 1(B)]. Each motif is composed of a α helix followed by a four stranded β sheet. One sheet is composed of $\beta_1\beta_2\beta_7\beta_8$

(motif 1) and the other of $\beta_3\beta_4\beta_5\beta_6$ (motif 2). AFV1-102 is not a repetition of the two motifs, since the second continuous motif is inserted between the second (β_2) and third strand (β_7) of motif 1. The main difference between the two motifs consists of an insertion of a third 14 residue helix (α_2) between β_4 and β_5 in motif 2 not present in motif 1 [Fig. 1(C)]. Helix α_3 (motif 1) and α_1 (motif 2) connect two strands from different β sheets (α_1 connects β_2 to β_3 and α_3 connects β_6 to β_7). Both motifs pack together on their helical sides. Helices α_1 and α_3 have the same length and pack against the sheets via hydrophobic interactions. Furthermore, both helices are kinked near their C terminus. The residues in the kink of helix α_1 form main chain hydrogen bonds with the residues in the connection to β_3 . In addition, a hydrogen bond between the NH1 of Arg43 located in the linker between β_4 and α_2 , and the main chain carbonyl of Lys25 at C terminus of α_1 may further stabilize the kink. The kink in helix α_3 is stabilized by very similar interactions: main chain hydrogen bonds between the C terminus of the helix and the α_3 β_7 connections and a hydrogen bond between the NH1 of the C terminal Arg95 and the carbonyl of a Lys78 at C terminus of α_3 . Despite very low sequence identity (10%) the two motifs superpose well with an rms deviation of 1.34 Å for 29 aligned C α positions [Fig. 1(C)]. The identical structures of the two motifs together with the very similar ways their helices are kinked and stabilized suggests that they originate from a common ancestral motif. AFV1-102 probably has evolved by duplication of an ancestral motif and subsequent insertion of motif 2 into motif 1.

The gateway tag is firmly bound

No structural analogs could be identified for either the intact AFV1-102 protein or the separate motifs using DALI or the SSM server of EBI. Since no homologs are identified for AFV1-102, combining structural and

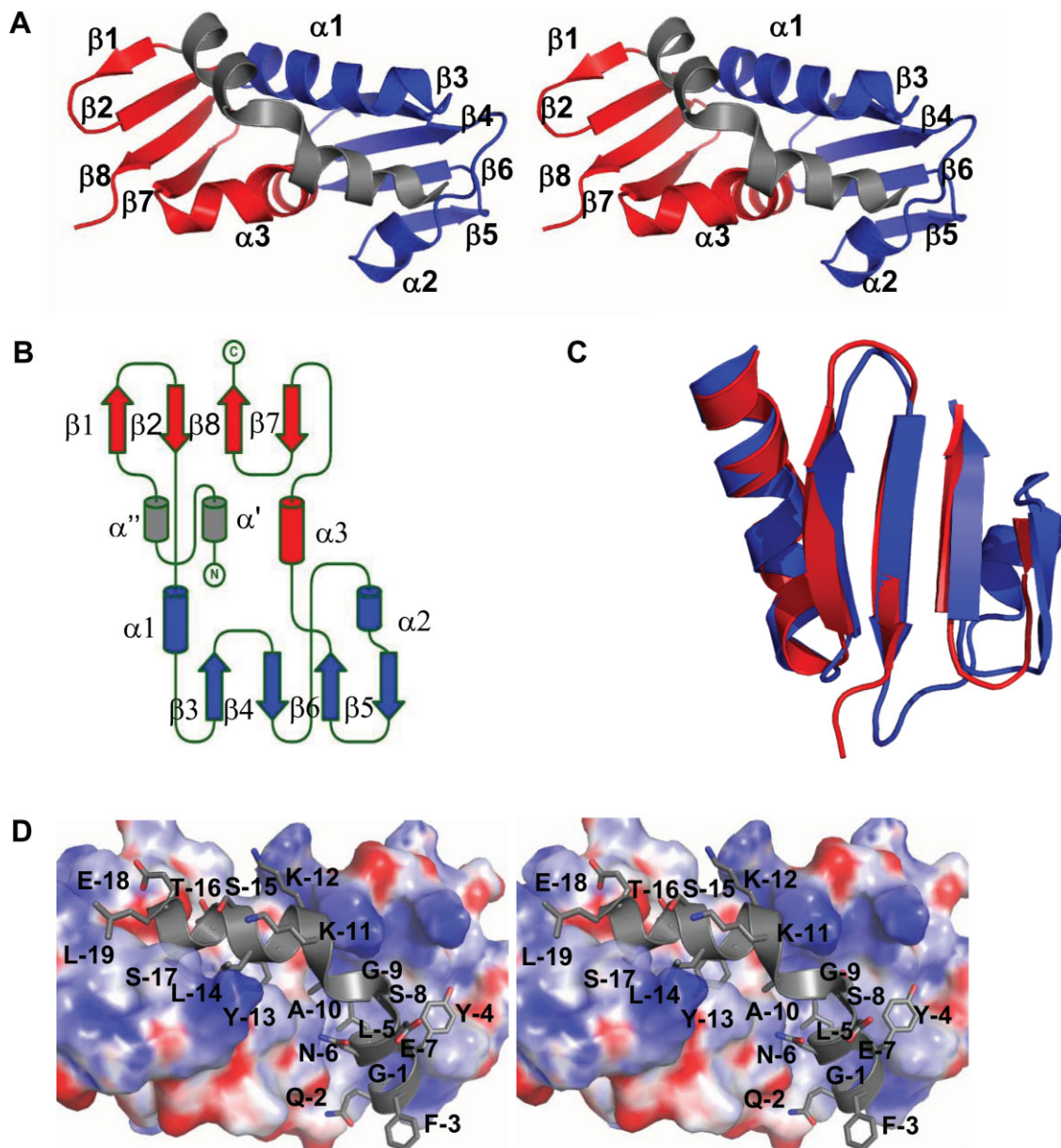


Figure 1. Structure of AFV1-102. (A) Stereo view of the cartoon representation of AFV1-102. The two motifs of the protein are colored in red and blue, the N-terminal tag is colored in gray. The secondary structure elements are indicated. (B) Topology of AFV1-102. Motifs and tag are colored as in A (diagram performed by Topdraw²¹). (C) Superposition of the two motifs of AFV1-102 in cartoon representation. (D) Focused stereo view on the interaction between the tag (in cartoon and sticks) and AFV1-102 (in electrostatic surface representation with 20% transparency). The residues of the tag are labeled with a one letter code (electrostatic surface calculation performed by APBS program²²).

sequence information could not be used to infer biochemical function. However, by serendipity the tag sequence of the genetic expression construct is strongly bound to the core protein in the crystal, suggesting that AFV1-102 is involved in protein protein interactions. The recombinant construction of AFV1-102 that yielded crystals still contains 30 residues at the N terminus encompassing the histidine tag, the attB site from the Gateway recombination system and the Tev protease cleavage site. After refinement of the

structure, unexplained strong residual electron density was observed in a canyon lying between the two motifs. We could readily accommodate this residual density by a peptide that corresponds to 19 residues of the tag (sequence LESTSLYKKAGSENLYFQG, numbered from -19 to -1 in the structure). This peptide adopts a kinked helical conformation and packs snugly into the groove [Fig. 1(D)]. Seven residues from the tag (Leu-19, Ser-17, Thr-16, Leu-14, Tyr-13, Tyr-4, Leu-5) make contacts with the protein core, 3 of which are

entirely buried (Ser-17, Tyr-13, Leu-5) [Fig. 1(D)]. Leu-19 interacts with Val44 on the edge of the groove; Ser-17 is inserted in a cleft made by Ile23, Leu26, Val28, and Trp49. In addition, the N ϵ 1 from Trp49 makes a hydrogen bond with the O γ of Ser-17. The methyl group of Thr-16 contacts Trp49, Leu52, and Leu72, whereas the O γ 1 and main chain N make hydrogen bonds with Asp48 OD2. Leu-14 mainly contacts Val22, Ile23, Leu26, and the aliphatic chain of Lys25. Tyr-13 is deeply inserted in the hydrophobic core of AFV1-102 in a cleft made by Met19, Val22, Ile23, Trp49, Phe69, and Val73. The OH of Tyr-13 points at the O of Met19 with which it makes a hydrogen bond. Leu-5 contacts Ile8, Ile10, Phe15, Met19, and Leu92. The aromatic cycle of Tyr-4 is not buried but makes contacts with Val3, Ile8, and Phe81 at the edge of the groove. At the C terminus of the tag Gln-2 NE2 makes hydrogen bonds with Ser18 OG and Glu14 O. Tyr-13 occupies a central position in the groove where it is inserted in a very hydrophobic pocket.

The Gateway recombination system is now routinely used for high throughput cloning. This cloning strategy adds a long tag to the target sequence and is usually cleaved after protein purification for structural studies. In our general structural genomics strategy, we decided to keep the tag to avoid the uncertain and time costly protease cleavage steps. Usually protein tags are disordered in the crystal, but this is clearly not the case in the present crystal structure. We could unambiguously determine the conformation of the gateway tag.

To determine the structure of AFV1-102 without the gateway tag, we tried two strategies. We incubated the tagged protein with Tev protease and alternatively we performed the cloning of AFV1-102 with a 6His tag attached at the C terminus. In both case it was not possible to produce soluble protein: the protein precipitated during tag cleavage and the 6His tag version of AFV1-102 was produced as inclusion bodies. This clearly shows that the gateway tag helps to solubilize the protein.

Search for an AFV1-102 partner

The unusual structuring of the tag and the many interactions it engages with the protein, suggests that the groove of AFV1-102 may be used for interaction with a protein partner. A Blast search with the tag-sequence (LESTSLYKKAGSENLYFQG) did not reveal nontrivial hits. However, in one of the proteins encoded by the AFV1 genome (AFV1-94), we identified the presence of a pentapeptide with sequence TSMYK. This peptide is very similar to that part of the gateway peptide (TSLYK) that strongly interacts with the core of AFV1-102. Experiments are underway to test whether AFV1-94 and AFV1-102 interact in solution. Straightforward testing of this interaction was hampered by the insolubility of nontagged AFV1-102.

Methods

Cloning, expression, and purification

The sequence of AFV1-102 was amplified by PCR from cDNA using two primers containing the attB sites of the Gateway recombination system (Invitrogen). The cDNA was cloned into the pDEST17 plasmid using the Gateway technology (Invitrogen).²³ A TEV protease cleavage site encoding sequence was inserted between the attB1 and the AFV1-102 gene. Expression was done at 37°C using the *E.coli* Rosetta (DE3) pLysS strain and the 2xYT medium (BIO 101 Inc.). When the cell culture reached an OD₆₀₀ nm of 1, overnight induction at 15°C was performed with 0.5 mM IPTG (Sigma). Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM β -mercaptoethanol). Cell lysis was completed by sonication and the lysate was heated for 20 min at 50°C before centrifugation at 20,000 rpm. The soluble fraction was loaded on a NiNTA column (Qiagen Inc.) equilibrated with buffer A. The protein was eluted with imidazole and loaded on a Superdex75 column (Amersham Pharmacia Biotech) equilibrated against buffer A with 10 mM β -mercaptoethanol. Selenomethionine substituted protein was produced and purified as the native protein. The homogeneity of the proteins was checked by SDS-PAGE.

Structure resolution

AFV1-102 native and seleno-substituted crystals were grown from a 1:1 μ L mixture of protein (14 mg/ml) with 1.5–2M ammonium sulfate, 0.1M HEPES pH 7.5, 2% PEG400, using hanging drop method at 23°C. For cryoprotection, crystals were soaked in a mixture of mother liquor and 30% glycerol. Crystals were then flash frozen at 100 K.

X-ray diffraction data were collected from a SeMet substituted AFV1-102 crystal on beamline BM30A (ESRF) at the Se K-edge. The crystal diffracted to 2.5 Å and belongs to the P₆,22 space group with two molecules per asymmetric unit, corresponding to 41% solvent content. Native data were collected on the ID29 beamline (ESRF) to 1.95 Å resolution. Crystal belongs to P₆,22 space group with one molecule per asymmetric unit, corresponding to 41.7% solvent content. Data processing was carried out with the program MOSFLM²⁴ and scaling and merging with SCALA.²⁵

The structure was solved using SAD X ray diffraction data at 2.5 Å resolution. Four Selenium atom sites were found with the program SHELXD²⁶ in the 20–3 Å resolution range. These sites were used for phasing with the program SOLVE.²⁷ After solvent flattening with the program RESOLVE,²⁷ the quality of the electron density map allowed automated construction of 90% of the model. This model was then refined against the 1.95 Å data set with the Arp/Warp²⁸ program that allowed automated building of the missing residues. The model was refined with REFMAC₅²⁹ and

manually corrected using O.³⁰ 19 residues of the tag and the 95 first residues are well defined in electron density map and fall within the allowed regions of the Ramachandran plot, as defined by Molprobit.³¹

Conclusions

The structure of AFV1-102 for which no homologues could be identified reveals a new $\beta\alpha\beta$ sandwich that was created by duplication of two almost identical motifs. No biological function could be deduced from the structure, but very clear binding of the tag peptide into an interdomain groove suggests that AFV1-102 interacts with a protein partner.

Coordinates

Coordinates and structure factors have been deposited in the Protein Data Bank with the accession code [2WB6](#).

References

- Ortmann AC, Wiedenheft B, Douglas T, Young M (2006) Hot Crenarchaeal viruses reveal deep evolutionary connections. *Nat Rev Microbiol* 4:520–528.
- Prangishvili D, Garrett RA (2005) Viruses of hyperthermophilic *Crenarchaea*. *Trends Microbiol* 13:535–542.
- Prangishvili D, Forterre P, Garrett RA (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* 4:837–848.
- Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* 117:52–67.
- Prangishvili D, Klenk HP, Jakobs G, Schmiechen A, Hanselmann C, Holz I, Zillig W (1998) Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. *J Biol Chem* 273:6024–6029.
- Janekovic D, Wunderl S, Holz I, Zillig W, Gierl A, Neumann H (1983) TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic sulfur reducing archaeobacterium *Thermoproteus tenax*. *Mol Gen Genet* 192:39–45.
- Bettstetter M, Peng X, Garrett RA, Prangishvili D (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. *Virology* 315:68–79.
- Vestergaard G, Aramayo R, Basta T, Haring M, Peng X, Brugger K, Chen L, Rachel R, Boisset N, Garrett RA, Prangishvili D (2008) Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. *J Virol* 82:371–381.
- Schleper C, Kubo K, Zillig W (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci USA* 89:7645–7649.
- Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res Microbiol* 154:295–302.
- Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. *J Virol* 78:1954–1961.
- Arnold HP, Ziese U, Zillig W (2000) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* 272:409–416.
- Haring M, Peng X, Brugger K, Rachel R, Stetter KO, Garrett RA, Prangishvili D (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: a novel virus family, the Globuloviridae. *Virology* 323:233–242.
- Ahn DG, Peng X, Brugger K, Rachel R, Stetter ICC, Garrett RA, Prangishvili D (2006) TTSV1, a new virus-like particle isolated from the hyperthermophilic crenarchaeote *Thermoproteus tenax*. *Virology* 351:280–290.
- Haring M, Rachel R, Peng X, Garrett RA, Prangishvili D (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. *J Virol* 79:9904–9911.
- Haring M, Vestergaard G, Rachel R, Chen L, Garrett RA, Prangishvili D (2005) Virology: independent virus development outside a host. *Nature* 436:1101–1102.
- Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci USA* 101:7716–7720.
- Xiang X, Chen L, Huang L, Luo Y, She Q, Huang L (2005) *Sulfolobus tengchongensis* spindle-shaped virus STSV1: virus-host interactions and genomic features. *J Virol* 79:8677–8686.
- Geslin C, Le Romancer M, Gaillard M, Erauso G, Prieur D (2003) Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. *Res Microbiol* 154:303–307.
- Leslie AGW (1992) Joint CCP4 and EACMB newsletter protein crystallography. Warrington, United Kingdom: Daresbury Laboratory, p 26.
- Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19:311–312.
- Holst M, Saied F (1995) Numerical solution of the non-linear Poisson-Boltzmann equation: developing more robust and efficient methods. *J Comput Chem* 16:337–364.
- Walhout AI, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, Vidal M (2000) GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* 328:575–592.
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62(Part 1):72–82.
- Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58(Part 10, Part 2):1772–1779.
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55(Part 4):849–861.
- Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56(Part 8):965–972.
- Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47:110–119.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Part 3):240–255.
- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by $C\alpha$ geometry: phi, psi, and $C\beta$ deviation. *Proteins* 50:437–450.