STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Cloning, expression, purification, characterization, crystallization and X-ray crystallographic analysis of recombinant Der f 21 (rDer f 21) from *Dermatophagoides farinae*

Sze Lei Pang,[a] Kok Lian Ho,[b] Jitka Waterman,[c] Aik-Hong Teh,[d] Fook Tim Chew[e] and Chyan Leong Ng[a]*

[a]Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia, [b]Department of Pathology, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia, [c]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, England, [d]Centre for Chemical Biology, Universiti Sains Malaysia, 10 Persiaran Bukit Jambul, 11900 Bayan Lepas, Penang, Malaysia, and [e]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore.
*Correspondence e-mail: clng@ukm.edu.my

Received 6 August 2015
Accepted 29 September 2015

Edited by R. Sankaranarayanan, Centre for Cellular and Molecular Biology, Hyderabad, India

Keywords: *Dermatophagoides farinae*; rDer f 21; house dust-mite allergen.

Supporting information: this article has supporting information at journals.iucr.org/f

Dermatophagoides farinae is one of the major house dust mite (HDM) species that cause allergic diseases. N-terminally His-tagged recombinant Der f 21 (rDer f 21), a group 21 allergen, with the signal peptide truncated was successfully overexpressed in an *Escherichia coli* expression system. The purified rDer f 21 protein was initially crystallized using the sitting-drop vapour-diffusion method. Well diffracting protein crystals were obtained after optimization of the crystallization conditions using the hanging-drop vapour-diffusion method with a reservoir solution consisting of 0.19 *M* Tris–HCl pH 8.0, 32% PEG 400 at 293 K. X-ray diffraction data were collected to 1.49 Å resolution using an in-house X-ray source. The crystal belonged to the *C*-centered monoclinic space group *C*2, with unit-cell parameters a = 123.46, b = 27.71, c = 90.25 Å, β = 125.84°. The calculated Matthews coefficient (V_M) of 2.06 Å^3 Da^{-1} suggests that there are two molecules per asymmetric unit, with a solvent content of 40.3%. Despite sharing high sequence identity with Blo t 5 (45%) and Blo t 21 (41%), both of which were determined to be monomeric in solution, size-exclusion chromatography, static light scattering and self-rotation function analysis indicate that rDer f 21 is likely to be a dimeric protein.

1. Introduction

The house dust mite (HDM) is a major cause of allergic diseases, affecting 85% of asthmatics from North and South America, Europe, South East Asia and Australasia (Platts-Mills *et al.*, 1989). 33 groups of HDM protein allergens have been isolated and accepted by the World Health Organization/International Union of Immunological Societies Subcommittee of Allergen Nomenclature. Following the identification of the group 21 mite allergen Blo t 21 in *Blomia tropicalis* (Gao *et al.*, 2007), two other group 21 allergens isolated from *Dermatophagoides pteronyssinus* and *D. farinae*, Der p 21 (Weghofer *et al.*, 2008) and Der f 21 (Cui *et al.*, 2014), respectively, were subsequently reported. Previous immunological studies carried out on 494 atopic patients showed that Blo t 21 is able to elicit positive allergic response in 57.9% of patients (Gao *et al.*, 2007), whereas IgE towards Der p 21 was identified in 26% of sera from 30 *D. pteronyssinus*-allergic patients in Central Europe (Weghofer *et al.* 2008). These studies suggested the prevalence of group 21 mite allergens in causing allergic diseases.

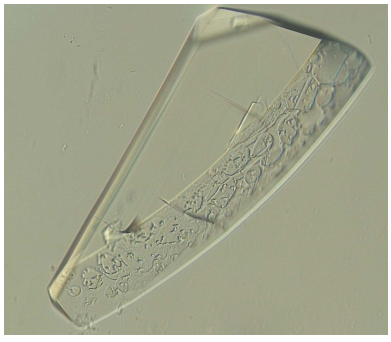

Table 1
Macromolecule-production information.

Source organism	*D. farinae*
DNA source	cDNA
Forward primer	5′-GCGGGATCCGAAGATAAATGGCGTAATGCAT-3′
Reverse primer	5′-GCCGAATTCTTAATCATCCGATTTTACAGCT-TTAAC-3′
Cloning vector	pET-M
Expression vector	pET-M
Expression host	*E. coli* strain Rosetta-gami (DE3)
Complete amino-acid sequence of the construct produced†	**MHHHHHHSSGLVPRGS**EDKWRNAFDHMLMEEFEE-KMDQIEHGLLMLSEQYKELEKTKSKELKEQIL-RELTIAENYLRGALKFMQQEAKRTDLNMFERY-NFETAVSTIEILVKDLAELAKKVKAVKSDD

† The His-tag fusion sequence at the N-terminus of rDer f 21 is shown in bold, followed by residues 25–136 of Der f 21.

Structural comparison of the NMR structure of Blo t 21 (PDB entry 2lm9) explicitly revealed its high structural similarity to the homologous group 5 proteins Blo t 5 (PDB entries 2jmh and 2jrk) and Der p 5 (PDB entry 3mq1) (Tan *et al.*, 2012; Naik *et al.*, 2008; Chan *et al.*, 2008; Mueller *et al.*, 2010). Despite having a highly conserved three antiparallel α-helical structure and a well conserved IgE epitope region, the group 21 mite allergen Blo t 21 shows low IgE cross-reactivity with group 5 mite allergens (Tan *et al.*, 2012). In addition, IgE cross-inhibition studies also showed a lack of IgE cross-reactivity of Blo t 21 with Der f 21 (Tan *et al.*, 2012). Nevertheless, a recent study reported a high IgE cross-reactivity of Der f 21 with Blo t 21 and other group 5 allergens (Kim *et al.*, 2015). High IgE cross-reactivity has also been found between Der f 21 and Der p 21 (unpublished results). Here, we describe the cloning, overexpression, purification, protein identification and characterization, crystallization and X-ray crystallographic analysis of recombinant Der f 21 (rDer f 21). Our interest is in solving the Der f 21 protein structure and determining the antigenic determinant. The crystal structure elucidation of Der f 21 will provide molecular details for IgE epitope mapping and will resolve the IgE cross-reactivity ambiguity among group 21 mite allergens.

2. Materials and methods

2.1. Protein production

Specific DNA primers were designed based on the cDNA sequence of Der f 21 (accession No. AY 800349) with the predicted signal peptide (the N-terminal 17 amino-acid residues) excluded. Using the forward primer 5′-GCGGGATCC-GAAGATAAATGGCGTAATGCAT-3′ and the reverse primer 5′-GCCGAATTCTTAATCATCCGATTTTACAGCT TTAAC-3′, Der f 21 with the N-terminal signal peptide and the preceding seven amino-acid residues removed was amplified by polymerase chain reaction (PCR). The amplified DNA fragment was then cloned between the BamHI and EcoRI restriction-endonuclease sites (underlined in the forward and reverse primer sequences, respectively) of a modified pET-32 vector (pET-M) to produce the pET-M-tDer f 21 clone. Notably, the rDer f 21 construct, which includes 16 residues from a 6×His-tag fusion and 112 residues of Der f 21 (residues 25–136), is slightly different from the previously reported Der f 21 construct (Cui *et al.*, 2014) consisting of residues 18–136 (Supplementary Table S1). The pET-M-tDer f 21 clone was transformed into the expression host *Escherichia coli* strain Rosetta-gami (DE3) to produce truncated rDer f 21 with an N-terminal 6×His-tag fusion (Table 1). Transformed bacterial cells were grown overnight in Luria–Bertani (LB) medium containing ampicillin (50 µg ml^{-1}) at 310 K. Overnight bacterial cultures were then inoculated and grown in 2 l LB medium until the OD_{600} reached 0.6. Recombinant protein expression was induced by the addition of 1 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) to the bacterial cultures, which were grown at 310 K for 4 h before harvesting by centrifugation at 5465*g*. The pellet was resuspended in binding buffer (20 m*M* Tris–HCl pH 7.9, 0.5 *M* NaCl, 20 m*M* imidazole, 20 m*M* β-mercaptoethanol) and lysed by sonication at an amplitude of 38% for 15 min (30 s pulse on and 30 s pulse off) followed by centrifugation at 14 000*g*. The supernatant containing the rDer f 21 protein was filter-sterilized using a 0.2 µm PVDF membrane filter and then applied onto an Ni–NTA-coupled HisTrap HP 5 ml column (GE Healthcare, UK) which had been pre-equilibrated with 15 ml binding buffer. The rDer f 21 protein was eluted using a linear gradient of washing buffer (20 m*M* Tris–HCl pH 7.9, 0.5 *M* NaCl, 0.5 *M* imidazole, 20 m*M* β-mercaptoethanol). Fractions containing soluble rDer f 21 protein were pooled and further purified by size-exclusion chromatography using a HiLoad 16/600 Superdex 75 pg gel-filtration column (GE Healthcare, UK) pre-equilibrated with size-exclusion buffer (20 m*M* Tris–HCl pH 7.9, 0.5 *M* NaCl, 20 m*M* β-mercaptoethanol).

2.2. Mass-spectrometric analysis

The Coomassie Brilliant Blue R-250-stained rDer f 21 SDS–PAGE gel band was excised and used for protein identification (First BASE Laboratories Sdn Bhd, Malaysia). Trypsin-digested peptides were extracted and analyzed by matrix-assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF MS) using a 5800 Proteomics Analyzer (Applied Biosystems–SCIEX; Bringans *et al.*, 2008). The protein of interest was identified with the *Mascot* sequence-matching software (Matrix Science) using the Ludwig NR Database. The analysis was performed by Proteomics International Pty Ltd, Australia. Independently, the molecular weight of rDer f 21 was also identified using a Bruker Daltonics ultraflex II TOF/TOF mass spectrometer (Agro Biotechnology Institute, Malaysia).

2.3. Static light scattering

rDer f 21 protein samples were prepared at three different concentrations (0.50, 1.08 and 1.61 mg ml^{-1}) in sample buffer (1 ml) consisting of 20 m*M* Tris–HCl pH 7.9, 50 m*M* NaCl. The intensity of the scattered light was measured at a single angle of 173° at 298 K using a Zetasizer Nano ZSP machine (Malvern, England). A Debye plot was generated from the

Table 2
Crystallization information.

Method	Sitting-drop vapour diffusion for initial crystal screening, hanging-drop vapour diffusion for crystal optimization
Plate type	96-well MRC plates for initial crystal screening, 24-well plates for crystal optimization
Temperature (K)	293
Protein concentration (mg ml^{-1})	∼10
Buffer composition of protein solution	20 m*M* Tris–HCl pH 7.9, 50 m*M* NaCl, 20 m*M* β-mercaptoethanol
Composition of reservoir solution	0.19 *M* Tris–HCl pH 8.0, 32% polyethylene glycol 400
Volume and ratio of drop	2 µl; 1:1 protein:reservoir solution
Volume of reservoir	80 µl for initial crystal screening, 1.0 ml for optimization

obtained data. The intercept of the Debye plot indicates the molecular weight of the protein sample and the slope represents the second virial coefficient, A_2. The Rayleigh equation used to obtain the molecular weight is

$$\frac{kc}{R_\theta} = \left(\frac{1}{M} + 2A_2c\right), \quad (1)$$

where $k = 4\pi^2 n_0^2 (\mathrm{d}n/\mathrm{d}c)^2/(\lambda^4 N_\mathrm{A})$, n_0 is the refractive index of the buffer, $(\mathrm{d}n/\mathrm{d}c)$ is the refractive-index increment, λ is the wavelength of light in vacuum, N_A is Avogadro's number, M is the protein molecular weight, c is the protein sample concentration and R_θ is the excess Rayleigh ratio of the protein sample. The Rayleigh ratio value can be obtained as $R_\theta = (I_\mathrm{s} - I_{\mathrm{s},0})/I_{\mathrm{s,T}}(n_0^2/n_\mathrm{T}^2)R_{\theta,\mathrm{T}}$, where I_s is the scattered light intensity of the solution, $I_{\mathrm{s},0}$ is the scattered light intensity of the solvent and $I_{\mathrm{s,T}}$, n_T and $R_{\theta,\mathrm{T}}$ are the scattered light intensity, the available refractive index and the Rayleigh ratio of toluene.

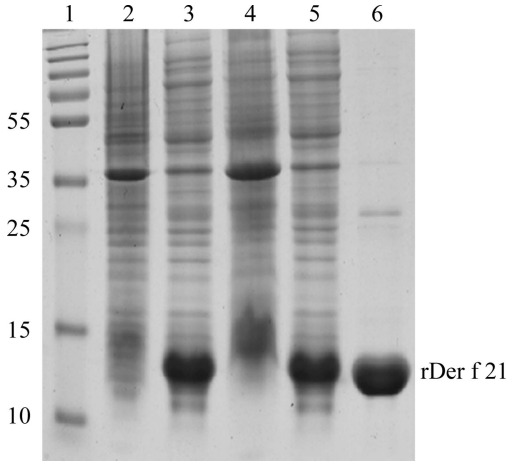

Figure 1
Overexpressed and purified rDer f 21 analyzed using 12% SDS–PAGE followed by Coomassie Blue staining. Lane 1, PageRuler Plus Prestained protein ladder (labelled in kDa; Thermo Scientific, USA). Lane 2, pellet of crude extract before IPTG induction as a negative control. Lanes 3–5, total protein, pellet and supernatant of crude extract after IPTG induction, respectively. Lane 6, purified rDer f 21 protein using Ni–NTA affinity chromatography.

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	MicroMax-007 HF
Wavelength (Å)	1.5418
Temperature (K)	100
Detector	R-AXIS IV^{++}
Crystal-to-detector distance (mm)	80
Rotation range per image (°)	0.5
Total rotation range (°)	185
Exposure time per image (s)	180
Space group	*C*2
a, *b*, *c* (Å)	123.46, 27.71, 90.25
α, β, γ (°)	90.00, 125.84, 90.00
Mosaicity (°)	0.16
Resolution range (Å)	19.09–1.49 (1.52–1.49)
Total No. of reflections	135651
No. of unique reflections	41017
Completeness (%)	99.5 (95.8)
Multiplicity	3.3 (2.5)
$\langle I/\sigma(I)\rangle$	22.4 (3.1)
R_meas†	0.031 (0.329)
Overall *B* factor from Wilson plot (Å^2)	17.5

† $R_\mathrm{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $N(hkl)$ is the multiplicity of reflection *hkl*.

2.4. Crystallization

The purified rDer f 21 protein was buffer-exchanged into crystallization buffer (20 m*M* Tris–HCl pH 7.9, 50 m*M* NaCl, 20 m*M* β-mercaptoethanol) and concentrated to 10 mg ml^{-1} using a Vivaspin 2 polyethersulfone concentrator fitted with a 3 kDa molecular-weight cutoff filter (Sartorius, Germany). Initial crystallization screening was carried out using the sitting-drop vapour-diffusion method in 96-well MRC crystallization plates (Molecular Dimensions, USA) with the PEGRx screen (Hampton Research, USA). Drops consisting of 1.0 µl rDer f 21 solution and 1.0 µl reservoir solution were equilibrated against 80 µl reservoir solution at 293 K. Initial crystal hits were obtained after approximately 48 h from a reservoir solution consisting of 0.1 *M* Tris–HCl pH 8.0, 30% polyethylene glycol (PEG) 400. Crystallization was further optimized in 24-well plates using the hanging-drop vapour-diffusion method at 293 K with drops that consisted of 1.0 µl protein solution and 1.0 µl reservoir solution (Table 2). Single rDer f 21 protein crystals suitable for X-ray diffraction analysis were obtained from 0.19 *M* Tris pH 8.0, 32% PEG 400. The best crystal of rDer f 21 grew to dimensions of approximately 600 × 300 × 80 µm after 4 d.

2.5. Data collection and processing

The rDer f 21 crystals were flash-cooled in liquid nitrogen without cryoprotectant. Preliminary diffraction data were collected at 100 K under a nitrogen-gas stream using a MicroMax-007 HF X-ray generator (Rigaku) with a copper anode coupled with VariMax HF optics and an R-AXIS IV^{++} image-plate area detector at a wavelength of 1.5418 Å. A total of 370 images were collected with an oscillation of 0.5° per image. The data-collection and processing statistics are summarized in Table 3. The data were indexed using *iMosflm* (Battye *et al.*, 2011) and were integrated and scaled with *XDS*

(Kabsch, 2010). *POINTLESS* (Evans, 2006) confirmed that the data were not twinned and that the crystal belonged to space group *C*2. The data were merged with *AIMLESS* (Evans & Murshudov, 2013). The general self-rotation function shows the presence of a noncrystallographic twofold axis, which indicates that rDer f 21 might exist as a dimer in the crystal.

3. Results and discussion

The rDer f 21 protein was purified to homogeneity as described in §2.1. The molecular weight of rDer f 21 was calculated to be approximately 15.2 kDa using the *ProtParam* tool (Gasteiger *et al.*, 2005). However, SDS–PAGE analysis showed the molecular weight of purified rDer f 21 to be approximately 13 kDa (Fig. 1). Mass spectrometry using MALDI-TOF/TOF MS confirmed the identity of the purified protein to be rDer f 21, with 41% of the peptide sequence matching that of Der f 21 (Supplementary Fig. S1). The results obtained using a Bruker Daltonics ultraflex II TOF/TOF mass spectrometer further confirmed that the molecular weight of the rDer f 21 protein was 15.2 kDa (Supplementary Fig. S2). The rDer f 21 protein was further characterized using size-exclusion chromatography (SEC) and static light scattering (SLC). Based on the retention time, the SEC results indicate that the rDer f 21 protein may exist either as a monomer or a dimer in solution, with a molecular weight greater than 17 kDa (Supplementary Fig. S3). Secondary-structure prediction using *Phyre*2 (Kelley *et al.*, 2015) suggests that rDer f 21 is likely to be an elongated helical bundle protein, similar to its homologues Blo t 21, Blo t 5 and Der p 5 (Supplementary Fig. S4). Elongated helical proteins have been shown to elute earlier in SEC, which results in the estimated molecular weight being higher by about 25–30% compared with the globular protein markers (Yousef, 2010). Hence, it is possible that rDer f 21 eluted earlier in the SEC in this study. Nonetheless, static light scattering, which showed Der f 21 with a molecular weight of ∼30 kDa, further suggests that rDer f 21 is likely to exist in a dimeric conformation in solution (Supplementary Fig. S5).

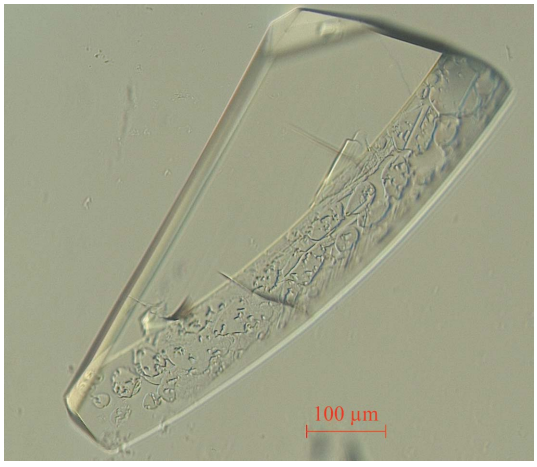

Figure 2
Crystal of rDer f 21 grown in reservoir solution consisting of 0.19 *M* Tris–HCl pH 8.0, 32% PEG 400.

Purified rDer f 21 (∼10 mg ml^{-1}) was subjected to crystallization screening. Crystals suitable for diffraction analysis were grown from an optimized reservoir solution composed of 0.19 *M* Tris–HCl pH 8.0, 32% PEG 400 (Fig. 2). The rDer f 21 protein crystals were flash-cooled without cryoprotectant and diffracted to 1.49 Å resolution (Fig. 3) on an in-house X-ray source. A native data set was collected with 99.5% completeness. Indexing by *POINTLESS* (Evans, 2006)

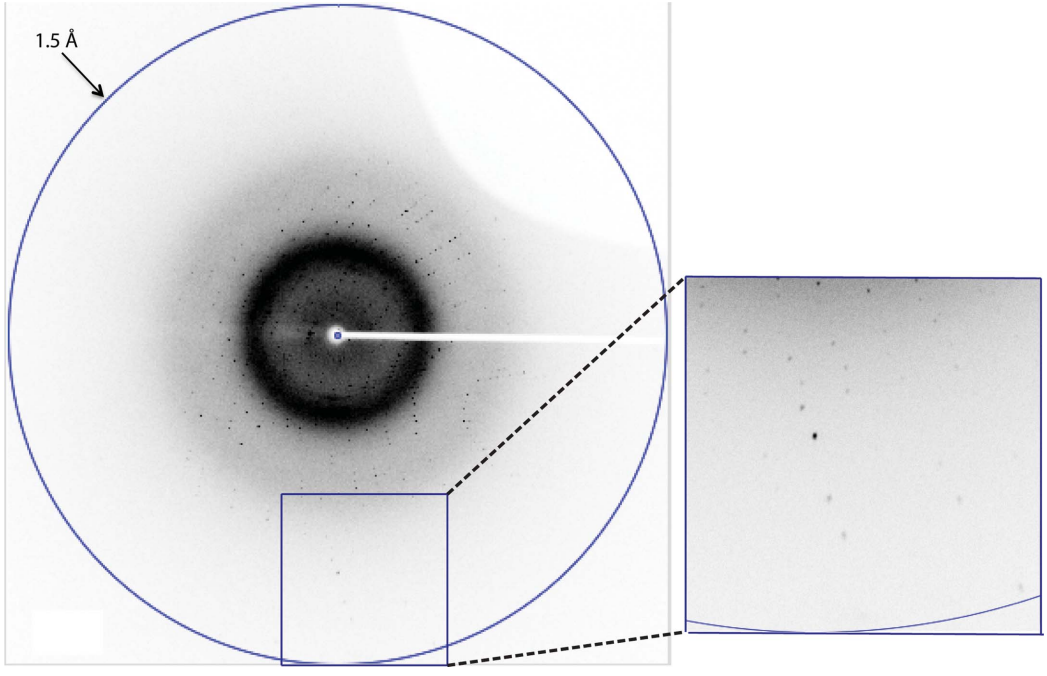

Figure 3
The rDer f 21 protein crystal diffracted to 1.49 Å resolution.

suggested that the crystal belonged to the *C*-centered monoclinic space group *C*2, with unit-cell parameters a = 123.46, b = 27.71, c = 90.25 Å, β = 125.84°. The crystallographic parameters and data-collection statistics are shown in Table 3. The calculated Matthews coefficient (V_M) of 2.06 Å^3 Da^{-1} (Matthews, 1968) indicates that there are two Der f 21 molecules per asymmetric unit, with an estimated solvent content of approximately 40.3%. A self-rotation function (Crowther, 1972) was calculated from the scaled data with a high-resolution data cutoff at 2.5 Å using *MOLREP* (Vagin & Teplyakov, 2010), as shown in Supplementary Fig. S6. In the κ = 180° section, the self-rotation map shows the twofold crystallographic axis of the assigned space group *C*2. The map also clearly shows two noncrystallographic symmetry (NCS) twofold axes located near the crystallographic axis, with corresponding self-rotation function peaks at approximately θ = 80°, $\varphi = \pm 100°$ in the χ = 180° section. The presence of a NCS twofold axis confirms our conclusion that rDer f 21 may be present as a dimer in solution, as indicated by size-exclusion chromatography and static light scattering. Compared with its monomeric homologues with high sequence identity, Blo t 5 and Blo t 21, rDer f 21 is likely to form a dimer, similar to its homologue Der p 21, which shares 70% sequence identity (Weghofer *et al.*, 2008). This suggests that Der f 21 and Der p 21 may share a similar stuctural conformation.

Currently, we are working towards the structural determination of rDer f 21 by molecular replacement. Comparing the rDer f 21 structure with the structures of homologous group 5 and group 21 allergens, we aim to resolve the IgE cross-reactivity ambiguity within group 21 mite allergens through structure comparison, site-directed mutagenesis and IgE-binding experiments.

Acknowledgements

The authors would like to acknowledge Universiti Kebangsaan Malaysia for financial support through the GUP-2014-010 grant. We thank Dr Gao Yun Feng from the National University of Singapore for the clone of the rDer f 21 gene. We thank Mr Ling Yok Ung (DKSH Technology, Malaysia) and Ms Norasfaliza Binti Rahmad (Agro Biotechnology Institute) for help with static light scattering and mass spectrometry. We also thank Dr Miguel Ortiz-Lombardía for helpful discussion of the crystallographic analysis.

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Bringans, S., Eriksen, S., Kendrick, T., Gopalakrishnakone, P., Livk, A., Lock, R. & Lipscombe, R. (2008). *Proteomics*, **8**, 1081–1096.
Chan, S. L., Ong, T. C., Gao, Y. F., Tiong, Y. S., Wang, D. Y., Chew, F. T. & Mok, Y. K. (2008). *J. Immunol.* **181**, 2586–2596.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Cui, Y., Jiang, Y., Ji, Y., Zhou, Y., Yu, L., Wang, N., Yang, L. & Zhang, C. (2014). *Am. J. Transl. Res.* **6**, 786–792.
Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Gao, Y. F., Wang, D. Y., Ong, T. C., Tay, S. L., Yap, K. H. & Chew, F. T. (2007). *J. Allergy Clin. Immunol.* **120**, 105–112.
Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. (2015). *Nature Protoc.* **10**, 845–858.
Kim, C.-R., Jeong, K. Y., Yi, M.-H., Kim, H.-P., Shin, H.-J. & Yong, T.-S. (2015). *Mol. Med. Rep.* **12**, 5467–5474.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mueller, G. A., Gosavi, R. A., Krahn, J. M., Edwards, L. L., Cuneo, M. J., Glesner, J., Pomés, A., Chapman, M. D., London, R. E. & Pedersen, L. C. (2010). *J. Biol. Chem.* **285**, 25394–25401.
Naik, M. T., Chang, C., Kuo, I., Kung, C. C.-H., Yi, F.-C., Chua, K.-Y. & Huang, T.-H. (2008). *Structure*, **16**, 125–136.
Platts-Mills, T. A. E. *et al.* (1989). *J. Allergy Clin. Immunol.* **83**, 416–427.
Tan, K. W., Ong, T. C., Gao, Y. F., Tiong, Y. S., Wong, K. N., Chew, F. T. & Mok, Y. K. (2012). *J. Biol. Chem.* **287**, 34776–34785.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Weghofer, M. *et al.* (2008). *Allergy*, **63**, 758–767.
Yousef, M. S. (2010). *J. Genet. Eng. Biotechnol.* **8**, 35–38.