The HKU Scholars Hub The University of Hong Kong 香港大學學術庫

Title	Protein-complex structure completion using IPCAS (Iterative Protein Crystal structure Automatic Solution)
Author(s)	Zhang, WZ; Zhang, H; Zhang, T; Fan, HF; Hao, Q
Citation	Acta Crystallographica Section D Biological Crystallography, 2015, v. 71, p. 1487-1492
Issued Date	2015
URL	http://hdl.handle.net/10722/214504
Rights	Creative Commons: Attribution 3.0 Hong Kong License

BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Protein-complex structure completion using *IPCAS* (*Iterative Protein Crystal structure Automatic Solution*)

Weizhe Zhang,[a,b] Hongmin Zhang,[a,c] Tao Zhang,[d] Haifu Fan[d] and Quan Hao[a]*

[a]Department of Physiology, University of Hong Kong, Hong Kong, [b]Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China, [c]Department of Biology and Shenzhen Key Laboratory of Cell Microenvironment, South University of Science and Technology of China, Shenzhen 518055, People's Republic of China, and [d]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China. *Correspondence e-mail: qhao@hku.hk

Received 21 November 2014
Accepted 2 May 2015

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

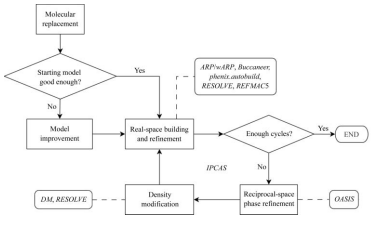
Keywords: *IPCAS*; protein complexes; iterative phasing; model building.

Protein complexes are essential components in many cellular processes. In this study, a procedure to determine the protein-complex structure from a partial molecular-replacement (MR) solution is demonstrated using a direct-method-aided dual-space iterative phasing and model-building program suite, *IPCAS* (*Iterative Protein Crystal structure Automatic Solution*). The *IPCAS* iteration procedure involves (i) real-space model building and refinement, (ii) direct-method-aided reciprocal-space phase refinement and (iii) phase improvement through density modification. The procedure has been tested with four protein complexes, including two previously unknown structures. It was possible to use *IPCAS* to build the whole complex structure from one or less than one subunit once the molecular-replacement method was able to give a partial solution. In the most challenging case, *IPCAS* was able to extend to the full length starting from less than 30% of the complex structure, while conventional model-building procedures were unsuccessful.

1. Introduction

Protein complexes have been found to be essential components of almost every cellular process (Hartwell *et al.*, 1999; Phizicky & Fields, 1995). Individual proteins can connect in the protein-interaction network through complexity and modularity and work together to accomplish a common function (Newman, 2006; Pereira-Leal *et al.*, 2006; Barabási & Oltvai, 2004). Proteins with more than one polypeptide or subunit are also found in many different protein families. Many protein complexes have been well characterized, among which the classical examples include haemoglobin (Hardison, 1996), tryptophan synthetase (Raboni *et al.*, 2009) and RNA polymerase (Hurwitz, 2005). Other well known multi-subunit proteins include metabolic enzymes, the DNA-replication complex, the nuclear pore complex, ribonucleoproteins *etc.*, which are usually called protein machines (Alberts & Miake-Lye, 1992). Another different kind of protein complex is a transient protein complex that controls numerous cellular processes, such as protein modification (phosphorylation, glycosylation and acylation *etc.*), transcription-complex recruitment and assembly, the transportation of proteins across membranes, chaperonin-aided native protein folding, translation-cycle regulation and cellular communication (Phizicky & Fields, 1995).

X-ray crystallographic structures of protein complexes can reveal high-resolution details of the structural interactions, give clues to interaction mechanisms and provide a basis and a



© 2015 International Union of Crystallography

Table 1
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Hha–H-NS	NDM-1–MBP	CD38–nanobody
Wavelength (Å)	0.9793	0.9792	0.9793
Resolution range (Å)	50–1.8 (1.86–1.80)	50–2.2 (2.28–2.20)	50–2.3 (2.38–2.30)
Space group	*P*1	*C*121	$P2_12_12_1$
Unit-cell parameters			
a (Å)	41.4	207.3	88.6
b (Å)	45.3	74.0	96.2
c (Å)	47.5	53.3	133.8
α (°)	68.4	90	90
β (°)	88.1	104.2	90
γ (°)	85.8	90	90
Total reflections	195794	417531	2876430
Unique reflections	28838 (2852)	72688 (6963)	51358 (5075)
Completeness (%)	97.55 (96.0)	98.8 (97.1)	99.7 (100.0)
Mean $I/\sigma(I)$	41.31 (3.38)	15.57 (2.33)	21.81 (3.39)
Wilson *B* factor (Å^2)	39.9	41.6	41.5
R_{sym} or R_{merge}	0.033 (0.309)	0.070 (0.465)	0.092 (0.576)
R_{work}	0.1800	0.1883	0.2075
R_{free}	0.2197	0.2349	0.2361
No. of non-H atoms	2621	4805	6094
Protein residues	261	598	727
R.m.s.d., bonds (Å)	0.011	0.009	0.011
R.m.s.d., angles (°)	1.38	1.15	1.35

template for many experimental and computational approaches such as interruption of interaction, binding-affinity maturation and the computational prediction or design of protein–protein interactions (Aloy *et al.*, 2005; Kortemme & Baker, 2004).

Identifying the structure of one of the subunits is a good starting point to study the whole complex. Here, we report a demonstration of *IPCAS* (*Iterative Protein Crystal structure Automatic Solution*), which derives the complete complex structure from a partial molecular-replacement (MR) solution.

The *IPCAS* pipeline is a direct-method-aided dual-space iterative phasing and model-building procedure which consists of SAD/SIR direct phasing and fragment extension (Hao *et al.*, 2000; Wang *et al.*, 2004) and partial-model extension without SAD/SIR information (He *et al.*, 2007). The graphical user interface (GUI) was published in 2010 (Zhang *et al.*, 2010). This study only utilizes the feature of partial model extension without SAD/SIR information. The iteration procedure involves (i) real-space model building and refinement by *ARP*/*wARP* (Langer *et al.*, 2008), *Buccaneer* (Cowtan, 2006), *phenix.autobuild* (Terwilliger *et al.*, 2008), *RESOLVE* (Terwilliger, 2003) and *REFMAC*5 (Murshudov *et al.*, 2011), (ii) direct-method-aided reciprocal-space phase refinement by *OASIS* (Zhang *et al.*, 2010) and (iii) phase improvement through density modification by *DM* (Winn *et al.*, 2011) or *RESOLVE* (Terwilliger, 2000). To apply a direct method in reciprocal-space phase refinement based on the built model, the conventional direct-method probability formula has been revised (He *et al.*, 2007).

To evaluate the *IPCAS* pipeline, the following three test cases were carefully selected from a practical perspective and one case was randomly selected from the PDB to demonstrate the capability of *IPCAS* to deal with low-resolution data.

(i) The Hha–H-NS complex. α-Haemolysin expression-modulating protein (Hha) and histone-like nucleoid structuring protein (H-NS) are both nucleoid-associated proteins from bacteria. In *Escherichia coli*, Hha can bind to H-NS and the formation of the complex is essential for silencing the expression of the toxin α-haemolysin (Ali *et al.*, 2013; Nieto *et al.*, 2002). In this work, the Hha–H-NS$_{64}$ complex (1.8 Å resolution) represents a difficult case in which the search model (an H-NS$_{46}$ dimer; PDB entry 1ov9; Cerdan *et al.*, 2003) shares 65% sequence identity with the H-NS part of the target protein and is less than 30% of the overall complex structure in size (details of data collection and test results are listed in Tables 1 and 2). This structure was previously unknown and the full structure with biological functions will be published elsewhere.

(ii) NDM-1 protein with an MBP tag. Maltose-binding protein (MBP) is often used in *E. coli* expression systems to help in the folding and to increase the solubility of the target protein and is also used as an affinity tag to help in purification (Lebendiker & Danieli, 2011). In crystallography, MBP can also help to phase the MBP-fusion protein by the molecular-replacement (MR) method. New Delhi metallo-β-lactamase-1 (NDM-1) is one of the class B β-lactamases. The first NDM-1 structure was solved by single-wavelength anomalous dispersion (SAD) with data collected at the zinc absorption edge (Zhang & Hao, 2011). In this complex at 2.2 Å resolution, NDM-1 composes about 36% of the length of the sequence and MBP makes up 57% of the length of the sequence. Here, we demonstrated that the NDM-1 structure could be obtained by *IPCAS* starting from the MR solution using MBP as a search model. Similarly, using the NDM-1 structure as an MR search model, the longer MBP part could also be built by *IPCAS*.

(ii) CD38–nanobody complex. Antibody–antigen interaction is a central part of the immune system and a typical form of protein complexes. However, the structural variability of antibodies leads to difficulties in structural determination (Davies *et al.*, 1990). Although the use of the MR method to determine the structures of antibodies or antibody–antigen complexes has long been known (Brünger, 1993), it is still not straightforward to determine the structure of novel or designed antibodies in practice. CD38 is a multi-functional enzyme involved in many cellular processes (Lee, 2006) and is a promising therapeutic target for antibody therapy in cancer treatment (Chillemi *et al.*, 2013; Stevenson, 2006). In this study, we report the structure of a CD38–nanobody complex (2.3 Å resolution) obtained using the CD38 molecule (55% of the sequence) as an MR search model and the designed nanobody structure could be built by *IPCAS*. The biological functions of this complex will be reported elsewhere.

(iv) PDB entry 4owr, a randomly selected lower resolution case (Quan *et al.*, 2014). Model building at low resolution is always a challenge owing to the limited number of observations compared with the large number of parameters to be defined (Karmali *et al.*, 2009). The 4owr complex (3.15 Å) was randomly selected from the PDB to test how *IPCAS* was able to deal with low-resolution data. The 4owr complex consists

of three chains: chain *A* (57%), chain *B* (10%) and chain *C* (32%), in which chain *A* was taken as the starting model for *IPCAS* model completion after *Phaser* had placed it within the unit cell. Finally, we demonstrated that model completion at low resolution can also be successful.

2. Method

Before the structure-completion procedure, a molecular-replacement (MR) solution from a partial model should be identified. If the MR solution is not of sufficient quality to allow subsequent model building, manual improvement of the solution structure might be necessary, such as manual refinement based on the fit between the model structure and the electron-density map using *Coot* (Emsley *et al.*, 2010; a detailed procedure with an example will be described in §3.1). The MR solution structure and the initial map from this structure are then taken as a starting point for *IPCAS* to perform model completion by iteration, including real-space refinement, direct-method-aided reciprocal-space refinement and model building, with sequence and solvent content assigned. The entire workflow is shown in Fig. 1. The individual program parameters that *IPCAS* calls can be adjusted in a graphical user interface (GUI); however, in this study the default values of these parameters are used in all four cases. *IPCAS* usually runs ten iterations of real-space refinement, direct-method-aided reciprocal-space refinement and model building, and the user can then decide whether or not more iterations are required depending on whether the results are converging.

Table 2
Model-completion test results.

Case	Hha–H-NS	NDM-1–MBP		CD38–nanobody	4owr
Residues in asymmetric unit	302	642		828	586
MR model (PDB entry)	H-NS dimer (1ov9)	MBP monomer (1y4c)	NDM-1 monomer (3q6x)	CD38 monomer (3ops)	Chain *A* (4owr)
Sequence identity between starting and search models (%)	65	100	100	100	100
No. of residues in the starting model (% of the complex)	89 (29.5%)	367 (57.2%)	229 (35.7%)	456 (55.0%)	335 (57.0%)
Phaser (*CCP*4)					
RFZ	9.3	8.1	24.2	6.8	3.9
TFZ	†	16.1	16.0	26.6	12.6
LLG	74	2383	1305	1847	2093
Starting FOM	0.456	0.522	0.426	0.508	0.651
IPCAS					
No. of residues built	269 (89%)	603 (94%)	602 (94%)	747 (90%)	566 (97%)
R factor/R_{free} (%)	24.6/28.9	22.7/27.9	23.5/27.9	24.2/28.2	25.7/32.2
No. of $\Delta C^{\alpha} < 1$ Å‡	250	586	588	712	429
Buccaneer (*CCP*4)					
No. of residues built	188 (62%)	606 (94%)	628 (98%)	773 (93%)	394 (67%)
R factor/R_{free} (%)	44.3/51.8	25.8/31.6	24.7/29.7	26.8/31.0	35.9/45.5
No. of $\Delta C^{\alpha} < 1$ Å‡	76	564	583	707	348

† The TFZ score is not applicable in space group *P*1. ‡ ΔC^{α} is the positional deviation of C^{α} atoms in the built model from those of the final structure.

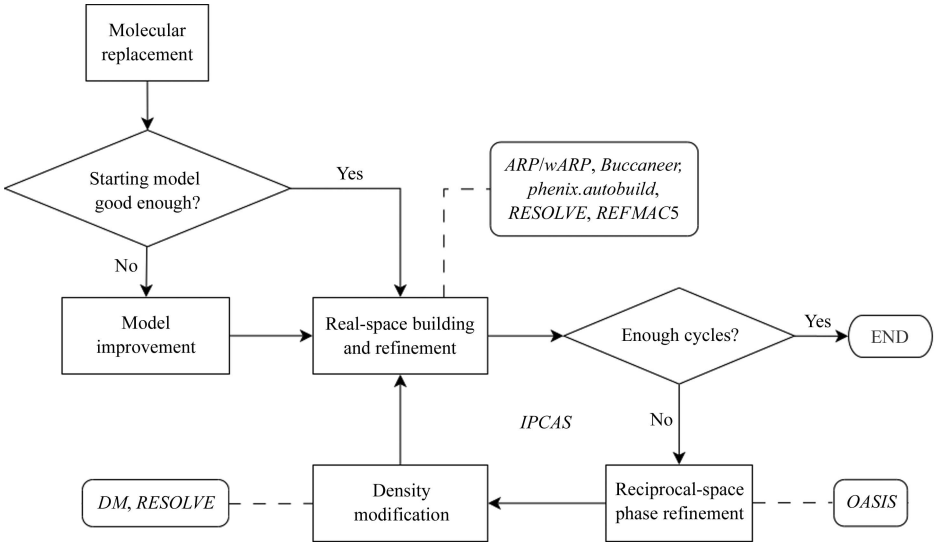


Figure 1
Workflow of *IPCAS* model completion from a partial molecular-replacement solution.

3. Test results

Four protein complexes were tested with *IPCAS*: the Hha–H-NS complex, the NDM-1–MBP complex, the CD38–nanobody complex and PDB entry 4owr. The data-collection and refinement statistics for the first three cases are listed in Table 1. *Phaser* (McCoy *et al.*, 2007) was first applied to obtain a partial MR solution. The iteration control in *IPCAS* was set as '*OASIS–DM–Buccaneer*' for all four cases. As a comparison, the same (or a greater) number of iterations of *Buccaneer* alone were carried out. The details of the four test cases and the test results are listed in Table 2.

3.1. Model completion of the Hha–H-NS complex

There are four chains in an Hha–H-NS complex, with one H-NS dimer in the core region and two Hha molecules binding at each side of the H-NS dimer. In our tests, an $H\text{-}NS_{46}$ dimer (PDB entry 1ov9; 65% sequence identity) was taken as a search model for MR. From the *Phaser* result listed in Table 2, the MR solution was not quite straightforward, with an LLG value of less than 100. Thus, the first *IPCAS* trial directly using the MR solution as

a starting structure for model completion could not proceed at the density-modification step as *DM* stopped automatically because the figure-of-merit (FOM) criterion was not met. Therefore, the MR solution from *Phaser* was manually

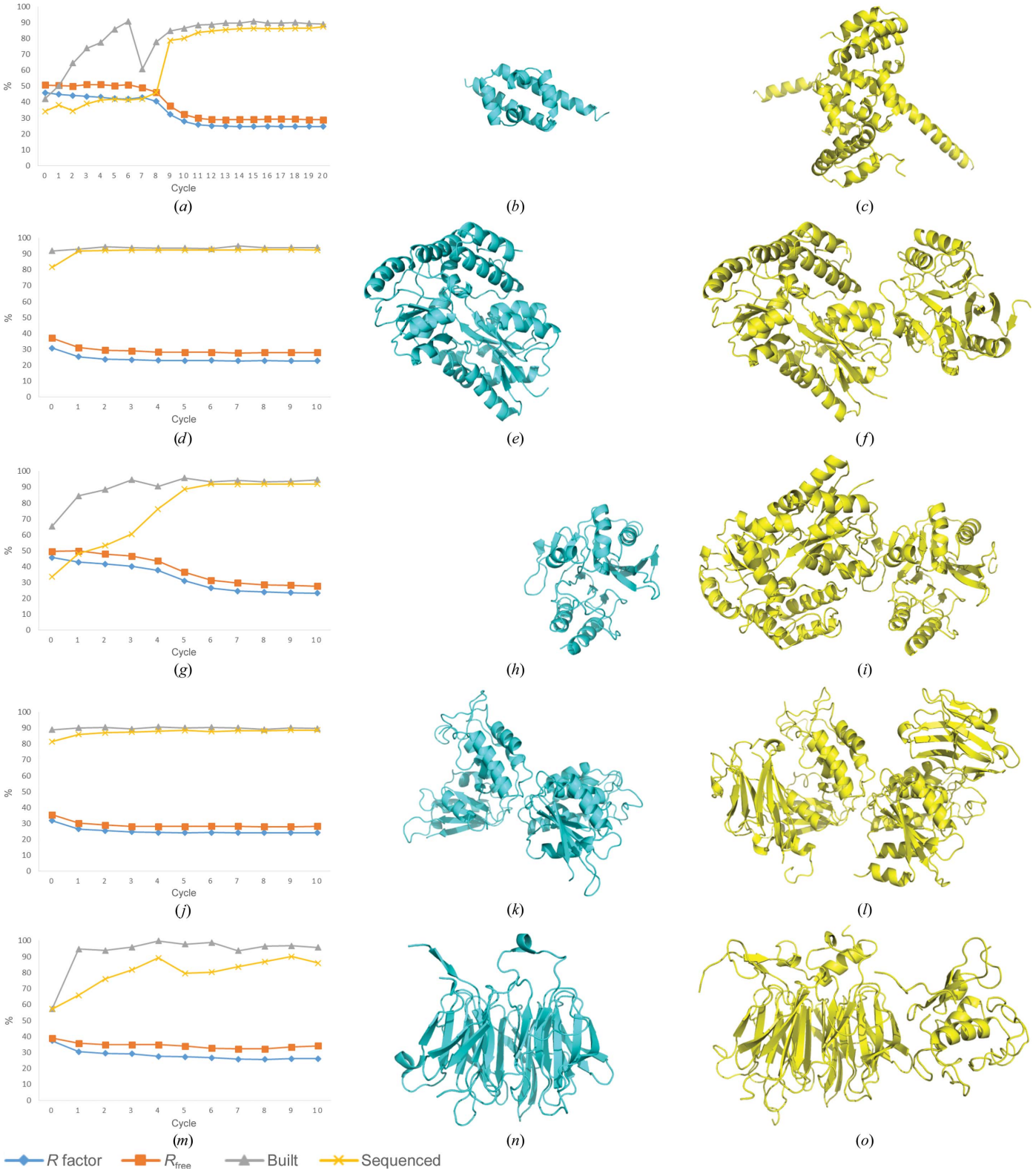


Figure 2
IPCAS model-completion results for four cases. The left five figures are model-completion status (percentage built/sequenced/*R* factor/R_{free}) per cycle for the Hha–H-NS complex (*a*), the NDM-1–MBP complex, starting from MBP (*d*) and starting from NDM-1 (*g*), the CD38–nanobody complex (*j*) and the 4owr complex (*m*). The middle five figures (*b*, *e*, *h*, *k*, *n*) are the initial models for each case (cyan). The right five figures (*c*, *f*, *i*, *l*, *o*) are the *IPCAS* model-completion results (yellow) for the five structures.

adjusted *via Coot* and *REFMAC*5. The model sequence was replaced by the target sequence and several terminal residues were removed as they did not fit well into the density. After several rounds of manual refinement, the R and $R_{\rm free}$ values decreased from 53.36 and 54.49% (the MR solution) to 46.18 and 50.63%, respectively, and the FOM increased from 0.313 to 0.456. The refined model was then delivered to *IPCAS* for extension. After 20 cycles of iteration, 269 residues (about 89% of the full sequence) were correctly modelled starting from the 89 residues of the MR solution. The model-completion rate became convergent after about the 15th cycle (Fig. 2*a*). The initial structure is shown in Fig. 2(*b*) and the final structure after cycle 20 is shown in Fig. 2(*c*). The widely used autobuilding program *Buccaneer* (v.1.5.2 in *CCP*4 v.6.3.0) alone did not give a correct result using either the *Phaser* solution or the manually refined initial structure.

We also attempted to use DEN refinement (Schröder *et al.*, 2010) and model morphing (Terwilliger *et al.*, 2012) with default parameters to improve the initial MR model quality. The R and $R_{\rm free}$ values decreased from 53.36 and 54.49% to 48.45 and 50.34%, respectively, in DEN refinement, and to 49.13 and 53.30,% respectively, in model morphing. However, neither *Buccaneer* nor *IPCAS* could extend the refined model any further. The r.m.s.d. of C^{α} atoms between the DEN-refined model and the manually refined model was 3.74 Å (part of the secondary structure collapsed after DEN refinement in this case) and the r.m.s.d. between the morphed model and the manually refined model was 1.08 Å. The r.m.s.d.s of C^{α} atoms between the final refined structure and the manually refined model, the DEN-refined model and the morphed model were also calculated, and the values were 0.63, 3.85 and 1.29 Å, respectively.

3.2. Model completion of the NDM-1–MBP complex

The NDM-1–MBP complex has two chains in the structure: one molecule of NDM-1 and one molecule of MBP. NDM-1 is about half the length of MBP. To obtain the NDM-1 structure, MBP (PDB entry 1y4c; LaPorte *et al.*, 2004) was taken as an MR search model and the solution given by *Phaser* was of sufficient quality to allow *IPCAS* to build about 94% of the sequence. As shown in Fig. 2(*d*), the model-completion rate became convergent after about the third cycle. The initial structure of MBP is shown in Fig. 2(*e*) and the final structure after cycle 10 is shown in Fig. 2(*f*). To further test the model-completion capability of *IPCAS*, the NDM-1 molecule was used as the MR search model. The result was similar apart from that building from the smaller subunit requires more cycles of iteration: the model-completion rate became convergent after about the seventh cycle (Fig. 2*g*). The initial structure of NDM-1 is shown in Fig. 2(*h*) and the final structure after cycle 10 is shown in Fig. 2(*i*). In comparison, *Buccaneer* alone could complete model building within ten cycles starting from MBP and within 20 cycles starting from NDM-1.

3.3. Model completion of the CD38–nanobody complex

The CD38–nanobody complex has four chains per asymmetric unit with two molecules of CD38 and two molecules of nanobody. In the test, a CD38 monomer was sent to *Phaser* to search for two copies and the solution was quite clear. Starting from the *Phaser* solution, *IPCAS* could build about 90% of the complex sequence and the model-completion rate became convergent after about the third cycle (Fig. 2*j*). The initial structure is shown in Fig. 2(*k*) and the final structure after ten cycles is shown in Fig. 2(*l*). In comparison, *Buccaneer* alone could finish model building within 15 cycles. The r.m.s.d. of the C^{α} atoms between nanobody-bound CD38 and apo CD38 (PDB entry 1yh3; Liu *et al.*, 2005) is 1.26 Å, suggesting a small conformational change upon ligand binding.

3.4. Model completion of the 4owr complex

The complex with PDB code 4owr is composed of three chains in an asymmetric unit with space group $P42_12$. In the test, chain *A* (57%) was sent to *Phaser* to search for one copy and the solution was quite clear. The MR solution was then sent to *IPCAS*, and after ten cycles of iteration about 97% of the sequence including the main parts of all three chains could be built (Fig. 2*m*). The initial structure is shown in Fig. 2(*n*) and the final structure after ten cycles is shown in Fig. 2(*o*). In comparison, *Buccaneer* alone did not extend the initial model further.

3.5. Summary

In our tests, *IPCAS* was able to extend the starting structures from as low as 30% of the complex to almost complete for all three cases, with reasonable R and $R_{\rm free}$ values. For the Hha–H-NS complex *IPCAS* initially failed to extend the structure directly from the *Phaser* solution, but after the FOM value of the starting model was increased from 0.313 to 0.456 by manual refinement *IPCAS* could eventually extend the complex structure from less than 30% to almost 90%. In contrast, the widely used autobuilding program *Buccaneer* (from the *CCP*4 suite) could successfully handle only the two cases with promising MR solutions (high starting fraction and high LLG) and not the toughest case of the Hha–H-NS complex nor the lower resolution case PDB entry 4owr.

4. Discussion

Structural studies of protein complexes are an important approach in understanding related cellular processes. In this study, we report a procedure that is particularly suitable for solving crystallographic protein-complex structures based on direct-method-aided dual-space iterative phasing and model building implemented in *IPCAS*. Our procedure shows an advantage, particularly for a test case in which only a small fraction/subunit (less than 30% of the complex) has a known homologue structure, compared with the widely used model-building approach *Buccaneer*. With an MR solution of sufficient quality, the *IPCAS* workflow can be quite automatic. In the challenging case where the MR solution does not have

sufficiently high quality, the starting model should be improved (until the FOM is larger than about 0.4 in our test) before being delivered to *IPCAS*. Also, our tests show that *IPCAS* is capable of low-resolution phasing (lower than 3 Å) with a sufficient known fraction (greater than 50%) of the complex in the starting model. We hope that our procedure may provide an option for solving protein-complex structures, especially for difficult cases.

5. Program availability

The program *IPCAS* can be downloaded at the website http://cryst.iphy.ac.cn.

Acknowledgements

This work was supported by grants 771011M and 766911M from the Research Grant Council of Hong Kong (to QH), grant 2011CB911101 from the Ministry of Science and Technology of China (to HF) and grant ZDSYS20140509142721429 from Shenzhen Municipal Government (to HZ). The X-ray diffraction data were collected on beamline 17U of the Shanghai Synchrotron Radiation Facility.

References

Alberts, B. & Miake-Lye, R. (1992). *Cell*, **68**, 415–420.
Ali, S. S., Whitney, J. C., Stevenson, J., Robinson, H., Howell, P. L. & Navarre, W. W. (2013). *J. Biol. Chem.* **288**, 13356–13369.
Aloy, P., Pichaud, M. & Russell, R. B. (2005). *Curr. Opin. Struct. Biol.* **15**, 15–22.
Barabási, A. L. & Oltvai, Z. N. (2004). *Nature Rev. Genet.* **5**, 101–113.
Brünger, A. T. (1993). *Immunomethods*, **3**, 180–190.
Cerdan, R., Bloch, V., Yang, Y., Bertin, P., Dumas, C., Rimsky, S., Kochoyan, M. & Arold, S. T. (2003). *J. Mol. Biol.* **334**, 179–185.
Chillemi, A., Zaccarello, G., Quarona, V., Ferracin, M., Ghimenti, C., Massaia, M., Horenstein, A. L. & Malavasi, F. (2013). *Mol. Med.* **19**, 99–108.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Davies, D. R., Padlan, E. A. & Sheriff, S. (1990). *Annu. Rev. Biochem.* **59**, 439–473.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
Hardison, R. C. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 5675–5679.
Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999). *Nature (London)*, **402**, C47–C52.
He, Y., Yao, D.-Q., Gu, Y.-X., Lin, Z.-J., Zheng, C.-D. & Fan, H.-F. (2007). *Acta Cryst.* D**63**, 793–799.
Hurwitz, J. (2005). *J. Biol. Chem.* **280**, 42477–42485.
Karmali, A. M., Blundell, T. L. & Furnham, N. (2009). *Acta Cryst.* D**65**, 121–127.
Kortemme, T. & Baker, D. (2004). *Curr. Opin. Chem. Biol.* **8**, 91–97.
Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
LaPorte, S. L., Forsyth, C. M., Cunningham, B. C., Miercke, L. J., Akhavan, D. & Stroud, R. M. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 1889–1894.
Lebendiker, M. & Danieli, T. (2011). *Methods Mol. Biol.* **681**, 281–293.
Lee, H. C. (2006). *Mol. Med.* **12**, 317–323.
Liu, Q., Kriksunov, I. A., Graeff, R., Munshi, C., Lee, H. C. & Hao, Q. (2005). *Structure*, **13**, 1331–1339.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Newman, M. E. J. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 8577–8582.
Nieto, J. M., Madrid, C., Miquelay, E., Parra, J. L., Rodríguez, S. & Juárez, A. (2002). *J. Bacteriol.* **184**, 629–635.
Pereira-Leal, J. B., Levy, E. D. & Teichmann, S. A. (2006). *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 507–517.
Phizicky, E. M. & Fields, S. (1995). *Microbiol. Rev.* **59**, 94–123.
Quan, B., Seo, H. S., Blobel, G. & Ren, Y. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 9127–9132.
Raboni, S., Bettati, S. & Mozzarelli, A. (2009). *Cell. Mol. Life Sci.* **66**, 2391–2403.
Schröder, G. F., Levitt, M. & Brunger, A. T. (2010). *Nature (London)*, **464**, 1218–1222.
Stevenson, G. T. (2006). *Mol. Med.* **12**, 345–346.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Terwilliger, T. C., Read, R. J., Adams, P. D., Brunger, A. T., Afonine, P. V., Grosse-Kunstleve, R. W. & Hung, L.-W. (2012). *Acta Cryst.* D**68**, 861–870.
Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2004). *Acta Cryst.* D**60**, 1991–1996.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, T., Gu, Y.-X., Zheng, C.-D. & Fan, H.-F. (2010). *Chin. Phys. B*, **19**, 086103.
Zhang, H. & Hao, Q. (2011). *FASEB J.* **25**, 2574–2582.