



# Crystallization and initial X-ray crystallographic analysis of a *de novo*-designed protein with left-handed $\beta\alpha\beta$ units

Naoki Tomita,<sup>a,†</sup> Riu Hirano,<sup>a,†</sup> Hiroto Murata,<sup>a,†</sup> Yasufumi Umena,<sup>b</sup>  
Hiroki Onoda,<sup>b</sup> George Chikenji<sup>a,\*</sup> and Leonard M. G. H. Chavas<sup>a,b,\*</sup>

Received 13 January 2025

Accepted 3 April 2025

Edited by D. Liebschner, Lawrence Berkeley National Laboratory, USA

† These authors contributed equally to this work.

**Keywords:** *de novo* protein design; left-handed  $\beta\alpha\beta$  motif; *AlphaFold*; X-ray crystallography.

<sup>a</sup>Department of Applied Physics, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, and <sup>b</sup>Synchrotron Radiation Research Center, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan. \*Correspondence e-mail: chikenji@nagoya-u.jp, l.chavas@nusr.nagoya-u.ac.jp

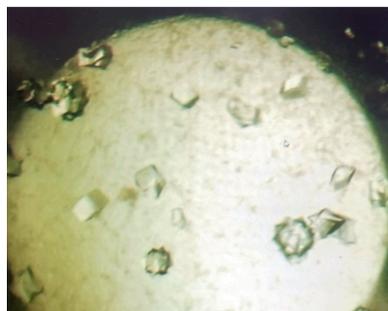
A newly designed protein featuring a rare left-handed  $\beta\alpha\beta$  motif has successfully been crystallized and characterized by preliminary X-ray diffraction. The computational design was conducted using a combination of *Rosetta BlueprintBDR*, *ProteinMPNN* and *AlphaFold2*, generating eight candidates based on predicted stability and folding accuracy. The final construct was expressed, purified and crystallized in space group  $P2_1$ . Complete X-ray diffraction data were collected on the BL2S1 beamline at the Aichi Synchrotron and processed to 1.95 Å resolution. Despite multiple attempts, molecular replacement using the *AlphaFold2* model did not yield a conclusive solution, suggesting that alternative phasing methods or refined modeling approaches will be needed. This work highlights both the promise and the challenges of pushing protein biodesign into underexplored structural motifs and provides a foundation for future structural and functional investigations.

## 1. Introduction

The design and synthesis of *de novo* proteins have emerged as a powerful approach to elucidating protein structure and function (Huang *et al.*, 2016). Such *de novo*-designed proteins enable the exploration of novel folds and functionalities that are not found in natural protein repertoires (Koga *et al.*, 2012). This study reports the crystallization and initial X-ray crystallographic analysis of a newly designed protein featuring a rare left-handed  $\beta\alpha\beta$  motif.

The left-handed  $\beta\alpha\beta$  motif is only rarely observed in the Protein Data Bank (PDB), making it an intriguing target for *de novo* design (Murata *et al.*, 2021). In principle,  $\beta\alpha\beta$  substructures can adopt either a right-handed or left-handed twist; however, surveys of the PDB reveal that the left-handed variant is almost entirely prohibited (Richardson, 1976). Because  $\beta\alpha\beta$  motifs occur so frequently in protein structures, appearing in more than half of all known domains, this absence severely restricts the diversity of naturally observed  $\alpha/\beta$  folds (Finkelstein & Ptitsyn, 1987). Incorporating a left-handed  $\beta\alpha\beta$  motif could therefore unlock new structural possibilities that cannot be accessed by the right-handed motif alone, making it a key frontier for expanding the repertoire of functional protein designs.

Recent advances in computational protein design, including *AlphaFold2*, have significantly enhanced our ability to predict and model novel protein structures with high accuracy (Jumper *et al.*, 2021). These breakthroughs were further recognized in 2024 with the award of the Nobel Prize in Chemistry to David Baker, Demis Hassabis and John Jumper for their pioneering work in computational protein design and



protein structure prediction, a testament to the transformative potential of engineering novel protein architectures.

In this work, we employed *AlphaFold2* alongside other computational tools to design a protein with a rare left-handed  $\beta\alpha\beta$  fold. The protein was purified, crystallized and subjected to X-ray diffraction experiments to characterize its structure and verify this unusual motif. While molecular replacement did not yield a solution, the crystallographic data presented here offer promising insights for future structural determination.

Our findings highlight the difficulties of solving protein motifs that are not routinely encountered in nature, whether mirror-image proteins or unusual structural motifs such as the left-handed  $\beta\alpha\beta$  fold. Further refinement of structure-determination methods and design strategies will be critical to overcome these challenges. By exploring unique secondary structures, we aim to expand the protein-design toolkit and uncover new possibilities in protein engineering.

## 2. Materials and methods

### 2.1. Protein-sequence design and model generation

The *de novo* protein featuring a left-handed  $\beta\alpha\beta$  motif was designed using a multifaceted computational pipeline. Firstly, *Rosetta BluePrintBDR* was employed with explicit constraints to favor backbone dihedral angles consistent with a left-handed  $\beta\alpha\beta$  geometry, ensuring that potential scaffolds would adopt this unusually twisted conformation (Fleishman *et al.*, 2011). These backbones were then refined using *ProteinMPNN*, which optimized the side-chain composition to favor stable folding and solubility (Dauparas *et al.*, 2022). Finally, *AlphaFold2* assessed the likelihood of each candidate adopting the intended fold, with pLDDT (predicted local distance difference test) and pTM (predicted template modeling) scores guiding the selection of the most promising designs (Jumper *et al.*, 2021). A total of eight sequences emerged from this workflow, each showing high predicted accuracy. Throughout iterative rounds of refinement, we closely monitored charge distribution, potential steric clashes and secondary-structure geometry, all to preserve a robust left-handed  $\beta\alpha\beta$  substructure. The top-performing construct, which scored highest on both the pLDDT and pTM metrics, was chosen for experimental validation (Takei & Ishida, 2022).

### 2.2. Molecular cloning, protein expression and purification

The gene encoding the designed protein (9.13 kDa) was synthesized (GenScript) and cloned into the pET-24a(+) vector (Novagen), adding a C-terminal His-tag. The recombinant vector was transformed into *Escherichia coli* T7 Express cells (New England Biolabs) grown in Luria–Bertani (LB) medium at 37°C. Protein expression was induced at an optical density (600 nm) of 0.6–0.8 by the addition of 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (Sigma–Aldrich) and was maintained for 4–6 h.

After harvesting cells by centrifugation at 8000g for 20 min at 4°C, the pellet was resuspended in phosphate-buffered

saline (PBS; Thermo Fisher Scientific). Following sonication, the soluble fraction was clarified by centrifugation at 20 000g for 20 min (repeated twice) at 4°C. Ni–NTA affinity chromatography (Qiagen) was used to isolate the target protein, which was subsequently concentrated using a Vivaspinn 3K ultrafiltration device (Cytiva). The His-tag remained uncleaved, as preliminary tests showed no detrimental effect on stability. The protein, at a concentration of  $\sim 120$  mg ml<sup>-1</sup>, was stored at 4°C short-term and frozen at –80°C long-term without loss of solubility.

### 2.3. Crystallization

Initial crystallization trials used the sitting-drop vapor-diffusion method in 96-well plates (Molecular Dimensions). A 1:1 mixture of the protein solution ( $\sim 120$  mg ml<sup>-1</sup>) and the reservoir solution was screened at 20°C using Crystal Screen from Hampton Research. Small crystals appeared after a few days under several PEG-containing conditions. Subsequent manual optimization yielded a reproducible condition consisting of 100 mM NaCl, 20% (w/v) PEG 5000 pH 7.0. Within a week, the crystals reached dimensions of up to 300  $\mu$ m in their longest axis (Fig. 1).

Crystals were harvested using LithoLoops (Molecular Dimensions) and briefly soaked in 20% (v/v) glycerol as a cryoprotectant before flash-cooling in liquid nitrogen. This protocol minimized ice formation and crystal cracking, resulting in improved diffraction at the synchrotron.

### 2.4. Synchrotron data collection and analysis

X-ray data were collected on the BL2S1 beamline at the Aichi Synchrotron using a PILATUS 1M detector (Watanabe *et al.*, 2017). An initial assessment showed diffraction to about 1.8 Å resolution. However, full data acquisition revealed significant radiation damage, reducing the effective resolution to 1.95 Å (Fig. 2). Data were collected at a wavelength of 0.7233 Å, employing a 0.1° oscillation with a 1.0 s exposure per image to cover 360° of total rotation. The total absorbed dose was estimated to be approximately 0.34 MGy, about one-tenth of the 3.0 MGy Garman dose limit, suggesting that the observed crystal decay may reflect an unusual sensitivity rather than purely dose-related damage (Zeldin *et al.*, 2013).

The data were indexed, integrated and scaled using the *xia2* pipeline and related software from the *CCP4* suite (Agirre *et al.*, 2023). The crystals belonged to space group  $P2_1$ , with unit-cell parameters  $a = 34.28$ ,  $b = 53.41$ ,  $c = 37.63$  Å,  $\beta = 113.10^\circ$ . Analysis suggested the presence of one molecule in the asymmetric unit, consistent with the molecular weight of the protein. Attempts to solve the structure by molecular replacement using the *AlphaFold2* model in multiple pipelines provided by *CCP4* Cloud (Krissinel *et al.*, 2022), including *MOLREP* (Vagin & Teplyakov, 2010) and *Phaser* (McCoy *et al.*, 2007), were inconclusive. For the generation of the model coordinates, the *B* factors were recalculated by *MOLREP*, assuming the *AlphaFold* model. Additional optimizations and further model refinements may be required as no solution was obtained, or alternative phasing methods should be

**Table 1**

Data-quality and refinement statistics.

Values in parentheses are for the outer shell.

Beamline	BL2S1, Aichi Synchrotron
Detector	PILATUS 1M
Wavelength (Å)	0.72333
Temperature (K)	120.0
Crystal-to-detector distance (mm)	200.0
Oscillation per image (°)	0.1
Oscillation range (°)	360.0
Mosaicity (°)	0.61
Space group	$P2_1$
$a, b, c$ (Å)	34.28, 53.41, 37.63
$\alpha, \beta, \gamma$ (°)	90, 113.10, 90
Resolution range (Å)	29.05–1.95 (1.99–1.95)
Total No. of reflections	59960 (3297)
No. of unique reflections	8930 (476)
$R_{\text{merge}}$	0.1576 (0.7670)
$CC_{1/2}$	0.9950 (0.8770)
Completeness (%)	97.20 (97.90)
$\langle I/\sigma(I) \rangle$	6.70 (1.20)
Multiplicity	6.70 (6.90)
Wilson $B$ factor (Å <sup>2</sup> )	18.79

considered. Data-quality and refinement statistics are given in Table 1.

### 3. Results

#### 3.1. Model generation

Of the eight *de novo* sequences generated by our computational pipeline, the one demonstrating the highest *AlphaFold2* confidence scores (pLDDT and pTM) was selected for experimental testing (Fig. 3). *In silico* predictions indicated that this protein would adopt a well packed core consistent with a left-handed  $\beta\alpha\beta$  conformation, featuring extensive side-chain networks predicted to stabilize the backbone twist. Notably, no major steric clashes or destabilizing motifs were identified, suggesting that the designed sequence was capable of folding and remaining soluble under typical laboratory conditions.



**Figure 1**

Optical micrograph of the *de novo*-designed protein crystals grown by sitting-drop vapor diffusion at 20°C. The crystals reached dimensions of up to 300  $\mu\text{m}$  within a week. The octahedral shapes visible here reflect the reproducible crystallization conditions, which provided samples suitable for X-ray diffraction experiments.

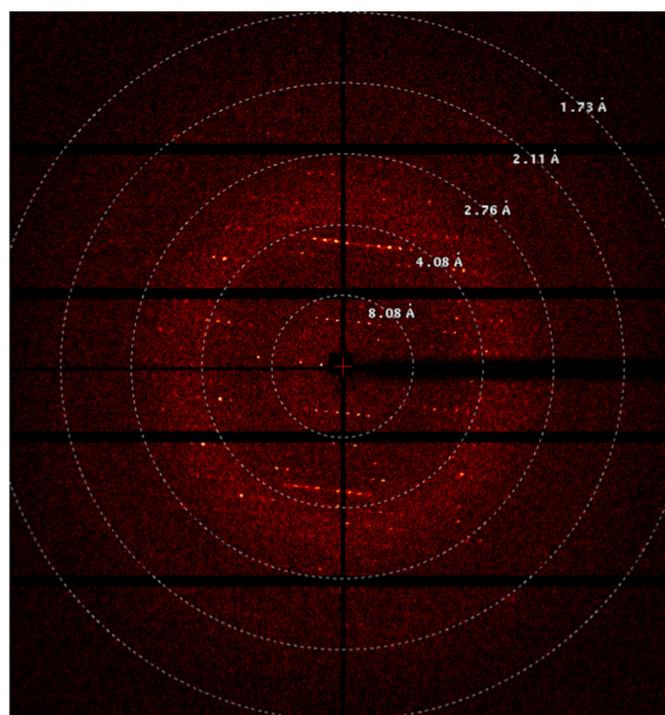
#### 3.2. Purification, crystallization and diffraction

Upon expression in *E. coli*, the protein appeared as a single major band on SDS-PAGE, and its high solubility permitted its concentration to  $\sim 120 \text{ mg ml}^{-1}$  without precipitation. In contrast to many *de novo* designs that can suffer from aggregation or instability, the present construct remained stable throughout purification and storage. Crystals reproducibly formed within a week under optimized conditions (100 mM NaCl, 20% PEG 5000 pH 7.0) and reached typical dimensions of  $\sim 300 \mu\text{m}$ .

Data collection at the Aichi Synchrotron revealed diffraction spots extending to approximately 1.8 Å resolution. However, noticeable radiation-induced decay caused the final data set to be truncated to 1.95 Å resolution to give a reliable signal-to-noise ratio. This finding underscores the X-ray sensitivity of the crystals and suggests that multi-crystal or shorter-exposure strategies might be essential for future structure-determination efforts.

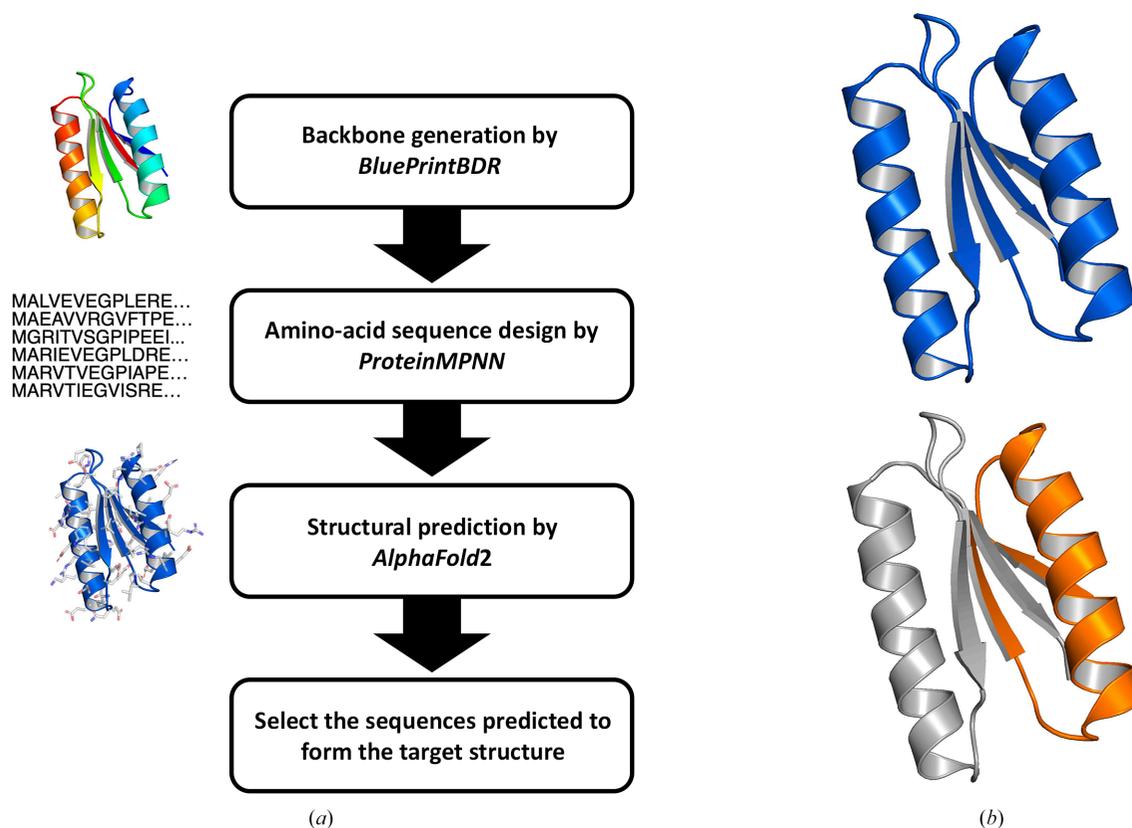
#### 3.3. Data analysis and molecular replacement

Detailed data processing in *DIALS* confirmed the space group to be  $P2_1$  with one molecule per asymmetric unit. Although the high-resolution shell revealed some reflections approaching 1.8 Å resolution, consistent beam damage



**Figure 2**

A representative diffraction image of the *de novo*-designed protein crystal collected on the BL2S1 beamline at the Aichi Synchrotron. Clear diffraction spots are visible at lower resolutions, but they diminish in intensity toward the higher resolution range. The relatively high mosaicity of these crystals is evident in the broader, less distinct diffraction spots. This image captures one still from the overall fine-sliced data collection, highlighting both the diffraction quality and the challenges posed by mosaicity.


**Figure 3**

Schematic representation of the *de novo* design workflow and the final *AlphaFold2*-predicted structure. (a) A flowchart outlines the sequential use of *Rosetta BluePrintBDR*, *ProteinMPNN* and *AlphaFold2* to generate candidate sequences featuring a left-handed  $\beta\alpha\beta$  motif. (b) The top-scoring *AlphaFold2* model is shown in two parts. In the upper ribbon diagram, the protein is color-coded by pLDDT (orange  $\rightarrow$  yellow  $\rightarrow$  light blue  $\rightarrow$  dark blue) to indicate increasing local confidence; the average pLDDT value is 97.0 and the predicted TM-score (pTM) is 0.84, reflecting high overall confidence. Lighter, warmer hues mark lower-confidence regions that may deviate from the *in silico* model in the crystal structure. In the lower ribbon diagram, the left-handed  $\beta\alpha\beta$  motif is highlighted in orange, emphasizing its location and potential structural significance.

required the final data set to be limited to 1.95 Å resolution to give a reliable signal-to-noise ratio. During data analysis, attempts were also made to integrate less than 360° of rotation, but no tangible improvements in data quality were observed (data not shown). Molecular-replacement attempts using the *AlphaFold2*-predicted model in multiple CCP4 Cloud pipelines (including *MOLREP* and *Phaser*) did not yield a definitive solution, pointing to possible deviations between the *in silico* model and the actual fold or limitations arising from radiation damage. Despite recalculating *B* factors for the model and performing additional optimizations of the *AlphaFold2*-generated model, no solution was obtained, suggesting the need for alternative phasing methods or further model refinements. Notably, no suitable homologous model could be identified in the PDB for this left-handed  $\beta\alpha\beta$  motif, leaving the *AlphaFold2*-based construct as our only feasible search template; attempts to use naturally occurring right-handed  $\beta\alpha\beta$  structures as proxies were similarly unsuccessful.

Interestingly, a recent large-scale study (Keegan *et al.*, 2024) reported that only 3% of approximately 400 structures solved by SAD failed to yield a solution by molecular replacement when using an *AlphaFold*-based model. Our inability to obtain a clear solution therefore underscores the unusual difficulties posed by this system. Given the rarity of the left-handed  $\beta\alpha\beta$

motif in current structure databases, conventional MR search procedures may struggle to accommodate such unique structural features. Alternative phasing strategies, such as selenomethionine labeling or heavy-atom soaking, will be investigated in future work to address these challenges.

#### 4. Conclusion

In summary, we have successfully designed, purified and crystallized a *de novo* protein predicted to feature a rare left-handed  $\beta\alpha\beta$  motif. Although X-ray data were collected to an effective resolution of 1.95 Å in space group  $P2_1$ , molecular replacement using the *AlphaFold2* model was not conclusive. Future efforts will focus on mitigating radiation damage during data collection and pursuing alternative phasing methods (for example SeMet labeling). Resolving this unusual fold will be a key advance in protein bio-design, broadening our understanding of engineering novel structural motifs and opening doors to innovative applications in protein science.

#### Acknowledgements

The X-ray diffraction experiments were performed with the approval of Aichi Synchrotron Radiation Center, Aichi

Science and Technology Foundation, Aichi, Japan (Proposal No. NU-023). Author contributions were as follows. HM and NT designed the *de novo* protein sequences, while RH and NT performed the biochemical experiments. YU and HO carried out the diffraction experiments. The concept and direction of the project were provided by GC and LC. All authors contributed to data analysis, interpretation and manuscript preparation.

### Conflict of interest

The authors declare no conflicts of interest.

### Data availability

The data are available from the corresponding authors upon request.

### Funding information

This work was partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research; BINDS) from AMED under Grant No. JP23ama121001 (LMGC).

### References

Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley, P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito,

M., Sánchez Rodríguez, F., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M. & Yamashita, K. (2023). *Acta Cryst. D* **79**, 449–461.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A. K., King, N. P. & Baker, D. (2022). *Science*, **378**, 49–56.

Finkelstein, A. V. & Ptitsyn, O. B. (1987). *Prog. Biophys. Mol. Biol.* **50**, 171–190.

Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E.-M., Khare, S. D., Koga, N., Ashworth, J., Murphy, P., Richter, F., Lemmon, G., Meiler, J. & Baker, D. (2011). *PLoS One*, **6**, e20161.

Huang, P.-S., Boyken, S. E. & Baker, D. (2016). *Nature*, **537**, 320–327.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodensteiner, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.

Keegan, R. M., Simpkin, A. J. & Rigden, D. J. (2024). *Acta Cryst. D* **80**, 766–779.

Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T. & Baker, D. (2012). *Nature*, **491**, 222–227.

Krissinel, E., Lebedev, A. A., Uski, V., Ballard, C. B., Keegan, R. M., Kovalevskiy, O., Nicholls, R. A., Pannu, N. S., Skubák, P., Berrisford, J., Fando, M., Lohkamp, B., Wojdyr, M., Simpkin, A. J., Thomas, J. M. H., Oliver, C., Vornrhein, C., Chojnowski, G., Basle, A., Purkiss, A., Isupov, M. N., McNicholas, S., Lowe, E., Triviño, J., Cowtan, K., Agirre, J., Rigden, D. J., Uson, I., Lamzin, V., Tews, I., Bricogne, G., Leslie, A. G. W. & Brown, D. G. (2022). *Acta Cryst. D* **78**, 1079–1089.

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.

Murata, H., Imakawa, H., Koga, N. & Chikenji, G. (2021). *PLoS One*, **16**, e0256895.

Richardson, J. S. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 2619–2623.

Takei, Y. & Ishida, T. (2022). *bioRxiv*, 2022.04.05.487218.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.

Watanabe, N., Nagae, T., Yamada, Y., Tomita, A., Matsugaki, N. & Tabuchi, M. (2017). *J. Synchrotron Rad.* **24**, 338–343.

Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Appl. Cryst.* **46**, 1225–1230.