Supporting information

Structure-Based Approach for the Discovery of Pyrrolo[3,2-*d*]pyrimidine-based EGFR T790M/L858R Mutant Inhibitors

Satoshi Sogabe,*,‡ Youichi Kawakita,‡ Shigeru Igaki,‡ Hidehisa Iwata, Hiroshi Miki, Douglas R. Cary, Terufumi Takagi, Shinji Takagi, Yoshikazu Ohta, Tomoyasu Ishikawa*

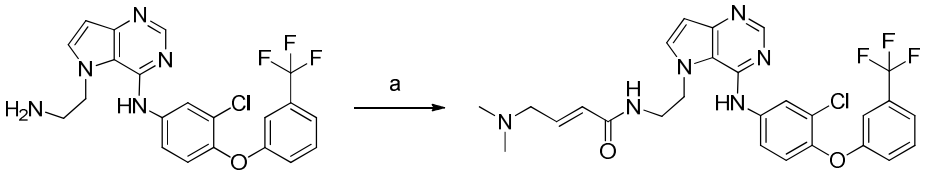
Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited: 26-1, Muraoka-Higashi 2-chome, Fujisawa, Kanagawa 251-8555, Japan.

Experimental Procedures

Chemical synthesis

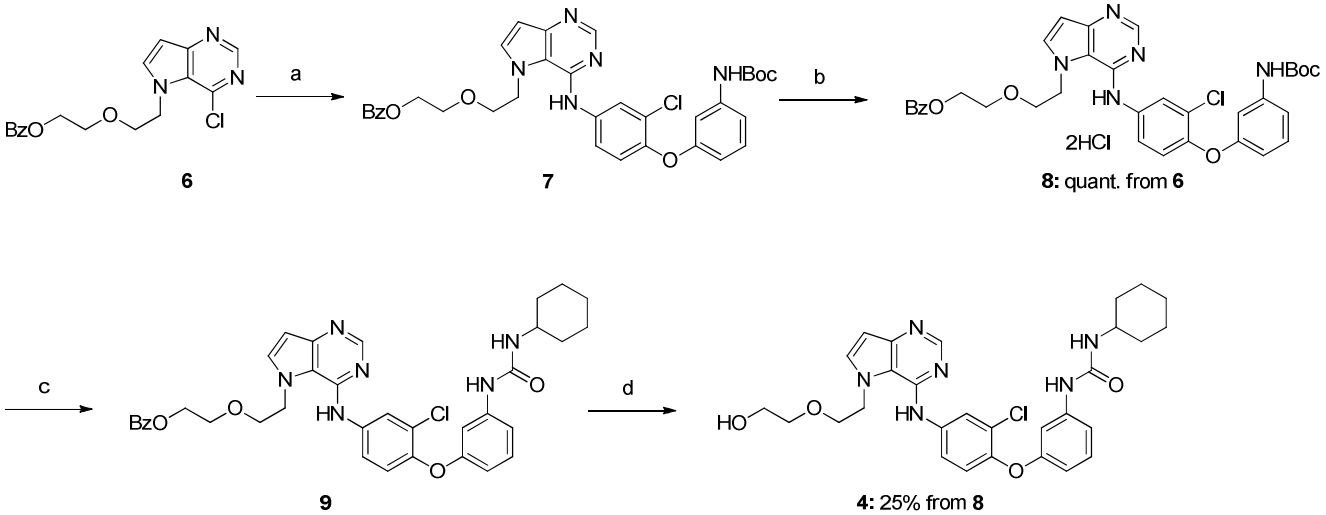
Neratinib was synthesized by a reported method.[1] Compounds **2** and **4** were synthesized as shown in Scheme 1 and 2.

Scheme 1. Synthesis of **2**



Reagents: (a) (*E*)-4-(dimethylamino)but-2-enoic acid, EDC, HOBt, DIEA, DMF, rt.

Scheme 2. Synthesis of **4**



Reagents: (a) *tert*-butyl[3-(4-amino-2-chlorophenoxy)phenyl]carbamate, 2-propanol, 80°C; (b) 4N HCl in EtOAc, MeOH, rt; (c) cyclohexylisocyanate, Et_3N, toluene, 120°C; (d) 1N NaOH, THF, MeOH, rt.

General Procedure

Melting points were determined on a Yanagimoto micro melting point apparatus or SRS OptiMelt melting point apparatus, and are uncorrected. Proton nuclear magnetic resonance (^{1}H-NMR) spectra were recorded on a Varian Gemini-200 (200 MHz) spectrometer or Varian Mercury-300 (300 MHz) spectrometer. Chemical shifts are given in parts per million (ppm) with tetramethylsilane as an internal standard, and coupling constants (*J* values) are given in Hertz (Hz). Splitting patterns and apparent multiplicities are designated as s (singlet), d (doublet), dd (double doublet), t (triplet), dt (double triplet), q (quartet), m (multiplet), br s (broad singlet). Elemental analyses were carried out by Takeda Analytical Research Laboratories, Ltd., and the results obtained were within ±0.4% of the theoretical values. MS spectra were collected with a Waters LC-MS system (ZMD-1) and were used to confirm ≥95% purity of each compound. The column used was an L-column 2 ODS (3.0 x 50 mm I.D., CERI, Japan) with a temperature of 40°C and a flow rate of 1.2 mL/min. Mobile phase A was 0.05% TFA in ultrapure water. Mobile phase B was 0.05%

TFA in acetonitrile which was increased linearly from 5% to 90% over 2 minutes, 90% over the next 1.5 minutes, after which the column was equilibrated to 5% for 0.5 minutes. Column chromatography was carried out on a silica gel column (Kieselgel 60, 63-200 mesh, Merck or Chromatorex® NH-DM1020, 100-200 mesh, Fuji Silysia chemical). Yields were not optimized.

(*E*)-*N*-(2-(4-((3-Chloro-4-(3-(trifluoromethyl)phenoxy)phenyl)amino)-5*H*-pyrrolo[3,2-*d*]pyrimidin-5-yl)ethyl)-4-(dimethylamino)but-2-enamide (2). A mixture of *N*-{3-chloro-4-[3-(trifluoromethyl)phenoxy]phenyl}-5-methyl-5*H*-pyrrolo[3,2-*d*]pyrimidin-4-amine **5** (213 mg, 0.48 mmol),[2] *tert*-butyl[3-(4-amino-2-chlorophenoxy)phenyl]carbamate (65.4 mg, 0.51 mmol), *N*,*N*-diisopropylethylamine (DIEA) (124 mg, 0.96 mmol), 1-hydroxybenzotriazole monohydrate (HOBt) (74.3 mg, 0.55 mmol) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (105 mg, 0.55 mmol) in *N*,*N*-dimethylformamide (DMF) (5 mL) was stirred at room temperature for 64 h. Saturated sodium hydrogen carbonate (40 mL) was added to the reaction mixture, and the mixture was extracted with EtOAc (40 mL). The organic layer was washed successively with water (40 mL) and brine (40 mL), and dried over $MgSO_4$. The solvent was evaporated under reduced pressure and the residue was subjected to silica gel column chromatography (eluent, MeOH/EtOAc=0:100 to 5:95) to give **2** (99.7 mg, 37%) as a white crystalline solid. mp 172-174°C. ^{1}H NMR (DMSO-d_6) δ: 2.10 (6H, s), 2.94 (2H, d, J = 6.0 Hz), 3.45 (2H, d, J = 6.0 Hz), 4.57 (2H, t, J = 6.6 Hz), 5.97 (1H, d, J = 15.4 Hz), 6.40-6.63 (2H, m), 7.15-7.37 (3H, m), 7.47 (1H, d, J = 7.9 Hz), 7.56-7.71 (2H, m), 7.84 (1H, dd, J = 9.0, 2.3 Hz), 8.09 (1H, d, J = 2.6 Hz), 8.25-8.48 (2H, m), 8.78 (1H, s). *Anal.* Calcd for $C_{27}H_{26}ClF_3N_6O_2$: C, 58.00; H, 4.69; N, 15.03. Found: C, 57.86; H, 4.64; N, 14.96.

2-(2-(4-((4-(3-Aminophenoxy)-3-chlorophenyl)amino)-5*H*-pyrrolo[3,2-*d*]pyrimidin-5-yl)ethoxy)ethyl benzoate dihydrochloride (8). A mixture of *tert*-butyl [2-(4-chloro-5*H*-pyrrolo[3,2-*d*]pyrimidin-5-yl)ethyl]carbamate[1] (**6**, 206 mg, 0.60 mmol) and *tert*-butyl[3-(4-amino-2-chlorophenoxy)phenyl]carbamate[3] (300 mg, 0.90 mmol) in 2-propanol (7 mL) was stirred at 80°C for 12 h. To the mixture was added saturated sodium hydrogen carbonate (20 mL), and the mixture was extracted with EtOAc (30 mL). The organic layer was washed with brine (3 mL), and dried over $MgSO_4$. Insoluble $MgSO_4$ was filtered off, and the filtrate was concentrated in vacuo. The obtained residue was subjected to silica gel column chromatography (eluent, EtOAc/hexane=20:80 to 100:0). The objective fractions were collected and concentrated under reduced pressure. To the residue were added 4N HCl in EtOAc (8 mL) and MeOH (8 mL), and the mixture was stirred at rt for 5 h. To the solution were added 8N NaOH (8 mL) and water (10 mL), and the mixture was extracted with dichloromethane (30 mL). The organic layer was washed with brine (3 mL), and dried over $MgSO_4$. Insoluble $MgSO_4$ was filtered off, and the filtrate was concentrated in vacuo. The obtained residue was subjected to silica gel column chromatography (eluent, MeOH/EtOAc=0:100 to 20:80) to give **8** (370 mg, quant.) as a white powder. ^{1}H-NMR (DMSO-d_6) δ: 3.76-3.80 (2H, m), 3.87-3.94 (2H, m), 4.22-4.35 (2H, m), 4.85-4.93 (2H, m), 6.62-6.77 (3H, m), 6.87-7.18 (3H, m), 7.30-7.71 (8H, m), 7.90 (1H, s), 8.01 (1H, s), 8.65 (1H, s).

1-(3-(2-Chloro-4-((5-(2-(2-hydroxyethoxy)ethyl)-5*H*-pyrrolo[3,2-*d*]pyrimidin-4-yl)amino)phenoxy)phenyl)-3-cyclohexylurea (4). To a suspension of 2-(2-(4-((4-(3-aminophenoxy)-3-chlorophenyl)amino)-5*H*-pyrrolo[3,2-*d*]pyrimidin-5-yl)ethoxy)ethyl benzoate dihydrochloride (**8**, 240 mg, 0.39 mmol) in toluene (15 mL), triethylamine (2.0 mL) and cyclohexylisocyanate (137 mg, 1.10 mmol) were added successively, and the reaction mixture was stirred at 120°C for 6 h. Water (20 mL) was added to the reaction mixture, and the mixture was extracted with EtOAc (50 mL). The organic layer was washed with saturated brine (20 mL) and dried over $MgSO_4$. Insoluble $MgSO_4$ was filtered off, and the filtrate was concentrated in vacuo. The obtained residue was subjected to amine-functionalized silica gel column chromatography (eluent, MeOH/EtOAc=0:100 to 15:85). The objective fractions were concentrated under reduced pressure. The residue was dissolved in MeOH (6.0 mL) and tetrahydrofuran (6.0 mL). To the solution was added 1N NaOH (3.0 mL) and the mixture was stirred at rt for 24 h. Water (20 mL) was added to the reaction mixture, and the mixture was extracted with EtOAc (30 mL). The organic layer was washed with saturated brine and dried over $MgSO_4$. Insoluble $MgSO_4$ was filtered off, and the filtrate was concentrated in vacuo. The obtained residue was subjected to amine-functionalized silica gel column chromatography (eluent, MeOH/EtOAc=0:100 to 10:90) to give compound **4** (56 mg, 25%) as white crystals. mp 221-223°C. ^{1}H-NMR (DMSO-d_6) δ: 1.09-1.82 (10H, m), 3.34-3.51 (1H, m), 3.49 (4H, s), 6.00 (1H, d, J = 9.0 Hz), 6.46-6.52 (2H, m), 6.96-7.21 (4H, m), 7.61-7.69 (2H, m), 7.98 (1H, d, J = 2.7 Hz), 8.34 (1H, s), 8.41 (1H, s), 8.93 (1H, s). *Anal*. Calcd for $C_{29}H_{33}ClN_6O_4.0.5H_2O$: C, 60.67; H, 5.97; N, 14.62. Found: C, 60.94; H, 6.09; N, 14.28.

Protein preparation

Human EGFR kinase domain genes bearing the T790M and L858R mutations were expressed and purified by using a baculovirus/insect cell system, as was the wild-type EGFR kinase described previously.[4] DNA encoding residues 696–1022, was amplified by PCR method using the mutated EGFR cytosolic region as a template. Amplified DNA was cloned into a pFastBac1 vector (Invitrogen, USA) to acquire the hexa-histidine tag without a TEV-protease cleavage site at the N-terminus. The obtained recombinant transfer vector (Bac-to-Bac expression system, Invitrogen, USA) was transfected into *Spodoptera frugiperda* (Sf9) cells to generate recombinant baculovirus. Large scale production of recombinant protein was carried out in Sf9 cells. Cells were harvested and rapidly frozen for storage at −80 °C. The thawing cell pellets were suspended in the Ni buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 mM DTT, 20 % glycerol) and Complete Protease Inhibitor Cocktail (Roche Molecular Biochemicals, Germany), and lysed with a high-pressure fluids processor (Microfluidizer, M-110EH, Microfluidics, USA). Cell debris was removed by centrifugation at 18,000 × g for 30 min. The supernatant was collected and mixed with Ni-NTA agarose (Qiagen, USA) pre-equilibrated with the Ni buffer. Protein bound to Ni-NTA resin was packed into an Econo-column (BioRad, USA) and washed with the Ni buffer, followed by the buffer containing 20 mM imidazole. The protein was eluted with the Ni buffer containing 500 mM imidazole. Further purification by size-exclusion chromatography with Superdex 200pg 26×60 (GE healthcare, UK) was carried out with the

final buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM DTT, 10% glycerol. Purified protein solution with more than 95% purity as estimated by SDS-PAGE was concentrated up to 5–10 mg/ml by ultrafiltration and flash-frozen in liquid nitrogen for storage at −80°C. Mass spectrometric analysis confirmed that the protein was prepared as the non-phosphorylated form.

Enzyme assay

The EGFR kinase assays were performed using radiolabeled [γ-^{32}P] ATP (GE Healthcare, USA) in 96-well plates. The kinase reactions were performed in 50 mmol/L Tris-HCl, pH7.5, 5 mmol/L MnCl2, 0.01% Tween-20 and 2 mmol/L DTT containing 0.9 μCi of [γ-^{32}P] ATP per reaction, 3 μmol/L ATP (K_m^{app}), 5 μg/mL poly-Glu-Tyr(4:1) and 0.25 μg/mL of purified EGFR kinase domain in a total volume of 50 μL. To measure the IC_{50} value for enzyme inhibition, compounds were incubated with the enzyme for 5 minutes prior to the reaction at room temperature. The kinase reactions were initiated by adding ATP. After the kinase reaction incubated for 10 min at rt, reactions were terminated by the addition of 10% (final concentration) trichloroacetic acid. The [γ-^{32}P]-phosphorylated proteins were filtered in a harvest plate (Millipore, USA) with a cell harvester (PerkinElmer, USA) and washed free of [γ-^{32}P] ATP with 3% phosphoric acid. The plates were dried, followed by the addition of 25 μL of MicroScintO (PerkinElmer, USA). Radioactivity was counted by a Topcount scintillation counter (PerkinElmer, USA). IC_{50} values and 95% confidence intervals were calculated by nonlinear regression analysis. For determining IC_{50} value toward the EGFR enzyme harboring T790M and L858R mutation, kinase assays were performed as described above with 0.5 μmol/L ATP (K_m^{app}).

Crystallography

Protein at a concentration of 10 mg/ml was complexed by incubation of with a 3-fold molar excess of compound on ice for a few hours prior to crystallization experiments. Solid samples of inhibitors were freshly dissolved in DMSO to give 50 mM final concentration. The mixtures were centrifuged for 10 min at 10,000 × g and the clear supernatants used for crystallization. Crystals of the T790/L858R mutants complexed with TAK-285, compound **2** and neratinib grew against a reservoir solution containing 0.1 M Bis-tris propane (pH 6–7), 34–36% Tacsimate at 20°C, by the sitting drop vapor diffusion method. Crystals of the T790M/L858R mutant with compound **4** were grown from a reservoir containing 0.1 M Mes (pH 5.5), 24% PEG 3350 at 20°C, whereas crystals of the wild-type complexed with compound **4** were obtained from a reservoir solution of 0.1 M Bis-tris (pH 6.5), 0.2 M lithium sulfate, 24% PEG 3350 at 20 °C. Prior to data collection, crystals were immersed in the reservoir solution with addition of a cryoportectant such as glycerol, ethylene glycol or PEG 3350 and were flash-frozen in liquid nitrogen. Diffraction data were collected from a single crystal using the CCD detector Quantum 315 (ADSC) at beamline 5.0.3 of the Advanced Light Source (Berkeley, CA) or Quantum 270 (ADSC) at beamline NE3A of Photon Factory (KEK, Japan) under a 100 K nitrogen cryostream. The data were reduced and scaled with HKL2000.[5] The high resolution limit of data used in refinement was determined based on a signal-to-noise criterion (I/σ(I)) rather than on agreement factors. The structures were solved by the molecular replacement method with Molrep[6] of the CCP4 program suites[7] using the active or inactive EGFR structures (PDB code: 1M14 or 1XKK) as a search model. The structures were refined through an iterative procedure utilizing REFMAC[8] followed by model building in COOT.[9] The dictionary files for the ligands were prepared using AFITT (OpenEye Scientific Software, USA). The final models were validated using Molprobity.[10] Crystallographic processing and refinement statistics are summarized in Table S1. All structural figures were generated using PyMOL (Schrödinger, USA).

Growth inhibition assay

H1975 cells harboring EGFR (T790M/L858R) mutation were plated in 96-well plates. On the following day, cells were exposed to compounds. After 3 days incubation, relative cell number was estimated using sulforhodamine B (SRB, Sigma-Aldrich) staining method. The cells were fixed with 5.6% (v/v) trichloroacetic acid and then washed with 1% acetic acid (v/v) to remove the excess dye. The protein-bound dye was dissolved in 10 mM Tris base solution for OD determination at 550 nm using MTP-450 microplate reader (CORONA). The IC_{50} value was calculated using SAS software (version 5.0).

Accession Codes

Atomic coordinates and structure factors have been deposited in the Protein Data Bank as shown in Table S1.

REFERENCES

(1) Tsou, H-R.; Overbeek-Klumpers, E. G.; Hallett, W. A.; Reich, M. F.; Floyd, M. B.; Johnson, B. D.; Michalak, R. S.; Nilakantan, R.; Discafani, C.; Golas, J.; Rabindran, S. K.; Shen, R.; Shi, X.; Wang, Y-F.; Upeslacis, J.; Wissner, A. Optimization of 6,7-Disubstituted-4-(arylamino)quinoline-3-carbonitriles as Orally Active, Irreversible Inhibitors of Human Epidermal Growth Factor Receptor-2 Kinase Activity. *J. Med. Chem.* **2005**, *48*, 1107-1131.

(2) Ishikawa, T.; Seto, M,; Banno, H.; Kawakita, Y.; Oorui, M.; Taniguchi, T.; Ohta, Y.; Tamura, T.; Nakayama, A.; Miki, H.; Kamiguchi, H.; Tanaka, T.; Habuka, N.; Sogabe, S.; Yano, J.; Aertgeerts, K.; Kamiyama K. Design and Synthesis of Novel Human Epidermal Growth Factor Receptor 2 (HER2)/Epidermal Growth Factor Receptor (EGFR) Dual Inhibitors Bearing a Pyrrolo[3,2-d]pyrimidine Scaffold. *J. Med. Chem.* **2011**, *54*, 8030–8050.

(3) Ishikawa, T.; Miwa, K.; Seto, M.; Banno, H.; Kawakita, Y. FUSED HETEROCYCLIC COMPOUND. PCT Int. Appl. WO 2007064045 (A1), June, 2007.

(4) Aertgeerts, K.; Skene, R., Yona, J.; Sang, B.; Zou, H.; Snell, G.; Jennings, A.; Iwamoto, K.; Habuka, N.; Hirokawa, A.; Ishikawa, T.; Tanaka, T.; Miki, H.; Ohta, Y.; Sogabe, S. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2. *J. Biol. Chem.* **2011**, *286*, 18756–18765.

(5) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307–326.
(6) Vagin, A.; Teplyakov, A. MOLREP: an automated program for molecular replacement, *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.
(7) Bailey, S. The CCP4 suite – programs for protein crystallography, Acta Crystallogr. Sect. D **1994**, *50*, 760–763.
(8) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method *Acta Crystallogr.* **1997**, *D53*, 240–255.
(9) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of coot, *Acta Crystallogr.* **2010**, *D66*, 486–501.
(10) Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson, J.S.; Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr.* **2010**, *D66*, 12-21.

(a) (b) (c)

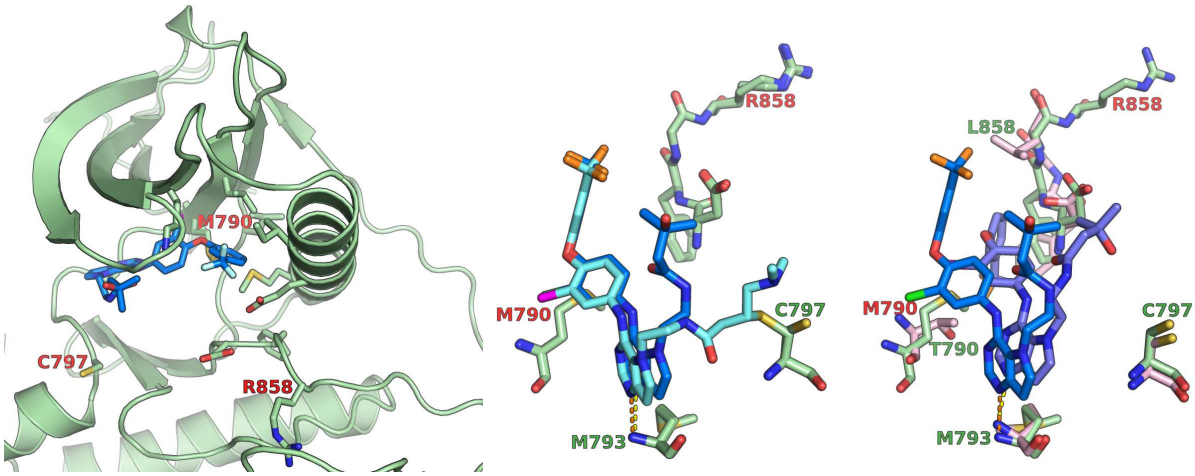


Figure S1. (a) Crystal structure of TAK-285 bound to the T790M/L858R mutant, colored in green. Two mutation sites and C797 are labeled. (b) Superposition of TAK-285 (blue) and compound **2** (cyan) from the mutant protein structures with key residues surrounding the binding site. Hydrogen bonds are indicated by orange and yellow dashed lines, respectively. (c) Superposition of TAK-285 from the mutant structure (blue) with TAK-285 from the wild type structure (purple-blue) with key residues surrounding the binding site. Hydrogen bonds are indicated by orange and yellow dashed lines, respectively.

Table S1. Data collection and refinement statistics

Crystal	Cmpd **1**	Cmpd **2**	Cmpd **3**	Cmpd **4**	Cmpd **4** (wild-type)
Data collection					
Beamline	5.0.3 (ALS)	5.0.3 (ALS)	5.0.3 (ALS)	NE3A (PF)	5.0.3 (ALS)
Space group	I23	I23	I23	$P4_32_12$	$P2_12_12_1$
Unit cell dimensions					
a, b, c (Å)	144.0, 144.0, 144.0	145.0, 145.0, 145.0	144.1, 144.1, 144.1	61.2, 61.2, 169.8	46.8, 68.4, 105
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	50–2.35	50–2.05	50–2.20	50–2.05	50–1.90
	(2.39–2.35)	(2.09–2.05)	(2.25–2.20)	(2.12–2.05)	(1.93–1.90)
Observed reflections	165812	174039	107138	113722	171444
Unique reflections	20719	31848	24849	20893	26379
Redundancy	8.0 (8.1)	5.5 (5.5)	4.3 (2.0)	5.4 (5.4)	6.5 (5.1)
Completeness (%)	100.0 (100.0)	99.9 (99.9)	97.5 (76.5)	98.8 (97.8)	96.6 (72.4)
I/σ	29.1 (2.6)	25.6 (2.0)	20.5 (2.1)	26.3 (2.0)	23.9 (2.3)
R_{sym}[a]	0.072 (0.878)	0.058 (0.910)	0.062 (0.333)	0.072 (0.917)	0.066 (0.781)
R_{meas}[b]	0.077 (0.941)	0.064 (0.990)	0.070 (0.429)	0.080 (0.997)	0.072 (0.879)
R_{pim}[b]	0.028 (0.329)	0.027 (0.423)	0.032 (0.260)	0.035 (0.430)	0.028 (0.377)
Molecules in ASU	1	1	1	1	1
Refinement					
Resolution (Å)	40–2.35	40–2.05	40–2.20	40–2.05	40–1.90
	(2.41–2.35)	(2.10–2.05)	(2.25–2.20)	(2.10–2.05)	(1.95–1.90)
Reflections	19606	30189	23565	19581	24920
R_{work}[b]	0.180 (0.240)	0.179 (0.268)	0.175 (0.228)	0.217 (0.506)	0.204 (0.347)
R_{free}[b]	0.216 (0.299)	0.206 (0.330)	0.200 (0.241)	0.275 (0.582)	0.258 (0.384)
Number of atoms					
Protein	2451	2480	2473	2394	2494
Ligand/Ion	38	39	52	44	50
Water	37	104	96	76	50
Average B factor (Å^2)[d]	54.6	47.8	43.4	53.4	51.4
Rms deviation from ideal geometry					
bond lengths (Å)	0.009	0.010	0.008	0.009	0.010
bond angles (°)	1.366	1.334	1.270	1.425	1.305
Ramachandran plot (%)[e]					
Preferred regions	96.4	97.0	97.0	95.2	97.4
Allowed regions	3.3	2.7	2.7	3.8	2.3
Outliers	0.3	0.3	0.3	1.0	0.3
PDB code	3W2O	3W2P	3W2Q	3W2R	3W2S

[a] $R_{sym}=\Sigma_h\Sigma_i|I(h)_i-<I(h)>|/\Sigma_h\Sigma_i<I(h)>$, where $<I(h)>$ is the mean intensity of symmetry-related reflections. [b] R_{meas} (=R_{rim}) and R_{pim} were calculated with SCALA in the CCP4 program suite. $R_{meas}=\Sigma_h[N/(N-1)]^{1/2}\Sigma_i|I(h)_i-<I(h)>|/\Sigma_h\Sigma_i<I(h)>$, $R_{pim}=\Sigma_h[1/(N-1)]^{1/2}\Sigma_i|I(h)_i-<I(h)>|/\Sigma_h\Sigma_i<I(h)>$. [c] $R_{work}=\Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$. R_{free} was calculated for randomly chosen 5% of reflections excluded from refinement. [d] B-factor includes contributions from TLS parameters. [e] Calculated with Coot. Values in parentheses are for the highest resolution shell.