

Research Article

The structure of human GCN2 reveals a parallel, back-to-back kinase dimer with a plastic DFG activation loop motif

 Taiana Maia de Oliveira¹, Victoria Korboukh², Sarah Caswell³, Jon J. Winter Holt⁴, Michelle Lamb⁵, Alexander W. Hird⁵ and Ross Overman³

¹Structure, Biophysics and FBLG, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; ²Mechanistic Biology and Profiling, Discovery Sciences, R&D, AstraZeneca, Boston, US; ³Discovery Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK; ⁴Medicinal Chemistry, Oncology R&D, AstraZeneca, Cambridge, UK; ⁵Medicinal Chemistry, Oncology R&D, AstraZeneca, Waltham, MA, USA

Correspondence: Taiana Maia de Oliveira (taiana.maiadeoliveira@astrazeneca.com)



When activated by amino acid starvation, the stress sensing protein kinase GCN2 phosphorylates the eukaryotic initiation factor 2 alpha, inhibiting translation to conserve energy and facilitate cell survival. Amino acid starvation, particularly of tryptophan and arginine, affects immune tolerance by suppressing differentiation and proliferation of T-cells via activation of GCN2 kinase. In addition, the GCN2 pathway mediates cancer survival directly within the context of metabolic stress. Here, we report the first crystal structures of the human GCN2 kinase domain (KD) in complex with two inhibitors of different size, shape, and chemical scaffold. Three novel activation loop conformations representative of different activation states of the kinase are described. In addition, a novel dimerization organization for GCN2 is observed. This arrangement is consistent with the hypothesis that the GCN2 KD forms an antiparallel inactive dimer until uncharged tRNA binds to it and triggers conformational changes that shift the equilibrium to the active parallel dimer.

Introduction

GCN2, general control nonderepressible 2 kinase, is one of four known vertebrate kinases — GCN2, PERK, PKR, and HRI — that respond to cellular stress by deactivating the eukaryotic initiation factor 2 alpha (eIF2 α). More specifically, GCN2 acts as a metabolic sensor by binding to uncharged tRNA, whose levels are enhanced by nutrient starvation, and eliciting amino acid deficiency stress response via phosphorylation of Serine 51 of eIF2 α , which culminates in a reduction in overall translation. This adaptive response saves energy and regulates signaling pathways to facilitate survival [1].

Numerous studies implicate GCN2 as instrumental to medically relevant events as diverse as cancer immune evasion and long-term memory formation. In fact, GCN2 has a dual pertinence in cancer biology: in addition to the suppressing of T cell differentiation that supports immune tolerance within the tumor microenvironment; as a stress response kinase, it promotes tumor survival and aggressiveness by maintaining tumor metabolic homeostasis [2].

The molecular mechanism of GCN2 activation is elusive and intricate as it results from the interplay of five defined domains of the protein: a RWD region that acts as a regulatory site, a PKD (pseudokinase domain), a KD (kinase domain), a HisRS (histidyl-tRNA synthetase) related domain that binds uncharged tRNA and a CTD (C-terminal dimerization domain) that in addition contributes to tRNA binding [3]. Moreover, ribosome binding has also been implicated in the activation of GCN2 [4]. To date, structural information on GCN2 has been limited to the yeast and murine CTD [5], the yeast KD [6], and the murine RWD [7]. The crystal structure of the CTD shows an interdigitated intimately connected dimer in which its core beta sheet is formed through interactions of strands

Received: 11 March 2019
Revised: 10 December 2019
Accepted: 23 December 2019

Accepted Manuscript online:
23 December 2019
Version of Record published:
17 January 2020

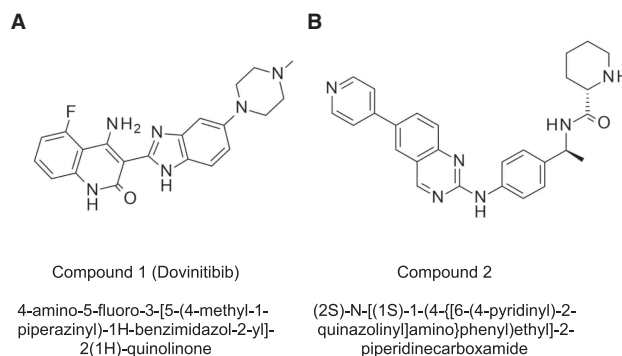


Figure 1. Chemical structures of GCN2 binding compounds co-crystallized with human GCN2.

(A) Compound 1 or dovinitib and (B) Compound 2 or (2S)-N-[(1S)-1-(4-[[6-(4-pyridinyl)-2-quinazolyl]amino]phenyl)ethyl]-2-piperidinecarboxamide.

from both polypeptide chains [5]. This supports the hypothesis that GCN2 is a constitutive dimer and activation upon tRNA binding would require a conformational change from the auto-inhibited state into the active one. Previous reports also implicate the KD and HisRS domain in the mediation of dimerization [8,6]. In addition, the PKD of interacts with the catalytic KD domain and inhibits its catalytic activity under non-starvation conditions [9].

Currently, the structural notions surrounding the mechanisms of eIF2 α kinases derive mainly from the crystal structures of the human PKR KD in complex with eIF2 α [10], yeast GCN2 KD in the hyperactive and native form [6], and human PERK KD [11]. Recognition of eIF2 α by these proteins seems to be achieved by a noncanonical orientation of the helix α G that accommodates eIF2 α by an induced fit mechanism that exposes its previously buried Ser51, allowing it to be phosphorylated [12]. Curiously, in these structures PKR organizes itself in a parallel dimer while, the yeast GCN2 is markedly different, being organized as an antiparallel dimer through N-lobe interactions [13].

Here, we report the first crystal structures of the human GCN2 KD in complex with two inhibitors of different size, shape, and chemical scaffold (Figure 1). In addition, three activation loop conformations representative of different activation states of the kinase are described. We show a range of conformational states can be stabilized by different ligands bound to the ATP binding pocket. This diversity can be potentially exploited for the development of conformation selective inhibitors for the treatment of cancer. Finally, these structures reveal a novel dimerization organization for GCN2 consistent with the hypothesis that the GCN2 KD would exist as an antiparallel inactive dimer until uncharged tRNA binds to it and triggers conformational changes that shift the equilibrium to the active parallel dimer.

Materials and methods

Molecular biology

For crystallographic studies, the constructs for the human eIF2 α kinase GCN2 (UniProtKB — Q9P2K8, E2AK4_HUMAN) were synthesized by GeneArt (Thermo Fisher Scientific) and cloned into a modified version of the pET28b vector. All constructs contained an N-terminal His6 tag followed by a TEV protease cleavage site upstream of the KD sequence (S577-T1020), together with an internal deletion to remove A663-P788 of the N-terminal lobe insertion region. In addition, one construct contained a D848N mutation to catalytically inactivate the kinase, and a second contained both T899A and T904A mutations to remove putative autocatalytic sites [14]. Protein sequences used for both constructs in this study are shown below.

6His-TEV-GCN2(S577-T1020, Δ A663-P788)

MHHHHHHGGGENLYFQGSSETQRQFSRYFIEFEELQLLGKGAFGAVIKVQNKLDGCCYAVKRPINPASRQ-FRRIKGEVTLLSRLHHENIVRYYNAWIERHERPSVTTEAVHYLYIQMEYCEKSTLRDITDQGLYRDTVRLWRLFREILDGLAYIHEKGMHRDLKPVNIFLDSDDHVKIGDFGLATDHLAFSADSKQDDQTGDLIKSDPSGHILTGMVGTALYVSPEVQGSTKSAYNQKVDLFSLGIIFFEMSYPHVMVTASERIFVLNQLRDPTSPKFPEDFDDGEH-AKQKSVISWLLNHPAKRPTATELLKSELLPPPQMESELHEVLHHTLT

6His-TEV-GCN2(S577-T1020,ΔA663-P788,D848N)

MHHHHHHGGGENLYFQGSSETQRQFSRYFIEFEELQLLGKGAFGAVIKVQNKLDGCCYAVKRIPINPASRQ-FRRIKGEVTLLSRLHHENIVRYYNWIERHERPSVTTEAVHYLYIQMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIHEKGMHRNLKPVNIFLDSDDHVKIGDFGLATDHLAFSADSKQDDQTGDLIKSDPSGHLMVGTALYVSPEVQGSTKSAYNQKVDLFLSLGIFFEMSYHPMVTASERIFVLNQLRDPTSPKFPEDFDDGEHAKQKSVISWLLNHDPAKRPTATELLKSELLPPPQMESELHEVLHHTLT

6His-TEV-GCN2(S577-T1020,ΔA663-P788, T899a,T904A)

MHHHHHHGGGENLYFQGSSETQRQFSRYFIEFEELQLLGKGAFGAVIKVQNKLDGCCYAVKRIPINPASRQ-FRRIKGEVTLLSRLHHENIVRYYNWIERHERPSVTTEAVHYLYIQMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIHEKGMHRDLKPVNIFLDSDDHVKIGDFGLATDHLAFSADSKQDDQTGDLIKSDPSGHLAGMVGAAALYVSPEVQGSTKSAYNQKVDLFLSLGIFFEMSYHPMVTASERIFVLNQLRDPTSPKFPEDFDDGEHAKQKSVISWLLNHDPAKRPTATELLKSELLPPPQMESELHEVLHHTLT

Mass spectrometry analysis for phosphorylation state

To obtain quantifiable phosphorylation data, kinase samples at 1 mg/ml in final buffer were loaded onto a Micromass LCT ES-TOF (liquid chromatography electrospray ionization time-of-flight) mass spectrometer, using a Waters 2790 HPLC as the inlet. Fifteen micrograms of protein was injected for each measurement onto a Phenomenex Jupiter 5 m C5 300A column, 150 × 2.0 mm. Protein was eluted using a fast gradient (0–90% B over 45 min at 120 ml/min; eluent A was aqueous 0.1% TFA, eluent B was 90% acetonitrile 0.1% TFA). Electrospray mass spectrometer data were collected between 12 and 25 min post injection, and deconvoluted using MaxEnt1 software (Waters). Theoretical protein masses were calculated using the MassLynx™ software (Waters).

Biochemical activity assay

Reagents

GFP-eIF2S1 and Tb-anti-peIF2alpha [pSer52] antibody were from ThermoFisher (PV4809 & PV4815, respectively), *Escherichia coli* tRNA was from Roche, sold by Sigma (10109541001).

Protein constructs used

G-GCN2-(S577-T1020, dA663-P788), ~7 phosphorylation sites.

GSGGG-6His-GCN2 (S572-G1024)-Avi, phosphorylated (10–15 sites)

6His-GCN2 (S572-G1024)-Avi, has ~5 phosphorylation sites.

GSGGG-6His-GCN2 (S572-G1024)-Avi, dephosphorylated.

Full-length GCN2 (lack of threonine 899 phosphorylation confirmed by Western Blot).

Experimental part

In brief, 1 μl of 2× GCN2 solution in 1× reaction buffer (20 mM Tris pH 7.5, 5 mM MgCl₂, 1 mM DTT, 0.005% Brij-35, and 0.5 mM EGTA) was added to a 1536-well plate (20 and 5 nM final concentration in 2 μl reaction for KD and FL construct, respectively). Reactions were initiated by adding 1 μl of 2× ATP/GFP-eIF2S1 (100 μM/25 nM final concentration) and ATP/tRNA/GFP-eIF2S1 (100 μM/0.2 mg/ml/25 nM final concentration) for one of the FL reactions. Plates were gently tapped to assure proper mixing of reagents. Reactions were then incubated for indicated time periods and quenched by the addition of 0.5 μl EDTA solution to final concentration of 20 mM (in 1× reaction buffer + 0.15% BSA). After the last time point, 0.5 μl of antibody solution (in 1× reaction buffer + 0.15% BSA) was added to all wells to 2 nM final concentration. To ensure proper mixing of reagents plate was first gently tapped and then spun for 1 min at 1000 rpm at RT. Plate was then sealed and incubated for 90 min at RT in the dark. Plates were then read on a BMG PHERAstar FSX (Ex.340, Em. 490/520, fly mode).

IC50 determination

Twenty nanoliters of compounds covering concentration range from 30 μM to 1 nM were added to 1536 well plates (Greiner 782075). One microliter of 2× GCN2 solution in 1× reaction buffer (20 mM Tris pH 7.5, 5 mM MgCl₂, 1 mM DTT, 0.005% Brij-35, and 0.5 mM EGTA) was added to each well of a 1536-well plate (1 nM

final concentration in 2 μ l final volume). Reactions were initiated by adding 1 μ l of 2 \times ATP/tRNA/GFP-eIF2 α (100 μ M/0.2 mg/ml/25 nM final concentration). Plates were gently tapped to ensure proper mixing of reagents. Reactions were then incubated for 70 min at RT and quenched by addition of 1 μ l 3 \times quench solution (3 μ M of potent GCN2 inhibitor 4-(5-((1-(tetrahydro-2H-pyran-4-yl)-1H-pyrazol-4-yl)amino)-3H-[1,2,3]triazolo [4,5-d]pyrimidin-3-yl) benzamide, 6 nM p-GFP-eIF2 α (phospho S52) Tb-Antibody 0.15% BSA in reaction buffer). To ensure proper mixing of reagents plates were gently tapped and then sealed and incubated for 90 min at RT in the dark. Plates were then read on a BMG PHERAstar FSX (Ex.340, Em. 490/520, fly mode). For artifact plates 2 μ l of fully phosphorylated GFP-eIF2 α solution (1.5 nM) were added to compound wells, followed by 1 μ l of 3 \times quench solution. Data were analyzed using Genedata Screener, for artifact plates to correct for compound fluorescence correction, a two-point normalization based on central (Neutral) and scale reference (stimulator/inhibitor control) well types correction was performed [15] In brief, it attempts to correct for the interference caused by compounds by either subtracting excess signal or by multiplying to add back reduced signal measured in separate compound-only plate (artifact plate). Correction of the assay plates by the artifact plates is limited to a two-fold correction (increase or decrease in signal). If the compound interference exceeds this limit, the well in question is masked for the purposes of fitting.

Protein expression and purification

The 6His-TEV-GCN2(S577-T1020, Δ A663-P788,D848N) construct was transformed into BL21 GOLD (DE3) cells (Agilent Technologies) and expressed overnight at 18°C, using 0.1 mM IPTG. Cell pellets were resuspended in a lysis buffer consisting of 40 mM HEPES pH 8.0, 500 mM NaCl, 1 mM TCEP, 20 mM imidazole, Benzonase[®] Nuclease (Merck KGaA) and cComplete EDTA-free protease inhibitors (Roche) and lysed by a single pass through a cell disruptor (Constant Systems Ltd) at 25 kpsi. The cell lysate was clarified by centrifugation at 70 000 \times g and proteins purified from the supernatant by batch-binding using Ni-NTA Superflow (QIAGEN). Following elution in 40 mM HEPES pH8.0, 300 mM NaCl, 1 mM TCEP, 500 mM imidazole, the N-terminal His6 tag was removed by overnight dialysis (40 mM HEPES pH8.0, 300 mM NaCl, 0.5 mM TCEP) in the presence of TEV protease (3 mg TEV/100 mg target protein). Samples were further purified by reverse IMAC followed by size exclusion chromatography using a Superdex 75 column (GE Healthcare) into a final buffer composed of 40 mM HEPES pH 7.9, 200 mM NaCl, 1 mM TCEP. An identical purification method was followed for 6His-TEV-GCN2(S577-T1020, Δ A663-P788) and 6His-TEV-GCN2(S577-T1020, Δ A663-P788, T899A,T904A), however, the latter construct was co-expressed with full length, untagged Lambda Protein Phosphatase (UniProtKB — P03772) on a pCDF-1b plasmid (Novagen), using *E. coli* BL21 Star[™] (DE3) cells (ThermoFisher).

Crystallization of human GCN2 KD

The protein samples used for crystallization were concentrated to 9 mg/ml in the purification buffer 40 mM HEPES pH 7.9, 200 mM NaCl, 1 mM TCEP. The compounds were added to final concentration of 50 μ M (5% DMSO-Protein-Buffer) incubated for 30 min and centrifuged for 10 min at 9000g before plates were setup. Crystals were obtained by the hanging-drop vapor-diffusion method under conditions consisting of 0.1 M MES pH 6, 20% PEG 8000 with 0.2 M calcium acetate (GCN2 + dovitinib) and 0.1 M BIS-TRIS pH 5.5–6.5, 20–25% PEG335 with 0.2 M ammonium sulfate (GCN2+(2S)-N-[(1S)-1-(4-{[6-(4-pyridinyl)-2-quinazolinyl]amino} phenyl)ethyl]-2 piperidinecarboxamide at room temperature. The drops were composed of 2 μ l of the protein-compound mixture and 2 μ l of the well solution. Finally, the crystals were soaked in cryoprotectant consisting of 25% ethylene glycol and immediately frozen in liquid nitrogen.

Data collection and structure determination

Diffraction data from the GCN2 crystals were collected remotely at the Diamond Beamline I04. Data were integrated with XDS, scaled with AIMLESS and the structures were determined by the molecular-replacement method using the program MolRep. The models of GCN2 KD were manually built using COOT [16] and further refined using BUSTER [17]. X-ray data collection and refinement statistics are presented in Table 1. The structure of GCN2:aminoquinazoline at 2.3 Å resolution displays R_{factor} and R_{free} of 0.21 and 0.24 while the structure of GCN2:dovinitib at 2.8 Å resolution displays R_{factor} and R_{free} of 0.21 and 0.26. The relatively large difference between R factor and R free of the latter is probably a consequence of the resolution. No residues are in the disallowed region.

Table 1 X-ray data collection and refinement statistics

	GCN2:dovinitib	GCN2:aminoquinazoline
Space group	C222 ₁	P1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	<i>a</i> = 155, <i>b</i> = 162, <i>c</i> = 123	<i>a</i> = 73.8, <i>b</i> = 77.79, <i>c</i> = 101.69
α , β , γ (°)	90, 90, 90	89.86, 90.05, 68.58
Resolution (Å)	49.36–2.8	2.3
<i>R</i> _{merge}	0.11	0.07 (0.67)
<i>I</i> / σ <i>I</i>	10.7 (0.9)	3.6 (0.9)
Completeness (%)	99.5 (94.1)	96.3 (85)
Number of reflections	38 997	90 485
<i>R</i> _{factor} / <i>R</i> _{free}	0.21/0.24	0.21/0.24

Results and discussion

Crystallization of human GCN2

As seen in other eIF2 α kinases, GCN2 features a unique loop inserted within the N-lobe of the KD. However, in GCN2, the loop is over 100 amino acids long and was thus engineered out of the crystallizable constructs (Δ 663–788). Optimization of domain boundaries was based on sequence alignments, secondary structure prediction, and analysis of the yeast GCN2 and human PERK structures. Construct 6His-TEV-GCN2(S577-T1020, Δ A663-P788) was expressed in *E. coli* but MS analysis revealed the protein to be highly autophosphorylated (Supplementary Figure S1). We opted then to introduce inactivating mutations at the catalytic loop and at the activation loop and expressed the following (respectively): GCN2 S577-T1020, Δ A663-P788, D848N and GCN2 S577-T1020, Δ A663-P788, T899A, T904A. Sample homogeneity was accessed by SEC-MALS (Supplementary Figure S2) and activity was probed by a biochemical assay (Supplementary Figure S3) as described in the Supplementary Methods

This allowed us to obtain stable protein, which was, however, still refractory to crystallization in the absence of ligands. Pre-incubation of GCN2 S577-T1020, Δ A663-P788, D848N and S577-T1020 construct with Dovitinib (Compound 1; IC₅₀ 0.9 μ M) and GCN2 S577-T1020, Δ A663-P788, T899A, T904A with compound (2*S*)-*N*-[(1*S*)-1-(4-[[6-(4-pyridinyl)-2-quinazoliny]amino}phenyl)ethyl]-2-piperidinecarboxamide (Compound 2, IC₅₀ 6.8 μ M) resulted in crystals diffracting between 2.3 and 3.5 Å. Compound 1 is a kinase inhibitor that binds to multiple kinases, including FLT3, PDGFR, and FGFR. It was developed by Novartis and was subjected to clinical trials for the treatment of diverse types of cancer [18]. Crystals of the complexes were obtained in 0.1 M MES pH 6, 20% PEG 80000 with 0.2 M calcium acetate and 0.1 M BIS-TRIS pH 5.5–6.5, 20–25% PEG3350 with 0.2 M ammonium sulfate, respectively.

Crystal structure of human GCN2

The structures were solved by molecular replacement with yeast GCN2 as the search model (PDB code:1ZYD) [6] (N-lobe and C-lobes replaced separately). As shown in Figure 2, GCN2 largely adopts the canonical protein kinase fold consisting of a beta-stranded N-terminal lobe connected via the flexible hinge region (residues 801–806) to the alpha-helical C-terminal lobe. Similarly to PKR, the smaller N-lobe (residues 590–880) is preceded by a short noncanonical α -helix α 0 (residues 584–589) and contains a five-stranded (β 1– β 5) twisted antiparallel β -sheet with a conserved helix α C (residues 628–642) linking β 3– β 4. The α C-helix is displaced from the canonical active arrangement and the large loop insert characteristic of the eIF2 α kinase family members, truncated here to aid crystallization (residues 663–788), occurs between β 4 and β 5. The bulkier C-lobe (residues 807–1006) containing amino acids implicated in activation (Activation loop, A loop), and peptide substrate recognition (catalytic loop) is predominantly α -helical (α D– α I). The DFG motif in the N-terminal part of the activation loop adopts diverse conformations (as detailed subsequently) while the HRD motif (846–848) characteristic of the catalytic loop that anchors the substrate to be phosphorylated retains the classical active-like state independently of the arrangement of the A loop. A notable exception deviating from the canonical

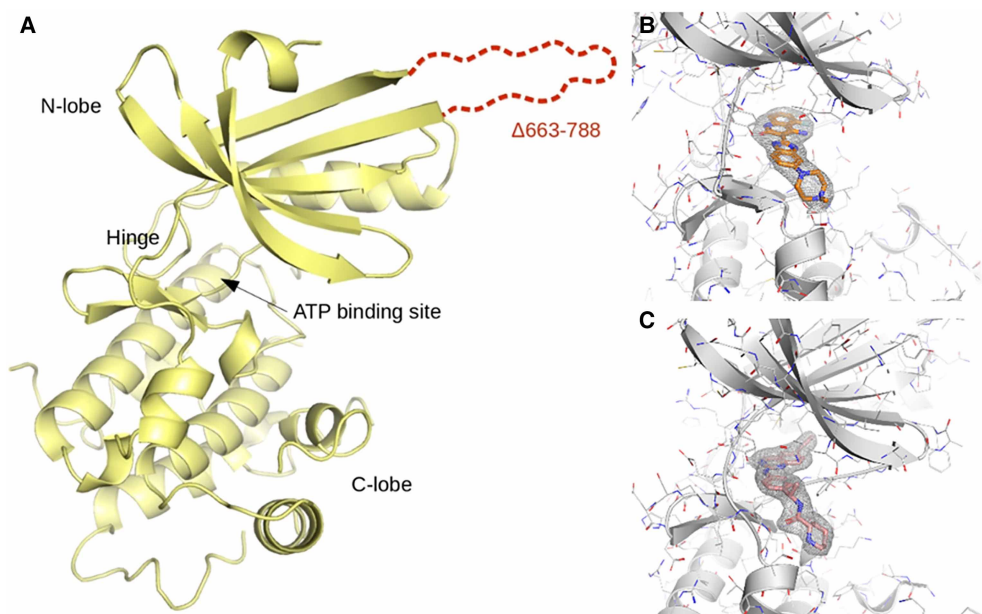


Figure 2. The crystal structure of human GCN2 kinase domain.

(A) Kinase domain canonical elements the N-terminal lobe, the C-terminal lobe and the hinge that connects the two kinase lobes are highlighted in the structure of GCN2. The insertion loop represented in red was removed to improve crystallizability. (B) Human GCN2 complexed to Compound 1. (C) Human GCN2 complexed to Compound 2.

configuration in the C-lobe is helix α G which is notably shifted, similarly to the structures of PKR and yeast GCN2 (Figure 2). Both inhibitors are clearly defined by difference electron density (Figure 2); they occupy the ATP binding site in the deep cleft between the N- and C-lobes of the kinase.

GCN2:Compound 1

4-amino-5-fluoro-3-[5-(4-methyl-1-piperazinyl)-1H-benzimidazol-2-yl]-2(1H)-quinolinone

GCN2 Complexed to dovitinib crystallized in space group $C222_1$ with 4 molecules in the asymmetric unit (cell parameters $A = 160 \text{ \AA}$, $B = 165 \text{ \AA}$, $C = 121 \text{ \AA}$). Compound 1 occupies the ATP binding site lined by an active-like conformation of the conserved N-terminal region of the activation loop of the kinase (Figure 2B). This conformation, so called, DFG-in, is characterized by the side chain of the aspartic acid pointing towards the ATP site where the phosphates of ATP would be and the phenylalanine pointing away from the site. The quinolinone group establishes hydrogen bonds with the hinge through the backbone carbonyl of Glu 803 and the amine of Cys 805 while a nitrogen from the benzimidazole interacts with the amide of Cys 805. In addition, there are hydrophobic contacts between the quinolone with Phe 855 and the gatekeeper residue Met 802. The methyl-piperazine tail extends out into the solvent channel; in two of the four molecules of the asymmetric unit, the piperazine is in position to establish a polar interaction with Asp 812. Protein:compound interactions can be seen in Supplementary Figure S4.

In two of the molecules, the activation loop is partially ordered (Asp 866 to Ala 874) and extend out into the solvent. There is the clear density for the DFG motif (Asp 866, Phe 867, and Gly 868) in all four molecules, and they are all poised in a DFG-in configuration. Interestingly, the critical catalytic salt bridge between Lys 619 in β -strand 3 and Glu 636 in the regulatory α C-helix, required by catalytic competent configurations of the KD, is absent. Instead, in all four molecules, Lys 619 is engaged in a salt bridge with the DFG-Asp 866 residue as in the dormant version of Src kinase [19]. Crystal structures of dovitinib complexes are available for FGFR1 [20], FGFR4 [21], and STK10 [22]. The superimposition of these structures reveals a similar binding mode, with most of the variation located in the solvent-exposed piperazine. Inhibitors that reach into the hydrophobic back pocket between the α C-helix and the β -sheet that forms the N-lobe of the kinase have been reported to stabilize kinase configurations where the catalytic bridge is broken and the C-helix is disengaged (α C-out).

Though dovitinib does not extend far into this pocket, the structures with FGFR4 and STK10 also lack the catalytic bridge. In these structures dovitinib behaves as a Type 1.5 inhibitor, that is, it binds to a DFG-in which notwithstanding represents an inactive conformation of the kinase.

GCN2:Compound 2

The GCN2:aminoquinazoline complex crystallized in space group P1 ($a = 73.80$, $b = 77.79$, $c = 101.69$; 89.86 , 90.05 , 68.58) with eight molecules in the ASU and it has been solved at 2.3 \AA resolution. Six of these molecules (chains A,B,C,D,E,G) exhibit a DFG-out conformation, with a specific inactive arrangement of the DFG motif (Figure 2). Since the serendipitous discovery of type II inhibitor imatinib there has been great interest toward the development of inhibitors with comparable binding mode. Type II inhibitors additionally extend into an allosteric pocket directly adjacent to the ATP binding site by replacing the Phe of the conserved DFG motif, stabilizing the aforementioned 'DFG-out' configuration. This induced fit translates to slow binding kinetics and non-ATP competitive mode of inhibition. The susceptibility of kinases to type II inhibitors varies, and it remains unclear whether all kinases are able to undergo the required conformational rearrangement of the DFG motif. Our crystal structure shows GCN2 is able to adopt this conformation and therefore can accommodate type II inhibitors. The electron density for chain D allows for the substantial model building of activation loop: we can trace the segment from the Asp 866 until Ala875 and from Gln903 onwards.

In the other two molecules (chains F and H), the activation loop adopts an intermediate state between classical DFG-in and classical DFG-out (Figure 3). In typical DFG-in arrangements the aspartic acid points towards the ATP cleft, where the β phosphate would be located, while the phenylalanine points away from the cleft. In contrast, in typical DFG-out configurations the aspartic acid and phenylalanine point in almost opposite directions of DFG-in conformation. In chains F and H of human GCN2: Compound 2 complex, Asp 866 and Phe 867 point to the same overall direction: when compared with the DFG-in arrangement of the GCN2: Compound 1 structure, Asp 866 is rotated 40° and points towards the gatekeeper 802, while Phe867 is rotated upwards 100° and occupies a pocket above the side chain of C-helix Glu636 rather than stacking against the HRD His-846.

Structures capturing intermediates of the DFG flipping process have been reported elsewhere and named as DFG-up, pseudo DFG-out or atypical DFG-out [23]. They have reduced pocket volume in average when compared with classical DFG-out arrangements and are not compatible with a typical type II inhibitor [23]. In contrast with other non-classical DFG structures [24,25], the activation loop of this atypical DFG arrangement is positioned closer to the DFG-in conformation (RMSD 3.3 \AA) (shown on the GCN2:dovitinib structure) than the classical DFG-out (RMSD 4.34 \AA) GCN2:aminoquinazoline compound — chains A,B,C,D,E,G) (Figure 4).

The x-ray structure indicates that the aminoquinazoline group makes two hydrogen bonds to the hinge of the kinase through the backbone of the Cys 805. The pyridine extends to the start of the back cleft occupying the space in the vicinity of the gatekeeper Met 802 and can have a polar interaction with the backbone of the DFG-Phe (chain H) when the protein is in the DFG-up position. The piperazine ring is exposed to the solvent channel and can make a hydrogen bond with Asp 812. Interactions can be seen in Supplementary Figure S5.

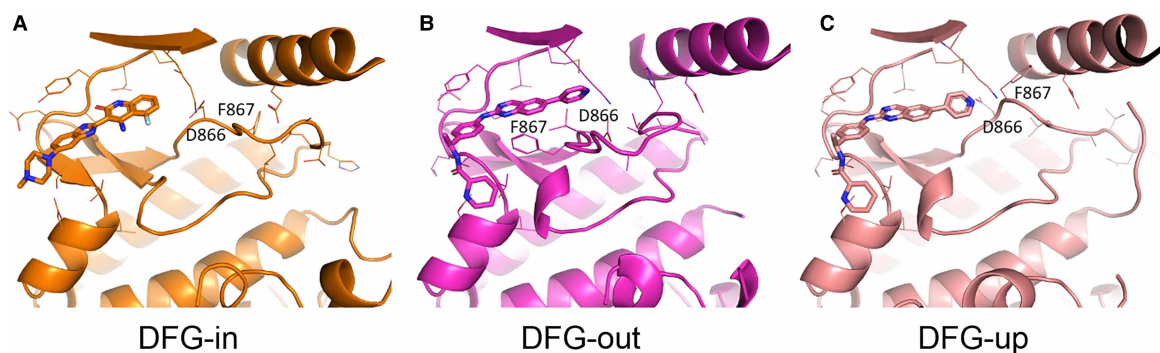


Figure 3. The different conformations adopted by the activation loop of human GCN2.

(A) Co-crystal structure of Compound 1 bound to GCN2 – the activation loop adopts the active DFG-in conformation. (B) and (C) Co-crystal structure of Compound 2 bound to GCN2 – in the different molecules within the asymmetric unit of this structure, the activation loop adopts either an inactive-DFG-out (B) or a DFG-up conformation (C).

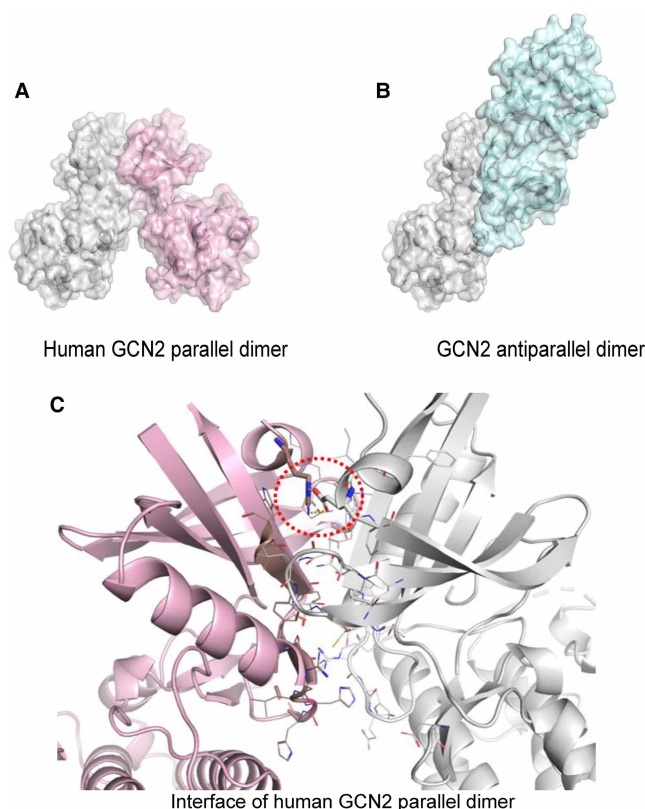


Figure 4. GCN2 dimeric organization.

The two top panels show a comparison of the dimeric structures of the kinase domains of **(A)** human GCN2 and **(B)** yeast GCN2. Human GCN2 dimerizes in a parallel orientation similar to PKR. Though yeast GCN2 dimerizes via a partially overlapping interface, the monomers are organized in antiparallel fashion. **(C)** Shows the interface of the back to back dimer highlighting the crucial electrostatic interaction between Arg 585 and Glu 589 (red dotted circle).

The quality of the density for the latter varies on different molecules in the asymmetric unit, and it is not clear the group is always in a good position to establish this interaction.

In short, this inhibitor binds to classical and non-classical DFG-out conformations of GCN2 while not extending far into the back cleft, and might be classified as a type II B inhibitor [26]. In general, these have a shorter residence time in relation to typical type II inhibitors (so-called type II A).

Regulation of human GCN2 activity

Structural studies based on yeast GCN2 postulate the kinase is inherently inactive because of the following characteristics: the catalytic salt bridge is broken and the C-helix is displaced out of its active position and there is an asparagine flap (Asn 793) that points toward the ATP binding site and sterically restricts access to it [6]. In addition, a network of interactions of residues in the vicinity of the hinge rigidifies the hinge interfering with the opening of the ATP cleft. In the human counterpart, one of the critical residues is an asparagine, which is replaced by a similarly bulky residue, Lys 807, that could perform a comparable function. The C-terminal portion of the hinge backbone that contains this residue in our structures is shifted 6 Å outwards when compared with the yeast GCN2 structure and in addition, the side chain of the lysine points away from the pocket, suggesting that this mechanism of autoinhibition might not be shared by humans. As we were not able to obtain a structure of human GCN2 in the absence of compounds, it could be possible that a similar mechanism does exist in humans and we do not observe the conformation simply because both inhibitors stabilize a version of the enzyme in which the flap is released. We, however, add that the MS analysis revealed autophosphorylation of the protein produced in *E. coli*, corroborating that the wild type KD of the human protein is not in an auto-inhibited state, at least when the insertion loop is not present.

In yeast GCN2, a single mutation from the arginine located in the C-terminal part of the hinge to Glycine (R794G — yeast numbering) increases the activity of the protein 75-fold [6]. This change was attributed to an enhancement of the flexibility of the hinge that interferes with the intramolecular interactions that keep the flap inside the ATP binding cleft and interfere with the interlobal movement required for substrate binding. Human GCN2 contains a serine (Ser 808) instead of an arginine at this position. This substitution might already be enough to decrease the rigidity of the hinge sufficiently to make the human GCN2 kinase more active than the yeast protein.

Dimerization has been reported to play a role in the activation of all four eIF2 α kinases [27]: Previous crystal structures of the KDs of PERK and PKR revealed a back-to-back parallel dimer is formed by predominantly N-lobe interactions between residues conserved within the kinase family [28,10]. A crucial electrostatic interaction involves Arg 585 from one protomer with Glu 589 from the second protomer (human GCN2 numbering). Mutagenesis studies replacing either of the residues and thus interfering with the salt bridge formation decreased the activity of the GCN2, PERK, and PKR. Interestingly, the yeast GCN2 structure features symmetric back to back dimer which is distinct and instead organized in an antiparallel fashion (Figure 4). While most of the interface is common between the two dimeric configurations (which in both cases involve primarily the N-lobes of the kinases), the Arg–Glu pair are over 10 Å away from each other in the antiparallel dimer. Given that the yeast GCN2 KD is inherently inactive and thus adopts a non-productive arrangement, the current prevalent hypothesis is that the GCN2 dimerizes as an inactive antiparallel dimer similar to the PKR dimer and stays that way until the concentration of uncharged tRNA is increased and the HisRS histidyl-tRNA synthetase related domain binds to it, causing conformational change in the dimer that then re-organizes in the productive parallel fashion. Our structures show for the first time that a dimeric parallel conformation consistent with active PKR (Figure 4) can be indeed adopted by human GCN2 and strengthens the notion that GCN2 undergoes a transition from antiparallel to parallel dimer during its process of activation.

The dimeric PKR-like arrangement of GCN2 can be observed in both crystal forms: the dimers seen with compounds 1 and 2 look remarkably similar and can be superimposed with a low RMSD of 1.08 Å. When compared with the PKR parallel dimer, the GCN2 dimer is more closed (Supplementary Figure S6): the loops that connect the α C-helix and β 4 are next to each other and engage into polar and apolar interactions, amounting to 13–17 intermolecular hydrogen bonds that include a network between Arg 650 of one protomer and the main chains of Tyr 651 and Leu 543 (in the GCN2:Compound 2 structure). The buried surface area of the interface amounts to 1958–2202 Å² in GCN2, depending on the crystal form, and is considerably larger than in PKR (1370–1760 Å²).

In conclusion, the structures of human GCN2 presented in this study reveal a dimerization arrangement distinct from the yeast GCN2 dimeric assembly, corroborating the hypothesis that the GCN2 dimer transitions from antiparallel to parallel configuration upon activation. Activation of the GCN2 in response to amino acid deprivation is one mechanism by which cancer cells thrive in the nutrient stress setting of the tumor micro-environment [29]. Within this context, GCN2 remains interesting as an oncology target. Our structures highlight the plasticity of the activation loop and exhibit a rarer conformation that could be targeted in an effort to enhance the selectivity of GCN2 inhibitors.

Abbreviations

CTD, C-terminal dimerization domain; eIF2 α , eukaryotic initiation factor 2; HisRS, histidyl-tRNA synthetase; KD, kinase domain; LCT ES-TOF, liquid chromatography electrospray ionization time-of-flight; PKD, pseudokinase domain.

Author Contribution

T.M.O. designed the constructs and carried out crystallization experiment and solved the structures. V.K. carried the biochemical experiments. S.C. performed sample characterization by SEC-MALS. J.J.W.H., M.L., and A.W.H. selected the compounds and progressed the chemistry. R.O. designed the constructs, developed protocols for purification, purified, and characterized the samples. The manuscript was written by T.M.O. with contributions from all authors.

Funding

This project was funded by AstraZeneca R&D.

Competing Interests

All authors are employees and shareholders of AstraZeneca, the funder of this study.

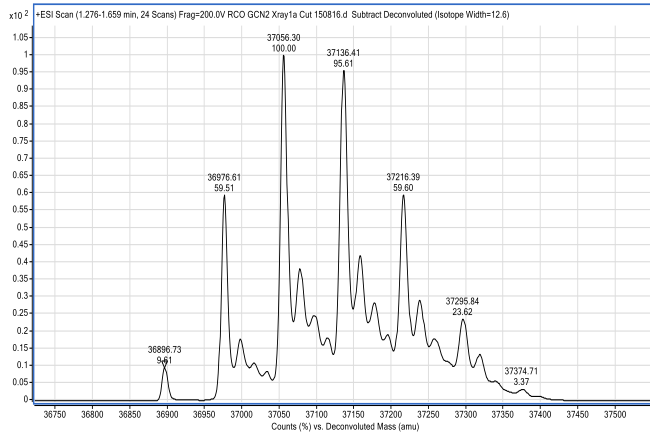
References

- Shenolikar, S. (2012) *Protein Phosphorylation in Health and Disease*, Academic Press, Orlando, FL.
- Timosenko, E., Hadjinicolaou, A.V. and Cerundolo, V. (2017) Modulation of cancer-specific immune responses by amino acid degrading enzymes. *Immunotherapy* **9**, 83–97 <https://doi.org/10.2217/imt-2016-0118>
- Castilho, B.A., Shanmugam, R., Silva, R.C., Ramesh, R., Himme, B.M. and Sattlegger, E. (2014) Keeping the eIF2 alpha kinase Gcn2 in check. *Biochim. Biophys. Acta* **1843**, 1948–1968 <https://doi.org/10.1016/j.bbamcr.2014.04.006>
- Inglis, A.J., Masson, G.R., Shao, S., Perisic, O., McLaughlin, S.H., Hegde, R.S. et al. (2019) Activation of GCN2 by the ribosomal P-stalk. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4946–4954 <https://doi.org/10.1073/pnas.1813352116>
- He, H., Singh, I., Wek, S.A., Dey, S., Baird, T.D., Wek, R.C. et al. (2014) Crystal structures of GCN2 protein kinase C-terminal domains suggest regulatory differences in yeast and mammals. *J. Biol. Chem.* **289**, 15023–15034 <https://doi.org/10.1074/jbc.M114.560789>
- Padyana, A.K., Qiu, H., Roll-Mecak, A., Hinnebusch, A.G. and Burley, S.K. (2005) Structural basis for autoinhibition and mutational activation of eukaryotic initiation factor 2 α protein kinase GCN2. *J. Biol. Chem.* **280**, 29289–29299 <https://doi.org/10.1074/jbc.M504096200>
- Nameki, N., Yoneyama, M., Koshiba, S., Tochio, N., Inoue, M., Seki, E. et al. (2004) Solution structure of the RWD domain of the mouse GCN2 protein. *Protein Sci.* **13**, 2089–2100 <https://doi.org/10.1110/ps.04751804>
- Qiu, H., Dong, J., Hu, C., Francklyn, C.S. and Hinnebusch, A.G. (2001) The tRNA-binding moiety in GCN2 contains a dimerization domain that interacts with the kinase domain and is required for tRNA binding and kinase activation. *EMBO J.* **20**, 1425–1438 <https://doi.org/10.1093/emboj/20.6.1425>
- Lageix, S., Rothenburg, S., Dever, T.E. and Hinnebusch, A.G. (2014) Enhanced interaction between pseudokinase and kinase domains in Gcn2 stimulates eIF2 α phosphorylation in starved cells. *PLoS Genet.* **10**, e1004326 <https://doi.org/10.1371/journal.pgen.1004326>
- Dar, A.C., Dever, T.E. and Sicheri, F. (2005) Higher-order substrate recognition of eIF2 α by the RNA-dependent protein kinase PKR. *Cell* **122**, 887–900 <https://doi.org/10.1016/j.cell.2005.06.044>
- Axten, J.M., Medina, J.R., Feng, Y., Shu, A., Romeril, S.P., Grant, S.W. et al. (2012) Discovery of 7-methyl-5-(1-[[3-(trifluoromethyl)phenyl]acetyl]-2,3-dihydro-1H-indol-5-yl)-7H-pyrrolo[2,3-d]pyrimidin-4-amine (GSK2606414), a potent and selective first-in-class inhibitor of protein kinase R (PKR)-like endoplasmic reticulum kinase (PERK). *J. Med. Chem.* **55**, 7193–7207 <https://doi.org/10.1021/jm300713s>
- Hinnebusch, A.G. (2005) eIF2 α kinases provide a new solution to the puzzle of substrate specificity. *Nat. Struct. Mol. Biol.* **12**, 835–838 <https://doi.org/10.1038/nmsb1005-835>
- Cole, J.L. (2007) Activation of PKR: an open and shut case? *Trends Biochem. Sci.* **32**, 57–62 <https://doi.org/10.1016/j.tibs.2006.12.003>
- Deng, J., Harding, H.P., Raught, B., Gingras, A.-C., Berlanga, J.J., Scheuner, D. et al. (2002) Activation of GCN2 in UV-irradiated cells inhibits translation. *Curr. Biol.* **12**, 1279–1286 [https://doi.org/10.1016/S0960-9822\(02\)01037-0](https://doi.org/10.1016/S0960-9822(02)01037-0)
- Shapiro, A.B., Walkup, G.K. and Keating, T.A. (2009) Correction for interference by test samples in high-throughput assays. *J. Biomol. Screen.* **14**, 1008–1016 <https://doi.org/10.1177/1087057109341768>
- Emsley, P. and Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 <https://doi.org/10.1107/S0907444904019158>
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W. et al. (2017) *Buster Version 2.11.2*, Global Phasing Ltd, Cambridge, U.K.
- Musolino, A., Campane, M., Neven, P., Denduluri, N., Barrios, C.H., Cortes, J. et al. (2017) Phase II, randomized, placebo-controlled study of dovitinib in combination with fulvestrant in postmenopausal patients with HR+, HER2– breast cancer that had progressed during or after prior endocrine therapy. *Breast Cancer Res.* **19**, 18 <https://doi.org/10.1186/s13058-017-0807-8>
- Roskoski, R. (2015) Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. *Pharmacol. Res.* **94**, 9–25 <https://doi.org/10.1016/j.phrs.2015.01.003>
- Klein, T., Vajpai, N., Phillips, J.J., Davies, G., Holdgate, G.A., Phillips, C. et al. (2015) Structural and dynamic insights into the energetics of activation loop rearrangement in FGFR1 kinase. *Nat. Commun.* **6**, 7877 <https://doi.org/10.1038/ncomms8877>
- Lesca, E., Lammens, A., Huber, R. and Augustin, M. (2014) Structural analysis of the human fibroblast growth factor receptor 4 kinase. *J. Mol. Biol.* **426**, 3744–3756 <https://doi.org/10.1016/j.jmb.2014.09.004>
- Elkins, J. M. 2017. Human STK10 bound to dovitinib. PDB databank (to be published) <https://doi.org/10.2210/pdb5owq/pdb>
- Vijayan, R.S.K., He, P., Modi, V., Duong-Ly, K.C., Ma, H., Peterson, J.R. et al. (2015) Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *J. Med. Chem.* **58**, 466–479 <https://doi.org/10.1021/jm501603h>
- Dodson, C.A., Kosmopoulou, M., Richards, M., Atrash, B., Bavetsias, V., Blagg, J. et al. (2010) Crystal structure of an Aurora-A mutant that mimics Aurora-B bound to MLN8054: insights into selectivity and drug design. *Biochem. J.* **427**, 19–28 <https://doi.org/10.1042/BJ20091530>
- Kuglstatler, A., Wong, A., Tsing, S., Lee, S.W., Lou, Y., Villaseñor, A.G. et al. (2011) Insights into the conformational flexibility of Bruton's tyrosine kinase from multiple ligand complex structures. *Protein Sci.* **20**, 428–436 <https://doi.org/10.1002/pro.575>
- Roskoski, R. (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol. Res.* **103**, 26–48 <https://doi.org/10.1016/j.phrs.2015.10.021>
- Hernández, G. and Jagus, R. (2016) *Evolution of the Protein Synthesis Machinery and Its Regulation*, Springer, Cham, Switzerland
- Cui, W., Li, J., Ron, D. and Sha, B. (2011) The structure of the PERK kinase domain suggests the mechanism for its activation. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 423–428 <https://doi.org/10.1107/S0907444911006445>
- Lehman, S.L., Ryeom, S. and Koumenis, C. (2015) Signaling through alternative Integrated Stress Response pathways compensates for GCN2 loss in a mouse model of soft tissue sarcoma. *Sci. Rep.* **5**, 11781 <https://doi.org/10.1038/srep11781>

Supplementary Figures

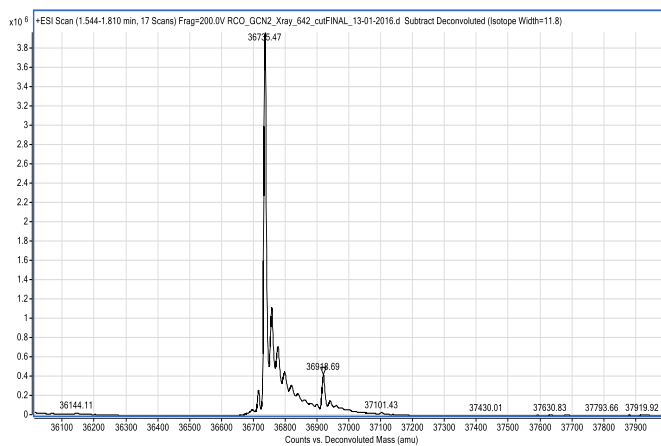
Supplementary Figure 1A

LC-MS of GCN2(S577-T1020,ΔA663-P788) to highlight multiple phosphorylations.



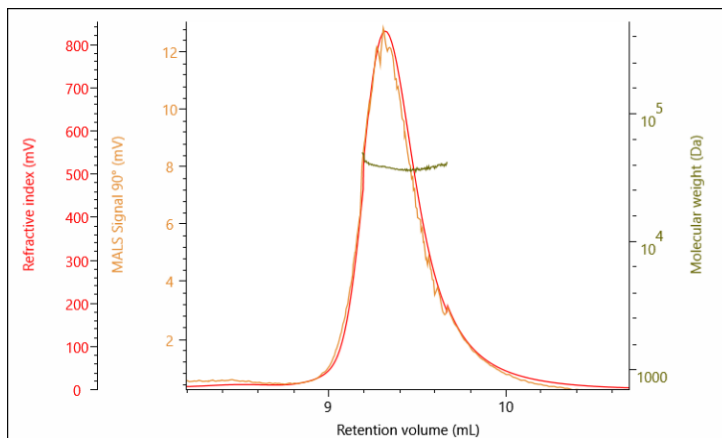
Supplementary Figure 1B

LC-MS of GCN2(S577-T1020,ΔA663-P788) to highlight lack of phosphorylations.



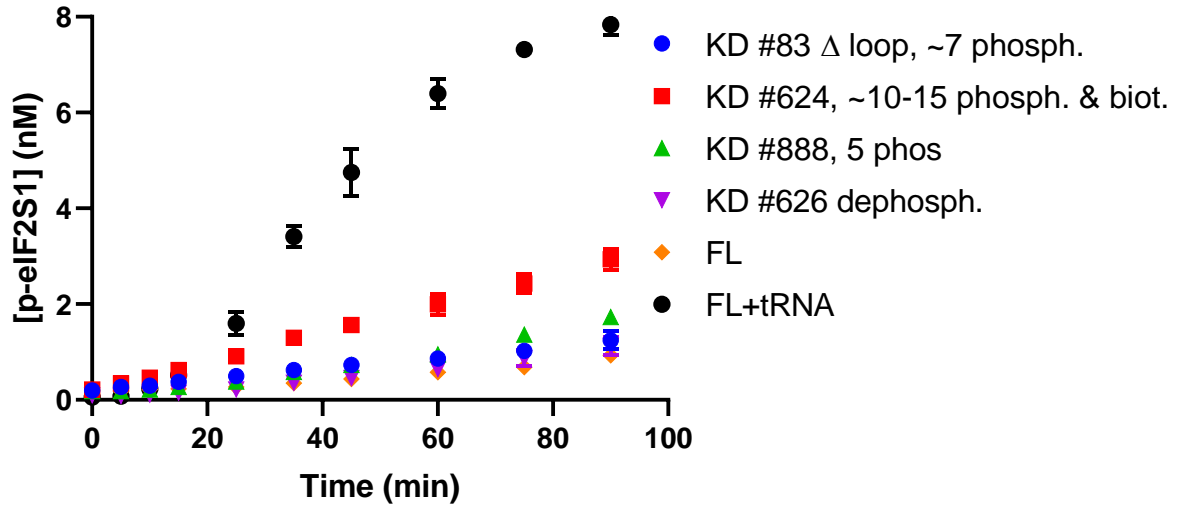
Supplementary Figure 2

SEC-MALS of human GCN2 Kinase Domain crystallization construct, GCN2 S577-T1020, Δ A663-P788 (2mg/mL), showing an average molecular weight of 38.5 kDa (consistent with a monomer)



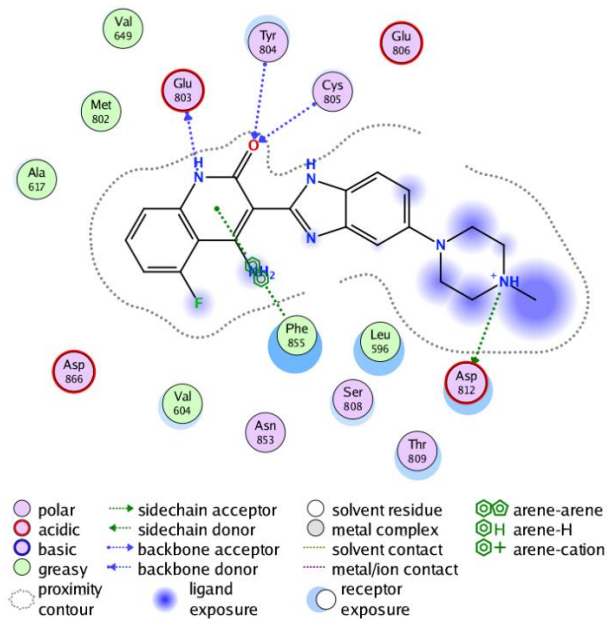
Supplementary Figure 3

Reaction progress curves for various GCN2 constructs determined according to the Biochemical Activity Assay session (Supplementary Materials). All four kinase domain constructs show relatively low levels of activity, with the highly phosphorylated construct #624 being most active, followed by partially phosphorylated, Δ loop & dephosphorylated KD constructs. Full length protein (FL) was tested at 4X lower concentration (5 nM) and showed minimal activity. However, in the presence of tRNA and after an ~20 minute lag which we attribute to an autophosphorylation step, the FL protein showed a dramatic increase in activity. (Datapoints above 5 nM p-eIF2S1 may be out of linear detection range). Data shown is an average of 3 independent experiments. Error bars are \pm SEM.



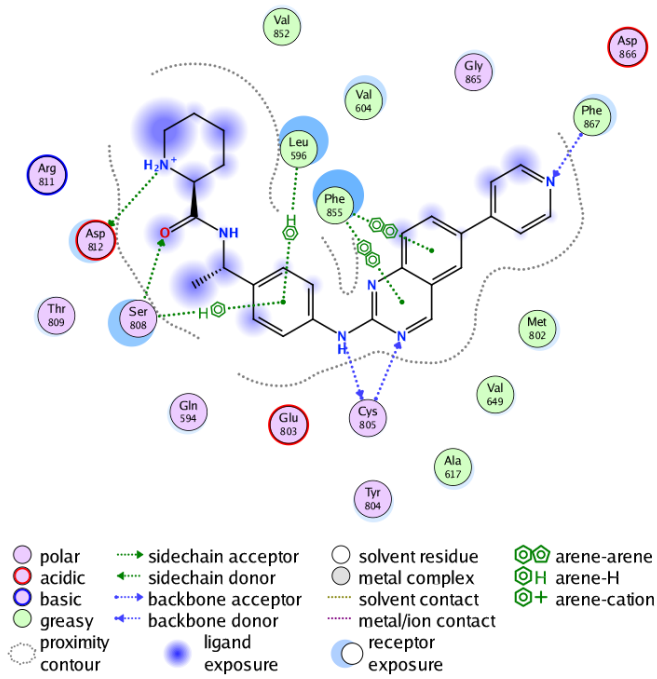
Supplementary Figure 4

Interaction map of Dovinitib:GCN2



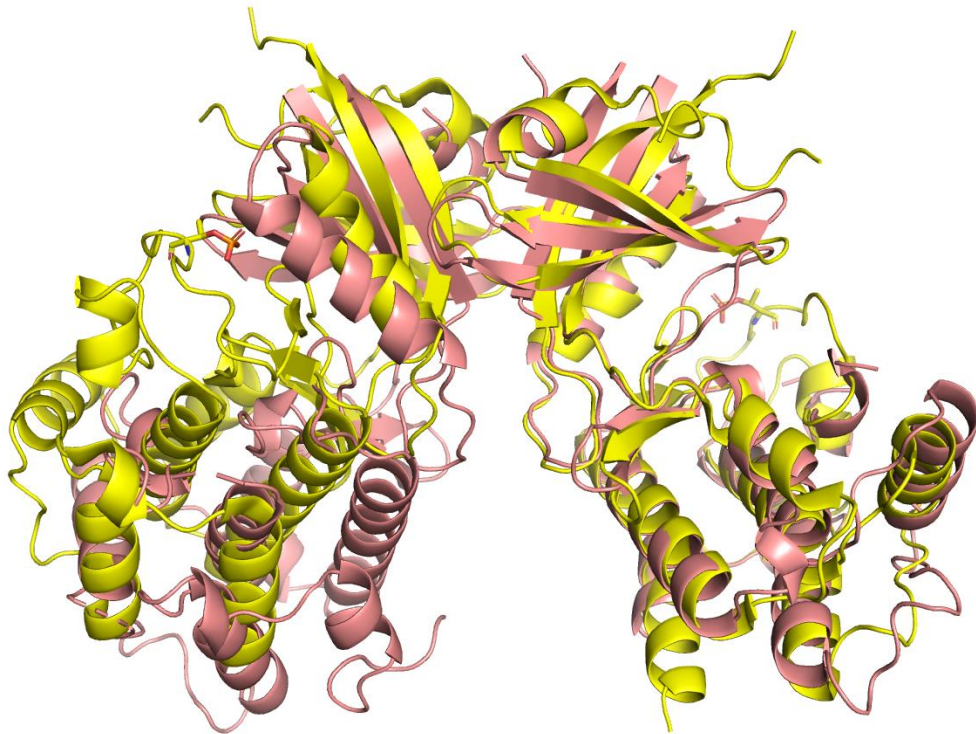
Supplementary Figure 5

Representative interactions of GCN2:(2S)-N-[(1S)-1-(4-[[6-(4-pyridinyl)-2-quinazolinyl]amino]phenyl)ethyl]-2 piperidinecarboxamide



Supplementary Figure 6

Human GCN2 (pink) adopts a dimeric arrangement similar to the one adopted by PKR (yellow)¹³. When compared to the parallel dimer of PKR, GCN2 is more closed.





Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Feb 20, 2019 – 09:53 am GMT

Deposition ID : D_1292100734
PDB ID : *(not yet assigned)*

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix)	:	1.13
EDS	:	2.1
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.1

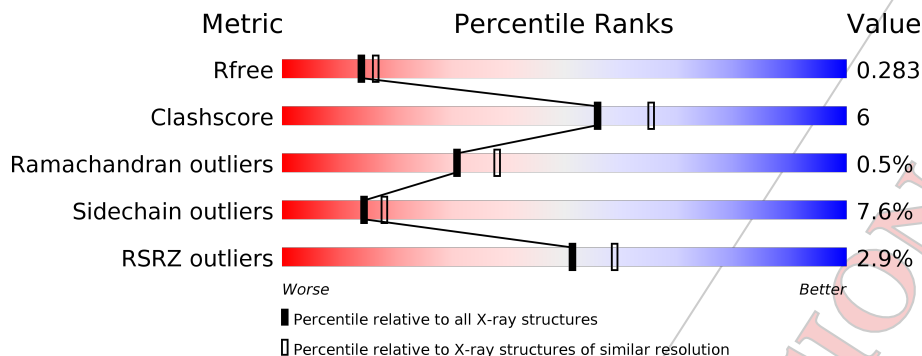
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



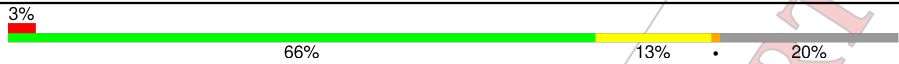

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	4477 (2.30-2.30)
Clashscore	122126	5072 (2.30-2.30)
Ramachandran outliers	120053	5022 (2.30-2.30)
Sidechain outliers	120020	5021 (2.30-2.30)
RSRZ outliers	108989	4374 (2.30-2.30)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	320	<div style="display: flex; align-items: center;"> <div style="width: 2%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 64%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 11%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 23%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">2% 64% 11% • 23%</p>
1	B	320	<div style="display: flex; align-items: center;"> <div style="width: 3%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 66%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 11%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 22%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">3% 66% 11% • 22%</p>
1	C	320	<div style="display: flex; align-items: center;"> <div style="width: 3%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 65%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 13%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 22%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">3% 65% 13% • 22%</p>
1	D	320	<div style="display: flex; align-items: center;"> <div style="width: 2%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 68%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 11%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 21%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">2% 68% 11% • 21%</p>
1	E	320	<div style="display: flex; align-items: center;"> <div style="width: 4%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 59%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 18%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 22%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">4% 59% 18% • 22%</p>
1	F	320	<div style="display: flex; align-items: center;"> <div style="width: 0%; height: 10px; background-color: red; margin-right: 2px;"></div> <div style="width: 65%; height: 10px; background-color: green; margin-right: 2px;"></div> <div style="width: 13%; height: 10px; background-color: yellow; margin-right: 2px;"></div> <div style="width: 21%; height: 10px; background-color: grey;"></div> </div> <p style="margin-left: 20px;">% 65% 13% • 21%</p>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	320	
1	H	320	

PRELIMINARY

VALIDATION

REPORT

2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 16055 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called general control nonderepressible 2 (GCN2).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	245	Total	C	N	O	S	0	0	0
			1865	1205	314	339	7			
1	B	249	Total	C	N	O	S	0	0	0
			1934	1248	330	349	7			
1	C	249	Total	C	N	O	S	0	0	0
			1928	1249	326	346	7			
1	D	254	Total	C	N	O	S	0	0	0
			1993	1289	336	361	7			
1	E	249	Total	C	N	O	S	0	0	0
			1952	1262	334	349	7			
1	F	254	Total	C	N	O	S	0	0	0
			1990	1285	339	358	8			
1	G	257	Total	C	N	O	S	0	0	0
			2005	1293	339	365	8			
1	H	259	Total	C	N	O	S	0	0	0
			2024	1306	342	367	9			

- Molecule 2 is a ligand with the chemical component id INH but its atom names do not match the existing wwPDB Chemical Component Dictionary definition for INH. Consequently no firm identification of ligand chemistry can be made. Once the structure is annotated then an identification and diagram will be given here.

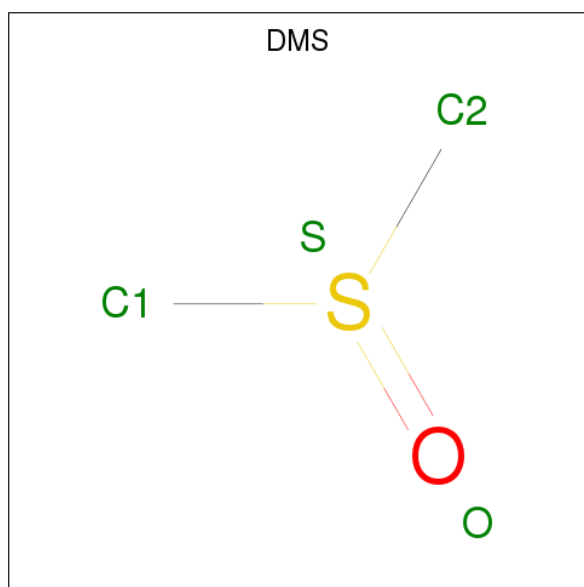
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		
2	I	1	Total	C	N	O	0	0
			34	27	6	1		

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C₂H₆OS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
3	K	1	Total	C	O	S	0	0
			4	2	1	1		
3	K	1	Total	C	O	S	0	0
			4	2	1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	W	84	Total	O	0
			84	84	

4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	73.80Å 77.80Å 101.69Å 89.86° 90.05° 68.58°	Depositor
Resolution (Å)	39.43 - 2.30 39.43 - 2.30	Depositor EDS
% Data completeness (in resolution range)	96.5 (39.43-2.30) 95.8 (39.43-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.38 (at 2.29Å)	Xtrriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.212 , 0.263 0.222 , 0.283	Depositor DCC
R_{free} test set	4728 reflections (5.24%)	wwPDB-VP
Wilson B-factor (Å ²)	38.5	Xtrriage
Anisotropy	0.321	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 47.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.448 for -h,-k,l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	16055	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.06 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.8623e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: DMS, INH

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.50	0/1909	0.70	0/2598
1	B	0.52	0/1979	0.68	0/2686
1	C	0.49	0/1976	0.69	0/2686
1	D	0.52	0/2041	0.69	0/2770
1	E	0.51	0/1998	0.70	0/2709
1	F	0.53	0/2035	0.73	0/2759
1	G	0.51	0/2049	0.69	0/2779
1	H	0.52	0/2071	0.70	0/2809
All	All	0.51	0/16058	0.70	0/21796

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1865	0	1732	17	0
1	B	1934	0	1821	19	0
1	C	1928	0	1816	16	0
1	D	1993	0	1897	18	0
1	E	1952	0	1873	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	1990	0	1916	24	0
1	G	2005	0	1920	22	0
1	H	2024	0	1938	24	0
2	I	272	0	44	20	0
3	K	8	0	12	1	0
4	W	84	0	0	2	0
All	All	16055	0	14969	171	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (171) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:636:GLU:HG2	1:F:868:GLY:HA2	1.59	0.83
1:D:609:ASN:HD22	1:D:612:ASP:H	1.25	0.83
1:G:866:ASP:HB2	1:G:870:ALA:HA	1.58	0.83
1:G:609:ASN:HD22	1:G:612:ASP:H	1.28	0.81
1:H:645:HIS:HD2	1:H:647:ASN:H	1.30	0.80
1:B:853:ASN:ND2	1:B:867:PHE:H	1.82	0.76
1:D:853:ASN:ND2	1:D:867:PHE:H	1.84	0.76
1:F:855:PHE:CE2	2:I:4:INH:C14	2.70	0.75
1:B:853:ASN:HD22	1:B:867:PHE:H	1.33	0.74
1:E:920:TYR:HB3	1:E:924:VAL:HG21	1.70	0.74
1:D:639:LEU:HD12	1:D:875:ALA:HB1	1.69	0.73
1:C:609:ASN:HD22	1:C:612:ASP:H	1.36	0.72
1:A:649:VAL:HG23	1:A:803:GLU:HG2	1.72	0.72
1:H:609:ASN:HD22	1:H:612:ASP:H	1.37	0.70
1:F:645:HIS:HD2	1:F:647:ASN:H	1.39	0.70
1:B:609:ASN:HD22	1:B:612:ASP:H	1.39	0.70
1:H:645:HIS:CD2	1:H:647:ASN:H	2.09	0.70
1:B:604:VAL:HG23	1:B:869:LEU:HD21	1.75	0.69
1:A:609:ASN:HD22	1:A:612:ASP:H	1.39	0.68
1:F:609:ASN:HD22	1:F:612:ASP:H	1.39	0.68
1:E:609:ASN:HD22	1:E:612:ASP:H	1.40	0.67
1:D:645:HIS:HD2	1:D:647:ASN:H	1.41	0.66
1:B:645:HIS:HD2	1:B:647:ASN:H	1.43	0.66
1:D:795:VAL:HG11	1:F:1005:GLN:HB3	1.77	0.66
1:D:853:ASN:HD22	1:D:867:PHE:H	1.42	0.66
1:H:960:PRO:HG2	1:H:979:SER:HA	1.79	0.65
1:E:649:VAL:HG23	1:E:803:GLU:HG2	1.79	0.65

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:645:HIS:CD2	1:B:647:ASN:H	2.15	0.64
1:D:609:ASN:ND2	1:D:612:ASP:H	1.95	0.64
1:F:645:HIS:CD2	1:F:647:ASN:H	2.16	0.64
1:D:645:HIS:CD2	1:D:647:ASN:H	2.16	0.63
1:F:855:PHE:CZ	2:I:4:INH:C13	2.82	0.63
1:E:609:ASN:HB3	1:E:612:ASP:OD1	1.99	0.63
1:G:859:ASP:HB3	1:G:861:HIS:HD2	1.64	0.63
1:G:867:PHE:HA	2:I:3:INH:H48	1.80	0.62
1:H:619:LYS:HE2	1:H:621:ILE:HD11	1.81	0.62
1:G:609:ASN:ND2	1:G:612:ASP:H	1.96	0.61
1:E:593:LEU:HD11	1:E:608:GLN:HB2	1.83	0.61
1:D:604:VAL:HG23	1:D:869:LEU:HD21	1.83	0.61
1:C:816:GLN:HB3	1:H:806:GLU:HG2	1.82	0.60
1:F:648:ILE:HG22	1:F:866:ASP:OD2	2.01	0.60
1:A:645:HIS:HD2	1:A:647:ASN:H	1.49	0.60
1:G:609:ASN:HB3	1:G:612:ASP:OD1	2.01	0.60
1:F:855:PHE:CE2	2:I:4:INH:C15	2.85	0.59
1:E:645:HIS:HB3	1:E:648:ILE:HG12	1.85	0.59
1:G:650:ARG:HG2	1:G:652:TYR:CE1	2.38	0.59
1:C:645:HIS:HD2	1:C:647:ASN:H	1.49	0.59
1:B:909:SER:HB2	1:B:924:VAL:HG13	1.85	0.58
1:E:609:ASN:ND2	1:E:612:ASP:H	2.02	0.58
1:F:960:PRO:HG2	1:F:979:SER:HA	1.86	0.57
1:C:931:ILE:HG12	1:C:982:LEU:HD21	1.86	0.57
1:G:645:HIS:HD2	1:G:647:ASN:H	1.53	0.56
1:C:629:GLN:O	1:C:633:ILE:HG12	2.05	0.56
1:D:650:ARG:HG2	1:D:652:TYR:CE1	2.41	0.56
1:D:909:SER:HB2	1:D:924:VAL:HG13	1.87	0.56
1:H:855:PHE:CE1	2:I:2:INH:C14	2.89	0.55
1:C:645:HIS:CD2	1:C:647:ASN:H	2.24	0.55
1:G:645:HIS:CD2	1:G:647:ASN:H	2.24	0.55
1:B:845:ILE:HD11	1:B:922:GLN:HG2	1.88	0.55
1:H:648:ILE:HG23	1:H:866:ASP:OD2	2.06	0.55
1:D:583:PHE:N	1:D:588:ILE:HD11	2.22	0.55
1:B:645:HIS:HB3	1:B:648:ILE:HG12	1.88	0.54
1:F:855:PHE:CZ	2:I:4:INH:C14	2.90	0.54
1:G:944:ALA:O	1:G:948:ILE:HG12	2.09	0.53
1:F:944:ALA:O	1:F:948:ILE:HG12	2.10	0.52
1:G:920:TYR:HB3	1:G:924:VAL:HG21	1.90	0.52
1:C:649:VAL:HG12	1:C:864:ILE:O	2.09	0.52
1:B:816:GLN:HG3	1:E:806:GLU:HG3	1.92	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:591:GLU:HG3	1:E:610:LYS:HE2	1.92	0.52
1:B:619:LYS:HE2	1:B:621:ILE:HD11	1.91	0.51
2:I:5:INH:C4	2:I:5:INH:N2	2.73	0.51
1:H:593:LEU:HD11	1:H:608:GLN:HB2	1.92	0.51
1:E:845:ILE:HD13	1:E:847:ARG:HG3	1.92	0.51
1:G:853:ASN:ND2	1:G:866:ASP:O	2.43	0.51
1:E:844:MET:HA	1:E:914:GLY:HA3	1.93	0.50
2:I:4:INH:C4	2:I:4:INH:N2	2.75	0.50
1:C:853:ASN:ND2	1:C:867:PHE:H	2.09	0.49
1:C:826:TRP:CZ2	1:C:974:GLN:HG3	2.47	0.49
1:B:824:ARG:HD2	1:E:807:LYS:HD3	1.93	0.49
1:E:650:ARG:HG2	1:E:652:TYR:CE1	2.47	0.49
2:I:3:INH:C4	2:I:3:INH:N2	2.75	0.49
2:I:8:INH:N2	2:I:8:INH:C4	2.76	0.49
1:D:920:TYR:C	1:D:922:GLN:H	2.16	0.49
1:H:855:PHE:CZ	2:I:2:INH:C13	2.96	0.49
1:H:650:ARG:HG2	1:H:652:TYR:CE1	2.48	0.49
1:H:903:GLY:HA2	1:H:906:LEU:HD12	1.95	0.48
1:H:855:PHE:CE1	2:I:2:INH:C15	2.96	0.48
1:A:828:LEU:HD22	1:A:862:VAL:HG23	1.96	0.48
1:C:609:ASN:ND2	1:C:612:ASP:H	2.10	0.48
1:A:584:SER:O	1:A:588:ILE:HG12	2.13	0.48
1:H:831:GLU:HB3	1:H:862:VAL:HB	1.94	0.48
1:A:645:HIS:CD2	1:A:647:ASN:H	2.29	0.48
1:E:980:TRP:CZ2	1:E:990:PRO:HB3	2.48	0.48
1:G:602:GLY:HA3	1:G:621:ILE:HD13	1.96	0.48
1:H:898:LEU:O	1:H:903:GLY:HA3	2.14	0.48
1:F:650:ARG:HD2	4:W:4:HOH:O	2.14	0.48
1:G:866:ASP:O	1:G:867:PHE:CB	2.61	0.47
1:H:609:ASN:ND2	1:H:612:ASP:H	2.08	0.47
1:E:645:HIS:CD2	1:E:647:ASN:H	2.31	0.47
1:E:853:ASN:ND2	1:E:867:PHE:H	2.12	0.47
1:A:813:THR:HG22	1:A:818:LEU:HB2	1.96	0.47
1:E:645:HIS:HD2	1:E:647:ASN:H	1.61	0.47
2:I:6:INH:C4	2:I:6:INH:N2	2.75	0.47
1:C:650:ARG:HG2	1:C:652:TYR:CE1	2.50	0.47
1:E:931:ILE:HG23	1:E:951:LEU:HD22	1.97	0.47
1:B:649:VAL:HG13	1:B:803:GLU:HG2	1.97	0.47
1:D:639:LEU:CD1	1:D:875:ALA:HB1	2.40	0.47
1:F:825:LEU:CD2	1:F:936:MET:HB3	2.45	0.47
1:D:650:ARG:HD2	4:W:8:HOH:O	2.15	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:943:THR:HG22	1:A:946:GLU:H	1.80	0.46
1:B:980:TRP:CD1	1:B:990:PRO:HD3	2.50	0.46
1:F:650:ARG:HG2	1:F:652:TYR:CE1	2.50	0.46
1:H:833:LEU:HD11	1:H:995:LEU:HG	1.97	0.46
2:I:2:INH:C4	2:I:2:INH:N2	2.78	0.46
1:A:612:ASP:OD1	1:A:614:CYS:HB2	2.15	0.46
1:F:845:ILE:HG12	1:F:871:THR:CG2	2.46	0.46
1:A:629:GLN:O	1:A:633:ILE:HG12	2.16	0.46
1:G:649:VAL:HG12	1:G:864:ILE:O	2.16	0.46
1:E:866:ASP:O	2:I:5:INH:C17	2.64	0.46
1:D:980:TRP:CD1	1:D:990:PRO:HD3	2.51	0.46
1:E:826:TRP:CZ2	1:E:974:GLN:HG3	2.51	0.46
1:G:950:VAL:HG13	1:G:963:PRO:HG3	1.98	0.46
1:C:944:ALA:O	1:C:948:ILE:HG12	2.16	0.46
1:H:836:LEU:HD11	1:H:925:ASP:HB3	1.98	0.45
1:A:654:ALA:HA	1:A:799:TYR:O	2.17	0.45
1:H:944:ALA:O	1:H:948:ILE:HG12	2.16	0.45
1:G:906:LEU:HD13	1:G:951:LEU:HD12	1.98	0.45
1:B:950:VAL:HG13	1:B:963:PRO:HG3	1.97	0.45
1:F:833:LEU:HD11	1:F:995:LEU:HG	1.98	0.45
1:E:648:ILE:HD12	1:E:839:ILE:HD11	1.98	0.45
1:F:828:LEU:HD22	1:F:856:LEU:HD21	1.98	0.44
1:H:855:PHE:CZ	2:I:2:INH:C14	3.00	0.44
1:B:650:ARG:HG2	1:B:652:TYR:CE1	2.52	0.44
1:A:948:ILE:O	1:A:952:ASN:HB2	2.18	0.44
1:F:902:VAL:HG12	1:F:948:ILE:HD12	1.99	0.44
2:I:1:INH:C4	2:I:1:INH:N2	2.79	0.44
2:I:7:INH:C4	2:I:7:INH:N2	2.79	0.44
1:H:852:VAL:HG22	3:K:1:DMS:H21	2.00	0.44
1:C:595:LEU:HD21	1:C:598:LYS:HB2	2.00	0.44
1:F:976:SER:O	1:F:980:TRP:HB2	2.18	0.44
1:A:650:ARG:HG2	1:A:652:TYR:CE1	2.53	0.43
1:B:609:ASN:ND2	1:B:612:ASP:H	2.13	0.43
1:D:985:ASP:HB2	1:G:876:PHE:O	2.18	0.43
1:E:829:PHE:HE2	1:E:1001:LEU:HD11	1.83	0.43
1:F:903:GLY:HA2	1:F:906:LEU:HD12	2.01	0.42
1:H:812:ASP:OD1	2:I:2:INH:N5	2.52	0.42
1:B:583:PHE:HB3	1:B:588:ILE:HD11	2.01	0.42
1:D:654:ALA:HA	1:D:799:TYR:O	2.18	0.42
1:E:639:LEU:HD21	1:E:915:SER:HB2	2.00	0.42
1:F:845:ILE:HG12	1:F:871:THR:HG21	2.02	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:980:TRP:CZ2	1:C:990:PRO:HB3	2.54	0.42
1:G:645:HIS:HB3	1:G:648:ILE:HG12	2.02	0.42
1:E:973:LYS:O	1:E:977:VAL:HG23	2.19	0.42
1:F:847:ARG:HD3	1:F:869:LEU:CB	2.50	0.42
1:C:654:ALA:HA	1:C:799:TYR:O	2.20	0.42
1:C:645:HIS:HB3	1:C:648:ILE:HG12	2.02	0.42
1:B:931:ILE:HG23	1:B:951:LEU:HD22	2.02	0.41
1:E:819:TYR:CD1	1:E:939:HIS:HA	2.55	0.41
1:H:833:LEU:HD21	1:H:995:LEU:HD23	2.02	0.41
1:A:985:ASP:HA	1:A:986:PRO:HD2	1.76	0.41
1:E:614:CYS:HB3	1:E:616:TYR:CE2	2.56	0.41
1:A:649:VAL:HG12	1:A:864:ILE:O	2.20	0.41
1:A:991:THR:OG1	1:A:994:GLU:HB2	2.21	0.41
1:E:859:ASP:HB3	1:E:861:HIS:HD2	1.85	0.41
1:G:846:HIS:O	1:G:847:ARG:HB2	2.20	0.40
1:G:867:PHE:HA	2:I:3:INH:C13	2.47	0.40
1:G:991:THR:OG1	1:G:994:GLU:HB2	2.21	0.40
1:H:825:LEU:HD23	1:H:936:MET:HB3	2.03	0.40
1:A:980:TRP:CZ2	1:A:990:PRO:HB3	2.57	0.40
1:E:856:LEU:HD23	1:E:856:LEU:HA	1.91	0.40
1:F:899:ALA:CB	1:F:948:ILE:HD11	2.51	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [\(i\)](#)

5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	237/320 (74%)	221 (93%)	14 (6%)	2 (1%)	21 25
1	B	241/320 (75%)	230 (95%)	11 (5%)	0	100 100
1	C	241/320 (75%)	227 (94%)	14 (6%)	0	100 100
1	D	246/320 (77%)	234 (95%)	10 (4%)	2 (1%)	21 25

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	239/320 (75%)	229 (96%)	8 (3%)	2 (1%)	21	25
1	F	244/320 (76%)	228 (93%)	13 (5%)	3 (1%)	14	15
1	G	245/320 (77%)	227 (93%)	17 (7%)	1 (0%)	36	45
1	H	247/320 (77%)	237 (96%)	10 (4%)	0	100	100
All	All	1940/2560 (76%)	1833 (94%)	97 (5%)	10 (0%)	31	38

All (10) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	848	ASP
1	D	848	ASP
1	E	917	LYS
1	F	902	VAL
1	F	970	GLU
1	A	867	PHE
1	D	873	HIS
1	E	919	ALA
1	G	942	VAL
1	F	869	LEU

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	184/284 (65%)	173 (94%)	11 (6%)	21	28
1	B	194/284 (68%)	181 (93%)	13 (7%)	18	23
1	C	194/284 (68%)	176 (91%)	18 (9%)	10	11
1	D	204/284 (72%)	190 (93%)	14 (7%)	17	22
1	E	200/284 (70%)	183 (92%)	17 (8%)	12	14
1	F	205/284 (72%)	188 (92%)	17 (8%)	12	15
1	G	206/284 (72%)	188 (91%)	18 (9%)	11	13
1	H	210/284 (74%)	197 (94%)	13 (6%)	20	27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1597/2272 (70%)	1476 (92%)	121 (8%)	14 18

All (121) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	585	ARG
1	A	601	PHE
1	A	642	ARG
1	A	650	ARG
1	A	653	ASN
1	A	867	PHE
1	A	916	THR
1	A	943	THR
1	A	952	ASN
1	A	959	SER
1	A	984	HIS
1	B	598	LYS
1	B	623	ILE
1	B	653	ASN
1	B	795	VAL
1	B	807	LYS
1	B	836	LEU
1	B	845	ILE
1	B	867	PHE
1	B	873	HIS
1	B	909	SER
1	B	926	LEU
1	B	943	THR
1	B	984	HIS
1	C	584	SER
1	C	598	LYS
1	C	642	ARG
1	C	653	ASN
1	C	795	VAL
1	C	806	GLU
1	C	813	THR
1	C	825	LEU
1	C	862	VAL
1	C	867	PHE
1	C	912	VAL
1	C	922	GLN
1	C	926	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	943	THR
1	C	952	ASN
1	C	984	HIS
1	C	993	THR
1	C	1001	LEU
1	D	598	LYS
1	D	649	VAL
1	D	653	ASN
1	D	795	VAL
1	D	836	LEU
1	D	845	ILE
1	D	852	VAL
1	D	867	PHE
1	D	909	SER
1	D	926	LEU
1	D	943	THR
1	D	984	HIS
1	D	999	GLU
1	D	1001	LEU
1	E	588	ILE
1	E	605	ILE
1	E	608	GLN
1	E	620	ARG
1	E	623	ILE
1	E	633	ILE
1	E	642	ARG
1	E	653	ASN
1	E	806	GLU
1	E	830	ARG
1	E	867	PHE
1	E	909	SER
1	E	921	ASN
1	E	947	ARG
1	E	953	GLN
1	E	983	ASN
1	E	998	SER
1	F	585	ARG
1	F	623	ILE
1	F	642	ARG
1	F	653	ASN
1	F	795	VAL
1	F	807	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	825	LEU
1	F	842	LYS
1	F	845	ILE
1	F	850	LYS
1	F	867	PHE
1	F	871	THR
1	F	909	SER
1	F	921	ASN
1	F	926	LEU
1	F	980	TRP
1	F	998	SER
1	G	585	ARG
1	G	598	LYS
1	G	620	ARG
1	G	636	GLU
1	G	642	ARG
1	G	813	THR
1	G	825	LEU
1	G	859	ASP
1	G	867	PHE
1	G	874	LEU
1	G	909	SER
1	G	926	LEU
1	G	942	VAL
1	G	947	ARG
1	G	965	ASP
1	G	976	SER
1	G	984	HIS
1	G	1007	GLU
1	H	585	ARG
1	H	642	ARG
1	H	648	ILE
1	H	653	ASN
1	H	807	LYS
1	H	813	THR
1	H	825	LEU
1	H	869	LEU
1	H	909	SER
1	H	921	ASN
1	H	928	SER
1	H	999	GLU
1	H	1003	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (32) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	609	ASN
1	A	645	HIS
1	A	853	ASN
1	A	984	HIS
1	B	609	ASN
1	B	624	ASN
1	B	645	HIS
1	B	853	ASN
1	C	609	ASN
1	C	645	HIS
1	C	853	ASN
1	D	609	ASN
1	D	645	HIS
1	D	853	ASN
1	E	609	ASN
1	E	645	HIS
1	E	853	ASN
1	E	861	HIS
1	E	913	GLN
1	E	921	ASN
1	F	609	ASN
1	F	645	HIS
1	F	853	ASN
1	G	609	ASN
1	G	645	HIS
1	G	853	ASN
1	G	861	HIS
1	H	609	ASN
1	H	645	HIS
1	H	853	ASN
1	H	921	ASN
1	H	974	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 8 could not be matched to an existing wwPDB Chemical Component Dictionary definition at this stage - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	DMS	K	1	-	3,3,3	0.38	0	3,3,3	0.38	0
3	DMS	K	2	-	3,3,3	0.35	0	3,3,3	0.78	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	DMS	K	1	-	-	0/0/0/0	0/0/0/0
3	DMS	K	2	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	K	1	DMS	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

6 Fit of model and data i

6.1 Protein, DNA and RNA chains i

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	A	245/320 (76%)	-0.01	6 (2%) 59 66	23, 50, 89, 102	0
1	B	249/320 (77%)	-0.08	8 (3%) 47 55	23, 45, 76, 97	0
1	C	249/320 (77%)	0.11	11 (4%) 34 41	24, 51, 96, 111	0
1	D	254/320 (79%)	-0.02	6 (2%) 59 66	23, 46, 80, 99	0
1	E	249/320 (77%)	0.07	12 (4%) 30 38	28, 51, 83, 102	0
1	F	254/320 (79%)	-0.09	3 (1%) 79 83	23, 46, 76, 100	0
1	G	257/320 (80%)	0.07	11 (4%) 35 42	29, 52, 84, 100	0
1	H	259/320 (80%)	-0.15	1 (0%) 92 95	23, 46, 74, 110	0
All	All	2016/2560 (78%)	-0.01	58 (2%) 51 59	23, 49, 84, 111	0

All (58) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	584	SER	7.4
1	B	966	PHE	6.3
1	G	982	LEU	5.0
1	G	584	SER	4.8
1	F	795	VAL	4.4
1	B	968	ASP	4.3
1	C	968	ASP	4.3
1	G	626	ALA	3.8
1	C	626	ALA	3.7
1	E	625	PRO	3.7
1	G	963	PRO	3.6
1	D	968	ASP	3.5
1	G	984	HIS	3.4
1	C	960	PRO	3.3
1	F	968	ASP	3.3
1	A	626	ALA	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	970	GLU	3.2
1	E	984	HIS	3.2
1	F	969	GLY	3.1
1	E	626	ALA	3.1
1	C	942	VAL	3.1
1	D	966	PHE	3.0
1	C	1000	LEU	3.0
1	D	965	ASP	3.0
1	E	601	PHE	2.9
1	E	982	LEU	2.9
1	A	942	VAL	2.9
1	B	942	VAL	2.9
1	C	949	PHE	2.9
1	E	915	SER	2.8
1	E	630	PHE	2.8
1	G	601	PHE	2.8
1	G	949	PHE	2.8
1	B	969	GLY	2.7
1	E	992	ALA	2.6
1	H	600	ALA	2.5
1	A	845	ILE	2.5
1	G	969	GLY	2.5
1	G	942	VAL	2.5
1	D	909	SER	2.5
1	B	970	GLU	2.4
1	D	948	ILE	2.4
1	E	948	ILE	2.4
1	C	909	SER	2.3
1	C	951	LEU	2.3
1	A	600	ALA	2.3
1	B	847	ARG	2.3
1	E	797	TYR	2.2
1	C	965	ASP	2.2
1	C	845	ILE	2.2
1	E	588	ILE	2.2
1	D	949	PHE	2.1
1	B	941	MET	2.1
1	A	958	THR	2.1
1	G	633	ILE	2.1
1	G	659	HIS	2.0
1	A	968	ASP	2.0
1	B	626	ALA	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	INH	I	5	34/?	0.91	0.17	33,46,65,67	0
2	INH	I	4	34/?	0.92	0.13	29,34,40,42	0
2	INH	I	2	34/?	0.93	0.13	28,33,47,49	0
2	INH	I	3	34/?	0.93	0.15	32,45,70,71	0
3	DMS	K	1	4/?	0.93	0.21	76,79,80,81	0
3	DMS	K	2	4/?	0.94	0.20	78,81,81,83	0
2	INH	I	8	34/?	0.96	0.12	21,28,40,43	0
2	INH	I	7	34/?	0.96	0.11	25,36,42,43	0
2	INH	I	1	34/?	0.97	0.12	20,28,44,47	0
2	INH	I	6	34/?	0.97	0.12	24,32,43,43	0

6.5 Other polymers [i](#)

There are no such residues in this entry.



Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Feb 20, 2019 – 12:59 pm GMT

Deposition ID : D_1292100769
PDB ID : *(not yet assigned)*

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtrriage (Phenix)	:	1.13
EDS	:	2.1
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.1

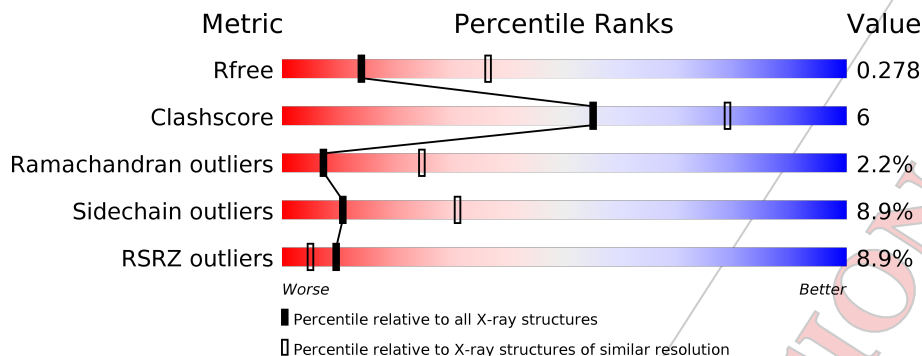
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	111664	2792 (2.80-2.80)
Clashscore	122126	3209 (2.80-2.80)
Ramachandran outliers	120053	3158 (2.80-2.80)
Sidechain outliers	120020	3160 (2.80-2.80)
RSRZ outliers	108989	2726 (2.80-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	320	 6% 66% 17% • 16%
1	B	320	 9% 67% 18% • 14%
1	C	320	 8% 63% 19% • 17%
1	D	320	 7% 67% 14% • 19%

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 8692 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called GCN2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	269	Total 2124	C 1365	N 363	O 388	S 8	0	0	0
1	B	275	Total 2188	C 1405	N 376	O 398	S 9	0	0	0
1	C	267	Total 2133	C 1364	N 368	O 393	S 8	0	0	0
1	D	260	Total 2086	C 1341	N 361	O 376	S 8	0	0	0

- Molecule 2 is a ligand with the chemical component id INH but its atom names do not match the existing wwPDB Chemical Component Dictionary definition for INH. Consequently no firm identification of ligand chemistry can be made. Once the structure is annotated then an identification and diagram will be given here.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
2	I	1	Total 29	C 21	F 1	N 6	O 1	0	0
2	I	1	Total 29	C 21	F 1	N 6	O 1	0	0
2	I	1	Total 29	C 21	F 1	N 6	O 1	0	0
2	I	1	Total 29	C 21	F 1	N 6	O 1	0	0

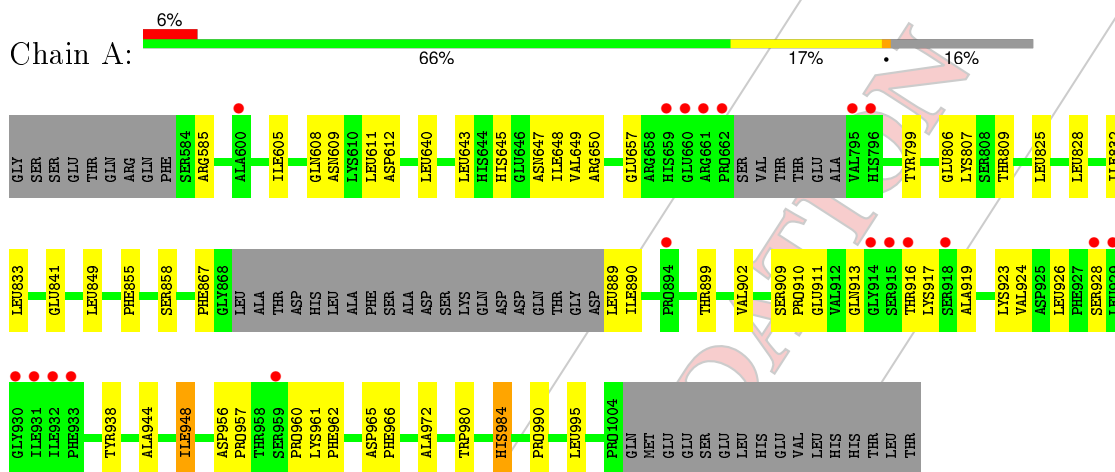
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	W	45	Total 45	O 45	0	0

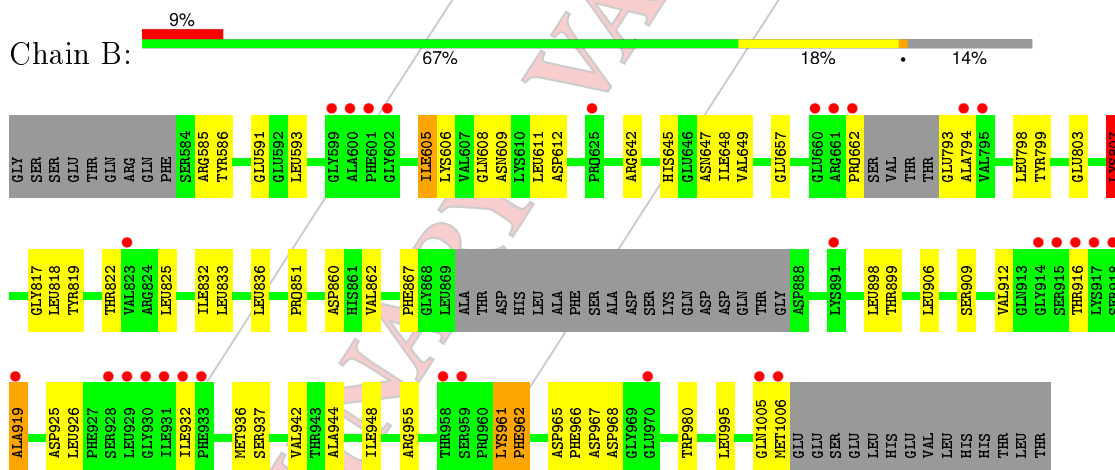
3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

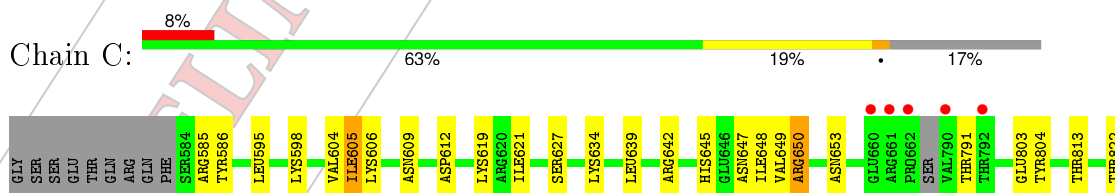
- Molecule 1: GCN2



- Molecule 1: GCN2



- Molecule 1: GCN2



4 Data and refinement statistics i

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	155.42Å 162.60Å 123.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.64 – 2.80 24.64 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (24.64-2.80) 99.7 (24.64-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.36 (at 2.80Å)	Xtrriage
Refinement program	BUSTER 2.11.7	Depositor
R, R_{free}	0.211 , 0.258 0.224 , 0.278	Depositor DCC
R_{free} test set	1946 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	92.5	Xtrriage
Anisotropy	0.204	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 70.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.062 for -k,-h,-l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8692	wwPDB-VP
Average B, all atoms (Å ²)	91.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.35% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: INH

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.51	0/2175	0.76	0/2950
1	B	0.50	0/2240	0.74	0/3034
1	C	0.50	0/2183	0.75	0/2960
1	D	0.54	0/2136	0.76	0/2893
All	All	0.51	0/8734	0.75	0/11837

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2124	0	2052	23	0
1	B	2188	0	2147	24	0
1	C	2133	0	2073	32	0
1	D	2086	0	2028	17	0
2	I	116	0	4	4	0
3	W	45	0	0	1	0
All	All	8692	0	8304	95	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 6.

All (95) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:650:ARG:HG2	1:C:650:ARG:HH11	1.32	0.92
1:A:645:HIS:HB3	1:A:648:ILE:HG12	1.55	0.89
1:C:645:HIS:HB3	1:C:648:ILE:HG12	1.64	0.80
1:B:645:HIS:HB3	1:B:648:ILE:HG12	1.62	0.79
1:B:645:HIS:HD2	1:B:647:ASN:H	1.31	0.76
1:A:645:HIS:HD2	1:A:647:ASN:H	1.33	0.76
1:A:944:ALA:O	1:A:948:ILE:HG12	1.90	0.71
1:A:645:HIS:CD2	1:A:647:ASN:H	2.11	0.69
1:C:609:ASN:HD22	1:C:612:ASP:H	1.41	0.68
1:C:650:ARG:HG2	1:C:650:ARG:NH1	2.09	0.68
1:D:609:ASN:HD22	1:D:612:ASP:H	1.43	0.65
1:A:650:ARG:NH2	1:D:650:ARG:HB3	2.12	0.64
1:B:944:ALA:O	1:B:948:ILE:HG12	1.98	0.63
1:C:595:LEU:HD21	1:C:598:LYS:HB2	1.82	0.62
1:C:645:HIS:CD2	1:C:647:ASN:H	2.18	0.62
1:C:645:HIS:HD2	1:C:647:ASN:H	1.48	0.61
1:D:645:HIS:HB3	1:D:648:ILE:HG12	1.83	0.61
1:A:609:ASN:HD22	1:A:612:ASP:H	1.50	0.60
1:C:980:TRP:HZ3	1:C:988:LYS:O	1.86	0.58
1:D:821:ASP:O	1:D:825:LEU:HB2	2.03	0.58
1:A:923:LYS:HA	1:A:926:LEU:HD12	1.86	0.58
1:B:609:ASN:HD22	1:B:612:ASP:H	1.50	0.58
1:B:916:THR:HG23	1:B:919:ALA:HB3	1.87	0.56
1:B:609:ASN:HB3	1:B:612:ASP:OD1	2.06	0.55
1:A:890:ILE:HD13	1:A:948:ILE:HG21	1.88	0.55
1:D:593:LEU:HD11	1:D:608:GLN:HB2	1.88	0.55
1:C:828:LEU:O	1:C:832:ILE:HG13	2.07	0.55
1:D:607:VAL:O	1:D:615:CYS:HA	2.07	0.54
1:D:645:HIS:HD2	1:D:647:ASN:H	1.54	0.54
1:B:818:LEU:HD23	1:B:936:MET:HG2	1.89	0.54
1:C:595:LEU:HD12	1:C:605:ILE:HG23	1.89	0.54
1:B:657:GLU:HB2	1:B:799:TYR:HE1	1.73	0.53
1:C:848:ASN:HB2	1:C:869:LEU:HD11	1.90	0.53
1:B:662:PRO:HG3	1:B:793:GLU:HA	1.90	0.53
1:C:851:PRO:HD3	1:C:932:ILE:HG12	1.91	0.52
1:B:825:LEU:HD11	1:B:937:SER:HA	1.90	0.52
1:B:585:ARG:HD2	1:C:585:ARG:HD2	1.91	0.52
1:A:609:ASN:HB3	1:A:612:ASP:OD1	2.10	0.52
1:A:640:LEU:HA	1:A:643:LEU:HD12	1.92	0.52
1:A:833:LEU:HD11	1:A:995:LEU:HD23	1.92	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:586:TYR:OH	1:C:605:ILE:HG13	2.11	0.51
1:D:980:TRP:CE3	1:D:990:PRO:HD3	2.45	0.50
1:C:977:VAL:HG13	1:C:995:LEU:HD11	1.94	0.50
1:A:910:PRO:HA	1:A:913:GLN:HE21	1.77	0.50
1:A:980:TRP:CD1	1:A:990:PRO:HD3	2.46	0.50
1:B:645:HIS:CD2	1:B:647:ASN:H	2.20	0.50
1:C:909:SER:HB2	1:C:912:VAL:HG23	1.94	0.49
1:D:981:LEU:HD21	1:D:995:LEU:HD22	1.94	0.48
1:B:833:LEU:HD11	1:B:995:LEU:HD23	1.96	0.48
1:C:980:TRP:CZ3	1:C:988:LYS:O	2.67	0.47
1:C:981:LEU:HD21	1:C:995:LEU:HD22	1.96	0.47
1:B:961:LYS:O	1:B:962:PHE:HB2	2.15	0.47
1:D:833:LEU:HD11	1:D:995:LEU:HD23	1.97	0.47
1:C:619:LYS:HG2	1:C:621:ILE:HD11	1.97	0.47
1:C:828:LEU:HD22	1:C:862:VAL:HG23	1.97	0.47
1:A:828:LEU:O	1:A:832:ILE:HG13	2.16	0.45
1:A:611:LEU:HD12	1:D:634:LYS:HE2	1.97	0.45
2:I:1:INH:N5	2:I:1:INH:N2	2.64	0.45
1:C:911:GLU:HB3	1:C:918:SER:HB2	1.97	0.45
1:D:950:VAL:HG13	1:D:963:PRO:HG3	1.98	0.45
1:B:593:LEU:HD12	1:B:606:LYS:HE2	1.99	0.45
1:B:611:LEU:HD12	1:C:634:LYS:HE2	1.99	0.45
1:A:957:PRO:HB3	1:A:984:HIS:CD2	2.52	0.44
1:C:973:LYS:O	1:C:977:VAL:HG23	2.16	0.44
1:C:639:LEU:HB2	1:C:867:PHE:HE2	1.81	0.44
2:I:4:INH:N5	2:I:4:INH:N2	2.64	0.44
1:C:649:VAL:HG13	1:C:803:GLU:HG2	2.00	0.44
1:D:645:HIS:CD2	1:D:647:ASN:H	2.34	0.43
1:C:968:ASP:O	1:C:972:ALA:HB2	2.18	0.43
1:B:836:LEU:HD11	1:B:925:ASP:HB3	2.01	0.43
1:C:645:HIS:HB3	1:C:648:ILE:CG1	2.42	0.43
1:C:991:THR:H	1:C:994:GLU:HB3	1.83	0.43
2:I:2:INH:N5	2:I:2:INH:N2	2.66	0.43
1:A:960:PRO:HB2	3:W:27:HOH:O	2.18	0.43
1:A:909:SER:HB2	1:A:924:VAL:HG13	2.02	0.42
1:B:912:VAL:HA	1:B:919:ALA:HB1	2.02	0.42
1:B:906:LEU:O	1:B:955:ARG:HD2	2.18	0.42
1:B:898:LEU:HD13	1:B:948:ILE:HB	2.01	0.42
1:C:604:VAL:HG22	1:C:619:LYS:HG3	2.01	0.42
1:D:833:LEU:HD21	1:D:995:LEU:HD23	2.01	0.42
1:B:851:PRO:HD3	1:B:932:ILE:HG12	2.02	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:609:ASN:ND2	1:C:612:ASP:H	2.11	0.41
1:D:633:ILE:O	1:D:637:VAL:HG23	2.21	0.41
1:B:832:ILE:HG12	1:B:862:VAL:HG11	2.03	0.41
1:A:657:GLU:HB2	1:A:799:TYR:HE1	1.85	0.41
1:D:985:ASP:HB3	1:D:988:LYS:HG2	2.02	0.41
1:C:934:PHE:CG	1:C:954:LEU:HD21	2.56	0.41
1:A:809:THR:HG22	1:A:855:PHE:CD1	2.55	0.41
1:A:938:TYR:HD1	1:A:966:PHE:HD1	1.67	0.40
1:B:807:LYS:HE2	1:B:860:ASP:HB2	2.03	0.40
1:C:606:LYS:HD2	1:C:804:TYR:CZ	2.56	0.40
1:A:585:ARG:HD2	1:D:585:ARG:HD2	2.04	0.40
1:A:902:VAL:HG23	1:A:948:ILE:HD12	2.02	0.40
2:I:3:INH:N2	2:I:3:INH:N5	2.65	0.40
1:B:586:TYR:OH	1:B:605:ILE:HG13	2.22	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	263/320 (82%)	240 (91%)	19 (7%)	4 (2%)	11	36
1	B	269/320 (84%)	237 (88%)	24 (9%)	8 (3%)	5	16
1	C	261/320 (82%)	229 (88%)	27 (10%)	5 (2%)	9	28
1	D	252/320 (79%)	229 (91%)	17 (7%)	6 (2%)	6	22
All	All	1045/1280 (82%)	935 (90%)	87 (8%)	23 (2%)	7	25

All (23) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	919	ALA

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	961	LYS
1	A	972	ALA
1	C	870	ALA
1	D	869	LEU
1	B	807	LYS
1	B	819	TYR
1	C	872	ASP
1	D	660	GLU
1	A	962	PHE
1	B	794	ALA
1	B	961	LYS
1	B	919	ALA
1	C	627	SER
1	C	962	PHE
1	D	867	PHE
1	D	920	TYR
1	D	954	LEU
1	B	817	GLY
1	B	967	ASP
1	C	961	LYS
1	B	962	PHE
1	D	962	PHE

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	224 / 286 (78%)	204 (91%)	20 (9%)	11 31
1	B	235 / 286 (82%)	215 (92%)	20 (8%)	12 33
1	C	228 / 286 (80%)	206 (90%)	22 (10%)	9 27
1	D	221 / 286 (77%)	202 (91%)	19 (9%)	11 33
All	All	908 / 1144 (79%)	827 (91%)	81 (9%)	11 31

All (81) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	605	ILE
1	A	608	GLN
1	A	649	VAL
1	A	806	GLU
1	A	807	LYS
1	A	825	LEU
1	A	841	GLU
1	A	849	LEU
1	A	858	SER
1	A	867	PHE
1	A	889	LEU
1	A	899	THR
1	A	911	GLU
1	A	916	THR
1	A	917	LYS
1	A	928	SER
1	A	948	ILE
1	A	956	ASP
1	A	965	ASP
1	A	984	HIS
1	B	591	GLU
1	B	605	ILE
1	B	608	GLN
1	B	642	ARG
1	B	649	VAL
1	B	798	LEU
1	B	803	GLU
1	B	807	LYS
1	B	822	THR
1	B	867	PHE
1	B	899	THR
1	B	909	SER
1	B	926	LEU
1	B	942	VAL
1	B	965	ASP
1	B	966	PHE
1	B	968	ASP
1	B	980	TRP
1	B	1005	GLN
1	B	1006	MET
1	C	605	ILE
1	C	642	ARG
1	C	650	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	653	ASN
1	C	791	THR
1	C	813	THR
1	C	822	THR
1	C	842	LYS
1	C	869	LEU
1	C	873	HIS
1	C	911	GLU
1	C	916	THR
1	C	945	SER
1	C	953	GLN
1	C	964	GLU
1	C	965	ASP
1	C	966	PHE
1	C	974	GLN
1	C	980	TRP
1	C	997	LYS
1	C	998	SER
1	C	1006	MET
1	D	595	LEU
1	D	800	ILE
1	D	807	LYS
1	D	813	THR
1	D	825	LEU
1	D	841	GLU
1	D	845	ILE
1	D	848	ASN
1	D	849	LEU
1	D	869	LEU
1	D	873	HIS
1	D	921	ASN
1	D	942	VAL
1	D	947	ARG
1	D	958	THR
1	D	970	GLU
1	D	980	TRP
1	D	1005	GLN
1	D	1006	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	609	ASN
1	A	645	HIS
1	A	913	GLN
1	A	984	HIS
1	B	609	ASN
1	B	645	HIS
1	C	609	ASN
1	C	645	HIS
1	D	609	ASN
1	D	645	HIS
1	D	659	HIS
1	D	848	ASN
1	D	921	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 could not be matched to an existing wwPDB Chemical Component Dictionary definition at this stage - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	269/320 (84%)	0.19	19 (7%) 16 9	57, 82, 115, 145	0
1	B	275/320 (85%)	0.35	29 (10%) 6 3	62, 88, 125, 150	0
1	C	267/320 (83%)	0.45	25 (9%) 8 4	65, 89, 132, 175	0
1	D	260/320 (81%)	0.40	22 (8%) 11 5	62, 92, 139, 167	0
All	All	1071/1280 (83%)	0.35	95 (8%) 9 5	57, 88, 130, 175	0

All (95) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	916	THR	9.1
1	A	662	PRO	7.7
1	C	1006	MET	7.4
1	C	1005	GLN	7.1
1	B	1006	MET	6.7
1	D	1005	GLN	6.1
1	D	1006	MET	5.9
1	B	1005	GLN	5.1
1	D	958	THR	4.7
1	D	930	GLY	4.5
1	B	917	LYS	4.5
1	D	919	ALA	4.4
1	B	600	ALA	4.2
1	A	659	HIS	4.1
1	D	1004	PRO	4.0
1	D	959	SER	4.0
1	C	932	ILE	4.0
1	A	795	VAL	4.0
1	C	919	ALA	4.0
1	D	932	ILE	3.9
1	A	600	ALA	3.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	660	GLU	3.7
1	A	929	LEU	3.7
1	B	959	SER	3.7
1	B	929	LEU	3.7
1	B	919	ALA	3.7
1	A	915	SER	3.6
1	B	794	ALA	3.5
1	B	662	PRO	3.5
1	A	918	SER	3.4
1	B	958	THR	3.4
1	C	903	GLY	3.4
1	B	970	GLU	3.3
1	D	929	LEU	3.3
1	B	823	VAL	3.2
1	A	661	ARG	3.1
1	C	968	ASP	3.1
1	B	661	ARG	3.0
1	B	914	GLY	2.9
1	C	958	THR	2.9
1	C	928	SER	2.8
1	D	660	GLU	2.8
1	B	932	ILE	2.8
1	B	601	PHE	2.8
1	A	928	SER	2.8
1	A	959	SER	2.8
1	B	931	ILE	2.8
1	B	625	PRO	2.7
1	B	891	LYS	2.7
1	C	929	LEU	2.7
1	C	790	VAL	2.6
1	D	794	ALA	2.6
1	D	969	GLY	2.6
1	A	932	ILE	2.6
1	A	930	GLY	2.5
1	C	913	GLN	2.5
1	D	918	SER	2.5
1	B	915	SER	2.5
1	B	660	GLU	2.5
1	B	930	GLY	2.5
1	A	931	ILE	2.4
1	A	914	GLY	2.4
1	D	913	GLN	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	599	GLY	2.4
1	B	928	SER	2.4
1	D	662	PRO	2.3
1	D	968	ASP	2.3
1	C	920	TYR	2.3
1	C	957	PRO	2.3
1	B	918	SER	2.3
1	D	928	SER	2.3
1	C	997	LYS	2.3
1	D	964	GLU	2.3
1	D	832	ILE	2.3
1	C	792	THR	2.3
1	C	930	GLY	2.3
1	B	916	THR	2.2
1	D	659	HIS	2.2
1	C	915	SER	2.2
1	D	933	PHE	2.2
1	C	661	ARG	2.2
1	C	918	SER	2.1
1	A	933	PHE	2.1
1	C	660	GLU	2.1
1	D	931	ILE	2.1
1	A	894	PRO	2.1
1	B	602	GLY	2.1
1	C	662	PRO	2.1
1	A	916	THR	2.1
1	C	904	THR	2.1
1	A	796	HIS	2.0
1	B	795	VAL	2.0
1	C	931	ILE	2.0
1	B	933	PHE	2.0
1	C	933	PHE	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	INH	I	1	29/?	0.87	0.22	66,80,109,109	0
2	INH	I	4	29/?	0.87	0.22	66,79,115,117	0
2	INH	I	2	29/?	0.92	0.19	58,69,102,102	0
2	INH	I	3	29/?	0.93	0.19	72,83,117,119	0

6.5 Other polymers [i](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION