Supporting information

O-GlcNAcase Fragment Discovery with Fluorescence Polarimetry

Vladimir S. Borodkin[1,*], Karim Rafie[1], Nithya Selvan[1,4], Tonia Aristotelous[3], Iva Navratilova[3], Andrew T. Ferenbach[1] and Daan M. F. van Aalten[1, 2,*]

[1]Division of Gene Regulation and Expression, [2]Division of Molecular Microbiology and [3]Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, United Kingdom. [4]Present address: Complex Carbohydrate Research Center, University of Georgia, Athens, United States.

List of Contents

Synthetic procedures and spectral data for all new compounds

The details of the instrumentation, general synthetic methods, and analytical techniques are as reported previously [1].

Compound 6a: A solution of diisopropyl azodicarboxylate (DIAD; 0.104 mL, 0.3 mmol) in THF (1 mL) was added dropwise to a hot (65 °C) solution of **7** (0.352 g, 0.41 mmol) and triphenylphosphine (PPh_3; 0.128 g, 0.49 mmol) in THF (5 mL). The reaction was kept at the specified temperature for 1 h. The reaction mixture was cooled down and concentrated. The residue was purified by flash column chromatography in PE-EE 10-30% gradient to give 0.316 g (0.375 mmol, 91%) of the target product as foam.

The reaction performed with 3.29 g (3.82 mmol) of **7** in THF (40 mL) gave 2.88 g (3.42 mmol, 90%) of the target compound.

Spectral data were in agreement with the reported previously [1].

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.28-7.23 (m, 2H), 6.78-6.74(m, 2H), 4.94 (dd, *J* = 10.7, 6.5 Hz, 1H), 4.69 and 4.59 (AB spectrum, *J* = 11.4 Hz, 2H), 4.61 (d, *J* = 2.8 Hz, 1H), 4.26 (dd, *J* = 10.7, 4.8 Hz, 1H), 4.03 (ddd, *J* = 6.7, 4.8, 2.5 Hz, 1H), 3.87 (dd, *J* = 10.7, 2.5 Hz, 1H), 3.80 (dd, *J* = 10.6, 2.8 Hz, 1H), 3.73 (3H, s), 3.26 (3H, s), 3.15 (3H, s), 1.31 (3H, s), 1.26 (3H, s), 0.88-0.77 (21H, m).

^{13}C NMR (126 MHz, $CDCl_3$) δ 158.9, 150, 130.3, 129.6, 113.4, 99.8, 99.6, 97.4, 80.1, 71.4, 70.1, 68.9, 63.6, 62.9, 62.1, 55.3, 48.2, 48.1, 17.8, 17.7, 17.6, 11.8.

$[\alpha]_D$ −28.5° c 1.0, $CHCl_3$.

HRMS-TOF (+): m/z = 843.1421, expected 843.1399 $[M+H]^+$

Compound 8: Aqueous trifluoroacetic acid 9:1 v/v (1 mL) was added to a solution of **6** (0.426 g, 0.59 mmol) in DCM (10 mL) at RT. The reaction was kept for 90 min. The reaction was diluted with toluene (10 mL) and concentrated. The residue was dissolved in a mixture of $CHCl_3$ and toluene and

concentrated. The residue was purified by flash column chromatography in [PE-DCM 4:1]-EA 15→40% to give 0.31 g (0.52 mmol, 88 %) of the target product as white-off foam.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.37 (s, 1H), 4.94 (d, *J* = 3.6 Hz, 1H), 4.24 (dd, *J* = 10.6, 9.5 Hz, 1H), 4.17 (dd, *J* = 10.8, 1.8 Hz, 1H), 3.98 (ddd, *J* = 9.4, 7.0, 1.9 Hz, 1H), 3.89 (dd, *J* = 10.9, 7.0 Hz, 1H), 3.88 (dd, *J* = 10.6, 3.6 Hz, 1H), 3.21 (s, 3H), 3.18 (s, 3H), 1.31 (s, 3H), 1.26 (s, 3H), 1.09 – 1.02 (m, 3H), 1.00 (dd, *J* = 6.7, 3.8 Hz, 18H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 146.09, 124.29, 99.92, 99.48, 81.79, 68.21, 64.01, 63.03, 61.52, 60.69, 48.15, 17.93, 17.58, 11.87.

$[\alpha]_D$ = 56.4° c 1.0 $CHCl_3$

HRMS-TOF (+): m/z = 597.1849, expected 597.1857 [M+H]+

Compound 9: Diphenylphosphoryl azide (DPPA; 0.338 mL, 1.57 mmol) and diazobicycloundecene (DBU; 0.235 mL, 1.57 mmol) were added in succession to a solution of **8** (0.31 g, 0.52 mmol) in toluene (5 mL) at RT. The reaction was kept for 30 min at RT. The reaction was placed in a preheated (90 °C) oil bath and kept for 40 min. The reaction was cooled, diluted with Me_2CO and concentrated. The brownish residue was dissolved in Me_2CO, absorbed on silica and purified by flash column chromatography in PE-EA 5→30% to give 0.304 g (0.49 mmol, 94 %) of the target product as yellow syrup.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.21 (s, 1H), 4.62 (d, *J* = 8.6 Hz, 1H), 4.14 (d, *J* = 10.0 Hz, 1H), 4.00 – 3.81 (m, 4H), 3.27 (s, 3H), 3.19 (s, 3H), 1.31 (s, 3H), 1.26 (s, 3H), 1.10 – 0.83 (m, 18H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 150.41, 143.22, 123.56, 99.60, 82.68, 70.74, 64.19, 62.10, 59.68, 57.25, 48.40, 17.95, 17.45, 11.91.

$[\alpha]_D$ = 66.2° c 1.0 $CHCl_3$

HRMS-TOF (+): m/z = 622.1931, expected 622.1922 $[M+H]^+$

Compound 5: Triphenylphosphine (PPh_3; 0.141 g, 0.54 mmol) was added to a solution of **9** (0.303 g, 0.49 mmol) in THF-water 20:1 (15 mL) at RT. The reaction was kept for 30 min at RT. The reaction was placed into preheated (65 °C) oil bath and kept for 3 h. The reaction was cooled down, N,N diisopropylethylamine (DIPEA; 0.348 mL, 2 mmol) and acetic anhydride (Ac_2O; 0.095 mL, 1 mmol) were added in succession and the reaction was stirred overnight at RT. The reaction was quenched with MeOH, stirred for 20 min and concentrated. The residue was partitioned between EA and 1M HCl and the layers were separated. The organic layer was washed successively with water, and a mixture of brine and concentrated aqueous $NaHCO_3$ solution. The aqueous layers were back extracted with EA. The combined organic layer was dried and concentrated to give purplish oil. The residue was purified by flash column chromatography in [PE-DCM 4:1]-Me_2CO 10→40% to give 0.385 g of a mixture of the target product and Ph_3PO as white-off foam. The residue was further purified by RP-C18 flash column chromatography in H_2O-MeOH 30→95% gradient to give 0.243 g (0.38 mmol, 78% over two steps) of the target product as white-off foam.

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.32 (d, *J* = 6.6 Hz, 1H), 7.32 (s, 1H), 4.61 – 4.48 (m, 2H), 4.30 (d, *J* = 11.0 Hz, 1H), 4.20 – 4.14 (m, 1H), 4.00 (dd, *J* = 11.1, 5.8 Hz, 1H), 3.82 (t, *J* = 9.5 Hz, 1H), 3.26 (s, 3H), 3.25 (s, 3H), 2.03 (s, 3H), 1.34 (s, 3H), 1.32 (s, 3H), 1.20 – 1.04 (m, 21H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.57, 146.14, 123.15, 99.42, 99.27, 81.19, 67.97, 65.46, 62.45, 59.54, 49.07, 48.14, 47.82, 23.29, 17.98, 17.93, 17.69, 17.42, 11.94.

$[\alpha]_D$ = 87.3° c 1.0 $CHCl_3$

HRMS-TOF (+): m/z = 622.2117, expected 638.2122 $[M+H]^+$

Compound 4: A solution of **5** (0.19 g, 0.3 mmol), *N*-Boc-propynyl amine (0.135 g, 0.9 mmol) in DMF (4.5 mL) was degassed by freezing, evacuating and filling with argon three times. Triethylamine (Et_3N; 0.21 mL, 1.5 mmol), cuprous iodide (CuI; 0.006 g, 0.03 mmol) and $Pd(PPh_3)_4$ (0.0035 g, 0.03 mmol) were then added. The reaction was placed in a preheated oil bath (80 °C). After approximately one

minute the heterogeneous yellowish coloured mixture turned into a colourless solution, which then started to develop a progressively deepening yellowish hue. The reaction was kept for 16 h at the specified temperature (the final colour was brown), cooled down and concentrated in vacuum. The brown residue was co-evaporated with toluene and purified by flash column chromatography in [PE-DCM 3:2]-Me_2CO 15→50% to give 0.194 g of the contaminated target product as amber amorphous compound. This material was further purified by RP C18 flash column chromatography to give 0.108 g (0.16 mmol, 53%) of the target compound as yellowish foam.

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.09 (s, 1H), 5.05 (t, *J* = 8.9 Hz, 1H), 4.21 (t, *J* = 10.1 Hz, 1H), 4.29-4.19 (broad dd, 1H), 4.13 (dd, *J* = 11.0, 1.8 Hz, 1H), 4.09 – 4.03 (m, 1H), 4.02 – 3.89 (m, 3H), 2.01 (s, 3H), 1.39 (s, 9H), 1.28 (s, 3H), 1.26 (s, 3H), 1.04 – 0.90 (m, 21H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 170.52, 155.49, 143.49, 122.76, 122.11, 99.50, 86.50, 79.57, 71.01, 64.56, 61.56, 59.17, 48.11, 47.89, 47.64, 30.99, 28.44, 23.52, 17.96, 17.87, 17.71, 17.43, 11.94.

HRMS-TOF (+): m/z = 665.3937 expected 665.3946 $[M+H]^+$

Compound 10: A solution of **4** (0.083 g, 0.12 mmol) in MeOH (2.5 mL) was stirred under slight overpressure (20 psi) of hydrogen gas in the presence of Pd-10%/C catalyst (0.02 g) at RT for 2 h. The catalyst was filtered off over a pad of Celite with the aid of MeOH (15 mL) and the filtrate was concentrated to give 0.078 g (0.117 mmol, 98%) of the target product as hard foam. The product was found to be more than 95% pure by NMR and was used in the next step as such.

^{1}H NMR (400 MHz, $CDCl_3$) δ 6.95 (s, 1H), 4.82 (bs, 1H), 4.48 (m, 2H), 4.27 (d, *J* = 10.9 Hz, 1H), 4.08 (dd, *J* = 9.8, 5.6 Hz, 1H), 3.99 (dd, *J* = 10.9, 5.6 Hz, 1H), 3.84 (t, *J* = 9.8 Hz, 1H), 3.24 (s, 3H), 3.23 (s, 3H), 3.15 (d, *J* = 6.6 Hz, 2H), 2.54 (t, *J* = 7.5 Hz, 2H), 2.03 (s, 3H),1.89 -1.73 (m, 2H), 1.44 (s, 9H), 1.31 (s, 6H), 1.04 – 0.90 (m, 21H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 170.30, 155.94, 143.40, 141.50, 113.65, 99.19, 78.89, 68.19, 65.62, 62.63, 59.11, 49.28, 48.09, 47.73, 40.19, 29.47, 28.43, 25.66, 23.37, 17.99, 17.93, 17.66, 17.48, 11.95.

HRMS-TOF (+): m/z = 669.4264 expected 669.4259 $[M+H]^+$

Compound 2: To a solid **10** (0.058 g, 0.087 mmol) trifluoroethanol (TFE) (2 mL) and 12 N HCl (0.04 mL) were added to form a faintly tan coloured solution. The reaction was placed into a preheated oil bath (50 °C) and kept for 2 h. The reaction was cooled down, diluted with toluene (2 mL) and concentrated. The residue was successively co-evaporated with toluene (2 mL), MeOH (3 mL) and briefly dried in vacuum. The residue was dissolved in DMF (1 mL) N-methyl morpholine (NMM; 0.05 mL) was added followed by solid FITC (0.023 g) and the reaction was kept at RT for 30 min. More FITC (0.01g) and saturated aqueous $NaHCO_3$ solution (0.1 mL) were added; after an additional 20 min, LC-MS confirmed disappearance of the starting material. The reaction mixture was diluted with 95% water-MeOH (3 mL), loaded onto RP C18 column and eluted with a water-MeOH gradient 5-95%. The appropriate fractions were pooled and concentrated to give the crude target product. This was re-purified by HPLC chromatography using Waters Xselect CSH 10×150 column with gradient 5 to 95 % buffer B (MeCN, 0.1 % formic acid) in buffer A (water, 0.1 % formic acid). Appropriate fractions were pooled and freeze dried to give 0.038 g (0.056 mmol, 65%) of the target compound as fluffy orange material.

^{1}H NMR (400 MHz, DMSO-d_6) δ 10.13 (s, 2H), 9.93 (s, 1H), 8.40 – 8.24 (m, 2H), 8.15 (d, *J* = 8.9 Hz, 1H), 7.76 (d, *J* = 8.3 Hz, 1H), 7.18 (d, *J* = 8.3 Hz, 1H), 7.10 (s, 1H), 6.68 (s, 2H), 6.67 – 6.50 (m, 4H), 5.56 – 5.41 (m, 1H), 5.36 (d, *J* = 4.8 Hz, 1H), 5.06 – 4.96 (m, 1H), 4.74 (t, *J* = 8.8 Hz, 1H), 4.01 (d, *J* = 11.1 Hz, 1H), 3.73 (d, *J* = 11.1 Hz, 1H), 3.70 – 3.63 (m, 2H), 3.58 (m, 3H), 2.56 – 2.48 (m, 2H, overlapped with DMSO), 1.97 – 1.75 (m, 2H), 1.88 (s, 3H).

^{13}C NMR (101 MHz, DMSO) δ 180.81, 169.72, 169.05, 160.39, 152.48, 143.90, 141.95, 129.57, 124.57, 117.09, 113.96, 113.32, 110.37, 102.70, 84.43, 73.23, 68.99, 61.61, 60.55, 49.89, 44.05, 28.60, 25.39, 23.46.

HRMS-TOF (+): m/z = 688.2073 expected 688.2077 $[M+H]^+$

Cloning and Protein Purification

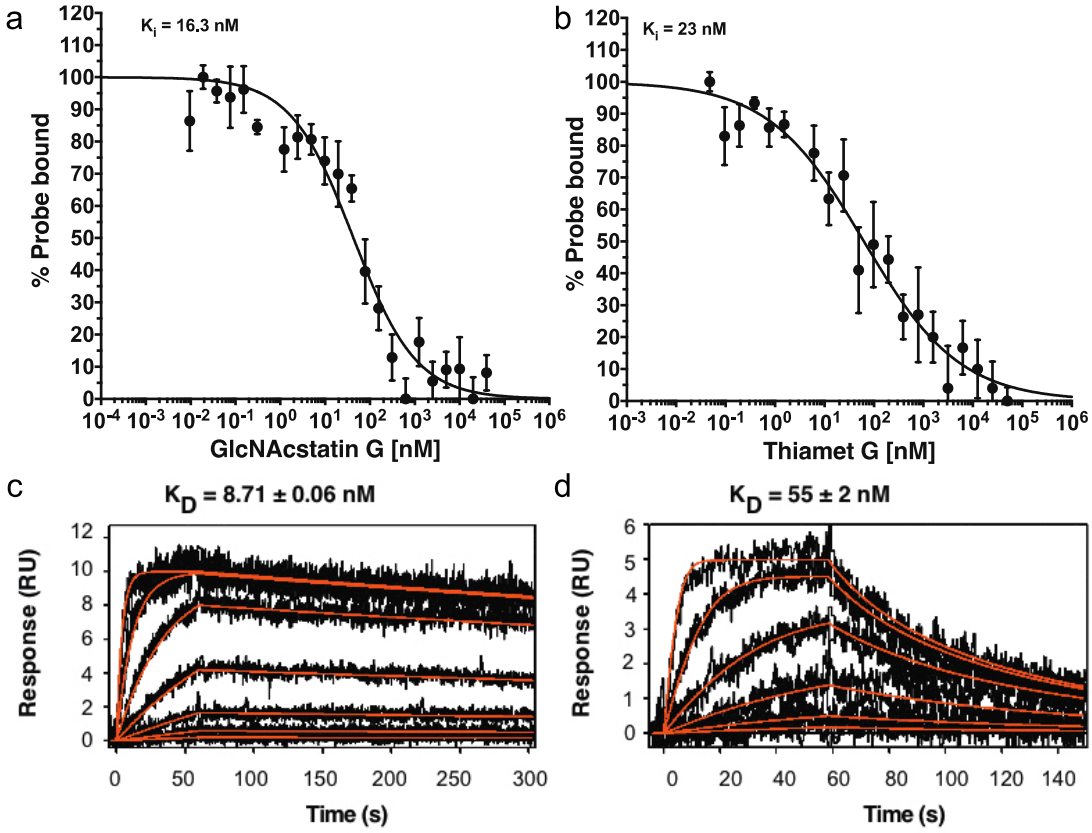
*Cp*OGA (31-618) was purified as described previously[2]. hOGA (60-916), codon optimized for bacterial protein expression was cloned with an N-terminal 6 x His tag by a previously described restriction-free cloning method [3, 4] into the pGEX6P1 (GE healthcare) backbone, which served to remove the GST tag encoded by the vector. The resulting construct was transformed into BL21(DE3)-Gold (Agilent) cells. Cells were grown overnight at 37 ˚C in Luria-Bertani medium containing 50 µg/ml Ampicillin (LB-Amp) and used at 10 mL L^{-1} to inoculate 6 L of fresh LB- Amp. Cells were grown to an OD_{600} of 0.6-0.8, transferred to 16 ˚C and induced with 250 µM of IPTG and harvested after 16 h by centrifugation for 30 min at 3500 rpm (4 ˚C). Cell pellets were resuspended in 10-20 mL L^{-10} of 50 mM Tris, and 250 mM NaCl (lysis buffer) at pH 7.5 supplemented with protease inhibitors (1 mM benzamidine, 0.2 mM PMSF and 5 µM leupeptin), DNAse and lysozyme. Cells were lysed using a continuous flow cell disrupter (Avestin, 3 passes at 20 kpsi) and the lysate was cleared by centrifugation (30 min, 15,000 rpm, 4 ˚C). Supernatants were collected and loaded on to 5 mL of IMAC sepharose (GE Healthcare) charged with $NiSO_4$ and pre-equilibrated with lysis buffer. Loaded resin was washed with 500 mL of lysis buffer containing 30 mM imidazole. Elution of the protein was achieved using 25 mM Tris pH 7.5, 150 mM NaCl (1 x TBS) supplemented with 200 mM imidazole. Eluates were dialysed into 1 x TBS following which cleavage of the 6 x His tag was performed using 6 x His tagged PreScission™ protease at 4˚C for 16 h. Negative IMAC was performed to remove the protease and the protein dialysed into 25 mM Tris pH 7.5 for anion exchange chromatography (AIEX). Protein was loaded onto a 5 mL Hi-Trap QFF column (GE Healthcare) and elution performed over a linear gradient of NaCl concentrations peaking at 500 mM at 25 column volumes of buffer (Buffer A – 25 mM Tris pH 7.5, Buffer B- Buffer A supplemented with 500 mM NaCl). Fractions containing relatively pure protein were pooled and subjected size exclusion chromatography using a Superdex 200, 26/60 column and 1 x TBS buffer. Fractions containing pure protein were pooled and concentrated using spin concentrators and purity was assessed by SDS-PAGE followed by Coomassie staining. Protein was snap frozen with a final concentration of 20 % glycerol until use.

Surface Plasmon Resonance

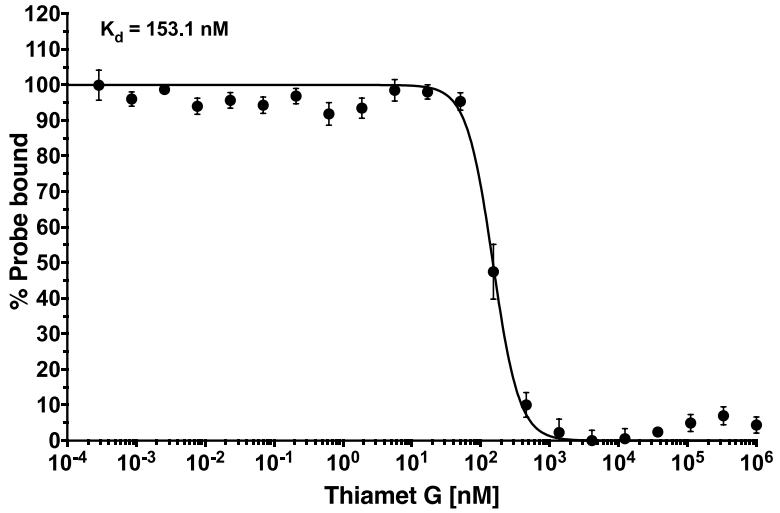
*Cp*OGA was chemically biotinylated using the EZ-Link NHS-PEG4-Biotin kit (Thermo) according to the manufacturer's instructions, except a 1:1 molar ratio of biotinylation reagent to protein was used. Protein was captured on a neutravidin surface prepared on high capacity amine sensor chip of a Mass-1 instrument (Sierra Sensors) at densities ~3,600–3,900 RU. All experiments were performed at 25 °C. Ligands were injected over captured protein at flow rate 30 µL min^{-1} in running buffer (1 x TBS buffer pH 7.5 containing 0.05% Tween20), with each compound injected in duplicates in concentration series adjusted specifically around their affinities. Association was measured for 60 s and dissociation for 120 s. All data were double referenced for blank injections of buffer and biotin-blocked Streptavidin surface. Analyser 2 (Sierra Sensors) and Scrubber 2 (BioLogic Software) were used to process and analyse the data.

Binding affinity of GBF to the hexosaminidases HexA/B

β-N-acetylglucosaminidase isolated from bovine kidneys (HexA/B) was purchased from Sigma-Aldrich (A2415-5UN). Reactions were carried out on black, 384-well plates (PerkinElmer) and solutions contained 1 nM GlcNAcstatin and varying concentrations of HexA/B in 0.1 M Tris-HCl pH 7.4, 150 mM NaCl, 1% DMSO in a total volume 25 µL. Polarization was measured on Pherastar FS plate reader (BMG Labtech) at excitation and emission wavelengths of 485 nm and 520 nm, respectively. Binding affinities were determined by non-linear regression curve fitting (One site – specific binding).



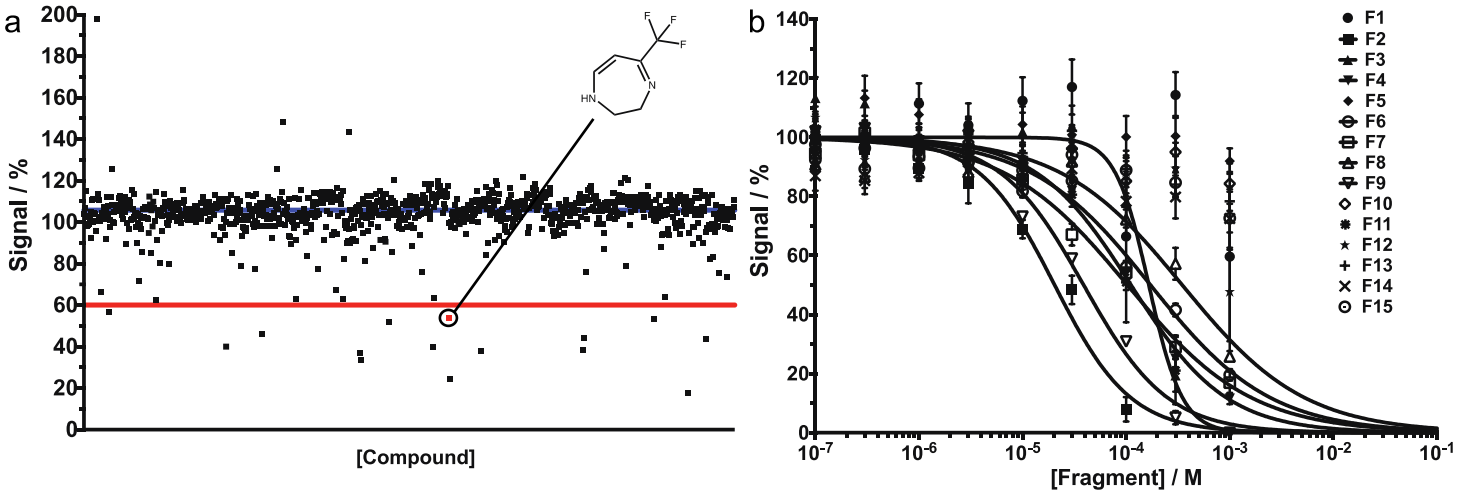
Binding affinity of GlcNAcstatin G and Thiamet G to *Cp*OGA by FP and SPR. Dose-response curves from FP assay showing the displacement from *Cp*OGA of a fixed concentration of fluorescent probe by increasing concentrations of **(a)** GlcNAcstatin G or **(b)** Thiamet G. Highest amount of probe bound to enzymes in the absence of inhibitors was set as 100%. Data points were fitted to a four-parameter equation for dose-dependent inhibition using Prism (GraphPad). Experiments were performed in triplicate and error bars represent standard error of the mean. **(c)** SPR sensorgram for the binding of GlcNAcstatin G to *Cp*OGA. GlcNAcstatin G was injected in duplicates at various concentrations (0.0046-3.33 uM, 0.0046-3.3 uM, 0.9-666 uM and 0.9-666 uM). **(d)** SPR sensorgram for the binding of Thiamet G to *Cp*OGA. Thiamet G was injected in duplicates at various concentrations (0.004-1 uM, 0.04-10 uM, 8.2-2000 uM and 8.2-2000 uM respectively). RU, relative units.



Binding affinity of Thiamet G to hOGA. Dose-response curves from FP assay showing the displacement from hOGA of a fixed concentration of fluorescent probe (50 nM) by increasing concentrations of Thiamet G. Highest amount of probe bound to enzymes in the absence of inhibitors was set as 100%. Data points were fitted to a four-parameter equation for dose-dependent inhibition using Prism (GraphPad). Experiments were performed in triplicate and error bars represent standard error of the mean.

Summary of measured IC_{50} values and calculated K_i values for known OGA inhibitors. Inhibition constants were calculated using the equation reported by Nikolovska-Coleska *et al.*[5]. All values are given in nM.

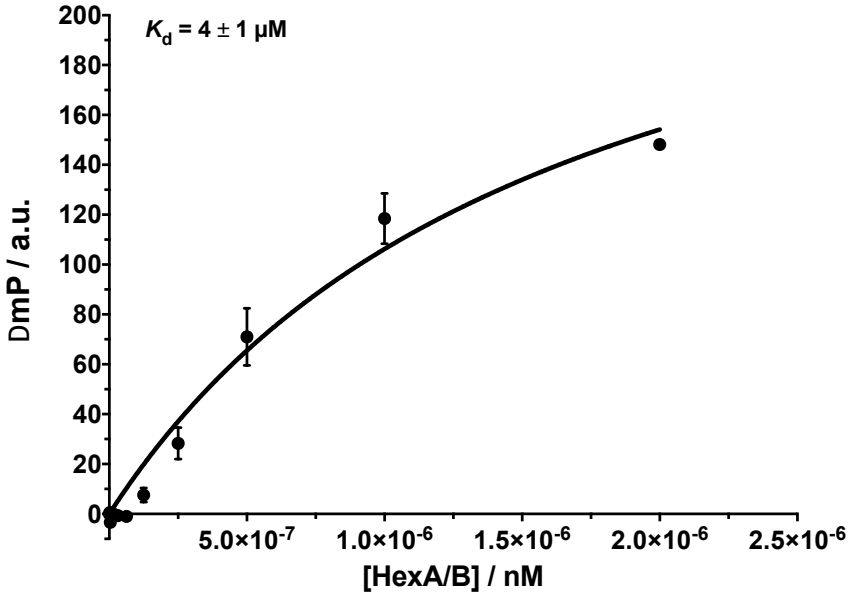
CpOGA	**GlcNAcstatin B**	**GlcNAcstatin G**	**Thiamet G**
HillSlope	-0.9 ± 0.1	-0.6 ± 0.1	-0.45 ± 0.1
IC50	14.5	42.9	62.6
95% CI (profile likelihood) for IC50	10.6 to 20.2	27.4 to 65.8	40.6 to 95.9
K_i calculated	6.6	16.3	37
hOGA	**GlcNAcstatin B**	**GlcNAcstatin G**	**Thiamet G**
HillSlope	-0.97 ± 0.2	-0.97 ± 0.2	-2.3 ± 0.4
IC50	63.2	57.8	150.1
95% CI (profile likelihood) for IC50	43 to 93.2	36.2 to 94.9	130.7 to 173.5
K_i calculated	86.8	81.5	153.1



High-throughput screen of the Maybridge Ro3 1000 fragment library. Scatter plot showing the reduction of FP in the presence of 200 µM fragment at a fixed concentration of labeled probe and *Cp*OGA. Scatter plot was generated using Prism (GraphPad). The median is shown as a blue line and the > 40% reduction in FP cut-off is shown as a red line.

Summary of measured IC_{50} values and calculated apparent K_i values for binders identified from a high-throughput screen of the Maybridge Ro3 1000 library. A 10 M fictive concentration was introduced to simulate 100% displacement needed for K_i calculation. Inhibition constants were calculated using the equation reported by Nikolovska-Coleska *et al.*[5]. N.d. = not determinable.

***Cp*OGA**	**HillSlope**	**IC_{50} / M**	**app. K_i calculated / µM**	**95% CI (profile likelihood) / IC_{50}**
F1	~ -11.35 ± 128342	~ 1.0E-03	n.d.	(Very wide)
F2	-1.2 ±0.2	2.0E-05	9.3	-4.9 to -4.5
F3	-2.7 ± 0.6	1.7E-04	61	-3.9 to -3.7
F4	1.0 ± 0.1	1.1E-04	42	-4.0 to -3.9
F5	~ -7 ± 2487	~ 1.4E-03	n.d.	(Very wide)
F6	-0.8 ± 0.1	1.7E-04	64	-3.9 to -3.7
F7	-0.7 ± 0.1	1.0E-04	36	-4.1 to -3.9
F8 (5TFD)	-0.7 ± 0.1	3.5E-04	146	-3.7 to -3.2
F9	-1.1 ± 0.2	3.9E-05	14	-4.6 to -4.3
F10	-0.4 ± 0.1	4.5E-02	n.d.	-2.3 to -0.6
F11	-1.3 ± 0.3	9.7E-05	33.4	-4.2 to -3.8
F12	-1.4 ± 0.4	9.9E-04	n.d	-3.1 to -2.6
F13	-0.5 ± 0.1	1.4E-02	n.d.	-2.5 to -1.1
F14	-0.2 ± 0.1	2.4E-02	n.d.	-2.4 to -0.7
F15	-0.3 ± 0.1	2.5E-02	n.d.	-2.3 to -0.9



Binding affinity of GlcNAcstatin BF to HexA/B. FP assay showing the binding of GlcNAcstatin BF to HexA/B Binding was measured by incubating a fixed concentration of labeled probe with varying concentrations of enzyme. Data points were fitted to a saturation binding equation using Prism (GraphPad). Experiments were performed in triplicate and error bars represent standard error of the mean.

Supplementary Material Table T3

X-ray diffraction data collection and structure refinement statistics. Values for the highest resolution shell in parenthesis.

	*Cp*OGA + 5TFD
Data collection	
Beamline, wavelength	ID30-A3, 0.9677 Å
Space group	*P61*
Cell dimensions (Å, °)	*a*=*b*=117.79, *c*=147.23, α=β=90, γ=120
Resolution (Å)	50 - 2.60 (2.72-2.60)
R_{merge}	0.24 (2.49)
$I/\sigma I$	7.9 (1.0)
$CC_{1/2}$	0.99 (0.705)
R_{meas}	0.25 (2.6)
R_{pim}	0.088 (0.923)
Completeness (%)	99.1 (99.5)
Redundancy	7.9 (8.2)
Refinement	
Resolution (Å)	50 – 2.6
No. total reflections	279056
No. unique reflections	35270
R_{work}, R_{free}	0.17 / 0.22
No. atoms	
Protein	4587
Ligand	11
Water	71
B-factor average	
Protein	66.34
Ligand	122.33
R.m.s. deviations	
Bond lengths (Å)	0.0153
Bond angles (°)	1.778
PDB ID	5OXD

Supplementary References

[1] Borodkin, V. S., and van Aalten, D. M. F. (2010) An efficient and versatile synthesis of GlcNAcstatins-potent and selective O-GlcNAcase inhibitors built on the tetrahydroimidazo[1,2-a]pyridine scaffold, *Tetrahedron 66*, 7838-7849.

[2] Schimpl, M., Schuttelkopf, A. W., Borodkin, V. S., and van Aalten, D. M. (2010) Human OGA binds substrates in a conserved peptide recognition groove, *Biochem. J. 432*, 1-7.

[3] van den Ent, F., and Lowe, J. (2005) RF cloning: A restriction-free method for inserting target genes into plasmids, *Journal of Bichemistry and Biophysics Methods 67*, 67-74.

[4] Selvan, N., Mariappa, D., van den Toorn, H. W., Heck, A. J., Ferenbach, A. T., and van Aalten, D. M. (2015) The Early Metazoan Trichoplax adhaerens Possesses a Functional O-GlcNAc System, *J. Biol. Chem. 290*, 11969-11982.

[5] Nikolovska-Coleska, Z., Wang, R., Fang, X., Pan, H., Tomita, Y., Li, P., Roller, P. P., Krajewski, K., Saito, N. G., Stuckey, J. A., and Wang, S. (2004) Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization, *Anal Biochem 332*, 261-273.

JOC *The Journal of Organic Chemistry*

COMPOUND CHARACTERIZATION CHECKLIST

Corresponding Author	jo-
Title	

TO DISPLAY INSTRUCTIONS: Double-click HERE (Formula Bar must be visible: Select View>Formula Bar) TO CLOSE: Enter/Return

COMPOUND	IDENTITY														PURITY					COMPUTATIONAL DATA in SI*			AUTHOR REMARKS
Compound, structure, or table-entry number	New	Known made by new/modified method	Weight and percentage yield	Physical state / mp range if cryst. solid	IR	UV-Vis	1H NMR	13C NMR	MS	MS Accurate mass (HRMS)	Optical rotation/ORD/CD	Enantiomeric/Diastereomeric ratio	X-ray [ORTEP and CIF in SI*]	<>	Copy of 1H/13C NMR spectrum in SI*	Copy of chromatogram in SI*	Quant. GC, HPLC, electrophoresis	Elemental analysis	<>	Cartesian coordinates or Z-matrix	# of imaginary frequencies	Total energy	(* SI = Supporting Information)
2*	X		X	X			X	X		X	X												
4*	X		X	X			X	X		X													
5*	X		X	X			X	X		X	X												
6a*		X	X	X			X	X		X	X												
8*	X		X	X			X	X		X	X												
9*	X		X	X			X	X		X	X												
10*	X		X	X			X	X		X													

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Sep 6, 2017 – 11:52 AM BST

Deposition ID : D_1200006533
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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

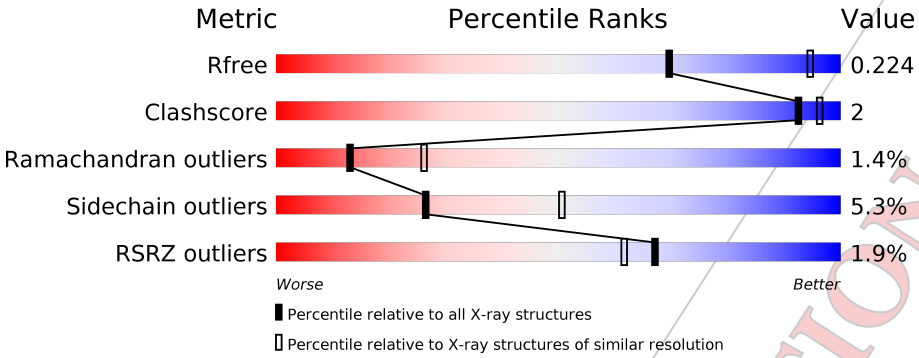
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029824

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 2.60 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}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)

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 <=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	579	2% 90% 10%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4690 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	579	4580	2879	752	932	17	0	0	0

- Molecule 2 is SERINE (three-letter code: SER) (formula: unknown).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	6	3	1	2	0	0

- Molecule 3 is CADMIUM ION (three-letter code: CD) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cd		
3	A	1	1	1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0
3	A	1	Total 1	Cd 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	A	1	Total Cd 1 1	0	0
3	B	1	Total Cd 1 1	0	0
3	B	1	Total Cd 1 1	0	0
3	B	1	Total Cd 1 1	0	0
3	B	1	Total Cd 1 1	0	0
3	B	1	Total Cd 1 1	0	0

- Molecule 4 is 5,6-DIHYDRO-BENZO[H]CINNOLIN-3-YLAMINE (three-letter code: DRG) (formula: unknown).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	X	1	Total C F N 11 6 3 2	0	0

- Molecule 5 is water.

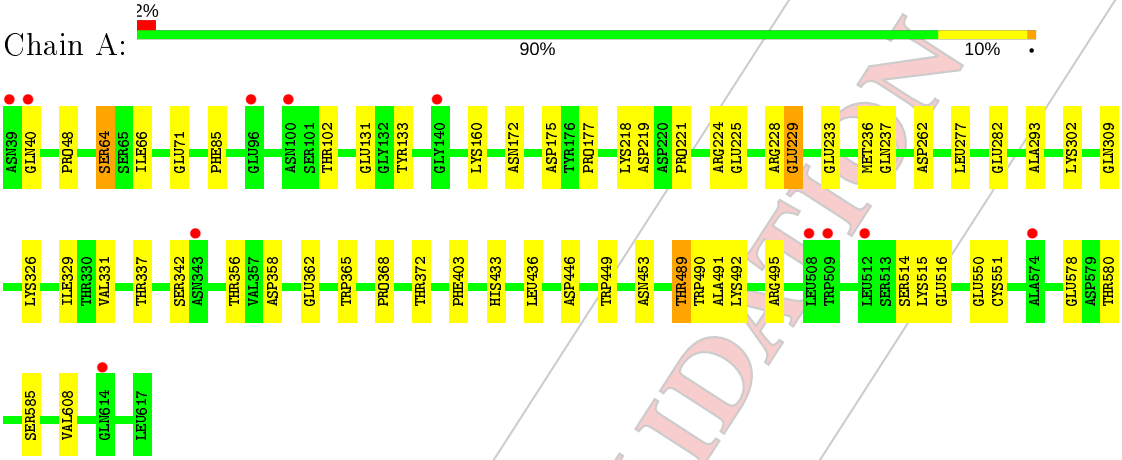
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	71	Total O 71 71	0	0

3 Residue-property plots

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:

Chain A:



GLOBAL-STATISTICS INFOmissingINFO

4 Model quality

4.1 Standard geometry

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

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	0.76	1/4675 (0.0%)	0.88	1/6348 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	131	GLU	CD-OE2	5.86	1.32	1.25

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	495	ARG	NE-CZ-NH2	-5.58	117.51	120.30

There are no chirality outliers.

There are no planarity outliers.

4.2 Too-close contacts

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	4580	0	4366	15	0
2	A	6	0	4	0	0
3	A	17	0	0	0	1
3	B	5	0	0	0	1
4	X	11	0	0	1	0
5	C	71	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4690	0	4370	16	1

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 2.

All (16) 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)
4:X:1:DRG:CAH	4:X:1:DRG:FAK	1.56	1.41
1:A:221:PRO:O	1:A:225:GLU:O	2.18	0.61
1:A:66:ILE:HG22	1:A:102:THR:HB	1.94	0.50
1:A:85:PHE:CD1	1:A:160:LYS:HG2	2.49	0.48
1:A:436:LEU:HD11	1:A:551:CYS:HB3	1.95	0.46
1:A:228:ARG:O	1:A:229:GLU:CB	2.63	0.46
1:A:177:PRO:HB3	1:A:449:TRP:CE3	2.52	0.45
1:A:48:PRO:HB3	1:A:446:ASP:HA	2.01	0.43
1:A:262:ASP:C	1:A:262:ASP:OD1	2.59	0.41
1:A:337:THR:HB	1:A:368:PRO:HA	2.02	0.41
1:A:224:ARG:C	1:A:225:GLU:O	2.56	0.41
1:A:514:SER:O	1:A:516:GLU:N	2.54	0.41
1:A:490:TRP:CE3	1:A:491:ALA:HB2	2.56	0.41
1:A:133:TYR:CE1	1:A:175:ASP:HB3	2.55	0.41
1:A:293:ALA:HA	1:A:329:ILE:O	2.21	0.40
1:A:433:HIS:HB3	1:A:550:GLU:OE1	2.21	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:1624:CD:CD	3:B:6:CD:CD[2_565]	1.50	0.70

4.3 Torsion angles (i)

4.3.1 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	577/579 (100%)	541 (94%)	28 (5%)	8 (1%)	13 26

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	515	LYS
1	A	64	SER
1	A	233	GLU
1	A	403	PHE
1	A	578	GLU
1	A	229	GLU
1	A	489	THR
1	A	608	VAL

4.3.2 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	494/494 (100%)	469 (95%)	25 (5%)	28 52

All (25) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	40	GLN
1	A	64	SER
1	A	71	GLU
1	A	172	ASN
1	A	218	LYS
1	A	219	ASP
1	A	236	MET
1	A	237	GLN
1	A	277	LEU
1	A	282	GLU
1	A	302	LYS
1	A	309	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	326	LYS
1	A	331	VAL
1	A	342	SER
1	A	356	THR
1	A	358	ASP
1	A	362	GLU
1	A	365	TRP
1	A	372	THR
1	A	453	ASN
1	A	489	THR
1	A	492	LYS
1	A	580	THR
1	A	585	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	390	ASN
1	A	442	HIS
1	A	453	ASN
1	A	538	ASN

4.3.3 RNA (i)

There are no RNA molecules in this entry.

4.4 Non-standard residues in protein, DNA, RNA chains (i)

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

4.5 Carbohydrates (i)

There are no carbohydrates in this entry.

4.6 Ligand geometry (i)

Of 24 ligands modelled in this entry, 22 are modelled with single atom - 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	$\#\|Z\| > 2$
2	SER	A	618	-	5,?,?	2.13	1 (20%)	1,?,?	1.27	0
4	DRG	X	1	-	7,?,?	3.11	4 (57%)	7,?,?	1.51	1 (14%)

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
2	SER	A	618	-	-	0/2/?/?	0/0/?/?
4	DRG	X	1	-	-	0/6/?/?	0/0/?/?

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	X	1	DRG	FAI-CAH	2.62	1.43	1.32
4	X	1	DRG	CAC-NAB	2.98	1.52	1.46
2	A	618	SER	CA-C	4.16	1.55	1.50
4	X	1	DRG	CAA-CAG	4.38	1.40	1.33
4	X	1	DRG	FAK-CAH	5.40	1.56	1.32

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	X	1	DRG	CAA-CAG-CAF	-2.56	128.27	131.87

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
4	X	1	DRG	1	0

4.7 Other polymers

There are no such residues in this entry.

4.8 Polymer linkage issues

There are no chain breaks in this entry.

5 Fit of model and data

5.1 Protein, DNA and RNA chains

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, 95^{th} 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	579/579 (100%)	0.11	11 (1%) 67 61	43, 61, 91, 114	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	40	GLN	2.7
1	A	39	ASN	2.6
1	A	509	TRP	2.5
1	A	100	ASN	2.4
1	A	574	ALA	2.3
1	A	343	ASN	2.2
1	A	614	GLN	2.2
1	A	508	LEU	2.1
1	A	140	GLY	2.1
1	A	512	LEU	2.0
1	A	96	GLU	2.0

5.2 Non-standard residues in protein, DNA, RNA chains

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

5.3 Carbohydrates

There are no carbohydrates in this entry.

5.4 Ligands

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains.

The B-factors column lists the minimum, median, 95^{th} 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	LLDF	B-factors(Å^2)	Q<0.9
3	CD	A	1623	1/?	0.99	0.17	1.01	75,75,75,75	0
3	CD	B	6	1/?	1.00	0.17	0.66	65,65,65,65	0
3	CD	B	2	1/?	0.95	0.13	-0.92	183,183,183,183	0
3	CD	A	1629	1/?	0.88	0.09	-1.39	188,188,188,188	0
3	CD	A	1628	1/?	1.00	0.15	-1.68	75,75,75,75	0
3	CD	A	1637	1/?	0.88	0.08	-2.22	185,185,185,185	0
3	CD	B	1	1/?	0.95	0.06	-4.47	156,156,156,156	0
3	CD	A	1625	1/?	1.00	0.16	-	63,63,63,63	0
3	CD	A	1631	1/?	0.96	0.13	-	126,126,126,126	0
3	CD	A	1626	1/?	0.98	0.05	-	104,104,104,104	0
3	CD	B	5	1/?	0.77	0.13	-	189,189,189,189	0
3	CD	A	1627	1/?	0.57	0.13	-	182,182,182,182	0
4	DRG	X	1	11/?	0.79	0.28	-	98,114,140,141	0
2	SER	A	618	6/?	0.68	0.49	-	83,103,113,126	0
3	CD	A	1634	1/?	0.73	0.07	-	158,158,158,158	0
3	CD	A	1619	1/?	0.99	0.15	-	105,105,105,105	0
3	CD	A	1632	1/?	0.88	0.18	-	187,187,187,187	0
3	CD	B	3	1/?	0.72	0.10	-	185,185,185,185	0
3	CD	A	1636	1/?	0.27	0.28	-	207,207,207,207	0
3	CD	A	1624	1/?	1.00	0.15	-	212,212,212,212	0
3	CD	A	1622	1/?	0.96	0.04	-	134,134,134,134	0
3	CD	A	1635	1/?	0.65	0.33	-	230,230,230,230	0
3	CD	A	1621	1/?	0.98	0.11	-	94,94,94,94	0
3	CD	A	1630	1/?	1.00	0.16	-	65,65,65,65	0

5.5 Other polymers (i)

There are no such residues in this entry.