Electronic Supplementary Information

Protein crystal structures with ferrocene and ruthenocene-based enzyme inhibitors

Adam J. Salmon, Michael L. Williams, Andreas Hofmann and Sally-Ann Poulsen*

*Corresponding author e-mail: s.poulsen@griffith.edu.au

Materials and Methods

Production of recombinant human carbonic anhydrase (CA) II

Human recombinant CA II was expressed in *Escherichia coli* BL21(DE3) using the plasmid pACA-I (kindly provided by Prof Carol Fierke) and an in-house implementation of the auto-induction method[1]. After cell lysis, soluble protein was purified by anion exchange chromatography with Q-Sepharose resin. Protein quality was checked by SDS-PAGE throughout.

Crystallisation

The concentrated protein (18 mg mL^{-1}) was subjected to co-crystallisation with ligands by the hanging drop vapour diffusion method, using a selection of ammonium sulfate containing conditions from our in-house factorial collection (2.6 or 2.9 M $(NH_4)_2SO_4$, and 0.1 M MES pH 6.5 or 0.1 M glycine pH 9.5). Crystallisation droplets consisted of 3 µL protein, 2 µL reservoir and 1 µL ligand solution (60 mM in MeOH). Monoclinic crystals appeared after 1-2 weeks.

Protein X-ray crystallography

X-ray diffraction data were obtained under cryo conditions at the Australian Synchrotron beam line MX1 equipped with a Quantum ADSC CCD detector. Data were processed with MOSFLM[2] and SCALA from the CCP4 suite[3]. Crystal structures were determined using the isomorphous structure of native human CA II. Ligand constraints for cyclopentadienyl-substituted triazole benzenesulfonamides were obtained by PRODRG and manually edited to obtain the full

metallocene containing ligand topologies, guided by parameters obtained from small molecule structures (ferrocene: Fe-C distances 2.05 Å, staggered conformation of Cp rings; ruthenocene: Ru-C distances 2.29 Å, eclipsed conformation of Cp rings). Computational refinement of the ligand-bound crystal structures was carried out with CNS[4] or REFMAC[5], interspersed with visual inspection and manual model building with O[6] and Coot[7]. The geometry of refined structures was evaluated with PROCHECK[8].

Table S1: Data collection and refinement statistics

Crystal	CA II:**1**	CA II:**2**	CA II:**3**	CA II:**4**
Data collection				
X-ray source	AS MX1	AS MX1	AS MX1	AS MX1
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions: *a*, *b*, *c* (Å); β□(°)	42.4, 41.6, 72.2; 104.6	42.5, 41.6, 72.2; 104.8	42.5, 41.6, 72.3; 104.6	42.5, 41.6, 72.5; 104.9
Wavelength (Å)	1.509	0.9568	1.509	0.9568
Resolution (Å)	2.0	1.5	2.2	1.6
R_{sym}	0.078 (0.157)	0.059 (0.174)	0.077 (0.258)	0.069 (0.256)
Completeness	0.946 (0.924)	0.996 (0.985)	0.988 (0.954)	0.991 (0.943)
Redundancy	3.8 (3.8)	4.0 (3.9)	3.5 (3.3)	7.1 (5.7)
Refinement				
Resolution (Å)	2.0	1.5	2.2	1.6
No. reflections: work / test	15173 / 780	37150 / 1969	11774 / 630	28932 / 1596
R_{work} / R_{free}	0.187 (0.217) / 0.223 (0.235)	0.176 (0.192) / 0.201 (0.220)	0.186 (0.204) / 0.241 (0.275)	0.193 (0.214) / 0.203 (0.223)
No. atoms				
Protein	2049	2059	2049	2049
Ion	1 (Zn^{2+})	1 (Zn^{2+})	1 (Zn^{2+})	1 (Zn^{2+})
Ligand	26 (**1**), 12 (2x glycerol)	26 (**2**), 6 (glycerol)	26 (**3**), 12 (2x glycerol)	26 (**4**), 6 (glycerol)
Water	172	369	122	191
Average B-factors				
Protein (Å^2)	16.3	5.57	23.8	15.0
Ion (Å^2)	11.1	2.00	15.0	8.27
Ligand (Å^2)	22.0 (**1**), 25.7 (glycerol)	12.7 (**2**), 10.3 (glycerol)	45.5 (**3**), 32.5 (glycerol)	24.4 (**4**), 23.3 (glycerol)
Water (Å^2)	29.1	19.3	33.9	25.5
R.m.s deviations				
bond lengths (Å)	0.005	0.011	0.005	0.005
bond angles (°)	1.403	1.251	1.410	1.420
B-factor for bonded protein atoms	2.72	1.23	3.11	2.35
B-factor for bonded ligand atoms	2.11 (**1**), 3.24 (glycerol)	2.54 (**2**), 2.42 (glycerol)	4.56 (**3**), 4.93 (glycerol)	2.61 (**4**), 3.35 (glycerol)
Ramachandran plot[a]	0.870, 0.120, 0.009, 0.0	0.889, 0.106, 0.005, 0.0	0.880, 0.111, 0.009, 0.0	0.894, 0.106, 0.0, 0.0

Datasets were obtained from one crystal each. Values in parentheses are for highest-resolution shell. Coordinates and structure factors have been deposited with the PDB accession codes: 3P55 (CA II:**1**), 3P3H (CA II:**2**), 3P44 (CA II:**3**) and 3P3J (CA II:**4**).

[a]Residues in most favoured regions, additional allowed regions, generously allowed regions, disallowed regions.

References

1. Studier, F., Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 2005, *41*, 207-234.

2. Leslie, A., Recent changes to the MOSFLM pacakge for processing film and image place data. Joint CCP4 + ESF-EAMCB. *Newsletter on Protein Crystallography* 1992, *26*.

3. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 1994, *50*, 760-763.

4. Brünger, A.; Adams, P.; Clore, G.; Delano, W.; Gros, P.; Grosse-Kunstleve, R.; Jiang, J.; Kuszewski, J.; Nilges, N.; Pannu, N.; Read, R.; Rice, L.; Simonson, T.; Warren, G., Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr. D* 1998, *54*, 905-921.

5. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr.* 1997, *D 53*, 240-255.

6. Jones, T. A.; Zou, J. Y.; Cowan, S.; Kjeldgaard, M., Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr. A* 1991, *47*, 110-119.

7. Emsley, P.; Cowton, K., Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004, *60*, 2126-2132.

8. Laskowski, R.; MacArthur, M.; Moss, D.; Thornton, J., PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 1993, *26*, 283-291.