CHEMBIOCHEM

Supporting Information

One Crystal, Two Temperatures: Cryocooling Penalties Alter Ligand Binding to Transient Protein Sites

Marcus Fischer,[b] Brian K. Shoichet,[b] and James S. Fraser*[a]

cbic_201500196_sm_miscellaneous_information.pdf

Supplemental Methods

Derivation of equation 1:

(1) $\Delta G_{app} = \Delta G_{site} + \Delta G_{penalty}$

(2) $fraction_{bound} = \frac{[L]/Kd_{app}}{1+[L]/Kd_{app}}$

$-RT\,ln\,Kd_{app} = -RT\,ln\,Kd_{site} + \Delta G_{penalty}$

$Kd_{app} = e^{\ln\,Kd_{site} + (\Delta G_{penalty}/-RT)} = e^{x}$

Assuming a minimum occupancy of 5% to observe binding

$occ_{min} > \frac{\frac{[L]}{e^{x}}}{1+\frac{[L]}{e^{x}}}$ (equation 2)

$0.05 > \frac{\frac{[L]}{e^{x}}}{1+\frac{[L]}{e^{x}}}$

$0.05 + 0.05\,[L]/e^{x} > [L]/e^{x}$

$0.05\,e^{x} + 0.05\,[L] > [L]$

$e^{x} > 19\,[L]$

$x > ln\,(19\,[L])$

$\ln\,Kd_{site} + (\Delta G_{penalty}/-RT) > \ln(19\,[L])$

$\Delta G_{penalty} > -RT \ln \frac{19\,[L]}{Kd_{site}}$

Generalization: $\Delta G_{penalty} > -RT\ln\frac{\frac{(1-occ_{min})}{occ_{min}}[L]}{Kd_{site}}$ (equation 1)

Calculation of temperature-dependence of occupancies

The standard Gibbs energetic driving force for benzimidazole binding is 2.16 kcal/ mol at room temperature. We first calculate kd_{app} from $\Delta G = -RT\ln Kd_{app}$ for each temperature and then determine the occupancies via *equation 2* as:

	RT	*200K*	*100K*
[L] (M)	0.033	0.033	0.033
kd_{app} (M)	0.026	0.00435	0.0000189
occupancy	**0.559**	**0.884**	**0.999**

Protein purification and crystallography

The protein was purified and crystallized as described[1]. Crystals were soaked at RT at concentrations indicated in the pdb file with compounds directly dissolved in the cryoprotectant 25% MPD. To minimize differences between the datasets a) MPD (2-methyl-2,4-pentanediol) was used as a precipitant and cryoprotectant to allow immediate re-collection of data on the same crystal at cryogenic temperature, b) data were collected on the same crystal volume, which matched the beam size. Matching the crystal to the beam size and using the same spot at both temperatures should keep any differences in the local distribution of soaked compounds from the rim to the core of the center of the crystal constant between both datasets. Therefore differences due to this effect can be excluded. Data were collected at the ALS Berkeley beamline 8.3.1 and automatically processed using the xia2 pipeline [2] to ensure that unit cell dimensions are

determined individually. Structures were solved by molecular replacement with Phaser [3] using the same model and R_{free} for all structures. Alternating cycles of refinement and model building were carried out in Refmac [4] or PHENIX [5], and Coot [6]. Alternative conformations and the respective ligand were added at the late stages of refinement. Phenix.refine [5] was used for occupancy refinement (strategy=individual_adp+occupancies). RT models were collected on the same crystal as previously deposited structures collected at cryo and deposited to the PDB as listed in SI Table 1. The wt structure was solved using the AutoMR and AutoBuild functions in PHENIX without manual intervention to avoid bias. There were two lines of evidence for ligand binding: The His96 appeared in the open conformation and there was density for the ligand. Therefore the ligand, benzimidazole, was included in the final stage of the occupancy refinement using phenix.refine (deposited to the pdb as 4XVA). Prior to deposition all structures were validated using validation tools within Coot, PHENIX, CCP4 and the ADIT deposition server.

We used temperature as a perturbation to modulate the state of the protein – and temperature affects related parameters like mosaicity and resolution. For the temperature pairs, consistent with previous studies [7-9], the unit cell size shrinks (by 8%) and mosaicity increases upon cryocooling (SI Table 3). Mosaicity results from differences in unit cell parameters (likely in small blocks within the crystal) and generally increases during cryocooling. Our cryocooled datasets have low mosaicity values that are indeed slightly larger than the RT datasets (SI Table 3). This suggests that our cryocooling procedure has not disrupted the crystal quality. Additionally, the resolution is slightly higher at cryogenic temperatures, likely reflecting the reduced thermal motion. Moreover, neither the slight increases in mosaicity nor gains in resolution should significantly alter the ability resolve the contents of the unit cell.

Despite the best efforts to unify the freezing protocol, these differences are a generally unavoidable phenomenon of cryo crystallography and therefore representative of cryogenic structures in the pdb, which comprise ~95% of the pdb. Experimental differences are rarely documented and hard to account for, even in systematic studies. Differences we considered include: differences in sample and loop size, residual liquid around the crystal, crystal age, temperature and composition, changes in the relative humidity, cooling rate, thickness of cold gas layer above LN2 dewar, the choice of cooling liquid and the velocity of gas stream. As we ultimately we care about finding ligands, awareness of such confounding factors can help to plan experiments and motivate careful analysis, especially if the outcome is (thermodynamically) unexpected and not entirely understood.

When applying the same stringency in the crystallographic refinement, the final model had a significantly lower average R_{free} of 0.14 at RT compared to 0.17 at cryogenic temperatures. The ~3% difference in R_{free} indicates superior model quality at RT although data, on average, were collected to similar resolution and showed no obvious signs of radiation damage, judged by an average R_{merge} of <6% at either temperature (SI Table 1-3).

Electron density sampling to reveal alternative side-chain conformations was performed using Ringer[10]. CONTACT [11] was used to identify dynamic contact networks between the cryptic site and the heme. Center of mass (C.O.M.) calculations were analogous to Frauenfelder et al.[8], using http://bioinformatica.isa.cnr.it/CALCOM/ [12] and Microsoft Excel for plotting. All structural figures were generated in PyMOL.

Binding assays

Affinities were measured by monitoring the shift of the heme Soret band near 410nm as described previously [1,13]. Ligand binding was measured by endpoint titration in

100 mM citric acid buffer adjusted to pH 4.5 with Bis-Tris propane. Ligand stocks were made up to 1 M in dimethyl sulfoxide.

Supplemental Results and Discussion

Multiple complete, high-resolution datasets can be collected on a single crystal at room temperature.

It is commonly believed that multiple crystals are required to collect one complete dataset at RT. To show that this is not necessarily true we used a single crystal to collect multiple datasets at RT on a third generation synchrotron source. To facilitate consecutive data collection on a single crystal we attenuated the x-ray beam by using an Al foil and shifting the energy to 13000 eV to reduce flux. To avoid evaporation, we covered the crystal with a MiTeGen plastic sleeve containing 10 µl of 75% reservoir solution and 25% water. We collected eight consecutive datasets on the same crystal volume, with three consecutive datasets collected to a resolution of <2.1Å and 7 consecutive datasets collected to a resolution of <3 Å; when insisting on >99% completeness, high multiplicity, and I/σI > 2 for the highest resolution shell (SI Fig. 8A). This illustrates that multiple datasets can be collected on a single crystal at RT but we note that the number may vary for lower symmetry space groups than P212121. However, CcP contains a radiation-sensitive heme iron, which we expected to impede data quality. We observe similar Fe difference densities throughout collecting 8 redundant datasets and, using Ringer, we found no differences in side-chain conformations of nearby residues His175, that coordinates the iron, and Asp233, which is responsible for binding primarily cationic ligands (SI Fig. 8B)

To investigate which residues are most affected in the binding site we use local density correlation (LDC) to compare the electron density distribution at each residue between

two datasets (SI Fig. 8C). Calculating the correlation coefficient between the first and all subsequent datasets show high correlation of local electron densities across the protein. Globally, the correlation coefficients (CC) between datasets decreased continuously from 95.5% to 64.6%, while CC standard deviations increased from 1.3% to 7.2% (SI Fig. 8C). Both trends indicate increasing structural heterogeneity with progressing damage from the first consecutive to the final dataset. Locally, we further looked at residues with proximity to the heme iron, functional relevance, higher propensity to suffer from radiation damage, or localization on the protein. None of these residues showed a noteworthy deviation from the average LDC. In the final dataset the three most affected residues (with the lowest CC) are V45, G41 and V169, and the least affected (with the highest CC) are S225, F274 and E35; all three are on the exterior of the protein (SI Fig. 8C).

Supplemental Figures and Tables

SI Table 1 Overview of PDB structure pairs collected on the same crystals at different temperatures.

ligand ZINC-ID	cryo	RT
(apo)	4OQ7	4NVA
8652421	4NVB	4XV4
36634	4NVC	4XV8
331945	4NVD	4XV7
331902	4NVE	4XV5
1583444	4NVF	4XV6

SI Table 2 Data collection and refinement statistics

PDB ID	**4XV5**	**4XV4**	**4XVB**	**4XV7**
Ligand	Benzimidazole	2-Amino-5-methylthiazole	Benzamidine	4-Amino-quinazoline
ZINC ID	331902	8652421	36634	331945
Data collection				
Space group	P212121	P212121	P212121	P212121
Cell dimensions				
a, *b*, *c* (Å)	51.53, 76.83, 107.57	51.53, 76.52, 107.30	51.46, 76.95, 107.71	51.48, 76.76, 107.69
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	39.76-1.65 (1.70-1.65)	51.53-1.69 (1.73-1.69)	39.76-1.57 (1.61-1.57)	39.74-1.62 (1.66-1.62)
R_{merge}	0.054 (0.692)	0.068 (0.707)	0.053 (0.555)	0.062 (0.695)
I / σI	17.9 (2.1)	13.7 (2.1)	14.9 (2.4)	13.3 (2.0)
Completeness (%)	99.7 (99.9)	99.9 (99.9)	99.6 (99.8)	99.9 (100)
Redundancy	4.2 (4.2)	4.1 (4.2)	4.3 (4.2)	4.3 (4.2)
Mosaicity	0.109	0.153	0.171	0.276
Wilson B	19.3	19.7	19.1	19.1
Refinement				
Resolution (Å)	39.76-1.65	51.53-1.69	39.76-1.57	39.74-1.62
No. reflections	51656 (5113)	48203 (4753)	60162 (5926)	54844 (5385)
R_{work} / R_{free}	0.1125/ 0.1503	0.1099/ 0.1495	0.1082/ 0.1399	0.1150/ 0.1477
No. atoms				
Protein	2558	2549	2553	2500
Ligand/ion	18	14	9	11
Water	253	268	298	262
B-factors				
Protein	22.0	21.1	21.4	22.3
Ligand/ion	29.5	26.2	14.8	17.2
Water	38.8	39.4	41.3	38.4
R.m.s. deviations				
Bond lengths (Å)	0.018	0.018	0.017	0.017
Bond angles (°)	1.49	1.54	1.59	1.46

*Values in parentheses are for highest-resolution shell.

PDB ID	**4XV6**	**4XVA**
Ligand	MES	Benzimidazole/ wtCcP complex
ZINC ID	1583444	331902
Data collection		
Space group	P212121	P212121
Cell dimensions		
a, *b*, *c* (Å)	51.48, 76.52, 107.3	84.11, 104.76, 185.25
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	33.42-1.55 (1.59-1.55)	92.63-2.66 (2.73-2.66)
R_{merge}	0.04 (0.65)	0.08 (0.70)
I / σI	17.0 (2.2)	12.3 (2.1)
Completeness (%)	99.7 (99.7)	99.6 (97.9)
Redundancy	4.0 (4.1)	4.1 (4.0)
Mosaicity	0.237	0.258
Wilson B	19.9	64.5
Refinement		
Resolution (Å)	33.42-1.55	92.63-2.66
No. reflections	62014 (6158)	47484 (4699)
R_{work} / R_{free}	0.1035/ 0.1355	0.2326/ 0.2764
No. atoms		
Protein	2543	9464
Ligand/ion	12	208
Water	272	284
B-factors		
Protein	24.4	54.4
Ligand/ion	21.8	44.5
Water	43.4	45.0
R.m.s. deviations		
Bond lengths (Å)	0.019	0.016
Bond angles (°)	1.65	1.39

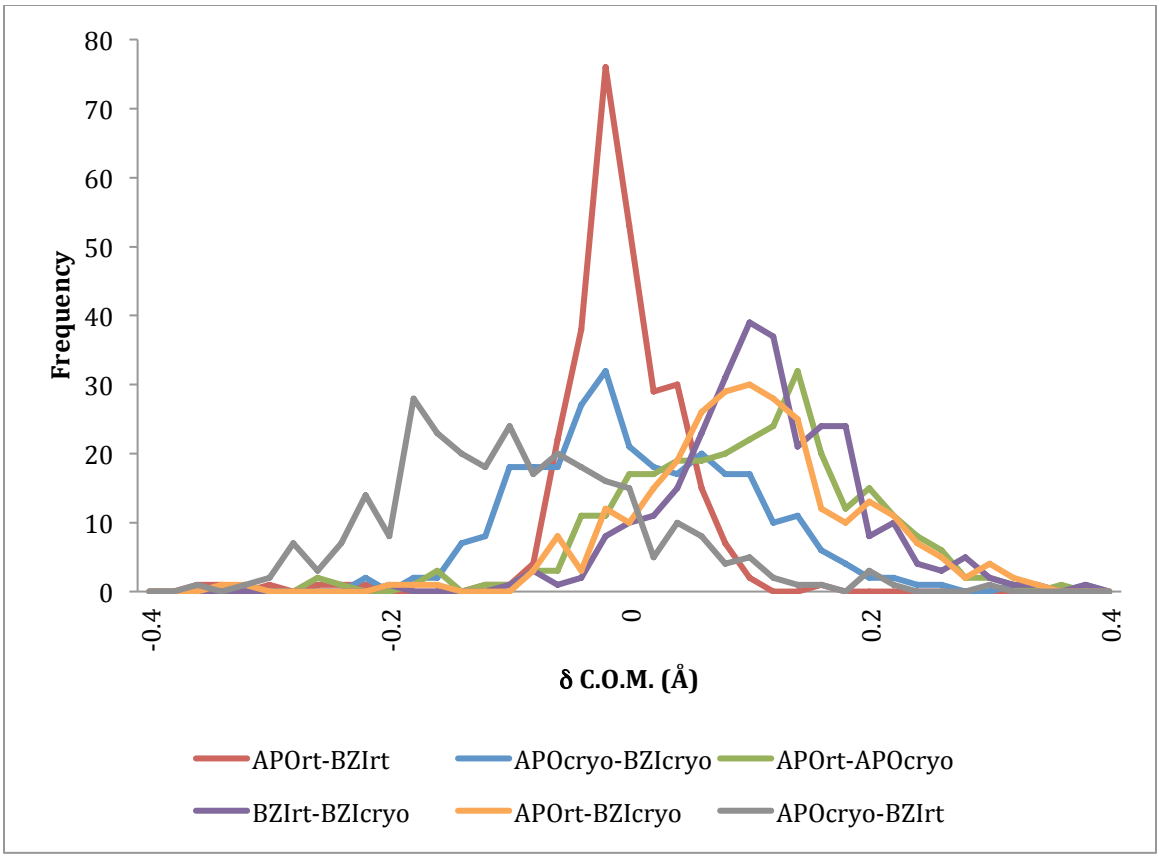
SI Table 3 (a) Values for benzimidazole are comparable at both temperatures. (b) Standard deviations (bold) are larger for cryo data collected on the same crystal; averaged over 6 crystals from SI Table S1.

a)

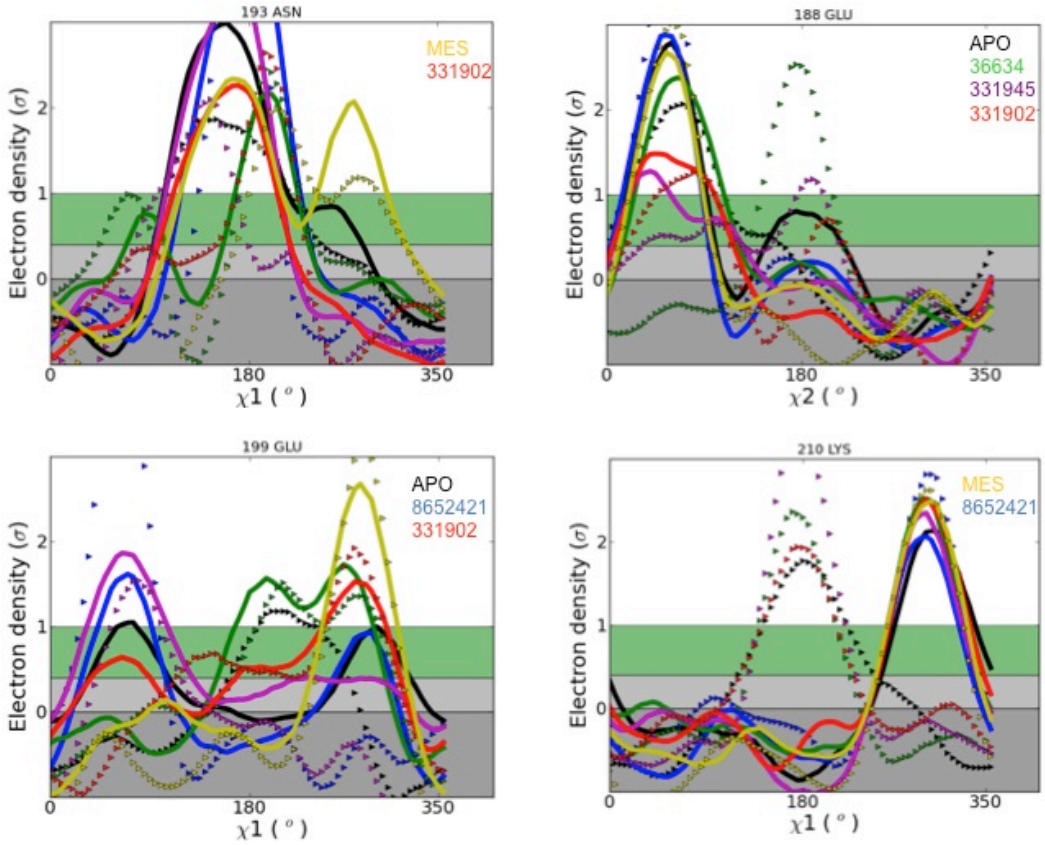
benzimidazole	RT	cryo
resolution	1.65 Å	1.54 Å
R_{merge}	0.054	0.042
mosaicity	0.109	0.175
final R_{free}	0.150	0.180

b)

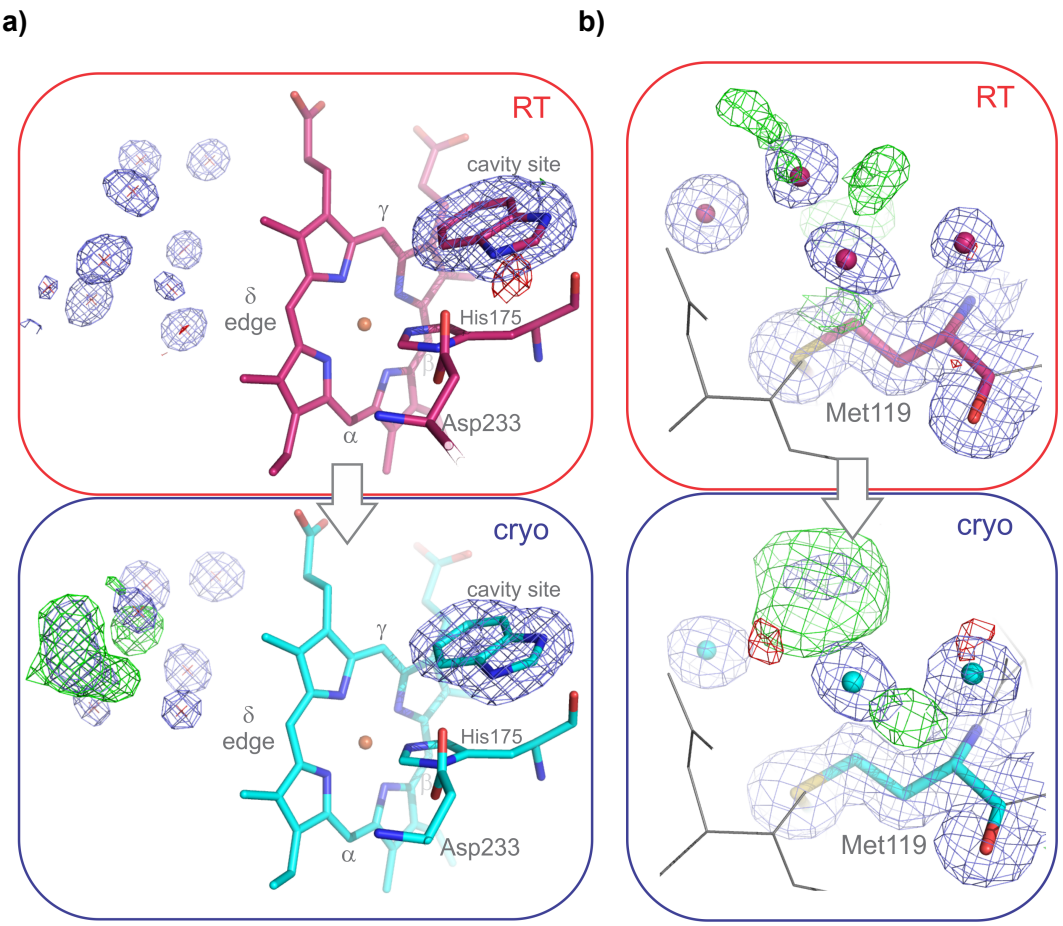
Average	Res	mosaicity	Wilson B	Rfree	UC vol	a	b	c	Protein vol
RT	1.61	0.17	19.4	0.146	424481 **±1678**	51.49 **±0.03**	76.68 **±0.19**	107.5 **±0.18**	50078 **±417**
cryo	1.42	0.28	11.7	0.169	395543 **±6702**	50.83 **±0.11**	74.00 **±0.49**	105.15 **±1.04**	54403 **±3331**



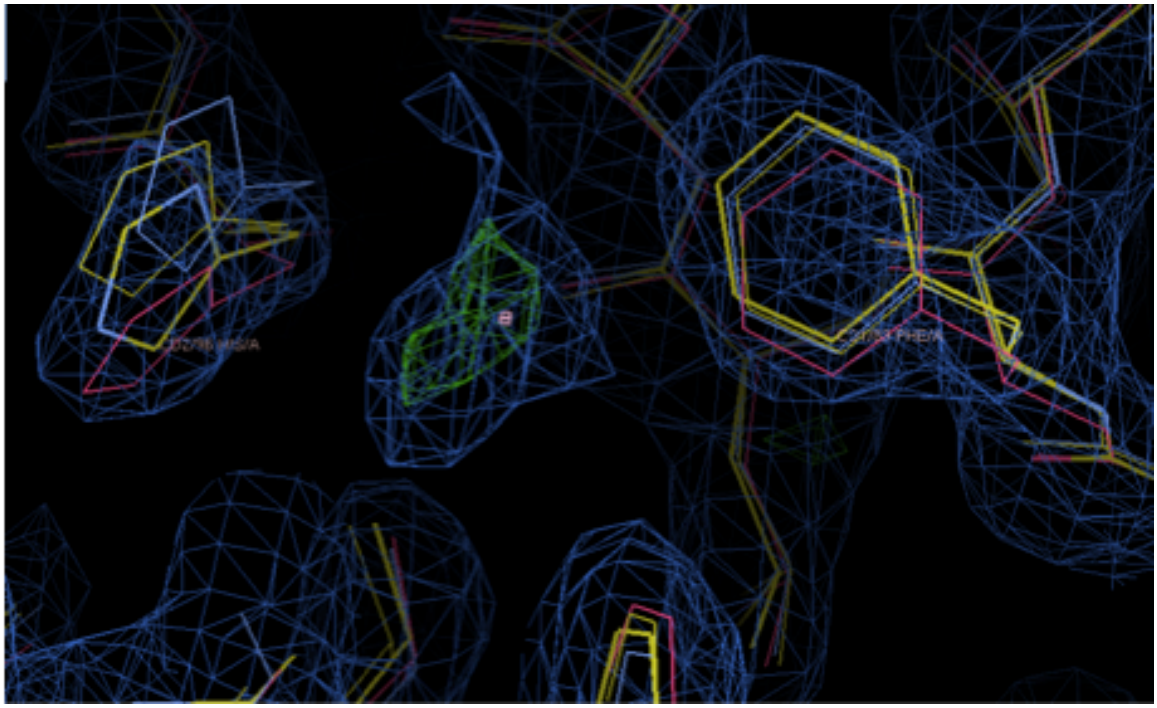
SI Figure 1 As main text Figure 1a with all other combinations.



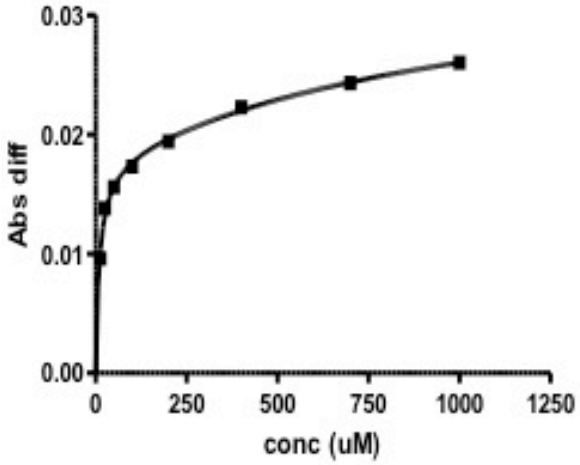
SI Figure 2 Non-systematic changes for several flexible side-chains in response to temperature. Ringer plots show differences in side-chain conformations (chi angles as a function of electron density) when datasets were collected on the same crystal at two temperatures. Temperature pairs of the same complex (or apo structure) are matched in color, with RT data shown as a continuous line and cryogenic data shown as a dashed line. Labels in the top right hand side corner of each plot denote complexes where the corresponding residue conformation responds non-systematically to temperature.



SI Figure 3 Electron density evidence for ligand binding to a) the delta site and the primary cavity site, and b) the M119 site; data of the benzimidazole-CcPga complex were collected on the same crystal at room-temperature (red box) and cryogenic temperature (blue).

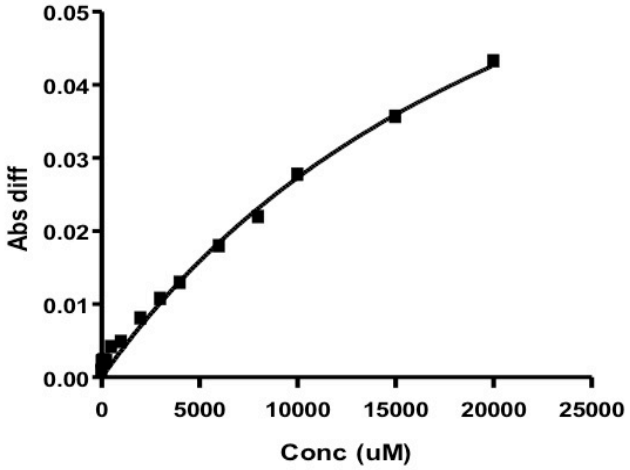


SI Figure 4 Automatic refinement of wildtype CcP data reveals His96 (left) in conformations that open the cryptic site for benzimidazole binding; difference electron density appears with the ligand excluded from refinement.



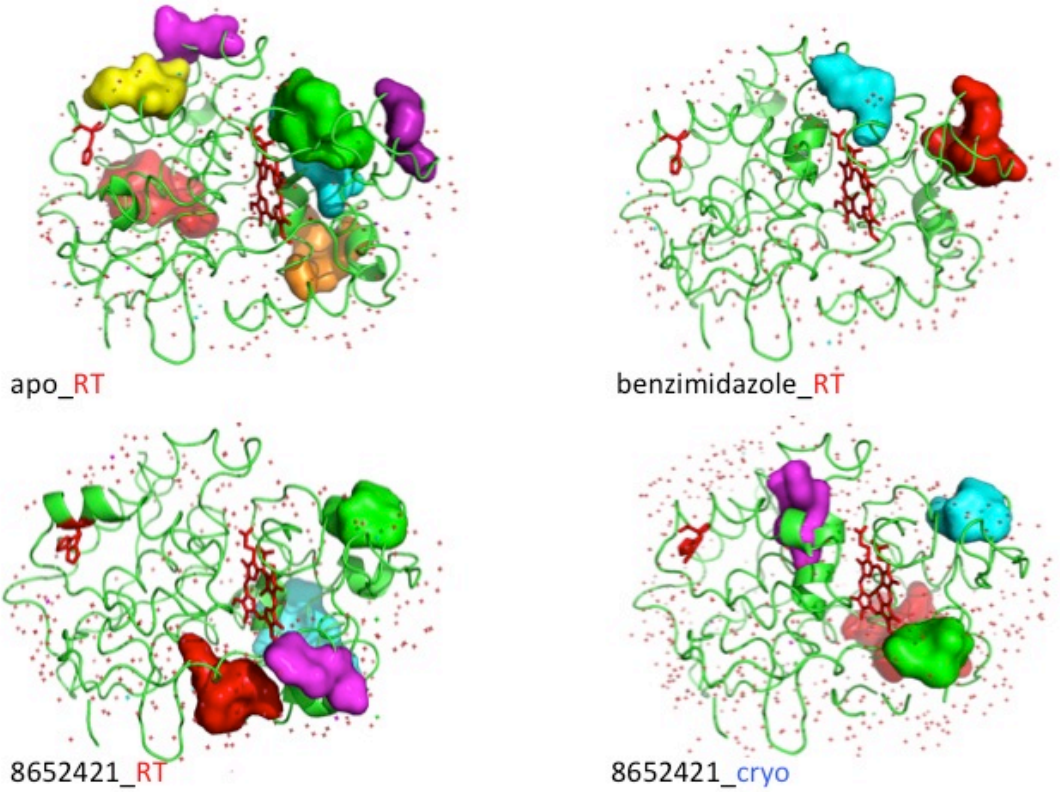
SI Figure 5 Soret band affinity data for CcP cavity mutant and benzimidazole.

A two-site binding model fits a kd1 of 10 μM for the cavity site and kd2 of >1 mM (1169 μM) for the secondary site. The R^2 for the two-site binding hyperbola is 0.9955 compared to 0.9037 for the single site fit, where only the kd1 for the primary site was fitted as 23.5 μM.



SI Figure 6 Soret band affinity data for CcP wildtype and benzimidazole.

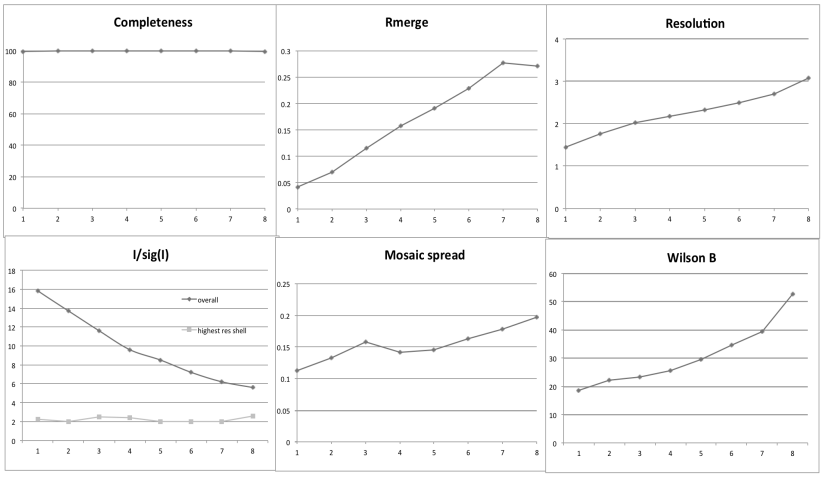
As the engineered cavity is occupied by the wildtype tryptophan W191 in this case, a single site fit was used to determine a kd of 25.6 mM for the secondary site. The R^2 for this single-site binding hyperbola is 0.9908.



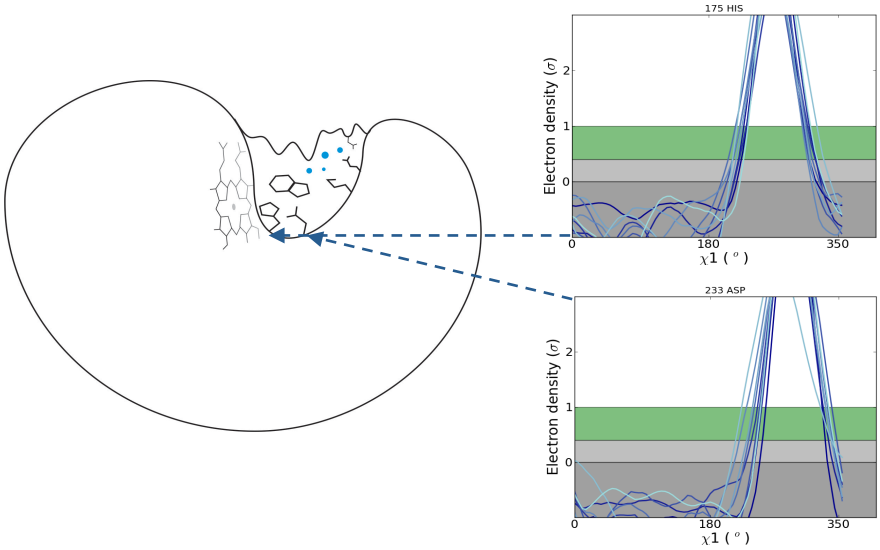
SI Figure 7 CONTACT network analysis[14] shows no steric coupling of the cryptic site and the heme (both shown in red sticks).

SI Figure 8

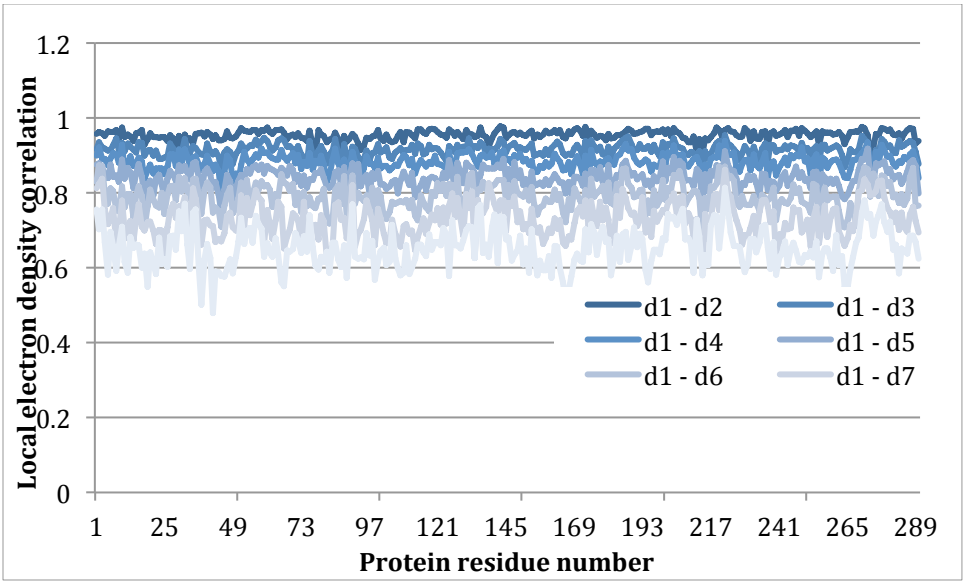
A)



B)



C)



SI Figure 8 A) Data collection statistics across 8 consecutively collected datasets. B) Ringer plots show no significant change in side-chain conformations of residues near the heme-Fe in the course data collection with increasing doses of radiation. C) Local electron density correlation (LDC) shows a negligible impact of radiation damage on local structure at RT between consecutive datasets. Correlation coefficients between the electron densities of 8 datasets continuously collected on a single crystal at room temperature, compared to the first dataset collected, enumerated over all residues. Although the initial unrefined map was used the correlation is very good excluding radiation damage as a source for the differences between the RT-cryo pairs.

Supplemental References

1. Rocklin, G.J. et al. Blind prediction of charged ligand binding affinities in a model binding site. *J Mol Biol* (2013).
2. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta crystallographica. Section D, Biological crystallography* **69**, 1260-73 (2013).
3. McCoy, A.J. et al. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
4. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **67**, 355-367 (2011).
5. Afonine, P.V. et al. Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallographica Section D* **66**, 1153-1163 (2010).
6. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132 (2004).
7. Fraser, J.S. et al. Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16247-52 (2011).
8. Frauenfelder, H. et al. Thermal expansion of a protein. *Biochemistry* **26**, 254-61 (1987).
9. Frauenfelder, H., Petsko, G.A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558-63 (1979).
10. Lang, P.T. et al. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein science : a publication of the Protein Society* **19**, 1420-31 (2010).
11. van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E. & Fraser, J.S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nature methods* **10**, 896-902 (2013).
12. Chelvanayagam, G., Knecht, L., Jenny, T., Benner, S.A. & Gonnet, G.H. A combinatorial distance-constraint approach to predicting protein tertiary models from known secondary structure. *Folding & design* **3**, 149-60 (1998).
13. Barelier, S. et al. Roles for ordered and bulk solvent in ligand recognition and docking in two related cavities. *PLOS ONE* (2013).
14. van den Bedem, H., Bhabha, G., Yang, K., Wright, P.E. & Fraser, J.S. Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat Methods* **10**, 896-902 (2013).