A benchmark for homomeric enzyme active site structure prediction highlights the importance of accurate modeling of protein symmetry.

Stephanie C. Contreras[1], Steve J. Bertolani[1], Justin B. Siegel[1,2,3*]

1. Department of Chemistry, University of California, Davis, Davis, California, United States of America

2. Department of Biochemistry and Molecular Medicine, University of California, Davis, California, United States of America

3. Genome Center, University of California, Davis, Davis, California, United States of America

KEYWORDS: Structure prediction, Rosetta Molecular Modeling Suite, Protein Symmetry.

Figure S1 Plot showing the C alpha distance deviation from the target crystal structure to their corresponding templates versus the template percent identity

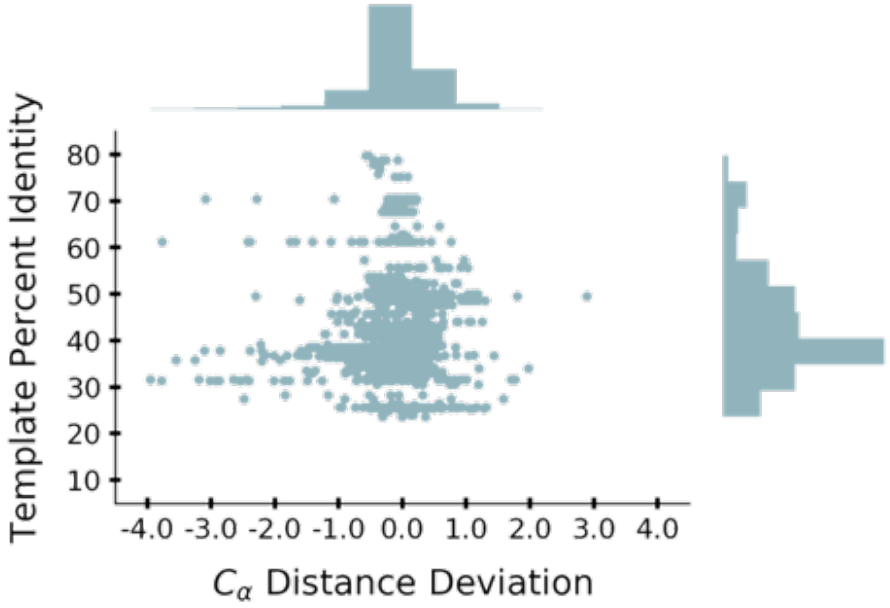
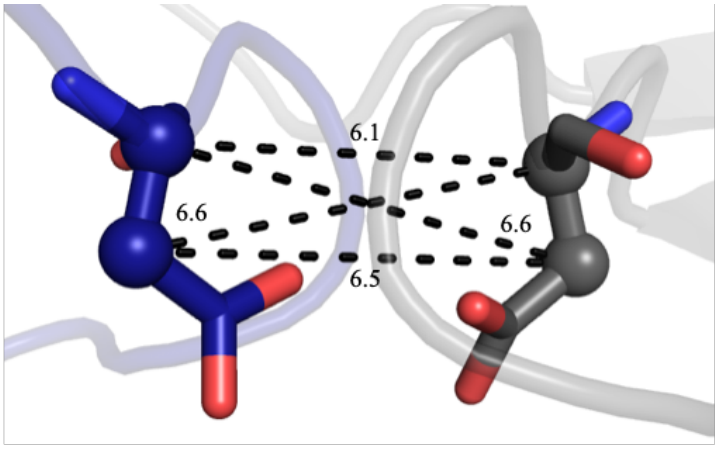


Figure S2 Visual representation as to how the distances used in the modeling were determined




```
AtomPair CA 25 CA 124 SCALARWEIGHTEDFUNC 1 HARMONIC 6.08 0.6
AtomPair CB 25 CB 124 SCALARWEIGHTEDFUNC 1 HARMONIC 6.64 0.7
AtomPair CA 25 CB 124 SCALARWEIGHTEDFUNC 1 HARMONIC 6.57 0.7
AtomPair CA 124 CB 25 SCALARWEIGHTEDFUNC 1 HARMONIC 6.50 0.7
```

Table S1. The target pdb codes that had entries on the mechanism and catalytic atlas database

catalytic site atlas (reference pdb: 1ctt)	pdb: 1ctu
His 102A	His 102A
Glu 104A	Glu 104A
Cys 129A	Cys 129A
Cys 132A	Cys 132A
catalytic site atlas (reference pdb: 1dbt)	pdb: 1dqx
Lys 33A	Lys 59A
Asp 60A	Asp 91A
Lys 62A	Lys 93A
Thr 123A	Ser 154A
Asp 65B	Asp 96B
catalytic site atlas (reference pdb: 1qin)	pdb: 1qin
Gln 33A	Gln 33A
Glu 99A	Glu 99A
His 126B	His 126B
Glu 172B	Glu 172B

catalytic site atlas (reference pdb: 1a30)	pdb: 2o4p
Asp 25A	Asp 25A
Asp 25B	Asp 25B
catalytic site atlas (reference pdb: 1rr6)	pdb: 2q7o
His 86A	His 86A
Ser 220A	Ser220A
Asn 243A	Asn 243A
catalytic site atlas (reference pdb: 1rr6)	pdb: 3bgs
His 86A	His 86A
Ser 220A	Ser220A
Asn 243A	Asn 243A
catalytic site atlas (reference pdb: 1mcz)	pdb: 3fzn
Glu 28E	Glu 28A
His 70E	His 70A

catalytic site atlas (reference pdb: 5icd)	pdb: 4bnp
Tyr 160A	Tyr 160A
Lys 230A	Lys 230B
Asp 283A	Asp 283B
Asp 307A	Asp 307A
catalytic site atlas (reference pdb: 2fua)	pdb: 4fua
Glu 73A	Glu 73A
His 92A	His 92A
His 94A	His 94A
His 155A	His 155A
Tyr 113A	Tyr 113B
catalytic site atlas (reference pdb: 1trk)	pdb: 5nd5
His 263A	His 315A
Glu 418B	Glu 463B

Table S2. The target pdb codes with the templates that were used for modeling

Target	Templates used									
1ctu	1ux0	1ux1	1zab	2d30	3dmo	3ijf	3r2n	4eg2	4wif	4wig
1dqx	1eix	2eaw	2jgy	2qcc	2qcf	3bk0	3bvj	3ewx	3ldv	3tr2
1nki	1npb	1r9c	2p7q	4ir0	4jh1	4nay	4nb1	5f6q	5v3d	5vb0
1ovm	1pvd	1pyd	1qpb	1zpd	2vbf	2vjy	2vk1	2vk8	2w93	4cok
1qin	1f9z	2c21	4hc5	4mtq	5d7z					
1a59	3hwk	3o8j	3tqg	1iom	1ixe					
2o4p	2p3d	3ggu	3mws	3t3c	3ttp	3u7s	3ucb	4z4x	5b18	5t2e
2q7o	1tcu	1vmk	2p4s	3khs	3la8	4lna	4m1e	5cxq	5cxs	5ifk
2vbg	1ovm	1pvd	1pyd	1qpb	2vjy	2vk1	2vk8	2w93	4cok	5euj
3bgs	1tcu	1vmk	2p4s	3khs	3la8	4lna	4m1e	5cxq	5cxs	5ifk
3fuc	1tcu	2p4s	3khs	3la8	4lna	4m1e	4ns1	5cxq	5cxs	5ifk
3fzn	1jsc	2ag0	2pan	2uz1	3d7k	3iae	4k9q	4q9d	4qpz	5fem
3mng	1tp9	2wfc	2xhf	3uma	4f82	5k1g	5k2i			
4bnp	1hqs	1x0l	1xkd	2d4v	2dht	2iv0	3ah3	3asj	3dms	
4fua	1jdi	1k0w	2fk5	2irp	2opi	3ocr	4c24	4c25	4m6r	4xxf
4hgo	1j8d	2r8x	3i6b	3ij5	3mmz	3mn1	3n07	3n1u	4nav	4um5
5nd5	1ay0	1gpu	1itz	1qgd	1r9j	1tka	2e6k	2r8o	3hyl	5hht

Table S3. The target pdb codes with the number of chains the biological unit has and the corresponding top template used for modeling with the number of chains the biological unit the template has

target	bio chains	top template	pid	bio chains
1ctu	2	4eg2	48.6	2
1dqx	2	2qcc	52.3	2
1nki	2	5v3d	67.7	2
1ovm	4	2vbf	41.4	2
1qin	2	4mtq	42.1	2
1a59	2	3hwk	59.8	2
2o4p	2	5t2e	79.8	2
2q7o	3	2p4s	55.8	3
2vbg	2	1ovm	41.4	4
3bgs	3	2p4s	55.8	3
3fuc	3	2p4s	55.6	3
3fzn	4	4q9d	44.5	4
3mng	2	2wfc	64.6	2
4bnp	2	3dms	75.2	2
4fua	4	4c24	43.4	4
4hgo	4	3mn1	43.1	4
5nd5	2	1itz	68.5	2

Table S4. Details about the sampling of the conformational library and constraints.
The atoms in the ligands were allowed to sample all conformer space with the exception that contained the cofactor thiamine pyrophosphate. The pyrophosphate functional group atoms were frozen. The conformation library of the ligands generated was used to dock into the models.

pdb code ligand	Description for conformational library
1ovm, 2vbg, 3fzn, 5nd5	The pyrophosphate region was frozen
4hgo	The vanadium was changed to a phosphorous atom because Rosetta does not have that atom type in it's library

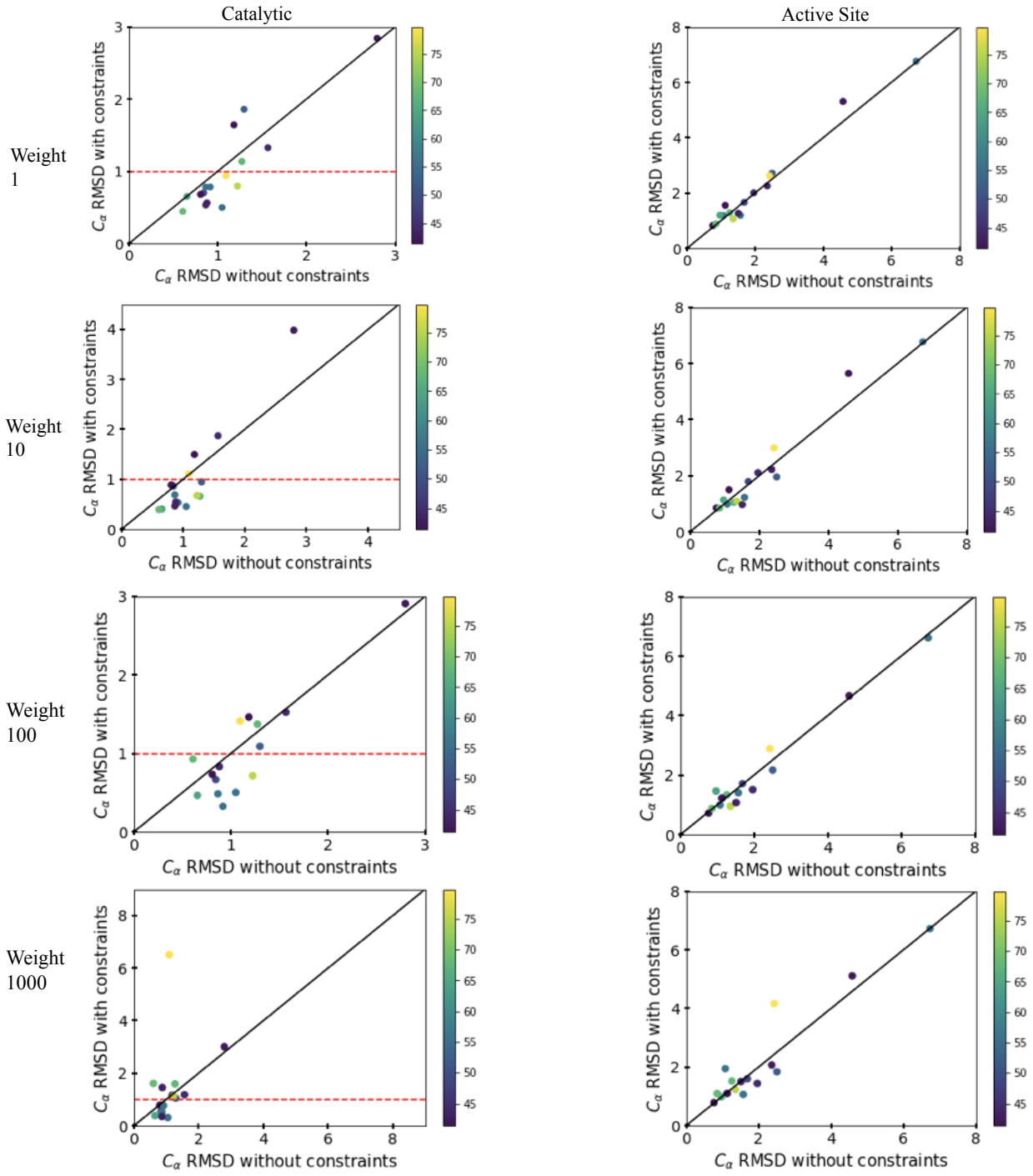
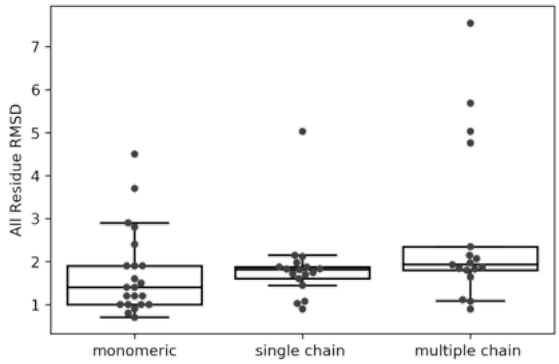
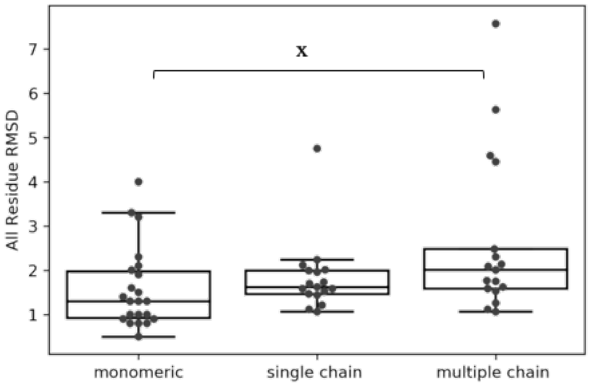


Figure S3 The analysis performed on the models with the weights of 1, 10, 100, 1000

(A) Models without catalytic constraints
All Residue RMSD



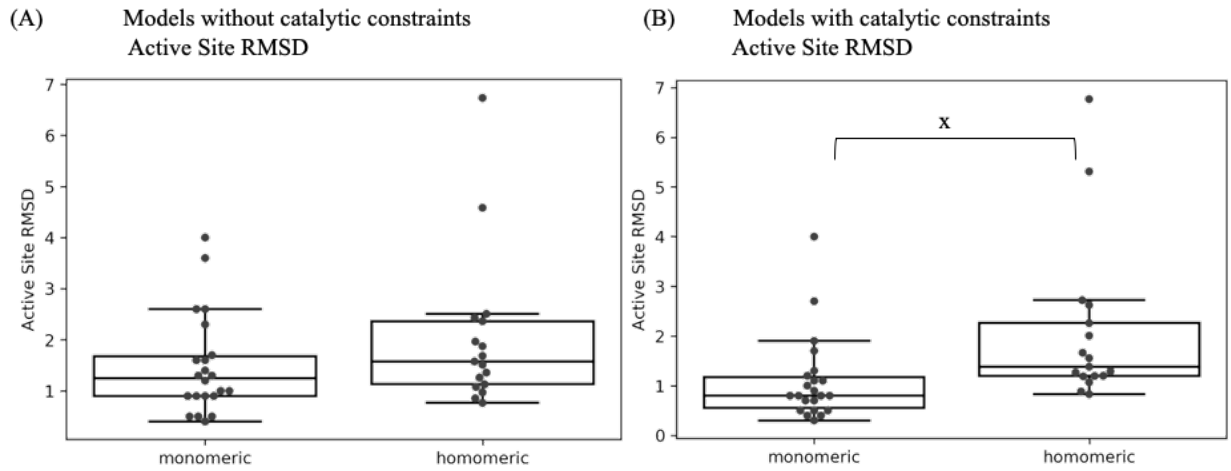
(B) Models with catalytic constraints
All Residue RMSD



(C)

All residues		monomeric	single chain	multiple chain
Models without catalytic constraints	average	1.72	1.86	2.67
	standard deviation	0.98	0.92	1.87
Models with catalytic constraints	average	1.58	1.83	2.64
	standard deviation	0.91	0.83	1.82

Figure S4. Boxplots comparing the RMSD of all residues between the monomeric and homomeric benchmarks. Each point in (A) and (B) represents the average RMSD of all residues for the lowest five models. The single chain represents the RMSD of only the single chain between the crystal structures or models within the homomeric benchmark. The multiple chains represent the RMSD analysis between all chains of the crystal structures or models within the homomeric benchmark. The X correspond to a p-value < 0.05. (A) Analysis between the target crystal structure to models without catalytic constraints (B) Analysis between the target crystal structure to models with catalytic constraints (C) A table of the average and standard deviation for the boxplots in A - B



(C)

Active site		monomeric	homomeric
Models without catalytic constraints	average	1.48	2.04
	standard deviation	0.97	1.51
Models with catalytic constraints	average	1.09	2.07
	standard deviation	0.85	1.62

Figure S5. Boxplots comparing the RMSD of all residues between the monomeric and homomeric benchmarks. Each point in (A) and (B) represents the average RMSD of active site residues for the lowest five models. The single chain represents the RMSD of only the single chain between the crystal structures or models within the homomeric benchmark. The multiple chains represents the RMSD analysis between all chains of the crystal structures or models within the homomeric benchmark. The X correspond to a p-value < 0.05. (A) Analysis between the target crystal structure to models without catalytic constraints (B) Analysis between the target crystal structure to models with catalytic constraints (C) A table of the average and standard deviation for the boxplots in A – B

Table S5. Statistical analysis of the all residue and active site RMSD using a two tailed t-test (A) all residues (B) active site residues

(A)

All residues		Monomeric-single	Monomeric-multiple	Single-multiple
Models without catalytic constraints	p-value	0.64	0.069	0.075
Models with catalytic constraints	p-value	0.39	0.039	0.081

(B)

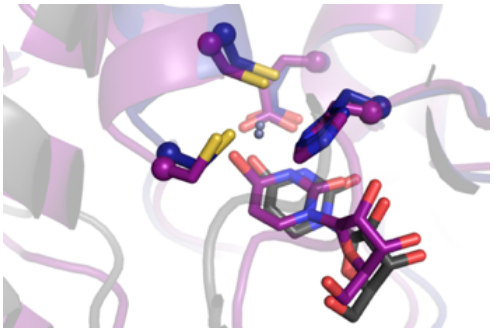
Active site		Monomeric-homomeric
Models without catalytic constraints	p-value	0.20
Models with catalytic constraints	p-value	0.033

Table S6. Number of conformations

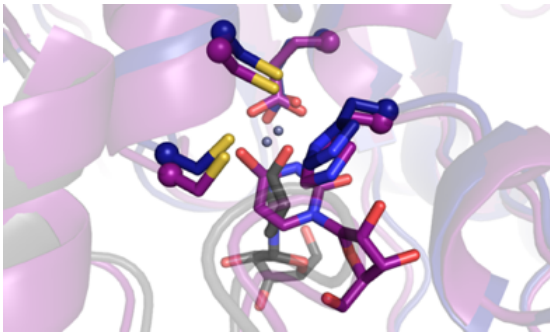
pdb	number of rotamers
1nki	1
3mng	3
1ovm	4
5nd5	4
1a59	10
2vbg	36
4bnp	81
1ctu	165
4fua	243
3fuc	319
3fzn	322
1dqx	1169
3bgs	1455
2q7o	3763
2o4p	20505
4hgo	44289
1qin	54278

Figure S6 The crystal structure is depicted in a deep purple color with the models overlaid on top. The models are depicted in blue and dark gray. The crystal structure ligand is in a deep purple color and the ligand docked into the model is in a dark gray.

(A) 1ctu

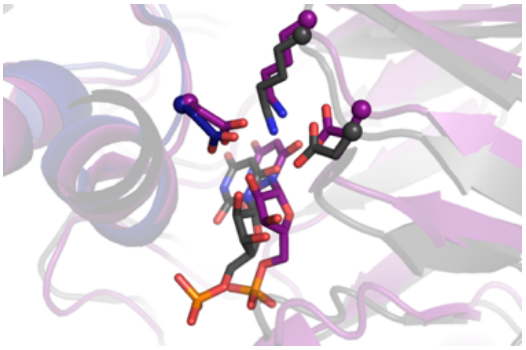


Ligand RMSD without constraints:
1.91

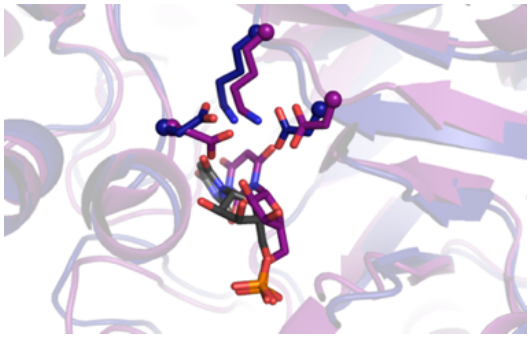


Ligand RMSD with constraints:
5.65

(B) 1dqx

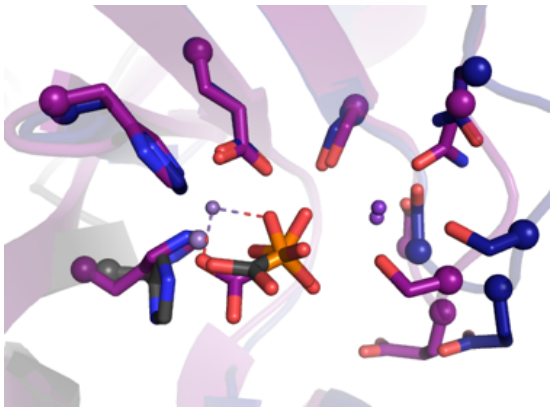


Ligand RMSD without constraints:
3.15

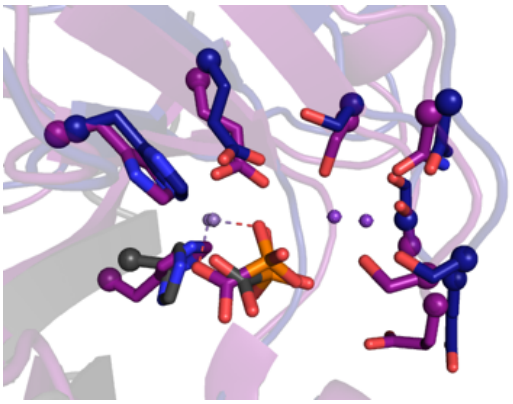


Ligand RMSD with constraints:
5.53

(C) 1nki

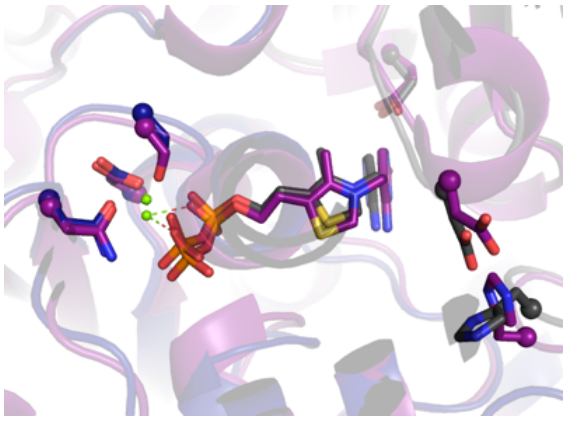


Ligand RMSD without constraints:
2.47

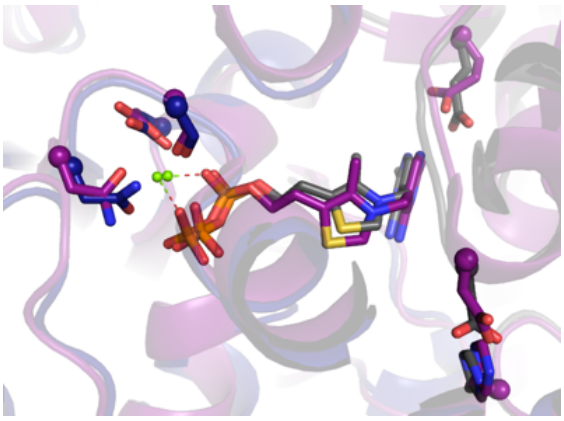


Ligand RMSD with constraints:
2.34

(D) 1ovm

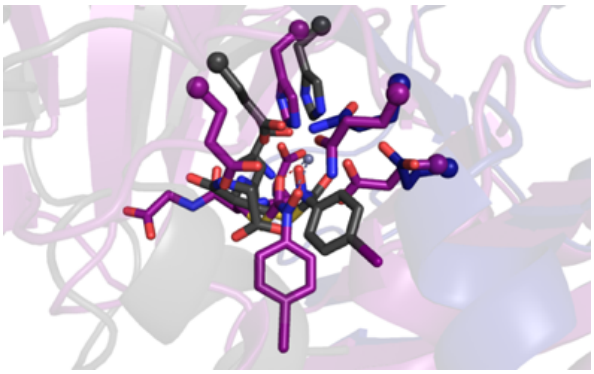


Ligand RMSD without constraints:
1.01

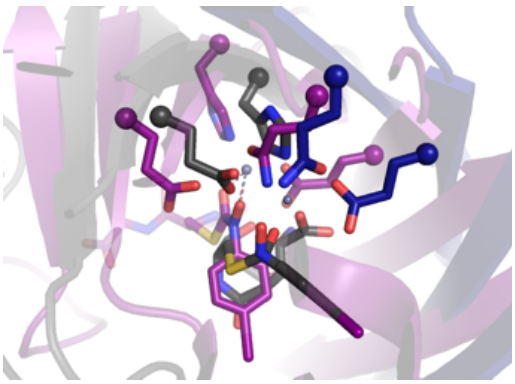


Ligand RMSD with constraints:
1.16

(E) 1qin

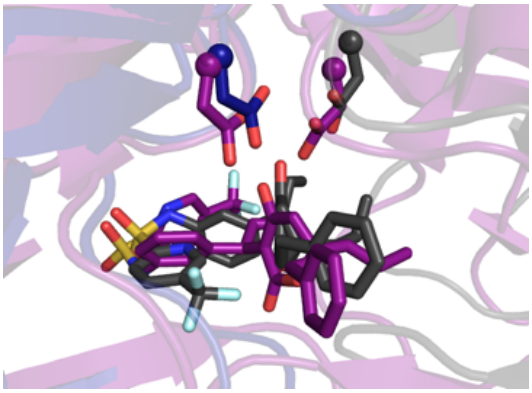


Ligand RMSD without constraints:
6.68

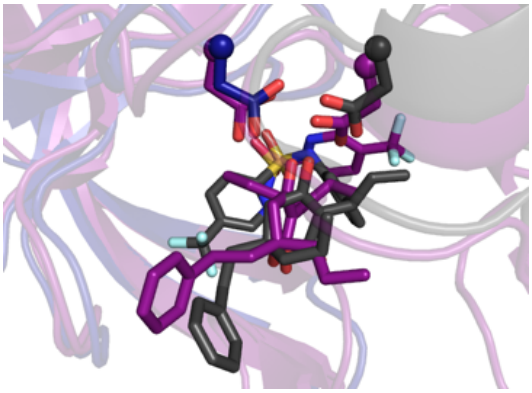


Ligand RMSD with constraints:
8.05

(F) 2o4p

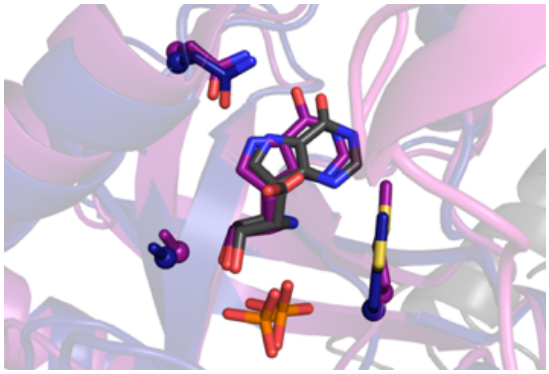


Ligand RMSD without constraints:
10.15

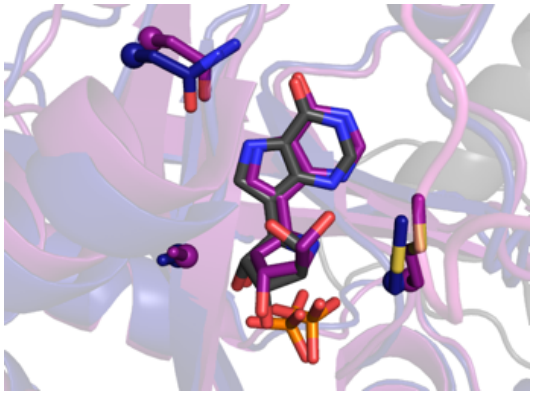


Ligand RMSD with constraints:
9.91

(G) 2q7o

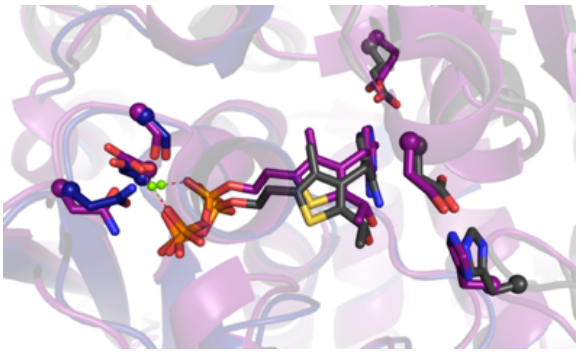


Ligand RMSD without constraints:
1.37

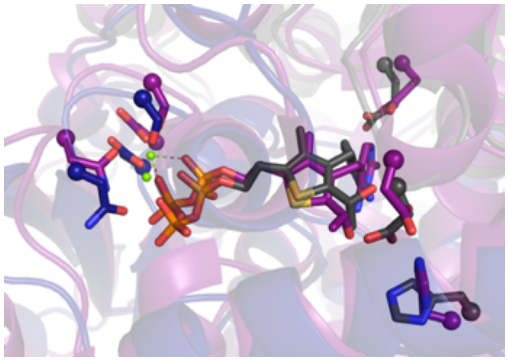


Ligand RMSD with constraints:
1.36

(H) 2vbg

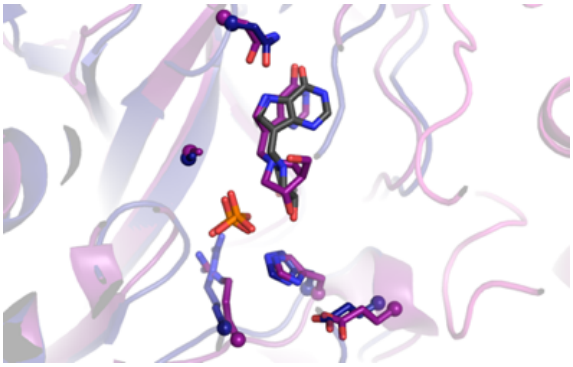


Ligand RMSD without constraints:
1.14

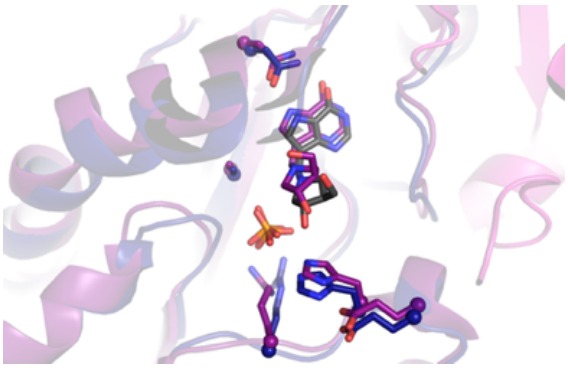


Ligand RMSD with constraints:
1.58

(I) 3bgs

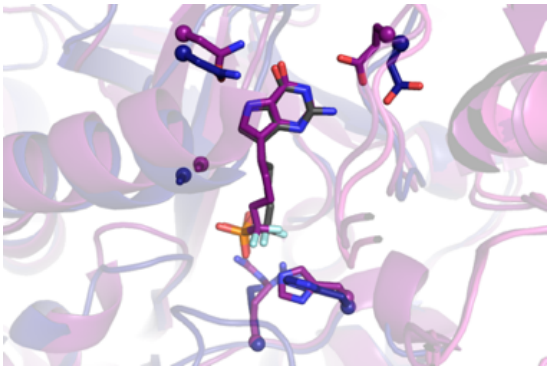


Ligand RMSD without constraints:
3.27

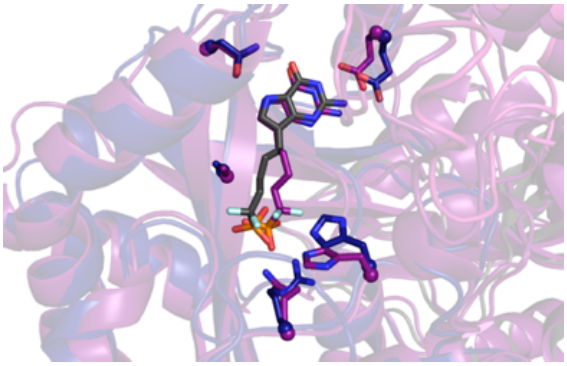


Ligand RMSD with constraints:
1.20

(J) 3fuc

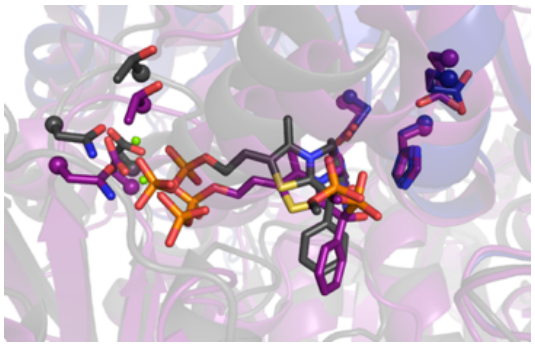


Ligand RMSD without constraints:
2.28

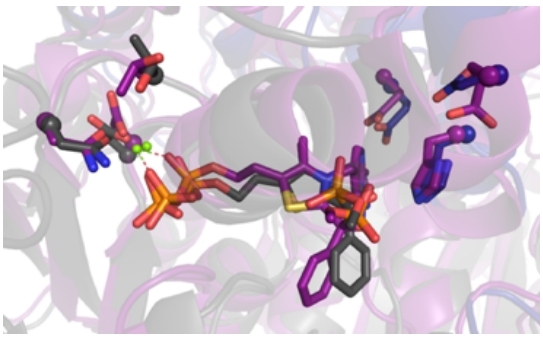


Ligand RMSD with constraints:
1.62

(K) 3fzn

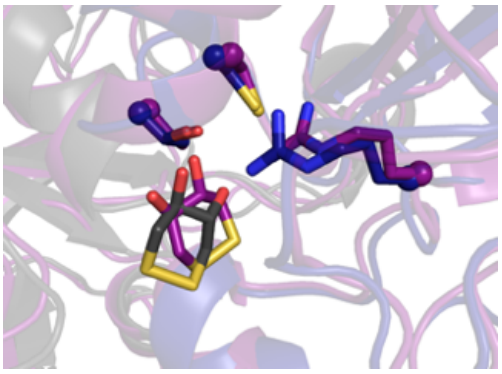


Ligand RMSD without constraints:
3.05

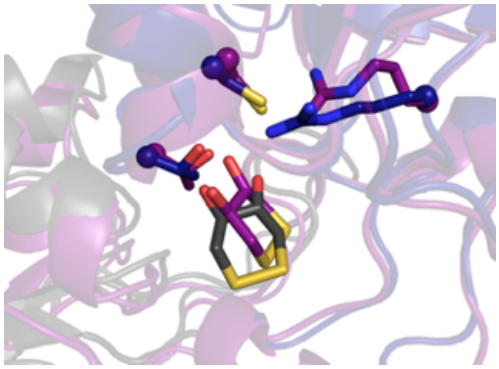


Ligand RMSD with constraints:
2.95

(L) 3mng

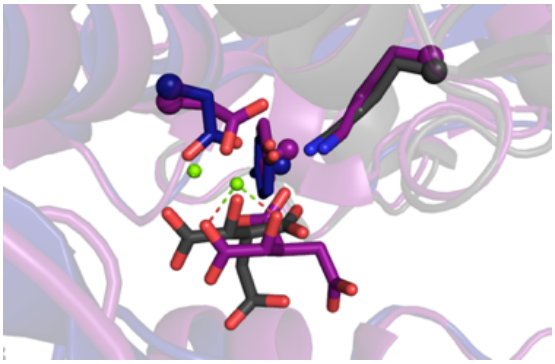


Ligand RMSD without constraints:
2.17

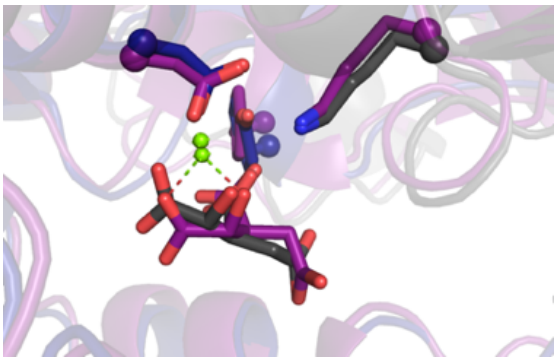


Ligand RMSD with constraints:
1.90

(M) 4bnp

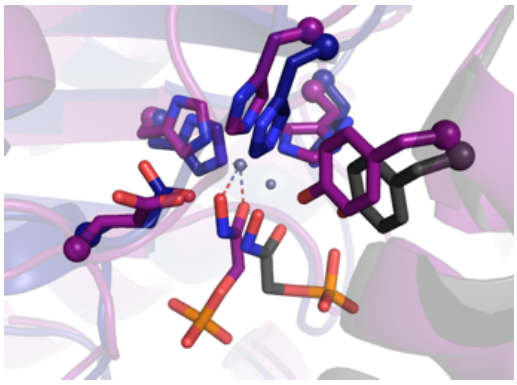


Ligand RMSD without constraints:
2.69

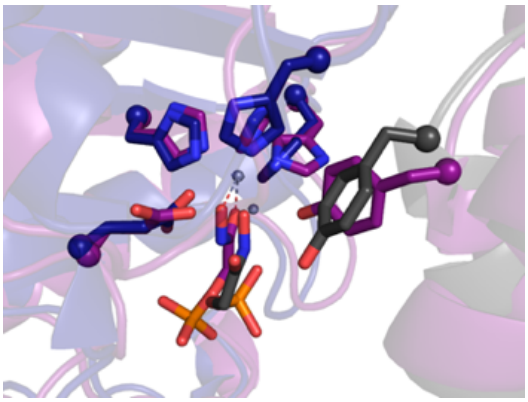


Ligand RMSD with constraints:
1.76

(N) 4fua

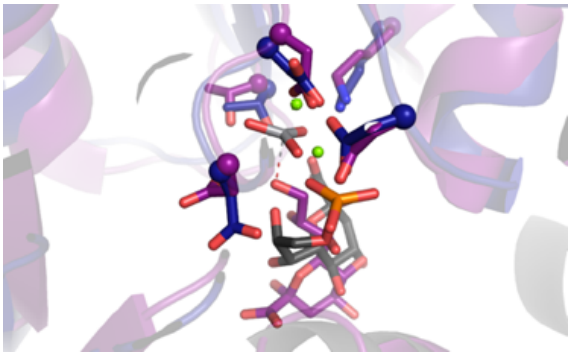


Ligand RMSD without constraints:
3.71

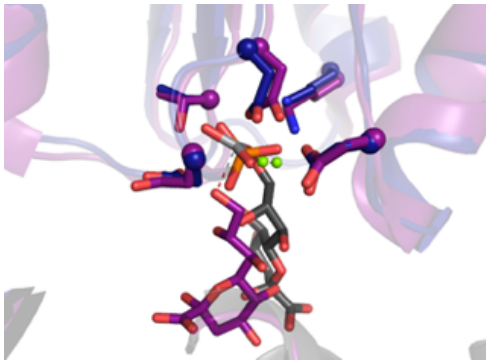


Ligand RMSD with constraints:
3.96

(O) 4hgo

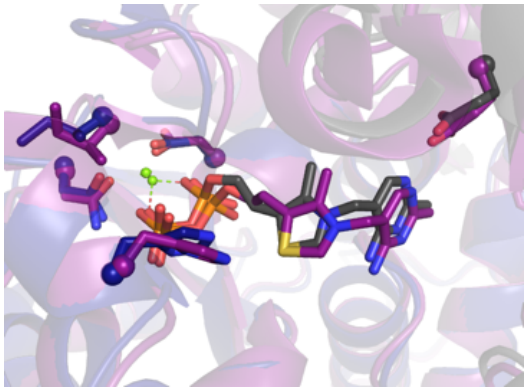


Ligand RMSD without constraints:
6.10

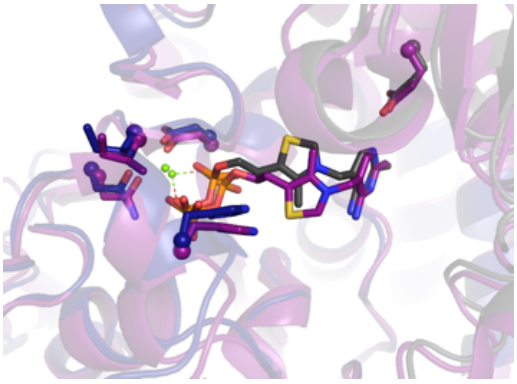


Ligand RMSD with constraints:
5.60

(P) 5nd5

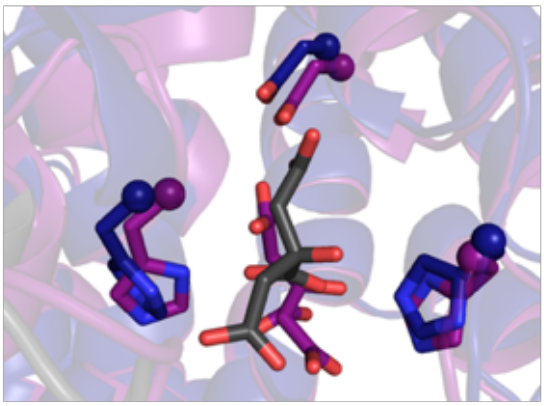


Ligand RMSD without constraints:
1.34

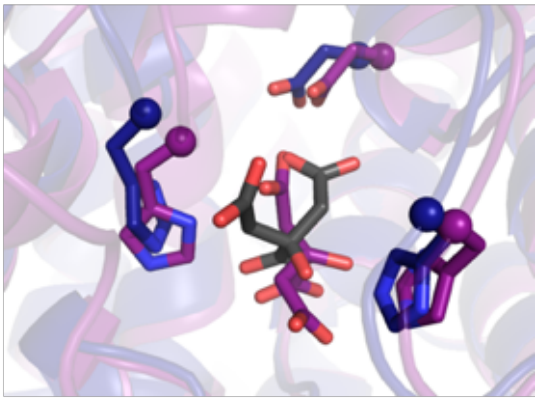


Ligand RMSD with constraints:
1.51

(Q) 1a59



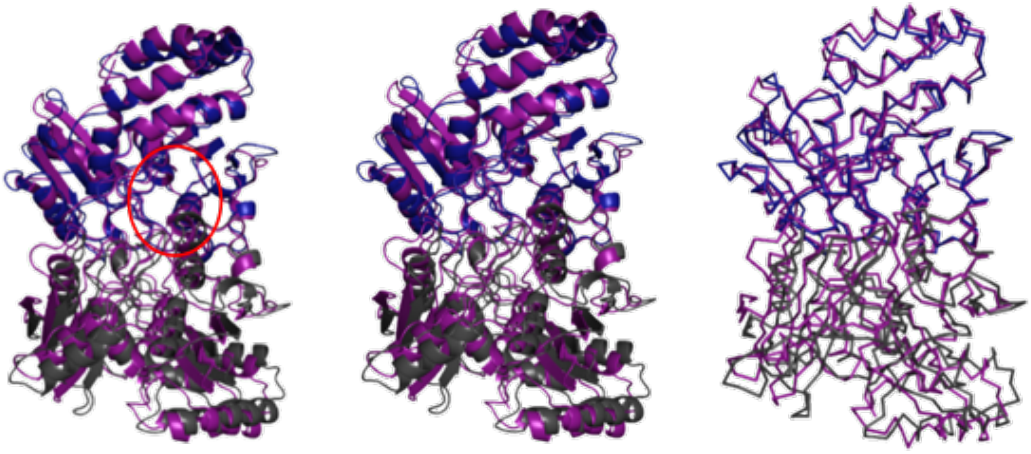
Ligand RMSD without constraints:
3.37



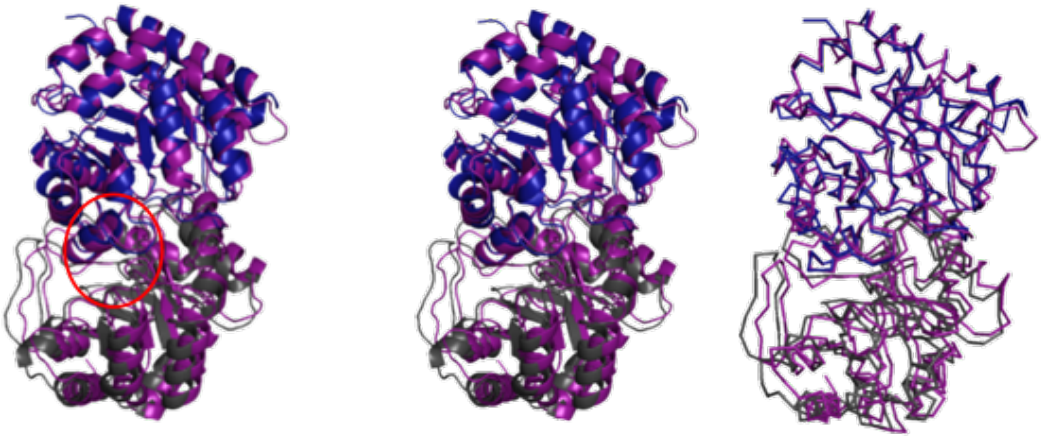
Ligand RMSD with constraints:
3.07

Figure S7 The area of the active site highlighted with a red circle on the first cartoon representation of protein. The crystal structure is depicted in a deep purple color with the models overlaid on top in cartoon and ribbon form. Depending on the number of chains, the models are depicted in blue (chain A), dark gray (chain B), light blue (chain C), and light gray (chain D).

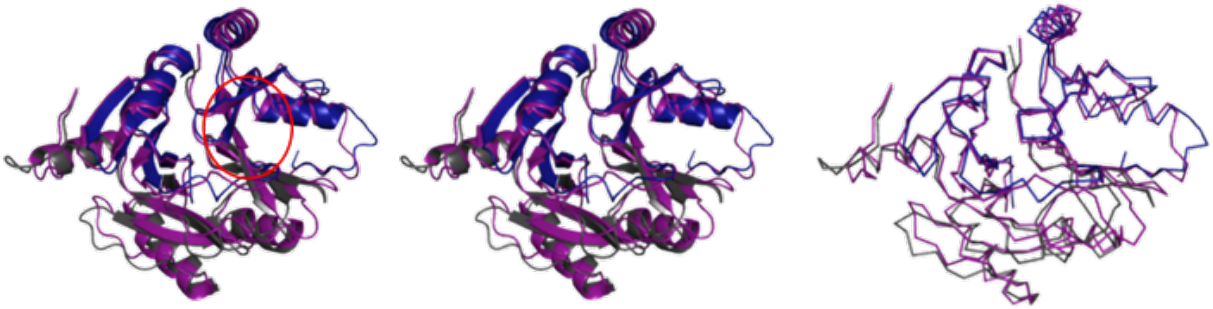
(a) 1ctu



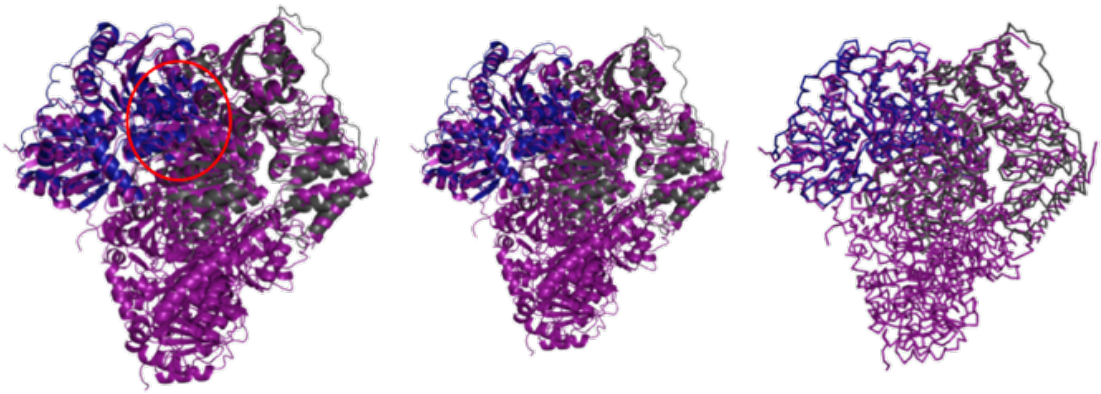
(b) 1dqx



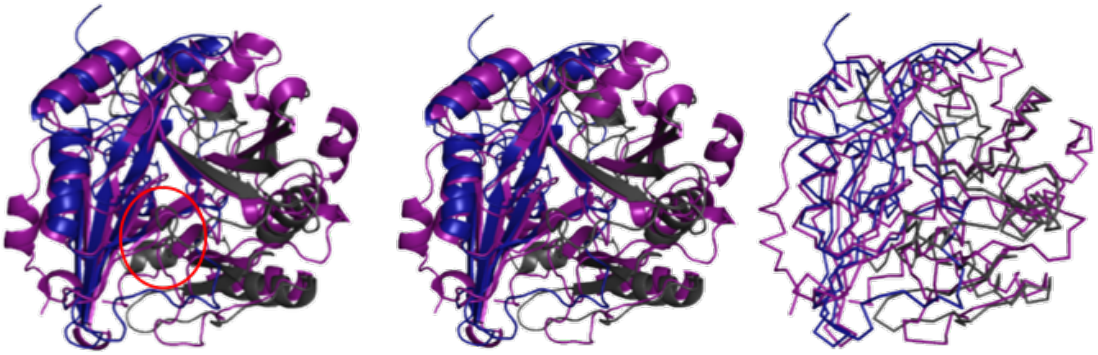
(c) 1nki



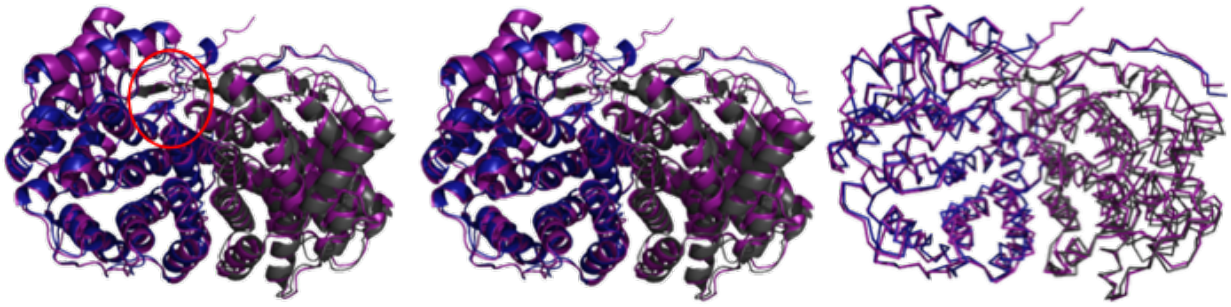
(d) 1ovm



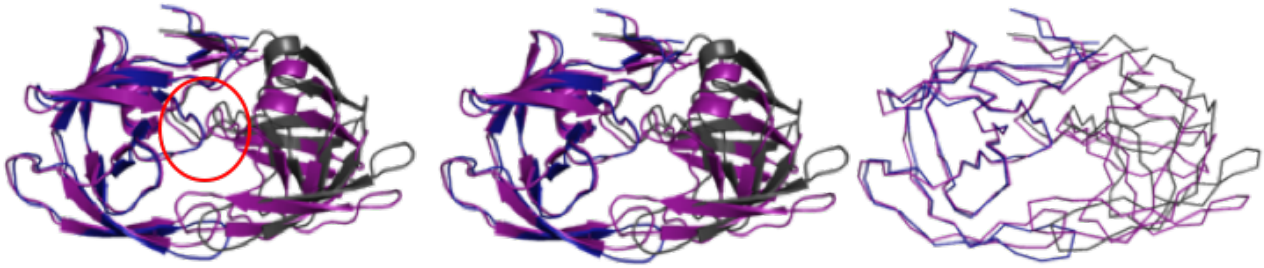
(e) 1qin



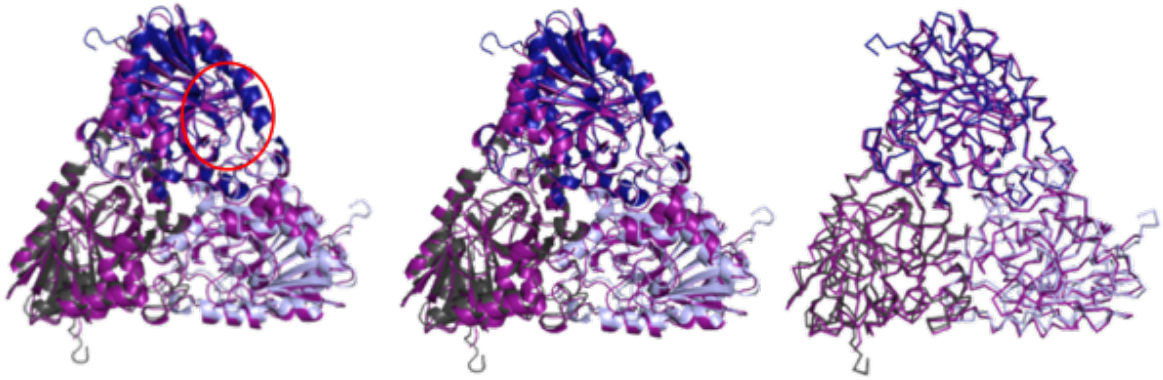
(f) 1a59



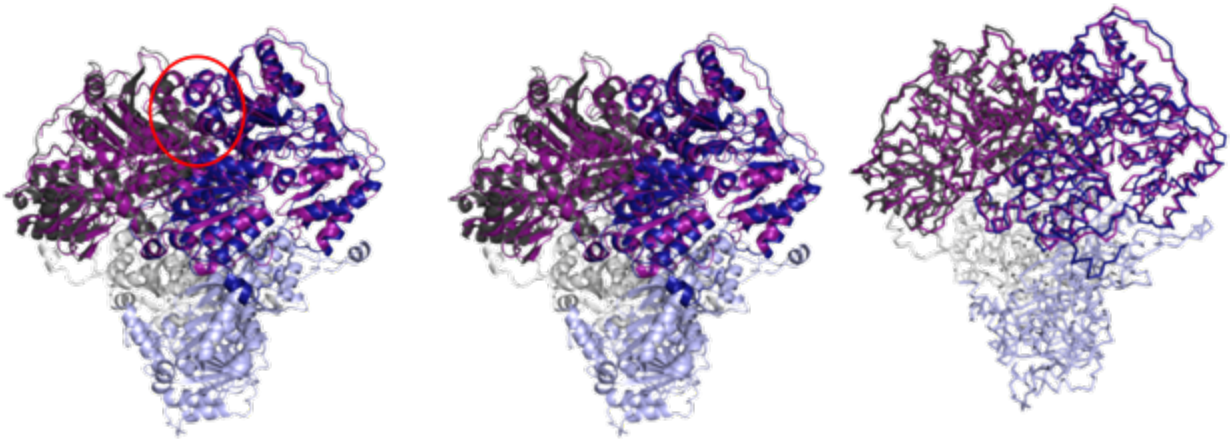
(g) 2o4p



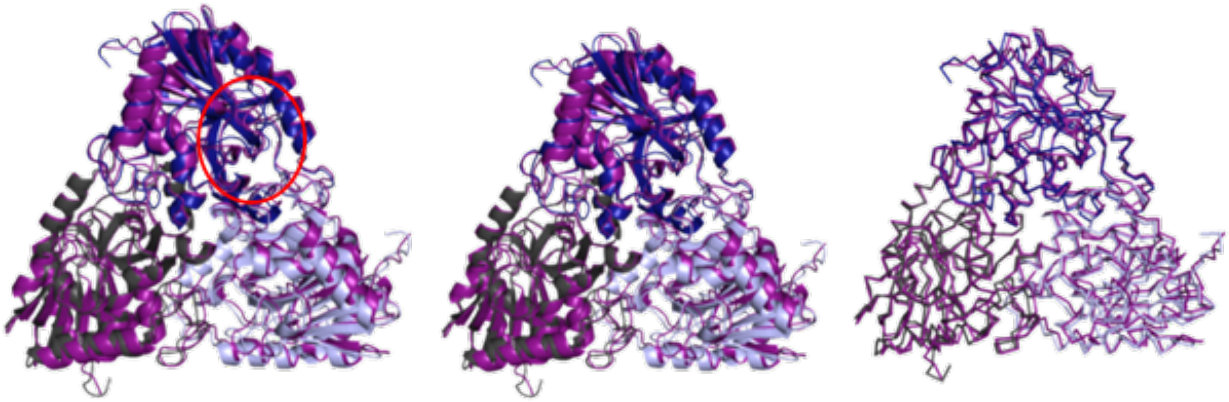
(h) 2q7o



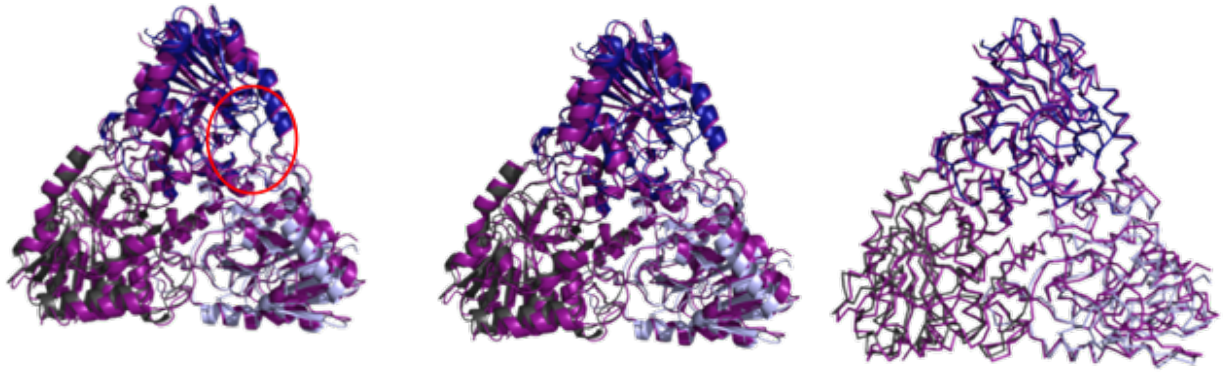
(i) 2vbg



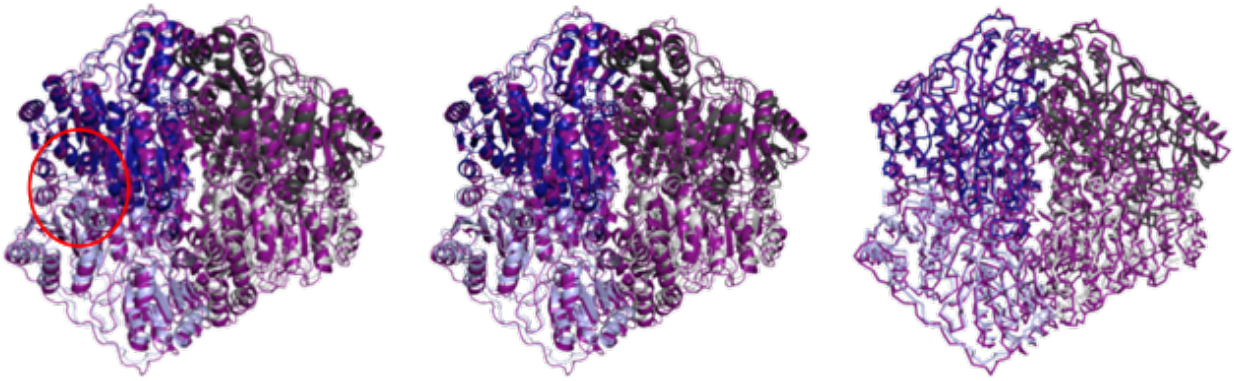
(j) 3bgs



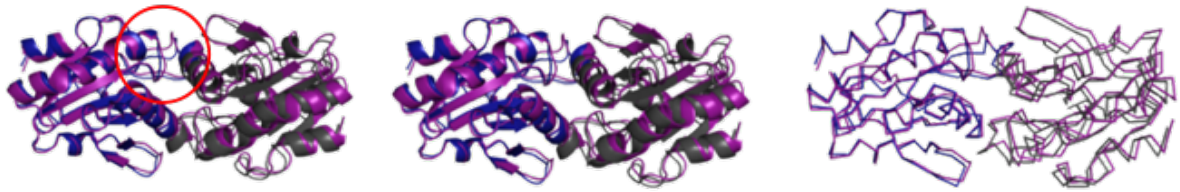
(k) 3fuc



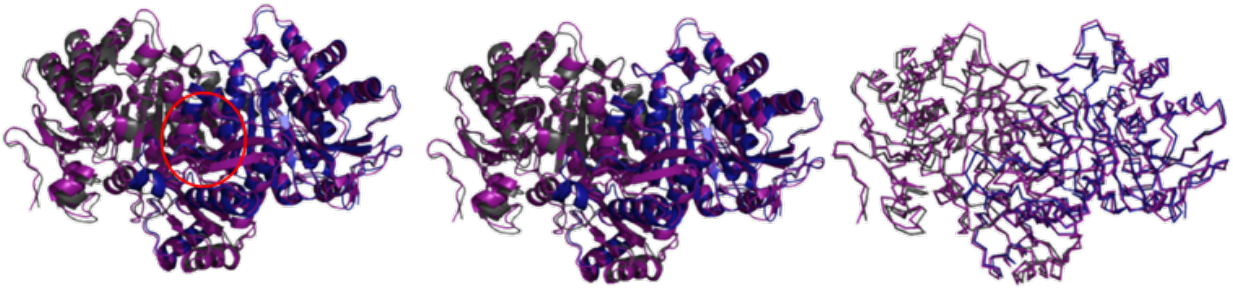
(l) 3fzn



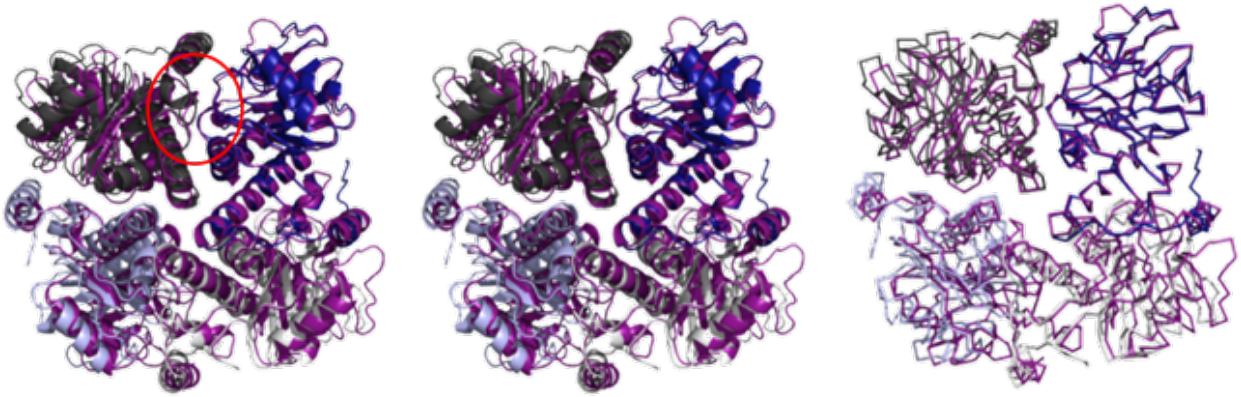
(m) 3mng



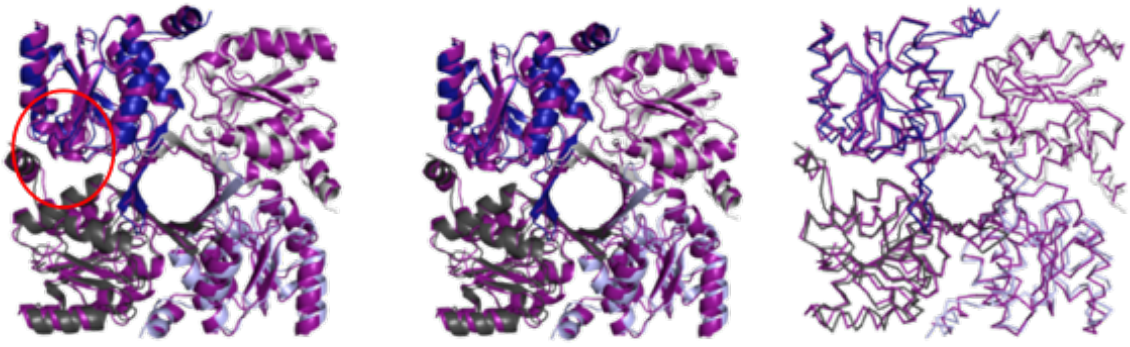
(n) 4bnp



(o) 4fua



(p) 4hgo



(q) 5nd5