STRUCTURAL
BIOLOGY

Volume 76 (2020)

Supporting information for article:

Development of a structure-analysis pipeline using multiple solvent crystal structures of barrier-to-autointegration factor

Sorabh Agarwal, Mychal Smith, Indhira De La Rosa, Kliment A. Verba, Paul Swartz, Miriam Segura-Totten and Carla Mattos

	Waters	*Total*	*%*		*Waters*	*Total*	*%*	*Diff*
Aqueous	174	174	100	**33% PDB**	172	172	100	0
Acetone	146	174	83.9	Acetone	114	172	66.3	17.6
Dioxane	115	174	66.1	Dioxane	90	172	52.3	13.8
DMSO	130	174	74.7	DMSO	94	172	54.7	20.1
Ethanol	115	174	66.1	Ethanol	92	172	53.5	12.6
Glycerol	110	174	63.2	Glycerol	85	172	49.4	13.8
Hexanediol	126	174	72.4	Hexanediol	93	172	54.1	18.3
Isobutanol	134	174	77.0	Isobutanol	93	172	54.1	22.9
Isopropanol	121	174	69.5	Isopropanol	90	172	52.3	17.2
Methanol	128	174	73.6	Methanol	100	172	58.1	15.4
RSR	151	174	86.8	RSR	103	172	59.9	26.9
tButanol	126	174	72.4	tButanol	96	172	55.8	16.6
TMAO	146	174	83.9	TMAO	113	172	65.7	18.2
Urea	139	174	79.9	Urea	103	172	59.9	20.0
Average	129.8	174	74.6	Average	97.4	172	56.6	18.0

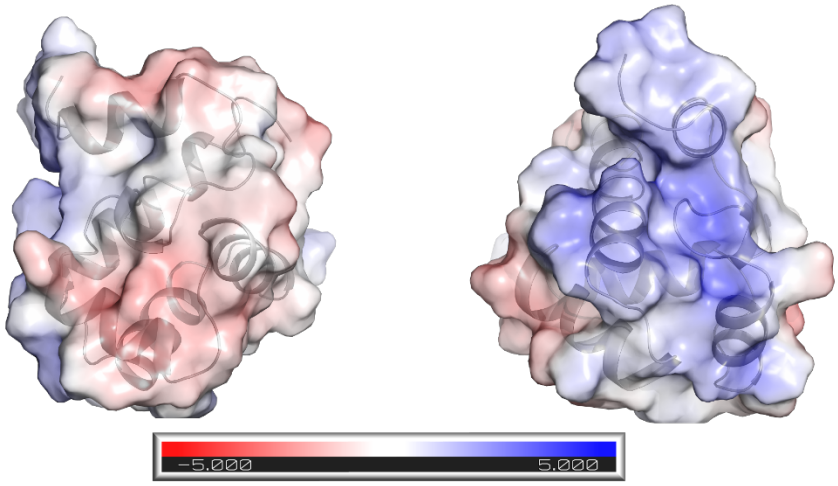
Supplementary Table S1. DRoP analysis showing conservation of waters from the aqueous model and the PDB derived water set. The left side of the table shows the fidelity of waters from the aqueous model for each solvent structure. The right side of the table with the 33% conserved water set shows less fidelity for waters in each solvent soaked structure. This analysis shows an average of 16.7% less conservation of waters across the organic sets.

PDB Files

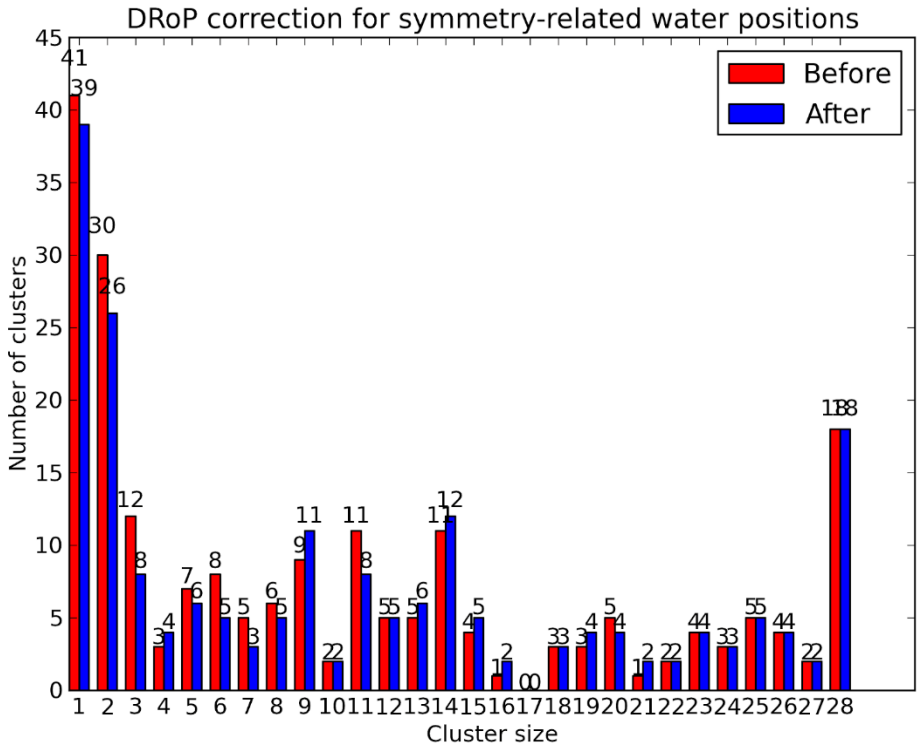
1	Fin1742DMSOv3ChainA.pdb	2	Fin1742DMSOv3ChainB.pdb	3	Fin1748AcetV2ChainA.pdb	4	Fin1748AcetV2ChainB.pdb	5	Fin1750GlycV3ChainA.pdb
6	Fin1750GlycV3ChainB.pdb	7	Fin1754IsopropV2ChainA.pdb	8	Fin1754IsopropV2ChainB.pdb	9	Fin1757tbutV2ChainA.pdb	10	Fin1757tbutV2ChainB.pdb
11	Fin1761DioV2ChainA.pdb	12	Fin1761DioV2ChainB.pdb	13	Fin1765MeOHV3ChainA.pdb	14	Fin1765MeOHV3ChainB.pdb	15	Fin1772EtOHv2ChainA.pdb
16	Fin1772EtOHv2ChainB.pdb	17	Fin1778RSRv3ChainA.pdb	18	Fin1778RSRv3ChainB.pdb	19	Fin1786IsoButV3ChainA.pdb	20	Fin1786IsoButV3ChainB.pdb
21	Fin1791TMAOv5ChainA.pdb	22	Fin1791TMAOv5ChainB.pdb	23	Fin1794UreaV2ChainA.pdb	24	Fin1794UreaV2ChainB.pdb	25	Fin1797s2AqueousV5ChainA.pdb
26	Fin1797s2AqueousV5ChainB.pdb	27	Final1775HexV2ChainA.pdb	28	Final1775HexV2ChainB.pdb				

Res Num	Average Position			RMSD	Files	#files
1	13.978	-3.346	19.119	0.628	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
2	37.458	10.792	25.875	0.981	13 15	(2)

Supplementary Table S2. DRoP output of conserved organics. The DRoP output above shows two identified organic binding sites along with detailed statistical information and which organic soak contributes waters to a selected organic site. The text output above is accompanied by a PDB file with the same sites. DRoP was able to identify 2 conserved sites. Site 1 (at BAF-emerin interface) showing a conserved ethanol in each of the chains and site 2 (at DNA binding surface) showing a methanol in 1 chain of 2 from the methanol soak and an ethanol in the same location from 1 chain of 2 in the ethanol soak. Figure output from DRoP program (1).



Supplementary Figure S1. Electrostatic surface of the BAF monomer. The electrostatics were contoured from −5 to +5 kT/e using APBS in PyMOL plotted on the solvent-accessible surface (2). On the left, is a view of the monomer consistent with the view of previous figures. Some acidic patches are detected on the surface of the protein. The right view shows a 90º rotated orientation of the molecule, as expected, a highly basic patch at the DNA binding interface is identified.



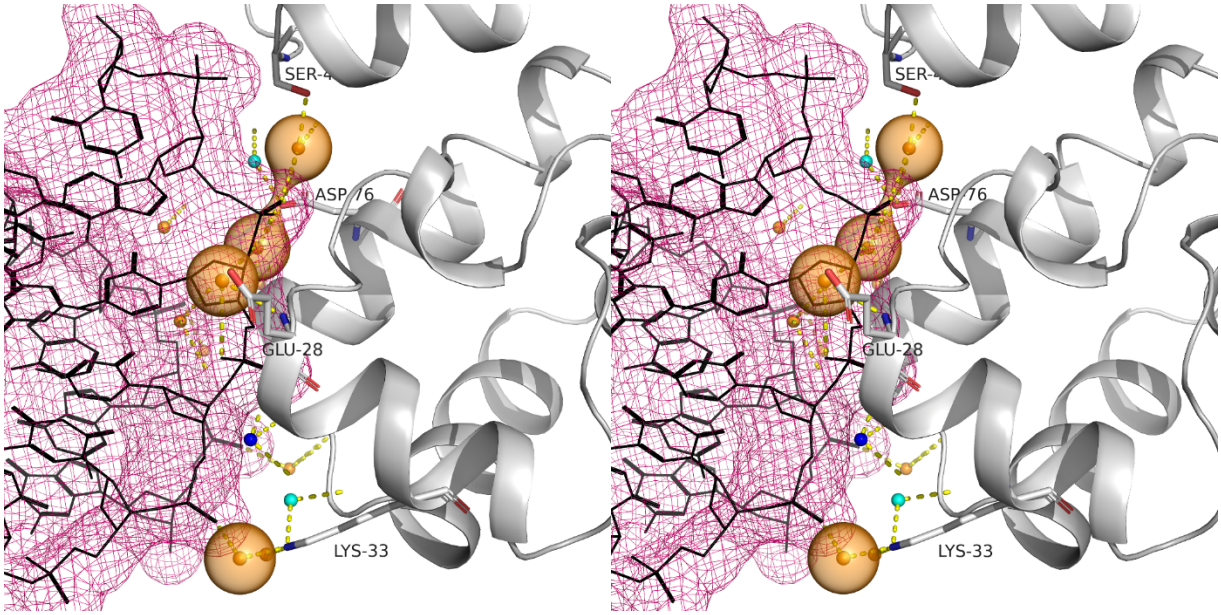
Supplementary Figure S2. DRoP output showing a histogram of symmetry corrected water molecules. Cluster size, on the x-axis represents the number of structures with a bound water molecule in a specific position. Number of clusters, shown on the y-axis, describes the number each cluster size. After running DRoP and moving symmetry related molecules, the number of clusters of each size changes, with red indicating clusters before and blue after. Most clusters remain constant and require no correction. As an example, an additional cluster of 21 water molecules is found after DRoP analysis. This is the result of symmetry related water(s) being moved and found to overlap with a cluster with less than 21 waters, bringing the cluster in question up to 21 conserved waters. Figure output from DRoP program (1).

PDB Files

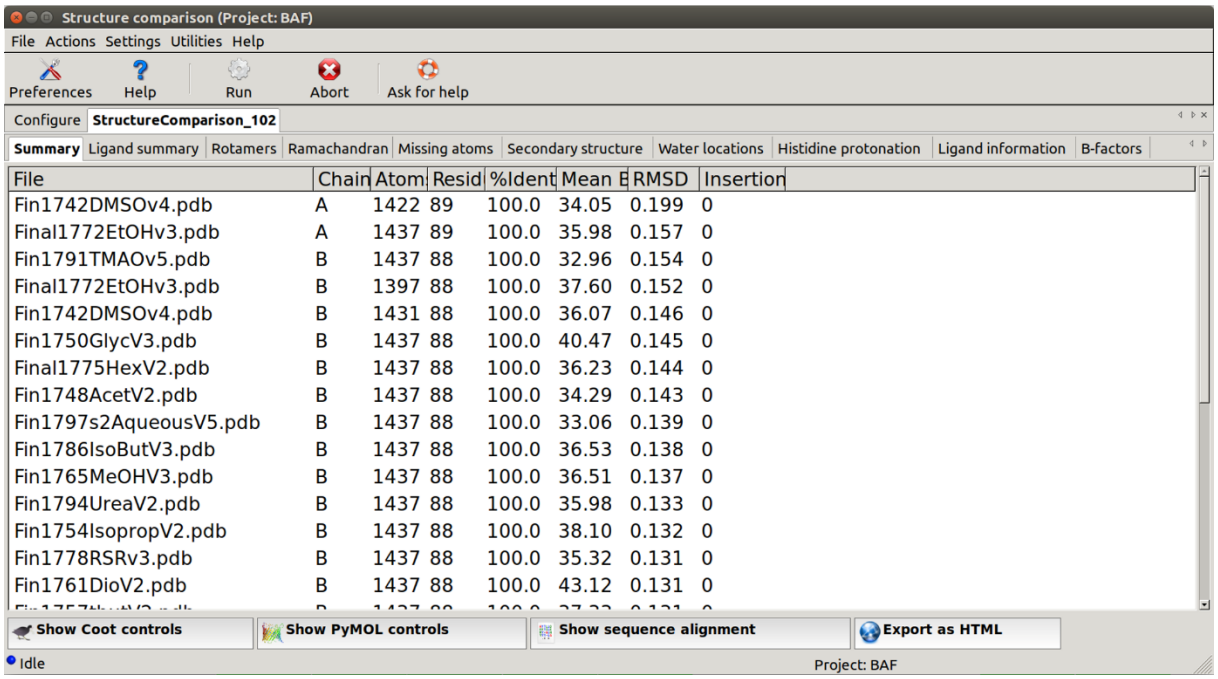
1	Fin1742DMSOv3ChainA.pdb	2	Fin1742DMSOv3ChainB.pdb	3	Fin1748AcetV2ChainA.pdb	4	Fin1748AcetV2ChainB.pdb	5	Fin1750GlycV3ChainA.pdb
6	Fin1750GlycV3ChainB.pdb	7	Fin1754IsopropV2ChainA.pdb	8	Fin1754IsopropV2ChainB.pdb	9	Fin1757tbutV2ChainA.pdb	10	Fin1757tbutV2ChainB.pdb
11	Fin1761DioV2ChainA.pdb	12	Fin1761DioV2ChainB.pdb	13	Fin1765MeOHV3ChainA.pdb	14	Fin1765MeOHV3ChainB.pdb	15	Fin1772EtOHv2ChainA.pdb
16	Fin1772EtOHv2ChainB.pdb	17	Fin1778RSRv3ChainA.pdb	18	Fin1778RSRv3ChainB.pdb	19	Fin1786IsoButV3ChainA.pdb	20	Fin1786IsoButV3ChainB.pdb
21	Fin1791TMAOv5ChainA.pdb	22	Fin1791TMAOv5ChainB.pdb	23	Fin1794UreaV2ChainA.pdb	24	Fin1794UreaV2ChainB.pdb	25	Fin1797s2AqueousV5ChainA.pdb
26	Fin1797s2AqueousV5ChainB.pdb	27	Final1775HexV2ChainA.pdb	28	Final1775HexV2ChainB.pdb				

Res Num	Average Position			RMSD	Avg B.	Sig. B.	Files	#files
1	36.740	11.174	16.739	0.104	24.450	5.014	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
2	37.138	10.475	14.101	0.114	28.646	5.117	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
3	33.180	9.419	11.553	0.117	23.754	5.237	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
4	33.328	1.993	12.583	0.127	25.454	5.083	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
5	19.194	-0.352	28.054	0.137	38.161	5.849	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
6	39.454	1.646	20.068	0.146	32.718	5.824	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
7	37.799	5.334	29.005	0.153	29.621	8.406	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
8	39.409	11.347	16.996	0.162	35.611	6.451	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
9	33.632	5.177	33.692	0.180	31.657	8.028	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
10	32.018	18.390	12.874	0.211	33.461	5.330	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
11	24.724	16.565	12.568	0.214	33.286	5.180	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
12	31.717	15.936	23.246	0.243	33.025	5.224	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
13	29.628	14.324	22.225	0.249	25.007	5.449	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
14	36.939	5.198	31.690	0.258	30.275	8.243	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
15	17.651	10.367	19.759	0.341	41.836	5.917	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
16	35.844	3.403	33.402	0.374	32.489	8.160	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
17	18.296	11.121	5.585	0.400	35.554	5.348	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
18	15.785	9.866	23.458	0.445	38.339	5.001	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(28)
19	37.284	3.150	7.037	0.220	42.748	4.873	1 2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(27)
20	17.889	11.386	17.104	0.314	37.985	5.772	1 2 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28	(27)

Supplementary Table S3. DRoP output of highly conserved waters. The selected DRoP output above shows the most highly conserved water sites along with detailed statistical information and which organic soak contributes waters to a selected conserved site. Waters are moved from symmetry related positions to a position closest to the protein surface and the text output above is accompanied by a PDB file with the same sites. Sites are arranged by the number of conserved waters and then for those sites with the same number of contributors, ranked by RMSD derived from how tightly waters are arranged around a site (1). With this information we were able to identify highly conserved water networks. Figure output from DRoP program (1).



Supplementary Figure S3. Identification and analysis of BAF-DNA interface waters. Shown here in stereoview is the putative DNA molecule bound to the aqueous BAF structure with conserved interface waters extracted from DRoP. The accessible surface area of the DNA was determined, shown in pink mesh (1.4 Å probe). Waters outside the accessible surface area, within 6Å of the superimposed DNA molecule were extracted as “interface waters”. This resulted in 13 waters. 4 weakly conserved waters were identified that mediate direct interactions between BAF (residues shown in sticks and labeled) and the putative DNA.



Supplementary Figure S4. Phenix structure comparison tool output. Shown here the structures are arranged by decreasing RMSD. The structure comparison tool allowed comparison of BAF structures in relation to a wide number of protein metrics including rotamers, Ramachandran angles, secondary structure and B-factors. Most differences between BAF datasets were minor, however the B-factor differences were found to be significant, warranting normalization and further investigation. Figure output from phenix structure comparison tool (3).

Supplementary References

1. Kearney, B.M., Schwabe, M., Marcus, K., Roberts, D.M., Dechene, M., Swartz, P. and Mattos, C. (2019) DRoP: Automated detection of conserved solvent binding sites on proteins. *Proteins*.
2. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, **98**, 10037-10041.
3. Moriarty, N.W., Liebschner, D., Klei, H.E., Echols, N., Afonine, P.V., Headd, J.J., Poon, B.K. and Adams, P.D. (2018) Interactive comparison and remediation of collections of macromolecular structures. *Protein Sci*, **27**, 182-194.