Table of Contents

Supplementary Methods: A short mathematical description of the wRMSD technique. (See also: Damm & Carlson. *Biophysical J.* 2000;49:457-466)

Given two proteins, **X** and **Y**, the two PDB files are parsed with Biopython 1.42 (http://biopython.org/) to compile a list of resolved residues in each crystal structure. Needle (EMBOSS) is used to align the sequences of **X** and **Y**. From the alignment, pairs of residues are matched and used in the overlay process. The superposition is based on paired Cα, but the code can easily be modified to incorporate more atoms if the user wishes.

First, the centers of mass of both proteins are placed at the origin, and a standard RMSD fit is used to give a rough, initial orientation for the overlay. Without this first step, the proteins would be too far apart and all weights would be zero. **X** and **Y** have *n* residues paired, and we calculate a Gaussian-weighting factor (w_n) for each pair, based on the distance between them.

$$w_n = e^{-(d_n)^2/c} \quad (1)$$

where c is an arbitrary scaling factor and d_n is determined as

$$d_n = \left((y_{nx} - x'_{nx})^2 + \left(y_{ny} - x'_{ny}\right)^2 + (y_{nz} - x'_{nz})^2\right)^{\frac{1}{2}} \quad (2)$$

The scaling factor, c, was set to the RMSD of the initial sRMSD the weights need to have sufficient power to overcome the initial differences in the superposition. When the weights were calculated for alignments of other programs a scaling factor of 5 was used for consistency.

A wRMSD fit is an iterative process: after a rotation is applied to a protein, the distances between the residues change, which in turn changes the weights, which requires a recalculation. Convergence is straightforward. Each iteration starts by placing the Gaussian-weighted center of mass (*wCM*) of each protein at the origin.

$$wCM_x = \frac{\sum_n w_n m_n x_n}{\sum_n m_n} \quad \text{and} \quad wCM_y = \frac{\sum_n w_n m_n y_n}{\sum_n m_n} \quad (3)$$

Weighting terms are used in the RMSD fit by simply incorporating them to the 3x3 covariance matrix (r_{ij}).

$$r_{ij} = \sum_n w_n y_{ni} x_{nj} \quad (4)$$

At this point, the rotation of the protein **X** onto **Y** is determined via the eigenvalues and eigenvectors of the square of the covariance matrix, as is standard practice. Rather than minimizing the sum of d_n^2, as is done in a standard RMSD fit, a wRMSD fit minimizes the sum of $w_n d_n^2$.

The goodness of fit can be measured in a sum of all weights. The maximum value occurs when all weighs are 1.0 and the sum is *n* (all atom pairs are perfectly overlaid). We write the sum of all weights (wSUM), normalized for the number of paired residues, as

$$wSUM = \frac{1}{n} \sum_n w_n \quad (5)$$

Supplementary Table 1: The complete references for crystal structures used in the study.

PDB ID	Title	Primary Citation Author	Pub. Year	Journal Name	Volume	First Page	PubMed ID
1A3G	Three-dimensional structure of Escherichia coli branched-chain amino acid aminotransferase at 2.5 A resolution.	Okada, K., Hirotsu, K., Sato, M., Hayashi, H., Kagamiyama, H.	1997	J.Biochem.(Tokyo)	121	637	9163511
1A9N	Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA.	Price, S.R., Evans, P.R., Nagai, K.	1998	Nature	394	645	9716128
1AB4	Crystal structure of the breakage-reunion domain of DNA gyrase.	Cabral, J.H., Jackson, A.P., Smith, C.V., Shikotra, N., Maxwell, A., Liddington, R.C.	1997	Nature	388	903	9278055
1AP8	Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein.	Matsuo, H., Li, H., McGuire, A.M., Fletcher, C.M., Gingras, A.C., Sonenberg, N., Wagner, G.	1997	Nat.Struct.Biol.	4	717	9302999
1AU1	The crystal structure of human interferon beta at 2.2-A resolution.	Karpusas, M., Nolte, M., Benton, C.B., Meier, W., Lipscomb, W.N., Goelz, S.	1997	Proc.Natl.Acad.Sci.USA	94	11813	9342320
1AUK	Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis.	Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., Saenger, W.	1998	Biochemistry	37	3654	9521684
1B74	Structure and mechanism of glutamate racemase from Aquifex pyrophilus.	Hwang, K.Y., Cho, C.S., Kim, S.S., Sung, H.C., Yu, Y.G., Cho, Y.	1999	Nat.Struct.Biol.	6	422	10331867
1BJT	Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands.	Fass, D., Bogden, C.E., Berger, J.M.	1999	Nat.Struct.Biol.	6	322	10201398
1BK0	Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation.	Roach, P.L., Clifton, I.J., Hensgens, C.M., Shibata, N., Schofield, C.J., Hajdu, J., Baldwin, J.E.	1997	Nature	387	827	9194566
1BOO	Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment.	Gong, W., O'Gara, M., Blumenthal, R.M., Cheng, X.	1997	Nucleic Acids Res.	25	2702	9207015

1CB8	Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes.	Fethiere, J., Eggimann, B., Cygler, M.	1999	J.Mol.Biol.	288	635	10329169
1D0B	Structure of the InlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen L. monocytogenes.	Marino, M., Braun, L., Cossart, P., Ghosh, P.	1999	Mol.Cell	4	1063	10635330
1DCS	Structure of a cephalosporin synthase.	Valegard, K., van Scheltinga, A.C., Lloyd, M.D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H.J., Baldwin, J.E., Schofield, C.J., Hajdu, J., Andersson, I.	1998	Nature	394	805	9723623
1E42	The Structure and Function of the Beta2-Adaptin Appendage Domain	Owen, D.J., Vallis, Y., Pearse, B.M.F., Mcmahon, H.T., Evans, P.R.	2000	Embo J.	19	4216	10944104
1EGU	Structural basis of hyaluronan degradation by Streptococcus pneumoniae hyaluronate lyase.	Li, S., Kelly, S.J., Lamani, E., Ferraroni, M., Jedrzejas, M.J.	2000	EMBO J.	19	1228	10716923
1EJH	Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G.	Marcotrigiano, J., Gingras, A.C., Sonenberg, N., Burley, S.K.	1999	Mol.Cell	3	707	10394359
1EPW	Structural analysis of the catalytic and binding sites of Clostridium botulinum neurotoxin B.	Swaminathan, S., Eswaramoorthy, S.	2000	Nat.Struct.Biol.	7	693	10932256
1FFV	The effect of intracellular molybdenum in Hydrogenophaga pseudoflava on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase.	Hanzelmann, P., Dobbek, H., Gremer, L., Huber, R., Meyer, O.	2000	J.Mol.Biol.	301	1221	10966817
1FJM	Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1.	Goldberg, J., Huang, H.B., Kwon, Y.G., Greengard, P., Nairn, A.C., Kuriyan, J.	1995	Nature	376	745	7651533
1FO4	Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: structure-based mechanism of conversion.	Enroth, C., Eger, B.T., Okamoto, K., Nishino, T., Nishino, T., Pai, E.F.	2000	Proc.Natl.Acad.Sci.USA	97	10723	11005854

1FOH	The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis.	Enroth, C., Neujahr, H., Schneider, G., Lindqvist, Y.	1998	Structure	6	605	9634698
1FSU	Structure of a human lysosomal sulfatase.	Bond, C.S., Clements, P.R., Ashby, S.J., Collyer, C.A., Harrop, S.J., Hopwood, J.J., Guss, J.M.	1997	Structure	5	277	9032078
1GLN	Architectures of class-defining and specific domains of glutamyl-tRNA synthetase.	Nureki, O., Vassylyev, D.G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S., Morikawa, K.	1995	Science	267	1958	7701318
1GZ0	The Structure of the Rlmb 23S Rrna Methyltransferase Reveals a New Methyltransferase Fold with a Unique Knot	Michel, G., Sauve, V., Larocque, R., Li, Y., Matte, A., Cygler, M.	2002	Structure	10	1303	12377117
1I7B	The structural basis for substrate specificity and inhibition of human S-adenosylmethionine decarboxylase.	Tolbert, W.D., Ekstrom, J.L., Mathews, I.I., Secrist 3rd., J.A., Kapoor, P., Pegg, A.E., Ealick, S.E.	2001	Biochemistry	40	9484	11583147
1IPA	An enzyme with a deep trefoil knot for the active-site architecture.	Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D.G., Shibata, T., Inoue, Y., Kuramitsu, S., Yokoyama, S.	2002	Acta Crystallogr.,Sect.D	58	1129	12077432
1IQ8	Crystal structure of archaeosine tRNA-guanine transglycosylase.	Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., (Morais Cabral, Jackson et al. 1997)Nameki, N., Watanabe, M., Kondo, H., Sekine, M., Okada, N., Nishimura, S., Yokoyama, S.	2002	J.Mol.Biol.	318	665	12054814
1ITF	The three-dimensional high resolution structure of human interferon alpha-2a determined by heteronuclear NMR spectroscopy in solution.	Klaus, W., Gsell, B., Labhardt, A.M., Wipf, B., Senn, H.	1997	J.Mol.Biol.	274	661	9417943

1JFL	Crystal structure of aspartate racemase from Pyrococcus horikoshii OT3 and its implications for molecular mechanism of PLP-independent racemization.	Liu, L., Iwata, K., Kita, A., Kawarabayasi, Y., Yohda, M., Miki, K.	2002	J.Mol.Biol.	319	479	12051922
1K4G	De novo design, synthesis, and in vitro evaluation of inhibitors for prokaryotic tRNA-guanine transglycosylase: a dramatic sulfur effect on binding affinity.	Meyer, E.A., Brenk, R., Castellano, R.K., Furler, M., Klebe, G., Diederich, F.	2002	ChemBioChem	3	250	11921407
1M0W	Large Conformational Changes in the Catalytic Cycle of Glutathione Synthase	Gogos, A., Shapiro, L.	2002	Structure	10	1669	12467574
1MHM	Monomeric S-Adenosylmethionine Decarboxylase from Plants Provides an Alternative to Putrescine Stimulation	Bennett, E.M., Ekstrom, J.L., Pegg, A.E., Ealick, S.E.	2002	Biochemistry	41	14509	12463749
1OYC	Old yellow enzyme at 2 A resolution: overall structure, ligand binding, and comparison with related flavoproteins.	Fox, K.M., Karplus, P.A.	1994	Structure	2	1089	7881908
1PBE	Crystal structure of the p-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 A resolution. Analysis of the enzyme-substrate and enzyme-product complexes.	Schreuder, H.A., Prick, P.A., Wierenga, R.K., Vriend, G., Wilson, K.S., Hol, W.G., Drenth, J.	1989	J.Mol.Biol.	208	679	2553983
1QTQ	How glutaminyl-tRNA synthetase selects glutamine.	Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S., Steitz, T.A.	1998	Structure	6	439	9562563
1QTS	Crystal structure of the alpha appendage of AP-2 reveals a recruitment platform for clathrin-coat assembly.	Traub, L.M., Downs, M.A., Westrich, J.L., Fremont, D.H.	1999	Proc.Natl.Acad.Sci.USA	96	8907	10430869
1TCO	X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex.	Griffith, J.P., Kim, J.L., Kim, E.E., Sintchak, M.D., Thomson, J.A., Fitzgibbon, M.J., Fleming, M.A., Caron, P.R., Hsiao, K., Navia, M.A.	1995	Cell(Cambridge,Mass.)	82	507	7543369
1TDJ	Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase.	Gallagher, D.T., Gilliland, G.L., Xiao, G., Zondlo, J., Fisher, K.E., Chinchilla, D., Eisenstein, E.	1998	Structure	6	465	9562556
2ENT	Solution structure of the second C2H2-type zinc finger domain from human Krueppel-like factor 15	Nagashima, T., Hayashi, F., Yokoyama, S.		To be Published			

2HGS	Molecular basis of glutathione synthetase deficiency and a rare gene permutation event.	Polekhina, G., Board, P.G., Gali, R.R., Rossjohn, J., Parker, M.W.	1999	EMBO J.	18	3204	10369661
2TMD	Correlation of x-ray deduced and experimental amino acid sequences of trimethylamine dehydrogenase.	Barber, M.J., Neame, P.J., Lim, L.W., White, S., Matthews, F.S.	1992	J.Biol.Chem.	267	6611	1551870
2TYS	Crystal structures of a mutant (betaK87T) tryptophan synthase alpha2beta2 complex with ligands bound to the active sites of the alpha- and beta-subunits reveal ligand-induced conformational changes.	Rhee, S., Parris, K.D., Hyde, C.C., Ahmed, S.A., Miles, E.W., Davies, D.R.	1997	Biochemistry	36	7664	9201907
3BTA	Crystal structure of botulinum neurotoxin type A and implications for toxicity.	Lacy, D.B., Tepp, W., Cohen, A.C., DasGupta, B.R., Stevens, R.C.	1998	Nat.Struct.Biol.	5	898	9783750
3PCG	Structures of competitive inhibitor complexes of protocatechuate 3,4-dioxygenase: multiple exogenous ligand binding orientations within the active site.	Orville, A.M., Elango, N., Lipscomb, J.D., Ohlendorf, D.H.	1997	Biochemistry	36	10039	9254599
5DAA	Effects of the E177K mutation in D-amino acid transaminase. Studies on an essential coenzyme anchoring group that contributes to stereochemical fidelity.	van Ophem, P.W., Peisach, D., Erickson, S.D., Soda, K., Ringe, D., Manning, J.M.	1999	Biochemistry	38	1323	9930994

Supplementary Table 2: Median RMSD differences (in Å)* between the structural superpositions generated utilizing sequence alignments with different parameters

Using global (Needleman-Wunsch) or local (Smith-Waterman) sequence alignment algorithm, coupled with standard (sRMSD) or weighted (wRMSD) superposition algorithm.

Homolog Proteins and PDB IDs	%ID	Smith-Waterman		Needlman-Wunsch	
		Standard	Weighted	Standard	Weighted
1FJM & 1TCO	39%	0.42	0.07	0.72	0.02
1M0W & 2HGS	37%	0.27	0.06	0.27	0.07
1AU1 & 1ITF	35%	1.48	0.12	1.43	0.17
1I7B & 1MHM	33%	0.25	0.09	0.24	0.03
1EPW & 3BTA	31%	0.64	0.08	0.69	0.09
1AUK & 1FSU	29%	0.91	0.24	0.54	0.09
1AP8 & 1EJH	29%	0.97	0.22	0.23	0.05
3PCG (chain A) & 3PCG (chain M)	28%	0.76	0.05	0.49	0.04
1A3G & 5DAA	27%	1.38	0.45	1.22	0.13
1IPA & 1GZ0	26%	5.83	1.30	2.59	0.23
1OYC & 2TMD	25%	2.26	0.73	0.30	0.08
1IQ8 & 1K4G	25%	0.76	0.50	0.34	0.08
1GLN & 1QTQ	24%	4.96	1.86	3.01	0.15
1BOO & 2ENT	23%	3.31	0.36	2.73	0.19
1AB4 & 1BJT	22%	1.46	0.55	0.48	0.43
1TDJ & 2TYS	21%	6.89	4.91	5.54	2.59
1BK0 & 1DCS	20%	11.08	1.77	0.92	0.51
1FFV & 1FO4	19%	1.21	0.66	1.08	0.49
1A9N & 1D0B	19%	2.33	0.56	7.85	8.81
1B74 & 1JFL	18%	1.52	0.51	1.36	0.22
1CB8 & 1EGU	18%	1.32	0.68	1.27	0.20
1FOH & 1PBE	17%	5.44	4.72	4.83	2.63
1E42 & 1QTS	16%	5.51	3.98	3.94	0.47

*The sequence alignments were altered by varying the similarity matrix and gap penalties. The variation across the superpositions was measured by pair wise RMSD (Å) between all the solutions. Median differences are reported, but all calculated pair-wise RMSDs are included in the supplementary material. Smaller value denotes a greater agreement between the superpositions.

Supplementary Table 3. Raw values of aligned residues used in comparison of HwRMSD to other Structural Alignment Program solutions.

	HwRMSD			CE			FATCAT			DaliLite			SSM			THESEUS*		
Protein Pair	%Cα<1Å	%Cα<2Å	**nalign**	%Cα<1Å	%Cα<2Å	**nalign**	%Cα<1Å	%Cα<2Å	**nalign**	%Cα<1Å	%Cα<2Å	**nalign**	%Cα<1Å	%Cα<2Å	**nalign**	%Cα<1Å	%Cα<2Å	**nalign**
1FJM & 1TCO	66.10%	85.0%	293	54.24%	87.4%	250	61.02%	88.5%	288	39.66%	84.7%	287	65.08%	87.8%	288	65.08%	85.0%	293
1M0W & 2HGS	43.13%	67.6%	460	45.45%	72.6%	328	43.76%	69.0%	462	37.21%	69.5%	459	43.34%	72.8%	456	45.03%	69.8%	460
1AU1 & 1ITF	25.75%	62.3%	162	22.16%	65.8%	98	17.96%	62.3%	159	2.40%	53.1%	160	0.60%	24.5%	143	25.75%	59.0%	161
1I7B & 1MHM	47.08%	80.7%	290	42.86%	82.9%	232	41.88%	80.3%	290	16.56%	66.8%	289	44.81%	81.7%	290	45.78%	80.3%	290
1EPW & 3BTA	34.59%	62.2%	516	17.58%	61.8%	316	18.53%	60.1%	519	10.40%	56.0%	505	27.41%	65.4%	509	34.22%	62.2%	516
1AUK & 1FSU	40.88%	64.6%	449	34.38%	70.9%	293	34.80%	68.5%	438	33.33%	63.3%	436	34.80%	68.2%	444	38.99%	62.9%	450
1AP8 & 1EJH	13.26%	37.6%	170	1.66%	29.1%	50	7.73%	35.5%	166	5.52%	32.9%	173	10.50%	38.3%	167	11.60%	34.7%	173
3PCG (A) & 3PCG (M)	46.27%	67.2%	186	17.91%	68.5%	111	34.33%	64.0%	189	11.94%	37.5%	184	43.28%	71.8%	181	45.27%	65.1%	186
1A3G & 5DAA	51.80%	88.8%	268	51.80%	89.2%	239	52.52%	89.9%	268	22.30%	83.9%	267	51.80%	90.2%	266	50.72%	86.9%	268
1IPA & 1GZ0	27.57%	52.9%	223	4.12%	29.7%	66	5.35%	33.8%	225	13.58%	45.7%	234	4.94%	34.5%	223	27.57%	48.9%	233
1OYC & 2TMD	29.03%	61.1%	332	25.81%	67.1%	214	20.82%	58.6%	333	7.33%	56.0%	327	21.11%	61.0%	323	31.38%	59.9%	332
1IQ8 & 1K4G	25.60%	70.4%	331	23.21%	69.3%	228	22.62%	68.1%	332	14.29%	59.4%	330	22.02%	68.9%	331	25.89%	69.2%	331
1GLN & 1QTQ	23.22%	60.5%	276	21.05%	64.0%	162	20.74%	60.1%	276	17.03%	59.2%	272	21.36%	63.6%	269	24.46%	58.7%	276
1BOO & 2ENT	40.96%	69.9%	246	33.95%	80.5%	169	28.41%	69.2%	234	14.76%	62.7%	241	40.96%	72.6%	241	40.96%	69.5%	246
1AB4 & 1BJT	14.01%	43.9%	435	11.25%	44.6%	183	10.62%	53.9%	432	3.82%	33.6%	432	5.94%	46.3%	408	16.77%	39.9%	431
1TDJ & 2TYS	14.71%	50.6%	326	12.31%	57.3%	173	11.11%	52.0%	325	15.02%	48.4%	322	16.52%	52.0%	321	16.82%	51.1%	327
1BK0 & 1DCS	15.71%	49.3%	268	11.79%	49.0%	122	13.57%	49.0%	257	9.29%	46.5%	254	14.64%	50.4%	260	16.07%	49.1%	267
1FFV & 1FO4	28.82%	62.1%	285	22.92%	63.6%	178	23.61%	62.7%	284	10.07%	46.1%	280	23.26%	63.8%	282	30.21%	61.1%	285
1A9N & 1D0B	23.31%	44.6%	148	6.13%	47.3%	62	20.25%	52.0%	148	1.84%	34.1%	132	17.79%	49.0%	143	26.38%	43.7%	151
1B74 & 1JFL	15.72%	45.0%	209	3.06%	35.0%	72	10.04%	40.4%	208	3.93%	32.7%	205	9.61%	49.2%	187	12.23%	42.1%	209
1CB8 & 1EGU	18.65%	51.1%	307	13.83%	55.7%	165	11.58%	55.7%	300	10.93%	43.9%	301	14.47%	55.2%	299	22.51%	48.5%	307
1FOH & 1PBE	12.76%	46.5%	368	5.10%	43.5%	158	20.15%	61.1%	365	15.82%	50.1%	367	18.88%	59.5%	363	NA	NA	NA
1E42 & 1QTS	9.83%	30.0%	227	2.14%	16.1%	34	10.68%	39.6%	217	0.00%	3.7%	215	4.70%	22.5%	204	7.26%	25.4%	228

* This is the use of THESEUS in place of wRMSD in our 3-step pipeline for comparing homologs.