Supplementary Material for

Super-resolution biomolecular crystallography with low-resolution data

Gunnar F. Schröder, Michael Levitt, and Axel T. Brunger

SUPPLEMENTARY TABLES

Supplementary Table 1: Properties of Low-Resolution PDB Structures[(a)]

PDB Identifier And Reference	X-ray Resolution (Å)	Number of Residues	Ramachandran Score of Deposited Entry	Deposited Values		Re-calculated	Difference in Values	Number Reflections		Reflections per Residue	
				R_{free}	R_{work}	R_{work}	R_{work}	Total	In R_{free} Test Set	Total	In R_{free} Test Set
1av1[41]	4.00	804	0.814	0.428	0.382	0.419	-0.037	15543	785	19.3	0.98
1isr[42]	4.00	448	0.948	0.259	0.237	0.270	-0.033	7182	577	16.0	1.29
1jl4[43]	4.30	557	0.922	0.453	0.420	0.367	0.053	6524	645	11.7	1.16
1pgf[44]	4.50	1102	0.913	0.267	0.254	0.258	-0.004	11169	902	10.1	0.82
1r5u[45]	4.50	3517	0.805	0.373	0.345	0.365	-0.020	57395	1721	16.3	0.49
1xdv[46]	4.10	1517	0.960	0.449	0.433	0.389	0.044	18877	1553	12.4	1.02
1xxi[47]	4.10	3532	0.937	0.369	0.366	0.395	-0.029	39838	4020	11.3	1.14
1ye1[48]	4.50	574	0.968	0.343	0.295	0.312	-0.017	3799	306	6.6	0.53
1yi5[49]	4.20	1356	0.870	0.378	0.331	0.319	0.012	19221	1596	14.2	1.18
1z9j[50]	4.50	821	0.752	0.338	0.299	0.281	0.018	12303	978	15.0	1.19
2a62[51]	4.50	319	0.749	0.346	0.271	0.322	-0.051	4252	323	13.3	1.01
2bf1[18]	4.00	304	0.680	0.388	0.385	0.393	-0.008	6122	280	20.1	0.92
2i36[52]	4.10	962	0.847	0.412	0.382	0.422	-0.040	16375	832	17.0	0.86
2qag[53]	4.00	702	0.895	0.392	0.376	0.446	-0.070	42484	2125	60.5	3.03
2vkz[54]	4.00	10941	0.935	0.268	0.268	0.303	-0.035	168779	8547	15.4	0.78
3bbw[55]	4.00	543	0.906	0.354	0.302	0.306	-0.004	9032	428	16.6	0.79
3crw[56]	4.00	485	0.768	0.319	0.237	0.283	-0.046	4853	367	10.0	0.76
3dmk[57]	4.19	2127	0.911	0.327	0.280	0.389	-0.109	32510	1618	15.3	0.76
3du7[58]	4.10	1839	0.709	0.265	0.215	0.319	-0.104	25513	1286	13.9	0.70
Average	4.19	1708	0.857	0.354	0.320	0.345	-0.025	26409	1520	16.6	1.02
Minimum	4.00	304	0.680	0.259	0.215	0.258	-0.109	3799	280	6.6	0.49
Maximum	4.50	10941	0.968	0.453	0.433	0.446	0.053	168779	8547	60.5	3.03

[(a)]Showing a summary of the properties of the selected low-resolution PDB structures and the corresponding observed diffraction data. R_{work} and R_{free} values were taken from the PDB file header. Ramachandran Score values were computed with Molprobity[30]. In each column, green shading marks the most favorable maximum or minimum value (high Ramachandran Score or low R-value); pink shading marks the least favorable value. The number of reflections per residue varies widely since the different crystal structures have very different amounts of solvent in the unit cell: when the solvent level is high, the unit cell is larger and there are more reflections per residue at a given resolution. The Ramachandran Score values of the deposited structures are often better than those of the structures obtained after either DEN and noDEN refinement (Table 2). This happens as many of the original PDB structures were optimized by rigid body refinement of previously determined higher resolution structures (1jl4, 1pgf, 1r5u, 1xdv, 1xxi, 1ye1, 1yi5, 2vkz) or refined in the presence of tight secondary structure restraints (2bf1) which

preserved the good secondary structure of the high resolution structures in the deposited low-resolution structures. For the remaining ten cases, the information provided in the deposited structures and corresponding publications was insufficient to determine if rigid body or heavily restrained refinement was used. Such a minimal refinement scheme produces good secondary structure but often results in unusually high R_{free} values or small R_{free}-R differences in the deposited structures. For five cases out of the nine (1jl4, 1r5u, 1xdv, 1ye1, 1yi5) we achieved substantial improvements in R_{free} values compared to the deposited structures with our automated re-refinements (cf. Table 2), emphasizing the notion that torsion angles should always be refined for macromolecular structures determined in the 4 -5 Å range. Thus, our DEN and noDEN re-refinements of the deposited structures ensured that all cases were treated the same way, allowing proper comparison between DEN and noDEN refinements.

Supplementary Table 2. Reference Models Used[(a)]

PDB Identifier	Number of Residues	Number of Chains	Template X-ray Resolution (Å)	% Residues in Reference Model	% sequence ID	Main Chain RMSD (Å)
1av1	804	4	2.40	100.0	100.0	12.70
1isr	448	1	2.20	85.7	93.9	0.50
1jl4	557	4	1.86	48.1	100.0	0.94
1pgf	1102	2	2.75	100.0	100.0	0.39
1r5u	3517	10	2.98	90.8	97.7	3.34
1xdv	1517	2	2.20	57.4	90.0	1.27
1xxi	3532	10	2.70	100.0	100.0	2.31
1ye1	574	4	1.65	100.0	99.6	0.23
1yi5	1356	10	2.05	100.0	82.2	1.39
1z9j	821	3	2.18	99.6	99.1	0.71
2a62	319	1	2.00	64.9	100.0	2.15
2bf1	304	1	2.00	100.0	35.8	9.91
2i36	962	3	3.40	100.0	98.4	1.13
2qag	702	3	2.60	60.5	56.1	2.21
2vkz	10941	6	3.10	100.0	99.9	12.42
3bbw	543	2	2.40	98.7	94.4	3.84
3crw	485	1	2.00	83.3	89.3	4.93
3dmk	2127	3	1.95	54.9	94.1	1.10
3du7	1839	5	3.58	99.8	72.5	2.26
Average	1708	4	2.42	86.5	89.6	3.35
Minimum	304	1	1.65	48.1	35.8	0.23
Maximum	10941	10	3.58	100.0	100.0	12.70

[(a)]Homology models were built for each separate protein chain from templates using sequence-alignment (Online Methods) and then combined into a composite reference model without concern for relative orientation of the chains (note that the automated generation of the DEN excludes atom pairs between chains). The percentage of residues in the reference model varies from 48.1% to 100%. The percentage sequence identity between the sequence modeled and the template varies from 35.8% to 100%. When residues are missing from the template, these residues are not modeled in the homology model and there are no DEN restraints involving atoms in these missing residues. The Main Chain RMSD between the reference model and the deposited PDB structure is calculated separately for each chain and then averaged (weighted by chain length). The Main Chain RMSD between varies from 0.23 Å for 1ye1 to 12.70 Å for 1av1 (all four chains of 1av1 involve a large conformational change). The Template X-ray Resolution value given is averaged over the used homologous structures weighted by chain length used. Similarly, the sequence identity and the RMSD between reference model and original PDB structure are chain-length weighted averages over all chains. In each column, green shading marks the most favorable maximum or minimum value (high Ramachandran Score or low R-value); pink shading marks the least favorable value.

Supplementary Table 3: R_{free} Ranges Found in Refinement Repeats[(a)]

PDB Identifier	DEN		noDEN	
	R_{free} min	R_{free} max	R_{free} min	R_{free} max
1av1	0.335	0.355	0.336	0.364
1isr	0.233	0.243	0.237	0.248
1jl4	0.353	0.382	0.354	0.378
1pgf	0.284	0.299	0.295	0.304
1r5u	0.334	0.349	0.335	0.345
1xdv	0.358	0.383	0.367	0.393
1xxi	0.407	0.415	0.465	0.482
1ye1	0.312	0.343	0.350	0.395
1yi5	0.323	0.335	0.336	0.352
1z9j	0.317	0.326	0.331	0.344
2a62	0.340	0.396	0.353	0.390
2bf1	0.479	0.513	0.492	0.559
2i36	0.387	0.400	0.401	0.433
2qag	0.392	0.415	0.401	0.418
2vkz	0.327	0.331	0.337	0.342
3bbw	0.304	0.327	0.334	0.360
3crw	0.324	0.350	0.338	0.377
3dmk	0.407	0.424	0.428	0.443
3du7	0.332	0.353	0.336	0.357

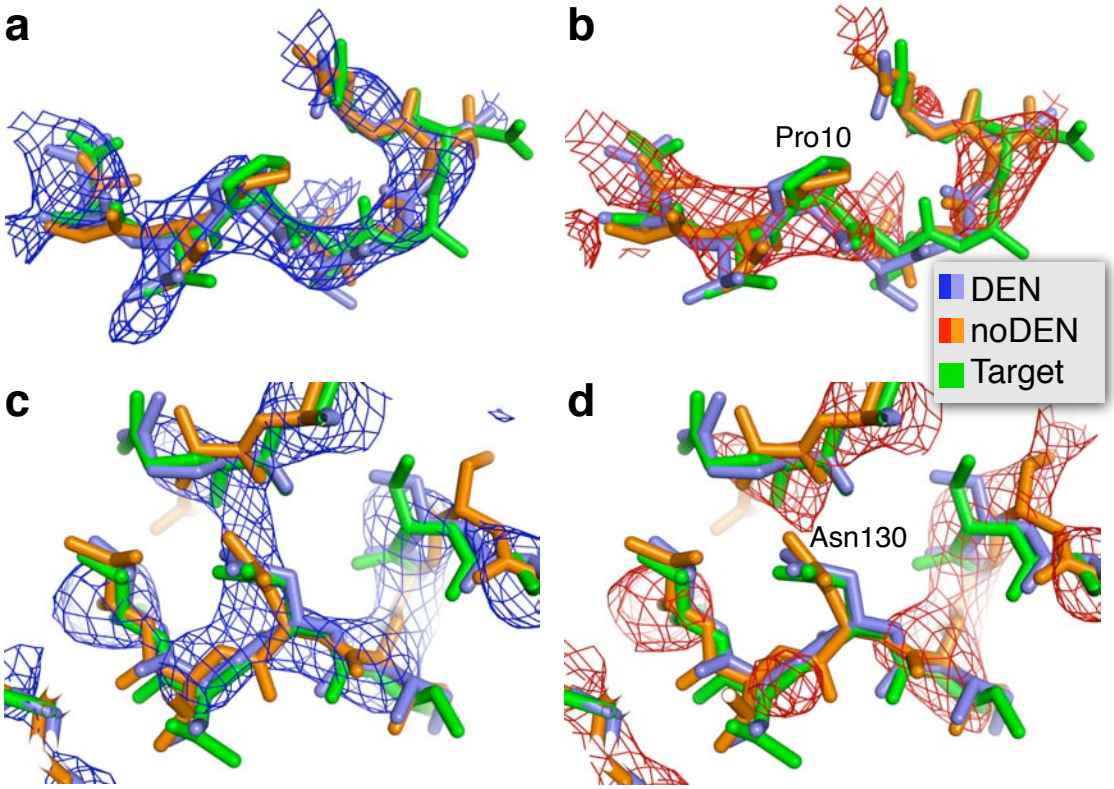
[(a)]Shown are the maximum and minimum R_{free} values obtained from the ten refinement repeats performed at the respective optimum (γ, w_{DEN}) parameter pair for each of the nineteen re-refinements of low-resolution PDB structures.

Supplementary Table 4: Optimum DEN Parameters[a]

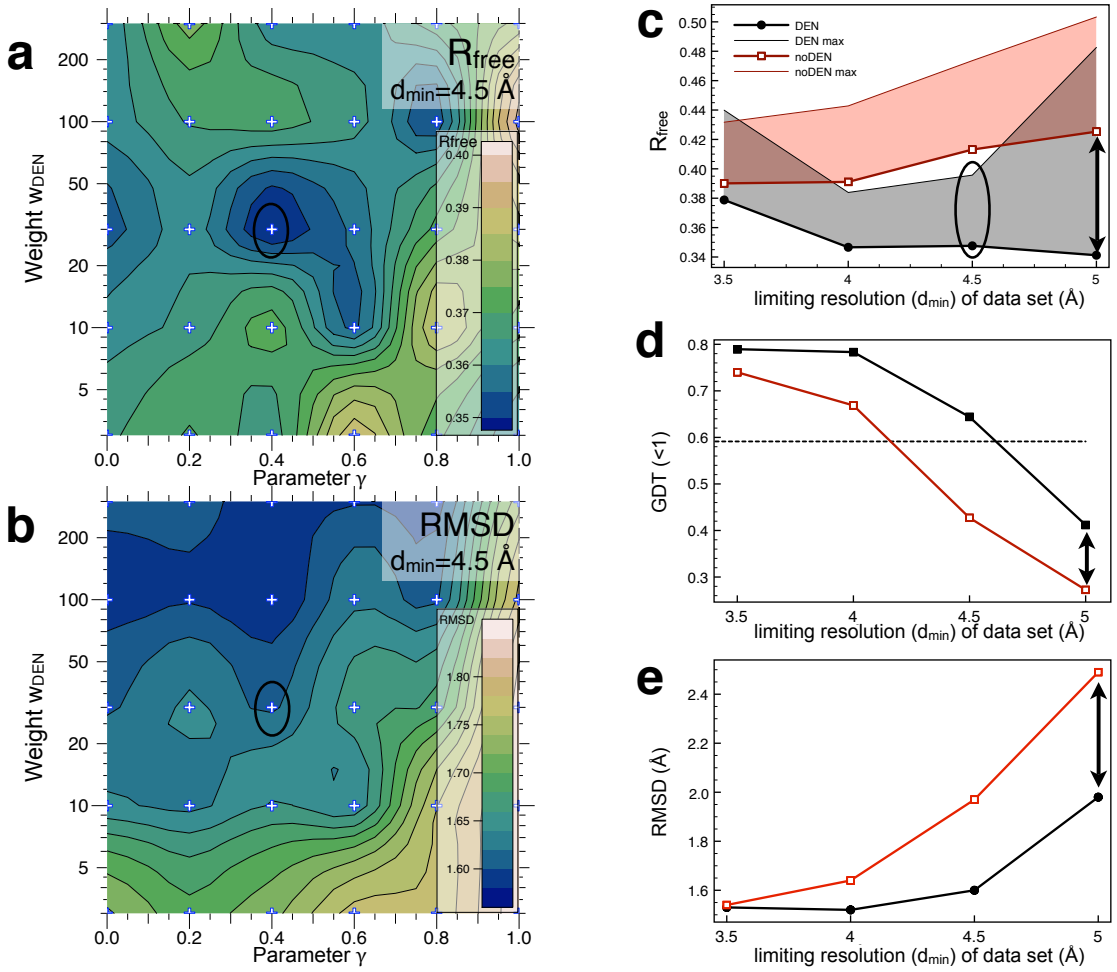
PDB Identifier	Optimum γ	Optimum w_{DEN}	R_{free} Improvement	γ = 1	γ < 1
1av1	1.0	100	0.0012	0.0012	
1isr	1.0	30	0.0043	0.0043	
1jl4	1.0	300	0.0009	0.0009	
1pgf	0.4	300	0.0108		0.0108
1r5u	1.0	3	0.0003	0.0003	
1xdv	1.0	300	0.0089	0.0089	
1xxi	0.2	300	0.0582		0.0582
1ye1	0.0	300	0.0381		0.0381
1yi5	0.6	10	0.0139		0.0139
1z9j	0.6	100	0.0135		0.0135
2a62	1.0	30	0.0131	0.0131	
2bf1	1.0	100	0.0131	0.0131	
2i36	0.2	30	0.0137		0.0137
2qag	1.0	100	0.0091	0.0091	
2vkz	0.6	300	0.0095		0.0095
3bbw	0.4	100	0.0304		0.0304
3crw	0.2	100	0.0136		0.0136
3dmk	0.4	300	0.0211		0.0211
3du7	1.0	10	0.0039	0.0039	
Average	0.66	148.05	0.0146	0.0061	0.0223
Minimum	0.0	3	0.0003	0.0003	0.0095
Maximum	1.0	300	0.0582	0.0131	0.0582

[a]Optimum (γ , w_{DEN}) parameters (Online Methods) obtained from the global grid search (Supplementary Fig. 4) for each of the nineteen re-refinements. The γ-values and DEN weights w_{DEN} for the DEN refined structures that yielded the lowest R_{free} are shown. The improvement in R_{free} is also listed. For easier reference, the cases with γ = 1 and γ < 1 are repeated in separate columns. In each column, green shading marks the most favorable maximum or minimum value (high Ramachandran Score or low R-value); pink shading marks the least favorable value. The parameter w_{DEN} weights the DEN energy term in the hybrid energy function (Eq. 1) and the γ-parameter balances the DEN between reference model and current model coordinates (Eq. 3 in Online Methods). Note that if γ equals 1, the term involving the distances in the reference model (d_{ij}^{ref}) is zero, which means that only topological information from the reference model, but no distance information is used. As explained in the Online Methods, DEN refinement is still able to improve these structures compared to noDEN refinement. Here nine proteins have an optimum γ value of 1.0 (1av1, 1isr, 1jl4, 1r5u, 1xdv, 2a62, 2bf1, 2qag, and 3du7) and the average improvement of R_{free} for DEN over noDEN (taken from Table 2) is 0.0061 which is about a quarter of the improvement for the cases with optimum γ value lower than 1 (0.0223). The average improvement for all structures is 0.0146.

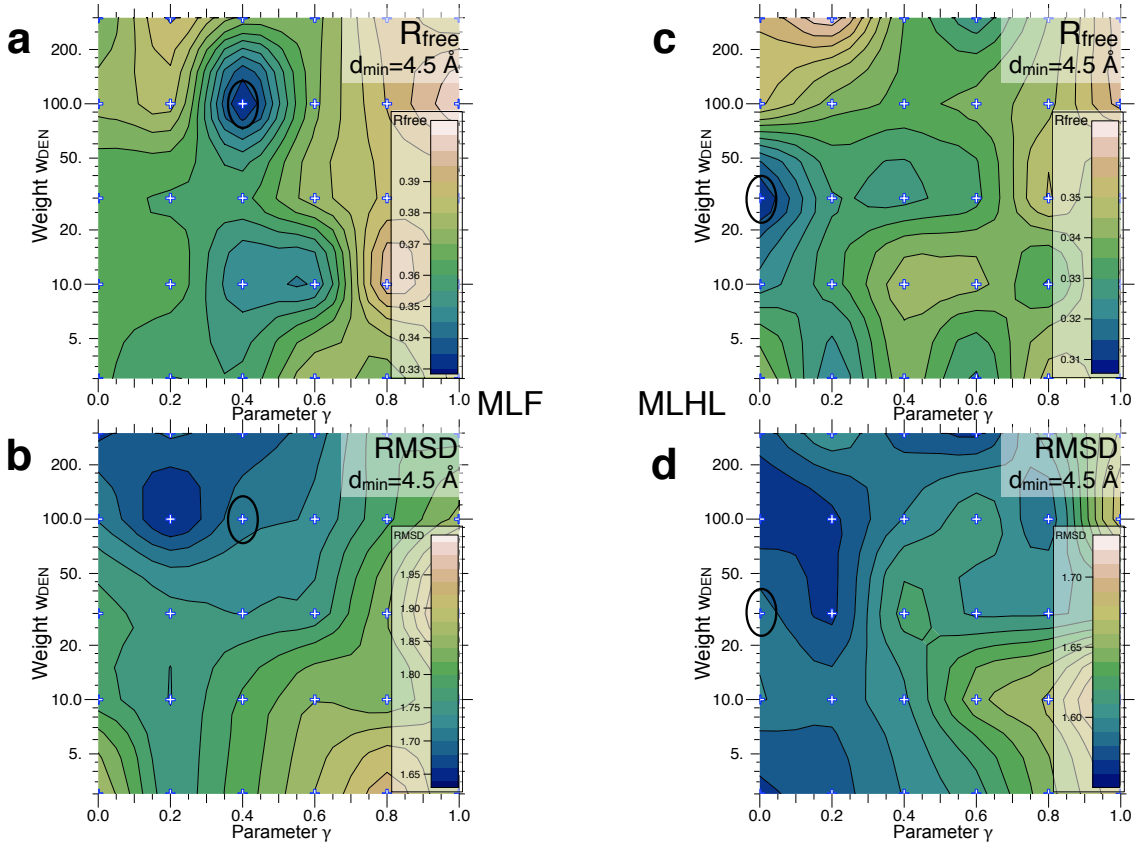
SUPPLEMENTARY FIGURES



Supplementary Figure 1: Improvement of electron density maps for penicillopepsin test calculations using the MLHL target function. Showing how DEN refinement (blue) improves the electron density map compared to noDEN refinement in the presence of experimental phase information. The models refined with and without DEN are superimposed (blue and red sticks, respectively). For comparison, the target structure (penicillopepsin, PDB ID 3app) is shown in green sticks. Refinement is started from the homology model (endothiapepsin, PDB ID 4ape) using the synthetic penicillopepsin diffraction data set with a limiting resolution d_{min} = 4.5 Å and a phase combined σ_A-weighted $2F_o$-F_c electron density map is calculated (superimposed on the model as a blue (red) mesh for refinement with (without) DEN). The experimental phases were obtained by single isomorphous replacement (SIR)[16]. The electron density maps were B-factor sharpened (B_{sharp}=-32.5 Å^2) and the contour level was set to 1.5 σ.

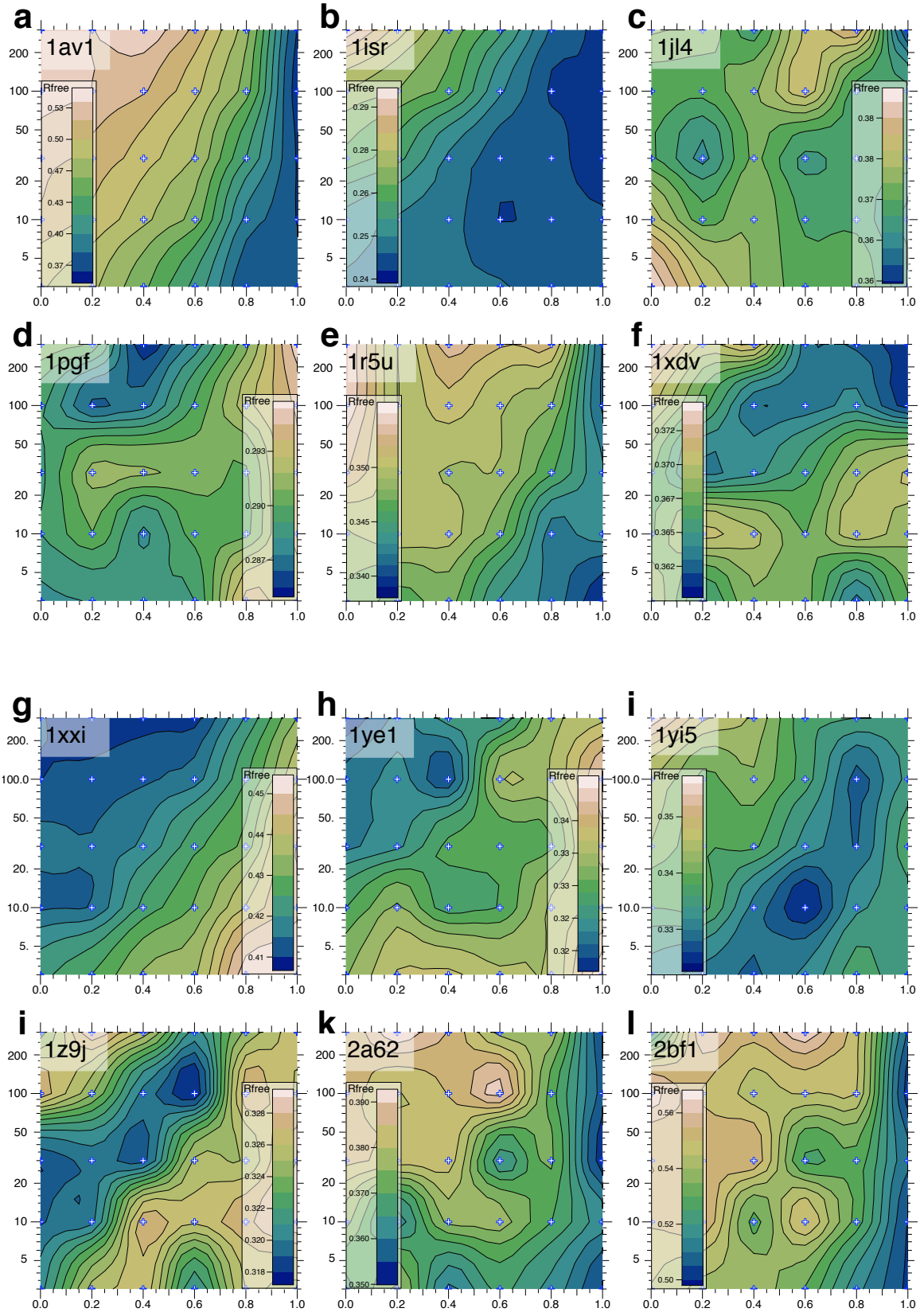


Supplementary Figure 2: Results for the penicillopepsin test calculations using the MLF target function (without experimental phase information). The panels are the same as in Fig. 1.

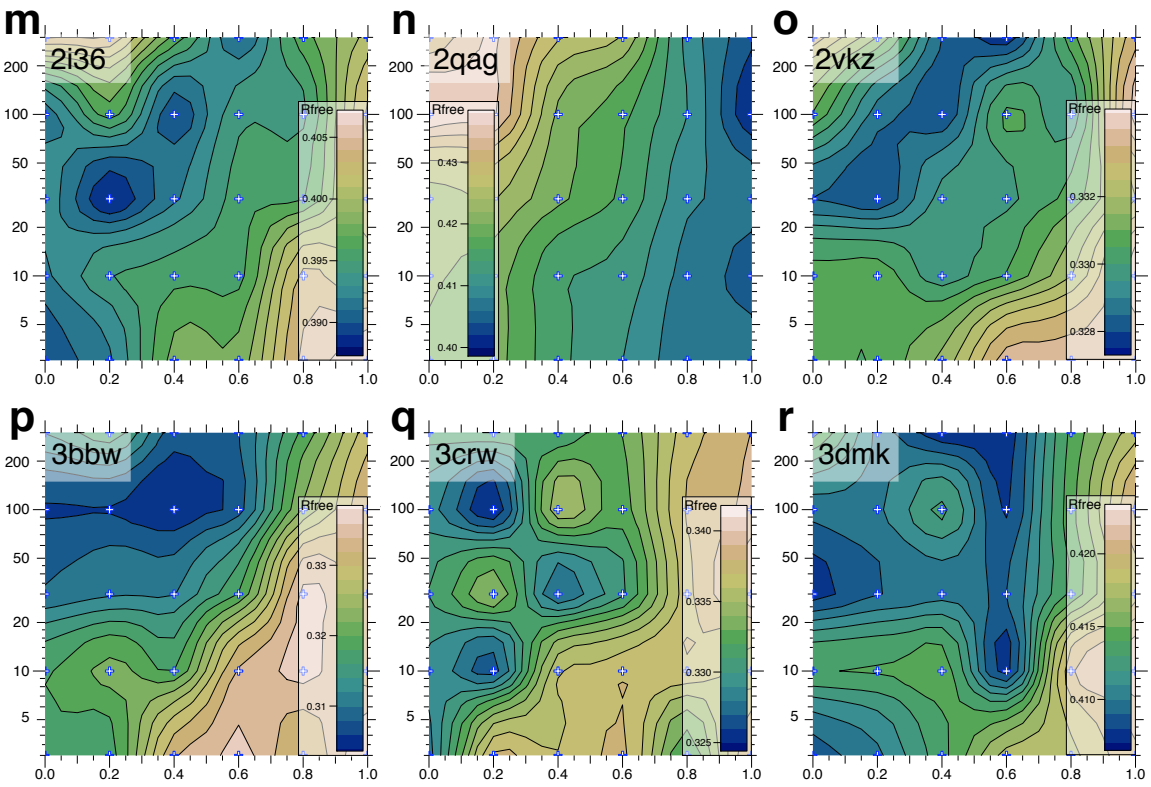


Supplementary Figure 3: Results for the penicillopepsin test calculations extended by two refinement cycles without DEN restraints. Showing how the (γ, w_{DEN}) grid-search determines the values that give the best R_{free} value for the synthetic diffraction data set at d_{min}=4.5 Å using the MLF (panels **a** & **b**) and MLHL (panels **c** and **d**) target functions. Compared to Fig. 1 and Supplementary Fig. 2, the refinement protocol was extended by two cycles without DEN restraints (Online Methods). Contour plots of the obtained R_{free} values ((**a**) and (**c**)) and corresponding RMSD values to the target structure ((**b**) and (**d**)) are shown, similar to Figs. 1a & 1b. The contour plots show clear minima for the R_{free} values. The black ellipses indicate the structures with the lowest R_{free} value. For the MLF target function with (without) DEN restraints the structure with the lowest R_{free} value has a GDT(<1) score of 0.63 (0.32), an RMSD to the target structure of 1.19 Å (1.84 Å), and 62.9% (53.7%) of residues fall within the favored region of the Ramachandran plot as determined by Molprobity. For the MLHL target function with (without) DEN restraints the structure with the lowest R_{free} has a GDT(<1) score of 0.73 (0.71), an RMSD of 1.05 Å (1.17 Å), and 64.7% (63.6%) of its residues fall within the favored region of the Ramachandran plot. The difference between these structures and the structures obtained without the two additional cycles without DEN restraints is small, their RMSD is 0.8 Å and 0.6 Å for the MLF and MLHL target functions, respectively. The R_{free} values drop even further: from 0.356 to 0.328 for MLF and from 0.311 to 0.307 for MLHL. This shows that DEN restraints are mostly helpful in the conformational search but do not bias the minimum of the target function significantly.

Supplementary Figure 4

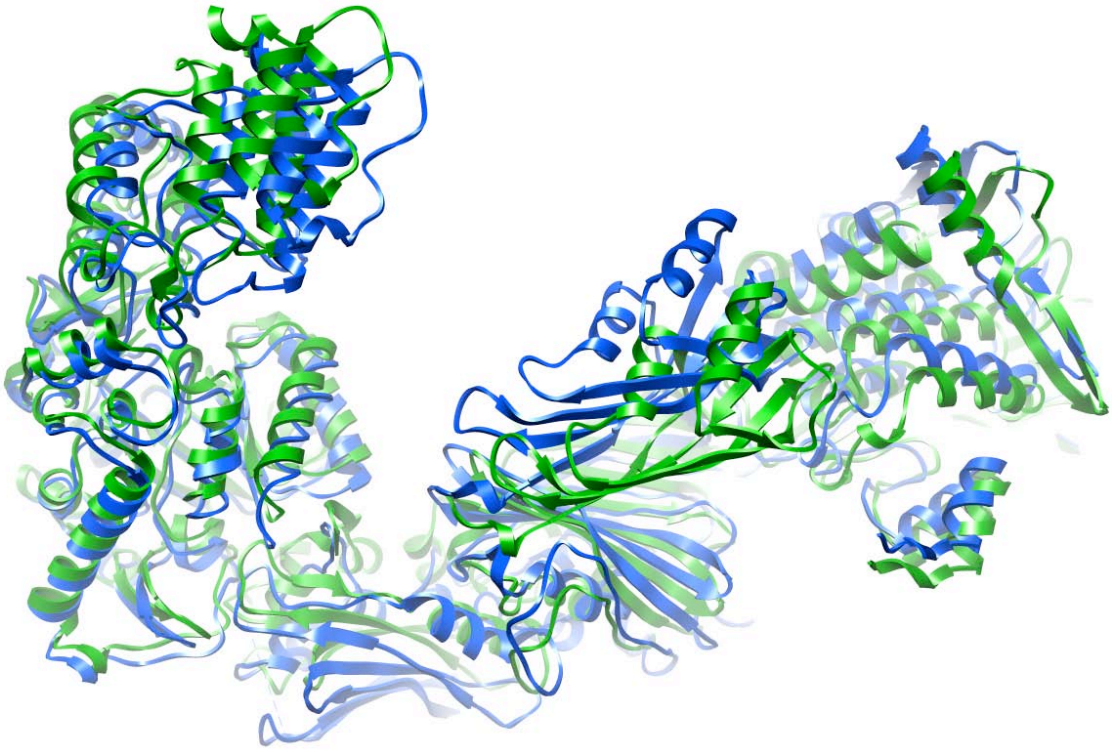


Supplementary Figure 4, cont'd

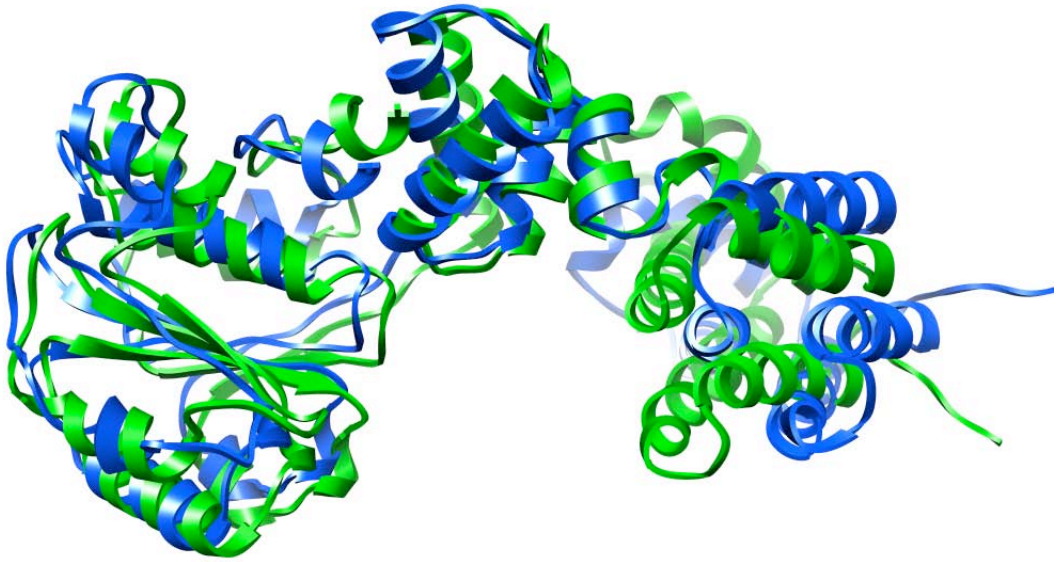


Supplementary Figure 4: R_{free} contour plots showing the parameter grid search for all nineteen re-refinements of low-resolution PDB structures. The R_{free} value is contoured using values calculated on a 6 x 5 grid (marked by small '+' signs) where the parameter γ, the deformability of the DEN, was [0.0, 0.2, 0.4, 0.06, 0.8, 1.0] and w_{DEN}, the weight of the DEN energy term, was [3, 10, 30, 100, 300]. For each grid point, ten refinement repeats were performed (each consisting of eight macrocycles, Online Methods) and the lowest R_{free} is plotted at each grid point.

(a)



(b)



Supplementary Figure 5: Examples of large differences between the reference models and the corresponding DEN-refined structures. DEN refinement can make use of reference models that exhibit significant domain motion compared to the crystal structure. (a) The reference model (green) used for PDB ID 2vkz is superimposed on the respective DEN-refined structure (blue). Only the H chain is shown and all backbone atoms were used for the superposition. (b) The reference model (green) used for PDB ID 1xxi is superimposed on the respective DEN-refined structure (blue). Only the B chain is shown and all backbone atoms were used for the superposition.

SUPPLEMENTARY REFERENCES

[41] Borhani, D.W., Rogers, D.P., Engler, J.A., & Brouillette, C.G. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation in *Proc Natl Acad Sci USA* (1997), Vol. 94, pp. 12291-12296.

[42] Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H., & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+ in *Proc Natl Acad Sci USA* (2002), Vol. 99, pp. 2660-2665.

[43] Wang, J.H. *et al.* Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule in *Proc Natl Acad Sci USA* (2001), Vol. 98, pp. 10799-10804.

[44] Loll, P.J., Picot, D., Ekabo, O., & Garavito, R.M. Synthesis and use of iodinated nonsteroidal antiinflammatory drug analogs as crystallographic probes of the prostaglandin H2 synthase cyclooxygenase active site in *Biochemistry* (1996), Vol. 35, pp. 7330-7340.

[45] Bushnell, D.A., Westover, K.D., Davis, R.E., & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms in *Science* (2004), Vol. 303, pp. 983-988.

[46] Sondermann, H. *et al.* Structural analysis of autoinhibition in the Ras activator Son of sevenless in *Cell* (2004), Vol. 119, pp. 393-405.

[47] Kazmirski, S.L., Podobnik, M., Weitze, T.F., O'Donnell, M., & Kuriyan, J. Structural analysis of the inactive state of the Escherichia coli DNA polymerase clamp-loader complex in *Proc Natl Acad Sci USA* (2004), Vol. 101, pp. 16750-16755.

[48] Kavanaugh, J.S., Rogers, P.H., & Arnone, A. Crystallographic evidence for a new ensemble of ligand-induced allosteric transitions in hemoglobin: the T-to-T(high) quaternary transitions in *Biochemistry* (2005), Vol. 44, pp. 6101-6121.

[49] Bourne, Y., Talley, T.T., Hansen, S.B., Taylor, P., & Marchot, P. Crystal structure of a Cbtx-AChBP complex reveals essential interactions between snake alpha-neurotoxins and nicotinic receptors in *EMBO J* (2005), Vol. 24, pp. 1512-1522.

[50] Thielges, M. *et al.* Design of a redox-linked active metal site: manganese bound to bacterial reaction centers at a site resembling that of photosystem II in *Biochemistry* (2005), Vol. 44, pp. 7389-7394.

[51] Patel, S. *et al.* Type II cadherin ectodomain structures: implications for classical cadherin specificity in *Cell* (2006), Vol. 124, pp. 1255-1268.

[52] Salom, D. *et al.* Crystal structure of a photoactivated deprotonated intermediate of rhodopsin in *Proc Natl Acad Sci USA* (2006), Vol. 103, pp. 16123-16128.

[53] Sirajuddin, M. *et al.* Structural insight into filament formation by mammalian septins in *Nature* (2007), Vol. 449, pp. 311-315.

[54] Johansson, P. *et al.* Inhibition of the fungal fatty acid synthase type I multienzyme complex in *Proc Natl Acad Sci USA* (2008), Vol. 105, pp. 12803-12808.

[55] Qiu, C. *et al.* Mechanism of activation and inhibition of the HER4/ErbB4 kinase in *Structure* (2008), Vol. 16, pp. 460-467.

[56] Fan, L. *et al.* XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations in *Cell* (2008), Vol. 133, pp. 789-800.

[57] Sawaya, M. *et al.* A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms in *Cell* (2008), Vol. 134, pp. 1007-1018.

[58] Cormier, A., Marchand, M., Ravelli, R.B.G., Knossow, M., & Gigant, B. Structural insight into the inhibition of tubulin by vinca domain peptide ligands in *EMBO Rep* (2008), Vol. 9, pp. 1101-1106.