Supplementary Information

Sampling Conformational Changes Prior and After Complexation Using ClustENM and HADDOCK for protein-protein and protein-DNA systems

Zeynep Kurkcuoglu[1] and A. M. J. J. Bonvin[1]

[1]Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands

E-mail: zeynepkurkcuoglu@gmail.com (ZK), a.m.j.j.bonvin@uu.nl (AMJJB)

Docking settings

Table S1. Flexible multidomain docking benchmark settings

PDB	Receptor active residues	Ligand active residues
1IRA	12- 15, 30, 31, 112, 114-116, 120, 121, 127, 129, 201- 203, 205, 236, 238, 240, 270, 271, 298-303	39, 40 ,42, 43, 50-54, 56, 102, 107, 126-130, 147, 149, 150, 151
1H1V	428, 429, 465, 467, 473, 474, 476, 477, 478, 480-482, 484, 485, 487, 488, 491-498, 646-651, 654-656, 658, 669	425, 533, 543-548, 567-569, 621, 692, 696, 711, 714, 715, 718, 726-728, 734, 741, 742, 745, 746, 748, 749
1Y64	1359, 1360, 1362, 1402-1409, 1411, 1412, 1414, 1415, 1424, 1427, 1428, 1430-1432, 1434, 1437, 1460, 1463, 1508, 1632, 1639, 1641	4-6, 99, 100, 102, 124, 125, 128, 130, 143-148, 167, 225, 318, 319, 321-324, 328, 341, 345, 346, 348-354, 359, 360, 362-365, 366, 367
1F6M	37, 39, 44, 81-83, 85, 99, 100, 128-131, 133, 137-139, 141-143, 215-217, 237, 238	31-35, 37, 40, 41, 44, 60, 67, 70-75, 77, 91, 93-98
1FAK	275-278, 304-309, 312, 325, 379, 382, 1062-1065, 1069-1074, 1076-1082, 1085, 1088, 1090, 1092-1094	17, 18, 20, 22, 24, 37, 39-48, 50, 51, 56, 58, 61, 74, 76, 90-96, 109, 110, 112, 130-133, 135, 140, 203
1ZLI	1, 4-12, 15, 26-29, 34, 35, 37, 41-46, 48, 49, 52-55, 57, 72-74	14, 69, 71-73, 119-125, 127, 145, 154-157, 161-164, 196-199, 201, 210, 244-249, 270, 276, 277, 279, 280, 283
1E4K	235-239, 326-330, 332, 1234-1239, 1265-1267, 1269, 1298, 1299, 1327	85-87, 110, 113, 114, 116, 117, 119, 126-129, 131, 132, 152, 155-158
1IBR	10-13, 15, 18, 22, 25, 51, 52, 55, 56, 59, 60, 62, 63, 67-69, 72, 105-107, 111, 156, 159, 160, 199, 232, 235, 239, 242, 246, 274, 277, 278, 281, 284, 285, 288, 336, 338, 340-342, 350, 354	12, 45-47, 64, 70, 74-79, 81, 82, 103, 104, 106, 107, 109-114, 137, 139-147, 156, 159, 162, 163, 166, 172, 173
1KKL	136-138, 140, 157, 178-180, 182, 198, 199, 204, 210, 1297, 1298, 1301, 1302, 1304, 1305, 1308-1310	11-16, 32, 37, 40-43, 45-49, 51-57
1NPE	37-39, 42-47, 73-79, 82, 89-91	921, 944, 946-949, 951, 967, 968, 992, 994, 1010, 1011, 1035, 1037, 1053, 1055, 1080, 1082, 1098, 1122, 1124, 1138, 1160, 1161, 1163, 1164, 1166
1DFJ	6, 7, 31, 32, 60, 89, 117, 146, 202, 228, 257, 259, 283, 285, 314, 316, 342, 379, 397, 399, 401, 402, 404-408, 426, 428, 430, 431, 432-436, 453, 455, 456	4, 7, 11, 12, 23, 24, 28, 31, 32, 35, 38-44, 65-67, 69, 71, 86,88-91, 109-111, 118-120

Table S2. Protein-DNA docking settings

PDB	**Protein ambiguous interaction restraints**	**DNA ambiguous interaction restraints**
1BY4	Act: 31, 32 -> 5, 6, 25, 26 24, 27 -> 3, 4, 27, 28 72, 73, 80 -> 2-4 Pas: 34, 75, 76, 77, 55, 56, 59, 62	Act: 2, 3,4, 5, 6, 23-29
3CRO	Act: 29, 31, 32, 35, 42-44 -> 14, 15, 23, 33 Pas: 9, 18-20, 27, 28, 30, 34, 36, 37, 40 ,41, 45, 46	4-7, 13-18, 22-25, 31-34, 36
1AZP	Act: 24 -> 3, 15 26, 29, 31, 45 -> 2-4, 13-15 22, 33, 42 -> 5-7, 10-12 Pas: 21, 25, 27, 28 39, 40, 44, 46, 47	2-7, 10-15
1JJ4	Act: 13, 16, 17, 19-21 Pas: 34, 35, 37 -> 26, 27	3-5, 25-30
1A74	Act: 97,122 -> 35, 36 54-56, 59, 60, 65, 73, 75 -> 1-7 Pas: 51, 57, 58, 66, 71, 77, 100, 119	1-7, 35, 36, 40
1ZME	Act: 9, 11, 12, 80, 82, 83 Pas: 4, 14, 39-43, 75, 85, 100-114	2-4, 9-15, 17, 18, 20-22, 26-35

Supplementary Figures

Figure S1. i-RMSD distributions of the generated complex structures from start (docking models, "gen0") to the final CustENM models ("gen2"). We used two starting models obtained from docking (top 2 best scoring clusters from HADDOCK, named as "gen0") and applied post-docking ClustENM for two generations to simulate induced-fit conformational changes. Yellow and green dots distinguish the conformations originating from the two starting docked models. The grey area highlights i-RMSD values below 4 Å. The i-RMSD changes due to ClustENM sampling is more visible in the case of 1Y64, 1F6M, 1FAK and 1E4K, but not so pronounced in the remaining cases. The quality of the starting models is the determining factor for the quality of the complexes after ClustENM sampling.

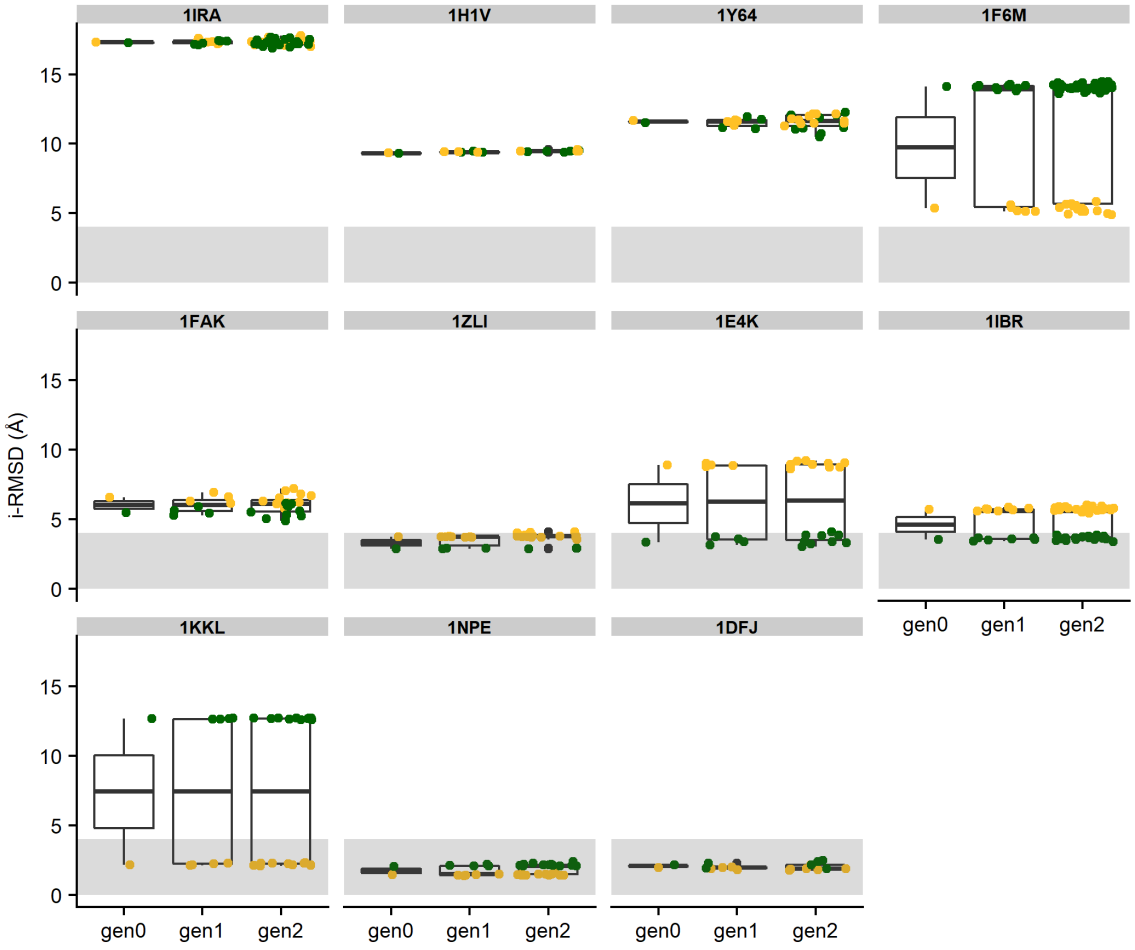

Figure S2. L-RMSD distributions of the generated complex structures of protein-protein systems from start (docking models, "gen0") to the final ClustENM models ("gen2"). The remaining dots in "gen1" and "gen2" are colored based on their "descendance" (i.e. green dots are the children of green dot in "gen0" and the same for yellow). The grey area is where l-RMSD is less than 10 Å. As expected, the change in l-RMSD is more visible compared to i-RMSD, illustrated in cases of 1Y64, 1F6M, 1FAK, 1ZLI, 1E4K, 1IBR, 1KKL and 1NPE.

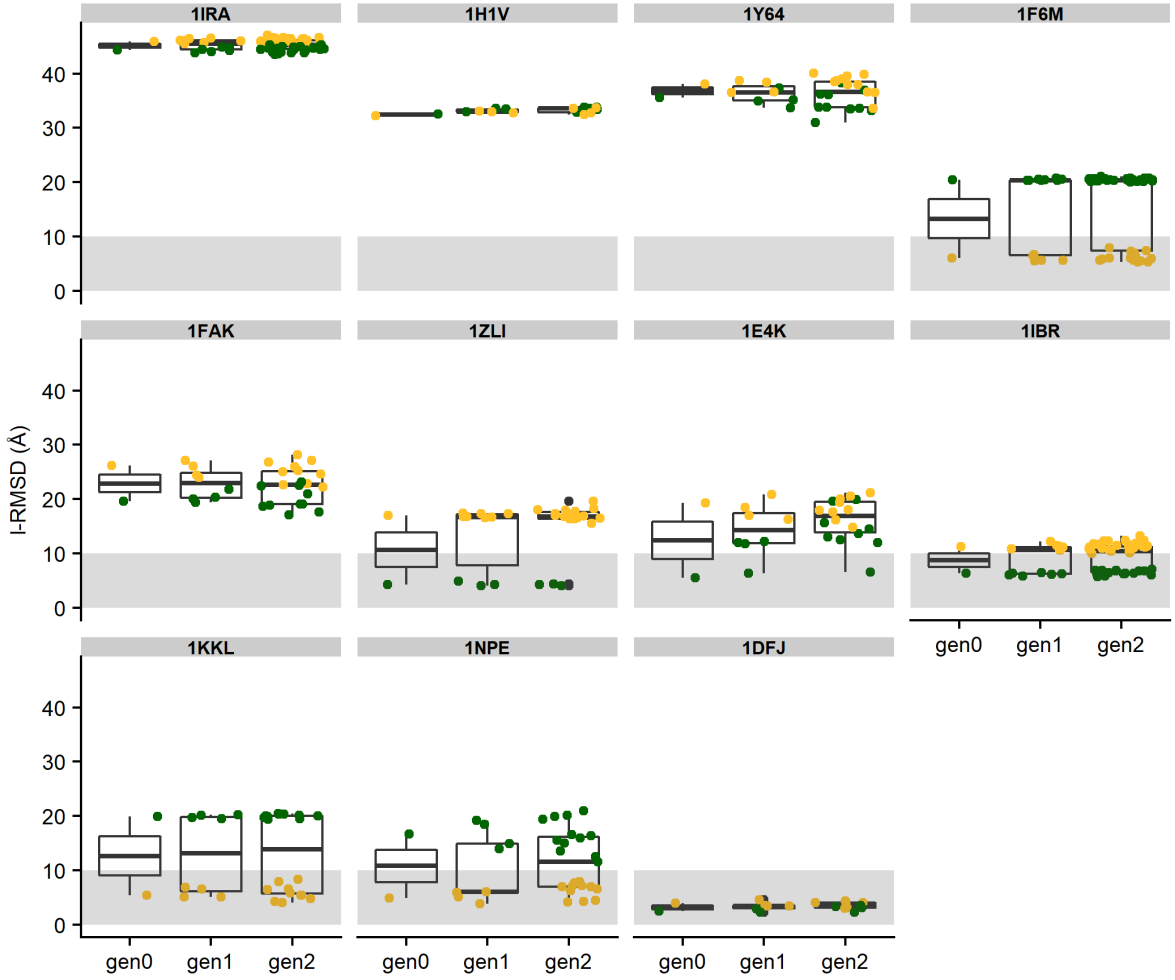

Figure S3. fnat of generated structured from start ("gen0") to final ("gen2") for protein-protein systems. Similar to i-RMSD and l-RMSD figures, two starting complex structures obtained from docking are denoted as "gen0" and shown as yellow and green dots. The remaining dots in "gen1" and "gen2" are colored based on their descendance. Grey colored area is where fnat is more than 0.1. The change in fnat is pronounced in cases of 1F6M, 1FAK, 1ZLI, 1E4K, 1IBR, 1KKL, 1NPE and 1DFJ.

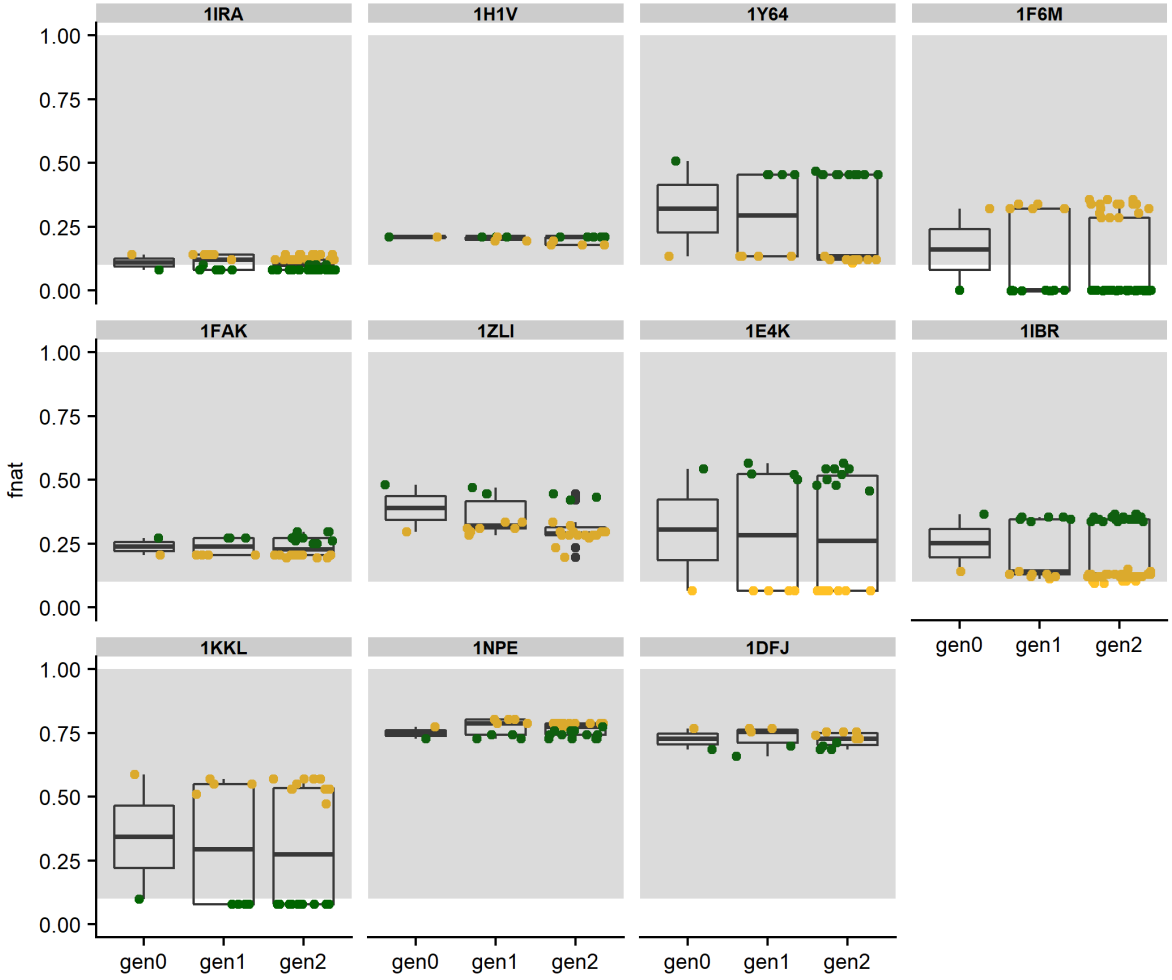

Figure S4. i-RMSD distributions of generated models from start ("gen0") to final ("gen2"). Green and yellow dots in "gen0" are the starting complex structures obtained from the docking. Remaining dots in "gen1" and "gen2" are colored based on their "descendance". The grey area highlights the region with i-RMSD less than 4 Å. The change in i-RMSD in protein-DNA complexes due to post-docking sampling is visible in all cases. The sampled structures show deviations resulting in both smaller and larger changes in i-RMSD. We can see that the post-docking ClustENM sampling does improve the quality of the complexes, thereby mimicking induced-fit effects.

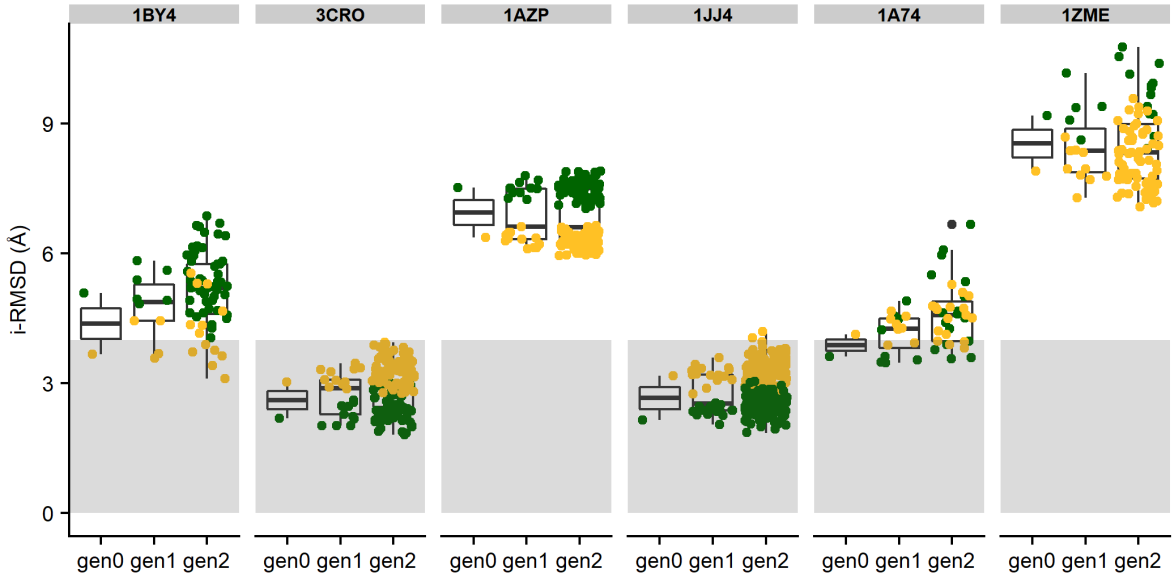

Figure S5. l-RMSD of generated structured from start ("gen0") to final ("gen2") for protein-DNA systems. Green and yellow dots in "gen0" are the starting complex structures obtained from docking and the remaining dots are colored based on their descendance. Like in the i-RMSD figure for protein-DNA cases, the change in l-RMSD is clear in all cases. Especially in 1ZME case, some of sampled conformers in gen2 have favorable l-RMSD thanks to post-sampling (grey region where l-RMSD < 10 Å).

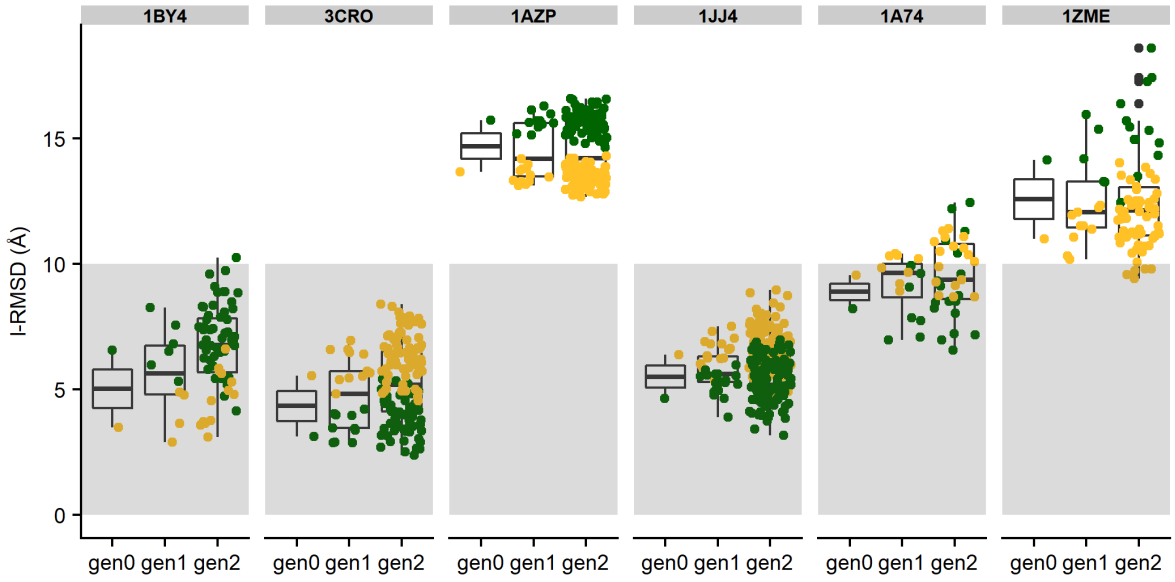

Figure S6. fnat of generated structured from start ("gen0") to final ("gen2") for protein-DNA systems. The color code is the same as i-RMSD and l-RMSD figures; green and yellow dots in "gen0" are the starting complex structures obtained from docking and the remaining dots are colored based on their descendance. The change in fnat for protein-DNA cases is also visible in all cases, in both favorable and unfavorable directions.