Supporting Information

An Accurate Metalloprotein-Specific Scoring Function and Molecular Docking Program Devised by a Dynamic Sampling and Iteration Optimization Strategy

Fang Bai[†,‡,¶], Sha Liao[§,¶], Junfeng Gu[†], Hualiang Jiang[⊥], Xicheng Wang[†], Honglin Li[*,§]

[†]Department of Engineering Mechanics, State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian, Liaoning 116023, China

[‡]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA

[§] State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

[⊥]Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica,Chinese Academy of Sciences, Shanghai 201203, China.

[*]To whom correspondence should be addressed. E-mail: hlli@ecust.edu.cn

Table S1. The training set of 434 zinc protein-ligand complexes (PDB codes).

1AZM 1B57 1BAV 1BCD 1BIW 1BKC 1BN3 1BN4 1BNN 1BNQ 1BNT 1BNV 1BQO 1BSK 1BZS 1C1R 1C1V 1C1W 1C2D 1C2F 1C2G 1C2H 1C2K 1C3I 1C3S 1C8T 1CAQ 1CBX 1CGL 1CIL 1CIN 1CIZ 1CNW 1CNY 1CPS 1CXV 1CZM 1D5J 1D7X 1D8M 1DD6 1DE5 1DMT 1DMY 1DPM 1DTH 1E48 1E51 1EOU 1EZ2 1F57 1FBE 1FT7 1G1D 1G4K 1G4O 1G9A 1G9B 1G9D 1GHY 1GI4 1GKC 1GVF 1GW6 1H2M 1H48 1H8L 1HDQ 1HDU 1HFC 1HFS 1HLK 1HS6 1HWW 1HXK 1HYT 1I76 1I8Z 1I90 1I9L 1I9M 1I9N 1I9O 1I9P 1IF4 1IF5 1IF6 1IF8 1IF9 1IGB 1IX1 1IY7 1J36 1JAQ 1JCQ 1JD0 1JH1 1JJ9 1JJE 1JY8 1KAR 1KQ0 1KR3 1KR6 1KWQ 1KWR 1LD7 1LFW 1LRH 1LRU 1LT8 1M2X 1MMB 1MMP 1MMQ 1MMR 1N95 1NI1 1NL4 1O1T 1O86 1OHL 1OKL 1OKN 1OQ5 1P6C 1P6E 1P6O 1PE5 1PS3 1PS6 1PVY 1PWU 1Q3A 1QF0 1QF1 1QF2 1QJI 1R1H 1R1J 1R55 1RJ6 1RM8 1RMZ 1S63 1S64 1SA4 1SLN 1SNN 1T64 1T69 1TF9 1THL 1TKH 1TQS 1TQT 1TQU 1TQW 1TTM 1TXR 1U4G 1USN 1UZE 1VEV 1VGN 1VKG 1W22 1XAI 1XOR 1XPZ 1XRY 1Y3G 1Y8J 1Y9Q 1YDA 1YDB 1YDD 1YOU 1YQY 1Z9G 1ZDP 1ZE8 1ZFK 1ZFQ 1ZG8 1ZG9 1ZGE 1ZH9 1ZP5 1ZS0 1ZTQ 1ZVX 1ZXV 1ZZ3 2A8H 2AFW 2AFX 2AIO 2AW1 2D1N 2D1O 2DOO 2DQM 2DVU 2DW1 2E25 2E3X 2EG7 2EK9 2ERP 2EU2 2EU3 2F14 2F92 2FOS 2FOV 2FOY 2FU8 2FV5 2G7Q 2GC0 2GD8 2GFK 2GKL 2H15 2HB9 2HD6 2HL4 2HNC 2HOC 2HPT 2I57 2ILP 2IMA 2IMB 2ISV 2IWE 2J83 2JBJ 2JEW 2JIG 2NMX 2NN1 2NN7 2NNO 2NNS 2NNV 2O3K 2O4H 2OC2 2OKL 2OVX 2OVZ 2OW0 2OW1 2OW7 2P53 2PIY 2PJ0 2PJ1 2PJ2 2PJ4 2PJ5 2PJ6 2PJ7 2PJ9 2PJA 2PJB 2PJC 2POU 2POV 2POW 2PVW 2Q1Q 2Q38 2QDS 2QO8 2QOA 2QP6 2QPJ 2R2L 2R59 2RFH 2RJQ 2TCL 2TMN 2V5X 2V77 2V8H 2VCG 2VQJ 2VQQ 2W0D 2W12 2W13 2W14 2WD3 2WEG 2WEH 2WEO 2WT9 2Z3H 2ZIR 2ZOG 2ZXG 3AIG 3B3C 3B4F 3B7R 3B8Z 3B92 3BET 3BKK 3BKL 3BL0 3BLB 3BOF 3BOL 3BQ5 3C10 3C52 3C56 3C7P 3CA2 3CAJ 3CHQ 3CHS 3CZN 3CZV 3D51 3D7D 3D7F 3D7H 3D8C 3DA2 3DBK 3DCC 3DCW 3DD0 3DD8 3DDG 3DHC 3DX3 3DX4 3E32 3E33 3E34 3E37 3EDZ 3EFT 3EHX 3EHY 3EJS 3ELM 3EW8 3EWJ 3EZP 3EZT 3F06 3F15 3F16 3F17 3F19 3F1A 3F28 3F2P 3F4X 3F7U 3FCQ 3FFP 3FH7 3FHE 3FLF 3FOR 3FUE 3FUF 3FUH 3FV4 3FVL 3FVP 3FW3 3FWD 3FXP 3G42 3GAY 3GIQ 3GNH 3H67 3HK5 3HK7 3HK8 3HK9 3HKA 3HKQ 3HKT 3HKU 3HLJ 3HS4 3HXD 3HXF 3HY7 3I1U 3IBI 3IBL 3IBN 3IBU 3IKF 3IVT 3IWW 3K2F 3K5X 3KSQ 3L0V 3LY0 3ZNC 456C 4AIG 4FUA 6CPA 7CPA 7TLN 830C 8CPA 966C

Table S2. The test set of 106 zinc protein-ligand complexes (PDB codes).

1B3D 1BN1 1BNM 1BNU 1BZM 1C1U 1C2E 1C2I 1C3R 1CIM 1CNX 1D8F 1DE6 1E47 1FBL 1G4J 1G9C 1GT7 1H2B 1HEE 1HY7 1I91 1I9Q 1IF7 1J37 1JIZ 1JJT 1LD8 1LRY 1MNC 1NVD 1OK M 1P6D 1PE7 1PE8 1Q1Y 1R1I 1ROS 1SA5 1T67 1TKF 1UZF 1X8I 1XUG 1YBQ 1YEJ 1Z9Y 1ZG7 1ZGF 1ZSB 1ZXC 2AFU 2C6C 2DDF 2DW0 2EG8 2F0Y 2FOQ 2FU9 2H4N 2IEJ 2ISW 2NNG 2OI0 2OW2 2PIZ 2PJ8 2PJT 2Q1B 2QDT 2QVV 2RJP 2V2A 2VQO 2W15 2WEJ 2YZ3 2ZIS 3B7U 3BHX 3BL1 3C0Z 3CHP 3D7G 3DAZ 3DCS 3DHB 3E30 3E8R 3ELF 3F0R 3F18 3F7B 3FGD 3FTX 3FUK 3FX6 3GB6 3H68 3HKN 3HXC 3HY9 3IGP 3JVH 4TLN 5TLN

Table S3. The set of 142 zinc protein-ligand complexes which have the reported experimental binding affinities.

1AZM 1B57 1BN1 1BN3 1BN4 1BNM 1BNN 1BNQ 1BNT 1BNU 1BNV 1BQO 1C1R 1C1U 1C1V 1C2D 1CBX 1CGL 1CIL 1CIM 1CIN 1DMY 1EOU 1F57 1FBL 1G1D 1G4J 1G4O 1GHY 1GI4 1HFC 1HFS 1I8Z 1I91 1I9L 1I9M 1I9N 1I9O 1I9P 1I9Q 1IY7 1JD0 1MMP 1MMQ 1MMR 1MNC 1O86 1OQ5 1P6D 1P6E 1QF0 1QF1 1QF2 1R1H 1R1J 1S63 1T69 1TF9 1THL 1TTM 1USN 1UZE 1UZF 1XUG 1Y3G 1YDA 1YDB 1YDD 1Z9G 1Z9Y 1ZDP 1ZE8 1ZFQ 1ZGE 1ZGF 1ZS0 1ZSB 1ZVX 1ZXC 1ZXV 2AFW 2AFX 2AW1 2D1N 2F92 2FV5 2GC0 2H15 2H4N 2HD6 2HH5 2HNC 2HOC 2IMA 2JBJ 2JEW 2NN1 2NNG 2NNO 2OI0 2POU 2POV 2POW 2PVW 2Q1B 2Q1Q 2Q38 2QO8 2QOA 2TMN 2WEJ 3B4F 3B92 3BHX 3BL1 3CAJ 3D7H 3DCC 3DCW 3DD0 3DD8 3E8R 3EDZ 3ELM 3EWJ 3F4X 3FFP 3FW3 3HKU 3IBI 3IBL 3IBN 3IBU 456C 4FUA 4TLN 5TLN 6CPA 7CPA 830C 8CPA 966C

Table S4. The detailed docking results for the 106 zinc metalloproteins in the test set performed by $MpSDock_{Zn}$.

PDB ID	Score Best Pose		Score Worst Pose		RMSD Best Pose	
	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)
1B3D	2.5	-438.50	4.3	-40.41	2.4	-228.10
1BN1*	5.2	-172.44	3.4	94.64	1.9	22.08
1BNM*	7.6	-187.75	3.2	2.25	1.3	-56.39
1BNU*	1.5	-279.51	2.5	58.14	0.5	-236.72
1BZM	2.4	-394.30	3.4	56.11	1.3	14.52
1C1U	1.2	-155.65	1.9	25.47	1.2	-155.65
1C2E	2.3	-337.66	4.3	71.43	1.0	-251.28

PDB ID	Score-best Pose		Score-worst Pose		RMSD-best Pose	
	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)
1C2I	1.5	-341.99	5.1	206.13	1.4	-119.05
1C3R	1.8	-237.54	4.6	15.61	1.8	-237.54
1CIM*	1.5	-313.72	3.2	26.33	1.2	-164.74
1CNX	2.0	-439.53	5.4	4.22	1.7	-112.79
1D8F	2.9	-414.77	6.5	-136.61	1.6	-199.57
1DE6	3.3	-358.96	1.9	-80.69	1.2	-228.79
1E47	0.9	-189.20	3.0	46.02	0.9	-189.20
1FBL	6.1	-289.28	3.1	-19.65	1.7	-30.54
1G4J	3.1	-205.84	3.8	44.82	1.5	-140.89
1G9C	1.7	-193.06	2.1	3.76	1.6	-187.75
1GT7	4.9	-159.03	3.5	99.05	1.8	-11.20
1H2B	2.5	-383.94	3.3	37.57	1.9	-75.03
1HEE	3.6	-345.51	3.0	16.42	0.8	-243.67
1HY7	1.7	-236.61	6.0	22.08	1.7	-236.61
1I9Q*	1.6	-47.38	4.1	20.84	1.5	-44.34
1I91*	3.7	-290.66	4.1	31.46	1.6	-85.94
1IF7*	3.9	-417.23	4.4	-8.96	1.5	-229.80
1J37*	2.8	-260.23	3.0	-64.71	0.9	-248.40
1JIZ	1.4	-351.35	4.9	-65.12	1.4	-351.35
1JJT	1.9	-218.90	2.0	-37.66	1.9	-72.31
1LD8	3.9	-271.07	2.8	57.59	1.5	-203.26
1LRY	1.9	-74.68	5.0	-41.44	1.9	-74.68
1MNC	1.8	-229.33	1.6	-85.84	1.6	-85.84
1NVD	5.0	-413.22	4.3	-93.78	2.5	-144.57
1OKM	3.2	-310.17	2.3	-17.67	1.2	-46.38
1P6D	6.6	-317.09	9.1	-23.62	6.6	-317.09
1PE7	4.2	-418.44	1.3	30.15	1.3	30.15
1PE8	3.9	-418.20	2.5	6.19	0.9	-184.06
1Q1Y	1.9	-252.03	3.3	-20.67	1.9	-252.03
1R1I*	1.7	-345.41	1.7	-345.41	1.7	-345.41
1ROS	1.3	-401.27	7.1	36.37	1.3	-401.27
1SA5	5.0	-165.51	2.7	33.07	1.2	-19.59
1T67	2.2	-360.75	4.2	-40.91	1.7	-263.52
1TKF	0.9	-373.13	1.6	72.21	0.9	-373.13
1UZF*	4.4	-271.76	5.1	-58.26	1.1	-191.96
1X8I	5.2	-366.27	4.1	-19.14	1.6	-81.08
1XUG	0.7	-387.17	2.9	-21.56	0.7	-349.34
1YBQ	2.9	-421.52	2.2	-133.94	1.1	-275.69
1YEJ	3.9	-277.83	8.7	46.06	1.2	-59.80

PDB ID	Score-best Pose		Score-worst Pose		RMSD-best Pose	
	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)
1Z9Y	3.1	-435.19	3.5	-14.05	1.5	-156.17
1ZG7*	3.0	-350.98	2.8	71.93	1.8	-88.59
1ZGF	1.6	-320.16	3.0	43.37	0.8	-3.56
1ZSB	0.7	-324.27	2.1	99.03	0.7	-275.63
1ZXC	1.5	-392.13	3.4	-65.24	1.1	-319.27
2AFU	2.0	-288.10	3.9	-52.69	1.1	-168.78
2C6C	1.9	-642.62	1.5	-156.71	1.4	-436.89
2DDF	3.6	-468.13	3.8	-41.72	2.0	-246.55
2DW0	2.8	-246.60	4.5	-9.52	2.8	-246.60
2EG8	3.2	-394.97	3.9	135.57	0.8	-19.89
2F0Y	4.1	-332.95	5.0	-8.27	3.1	-245.95
2FOQ	1.7	-282.94	1.8	-4.71	1.0	-144.52
2FU9*	3.3	-380.07	5.9	15.67	1.2	-276.63
2H4N	1.5	-188.37	1.4	90.19	1.0	12.37
2IEJ	6.8	-327.61	6.7	-69.90	3.0	-171.14
2ISW	3.0	-208.85	2.5	119.94	1.6	-7.76
2NNG	2.0	-382.98	3.8	22.38	1.0	-164.38
2OI0*	3.3	-126.71	6.9	45.88	1.8	-122.33
2OW2	2.0	-394.74	2.5	-10.71	1.0	-89.75
2PIZ	2.0	-516.58	4.5	-127.90	1.8	-416.33
2PJ8	1.6	-437.19	7.0	-163.62	1.6	-437.19
2PJT	1.9	-488.26	6.5	-59.15	1.9	-488.26
2Q1B	1.3	-284.53	2.0	32.26	0.5	-86.94
2QDT*	3.0	-267.79	4.1	126.31	0.5	-108.50
2QVV	1.6	-75.95	1.1	-0.69	1.1	-0.96
2RJP	1.7	-284.62	1.7	-143.26	1.7	-143.26
2V2A	4.0	-204.56	3.9	51.64	1.5	-14.93
2VQO	9.4	-309.70	0.9	45.66	0.9	45.66
2W15	5.0	-299.32	4.2	-73.63	4.2	-104.36
2WEJ	1.4	-449.83	3.2	43.54	0.9	-83.11
2YZ3*	2.5	-514.47	2.6	-64.25	1.3	-271.80
2ZIS	6.5	-263.93	5.9	56.49	1.8	-29.44
3B7U	1.0	-341.71	6.1	-74.74	1.0	-341.71
3BHX	1.8	-590.81	1.6	-158.67	1.0	-404.00
3BL1	7.5	-161.91	3.3	-5.06	1.4	-87.92
3C0Z	2.9	-362.83	1.4	-23.93	1.1	-33.95
3CHP	1.5	-243.78	1.5	-32.48	1.5	-243.78
3D7G	0.9	-454.54	2.4	-259.30	0.9	-440.98
3DAZ	2.6	-281.51	2.4	74.93	1.3	-137.29

PDB ID	Score-best Pose		Score-worst Pose		RMSD-best Pose	
	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)	RMSD(Å)	Score(k_BT)
3DCS	1.4	-274.38	2.3	115.60	1.1	-236.04
3DHB	1.8	-424.55	3.8	3.22	1.8	-424.55
3E8R	1.5	-435.20	1.2	-169.46	1.2	-169.46
3E30	6.4	-359.91	7.2	-70.95	3.3	-72.22
3ELF	4.9	-167.63	5.0	80.24	2.0	-54.62
3F0R	3.7	-249.51	4.2	23.91	1.7	-74.01
3F7B	1.0	-291.81	6.8	31.47	1.0	-291.81
3F18	1.5	-264.26	3.0	-23.24	0.9	-127.51
3FGD	1.7	-409.87	2.3	55.38	1.0	-157.17
3FTX	1.9	-475.12	1.7	-276.71	0.9	-290.26
3FUK	3.8	-257.80	3.9	-178.43	1.7	-225.44
3FX6	1.9	-413.99	3.7	-56.91	1.9	-413.99
3GB6	5.9	-426.96	2.6	69.15	2.6	69.15
3H68	1.6	-569.74	1.4	77.75	1.0	-228.06
3HKN	NA	NA	NA	NA	NA	NA
3HXC	NA	NA	NA	NA	NA	NA
3HY9	1.7	-373.87	5.8	-159.73	1.4	-260.29
3IGP	1.0	-340.16	1.9	27.51	0.9	-104.15
3JVH	4.4	-258.31	4.0	-5.37	1.2	-234.78
4TLN	1.7	-298.95	1.8	-55.52	1.6	-149.82
5TLN	1.8	-177.60	1.9	-158.85	1.6	-168.48

* Zinc metalloproteins involved sulfur ligands.

Table S5. The test set of 15 zinc protein-ligand complexes used in KScore (PDB codes).

1CBX	1MNC	1TLP	1TMN	2TMN	3CPA	3TMN	4TLN	4TMN	5TLN
5TMN	6CPA	6TMN	7CPA	8CPA					

Table S6. The test set of 286 ordinary protein-ligand complexes used in KScore (PDB codes).

10GS 1A46 1A5G 1A69 1A9M 1AAQ 1ABE 1ABF 1AI5 1AJP 1AJQ 1AJV 1AJX 1APB 1B11 1B5G 1B8O 1BA8 1BAP 1BB0 1BCU 1BDQ 1BMA 1BRA 1BV7 1BV9 1BWA 1BWB 1BXO 1C5P 1C5Q 1C5S 1C5T 1C84 1CE5 1CLA 1D09 1D3D 1D3P 1D4K 1D4L 1D4Y 1DHF 1DIF 1DMP 1DR1 1DRF 1E1V 1E66 1ELA 1ELB 1ETR 1ETS 1EXW 1F0T 1F0U 1F4E 1F4F 1F4G 1F5K 1FKB 1FKF 1FKI 1FMO 1G35 1G36 1G3B 1GHZ 1GI1 1GI6 1GJ6 1GNM 1GNN 1GNO 1GPK 1H23 1H4W 1HBV 1HEG 1HI4 1HIH 1HOS 1HPO 1HPS 1HPV 1HPX 1HSH 1HVH 1HVI 1HVJ 1HVK 1HVL 1HVR 1HVS 1HWR 1HXB 1HXW 1INC 1J14 1J16 1J17 1JQD 1JQE 1K1I 1K1J 1K1L 1K1M 1K1N 1K4G 1K9S 1KV5 1L2S 1LOL 1M0N 1M0Q 1M2Q 1MES 1MET 1MEU 1MQ6 1MTR 1N2V 1NC1 1NFY 1NHU 1NNY 1NVQ 1NWL 1O2H 1O2J 1O2K 1O2N 1O2O 1O2Q 1O2S 1O2W 1O2X 1O2Z 1O30 1O33 1O36 1O38 1O3D 1O3F 1O3H 1O3I 1O3J 1O3K 1O3P 1ODY 1OM1 1OSS 1OYQ 1PB9 1PBQ 1PPC 1PPH 1PPM 1PR5 1PRO 1PXO 1QB1 1QB6 1QB9 1QBN 1QBO 1QBR 1QBS 1QBU 1RBP 1RE8 1RGK 1RGL 1RNT 1S39 1SBG 1SL3 1SQA 1SRE 1SV3 1SYH 1TET 1THA 1TNG 1TNH 1TNI 1TNJ 1TNK 1TNL 1TRD 1TX7 1U2Y 1U33 1UTJ 1UTL 1UTM 1UTN 1UTO 1UTP 1UWT 1V2J 1V2K 1V2L 1V2N 1V2O 1V2Q 1V2R 1V2S 1V2T 1V2U 1V2W 1V48 1VFN 1VZQ 1W5V 1X1Z 1XD1 1XGJ 1Y1M 1Y6Q 1YDT 1YYY 1ZOE 1ZZZ 2AOU 2AYR 2B1V 2B7D 2BOK 2BPV 2BPY 2BQV 2BRB 2BRM 2BZ6 2BZA 2CEQ 2CGR 2CSC 2D3U 2D3Z 2F80 2FAI 2FDP 2FLB 2FX6 2FZC 2G94 2GBP 2GSS 2HS1 2HS2 2I0D 2QWB 2QWC 2QWD 2QWE 2QWF 2QWG 2SNS 2STD 2XIM 2XIS 3CLA 3FX2 3GSS 3PTB 4CLA 4FIV 4TIM 4XIA 5ABP 5CNA 5ER1 6ABP 6FIV 6RNT 6STD 6TIM 7ABP 7EST 7TIM 7UPJ 8ABP 9AAT 9ABP

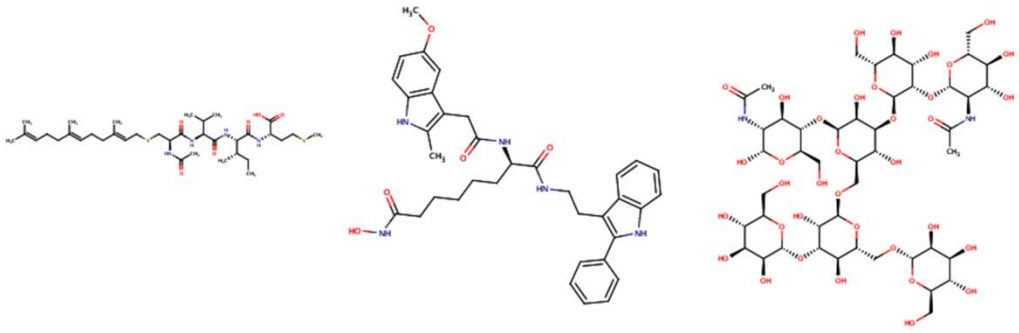

Figure S1. The structures of example ligands whose binding conformations were failed to be reproduced in iteration optimization process. These ligands are from PDBs (from left to right): 1O1T, 2V5X and 3CZN.