Supporting Information

Ligand-Binding Site Structure Refinement Using Molecular Dynamics with Restraints Derived from Predicted Binding Site Templates

Hugo Guterres[1], Hui Sun Lee[1], and Wonpil Im[1,2]*

[1]Department of Biological Sciences, Lehigh University, Bethlehem PA, 18015 USA
[2]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

Corresponding Authors
*wonpil@lehigh.edu

Table S1. Predicted protein structures obtained using I-TASSER with benchmark setting that exclude homologous templates with greater than 30% sequence identity. Protein sequences less than 300 amino acids are obtained from the Astex list. Their Cα-RMSD and I-TASSER confidence scores are listed.

No.	PDB	Residues	RMSD	C-score*
1	1towA	131	1.20	1.47
2	1lpzB	234	1.69	0.32
3	1tz8A	114	1.69	1.04
4	1g9vA	141	1.72	1.23
5	1r55A	203	1.90	1.32
6	1hq2A	158	2.18	1.15
7	1oweA	245	2.44	0.98
8	1q1gA	243	2.51	-0.16
9	1tt1A	251	2.53	0.82
10	1ke5A	298	2.54	0.76
11	1oytH	257	2.60	0.85
12	1p62B	241	2.64	-0.04
13	1jjeA	220	2.77	1.08
14	2bm2A	242	3.16	0.85
15	1ygcH	254	3.18	0.30
16	1s3vA	186	3.22	0.93
17	1k3uA	268	3.34	1.07
18	1u1cA	253	3.35	0.47
19	1unlA	292	3.41	0.40
20	1v0pA	286	3.68	0.56
21	1w2gA	208	3.73	1.25
22	1sqnA	251	4.02	-0.20
23	2br1A	271	4.10	-1.74
24	1m2zA	255	4.12	0.03
25	1s19A	253	4.14	0.45
26	1ia1A	192	4.41	0.65
27	1v48A	283	4.51	-0.33
28	1u4dA	273	4.53	0.41
29	1n46A	251	4.60	0.27
30	1gkcA	163	4.63	-0.60
31	1z95A	246	5.15	0.06
32	1kzkA	99	5.52	0.61

33	1sj0A	245	6.21	0.85
34	1ig3A	254	6.30	1.01
35	1hnnA	261	6.39	-1.14
36	1navA	263	6.45	-0.40
37	1hvyA	288	7.24	0.02
38	2bsmA	208	8.76	0.38
39	1sg0A	230	9.41	0.01
40	1lrhA	160	10.96	-0.52
	Avg.	*229.3*	*4.17*	*0.41*

**C-score is a confidence score from I-TASSER predictions that ranges from -5 to 2*

Table S2. Representative structures used for force constants optimization without Cα positional restraint. Cα -RMSD (Å) of the model structures from their respective experimental structures are shown for both the protein and their ligand-binding sites (LBS). Improved cases with lower RMSD are highlighted in bold.

	Initial Models		Refined_1.0Fc		Refined_0.5Fc		Refined_0.1Fc	
PDB	Protein	LBS	Protein	LBS	Protein	LBS	Protein	LBS
1towA	1.20	0.87	1.26	**0.71**	**1.19**	**0.78**	1.29	1.03
1g9vA	1.72	1.38	2.16	**0.65**	2.60	**0.66**	2.48	**1.04**
1r55A	1.90	0.88	3.54	**0.38**	4.70	**0.41**	3.54	**0.59**
1hq2A	2.18	2.01	**2.12**	**0.36**	**2.06**	**0.47**	**1.98**	**0.77**
2bm2A	3.16	3.02	4.54	**0.41**	3.91	**0.61**	4.71	**1.63**
1ygcH	3.18	2.80	3.74	**0.39**	3.61	**0.63**	3.54	**1.17**
1s19A	4.14	1.96	4.31	**0.53**	4.51	**0.96**	4.96	**1.37**
1gkcA	4.63	2.51	5.14	**0.36**	5.19	**0.51**	5.90	**1.31**
1z95A	5.15	1.32	**4.78**	**0.27**	5.00	**0.41**	**4.94**	**0.67**
Average	3.03	1.86	3.51	**0.41**	3.64	**0.60**	3.70	**1.06**

Table S3. Representative structures used for force constants optimization with a Cα positional restraint force constant of 0.5 kcal/(mol·Å^2). Cα -RMSD (Å) of the model structures from their respective experimental structures are shown for both the protein and the ligand-binding sites (LBS). Improved cases with lower RMSD are highlighted in bold.

	Initial Models		Refined_1.5Fc		Refined_1.0Fc		Refined_0.5Fc	
PDB	Protein	LBS	Protein	LBS	Protein	LBS	Protein	LBS
1towA	1.20	0.87	**1.09**	**0.66**	**1.08**	**0.71**	**0.96**	**0.69**
1g9vA	1.72	1.38	**1.24**	**0.34**	**1.27**	**0.40**	**1.47**	**0.60**
1r55A	1.90	0.88	**1.85**	**0.28**	**1.89**	**0.37**	**1.83**	**0.37**
1hq2A	2.18	2.01	**1.66**	**0.41**	**1.62**	**0.53**	**1.68**	**0.77**
2bm2A	3.16	3.02	**2.92**	**0.48**	**3.01**	**0.55**	**3.01**	**0.96**
1ygcH	3.18	2.80	**3.03**	**0.36**	**3.00**	**0.53**	**3.02**	**0.45**
1s19A	4.14	1.96	**4.06**	**0.67**	**4.04**	**0.70**	**4.11**	**1.03**
1gkcA	4.63	2.51	**4.44**	**0.28**	**4.45**	**0.56**	**4.50**	**0.69**
1z95A	5.15	1.32	**4.67**	**0.33**	**4.63**	**0.35**	**4.65**	**0.37**
Average	3.03	1.86	**2.77**	**0.42**	**2.78**	**0.52**	**2.80**	**0.66**

Table S4. Overview of refinement targets from 40 targets with bound ligands. Ligands were docked to the initial unrefined structures using AutoDock Vina. Group 1 (1-5), group 2 (6-21), group 3 (22-32), and group 4 (33-40). Ligand RMSDs relative to the experimental structures are shown. Improved cases with lower RMSD are highlighted in bold.

		LBS RMSD (Å)			Protein RMSD (Å)			Ligand RMSD (Å)		
No	PDB	Init.	Ref.	Cont.	Init.	Ref.	Cont.	Init.	Ref.	Cont.
1	1towA	0.87	**0.34**	1.93	1.20	**0.95**	1.42	3.73	4.10	5.57
2	1lpzB	0.73	**0.51**	0.93	1.69	**1.66**	**1.65**	6.60	7.59	8.60
3	1tz8A	0.50	**0.49**	0.52	1.69	**1.61**	1.70	3.17	**1.35**	3.42
4	1g9vA	1.38	**0.58**	1.84	1.72	**1.51**	1.95	6.95	**5.94**	**6.13**
5	1r55A	0.88	**0.71**	**0.54**	1.90	**1.81**	**1.85**	3.42	5.16	5.27
Group 1 avg.		*0.87*	***0.53***	*1.15*	*1.64*	***1.51***	*1.71*	*4.77*	*4.83*	*5.80*
6	1hq2A	2.01	**1.35**	**2.00**	2.18	**1.80**	2.21	6.17	6.52	6.45
7	1oweA	1.44	**0.62**	1.54	2.44	**2.26**	2.46	8.41	8.74	9.22
8	1q1gA	1.16	1.37	1.45	2.51	**2.43**	2.59	3.32	3.67	5.34
9	1tt1A	1.07	**0.50**	1.24	2.53	**2.32**	**2.32**	2.64	3.09	**2.35**
10	1ke5A	1.15	**0.87**	**0.74**	2.54	**2.41**	**2.39**	4.51	**3.58**	1.50
11	1oytH	3.04	**0.52**	3.58	2.60	**2.30**	2.63	7.41	**6.81**	9.07
12	1p62B	1.70	**1.42**	1.90	2.64	**2.51**	**2.58**	2.50	**2.35**	6.87
13	1jjeA	2.42	**2.38**	2.87	2.77	**2.66**	2.81	5.25	**5.21**	5.30
14	2bm2A	3.02	**1.32**	3.14	3.16	**3.01**	3.28	8.32	**5.60**	9.76
15	1ygcH	2.80	**0.64**	2.99	3.18	**3.08**	3.24	6.54	**6.28**	12.31
16	1s3vA	1.70	**0.68**	1.96	3.22	**3.09**	3.23	7.24	**6.68**	**7.02**
17	1k3uA	5.80	**0.50**	**5.52**	3.34	**2.00**	**3.21**	3.65	4.59	7.53
18	1u1cA	5.08	**3.38**	**4.76**	3.35	**3.10**	**3.31**	2.76	**2.45**	**1.73**
19	1unlA	1.27	**0.94**	2.30	3.41	**3.39**	3.45	4.48	6.69	8.13
20	1v0pA	1.94	**1.40**	1.94	3.68	**3.64**	3.78	7.50	8.32	9.47
21	1w2gA	1.54	**0.96**	2.19	3.73	3.80	3.90	4.91	**4.14**	5.76
Group 2 avg.		*2.32*	***1.18***	*2.51*	*2.96*	***2.74***	*2.96*	*5.35*	***5.30***	*6.74*
22	1sqnA	0.95	**0.91**	0.97	4.02	4.03	**3.99**	6.51	**6.31**	6.52
23	2br1A	1.11	**0.91**	**0.99**	4.10	**4.06**	**4.04**	5.83	5.88	5.99
24	1m2zA	1.42	**1.39**	1.63	4.12	**4.04**	4.14	7.21	**5.65**	**7.02**
25	1s19A	1.96	**0.74**	1.97	4.14	**4.09**	**4.12**	3.13	**2.87**	**3.11**
26	1ia1A	1.81	**1.54**	2.20	4.41	**4.33**	4.46	6.19	**6.00**	6.29
27	1v48A	3.52	**1.06**	2.97	4.51	**4.33**	4.59	6.58	**5.94**	6.60
28	1u4dA	1.29	**0.85**	1.42	4.53	**4.46**	4.58	5.81	**4.78**	6.81
29	1n46A	1.62	**1.21**	1.20	4.60	**4.55**	4.67	3.82	5.07	**3.53**
30	1gkcA	2.51	**1.32**	3.14	4.63	**4.44**	4.76	6.09	**5.87**	8.31

31	1z95A	1.32	**1.10**	**1.20**	5.15	**4.65**	**4.59**	4.92	**4.20**	**3.67**
32	1kzkA	7.83	**1.90**	8.32	5.52	**3.72**	5.63	7.36	12.30	14.05
Group 3 avg.		*2.30*	***1.18***	*2.36*	*4.52*	***4.25***	***4.51***	*5.77*	*5.90*	*6.54*
33	1sj0A	3.13	**1.73**	3.01	6.21	**5.89**	**6.20**	2.10	**0.83**	**1.23**
34	1ig3A	1.86	2.01	**1.49**	6.30	**5.62**	**5.64**	3.80	19.60	22.76
35	1hnnA	5.52	**5.44**	5.75	6.39	**6.37**	6.42	4.84	6.21	5.79
36	1navA	1.92	**1.51**	2.15	6.45	**6.05**	**6.11**	7.37	**4.34**	**3.06**
37	1hvyA	8.07	**7.80**	8.12	7.24	**7.22**	7.26	6.14	11.40	20.96
38	2bsmA	3.97	**2.60**	4.06	8.76	**8.68**	8.89	5.95	6.63	6.52
39	1sg0A	1.27	**1.10**	1.38	9.41	9.45	9.44	5.61	8.18	9.27
40	1lrhA	7.58	7.77	**7.49**	10.96	10.97	11.10	3.15	3.59	**2.87**
Group 4 avg.		*4.17*	***3.75***	*4.18*	*7.72*	***7.53***	***7.63***	*4.87*	*7.60*	*9.06*
Total avg.		*2.50*	***1.61***	*2.63*	*4.17*	***3.96***	*4.16*	*5.30*	*5.86*	*7.03*

***Cont.** = simulations without distance restraints and with positional restraints.*

Using initial unrefined structures with bound crystal ligands, we ran the same restrained MD simulations for all 40 targets. The overall results, in terms of structure refinements, show similar values as observed in the main text, refinement without bound ligand. The average improvement is 0.89-Å, a slight worse result than observed in the refinement without bound ligand, which has 0.90-Å average improvement. Although, one might expect that the docked ligand should improve structure refinement, most of the docked ligands have high ligand RMSDs relative to the crystal structures (average 5.30 Å) that do not help local structure refinement (**Table S4**).

The bound ligands have 50-ns simulation time to adjust their binding poses. Although, some ligands manage to slightly improve their binding poses, most ligands stay in similar binding modes as compared to their initial structures. A few ligands that are bound in opened and shallow binding sites, end up leaving the binding site during the simulation time. They are shown in model 32, 34, 37 and 15. A lot of the improvements in the ligand RMSD are not significant, because the initial docked ligands have incorrect positioning of ligand moieties in the binding site. For example, in model 11, the ligand is incorrectly docked with its benzamidine moiety in the place its fluorobenzyl group when compared to the crystal structure (**Figure S4A**). During the simulation the ligand does not undergo major conformational change to correct this binding mode. As a result, the ligand RMSD stays about the same (from 7.41 to 6.81 Å). Overall, we observe that most ligands have very limited space to move in buried binding sites during the short simulation time.

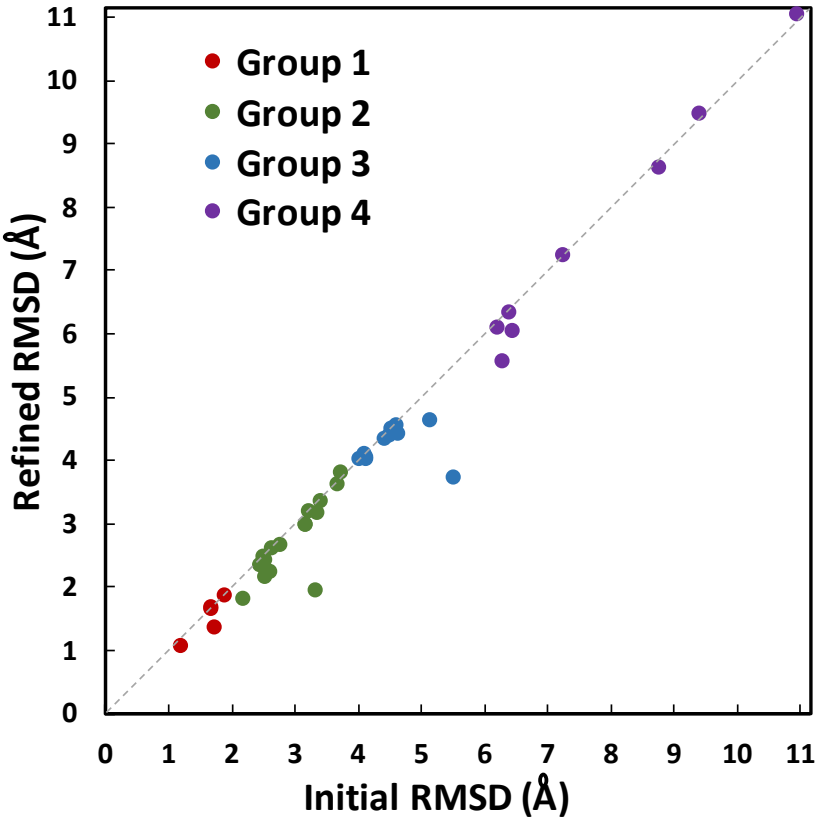

Figure S1. Protein Cα- RMSD values comparing initial unrefined and refined model structures against the experimental structures. The structures are separated into 4 groups based on their initial unrefined protein model RMSD relative to the experimental structures: group 1 (1-2 Å, red), group 2 (2-4 Å, green), group 3 (4-6 Å, blue), and group 4 (>6 Å, purple). The average improvement for protein Cα RMSD is 0.21 Å. This shows that a targeted protein structure refinement that focuses on protein ligand-binding site can consistently improve the overall protein structure.

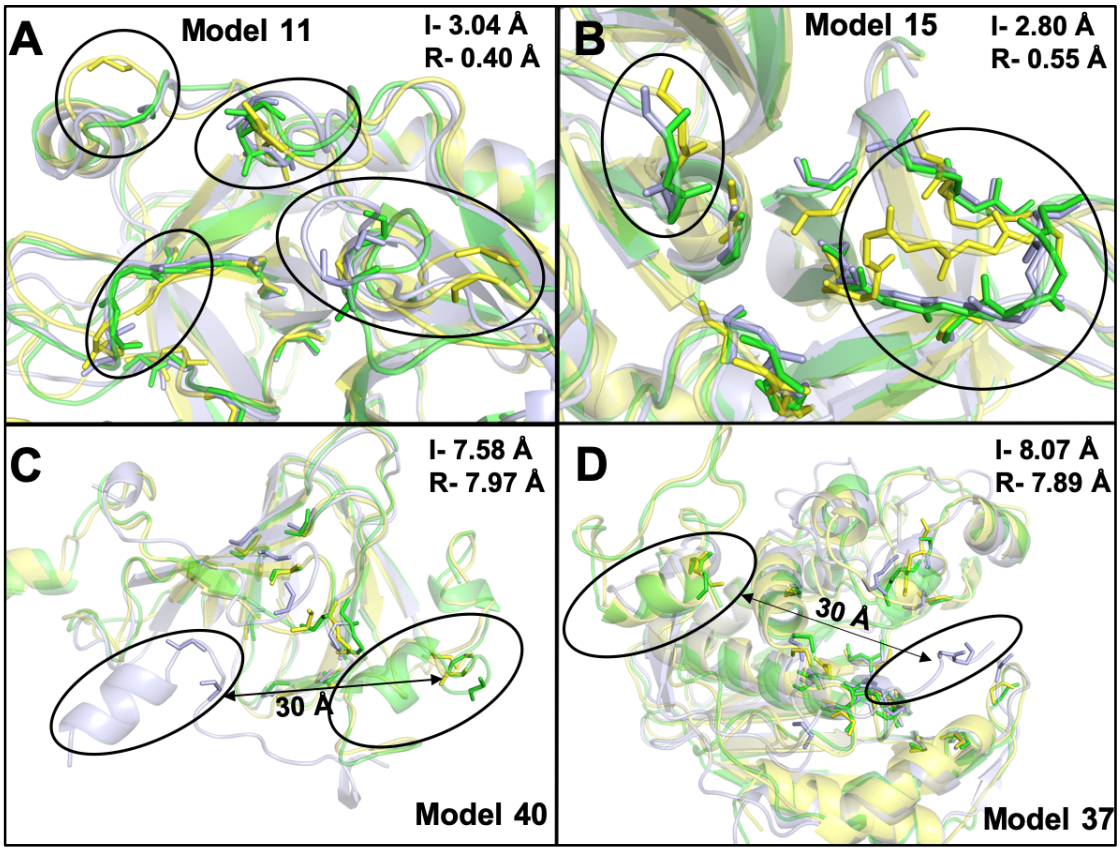

Figure S2. Aligned structures of initial unrefined models (yellow) and refined models (green) on the crystal structures (light blue). The backbones of ligand-binding site residues are shown in sticks. I- is the initial unrefined structure RMSD and R- is the refined structure RMSD relative to the crystal structure. Structures discussed in the text are shown, (A) Model 11- thrombin (PDB 1oyt). (B) Model 15- serine protease factor VII (PDB 1ygc). (C) Model 40- auxin-binding protein (PDB 1lrh). (D) Model 37- thymidylate synthase (PDB 1hvy).

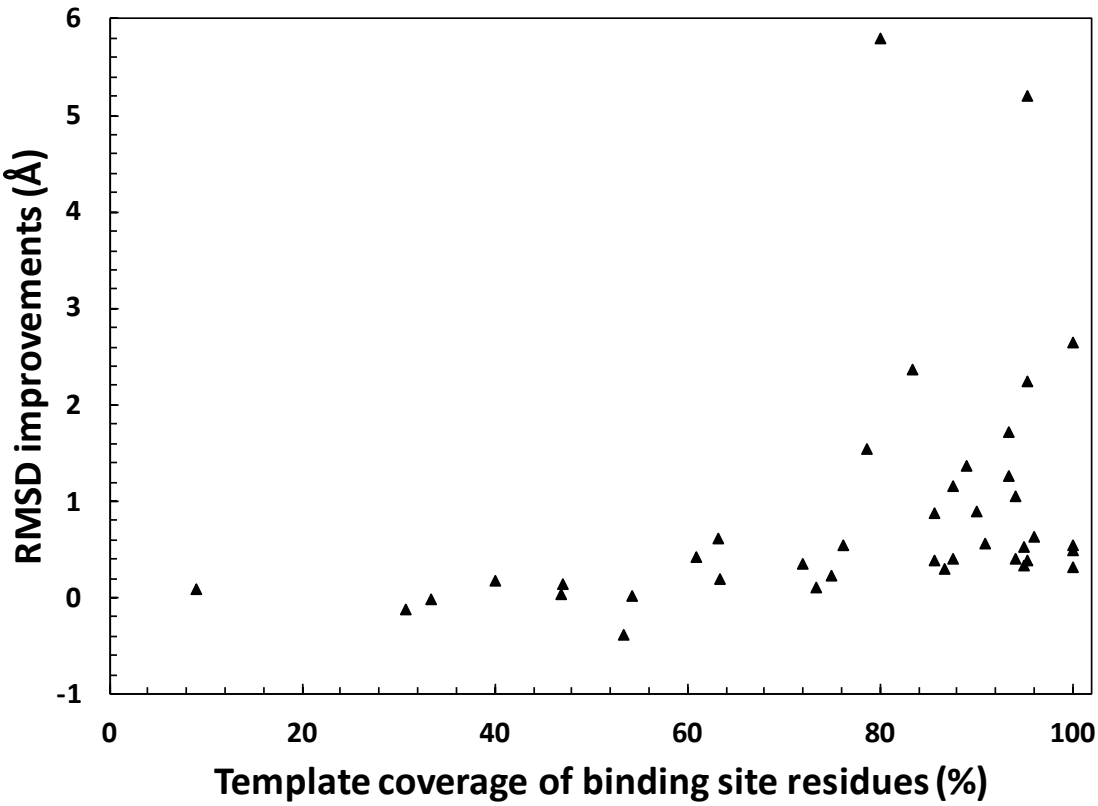

Figure S3. Ligand-binding site Cα RMSD improvements as a function of template coverage of binding site residues. Successful refinements are observed consistently with templates that have greater than 60% binding site residues coverage.

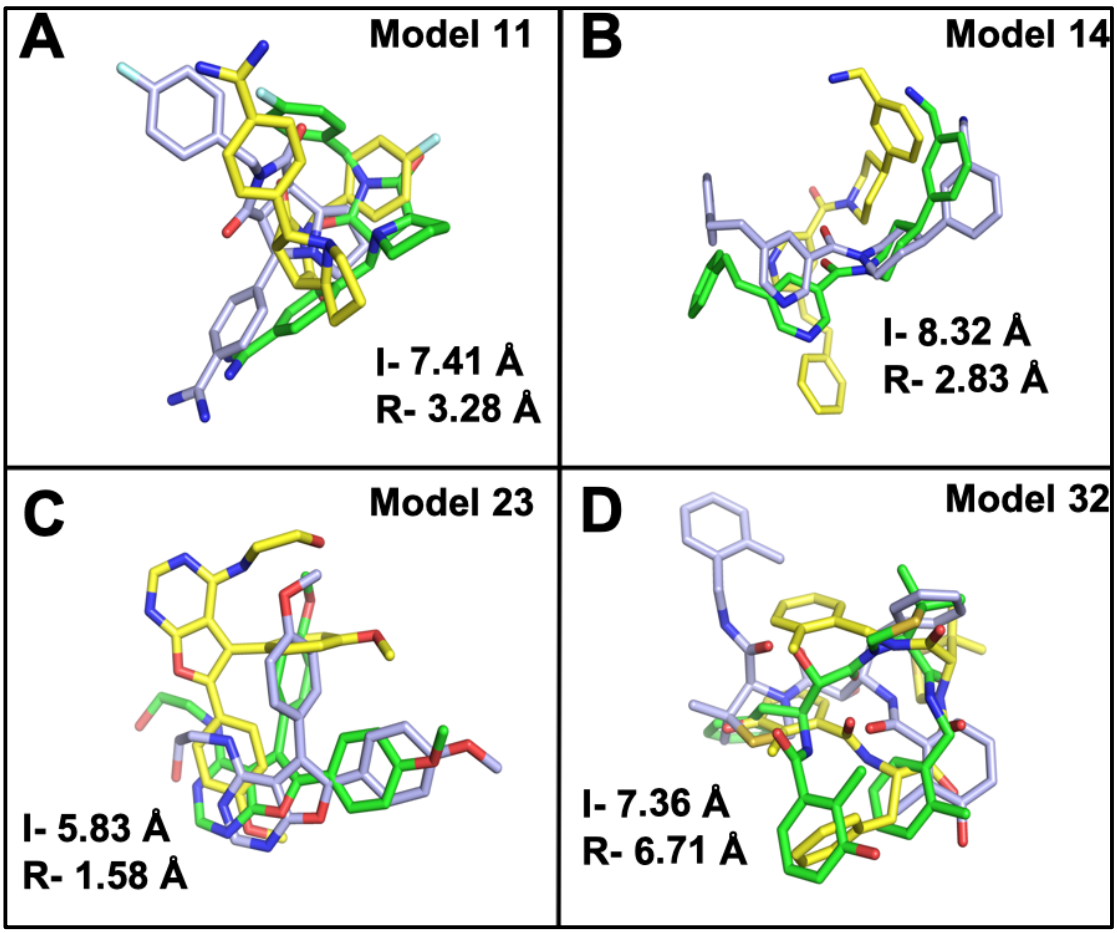

Figure S4. Ligand binding modes obtained from docking are compared to the crystal structures (light blue). Ligand poses docked to the initial unrefined structures (yellow) and to the refined structure (green). I- is the initial unrefined structure ligand RMSD, and R- refers to refined structure ligand RMSD. Ligands discussed in the text are shown, (A) model 11- thrombin (PDB 1oyt). (B) Model 14- human beta2 tryptase (PDB 2bm2). (C) Model 23- serine/threonine protein kinase chk1 (PDB 2br1). (D) Model 32- HIV protease (PDB 1kzk)