Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

SUPPLEMENTARY MATERIAL

Fragment Dissolved Molecular Dynamics: A systematic and efficient method to locate binding sites

Cristian Privat[1], José M. Granadino-Roldán[2], Jordi Bonet[1], Maria Santos Tomas[3], Juan J. Perez[4], Jaime Rubio-Martinez[*1]

[1] *Departament de Ciència dels Materials i Química Física, Universitat de Barcelona (UB) and the Institut de Quimica Teorica i Computacional (IQTCUB), Martí i Franqués 1, 08028 Barcelona, Spain*

[2] *Departamento de Química Física y Analítica, Facultad de Ciencias Experimentales, Universidad de Jaén, Campus "Las Lagunillas" s/n, 23071, Jaén, Spain*

[3]*Department of Architecture Technology, Universitat Politecnica de Catalunya, Av. Diagonal 649, 08028 Barcelona, Spain*

[4]*Deparment of Chemical Engineering, Universitat Politecnica de Catalunya, Av. Diagonal 647, 08028 Barcelona, Spain*

Detailed description of *step 1*

Our method starts by constructing a box containing one specific ligand embedded in waters that fill the entire volume of that box. We name this box as LIGAND_BOX. We have a box for each ligand and each LIGAND_BOX has a different size because it depends on the size of the ligand. The size of the box is decided by the LEaP module of the Amber suite using as an external parameter "*the minimum distance from the edge of the BOX to the ligand* (d_min)". Once this parameter is selected, the LEaP module constructs automatically the box (its dimensions: ([x_box, y_box, z_box]) and fills it with waters (see Figure Description 1).

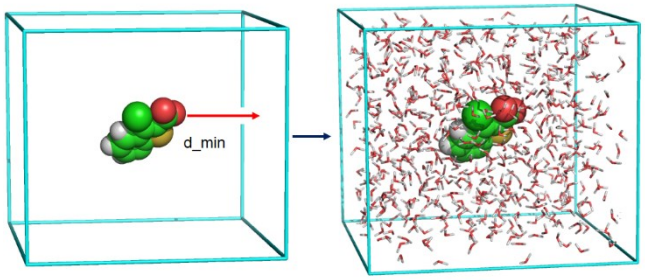


Figure Description 1. Example of the addition of water molecules in the BOX by the LEaP module of AMBER.

In particular the commands used in the LEaP module were:

ligand = loadpdb name_of_my_ligand.pdb # Here we read the ligand alone
solvatebox ligand TIP3PBOX d_min 1.0 # Here we define the parameters for the LIGAND_BOX construction.

Where d_min 1.0 implies that "if any inserted water molecule has a distance to any atom of the ligand lower than 1.0 Å the water molecule is removed". This parameter is intended to avoid bad clashes between the ligand and the waters we are adding to the protein. The TIP3PBOX is placed as many times as necessary to fill the entire defined box (see Figure Description 1). The information about each of the constructed LIGAND_BOXes is described in Table S1.

Detailed description of *step 3*

To construct all the selected systems (protein+ligands+water), we start with the prepared protein alone (step 2, see Figure 1). Then, we use the LEaP module to "solvate with ligands" the protein. The procedure is similar to the previous preparation of the LIGAND_BOX, although in this case, instead of using the TIP3PBOX to solvate, we use the LIGAND_BOX generated in the first step (see Figure Description 2). Each of the proteins studied is solvated with its corresponding ligand. For example, in the case of the endothiapepsin protein, we generated four different systems using four different ligand boxes corresponding to the four ligands studied. We used the same commands as before:

protein = loadpdb name_of_my_protein.pdb # Here we read the protein alone
solvatebox protein LIGAND_BOX d_min 1.0 # Here we define the parameters for the PROTEIN_BOX construction.

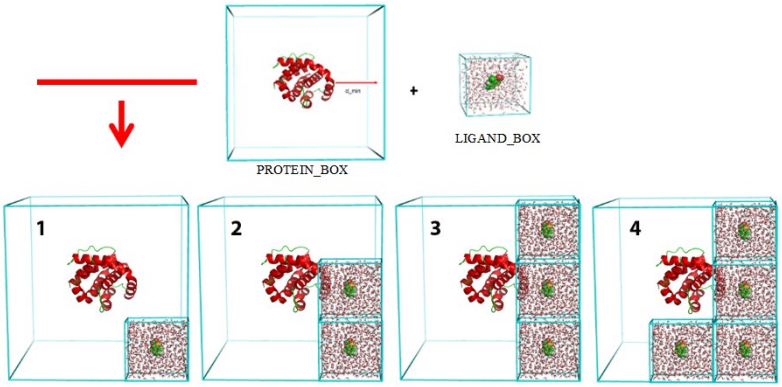


Figure Description 2. Process to solvate the protein with boxes of LIGAND_BOX. LEaP will surround the protein with as many LIGAND_BOXes as needed to fill the complete defined space. That number will depend on the size of the LIGAND_BOX and on the size of the PROTEIN_BOX.

The solvatebox command defines the size of the box for each protein. The information about each of the constructed protein systems is described in Table S2.

Table S1. Dimension of the LIGAND_BOX in each direction (X_{BOX}, Y_{BOX} and Z_{BOX}), number of water molecules ($\#_{WAT}$) at the end of the solvation process and distance values ($d_{\_min}$) used to solvate the ligands.

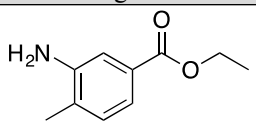
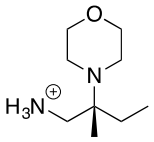
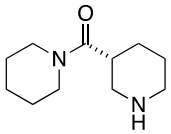
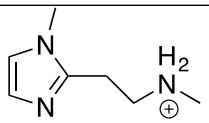
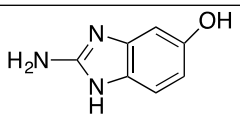
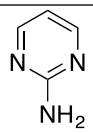
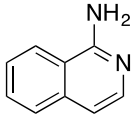
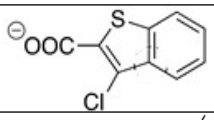
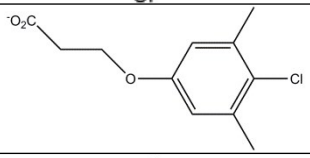
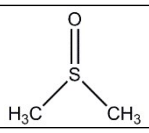
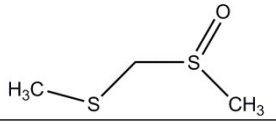
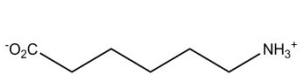
	Ligand	X_{BOX} (Å)	Y_{BOX} (Å)	Z_{BOX} (Å)	$\#_{WAT}$	d_{min} (Å)
SET I		44.537	41.891	38.617	1635	15.0
		41.545	42.118	40.136	1611	15.0
		43.508	43.772	42.040	1812	15.0
		42.113	41.432	38.379	1502	15.0
SET II		41.123	42.620	38.379	1515	15.0
		40.176	38.307	40.136	1406	15.0
		39.456	42.620	40.862	1562	15.0
SET III		34.207	31.785	28.109	641	10.0
		34.986	32.158	31.855	777	10.0
SET IV		34.195	34.807	35.600	1429	15.0
		34.122	37.852	35.524	1510	15.0
		38.535	35.785	37.745	1700	15.0

Table S2. Dimension of the PROTEIN_BOX in each direction (X_{BOX},Y_{BOX} and Z_{BOX}), number of water molecules ($\#_{WAT}$) and number of ligands (#LIG) at end of the solvation process and distance values (d_{min}) used to solvate with the LIGAND_BOX the proteins.

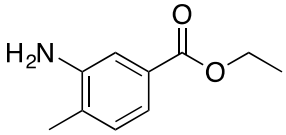
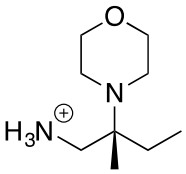
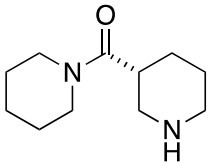
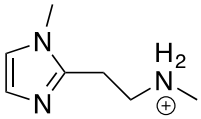
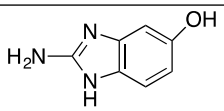
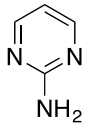
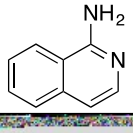
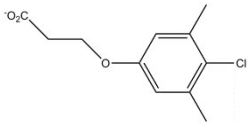
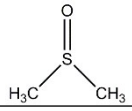
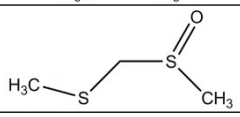
	Ligand	PDB ID	#LIG	DIM (Å)	$\#_{WAT}$	d_{min} (Å)
SET I		5DPZ	9	127.568 120.117 135.738	56618	17.5
		5DR1		118.885 113.045 126.581	45142	13.0
		4Y3W	12	118.941 112.160 126.354	44296	13.0
		5DR1		116.869 111.003 124.469	44268	12.0
		4Y3T	12	122.534 116.935 130.727	50176	15.0
		5DR1		122.930 117.004 130.592	50359	15.0
		4YD6	8	114.811 106.843 122.700	39260	11.0
		5DR1	12	119.716 114.039 127.610	46294	13.5
SET II		1FV9[24]	26	87.650 96.568 81.407	17562	16.0
		2JJC[25]	25	113.284 110.627 109.948	36531	13.0
		2OHK[26]	25	103.592 114.290 96.099	29937	20.0
SET III		4OQ5[27]	20	87.328 85.653 86.177	17132	20.0
			26	87.126 85.614 86.423	17004	20.0
SET IV		1D6O	8	98.123 85.025 84.719	19266	22.0
			6	98.082 85.336 84.557	19363	22.0
		1PMK	8	85.227 78.223 82.684	14709	22.0

Table S3. Results calculated for reactive trajectories of set I. Pockets are defined in reference 32. MMGBSA and K_{DEEP} energies in kcal mol^{-1}.

	System	MD run number														
		1				2				3				4		
Natural Receptor	f031	Binding site Pocket	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site Pocket	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site Pocket	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site Pocket	RT[a]	MMGBSA[b]	K_{DEEP}[c]
		S1'	50	-9.6	-5.8	S1'	180	-12.2	-6.4	S1'	125	-15.1	-5.8	S1'	100	-11.8	-5.9
		Pocket S6	75	-10.3	-7.2									Pocket S6	50	-10.6	-6.6
	f035									Pocket S3	120	-24.4	-5.46	Pocket S6	20	-11.5	-4.23
	f207	Dyad[d]	40	-11.9	-5.72	Dyad	180	-18.9	-6.1					Dyad	20	-10.4	-5.4
						Pocket S6	80	-9.6	-5.2					Pocket S6	75	-12.2	-4.8
	f240	Pocket S6	40	-8.7	-4.4	Pocket S1	140	-15.5	-5.82	Pocket S1	50	-18.3	-5.93				
5DR1 Receptor	f031_5DR1	Pocket S1'	180	-12.4	-4.6	Pocket S1'	40	-11.8	-5.3	Pocket S1'	120	-11.8	-4.71	Pocket S1'	25	-11.6	-6.2
						Pocket S6	20	-10.7	-6.4								
	f035_5DR1	Pocket S3	160	-29.3	-5.8	Pocket S6	180	-15.7	-5.2	Pocket S3	170	-27.2	-6.3	Pocket S3	60	-31.6	-5.3
	f207_5DR1	Dyad[d]	40	-18.5	-5.8	Dyad	170	-13.9	-5.9	Dyad	140	-13.4	-6.2	Dyad	70	-15.5	-3.2
						Pocket S6	40	-11.4	-5.6								
	f240_5DR1	Pocket S1	150	-14.4	-6.4	Pocket S1	60	-17.3	-6.1	Pocket S1	80	-14.4	-6.0	Pocket S1	40	-15.7	-6.1

[a] In ns. [b]MMPBSA binding energy. [c]K_{DEEP} binding energy.[d] Catalytic dyad

Table S4. Descriptors calculated for reactive trajectories of systems f031_5DR1 and f035_5DR1 of set I after 300 ns of MD. Pockets are defined in reference 32. MMGBSA and K_{DEEP} energies in kcal mol^{-1}.

System[a]	Lenght[b]	Binding site	Ave. MMGBSA[c]	Ave. K_{DEEP}[d]	Best MMGBSA	Best K_{DEEP}	N. react.[e]	Best RT[b]
f031	200	Pocket S1´	**-11.9**	**-5.2**	**-12.4**	-6.2	**4**	**180**
		Pocket S6	-2.7	-1.6	-10.7	**-6.4**	1	20
	300	Pocket S1´	**-2.7**	**-1.2**	**-10.6**	**-5.6**	**1**	**125**
f035	200	Pocket S3	**-22.0**	**-4.3**	**-31.6**	**-6.3**	**3**	**170**
		Pocket S6	-3.9	-1.3	-15.7	-5.2	1	180
	300	Pocket S3	**-20.2**	**-2.7**	**-44.8**	**-5.5**	**2**	**270**

[a] Using the 5DR1 receptor. [b] In ns. [c]Average MMGBSA binding energy (see Equation 1). [d] Average K_{DEEP} binding energy. [e] Number of reactive trajectories interacting with that binding pocket. In bold the best value for each descriptor

Table S5. Results calculated for reactive trajectories of systems f031_5DR1 and f035_5DR1 of set I after 300 ns of MD. Pockets are defined in reference 32. MMGBSA and K_{DEEP} energies in kcal mol^{-1}.

System	MD run number															
	1				2				3				4			
	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]
f031_5DR1													Pocket S1'	125	-10.6	-5.6
f035_5DR1	Pocket S3	260	-44.8	-5.3					Pocket S3	270	-36.1	-5.5				

[a] In ns. [b]MMPBSA binding energy. [c]K_{DEEP} binding energy.[d] Catalytic dyad

Table S6. Definition of binding sites. Each binding site is composed of those residues that are within 3.6 Å from the ligand in the last snapshot of the molecular dynamics run.

System	Binding site							
1fv9	Experimental	ASP191	SER192	CYS193	VAL215	GLY220	GLY228	
	BS 1	TYR126	ASP128	ARG233	SER235	HIS236		
	BS 2	GLU166	TYR172	PRO187	PRO228			
	BS 3	HIS84	LYS85	TRP240	HIS244			
	BS 4	MET156	THR157	ASP186	THR191			
	BS 5	THR28	TYR29	ASN67	THR150			
	BS 6	ARG21	ARG23	ASP56				
	BS 7	LYS75	SER105	LYS106				
	BS 8	TYR29	TYR150	GLN197	SER198			
2jjc	Experimental	SER52	ALA55	ASP93	GLY97	MET98	LEU107	THR184
	BS 1	ILE43	SER47	LEU54	LEU60	ARG166		
2ohk	Experimental	ASP32	GLY34	TYR71	ILE118	ASP228	GLY230	THR231
	S1	GLY11	GLN12	GLY13	LEU30	TYR71	PHE108	ILE110
		TRP115	ILE118	GLY230	THR232			
	BS 1	THR26	GLN25	ARG50	TYR51	GLN53		
	BS 2	PHE159	HIS362					
	BS 3	PHE47	PHE109	ASN111	SER113			
	BS 4	ARG7	ASN5	TYR15	GLU17	HIS89		
	BS 5	ARG54	GLN55	ASP62	GLU79			
	BS 6	LEU213	LYS214	LYS246	SER247			
class1 & 2	BS 1	ARG233	LYS234	THR280				
	BS 2	ARG207	LYS208	GLU211				
	BS 3	ASP172	GLU173	LEU174	ARG303	VAL307		
	BS 4	ARG187	LYS197	LYS279	GLU284	GLU288		
	BS 5	SER247	ARG248	ILE251	ASP296	ARG300		
	BS 6	ARG184	LYS194	PRO198	MET199	ARG207	ARG214	
	BS 7	GLY219	ARG222	ASP223				
	BS 8	GLN221	ARG222	GLU225	PHE273	LYS276		
	BS 9	GLN229	ARG233	LYS238	HIS277			
	BS 10	ARG207	LYS208	ARG215	HIS320			
	BS 11	THR259	ARG263	LYS302	VAL316			
	BS 12	GLU180	ARG184	ARG187				
	BS 13	LYS234	ARG248					
	BS 14	LYS244	CYS286					
	BS 15	LEU235	ASP242	SER245	LEU246			

Table S7. Results calculated for reactive trajectories of set II. MMGBSA and K_{DEEP} energies in kcal mol^{-1}. Pockets are defined in Table S6 and reference 57.

Simulation time	System	MD run number															
		1				2				3				4			
		Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]
200 ns	1fv9	Experimental	190	-17.8	-4.8	Experimental	170	-15.7	-4.7	Experimental	151	-13.0	-5.2	BS 3	138	-17.0	-4.6
		BS 1	190	-22.5	-5.2	BS 4	30	-12.1	-5.0	BS 1	118	-25.4	-5.7	BS 8	38	-11.2	-5.8
		BS 2	40	-12.3	-5.6	BS 5	152	-13.1	-5.0	BS 7	54	-10.9	-5.7				
		BS 3	120	-16.9	-4.9	BS 6	49	-16.0	-3.2								
300 ns		Experimental	290	-21.1	-4.6	Experimental	270	-15.7	-5.5	Experimental	251	-25.4	-5.4	BS 3	238	-16.6	-5.4
		BS 1	290	-22.7	-5.6												
		BS 3	220	-16.8	-5.6												
200 ns	2jjc	Experimental	90	-8.9	-4.8					Experimental	190	-11.0	-4.8	Experimental	70	-7.6	-4.9
		BS 1	170	-18.2	-4.0												
200 ns	2ohk	BS 1	39	-9.9	-4.4	S1	105	-11.6	-5	BS 4	29	-10.5	-3.2	S1	161	-11.8	-5.2
		BS 2	20	-8.0	-3.6	BS 4	70	-13.8	-5.6	BS 6	36	-8.6	-3.6	BS 4	141	-12.8	-5.5
		BS 3	180	-14.4	-3.6	BS 5	62	-6.7	-5.6					BS 6	47	-8.8	-5.5
300 ns		BS 3	280	-16.4	-4.5	S1	205	-11.0	-5.5	BS 4	129	-12.8	-3.8	S1	261	-10.2	-5.7
						BS 4	170	-12.7	-5.5					BS 4	241	-12.7	-5.6
1 μs						Experimental	28	-9.9	-5.9	BS4	829	-13.2	-4.2				

[a] In ns.[b] MMPBSA binding energy.[c] K_{DEEP} binding energy

Table S8. Results calculated for reactive trajectories of set III. MMGBSA and K_{DEEP} energies in kcal mol^{-1}. Pockets are defined in Table S6.

Simulation time	System	MD run number															
		1				2				3				4			
		Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]
200 ns	class1	Experimental	170	-22.3	-5.5	Experimental	195	-21.0	-6.0	Experimental	195	-19.4	-6.2	Experimental	190	-21.8	-5.6
		BS 1	25	-9.7	-4.5	BS 3	40	-13.5	-4.17	BS 2	50	-18.7	-4.7	BS 4	140	-17.4	-3.8
		BS 2	60	-12.4	-5.2	BS 4	60	-18.2	-3.56	BS 4	160	-13.6	-3.4				
		BS 3	50	-14.1	-4.2	BS 5	60	-17.6	-3.34	BS 6	100	-24.0	-4.9				
		BS 4	50	-16.2	-3.8					BS 7	50	-13.4	-5.5				
		BS 5	25	-13	-4.1												
300 ns		Experimental	270	-22.3	-5.3	Experimental	295	-21.7	-5.3	Experimental	295	-23.9	-5.5	Experimental	290	-24.3	-6.4
		BS 2	160	-12.4	-5.0	BS 4	160	-16.1	-3.4	BS 4	260	-18.2	-3.8	BS 4	240	-15.3	-4.0
		BS 4	150	-21.3	-3.2					BS 6	200	-28.1	-3.8				
200 ns	class2	Experimental	140	-28.8	-6.3	Experimental	132	-28.0	-6.0	Experimental	175	-21.1	-6.2	BS 10	110	-15.8	-4.6
		BS 4	20	-22.7	-3.5	BS 4	33	-24.9	-3.8	BS 5	135	-12.9	-3.2	BS 13	20	-16.6	-4.6
		BS 5	50	-19.9	-3.2	BS 15	30	-12.4	-3.8	BS 6	30	-14.6	-3.7	BS 14	122	-15.3	-3.7
		BS 6	192	-31.7	-4.8					BS 10	78	-20.2	-4.7				
		BS 8	60	-14.6	-4.3					BS 11	70	-20.6	-2.9				
		BS 9	25	-14.6	-4.1					BS 12	110	-28.3	-5.0				
300 ns		Experimental	240	-29.8	-6	Experimental	232	-22.5	-4.7	Experimental	275	-21.6	-5	BS 14	222	-16.9	-3.6
		BS 6	292	-29	-3.9	BS 4	133	-24	-3.9	BS 10	178	-16.6	-4.1				
										BS 11	170	-20.2	-3.2				
1 μs		Experimental	940	-23.2	-6.5	Experimental	932	-22	-6.2	Experimental	975	-18.4	-5.8				

[a] In ns. [b] MMPBSA binding energy. [c] K_{DEEP} binding energy

Table S9. Results calculated for reactive trajectories of set IV. MMGBSA and K_{DEEP} energies in kcal mol^{-1}.

Apo Receptor System	MD run number															
	1				2				3				4			
	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]	Binding site	RT[a]	MMGBSA[b]	K_{DEEP}[c]
dmso																
dss					Experimental	155	-14.2	-5.9	Experimental	149	-14.4	-6.0				
aca	Experimental	17	-11.2	-6.2	Experimental	3	-2.2	-5.7	Experimental	13	-10.0	-5.8	Experimental	23	-13.2	-5.9

[a] In ns. [b] MMPBSA binding energy. [c] K_{DEEP} binding energy

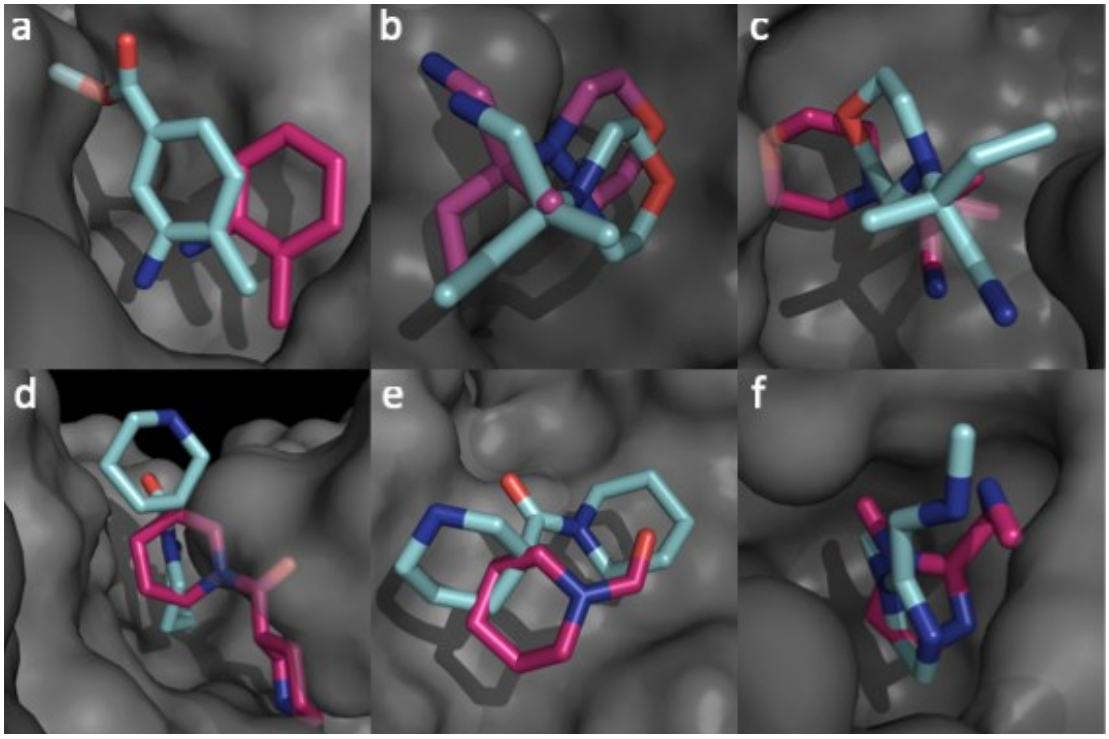


Figure S1. Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the different systems studied in set I with the 5DR1 receptor. a) f031_5DR1, pocket S1´, b) f035_5DR1, pocket S3, c) f035_5DR1, pocket S6, d) f207_5DR1, catalytic dyad, e) f207_5DR1, pocket S6, f) f240_5DR1, pocket S1. Hydrogen atoms are omitted for clarity.

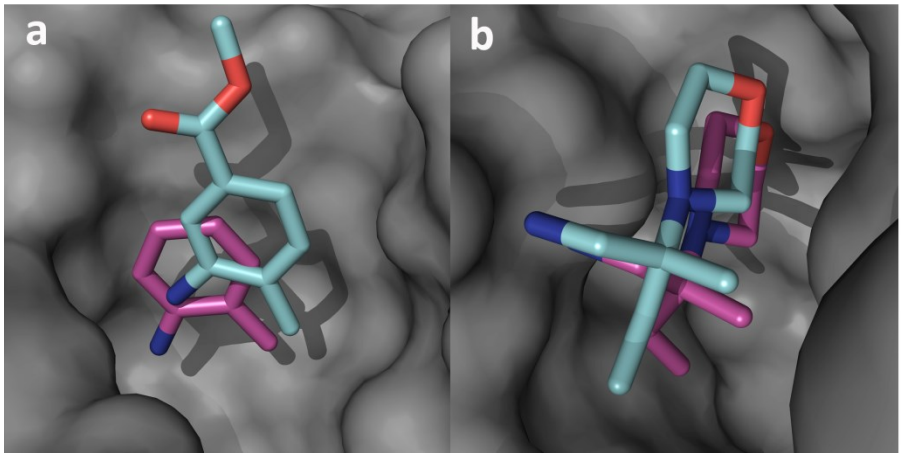


Figure S2. Comparison of experimental (carbon atoms magenta) and predicted (carbon atoms cyan) binding poses for the f031 and f035 systems studied in set I with the 5DR1 receptor after 300ns of MD. a) f031_5DR1, pocket S1´, b) f035_5DR1, pocket S3. Hydrogen atoms are omitted for clarity

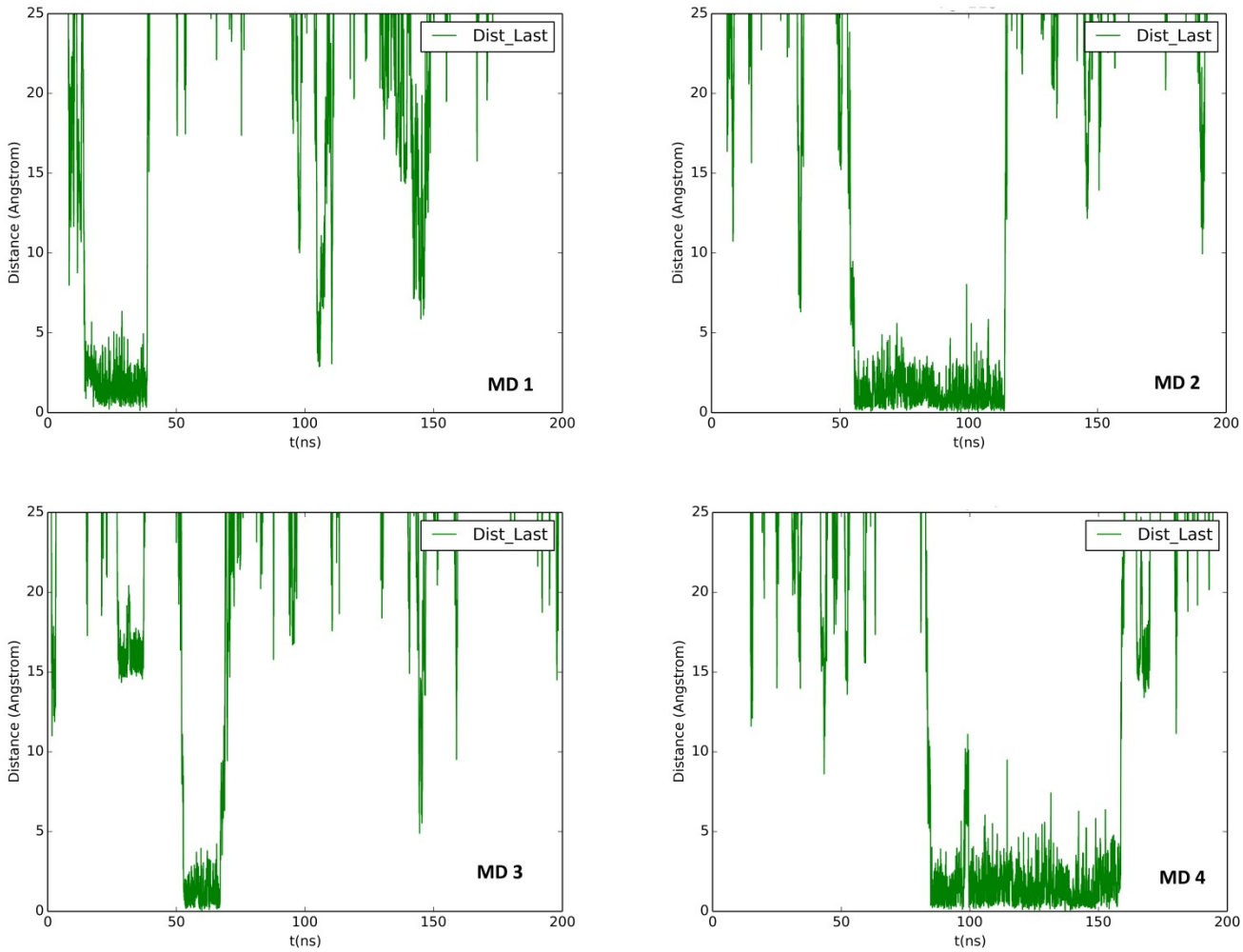


Figure S3. Plots of the distance between the S99 atom of the DMSO ligand to the same atom of the X-ray structure along the four molecular dynamics.