Electronic Supplementary Material (ESI) for Molecular BioSystems.

Supplementary Material

Target specific proteochemometric model development for BACE1– protein flexibility and structural water are critical in virtual screening

Prabu Manoharan,[a] Kiranmai Chennoju,[b] Nanda Ghoshal*[a]

[a]Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata – 700032, India.

[b]National Institute of Pharmaceutical Education and Research, Kolkata -700032, India.

Statistical Validation

The statistical parameters, non-cross-validated correlation coefficient (R^2), LOO cross-validated correlation coefficient (R^2_{cv}) and Predicted REsidual Sums of Squares (PRESS) were calculated accordingly to the following equations.

$$R^2 = \frac{\sum (Y_{pred} - Y)^2}{\sum (Y - \bar{Y})^2} \qquad (1).$$

$$R^2_{cv} = 1 - \frac{PRESS}{\sum (Y - \bar{Y})^2} \qquad (2).$$

$$PRESS = \sum (Y - Y_{pred})^2 \qquad (3).$$

Where, Y is the Experimental value, $\bar{Y}$ is the mean of the experimental values.

To validate the derived QSAR models, the overall predictive ability of the analysis was evaluated by the term R^2_{pred}, which was calculated according to eq.4.

$$R^2_{pred} = \frac{1 - PRESS}{SD} \qquad (4).$$

Where, SD is the sum of square deviations between the biological activities of each molecule in the test set, and the mean activity of the training set molecules, and PRESS is the sum of square deviations between the predicted and the actual activities of molecules in the test set. All the obtained models have acceptable levels of predictive quality. More rigorous validation was carried out using the other parameters as proposed by Tropsha *et al*, stated in eq.5 and 6.

$$(R^2 - R^2{}_o)/R^2 < 0.1 \text{ or } (R^2 - R'^2{}_o)/R^2 < 0.1 \qquad (5).$$

$$k \text{ or } k' \text{ close to } 1 \qquad (6).$$

Where, R^2 represents the square correlation coefficient between the predicted and observed activities. $R^2{}_O$ and $R^{2\prime}{}_O$ are quantities characterizing the square correlation between predicted vs observed and observed vs predicted activity, respectively, with Y-intercept set to zero and k, k′ are their corresponding slopes. The models developed in this study satisfy above-mentioned condition.

Model

$$pIC_{50} = 3.46786 + -0.023945607 * HB + 0.0525335 * HYB \qquad (7)$$

N = 81, R = 0.79895, R^2 = 0.638321, R^2_{CV} = 0.608773, F = 68.8304, R^2_{pred} = 0.62, R^2_{pred} = 0.646 (External test set), R^2_{pred} = 0.687 (Additional validation set).

Table S1. BACE1 PDB ID with Experimental pIC_{50}, Predicted pIC_{50} and descriptors used in this study.

S. No	PDB ID	EXP pIC_{50}	PRED pIC_{50}	¥HB count	€HYD count
1	3ixj	9.49485	9.75916	16	128
2	2vkm	9	9.08452	14	114
3	3k5g	8.69897	7.58156	7	82
4	2xfk	8.52288	8.22402	11	96
5	2qmf	8.52288	7.49889	9	81
6	2vnm	8.52288	7.78351	7	86
7	2iqg	8.30103	7.29695	9	77
8	2qmd	8.30103	7.19597	9	75
9	3lpi	8.30103	6.75989	11	67
10	3ixk	8.18046	9.08452	14	114
11	3lpk	8.1549	8.02207	11	92
12	2qzl	8.09691	8.74498	10	106
13	3l58	8.09691	7.65257	12	85
14	2qk5	8.09691	6.90914	8	69
15	1ym4	8	7.56547	8	82
16	2p83	7.95861	7.29695	9	77
17	2oah	7.95861	7.56991	14	84
18	3ivi	7.92082	6.59013	9	63
19	2vj6	7.88606	8.03816	10	92

20	2hm1	7.88606	7.50111	12	82
21	3ooz	7.85387	5.94767	5	49
22	3cib	7.85387	6.89305	9	69
23	2xfi	7.82391	7.73968	16	88
24	3s7m	7.82391	6.21841	7	55
25	4dh6	7.77469	6.95741	5	69
26	2wf3	7.72125	7.1477	12	75
27	2wf2	7.58503	7.6043	15	85
28	2ph6	7.56864	7.41845	14	81
29	2qzk	7.56864	7.1477	12	75
30	3dv5	7.56864	6.90914	8	69
31	2vie	7.48149	7.54716	6	81
32	2vij	7.46852	7.73302	7	85
33	3inf	7.39794	6.20232	8	55
34	2vj7	7.39794	7.73302	7	85
35	3ivh	7.3279	6.31938	7	57
36	3k5d	7.27572	6.23672	9	56
37	4acx	7.25	6.14961	5	53
38	3in3	7.22185	5.52768	6	41
39	3s7l	7.09691	5.81673	10	48
40	3l38	7	6.48915	9	61
41	2hiz	6.95078	7.3957	6	78
42	3ufl	6.82391	7.69641	3	83
43	2f3e	6.80688	7.12939	10	74
44	3ine	6.76955	5.4084	4	38
45	2f3f	6.72125	6.64061	9	64
46	3kyr	6.72125	5.93823	15	52
47	2is0	6.69897	8.14357	16	96
48	3u6a	6.48946	6.52133	7	61
49	3l3a	6.37675	7.05838	5	71
50	3nsh	6.3279	5.8994	8	49
51	2va6	6.22915	5.72741	3	44
52	3dv1	6.22915	6.62452	10	64
53	2qu3	6.22915	5.94767	5	49
54	3l5f	6.21824	5.0733	6	32
55	2va7	6.19382	5.98206	6	50
56	3pi5	6.02228	6.25059	5	55
57	3lnk	5.9393	6.28942	12	58
58	1tqf	5.85387	7.12218	23	78
59	3ind	5.81531	5.27303	3	35
60	2ph8	5.74473	6.20676	14	57
61	3h0b	5.49485	5.84669	5	47
62	2qu2	5.4318	5.39231	5	38
63	3ckr	5.30094	7.04451	9	72
64	3udy	5.20343	6.35156	5	57
65	3s2o	5.1549	4.39866	4	18
66	3l5c	5.1549	5.27524	6	36
67	2oht	5.04096	5.0733	6	32
68	3udn	4.91364	6.13131	3	52

69	3exo	4.61979	5.67914	6	44
70	2ohq	4.60206	5.81229	4	46
71	3udr	4.59517	6.33769	9	58
72	2ohs	4.39794	5.72963	6	45
73	3l5d	4.10791	5.16262	13	36
74	2ohp	4.02687	4.60282	7	23
75	2ohr	4	5.30964	7	37
76	3udk	3.9431	5.18814	2	33
77	2ohm	3.50864	4.55012	4	21
78	2ohn	3.30103	4.68327	2	23
79	3msk	3.11351	4.55234	7	22
80	3hw1	3.06048	4.61669	3	22
81	2ohk	2.69897	4.34817	4	17
82*	2wf1	8.69897	7.70084	9	85
83*	3cic	8.52288	7.28086	10	77
84*	3cid	8.30103	7.17988	10	75
85*	2qp8	8.09691	7.29695	9	77
86*	1ym2	8	7.10165	18	76
87*	2irz	7.92082	7.6043	15	85
88*	2xfj	7.88606	7.71693	8	85
89*	2b8l	7.82391	8.74498	10	106
90*	3k5c	7.76955	7.21206	8	75
91*	3in4	7.39794	5.63422	6	44
92*	4acu	7.39	6.97572	7	70
93*	3n4l	7.22915	6.75546	5	65
94*	2b8v	7.00877	8.00598	12	92
95*	3qbh	6.82391	7.53329	10	82
96*	2vj9	6.74473	7.86839	8	88
97*	2wf0	6.67778	7.66644	8	84
98*	1w51	6.30103	7.22815	7	75
99*	2ze1	6.22185	6.604	5	62
100*	2zdz	6.1549	6.70497	5	64
101*	3duy	5.85387	6.57404	10	63
102*	3fkt	5.55284	6.59234	12	64
103*	2ohu	5.37675	5.89718	5	48
104*	3udp	4.64589	6.47084	7	60
105*	3igb	4.42022	5.07108	3	31
106*	2va5	4.0655	4.38257	5	18
107*	3kmx	3.67778	5.29577	11	38
108*	2ohl	2.69897	4.16231	3	13

*test set, ¥HB= Hydrogen bonds, €HYD = Hydrophobic interactions

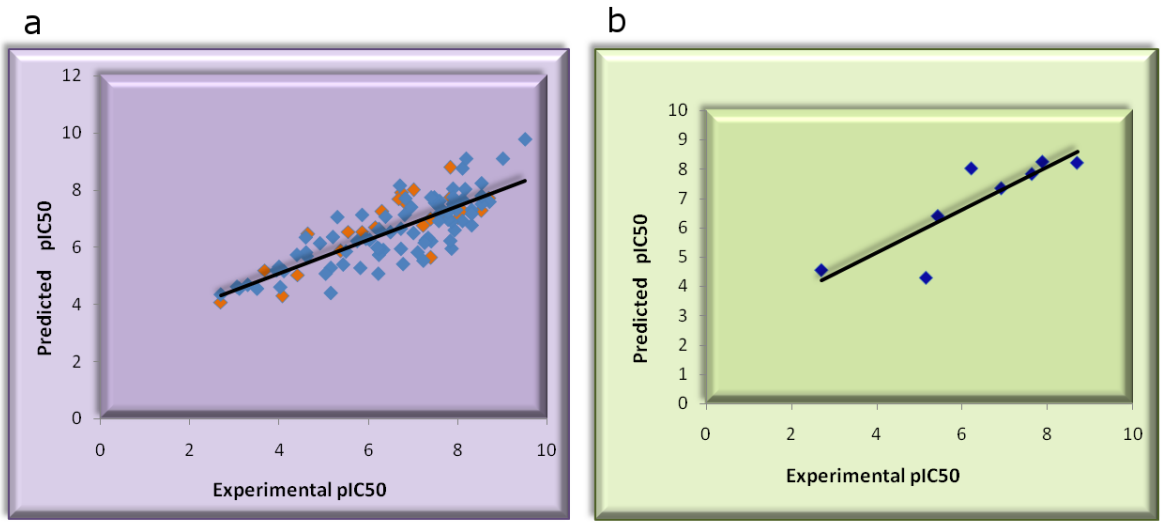

Figure S1. Scatter plot showing the correlation between experimental and predicted pIC_{50} of both training set (blue dimond) and test set (orange dimond) (a) and external test set (b)

Table S2. External test set (BACE1 PDB ID) with experimental pIC_{50}, predicted pIC_{50} and descriptors used in this study

SI. No	PDB ID	Exp pIC_{50}	Pred pIC_{50}	HB count	HYD count
1	2vnn	8.69897	8.162	12	96
2	2wf4	7.886057	8.214	15	98
3	3lpj	6.920819	7.328	11	79
4	2viz	6.218245	7.96	9	91
5	3k5f	5.431798	6.392	7	59
6	3hvg	2.69897	4.624	5	23
7	2wez	7.638272	7.758	6	86
8	3msl	5.154902	4.374	5	18

Table S3. Additional validation set with experimental pIC_{50}, predicted pIC_{50} and descriptors used in this study.

SI. No	BindingDB ID	Exp pIC_{50}	Pred pIC_{50}	HB count	HYD count
1	1103	8.30103	7.94715	6	88
2	1436	5.154902	5.81721	5	47
3	2324	7.136677	6.86324	3	66

4	271	7.136677	6.94437	4	68
5	494	6.756962	6.91578	3	67
6	506	6.744727	7.56011	9	82
7	5d	7.161151	7.30746	2	74
8	846	6.221849	6.28538	3	55
9	859	6.19382	6.37114	6	58
10	887	6.142668	6.65775	5	63
11	910	6.09691	6.54804	3	60

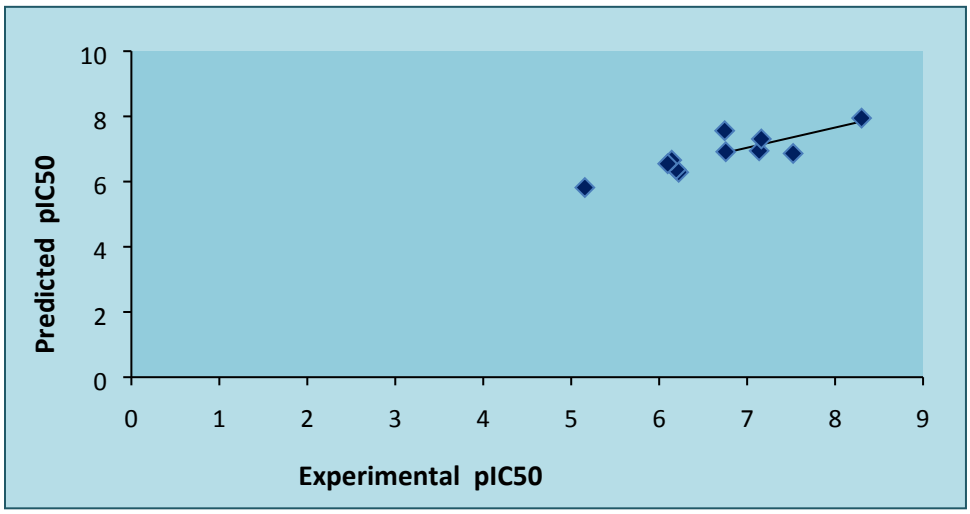

Figure S2. Scatter plot showing the relation between experimental and predicted pIC_{50} of additional validation set.

Table S4 BACE1 PDB ID with experimental and predicted pIC_{50} (nm) (model 2) obtained from SPLC model.

S. No	PDB ID	Exp pIC_{50}(nm)	Pred pIC_{50}(nm)
1	2F3E	6.81	7.25
2	2F3F	6.72	6.73
3	2HIZ	6.95	7.10

4	2HM1	7.89	7.54
5	2IQG	8.30	7.27
6	2OHP	4.03	4.84
7	2OHT	5.04	5.43
8	2P83	7.96	7.29
9	2PH6	7.57	7.60
10	2PH8	5.74	6.55
11	2QMD	8.30	7.65
12	2QMF	8.52	7.63
13	2QU2	5.43	5.72
14	2QU3	6.23	6.08
15	2QZK	7.57	7.14
16	2QZL	8.10	8.69
17	2VA6	6.23	5.82
18	2VIE	7.48	7.41
19	2VIJ	7.47	7.63
20	2VJ6	7.89	8.10
21	2VJ7	7.40	8.16
22	2VKM	9.00	9.35
23	2VNM	8.52	7.66
24	2WF2	7.59	7.90
25	2WF3	7.72	7.37
26	2XFI	7.82	8.25
27	2XFK	8.52	8.36
28	3CIB	7.85	7.16
29	3DV1	6.23	6.66
30	3DV5	7.57	6.77
31	3EXO	4.62	5.55
32	3H0B	5.49	5.97
33	3IND	5.82	5.39
34	3INF	7.40	6.38
35	3IVH	7.33	6.78
36	3IXJ	9.49	10.05
37	3KYR	6.72	6.79
38	3L38	7.00	6.82
39	3L3A	6.38	7.09
40	3L58	8.10	8.02
41	3L5C	5.15	5.40
42	3L5D	4.11	5.26
43	3LNK	5.94	6.53
44	3LPK	8.15	8.46

45	3NSH	6.33	5.93
46	3PI5	6.02	5.92
47	3S2O	5.15	4.30
48	3U6A	6.49	6.96
49	3UDK	3.94	4.91
50	3UDQ	6.00	6.34
51	3UDY	5.20	6.22
52	3UFL	6.82	7.21
53	4AZY	6.70	5.67
54	4D88	7.31	6.99
55	4DH6	7.77	6.86
56	4DI2	8.26	7.37
57	4DPI	7.11	7.86
58	4EWO	7.36	6.46
59*	1W51	6.30	7.17
60*	1YM2	8.00	8.03
61*	2B8L	7.82	8.73
62*	2B8V	7.01	8.15
63*	2IRZ	7.92	7.82
64*	2OHU	5.38	6.18
65*	2VA5	4.07	4.68
66*	2VJ9	6.74	7.87
67*	2WF0	6.68	7.57
68*	2WF1	8.70	7.69
69*	2XFJ	7.89	7.70
70*	2ZDZ	6.15	6.86
71*	2ZE1	6.22	7.06
72*	3CIC	8.52	7.62
73*	3CID	8.30	7.53
74*	3DUY	5.85	6.51
75*	3IGB	4.42	5.28
76*	3K5C	7.77	7.13
77*	3N4L	7.23	6.80
78*	3QBH	6.82	7.22
79*	4ACU	7.39	7.10

*****_Indicates testset_

Table S5 BACE1 protein-ligand crystal complexes considered as external test set, is provided with PDB ID, experimental and predicted pIC_{50} obtained from SPLC model .

S.No	PDB ID	Exp pIC_{50} (nm)	Pred pIC_{50} (nm)
1	2VNN	8.70	8.36
2	2WEZ	7.64	7.56
3	2WF4	7.89	8.58
4	3K5F	5.43	6.41
5	3LPJ	6.92	7.86
6	3MSL	5.15	4.97
7	3VEU	7.26	7.05
8	3VF3	5.38	5.65
9	3VG1	7.26	6.24
10	3VV6	3.80	4.47
11	3VV7	5.34	5.71
12	3VV8	4.77	4.96
13	4B1C	6.72	6.26
14	4D85	6.85	6.27
15	4D89	6.64	6.09
16	4FM8	5.62	5.78

Table S6. Descriptors that explain BACE1 biological activity.

S.No	Descriptor	Explanation in detail
1	***Glide ligand efficiency***	GlideScore / (number of heavy atoms)
2	***XP LipophilicEvdW***	Lipophilic term derived from hydrophobic grid potential and fraction of the total protein-ligand vdW energy.
3	***XP Sitemap***	SiteMap ligand-receptor non-Hydrogen bonding polar-hydrophobic terms.
4	***XP HBond***	ChemScore H-bond pair term.
5	***XP Penalties***	Polar atom burial and desolvation penalties, and penalty for intra-ligand contacts.
6	***XP PhobEn***	Hydrophobic enclosure reward. Hydrophobic enclosure: Locate regions (“enclosed” by protein lipophilic groups) where water molecules have unfavorable chemical potentials due to inability to make full complement of H bonds and/or substantial loss of entropy; then, occupation of these regions with suitable ligand groups provides exceptional gains in potency with a relatively small number of atoms.
7	***XP PhobicPenal***	Penalty for solvent-exposed ligand groups; cancels van der Waals terms.
8	***XP Eint***	The sum of the 'vdw' and 'coul' terms. This is a convenient measure of the non-bonded interaction energy.
9	**glide_lipo**	
10	**glide_hbond**	
11	**glide_erotb**	

e-Energetic predictive model results

Table S7. BACE1 protein-ligand crystal complexes used in this study for the purpose of deriving predictive model is provided with PDB ID, experimental and predicted ΔG obtained from e-Energetic model 4 (glide XP) and model5 (glide SP)

S. No	PDB ID	Exp ΔG (Kcal/mol) Glide XP	Pred ΔG (Kcal/mol) Glide XP	S. No	PDB ID	Exp ΔG (Kcal/mol) Glide SP	Pred ΔG (Kcal/mol) Glide SP
1	2F3E	-9.28	-9.44	1	2F3E	-9.28	-10.76
2	2F3F	-9.16	-10.20	2	2F3F	-9.16	-9.31
3	2HIZ	-9.47	-9.16	3	2HIZ	-9.47	-9.69
4	2HM1	-10.75	-11.62	4	2HM1	-10.75	-10.56
5	2IQG	-11.31	-10.55	5	2IQG	-11.31	-10.65
6	2OAH	-10.84	-9.78	6	2IS0	-9.13	-10.29
7	2OHK	-3.68	-4.85	7	2OHK	-3.68	-4.26
8	2OHM	-4.78	-5.96	8	2OHM	-4.78	-4.58
9	2OHN	-4.50	-5.11	9	2OHN	-4.50	-3.73
10	2OHP	-5.49	-5.52	10	2OHP	-5.49	-5.17
11	2OHQ	-6.27	-6.29	11	2OHQ	-6.27	-6.85
12	2OHR	-5.45	-6.92	12	2OHR	-5.45	-6.99
13	2OHS	-5.99	-6.86	13	2OHS	-5.99	-6.90
14	2OHT	-6.87	-7.06	14	2OHT	-6.87	-7.76
15	2P83	-10.84	-11.06	15	2P83	-10.84	-9.96
16	2PH6	-10.31	-10.84	16	2P8H	-9.74	-11.04
17	2QK5	-11.03	-10.52	17	2PH6	-10.31	-9.71
18	2QMD	-11.31	-11.55	18	2PH8	-7.83	-8.82
19	2QMF	-11.61	-11.57	19	2QK5	-11.03	-10.30
20	2QU2	-7.40	-7.15	20	2QMD	-11.31	-11.32
21	2QU3	-8.49	-8.01	21	2QMF	-11.61	-11.28
22	2QZK	-10.31	-9.94	22	2QU2	-7.40	-7.78
23	2QZL	-11.03	-12.68	23	2QU3	-8.49	-8.85
24	2VA6	-8.49	-7.26	24	2QZK	-10.31	-10.47
25	2VIE	-10.19	-10.73	25	2QZL	-11.03	-12.13
26	2VIJ	-10.18	-10.76	26	2VA6	-8.49	-7.85
27	2VIY	-7.83	-9.60	27	2VA7	-9.67	-9.43
28	2VJ6	-10.75	-9.84	28	2VIE	-10.19	-10.94
29	2VJ7	-10.08	-10.98	29	2VIJ	-10.18	-10.51
30	2VKM	-12.26	-12.41	30	2VIY	-7.83	-8.97
31	2VNN	-11.85	-11.59	31	2VJ6	-10.75	-10.28
32	2WF2	-10.34	-10.29	32	2VJ7	-10.08	-10.70

33	2WF3	-10.52	-10.52	33	2VKM	-12.26	-12.87
34	2XFI	-10.66	-9.32	34	2VNN	-11.85	-10.34
35	2XFK	-11.61	-10.08	35	2WF2	-10.34	-10.32
36	2ZDZ	-8.39	-10.45	36	2WF3	-10.52	-10.22
37	3CIB	-10.70	-11.30	37	2XFI	-10.66	-9.97
38	3CKR	-7.22	-8.84	38	2XFK	-11.61	-10.88
39	3DM6	-10.13	-11.78	39	3CIB	-10.70	-10.80
40	3DV1	-8.49	-9.01	40	3CKR	-7.22	-9.12
41	3DV5	-10.31	-10.04	41	3DV1	-8.49	-8.64
42	3EXO	-6.30	-7.40	42	3DV5	-10.31	-9.93
43	3H0B	-7.49	-6.69	43	3EXO	-6.30	-7.20
44	3IN4	-10.08	-8.13	44	3H0B	-7.49	-7.14
45	3IND	-7.92	-8.05	45	3IN3	-9.84	-8.59
46	3INH	-10.49	-9.28	46	3IN4	-10.08	-8.71
47	3IVH	-9.99	-9.50	47	3IND	-7.92	-8.46
48	3IVI	-10.79	-9.09	48	3INE	-9.22	-9.10
49	3IXJ	-12.94	-12.76	49	3INF	-10.08	-9.33
50	3IXK	-11.15	-10.74	50	3IVH	-9.99	-8.91
51	3K5D	-9.91	-9.57	51	3IVI	-10.79	-9.62
52	3K5G	-11.72	-10.68	52	3IXJ	-12.94	-14.56
53	3KYR	-9.16	-8.64	53	3K5D	-9.91	-9.37
54	3L38	-9.54	-8.74	54	3K5G	-11.72	-10.64
55	3L3A	-8.69	-8.14	55	3KYR	-9.16	-9.20
56	3L58	-11.03	-10.61	56	3L38	-9.54	-10.17
57	3L5C	-7.02	-7.12	57	3L3A	-8.69	-10.81
58	3L5D	-5.60	-7.36	58	3L58	-11.03	-10.67
59	3L5E	-10.31	-9.29	59	3L5C	-7.02	-7.81
60	3L5F	-8.47	-7.58	60	3L5E	-10.31	-9.41
61	3LHG	-10.49	-9.15	61	3L5F	-8.47	-7.34
62	3LNK	-8.09	-8.90	62	3LHG	-10.49	-9.07
63	3LPI	-11.31	-11.30	63	3LNK	-8.09	-8.81
64	3LPK	-11.11	-11.05	64	3LPI	-11.31	-11.00
65	3MSJ	-4.24	-5.86	65	3LPK	-11.11	-11.29
66	3MSK	-6.25	-5.51	66	3MSJ	-4.24	-5.54
67	3NSH	-8.62	-9.15	67	3MSK	-6.25	-7.70
68	3OOZ	-10.70	-10.13	68	3NSH	-8.62	-8.13
69	3PI5	-8.21	-8.67	69	3OOZ	-10.70	-9.14
70	3S2O	-7.02	-6.95	70	3PI5	-8.21	-8.37
71	3S7L	-9.67	-8.10	71	3S2O	-7.02	-8.06
72	3UDK	-6.12	-6.23	72	3S7L	-9.67	-7.76
73	3UDN	-7.02	-8.91	73	3U6A	-8.84	-8.77

74	3UDP	-7.12	-7.75	74	3UDH	-4.03	-4.06
75	3UDQ	-8.18	-9.12	75	3UDJ	-4.93	-5.97
76	3UDY	-7.77	-7.98	76	3UDK	-6.12	-6.70
77	3UFL	-9.30	-10.36	77	3UDN	-7.02	-8.20
78	4ACX	-9.69	-9.71	78	3UDP	-7.12	-8.16
79	4AZY	-9.13	-8.98	79	3UDQ	-8.18	-8.07
80	4D88	-9.96	-10.56	80	3UDR	-6.76	-6.60
81	4DH6	-10.59	-10.03	81	3UDY	-7.77	-8.02
82	4DI2	-11.26	-9.84	82	3UFL	-9.30	-10.63
83	4DPF	-8.72	-8.16	83	4ACX	-9.69	-9.40
84	4EWO	-10.02	-8.32	84	4AZY	-9.13	-8.24
85	4EXG	-11.00	-8.90	85	4B00	-11.74	-10.65
86*	1W51	-8.59	-10.36	86	4D88	-9.96	-9.89
87*	2B8L	-10.66	-11.37	87	4DH6	-10.59	-9.97
88*	2B8V	-9.55	-10.02	88	4DI2	-11.26	-10.72
89*	2IRZ	-10.79	-10.95	89	4DUS	-11.31	-9.36
90*	2OHL	-3.68	-3.02	90	4EWO	-10.02	-8.92
91*	2OHU	-7.33	-8.03	91	4EXG	-11.00	-9.06
92*	2QP8	-11.03	-9.74	92*	1W51	-8.59	-9.80
93*	2VA5	-5.54	-7.61	93*	2B8L	-10.66	-10.36
94*	2VJ9	-9.19	-10.27	94*	2B8V	-9.55	-8.89
95*	2WF1	-11.85	-10.94	95*	2IRZ	-10.79	-9.67
96*	2XFJ	-10.75	-10.57	96*	2OHL	-3.68	-5.70
97*	2ZE1	-8.48	-8.10	97*	2OHU	-7.33	-10.01
98*	3CIC	-11.61	-10.39	98*	2QP8	-11.03	-9.76
99*	3CID	-11.31	-11.19	99*	2VA5	-5.54	-6.24
100*	3DUY	-7.98	-9.31	100*	2VJ9	-9.19	-9.71
101*	3IGB	-6.02	-5.78	101*	2WF0	-9.10	-10.38
102*	3K5C	-10.59	-10.42	102*	2WF1	-11.85	-10.38
103*	3KMX	-5.04	-6.64	103*	2XFJ	-10.75	-10.64
104*	3N4L	-9.85	-9.23	104*	2ZE1	-8.48	-9.62
105*	3QBH	-9.30	-11.05	105*	3CIC	-11.61	-11.21
106*	4ACU	-10.08	-8.58	106*	3CID	-11.31	-10.69
-	-	-	-	107	3DUY	-7.98	-8.42
-	-	-	-	108	3FKT	-7.57	-9.01
-	-	-	-	109	3IGB	-6.02	-7.21
-	-	-	-	110	3K5C	-10.59	-10.68
-	-	-	-	111	3KMX	-5.04	-6.19
-	-	-	-	112	3N4L	-9.85	-8.81
-	-	-	-	113	3QBH	-9.30	-9.50
-	-	-	-	114	4ACU	-10.08	-10.21

**Indicates testset*

Table. S8. External test set used in this study with experimental and predicted ΔG (obtained from model 4 (glide xp) and model5 (glide sp)

S.No	PDB ID	Exp ΔG (Kcal/mol) glide xp	Pred ΔG (Kcal/mol) glide xp	S.No	PDB ID	Exp ΔG (Kcal/mol) glide sp	Pred ΔG (Kcal/mol) glide sp
1	2VIZ	-8.47	-9.92	1	2VNM	-11.61	-10.47
2	2VNM	-11.61	-11.80	2	2WEZ	-10.41	-10.15
3	2WEZ	-10.41	-10.15	3	2WF4	-10.75	-9.50
4	2WF4	-10.75	-10.21	4	3K5F	-7.40	-9.11
5	3HVG	-3.68	-4.38	5	3MSL	-7.02	-7.49
6	3MSL	-7.02	-6.71	6	3VEU	-9.89	-10.48
7	3VEU	-9.89	-10.65	7	3VF3	-7.33	-7.82
8	3VG1	-9.89	-10.24	8	3VG1	-9.89	-9.24
9	3VV6	-5.18	-5.80	9	3VV6	-5.18	-5.18
10	3VV7	-7.27	-7.87	10	3VV7	-7.27	-8.46
11	3VV8	-6.50	-6.65	11	3VV8	-6.50	-7.49
12	4B1C	-9.16	-8.48	12	4B1C	-9.16	-7.87
13	4D89	-9.05	-9.92	13	4D85	-9.34	-10.49
14	4FM7	-9.54	-10.16	14	4D89	-9.05	-9.55
15	4FM8	-7.66	-9.32	15	4FM7	-9.54	-8.90
-	-	-	-	-	4FM8	-7.66	-9.28

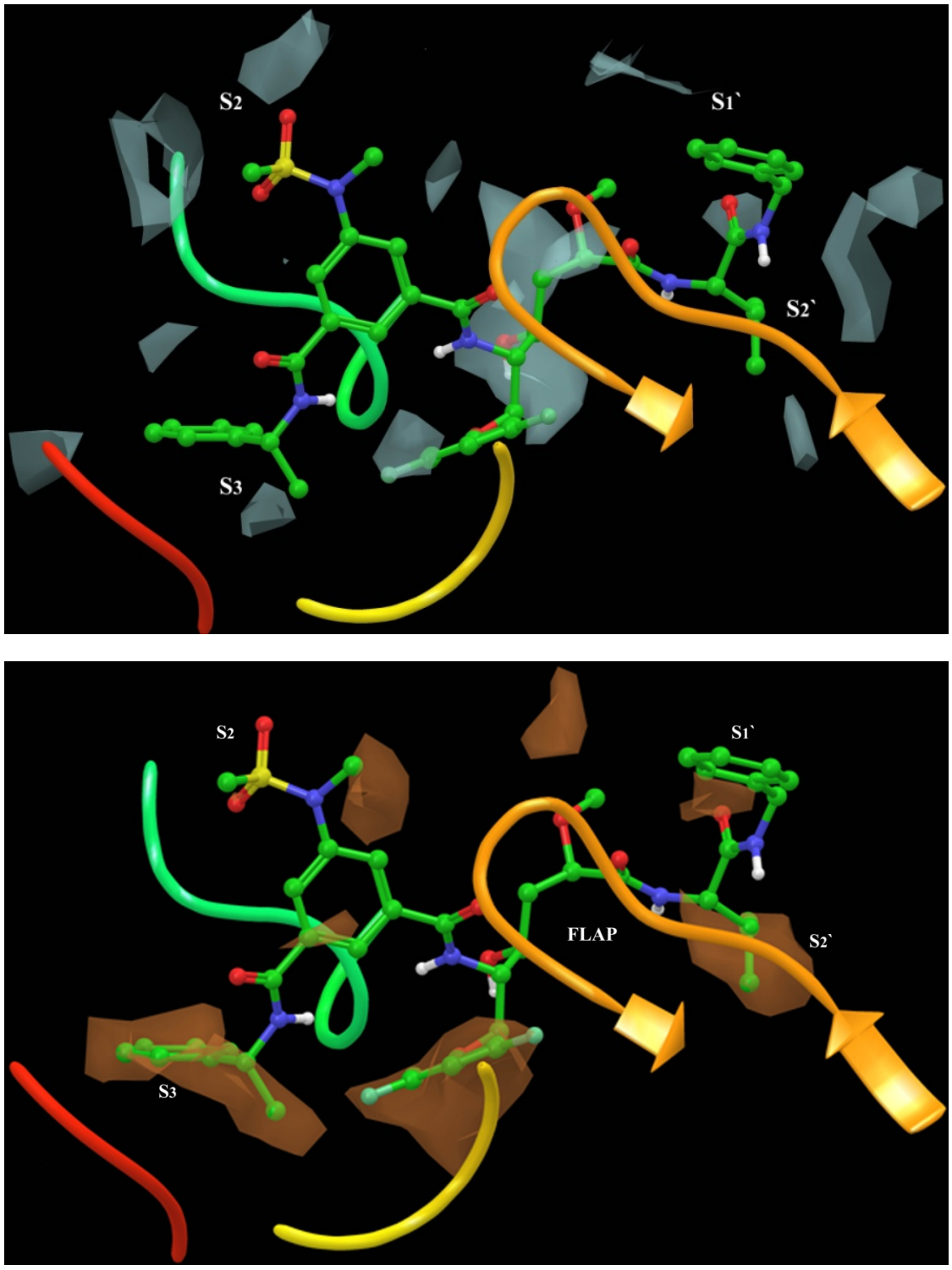

Figure S3. Site map depicting hydrophilic (top) and hydrophobic (bottom) contours calculated from BACE1 active site, bound with hydroxyethylene derivative ligand (PDB ID 3IXJ). The binding site residues are represented as cartoon. The subsites and flap region in the active site are labelled.

Table. S9 The BACE1 crystal structures used in PLIF model building (training and test sets) with experimental and predicted pIC_{50} values

S.No	PDB ID	Exp pIC_{50} (nm)	Pred pIC_{50} (nm)
1	1TQF	5.85	5.47
2	2F3F	6.72	6.97
3	2HM1	7.89	8.63
4	2IQG	8.30	7.43
5	2IS0	6.70	6.88
6	2OHK	2.70	2.83
7	2OHM	3.51	3.11
8	2OHN	3.30	3.23
9	2OHS	4.40	5.12
10	2OHT	5.04	6.22
11	2P8H	7.15	8.04
12	2P83	7.96	7.15
13	2PH6	7.57	6.84
14	2QK5	8.10	9.06
15	2QMD	8.30	7.12
16	2QMF	8.52	8.21
17	2QU2	5.43	5.27
18	2QU3	6.23	6.69
19	2QZK	7.57	7.04
20	2QZL	8.10	7.36
21	2VA6	6.23	5.17
22	2VA7	7.10	7.08
23	2VIE	7.48	8.05
24	2VIJ	7.47	7.55
25	2VIY	5.74	6.62
26	2VJ6	7.89	7.84
27	2VKM	9.00	8.15
28	2WF2	7.59	7.61
29	2WF3	7.72	7.93
30	2XFI	7.82	6.73
31	2XFK	8.52	8.05
32	3CIB	7.85	7.43
33	3DM6	7.43	8.60
34	3DV5	7.57	7.69
35	3EXO	4.62	6.02
36	3IN3	7.22	6.99
37	3IN4	7.40	7.06

38	3INF	7.40	7.64
39	3INH	7.70	8.03
40	3IXJ	9.49	8.89
41	3K5D	7.28	7.27
42	3K5G	8.60	8.46
43	3KN0	3.60	4.42
44	3KYR	6.72	7.83
45	3L3A	6.38	6.75
46	3L38	7.00	6.81
47	3L5C	5.15	5.93
48	3L5D	4.11	4.77
49	3L5E	7.57	7.27
50	3L5F	6.22	6.68
51	3L58	8.10	7.63
52	3LHG	7.70	6.53
53	3LNK	5.94	6.52
54	3LPI	8.30	9.00
55	3LPK	8.15	7.54
56	3NSH	6.33	6.57
57	3OOZ	7.85	7.16
58	3PI5	6.02	6.48
59	3S2O	5.15	5.54
60	3S7L	7.10	7.44
61	3U6A	6.49	5.16
62	3UDH	2.96	3.79
63	3UDJ	3.62	3.63
64	3UDK	4.49	5.75
65	3UDM	4.48	5.66
66	3UDN	5.15	4.65
67	3UDP	5.22	5.56
68	3UDQ	6.00	5.36
69	3UDR	4.96	5.38
70	3UDY	5.70	5.67
71	3UFL	6.82	6.17
72	4ACX	7.11	6.86
73	4B1D	7.62	6.68
74	4D88	7.31	6.04
75	4DH6	7.77	6.54
76	4DPF	6.40	6.64
77*	1W51	6.30	7.27
78*	2B8V	7.01	7.33

79*	2OHU	5.38	5.26
80*	2VJ9	6.74	5.62
81*	2WF0	6.68	6.87
82*	2WF1	8.70	7.68
83*	2XFJ	7.89	6.98
84*	2ZDZ	6.15	6.82
85*	2ZE1	6.22	5.14
86*	3CIC	8.52	7.74
87*	3CID	8.30	7.43
88*	3FKT	5.55	5.31
89*	3IGB	4.42	5.64
90*	3K5C	7.77	7.82
91*	3KMX	3.70	2.63
92*	3N4L	7.23	6.05

*test set

Table S**10**. External test set with experimental and predicted pIC_{50}, obtained from PLIF model based prediction.

S.No	**PDB ID**	**Exp pIC_{50}**	**Pred pIC_{50}**
1	2VIZ	6.22	7.36
2	2VNM	8.52	7.41
3	2WEZ	7.64	7.55
4	2WF4	7.89	7.03
5	3HVG	2.70	4.05
6	3MSL	5.15	5.12
7	3VF3	5.38	5.82
8	3VG1	7.26	7.28
9	3VV7	5.34	5.81
10	3VV8	4.77	6.53
11	4D89	6.64	6.54
12	4FM7	7.00	6.60
13	4FM8	5.62	5.20

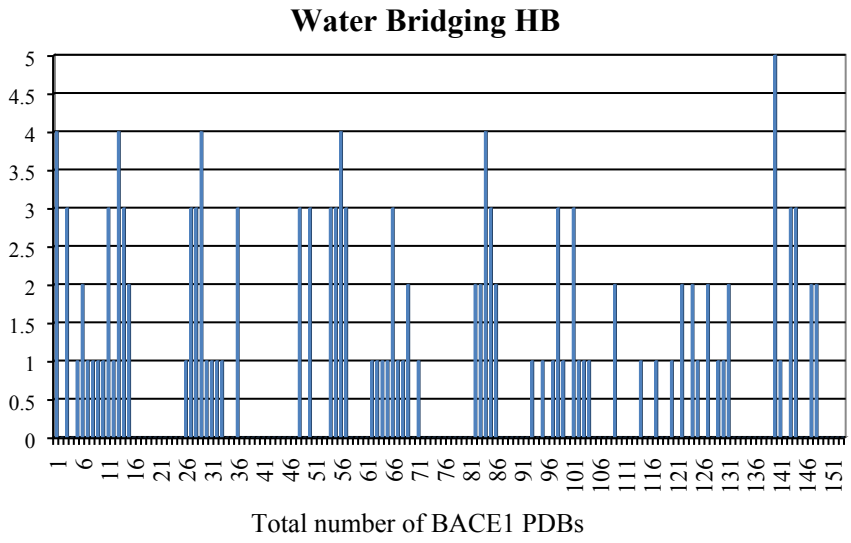

Figure S4. Number of bridging hydrogen bonds observed in BACE1-ligand crystal complexes

Ensemble docking based results

Table S11. BACE1 X-ray crystal structures used in ensemble docking and the RMS deviation observed between each structure.

S. No	PDB ID	1TQF (Å)	2VNM(Å)	3CIC(Å)	2HM1(Å)	3KMX (Å)
1	**1TQF (Å)**	0	0.710	0.530	0.703	0.536
2	**2VKM (Å)**	0.710	0	0.564	0.517	0.552
3	**3CIC (Å)**	0.530	0.564	0	0.497	0.182
4	**2HM1 (Å)**	0.703	0.517	0.497	0	0.562
5	**3KMX (Å)**	0.536	0.552	0.182	0.562	0

Table S12. Ensemble docking based prediction result obtained from SPLC model, BACE1 inhibitors BindingDB monomer identification number, PDB ID, experimental and predicted pIC_{50}.

S.No	BACE1 inhibitors BindingDB monomer ID	BACE1 PDB ID	Exp pIC_{50} (nM)	Pred pIC_{50} (nM)
1	16029	1TQF	6.62	7.40
2	50300230	1TQF	5.15	6.39
3	50254665	1TQF	5.80	7.01

4	50254666	1TQF	5.60	6.77
5	50271632	1TQF	6.29	6.38
6	50300231	1TQF	4.11	5.19
7	50300235	1TQF	5.52	6.59
8	50313895	1TQF	7.70	8.17
9	50196732	2HM1	6.31	6.95
10	50196742	2HM1	7.82	7.50
11	50196748	2HM1	7.66	6.98
12	50226458	2HM1	4.10	4.61
13	50300236	2HM1	6.46	6.57
14	16030	2VKM	6.74	7.78
15	50196734	2VKM	8.40	7.95
16	50196750	2VKM	8.30	7.93
17	50254713	2VKM	6.28	7.55
18	50196757	2VKM	8.40	8.00
19	50196740	3CIC	8.40	7.50
20	50196753	3CIC	8.30	8.09
21	50226440	3CIC	5.46	5.68
22	50226451	3CIC	3.32	4.38
23	50271626	3CIC	8.52	7.83
24	50271768	3CIC	6.28	6.56
25	50271810	3CIC	6.21	6.77
26	50271861	3CIC	8.00	7.24
27	50300232	3KMX	5.30	5.82

Table: S13 Ensemble docking based prediction results obtained from Glide XP and Glide SP models, BACE1 inhibitors BindingDB monomer identification number, PDB ID, experimental and predicted ΔG.

S.No	BACE1 inhibitors-BindingDB monomer ID	BACE1 PDB ID	Exp ΔG (Kcal/mol) Glide XP	Pred ΔG (Kcal/mol) Glide XP	S.No	BACE1 inhibitors-BindingDB monomer ID	BACE1 PDB ID	Exp ΔG (Kcal/mol) Glide SP	Pred ΔG (Kcal/mol) Glide SP
1	16029	1TQF	-9.02	-8.83	1	1TQF	50196742	-10.66	-11.13
2	50226458	1TQF	-5.59	-6.35	2	1TQF	50231941	-8.12	-8.62
3	50254665	1TQF	-7.90	-8.53	3	1TQF	50271632	-8.57	-8.70
4	50254666	1TQF	-7.63	-9.57	4	2HM1	16690	-7.13	-7.55
5	50254747	1TQF	-8.18	-9.73	5	2HM1	50196732	-8.59	-8.71
6	50271632	1TQF	-8.57	-8.87	6	2HM1	5022644	-7.43	-6.37

7	50271810	1TQF	-8.46	-8.75	7	2HM1	50226451	-4.52	-4.78
8	50300235	1TQF	-7.53	-7.37	8	2HM1	50254665	-7.90	-7.76
9	50196739	2HM1	-11.31	-10.96	9	2HM1	50254667	-8.25	-7.73
10	50196740	2HM1	-11.44	-10.83	10	2HM1	50254709	-8.11	-8.55
11	50196742	2HM1	-10.66	-11.33	11	2HM1	50300230	-7.02	-7.89
12	50196747	2HM1	-11.85	-10.41	12	2HM1	50300231	-5.60	-6.42
13	50196748	2HM1	-10.43	-10.29	13	2HM1	50300232	-7.22	-7.14
14	16030	2VKM	-9.19	-10.05	14	2HM1	50300235	-7.53	-7.94
15	50196734	2VKM	-11.44	-11.71	15	2HM1	50300236	-8.80	-8.85
16	50196750	2VKM	-11.31	-11.24	16	2VKM	16020	-8.19	-8.09
17	50196754	2VKM	-11.44	-10.28	17	2VKM	50196740	-11.44	-10.14
18	50196757	2VKM	-11.44	-11.04	18	2VKM	50196748	-10.43	-10.49
19	50271629	2VKM	-11.85	-10.84	19	2VKM	50196750	-11.31	-10.52
20	50313895	2VKM	-10.49	-10.61	20	2VKM	50196753	-11.31	-11.54
21	50196744	3CIC	-11.44	-10.79	21	2VKM	50196754	-11.44	-10.22
22	50196753	3CIC	-11.31	-9.97	22	2VKM	50212195	-10.11	-10.22
23	50226451	3CIC	-4.52	-5.40	23	2VKM	50254713	-8.56	-9.15
24	50254709	3CIC	-8.11	-9.31	24	2VKM	50271810	-8.46	-8.46
25	50271626	3CIC	-11.61	-11.42	25	3CIC	50196744	-11.44	-10.60
26	50271628	3CIC	-12.26	-11.26	26	3CIC	50254666	-7.63	-7.26
27	50271768	3CIC	-8.55	-8.79	27	3CIC	50271627	-8.12	-9.15
28	50271861	3CIC	-10.90	-9.73	28	3CIC	50271628	-12.26	-11.29
29	50300230	3CIC	-7.02	-8.36	29	3CIC	50271768	-8.55	-7.72
30	50300231	3CIC	-5.60	-6.30	30	3CIC	50271861	-10.90	-9.83
31	50300236	3CIC	-8.80	-8.63	31	3KMX	16029	-9.02	-8.27
32	50300232	3KMX	-7.22	-9.08	32	3KMX	50196734	-8.18	-9.10
-	-	-	-	-	33	3KMX	50226458	-5.59	-5.63

Table S14. Ensemble docking based prediction results obtained from PLIF model, BACE1 inhibitors BindingDB monomer identification number, PDB ID, experimental and predicted ΔG.

S.No	BACE1 inhibitors-BindingDB monomer ID	BACE1 PDB ID	Exp pIC50	Pred pIC50
1	16020	1TQF	6.01	7.32
2	16029	1TQF	6.62	7.20
3	50196750	1TQF	8.30	7.64
4	50226451	1TQF	3.32	4.93

5	50254713	1TQF	6.28	7.72
6	50271628	1TQF	9.00	8.13
7	50271632	1TQF	6.29	6.96
8	50271768	1TQF	6.28	6.95
9	50271810	1TQF	6.21	6.79
10	50313895	1TQF	7.70	7.87
11	50196739	2HM1	8.30	8.65
12	50196740	2HM1	8.40	8.76
13	50196742	2HM1	7.82	8.74
14	50196747	2HM1	8.70	9.20
15	50196748	2HM1	7.66	9.01
16	50196751	2HM1	9.00	9.03
17	50271626	2HM1	8.52	7.79
18	50300236	2HM1	6.46	7.46
19	16030	2VKM	6.74	5.37
20	16412	2VKM	4.82	5.66
21	16690	2VKM	5.23	5.61
22	50226458	2VKM	4.10	3.60
23	50231941	2VKM	5.96	5.61
24	50254665	2VKM	5.80	5.68
25	50254666	2VKM	5.60	5.65
26	50254667	2VKM	6.05	5.69
27	50254713	2VKM	6.28	5.66
28	50254747	2VKM	6.00	5.67
29	50271627	2VKM	5.96	6.08
30	50300225	2VKM	3.70	3.60
31	50313887	2VKM	4.65	6.14
32	50300234	3CIC	4.30	4.53

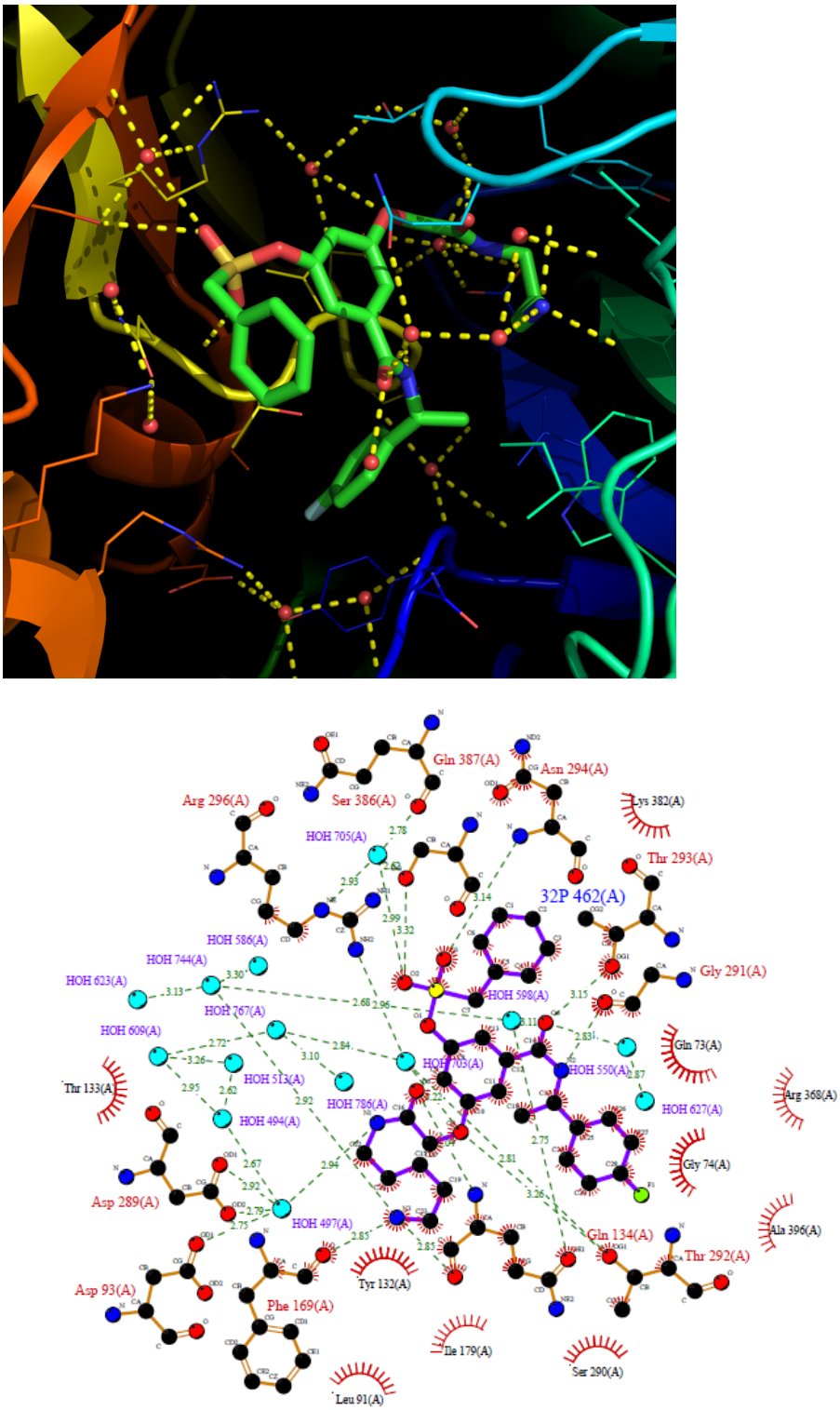

Figure S5. 3D (top) and 2D (bottom) pictures depicting the solvent mediated hydrogen bond network observed in the active sites of BACE1 crystal (1TQF).

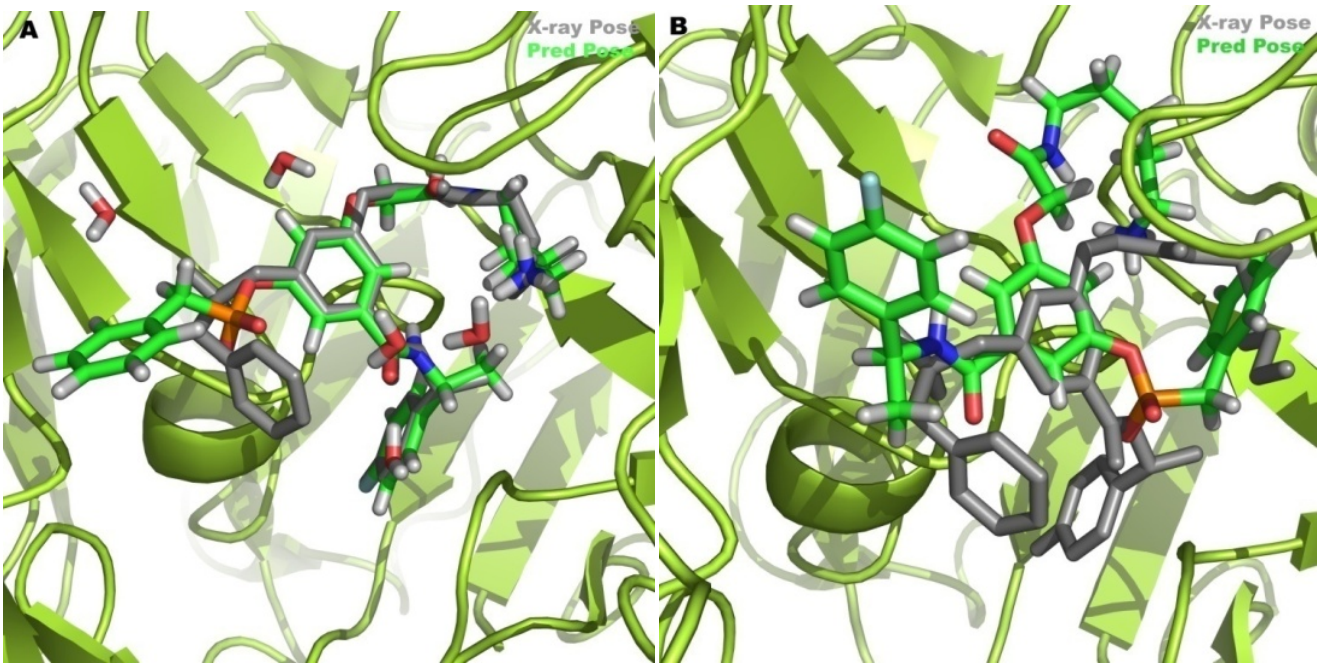

Figure S6. Self-docking results for BACE1 (PDB ID 1TQF) with solvated (A) and desolvated (B) active site. (A). Overlay of docked ligand (atom color) with x-ray crystal structure ligand (gray) in the BACE1 active site with five structural waters (HOH 550, 598, 703, 705, 744), (B) Overlay of docked ligand (atom color) with x-ray crystal structure ligand (gray) in the BACE1 active site without the structural waters.

Table S15. Self-docking, bioactive pose ranking by Glide XP score predicted for BACE1 native ligands.

S.No	Crystal structure	Native ligand bioactive pose ranking (Solvated docking)	Glide XP score (Solvated docking)	RMS Deviation from bioactive pose (Solvated docking)	Native ligand bioactive pose ranking (Desolvated docking)	Glide XP score (Desolvated docking)	RMS Deviation from bioactive pose (Desolvated docking)
1	1TQF	1	-14.40	2.4034	1	-10.89	2.4709
2	2HM1	1	-12.82	1.7562	2	-9.23	1.7172
3	2VKM	1	-13.92	1.8787	1	-14.11	2.5900
4	3CIC	1	-14.52	2.7019	1	-14.70	1.2669
5	3KMX	1	-8.632	0.7621	5	-6.900	0.7343