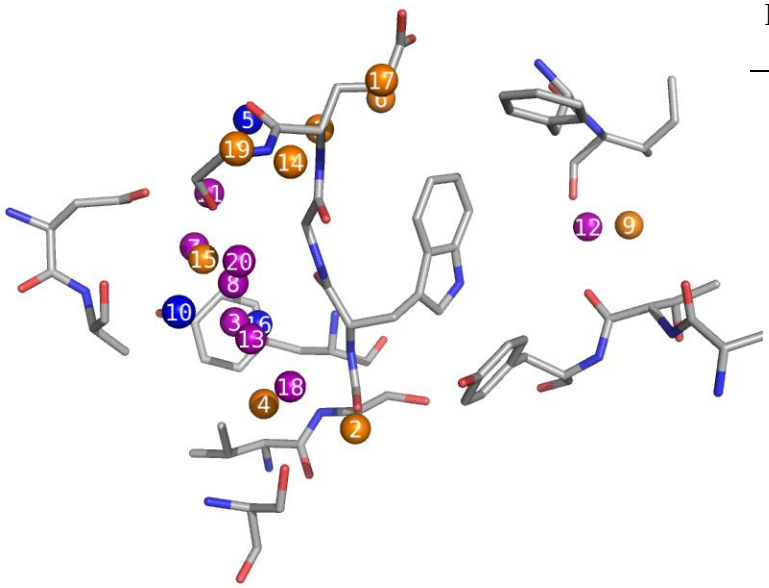
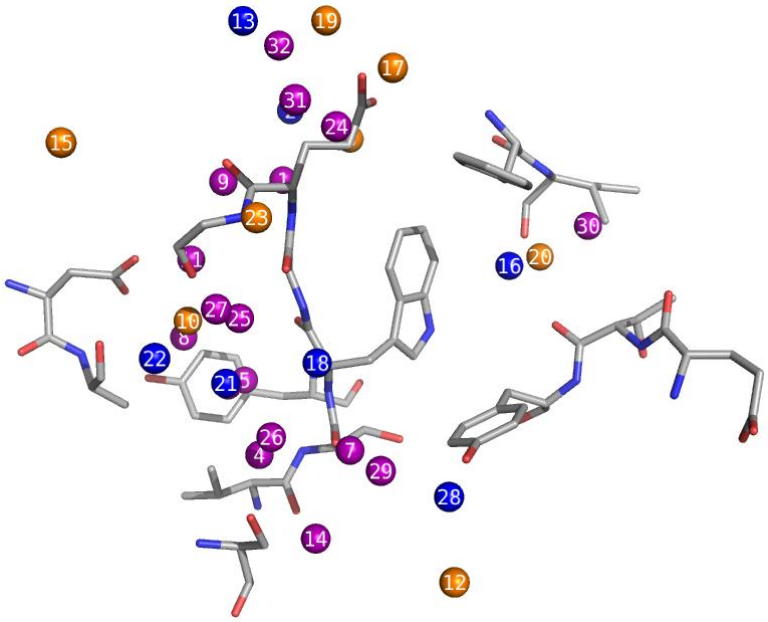
Supporting Information S1. Estimated free energies for the hydration sites of the studied protein systems. The hydration sites in the left figures are numbered according to the first column of the right tables. Hydration sites that contribute both entropically and enthalpically if replaced by a ligand are shown in blue. The hydration sites whose entropic gain surpasses the enthalpic loss when replaced by ligands are shown in purple. The hydration sites whose enthalpic loss is larger than the entropic gain when replaced by ligands are shown in orange. The energy values in the right table are all in kcal/mol. The estimated configurational entropy using both 3×3 and 6×6 matrices were reported. The ΔG_{hs} listed is calculated using the estimated entropy from the 3×3 matrix. Please refer to the "Material and Methods" for the definition of the ΔS_{hs}, ΔE_{hs} and ΔG_{hs}.

fXa: 1F0S



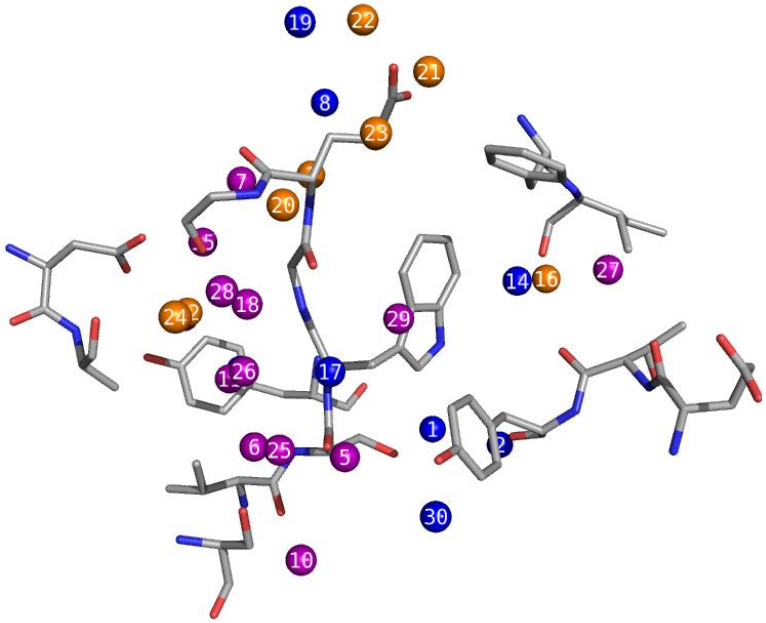
Hydration Site	$-T\Delta S_{hs}$ (3×3)	ΔE_{hs}	ΔG_{hs}	$-T\Delta S_{hs}$ (6×6)
1	2.81	-3.58	-0.77	2.78
2	2.05	-2.34	-0.29	2.04
3	2.88	-2.58	0.30	2.91
4	3.17	-3.26	-0.09	3.19
5	2.02	0.39	2.41	2.06
6	2.54	-3.77	-1.22	2.52
7	2.92	-2.55	0.37	2.93
8	2.14	-0.82	1.32	2.15
9	2.56	-3.46	-0.90	2.57
10	2.22	1.29	3.51	2.24
11	2.56	-0.77	1.79	2.48
12	1.67	-0.85	0.82	1.65
13	1.84	-1.14	0.69	1.81
14	1.82	-2.40	-0.57	1.79
15	3.05	-3.27	-0.22	3.11
16	2.72	1.15	3.87	2.81
17	1.72	-1.78	-0.05	1.75
18	1.80	-1.33	0.48	1.80
19	1.96	-3.54	-1.58	1.97
20	1.27	-0.16	1.11	1.28

fXa: 1MQ6



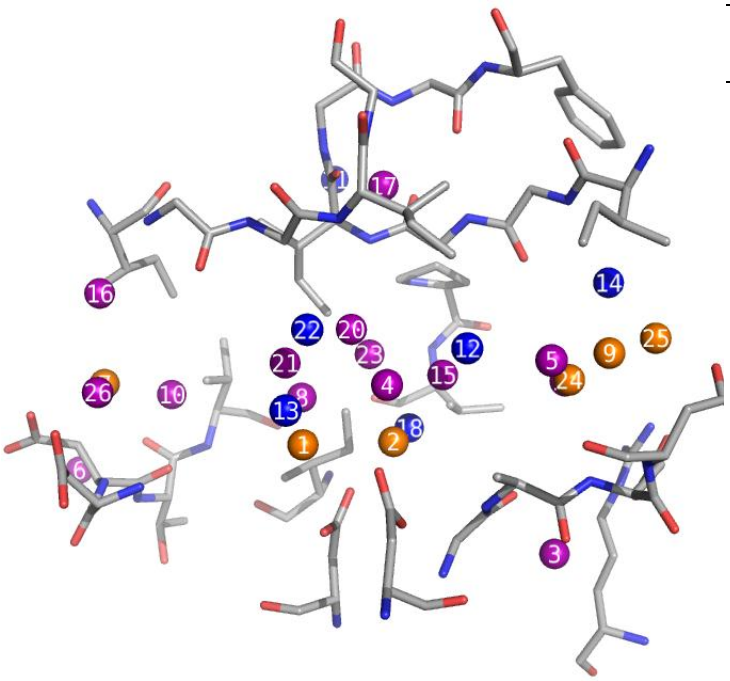
Hydration Site	$-T\Delta S_{hs}$ (3×3)	ΔE_{hs}	ΔG_{hs}	$-T\Delta S_{hs}$ (6×6)
1	4.10	-4.09	0.02	4.13
2	3.56	2.69	6.25	3.57
3	2.31	-3.67	-1.35	2.33
4	3.98	-3.88	0.10	3.96
5	3.33	-0.38	2.96	3.28
6	3.13	-2.44	0.70	3.10
7	2.85	-1.99	0.86	2.84
8	3.05	-0.17	2.88	3.07
9	2.22	-0.39	1.83	2.24
10	2.88	-3.61	-0.72	2.89
11	2.54	-1.37	1.17	2.48
12	1.82	-2.08	-0.27	1.86
13	2.79	0.24	3.03	2.71
14	1.87	-1.83	0.05	1.90
15	3.69	-5.01	-1.31	3.66
16	2.09	1.24	3.33	2.12
17	1.93	-2.54	-0.61	1.90
18	1.58	1.03	2.61	1.53
19	1.71	-3.04	-1.33	1.74
20	2.44	-2.50	-0.06	2.40
21	1.72	0.23	1.95	1.68
22	1.86	1.21	3.07	1.88
23	1.63	-2.49	-0.86	1.66
24	1.92	-1.90	0.02	1.86
25	2.00	-0.96	1.04	1.96
26	1.67	-0.36	1.31	1.64
27	1.24	-0.24	1.00	1.28
28	1.57	0.32	1.89	1.68
29	1.91	-0.33	1.58	1.87
30	1.61	-0.81	0.80	1.63
31	1.49	-0.82	0.66	1.50
32	2.47	-0.25	2.22	2.42

fXa: 1NFU



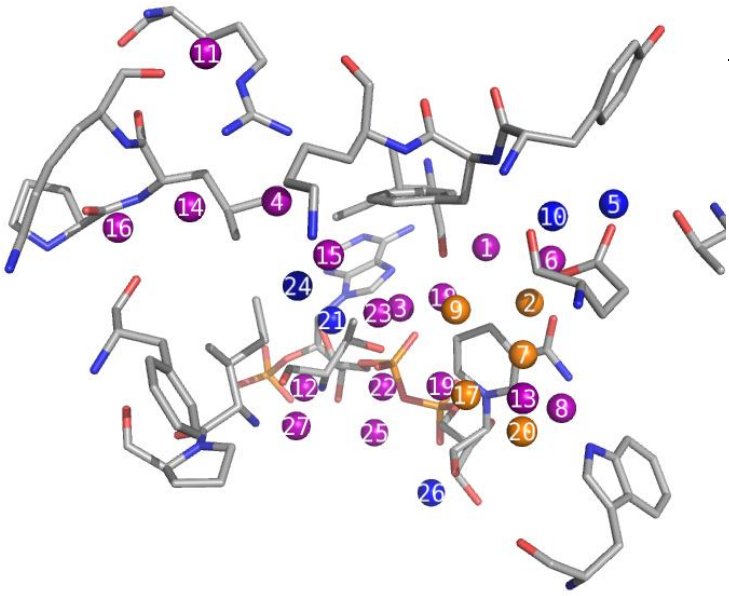
Hydration Site	$-T\Delta S_{hs}$ (3×3)	ΔE_{hs}	ΔG_{hs}	$-T\Delta S_{hs}$ (6×6)
1	2.16	2.95	5.11	2.19
2	3.79	1.41	5.19	3.82
3	2.98	-3.06	-0.08	2.92
4	2.08	-4.12	-2.04	2.11
5	2.65	-2.09	0.56	2.67
6	3.81	-3.25	0.56	3.84
7	1.89	-1.36	0.53	1.88
8	2.34	3.93	6.26	2.35
9	2.82	0.61	3.43	2.75
10	1.84	-1.15	0.68	1.86
11	3.24	-2.67	0.57	3.19
12	2.90	-3.44	-0.55	2.84
13	2.70	-0.38	2.32	2.66
14	2.54	1.14	3.68	2.61
15	2.27	-1.23	1.03	2.26
16	2.85	-3.44	-0.59	2.86
17	1.90	0.77	2.68	1.85
18	1.69	-0.63	1.06	1.68
19	2.52	0.16	2.68	2.51
20	1.74	-2.86	-1.12	1.71
21	1.63	-2.93	-1.29	1.64
22	1.71	-3.03	-1.32	1.67
23	1.78	-1.82	-0.04	1.76
24	2.02	-2.20	-0.18	2.03
25	1.60	-0.27	1.33	1.58
26	1.78	-0.36	1.42	1.73
27	2.12	-1.28	0.84	2.09
28	1.35	-0.42	0.93	1.36
29	1.90	-0.20	1.69	1.86
30	1.91	0.63	2.54	1.86

HIVPR: 1AJV



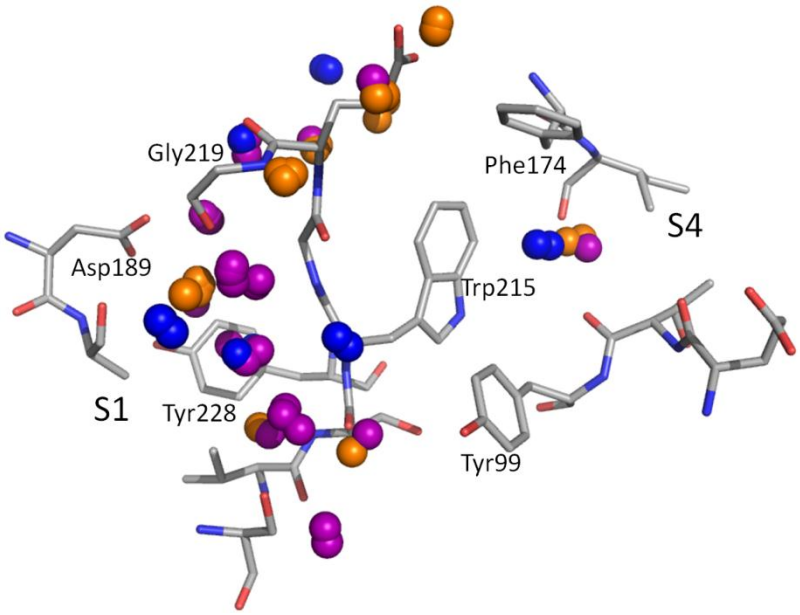
Hydration Site	$-T\Delta S_{hs}$ (3×3)	ΔE_{hs}	ΔG_{hs}	$-T\Delta S_{hs}$ (6×6)
1	3.83	-5.82	-2.00	3.82
2	3.52	-6.03	-2.51	3.49
3	4.03	-0.43	3.60	3.90
4	1.77	-1.35	0.42	1.77
5	1.95	-0.42	1.53	1.91
6	3.78	-1.84	1.95	3.77
7	2.17	-4.70	-2.52	2.22
8	1.76	-1.38	0.38	1.74
9	2.80	-5.26	-2.46	2.69
10	2.14	-0.11	2.03	2.14
11	1.84	0.83	2.67	1.85
12	1.60	1.89	3.49	1.58
13	1.94	1.70	3.64	1.92
14	1.61	0.32	1.93	1.59
15	1.55	-1.36	0.19	1.49
16	1.66	-0.42	1.24	1.68
17	2.33	-1.20	1.12	2.33
18	1.59	3.00	4.59	1.58
19	2.20	-0.43	1.78	2.27
20	1.47	-0.08	1.38	1.46
21	1.85	-1.61	0.24	1.81
22	1.56	1.25	2.81	1.57
23	1.52	-0.26	1.26	1.49
24	2.47	-3.12	-0.66	2.39
25	1.72	-3.85	-2.14	1.69
26	1.67	-0.91	0.76	1.75

pcDHFR: 1DAJ

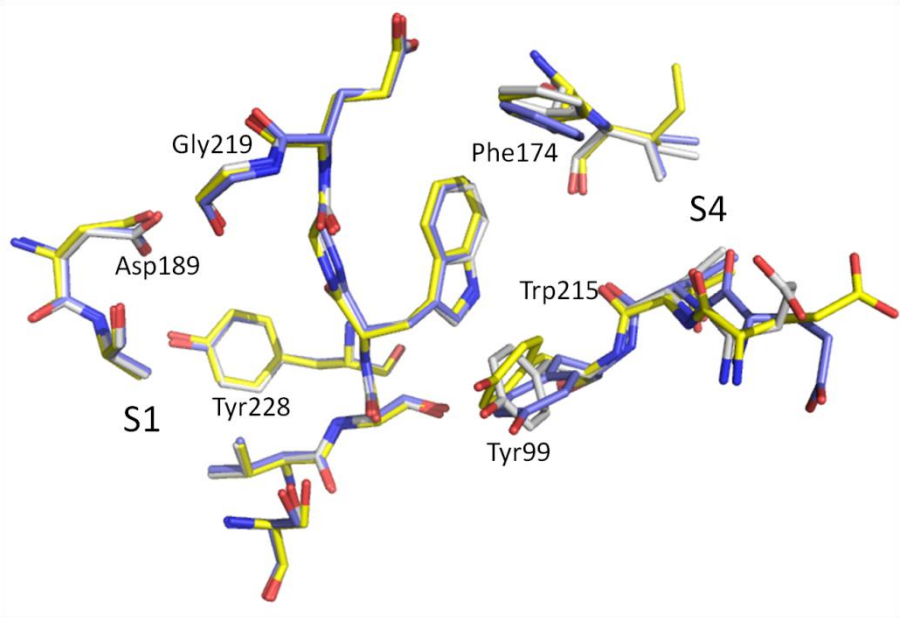


Hydration Site	$-T\Delta S_{hs}$ (3×3)	ΔE_{hs}	ΔG_{hs}	$-T\Delta S_{hs}$ (6×6)
1	4.30	-4.28	0.02	4.27
2	3.46	-4.15	-0.69	3.45
3	2.21	-0.29	1.91	2.19
4	1.45	-0.79	0.65	1.47
5	2.75	2.09	4.84	2.71
6	2.75	-1.31	1.44	2.76
7	2.33	-2.72	-0.39	2.35
8	2.68	-1.62	1.06	2.68
9	1.75	-2.12	-0.38	1.69
10	3.46	1.85	5.30	3.42
11	1.56	-0.97	0.58	1.58
12	2.24	-1.81	0.44	2.26
13	2.91	-0.19	2.72	2.94
14	1.92	-1.78	0.14	1.95
15	2.51	-2.39	0.12	2.52
16	2.32	-0.82	1.50	2.37
17	1.72	-3.26	-1.54	1.72
18	1.86	-0.37	1.48	1.80
19	1.72	-0.19	1.53	1.74
20	1.70	-3.46	-1.76	1.74
21	1.78	1.13	2.91	1.80
22	1.53	-1.05	0.49	1.56
23	1.61	-0.04	1.57	1.59
24	1.66	0.85	2.52	1.70
25	1.93	-0.36	1.57	1.84
26	1.70	2.76	4.47	1.69
27	1.56	-0.79	0.78	1.62

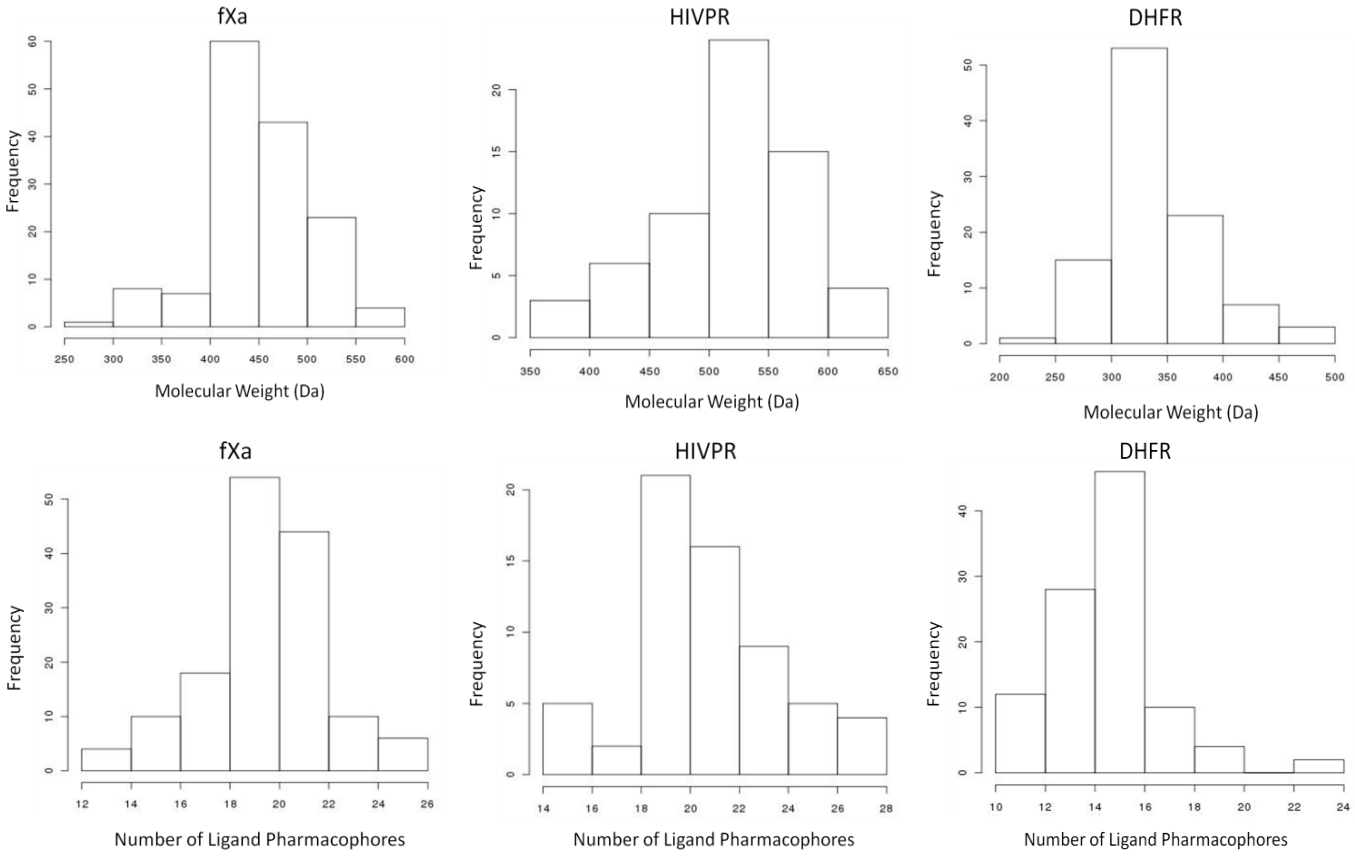
Supporting Information S2. Overlay of the hydration sites identified in the three structures of fXa showing there overlapping in the free energy estimation. The hydration sites of each fXa structures are color coded as in S1: Hydration sites that contribute both entropically and enthalpically when replaced by ligands are shown in blue. The hydration sites whose entropic gain surpasses the enthalpic loss when replaced by ligands are shown in purple. The hydration sites whose enthalpic loss is larger than the entropic gain when replaced by ligands are shown in orange.



Supporting Information S3. Alignment of the binding site residues in the three fXa structures (1F0S: yellow; 1MQ6: blue; 1NFU: grey). The RMSDs between the heavy atoms of the binding site residues in the three structures are as follows: 1F0S-1MQ6: 0.95 Å; 1F0S-1NFU: 0.93 Å; 1MQ6-1NFU: 0.83 Å.



Supporting Information S4. Distribution of molecular weight (top line) and number of ligand pharmacophore elements (bottom line) for all actives of fXa, HIVPR and DHFR.



Supporting Information S5. Comparison of the optimal Hydration Site Restricted Pharmacophore (HSRP) model (i.e. highest AUC) of each protein structures to Randomly-Selected Protein Pharmacophore (RSPP) models with identical number of pharmacophore elements. Corresponding to the optimal HSRP model of each protein structure, we randomly selected the same number of pharmacophore elements from the full set of protein pharmacophores. The virtual screening was then performed using the RSPP models. The AUC differences between the RSPP models and the optimal HSRP models are calculated. The random selection and screening process was performed three times. The average of the AUC differences for each protein structures and the standard errors are displayed. Negative values mean that the HSRP displays higher AUC values than the RSPP.