CReM-pharm: de novo design using 3D pharmacophores

Alina Denzler, Dinesh Kumar Sriramulu, Jozef Pecha, Pavel Polishchuk

Supplementary materials

Example to run the code

An example how to run structure generation with CReM-pharm for 8DV7 pharmacophore model. All necessary files can be found at links given in the Data availability section.

```
crempharm -q 8dv7_full.xyz --ids 2 6 7 -o 8dv7_out_dir -t 3 -f
starting_structures.dat -d chembl22_sa25_hac12.db-r 3 -x -n 20 --
conf_gen cdpkit --dist 1.5 -e 2.2 --mw 450 --tpsa 120 --rtb 7 --logp 4 -
c 4 -w 12
```

Values for `-w` (number of molecules processed in parallel) and `-c` (number of cores used for conformer enumeration for each molecule) can be adjusted based on a particular number of available CPUs. Their product can be greater than the total number of available CPUs.

Please note that starting feature ids are enumerated from 0 in the command line, but in the main text they are started from 1.

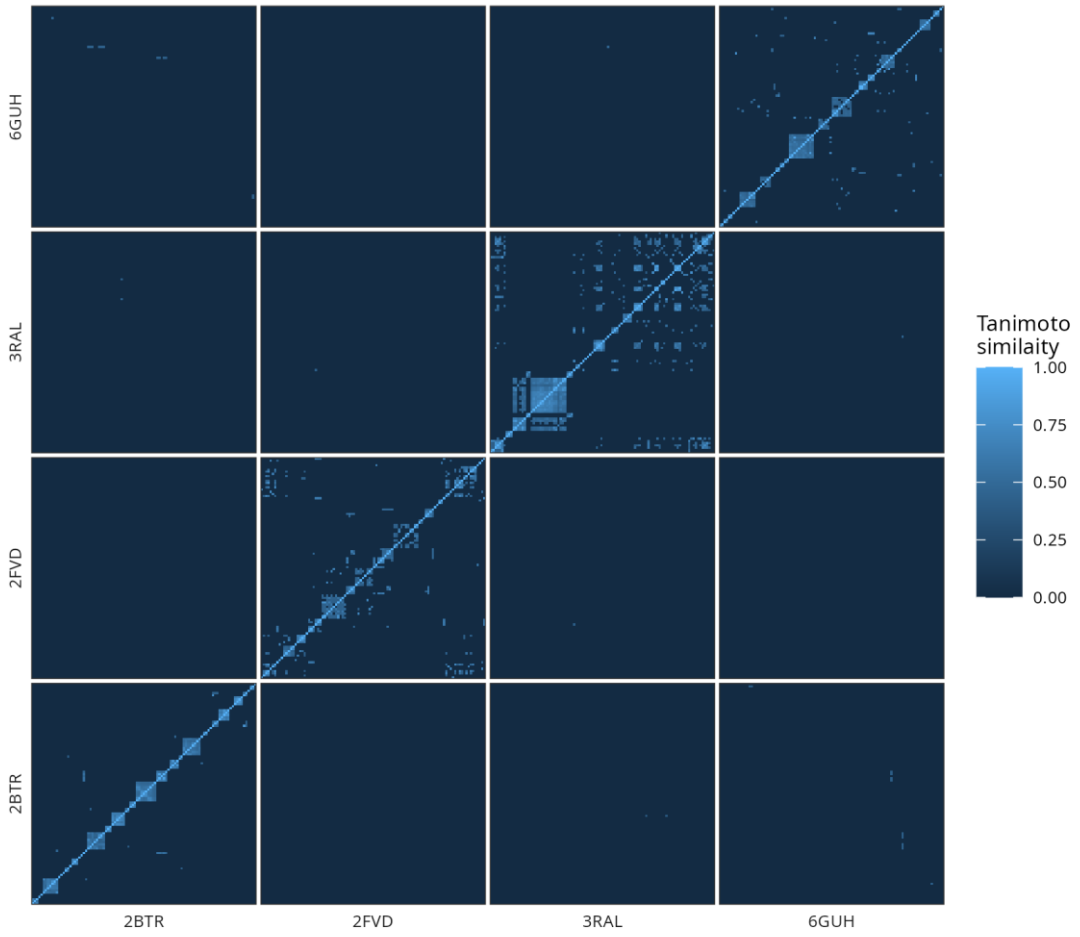


Figure S1. Pairwise Tanimoto (2048-bit Morgan fingerprints of radius 2) for top 100 compounds selected by best docking score, which were generated for different pharmacophore models of CDK2. Similarity values below 0.4 was assigned 0 to better highlight low similarity across different runs.

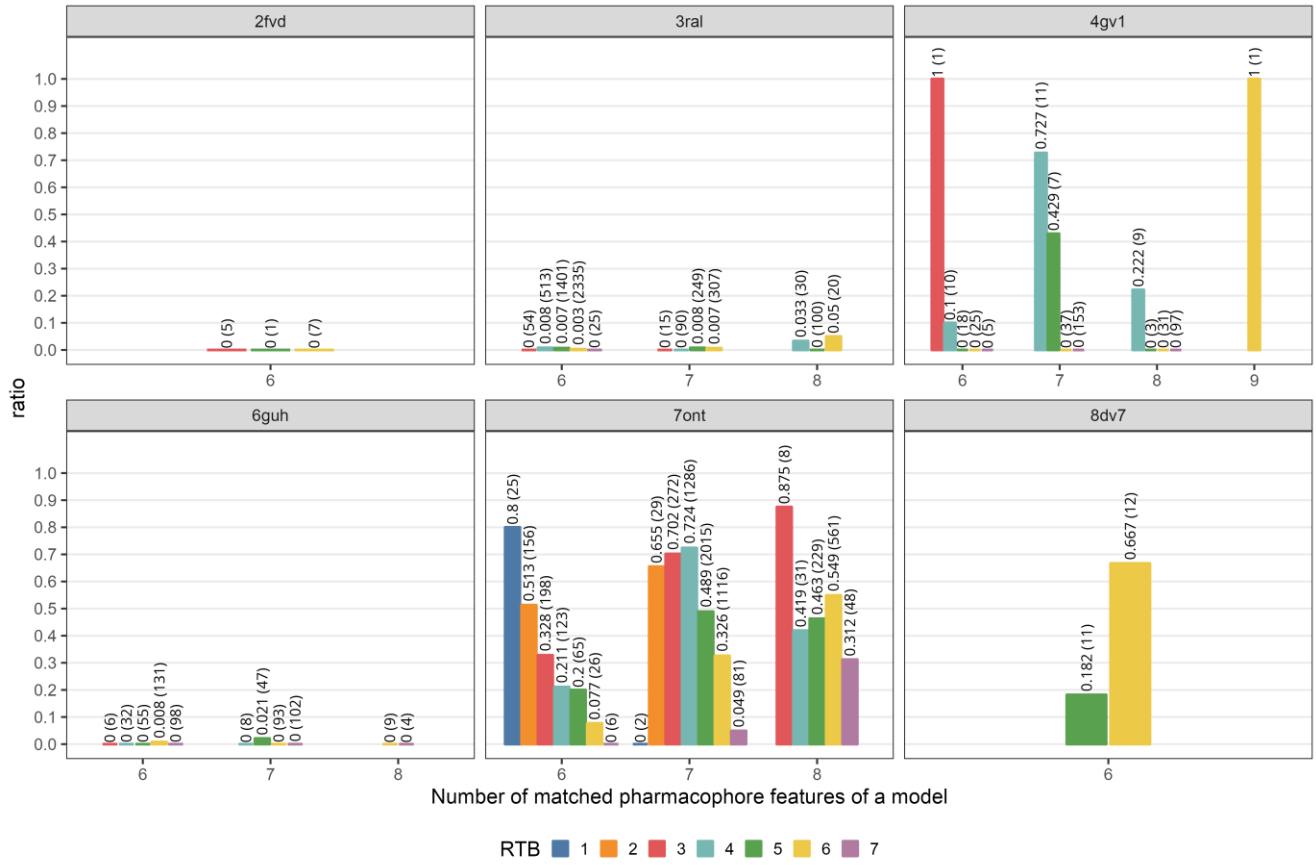


Figure S2. Reproducibility of the poses of de novo generated compounds in molecular docking (RMSD ≤ 2Å). Numbers designate the ratio of reproduced poses, numbers in brackets are the total number of molecules in a bin (with reproduced and non-reproduced poses).

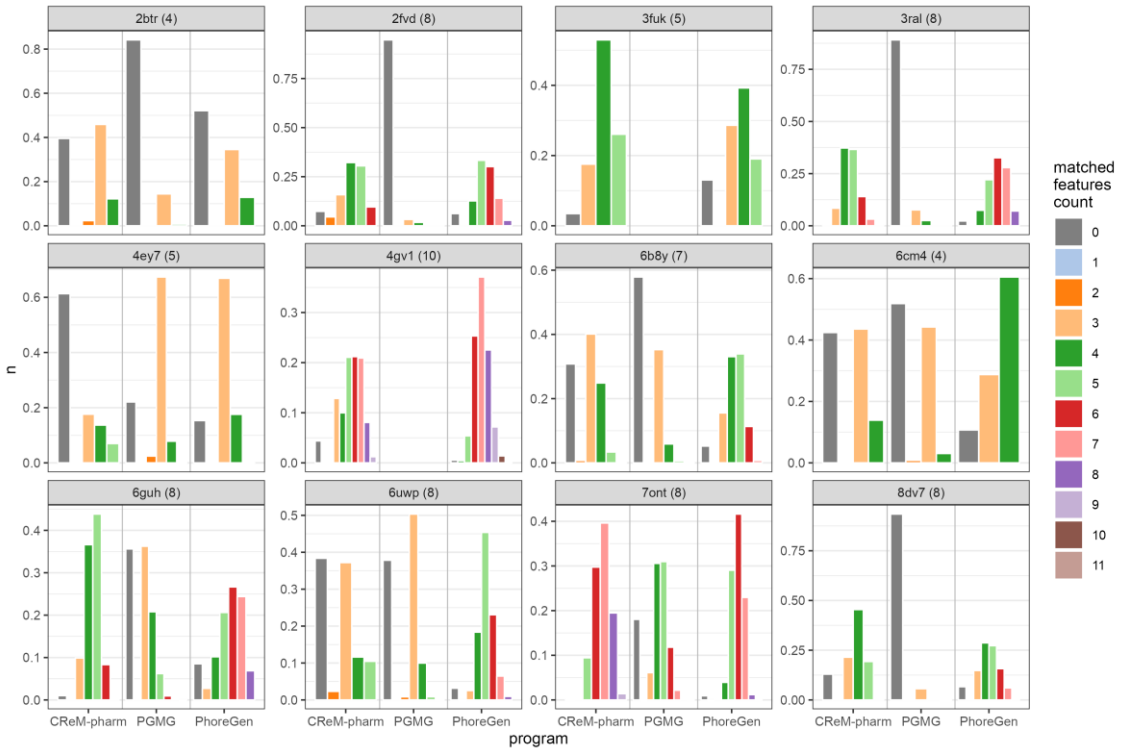


Figure S3. The number of structures matched a specific number of features in directed 3D pharmacophore models using conventional virtual screening by means of LigandScout.

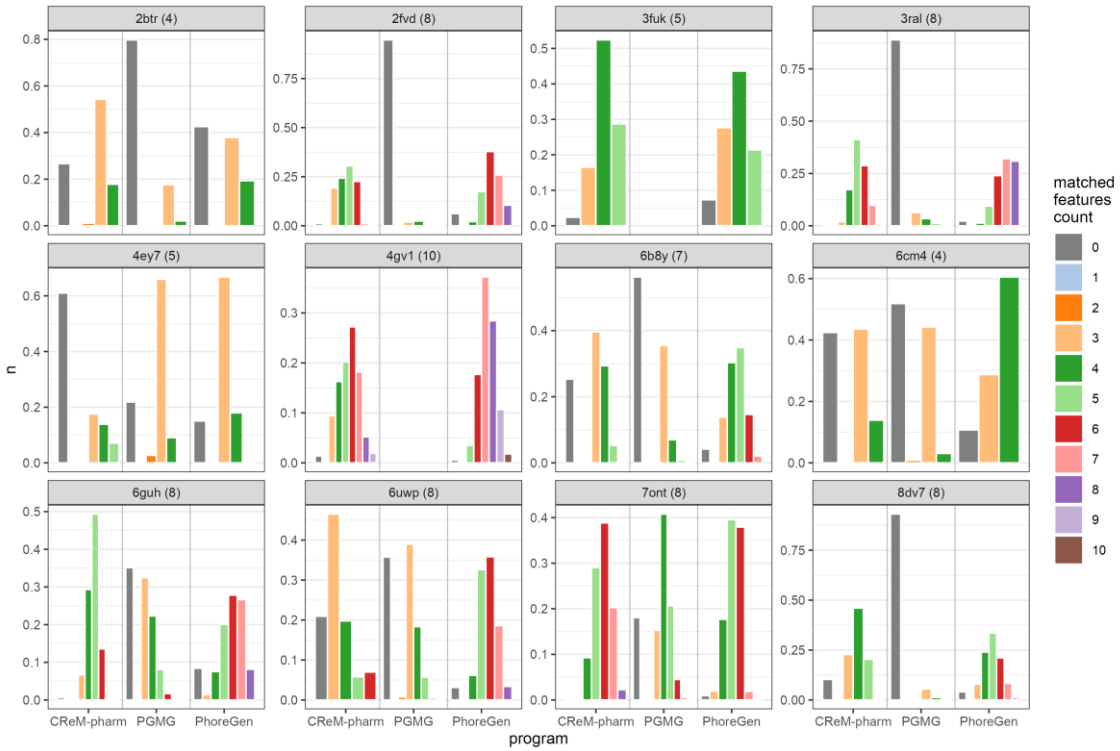


Figure S4. The number of structures matched a specific number of features in undirected 3D pharmacophore models using conventional virtual screening by means of LigandScout.

Table S1. Starting features and the total number of generated compounds matching some of features of the corresponding pharmacophore model

protein	starting features	n
3FUK	3,4	376
4EY7	4,5	2585
4GV1	1,6,10	731
6B8Y	3,7	6379
6CM4	1,3	13469
6UWP	1,2,6	86
7ONT	1,5,6,7	6565
8DV7	3,7,8	7098

Table S2. The number of compounds generated by PGMG and the number of compounds satisfying different criteria: (i) the number of compounds having the desired number of corresponding features disregarding their spatial arrangement, (ii) the number of compounds satisfying broader drug-like criteria (MW <= 500, logP <=5 and TPSA <= 120) and (iii) the number of compounds satisfying criteria applied for CReM (MW <= 450, logP <=4 and TPSA <= 120).

protein	n	n (satisfying feature counts)	ratio (satisfying feature counts)	n (broader physchem criteria)	ratio (broader physchem criteria)	n (crem physchem criteria)	ratio (crem physchem criteria)
2btr	7974	4394	0.551	3708	0.465	1926	0.242
2fvd	8301	6286	0.757	69	0.008	14	0.002
3ral	8662	5097	0.588	1240	0.143	470	0.054
4ey7	7452	5192	0.697	2889	0.388	1060	0.142
6b8y	7778	6274	0.807	640	0.082	109	0.014
6cm4	7405	7026	0.949	680	0.092	100	0.014
6guh	7709	4697	0.609	976	0.127	191	0.025
6uwp	8116	2869	0.353	1879	0.232	409	0.05
7ont	8489	7415	0.873	608	0.072	60	0.007
8dv7	8738	4641	0.531	73	0.008	11	0.001

Table S3. The number of features available in structures generated by PGMG corresponding the pharmacophore features of a model. This count disregards spatial arrangement of features. For each feature count there is the number of structures having this number of features and the number of structures missing at least one polar or non-polar feature. The sum of structures missing one polar or non-polar feature may be not equal to the total number of structures because in some structures both, polar and non-polar, features may be missing.

		number of structures		
protein	number of features	having the given number of features	missing at least one polar feature	missing at least one non-polar feature
2btr	2	108	108	3
2btr	3	3472	3456	16
2btr	4	4394	0	0
2fvd	3	24	24	0
2fvd	4	14	14	1
2fvd	5	37	36	7
2fvd	6	530	503	45
2fvd	7	1410	864	546
2fvd	8	6286	0	0
3ral	2	65	65	0
3ral	3	37	37	0
3ral	4	134	134	1
3ral	5	360	360	3
3ral	6	1118	1118	1
3ral	7	1851	1771	80
3ral	8	5097	0	0
4ey7	0	1	1	1
4ey7	1	1	1	1
4ey7	2	2	2	2
4ey7	3	55	50	33
4ey7	4	2201	2194	7
4ey7	5	5192	0	0
6b8y	1	2	2	2
6b8y	2	2	2	2
6b8y	3	1	1	1
6b8y	4	10	7	10
6b8y	5	234	184	87
6b8y	6	1255	840	415
6b8y	7	6274	0	0
6cm4	3	379	377	2
6cm4	4	7026	0	0
6guh	4	17	17	0
6guh	5	125	125	1
6guh	6	502	502	4
6guh	7	2368	2353	15
6guh	8	4697	0	0
6uwp	3	92	92	0

6uwp	4	162	162	0
6uwp	5	508	508	0
6uwp	6	1614	1614	1
6uwp	7	2871	2870	1
6uwp	8	2869	0	0
7ont	3	13	13	13
7ont	4	2	2	2
7ont	5	5	5	2
7ont	6	13	13	2
7ont	7	1041	1041	0
7ont	8	7415	0	0
8dv7	4	2	2	2
8dv7	5	104	104	2
8dv7	6	545	537	39
8dv7	7	3446	3400	46
8dv7	8	4641	0	0

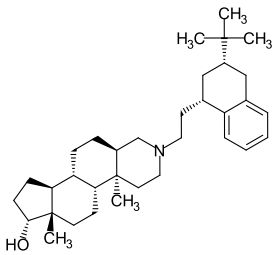
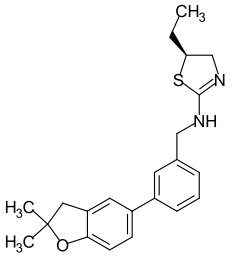
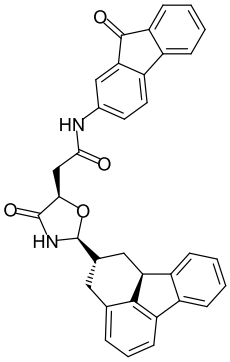
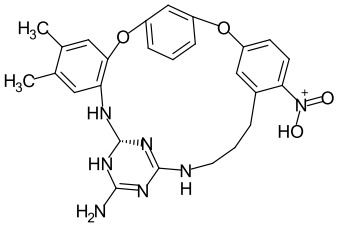
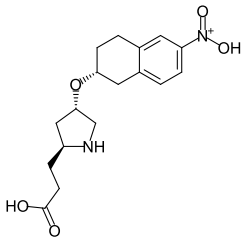
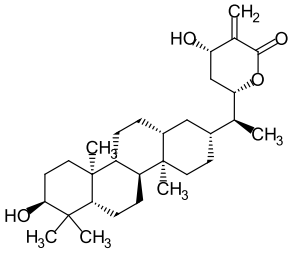
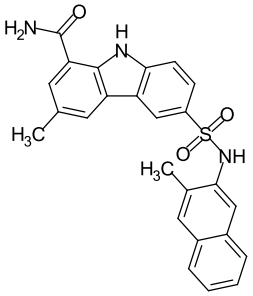
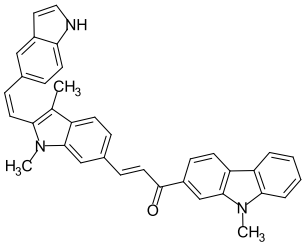
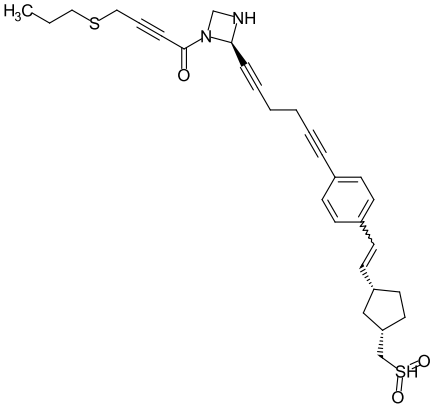
Table S4. The number of enumerated ligand structures for each set of settings, the number of generated structures matched a particular number of features and running time of individual generation tasks.

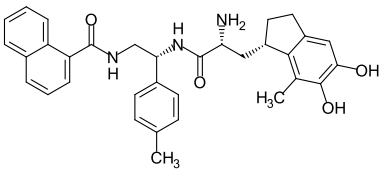
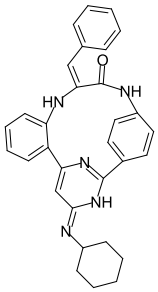
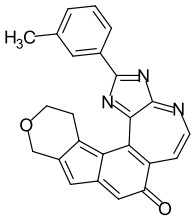
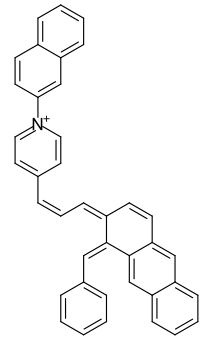
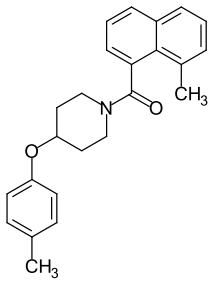
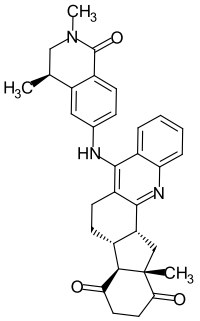
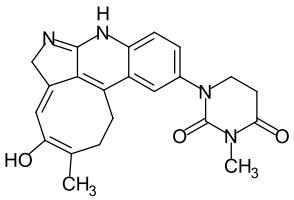
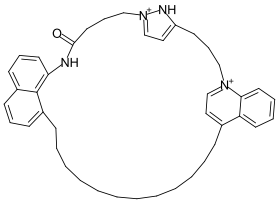
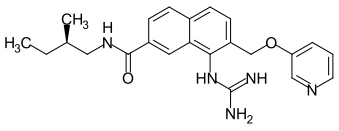
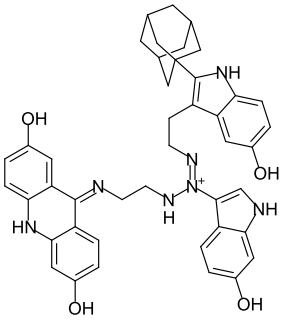
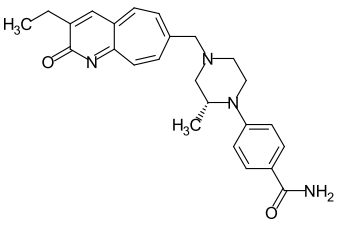
PDB model	starting features	CReM DB	radius	number of generated structures matching a particular number of features									total number of generated structures matching a pharmacophore model	total number of enumerated and embedded structures	generation time
				2	3	4	5	6	7	8	9	10			
3RAL	3,6,7	all	1	-	55	3283	3	2	2	30	-	-	3375	894601	126m (interrupted)
3RAL	3,6,7	all	2	-	55	5395	2810	2617	1168	309	-	-	12354	1882607	720m
3RAL	3,6,7	all	3	-	55	2035	9190	2866	935	752	-	-	15833	823964	720m
3RAL	3,6,7	all	4	-	55	735	2215	799	224	21	-	-	4049	139716	253m
3RAL	3,6,7	all	5	-	55	74	177	20	6	0	-	-	332	7500	7m 11s
3RAL	3,6,7	SA2.5	1	-	55	5692	6404	3873	1215	914	-	-	18153	2864314	720m
3RAL	3,6,7	SA2.5	2	-	55	4086	9279	4609	1028	225	-	-	19282	1812272	720m
3RAL	3,6,7	SA2.5	3	-	55	1156	3845	4328	661	150	-	-	10195	428929	720m
3RAL	3,6,7	SA2.5	4	-	55	488	1477	449	75	0	-	-	2544	77357	90m
3RAL	3,6,7	SA2.5	5	-	55	63	118	12	4	0	-	-	252	4518	4m 58s
3RAL	3,6,7	SA2	1	-	55	4421	9617	5447	743	280	-	-	20563	1720462	720m
3RAL	3,6,7	SA2	2	-	55	1993	10060	2558	1407	218	-	-	16291	845327	720m
3RAL	3,6,7	SA2	3	-	55	329	1030	918	118	29	-	-	2479	98708	79m 50s
3RAL	3,6,7	SA2	4	-	55	118	392	97	16	0	-	-	678	21015	11m 30s
3RAL	3,6,7	SA2	5	-	55	10	72	0	2	0	-	-	139	944	3m 25s
3RAL	1,6	SA2.5	3	1642	1595	1170	545	182	83	5	-	-	5222	441756	720m
3RAL	4,5,8	SA2.5	3	-	215	405	508	294	304	22	-	-	1748	241622	161m 46s
2BTR	3,4	SA2.5	3	1373	2641	2547	-	-	-	-	-	-	6561	1244015	720m
2FVD	4,7,8	SA2.5	3	-	36	41	87	13	0	0	-	-	177	28414	29m 22s
6GUH	1,5,8	SA2.5	3	-	232	8910	16800	322	250	13	-	-	26527	1276803	720m
3FUK	3,4	SA2.5	3	7	14	58	297	-	-	-	-	-	376	95904	102m 8s
4EY7	4,5	SA2.5	3	1605	342	11	627	-	-	-	-	-	2585	179798	502m
4GV1	1,6,10	SA2.5	3	-	166	6	151	59	208	140	1	0	731	89710	81m 12s
6B8Y	3,7	SA2.5	3	683	5206	399	91	0	0	-	-	-	6379	2075364	720m
6CM4	1,3	SA2.5	3	1222	7479	4768	-	-	-	-	-	-	13469	450532	720m
6UWP	1,2,6	SA2.5	3	-	78	6	2	0	0	0	-	-	86	1669	2m 10s
7ONT	1,5,6,7	SA2.5	3	-	-	26	262	599	4801	877	-	-	6565	252633	720m
8DV7	3,7,8	SA2.5	3	-	580	5831	664	23	0	0	-	-	7098	717505	390m 42s

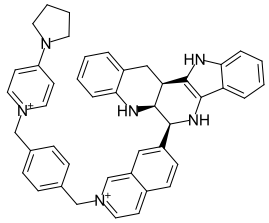
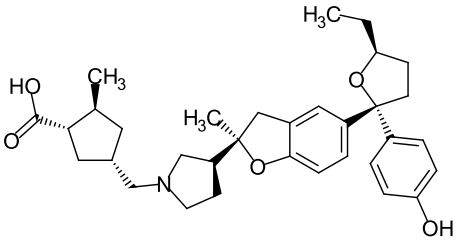
Table S5. The number of distinct compounds generated by PhoreGen and the number of compounds satisfying different criteria: (i) the number of compounds having the desired number of corresponding features disregarding their spatial arrangement, (ii) the number of compounds satisfying broader drug-like criteria (MW <= 500, logP <=5 and TPSA <= 120) and (iii) the number of compounds satisfying criteria applied for CReM (MW <= 450, logP <=4 and TPSA <= 120).

protein	n (distinct)	n (broader physchem criteria)	ratio (broader physchem criteria)	n (crem physchem criteria)	ratio (crem physchem criteria)
2btr	6455	6420	0.995	6165	0.955
2fvd	9565	5324	0.557	4073	0.426
3fuk	6527	6265	0.96	6004	0.92
3ral	8627	2124	0.246	1639	0.19
4ey7	8837	5890	0.667	4217	0.477
4gv1	9058	1942	0.214	1592	0.176
6b8y	9562	9183	0.96	7925	0.829
6cm4	6278	6102	0.972	5016	0.799
6guh	8043	7676	0.954	6198	0.771
6uwp	9435	7531	0.798	6446	0.683
7ont	9189	7726	0.841	5879	0.64
8dv7	9990	569	0.057	47	0.005

Table S6. Top scored structures among all structures generated by PGMG and PhoreGen

	PGMG	PhoreGen
2BTR	 -12.0 / 4.54	 -10.6 / 3.19
2FVD	 -13.9 / 4.04	 -12.1 / 6.12
3FUK	-	 -11.3 / 4.06
3RAL	 -13.7 / 5.07	 -13.2 / 2.39
4EY7	 -16.8 / 2.97	 -11.6 / 5.22

4GV1	-	 -12.0 / 3.7
6B8Y	 -14.2 / 4.67	 -13.4 / 3.31
6CM4	 -16.2 / 3.12	 -12.3 / 2.07
6GUH	 -15.1 / 4.42	 -12.2 / 3.64
6UWP	 -13.9 / 5.69	 -9.63 / 3.05
7ONT	 -14.9 / 5.21	 -12.2 / 2.85

8DV7	 -13.6 / 4.45	 -11.9 / 4.83