

Supplementary Information

Block-based characterization of protease specificity from substrate sequence profile

Enfeng Qi¹, Dongyu Wang², Bo Gao¹, Yang Li¹, Guojun Li^{1,*}

¹School of Mathematics, Shandong University, Jinan 250100, China

²The State Key Laboratory of Microbial Technology, Shandong University, Jinan 250100, China

*Corresponding authors: G. Li

Figure legend

Figure S1. Distribution of principal components.

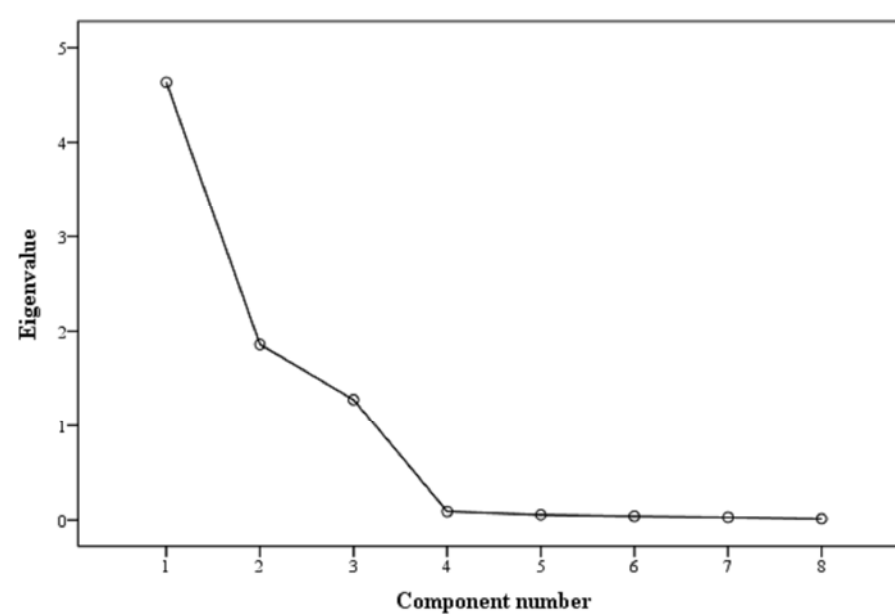


Figure S1. Distribution of principal components. The horizontal axis represents the principal components, and the vertical axis represents eigenvalues. The component 1, component 2 and component 3 are chosen for the subsequent analysis for their eigenvalues more than 1.

Table legends

Table S1. The entropies of eight blocks of 61 proteases.

Table S2. The distance matrix of 61 proteases.

Table S3. The Eigenvalues and corresponding contributions of the principal components.

Table S4. Principal components load matrix.

Table S1. The entropies of eight blocks of 61 proteases.

	Protease	MEROPS ^a	Catalytic Type	Number ^b	E ₄	E ₃	E ₂	E ₁	E ₁ '	E ₂ '	E ₃ '	E ₄ '
1	Pepsin A	A01.001	Aspartic	387	8.467	8.384	7.288	3.893	4.058	7.387	8.317	8.238
2	Cathepsin D	A01.009	Aspartic	727	9.407	8.759	6.352	2.706	3.767	7.206	9.134	9.443
3	Cathepsin E	A01.010	Aspartic	1287	10.218	9.346	6.517	2.802	3.516	7.043	9.543	10.214
4	Rhizopuspepsin	A01.012	Aspartic	240	7.850	7.826	7.100	4.113	4.088	6.986	7.701	7.712
5	Aspergillopepsin	A01.016	Aspartic	156	7.233	7.237	6.648	3.976	4.031	6.731	7.225	7.201
6	Necepsin 1	A01.053	Aspartic	117	6.853	6.830	6.417	3.732	3.903	6.383	6.802	6.853
7	HIV-1 Retropepsin	A02.001	Aspartic	632	8.280	8.306	6.290	3.350	3.582	6.371	8.586	9.108
8	Cathepsin L	C01.032	Cysteine	1065	8.536	9.056	6.569	3.809	3.646	7.301	9.446	9.938
9	Cathepsin L1	C01.033	Cysteine	192	7.483	7.452	6.689	3.913	4.006	6.835	7.477	7.539
10	Cathepsin S	C01.034	Cysteine	732	8.236	8.729	6.346	3.636	3.661	7.211	9.078	9.464
11	Falcipain 2	C01.046	Cysteine	137	6.656	6.797	6.254	3.842	3.919	6.276	6.753	6.784
12	Cathepsin B	C01.060	Cysteine	589	8.372	8.460	6.804	3.511	3.783	6.928	8.080	8.644
13	Falcipain 3	C01.063	Cysteine	114	6.319	6.710	6.113	3.896	3.856	5.924	6.403	6.365
14	Cathepsin K	C01.036	Cysteine	121	5.254	6.117	5.575	3.427	3.683	6.190	6.671	6.704
15	Calpain 1	C02.001	Cysteine	100	6.476	6.396	5.785	3.827	3.510	5.941	6.440	6.421
16	Calpain 2	C02.002	Cysteine	156	7.242	7.140	6.329	3.957	3.709	6.437	7.114	7.191
17	Caspase 1	C14.001	Cysteine	170	7.247	6.694	4.606	1.265	3.173	6.115	7.252	7.327
18	Caspase 3	C14.003	Cysteine	571	8.171	6.825	3.621	0.037	3.353	6.857	8.819	9.129
19	Caspase 7	C14.004	Cysteine	149	6.585	5.802	3.541	0.058	2.820	5.746	7.021	7.206
20	Caspase 6	C14.005	Cysteine	189	7.178	5.789	4.057	0.000	3.871	6.619	7.425	7.541
21	Thimet Oligopeptidase	M03.001	Metallo	46	4.665	4.863	4.648	3.321	3.137	4.735	5.081	3.644
22	Neurolysin	M03.002	Metallo	44	4.535	4.726	4.042	2.730	3.062	4.172	4.195	1.778
23	Thermolysin	M04.001	Metallo	260	7.697	7.740	7.063	4.067	2.951	6.299	7.593	7.675
24	MMP 8	M10.002	Metallo	104	6.332	6.275	5.683	3.395	3.407	5.804	6.284	6.381
25	MMP 2	M10.003	Metallo	2502	10.404	9.781	7.432	3.863	3.037	6.865	9.655	11.021
26	MMP 9	M10.004	Metallo	340	7.868	7.637	6.651	3.572	3.636	6.897	8.017	8.290
27	MMP 3	M10.005	Metallo	171	7.213	7.033	6.209	3.582	3.682	6.434	7.162	7.292
28	MMP 7	M10.008	Metallo	174	7.317	7.112	6.297	3.705	2.542	5.586	6.983	7.222
29	MMP 12	M10.009	Metallo	171	7.071	6.645	5.897	3.390	3.482	6.257	6.972	7.304
30	MMP 13	M10.013	Metallo	138	5.610	6.194	5.647	3.039	3.408	6.062	6.536	5.663
31	Membrane-MMP 1	M10.014	Metallo	127	6.957	6.863	6.343	3.632	3.453	6.145	6.957	6.973
32	Astacin	M12.001	Metallo	206	7.634	7.570	6.763	3.891	2.791	5.830	7.354	7.657
33	Meprin	M12.002	Metallo	766	9.496	9.297	7.556	3.867	3.555	7.102	9.122	9.522
34	LAST MAM Peptidase	M12.033	Metallo	429	8.726	8.545	7.236	3.898	3.130	6.179	8.113	8.641
35	Neprilysin	M13.001	Metallo	102	5.953	6.238	6.001	3.872	2.976	5.610	5.756	5.190
36	Peptidyl-Lys Metallopeptidase	M35.004	Metallo	1869	10.823	10.464	8.117	4.123	0.000	4.174	8.125	10.456
37	Chymotrypsin A	S01.001	Serine	711	9.274	8.758	6.287	2.536	3.995	7.573	9.086	9.247
38	Granzyme B	S01.010	Serine	1568	9.884	8.182	5.119	1.525	3.727	7.362	9.795	10.475
39	Elastase 2	S01.131	Serine	364	7.936	8.147	6.298	2.860	4.015	7.223	8.354	8.465
40	Cathepsin G	S01.133	Serine	244	7.034	7.526	6.303	3.021	3.944	6.919	7.707	7.795
41	Granzyme A	S01.135	Serine	286	8.036	7.494	5.150	1.520	3.536	6.329	7.814	7.978
42	Granzyme B rodent	S01.136	Serine	341	8.158	7.182	4.411	0.969	3.698	6.641	8.022	8.329
43	Chymase	S01.140	Serine	100	6.624	6.389	4.875	1.716	3.908	5.737	6.448	6.449
44	Trypsin 1	S01.151	Serine	9014	12.162	9.111	5.095	0.996	4.054	8.116	11.601	12.897
45	Thrombin	S01.217	Serine	136	6.512	6.301	3.317	0.570	3.255	6.047	6.791	6.928
46	Plasmin	S01.233	Serine	121	6.797	6.718	5.031	1.240	3.761	6.199	6.774	6.791
47	KLK 4	S01.251	Serine	112	6.390	6.297	4.537	1.534	3.833	6.001	6.513	6.591
48	Glutamyl Peptidase I	S01.269	Serine	1045	9.115	8.017	4.586	0.549	3.992	7.731	9.675	9.981
49	Lysyl Peptidase	S01.280	Serine	778	9.371	7.851	4.112	0.000	4.071	7.869	9.431	9.580
50	Lactocepin 1	S08.019	Serine	101	6.638	6.619	5.980	3.737	3.960	6.106	6.619	6.611
51	Kexin	S08.070	Serine	171	6.737	4.516	0.896	0.052	3.011	5.863	7.063	7.359
52	Furin	S08.071	Serine	172	6.012	5.494	2.540	0.211	3.491	5.855	7.083	7.312
53	PCSK2 Peptidase	S08.073	Serine	166	6.659	5.628	2.393	0.457	3.734	6.214	7.108	7.221
54	PCSK4 Peptidase	S08.074	Serine	94	5.551	5.122	2.083	0.204	3.335	5.234	6.419	6.504
55	PCSK6 Peptidase	S08.075	Serine	94	5.543	5.121	2.009	0.204	3.391	5.304	6.448	6.533
56	PCSK5 Peptidase	S08.076	Serine	116	5.904	5.389	2.333	0.216	3.546	5.583	6.724	6.789
57	PCSK7 Peptidase	S08.077	Serine	102	5.619	5.172	2.099	0.239	3.333	5.351	6.547	6.594
58	KPC2-type Peptidase	S08.109	Serine	104	4.384	3.758	1.616	0.534	3.079	5.318	6.359	6.566
59	Lactocepin 3	S08.116	Serine	158	7.291	7.266	6.364	3.853	4.031	6.743	7.240	7.291
60	Signal Peptidase 1	S26.001	Serine	297	7.261	5.297	4.037	0.518	2.970	5.788	7.538	7.882
61	Signalase 21kDa	S26.010	Serine	316	8.081	7.136	5.808	2.398	3.672	6.946	8.034	8.186

^aThe index of 61 proteases in the MEROPS database. ^bThe number of substrate sequences which have been removed redundancy.

Table S2. The distance matrix of 61 preteases^a.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	
1	0.000	2.356	3.406	1.263	2.441	3.324	1.850	2.308	1.923	1.844	3.562	0.951	4.261	5.025	4.404	2.747	4.777	5.715	6.778	5.992	8.580	10.526	2.083	4.728	4.072	1.392	2.818	3.509	3.385	5.326	3.304	2.633	2.144	1.619	5.724	6.509	2.299	4.434	1.571	2.307	3.591	4.377	5.233	7.773	6.439	4.866	5.454	4.904	5.443	3.922	8.907	7.452	7.115	8.463	8.478	7.860	8.320	9.665	2.442	6.118	2.562	
2	2.356	0.000	1.375	3.322	4.209	4.914	1.743	1.574	3.681	1.501	5.133	1.960	5.814	6.322	5.815	4.316	4.861	4.511	6.643	5.769	9.875	11.756	3.648	6.013	2.821	2.676	4.207	4.666	4.565	6.640	4.746	3.916	1.777	2.420	7.242	5.955	0.528	2.237	2.048	3.482	3.324	3.694	5.907	5.581	6.451	5.339	6.025	3.053	3.716	5.390	8.542	7.322	6.937	8.525	8.529	7.844	8.357	9.870	4.065	5.863	2.767	
3	3.406	1.375	0.000	4.468	5.438	6.158	2.790	2.024	4.903	2.421	6.375	3.077	7.040	7.594	7.037	5.526	6.037	5.324	7.729	6.909	11.101	12.962	4.686	7.242	1.765	3.889	5.449	5.732	5.799	7.919	5.960	4.940	1.844	1.777	3.162	8.437	5.347	1.737	2.309	3.384	4.793	4.495	4.774	7.153	4.691	7.593	6.590	7.279	3.545	4.242	6.635	9.516	8.423	8.039	9.637	9.642	8.960	9.468	11.002	5.313	6.917	4.068
4	1.263	3.322	4.468	0.000	1.228	2.131	2.240	3.265	0.770	2.638	2.360	1.495	3.049	3.916	3.229	1.588	4.234	5.800	6.219	5.574	7.413	9.399	1.346	3.589	5.139	1.085	1.746	2.591	2.360	4.222	2.178	1.849	3.276	1.988	4.546	7.011	3.267	5.213	1.839	1.618	3.390	4.288	4.362	8.795	5.831	4.161	4.625	5.502	5.952	2.738	8.374	6.857	6.564	7.733	7.754	7.191	7.607	8.853	1.299	5.638	2.378	
5	2.441	4.209	5.438	1.228	0.000	0.928	2.929	4.237	0.552	3.517	1.147	2.439	1.861	2.755	2.041	0.561	3.599	5.698	5.574	4.982	6.220	8.247	1.537	2.412	1.639	1.065	0.783	1.951	1.373	3.046	1.109	1.694	4.222	2.876	3.412	7.725	4.128	5.838	2.410	1.335	2.390	4.116	3.394	9.624	5.042	3.360	3.607	5.907	6.247	1.534	7.615	6.056	5.812	6.800	6.825	6.318	6.690	7.839	0.330	5.012	2.394	
6	3.324	4.914	6.158	2.131	0.928	0.000	3.571	5.024	1.426	4.275	0.313	3.212	1.015	1.999	1.186	0.747	3.252	5.743	4.966	4.649	5.340	7.390	2.070	1.519	6.991	2.375	0.782	1.723	0.946	2.205	0.571	1.949	5.247	3.564	2.583	8.176	4.837	6.375	3.068	1.741	3.281	4.155	2.720	10.257	4.533	2.865	3.028	6.315	6.574	0.675	7.127	5.538	5.357	6.145	6.179	5.728	6.054	7.135	0.964	4.620	2.724	
7	1.850	1.743	2.790	2.240	2.929	3.571	0.000	1.794	2.434	1.161	3.747	1.069	4.392	4.895	4.405	2.945	4.005	4.545	5.881	5.264	8.523	10.508	2.284	4.624	3.688	1.439	2.857	3.210	3.204	5.320	3.346	2.453	2.307	1.461	5.861	5.895	1.934	3.410	1.328	2.281	2.692	3.391	4.822	7.029	5.648	4.413	4.982	3.937	4.560	4.014	7.918	6.537	6.258	7.617	7.632	7.008	7.466	8.857	2.788	5.149	2.011	
8	2.308	1.574	2.024	3.265	4.237	5.024	1.794	0.000	3.700	0.804	5.203	2.056	5.883	6.321	5.934	4.379	5.553	5.419	7.355	6.610	10.100	12.109	3.577	6.206	2.558	2.730	4.359	4.792	4.772	6.823	4.874	3.943	1.511	2.395	7.377	5.899	1.771	3.212	2.381	3.626	4.172	4.630	6.458	6.121	7.113	5.978	6.562	4.042	4.830	5.522	9.217	7.928	7.616	9.102	9.104	8.458	8.934	10.283	4.140	6.605	3.415	
9	1.923	3.681	4.903	0.770	0.552	1.426	2.434	3.700	0.000	2.987	1.650	1.904	2.372	3.178	2.515	0.888	3.646	5.482	5.587	5.017	6.726	8.746	1.313	2.860	5.685	1.139	1.008	2.095	1.626	3.508	1.482	1.600	3.889	2.420	3.916	7.373	3.608	5.368	1.931	1.132	3.031	3.930	3.668	9.113	5.181	3.525	3.921	5.506	5.894	2.019	7.732	6.193	5.925	7.025	7.047	6.505	6.904	8.110	0.555	5.054	2.093	
10	1.844	1.501	2.421	2.638	3.517	4.275	1.161	0.804	2.987	0.000	4.451	1.442	5.132	5.551	5.172	3.640	4.808	4.941	6.645	5.921	9.324	11.331	2.951	5.432	3.211	2.008	3.597	4.096	4.002	6.039	4.124	3.294	1.859	2.023	6.628	6.131	1.607	3.247	1.637	2.847	3.475	3.999	5.672	6.575	6.374	5.204	5.775	3.862	4.584	4.760	8.561	7.219	6.911	8.369	8.372	7.731	8.204	9.554	3.397	5.930	2.680	
11	3.562	5.133	6.375	2.360	1.147	0.313	3.747	5.203	1.650	4.451	0.000	3.434	0.755	1.771	0.961	0.920	3.294	5.843	4.942	4.693	5.118	7.190	2.262	1.332	7.201	2.596	0.972	1.782	1.019	2.021	0.678	2.101	5.471	3.774	2.396	8.337	5.060	6.564	3.275	1.923	3.421	4.271	2.662	10.459	4.485	2.886	2.962	6.485	6.737	0.444	7.051	5.463	5.300	6.023	6.057	5.630	5.937	6.978	1.167	4.642	2.924	
12	0.951	1.960	3.077	1.495	2.439	3.212	1.069	2.056	1.904	1.442	3.434	0.000	4.114	4.800	4.204	2.596	4.209	5.088	6.193	5.494	8.402	10.357	1.901	4.468	3.885	1.054	2.582	3.129	3.065	5.147	3.092	2.275	2.144	1.207	5.593	6.108	1.989	3.954	1.103	2.011	2.946	3.750	4.818	7.484	5.881	4.415	5.019	4.412	4.961	3.756	8.338	6.888	6.574	7.921	7.937	7.316	7.777	9.174	2.353	5.540	2.073	
13	4.261	5.814	7.040	3.049	1.861	1.015	4.392	5.883	2.372	5.132	0.755	4.114	0.000	1.511	0.601	1.600	3.494	6.257	4.944	4.890	4.439	6.511	2.825	0.926	7.854	3.311	1.655	2.035	1.526	1.566	1.201	2.571	6.141	4.373	1.744	8.722	5.749	7.209	3.965	2.593	3.873	4.701	2.556	11.131	4.465	2.977	2.883	7.081	7.290	0.553	7.010	5.424	5.319	5.810	5.852	5.500	5.747	6.709	1.892	4.796	3.561	
14	5.025	6.322	7.594	3.916	2.755	1.999	4.895	6.321	3.178	5.551	1.771	4.800	1.511	0.000	1.415	2.519	3.303	5.947	4.322	4.399	4.135	6.336	3.683	1.296	8.486	3.893	2.398	2.807	2.045	1.217	2.073	3.422	6.857	5.203	2.202	9.522	6.224	7.409	4.390	2.952	4.053	4.652	2.396	11.373	3.892	2.802	2.466	7.072	7.260	1.589	6.260	4.605	4.650	4.980	5.014	4.709	4.912	5.623	2.715	4.283	3.843	
15	4.404	5.815	7.037	3.229	2.041	1.186	4.405	5.934	2.515	5.172	0.961	4.204	0.601	1.415	0.000	1.669	3.203	5.996	4.591	4.616	4.282	6.383	2.886	0.537	7.886	3.347	1.656	1.870	1.335	1.429	1.190	2.564	6.231	4.441	1.652	8.691	5.762	7.092	4.017	2.673	3.718	4.441	2.347	11.036	4.143	2.801	2.641	6.953	7.134	0.648	6.579	5.084	4.970	5.464	5.506	5.162	5.396	6.338	2.017	4.414	3.443	
16	2.747	4.316	5.526	1.588	0.561	0.747	2.945	4.379	0.888	3.640	0.920	2.596	1.600	2.519	1.669	0.000	3.293	5.534	5.145	4.788	5.915	7.954	1.476	2.037	6.323	1.770	0.425	1.476	0.948	2.787	0.698	1.422	4.602	2.917	3.108	7.594	4.273	5.854	2.564	1.440	3.039	3.951	3.099	9.696	4.723	3.127	3.383	5.919	6.235	1.249	7.263	5.746	5.515	6.449	6.479	5.991	6.342	7.493	0.503	4.701	2.346	
17	4.777	4.861	6.037	4.234	3.599	3.252	4.005	5.553	3.646	4.808	3.294	4.209	3.494	3.303	3.203	3.293	0.000	3.104	2.039	1.882	7.591	10.382	3.938	2.949	7.365	3.622	2.903	3.124	2.532	3.241	3.004	3.580	6.062	4.684	4.210	8.479	4.761	5.267	3.527	2.894	1.595	1.829	1.829	1.691	9.222	1.795	1.120	1.583	4.686	4.627	3.148	4.513	2.923	2.718	3.875	3.903	3.282	3.735	5.321	3.355	1.837	2.431
18	5.715	4.511	5.324	5.284	5.698	5.743	4.545	5.419	5.482	4.941	5.843	5.088	6.257	5.947	5.996	5.534	3.104	0.000	3.467	2.653	8.751	10.382	5.765	5.857	6.876	4.884	5.161	5.552	4.979	6.160	5.517	5.591	6.194	5.700	7.227	8.824	4.395	3.521	4.263	4.611	2.766	1.750	4.620	6.948	3.609	3.948	4.389	2.478	2.204	5.810	4.705	3.869	3.540	5.239	5.216	4.519	5.046	6.564	5.422	2.838	3.475	
19	6.778	6.643	7.729	6.219	5.474	4.966	5.881	7.355	5.587	6.645	4.942	6.193	4.944	4.322	4.591	5.145	2.039	3.467	0.000	1.663	5.832	7.263	5.860	4.193	9.102	5.585	4.764	4.823	4.280	4.182	4.744	5.443	7.967	6.637	5.186	9.968	6.542	6.559	5.455	4.801	3.432	3.131	2.639	10.289	0.989	2.414	2.277	5.719	5.461	4.701	2.958	1.388	1.605	2.250	2.285	1.714	2.112	3.749	5.248	1.386	4.335	
20	5.992	5.769	6.909	5.574	4.98																																																									

Table S3. The Eigenvalues and corresponding contributions of the principal components.

Component	Total	% of Variance	Cumulative %
1	4.635	57.938	57.938
2	1.863	23.284	81.223
3	1.277	15.960	97.183
4	0.090	1.124	98.307
5	0.055	0.682	99.989
6	0.040	0.494	99.483
7	0.028	0.351	99.834
8	0.013	0.166	100.000

Table S4. Principal components load matrix^a.

	Component		
	1	2	3
E ₄	0.955	-	-
E ₃	0.803	-	-
E ₂	-	0.939	-
E ₁	-	0.983	-
E ₁ '	-	-	0.991
E ₂ '	0.692	-	0.690
E ₃ '	0.973	-	-
E ₄ '	0.982	-	-

^aThe coefficients < 0.6 are not shown in the load matrix for their weak influencing to entropies.