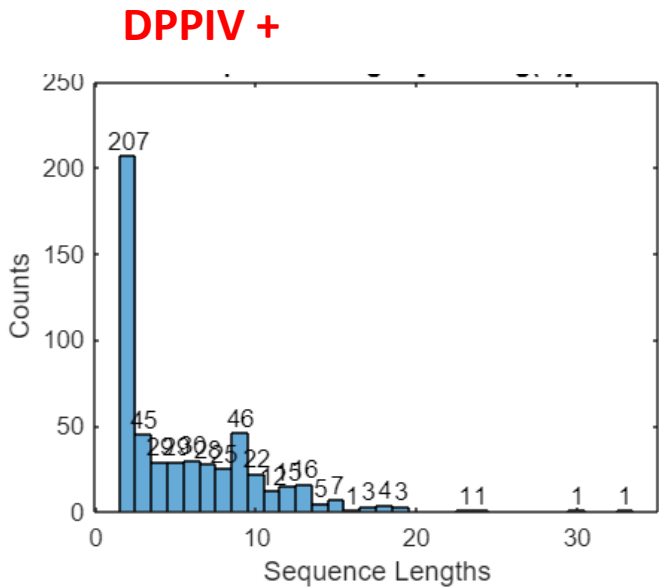


**Fig S1:** The bar plot of count of peptides at each sequence length for both classes—DPP-IV inhibitory (DPPIV +) and non-inhibitory (DPPIV -)—from the original dataset (Wang et al., 2024).

DPPIV +

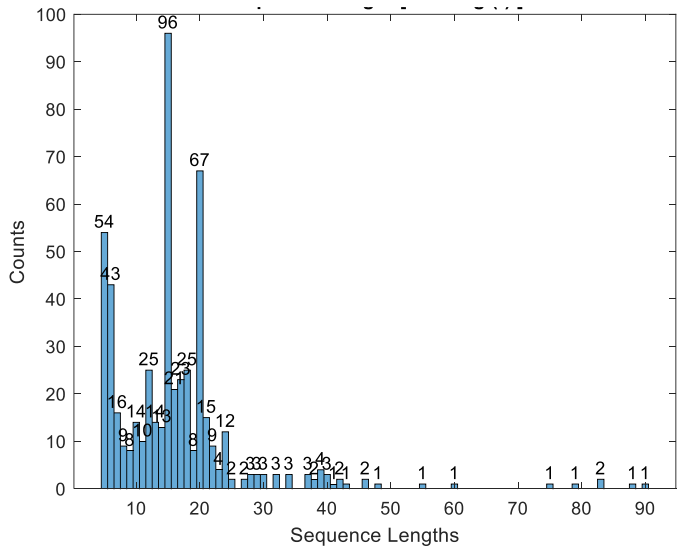
Sequence Length	Counts
2	207
3	45
4	29
5	29
6	30
7	28
8	25
9	46
10	22
11	12
12	15
13	16
14	5
15	7
16	1
17	3
18	4
19	3
23	1
24	1
30	1
33	1



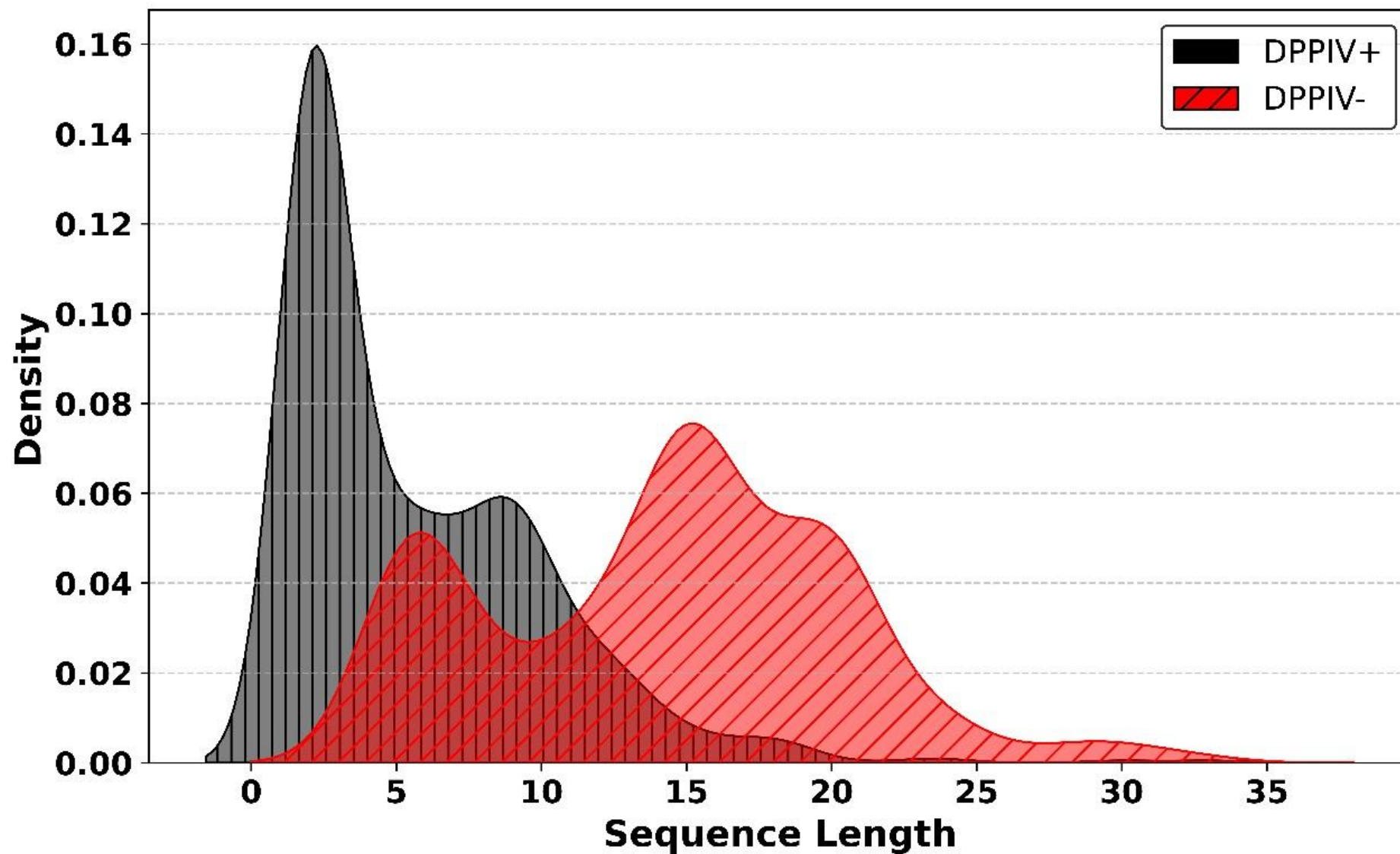
DPPIV -

Sequence Length	Counts
27	2
28	3
29	3
30	3
32	3
34	3
37	3
38	2
39	4
40	3
41	1
42	2
43	1
46	2
48	1
55	1
60	1
75	1
79	1
83	2
88	1
90	1

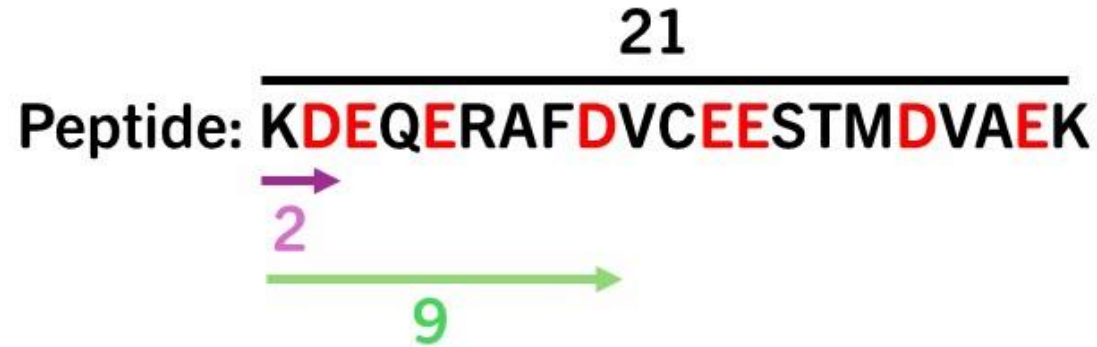
Sequence Length	Counts
5	54
6	43
7	16
8	9
9	8
10	14
11	10
12	25
13	14
14	13
15	96
16	21
17	23
18	25
19	8
20	67
21	15
22	9
23	4
24	12
25	2



**Fig S2: The peptide sequence length distribution from whole dataset (DPPIV +: 664; DPPIV -: 629)**



**Fig S3: Illustration of the calculations of distribution descriptor (D) with a sample peptide sequence**



The first residue of the **negative charge class** at the position from the N-terminal of the peptide sequence:  $2/21 = 0.095$

The 50% residue of the **negative charge class**: total no. of negative residue  $\times 0.5 = 8 \times 0.25 = 4$

The 50% residue of the **negative charge class** (2 residues) at the position from the N-terminal of the peptide sequence:  $9/21 = 0.429$

Residue = amino acid

Fig S4: CTD-C descriptor differences between DPPIV+ and DPPIV- classes

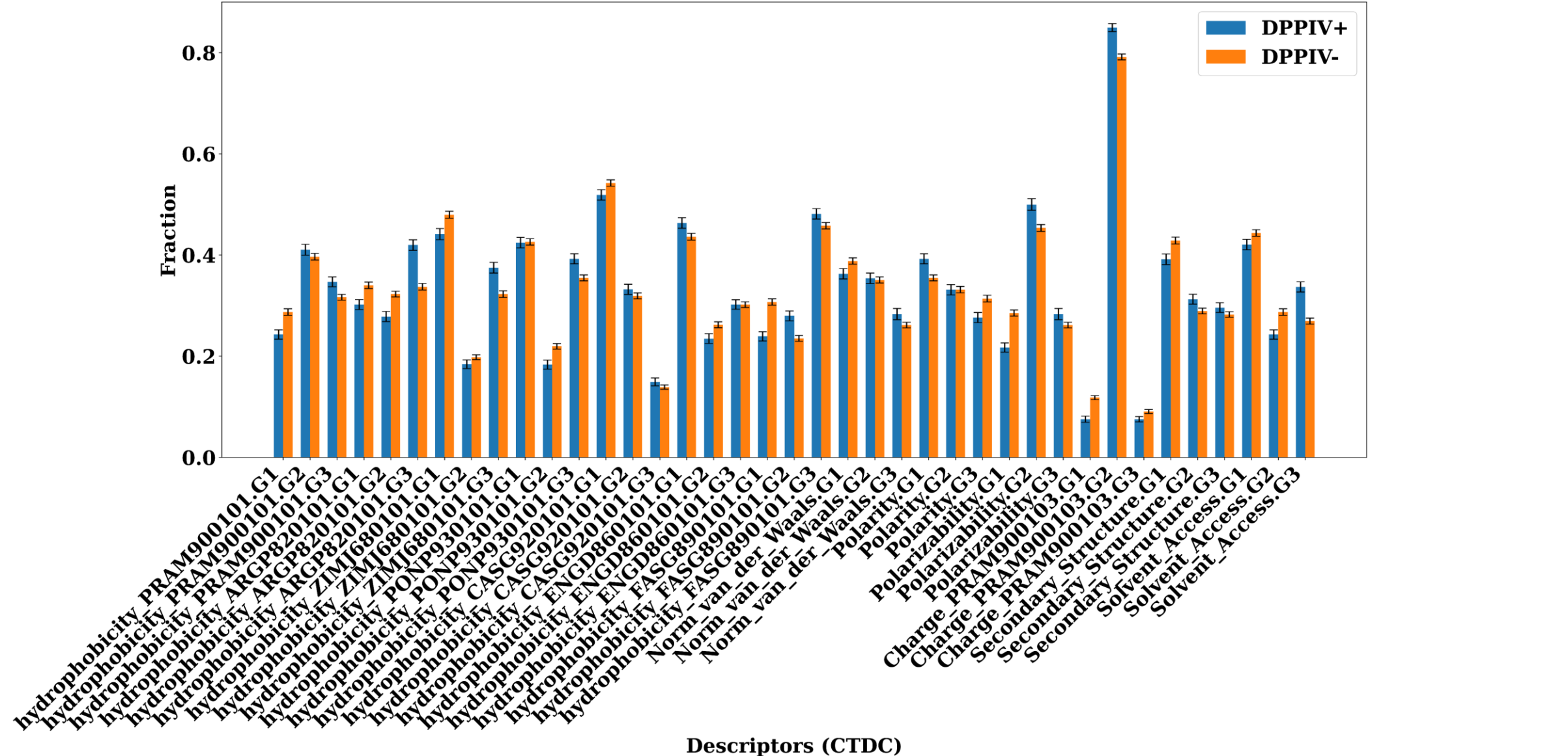


Fig S5: CTD-T descriptor differences between DPPIV+ and DPPIN- classes

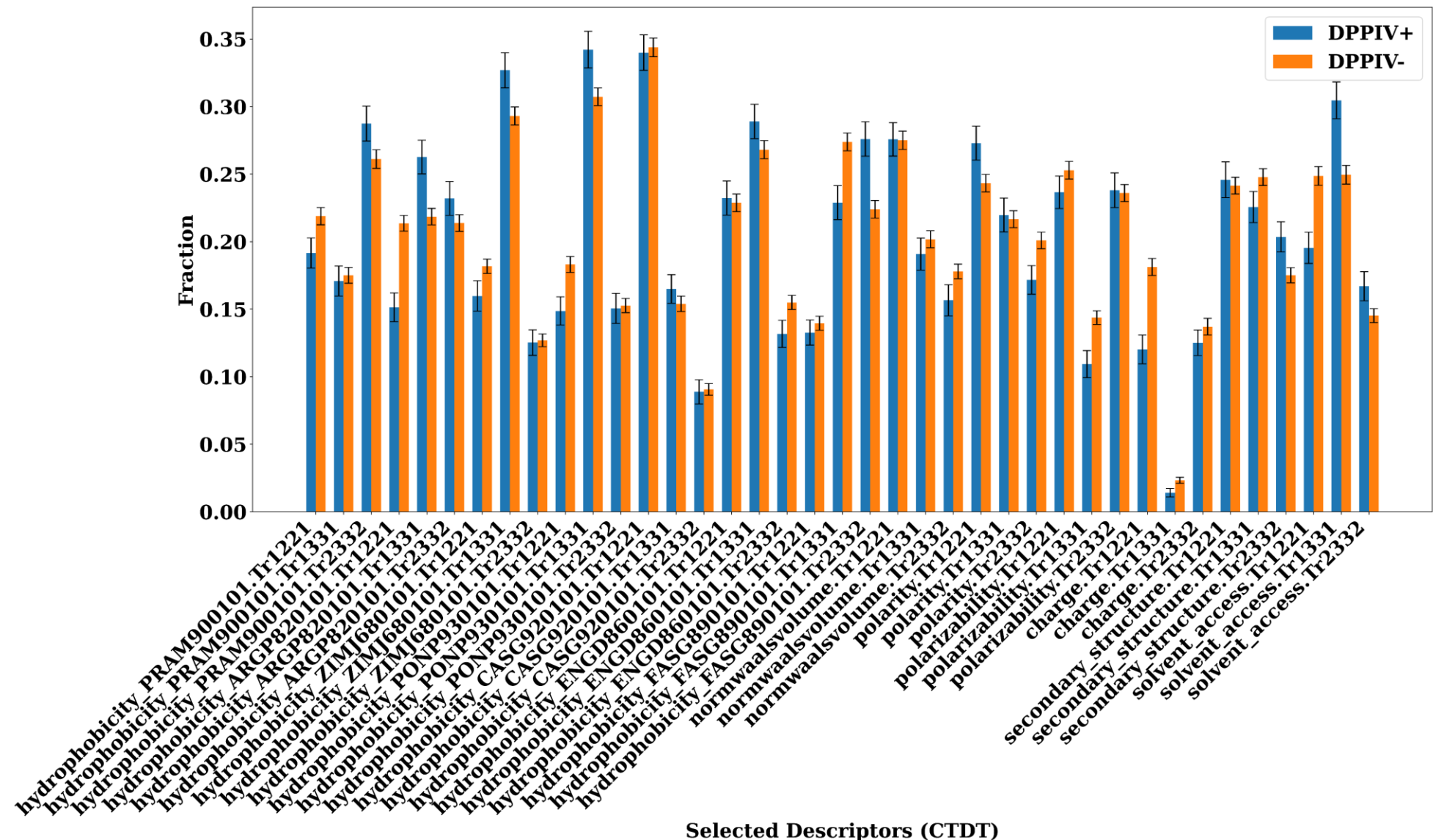




Fig S6: CTD-D descriptor differences between DPPIV+ and DPPIV- classes

