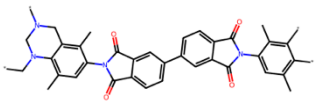
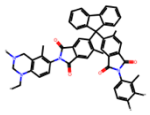
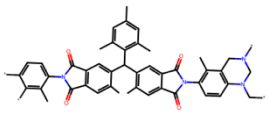
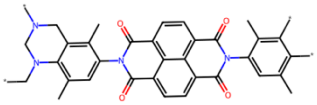
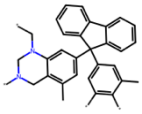
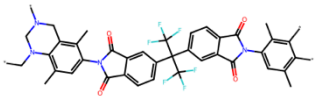
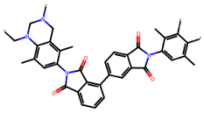
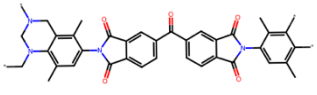
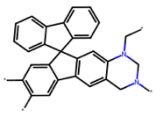
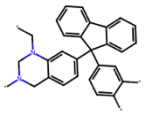
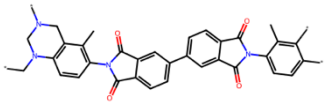
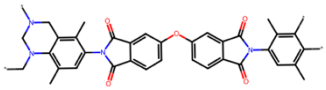
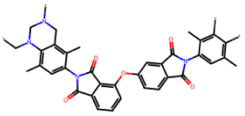
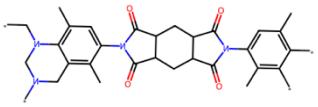
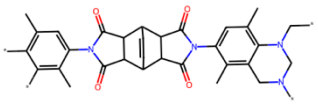
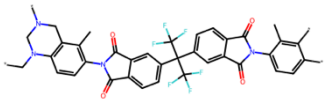
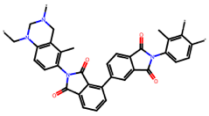
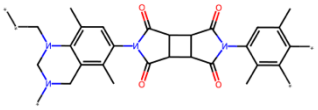
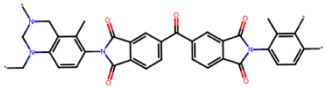
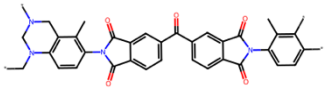
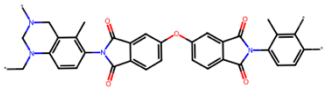
Predicted fractional free volume (FFV) of PIMs

Note: This RF model predicted the FFV of the PIM structure in this work to be approximately 0.40, positioning it within the top 20% among a series of PIMs. Despite this relatively high prediction, it may still be underestimated. Machine learning model is tend to capture general patterns, both latent and non-latent (refer to the SHAP analysis in RF.ipynb). Based on the general patterns, the model is inclined to predict that while the rigid and twisted structure of TB and NDI can increase the free volume, the large coplanar structure and strong interaction between NDI units will lead to local tight packing. However, the use of H-shaped TB-NDI-TB molecular stents suppresses the local tight packing to break through the inherent limitation of the "double-edged sword" effect of aromatic rings. Although the dataset comprises over 8000 polymers, this particular mechanism for increasing free volume is not universal to polymers and therefore may not be recognized by the model.

Index	SMILES	Repeat units	Predicted FFV
1	*CN1CN(*)Cc2c(C)c(N3C(=O)c4ccc(-c5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)cc4C3=O)cc(C)c21		0.40550893
2	*CN1CN(*)Cc2c1ccc(N1C(=O)c3cc(C)c(C4(c5cc6c(cc5C)C(=O)N(c5ccc(*)c(*)c5C)C6=O)c5ccccc5-c5ccccc54)cc3C1=O)c2C		0.40217199
3	*CN1CN(*)Cc2c1ccc(N1C(=O)c3cc(C)c(C(c4cc5c(cc4C)C(=O)N(c4ccc(*)c(*)c4C)C5=O)c4c(C)cc(C)cc4C)cc3C1=O)c2C		0.39995992
4 (This work)	*CN1CN(*)Cc2c(C)c(N3C(=O)c4ccc5c6c(ccc(c46)C3=O)C(=O)N(c3cc(C)c(*)c(*)c3C)C5=O)cc(C)c21		0.39931013
5	*CN1CN(*)Cc2c(C)cc(C3(c4cc(*)c(*)c(C)c4)c4ccccc4-c4ccccc43)cc21		0.39914182
6	*CN1CN(*)Cc2c(C)c(N3C(=O)c4ccc(C(c5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)(C(F)(F)F)C(F)(F)F)cc4C3=O)cc(C)c21		0.39881561
7	*CN1CN(*)Cc2c(C)c(N3C(=O)c4cccc(-c5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)c4C3=O)cc(C)c21		0.39667466
8	*CN1CN(*)Cc2c(C)c(N3C(=O)c4ccc(C(=O)c5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)cc4C3=O)cc(C)c21		0.39617291
9	*CN1CN(*)Cc2cc3c(cc21)C1(c2ccccc2-c2ccccc21)c1cc(*)c(*)cc1-3		0.39494237

10	*CN1CN(*)Cc2ccc(C3(c4ccc(*)c(*)c4)c4ccccc4-c4ccccc43)cc21		0.39436355
11	*CN1CN(*)Cc2c1ccc(N1C(=O)c3ccc(-c4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)cc3C1=O)c2C		0.39416606
12	*CN1CN(*)Cc2c(C)c(N3C(=O)c4ccc(Oc5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)cc4C3=O)cc(C)c21		0.39357151
13	*CN1CN(*)Cc2c(C)c(N3C(=O)c4cccc(Oc5ccc6c(c5)C(=O)N(c5cc(C)c(*)c(*)c5C)C6=O)c4C3=O)cc(C)c21		0.39126067
14	*CN1CN(*)Cc2c(C)c(N3C(=O)C4CC5C(=O)N(c6cc(C)c(*)c(*)c6C)C(=O)C5CC4C3=O)cc(C)c21		0.39091107
15	*CN1CN(*)Cc2c(C)c(N3C(=O)C4C5C(=O)N(c6cc(C)c(*)c(*)c6C)C(=O)C5C4C3=O)cc(C)c21		0.39051714
16	*CN1CN(*)Cc2c1ccc(N1C(=O)c3ccc(C(c4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)(C(F)(F)F)C(F)(F)F)cc3C1=O)c2C		0.3900206
17	*CN1CN(*)Cc2c1ccc(N1C(=O)c3cccc(-c4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)c3C1=O)c2C		0.38929529
18	*CN1CN(*)Cc2c(C)c(N3C(=O)C4C5C=CC(C6C(=O)N(c7cc(C)c(*)c(*)c7C)C(=O)C56)C4C3=O)cc(C)c21		0.38691941
19	*CN1CN(*)Cc2c1ccc(N1C(=O)c3ccc(C(=O)c4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)cc3C1=O)c2C		0.38526852
20	*CN1CN(*)Cc2c1ccc(N1C(=O)c3cccc(Oc4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)c3C1=O)c2C		0.38371944
21	*CN1CN(*)Cc2c1ccc(N1C(=O)c3ccc(Oc4ccc5c(c4)C(=O)N(c4ccc(*)c(*)c4C)C5=O)cc3C1=O)c2C		0.38238191