

No.	Dataset	Initial Size	Remarks	Final set used in the study	
				Proteins	Sites
i	PDBBind	1091	One representative ligand of each type is considered in each protein; filtered to remove sites for small and covalently bound ligands; retained proteins common with the SCOP database	786	893
ii	PDBBind	1091	Filtered to remove sites for small , covalently bound ligand and all suggested by Jackson and co-workers (10) to contribute to noise; considered sites for all ligand types for the remaining	456	1146
iii	Curated dataset of CIT,MTX,MK1 and PGA	27	Ligands for varying sizes and types from PDBBind with multiple sites for a ligand in different proteins	27	51
iv	Tetramer	3768	Multiple sites for each ligand in the same protein; filtered to remove small and covalently bound ligands	1525	11301

Table 1. Different datasets used in the study