SUPPORTING INFORMATION.

Computational Analysis of Crystallization Additives for the Identification of New Allosteric Sites

Jade Fogha[1], Julien Diharce[1], Alan Obled[1], Samia Aci-Sèche[1], Pascal Bonnet[1*].

[1] Institut de Chimie Organique et Analytique, Université d'Orléans, UMR CNRS 7311, BP6759, 45067 Orléans, Cedex 2, France.

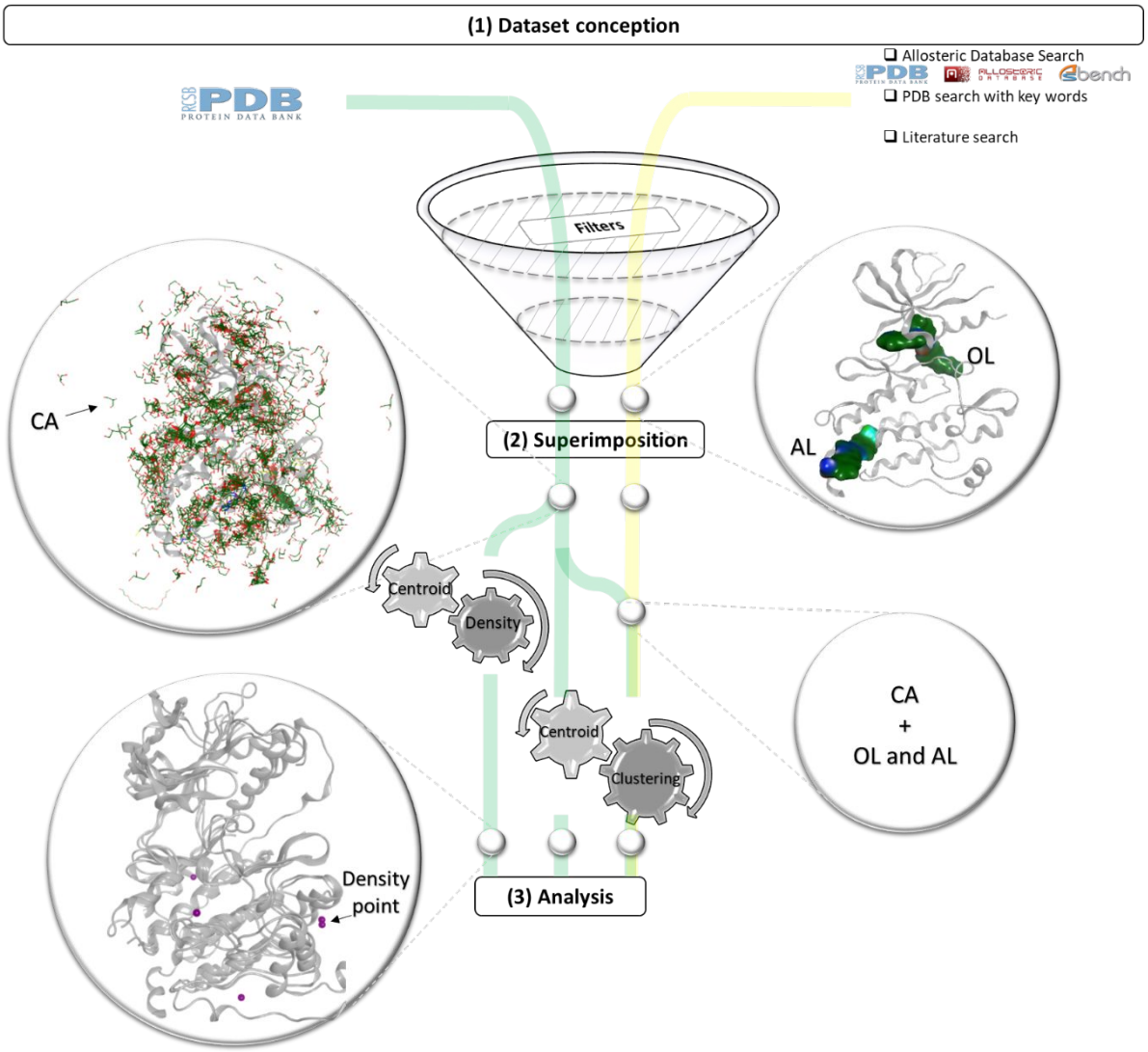
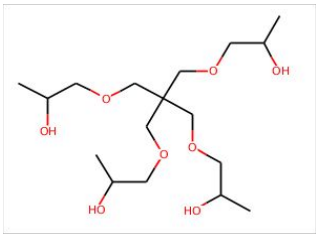
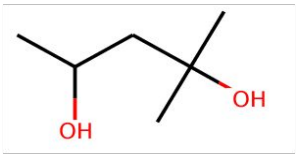
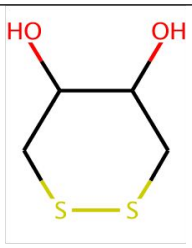
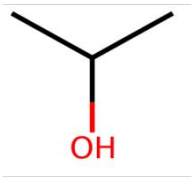
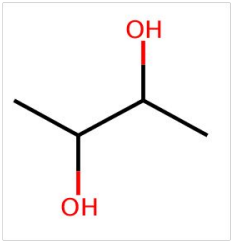
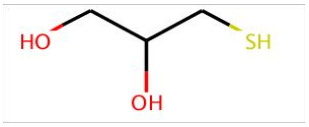
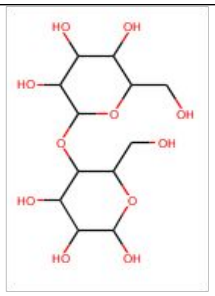
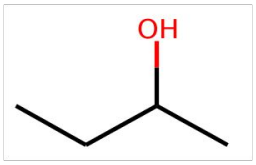
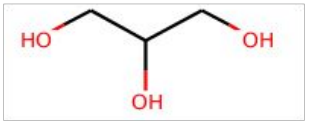
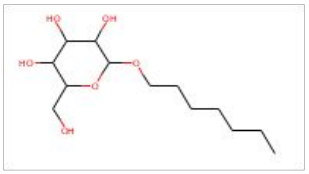
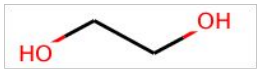
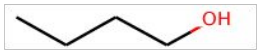
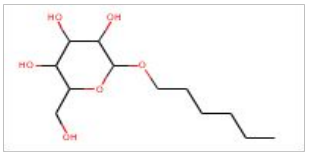
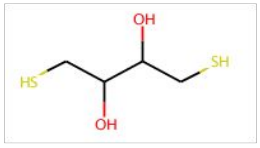
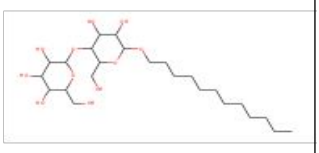
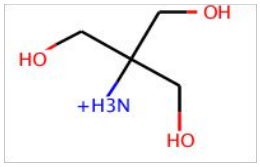
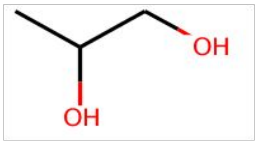
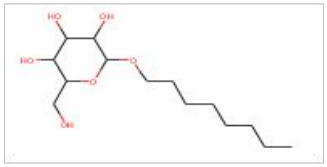


Figure S1. General workflow used in the study. For a given protein family (protein kinases or nuclear receptors), the process is divided in three main steps (1) conception of the two datasets (CA in green line and AL + OL in yellow line), (2) superimposition of the datasets on a reference protein, (3) clustering and density analysis. CA represent crystallographic additives, OL orthosteric ligands and AL allosteric ligands**.**

Table S1. List of crystallographic additives retrieved in the crystallographic structures of protein kinases and nuclear receptors and considered in the CA dataset.

(2s)-1-[3-{[(2r)-2-hydroxypropyl]oxy}-2,2-bis({[(2r)-2-hydroxypropyl]oxy}methyl)propoxy]propan-2-ol		**beta-mercaptoethanol**	
(4s)-2-methyl-2,4-pentanediol		**hexaethylene glycol**	
(4s,5s)-1,2-dithiane-4,5-diol		**isopropyl alcohol**	
(r,r)-2,3-butanediol		**monothioglycerol**	
1,3-propandiol		**triethylene glycol**	

1-ethoxy-2-(2-ethoxyethoxy)ethane		**maltose**	
2-butanol		**glycerol**	
(2s)-1-[3-{[(2r)-2-hydroxypropyl]oxy}-2,2-bis({[(2r)-2-hydroxypropyl]oxy}methyl)propoxy]propan-2-ol		**heptyl-beta-d-glucopyranoside**	
1,2-ethanediol		**hexane-1,6-diol**	
1-butanol		**hexyl beta-d-glucopyranoside**	
2,3-dihydroxy-1,4-dithiobutane		**dodecyl-alpha-d-maltoside**	

2-{2-[2-(2-{2-[2-(2-ethoxy-ethoxy)-ethoxy]-ethoxy}-ethoxy)-ethoxy]-ethoxy}-ethanol		**pentaethylene glycol**	
2-amino-2-hydroxymethyl-propane-1,3-diol		**s-1,2-propanediol**	
b-octylglucoside			

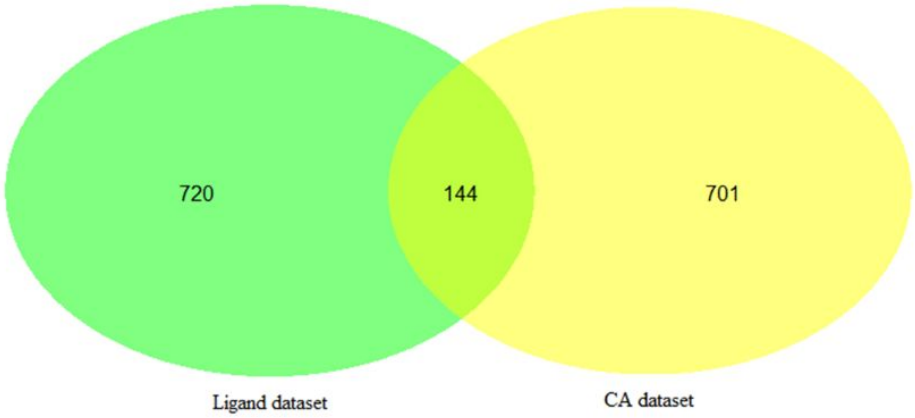


Figure S2. Datasets representation by Venn diagram in terms of PDB files of protein kinases.

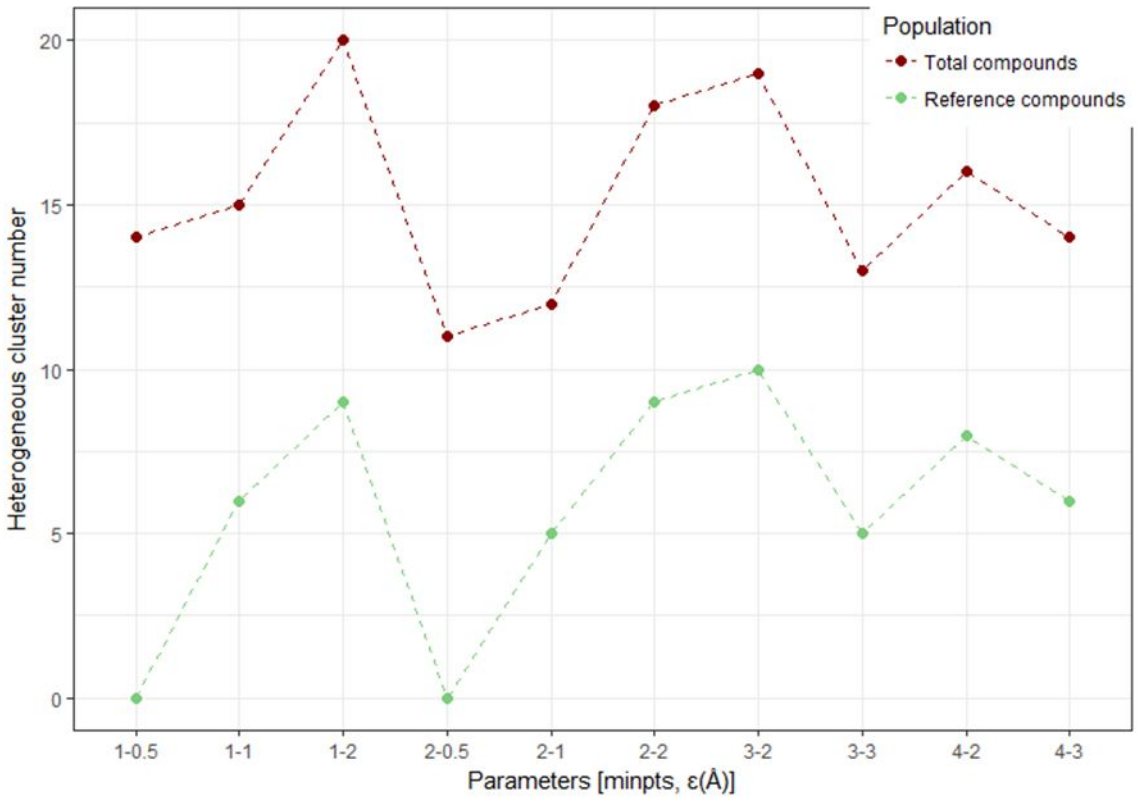


Figure S3. Cluster population evaluation in function of the different couples of parameters formed by the distance between two points in a cluster (ε in Å) and the minimal number of points to form a cluster (minpts) for protein kinases.

Table S2. Position of the 18 reference allosteric ligands of protein kinases after the clustering procedure with different couples of parameters. Red color represents homogeneous clusters or singletons (associated with "S") having a wrong prediction. Green color represented heterogeneous clusters (crystallization additives and ligands) having a good prediction. Yellow correspond to heterogeneous clusters related to orthosteric site having a wrong prediction.

	DEF pocket		Back Pocket							Myristate pocket			PIF pocket		Substrat pocket		P-loop pocket	
[minpts - ε(Å)]	4E6C (p38α)	3O2M (JNK1)	3O96 (AKT1)	3LW0 (IGF1R)	1S9J (MEK)	4LMN (MEK)	3EQC (MEK)	4ITH (RIPK1)	4ZJI (PAK1)	3MS9 (ABL)	3K5V (ABL)	3PYY (ABL)	3PXF (CDK2)	3HRF (PDK1)	3JVR (CHK1)	3F9N (CHK1)	3H30 (CK2a1)	4CFE (AMPKα1/2)
2 - 0.5	S	S	S	S	12	145	12	S	S	S	S	S	93	45	S	S	S	S
1 - 2	32	115	0	0	0	0	0	0	0	259	18	259	77	10	261	1	6	96
1 - 1	28	S	S	0	0	0	0	24	S	S	49	S	117	21	323	263	S	93
2 - 2	32	115	0	0	0	0	0	0	0	S	18	S	77	10	S	1	6	95
2 - 1	28	S	S	S	0	0	0	24	S	S	49	S	117	21	S	S	S	93
3 - 2	34	117	18	S	0	0	0	0	0	S	19	S	78	9	S	1	6	93
1 - 0.5	S	S	S	S	12	145	12	S	S	S	S	S	93	45	298	S	S	S
3 - 3	9	16	0	0	0	0	0	0	0	6	6	6	4	4	0	0	0	0
4 - 3	11	51	0	0	0	0	0	0	0	7	7	7	6	6	S	0	5	5
4 - 2	34	S	69	S	0	0	0	0	0	S	18	S	77	7	S	1	8	0
1 - 2 [a]	35	118	0	0	0	0	0	0	0	263	19	263	80	10	265	1	6	99
3 - 2 [a]	37	120	19	S	0	0	0	0	0	S	20	S	82	9	S	1	6	96

[a] clustering performed without orthosteric ligand extracted using a pharmacophore search.

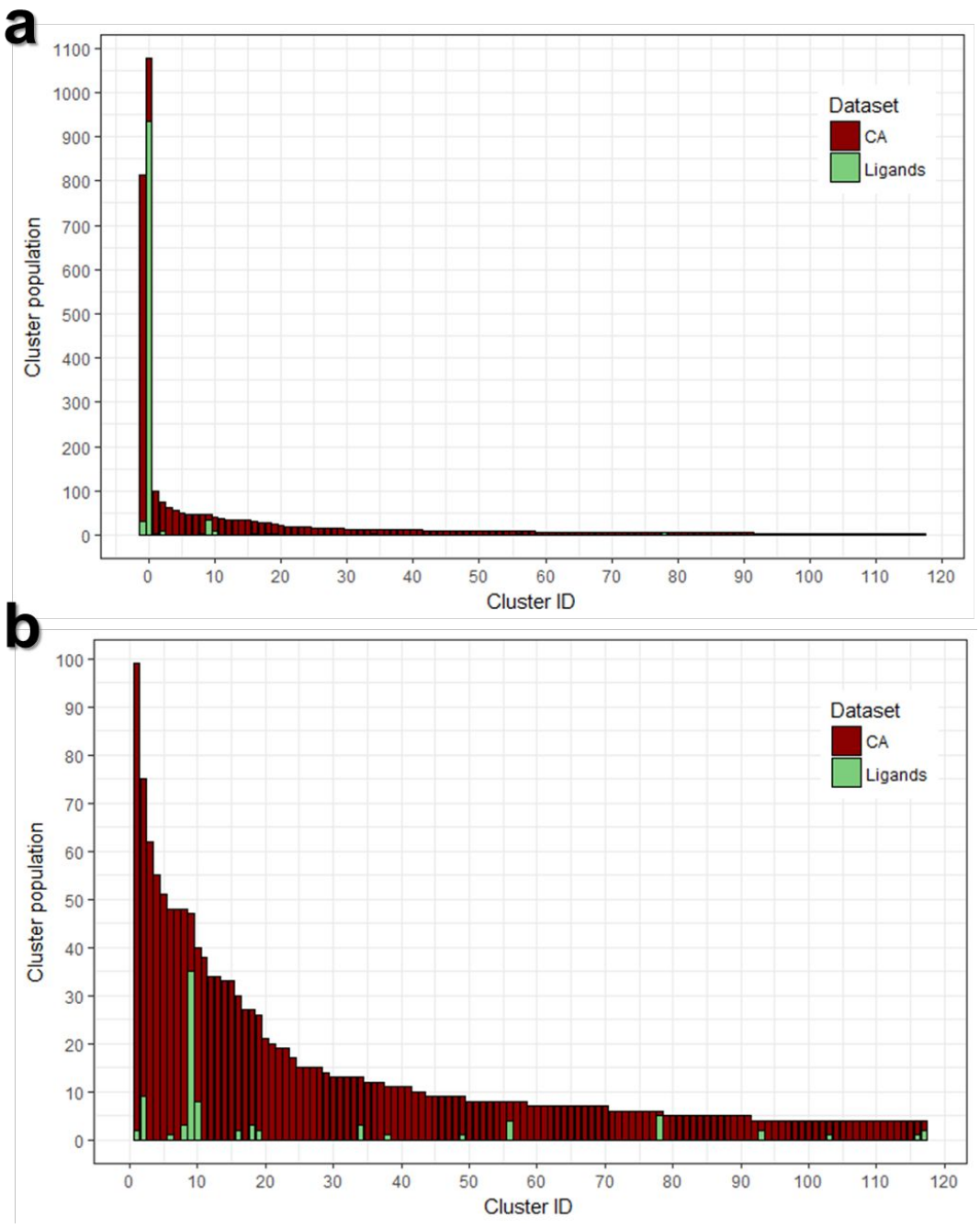


Figure S4. Clustering results with the couple of parameters [3 – 2] for protein kinases, ranked by their population. The singletons (cluster ID -1) and the different clusters are represented in (a), the singletons and cluster 0 (orthosteric identifier) are removed in (b). CA are crystallization additives.