

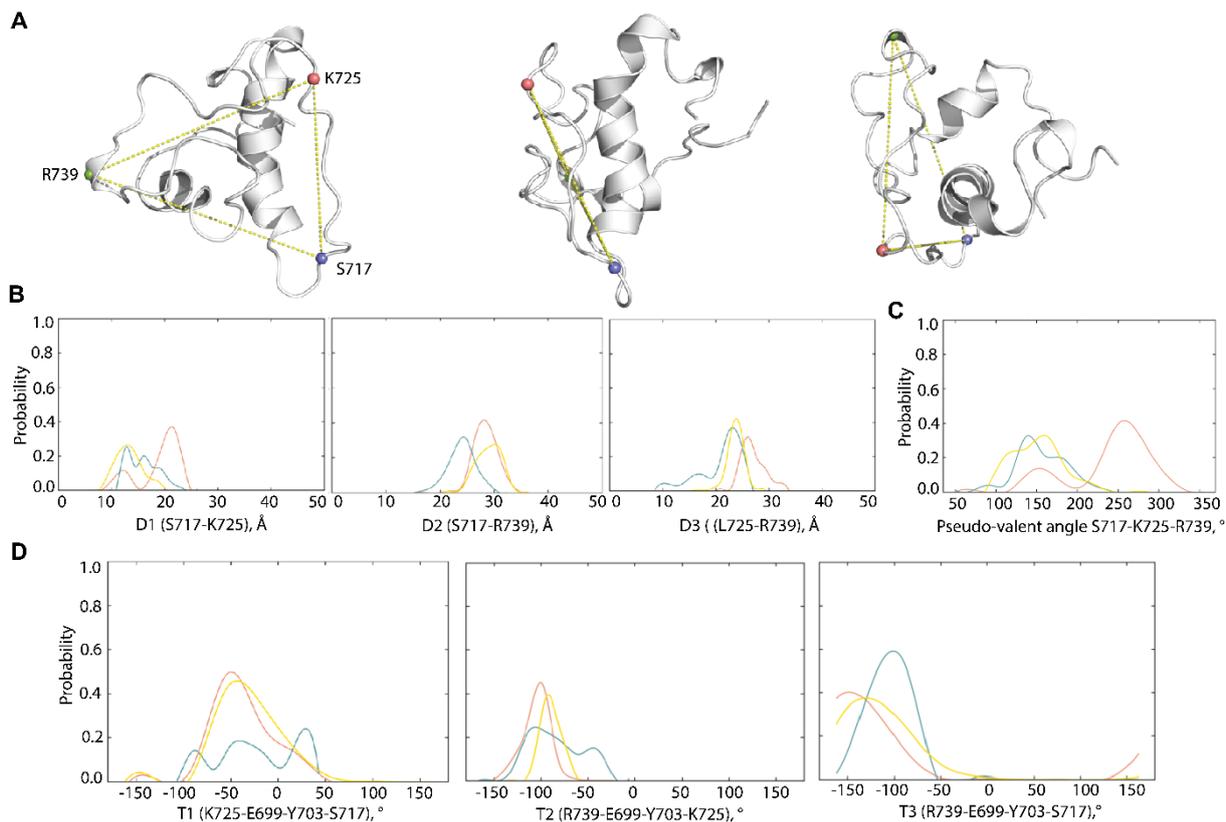
## **Supplementary Materials**

### **Does generic cyclic Kinase Insert Domain of Receptor Tyrosine Kinase KIT clone its native homologue?**

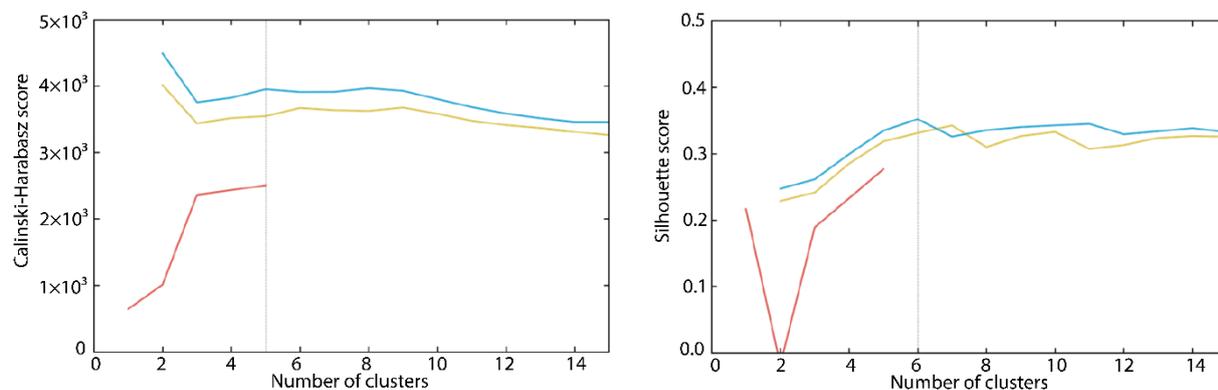
Julie Ledoux<sup>1</sup> and Luba Tchertanov<sup>1\*</sup>

#### **Content:**

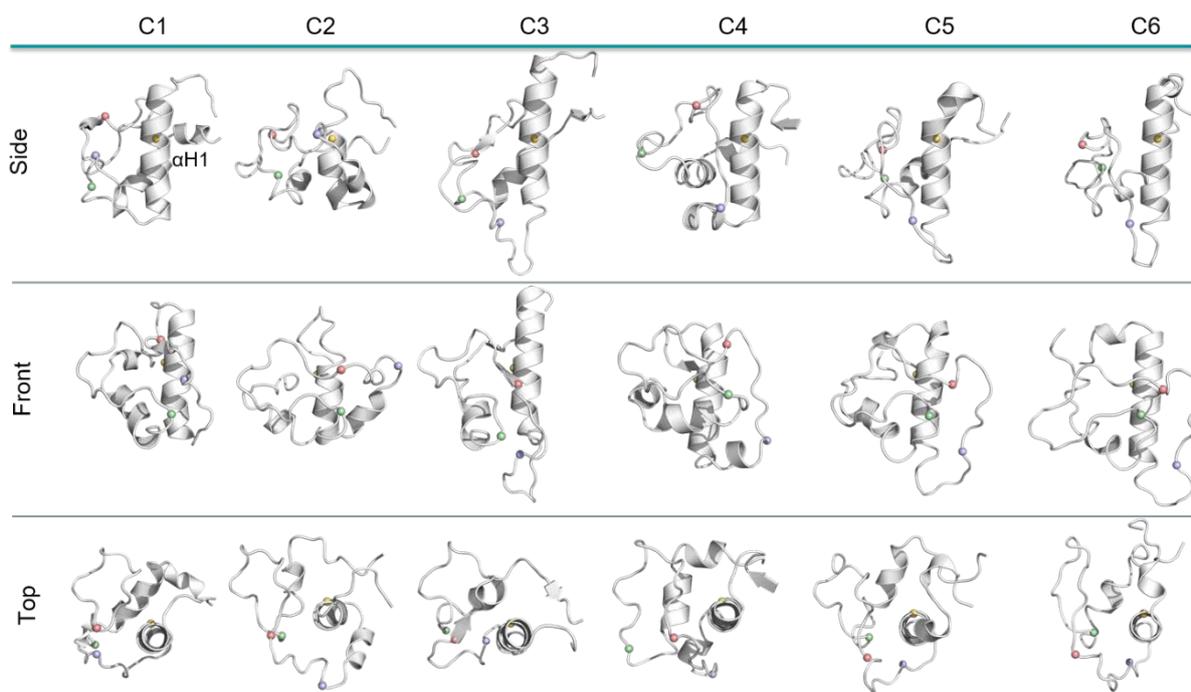
3 Figures: Figure S1; Figure S2; Figure S3



**Figure S1.** Geometry metrics of the most fluctuating residues. (A) 3D structure. Protein is shown as cartoon; the most fluctuating residues are displayed as balls. (B) The distance values probability. (C) The pseudo-valent angle probability. (D) Pseudo-torsion (dihedral) angle probability. (B-C) Metrics were calculated for each KID entity: KID<sup>CC</sup> in yellow, KIDD in red and KIDC in teal.



**Figure S2.** Performance scores were obtained during hyperparameters tweaking to find the better clustering method. Number of clusters obtained by DBSCAN [47] (in red), K-means [48] (in blue) and Ward's [49] methods (in yellow). Scores were computed with Calinski-Harabasz [51] (left) and Silhouette [50] (right).



**Figure S3.** Representative conformations from each cluster (C1 – C6). The calculation was performed on the concatenated trajectories after fitting all data on  $\alpha$ H1-helix taken from KID conformation at  $t=0$ . Three views – side, front and top – concerning  $\alpha$ H1-helix are shown. Protein is displayed as cartoon, and the position of each tyrosine residue is shown by ball: Y703 in yellow, Y721 in lilac, Y730 in coral and Y747 in green.

3 Tables: Table S1; Table S2

**Table S1.** Non-overlapping metrics (features) value intervals in each most populated KID cluster show their differences in geometric properties.

Cluster		Metric (Feature)	Value range	
C <sub>N1</sub>	C <sub>N2</sub>		C <sub>N1</sub>	C <sub>N2</sub>
C1	C2	Distance S717-K725	$17.1 \pm 2.7 \text{ \AA}$	$12.4 \pm 2.0 \text{ \AA}$
		Distance Y721-Y747	$9.6 \pm 2.0 \text{ \AA}$	$14.4 \pm 1.7 \text{ \AA}$
	C3	Radius of gyration	$12.4 \pm 0.4 \text{ \AA}$	$13.7 \pm 0.2 \text{ \AA}$
		$\psi$ Y721	$-30.7 \pm 34.9^\circ$	$73.9 \pm 30.9^\circ$
		$\psi$ Y730	$-18.0 \pm 24.1^\circ$	$140.8 \pm 29.3^\circ$
	C5	Distance S717-K725	$17.1 \pm 2.7 \text{ \AA}$	$21.5 \pm 1.7 \text{ \AA}$
Distance Y730-Y747		$13.1 \pm 3.9 \text{ \AA}$	$6.5 \pm 0.8 \text{ \AA}$	

		$\psi$ Y721	$-30.7 \pm 34.9^\circ$	$152.0 \pm 22.0^\circ$
<b>C2</b>	<b>C3</b>	Distance Y703-Y747	$12.0 \pm 1.4 \text{ \AA}$	$16.8 \pm 1.2 \text{ \AA}$
		Distance Y721-Y747	$14.4 \pm 1.7 \text{ \AA}$	$8.7 \pm 1.0 \text{ \AA}$
		Radius of gyration	$12.4 \pm 0.3 \text{ \AA}$	$13.7 \pm 0.2 \text{ \AA}$
		$\psi$ Y721	$-28.2 \pm 18.6^\circ$	$73.9 \pm 30.9^\circ$
		$\psi$ Y730	$-11.6 \pm 35.1^\circ$	$140.8 \pm 29.3^\circ$
	<b>C5</b>	Distance S717-K725	$12.4 \pm 2.0 \text{ \AA}$	$21.5 \pm 1.7 \text{ \AA}$
		Distance Y721-Y747	$14.4 \pm 1.7 \text{ \AA}$	$10.0 \pm 1.1 \text{ \AA}$
$\psi$ Y721		$-28.2 \pm 18.6^\circ$	$152.0 \pm 22.0^\circ$	
<b>C3</b>	<b>C5</b>	Distance S717-K725	$13.0 \pm 2.0 \text{ \AA}$	$21.5 \pm 1.7 \text{ \AA}$
		Distance Y703-Y747	$16.8 \pm 1.2 \text{ \AA}$	$11.3 \pm 0.8 \text{ \AA}$
		Distance Y730-Y747	$9.0 \pm 1.6 \text{ \AA}$	$6.5 \pm 0.8 \text{ \AA}$
		Radius of gyration	$13.7 \pm 0.2 \text{ \AA}$	$12.6 \pm 0.3 \text{ \AA}$
		$\psi$ Y721	$73.9 \pm 30.9^\circ$	$152.0 \pm 22.0^\circ$
		$\psi$ Y730	$140.8 \pm 29.3^\circ$	$-6.9 \pm 33.6^\circ$

**Table S2.** Contingency table of agreement (%) on cluster population between K-means run with  $k = 5$  and  $k = 6$ . The best population sizes are in bold.

		<b>k = 6</b>						
		C1	C2	C3	C4	C5	C6	<b>Total</b>
<b>k = 5</b>	C1	0.0	0.0	0.0	<b>0.2</b>	<b>21.2</b>	0.0	<b>21.4</b>
	C2	0.0	0.0	<b>22.5</b>	<b>11.9</b>	0.1	0.0	<b>34.5</b>
	C3	0.0	0.0	0.0	0.0	0.0	<b>7.1</b>	<b>7.1</b>
	C4	0.0	<b>22.9</b>	0.0	<b>0.2</b>	0.0	0.0	<b>23.1</b>
	C5	<b>13.8</b>	0.0	0.0	0.0	0.0	0.0	<b>13.8</b>
	<b>Total</b>	<b>13.8</b>	<b>22.9</b>	<b>22.5</b>	<b>12.3</b>	<b>21.3</b>	<b>7.1</b>	<b>100</b>