## **Supplementary Materials**

### Structure-Based Discovery and Bioactivity Evaluation of Novel Aurora-A Kinase Inhibitors as Anti-Cancer Agents via Docking-based Comparative Intermolecular Contacts Analysis (dbCICA)

Majd S. Hijjawi<sup>1</sup>, Reem Fawaz Abutayeh<sup>2</sup> and Mutasem O. Taha<sup>3,\*</sup>

<sup>1</sup> Department of Pharmacology, Faculty of Medicine, The University of Jordan, Amman, Jordan; <u>mjd8120578@fgs.ju.edu.jo</u>; (M.H.) <sup>2</sup> Department of Pharmaceutical Chemistry and Pharmacognosy, Faculty of Pharmacy, Applied Science Private University, Amman, Jordan; r\_abutayeh@asu.edu.jo; (R.A.)

<sup>3</sup>Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan; <u>mutasem@ju.edu.jo</u>; (M.T.) \*Correspondence: <u>mutasem@ju.edu.jo</u>; Tel.: +962-6-5355000

	Table of Contents
Tables	• Table S 1: The structures of the (79) seventy-nine Aurora-A Kinase inhibitors collected from the (ChEMBL), utilized in modeling.
	• Table S 2: Refinement list for steric refinement of Hypo(SB-1)
	• Table S 3: The 75 high-ranking hits captured by Refined-Hypo(SB-1) 3D search query-
	obtained from NCI and their anti-Aurora-A kinase inhibition% at 10 $\mu M$ using Z'-LYTE
	kinase assay, sum of their critical contacts and their predicted activity.
Sections	<ul> <li>Section S 1: LibDock Docking Engine and Scoring Settings</li> </ul>
	<ul> <li>Section S 2: Scoring of Docked Ligand Poses.</li> </ul>
	<ul> <li>Section S 3: Genetic Algorithm Implementation in dbCICA Modeling</li> </ul>
	<ul> <li>Section S 4: Receiver-operating characteristic (ROC) curve analysis</li> </ul>
	Section S 5: Steric Refinement of pharmacophores
Figures	• Figure S 1: Three-dimensional plot showing three main principal components calculated for the Testing Set of for HIPHOP-REFINE modelling
	• Figure S 2: The chemical structures of the tested highest-ranking hits
	• Figure S 3- S 10: "NMR" & Mass Spectrum for Hit 88 (NCI 1576)
	• Figure S 11- S 18: "NMR" & Mass Spectrum for Hit 130 (NCI 12849)
	• Figure S 19- S 25: "NMR" & Mass Spectrum for Hit 85 (NCI 14040)
	• Figure S 26- S 32: "NMR" & Mass Spectrum for Hit 112 (NCI 12415)
	• Figure S 33- S 38: "NMR" & Mass Spectrum for Hit 141(NCI 12492)

# • **Table S1:** The structures of the (79) seventy-nine Aurora-A Kinase inhibitors collected from (ChEMBL), utilized in modeling.

List for the structures of (79) Aurora-A Kinase inhibitors utilized in modeling and their reported Ki values (expressed in Nm-nanoMolar) collected from the European Bioinformatics Institute database (ChEMBL)



Compound	R	Ki (nM)	Reference
1	-H	1600	( Mortlock, et al., 2007)
2	2-F	1400	( Mortlock, et al., 2007)
3	3-F	450	( Mortlock, et al., 2007)
4	4-F	2000	( Mortlock, et al., 2007)
5	2,3-di-F	410	( Mortlock, et al., 2007)
6	3,5-di-F	230	( Mortlock, et al., 2007)
7	3-C1	450	( Mortlock, et al., 2007)
8	3-CN	2600	( Mortlock, et al., 2007)
9	3-OH	4000	( Mortlock, et al., 2007)
10	3-OMe	1900	( Mortlock, et al., 2007)
11	3-CF <sub>3</sub>	510	( Mortlock, et al., 2007)
		N−NH	



Х

Compound	R	Х	Ki (nM)	Reference
12	2,3-di-F	HONN	55	( Mortlock, et al., 2007)
13	2,3-di-F	HON	220	(Mortlock, et al., 2007)
14	2,3-di-F	HO	280	( Mortlock, et al., 2007)
15	3-F	HO	690	(Mortlock, et al., 2007)

Compound	R	X	Ki (nM)	Reference
16	3-F	OH   N , s <sup>s</sup>	65	( Mortlock, et al., 2007)
17	3-F	HO	190	( Mortlock, et al., 2007)
18	3-F	HO N-E	220	( Mortlock, et al., 2007)
19	2,3-di-F	HON~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	360	( Mortlock, et al., 2007)
20	2,3-di-F	HON	87	( Mortlock, et al., 2007)
21	2,3-di-F	HON	18	( Mortlock, et al., 2007)



Compound	R	X	Ki (nM)	Reference
22	2,3-di-F	HO	160	( Mortlock, et al., 2007)
23	2,3-di-F	HON	350	(Mortlock, et al., 2007)
24	3-F	HONN	1400	(Mortlock, et al., 2007)
25	3-F	HONN	980	(Mortlock, et al., 2007)

Compound	Structure	Ki (nM)	Reference
26	HN HN O HN N S O	0.6	(Pollard and Mortimore, 2009)
27		58	(Pollard and Mortimore, 2009)
28	HN - N $HN - N$ $H$	4	(Pollard and Mortimore, 2009)
Compound	Structure	Ki (nM)	Reference
29		81	( Bebbington, et al., 2009)
		NH	
Compound	× N 1	x ~ Ki (nM)	Reference
30	NH	24	(Bebbington, et al., 2009)
31	NMe	17	(Bebbington, et al., 2009)
32	0	36	(Bebbington, et al., 2009)
33	S	20	(Bebbington, et al., 2009)



Ar	Ki (nM)	Reference
2-Cl-Ph	5	(Bebbington, et al., 2009)
3-Cl-Ph	4	(Bebbington, et al., 2009)
4-Cl-Ph	6	(Bebbington, et al., 2009)
2,3-Di-Cl-Ph	3	(Bebbington, et al., 2009)
2,4-Di-Cl-Ph	2	(Bebbington, et al., 2009)
2,6-Di-Cl-Ph	5	(Bebbington, et al., 2009)
3,4-Di-Cl-Ph	2	(Bebbington, et al., 2009)
2-OMe-Ph	24	(Bebbington, et al., 2009)
4-OMe-Ph	9	(Bebbington, et al., 2009)
3,4-Di-OMe-Ph	17	(Bebbington, et al., 2009)
2-Naphthyl	1	(Bebbington, et al., 2009)
	Ar           2-Cl-Ph           3-Cl-Ph           4-Cl-Ph           2,3-Di-Cl-Ph           2,4-Di-Cl-Ph           2,6-Di-Cl-Ph           3,4-Di-Cl-Ph           2-OMe-Ph           4-OMe-Ph           3,4-Di-OMe-Ph           2-Naphthyl	Ar         Ki (nM)           2-Cl-Ph         5           3-Cl-Ph         4           4-Cl-Ph         6           2,3-Di-Cl-Ph         3           2,4-Di-Cl-Ph         2           2,6-Di-Cl-Ph         5           3,4-Di-Cl-Ph         2           2-OMe-Ph         24           4-OMe-Ph         9           3,4-Di-OMe-Ph         17           2-Naphthyl         1



Compound	<b>R</b> 1	Ar	Ki (nM)	Reference
45	Me	4-(NHSO <sub>2</sub> Me)-Ph	2	(Bebbington, et al., 2009)
46	Me	4-(NHC(O)OtBu)-Ph	9	(Bebbington, et al., 2009)
47	Me	4-(NHC(O)Et)-Ph	1	(Bebbington, et al., 2009)
48	Н	4-(NHC(O)Me)-Ph	24	(Bebbington, et al., 2009)



Compound	R1	R2	Ki (nM)	Reference
49	Н	Me	86	(Bebbington, et al., 2009)
50	Me	Me	18	(Bebbington, et al., 2009)
51	Ph	Me	2.4	(Bebbington, et al., 2009)
52	Me	Et	5.1	(Bebbington, et al., 2009)
53	CyPr	Et	3.9	(Bebbington, et al., 2009)
54	tBu	Et	10	(Bebbington, et al., 2009)
55	3-Py	Et	3.9	(Bebbington, et al., 2009)
56	N-Z	Et	2.4	(Bebbington, et al., 2009)
57	₹-N	Et	1.6	(Bebbington, et al., 2009)
58	ξ−NO	Et	1.7	(Bebbington, et al., 2009)
59	ξ−NNH	Et	3.7	(Bebbington, et al., 2009)
60	ξ−N_N	Et	1.6	(Bebbington, et al., 2009)
61	ξ− <b>N</b>	Et	1.3	(Bebbington, et al., 2009)
Compound	Stru	ıcture	Ki (nM)	Reference





490

(Adams, et al., 2010)

Compound		Structure		Ki (nN	<b>A</b> )	Reference
63		N-NH N-NH H N		2.5		(Coumar, et al., 2008)
		N	$ \begin{array}{c}                                     $			
Compound	R1	R2	R3	<b>R4</b>	Ki (nM)	Reference

Compound	KI	<b>R</b> 2	K3	K4	<b>KI (NM)</b>	Reference
64	NH <sub>2</sub>	Н	$NO_2$	Н	73.0	Wang, et al., 2010
65	Me	NMe <sub>2</sub>	Н	Н	6.9	Wang, et al., 2010
66	Me	NHMe	Н	Н	18.0	Wang, et al., 2010
67	Me	$\mathbf{NH}_2$	Н	Н	31.0	Wang, et al., 2010
68	Me	OMe	OMe	OMe	0.4	Wang, et al., 2010
69	NHMe	OMe	OMe	OMe	1.0	Wang, et al., 2010
70	NHEt	OMe	OMe	OMe	3.0	Wang, et al., 2010
71	NHEt	CH <sub>2</sub> NHAc	Н	Н	5.0	Wang, et al., 2010
72	Me	morpholin-1-yl	Н	Н	4.0	Wang, et al., 2010
73	$NH_2$	morpholin-1-yl	Н	Н	8.0	Wang, et al., 2010
74	NHMe	morpholin-1-yl	Н	Н	19.0	Wang, et al., 2010
75	NHEt	morpholin-1-yl	Н	Н	14.0	Wang, et al., 2010
76	NH <sub>2</sub>	4-acetylpiperazin-1-yl	Н	Н	7.6	Wang, et al., 2010
77	NHMe	4-acetylpiperazin-1-yl	Н	Н	12.0	Wang, et al., 2010
78	Me	4-acetylpiperazin-1-yl	Н	Н	8.2	Wang, et al., 2010
79	Me	4-(methylsulfonyl) piperazin-1-yl	Н	Н	9.2	Wang, et al., 2010

#### • Section S 1: LibDock Docking Engine and Scoring Settings

LibDock enables rapid screening of combinatorial libraries where conformations of the ligands are aligned to polar and apolar receptor interactions sites (i.e., hotspots). Conformations can be either pre-calculated or generated on the fly. Since some of the output poses may have hydrogen atoms in close proximity to the receptor, a CHARMm minimization step can be optionally enabled to further optimize the docked poses.

LibDock docking follows the following steps: (1) Remove hydrogen atoms, (2) Rank ligand conformations and prune by solvent accessible surface area (SASA), (3) find hotspots using a grid that is placed into the binding site and using polar and non-polar probes. The numbers of hotspots are pruned by clustering to a user defined value, (4) Dock ligand poses by aligning to binding site hotspots. This is performed by using triplets (i.e., three ligand atoms are aligned to three receptor hotspots). (5) Poses which result in protein clashes are removed. (6) A final BFGS pose optimization stage is performed using a simple pair-wise score (similar to Piecewise Linear Potential). The top scoring ligand poses are retained (7) Hydrogen atoms are added back to the docked ligands. (8) Optionally, CHARMm minimization can be carried out to reduce steric clashes caused by added hydrogen atoms.

The following LibDock parameters were implemented in the presented project: (1) Prior to docking, the DiscoveryStudio 2.5.5 module CAT-CONFIRM was used to generate a maximum of 255 conformers (not exceeding an energy threshold of 20 kcal/mol from the most stable conformer) for each ligand employing "CAESAR" conformation generation option. (2) A binding site sphere of 12.2 Å radius surrounding the centre of the co-crystallized ligand was used to define the binding site. (3) The maximum number of binding site hotspots (polar and apolar) was set to 100. (4) The ligand-to-hotspots matching RMSD tolerance value was set to 0.25Å. (5) The maximum number of poses saved for each ligand during hotspots matching before final pose

minimization = 100. (6) Maximum number of poses to be saved for each ligand in the binding pocket = 100. (7) Minimum LibDock score (poses below this score are not reported) = 100. (8) Maximum number of rigid body minimization steps during the final pose optimization (using BFGS method) = 50. (9) Maximum number of steric clashes allowed before the pose-hotspot alignment is terminated (specified as a fraction of the heavy atom count) = 0.1.

(10) Maximum value for nonpolar solvent accessible surface area for a particular pose to be reported as successful =  $15.0\text{\AA}^2$ . (11) Maximum value for polar solvent accessible solvent area for a particular pose to be reported as successful =  $5.0\text{\AA}^2$ . (12) No final ligand minimization was implemented (i.e., in the binding pocket).

#### • Section S 2: Scoring of Docked Ligand Poses.

Highest ranking docked conformers/poses generated by LibDock were scored using 7 scoring functions: Jain (Jain, 1996), LigScore1, LigScore2 (Venkatachalam, et al., 2003), PLP1, PLP2 (Gehlhaar, et al., 1999), PMF and PMF04 (Muegge and Martin 1999; Muegge, 2002).

LigScore1 and LigScore2 scores were calculated employing CFF force field (version 1.02) and using grid-based energies with a grid extension of 7.5 Å across the binding site. PMF scores were calculated employing cutoff distances of 12.0 Å for carbon-carbon interactions and other atomic interactions, while PMF04 scores were calculated employing cutoff values of 6.0 and 9.0Å for carbon-carbon interactions and other atomic interactions, respectively.

#### • Section S 3: Genetic Algorithm Implementation in dbCICA Modeling

The GA toolbox within MATLAB (Version R2007a) was adapted by implementing the following four basic components: the creation function, cross-over function, mutation function, and fitness function. The creation function randomly generates a population of chromosomes of a

predefined size (number of summed contacts columns, as mentioned in steps 5 and 6 in the section of Docking-Based Comparative Molecular Contacts Analysis (dbCICA) under Materials and Methods) in which every chromosome encodes for certain possible column summation model. Chromosomes differ from one another by the set of summed columns and their weights. Crossover children are the offspring created by selecting vector entries (i.e., genes) from a pair of individual chromosomes in the first generation and combining them to form two complementary children, while mutation children are those created via applying random changes to corresponding parents, i.e., each single parent chromosome is mutated to give a single child by randomly replacing selected gene in the parent chromosome with another from the chromosome

population. Each chromosome is associated with a fitness value that reflects how good the summation of its encoded genes compares to other chromosomes. The fitness functions in dbCICA can be the correlation coefficient ( $r^2$ ), leave-one-out  $r^2$ , or K-fold  $r^2$ .

In this project (dbCICA of Aurora-A Kinase) we implemented a 5-fold  $r^2$  as fitness criterion. In this procedure, each chromosome is ranked as follows: The training set is divided into two subsets: fit and test subsets. The test subset is randomly selected to represent *ca.* 20% of the training compounds. This procedure is repeated over 5 cycles; accordingly, 5 test subsets with their complementary fit subsets are selected for each chromosome (i.e., column summation model).

The 5 test subsets should cover *ca.* 100% of the training compounds by avoiding selecting the same compound in more than one test subset. The fit sets are then utilized to generate 5 submodels employing the same chromosome. The resulting sub-models are then utilized to predict the bioactivities of the corresponding testing subsets. Finally, the predicted values of all 5 test subsets are correlated with their experimental counterparts to determine corresponding 5-fold  $r^2$ . The best column-summation model is selected as representative db-CICA model. The fitness function in the current db-CICA modelling project was 5-fold  $r^2$ .

The following parameters were implemented for GA search of best models: (1) Size of chromosome population = 100, (2) Rate of mating (crossover fraction): 80% (3) Elite count = 1 (4) Maximum number of generations which is needed to exit from GA iteration cycles and completion of the algorithm = 1000.

Based on these settings, the numbers of each type of children in the offspring generation is as follows: There is 1 elite child (corresponding to the individual in the parents' generation with the best fitness value), and there are 199 individual children other than the elite child. The algorithm rounds 0.8 (crossover fraction) x 199 = 159.2 to 159 to get the number of crossover children and the remaining 40 (i.e., 199-159) are the mutation population. The elite child is passed to the offspring population without alteration.

#### • Section S 4: Receiver-operating characteristic (ROC) curve analysis

The Testing Set: The classification power of the resulting pharmacophores was validated using receiver-operating characteristic (ROC) curve analysis by testing the ability of a particular pharmacophore to selectively capture diverse Aurora-A kinase inhibitors from large list of inactive compounds. The testing set implemented in this project was entirely composed of experimentally validated active and inactive Aurora-A kinase inhibitors extracted from the European Bioinformatics Institute database (ChEMBL, https://www.ebi.ac.uk/chembl). It included 86 experimentally-validated active compounds (anti-Aurora-A Kinase with Ki values  $\leq 10$  nM) & 248 less-active compounds (anti-Aurora-A Kinase with Ki values > 500 nM considered as decoy list). To insure that active testing compounds closely resemble the diversity

of less-active members we computed top three principal components based on 12 physicochemical descriptors (i.e., LogP, molecular weight, hydrogen bond donors and acceptors, rotatable bonds, rings, aromatic rings, fractional polar surface area surface area, polar surface area and number of fragments) for active testing compounds and compared them with corresponding principal components calculated for the inactive testing inhibitors in each case. Below figure shows three-dimensional plot of the principal components representing active and inactive testing compounds testing set.



**Figure S 1:** Three-dimensional plot showing three main principal components calculated for the Testing Set (based on 12 physicochemical descriptors, see text). Amber spheres (•) represent active compounds (Ki  $\leq$  10) while blue spheres (•) represent inactive compounds (Ki  $\geq$  500) as enlisted in ChEMBL database.

Conformational ensembles were generated for the testing set using "CESEAR" conformation generation option implemented in DiscoveryStudio 2.5.5. The results are presented in the form of ROC curves. A ROC curve is plotted by considering the highest score (fit value against the

tested pharmacophore) of an active molecule as the first threshold then counting the number of decoy compounds within this cut-off value.

The corresponding sensitivity (SE, also known as True Positive Rate) and specificity (SP, also known as True Negative Rate) are calculated using equation 1 and equation 2, respectively, and plotted. This process is repeated using the active molecule possessing the second highest score and so on, until the scores of all active compounds are considered as selection cut-off values (Triballeau, et al., 2006 and Kirchmair, et al., 2008).

Where, TP (true positive) is the number of active compounds that are captured by the pharmacophore under concern, FN (false negative) is the number of active compounds discarded from the hits list by the virtual screening method, TN (true negative) is the number of discarded decoys, while FP (false positive) is the number of captured decoys (presumably inactive) (Irwin and Shoichet, 2005; Triballeau et al., 2006; Kirchmair et al., 2008).

If all molecules scored by a virtual screening (VS) protocol with sufficient discriminatory power are ranked according to their score (i.e., fit values), starting with the best-scored molecule and ending with the molecule that got the lowest score, most of the actives will have a higher score than the decoys. Since some of the actives will be scored lower than decoys, an overlap between the distribution of active molecules and decoys will occur, which will lead to the prediction of false positives and false negatives. (Irwin and Shoichet, 2005; Triballeau et al., 2006; Kirchmair et al., 2008). The selection of one score value as a threshold strongly influences the ratio of actives to decoys and therefore the validation of a VS method. The ROC curve method avoids the selection of a threshold by considering all SE and SP pairs for each score threshold (Irwin and Shoichet, 2005; Triballeau et al., 2006; Kirchmair et al., 2008). A ROC curve is plotted by setting the score of the active molecule as the first threshold. Afterwards, the number of decoys within this cutoff is counted and the corresponding SE and SP pair is calculated. This calculation is repeated for the active molecule with the second highest score and so forth, until the scores of all actives are considered as selection thresholds.

In practice, the ROC curve for a set of actives and inactive decoys with randomly distributed scores tends towards the SE = 1-SP line asymptotically with increasing number of actives and decoys (Triballeau, et al., 2005). The success of particular virtual screening workflow depending on ROC analysis evaluation can be provided as follow:

- 1) Area under the ROC curve (AUC): optimal ROC curve has a value of 1 and random distribution with 0.5. Any virtual screening that performs better than a random discrimination of actives and inactives get an AUC value between 0.5 and 1, whereas an AUC value lower than 0.5 represents the unfavourable case of a virtual screening method that has a higher probability to assign the best scores to decoys than to actives (Irwin and Shoichet, 2005; Triballeau, et al., 2006; Taha, 2012).
- Overall accuracy (ACC): describes the percentage of compounds that were correctly classified by the screening protocol (equation 3). Testing compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) (Triballeau, et al., 2006).

$$ACC = \frac{TP + TN}{N} = \frac{A}{N} \cdot Se + \left(1 - \frac{A}{N}\right) \cdot Sp \dots (3)$$

Where, N is the number of all compounds in the testing database, A is the number of true actives in the testing database.

3) Overall true negative rate (TNR) or overall specificity (SP): that describes the fraction percentage of discarded inactive by the virtual screening tool. Discarded inactive test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values (Jacobsson et al., 2003; Irwin and Shoichet, 2005; Triballeau et al., 2006; Kirchmair et al., 2008, Taha, 2012).

4) Overall true positive rate (TPR) or overall sensitivity (SE): describes the fraction percentage of captured actives from the total number of actives. Active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values (Jacobsson et al., 2003; Irwin and Shoichet, 2005; Triballeau et al., 2006; Kirchmair et al., 2008, Taha, 2012).

#### • <u>Section S5: Steric Refinement of pharmacophores</u>

Based on ROC results **Hypo(SB-1**) model had better behavior over **Hypo(SB-2**), as detailed in results section. In order to improve the classification properties of **Hypo(SB-1**) model, it was complemented with exclusion spheres by employing HIPHOP-REFINE module of DiscoveryStudio 2.5.5 (Khanfar and Taha, 2013). HIPHOP-REFINE identifies spaces occupied by the conformations of inactive compounds and free from conformations of active ones. These areas are filled with exclusion volumes to represent the steric constrains of the binding pocket (Hahn, 1997; Taha, et al., 2011; Khanfar and Taha, 2013). A subset of 32 training compounds

was carefully selected from the molecules in shown below in **Table S 2** for HIPHOP-REFINE modelling or construction of appropriate exclusion regions around **Hypo(SB-1)**.

The Principal and Maximum Omitted Features (MaxOmitFeat) parameters are used to define how many molecules fit the selected pharmacophore hypothesis (partially or completely) for steric refinement purposes. Active compounds are assigned MaxOmitFeat parameter of zero and Principal value of 2 to direct the software to fit all their chemical moieties against all the pharmacophoric features of the particular hypothesis. However, inactive compounds are allowed to miss one or two features by assigning them a MaxOmitFeat of 1 or 2, respectively. Moreover, inactives were assigned Principal value of zero to indicate their inferior bioactivities. However, intermediate active compounds are normally assigned a principal value of 1 and a MaxOmitFeat of zero or 1 in accordance with the number of features the compound loses, to indicate their intermediate status (Taha, et al., 2011).

In this project it was decided to consider the Ki value of 510 nM as an arbitrary activity/inactivity threshold, such that compounds with values equal to or more than 510 nM considered "inactives" with Principal value of zero, and were carefully evaluated to assess whether their lower potencies are attributable to missing one or more pharmacophoric features (MaxOmitFeat = 1 or 2), or only related to steric clashes within the binding pocket (MaxOmitFeat = 0).

However, compounds of Ki values ranging from 5 nM to less than 510 nM were considered moderately active and were assigned a principal value of 1 and MaxOmitFeat of 1 or zero (according to their number of missed feature). Compounds of Ki values less than 5.0 nM were considered active, and were assigned Principal value of 2, and MaxOmitFeat of zero. The conformational spaces of training lists were generated using "BEST" conformation generation option in DiscoveryStudio 2.5.5. The training compounds employed for steric refinement of the

generated pharmacophore **Hypo(SB-1)**, and their corresponding Principal and MaxOmitFeat parameters are sown below in **Table S 2**.

HIPHOP-REFINE was configured to permit a maximum of 100 exclusion spheres to be added to pharmacophoric hypothesis of **Hypo(SB-1)**. The HIPHOP-REFINE process resulted in adding 93 exclusion volumes to **Hypo(SB-1)**, and the sterically refined pharmacophore was named **Refined-Hypo (SB-1)**.

Compound <sup>a</sup>	Ki (nM)	Principal value	MaxOmitFeat <sup>b</sup>
1	1,600	0	2
2	1,400	0	2
4	2,000	0	2
8	2,600	0	2
9	4,000	0	0
10	1,900	0	2
11	510	0	2
15	690	0	2
20	87	1	1
24	1,400	0	2
25	980	0	2
26	0.6	2	0
28	4	2	0
34	5	1	1
39	5	1	1
42	9	1	0
43	17	1	0
45	2	2	0
46	9	1	0
47	1	2	0
49	86	1	0
50	18	1	0
52	5.1	1	0
54	10	1	0
57	1.6	2	0
58	1.7	2	0
60	1.6	2	0
61	1.3	2	0
66	18	1	1
71	5	1	1
78	8.2	1	1
79	9.2	1	1

Table S 2: Refinement list for steric refinement of Hypo(SB-1)

<sup>a</sup>Compounds' numbers are as in **Table S3** <sup>b</sup>MaxOmitFeat: Maximum omitted features.

Table S3: The 75 high-ranking hits and their anti-Aurora-A kinase inhibition% at 10 µM using Z'-LYTE kinase assay, sum of their critical contacts and their predicted activity. These high-ranking hits captured by Refined-Hypo(SB-1) -3D search query- derived from Hypo(SB-1) pharmacophore were docked into (3w2c) using the docking-scoring settings of (SB-1) and their docked poses were analyzed to identify their critical binding contacts (marked by dbCICA model Table 2.) were used to predict their K<sub>i</sub> values by substituting the sum of binding contacts in the respective dbCICA-regression equations (Table 1.). Steps of activity prediction was also employed using docking/scoring settings of (SB-2) dbCICA model to assess the similarity extent in "predicted activity" of both (SB-2) model and the better performing (SB-1) model

		SB-1			9/	
Hitª	NCI Code	Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Inhibition at 10 μM
80	19024	15	0.27	14	6.96	20
81	34607	15	0.27	14	6.96	3
82	1987	14	0.58	12	18.72	5
83	22650	14	0.58	13	11.41	7
84	4293	14	0.58	13	11.41	17
<u>85*</u>	<u>14040*</u>	<u>13</u>	<u>1.20</u>	<u>13</u>	<u>11.41</u>	<u>56</u>
86	22651	13	1.20	13	11.41	12
87	24666	13	1.20	11	30.71	-4
<u>88</u>	<u>1576</u>	<u>12</u>	<u>2.51</u>	<u>14</u>	<u>6.96</u>	<u>75</u>
89	18100	12	2.51	13	11.41	2
90	23953	12	2.51	13	11.41	-2
91	33654	12	2.51	13	11.41	-2
92	34311	12	2.51	12	18.72	3
93	34595	12	2.51	13	11.41	14
94	35036	12	2.51	10	50.37	-2
95	4356	12	2.51	12	18.72	30
96	4721	12	2.51	11	30.71	0
97	7506	12	2.51	12	18.72	-4
98	10637	11	5.22	9	82.61	-5
99	11196	11	5.22	12	18.72	5

<sup>a</sup>Hits are as in **Figure S 3**. <sup>b</sup>Contacts summations according to corresponding dbCICA model (Tables 1 and 2). <sup>c</sup>Predicted Ki (nM) by substituting the number of contacts of each docked compound in the regression equation of the corresponding dbCICA model. \*Underlined hits with highest Inhibition%

Table S3: The 75 high-ranking hits and their anti-Aurora-A kinase inhibition% at 10 μM using Z'-LYTE kinase assay, sum of their critical contacts and their predicted activity. These high-ranking hits captured by Refined-Hypo(SB-1) -3D search query- derived from Hypo(SB-1) pharmacophore were docked into (3w2c) using the docking-scoring settings of (SB-1) and their docked poses were analyzed to identify their critical binding contacts (marked by dbCICA model Table 2.) were used to predict their K<sub>i</sub> values by substituting the sum of binding contacts in the respective dbCICA-regression equations (Table 1.). Steps of activity prediction was also employed using docking/scoring settings of (SB-2) dbCICA model to assess the similarity extent in "predicted activity" of both (SB-2) model and the better performing (SB-1) model

Hitª	NCI Code	SB-1		SB-2		%	
		Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Contac atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Inhibition at 10 μM	
100	12847	11	5.22	11	30.71	7	
101	17288	11	5.22	13	11.41	6	
102	18099	11	5.22	12	18.72	6	
103	22645	11	5.22	12	18.72	12	
104	23413	11	5.22	12	18.72	23	
105	33564	11	5.22	13	11.41	1	
106	34873	11	5.22	12	18.72	6	
107	4354	11	5.22	12	18.72	25	
108	6919	11	5.22	11	30.71	-7	
109	7501	11	5.22	12	18.72	-142	
110	9293	11	5.22	10	50.37	3	
111	10515	10	10.87	13	11.41	1	
<u>112</u>	<u>12415</u>	<u>10</u>	<u>10.87</u>	<u>14</u>	<u>6.96</u>	<u>55</u>	
113	14341	10	10.87	12	18.72	8	
114	22676	10	10.87	10	50.37	4	
115	22677	10	10.87	11	30.71	-1	
116	23825	10	10.87	13	11.41	21	
117	26690	10	10.87	15	4.24	7	
<sup>a</sup> Hits are as in <b>Figure S 3</b> . <sup>b</sup> Contacts summations according to corresponding dbCICA model (Tables 1 and 2). <sup>c</sup> Predicted Ki (nM) by substituting the number of contacts of each docked compound in the regression equation of the corresponding dbCICA model. *Underlined hits with highest Inhibition%							

Table S3: The 75 high-ranking hits and their anti-Aurora-A kinase inhibition% at 10 μM using Z'-LYTE kinase assay, sum of their critical contacts and their predicted activity. These high-ranking hits captured by Refined-Hypo(SB-1) -3D search query- derived from Hypo(SB-1) pharmacophore were docked into (3w2c) using the docking-scoring settings of (SB-1) and their docked poses were analyzed to identify their critical binding contacts (marked by dbCICA model Table 2.) were used to predict their K<sub>i</sub> values by substituting the sum of binding contacts in the respective dbCICA-regression equations (Table 1.). Steps of activity prediction was also employed using docking/scoring settings of (SB-2) dbCICA model to assess the similarity extent in "predicted activity" of both (SB-2) model and the better performing (SB-1) model

		SB-1		SB-2	%		
Hitª	NCI Code	Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Inhibition at 10 μM	
118	29057	10	10.87	14	6.96	7	
119	31475	10	10.87	13	11.41	-3	
120	31937	10	10.87	10	50.37	-2	
121	3343	10	10.87	11	30.71	-2	
122	33550	10	10.87	13	11.41	-1	
123	34304	10	10.87	13	11.41	4	
124	34692	10	10.87	14	6.96	-2	
125	6807	10	10.87	14	6.96	-1	
126	6848	10	10.87	12	18.72	-2	
127	10868	9	22.62	13	11.41	0	
128	11191	9	22.62	13	11.41	-5	
129	12840	9	22.62	13	11.41	3	
<u>130</u>	<u>12849</u>	<u>9</u>	<u>22.62</u>	<u>13</u>	<u>11.41</u>	<u>86</u>	
131	23575	9	22.62	12	18.72	3	
132	26084	9	22.62	13	11.41	22	
133	28316	9	22.62	12	18.72	3	
134	31011	9	22.62	13	11.41	8	
135	3289	9	22.62	13	11.41	7	
136	5769	9	22.62	13	11.41	5	
137	6888	9	22.62	12	18.72	-30	
<sup>a</sup> Hits are as in <b>Figure S 3</b> . <sup>b</sup> Contacts summations according to corresponding dbCICA model (Tables 1 and 2). <sup>c</sup> Predicted Ki (nM) by substituting the number of contacts of each docked compound in the regression equation of the corresponding dbCICA model. *Underlined hits with highest Inhibition%							

Table S3: The 75 high-ranking hits and their anti-Aurora-A kinase inhibition% at 10 μM using Z'-LYTE kinase assay, sum of their critical contacts and their predicted activity. These high-ranking hits captured by Refined-Hypo(SB-1) -3D search query- derived from Hypo(SB-1) pharmacophore were docked into (3w2c) using the docking-scoring settings of (SB-1) and their docked poses were analyzed to identify their critical binding contacts (marked by dbCICA model Table 2.) were used to predict their K<sub>i</sub> values by substituting the sum of binding contacts in the respective dbCICA-regression equations (Table 1.). Steps of activity prediction was also employed using docking/scoring settings of (SB-2) dbCICA model to assess the similarity extent in "predicted activity" of both (SB-2) model and the better performing (SB-1) model

	NCI Code	SB-1		SB-2		0/2
Hit <sup>a</sup>		Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	Contact atoms Summation <sup>b</sup>	Predicted Ki (nM) <sup>c</sup>	<sup>70</sup> Inhibition at 10 μM
138	6924	8	47.07	13	11.41	-1
139	10188	8	47.07	13	11.41	6
140	11193	8	47.07	12	18.72	-7
141	<u>12492</u>	<u>8</u>	47.07	<u>11</u>	<u>30.71</u>	<u>51</u>
142	12990	8	47.07	13	11.41	-3
<u>143</u>	26037	8	47.07	11	30.71	-72
144	32263	8	47.07	13	11.41	0
145	33974	8	47.07	12	18.72	0
146	34870	8	47.07	13	11.41	4
147	4355	8	47.07	12	18.72	14
148	7959	8	47.07	13	11.41	-3
149	7960	8	47.07	13	11.41	-1
150	8793	8	47.07	13	11.41	-3
151	11155	7	97.94	13	11.41	-7
152	21194	7	97.94	10	50.37	1
153	3021	7	97.94	13	11.41	-6
154	34688	7	97.94	13	11.41	-2
<sup>a</sup> Hits are as in <b>Figure S 3</b> . <sup>b</sup> Contacts summations according to corresponding dbCICA model (Tables 1 and 2). <sup>c</sup> Predicted Ki (nM) by substituting the number of contacts of each docked compound in the regression equation of the corresponding dbCICA model. *Underlined hits with highest Inhibition%						











115 (NCI 22677)



107 (NCI 4354)



110 (NCI 9293)



108 (NCI 6919)



111 (NCI 10515)

Ö

0

 $\mathbf{O}$ 



113 (NCI 14341)



116 (NCI 23825)





117 (NCI 26690)



118 (NCI 29057)



121 (NCI 3343)



124 (NCI 34692)



127 (NCI 10868)



119 (NCI 31475)



122 (NCI 33550)



120 (NCI 31937)



123 (NCI 34304)





125 (NCI 6807)







128 (NCI 11191)

129 (NCI 12840)







144 (NCI 32263)



146 (NCI 34870)



9.



148 (NCI 7959)

147 (NCI 4355)





150 (NCI 8793)



153 (NCI 3021)



154 (NCI 34688)



# "NMR" Charts & Mass Spectrum

### "NMR" Charts & Mass Spectrum For Hit 88 (NCI 1576)

### Figure S 3: <sup>1</sup>H-NMR Charts for hit 88 (NCI 1576)









SSB LB GB PC 0 1.00 Hz 0

1.00








## "NMR" Charts & Mass Spectrum For Hit 130 (NCI 12849)









<u>∩</u> 1.50







Analysis Info Analysis Name Method Sample Name Comment	F:\Data\2017MAY03\AREEJ_000052.d ESI-POS-2017 16(12849) MEOH			Operator Instrument	Bruker_PC apex-IV
Acquisition	Parameter				
		Sec. Sec.			
ntens.		355.23760	)		
2.5					
201					
2.0					
1.5					
1.0				356 24102	
0.5				555.24102	
0.5		4			
0.01,	354.5	355.0	355.5	356.0	356.5 n
ADEE	1 000052 d: +MS				
AREE	J_000052.0. +WS				
	and the second				
# 1	n/z I	36			
2 179.118	314 3951963	2.5			
3 205.133	4377639	2.8			
4 256.133	362 3995775	2.6			
5 269.165	512 10425320	6.7			
6 301.14	3768547	2.4			
7 301.190	96 53069171	34.0			
8 302.194	147 10001374	6.4			
9 313.19	25804915	16.5			
10 314.194	4/19034	42.0			
11 327.200	19990003	2.6			
13 353 254	343 3851968	2.5			
14 353 259	6983936	4.5			1
15 353.265	66 96340992	61.6			/
16 353.272	6727936	4.3			
17 354.269	23 20478344	13.1			$\sim$
18 355.237	29100815	18.6		F V N	
19 356.24	6540407	4.2			민민기
20 381.284	6382676	4.1	0		$\checkmark \checkmark$
21 381.289	11449492	1.3	0		
22 381.296	156310164	6.9			
23 381.304	14 10663060	2.5			
24 381.308	10 35555222	227			
26 383 301	325 3833897	2.5			
27 591 493	351 5448197	3.5			
28 619.525	9776843	6.3			
29 647.556	3885048	2.5			
30 685.433	360 7750803	5.0			
		- 725 TV		F/40/0047 0.40.48 AM	Dego 1 of 1

## "NMR" Charts & Mass Spectrum For Hit 85 (NCI 14040)











1.80



Figure S 24: Mass Spectrum for Hit 85 (NCI 14040)



## "NMR" Charts & Mass Spectrum For Hit 112 (NCI 12415)















## "NMR" Charts & Mass Spectrum For Hit 141 (NCI 12492)







1.20





