J Comput Aided Mol Des (2015) 29:561–581
DOI 10.1007/s10822-015-9848-1

Combining docking-based comparative intermolecular contacts analysis and k-nearest neighbor correlation for the discovery of new check point kinase 1 inhibitors

Nour Jamal Jaradat[1] · Mohammad A. Khanfar[2] · Maha Habash[3] · Mutasem Omar Taha[2]

Received: 26 December 2014 / Accepted: 3 May 2015 / Published online: 9 May 2015

Abstract Check point kinase 1 (Chk1) is an important protein in G2 phase checkpoint arrest required by cancer cells to maintain cell cycle and to prevent cell death. Therefore, Chk1 inhibitors should have potential as anti-cancer therapeutics. Docking-based comparative intermolecular contacts analysis (dbCICA) is a new three-dimensional quantitative structure activity relationship method that depends on the quality and number of contact points between docked ligands and binding pocket amino acid residues. In this presented work we implemented a novel combination of k-nearest neighbor/genetic function algorithm modeling coupled with dbCICA to select critical ligand-Chk1 contacts capable of explaining anti-Chk1 bioactivity among a long list of inhibitors. The finest set of contacts were translated into two valid pharmacophore hypotheses that were used as 3D search queries to screen the National Cancer Institute's structural database for new Chk1 inhibitors. Three potent Chk1 inhibitors were discovered with IC_{50} values ranging from 2.4 to 69.7 μM.

Keywords Check point kinase 1 · k nearest neighbor · dbCICA · Pharmacophore · In silico mining

Electronic supplementary material The online version of this article (doi:10.1007/s10822-015-9848-1) contains supplementary material, which is available to authorized users.

✉ Mutasem Omar Taha
mutasem@ju.edu.jo

[1] Faculty of Pharmacy, Zarqa University, Zarqa, Jordan

[2] Department of Pharmaceutical Sciences, Faculty of Pharmacy, The University of Jordan, Amman, Jordan

[3] Faculty of Pharmacy, Applied Sciences University, Amman, Jordan

Introduction

Check point kinase 1 (Chk1) is a serine-threonine kinase that has important role repairing DNA damage. Cell cycle checkpoints temporarily stop the progression of cell cycle to allow time for the repair of the DNA damage in order to maintain the genomic integrity and the survival of cells [1, 2].

Apparently, from therapeutic point of view, Chk1-mediated S or G2 checkpoint arrests and subsequent DNA repair in tumor cells limit the efficacy of radiation therapy and cytotoxic drugs leading to drug resistance. It has been suggested that Chk1 inhibition would preferentially sensitize tumors to DNA damaging agents [3, 4].

Normal cells depend on arrest at G1 phase for repairing DNA damage; while tumor cells (particularly p53-mutated) perform their DNA repairs during cell cycle arrest via S and G2 checkpoints due to their incapability of G1 arrest. As a result, specific inhibition of S and G2 arrests through inhibition of Chk1 is expected to selectively drive p53-deficient tumor cells to enter mitotic catastrophe, and eventually apoptosis following DNA damage [5, 6]. This assumption has been recently confirmed by the reported increase in effectiveness of DNA-damaging agents when given to p53-mutated tumor cells after knocking down their Chk1 using small interfering RNA (siRNA) or Chk1 antisense technique [7–9], or even if inhibited by a natural product such as 7-hydroxystaurosporine [10], which proves Chk1 to be a target for selective chemosensitization.

Docking-based comparative intermolecular contacts analysis (dbCICA) is a novel 3D QSAR analysis methodology developed by us to validate docking settings and to extract valid structure-based pharmacophores [11–13]. In this approach the interest is directed on recognizing a group of atoms in the active site that selectively contact with active ligands while evade inactive ones. If such a set of binding

site contact atoms is identified for a group of docked ligands, then it can be safely assumed that the corresponding docking settings are successful, i.e., it managed to arrange the molecules in the binding site in a way that justifies the variation in their bioactivities [11–13]. Furthermore, critical contact points identified by dbCICA modeling can be transformed into pharmacophore model(s) that can be used for in silico mining for novel hits [11–13].

kNN based quantitative structure activity relationship (QSAR) methodology relies on a distance learning methodology for calculating and predicting ligands bioactivity; that is the activity value of an unknown ligand structure can be predicted from the activity values of a certain number (k) of nearest neighbors to that ligand (kNNs) in the training set. The neighborhood is defined depending on certain selected descriptors (in this case ligand-receptor contact points) while the nearness is measured by an appropriate distance metric (e.g., molecular similarity measure) [14–16].

The greatest advantage of kNN-based QSAR modeling is its ability to non-linearly correlate bioactivity with various molecular descriptors [14–16], which makes it handy in cases of complicated structure–activity relationships.

From our experience in dbCICA modeling [11–13], we noticed that in some cases the relationship between ligand-receptor contacts and corresponding ligands' bioactivities can be rather complex and hard to model using conventional linear regression methods, which can lead to false conclusions about the optimal docking conditions for a particular list of inhibitors. This trend prompted us to attempt using kNN methodology as correlation platform in dbCICA modeling.

We coupled kNN with genetic function algorithm (GFA) to find the finest combination of ligand-receptor contacts that can explain bioactivity variations among training compounds. We decided to test the new methodology (kNN-dbCICA) to search for new Chk1 inhibitors. Our decision was driven by interesting anti-cancer properties of Chk1 inhibitors and the availability of proper crystallographic structures for this protein. However, we were inclined to perform the study on a single crystallographic example to avoid unnecessarily extending the study (Fig. 1).

Experimental

Molecular modeling

Software and hardware

The molecular modeling software packages used in the current research were installed on standard personal computers:

- CS ChemDraw Ultra (Version 7.0.3), Cambridge Soft Corp., USA.
- MarvinView (Version 5.1.4), ChemAxon Ltd., Budapest, Hungary.
- Ligandfit implemented in Discovery Studio 2.5.5 (DS2.5.5), Accelrys Inc., San Diego.
- MATLAB (Version R2007a), The MathWorks Inc., USA.
- Cat.Search implemented in Discovery Studio (version 2.5.5), Accelrys Inc., USA.

Data set

The structures of 192 Chk1 inhibitors were collected from published literature [17–20]. The ligands were carefully gathered in such away to make sure that their anti-Chk1 activities were determined via identical bioassay procedure, which should permit appropriate QSAR correlation. The bioactivities of the collected inhibitors were expressed as IC_{50} values (Table A under the Supplementary Material).

Preparation of ligands

The 2D structures of the Chk1 ligands (1–192, Table A under the Supplementary Material) were sketched in ChemDraw Ultra (Version 7.0.3). We used MarvinView to guide us in identifying ionized and un-ionized states for ionizable inhibitors. Subsequently the structures were translated into sound 3D representations by using rule-based methods implemented in Discovery Studio and were saved in SD format for subsequent docking experiments. The Log(IC_{50}) (nM) values were applied in dbCICA analysis, hence linearly correlating the biological data to the free energy change.

Preparation of Chk1 crystal structure

The 3D structure of Chk1 was taken from the Protein Data Bank (PDB code: 3TKI, resolution: 1.6 Å). We selected this particular protein structure based on two important issues: (1) its superior resolution (1.6 Å), and (2) the fact that the co-crystallized ligand in this protein has the highest affinity ($IC_{50} = 0.05$ nM) among all high-resolution (<2.0 Å) Chk1 crystallographic complexes.

Hydrogen atoms and Consistent Force Field charges were added to the protein atoms as implemented within Discovery Studio. Docking experiments were repeated on hydrous and anhydrous versions of the binding pocket (ATP binding site).

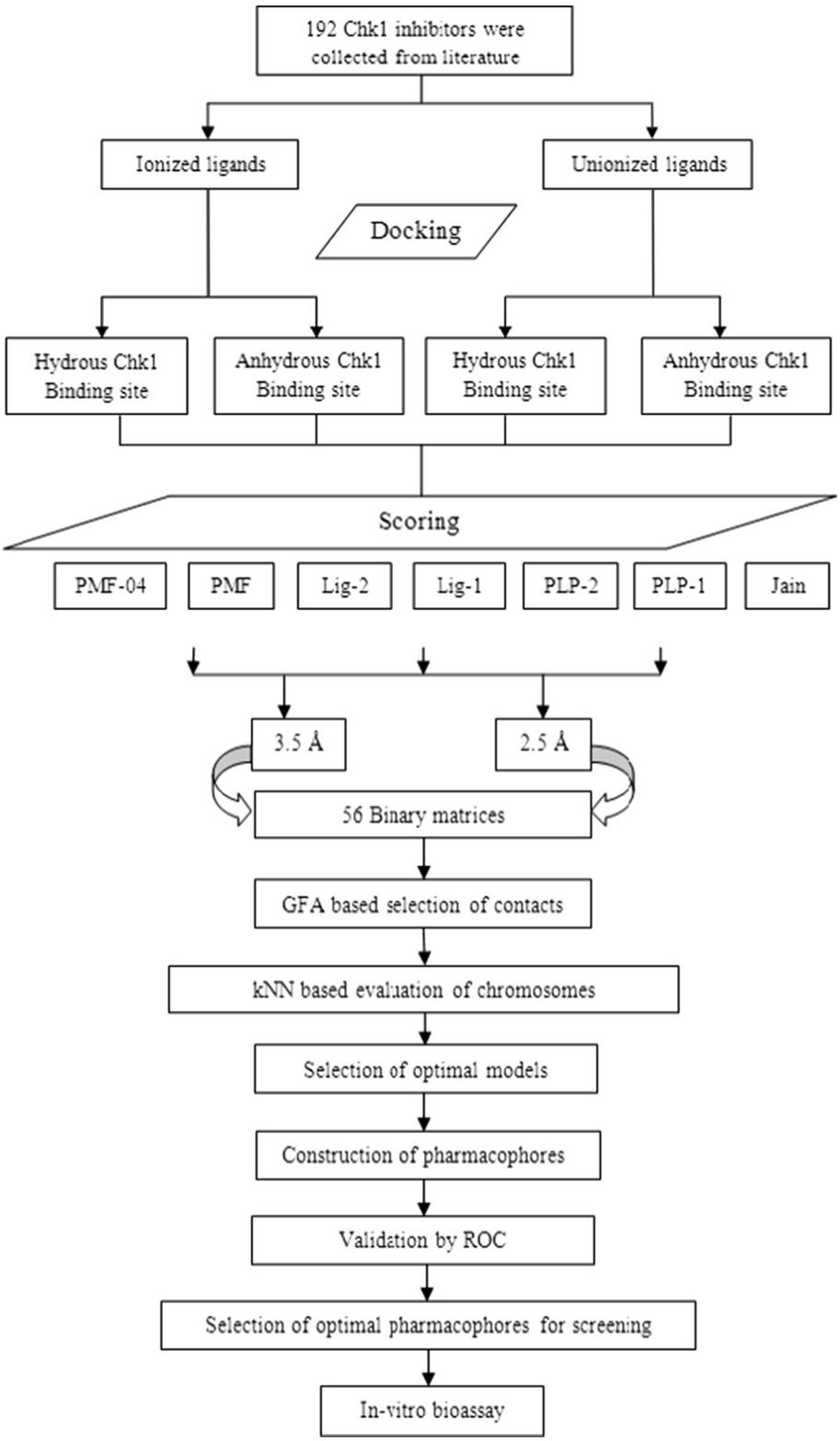


Fig. 1 General computational workflow implemented herein for discovering new Chk1 inhibitors by kNN-dbCICA methodology. *GFA* genetic function approximation, *kNN* k nearest neighbour, *ROC* receiver operating characteristic curve analysis

LigandFit docking

LigandFit docking engine considers the flexibility of the ligand while treats the receptor as rigid [21]. The steps that were implemented in the LigandFit docking process and corresponding docking settings are described in Section SM-1 in the Supplementary Material. High ranking docked conformers/poses were scored using 7 scoring functions,

namely: Jain, LigScore1, LigScore2, PLP1, PLP2, PMF and PMF04 [21]. Accordingly, we performed a total of 28 docking experiments corresponding to two ligand ionization state, two protein hydration states and seven scoring functions.

kNN-implemented docking-based comparative intermolecular contacts analysis (kNN-dbCICA)

This approach is described by the following sequential points and summarized in Fig. 1:

(i) *Assigning contacts-based binary code* The docked poses/conformers of each inhibitor (based on a certain docking configuration) are assessed to determine their nearby atoms within the ATP-binding site. Atomic neighbors that lie closer than (or equal to) definite predefined distance threshold are allocated an intermolecular contact value of "one", otherwise they are given a contact value of "zero". For example, if the distance between atom A in the docked ligand and atom B in the binding pocket was less than or equals the predefined threshold, then this contact (B) will be given value of 1. Distance evaluations were automatically performed employing the Intermolecular Monitor of DS 2.5.5 at two distance thresholds: 3.5 and 2.5 Å.

Ultimately, this step generates a 2D matrix for each docking-scoring configuration. Each matrix is composed of row labels corresponding to docked ligands and column labels corresponding to different binding site atoms. The matrix is filled with binary code, whereby "zeros" correspond to inter-atomic distances that exceed the predetermined threshold and "ones" for distances below (or equal) the predefined threshold cutoffs.

As a result, two binary matrices (corresponding to each distance threshold) were constructed for each docking configuration (i.e., ligand ionization state, and binding site hydration state and scoring function). Accordingly, a total of 56 binary files were generated in the study, i.e., corresponding to 2 ligand ionization states × 2 protein hydration states x 7 scoring functions × 2 distance thresholds. This complication forced us to perform the study on a single crystallographic example to avoid unnecessarily extending the study.

(ii) *Removing sparse contacts* The resulting binary matrices were inspected for sparse contacts, such that any binding site atom having ≤2 contacts with docked ligands was omitted from the respective binary matrix to minimize statistical noise in subsequent analyses. For example, in one case (docking ionized ligands into the hydrous binding site using LigScore2 scoring function and collecting binding site contacts at 3.5 Å distance threshold), the total number of contacts were 254, out of which 63 were sparse contacts.

(iii) *GFA-kNN-based contacts selection* The resulting binary matrices were then presented as input files to GFA-based search for best contacts (columns) capable of explaining bioactivity variation via k nearest neighbor correlation methodology: GFA relies on the evolutionary operations of "crossover and mutation" to choose combination of contacts (columns) and then test their ability to explain bioactivity variation across the training compounds using k nearest neighbors analysis.

(A) We used the GFA toolbox within MATLAB (Version R2007a) for chromosome creation, mutation function, cross-over function and fitness function.

GFA operates in a cycle of four phases: (1) encoding mechanism; (2) defining a suitable fitness function; (3) creating a population of chromosomes (vectors); (4) genetic manipulation of chromosomes [22]. We employed a gene-based encoding system where the created models (chromosomes) differ from each other by the set of independent variables (intermolecular contacts) that comprise each model. If the general number of independent variables (contacts) is equal to P (in this particular case, P corresponds to contacts columns in a binary matrix), the chromosome corresponding to any model consists of a string of P binary digits (bits) called "genes". Each value in the string represents an independent variable (intermolecular contact, 0 = absent, 1 = present). Each chromosome is associated with a fitness value that reflects how good it is compared with other chromosomes. The following are important control parameters in GFA to select the finest descriptors:

- *Creating an initial population* In our study we implemented an initial population size of 120 chromosomes.
- *Mating population* Mating is a process during which two parent chromosomes are joined to generate new solutions (offspring). The probability of mating can take values in the range between 0

and 100 %. In our study we set the mating probability to be 80 % of population chromosomes.

- *Mutation operator* A mutation operator changes one or more bits (or genes) in the chromosome to its complement according to a given probability. In our study we implemented Gaussian-based random mutation.
- *Maximum number of generations* This is needed to exit from a basic cycle and complete the algorithm [22]. In our study we specified a maximum number of genetic iterations (generations) of 1500.
- GFA can be controlled to yield best kNN-dbCICA models resulting from any specified number of ligands'-receptor intermolecular contacts. In this study we employed GFA to search for the best kNN-dbCICA models resulting from 5 contacts and repeat the process to reach for the best models for 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15 contacts. GFA was employed to search for the best possible combination of contacts capable of explaining variation in biological activities among the training compounds.

(B) kNN-based Correlation method relies on a distance learning approach such that the activity value of an unknown ligand is predicted from the activity values of certain number (k) of its nearest neighbors (kNNs) in the training set. The similarity is measured using a distance metric that measures the distance of the unknown compound to its nearest neighbors with regards to the GFA-selected contacts within the binding site. In the current study the Euclidean distance is considered. The standard kNN method is implemented as the following workflow: (1) calculate distances between an unknown object (e.g., x) and all the objects in the training list (2) select k objects from the training set closest to object x, with respect to their calculated distances from each other; (3) use weighted average to calculate the activity value of object x referring to its kNNs. The best k value was found experimentally to lie between 1 and 5 [14, 15]. However, in our kNN approach we scanned k values ranging from 3 to 10. (4) The fitness function employed herein is $r^2_{L20\ \%O}$. In this method 20 % of the observations are removed of the training set, and their activities are predicted using weighted average of k closest neighbors. The process is repeated over five cycles and in each cycle the selected testing set is different from those for the other cycles. The predicted activity value of each compound based on the weighted average of its nearest neighbors is calculated using Eq. (1).

$$\overline{Y}x = \frac{\sum_{k-nearest\ \ neighbors} y_i d_i}{\sum_{k-nearest\ \ neighbors} d_i} \quad (1)$$

where $\overline{Y}_x$ is the predicted activity of ligand x, y_i represent the activities of the closest k-neighbors, and d_i is the Euclidean distance of the compound from its kNNs. The leave 20 %-out cross-validation coefficient is calculated by Eq. (2) over five cycles and the final reported value is the average.

$$r^2_{L20\%O} = 1 - \frac{\sum_{x=1}^{Testing\ set} (y_x - \overline{y_x})^2}{\sum_{x=1}^{Testing\ set} (y_x - y_{avg.tr})^2} \quad (2)$$

where y_x is the experimental bioactivity of compound x and $y_{avg.tr}$ is the average bioactivity of training compounds (i.e., after removing the testing set).

(iv) *Determination of Directly and Inversely Proportional Contacts To Bioactivity* After identifying optimal kNN-dbCICA model(s) we scanned the corresponding intermolecular contacts using multiple linear regression analysis to identify directly and inversely proportional binding site contacts with ligand's bioactivities. Correlations were performed by collectively regressing contact points, identified by a particular high ranking dbCICA-kNN model, against bioactivity (as log(1/IC50)) in a single multi-term equation. We then defined directly proportional contacts points (favorable contacts) as those having positive regression slopes, while contact points having negative regression slopes were defined as being disfavored (clash points). Accordingly, contacts were classified as pharmacophoric (positive) contacts and forbidden (exclusion) contacts. Exclusion contacts influence bioactivity in a negative way as they probably resemble sterically inaccessible regions within the binding pocket.

Manual generation of pharmacophores guided by successful kNN-dbCICA models

The best kNN-dbCICA models were used to guide the building of corresponding pharmacophores to be applied as search queries for the discovery of new Chk1 inhibitors. The pharmacophore models were developed via the following steps:

1. The docking conditions that gave the best kNN-dbCICA models were selected, e.g., Ligandfit docking of ionized ligands into the binding pocket of a hydrated protein and using PMF04 as a scoring function (see Results and Discussion). The corresponding docked poses/conformers of the most potent-well behaved training compounds were kept in the binding pocket while other less potent compounds were removed. Well-behaved compounds are defined as those training compounds with well-predicted bioactivities by the selected optimal kNN-dbCICA model, i.e., they have the least residual difference between fitted and experimental bioactivities as predicted by the respective kNN-dbCICA model.
2. Significant positive molecular atomic contacts within the binding pocket were displayed and carefully inspected to determine their close ligands' moieties. Agreement among potent, well-behaved training compounds on placing features of common physicochemical properties close to significant contact atom (as determined by the kNN-dbCICA model) guided us to place a corresponding pharmacophoric feature onto that position. For example, if potent, well-behaved docked compounds have agreed on placing aromatic rings adjacent to certain kNN-dbCICA significant contact atoms (within the predefined distance threshold in the binding pocket) then a hydrophobic aromatic feature was placed onto the aromatic rings. The Pharmacophoric features of the optimal models were added manually from DS2.5.5 feature library and using default feature tolerance radii (1.6 or 2.2 Å). It has to be mentioned that the presence of certain significant contact atom in the binding site vicinity does not necessarily indicate significant ligand interaction(s) with that particular contact atom in the binding pocket, but it probably indicates significant interaction in the neighborhood of that atom.
3. In order to account for the steric constraints of the binding pocket, binding site atoms that present significant kNN-based contacts and have negative linear regression correlations with bioactivity were presented in the corresponding pharmacophoric model(s) as centers of exclusion spheres. Negative contacts point to spaces occupied by docked conformers/poses of inactive compounds and free from active ones and therefore can be filled with exclusion volumes. Pharmacophoric exclusion spheres were added manually from DS 2.5.5 feature library and employing default feature radii (1.2 Å).

Receiver operating characteristic (ROC) curve analysis

kNN-dbCICA-based pharmacophores were validated by evaluating their abilities to selectively discriminate and identify anti-Chk1 active compounds from a large testing decoy list composed of few known active ligands and a large number of decoys [23, 38, 39]. Description on how to prepare this list together with detailed information about ROC analysis are found in section SM-2 under the Supplementary Material.

In-silico screening of the NCI-database for new Chk1 inhibitors

Pharmacophore models derived from best kNN-dbCICA models were implemented as 3D search queries to mine the NCI database (238,819 compounds) to discover new anti-Chk1 inhibitors. Virtual screening was performed using the "Best Flexible Database Search" option available within CATALYST.

NCI hits were then filtered based on Lipinski's rule of five such that only hits of molecular weights $\leq$500 dalton, H-bond donors less $\leq$5, H-bond acceptors $\leq$10, and log $p \leq 5$ were retained. Remaining hits were fitted against corresponding kNN-dbCICA pharmacophores using the "best fit" option implemented within CATALYST. Top ranking hits were obtained from the NCI and later examined against Chk1.

Biological evaluation of captured hits

Human recombinant Chk1 was purchased from Life Technology (Carlsbad, CA). The bioassay was conducted using Invitrogen Z'-LYTE® Ser/Thr 19 Peptide assay kit. The bioassay was performed as directed in provider protocol using Chk1 and ATP concentrations of 14 nM and 100 μM, respectively. Hits were tested at different concentrations ranged from 100 nM to 100 μM. At least two data points from each concentration were collected. The IC_{50} value of each hit was calculated using nonlinear regression of the log(concentration) vs % inhibition values using GraphPad Prism 5.0. The settings of the assay were validated using positive (Staurosporine) and negative (provided in Z'-LYTE™ Kinase Assay kit) controls.

Results and discussion

Background

Docking is essentially a conformational searching tool to find the optimal ligand pose/conformation with certain binding pocket. However, it can be misleading because of the inability of docking engines to determine the entropic cost and therefore free energy of binding. This point combined with other problems related to the quality of crystallographic receptor structures (e.g., resolution), the significance of explicit water molecules in the binding process, and ionization/flexibility of the binding site can render docking experiments unreliable and warrant strict validation [24–27]. However, despite their limitations, docking engines normally succeed in yielding the experimental ligand pose/conformation among high-ranking docked solutions [28–31]. As a result it's reasonably possible to find docked inhibitor 3D poses/conformers consistent with corresponding bioactivities [24–26].

In dbCICA, the concern is directed on recognizing a group of atoms within the binding site that come in contact with active inhibitors while avoid inactive ones. If these contact atoms are identified for a docked list of inhibitors, then one can assume that the docking setting is successful, i.e., it managed to put the ligands in an arrangement that can explain bioactivity variation [11–13, 24–26].

Certainly, high ligand-receptor affinity is mediated by certain list of critical interactions (hotspots). However, because docking packages and scoring algorithms assess large number of ligand-receptor interactions while generating their docking results, the impact of these critical interactions on affinity estimation can be diluted [11–13]. In this context, determining a group of affinity-discriminating contact atoms in the binding site (i.e., dbCICA) is not only supposed to help in validating a particular docking setting, but also it should help to pinpoint critical ligand-receptors interactions that correlate with affinity [11–13]. Indeed, the generated dbCICA models can be used to build pharmacophores of limited number of binding features that are computationally efficient to search for new hits [11–13].

Unfortunately, our experience with dbCICA modeling tells us that this method occasionally fails to find sets of binding site contacts capable of explaining bioactivity variations. We propose that such failures are related to failure of the combined genetic algorithm-multiple linear regression analysis (GFA-MLR) implemented within dbCICA, i.e., in finding optimal binding site contacts capable of explaining bioactivity. Apparently, the basic fault behind such failures is related to assuming, and therefore, enforcing linear relationships between binding site contacts and corresponding ligands' bioactivities.

On the other hand, the combination of kNN with GFA allows efficient search for linear and/or nonlinear QSAR correlations among extended sets of explanatory descriptors. Accordingly, we assumed that implementing GFA-kNN in dbCICA analysis should allow access to complex nonlinear relationships that are otherwise inaccessible using GFA-MLR-dbCICA modeling. Significant contacts can be subsequently used as reference points to build binding pharmacophoric features.

Application of kNN-dbCICA for discovery of new Chk1 inhibitors

The availability of high-quality crystal structures for the kinase domain of human Chk1 puts structure-based inhibitor design strategies against human Chk1 kinase on firm footing towards developing new anti-Chk1 inhibitors as potential anticancer agents [40].

However, structural analysis of single (or few) receptor–ligand complexes is not adequate to identify critical binding interactions if it is not followed by appropriate SAR analysis [34]. Therefore, we decided to implement kNN-dbCICA modeling to find the optimum docking settings necessary for effective docking of 192 Chk1 ligands and to identify critical binding interactions involved in ligand-

Table 1 Highest-ranking Chk1 inhibitors kNN-dbCICA models, their corresponding parameters and statistical criteria

kNN-dbCICA model	Ligand ionization state	Explicit water[a]	Scoring function	Contact distance threshold (Å)[b]	Optimal number of nearest neighbors (k)	Number of positive contacts	Number of negative contacts	Correlation statistics		
								r^2	r^{2c}_{LOO}	$r^{2d}_{L\text{-}20\ \%}$
1	Ionized	Present	Ligscore2	3.5	3	8	6	0.76	0.62	0.57
2	Ionized	Present	PMF04	3.5	4	9	6	0.76	0.67	0.65

[a] Crystallographically explicit water of hydration

[b] Distance thresholds used to define ligand-binding site contacts

[c] Cross-validation correlation coefficients determined by the leave-one-out technique

[d] Cross-validation correlation coefficients determined by the leave-20 %-out technique repeated five times

Table 2 Critical binding site contact atoms proposed by optimal kNN-dbCICA models

kNN-dbCICA model	Amino acids and corresponding atom identities[a]	
	Positive contacts	Negative contacts
1	ALA36:HB1	
	ASP94:OD2	
	GLU134:HG1	GLN13:OE1
	LEU15:HA	LEU137:HD11
	LEU84:HB2	LEU84:CD1
	VAL23:HB	SER88:CA
	Water104:O	TYR86:CA
	Water186:H1	VAL23:HG12
2	ARG95:HH21	
	ASP94:CG	
	ASP94:HB2	LYS38:HB1
	GLY16:CA	VAL23:HG11
	GLY16:HA1	Water163:H2
	GLY90:CA	Water165:O
	THR14:HB	Water25:H1
	Water209:H2	Water64:H1
	Water80:O	

[a] Binding site amino acids and their significant atomic contacts. Atom codes are as provided by the protein data bank file format (e.g., THR14:HB encodes for Hydrogen atom (B) of Threonine number 14) except for hydrogen atoms which were coded by DS2.5

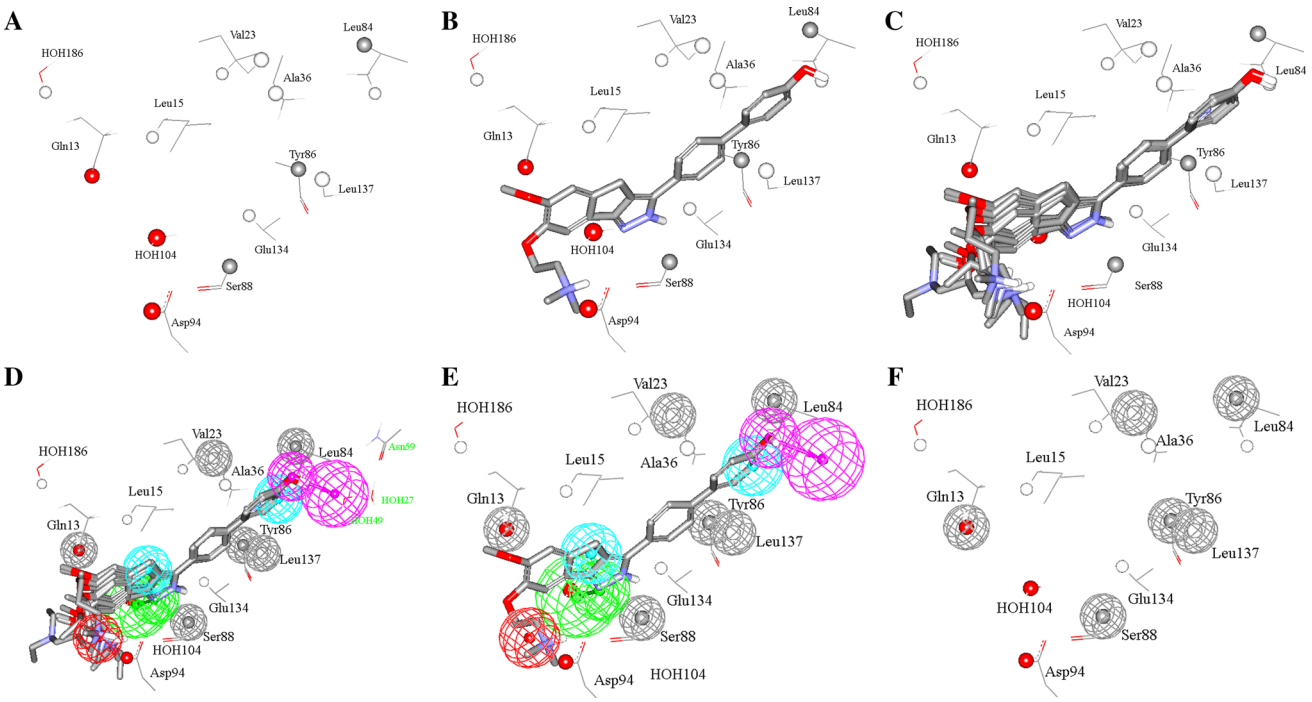


Fig. 2 Steps for manual generation of binding hypothesis Hypo1 as guided by kNN-dbCICA model **1** (Tables 2 and 3): **a** The binding site moieties in kNN-dbCICA model **1** with significant contact atoms shown as spheres. **b** The docked pose of the well-behaved compound **33** (IC_{50} = 0.8 nM) within the binding pocket **c** The docked poses of the well-behaved and potent compounds **33**, **34, 41, 4, 18** and **90**. **d** Manually placed pharmacophoric features onto chemical moieties common among docked well-behaved potent compounds **33**, **34, 41, 4, 18** and **90**. **e** The docked pose of **33** (well-behaved and potent, IC_{50} = 0.8 nM) and how it relates to the proposed pharmacophoric features. **f** Exclusion spheres fitted against binding site atoms showing negative correlations with bioactivity (as emergent in kNN-dbCICA model **1**), atom colors were used according to CPK system (*H white*, C *grey*, O *red*)

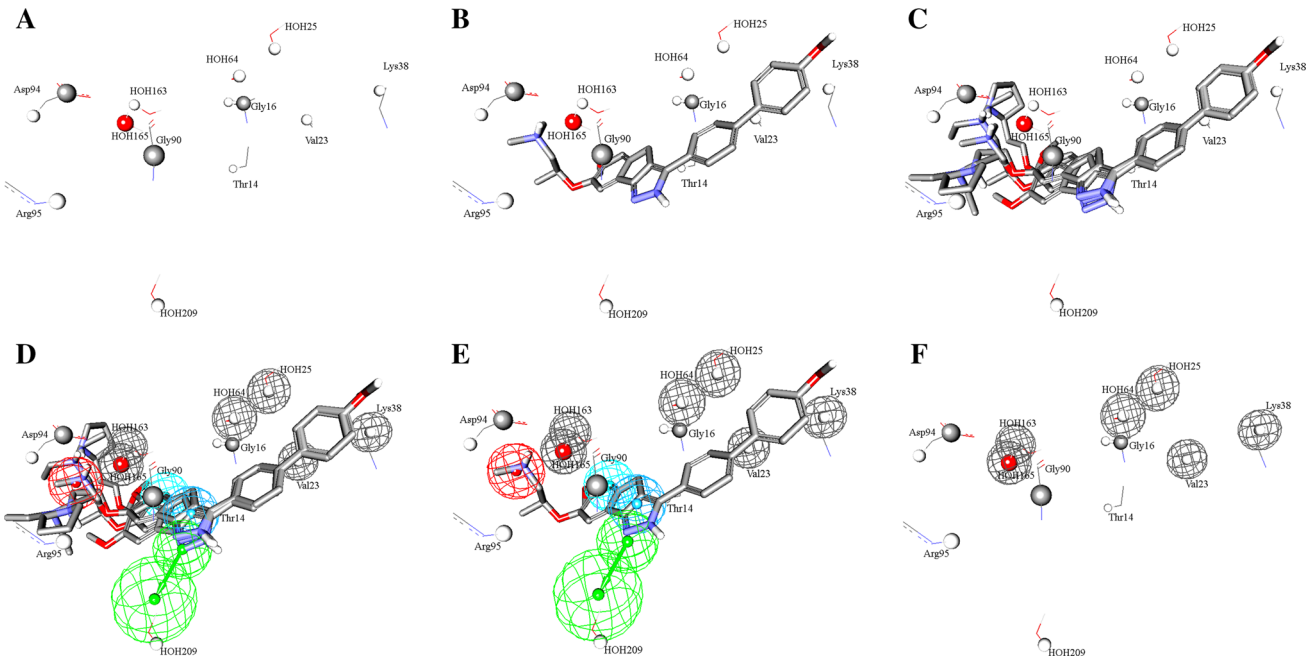


Fig. 3 Steps for manual generation of binding hypothesis Hypo2 as guided by kNN-dbCICA model **2** (Tables 2 and 3): **a** The binding site moieties in kNN-dbCICA model **2** with significant contact atoms shown as spheres. **b** The docked pose of the well-behaved compound **34** ($IC_{50} = 2$ nM) within the binding pocket **c** The docked poses of the well-behaved and potent compounds **34, 19, 36** and **41**. **d** Manually placed pharmacophoric features onto chemical moieties common among docked well-behaved potent compounds **34, 19, 36** and **41**. **e** The docked pose of **34** (well-behaved and potent, $IC_{50} = 2$ nM) and how it relates to the proposed pharmacophoric features. **f** Exclusion spheres fitted against binding site atoms showing negative correlations with bioactivity (as emergent in kNN-dbCICA model 2), atom colors were used according to CPK system (H: *white*, C: *grey*, O: *red*)

Chk1 complexation. Furthermore, we translated the optimal kNN-dbCICA models into pharmacophore hypotheses and used them as 3D search queries to search for new anti-Chk1 inhibitors from NCI database.

Initially all 192 compounds were docked into the active site of Chk1 applying LigandFit [21]. We selected a particular Chk1 protein structure (3TKI) as docking template based on its superior resolution (1.6 Å) and the high affinity of its co-crystallized ligand ($IC_{50} = 0.05$ nM) (see Sect. 2.1.4. under Experimental). Needless to say that high affinity ligands imprint their corresponding proteins in such a way that exposes critical binding site hotspots to docked ligands. Critical binding hotspots need to be readily accessible so that the docking engine can evaluate interaction energies involving these hotspots with docked ligands, and therefore, successfully explain biological variability among ligands. Moreover, we strongly believe crystallographic structures corresponding to high affinity ligands can reasonably replace consideration of receptor flexibility during docking simulations because high-affinity ligands modify the conformation of the binding pocket in the best possible way.

Four docking settings were implemented: Ionized ligands *versus* unionized ligands, and hydrous binding site versus anhydrous binding site. This workflow is necessary since it is hard to predict the influence of ligand ionization and the hydration state of the binding site on the outcome of docking studies prompting us to evaluate the docking outcomes of all possible combinations resulting from ligand ionization and ATP-binding site hydration.

Consequently, highest-ranking docking solutions were scored by 7 different scoring functions; namely: LIGSCORE1, JAIN, LIGSCORE2, PLP1, PLP2, PMF, and PMF04 [21].

The highest-ranking conformers/poses, according to each docking configuration/scoring function, were aligned together within the binding pocket. Subsequently, intermolecular inhibitor-binding site contacts were identified applying two distance thresholds: 2.5 and 3.5 Å. As a result, two contacts binary matrices were generated for each docking solution.

Subsequent implementation of GFA/kNN-based exploration identified many kNN-dbCICA models. However, the highest ranking models were chosen based on their statistical parameters.

Table 1 shows the contacts' distance thresholds, list of positive and negative contacts, and statistical measures of optimal kNN-dbCICA models. Clearly from the table, hydrous binding site and ligand ionization enhanced the qualities of the kNN-dbCICA models as both successful models were based on docking configurations performed on ionized

Table 3 Pharmacophoric features, corresponding tolerances and 3D coordinates (X, Y, Z) of optimal kNN-dbCICA-based pharmacophore models

Model[a]	Definitions	Chemical features						
		Hbic[c]	Hbic	HBD[d]		HBA[e]		PosIon[f]
Hypo1[h]	Tolerances[b]	1.60	1.60	2.20	1.60	2.20	1.60	1.60
	Coordinates							
	x	20.84	11.70	9.06	10.30	22.30	21.65	26.77
	y	−3.32	−0.38	0.83	−1.65	0.90	−0.18	−4.93
	z	9.08	10.81	13.53	12.37	4.68	7.40	12.06
		HBA		Hbic-Arom[g]		Hbic	PosIon	
Hypo2[i]	Tolerances	2.20	1.60	1.60		1.60	1.60	
	Coordinates							
	x	23.64	21.65	20.74		22.23	26.92	
	y	1.62	−0.12	−1.87		−5.15	−4.12	
	z	5.94	7.37	8.49		8.78	11.58	

[a] As in Table 2

[b] Refer to the radii of the feature spheres (Å)

[c] Hydrophobic feature

[d] Hydrogen Bond Donor feature

[e] Hydrogen Bond Acceptor feature

[f] Positive Ionizable feature

[g] Aromatic hydrophobic feature

[h] Number of Associated Exclusion Spheres = 6 of 1.2 Å tolerance, at the following X, Y, Z coordinates: (22.02, −3.43, 1.94), (15.30, 0.12, 12.85), (8.32, −1.12, 9.73), (20.96, 3.66, 6.53), (14.80, 3.19, 6.30) and (12.41, −4.24, 7.97)

[i] Number of Associated Exclusion Spheres = 6 of 1.2 Å tolerance, at the following X, Y, Z coordinates: (9.23, −4.08, 9.40), (13.80, −3.86, 9.02), (23.40, −5.05, 12.61), (23.61, −2.45, 12.25), (14.75, −3.25, 13.59), (17.00, −2.20, 12.93)

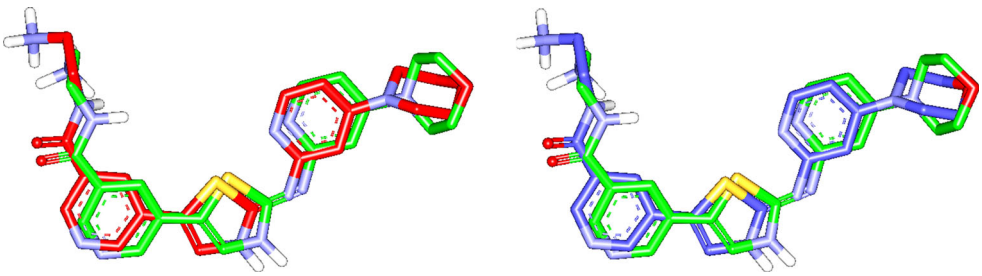


Fig. 4 Comparison between the co-crystallized pose of 3TKI ligand (*Green*, PDB code: 3TKI, Resolution 1.6 Å) and its docked pose according to docking conditions of kNN-dbCICA models 1 (on the *left*, *Red*) and 2 (on the *right*, *Blue*) as in Table 1

ligands into hydrous binding site. Accordingly, it can be safely assumed that ligand ionization and hydrous binding site yield the most realistic docking conditions. Moreover, PMF04 scoring function seems to be the most successful in aligning the docked molecules in such a way to correlate the docked poses with corresponding bioactivities, as seen in the statistical criteria of kNN-dbCICA model 2 (Table 1).

Table 2 shows important Chk1 active site contact atoms (both positive and negative ones) suggested by optimal kNN-dbCICA models 1 and 2. Clearly from the table, the highest ranking kNN-dbCICA models point to almost completely different sets of binding site contacts, suggesting two different corresponding binding modes.

Figures 2 and 3 show how dbCICA models 1 and 2 were translated into corresponding pharmacophores (hypo1 and hypo2, respectively) within Discovery Studio 2.5 environment (see Sect. 2.1.7 under Experimental). Initially, the significant contact atoms in the binding pockets were marked in spherical forms (see a in Figs. 2, 3). Subsequently, we kept only potent ($IC_{50} < 15$ nM) and well-behaved docked compounds in the binding pocket, i.e., those of least difference between experimental and fitted bioactivities (determined by kNN-dbCICA models 1 or 2) as shown in b and c in Figs. 2 and 3. Thereafter, proper pharmacophoric features were placed onto common aligned chemical functionalities

among docked compounds in such a way to highlight the interactions encoded by the critical contacts, as in d and e in Figs. 2 and 3. For example, in kNN-dbCICA model 1, emergence of significant contact at the carboxylic oxygen of ASP94, combined with the consensus of well-behaved, potent docked ligands on placing nearby quaternary ammonium groups, prompted us to place a positive ionizable feature onto the ligand's ammonium groups (Fig. 2d). Similarly, agreement of docked, potent and well-behaved compounds on placing hydrophobic moieties near the hydrogen atoms HB1 and HA of ALA36 and LEU15, respectively (both emerged as significant contacts in model 1, Table 2), prompted us to place hydrophobic pharmacophoric features onto both hydrophobic positions within the ligands, as in Fig. 2d. Comparably, agreement among potent, well-behaved compounds in positioning their pyrazole rings close to the oxygen atom of Water104 (which is a critical contact in kNN-dbCICA model 1) in such a way that the sp2 hybridized pyrazole nitrogen is projected towards Water104 suggested placing hydrogen bond acceptor feature in that location, as in Fig. 2d, e.

Finally, the agreement among docked, potent and well-behaved training compounds in placing phenol groups at close proximity (i.e., within contact threshold distance of 3.5 Å) to the critical hydrogen contact of LEU84 prompted us to look for a nearby interaction responsible for this contact. We quickly realized that a hydrogen-bonding interaction tying the ligands' phenolic hydroxyls to Water49 (shown in Fig. 2d) is most likely responsible for this significant contact. Accordingly, we positioned a hydrogen bond donor feature onto the ligands' phenolic hydroxyls to encode for this interaction. Incidentally, Water49 is fixed by hydrogen-bonding to ASN59 via bridging Water27 (hydrogen bonding bridge components are shown in green labels) as shown in Fig. 2d. A similar approach was implemented for building Hypo2 pharmacophore based on kNN-dbCICA model 2, as in Fig. 3. Table 3 shows the X, Y, Z coordinates of the two pharmacophores.

It remains to be mentioned that significant inverse intermolecular contacts identified by kNN-dbCICA modeling (Table 2) were annotated as exclusion spheres in the pharmacophore models to indicate sterically disfavored areas for ligand binding within the binding pocket (Figs. 2, 3).

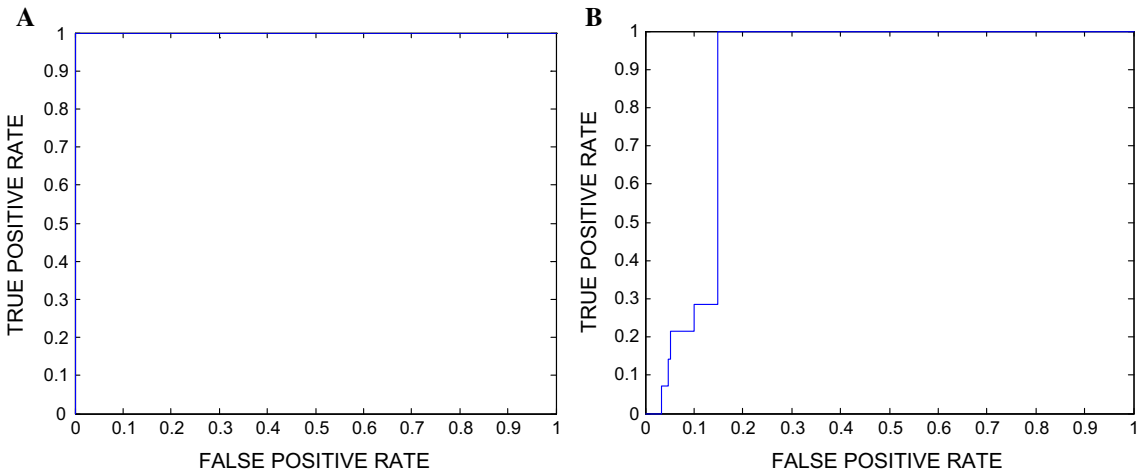


Fig. 5 Receiver operating characteristic (ROC) curves of kNN-dbCICA-based pharmacophores. **a** Hypo1, **b** Hypo2

Table 4 ROC[a] performances of kNN-dbCICA selected pharmacophores as 3D search queries

Pharmacophore model	ROC[a]–AUC[b]	ACC[c]	SPC[d]	TPR[e]	FNR[f]
Hypo1	1.00	0.97	0.99	0.21	0.0073
Hypo2	0.88	0.97	0.99	0.29	0.0097

[a] *ROC* receiver operating characteristic

[b] *AUC* area under the curve

[c] *ACC* overall accuracy

[d] *SPC* overall specifity

[e] *TPR* overall true positive rate

[f] *FNR* overall false negative rate

Table 5 High-ranking captured hits by kNN-dbCICA models 1 and 2, predicted IC_{50} values according to QSAR Eq. (1) and in vitro bioactivities

Hit[a]	NCI code	Hit captured by	Predicted IC_{50} (nM) based on kNN-dbCICA models[b]		% inhibition at 10 μM	IC_{50} (μM)
			1	2		
193	35839	Hypo1	NA[c]	NA[c]	−28	–
194	38278	Hypo1	NA[c]	NA[c]	−21	–
195	38279	Hypo1	96.9	13.9	−23	–
196	50464	Hypo1	NA[c]	NA[c]	5	–
197	53212	Hypo1	NA[c]	NA[c]	−22	–
198	57149	Hypo1	NA[c]	NA[c]	0	–
199	63689	Hypo1	NA[c]	NA[c]	−23	–
200	63697	Hypo1	NA[c]	NA[c]	0	–
201	65373	Hypo1	NA[c]	NA[c]	6	–
202	67544	Hypo1	NA[c]	NA[c]	3	–
203	70706	Hypo1	NA[c]	NA[c]	0	–
204	72378	Hypo1	NA[c]	NA[c]	0	–
205	79546	Hypo2	2.0	675.3	0	–
206	79549	Hypo2	1.7	3.1	0	–
207	79550	Hypo2	1.7	3.1	−27	–
208	80117	Hypo1	20.0	96.9	16	69.7
209	86367	Hypo2	1.0	4.0	80	2.4
210	86370	Hypo2	8.6	1.0	89	2.6
211	94026	Hypo2	5.4	2.9	0	–
212	138659	Hypo2	11.0	6.3	2	–
213	138660	Hypo2	1.0	1.6	0	–
214	138661	Hypo2	11.0	1.6	0	–
215	143704	Hypo2	1.8	2.9	−34	–
216	153422	Hypo2	1.4	3.9	−31	–
217	166674	Hypo2	1.5	1.9	−33	–
218	168201	Hypo2	1.0	2.7	−29	–
219	171554	Hypo2	3.6	4.8	−33	–
220	179207	Hypo2	1.0	2.7	−39	–
221	212051	Hypo2	1.0	3.1	−22	–
222	240871	Hypo2	1.8	2.7	−21	–
223	246014	Hypo2	1.0	4.8	0	–
224	287977	Hypo2	144.5	2.8	−22	–
225	355457	Hypo2	1.5	1.1	−30	–
226	363250	Hypo2	2.8	6.0	0	–
227	378695	Hypo2	1.3	6.4	−39	–
228	405323	Hypo2	144.5	29.3	−29	–
229	653279	Hypo2	1.0	160.5	−22	–
Staurosporine[g]				–	100	0.068[d]

[a] Hits numbers are as in Fig. 6

[b] Predicted anti-Chk1 IC_{50} values based on each kNN-dbCICA model through Eq. (1)

[c] Not predicted as it failed to dock

[d] Standard positive control inhibitor [37]

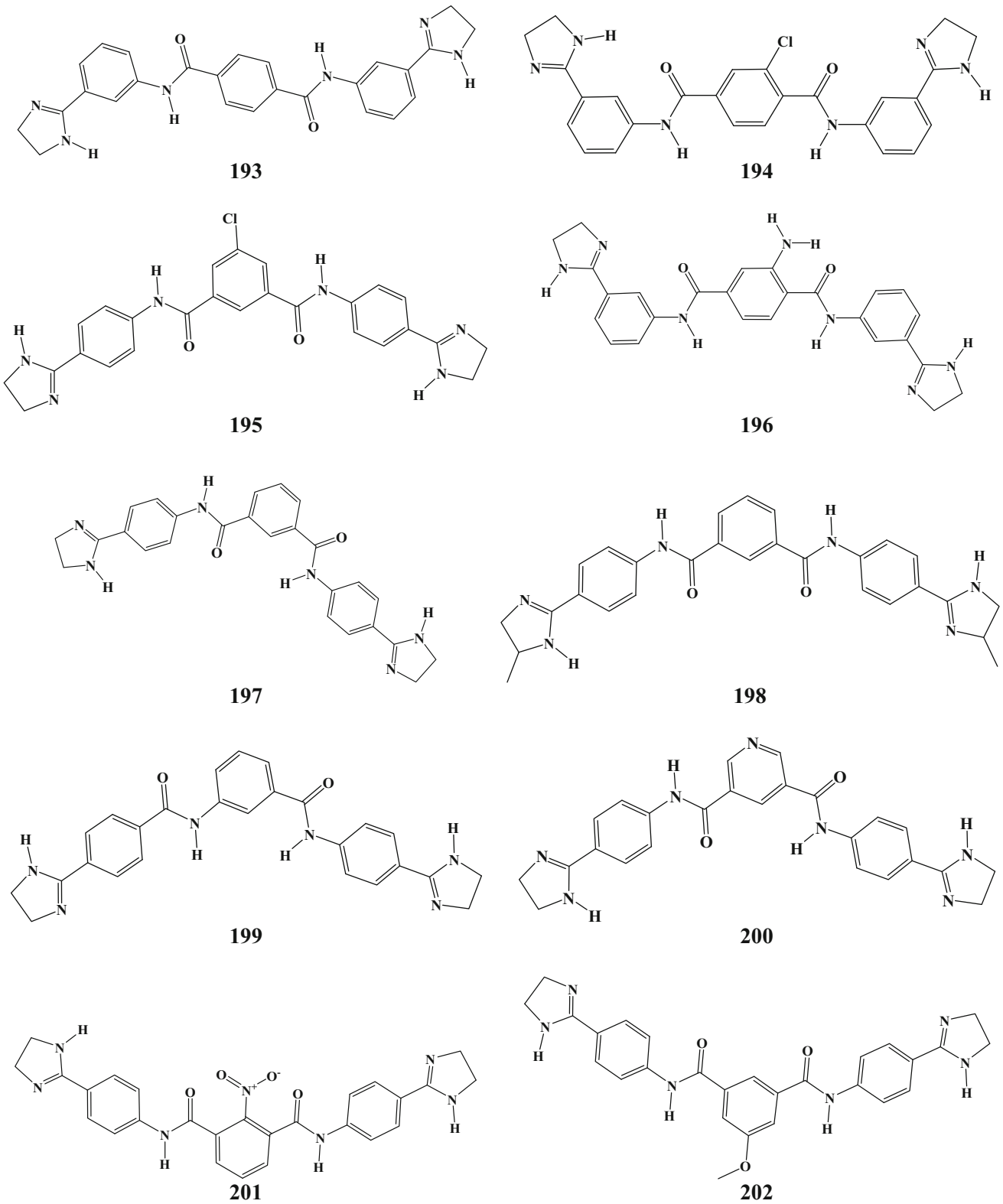
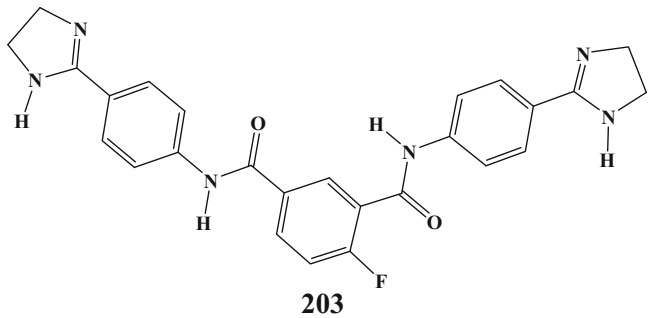
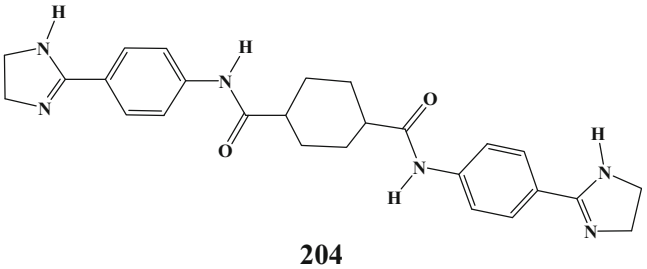


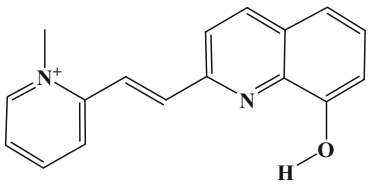
Fig. 6 The chemical structures of the tested highest ranking hits captured by pharmacophoric models 1 and 2



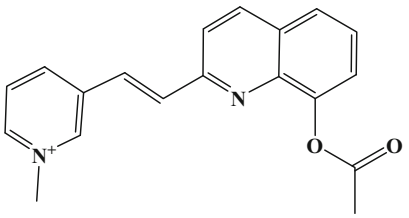
203



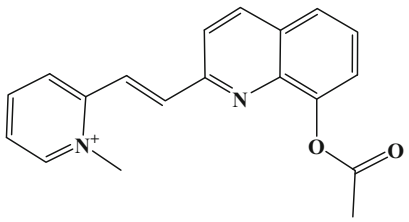
204



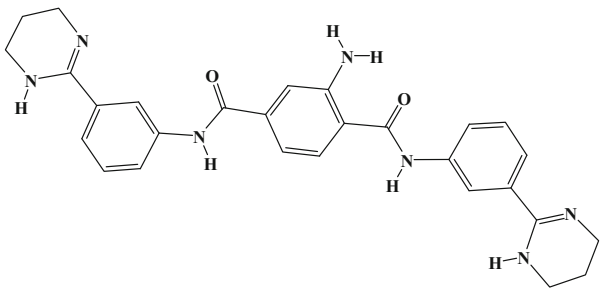
205



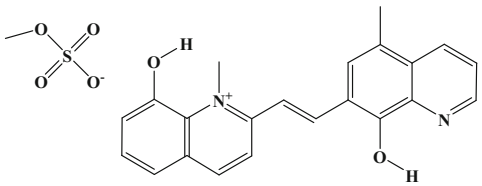
206



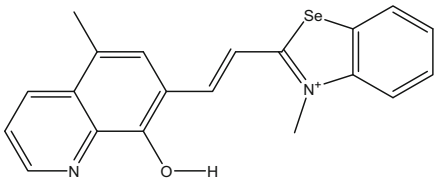
207



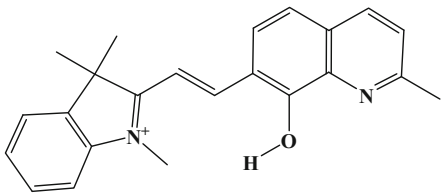
208



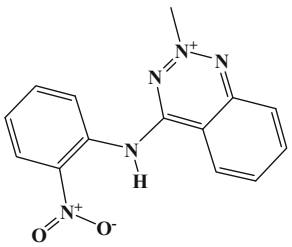
209



210



211



212

Fig. 6 continued

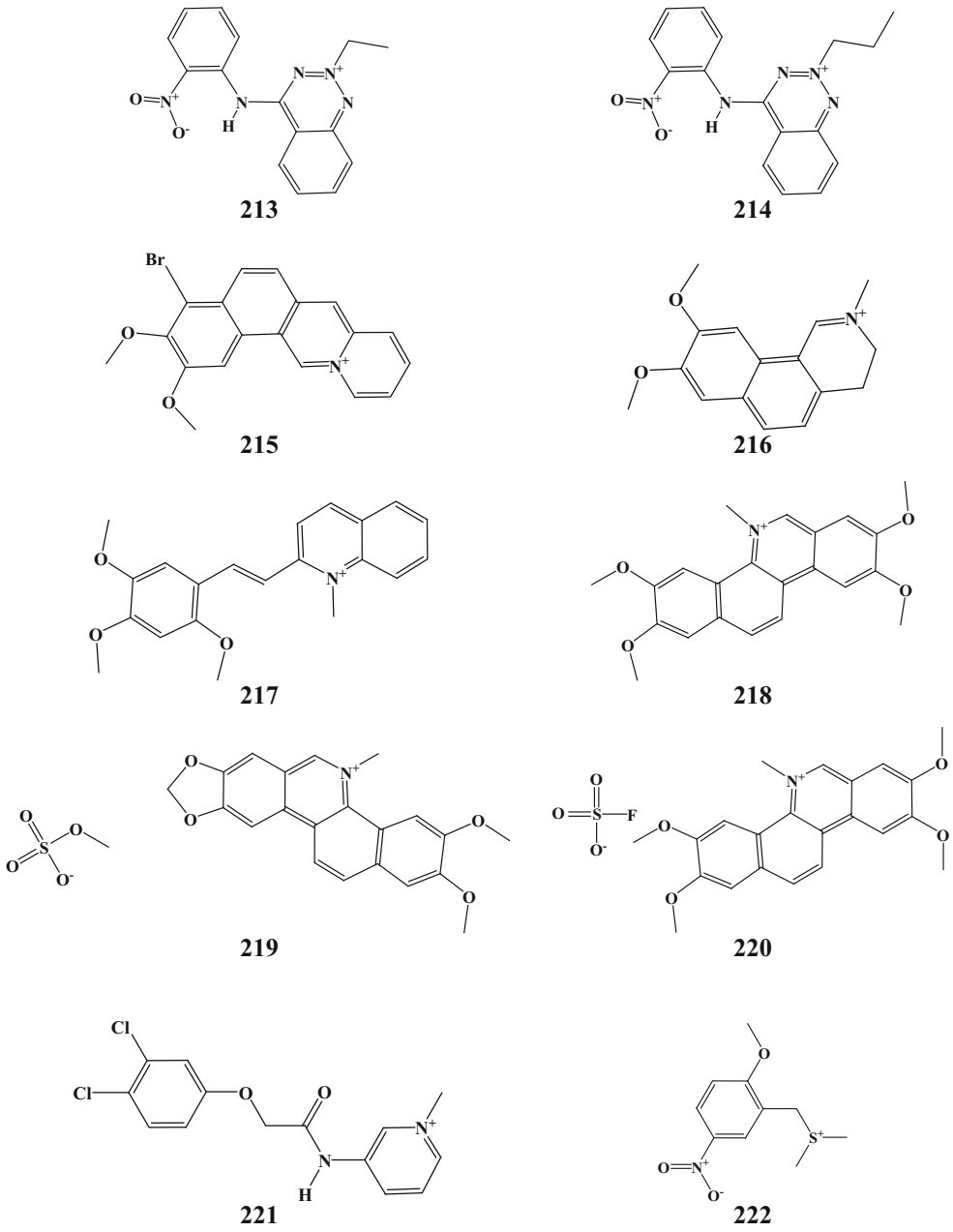
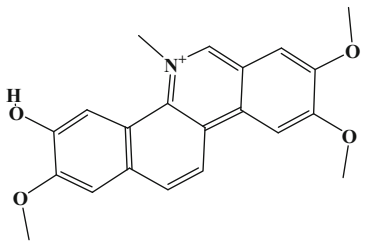
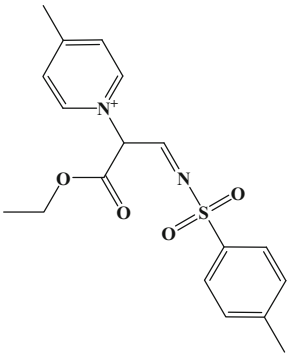


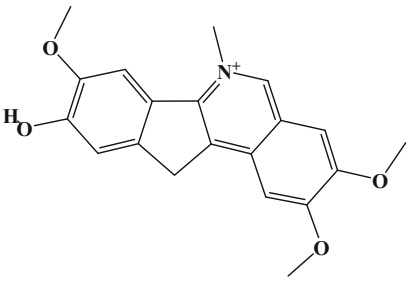
Fig. 6 continued



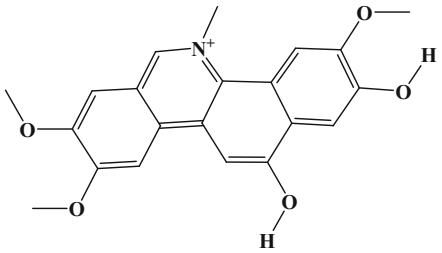
223



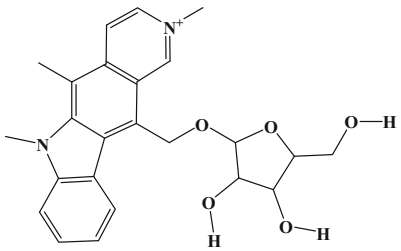
224



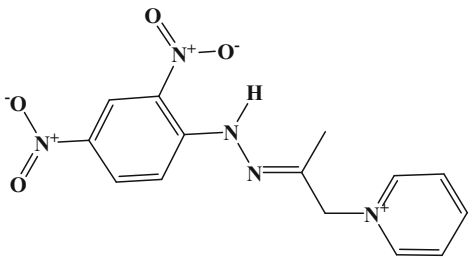
225



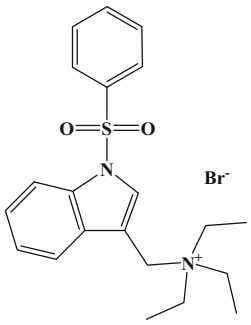
226



227



228



229

Fig. 6 continued

Validation of kNN-dbCICA models

Self docking

To evaluate the docking conditions proposed by kNN-dbCICA models 1 and 2 in Table 1, we compared the experimental crystallographic pose of a co-crystallized ligand within Chk1 (PDB code: 3TKI, Resolution 1.6 Å) with the corresponding docked poses of the ligand resulting from the docking/scoring conditions of kNN-dbCICA models (1) and (2). Figure 4 shows a comparison between the two cases. Clearly from the figure, both docking conditions closely reproduced the experimental co-crystallized pose of the ligand giving further confidence to our kNN-dbCICA modeling results, where the RMSD values between the co-crystallized ligand and the two poses were equal to 1.02 Å, i.e., in both models. This probably happened because both docking conditions converged on the same docked pose for the 3TKI ligand.

Receiver operating characteristic (ROC) curve analysis of Hypo1 and Hypo2

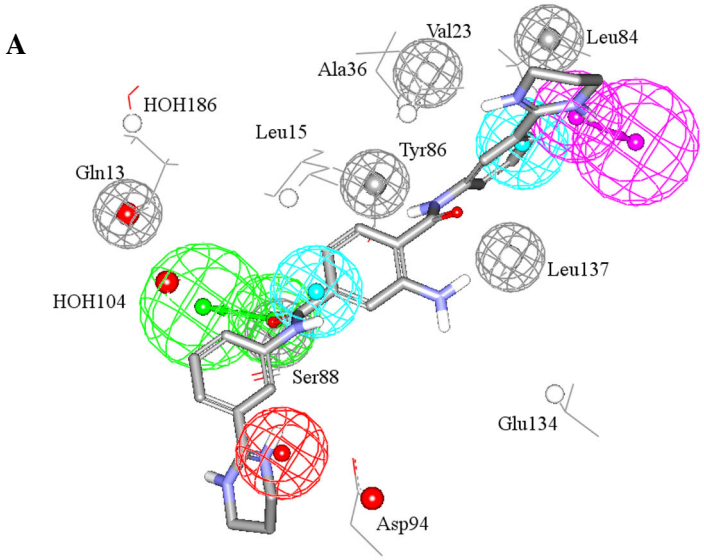
To further validate the resulting models (both kNN-dbCICA-based docking settings and corresponding pharmacophores), Hypo1 and Hypo2 were subjected to receiver-operating characteristic (ROC) analysis. In ROC analysis, the ability of a particular pharmacophore model to correctly rank a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the resulting ROC in addition to other parameters: overall accuracy, overall specificity, overall true positive rate and overall false negative rate (see Sect. 2.1.8 for more details). Figure 5 and Table 4 show the ROC performances of our dbCICA-based pharmacophores.

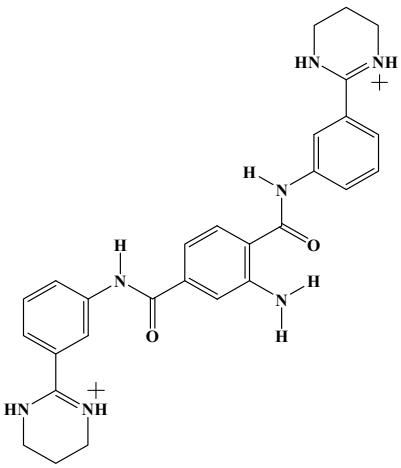
Looking at Table 4, it can be concluded that Hypo1 performs better than Hypo2 based on ROC-AUC, ACC, SPC, and FNR, except for TPR value which was slightly better for Hypo2. Still, both models illustrated excellent overall ROC profiles warranting their use as 3D search queries to mine the NCI database for new Chk1 inhibitors.

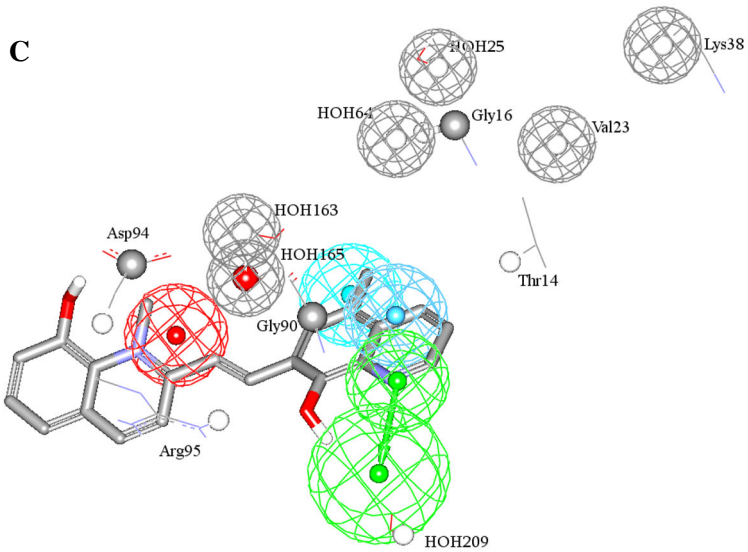
In silico screening of databases and subsequent in vitro bioassay

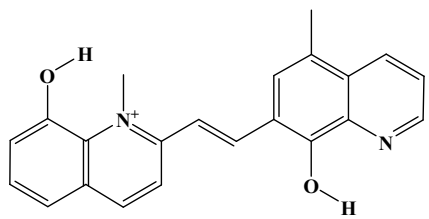
The ultimate validation of kNN-dbCICA methodology is to assess the ability of the pharmacophore models (Hypo1 and Hypo2) as successful virtual 3D search queries capable of catching new Chk1 inhibitors. Therefore, we employed Hypo1 and Hypo2 to search the national cancer institute (NCI) list of compounds (contains 238,819 compounds) for new Chk1 inhibitors. Captured hits were subsequently filtered by a molecular weight threshold of 500 dalton in order to remove large, non-drug-like compounds [35, 36]. The remaining hits were ranked according to their goodness-of-fit against their corresponding capturing pharmacophores (i.e., Hypo1 and Hypo2) and the best 37 were acquired from the NCI for in vitro evaluation against Chk1. We attempted to predict the bioactivities of these hits by docking them into Chk1 protein binding pocket (PDB code: 3TKI) employing the same docking conditions of kNN-dbCICA models 1 and 2 (Table 1) followed by analyzing the docked poses for critical contacts according to kNN-dbCICA models 1 and 2 (Table 2). However, some hits failed to dock within Chk1 binding site. We believe the main reason for this observation is related to the fact that pharmacophore features are expressed as spheres (not points) allowing certain margins of spatial tolerances for corresponding overlapping chemical features of molecules captured by virtual screening. That is, chemical features of hit molecules are tolerated to spatially fit within pharmacophoric spheres rather than pharmacophoric points. Tolerance spheres in pharmacophore models are intended to compensate for the conformational perturbations of protein receptors at physiological temperatures. Accordingly, it is not necessary for a particular hit compound captured by certain structure-based pharmacophore model to dock neatly within the corresponding crystallographic binding pocket of the receptor because pharmacophore modeling attempts to simulate different conformational states of the respective binding pocket, while the corresponding crystallographic structure is a rigid representation of the binding pocket in a single conformational state. Another problem that might cause this discrepancy, i.e., failure of pharmacophoric hits to dock into corresponding crystallographic binding pockets, is the fact that pharmacophore models lack proper presentation of the steric constrains of the binding pocket [16, 41–43], meaning that some hits are "larger" than the binding pocket despite being successfully fitted against the corresponding pharmacophores. This weakness can be remedied by decorating the pharmacophoric model with appropriate exclusion spheres to characterize the steric limits of the binding pocket. However, we decided to limit the number of exclusion spheres in the models to those suggested by optimal kNN-dbCICA models to provide balanced analysis of the capabilities of this novel approach, i.e., without implementing extra techniques to improve the performance of the resulting pharmacophores such as use of external algorithms to add more exclusion spheres.

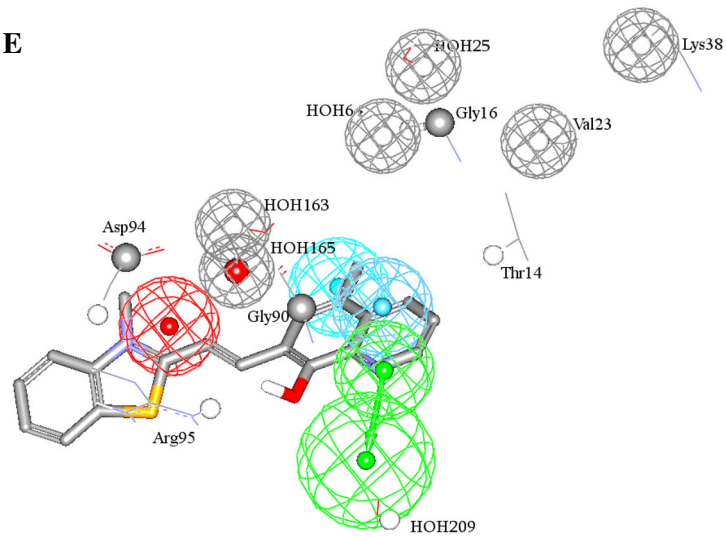
Table 5 and Fig. 6 show captured hits of best fits against corresponding pharmacophores (Hypo1 or Hypo2), their kNN-dbCICA based predictions, as well as their experimental in vitro bioactivities. Several hits showed

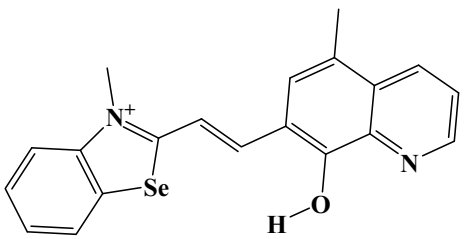

A
Val23
Leu84
Ala36
HOH186
Leu15
Tyr86
Gln13
Leu137
HOH104
Ser88
Glu134
Asp94


B
HN
NH
H
N
O
H
N
O
N
H
H
HN
NH


C
HOH25
Lys38
HOH64
Gly16
Val23
HOH163
Asp94
HOH165
Thr14
Gly90
Arg95
HOH209


D
H
O
N+
N
O
H


E
HOH25
Lys38
HOH64
Gly16
Val23
HOH163
Asp94
HOH165
Thr14
Gly90
Arg95
HOH209


F
N+
Se
N
H
O

◀**Fig. 7** **a**, **c**, **e** Hypo 1 and 2 fitted against active hits **208**, **209** and **210,** respectively, as they dock within the binding site of Chk1 (PDB code: 3TKI), **b**, **d** and **f** show the chemical structure of active hits **208**, **209** and **210**, respectively

"negative" inhibitory readouts at 10 μM probably due to bioassay artifacts. The fact that we implemented a FRET–based bioassay suggests that hit molecules of electron-deficient extended aromatic ring systems interfered in the fluorescence of assay fluorophores, thus causing "negative" inhibitory readouts. Such interferences are not uncommon in FRET-based bioassay results [45]. However, to avoid erroneously assigning bioactivities to these compounds, we decided not to evaluate their IC50 s.

Still, three hits showed significant anti-Chk1 properties at 10 μM, namely, **208**, **209**, and **210.** Subsequent follow up showed they have anti-Chk1 IC_{50} values ranging from 2.4-69.7 μM (NMR and mass spectroscopic charts of the compounds are shown in Supplementary Materials). It's noteworthy to mention that the two most active hits were captured by Hypo2 and many inactive hits failed to dock inside the active site of Chk1. Figure 7 shows the three anti-*C*hk1 hits **208**, **209** and **210**, and how they map Hypo1 and Hypo2 overlaid onto Chk1 binding site (showing critical contacts corresponding to models 1 and 2, Tables 1 and 2). Interestingly, the three active hits show similar interactions to those seen with training compounds as in Figs. 2 and 3 and related discussion.

Although **208**, **209** and **210** show mediocre anti-Chk1 bioactivities, i.e., compared to published inhibitors [44], their significance stems out from the fact that they present new chemical scaffolds as Chk1 inhibitors.

Interestingly, our assay conditions yielded an IC_{50} value of 68 nM for the standard pan kinase inhibitor staurosporine, which is reasonably comparable with the literature value for this inhibitor (i.e., 2.1 nM) [37].

Conclusion

As the understanding of signaling events at the molecular level leading to the checkpoint arrests during DNA damage started to emerge during the last few years, Chk1 has been considered as one of the most attractive targets for drug discovery of anticancer agents.

The drawbacks of the structure-based design combined with the inadequacies of corresponding validation methods prompted us to envisage a novel validation approach based on 3D-QSAR analysis, namely Docking-Based Comparative Intermolecular Contacts Analysis with the statistical tool k-Nearest Neighbour (kNN-dbCICA). In the current project we employed a wide range of docking configurations (7 scoring functions) to dock 192 published inhibitors into the binding pocket of Chk1 (PDB code: 3TKI). Moreover, the inhibitors were docked in their ionized and unionized forms into the hydrous and anhydrous versions of the binding pocket.

The novel methodology; kNN-dbCICA, was very useful in identifying and validating the optimal docking configurations. Interestingly, PMF04 and LigScore2 based scoring were superior over other scoring functions; ligand ionization and protein hydration seem to enhance the strength of the resulting kNN-dbCICA models.

Two docking configurations were found to achieve self-consistent kNN-dbCICA models that were consequently used to construct corresponding pharmacophoric hypotheses that were utilized to screen the NCI list of compounds. Experimental validation by in vitro assay using Chk1 kit proved anti Chk1 activities for three hit molecules, with the most active having IC_{50} of 2.4 μM.

Acknowledgments The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. The authors are also thankful to the National Cancer Institute for freely providing NCI hits. (Additional Supporting Information may be found in the online version of this article.).

References

1. Carrassa L, Damia G (2011) Unleashing Chk1 in cancer therapy. Cell Cycle 10:2121–2128. doi:10.4161/cc.10.13.16398
2. Wang GT, Li G, Mantei RA, Chen Z, Kovar P, Gu W, Xiao Z, Zhang H, Sham HL, Sowin T, Rosenberg SH, Lin NH (2005) 1-(5-Chloro-2-alkoxyphenyl)-3-(5-cyanopyrazin-2-yl)ureas [correction of cyanopyrazi] as potent and selective inhibitors of Chk1 kinase: synthesis, preliminary SAR, and biological activities. J Med Chem 48:3118–3121. doi:10.1021/jm048989d
3. Zhou BB, Bartek J (2004) Targeting the checkpoint kinases: chemosensitization versus chemoprotection. Nat Rev Cancer 4:216–225. doi:10.1038/nrc1296
4. Hirose Y, Berger MS, Pieper RO (2001) Abrogation of the Chk1-mediated G(2) checkpoint pathway potentiates temozolomide-induced toxicity in a p53-independent manner in human glioblastoma cells. Cancer Res 61:5843–5849
5. Powell SN, DeFrank JS, Connell P, Eogan M, Preffer F, Dombkowski D, Tang W, Friend S (1995) Differential sensitivity of p53(−) and p53(+) cells to caffeine-induced radiosensitization and override of G2 delay. Cancer Res 55:1643–1648
6. Kawabe T (2004) G2 checkpoint abrogators as anticancer drugs. Mol Cancer Ther 3:513–519
7. Koniaras K, Cuddihy AR, Christopoulos H, Hogg A, O'Connell MJ (2001) Inhibition of Chk1-dependent G2 DNA damage checkpoint radiosensitizes p53 mutant human cells. Oncogene 20:7453–7463. doi:10.1038/sj.onc.1204942
8. Wang H, Wang X, Zhou XY, Chen DJ, Li GC, Iliakis G, Wang Y (2002) Ku affects the ataxia and Rad 3-related/CHK1-dependent S phase checkpoint response after camptothecin treatment. Cancer Res 62:2483–2487
9. Chen Z, Xiao Z, Chen J, Ng SC, Sowin T, Sham H, Rosenberg S, Fesik S, Zhang H (2003) Human Chk1 expression is dispensable

for somatic cell death and critical for sustaining G2 DNA damage checkpoint. Mol Cancer Ther 2(6):543–548
10. Graves PR, Yu L, Schwarz JK, Gales J, Sausville EA, O'Connor PM, Piwnica-Worms H (2000) The Chk1 protein kinase and the Cdc25C regulatory pathways are targets of the anticancer agent UCN-01. J Biol Chem 275:5600–5605. doi:10.1074/jbc.275.8.5600
11. Taha MO, Habash M, Al-Hadidi Z, Al-Bakri A, Younis K, Sisan S (2011) Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. J Chem Inf Model 51:647–669. doi:10.1021/ci100368t
12. Taha MO, Habash M, Khanfar MA (2014) The use of docking-based comparative intermolecular contacts analysis to identify optimal docking conditions within glucokinase and to discover of new GK activators. J Comput Aided Mol Des 28:509–547. doi:10.1007/s10822-014-9740-4
13. Al-Sha'er MA, Taha MO (2012) Application of docking-based comparative intermolecular contacts analysis to validate Hsp90alpha docking studies and subsequent in silico screening for inhibitors. J Mol Model 18:4843–4863. doi:10.1007/s00894-012-1479-z
14. Sharaf MA, Illman DL, Kowalski BR (1986) Chemometrics. Wiley, New York
15. Zheng W, Tropsha A (2000) Novel variable selection quantitative structure—property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 40:185–194. doi:10.1021/ci980033m
16. Khanfar MA, Taha MO (2013) Elaborate ligand-based modeling coupled with multiple linear regression and k nearest neighbor QSAR analyses unveiled new nanomolar mTOR inhibitors. J Chem Inf Model 53(10):2587–2612. doi:10.1021/ci4003798
17. Tao ZF, Chen Z, Bui MH, Kovar P, Johnson E, Bouska J, Zhang H, Rosenberg S, Sowin T, Lin NH (2007) Macrocyclic ureas as potent and selective Chk1 inhibitors: an improved synthesis, kinome profiling, structure-activity relationships, and preliminary pharmacokinetics. Bioorg Med Chem Lett 17(23):6593–6601. doi:10.1016/j.bmcl.2007.09.063
18. Tao ZF, Li G, Tong Y, Stewart KD, Chen Z, Bui MH, Merta P, Park C, Kovar P, Zhang H, Sham HL, Rosenberg SH, Sowin TJ, Lin NH (2007) Discovery of 4′-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-benzonitriles and 4′-(1,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-pyridine-2′-carbonitriles as potent checkpoint kinase 1 (Chk1) inhibitors. Bioorg Med Chem Lett 17(21):5944–5951. doi:10.1016/j.bmcl.2007.07.102
19. Tao ZF, Li G, Tong Y, Chen Z, Merta P, Kovar P, Zhang H, Rosenberg SH, Sham HL, Sowin TJ, Lin NH (2007) Synthesis and biological evaluation of 4′-(6,7-disubstituted-2,4-dihydro-indeno[1,2-c]pyrazol-3-yl)-biphenyl-4-ol as potent Chk1 inhibitors. Bioorg Med Chem Lett 17(15):4308–4315. doi:10.1016/j.bmcl.2007.05.027
20. Tao ZF, Wang L, Stewart KD, Chen Z, Gu W, Bui MH, Merta P, Zhang H, Kovar P, Johnson E, Park C, Judge R, Rosenberg S, Sowin T, Lin NH (2007) Structure-based design, synthesis, and biological evaluation of potent and selective macrocyclic checkpoint kinase 1 inhibitors. J Med Chem 50(7):1514–1527. doi:10.1021/jm061247v
21. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 21:289–307. doi:10.1016/S1093-3263(02)00164-X
22. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure-Activity-relationships and quantitative structure-property relationships. J Chem Inf Comput Sci 34(4):854–866. doi:10.1021/Ci00020a020
23. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P (2004) Virtual screening using protein-ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci 44(3):793–806. doi:10.1021/ci034289q
24. Abu Hammad AM, Afifi FU, Taha MO (2007) Combining docking, scoring and molecular field analyses to probe influenza neuraminidase-ligand interactions. J Mol Graph Model 26(2):443–456. doi:10.1016/j.jmgm.2007.02.002
25. Abu-Hammad A, Zalloum WA, Zalloum H, Abu-Sheikha G, Taha MO (2009) Homology modeling of MCH1 receptor and validation by docking/scoring and protein-aligned CoMFA. Eur J Med Chem 44(6):2583–2596. doi:10.1016/j.ejmech.2009.01.031
26. Taha MO, AlDamen MA (2005) Effects of variable docking conditions and scoring functions on corresponding protein-aligned comparative molecular field analysis models constructed from diverse human protein tyrosine phosphatase 1B inhibitors. J Med Chem 48(25):8016–8034. doi:10.1021/jm058047o
27. Homans SW (2007) Water, water everywhere–except where it matters? Drug Discov Today 12(13–14):534–539. doi:10.1016/j.drudis.2007.05.004
28. Song CM, Lim SJ, Tong JC (2009) Recent advances in computer-aided drug design. Brief Bioinform 10(5):579–591. doi:10.1093/bib/bbp023
29. Jorgensen WL (2009) Efficient drug lead discovery and optimization. Acc Chem Res 42(6):724–733. doi:10.1021/ar800236t
30. Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. J Med Chem 49(20):5851–5855. doi:10.1021/jm060999m
31. Krissinel E (2010) Crystal contacts as nature's docking solutions. J Comput Chem 31(1):133–143. doi:10.1002/jcc.21303
32. Ni ZJ, Barsanti P, Brammeier N, Diebes A, Poon DJ, Ng S, Pecchi S, Pfister K, Renhowe PA, Ramurthy S, Wagman AS, Bussiere DE, Le V, Zhou Y, Jansen JM, Ma S, Gesner TG (2006) 4-(Aminoalkylamino)-3-benzimidazole-quinolinones as potent CHK-1 inhibitors. Bioorg Med Chem Lett 16(12):3121–3124. doi:10.1016/j.bmcl.2006.03.059
33. Foloppe N, Fisher LM, Howes R, Kierstan P, Potter A, Robertson AG, Surgenor AE (2005) Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein-ligand affinity. J Med Chem 48(13):4332–4345. doi:10.1021/jm049022c
34. Triballeau N, Bertrand H-O, Acher F (2006) Are you sure you have a good model? In: Hoffmann RD (ed) Langer T. Pharmacophores and Pharmacophore Searches, Wiley, pp 325–364
35. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1–3):3–26. doi:10.1016/S0169-409X(00)00129-0
36. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45(12):2615–2623. doi:10.1021/jm020017n
37. Reader JC, Matthews TP, Klair S, Cheung KM, Scanlon J, Proisy N, Addison G, Ellard J, Piton N, Taylor S, Cherry M, Fisher M, Boxall K, Burns S, Walton MI, Westwood IM, Hayes A, Eve P, Valenti M, de Haven Brandon A, Box G, van Montfort RL, Williams DH, Aherne GW, Raynaud FI, Eccles SA, Garrett MD, Collins I (2011) Structure-guided evolution of potent and selective CHK1 inhibitors through scaffold morphing. J Med Chem 54(24):8328–8342. doi:10.1021/jm2007326
38. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO (2005) Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem 48(7):2534–2547. doi:10.1021/jm049092j
39. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008) Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection–what can we learn from earlier mistakes? J Comput Aided Mol Des 22(3–4):213–228. doi:10.1007/s10822-007-9163-6

40. Dudkin VY, Rickert K, Kreatsoulas C, Wang C, Arrington KL, Fraley ME, Hartman GD, Yan Y, Ikuta M, Stirdivant SM, Drakas RA, Walsh ES, Hamilton K, Buser CA, Lobell RB, Sepp-Lorenzino L (2012) Pyridyl aminothiazoles as potent inhibitors of Chk1 with slow dissociation rates. Bioorg Med Chem Lett 22(7):2609–2612. doi:10.1016/j.bmcl.2012.01.110
41. Taha MO, Al-Sha'er MA, Khanfar MA, Al-Nadaf AH (2014) Discovery of nanomolar phosphoinositide 3-kinase gamma (PI3Kgamma) inhibitors using ligand-based modeling and virtual screening followed by in vitro analysis. Eur J Med Chem 84:454–465. doi:10.1016/j.ejmech.2014.07.056
42. Al-Sha'er MA, Khanfar MA, Taha MO (2014) Discovery of novel urokinase plasminogen activator (uPA) inhibitors using ligand-based modeling and virtual screening followed by in vitro analysis. J Mol Model 20(1):2080. doi:10.1007/s00894-014-2080-4
43. Abuhamdah S, Habash M, Taha MO (2013) Elaborate ligand-based modeling coupled with QSAR analysis and in silico screening reveal new potent acetylcholinesterase inhibitors. J Comput Aided Mol Des 27(12):1075–1092. doi:10.1007/s10822-013-9699-6
44. Matthews TP, Jones AM, Collins I (2013) Structure-based design, discovery and development of checkpoint kinase inhibitors as potential anticancer therapies. Expert Opin Drug Discov 8(6):621–640. doi:10.1517/17460441.2013.788496
45. Imbert P-E, Unterreiner V, Siebert D, Gubler H, Parker C, Gabriel D (2007) Recommendations for the reduction of compound artifacts in time-resolved fluorescence resonance energy transfer assays. Assay Drug Dev Technol 5(3):363–372. doi:10.1089/adt.2007.073