

Elaborate ligand-based modeling reveal new submicromolar Rho kinase inhibitors

Rand Shahin · Saja AlQtaishat · Mutasem O. Taha

Received: 17 April 2011 / Accepted: 1 December 2011
© Springer Science+Business Media B.V. 2011

Abstract Rho Kinase (ROCKII) has been recently implicated in several cardiovascular diseases prompting several attempts to discover and optimize new ROCKII inhibitors. Towards this end we explored the pharmacophoric space of 138 ROCKII inhibitors to identify high quality pharmacophores. The pharmacophoric models were subsequently allowed to compete within quantitative structure–activity relationship (QSAR) context. Genetic algorithm and multiple linear regression analysis were employed to select an optimal combination of pharmacophoric models and 2D physicochemical descriptors capable of accessing self-consistent QSAR of optimal predictive potential ($r_{77} = 0.84$, $F = 18.18$, $r_{LOO}^2 = 0.639$, r_{PRESS}^2 against 19 external test inhibitors = 0.494). Two orthogonal pharmacophores emerged in the QSAR equation suggesting the existence of at least two binding modes accessible to ligands within ROCKII binding pocket. Receiver operating characteristic (ROC) curve analyses established the validity of QSAR-selected pharmacophores. Moreover, the successful pharmacophores models were found to be comparable with crystallographically resolved ROCKII binding pocket. We employed the pharmacophoric models and associated QSAR equation to screen the national cancer institute (NCI) list of compounds. Eight submicromolar ROCKII inhibitors were identified. The most potent gave IC_{50} values of 0.7 and 1.0 μ M.

Keywords Rho kinase II · ROCK II · Pharmacophore modeling · Quantitative structure–activity relationship · In silico screening · Cardiovascular diseases

Introduction

Rho-kinases (or Rho associated protein kinases, ROCKs) are downstream targets of the Rho G-proteins. They are involved in the regulation of cytoskeletal reorganization and gene expression [1].

There are two isoforms of ROCK: ROCK I (or ROK β) and ROCK II (or ROK α) [2]. Both isoforms are highly homologous, sharing 65% homology in amino acid sequence and 92% homology in their kinase domains. However, although both isoforms are ubiquitously expressed, ROCK II is highly expressed in the brain and the heart, whereas ROCK I is expressed preferentially in the lung, liver, spleen, kidney and testis. Nevertheless, evidence of the functional differences between ROCK I and ROCK II is still lacking [3].

The catalytic kinase domain of the two ROCK isoforms is located at the amino terminus, followed by coiled-coil-forming region that encompasses the Rho-binding domain (RBD), which is followed in turn by pleckstrin-homology domain (PH). The cysteine-rich repeat domain (CRD) is present at the carboxyl terminus [4].

The carboxyl terminus of ROCKs folds back onto the kinase domain, thereby forming an auto-inhibitory loop that maintains the enzyme in an inactive state. Binding of active Rho to ROCKs disrupts this negative regulatory interaction resulting in activation of the enzyme in response to extracellular signals [4].

ROCK II is involved in the regulation of vascular tone, endothelial function, inflammation and remodeling of cardiac cells [4, 5]. In particular, abnormal activation of the

Electronic supplementary material The online version of this article (doi:10.1007/s10822-011-9509-y) contains supplementary material, which is available to authorized users.

R. Shahin · S. AlQtaishat · M. O. Taha (✉)
Department of Pharmaceutical Sciences, Faculty of Pharmacy,
University of Jordan, Amman, Jordan
e-mail: mutasem@ju.edu.jo

Rho/ROCK II pathway has been shown to play a role in hypertension [5].

ROCK II appears to increase the force and velocity of actinomyosin cross-bridging in smooth muscle and non-muscle cells by maintaining the phosphorylated state of myosin light chain (MLC). This is achieved through direct phosphorylation of MLC and via inhibiting myosin light chain phosphatase (MLCPh) [3].

Accordingly, inhibition of ROCK II should have beneficial effects in a variety of cardiovascular disorders including hypertension, atherosclerosis, ischemia–reperfusion injury, stroke, myocardial hypertrophy, heart failure, cardiac allograft vasculopathy and vein graft disease [6]. The therapeutic significance of ROCK II in cardiovascular medicine has recently attracted a great deal of attention [1].

The main focus of recent efforts towards the development of new ROCK II inhibitors concentrate on structure-based ligand design [7–10]. To date, several human ROCK II complexes have been resolved by X-ray crystallography and solution NMR, e.g., PDB codes: 2H9V, 2ROW, 2ROV, 2F2U [11–14]. However, crystallographic structures are limited by inadequate resolution [15] and crystallization-related artifacts of the ligand–protein complex [16–18]. Moreover, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substates [19].

The continued interest in designing new ROCK II inhibitors and lack of adequate ligand-based computer-aided drug discovery efforts and the significant induced fit flexibility observed for ROCK II [20] prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore(s) integrated within self-consistent QSAR model. This approach avoids the pitfalls of structure-based techniques; furthermore, the pharmacophore model(s) can be used as 3D search queries to discover new ROCKII inhibitory scaffolds. We previously reported the use of this approach towards the discovery of new inhibitory leads against glycogen synthase kinase-3 β , [21] bacterial MurF [22], protein tyrosine phosphatase [23], DPP IV, [24] hormone sensitive lipase [25], β -secretase [26], influenza neuraminidase [27], cholesteryl ester transfer protein [28], CDK1 [29], Hsp90 [30], estrogen receptor β , [31] β -D-Glucosidase, [32] β -D-Galactosidase [33] and glycogen phosphorylase [34].

We employed the CATALYST-HYPOGEN module of the software package Discovery Studio [35] to construct plausible binding hypotheses for a diverse list of ROCKII inhibitors [7–9]. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for an optimal QSAR that combine high-quality binding pharmacophores with other molecular descriptors capable of explaining bioactivity variation across a collection of diverse ROCK II inhibitors. The

optimal pharmacophores were subsequently used as 3D search queries to screen the national cancer institute (NCI) list of compounds for new ROCK II inhibitory leads.

CATALYST-HYPOGEN models drug-receptor interaction using information derived only from the ligand structure. It identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes (Hbic), positive and negative ionizable (PosIon and NegIon) groups and aromatic planes (RingArom). The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range. CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies [21–31, 36, 37].

Experimental

Molecular modeling

Software and hardware

The following software packages were utilized in the present research.

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.
- CERIOUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA [38].
- CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (<http://www.cambridgesoft.com>), USA.
- Discovery Studio 2.5, Accelrys Inc. (www.accelrys.com), USA [39].

Pharmacophore and QSAR modeling studies were performed using CATALYST (HYPOGEN module), CERIOUS2 software suites from Accelrys Inc. (San Diego, California, www.accelrys.com) and Discovery Studio. Structure drawing was performed employing ChemDraw Ultra 6.0 which was installed on a Pentium 4 PC.

Data set

The structures of ROCK II inhibitors (**1–138**, Tables A1 and A2 under supplementary material) were collected from recently published literature [7–9]. Despite that the collected inhibitors were gathered from three separate articles, they were bioassayed employing the same bioassay methodology. The bioactivities were expressed as the concentrations of the

test compounds that inhibited the activity of ROCK II by 50% (IC_{50} in μM). The logarithm of measured IC_{50} values were used in QSAR and pharmacophores analyses, thus approximately correlating the data linear to the free energy change (assuming $IC_{50} \approx K_i$).

In cases where IC_{50} is expressed as being higher than 10 μM (e.g., compounds **4**, **5**, **24**, **44** and **58**, see tables A1 and A2 under Supplementary Materials) it was assumed it equals 35 μM . This assumption is necessary to allow 3.5 log cycles separation between the most potent training compounds (i.e., **113** and **120**) and the least active training compounds (i.e., **4**, **5**, **24**, **44** and **58**). However, this assumption was applied only for pharmacophore modeling purposes since poorly active compounds are utilized by the CATALYST-HYPGEN algorithm in the subtractive phase to exclude non-discriminating pharmacophores (i.e., those that lack the ability to differentiate active training compounds from inactive ones). However, these compounds were excluded in QSAR modeling to avoid errors in statistical regression.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra and saved in MDL-molfile format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum using the molecular mechanics CHARMM force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for CATALYST conformational analysis. CATALYST automatically deals with different ligand states such as protonation states.

Conformational analysis

The molecular flexibilities of the collected compounds were taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1–138**, Tables A1 and A2 under Supplementary Material) was explored adopting the “best conformer generation” option within CATALYST [40] based on the generalized CHARMM force field implemented in the program. Default parameters were employed in the conformation generation procedure of training compounds, i.e., a conformational ensemble was generated with an energy threshold of 20 kcal/mol from the local minimized structure at which has the lowest energy level and a maximum limit of 250 conformers per molecule [40].

Pharmacophoric hypotheses generation

All 138 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an “Uncertainty” value of

3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor [41, 42]. Subsequently, six structurally diverse training subsets: sets **I**, **II**, **III**, **IV**, **V** and **VI** (in table B under supplementary material) were carefully selected from the collection for pharmacophore modeling. Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least three and a half logarithmic cycles. Lesser training lists could lead to chance correlation and thus faulty models.

The selected training sets were utilized to conduct 36 modeling runs to explore the pharmacophoric space of ROCK II inhibitors (table C under supplementary material). The exploration process included altering interfeature spacing parameter (100 and 300 pm) and the maximum number of allowed features in the resulting pharmacophore hypotheses, i.e., they were allowed to vary from 4 to 5 for first and second runs and from 5 to 5 for third and fourth runs of each training set, as shown in table C under supplementary material.

Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, subtractive phase and optimization phase (see CATALYST Modeling Algorithm under section SM-1 in Supplementary Materials) [40–45].

Assessment of the generated hypotheses

When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: Weight cost, Error cost and Configuration cost (see CATALYST Cost Analysis in Assessment of Generated Binding Hypotheses in section SM-2 under Supplementary Materials).

An additional approach to assess the quality of CATALYST-HYPOGEN pharmacophores is to cross-validate them using the Cat-Scramble program implemented in CATALYST. This validation procedure is based on Fischer’s randomization test [46]. In this validation test; we selected a 90% confidence level, which instruct CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST-HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses. Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models [40, 46, 47].

Clustering of the generated pharmacophore hypotheses

The successful models (335) were clustered into 69 groups utilizing the hierarchical average linkage method available

in CATALYST. Subsequently, the highest-ranking representatives, as judged based on their significance r^2 -values, were selected to represent their corresponding clusters in subsequent QSAR modeling. Table 1 shows the statistical criteria of representative pharmacophores including their pharmacophoric features, success criteria and differences from corresponding null hypotheses. The table also shows the corresponding Cat. Scramble confidence levels for each representative pharmacophore.

QSAR modeling

A subset of 96 compounds from the total list of inhibitors (1–138) was utilized as a training set for QSAR modeling. Only compounds of well-defined IC_{50} values were considered in QSAR modeling. However, since it is essential to assess the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining 19 molecules (ca. 20% of the dataset) were employed as an external test subset for validating the QSAR models. The test molecules were selected as follows: collected inhibitors of defined IC_{50} values (Tables A1 and A2 in Supplementary Materials) were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set.

The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations in SD format. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, energies of the lowest and highest occupied molecular orbitals, etc.) [48]. Furthermore, the training compounds were fitted (using the Best-fit option in CATALYST) against the representative pharmacophores (69 models, Table 1), and their fit values were added as additional descriptors. The fit value for any compound is obtained automatically via equation (D) (under SM-1 in Supplementary Materials) [40].

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable

of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF) [48]. However, to avoid overwhelming GFA-MLR with large number of poor descriptors; we removed 10% of those showing lowest variance prior to QSAR analysis.

Our preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic and spline equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30,000 and lack-of-fit (LOF) smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 5 to 25 explanatory terms.

All QSAR models were validated employing leave one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS}) and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2_{PRESS} is defined as:

$$r^2_{PRESS} = SD - PRESS/SD \quad (1)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.

Receiver operating characteristic (ROC) curve analysis

QSAR-selected pharmacophore models (i.e., Hypo4/15 and Hypo 6/35) were validated by assessing their abilities to selectively capture diverse ROCK II active compounds from a large testing list of actives and decoys.

The testing list was prepared as described by Verdonk and co-workers [49, 50]. Briefly, decoy compounds were selected based on three basic one-dimensional (1D) properties that allow the assessment of distance (D) between two molecules (e.g., i and j): (1) the number of hydrogen-bond donors (NumHBD); (2) number of hydrogen-bond acceptors (NumHBA) and (3) count of nonpolar atoms (NP, defined as the summation of Cl, F, Br, I, S and C atoms in a particular molecule). For each active compound in the test set, the distance to the nearest other active compound is assessed by their Euclidean Distance Eq. 2:

$$D(i,j) = \sqrt{(\text{NumHBD}_i - \text{NumHBD}_j)^2 + (\text{NumHBA}_i - \text{NumHBA}_j)^2 + (\text{NP}_i - \text{NP}_j)^2} \quad (2)$$

Table 1 Success criteria of representative pharmacophoric hypotheses (cluster centers)

RUN ^a	Hypotheses ^b	Features	Cost				R^d	Global R^e	F -statistic ^f	Cat. scramble (%)
			Config.	Total	Null	Residual ^c				
1	3	HBDN, Hbic, 2×RingArom	14.1	108.5	125.4	16.9	0.85	0.09	14.2	95
	10	HBA, HBDN, Hbic, RingArom	14.1	110.6	125.4	14.8	0.82	0.03	4.6	95
2	4	HBA, HBDN, Hbic, RingArom	10.9	102.4	125.4	23	0.9	0.18	30.3	95
	8	HBA, HBDN, Hbic, RingArom	10.9	106.6	125.4	18.8	0.83	0.22	38.6	95
	10	3×HBDN, Hbic,	10.9	106.7	125.4	18.8	0.83	0.18	30.6	95
3	3	HBDN, 3×Hbic, RingArom	9.2	110.8	125.4	14.6	0.78	0.06	9	95
	4	2×HBDN, 3×Hbic	9.2	112.8	125.4	12.6	0.77	0.13	20.7	90
	6	HBA, HBDN, 3×Hbic	9.2	113.3	125.4	12.2	0.76	0.14	21.4	90
5	4	2×HBDN, Hbic, RingArom	12.5	108.6	125.4	16.9	0.83	0.27	49.6	90
	8	2×HBDN, 2×Hbic	12.5	109.8	125.4	15.6	0.83	0.22	37.8	90
6	9	HBA, 2×HBDN, Hbic	9.6	106.9	125.4	18.6	0.82	0.28	53.1	95
7	2	2×HBDN, 3×Hbic	7.4	104.9	125.4	20.6	0.83	0.11	17.3	90
9	7	2×Hbic, RingArom, PosIon	17	109.2	186.6	77.4	0.97	0.23	41.5	95
	8	Hbic, 2×RingArom, PosIon	17	109.1	186.6	77.5	0.97	0.2	34.7	95
10	2	HBA, Hbic, RingArom, PosIon	15.1	104.7	186.6	81.8	0.98	0.23	40	95
	3	HBA, Hbic, RingArom, PosIon	15.1	104.9	186.6	81.7	0.98	0.26	47.9	95
	7	HBA, 2×Hbic, PosIon	15.1	106.3	186.6	80.3	0.97	0.25	44.6	95
	9	HBD, 2×RingArom, PosIon	15.1	106.6	186.6	80	0.97	0.22	39.4	95
	10	HBA, 2×RingArom, PosIon	15.1	106.8	186.6	79.8	0.97	0.23	40.2	95
12	2	2×HBD, 2×Hbic, PosIon	9.7	107.7	186.6	78.9	0.94	0.26	48	95
14	2	2×HBD, Hbic, PosIon	12.2	95.7	149.4	53.7	0.95	0.23	39.7	95
	3	HBA, 2×Hbic, PosIon	12.2	98.7	149.4	50.7	0.93	0.26	46.6	95
	7	HBA, 2×RingArom, PosIon	12.2	100.5	149.4	48.9	0.91	0.24	41.8	95
	4	HBD, 2×Hbic, RingArom, PosIon	12.7	93.5	149.4	55.9	0.96	0.28	53.3	95
15	7	HBD, 3×Hbic, PosIon	12.7	95.8	149.4	53.6	0.96	0.22	38.8	95
	5	HBA, HBD, 3×Hbic	3.7	102	149.4	47.4	0.83	0.16	25.6	95
16	6	HBA, HBD, 2×Hbic, PosIon	3.7	102.6	149.4	46.8	0.83	0.31	61.8	95
	3	2×HBD, Hbic, PosIon	16.1	94.8	113.9	19.1	0.95	0.11	16.8	95
17	7	HBA, HBD, Hbic, PosIon	16.1	95.1	113.9	18.9	0.95	0.28	52.9	95
	8	HBA, 2×HBD, PosIon	16.1	95.2	113.9	18.7	0.95	0.09	12.6	95
	1	HBA, HBD, Hbic, PosIon	13.8	92.1	113.9	21.8	0.96	0.09	14.1	95
	4	HBA, HBD, Hbic, PosIon	13.8	92.1	113.9	21.8	0.96	0.07	10.2	95
18	7	HBA, HBD, RingArom, PosIon	13.8	92.4	113.9	21.5	0.96	0.19	30.9	95
	8	HBA, HBD, RingArom, PosIon	13.8	92.6	113.9	21.3	0.96	0.33	65.8	95
	1	HBD, 2×Hbic, RingArom, PosIon	13.7	94	113.9	0	0.94	0.29	54.9	95
	10	HBD, 2×Hbic, RingArom, PosIon	13.7	89.6	113.9	19.9	0.92	0.22	38.5	95
19	4	HBA, HBD, 2×Hbic, PosIon	8.6	89.7	113.9	24.3	0.93	0.21	36.4	95
	6	HBA, HBD, 2×Hbic, PosIon	8.6	94.2	113.9	0	0.87	0.14	21.3	95
	9	HBA, HBD, 2×Hbic, PosIon	8.6	96.4	113.9	24.2	0.85	0.27	51.5	95
22	3	2×HBD, Hbic, PosIon	11.9	91.5	113.9	19.8	0.94	0.04	6.2	95
23	8	2×HBD, 2×Hbic, PosIon	11.4	97	113.9	17.5	0.89	0.22	38.1	95
24	3	2×HBD, 2×Hbic, PosIon	2.6	98.5	113.9	22.4	0.77	0.05	6.6	95
	6	2×HBD, 3×Hbic	2.6	109.6	113.9	16.9	0.59	0.01	1	95
25	1	HBD, 2×Hbic, PosIon	15.8	93.7	161.6	67.8	0.99	0.2	34.6	95
	5	HBA, HBD, RingArom, PosIon	15.8	96.2	161.6	65.4	0.97	0.29	54.1	95
	8	HBD, Hbic, RingArom, PosIon	15.8	96.8	161.6	64.8	0.97	0.19	31.7	95

Table 1 continued

RUN ^a	Hypotheses ^b	Features	Cost				R^d	Global R^e	F -statistic ^f	Cat. scramble (%)
			Config.	Total	Null	Residual ^c				
26	2	HBA, Hbic, RingArom, PosIon	12.9	92.4	161.6	69.2	0.98	0.28	51.7	95
	9	HBA, HBD, RingArom, PosIon	12.9	94.6	161.6	67	0.96	0.25	45.5	95
	10	2×HBD, Hbic, PosIon	12.9	94.8	161.6	66.8	0.96	0.21	36.3	95
27	9	HBD, 2×Hbic, RingArom, PosIon	14.1	96.3	161.6	65.2	0.97	0.26	48.7	95
28	8	HBA, HBD, 2×Hbic, PosIon	6	98.3	161.6	63.3	0.91	0.31	62.1	95
29	3	HBAF, 2×RingArom, PosIon	17.8	91	117.8	26.8	0.94	0.28	52	95
	5	2×HBAF, RingArom, PosIon	17.8	91.1	117.8	26.7	0.92	0.11	16.9	95
	8	2×HBAF, RingArom, PosIon	17.8	91.9	117.8	25.9	0.91	0.09	13.9	95
	9	HBDN, 2×RingArom, PosIon	17.8	92	117.8	25.8	0.91	0.23	39.6	95
	10	HBAF, HBDN, RingArom, PosIon	15.8	87.5	117.8	30.3	0.94	0.26	47.8	95
30	2	3×HBAF, Hbic	15.8	86.8	117.8	31	0.95	0.06	8.5	95
	4	HBAF, HBDN, Hbic, RingArom	15.8	86.8	117.8	30.9	0.95	0.21	35.4	95
	7	2×HBAF, Hbic, PosIon	15.8	87.1	117.8	30.6	0.94	0.11	16.9	95
31	3	HBAF, 2×Hbic, RingArom, PosIon	15.5	89	117.8	28.8	0.92	0.31	60.4	95%
	8	HBDN, 2×Hbic, RingArom, PosIon	15.5	91	117.8	26.7	0.9	0.22	39.2	95
	10	2×HBDN, 2×Hbic, PosIon	15.5	92	117.8	25.8	0.9	0.23	40.7	95
32	1	HBAF, HBDN, 2×Hbic, PosIon	11.4	88.1	117.8	29.7	0.89	0.28	51.9	95
	8	2×HBAF, 2×Hbic, PosIon	11.4	91.6	117.8	26.1	0.86	0.07	10.3	95
33	1	HBAF, Hbic, RingArom, PosIon	13.3	80.4	126.2	45.8	0.98	0.31	60.6	95
34	5	HBAF, Hbic, RingArom, PosIon	24.9	92.2	126.2	33.9	0.98	0.3	59	95%
35	6	HBAF, Hbic, HbicArom, PosIon	11.4	80.5	126.2	45.6	0.97	0.23	40.6	95
	10	HBAF, Hbic, HbicArom, PosIon	11.4	81.3	126.2	44.8	0.96	0.23	41.1	95

^a Correspond to runs in Table C under supplementary materials

^b High ranking representative hypotheses (in their corresponding clusters, see “Clustering of the generated pharmacophore hypotheses”)

^c Difference between total cost and the cost of the corresponding null hypotheses

^d Correlation coefficients between pharmacophore-based bioactivity estimates (calculated from equation (C) in SM-1 under Supplementary Materials) and bioactivities of corresponding training compound (subsets in table B under supplementary material)

^e Correlation coefficients between pharmacophore-based bioactivity estimates and bioactivities of all collected compounds

^f Fisher statistic calculated based on the linear regression between the fit values of all collected inhibitors (1–138, table A under supplementary material) against pharmacophore hypothesis (employing the “best fit” option and equation (D)) and their respective anti-ROCK II bioactivities (log (1/IC₅₀) values)

^g Ranking of hypotheses is as generated by CATALYST in each automatic run

^h Bolded pharmacophores appeared in the best QSAR equations

The minimum distances are then averaged over all active compounds (Dmin). Subsequently, for each active compound in the test set, around 30 decoys were randomly chosen from the ZINC database [51]. The decoys were selected in such a way that they did not exceed Dmin distance from their corresponding active compound.

To diversify active members in the list, we excluded any active compound having zero distance ($D(i,j)$) from other active compound(s) in the test set. Active testing compounds were defined as those possessing ROCK II affinities ranging from 0.01 to 2.2 μ M. The test set included 18 active compounds and 642 ZINC decoys.

The test set (660 compounds) was screened by each particular pharmacophore employing the “Best flexible search” option implemented in CATALYST, while the conformational spaces of the compounds were generated employing the “Fast conformation generation option” implemented in CATALYST. Compounds missing one or more features were discarded from the hit list. In silico hits were scored employing their fit values as calculated by Eq. D) in Supplementary Materials.

The ROC curve analysis describes the sensitivity (Se or true positive rate, Eq. 3) for any possible change in the

number of selected compounds (n) as a function of $(1-Sp)$. Sp is defined as specificity or true

$$Se = \frac{\text{Number of Selected Actives}}{\text{Total Number of Actives}} = \frac{TP}{TP + FN} \quad (3)$$

negative rate Eq. 4 [50, 52].

$$Sp = \frac{\text{Number of Discarded Inactives}}{\text{Total Number of Inactives}} = \frac{TN}{TN + FP} \quad (4)$$

where, TP is the number of active compounds captured by the virtual screening method (true positives), FN is the number of active compounds discarded by the virtual screening method, TN is the number of discarded decoys (presumably inactives), while FP is the number of captured decoys (presumably inactive) [50, 52].

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 [50, 52]. 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 [50, 52]. 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.

The ROC curve representing ideal distributions, where no overlap between the scores of active molecules and decoys exists, proceeds from the origin to the upper-left corner until all the actives are retrieved and Se reaches the value of 1. In contrast to that, the ROC curve for a set of actives and decoys with randomly distributed scores tends towards the $Se = 1-Sp$ line asymptotically with increasing number of actives and decoys [50, 52]. The success of a particular virtual screening workflow can be judged from the following criteria (shown in Table 3):

1. Area under the ROC curve (AUC) [50, 52, 53]. In an optimal ROC curve an AUC value of 1 is obtained; however, random distributions cause an AUC value of 0.5. Virtual screening that performs better than a random discrimination of actives and decoys retrieve an AUC value between 0.5 and 1, whereas an AUC

value lower than 0.5 represents the unfavorable case of a virtual screening method that has a higher probability to assign the best scores to decoys than to actives [50, 52].

2. Overall Accuracy (ACC): describes the percentage of correctly classified molecules by the screening protocol Eq. 5. Testing compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) [50, 52, 53].

$$ACC = \frac{TP + TN}{N} = \frac{A}{N} \cdot Se + \left(1 - \frac{A}{N}\right) \cdot Sp \quad (5)$$

where, N is the total number of compounds in the testing database, A is the number of true actives in the testing database.

3. Overall specificity (SPC): describes the percentage of discarded inactives by the particular virtual screening workflow. Inactive test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [50, 52, 53].
4. Overall True Positive Rate (TPR or overall sensitivity): 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 [50, 52, 53].
5. Overall False Negative Rate (FNR or overall percentage of discarded actives): describes the fraction percentage of active compounds discarded by the virtual screening method. Discarded active test compounds are assigned a binary score value of zero (compound not captured) or one (compound captured) regardless to their individual fit values [50, 52, 53].

In silico screening for new ROCKII inhibitors

Hypo4/15 and Hypo6/35 were employed as 3D search queries to screen the National Cancer Institute (NCI) 3D flexible structural database. The screening was done employing the “Best Flexible Database Search” option implemented within CATALYST. Captured hits were filtered according to Lipinski’s [54] and Veber’s [55] rules. Remaining hits were fitted against the two pharmacophores using the “best fit” option within CATALYST. The fit values together with the relevant molecular descriptors of each hit were substituted in the optimal QSAR Eq. 6. The highest ranking molecules based on QSAR predictions were acquired and tested in vitro. Table 5 shows active hits and their QSAR-predictions and experimental bioactivities.

In vitro experimental studies

Materials

All of the chemicals used in these experiments were of reagent grade and obtained from commercial suppliers. NCI samples were kindly provided by the national cancer institute. Rho-kinase drug discovery kit was purchased from Cyclex (Japan), the standard inhibitor Y-27632, water and DMSO for bioanalysis were all purchased from Sigma-Aldrich (USA).

Preparation of hit compounds for in vitro assay

The tested compounds were provided as dry powders in variable quantities (5–10 mg). They were initially dissolved in DMSO to give stock solutions of fixed concentrations. Subsequently, they were diluted to the required concentrations with deionized water for bioassay.

Quantification of Rho-kinase activity in a spectrophotometric assay

The activity of the in silico hits were quantified by rho-kinase drug discovery kit (Cyclex, Japan). The 96-well plate of the kit is pre-coated with recombinant C terminus of the myosin-binding subunit of myosin phosphatase (MBS) phosphorylated by rho-kinase (at threonine-696). The detector antibody is conjugated to horseradish peroxidase and specifically detects the phosphorylated form of MBS.

To perform the assay, the tested hits were first diluted with the kinase buffer provided with the kit. The hits' solutions were then pipetted into assay wells to yield final concentrations of 0.1–10 μM . Subsequently, rho-kinase was added to each well as aqueous solution (0.01 units in 10 μL) and allowed to phosphorylate the bound substrate in the presence of Mg^{2+} and ATP. The amount of phosphorylated substrate was measured by binding it with the detector antibody conjugate to horseradish peroxidase, which then catalyzes the conversion of the chromogenic substrate tetra-methylbenzidine (TMB) from a colorless solution to a blue solution. The color is quantified by spectrophotometry and reflects the relative activity of rho-kinase. Samples' absorbances were determined at a wavelength of 650 nm using a plate reader (Bio-Tek instruments ELx 800, USA). Inhibition of Rho-kinase was calculated as percent activity of the uninhibited Rho-kinase enzyme control. Y-27632 was tested as positive control. Negative controls were prepared by adding the enzyme to the reaction with the kinase reaction buffer only [56].

Results and discussion

Exploration of ROCKII pharmacophoric space

The fact that we have an informative list of 138 ROCKII inhibitors of evenly spread bioactivities over more than 3.5 orders of magnitude, prompted us to employ CATALYST-HYPOGEN to identify possible pharmacophoric binding modes assumed by ROCKII inhibitors [7–9].

HYPOGEN implements an optimization algorithm that evaluates large number of potential binding models for a particular target through fine perturbations to hypotheses that survived the constructive and subtractive phases of the modeling algorithm (see section SM-1 Pharmacophoric Hypotheses Generation in Supplementary Materials) [57]. The extent of the evaluated pharmacophoric space is reflected by the configuration (Config.) cost calculated for each modeling run. It is generally recommended that the Config. cost of any HYPOGEN run not to exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST) to guarantee thorough analysis of all models [58]. The size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features and other CATALYST control parameters [58].

Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization via allowing effective evaluation of limited number of pharmacophoric models. On the other hand, rigorous restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses, as they might occur outside the “boundaries” of the pharmacophoric space.

Moreover, The fact that pharmacophore modeling requires limited number of carefully selected training compounds (from 16 to 45 compounds only) [40] that exhibit bioactivity variations attributable solely to the presence or absence of pharmacophoric features, i.e., not due to steric or electronic factors, makes it impossible to explore the pharmacophore space of large training sets in one shot (e.g., 138 compounds), partly because CATALYST-HYPOGEN is not suited to handle large number of compounds and partly because pharmacophore modeling is generally confused by electronic and steric bioactivity modifying factors commonly encountered in SAR data. This dilemma prompted us to break the collected compounds into smaller training subsets compatible with pharmacophore modeling, i.e., of bioactivity variations attributable solely to the presence or absence of pharmacophoric features (3D SAR). Nevertheless, the basic problem in this approach is to identify a particular training set capable of representing the whole list of collected compounds. This problem can be very significant in cases of large SAR lists, as in our case. We found that the best

way to solve this problem is by exploring the pharmacophoric space of several carefully selected training subsets, i.e., from the whole list of collected compounds, followed by allowing the resulting pharmacophores to compete within the context of GFA-QSAR analysis such that the best pharmacophore(s) that are capable of explaining bioactivity variations across the whole list of collected compounds is selected. However, since pharmacophore models fail in explaining electronic and steric bioactivity-modulating effects, the GFA-QSAR process should be allowed to select other 2D physicochemical descriptors to complement the selected pharmacophore(s). We have applied this approach in several previous publications [21–31].

Therefore, we decided to explore the pharmacophoric space of ROCKII inhibitors under reasonably imposed “boundaries” through 36 HYPOGEN automatic runs and employing six carefully selected training subsets: subsets I–VI in table B under supplementary material. The training compounds in these subsets were selected in such a way to guarantee maximal 3D diversity and continuous bioactivity spread over more than 3.5 logarithmic cycles. Moreover, training subsets were selected in such a way that their member compounds apparently share certain 3D SAR rules (by visual evaluation). We gave special emphasis to the 3D diversity of the most active compounds in each training subset (Table B under supplementary material) because of their significant influence on the extent of the evaluated pharmacophoric space during the constructive phase of HYPOGEN algorithm. However, it must be mentioned that not all collected compounds were incorporated in the pharmacophore training subsets, in fact, compounds that exhibit limited diversity or significant bioactivity-modifying steric or electronic influences were excluded from the training subsets (see “[Pharmacophoric hypotheses generation in experimental](#)” and SM-1 under Supplementary Materials) [57].

Guided by our rationally restricted pharmacophoric exploration concept, we restricted the software to explore pharmacophoric models incorporating from zero to one PosIon feature, from zero to three HBA, Hbic, and RingArom features instead of the default range of zero to five, as shown in Table C under supplementary material. Furthermore, we instructed HYPOGEN to explore only 4- and 5-featured pharmacophores, i.e., ignore models of lesser number of features in order to further narrow the investigated pharmacophoric space and to represent the feature-rich nature of known ROCK II ligands (as shown in Table C under supplementary material).

In each run, the resulting binding hypotheses were automatically ranked according to their corresponding “total cost” value, which is defined as the sum of error cost, weight cost and configuration cost (see “[Assessment of the generated hypotheses](#)” in Experimental and section

SM-2 under Supplementary Materials) [37, 40, 42, 57–59]. Error cost provides the highest contribution to total cost and it is directly related to the capacity of the particular pharmacophore as 3D-QSAR model, i.e., in correlating the molecular structures to the corresponding biological responses [40, 42, 57–59]. HYPOGEN also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 1) the more likely that the hypothesis does not reflect a chance correlation. An additional validation technique based on Fischer’s randomization test [46] was recently introduced into CATALYST: Cat. Scramble [40]. In this test the biological data and the corresponding structures are scrambled several times and the software is challenged to generate pharmacophoric models from the randomized data. The confidence in the parent hypotheses (i.e., generated from unscrambled data) is lowered proportional to the number of times the software succeeds in generating binding hypotheses from scrambled data of apparently better cost criteria than the parent hypotheses (see “[Assessment of the generated hypotheses](#)” in Experimental and section SM-2 under Supplementary Materials) [37, 40, 42, 57–59]. Fortunately, all generated pharmacophores illustrated $\geq 90\%$ Cat.Scramble significance.

Eventually, 336 pharmacophore models emerged from 36 automatic HYPOGEN runs. These were subsequently clustered and their best representatives (68 models, Table 1, see “[Clustering of the generated pharmacophore hypotheses](#)”) were used in subsequent QSAR modeling. Clearly from Table 1, the representative models shared comparable features and excellent statistical success criteria. Emergence of several statistically comparable pharmacophore models suggests the ability of ROCKII ligands to assume multiple pharmacophoric binding modes within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the binding process.

QSAR modeling

Despite that pharmacophoric hypotheses provide excellent insights into ligand-macromolecule recognition and can be used to mine for new biologically interesting scaffolds, their predictive value as 3D-QSAR models is limited by steric shielding and bioactivity-enhancing or reducing auxiliary groups [21–31]. This point combined with the fact that pharmacophore modeling of ROCKII inhibitors furnished several binding hypotheses of comparable success criteria prompted us to employ classical QSAR analysis to search for the best combination of

pharmacophore(s) and other 2D descriptors capable of explaining bioactivity variation across the collected inhibitors (tables A1 and A2 under Supplementary Materials). However, we removed compounds that were reported without specific IC₅₀ values (e.g., IC₅₀ > 10 μM), leaving a total of 96 compounds for QSAR modeling. We employed genetic function approximation and multiple linear regression QSAR (GFA-MLR-QSAR) analysis to search for an optimal QSAR equation(s).

The fit values obtained by mapping representative hypotheses (68 models) against modeled ROCK II inhibitors were enrolled, together with around 100 other physicochemical descriptors, as independent variables (genes) in GFA-MLR-QSAR analysis (see “[QSAR modeling in experimental](#)”) [21–31, 48, 60]. Moreover, we enrolled the multiplication products of different descriptors as additional explanatory variables to achieve self-consistent and predictive QSAR models.

However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 19 molecules (shown in table A2 under Supplementary Materials) and employed them as external test molecules for validating the QSAR models (r_{PRESS}^2). Figure B under Supplementary Materials shows the molecular similarity/diversity profiles among training and testing compounds. All QSAR models were cross-validated automatically using the leave-one-out cross-validation in CERIU2 [48, 60]. Equation 6 shows the details of the optimal QSAR model. Figure 1 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors.

$$\begin{aligned} \text{Log}(1/\text{IC}_{50}) = & 0.454 + 6.59 \times 10^{-4}(\text{Hypo6}/35)^3 \\ & + 3.72 \times 10^{-3}(\text{Hypo4}/15)^3 - 1.24 \\ & \times 10^{-7}(\text{Wiener})^2(\text{AtypeH53}) \\ & - 8.65 \times 10^{-2}(\text{JX})^2(\text{Hypo4}/15) \\ & - 9.60 \times 10^{-3}(\text{Hypo6}/35)^2(\text{AtypeH53}) \\ & - 8.25 \times 10^{-2}(\text{LUMO})(\text{JX})(\text{AtypeH53}) \\ & - 7.56 \times 10^{-2}(\text{JX})(\text{AtypeC24})(\text{AtypeN72}) \\ & + 1.50 \times 10^{-2}(\text{Hypo6}/35)(\text{AtypeC24}) \\ & (\text{AtypeH53}) + 1.03 \times 10^{-4}(\text{Wiener}) \\ & (\text{AtypeC24})(\text{AtypeN72}) \end{aligned}$$

$$r_{77} = 0.842, \quad F\text{-statistic} = 18.18, \quad r_{\text{LOO}}^2 = 0.639, \quad (6)$$

$$r_{\text{PRESS}(19)}^2 = 0.494$$

where, r_{77} is the correlation coefficient against 77 training compounds, r_{LOO}^2 is the leave-one-out correlation coefficient, and r_{PRESS}^2 is the predictive r^2 determined for the 19 test compounds [48, 60].

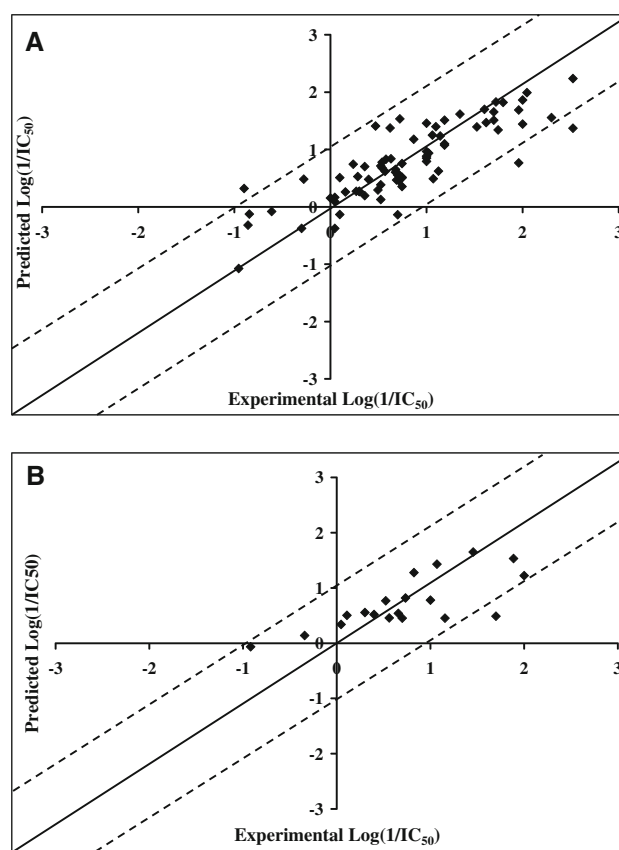


Fig. 1 Experimental versus **a** fitted (77 compounds, $r_{\text{LOO}}^2 = 0.639$), and **b** predicted (19 compounds, $r_{\text{PRESS}}^2 = 0.530$) bioactivities calculated from the best QSAR model Eq. 1. The *solid lines* are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the *dotted lines* indicate the 1.0 Log(1/IC₅₀) error margins

Hypo4/15 and Hypo6/35 represent the fit values of the training compounds against these two pharmacophores (bolded models in Table 1 and Figs. 2 and 3) as calculated from equation (D) under section SM-2 in Supplementary Materials. Atype descriptors are atom-type-based AlogP thermodynamic descriptors: AtypeN72 encodes for the number of amidic NH, AtypeC24 encodes for the number of aromatic CH atoms (excluding symmetrical ones), AtypeH53 encodes for the substitution patterns of alicyclic rings with 1,4-disubstituted cyclohexyls given the highest values while 1,3-disubstituted analogues assigned the least values. LUMO is the energy of the lowest unoccupied molecular orbital calculated by semiempirical quantum mechanical method (MOPAC). JX is a Balaban index related to the shape and atomic electronegativities of a particular compound. Wiener is the sum of the chemical bonds existing between all pairs of heavy atoms in the molecule. Table D under Supplementary Materials shows the values of the different descriptors in Eq. 6 for modeled

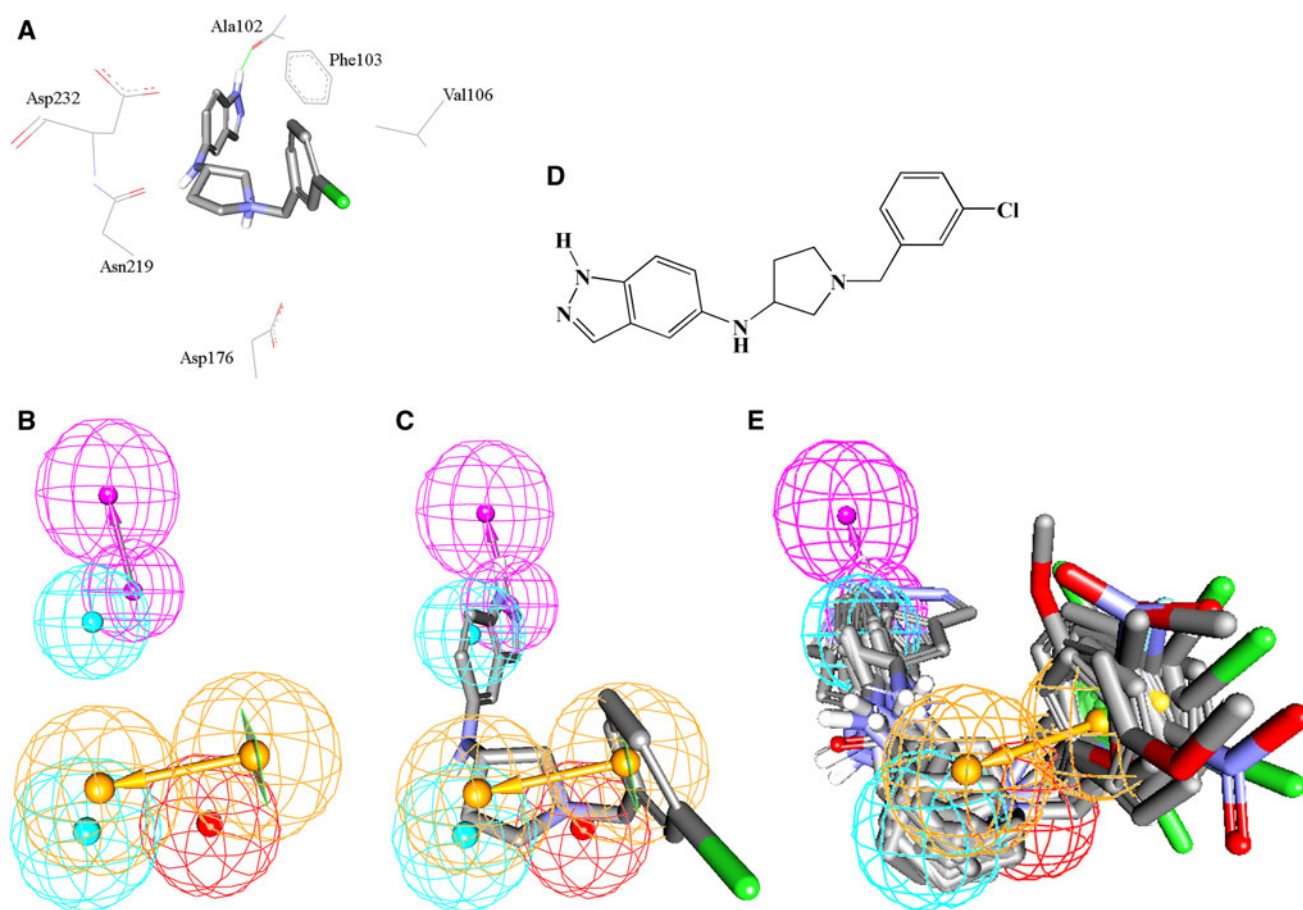


Fig. 2 **a** Docked structure of training compound **98** ($IC_{50} = 0.02 \mu\text{M}$, table A1 under Supplementary Materials) into ROCK II (PDB code: 2H9 V, resolution 1.85 Å). **b** Pharmacophoric features of Hypo4/15: HBD as *pink vectored spheres*, Hbic as *light blue spheres*, RingArom as *vectored orange spheres*, PosIon as *red*

spheres. **c** Hypo4/15 fitted against **98**, **d** Chemical structure of **98**. **e** Hypo4/15 fitted against a collection of training compounds, namely: **41, 43, 88, 96, 97, 98, 99, 101, 104, 105, 106, 107, 108, 109, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121** and **138**

compounds (Tables A1 and A2 under Supplementary Materials).

Emergence of two orthogonal pharmacophoric models, i.e., Hypo4/15 and Hypo6/35 of cross-correlation $r^2 \leq 0.28$, in Eq. 6 suggests they represent two complementary binding modes accessible to ligands within the binding pocket of ROCK II, i.e., one of the pharmacophores explains the bioactivities of some training inhibitors while the other explains the remaining inhibitors. Similar conclusions were reached about the binding pockets of other targets based on QSAR analysis [21–31]. Figures 2c and 3c show Hypo6/35 and Hypo4/15 and how they map **107** ($IC_{50} = 0.009 \mu\text{M}$) and **98** ($IC_{50} = 0.02 \mu\text{M}$), respectively. Figures 2e and 3e show how the two pharmacophores fit two distinct collections of training compounds, which further supports the notion of at least two binding modes. The X, Y, and Z coordinates of the two pharmacophores are given in Table 2.

Other descriptors emerged the QSAR equation as complex terms composed of the multiplication products of two

or three descriptors, which suggest they exert complex effects on ROCKII binding. For example, the descriptors AtypC24, AtypN72 and Wiener index emerged in Eq. 6 as complex products combined with positive and negative slopes.

However, AtypH53 appeared generally in the QSAR equation combined with negative slopes, which seem suggest that ligands possessing 1,3-disubstituted cyclohexyl (or cyclopentyl) cores seem to better fit the binding site compared to other substitution patterns, e.g., 1,4-disubstituted cyclohexyl-based cores.

Similarly, emergence of LUMO in Eq. 6 as part of complex product with negative slope suggests that ligand/ROCK II affinity favors electrophilic ligands probably due to π -stacking with certain electron-rich aromatic centers in the binding pocket, probably the aromatic rings of Phe103 and Phe384 (Fig. 3a).

Finally, emergence of JX in Eq. 6 illustrates certain role played by the ligands' topology in the binding process. However, despite their predictive significance, the information content of topological descriptors is quite obscure.

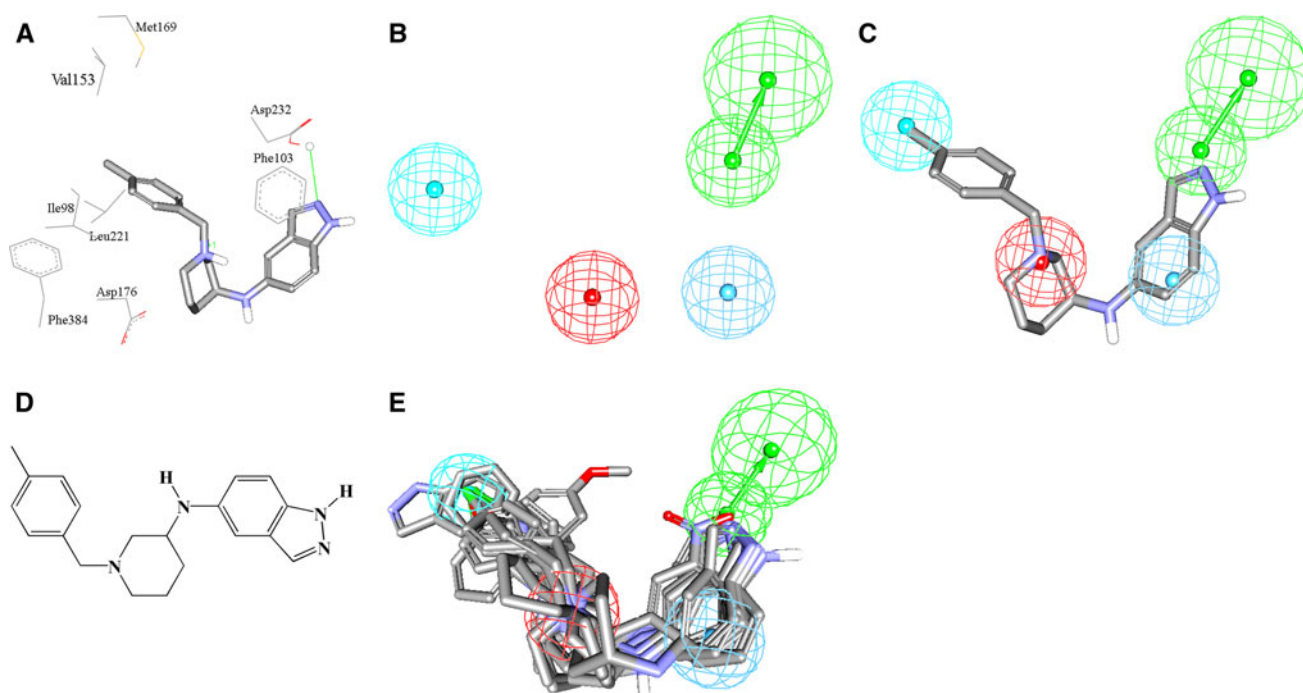


Fig. 3 **a** Docked structure of training compound **107** ($IC_{50} = 0.009 \mu\text{M}$, table A1 under Supplementary Materials) into ROCK II (PDB code: 2H9 V, resolution 1.85 Å) binding pocket. **b** Pharmacophoric features of Hypo6/35: HBA as green vector spheres, Hbic as light blue spheres, HbicArom as dark blue spheres,

PosIon as red spheres. **c** Hypo6/35 fitted against **107**, **d** Chemical structure of **107**. **e** Hypo6/35 fitted against a collection of training compounds, namely: **68**, **69**, **95**, **96**, **98**, **99**, **101**, **102**, **107**, **108**, **109**, **113**, **126**, **129** and **130**

Table 2 Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo4/15 and Hypo6/35

Model	Definition	Chemical features						
		HBD		RingArom	Hbic	HbicArom	PosIon	
Hypo4/15 ^a	Weights	1.93610		1.93610	1.93610	1.93610	1.93610	
	Tolerances	1.60	2.20	1.60	2.20	1.60	1.60	
	Coordinates							
	X	4.21	4.77	-3.72	-1.99	2.94	-0.84	-1.77
	Y	1.57	1.00	-1.13	-3.49	-0.67	-3.10	-0.43
Z	-3.13	-6.03	-0.36	0.32	-2.35	1.72	1.88	
Hypo6/35 ^b		HBAN			Hbic	HbicArom	PosIon	
	Weights	2.22164			2.22164	2.22164	2.22164	
	Tolerances	1.60		2.20	1.60	1.60	1.60	
	Coordinates							
	X	0.78	2.13	0.38	-2.03	-0.57	-0.57	
Y	-5.11	-7.59	-0.38	-4.73	-0.37	-0.37		
Z	4.49	5.53	4.40	-5.40	-0.15	-0.15		

^a Hypo4/15: the 4th pharmacophore hypothesis generated in the 15th HYPOGEN run (Table 1)

^b Hypo6/35: the 6th pharmacophore hypothesis generated in the 35th HYPOGEN run (Table 1)

It remains to be mentioned that although QSAR Eq. 6 and its associated pharmacophore models succeeded in explaining the bioactivities of most training and testing ROCKII inhibitors, they failed in explaining the bioactivities of some testing compounds (i.e., outliers), in particular, Eq. 6 significantly underestimated the excellent

potency of **81** ($IC_{50} = 0.02 \mu\text{M}$, table A2 under Supplementary Materials), which appeared as outlier outside the ± 1.0 log cycle error margin in Fig. 1b (upper right quadrant). The fact that this compound failed to map any of the pharmacophores in Eq. 6 (as can be seen in table D under Supplementary Materials), despite its excellent potency,

suggests the existence of additional binding modes available for ROCK II ligands, i.e., extra to those described in Eq. 6. In fact, removing this compound from the testing set raised r_{PRESS}^2 value from 0.494 to 0.623. Similar arguments seem to apply for testing compounds **76** and **91** (table A2 under Supplementary Materials) albeit they are of much less influence on r_{PRESS}^2 .

Receiver operating characteristic (ROC) curve analysis

To further validate the resulting models (both QSAR and pharmacophores), we subjected our QSAR-selected pharmacophores to receiver-operating curve (ROC) analysis. In ROC analysis, the ability of a particular pharmacophore model to correctly classify a list of closely-related compounds (i.e., of similar H-bonding propensities and lipophilicities) as actives or inactives is tested and indicated by the area under the curve (AUC) of the corresponding ROC together with other parameters, namely, overall accuracy, overall specificity, overall true positive rate and overall false negative rate (see “Receiver operating characteristic curve analysis” under Experimental for more details) [49–52]. Optimal ROC parameters indicate the superiority of the corresponding pharmacophore model over simpler molecular similarity-based search queries.

Table 3 and Fig. 4 show the ROC results of our QSAR-selected pharmacophores. Hypo4/15 and Hypo6/35 illustrated excellent overall performances with AUC values of 99.5 and 99.4% respectively. This is not unexpected, as the presence of positive ionizable features in both models should significantly enhance their selectivities (Fig. 5).

Comparison of pharmacophore model with the active site of ROCKII

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the pharmacophoric features of Hypo4/15 and Hypo6/35 and how they map

Table 3 ROC curve analysis criteria for QSAR-selected pharmacophores and their sterically-refined versions

Pharmacophore model	ROC ^a –AUC% ^b	ACC% ^c	SPC% ^d	TPR% ^e	FNR% ^f
Hypo4/15	99.5	97.3	99.7	11.1	0.3
Hypo6/35	99.4	97.3	99.2	27.8	0.8

^a ROC receiver operating characteristic curve

^b AUC area under the curve

^c ACC overall accuracy

^d SPC overall specificity

^e TPR overall true positive rate

^f FNR overall false negative rate

training compounds **98** and **107**, respectively, with optimal docked poses of the two compounds. The docking experiments were conducted employing LigandFit docking engine and through default docking settings [38]. Figures 2 and 3 show the pharmacophores, docked poses and corresponding mapped conformers.

We compared the structure of **98** (IC₅₀ = 0.02 μM, table A1 under supplementary material) docked into ROCK II (PDB code: 2H9 V) with Hypo4/15. Figure 2 shows the chemical structure of the ligand and compares its ROCK II docking picture with the way it maps Hypo4/15. Mapping the NH of the indazole ring in **98** against HBD feature in Hypo4/15 (Fig. 2c) corresponds to hydrogen bonding interactions tying the indazole NH and the peptidic carbonyl of Ala102 (Fig. 2a) in the docked complex. Similarly, mapping the benzene ring of the ligand’s indazole against Hbic feature in Hypo4/15 (Fig. 2c) corresponds to hydrophobic stacking against the aromatic ring of Phe103. Moreover, fitting the pyrrolidine core in **98** against PosIon and Hbic features in Hypo4/15 (Fig. 2c) seems to encode for electrostatic and hydrophobic attractive interactions connecting the pyrrolidine NH and carbon ring with the carboxylate and CH₂ linker of Asp232,

Fig. 4 ROC curves of: **a** Hypo4/15, **b** Hypo6/35

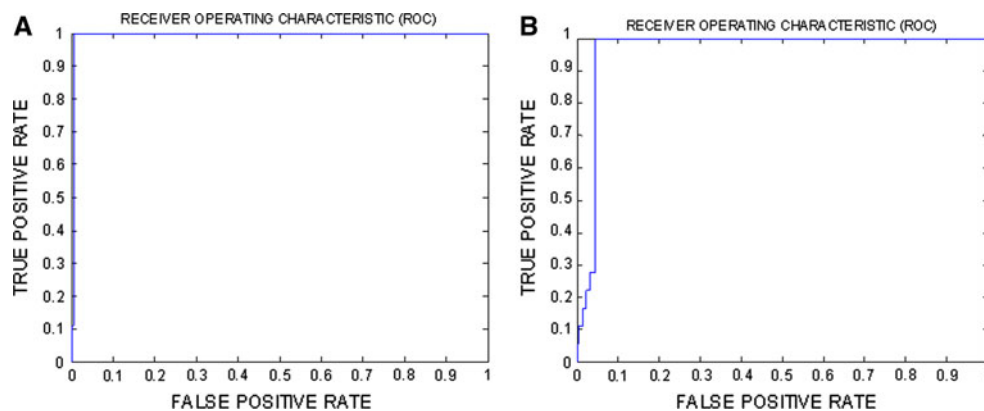
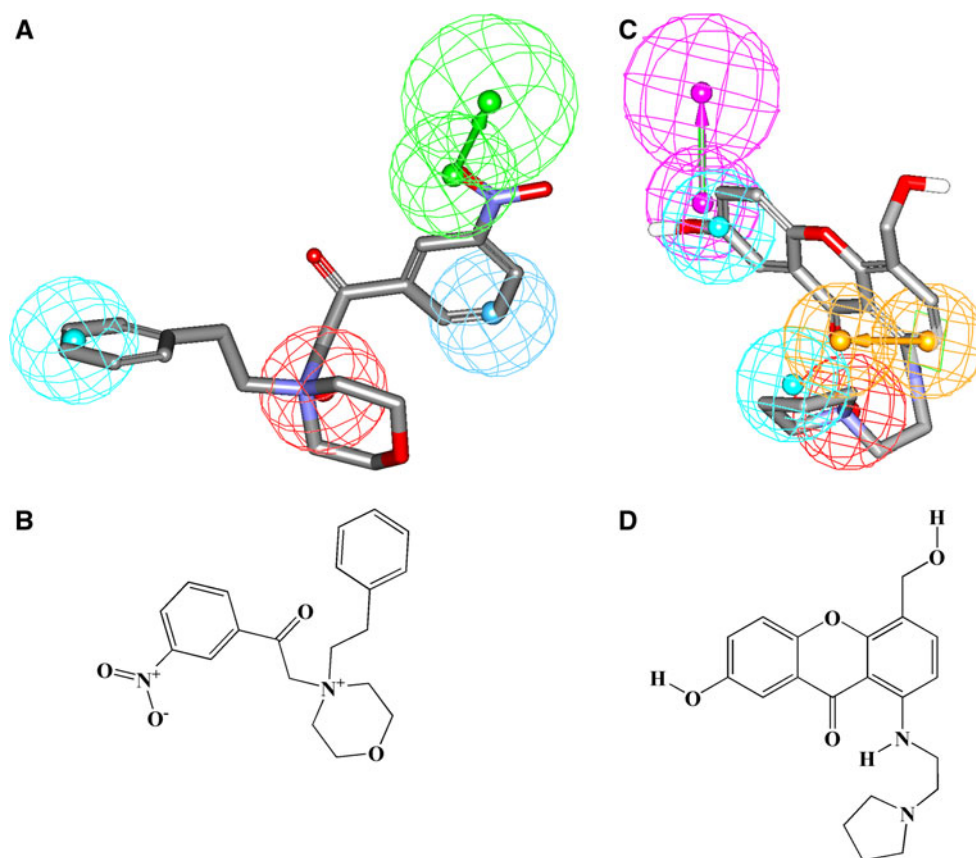


Fig. 5 **a** show Hypo6/35 fitted against hit 145 ($IC_{50} = 1 \mu\text{M}$, Table 5) **b** Chemical structure of 145 **c** show Hypo4/15 fitted against hit 144 ($IC_{50} = 4.77 \mu\text{M}$, Table 5) **d** Chemical structure of 144



respectively (Fig. 2a). Finally, fitting the chlorobenzene in **98** against a RingArom feature in Hypo4/15 correlates with stacking against the adjacent hydrophobic side chain of Val106.

Similar analogies can be drawn between the docked conformer of **107** ($IC_{50} = 0.009 \mu\text{M}$, table A1 under supplementary material) and the way it maps Hypo6/35, as shown in Fig. 3. Mapping the amino functionality of the piperidine core of **107** against PosIon feature in Hypo6/35 (Fig. 3c) agrees with electrostatic attraction connecting this group and the carboxylate of Asp176 in the docked pose (Fig. 3a). Likewise, positioning the aromatic methyl substituent of **107** within a hydrophobic pocket comprised of the side chains of Leu221, Ile98, VAL153, Phe384 and Met169 in the docked pose (Fig. 3a) agrees with mapping this moiety against Hbic feature in Hypo6/35 (Fig. 3c). Similarly, mapping the sp^2 nitrogen and benzene ring of the ligand's indazole group against HBA and HbicArom in Hypo6/35 (Fig. 3c) points to hydrogen bonding and aromatic stacking interactions connecting the indazole ring with the carboxylic acid and aromatic side chains of Asp232 and Phe103 in the docked pose, respectively, as in Fig. 3a.

Clearly from the above discussion, Hypo4/15 and Hypo6/35 represent two valid binding modes assumed by ligands within ROCK II. Furthermore, these models point

Table 4 Numbers of selected, filtered, tested, active and inactive hits captured by Hypo4/15 and Hypo6/35 from NCI list of compounds

	Number of in silico hits captured by	
	Hypo4/15	Hypo6/35
Post screening filtering ^a		
Before	674	4692
After	172	215
Selected based on QSAR predictions	40	42
Acquired from the NCI ^b	29	5
Number of actives ^c	5	16
Assayed to determine IC_{50}	2	6

NCI national cancer institute list of available compounds includes 238,819 structures

^a Using Lipinski's and Veber's rules. A maximum of two Lipinski's violations were tolerated

^b 23 acquired compounds were mutually captured by both pharmacophores (Hypo4/15 and Hypo6/35), the total number of acquired compounds from the NCI is 51 compounds. See table E in supplementary Materials to identify the origin of captured compounds in each case

^c Compounds illustrating inhibition percentage $\geq 20\%$ were counted as actives

to limited number of critical interactions required for high ligand- ROCK II affinity in each of the binding modes. In contrast, docked complexes reveal many bonding

interactions without highlighting critical ones. Figures 2a and 3a only show interactions corresponding to pharmacophoric features while other binding interactions were hidden for clarity.

In silico screening and subsequent in vitro evaluation

Hypo4/15 and Hypo6/35 were employed as 3D search queries against the NCI (238,819 structures) using the “Best Flexible Database Search” option implemented within CATALYST. Compounds that have their chemical groups spatially overlap (map) with corresponding features of the particular pharmacophoric model were captured (hits). Table 4 summarizes the numbers of captured hits by each pharmacophore.

NCI hits were filtered based on Lipinski’s [54] and Veber’s [55] rules. Enforcing some kind of drug-likeness pre-filters should help in finding hits more amenable for subsequent optimization into leads.

Surviving hits were fitted against Hypo4/15 and Hypo6/35 and their fit values, together with other relevant molecular descriptors, were substituted in QSAR Eq. 6 to

predict their anti-ROCK II IC_{50} values. The highest-ranking hits compounds were evaluated in vitro against recombinant ROCK II assay kit (Cyclex, Japan). Table 4 shows the number of selected, filtered, tested, active and inactive hits captured by Hypo4/15 and Hypo6/35 from NCI list of compounds, while table E under Supplementary Materials summarizes all information about captured hit molecules acquired from the NCI, their fit values against Hypo4/15 and Hypo 6/35, their corresponding QSAR estimates from Eq. 6 and there in vitro anti- ROCK II inhibition percentage at 10 μ M. Figure A under Supplementary Materials illustrates the dose/response plots of the active hits. To validate our bioassay settings, we determined the IC_{50} value of standard inhibitor Y-27632 under the same conditions [61].

Initially, hits were screened at 10 μ M concentrations, subsequently; compounds showing anti-ROCK II inhibitory percentages $\geq 35\%$ at 10 μ M were further assessed to determine their IC_{50} values. Table 5 shows active hits and their corresponding estimated and experimental anti-ROCK II bioactivities. Interestingly, their corresponding dose/response regression lines (Figure A under

Table 5 High-ranking hit molecules with their fit values against (Hypo4/15; Hypo 6/35) their corresponding QSAR estimates from Eq. 6 and there in vitro anti- ROCK II bioactivities

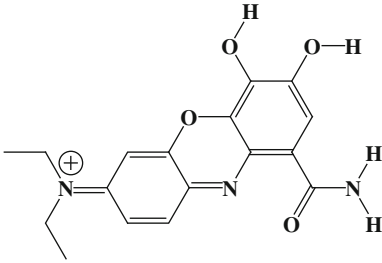
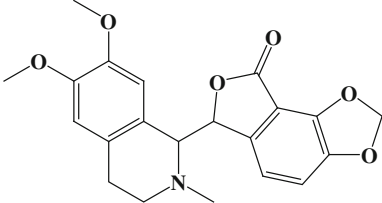
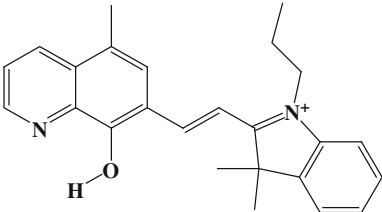
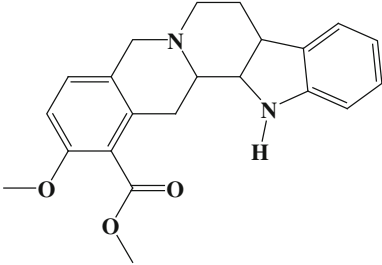
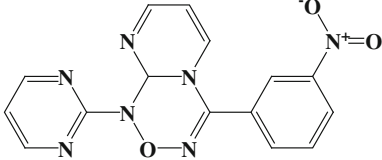
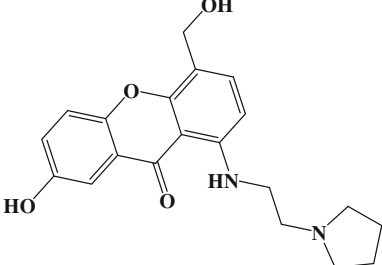
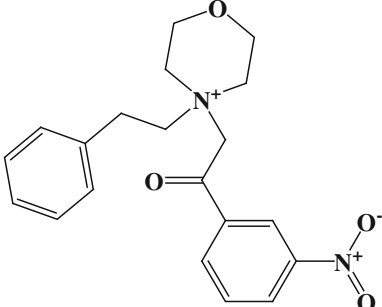
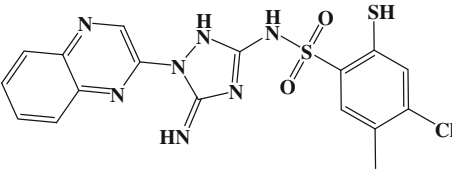
No. ^a	NCI code	Structure	Fit values ^b		QSAR predictions ^c		Experimental ^d
			Hypo4/15	Hypo6/35	Log (1/ IC_{50})	IC_{50} (μ M)	
139	NCI0007823		0	8.12	0.66	0.22	5.57
140	NCI0032983		0	8.25	0.82	0.15	1.41
141	NCI0104542		4.54	3.51	-0.42	2.6	1.69

Table 5 continued

No. ^a	NCI code	Structure	Fit values ^b		QSAR predictions ^c		Experimental ^d
			Hypo4/15	Hypo6/35	Log (1/IC ₅₀)	IC ₅₀ (μM)	
142	NCI0151684		0	7.81	0.77	0.17	45.35
143	NCI0380794		0	7.4	0.86	0.14	0.70
144	NCI0377095		6.45	0	0.17	0.67	4.77
145	NCI0401383		0	7.88	0.78	0.17	1.00
146	NCI0674004		0	8.1	3.13	0.00075	1.65

Further information about these compounds and other tested hits are shown in Table E under Supplementary Materials

^a Compound numbers

^b Best-fit values calculated by equation (D) in section SM-1 under Supplementary Materials

^c Predictions based on optimal QSAR model (6)

^d Experimental in vitro validation against rho-kinase

Supplementary Materials) illustrated excellent correlation coefficients, strongly suggestive of non-promiscuous inhibitory behavior.

It remains to be mentioned that QSAR predictions of tested hits were generally rather accurate (Table 5), i.e., within 1.0 log cycle from experimental values, except in

two cases: **142** and **146**, which were significantly overestimated by the QSAR model. We believe these prediction errors are because training compounds used in QSAR and pharmacophore modeling are significantly structurally different from these two molecules, which limits the extrapolatory potential of the QSAR equation. Furthermore, the fact that we implemented different bioassay method from that used for training compounds can also explain part of the predicted-to-experimental differences in hit bioactivities. Additionally, QSAR-based predictions have their errors; in fact an r^2_{PRESS} value of 0.494 suggests certain level of uncertainty in predictions.

Conclusions

ROCK II inhibitors are currently considered as potential treatments for hypertension. The pharmacophoric space of ROCK II inhibitors was explored via six diverse sets of inhibitors and using CATALYST-HYPOGEN to identify high quality binding model(s). Subsequently, genetic algorithm and multiple linear regression analysis were employed to access optimal QSAR model capable of explaining anti-ROCK II bioactivity variation across 138 collected ROCK II inhibitors. Two orthogonal pharmacophoric models emerged in the QSAR equation suggesting the existence of at least two distinct binding modes accessible to ligands within ROCK II binding pocket. The QSAR equation and the associated pharmacophoric models were experimentally validated by the identification of several ROCK II inhibitors retrieved via *in silico* screening, out of which 8 inhibitors illustrated micromolar potencies. Our results suggest that the combination of pharmacophoric exploration and QSAR analyses can be useful tool for finding new diverse ROCK II inhibitors.

Acknowledgments This project was partially sponsored by the Faculty of Graduate Studies (This work is part of PhD. Thesis of Rand Shahin). The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds.

References

- Shimokawa H, Rashid M (2007) *Trends Pharmacol Sci* 28:296–302
- Olson K (2008) *Curr Opin Cell Biol* 20:242–248
- Kumar R, Singh V, Baker K (2007) *J Mol Cell Cardiol* 42:1–11
- Muller B, Mack H, Teusch N (2005) *Nat Rev Drug Discov* 4:387–399
- Offermanns S, Wettschureck N (2002) *J Mol Med* 80:629–638
- Dong M, Bryan P, James K, Yip Y-Y, Gabriel WK, Yu C (2010) *Drug Discov Today* 15:622–629
- Takami A, Iwakubo M, Okada Y, Kawata H, Takahashi N, Shindo K, Kimura K, Tagami Y, Miyake M, Fukushima K, Inagaki M, Amano M, Kaibuchi K, Iijima H (2004) *Bioorg Med Chem Lett* 12:2115–2137
- Iwakubo M, Takami A, Okada Y, Kawata T, Tagami Y, Ohashi H, Sato M, Sugiyama T, Fukushima K, Iijima H (2007) *Bioorg Med Chem Lett* 15:350–364
- Iwakubo M, Takami A, Okada Y, Kawata T, Tagami Y, Sato M, Sugiyama T, Fukushima K, Shinichiro T, Kaibuchib K, Iijima H (2007) *Bioorg Med Chem Lett* 15:1022–1033
- Ho K, Beasley J, Belanger L, Black D, Chan J, Dunn D, Hu B, Klön A, Kultgen S, Ohlmeyer M, Parlato S, Ray P, Pham Q, Rong Y, Roughton A, Walker T, Wright J, Xu K, Xu Y, Zhang L, Webba M (2009) *Bioorg Med Chem Lett* 19:6027–6031
- Yamaguchi H, Miwa Y, Kasa M, Kitano K, Amano M, Kaibuchi K, Hakoshima T (2006) *J Biochem* 140:305–311
- Wen W, Liu W, Yan J, Zhang M. The C1 domain of ROCK II. Protein data bank code: 2ROW
- Wen W, Liu W, Yan J, Zhang M. The split PH domain of ROCK II. Protein data bank code: 2ROV
- Yamaguchi H, Kasa M, Amano M, Kaibuchi K, Hakoshima T (2006) *Structure* 14:589–600
- Beeley NRA, Sage C (2003) *Targets* 2:19–25
- Klebe G (2006) *Drug Discov Today* 11:580–594
- Steuber H, Zentgraf M, Gerlach C, Sottriffer CA, Heine A, Klebe G (2006) *J Mol Biol* 363:174–187
- Stubbs MT, Reyda S, Dullweber F, Moller M, Klebe G, Dorsch D, Mederski W, Wurziger H (2002) *ChemBioChem* 3:246–249
- DePristo MA, de Bakker PIW, Blundell TL (2004) *Structure* 12:831–838
- Gohda K, Hakoshima T (2008) *J Comput Aid Mol Des* 22(11): 789–797
- Taha MO, Bustanji Y, Al-Ghoussein MAS, Mohammad M, Zalloum H, Al-Masri IM, Atallah N (2008) *J Med Chem* 51: 2062–2077
- Taha MO, Atallah N, Al-Bakri AG, Paradis-Bleau C, Zalloum H, Younis K, Levesque RC (2008) *Bioorg Med Chem* 16:1218–1235
- Taha MO, Bustanji Y, Al-Bakri AG, Yousef M, Zalloum WA, Al-Masri IM, Atallah N (2007) *J Mol Graph Model* 25:870–884
- Al-masri IM, Mohammad MK, Taha MO (2008) *Chem Med Chem* 3:1763–1779
- Taha MO, Dahabiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008) *J Med Chem* 51:6478–6494
- Al-Nadaf A, Abu Sheikha G, Taha MO (2010) *Bioorg Med Chem* 18:3088–3115
- Abu-Hammad AM, Taha MO (2009) *J Chem Inf Model* 49: 978–996
- Abu Khalaf R, Abu Sheikha G, Bustanji Y, Taha MO (2010) *Eur J Med Chem* 45:1598–1617
- Al-Sháer M, Taha MO (2010) *Eur J Med Chem* 45:4316–4330
- Al-Sháer M, Taha MO (2010) *J Chem Inf Model* 50:1706–1723
- Taha MO, Trarairah M, Zalloum H, Abu Sheikha G (2010) *J Mol Graph Model* 28:383–400
- Abu Khalaf R, Abdula AM, Mubarak M, Taha MO (2011) *J Mol Model* 17:443–482
- Abdula AM, Abu Khalaf R, Mubarak M, Taha M (2011) *J Comput Chem* 3:463–482
- Tamura M, Nakao H, Yoshizaki H, Shiratsuchi M, Shigyo H, Yamada H, Ozawa T, Totsuka T, Hidaka H (2005) *Biochim Biophys Acta* 1754:245–252
- Discovery Studio version 2.5 (DS 2.5) User Manual (2009) Accelrys Inc, San Diego
- Van Drie JH (2003) *Curr Pharm Des* 9:1649–1664
- Poptodorov K, Luu T, Langer T, Hoffmann R (2006) In methods and principles in medicinal chemistry. In: Hoffmann RD (ed)

- Pharmacophores and Pharmacophores Searches, vol 2. Wiley-VCH, Weinheim, pp 17–47
38. CERIOUS2 4.10 LigandFit User Manual (2000) Accelrys Inc., San Diego
 39. Discovery Studio 2.5.5 User Guide (2010) Accelrys Inc., San Diego
 40. CATALYST 4.11 Users' Manual (2005) Accelrys Software Inc San Diego, CA
 41. Sutter J, Güner O, Hoffmann R, Li H, Waldman M (2000) In: Güner OF (ed) Pharmacophore perception, development, and use in drug design. International University Line, La Jolla, pp 501–511
 42. Kurogi Y, Güner OF (2001) *Curr Med Chem* 8:1035–1055
 43. Poptodorov K, Luu T, Langer T, Hoffmann R (2006) In: Hoffmann RD (ed) *Methods and principles in medicinal chemistry. Pharmacophores and Pharmacophores Searches, vol 2.* Wiley-VCH, Weinheim, pp 17–47
 44. Li H, Sutter J, Hoffmann R (2000) In: Güner OF (ed) *Pharmacophore perception, development, and use in drug design.* International University Line, La Jolla, pp 173–189
 45. Bersuker IB, Bahçeci S, Boggs JE (2000) In: Güner OF (ed) *Pharmacophore perception, development, and use in drug design.* International University Line, La Jolla, pp 457–473
 46. Fischer R (1966) *The principle of experimentation illustrated by a psycho-physical.* ExpeHafner Publishing Co, 8th ed. Hafner Publishing, New York Chapter II
 47. Krovat EM, Langer T (2003) *J Med Chem* 46:716–726
 48. CERIOUS2 (2005) *QSAR Users' Manual, Version 4.10;* Accelrys Inc., San Diego, pp 43–88, 221–235, 237–250
 49. Verdonk ML, Marcel L, Berdini V, Hartshorn MJ, Mooij WTM, Murray CW, Taylor RD, Watson P (2004) *J Chem Inf Comput Sci* 44:793–806
 50. Kirchmair J, Markt P, Distinto S, Wolber G, Langer T (2008) *J Comput Aided Mol* 22:213–228
 51. Irwin JJ, Shoichet BK (2005) *J Chem Inf Comput Sci* 45:177–182
 52. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O (2005) *J Med Chem* 48:2534–2547
 53. Jacobsson M, Liden P, Stjernschantz E, Bostroem H, Norinder U (2003) *J Med Chem* 46:5781–5789
 54. 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 Del Rev* 46:3–26
 55. 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* 45:2615–2623
 56. CycLex, Rho-Kinase Assay Kit (Cat# CY-1160) Users' Manual (2009) CycLex Co, Ltd, Ina, Nagano, Japan
 57. Li H, Sutter J, Hoffmann R (2000) In: Güner OF (ed) *Pharmacophore perception, development, and use in drug design.* International University Line, La Jolla, pp 173–189
 58. Sutter J, Güner O, Hoffmann R, Li H, Waldman M (2000) In: Güner OF (ed) *Pharmacophore perception, development, and use in drug design.* International University Line, La Jolla, pp 501–511
 59. Bersuker IB, Bahçeci S, Boggs JE (2000) In: Güner OF (ed) *Pharmacophore perception, development, and use in drug design.* International University Line, La Jolla, pp 457–473
 60. Ramsey LF, Schafer WD (1997) *The Statistical Sleuth, 1st edn.* Wadsworth Publishing Company, Belmont CA
 61. 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