Medicinal Chemistry Research (2019) 28:1368–1387
https://doi.org/10.1007/s00044-019-02377-7

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

QSAR-guided pharmacophore modeling and subsequent virtual screening identify novel TYK2 inhibitor

Elham Y. Al-Barghouthy[1] · Areej Abuhammad[1] · Mutasem O. Taha[1]

Received: 9 March 2019 / Accepted: 23 May 2019 / Published online: 3 June 2019

Abstract

Tyrosine Kinase 2 (TYK2) inhibition is of potential therapeutic value for treating autoimmune diseases. An elaborate ligand-based computational workflow was employed to explore structural requirements for TYK2 inhibition. Genetic function algorithm (GFA) was coupled to k-nearest neighbor (kNN) and multiple linear regression (MLR) analyses to search for predictive QSAR models based on optimal pharmacophore(s)/physicochemical descriptors combinations. QSAR-selected pharmacophores were validated by receiver operating characteristic (ROC) curve analysis and by comparison with crystallographic structures of known inhibitors complexed within the TYK2 binding pocket. Optimal QSAR models and their associated pharmacophore hypotheses were used to identify new TYK2 inhibitory leads retrieved from the National Cancer Institute (NCI) structural database. The most potent hit exhibited experimental anti-TYK2 IC_{50} of 7.1 μM.

Keywords TYK2 · Pharmacophore · QSAR · Multiple linear regression · k-nearest neighbor

Introduction

Autoimmune diseases (AIDs) have high incidence and prevalence rates worldwide (Lerner et al. 2015). Traditional treatments of AIDs can expose patients to life-threatening opportunistic infections and long-term risk of malignancy (Gillett and Chan, 2000). Therefore, the development of more specific treatments that lower the risk of systemic immune suppression and improve tolerability is a desirable goal.

One of the newer approaches in managing AIDs is the inhibition of protein kinases by small molecules (Hernández-Flórez and Valor, 2016). One of the most relevant intracellular signaling pathways involved in autoimmune disorders is the janus kinase pathway (JAK-STAT) (Hernández-Flórez and Valor, 2016).

Janus kinases (JAKs) are non-receptor protein tyrosine kinase family comprised of four members: Tyrosine Kinase 2 (TYK2), Janus Kinase 1 (JAK1), Janus Kinase 2 (JAK2) and Janus Kinase 3 (JAK3). JAKs are associated with the intracellular domains of a wide range of cytokines and growth factor receptors involved in inflammation and hematopoiesis (Liang et al. 2013; Patterson et al. 2014). TYK2 was the first member of the JAK family to be identified (Stark et al. 1998). It is involved in interleukin (ILs) and interferon signaling (Roskoski 2016), in particular, IL-12 and IL-23. IL-12 is related to T-helper cells type 1 (Th1) differentiation whilst IL-23 has a key role in T-helper cells type 17 (Th17) differentiation. Th1 and Th17 are responsible for inflammatory reactions. Overexpression of these cells is associated with certain autoimmune diseases such as the irritable bowel disease (Duerr et al. 2006; Norman 2012; Bernstein et al. 2005; De Mesquita et al. 2009; Ellinghaus et al. 2012; Lees et al. 2011), multiple sclerosis disease (Matusevicius et al. 1999) and psoriasis (Ma et al. 2008). The association between TYK2 and rheumatic diseases has also been reported (Sigurdsson et al. 2005; Sigurdsson et al. 2007).

These findings prompted the exploration of TYK2 inhibition as a potential therapeutic rationale for treating autoimmune diseases (Norman 2012). However, the development of TYK2 inhibitors is still in early stages (Liang et al. 2013; Kettle et al. (2017)). However, several

Supplementary information The online version of this article (https://doi.org/10.1007/s00044-019-02377-7) contains supplementary material, which is available to authorized users.

✉ Mutasem O. Taha
mutasem@ju.edu.jo

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

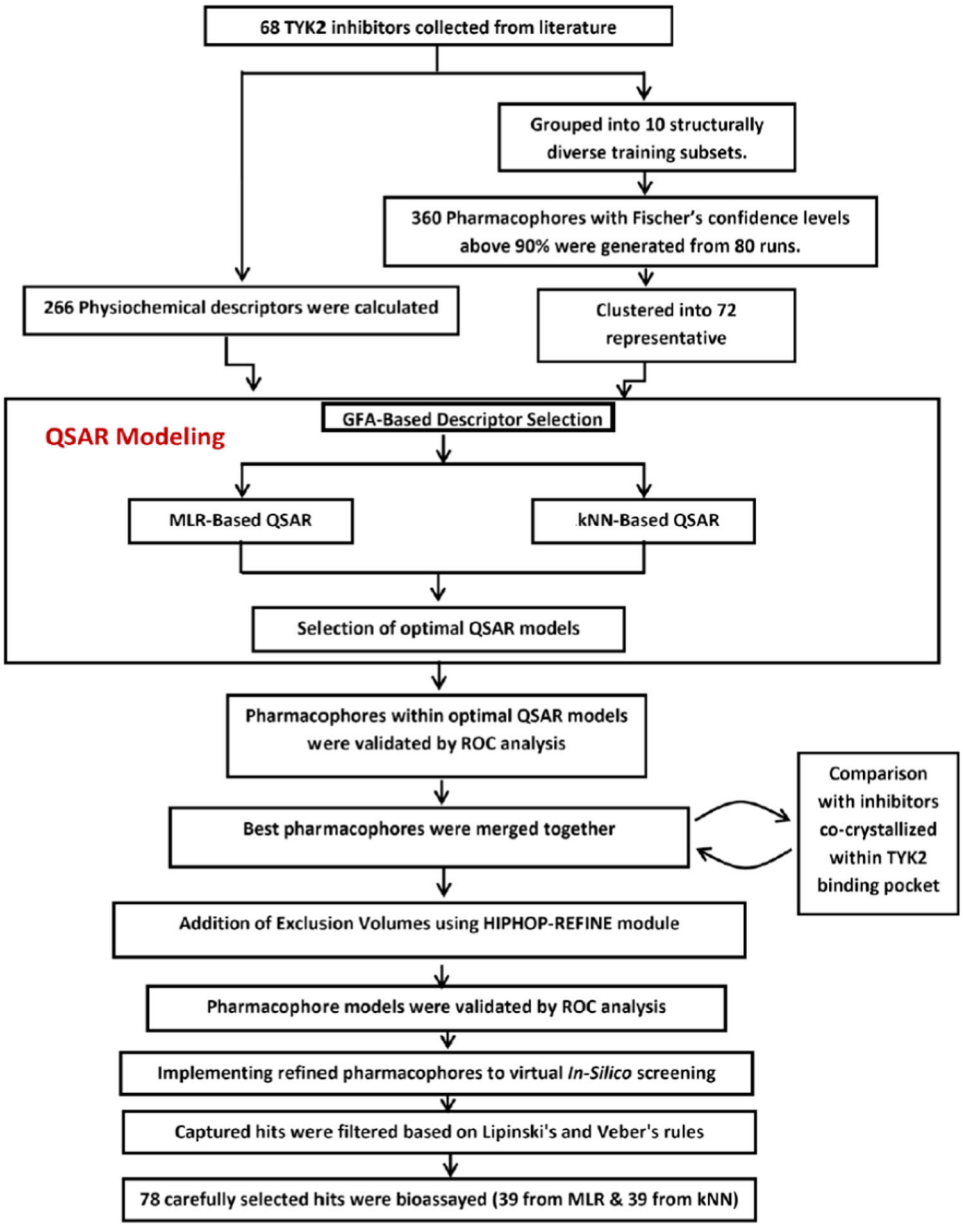


Fig. 1 Summary of the computational workflow of the project

TYK2 inhibitors have been recently patented (Norman 2012; Kettle et al. (2017); Blench et al, 2012; Blench et al. 2012).

Discovery efforts of new TYK2 inhibitors focused on structure-based ligand design (Liang et al. 2013a; Liang et al. 2013b; Tsui et al. 2011; Fensome et al. 2018; Hamaguchi et al. 2018; Liang et al. 2017; Moslin et al. 2017; Yogo et al. 2016; Tokarski et al. 2015; Galatsis et al. 2014; Lupardus et al. 2014; Chrencik et al. 2010; Frush et al. 2017). To date, several crystallographic structural studies on TYK2-inhibitor complexes have been reported (Liang et al. 2013; Tsui et al. 2011; Fensome et al. 2018; Hamaguchi et al. 2018; Liang et al. 2017; Moslin et al. 2017; Yogo et al. 2016; Tokarski et al. 2015; Galatsis et al. 2014; Lupardus et al. 2014; Chrencik et al. 2010). However, reliance solely on crystallographic structures for ligand design has limitations related to inadequate resolution (Beeley and Sage 2003) and crystallization-related artifacts of the ligand–protein complex (Homans 2007; Klebe 2006; Steuber et al. 2006; Stubbs et al. 2002). Moreover, crystallographic structures ignore protein discrete conformational states and anisotropic motion (Depristo 2004; Srivastava et al. 2018).

The continued interest in designing new TYK2 inhibitors prompted us to implement our ligand-based QSAR-guided pharmacophore selection methodology (Taha 2012; Abuhammad and Taha 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011;

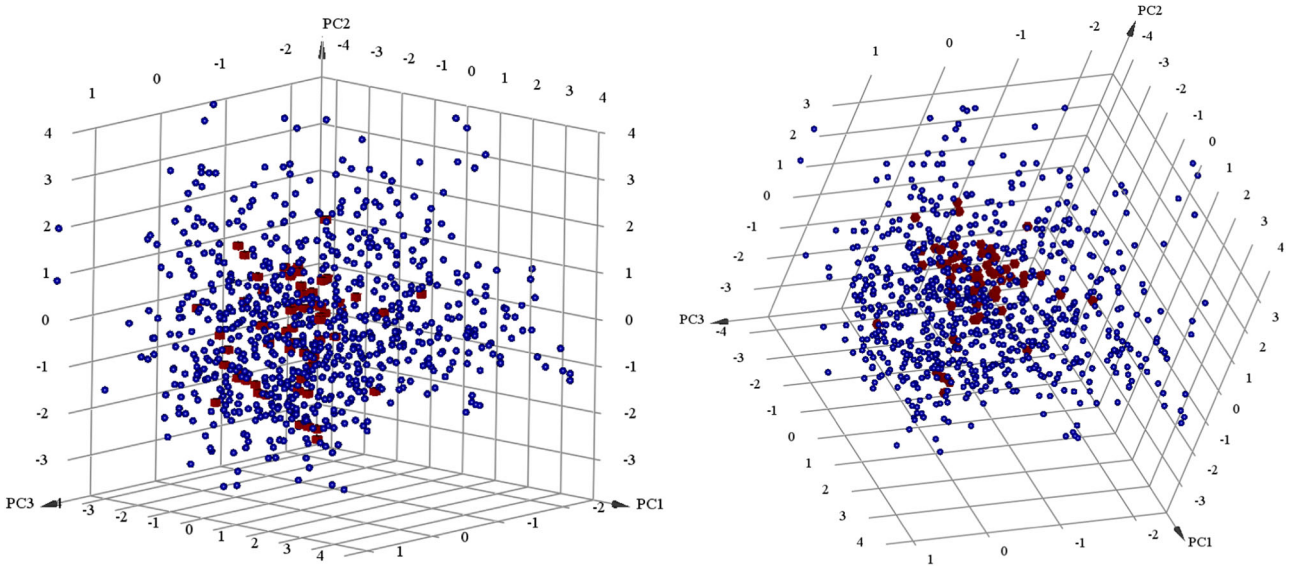


Fig. 2 Two views of 3D plot showing three main principal components calculated for the modeled training TYK2 inhibitors (**1**–**68**, Table S1) compared to 892 TYK2 inhibitors reported in ChEMBL (based on 12 physicochemical descriptors, see text). Red cubes (■) represent modeled training TYK2 inhibitors while blue spheres (●) represent TYK2 inhibitors reported in ChEMBL

Aboalhaija et al. 2016; Alabed et al. 2016; Khanfar and Taha, 2013) to explore the possibility of developing ligand-based pharmacophore(s) integrated within self-consistent QSAR model(s). Ligand-based methods have the advantage of being able to bypass pitfalls of structure-based drug design (Taha et al. 2008b; Al-Nadaf et al. 2010; Taha et al. 2011; Shahin et al. 2012; Aboalhaija et al. 2016; Alabed et al. 2016). The resulting pharmacophores and QSAR models were used to virtually screen the National Cancer Institute (NCI) Structural Database for new TYK2 inhibitors.

Material and methods

Figure 1 summarizes the overall computational workflow.

Software

Discovery Studio (Version 2.5.5), BIOVIA Inc. (http://www.3dsbiovia.com/), USA.

Data preparation and pharmacophore exploration

Training compounds were collected from the European Bioinformatics Institute database ChEMBL (March/2016). Only compounds of explicit stereochemistry, published in peer reviewed journal articles and have their inhibitory bioactivities reported as Ki values were included in this study to ensure data consistency (Taha 2012). Therefore, only 68 inhibitors (**1**-**68**, Table S1 under Supporting Materials) were recruited in modeling (Liang et al. 2013a; Liang et al. 2013b, Kulagowski et al. 2012).

The collected set of TYK2 inhibitors were categorized into 10 training subsets (shown in Table S2 in Supporting Material) and used to explore the pharmacophoric space of TYK2 binding through 80 pharmacophore-generating automatic runs using Discovery Studio. Training compounds in each subset were selected in such a way that they conform to certain envisaged common binding mode and that their differences in anti-TYK2 bioactivities are primarily attributable to presence or absence of pharmacophoric features, i.e., not due to steric shielding or bioactivity reducing or enhancing auxiliary groups (e.g., electron withdrawing or donating groups) (Taha 2012; Abuhammad and Taha, 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Shahin et al. 2012; Aboalhaija et al. 2016; Alabed et al. 2016). Only pharmacophore models of 4 and 5 binding features were explored, while models of less than 4 features were ignored (as shown in Supporting Table S3). This restriction is aimed at narrowing the evaluated pharmacophoric space. Nevertheless, to have more pharmacophore hypotheses delivered to the subsequent QSAR competition step, it was decided to implement not-too-strict filters on the resulting binding hypotheses: Pharmacophore models of ≥ 90% Fisher's randomization confidence (Fisher 1971) were retained for subsequent processing (i.e., clustering and QSAR analyses). Our experience suggests that tighter confidence restrictions (i.e., randomization confidence ≥ 95% or 98%) might overstep significant pharmacophores (Taha 2012; Abuhammad and Taha, 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016; Khanfar and Taha, 2013; Abutayeh and Taha, 2019).

Supplementary sections S1 to S4 and supplementary Fig. S1 describe the computational workflow in details (Taha 2012; Abuhammad and Taha, 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Shahin et al. 2012; Aboalhaija et al. 2016; Alabed et al. 2016). Supplementary Table S4 shows the statistical criteria of pharmacophores generated during pharmacophore-exploration phase.

QSAR modeling

Genetic function algorithm (GFA) was employed to select different combinations of pharmacophores and molecular descriptors and submit them to either multiple linear regression (MLR) analysis or k-nearest neighbor (kNN) regression to evaluate abilities to explain bioactivity variations within modeled TYK2 inhibitors (**1**-**68**, Table S1). Supporting section S5 details GFA computational workflow, while Fig. S2 summarizes the main GFA steps. MLR is intended to evaluate linear correlations tying structural descriptors with ligands' bioactivities, whilst kNN modeling is non-linear nonparametric correlation method that assumes the bioactivity of certain ligand to be the distance weighted-average of bioactivities of its nearest neighbors. The neighborhood is defined based on the considered combination of descriptor(s) through certain distance metric to measure nearness (Alabed et al. 2016; Zheng and Tropsha, 2000; Khanfar and Taha, 2013; Abutayeh and Taha, 2019)

QSAR modeling was performed using a training set of 55 compounds from the collected list of TYK2 inhibitors and validated using leave-one-out r^2 (r^2_{LOO}) and predictive r^2 (r^2_{PRESS}) against a randomly selected testing set of 13 inhibitors. The testing set was selected by ranking the collected inhibitors according to their Ki values, then picking every fifth compound starting from the high-potency end. Supporting sections S6 and S7 detail the computational procedure of MLR and kNN QSAR modeling.

ROC validation, steric refinement of pharmacophore models

QSAR-selected pharmacophores were evaluated using receiver operating characteristic (ROC) curve analysis to assess their abilities to capture diverse TYK2 inhibitors from a large list of actives and decoys (Taha et al. 2008b; Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016; Triballeau et al. 2005). Two testing sets were implemented: Testing Set 1 is composed of active TYK2 inhibitors (experimentally validated) and virtual decoys (see supporting section S8 for more details), while Testing Set 2 is composed entirely of experimentally validated active and inactive TYK2 inhibitors extracted from the European Bioinformatics Institute database (ChEMBL) (see supporting section S8 and Fig. S3 for more details).

The HIPHOP-REFINE module of Discovery Studio was used to complement the resulting pharmacophores with exclusion spheres to improve their classification properties. HIPHOP-REFINE identifies spaces occupied by inactive compounds and free from active ones and fills them with exclusion volumes to resemble steric constrains of the binding pocket (see supporting section S10, supporting Table S5 and supporting Fig. S4 for more details) (Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Shahin et al. 2012; Aboalhaija et al. 2016; Khanfar and Taha, 2013).

In silico screening for TYK2 inhibitors

Successful pharmacophores were used as 3D search queries to screen the National Cancer Institute (NCI) database (253,368 compounds) employing the "best flexible database search" protocol implemented within Discovery Studio.

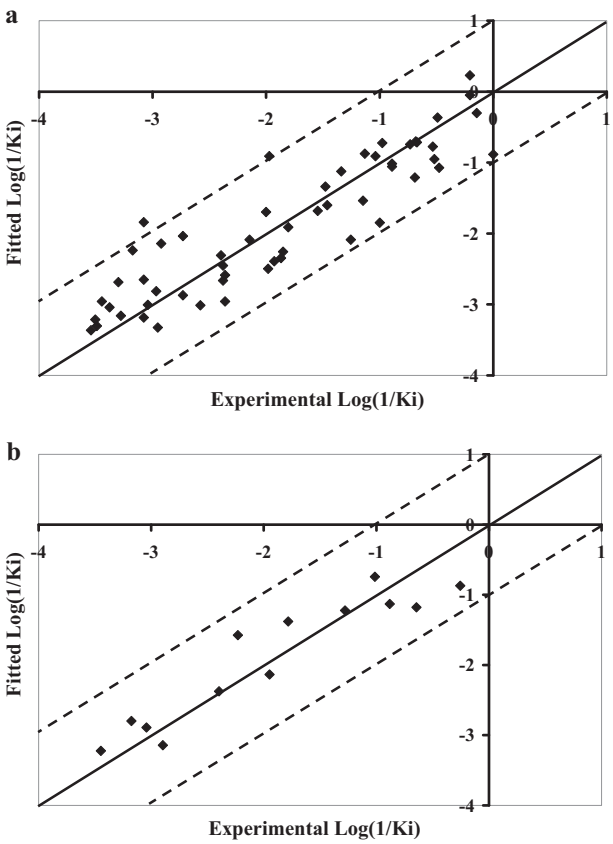


Fig. 3 Experimental versus (**a**) fitted, 55 compounds, $r^2_{LOO} = 0.76$, and (**b**) predicted, 13 compounds, $r^2_{PRESS} = 0.87$, bioactivities calculated from the best MLR-QSAR model Eq. (S1). The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log point error margins

Table 1 Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of **Hypo(I-T7-5)**, and **Hypo(F-T5-8)** selected by MLR-QSAR method

Model[a]	Definition	Chemical features							
			HBD		HBA		RingArom		Hbic
Hypo(I-T7-5)[a]	Weight		1.84		1.84		1.84		1.84
	Tolerances		1.60	2.20	1.60	2.20	1.60	1.60	1.60
	Coordinates	X	0.22	−1.74	2.66	2.22	2.87	2.81	3.52
		Y	−0.19	−1.13	−3.85	−1.94	−0.10	0.11	−4.68
		Z	−0.08	−2.16	1.03	3.30	−0.01	2.99	−1.00
Hypo(F-T5-8)[b]			HBD		Hbic		RingArom		HBA
	Weight		1.56		1.56	1.56		1.56	
	Tolerances		1.60	2.20	1.60	1.60	1.60	1.60	2.20
	Coordinates	X	−0.13	−1.52	−2.04	5.91	3.67	4.18	7.18
		Y	0.37	−1.65	3.34	−2.93	−4.25	−0.002	0.00
		Z	0.22	1.95	−0.78	−2.07	−3.57	0.00	0.00

[a]**Hypo(I-T7-5)** corresponds to the 5th pharmacophore model generated by subset I (Supporting Table S2), run number 7 (Supporting Tables S3 and S4). Model shown in Fig. 3a

[b]**Hypo(F-T5-8)** corresponds to 8th pharmacophore model generated by subset F (Supporting Table S2), run number 5 (Supporting Tables S3 and S4). Model shown in Fig. 4a

Hits were subsequently filtered based on Lipinski's rule to retain drug like molecules (Lipinski et al. 2012), and SMARTS (SMILES Arbitrary Target Specification pattern) filter to exclude hits exhibiting reactive groups (e.g., alkyl-halides, see Table S6 under supplementary material). Surviving hits were fitted against corresponding pharmacophores, and their fit values (determined as in supplementary equation S4) were substituted, together with other 2D and 3D molecular descriptors, in the corresponding QSAR models (MLR or kNN-based) to calculate predicted anti-TYK2 bioactivities. Highest ranking available hits (i.e., from the NCI) based on QSAR predictions were ordered from the NCI and tested *in vitro*. Fig. S7 and Table S7 under Supplementary Materials show the chemical structures of acquired hits from the NCI, their predicted bioactivities by MLR-QSAR and kNN-QSAR models and experimental bioactivities.

TYK2 enzyme inhibition assay

Bioassay was performed using Invitrogen Z'-LYTE® Kinase Assay (Ma et al. 2008). Each hit compound was dissolved in DMSO to form 10 mM stock solution (ca. 200 µL). Hits were initially tested at 10 µM. The bioassay was performed in black 384-well plates. For each assay a 100-nL of tested compound stock solution (10 mM) was mixed with 2.4 µL kinase buffer, 5 µL peptide substrate/TYK2 mixture and 2.5 µL ATP solution. The final 10 µL kinase reaction included ≥ 1% DMSO, .ca. 70 ng TYK2 kinase, 2 µM Tyr03 (kinase peptide substrate from Z'-LYTE®) and 25 µM ATP in 50 mM HEPES pH 6.5, 0.01% BRIJ-35, 10 mM $MgCl_2$, 1 mM EGTA and 0.02% NaN_3. The mixture was shaken for 30 s then incubated at room temperature for 1 h to complete the kinase reaction. Subsequently, 5 µL development reagent A (from Z'-LYTE®, 1:4096 dilution) was added to each reaction mixture and shaken over 30 s. The mixture was incubated at room temperature for another 60 min before measuring the fluorescence at λ_{Ex} of 445 and λ_{Em} of 520 nm.

Initially, hits were screened at 10 µM. A hit showed TYK2 inhibition percentage >50% (73%) at 10 µM (i.e., **72**, see Table 6). It was further tested at 1 µM, 10 µM, 50 µM, 500 µM and 1000 µM to determine its IC_{50} value. The tests were conducted in duplicates. Staurosporine was used as positive control (IC_{50} of 1.03 nM) (Thermo Fisher Scientific, 2016). IC_{50} values were calculated using GraphPad Prism 5.1.

Results and discussion

QSAR-guided phramacophore selection

A diverse list of 68 TYK2 inhibitors (**1**–**68**, Table S1 under Supporting Materials) were collected from ChEMBL for modeling. Comparison with other drug-like TYK2 inhibitors from ChEMBL (892 compounds) with respect to top three principal components based on 12 physicochemical descriptors (i.e., LogP, molecular weight, hydrogen bond donors and acceptors, rotable bonds, rings, aromatic rings, fractional polar surface area surface area, polar surface area and number of fragments) showed the modeled list to represent significant portion of the chemical space of reported TYK2 inhibitors. Figure 2 shows three-dimensional plot of the principal components representing the modeled list compared to reported TYK2 inhibitors.

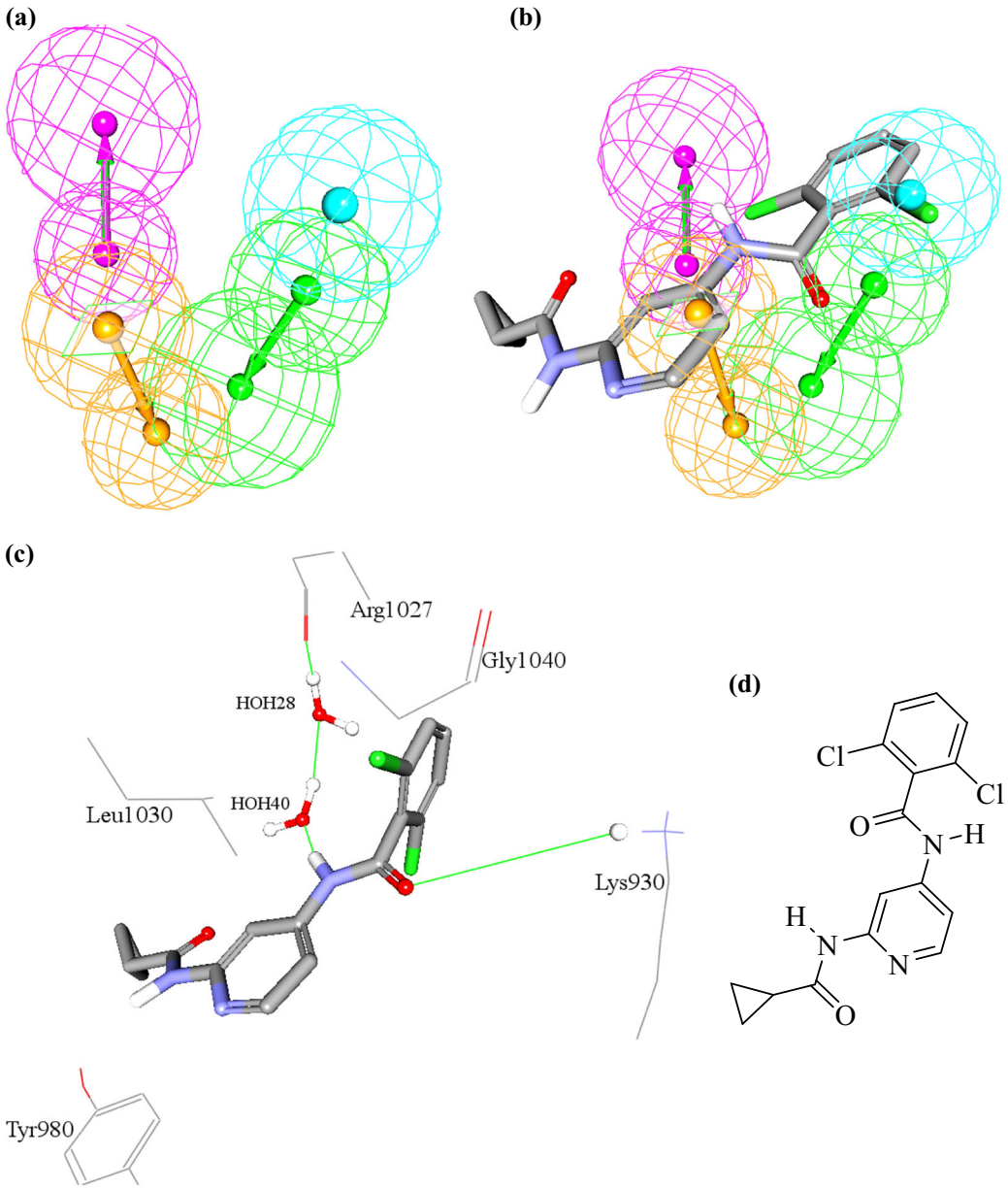


Fig. 4 **a** Pharmacophoric features of **Hypo(I-T7-5)**. HBA shown as green vectored spheres, Hbic shown as blue sphere, RingArom as orange vectored spheres, and HBD as pink vectored sphere. **b** **Hypo(I-T7-5)** fitted against the co-crystallized pose of inhibitor **11** (Supporting Information Table S1, Ki = 4.8 nM) bound within TYK2 (PDB code: 4GIH, resolution 2.0Å). **c** The co-crystallized ligand **11** within the binding pocket of TYK2. The figures shows corresponding ligand-TYK2 interactions. Hydrogen bonds are shown as green lines. **d** The chemical structure of the co-crystallized ligand **11** (Supporting Information Table S1)

The collected list was then used to explore the pharmacophoric space of TYK2 inhibitors using Discovery Studio through series of well-established steps. Representative pharmacophores (Supplementary Table S4) were allowed to compete within QSAR context to select optimal combination of pharmacophores and other molecular descriptors capable of explaining bioactivity variations across collected inhibitors (Taha 2012; Abuhammad and Taha, 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c, Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016; Abutayeh and Taha, 2019).

QSAR modeling complements pharmacophore modeling and should correct inherent flaws in pharmacophores related to explaining bioactivity effects associated with electron-donating and/or electron-withdrawing functionalities (Taha 2012; Abuhammad and Taha, 2016; Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c, Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016;

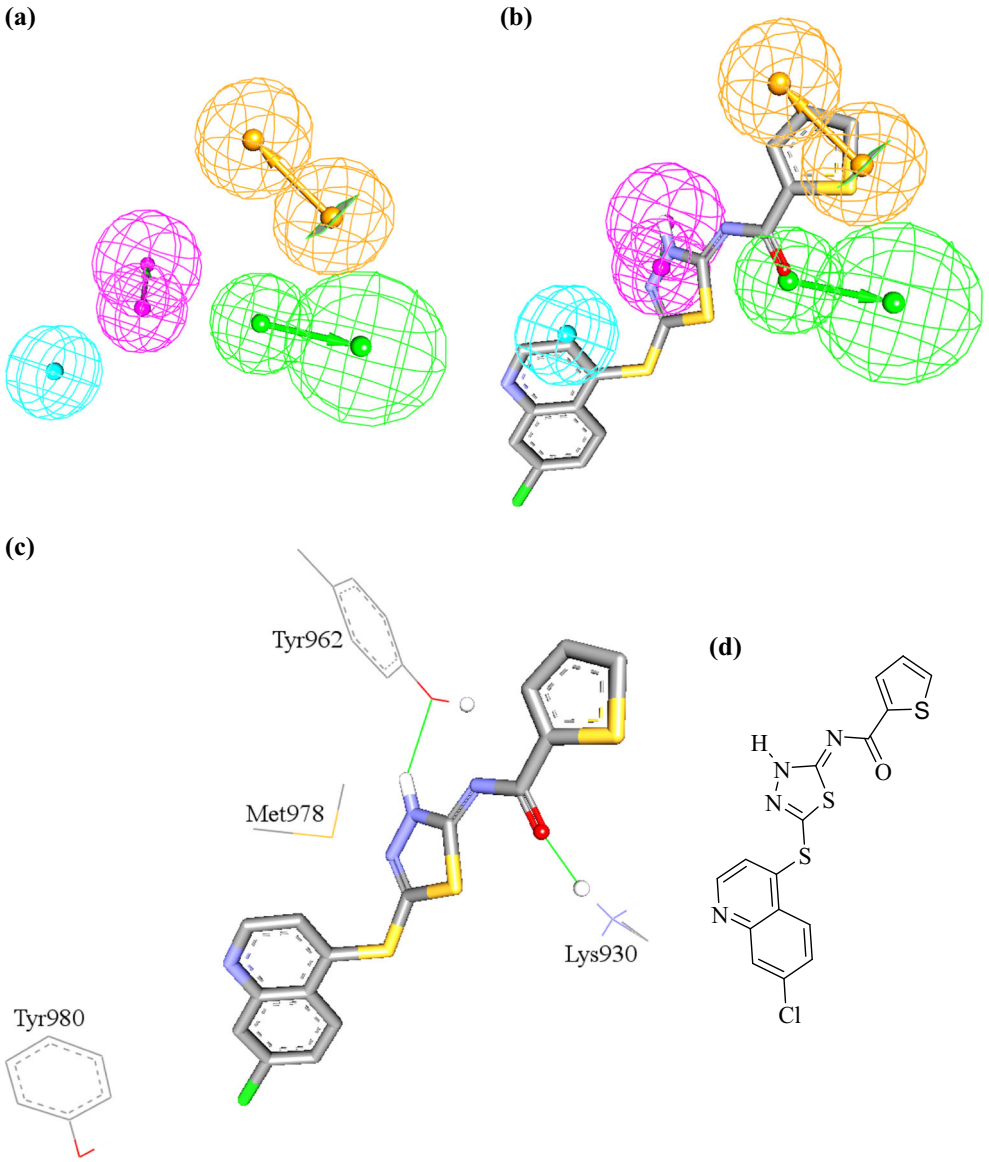


Fig. 5 **a** The pharmacophoric features of **Hypo(F-T5-8)**. HBA as green vectored spheres, Hbic as blue sphere, RingArom as orange vectored spheres, and HBD as pink vectored spheres. **b** **Hypo(F-T5-8)** fitted against the co-crystallized pose of a TYK2 inhibitor (Ki = 32 nM) as it binds within TYK2 (PDB code: 3NYX, resolution 2.5 Å). **c** The co-crystallized ligand within the binding pocket of TYK2 showing corresponding ligand-TYK2 interactions. **d** The chemical structure of bound ligand

Abutayeh and Taha, 2019) and ligand-hydration issues (Homans 2007).

QSAR modeling was performed by fitting the training compounds against 72 representative pharmacophores. The resulting fit values were enrolled together with a selection of 2D descriptors as independent variables in iterative cycles of genetic function algorithm (GFA) (Whitley 1994) coupled with multiple linear regression (GFA/MLR) or k-nearest neighbor (GFA/kNN) QSAR analysis.

Multiple linear regression-based QSAR modeling

Equation (1) represents the optimal GFA/MLR-based QSAR model, while Fig. 3 shows the corresponding scatter plots of experimental versus estimated bioactivities for both training and testing inhibitors.

$$\begin{aligned} &\mathrm{Log}(1/\mathrm{K_i}) = +0.47 - 0.79(\mathrm{NumAromaticBonds}) \\ &+4.25(\mathrm{NumAromaticRings}) + 1.73(\mathrm{NumRings3}) \\ &-8.00 \times 10^{-3}(\mathrm{JursTASA}) \\ &+0.29\mathrm{Hypo}(\mathrm{F-T5-8}) + 0.25\mathrm{Hypo}(\mathrm{I-T7-5}) \\ &\mathrm{n} = 55, \mathrm{r}^2 = 0.82, \mathrm{F-statistic} = 36.2, \mathrm{r}^2_{\mathrm{LOO}} = 0.76, \mathrm{r}^2_{\mathrm{PRESS}} = 0.87 \end{aligned} \quad (1)$$

Table 2 Optimal kNN-QSAR model including its corresponding descriptors, number of nearest neighbors and statistical criteria

Selected descriptors	Number of nearest neighbors	Statistical criteria[a,b,c,d]			
		r^2	r^2_{LOO}	$r^2_{L20\%Out}$	r^2_{PRESS}
Hypo(D-T5-10) **Hypo(D-T7-8)** ESCount_tN[e] SC3CH[f] JursDPSA3[g] ShadowYZfrac[h]	4	0.84	0.77	0.90	0.61

[a]$\mathbf{r^2}$ is the correlation coefficient between the predicted and experimental log(1/Ki) values

[b]$\mathbf{r^2_{LOO}}$ is the leave one out cross correlation coefficient

[c]$\mathbf{r^2_{L20\%Out}}$ is the leave-20%-out cross-correlation coefficient

[d]$\mathbf{r^2_{PRESS}}$ is the predictive r^2 determined for 13 randomly selected test compounds (Supporting Table S1)

[e]ESCount_tN count of cyano groups

[f]SC3CH count of third order ring and chain subgraph Count Index

[g]Jurs descriptor corresponds to difference between charge weighted partial positive surface area and charge weighted partial negative surface area

[h]ShadowYZfrac is the area of the molecular shadow in the YZ plane (Dassault Systèmes BIOVIA, 2009)

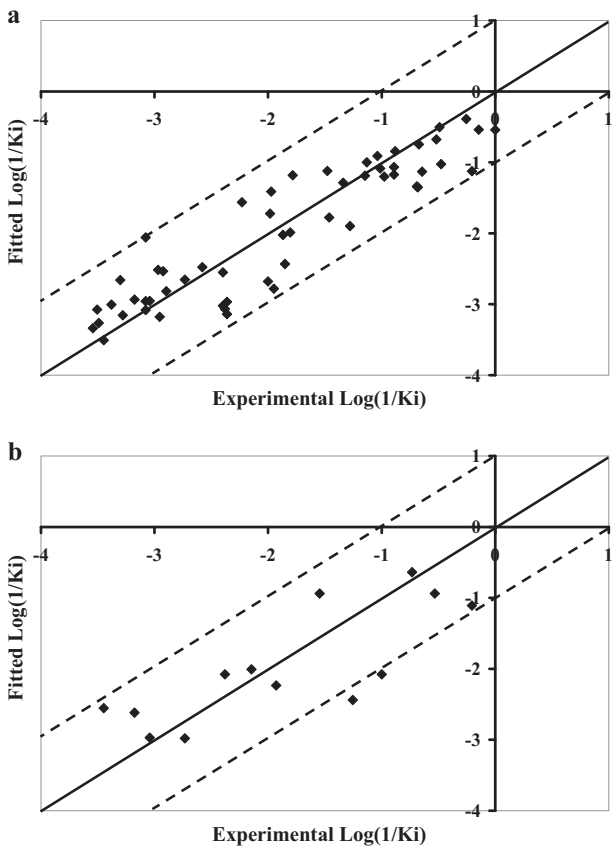


Fig. 6 Experimental versus (**a**) fitted 55 training compounds ($r^2_{LOO} = 0.77$) and (**b**) predicted 13 testing compounds ($r^2_{PRESS} = 0.61$), bioactivities calculated from the best kNN-QSAR model as in Table 2. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1.0 log point error margins

where, n is the number of training compounds, F-statistic is Fischer statistic, r^2_{LOO} is the leave-one-out cross validation correlation coefficient, and r^2_{PRESS} is the predictive r^2 determined for 13 randomly selected test compounds (see Supporting Table S1). NumAromaticBond and NumAromaticRings are the number of aromatic bonds and aromatic rings, respectively, in a particular training molecule, NumRings3 is the number of 3-membered rings (i.e., cyclopropyls), JursTASA is a Jurs-fractional charged partial surface area computed as the sum of the solvent-accessible surface area of atoms with absolute partial charge less than 0.2 (Dassault Systèmes BIOVIA 2009). **Hypo(F-T5-8)** and **Hypo(I-T7-5)** represent the fit values of training compounds against these two pharmacophores (see supporting Table S4). Fit values calculated as described in supportting section S3 (Eq. S4). The three-dimensional coordinates of **Hypo(F-T5-8)** and **Hypo(I-T7-5)** are shown in Table 1, while Figs. 4 and 5 show the two models and how they fit potent inhibitors co-crystallized within TYK2.

Appearance of NumAromaticBonds and NumAromaticRings in Eq. (1) combined with negative and positive slopes, respectively, suggests that anti-TYK2 bioactivity is related to the number of aromatic rings within ligand molecules. However, the apparent contradictory contribution of these two rather related descriptors in bioactivity is indicative of certain complex influence imparted by ligand aromaticity on ligand affinity to TYK2. Still, the significantly larger positive contribution (slope) of NumAromaticRings indicates that ligand aromaticity has positive overall effect on anti-TYK2 bioactivity. This trend is probably because TYK2 binding pocket includes several amino-acid residues capable of stacking against ligands' aromatic rings, e.g., the phenol of Tyr980, guanidine groups of Arg901, Arg1027 and carboxylate of Asp1041. Moreover, the binding pocket includes several peptidic amide groups that can also engage in stacking against ligand's aromatic rings.

Interestingly, QSAR Eq. (1) shows NumRings3 to be one of the significant positive contributors to anti-TYK2 bioactivity. The fact that this descriptor points to the number of 3-membered rings in ligand molecules indicate that cyclopropyl fragments are significantly involved in ligand binding within TYK2 binding pocket. Cyclopropyl groups are known to be involved in hydrogen bonding interactions with hydrogen bond donors (e.g., hydroxyls) (Joris et al. 1968). Accordingly, it can be concluded that cyclopropyl groups aid ligand-TYK2 binding via hydrogen bonding. However, since pharmacophore models generally fail to

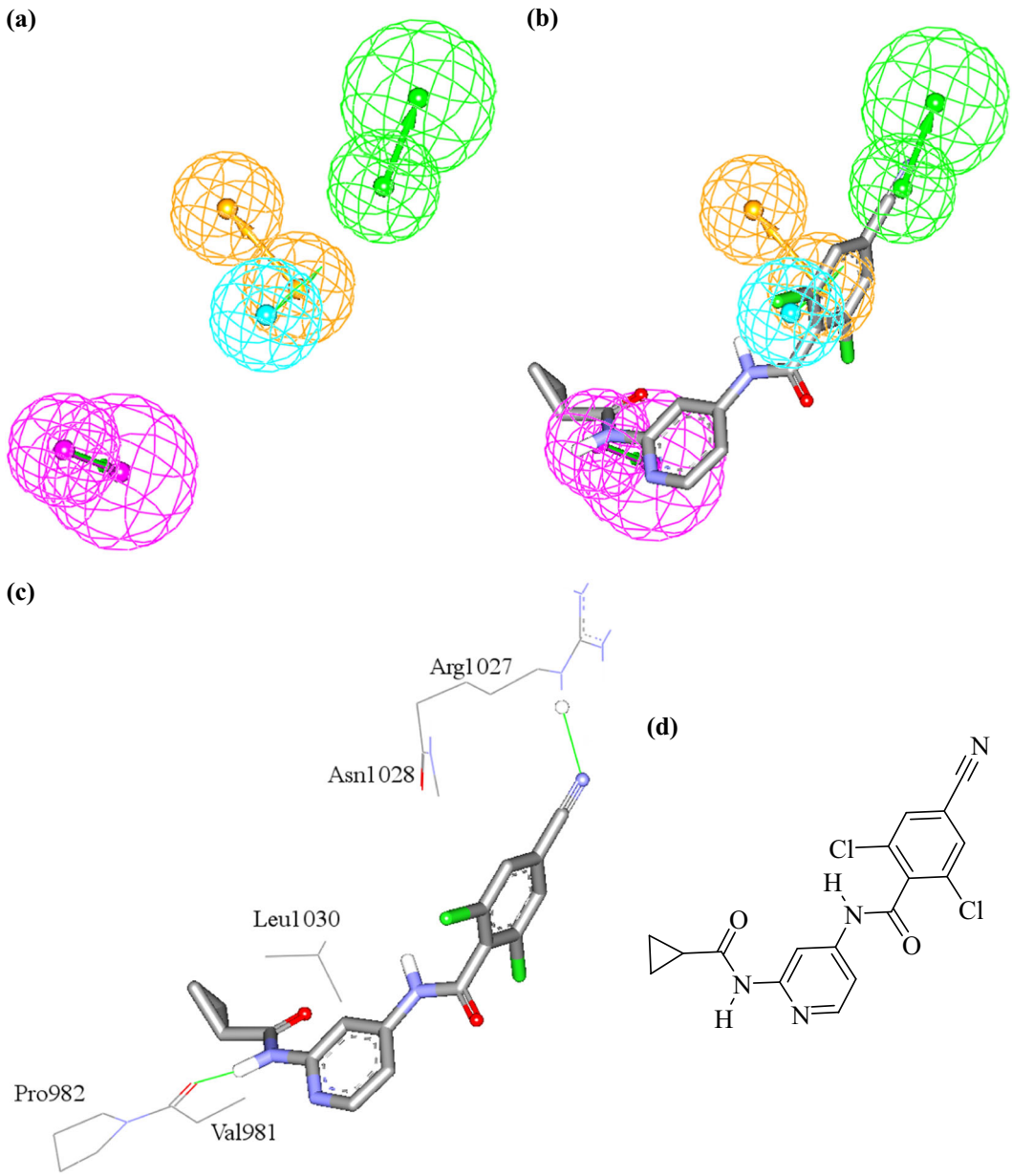


Fig. 7 **a** Pharmacophoric features of **Hypo(D-T5-10)**. HBA shown as green vectored spheres, Hbic shown as blue sphere, RingArom as orange vectored spheres, and HBD as pink vectored sphere. **b Hypo (D-T5-10)** fitted against the co-crystallized pose of inhibitor **5** (supporting Table S1, Ki = 1.8 nM) bound within TYK2 (PDB code: 4GII, resolution 2.31 Å). **c** The co-crystallized ligand **5** within the binding pocket of TYK2. The figure shows corresponding ligand-TYK2 interactions. Hydrogen bonds are shown as green lines. **d** The chemical structure of the co-crystallized ligand **5**

represent the hydrogen-bonding capacities of cyclopropyls, this forced GFA/MLR analysis to select the corresponding descriptor (i.e., NumRings3) in QSAR model (1) to explain the enhanced activities of molecules exhibiting this group compared to their counterparts lacking this feature.

Finally, emergence of JursTASA (indicates atomic neutrality) in Eq. (1) combined with negative slope is suggestive that anti-TYK2 bioactivity is generally more pronounced with ligands having molecular fragments of significant electronic imbalances (i.e., containing more dipolar groups) compared to their electronically-homogenous counterparts. This is not surprising as dipolar fragments can better interact with complementary binding site groups.

Although two pharmacophores emerged in Eq. (1), the two models are rather related as they were successfully merged in a single binding model (see section 4 and Fig. 9). Figures 4 and 5 show how the two models fit bound poses of two inhibitors co-crystallized within TYK2. Figure 4 shows how **Hypo(I-T7-5)** fits the co-crystallized pose of

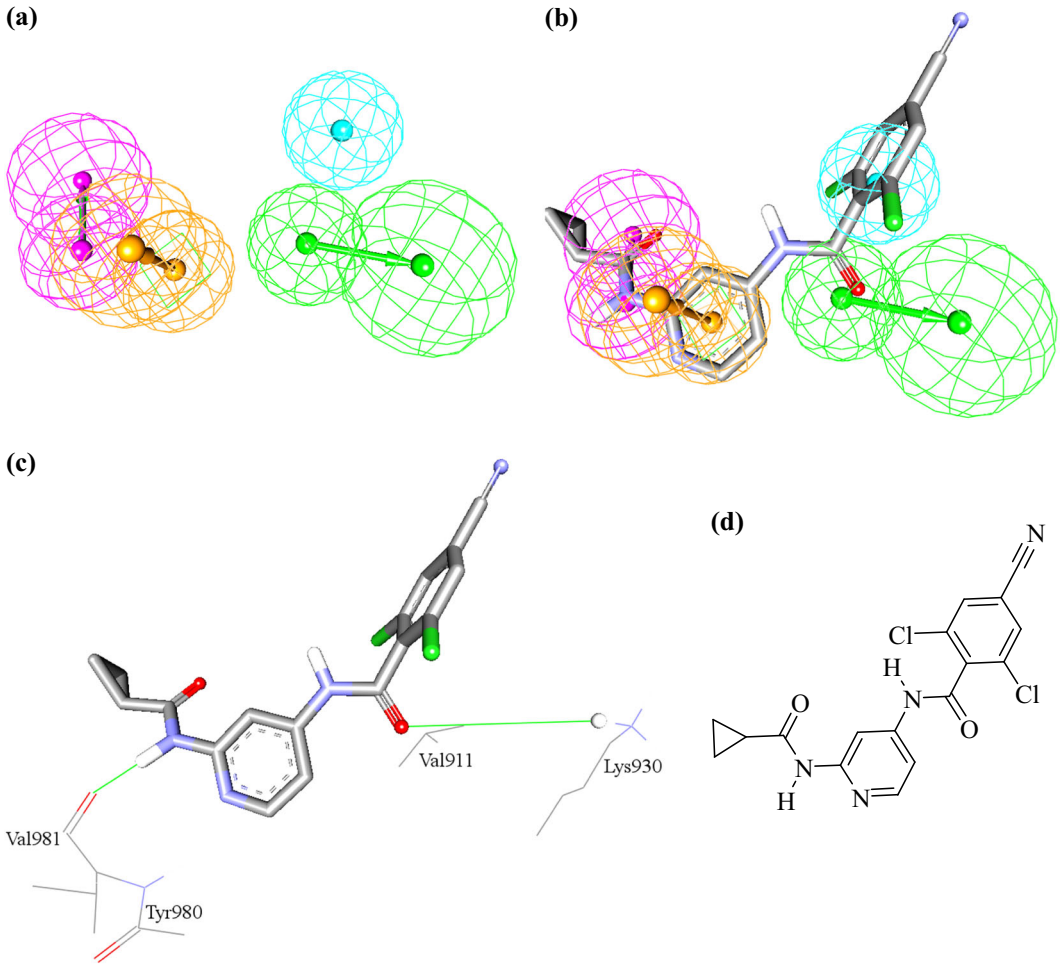


Fig. 8 **a** Pharmacophoric features of **Hypo(D-T7-8)**. HBA shown as green vectored spheres, Hbic shown as blue sphere, RingArom as orange vectored spheres, and HBD as pink vectored sphere. **b** **Hypo (D-T7-8)** fitted against the co-crystallized pose of inhibitor **5** (Table S1, Ki = 1.8 nM) bound within TYK2 (PDB code: 4GII, resolution 2.31Å). **c** The co-crystallized ligand **5** within the binding pocket of TYK2. The figure shows corresponding ligand-TYK2 interactions. Hydrogen bonds are shown as green lines. **d** The chemical structure of the co-crystallized ligand

inhibitor **11** (Supplementary Table S1) (Liang et al. 2013a), bound within the binding pocket of TYK2 (PDB code: 4GIH, resolution 2.0Å). Mapping the pyridine ring of the ligand against aromatic ring (RingArom) feature in **Hypo(I-T7-5)** (Fig. 4b) corresponds to stacking this group against the aromatic ring of Tyr980 in the binding pocket (Fig. 4c). Similarly, mapping the central amidic NH of the ligand against hydrogen bond donor (HBD) in **Hypo(I-T7-5)** (Fig. 4b) correlates with hydrogen bonding tying this fragment with carbonyl residue of Arg1027 via two bridging water molecules (Fig. 4c). Likewise, mapping the central amidic carbonyl of co-crystallized ligand against hydrogen bond acceptor (HBA) feature in **Hypo(I-T7-5)** agrees with hydrogen bonding connecting this group with the ammonium side chain of Lys930 (Fig. 4c). Finally, fitting one of the ligand's chloro substituents within a hydrophobic pouch composed of the side chain of Leu1030 and the α-carbon of Gly1040 (Fig. 4c) agrees with mapping the ligand's chlorine atom against hydrophobic (Hbic) feature in of **Hypo(I-T7-5)** (Fig. 4b).

Comparably, **Hypo(F-T5-8)** fitted the bound pose of another co-crystallized ligand within TYK2, as shown in Fig. 5. Mapping the thiophene ring of the ligand against RingArom feature in **Hypo(F-T5-8)** (Fig. 5b) agrees with π-stacking against the phenol side chain of Tyr962 (Fig. 5c). Similairly, the hydrogen bonding interactions connecting the ligand's carbonyl and thiadiazole fragments (in one tautomeric form) with the side chains of Lys930 and Try962, respectively (Fig. 5c), correlate with mapping these ligand fragments against HBA and HBD features, respectively (Fig. 5b). Likewise, the close proximity between the methylsulfide of Met978 and the quinoline fragment of the co-crystallized ligand (Fig. 5c) is indicative of mutual hydrophobic attraction, which was nicely represented

Table 3 Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of **Hypo(D-T5-10)** and **Hypo(D-T7-8)** selected by kNN-QSAR method

Model[a]	Definition	Chemical features							
			HBD		Hbic	RingArom		HBA	
Hypo(D-T5-10)[a]	Weight		1.81		1.81	1.81		1.81	
	Tolerances		1.60	2.20	1.60	1.60	1.60	1.60	2.20
	Coordinates	X	4.47	3.94	−1.85	−1.40	0.88	−2.97	−3.95
		Y	−3.05	−1.96	0.022	2.15	3.40	5.42	7.20
		Z	2.04	4.80	−2.59	−0.40	−1.90	−1.22	−3.44
			HBD		HBA		RingArom		Hbic
Hypo(D-T7-8)[b]	Weight		2.02		2.02		2.02		2.02
	Tolerances		1.60	2.20	1.60	2.20	1.60	1.60	1.60
	Coordinates	X	1.37	2.88	5.50	7.05	2.75	1.45	7.60
		Y	0.00	2.60	−4.66	−7.24	−2.42	−3.16	−3.98
		Z	0.00	0.00	0.55	0.47	0.001	2.60	2.80

[a]**Hypo(D-T5-10)** corresponds to the 10th pharmacophore model generated by subset D (Supporting Table S2), run number 5 (Supporting Tables S3 and S4). Model shown in Fig. 7a

[b]**Hypo(D-T7-8)** corresponds to the 8th pharmacophore model generated by subset D (Supporting Table S2), run number 7 (Supporting Tables S3 and S4). Model shown in Fig. 8a

Table 4 ROC performances of QSAR pharmacophores and their merged and satirically refined versions as 3D search queries using Testing Set 1 and Testing Set 2 lists

		Testing set 1				Testing set 2			
QSAR model	Pharmacophore model	ROC[a]-AUC[b]	ACC[c]	SPC[d]	TPR[e]	ROC[a]-AUC[b]	ACC[c]	SPC[d]	TPR[e]
MLR	**Hypo(I-T7-5)**	0.782	0.242	0.215	1.000	0.686	0.265	0.249	0.857
	Hypo(F-T5-8)	0.658	0.301	0.279	0.917	0.625	0.414	0.409	0.619
kNN	**Hypo(D-T5-10)**	0.779	0.584	0.584	0.583	0.589	0.449	0.447	0.524
	Hypo(D-T7-8)	0.951	0.264	0.238	1.000	0.710	0.257	0.237	1.000
Merged models	**Hypo(I-T7-5/F-T5-8)**	0.967	0.933	0.951	0.417	0.934	0.906	0.926	0.190
	Hypo(D-T5-10/D-T7-8)	0.955	0.919	0.939	0.333	0.911	0.873	0.889	0.286
	Refined Hypo(D-T5-10/D-T7-8)	0.984	0.955	0.977	0.333	0.949	0.923	0.947	0.048

[a]ROC: receiver operating characteristic

[b]AUC: area under the curve

[c]ACC: overall accuracy

[d]SPC: overall specificity

[e]TPR: overall true positive rate

through mapping part of the the quinoline ring against Hbic feature in **Hypo(F-T5-8)**, as in Fig. 5b.

kNN-based QSAR modeling

Since MLR-based QSAR analysis assumes the presence of linear correlation between generated molecular descriptors and corresponding bioactivities, we decided to evaluate the structure-bioactivity profile of the collected list using a nonlinear QSAR modeling methodology by coupling genetic function algorithm with k-nearest neighbor (kNN). Table 2 shows the selected descriptors, number nearest neighbors, and statistical criteria of highest-ranking kNN-based QSAR model and Fig. 6 shows the corresponding scatter plots of experimental versus estimated bioactivities for both training and testing inhibitors.

This example shows pharmacophore/QSAR modeling to yield reasonable models via linear and nonlinear approaches (Khanfar and Taha, 2013). Although kNN models fail to explicitly characterize the correlation nature between a particular molecular descriptor(s) in the model and ligands' bioactivities, the presence of certain molecular descriptor in the successful kNN-QSAR model is indicative of significant role played by that descriptor in ligands' bioactivity. Accordingly, emergence of descriptors related to cyano groups count (ESCount_tN), charged molecular surface

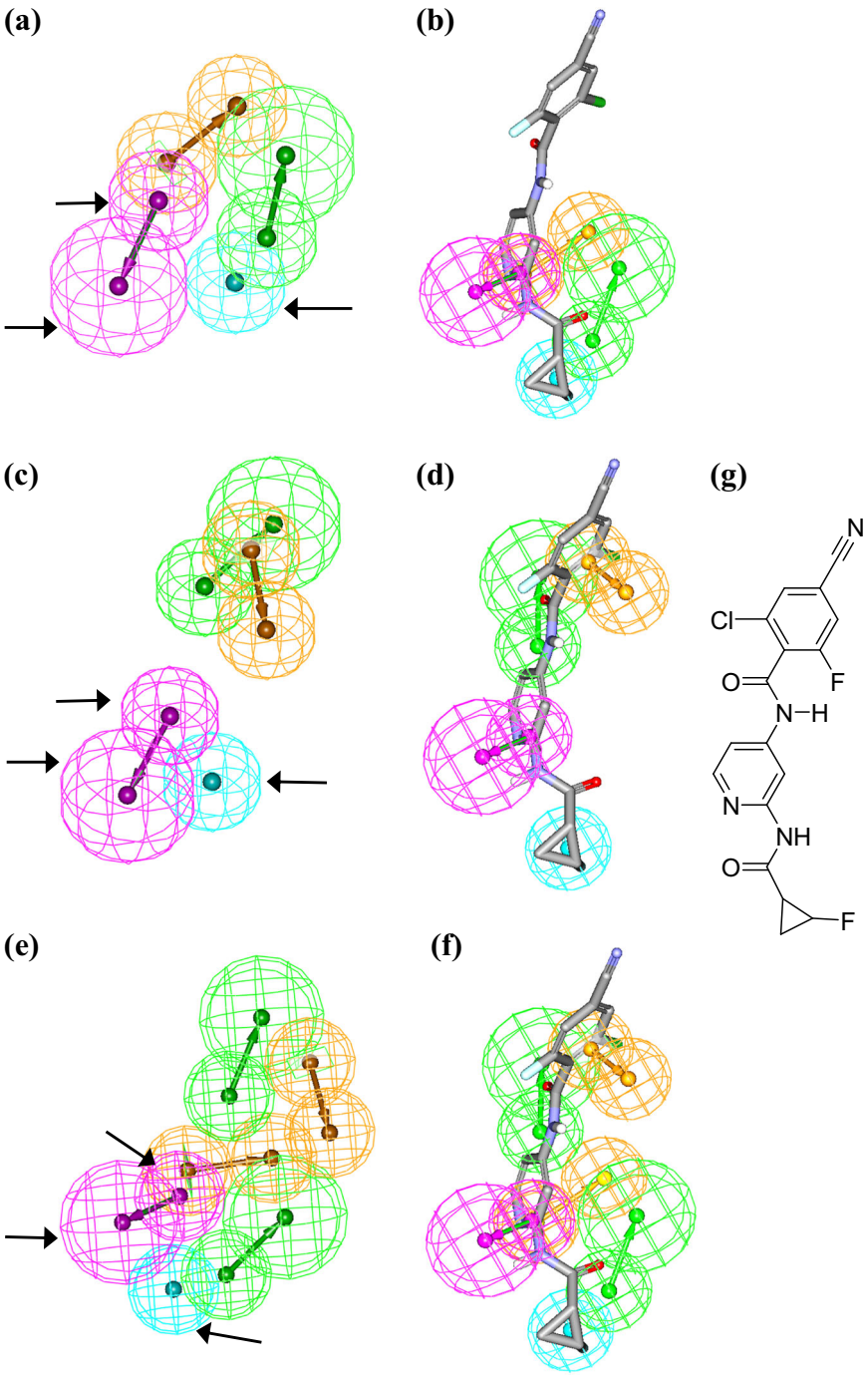


Fig. 9 Pharmacophoric features of (**a**) **Hypo(I-T7-5)**, (**c**) **Hypo (F-T5-8)**, and (**e**) Merged model **Hypo(I-T7-5/F-T5-8)**. Arrows point to closely positioned common features in **Hypo(I-T7-5)** and **Hypo(F-T5-8)** allowing merging into merged pharmacophore model **Hypo(I-T7-5/F-T5-8)**. (**b**), (**d**) and (**f**) represent the three pharmacophores fitted against **4** (Supporting Table S1, Ki = 1.6 nM), (**g**) Chemical structure of **4**. HBA as green vectored spheres, Hbic as blue spheres, and RingArom as orange vectored spheres, HBD as pink vectored spheres

area (JursDPSA3), ring and chain subgraph index (SC3CH), and molecular shadow (ShadowYZfrac) is suggestive of certain relationship connecting molecular shape and charge properties with anti-TYK2 bioactivity albeit of complex nature.

Two pharmacophore models emerged in the optimal kNN QSAR model (Table 2). The three-dimensional coordinates of these pharmacophores are shown in Table 3, while Figs 7 and 8 show how the two models map bound pose of inhibitor **5** (see supporting Table S1) within the binding pocket of TYK2 (PDB code: 4GII, resolution 2.31Å).

Figure 7 shows how **Hypo(D-T5-10)** fits the co-crystallized ligand **5**. Clearly from the Figure, mapping the dichlorobenzene ring of **5** against RingArom feature in **Hypo(D-T5-10)** (Fig. 7b) corresponds to π-stacking against the peptidic amidic bond connecting Asn1028 and Arg1027 in the binding pocket (Fig. 7c). In addition, mapping the chlorine atom against Hbic feature in **Hypo(D-T5-10)** (Fig. 7b) agrees with apparent hydrophobic attraction connecting

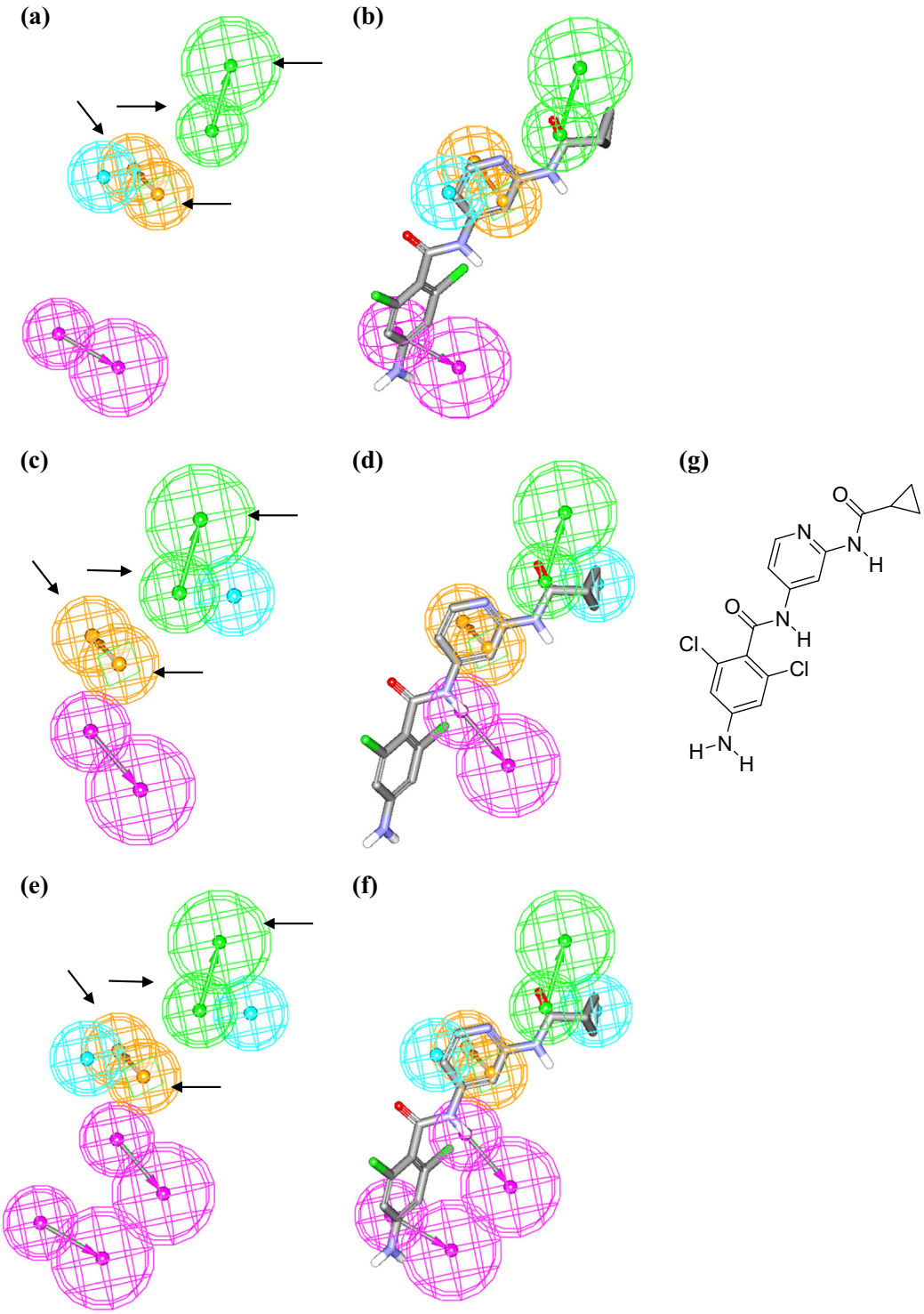


Fig. 10 Pharmacophoric features of (**a**) **Hypo(D-T5-10)**, (**c**) **Hypo(D-T7-8)**, and (**e**) Merged model **Hypo(D-T5-10/D-T7-8)**. Arrows point to closely positioned common features in **Hypo(D-T5-10)** and **Hypo(D-T7-8)** allowing merging into merged pharmacophore model **Hypo(D-T5-10/D-T7-8)**. (**b**), (**d**) and (**f**) represent the three pharmacophores fitted against **1** (Table 2, Ki = 1 nM), (**g**) Chemical structure of **1**. HBA as green vectored spheres, Hbic as blue spheres, and RingArom as orange vectored spheres, HBD as pink vectored spheres

the same chlorine atom with the hydrophobic side chain of Leu1030 (Fig. 7c). Similarly, mapping the terminal amidic NH of **5** against HBD in **Hypo(D-T5-10)** (Fig. 7b) correlates with hydrogen bonding tying this fragment with carbonyl of Val981 (Fig. 6c). Finally, mapping the cyano group of co-crystallized **5** against HBA feature in **Hypo(D-**

Table 5 Features and corresponding tolerances and 3D coordinates of merged pharmacophore models

Model	Definition	Chemical features										
		HBD		Hbic	RingArom		HBA		HBA		RingArom	
Hypo(I-T7-5/ F-T5-8)[a]	Tolerances	2.20	1.60	1.60	1.60	1.60	1.60	2.20	1.60	2.20	1.60	1.60
	Coordinates X	−1.80	0.47	3.27	2.87	2.81	2.67	2.23	1.69	2.60	−0.50	−2.38
	Y	−1.08	−0.64	−4.29	−0.10	0.11	−3.85	−1.94	2.57	5.11	3.81	1.55
	Z	−2.16	−0.23	−0.93	−0.02	2.97	1.01	3.28	1.76	3.09	4.85	5.43
Refined Hypo(D-T5-10/ D-T7-8)[b,c]		HBD		HBD	RingArom			Hbic	Hbic	HBA		
	Tolerances	1.60	2.20	1.60	2.20	1.60	1.60	1.60	1.60	2.20	1.60	
	Coordinates X	1.62	3.08	−2.36	−0.09	2.29	3.27	8.09	1.21	7.54	6.14	
	Y	−0.10	2.53	2.82	4.79	−3.61	−2.50	−4.06	−3.82	−7.40	−4.74	
	Z	−0.13	−0.12	1.43	1.45	2.09	−0.52	2.09	−2.32	−0.16	−0.16	

[a]Model Shown in Fig. 9

[b]Model Shown in Fig. 10

[c]Steric refinement added 4 exclusion spheres (of 1.20 Å radii) to this model at the following X, Y, and Z coordinates: (7.97, −4.36, −2.15), (10.00, −2.30, 1.20), (5.90, −1.30, 1.90), and (7.97, −3.34, 4.99)

T5-10) (Fig. 7b) agrees with hydrogen bonding interaction anchoring this group with the guanidine side chain of Arg1027 (Fig. 7c).

Comparably, **Hypo(D-T7-8)** fitted the bound pose of the same co-crystallized ligand (i.e, **5**) within TYK2 albeit representing different set of attractive interactions, as shown in Fig. 8. Mapping the pyridine ring of the ligand against RingArom feature in **Hypo(D-T7-8)** (Fig. 8b) agrees with π-stacking against the peptidic bond tying Tyr980 and Val981 (Fig. 8c). Similarly, hydrogen bonding interactions connecting the ligand's amides with the side chains of Lys930 and Val981 (Fig. 8c) correlate with mapping the same moieties against HBA and HBD features, respectively (Fig. 8b). Finally, fitting the ligand's dichlorobenzene against Hbic feature in **Hypo(D-T7-8)** (Fig. 8b) agrees with hydrophobic interaction with the nearby isopropyl side chain of Val911 (Fig. 8c).

Receiver operating characteristic curve analysis and merging of complementary pharmacophores

Receiver operating characteristic (ROC) analysis was used to assess the abilities of resulting pharmacophores to selectively capture diverse TYK2 inhibitors from a large pool of actives and decoys (Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c, Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016; Abutayeh and Taha, 2019; Khanfar and Taha, 2013; Triballeau et al. 2005). Two testing sets were used in the evaluation: Testing Set 1 composed of experimentally validated active TYK2 inhibitors and virtual decoys, while Testing Set 2 is entirely composed of experimentally validated active and inactive TYK2 inhibitors extracted from ChEMBL database. Table 4 summerizes the resulting ROC performances. Apparently, the pharmacophores showed mediocre classification powers particularly with Testing Set 2.

The mediocre performances of the resulting pharmacophores combined with apparent similarities among their binding features, prompted us to attempt merging analogous models, as in Figs 9 and 10. Needless to say that pharmacophores of more binding features should be more selective.

Careful assessment of **Hypo(I-T7-5)** and **Hypo(F-T5-8)** (Fig. 9) shows the two pharmacophores to share two features: hydrophobic and hydrogen bond donor (Fig. 9a, c). In fact, the close resemblance between the two models combined with their equivalent contributions to bioactivity (their slopes in QSAR Eq. (1)) suggested they represent the same ligand binding mode within TYK2. The two models were successfully merged in a single pharmacophore, namely, **Hypo(I-T7-5/F-T5-8)** (Fig. 9e). Similarly, the close similarity between kNN-QSAR selected pharmacophores, i.e., **Hypo(D-T5-10)** and **Hypo(D-T7-8)** (shared HBA and RingArom features) suggested they encode for the same binding mode and justified their merge in a single binding model, i.e., **Hypo(D-T5-10/D-T7-8)**, as in Fig. 10. Table 5 shows the X, Y, and Z coordinates of the merged models. As expected, both hybrid pharmacophores demonstrated superior ROC performances compared to their individual component pharmacophores, as in Table 4. This is rather anticipated since merged pharmacophores included more features and therefore are more selective (See Supporting Figs. S5 and S6).

Hypo(I-T7-5/F-T5-8) and Hypo(D-T5-10/D-T7-8) probably represent two distinct binding modes accessible to TYK2 ligands.

To further improve the ROC performances of merged models, we attempted to compliment them with exclusion spheres using the HIPHOP-REFINE module of Discovery

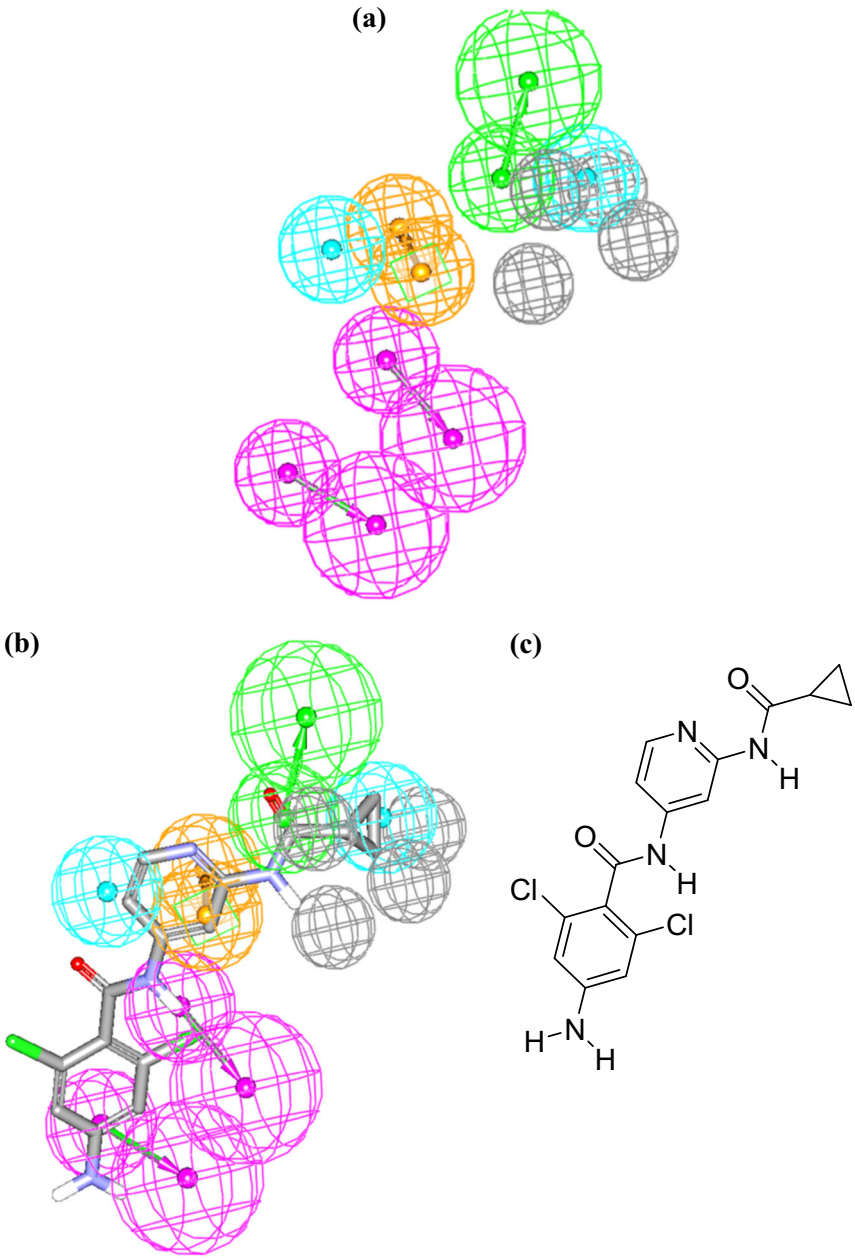


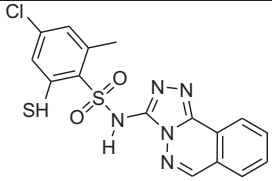
Fig. 11 **a** Sterically-refined version of hybrid pharmacophore **Hypo(D-T5-10/D-T7-8)**, (**b**) Sterically refined **Hypo(D-T5-10/D-T7-8)** fitted against **1** (Ki = 1, Table 2), (**c**) Chemical structure of **1**. HBA as green vectored spheres, Hbic as blue spheres, and RingArom as orange vectored spheres, HBD as pink vectored spheres

Studio (Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Khanfar and Taha, 2013). Exclusion spheres simulate the steric constrains of the binding pocket. However, our attempts to sterically-refine **Hypo(I-T7-5/F-T5-8)** were futile, probably because the poor affinities of inactive compounds are totally explainable by missing pharmacophoric features and not by steric clashes with the binding pocket. Still, fortunately, it was possible to add 4 exclusion spheres to **Hypo(D-T5-10/D-T7-8)**, as in Fig. 11, leading to significant ROC improvements, as in Table 4 and supporting Figs. S5 and S6.

In silico screening and subsequent in vitro evaluation

Hypo(I-T7-5/F-T5-8) and sterically refined **Hypo(D-T5-10/D-T7-8)** were employed as 3D search queries to screen the National Cancer Institute (NCI) list of compounds for new TYK2 inhibitors. Captured hits were filtered by Lipinski's criteria (Lipinski et al. 2012), and fitted against the corresponding component pharmacophores (i.e., **Hypo (I-T7-5)**, **Hypo(F-T5-8)**, **Hypo(D-T5-10)**, and **Hypo(D-T7-8)**). The fit values were substituted in MLR-QSAR Eq. (1) or kNN-based QSAR model (Table 2) for the determination of their predicted bioactivities.

Table 6 Active hit captured by merged pharmacophore **Hypo(I-T7-5/F-T5-8)** and fit values against component pharmacophores, QSAR predictions, and anti-TYK2 bioactivity

Hit[a]	NCI code	Structure[b,c,d]	Fit values against pharmacophores[e]		QSAR Predictions (nM)	Anti-TYK2 bioactivity	
			Hypo(F-T5-8)	Hypo(I-T7-5)		% Inhibition at 10 µM	IC_{50} (µM)[g]
72	696894		4.99	5.85	3.10[f]	73	7.1

[a]All captured hits are listed in Supplementary Table S7

[b]The chemical structures of other captured hits are in Supplementary Fig. S7

[c]Captured by hybrid pharmacophore model **Hypo(I-T7-5/F-T5-8)** shown in Fig. 9

[d]The identity and purity of **72** were established by ^{1}H, ^{13}CNMR and mass spectroscopy (see Supporting Figs. S8 to S10) (Brzozowski 1998)

[e]Fit values as calculated by Supporting equation S4 under Supporting section S3

[f]Predicted by QSAR Eq. (1)

[g]IC_{50} values determined as average of duplicate measurements at 6 concentration points as in Fig. 12. Staurosporine was used as positive control (IC50 of 1.03 nM) (Thermo Fisher Scientific 2016)

Highest ranking 77 hits were ordered from the NCI and their anti-TYK2 bioactivities evaluated in vitro at 10 µM. Supporting Fig. S7 and Table S7 show the chemical structures of tested hits, their fit values against corresponding pharmacophores, and experimental TYK2 inhibitory percentages at 10 µM. Table 6 shows the results of the most potent hit **72**, while Fig. 12 shows its dose-response curve. Interestingly, this active hit was captured by hybrid pharmacophore **Hypo(I-T7-5/F-T5-8)**. Figure 13 shows how **72** fits **Hypo(I-T7-5/F-T5-8)**.

Clearly from Fig. 13 **72** tightly fits the merged pharmacophore as will as component pharmacophores. Still, QSAR predictions of captured high-ranking hits (including **72**) deviate significantly from the corresponding experimentally-measured IC_{50} values (as in Table 6 and Supporting Table S7). This discrepancy is probably due to the significant structural difference between the training list used in pharmacophore and QSAR modeling compared to tested hits. Figure 14 shows the physicochemical space of modeled TYK2 inhibitors (**1**–**68**, Supporting Table S1) compared to that of the tested high-ranking hits (77 compounds including hit **72**) with respect to top three principal components based on 12 descriptors (i.e., LogP, molecular weight, hydrogen bond donors and acceptors, rotatable bonds, rings, aromatic rings, fractional polar surface area surface area, polar surface area and number of fragments). Clearly from Fig. 14, the tested hits occupy distinct physicochemical space from that of the training compounds, which should undermine the extrapolatory potential of the QSAR model and associated pharmacophore models. However, we were forced to use this particular list of training compounds because they are the largest set of TYK2 inhibitors that have their bioactivities determined by consistent bioassay procedure (Needless to say that proper QSAR and pharmacophore modeling necessitate the use of training compounds of consistent bioactivity measurements). Moreover, careful evaluation of structure-activity data available at ChEMBL database (Gaulton et al. 2017), including supporting table S1, shows the existence of several activity cliffs within TYK2 inhibitors population (examples shown in Fig. 15). Activity cliffs are closely similar structural analogs of significantly dissimilar inhibitory profiles (Stumpfe and Bajorath, 2012; Stumpfe et al. 2014; Guha and Medina-Franco, 2014; Dimova et al. 2015). QSAR and pharmacophore models rely on the fundamental assumption that bioactivity is a smooth function of structure (Husby et al. 2015; Furtmann et al. 2015). Accordingly, the exsitence of activity cliffs in a particular SAR data is detrimental to the predictive capacity of corresponding QSAR and pharmacophore models, which should add further error to the extrapolatory potential of our TYK2 QSAR models. Another reason for the observed experimental/predicted bioactivity difference is the variability in bioassay conditions of modeled compounds compared to captured hits. Nevertheless, success in capturing one micromolar bioactive hit from a list of 77 hits is considered excellent enrichment (Taha 2012; Abuhammad and Taha, 2016;

Taha et al. 2008a; Taha et al. 2008b; Taha et al. 2008c; Al-Nadaf et al. 2010; Taha et al. 2011; Aboalhaija et al. 2016; Alabed et al. 2016; Khanfar and Taha, 2013; Abutayeh and Taha, 2019).

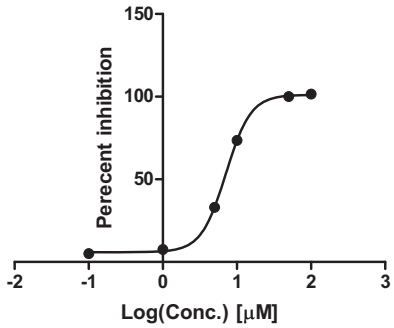


Fig. 12 Dose-response plot of Hit **72**. Each point represent the average of duplicate measurements

Conclusion

A ligand-based computational workflow was employed to explore the structural requirements for potent TYK2 inhibition using 68 known TYK2 inhibitors. Genetic function algorithm (GFA) was coupled with k nearest neighbor (kNN) or multiple linear regression (MLR) analyses to generate predictive QSAR models based on optimal combinations of pharmacophores and physicochemical descriptors. The optimal QSAR-selected pharmacophore models were validated by receiver operating characteristic (ROC) curve analysis and by comparison with crystallographic structures of known inhibitors co-crystallized within TYK2 binding pocket. Optimal QSAR models and their associated pharmacophore hypotheses were employed to identify new micromolar anti-TYK2 lead retrieved from the NCI database.

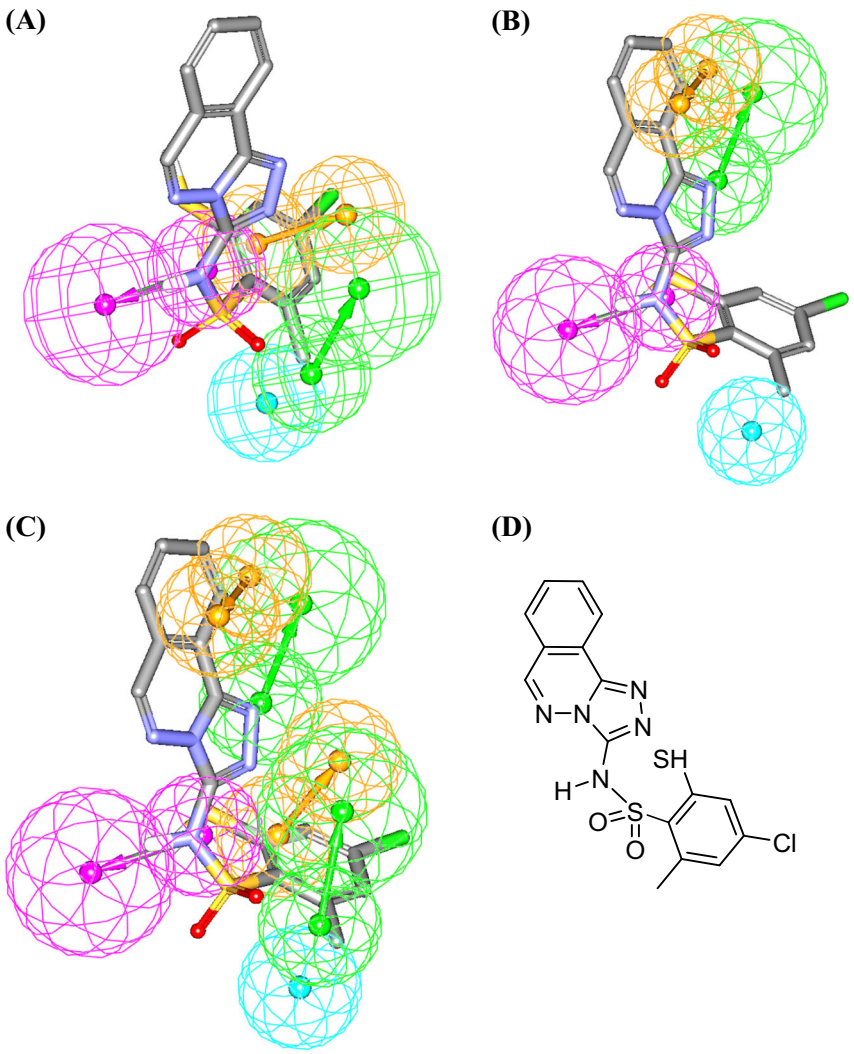


Fig. 13 Hit **72** (Table 5) fitted against (**a**) **Hypo(I-T7-5)**, (**b**) **Hypo(F-T5-8)**, and (**c**) Capturing merged pharmacophore **Hypo(I-T7-5/F-T5-8)**. HBA as green vectored spheres, Hbic as blue sphere, RingArom as orange vectored spheres, HBD as a pink sphere. **d** Chemical structure of **72**

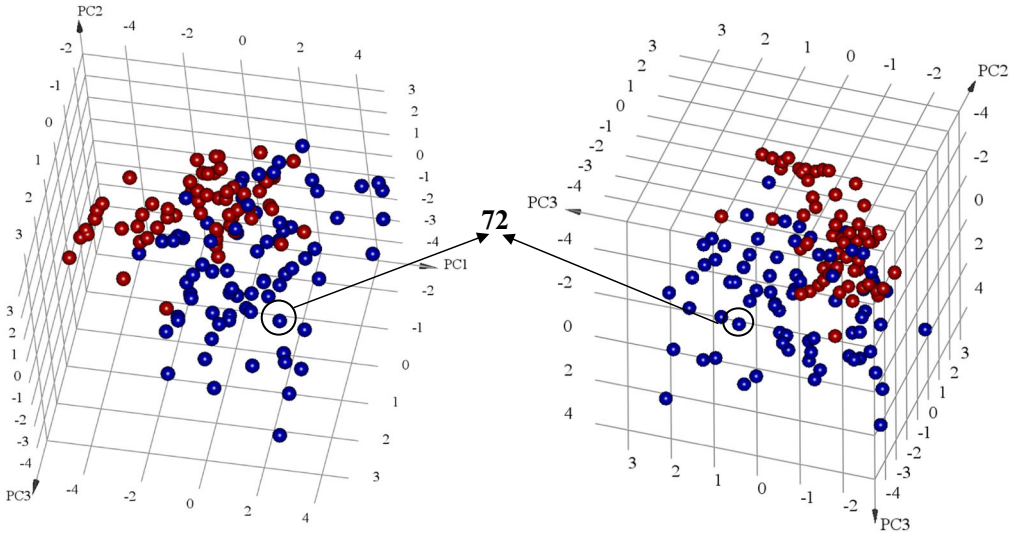


Fig. 14 Two views of 3D plot showing three main principal components calculated for the modeled training TYK2 inhibitors (**1**–**68**, Table S1) compared to those calculated for the 77 high-ranking hits captured by the resulting pharmacophores (based on 12 physicochemical descriptors, see text). Red spheres (●) represent modeled training inhibitors while blue spheres (●) represent high-ranking hits. Arrows point to the position of hit **72** within the 3D plot

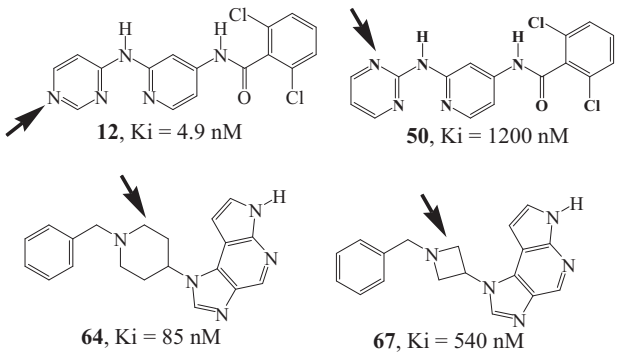


Fig. 15 Structure-activity data collected from supporting Table S1 showing activity cliffs in TYK2 training data. Arrows indicate positions of structural differences among activity cliff pairs

Acknowledgements The authors thank the Deanship of Scientific Research at the University of Jordan for funding this project.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Aboalhaija NH, Zihlif MA, Taha MO (2016) Discovery of new selective cytotoxic agents against Bcl-2 expressing cancer cells using ligand-based modeling. Chem-Biol Interact 250:12–26

Abuhammad A, Taha MO (2016) QSAR studies in the discovery of novel type-II diabetic therapies. Expert Opin Drug Discov 11:197–214

Abutayeh RF, Taha MO (2019) Discovery of novel Flt3 inhibitory chemotypes through extensive ligand-based and new structure-based pharmacophore modelling methods. J Mol Graph Model 88:128–151

Alabed SJ, Khanfar M, Taha MO (2016) Computer-aided discovery of new FGFR-1 inhibitors followed by in vitro validation. Future Med Chem 8:1841–1869

Al-Nadaf A, Sheikha GA, Taha MO (2010) Elaborate ligand-based pharmacophore exploration and QSAR analysis guide the synthesis of novel pyridinium-based potent β-secretase inhibitory leads. Bioorganic. Med Chem 18:3088–3115

Beeley N, Sage C (2003) GPCRs: an update on structural approaches to drug discovery. Targets 2:19–25

Bernstein CN, Wajda A, Blanchard JF (2005) The clustering of other chronic inflammatory diseases in inflammatory bowel disease: a population-based study. Gastroenterology 129:827–836

Blench TJ, Ellwood C, Goodacre SC, Lai Y, Liang J, Macleod C, Magnuson SR, Tsui VH, Williams K, Zhang B (2012) Aza-benzothiazole compounds, compositions and methods of use, US Patent 2012/O2O2788, filled March 22, 2012, issued January 8, 2013

Blench T, Goodacre S, Lai Y, Liang Y, Macleod C, Magnuson S, Tsui V, Williams K, Zhang B (2012) Pyrazolopyridines and pyrazolopyridines and their use as TYK2 inhibitors. WO Patent Appl. WO2012/066061, filled May 24, 2012, issued May 19, 2013.

Brzozowski Z (1998) 2-Mercapto N-(azolyl)benzenesulfonamides. VI. Synthesis and anti-HIV activity of some new 2-mercapto-N-(1,2,4-triazol-3-yl)benzenesulfonamide derivatives containing the 1,2,4-triazole moiety fused with a variety of heteroaromatic rings. Acta Pol Pharm 55:473–480

Chrencik J, Patny A, Leung I, Korniski B, Emmons T, Hall T, Weinberg R, Gormley J, Williams J, Day J, Hirsch J, Kiefer J, Leone J, Fischer H, Sommers C, Huang H, Jacobsen E, Tenbrink R, Tomasselli A, Benson T (2010) Structural and thermodynamic characterization of the TYK2 and JAK3 kinase domains in complex with CP-690550 and CMP-6. J Mol Biol 400:413–433

Dassault Systèmes BIOVIA (2009) Discovery studio modeling environment, Release 2.5.5, San Diego, USA

De Mesquita MB, Ferrante M, Henckaerts L, Joossens M, Janssens V, Hlavaty T, Pierik M, Joossens S, Van Schuerbeek N, Van Assche

G (2009) Clustering of (auto) immune diseases with early-onset and complicated inflammatory bowel disease. Eur J Pediatr 168:575–583

Depristo M (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure 12:831–838

Dimova D, Stumpfe D, Hu Y, Bajorath J (2015) Activity cliff clusters as a source of structure–activity relationship information. Expert Opin Drug Discov 10:441–447

Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 314:1461–1463

Ellinghaus D, Ellinghaus E, Nair RP, Stuart PE, Esko T, Metspalu A, Debrus S, Raelson JV, Tejasvi T, Belouchi M (2012) Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. Am J Hum Genet 90:636–647

Fensome A, Ambler C, Arnold E, Banker M, Brown M, Chrencik J, Clark J, Dowty M, Efremov I, Flick A, Gerstenberger B, Gopalsamy A, Hayward M, Hegen M, Hollingshead B, Jussif J, Knafels J, Limburg D, Lin D, Lin T, Pierce B, Saiah E, Sharma R, Symanowicz P, Telliez J, Trujillo J, Vajdos F, Vincent F, Wan Z, Xing L, Yang X, Yang X, Zhang L (2018) Dual inhibition of TYK2 and JAK1 for the treatment of autoimmune diseases: discovery of ((S)-2,2-Difluorocyclopropyl)((1R,5S)-3-(2-((1-methyl-1H-pyrazol-4-yl)amino)pyrimidin-4-yl)-3,8-diazabicyclo[3.2.1]octan-8-yl) methanone (PF-06700841). J Med Chem 61:8597–8612

Fisher R (1971) The principle of experimentation illustrated by a psycho-physical experiment. The Design of Experiments. Hafner Publishing Company, New York

Frush E, Sekharan S, Keinan S (2017) In silico prediction of ligand binding energies in multiple therapeutic targets and diverse ligand sets—a case study on BACE1, TYK2, HSP90, and PERK proteins. J Phys Chem B 121:8142–8148

Furtmann N, Hu Y, Gütschow M, Bajorath J (2015) Identification of interaction hot spots in structures of drug targets on the basis of three-dimensional activity cliff information. Chem Biol Drug Des 86:1458–1465

Galatsis P, Henderson J, Kormos B, Han S, Kurumbail R, Wager T, Verhoest P, Noell G, Chen Y, Needle E, Berger Z, Steyn S, Houle C, Hirst W (2014) Kinase domain inhibition of leucine rich repeat kinase 2 (LRRK2) using a [1,2,4]triazolo[4,3-b]pyridazine scaffold. Bioorganic Med Chem Lett 24:4132–4140

Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, Davies M, Dedman N, Karlsson A, Magariños MP, Overington JP, Papadatos G, Smit I, Leach AR (2017) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954

Gillett N, Chan C (2000) Applications of immunohistochemistry in the evaluation of immunosuppressive agents. Hum Exp Toxicol 19:251–254

Guha R, Medina-Franco JL (2014) On the validity versus utility of activity landscapes: are all activity cliffs statistically significant? J Cheminform 6:11

Hamaguchi H, Amano Y, Moritomo A, Shirakami S, Nakajima Y, Nakai K, Nomura N, Ito M, Higashi Y, Inoue T (2018) Discovery and structural characterization of peficitinib (ASP015K) as a novel and potent JAK inhibitor. Bioorganic Med Chem 26:4971–4983

Hernández-Flórez D, Valor L (2016) Protein-kinase inhibitors: a new treatment pathway for autoimmune and inflammatory diseases? Reumatol Clin 12:91–99

Homans SW (2007) Water, water everywhere—except where it matters? Drug Discov Today 12:534–539

Husby J, Bottegoni G, Kufareva I, Abagyan R, Cavalli A (2015) Structure-based predictions of activity cliffs. J Chem Inf Model 55:1062–1076

Joris L, Rague Schleyer P, Gleiter R (1968) Cyclopropane rings as proton-acceptor groups in hydrogen bonding. J Am Chem Soc 90:327–336

Kettle JG, Åstrand A, Catley M, Grimster NP, Nilsson M, Su Q, Woessner R (2017) Inhibitors of JAK-family kinases: an update on the patent literature 2013-2015, Part 1. Expert Opin Ther Pat 27:127–143

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 53:2587–2612

Klebe G (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today 11:580–594

Kulagowski JJ, Blair W, Bull RJ, Chang C, Deshmukh G, Dyke HJ, Eigenbrot C, Ghilardi N, Gibbons P, Harrison TK (2012) Identification of imidazo-pyrrolopyridines as novel and potent JAK1 inhibitors. J Med Chem 55:5901–5921

Lees C, Barrett J, Parkes M, Satsangi J (2011) New IBD genetics: common pathways with other diseases. Gut 60:1739–1753

Lerner A, Jeremias P, Matthias T (2015) The world incidence and prevalence of autoimmune diseases is increasing. Int J Celiac Dis 3:151–155

Liang J, Tsui V, Van Abbema A, Bao L, Barrett K, Beresini M, Berezhkovskiy L, Blair WS, Chang C, Driscoll J, Eigenbrot C, Ghilardi N, Gibbons P, Halladay J, Johnson A, Kohli PB, Lai Y, Liimatta M, Mantik P, Menghrajani K, Murray J, Sambrone A, Xiao Y, Shia S, Shin Y, Smith J, Sohn S, Stanley M, Ultsch M, Zhang B, Wu LC, Magnuson S (2013a) Lead identification of novel and selective TYK2 inhibitors. Eur J Med Chem 67:175–187

Liang J, Van Abbema A, Balazs M, Barrett K, Berezhkovsky L, Blair W, Chang C, Delarosa D, Devoss J, Driscoll J (2013b) Lead optimization of a 4-aminopyridine benzamide scaffold to identify potent, selective, and orally bioavailable TYK2 inhibitors. J Med Chem 56:4521–4536

Liang J, Van Abbema A, Balazs M, Barrett K, Berezhkovsky L, Blair W, Chang C, Delarosa D, Devoss J, Driscoll J, Eigenbrot C, Goodacre S, Ghilardi N, Macleod C, Johnson A, Bir Kohli P, Lai Y, Lin Z, Mantik P, Menghrajani K, Nguyen H, Peng I, Sambrone A, Shia S, Smith J, Sohn S, Tsui V, Ultsch M, Williams K, Wu L, Yang W, Zhang B, Magnuson S (2017) Identification of an imidazopyridine scaffold to generate potent and selective TYK2 inhibitors that demonstrate activity in an in vivo psoriasis model. Bioorganic Med Chem Lett 27:4370–4376

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 64:4–17

Lupardus P, Ultsch M, Wallweber H, Bir Kohli P, Johnson A, Eigenbrot C (2014) Structure of the pseudokinase-kinase domains from protein kinase TYK2 reveals a mechanism for Janus kinase (JAK) autoinhibition. Proc Natl Acad Sci USA 111:8025–8030

Ma H, Deacon S, Horiuchi K (2008) The challenge of selecting protein kinase assays for lead discovery optimization. Expert Opin Drug Discov 3:607–621

Ma H-L, Liang S, Li J, Napierata L, Brown T, Benoit S, Senices M, Gill D, Dunussi-Joannopoulos K, Collins M (2008) IL-22 is required for Th17 cell–mediated pathology in a mouse model of psoriasis-like skin inflammation. J Clin Invest 118:597–607

Matusevicius D, Kivisäkk P, He B, Kostulas N, Özenci V, Fredrikson S, Link H (1999) Interleukin-17 mRNA expression in blood and

CSF mononuclear cells is augmented in multiple sclerosis. Mult Scler 5:101–104

Moslin R, Gardner D, Santella J, Zhang Y, Duncia J, Liu C, Lin J, Tokarski J, Strnad J, Pedicord D, Chen J, Blat Y, Zupa-Fernandez A, Cheng L, Sun H, Chaudhry C, Huang C, D'arienzo C, Sack J, Muckelbauer J, Chang C, Tredup J, Xie D, Aranibar N, Burke J, Carter P, Weinstein D (2017) Identification of imidazo[1,2-b] pyridazine TYK2 pseudokinase ligands as potent and selective allosteric inhibitors of TYK2 signalling. MedChemComm 8:700–712

Norman P (2012) Selective JAK1 inhibitor and selective TYK2 inhibitor patents. Expert Opin Ther Pat 22:1233–1249

Patterson H, Nibbs R, Mcinnes I, Siebert S (2014) Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases. Clin Exp Immunol 176:1–10

Roskoski R (2016) Janus kinase (JAK) inhibitors in the treatment of inflammatory and neoplastic diseases. Pharmacol Res 111:784–803

Shahin R, AlQtaishat S, Taha MO (2012) Elaborate ligand-based modeling reveal new submicromolar Rho kinase inhibitors. J Comput Aided Mol Des 26:249–266

Sigurdsson S, Nordmark G, Göring HH, Lindroos K, Wiman A-C, Sturfelt G, Jönsen A, Rantapää-Dahlqvist S, Möller B, Kere J (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet 76:528–537

Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Rönnelid J, Klareskog L, Huizinga TW (2007) Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. Arthritis Rheum 56:2202–2210

Srivastava A, Nagai T, Srivastava A, Miyashita O, Tama F (2018) Role of COmputational Methods in Going Beyond X-ray Crystallography to Explore Protein Structure and Dynamics. Int J Mol Sci 19:3401

Stark GR, Kerr IM, Williams BR, Silverman RH, Schreiber RD (1998) How cells respond to interferons. Annu Rev Biochem 67:227–264

Steuber H, Zentgraf M, Gerlach C, Sotriffer C, Heine A, Klebe G (2006) Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. J Mol Biol 363:174–187

Stubbs M, Reyda S, Dullweber F, Möller M, Klebe G, Dorsch D, Mederski W, Wurziger H (2002) pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. Chembiochem 3:246–249

Stumpfe D, Bajorath J (2012) Exploring activity cliffs in medicinal chemistry: miniperspective. J Med Chem 55:2932–2942

Stumpfe D, Dimova D, Bajorath J (2014) Composition and topology of activity cliff clusters formed by bioactive compounds. J Chem Inf Model 54:451–461

Taha MO (2012) Mixing pharmacophore modeling and classical QSAR analysis as powerful tool for lead discovery. In: Taha MO (ed) Virtual screening. InTech, Shanghai, China

Taha MO, Atallah N, Al-Bakri AG, Paradis-Bleau C, Zalloum H, Younis KS, Levesque RC (2008a) Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening. Bioorg Med Chem 16:1218–1235

Taha MO, Bustanji Y, Al-Ghussein MA, Mohammad M, Zalloum H, Al-Masri IM, Atallah N (2008b) Pharmacophore modeling, quantitative structure–activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3β inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. J Med Chem 51:2062–2077

Taha MO, Dahabiyeh LA, Bustanji Y, Zalloum H, Saleh S (2008c) Combining ligand-based pharmacophore modeling, quantitative structure−activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors. J Med Chem 51:6478–6494

Taha MO, Qandil AM, Al-Haraznah T, Khalaf RA, Zalloum H, Al-Bakri AG (2011) Discovery of New Antifungal Leads via Pharmacophore Modeling and QSAR analysis of fungal n-myristoyl transferase inhibitors followed by in silico screening. Chem Biol Drug Des 78:391–407

Thermo Fisher Scientific (2016) SelectScreen™ Biochemical Kinase Profiling Service Z′-LYTE™ Screening Protocol and Assay Conditions. 1–38

Tokarski J, Zupa-Fernandez A, Tredup J, Pike K, Chang C, Xie D, Cheng L, Pedicord D, Muckelbauer J, Johnson S, Wu S, Edavettal S, Hong Y, Witmer M, Elkin L, Blat Y, Pitts W, Weinstein D, Burke J (2015) Tyrosine kinase 2-mediated signal transduction in t lymphocytes is blocked by pharmacological stabilization of its pseudokinase domain. J Biol Chem 290:11061–11074

Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O (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 48:2534–2547

Tsui V, Gibbons P, Ultsch M, Mortara K, Chang C, Blair W, Pulk R, Stanley M, Starovasnik M, Williams D (2011) A new regulatory switch in a JAK protein kinase. Proteins 79:393–401

Whitley D (1994) A genetic algorithm tutorial. Stat Comput 4:65–85

Yogo T, Nagamiya H, Seto M, Sasaki S, Shih-Chung H, Ohba Y, Tokunaga N, Lee G, Rhim C, Yoon C, Cho S, Skene R, Yamamoto S, Satou Y, Kuno M, Miyazaki T, Nakagawa H, Okabe A, Marui S, Aso K, Yoshida M (2016) Structure-based design and synthesis of 3-amino-1,5-dihydro-4h-pyrazolopyridin-4-one derivatives as tyrosine kinase 2 inhibitors. J Med Chem 59:733–749

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