Traditional Chinese Medicine as Dual Guardians against Hypertension and Cancer?

Weng Ieong Tou[1, 2], Calvin Yu-Chian Chen[2, 3, 4, 5, 6*]

[1]*School of Medicine, China Medical University, Taichung, 40402, Taiwan*

[2]*Laboratory of Computational and Systems Biology, China Medical University, Taichung, 40402, Taiwan.*

[3]*Department of Medical Research, China Medical University Hospital, Taichung, 40402, Taiwan.*

[4]*Department of Biotechnology, Asia University, Taichung, 41354, Taiwan.*

[5]*Department of Biomedical Informatics, Asia University, Taichung, 41354, Taiwan.*

[6]*China Medical University Beigang Hospital, Yunlin,65152, Taiwan.*

[*]Corresponding author. Tel.: +886-4-2205-3366 ext. 4124

E-mail address: ycc929@MIT.EDU (C.Y.-C. Chen)

Abstract

This study utilizes the comprehensive traditional Chinese medicine database TCM Database@Taiwan (http://tcm.cmu.edu.tw/) in conjunction with structure-based and ligand-based drug design to identify multi-function Src inhibitors. The three potential TCM candidates identified as having suitable docking conformations and bioactivity profiles were angeliferulate, (3R)-2'-hydroxy-3',4'-dimethoxyisoflavan-7-O-beta-D-glucoside(HMID) and 3-[2',6-dihydroxy-5'-(2-propenyl)[1,1'-biphenyl]3-yl]-(E)-2-propenoic acid(3PA). Molecular dynamics (MD) simulation demonstrated that the TCM candidates have more stable interactions with the cleft and in complex with Src kinase compared to Saracatinib. Angeliferulate and HMID, both originated from *Angelica sinensis*, not only interact with Lys298 and amino acids from different loops in the cleft, but with Asp407 located on the activation loop. These interactions are important to reduce the opening of the activation loop due to phosphorylation, hence stabilize the Src kinase cleft structure and inhibit activation. The TCM candidates also exhibited high affinity to other cancer-related target proteins (EGFR, HER2, HSP90). Our observations suggest that the TCM candidates might have multi-targeting effects in hypertension and cancer.

Keywords: cancer; Src; traditional Chinese medicine (TCM); docking; QSAR; molecular dynamics (MD)

Introduction

Src-family protein kinases are proto-oncogenes that participate in important physiological functions such as cellular differentiation, mobility, and proliferation. Src, a member of the Src-family protein kinases, functions as a signal protein and is implicated in various diseases. Src is ubiquitously expressed within the body, but distribution levels vary depending on individual tissue and organs (1,2). The role of Src in cancer was initially established by its overexpression in colon (3-7) and breast cancer (8-20) and is now well established. Figure 1 illustrates the major cancer types in which Src plays a significant role, and important substrates and factors involved in each pathway (21-39). The critical role of Src in cancer makes it an attractive target for designing novel cancer therapy.

The carboxyl terminal of the Src kinase (CSK) is important in regulating conformation and activity of Src. Under normal inactive conditions, the Src protein is locked as an inward-folding conformation through binding between the phosphorylated Tyr527 and the SH2 domain. However, in mitotic cells, Src becomes activated due to thedephosphorylation of Tyr527 and phosphorylation of Tyr419.

From N-terminal to C-terminal, Src is composed of a smaller amino-terminal lobe (residues 270-340) which binds ATP, and a larger carboxyl-terminal lobe (residues 345-523) which binds with substrates. The ATP binding site is also partially located in the larger lobe. By regulating the structure of the alpha-helix, the large lobe can move toward or away from the small lobe, resulting in opening or closing of the cleft between the two lobes. The Src catalytic site is located within the cleft. An open conformation allows the entrance of ATP into the cleft and exit of ADP from the cleft. Drugs that can either interact with the residues (404-432) on the

activation loop or inhibit the activation loop from moving away and opening the cleft as a result of Tyr419 phosphorylation can effectively inhibit Src activity.

The importance of Src is established in cancer, but increasing studies are also suggesting its possible involvement in hypertension (40).Based on the pathophysiology of hypertension, endogenous ouabain (EO) and mutant α-adducin have been shown to increase binding of Na-K-ATPase with the SH2 domain and phosphorylation of Tyr416, thus activating subsequent Na,K-ATPase-Src-EGFR-ERK1/2 signaling cascade (41-43). As Src kinase is a key protein regulating the activation of downstream pathways, inhibiting Src kinase activation and Tyr416 phosphorylation may be a novel approach to controlling hypertension. Observations showing that lung cancer patients also have the tendency to develop hypertension and other cardiovascular diseases (44) also suggest possible linkage between the two diseases. In view of this, we speculate that Src may be a common factor for the development of hypertension and cancer, and inhibiting Src may have multi-targeting effects for both diseases. Though many Src inhibitors have been developed, limited pharmaceutical effectiveness has been observed. Recently, computational techniques are commonly used to screen or predict potent drugs targeting specific diseases (45-58).In view of this shift in drug design, and considering the vast pharmaceutical potential of traditional Chinese medicine (TCM), we constructed the world's most comprehensive TCM database TCM Database@Taiwan (http://tcm.cmu.edu.tw/) (59,60) and its accompanying cloud-computing webserver iScreen (61) and integrative computational design portal "**i**ntegrated **S**yste**M**s Biology **A**ssociated **R**esearch with **T**CM" (iSMART) (62) to facilitate drug development from TCM. TCM Database@Taiwan has been successfully used to identify lead compounds for a variety of important diseases

(63-82). In our previous study, we utilized TCM Database@Taiwan to identify potential Src inhibitors (83). Based on the essential role of Src in cancer, we deemed it important to evaluate the effect of TCM compounds with Src inhibitory potential against other cancer target proteins. This study targets not only Src, but also validates the docking potential of these TCM candidates against established cancer targets EGFR (84,85), HER2 (86,87), and HSP90 (88-91). Since Src is also involved in the pathological mechanisms of EGFR and HER2 (92,93), ourTCM candidates may have potential as multi-targeting inhibitors for different cancers.

Materials and Methods

Docking and Candidate Screening

The protein structure used in this study was downloaded from Protein Data Bank (PDB: 2H8H) (94). The binding site was defined as the space occupied by Saracatinib within the 2H8H crystal structure, and TCM compounds from TCM Database@Taiwan docked and screened. Using Discovery Studio 2.5, the ligands were first passed through Lipinski's Rule of Five, and then screened for contour of TCM ligands with the Src kinase binding site using LigandFit. LigandFit is a receptor-rigid docking algorithm that uses Monte Carlo simulation to match ligands with designate binding sites on a given protein. Results from docking are ranked by binding energy and ligand similarity to the Saracatinib within the 2H8H crystal structure, and three candidates are selected for further analysis. Each candidate ligand was minimized with smart minimizer setting in Minimization Algorithm under the force field of CHARMm (95), and a maximum of five docking poses were generated. Each generated docking pose was redocked into the protein for a second time, and five poses from each re-dock were generated. The 25 poses for each ligand are then visually compared with that of Saracatinib within the crystal

structure, and docking poses that are most similar to Saracatinib were selected. Hydrophobic contacts between the ligands and Src kinase are calculated with LigPlot v.2.2.25(96). The absorption, distribution, metabolism, excretion and toxicity (ADMET) analysis in D.S. 2.5 was used to calculate pharmacology and toxicity of the derivatives in human bodies. To assess the general applicability of the candidates on other cancer target proteins, the top three candidates from screening were further docked into EGFR, HER2, and HSP. Protein structures used for this application spectrum verification were EGFR (PDB: 2ITY) (97) and HSP90 (PDB: 3K97) (98). The HER2 model used was adopted from our previous study (72,99) and was built from structures 2ITY and 2J5E (97,100).

Bioactivity Prediction by Multiple Linear Regression (MLR) and Support Vector Machine (SVM) Models

Linear MLR (101) and nonlinear SVM (102) were used to construct quantitative structure-activity relationship (QSAR) models for predicting the bioactivity of the TCM compounds. A total of 20 Src ligands (103) were randomly separated into a training set of 15 compounds and an external validation test set of five compounds.

Prior to constructing QSAR models, genetic function approximation (GFA) (104) was applied to identify representative descriptor sets from the large pool of descriptors generated from the training set. The MLR model was built based on the representative descriptors using MATLAB in the form of **equation [1]**:

$$pIC_{50} = \alpha_0 + \sum_{1}^{n} \alpha_n x_n \qquad \textbf{[1]}$$

where α_0 is a constant value and α_n is the coefficient value of descriptor X_n.

The generated MLR model was validated with cross-validation and independent

tests and verified by calculating the square correlation coefficients (R^2) between predicted and actual pIC_{50} of the training set.

SVM are groups of supervised methods that allow categorization of hard-to separate patterns through the use nonlinear of generalized portrait algorithms (105). The SVM algorithm was extended for regression (SVMR) from its original use for classification through the use of a ε-insensitive loss function (106) with the goal of identifying a function *f(x)* in which all training points has a maximum deviation ε from experimental values and has a maximum margin (107). A final nonlinear SVM regression giving the modeled property for a pattern *x* was obtained by introducing Lagrange multipliers and kernels to map input patterns into a higher dimension space, the formula being equation **[2]**:

$$f(x_k) = \sum_{i=1}^{m} \left(\lambda_i^- - \lambda_i^+\right) K\left(x_i, x_k\right) + b \qquad \textbf{[2]}$$

where λ_i^-, λ_i^+ are Lagrange multipliers and $K(x_i, x_k)$ is the kernel function.

Our SVM model was constructed using the LibSVM program(108). Key parameters determining the SVM model fit are C cost, ε, γ, the kernel type, and the corresponding kernel parameters. The kernel selected for training the SVM model was the Gaussian radial basis function kernel equation **[3]**:

$$K\left(x_i, x_k\right) = \exp\left(-\frac{\left\|x - y\right\|^2}{2\sigma^2}\right) \qquad \textbf{[3]}$$

Optimum C, ε, γ were determined using the gridregression.py command within LibSVM. Cross-validation was conducted according to default settings in LibSVM. The validated MLR and SVM models to the TCM candidates to predict individual pIC_{50} values of the compounds.

3D-QSAR Modeling and Analysis

123 3D-QSAR methods such as comparative molecular field analysis (CoMFA)
124 (109) and comparative molecular similarity analysis (CoMSIA) (110) are widely
125 used as activity prediction tools in drug design. Since traditional QSAR (MLR and
126 SVM) do not take into account the 3D structure of the compounds, CoMFA and
127 CoMSIA models were constructed to further test the robustness of TCM candidates
128 as ligands with biological activity against Src. The Src ligands from (103) were
129 randomly divided into a training set of 15 compounds and an external validation test
130 set of five compounds.

131 Partial least square (PLS) analysis is a statistical tool used for establishing a
132 linear model describing the correlation between dependent and independent
133 variables and has the advantage of being directly applicable for prediction(111). In
134 this section, CoMFA and CoMSIA descriptors were used as independent variables
135 and pIC_{50} values were the dependent variables.

136 The cross-validated coefficient, q^2, which is calculated by equation **[4]**, was used to
137 evaluate the prediction accuracy:

138 $$q^2 = 1 - \frac{\sum \left(pIC_{50\_predicted} - pIC_{50\_actual} \right)^2}{\sum \left(pIC_{50\_actual} - pIC_{50\_mean} \right)^2} \qquad \textbf{[4]}$$

139 Conventional correlation coefficient r^2 and the standard error, SEE, were also
140 computed for each PLS model. Models with the highest q^2, r^2, and lowest SEE were
141 selected as the optimum CoMFA and CoMSIA model. Once the optimum model is
142 established, relevant compound descriptors are projected into the PLS model to
143 make external predictions on the test set.

144 CoMFA and CoMSIA structure building was performed using the SYBYL
145 program. The CoMFA descriptors steric and electrostatic field energies were

calculated by Lennard Jones function (112) and Coulombic function (113), respectively, using the SYBYL default parameters: avan der Waals (vdW) radius of 1.52 Å, a C^{1+} probe atom, grid point spacing of 2Å, and energy cut-off value of 30 kcal/mol. The field contributions in CoMSIA, namely steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor descriptors, were calculated with identicalC^{1+} probe atom and grid spacing parameters. The probe radius was set at 1.0 Å. The default attenuation factor (R) value of 0.3 was used. Column filtering was set at2.0 kcal/mol. The TCM ligands were overlaid against the generated models to evaluate biological activities based on the CoMFA and CoMSIA models.

Molecular Dynamics (MD) Simulation

To verify the stability of TCM candidates under dynamic conditions, molecular dynamics (MD) simulations were conducted on the Src-candidate complexes using DS2.5. The energy of each complex was minimized with 500 steps each of Steepest Descent and Conjugate Gradient. The system was heated for 50 ps to increase temperature from 50K to 310K and allowed to equilibrate for 200 ps once the target temperature was reached. Canonical ensemble (NVT; constant temperature) was selected for the 40 ns production process and snapshots were taken at 20 ps intervals. Time steps were set at 2fs. Electrostatic interactions (114) were calculated using Particle Mesh Ewald (PME) method. The MD results were used to analyze energy trajectories, H-bond formation and distances, and torsion angles which provide insights to the interaction between TCM candidates and Src kinase.

Results and Discussion

Docking and Candidate Screening

. Table I lists the top ten TCM candidates with the lowest Binding Energy and

Ligand Internal Energy. The complete list of top 100 TCM ligands based on Binding Energyand Ligand Internal Energy can be viewed in Supplementary Table I. Considering the ability to form multiple bonds with Src and their chemical structure, Angeliferulate,(3R)-2'-hydroxy-3',4'-dimethoxyisoflavan-7-O-beta-D-glucoside (HMID), and 3-[2',6-dihydroxy-5'-(2-propenyl)[1,1'-biphenyl]3-yl]-(E)-2-propenoic acid (3PA) were selected as our top candidates. The structural scaffolds of the TCM candidates and the control Saracatinib are shown in Figure 2.Significantly lower binding energy and LIE values were estimated for the TCM candidates (Table I).A higher binding energy, such as the case with Saracatinib, implies that the ligand binding with the protein is more unstable. Results for ADMET are shown in Supplementary Figure 1.

Figures 3-6 illustrates interactions and Src protein residues that may be of importance for each ligand. Figure 7 highlights amino acids that form hydrophobic contacts with each test ligand. Saractatinib formed two types of interactions with Src. Pi-interaction was formed with the positively charged alkyl group in Lys298 (Figure 3A), and hydrophobic interactions were formed with amino acids located on different loops (Figure 7A). The low calculated binding energy and LIE for Angeliferulate (Top 1) could be the result of H-bonds formed with Lys298, Asp351, Asp407 (Figure 4A), and hydrophobic interactions with Val284, Gly347, and Leu396 (Figure 7B). HMID is docked within Src through H-bonds with Leu276, Lys298, and Asp407 (Figure 5A) and hydrophobic interactions with seven amino acids (Figure 7C). These interactions are formed with amino acids either on the

activation loop (Ala407, Asp407) or on different loops (Leu276, Ser348), helping to maintain the stability of the cleft. Similarly, 3PA also docked in Src through the formation of H-bonds (Lys298, Ser348, and Asp351) (Figure 6A) and hydrophobic interactions formed with amino acids on different loops (Figure 7D).

Lys298 is an amino acid of critical importance in Src activation. Under open conformations, ATP will enter the cleft, bind atLys298, undergo hydrolysis to release a phosphate and exit as ADP. The released phosphate is used to phosphorylate downstream reactions. A ligand that can effectively interact with Lys298and form stable and permanent interactions with neighboring amino acid residues has potential to block ATP from the binding site and inhibit activation of downstream reactions. Whilst Saracatinib formed a pi-interaction with Lys298, lack of H-bonds with other residues renders Saracatinib unstable. Higher stability of TCM candidates over Saracatinib may be likely due to their ability to form H-bonds with multiple amino acids including Lys298. Hydrophobic contacts, through significantly weaker than pi-interactions and H-bonds, may also contribute to stability. In particular, hydrophobic interactions formed by Saracatinib, HMGF, and 3PA with Leu276 and Ser348 (located on different amino acid chains) (Figure 7) can increase stability of the cleft. HMID also forms hydrophobic interactions with activation loop amino acids Ala406 and Asp407 which can resist removal of the activation loop from the cleft during Tyr419 activation.

Bioactivity Prediction by Multiple Linear Regression (MLR)and Support Vector Machine (SVM) Models

The MLR and SVM models were developed with ten structural descriptors identified by GFA, namely *ALogP*, *ES_Sum_dsCH*, *ES_Sum_aaCH*, *ES_Sum_sCl*, *Molecular_Weight*, *Molecular_SurfaceArea*, *Molecular_PolarSurfaceArea*,

Jurs_RPCG, *Jurs_WNSA_1*, and *PMI_X.ALogP*is a measurement of molecular hydrophobicity based on Ghose and Crippen's method (115); *ES_Sum_dsCH*is theelectrotopological count for carbons with one single bond and one double bond,*ES_Sum_aaCH* stand for the electrotopological count for carbons with two aromatic bonds; *ES_Sum_sCl* is related to the electrotopological count of chlorides with a single bond; *Molecular_Weight* is the sum of atomic weight; *Molecular_SurfaceArea* and *Molecular_PolarSurfaceArea*calculate total surface area and polar surface area, respectively; *Jurs_RPCG*describes relative positive charge; *Jurs_WNSA_1*describes the total molecular solvent-accessible surface, and *PMI_X* is a spatial descriptor related to the orientation and conformational rigidity of the ligand.

The generated MLR model is expressed as follows and has good prediction ability (R^2= 0.8043; Figure 8A):

$$\begin{aligned} pIC_{50} &= 387.7 - 0.2414 \times ALogP + 9.2497 \times ES\_Sum\_dsCH \\ &- 0.879 \times ES\_Sum\_aaCH - 33.9470 \times ES\_Sum\_sCl + \\ &0.0433 \times Molecular\_Weight + 0.0107 \times Molecular\_SurfaceArea \\ &- 0.0567 \times Molecular\_PolarSurfaceArea - 7.3521 \times Jurs\_RPCG \\ &+ 0.0050 \times Jurs\_WNSA\_1 + 0.0060 \times PMI\_X \end{aligned}$$

The SVM model constructed using the aforementioned descriptors also generated a prediction model where predicted values were highly correlated to actual observed values (R^2=0.937; Figure 8B). The robustness of our models were validated through external validation tests using the test set. Good correlation between observed and predicted pIC_{50} values were observed for both models.

Predicted pIC_{50}values of the TCM candidates and Saracatinib using the MLR and SVM models are listed in Table I. Results suggest that the TCM candidates have good bioactivity towards Src.

3D-QSAR Modeling and Analysis

Table II shows the results of the CoMFA and CoMSIA models generated through PLS algorithm. PLS statistics led to a CoMFA model in which steric features were the dominant contributing factor. Several CoMSIA model combinations were generated and the model considering electrostatic (E), hydrophobic (H), and H-bond donor (D) properties was selected as the optimum model based on high cross-validation and non-cross validation correlation coefficients (q^2=0.482, r^2=0.877).

To test the predictive capabilities of the models, they were used to predict the pIC_{50} values of an external test set excluded from the original training set. Bioactivity predictions of the training and test set are listed in Table III.The correlation curves show high square correlation coefficients of R^2=0.9721 for CoMFA (Figure 9A) and R^2=0.9414 for CoMSIA (Figure 9B), implying models of good prediction power.

In Figure 10, the heteroaromatic ring of Saracatinib, the benzene of Angeliferulate, and the hydroxyl group of HMID, and partial regions of the hexane ring in 3PA fall within the steric favor region, but only Angeliferulate forms interaction bonds. All test ligands fall within the regions between Leu276/Asp351 and Lys298/Asp407, satisfying the CoMFA steric favoring region located at Leu276/Asp351. However, only TCM candidates have interactions with the amino acids located within the Lys298/Asp407 region. When the CoMSIA contour was superimposed, Lys298 was located within the region where hydrophilic interactions were desirable for bioactivity (white)(Figure 11). The ability of Angeliferulate, HMID and 3PA to form H-bonds at Lys298 contour to this hydrophilic region and most likely contribute to higher bioactivities. The hydroxyl group of Angeliferulate falls adjacent to the electrostatic favoring (orange) region of Leu276/Asp351

(Figure 11B). The hydrophobic benzene ring moiety of 3PA is located close to the hydrophobic disfavoring region (white) of Leu276/Asp351 (Figure 11D). Violation of the hydrophobic disfavor region contour matches the lower bioactivity predicted by our SVM/MLR models (Table I).

Molecular Dynamics (MD) Simulation

RMSDs and total energy trajectories

Protein-ligand complex RMSDs, ligand RMSDs, and total energy level changes during MD are shown in Figure 12. Saracatinib has the highest complex, ligand RMSDs, and the highest total energy of the four test compounds. Most notably, the significant increase in ligand RMSD (Figure 12B) indicates high instability. Total energy of the TCM candidates by increasing order were3PA<Angeliferulate<HMID, all of which were lower than that of Saracatinib (Figure 12C). The lower total energy trajectories indicate a more stable state of the ligand-protein complex during dynamic situations.

Saracatinib-Src interactions during MD

Saracatinib formed the least amount of interactions with Src during docking, a phenomenon also observed during MD (Figure 3A).As indicated by the H-bond distance trajectories (Figure 3B) and H-bond occupancy analysis (Table IV), the only stable H-bonds formed during MD were with Lys298 and Ser348. Saracatinib could not form H-bonds with neighboring residues Leu276, Asp351, and Asp407 were the underlying reasons for the inability to form H-bonds.

Angeliferulate-Src interactions during MD

The primary stabilizing interactions formed between Angeliferulate and Src were at Ser348, Asp351, and Asp407 (Figure 4B). H-bonds with Lys298 were also initially observed, but rotations on Angeliferulate increased the distance and

290 discouraged interactions with Lys298 after 9.86 ns. Distance from Asp351 reduced
291 from 4.47Å to 2.45Å at 7.64ns, enabling the formation of multiple H-bonds. Small
292 fluctuations on the H-bond distanceswithSer348 and Asp407 were observed, but
293 since the distance was greater than the 2.50Å cut-off distance, the H-bonds were
294 presumed to be weaker.
295 *HMID-Src interactions during MD*
296 A significant directional shift of HMID was observed during MD due to the
297 formation of H-bonds with new amino acid residues (Table IV). Initially, the
298 H-bond with Leu276 during directed HMID towards the small lobe (Figure 5A).
299 During MD, the loss of the Leu276 H-bond and the formation of H-bond with
300 Ser348 (Figure 5B) shifted HMID towards the large lobe (Figure 5A). HMID
301 formed seven H-bonds with Lys298, but as the HZ1, HZ2, and HZ3 on Lys298
302 continuously rotated, only a 50.65% occupancy was recorded. The H-bond with
303 Asp407 observed during docking was stable throughout MD.
304 *3PA-Src interactions during MD*
305 3PA was primarily stabilized by Lys298 and Asp351 during MD (Table IV).
306 Multiple H-bonds were formed with Lys298. Distance fluctuations recorded in
307 Figure 6B were normal circumstances caused by the constant rotation of the Lys298
308 H atoms. A stable H-bond with occupancies of greater than 99% was maintained
309 with Asp351. Contrary to the docking pose, no high occupancy H-bonds were
310 observed for Ser348.
311 *Torsion angle changes during MD*
312 More information on bond formation during MD can be explained through
313 torsion angle changes. As shown in Figure 3C, torsion of **a** and **b** contribute to the
314 relative spatial angle of the chloride-containing moiety of Saracatinib. The inability

of Saracatinib to maintain interactions with Lys298 and Asp407 could be attributed to conformational shifts brought on by these torsions which increase the distance of these residues from Saracatinib. Oxane could not form interactions due to large torsion fluctuations in **c** and **d** and its orientation away from the pocket. Piperazine groups could not form H-bonds due to the large torsion changes observed in **e**, **f**, **g**, and **h**.

Angeliferulate was a largely flexible ligand as indicated by the torsion angles (Figure 4C). Large torsion changes at **a**-**d** disrupted the ability of the terminal aromatic ring to form stable H-bonds. Torsion angles at **e**-**h** also showed that constant fluctuations. Though **m** and **n** were more stable, the dynamic outer regions of the cleft limited the methoxy group and hydroxyl groups of **m** and **n** from forming H-bonds.

Fluctuations of the torsion angles were also observed in HMID, but H-bond formations were rarely affected (Figure 5C). The stability of **b**, **e**, and **f** enabled the formation of stable H-bonds with Lys298 and Asp407 through O17 and O28, respectively. Other recorded fluctuations at **a-d**, and **g** did not affect stability of these two H-bonds. Two large torsion changes were observed at **h** (8.98 ns and 15 ns), both of which directly caused H-bond distance changes between H55 and Leu276, Ser348, and Asp351 (Figure 5B).

Figure 6C shows that **c**, which connects the two benzene moieties in 3PA were very stable. Torsions at **e**, **f**, **g** did not fluctuate greatly, resulting in the stable H-bond with Lys298. 3PA remained in a relatively linear conformation until changes observed at **d**. Nonetheless, aromatic rings remained in a planar state after 22 ns.

Discussion

Based on the results of docking and MD, Leu276, Lys298, Ser348, Asp351,

and Asp407 are important amino acid residues for stability within Src (Table V). MD results generally supported the findings of docking but showed additional H-bond formation. This implies that under dynamic physiological conditions, all ligands tested could form and remain in complex with Src. Comparison between ligands show that the TCM candidates were more stable and had lower total energy than Saracatinib. The higher stability of the TCM candidates can be attributed to the formation of more H-bonds that remain stable throughout MD. Angeliferulate and HMID not only formed H-bond with the ATP binding site Lys298, but also with Asp407. This can effectively limit the phosphorylated Tyr419 from moving away from the cleft, thus exhibiting important characteristics for being potential Src inhibitors.

Conclusion

This study utilizes computational methods to virtually screen small molecules found in TCM for potential Src inhibitors. Potential of each candidate was validated using structure-based and ligand-based drug design. As shown in Supplementary Video 1, Angeliferulate and HMID have multiple stable interactions with the two Src cleft loops while simultaneously interacting with Asp407, hindering the activation loop from activation. 3PA also exhibits drug-like potential by primarily interacting with Src through via cleft loop amino acids, but may be less potent than its TCM counterparts due to the lack of direct interaction with the activation loop. Considering the aforementioned interactions with Src and high affinity with EGFR, HER2, and HSP90, we suggest that Angeliferulate and HMID which both originate from the TCM *Angelica sinensis* may have potential as multi-targeting drug leads.

Supplementary Materials

Supplementary material include the top 100 candidates from screening, ADMET

365 results of the top three TCM candidates, and a video depicting the mode of
366 inhibition of the TCM candidates on Src.

367 ***Acknowledgements***

368 The research was supported by grants from the National Science Council of
369 Taiwan (NSC 100-2325-B-039-001), Committee on Chinese Medicine and
370 Pharmacy (CCMP100-RD-030), China Medical University and Asia University
371 (DMR-101-094). This study is also supported in part by Taiwan Department of
372 Health Clinical Trial and Research Center of Excellence (DOH101-TD-B-111-004)
373 and Taiwan Department of Health Cancer Research Center of Excellence
374 (DOH101-TD-C-111-005). We are grateful to the National Center of
375 High-performance Computing for computer time and facilities and Dr. Su-sen
376 Chang for discussions and technical assistance with the manuscript preparations. We
377 also wish to express thanks to cloud-computing facilities at Asia University.

References

1. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. *Science 298*, 1912-1934 (2002).
2. S. M. Thomas and J. S. Brugge. *Annu Rev Cell Dev Biol 13*, 513-609 (1997).
3. C. Oneyama, E. Morii, D. Okuzaki, Y. Takahashi, J. Ikeda, N. Wakabayashi, H. Akamatsu, M. Tsujimoto, T. Nishida, K. Aozasa, and M. Okada. *Oncogene*, (Advanced Online Publication) doi: 10.1038/onc.2011.1367 (2011).
4. J. Huang, L. Yao, R. Xu, H. Wu, M. Wang, B. S. White, D. Shalloway, and X. Zheng. *Embo J 30*, 3200-3211 (2011).
5. G. Swaminathan and C. A. Cartwright. *Oncogene*, (Advanced Online Publication) doi:10.1038/onc.2011.1242 (2011).
6. R. R. Brady, C. J. Loveridge, M. G. Dunlop, and L. A. Stark. *Carcinogenesis 32*, 1069-1077 (2011).
7. D. Gianni, N. Taulet, C. DerMardirossian, and G. M. Bokoch. *Mol Biol Cell 21*, 4287-4298 (2010).
8. B. Pohorelic, R. Singh, S. Parkin, K. Koro, A. D. Yang, C. Egan, and A. Magliocco. *Breast Cancer Res Tr*, (Advanced Online Publication) doi: 10.1007/s10549-10011-11753-10542 (2011).
9. J. Jian, Q. Yang, and X. Huang. *J Biol Chem 286*, 35708-35715 (2011).
10. A. J. Bernier, J. Zhang, E. Lillehoj, A. R. Shaw, N. Gunasekara, and J. C. Hugh. *Mol Cancer 10*, 93 (2011).
11. V. Ratushny, H. B. Pathak, N. Beeharry, N. Tikhmyanova, F. Xiao, T. Li, S. Litwin, D. C. Connolly, T. J. Yen, L. M. Weiner, A. K. Godwin, and E. A. Golemis. *Oncogene*, doi: 10.1038/onc.2011.1314 (2011).
12. N. Tikhmyanova and E. A. Golemis. *PLoS One 6*, e22102 (2011).
13. T. Miyake and S. J. Parsons. *Oncogene*, doi: 10.1038/onc.2011.1332 (2011).
14. M. E. Irwin, N. Bohin, and J. L. Boerner. *Cancer Biol Ther 12*, 718-726 (2011).
15. T. J. Shackleford, Q. Zhang, L. Tian, T. T. Vu, A. L. Korapati, A. M. Baumgartner, X. F. Le, W. S. Liao, and F. X. Claret. *Breast Cancer Res 13*, R65 (2011).
16. J. D. Bjorge, A. S. Pang, M. Funnell, K. Y. Chen, R. Diaz, A. M. Magliocco, and D. J. Fujita. *PLoS One 6*, e19309 (2011).
17. Z. Hou, D. J. Falcone, K. Subbaramaiah, and A. J. Dannenberg. *Carcinogenesis 32*, 695-702 (2011).
18. C. C. Mader, M. Oser, M. A. Magalhaes, J. J. Bravo-Cordero, J. Condeelis, A. J. Koleske, and H. Gil-Henn. *Cancer Res 71*, 1730-1741 (2011).

19. K. K. Haenssen, S. A. Caldwell, K. S. Shahriari, S. R. Jackson, K. A. Whelan, A. J. Klein-Szanto, and M. J. Reginato. *J Cell Sci 123*, 1373-1382 (2010).
20. K. E. Reeder-Hayes, L. A. Carey, and W. M. Sikov. *Breast Dis 32*, 123-136 (2010).
21. S. Thomas, J. B. Overdevest, M. D. Nitz, P. D. Williams, C. R. Owens, M. Sanchez-Carbayo, H. F. Frierson, M. A. Schwartz, and D. Theodorescu. *Cancer Res 71*, 832-841 (2011).
22. N. Said and D. Theodorescu. *Cancer Metastasis Rev 28*, 327-333 (2009).
23. A. Prinetti, T. Cao, G. Illuzzi, S. Prioni, M. Aureli, N. Gagliano, G. Tredici, V. Rodriguez-Menendez, V. Chigorno, and S. Sonnino. *J Biol Chem 47*, 40900-40910 (2011).
24. S. Charoenfuprasert, Y. Y. Yang, Y. C. Lee, K. C. Chao, P. Y. Chu, C. R. Lai, K. F. Hsu, K. C. Chang, Y. C. Chen, L. T. Chen, J. Y. Chang, S. J. Leu, and N. Y. Shih. *Oncogene 30*, 3570-3584 (2011).
25. H. S. Kim, H. D. Han, G. N. Armaiz-Pena, R. L. Stone, E. J. Nam, J. W. Lee, M. M. Shahzad, A. M. Nick, S. J. Lee, J. W. Roh, M. Nishimura, L. S. Mangala, J. Bottsford-Miller, G. E. Gallick, G. Lopez-Berestein, and A. K. Sood. *Clin Cancer Res 17*, 1713-1721 (2011).
26. E. L. Leung, J. C. Wong, M. G. Johlfs, B. K. Tsang, and R. R. Fiscus. *Mol Cancer Res 8*, 578-591 (2010).
27. X. G. Liu, Y. Guo, Z. Q. Yan, M. Y. Guo, Z. G. Zhang, and C. A. Guo. *Zhonghua Zhong Liu Za Zhi 33*, 340-344 (2011).
28. Y. Ding, X. Wang, A. Xu, X. Xu, K. Tian, C. Y. Young, and H. Yuan. *J Cell Biochem 112*, 818-828 (2011).
29. T. Kobayashi, T. Inoue, Y. Shimizu, N. Terada, A. Maeno, Y. Kajita, T. Yamasaki, T. Kamba, Y. Toda, Y. Mikami, T. Yamada, T. Kamoto, O. Ogawa, and E. Nakamura. *Mol Endocrinol 24*, 722-734 (2010).
30. S. Zhang, H. E. Zhau, A. O. Osunkoya, S. Iqbal, X. Yang, S. Fan, Z. Chen, R. Wang, F. F. Marshall, L. W. Chung, and D. Wu. *Mol Cancer 9*, 9 (2010).
31. J. DaSilva, D. Gioeli, M. J. Weber, and S. J. Parsons. *Cancer Res 69*, 7402-7411 (2009).
32. G. Pandini, M. Genua, F. Frasca, R. Vigneri, and A. Belfiore. *Ann N Y Acad Sci 1155*, 263-267 (2009).
33. F. Leve, T. G. Marcondes, L. G. Bastos, S. V. Rabello, M. N. Tanaka, and J. A. Morgado-Diaz. *Eur J Pharmacol 671*, 7-17 (2011).
34. L. M. Sturla, P. O. Zinn, K. Ng, M. Nitta, D. Kozono, C. C. Chen, and E. M. Kasper. *Brit J Cancer 105*, 1235-1243 (2011).

35. A. E. Al Moustafa, A. Yasmeen, and A. Achkhar. *Med Hypotheses 77*, 812-814 (2011).
36. S. Saini, S. Arora, S. Majid, V. Shahryari, Y. Chen, G. Deng, S. Yamamura, K. Ueno, and R. Dahiya. *Cancer Prev Res 4*, 1698-1709 (2011).
37. X. F. Le, W. Mao, G. He, F. X. Claret, W. Xia, A. A. Ahmed, M. C. Hung, Z. H. Siddik, and R. C. Bast, Jr. *J Natl Cancer I 103*, 1403-1422 (2011).
38. A. Migliaccio, G. Castoria, and F. Auricchio. *Methods Mol Biol 776*, 361-370 (2011).
39. J. C. Lee, M. C. Maa, H. S. Yu, J. H. Wang, C. K. Yen, S. T. Wang, Y. J. Chen, Y. Liu, Y. T. Jin, and T. H. Leu. *Mol Carcinog 43*, 207-214 (2005).
40. M. Ferrandi, I. Molinari, L. Torielli, G. Padoani, S. Salardi, M. P. Rastaldi, P. Ferrari, and G. Bianchi. *Sci Transl Med 2*, 59ra86 (2010).
41. P. M. Kearney, M. Whelton, K. Reynolds, P. Muntner, P. K. Whelton, and J. He. *Lancet 365*, 217-223 (2005).
42. P. Ferrari. *Biochim Biophys Acta 1802*, 1254-1258 (2010).
43. P. Ferrari, M. Ferrandi, G. Valentini, and G. Bianchi. *Am J Physiol-Reg I 290*, R529-R535 (2006).
44. P. C. Chen, C. H. Muo, Y. T. Lee, Y. H. Yu, and F. C. Sung. *Stroke 42*, 3034-3039 (2011).
45. A. P. Guimaraes, A. A. Oliveira, E. F. da Cunha, T. C. Ramalho, and T. C. Franca. *J Biomol Struct Dyn 28*, 455-469 (2011).
46. E. P. Semighini, J. A. Resende, P. de Andrade, P. A. B. Morais, I. Carvalho, C. A. Taft, and C. H. T. P. Silva. *J Biomol Struct Dyn 28*, 787-796 (2011).
47. Y. D. Cai, J. F. He, and L. Lu. *J Biomol Struct Dyn 28*, 797-804 (2011).
48. Z. H. Mei, J. Liu, and H. W. Yu. *J Biomol Struct Dyn 28*, 871-879 (2011).
49. K. C. Chen and C. Y. C. Chen. *Soft Matter 7*, 4001-4008 (2011).
50. H. J. Huang, K. J. Lee, H. W. Yu, H. Y. Chen, F. J. Tsai, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 187-200 (2010).
51. C. Y. C. Chen. *J Biomol Struct Dyn 27*, 627-640 (2010).
52. K. Bhargavi, P. Kalyan Chaitanya, D. Ramasree, M. Vasavi, D. K. Murthy, and V. Uma. *J Biomol Struct Dyn 28*, 379-391 (2010).
53. M. T. Cambria, D. Di Marino, M. Falconi, S. Garavaglia, and A. Cambria. *J Biomol Struct Dyn 27*, 501-509 (2010).
54. G. Ompraba, D. Velmurugan, P. A. Louis, and Z. A. Rafi. *J Biomol Struct Dyn 27*, 489-499 (2010).
55. E. F. F. da Cunha, E. F. Barbosa, A. A. Oliveira, and T. C. Ramalho. *J Biomol Struct Dyn 27*, 619-625 (2010).
56. C. Y. C. Chen. *J Taiwan Inst Chem E 41*, 143-149 (2010).

57. C. Y. C. Chen. *J Taiwan Inst Chem E 40*, 55-69 (2009).
58. C. Y. C. Chen. *J Chin Chem Soc-Taip 54*, 653-658 (2007).
59. C. Y. C. Chen. *PLoS One 6*, e15939 (2011).
60. K. Sanderson. *Nat Med 17*, 1531 (2011).
61. T. Y. Tsai, K. W. Chang, and C. Y. C. Chen. *J Comput Aid Mol Des 25*, 525-531 (2011).
62. K. W. Chang, T. Y. Tsai, K. C. Chen, S. C. Yang, H. J. Huang, T. T. Chang, M. F. Sun, H. Y. Chen, F. J. Tsai, and C. Y. C. Chen. *J Biomol Struct Dyn 29*, 243-250 (2011).
63. S. S. Chang, H. J. Huang, and C. Y. C. Chen. *PLoS Comput Biol 7*, e1002315 (2011).
64. S. S. Chang, H. J. Huang, and C. Y. C. Chen. *Mol Biosyst 7*, 3366-3374 (2011).
65. K. C. Chen, M. F. Sun, S. C. Yang, S. S. Chang, H. Y. Chen, F. J. Tsai, and C. Y. C. Chen. *Chem Biol Drug Des 78*, 679-688 (2011).
66. T. T. Chang, K. C. Chen, K. W. Chang, H. Y. Chen, F. J. Tsai, M. F. Sun, and C. Y. C. Chen. *Mol Biosyst 7*, 2702-2710 (2011).
67. M. F. Sun, H. Y. Chen, F. J. Tsai, S. H. Lui, and C. Y. C. Chen. *J Biomol Struct Dyn 29*, 325-337 (2011).
68. M. F. Sun, T. T. Chang, K. W. Chang, H. J. Huang, H. Y. Chen, F. J. Tsai, J. G. Lin, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 895-906 (2011).
69. T. T. Chang, M. F. Sun, H. Y. Chen, F. J. Tsai, M. Fisher, J. G. Lin, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 773-786 (2011).
70. C. H. Lin, T. T. Chang, M. F. Sun, H. Y. Chen, F. J. Tsai, K. L. Chang, M. Fisher, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 471-482 (2011).
71. P. C. Chang, J. D. Wang, M. M. Lee, S. S. Chang, T. Y. Tsai, K. W. Chang, F. J. Tsai, and C. Y. C. Chen. *J Biomol Struct Dyn 29*, 471-483 (2011).
72. S. C. Yang, S. S. Chang, and C. Y. C. Chen. *PLoS One 6*, e28793 (2011).
73. S. C. Yang, S. S. Chang, H. Y. Chen, and C. Y. C. Chen. *PLoS Comput Biol 7*, e1002189 (2011).
74. T. T. Chang, H. J. Huang, K. J. Lee, H. W. Yu, H. Y. Chen, F. J. Tsai, M. F. Sun, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 309-321 (2010).
75. C. Y. Chen and C. Y. C. Chen. *J Mol Graph Model 29*, 21-31 (2010).
76. H. J. Huang, K. J. Lee, H. W. Yu, C. H. Hsu, H. Y. Chen , F. J. Tsai, and C. Y. C. Chen. *J Biomol Struct Dyn 28*, 23-37 (2010).
77. C. Y. C. Chen. *J Taiwan Inst Chem E 40*, 155-161 (2009).
78. C. Y. Chen, Y. H. Chang, D. T. Bau, H. J. Huang, F. J. Tsai, C. H. Tsai, and C. Y. C. Chen. *Acta Pharmacol Sin 30*, 1186-1194 (2009).

79. C. Y. C. Chen. *J Chin Inst Chem Eng 39*, 291-299 (2008).
80. C. Y. C. Chen. *J Chin Inst Chem Eng 39*, 663-671 (2008).
81. C. Y. C. Chen, G. W. Chen, and W. Y. C. Chen. *J Chin Chem Soc-Taip 55*, 297-302 (2008).
82. C. Y. Chen, H. J. Huang, F. J. Tsai, and C. Y. C. Chen. *J Taiwan Inst Chem E 41*, 8-15 (2010).
83. W. I. Tou and C. Y. C. Chen. *PLoS One In press.* (2012).
84. H. W. Lo, S. C. Hsu, W. Xia, X. Cao, J. Y. Shih, Y. Wei, J. L. Abbruzzese, G. N. Hortobagyi, and M. C. Hung. *Cancer Res 67*, 9066-9076 (2007).
85. H. W. Lo, S. C. Hsu, M. Ali-Seyed, M. Gunduz, W. Xia, Y. Wei, G. Bartholomeusz, J. Y. Shih, and M. C. Hung. *Cancer Cell 7*, 575-589 (2005).
86. M. Casimiro, O. Rodriguez, L. Pootrakul, M. Aventian, N. Lushina, C. Cromelin, G. Ferzli, K. Johnson, S. Fricke, F. Diba, B. Kallakury, C. Ohanyerenwa, M. Chen, M. Ostrowski, M. C. Hung, S. A. Rabbani, R. Datar, R. Cote, R. Pestell, and C. Albanese. *Cancer Res 67*, 4364-4372 (2007).
87. C. Bartholomeusz, H. Itamochi, L. X. Yuan, F. J. Esteva, C. G. Wood, N. Terakawa, M. C. Hung, and N. T. Ueno. *Cancer Res 65*, 8406-8413 (2005).
88. Y. Yufu, J. Nishimura, and H. Nawata. *Leukemia Res 16*, 597-605 (1992).
89. M. Ferrarini, S. Heltai, M. R. Zocchi, and C. Rugarli. *Int J Cancer 51*, 613-619 (1992).
90. A. Jameel, M. Law, R. Coombes, and Y. Luqmani. *Int J Oncol 2*, 1075-1080 (1993).
91. L. Whitesell, E. G. Mimnaugh, B. De Costa, C. E. Myers, and L. M. Neckers. *P Natl Acad Sci U S A 91*, 8324-8328 (1994).
92. T. H. Leu and M. C. Maa. *Front Biosci 8*, s28-38 (2003).
93. E. M. Poole, K. Curtin, L. Hsu, R. J. Kulmacz, D. J. Duggan, K. W. Makar, L. Xiao, C. S. Carlson, M. L. Slattery, B. J. Caan, J. D. Potter, and C. M. Ulrich. *Int J Mol Epidemiol Genet 2*, 300-315 (2011).
94. L. F. Hennequin, J. Allen, J. Breed, J. Curwen, M. Fennell, T. P. Green, C. Lambert-van der Brempt, R. Morgentin, R. A. Norman, A. Olivier, L. Otterbein, P. A. Ple, N. Warin, and G. Costello. *J Med Chem 49*, 6465-6488 (2006).
95. B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. *J Comput Chem*

30, 1545-1614 (2009).
96. R. A. Laskowski and M. B. Swindells. *J Chem Inf Model 51*, 2778-2786 (2011).
97. C. H. Yun, T. J. Boggon, Y. Li, M. S. Woo, H. Greulich, M. Meyerson, and M. J. Eck. *Cancer Cell 11*, 217-227 (2007).
98. P. P. Kung, B. Huang, G. Zhang, J. Z. Zhou, J. Wang, J. A. Digits, J. Skaptason, S. Yamazaki, D. Neul, M. Zientek, J. Elleraas, P. Mehta, M. J. Yin, M. J. Hickey, K. S. Gajiwala, C. Rodgers, J. F. Davies, and M. R. Gehring. *J Med Chem 53*, 499-503 (2010).
99. M. F. Sun, S. C. Yang, K. W. Chang, T. Y. Tsai, H. Y. Chen, F. J. Tsai, J. G. Lin, and C. Y. C. Chen. *Mol Simulat 37*, 884-892 (2011).
100. J. A. Blair, D. Rauh, C. Kung, C. H. Yun, Q. W. Fan, H. Rode, C. Zhang, M. J. Eck, W. A. Weiss, and K. M. Shokat. *Nat Chem Biol 3*, 229-238 (2007).
101. B. K. Slinker and S. A. Glantz. *Circulation 117*, 1732-1737 (2008).
102. R. Burbidge, M. Trotter, B. Buxton, and S. Holden. *Comput Chem 26*, 5-14 (2001).
103. D. H. Boschelli, D. Wang, Y. Wang, B. Wu, E. E. Honores, A. C. Barrios Sosa, I. Chaudhary, J. Golas, J. Lucas, and F. Boschelli. *Bioorg Med Chem Lett 20*, 2924-2927 (2010).
104. M. Tufail and L. E. Ormsbee. *J Hydroinform 8*, 193-206 (2006).
105. V. Vapnik and A. Lerner. *Automat Remote Contr 24*, 774-780 (1963).
106. V. Vapnik. *The Nature of Statistical Learning Theory*. (Springer, 1995).
107. O. Ivanciuc. *Reviews in Computational Chemistry 23*, 291-400 (2007).
108. C.-C. Chang and C.-J. Lin. *ACM TSIT*, 2:27:21--27:27 (2011).
109. R. D. Cramer, 3rd, D. E. Patterson, and J. D. Bunce. *Prog Clin Biol Res 291*, 161-165 (1989).
110. G. Klebe, U. Abraham, and T. Mietzner. *J Med Chem 37*, 4130-4146 (1994).
111. T. Langer and S. D. Bryant. in *The Practice of Medicinal Chemistry* (ed C. G. Wermuth) Ch. 29, 587-604 (Path Press, 2008).
112. R. J. Gillespie, D. Bayles, J. Platts, G. L. Heard, and R. F. W. Bader. *J Phys Chem A 102*, 3407-3414 (1998).
113. G. Tang and B. K. Ma. *Chinese Phys 9*, 737-741 (2000).
114. M. Kawata and M. Mikami. *Chem Phys Lett 313*, 261-266 (1999).
115. A. K. Ghose and G. M. Crippen. *J Chem Inf Comput Sci 27*, 21-35 (1987).