Available online at www.tox.or.kr

Topomer-CoMFA Study of Tricyclic Azepine Derivatives - EGFR Inhibitors

Jae Yoon Chung[1], **F. A. Pasha**[1], **Hwanwon Chung**[1], **Beom-Seok Yang**[1], **Cheolju Lee**[1], **Jung Soo Oh**[1], **Myoung-Woon Moon**[1], **Seung Joo Cho**[1] **& Art E. Cho**[2]

[1]Computational Science Center, Future Fusion Technology Division, Korea Institute of Science and Technology, PO Box 131, Seoul 130-650, Korea
[2]Department of Biotechnology and Bioinformatics, Korea University, Jochiwon-Eup, Yeongi-Gun, Chungnam 339-700, Korea
Correspondence and requests for materials should be addressed to S. J. Cho (chosj@kist.re.kr)

Accepted 24 January 2008

Abstract

EGFR has been intensively investigated as a target to block the signal transduction pathway which stimulates cancer growth and metastasis. Studies about structure-activity relationship for tricyclic azepine derivatives were performed with topomer-CoMFA. The derived topomer-CoMFA model with steric and electrostatic field parameters based on fragment units gave reasonable statistics (q^2=0.561, r^2=0.679). The model explains why a halogen atom at the meta position of aniline is important to increases inhibitory activity. This comes from an electrostatically negative groups are favored near this region. The model also shows that there are sterically favored regions around methoxy group extended from oxazepine derivatives. The findings about steric and electrostatic effects can be utilized for designing new inhibitors.

Keywords: Topomer-CoMFA, EGFR, QSAR

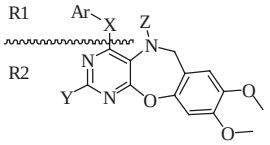
Sizable and effective progress in cancer therapy has been achieved due to extensive clinical research on malignant tumor[1]. Cancer treatments such as chemotherapy and radiation have been used to cure many patients[2]. These treatments, however, cannot be effective against advanced tumor cells because the conventional therapies give drug resistance to tumor cells[3]. The intensive studies of epidermal growth factor (EGF) and EGF receptor (EGFR) have enabled to develop new strategies of cancer cures[4]. One of the strategies is blocking the pathway of cancerous growth and metastasis[5,6].

EGFR, a 170-kDa protein at an epithelial cell surface, consists of three major parts such as extra-membrane ligand-binding sites, single trans-membrane parts, and intracellular tyrosine kinase domains[7]. EGFR belongs to the erbB family including EGFR (HER1 or erbB1), erbB2 (HER2), erbB3 (HER3), and erbB4 (HER4)[8]. EGFR is known to control the signal transduction pathway from the membrane to nucleus which is involved in cell proliferation, apoptosis, and metastasis. EGFR in tumor cells is more over-expressed than that in the normal[9], and the signal transduction pathway of EGFR in tumor cells affects their growth and metastasis. EGFR became the attractive target for the efficient cancer-therapy to block this signal pathway in tumor as mentioned above.

Before the signal transduction begins at EGFR, this receptor generally becomes activated by dimerization[10]. This dimerization occurs after binding proper ligands on the extra-membrane domain. These ligands for dimerization can be EGFs and transforming growth factor-α (TGF-α), as well as hormone-like polypeptides[6]. Dimerization plays crucial roles in transmitting a variety of functions from cell surfaces to nucleuses. Dimerization causes autophorsphorylation, and thereby Protein Tyrosine Kinases (PTKs) are activated. As a result of EGFR-ligand binding, the activation on PTKs initiates the signal transduction pathway which follows complicated cascades transducing small signal-molecules to the nucleus. The growth of tumor cells could be decreased by blocking more than one step among the sequential steps from the extra-membrane to the PTKs.

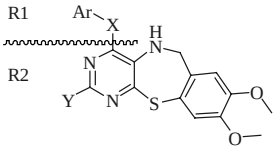
Deregulation of EGFR is commonly observed in various tumors including prostate, breast, and gastric cancers[11]. This deregulation causes PTKs to overactivate and to produce continuously an improper signaling transduction. Normal cells generally have self-adopting ability to keep suitable balance. If mem-

Table 1. Oxazepines as inhibitors of EGFR and their inhibition activates[11].



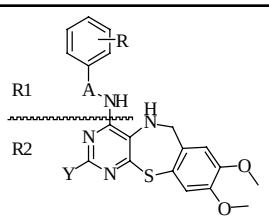
Compound	Structure				pIC_{50}
	Ar	X	Y	Z	
1c	3-Br-Ph	NH	H	H	6.52
2a	3-Me-Ph	NH	H	H	5.13
2b	3-Ethynhyl-Ph	NH	H	H	5.49
2c	4-Br-Ph	NH	H	H	5.06
2d	4-F-Ph	NH	H	H	5.15
2e	3-Cl-4-F-Ph	NH	H	H	5.92
2f	3-Cl-2-F-Ph	NH	H	H	6.52
2g	5-Cl-4-F-Ph	NH	H	H	5.96
2h	2-Cl-4-F-Ph	NH	H	H	5.47
2i	6-Indazolyl	NH	H	H	5.92
2j	2-Naphthyl	NH	H	H	6.30
2k	6-Benzthiazolyl	NH	H	H	5.25
2n	3-Br-Ph	O	H	H	6.15
2o	3-Cl-4-F-Ph	O	H	H	5.92
2p	3-Br-Ph	O	H	Me	5.34
2q	3-Cl-4-F-Ph	O	H	Me	4.97
2r	3-Br-Ph	S	H	H	5.89
2s	3-Cl-Ph	S	H	H	6.00

Table 2. Thiazepines as inhibitors of EGFR and their inhibition activities[11].



Compound	Structure			pIC_{50}
	Ar	X	Y	
3a	3-Br-Ph	NH	H	5.23
3b	3-Cl-4-F-Ph	NH	H	5.09
3c	3-Cl-2-F-Ph	NH	H	4.20
3d	6-Indazolyl	NH	H	4.85
3h	3-Br-Ph	NH	Me	4.51
3j	3-Br-4-Me-Ph	NH	Me	4.34
3m	2-Naphthyl	NH	Me	4.77
3n	5-Benzimidazolyl	NH	Me	4.84
3o	6-Benzthiazolyl	NH	Me	4.82
3p	6-Indazolyl	NH	Me	4.61
3q	5-Indazolyl	NH	Me	4.79
3r	3-Br-Ph	O	H	4.66
3s	3-Cl-2-F-Ph	O	H	4.71
3t	3-Br-Ph	S	H	5.80
3u	3-Cl-Ph	S	H	5.54

Table 3. Diazepines as inhibitors of EGFR and their inhibition activities[11].



Compound	Structure		pIC_{50}
	R	A	
4a	3-Br	–	5.82
4b	3-Cl-4-F	CH_2	6.52
4c	3-MeO	CH_2	6.10
4d	H	CH_2	6.70
4e	H	(R)-MeCH	6.82

brane receptors such as EGFR are over-activated, then cells will reduce the stimulation by decreasing the number of the membrane receptors. This recovery ability, however, does not seem to work effectively in most tumor cells. Some studies report that abnormally high expression and mutation of receptor and /or ligand can deregulate EGFR[7,10]. After binding between abnormal EGFRs and ligands, aberrant hetero-dimerization occurs between EGFR and other erbB family, mainly HER2. This hetero-dimerization induces over-activation loop on PTKs, and it can trigger to grow cancerous cells.

There are two main parts of EGFR for the cancer cure as the following: 1) the ligand-biding site and 2) the PTKs domain. These parts are effective targets to block the abnormal flow of activation of PTKs. In the early stage of EGFR studies, the ligand-binding site was mainly investigated as the target because (structural) information about PTKs domain was not sufficient[12]. Availability of more than 40 X-ray structures stimulated studies targeting ATP-binding site of PTKs for the cancer cure[12]. Several studies indicate that mutations in the ATP-binding site are important to block the downstream of the signal pathway[13]. The ATP-binding inhibitors bind to PTKs domain to block the abnormal signal transduction pathway. Diverse therapies targeting at this ATP-binding site are in train. Several clinical chemical-compounds such as Iressa and Tarceva were investigated and already are in preclinical stage[9]. Smith and his collaborators suggested new tricyclic azepine derivatives as EGFR inhibitors similar with Iressa and Tarceva[14].

This study aims to understand how steric and electrostatic influences interact with tricyclic azepine derivatives (Table 1, 2 and 3)[14]. The major difference among Table 1, 2 and 3 is R2 fragment which differs

only at one atom (O, S, NH) position on the central azepine ring. The tricyclic azepine derivatives have been applied to topomer comparative molecular field analysis (topomer-CoMFA), one of 3D-QSAR techniques. Unlike the original CoMFA, topomer-CoMFA does not require the demanding alignment work while making models. This is because topomer-CoMFA converts fragments of molecular series into only unique topomer shapes. Molecular modeling calculations were performed using SYBYL 7.3.5.

Data Sets and Preparation

Thirty-nine non-quinazoline oxazepine derivatives among the dataset reported by Smith *et al*. were selected as input data for 3D-QSAR analysis[14]. The structures of the compounds and their biological inhibitory data are given in table 1, 2 and 3. The biological inhibitory data in the report of Smith *et al*. used the inhibition constant IC_{50} value which is the concentration (μM) of inhibitor producing 50% inhibition of EGFR PTK. In this dataset, the IC_{50} values were converted into pIC_{50} ($-\log IC_{50}$) as the dependent variable in the topomer-CoMFA analysis.

Topomer-CoMFA analysis should be compared based on 3D topologically aligned structure within sets of fragments. Before progressing it, each molecule is separated into two fragments at an acyclic bond except for bonds in ring structures. This molecular series can be properly separated into roughly homologous fragments at a commonly located acyclic bond as Figure 1(b). The selected acyclic bond (fragment bond) to separate this molecular series is presented in Table 1, 2 and 3. Figure 2 shows the alignment of data set noted by molecular superposition of the pyrimidine ring before topomer-CoMFA. This molecular series were separated into R1 and R2, and each of R1 and R2 fragments is topologically aligned based on the fragment bond and its orientation (Figure 3).

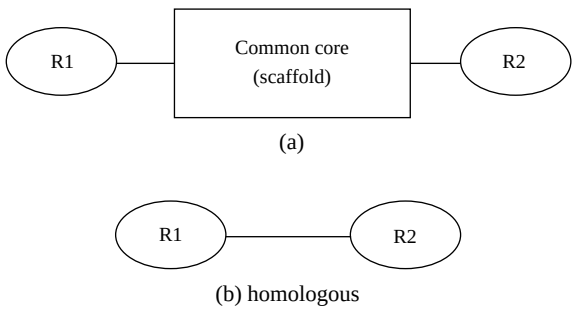


Figure 1. Two different molecular cases for topomer-CoMFA.

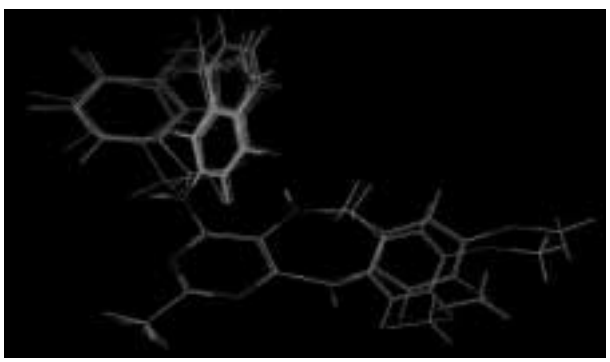


Figure 2. Alignment of training set.

Topomer-CoMFA Model

The steric and electrostatic fields for topomer-CoMFA were calculated at each lattice intersection of regular grid-space of 2Å. The lattice was fixed automatically into $10 \times 10 \times 10$ grid format to enclose topomers. The fields of topologically aligned fragments were calculated as in the original CoMFA with sp^3 carbon atom as the probe for the steric fields and a negative oxygen atom as the probe for the electro-

Table 4. The result of topomer-CoMFA.

#Comp	Stderr	q^2	r^2
2	0.481	0.561	0.679

#Comp=The number of components that provides the highest q^2
Stderr=cross-validated standard error of estimate
q^2=LOO cross-validated correlation coefficient
r^2=noncross-validated correlation coefficient

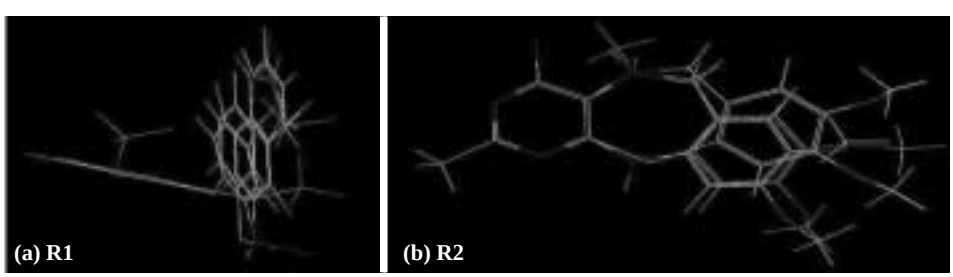


Figure 3. Topologically aligned fragments R1 and R2.

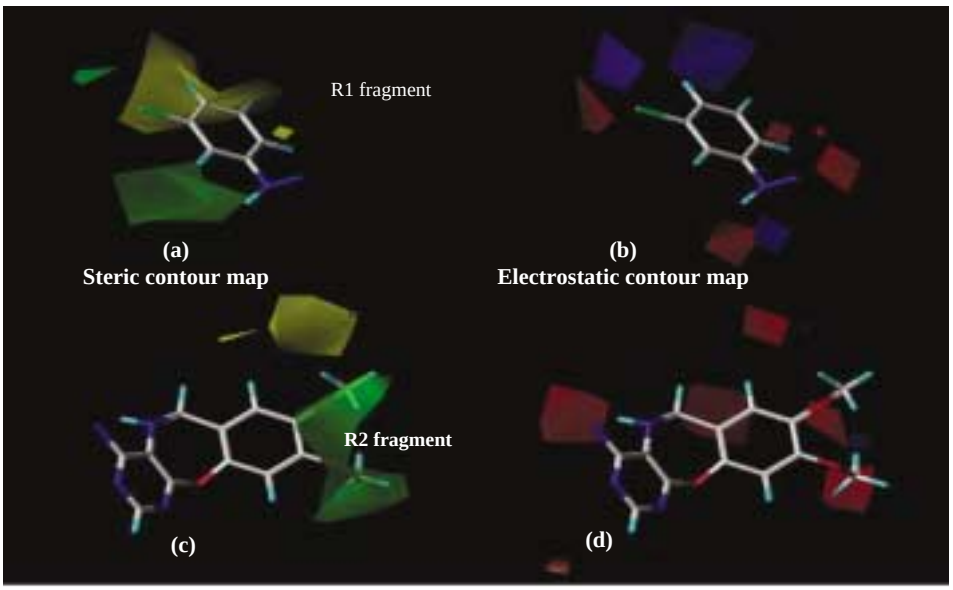


Figure 4. Topomer-CoMFA contour maps represented with compound 1C. (a) and (b) reflect steric and electrostatic contours of fragment R1. (c) and (d) reflect steric and electrostatic contours of fragment R2.

Table 5. Topomer-CoMFA contour map for R1 and R2 of Table 1, 2 and 3.

Contour	R1			R2		
	Contour level	Color	Volume estimate	Contour level	Color	Volume estimate
Steric	−0.004	YELLOW	19.3	−0.014	YELLOW	8.9
	0.007	GREEN	16.3	0.011	GREEN	13.1
Electrostatic	−0.003	RED	3.5	−0.002	RED	10.3
	0.003	BLUE	7.8	0.018	BLUE	0.1

static field. The original CoMFA uses only one column, but topomer-CoMFA uses two CoMFA columns of R1 and R2. Partial Least-Squares (PLS) is used to generate a topomer-CoMFA model. PLS produces a set of coefficients displaying influence of positions on biological properties. PLS analysis yielded a cross-validated correlation coefficient q^2 of 0.561 and r^2 of 0.679 with cross-validated standard error of estimate 0.481. The number of components that provides the highest q^2 is 2. The results of topomer-CoMFA analysis are summarized in Table 4. These values show acceptable statistical correlation and predictability of this topomer-CoMFA model.

Topomer-CoMFA interaction maps (steric and electrostatic interactions) for both fragments (R1 and R2) are shown in Figure 4, and the levels of contour for both electrostatic and steric fields are listed along with their color schemes (Table 5). Figure 5(a) and (c) show why the halogen atoms (Br and Cl) at meta position of the phenyl ring in R1 increase activity values. The region near the meta position favors electronegative substituent as indicated by the red color (Figure 5). The halogen atom at the para position of the ring as Figure 5(b) and (d) occupies the yellow steric unfavored region, so their activity values are decreased. This explains why 3-bromoanilino moiety would increase the inhibitory activity as the EGFR inhibitor[15].

In Figure 4(c), the green color around two OCH_3 groups at the end of R2 indicates that steric bulk is favored there. The major difference between Table 1 and 2 is R2 moiety which differs only one atom (O vs. S or oxazepine vs. thiazepine). When the substituent at R1 are the same for both Table 1 and 2, the activities are always higher for Table 1. Therefore the activity difference should come from the change of a heteroatom in the 7-membered ring.

In Figure 6, (a) and (b) explains reasons for different activity values between Table 1 and 2. Oxazepine leads two OCH_3 groups to the green steric favored region in Figure 6(a). Thiazepine leads the OCH_3 groups to the yellow steric unfavored region in Figure 6(b). It seems that steric influence is important to differentiate the activity for this region (R2).

In Figure 6(c), diazepines (molecules in Table 3) seem to have less activity values because OCH_3 groups of (c) are not bent to the green steric favored region. Compounds 4b-4e in Table 3, however, have

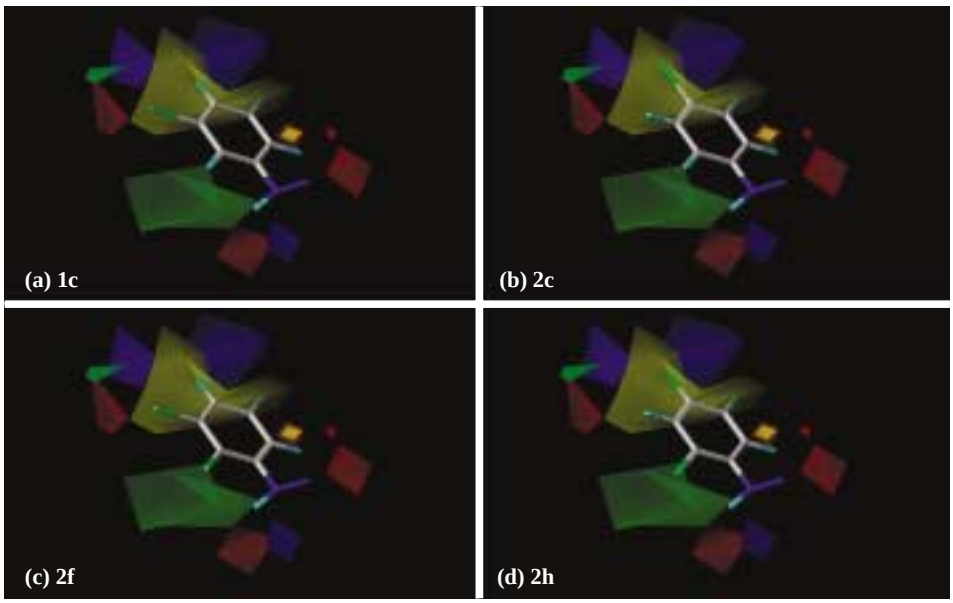


Figure 5. Contour maps of R1 of (a) 1c, (b) 2c, (c) 2f, and (d) 2h.

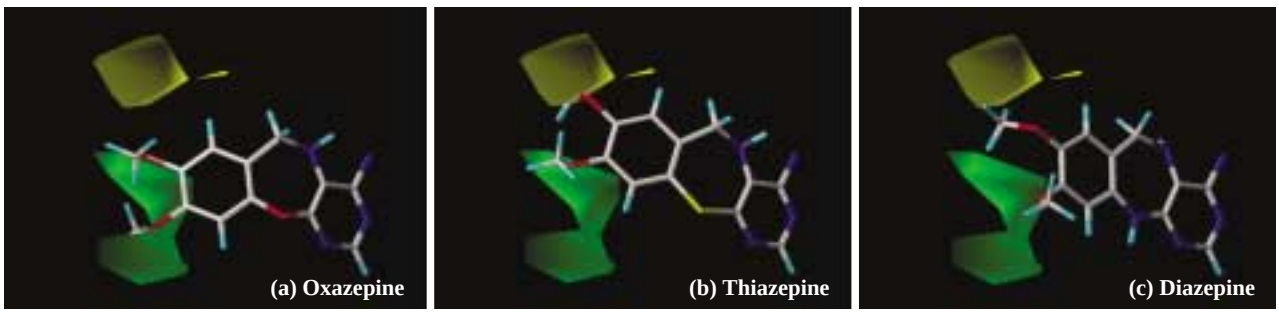


Figure 6. Topomer-CoMFA contour maps of fragment R2s in Table 1, 2, and 3. (a) 1c. The two OC_3s at the end of R2 is located at green steric favored region. (b) 3a. One OC_3 is at yellow steric unfavored region. (c) 4a. None of OC_3s are located at either the steric favored or unfavored region.

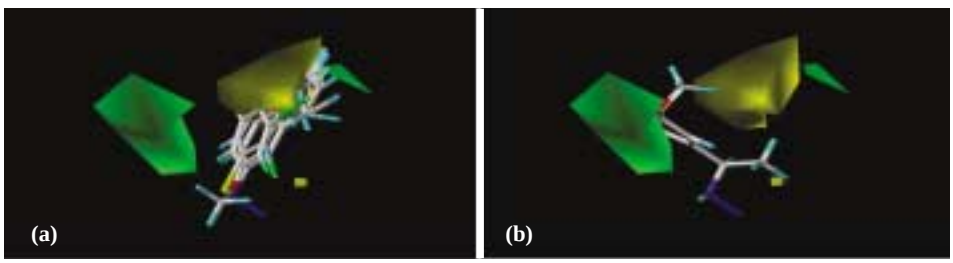


Figure 7. Steric contour of superimposed fragment R1. (a) R1 of molecules in Table 1 and 2. (b) R1 of molecules in Table 3 (except for 4a).

great activity values. Figure 7(b) shows a structural feature that compounds except for 4a in Table 3 have. Compounds except for 4a in Table 3 have substituent (methylene or ethylene) of a position in Table 3. 4a is to understand the effect of diazepine by comparing with its counterparts, 1c and 3a in Table 1 and 2. The R1 fragment of 4a belongs to Figure 7(a) because 4a has the same kind R1 fragment as compounds Table 1 and 2. In Figure 7, R1 fragments (a) from Table 1 and 2 are different from R1 fragments (b) from Table 3. R1 fragments (b) in Table 3 avoid the yellow steric unfavored region and are bent to the left green steric favored region (Figure 7). The influence of the steric favor of R1 fragments from Table 3 increases their

activity values.

Discussion

Using a 3D-QSAR methodology, the topomer-CoMFA, a 3D-QSAR model was developed on tricyclic azepine derivatives for the EGFR inhibitor. The topomer-CoMFA model was used to explain the observed structure-activity relationship for these azepines with steric and electrostatic field parameters. The resultant model was statistically reasonable (q^2= 0.561, r^2=0.679). Some minor electrostatic effects could be found around R1 (Figure 5), and the sterically important regions are mainly localized at the ends of molecule structures (Figure 6 and 7). A halogen atom at the meta position of the phenyl ring in R1 and OC_3 groups from oxazepine (R2) are favored by steric bulkiness. These findings can be utilized for designing new inhibitors.

Methods

Topomer-CoMFA

Comparative Molecular Field Analysis (CoMFA) has been widely used as 3D-QSAR methodology in analyzing structure and activity. Despite powerful visual-expression of structure and activity, CoMFA has its inherent problems. The major problem is input precondition by subjective alignment of molecular series[16]. Slight and partial shift in a lattice by alignment of molecular series can produce misleading analysis of interaction between structure and activity. This problem could be reduced by an objective methodology which determines unique conformations and superposition not for the whole of molecules but for fragments of molecules.

A topomer is an invariant 3D representation of molecular subunit, and it is generated from its 2D topology by topomer alignment in topomer-CoMFA. Because of the invariant representation, topomers can have objective conformations for each subunit (fragment)[17]. Topomer-CoMFA is a comparative molecular field analysis calculating the steric and electrostatic fields with topomerically aligned fragments[18]. Topomer-CoMFA modeling performs two important steps[19]. First, invariant 3D representations of molecule fragments by fragmentation rule and topomer alignment are generated in lattices of energy fields. Second, original CoMFA is applied in the fragments.

In topomer-CoMFA, every molecule should be separated into two fragments which are referred to as R series, and this is shown schematically in Figure 1. Figure 1(a) indicates that compounds have two side chains attached to a common core, and Figure 1(b) describes that compounds consist of two roughly homologous series of molecules. Fragments of template molecule separated by user are converted into “minimal discriminating substructure[20]”. “Fragmentation algorithm” uses this minimal discriminating substructure of the template molecule to find each unique substructure in the left molecular series[20]. These separated fragments will be used to calculate molecular field in lattice as structural variations.

The fragments separated in the fragmentation algorithm are applied into topomer alignment to make 3D invariant representation[20]. Before the topomer alignment, fragments should be adjusted to match their orientation with their open valance (attachment bond) in Cartesian space. Topomer alignment provides ordering to atoms of each fragment, and this enables to express invariant strings in 2 dimensions, e.g. SLN pattern. Topomer alignment also uniquely modifies conformation of each molecule in 3 dimensions by a protocol which can be applied in various circumstances. Topomer alignment protocol describes about special circumstance such as torsions, chirality, and ring puckers. These factors largely make the alignment in the original CoMFA variant and difficult. As a result, topomer alignment generates invariant 3D topomer representations of fragments through specific rules applying to these factors.

Topomer-CoMFA cannot be related in every class of molecule structures. Two classes of molecules may not be difficult to generate effective fragments[19]. The first is compounds containing many homologous side chains with one large common core such as steroid data sets. The second compounds have none of noticeable common core, and it is difficult to apply the fragmentation algorithm.

The left operation after the topomer alignment in topomer-CoMFA is calculation of CoMFA fields. The original CoMFA calculates steric and electrostatic fields between probe and atoms, and topomer-CoMFA follows this calculation in almost the same way. Topomer-CoMFA has two different things from the calculation of molecular fields. First, there is an “at-tenuation factor” which decreases the field contribution of the fragment atoms according to distance from the attachment bond having the open valance[18]. The steric and electrostatic contribution of atoms multiplied by the attenuation factor, usually 0.85, makes 0.85^N where N is the number of rotatable bonds between a specific atom and the attachment bond. Second, steric and electrostatic values at lattice points for the computational convenience are not continuous but disperse numbers as follows[18]. Steric field values

can be 0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30 kcal/mol. Electrostatic values can be −13, −11, −9, −7, −5, −3, −1, 1, 3, 5, 7, 9, 11, 13, or 15 kcal/mol. If a value is higher than shown above, it will be rounded down.

PLS Calculations and Validations

Partial least squares (PLS)[21,22] methodology for cyclic cross-validation with leave-one-out (LOO) method is used in topomer-CoMFA to produce a series of coefficients as in the original CoMFA. Topomer-CoMFA field is used as an independent variable, and the pIC_{50} activity value used as dependent variable. PLS solves a series of equation with much more unknown quantities than equations. Cross-validation procedure evaluates topomer-CoMFA model according to how well the model predicts. This procedure iteratively rederives new models with topomer-CoMFA table where one or more compounds (rows) are omitted, and predicts the omitted observations. The cross-validated coefficient, q^2, is calculated using Eq. 1 below:

$$q^2 = 1 - \frac{\sum(Y_{predicted} - Y_{observed})^2}{\sum(Y_{observed} - Y_{mean})^2} \quad (1)$$

Where $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively. $\sum(Y_{predicted} - Y_{observed})^2$ is the predictive sum of squares (PRESS). Deciding the best number of PLS components is needed to avoid overfitting the data, and the number corresponding with the lowest PRESS value is proper to derive the final PLS regression models.

References

1. Kurup, A., Garg, R. & Hansch, C. Comparative QSAR study of tyrosine kinase inhibitors. *Chem Rev* **101**: 2573-2600 (2001).
2. Sorrentino, B. P. Gene therapy to protect haematopoietic cells from cytotoxic cancer drugs. *Nat Rev Cancer* **2**:431-441 (2002).
3. Katsman, A., Umezawa, K. & Bonavida, B. Reversal of resistance to cytotoxic cancer therapies: DHMEQ as a chemo-sensitizing and immuno-sensitizing agent. *Drug Resistance Updates* **10**:1-12 (2007).
4. Cohen, P. Protein kinases-the major drug targets of the twenty-first century? *Nature Reviews Drug Discovery* **1**:309-315 (2002).
5. Hubbard, S. R. & Till, J. H. Protein tyrosine kinase structure and function. *Annual Review of Biochemistry* **69**:373-398 (2000).
6. Lo, H-W., Hsu, S. C. & Hung, M. C. EGFR signaling pathway in breast cancers: from traditional signal transduction to direct nuclear translocalization. *Breast Cancer Research and Treatment* **95**:211-218 (2006).
7. Janmaat, M. L. & Giaccone, G. Small-molecule epidermal growth factor receptor tyrosine kinase inhibitors. *Oncologist* **8**:576-586 (2003).
8. Toschi, L. & Cappuzzo, F. Understanding the new genetics of responsiveness to epidermal growth factor receptor tyrosine kinase inhibitors. *Oncologist* **12**: 211-220 (2007).
9. Noble, M. E. M., Endicott, J. A. & Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Sience* **303**:1800-1805 (2004).
10. Dawson, J. P. *et al*. Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol Cell Biol* **25**:7734-7742 (2005).
11. Baselga, J. Why the epidermal growth factor receptor? The rationale for cancer therapy. *Oncologist* **7**:2-8 (2002).
12. Traxler, P. *et al*. Tyrosine kinase inhibitors: from rational design to clinical trials. *Medicinal Research Reviews* **21**:499-512 (2001).
13. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem* **277**:46265-46272 (2002).
14. Smith Ii, L. *et al*. Novel tricyclic azepine derivatives: biological evaluation of pyrimido[4,5-b]-1,4-benzoxazepines, thiazepines, and diazepines as inhibitors of the epidermal growth factor receptor tyrosine kinase. *Bioorganic & Medicinal Chemistry Letters* **16**:5102-5106 (2006).
15. Bridges, A. J. *et al*. Tyrosine kinase inhibitors. 8. an unusually steep structure-activity relationship for analogues of 4-(3-Bromoanilino)-6,7-dimethoxyquinazoline (PD 153035), a potent inhibitor of the epidermal growth factor receptor. *J Med Chem* **39**:267-276 (1996).
16. Cramer, R. D. Topomer CoMFA: A design methodology for rapid lead optimization. *J Med Chem* **46**: 374-388 (2003).
17. Jilek, R. J. & Cramer, R. D. Topomers: a validated protocol for their self-consistent generation. *J Chem Inf Comput Sci* **44**:1221-1227 (2004).
18. Cramer, R. D. & Jilek, R. J. Comparative field analysis (CoMFA) utilizing topomeric alignment of molecular fragments. United States. (2003).
19. Cramer, R. D. & Patterson, D. E. Further method of creating and rapidly searching a virtual library of potential molecules using validated molecular structural descriptors. United States. (2001).
20. Tripos Bookshelf 7.3. Tripos Inc. (1699)
21. Bush, B. L. & Nachbar, R. B. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *Journal of Computer-Aided Molecular Design* **7**:587-619 (1993).
22. Wold, S., Sjostrom, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**:109-130 (2001).