STRUCTURE AND LIGAND-BASED DESIGNING OF SOME NOVEL ENZYME INHIBITORS

A Thesis Submitted to

THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA

For the award of the degree of

Doctor of Philosophy

IN

PHARMACY

BY

VISHAL P. ZAMBRE

Under the Guidance of
Prof. M. R. Yadav

Pharmacy Department
Faculty of Technology and Engineering
The Maharaja Sayajirao University of Baroda
Vadodara-390 001

OCTOBER 2011

Pharmacy Department
Faculty of Technology and Engineering
The Maharaja Sayajirao University of Baroda
Post Box No. 51, Kalabhavan
Vadodara – 390 001, Gujarat, India

Date:

CERTIFICATE

This is to certify that the thesis entitled, "***Structure and Ligand-Based Designing of Some Novel Enzyme Inhibitors***" submitted for the award of Ph.D. degree in Pharmacy by Mr. Vishal Prakash Zambre incorporates the original research work carried out by him under my supervision.

Supervisor

The Head
Pharmacy Department

The Dean
Faculty of Technology & Engineering
The Maharaja Sayajirao University ofBaroda
Vadodara-390 001

DECLARATION

I hereby declare that the thesis entitled, "***Structure and Ligand-Based Designing of Some Novel Enzyme Inhibitors***" submitted herewith to The Maharaja Sayajirao University of Baroda, Vadodara for the fulfilment of the award of the degree of **DOCTOR OF PHYLOSOPHY** is the result of work carried out by me in Pharmacy Department, Faculty of Technology and Engineering, The M. S. University of Baroda, Vadodara.

The results of this work have not been submitted previously for any degree/fellowship.

Date:

Place: **Vishal P. Zambre**

Dedicated to Papa & Mummy…..

Acknowledgements

I take this opportunity to express my greatest sense of gratitude to my supervisor **Prof. M. R. Yadav**, Head, Pharmacy Department, Faculty of Technology & Engineering, The M. S. University of Baroda, for the wonderful introduction to the subject Computational Chemistry and Molecular Modelling. I am also thankful to him for his novel ideas, patient guidance and continuous encouragement throughout my research. He inculcated in me a logical approach to tackle scientific problems. I really enjoyed working with him during these years.

I also thank **Prof. (Mrs.) R. Giridhar**, Pharmacy Department for her kind support and valuable suggestions in departmental scientific activities.

I owe my thanks to all my senior colleagues **Dr. Devendra**, **Dr. Anwar**, **Dr. Das Gupta**, **Dr. Sable**, **Mr. Dhaval** and **Mr. Prashant Naik** for their guidance and help. I would like to say very special thanks to **Dr. Prashant Murumkar** for his healthy scientific discussions during my research work.

I am thankful to my colleagues **Anand**, **Palash**, **Vijay**, **Premlal**, **Riyaj**, **Amit**, **Mukesh**, **Yogish** and post-graduate students for creating such a good working atmosphere and for their help. Entire staff of our department is greatly appreciated for their timely help. A swift help provided by **Mr. Kaushik bhai** for our workstations when they misbehaved is highly appreciated.

Special thanks are due to **Dr. M. K. Mohan, Dr. Koiremba Singh, Mr. Riyaj** and **Mr. Palash** for their continuous support and making hostel life really, really beautiful. Kind support provided by **Mr. Vishal** & **Mr. Nilesh Bhai** is highly appreciated.

The financial assistance provided by All India Council for Technical Education (AICTE), New Delhi in the form of **National Doctoral Fellowship (NDF)**, is humbly acknowledged.

I thank various Government funding agencies like Department of Science and Technology (**DST**), Department of Biotechnology (**DBT**) and Council for Scientific and Industrial Research (**CSIR**) for providing me full Travel Grant for presenting my research work at international conferences.

My lovely brother **Ajit dada** and sister-in-law **Duhita vahini** deserves special thanks for their continuous support. I would like to express my sincere thanks to my parents **Dr. Prakash Zambre** and **Mrs. Kalpana Zambre**, for their continuous support and for their motivation and encouragement to achieve higher goals. They raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

My late grandfather **Ajoba** and late grandmother **Ajji** are remembered as they would have been the happiest people to see me at this stage, who always supported education.

Vishal P Zambre

CONTENTS

1. INTRODUCTION

1.1 Molecular Modeling & Drug Discovery

The average cost of developing new drug molecules and the time taken to market them is pretty high. Although the time factor is getting reduced but the cost factor is increasing further. Reasons for the growing expenses of pharmaceutical companies include the investments needed for the new high throughput research technologies [1] and for increase in the number of studies required for new drug molecules [2]. However, even with the growing cost, the number of new drugs coming to market has shown only a marginal increase. One of the reasons for this trend could be lack of structural information about the target molecules. In the drug design process enzymes are frequently the target of choice because of their involvement in various biochemical pathways in human physiology. Even with enzymes, there can be problems in obtaining their structural information. Sometimes it is difficult to isolate or produce sufficient quantities of the target enzyme to study it directly. These obstacles hinder the successful entry of drug candidates into market. Therefore, there is an imperative demand for efficient methods that could enhance the drug discovery process.

Computational methods or molecular modeling techniques can be utilized to accelerate drug discovery process for obtaining new drug molecules. Today, almost every multi-national drug company and Contract Research Organization (CRO) involved in drug discovery has adopted computational methodology in different stages of the design process. Many computational methods complement one another and may be combined to help rationalize the drug discovery process. The ultimate challenge in drug design is to predict and explain activities of new drug molecules [3]. Modern drug discovery is a multidisciplinary project, where the role of various computational methods is to utilize experimental and predicted information in designing new active compounds thereby facilitating and enhancing the rate of discovery of appropriate chemical entities for lead optimization. As for as success stories of drug molecules generated through molecular modeling is concerned, it has been claimed that structure-based drug design methods have already contributed to the introduction of some drug compounds into clinical trials and for drug approval [4, 5].

Currently two major molecular modeling strategies are employed in drug design process, ligand-based drug design and structure-based drug design.

1.1.1 Ligand-based drug design

1.1.1.1 History of QSAR

To rationalize the drug design process, medicinal chemists frequently rely on structure–activity relationships (SAR). An SAR is study of structural changes made to a common core structure and observing how these changes affect activity. The resulting SAR model is then useful in lead optimization, where functional groups present on the core structure with physicochemical properties for high potency may be further explored.

A natural extension of the traditional SAR is the quantitative structure–activity relationship (QSAR), where some measures of chemical properties are correlated with biological activity to derive a mathematical representation of the underlying SAR. The first QSAR model was reported by Richardson in 1869 where the narcotic effect of a series of alcohols was correlated with molecular weight [6]. QSAR was revolutionized by Free-Wilson [7] and Hansch analysis [8] methods independently. A more detailed review describing the development of QSAR has been reported by Kubinyi [9, 10]. Although, the traditional QSAR models are useful for correlating chemical properties of structure with the biological activity, they do not account for how changes in three-dimensional molecular shape affect biological response. To include shape-related descriptors in QSAR modeling, attention was given to the development of 3D-QSAR. The first applicable 3D-QSAR method was introduced by Cramer et al. [11] which compares the steric and electrostatic molecular fields between a series of aligned molecules using PLS statistical analysis. The key advantage in using molecular fields is that their contributions to the model can be directly visualized in 3D, highlighting the structural features where changes can be made to improve potency. The power of the approach is indicated by hundreds of references that can be found in the literature applying the approach successfully to drug design, as well as many extensions of the original methods (e.g. CoMSIA [12], HINT [13], GRID/GOLPE [14], HASL [15], COMPASS [16] and AFMoC [17].

1.1.1.2 CoMFA Methodology

Comparative molecular field analysis (CoMFA) is the first approach with electrostatic and steric interactions of molecules with their environment, taking into account the 3D shape of the molecules. During the early stages of development of 3D-QSAR, researchers recognized that protein–ligand interactions were mainly non-covalent in nature. Based on this, it was thought that the electrostatic and steric components from molecular mechanics force fields may provide sufficient information to allow derivation of a predictive 3D-QSAR model. This was evaluated and confirmed in 1988 with the successful development of predictive CoMFA model for a set of steroids [11].

In CoMFA methodology the molecules of the data set need to be aligned geometrically. Different alignment rules can be applied on the molecules of the data set under study. The aligned molecules are then placed into cubic lattice one-by-one and interaction energies between the molecule under study and a defined probe atom (usually a proton or sp^3 hybridized carbon atom with a positive charge) are calculated for each lattice point. Each lattice point defines a position in space relative to the molecule. The steric van der Walls interactions are derived from the Lennard-Jones potentials as per Equation 1 and electrostatic forces utilize a simple Coulombic interaction term as per Equation 2. These potentials contribute only to the enthalpic term of the free energy of binding, although it may be important to take into account the entropic effects (e.g. hydropobicity) as well [18, 19].

$$E_{vdw} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} \quad \text{.....(1)}$$

$$E_{elect} = \left(\frac{q_1 q_2}{\varepsilon r}\right) \quad \text{.....(2)}$$

After calculation of interaction energies at each lattice intersection for each compound, data must be analyzed to find the correlation between changes in interaction energy and potency. Depending on the size and shape of the biologically active ligands, the number of lattice intersections may number in thousands and analysis of the data goes well beyond the limitations of multiple linear regression (MLR). PLS, on the other hand, is well suited to finding the correlation between a

small number of dependent variables and a large number of independent variables and make the derivation of CoMFA models possible. In the iterative procedure of PLS, new components (latent variables) are extracted so that each time, the degree of commonality between the dependent and independent variables is minimized. Usually, a maximum of five or six components are enough to generate a realistic model. The optimum number of components is traditionally determined by cross-validation, a technique that assesses the ability of a QSAR model to predict the biological data. In this technique one or more compounds are left out from the model and their biological activities are predicted on the basis of the model derived with the remaining compounds. This procedure is repeated until each compound has been predicted once [20]. The resulting statistical values are q^2 (square of the cross-validated correlation coefficient, Equation 3) and S_{PRESS} (standard deviation of the error of predictions, Equation 4).

$$q^2 = 1 - \frac{PRESS}{\Sigma (Y - Y_{mean})^2} = 1 - \frac{\Sigma (Y - Y_{pred})^2}{\Sigma (Y - Y_{mean})^2} \quad \text{.....(3)}$$

Where, PRESS = the sum of the squared error of predictions; Y= experimental value; Y_{pred} = predicted value; Y_{mean} = mean value of the experimentally observed activities.

$$S_{PRESS} = \sqrt{\frac{PRESS}{N - c - 1}} \quad \text{.....(4)}$$

The number of components used in the development of final model is recommended to be taken as that in which the S_{PRESS} value is smaller or it no longer decreases significantly [21]. This helps to avoid adding noise to the model. In fact, it has been shown that an excessive amount of noise (i.e. irrelevant grid points that do not contribute to explaining the biological activity) can be detrimental to the PLS prediction. Therefore, different variable selection techniques have been developed to eliminate irrelevant X-variables. The 'minimum sigma' filter eliminates those grid points that express lower variance in the field values than the user-defined cut-off value (e.g. 2.0 kcal/mol). More advance variable selection approaches include, for example, the GLOPE procedure [22] and the smart region definition method [23].

The non-cross validated r^2 (r^2_{ncv}) shows goodness of fit of the developed model i.e how well this particular model fits the input data. In an analogous way with

q^2 and S_{PRESS}, r^2 and the corresponding standard deviation can be calculated by replacing Y_{pred} with Y_{cal} (activity value calculated by the model) in Equation 3 and 4. The predictive r^2 (r^2_{pred}) is obtained if the model is used to predict the activities of a set of compounds not included in the model [20].

1.1.1.3 Modifications in CoMFA methodology

Perhaps the single most important drawback for CoMFA analyses is the sensitivity of the technique to the alignment of the input structures, and usually a structural scaffold is required to make this consistent, which also makes CoMFA most applicable within congeneric series. To avoid the alignment problem, several modifications have been made into the original method. Pastor et al. have developed the Grid independent (GRIND) descriptors approach [24], whereby GRID MIFs for each molecule are filtered by energy and MIF node pair distance. The GRIND descriptors represent the product of the interaction energy for each pair of nodes, binned by distance, with only the maximum products being retained. After statistical analysis such as PLS, the QSAR model is obtained and the strongest contributions can be traced back to the node pairs (analogous to pairs of pharmacophore points) which can then give insight into improving activity. Recently, a modification that increases the specificity of the descriptors has also been described [25], and it also improves the MIF point sampling. It should be noted that in such an approach, finding appropriate molecular conformation is still a challenging task. Additionally, another GRID-based method has been developed that compresses the 3D GRID maps into a few quantitative 2D descriptors; typically more relevant for ADME modelling, it has also been used successfully to model the SAR of anti-HIV quinolones [26]. The topomer descriptors discussed above have also been used in the CoMFA context, solving the alignment and conformation problems. The topomers are generated by deterministic rules that consistently describe fragments in 3D and are aligned according to predetermined rules. However, this does require the compounds to share some kind of common core or 'equivalent' acyclic bond where the structures are split into fragments for comparison. Once the topomers have been generated (and with them the alignments), the CoMFA statistical analysis is performed with a few minor modifications to 'standard CoMFA' [27]. The approach by Vinter et al. for solving the conformational problem is to compare the XED (Extended Electron Distributions) -derived field minima points between pairs of active structures, in each case

considering several generated conformations. Cross-correlating these pairs, or duos, from several active molecules yields the bioactive conformation hypothesis. Aligning other active structures to this field template provides a combined field and volume score that has been shown to correlate well with activity and exhibits reasonable predictivity [28].

1.1.1.4 CoMSIA Methodology

The comparative molecular similarity indices analysis (CoMSIA) method was developed to improve the limitation of the potentials used as steric and electrostatic fields in CoMFA. Thus, CoMSIA method is useful to evaluate hydrophobic, hydrogen bond donor and hydrogen bond acceptor characters along with steric and electrostatic properties of molecules. In-fact, the original version of CoMSIA includes the calculation of steric, electrostatic and hydrophobic fields; H-bond donor and acceptor fields have been introduced later into the method [29]. In CoMSIA methodology, instead of field descriptors based on Lennard-Jones- and Coulomb-type potentials, molecular descriptors based on similarity indices of the aligned molecules are computed. The energy potentials used in COMFA are very steep near the van der Waals surface of the molecules and they produce singularities at the atomic centers. In order to avoid too large energy values, arbitrary cut-off values have to be defined and evaluation of the potentials are restricted to regions outside the molecules. To overcome such problem, CoMSIA utilizes a Gaussian-type function for the distance dependence between the probe and the atoms of the data set molecules (Equation 5).

$$A_{F,k}{}^{q}(j) = -\sum_{i=1}^{n} w_{\text{probe},k} w_{ik} \mathrm{e}^{-\alpha r_{iq}{}^{2}} \quad \text{.....(5)}$$

where ω_{ik} is the value of property *k* of atom *i*. The properties include a steric contribution as a third power of atomic radii, electrostatic properties as pre-calculated point charges, atom-based hydrophobicity parameters and representative positions of H-bond donor and acceptor distribution. The properties $\omega_{probe,k}$ of the utilized probe include a radius of 0.1 nm, +1 charge, +1 hydrophobicity, and +1 for both H-bond donor and acceptor properties.

The similarity indices can be calculated at all grid points inside and outside the molecule using a common probe which is placed at intersections of a regularly sampled 3D grid box similar to CoMFA. A grid resolution of 0.1 nm, a grid box extension of 4.0 nm and an attenuation factor α of 0.3 were used in the original CoMSIA report [12]. This makes CoMSIA relatively intensive to changes in grid spacing or orientation of the aligned molecules with respect to the lattice [30]. Statistical evaluation of similarity indices is done by PLS in the same way as in CoMFA.

1.1.1.5 Advantages of CoMSIA over CoMFA

The Gaussian function used in CoMSIA methodology for calculating distance dependence (r_i - r_j), requires no definition of a cut-off for field values. CoMSIA is not so sensitive to changes in orientation of the superimposed molecules than CoMFA, or to translations and rotations of the compound set with respect to the grid box [30]. Hence, CoMSIA may result in better predictive values. A third benefit for CoMSIA over CoMFA is the better visualization and interpretation ability for the regions that are predicted to be important for activity. In CoMFA these regions are highlighted as contour maps in the 3D space that surrounds molecules [11], whereas in CoMSIA, the regions are located within space that is occupied by the molecules and thus, directly pinpointing structural features that are important for activity [31].

1.1.1.6 Validation of QSAR models

The process of QSAR model development is divided into three steps. The first stage includes the selection of data set for QSAR studies and the calculation of molecular descriptors. The second stage deals with the selection of a statistical data analysis technique for correlation. The correlation process can not be done using standard multiple linear regression (MLR) methods due to the huge number of investigated independent variables [9]. Instead, a partial least square (PLS) analysis [32] is the method of choice. PLS extracts principle component vectors (PCs), which are obtained as linear combinations of the original independent variables in order to maximize the correlation between independent and dependent variables. All PCs are orthogonal; therefore, a new PC can be used in explaining only the data that is not already described using the existing PCs. The final part of QSAR model development is model validation in which estimates of the predictive power of the model are

calculated. The validation of a method is done to establish the reliability and relevance of the method for a particular purpose. Reliability refers to the reproducibility of results, the relevance is related to the scientific use and practical usefulness and the purpose refers to the intended application. This predictive power is one of the most important characteristics of QSAR models.

To evaluate the predictive ability, 3D-QSAR model need to be cross-validated using optimum number of principle components [9, 20, 33]. Cross-validation can be carried out by using leave-one-out procedure. In leave-one-out procedure, cross-validation is carried out by omitting one compound at a time from the model building and then predicting its activity with a model that is generated from the rest of the compounds. The same procedure has been repeated for each individual compound. The outcome of this procedure is a cross-validated correlation coefficient q^2, which is calculated according to the Equation 3

A random group cross-validation procedure should be repeated several times as the difference in distribution of the compounds in the group during each PLS analysis affects the q^2 value. Clark et al. studied that a q^2 value that is greater than 0.25 is very unlikely to result from a chance correlation [34] and q^2 above 0.3 indicates that probability of chance correlation is less than 5% [35]. Many authors consider high q^2 (> 0.5) as an indicator or the proof of higher predictive power of the developed QSAR model. They do not test the model for their ability to predict the activity of the compounds outside the training set (test set) and they claimed that these models were highly predictive. Thus, every QSAR model should be characterized by a reasonably high q^2 for their ability to accurately predict the biological activities of compounds not included in the training set.

To establish model robustness, *Y*-Randomization of response is another important validation criterion. The method consists of repeating the QSAR model derivation calculation procedure, but with randomized activities. The subsequent probability assessment of the resultant statistics is then used to gauge the robustness of the model developed with the actual activities. It is often used along with the cross-validation. If all QSAR models obtained in the *Y*-randomization test have relatively high R^2 and LOO q^2, it implies that an acceptable QSAR model cannot be obtained for the given data set by the current modeling methods.

It is still common not to test QSAR models (characterized by a reasonably high LOO q^2) for their ability to predict accurately biological activities of compounds from an external test dataset, i.e. those compounds, which were not used for the model development. The high q^2 does not imply automatically a high predictive ability of the model. In order to both develop the model and validate it, one needs to split the whole available dataset into the training and test set. In fact, the lack of correlation between the high value of the training set q^2 and the high predictive ability of a QSAR model has been noticed earlier in the case of 3D QSAR [36-38]. These studies indicated that while high q^2 is the necessary condition for a model to have a high predictive power, it is not a sufficient condition. The only way to estimate the true predictive power of a model is to test it on a sufficiently large collection of compounds from an external test set.

Even today, many studies continue to consider q^2 as the only parameter characterizing the predictive power of QSAR models. Tropsha et al. have demonstrated the insufficiency of the training set statistics for developing externally predictive QSAR models and formulated the main principles of model validation [39], and incorporated new rigorous validation criteria. According to Tropsha et al. a good model should qualify in the following criteria:

a) $q^2 > 0.5$

b) $r^2 > 0.6$

c) $[(R^2 - R_0^2) / R^2] < 0.1$ $[(R^2 - R'_0{}^2) / R^2] < 0.1$

d) $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$

where q^2 is the cross-validated correlation coefficient; R^2 or r^2 is the correlation coefficient for the experimental (y) vs. predicted (ỹ) activities for the test set molecules; R_0^2 and $R'_0{}^2$ are the correlation coefficients for the regression line passing through the origin for y vs. ỹ and ỹ vs. y respectively. k and k' are the slopes for regression line passing through origin $y^{r0} = k\tilde{y}$ and $\tilde{y}^{r0} = k'y$ and were calculated by equations (6) and (7).

$$k = \frac{\sum y_i \tilde{y}_i}{\tilde{y}^2} \quad \ldots\ldots\ldots\ldots (6)$$

$$k' = \frac{\sum y_i \hat{y}_i}{\hat{y}^2} \quad \text{..... (7)}$$

An additional parameter, modified R^2 (R^2_m) [40, 41] for internal R_m^2 (LOO), R_m^2 (overall) and external R_m^2 (Test) validation was studied by Roy et al. The R^2_m can be described as follows:

$$R_m^2 = R^2 \left(1 - \sqrt{|R^2 - R_0^2|}\right) \text{.....(8)}$$

1.1.1.7 Pharmacophore modeling

Pharmacophore modeling is considered to be an important tool in drug discovery process. A 3D-pharmacophore is defined as that set of properties and their arrangement in 3D space that a compound must posses for it to be active in a particular biological test [42, 43]. A pharmacophore can also be defined as the 3D arrangement of features that is crucial for a ligand molecule in order to interact with a target receptor in a specific binding site [44]. Many computational methods for pharmacophore identification have been introduced [45]. Some methods, which are called the direct methods, use both ligand and receptor information. However, in most cases the structure of the target receptor is unknown and the only available information is a set of ligands that have been observed to interact with the receptor. In such cases, only the second type of methods, the indirect methods are applicable, since they are based on ligand information alone.

A pharmacophore model can be used to explain the SAR of a series of ligands, to guide the molecular alignment process in 3D-QSAR and to search for novel structures with the same pharmacophore. It can serve as an important model for virtual screening, especially in the case where the 3D structure of the receptor is unknown and docking techniques are not applicable. Besides screening, pharmacophore is a powerful model used in other applications of drug development, like *de novo* design, lead optimization, ADME/Tox studies and Chemogenomics [46, 47].

While developing a pharmacophore model, it is of utmost importance to use correct structures as well as the use of consistent and mechanistically homogeneous biological data collected following the same protocol for all the compounds having a similar binding mode [48]. It is also important to remember that usually multiple

pharmacophore alignments are possible for a data set. The quality of a pharmacophore model can only be measured by the models' success in prospective application to drug design i.e. how well it can facilitate lead optimization and the accuracy in selection of active compounds to be synthesized, or how many active hits can be found in virtual screening with a query based on the model [48].

Drug-like molecules may adopt many possible conformations. The specific conformations that the input ligands adopt in the active site of the receptor are usually unknown. Additionally, they cannot be assumed to be the ones with the lowest energy [49]. Therefore, all the feasible conformations of the ligands should be considered during the search for the common pharmacophore. Most methods perform the conformational search in the initial stage. Examples for such methods include RAPID [50] MPHIL [51] DISCO [52, 53] PHASE [54] Catalyst/HipHop [55-59] Catalyst/HypoGen [55, 59, 60] and others [61-63]. These methods generate a discrete set of conformations for each ligand with the goal of covering its whole conformational space. Recently, Schneidman-Duhovny et al. [64] reported first web server for detecting 3D pharmacophores shared by known active ligands in the absence of structural information on the target receptor. It has been demonstrated that the deterministic and efficient algorithm behind the server allows a fast and reliable detection of pharmacophores with explicit consideration of the flexibility of the ligands. Dror et al. [65] showed that pharmacophore hypothesis generated by PharmaGist web server were found to be similar to a pharmacophore hypothesis computed by Catalyst which confirms the efficiency and reliability of PharmaGist web server.

1.1.2 Structure-based drug design

1.1.2.1 Molecular Docking

"Molecular docking" explores the binding modes of two interacting molecules, depending upon their topographic features or energy-based considerations [66], and aims to fit them into conformations that lead to favourable interactions. It therefore constitutes an essential step in determining the active conformation of a drug, i.e. its conformation when bound to the receptor. Hence, prediction of binding orientations of small molecules in a protein/DNA binding site has become increasingly important in drug design. Although it is possible experimentally to elucidate the structure of

protein–ligand complexes using crystallographic or nuclear magnetic resonance (NMR) methods, it would be difficult to do so for all ligands in a medicinal chemistry project. Molecular docking may provide this important information. Docking is often approximated to a Lock and Key process where the conformation of a ligand and receptor do not change during binding. This is the simplest to simulate, but is generally thought to be unreasonable. Ligands are often flexible and occupy multiple conformations in solution. Although, the conformations of receptors are better defined, they too can change, particularly on ligand binding in the so called Induced Fit model. Molecular Docking has a wide variety of uses and applications in Drug Discovery, including structure activity studies, lead optimization, finding potential lead by virtual screening, providing binding hypothesis to facilitate prediction of mutagenesis studies, assisting X-ray crystallography in the fitting of substrate and inhibitors to electron density, chemical mechanism studies and combinatorial library design.

Large number of docking program are available for use in virtual screening and every program has different algorithms to handle ligand and protein flexibility, scoring functions and CPU time to dock a molecule to a given target. If in the process of docking, both ligand and the target are treated as rigid bodies [67], the conformational flexibility of ligands can be taken into account by creating a collection of conformers and docking each one of them separately into target site. Instead, the conformational search of ligands can be explored during the docking process. Some examples of semi-flexible docking approaches also exist viz. incremental growth methods [68], genetic algorithm (e.g. GOLD) [69], Tabu search (e.g. PRO_LEADS) [70], and combined Monte Carlo and simulated annealing methods (e.g. Dock Vision) [71]. Molecular docking using rigid protein is sometimes an inaccurate approximation, since the binding of a ligand can induce large conformational changes in a receptor binding site [72, 73]. One option to consider protein flexibility is to use an ensemble of protein conformers. Example of such a docking program is FlexE, wherein various protein conformers are superimposed and it treats the dissimilar protein regions as distinct alternatives [74].

Once a pose has been generated for a ligand in the binding site, scoring function needs to be applied to rank the quality of the pose with respect to other poses of the compound based on binding energy of association of each pose. Scoring

function estimates the free energy of binding of a ligand in a target-ligand complex. Scoring functions are used to optimize the placement of the ligands during the docking process and are applied to rank the resulting ligand poses with respect to the other poses and ligands [75, 76]. There are wide choices of scoring functions available that are grouped into force-field based, empirical and knowledge-based functions [77]. It is a well known fact that these fast scoring methods do not perform so accurately as the time-consuming free energy perturbation technique [78]. Using a combination of one or more scoring functions (i.e. consensus scoring) has been reported to improve the results [79-81].

Some of the docking programs which are used in drug designing are AUTODOCK [82], CDOCKER [83], DOCK [84], FlexX [85, 86], GOLD [87], GLIDE [88, 89].

1.2 Telomerase and Cancer

1.2.1 Telomerase

Telomerase enzyme was first discovered in the unicellular ciliate, Tetrahymena where it was shown to use an unusual mode of DNA synthesis for polymerization of telomeric DNA [90, 91]. It is a catalytically active ribonucleoprotein enzyme that consists of two major components, a reverse transcriptase catalytic subunit (hTERT 127 kD; Gene ID 7015) and an RNA subunit (hTR; Gene ID 7012). The RNA subunit functions as a template for synthesis of telomeric DNA with the help of TERT directly on to the 3′ end of the chromosomes. The amino-terminal moiety of the TERT protein is essential for the nucleolar localization, and multimerisation. The COOH-terminal region is involved in the processivity of the enzyme and is indispensable for *in vivo* activity [92]. The central region contains the motifs characteristic of reverse transcriptase proteins and a conserved RNA-binding domain required for specific binding of hTR by the hTERT subunit [93]. The human telomerase RNA (hTR) extends on 451 nucleotides and contains 11-nucleotide long template sequence for telomeric repeat synthesis. Ten conserved helical regions were proposed in vertebrate telomerase RNA including four distinct structural domains: the pseudoknot domain, the CR4eCR5 domain, the Box H/ACA domain and the CR7 domain [94]. Earlier it was shown [95] that dyskerin

(57 kD) interacts with hTR; however, it has been proved very recently that the presence of dyskerin protein is essential for the catalytic activity of the enzyme [96]. There is considerable evidence from *in vitro* reconstitution studies that telomerase exists as a dimer [97, 98]; therefore, it was proposed that the catalytically active human telomerase is composed of two molecules each of hTR, hTERT and dyskerin (about 670 kD). The enzyme is almost undetectable in most normal somatic cells except in proliferative cells of renewing tissue [99, 100]. However, recent evidence shows that a small amount of the enzyme activity is detectable in normal cells also as they enter S-phase but disappear when the cells go into G2-phase [101]. Telomerase enzyme is involved in telomere capping [102] and in the DNA-damage response [103].

In addition to TR and TERT, more than 30 proteins have been proposed to be associated with the enzymatic telomerase complex [104].

1.2.2 Telomeres

Human telomere (telos = end; meros = part) consists of tandem repeats of the hexameric DNA sequence TTAGGG, ranging from 15 kb at birth and less than 5 kb in chronic diseases. Telomeres are dynamic nucleoprotein complexes that cap the ends of linear eukaryotic chromosomes [105]. The role of telomeres for chromosome stability was reported in the early 1940s by experiments in maize [106]. Telomeres have two important duties. First, they protect the chromosome ends from destructive nucleases and other damaging events e.g. end-joining. Second, they enable the ends to be completely replicated. Due to the nature of lagging strand DNA synthesis, conventional DNA polymerases are unable to completely replicate the ends of chromosomes resulting in loss of 50-100 bp of telomeric DNA replication [107]

The fact that these bases do not code for any genetic information does not diminish their importance. It is now know that they are a site of dynamic activity beyond being the biologic timepiece [108]. They have a unique T-looped configuration where the telomere bends back on itself [109]. The overhanging guanine-rich single strand is placed into the double stranded telomere. This creates a second smaller d-loop by displacing one of the telomere strands. This structure appears to protect the telomeres from end to end fusion with other chromosomes and

from cell cycle checkpoints that would otherwise recognize the telomeres as chromosome breaks requiring repair (reviewed in [110]).

1.2.3 Role of telomerase in cancer

1.2.3.1 Experimental evidences

During the past decade, more than 500 research/review articles have been published on telomerase. The first detection of telomerase was done in ovarian cancer in 1994 [111]. Nearly the complete spectrum of human tumors has been shown to be telomerase positive (Figure 1.1). One of the largest compilations of telomerase activity in human tumor tissues has been recently published [112]. In this series which include 601 human tumor samples, telomerase was detected in 476 samples (79%). So much high percentage makes telomerase both an attractive cancer marker and a cancer target [113]. Experimental result of a wide series of 100 neuroblastoma suggests that the telomerase expression could be responsible in the evolution of neuroblastoma [114]. Considering normal tissues, telomerase is present in normal human white blood cells at very low level. Telomerase is also detected at low levels in germline tissues and in some mitotic stem cells from normal epithelia [115].

Taken all together, clinical pharmacology data support the concept that telomerase is a relevant target in oncology. The expression of telomerase between normal and tumor tissues is indeed remarkable and largely superior to those observed for classical chemotherapeutic targets such as topoisomerases, tubulin/microtubules, enzymes implied in DNA metabolism/replication. This suggests that antitelomerase treatment would turn in low toxicity for normal tissues [117].

To demonstrate whether telomerase is required for cell viability and tumor formation, knockout mice were reared by elimination of the gene encoding the telomerase RNA component [118] which gave an additional clue that the telomere-associated events are indeed relevant to carcinogenesis [119].

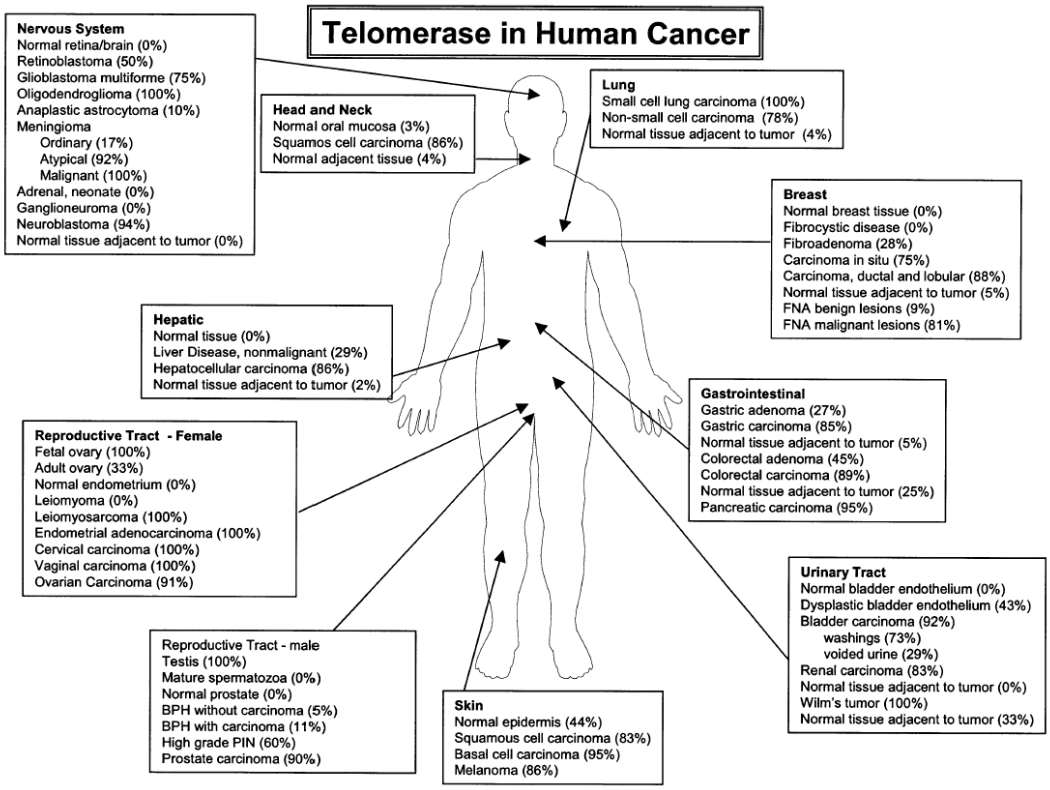


Figure 1.1: Summary of telomerase activity expression in human cancers from a review of the literature. Tumor samples were assayed by the TRAP assay. Percentages in parentheses refer to the number of samples that were telomerase positive compared to matched control tissue [116].

The human telomerase reverse transcriptase (hTERT) catalytic subunit has been cloned and telomerase activity can be transiently reconstituted in normal human diploid cells which express low levels of the template RNA component of telomerase but which do not express hTRT [120-123]. Thus cell transfection experiments further confirm the hypothesis that telomere shortening causes cellular senescence.

To examine the functions of telomerase in an immortalized cell, the antisense approach has also been used [124]. Antisense hTR expression constructs were introduced into Hela cells; in parallel, control expression vectors lacking the antisense hTR coding sequences were electroporated into the same cells. The ability to induce cell crisis in a tumor cell line such as Hela cells by expressing an antisense against the hTR, supports telomerase inhibition to be a 'cause' rather than a 'consequence' and also to be a potential therapeutic approach for fighting cancer [117].

1.2.4 Approaches for targeting telomerase in cancer therapy

1.2.4.1 Targeting the RNA component of telomerase (hTR)

1.2.4.1.1 Antisense oligodeoxynucleotides

Oligodeoxynucleotides (ODN) consist of short stretches of DNA that are complementary to a target RNA. The mechanism of action for most of the applications is to hybridize the complementary RNA by Watson–Crick base pairing and inhibit the translation of the RNA by a passive and/or active mechanism. Telomerase presents itself as an interesting therapeutic target for these drugs because it possesses a functional RNA component as part of its structure. The template region of hTR must be exposed to add new telomeric repeats onto the chromosome, making this an accessible target for the ODN activity. However ODNs have several drawbacks in drug development. The major problem is their cellular delivery because without a transfecting agent, these drugs do not easily enter cells in culture. It was observed that *in vivo* they cross the cell membrane by a poorly understood endocytic mechanism. Once inside the cell they are subjected to undergo degradation by a variety of exo- and endonucleases [125].

Numbers of studies have been published on telomerase inhibition using antisense approaches aimed both at the template and at non-template regions of hTR. The first report was by Feng et al [124.] who used a construct expressing an antisense transcript to the first 185 nucleotides of the RNA and introduced them into HeLa cells.

1.2.4.1.2 Peptide Nucleic Acids (PNAs)

PNAs are analogs of RNA and DNA, in which the pentose-phosphate backbone is replaced by an oligomer of *N*-(2-aminoethyl)glycine, making them resistant to degradation by endo- and exonucleases. This neutral backbone additionally enhances the affinity and specificity of hybridization to the RNA targets. Human telomerase can be inhibited in cells by PNA oligonucleotides complementary to the telomere templating region of hTR [126].

1.2.4.1.3 Hammerhead ribozymes

Hammerhead ribozymes are small RNA molecules that possess specific endo-ribonuclease activity. They consist of a catalytic core flanked by anti-sense sequences that function in the recognition of the target site. Yokoyama et al. used this approach and concluded that the ribozyme targeting the template region proved to be the most efficacious in reducing telomerase activity and additionally led to telomere shortening over a four-week period [127]. A similar study of targeting the hTR template in melanoma cells using a ribozyme, showed a reduction in telomerase activity [128].

1.2.4.2 Targeting telomerase catalytic protein subunit-dominant-negative mutant telomerase hTERT

Certainly, dominant-negative mutant telomerase hTERT (mutants that are catalytically inactive but still able to bind and sequester hTR) has also been considered as a target for cancer therapy. A study from Hahn et al. has shown to shorten telomere and induce apoptosis and reduce tumorigenesis in mice [129, 130]. Adenoviral delivery of anti-hTERT ribosomes, small catalytically active RNA molecules that cleave their RNA substrate in a sequence dependent manner in ovarian cancer cells results immediate apoptosis without causing telomere shortening [131]. Small molecules and natural compounds have been shown to act as potent telomerase inhibitors [132, 133]. As most of the attempts telomerase inhibition involve slow process of telomerase shortening, long treatment time is required before therapeutic effect is achieved. This problem can be solved by combining telomerase inhibitors with DNA-damaging chemotherapy.

1.2.4.3 Targeting G-quadruplex DNA

Guanine-rich (G-rich) stretches of DNA have a high tendency to self-associate into planar guanine quartets (G-quartets) to give unusual structures called G-quadruplexes. It was first reported by Davies and co-workers in 1962 [134]. G-Quadruplexes are a family of nucleic acid secondary structures stabilized by G-quartets formed in the presence of cations. With the advent of X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and other powerful technologies, the structures of many G-quadruplexes have been resolved. Each quartet is composed

of four guanines held together by a cyclic arrangement of 8 hydrogen bonds (Figure 1.2). The presence of a central cation (K^+ or Na^+) helps to maintain the stabi-

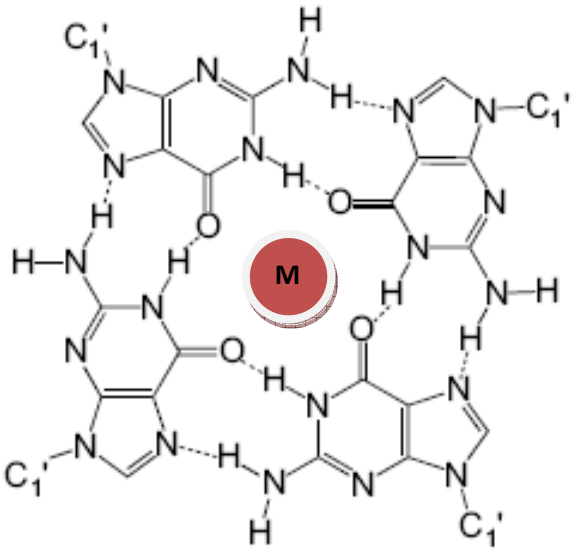


Figure 1.2: Structure of G-quadruplex with a centrally placed metal ion.

lity of the structure, which may be very stable under physiological conditions. Quadruplexes have now been studied in organisms from E. coli [135] to humans. The existence of G-quadruplexes *in vivo* was initially established in ciliates using specific antibodies [136, 137].

Optimal telomerase activity requires an unfolded single stranded substrate, because G-quadruplex formation directly inhibits telomerase elongation *in vitro* [138]. Therefore, ligands that selectively bind to and stabilize G-quadruplex structures may interfere with telomere conformation and telomere elongation. The formation of quadruplexes in telomeric DNA results in the inhibition of telomerase to extend the telomeres. Thus, these structures have become important drug-design targets for the treatment of Cancer. The number of known G-quadruplex ligands has grown rapidly over the past few years. The demonstration that telomeres are a significant target for these ligands was provided by a study using a radiolabeled Gquadruplex ligand. The results showed that the radiolabeled molecules preferentially bind to the terminal parts of chromosomes [139].

1.2.5 Structure and Topology of G-quadruplex

DNA is considered to be an important drug target in anticancer therapies. Conventionally, the development of alkylating agents as anticancer agents is highly dependent on the discovery and evolution of the DNA duplex and its associated processes. Unfortunately, these drugs have drawbacks like extreme cytotoxicity and nonspecificity. To solve these problems extensive efforts have been directed toward the discovery of new agents with improved selectivity and lesser cytotoxicity [140]. Other than the typical double helix DNA proposed by Watson and Crick, DNA can self-associate into other biologically relevant structures, known as G-quadruplexes. This secondary DNA structure represents a new drug target for DNA-binding compounds.

1.2.5.1 The building blocks of G-quadruplexes

G-Quadruplex structure was first reported by Gellert et al. over 40 years ago [134]. The building block of G-quadruplexes is the G-quartet (G-tetrad). Four guani-

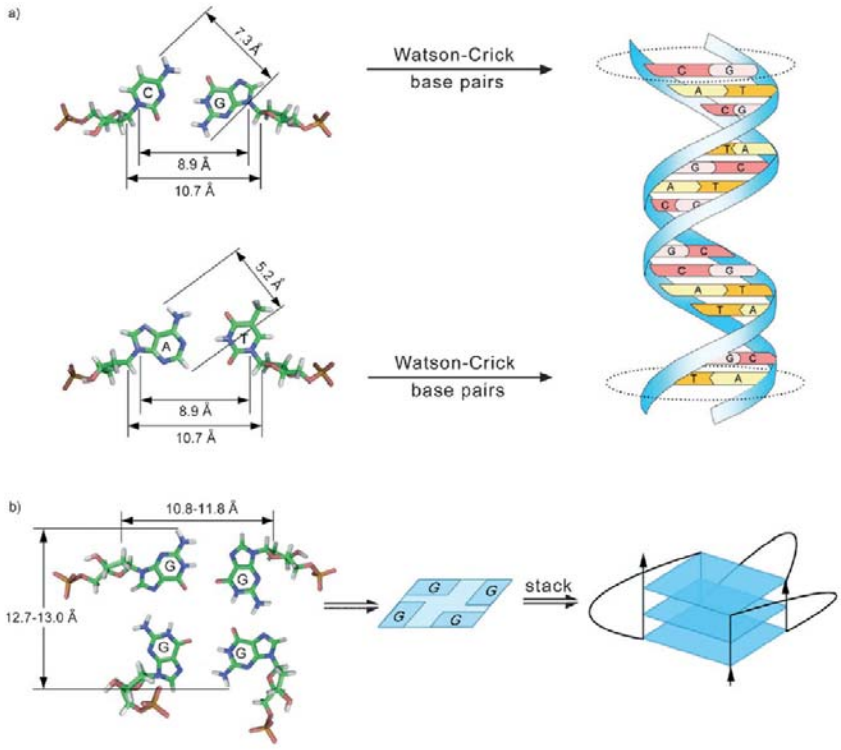


Figure 1.3 : Comparison of the dimensions of duplex and G-quadruplex DNA structures. a) The double helix and its base pair surface. b) The quadruplex structure and the G-quartet surface. [141]

nes are associated into a cyclic Hoogsteen hydrogen bonding motif in which each guanine base forms two hydrogen bonds with its neighbors as shown in Figure 1.3.

The advent of availability of crystal structures of G-quadruplex has shown that the G-quartet has a square aromatic surface, the dimensions of which are much bigger than the Watson-Crick base pairs (Figure 1.3) and this difference constitutes the basis for designing G-quadruplex specific ligands [142, 143].

1.2.5.2 The basic topology and structure of G-quadruplexes

G-Quadruplex structures exhibit extensive structural diversity and polymorphism. The structural polymorphism occurs mostly from the nature of the loop, such as variations of strand stoichiometry, strand polarity, glycosidic torsion

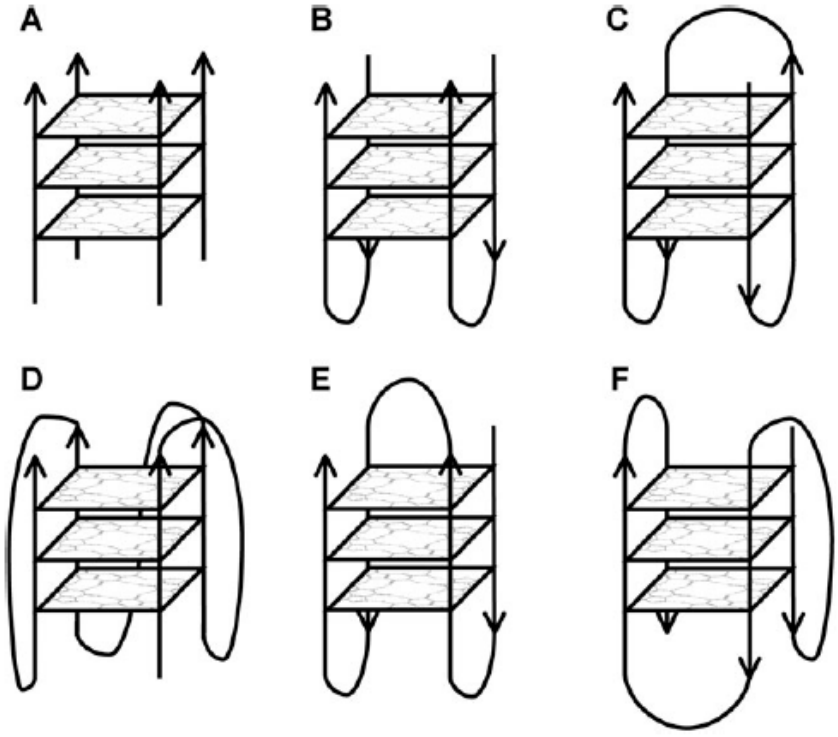


Figure 1.4: G-Quadruplexes can adopt a range of different stoichiometries and folding patterns. (A) Tetramolecular structure with all strands parallel; (B) bimolecular antiparallel structure with adjacent parallel strands; (C) unimolecular antiparallel structure with alternating parallel strands; (D) unimolecular parallel structure with three double chain reversal loops; (E) unimolecular antiparallel structure with adjacent parallel strands and a diagonal loop; (F) unimolecular mixed structure with three parallel and one antiparallel strands. All three structures (D), (E) and (F) have been observed for the human telomeric repeats.

angle and the location of the loops that link the guanine strands. Due to presence of directionality to the strands, from the 5′ end to the 3′ end, there are topological variants possible for these four strands. All four strands may be parallel, three parallel and one in the opposite direction (antiparallel), or there may be two in one direction and two in the other, either with the parallel pairs adjacent to each other or opposite to each other. Some examples of different stoichiometries and folding patterns are shown in Figure 1.4.

At molecular level, different directionality of the strands relates to the conformational state of the glycosidic bond between the guanine base and the sugar. This may be either *syn* or *anti*, as depicted in Figure 1.5. When all four strands are parallel, all of the bases are in the *anti*-conformation and the grooves between the backbones are all of equal size. When any of the strands are antiparallel, the bases must be in the *syn* form in order for the hydrogen bonds to be formed correctly; this results in grooves of different sizes. When successive guanines are both *anti* or both *syn*, the groove is medium in size; if the first is *anti* and the second *syn*, the groove is wider, and if the first is *syn* and the second *anti*, then the groove is narrower. G-Quadruplexes may be comprised of four separate strands, forming tetramolecular G-quadruplexes. Alternatively, they may be formed from two strands, each with two sets of contiguous guanines, or just from one strand folding back on itself to form an intramolecular structure. In either of these cases, there will be loops that serve to connect the strands of the structure together [144].

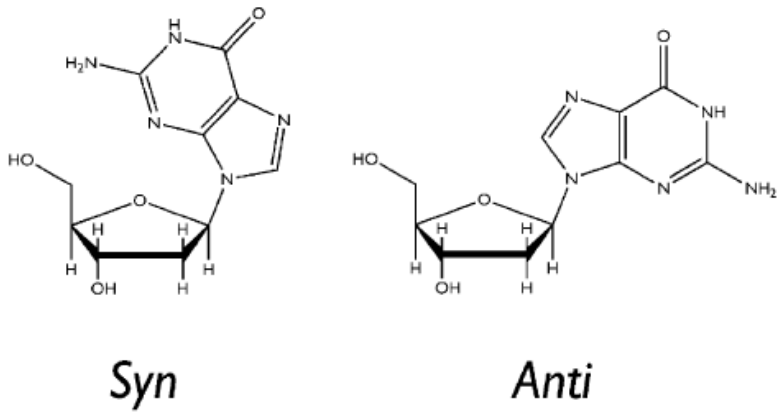


Figure 1.5: The bond between the base and sugar can rotate. It has two preferred conformations, *syn* and *anti*.

Many G-quadruplexes can coordinate with cations such as K^+ and Na^+. The hole between G-tetrads is well suited to coordinating cations of this size because the two planes of tetrads are lined by eight carbonyl oxygen atoms (with strong negative electrostatic potential) that create a central negatively charged channel inside the G-tetrad stack (Figure 1.6) [141]. On the basis of the finding reported by Zahler and colleagues [138], substantial efforts have been made to identify synthetic and natural compounds that lock telomeric DNA in a G-quadruplex conformation and thus impede telomere elongation *in vivo*. Given the requirement for telomere maintenance in the indefinite proliferation of cancer cells, such molecules are promising candidates as anticancer drugs.

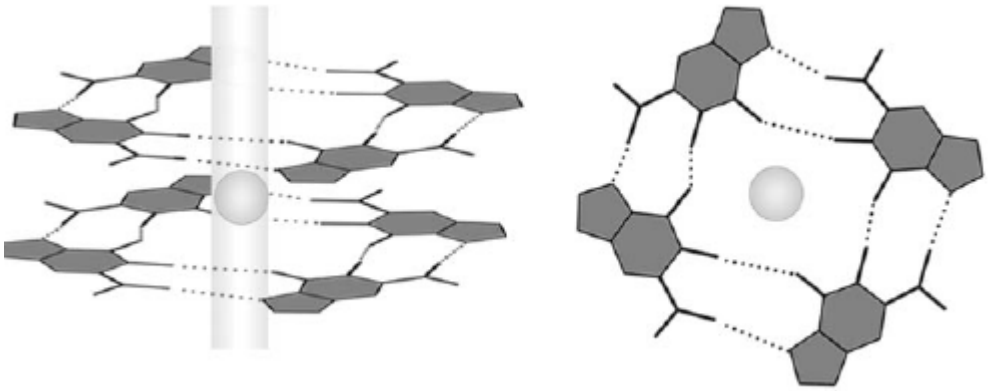


Figure 1.6: G-Quadruplex ion channel

A large number of G-quadruplex-interacting ligands from many chemical classes have been reported as described in the next chapter.

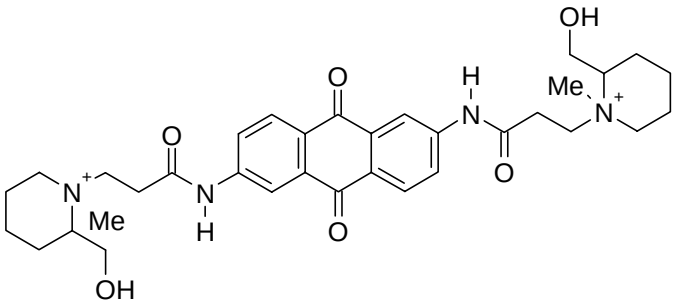
2. REVIEW OF LITERATURE

2.1 G-Quadruplex stabilizing ligands as telomerase inhibitors

G-Quadruplex structures have materialized as a significant biological target for telomerase inhibitors. Search for selective G-quadruplex stabilizing ligands to block the telomerase activity has been taken up by several research groups. Although the discovery of a potent and selective G-quadruplex stabilizing ligand is still a long way, some promising lead compounds have been reported in the literature.

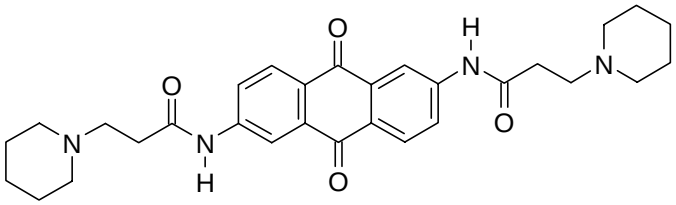
2.1.1 Anthraquinone and Fluorenone Derivatives

Sun et al. published first report of a non-nucleoside small molecule telomerase inhibitor (**1**) (IC_{50} 23 μM) and proved that this compound inhibited the telomerase enzyme by interaction with G-quadruplex [145].



(**1**)

In continuation of the above study Neidle group prepared a range of isomeric 1,4- and 2,6-diamidoanthraquinone derivatives and examined them for telomerase

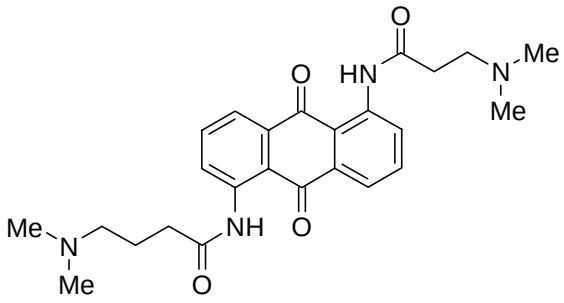


(**2**)

inhibitory activity. Structure-activity relationship studies showed that compounds containing piperidine or substituted piperidines as end groups were more potent

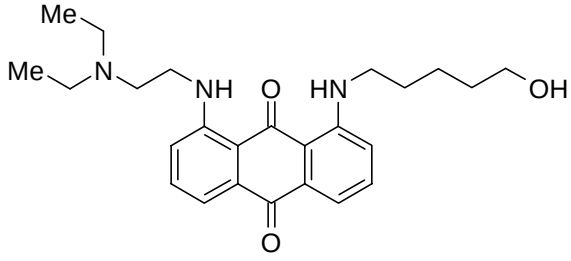
inhibitors. Among all of the tested compounds, compound (**2**) was found to be the most potent telomerase inhibitor with IC_{50} value of 4.5 μM [146].

To study the structural requirements a series of regioisomeric 1,5-, 1,8- and 2,7-anthraquinone derivatives were synthesized and evaluated for telomerase inhibitory activity. It was found that telomerase inhibition (IC_{50} 1.3-3.7 μM) was achieved within each isomeric series. Among the three isomeric series of compounds, compound (**3**) (1,5- series) was found to be the most potent telomerase inhibitor with IC_{50} value of 1.3 μM.



(**3**)

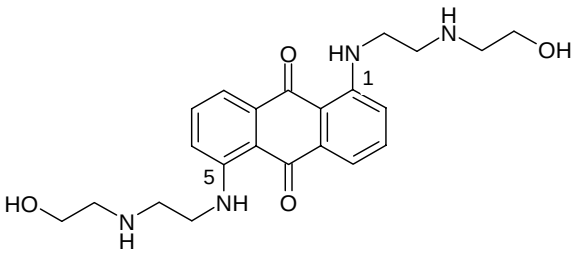
Molecular modeling studies have been carried out for number of mono- and bis-(amino)-substituted anthraquinone derivatives [147], wherein the ability of these compounds to stabilize G-tetraplex DNA has been examined. The results obtained were compared with experimental data for telomerase inhibition that showed good correlation. Compound (**4**) showed a strong inhibition (43.3%) at a concentration of 10 μM. 1,8-Disubstituted anthraquinone derivatives inhibited telomerase enzyme through intercalation of G-quadruplex DNA.



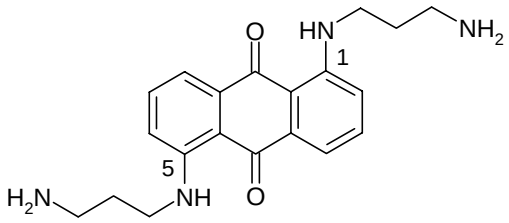
(**4**)

Mitoxantrone and ametantrone related 1,4- and 1,5-difunctionalized tricyclic anthraquinone derivatives have been reported as human telomerase inhibitors [148].

Compounds (**5** and **6**) were found to be the most potent telomerase inhibitors with IC_{50} 0.2 and 0.5 μM, respectively. The study revealed that position of the side chains



(**5**)

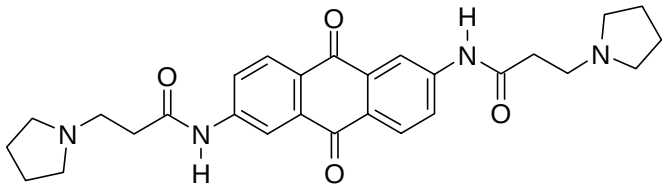


(**6**)

on the anthraquinone core is important for telomerase inhibition. Further, primary or tertiary amino groups are required for the telomerase inhibitory activity.

A small library of 2,6- and 2,7-disubstituted anthraquinone-amino acid conjugates was synthesized to identify structural requirements for telomerase inhibition [149]. Structural requirements were studied with the consideration of direction of the amide linkage between the linker and the anthraquinone molecules i.e. amide or "reversed" amide, length of the linker, charge, lipophilicity and bulkiness of the amino acid in the side chain. In all the classes of derivatives, Lys or Arg residue containing side chains resulted in optimal telomerase inhibition in the range of 0.8-1.5 μM (Table 2.1)

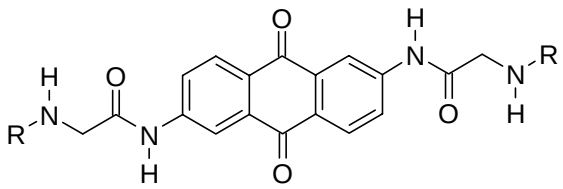
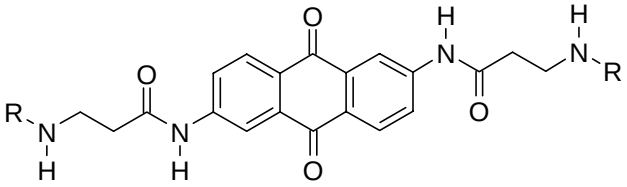
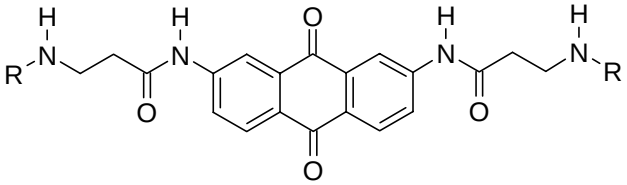
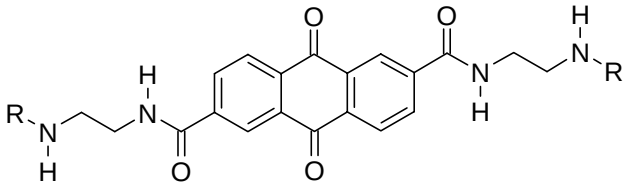
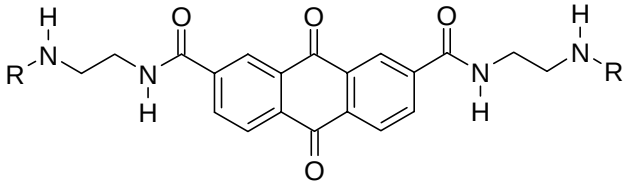
It was established that appropriately substituted amidoanthraquinone derivatives were effective G-quadruplex stabilizers thereby inhibiting telomerase enzyme. Further, effect of direction of amide bond was studied by Zagotto et al. for

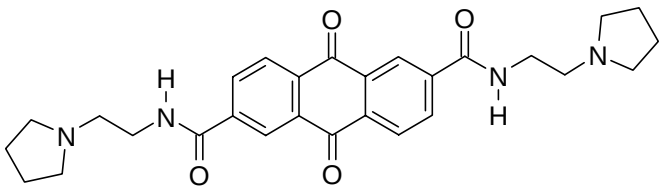


(**12**) AQ-NH-CO (IC_{50} 2.5 μM)

modulation of G-quadruplex recognition and telomerase inhibition [150]. Interestingly, it was observed that the AQ-NH-CO (**12**) performed better than AQ-CO-NH (**13**).

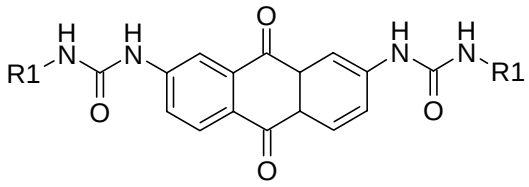
Table: 2.1 Telomerase inhibition data for disubstituted anthraquinone-amino acid conjugates

Compd.	Scaffold	General Structure	R	IC_{50} (μM)
7	2,6-AQ-Gly		-Lys-NH_2	0.8
			-Arg-NH_2	0.8
8	2,6-AQ-βAla		-Lys-NH_2	0.8
			-Arg-NH_2	0.8
9	2,7-AQ-βAla		-H	1.2
			-Lys-NH_2	1.5
10	2,6-AQ-ED		-Lys-NH_2	0.8
11	2,7-AQ-ED		-Arg-NH_2	1.2

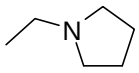


(**13**) AQ-CO-NH (IC_{50} 10.4 μM)

Another series of 2,7-bis-substituted amido-anthraquinone derivatives was synthesized and screened for human telomerase inhibitory activity. In the series, compounds (**14-23**) showed potent telomerase inhibitory activity [151].



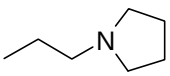
R1

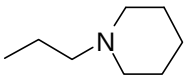
(**14**) = 

(**15**) = $CH_2NHCH_2CH_3$

(**16**) = $CH_2NHCH_2CH_3$

(**17**) = $CH_2CH_2N(CH_2CH_3)_2$

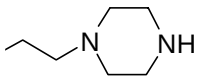
(**18**) = 

(**19**) = 

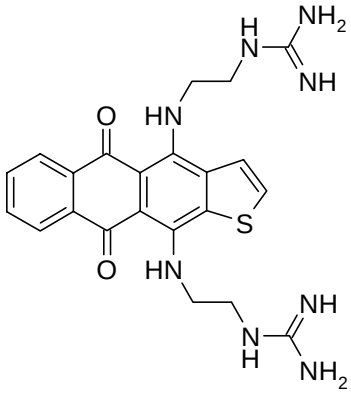
(**20**) = $CH_2CH_2N(CH_3)_2$

(**21**) = $CH_2CH_2NHCH_2CH_2CH_3$

(**22**) = $CH_2CH_2NHCH_2CH_3$

(**23**) = 

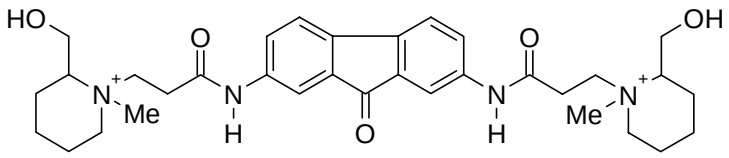
Shchekotikhin et al. reported a series of thiophene-fused tetracyclic analogs of the antitumor drug ametantrone as telomerase inhibitors. The guanidine derivative



(**24**)

(**24**) was found to be the most potent (IC_{50} 1 μM) compound among the whole series [152].

Perry et al. synthesized fluorenone series of compounds that exhibited a broad range of telomerase inhibitory activity in the range of 8-12 μM. Compound (**25**) (IC_{50} 8 μM) was found to be the most potent one out of all of the synthesized fluorenone derivatives [153]. However, fluorenone derivatives are comparatively less potent than

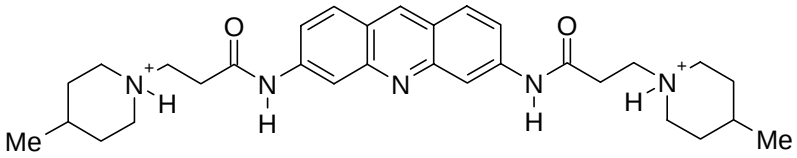


(**25**)

anthraquinone derivatives. This could be because of the change of six membered to five membered central ring of the aromatic core. This change reduces the electron deficiency of the chromophore which leads to weakening of the strength of π-stacking interactions and thus contributing to the reduced affinity of the fluorenone analogues for the G-quadruplex structure [153].

2.1.2 Acridine derivatives

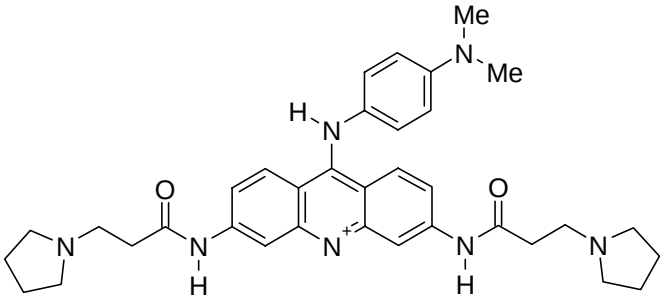
Harrison et al. has described a series of 3,6-disubstituted acridine derivatives in search of increased telomerase inhibitory activity by a rational approach. [154]. The molecular modeling study revealed that compounds containing a planar aromatic chromophore inhibits telomerase by stabilizing and binding to G-quadruplex structure [155]. It has been predicted that the acridine moiety has a planar chromophore with a



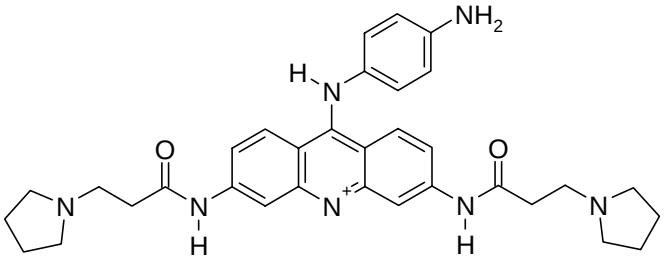
(**26**)

central heterocyclic nitrogen atom which got protonated at physiological *p*H. This would increase electron deficiency enabling enhanced G-quadruplex interactions. The 4-methyl piperidine derivative (**26**) (IC_{50} 1.35 μM) was found to be the most potent telomerase inhibitor from the series of the synthesized compounds.

Structure-based design approach to discover selective and potent small molecule telomerase inhibitors has been reported by Read et al. [156]. It was found that the 3,6,9-trisubstituted acridines had potent *in vitro* inhibitory activity against human telomerase compared with the earlier reported 3,6-disubstituted compounds. Compounds [(**27**) (BRACO19) and (**28**)] showed significantly higher telomerase inhibitory activity with IC_{50} values of 0.095 and 0.060 μM respectively. The anilino substituent at the 9th position of the acridine chromophore is predicted to lie in a third



(**27**) BRACO19



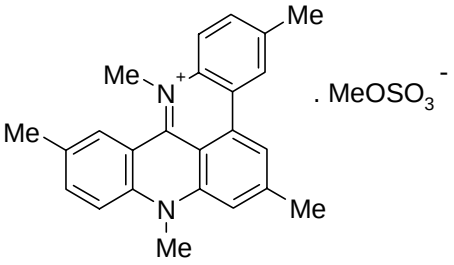
(**28**)

groove of the quadruplex leading to increased selectivity compared to 3,6-disubstituted compounds.

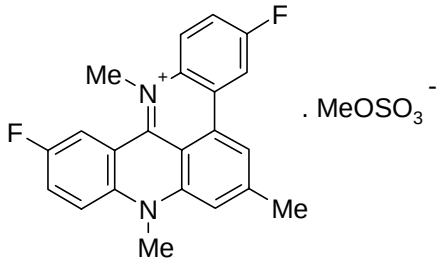
Heald et al. reported novel methylated pentacyclic quinoacridinium salts and evaluated them for telomerase inhibitory activity. Compounds (**29** and **30**) were found to be the most potent telomerase inhibitors with IC_{50} values of 0.33 μM and 0.25 μM respectively. Compound (**30**) has good pharmaceutical properties like it is soluble in water, stable in the *p*H range of 5-9 and efficiently transported into tumor cells [157].

Heald et al. further synthesized polycyclic acridines and tested them for their ability to inhibit telomerase activity. Among all of the synthesized compounds, 3-

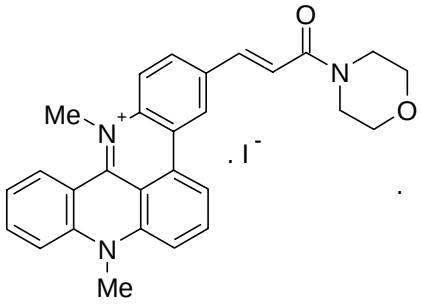
acryloylmorpholinyl polycyclic acridine (**31**) (IC_{50} 0.37 μM) was found to be a reasonably potent telomerase inhibitor [158].



(**29**) RHPS3

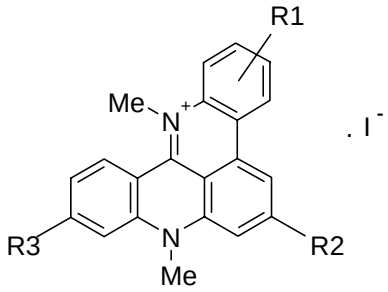


(**30**) RHSP4



(**31**)

A series of quaternized quino[4,3,2-*kl*]acridinium salts was synthesized by Cheng et al. The growth inhibitory activities of these compounds against tumor cell

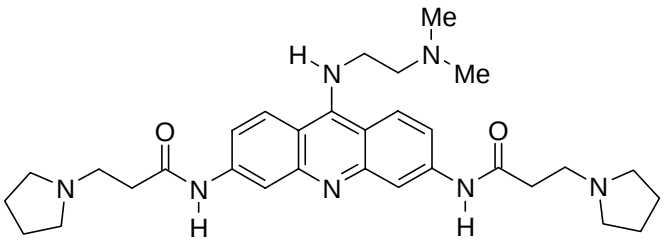


(**32**) R_1 = 2-NHAc; R_2 = H; R_3 = H

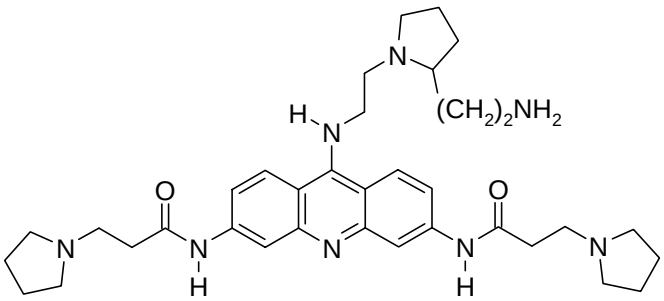
(**33**) R_1 = 3-NHAc; R_2 = H; R_3 = H

lines were measured. Selectivity for quadruplex DNA binding and stabilization by these compounds have also been explored [159]. It was noticed that two of the compounds (**32**) (IC_{50} 0.38 μM) and (**33**) (IC_{50} 0.41 μM) have potent telomerase inhibitory activity (IC_{50} < 0.5 μM) within the series of compounds.

The synthesis and evaluation of telomerase inhibitory activity of 2,6,9-, 2,7,9-, and 3,6,9- trisubstituted acridine derivatives have been reported by Harrison et al. Two compounds from 3,6,9-trisubstituted acridine series, compounds (**34** and **35**) were identified as most potent telomerase inhibitors with IC_{50} value of 0.018 μM [160].

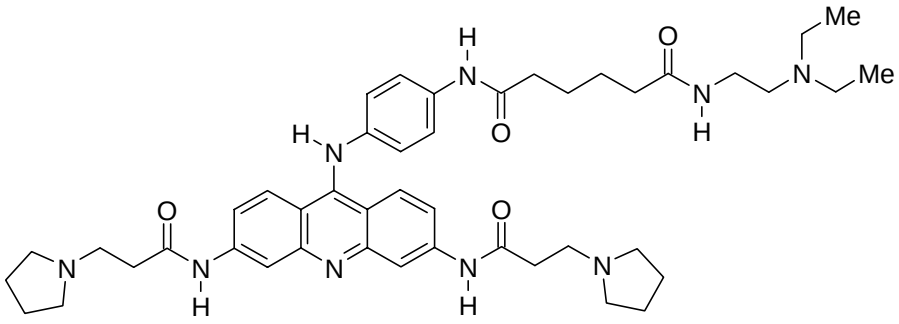


(**34**)

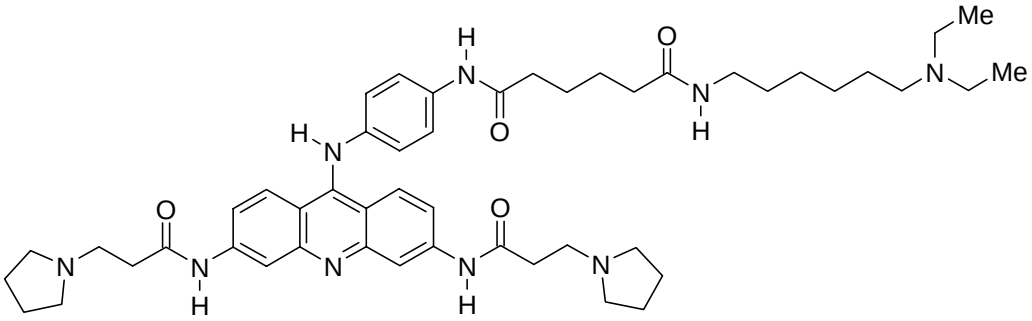


(**35**)

Another series of 3,6,9-trisubstituted acridine derivatives was synthesized and evaluated for telomerase inhibitory activity by Schultes et al. Within the series, two compounds with diethyl amine side chains (**36** and **37**) showed the best telomerase



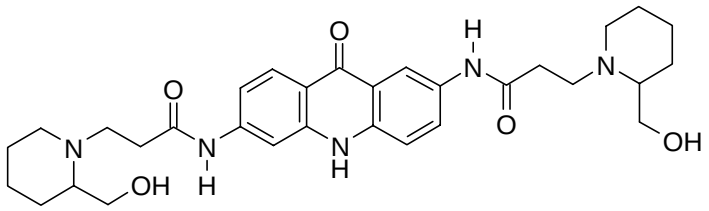
(**36**)



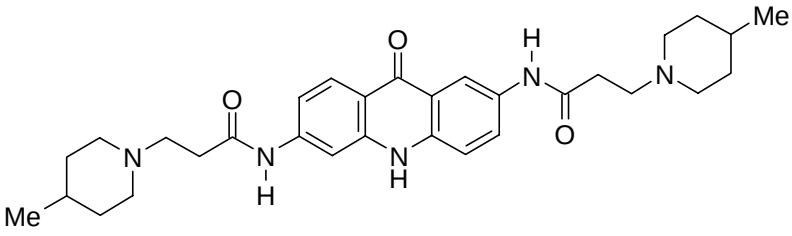
(**37**)

inhibition with IC_{50} values of 98 nM and 80 nM respectively [161].

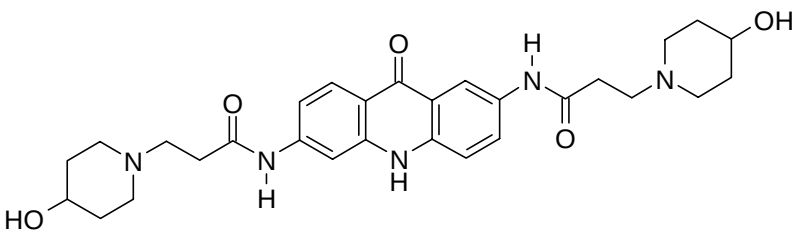
Harrison et al. reported synthesis and evaluation of 2,6-, 2,7- and 3,6-bis-aminoalkylamidoacridones, which showed similar level of *in vitro* activity against human telomerase compared to their acridine derivatives. The 2,6 derivatives (**38-40**) were found to be the most potent acridone derivatives within the series with IC_{50} values of 0.2 μM for each derivative [162].



(**38**)



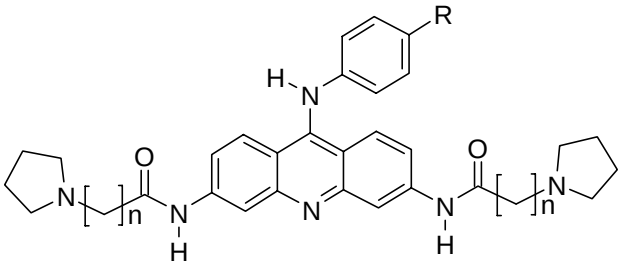
(**39**)

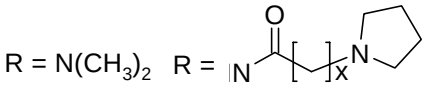


(**40**)

Effect of extensions of the 3,6- and 9- substituted side chains of trisubstituted acridines on telomerase activity was studied by Moore et al. Compound (**46**) was

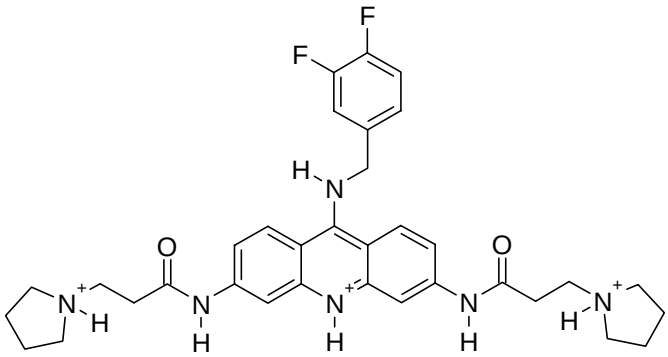
found to be the most potent telomerase inhibitor (IC_{50} 67.4 nM) among the series [163].





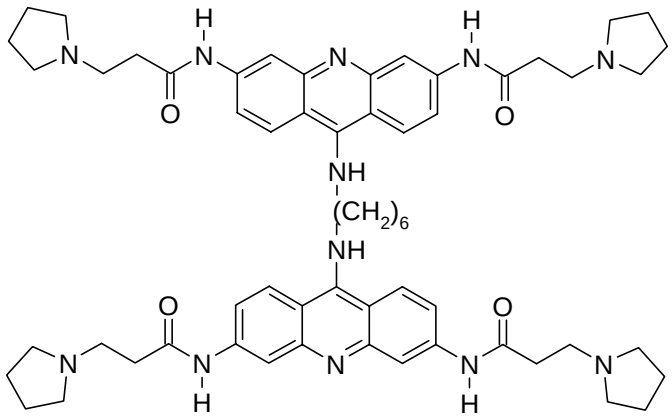
(**41**)	n = 2	(**45**)	n = 2 x = 1
(**42**)	n = 3	(**46**)	n = 2 x = 2
(**43**)	n = 4	(**47**)	n = 2 x = 3
(**44**)	n = 5	(**48**)	n = 3 x = 2
		(**49**)	n = 4 x = 2
		(**50**)	n = 5 x = 2

A new series of benzylaminoacridine compounds has been designed by structure-based drug design approach. It was observed that replacement of aniline substituent by benzylamino group resulted in increase in telomerase inhibitory activity only for one compound i.e compound (**51**) (IC_{50} 0.03 μM) [164].



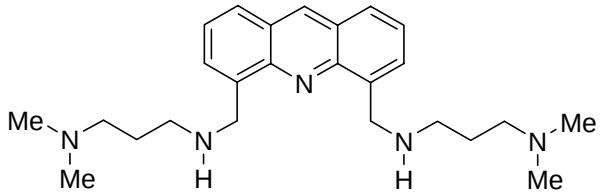
(**51**)

It was hypothesized that multiple interactions between the G-quadruplex and the binding ligands would be effective for inhibition of telomerase enzyme. To test this hypothesis, a small series of 3,6-disubstituted diacridine dimers, BAD compounds (BRACO19 Analog Dimer) was synthesized. It was observed that the BAD compounds showed human telomerase inhibition with IC_{50} values similar to or lower than BRACO19 (IC_{50} 5.4 μM). Compound BAD1 (**52**) (IC_{50} 4 μM) was found to be the most potent telomerase inhibitor [165].



(**52**) BAD1

A series of 4,5-bis(dialkylaminoalkyl)-substituted acridines have been synthesized and evaluated for telomerase inhibitory activity. Among all of the derivatives the dimethylaminopropyl substituted derivative (**53**) (IC_{50} 0.15 μM) was found to be the most active telomerase inhibitor [166].

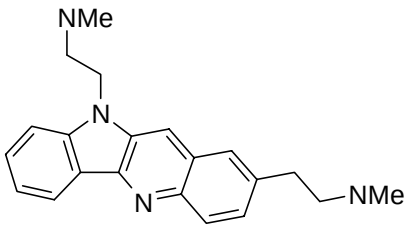


(**53**)

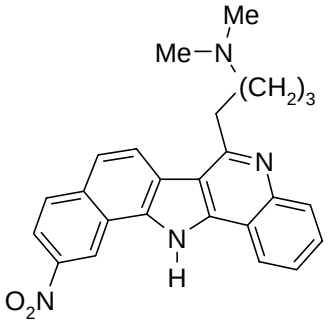
2.1.3 Quindoline derivatives

Caprio et al. have developed a novel telomerase inhibitor (**54**) (IC_{50} 16 μM) from 10*H*-Indolo[3,2-*b*]quinoline, an alkaloid obtained from the West African shrub *Cryptolepis sanguinolenta*. It might have been expected that tetracyclic chromophore will enhance stacking interactions with G-quadruplex and increase telomerase inhibitory activity. But, these compounds were found to be lesser active than the best quadruplex-mediated telomerase inhibitors based on tricyclic chromophores [145, 146, 153, 154, 167].

A small series of benzoindoloquionoline derivatives was synthesized and evaluated for telomerase inhibitory activity. A pentacyclic chromophore containing compound PSI99A (**55**) was found to be the most potent telomerase inhibitor with IC_{50} value of 0.5 μM [168].

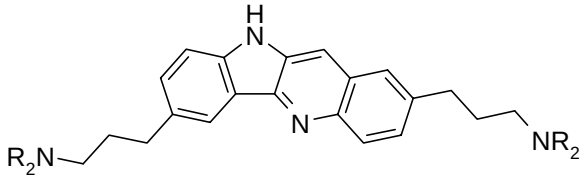


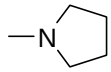
(**54**)

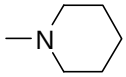


(**55**) PSI99A

Guyen et al. reported the synthesis and evaluation of telomerase inhibitory activity of rationally designed analogues of 10*H*-Indolo[3,2-*b*]quinoline. Compounds (**56** and **57**) exhibited IC_{50} value of 6.3 μM and 11.8 μM, respectively. It has been observed that these compounds are more potent than the previously reported 10*H*-Indolo[3,2-*b*]quinoline derivative (**54**) (IC_{50} 16 μM) [169]. This may be due to strong



(**56**) R_2 = 

(**57**) R_2 = 

interaction between the side chains of compounds and backbone phosphate oxygen atom as indicated by molecular modeling study. (Figure 2.1)

Zhou et al. developed a novel series of quindoline analogues as potent telomerase inhibitors. It was observed that compound SYUIQ-5 (**58**) was found to be the most potent telomerase inhibitor with IC_{50} value of 0.44 μM.

Structure-activity relationship studies revealed that telomerase inhibitory activity of these derivatives was significantly improved upon introduction of electron donating groups at C-11 position of quindoline. It was also observed that increase in potency of telomerase inhibitors was associated with increase of charge density at 5-*N* position [170].

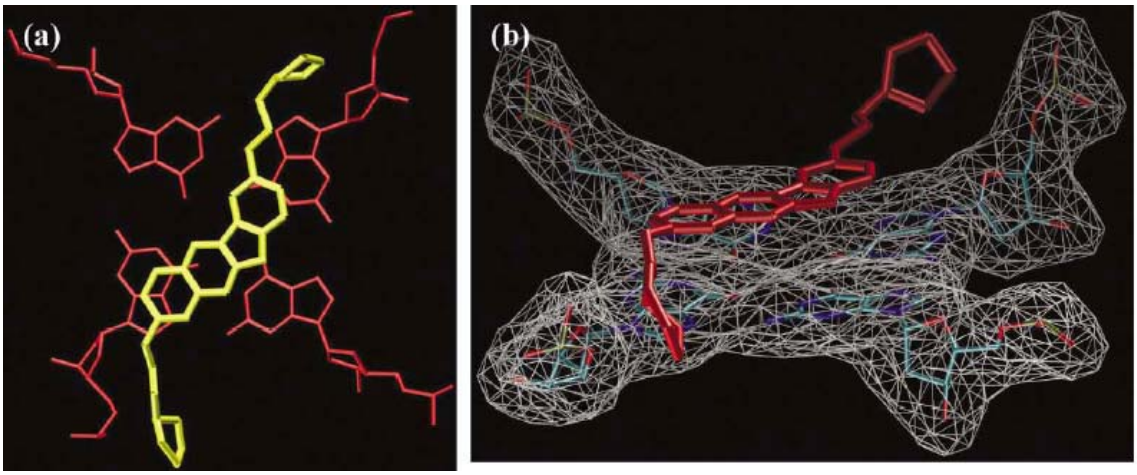
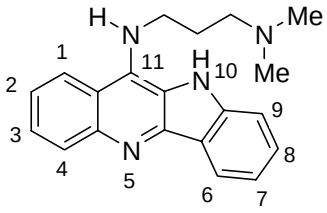
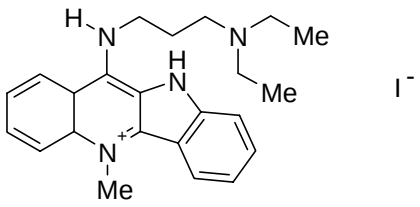


Figure 2.1: Docking pose of compound (**56**) showing stacking and side chain interaction with G-quadruplex [169].

A new series of quindoline derivatives as G-quadruplex stabilizing telomerase inhibitors was reported by Zhou et al. It was found that compound SYUIQ-5 (**58**) reproducibly inhibited telomerase activity in cancer cell lines, resulting in telomere shortening, followed by cellular senescence. Thus the authors proposed that SYUIQ-5 (**58**) was highly specific, novel and potent G-quadruplex-interactive agent and could be a promising agents for cancer treatment [171].



(**58**) SYUIQ-5

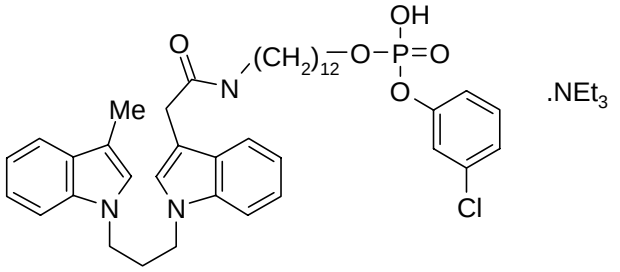


(**59**)

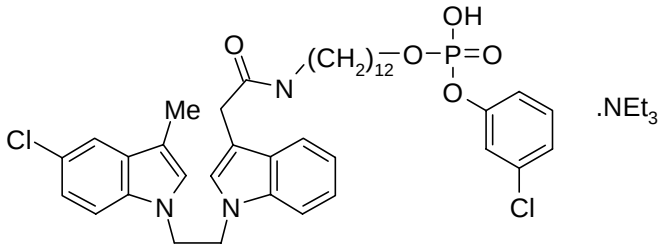
Lu et al. have described effect of 5*N*-methylated quindoline derivatives on quadruplex binding affinity and telomerase inhibition. A series of 5*N*-methylquindoline derivatives were designed. It was proposed that introduction of a positive charge favored electrostatic interactions and improved its inhibitory ability of telomerase enzyme. It was observed that the introduction of positive charge by methylation on the 5*N* position led to increase in π- stacking interaction due to the reduction of the electron density of the aromatic core of ligands. Out of all of the compounds reported in the series, compound (**59**) (0.16 μM) was found to be the most potent telomerase inhibitor [172].

2.1.4 Bisindole derivatives

Sasaki et al. have reported a novel bisindole scaffold as telomerase inhibitor. They have synthesized a series of bisindole derivatives and evaluated them for human telomerase inhibition using PCR assay. Compounds (**60**) (3.4 μM) and (**61**) (2.5 μM) comprising of two indole units, a long alkyl chain and phosphodiester group were found to be the most potent telomerase inhibitors [173]. It was observed that a hydrophobic group may be an essential factor for human telomerase inhibition.

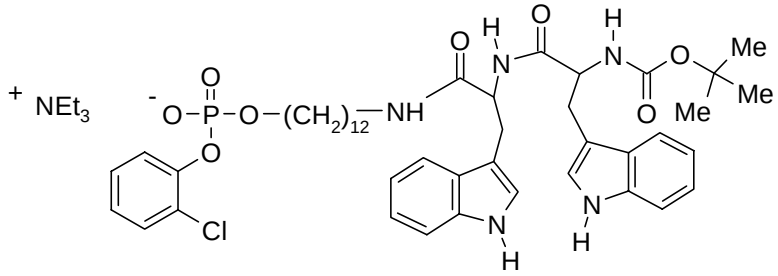


(**60**)



(**61**)

Further, the same group has reported solid-phase synthesis of a library of telomerase inhibitors with same structural features viz. aromatic phosphate, long alkyl

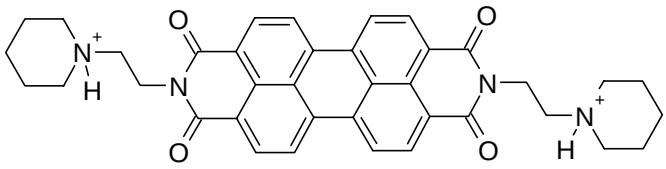


(**62**)

chain and tryptophan components. Out of these compounds, a *D,D*-ditryptophan derivative (**62**) has been identified as the most potent telomerase inhibitor with IC_{50} of 0.3 μM [174].

2.1.5 Perylene derivatives

The first potent human telomerase inhibitor based on perylene scafold was identified by Yu et al. with IC_{50} value in low μM range. [175]. The authors described G-quadruplex-binding studies by the telomerase-inhibiting compound *N,N'*-bis[2-(1-piperidino)-ethyl]-3,4,9,10-perylenetetracarboxilicdiimide designated as PIPER (**63**). It was observed that PIPER (**63**) was bound to G-quadruplex with stoichiometries of 1:2, 1:1 and 2:1 out of which 1:2 PIPER:DNA stoichiometry model was found to be a reasonable model which formed a sandwich-type complex. The π-π stacking interactions between two guanine tetrad and the aromatic chromophore of the drug play a crucial role in stabilizing the ligand-bis-quadruplex complex. Subsequently,

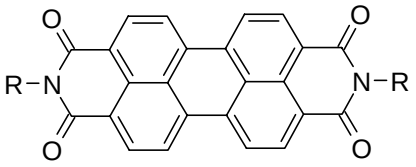


(**63**) PIPER

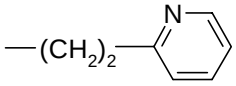
Hurley group stated that the formation of G-quadruplex structure by PIPER took place in a similar way as that of chaperone proteins [176].

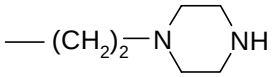
Kerwin et al. reported that G-quadruplex DNA binding selectivity of PIPER (**63**) is *p*H dependent. It was found that perylenediimide-G-quadruplex DNA binding selectivity is mediated by ligand aggregation [177].

The effect of different side chains of perylene derivatives have been studied for their ability in stabilization of G-quadruplex DNA structures. It was found that electrostatic interactions between ligands' side chains and G-quadruplex groove play major role in telomerase inhibition. Four perylene derivatives [PIPER (**63**), DAPER (**64**), PIPER2 (**65**) and PIPER3 (**66**)] were synthesized and tested for telomerase inhibition. Out of these derivatives, DAPER (**64**) and PIPER3 (**66**) were found to inhibit telomerase enzyme with IC_{50} values in the range of 10-20 μM [178].

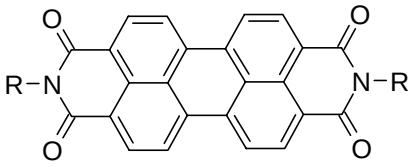


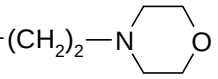
(**64**) DAPER, R = $-(CH_2)_3N(CH_3)_2$

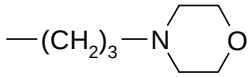
(**65**) PIPER2, R = 

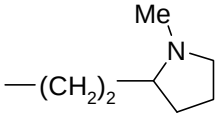
(**66**) PIPER3, R = 

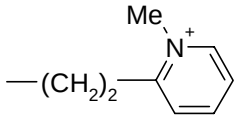
Rossetti et al. synthesized a small series of perylene derivatives to understand the role of different features of side chains in inter- and intra-molecular G-quadruplex structures' formation and in the inhibition of telomerase. PIPER3 (**66**), PIPER6 (**69**) and PIPER7 (**70**) (IC_{50} between 5-10 μM) were found to be more efficient telomerase inhibitors than PIPER (**63**), PIPER4 (**67**) and PIPER5 (**68**). It was suggested that too short side chains and/or a weak basicity of side chain amines lead to poor telomerase inhibitory activity [179].



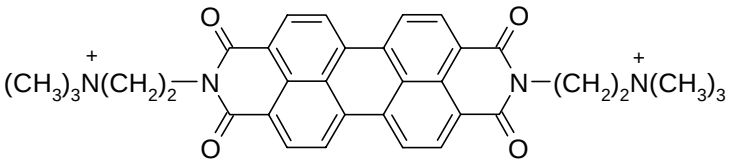
PIPER4 (**67**), R = 

PIPER5 (**68**), R = 

PIPER6 (**69**), R = 

PIPER7 (**70**), R = 

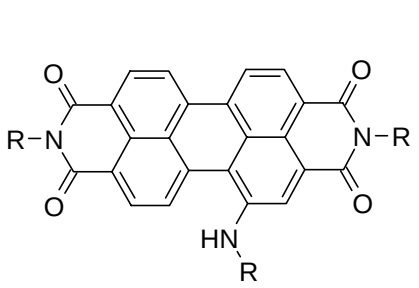
Sissi et al. synthesized a series of PIPER (**63**) related compounds as telomerase inhibitors by modifying condensed aromatic rings and introducing differ-



(**71**)

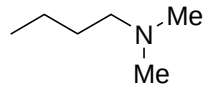
ent side chains which gets protonated at physiological *p*H. Among all of the tested perylene derivatives, heptacyclic compound (**71**) was found to be the most efficient telomerase inhibitor [180].

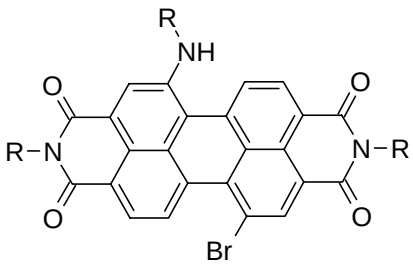
Franceschin et al. have reported synthesis of new perylene derivatives and evaluated them for their telomerase inhibitory potential. Compounds [(DAPER3C (**72**), DAPER3C-Br (**73**), DAPER4C(1,6) (**74**), and DAPER4C(1,7) (**75**)] inhibited 50% of telomerase enzyme at about 5 μM concentration. This could be due to greater number of positively charged side chains present on the perylene scaffold [181].



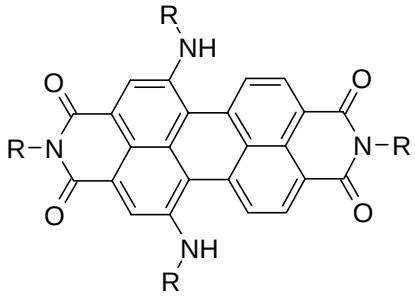
(**72**) DAPER3C

R =

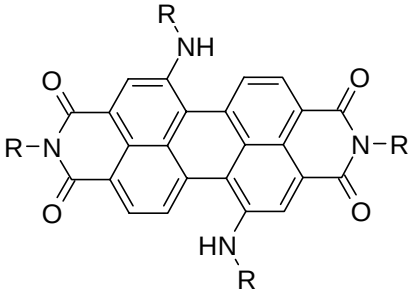




(**73**) DAPER3C-Br

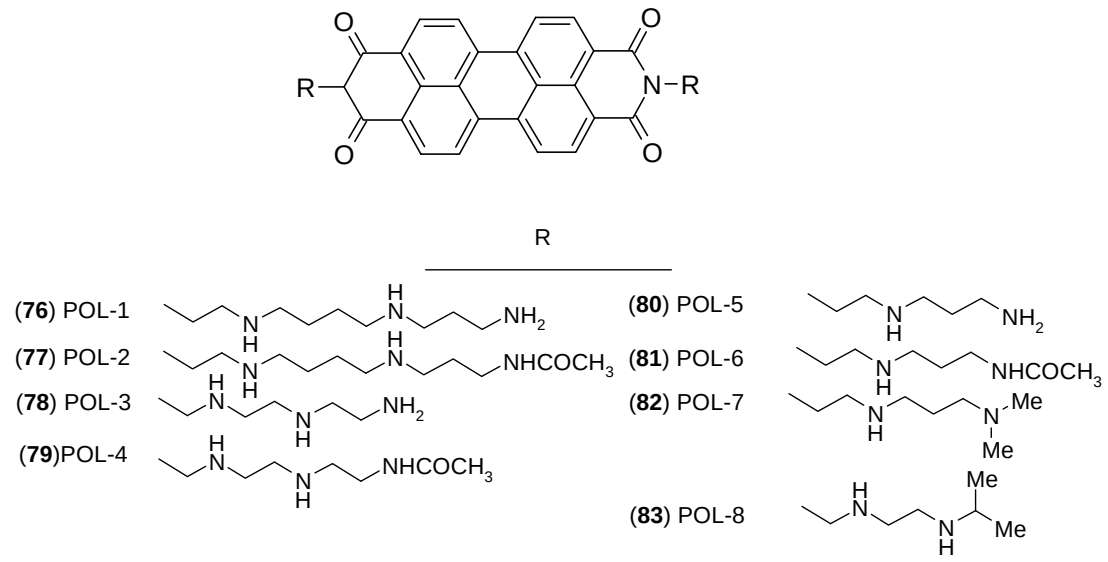


(**74**) DAPER4C(1,6)



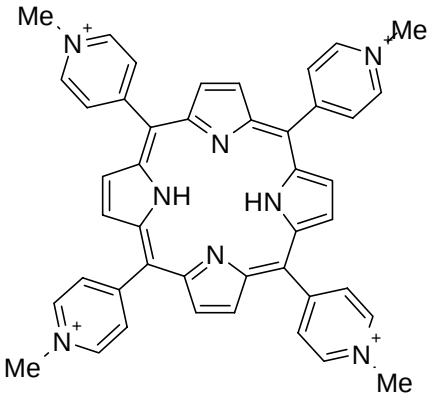
(**75**) DAPER4C(1,7)

Franceschin et al. have further synthesized eight polyamine perylene diimides. The aim of the study was to evaluate the role of the number and the distance of positive charges in the side chains on G-quadruplex stabilization and telomerase inhibition [182]. Of these polyamine perylenediimide derivatives compound POL-3 (**78**) and POL-5 (**80**) were found to inhibit telomerase enzyme in the concentration range of 7-10 μM. Further, it was reported that the selectivity of POL-3 (**78**) was regulated by side chain charge density [183].



2.1.6 Porphyrine derivatives

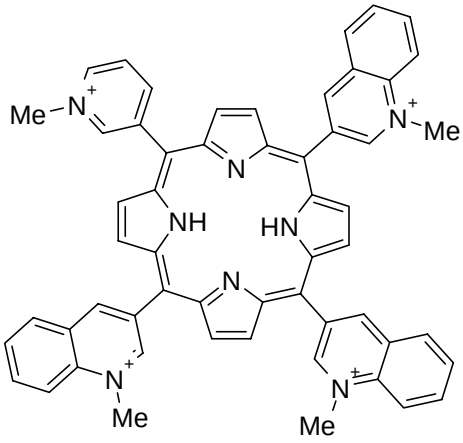
Very first report on cationic porphyhrines as telomerase inhibitors has been described by Hurley group. It was found that 5,10-15,20-tetra-(*N*-methyl-4-pyridyl)porphine TMPyP4 (**84**) stacked with G-tetrad that stabilized quadruplex DNA and effectively inhibited human telomerase with IC_{50} value of 6.5 ±1.4 μM [184].



(**84**) TMPyP4

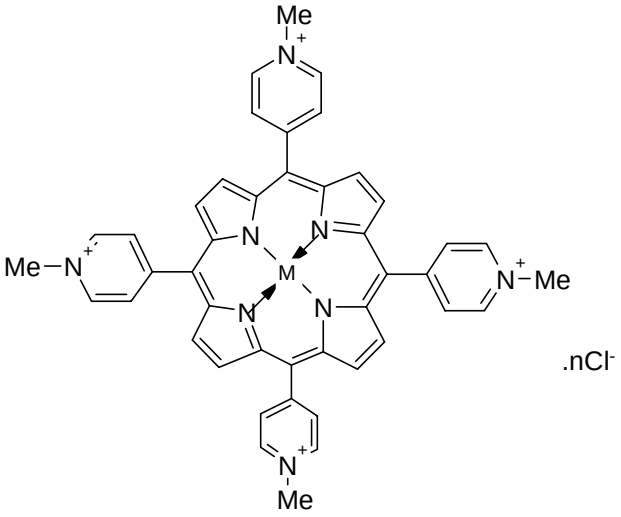
Further, the same group has reported wide range of analogues of TMPyP4 (**84**) as telomerase inhibitors and studied basic structure-activity relationship for the cationic porphyrins. Among all of the analogues of TMPyP4 (**84**), compound (**85**) (5 μM) was found to be the most potent telomerase inhibitor. Following structural insights have been reported for the effective telomerase inhibition: (1) The face of the porphyrin must be available for stacking, (2) the positive charged substituents are important but may be interchanged and combined with hydrogen-bonding groups, (3)

substitution is tolerated only on *meso* positions of the porphyrin and size of the substitution should be matched to the width of the grooves in which they lie [185].



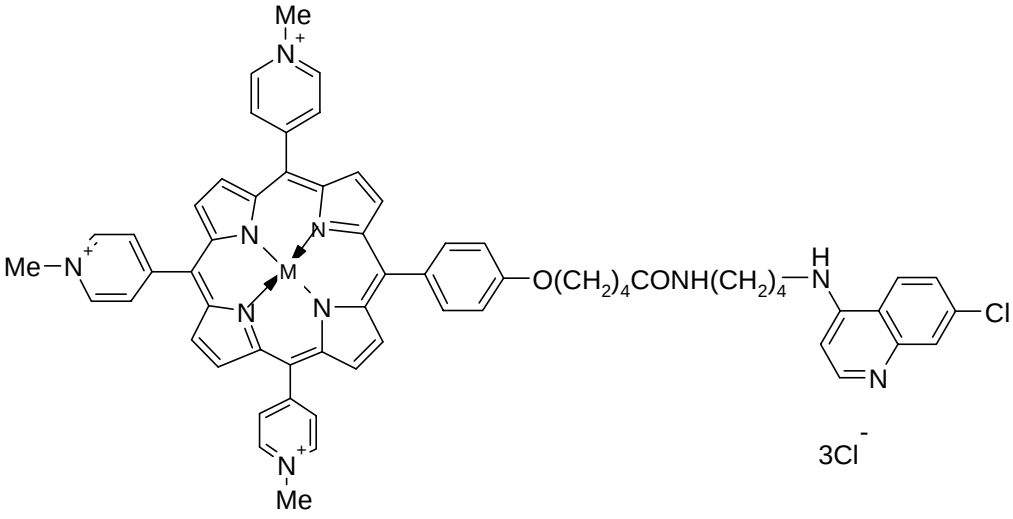
(**85**)

Maraval et al. synthesized a series of metalloporphyrins as G-quadruplex DNA targeted telomerase inhibitors. It was found that all of the porphyrin complexes

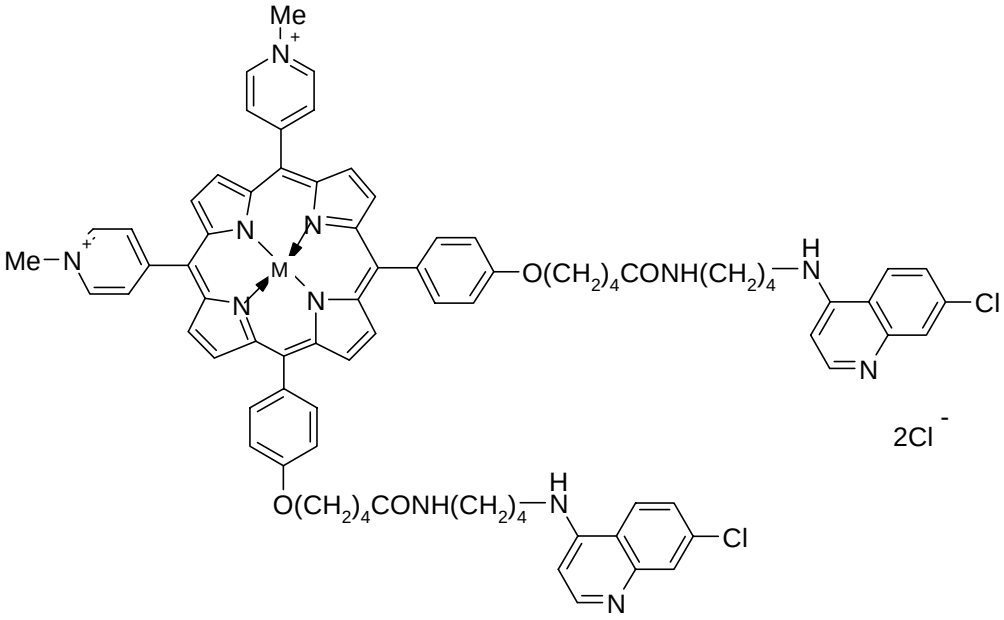


TMPyP4
(**86**) M = Mn^{+3}; n=5
(**87**) M = Ni^{+2}; n = 4

were capable of inhibiting the telomerase enzyme with IC_{50} values in micromolar ranges (Table 2.2) [186].



(**88**) M = Mn^{+3}
(**89**) M = Ni^{+2}



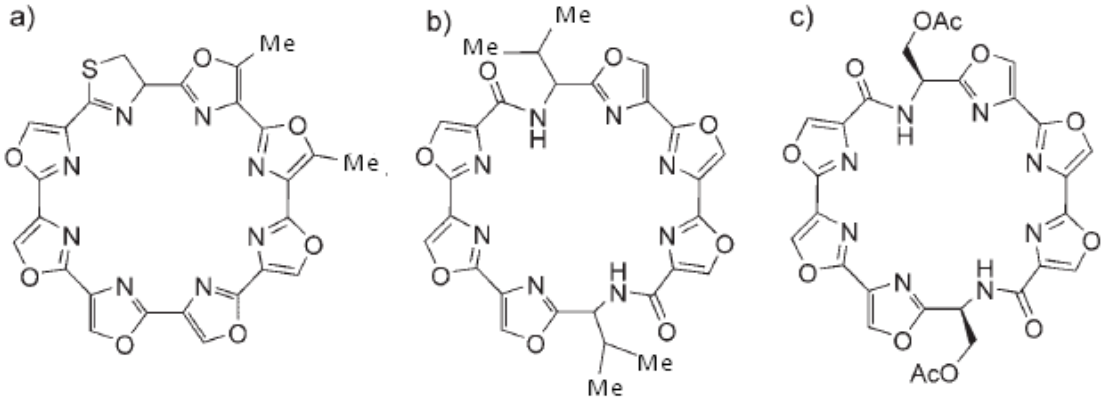
(**90**) M = Mn
(**91**) M = Ni

Table 2.2 Inhibition of telomerase by metalloporphyrin derivatives

Compound	Porphyrin derivative	IC_{50} (μM)
86	Mn-TMPyP	25.9
87	Ni-TMPyP	5.00
88	Mn-4a	11.5
89	Ni-4a	7.30
90	Mn-4b	8.6
91	Ni-4b	39

2.1.7 Macrocyclic compounds

Telomestatin (**92**), a macrocyclic natural product isolated from the actinomycete *Streptomyces anulatus* 3533- SV4 is currently the most efficient *in vitro* telomerase inhibitor with an IC_{50} value of 5 nM [187, 188]. Telomestatin consists of seven oxazole rings and one thiazoline ring, and is found to interact with G-quadruplex structures. This compound has 70-fold selectivity for intramolecular G-quadruplexes than duplex DNA [189].

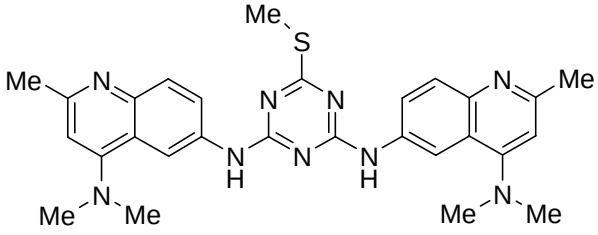


(**92**) Telomestatin (**93**) HXDV (**94**) Bistrioxazole acetate

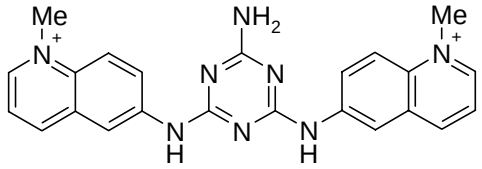
Macrocyclic hexaoxazole HXDV (**93**) and bistrioxazole acetate (**94**) are synthetic derivatives of telomestatin reported by Barbieri et al. and Tera et al. respectively [190, 191]. HXDV (**93**) has a strong selectivity for quadruplex over duplex or triplex DNA. Bistrioxazole acetate (**94**), which has a macrocyclic bisamide structure, showed telomerase inhibitory activity with an IC_{50} value of 2 μM.

2.1.8 Triazine derivatives

A series of triazine compounds as G-quadruplex stabilizing telomerase inhibitors have been reported by Riou et al. [192]. Among all of the triazine derivative



(**95**) 115405

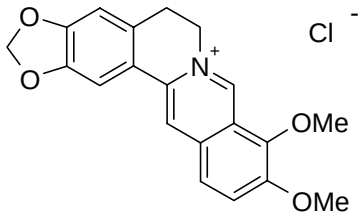


(**96**) 12459

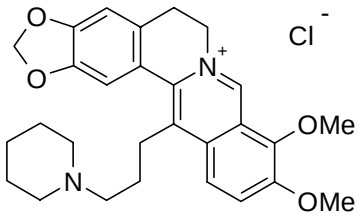
bisquinoline-substituted triazines, compound 115405 (**95**) (IC_{50} 0.041 μM) and 12459 (**96**) (IC_{50} 0.13 μM), were found to be active against human telomerase at nanomolar concentrations.

2.1.9 Berberine derivatives

Berberine is chemically an alkaloid isolated from Chinese herb known for its antimicrobial activity. Berberine (**97**) as telomerase inhibitor was first reported in 1999 [193]. Subsequently, a number of 13-substituted berberine derivatives have been identified as inhibitors of telomerase activity by stabilizing G-quadruplex DNA. Compound (**98**) was found to be a potent compound out of all of the 13-substituted berberine derivatives [194].

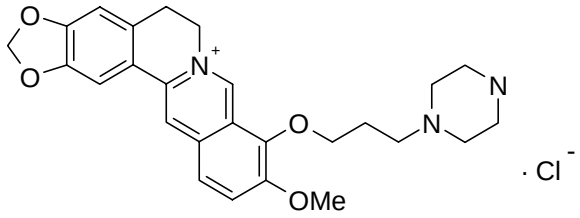


(**97**) Berberine



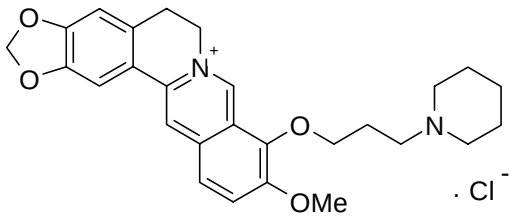
(**98**)

Zhang et al. have studied berberine (**97**) and 9-substituted berberine derivatives for their telomerase inhibitory activity [195]. Among the synthesized berberine derivatives, compounds (**99** and **100)** were found to be the most potent



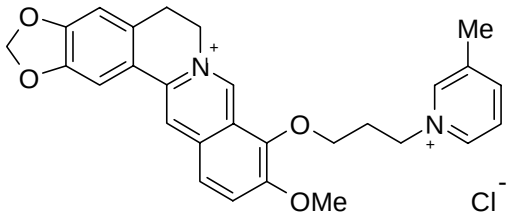
(**99**)

telomerase inhibitors with IC_{50} values 7.5 μmol/L and 12.5 μmol/L respectively. It

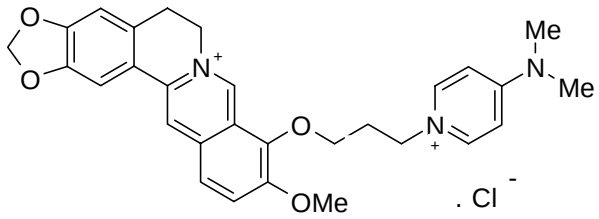


(**100**)

was observed that the 9-substituted derivatives showed stronger binding affinity for G-quadruplexes and higher telomerase inhibitory activity than that of berberine.



(**101**)



(**102**)

A new series of 9-substituted berberine derivatives has been reported as telomeric quadruplex ligands by Ma et al. Compounds (**101** and **102)** were found to be equipotent having IC_{50} values of 14 μM. Biophysical and biochemical assay of the synthesized berberine derivatives indicated that introduction of positively charged aza-aromatic terminal group into the side chain of position-9 of berberine significantly improved their binding ability with G-quadruplex and exhibited the inhibitory effect on telomerase activity [196].

2.2 Molecular modeling studies of G-quadruplex stabilizing telomerase inhibitors

There are currently 124 entries in the Protein data bank PDB for three-dimensional quadruplex structures. Majority of these have been resolved by solution NMR methods. The resolution of the structures varies between 1.10 and 3.20 Å. Views of ligand-quadruplex crystal structures are shown in Figure 2.2.

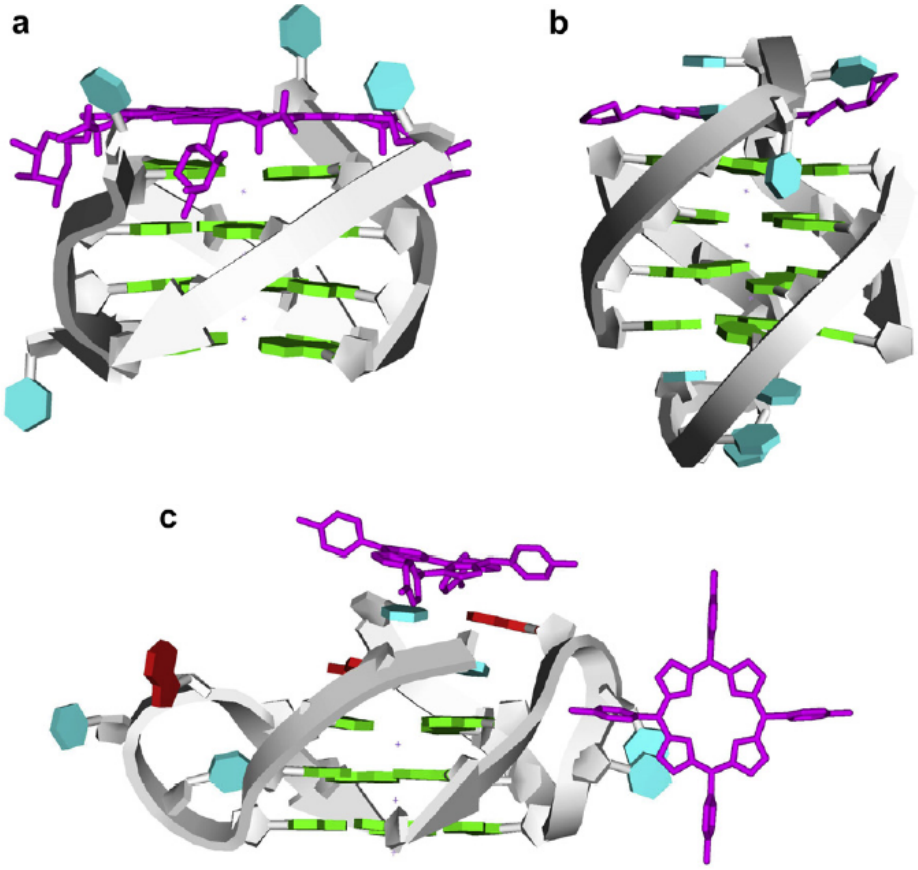


Figure 2.2: Views of three ligand-quadruplex crystal structures. (a) The daunomycin-d(TG4T) complex, PDB Id. 1O0K. (b) The acridine-d(G4T4G4) complex, PDB Id. 1L1H. (c) The porphyrin TMPyP4- d(TAG3T2AG3) complex, PDB Id. 2HRI [197].

Initial molecular modeling studies were carried out by Read et al. to test the hypothesis that compounds possessing a planar aromatic chromophore inhibited telomerase via stabilization of, and binding to, a guanine quadruplex structure (Figure 2.3). 2,6-Disubstituted amidoanthracene-9,10-dione and 3,6-disubstituted acridine derivatives were used in testing this hypothesis. The relative binding energies between these compounds and the human telomere DNA quadruplex were determined using

molecular simulation methods. It was observed that the results obtained were in excellent agreement with the biological activity as measured *in vitro* using a modified telomeric repeat amplification protocol assay (TRAP). Hence, it was proposed that

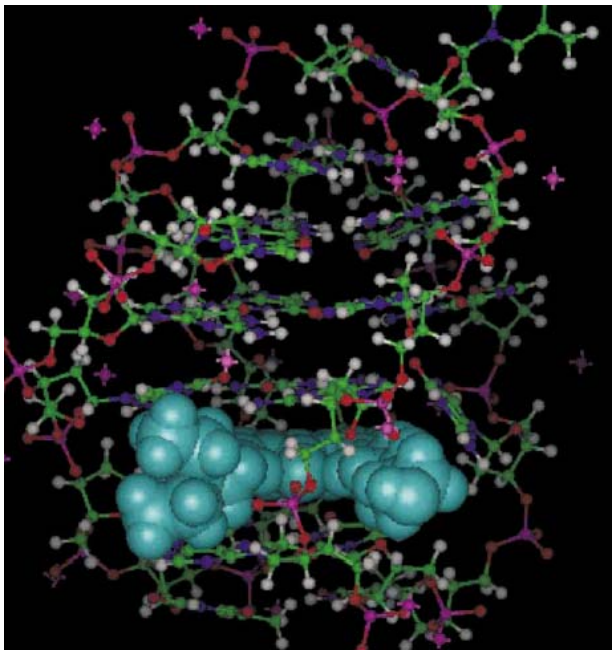


Figure 2.3: Structure of 1,8-disubstituted acridine derivative intercalated into G-quadruplex.

this experimental evidence provides strong support for the hypothesis that G-quadruplexes were the primary targets for telomerase inhibitors with extended planar chromophores [155].

Neidle and co-workers have reported first structure-based design of selective and potent G-quadruplex-mediated telomerase inhibitors. It was proposed that 3,6,9-trisubstituted acridine inhibitors interacted selectively with the human DNA quadruplex structure, as a means of specifically inhibiting the action of human telomerase enzyme. The results showed that the anilino substituent at the 9-position of the acridine chromophore was predicted to lie in a third groove of the quadruplex (Figure 2.4). These acridine derivatives have potent *in vitro* inhibitory activity against human telomerase, with EC_{50} values of up to 60 nM [156].

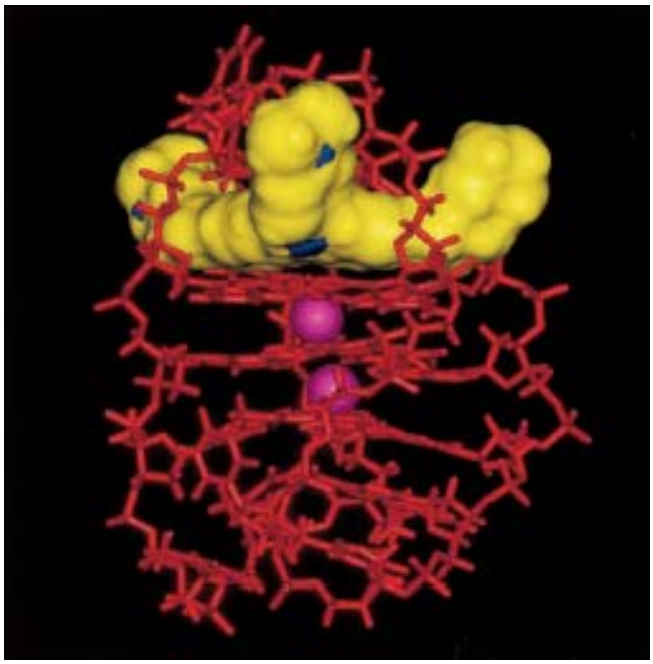
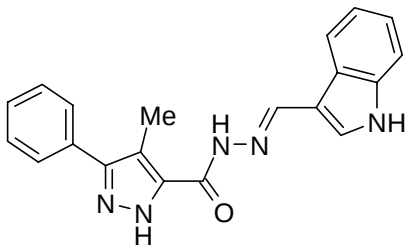


Figure 2.4: Solvent-accessible surface area of 3,6,9-trisubstituted acridine derivative

Earlier, the application of virtual screening methods against nucleic acids as target was unexplored due to the lack of structural studies and understanding of the biological role of G-quadruplex DNA. With the advent of structural information by crystallographic and NMR resolved G-quadruplex structures, molecular modeling and virtual screening of chemical libraries against this target offered the possibility of identifying novel chemotherapeutics possessing high selectivity.

Ma et al. have reported the first example of virtual screening to identify G-quadruplex-binding ligands in 2008 [198]. In this work virtual ligand screening was used in two step process to identify selective G-quadruplex ligands from a library of drug-like structures. PDB 1KF1 [199] was used as a starting model for their virtual screening campaign. In the first step, Lipinski filters were applied to a library of 1,00,000 drug-like compounds to remove compounds that would not be predicted to display ideal pharmacokinetic properties. The remaining compounds were docked into a grid receptor and ranked according to scores of each ligand-quadruplex complex in ICM method (Molsoft) [200]. From the library of drug-like compounds, 1*H*-pyrazole-3-carboxy-4-methyl-5-phenyl-(1*H*-indol-3-ylmethylene)hydrazide (**103**) was identified as a top candidate. Additional molecular modeling (Figure 2.5) studies sho-



(**103**)

wed that the compound was bound at the 3′ terminus of the G-quadruplex with extensive interactions between the aromatic phenyl and indole scaffolds of (**103**) and the terminal G-quartet.

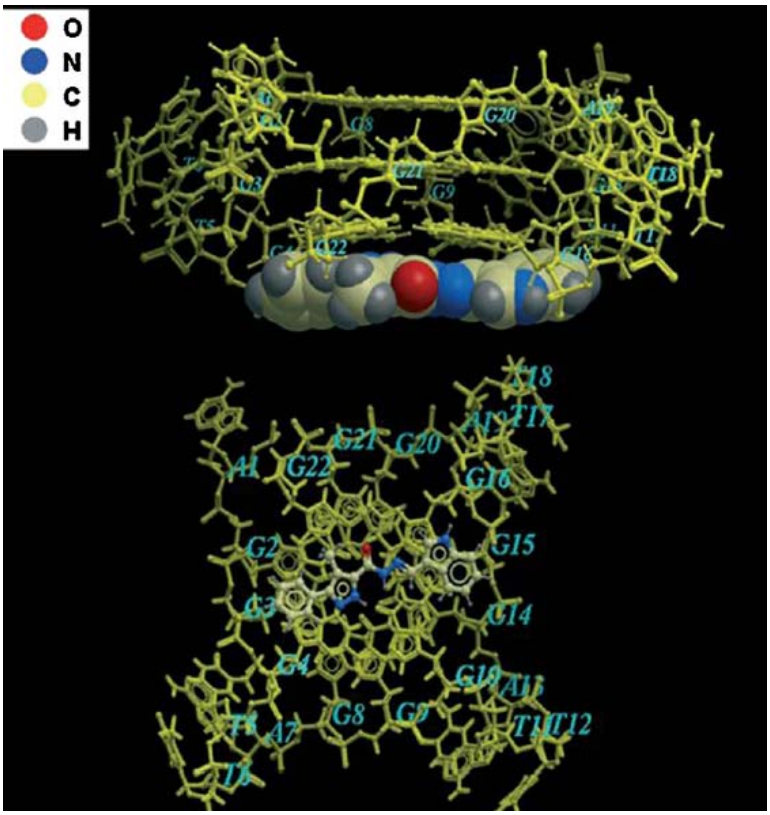
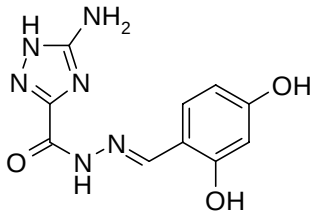


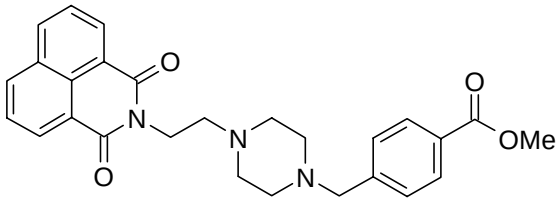
Figure 2.5: Molecular models showing the interaction of compound (**103**) with the intramolecular human telomeric G-quadruplex [198].

Cosconati et al. have recently reported virtual ligand screening to identify G-quadruplex groove binders [201]. The author used AutoDOCK and a crystal structure of the G-quadruplex d(TGGGT)4 (PDB:1S45 [202]); a library of 6000 structures was screened to identify new hits as G-quadruplex groove-binders. The structures were ranked according to their predicted free binding energies, out of which 137

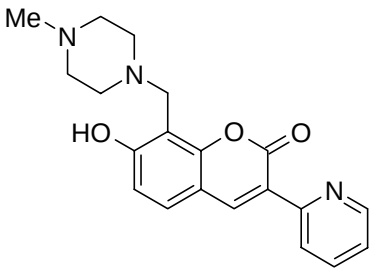
compounds possessed predicted binding energies lower than the cut-off energy of -6.0 kcal/mol. In a second round of selection, compounds were screened by visual inspection of the predicted poses and their interactions with G-quadruplex and found 30 compounds. After experimental study, only six compounds (**104-109**) were confirmed as G-quadruplex groove binders.



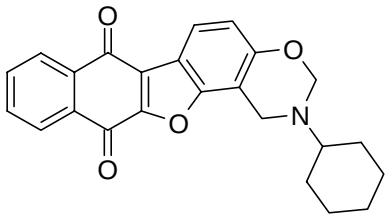
(**104**)



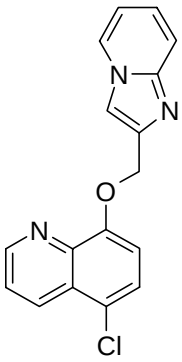
(**105**)



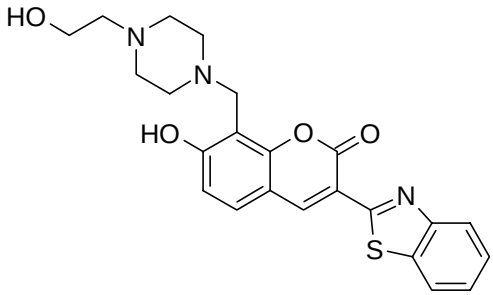
(**106**)



(**107**)



(**108**)



(**109**)

Ma and co-workers have conducted virtual ligand screening of a natural products database in order to identify novel natural product scaffolds as selective G-quadruplex stabilizing agents [203]. The study was started with in house molecular modeling of the c-myc G-quadruplex based on the X-ray crystal structure of the human telomeric G-quadruplex. A library of 20,000 natural product and natural

product-like structures were screened *in silico* and four hits were identified as selective G-quadruplex stabilizers.

The detailed highlights of some successful applications of molecular modeling in virtual ligand screening and structure-based designing of organic and inorganic molecules that target G-quadruplex have been reviewed recently [204].

Molecular modeling and molecular dynamics simulations studies have been reported by Li et al. to investigate interactions between dimeric G-quadruplex structure and a planar ligand of a perylene derivative. The results show that perylene derivative can bind to the dimer efficiently through end stacking, and the binding mode of the ligand stacked with the 3′-terminal thymine base is most favorable (Figure 2.6) [205].

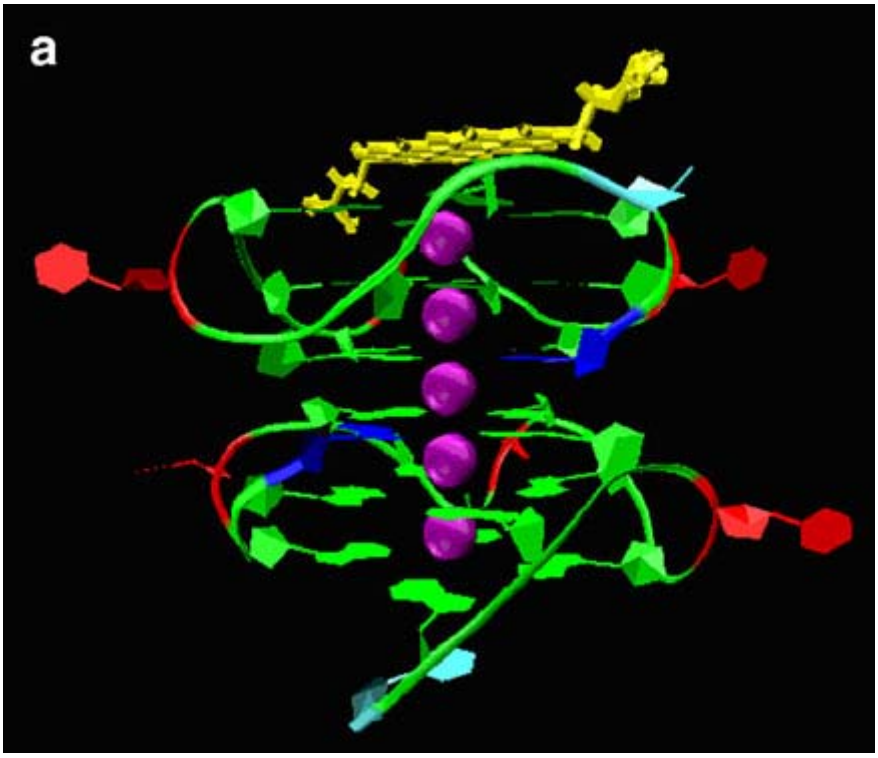


Figure 2.6: Stacking view of perylene derivative with dimeric G-quadruplex structure [205].

Aixiao et al. presented molecular dynamics (MD) simulations of four-stranded G-quadruplex DNA complexes formed by the sequence $d(TTAGGG)_4$ and the TQMP ligand to investigate the best possible binding mode. It was proposed that in gas phase, the interaction between TQMP and G-quadruplex DNA is mainly electrostatic.

But in water medium, the TQMP ligand preferentially binds in the diagonal loop position [206].

Yang et al. described a simulated annealing-based docking and molecular dynamics simulation to define the stable binding mode of 3,6-bis(1-methyl-4-vinyl-

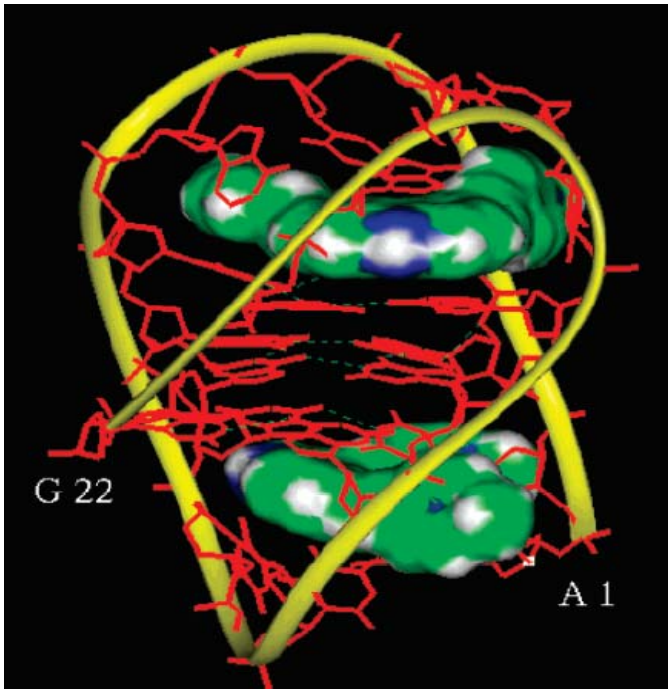
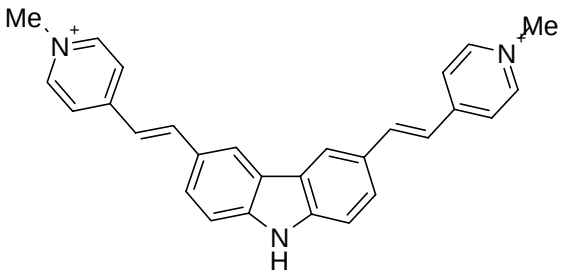


Figure 2.7: Structure of the 2:1 model where the two BMVCs (**110**) are bound with one G quadruplex in the end-stacked model.



BMVC (**110**)

pyridinium iodide) carbazole (BMVC), which stabilized the quadruplex structure of the human telomeric DNA and inhibited telomerase activity. It was found that the compound selectively interacted with G-quadruplex structure. Molecular modeling shows that a 2:1 binding model (Figure 2.7) involving the external binding of BMVC to both ends of the G-quartet is the most stable binding mode.

Neidle group recently reported X-ray crystallographic study to understand structural basis of DNA quadruplex recognition by an acridine derivative and highlighted some selectivity related issues for G-quadruplex binders [207, 208].

Chen et al. have reported pharmacophore-based discovery of triaryl-substituted imidazoles as new telomeric G-quadruplex ligand. Triaryl-substituted imidazole was confirmed to be a new telomeric G-quadruplex ligand with potent binding and stabilizing activity to G-quadruplex DNA, as well as a 8.7-fold selectivity towards telomeric G-quadruplex DNA over duplex DNA [209].

Molecular modeling studies carried out by Alcaro et al. with experimental DNA models with the quadruplex sequence suggested that the introduction of a net positive charge onto the side chain of a series of fluorenonecarboxamides can improve G-quadruplex binding [210].

3. AIMS AND OBJECTIVES

Cancer is one of the most feared of all the diseases and has high impact on society in terms of suffering and early mortality. According to a recent report by Merrill et al. [211] 1,040,381 male and 1,011,355 female incidents of cancer cases have been diagnosed during 2000 to 2007.

Although significant advances in early diagnosis and treatment of cancer have been made in the past decade, there remains a compelling need for more effective therapies for the treatment of many different types of cancers. The identification of more specific and thereby less toxic therapies for these tumors will provide significant advantages for cancer control. Earlier, screening and testing of large number of chemical and natural substances for potential therapeutic activity was random and time consuming because drug targets were unknown in most of the cases. The success of human genome project has provided a whole range of new opportunities for the discovery of drugs in terms of newly identified drug targets. The number of putative drug targets has now increased considerably.

For a medicinal product to reach patients, it commonly takes about 14-15 years and $ 800 million to end up the long tedious drug development process. Furthermore, only few candidates finish the clinical studies successfully and pass the strict inspection of drug regulatory agency; despite that thousands of new therapeutic candidates are being investigated in laboratories every year. However, the recent advances in technologies, like computational chemistry and Computer-aided drug design are now offering a fast track to deal with the problems associated with traditional drug discovery techniques. Computer-aided drug design offers an *in silico* alternative to medicinal chemistry techniques for studying the structure and predicting the biological activity of drug candidates. It has the advantages of both speed & low cost, and is becoming an indispensable tool for major pharmaceutical companies.

Over the past several years, tremendous efforts have been made to understand the role of telomerase in the development and progression of human cancers. Telomerase activity is nearly universal in human cancer cell lines, and is found in about 85–90% of primary tumors. Telomerase prevent cells from undergoing replicative senescence and contributes to their unlimited proliferation, telomerase can protect cells from DNA damage and subsequent apoptosis and telomerase activity

may correlate with the aggressive behavior (e.g migration and invasion) of tumor cells. Based on these findings, telomerase has been proposed as a potential target not only for cancer diagnosis but also for anticancer therapies.

Targeting G-quadruplex has become valid and proven approach for inhibiting telomerase enzyme. Many leading compounds that target these structures have been reported, and a few of them have entered preclinical or clinical trials. However, the selectivity of G-quadruplex stabilizing compounds has yet to be improved in order to discover more acceptable and potent telomerase inhibitors.

With a better understanding of the structural properties of G-quadruplexes and information obtained from previously known inhibitors, it is expected that a wealth of new potent and selective G-quadruplex stabilizing telomerase inhibitors could be identified. Identification of minimum requirements of structural features and pharmacophoric groups for G-quadruplex stabilizing telomerase inhibitors were hampered due to lack of molecular modeling studies. Hence it was planned to rationalize the structural requirements for G-quadruplex ligands based on 3D-QSAR (CoMFA and CoMSIA) and pharmacophore modeling which could be used in identification and optimization of newer telomerase inhibitors.

Considering the above it was planned to undertake the following studies:

3.1 Determining structural requirements of acridine derivatives as G-quadruplex mediating telomerase inhibitors by 3D-QSAR CoMFA and CoMSIA modeling.

3.2 Development of predictive 3D-QSAR CoMSIA models for anthraquinone and acridone derivatives as telomerase inhibitors targeting G-quadruplex DNA telomere

3.3 Molecular modeling studies of some triazine derivatives as G-quadruplex binders for inhibition of telomerase enzyme

3.4 Efforts towards developing a universal 3D-QSAR model for G-quadruplex mediating telomerase inhibitors.

4. RESULTS AND DISCUSSION

The results obtained for various studies have been discussed under the following heads:

4.1 Determining structural requirements of acridine derivatives as G-quadruplex mediating telomerase inhibitors by 3D-QSAR CoMFA and CoMSIA modeling.

4.2 Development of predictive 3D-QSAR CoMSIA models for anthraquinone and acridone derivatives as telomerase inhibitors targeting G-quadruplex DNA telomere

4.3 Molecular modeling studies of some triazine derivatives as G-quadruplex binders for inhibition of telomerase enzyme

4.4 Efforts towards developing a universal 3D-QSAR model for G-quadruplex mediating telomerase inhibitors.

4.1 Determining structural requirements of acridine derivatives as G-quadruplex mediating telomerase inhibitors by 3D-QSAR CoMFA and CoMSIA modeling.

In the current study it was planned to identify the structural requirements in trisubstituted acridines for their G-quadraplex stabilizing telomerase inhibiting activity using 3D-QSAR techniques. CoMFA and CoMSIA, being the most widely used ones, were used as 3D-QSAR techniques for the chosen series of compounds.

Although the aim of the study was simply to find structural requirements for selective G-quadruplex binding telomerase inhibitors having trisubstituted acridines as a basic scaffold and develop a 3D-model for the purpose of optimization of the leads, certain interesting observations were made during the study concerning the structures of these molecules and their interactions with the G-quadruplex of the enzyme telomerase.

4.1.1 Data set used

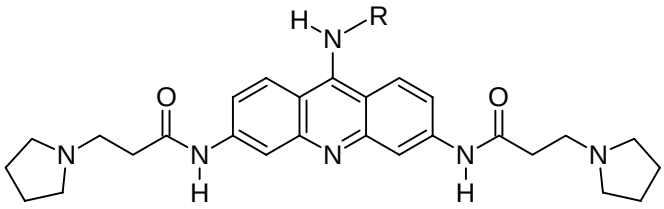
In the present work, a total of 79 compounds with their telomerase inhibitory activity from the reported work [154, 155, 157, 160, 161, 163, 164] were selected. The reported compounds showed wide variations in their structures and potency

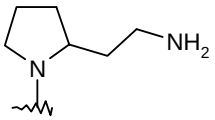
profile. Considering a high deviation in the biological activity and structural variations among the compounds of the series it was considered as an ideal series for performing QSAR analysis. Biological data with negative logarithm of minimum inhibitory concentration (MIC) expressed in moles/liter was used as a dependent variable in the 3D-QSAR study, thus correlating the data linearly to the free energy change.

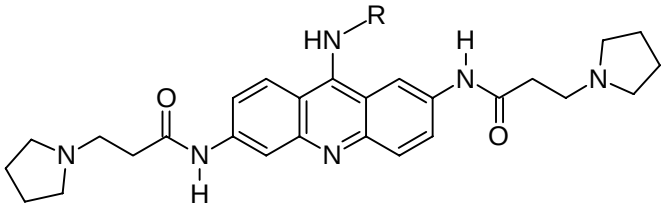
4.1.2 Selection of training and test Sets

In view of the finding that q^2 appears to be a necessary but not a sufficient condition for a model to have high predictive power [46] an emphasis has been given in the present study for validation of the developed model using an external test set. The whole set of 79 compounds was divided into training set (61) and test set (18) compounds. In the training set, most potent, moderately active and lowly active compounds were included to spread the activity range. The test set compounds were selected in such a manner that at least one structural analog of the training set was chosen for the test set.

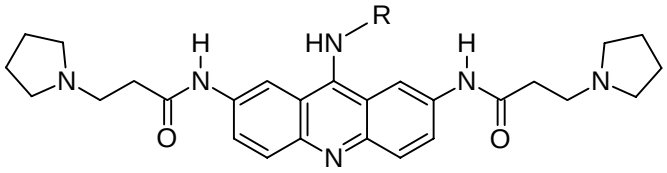
Table 4.1.1 Molecular structures and telomerase inhibitory activity for substituted acridine derivatives



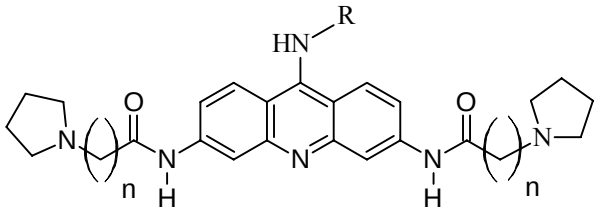
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
1a	-$C_6H_4NMe_2$ (*p*)	0.115	6.94	6.425
2	-$C_6H_4NH_2$ (*p*)	0.074	7.13	7.233
3	-$CH_2CH_2CH_2NHMe_2$	0.06	7.22	7.486
4	-$CH_2CH_2C_5H_{10}N$ (*c*)	0.05	7.30	7.250
5	-$C_6H_4NH_2$ (*m*)	0.06	7.22	7.185
6	-$C_6H_4NH_2$ (*o*)	0.02	7.70	7.309
7	-$C_6H_4NMe_2$ (*m*)	0.1	7.00	7.044
8	-C_6H_{11} (*c*)	0.09	7.05	7.007
9	-CH_2CH_2OMe	0.14	6.85	7.058
10	-C_7H_{13} (*c*)	0.21	6.68	6.939
11	-$C_6H_4COCH_3$ (*p*)	0.04	7.40	7.112
12	-$CH_2CH_2NHMe_2$	0.018	7.74	7.515
13a		0.018	7.74	7.125
14	-$CH_2C_5H_4N$ (*m*) (*c*)	0.066	7.18	7.295
15	-$C_6H_4NHCOCH_3$ (*m*)	0.1	7.00	7.249
16	-C_3H_5 (*c*)	0.05	7.30	7.055
17	-C_6H_4F (*p*)	0.07	7.15	6.989
18a	-C_6H_4SMe (*o*)	0.15	6.82	6.996
19	-C_6H_4SMe (*m*)	0.1	7.00	6.996

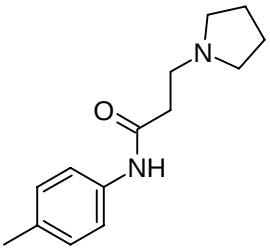
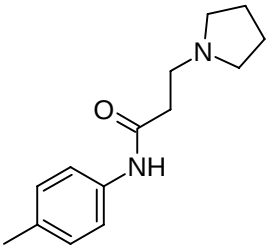
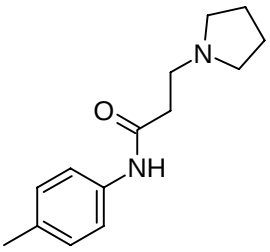


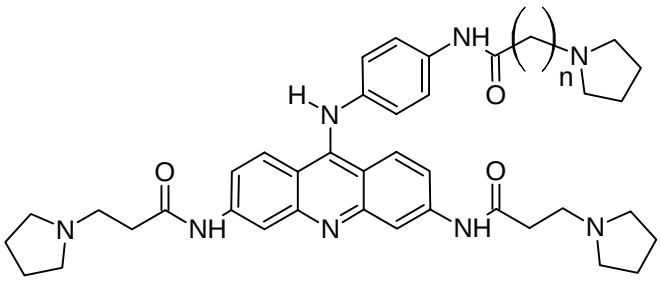
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
20	-$C_6H_4NH_2$ (*p*)	0.08	7.10	7.127
21	-$C_6H_4NMe_2$ (*p*)	0.17	6.77	6.877
22	-$CH_2CH_2NHMe_2$	0.27	6.57	6.037
23a	-$C_6H_4NH_2$ (*m*)	0.21	6.68	6.813
24	-$C_6H_4NH_2$ (*0*)	0.11	6.96	7.169
25*	-C_6H_5	1.33	5.88	-
26a	-$CH_2CH_2CH_2NHMe_2$	0.08	7.10	7.045
27	-C_6H_{11} (*c*)	0.21	6.68	6.713



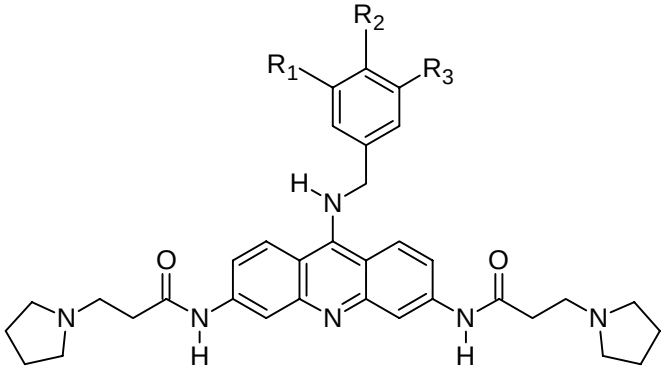
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
28	-C_6H_4OMe(*p*)	0.46	6.34	6.091
29	-$C_6H_4NH_2$ (*o*)	0.17	6.77	6.237
30	-$C_6H_4NH_2$ (*m*)	1.09	5.96	6.182
31a	-$C_6H_4NMe_2$ (*m*)	0.6	6.22	6.059
32	-$C_6H_4NH_2$ (*p*)	0.2	6.70	6.274
33	-$C_6H_4NMe_2$ (*p*)	0.5	6.30	6.070
34	-C_6H_5	1.29	5.89	6.096
35	-C_6H_4OMe (*m*)	2.73	5.56	6.047
36	-C_6H_4OH (*o*)	1.03	5.99	6.105
37	-$CH_2CH_2NHMe_2$	0.57	6.24	6.213



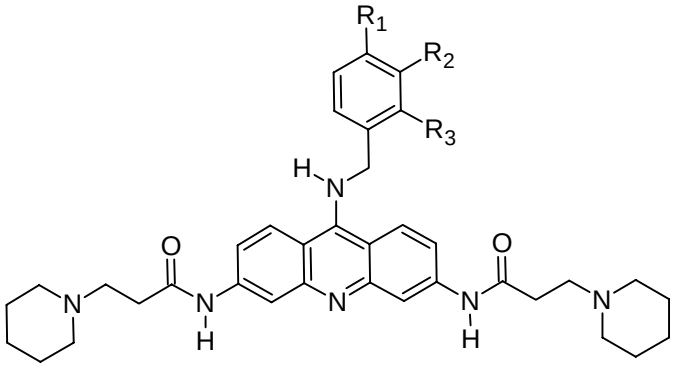
Compound	R	n	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
38	-$C_6H_4N(Me)_2$ (*p*)	3	0.099	7.00	6.635
39	-$C_6H_4N(Me)_2$ (*p*)	4	1.93	5.71	5.805
40	-$C_6H_4N(Me)_2$ (*p*)	5	6.91	5.16	5.302
41		3	0.326	6.49	6.796
42		4	0.255	6.59	6.844
43a		5	0.146	6.84	6.815



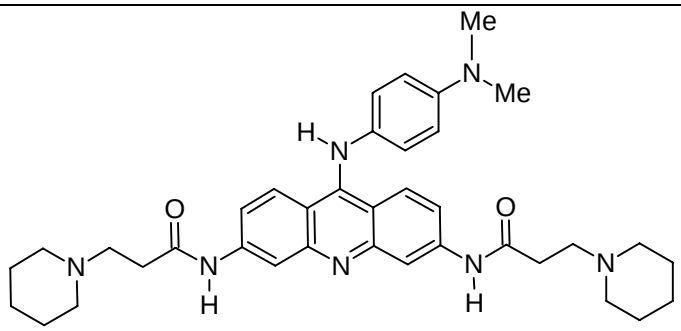
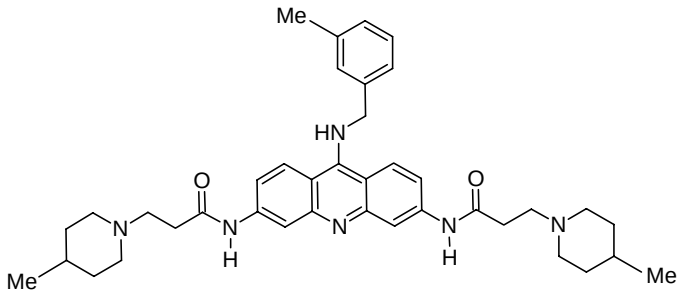
Compound	n	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
44	1	0.167	6.78	6.497
45	2	0.067	7.17	6.485
46	3	0.117	6.93	6.547

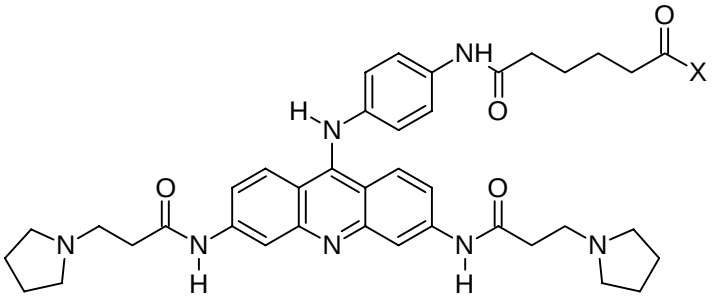


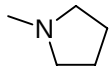
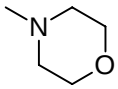
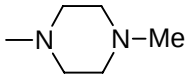
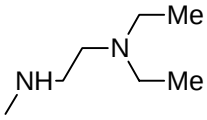
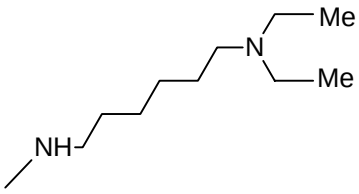
Compound	R_1	R_2	R_3	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
47*	F	F	H	0.03	7.52	-
48a	OCH_3	H	OCH_3	0.35	6.46	6.289
49	CH_3	H	H	1.00	6.00	6.174
50a	CF_3	H	F	0.24	6.62	6.142
51	CF_3	H	CF_3	0.86	6.07	6.092

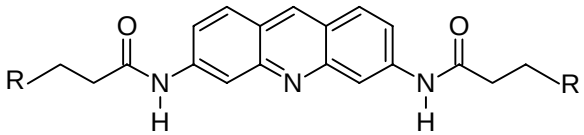


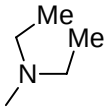
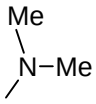
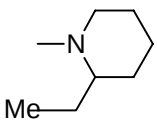
Compound	R_1	R_2	R_3	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
52	F	F	H	0.88	6.06	6.462
53a	OCH_3	H	OCH_3	0.44	6.36	6.301
54	CH_3	H	H	0.23	6.64	6.186

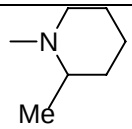
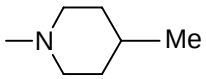
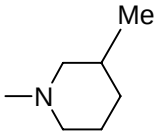
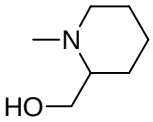
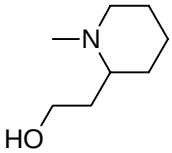
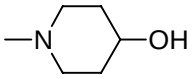
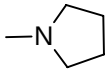
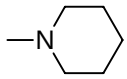
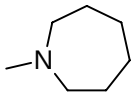
Compound	Structure	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
55		0.36	6.44	6.951
56		0.39	6.41	6.202

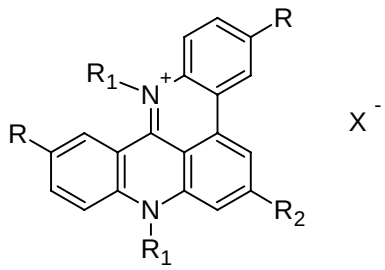


Compound	X	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
57		0.318	6.50	6.630
58a		0.267	6.57	6.686
59		0.165	6.78	6.672
60		0.098	7.01	6.912
61a		0.080	7.10	6.757



Compound	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
62		5.8	5.24	5.506
63		8.2	5.09	5.482
64		2.7	5.57	5.409

Compound	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
65		2.6	5.58	5.385
66		1.35	5.87	5.472
67a		4.4	5.36	5.420
68		5.4	5.27	5.337
69a		4.1	5.39	5.336
70		8.0	5.10	5.364
71		5.2	5.28	5.379
72		2.8	5.55	5.424
73a		3.1	5.51	5.509



Compound	R	R_1	R_2	X^-	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
74a	H	Me	H	I	0.38	6.42	6.229
75	H	Et	Me	I	2.00	5.70	6.191
76	Me	Me	Me	$MeOSO_3$	0.25	6.60	6.218
77a	F	Me	Me	$MeOSO_3$	0.33	6.48	6.247
78	H	Me	Me	$MeOSO_3$	0.76	6.12	6.206
79	Cl	Me	Me	$MeOSO_3$	0.25	6.60	6.272

c = cyclic, *o* = ortho, *m* = meta, *p* = para, a = test set compounds, * = outlier.

4.1.3 Docking studies

Docking studies were performed with different protonation strategies (Figure 4.1.1) of the most active compound (**12**) in order to understand the cause of the significant difference in molecular structure with different protonation sites on binding interactions with G-quadruplex. The decision of right protonation strategy was taken on the basis of docking results.

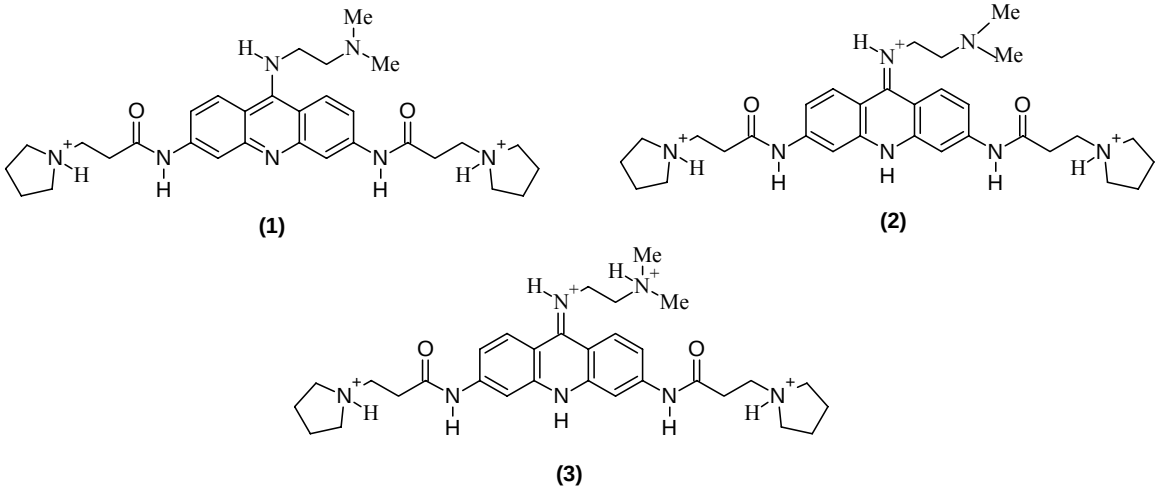
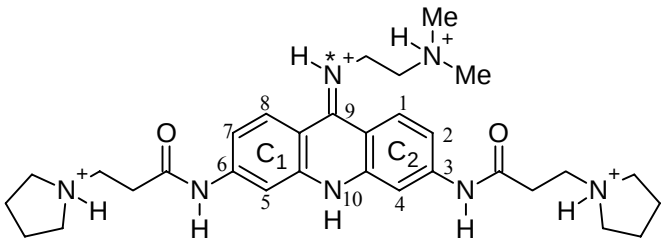


Figure 4.1.1: Different protonation strategies

4.1.4 Selection of Template and Molecular Alignment

In the development of 3D-QSAR models, the choice of the template conformation is the most important factor to provide a reliable 3D-QSAR model. This renders the spatial alignment of molecules under study as one of the most sensitive and determining factors in obtaining a robust and meaningful model. The energy minimized docked conformation [83] of the most active compound (**12**) from the series was used as a template. The molecular alignment was carried out using the centroid and atom-based rms fit. These options use alignment of structures through defining the centroid and pairwise superimposition of these centroids and atoms which render all structures in the database in the same frame of reference as the template compound. The centroids and atoms considered for the alignment are marked with C1, C2 and an asterisk (*), respectively, in Figure 4.1.2. The superimposition of all the compounds on template (compound **12**) is shown in Figure 4.1.3



C1, C2 Centroids and *Atom considered for alignment

Figure 4.1.2: Template used for the alignment

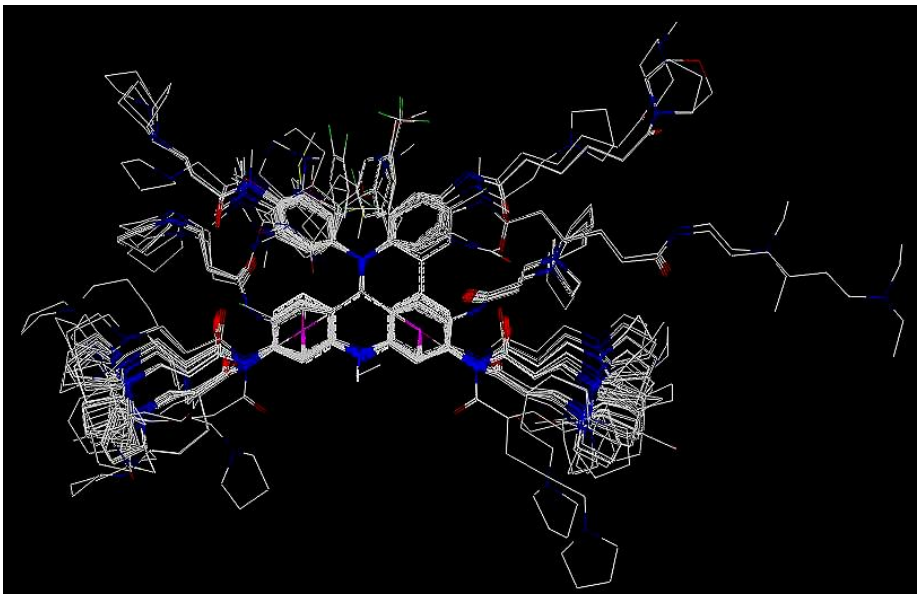


Figure 4.1.3: Superimposition of compounds on template

CoMFA analysis initially was performed on the reported compounds of the training set as such. The model obtained thus, although observed to be on the edge of statistically acceptable limits, offered seriously erroneous contour diagrams wherein the central nitrogen of the acridine ring was shown to have H-bond acceptor region as negatively charged electrostatic contour. The central part of the G-quadruplex quartet is already rich in electron density converting it into an electronegative ion channel that runs along the central region of a DNA G-quadruplex. For inhibition of the enzyme telomerase, the inhibitor has to stack in such a way that the nitrogen atom, of the acridine ring, overlapped with the central electronegative channel. This was not possible with this model as there would be repulsion between the two electron rich species. Since the molecules contained many nitrogen atoms in their structures, it was speculated that some of them could exist in protonated form in a biological environment. So, various protonation strategies (Figure 4.1.1) were applied to the molecules of the series. Nitrogen atoms of the amides and those that were directly connected to phenyl ring were spared from this protonation strategy as these were not supposed to be protonated at physiological *p*H due to their lower basicity.

4.1.5 Docking results

To elucidate the interaction mechanism and right protonation strategy of G-quadruplex ligands, compound (**12**) the most potent ligand in protonated forms with three different protonation strategies (Figure 4.1.1) were independently docked with the G-quadruplex. Interestingly, results of docking studies supported our protonation 'strategy 3' as its model showed favorable binding interactions with G-quadruplex DNA as shown in Figure 4.1.4. It was observed that a side chain interacted better with particular nucleotides Thy 1007 and Thy 1008 present in the X-ray crystal structure of G-quadruplex. A particular conformational geometry adopted by the side chains on the two phenyl rings brings aliphatic substituents present on the ninth position of acridine ring significantly deeper into the cavity of the DNA groove. This leads to better interaction with the surface of the groove when the terminal nitrogen atom is in the protonated state, resulting in low interaction energy compared to other strategies (Table 4.1.2). The docking experiment showed that the terminal protonated nitrogen atom and the 9-ammonium proton of compound (**12**) (strategy 3) can form a strong hydrogen bond with Gua 1009, whereas such interactions were not observed in strategies 1 and 2. In strategy 2, though the nitrogen atom present on the C-9 position

is in a protonated form, it is not a sufficient structural requirement to produce favorable binding interactions.

Table 4.1.2 Interaction energies and CDOCKER energy (score) for the three different strategies using compound (**12**) with quadruplex, in kcal mole^{-1}

Strategy	Interaction energy	CDOCKER energy
1	-92.549	-45.767
2	-88.629	-50.717
3	-104.321	-58.000

However, protonation of a terminal nitrogen atom along with nitrogen present on the C-9 position (strategy 3) resulted in an ideal binding conformation which offered favorable binding interactions of proton on the *N*-10 nitrogen of acridine with the electronegative oxygen of Gua 1009 (Figure 4.1.5). The docked view of compound (**12**) with a solvent accessible surface area is shown in Figure 4.1.6, where the white color of the ligand complements the surface area of the binding site, the blue region indicates nitrogen atoms in the ligands, and potassium ions are shown in yellow balls. Thus, strategy 3 was considered ideal for further development of predictive 3D-QSAR CoMFA and CoMSIA models.

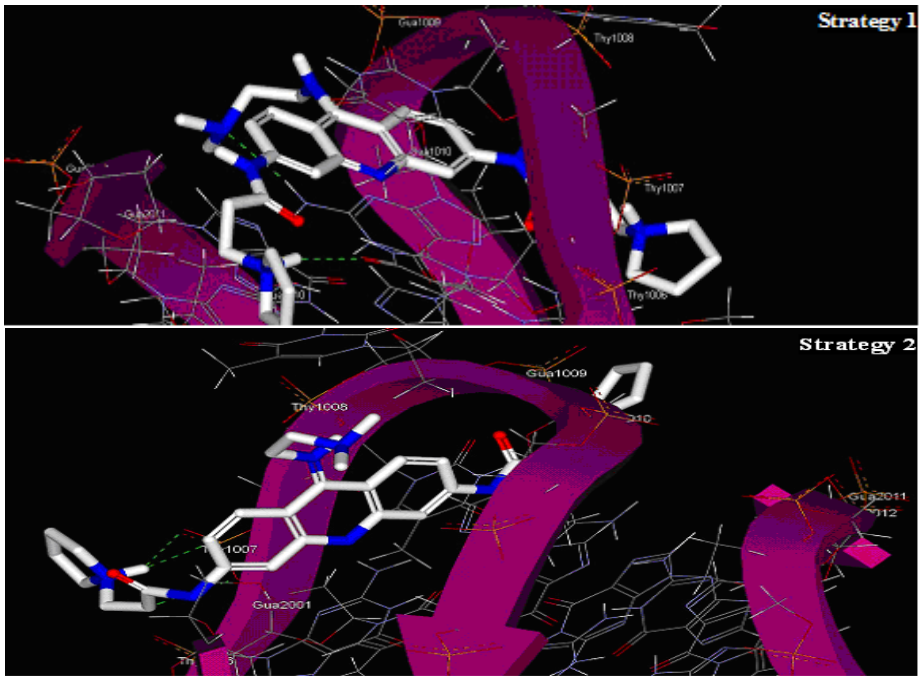


Figure 4.1.4: Interacting docking models for first two strategies

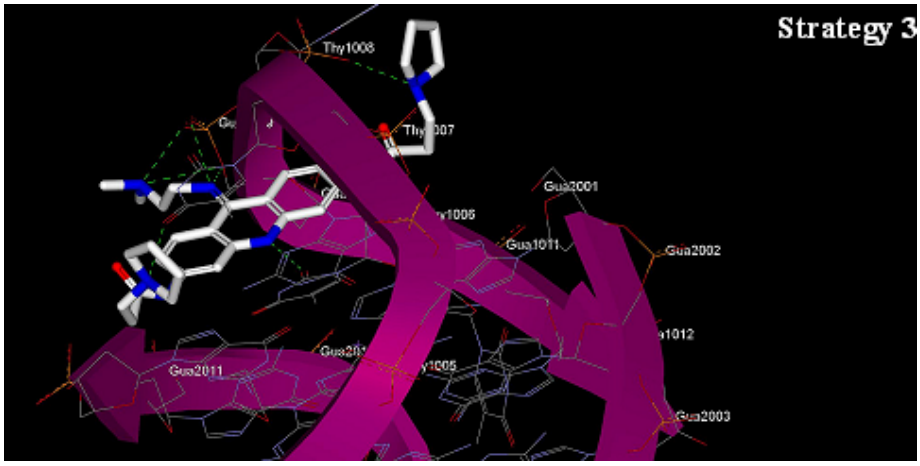


Figure 4.1.5: Interacting docking models for the third strategy

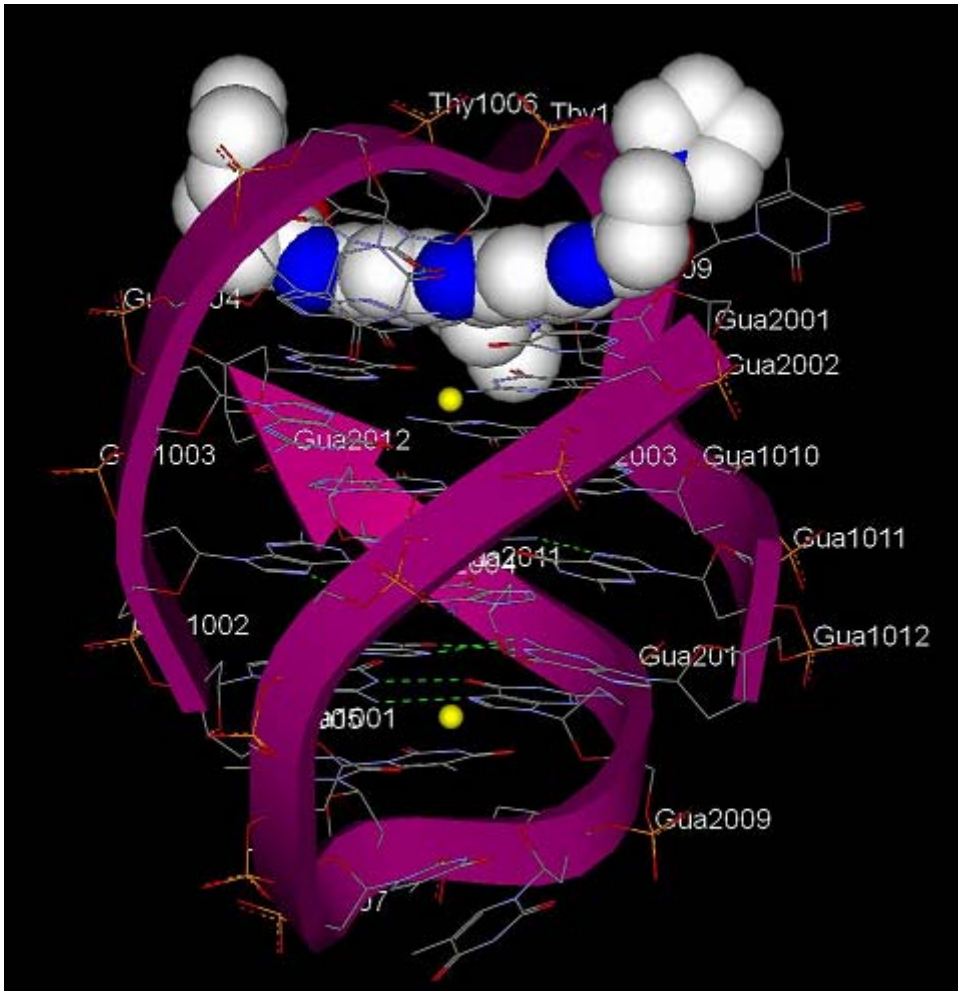


Figure 4.1.6: Docked view of compound (**12**) shown with solvent accessible surface with nitrogen atom in blue and potassium ions in the central channel of the G-quartet in yellow color.

4.1.6 CoMFA Analysis

After alignment of all training set compounds, Analysis-A (Table 4.1.3) yielded a correlation with an r^2_{cv} of 0.540 (4 optimum number of components), a conventional r^2 of 0.778, an *F*-value of 49.054, and an external predictivity with an r^2_{pred} of 0.660. The steric and electrostatic contributions were 67.50% and 32.50%, respectively.

Table 4.1.3 PLS statistical results of CoMFA (Strategy 3)

Parameter	CoMFA	
	Analysis-A	**Analysis-B**
r^2_{cv} [a]	0.540	0.552
SEP[b]	0.477	0.474
ONC[c]	4	3
r^2_{ncv} [d]	0.778	0.783
SEE[e]	0.332	0.330
F-value	49.054	67.184
Prob $r^2 = 0$	0	0
r^2_{pred}	0.660	0.752
r^2_{bs} [f]	0.917	0.919
SD_{bs} [g]	0.018	0.019
Steric contribution	67.50%	62.90%
Electrostatic contribution	32.50%	37.10%

[a]Cross-validated r^2; [b]Standard error of estimate
[c]Optimum number of components; [d]Non-cross validated r^2
[e]Standard error of estimate; [f]Boot-strapped r^2
[g]Boot-strapping standard deviation

In order to increase the predictive power of the model further experiments were performed. Based on the results of QSAR studies from partial least-squares (PLS) analysis, 2 molecules (compounds **25** and **47**) from the training set with high residual values (Figure 4.1.7) were omitted. Compounds (**34)** and (**52)** with similar functional groups at the C9 position as that of compounds (**25)** and (**47)**, respectively, are comparatively well predicted. Hence, it is most likely that outliers are a result of

an experimental error that could happen when analyzing a large data set [212]. After removing outliers, CoMFA results obtained using the training set of 60 compounds showed a better confidence level of higher statistical significance (Analysis-B Table 4.1.3). In 3D-QSAR studies r^2_{cv} of 0.3 is considered statistically significant [30]. In view of it, the models having $r^2_{cv} > 0.5$ can be considered to be much better and statistically significant.

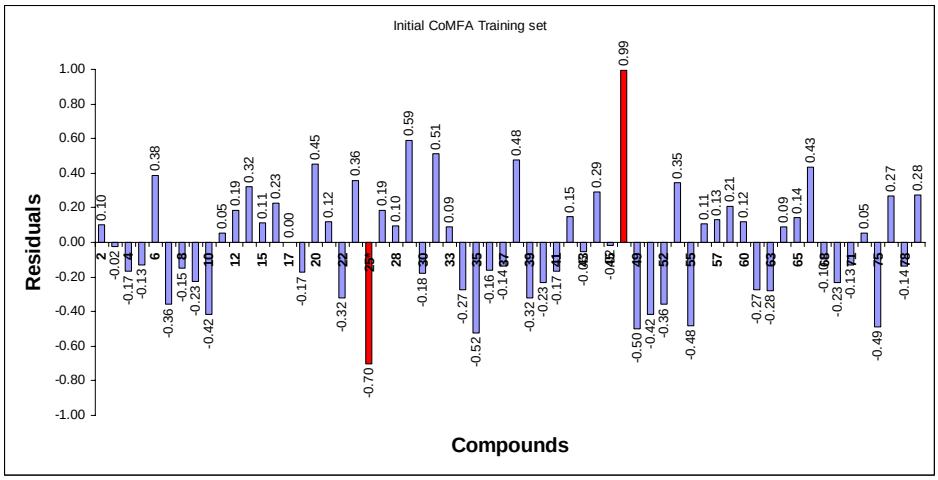
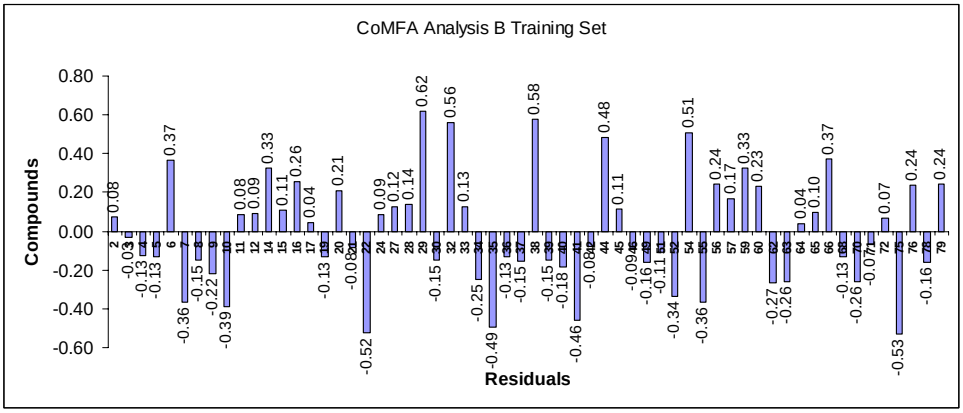
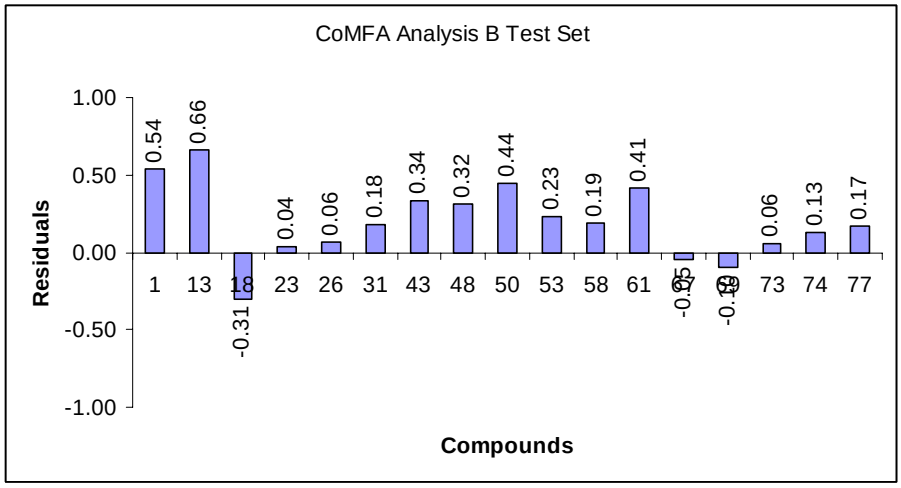


Figure 4.1.7: Histogram of CoMFA residual value for initial training set (Analysis-A)

Analysis-B showed an improved cross-validated r^2 of 0.552 (3 optimum number of components), a conventional r^2 of 0.783, an *F*-value of 67.184, and a predictive r^2 of 0.752. Histograms of residual values of training and test sets are shown in Figure 4.1.8.



(a)



(b)

Figure 4.1.8: Histogram of CoMFA residual values for training set and test set a) Training set b) Test set (Analysis-B)

To further assess the robustness of the model, bootstrapping analysis (100 runs) was performed and an r^2_{bs} of 0.919 (SD_{bs} 0.019) was obtained, further establishing the strength of the model. The steric and electrostatic contributions were found to be 62.90% and 37.10%, respectively. Data set and alignment of CoMFA was used further for CoMSIA analysis.

4.1.7. CoMSIA Analysis

CoMSIA is similar to CoMFA but uses a Gaussian function rather than Coulombic and Lennard-Jones potentials to assess the contribution from different fields. CoMSIA was performed using steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields. 3D-QSAR models were generated using the above fields in different combinations, and the results of the study are summarized in Table 4.1.4.

Table 4.1.4 Summery of CoMSIA analysis

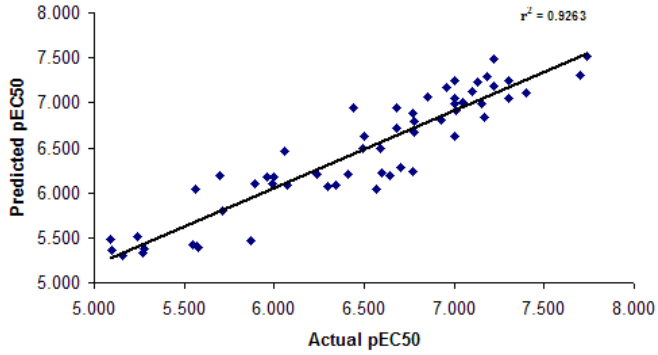
	HAD	SED	SEA	SHE	EHD	SEDA	**SEHD**	SEHA	SEHDA
r^2_{cv}	0.574	0.568	0.532	0.607	0.585	0.552	**0.606**	0.562	0.613
ONC	3	3	2	3	3	3	**3**	3	3
SEP	0.466	0.469	0.484	0.448	0.456	0.478	**0.444**	0.473	0.444
r^2_{ncv}	0.806	0.821	0.658	0.794	0.826	0.816	**0.812**	0.827	0.832
SEE	0.314	0.302	0.414	0.324	0.295	0.306	**0.300**	0.297	0.274
***F*-Value**	76.289	84.011	53.768	70.806	88.437	81.515	**80.440**	87.479	102.024
$P_{r2=0}$	0	0	0	0	0	0	**0**	0	0
Contribution (fraction)									
S	-	0.164	0.159	0.195	-	0.148	**0.122**	0.154	0.109
E	-	0.329	0.322	0.406	0.284	0.253	**0.249**	0.264	0.191
H	0.323	-	-	0.399	0.286	-	**0.237**	0.280	0.218
D	0.395	0.507	-	-	0.430	0.347	**0.392**	-	0.267
A	0.282	-	0.519	-	-	0.252	-	0.303	0.215
r^2_{pred}	0.567	0.823	-	0.895	0.870	0.514	**0.890**	0.644	0.643
r^2_{bs}	0.929	0.917	0.865	0.918	0.890	0.915	**0.908**	0.910	0.923
SD_{bs}	0.016	0.018	0.029	0.019	0.020	0.032	**0.016**	0.017	0.015

ONC = Optimum number of components, SEP = Standard error of prediction, SEE= Standard error of estimate, S = Steric, E = Electrostatic, D = H-bond donor, A = H-bond acceptor and H = Hydrophobic

CoMSIA models showed higher correlation and high predictive properties. In most of the models, hydrophobic field was a common factor indicating the importance of lipophilicity for the present series of molecules. We found that the CoMSIA descriptors such as steric, electrostatic, hydrophobic, and hydrogen bond donor fields played a significant role in the prediction of biological activity. An excellent value of 0.890 for r^2 prediction and 0.606 for r^2 cross-validation with 3 optimum numbers of components were obtained for this model. A good r^2_{ncv} 0.812 was observed for internal prediction of the model. The contribution of steric, electrostatic, hydrophobic, and hydrogen bond donor fields of this model were 12.2%, 24.9%, 23.7% and 39.2%, respectively. Incorporation of hydrogen bond acceptor field descriptor produced no significant change in internal predictivity. However, a substantial decrease in r^2 prediction from 0.890 to 0.644 was observed which was noteworthy. Thus,

incorporation of the hydrogen bond acceptor descriptor to steric, electrostatic, hydrophobic, and hydrogen bond donors in different combinations gave statistically poor models. Exclusion of the steric descriptor from the highly predictive model (SEHD) resulted in slightly improved internal predictivity. However, a little decrease in r^2 prediction from 0.890 to 0.870 was observed. Above findings indicate that steric, electrostatic, hydrophobic and hydrogen bond donor fields can be more favorable for G-quadruplex stabilization, and they could make an important contribution to the binding energy of association. Overall, positively charged ligands with hydrogen bonding ability and large lipophilic patches could offer good binding properties for G-quadruplex indicating the importance of hydrophobic fields and H-bond donor functional groups for the biological activity. The graphs of actual vs. predicted activities for training and test set molecules from the best CoMSIA model (Steric, Electrostatic, Hydrophobic, and Hydrogen bond donor descriptors) are shown in Figure 4.1.9a,b and the histogram of residual values for training and test set molecules are shown in Figure 4.1.10a,b

(a)



(b)

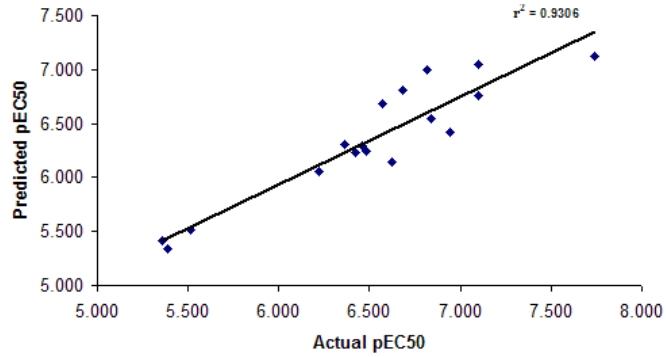
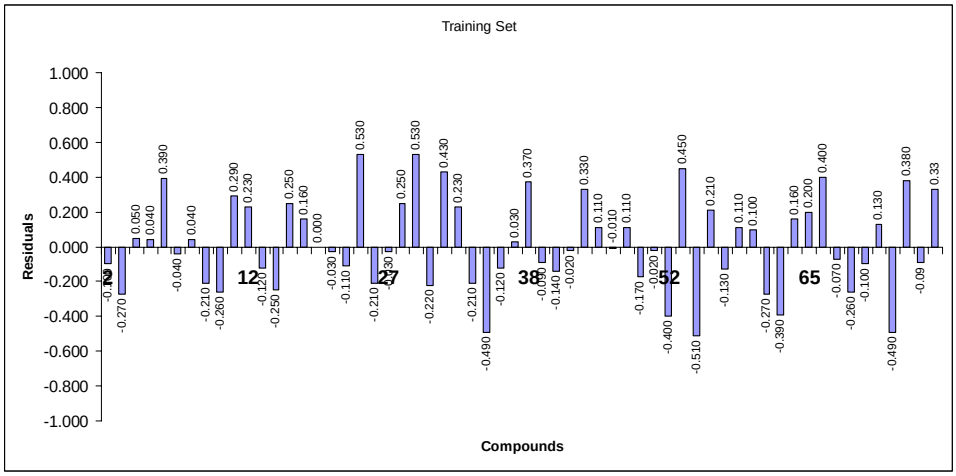
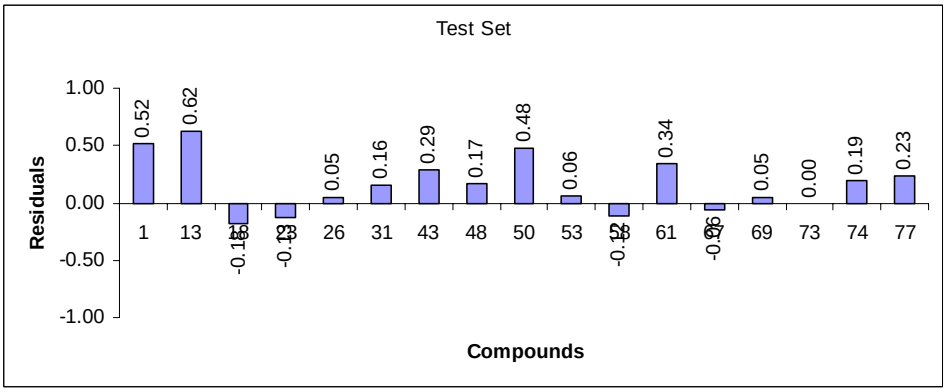


Figure 4.1.9: Graph of Actual vs. Predicted activities for training and test set molecules from the best predictive CoMSIA model. a) Training Set b) Test Set



(a)



(b)

Figure 4.1.10: Histogram of CoMSIA residual values for training set and test set a) Training set b) Test set

4.1.8 Model Validation

4.1.8.1 Internal Validation

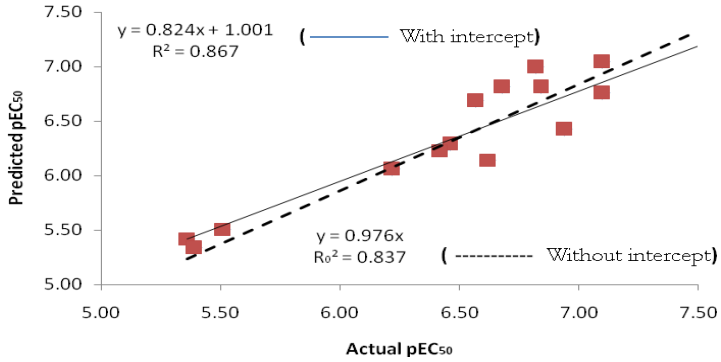
In model validation we examined the internal predictive power of the models and their ability to reproduce biological activities of the compounds for the training set. The computed telomerase inhibitory activity from the CoMFA and CoMSIA studies showed a good correlation with experimental telomerase inhibitory activity.

4.1.8.2 External Validation

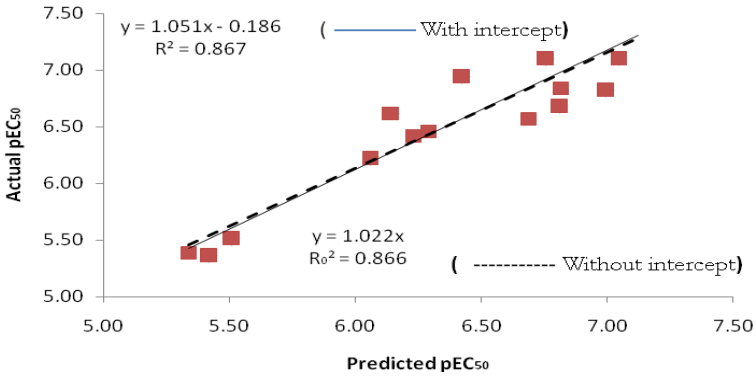
External validation of QSAR models was done to verify the excellent statistical parameters that were obtained and also to investigate whether the activity of substituted acridines from external data could be predicted well with this model. A large data set of compounds containing 17 molecules, as a test set (Table 4.1.1), was

selected from the data set of 77 compounds for the validation experiments. The ultimate test for predictability of QSAR analysis in the drug design process is to predict the biological activity of new compounds that have not been included in the training set. The values of r^2_{pred} were calculated for the test set and gave the best results for CoMFA and CoMSIA with the values of 0.752 and 0.890, respectively. Thus, the CoMSIA model displays higher predictivity both in regular cross-validation and in the prediction of the test compounds.

Besides this, the regression plots of actual vs. predicted activities and predicted vs. actual activities were constructed (Figure 4.1.11a,b) to apply rigorous validation parameters reported by Trosha et al. that have been described in chapter 1, section 1.1.1.6.



(a)



(b)

Figure 4.1.11: External validation of CoMSIA model. (a) Actual vs. Predicted activities (b) Predicted vs. Actual activities, of Test Set compounds

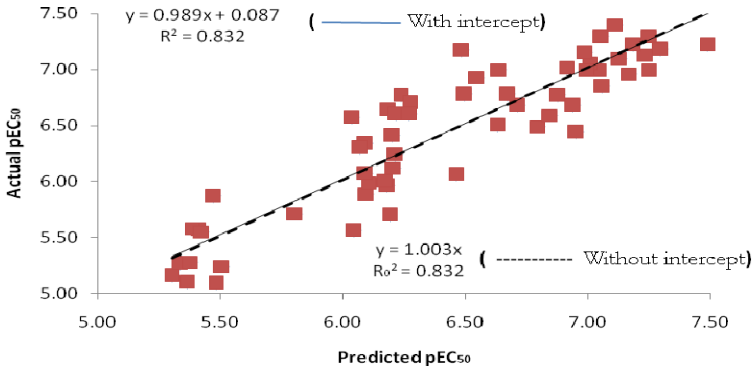


Figure 4.1.12: Internal Validation of best CoMSIA model. Predicted vs Actual. activities of Training set compounds

4.1.8.3 R^2_m as a validation criterion

Additionally, the model was validated with validation criterion R^2_m reported by Roy et.al [40, 41]. The R^2_m (Test) value penalizes a model for large differences between observed and predicted values of test set compounds. The CoMSIA model was found to be superior with R^2_m (Test) = 0.866. Since this value is based on limited number of test set compounds, to further confirm its reliability, R^2_m values were calculated for training set compounds using LOO predictions. The value of R^2_m (training) 0.832 confirms that the model built so is a reliable one. The results obtained are shown in Table 4.1.5. In case of best external prediction $R^2_m = R^2$ while in the worst case, $R^2_m = 0$. In our case for external test set compound value of R^2_m 0.866 was found to be quite close to R^2 value 0.677 and for training set compound value of R^2_m 0.832 was found to be exactly equal to R^2 value 0.832 (Figure. 4.1.12), indicating that the derived model is a highly predictive one.

4.1.8.4 Percentage of residual values (% Residual)

In the present study, a new validation parameter was analyzed and investigated to check the reliability of 3D-QSAR model. The percentage residual value is mainly governed by the sum of actual activities of test set compound and absolute sum of differences between actual and predicted values of the test set compounds. The percentage of residuals of differences between actual and predicted values of test set compounds was calculated as a percentage of the summed up observed activities of the test set compounds. The above validation criterion was fulfilled when applied to the present 3D-QSAR model. Thus, percentage residual value could be important

criterion to test the reliability of a newly developed QSAR model. The results are shown in Table 4.1.5

Table 4.1.5 Summary of additional validation for CoMSIA model

Statistics	CoMSIA model (SEHD)
R^{2a}	0.867
R_0^{2b}	0.837
k^{c}	0.976
$(R^2 - R_0^2)/R^2$	0.034
$R'_0{}^{2b}$	0.866
k'^{c}	1.022
$(R^2 - R'_0{}^2)/R^2$	0.001
$R_m^2 = R^2 (1-\sqrt{\lvert R^2 - R_0^2 \rvert})$	0.866
Absolute sum of observed *p*IC_{50} (Test set)	110.61
Absolute sum of residuals (Test set)	3.38
% residual (Test Set)	3.05%
Absolute sum of observed *p*IC_{50} (Training set)	388.06
Absolute sum of residuals (Training set)	13.82
% residual (Training set)	3.56%

[a] Correlation coefficient derived from predicted pIC_{50} of test set compounds
[b] Correlation coefficient for the regression line passing through origin for actual vs. predicted and predicted vs. actual activity, respectively
[c] Slope for regression line passing through origin obtained from actual vs. predicted and predicted vs. actual activity, respectively

4.1.8.5 Fischer Statistics (*F*-Test)

Fischer statistics (*F*) is the ratio between explained and unexplained variance for a given number of degrees of freedom. F-test is a variance related statistics that compares two models differing by one or more variable to see if the more complex model is more reliable than the less complex one. The model is supposed to be good if the *F*-test is above a threshold value, i.e. tabulated value. The larger the value of *F*, the greater the probability that the QSAR equation is significant. The *F* values for the CoMFA and CoMSIA models were 67.184 and 80.440, respectively, [CoMFA/CoMSIA F_{05} (3, 54) = 2.84 (Tab)] at 95% confidence level, which suggests that these models are statistically significant with CoMSIA having an upper hand over CoMFA.

4.1.9 Visualization of Contour Maps

CoMFA steric and electrostatic contour maps are not much different than CoMSIA, hence these are not discussed. CoMSIA steric fields are shown in Figure 4.1.13. The green contours represent regions of high steric tolerance (80% contribution), while the yellow contours represent regions of low steric bulk tolerance (20% contribution).

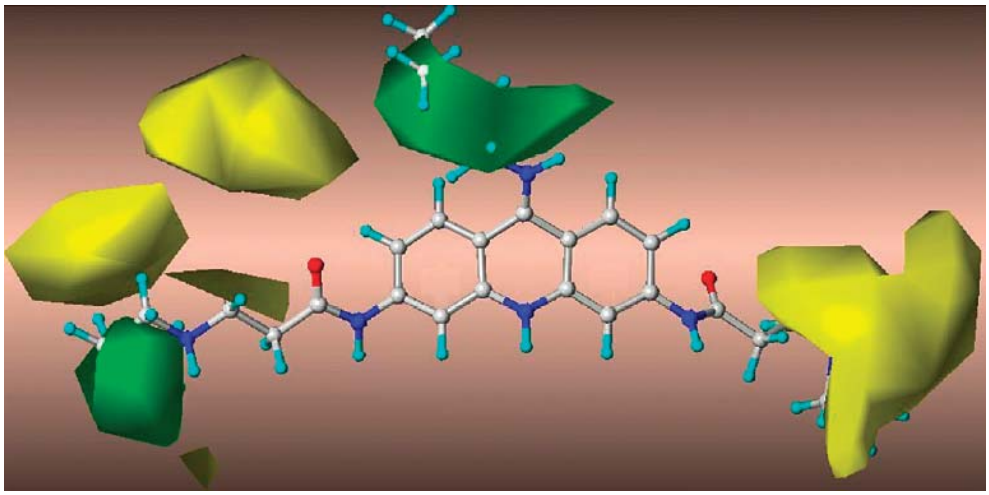


Figure 4.1.13 CoMFA stdev*coeff steric contour maps. The most active compound (**12**) is displayed in the background. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the activity.

Prominent green contours present in the vicinity of the ninth position of the acridine ring and small contours present surrounding the pyrrolidine rings indicate that generally steric bulk is favored at these sites. The good inhibitory potency of compounds (**12**) ($^{tel}EC_{50}$ 0.018) and (**13**) ($^{tel}EC_{50}$ 0.018) is due to orientation of the C-9 substituted bulkier group toward the sterically favored regions. While in the case of compounds (**63**) ($^{tel}EC_{50}$ 8.2), (**68**) ($^{tel}EC_{50}$ 5.4), (**70**) ($^{tel}EC_{50}$ 8.0), (**71**) ($^{tel}EC_{50}$ 5.2), and (**73**) ($^{tel}EC_{50}$ 3.1) there was an absence of a bulkier group in this region. The cycloheptane ring present at the C9 position of compound (**10**) ($^{tel}EC_{50}$ 0.21) was oriented away from the prominent green contour, and the C3 side chain was directed toward a sterically unfavorable yellow region that may lead to steric clashes in these regions. This makes compound (**10**) comparatively less active among the 3,6,9-substituted acridine derivatives. 2,6,9-Substituted acridine compounds (**21** to **27**) show less activity than that of 3,6,9-substituted acridine as the side chain substituted

at the C2 position is directed toward sterically unfavorable regions. Compound (**20**) ($^{tel}EC_{50}$ 0.08) showed good inhibitory potency as a sterically favored space was occupied by the ninth substituted aryl ring. In the case of the 2,7,9-substituted acridine side chain at the C-7 position of acridines; it directs toward yellow contours which restrict the steric substitution. This is in line with the SAR indicating lower activities for compounds (**30**) ($^{tel}EC_{50}$ 1.09), (**34**) ($^{tel}EC_{50}$ 1.29), (**35**) ($^{tel}EC_{50}$ 2.73), and (**36**) ($^{tel}EC_{50}$ 1.03). The case of compound (**38**) ($^{tel}EC_{50}$ 0.099) with three carbons in C-3, C-6 substituted side chains is more favorable as these side chains are perfectly entrenched in sterically favored regions. Incorporation of an additional -CH_2 group in the carbon spacer chain directs the pyrrolidine ring of either side into disfavorable yellow contours, e.g. compounds (**39**) ($^{tel}EC_{50}$ 1.93) and (**40**) ($^{tel}EC_{50}$ 6.91). Compounds (**49**) ($^{tel}EC_{50}$ 1.00), (**51**) ($^{tel}EC_{50}$ 0.86) and **52** ($^{tel}EC_{50}$ 0.88) are relatively less active because the benzylamino substituent changes its conformation and directs away from the sterically favorable green contours. Also, the 3,6-substituted side chain changes its orientation to sterically unfavorable yellow contours. This could be one of the reasons why these compounds showed lower activity than anilino substituted acridines. Among the polycyclic acridines, compound (**75)** ($^{tel}EC_{50}$ 2.0) showed lower activity because of the protrusion of the -CH_2CH_3 group away from the sterically favorable green contours.

CoMSIA electrostatic fields are shown in Figure 4.1.14. The increase in positive charge is favored in the blue regions, while the increase in negative charge is favored in the red regions. The electrostatic contours of CoMSIA show blue regions surrounding the central ring nitrogen atom of acridine indicating that incorporation of electropositive substituents would enhance the activity. Blue contours in the vicinity of pyrrolidine rings of both of the side chains substituted at C-3 and C-6 of the acridine ring have been observed. A blue favorable electrostatic region is observed above the acridine ring in the model. This particularly explains the importance of the acridine chromophore for π-π or van der Waals interactions with G-quadruplex. These findings could be used for development of new generation of acridine analogues by introducing more electropositive atoms at these positions.

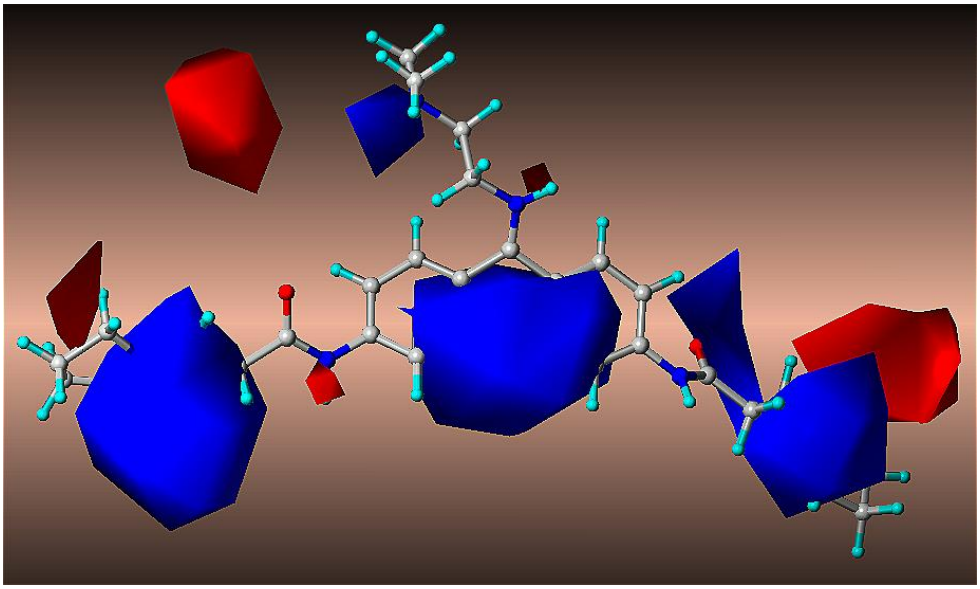


Figure 4.1.14: CoMFA stdev*coeff electrostatic contour maps. The most active compound (**12**) is displayed in the background. Blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity.

Though, the molecular structure of acridine derivatives was built with a general structure (**4**) (Figure 4.1.15), the compounds predominantly exist in the structural form (**5**) as shown in Figure 4.1.15.

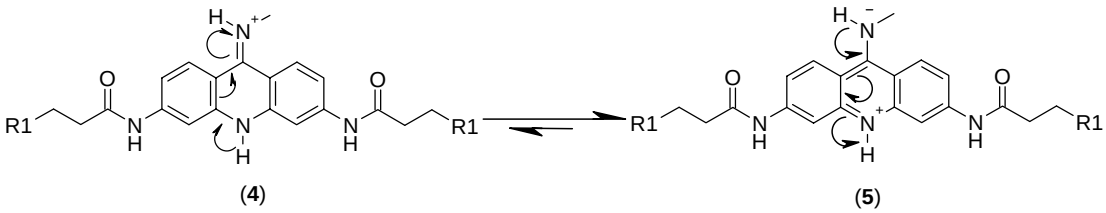


Figure 4.1.15 Predominant structure of acridine

A big blue contour surrounding *N*-10 and a red contour at the nitrogen attached to C-9 is indicative of a positively charged nitrogen (*N*-10) and a lone pair of electrons on a nitrogen attached to C-9 respectively. This is a very important finding which shows that *N*-10 is much more basic than nitrogen attached to C-9 and gets protonated faster. Aromaticity of all the three rings of acridine is also preserved, making structure (**5)** to be much more stable than structure (**4)** (Figure 4.1.15).

Compound (**12)** ($^{tel}EC_{50}$ 0.018) was found to be the most active one among the series due to the reason that the terminal nitrogen atom on ninth substituent, pyrrolidine ring nitrogen atoms on both the C3 and C6 side chains and the central ring

nitrogen atom of the acridine are perfectly entrenched into positively charged favored blue contours. Among the 3,6,9-trisubstituted acridines compound (**9**) ($^{tel}EC_{50}$ 0.14) showed poor activity probably due to the fact that the terminal -OCH_3 group is embedded in an electropositive blue region. The influence of the positively charged favored blue region was seen in compound (**20**) ($^{tel}EC_{50}$ 0.08), which is the most potent compound among 2,6,9-trisubstituted acridines. Compound (**20**) showed better telomerase inhibitory activity than compound (**22**) ($^{tel}EC_{50}$ 0.20) among the 2,6,9-trisubstituted acridines due to the reason that the *p*-NH_2 group (which might not get protonated due to its lesser basicity) on the phenyl ring is oriented in the blue region. However, the terminal nitrogen on the C-9 substituted aliphatic chain in compound (**22**) is oriented away from the positively charged favored region.

Compounds (**63** to **72**) exhibit comparatively low activity owing to the absence of electropositive side chains at the C-9 position of the acridine ring. Compounds (**34**) ($^{tel}EC_{50}$ 1.29) and (**35**) ($^{tel}EC_{50}$ 2.73) exhibit low activity as an electropositive pyrrolidine nitrogen present on the C-7 substituted side chain is embedded in electronegative red contours. In addition to this, in case of compound (**35**) the -OCH_3 group present on the phenyl ring is also oriented toward electropositive blue contours. This could be the reason for poorer activity of compound (**35**) than compound (**34**). Electropositive hydrogen atoms of the C-3 and C-6 substituted side chain in compound (**40**) ($^{tel}EC_{50}$ 6.91) embedded in negatively charged favored red regions could be the reason for its lower activity.

Hydrophobic maps shown in Figure 4.1.16 indicate that the lipophilic favorable yellow region is found surrounding the aromatic acridine ring. This suggests that lipophilicity of this portion of the molecule is an important factor for the activity. The white region under yellow contour at the ninth position of the acridine ring indicates that a hydrophilic atom in this zone favors the inhibitory activity. Compounds with aniline and benzylaniline substituents at the C-9 position of acridine show good activity as an aromatic ring fits well into the yellow region which is favorable for hydrophobic groups.

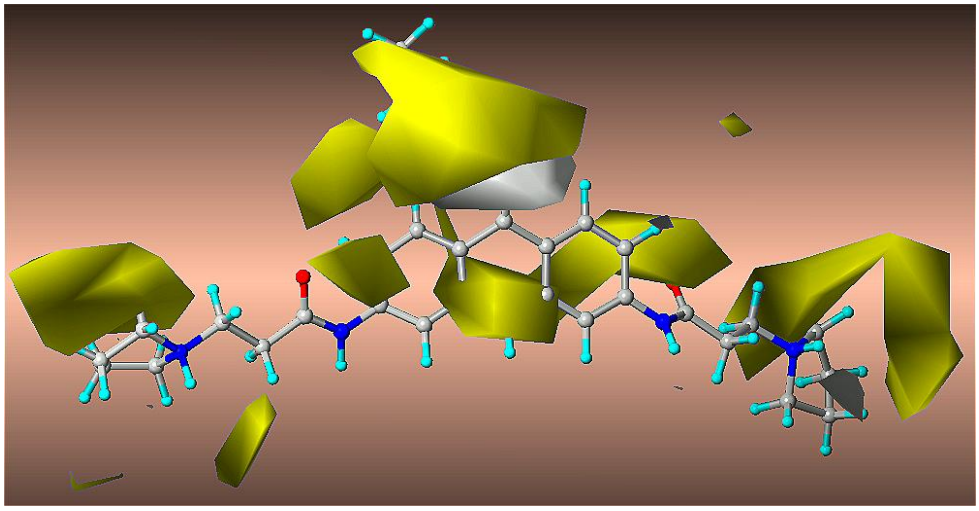


Figure 4.1.16: CoMSIA stdev*coeff hydrophobic contour maps. The most active compound (**12**) is displayed in the background. Yellow contours indicate regions where hydrophobic groups increase activity, white contours indicate regions where hydrophobic groups decrease activity.

In compounds (**62** to **73**), absence of lipophilic substituents at the C-9 position of acridine could be one of the reasons for their lower range of activity. The lipophilic ethyl group of compound (**75**) ($^{tel}EC_{50}$ 2.00) embedded in an unfavorable white region is responsible for its low activity.

The graphical interpretation of the field's contributions of the H-bond donor properties is shown in Figure 4.1.17. In principle H-bond donor contours should highlight areas beyond the ligands where putative hydrogen partners in the target could form H-bonds to influence binding affinity. A cyan colored hydrogen bond donor region surrounds the C-9 position terminal nitrogen atom, pyrrolidine ring nitrogen atoms and the central ring nitrogen atom of the acridine ring indicating a favorable region for H-bond donor fields. Purple disfavorable regions are found in the vicinity of the pyrrolidine ring substituted on C-6 and away from the amido alkyl chain substituted at C-3 of the acridine ring. Compound (**20**) ($^{tel}EC_{50}$ 0.08) which contains a *p*-NH_2 group on the phenyl ring oriented toward the cyan region has a high potency among 2,6,9-trisubstituted acridines. Compound (**34**) ($^{tel}EC_{50}$ 1.29) and compound (**35**) ($^{tel}EC_{50}$ 2.73) show lower activity as the C-2 and C-7 substituted side chains are oriented away from the favorable cyan region.

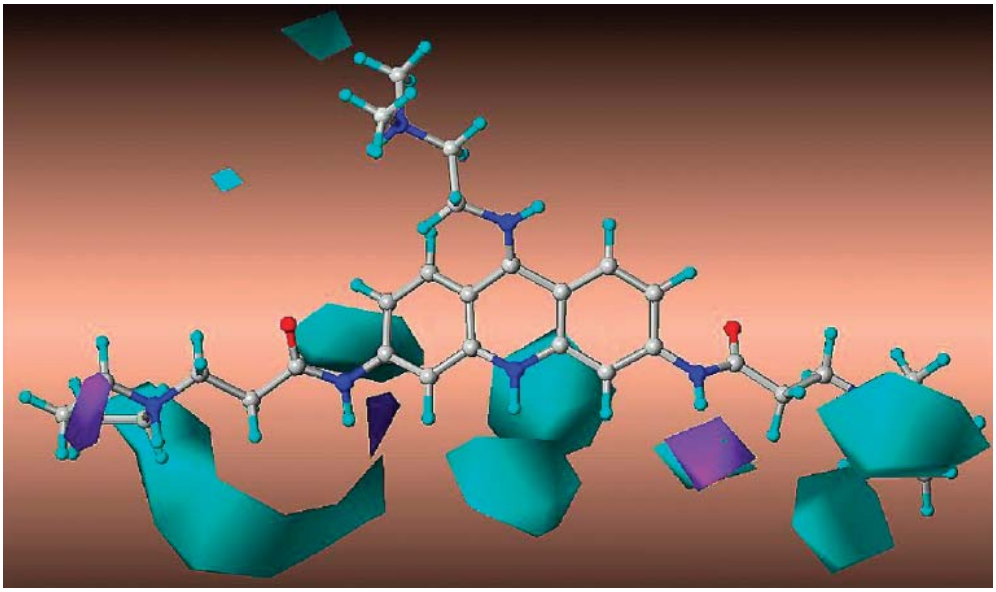


Figure 4.1.17: CoMSIA stdev*coeff hydrogen bond donor contour maps. The most active compound (**12**) is displayed in the background. Cyan contours indicate regions where H-bond donor group increases activity, whereas purple contours indicate regions where H-bond donor group decreases activity.

The role of hydrogen bond donor contributions can be clearly explained by analyzing the compounds (**39** and **40**). Pyrrolidine ring nitrogen atoms present on the C-3 and C-6 positions of acridine in compound (**39**) ($^{tel}EC_{50}$ 1.93) and compound (**40**) ($^{tel}EC_{50}$ 6.91) cross the hydrogen bond donor favorable cyan region due to an increase in side chain length. This could be the reason for their low activities. Compounds (**62** to **73**) exhibit lower activity due to an unsubstituted C-9 position which plays a key role in hydrophobic and hydrogen bond donors fields.

4.1.10 Pharmacophore Model

Using the results obtained by the best predictive CoMSIA model, a putative pharmacophore model has been proposed. To propose the pharmacophore model, the distance of contour maps surrounding the H-bond donor groups was measured. In the proposed preliminary pharmacophore model, pharmacophoric distances using the G-quadruplex stabilizing ligands resulted in the distances of 14.55 Å (**A1-A2**), 9.982 Å (**A1-A3**), 18.840 Å (**A2-A3**), 11.343 Å (**A2-A4**), 9.618 Å (**A3-A4**), 8.005 Å (**R-A1**), 9.822 Å (**R-A2**) and 9.674 Å (**R-A3**).

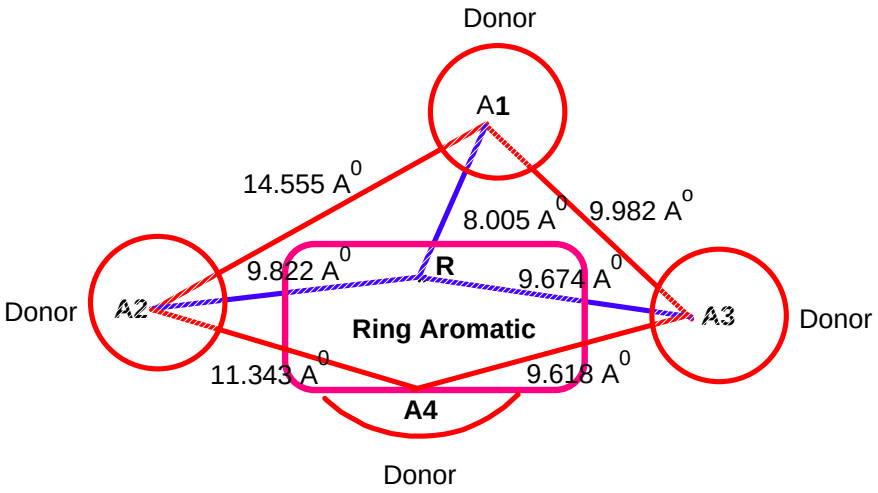


Figure 4.1.18: Pharmacophore model derived from CoMSIA study

The 5 point pharmacophore model along with the interfeature distances are shown in Figure 4.1.18. The key structural components for G-quadruplex stabilization are bulky hydrophobic and electropositive C-9 substitution, cationic sites, and a tricyclic aromatic proton donor pharmacophore which may be essential for perfect orientation of side chains into quadruplex grooves.

After a detailed study of CoMFA and CoMSIA for substituted acridines, it could be concluded that:

1. Functionality attached to the ninth position of the acridine ring having some steric bulk with a positive charge may lead to compounds with higher activity. Bulky electron withdrawing groups along with high hydrophobicity may lead to improved telomerase inhibitory activity.

2. The number of protonated nitrogen atoms in the molecule brings charge transfer electrostatic interactions with higher chemical stability in the biological system and crucially govern the telomerase inhibitory activity.

3. Maintaining the bulkiness of the molecule and providing specific conformation to the molecule can also achieve an optimum G-quadruplex stabilization thereby increasing telomerase inhibition.

4. Hydrogen bonds are considered to make an important discriminatory contribution in the binding of ligands to G-quadruplex; this could form a basis for a strategy to design new G-quadruplex-binding ligands.

It is clear that 3D-QSAR information obtained from this kind of analysis provides important guidelines for the drug design process. As a consequence, the outcome of this study could be used as guidelines for further development of selective and more potent telomerase inhibitors. These proposed models can also be used to predict the activity of newly designed analogs, prior to synthesis. It is hoped that these models could help in design and optimization of acridine derivatives as better telomerase inhibitors as promising anticancer agents.

4.2 Development of predictive 3D-QSAR CoMSIA models for anthraquinone and acridone derivatives as telomerase inhibitors targeting G-quadruplex DNA telomere

4.2.1 Data set used

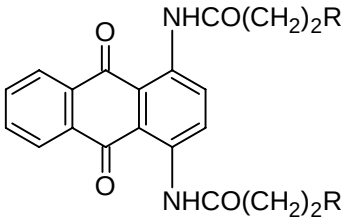
The molecular structures and activities of 61 telomerase inhibitors were taken from the literature [146, 155, 162, 167]. The telomerase inhibitory activity of compounds is reported as EC_{50} values in the micro molar (μM) range. The selected compounds cover a wide range of biological activity (0.2–50 μM) and diverse structural features. The reported EC_{50} values were converted into –log (EC_{50}), i.e. pEC_{50} for use in the QSAR studies. The whole set of 61 compounds (Table 4.2.1) was divided into training set of 49 and test set of 12 compounds in the process of model refinement for all CoMSIA models reported herein. In the training set, most potent, moderately active and lowly active compounds were included to spread the activity range. The test set compounds were selected in such a manner that at least one structural analog remained in the training set.

4.2.2 Selection of template and molecular alignment

Three different molecular alignments were carried out in the present study. Initially, two molecular alignments, based on MacroModel Monte Carlo conformational search-derived templates were derived using centroid and atom-based alignment rule in SYBYL. The two conformational search-based templates were derived as follows: The most active compound (**61**) was constructed in SYBYL. Its structure was minimized, and then used as a starting point for a Monte Carlo conformational search employing MacroModel version 7.0 (Schrödinger, Inc.). The conformational search was carried out using MMFF94 force field in MacroModel for

5000 iterations using water as a solvent. The global minimum conformation, designated as Conf1, and the second lowest energy conformation from the global minimum, designated as Conf2 were used as templates to construct the rest of the molecules.

Table 4.2.1 Molecular structures and telomerase inhibitory activity of anthraquinone and acridone derivatives



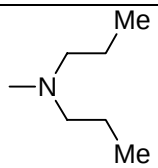
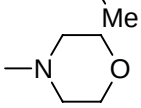
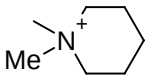
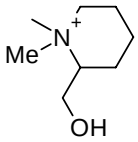
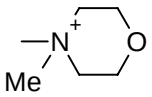
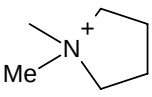
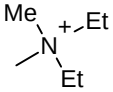
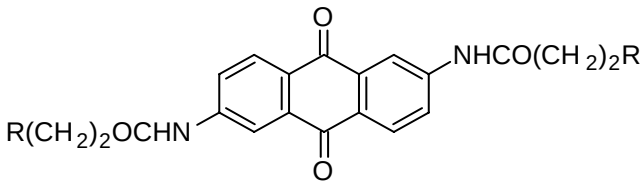
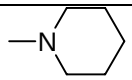
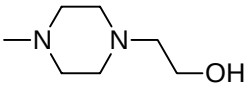
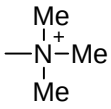
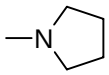
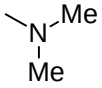
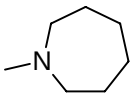
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
1		50	4.30	4.220	0.08
2		33.5	4.47	4.469	0.00
3		11.1	4.95	5.260	-0.31
4		9.4	5.03	5.299	-0.27
5		34.5	4.46	4.336	0.12
6		1.8	5.74	5.762	-0.02
7a		5.0	5.30	5.265	0.04
8		7.0	5.15	5.311	-0.16
9		3.1	5.51	5.362	0.15

Table 4.2.1 (*continued*)

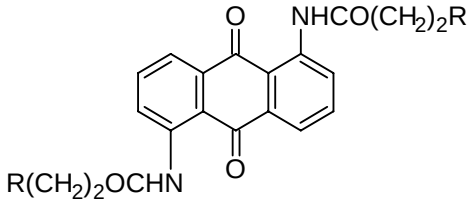


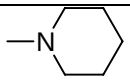
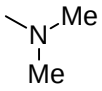
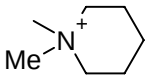
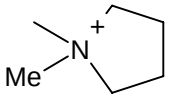
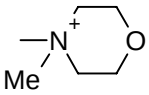
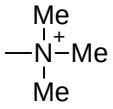
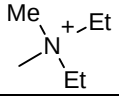
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
10a		4.5	5.35	5.443	-0.09
11		16.5	4.78	4.703	0.08
12a		17.3	4.76	5.033	-0.27
13a		1.8	5.74	5.216	0.52
14		4.1	5.39	5.403	-0.01
15		3.5	5.46	5.443	0.02
16		13.0	4.89	5.084	-0.19

A third alignment was derived as discussed below:

The conformation of a disubstituted acridine derivative (biologically active moiety) was extracted from the co-crystallized structure of G-quadruplex PDB 1L1H [213]. Conf1 of the most active compound (**61**) was superimposed on the conformation of the acridine derivative extracted from the co-crystallized structure. As a measure of superimposition reliability, the RMSD between these two conformations was taken into the consideration (0.9Å obtained herein).

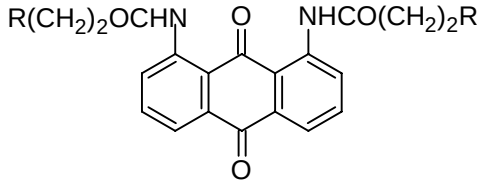
Table 4.2.1 (*continued*)



Compound	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
17		2.3	5.64	5.544	0.10
18a		1.3	5.89	5.444	0.45
19		2.7	5.57	5.486	0.08
20		8.6	5.07	5.097	-0.03
21		8.8	5.06	5.100	-0.04
22a		14.0	4.85	4.537	0.31
23		13.2	4.88	4.716	0.16
24		16.8	4.77	4.682	0.09

It was interesting to note that the resulting conformation (conf3) of compound (**61**) fully superimposed (RMSD 0.00 Å) on conformation (conf1). Hence, conformations (conf1 and conf2) only were used as templates for further study. The partial atomic charges required for the electrostatic interactions were computed by the semi empirical molecular orbital method using Molecular Orbital PACkage (MOPAC) [214] with Austin Model 1 (AM1) Hamiltonian [215]. The centroids and atoms considered for alignment are marked with C1, C2 and an asterisk (*), respectively in Figure 4.2.1. The superimposition of all the compounds on template (compound **61**) is shown in Figure 4.2 2.

Table 4.2.1 (*continued*)



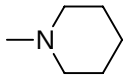
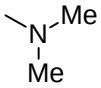
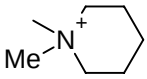
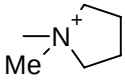
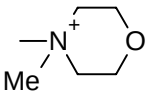
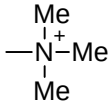
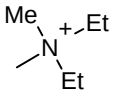
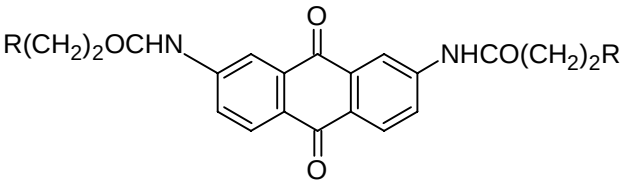
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
25		3.7	5.43	5.434	0.00
26		6.4	5.19	5.368	-0.18
27		4.2	5.38	5.382	0.00
28a		7.8	5.11	5.153	-0.04
29		8.2	5.09	5.152	-0.06
30		10.0	5.00	4.984	0.02
31		4.4	5.36	5.174	0.19
32		7.5	5.12	5.173	-0.05

Table 4.2.1 (*continued*)



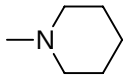
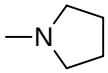
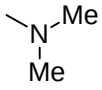
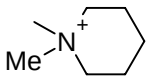
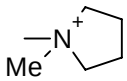
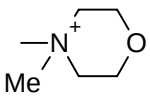
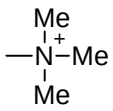
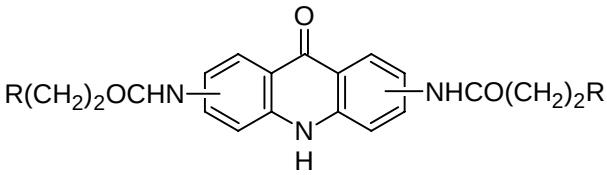
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
33		3.1	5.51	5.764	-0.25
34		2.0	5.70	5.736	-0.04
35		4.7	5.33	5.154	0.18
36		4.3	5.37	5.198	0.17
37a		7.8	5.11	4.986	0.12
38a		16.0	4.80	4.979	-0.18
39		16.5	4.78	4.853	-0.07
40		14.5	4.84	4.812	0.03

Table 4.2.1 (*continued*)



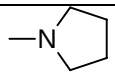
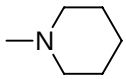
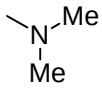
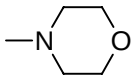
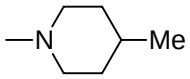
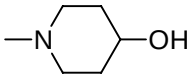
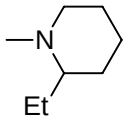
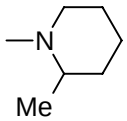
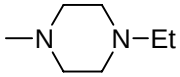
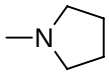
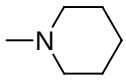
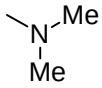
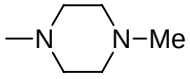
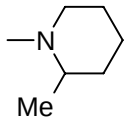
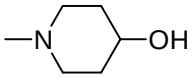
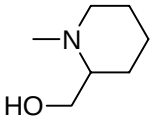
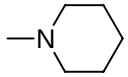
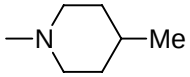
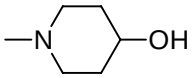
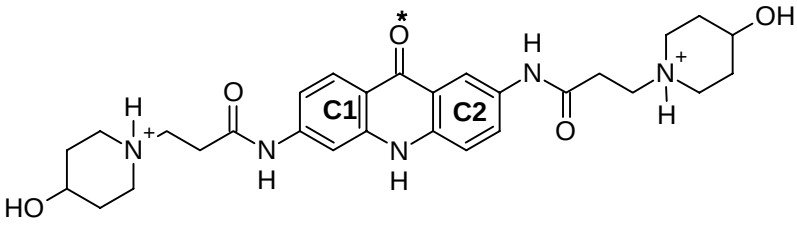
Compound	Position	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
41	3,6		8.1	5.09	5.326	-0.24
42	3,6		5.9	5.23	5.460	-0.23
43	3,6		4.3	5.37	5.381	-0.01
44	3,6		5.7	5.24	5.416	-0.18
45	3,6		49	4.31	4.532	-0.22
46	3,6		1.7	5.77	5.536	0.23
47	3,6		1.7	5.77	5.428	0.34
48	3,6		2.3	5.64	5.557	0.08
49a	3,6		2.3	5.64	5.507	0.13
50	3,6		1.9	5.72	5.634	0.09
51	2,7		5.8	5.24	5.359	-0.12
52	2,7		1.9	5.72	5.882	-0.16
53	2,7		0.6	6.22	5.906	0.31
54a	2,7		1.9	5.72	5.643	0.08

Table 4.2.1 (*continued*)

Compound	Position	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}	Residual
55	2,7		1.5	5.82	5.934	-0.11
56	2,7		2.3	5.64	5.716	-0.08
57	2,6		0.2	6.70	6.638	0.06
58	2,6		0.7	6.15	6.175	-0.03
59	2,6		0.4	6.40	6.164	0.24
60a	2,6		0.2	6.70	6.202	0.5
61	2,6		0.2	6.70	6.716	-0.02

[a] Test set compounds



C1, **C2** Centroids and * Atom considered for alignment

Figure 4.2.1: Template for alignment

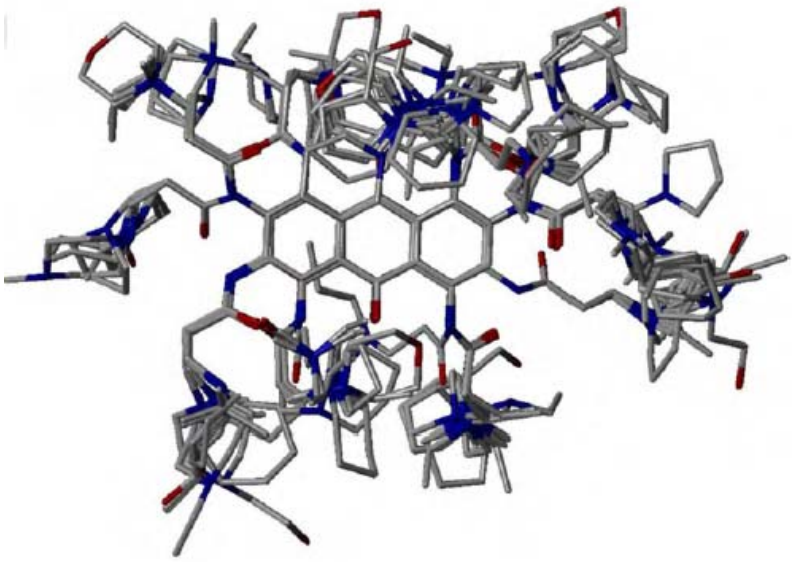
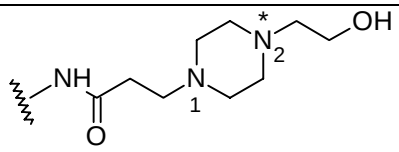
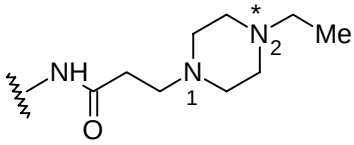
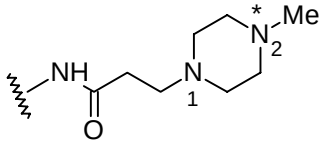


Figure 4.2.2: Alignment of all compounds

Table 4.2.2 *p*Ka values of nitrogens of piperazine ring system in compounds

Compound No	Anthraquinone/Acridone	Substitution	*p*Ka 1	*p*Ka 2
11	2,6-Disubstituted anthraquinone		3.49	7.75
50	3,6-Disubstituted acridone		3.49	8.61
54	2,7-Disubstituted acridone		3.50	7.59

* Nitrogens used for protonation strategy in the study

4.2.3 *p*Ka value calculation

As there were structures (compounds **11**, **50**, **54**) with two nitrogen atoms in piperazine ring, the decision about protonating the right nitrogen was taken on the basis of *p*Ka values. Table 4.2.2 contains structures and *p*Ka values of the studied compounds. Pallas 3.7.1.1 software from CompuDrug International Inc. [216] was used to calculate *p*Ka values for side chain substituent piperazine (Table 4.2.2). This program provides an important resource for predicting *p*Ka values before synthesis in QSAR studies.

4.2.4 CoMSIA results

A representative set of 61 anthraquinone and acridone derivatives was extracted from the literature [146, 155, 162, 167]. Great care was taken while building the structures of compounds to determine physiological ionization state and to check *in situ* protonation sites present on ligands. CoMSIA models were developed using individual single fields and also various fields in combination. Qualities of the models so developed were gauzed by the statistics obtained for each model. Data of CoMSIA models using different fields for Conf1 and Conf2 is presented in Tables 4.2.3 and 4.2.4, respectively. It was observed that CoMSIA model for Conf1 (Table 4.2.3), involving a combination of hydrophobic, acceptor and donor (HAD) fields, played

Table 4.2.3 Summery of CoMSIA for Conf1

Parameter	**HAD**	SED	SEA	SHE	EHD	SEDA	SEHD	SEHA	SEHDA
r^2_{cv}	**0.489**	0.403	0.336	0.347	0.425	0.420	0.404	0.329	0.410
ONC	**5**	4	5	4	4	4	4	4	4
SEP	**0.425**	0.433	0.456	0.453	0.425	0.427	0.433	0.459	0.430
r^2_{ncv}	**0.923**	0.785	0.846	0.822	0.832	0.860	0.821	0.855	0.861
SEE	**0.157**	0.260	0.220	0.237	0.230	0.210	0.237	0.213	0.209
F-Value	**103.221**	40.25	60.256	50.713	54.517	67.353	50.386	64.994	67.999
$P_{r2=0}$	**0**	0	0	0	0	0	0	0	0
r^2_{pred}	**0.721**	0.657	0.789	0.633	0.665	0.721	0.662	0.666	0.701
r^2_{bs}	**0.934**	-	-	-		-	-	-	-
SD_{bs}	**0.020**	-	-	-		-	-	-	-
				Contribution (fraction) (%)					
S	-	20.4	23.0	20.7	-	15.4	14.4	16.0	11.8
E	-	38.6	42.3	41.9	33.4	27.9	28.5	30.4	22.2
H	**31.7**	-	-	37.4	31.2	-	26.3	27.4	20.6
D	**40.1**	41.0	-	-	35.4	31.4	30.7	-	25.1
A	**28.2**	-	34.8	-	-	25.2	-	26.2	20.3

r^2_{cv}- cross-validated correlation coefficient, ONC- optimum number of components from PLS analysis, SEP-standard error of prediction, r^2_{ncv}- non-crossvalidated correlation coefficient, SEE-standard error of estimate, *F*- Fischer ratio, S- steric, E- electrostatic, H- hydrophobic, D-donor, A-acceptor, r^2_{pred}- predictive correlation coefficient, r^2_{bs}- bootstrapping correlation, SD_b-bootstrapping standard deviation.

significant role in the prediction of biological activity and offered a highly predictive 3D-model for substituted anthraquinone and acridone derivatives.

Table 4.2.4 Summery of CoMSIA for Conf2

Parameter	HAD	SED	SEA	SHE	EHD	SEDA	SEHD	SEHA	SEHDA
r^2_{cv}	0.302	0.329	0.338	0.256	0.356	0.294	0.273	0.307	0.290
ONC	2	4	2	2	2	2	2	2	2
SEP	0.457	0.459	0.446	0.473	0.440	0.460	0.467	0.456	0.462
r^2_{ncv}	0.663	0.737	0.627	0.607	0.632	0.629	0.620	0.648	0.644
SEE	0.318	0.287	0.335	0.344	0.332	0.334	0.338	0.325	0.327
F-Value	45.247	30.776	38.649	35.455	39.517	39.018	37.479	42.400	41.644
$P_{r2=0}$	0	0	0	0	0	0	0	0	0
r^2_{pred}	0.649	0.645	0.513	0.606	0.642	0.573	0.621	0.588	0.610
r^2_{bs}	-	-	-	-		-	-	-	-
SD_{bs}	-	-	-	-		-	-	-	-
				Contribution (fraction) (%)					
S	-	21.2	22.1	22.8	-	15.8	16.2	15.4	12.1
E	-	33.9	37.1	36.3	30.7	26.1	25.6	25.6	19.9
H	34.0	-	-	40.9	34.8	-	29.0	28.8	22.4
D	33.4	44.9	-	-	34.5	29.5	29.3	-	22.9
A	32.6	-	40.8	-	-	28.5	-	30.2	22.7

r^2_{cv}- cross-validated correlation coefficient, ONC- optimum number of components from PLS analysis, SEP-standard error of prediction, r^2_{ncv}- non-crossvalidated correlation coefficient, SEE-standard error of estimate, *F*-Fischer ratio, S- steric, E- electrostatic, H- hydrophobic, D-donor, A-acceptor, r^2_{pred}- predictive correlation coefficient, r^2_{bs}- bootstrapping correlation, SD_{bs}- bootstrapping standard deviation.

For Conf1 an excellent value of 0.721 (test set) for r^2 prediction and 0.489 for r^2 cross-validation with 5 optimum numbers of components were obtained for this model. A good r^2_{ncv} 0.923 (training set) was observed for internal prediction of the model. The contribution of hydrophobic, acceptor and donor fields of this model were 31.7%, 28.2%, and 40.1%, respectively. CoMSIA PLS analysis for Conf1 (Table 4.2.3) using hydrophobic, acceptor, and donor descriptors was found to be statistically acceptable. However, CoMSIA models for Conf2 (Table 4.2.4) was statistically inferior to models developed for Conf1 (Table 4.2.3). It was observed from the statistical quality of two different conformational models that Conf1 afforded the bioactive conformation for the development of predictive CoMSIA model for anthraquinones and acridones as telomerase inhibitors.

Above findings indicate that hydrophobic, hydrogen bond acceptor and donor fields can be more favorable for G-quadruplex stabilization in case of anthraquinone and acridone derivatives, and they could make an important contribution to the

binding energy of association. The graphs of observed vs. predicted activities for training and test set molecules from the best CoMSIA model (hydrophobic, hydrogen bond acceptor and hydrogen bond donor descriptors) are shown in Figure 4.2. 3

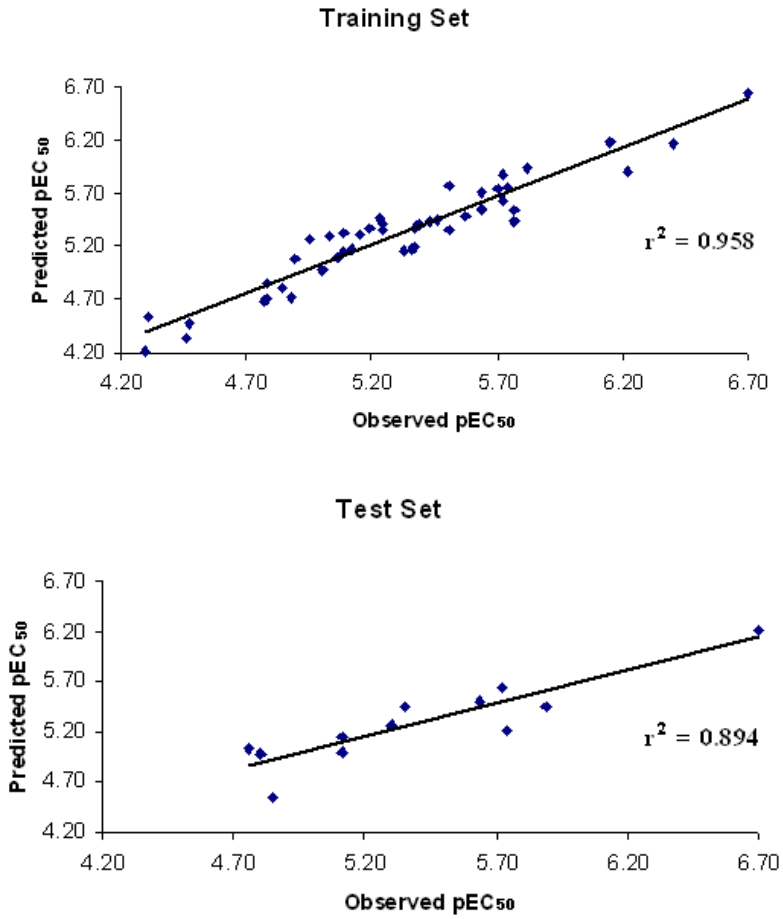


Figure 4.2.3: Graph of observed vs. predicted activities for training and rest set molecules from the best predicitive CoMSIA model (HAD)

4.2.5 Validation of best CoMSIA model (Conf1)

4.2.5.1 Internal validation (goodness of fit)

In model development we examined the internal predictive power of the models and their ability to reproduce biological activities of the compounds for the training set. The computed telomerase inhibitory activity from CoMSIA model showed a good correlation (r^2_{ncv} 0.923) with experimental telomerase inhibitory activity.

4.2.5.2 Fischer statistics (*F*-test)

The model is supposed to be good if the *F*-test is above a threshold value, i.e. tabulated value. The larger value of *F* indicates that the developed QSAR model is reliable and robust. The *F*-value for the best CoMSIA model was 103.221 [F_{05} (5, 60) = 2.3688 (Tab)] at 95% confidence level, which suggests that the model is statistically significant.

4.2.5.3 *Y*-Randomization test

In this test, the dependent variable is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The process is repeated several times, and the poor values of r^2 and q^2 for the new models ensure the robustness of the original QSAR models [38–41]. In the present case, 10 random trials were run for best CoMSIA model (HAD). None of the random trials could match the original model. The non-cross-validated and cross-validated r^2 for the model from randomized data was found in the range of 0.182 – 0.238 and −0.581 to −0.143, respectively. The results of randomization test have been presented in Table 4.2.5.

Table 4.2.5 Randomization test results for CoMSIA model (HAD)

Parameter	**Value**
No. of random trials	10
No. of trials with r^2_{ncv} and r^2_{cv} greater than non-random trial	0
No. of trials with r^2_{ncv} and r^2_{cv} lesser than non-random trials	10
Range of r^2_{ncv} from randomized data	0.182 to 0.238
Range of r^2_{cv} from randomized data	-0.581 to -0.143

4.2.5.4 Bootstrapping

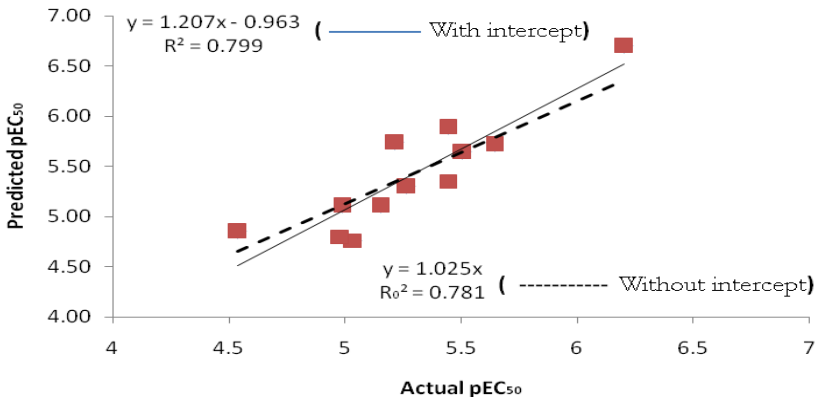
Further, to establish strength and robustness of the model, bootstrapping [217, 218] (100 runs) was performed. A statistical confidence was determined by the standard deviation of the statistics obtained from bootstrapping. A bootstrapped value (Table 4.2.3) (r^2_{bs} of 0.934) and bootstrapped standard deviation (SD_{bs} of 0.020) are obtained for CoMSIA (HAD) model. The results indicate that a reliable model has been successfully developed.

4.2.5.5 External validation

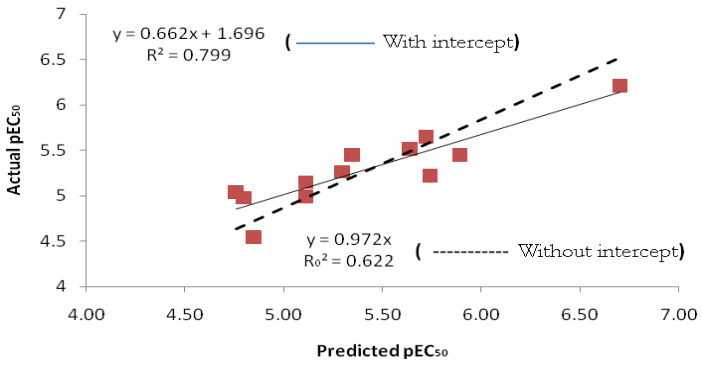
LOO cross-validation alone is not a reliable indicator of the prediction capability of QSAR model for two well-known reasons. LOO tends to reflect more of model robustness than its prediction power and LOO yields increasingly optimistic results with increasing data size (simply because removing one compound from a large data set will generally result in little change if any) [22]. Moreover, LOO cross-validation does not show the correct symptomatic trend with increasing number of observations [219]. Certainly, the high value of LOO cross-validation appears to be the necessary but not the sufficient condition for the models to have good predictive power. Golbraikh and Tropsha have emphasized that the actual predictive ability of a QSAR model can only be estimated using an external test set of compounds that were not used for building the model [39].

The value of r^2_{pred} was calculated for the test set and gave the best results for CoMSIA model with hydrophobic, acceptor and donor as descriptors with a value of 0.721. Thus, the CoMSIA model (HAD) displays higher predictivity both in cross-validation and in the prediction of the test compounds.

Besides this, the regression plots of actual vs. predicted activities and predicted vs. actual activities were constructed (Figure 4.2.4a,b) to apply rigorous validation parameters reported by Tropsha et al. that has been described in chapter 1, section 1.1.1.6.



(a)



(b)

Figure 4.2.4: External validation of CoMSIA model. (a) Actual vs. Predicted activities (b) Predicted vs. Actual activities, of Test Set compounds

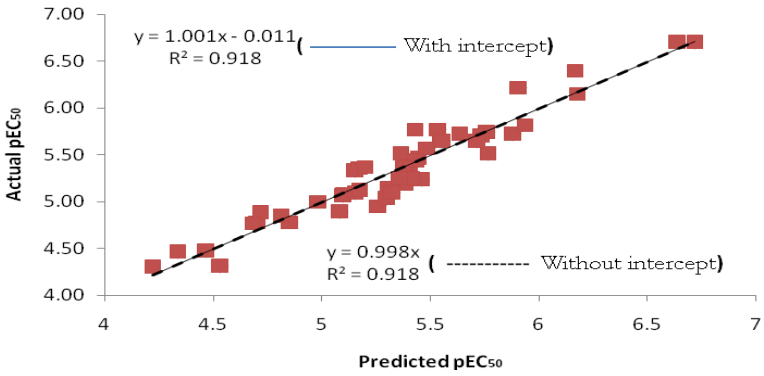


Figure 4.2.5: Internal Validation of best CoMSIA model. Predicted vs. Actual activities of Training set compounds

4.2.5.6 R^2_m as a validation criterion

Additionally, the model was validated with validation criterion R^2_m reported by Roy et.al [40, 41]. The R^2_m (Test) value penalizes a model for large differences between observed and predicted values of test set compounds. The CoMSIA model was found to be superior with R^2_m (Test) = 0.791. Since this value is based on limited number of test set compounds, to further confirm its reliability, R^2_m values were calculated for training set compounds using LOO predictions. The value of R^2_m (training) 0.918 confirms that the model built so is a reliable one. The results obtained are shown in Table 4.2.6. In case of best external prediction $R^2_m = R^2$ while in the worst case, $R^2_m = 0$. In present study for external test set compounds, the value of R^2_m 0.791 was found to be almost equal to R^2 value 0.799 and for training set compounds,

the value of R^2_m 0.918 was found to be exactly equal to R^2 value 0.918 (Figure 4.2.5), indicating that the derived model is a highly predictive one.

Table 4.2.6 Summary of validation for CoMSIA model

Statistics	CoMSIA model (HAD)
R^{2a}	0.799
R_0^{2b}	0.781
k^c	1.025
$(R^2 - R_0^2)/R^2$	0.221
R'^{2b}_0	0.622
k'^c	0.972
$(R^2 - R'^2_0)/R^2$	0.022
$R_m^2 = R^2 (1-\sqrt{\lvert R^2 - R_0^2 \rvert})$	0.791
Absolute sum of observed *p*IC_{50} (Test set)	64.97
Absolute sum of residuals (Test set)	2.73
% residual (Test Set)	4.20%
Absolute sum of observed *p*IC_{50} (Training set)	261.95
Absolute sum of residuals (Training set)	5.90
% residual (Training set)	2.25%

[a] Correlation coefficient derived from predicted *p*IC_{50} of test set compounds
[b] Correlation coefficient for the regression line passing through origin for actual vs. predicted and predicted vs. actual activity, respectively
[c] Slope for regression line passing through origin obtained from actual vs. predicted and predicted vs. actual activity, respectively

4.2.7 Interpretation of the best CoMSIA model

The contour plot of the CoMSIA hydrophobic, H-bond acceptor, and H-bond donor fields (stdev*coeff) are presented in Figure 4.2.6. Favored and disfavored levels fixed at 80% and 20%, respectively, were used. The most active compound (**61**) ($^{tel}EC_{50}$ 0.2) from the series is shown inside the fields.

Hydrophobic maps shown in Figure 4.2.6a indicate that the lipophilic favorable yellow region is found surrounding the aromatic planar chromophore and alkyl side chains. This suggests that lipophilicity of this portion of the molecule is an important factor for the activity.

The white region at the end of the side chains indicates that hydrophilic substitution in this zone favors the inhibitory activity. Aromatic chromophore of anthraquinone and acridone derivatives fits well in favorable yellow hydrophobic region. This indicates the importance of lipophilicity for telomerase inhibitory

activity. Alkylamido side chains of compounds (**57**) ($^{tel}EC_{50}$ 0.2), (**60**) ($^{tel}EC_{50}$ 0.2) and (**61**) ($^{tel}EC_{50}$ 0.2) are perfectly entrenched in favorable yellow region, hence, found to be the most potent in the series. Side chains of compound (**1**) ($^{tel}EC_{50}$ 4.30) were directed away from the favorable hydrophobic region. This could be one of the

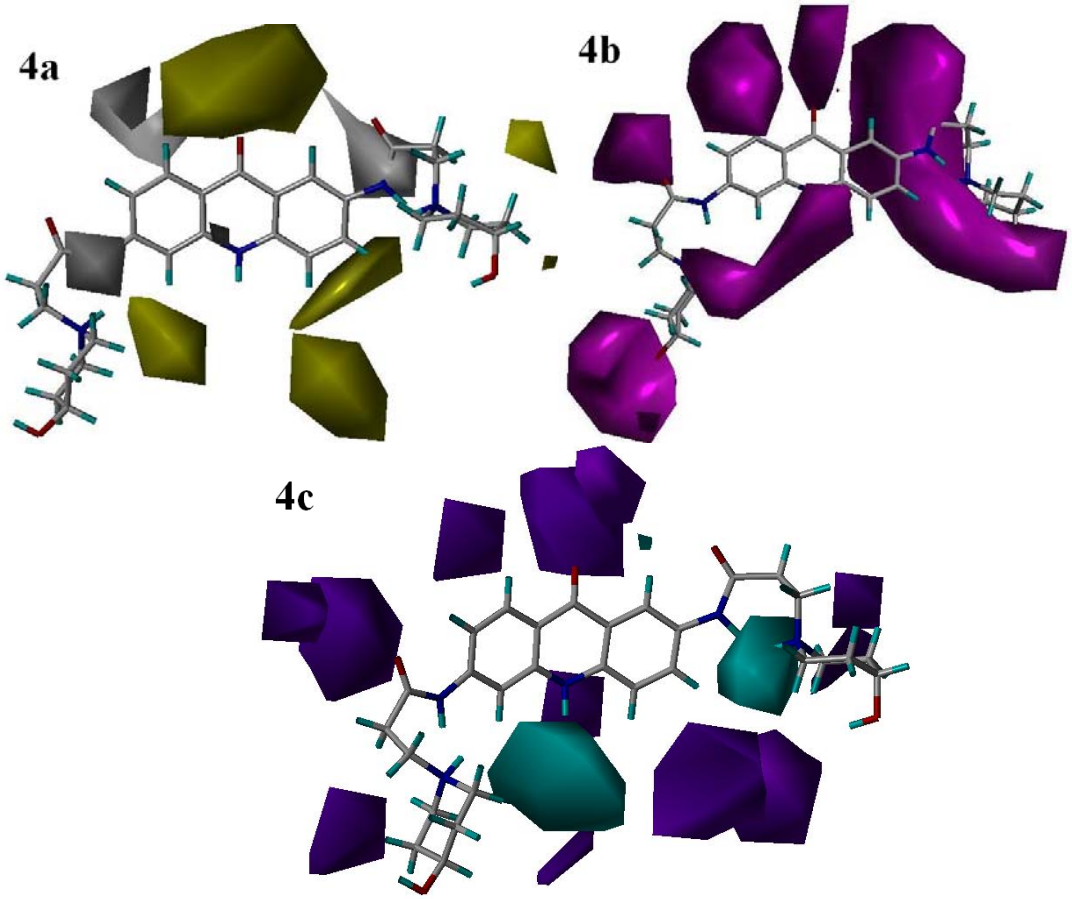


Figure 4.2.6: CoMSIA stdev*coeff contour maps. Compound (**61**) is shown inside the field. (4a) hydrophobic field, (4b) H-bond acceptor field and (4c) H-bond donor field. Yellow and white contours indicate regions where hydrophobic groups favored and disfavored the activity, respectively. Magenta contours represent areas where H-bond acceptors are favored. Cyan and purple contours represent areas where H-bond donors are favored and disfavored, respectively

reasons why 1,4-substituted anthraquinone derivatives are less potent than the ones with 2,6-substitution. In case of compounds (**11**) ($^{tel}EC_{50}$ 16.5), (**50**) ($^{tel}EC_{50}$ 1.9) and (**54**) ($^{tel}EC_{50}$ 1.9), the well defined white region is associated with nitrogen atom in piperazine moiety and indicates that more hydrophilic piperazine moiety is preferable. This result explains why nitrogen atoms belonging to the piperazine has positive contribution to telomerase inhibitory activity in 3D-QSAR CoMSIA analysis.

CoMSIA H-bond acceptor field (Figure 4.2.6b) shows the regions where H-bond donors in the binding site are predicted to enhance (magenta) the binding of ligands to G-quadruplex.

According to Lewis definition of acidity and basicity, where acids accept and bases donate electron density, the H-bond acceptor is a Lewis base since it donates electron density. It is interesting to observe that the nitrogen atom of the acridone ring system functions as a hydrogen bond acceptor as it carries non-bonding pair of electrons. There is prominent magenta contour surrounding the central nitrogen atom of acridone, which indicates that an H-bond acceptor is favored at this position. In anthraquinons, the nitrogen of acridone is replaced by a ketonic oxygen. This oxygen is also indicated as a favorable region for H-bond acceptor group along with the amide keto oxygens of the side chains.

It may be inferred from the above observations that the central nitrogen atom of acridone ring should have more electron density, so that it can act as H-bond acceptor and the substitutions should be made accordingly at this position without diminishing its relative basicity (i.e., strength) as hydrogen bond acceptor group. Molecular orbital considerations suggest that the acridone chromophoric group is not a fully delocalized system. Hence, central ring acridone nitrogen atom does not contribute to effective π–π staking interaction with adjacent base pairs, as seen in acridine derivatives. However, it has not been made very clear whether its interaction with the G-quadruplex is ionic or not [162]. The Current study reveals that the acridone ring nitrogen remains non-protonated and acts as H-bond acceptor group in the stacking process. Compound (**61**) is most active as its oxygen atom of carbonyl group at position-9 and central ring nitrogen atom of acridone chromophore fall exactly in the region of favorable H-bond acceptor contours. H-bond donor alkyl side chains of compound (**1**) ($^{tel}EC_{50}$ 4.30) are oriented away from favorable H-bond donor contour region. This is the reason it has got lowest activity in the series.

The graphical interpretation of the field's contributions of the H-bond donor properties is shown in Figure 4.2.6c. A cyan colored H-bond donor region surrounds the piperidine ring nitrogen atoms showing the favorable regions for H-bond donor fields. A prominent purple disfavorable region was found in the vicinity of oxygen atom of carbonyl group present on C9 position of acridone chromophore. In

compound (**1**) ($^{tel}EC_{50}$ 4.3) NH of amide groups substituted at C1 and C4 position of anthraquinone were perfectly entrenched towards H-bond donor disfavorable region along with oxygen atom of amide protruding towards H-bond donor favorable region hence it showed poor telomerase inhibitory activity. This explains the involvement of amide functionality present in anthraquinone derivatives for stabilization of G-quadruplex. In case of compound (**41**) ($^{tel}EC_{50}$ 8.1), NH of amide groups in alkyl side chains of acridone ring are oriented away from the favorable H-bond donor region. This could be the reason for its lower activity than other 3,6-disubstituted acridone derivatives.

It is remarkable to note that the nitrogens on the basic side chains act as H-bond donors after protonation. Compounds with simple ammonium groups (**14**, **17**, **18**, **19**, **25**, **27**, **33**, **34**, and **35**) possess higher biological activity over the corresponding quarternary ammonium compounds (**12**, **20**, **23**, **24**, **28**, **32**, **37**, **38**, and **40**) because the quarternary ammonium group can not act as H-bond donor.

4.3 Molecular modeling studies of some triazine derivatives as G-quadruplex binders for inhibition of telomerase enzyme

4.3.1 Data set used

The data set was chosen from a series of 34 triazine derivatives as telomerase inhibitors reported by Mailliet et al. [220]. The telomerase inhibitory activity of compounds is reported as the concentration capable of inhibiting 50% of the enzyme activity (IC_{50}) in micromolar (μM) range. The selected compounds cover a wide range of biological activities (0.041 to 3.9 μM) and possess diversity in structural features. The IC_{50} (μM) values were converted into molar units and then expressed in negative logarithmic units (pIC_{50}). For carrying out 3D-QSAR CoMSIA study, 31 compounds (Table 4.3.1) were selected from the series to remove redundancy in terms of chemical structures. The compounds were divided into training (22) and test (9) compounds keeping in mind that the models should be tested on a sufficiently large test set (25–33% of total) to establish a reliable QSAR model [221]. In the training set, most potent, moderately active and lowly active compounds were included to spread the activity range. The test set compounds were selected in such a manner that at least one structural analog remained in the training set.

4.3.2 Selection of template and molecular alignment

In this study, simulated annealing-based [222] global minimum energy structure of the most active compound (**11**) was used as a template. It is known that the conformation representing the global minimum of the ligand may not bind to the

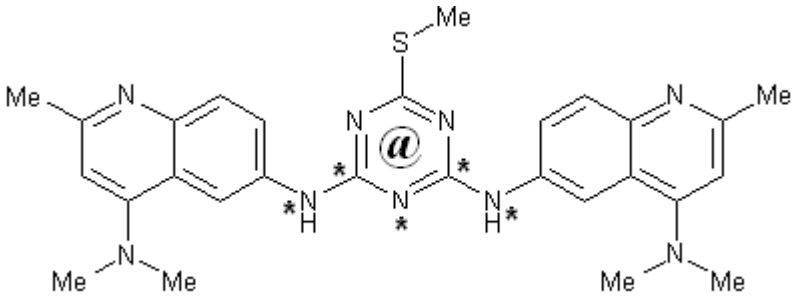


Figure 4.3.1: Template and Atoms considered for the alignment

receptor and some degree of torsional freedom is required for the drug to adopt to the receptor-binding site to yield a bioactive conformation [223]. But in the present study, structures possess sufficient rigidity to limit the possible active conformations and the simulated annealing-based global minimum energy conformation is an excellent starting point to identify possible bioactive conformation for the chosen template. The most active compound (**11**) was constructed in SYBYL. A preliminary minimization was performed to remove close atom contacts using standard Tripos force field with Gasteiger-Huckel charges, and repeated minimization was performed using steepest-descent and conjugate gradient methods until the root-mean-square (rms) deviation of 0.001 kcal/mol was achieved. The structure was next subjected to molecular dynamic simulation, where it was heated to 700 K for 1 ps and annealed slowly to 300 K in steps of 100 K for 1 ps at each temperature. The exponential annealing function was used and 10 such cycles were run. The least energy conformation was determined and subsequently subjected to minimization with the same criteria as mentioned above, and used as a template to build structures of all other training and test set compounds. Each built structure was aligned to the template compound **11** by atom and centroid-based alignment. The atoms and centroid considered for the alignment are marked with an asterisk (*) and (@) respectively in Figure 4.3.1.

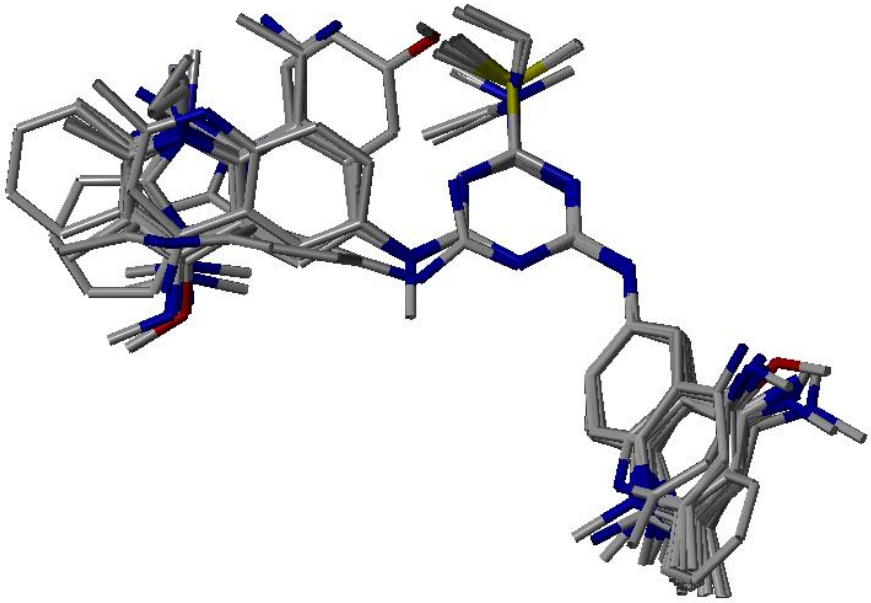


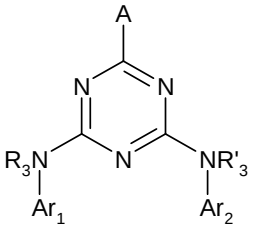
Figure 4.3.2: Alignment of all compounds

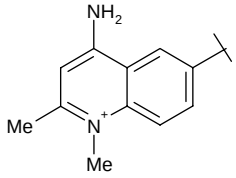
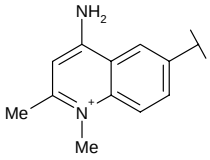
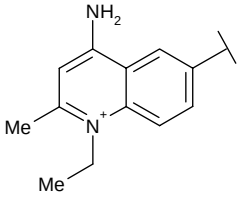
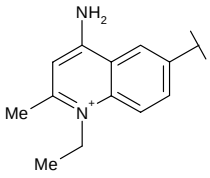
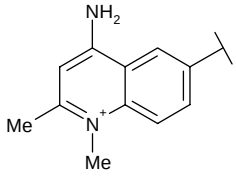
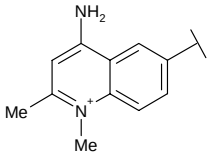
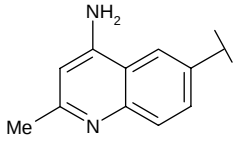
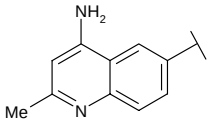
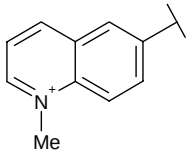
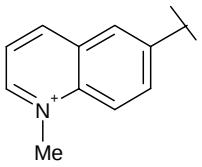
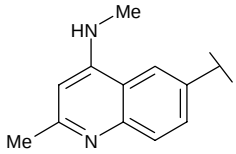
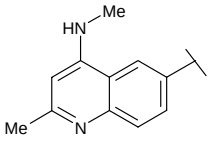
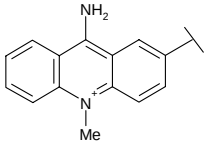
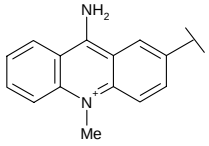
The superimposition of all the compounds on template (compound **11**) is shown in Figure 4.3.2.

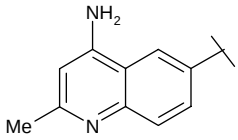
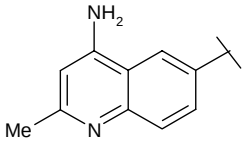
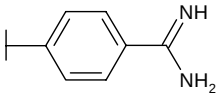
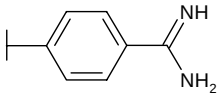
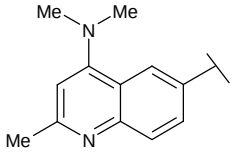
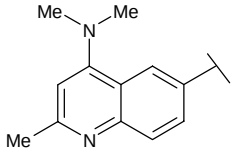
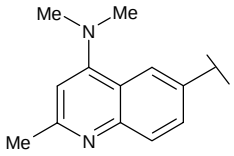
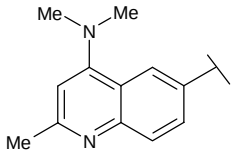
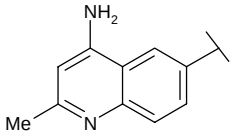
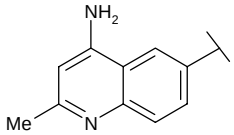
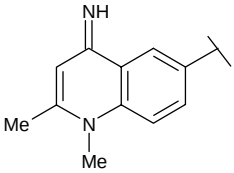
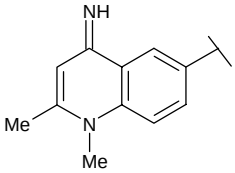
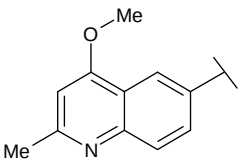
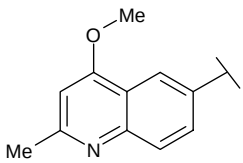
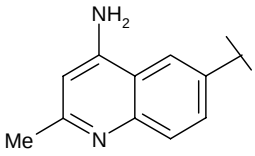
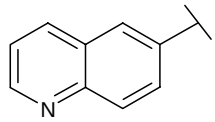
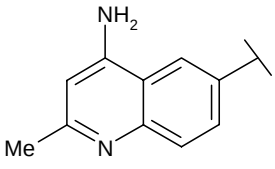
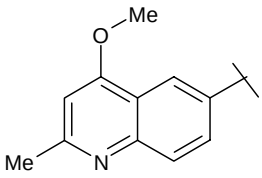
4.3.3 CoMSIA results

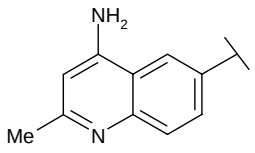
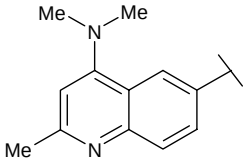
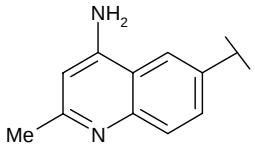
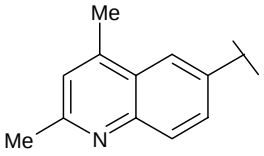
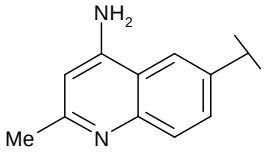
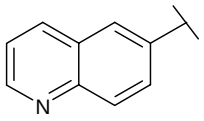
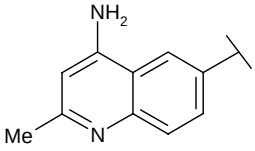
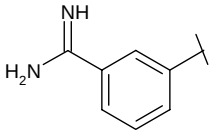
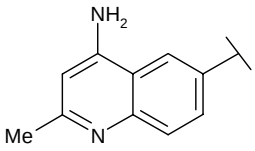
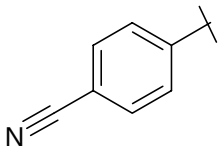
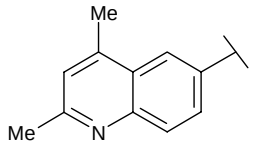
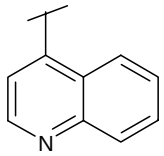
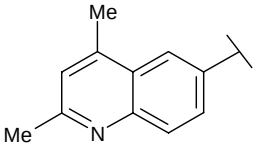
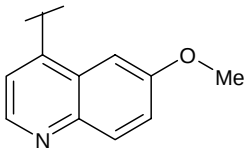
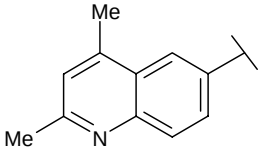
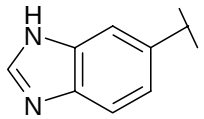
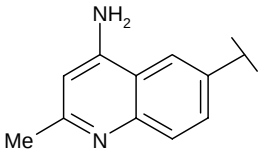
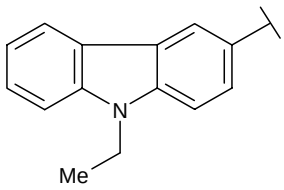
Initially, CoMSIA models were constructed with 22 compounds in training set using individual single fields and also various fields in combination. The qualities of the models so developed were judged by the statistical quality for each model. According to this analysis, it was detected that two compounds (**10** and **12**) presented large residuals (greater than a unit value) in all models. Although, there is no experimental evidence to explain the 'outlier' behavior of these compounds, apparently in case of compound (**10**) it was the only compound in the series with 4-chlorosubstituted triazine amongst the training set molecules. This structural uniqueness in the training set molecules might be the reason for its poor prediction. In case of compound (**12**), it might be a result of a random experimental error that could be significant when analyzing the large datasets [212].

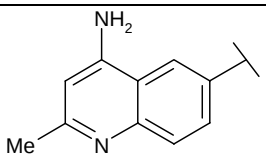
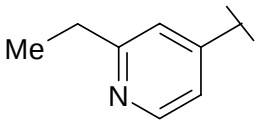
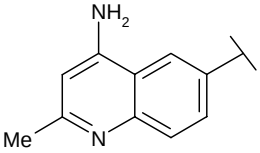
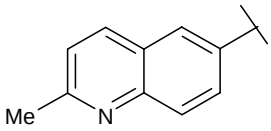
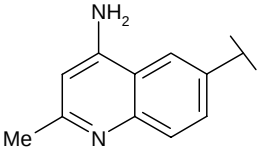
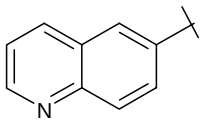
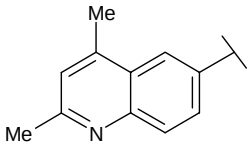
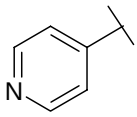
Table 4.3.1 Molecular structures and telomerase inhibitory activity of triazine derivatives

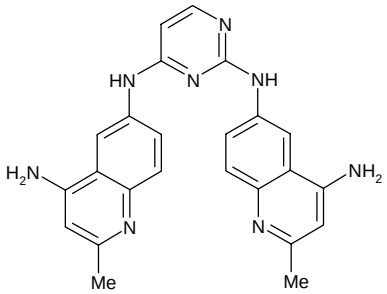
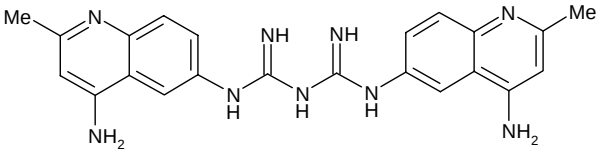


Compound	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (μM)	Actual pIC_{50}	Predicted pIC_{50}	Residuals
1[a]	$-NH_2$		H		H	0.25	6.60	6.84	-0.24
2[a]	$-NH_2$		H		H	0.056	7.25	7.03	0.22
3[a]t	$-N(Me)_2$		H		H	0.22	6.66	6.33	0.33
4[a]t	-NHMe		H		H	0.51	6.29	6.69	-0.40
5[a]t	$-NH_2$		H		H	0.13	6.89	6.50	0.39
6[a]	-NHMe		H		H	0.3	6.52	6.49	0.03
7	$-NH_2$		H		H	0.89	6.05	6.10	-0.05

Compd.	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (μM)	Actual pIC_{50}	Predicted pIC_{50}	Residuals
8[a]t	- NH_2		H		H	0.051	7.29	6.78	0.51
9t	- NH_2		H		H	0.74	6.13	6.21	-0.08
10*	-Cl		H		H	3.0	5.52	-	-
11[a]	- SMe		H		H	0.041	7.39	7.35	0.04
12*	- SMe		H		H	0.049	7.31	-	-
13	- SMe		H		H	0.95	6.02	5.99	0.03
14	- SMe		H		H	3.9	5.41	5.35	0.06
15t	- SMe		H		Me	3.5	5.46	5.59	-0.13
16	- SMe		H		H	2.3	5.64	5.68	-0.04

Compd.	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (μM)	Actual pIC_{50}	Predicted pIC_{50}	Residuals
17[a]	-SMe		H		H	0.34	6.47	6.54	-0.07
18	-SMe		H		H	2.2	5.66	5.56	0.10
19t	-SMe		H		H	3.4	5.47	5.79	-0.32
20	-SMe		H		H	2.6	5.59	5.64	-0.05
21	-SMe		H		H	3.3	5.48	5.51	-0.03
22	-SMe		H		H	3.4	5.47	5.47	0.00
23	-SMe		H		H	3.4	5.47	5.39	0.08
24t	-SMe		H		H	3.3	5.48	6.03	-0.55
25t	-SMe		H		H	3.3	5.48	5.42	0.06

Compd.	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (μM)	Actual pIC_{50}	Predicted pIC_{50}	Residuals
26	-SMe		H		H	2.8	5.55	5.56	-0.01
27	-SMe		H		H	3.4	5.47	5.55	-0.08
28	$N(Et)_2$		H		H	1.0	6.00	5.98	0.02
29	$N(Et)_2$		H		H	2.5	5.60	5.67	-0.07

Compd.	Compound	IC_{50}	Actual pIC_{50}	Predicted pIC_{50}	Residuals
30[a]		0.49	6.31	6.31	0.00
31		2.0	5.70	5.64	0.06

[t] Test set compounds, [a] Compounds considered for development of Pharmacophore, *Identified outliers

Table 4.3.2 Summary of CoMSIA models

No.	Models	Fields	r^2_{cv}	SEP	ONC	r^2_{ncv}	SEE	*F*-value	r^2_{pred}	r^2_{bs}	SD
1	S		0.100	0.581	1	0.648	0.364	33.170	-	-	-
2	E		0.202	0.548	1	0.453	0.453	14.936	-	-	-
3	H		0.445	518.000	5	0.978	0.102	126.182	-	-	-
4	D		0.091	0.633	5	0.923	0.193	33.381	-	-	-
5	A		0.374	0.485	1	0.576	0.399	24.472	-	-	-
6	SE	S32.8; E67.2	0.347	0.562	5	0.970	0.121	89.858	-	-	-
7	SH	S34.5; H65.5	0.511	0.486	5	0.987	0.079	214.887	0.608		
8	SA	S42.4; A57.6	0.527	0.478	5	0.970	0.120	91.050	0.220	-	-
9	SD	S39.3; D60.7	0.181	0.629	5	0.967	0.127	80.947	0.614	-	-
10	EH	E47.9; H52.1	0.357	0.557	5	0.972	0.117	96.264	0.480	-	-
11	EA	E63.3; A34.7	0.305	0.511	1	0.551	0.411	22.126	0.118	-	-
12	ED	E56.4; 43.6	0.274	0.592	5	0.933	0.180	38.961	-	-	-
13	HA	H58.6; A41.4	0.411	0.533	5	0.976	0.109	111.915	0.542	-	-
14	HD	H56.4; D43.6	0.609	0.435	5	0.971	0.119	92.665	0.695	-	-
15	SEA	S13.5; E56.5; A30.0	0.330	0.502	1	0.597	0.389	26.635	-	-	-
16	SED	S19.9; E46.6; D33.5	0.293	0.585	5	0.950	0.155	53.619	-	-	-
17	SHE	S21.6; H41.3; 37.1	0.355	0.558	5	0.976	0.108	112.276	0.475	-	-
18	EAD	E49.0; A26.0; D25.0	0.328	0.502	1	0.582	0.396	25.079	-	-	-
19	EAH	E48.9; A26.0; H25.2	0.303	0.512	1	0.512	0.382	28.286	-	-	-
20	ADH	A26.9; D32.9; H40.2	0.426	0.527	5	0.969	0.123	87.271	0.671	-	-
21	ADS	A40.7; D36.3; S23.0	0.360	0.520	3	0.876	0.229	37.574	-	-	-
22	SAH	S22.8; 30.6; H46.6	0.422	0.529	5	0.982	0.093	154.117	0.558	-	-
23	**SDH**	**S19.7; D33.6; H46.7**	**0.593**	**0.443**	**5**	**0.976**	**0.167**	**115.130**	**0.738**	**0.994**	**0.006**
24	SEAD	S15.0; E34.0; A26.6; D24.4	0.281	0.535	2	0.781	0.295	30.393	-	-	-
25	SEAH	S10.5; E43.8; A23.2; H22.5	0.298	0.514	1	0.647	0.364	32.984	-	-	-
26	SEDH	S14.0; E28.7; D25.0; H32.4	0.482	0.500	5	0.968	0.124	85.161	0.633		
27	EADH	E27.2; A19.8; D24.3; H28.7	0.380	0.547	5	0.968	0.125	83.737	0.602	-	-
28	ADHS	A22.8; D27.0; H35.3; S14.9	0.580	0.451	5	0.975	0.111	107.007	0.684	-	-
29	SEHAD	S11.1; E24.4; H26.0; A17.5; D21.0	0.352	0.560	5	0.976	0.108	112.733	0.620	-	-

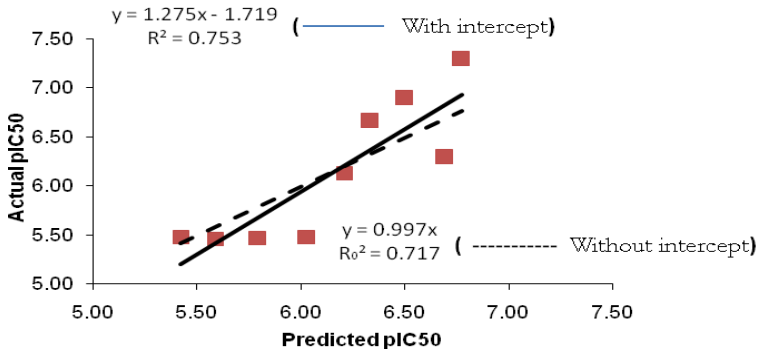
S-Steric; E- Electrostatic; H- Hydrophobic; D- Donor; A- Acceptor; r^2_{cv}- cross-validated correlation coefficient; SEP- standard error of prediction; ONC- optimum number of components; r^2_{ncv}- non-cross validated correlation coefficient; *F*- Fisher value; r^2_{pred}- predictive correlation coefficient; r^2_{bs}- bootstrapped correlation coefficient; SD-stand dev.

CoMSIA models with 20 compounds (excluding **10** and **12**) were constructed. Results showed a remarkable enhancement in statistical quality of each new model. The step-wise development of CoMSIA models using different fields has been shown

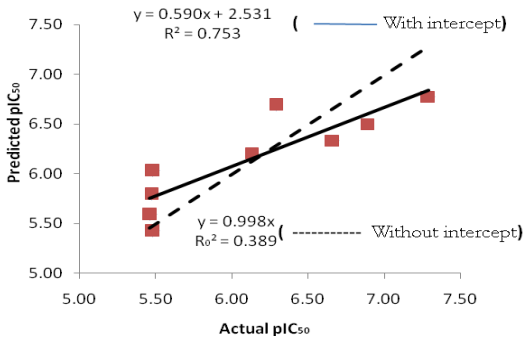
in Table 4.3.2. It was observed that CoMSIA model, involving a combination of steric, donor and hydrophobic (SDH) fields, played significant role in the prediction of telomerase inhibitory activity and offered a highly predictive 3D-model for triazine derivative. Experimental and predicted activities are presented in Table 4.3.1. An excellent value of 0.738 (test set) for r^2 prediction and 0.593 for r^2 cross-validation with 5 optimum numbers of components were obtained for this model. A good r^2_{ncv} 0.976 (training set) was observed for internal prediction of the model. The contribution of steric, donor and hydrophobic (SDH) fields of this model were 19.7%, 33.6%, and 46.7%, respectively. The above findings indicate that the steric, hydrophobic, and donor fields are more favorable for G-quadruplex stabilization in case of triazine derivatives and these three fields could effectively contribute in the binding energy of association.

4.3.3.1 External validation of best CoMSIA model (SDH)

The best CoMSIA model (SDH) was validated giving a special emphasis on statistical significance and predictive ability as external validation is a most crucial characteristic of any 3D-QSAR model. Generally, cross-validation is used as a reliable indicator of the prediction capability of QSAR model. However, models having higher cross-validated r^2 ($q^2 > 0.5$) are not always predictive. Hence, the only approach to gauze the true predictive power of a model is to test it on a sufficiently large set of compounds from an external test set. The predicted r^2 value for CoMSIA (SDH) model was found to be 0.738 (Table 4.3.2). Besides this, the regression plots of predicted vs. actual activities and actual vs. predicted activities were constructed (Figure 4.3.3a,b) to apply the rigorous validation criteria.



(a)



(b)

Figure 4.3.3: External validation of CoMSIA model (SDH). (a) Predicted vs. Actual activities (b) Actual vs. Predicted activities, of Test Set compounds

4.3.3.2 R^2_m as a validation criterion

Additionally, the model was validated with validation criterion R^2_m reported by Roy et.al [40, 41]. The R^2_m (Test) value penalizes a model for large differences between observed and predicted values of test set compounds. The CoMSIA model was found to be superior with R^2_m (Test) = 0.739. Since this value is based on limited number of test set compounds, to confirm further its reliability, R^2_m values were calculated for training set compounds using LOO predictions. The value of R^2_m (training) 0.976 confirms that

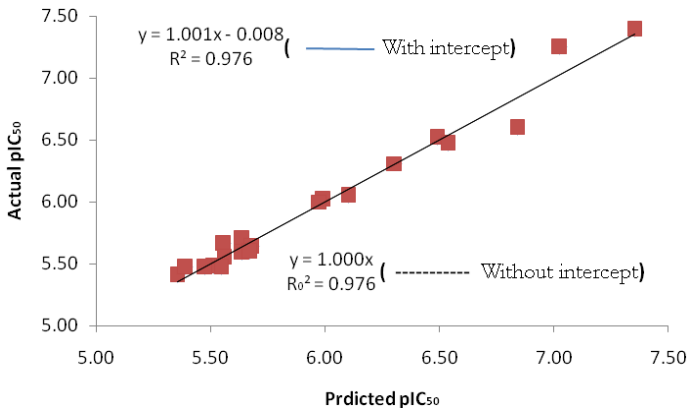


Figure: 4.3.4: Internal validation of best CoMSIA model Predicted vs. Actual activities of training set compounds

the model built so is a reliable one. The results obtained are shown in Table 4.3.3. In case of best external prediction $R^2_m = R^2$ while in the worst case, $R^2_m = 0$. In present study for external test set compounds, a value of R^2_m 0.739 was found to be almost equal to R^2 value 0.753 and for training set compound, a value of R^2_m 0.976 was found to be exactly equal to R^2 value 0.976 (Figure 4.3.4), indicating that the derived model is a highly predictive one.

Table 4.3.3 Summary of validation for CoMSIA model

Statistics	CoMSIA model (SDH)
R^{2a}	0.753
R_0^{2b}	0.389
k^{c}	0.998
$(R^2 - R_0^2)/R^2$	0.048
R'^{2b}_0	0.717
k'^{c}	0.997
$(R^2 - R'^2_0)/R^2$	0.483
$R_m^2 = R^2 (1-\sqrt{\lvert R^2 - R_0^2 \rvert})$	0.739
Absolute sum of observed *p*IC$_{50}$ (Test set)	55.15
Absolute sum of residuals (Test set)	2.77
% residual (Test Set)	5.02%
Absolute sum of observed *p*IC$_{50}$ (Training set)	119.65
Absolute sum of residuals (Training set)	1.48
% residual (Training set)	1.24%

[a] Correlation coefficient derived from predicted *p*IC$_{50}$ of test set compounds
[b] Correlation coefficient for the regression line passing through origin for actual vs. predicted and predicted vs. actual activity, respectively
[c] Slope for regression line passing through origin obtained from actual vs. predicted and predicted vs. actual activity, respectively

4.3.3.3 *Y*-Randomization test

In *Y*-randomization test, 10 random trials were run for the best CoMSIA model (SDH). None of the random trials could match the original model. The non-cross-validated and cross-validated r^2 for the model from randomized data was found in the range of 0.201 to 0.423 and -0.349 to 0.191, respectively. The results of randomization test have been presented in Table 4.3.4**.**

Table 4.3.4 Randomization test results for CoMSIA model (SDH)

Parameter	Value
No. of random trials	10
No. of trials with r^2_{ncv} and r^2_{cv} greater than non-random trial	0
No. of trials with r^2_{ncv} and r^2_{cv} lesser than non-random trials	10
Range of r^2_{ncv} from randomized data	0.201 to 0.423
Range of r^2_{cv} from randomized data	-0.349 to 0.191

4.3.3.4 Percentage of residual values (% Residual)

In the present study a new validation parameter was analyzed and investigated to check the reliability of 3D-QSAR model. The percentage residual value is mainly governed by the sum of actual activities of test set compound and absolute sum of differences between actual and predicted values of the test set compounds. The percentage of residuals of differences between actual and predicted values of test set compounds was calculated as a percentage of the summed up observed activities of the test compounds. The % residual values of the differences between actual and predicted values of the training and test set compounds were found to be 1.24% and 5.02 % respectively. The results are shown in Table 4.3.3

4.3.3.5 Fischer statistics (*F*-test)

Fischer value (*F*) is the ratio between explained and unexplained variance for a given number of degree of freedom. The model is supposed to be good if the *F*-test is above a threshold value, i.e. tabulated value. The larger the value of *F*, the greater is the probability that the QSAR model is a realistic one. The *F*-value for the best CoMSIA model was 115.130 [F_{05} (5, 30) = 3.6990 (Tab)] at 99% confidence level, which suggests that the model is statistically significant.

4.3.4 Pharmacophore elucidation

The development of pharmacophore model was carried out using web-based server called PharmaGist [64]. Pharmacophore models were derived by using 10 active (pIC$_{50}$ $\geq$ 6.00) compounds from the series Table 4.3.1. The results were achieved using a set of standard default parameters. All the compounds are flexibly aligned on the pharmacophore and showed a satisfactory superimposition of the pharmacophoric points (Figure 4.3.5). The highest-scoring pharmacophore hypothesis was considered as best pharmacophore model for triazine derivatives displayed in

Figure 4.3.6. It comprises of three aromatic, one hydrophobic, two donors and two acceptor features. These findings along with the recent molecular modeling study [209] suggest that hydrophobicity is a dominant feature for G-quadruplex mediating telomerase inhibitors. The distances and angles among the pharmacophoric features are depicted in Figure 4.3.6a,b respectively.

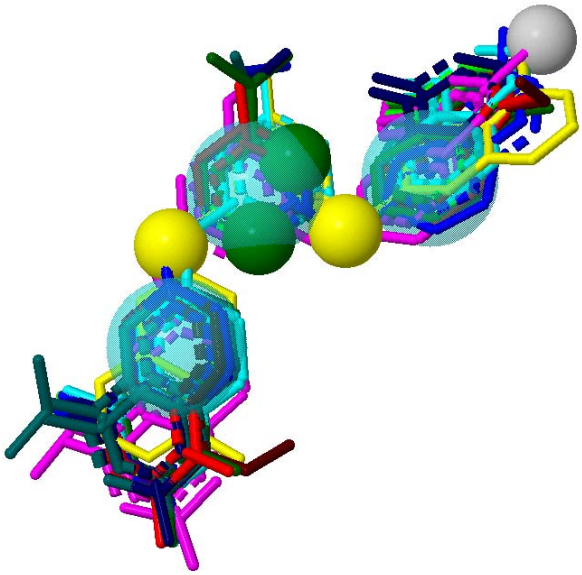


Figure 4.3.5: Alignment of ligands on Pharmacophoric points

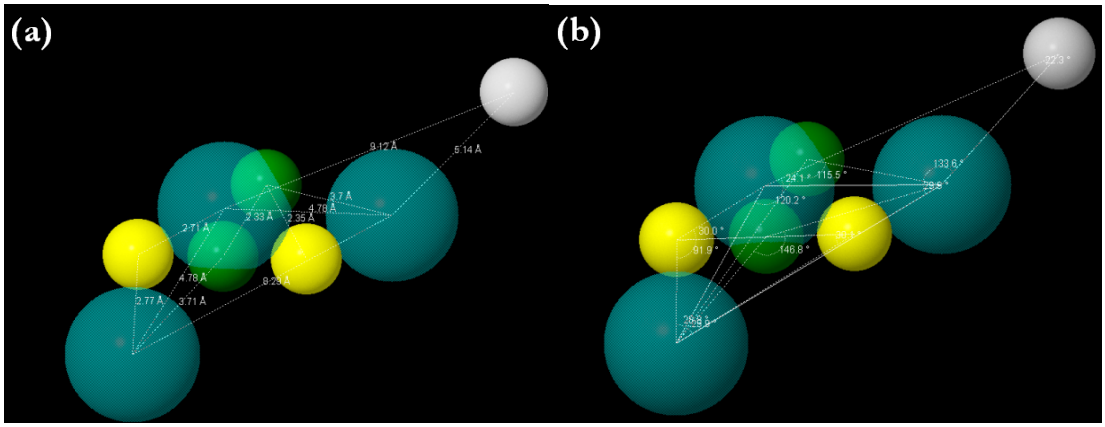


Figure 4.3.6: Geometry of the pharmacophore. Cyan spheres aromatic feature, yellow spheres donor features, green sphere acceptor features and white sphere hydrophobic features. (a) Distances; (b) angles.

In the present study both 3D-QSAR and pharmacophore approaches have been employed to develop meaningful computational models for designing of more

potent telomerase inhibitors. The reliability of pharmacophore model was judged by comparing inter-feature distances obtained from both of the approaches.

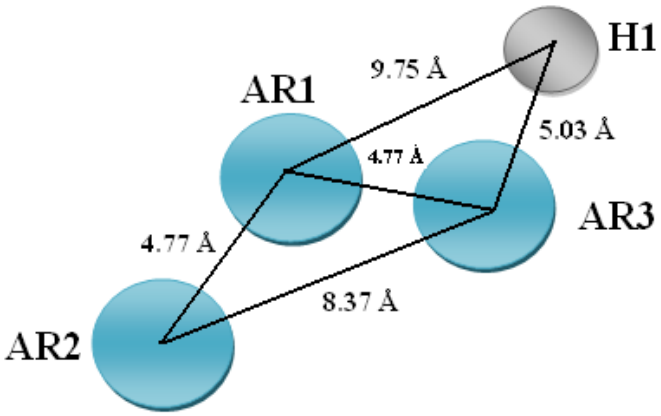


Figure 4.3.7: Distances obtained from 3D-QSAR CoMSIA study

The inter-feature distances obtained from 3D-QSAR CoMSIA study are shown in Figure 4.3.7 and the distance comparison of both the approaches is shown in Table 4.3.5.

Table 4.3.5 Comparison of pharmacophoric and 3D-QSAR contour maps distances

	Pharmacophore	**3D-QSAR**
Features	**Distance**	**Distance**
AR1-AR2	4.78 Å	4.77 Å
AR1-AR3	4.78 Å	4.77 Å
AR2-AR3	8.29 Å	8.37 Å
AR3-H1	5.14 Å	5.03 Å
AR1-H1	9.12 Å	9.75 Å

From the above results it was observed that inter-feature distances obtained by 3D-QSAR and pharmacophore modeling approaches are almost equal. Thus, the developed 3D-QSAR and the pharmacophore models are found to be same and could be very useful for lead optimization.

4.3.5 Interpretation of the best CoMSIA model

The contour plots of the CoMSIA steric, H-bond donor and hydrophobic fields (stdev*coeff) are presented in Figure 4.3.8. Favored and disfavored levels fixed at 80% and 20%, respectively, were used. The most active compound (**11**) ($^{tel}EC_{50}$ 0.041 μM) from the series is shown inside the fields. The green contours (favored)

represent regions of high steric tolerance, while the yellow contours (disfavored) represent regions of low steric bulk tolerance.

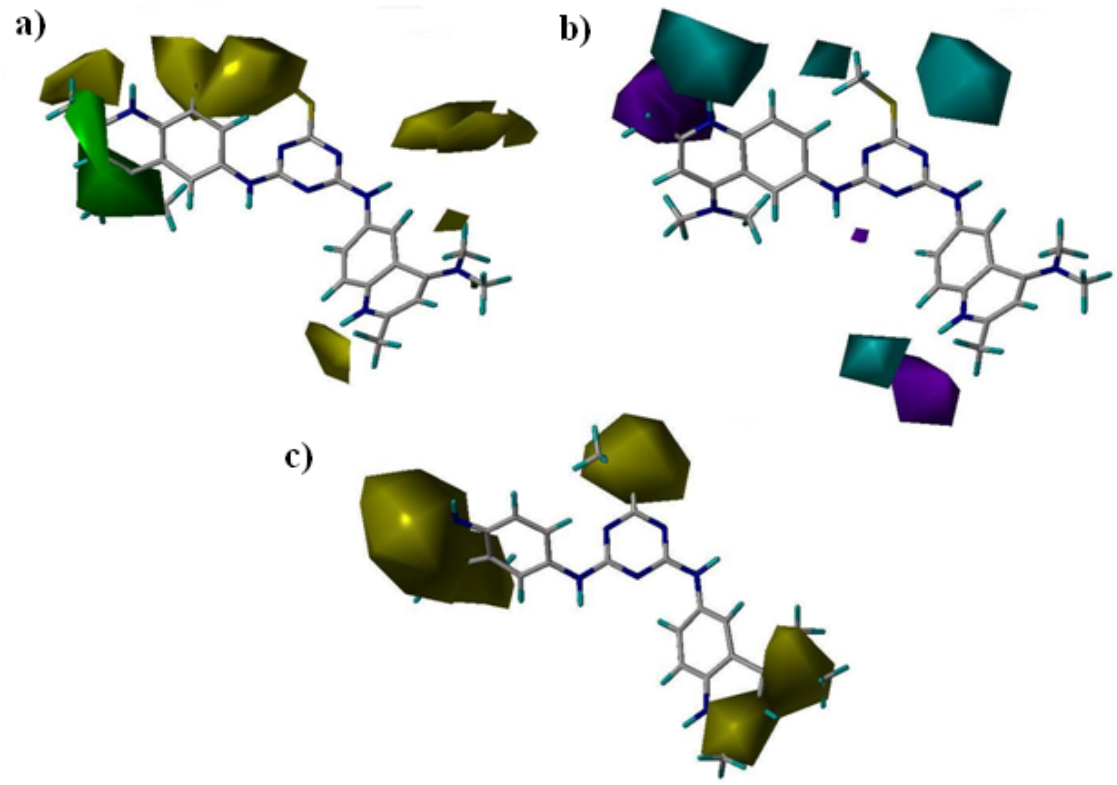


Figure4.3.8: CoMSIA stdev*coeff contour maps. Compound (**11)** is shown inside the field. (a) Steric field, (b) H-bond donor field and (c) Hydrophobic field. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the activity. Cyan contours indicate regions where the H-bond donor group increases activity, whereas purple contours indicate regions where the H-bond donor group decreases activity. Yellow contours indicate regions where hydrophobic groups increase activity

Prominent green contours present in the vicinity of the left quinoline ring indicated that generally steric bulk is favored at these sites (Figure 4.3.8a). It was observed that a good inhibitory potency of compound (**11**) ($^{tel}EC_{50}$ 0.041 µM) was due to orientation of bulkier quinoline ring toward the sterically favored regions. The same orientation has also been observed in compound (**2**) ($^{tel}EC_{50}$ 0.056 µM) and compound (**8**) ($^{tel}EC_{50}$ 0.051 µM). While in the case of compounds (**20**) ($^{tel}EC_{50}$ 2.6 µM), (**21**) ($^{tel}EC_{50}$ 3.3 µM), (**26**) ($^{tel}EC_{50}$ 2.8 µM) and (**29**) ($^{tel}EC_{50}$ 2.5 µM) there was an absence of a bulkier quinoline ring in this region. Although, compound (**22**) ($^{tel}EC_{50}$ 3.4 µM), (**23**) ($^{tel}EC_{50}$ 3.4 µM) and (**25**) ($^{tel}EC_{50}$ 3.3 µM) possess bulkier groups in this region, these groups are shifted away from sterically favored green

contour maps and oriented towards sterically disfavored yellow contour maps. This makes the ground for their poor activity among the series of compounds. Another important factor which contributes to the better telomerase activity is bulkier substitution at fourth position of quinoline ring. It was perceived that compounds (**2**) ($^{tel}EC_{50}$ 0.056 μM), (**11**) ($^{tel}EC_{50}$ 0.041 μM) and (**17**) ($^{tel}EC_{50}$ 0.34 μM) with bulkier substitution on fourth position of quinoline ring are comparatively more potent than compounds (**18**) ($^{tel}EC_{50}$ 2.2 μM), (**19**) ($^{tel}EC_{50}$ 3.4 μM) and (**27**) ($^{tel}EC_{50}$ 3.4 μM) having smaller substitutions at this position.

The graphical interpretation of H-bond donor properties is shown in Figure 4.3.8b. A cyan colored H-bond donor region towards the end of all the three side chains of triazine ring showed the favorable regions for H-bond donor fields. A prominent purple disfavorable region was found in the vicinity of 2-CH_3 group present on both the quinoline rings. In the most potent compound (**11)** ($^{tel}EC_{50}$ 0.041 μM) S-CH_3 group, protonated nitrogen and –$N(CH_3)_2$ substituted at fourth position of quinoline ring are perfectly entrenched in the favorable cyan colored contours. H-bond donor groups –NC_2H_5 on left side quinoline moiety and 4-NH_2 on central triazine ring of compound (**2**) ($^{tel}EC_{50}$ 0.056 μM) are oriented towards favored H-bond donor contours. Compounds (**31**) ($^{tel}EC_{50}$ 2.0 μM), (**14**) ($^{tel}EC_{50}$ 3.9 μM) and (**15**) ($^{tel}EC_{50}$ 3.5 μM) show lower activity as protonated nitrogens of their quinoline rings are oriented away from the favorable cyan region.

Hydrophobic maps shown in Figure 4.3.8c indicate that the lipophilic favorable yellow region is found surrounding both of the quinoline rings. This suggests that lipophilicity of this portion of the molecule is an important factor for the activity. Compound (**11**) ($^{tel}EC_{50}$ 0.041 μM) shows higher telomerase inhibitory activity among the series because three hydrophobic substituents on the three side chains of triazine ring are perfectly embedded in favored yellow hydrophobic contour maps. Deviation of these substituents, away from favored regions lead to loss of telomerase inhibitory activity of triazine derivatives viz. compound (**31**) ($^{tel}EC_{50}$ 2.0 μM), (**14**) ($^{tel}EC_{50}$ 3.9 μM), (**19**) ($^{tel}EC_{50}$ 3.4 μM) and (**22**) ($^{tel}EC_{50}$ 3.4 μM).

4.4. Efforts towards developing a universal 3D-QSAR model for G-quadruplex mediating telomerase inhibitors

After developing independent 3D-QSAR models for G-quadruplex mediating telomerase inhibitors belonging to different chemical classes, it was planned to study the feasibility of developing a universal 3D-QSAR model for telomerase inhibitors. It was though that if such a model could be developed then it could be employed in future for predicting the activity of compounds (i.e. compounds having aromatic central core with two side chains on either side) before their actual synthesis.

4.4.1 Data set used

A data set of 169 G-quadruplex mediating telomerase inhibitors having four different common scaffolds (Acridine, Anthraquinone, Acridone and Triazine) was taken from literature [146, 154,155, 157, 160-164, 167, 220]. The structures of training and test set molecules are given in Table 4.4.1. The biological activity values were converted into molar units and then expressed in negative logarithmic values. For developing 3D-QSAR CoMSIA model, training set and test set comprise of 127 and 42 compounds, respectively.

4.4.2 Selection of template and molecular alignment

Various alignement rules were tried, but a combination of atom and centroid-based alignment yielded best results. The template structure compound (**12**) used for the alignment is shown in Figure 4.4.1. The atoms and centroid considered for the alignment are marked with an asterisk (*) and C1 respectively. The Superimposition of all the compounds on template is shown in Figure 4.4.2.

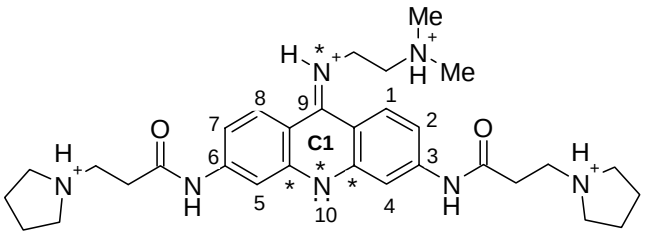
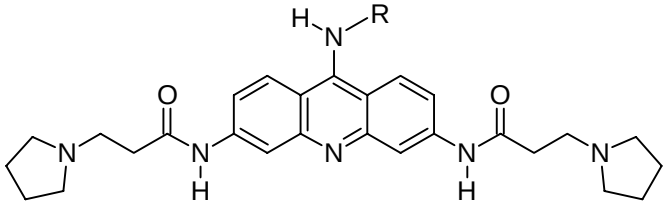
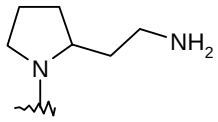
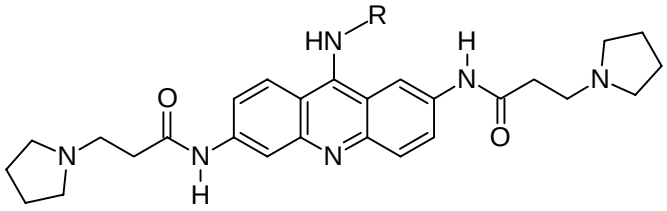


Figure 4.4.1: Template structure

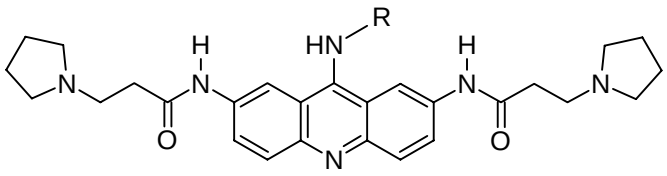
Table 4.1.1 Telomerase inhibition data for substituted acridine compounds



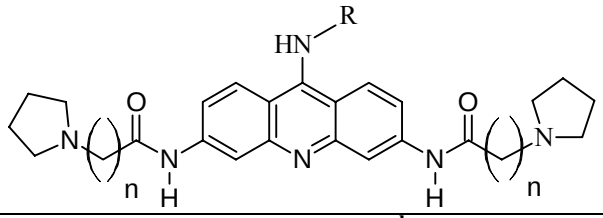
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
1	-$C_6H_4NMe_2$ (*p*)	0.115	6.94	6.976
2	-$C_6H_4NH_2$ (*p*)	0.074	7.13	7.091
3	-$CH_2CH_2CH_2NHMe_2$	0.06	7.22	7.231
4[a]	-$CH_2CH_2C_5H_{10}N$ (*c*)	0.05	7.30	7.099
5	-$C_6H_4NH_2$ (*m*)	0.06	7.22	7.033
6	-$C_6H_4NH_2$ (*o*)	0.02	7.70	7.315
7[a]	-$C_6H_4NMe_2$ (*m*)	0.1	7.00	7.045
8[a]	-C_6H_{11} (*c*)	0.09	7.05	6.860
9	-CH_2CH_2OMe	0.14	6.85	7.087
10[a]	-C_7H_{13} (*c*)	0.21	6.68	6.458
11	-$C_6H_4COCH_3$ (*p*)	0.04	7.40	7.174
12	-$CH_2CH_2NHMe_2$	0.018	7.74	7.480
13		0.018	7.74	7.565
14[a]	-$CH_2C_5H_4N$ (*m*) (*c*)	0.066	7.18	7.022
15[a]	-$C_6H_4NHCOCH_3$ (*m*)	0.1	7.00	7.064
16	-C_3H_5 (*c*)	0.05	7.30	6.999
17	-C_6H_4F (*p*)	0.07	7.15	7.072
18	-C_6H_4SMe (*o*)	0.15	6.82	7.028
19	-C_6H_4SMe (*m*)	0.1	7.00	7.032

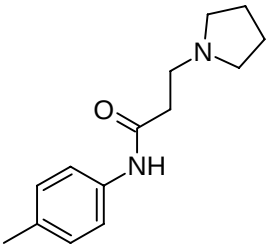
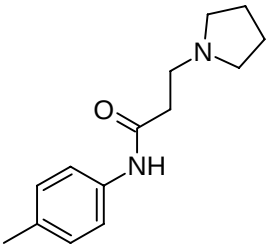
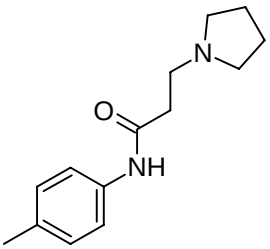


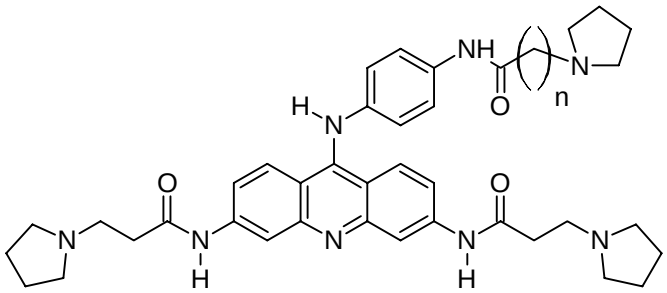
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
20	-$C_6H_4NH_2$ (*p*)	0.08	7.10	6.909
21[a]	-$C_6H_4NMe_2$ (*p*)	0.17	6.77	6.718
22[a]	-$CH_2CH_2NHMe_2$	0.27	6.57	6.971
23	-$C_6H_4NH_2$ (*m*)	0.21	6.68	6.993
24	-$C_6H_4NH_2$ (*o*)	0.11	6.96	6.953
25[a]	-C_6H_5	1.33	5.88	6.715
26	-$CH_2CH_2CH_2NHMe_2$	0.08	7.10	6.988
27[a]	-C_6H_{11} (*c*)	0.21	6.68	6.837



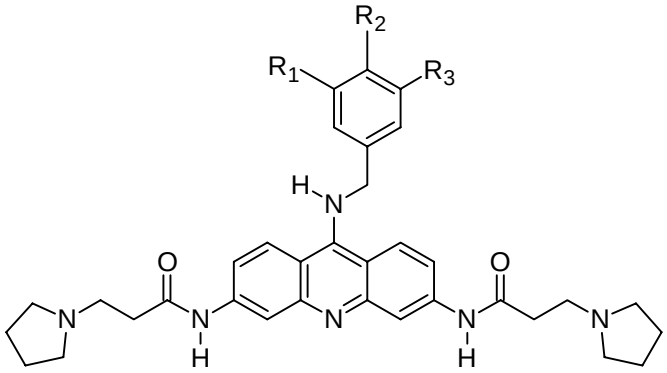
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
28	-C_6H_4OMe(*p*)	0.46	6.34	6.116
29	-$C_6H_4NH_2$ (*o*)	0.17	6.77	6.392
30[a]	-$C_6H_4NH_2$ (*m*)	1.09	5.96	6.430
31	-$C_6H_4NMe_2$ (*m*)	0.6	6.22	6.081
32	-$C_6H_4NH_2$ (*p*)	0.2	6.70	6.335
33[a]	-$C_6H_4NMe_2$ (*p*)	0.5	6.30	6.150
34	-C_6H_5	1.29	5.89	6.150
35	-C_6H_4OMe (*m*)	2.73	5.56	6.060
36	-C_6H_4OH (*o*)	1.03	5.99	6.203
37	-$CH_2CH_2NHMe_2$	0.57	6.24	6.381



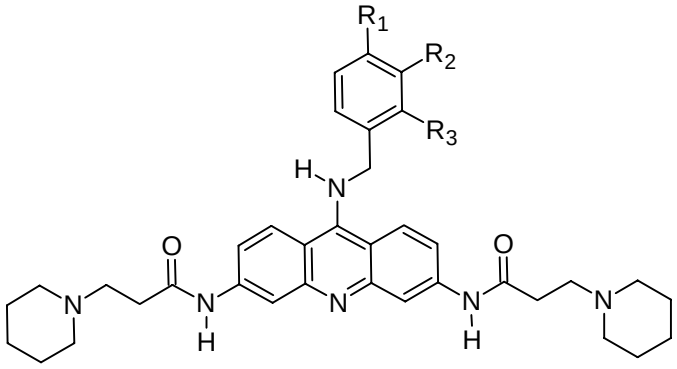
Compound	R	n	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
38	-$C_6H_4N(Me)_2$ (*p*)	3	0.099	7.00	6.519
39[a]	-$C_6H_4N(Me)_2$ (*p*)	4	1.93	5.71	6.356
40	-$C_6H_4N(Me)_2$ (*p*)	5	6.91	5.16	5.830
41		3	0.326	6.49	6.759
42[a]		4	0.255	6.59	6.557
43		5	0.146	6.84	6.162



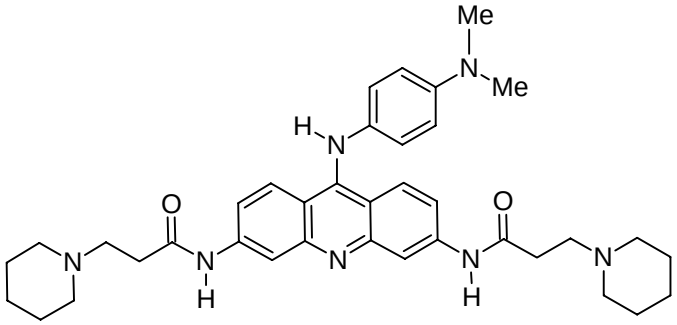
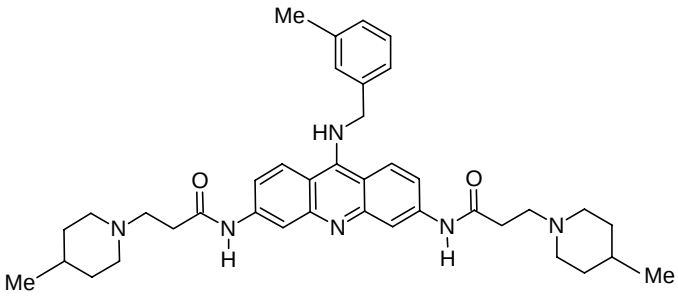
Compound	n	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
44	1	0.167	6.78	7.077
45	2	0.067	7.17	7.119
46	3	0.117	6.93	7.029

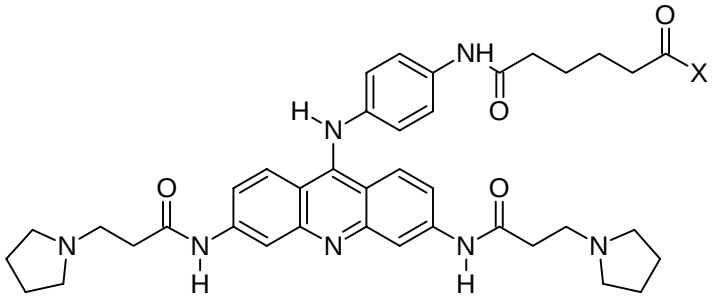


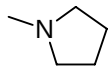
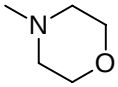
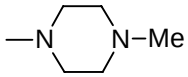
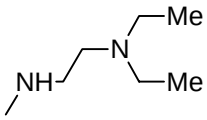
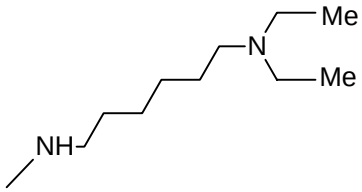
Compound	R_1	R_2	R_3	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
47[a]	F	F	H	0.03	7.52	6.670
48	OCH_3	H	OCH_3	0.35	6.46	6.783
49	CH_3	H	H	1.00	6.00	6.597
50	CF_3	H	F	0.24	6.62	6.576
51[a]	CF_3	H	CF_3	0.86	6.07	6.513

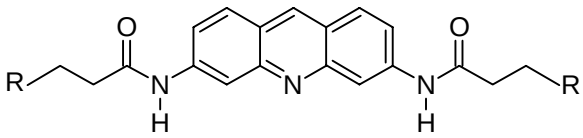


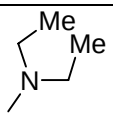
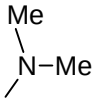
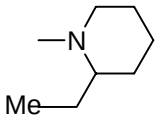
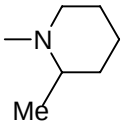
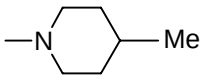
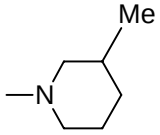
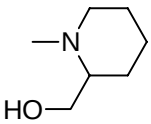
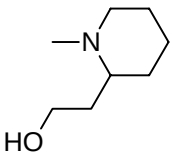
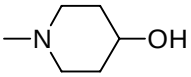
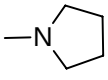
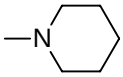
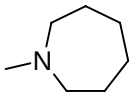
Compound	R_1	R_2	R_3	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
52	F	F	H	0.88	6.06	6.195
53[a]	OCH_3	H	OCH_3	0.44	6.36	6.310
54	CH_3	H	H	0.23	6.64	6.170

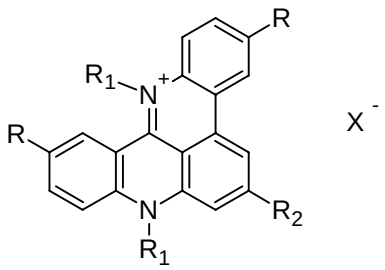
Compound	Structure	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
55		0.36	6.44	6.584
56[a]		0.39	6.41	6.138



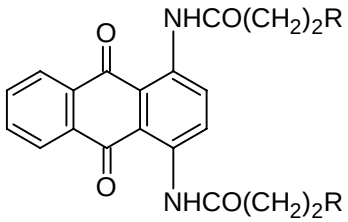
Compound	X	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
57		0.318	6.50	6.622
58		0.267	6.57	6.457
59		0.165	6.78	6.740
60		0.098	7.01	6.769
61[a]		0.080	7.10	6.812

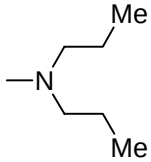
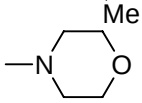
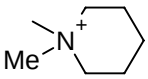
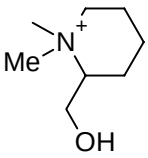
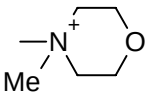
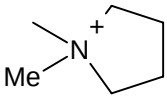
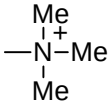
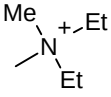


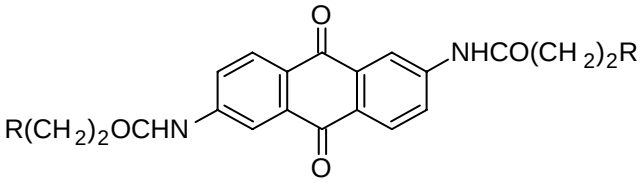
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
62		5.8	5.24	5.368
63		8.2	5.09	5.479
64		2.7	5.57	5.639
65		2.6	5.58	5.456
66		1.35	5.87	5.675
67		4.4	5.36	5.510
68		5.4	5.27	5.027
69		4.1	5.39	5.603
70		8.0	5.10	5.230
71		5.2	5.28	5.386
72[a]		2.8	5.55	5.672
73		3.1	5.51	5.492

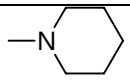
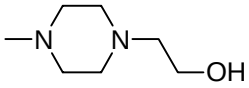
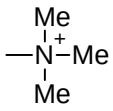
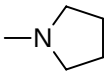
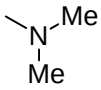
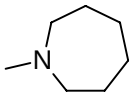


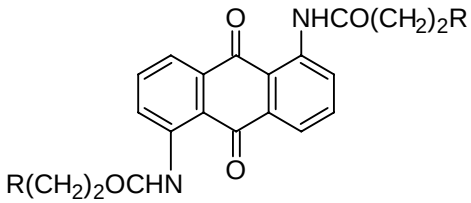
Compound	**R**	**R_1**	**R_2**	**X^-**	**$^{tel}EC_{50}$ (µM)**	**Actual *p*EC_{50}**	**Predicted *p*EC_{50}**
74	H	Me	H	I	0.38	6.42	6.392
75	H	Et	Me	I	2.00	5.70	6.316
76	Me	Me	Me	$MeOSO_3$	0.25	6.60	6.460
77	F	Me	Me	$MeOSO_3$	0.33	6.48	6.459
78	H	Me	Me	$MeOSO_3$	0.76	6.12	6.393
79	Cl	Me	Me	$MeOSO_3$	0.25	6.60	6.558

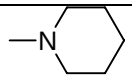
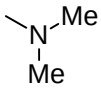
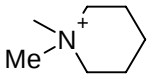
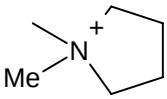
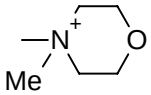
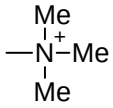
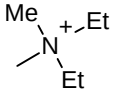


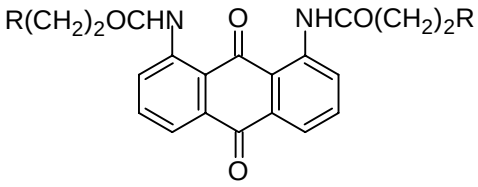
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
80		50	4.30	4.171
81		33.5	4.47	4.576
82		11.1	4.95	5.192
83		9.4	5.03	5.095
84		34.5	4.46	4.511
85[a]		1.8	5.74	5.312
86		5.0	5.30	5.271
87		7.0	5.15	5.198
88		3.1	5.51	5.279

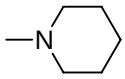
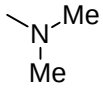
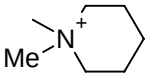
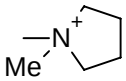
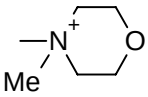
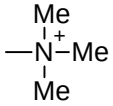
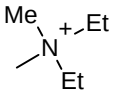


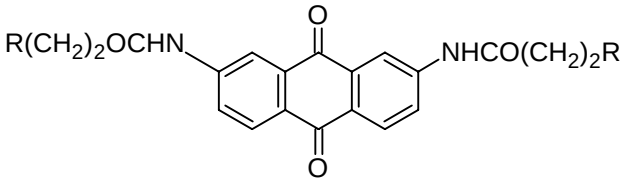
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
89		4.5	5.35	5.345
90		16.5	4.78	4.640
91		17.3	4.76	4.922
92		1.8	5.74	5.172
93		4.1	5.39	5.265
94[a]		3.5	5.46	5.266
95		13.0	4.89	5.146

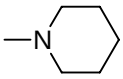
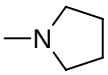
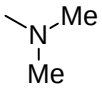
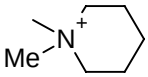
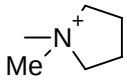
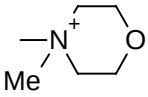
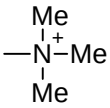


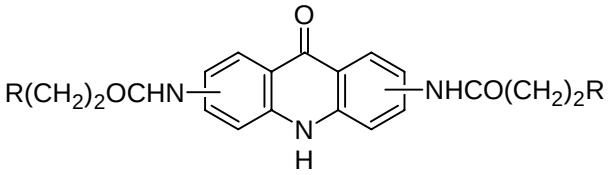
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
96		2.3	5.64	5.533
97		1.3	5.89	5.676
98		2.7	5.57	5.650
99		8.6	5.07	5.262
100[a]		8.8	5.06	5.425
101		14.0	4.85	4.712
102		13.2	4.88	4.824
103		16.8	4.77	4.697

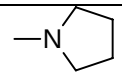
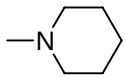
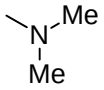
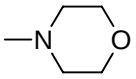
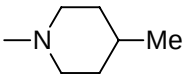
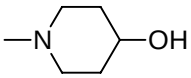
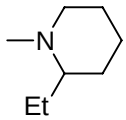
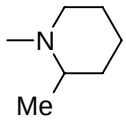
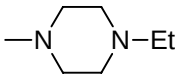
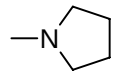
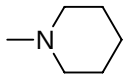
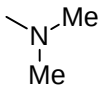
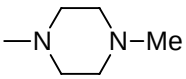


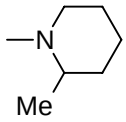
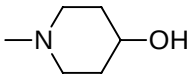
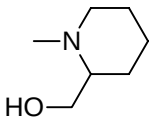
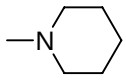
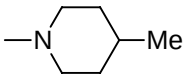
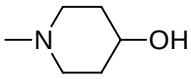
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
104		3.7	5.43	5.306
105[a]		6.4	5.19	5.326
106		4.2	5.38	5.304
107		7.8	5.11	4.985
108		8.2	5.09	5.247
109		10.0	5.00	5.155
110		4.4	5.36	5.253
111		7.5	5.12	5.244

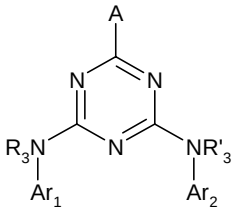


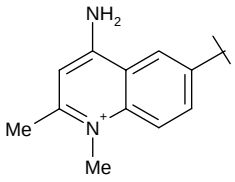
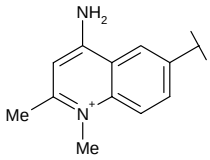
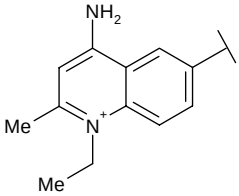
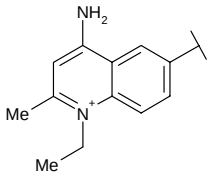
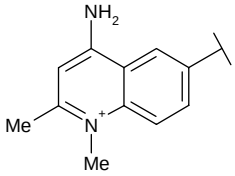
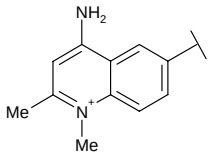
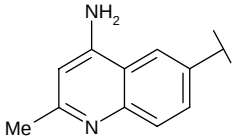
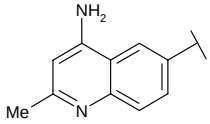
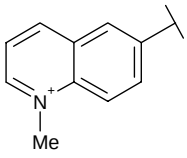
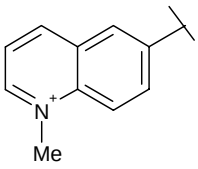
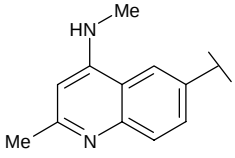
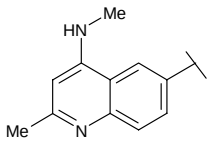
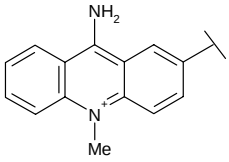
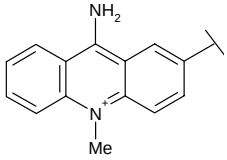
Compound	R	$^{tel}EC_{50}$ (μM)	Actual pEC_{50}	Predicted pEC_{50}
112[a]		3.1	5.51	5.154
113[a]		2.0	5.70	5.168
114		4.7	5.33	5.311
115		4.3	5.37	5.334
116[a]		7.8	5.11	4.829
117		16.0	4.80	4.714
118		16.5	4.78	4.555
119[a]		14.5	4.84	4.984

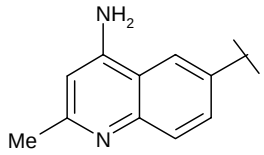
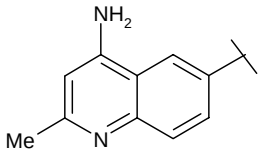
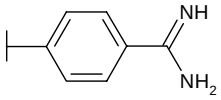
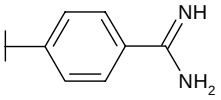
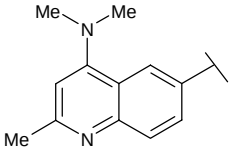
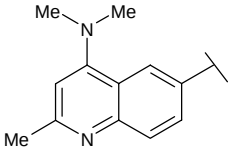
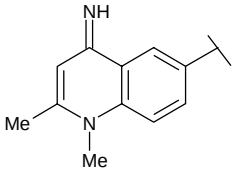
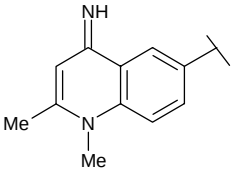
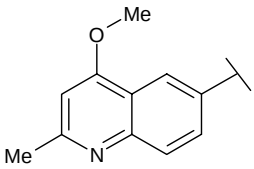
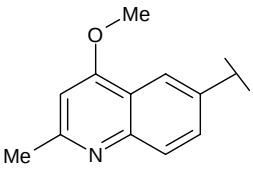
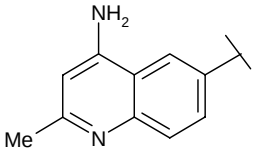
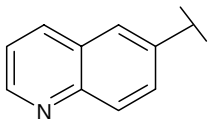
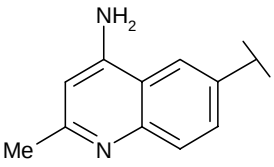
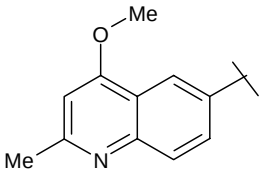
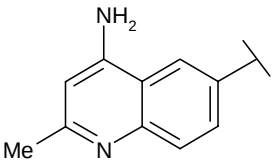
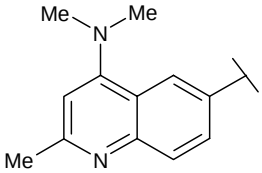
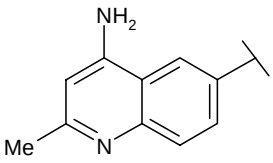
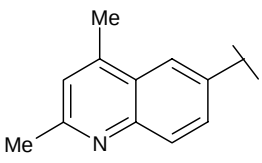
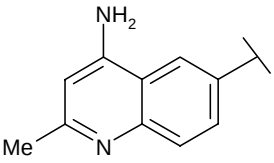
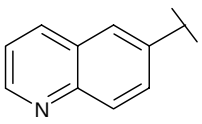


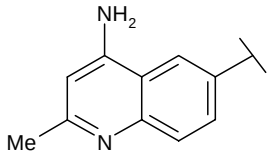
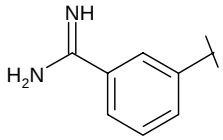
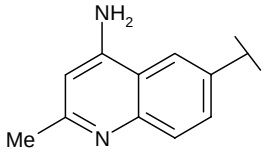
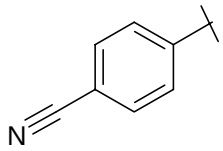
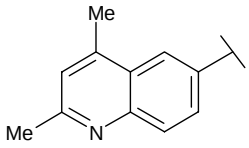
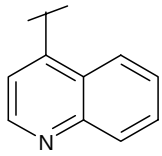
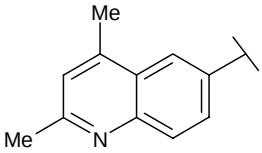
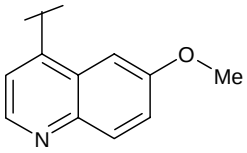
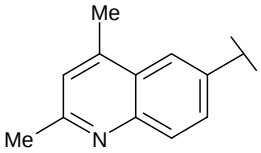
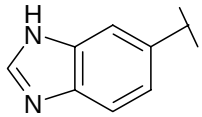
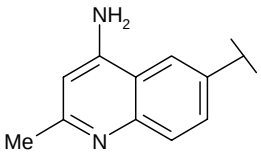
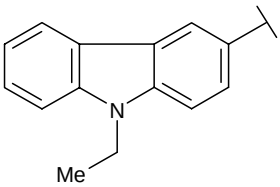
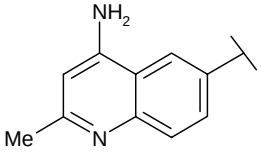
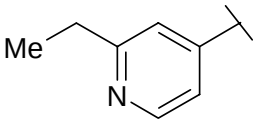
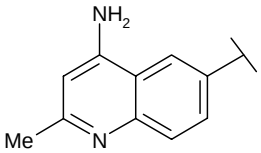
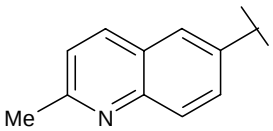
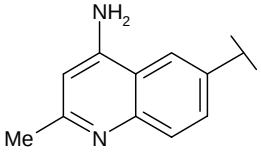
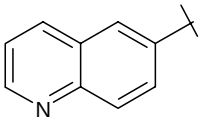
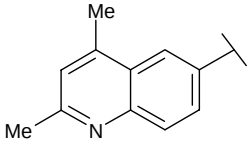
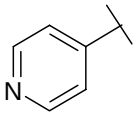
Compound	Position	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
120	3,6		8.1	5.09	5.430
121[a]	3,6		5.9	5.23	5.633
122	3,6		4.3	5.37	5.445
123	3,6		5.7	5.24	5.512
124*	3,6		49	4.31	-
125	3,6		1.7	5.77	5.714
126[a]	3,6		1.7	5.77	5.411
127[a]	3,6		2.3	5.64	5.743
128	3,6		2.3	5.64	5.658
129	3,6		1.9	5.72	5.639
130	2,7		5.8	5.24	5.310
131[a]	2,7		1.9	5.72	5.957
132	2,7		0.6	6.22	5.899
133	2,7		1.9	5.72	5.697

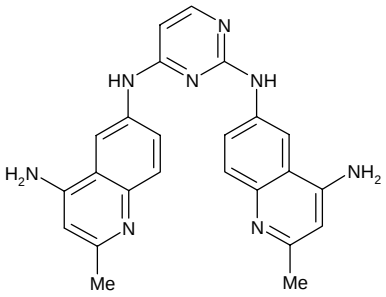
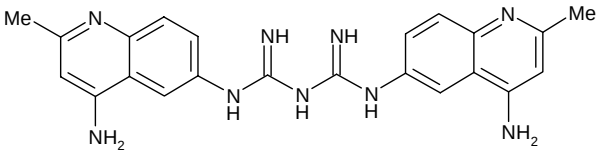
Compound	Position	R	$^{tel}EC_{50}$ (µM)	Actual pEC_{50}	Predicted pEC_{50}
134[a]	2,7		1.5	5.82	5.923
135	2,7		2.3	5.64	5.632
136	2,6		0.2	6.70	6.472
137	2,6		0.7	6.15	6.405
138	2,6		0.4	6.40	6.457
139	2,6		0.2	6.70	6.529
140	2,6		0.2	6.70	6.770



Compd	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (μM)	Actual pIC_{50}	Predicted pIC_{50}
141	$-NH_2$		H		H	0.25	6.60	6.571
142[a]	$-NH_2$		H		H	0.056	7.25	7.206
143	$-N(Me)_2$		H		H	0.22	6.66	6.840
144	-NHMe		H		H	0.51	6.29	6.301
145	$-NH_2$		H		H	0.13	6.89	7.038
146	-NHMe		H		H	0.3	6.52	6.892
147[a]	$-NH_2$		H		H	0.89	6.05	6.332

Compd.	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (µM)	Actual pIC_{50}	Predicted pIC_{50}
148	- NH_2		H		H	0.051	7.29	7.123
149[a]	- NH_2		H		H	0.74	6.13	6.148
150	- SMe		H		H	0.041	7.39	6.885
151	- SMe		H		H	0.95	6.02	6.358
152*	- SMe		H		H	3.9	5.41	-
153	- SMe		H		Me	3.5	5.46	5.556
154	- SMe		H		H	2.3	5.64	5.563
155[a]	- SMe		H		H	0.34	6.47	6.652
156	- SMe		H		H	2.2	5.66	5.458
157	- SMe		H		H	3.4	5.47	5.446

Compd.	A	Ar_1	R_3	Ar_2	R'_3	IC_{50} (µM)	Actual pIC_{50}	Predicted pIC_{50}
158	-SMe		H		H	2.6	5.59	5.404
159[a]	-SMe		H		H	3.3	5.48	5.684
160	-SMe		H		H	3.4	5.47	5.403
161	-SMe		H		H	3.4	5.47	5.561
162	-SMe		H		H	3.3	5.48	6.072
163*	-SMe		H		H	3.3	5.48	-
164[a]	-SMe		H		H	2.8	5.55	5.416
165	-SMe		H		H	3.4	5.47	5.484
166	-$N(Et)_2$		H		H	1.0	6.00	5.687
167[a]	-$N(Et)_2$		H		H	2.5	5.60	5.788

Compound	Structure	IC_{50} (µM)	Actual pIC_{50}	Predicted pIC_{50}
168		0.49	6.31	6.197
169		2.0	5.70	5.590

[a]Test set compounds; *Identified outliers

4.4.3 CoMSIA results

Initially, CoMSIA models were constructed with 127 compounds in training and 42 compounds in test set using various fields in combination. The qualities of the models so developed were judged by the statistical significance of each model. According to this analysis it was detected that compounds (**124**, **154** and **165)** offered large residuals (greater than a unit value). CoMSIA models with 126 compounds in training set and 40 compounds in test set were constructed and the results showed a remarkable enhancement in statistical quality of each model. The summery of each CoMSIA model using different descriptor fields has been shown in Table 4.4.2

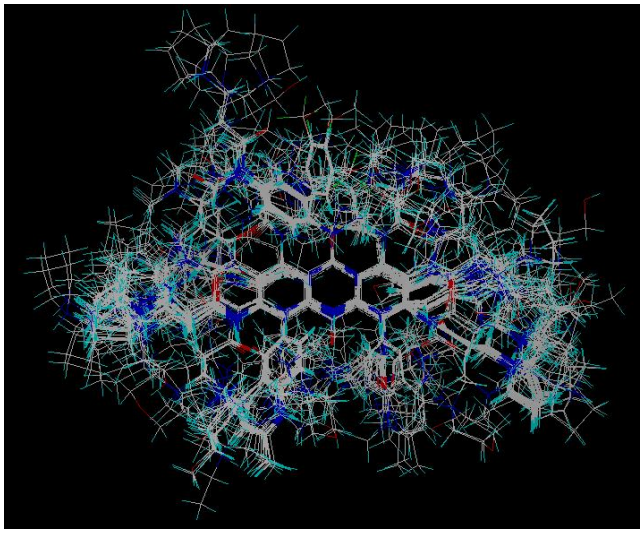


Figure 4.4.2: Superimposition of all compounds

Table 4.4.2 Summery of CoMSIA models

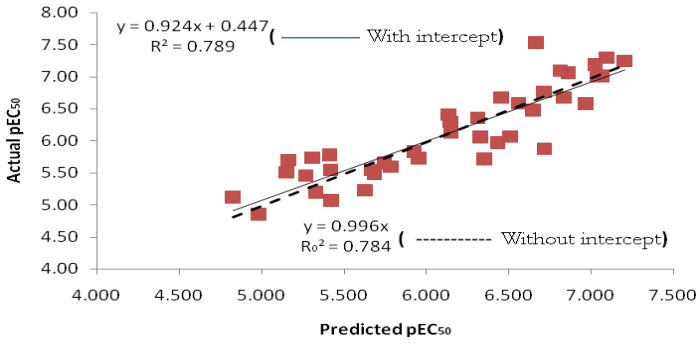
	HAD	SED	SEA	SHE	EHD	SEDA	SEHD	SEHA	**SEHDA**
r^2_{cv}	0.694	0.604	0.666	0.655	0.638	0.677	0.658	0.680	**0.683**
ONC	4	5	4	6	6	6	6	4	**6**
SEP	0.468	0.525	0.480	0.493	0.504	0.476	0.490	0.471	**0.472**
r^2_{ncv}	0.909	0.872	8.831	0.890	0.913	0.908	0.912	0.849	**0.917**
SEE	0.252	0.298	0.341	0.278	0.247	0.255	0.249	0.323	**0.241**
***F*-Value**	238.948	163.836	149.130	160.480	209.327	195.225	204.487	170.457	**220.342**
$P_{r2=0}$	0	0	0	0	0	0	0	0	**0**
Contribution (fraction)									
S	-	24.4%	24.0%	22.0%	-	16.5%	15.5%	16.2%	**12.0%**
E	-	35.9%	35.5%	33.6%	27.6%	22.5%	23.5%	24.9%	**17.6%**
H	34.0%	-	-	44.3%	39.3%	-	32.6%	29.6%	**23.2%**
D	32.0%	39.6%	-	-	33.1%	28.9%	28.3%	-	**22.6%**
A	33.9%	-	40.5%	-	-	32.1%	-	29.3%	**24.6%**
r^2_{pred}	0.758	0.736	0.747	0.715	0.782	0.770	0.783	0.766	**0.793**
r^2_{bs}	-	-	-	-	-	-		-	**0.949**
SD_{bs}	-	-	-	-	-	-		-	**0.009**

ONC = Optimum number of components, SEP = Standard error of prediction, SEE=Standard error of estimate, S = Steric, E = Electrostatic, D = H-bond donor, A = H-bond acceptor and H = Hydrophobic

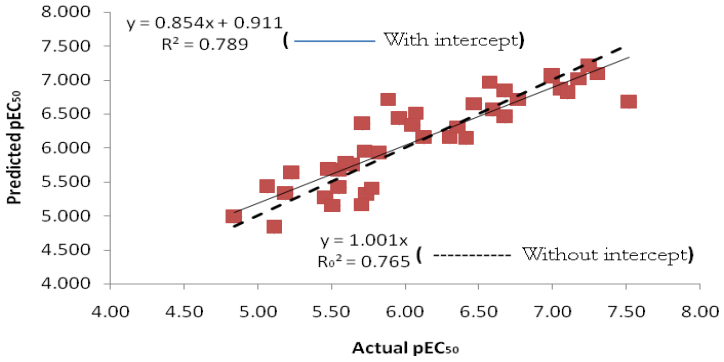
4.4.4 Validation of best CoMSIA model

It was observed that in CoMSIA model involving steric, electrostatic, hydrophobic, donor and acceptor fields all five descriptors played significant role in predicting telomerase inhibitory activity. CoMSIA model generated using the above fields showed cross-validated r^2 0.683 with six optimum number of components, non-cross validated r^2 0.917, *F*-value 220.342, low standard error of estimate 0.241, bootstrapped r^2 0.949 and predictive r^2 0.793. The predicted activity of training and test set compounds are given in Table 4.4.1.

The regression plots of predicted vs. actual activities and actual vs. predicted activities were constructed (Figure 4.4.3a,b) to apply rigorous validation parameters reported by Tropsha et al. that have been described in chapter 1, section 1.1.1.6.



(a)



(b)

Figure 4.4.3: External validation of CoMSIA model. (a) Predicted vs. Actual activities (b) Actual vs. Predicted activities, of Test Set compounds

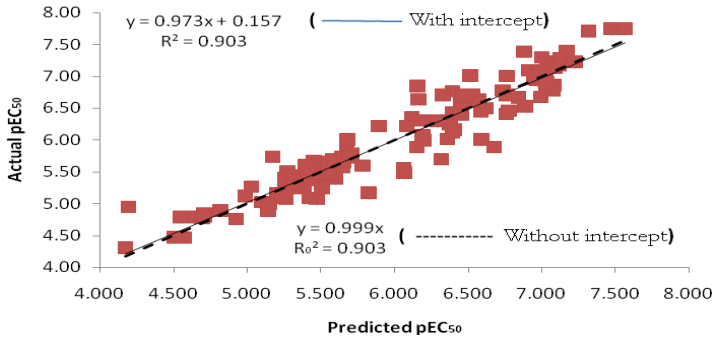


Figure 4.4.4: Internal Validation of best CoMSIA model. Predicted vs. Actual activities of Training set compounds

4.3.3.1 R^2_m as a validation criterion

Additionally, the model was validated with validation criterion R^2_m reported by Roy et.al [28, 29]. The R^2_m (Test) value penalizes a model for large differences between observed and predicted values of test set compounds. The CoMSIA model was found to be superior with R^2_m (Test) = 0.787. Since this value is based on limited

number of test set compounds, to further confirm its reliability, R^2_m values were calculated for training set compounds using LOO predictions. The value of R^2_m (training) 0.903 confirms that the model built so is a reliable one. The results obtained are shown in Table 4.4.3. In case of best external prediction $R^2_m = R^2$ while in the worst case, $R^2_m = 0$. In the present study for external test set compounds, a value of R^2_m 0.787 was found to be almost equal to R^2 value 0.789 and for training set compounds, a value of 0.903 was found to be exactly equal to R^2 value 0.903, Figure 4.4.4, indicated that the derived model is a highly predictive one. The results are summarized in Table 4.4.3

Table 4.4.3 Summery of addition validation parameter

Statistics	CoMSIA model (SEHDA)
R^{2a}	0.789
R_0^{2b}	0.765
k^c	1.001
$(R^2 - R_0^2)/R^2$	0.006
R'^{2b}_0	0.784
k'^c	0.996
$(R^2 - R'^2_0)/R^2$	0.030
$R_m^2 = R^2 (1-\sqrt{\lvert R^2 - R_0^2 \rvert})$	0.787
Absolute sum of observed pIC_{50} (Test set)	245
Absolute sum of residuals (Test set)	10.343
% residual (Test Set)	4.22%
Absolute sum of observed pIC_{50} (Training set)	764.64
Absolute sum of residuals (Training set)	24.168
% residual (Training set)	3.16%

[a] Correlation coefficient derived from predicted pIC_{50} of test set compounds
[b] Correlation coefficient for the regression line passing through origin for actual vs. predicted and predicted vs. actual activity, respectively
[c] Slope for regression line passing through origin obtained from actual vs. predicted and predicted vs. actual activity, respectively

4.4.5 Interpretation of best CoMSIA model

The contour maps of the CoMSIA steric, electrostatic, hydrophobic, donor and acceptor fields are presented in Figure 4.4.5

The contour plot of the CoMSIA steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor fields (stdev*coeff) are presented in Figure 4.4.5. Favored and disfavored levels fixed at 80% and 20%, respectively, were used.

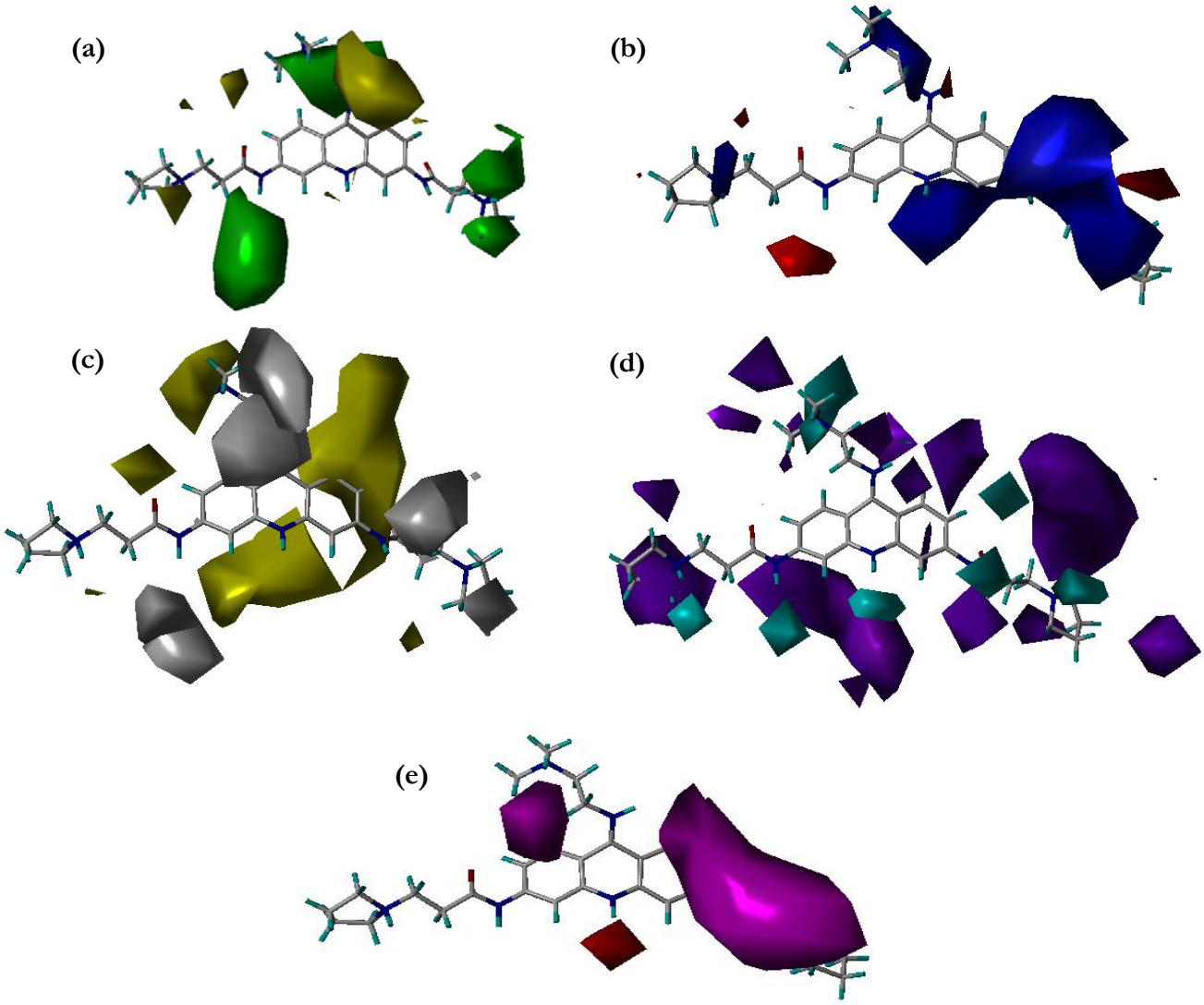


Figure 4.4.5: CoMSIA stdev*coeff contour maps. Compound (**12**) is shown inside the field. (a) Steric, (b) Electroststic, (c) Hydrophobic field, (d) Donor, (e) Acceptor fields. Green and yellow polyhedra indicate regions where more steric bulk or less steric bulk, respectively, will enhance the activity. Blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity, Yellow contours indicate regions where hydrophobic groups increase activity, Cyan contours indicate regions where the H-bond donor group increases activity, whereas purple contours indicate regions where the H-bond donor group decreases activity, Magenta contours represent areas where H-bond acceptors are favored.

The most active compound (**12**) ($^{tel}EC_{50}$ 0.018) from the series is shown inside the fields. Significant green contours surrounding 9th position of acridine ring and small contours in the vicinity of the pyrrolidine rings represent favorable steric area to

increase the telomerase inhibitory activity. This is the reason why compounds (**12**) ($^{tel}EC_{50}$ 0.018) and (**13**) (**(**$^{tel}EC_{50}$ 0.018**)** showed high telomerase inhibitory activity. Whereas in case of disubstituted acridine derivatives (**63**) ($^{tel}EC_{50}$ 8.2), (**68**) ($^{tel}EC_{50}$ 5.4), (**70**) ($^{tel}EC_{50}$ 8.0), (**71**) ($^{tel}EC_{50}$ 5.2) and (**73**) ($^{tel}EC_{50}$ 3.1) there was an absence of bulkier group in this region. In case of anthraquinone derivatives, compounds (**139**) ($^{tel}EC_{50}$ 0.2) and (**140**) ($^{tel}EC_{50}$ 0.2) are the most potent telomerase inhibitors due to the reason that side chains of these molecules are perfectly entrenched in sterically favored green contours. Compounds (**80**) ($^{tel}EC_{50}$ 50) and (**81**) ($^{tel}EC_{50}$ 33.5) showed poor telomerase inhibitory activity as 1,4-disubstituted side chains of these compounds have oriented towards the sterically disfavorable yellow contours. In case of triazines, a good inhibitory potency of compound (**150**) ($^{tel}EC_{50}$ 0.041 μM) is due to the orientation of bulkier quinoline ring toward the sterically favored regions.

Blue contours surrounding the pyrrolidine rings of both of the side chains substituted at C-3 and C-6 of acridine ring represent favourable site for electropositive substitutions. Another favorable electrostatic region is observed above the acridine ring in the model. Compound (**20**) ($^{tel}EC_{50}$ 0.08) showed higher telomerase inhibitory activity due to the presence of *p*-NH_2 group on the phenyl ring oriented in blue region. Compounds (**139)** ($^{tel}EC_{50}$ 0.2) and (**140**) ($^{tel}EC_{50}$ 0.2) are the most potent telomerase inhibitors due to the reason that the side chains of these molecules are perfectly entrenched in electrostatic favorable blue contours. Compounds (**80**) ($^{tel}EC_{50}$ 50) and (**81**) ($^{tel}EC_{50}$ 33.5) showed poor telomerase inhibitory activity as their 1,4-disubstituted side chains were oriented towards disfavorable red contours.

For hydrophobic contour maps, lipophilic favorable yellow region is found surrounding the aromatic acridine ring and the white region under yellow contour at 9th position of acridine ring indicates that hydrophilic atoms in this zone favor the telomerase inhibitory activity. Hence, compounds with aniline and benzyl substituents at C-9 position of acridine moiety show good activity as the aromatic ring fits well into the yellow region which is favorable for hydrophobic groups. In case of anthraquinone derivatives, compounds (**136**) ($^{tel}EC_{50}$ 0.2), (**139**) ($^{tel}EC_{50}$ 0.2) and (**140**) ($^{tel}EC_{50}$ 0.2) were observed to be active anti-telomerase agents due to orientation of alkylamido side chains of these compounds perfectly entrenched in favorable yellow region. For triazine derivatives, compound (**141**) ($^{tel}EC_{50}$ 0.041) showed higher telomerase inhibitory activity among the series because three

hydrophobic substituents on the three side chains of triazine ring are perfectly embedded in favored yellow hydrophobic contour maps.

Cyan colored hydrogen bond donor region surrounds C-9 position terminal nitrogen atom, pyrrolidine ring nitrogen atoms and central ring nitrogen atom of acridine ring indicating favorable regions for H-bond donor fields. A *p*-NH_2 group on phenyl ring of compound (**20**) ($^{tel}EC_{50}$ 0.08) oriented towards cyan region, could be the reason why compound (**20**) has higher telomerase inhibitory activity. Compound (**141**) ($^{tel}EC_{50}$ 0.041 μM) showed a good telomerase inhibitory activity because S-CH_3 group, protonated nitrogen and –$N(CH_3)_2$ substituted at fourth position of quinoline ring are perfectly entrenched in the favorable cyan colored contours.

In case of H-bond acceptor contour maps, there is a prominent magenta contour surrounding the side chain substituted at C3 position; which indicates that an H-bond acceptor is favored at this position. Compound (**141)** ($^{tel}EC_{50}$ 0.041 μM) is most active as its oxygen atom of carbonyl group at position-9 and central ring nitrogen atom of acridone chromophore fall exactly in the region of favorable H-bond acceptor contours.

In overall conclusion, the main objective of any rational drug design approach is to design and discover new leads with appropriate shape and chemical features to be subsequently optimized to highly active and selective drugs. The present work has focused on the development of reliable 3D-QSAR models by CoMFA/CoMSIA and pharmacophore modeling studies of G-quadruplex mediating telomerase inhibitors.

5. EXPERIMENTAL

The tools, techniques and procedures/methods used for carrying out research work reported in this thesis have been described as follows:

5.1. Hardwares, Softwares and Web server used in Molecular modeling

5.1.1 Hardwares

Three-dimensional QSAR (3D-QSAR CoMFA and CoMSIA), studies were performed using Silicon Graphics Fuel workstation running on IRIX 6.5 operating system. The system configuration was as follows:

Work station Name	**Silicon Graphics Fuel**
Model	Onyx3 Infinite performance Fuel
CPU	1700 MHz MIPS R 1600 (IP35) Processor with MIPS R1 6010 FPU
Memory	512 MB
Data cache	32 KB
Instruction cache	32 KB
Operating system	IRIX 64 Release 6.5
Graphics	V10
Hardware	HUB in module 001c01 Slot 0 Revision 2 speed 200 MHz (enabled)

The molecular docking studies were performed on PC with Microsoft Windows XP Professional version 2002. The system configuration was as follows:

Model	**HP Compaque**
Processor	Pentium (R) 4, CPU 3.20 GHz
Memory	2 GB DDR2
Data cache	2 x 256 KB L2
Operating system	Windows XP

The pharmacophore modeling web server was used on PC with Microsoft Windows XP Professional version 2002. The system configuration was as follows:

Model	**Acer Aspire 4520**
Processor	AMD Athlon 64 x 2 Dual-core 1.8 GHz,
Memory	1 GB DDR2
Data cache	2 x 256 KB L2
Operating system	Windows XP
Graphics	NVIDIA GeForce 7000M

5.1.2 Softwares

- Three-dimensional QSAR studies were performed using QSAR module in SYBYL (version 7.0) procured from Tripos Inc., USA.
- Docking studies were performed using CDOCKER procured from Accelrys Inc., USA.
- Monte Carlo conformational search was performed using MacroModel version 7.0 procured from Schrödinger, Inc.
- *p*Ka values for the selected compounds were calculated using Pallas 3.7.1.1 software obtained from CompuDrug International Inc. with an academic license.

5.1.3 Web servers

A ligand-based pharmacophore elucidation was performed using PharmaGist web server [64].

5.2 Calculation of Biological Activities in Terms of pEC_{50}/pIC_{50} i.e. Negative logarithm of EC_{50}/IC_{50} in Molar Concentration

In all of the 3D-QSAR studies, the biological activity in EC_{50}/IC_{50} were converted to pEC_{50}/pIC_{50} using the following equation:

$$pEC_{50}/pIC_{50} = - \log (EC_{50}/IC_{50})$$

where, EC_{50}/IC_{50} is the concentration (in M) of the inhibitor producing 50% inhibition of enzyme. These values were taken as dependent variables in the 3D-QSAR analysis.

5.3 3D-QSAR studies

5.3.1 Structure building and energy minimization/optimization of structures

A Silicon Graphics Fuel workstation with IRIX 6.5 operating system running SYBYL 7.0 was used for three-dimensional structure building. Initial optimization of the structures was carried out using TRIPOS force field with Gasteiger-Huckel charges, and repeated minimization was performed using steepest-descent and conjugate gradient methods until the root-mean-square (rms) deviation of 0.001 kcal/mol was achieved. Conformational energies were computed with electrostatic terms; the lowest energy structures finally minimized were used in superimposition. The partial atomic charges required for the electrostatic interactions were computed by the semiempirical molecular orbital method using Molecular Orbital PACkage (MOPAC) [214] with Austin Model 1 (AM1) Hamiltonian [215].

5.3.2 Choice of bioactive template conformation

In case of 3D-QSAR studies of acridine derivatives as telomerase inhibitors, the co-crystalized ligand was extracted out from the holo structure (PDB code: 1L1H) [213] imported from the protein data bank. The energy minimized structure of the most active compound from the series was docked into active site of PDB 1L1H and all other compounds were built on it. The energy minimized docked conformation of the most active compound from the series was used as a template for molecular alignment.

In case of 3D-QSAR studies for anthraquinone and acridone derivatives as telomerase inhibitors bioactive conformation was obtained using MacroModel Monte Carlo conformational search. The two conformational search-based template conformations were derived as follows. The most active compound was constructed in SYBYL. Its structure was minimized, and then used as a starting point for a Monte Carlo conformational search employing MacroModel version 7.0 (Schrödinger, Inc.). The conformational search was carried out using MMFF94 force field in MacroModel for 5000 iterations using water as a solvent. The global minimum conformation, designated as Conf1, and the second lowest energy conformation from the global minimum designated as Conf2 were used as templates to construct the rest of the molecules. A third conformation was derived as discussed below:

The conformation of a disubstituted acridine derivative (biologically active moiety) was extracted from the co-crystallized structure of G-quadruplex PDB 1L1H. Conf1 of the most active compound was superimposed on the conformation of the acridine derivative, extracted from the co-crystallized structure. As a measure of superimposition reliability, the RMSD between these two conformations was taken into the consideration (0.9 Å obtained herein). It was interesting to note that the resulting conformation (conf3) of most active compound fully superimposed (rmsd 0.00 Å) on conformation conf1. Hence, conformations conf1 and conf2 only were used as templates for further study.

For search of bioactive template conformation in 3D-QSAR studies of triazine derivatives as telomerase inhibitors, a simulated annealing-based [222] global minimum energy structure of the most active compound was used as a template. It is known that the conformation representing the global minimum of the ligand may not bind to the receptor and some degree of torsional freedom is required for the drug to adopt to the receptor-binding site to yield a bioactive conformation [223]. But in our case, structures possess sufficient rigidity to limit the possible active conformations and the simulated annealing-based global minimum energy conformation is an excellent starting point to identify possible bioactive conformation for the template chosen. The most active compound was constructed in SYBYL. A preliminary minimization was performed to remove close atom contacts using standard Tripos force fields with Gasteiger-Huckel charges, and repeated minimizations were performed using steepest-descent and conjugate gradient methods until the root-mean-square (rms) deviation of 0.001 kcal/mol was achieved. The structure was next subjected to molecular dynamic simulation, wherein it was heated to 700 K for 1 ps and annealed slowly to 300 K in steps of 100 K for 1 ps at each temperature. The exponential annealing function was used and 10 such cycles were run. The least energy conformation was determined and subsequently subjected to minimization with the same criteria as mentioned above, and used as a template to build structures of all other training and test set compounds.

5.3.3 Molecular alignment

Various molecular alignment rules were tried but it was observed that the alignment carried out using the centroid and atom-based rms fit resulted into good

results. This option uses alignment of structures through defining the centroid and pairwise superimposition of these centroids and atoms which renders all structures in the database in the same frame of reference as the template compound.

5.3.4 Selection of Training and Test Sets

In view of the finding that q^2 appears to be a necessary but not a sufficient condition for a model to have high predictive power, [39] an emphasis has been given for external validation of the developed model using sufficiently large test set (25-33% of the total compounds) to establish a reliable 3D-QSAR model. In the training set, most potent, moderately active and low active compounds were included to spread the activity range. The test set compounds were selected in such a manner that at least one structural analog of the training set was chosen for the test set.

5.3.5 CoMFA Analysis

CoMFA of the selected series of molecules was carried out on the steric and electrostatic fields using the default values. A 3D cubic lattice, with a 2 Å grid spacing, was generated automatically around these molecules to ensure that the grid extended the molecular dimensions by 4 Å in all directions. A threshold column filtering of 2.0 kcal/mol was set to speed up the analysis and reduce the amount of noise. The steric and electrostatic fields were calculated separately for each molecule using sp^3 carbon atom probe with a +1 charge and computed energy cut-off values of 30 kcal/mol Å for both steric and electrostatic interactions with each atom in the molecule using CoMFA standard scaling. The detailed description of the settings used for CoMFA analysis are described in Table 5.3.1

5.3.6 CoMSIA analysis

The standard setting (probe with +1 charge , 1 Å radius, +1 hydrophobicity, +1 hydrogen bond donating, +1 hydrogen bond accepting, attenuation factor of 0.3 and grid spacing of 2 Å) was used in CoMSIA analysis to calculate five different fields viz. steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor.

Table 5.3.1 Settings used for CoMFA analysis

Setting Name	Setting description	Set to
CoMFA field class	Steric and/or electrostatic potentials are evaluated at every lattice point	Tripos standard
Field types	Field types to be used in the calculation	Both- steric and electrostatic
Dielectric function	A value to control the form of the Coulombic electrostatic energy calculation • Dielectric- energy falls as $1/r^2$ • Constant: energy falls as $1/r$, the pure Coulombic expression	Distance dielectric function
Smoothing	This option allows the field value used in CoMFA to be an average field value. None: Uses the actual values as the field value Box: Uses as the field value the average of field values at the vertices of a box centered on the actual value. Sphere: Uses as a field value, if region describes a known receptor.	None- Actual field values were considered
Drop Electrostatics	The Drop Electrostatics setting controls how electrostatic energies are treated at lattice points where steric energies are above their cutoffs.	Within steric cutoff for each row
Steric cutoff	Maximum possible steric energy value.	30 kcal/mol
Electrostatic cutoff	Maximum possible electrostatic energy value.	30 kcal/mol
Transition	• Smooth: Interpolates from 6 kcal/mol below the cutoff to the plateau (i.e. cutoff itself) • Abrupt: Discontinuous	Smooth
Region setting	Create new region file automatically as part of CoMFA column creation. It will extend at least 4Å beyond every molecule in all directions, will have 2Å spacing, and a probe atom with vdW properties of sp^3 hybridized carbon and a charge of +1	Create automatically

5.3.7 Partial Least Square (PLS) analysis

The CoMFA/CoMSIA descriptors were used as independent variables and $p\text{EC}_{50}/p\text{IC}_{50}$ values as dependent variables in PLS regression analysis for deducing 3D-QSAR models. Cross-validation was used to check the predictive power of the

derived model. The result of analysis corresponds to the regression equations with thousands of coefficients. The predictive values of models were evaluated using leave-one-out (LOO) cross-validation method. The number of components leading to the highest cross-validated r^2 and standard error of prediction (SEP), was set at the optimum number of components (*N*c) in PLS analysis. Using these optimum numbers of components, non-cross validated r^2 was derived. This analysis was saved and used to predict the biological activity of training and test set compounds under study. A minimum of 2.0 kcal/mol column filtering (σ) was used as the threshold column filtering value in PLS analysis. To further assess the robustness and statistical confidence of the developed model, the cross-validated results were analyzed by considering the fact that a value of r^2_{cv} above 0.3 indicated that probability of chance correlation was less than 5% [35].

5.3.8 Evaluation of the predictive ability of CoMFA and COMSIA models

The predictive ability of each analysis was determined from the test set molecules that were not included in the training set. These molecules were aligned and their activities were predicted by each PLS analysis. The predictive r^2 (r^2_{pred}) value was defined as follows:

$$r^2_{pred} = (SD\text{-}PRESS)/SD$$

where, SD is the sum of squared deviations between the biological activities of the test set and mean activity of the training set molecules and PRESS is the sum of squared deviation between actual and predicted activities of the test set molecules.

5.4 Docking studies

The X-ray crystal structure of the quadruplex-drug DNA complex (PDB 1L1H) [211] containing a co-crystallized disubsituted acridine derivative in the active site was used. Potassium ions in the central channel between the planes of each G-quartet were preserved, and all water molecules were deleted. The CHARMm force fields were assigned to a quadruplex, and the hydrogen atoms were automatically added. The structure of the drug-DNA complex after knocking out the cocrystalized ligand was used for docking studies.

Docking studies were performed using the molecular dynamics simulated annealing algorithm-based docking program CDOCKER to generate ligand poses wherein a set of ligand conformations are generated using high temperature molecular dynamics. Random orientations of the conformations are produced by translating the center of the ligand to a specified location within the receptor active site and performing a series of random rotations. A softened energy is calculated, and the orientation is kept if the energy is less than a specified threshold. This process continues until either the desired number of low-energy orientations is found or the maximum numbers of bad orientations have been tried. Each orientation is subjected to simulated annealing molecular dynamics. The target is heated up to a temperature of 700 K and then cooled to 300 K. A final minimization of the ligand in the rigid receptor using non-softened potential is performed. For each final pose, the CHARMm energy (interaction energy plus ligand strain) and the interaction energy alone are calculated. The poses are sorted by CHARMm energy, and the top scoring poses are retained where lower values indicate more favorable binding [83].

5.5 Pharmacophore elucidation

A ligand-based pharmacophore elucidation was performed using PharmaGist webserver. The method consists of four major steps: (i) ligand representation, (ii) pairwise alignment, (iii) multiple alignments and (iv) solution clustering and output as shown in Figure 5.1.

In common pharmacophore development approach, large number of possible conformations for each ligand were generated, whereas PharmaGist uses the most active compound as ‘the pivot’ and considered within the search for the common pharmacophore. The benefit of this approach lies in the fact that when there is no information on the binding conformation of any of the ligands, a set of conformations for only one of them (the pivot) needs to be computed.

Unless a pivot is specified by the user, the algorithm iteratively tries each input ligand as a pivot. The algorithm identifies pharmacophores by computing multiple flexible alignments between the input ligands. The resulting multiple alignments reveal spatial arrangements of consensus features shared by different subsets of input ligands. The highest-scoring ones are potential pharmacophores [65].

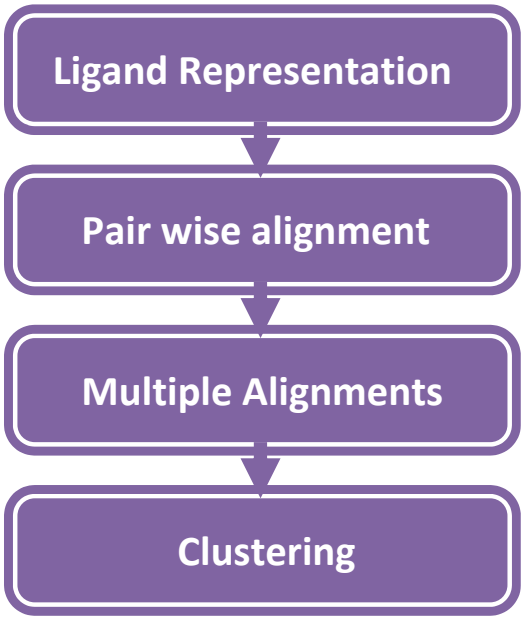


Figure 5. 1: PharmaGist Method Flow [64]

Ten active compounds from the series were selected to generate pharmacophore models. The most active and structurally rigid compound was assigned as the pivot molecule and default values were used for other settings. Models with all the 10 ligands with contribution to the pharmacophoric features were considered. Generated models were ranked with the scores and the one with highest score was selected as the best pharmacophore model.

6. CONCLUSION

A larger number of world population is affected by cancer. Wide varieties of chemotherapeutic agents are used for its cure but unfortunately cancer still remains an enigma for the world's science community. Newer targets are being identified to develop more efficacious and less toxic anticancer drugs. One such target is telomerase enzyme, responsible for making the cancerous cells almost immortal. The concentration of telomerase in the carcinogenic tissue is found very high in comparison to the normal tissues.

It was envisaged that arresting telomerase activity in cancer cells would force them to adopt apoptosis without affecting the normal tissues. So, telomerase inhibitors could offer a very big advantage over the existing anticancer agents which treat all the cells (normal or cancerous) alike. Telomerase inhibitors would arrest the growth of cancerous cells without affecting the normal cells.

A large numbers of chemical entities have been evaluated in the literature as telomerase inhibitors. These compounds bind physically with the guanine tetrad present in the G-quadruplex of telomerase thereby blocking biochemical role. These compounds belong to various categories like acridines, anthraquinones, acridones, triazines, perylenes, porphyrines etc.

In the current work, efforts have been made to develop reliable 3D-QSAR models for optimization of a few series of chemicals as telomerase inhibitors. Initially, 3D-QSAR CoMFA/CoMSIA models were developed for acridine and pentacyclic acridine derivatives. Molecular docking provided predicted bioactive conformations for alignment for derivation of 3D-QSAR models. Molecular docking studies also helped us to identify right protonating nitrogen atoms *in situ*. The QSAR models showed good correlative and predictive capabilities in terms of q^2 and r^2 values. High bootstrapped r^2 values and small standard deviations indicate that the derived models are statistically significant. 3D-QSAR models have been validated in various ways and were found to be reliable and robust models.

Another 3D-QSAR study was performed on anthraquinone and acridone derivatives to gain structural insights for selective G-quadruplex stabilization for telomerase inhibition. 3D-QSAR CoMSIA method has been applied successfully to

rationalize the anti-telomerase activity of these derivatives. The developed models showed good statistical significance in internal (q^2, cross-validation and bootstrapping) validation and performed very well in predicting the biological activity (pEC$_{50}$) of the compounds in the test set.

Next, 3D-QSAR studies were performed on triazine derivatives as G-quadruplex mediating telomerase inhibitors. In this work, CoMSIA studies were performed to gain insight into their telomerase enzyme inhibition. Highly predictive CoMSIA models using an alignment based on the bioactive conformation derived from molecular dynamics simulated annealing rather than from the minimum energy conformer of the most active member of the series. The models derived showed high quantitative predictive ability.

After developing independent 3D-QSAR models for G-quadruplex mediating telomerase inhibitors belonging to different chemical classes, a universal model for all these derivatives was derived using 169 telomerase inhibitors with different chemical classes. The data set comprises of large structural diversity and variable biological activities. This makes the data set ideal for the development of a universal model. The CoMSIA studies produced equally good models expressed in terms of the q^2 values. The predictive powers of the derived models were demonstrated to be reliable. It is hoped that this model could be useful for predicting the telomerase inhibitory activity of any compounds having aromatic central core with two side chains on either side before their actual synthesis.

All the models developed using CoMFA/CoMSIA 3D-QSAR techniques were validated using external data set, and by using well established statistical tools. The universal model developed using four different classes of chemical compounds could be used for prediction of activity of any unknown compound from these categories for its telomerase inhibiting activity.

7. REFERENCES

1. Service R.F., Surviving the blockbuster syndrome., **Science**, 2004, **303**, 1796-1799

2. Rawlins M.D., Cutting the cost of drug development?, **Nat. Rev. Drug. Discov**. **3**, 2004, 360-364

3. Andrews P.R., In Kubinyi H., Ed., 3D-QSAR In Drug Design: Theory, Methods and Applications., pp. 13-40, ESCOM, Leiden, 1993

4. Jorgensen W.L., The many roles of computation in drug discovery., **Science**, 2004, **303**, 1813-1818

5. Itzstein von M.; Wu W.-Y.; Kok G.B.; Pegg M.S.; Dyason J.C.; Jin B.; Phan T.V.; Smyth M.L.; White H.F.; Olilver S.W; Colman P.M.; Varghese J.N.; Ryan D.M.; Woods J.M.; Bethell R.C.; Hotham V.J.; Cameron J.M.; Penn C.R., Rational design of potent sialidase-based inhibitors of influenza virus replication., **Nature,** 1993, **363**, 418-423

6. Richardson, B. W., Lectures on Experimental and Practical Medicine Medical Times and Gazette., 1869, 2, 703-705

7. Free, S. M., Jr.; Wilson, J. W., A Mathematical Contribution to Structure-Activity Studies., **J. Med. Chem**., 1964, **7**, 395-399

8. Hansch, C.; Fujita, T., ρ-σ-π Analysis: A Method for the Correlation of Biological Activity and Chemical Structure., **J. Am. Chem. Soc.**, 1964, **86**, 1616-1626

9. Kubinyi, H., QSAR and 3D QSAR in drug design. Part 1: methodology., **Drug Discov., Today,** 1997, **2**, 457-467

10. Kubinyi, H., From Narcosis to Hyperspace: The History of QSAR., **Quant. Struct.-Act. Relat.,** 2002, **21**, 348-356

11. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D., Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins., **J. Am. Chem. Soc.,** 1988, **110**, 5959-5967

12. Klebe, G.; Abraham U.; Mietzner T., Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity., **J. Med. Chem.,** 1994, **37**, 4130–4146

13. Kellogg, G.E., Semus S.F; Abraham D.J., HINT: a new method of empirical hydrophobic field calculation for CoMFA., **J. Comput. Aid. Mol. Des.,** 1991, **5**, 545–552

14. Goodford, P., Multivariate characterization of molecules for QSAR analysis., **J. Chemom.,** 1996, **10**, 107–117

15. Doweyko, A.M., The hypothetical active site lattice- *in vitro* and *in vivo* explorations using a three-dimensional QSAR technique., **J. Math. Chem.,** 1991, **7**, 273–285

16. Jain, A.N.; Koile K.; Chapman D., Compass: predicting biological activities from molecular surface properties. Performance comparisons on a steroid benchmark., **J. Med. Chem.,** 1994, **37**, 2315–2327

17. Gohlke, H.; Klebe G., DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein., **J. Med. Chem.,** 2002, **45**, 4153–4170

18. Holtje, H.; Sippl, W.; Rognan, D.; Folkers, G., Molecular modeling basic principle and applications, 2nd ed.: Weinheim, Wiley-VCH Verlag Gmbh & Co. KGnA: Germany, 2003; pp 228

19 Klebe, G.; Abraham, U., On the prediction of biding properties of the drug molecules by comparative molecular field analysis., **J. Med. Chem**., 1993, **36**, 70-80

20. Cramer III, R.; DePriest, S.; Patterson, D.; Hecht, P., The developing practice of comparative molecular field analysis., In 3D QSAR in drug design: Theory Methods and Applications, Kubinyi, H, Ed.: ESCOM, Leiden, 1993; pp. 443-485

21. Kubinyi, H.; Abraham, U., Practical problems in PLS analysis. In: 3D-QSAR in drug design: Theory, Methods and Applications, Kubinyi, H., Ed; ESCOM, Leiden 1993; pp. 717-728

22. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S., Generating optimal linear PLS estimations (GLOPE): An advanced chemometric tool for handling 3D-QSAR problems**., Quant. Struct-Act. Relat.,** 1993, **12**, 9-20

23. Pastor, M.; Cruciani, G.; Clementi, S., Smart region definition: a new way to improve the predictive ability and inter-predictability of three-dimensional quantitative structure-activity relationships., **J. Med. Chem.,** 1997, **40**, 1455-1464

24. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S., Grid-Independent Descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors., **J. Med. Chem.,** 2000, **43**, 3233–3243

25. Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F., Anchor-GRIND: filling the gap between standard 3D QSAR and the GRid-Independent descriptors., **J. Med. Chem.,** 2005, **48**, 2687–2694

26. Filipponi, E.; Cruciani, G.; Tabarrini, O.; Cecchetti, V.; Fravolini, A., QSAR study and VolSurf characterization of anti-HIV quinolone library., **J. Comput. Aid. Mol. Des.,** 2001, **15**, 203–217

27. Cramer, R.D., Topomer CoMFA: a design methodology for rapid lead optimization., **J. Med. Chem.,** 2003, **46**, 374–388

28. Low C.M.; Vinter, J.G. Rationalizing the activities of diverse cholecystokinin 2 receptor antagonists using molecular field points., **J. Med. Chem.,** 2008, **51**, 565–573

29. Klebe, G.; Abraham, U. Comparative Molecular Similarity Index Analysis (CoMSIA) to Study Hydrogen Bonding Properties and to Score Combinatorial Libraries., **J. Comput.-Aided Mol. Des.,** 1999, **13**, 1-10

30 Böhm, M.; Sturzebecher, J.; Klebe, G., Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa., **J. Med. Chem.,** 1999, **42**, 458-477

31. Politzer P.; Murray J.S. In Lipkowitz K.B. and Boyd D.B., Eds., Revs. in Comp. Chem. Vol. 2, pp. 273-312,Wiley-VCH Publishers, Inc., UK, 1991

32. Wold, S.; Ruhe, A.; Wold, H.; Dunn, W.J., The partial least squares approach to generalized inverses., **SIAM J. Sci. Stat. Comput.,** 1984, **5**, 753-743

33. Leach, A.R.; Gillet, V.J., Eds. An Introduction to Chemoinformatics. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003

34. Clark, M.; Cramer, III R.D., The probability of Chance correlation using partialleast squares (PLS)., **Quant. Struct-Act. Relat.,** 1993, **12**, 137-145

35. Clark, M.; Cramer, R.D. III.; Jones, D.; Patterson, D.E.; Simeroth, P., Comparative molecular field analysis (CoMFA). 2. toward its use with 3D-structural databases., **Tetrahedron Comput. Methodol.,** 1990, **3**, 47–59

36. Novellino, E.; Fattorusso, C.; Greco, G., Use of comparative molecular field analysis and cluster analysis in series design., **Pharm. Acta Helv.,** 1995, **70**, 149–154

37. Norinder, U., Single and domain made variable selection in 3D QSAR applications., **J. Chemomet.,** 1996, **10**, 95–105

38. Kubinyi, H.; Hamprecht, F.A.; Mietzner, T., Three-dimensional quantitative similarity–activity relationships (3D QSAR) from SEAL similarity matrices., **J. Med. Chem.,** 1998, **41**, 2553–2564

39. Golbraikh, A.; Tropsha, A., Beware of q^2 !, **J. Mol. Graph. Mod.,** 2002, **20**, 269–276

40. Roy P. P.; Roy, K., On Some Aspects of Variable Selection for Partial Least Squares Regression Models., **QSAR Comb. Sci.,** 2008, **27**, 302-313

41. Roy, P. P.; Paul, S.; Mitra, I.; Roy, K., On Two Novel Parameters for Validation of Predictive QSAR Models., **Molecules,** 2009, **14**, 1660-1701

42. Gund, P., Three-dimensional pharmacophoric pattern searching., **Prog. Mol. Subcell Biol.,** 1977, **5**, 117-143

43. Kier, L.B., Molecular conformation. Molecular orbital theory in drug research; Academy press: New York, 1971, pp 164-195

44. Wermuth, C.; Ganellin, E.; Lindberg, P.; Mitscher, L., Glossary terms used in Medicinal Chemistry (IUPAC Recommendations 1998)., **Pure and Appl. Chem.,** 1998, **70**, 1129-1143

45. (a) Gu¨ner, O.F., (ed.) (2000) Pharmacophore Perception, Development, and Use in Drug Design. International University Line, La Jolla, CA, (b) Dror,O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H., Predicting molecular interactions in silico: I. an updated guide to pharmacophore identication and its applications to drug design., **Front. Med. Chem.,** 2006, **3**, 551–584

46. Rognan, D., Chemogenomic Approaches to Rational Drug Design., **Br. J. Pharmacol.,** 2007, **152**, 38–52

47. Klabunde, T., Chemogenomic Approaches to Drug Discovery: Similar Receptors Bind Similar Ligands., **Br. J. Pharmacol.,** 2007, **152**, 5–7

48. van Drie, J. Pharmacophore discovery-lessons learned., **Curr. Pharm. Des.,** 2003, **9**, 1649-1664

49. Gu¨nther, S.; Senger, C.; Michalsky, E.; Goede, A.; Preissner, R., Representation of target-bound drugs by computed conformers: implications for conformational libraries., **BMC Bioinf.,** 2006, **7**, 293

50. Finn, P. W.; Kavraki, L. E.; Latombe.; J.-C.; Motwani, R.; Shelton, C.; Venkatasubramanian, S.; Yao, A., RAPID: Randomized Pharmocophore Indentification for Drug Design., **Comput. Geom.: Theory Appl.,** 1998, **10**, 263–272

51. Holliday, J.; Willet, P., Using a Genetic Algorithm to Identify Common Structural Features in Sets of Ligands., **J. Mol. Graphics Modell.,** 1997, **15**, 203–253

52. Martin, Y.; Bures, M.; Dahaner, E.; DeLazzer, J.; Lico, I.; Pavlik, P., A fast new approach to pharmacophore mapping and its Application to Dopaminergic and Benzodiazepine agonists., **J. Comput.-Aided Mol. Des.** 1993, **7**, 83–102

53. Martin, Y. C., In Pharmacophore Perception, Development, and Use in Drug Design; International University Line: La Jolla, CA, U.S.A., 2000; Chapter DISCO: What We Did Right and What We Missed, pp 49-68

54. Dixon, S.; Smondyrev, A.; Knoll, E.; Rao, S.; Shaw, D.; Friesner, R., PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results., **J. Comput.-Aided Mol. Des.,** 2006, **20**, 647–671

55. Gu¨ner, O. F.; Clement, O.; Kurogi, Y., Pharmacophore Modeling and Three Dimensional Database Searching for Drug Design Using Catalyst: Recent Advances., **Curr. Med. Chem.** 2004, 11, 2991–3005

56. Clement, O. A.; Mehl, A. T., In Pharmacophore Perception, Development, and Use in Drug Design; International University Line: La Jolla, CA, U.S.A., 2000

57. Chapter HipHop: Pharmacophores Based on Multiple Common-Feature Alignments., pp 69-84

58. Barnum, D.; Greene, J.; Smellie, A.; Sprague, P., Identification of Common Functional Configurations among Molecules., **J. Chem. Inf. Comput. Sci.,** 1996, **36**, 563–571

59. Kurogi, Y.; Guner, O. F., Pharmacophore Modeling and Three-Dimensional Database Searching for Drug Design Using Catalyst., **Curr. Med. Chem.,** 2001, **8**, 1035–1055

60. Li, H.; Sutter, J.; Hoffmann, R., In Pharmacophore Perception, Development, and Use in Drug Design; International University Line: La Jolla, CA, U.S.A., 2000; Chapter HypGen: An Automated System for Generating 3D Predictive Pharmacophore Models., pp 171-189

61. Feng, J.; Sanil, A.; Young, S., PharmID: Pharmacophore Identification Using Gibbs Sampling., **J. Chem. Inf. Model.,** 2006, **46**, 1352–1359

62. Zhu, F.; Agrafiotis, D., Recursive Distance Partitioning Algorithm for Common Pharmacophore Identification., **J. Chem. Inf. Model.,** 2007, **47**, 1619–1625

63. Podolyan, Y.; Karypis, G., Common Pharmacophore Identification Using Frequent Clique Detection Algorithm., **J. Chem. Inf. Model.,** 2009, **49**, 13–21

64. Schneidman-Duhovny, D.; Dror, O.; Inbar, Y.; Nussinov, R.; Wolfson, H. J., PharmaGist: a webserver for ligand-based pharmacophore detection., **Nucleic Acids Res.,** 2008, Web Server issue W223–W228

65. Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J., Novel Approach for Efficient Pharmacophore-Based Virtual Screening: Method and Applications., **J. Chem. Inf. Model.,** 2009, **49**, 2333–2343

66. Shoichet, B.K.; Bodian , D.L.; Kuntz , I.D., Docking by least-squares fitting of molecular surface patterns., **J. Comput. Chem.,** 1992, **13**, 380-397

67. McGann, M.; Almond H.; Nicholls, A., Guassian docking functions. **Biopolymers**, 2003, **68**, 76-90

68. Rarey, M.; Kramer, B.; Lengauer, T., A last flexible docking method using an incremental construction algorithm., **J. Mol. Biol.,** 1996, **261**, 470-489

69. Jones, G.; Willett, P.; Glen R.; Leach A.; Taylor, R., Development and validation of a genetic algorithm for flexible docking., **J. Mol. Biol.,** 1997, **267**, 727-748

70. Baxter, C.; Murray, C.; Clark, D.; Westhead, D.; Eldridge, M., Flexible docking using Tabu search and an empirical estimate of binding affinity., **Proteins,** 1998, **33**, 367-382

71. Hart, T.; Read, R. A., A multiple-start Monte Carlo docking method., **Proteins**, 1992, **13**, 206-222

72. Davis A.; Teague, S. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis., **Angew. Chem.,** Intl. Ed., 1999, **38**, 737-749

73. Fritz, T.; Tondi, D.; Finer-Moore, J.; Costi, M.; Stroud, R., Predicting and harnessing protein flexibility in the design of species-specific inhibitors of thymidylate synthase., **Chem. Biol.,** 2001, **10**, 981-995

74. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking consideration protein structure variations., **J. Mol. Biol.,** 2001, **308**, 377-395

75. Jain AN., Scoring functions for protein-ligand docking., **Curr. Protein Pept. Sci.,** 2006, **7**, 407-420

76. Schenecke, V.; Kuhn, L., Virtual screening with salvation and ligand-induced complemetarity., **Perspect. Drug. Discov. Des.,** 2000, **20**, 171-190

77. Lyne, P., Structure-based virtual screening: an overview., **Drug Discov. Today,** 2002, **7**, 1047-1055

78. Kollman, P., Free energy calculations: Application to chemical and biochemical phenomena., **Chem. Rev.,** 1993, **93**, 2395-2417

79. Wang, R.; Lu, Y.; Wang, S., Comparative evaluation of 11 scoring functions for molecular docking., **J. Med. Chem.,** 2003, **46**, 2287-2303

80. Bissantz, C.; Folkers G.; Roganm, D., Protein-based virtual screening of chemical databases. I. Evaluation of different docking scoring combinations., **J. Med. Chem.,** 2000, **43**, 4759-4767

81. Charifson, P.; Corkery, J.; Murko, M.; Walters, W., Consensus scoring: A method for obtaining improved hit rates from docking databases of three dimensional structures into proteins., **J. Med. Chem.,** 1999, **42**, 5100-5109

82. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4., **J. Comput-Aid. Mol. Des.,** 1996, **10**, 293-304

83. Wu, G.; Robertson, D.; Brooks, C., III; Vieth, M., Detailed analysis of Grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm., **J. Comput. Chem.,** 2003, **24**, 1549–1562

84. Ewing, T.; Makino, S.; Skillman., A.; Kuntz, I., Search strategies for automated molecular docking of flexible molecule databases., **J. Comput. Aided. Mol. Des.,** 2001, **15**, 411-428

85. Sousa, S.F.; Fernandes, P.A.; Ramos, M.J., Protein-Ligand Docking: Current Status and Future Challenges., **Proteins,** 2006, **65**, 15-26

86. Kuo, C.; Assefa, H.; Kamath, S.; Brzozowski, Z.; Slawinski, J.; Saczewski, F.; Buolamwini, J.; Neamati, N., Application of CoMFA and CoMSIA 3D-QSAR and docking studies in optimization of mercaptobenzenesulphonamides as HIV-I integrase inhibitors., **J. Med. Chem.,** 2004, **47**, 385-399

87. Jones, G.; Willett, P.; Glen R., A genetic algorithm for flexible molecular overlay and pharmacophore elucidation., **J. Comput. Aided. Mol. Des.,** 1995, **9**, 532-549

88. Friesner, R.; Banks, J.; Murphy, R.; Halgren, T.; Klicic, J.; Mainz, D.; Repasky, M.; Knoll, E.; Shelly, M.; Perry, J.; Shaw, D.; Francis, P.; Shenkin, P., Glide: a new approach for rapid, accurate, docking and scoring. 1. Method and assessment of docking accuracy., **J. Med. Chem.,** 2004, **47**, 1739-1749

89. Halgren, T.; Murphy, R..; Friesner, R.; Bread, H.; Frye, L.; Pollard W.; Banks., J., Glide: A new approach for rapid, accurate, docking and scoring. 2.

Enrichment factors in database screening., **J. Med. Chem.,** 2004, **47**, 1750-1759

90. Greider C.W.; Blackburn E.H., Identification of specific telomere transferase terminal activity in Tetrahymena extract., **Cell,** 1985, **43**, 405-413

91. Greider C.W.; Blackburn E.H., A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis., **Nature,** 1989, **337**, 331-337

92. Banik, S.S.; Guo, C.; Smith, A.C.; Margolis, S.S.; Richardson, D.A.; Tirado, C.A.; Counter, C.M., C-terminal regions of the human telomerase catalytic subunit essential for in vivo enzyme activity., **Mol. Cell. Biol.,** 2002, **22**, 6234-6246

93. Kelleher, C.; Teixeira, M.T.; Forstemann, K.; Lingner, J., Telomerase: biochemical considerations for enzyme and substrate., **Trends Biochem. Sci.,** 2002, **27**, 572-579

94. Chen, J.L.; Blasco, M.A.; Greider, C.W., Secondary structure of vertebrate telomerase RNA., **Cell,** 2000, **100**, 503-514

95. Mitchell, J.R.; Wood, E.; Collins K., A telomerase component is defective in the human disease dyskeratosis congenital., **Nature,** 1999, **402**, 551-555

96. Cohen, S.B.; Graham, M.E.; Lovrecz, G.O.; Bache, N.; Robinson, P.J.; Reddel, R.R., Protein composition of catalytically active human telomerase from immortal cells., **Science,** 2007, **315**, 1850-1853

97. Wenz, C.; Enenkel, B.; Amacker, M.; Kelleher, C.; Damm, K.; Lingner, J., Human telomerase contains two cooperating telomerase RNA molecules., **EMBO J.,** 2001, **20**, 3526-3534

98. Beattie, T.L.; Zhou, W.; Robinson, M.O.; Harrington, L., Functional multimerization of the human telomerase reverse transcriptase., **Mol. Cell. Biol.,** 2001, **21**, 6151-6160

99. Forsyth, N.R.; Wright, W.E.; Shay, J.W., Telomerase and differentiation in multicellular organisms: turn it off, turn it on and turn it off again., **Differentiation,** 2002, **69**, 188-197

100. Wright, W.E.; Piatyszek, M.A.; Rainey, W.E.; Byrd, W.; Shay, J.W., Telomerase activity in human germline and embryonic tissues and Cells., **Dev. Genet.,** 1996, **18**, 173-179

101. Masutomi, K.; Yu, E.Y.; Khurts, S.; Ben-Porath, I.; Currier, J.L.; Metz, G.B.; Brooks, M.W.; Kaneko, S.; Murakami, S.; DeCaprio, J.A.; Weinberg, R.A.;

Stewart, S.A.; Hahn, W.C., Telomerase maintain telomere structure in normal human cells., **Cell,** 2003, **114**, 241-253

102. Read M.; Harrison, R.J.; Romagnoli, B.; Tanious, F.A.; Gowan, S.H.; Reszka, A.P.; Wilson, W.D.; Kelland, L.R.; Neidle S., Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors., **Proc. Natl. Acad. Sci. USA.,** 2001, **98**, 4844-9

103. Masutomi, K.; Possemato, R.; Wong, J.M.Y.; Currier, J.L.; Tothova, Z.; Manola, J.B.; Ganesan, S.; Lansdorp, P.M.; Collins, K.; Hahn, W.C., The telomerase reverse transcriptase regulates chromatin state and DNA damage responses., **Proc. Natl. Acad. Sci. USA,** 2005, **102**, 8222-8227

104. Cohen, S.B.; Graham, M.E.; Lovrecz, G.O.; Bache, N.; Robinson, P.J.; Reddel, R.R., Protein composition of catalytically active human telomerase from immortal cells., **Science,** 2007, **315**, 1850-1853

105. Shay J.W.; Wright, W.E., Senescence and immortalization: role of telomeres and telomerase., **Carcinogenesis**, 2005, **26**, 867-874

106. McClintock, B., The fusion of broken ends of Chromosomes Following Nuclear Fusion., **Proc. Natl. Acad. Sci. USA,** 1942, **28**, 458-463

107. Harley, C.; Futcher, A.B.; Greider, C.W. Telomeres shorten during ageing of human fibroblasts., **Nature,** 1990, **346**, 866–868

108. Pardue, M-L., Telomeres and telomerase: more than the end of the line., **Chromasoma,** 1999, **108**, 73–82

109. Griffith, J.D.; Comeau, L.; Rosenfield, S.; Stansel, R.M.; Bianchi, A.; Moss, H.; de Lange, T., Mammalian telomeres end in a large duplex loop., **Cell,** 1999, **97**, 503–514

110. Shay, J.W.; Wright, W.E., Telomeres and telomerase in the regulation of human cellular aging., **Mol. Biol. Aging,** 1999, **44**, 148–158

111. Counter, C.M.; Hirte, H.W.; Bacchetti, S.; Harley, C.B., Telomerase activity in human ovarian carcinoma., **Proc. Natl. Acad. Sci. USA,** 1994, **91**, 2900–2904

112. Raymond, E.; Sun, D.; Chen, S.F.; Windle, B.; Von Hoff, D.D., Agents that target telomerase and telomeres., **Curr. Opin. Biotech.,** 1996, **7**, 583–91

113. Rhyu, M.S., Telomeres, telomerase and immortality., **J. Natl. Cancer Inst.,** 1995, **87**, 884–94

114. Hiyama, E.; Hiyama, K.; Yokoyama, T.; Matsuura, Y.; Piatyszek, M.A.; Shay, J.W., Correlating telomerase activity levels with neuroblastoma outcomes., **Nat. Med.,** 1995, **1**, 249–55

115. (a)Greider, C.W., Telomerase activity, cell proliferation and cancer., **Proc. Natl. Acad. Sci. USA,** 1998, 95, 90–92, (b) Autexier, C.; Greider, C.W., Telomerase and cancer: revisiting the telomere hypothesis., **TIBS,** 1996, **21**, 387–91

116. Meaghan, P.G.; Woodring E.W.; Jerry, W.S., Telomerase in cancer and aging., **Critical Rev. Onc./Hemat.,** 2002, **41**, 29–40

117. Lavelle, F.; Riou, J.; Laoui, A.; Mailliet, P., Telomerase: a therapeutic target for the third millennium?, **Critical Rev. Onc/Hemat.,** 2000, **34**, 111–126

118. Blasco, M.A.; Lee, H.-W.; Hande, P.M.; Samper, E.; Lansdorp, P.M.; DePinho, R.A.; Greider, C.W., Telomere shortening and tumor formation by mouse cells lacking telomerase RNA., **Cell,** 1997, **91**, 25–34

119. Greenberg, R.A.; Chin, L.; Femino, A.; Lee, K.H.; Gottlieb, G.J.; Singer, R.H.; Greider, C.W.; DePinho, R.A., Short dysfunctional telomeres impair tumorigenesis in the INK4a(delta2/3) cancer-prone mouse., **Cell,** 1999, **97**, 515–525

120. Nakamura, T.M.; Morin, G.B.; Chapman, K.B.; Weinrich, S.L.; Andrews, W.H.; Lingner, J.; Harley, C.B.; Cech, T.R., Telomerase catalytic subunit homologs from fission yeast and human., **Science,** 1997, **277**, 955–9

121. Meyerson, M.; Counter, C.M.; Eaton, E.N.; Ellisen, L.W.; Steiner, P.; Caddle, S.D.; Ziaugra, L.; Beijersbergen, R.L.; Davidoff, M.J.; Liu, Q.; Bacchetti, S.; Haber, D.A.; Weinberg, R.A., hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells during immortalization., **Cell,** 1997, **90**, 785–795

122. Kilian, A.; Bowtell, D.D.; Abud, H.E.; Hime, G.R.; Venter, D.J.; Keese, P.K.; Duncan, E.L; Reddel, R.R.; Jefferson, R.A., Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types., **Human Mol. Gen.,** 1997, **6**, 2011–2019

123. Weinrich, S.L.; Pruzan, R.; Ma, L.; Ouellette, M.; Tesmer, V.M.; Holt, S.E.; Bodnar, A.G.; Lichtsteiner, S.; Kim, N.W.; Trager, J.B.; Taylor, R.D.; Carlos, R.; Andrews, W.H.; Wright, W.E.; Shay, J.W.; Harley, C.B.; Morin, G.B., Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT., **Nat. Genet.,** 1997, **17**, 498–502

124. Feng, J.; Funk, W.D.; Wang, S.S.; Weinrich, S.L.; Avilion, A.A.; Chiu, C.P.; Adams, R.R.; Chang, E.; Allsopp, R.C.; Yu, J.; et al., The RNA component of human telomerase., **Science,** 1995, **2**, 69, 1236–1241

125. White, L. K.; Wright, W. E.; Shay, J. W., Telomerase inhibitors., **Trends in Biotech.** 2001, 114-120

126. Artandi, S.E.; DePinho, R.A., Mice without telomerase: what can they teach us about human cancer?, **Nat. Med.,** 2000, **6**, 852–855

127. Yokoyama, Y.; Takahashi, Y.; Shinohara, A.; Lian, Z.; Wan, X.; Niwa, K.; Tamaya, T., Attenuation of telomerase activity by a hammerhead ribozyme targeting the template region of telomerase RNA in endometrial carcinoma cells., **Cancer Res.,** 1998, **58**, 5406–5410

128. Folini, M.; Colella, G.; Villa, R.; Lualdi, S.; Daidone, M.G.; Zaffaroni, N., Inhibition of telomerase activity by a hammerhead ribozyme targeting the RNA component of telomerase in human melanoma cells., **J. Invest. Dermatol.,** 2000, **114**, 259–267

129. Hahn, W.C.; Stewart, S.A.; Brooks, M.W.; York, S.G.; Eaton, E.; Kurachi, A.; Beijersbergen, R.L.; Knoll, J.H.; Meyerson, M.; Weinberg, R.A., Inhibition of telomerase limits the growth of human cancer cells., **Nat. Med.,** 1999, **5**, 1164-70

130. Zhang, X.; Mar, V.; Zhou, W.; Harrington, L.; Robinson, M.O., Telomere shortening and apoptosis in telomerase-inhibited human tumor cells., **Genes Dev.,** 1999, **13**, 2388-99

131. Saretzki, G.; Ludwig, A.; von Zglinicki, T.; Runnebaum, I.B., Ribozyme-mediated telomerase inhibition induces immediate cell loss but not telomere shortening in ovarian cancer cells., **Cancer Gene Ther.,** 2001, **8**, 827-34

132. Giridharan, P.; Somasundaram, S.T.; Perumal, K.; Vishwakarma, R.A.; Karthikeyan, N.P.; Velmurugan, R.; Balakrishnan, A., Novel substituted methelenedioxy lignan suppresses proliferation of cancer cells by inhibiting telomerase and activation of C-myc and caspases leading to apoptosis., **Br. J. Cancer**, 2002, **87**, 98-105

133. Naasani, I.; Seimiya, H.; Yamori, T.; Tsuruo, T., FJ5002: A potent telomerase inhibitor identified by exploiting the diseases oriented screening program with COMPARE analysis., **Cancer Res.,** 1999, **59**, 4004-11

134. Gellert, M.; Lipsett, M.N.; Davies, D.R., Helix formation by guanylic acid., **Proc. Natl. Acad. Sci. USA,** 1962, **48**, 2013–2018

135. Duquette, M.L.; Handa, P.; Vincent, J.A.; Taylor, A.F.; Maizels, N., Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA., **Genes and Dev.,** 2004, **18**, 1618-1629

136. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H.J.; Plu¨ckthun, A., In vitro generated antibodies specific for telomeric guanine quadruplex DNA react with Stylonychia lemnae macronuclei., **Proc. Nat. Acad. Sci. USA,** 2001, **98**, 8572-8577

137. Paeschke, K..; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H., Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo., **Nat. Struct. and Mol. Biol.,** 2005, **12**, 847-854

138. Zahler, A.M.; Williamson, J.R.; Cech, T.R.; Prescott, D.M., Inhibition of telomerase by G-quartet DNA structures., **Nature,** 1991, **350**, 718-720

139. Granotier, C.; Pennarun, G.; Riou, L. Hoffschir, F.; Gauthier, L.R.; DeCian, A.; Gomez, D.; Mandine, E.; Riou, J.F.; Mergny, J.L.; Mailliet, P.; Dutrillaux, B.; Boussin, F.D., Preferential binding of a Gquadruplex ligand to human chromosome ends., **Nucleic Acid Res.,** 2005, **33**, 4182-4190

140. Hurley, L. H., DNA and its associated processes as targets for cancer therapy., **Nat. Rev. Cancer,** 2002, **2**, 188–200

141. Tian-miao, O.; Yu-jing Lu; Jia-heng Tan; Zhi-shu Huang; Kwok-Yin Wong; Lian-quan Gu., G-Quadruplexes: Targets in Anticancer Drug Design., **ChemMedChem,** 2008, **3**, 690-713

142. Wheelhouse, R. T.; Jennings, S. A.; Phillips, V. A.; Pletsas, D.; Murphy, P. M.; Garbett, N. C.; Chaires, J. B.; Jenkins, T. C., Design, synthesis, and evaluation of novel biarylpyrimidines: a new class of ligand for unusual nucleic acid structures., **J. Med. Chem.,** 2006, **49**, 5187– 5198

143. De Cian, A.; DeLemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D. Highly efficient G-quadruplex recognition by bisquinolinium compounds., **J. Am. Chem. Soc.,** 2007, **129**, 1856–1857

144. Julian L.H., Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes., **Chem. Soc. Rev**., 2008, **37**, 1375–1384

145. Sun, D.; Thompson, B.; Cathers,; B. E., Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H., Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound., **J. Med. Chem.,** 1997, **40** , 2113–2116

146. Perry, P. J.; Gowan, S. M.; Reszka, A. P.; Polucci, P.; Jenkins, T. C.; Kelland, L. R.; Neidle, S., 1,4- and 2,6-Disubstituted Amidoanthracene-9,10-dione

Derivatives as Inhibitors of Human Telomerase., **J. Med. Chem.,** 1998, **41**, 3253–3260

147. Cairns, D.; Michalitsi, E.; Jenkins, T. C; Mackay, S. P., Molecular Modelling and Cytotoxicity of Substituted Anthraquinones as Inhibitors of Human Telomerase., **Bioorg. Med. Chem.,** 2002, **10**, 803-807

148. Huang, H. S.; Chou, C.; Guo, C. L.; Yuan, C. L.; Lu, Y. C.; Shieh, F. Y.; Lin, J. J., Human telomerase inhibition and cytotoxicity of regioisomeric disubstituted amidoanthraquinones and aminoanthraquinones., **Bioorg. Med. Chem.,** 2005, **13** 1435-1444

149. Sissi, G. Z, C.; Lucatello, L.; Pivetta, C.; Cadamuro, S. A.; Fox, K. R.; Neidle, S.; Palumbo, M., Aminoacyl−Anthraquinone Conjugates as Telomerase Inhibitors: Synthesis, Biophysical and Biological Evaluation., **J. Med. Chem.,** 2008, **51**, 5566–5574

150. Zagotto, G.; Sissi, C.; Moro, S.; Ben, D. D.; Parkinson, G. N.; Fox, K. R.; Neidle, S.; Palumbo M., Amide bond direction modulates G-quadruplex recognition and telomerase inhibition by 2,6 and 2,7 bis-substituted anthracenedione derivatives., **Bioorg. Med. Chem.,** 2008, **16**, 354-361

151. Huang, H. S.; Huang, K. F.; . Li, C. L.; Huang, Y. Y.; Chiang, Y. H.; Huang, F. C.; Lin, J. J., Synthesis, human telomerase inhibition and anti-proliferative studies of a series of 2,7-bis-substituted amido-anthraquinone derivatives., **Bioorg. Med. Chem.,** 2008, **16**, 6976-6986

152. Shchekotikhin, A. E.; Glazunova, V. A.; Dezhenkova, L. G.; Luzikov, Y. N.; Sinkevich, Y. B.; Kovalenko, L. V.; Buyanov, V. N.; Balzarini, J.; Huang, F. C.; Lin, J. J.; Huang, H. S.; Shtil, A. A.; Preobrazhenskaya, M. N., Synthesis and cytotoxic properties of 4,11-bis[(aminoethyl)amino]anthra[2,3-*b*]thiophene-5,10-diones, novel analogues of antitumor anthracene-9,10-diones., **Bioorg. Med. Chem.,** 2009, **17**, 9, 1861-1869

153. Perry, P. J.; Read, M. A.; Davies, R. T.; Gowan, S. M.; Reszka, A. P.; . Wood, A. A.; Kelland, L. R.; Neidle, S., 2,7-Disubstituted Amidofluorenone Derivatives as Inhibitors of Human Telomerase., **J. Med. Chem.,** 1999, **42**, 2679–2684

154. Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Neidle S., Human telomerase inhibition by substituted acridine derivatives., **Bioorg. Med. Chem. Lett.,** 1999, **9**, 2463-2468

155. Read, M. A.; Wood, A. A.; Harrison, R. J.; Gowan, S. M.; Kelland, L. R.; Dosanjh, H. S.; Neidle, S., Molecular Modeling Studies on G-Quadruplex Complexes of Telomerase Inhibitors: Structure−Activity Relationships., **J. Med. Chem.,** 1999, **42** , 4538–4546

156. Read, M.; Harrison, R.J.; Romagnoli, B.; Tanious, F.A.; Gowan, S.H.; Reszka, A.P.; Wilson, W.D.; Kelland, L.R.; Neidle, S., Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors., **Proc. Natl. Acad. Sci. USA**, 2001, **98**, 4844-4849

157. Heald, R. A.; Modi, C.; Cookson, J. C.; Hutchinson, I.; Laughton, C. A.; Gowan, S. M.; Kelland, L. R.; Stevens, M. F. G., Antitumor Polycyclic Acridines. 8.1 Synthesis and Telomerase-Inhibitory Activity of Methylated Pentacyclic Acridinium Salts., **J. Med. Chem.,** 2002, **45**, 590–597

158. Heald, R. A.; Stevens, M. F. G., Antitumour polycyclic acridines. Palladium(0) mediated syntheses of quino[4,3,2-*kl*]acridines bearing peripheral substituents as potential telomere maintenance inhibitors., **Org. Biomol. Chem.,** 2003, **1**, 3377-3389

159. Cheng, M. K.; Modi, C.; Cookson, J. C.; Hutchinson, I.; Heald, R. A.; McCarroll, A. J.; Missailidis, S.; Tanious, F.; Wilson, W. D.; Mergny, J. L.; Laughton, C. A.; Stevens, M. F. G., Antitumor Polycyclic Acridines. 20.(1) Search for DNA Quadruplex Binding Selectivity in a Series of 8,13-Dimethylquino[4,3,2-*kl*] acridinium Salts: Telomere- Targeted Agents., **J. Med. Chem.,** 2008, **51**, 963–975

160. Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S., Trisubstituted Acridine Derivatives as Potent and Selective Telomerase Inhibitors., **J. Med. Chem.,** 2003, **46**, 4463–4476

161. Schultes, C. M.; Guyen, B.; Cuesta, J.; Neidle, S., Synthesis, biophysical and biological evaluation of 3,6-bis-amidoacridines with extended 9-anilino substituents as potent G-quadruplex-binding telomerase inhibitors., **Bioorg. Med. Chem. Lett.,** 2004, **14**, 4347-4351

162. Harrison, R. J.; Reszka, A. P.; Haider, S. M.; Romagnoli, B.; Morrell, J.; Read, M. A.; Gowan, S. M.; Incles, C. M.; Kelland, L. R.; Neidle, S., Evaluation of by disubstituted acridone derivatives as telomerase inhibitors: the importance of G-quadruplex binding., **Bioorg. Med. Chem. Lett.,** 2004, **14**, 5845-5849

163. Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S., Trisubstituted Acridines as G-quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation., **J. Med. Chem.,** 2006, **49**, 582–599

164. Martins, C.; Gunaratnam, M.; Stuart, J.; Makwana, V.; Greciano, O.; Reszka, A. P.; Kelland, L. R.; Neidle, S., Structure-based design of benzylamino-

acridine compounds as G-quadruplex DNA telomere targeting agents., **Bioorg. Med. Chem. Lett.,** 2007, **17**, 2293-2298

165. Fu, Y. T.; Keppler, B. R.; Soares, J.; Jarstfer, M. B. BRACO19 analog dimers with improved inhibition of telomerase and hPot 1., **Bioorg. Med. Chem.,** 2009, **17**, 2030-2037

166. Cochard, M. L.; Kim, Y. M.; Brassart, B.; Riou, J. F.; Laronze, J. Y.; Sapi, J. Synthesis and biological evaluation of novel 4,5-bis(dialkylaminoalkyl)-substituted acridines as potent telomeric G-quadruplex ligands., **Eur. J. Med. Chem.,** 2009, **44**, 3880-3888

167. Perry, P. J.; Reszka, A. P.; Wood, A. A.; Read, M. A.; Gowan, S. M.; Dosanjh, H. S.; Trent, J. O.; Jenkins, T. C.; Kelland, L. R.; Neidle, S., Human Telomerase Inhibition by Regioisomeric Disubstituted Amidoanthracene-9,10-diones., **J. Med. Chem.,** 1998, **41**, 4873–4884

168. Alberti, P.; Schmitt, P.; Nguyen, C. H.; Rivalle, C.; Hoarau, M.; Grierson, D. S.; Mergny, J. L., Benzoindoloquinolines interact with DNA tetraplexes and inhibit telomerase., **Bioorg. Med. Chem. Lett.,** 2002, **12**, 1071-1074

169. Guyen, B.; Schultes, C. M.; Hazel, P.; Mann, J.; Neidle, S., Synthesis and evaluation of analogues of 10*H*-indolo[3,2-*b*]quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase., **Org. Biomol. Chem.,** 2004, **2**, 981-988

170. Zhou, J. L.; Lu, Y. J.; Ou, T. M.; Zhou, J. M.; Huang, Z. S.; Zhu, X. F.; Du, C. J.; Bu, X. Z.; Ma, L.; Gu, L. Q.; Li, Y. M.; Chan, A. S. C., Synthesis and Evaluation of Quindoline Derivatives as G-Quadruplex Inducing and Stabilizing Ligands and Potential Inhibitors of Telomerase., **J. Med. Chem.,** 2005, **48**, 7315–7321

171. Zhou, J.M.; Zhu, X.F.; Lu, Y.J.; Deng, R.; Huang, Z.S.; Mei, Y.P.; Wang, Y.; Huang, W.L.; Liu, Z.C.; Gu, L.Q.; Zeng, Y.X, Senescence and telomere shortening induced by novel potent G-quadruplex interactive agents, quindoline derivatives, in human cancer cell lines., **Oncogene,** 2006, **25**, 503-11

172. Lu, Y. J.; Ou, T. M.; Tan, J. H.; Hou, J. Q.; Shao, W. Y.; Sun, D. P. N.; Wang, X. D.; Wu, W. B.; Bu, X. Z.; Huang, Z. S.; Ma, D. L.; Wong, K. Y.; Gu, L. Q., 5-N-Methylated Quindoline Derivatives as Telomeric G-Quadruplex Stabilizing Ligands: Effects of 5-N Positive Charge on Quadruplex Binding Affinity and Cell Proliferation., **J. Med. Chem.,** 2008, **51**, 6381–6392

173. Sasaki, S.; Ehara ,T.; Sakata, I.; Fujino, Y.; Harada, N.; Kimura, J.; Nakamura, H.; Maeda, M., Development of novel telomerase inhibitors based on a Bisindole unit., **Bioorg. Med. Chem. Lett.,** 2001, **11**, 583-585

174. Sasaki, S.; Ehara, T.; Alam, Md. R.; Fujino, Y.; Harada, N.; Kimura, J.; Nakamura, H.; Maeda, M. Solid-Phase synthesis of a library constructed of aromatic phosphate, long alkyl chains and tryptophane components, and identification of potent dipeptide telomerase inhibitors., **Bioorg. Med. Chem. Lett.,** 2001, **11**, 2581-2584

175. Fedoroff, O. Y.; Salazar, M.; Han, H.; Chemeris, V. V.; Kerwin, S. M.; Hurley, L. H., NMR-Based Model of a Telomerase-Inhibiting Compound Bound to G-Quadruplex DNA., **Biochem.,** 1998, **37**, 12367–12374

176. Han, H.; Cliff, C. L.; Hurley, L. H., Accelerated Assembly of G-Quadruplex Structures by a Small Molecule., **Biochem.** 1999, **38**, 6981–6986

177. Kerwin, S. M.; Chen, G.; Kern, J. T.; Thomas, P. W., Perylene Diimide G-Quadruplex DNA Binding Selectivity is Mediated by Ligand Aggregation., **Bioorg. Med. Chem. Lett.,** 2002, **12**, 447-450

178. Rossetti, L.; Franceschin, M.; Bianco, A.; Ortaggi, G.; Savino, M. Perylene diimides with different side chains are selective in inducing different G-Quadruplex DNA structures and in inhibiting telomerase., **Bioorg. Med. Chem. Lett.,** 2002, **12**, 2527-2533

179. Rossetti, L.; Franceschin, M.; Schirripa, S.; Bianco, A.; Ortaggi, G.; Savino, M., Selective interactions of perylene derivatives having different side chains with inter- and intramolecular G-quadruplex DNA structures. A correlation with telomerase inhibition., **Bioorg. Med. Chem. Lett.,** 2005, **15**, 413-420

180. Sissi, C.; Lucatello, L.; Krapcho, A. P.; Maloney, D. J.; Boxer, M. B.; Camarasa, M. V.; Pezzoni, G.; Menta, E.; Palumbo, M. Tri-, tetra- and heptacyclic perylene analogues as new potential antineoplastic agents based on DNA telomerase inhibition., **Bioorg. Med. Chem.,** 2007, **15**, 555-562

181. Franceschin, M.; Pascucci, E.; Alvino, A.; D'Ambrosio, D.; Bianco, A.; Ortaggi, G.; Savino, M., New highly hydrosoluble and not self-aggregated perylene derivatives with three and four polar side-chains as G-quadruplex telomere targeting agents and telomerase inhibitors., **Bioorg. Med. Chem. Lett.,** 2007, **17**, 2515-2522

182. Franceschin, M.; Lombardo, C. M.; Pascucci, E.; D'Ambrosio, D.; Micheli, E.; Bianco, A.; Ortaggi, G.; Savino, M., The number and distances of positive charges of polyamine side chains in a series of perylene diimides significantly influence their ability to induce G-quadruplex structures and inhibit human telomerase., **Bioorg. Med. Chem.,** 2008, **16**, 2292-2304

183. Micheli, E.; Lombardo, C. M.; D'Ambrosio, D.; Franceschin, M.; Neidle, S.; Savino, M., Selective G-quadruplex ligands: The significant role of side chain

charge density in a series of perylene derivatives., **Bioorg. Med. Chem. Lett.**, 2009, **19**, 3903-3908

184. Wheelhouse, R. T.; Sun, D.; Han, H.; Han, F. X.; Hurley, L. H., Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(*N*-methyl-4-pyridyl)porphine with Quadruplex DNA., **J. Am. Chem. Soc.,** 1998, **120**, 3261–3262

185. Shi, D. F.; Wheelhouse, R. T.; Sun, D.; Hurley, L. H., Quadruplex-Interactive Agents as Telomerase Inhibitors: Synthesis of Porphyrins and Structure−Activity Relationship for the Inhibition of Telomerase., **J. Med. Chem.,** 2001, **44**, 4509–4523

186. Maraval, A.; Franco, S.; Vialas, C.; Pratviel, G.; Blasco, M. A.; Meunier, B. Porphyrin–aminoquinoline conjugates as telomerase inhibitors., **Org. Biomol. Chem.,** 2003, **1**, 921-927

187. Shin-ya, K.; Wierzba, K.; Matsuo, K. I.; Ohtani, T.; Yamada, Y.; Furihata, K.; Hayakawa, Y.; Seto, H., **J. Am. Chem. Soc.,** 2001, **123**, 1262–1263.

188. Kim, M.-Y.; Vankayalapati, H.; Shin-ya, K.; Wierzba, K.; Hurley, L. H., Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex., **J. Am. Chem. Soc.,** 2002, **124**, 2098–2099

189. Kim, M.-Y.; Gleason-Guzman, M.; Izbicka, E.; Nishioka, D.; Hurley, L. H., The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures., **Cancer Res.,** 2003, **63**, 3247–3256

190. Barbieri, C. M.; Srinivasan, A. R.; Rzuczek, S. G.; Rice, J. E.; LaVoie, E. J.; Pilch, D. S., Defining the mode, energetics and specificity with which a macrocyclic hexaoxazole binds to human telomeric G-quadruplex DNA., **Nucleic Acids Res.,** 2007, **35**, 3272–3286

191. Tera, M.; Sohtome, Y.; Ishizuka, H.; Doi, T.; Takagi, M.; Shin-ya, K.; Nagasawa, K., Design and synthesis of telomestatin derivatives and their inhibitory activity of telomerase., **Heterocycles,** 2006, **69**, 505–514

192. Riou, J. F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Megnin-Chanet, F.; Helene, C.; Mergny, J. L., Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands., **Proc. Natl. Acad. Sci. USA,** 2002, **99**, 2672–2677

193. Naasani, I.; Seimiya, H.; Yamori, T.; Tsuruo, T., FJ5002: a potent telomerase inhibitor identified by exploiting the disease-oriented screening program with COMPARE analysis., **Cancer Res.,** 1999, **59**, 4004–4011

194. Franceschin, M.; Rossetti, L.; D'Ambrosio, A.; Schirripa, S.; Bianco, A.; Ortaggi, G.; Savino, M.; Schultes, C.; Neidle, S., Natural and synthetic G-quadruplex interactive berberine derivatives., **Bioorg. Med. Chem. Lett.,** 2006, **16**, 1707–1711

195. Zhang, W. J.; Ou, T. M.; Lu, Y. J.; Huang, Y. Y.; Wu, W. B.; Huang, Z. S.; Zhou, J. L.; Wong, K. Y.; Gu, L. Q., 9-Substituted berberine derivatives as G-quadruplex stabilizing ligands in telomeric DNA., **Bioorg. Med. Chem.,** 2007, **15**, 5493–5501

196. Ma, Y.; Ou, T. M.; Tan, J. H.; Hou, J. Q.; Huang, S. L.; Gu, L. Q.; Huang, Z. S., Synthesis and evaluation of 9-O-substituted berberine derivatives containing aza-aromatic terminal group as highly selective telomeric G-quadruplex stabilizing ligands., **Bioorg. Med. Chem. Lett.,** 2009, **19**, 3414–3417

197. Neidle, S.; Parkinson, G. N., Quadruplex DNA crystal structures and drug design., **Biochimie.,** 2008, **90**, 1184-96

198. Ma, D. L.; Lai, T. S.; Chan, F. Y.; Chung, W. H.; Abagyan, R.; Leung, Y. C.; Wong, K. Y., Discovery of a drug-like G-quadruplex binding ligand by highthroughput docking., **ChemMedChem,** 2008, **3**, 881-884

199. Parkinson, G.N.; Lee, M.P.H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA., **Nature,** 2002, **417**, 876-880

200. Totrov, M.; Abagyan, R., Flexible protein-ligand docking by global energy optimization in internal coordinates., **Proteins,** (Suppl. 1), 1997, 215-220

201. Cosconati, S.; Marinelli, L.; Trotta, R.; Virno, A.; Mayol, L.; Novellino, E.; Olson, A.J.; Randazzo, A., Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders., **J. Am. Chem. Soc.,** 2009, **131**, 16336-16337

202. Caceres, C.; Wright, G.; Gouyette, C.; Parkinson, G.; Subirana, J.A., A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tlþ/Naþ ions., **Nucleic Acids Res.,** 2004, **32**, 1097-1102

203. Lee, H.; Chan, D.S. H.; Yang, F.; Lam, H. Y.; Yan, S. C.; Che, C. M.; Ma, D. L.; Leung, C. H., Identification of natural product fonsecin B as a stabilizing ligand of c-myc G-quadruplex DNA by high-throughput virtual screening., **Chem. Commun.,** 2010, **46**, 4680-4682

204. Ma, D.L.; Chan, D.S.H.; Lee, P.; Kwan, M.H.T.; Leung, C.H., Molecular modeling of drug-DNA interactions: Virtual screening to structure-based design., **Biochimie,** 2011, **93**, 1252-1266

205. Li, M.H.; Luo, Q.; Xue, X.G.; Li, Z.S, Molecular dynamics studies of the 3D structure and planar ligand binding of a quadruplex dimer., **J. Mol. Mod.,** 2011, **17**, 515-526

206. Aixiao, L.; François, M.; Florent, B.; Michel, D.; Baoshan, W.; Xiang, Z.; Ping, W., Molecular modeling study of binding site selectivity of TQMP to G-quadruplex DNA., **Eur. J. Med. Chem.,** 2010, **45**, 983-991

207. Campbell, N.H.; Parkinson, G.N.; Reszka, A.P.; Neidle, S., Structural basis of DNA quadruplex recognition by an acridine drug., **J. Am. Chem. Soc.,** 2008,**130**, 6722-6724

208. Campbell, N.H.; Patel, M.; Tofa, A.B.; Ghosh, R.; Parkinson, G.N.; Neidle. S., Selectivity in ligand recognition of G-quadruplex loops., **Biochemistry**, 2009, 48, 1675-80

209. Chen, S.B.; Tan, J.H.; Ou, T.M.; Huang, S.L.; An, L.K.; Luo, H.B.; Li, D.; Gu, L.Q.; Huang, Z.S., Pharmacophore-based discovery of triaryl-substituted imidazole as new telomeric G-quadruplex ligand., **Bioorg. Med. Chem. Lett.,** 2011, **21**, 1004-1009

210. Alcaro, S.; Artese, A.; Iley, J.N.; Missailidis, S.; Ortuso, F.; Parrotta, L.; Pasceri, R.; Paduano, F.; Sissi, C.; Trapasso, R.; Vigorita, M.G., Rational design, synthesis, biophysical and antiproliferative evaluation of fluorenone derivatives with DNA G-quadruplex binding properties., **ChemMedChem,** 2010, **5**, 575-583

211. Merrill, R.M.; Sloan. A.; Anderson, A.E.; Ryker, K., Unstaged Cancer in the United States: A Population-Based Study., **BMC Cancer,** 2011, **11**, 402

212. Devillers, J.; Lipinick, R., In Practical Applications of Quantitative-Structure ActiVity Reationships (QSAR) in EnVironmental chemistry and Toxicology., 1st ed; Karcher, W., Devillers, J., Eds.; Kluwer: Dordrecht, 1990; Vol. 1, pp 129-143.

213. Haider, S.M.; Parkinson, G.N.; Neidle, S., Structure of a G-quadruplex-ligand complex., **J. Mol. Biol.,** 2003, **326**, 117–125

214. Stewart J., MOPAC: a semi impirical molecular orbital program., **J. Comput. Aid-Mol. Des.,** 1990, **4**, 1–105

215. Dewar, M.; Zoebisch, E.; Healy, E.; Stewart, J., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model., **J. Am. Chem. Soc.,** 1985, **107**, 3902–3909

216. CompuDrug, Latest Upgrades, (accessed February 10, 2010), http://www.compudrug.com.

217. Stahle L., Wold S., Multivariate data analysis and experimental design in biomedical research., **Prog. Med. Chem.,** 1988, **25**, 291–338

218. Cramer, R.D.; Bunce, J.D.; Patterson, D.E.; Frank, I.E., Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies., **Quant. Struct. Act. Relat.,** 1988, **7**, 18–25

219. Shao, J. Linear model selection by cross-validation, **J. Am. Stat. Assoc.,** 1993, **88**, 486–494

220. Aventis Pharms S.A. US Patent Application, US 20040053966A1

221. Prathipati, P.; Saxena, A. K., Comparison of MLR. PLS and GA-MLR in QSAR analysis., **SAR QSAR Environ. Res.,** 2003, **14**, 433–446

222. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4., **J. Comput.-Aided Mol. Des.,** 1996, **10**, 293-304

223. Bush, C. A.; Martin-Pastor, M.; Imberty, A., Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides., **Annu. Rev. Biophys. Biomol. Struct.,** 1999, **28**, 269-293.

****

Publications:

1. Zambre, V.P.; Murumkar, P.R.; Giridhar, R.; Yadav, M.R. Structural investigations of acridine derivatives by CoMFA and CoMSIA reveal novel insight into their structures toward DNA G-quadruplex mediated telomerase inhibition and offer a highly predictive 3D-model for substituted acridines., **J. Chem. Inf. Model.,** 2009, **49**, 1298-311.
2. Zambre, V.P.; Murumkar, P.R.; Giridhar, R.; Yadav, M.R. Development of highly predictive 3D-QSAR CoMSIA models for anthraquinone and acridone derivatives as telomerase inhibitors which target G-quadruplex DNA telomere., **J. Mol. Graph. Model.,** 2010, **29**, 229–239.
3. Zambre, V.P.; Giridhar, R.; Yadav, M.R. "Molecular modeling studies of some triazine derivatives as G-quadruplex binders for inhibition of telomerase enzyme, **Eur. J. Med. Chem.** (Communicated)
4. Zambre, V.P.; Giridhar, R.; Yadav, M.R.; Development of a universal 3D-QSAR CoMSIA model for predicting G-quadruplex stabilizing telomerase inhibitory activity of anti telomerase compounds., **Chem. Biol. Drug. Des.** (Communicated)
5. Zambre, V.P.; Giridhar, R.; Yadav, M.R., G-Quadruplex stabilizing telomerase inhibitors-A decade of development., **Eur. J. Med. Chem.** (Communicated)

Presentations (International):

1. Zambre V.P., Murumkar P.R., Giridhar, R., Yadav M.R., DNA G−quadruplex ligands: anthraquinones and acridones: 3D − QSAR studies for a quantitative prediction of the telomerase inhibitory activity based on comparative molecular similarity indices analysis. In 239th **American Chemical Societys' National Meeting and Exposition**, Moscone Convention Centre, San Francisco, CA, USA 21-25 March 2010

Presentations (National)

1. Zambre V.P., Murumkar P.R., Giridhar, R., Yadav, M.R., Comparative Molecular Similarity Indices Analysis of DNA G-Quadruplex Mediating Telomerase Inhibitors. In **61th Indian Pharmaceutical Congress**, Ahmedabad, to be held in 11-13, December, 2009.

Book Chapter

1. Yadav M.R., Murumkar P.R., Zambre V.P. "Advances in studies on Collagenase Inhibitors" In Advances in Studies on Matrix Metalloproteinase Inhibitors, Series: **Experientia Supplementum, Springer Basel AG, Switzerland.** (In Press)