Targeting the 'Oligomerization Region' of the Epidermal Growth Factor Receptor

using Small Molecule Inhibitors as Novel Anticancer Agents

By

Zechariah Lanka Marting

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Pharmacy

University of Sydney

December 2015

Dedication

To my well-cherished family who collectively sacrificed to make this program a success

There is humor in everything – it makes the difference

Declaration

I declare that this submission is my work supervised by Professor David Edward Hibbs (Faculty of Pharmacy), Professor W Groundwater (Faculty of Pharmacy), and Assoc. / Professor Thomas Grewal (Faculty of Pharmacy). Unless otherwise stated, and that to the best of my knowledge and belief, it contains no materials submitted for any other degree.

December 2015

Acknowledgments

I am grateful to Professor David E Hibbs who has made an indelible mark of 'ever forward and never backward' in me by painstakingly nurturing me academically from somewhere to where I am today. I consider him a cross-cultural academician who has shown that 'true knowledge' breaks down human barriers to get the nut. In addition, I sincerely acknowledge Professor Paul Groundwater and Assoc. / Professor Thomas Grewal who are academic icons I was privileged to cross paths with as other supervisors. I acknowledge Monira Hogue (a PhD student at the time) and Linda Varadi (PhD) for their significant assistance to me in this Program while it lasted. Factually, I enjoyed the friendship of the 2015 set under shared supervisors. I am indebted to the Technical and Administrative staffs in the Faculty of Pharmacy of the University who made the going smoother, especially when their attention was needed. I equally thank my Employers, the University of Jos, Plateau State, Nigeria, who sponsored the Program. Indeed, the list is not exhaustive but a mention of Dr. Nia Wheate is necessary who helped me in my final desperate hours. I owe my family so much, particularly my wife (Hannatu Garos), who made definite sacrifices to make this academic milestone a reality. Finally, I give my Creator the uttermost praise for being there for me in my very low and high days while the program lasted.

Abstract

The v6.3Glide Program (Schrödinger Suite, Schrödinger Inc., New York NY, USA) was used to screen $>9 \times 10^6$ compounds from several ligand sources including ZINC Chemical and Ambinter Chemical Databases. The selection of the four compounds DPT1 **1** – DPT4 **4** was based on their potential inhibition in both the ectodomain (oligomerization region) and endodomain (cytoplasmic tyrosine kinase domain) regions of EGFR when docked into 1IVO |246 – 253|**S** and 2J6M grids and not on Gscores. Their MTS bioevaluation (in 8×10^3 cells/mL) revealed the following: DPT1 **1**, IG_{50} 3.980 µM, IC_{50} 8.900 nM; DPT2 **2**, IG_{50} 0.250 µM, IC_{50} 0.400 nM; DPT3 **3**, IG_{50} 1.590 µM, IC_{50} 1.500 nM; and DPT4 **4**, IG_{50} 1.590 µM, IC_{50} 1.100 nM. Moreover, their western blotting indicated the following: DPT1 **1**, 1.00, 10.00 & 100.00 µM showed strong $pY-EGFR$ signals; DPT2 **2** 1.00 µM showed strong $pY-EGFR$ signal while a weak $pY-ERK$ 1 & 2 signal; DPT3 **3** 10.00 µM showed strong $pY-EGFR$ signal while its 100.00 µM showed strong $pY-ERK$ 1 & 2 signal and DPT4 **4** 1.00 µM showed strong $pY-EGFR$ signal but its 100.00 µM showed strong pY - EGFR signal. DPT5 **5** & DPT6 **6** (RTKI-like compounds) and zm-2 **8** were MTS assayed (in 5×10^3 cells/mL) and found DPT5 **5**, IG_{75} 28.18 µM, IC_{50} 0.389 µM; DPT6 **6**, IG_{75} 40.46 µM, IC_{50} 30.460 µM; zm2 **8**, threshold conc. 0,130 µM, IC_{50} 0.405 µM.

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
Akt (PKB)	Protein Kinase B
ATP	Adenosine Triphosphate
BSA	Bovine Serum Albumin
BTC	Betacellulin
CADD	Computer Aided Drug Design
CPK	Corey/Pauling/Koltun atom color/space/filling models
DCM	Dichloromethane
DMSO	Dimethyl Sulfoxide
EGF - EGFR	Epidermal Growth Factor - Receptor
ER	Epiregulin
ErbB1 (HER1)	Human Epidermal Growth Factor Receptor 1
ErbB2 (HER2)	Human Epidermal Growth Factor Receptor 2

ErbB3 (HER3)	Human Epidermal Growth Factor Receptor 3
ErbB4 (HER4)	Human Epidermal Growth Factor Receptor 4
ERK	Extracellular regulated kinase
FAK (PTK2)	Focal Adhesion Kinase or Protein Tyrosine Kinase 2
FGF	Fibroblast Growth Factor
FGFR	Fibroblast Growth Factor Receptor
FTIR	Fourier Transform Infrared
GUI	Graphical User Interface
HB-EGF	Heparin-Binding EGF-Like Growth Factor
HB-EGF	Heparin-Binding Epidermal Growth Factor
HTVS	High-Throughput Virtual Screening
LBDD	Ligand-Based Drug Design
mAbs	Monoclonal Antibodies
MAKP	Mitogen-Activated Protein Kinase
MAP	Mitogen-Activated Protein

MAPK	Mitogen-Activated Protein Kinase
MEK	Mitogen Extracellular Kinase
MS	Mass Spectroscopy
MSA	Molecular surface area
PSA	Polar Surface area
HBD	Hydrogen bond donor
HBA	Hydrogen bond acceptor
LBDD	Ligand-based drug design
SBDD	Structure-based drug design
MW	Molecular Weight
NRG	Neuregulin
NSCLC	Non-Small Cell Lung Cancer
PBS	Phosphate Buffered Saline
PDB	Protein Data Bank
PI3K	Phosphatidylinositol-3-Kinase

PKA	cAMP-dependent protein kinase
PKC	Protein Kinase C
PLCG	Phospholipase C, gamma
QSAR	Quantity Structure-Activity Relationship
Ras	Rat sarcoma' (RAS viral (v-ras) oncogene homolog)
RMSD	Root Mean Square Deviation
RPTK	Receptor Protein Tyrosine Kinase
SBDD	Structure-Based Drug Design
SCCH	Squamous Carcinoma Cell of the Head and Neck
SDS	Sodium Dodecyl Sulphate
SP	Standard Precision
RTKI	Receptor tyrosine Kinase Inhibitor
WHO	World Health Organization
XP	Extra Precision

Table of contents

List of Figures

List of Tables

Chapter one

Introduction

1.0 Cancer- the emperor of all maladies

Cancer remains a global health burden despite elaborate high-tech approaches it has received to be tamed [1]. For instance, the WHO recorded 7.6×10^6 deaths in 2008 [2, 3] and 8.2×10^6 in 2012 [4, 5] and predicts that by 2020 it would be 1×10^7 deaths [6] and a 20.3×10^6 in 2030 [7]. Cancer is a dynamic molecular group of complex heterogeneous diseases [8, 9] that subvert the entire cell communication pathway [10] of the infected cells. Some known cancer hallmarks as it progresses are abnormally high cell proliferation, migration, invasion, and metastasis [11]. Additional ones are evasion to growth suppressors, resistance to cell death (anti-apoptotic), immortality replication, formation of angiogenesis, exhibition of genetic diversification due to genomic instability and mutation, creation of tumor-promoting inflammation and ability to generate a suitable tumor microenvironment [12]. Future cancer management remains a serious challenge globally [13] due to its complex nature and poor prognosis [14, 15], hence our introduction of the disease as the emperor of all maladies [16]. Chemotherapy is essential in the overall cancer management plan despite the substantial challenges and setbacks such as resistance to targeted therapies [17] from mutational changes [18] and life-threatening side effects on short and/or long term use [19]. Search for anticancer drugs remains an intensive research area [20]. Advances in cancer molecular biology has assisted to identify new ways to inhibit its growth, leading

to new, more selective, and less toxic forms of its therapy [21]. This information foregrounds the basis for the innovation of novel small molecules capable of inhibiting well established pathways involving specific aberrant receptors like the human epidermal growth factor receptor (hERBB1–ERBB4 or ErbB1 – ErbB4) subfamily [22] of receptor tyrosine kinase (RTK) family and their cognate ligands [23]. The definitive role of this exemplary cell surface receptor, EGFR-RTK (ERBB1-RTK) in cancer has made it a target template for drug design of several types of inhibitors [24, 25].

1.1 The cell surface receptors (membrane or transmembrane receptors)

The cell surface receptors are specialized complex integral proteins that play a key role in mediating communication or signal transduction between the interior and exterior of the cell. They are generally entrenched in the cell's plasma membrane and are also referred to transmembrane or membrane receptors [26]. However, those found in the cellular cytoplasm are described as cytoplasmic receptors [27] and those in the cell's nucleus as nuclear receptors [28, 29]. Each or a group of receptors within a given site does have specific signaling molecules attached [30, 31]. Cell surface receptors are central to the regulation of both cellular and systemic physiology and act as conveyor mediators between the extracellular and intracellular phases of the cell by facilitating protein trafficking and regulating most intracellular processes [32]. The binding molecule to the cell surface receptor (R) such as a neurotransmitter, hormone, pharmaceutical drug or a toxin is the ligand (L). All cell surface receptors share some common features as proteins, such as high

specificity for cognate ligands to form reversible complexes (R⇋L) with equilibrium constants K_1 (association) and K_2 (dissociation), whose response amplitude is dependent on their concentrations as shown:

$$[R] + [L] \leftrightarrows [R: L]$$
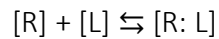
The formation of the [R: L] complex normally initiates major biochemical processes, for example, oligomerization, of the individual proteins via *trans* auto-phosphorylation of proteins as noticed in the protein tyrosine kinases [33].

1.2 EGF-induced human EGFR activation: basis for western blotting

The mature human EGFR (hEGFR) is a construct of 1186 amino acid (aa) residues, glycosylated, and has a modular structure comprised of a ligand-binding extracellular domain of 621 aa residues tethered to the plasma membrane by a single transmembrane region of 23 hydrophobic aa residues and is attached to a cytoplasmic domain constituted of 542 aa residues. The latter contains a 300 aa tyrosine kinase catalytic region flanked by a juxtamembrane domain and a C-terminal domain [34]. The auto-phosphorylation of the C-terminal domain tyrosine residues occurs when epidermal growth factor (EGF) binds the receptor [35]. Full members of the hErbB (hERBB) RTK (ErbB-RTK) subfamily are ErbB1 (EGFR, ERBB1, HER1), ErbB2 (ERBB2, HER2), ErbB3 (ERBB3, HER3) and ErbB4 (ERBB4, HER4) are structurally composed of the four domains [36] and are key regulators of cell-to-cell inductive processes and cell fate [37]. Using specific primary antibodies and appropriate secondary antibodies Western blotting (WB) may be used to predict the activities of the

generated hits [38]. The procession appears thus: when Ligand binds → oligomerization (dimerization) occurs → auto-phosphorylation of tyrosine kinase domain ensues → activation of downstream signaling pathways starts → cytoskeletal and gene transcriptional activities follow. Figure 1 illustrates this scheme. In carcinogenesis, the pathways are over actively deregulated [39, 40].
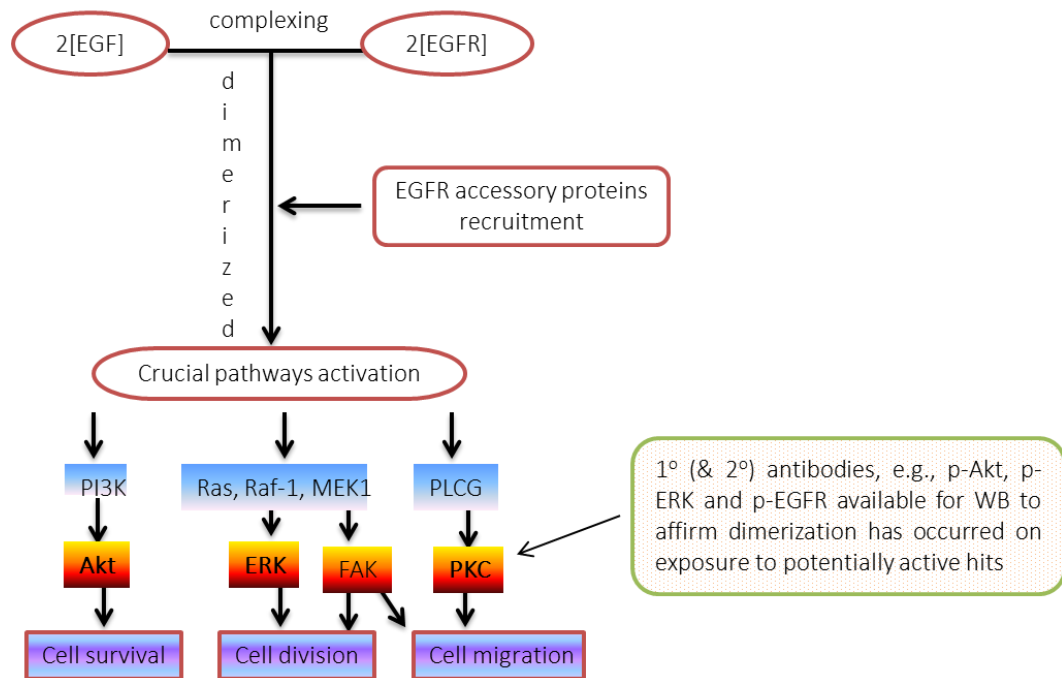
Figure 1. EGF-induced activation of EGFR.

Traditionally, two EGFs bind two EGFRs in a 2EGF:2EGFR fashion through a dimerization or an oligomerization process to become dimers (oligomers). The phosphorylation-dephosphorylation reversible process in the EGFR-RTK activation that leads to EGFR accessory proteins recruitment usually follows this pattern. Source: Revised from Schubbert et al., 2007.

1.3 The role of ErbB-RTK subfamily in carcinogenesis

Carcinogenesis defines the process in which normal cells turn into cancer cells following mutation [41] and epimutation [42] of the genetic material of normal cells, thereby upsetting the normal balance between proliferation and apoptosis [43]. For the imbalance to result to cancer, it may require a series of several mutations in one or more classes of the genes [44]. The carcinogenic role of the ErbB-RTK subfamily is well reported [45]. Each of the four domains in the family has a set number of aa residues that differ from each other [36], thus rendering the noticeable basic structural differences among them. This is evidenced in their unique ligand type and other features [36]. For instance, ErbB2 does not have a reported ligand of its own, yet, it is the most favored in heterodimerization participation with other subfamily members [46]. Furthermore, ErbB3 is devoid of the tyrosine kinase domain but has its unique or shared ligands [47]. The ErbB1 and ErbB4 are structurally the closest in the subfamily, yet, individually have distinct or shared cognate ligands [48] (Figure 12). The ErbB-RTK subfamily is regarded as central to most epithelial and mesenchymal cellular processes, thus clinically accepted as a target to anticancer drug development [49, 50]. The subfamily is implicated in a number of non-hematopoietic cancers such as glioma, prostate, CRC (colorectal cancer), NSCLC (non-small cell lung cancer), breast, etc. [51]. A few cognate ligands of the subfamily are known to be involved in a number of carcinomas [52, 53]. For examples, the transforming growth factor-alpha (TGFα) is a known active ligand in colon cancer [54] and epidermal growth factor (EGF) in

lung cancer [55]. The excessive ErbB subfamily receptor signaling brings about a wide variety

of carcinomas contained in Table 1 [52, 56].

Table 1. ErbB-RTK receptor subfamily members in carcinomas.

The reported carcinogenic involvement of the subfamily covers many types of human cells

in the body. The mechanisms may be a combination of an upregulation, expression, co-

expression, overexpression or a combination of two or more of these, due often to mutation.

Estimates are expressed in percentage and positive (p) or negative (n), meaning the member

plays a role or not.

Expression of ErbB subfamily in Carcinomas [57, 58]				
Carcinoma	erbB1^	erbB2ᵛ	erbB3ᵐ	erbB4ш
Breast	14-1%*	10-7%*	p	p
Ovary	30-5%*	20-2%*	p	p*
Renal	50-90%	24-40%	p	n
Lung (NSCLC)	40-0%*	3-56%*	p*	P
Head and Neck (SCCHN)	30-5%*	32-2%*	p*	P

Colorectal	25-75%	7%*	n	n
Pancreas	30-0%*	n	n	P
Glioma	40-0%*	n	n	n
Bladder	31-8%*	7-36%	n	n
Esophagus	P*	13-73%	n	n
Stomach	p*	5-55%*	n	n
Prostate	P	p*	p	P
Melanoma	P	p	p	P
Thyroid	P	n	p	P
Endometrial	p*	p*	n	n
Skin (squamous)	P	n	n	P
Lung (small)	P	n	n	P
Cervical	p*	n	n	n
Sarcomas	P	n	n	n

Chronic Myelogenous Leukemia	n	n	n	P

* Clinical studies have linked overexpression and/or mutation of the receptor to worse prognosis; ^Mutation and/or overexpression; ˅Overexpression and/or upregulation (gain-of-function); ᵐ Expression and/or co-expression with erbB2; ᵚ Expression and/or co-expression with erbB2; p positive ; n negative; SCCHN- Squamous carcinoma cell of the Head and neck; NSCLC- Non-small cell lung cancer. Source: http://www.cellsignal.com/reference/kinase_disease.html

1.4 Sampled classes of Anti-ErbB - RTK based on mechanisms

The ample presence of the subfamily members in human tissues and organs [52, 59] makes it a suitable drug target in search and development of novel anticancer [60]. There are two well-established mechanisms upon which some anti-EGFR agents have been developed, viz, the receptor tyrosine kinase inhibitors (RTKIs or TKIs) [61] and monoclonal antibodies (mAbs) [62]. Additionally, there are search for newer mechanisms ongoing [63].

1.4.1 Anti-heat-shock protein90s (Hsp90)

Hsp90 are abundant proteins known for maintenance of cellular homeostasis and are inducible under a variety of duress [64]. These proteins normally ensure other tyrosine proteins fold properly, hence are called molecular chaperons and accumulation occurs when there are pathological conditions like cancer [65]. Most RTKs such as EGFR require

chaperons to maintain their competent activation conformation as these interact and stabilize kinases as well as eliminate misfolded kinases [66]. This makes the Hsp90 antitumor target useful for kinase-related chaperone machinery. The anti-Hsp90 agents inhibit the function of the protein by binding to its (the protein) unusual ADP/ATP-binding pocket [67]. Investigational anti-Hsp90 drugs include compounds in Figure 2. Geldanamycin, for instance, affects ErbB2, EGF, v-Src and Raf-1. Other compounds include cisplatin, radicol, and novobiocin [68].
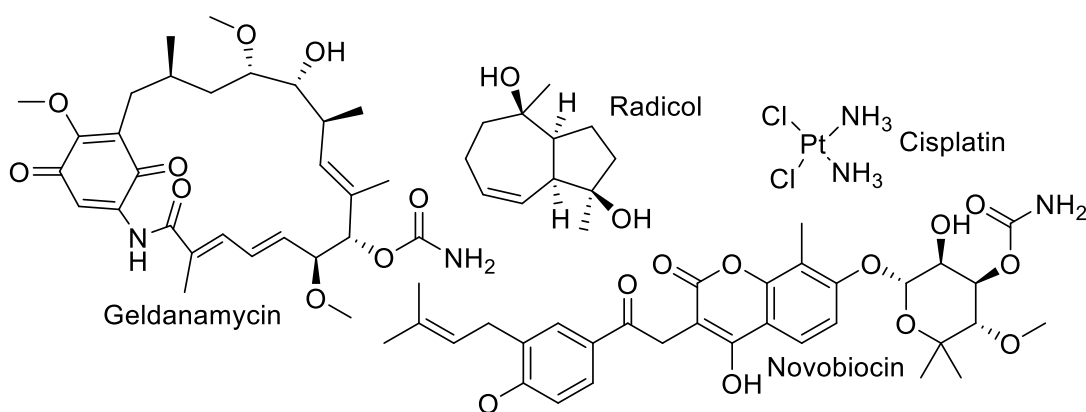
Figure 2. Samples of anti-Hsp90 anti-EGFR compounds.

1.4.2 Dimeric compounds as anti-EGFR drugs

Dimerization is an essential prerequisite in the ErbB − RTK subfamily signal transduction (Figure 9) and it is the lowest hierarchical possible and an exemplar of oligomerization mechanisms most studied as either a homodimer or heterodimer. The dimeric compounds are developed to interact with each monomer of the dimeric protein target as inhibitors for use as anticancer drugs [69]. It emphasizes the associations that can lead to a 'closed

structure' with its intrinsic symmetry unperturbed and therefore stable. Such dimers have mechanistically been successful in the development of orally active anticancer agents [70].

Some natural dimeric compounds with anti-proliferative activity are curcumin [71], artemisinin and courmemycin [72] shown in Figure 3.
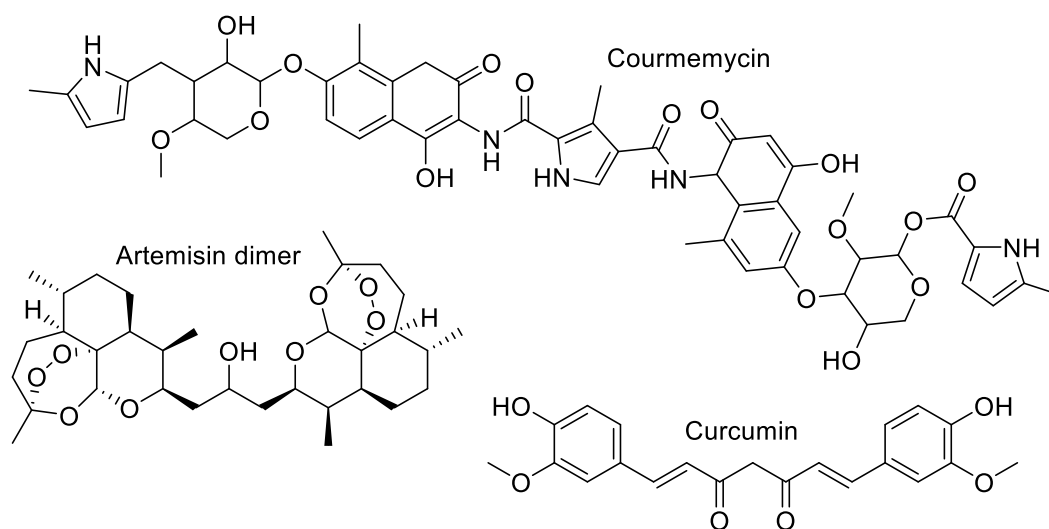
Figure 3. Examples of dimeric anti-cancer drugs from natural sources.

1.4.3 Antisense drugs as anti-EGFR dimerization agents

Antisense oligonucleotides have been investigated for therapy of various ailments such as cancer [73]. The group's activity is based on genetic sequence of a particular gene known to be involved in the cancer type and, thus designed to bind the messenger RNA (mRNA) [74] produced by that gene to inactivate it, thereby turning the gene 'off' [75]. The search for such drugs that decrease the expression of EGFR while regulating the cell proliferation for potential anticancer therapy defines the anti-sense phenomenon [76]. Typical antisense oligodeoxynucleotides include GEM 231 (Figure 4), a compound with a mixed-backbone of

an oligonucleotide structure that is said to have anti-tumor activity *in vitro* and *in vivo* through targeting PKA (protein kinase A) RIα-subunit [77]. It is used in the treatment of breast, lung, and prostate carcinomas. The expression of EGFR and type I cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKAI) is associated with neoplastic transformation that affects the receptor [78], accordingly the antisense target.

5'-GCG UGC CTC CTC AC U GGC-3'

Figure 4. GEM®231. PKA Riα antisense structure. The shows the bases that represent the 2'-*O*-methyl oligonucleosides as underlined. 18-mer mixed backbone (RNA-DNA) phosphorothioate oligonucleotide. Source: Mani et al.,

1.4.4 Inhibitors of angiogenesis as anti-EGFR

Tumor cells release specific pro-angiogenic factors such as angiogenin, vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF) and transforming growth factor-β (TGF-β), which stimulate endothelial cell proliferation, migration, and invasion [79]. Although the inhibition of the VEGF pathway has been the target against its vasculogenesis and angiogenesis, the key role of EGFR in angiogenesis is inhibited, thus, by blocking VEGFR EGFR is repressed. Figure 5 exemplified angiogenesis inhibitors with anti-EGFR activity. Cetuximab and panitumumab are known Anti-EGFR mAbs that block VEGFR activity as well [80].

Vandetanib, anti-EGFR and VEGFR

Curcumin

Imatinib, multikinase inhibitor

PKC 412, anti-EGFR, anti-VEGFR1, etc.

Gefitinib, anti-EGFR, angigenesis inhibitor
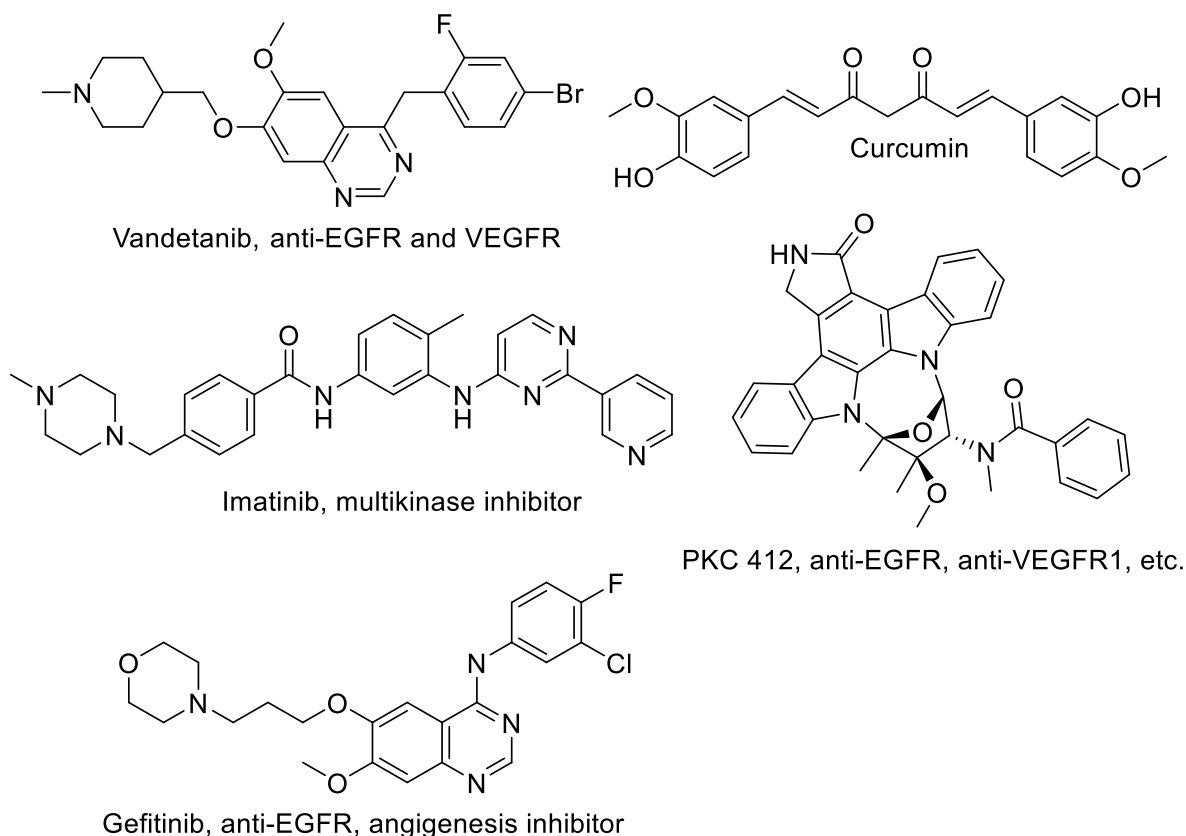
Figure 5. Samples of selected angiogenesis inhibitors (curcumin emphasized).

1.4.5 The small-molecule receptor tyrosine kinase inhibitors (RTKIs) and monoclonal antibodies (mAbs)

Small-molecule RTKIs and mAbs have been the most promising and widely used agents. Although all target EGFR, they have different mechanisms as the RTKIs act in the cytoplasmic tyrosine kinase domain while the mAbs act on the extracellular domain. The mAbs are often considered as first peptide anti-dimerization agents [81]. Some selected structures of RTKIs, lapatinib, canertinib, erlotinib, and gefitinib (section 1.4.4), [80] are shown in Figure 6. Among the mAbs in clinical use are cetuximab, rituximab, bevacizumab, trastuzumab, etc. [82].
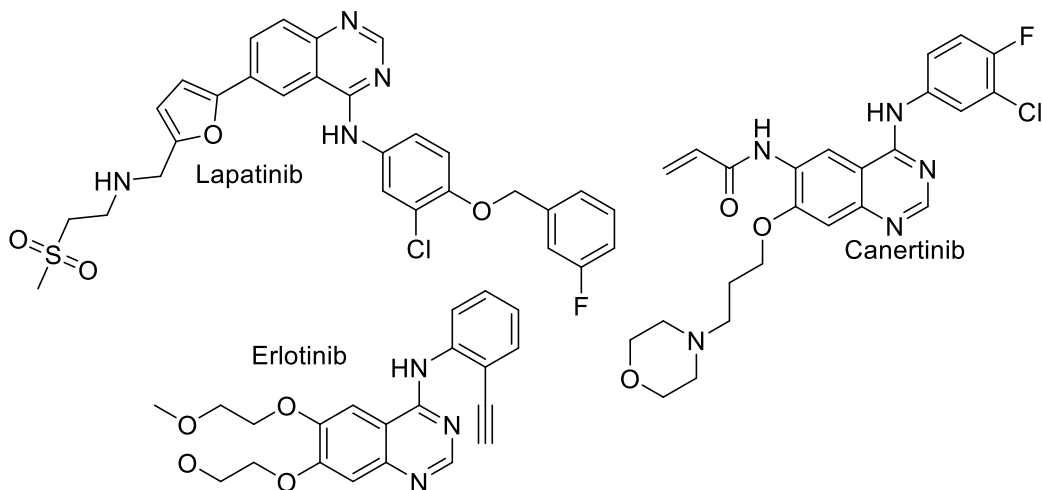
Figure 6. Structures of selected RTKIs in clinical application.

1.4.6 Antibody-based immune-conjugates or antibody-drug conjugates (ADCs)

This is a class of highly potent anticancer targeted therapeutics [83]. They are complex molecules made of an antibody (a whole mAb or its antibody fragment) linked through a stable chemical linker with labile bonds (cleavable bonds) to split at defined conditions such as certain enzymes or a specified pH [84], to a biologically active antitumor compound [85]. Examples of functional linkers in use the hydrazine bond linker, thioether bond linker, amide bond linker, disulfide bond linker, etc. [86]. These have been designed to improve the therapeutic window of chemotherapeutic agents or render the drug inactive (by acting as a prodrug) through altering its *in vivo* distribution due to conjugation with tumor-targeting monoclonal antibodies [83]. Trastuzumab-Emtansine (EQ75-ADR) is such an example that targets ErbB2 in breast cancer therapy [63]. Antibody-based immunoconjugates are largely investigational agents. Figure 7 shows a few examples.
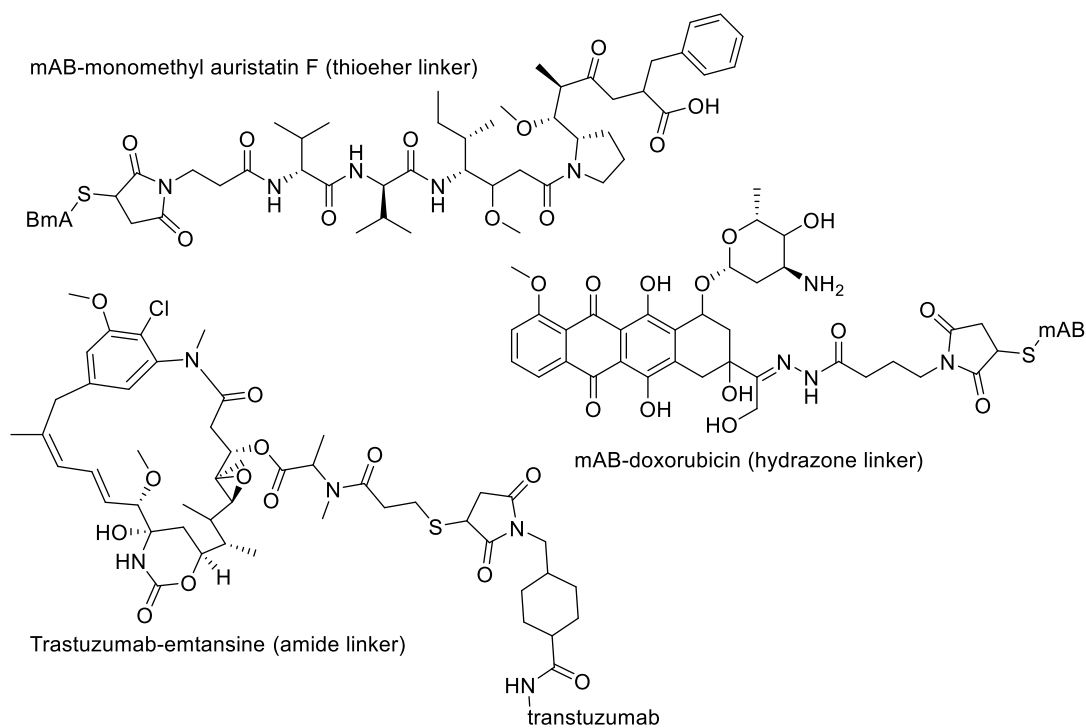
Figure 7. Samples of antibody-based immunoconjugates with their linkers.

1.4.7 Anti-EGFR anti-dimerization agents

Since members of the ErbB - RTK subfamily control vital signaling processes once activated by canonical ligands to cause oligomerization sequential to the catalytic activation of the intracellular kinase domains on recruitment of accessory proteins (Figure 1), it is logical to devise anti-dimerization agents that prevent dimerization from occurring even if the ligands are in place, on grounds of therapy. The use of small molecules as anti-dimerization (anti-oligomerization) agents remains yet to be fully exploited, thus only a few small molecules claim success against cancer [87] and another against psoriasis [88]. Figure 8 shows the structures of the two compounds tagged **P** and **Y.**

9-isobutyl-6-(((2-methylnaphthalen-1-yl)
methyl)thio)-9*H*-purin-2-amine (**P**)

(*E*)-7-benzyl-8-(2-(4-ethoxy-3-hydroxybenzylidene)hydrazinyl)-
1,3-dimethyl-3,7-dihydro-1*H*-purine-2,6-dione
(NSC56452) (**Y**)
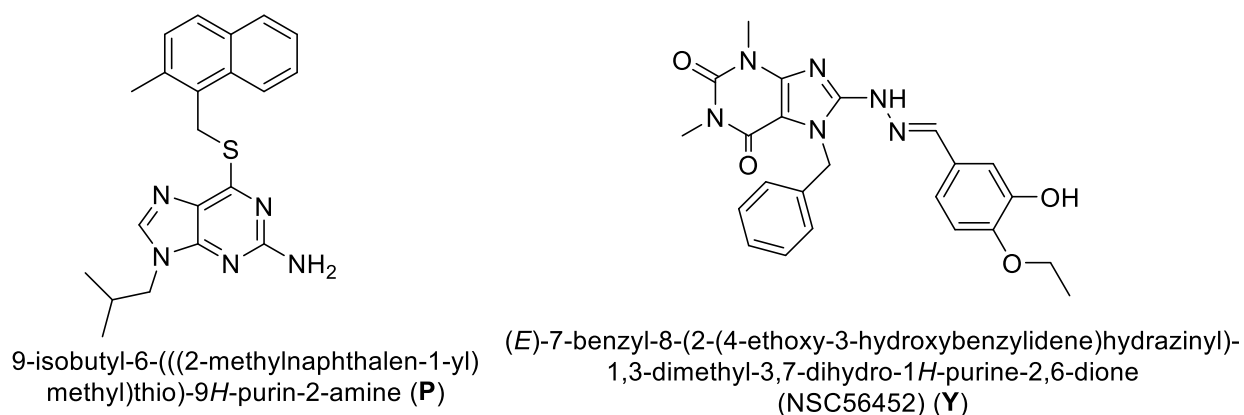
Figure 8. Two typical potentially active anti-ErbB anti-dimeric small molecules.

1.5 Dimerization: a prerequisite for protein posttranslational modifications

One of the posttranslational modifications (PTMs) in EGFR - RTKs is the reversible phosphorylation-dephosphoryaltion event in the cytoplasmic kinase domain. PTMs are the most ubiquitous natural protein-protein interactions in biological systems [89]. The key roles of the protein kinases in these cellular activities are exploited for drug targets [90]. Protein kinases transfer a phosphate, generally from ATP, which is reversed by the phosphatases under regulation, resulting in a fragile but sensitive equilibrium of phosphorylation-dephosphorylation process. Perturbation of this fragility can cause defects in key cellular functions. Some specific roles controlled by dimerization in protein-protein interactions generally (as shown in Figure 9) include facilitation of proximity and orientation in protein interactions [91], sets differential regulation for heterodimerization process [92], specifically regulates EGF-EGFR shed off [93], enhances specific protein-protein interaction [94], generates temporal and spatial boundaries [95] and regulates monomer to dimer transition [96]. These roles increase in complexity in oligomerization [97].
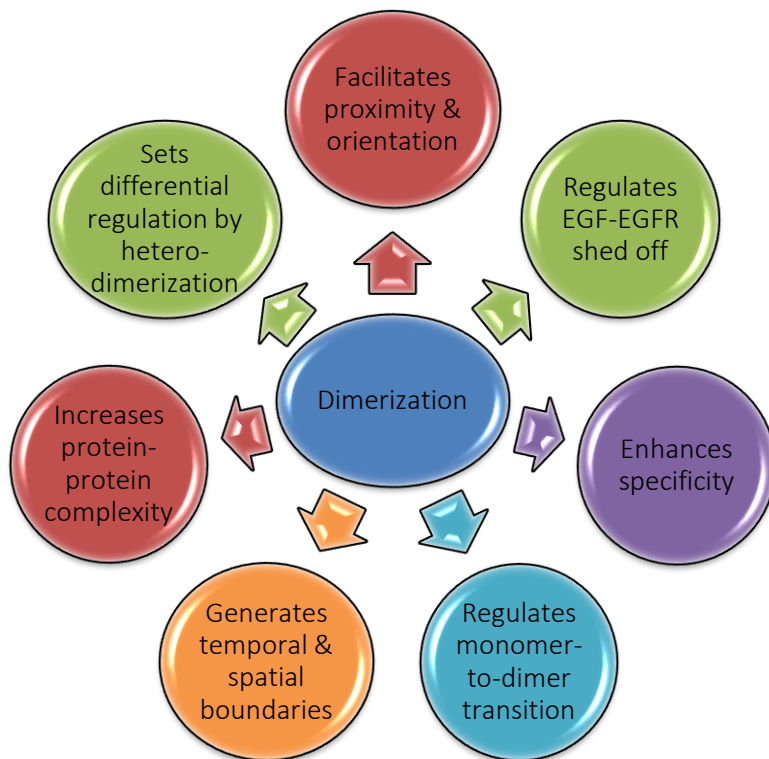
Figure 9. Some general regulatory roles of dimerization. Most biological systems show dimerization as exemplified by EGFR.

1.5.1 The mechanisms of dimerization in EGFR

The relevance of dimerization (oligomerization) is aforementioned. Several families of growth factors induce dimerization by their oligomeric disposition, e.g. platelet-derived growth factor (PDGF) while monomeric growth factors like EGF contain two binding sites for their receptors hence can cross-link two neighboring ErbB subfamily members [98]. The binding to the two receptors is simultaneous across a dimer interface, requiring additional stabilizing interactions between the monomer receptor molecules themselves. EGF, a protein consisting of 53 amino acid residues, has shown it can singly bind simultaneously two receptor molecules to appear like 2[EGFR] plus [EGF], described as a 2:1 fashion, to

effect dimerization of the extracellular domain of the EGFR [99, 100]. However, the more detailed fashion that has received global attention in a similar stoichiometric manner is 2[EGFR:EGF] or 2[EGFR] plus 2[EGF], that is, a 2:2 fashion [98, 101]. A number of models exists that attempt to explain the dynamic mechanisms of dimerization, most of which are based on some assumptions to arrive at a parametric conclusion [102]. One such is based on the 3D structures of the extracellular ligand binding domains of the EGFR and proposed as a mathematical model for the receptor-induced or ligand-induced dimerization and activation of ErbB3 [103], which is grouped as the receptor-mediated dimerization (RMD) [104] and ligand-mediated dimerization (LMD) [105, 106]. Ligand binding with the receptor is said to be a ligand-induced monomer-dimer transition [107] or ligand-induced conformational change with pre-formed dimers [102, 108]. In the case of the EGF-EGFR interaction, one method is the intramolecular mechanism, whence the EGF has induced a conformational change in the extracellular region that is transmitted via the transmembrane helix to activate the tyrosine kinase in a *cis* conformation, which mediates the autophosphorylation in the cytoplasmic region on multiple tyrosine residues [109]. The other is the intermolecular mechanism where information is transferred from the extracellular region to the cytoplasmic region because there is a lateral contact between the two receptors initiated by the ligand for dimerization to occur [110]. While the role of the ErbB4 may be unclear in oncogenesis those of EGFR, ERBB2 and ERBB3 are all implicated in the development and progression of cancer, and heterodimerization of the receptors plays a crucial part in their function [111]. Dimerization of the EGFR remains complex, therefore,

the need of models that offers closest explanation is required. Figure 10 is based on the

LMD and RMD [112, 113].
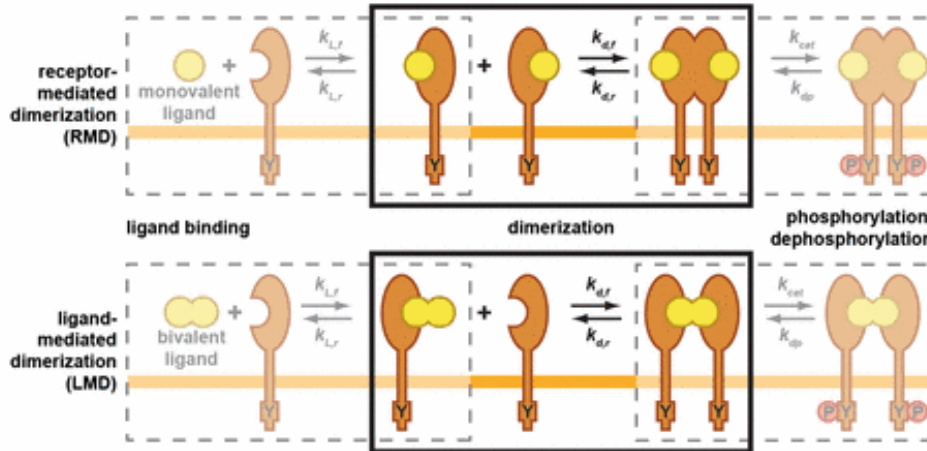
Figure 10. RMD and LMD model topologies. RMD occurs through direct extracellular

receptor–receptor interactions that are stabilized by the binding of monovalent ligand to

receptors. LMD occurs through indirect receptor–receptor interactions facilitated by an

interposed bivalent ligand, around which the receptors dimerize. In both cases, receptor

dimerization promotes phosphorylation (P) of a representative tyrosine (Y) present in the

cytoplasmic tail of the receptor. Where $K_{L,f}$ = Ligand association, $K_{L,r}$ = Ligand dissociation,

$K_{d,f}$ = Receptor coupling, $K_{d.,r}$ = Receptor uncoupling, Kcat = phosphorylation and K_{dp} =

Dephosphorylation. Revised source: Monast *et al*. (2014).

1.5.2 The key aa residues involved in the dimerization of EGFR

EGFR suits structure-based drug design (SBDD) for its current information [114, 115]. The

extracellular structure of the dimerized EGFR was elucidated in 2002 [101] using x-ray

crystallography and identified as PDB ID: 1IVO in the RCSB Protein Data Bank [46, 101]. The

four domains (I - IV) in the extracellular region of the dimer surface are mostly affiliated with domain II where a beta-hairpin arm of aa residues PRO242-GLY259 in this domain are said to be essential to maintaining the oligomeric structure [116]. This group of aa residues at the tip overlay ASN86, PHE230, PHE263, GLY264, ALA265, TYR275, THR278, CYS283, ARG285, and ALA286 set of aa residues in the cleft [116] through hydrophobic bonds (HYs) formation. The tip and cleft aa residues interact in such a predictive manner to form the HYs hexagonally sketched through the aa residues highlighted in Figure 11 (section 2.1). This β-hairpin loop structure is canonized as the 'dimerization) arm', and referred to as the 'dimerization or oligomerization region' for a dimer or an oligomer. However, the 'oligomerization region' is defined here as the region that embraces the binding aa residues in the backbone chain & side chain of the tip and the cleft of the oligomer. This is a feature that exists in all the subfamily members to enable formation of homodimers or homo-oligomers [92] & heterodimers or hetero-oligomers [117], particularly, as tetramers [118]. The key functions of the 'oligomerization arm' are to hold together the dimerizing (oligomerizing) receptors and thus, provide a sequential peptide framework for subsequent biological functions.
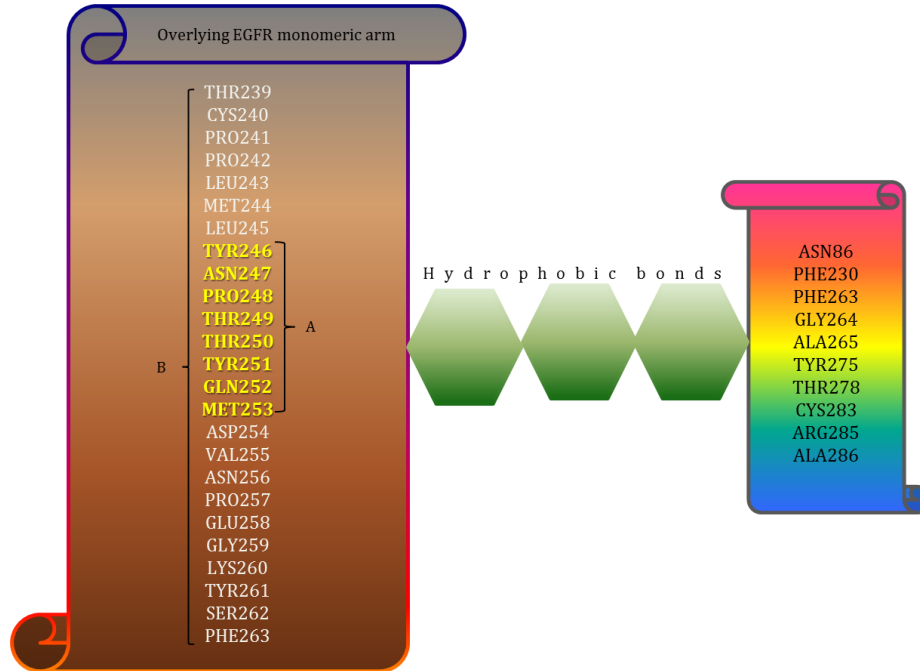
Figure 11. EGFR model of the 'oligomerization region' showing the aa residues.

The aa residues at the tip 'monomeric arm' (the left scroll) overlay or are linked to those below in the cleft (right scroll) by the HYs (represented in hexagonal shapes). The eight key aa residues at the tip of the 'dimerization arm' involved in binding those in the cleft of another EGFR during dimerization (oligomerization) are shown in yellow colour [A] while [B] describes the side chain aa residues in the 'monomeric arm region'. It is important to recognize that each monomer has its own [A] and [B]. Modified source: Ferguson et al., 2008.

The active participation of some of the ligands in realizing dimerization and carcinogenesis is already mentioned (section 1.3). Each cognate ligand plays a unique role in the dimerization of their canonical receptors. Thus, ligands are classified as mono -, bi - and tri – specific, exemplified by EGF, BTC, NRG2β, respectively, for HER1, HER1 & HER4, HER1,

HER3 & HER4, in that order [116], since members in the subfamily share one or more ligands [119, 120] as shown in Figure 12 that is responsible for inducing the formation of their homodimers, heterodimers or their higher oligomers [121]. However, this excludes the ErbB2 (HER2) for lack of a definite ligand despite its dominant participation in formation of hetero-oligomers with others in the group [122].
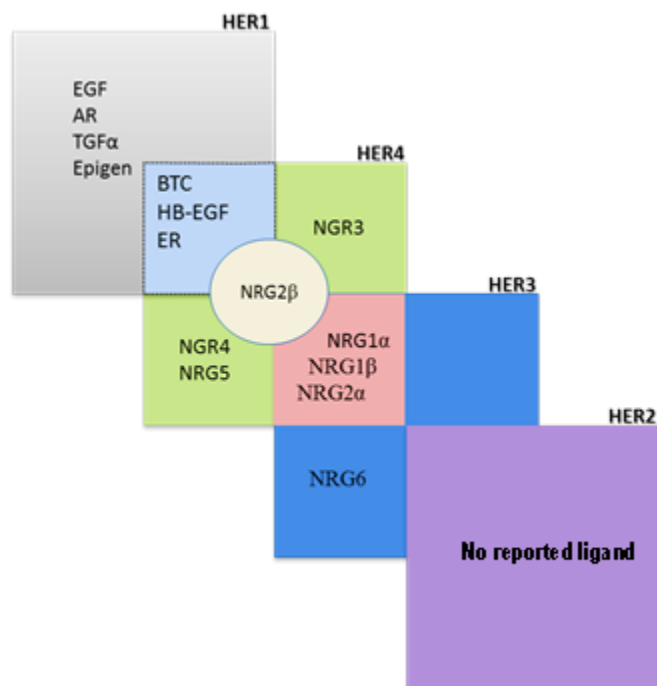
Figure 12. Model of the canonical ligands of the ErbB- RTK subfamily. The canonical ligands are unique to or shared among some of the receptors HER1, HER2, HER3 and HER4. The blue color square with BTC, HB-EGF and ER are ligands shared by HER1 & HER4, red color square containing NRG1α, NRG1β and NRG2α are ligands shared by HER3 & HER4 and cyclic tan color containing NRG2β is a ligand shared by HER1, HER3 & HER4. HER2 has no representative ligand. Oligomerization results from these interactions.

1.5.3 Dimerization is becoming insufficient to explain EGFR activation

EGFR is beheld as a critical super-node in cellular signaling pathway mechanisms pathways [123, 124] (sections 1.2 & 1.3). The oligomerization process includes forming higher numbers of monomers than just the traditional two [125, 126] . It is now clearer that the ErbB-RTK subfamily members activate each other through forming tetrameric molecules such as heterotetramers [127] and homotetramers [128], although EGFR trimers have been reported as well [121]. Schlessinger and coworkers in 1976, however, initiated a transforming perspective from the traditional dimerization mechanisms when they used fluorescence microscopic studies to achieve direct visualization of ErbB binding, aggregation and internalization of EGF on living fibroblast cells when suitable fluorescent derivatives of EGFR ligands and antagonists were developed [129]. This is the classic work that connected EGFR oligomerization, mobility, and internalization [130]. Further structural details were obtained when fluorescence microscopy and Förster Resonance Energy Transfer (FRET) were combined to elucidate the architectural and clustering of the subfamily RTKs [131]. For instance, an evidence that EGFR activation may not just occur within an EGFR dimer but depends also on interactions among EGFR dimers, a type of secondary dimerization, which is claimed to be modulated by multiple ligands and / or crosstalk with other receptors, which then propagates their activation [132, 133]. Thus, the principle of tetramerization is through transmodulation or transactivation of the heterodimers or homodimers [37]. Formation of higher oligomers (larger than dimers) is recognized among subfamily members, as shown for EGFR [106] with ErbB2 [134] and ErbB3 [135] in Figure 13 (section 1.5.3). However,

EGFR/ErbB2 and ErbB2/ErbB3 oligomers were proposed [136]. EGFR multimerization has been reported to be an important feature in regulation of its activities [137].The unresolved EGF-binding heterogeneity structurally displayed by the EGFR on the surface of cell lines shown on Scatchard plots depict negative binding cooperativity that is said to be linked primarily to high-affinity receptors involved in the mitogenic signaling process [132], which indicates a defense to oligomerization process. Queries such as 'are we missing a trick in cancer therapy' [138] or 'finding the missing links in EGFR' [139] can be answered on the basis that oligomerization, and not dimerization, mechanisms generate the EGFR-activated signals [140, 141].Furthermore, microscopic, image correlation spectroscopy and phosphorylation (western blots) using a model of mass-action chemical kinetics have been used to define the higher-order ligand-induced receptor-mediated EGFR oligomerization or clustering and its influence on signaling outputs, which proved to be distinct from EGFR dimers [142]. The ongoing investigations on tetrameric structures of the ErbB subfamily members [136] (depicted in Figure 13) using modified forms of FRET and microscopic imaging such as cited [34, 143, 144] makes the object of our study relevant.
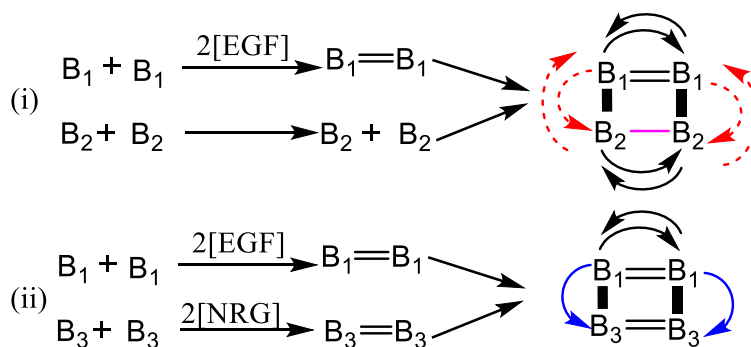
Figure 13. ErbB-RTK activates its members by hetero-tetramerization. B_1: EGFR, B_2: ErbB2, B_3: ErbB3; solid arrows (black color) = auto-phosphorylation (aY) between homodimerized moieties; solid arrow (blue color) = unidirectional transphosphorylation of B_3 by B_1 since B_3 is devoid of tyrosine kinase domain; broken arrows (red color) = potential transphosphorylation (tY) between two EGFRs, two ErbB2s, i.e., between heterodimers where tyrosine kinase domain exists. (i) EGF-induced EGFR homodimer $2[B_1]$ and two molecules of $[B_2]$ to form a heterotetramer complex. The two B_2 molecules dimerize through binding to an activated B_1 dimer. Arrows mark aY between B_1 / B_1 and B_2 / B_2 & tY between B_1 / B_2 and B_2 / B_1. Note that no ligand is indicated for B_2 and the purple color single bond shows it has been activated by homodimer $2[B_1]$. This is expected to be similar in B_2 & ErbB4 heterotetramer formation. (ii) EGF-induced EGFR homodimer $2[B_1]$ and a NRG-induced $2[B_3]$ homodimer to form the heterotetramer complex. Arrows mark aY between B_1 / B_1 only & tY of B_3 by B_1. This is expected to be similar when B_3 & ErbB4 form a heterotetramer. Source: modified from Schlessinger, 2000.

1.6 Aims/Objectives

EGFR activates its signal transduction through hetero-tetramer structures (Section 1.5.3). Trial and proven anti-EGFR approaches discussed in section 1.4 are summarized in Figure 14 to illustrate fundamental mechanisms that bring about the anti-EGFR. Other specific examples are combinations of mAbs [49] and RTKIs / mAbs [145], decoy receptors [146], DNA aptamers [147] and anti-ATP binding sites [148] agents. The circles overlap to indicate anti-EGFR activity, howbeit, from dissimilar mechanisms. Our focus area is arrowed.
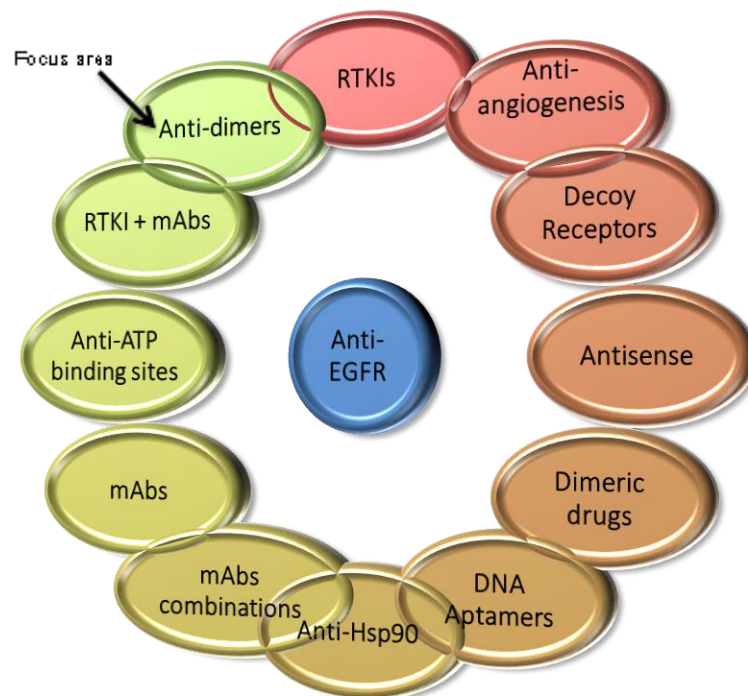
Figure 14. Aims/objectives defined based on Anti-EGFR mechanism.

These entire agents share in interfering with EGFR dimerization subsequently, however, through different mechanisms. For instance, the mAbs act at the extracellular domain but the RTKIs in the cytoplasmic domain, inhibiting the EGFR activity subsequently. The 'Focus area' points to our aims/objectives as we design, bioevaluation, and synthesis small molecules that target the 'oligomerization region' of EGFR to be novel anticancer agents on successful development.

The aims/objectives of this piece of work are:

a. To design novel small molecules that inhibits the oligomer formation in EGFR by targeting the monomeric arm in the 'oligomerization region' as defined.

b. To bioevaluate the hit (s) in an *in vitro* bioanalysis using colorimetric and western blotting in order to investigate the hit (s) influence on the i) EGF-induced pY-Akt activation, ii) EGF-induced pY-ERK1 and 2, and ii) EGF-induced pY-EGFR.

c. To optimize successful lead (s) as well as generate a library (or libraries) of potentially active compounds using the v3.9CombiGlide for further bioanalysis rounds.

d. To synthesize and / or design synthetic methods of individual and / or groups of compounds with clear potential anti-EGFR anti-oligomeric activities.

1.7 The approach to achieve the aims / objectives

Programs of the Schrödinger Release 2015-4 shall be used throughout the drug design procedures. Schrödinger is a leading provider of advanced molecular simulations and enterprise software solutions and services for life sciences and materials research [149]. Two EGFR 3D crystals of different domains - Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains (ectodomain), 1IVO (PDB ID 1IVO) [101] and Crystal structure of EGFR kinase domain in complex with AEE788, 2J6M (PDB ID 2J6M) (endodomain) [150] shall be imported from the RCSB Protein Data Bank and prepared. Grids of these crystals shall be generated and subsequently docked with 3D ligands sourced from Chemical Data Banks. The emerged 'hits' would be purchased and / or synthesized for bioevaluation. The lead candidate (s) will be optimized using the v3.9CombiGlide tool to generate a biased library (or libraries) of anti-EGFR anti-oligomeric compounds for bioassays. Synthesis shall apply to compounds of 'best' GlideScores (Gscores) not purchasable or

synthesizable. All synthesized compounds shall be bioevaluated and optimized again as lead candidates. The cycle of design, bioevaluation, and synthesis of lead candidates may continue, varying the physicochemical properties to obtain potentially more active anti-oligomeric compounds preparatory for drug development (Figures 28 & 45). Figure 15 shows the reductionist bow-tie model that attempts to simplify the complexity of the EGFR influence on biological systems. The input of a growth factor that functions through the EGFR receptor to achieve hetero- or homodimers activates common signaling cascades that are defined through a core collection process of biochemical interactions which are tightly coupled to each other and interfaced by the oligomerized EGFR, i.e., dimers such as 2[EGF:EGFR] or tetramers represented as |2[EGF:EGFR]:2[ErbB2.ErbB2]| [151] (Figure 13).

The EGFR dimer axis

Space-filled EGFR (1IVO), showing the two EGF ligands (red & green), the two EGFR dimers (orange & olive green). The two monomeric arms overlap each other (the orange over the olive green)

The outcomes include
- Cell survival
- Cell growth
- Cell migration
- Cell proliferation
- Cell ontogenesis

The β-hair loop of the EGFR, showing the template residues, PRO242-GLY259, on the backbone chain of the loop

The spheres represent the 10 underlying residues in the cleft that form hydrophobic bonds, shown in yellow dotted lines, with the 'tip' residues on the monomeric arm. Such bond is seen between THR249 and ASN86 in Domain II

Figure 15. The bow-tie model simplifies the 2[EGF:EGFR] network.

The model is modified to show the dimer axis or symmetry of the EGFR crystal (1IVO), the appearance of the overlapping aa residues shown in the space-filled crystal, its β-hair loop, and a possible show of the aa residues in the cleft and at the tip. The output of the core process following gene transcription is shown in cell proliferation, survival, etc. as inscribed. Source: Citri and Yarden, 2006; Ogiso et al., 2002; Ferguson et al., 2000 & 2008.

Crystal Structure of the Complex of Human Epidermal Growth Factor and Receptor Extracellular Domains (2 EGFs in orange/yellow, 2 EGFRs in green/dark red). PDBID- 1IVO (no backbone atoms displayed)

'Monomeric arm' of a monomer

EGF

EGF

An incised EGFR-EGF complex showing a monomeric arm

Domain II dimerization interface showing mainly the interacting amino acid residues

Y251
R285
H280
D279
Y251
R285

D 254 F 230 N 247 T 249 N 86
M 253 250
Q 252 G 264 F 248
246 263 Y 251
R 285
G 283 A 286 Y 275

Backbone amino acids in the 'dimeric arm' symmetry

Figure 16. Full and partial representations of the EGFR crystal (1IVO).

Top left is the 1IVO, in ribbon of a 2[EGF:EGFR] dimeric ectodomain structure without protein backbone chain; top right is the monomeric arm without the protein backbone; bottom left shows the interface of the two dimeric (oligomeric) arm region and their prominent interacting aa residues TYR251, ASP279,HIS280, and ARG285, and bottom right is the dimeric (oligomeric) arm showing most of the protein backbone chain residues and a few of the aa residues in the cleft are also noticeable, e.g., ASN86, PHE230, PHE263, TYR275, CYS283, ALA286, etc.

2.0 Introduction

Computer-aided drug design (CADD) is a rational technique [152] in search of novel biomolecular sites using known ligands or termed ligand-based drug design (LBDD) [153], or search for novel ligands using known biomolecular sites or termed structure-based drug design (SBDD) [154], although the duo can complement each other in drug discovery development [155, 156]. LBDD depends on the knowledge of the 3D ligand and SBDD relies on the information of the 3D target molecular structure. The design of small molecules that target the 'oligomerization region' of the EGFR is a SBDD style that envelops molecular docking, structure-based virtual screening, and molecular dynamics methods [157]. The use of molecular docking and scoring (*in silico* screening) and modeling methods has enhanced search for new therapeutic drugs [158] for its relative cheapness , fastness, and information-driven [159]. Some EGFR details are shown in Figure 16 and explained in section 2.1. The crystal structures of EGFR PDBID1IVO (Figure 17) and PDBID2J6M (Figure 18) were obtained from the RCSB PDB.
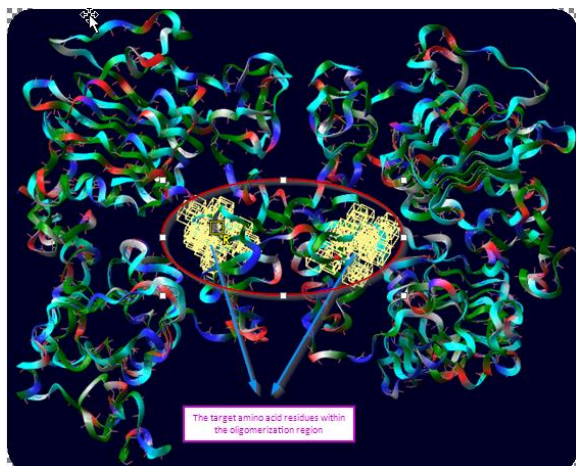
Figure 17. Ribbon structure of 1IVO crystal.

The oligomerization region is enclosed in the red ring and the target aa residues in yellow cubes.
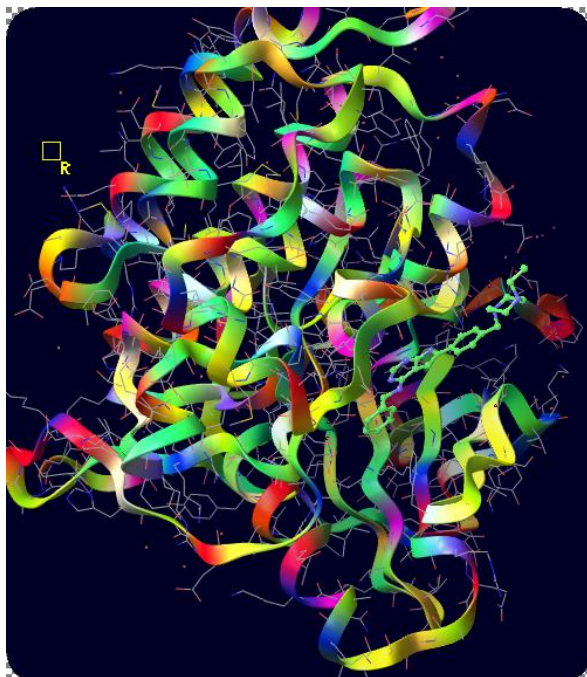
Figure 18. Ribbon structure of 2J6M crystal.

The cytoplasmic kinase site in complex with the RTKI, AEE788 in green ball & stick shown.

The 2J6M crystal prior to preparation showed it had several interactions involving water bridges, π-cation contacts between LYS754, H-bond (Hb) with MET793 carbonyl oxygen on the main side chain and another Hb from MET793 amino group to the the nitrogen on the pyrimidine moiety. The H-bonds (Hbs) and HYs are shown in Figures 19.
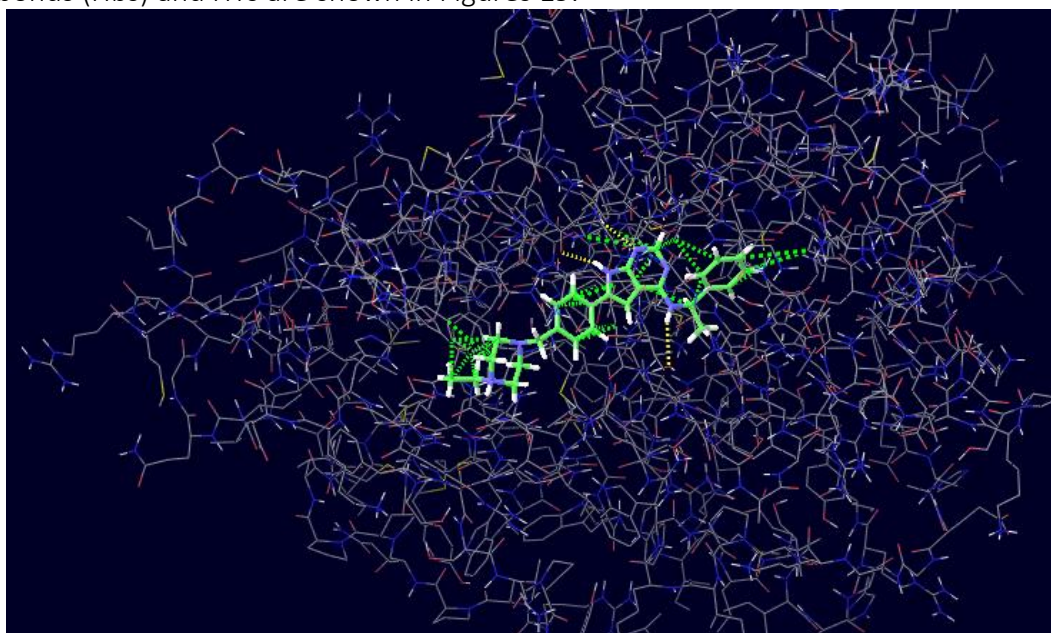
Figure 19. 2J6M crystal interfaced with Hbs and HYs in dotted yellow line.

2.1 Targeting the 'oligomerization region' of the EGFR-RTK

Section 1.5.2 has defined the oligomerization region. The protein backbone is absolutely necessary in anchoring and aligning the frontline aa residues to form the HYs and Hbs that effect oligomerization on binding with native ligands [160]. The ligand that binds induces a monomer–oligomer transition from an untethered monomeric receptor on exposure of a dimerization arm which partners with another induced receptor to cause their oligomerization, kinase activation and signaling [141]. The repositioned allosteric EGFR enhances the activity of the so-called eight key aa residues in the tip monomeric arm (**T** for 'at the tip or top') that associated with the second or underlying monomeric arm (**B** for 'below or in the receptor cleft') (Figure 11). This is described as:

a) TYR246, PRO248 and TYR251 [**T**] form hydrophobic bond (HY) with PHE230, PHE263, ALA265, TYR275, and ARG285 [**B**] in the cleft

b) PRO248 [**T**] forms selectively HY with PHE230 and ALA265 [**B**]

c) TYR246 side chain oxygen atom [**T**] forms Hb with CYS283 carbonyl oxygen and GLY264 nitrogen [**B**] in the cleft

d) THR249 [**T**] forms Hb with ASN86 in domain 1

e) TYR251 [**T**] oxygen on side chain forms Hb ARG285 nitrogen atom on side chain as well as HY with PHE263, GLY264 and TYR275 [**B**] in the cleft

f) GLN252 [T] amide main side chain forms Hb with ALA286 [B] carbonyl oxygen in the cleft and

g) MET253 [T] forms HY with THR278 [B] in the cleft [101, 161] (Figures 20 & 21).)

The interception prior to the dimerization is what we seek to achieve since interfering with any or all of these hydrophobic - bond & H − bonding - forming aa residues (Figure 22) imperils native processes. Our grids were generated based on this knowledge.

2.1.1 Descriptions of the 1IVO for grid design before and after.

Figure 20 shows the [T] and [B] aa residues on a monomeric arm with the line of symmetry between them. The purpose would be to examine the ligand impact on the [T] aa residues, the [B] aa residues, combinations of [T] and [B] aa residues, and some of the side chain and backbone aa residues (some observations not reported) associated with the anti-dimerization process [87]. Figure 21 shows that all the aa residues on [T] and [B] are located on each monomeric arm but must require another of its kind to dimerize. Figure 22 shows the hydrophobic / hydrophilic surface map on the [T] and in [B] of the receptor.

It is necessary to note that [T] from Chain **A** or **B** overlaps [B] only on its Chain **B** or **A** accordingly [99, 101, 161].
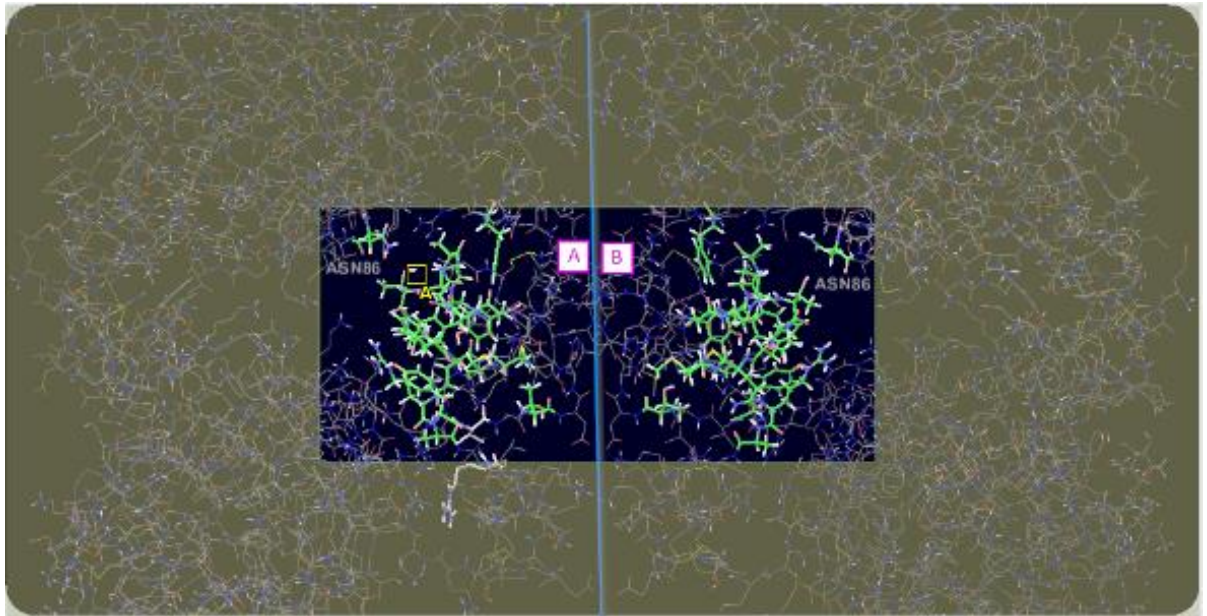
Figure 20. The 8 [T] and 10 [B] key aa residues in 1IVO.

[T] & [B] responsible for the oligomerization highlighted in green are shown cropped for |A| & |B|. |A| is symmetrical to |B| as delineated and parted by the symmetric line. ASN86 (shown) is in domain 1, thus way out of the apparently clustered residues.
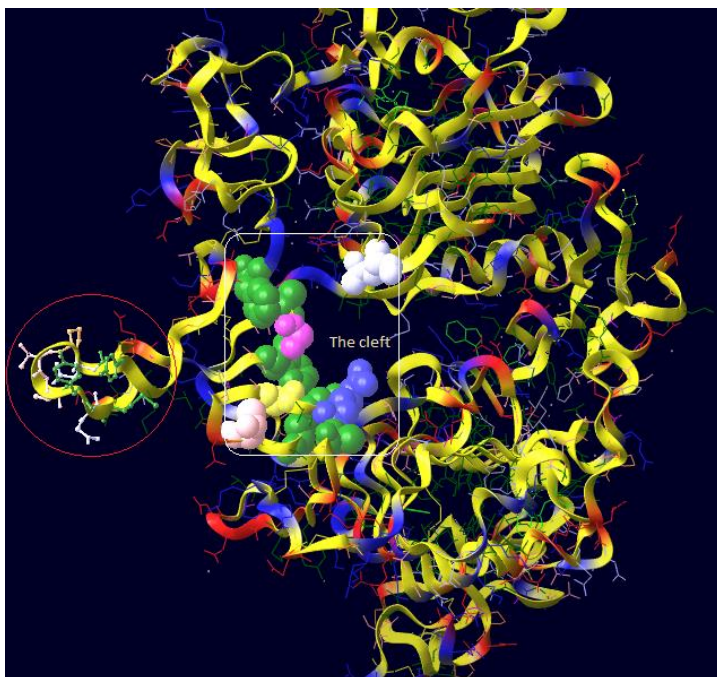
Figure 21. The oligomerization residues on a monomeric arm of 1IVO. The monomeric arm (in red ring) [T] (aa residues in ball & stick) overlays [B] (aa residues in white rectangle) of 2nd monomeric arm during oligomerization process. Note the cleft & ASN86 (in white).
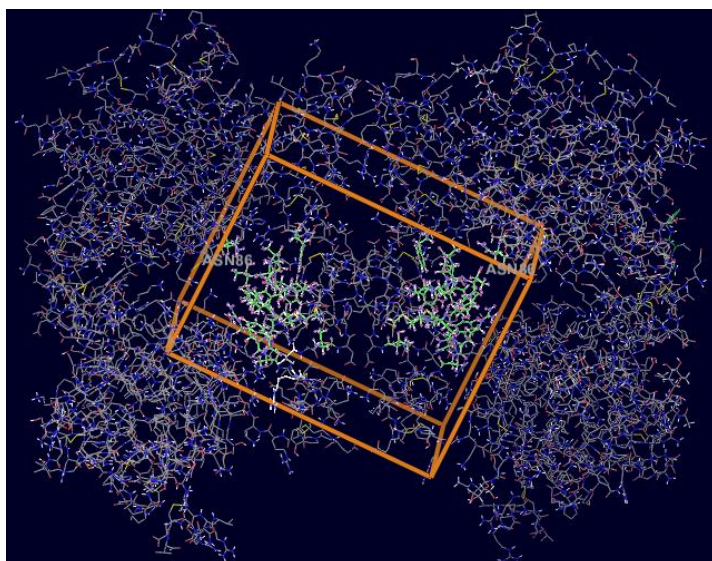
Figure 22. Hydrophobic / hydrophilic surface map of 1IVO. The symmetric structure of 1IVO showing the hydrophobic / hydrophilic surface map within 5Å.

2.2 Modifications on the imported 1IVO and 2J6M crystals from the RCSB PDB

The 2[EGFR:EGF] dimer crystal structure, 1IVO (extracellular domain), and AEE788:EGFR (cytoplasmic tyrosine kinase domain), 2J6M, were individually imported from the RCSB PDB (Research Collaboratory Structural Bioinformatics Protein Data Bank) and prepared with the Protein Preparation Wizard (ppw) with v3.4Epik, v4.2Prime and v6.9Impact (Schrödinger, LLC, New York, NY, 2015-4). The workspace structure was preprocessed to achieve the removal of water molecules of crystallization and other non-essential waters, the addition of hydrogen atoms, creation of disulfide bonds and filling of missing side chains and loops. Following the analysis of the workspace, hets such as N-acetyl-*D*-glucosamine (NAG) ligands were removed within 5Å by default during proton and metal charge generation. Optimization of the proteins for Hb, ASN, GLN and HIS states were achieved automatically using the Epik during protein preparation, which assigned the protein orientations for the minimization to follow for unrestrained frozen heavy atoms. The program was run at pH 7.0 ± 3, convergence RMSD of 0.30Å and Force Field OPLS_2005, all at default values [162].

2.2.1 2j6M crystal after protein preparation

The RTKI (AEE788) is now removed for a new ligand occupation (Figure 23).
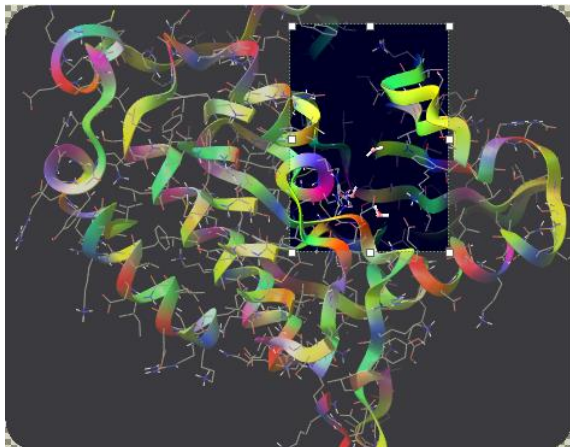
Figure 23. 2J6M grid with ligand AEE removed. The cropped area would be occupied by a new ligand (cf. Figure 16).

2.2.2 1IVO after protein preparation
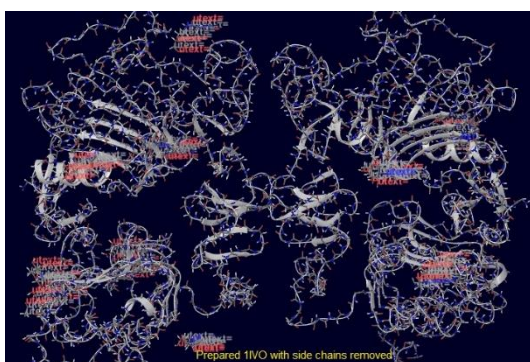
Figure 24. A prepared dimeric 1IVO crystal for grid generation. A dimeric structure of 1IVO and its dimers in place. The side chain aa residues are removed for clarity, set for grid generation.
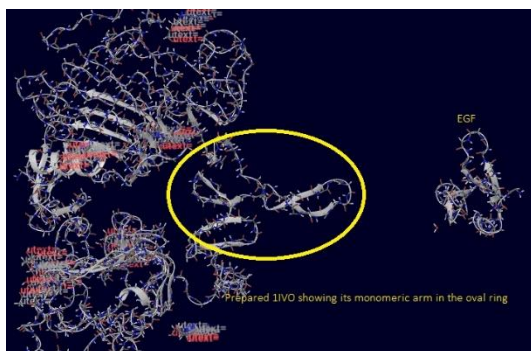
Figure 25. A prepared monomeric 1IVO crystal for grid generation.

A monomeric structure of 1IVO & its monomeric arm in the yellow ring. The side chain aa residues are removed for clarity.

2.3 Grid generation from the prepared 1IVO and 2J6M crystals

2.3.1 2J6M Grid

The grid of the 2J6M was generated using v6.9Glide (Schrödinger, LLC, New York, NY, 2015-4). The ligand AEE788 was picked to define the grid box (Figure 26).
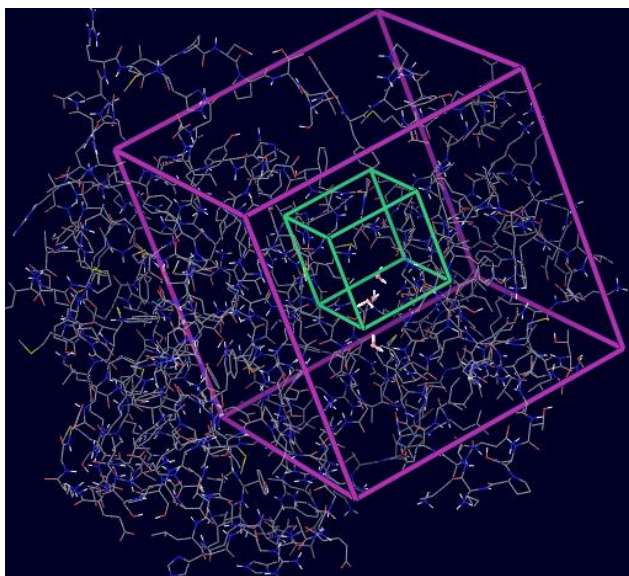
Figure 26. 2J6M grid at 10Å view. Grid generation removed AEE788. The green cube shows the molecular center of the incoming ligand while the purple envelopes the entire structure of the ligand.

2.3.2 The generated grids from 1IVO

2J6M helped to furnish us with helpful information about our focused receptor, 1IVO. Unlike 2J6M, several grids were generated from the 1IVO crystal for docking, as either probes and / or attempts to improve on the Gscores. The grids generated for docking were not only from the essential aa residues directly involved in the HY during oligomerization of EGFR but the grids extended to a combination of side chain and backbone aa residues. Among the grids generated were a) |246-253|S or **W**, comprising amino acid residues

TYR246, ASN247, PRO248, THR249, THR250, TYR251, GLN252 & MET253 [88], b) |TYR246 TYR251|S, c) |TYR 246 MET253|S, d) |ASN86 PHE230 PHE263 GLY264 ALA265 TYR275 THR278 CYS283 ARG285 ALA286|S or W and e) |TYR246 PHE230 TYR275 ALA286|W and f) |200-300|W, i.e., (ARG200, GLY201, LYS202, SER203, PRO204, ASP206, CSY207, CYS208, HIS209, ASN210...ASP290, SER291, TYR292, GLU293, MET294, GLU295, GLU296, ASP297, GLY298, VAL299 & ARG300) (Figure 11). Grids from aa residues of the 'monomeric arm' are tagged S (for single or one arm), dimeric region are described as W (for whole or complete dimer). Grids of W emphasized the symmetry of the dimer [163] (Figures 20 & 22). Grid 1IVO |246-253|S remained the most biased throughout the experiment unless stated otherwise. The numbers within the two vertical lines reflect the aa residues used to define the centroid of the grid which defines that of the docked ligand.

2.3.3 Similes of some generated 1IVO grids

The grids generated aimed at obtaining 'best' Gscores as well as for investigation. Figures 27A-27E show some of these grids ably used throughout the search for anti-oligomeric novel compounds against cancer. However, docked grids with very poor Gscores are not reflected in this report.
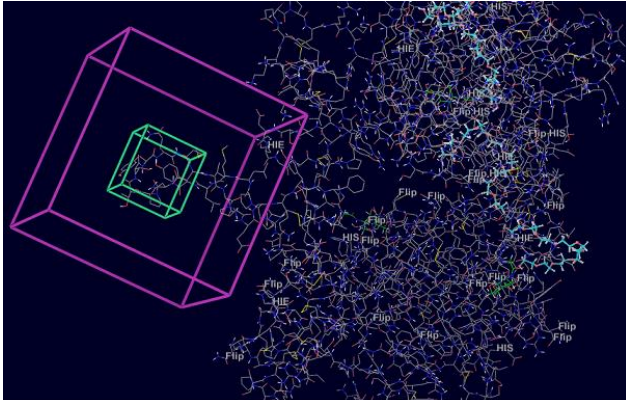
Figure 27A. 1IVO grid |246-253|S.
Grid of the single arm of the 1IVO
crystal. The green cube defines the
centroid of any docked ligand while the
purple box (10Å) its entire structure.
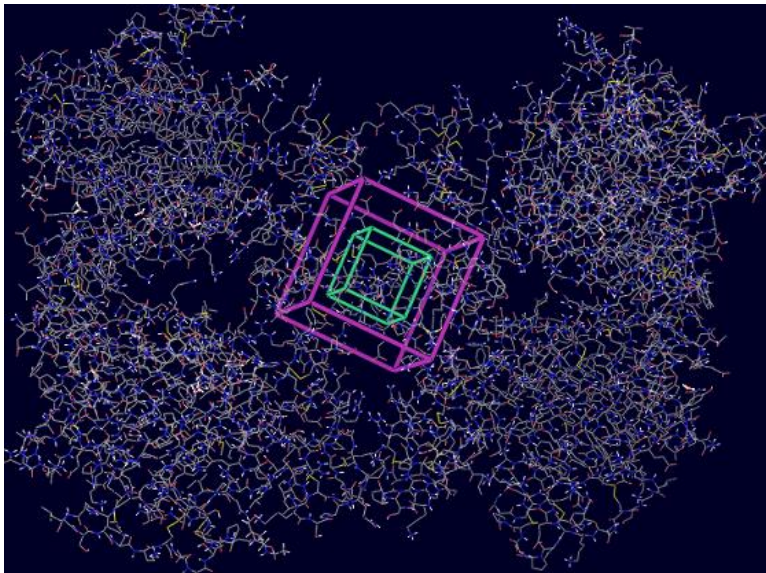
Figure 27B. 1IVO grid |246-253|W.
A dimeric grid of 1IVO crystal.
The green cube describes the
molecular center of any docked
ligand while the purple box (10Å)
encloses its entire structure.
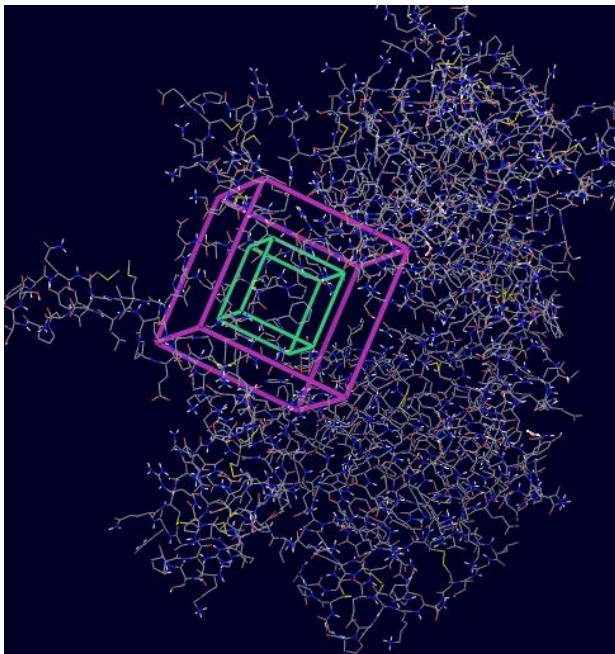
Figure 27C. 1IVO grid |86 230 263 264 265 275 278 283 285 286|S. 1IVO monomeric arm grid of the key aa residues in the cleft. The green cube articulates the molecular center of any docked ligand while the purple box (10Å) encloses its entire structure.
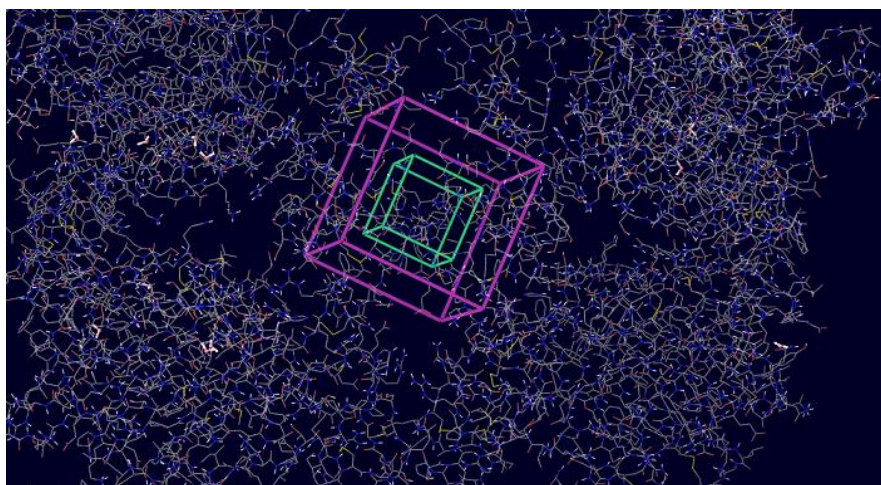
Figure 27D. 1IVO grid |86 230 263 264 265 275 278 283 285 286|W. A dimeric grid from two different aa residues of two clefts. The green cube contains the molecular center of any docked ligand while the purple box (10Å) encloses its entire structure.
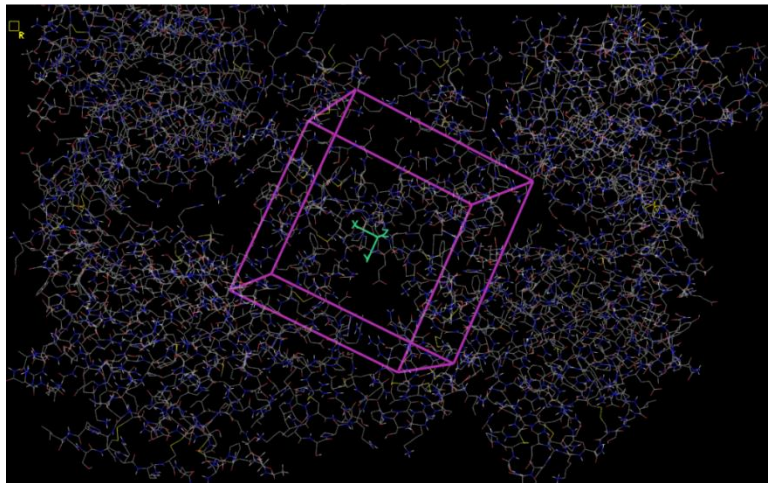
Figure 27E. 1IVO grid |200-300|W.

A complete or whole aa residues 200 – 300 that included residues from both side and backbone chains. The green X, Y & Z are coordinates that define the molecular center of any docked ligand while the purple box (10Å) encloses its entire structure.

2.4 The ligand sources and preparation

The chemical databases we obtained our ligands and their figures in brackets in ready-to-dock format (3D) were ZINC (www.zinc.docking.org) [4,439,415], Ambinter (www.ambinter.com) [5,053,503], SPECS (www.specs.net) [25,030], Cayman (www.caymanchem.com) 4,376], ChEBL (www.ebi.ac.uk) [34,383], Chembridge (www.chembridge.com) [24,971] and Mcule (https:/mcule.com) [5,000] [164]. Thirty one (31) known RTKI ligands were also constructed using v10.4Maestro GUI (graphic user interphase) which effected a 2D to 3D format atructures and minimized with v3.6LigPrep (Schrödinger, LLC, New York, NY, 2015-4). A total of $>9 \times 10^6$ compounds were imported and docked, however, only a significantly low figure had an average of $\geq$ - 5.0 Gscores . Figure 28 shows a description of the role docking process plays in drug discovery and

development [165]. Tables 4, 5, 6 & 13 and 8, 9, 10 & 12 show the top 20 ligands of docked 1IVO and 2J6M, respectively, with respective ligands. Table 13 gives examples of the RTKIs, curcumin, **P** and **Y** compounds docked into 1IVO |246-253|S showing their Gscores.
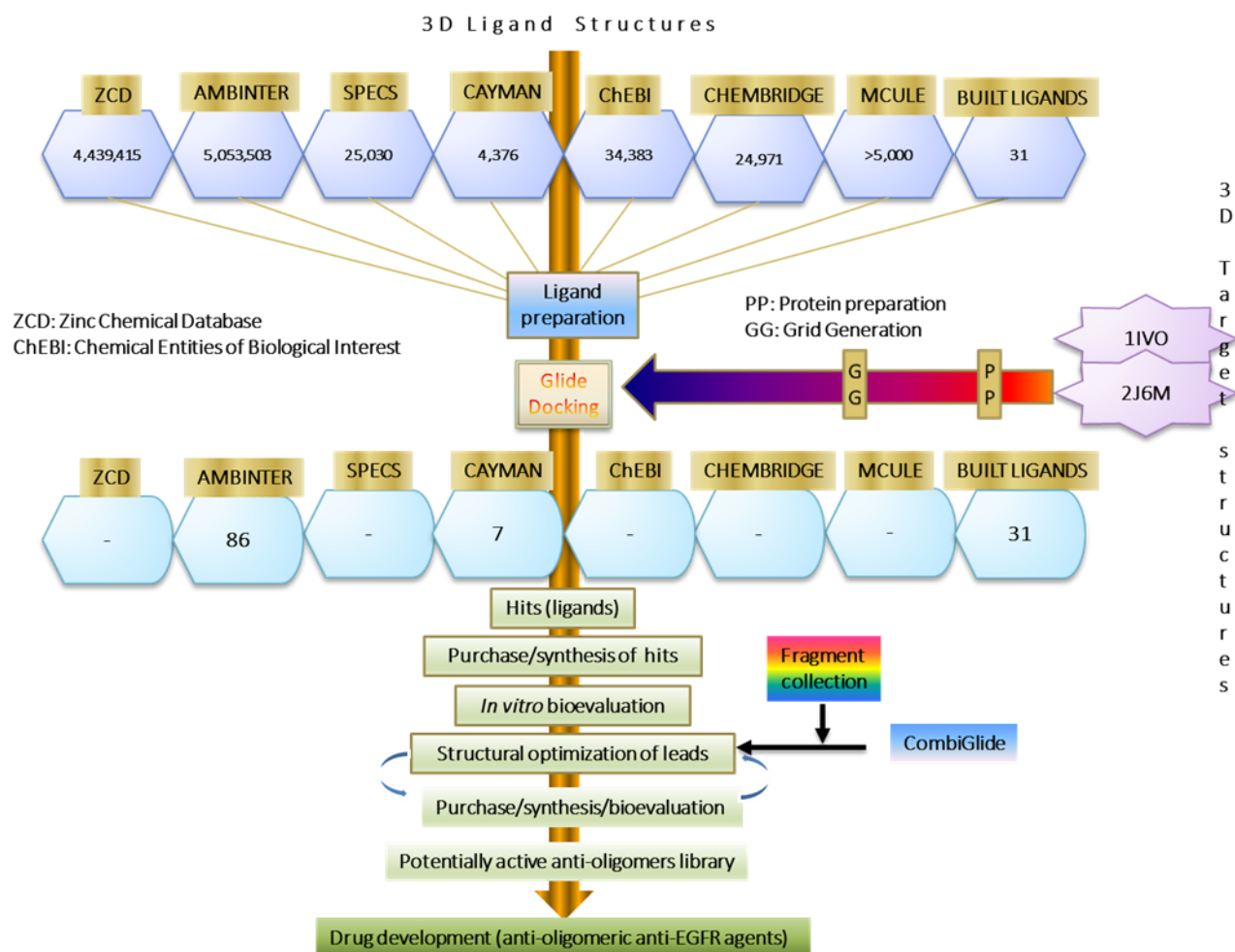
Figure 28. Drug design itinerary model Glide-based.

1IVO and 2J6M grids were ligand-docked on premeditation. 1IVO grids used often were |246-253|S, |246-253|W, |86,230,263,264,265,275,278,283,285,286|W, and |200-300|W. Total number of hits 124: 86 from Ambinter and 7 from Cayman chemical databases and 31 built ligands (RTKIs, curcumin, **P** and **Y**) docked into 1IVO |246-253|S, whose results alone are shown here. The number of ligands used from each source are shown. The 2J6M docked results with the built RTKIs and chemical databases are not reflected in the Table but are found in Section 2.5.3. In addition to considering Lipinski's "Rule of Five" (Ro5) filter Lipinski-like filters were theoretically instrumental in the hierarchical Glide screening of the ligands Table 2). While Table 3 contains DPT1 **1**, DPT2 **2**, DPT3 **3**, DPT4 **4**, DPT5 **5** & DPT6 **6** as ligands that emerged as hit leads, Tables 6 & 10 show that the basis of DPT1 **1**, DPT2 **2**, DPT3 **3** & DPT4 **4** selection was in their potential inhibition at the ectodomain (1IVO) and endodomain (2J6M) regions represented by these crystals, and not necessarily on Gscores. DPT5 **5** & DPT6 **6** are hit leads from 2J6M-ZINC docked ligands, and thus RTKI-like in activity. Enumeration and docking of core leads was achieved with 3.9CombiGlide in order to generate potentially more active anti-EGFR anti-oligomeric agents. CombiGlide offers a continuous method of generating more potentially active compounds for designed libraries of anti-oligomeric discovery and development. Modified from: Hughes et al., 2011 & Kalyaanamoorthy and Chen, 2011.

2.4.1 Modeling and selection of the 'best hit leads'

We saw it important to test-run our generated grid (s) from 1IVO with known compounds such as the RTKIs, cucurmin, **P** and **Y** (already described) in addition to the ligands from the chemical data bases, despite known and unknown mechanism differences. Docking of these compounds into 2J6M revealed their possible activities in the endodomain or cytoplasmic region. Since v6.9Glide allows the choice of filters during the docking process, the Lipinski's "Rule of Five" (Ro5), alongside other Lipinski-like filters have been considered (as Ro5 fails to predict pharmacological properties of an active compound that leads to drugs with higher MW, more rings, more rotatable bonds, and higher lipophilicity, following its molecular structural modification) were applied. The refining of Ro5 had led to development of more empirical rules that almost itemized the biological and physicochemical properties any potential active drug should feature. Among such are Bioavailability filters (BF) [166], Ghose filters (GF) [167], Lead-likeness filters (LLF) [168] , Muegge filters (MF) [169], and Veber filters (VF) [170]. Table 2 shows the Lipinski-like filters and their key features. The potential activities of of the 'hit-to-lead-candidates' were based on the MTS cytotoxicity assay and western blotting, however, the rules assist extra guide.

Table 2. The Lipinski' Ro5 and other Lipinski-like filters.

In addition to the Ro5, there other empirical rules that predict   absorption (A), distribution (D), metabolism (M), excretion (E), and toxicity level (T) (ADMET) of potentially active agents and their main features in drug discovery and development, which been described here.

Lipinski-like filters	Lipinski's rule (Ro5)	Bioavailability filters (BF)	Ghose filter (GF)	Lead likeness (LLF)	Muegge filter (MF)	Veber filter (VF)
Mass (MW)	≤500	≤500	≥160≤480	≥450	≥200≤600	-
logP	≤5	≤5	≥0.4≤5.6	≥7.4≤30	≥-2≤5	-
Hb donor (HBD)	≤5	≤5	-	≤5	≤5	-
Hb acceptor (HBA)	≤10	≤10	-	≤10	≤10	-
Rotatable count (RotC)	-	-	-	≤10	≤15	≤10
PSA	-	-	-	-	≤150	≤140
MSA (SASA)	-	≥200≤800	-	-	-	-
Fused ring count (FRC)	-	≥5≤6	-	-	-	-
Atom count (AC)	-	-	≥20≤70	-	≥5	-
Refractivity (R)	-	-	≥40≤130	-	-	-
logD	-	-	-	≥-4≤4	-	-
Ring count (RC)	-	-	-	≤4	≤7	-

SASA (MSA): solvent-accessible surface area or molecular surface area, PSA: polar surface area; other abbreviations have been suggested for space and use only)

2.4.2 Importance of the Glide-guided R-L bonding

Glide (grid-based ligand docking with energetics) is designed to perform an in-depth search for the positional, orientational, and conformational space available to the ligand as practicable as possible while considering enough computational speed to screen large libraries, [162], however, this was querried during this modeling process. This is accomplished via the use of a series of hierarchical filters (section 2.4.1, Table 2). Glide has three operational stages in the whole docking process: the high-throughput virtual screening (HTVS) wherein all compounds are screened rapidly with both ligand and

receptor being held rigid. The next stage is the Standard Precision (SP) which allows the flexibility of the ligand alone, and Extra Precision (XP) is only for the best scoring states. A pose was saved per ligand in 1000 poses permitted to pass to the *in situ* energy minimization (within the active site). Docking accuracy was assessed by redocking ligands from 150 − 300 cocrystallized PDB complexes starting from conformationally optimized ligand geometries that bear no memory of the correctly docked pose. A notable feature in compounds is their ability to display their rotamers, which are form a number of isomers of a molecule that can be interconverted by rotation of part of the molecule about a particular bond, can influence docking values as well. Geometrical errors are reportedly less than 1 Å for the top-ranked pose in nearly half of the cases and are greater than 2 Å in only about one-third of them [171]. For this reason drug modeling using Glide is currently ranked better among similar techniques [172, 173]. The flexible docking method was preferred to rigid or none (refinely only) or none (score in place only). Glide primary scoring function is Gscore (abbreviated as Gscore) or docking scores (abbreviated as DScores). All predicted binding affinities were within the default root mean square deviation (rmsd) of 0.3 kcal/mol. Although the Gscores should be identical to the DScore, a slight discrepancy was observed when 2J6M was docked with ZINC (Table 9). The docked outcomes were encouraging. Figure 29 illustrates the key functions in the Glide docking hierarchical pattern.
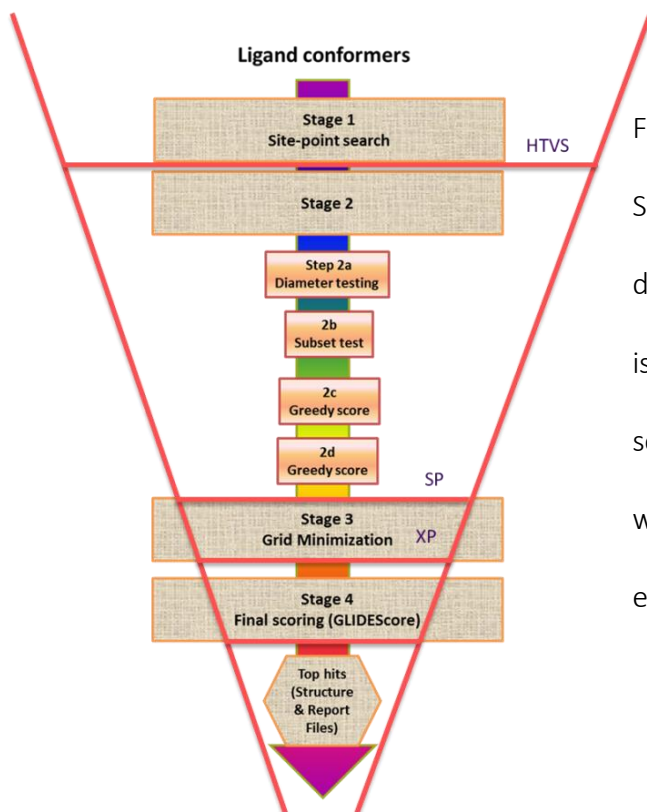
Figure 29. The Glide docking hierarchy. Stage 1 defines the HTVS, Stages 2a – 2d define the SP and Stage 3 defines the XP. It is notable to know that each stage has its scores, the poorest occurring at the HTVS when grid conformers have maximum energy. Basic source Friesner et al., 2004

2.4.3 The 'hit leads' in the light of the Glide parameters

Gscore is used to predict binding affinity and rank-ordering ligands in database screens, and in conjunction with the ligand - receptor molecular mechanics interaction energy and the docked ligand strain energy [162], the correct ligand pose (relative orientation) can be selected. However, Friesner and coworkers observed that Emodel (glide energy model or model energy, ME), which is the composite scoring function in Glide, is much better at selecting the correct pose than is either the molecular mechanics energy or Gscore alone. This supports our choice of ME for our comparison of data in Canvas or Script analysis from the Project Table. Glide is designed to offer an effective high-throughput virtual screening of potential ligands based on the mode of binding affinity for a given receptor

macromolecule. It allows studies and comparisons of scores and geometries of both test and reference ligands (where the two exist), as well as generates conformers that produce 'best' binding modes at lowest free energy (ΔG) possible for newly designed ligands. 1IVO does not have a reference ligand, unlike 2J6M. Ligand and protein preparation optimized as well as minimize structural requirements for a successful docking. The generation of receptor grid supported a rigid protein structure for ease of docking and scoring. Glide then allows the analysis and comparison of ligand binding affinities and their geometries, which can be used to predict ligand features. Sections 2.5.2 and 2.5.6 show some of these features. The Maestro GUI functions principally to facilitate the set up and submissions of jobs to Schrodinger's computational programs through the designing and simplification of modeling tasks, like molecular building and data analysis using v10.4Maestro (Schrödinger, LLC, New York, NY, 2015). Other Glide parameters that waved among themselves to make the docking a success or a failure include the docking score (DS) which defines the Gscore (GS) supplemented by Epik state penalties; Gscore (GS) is composed of the van der Waals (vdW), Coul (Coulomb), Lipo (lipophilicity), Hb, metal rewards, RotBonds and site; glide energy (GE), is the Modified Coulomb-van der Waals interaction energy; glide ligand efficiency (GLE), expresses the relationship between the docking score and number of heavy atoms represented mathematically as (*docking score*)/(*number of heavy atoms*); and glide ligand efficiency surface area (GLESA), is the efficiency metric approximates of the effect of surface area by relating the docking score to the heavy atoms in this manner

(*docking score*)/(*number of heavy atoms*)$^{2/3}$. Another important is ME (already mentioned)

is expressed as:

Gscore + internal ligand strain (Einternal) + Coulomb + van der Waals energy energies

ME is insightful for analytical purposes. LowMW is an example of Glide XP term, which

defines the rewards for ligands with low molecular weights [162, 171]. A few of these

metrics were employed to show if any correlation existed among selected compounds,

particularly the lead candidates DPT1 **1**, DPT2 **2**, DPT 3 **3** and DPT4 **4**, as a level top best

were utilized, which has been done for some compounds prior to and after their structural

modifications with the v3.9CombiGlide (Section 2.5.8). We also used this information to

view our compounds in the face of known compounds, with or without similar

mechanisms. The scatter plot platform was also used to correlate structurally some of

these properties using v2.6Canvas (Schrödinger, LLC, New York, NY, 2015) (Figures 50 – 54

and values were tabulated in Tables 24 – 28 for to view the relationship. The hierarchical

disposition of Glide predisposes its program to high accuracy in the prediction of the

binding mode of the ligand [171] and therefore analyses and views suitable.

2.5 The Docking outcomes, optimization and additional analyses

Although 124 ligands were considered to have emerged, including the 31 built RTKIs,

curcumin and **P** & **Y,** only compounds DPT1 **1** – DPT4 **4** in Table 3 have met our selection

criteria for further investigation as already explained. Compounds DPT5 **5** and DPT6 **6** are

ZINC ligands that had no score with 1IVO **|**246-253**|S** but when docked with 2J6M had the
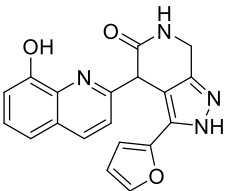
highest Gscores (Table 9). These shall be studied biologically. The lead candidates were optimized to improve potential potency (Tables 20-23). Scatter Plots was to check relationships using their correlation values as explained: if $R^2 = 1.00$, it is said to be a unique link; if $R^2 = 0.00$, no link exists; if $R^2 = \geq 0.50$, fairly good link exists; $R^2 = \leq 0.50$, poor link exists.
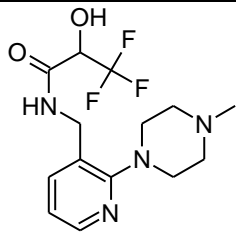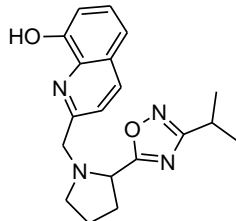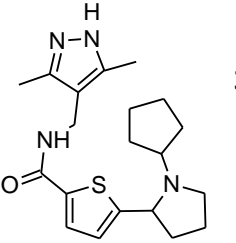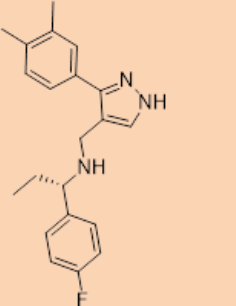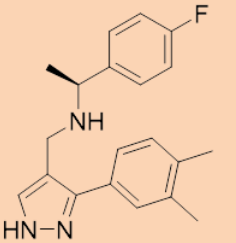
2.5.1 *In silico* results for the 1IVO |246-253|S

Grid |246-253|S prepared from 1IVO monomeric arm was first docked with each of the chemical databases. 2J6M grid was handled the same way. Table 3 shows a summary of the selected and coded hit molecules (section 2.4.3).

Table 3. The six codified selected hits. The selection of DPT1 **1**-DPT4 **4** is based on inhibition of both 1IVO and 2J6M. DPT5 **5** and DPT6 **6**, in orange, inhibited only 2J6M but for their 'best' scores and availability, would be bioevaluated.

NB: Ambinter used the coding as ZINC at the time of docking.

Compound	Structure (2D)	Gscore	Code name
ZINC 65372018	**1**	-2.7	DPT 1 **1**

ZINC 67919965	 **2**	-1.9	DPT2 **2**
ZINC 67851668		2.1	DPT3 **3**
ZINC 20537922	 **3** **4**	-3.5	DPT4 **4**
ZINC56992273		-10.6	DPT5 **5**
ZINC59454528		-10.3	DPT6 **6**

2.5.2 The 1IVO docking outcomes

The Gscores for known RTKIs are shown in 4. The Gscore values were generally low (a typical good Gscore range is -8 to -15). Similar observations occured in other docked ligands (cf. Tables 5 & 6). Tables 4, 5 & 6 show the Gscores and interacting aa residues for the top 20 ligands of RTKIs, Chembridge & Ambinter, respectively, when docked with 1IVO |246-253|S. 1IVO |246-253|S-ZINC ligands did not yield result (cf. Table 9). Figures 34 & 39A -39D show Scatter Plot analyses of Tables 4 & 5.

Table 4. Top 20 docked 1IVO |246-253|S-Built RTKI ligands.

Low Gscores a likely indication of different mechanisms (cf. Table 8).

Ligand	Gscore	Associated residues
CUDC-101	-4.3	TYR246 TYR251
WZ4002	-4.27	TYR246
WZ8040	-3.77	MET253 GLN252 ASP279
Lapatinib	-3.7	TYR246
Tyrphostin AG 478	-3.59	TYR246 MET244
AG-490	-3.47	MET244
WZ3146	-3.33	TYR261 ASP279
PD153035	-3.19	ASP279
AST-285	-3.08	
Mubritinib	-2.98	

Dasatinib	-2.92	HIS281
Gefitinib	-2.68	MET244 TYR246
Dabrafenib	-2.67	
AZ8931	-2.63	TYR251
Canertinib	-2.6	TYR246
Sunitinib	-2.52	TYR246 TYR251
Neratinib	-2.51	TYR246
Dacomitinib	-2.45	GLN252
Pelitinib	-2.45	
ARRY-380	-2.42	TYR246 TYR251

The most featured aa residue (s) linked to a ligand is shown in the pie chart in Figure 30.
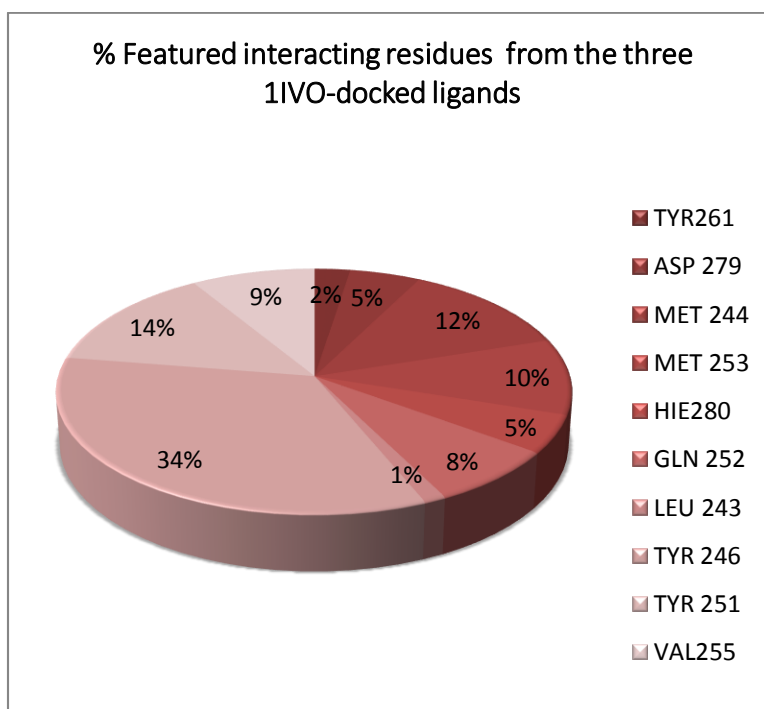
Figure 30. Predominant aa residues in 1IVO |246-253|S-RTKI complexes. The first four residues noticed in most of the interactions were TYR246 (34%), TYR251 (14%), MET244 (12%) and GLN252 (8%) despite low Gscores in the first 20. MET253 was absent. LEU243, HIS280, MET244, and VAL255 were protein side chain residues with unknown contributory mechanism to oligomerization. The information assisted in knowing more about the aa residues from other mechanisms.

(a) y = 5.17x +-22.99 (R^2 = 0.21)

(b) y= 0.02x +-0.03 (R^2 = 0.01)

(c) $y = 4.95x + -16.39$ ($R^2 = 0.28$)

(d) $y = 0.10x + 0.01$ ($R^2 = 0.71$)
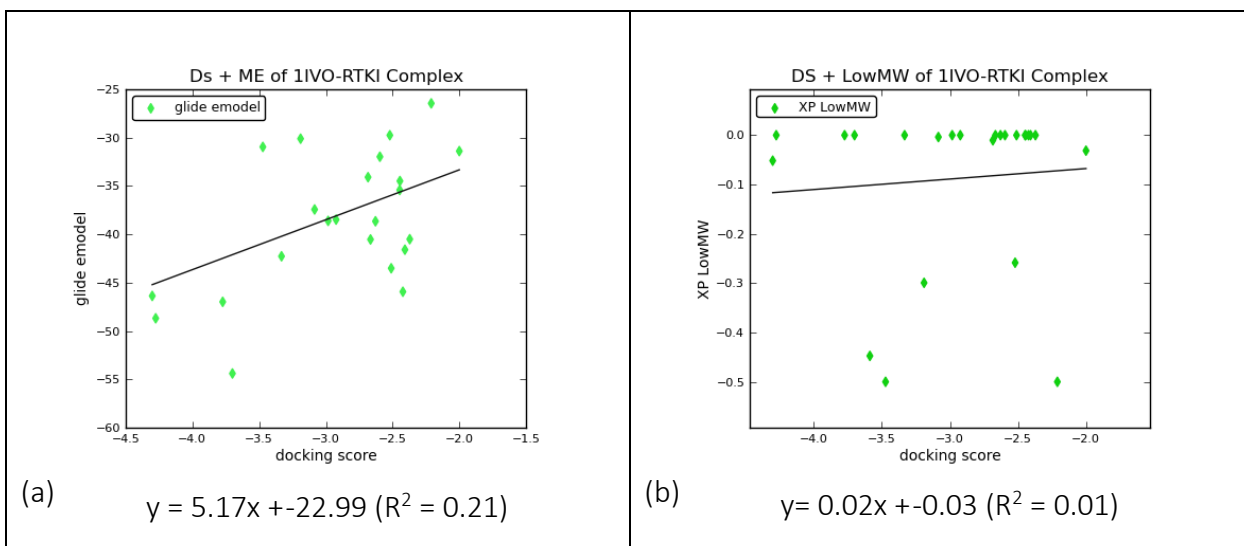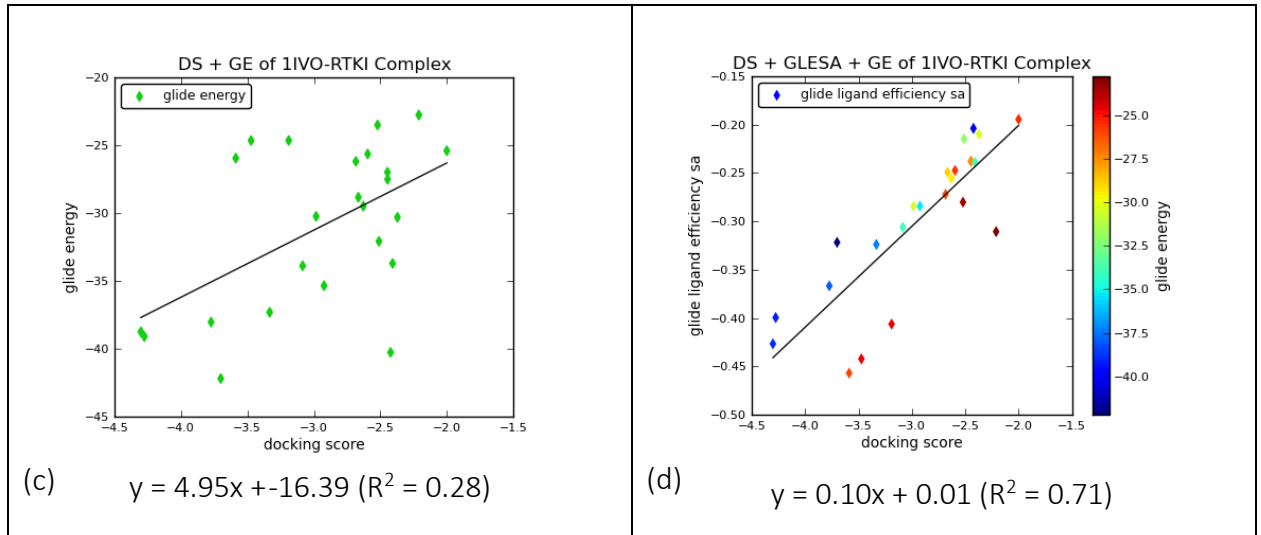
Figure 31. Scatter Plots for top 20 1IVO |246-253|S-RTKI ligands. The selected docking properties are shown. (a) A poor link in DS & ME, (b) A poor link in DS & LowMW exists, and (c) A poor link in DS & GE exists, (d) A fairly good link in DS & GLESA independent of GE.

Table 5. Top 20 docked 1IVO |246-253|S- Chembridge ligands.

Low Gscores and their interacting residues are shown (cf. Table 9).

Ligand	Gscore	Associated residues
ZINC72131899	-4.04	TYR246
ZINC67458838	-4	TYR246
ZINC65425691	-3.4	HIS280 MET244
ZINC67847696	-3.3	TYR246 PRO248
ZINC65461424	-3.04	MET244 THR239
ZINC72457980	-3	MET253 LEU243
ZINC67955977	-2.9	MET253 GLN252
ZINC67847695	-2.79	TYR246 THR239 CYS240
ZINC72457966	-2.74	HIS280 MET244
ZINC65425691-2	-2.71	TYR246 MET244
ZINC67446719	-2.7	TYR246 CYS240 THR239
ZINC23116769	-2.66	GLN252
ZINC72457976	-2.53	TYR261 ASP279
ZINC65461432	-2.47	VAL255 ASN247
ZINC67920036	-2.2	MET253
ZINC02897937	-2.16	MET253
ZINC72131820	-2.05	TYR251
ZINC55271940	-1.92	HIS280
ZINC65425691-3	-1.8	MET253 GLN252
ZINC67847708	-1.74	MET253

The Scatter Plot from the Project Table **v10.4Maestro** (Schrödinger, LLC, New York, NY, 2015-4) was used to analyze 1IVO |246-253|S – Chembridge complexes in Table 5 based on selected properties of DS, GE, GLESA, ME and LowMW [171]. These properties depicted a varying relationship with each other. Scatter Plots 32A to 32D give an insight into some of them.
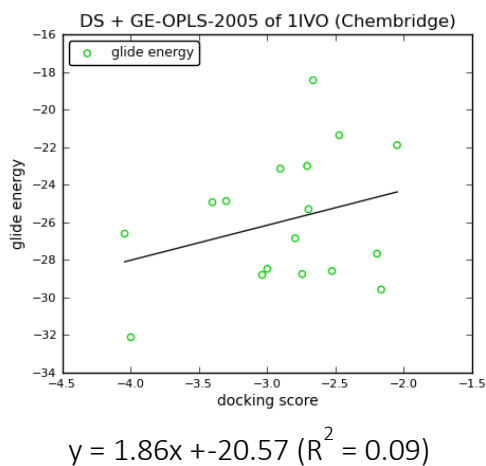
DS + GE-OPLS-2005 of 1IVO (Chembridge) $y = 1.86x + -20.57 (R^2 = 0.09)$	Figure 32A. DS v GE. DS & GE show poor link.
DS + ME OPLS2005 of 1IVO (Chembridge) $y = 1.95x + -24.44 (R^2 = 0.03)$	Figure 32B. DS v ME. DS & ME show poor link.
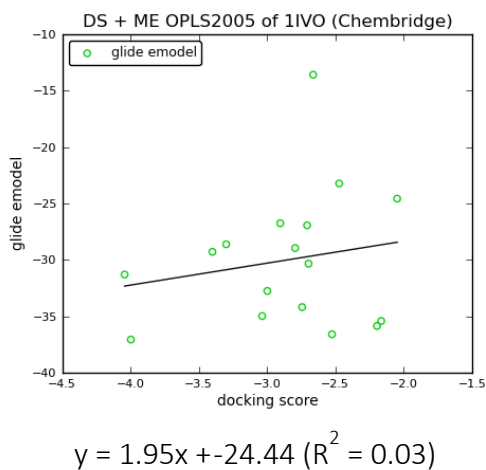
Figure 32C. DS v LowMW.

DS & XP LowMW show poor link

$y = 0.07x +-0.15 (R^2 = 0.11)$

Figure 32D. DS v GLESA.

DS & GLESA show strong link.

$y = 0.12x + 0.00 (R^2 = 0.92)$
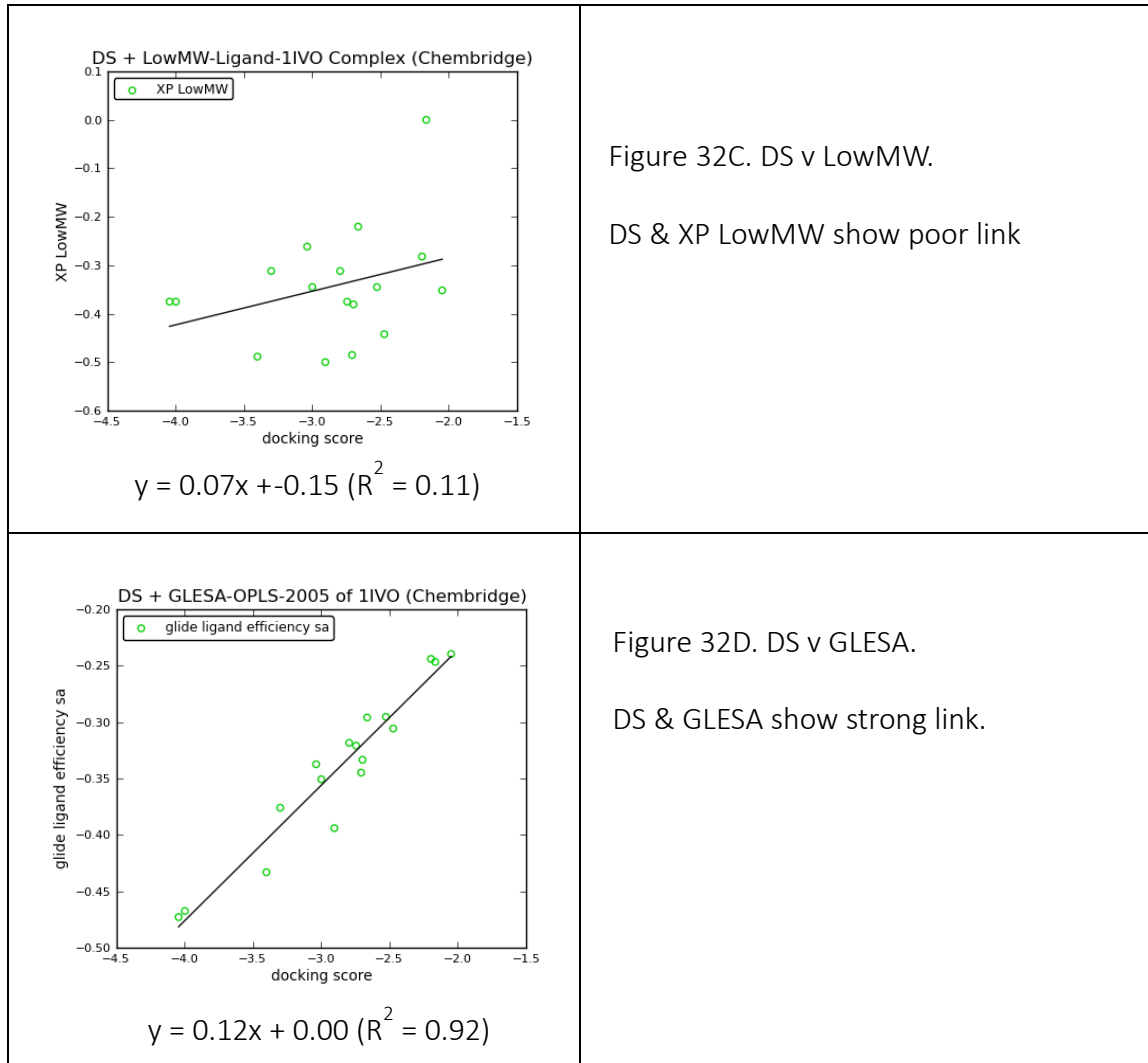
Table 6. Top 20 docked 1IVO |246-253|S- Ambinter ligands.

Low Gscores and their interacting residues are shown.

Colored rows indicate selected hits (cf. Table 10).

Ligand	Gscore	Associated residues	Hit code
ZINC72131821	-3.82	TYR246	
ZINC20537922	-3.53	TYR246 MET244	DPT4 **4**
ZINC20322301	-3.35	TYR251	
ZINC72131820	-3.3	TYR246 MET244	
ZINC67847695	-3.1	TYR246 MET253 GLN252	
ZINC67847694	-2.8	TYR246	
ZINC65372018	-2.73	TYR246 TYR251	DPT1 **1**
ZINC71771450	-2.66	TYR251	
ZINC67741022	-2.61	TYR246	
ZINC67955954	-2.3	TYR246	
ZINC67851668	-2.07	TYR246 TYR251	DPT3 **3**
ZINC72457948	-2.07		
ZINC67774117	-2.03	HIS280	
ZINC02887281	-1.96	MET244	
ZINC00052413	-1.95	TYR246	
ZINC67919965	-1.92	TYR246 TYR251	DPT2 **2**
ZINC20907761	-1.81		
ZINC67955955	-1.71	TYR246	
ZINC39515547	-1.59	TYR251	
ZINC22926613	-0.63		

The Gscores from the monomeric arm are low as mentioned. The v2.6Canvas program

(Schrödinger, LLC, New York, NY, 2015-4) was used to study a few of the properties of the

four selected hits (Tables 6 & 10) utilizing a scatter plot of the GLE $(y - axis)$, GE (in color of

diamond), ME (size of the diamond) and DS $(x - axis)$. A linear relationship was observed

between the DS & GLE (Figure 33). All other experimental properties were within the

calculated values shown in Table 7.
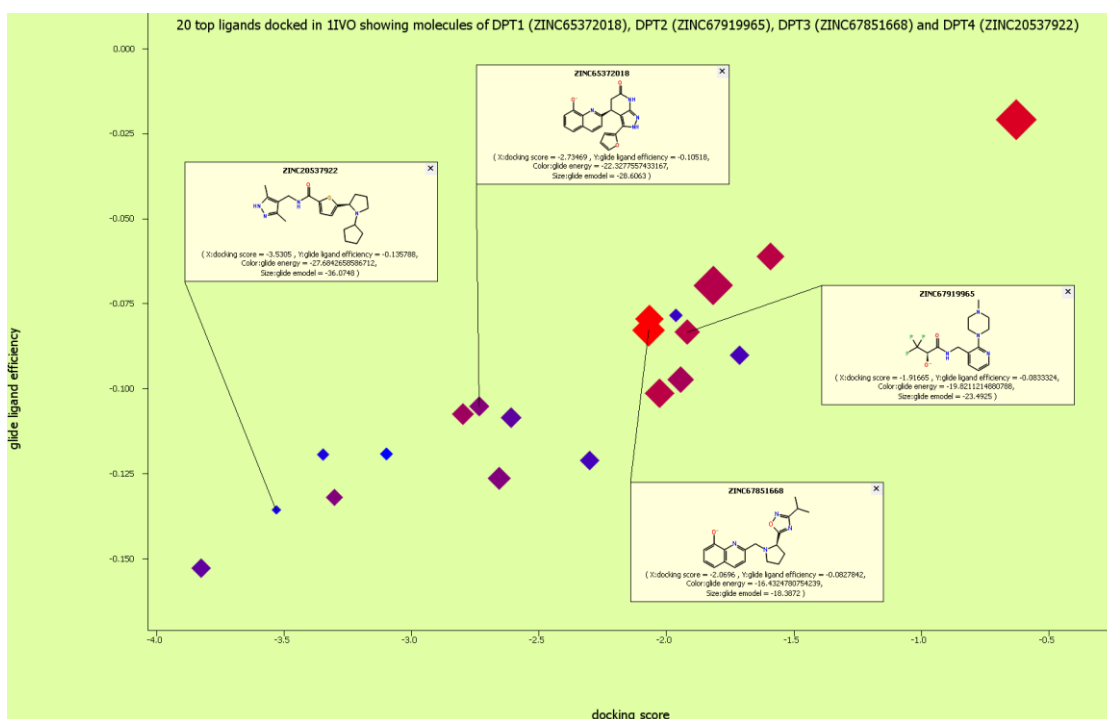
Figure 33. Top 20 1IVO |246-253|S- Ambinter ligands displaying DPT1 **1**- DPT4 **4**.

Using v2.6Canvas Program, the following were studied – DS, GLE, GE (color intensity

of the square) and ME (size of the square). The details of the Scatter Plot are

contained in Table 7.

Table 7. Detailed values for the selected properties of DPT1 **1** - DPT4 **4**. These estimated values shown are from Figure 33. The observed values were within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-36.075	-12.213	-28.944	-16.433
DPT1 (ZINC65372018)	-2.735	-0.105	-28.606		-22.308	
DPT2 (ZINC67919965)	-1.917	-0.083	-23.493		-19.821	
DPT3 (ZINC67851668)	-2.070	-0.083	-18.387		-16.432	
DPT4 (ZINC20537922)	-3.531	-0.136	-36.075		-27.684	

2.5.3 The 2J6M docking outcomes

The docking protocol was repeated with 2J6M and selected scores shown in Tables 8, 10 & 11. The Tables show the scores and type of interacting residues for the first 20 docked ligands.

Table 8. Top 20 docked 2J6M-Built RTKI ligands.

Gscores 3 – 4x higher than their counterparts in 1IVO |246-253|S and their interacting residues are shown (cf. Table 4).

Ligand	Gscore	Associated residues
CUDC-101	-9.24	MET793 ASP855 ASN842
Pelitinib	-9.19	THR854
Gefitinib	-8.95	MET793
Erlotinib	-8.92	MET793 CYS797
Dacomitinib	-8.56	MET793 CYS797
AEE788	-8.48	MET793
Tyrphostin AG 1478	-8.43	MET793
AZ8931	-8.41	MET793
PD153035	-8.28	MET793
Canertinib	-7.99	MET793
BMS 599626	-7.53	MET793 ASP855 LYS745
WZ3146	-7.1	MET793 ARG841
Lapatinib	-6.99	MET793 ARG841 ASP855
ARRY-380	-6.65	GLU804 ARG841
Mubritinib	-6.61	MET793 THR854
AG-490	-6.58	MET793 GLU762
WZ4002	-6.49	MET793
Suramin	-6.47	ARG841 ASP855 GLU758 GLU804
Dasatinib	-6.26	MET793 ASN842
AST-285	-6.13	MET793

The most linked residues on for 2J6M were MET793, GLN791, ASP855, and LYS745 shown

in the pie chart in Figure 32.
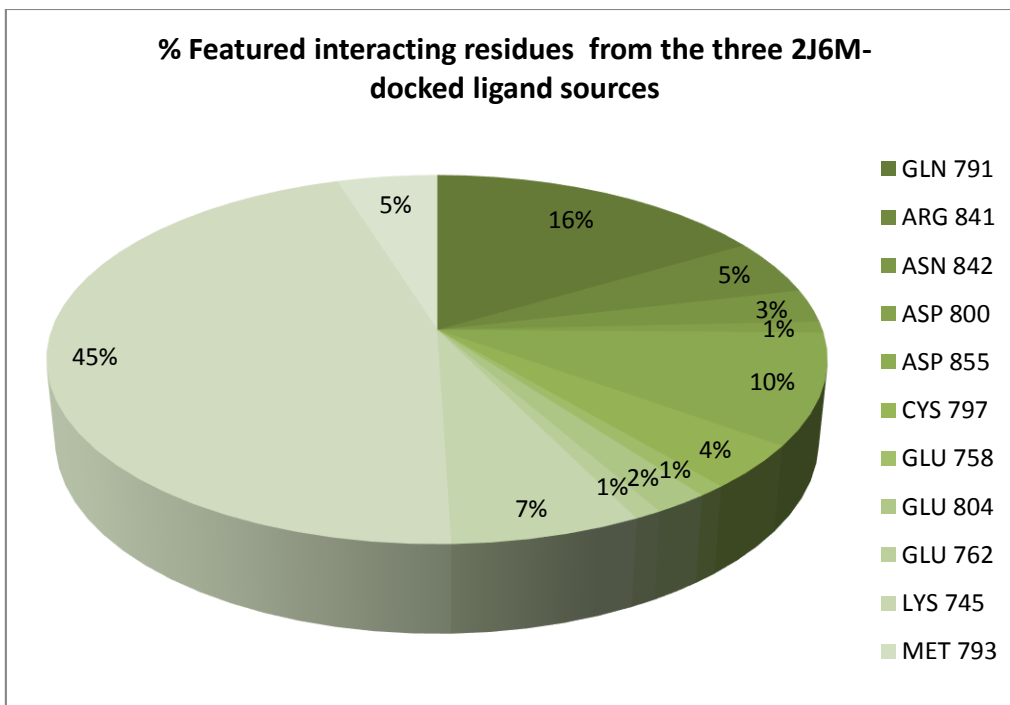
Figure 34. Predominant aa residues in 2J6M-RTKI complexes. The first four residues noticed

in most of the interactions were MET793 (45%), GLN791 (16%), ASP855 (10%) and LYS745

(7%). These residues had the 'best' Gscores among the first 20 in ranking. Other amino acid

residues formed bonds with the protein side chain residues of the kinase domain.

Table 9. Top 20 docked 2J6M-ZINC ligands.

Gscores are 4 - 5x higher than their counterparts in 1IVO |246-253|S

and their interacting residues are shown.

Ligand	GScore	DSocre	Interacting residues
ZINC12301510	-10.58	-9.36	MET793 GLN791
ZINC12302594	-10.33	-9.21	MET793 GLN791
ZINC12302594-2	-10.3	-9.18	MET793 GLN791
ZINC12301510-2	-9.72	-9.14	MET793 GLN791 LYS745
ZINC19774030	-9.71	-8.9	MET793 GLN791
ZINC12298615	-9.19	-8.78	MET793 GLN791
ZINC12298615-2	-9.18	-8.76	MET793 GLN791
ZINC19760143	-9	-8.59	MET793 GLN791
ZINC19760143-2	-8.96	-8.54	MET793 GLN791
ZINC12301510-3	-8.97	-8.39	MET793 GLN791
ZINC12301510-4	-9.59	-8.36	MET793 GLN791
ZINC19800528	-8.73	-8.32	MET793 GLN791
ZINC12301510-5	-8.9	-8.32	MET793 LYS745
ZINC12301510-6	-8.87	-8.29	MET793 LYS745
ZINC12302594-3	-8.84	-8.22	MET793 GLN791
ZINC12301510-7	-8.78	-8.2	MET793 GLN791
ZINC12301510-8	-8.77	-8.19	MET793
ZINC19760143-3	-8.55	-8.14	MET793 GLN791
ZINC11839243	-8.13	-8.12	MET793 ASP855
ZINC12302594-4	-8.74	-8.11	MET793 GLN791

Table 10. Top 20 docked 2J6M-Ambinter ligands.

Gscores 2 - 3x higher than their counterparts in 1IVO |246-253|S their interacted residues shown (cf. Table 6). Colored rows are selected hits based on their antagonistic effects on both 1IVO and 2J6M (cf. Table 6).

Ligand	Gscore	Associated residues	Hit code
ZINC71771450	-7.69	MET793 GLN791 LYS745	
ZINC72131820	-7.4	MET793	
ZINC67847694	-6.95	MET793 ASN842	
ZINC72131821	-6.77	CYS797 ASP800	
ZINC67847695	-6.45	ASP855 LYS745	
ZINC20537922	-6.21	MET793	DPT4 4
ZINC67741022	-5.81	MET793 THR854	
ZINC65372018	-5.63	MET793	DPT1 1
ZINC67955954	-5.62	MET793 ASP855	
ZINC67955955	-5.55	ASP 855 THR854	
ZINC67919965	-5.41	LYS745	DPT3 3
ZINC72457948	-5.16	CYS797	
ZINC20322301	-4.82	PRO794	
ZINC02887281	-4.58	MET793 ARG841	

ZINC39515547	-4.31	THR854 ASP855	
ZINC67774117	-3.99	ASP855	
ZINC22926613	-3.76	MET793	
ZINC20907761	-3.71	ASP855	
ZINC67851668	-3.35	-	DPT3 **3**
ZINC00052413	-1.82	MET793	

The v2.6Canvas program (Schrödinger, LLC, New York, NY, 2015-4) was used to study a few of the properties of the four selected hits (Tables 6 & 10) utilizing a scatter plot of the GLE (y − axis), GE (in color of cross), ME (size of cross) and DS (x − axis). A linear relationship was observed between the docking score and glide ligand efficiency (Figure 35). All other experimental properties were within the calculated values (as shown in Table 11)
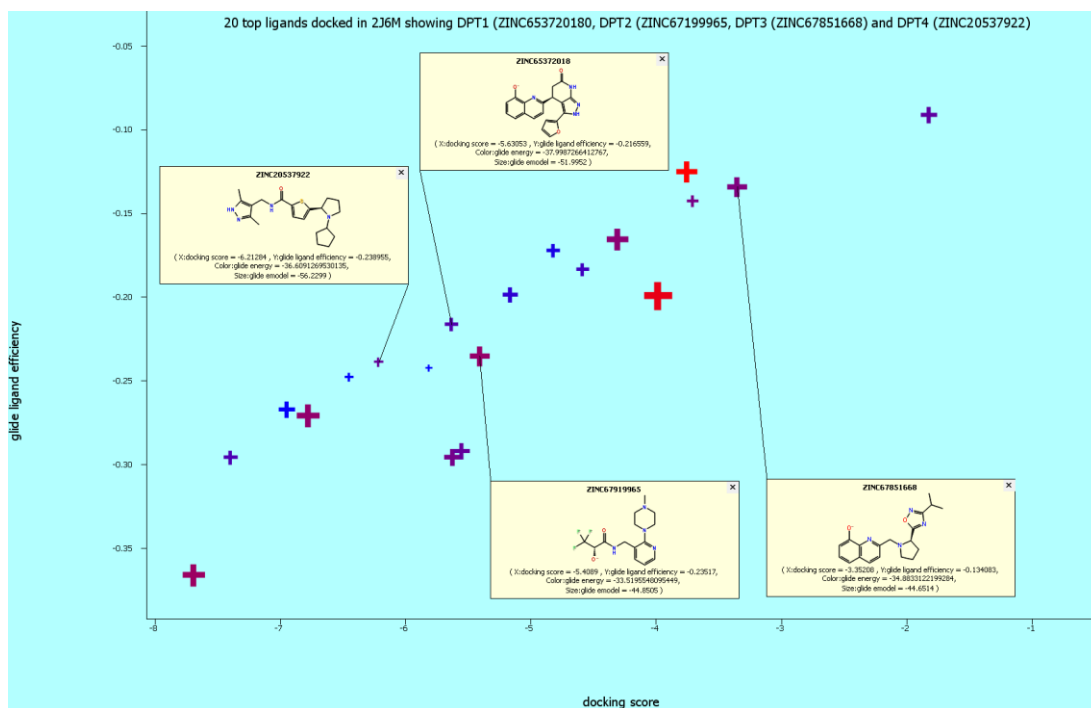
Figure 35. Top 20 2J6M-Ambinter ligands displaying DPT1 **1** - DPT4 **4**. Using v2.6Canvas Program, the following were studied – DS (x – axis), GLE (y – axis), GE (color intensity of the diamond) and ME (size of the diamond). The details of the scatter plot are contained in Table 11.

Table 11. Detailed values for the selected properties of DPT1 **1** - DPT4 **4**. These estimated values shown are from Figure 35. The observed values were within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-36.075	-12.213	-28.944	-16.433
DPT1 (ZINC65372018)	-5.631	-0.217	-52.000		-38.000	
DPT2 (ZINC67919965)	-5.409	-0.235	-44.851		-33.520	
DPT3 (ZINC67851668)	-3.352	-0.134	-44.651		-34.883	
DPT4 (ZINC20537922)	-6.213	-0.239	-56.230		-36.609	

Table 12. Top 20 docked 2J6M-Chembridge ligands.
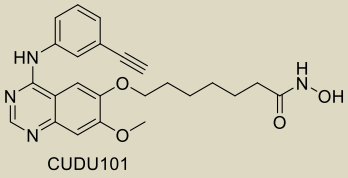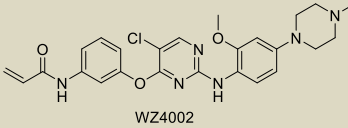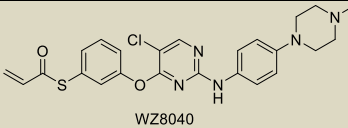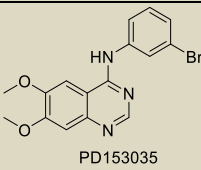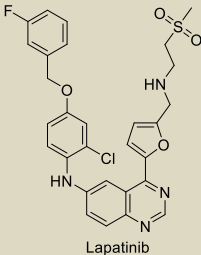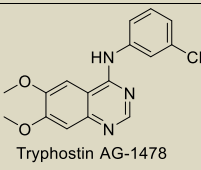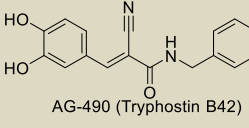
The Gscores 4 -5x those observed in 1IVO |246-253|S showing their linked residues and Gscores (cf. Table 5).
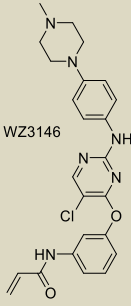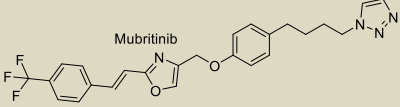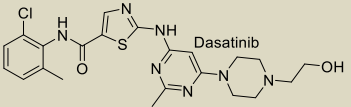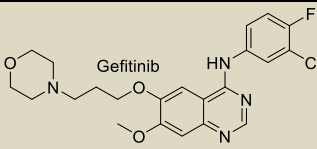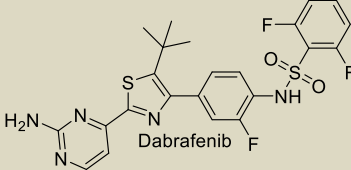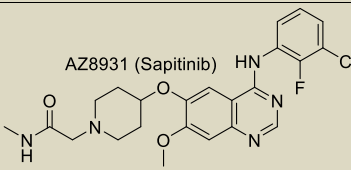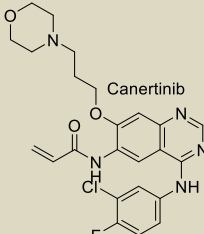
Ligand	GScore	Interacting aa residues
ZINC00064952	-9.09	ALA743 LEU788 MET793
ZINC00309912	-9.07	GLN791
ZINC17489032	-8.6	MET793
ZINC04812637	-8.26	MET793
ZINC16677424	-8.26	GLN791 MET793
ZINC19774030	-8.06	PRO794 LEU1001
ZINC19774030-2	-8	PRO794 LEU1001
ZINC12202453	-7.62	PRO794
ZINC19799281	-6.96	MET793 PHE997 PHE795
ZINC19791293	-6.85	GLU804 PHE997
ZINC12302594	-6.83	ASP994
ZINC19791293-2	-6.4	LEU1001
ZINC12302087	-6.37	LEU1001
ZINC19787192	-6.26	LEU1001
ZINC11788575	-6	LEU1001
ZINC09012972	-5.86	LEU1001
ZINC05041707	-5.79	LEU1001
ZINC12302736	-5.77	ASP800 CYS797
ZINC19772740	-5.67	PRO794
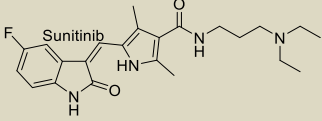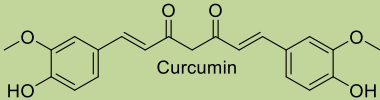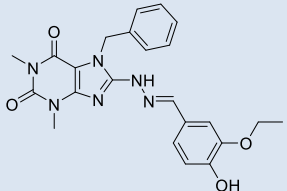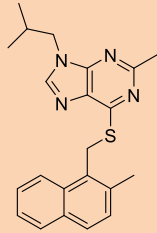
2.5.4 Selected docking properties for top 18 Built compounds for analysis

Fifteen RTKIs (sections 1.4.4 & 1.4.5), curcumin (sections 1.4.2 & 1.4.4) and two reportedly active anti-dimer agents **P** & **Y** (section 1.4.7) were built using the v10.4Maestro GUI, thus the conversion of 2D to 3D structures by LigPrep was accomplished. The ligands were prepared with the v3.6LigPrep (Schrödinger, LLC, New York, NY, 2015-4). These were docked into the monomeric 1IVO grid |246-253|**S** and a selected docking and non-docking properties of these compounds were examined. Curcumin had the highest Gscore (-9.3), however, **P** and **Y** surprisingly had Gscores of -4.8 and -4.3 respectively, similar to some of the docked RTKIs with 1IVO |246-253|**S** (Tables 4, 5 & 6). This offered a broader view for comparative purpose with the evolving anti-oligomeric agents. The docking properties were DS on the x-axis and ME, LowMW, GE on the y-axis as illustrated in Figure 34 (a), (b) and (c), respectively. However, (d) is scatter plot of three docking properties of DS (x-axis), GLESA (y-axis), GE (in rainbow color). There is no clear correlation between the DS and ME shown in (a). The (b) plot exhibits no relationship between the DS and LowMW while (c) shows a slight relationship between the DS and GE. The plot in (d) shows a correlation exists between the DS and the GLESA, however, with GE having no clear effect (Figure 34).

Table 13. 1IVO |246-253|**S**-Built ligands. These Built ligands were selectively built RTKIs (tan color), curcumin (olive green color), **P** (blue color) and **Y** (orange).

Ligand structure	Gscore*	MW$^{\alpha}$	HBD/HBA$^{\beta}$	PSA (Å^2)†	MSA (Å^2)‡
CUDU101	--4.3	434.488	3/8	106	762.93
WZ4002	-4.3	494.973	3/6	91.85	686.15
WZ8040	-3.8	481.013	2/7	99	717.29
PD153035	-3.2	360.205	1/5	56.57	425.65
Lapatinib	-3.7	581.057	2/6	106.35	776.04
Tryphostin AG-1478	-3.6	315.754	1/5	56.27	421.44
AG-490 (Tryphostin B42)	-3.5	294.305	3/4	93.35	389.46

	-3.3	464.947	2/8	82.62	638.59
WZ3146	-3.3	464.947	2/8	82.62	638.59
Mubritinib	-3.0	468.471	0/3	65.97	636.19
Dasatinib	-2.9	488.005	3/6	106.51	665.66
Gefitinib	-2.7	446.902	1/7	68.74	623.66
Dabrafenib	-2.7	519.562	3/7	110.86	658.92
AZ8931 (Sapitinib)	-2.6	473.928	2/8	88.61	649.86
Canertinib	-2.6	485.938	2/8	88.61	651.65

	*	α	β	†	‡
Sunitinib	-2.5	398.474	3/6	77.23	597.95
Curcumin	-9.3	368.380	2/6	93	509.73
(E)-7-benzyl-8-(2-(3-ethoxy-4-hydroxybenzylidene) hydrazinyl)-1,3-dimethyl-3,7-dihydro-1H-purine-2,6-dione. Petch et al, 2012. **P**	-4.8	448.475	2/7	112.29	733.59
9-isobutyl-2-methyl-6-(((2-methylnaphthalen -1-yl)methyl)thio)-9H-purine,Yang et al., 2010. **Y**	-4.2	376.518	0/3	43.60	684.91

*=Docking score, α = molecular weight, β = HBD (hydrogen bond donor) and HBA (hydrogen bond

acceptor), † = polar surface area, and ‡ = molecular surface area (solvent accessible surface area)
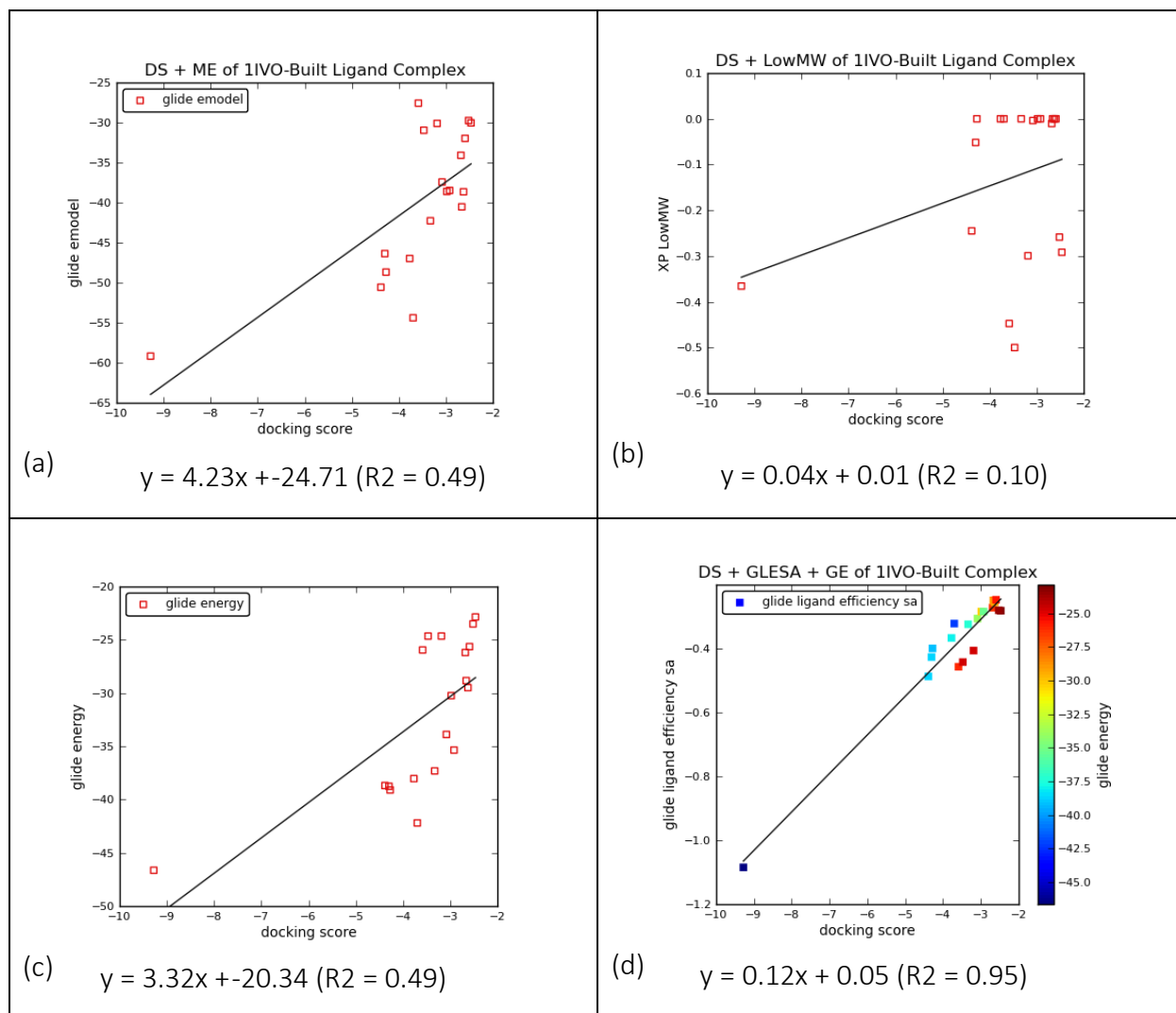
Figure 36. Scatter Plots of top 20 1IVO |246-253|S-Built ligands. (a) DS & ME show poor correlation, (b) DS & LowMW show poor correlation, (c) DS & GE show poor correlation, (d) DS, GLESA & GE show unique correlation.

2.5.5 Visuals of DPT1 1-DPT4 4 in complex with 1IVO |246-253|S

Visualization of the nature of Hbs / HYs formed between DPT1 **1**-DPT4 **4** and 1IVO |246-253|S through their complexing aa residues in Figures 37, 38, 39 & 40, respectively.

2.5.5.1 1IVO |246-253|S in complex with DPT1 1
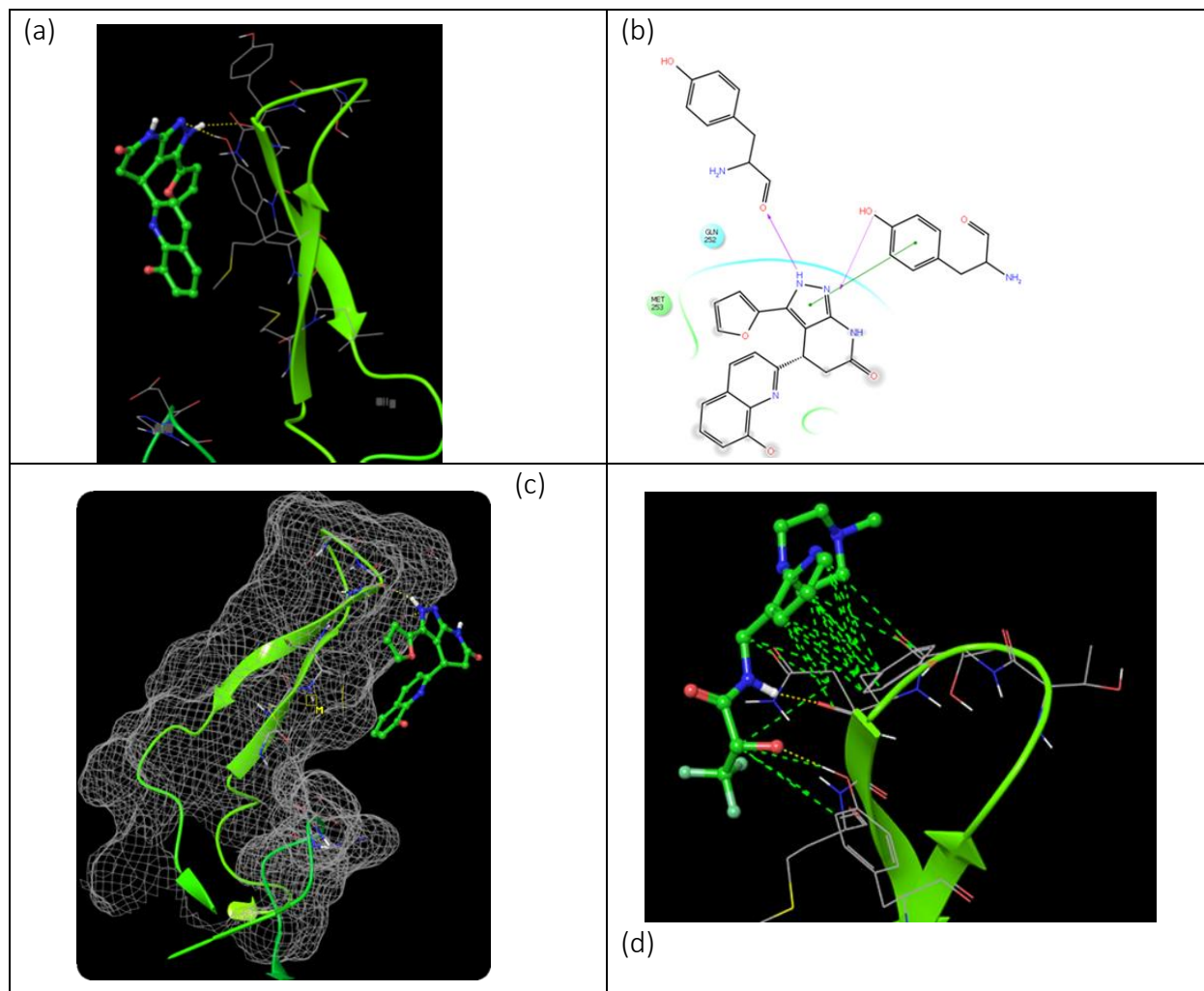
Figure 37. Screens of DPT1 **1** - |246-253|**S** complex. (a): DPT1 **1** formed Hb with primary residues, TYR246 and TYR251, at **T**; (b): an expanded 2D structure of DPT1 **1** showing the mechanism of the Hb and contribution of the π-π interactions formed between the TYR246 aromatic ring and that of the pyrazole ring of DPT1 **1**; (c) a mesh binding site surface created for the receptor and ligand, showing the Hb; and (d) the HY ≈ 10x denser than Hb.

2.5.5.2 Screens of the 1IVO in complex with DPT2 2
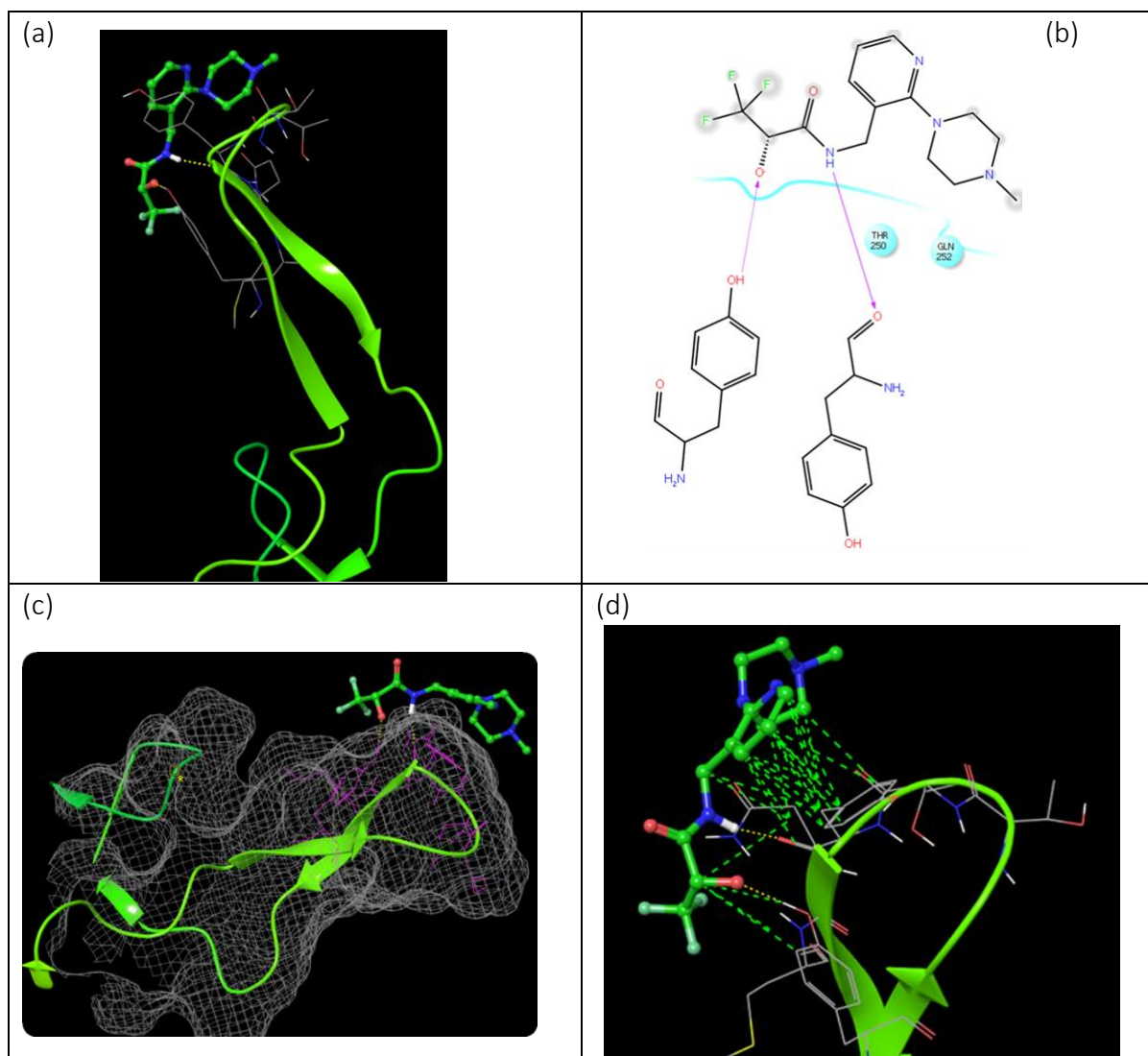
Figure 38. Screens of DPT2 **2** -|246-253|**S** complex. (a): DPT2 **2** formed Hb with primary residues, TYR246 and TYR251 at **T**; (b): an expanded 2D structure of DPT2 **2** showing the mechanism of the Hb that occurred between the residues; (c) a mesh binding site surface created for the receptor, showing the Hb; and (d) the HYs ≈ 10X denser than the Hb.

2.5.5.3 Screens of the 1IVO in complex with DPT3 3
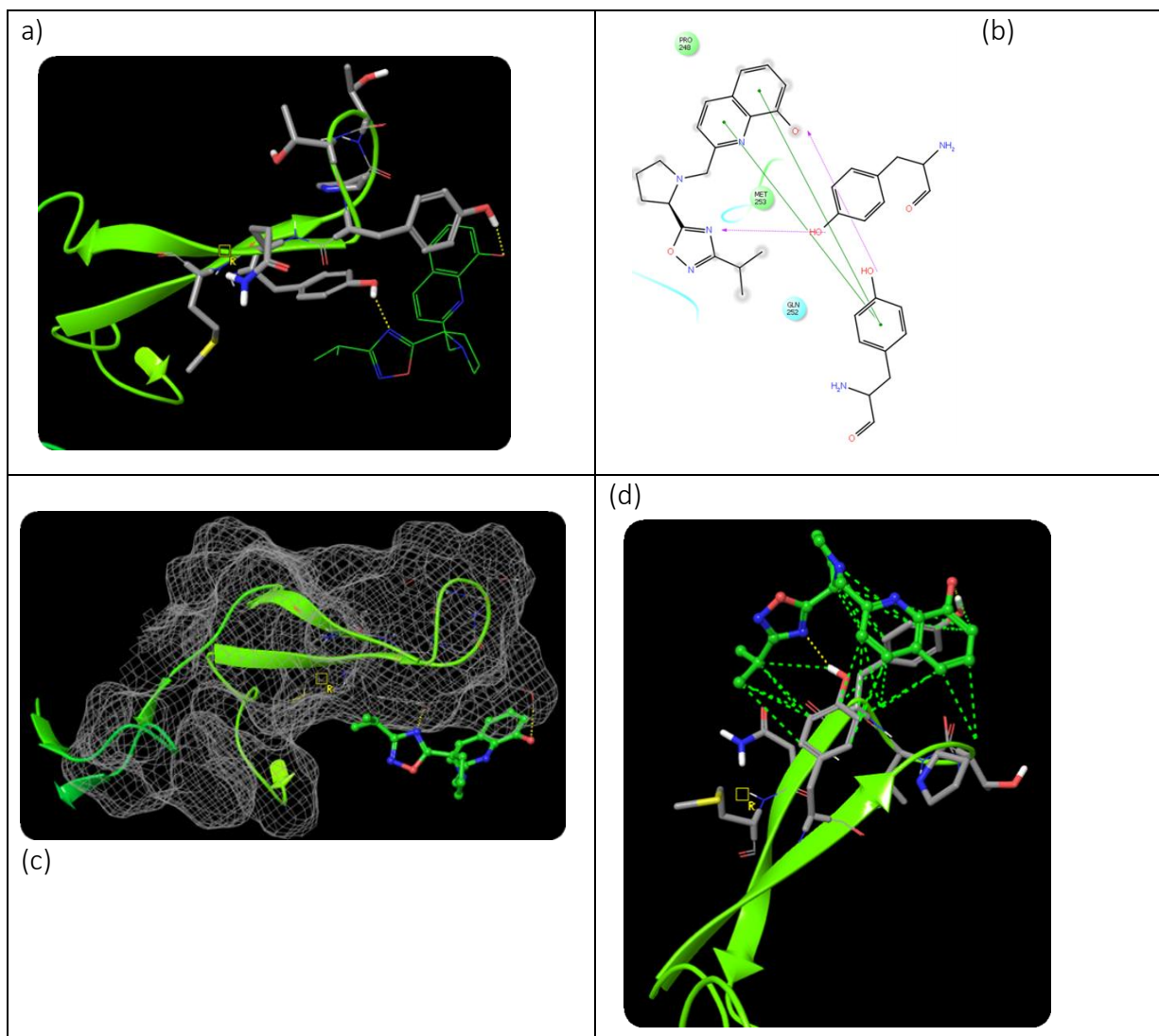
Figure 39. Screens for DPT3 **3**-|246-253|**S** complex. (a): DPT3 formed Hb with primary residues, TYR246 and TYR251 at **T**; (b): an expanded 2D structure of DPT3 showing the mechanisms of the Hb and π-π interactions between the TYR246 aromatic ring and the quinoline ring of the ligand; (c) a mesh binding site surface created for the receptor, showing the Hb, and (d) the HY is ≈ 10X denser than Hb.

2.5.5.4 Screens of the 1IVO in complex with DPT4 4
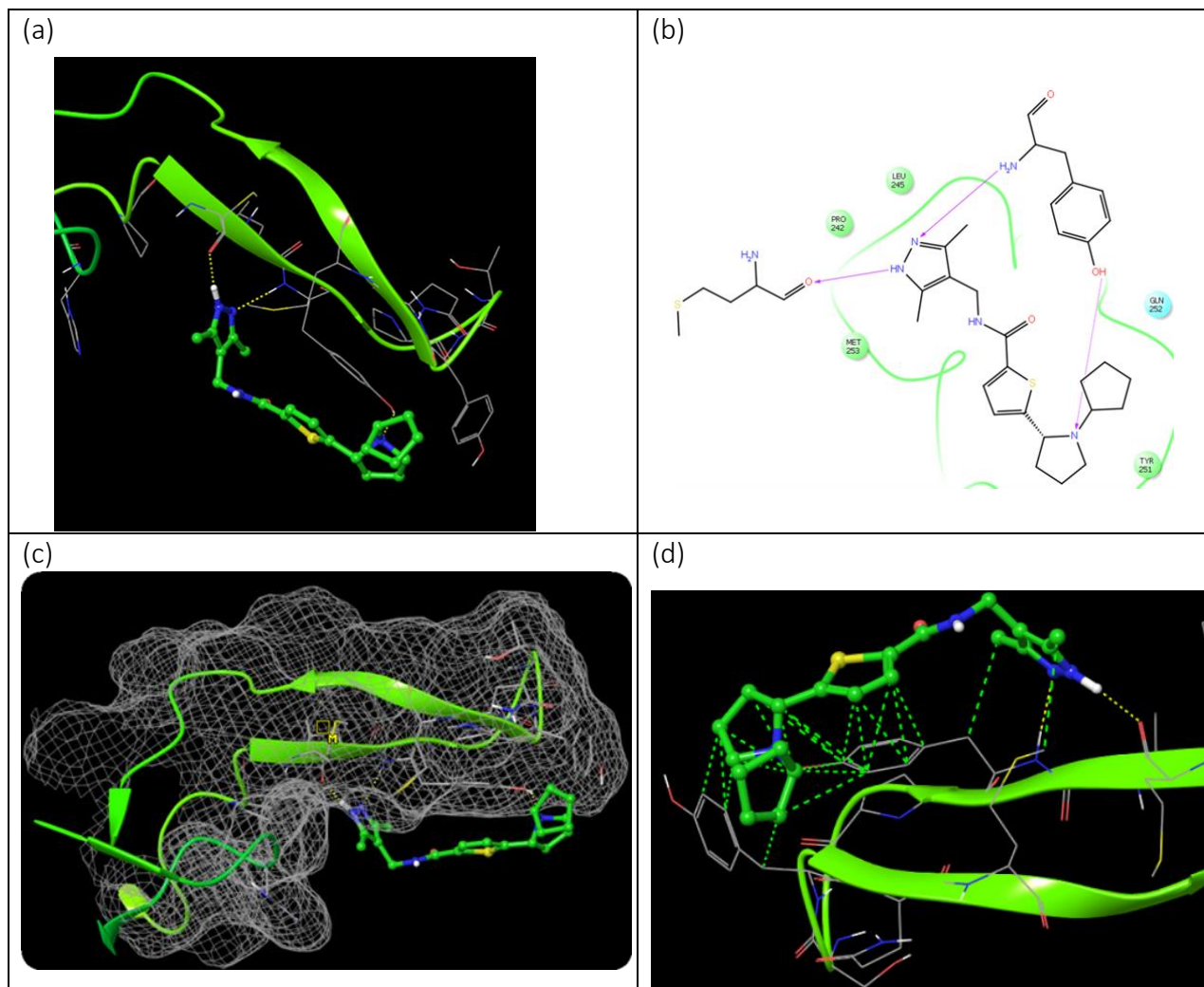
Figure 40. Screens for DPT4 **4**-|246-253|**S** complex. (a): DPT3 formed Hb with primary residues, TYR246 and TYR251 at **T**; (b): an expanded 2D structure of DPT3 **3** showing the mechanisms of the Hb with the TYR246 and MET244 residues; (c) a mesh binding site surface created for the receptor, showing the Hb, and (d) the HY ≈ 10X denser than Hb.

2.5.6 Molecular interactions of DPT1 1-DPT4 4

Glide assists in showing the molecular interactions between the binding residues and candidate compounds. An insight into how the lead candidates interact with the aa residues is useful in studying their mechanisms of action and structural modifications for improved potency. Figures 41A/B, 42A/B, 43A/B & 44A/B show some of these basic properties for DPT1 **1**, DPT2 **2**, DPT3 **3** & DPT4 **4**, respectively.

2.5.6.1 The 1IVO |246-253|S-DPT1 1-aa residues molecular interactions
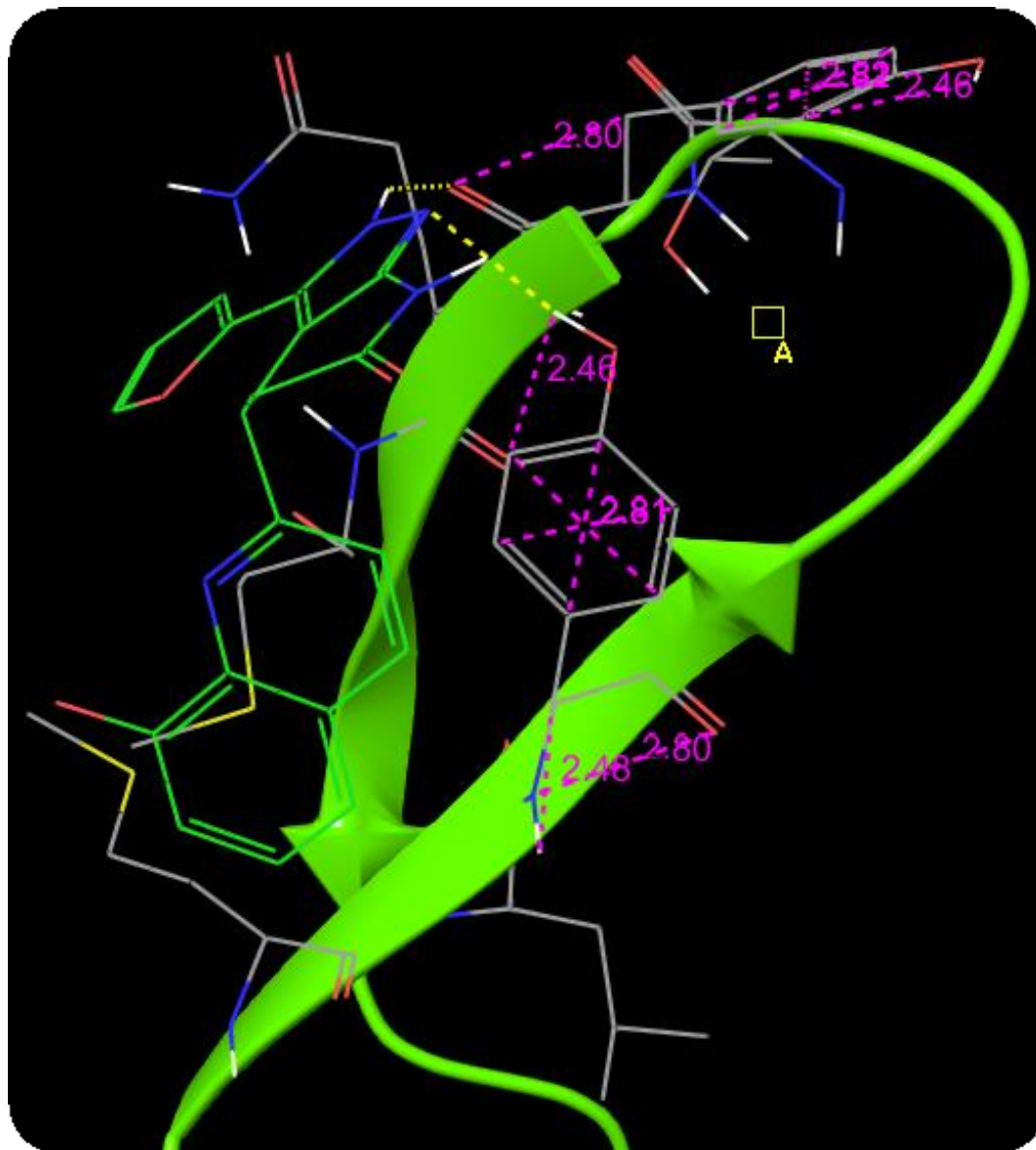
Figure 41A. The 3D picture of the interacting residues with DPT1 **1**.

The key residues are TYR246 and TYR246, their distances from the ligand

and the Hbs in yellow-dotted lines within a space of 4Å are shown. Details

in 2D format are shown in Figure 41B.
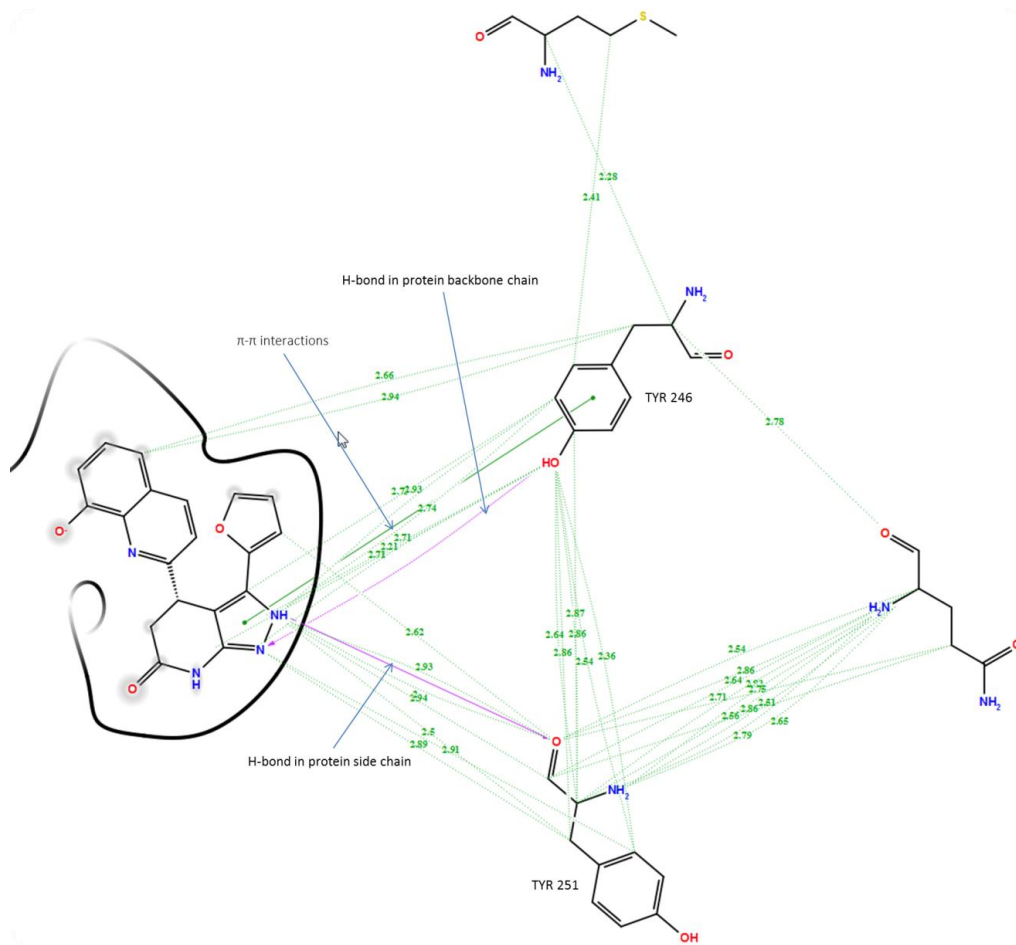
Figure 41B. Analysis of the interacting residues with DPT1 **1**.

TYR246 and TYR251 form Hbs (purple) and a π-π interaction (green) with DPT1 **1**. TYR246 protein side chain displays i) Hb between its H hydroxyl phenyl function and the N in N-20 of the 2,4,5,7-tetrahydro-6H-pyrazolo[3,4-b]pyridin-6-one moiety of DPT1 **1** & ii) π-π interactions between its phenyl ring and pyrazolo- moiety of DPT1 **1**; TYR251 protein backbone chain-a displays a Hb between its O carbonyl group and the H on N-19 pyrazolo-pyridin-6-one moiety of DPT1. These predict the mechanisms of DPT1 **1** as an anti-oligomerization agent within these approximate distances and angles summarized in Table 14.

2.5.6.2 The 1IVO |246-253|S residual interactions with DPT2 2
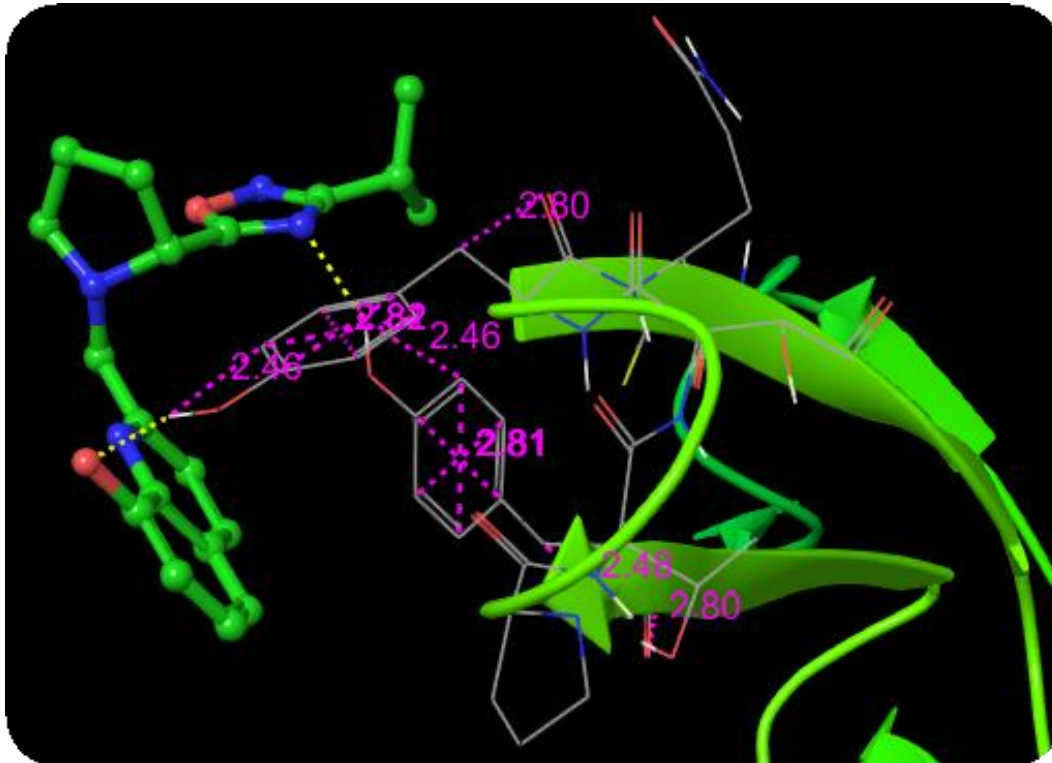
Figure 42A. The 3D picture of the interacting residues with DPT2 **2**.

The key residues are TYR246 and TYR251, distances away from the ligands, and

the Hbs in yellow-dotted lines within a space of 4Å. Details in 2D format are

shown in Figure 42B.
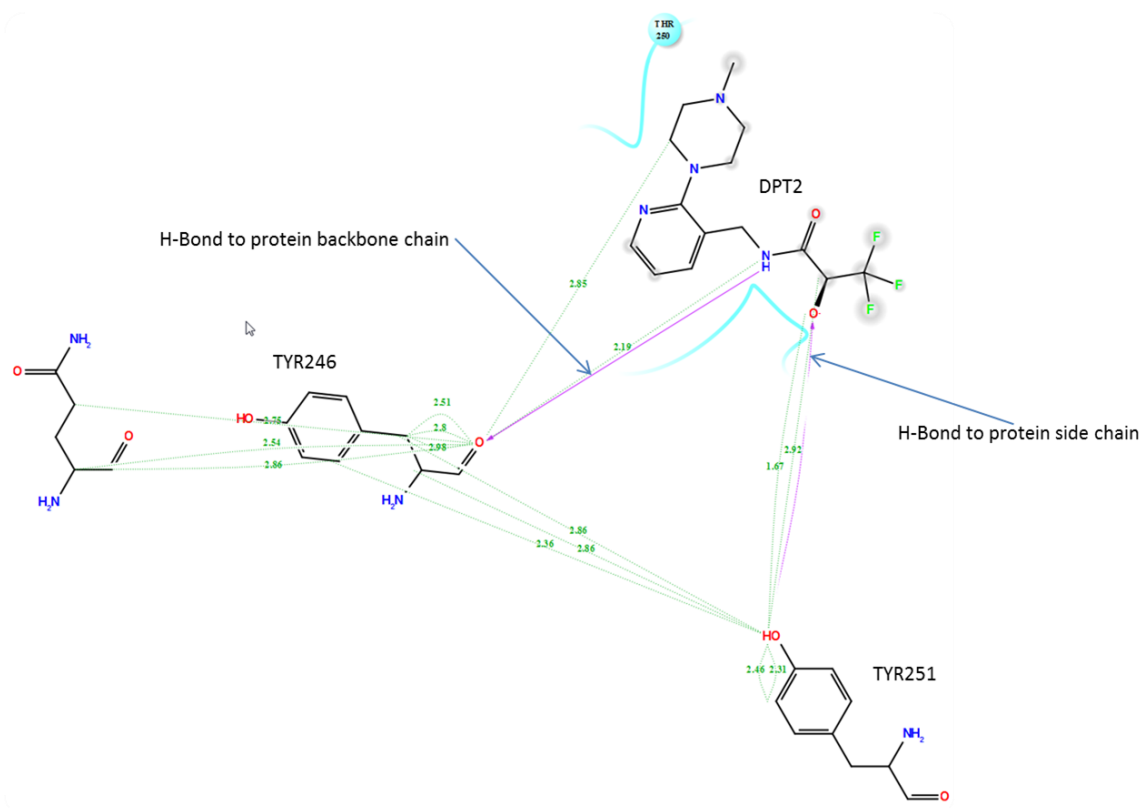
Figure 42B. Analysis of the interacting residues with DPT2 **2**.

The interacting residues with DPT2 **2** are TYR246 and TYR251. The Hb (in purple) is as a protein side chain (as with TYR251 in purple line) or protein backbone (as with TYR246 in purple line). These predict the mechanisms of DPT2 **2** as an anti-oligomerization agent within these approximate distances and angles summarized in Table 14.

2.5.6.3 The 1IVO |246-253|S residual interactions with DPT3 3
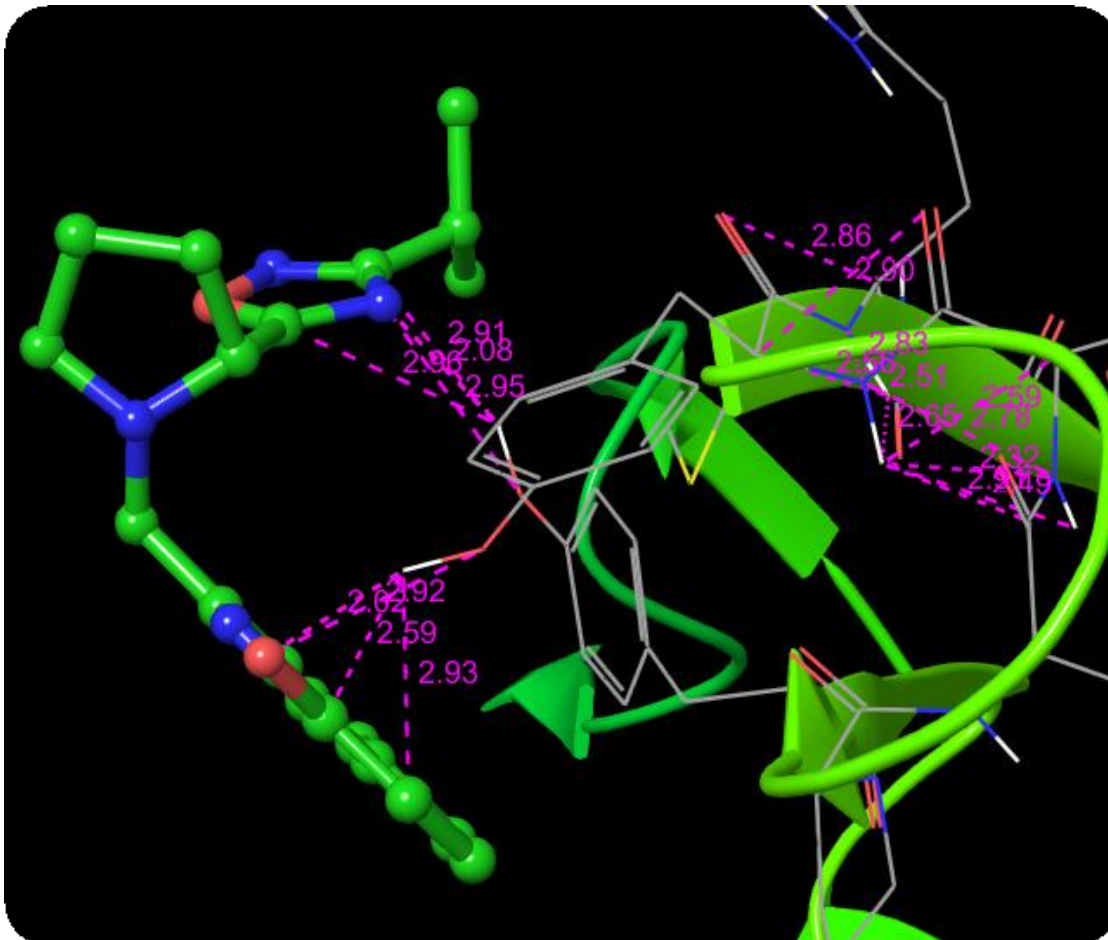
Figure 43A. The 3D picture of the interacting residues with DPT3 **3**.

The key residues are TYR246 and TYR251, distances away from the ligands, and

the HBs within a space of 4Å are shown. Details in 2D format are shown in Figure

43B.
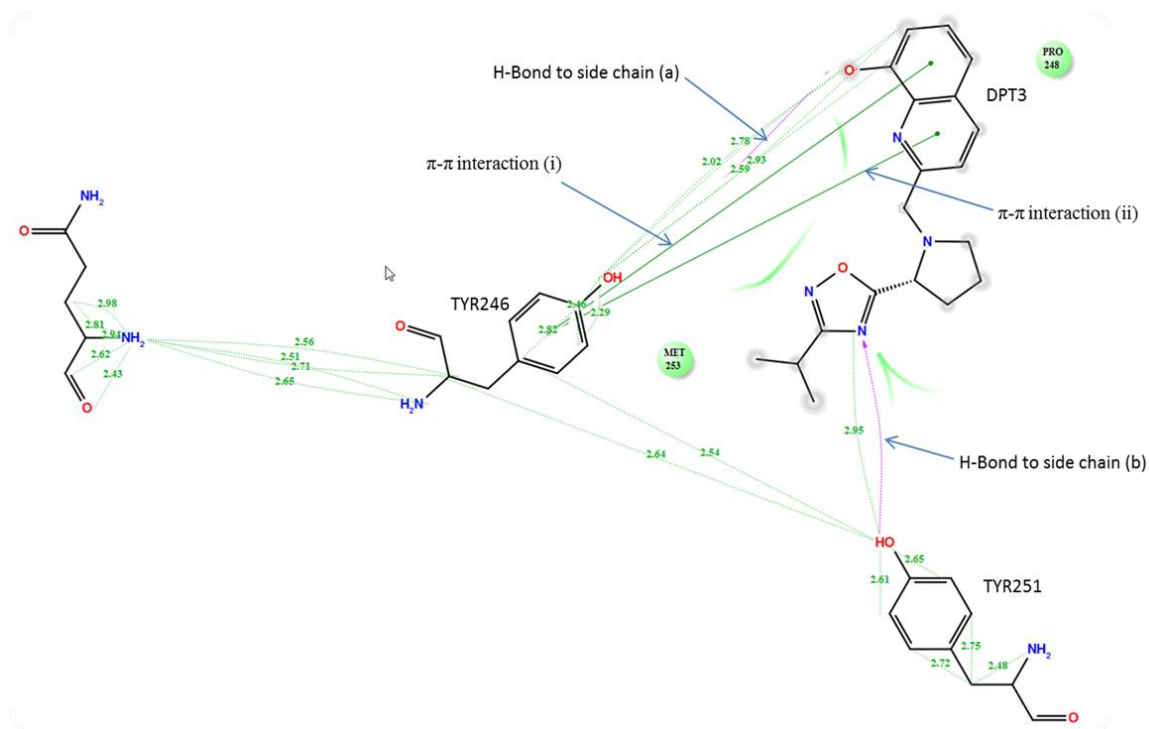
Figure 43B. Analysis of the interacting residues with DPT3 **3**.

TYR246 and TYR251 form Hbs (purple) and a pair of π-π interactions (green) with DPT3.

TYR246 protein side chain – *a* displays i) Hb between its H hydroxyl group and the O hydroxyl group on C-6 naphthyl ring of DPT3 **3** & ii) a pair of π-π interactions ((i) & (ii) as labeled in the diagram) between its phenyl ring and the naphthyl rings of DPT3 **3**; TYR251 protein side chain – *b* displays Hb between its H hydroxyl function and N-19 of the oxadiazole ring of DPT3 **3**. These predict the mechanisms of DPT3 **3** as an anti-oligomerization agent within these approximate distances and angles summarized in Table 14.

## 2.5.6.4 The 1IVO |246-253|S residual interactions with DPT4 4
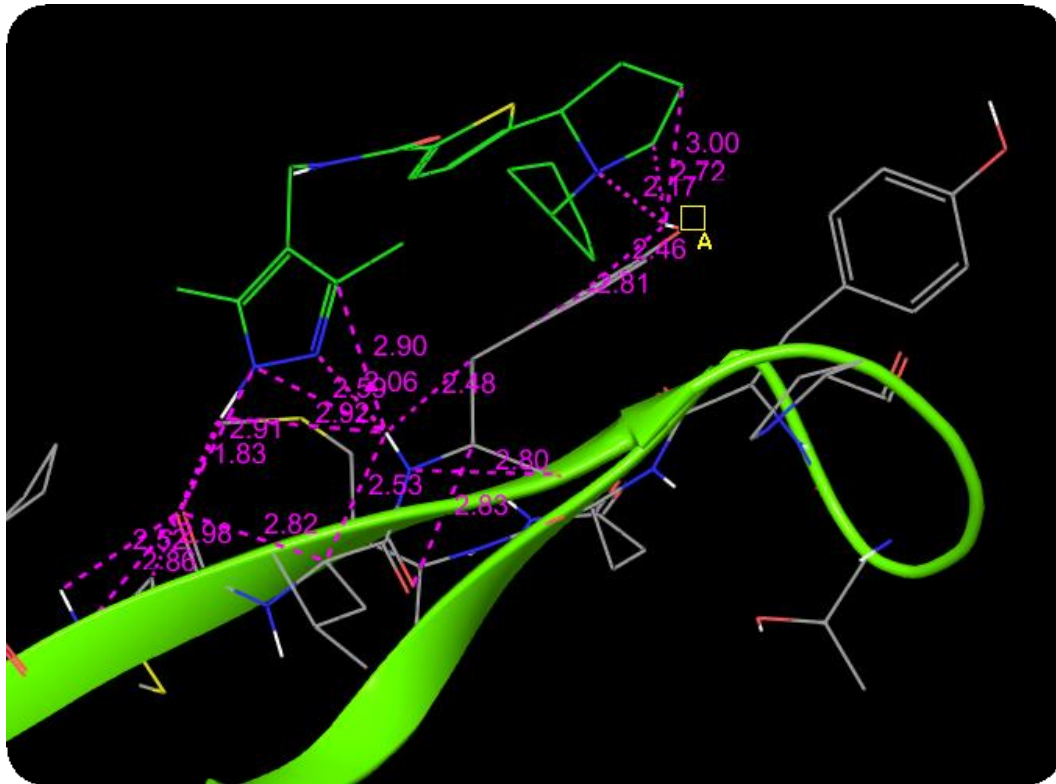
Figure 44A. The 3D picture of the interacting residues with DPT4 **4**.

The key residues are TYR246 and TYR251, distances away from the ligands, and

the Hbs within a space of 4Å are shown. Details in 2D format are in Figure 43B.
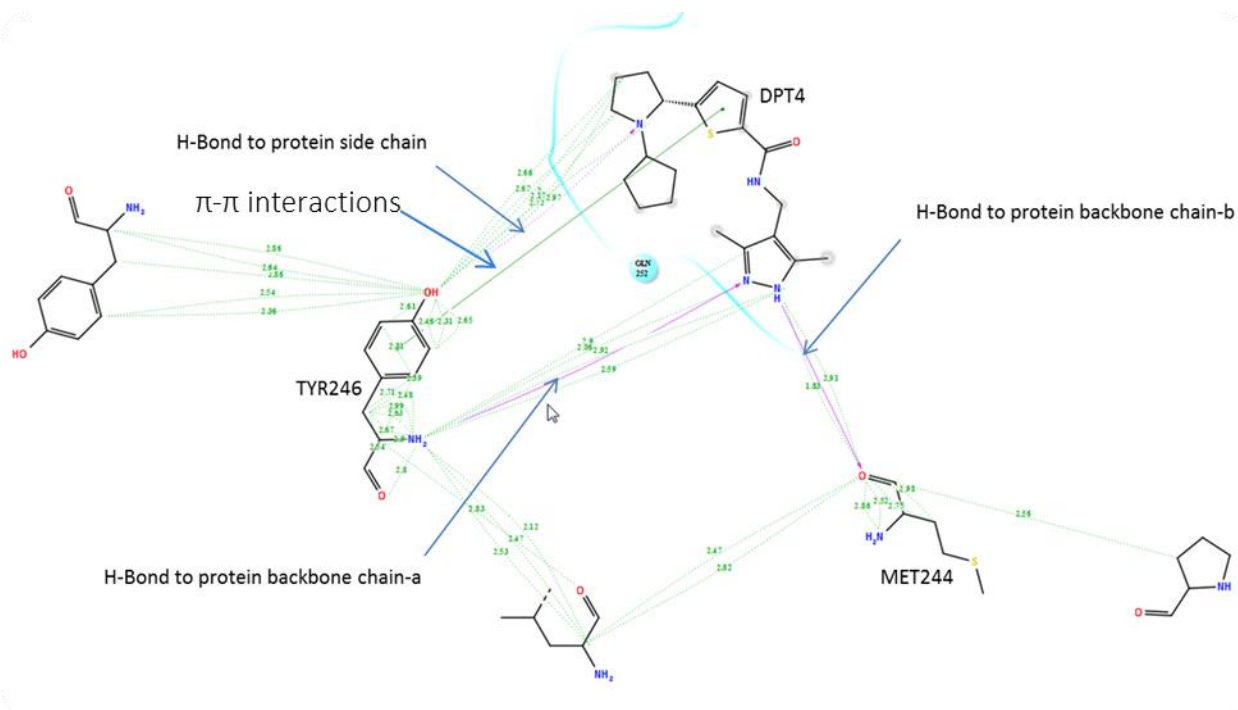
Figure 44B. Analysis of the interacting residues with DPT4 **4**.

MET244 and TYR246 form Hbs (purple) and a π-π interaction (green) with DPT4 **4**. MET244 protein backbone chain – *b* displays Hb between its O carbonyl group and the H of N-22 of the pyrazole ring of DPT4 **4**; TYR246 protein backbone chain – *a* displays Hb between its H amide group and N-21 of the pyrazole ring of DPT4 **4**; TYR246 protein side chain displays i) Hb between its H hydroxyl function and N-6 pyrrole ring of DPT4 **4** ii) a π-π interaction between its phenyl ring and thiol ring of DPT4 exists. These predict the mechanisms of DPT4 as an anti-oligomerization agent within these approximate distances and angles summarized in Table 14.

The summary of the residual interactions of DPT1 **1**, DPT2 **2**, DPT3 **3**, and DPT4 **4** are shown in Table 14.

Table 14. A précised molecular studies of DPT1 **1**-DPT4 **4** within 4 Ǻ.

Figures 40B, 41B, 42B, and 43B show that the mode of aa residual interactions with the ligands in terms of their distances and angles within a specified space, in this case 4Å, can be predicted. There are two types of Hbs formed - those with the protein backbone chain and those with the protein side chain. The protein side chain lacks an acceptor angle while the protein backbone chain possesses both acceptor and donor angles. The π-π interacts in a manner unique to the ligand, thus DPT2 **2** does not have any, DPT4 **4** interaction is face-to-face, DPT3 **3** exhibits both face-to-face and edge-to-face, and DPT1 **1** shows edge-to-face at the different distances with the face seemingly having shorter distance than the edge-to-face. Only DPT2 **2** lacks protein backbone chain. The largest acceptor angle 168.59^{0} from DPT4 **4** that already exhibits two donor angles 94.95^{0} and 137.59^{0} show its high potential for structural optimization. DPT2 **2** exhibits the highest donor protein side chain angle, 156.72^{0} and only DPT3 **3** shows two different protein side chain angles, 97.81^{0} and 118.50^{0}. These parameters render the binding affinities, hence their anti-EGFR, individually unique.

Feature		DPT1 1	DPT2 2	DPT3 3	DPT4 4
Interacting residue (s)		TYR246 TYR251	TYR246 TYR251	TYR246 TYR251	TYR246 MET244
Hb to protein side chain	Distance from ligand (Å)	2.210	1.667	(a) 2.016, (b) 2.078	2.169
	Donor angle (^{0})	96.15	156.72	(a) 97.81, (b) 118.50	117.69
	Acceptor angle (^{0})				
Hb to protein backbone chain	Distance from ligand (Å)	1.997	2.192		(a) 2.062, (b) 1.829
	Donor angle (^{0})	130.14	149.98		(a) 94.95, (b)137.32
	Acceptor angle (^{0})	151.51	133.76		168.59
π-π interactions	Nature of interaction	Edge to face		Edge to face (i), Edge to face (ii)	Face to face
	Distance (Å)	4.974		4.819 (i), 5.307 (ii)	3.818
	Angle (^{0})	76.40		67.75 (i), 68.48 (ii)	11.17

The role of MET 244 has an unknown anti-dimerization activity.

2.5.7 Docking outcomes of some 1IVO complete (W) grids
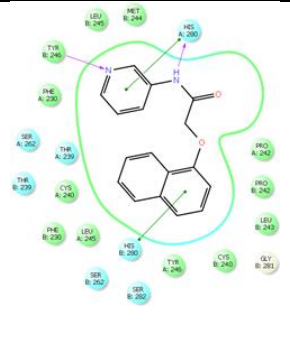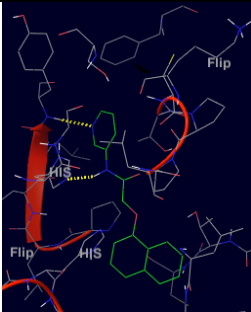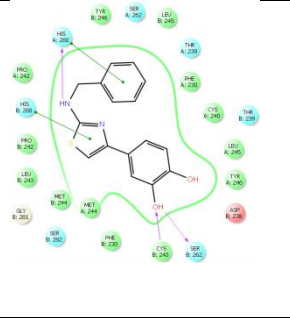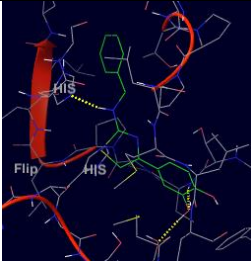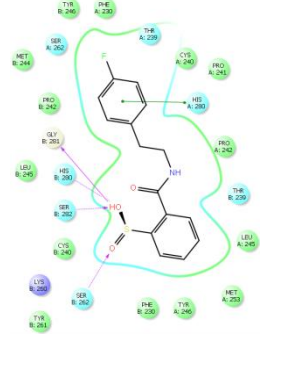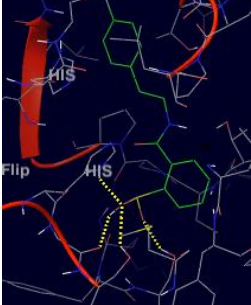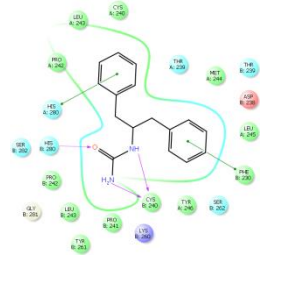
An initial no Gscore when |246-253|S was docked with ZINC ligands through obtaining low Gscores to deliberate investigational reasons prompted the need to generate 1IVO grids that spanned the entire 'oligomerization region' (already defined). Tables 4, 5 & 6 show

some of the low Gscores when docked with the monomeric arm. The opposite was observed when the same compounds were docked with 2J6M as 3 – 5x Gscores of the 1IVO monomeric grids were obtained (Tables 8, 9 & 10), particularly Table 9 of 2J6M-ZINC ligand complexes, whereas no score was observed with the 1IVO |246-253|**S**-ZINC ligand docking. The viable dimeric grids generated include |246-253|**W** and |86, 230, 263, 264, 265, 275, 278, 283, 285, 286|**W**, representing **T** and **B** residues, respectively. Another 1IVO grid that reflected the entire 'oligomerization region' was 1IVO |200-300|**W** since its aa residues were from both the side chain and backbone chain of the 1IVO. All these grids were docked with the same chemical databases with outcomes very encouraging since average cutoff Gscores ≥ -6.0, which is still 3 – 4x those of DPT1 **1** – DPT4 **4**. Only a selected few dimeric docking outcomes are shown in Tables 15 – 18.

2.5.7.1 1IVO |246-253|W-Chembridge ligands

Table 15. Top 4 of docked |246-253|**W**-Chembridge ligands.
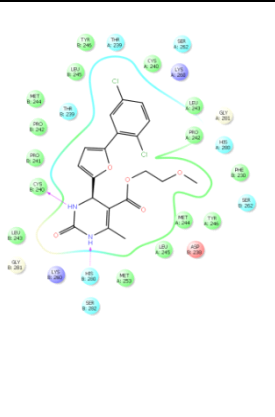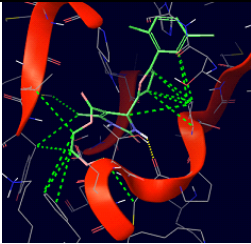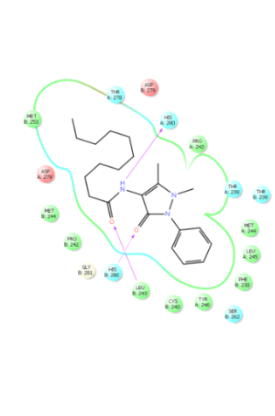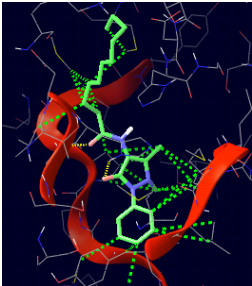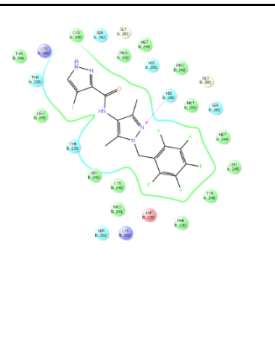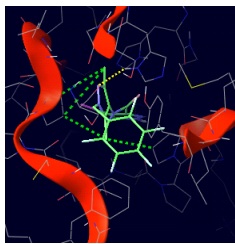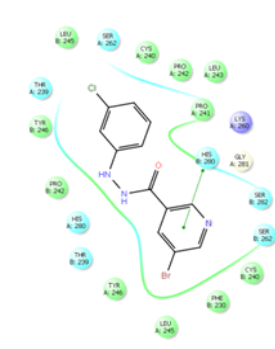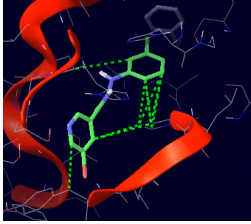
Gscores of 3- 4x higher than those from the |246-253|**S** and their aa residues are shown (cf. Table 5) (where L-R, ligand-receptor complex).

Ligand	Gscore	Interacting residue (s)	2D L-R complex	3D L-R complex
ZINC04661187	-8.1	TYR246 HIS280		
ZINC00784489	-7.6	CYS240 SER262 HIS280		
ZINC22018488	-7.6	GLY281 SER262		
ZINC05682347	-7.2	CYS240 HIS280 (B)		

2.5.7.2 1IVO |246-253|W -ZINC ligands
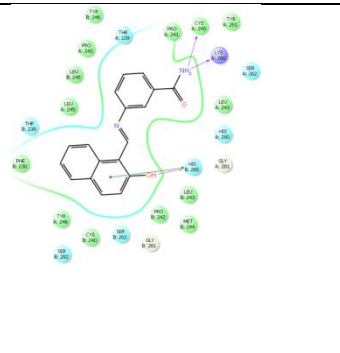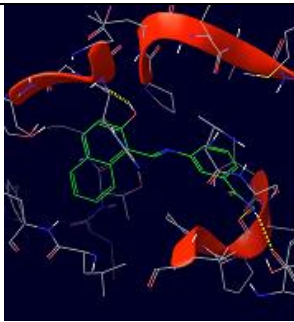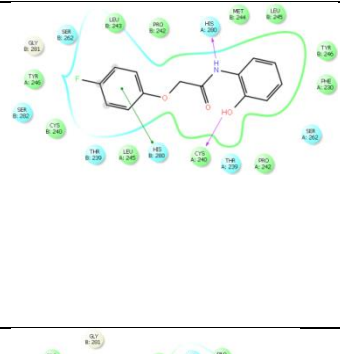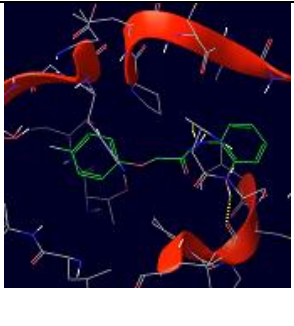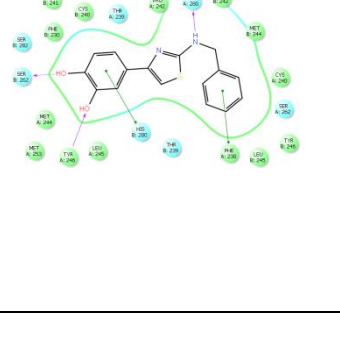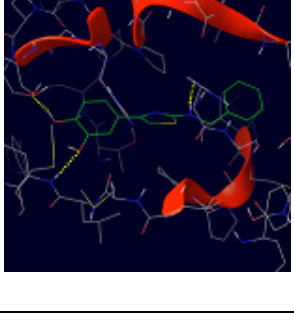
Table 16. Top 4 of docked |246-253|W-ZINC ligands.
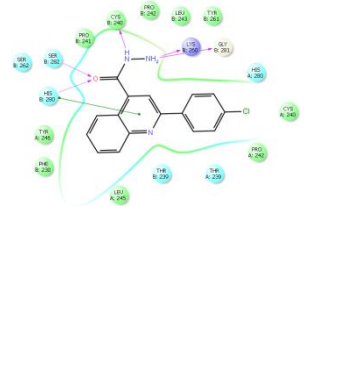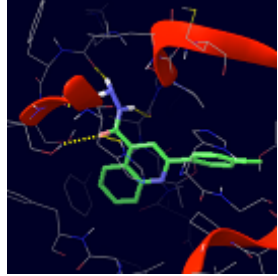
Gscores are 2-3x higher than those from the|246-253|S and their aa residues are

shown. There was no Gscores with the |246-253|S-ZINC ligands.

Ligand	Gscore	Interacting residue (s)	2D L-R complex	3D L-R complex
ZINC08441257	-6.12	HIS280 CYS240		
ZINC02062178	-5.98	HIS280 LEU243		
ZINC08441846	-5.03	HIS280		
ZINC20031600	-5.03	HIS280		

2.5.7.3 1IVO |86 230 263 264 265 275 278 283 285 286|W -Chembridge ligands

Table 17. |86 230 263 264 265 275 278 283 285 286|W-Chembridge complexes. Gscores and their interacting aa residues are shown. Monomeric grid was not docked.
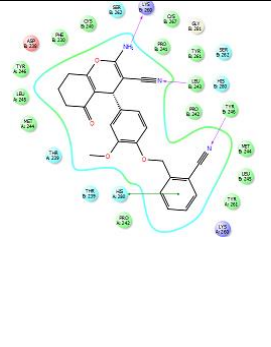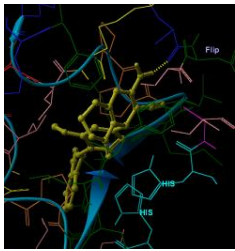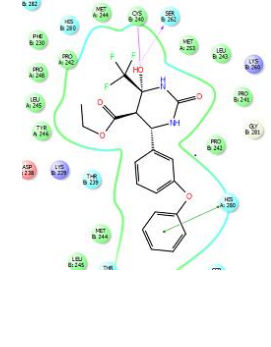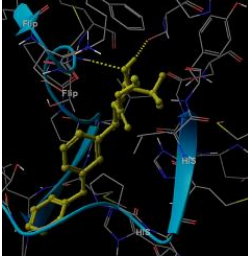
Ligand	Gscore	Interacting residues	2D L-R complex	3D L-R complex
ZINC00297549	-7.7	CYS240 LYS260 HIS280		
ZINC004784489	-7.3	CYS240 HIS280		
ZINC00784489	-7.3	TYR246 SER262 PHE230 HIS280		

ZINC00406587	-7.3	CYS240 LYS260 SER282 HIS280 GLY281		

2.5.7.4 1IVO |86 230 263 264 265 275 278 283 285 286|W-Ambinter ligands

Table 18. |86 230 263 264 265 275 278 283 285 286|W-Ambinter complexes.

Gscores and their interacting aa residues are shown. Monomeric arm was not docked.

Ligand	Gscore	Interacting residues	2D L-R complex	3D L-R complex
ZINC00661065	-8.6	TYR246 LYS260 LEU243		
ZINC08432896	-8.5	CYS240 SER262 HIS280 (π-cat)		

ZINC18153302	-7.1	CYS240 HIS280		
ZINC16972974	-6.9	CYS240		

These results are yet to be exploited due to time constraint.

2.5.8 Optimization of the lead hits (DPT1 1-DPT4 4)

The four lead candidates, DPT1 **1**-DPT4 **4**, having passed through the bioassay positively, were structurally modified using the v3.9CombiGlide (Schrödinger, LLC, New York, NY, 2015-4). Library design is one of the four methods used in *in silico* screening for drug design, besides the well-known structure-based, ligand alignment, pharmacophores, QSAR and the *de novo* approaches. Here, almost all were in place. By allowing flexible docking, each conformer for each candidate was generated and docked into the rigid active site of the receptor and binding affinity of each was assessed based on their final Gscore or DS,

which are invariably the same. Approximately 2.3×10^9 structurally modified compounds were generated from cores of each lead candidate, different fragment collections and with or without use of linkers, which are a collection of representative functional groups commonly found in pharmaceuticals and permits the user to effectively explore an active site and to identify types of functional groups that are likely to interact with the binding site in an energetically favorable manner, and only ≥100 had ≈ -7.0 Gscores. Five core structures were prepared from each lead candidate (DPT1 **1**-DPT4 **4**) preparatory for optimization with the fragment collection by the v3.9CombiGlide, which first enumerates the 2D core and fragment structures with or without linkers, transform the new 'synthesized' to 3D structures and dock them into the receptor grid of choice. The illustration in Table 19 is for DPT1 **1** only for space.

2.5.8.1 Model for optimization of lead candidates (DPT1 1-DPT4 4)

Figure 45 shows the model designed to highlight some key stages the v3.9CombiGlide utilizes to enumerate and dock. Addressing conformation at all stages of ligand sampling is core in Glide-docking. The Glide docking and scoring procedure permitted four approaches (already mentioned). Flexible ligand docking by default was used for the *in silico* screening throughout the SBDD, which allowed the conformer with optimum binding affinity (lowest energy) to dock and be reflected as a Gscore.
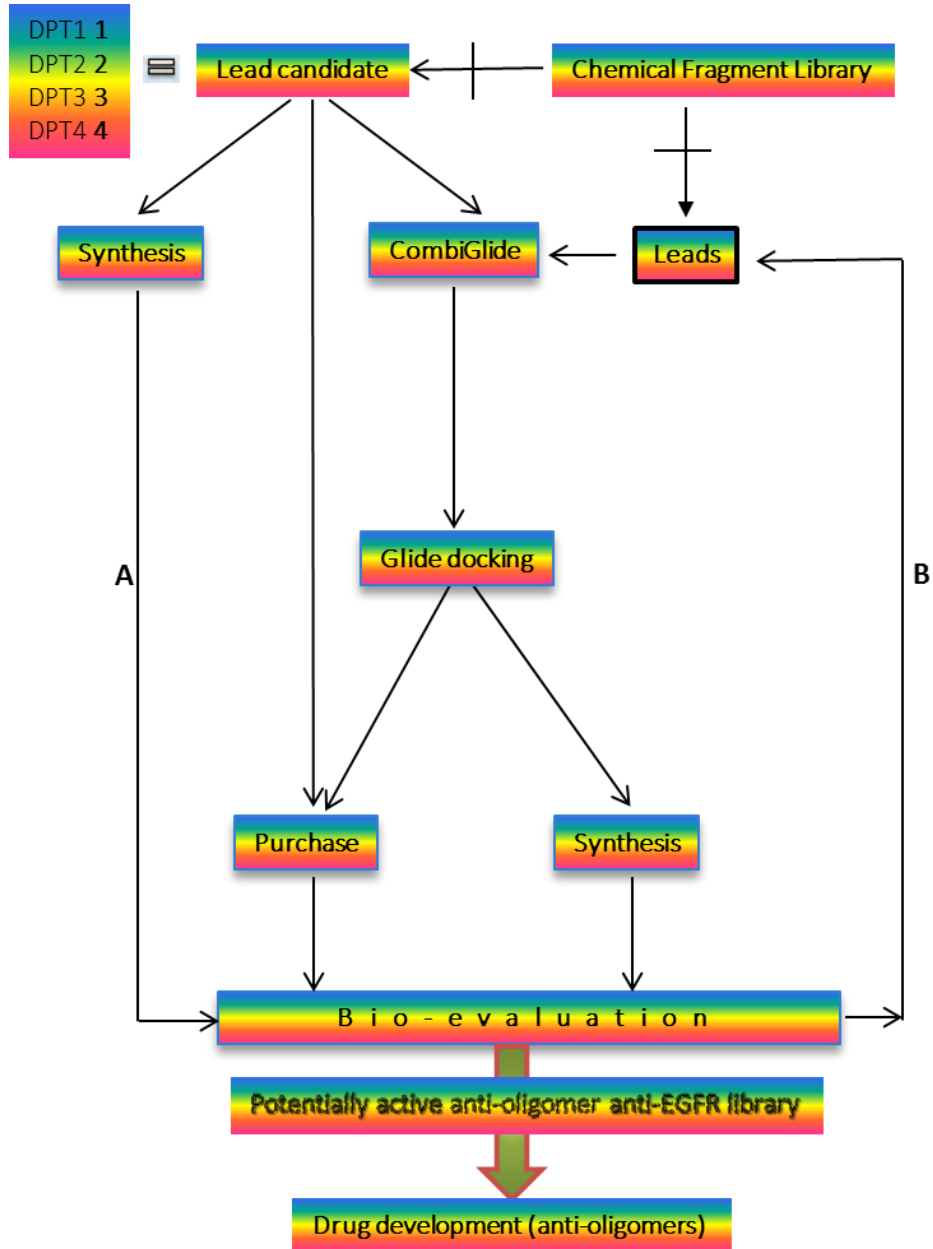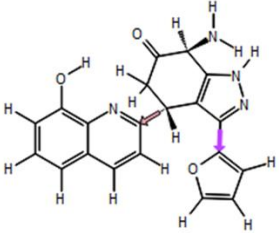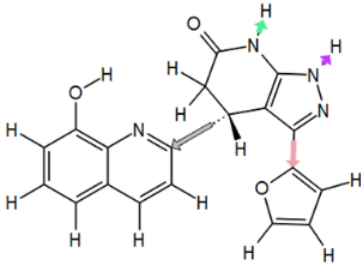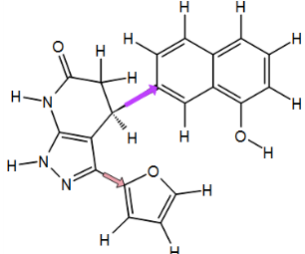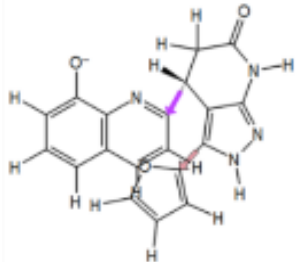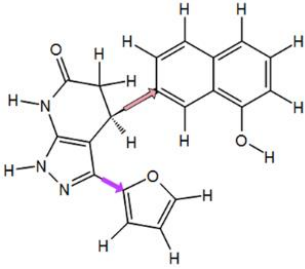
Figure 45. An Iterative model for optimization of DPT1 **1**-DPT4 **4**.

The v3.9CombiGlide first enumerates, synthesize using one or more linkers that bond the selected fragments from the fragment collection [indicated by the plus cross arrow] to the core structure. The enumerated is docked to the receptor grid hierarchically. The 'best' are purchased and/or synthesized for bioevaluation. The cycle **A→B** & **B→A** continues until a focused library of potentially active agents is generated for further drug development through the relevant transforming connections of the new modified structure (s), which are then purchased / synthesized→bioevaluation. Table 19 shows some of the core structures generated from DPT1 **1** only for space while Tables 20-23 & 28 show the top newly modified products of DPT4 **4** in the first cycle. However, zm1 and zm2 were found synthesizable (chapter three). The emerged leads can further be modified or made a fragment collection for a new or old core molecule in the second cycle to obtain more potent compounds predicted by their Gscores after bioevaluation. The Program is also suitable for a *de novo* structure-based drug design. Table 30 shows some of the synthesizable yields from the optimization. Modified source: Cavasotto et al., 2011.

Table 19. Optimization method for DPT1 **1**-DPT4 **4**: DPT1 **1** exemplified.

The arrows indicate the departing fragment (s) and where the new fragment (s) will be attached. The different colors define the positions of the outgoing and would-be fragment (s) on enumeration. The more the linkers the more compounds are enumerated and docked.

Cores generated for DPT1 1	The v3.9CombiGlide Information display
	Number of structures in combinatorial library: 43680 Combinatorial definition table <table><thead><tr><th></th><th>Attachment name</th><th>Number of structures</th><th>Minimum linkers</th><th>Maximum linkers</th><th>Collection file</th></tr></thead><tbody><tr><td>1</td><td>Pos-1</td><td>84</td><td>0</td><td>1</td><td>DPT111.bld</td></tr><tr><td>2</td><td>Pos-2</td><td>130</td><td>0</td><td>1</td><td>DPT121.bld</td></tr></tbody></table>
	Number of structures in combinatorial library: 23425600000000 Combinatorial definition table <table><thead><tr><th></th><th>Attachment name</th><th>Number of structures</th><th>Minimum linkers</th><th>Maximum linkers</th><th>Collection file</th></tr></thead><tbody><tr><td>1</td><td>Pos-2</td><td>200</td><td>0</td><td>10</td><td>DPT1_2014_Schro-Frags.bld</td></tr><tr><td>2</td><td>Pos-3</td><td>200</td><td>0</td><td>10</td><td>DPT1_2014_Schro-Frags.bld</td></tr><tr><td>3</td><td>Pos-4</td><td>200</td><td>0</td><td>10</td><td>DPT1_2014_Schro-Frags.bld</td></tr><tr><td>4</td><td>Pos-9</td><td>200</td><td>0</td><td>10</td><td>DPT1_2014_Schro-Frags.bld</td></tr></tbody></table>
	Number of structures in combinatorial library: 875 Combinatorial definition table <table><thead><tr><th></th><th>Attachment name</th><th>Number of structures</th><th>Minimum linkers</th><th>Maximum linkers</th><th>Collection file</th></tr></thead><tbody><tr><td>1</td><td>Pos-3</td><td>35</td><td>0</td><td>0</td><td>cdpt2-2603141111.bld</td></tr><tr><td>2</td><td>Pos-4</td><td>25</td><td>0</td><td>0</td><td>cdpt2-2603141112.bld</td></tr></tbody></table>
	Number of structures in combinatorial library: 800 Combinatorial definition table <table><thead><tr><th></th><th>Attachment name</th><th>Number of structures</th><th>Minimum linkers</th><th>Maximum linkers</th><th>Collection file</th></tr></thead><tbody><tr><td>1</td><td>Pos-5</td><td>10</td><td>0</td><td>1</td><td>dpt1a-20031411.bld</td></tr><tr><td>2</td><td>Pos-6</td><td>20</td><td>0</td><td>1</td><td>dpt1a-20031421.bld</td></tr></tbody></table>

	Number of structures in combinatorial library: 250

Combinatorial definition table
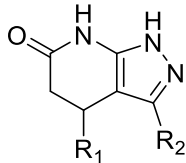
	Attachment name	Number of structures	Minimum linkers	Maximum linkers	Collection file
1	Pos-5	10	0	0	dpt1-03-03-1411.bld
2	Pos-6	25	0	0	dpt1-03-03-14211.bld

2.5.8.2 Optimized compounds from DPT1 1

DPT1 **1** (ZINC 65372018) had a Gscore of -2.7, but it has generated compounds with Gscores -7.0 following optimization, an indication binding affinity is 3x improved.

Table 20. Top 10 optimized 1IVO |246-253|S-DPT1 **1** with their DS & aa residues.

R$_1$ and R$_2$ are points of new attachments on the scaffold for the labeled moieties. The number of linkers used (if any), the Gscore, interacting residues, code name, and potential energy are reflected accordingly. The number of structures enumerated and docked by the v3.9CombiGlide is shown right to the scaffold.
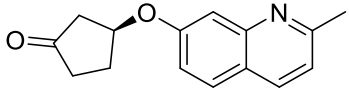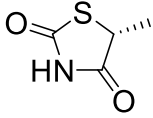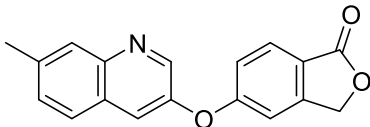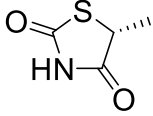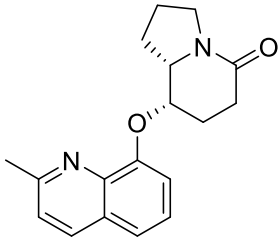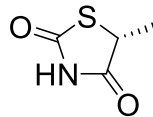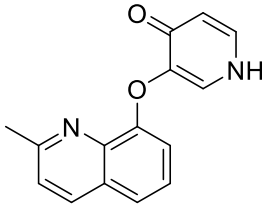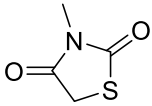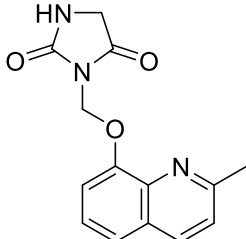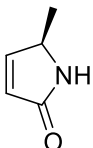
	Number of structures in combinatorial library: 43680

Combinatorial definition table

	Attachment name	Number of structures	Minimum linkers	Maximum linkers	Collection file
1	Pos-1	84	0	1	DPT111.bld
2	Pos-2	130	0	1	DPT121.bld

R$_1$	R$_2$	Code name	Linker (if any)	Gscore	Interacting residues	Potential Energy

		DPT1A	None	-7.34	CYS240 GLY264 HIS280 (π-Cat)	161.55
		DPT1B	None	-7.08	GLY264 CYS240	192.61
		DPT1C	None	-6.74	CYS240 GLY264	259.70
		DPT1D	None	-6.71	CYS283 CYS240 HIS (π-Cat)	169.74
		DPT1E	None	-6.6	CYS240 SER282 GLY264 PHE230 (π-Cat)	197.52

(structure)	(structure)	DPT1F	None	-6.58	CYS240 GLY264 HIS280 (π-Cat)	197.52
(structure)	(structure)	DPT1 G	None	-6.56	CYS240 GLY264	266.41
(structure)	(structure)	DPT1 H	None	-6.53	CYS240 GLY264 SER262 HIS280 (π-Cat)	127.29
(structure)	(structure)	DPT1I	None	-6.31	ASP238 CYS240 GLY264	227.65
(structure)	(structure)	DPT1J	None	-6.27	CYS240 GLY264	174.22

(a)

$$y = 5.04x + 228.05 \ (R^2 = 0.00)$$

(b)

$$y = 1.64x + -41.17 \ (R^2 = 0.01)$$

(c)
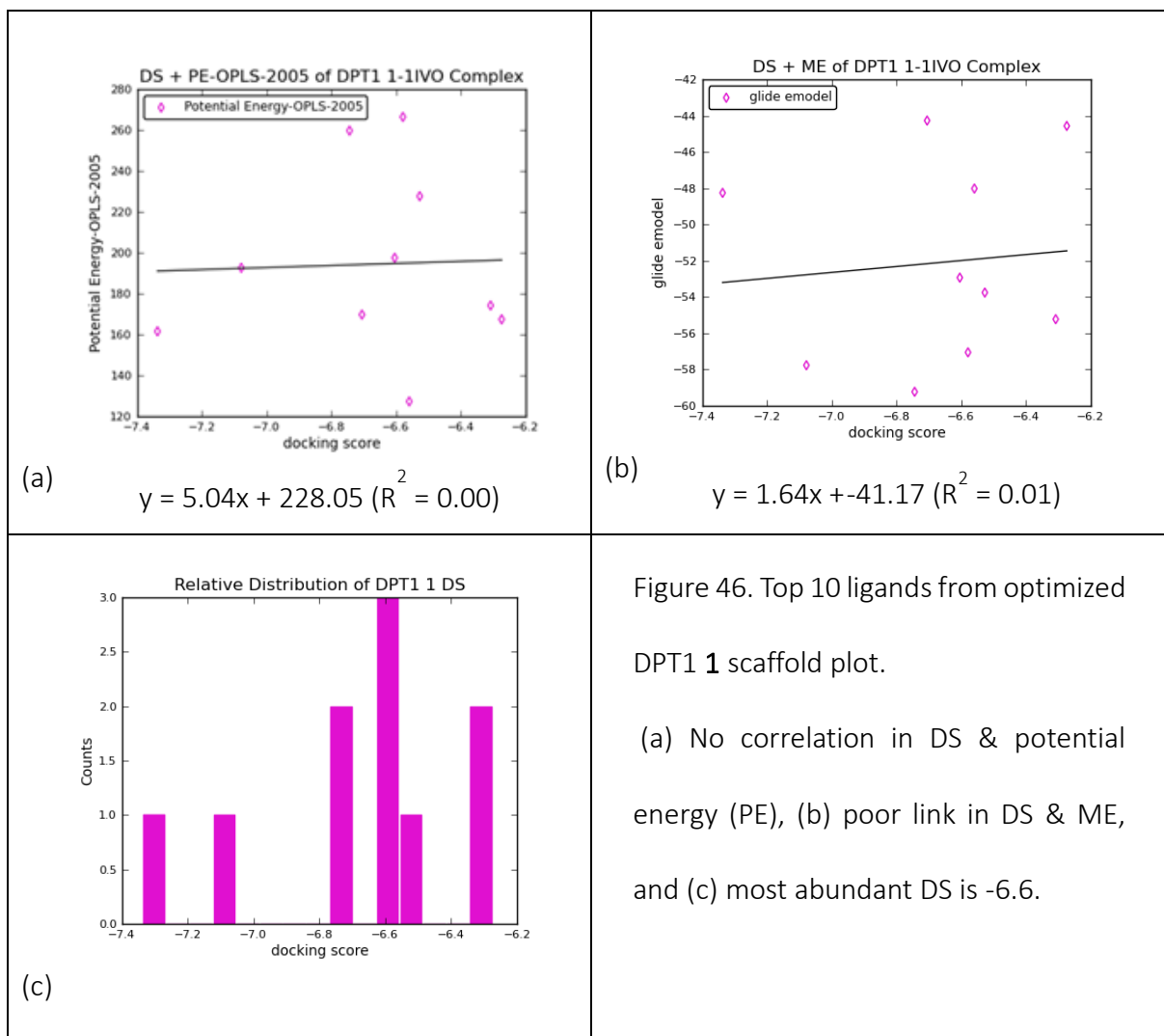
Figure 46. Top 10 ligands from optimized DPT1 **1** scaffold plot.
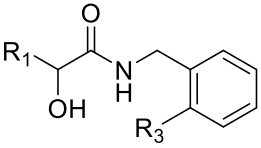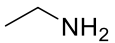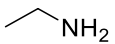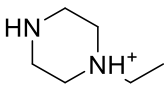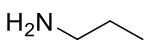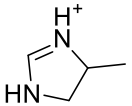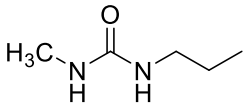
(a) No correlation in DS & potential energy (PE), (b) poor link in DS & ME, and (c) most abundant DS is -6.6.
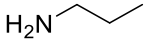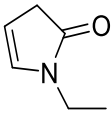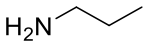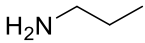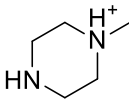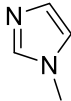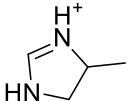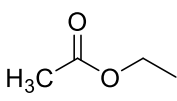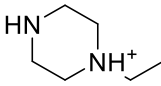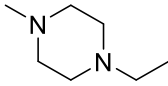
2.5.8.3 Optimized compounds from DPT2 2

DPT2 **2** (ZINC 67919965) had a Gscore of -1.9, but it has generated compounds with Gscores -7.0 following optimization, indication binding affinity is about 4x improved.

Table 21. Top 10 optimized 1IVO |246-253|S-DPT2 **2** with their DS & aa residues.

R$_1$ and R$_3$ are points of new attachments on the scaffold for the labeled moieties. The number

of linkers used (if any), the Gscore, interacting residues, code name, and potential energy are

reflected accordingly. The number of structures enumerated and docked by the v3.9CombiGlide

is shown right to the scaffold.

R$_1$	R$_3$	Code name	Gscore	Linkers (if any)	Interacting residue (s)	Potential energy
—SH	NH$_2$	DPT2A	-7.59	1	SER262 HIS280 SER282 LYS260 cys240	11.60
—NH$_3^+$	NH$_2$	DPT2B	-7.41	1	SER262 LYS260 TYR261 (π-Cat) GLY281 PRO241 HIS280	16.73
HN NH$^+$	H$_2$N	DPT2C	-7.13	2	SER282 SER262 CYS240 LYS260 HIS280 (π-Cat)	47.00
H$^+$ N HN	H$_3$C N H O N H	DPT2D	-6.97	2	CYS283 CYS240 HIS280 SER262	27.61

Scaffold label region:

Number of structures in combinatorial library: 600

Combinatorial definition table

	Attachment name	Number of structures	Minimum linkers	Maximum linkers	Collection file
1	Pos-1	10	0	2	dpt2-16031411.bld
2	Pos-3	10	0	1	dpt2-16031421.bld

					PHE230 (π-Cat)	
H₃C—	H₂N (chain)	DPT2E	-6.94	2	CYS240 SER262 LYS260 HIS280 HIS280 (π-Cat)	19.20
(N-ethyl dihydropyrrolone)	H₂N (chain)	DPT2F	-6.92	2	CYS240 SER262 LYS260	73.30
—SH	H₂N (chain)	DPT2G	-7.09	2	SER262 SER282 LYS260 CYS240 HIS280 (π-Cat)	10.33
(N-methyl piperazinium)	(N-methyl imidazole)	DPT2H	-6.91	None	SER262 CYS240 HIS280 GLY264 PHE230 (2 π-Cats)	104.94
(methyl imidazolinium)	H₃C (ethyl acetate)	DPT2I	-6.87	1	GLY264 SER262 CYS240 HIS280 PHE230 (π-Cat)	62.90
(N-ethyl piperazinium)	(N-methyl N'-ethyl piperazine)	DPT2J	-6.84	2	CYS240 SER262 SER282	158.17

(a)

$y = 110.31x + 832.65 (R^2 = 0.33)$

(b)

$y = -3.11x + -72.30 (R2 = 0.01)$
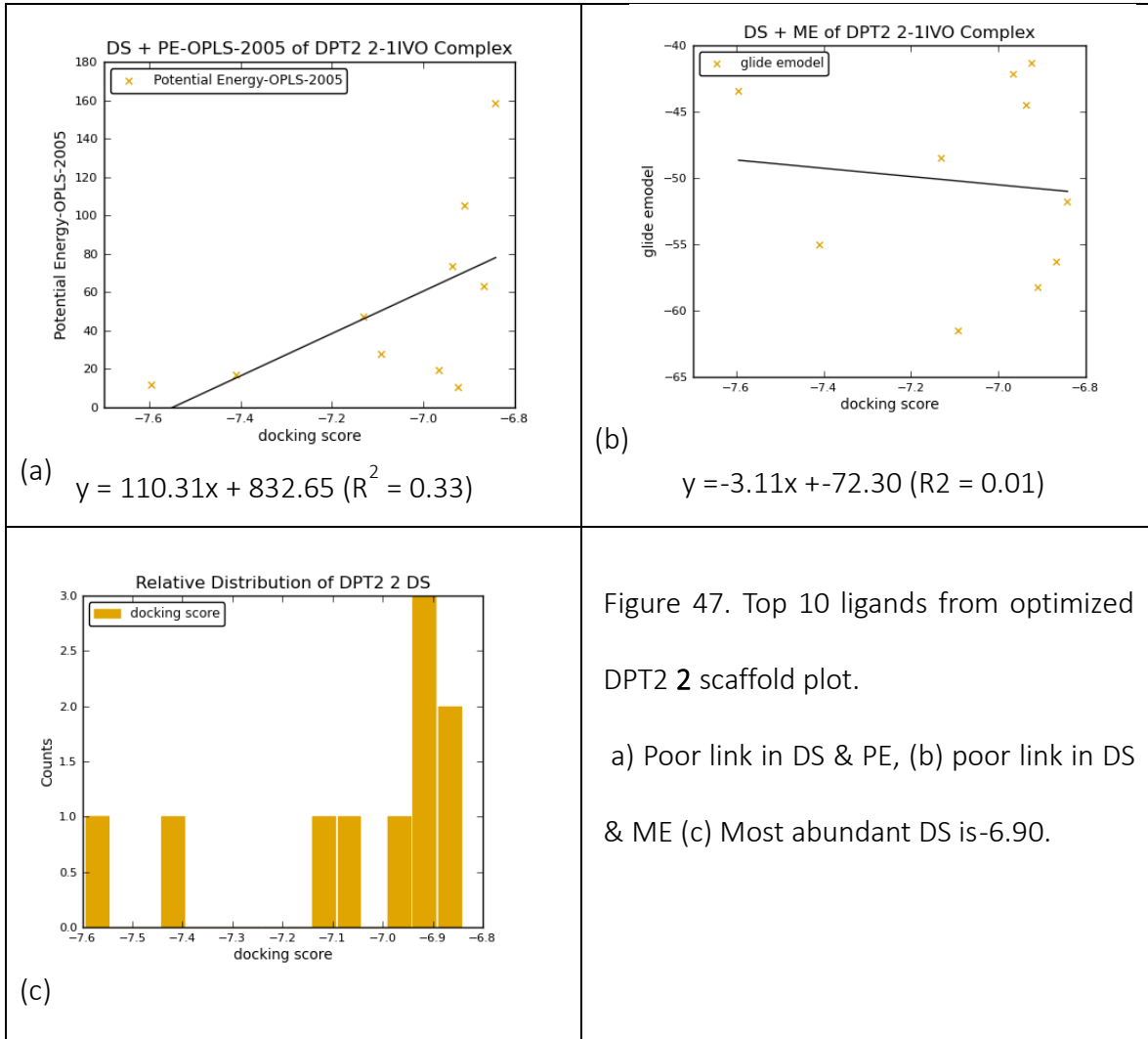
Figure 47. Top 10 ligands from optimized DPT2 **2** scaffold plot.
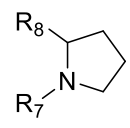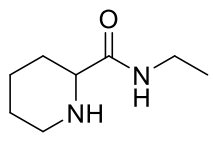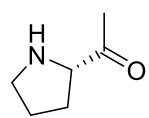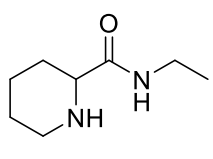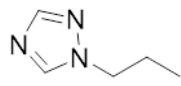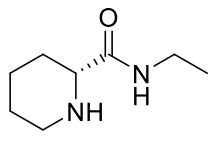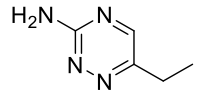
 a) Poor link in DS & PE, (b) poor link in DS & ME (c) Most abundant DS is -6.90.

(c)
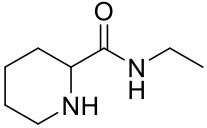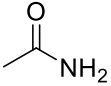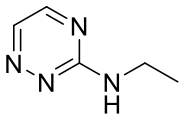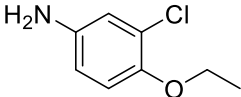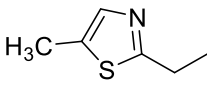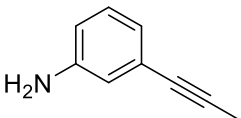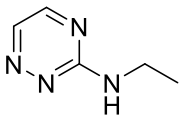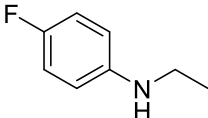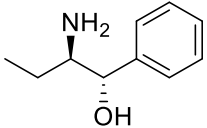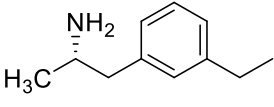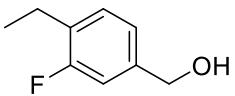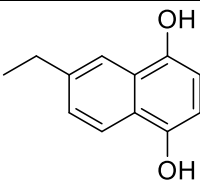
2.5.8.4 Optimized compounds from DPT3 3
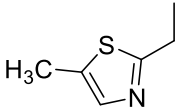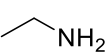
DPT3 **3** (ZINC 67851668) had a Gscore of -2.1, but it has generated compounds with Gscores -7.0 following optimization, an indication binding affinity is more than 3x improved.

Table 22. Top 10 optimized 1IVO |246-253|S-DPT3 **3** with their DS & aa residues.

R$_7$ and R$_8$ are points of new attachments on the scaffold for the labelled moieties. The number of linkers used (if any), the Gscore, interacting residues, code name, and potential energy are reflected accordingly. The number of structures enumerated and docked by the v3.9CombiGlide is shown right to the scaffold.

Number of structures in combinatorial library: 14400

Combinatorial definition table

	Attachment name	Number of structures	Minimum linkers	Maximum linkers	Collection file
1	Pos-7	60	0	1	DPT3_010714a_.bld
2	Pos-8	60	0	1	DPT3_010714a_.bld

R$_7$	R$_8$	Code name	GScore	Linkers (if any)	Interacting residues	Potential energy
		DPT3A	-8.5	1	TYR246 CYS240 GLY281	246.25
		DPT3B	-8.07	1	TYR246 SER282 HIS280	191.48
		DPT3C	-8.02	1	TYR246 LYS260 HIS280	191.11

				1	TYR246	
		DPT3D	-8.01		CYS240	150.07
		DPT3E	-8	1	TYR246 HIS280	168.31
		DPT3F	-7.99	1	GLY281 CYS240 HIS280 HIS280 (π-Cat) PHE230 (π-Cat)	125.16
		DPT3G	-7.82	1	TYR246	176.86
		DPT3H	-7.75	1	ASP279 CYS302 (A) CYS302 (B) THR278	313.10
		DPT3I	-7.69	1	GLN252 MET253	178.86

					ASP279	
					LYS303	
				1	CYS240	
H_3C [thiazole: S, N, ethyl]	[amine: NH_2]	DPT3J	-7.52		GLY281	174.89
					PRO241	
					LYS260	

(a)

y =-18.63x + 43.67 (R2 = 0.01)

(b)

y = 17.08x + 88.13 (R2 = 0.15)

(c)

Figure 48. Top 10 ligands from optimized DPT3 **3** scaffold plot.

a) Poor Link in DS & PE, (poor link in DS & ME, and (c) most abundant DS is -8.0.
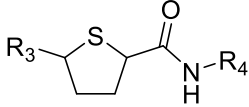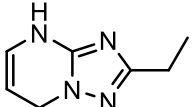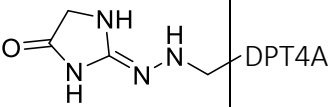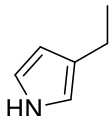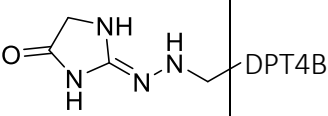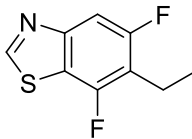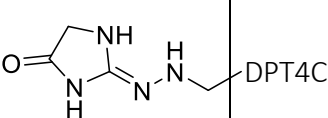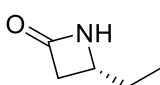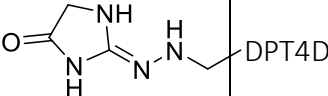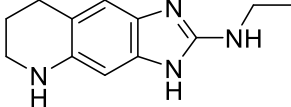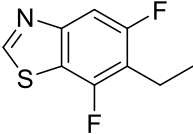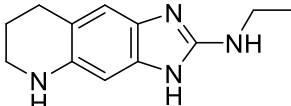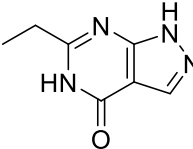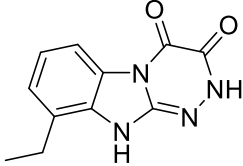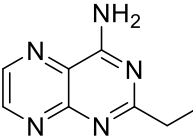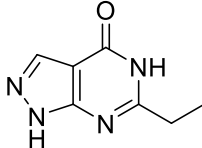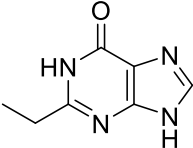
2.5.8.5 Optimized compounds from DPT4 4
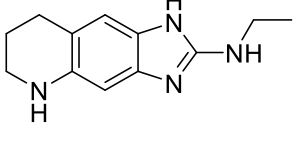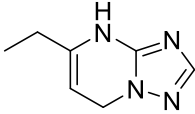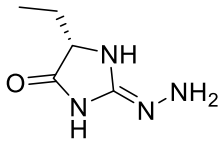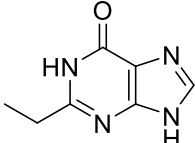
DPT4 **4** (ZINC 20537922) had a Gscore -3.5, but it has generated compounds with Gscores -8.0 following optimization, an indication binding affinity is about 3x improved.

Table 23. Top 10 optimized 1IVO |246-253|S-DPT4 **4** with their DS & aa residues.

R_3 and R_4 are points of new attachments on the scaffold for the labeled moieties. The number of linkers used (if any), the Gscore, interacting residues, code name, and potential energy are reflected accordingly. The number of structures enumerated and docked by the v3.9CombiGlide is shown right to the scaffold.

R₃ structure	Combinatorial library info
	Number of structures in combinatorial library: 14884 Combinatorial definition table | | Attachment name | Number of structures | Minimum linkers | Maximum linkers | Collection file | | 1 | Pos-3 | 61 | 0 | 1 | DPT4_300614.bld | | 2 | Pos-4 | 61 | 0 | 1 | DPT4_300614.bld |

R_3	R_4	Code name	Gscore	Linkers (if any)	Interacting residues	Potential energy-OPLS-2005
		DPT4A	-10.4	1	ASP238 CYS240 SER282 TYR246	100.45
		DPT4B	-9.68	1	ASP238 CYS240 SER282 TYR246	76.98
		DPT4C	-9.52	1	ASP238 CYS240 TYR246 SER282	83.40
		DPT4D	-9.34	1	ASP238 CYS240	77.49

					HIS280	
					SER282	
					TYR246	
		DPT4E	-9.11	1	TYR246 CYS240 ASP238 LYS229 (π-Cat)	57.20
		DPT4F	-8.95	1	Tyr246 SER262	113.51
		DPT4G	-8.92	1	TYR246 MET244 CYS240 LYS229 (π-Cat)	282.38
		DPT4H	-8.92	1	VAL255 TYR246 MET244 CYS240	172.03

		DPT4I	-8.88	1		64.35
		DPT4J	-8.88	1		154.09

(a)

$$y = 48.56x + 567.81 \ (R^2 = 0.12)$$

(b)

$$y = -4.48x + -102.47 \ (R^2 = 0.09)$$

(c)
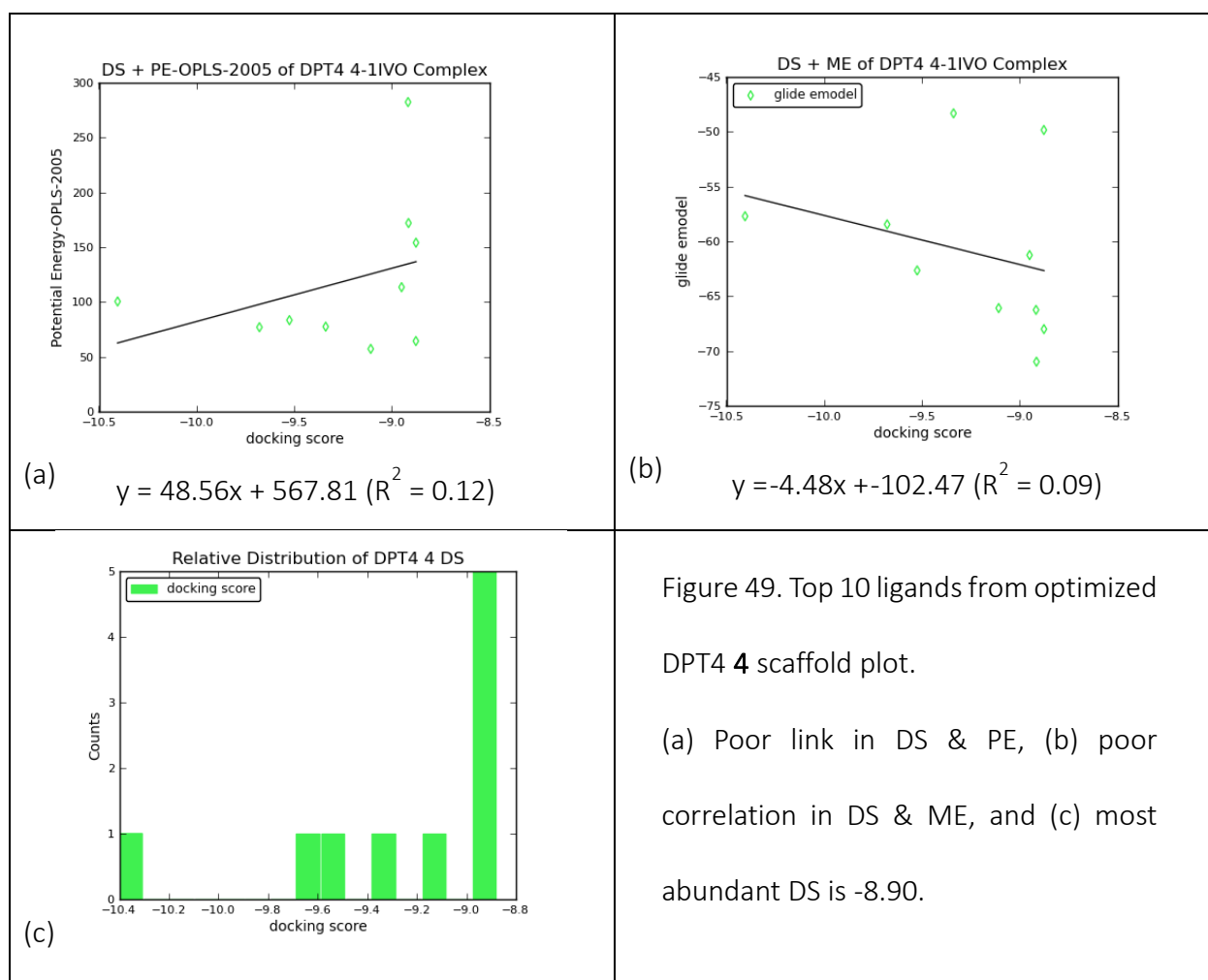
Figure 49. Top 10 ligands from optimized DPT4 **4** scaffold plot.

(a) Poor link in DS & PE, (b) poor correlation in DS & ME, and (c) most abundant DS is -8.90.

These structurally optimized compounds await more analyses as most of them were neither purchasable nor synthesizable. Therefore a complete approach to their chemical and biological

handling is required. However, a core of DPT4 **4** was modified using Cayman chemical fragments to obtain two synthesizable compounds zm1 **7** & zm2 **8** (Table 30).

2.5.8.6 Canvas studies of the optimized & docked DPT1 1 – DPT4 4 complexes

Lead optimization involved enumeration and docking of the selected potentially active compounds as described. The parameters involved in the docking contribute to the final 'best' binding affinity of the compound since binding energy is lowest. Using v2.6Canvas (Schrödinger, LLC, New York, NY, 2015-4), we can investigate more about these lead candidates on selected parameters of the docked 1IVO |246-253|**S**-Optimized DPT1 **1** – DPT4 **4** structures plus 1IVO |246-253|**S**-RTKI ligands for an observation. For each lead candidate, molecules with lowest and highest DS within the first top 10 were picked for their merited qualities and analyzed based on DS (x-axis), GLE (y-axis), ME (symbol size) and GE (color of symbol). Figures 46 - 49 show their Scatter Plot views with structures of the DPT$_H$'s lowest and highest DSs and other properties summarized in Tables 24 - 28, respectively. Data for the analyses were from Tables 20 – 23 & 4 in that order.

2.5.8.6.1 Top 10 1IVO |246-253|-DPT1 1 ligands
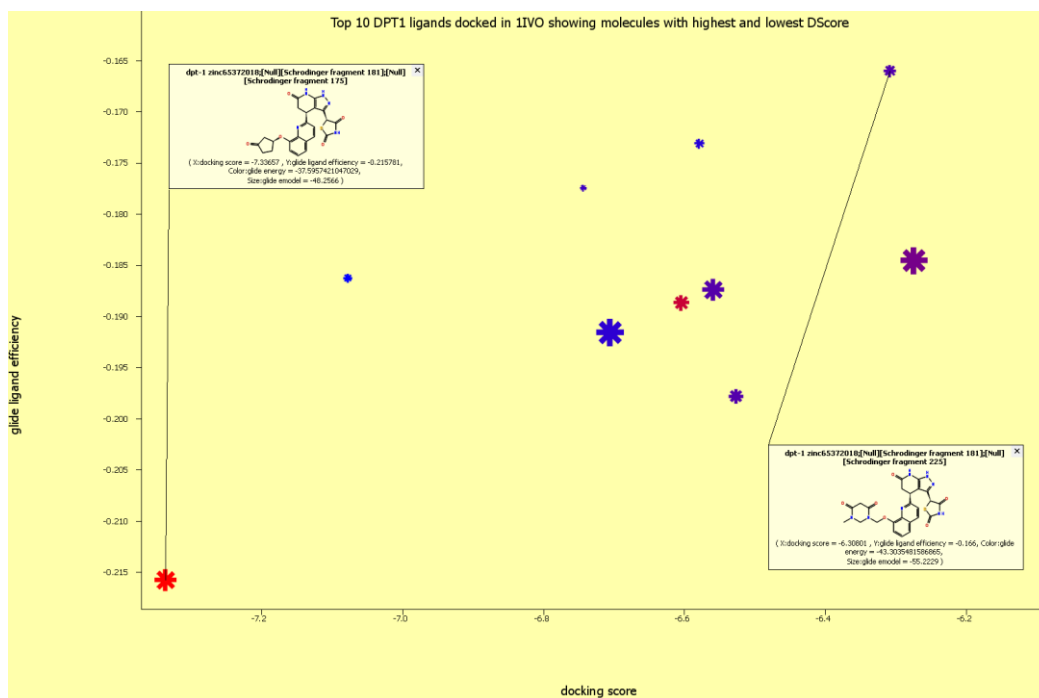
Figure 50. Top 10 optimized 1IVO |246-253|**S**-DPT1 **1** analysis. The highest and lowest DS are shown. GE (color of star) and ME (size of star). Details in are contained in Table 24. Source: Table 20.

Table 24. Detailed values for the selected properties of DPT1 **1** from Figure 50.

The observed values for the highest and lowest structures showing ME & GE within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-59.226	-44.271	-46.046	-37.600
DPT1A (DPT1-ZINC65372018)	-7.337	-0.216	-48.257		-37.596	
DPT1J (DPT1-ZINC65372018)	-6.308	-0.166	-55.223		-43.304	

2.5.8.6.2 Top 10 1IVO |246-253|-DPT2 2 ligands
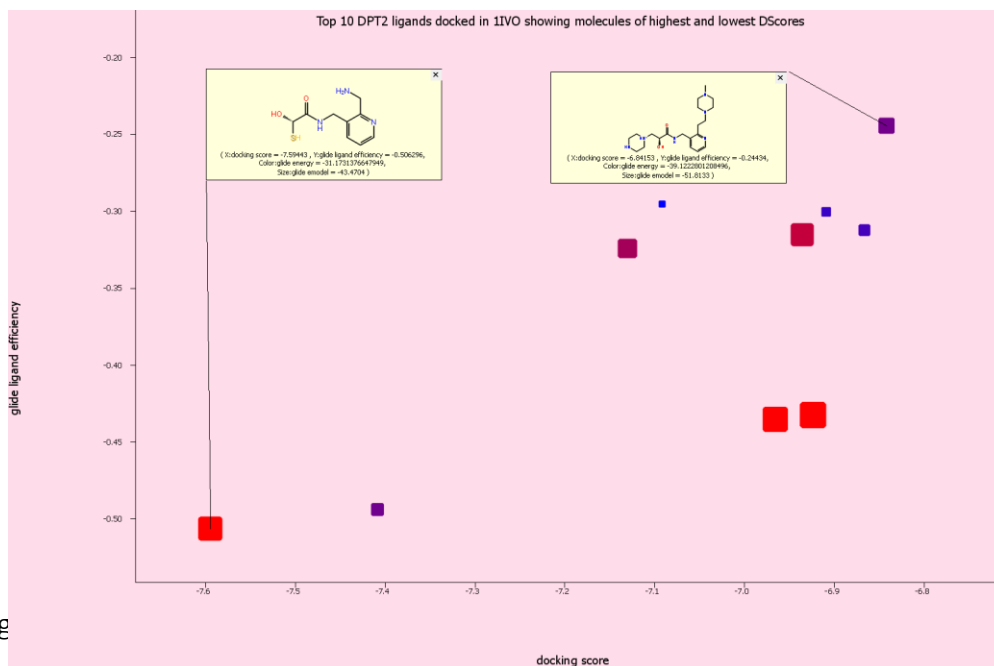
Fig

Structures with the highest and lowest DS are shown. GE (color of round rectangle) and ME (size of round rectangle). Details in are contained in Table 25. Source: Table 21.

Table 25. Detailed values for the selected properties of DPT2 **2** from Figure 51.

The observed values for the highest and lowest structures showing ME & GE within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-61.537	-41.357	-45.721	-31.173
DPT2A (DPT2-ZINC67919965)	-8.504	-O.387	-60.195		-42.483	
DPT2J (DPT2-ZINC67919965)	-7.523	-0.537	-25.924		-26.994	

2.5.8.6.3 Top 10 1IVO |246-253|-DPT3 3 ligands
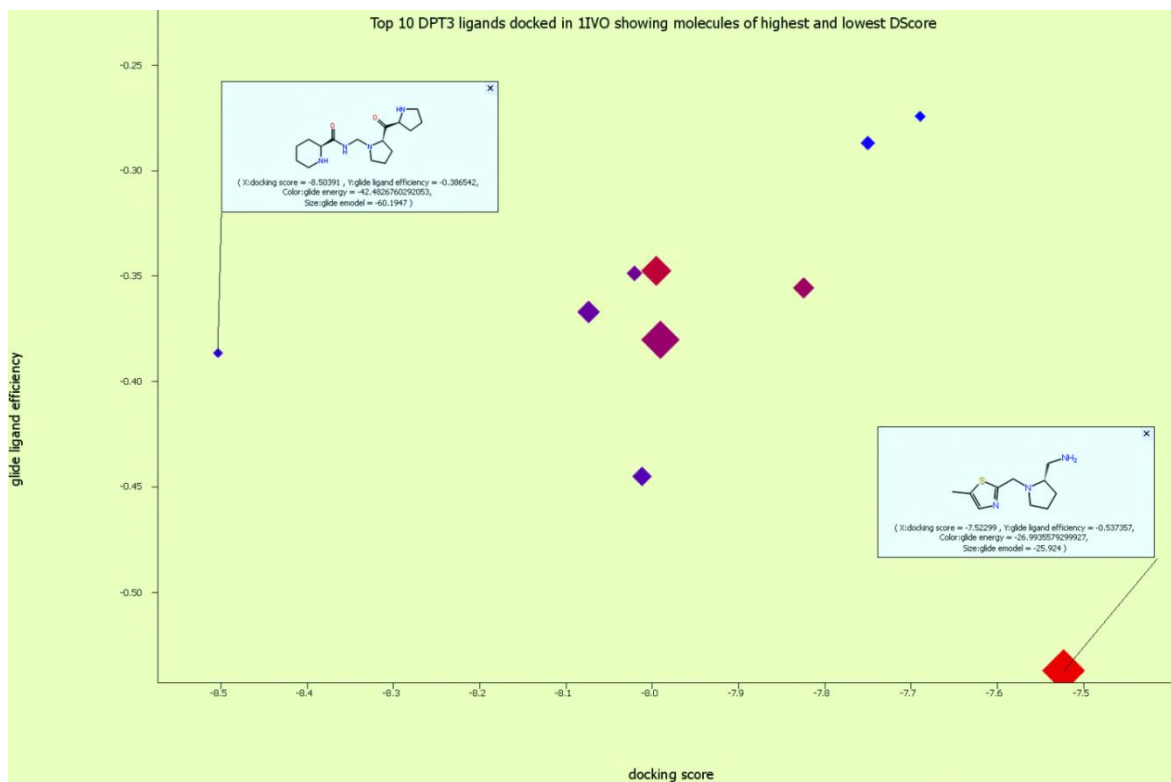
Figure 52. **Top 10 optimized 1IVO |246-253|S-DPT3 3 analysis.**

Structures with the highest and lowest DS are shown. GE color of diamond) and

ME (size of diamond). Details in are contained in Table 26. Source: Table 22.

Table 26. Detailed values for the selected properties of DPT3 **3** from Figure 52.

The observed values for the highest and lowest structures showing ME & GE within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-60.195	-25.924	-46.520	-26.994
DPT3A (DPT3-ZINC67851668)	-8.504	-0.387	-60.195		-42.483	
DPT3J (DPT2-ZINC67851668)	-7.523	-0.537	-25.924		-26.994	

2.5.8.6.4 Top 10 1IVO |246-253|-DPT4 4 ligands

Top 10 DPT4 ligands docked in 1IVO showing the molecules with highest and lowest DScores
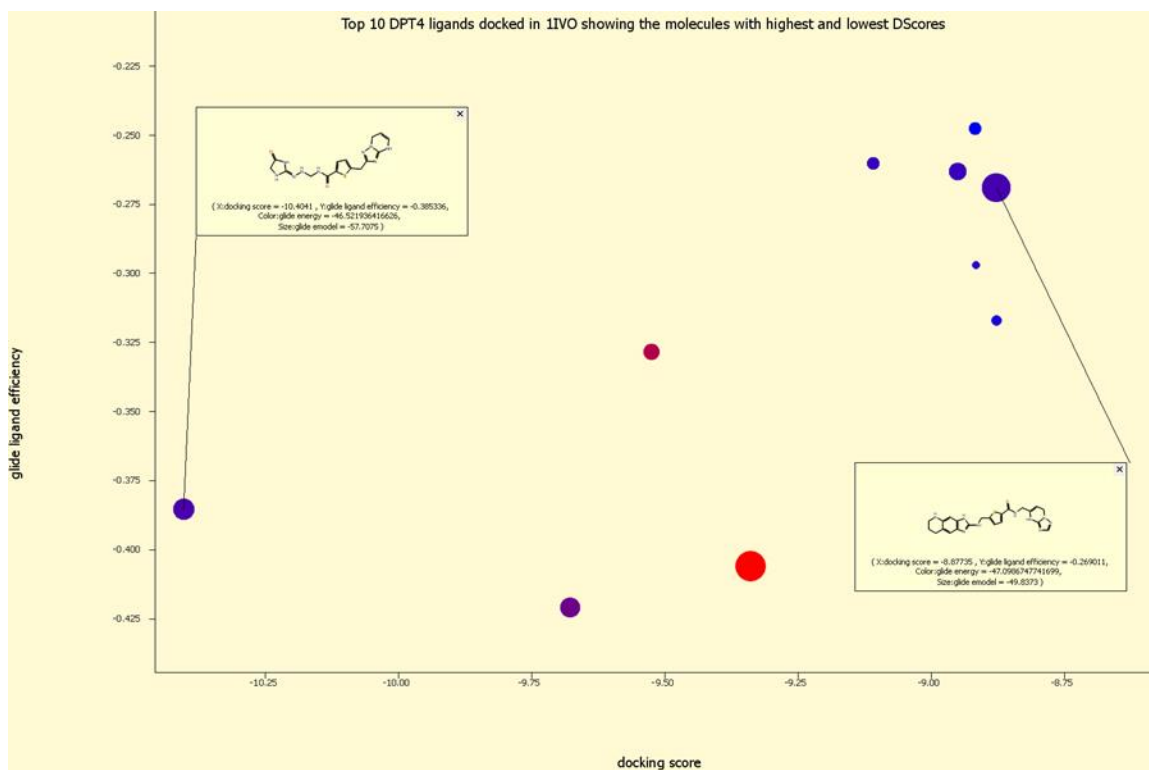
Figure 53. **Top 10 optimized 1IVO |246-253|S**-DPT4 **4** analysis.

Structures with the highest and lowest DS are shown. GE (color of circle) and ME

(size of circle). Details in are contained in Table 27. Source: Table 23.

Table 27. Detailed values for the selected properties of DPT4 **4** from Figure 53.

The observed values for the highest and lowest structures showing ME & GE within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-48.320	-70.980	-55.993	-23.260
DPT4A (DPT4-ZINC20537922)	-10.404	-0.385	-57.708		-46.522	
DPT4J (DPT4-ZINC20537922)	-8.877	-0.269	-49.837		-47.099	

2.5.8.6.5 Canvas analysis of 1IVO |246-253|S-Built RTKIs
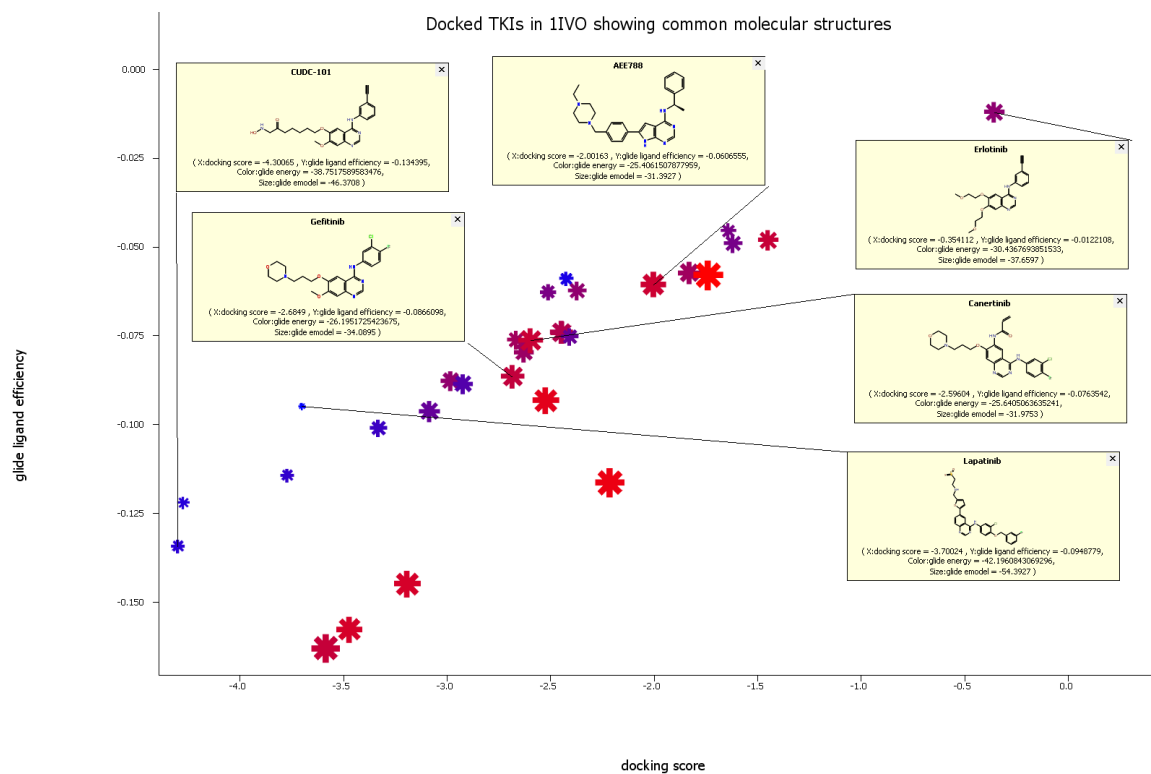
Figure 54. **Top 10 optimized 1IVO |246-253|S**-RTKI analysis.

Structures with the highest and lowest DS are shown. GE (color of star) and ME (size of star).

Table Details are expressed in Table 28. Source: Table 4.

Table 28. Detailed values for the selected properties of RTKIs from Figure 54.

The observed values for the highest and lowest structures showing ME & GE within range.

Ligand	Docking score (x-axis)	Glide ligand efficiency (y-axis)	Glide Emodel (size) Fixed = 16points		Glide Energy (color) Fixed = Dark blue	
			Minimum (12 points)	Maximum (48 points)	Minimum (blue)	Maximum (Red)
			-54.393	-25.487	-42.196	-21.224
CUDC-101	-4.301	-0.134	-46.371		-38.752	
AEE788	-2.002	-0.061	-31.393		-25.406	
Gefitinib	-2.685	-0.087	-34.090		-26.195	
Erlotinib	-0.354	-0.012	-37.660		-30.437	
Canertinib	-2.600	-0.076	-31.975		-25.641	
Lapatinib	-3.700	-0.095	-54.393		-42.196	

2.6 Some observed challenges in molecular docking processes

The design of the active constituents against specific targets has its challenges, some which are i) handling the unexpected results such as no scores as observed in 1IVO |246-253|S docking, ii) unsure of the limits of your docking software, iii) managing false positives and false fails, iv) validation of unliganded target biomolecules, especially when it requires developing such, v) failed structural enumerations and subsequent docking as noticed

during the combinatorial synthesis using CombiGlide, vi) enumerated and docked 'compound forests' that yielded no scores, vii) enumerated and docked molecules that yielded very low docking scores defiant to expectations, viii) optimized products major microspecies at pH 7.4 exist as ionic as obtained with DPT2 **2** in Table 21, ix) delayed deliveries from chemical vendors on order for those purchasable, x) enumerated and docked compounds that yielded good glide scores but hard to synthesize and / or purchase, and xi) anxiety if high scored ligands synthesized and / or purchased 'hits' to translate into 'lead' candidates following bio-evaluation fail. The whole process of the modeling is punctuated with real uncertainties, therefore to finally emerge with results is encouraging amidst the staircases ascended [174].

Tables 29 and 30 contained purchasable and synthesizable hits of coded DPT5 **5** and DPT6 **6** & optimized zm1 **7** and zm2 **8**, respectively.

Table 29. 2J6M- DPT5 **5** & DPT6 **6** showing their Gscores and aa residues.
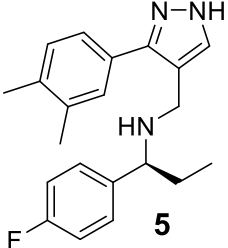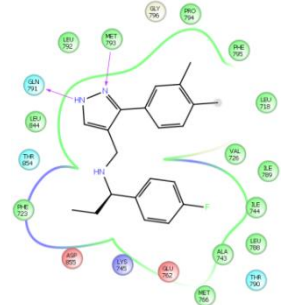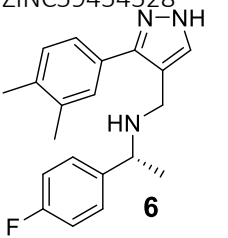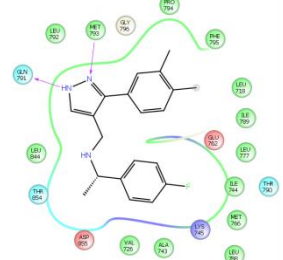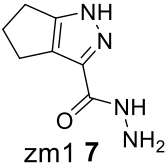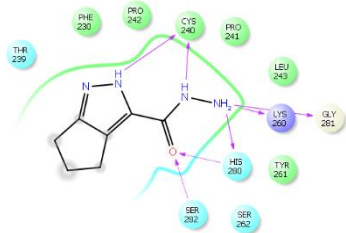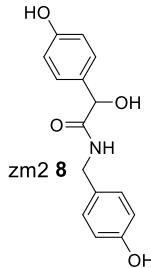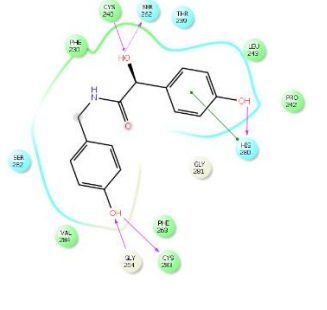
Ligand	Gscore	Linked aa residues	Code name	2D L-R Complex
ZINC56992273 **5**	-10.6	GLN791, MET793	DPT5 **5**	
ZINC59454528 **6**	-10.3	GLN791, MET793	DPT6 **6**	

Table 30. 1IVO |200-300|W-zm1 **7** & zm2 **8** showing their DS and aa residues.

Ligand	Gscore	Linked aa residues	2D L-R complex
Zm1 **7** zm1 **7**	-8.0	GLY281, LYS260, HIS280, SER282	
Zm2 **8** zm2 **8**	-9.2	CYS240, HIS280, CYS283, GLY264, SER264	

2.7 The docking computational details

The 2-D ligand structures of the known RTKIs were built using v10.4MaestroGUI, since it provided many viewing options to accommodate the varied needs of different applications. Thus, it gave the small molecules and large biomolecular complexes a wide range of clarity of modeled systems with full-featured 3D visualization and enhanced 2D ligand interaction diagrams. V3.6LigPrep generated accurate, energy minimized their 3D molecular structures on conversion from the 2D, and corrected the Lewis structures to eliminate mistakes in the 31 ligands in order to reduce downstream computational errors. The geometry minimizations of all the ligands constructed was by the OPLS_2005 force field. Optimizations were converged to a gradient RMSD at 0.05kj/mol with a maximum iterations of 2500 with an extended cutoff thus with minimal changes in RMSD gradients. These compounds were then docked into the grids generated from the prepared 1IVO and 2J6M with the Prep Wizard (v3.4Epik, v4.2Prime & v6.9Impact) using v6.9Glide hierarchically through a number of Lipinski's Ro5 embedded and Lipinski-like filters and / or on choice. The shape and properties of the binding site were characterized using some selected docking out properties of the top 4, 10 or 20 as the case warranted. An accomplished docking would be reflected in the glide ligand efficiency, emodel that combined the energy grid score, the binding affinity predicted by Gscore, and (for flexible docking) the internal strain energy for the model potential used to direct the conformational-search algorithm on one hand while the glide energy efficiency related to the glide score and the heavy atoms of the protein. The application of the v2.6Canvas
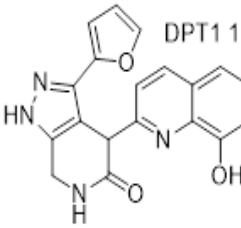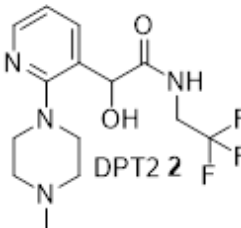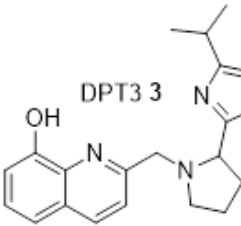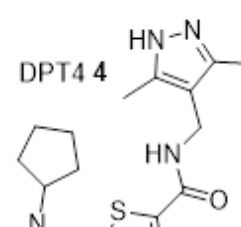
established the correlation of these properties. The v3.9CombiGlide enumerated as well as well docked newly generated compounds and their shape and properties of the binding site were characterized using the selected docking out properties for further characterization.
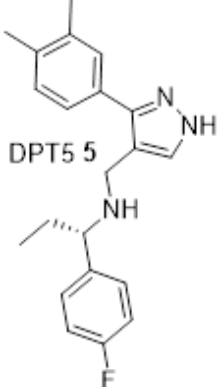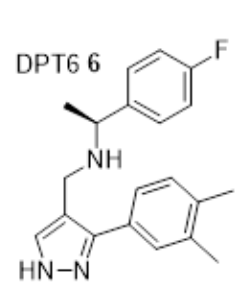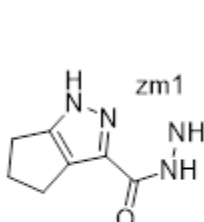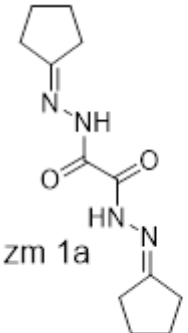
2.8 Drug-likeness of DPT1 1-DPT6 6 and zm1 7, zm1a 7 and zm2 8
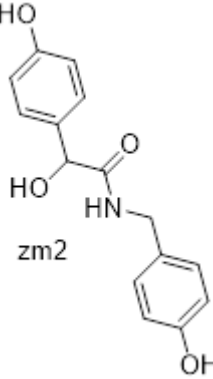
Sections 2.5.2, 2.5.6, 2.5.7, 2.5.8 and 2.9 reflect properties of DPT1 **1**-DPT4 **4** as our major focus as explained. The examination of the drug-likeness of these selected hits is assertive. Drug-likeness is a complex balance of various molecular properties and structural features, requiring that a molecule intended for therapy should resemble known drugs, where they exist. Properties common to drugs include hydrophobicity, electronic distribution, hydrogen bonding (Hb) characteristics, pharmacophoric features, and molecule size and flexibility are commonly identified as influential factors of the behavior of the small molecule in a living organism. The ChemAxon (http://www.chemicalize.org/) and its EGFRpred Software (http://crdd.osdd.net/oscadd/egfrpred/draw.php) predicted the Lipinski-like filters [175] and the anti-EGFR activity [176] of DPT1 **1**-DPT6 **6**, zm1 **7**, zm1a **7** and zm2 **8**. The results of their predictions are shown in Table 31.

Table 31. Lipinski-like filter predictions: DPT1 **1-**DPT6 **6,** zm1 **7**, zm1a **7** and zm2 **8**.

ChemAxon Software was used for their Lipinski-like filters and anti-EGFR predictions. 'Yes' = requirements met; 'No' = requirements unmet. All abbreviations are fully clarified in Table 2.

Structure/properties of hit	Ro5	BF	GF	LLF	MF	VF	EGFR-inhibition
DPT1 1 logP: 3.05 logD: 2.36 HBD: 3 HBA: 5 RotC: 2 PSA: 104.04 MSA: 435.13 AC: 40 RC: 5 R: 93.56 MW: 346.34	Yes	Yes	Yes	No	Yes	Yes	1. Probability score = 3 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
DPT2 2 logP: 0.60 logD: 0.35 HBD: 2 HBA: 5 RotC: 5 PSA: 68.70 MSA: 465.84 AC: 42 RC: 2 R: 79.51 MW: 332.32	Yes	Yes	Yes	Yes	Yes	Yes	1. Probability score = 3 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
DPT3 3 logP: 3.90 logD: 3.18 HBD: 1 HBA: 6 RotC: 4 PSA: 75.28 MSA: 506.24 AC: 47 RC: 4 R: 95.65 MW: 338.40	Yes	Yes	Yes	Yes	Yes	Yes	1. Probability score = 3 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
DPT4 4 logP: 4.95 logD: 2.30 HBD: 2 HBA: 3 RotC: 5 PSA: 61.02 MSA: 575.51 AC: 54 RC: 4 R: 106.92 MW: 372.53	Yes	Yes	Yes	Yes	Yes	Yes	1. Probability score = 0 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model

Structure / Properties								Prediction
DPT5 5 logP: 5.70 logD: 3.80 HBD:2 HBA: 1 RotC: 6 PSA: 40.71 MSA: 584.43 AC: 49 RC: 6 R: 101.61 MW: 337.20	No	Yes	No	No	No	Yes		1. Probability score = 1 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
DPT6 6 logP: 5.63 logD: 3.41 HBD: 2 HBA: 1 RotC: 5 PSA: 40.71 MSA: 500.05 AC: 46 RC: 3 R: 97.09 MW: 323.41	N0	Yes	Yes	Yes	No	Yes		1. Probability score = 1 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
zm1 logP: -0.37 logD: 0.07 HBD: 4 HBA: 3 RotC: 1 PSA: 83.80 MSA: 232.82 AC: 22 RC: 2 R: 45.85 MW: 166.09	Yes	Yes	Yes	Yes	No	Yes		1. Probability score = 1 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model
zm 1a logP: 1.12 logD: 1.94 HBD: 2 HBA: 4 RotC: 3 PSA: 82.92 MSA: 378.59 AC: 51 RC: 3 R: 66.29 MW: 250.30	Yes	Yes	Yes	Yes	Yes	Yes		1. Probability score = 1 2. Prediction = Non-inhibitor 3. Prediction on EGFR-10 model

| | logP: 1.57
logD: 1.47
HBD: 4
HBA: 4
RotC: 4
PSA: 89.79
MSA: 380.27
AC: 35
RC:2
R: 45.85
MW: 273.29 | Yes | Yes | Yes | Yes | Yes | Yes | 1. Probability score = 1
2. Prediction = Non-inhibitor
3. Prediction on EGFR-10 model |

All the compounds meet the Ro5 requirements except DPT1 **1**, DPT5 **5**, DPT6 **6** & zm1 **7** which failed in one or two Lipinski-like filters as indicated.

<center>Chapter Three</center>

<center>Syntheses of zm1 7 (1,4,5,6-tetrahydrocyclopenta[*c*]pyrazole-3-carbohydrazide), zm1a

7 (N'^1,N'^2-dicyclopentylideneoxalohydrazide) and zm2 8 (2-hydroxy-*N*-(4-

hydroxybenzyl)-2-(4-hydroxyphenyl)acetamide)</center>
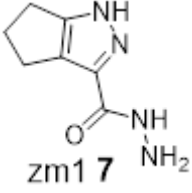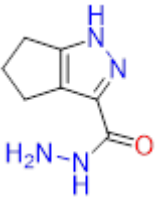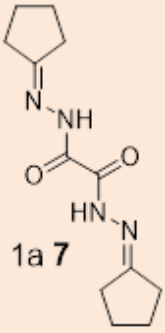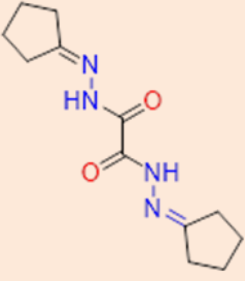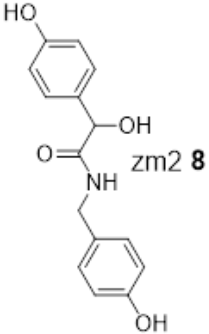
3.0 Introduction
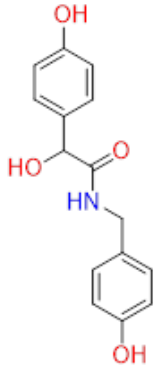
The structural optimization of a prepared DPT4 **4** core with Cayman fragment collection

without a linker produced two synthesizable compounds: zm1 **7** (Gscore -8.0), zm1a **7** and

zm2 **8** (Gscore -9.2) when docked into 1IVO |200-300|W that are synthesizable (Table 30).

However, in the process of synthesizing zm1 **7** compound zm1a **7** (N'^1,N'^2-

dicyclopentylideneoxalohydrazide) crystalized out and x-crystallographic information

obtained. The generation of these first two compounds against a validated targeted for

cancer therapy such as EGFR [177] is encouraging, despite chemical state. Figure 56 shows

their structures, however, Table 30 shows only zm1 **7** and zm2 **8** since these are optimized

docked compounds and their DS values. Table 31 shows the Lipinski-like filter properties

of these compounds, introducing them as having passed the drug-likeness and lead-

likeness based on the predictions by the ChemAxon Software [175]. The prediction of their

anti-EGFR activities by the Software in Table 31 and elaborated on in Table 32 is based on

statistical principles of random forest regression described by Singh and coworkers [176].

Singh and his group generated QSAR models using single scaffold-based analogues along

with data collected experimentally. The group asserted that a dataset of approximately

<center>154</center>

3500 diverse molecules was used to generate the QSAR-based prediction models. Random forest is machine-learning software that envisages the inhibition strength of a potential anti-EGFR candidate drug. Random forest was first developed by Breiman [178] with a wide application for various statistical purposes in cellular studies [179]. Tables 31 & 32 are based on this information, wherein both zm1 **7**, zm1a **7** and zm2 **8** are active inhibitors of EGFR with probability scores of 20, 28 and 23, respectively, on Prediction Model EGFR-1000. Figures 55 & 56 show the 3D structures of zm1 **7** and zm2 **8** and associated aa residue interactive properties.

Table 32. DS & full anti-EGFR features of zm1 **7**, zm1a **7** (in view) and zm2 **8**.

No docked or bioevaluated metrics for zm1a **7**. Zm1 **7** found IC_{50} was not established since its slope was negative when bioevaluated. Anti-EGFR predictions were done with ChemAxon Software based on statistical models (nM) of $IC_{50} \leq 10$, 100, and 1000. Their binding energy estimates (BEs) as defined by the molecular mechanics and implicit solvent energy were calculated using the prime energy for an optimized receptor-ligand complex. For zm1 **7** total tautomers is **7**, zm1a **7** 224, and zm2 **8** 4, however, only the dominant is shown.

Agent	Major tautomer shown	aa residues	Gscore	Anti-EGFR model[†]		
				10	100	1000
 zm1 **7** bee: -38.23; pred. IC$_{50}$: 4.6 μM; calc. IC$_{50}$: ?	Total: 7 	CYS240, MET244, LYS260, HIS280, GLY281, SER282	-8.0	1	5	20
 zm 1a **7**	 Total: 224	Yet to be docked		5	14	28
bee: - 28.13; zm2 **8**		CYS240, SER262, GLY264, HIS280, CYS283	-9.2	1	4	23

| predict. IC$_{50}$: 4.30 μM;

calc. IC$_{50}$: 0.63 μM |
Total: 4 | | | | | |

Figure 55. 3D structure of zm1 **7** in bonds with aa residues. Zm1 **7** (in ball & stick) showing its linked aa residues the distances (in purple dashed lines). The Hbs (yellow dashed lines) and the hydrophobic bonds (green dashed lines).
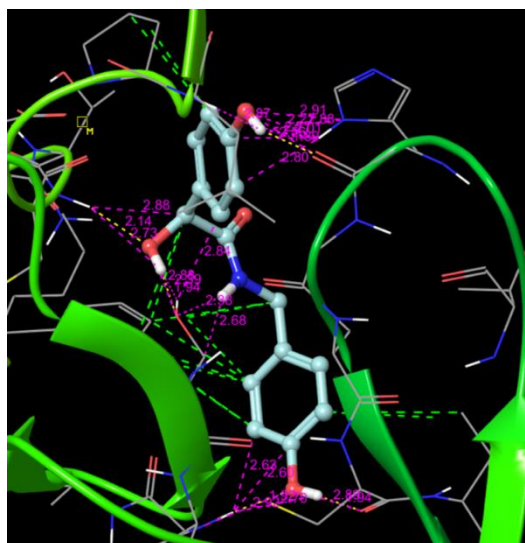
Figure 56. 3D structure of zm2 **8** in bonds with aa residues. Zm2 **8** (in ball & stick) and its linked aa residues show the distances (in purple dashed lines). The Hbs (yellow dashed lines) and the hydrophobic bonds (green dashed lines).

3.1 General instrumentation procedure for the syntheses of zm1 7 & zm2 8

The reagents were obtained from Sigma-Aldrich, Australia, and used as such without further treatment. The Buchi rotary evaporator was used to achieve an efficient and gentle (gradual) evaporation of solvents from samples performed under vacuum. The products were purified by recrystallization and / or column chromatography, which were all carried out using prepacked silica gel columns provided by Grace Reveleris[R] on the Grace Automated column using favorable separating systems earlier guided by thin layer chromatography (TLC). The TLC was performed on Merck aluminium plates, silica gel 60 F264. Components spotted on the TLC were visualized by observation under ultraviolet light (UV) 254 and/or 365 nm. Elemental analysis was carried out at Campbell Micro analytical Laboratory, Department of Chemistry, University of Otago, New Zealand. The IR was recorded on a Shimadzu FTIR–8400S apparatus. The melting points were determined on a Shimadzu SRS MPA100 OptiMelt Automated Melting Point apparatus. Proton and carbon nuclear magnetic resonance (1H and 13C NMR) spectra were recorded on a Varian 400MR using DMSO-d_6 as solvent. Chemical shifts were measured in parts per million (ppm) downfield from an internal tetramethylsilane (TMS) standard, and the residual solvent was used as a signal. Coupling constants are reported in Hz. NMR abbreviations used are s = singlet, d = doublet, dd = double doublet, q = quartet, m = multiplet, t = triplet and br = broad. The mass spectroscopy was performed on ESI Q1MS for low and high resolution.

3.1.2 Zm1 7 and [zm1a 7] syntheses

zm1 7

1,4,5,6-tetrahydrocyclopenta[*c*]
pyrazole-3-carbohydrazide

N'^1,N'^2-dicyclopentylideneoxalohydrazide
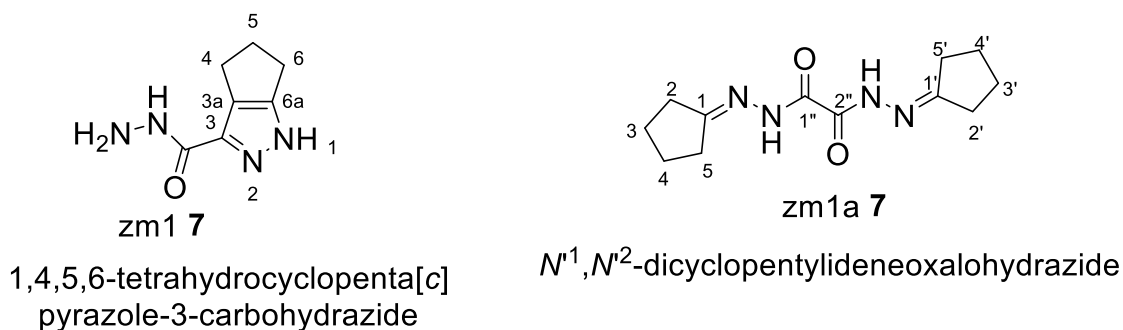
Figure 57. Structures of zm1 **7** and zm1a **7**.

Square brackets indicate an unintended product. Although the duo had the same starting materials their mechanisms are different. Zm1 **7** is a 5,5-fused pyrazole carbohydrazide derivative but zm1a **7** an aliphatic hydrazide. Reactions of carbonyl-containing compounds with hydrazides and semicarbazides have extensively been investigated via versatile synthetic methods and for specific use in structural elucidation by infrared (IR) spectroscopy [180]. An earlier synthetic approach for hydrazides was by Szmuszkovicz and Greig in 1961 formed hydrazone-hydrazide compounds such as aliphatic and aromatic hydrazides by reacting suitable oxalic acid derivatives with the hydrazine hydrate to yield oxalic acid *bis*hydrazide on hydrazinolysis [181]. Since then, similar reactions have appeared in literature such as described [182-184]. We adopt a modified method by Semple *et al.*, 2008 [185], to synthesize zm1 **7** and [zm1a **7**] which emerged as a crystal. An appropriate reagent to start with is cyclopentanone **9**, which was acylated with diethyl oxalate **10**, which on cyclization gives diketo ester **11**. This reacts with $H_2NNH_2.H_2O$ **12** yield the bicyclic pyrazole ester **13**. Excess $H_2NNH_2.H_2O$ yields both zm1 **7** and [zm1a **7**], howbeit, by different mechanisms (Scheme I).
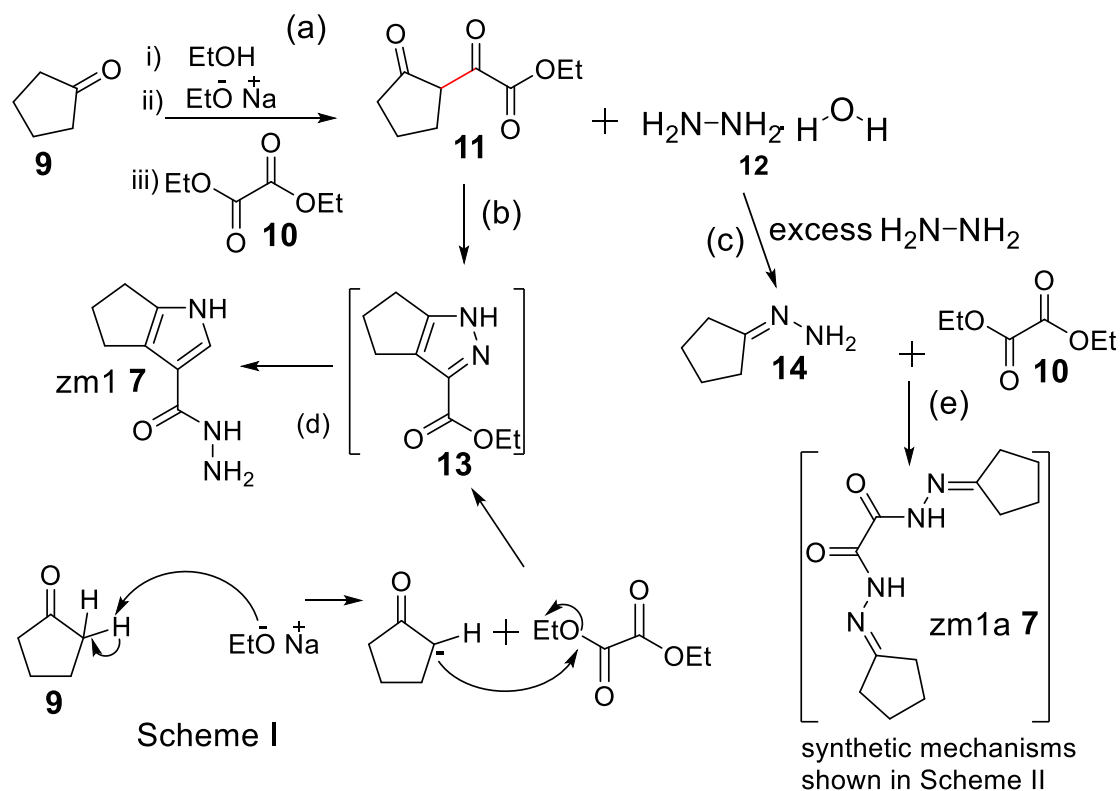
Figure 58. Scheme I: Syntheses of zm1 **7** and zm1a **7**.

The mechanisms for zm1 **7** and zm1a **7** synthesis are distinct due mainly to altered conditions. The conditions, reagents and/or process are (a) NaOET, EtOH, (EtOCO)$_2$, all mixed at room temperature (rt) before mixture was heated at 75 ^{0}C; (b) H$_2$NNH$_2$.H$_2$O, 75 ^{0}C, ≥ 2hr, cyclization; (c) excess H$_2$NNH$_2$.H$_2$O required; (d) Hydrazinolysis; (e) (EtOCO)$_2$, condensation. Zm1 **7** is a product of hydrazinolysis via the hydrazine hydrate process. First, EtO⁻ nucleophile extracts H from the active α-methylene C to the carbonyl ketone, the carbanion readily reacts with the carbonium generated by the leaving EtO⁻ , to form a new bond (in red colour) to get **11**. A two equivalent H$_2$NNH$_2$.H$_2$O is required to first form the transition product **13**, and subsequently zm1 **7**. Zm1a **7** followed a different path, as an excess of H$_2$NNH$_2$.H$_2$O was required to first form 14, and in condensation with (EtOCO)$_2$ to yield the product.
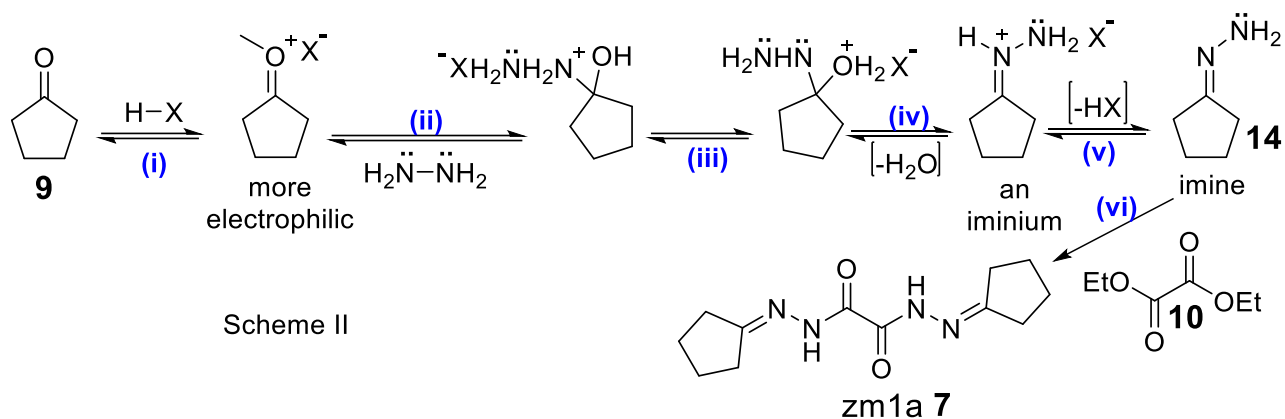
Figure 59. Scheme II: mechanisms for zm1a **7** synthesis over imine formation.

Summary: **(i)** Protonation of **9** by a H donor, **(ii)** **9** becomes more electrophilic for a [1,2] addition with the amine, hydrazine, **(iii)** proton transfer, **(iv)** a [1,2] elimination of [-H$_2$O] to form an iminium ion moiety, **(v)** deprotonation [-HX] with a proton acceptor to yield the imine **14**, **(vi)** a condensation reaction of **14** & **10** and removal of [-CH$_3$CH$_2$OH]$_2$ to produce zm1a **7**. Reaction starts with stirring the dissolved **9** with a proton-donating moiety H-X, e.g., some acid, which weakens the C=O, making the carbonyl carbon more electrophilic since it binds to the carbonyl oxygen. The H$_2$NNH$_2$, an amine & a nucleophile, then attacks the carbonyl electrophilic carbon whence a [1,2] addition occurs. The proton transfer is an intramolecular event that protonates the [OH] to give [H$_2$O$^+$], which is an excellent leaving group. This transfer also frees the lone pair on the amine nitrogen, which attacks the neighboring carbon, leading to eliminating the [H$_2$O] and iminium formation, considered the most positively charged organic molecules. A proton-accepting moiety, e.g., some base, deprotonates [-HX] the iminium to have **14**. Further condensation of this product in **10** yields zm1a **7**, which is a better alternative path to its synthesis.

3.1.3 Experimental details: syntheses of zm1 7 and [zm1a 7]

Square brackets indicate an unintended product. Cyclopentanone **9** (4 g, 47.55 mmol) was dissolved in ethanol (230 mL). Diethyl oxalate (7.75mL, 47.55 mmol) **10** was added with stirring, followed by the addition of sodium ethoxide (3.56 g, 47.55 mmol), at room temperature. The mixture was heated at 75 ^{0}C for 2hr and cooled to 4 ^{0}C in an ice bath, to give ethyl 2-oxo-2-(2-oxocyclopentyl) acetate **11,** as a pale yellow liquid product, which was neither isolated nor extracted. Aqueous hydrazine **12** (4.61 mL, 47.55 mmol) was added to **11** and heated at 75 ^{0}C for 2 hours in the ethanolic medium. Its cyclization yielded the bicyclic pyrazole ester, 1,4,5,6-tetrahydrocyclopenta[*c*]pyrazole-3-carboxylate **13**. The resulting aqueous crude product was extracted with dichloromethane (3 x 30 mL). The organic portion was dried over Na_2SO_4 and the DCM removed under reduced pressure. The brown fluffy solid obtained was purified by column chromatography on silica eluting with 50% EtOAc/n-hexane. Samples showing the same spot by TLC observed under UV light (254 nm) were pooled and evaporated under pressure and product recrystallized from EtOAc to obtain a white fluffy solid **7** of weight 208mg, 5.2%). 1HNMR data revealed that the isolated solid was a mixture of zm1 **7** and zm1a **7**; mp = 205-210 ^{0}C, no literature report on melting point; ^{1}H NMR (400 MHz, DMSO-d_6,) 1.62 - 1.76 (2H, m, CH_2, C-5), 2.16 - 2.25 (2H, m, CH_2, C-4), 2.37 (2H, dd, J=7.1 Hz, C-6); ^{13}CNMR (400 MHz, DMSO-d_6,) 24.6 (C-5), 28.8 (C-4), 33.58 (C-4), 156.2 (C-3a), 164.3 (C-3), 166.99 (C-6a), 174.21 (C-7); v_{max}, KBr/cm, 3264 (-N-H), 1683 (C=O), 1653 (-CONH-; [M + H]$^+$ = 167.09.

3.2 Zm2 8: synthetic approaches
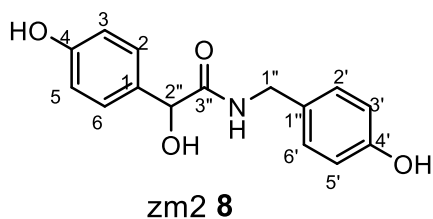
Figure 60. Structure of zm2 **8**.

zm2 **8**

2-hydroxy-*N*-(4-hydroxybenzyl)-2-(4-hydroxyphenyl)acetamide

Zm2 **8** is a new chemical entity that emerged when DPT4 **4** was optimized using Cayman fragment collection. Zm2 **8**, a hydroxylated benzoic acid amide derivative, was docked in 1IVO |200-300|**W** as a proceed of the optimized DPT4 **4** in search of anti-EGFR anti-oligomeric agent. The polyhydroxylated benzoic acid amide derivatives are known anticancer agents [186], bitter-masking compounds [187] as well as antioxidants and non-metal stabilizers essential in organisms, particularly, human skin [188]. Tables 29, 30 & 31 show its docked results, Lipinski filter-like properties and anti-EGFR inhibition prediction using ChemAxon Software, respectively. The synthesis of the agent will add to the new drug group with a defined mechanism against cancer when successfully achieved and bioevaluated.

3.2.1 Synthesis of zm2 8

The direct conversion of a carboxylic acid to an amide is difficult as amines are basic and tend to convert the carboxylic acids to their highly unreactive carboxylates. The addition

of the carboxylic acid to DCC (dicyclohexylcarbodiimide) **17** (EDC, (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide) can also be used) creates a good leaving group that can be displaced by an amine during nucleophilic substitution. Complete removal of moisturized air requires the dissolution of reagents in 1,4-dioxan under nitrogen. There are known synthetic approaches to obtain hydroxylated benzoic acid amide or phenolic amide derivatives [189], however, the Ley and Bertram method [186] is adopted with modifications [190, 191] for the zm2 **8** synthesis. Ley and coworker used the non-classical coupling of phenolic acids with acid chlorides and amines in basic mixtures since they observed the phenolic acids self-condense, to synthesize polyhydroxylated phenolic acid amides. The unprotected phenolic acid derivatives are generated using an activating agent such as DCC that induces coupling to form an unstable filterable precipitate amine-reactive O-acyl isourea intermediate from (2-(4-Hydroxyphenyl)acetic acid **15.** The addition of NHS stabilizes the amine-reactive intermediate in converting it to an amine-reactive NHS-ester, a semi-stable esterified 4-hydroxysuccinimide (NHS). This increases the efficiency of DCC-mediated coupling reactions to yield **8** when reacted with phenolic acid amine **19** through the formation of an amide bond (in red color in Scheme **III**). This non-classical coupling approach is shown in Scheme **III** in Figure 60. Scheme **IV** shows the synthesis of zm2 **8** in Figure 61.
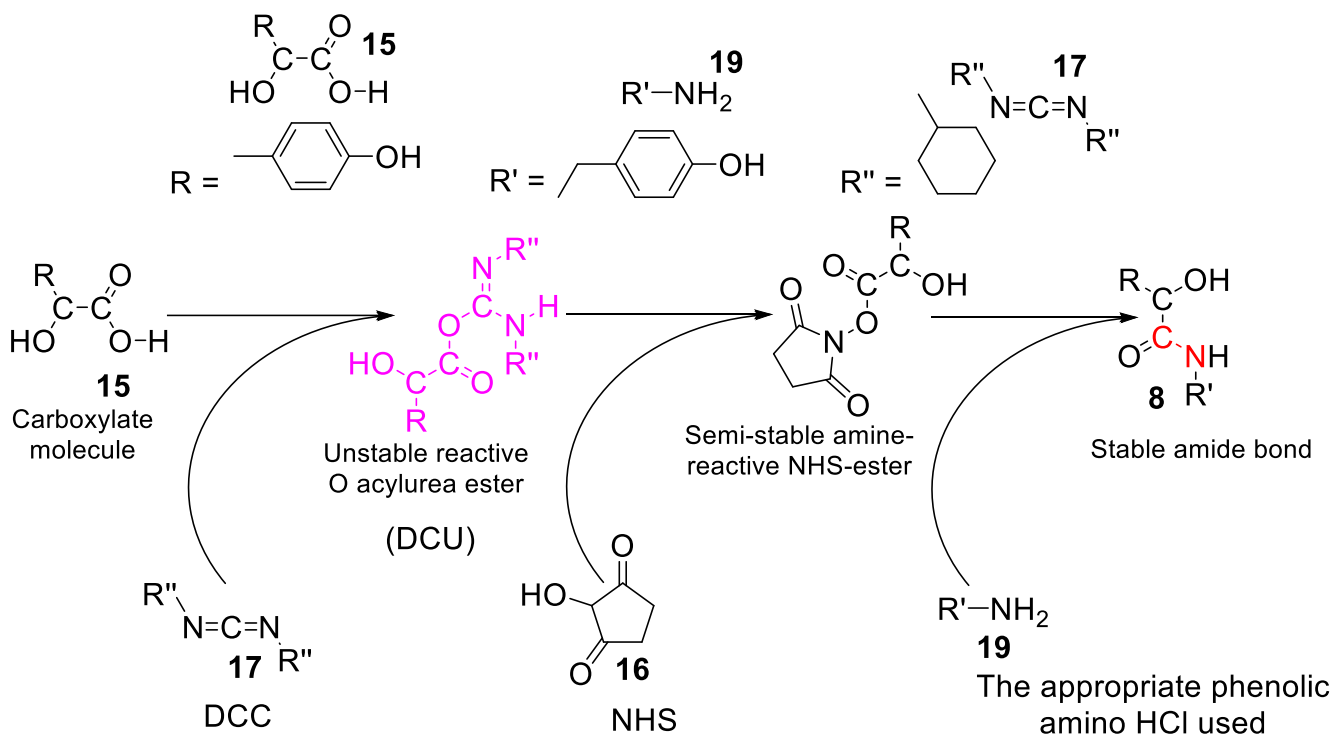
Figure 61. Scheme III: the non-classic coupling synthetic mechanism of zm2 **8**.

DCC–NHS chemistry: DCC reacts with a carboxylic-acid group to form an amine-reactive O-acyl isourea intermediate. This intermediate reacts with the amine to yield a conjugate of the two molecules joined by a stable amide bond. The addition of NHS stabilizes the amine-reactive intermediate in converting it to an amine-reactive NHS-ester, thus increasing the efficiency of DCC-mediated coupling reactions to yield **8** when reacted with phenolic acid amine **19.**
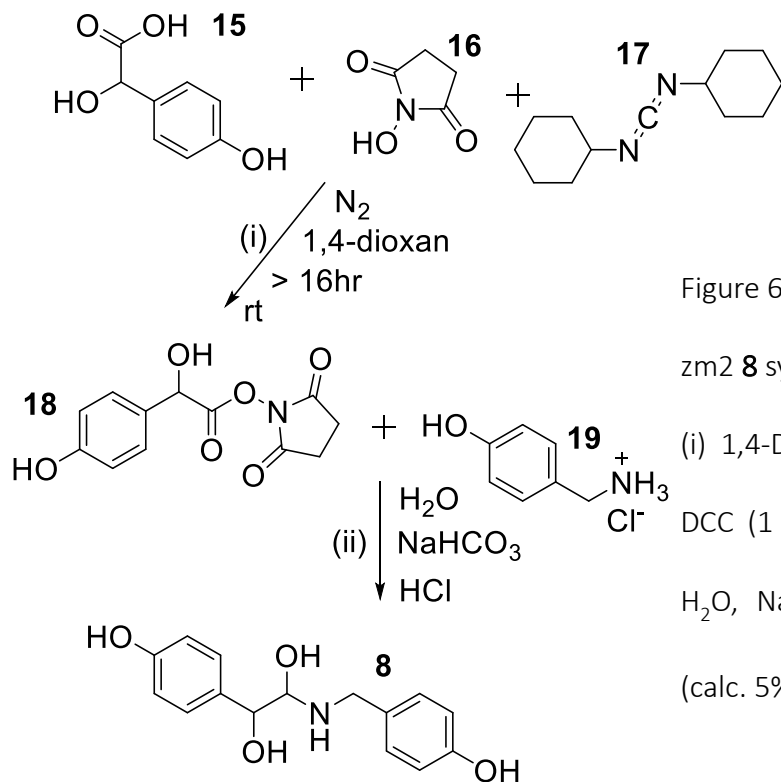
Figure 62. Scheme IV: Conditions for zm2 **8** synthesis.

(i) 1,4-Dioxan, NHS (1 equivalent), DCC (1 equivalent), rt, ≥ 16hr; (ii) H$_2$O, NaHCO$_3$, (1 equivalent), HCl (calc. 5%), ≥ 2hr.

Scheme IV

Zm2 **8** is synthesizable both in the non-classical (Scheme **IV**, mechanisms in Scheme **III**), which requires no protection of the phenolic acid amide –OH group (s) and the classical coupling methods using DCC (Scheme **V,** Figure 62, for mechanisms). This requires protection of the –OH group (s).

3.2.2 Synthesis of zm2 8 – Experimental details

2-(4-hydroxyphenyl)-2-oxoacetic acid (4-hydroxymandelic acid) **15** (2 g, 11.9 mmol) and 1-hydroxypyrrolidine-2,5-dione **16** (1.40 g, 11.9 mmol) were dissolved in 100 mL of dioxan (12.3 mL/mmol) and *N,N'*-dicyclohexylcarbodiimide (DCC) **17** (2.5 g, 11.9 mmol) was added with stirring, under N$_2$, at room temperature and left overnight (≥16 hr). The precipitated

by-product **18** was filtered off and 4-(amino methyl) phenol **19** (1.5 g, 11.9 mmol) dissolved in 100 mL of water was added to the filtrate. The resulting product was acidified using HCl (calc., 5%) and extracted with ethyl acetate (3 x 30 mL). The organic layer was evaporated under reduced pressure and dried over Na_2SO_4 to give (2-hydroxy-N-(4-hydroxybenzyl)-2-(4-hydroxyphenyl)acetamide) **8** which was recrystallized from EtOAc to give an off-white solid (1180 mg, 59%), after drying in the desiccator; mp = 196.5-197.3 0C; 1HNMR (400 MHz, DMSO-d_6,) 4.1- 4.2 (2H, m, CH2, H-1'), 4.78 (1H, br. s., H-2'), 5.89 (1H, d, J=3.9 Hz, CH-OH), 6.65 (2H, dd, J = 8.5 Hz, H-2', H-6'); 6.99 (1H, d, J = 8.6 Hz, 2 x H-3', 2 x H-5'), 7.15 (1H, d, J = 8.6 Hz, H-3', H-5'), 8.2 (1H, t, J = 6.16 Hz, N-H), 9.25 (1H, br. s., H-4'); 13CNMR (400 MHz, DMSO-d_6,) 41.7 (CH2, C-2H), 73.7 (CH-OH, C-2), 115.3 (CH: C-3, C-5, C-3, C-5), 128.3 (CH, C-2), 129 (CH, C-6), 130.20 (C-1'), 132.0 (C-1), 156.56 (C-4'), 157.0 (C-4), 173 (C, CO); vmax KBr/cm-1 3369.6 Ar-OH, 2504 O-H (OC-OH), 1545 (amide), 1512 (ArC=C); (ESI MS): 273.1 Found 274.0 $[M + H]^+$; found elemental analysis (%): C 65.7, H 5.59; N 5.1, calc. (%): C, 65.9; H, 5.5; N, 5.1

Reactants to products simplified
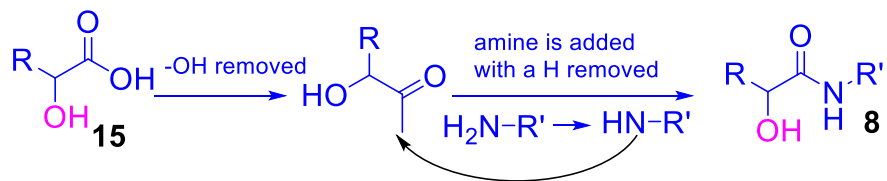
a) deprotonation:
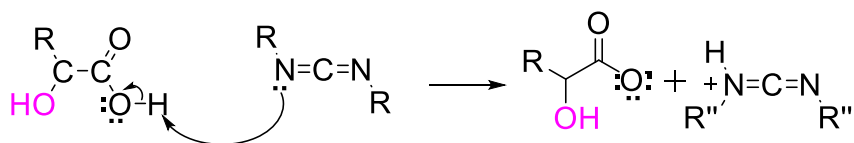
b) nucleophilic attack by carboxylate
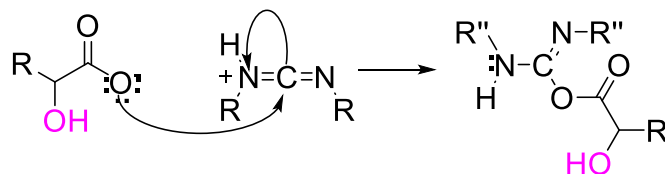
c) nucleophilic attack by amine
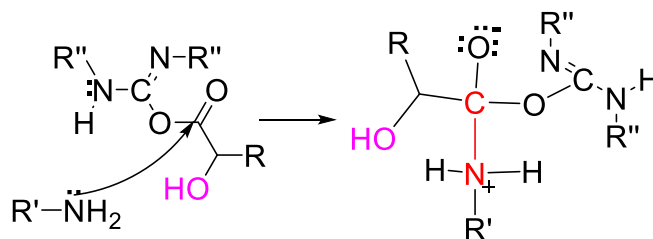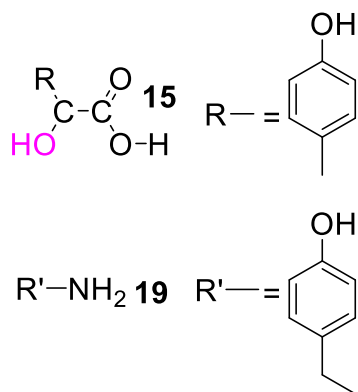
d) proton transfer
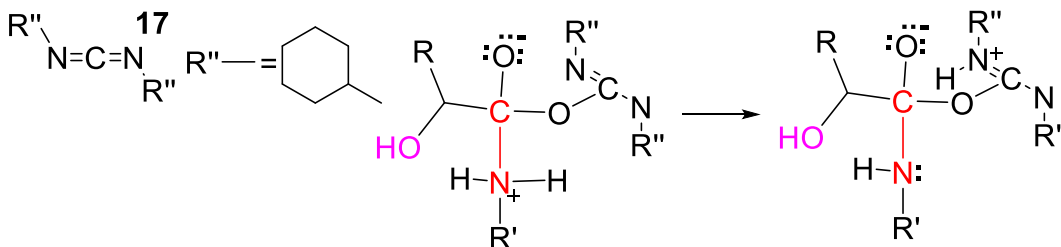
Scheme V

e) leaving group removal

zm2 **8**   Dicyclohexylurea
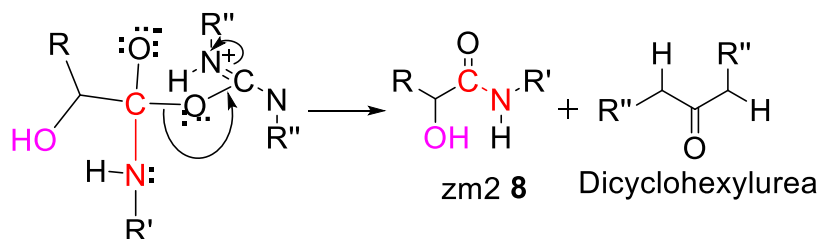
Figure 63. Scheme IV: a classical coupling for zm2 **8** synthesis with DCC.

The direct conversion of a carboxylic acid to an amide is difficult because amines are basic and tend to convert carboxylic acids to their highly unreactive carboxylates. When the carboxylic acid is added to the DCC molecule, a good leaving group is formed that is easy to be displaced by an amine during nucleophilic substitution. DCC - induced coupling to form an amide linkage is an important reaction in the synthesis of peptides or where an amide (-N-C-) is required. Reaction requires protection of vulnerable groups such as –OH (in purple colour) as in phenolic acid **15**. The stable amide bond is shown in red color.

Chapter Four

Bio-evaluation of the selected hits

4.0 Introduction

The screening of antitumor compounds utilizes a number of methods as *in vitro* or *in vivo*, for cytotoxicity, associated with some terms as named. For instance, a 'cytotoxic' compound is toxic to cells regardless whether they are normal or abnormal ones, and may be cytostatic (stops cell growth reversibly or irreversibly) or cytocidal (kills cells); an 'antitumor' compound is active in an *in vivo* tumor system and shows selectivity against the tumor cells; and an anticancer refers to an effective agent against human cancers. All viability screening assay methods have limitations. Typical examples of viability assays are: a) membrane integrity assays - which determine cell membrane integrity via dye exclusion from live cells, e.g., Exclusion dyes, Fluorescent assays or lactate dehydrogenase (LDH) leakage; b) functional assays - which examine metabolic components that are growth essentials, e.g., (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) (MTS), crystal sulphate /acid phosphatase or neutral red assay; c) Deoxyribonucleic acid (DNA) labeling assays- these carry out simultaneous cell selection and viability, e.g., Fluorescent conjugates; d) morphological- which determines any morphological change, e.g., microscopic observation; and e) reproductive- which defines rate of growth, e.g., colony formation. The functional assays using MTS, a colorimetric-based sensitivity, cytotoxicity, and quantitative assay of viable cells in proliferation shall be used. The reduction of MTS tetrazolium compound by viable cells

generates a colored formazan dye product that is soluble in cell culture media, a conversion effected by nicotinamide adenine dinucleotide (phosphate) hydrogen (NAD(P)H)-dependent dehydrogenase enzymes in metabolically active cells. The formazan dye produced can be quantified by measuring the absorbance at 490-500 nm with the microplate reader. It is rapid, non-radioactive, safe, easy to use, flexible, accurate, permits high-throughput analysis under reduced medium and plastics costs, and allows volume miniaturization in 96-well microliter plates [192]. The prepared MTS solution (MTS and an electron-coupling reagent (phenazine methosulfate) PMS) is added directly into the wells of the cell culture media without intermittent steps, making MTS to have an edge over its group members like 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) [193].

4.1 Some parametric terms used in the growth inhibition studies

The cellular response parameters applied are the median growth inhibitory concentration (dose) or growth inhibition of 50 % (IG_{50}), total growth inhibition (TGI) and half-maximal inhibitory concentration or concentration (dose) of inhibition of 50% (IC_{50}). They are incorporated values that represent the concentrations at which the term percentage growth (PG) is +50, 0, and -50, for IG_{50}, TGI and IC_{50}, respectively [194-196].

4.1.1 Data generation and treatments

This is termed the measured effect of the compound on A431 wild type (A431wt) cell line premeditated in this work as follows:

a. If (Mean A_H – Mean A_C) $\geq$ 0, then

PG = 100 x (Mean A_H – Mean A_C)/(Mean A_C – Mean A_B)

b. If (Mean A_H – Mean A_C) < 0, then

PG = 100 x (Mean A_H – Mean A_C)/Mean A_C,

c. If (Mean A_H – Mean A_C) = 0, then PG = TGI = 0, which implies A_H = A_C

where: Mean A_H = the average absorbance (A) of hit dose; Mean A_B = the average A of 0.1 % DMSO + A431wt; Mean A_C = the average A of the blanks (100 μM 0.1% DMSO or complete medium minus cells) [197] (Table 33). Concentrations are expressed by log10 μM. The values for mean ±SD of data from the triplicate wells are calculated using the GraphPad Prism and adds the error bars automatically (GraphPad-Prism Software Inc., San Diego, CA).

Table 33. The 96-well microliter plate format for colorimetry.

Letters A –H and Nos. 1-12 represent the microliter plate map. Six rows (A – F) were used for the two different concentrations (e.g., 5000 cells in A – C & 8000 cells in D – F) as divided by the colors orange for one and dark blue for another concentration. The 12 columns are as labeled the blank column 1 (complete medium, for A_B), control column 2 (0.1% DMSO & cells, for A_C) and concentration or dose columns 3- 12 (test drug or 'hit', A431wt cells & medium, for A_H). The mean absorbance values for A_B, A_C and A_H for the first cell concentration, e.g., 5000 cells, are each (A + B +C)/3 & for, e.g., 8000 cells, (D + E + F)/3. The mean absorbance value in each column for different A431wt cell concentration reflects the hit dose, which can be normalized to percentage (0%- 100%) in GraphPad Prism for calculation of IG_{50} and IC_{50}. Section 4.3 and Figures 65 – 72 show the dose –response curves (DR) of the analyzed compounds.

Row	1	2	3	4	5	6	7	8	9	10	11	12
	Blank	Control	Hit concentrations (µM) used all through									
A	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
B	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
C	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
D	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
E	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
F	100µL medium	10µL(.1%v/v DMSO) + 90µL (cells)	1	2.5	5	7.5	10	15	25	50	80	100
G												
H												

4.1.2 PG

This term measures the effect of the compound (hits) on (A431wt) cell line in percentile [197] and a useful parameter that defines other predictions [198].

4.1.3 IG$_{50}$

GI$_{50}$ value defines the concentration for which PG = +50, and demonstrates that a drug effect of this intensity is interpreted as primary growth inhibition, i.e., the drug concentration that gives a 50% reduction in the net protein increase. GI$_{50}$, a T/C-like parameter, where T = test growth (H, 'hit') & C = control growth, is capable of having values from +100 to-100, which makes it an acceptable measure of the growth inhibitory power of the tested compound (s) [197].

4.1.4 IC$_{50}$

This half-maximal inhibitory concentration measures the effectiveness of a substance that inhibits a specific biochemical function (minimal inhibition can be estimated as well), i.e.,

the molar concentration of inhibition that specifically reduces the response by 50% of the maximum attained in a biological or biochemical function [199]. It is equivalent to the concentration for which the PG = -50 [20].

4.1.5 TGI

This value shows that absorbance (number or mass of cells) in the hit dose well when an experiment ends equals the absorbance (number or mass of cells) in the control well. A drug with such intensity is cytostatic. In order words T (or H, hit dose) = C [197].

4.2 Hits for bioevaluation

DPT1 **1** – DPT6 **6** and zm1 **7** & zm2 **8** (Tables 5, 9 & 27) are slated for bioassay. However, zm1 **7** and zm2 **8** are the optimized products. Zm1 **7** had been bioevaluated following its spectral details before the unexpected crystallization of zm1a **7** from regular dimethyl sulfoxide (DMSO) (70 % by volume) and distilled water (30 %), and therefore yet to be assayed. Zm2 **8** is synthesized and characterized. The predicted drug-likeness, lead-likeness and a few other Lipinski-like filters of these compounds (DPT1 **1** – DPT6 **6** and zm1 **7**, zm1a **7** & zm2 **8**) using ChemAxon Software are shown in Table 31. Table 32 shows the predicted anti-EGFR of compounds zm1 **7**, zm1a **7** and zm2 **8**, using the EGFR$_{Pred}$, a ChemAxon Software, though the mechanism is not defined by the Software [176]. The downstream cascade resulting from the autophosphorylation of the tyrosine kinase aa residues requires more protein scaffolds and adapters to attain biochemical outcomes [200]. Akt [201] and ERK1/2 [202] are such key adaptor pathways among others. Anti-oligomerization activity is presumed to have occurred when phospho-Akt and phospho-ERK1/2 are missing and /

or smaller than the control, which are quantifiable [203, 204], which sets the pace for WB [205]. Figure 64 describes the model the bioassay is based.
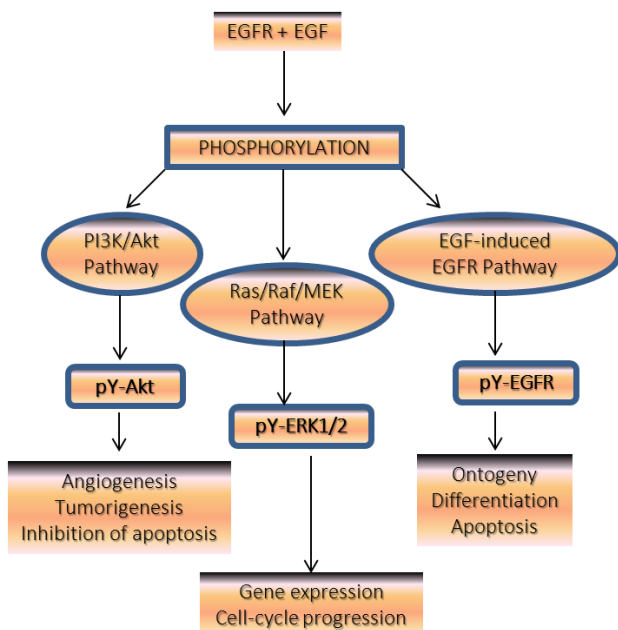
Figure 64. Model for the WB analysis. This is based on selecting the PI3K/Akt, Ras/Raf/MEKERK1/2, and EGF-induced EGFR activation as key signal transduction pathways to be blocked by DPT_H, if active.

A431wt cell line, a human epidermoid carcinoma cell line, is chosen for its extremely high number of EGF receptors on its cell surface (3 x 10^6/cell, ≈ 0.2% of total protein) [206], which makes protein quantification and WB viable at very low determination levels.

4.3 Experimental – Colorimetry results

We adopted the CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay protocol. Samples were solubilized in DMSO without filtration or sterilization of drug stocks at 100 – fold the desired final maximum 'hit' concentration and stored frozen prior to use. The A431wt cells were grown in DMEM-Dulbecco's Modified Eagle Medium; 10 % FBS; 2mM glutamine; 100 IU penicillin & 0.1 mg/mL streptomycin in T75 Flasks at 37 °C, 5 % CO_2, 95 % Air and 100 % relative humidity Incubators. The 96-well microliter plates were seeded

with the A431wt at cell strength range 4 x 10^3 - 1 x 10^4 cells/mL and incubated in a 100 %

humidified 5 % CO_2 Incubator over night at 37 ^{o}C. The old medium was removed and fresh

medium added. 10.0 μL containing various prepared concentrations (not serially) of 'hit'

or drug (DPT1 **1 -** DPT6 **6** & zm1 **7** or zm2 **8**) was added whereas positive/negative control

has standard 'hit' and no 'hit', respectively (as illustrated in Table 33). Plates were

incubated for 48hr at 37 ^{o}C in 100% humidified 5 % CO_2 Incubator. The medium was

removed from the wells at the end of incubation and 20 μL of combined MTS/PMS

prepared solution pipetted into each of the 96 wells assay plate containing the 100 μL of

cell line. Plates were incubated at 37 ^{o}C in humidified CO_2, 5% atmosphere and 100 %

relative humidity for 2 - 3hr and readings were taken at an absorbance of 490 nm using

Bio-Rad Colorimetric Assay Plate Reader [207]. Each graph was plotted using the

v6.0GraphPad Prism. GraphPad Prism makes provisions to statistically accommodate

outliners for 'best' and reproducible curves for tested compounds, as well as it is fast, can

express the results as fractions or percentages and the Software is easily manipulated, for

instance, calculations can be made for concentrations of 'hit' agents giving a H/C value of

10 %., or 10 % growth inhibition (IC_{90}), H/C value of 25 %, or 75 % growth inhibition (IC_{75})

and H/C value of 50 % and therefore growth inhibition 50 % (IC_{50}) [197]. Figures 65 – 71

show the dose – response curves (DR) of the compounds bioevaluated using the GraphPad

Prism. Table 34 shows the summary of found and calculated dose-response curves (DR)

and it is indicated where there is no observed result (nor).

4.3.1 Dose-response curves for DPT1 1
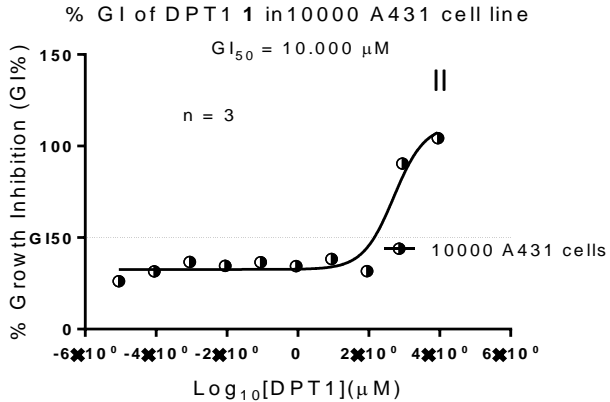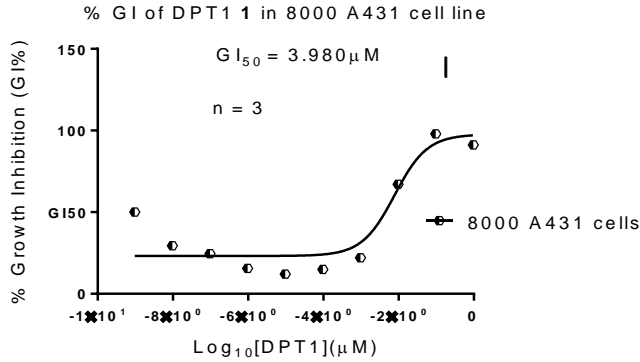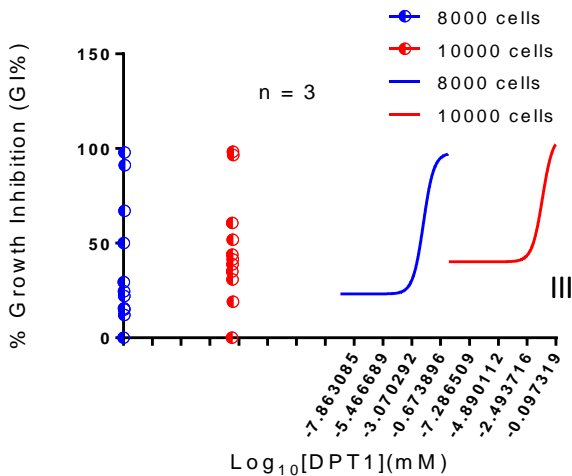
Figure 65. DR of DPT1 **1** vs cell line strength 8 x 103 & 1 x 104 cell/mL.

I: the concentration that gave 50% reduction in the net protein increase of A431wt cells in cell strength of 8 x 10^3 cells/mL, IG_{50} = 3.980 µM, SD ±30.69; **II:** IG_{50} of 1 x 10^4 cell/mL = 10.000 µM, ±25.25, for a triplicate determination per DPT1 **1** concentration; **III:** Grouped sketched of **I** & **II** shows the difference between cell strengths.

4.3.2 Representative dose-response curves for DPT2 2

% G I of D P T 2 **2** in 8 0 0 0 A 4 3 1 cell line

$IG_{50} = 0.251 \mu M$

I

n = 3

%Growth inhibition(%GI)

GI50

▽ % G I (8000 cells)

$Log_{10}[D P T 2]$ (μM)

% G I of D P T 2 **2** in 1 0 0 0 0 A 4 3 1 cell line

$IG_{50} = 1.585 \mu M$

II

% G r o w t h Inhibition (G I%)

GI50

▲ % G I (10000 cells)

$Log_{10}[D P T 2]$ (μM)

% G I of D P T 2 **2** in 8 0 0 0 & 1 0 0 0 0 A 4 3 1 cell line

IG_{50} (8000 cells) = $0.251 \mu M$; (10000 cells) = $1.585 \mu M$

% G I (8000 cells)
% G I (10000 cells)
% G I (8000 cells)
% G I (10000 cells)

n = 3

III

Growth Inhibition (%)

GI50

$Log_{10}[D P T 2]$ (μM)

Figure 66. DR of DPT2 **2** vs cell line strength 8 x 103 & 1 x 104 cells/mL.

I: the concentration that gave 50% reduction in the net protein increase of A431wt cells in cell strength of 8 x 10^3 cells/mL, IG_{50} = 0.251 μM, SD ±34.36; **II:** IG_{50} of 1 x 10^4 cell/mL = 1.585 μM, ±31.96, for a triplicate determination per DPT2 **2** concentration; **III:** Grouped sketched of **I** & **II** shows the difference between cell strengths.

4.3.3 Representative dose-response curves for DPT3 3
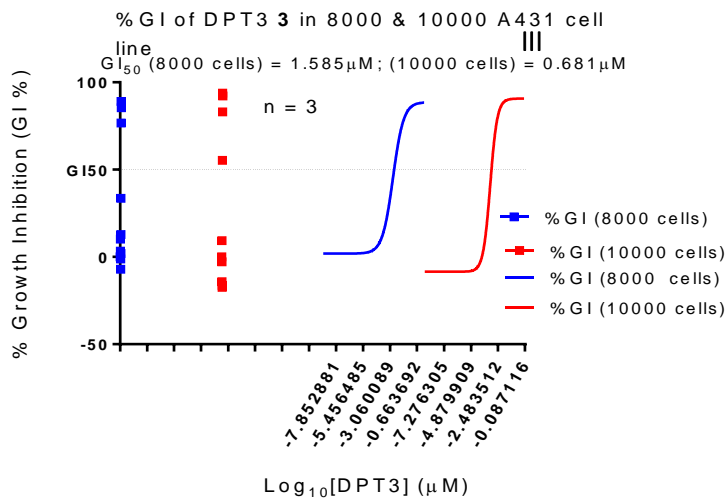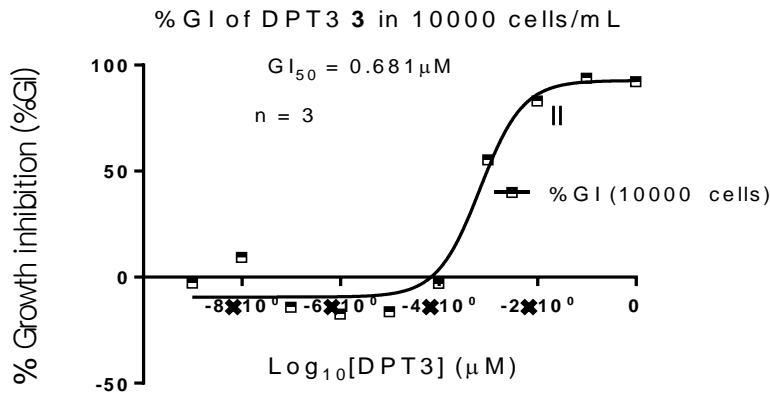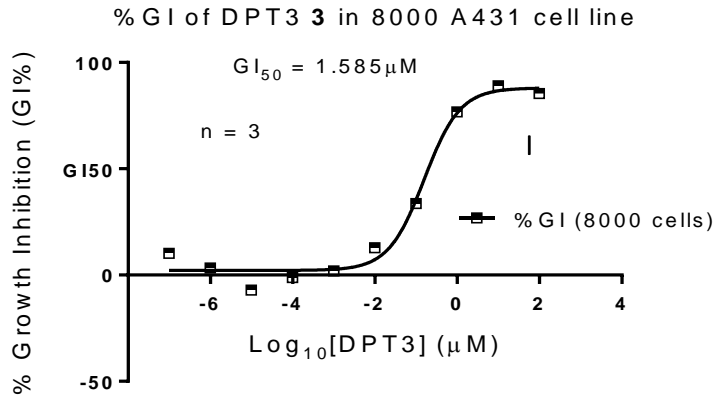
Figure 67. DR of DPT3 **3** vs cell line strength 8×10^3 & 1×10^4 cells/mL.

I: the concentration that gave 50% reduction in the net protein increase of A431wt cells in cell strength of 8×10^3 cells/mL, IG_{50} = 1.585 µM, ±36.45; **II**: IG_{50} of 1×10^4 cell/mL = 0.681 µM, ±45.04, for triplicate determination per DPT3 **3** concentration; **III**: Grouped sketched of **I** & **II** shows the difference between cell strengths.

4.3.4 Representative dose-response curves for DPT4 4

% G I of D P T 4 4 in 8000 A 431 cell line

$GI_{50} = 1.585 \mu M$

n = 3

I

% G I (8000 cells)

$Log_{10}[DPT4] (\mu M)$
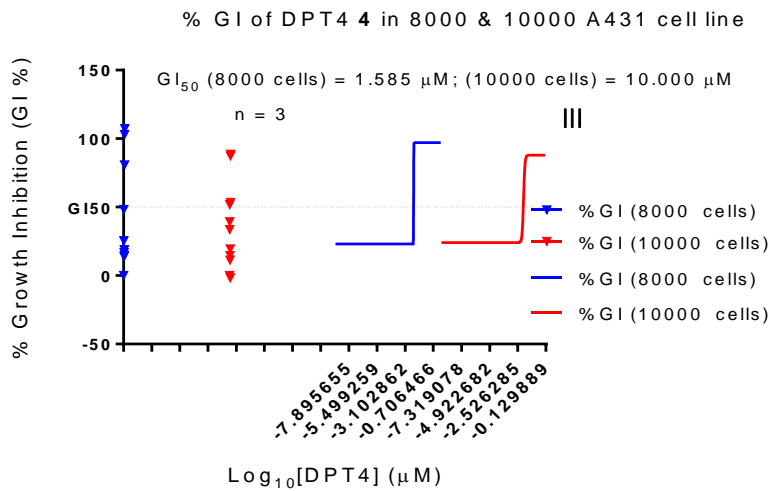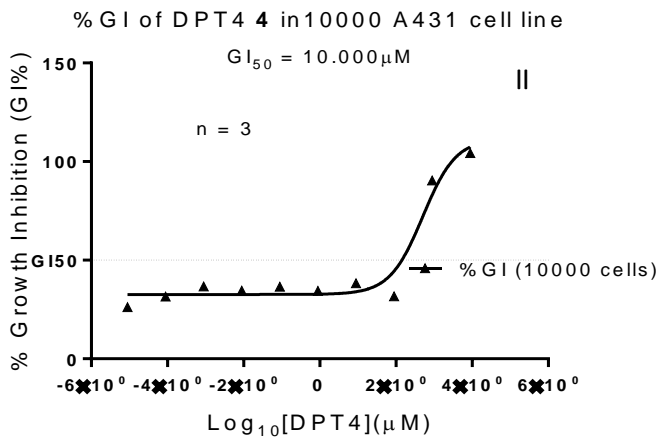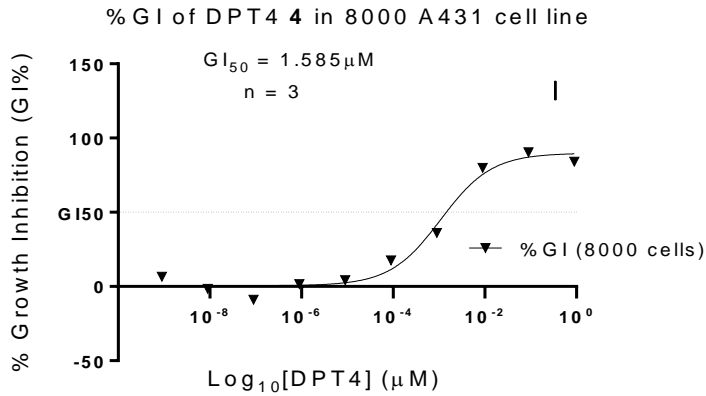
Figure 68. DR of DPT4 **4** on A431wt cell line strength 8×10^{3} & 1×10^{4} cells/mL.

I: the concentration that gave 50% reduction in the net protein increase of A431wt cells in cell strength of 8×10^{3} cells/mL, $IG_{50} = 1.585 \mu M$, SD ± 36.00; **II:** IG_{50} of 1×10^{4} cell/mL $= 10.000 \mu M$, ± 30.89, for triplicate determination per DPT4 **4** concentration; **III:** Grouped sketched of **I** & **II** shows the difference between cell strengths.

% G I of D P T 4 4 in 10000 A 431 cell line

$GI_{50} = 10.000 \mu M$

II

n = 3

% G I (10000 cells)

$Log_{10}[DPT4](\mu M)$

% G I of D P T 4 4 in 8000 & 10000 A 431 cell line

GI_{50} (8000 cells) $= 1.585 \mu M$; (10000 cells) $= 10.000 \mu M$

n = 3

III

% G I (8000 cells)
% G I (10000 cells)
% G I (8000 cells)
% G I (10000 cells)

$Log_{10}[DPT4] (\mu M)$

4.3.5 Representative dose-response curves for DPT5 5
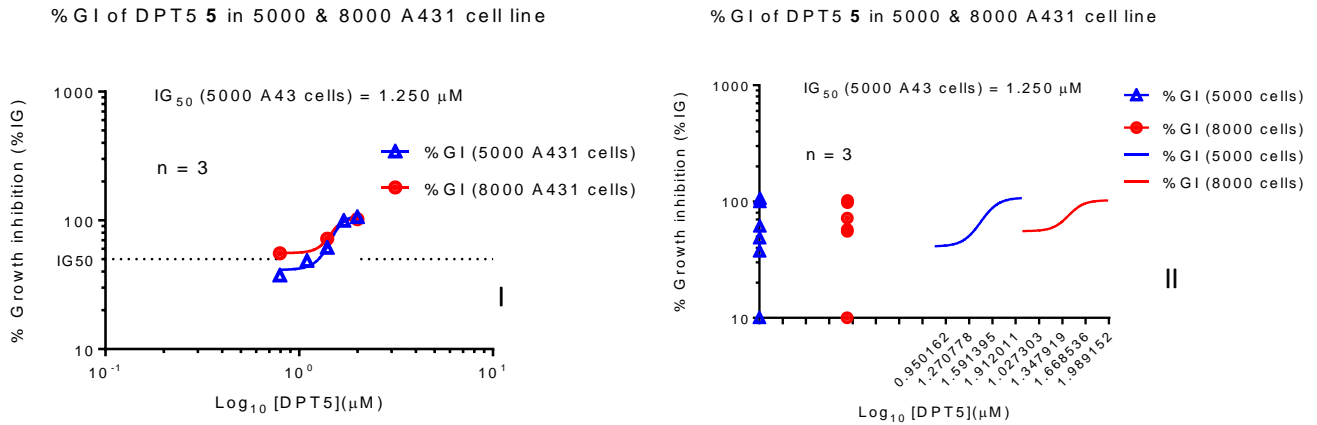
Figure 69. DR of DPT5 **5** vs cell line strength 5 x 10^3 & 8 x 10^3 cells/mL.

I: the concentration that gave 50% reduction in the net protein increase of A431wt cells in cell strength of 5 x 10^3 cells/mL, IG_{50} = 1.250 µM, SD ±27.47; IGtc (threshold concentration) of 8 x 10^3 cells/mL = 6.25 µM, SD ±19.71, for triplicate determination per DPT5 **5** concentration; **II**: A survival sketch enables the sigmoid shape and calculation of the IG_{50}. Grouped sketched of **I** shows the difference between cell strengths.

4.3.6 Dose-response curves for DPT6 6
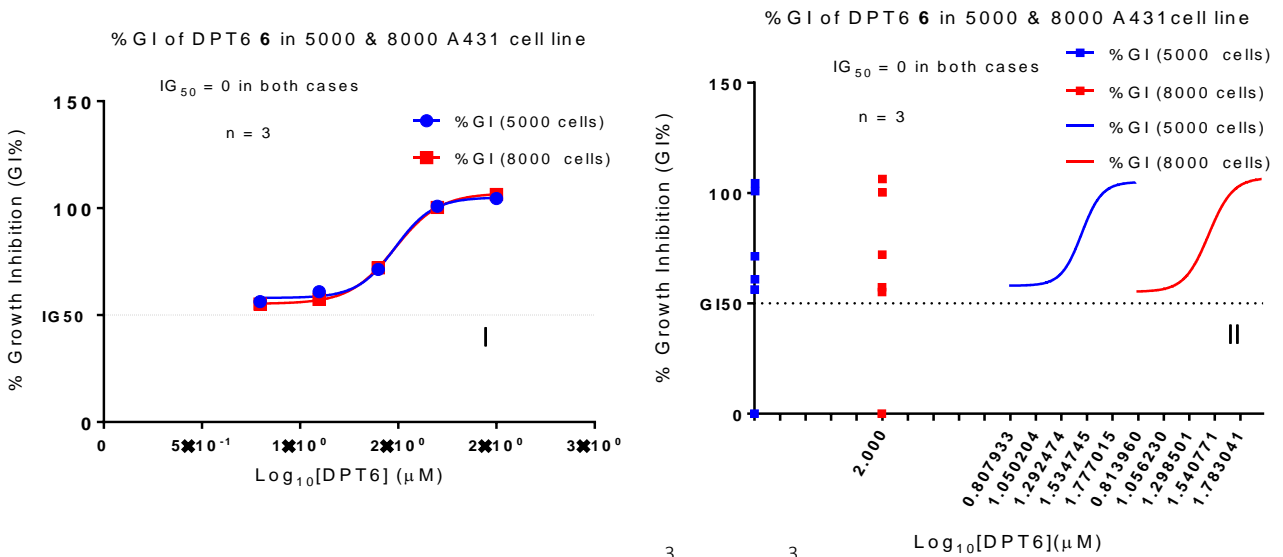
Figure 70. DR of DPT6 **6** vs cell line strength 5×10^3 & 8×10^3 cells/mL.

I: the concentration that gave 25% reduction in the net protein increase of A431wt cells in cell strength of 5×10^3 cells/mL, $IG_{75} = 40.460$ μM, SD ±20.17; IG_{75} in cell strength of $8 \times 10^3 = 40.46$, SD ±21.40, in a triplicate determination per DPT6 **6** concentration; **II**: Grouped sketched of **I** shows the narrow difference between cell strengths.

4.3.7 Representative dose-response curves for the mixed zm1 7
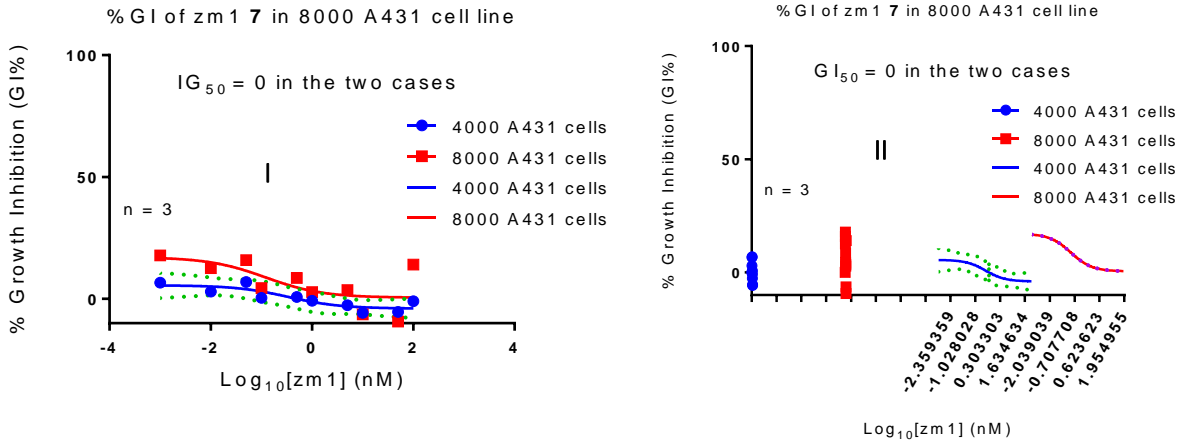
Figure 71. DR of zm1 **7** vs cell line strength 4 x 10^3 & 8 x 10^3 cells/mL.

I: the concentration that gave 25% reduction in the net protein increase of A431wt cells in cell strength of 4 x 10^3 cells/mL, IG_{75} = 30.460 μM, SD ±4.14; IG_{75} in cell strength of 8 x 10^3 = 30.300 ±8.70, for triplicate determination per zm1 **7** concentration; II: Grouped sketched of I shows the difference between cell strengths. Zm1 **7** slopes are different from the rest.

4.3.8 Representative dose-response curves for zm2 8
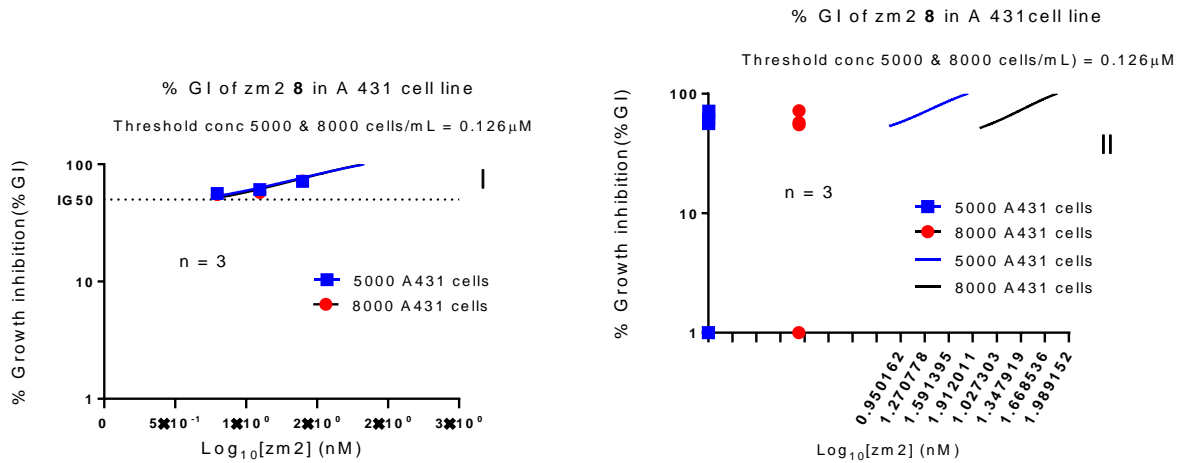
Figure 72. DR of zm2 **8** vs cell line strength 5 x 10^3 & 8 x 10^3 cells/mL.

I: the threshold concentration (tc) required to initiate a reduction in the net protein increase of A431wt cells in both cell strengths of 5 x 10^3 IG_{tc} (threshold concentration) = 0.130 μM, SD ±3.32; the estimated threshold in cell strength concentration of 8 x 10^3 cells/mL, IG_{tc} = 0.130 μM, ±6.31, for a triplicate determination per zm2 **8** concentration; II: Grouped sketched of I shows no difference between cell strengths.

The summary of the values described in the dose-response curves are shown in Table 34. The experimental IC_{50} and the predictive IC_{50} values are also included, in addition to other parametric information.

Table 34. Summary of found and calculated parametric DR values.

The % GI_{50}, % GI_{75}, calculated & investigational IC_{50} for all screened hits in cell strength range $4 \times 10^3 - 1 \times 10^4$ cell/mL used in the experiment are shown. Only zm1 **7** used 4×10^3 cell strength in the random cell strength sampling. The predicted IC_{50} for each candidate under consideration is indicated. The total growth inhibition (TGI) has a zero value since there are IG values.

Lead Candidate	Gscore	%GI_{50} or %GI_{75}[‡] (μM)			TGI^Δ	IC_{50} (found)[¤] (μM)			IC_{50} (μM) (predicted)[†] [208]
		5000 cells/mL	8000 cells/mL	10000 cells/mL		5000 cells/mL	8000 cells/mL	10000 cells/mL	
Anti-oligomerization lead candidates									
DPT1 **1**	-2.70		3.98	10.00	0		8.90	28.90	3.98
DPT2 **2**	-1.90		0.25	1.59	0		0.40	2.40	0.17
DPT3 **3**	-2.10		1.59	0.68	0		1.50	0.70	6.61
DPT4 **4**	-3.50		1.59	10.00	0		1.10	9.10	3.98
Cytoplasmic Tyrosine kinase-Targeted (RTKI-like)									
DPT5 **5**	-10.60	1.25	6.25		0	0.3892	6.25	4.50	6.31
DPT6 **6**	-10.30	40.46 (IG_{75})	40.458 (IG_{75})		0	30.460	30.30	4.80	7.59
Synthesized Anti-oligomeric anti-EGFR									
Zm-1 **7**[†]	-8.00				0	0.3404[‡]	0.1216	-	2.82
Zm-2 **8**	-9.2	0.13 (tc)	0.13 (tc)		0	0.4053	0.0	4.20	5.01

‡ = GI$_{50}$ or GI$_{75}$ was used; Δ = total growth inhibition is zero once drug acts; ¤ = predicted from QikProp (Schrodinger LLC, New York, NY USA); nor = no observed result; tc = threshold concentration; † = zm1 **7** is the crude mixture of zm1 **7** based on spectral information and zm1a **7** which crystallized out late and could not be assayed for time; ‡ = zm1 **7** was analyzed in cell strengths 4 x 10^3 & 8 x 10^3 only

4.4 Experimental: Immunoblotting (WB) results

The preliminary screening of the hits by MTS - colorimetric methods showed DPT1 **1** – DPT4 **4** could further be investigated (Figures 65 – 72 and Table 34). Immunoblot analysis of EGFR have been reported [209, 210]. The protocol to set the activity profile for each of the compounds DPT1 **1**- DPT4 **4** at selected concentrations of 1, 10 and 100 µM and Gefinitib 10 µM in a 2 x 6 – well plate was adhered to (format shown in Table 35).

Table 35. Format for activity profile of the 'hits for WB.

Where DPT$_H$ = the hits of DPT1 **1** – DPT4 **4** & A431wt = A 431 wide type. Hit was allowed an overnight contact with A431wt, however, only about minutes contact when EGF as added to cell well contents.

First six-well plate		
A431wt 1 µM DPT_H 1	A431wt 10 µM DPT_H 2	A431wt 100 µM DPT_H 3
A431wt 1 µM DPT_H +EGF 4	A431wt 10 µM DPT_H +EGF 5	A431wt 100 µM DPT_H +EGF 6
Second six-well plate		
A431 WT 1	A431 WT +EGF 2	A431 WT 10 µM Gefitinib +EGF 3
4	5	6

Each well had 2 x 10^6 cells grew to 80 % confluent, washed and serum-starved overnight in starving medium (0.1 % FBS/BSA, DMEM). The DPT_H (hits of DPT1 **1** – DPT4 **4**) and Gefitinib at said concentrations were added in complete medium (10 % FBS, 1 % PS, DMEM) and incubated at 100 % humidity, 37 °C, 5 % CO_2 and 95 % Air for 1 hr. The cells were stimulated with 10ng/mL EGF for 3 minutes, washed with ice cold PBS and cells lysed with 200 µL lysis enriched buffer per well. The precipitates and lysates were resolved on 7.5% SDS-PAGE, electrophoretically transferred to the polyvinylidene fluoride (PDVF) membrane, and blocked in 5% skim milk in Tris-Buffered Saline and Tween 20 (TBST) for 1hr, to prevent interactions between the membrane and the primary antibody used for detecting target protein. This was washed 3x for 5 minutes with PBS-Tween 0.01% and the

membrane probed with the designated primary antibody [p-Akt, Rabbit (Host), dil.1:1000, Cat. No. 4685 (Cell Signaling Technology, Danvers MA 01923, USA); p-44/42 Map Kinase (ERK1/2), Rabbit (Host), dil.1:1000, Cat. No. 9101 (Cell Signaling Technology, Danvers MA 01923, USA); or p-EGFR (tyrosine), EGFR (Cell Signaling Technology, Danvers MA 01923, USA)], in 5 % bovine serum albumin (BSA) and incubated overnight at 4 $^{\circ}$C on a shaker in the cold room (or for 1hr at rt). This was washed with PBS-Tween 0.01 % for 5 minutes 3x and then probed with a secondary, HRP-linked Anti-rabbit, Goat (Host), dil.1:10000, Cat. No. 7074 (Cell Signaling Technology, Danvers MA 01923, USA) for 1hr at rt and washed again 3x for 5 minutes with PBS-Tween 0.01 %. Then proteins were visualized using Western Lightning® Plus-ECL substrates (Perkin Elmer, Waltham, MA, USA) according to manufacturer instructions. Membranes incubated in ECL were exposed to X-ray film (Fujifilm, Tokyo, Japan). Densitometry was carried out using v1.47ImageJ [211].

4.4.1 Effects of DPT1 1-DPT4 4 on EGF-induced Akt activation

Conc (µM)	-	-	1	10	100	1	10	100	1	10	100	1	10	100
DPT	-	-	1	1	1	2	2	2	3	3	3	4	4	4
EGF	-	+	+	+	+	+	+	+	+	+	+	+	+	+

p-Akt

Total Akt

p-Akt

Figure 73. Effects of DPT1 **1** to DPT 4 **4** on EGF-induced pY-Akt1/2 activation.

The profile activity of the DPT1 **1** – DPT4 **4** was prepared according to the protocol at concentrations 1, 10 & 100 µM for each DPT. Cells were starved overnight in medium 0.1 % FBS/BSA, DMEM, and on rinsing were incubated with the compounds in non-starving medium10 % FBS,1 % penicillin & streptomycin (PS) , DMEM. The cells were then stimulated with 10 ng/mL EGF. WB assessed cell lysates.

4.4.2 Effects of DPT1 1-DPT4 4 on EGF-induced ERK1/ERK2 activation

Conc (µM)	-	-	1	10	100	1	10	100	1	10	100	1	10	100
DPT	-	-	1	1	1	2	2	2	3	3	3	4	4	4
EGF	-	+	+	+	+	+	+	+	+	+	+	+	+	+

Figure 74. Effects of DPT1 **1** to DPT 4 **4** on EGF-induced pY-ERK1/2 activation.

The profile activity of the DPT1 **1** – DPT4 **4** was prepared according to the protocol at concentrations 1, 10 & 100 µM for each DPT. Cells were starved overnight in medium 0.1 % FBS/BSA, DMEM, and on rinsing were incubated with the compounds in non-starving medium10 % FBS,1 % penicillin & streptomycin (PS) , DMEM. The cells were then stimulated with 10 ng/mL EGF. WB assessed cell lysates.

4.4.3 Effects of DPT1 1-DPT4 4 on EGF-induced pY-EGFR activation and representative graph for all
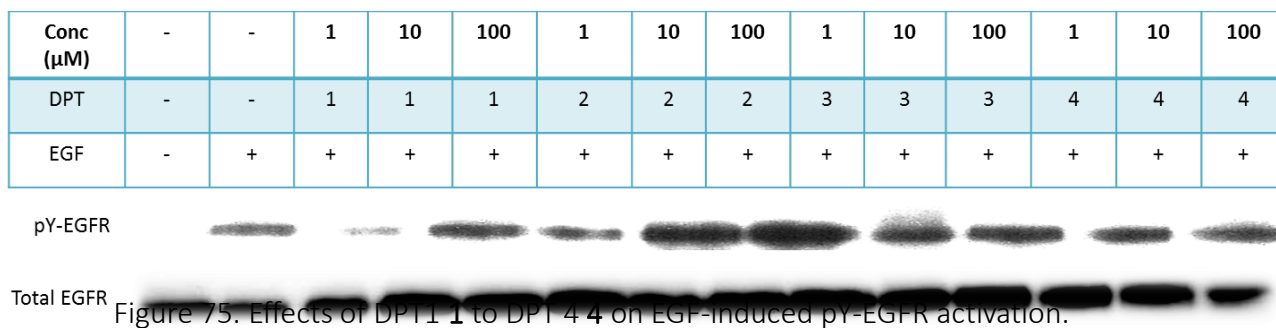
Conc (µM)	-	-	1	10	100	1	10	100	1	10	100	1	10	100
DPT	-	-	1	1	1	2	2	2	3	3	3	4	4	4
EGF	-	+	+	+	+	+	+	+	+	+	+	+	+	+

pY-EGFR

Total EGFR

Figure 75. Effects of DPT1 1 to DPT 4 4 on EGF-induced pY-EGFR activation.

The profile activity of the DPT1 **1** – DPT4 **4** was prepared according to the protocol at concentrations 1, 10 & 100 µM for each DPT. Cells were starved overnight in medium 0.1 % FBS/BSA, DMEM, and on rinsing were incubated with the compounds in non-starving medium10 % FBS,1 % penicillin & streptomycin (PS) , DMEM. The cells were then stimulated with 10 ng/mL EGF. WB assessed cell lysates.
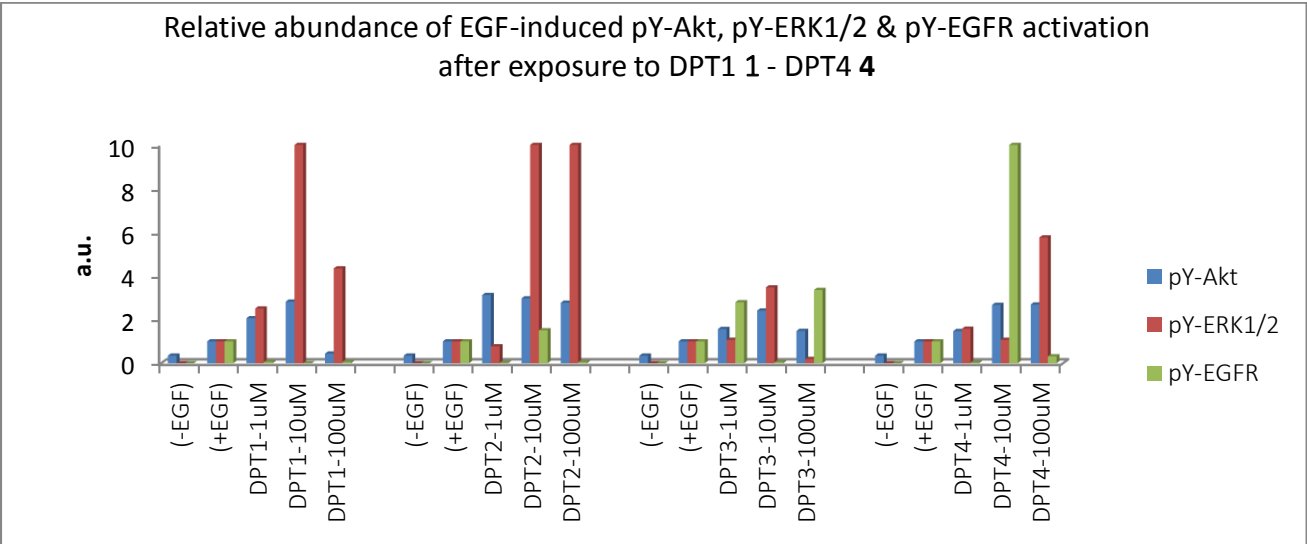
Figure 76. Graph showing normalized data from Figures 72 – 74 using ImageJ.

This was analyzed by densitometry using ImageJ Software program (available from http://imagej. nih.gov/ij/) that analyzed the selected image lanes on the film to measure the relative amount of the specific protein on the blot by comparing it with a control.

Chapter Five

Result analyses and discussions

5.0 Introduction

There are two established types of ErbB-RTK subfamily inhibitors, viz, a) the mAbs and b) the RTKIs, which act at different domains of the receptors. For instance, the mAbs act at the ectodomain locale either by interfering with the ligand such as EGF [212] or the receptor such as EGFR [213], in each case interference with dimerization occurs, while the RTKIs act in the cytoplasmic tyrosine kinase domain. However, little has been done in the dimerization region as per developing specifically anti-EGFR anti-dimeric small molecules that inhibit the dimerization processes in spite the ligands bound to their canonical ErbB receptors [87]. In 'Targeting the Oligomerization Region of the Epidermal Growth Factor Receptor', we have attempted to design small molecules that act in the 'oligomerization region' (Section 1.5.2). Additionally, available data show dimerization is failing to respond to fundamental questions, for which the concept of oligomerization mechanisms, particularly heterotetramerization, has dared to, about the ErbB-RTK subfamily members (Section 1.5.3). Once more, this has polished the scope of our work. In this pioneering piece of the study, EGFR, a known validated cancer target was aimed, specifically the 'oligomerization region' (already defined) using millions of small molecules from free and commercial chemical databases, however, only four were selected finally for purchase and preliminarily bioevaluated. One, zm2 **8**, obtained through the optimization of DPT4 **4**

molecule, was synthesized. It has undergone a preliminary colorimetric (MTS) bioassay with success, however, it still requires further bioanalysis. One more potentially active and synthesizable compound, zm1 **7**, awaits its synthesis and characterization for bioevaluation. A third synthesizable compound, zm1a **7**, inadvertently crystallized out while synthesizing zm1 **7**, which, hopefully will further be characterized and bioevaluated.

5.1 The Molecular modeling (MM)

MM incorporates altogether the theoretical methods and computational techniques used to model or mimic the behavior of molecules [214]. Molecular docking (MD) is key in MM as it predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex [215]. The knowledge of the preferred orientation in turn can be used to predict the strength of association or binding affinity between two molecules using, for example, scoring functions. MD interprets MM, hence it is an optimization problem when optimized conformation for both receptor and ligand is achieved, thus making it central to computer-aided rational drug design [216]. Our molecular modeling appreciated these briefs and we went on to achieve as described.

5.1.1 The Grids

Following the modification of the proteins (Section 2.2), a grid was generated from the 2J6M, which meant removal of AEE788 for new ligands to come in, however, five grids were generated from the 1IVO, namely, |246-253|S (Figure 27A), |246-253|W (Figure 27B), |86 230 263 264 265 275 278 283 285 286|S (Figure 27C), |86 230 263 264 265 275

278 283 285 286|W ((Figure 27D) and |200-300|W (Figure 27E) (Section 2.3.3). A total of about six grids were generated. When |246-253|W was docked in a similar way as |246-253|S, the Gscores were 2 – 3x higher as seen in section 2.5.7. Furthermore, |246-253|S did not show a Gscore when docked with ZINC ligands but |246-253|W had as shown in Table 9. The docking of |86 230 263 264 265 275 278 283 285 286|W with a number of different ligand sources yielded encouraging Gscores as contained in Tables 17 & 18 of section 2.5.7. A number of grids were not utilized sufficiently such as |86 230 263 264 265 275 278 283 285 286|S, |200-300|W & |86 230 263 264 265 275 278 283 285 286|W for time. For instance, the optimized lead DPT4 **4** yielded zm1 **7** and zm2 **8** when docked with |200-300|W, showing high Gscores (Table 30). 2J6M docked outcomes had Gscores generally 3 – 5x those of the |246-253|S as shown in Tables 8, 9, 10 & 12 (cf. Tables 4, 5, & 6). These observations were insightful as our docking outcomes could possibly be influenced by the quality and / or nature of grid generated. A number of factors may be considered for these observations.

5.1.2 The Ligands

It is important to note that Ambinter, ChemBridge and the curated ZINC Chemical Databases shared the same coding system, ZINC, at the time this Project commenced. A total of about ≥ 9 x 10^6 ready-to-dock ligands from seven chemical databases were used in the docking, with focus on the |246-253|S and the subsidiary grid 2J6M (Section 2.4). Not all ligands docked properly with this grid, e.g., from the ZINC Chemical Database (see Figure 28). The ligands, however, gave the 'best' scores with 2J6M (Table 9) which

ultimately led to selecting DPT5 **5** & DPT6 **6** as hit leads as well (Table 29), despite domain differences. Similarly, **|**246-253**|S**-ZINC ligands yielded but **|**246-253**|W**-ZINC ligands did as shown in Table 16. 2J6M generally excelled with most ligands docked and were used **W** (dimer) docking for time. These few observations indicate an adequate knowledge about ligand sources may save time and resources prior to docking. Moreover, there are specific or focused and general ligand sources. At the moment none is known for the oligomerization region.

5.1.3 Glide docking

Glide has been instrumental to the grid generation of the prepared receptors, ligand docking, and analyzing their parameters (Section 2.4.2). Glide can be set to dock flexibly or rigidly ligands into, usually, a rigid receptor structure as already explained and the nature of the grid matters as shown in Figures 28, 29 & 45 and reflective results in Tables 29 & 30. Glide in modeling stands out among others [172]. However, Glide can yield both false-positives and false-negatives as well [158], which can be treated by post docking processes. Or better, values of 15% and 2% should be adapted depending on the library size and preparedness to wait. For instance, if the library is made of 10,000 structures or less it is alright to directly use the XP procedure to avoid false-positive and false-negative results (with 10 jobs with 1000 structures for example, each job will take 5,000 minutes i.e. 3-5 days). However, Glide makes efforts to take care of the false-positives and false-negatives using its extra-precision (XP) mode that combines a powerful sampling protocol with a custom scoring function that is specifically designed to eliminate false positives [171], thus,

further improve enrichment. Time, size and nature of ligand and receptor were all observed to influence the verdicts of Glide. The fact that we have Schrödinger Release 2015-4, suggests affirmatively Schrödinger (Schrödinger, LLC, New York, NY), the sole Proprietors of the Schrödinger Suite, is addressing some of these issues, and to remain competitive.

5.1.4 Glide parameters for qualifying and quantifying docking outputs

It is a status that we rate the success or failure of our docked outcomes by the final GlideScore (Gscore, GS) in Glide, which may be interpreted as the final quantification or scoring of the functions in the Program. However, it is important to note that several parameters entwined for a successful scoring function. "Which is the least or the most important parameter" seems to be a right query to dabble with. In this piece of work, we appreciated deeply what we obtained as final Gscores, however, we needed another approach to handle our docking outcomes, which finally led to the selection of the 'hit leads' DPT1 **1** – DPT4 **4** despite their very low Gscores (already explained), which subsequently became our 'lead candidates' on successful preliminary bioevaluation. There are twenty-two Maestro properties generated by HTVS and SP during docking runs and an additional fifteen Glide XP terms (plus their subdivisions) and their visualizations, for which it is expected that docking score (DS) (all the additional terms) and Gscore (sum of XP terms) should have an equal figure, as an instance [171]. Fails are expected as seen in Table 9 where DS ≠ GS.   We deliberately chose docking properties such as DS or Gscore, GLE, ME and GLESA & a Glide XP term, LowMW, among other terms or parameters in v6.9Glide

program to characterize docked compounds. The Scatter Plots from the Project Table and v2.6Canvas Programs were used to see if correlations existed among a group of docked compounds (Section 2.5), the influence of the receptor (s), the docking software/scoring functions and ligand nature. The Scatter Plots for the lead candidate compounds were particularly focused on, with occasional comparison with known compounds. ME was the most featured for reason it appears to be the most intertwined with other parameters (Section 2.4.3). However, the use of v2.6Canvas showed that the comparison of DS, GLESA, ME and GE docking properties were within defined ranges Tables 7, 11, & 24 − 28. The Scatter Plots showed wide variations (Section 2.5 introduces correlation values). Figure 36 shows those of known compounds (Built ligands from Table 13), which had poor correlation (R^2 = ≤ 0.50) except in DS & GLESA independent of the GE; Figures 32A − 32D showed poor correlation except in DS & GLESA; Figure 36 had poor correlations in all except DS & GLESA independent of GE; Figures 46 – 49 showed poor correlations in all the optimized structures. The Canvas reports about the compounds tested with it were within the required range, as exemplified by Tables 7 & 11 containing the selected hit leads when Ambinter ligands were docked with 1IVO |246-253|S & 2J6M; Tables 24 − 28 show ME and GE are within range for the DPT1 **1** − DPT4 **4**, respectively. There is also a similar observation in the Canvas analyses of 1IVO |246-253|S-RTKI ligands from Tables 4 shown in Table 28. It thus shows Canvas is more relax with Glide parameters than the Scatter Plots.

5.1.5 The CombiGlide enumerated and docked results

We used v3.9CombiGlide to optimize the modified cores of DPT1 **1** – DPT4 **4** which enabled their synthetic feasibilities into designing a combinatorial libraries of the compounds. The workflow flow we harnessed was to instruct CombiGlide to enumerate, dock and analyze our results on a) submitting cores for each of the lead candidates in turn, b) made available the chemical fragment database, and c) chose whether to use a linkage (s) or not. It achieved this through the combination of accurate ligand-receptor scoring and an effective combinatorial docking algorithms (Section 2.5.8, Figure 45). We generated ≥ 2.3 x 10^9 compounds, which actually overwhelmed the software, therefore more than three-quarters could not be docked due mainly to designed properties because CombiGlide functions 'best' for small focused libraries. Only two grids were used, viz, 1IVO |246-253|**S** and |200-300|**W**, which were docked with optimized DPT1 **1**-DPT4 **4**. The results of docked |246-253|**S** are shown in Tables 20 – 23. Gscores improved 3-5x as shown in these Tables 20-23 & 30. Table 30 in particular resulted 1IVO |200-300|**W**-optimized DPT4 **4** core. Although, most of these new chemical entities were neither purchasable nor readily synthesized, the |200-300|**W**- yielded two synthesizable compounds zm1 **7** & zm2 **8** (Table 30). The 'best' optimized compounds were from DPT4 **4** (Table 30) while DPT2 **2** yielded the most ionized products (↑pKA) (Table 21). The physicochemical properties of the outcomes of the optimized lead candidates depended on the nature of the fragments and cores, some of which have been analyzed using Glide features as shown by their Scatter Plots in Figures 46 – 49, where variations in their DS were significant, though Canvas

analyses in Tables 24 – 27 depict their values as within range, which expressed a similarity with the known compounds in Table 28.

5.1.6 Drug-likeness and lead-likeness of the lead candidates

Gscore is a relatively expensive calculation when Glide is ran. In its virtual screening workflow, filtering requires an unconditional run of QikProp, prefiltration by Lipinski's Rule and the removal of ligands with reactive functional groups. Glide has ≥ 70 properties, ≥ 18 general attributes and close to 160 functional group counts to use in ligand filtering. This left us with the choice to design our compounds to our desires. Lipinski's Ro5 is the most popular filter, however, its limitations are known (Section 2.4.1). This weakness in Ro5 demanded a search for new empirical rules, commonly referred to as Lipinski-like filters shown in Table 2. These empirical rules have been used to predict some basic drug-likeness and lead-likeness of the lead candidates DPT1 **1**-DPT4 **4**, the RTKI-like DPT5 **5** & DPT6 **6**, zm1 **7**, [zm1a **7**] & zm2 **8**, although their selection procedures were manual. Table 31 shows that all the compounds made the Ro5 and only a few missed it with the Lipinski-like filters, as predicted by ChemAxon Software (already mentioned).

5.1.7 Other physicochemical properties of DPT1 1-DPT4 4

We started the visualization of the physicochemical properties of the four lead candidates in Section 2.5.5 subsections 2.5.5.1 -2.5.5.4 (Figures 37 – 40), where the emphasis was on the mechanisms on how each lead interacted with its bonding aa residues to form the key noticeable linkages, namely, i) hydrogen bonds, ii) hydrophobic bonds and iii) the $\pi - \pi$

interactions. The densest were the hydrophobic bonds which seemed to be influenced by the structure of the compound (section 2.1). Only DPT1 **1** & DPT3 **3** demonstrated their $\pi - \pi$ interactions which occurred between the TYR246 aromatic ring and the pyrazole of DPT1 **1** and the quinoline ring of DPT3 **3**. The hydrogen bonds were formed mostly with TYR246 & TTYR251 aa residues. Section 2.5.6 subsections 2.5.6.1 − 2.5.6.4 expanded on the section 2.5.5 to offer a more detailed molecular interactions that occurred between the aa residues and the lead hits. Table 14 shows a summary of the aa residual hydrogen bond, hydrophobic and $\pi - \pi$ interactions exhibited by the leads, and following carefully Figures 41A/B, 42A/B, 43A/B & 44A/B. This showed the uniqueness in these compounds as anti-EGFR anti-oligomers, which is can be manipulated chemically to improve potency. The function of MET244 is unclear in the oligomeric arm of the β-hair loop, however its significance in the 'oligomerization region' is predictable.

5.1.8 Limitations and challenges in protocols contributory to docking outcomes

In addition to those already mentioned (Section 2.6), a few factors could have influenced the docking outcomes in either way (poor or good docking), some of which are explained as follows: a) since glide docks flexible ligands into a rigid receptor structure by a rapid sampling of the conformational, orientation, and positional degrees of freedom of the ligand of the receptor, this could suggest that the rigidity and the interacting space could have been interfered with, leading to the poor outcome; b) further evidence to this is the increase in ligand-receptor interaction space and therefore Gscores increase in the oligomeric grid docking (Tables 15-18 & 30) as compared to 1IVO |246-253|S in Tables 4,

5, & 6; c) the software version could be a contributor to poor result, for instance, the maximum number of atoms in a ligand is 500 and the maximum number of rotatable bonds is 100 for the 2015-1 release v6.8Glide but our Project commenced with the limits of 300 atoms and 50 rotatable bonds, whereas, for a rigid docking and score-in-place, there is no limit on the number of rotatable bonds; d) Glide concerns mainly with generating an accurate pose for each ligand and docking enrichment (the separation of actives from inactives), however, the task of accurately estimating binding for a set of ligands, particularly if they are diverse in structure, remains a herculean challenge and an active area of research for Schrödinger, which could have influenced our outcome; e) Glide is slow and somewhat poor in handling very large chemical databases, a disposition that could have influenced the outcomes following the millions of compounds docked; f) opposed to undruggability is a good docking target, which should be deep with well-defined pocket as shallow pockets have too many options that guarantee no docking altogether. This is explained this way: the **T** (|246-253|**S**) is a shallow site and this could cause ligands to bypass docking, a feature that is seen in the wide differences in the Gscores of **W** and **S**; g) the target site for specific interactions should have well-ordered side chains as well as having their backbone chains in place, which was weak situation in the monomeric arm (**S**), and therefore could impact negatively on docking requirements; h) many positive charge-charge or Hb and less hydrophobic interactions, particularly, a minimized vdW bad interactions are conditions that assist in effective docking, a lack of these can lead to poor docking, which could be absent in **S** when 1IVO |246-253|**S**-ZINC

ligand docking was done; i) the mechanism of drug action matters, for instance, 1IVO |246-253|S, which resulted from excising its bonding partner in the cleft, did not only dock poorly, but there could have been a form mechanisms, for instance, docked 2J6M-ZINC ligands had Gscores 3-5x those obtained with other ligand sources, as shown in Tables 8, 9 & 12; and j) on the overall, docking and scoring algorithms are currently challenged with several limitations, some of which have been mentioned, which do affect effective docking leading to unexpected outcomes. Docking is a stair case that depends on so many factors to ascend successfully. Attempts to study some docking properties of known drugs vis-à-vis the same selected chemical properties, exemplified by the RTKIs, curcumin and anti-dimeric anti-EGFR **P** & **Y**), when docked with 1IVO |246-253|S, revealed that docking properties for these compounds fluctuate as well which is noticeable in Tables 3, 12 & 28 and Figures 36 & 54) despite differences in mechanisms. The Gscores and a few other physicochemical properties were observed to be close to each other, curcumin, though its mechanism of inhibition is unknown, had a Gscore of -9.3, which was 2-4x higher than the rest. These observations affirm our believe that DPT1 **1** – DPT4 **4**, in particular, will make good anti-EGFR anti-oligomeric agents.

5.1.9 Summary of drug modeling

The detailed analyses of DPT1 **1**, DPT2 **2**, DPT3 **3** and DPT4 **4** for their inhibition at the ectodomain (1IVO |246-253|S) and endodomain (2J6M) regions of the EGFR, supported by their predicted anti-EGFR, and their general successful passage of the Lipinski-like filter tests, prepared these potential pioneering anti-oligomeric compounds for bioevaluation.

The work lends credence to the report that 'good ligands' are hard to find' [217], for instance, ≥ 2.3 x 10^9 compounds were enumerated and only ≤ 4.4 x 10^{-8} % had Gscores ≥-6.50 and DPT4 **4** yielded the 'best' Gscore hits following its optimization (Table 30). Nevertheless, zm1 **7** and zm2 **8** from DPT4 **4** and Cayman fragment collection were found to be synthesizable. Emodel used as analytical tool of these compounds for reasons explained. DPT5 **5**-DPT6 **6** were chosen for their 'good' Gscores although they have RTKI-like activity, as they were 2J6M docked yields, which can be bioassayed.

5.1.10 Future work

a) The 'oligomerization region' remains a complex site to explore in the development of novel anti-cancer drugs, despite current approaches. This is because it is still being accurately defined [218].

b) The future of anti-oligomers is seemingly set vis-à-vis the emergence that dimerization mechanisms are giving way to oligomerization mechanisms, and we depend more on biomolecular information for improved drug designs, particularly, in SBDD.

c) The compounds that have been identified whether during docking with monomeric or dimeric grids are gold mines to this Project, in search for ways to synthesize and / or purchase them for preliminary bioassay in the drug discovery and development process.

d) DPT1 **1**-DPT4 **4** from 1IVO |246-253|**S** docking, DPT5 **5** & DPT6 **6** from ZINC chemical database docking, thus reflecting RTKI-like compounds, and zm1 **7** & zm2 **8** from the optimized core of DPT4 **4** with anti-EGFR anti-oligomeric activities, we believe have

unlocked their extensive studies in the nearby further using both computer-based and biologically-based approaches to obtain their 'best' as anti-EGFR 'anti-oligomeric' in cancer therapy.

5.2. Compounds zm1 7 & zm1a 7 and zm2 8

5.2.1 Compounds zm1 7 and zm1a 7

The code zm1 **7** and zm1a **7** show the duo had proceeded the same synthetic path until the later crystallized unexpectedly, making zm1 **7** a confirmed mixture of at least itself zm1a **7**. The synthesis of zm1 **7** was challenging, requiring a repeat of more than five times. First, our final yield was insoluble in most available solvents. The chromatographic (TLC, Column) processes and spectral (m/z, 1HNMR and 13CNMR) observations had z1 **7** as a single entity. The elemental results further showed zm1a **7** as the compound, however, there was no spectral data to confirm it. For instance, the elemental analysis of zm1 **7**, calculated (%) - C, 50.59; H, 6.07 & N, 33.71 [219], found (%) – C, 57.28; H, 42 & N, 22.70. The crystallization occurred on dissolution in a mixture of regular DMSO (70% by volume), distilled H_2O (30% by volume), and left over night. The crystallographic report on zm1a **7** indicated its elemental analysis as follows: calculated (%) - C, 57.6; H, 7.2 & N, 22.4, found (%) – C, 57.3; H, 7.4 & N, 22.70 [181], which aligned with those intended for zm1 **7**. Time and its quantity did not permit its further chemical characterization and bioevaluation, however, predictive information about it are shown in Tables 30 & 31. The mechanisms to obtain zm1 **7** and zm1a **7** showed that hydrazinolysis and condensation were required to

yield them, respectively. This was not clear at onset. Additionally, the quantities were infinitesimal and left to proceed despite a slight pale yellow color change when **11** was formed. The multi-component reactions (MRCs) approach utilized has its pros and cons as well. MRC is defined as the process where three or more reactants are combined in a single chemical step to produce products that incorporate substantial portions of the reactants [220]. One of its demerits the mechanisms of certain types of reactions may not be feasible while for infinitesimal quantities reactions leading to by-products may unnoticed. While zm1 **7** possesses the 1°, 2° & 3° amine functionalities and an aromatic ring zm1a **7** has 2° and 3°, we therefore expect zm1 **7** to fragment more. Scheme I in Figure 58 shows the synthesis of zm1 **7** without excess use of hydrazine monohydrate and, instead of obtaining zm1a **7** as a by-product it can be synthesized an entity according to Scheme II in Figure 59. Although the 1H NMR, 13C NMR and IR data for zm1 7 & zm1a 7 are expressed in Section 3.1.3 and included in the Appendix further discussion on this shall be suspended until the chemistry of these two potentially active anti-EGFR and anti-oligomeric compounds are resolved.

5.2.1.1 Summary on zm1 7 and zm1a 7 syntheses: 'divided and lost'

a) The syntheses of zm1 **7** and zm1a **7** may be described as a form of 'divided and lost' because of the inconclusive resolution of the two compounds.

b) Schemes I & II show that the compounds are preferentially synthesizable as zm1 **7** & zm1a **7**, respectively.

5.2.1.2 Future work on zm1 7 and zm1a 7

a) Zm1 **7** and zm1a **7** should be synthesized vis-à-vis their potential anti-EGFR anti-oligomeric activity for which literature has asserted to. Indeed, these are new chemical entities whose structures are easy to easy to manipulate to achieve more potent compounds.

b) Zm1 **7** and zm1a **7** should be synthesized separately noting their clear separate mechanisms as shown in Schemes I and Schemes II, respectively.

c) The bioevaluation of these compounds, zm1 **7** & zm1a **7**, is important because of their potential claims, and should therefore be pursued.

5.2.2 Discussion on zm2 8 synthesis

Zm2 **8** is a new chemical entity with potential anti-EGFR anti-oligomeric activity according to its predictive results in Table 30 and other properties in Table 32. We used a mono-*p*-hydroxy substituted mandelic acid instead of the disubstituted *o*- & *p*- compound. The room-temperature intermediate active semi-stable NHS-ester obtained when DCU has been filtered out, is a two-step reaction [191, 221] as shown in Schemes III & IV, Figures 61 & 62, respectively. Its synthesis is conveniently a simple preparation method, which is a basic requirement in drug discovery. NHS is a commonly used activating agent for carboxylic acids, basically esters, with a good leaving group that reacts with amines to form amides, otherwise the carboxylic acid would just form salt with the amines. It did not need protection of the available –OH functionality and is not harsh either as no heat was

required. It represents a typical non-classical coupling in amide formation [221] as shown by Scheme **III** Figure 61. The classical coupling is illustrated in Scheme **V** in Figure 63, which is a multi-step mechanisms. This approach is an alternative method to synthesis zm2 **8** without requiring NHS, however, the –OH species in the phenolic carboxylic acid must be protected [222, 223]. The spectral and chemical data of the compound are shown in Section 3.2.2 as well as included in the Appendix that affirmed zm2 **8** as a new chemical entity.

5.2.2.1 Summary on the synthesis of zm2 8

a) Zm2 **8** is synthesized.

b) Since two synthetic methods are available for exploitation, the two can be explored to improve yield as well as reduce cost and time to obtain zm2 **8**.

c) Zm2 **8** is set for bioevaluation.

5.2.2.2 Future work

The easiness in synthesizing zm2 **8** offers the opportunity to develop its analogs for a high throughput bioassay screening as anti-EGFR anti-oligomers.

5.2.3 Summary on Compounds zm1 7 & zm1a 7 and zm2 8

Zm2 **8** has been synthesized but not zm1 **7**. Mechanistically, there are two methods zm2 **8** can be synthesized (Scheme **III**, Figures 61 or Scheme **V**, Figure 63). Since zm1 **7** is likely a mixture of zm1 **7** and zm1a **7** based on crystallographic reports, these compounds can be

independently synthesized following Scheme I, Figure 58 and Scheme II, Figure 59, respectively.

5.3 Bioevaluation of hit leads (DPT1 1 – DPT4 4)

The lead hits have been grouped into three sets following assays by the colorimetric methods: i) potentially active anti-EGFR anti-oligomeric compounds, DPT1 **1** – DPT4 **4**; ii) potentially active RTKI-like compounds , DPT5 **5** & DPT6 **6**; and iii) the synthesized potentially active anti-EGFR anti-oligomeric compounds, zm1 **7** & zm2 **8**. Compound zm1 **7** can only be mentioned accordingly and referring to it occasionally, since it was an impure. The bioevaluation of impure compounds can be misleading even when results are tolerable, although the same could serve as a guide as well. The elemental analysis, melting point and the last minute inadvertent crystallization qualified zm1 **7** to be a mixt chemical entity.

5.3.1 MTS –Colorimetric analyses of the compounds

We observed the following when two different cell strengths were used: a) only DPT6 **6** and zm2 **8** showed a single IG values, b) DPT1 **1** and DPT4 **4** had the widest variation as indicated by their SD and expressed by IG values of between 6.00 – 8.00 , and c) DPT2 **2** and DPT3 **3** had close IG values. These points highlighted show cell strength used can influence outcomes in functional assays. Furthermore, different cell concentrations were used in each set. For instance, cell strengths used for DPT1 **1** – DPT4 **4** were 8 x 10^3 & 1 x 10^4 while the rest used 5 x 10^3 & 8 x 10^3 except zm1 **7** that used 4 x 10^3 & 8 x 10^3 cell/mL.

This assisted further to appreciate if variations could occur, which did as shown in some parametric differences in Figures 70– 72 where graph shapes and / or IG & IC estimations were difficult, however, DPT1 **1** –DPT4 **4** results in Figures 65 -68 were easy to estimate. Furthermore, for compounds in 1 x 10^4 cell/mL A431wt cell strength expressed a IC_{50} at lower µM concentration levels such as DPT3 **3**, 0.70 µM; DPT2 **2**, 2.40 µM; DPT4 **4** 9.10 µM; and DPT1 **1**, µM28.90, although this is a subjective explanation. IG_{50} calculations for some compounds was again difficult despite manipulations of the GraphPad Prism, thus the threshold values was resorted to as for zm2 **8** cell strength of 5 x 10^3 cell/mL. A significant observation also is the high IG_{75} values DPT6 **6**, indicating a 25 % concentration of their doses is enough to inhibit EGFR activity. This is important because these ZINC ligands (Table 9) had no prior anti-oligomeric activity when initially docked with 1IVO |246-253|**S**. The synthetic compound zm2 **8** expressed only threshold IG values, beckoning for more work on it. The found IC_{50} values for DPT1 **1** – DPT4 **5**, zm1 **7** & zm2 **8** are significantly lower than the calculated, indicating these compounds have noteworthy potential anti-EGFR anti-oligomeric activities, though DPT6 **6** showed about 5x the predicted value indicating higher doses are required to elicit activity. All the compounds expressed a zero value for TGI, signifying all have some measurable activities against the cell line and therefore, almost all (except zm1 **7**) can be modified structurally to have compounds that are more potent. Table 34 shows a summary colorimetric results. These observations could not have been without the influence of some of the protocols and instrumentation. For instance, it is known that colorimetric methods are not without limitations despite their

highlighted merits [224]. GraphPad Prism plotted DRs (drug response curves) that had anticipated formazan quantity produced and measured at 490nm absorbance was proportional to the number of living cells in the sample. However, this is a clean assumption since other colored metabolites can be produced in living systems in the presence and / absence of a substance at any time, capable of influencing the absorbance at the said λ nm (wavelength) [225]. This is true because it is noticed on a number of occasions where the actual cell number did not agree with assay reading. Formazan amplification by the commonly used culture media in large uncorrectable proportions is a common feature when microplate reader reads results, which do affect analyses in the end. Cell handling protocols can be stressful on the cells before commencement of experiment despite initial 'best' confluent cells. Production of the formazan had been dependent on the MTS concentration; however, the kinetics and degree of saturation of the composite metabolic processes are dependent on both cell type and form, which absorbance at λ 490 nm may not quantify. Assay becomes less effective as proliferation of cells slows down by increasing amount of dye and other metabolites, which goes on to influence analysis outcomes. Another limitation is that individual cell numbers are not quantitated and results are expressed as a percentage of control absorbance. This is in recognition that individual cancer cells in the same medium are not all equal. Finally, it requires a different experimental method to distinguish cytocidal from cytostatic activity of a compound even though a TGI = 0. However, the prescreening clues obtained from MTS assay do offer a helpful guide in bioevaluation processes. The inherent weakness in

colorimetric methods has shown that MTS can underestimate the anti-proliferation of a compound [226]. Some of these points only stress a fact that our result could be highly commendable. A key to obtaining a 'good' sigmoid DR curve in GraphPad Prism is to normalize a datum between 0 and 100 % with mathematically defined baselines. The results shown in Figures 69 – 72 had a challenge to normalize, as attempts to achieve normalization affected both shape and concentration designation. A case at hand is zm1 **7** in Figure 71 having a negative slope. Nonetheless, this can be explained on the basis that zm1 **7** was a crude product. Getting a 'good' dose-response curve still faces challenges, especially in biological assays [227]. Besides, neither cell culture nor animal model systems have provided reliable predictions of drug efficacy or toxicity [228]. Despite these limitations, almost all these compounds assayed by MTS seemed to have made our minimum requirements to proceed to the precipitation and immunoblotting analysis, except for zm1 **7**. DPT1 **1** – DPT4 **4**, in particular, as already mentioned, seem to align their predicted drug-likeness and lead-likeness features with their IG_{50} and IC_{50} values, to become promising anti-EGFR anti-oligomers. Similarly, zm2 **8** expressed an equal threshold concentration (0.130 µM) in both cell strengths and did meet all the drug-like, lead-like and anti-EGFR predictions in Tables 29, 30 & 31. Threshold concentration obtained can be used to measure other effects or toxicity concentrations, such as the concentrations at which a substance, not necessarily toxic, becomes detectable or noticeable by instruments, a human, or an animal. Threshold concentrations have been

used extensively in anticancer drug discovery and development [229, 230]. This emphasizes the need to explore zm2 **8** beyond its present look.

5.3.2 The immunoblotting of DPT1 1 – DPT4 4

The probes that were labeled and bound to the protein of interest needed to be detected on the western blot using one of several methods such as colorimetric, radioactive, and fluorescent methods. However, chemiluminescent method is the most commonly used. Enhanced chemiluminescence (ECL) is a sensitive quantitative method that relatively quantifies protein of interest [231]. It is an indirect enzymatic method that follows these basic principles: I) 2^0 antibodies are labeled with an enzyme reporter, horseradish peroxidase (HRP), II) chemiluminescent substrate (usually luminol) is applied to the blot, III) luminol is oxidized in the presence of HRP and H_2O_2 to form an excited state product that emits light (this defines the method ECL), IV) the light produced by this enzymatic reaction is quantitatively detected by exposure to x-ray film (or a digital imaging with a camera) as shown in Figure 77. The reaction product produces luminescence that is related to the amount of the protein [38]. The blot images in Figures 73 – 75 for DPT1 **1** - DPT4 **4** were achieved from the blot and analyzed by densitometry using ImageJ software program that analyzed the selected image lanes or bands on the film to measure the relative amount of each specific protein on the blot by comparing it with a control (Figure 76).
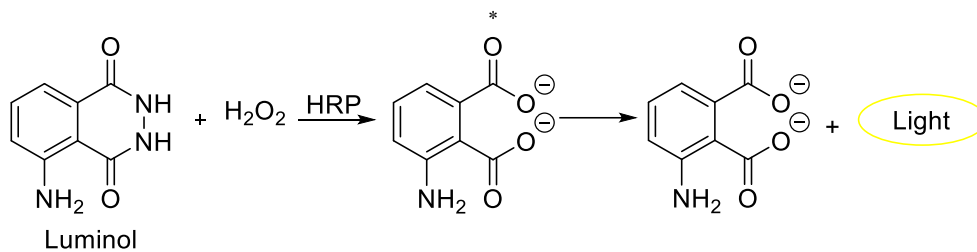
Figure 77. The chemiluminescence reaction in WB bioanalysis. Luminol is oxidized in the presence of HRP and hydrogen peroxide to form an excited state product (3-aminophthalate) that emits light at 425 as it decays to the ground state. Source: Alegria-Schaffer et al., 2009.

The concentrations used for each DPT$_H$ were 1, 10 100 μM. Observed inconsistencies include a) in Figures 73, 74, 75 & 76 the concentrations (in μM) seem to show increase (↑) stimulation instead as follows: (i): DPT1 **1** 1.00 ↑ pY - Akt, 10.00 ↑ of pY − Akt and pY − ERK 1 & 2, 100.00 ↑ pY − ERK 1 & 2, (ii): DPT2 **2** 1.00 ↑ pY − Akt, 10.00 ↑ pY − Akt, pY − ERK 1 & 2 & pY − EGFR, 100.00 ↑ pY − Akt and pY − ERK 1 & 2, (iii): DPT3 **3** 1.00 ↑ pY − Akt, pY − ERK 1 & 2and pY − EGFR, 10.00 ↑ pY − Akt & pY − ERK 1 & 2, 100.00 ↑ pY − Akt & pY - EGFR, (iv): DPT4 **4** 1.00 ↑ pY − Akt and pY − ERK 1 & 2, 10.00 ↑ pY − Akt & pY − EGFR, 100.00 ↑ pY − Akt and pY − ERK 1 & 2; b) the Total ERK 1 & 2 appeared fainter than pY − ERK 1 & 2 in Figure 74; and c) lanes appear disjointedly arranged in Figure 75 as if depicting a lane has been missed. However, the positive control or reference shown in Figure 76 (i.e., +EGF) indicates the results are analytically reliable despite the discrepancies observed. WB procedures are packed with protocols, not only hard to master but also prone to fail because it depends on the quality of the primary antibodies as well as whether polyclonal antibodies are used as the secondary antibodies, which may reveal non-

targeted proteins (as opposed to monoclonal antibodies) during development. WB is sensitive and specific, however, it can still misjudge results, a common feature inherent in all protocols, and the blotting itself that demands several repeat washings during and prior to ECL quantification can affect authenticity of the results. Moreover, ECL does produce light emission that fades exponentially over time, requires multiple exposures to capture optimal images, which can be subjective, especially when x-ray films are used to capture the resulting light emissions, and results can be hard to quantify, despite the several positive uses of ECL. Thus, when we merge these observations in WB and ECL we utilized for our work, some of the incongruities highlighted above may be explained [232, 233]. Therefore, the observations in the functional assays and immunoblotting results indicate there are positive conclusions to make.

5.3.3 Summary on bioevaluation

a) The colorimetric (MTS) analyzed gave clues on the potential activities of compounds assayed, viz, DPT1 **1** – DPT6 **6**, zm1 **7** (crude product) & zm2 **8**, all of which indicated a level of activity, supporting their next stage of bioevaluation. However, this was not to be due to time and chemical challenges particularly with zm1 **7** and zm1a **7**.

b) DPT1 **1** – DPT4 **4** are now lead candidates as shown by the WB results summary: DPT1 **1** 1.00, 10.00 & 100.00 µM showed strong pY − EGFR signals; DPT2 **2** 1.00 µM showed strong pY − EGFR signal but a weak pY − ERK 1 & 2 signal; DPT3 **3** 10.00 µM showed strong pY − EGFR signal but its 100.00 µM showed pY − ERK 1 & 2 signal and DPT4 **4** 1.00 µM

showed strong pY – EGFR signal but its 100.00 µM showed strong pY - EGFR signal. These compounds can be considered potentially active anti-oligomeric anti-EGFR new chemical entities

5.3.4 Future perspective

1) Zm1 **7**, zm1a **7**, and zm2 **8** should be synthesized and bioevaluated using simpler techniques such as ELISA or a repeat of the western blotting.

2) DPT5 **5** & DPT6 **6** can proceed to WB to establish their true RTKI-like activities.

3) DPT 1 **1** – DPT4 **4** are confirmed potentially active anti-EGFR anti-oligomeric lead candidates. We need to a) optimized these agents to obtain potentially more active agents, and b) the biological evaluation of these compounds should be expanded to include other analytical methods like HPLC–MS that chemically characterize their biological activities against the said target. Flow cytometry or ELISA as a simpler or a combination of bioassay techniques are employable.

4) We look forward to having a library or libraries of active anti-EGFR anti-oligomeric agents.

5.4 Conclusion

a) The consideration of other factors during the molecular modeling process, besides the Gscore status quo, seemed to assist effectively in DPT1 **1** – DPT4 **4** selection, despite their very low Gscores. The preliminary biolevaluation results show all the four have potential anti-EGFR anti-oligomeric activities.

b) The need to resynthesize compound (s) zm1 **7** (and zm1a **7**) is clear to commence their bioevaluation.

c) Zm2 **8** can be synthesized alongside its homologues for high-throughput bioevaluation screen since its colorimetric values indicated it as a strong potentially promising anti-EGFR anti-oligomer.

d) The need to expand biological and chemical processes on these compounds seem clear.

e) There are still more grids already generated and many compounds with 'good' DS that require exploitation.

f) The need for a more specific oligomerization region – targeted ligands is necessarily helpful, especially now that multimeric rather dimeric structures are taking center stage in biomolecular science.

h) The substantial number of docked compounds using different grids can still be explored.

References

1. Bailar, J.C. and H.L. Gornik, *Cancer undefeated.* New England Journal of Medicine, 1997. **336**(22): p. 1569-1574.
2. Jemal, A., et al., *Cancer statistics, 2008.* CA: a cancer journal for clinicians, 2008. **58**(2): p. 71-96.
3. Who, *World health statistics 2008*. 2008: World Health Organization.
4. Torre, L.A., et al., *Global cancer statistics, 2012.* CA: A Cancer Journal for Clinicians, 2015. **65**(2): p. 87-108.
5. Organization, W.H., *World cancer report, 2014.* WHO Report. Geneva: WHO, 2014.
6. Parkin, D.M., et al., *Cancer in Africa 2012.* Cancer Epidemiology Biomarkers & Prevention, 2014. **23**(6): p. 953-966.
7. Bray, F., et al., *Global cancer transitions according to the Human Development Index (2008–2030): a population-based study.* The lancet oncology, 2012. **13**(8): p. 790-801.
8. Ferlini, A. and S. Fini, *Genetic Heterogeneity and Human Disease.* Eur J Hum Genet, 2015. **23**(4): p. 559-559.
9. Nwosu, V., et al., *Heterogeneity of genetic alterations in prostate cancer: evidence of the complex nature of the disease.* Human molecular genetics, 2001. **10**(20): p. 2313-2318.
10. Kamińska, K., et al., *The role of the cell–cell interactions in cancer progression.* Journal of cellular and molecular medicine, 2015.
11. Tavazoie, S.F., et al., *Endogenous human microRNAs that suppress breast cancer metastasis.* nature, 2008. **451**(7175): p. 147-152.
12. Witz, I.P., *The tumor microenvironment: the making of a paradigm.* Cancer Microenvironment, 2009. **2**(1): p. 9-17.
13. Weir, H.K., et al., *The past, present, and future of cancer incidence in the United States: 1975 through 2020.* Cancer, 2015: p. n/a-n/a.
14. Leaf, C., *Why we're losing the war on cancer (and how to win it).* FORTUNE-EUROPEAN EDITION-, 2004. **149**(5): p. 42-55.
15. Carter, D., *New Global Survey Shows an Increasing Cancer Burden.* AJN The American Journal of Nursing, 2014. **114**(3): p. 17.
16. Pinto, G., A.A.M. Alhaiek, and J. Godovac-Zimmermann, *Proteomics reveals the importance of the dynamic redistribution of the subcellular location of proteins in breast cancer cells.* Expert review of proteomics, 2015(0): p. 1-14.
17. Wall, A.M. and D.J. Abraham, *Drug Resistance in Cancer Chemotherapy*, in *Burger's Medicinal Chemistry and Drug Discovery*. 2003, John Wiley & Sons, Inc.
18. Sharma, S.V., et al., *Epidermal growth factor receptor mutations in lung cancer.* Nature Reviews Cancer, 2007. **7**(3): p. 169-181.
19. Dodd, M.J., *Side effects of cancer chemotherapy.* Annual review of nursing research, 1993. **11**: p. 77-77.
20. Boyd, M.R. and K.D. Paull, *Some practical considerations and applications of the National Cancer Institute in vitro anticancer drug discovery screen.* Drug Development Research, 1995. **34**(2): p. 91-109.
21. Gibbs, J.B., *Mechanism-based target identification and drug discovery in cancer research.* Science, 2000. **287**(5460): p. 1969-1973.
22. Davies, D.M., et al., *Flexible targeting of ErbB dimers that drive tumorigenesis by using genetically engineered T cells.* Molecular Medicine, 2012. **18**(4): p. 565.

23.   Nicholson, R., J. Gee, and M. Harper, *EGFR and cancer prognosis.* European Journal of Cancer, 2001. **37**: p. 9-15.

24.   Mizuguchi, T., et al., *Evaluation of dimerization–inhibitory activities of cyclic peptides containing a β-hairpin loop sequence of the EGF receptor.* Bioorganic & medicinal chemistry, 2012. **20**(19): p. 5730-5737.

25.   Sequist, L.V., *The Anticipated Next Season of EGFR Inhibitors.* The oncologist, 2015. **20**(4): p. 335-336.

26.   Yarden, Y., et al., *Human proto-oncogene c-kit: a new cell surface receptor tyrosine kinase for an unidentified ligand.* The EMBO journal, 1987. **6**(11): p. 3341.

27.   Rothe, M., et al., *A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor.* Cell, 1994. **78**(4): p. 681-692.

28.   Mangelsdorf, D.J., et al., *The nuclear receptor superfamily: the second decade.* Cell, 1995. **83**(6): p. 835-839.

29.   Chawla, A., et al., *Nuclear receptors and lipid physiology: opening the X-files.* Science, 2001. **294**(5548): p. 1866-1870.

30.   Steck, T.L., *THE ORGANIZATION OF PROTEINS IN THE HUMAN RED BLOOD CELL MEMBRANE A Review.* The Journal of cell biology, 1974. **62**(1): p. 1-19.

31.   Higashiyama, S., et al., *Membrane-anchored growth factors, the epidermal growth factor family: Beyond receptor ligands.* Cancer Science, 2008. **99**(2): p. 214-220.

32.   Bar-Sagi, D. and J.R. Feramisco, *Induction of membrane ruffling and fluid-phase pinocytosis in quiescent fibroblasts by ras proteins.* Science, 1986. **233**(4768): p. 1061-1068.

33.   Basu, A., et al., *Inhibition of tyrosine kinase activity of the epidermal growth factor (EGF) receptor by a truncated receptor form that binds to EGF: role for interreceptor interaction in kinase regulation.* Molecular and cellular biology, 1989. **9**(2): p. 671-677.

34.   Tynan, C.J., et al., *A tale of the epidermal growth factor receptor: The quest for structural resolution on cells.* Methods, 2015.

35.   Downward, J., M.D. Waterfield, and P.J. Parker, *Autophosphorylation and protein kinase C phosphorylation of the epidermal growth factor receptor. Effect on tyrosine kinase activity and ligand binding affinity.* Journal of Biological Chemistry, 1985. **260**(27): p. 14538-14546.

36.   Needham, S.R., et al., *Structure-function relationships and supramolecular organization of the EGFR (epidermal growth factor receptor) on the cell surface.* Biochemical Society transactions, 2014. **42**(1): p. 114-119.

37.   Hackel, P.O., et al., *Epidermal growth factor receptors: critical mediators of multiple receptor pathways.* Current Opinion in Cell Biology, 1999. **11**(2): p. 184-189.

38.   Jensen, E.C., *The Basics of Western Blotting.* The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 2012. **295**(3): p. 369-371.

39.   Jorissen, R.N., et al., *Epidermal growth factor receptor: mechanisms of activation and signalling.* Experimental cell research, 2003. **284**(1): p. 31-53.

40.   Wakeling, A., et al., *Specific inhibition of epidermal growth factor receptor tyrosine kinase by 4-anilinoquinazolines.* Breast cancer research and treatment, 1996. **38**(1): p. 67-73.

41.   Vogelstein, B. and K.W. Kinzler, *Cancer genes and the pathways they control.* Nature medicine, 2004. **10**(8): p. 789-799.

42.   Suter, C.M., D.I. Martin, and R.L. Ward, *Germline epimutation of MLH1 in individuals with multiple cancers.* Nature genetics, 2004. **36**(5): p. 497-501.

43.   Pitot, H.C. and Y. Dragan, *Facts and theories concerning the mechanisms of carcinogenesis.* The FASEB journal, 1991. **5**(9): p. 2280-2286.

44.   Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-767.

45.   Jia, L.-T., et al., *Regulators of carcinogenesis: Emerging roles beyond their primary functions.* Cancer Letters, 2015. **357**(1): p. 75-82.

46.   Garrett, T.P., et al., *The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors.* Molecular cell, 2003. **11**(2): p. 495-505.

47.   Shi, F., et al., *ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation.* Proceedings of the National Academy of Sciences, 2010. **107**(17): p. 7692-7697.

48.   Bouyain, S., et al., *The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(42): p. 15024-15029.

49.   Mendelsohn, J. and J. Baselga, *The EGF receptor family as targets for cancer therapy.* Oncogene, 2000. **19**(56): p. 6550-6565.

50.   Baselga, J., *Why the epidermal growth factor receptor? The rationale for cancer therapy.* The oncologist, 2002. **7**(Supplement 4): p. 2-8.

51.   Grandis, J.R. and J.C. Sok, *Signaling through the epidermal growth factor receptor during the development of malignancy.* Pharmacology & therapeutics, 2004. **102**(1): p. 37-46.

52.   Yarden, Y., *The EGFR family and its ligands in human cancer: signalling mechanisms and therapeutic opportunities.* European journal of cancer, 2001. **37**: p. 3-8.

53.   Normanno, N., et al., *Epidermal growth factor receptor (EGFR) signaling in cancer.* Gene, 2006. **366**(1): p. 2-16.

54.   Babyatsky, M.W., G. Rossiter, and D.K. Podolsky, *Expression of transforming growth factors alpha and beta in colonic mucosa in inflammatory bowel disease.* Gastroenterology, 1996. **110**(4): p. 975-984.

55.   Choi, S., J. Mendrola, and M. Lemmon, *EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer.* Oncogene, 2007. **26**(11): p. 1567-1576.

56.   Alroy, I. and Y. Yarden, *The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions.* FEBS letters, 1997. **410**(1): p. 83-86.

57.   Holbro, T. and N.E. Hynes, *ErbB receptors: directing key signaling networks throughout life.* Annu. Rev. Pharmacol. Toxicol., 2004. **44**: p. 195-217.

58.   Harari, P., *Epidermal growth factor receptor inhibition strategies in oncology.* Endocrine-Related Cancer, 2004. **11**(4): p. 689-708.

59.   Burgess, A.W., *EGFR family: Structure physiology signalling and therapeutic targets The EGFR family members regulate the production, survival, movement, and shape of many embryonic and adult cell types. Signalling from three of the family members EGFR, erbB3 and erbB4) are regulated by ligand dependent oligomerization and fourth, erbB2, by oligomerization with ligated forms of the other family members. Three of the family members (EGFR, erbB2 and erbB4) signal via activated tyrosine kinase scaffolds, the fourth, erbB3, by ligand induced oligomerization and activation of the other family members.* Growth Factors, 2008. **26**(5): p. 263-274.

60.   Ghosh, S., et al., *Rational design of potent and selective EGFR tyrosine kinase inhibitors as anticancer agents.* Current cancer drug targets, 2001. **1**(2): p. 129-140.

61.   Kosaka, T., et al., *Mechanisms of resistance to EGFR TKIs and development of a new generation of drugs in non-small-cell lung cancer.* BioMed Research International, 2011. **2011**.

62. Kawamoto, T., et al., *Growth stimulation of A431 cells by epidermal growth factor: identification of high-affinity receptors for epidermal growth factor by an anti-receptor monoclonal antibody.* Proceedings of the National Academy of Sciences, 1983. **80**(5): p. 1337-1341.

63. Yewale, C., et al., *Epidermal growth factor receptor targeting in cancer: A review of trends and strategies.* Biomaterials, 2013. **34**(34): p. 8690-8707.

64. Li, Z. and P. Srivastava, *Heat-Shock Proteins.* Current Protocols in Immunology, 2004: p. A. 1T. 1-A. 1T. 6.

65. Rouse, J., et al., *A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins.* Cell, 1994. **78**(6): p. 1027-1037.

66. Georgopoulos, C. and W. Welch, *Role of the major heat shock proteins as molecular chaperones.* Annual review of cell biology, 1993. **9**(1): p. 601-634.

67. Marcu, M.G., T.W. Schulte, and L. Neckers, *Novobiocin and related coumarins and depletion of heat shock protein 90-dependent signaling proteins.* Journal of the National Cancer Institute, 2000. **92**(3): p. 242-248.

68. Sreedhar, A.S., C. So, and P. Csermely, *Inhibition of Hsp90: a new strategy for inhibiting protein kinases.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2004. **1697**(1): p. 233-242.

69. Richter, P. and E. Fahr, *Synthesis of trans/syn-and trans/anti-Dimeric Uracil.* Angewandte Chemie International Edition in English, 1969. **8**(3): p. 208-209.

70. Hadden, M. and B. Blagg, *Dimeric approaches to anti-cancer chemotherapeutics.* Anti-cancer agents in medicinal chemistry, 2008. **8**(7): p. 807.

71. Starok, M., et al., *EGFR Inhibition by Curcumin in Cancer Cells: a Dual Mode of Action.* Biomacromolecules, 2015.

72. Adams, B.K., et al., *Synthesis and biological evaluation of novel curcumin analogs as anti-cancer and anti-angiogenesis agents.* Bioorganic & Medicinal Chemistry, 2004. **12**(14): p. 3871-3883.

73. Goodchild, J., *Therapeutic oligonucleotides*, in *Therapeutic Oligonucleotides*. 2011, Springer. p. 1-15.

74. Bischoff, A., et al., *A global microRNA screen identifies regulators of the ErbB receptor signaling network.* Cell Communication and Signaling, 2015. **13**(1): p. 5.

75. Morcos, P.A., *Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos.* Biochemical and biophysical research communications, 2007. **358**(2): p. 521-527.

76. Uhlmann, E. and A. Peyman, *Antisense oligonucleotides: a new therapeutic principle.* Chemical Reviews, 1990. **90**(4): p. 543-584.

77. Mani, S., et al., *Clinical Studies in Patients with Solid Tumors using a Second-Generation Antisense Oligonucleotide (GEM® 231) Targeted against Protein Kinase A Type I.* Annals of the New York Academy of Sciences, 2003. **1002**(1): p. 252-262.

78. Ciardiello, F., et al., *Cooperative inhibition of renal cancer growth by anti-epidermal growth factor receptor antibody and protein kinase A antisense oligonucleotide.* Journal of the National Cancer Institute, 1998. **90**(14): p. 1087-1998.

79. Su, J.-L., et al., *The VEGF-C/Flt-4 axis promotes invasion and metastasis of cancer cells.* Cancer cell, 2006. **9**(3): p. 209-223.

80. Ciardiello, F., et al., *Antitumor effect and potentiation of cytotoxic drugs activity in human cancer cells by ZD-1839 (Iressa), an epidermal growth factor receptor-selective tyrosine kinase inhibitor.* Clinical Cancer Research, 2000. **6**(5): p. 2053-2063.

81. Zhu, L., et al., *B-cell epitope peptide vaccination targeting dimer interface of epidermal growth factor receptor (EGFR).* Immunology letters, 2013. **153**(1): p. 33-40.

82. Sato, J., et al., *Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors.* Molecular biology & medicine, 1983. **1**(5): p. 511-529.

83. Senter, P.D., *Potent antibody drug conjugates for cancer therapy.* Current opinion in chemical biology, 2009. **13**(3): p. 235-244.

84. Leriche, G., L. Chisholm, and A. Wagner, *Cleavable linkers in chemical biology.* Bioorganic & medicinal chemistry, 2012. **20**(2): p. 571-582.

85. Ducry, L. and B. Stump, *Antibody– drug conjugates: linking cytotoxic payloads to monoclonal antibodies.* Bioconjugate chemistry, 2009. **21**(1): p. 5-13.

86. Smaglo, B.G., D. Aldeghaither, and L.M. Weiner, *The development of immunoconjugates for targeted cancer therapy.* Nat Rev Clin Oncol, 2014. **11**(11): p. 637-648.

87. Yang, R.Y., et al., *Targeting the Dimerization of Epidermal Growth Factor Receptors with Small-Molecule Inhibitors.* Chemical biology & drug design, 2010. **76**(1): p. 1-9.

88. Petch, D., et al., *Design and synthesis of EGFR dimerization inhibitors and evaluation of their potential in the treatment of psoriasis.* Bioorganic & medicinal chemistry, 2012. **20**(19): p. 5901-5914.

89. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angewandte Chemie International Edition, 2005. **44**(45): p. 7342-7372.

90. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nat Rev Drug Discov, 2002. **1**(4): p. 309-15.

91. Hunter, T., *Signaling—2000 and beyond.* Cell, 2000. **100**(1): p. 113-127.

92. Muthuswamy, S.K., M. Gilman, and J.S. Brugge, *Controlled dimerization of ErbB receptors provides evidence for differential signaling by homo-and heterodimers.* Molecular and cellular biology, 1999. **19**(10): p. 6845-6857.

93. Ceresa, B.P., *Regulation of EGFR endocytic trafficking by rab proteins.* 2006.

94. Marianayagam, N.J., M. Sunde, and J.M. Matthews, *The power of two: protein dimerization in biology.* Trends in biochemical sciences, 2004. **29**(11): p. 618-625.

95. Blackwell, H.E., P.A. Clemons, and S.L. Schreiber, *Exploiting site-site interactions on solid support to generate dimeric molecules.* Organic letters, 2001. **3**(8): p. 1185-1188.

96. Al-Bassam, J., et al., *Distinct conformations of the kinesin Unc104 neck regulate a monomer to dimer motor transition.* The Journal of cell biology, 2003. **163**(4): p. 743-753.

97. Nooren, I. and J.M. Thornton, *Diversity of protein–protein interactions.* The EMBO journal, 2003. **22**(14): p. 3486-3492.

98. Lemmon, M.A., et al., *Two EGF molecules contribute additively to stabilization of the EGFR dimer.* The EMBO Journal, 1997. **16**(2): p. 281-294.

99. Ferguson, K.M., et al., *EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization.* Molecular cell, 2003. **11**(2): p. 507-517.

100. Kosmidis, E.K., et al., *Functional Aspects of the EGF-Induced MAP Kinase Cascade: A Complex Self-Organizing System Approach.* PloS one, 2014. **9**(11): p. e111612.

101. Ogiso, H., et al., *Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains.* Cell, 2002. **110**(6): p. 775-787.

102. Gadella, T. and T.M. Jovin, *Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation.* The Journal of cell biology, 1995. **129**(6): p. 1543-1558.

103. Klein, P., et al., *A structure-based model for ligand binding and dimerization of EGF receptors.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(4): p. 929-934.

104. Olayioye, M.A., et al., *ErbB receptor-induced activation of stat transcription factors is mediated by Src tyrosine kinases.* Journal of Biological Chemistry, 1999. **274**(24): p. 17209-17218.

105. Bishayee, S., et al., *Ligand-induced dimerization of the platelet-derived growth factor receptor. Monomer-dimer interconversion occurs independent of receptor phosphorylation.* J Biol Chem, 1989. **264**(20): p. 11699-705.

106. Clayton, A.H., et al., *Ligand-induced dimer-tetramer transition during the activation of the cell surface epidermal growth factor receptor-A multidimensional microscopy analysis.* Journal of Biological Chemistry, 2005. **280**(34): p. 30392-30399.

107. Yarden, Y. and J. Schlessinger, *Self-phosphorylation of epidermal growth factor receptor: evidence for a model of intermolecular allosteric activation.* Biochemistry, 1987. **26**(5): p. 1434-1442.

108. Martin-Fernandez, M., et al., *Preformed oligomeric epidermal growth factor receptors undergo an ectodomain structure change during signaling.* Biophysical journal, 2002. **82**(5): p. 2415-2427.

109. Gill, G.N., P.J. Bertics, and J.B. Santon, *Epidermal growth factor and its receptor.* Molecular and cellular endocrinology, 1987. **51**(3): p. 169-186.

110. Honegger, A., et al., *Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells.* Molecular and cellular biology, 1990. **10**(8): p. 4035-4044.

111. Baselga, J., *Targeting tyrosine kinases in cancer: the second wave.* Science, 2006. **312**(5777): p. 1175-1178.

112. King, A.C. and P. Cuatrecasas, *Exposure of cells to an acidic environment reverses the inhibiton by methylamine of the mitogenic response to epidermal growth factor.* Biochemical and biophysical research communications, 1982. **106**(2): p. 479-485.

113. Monast, C.S., N. Mehta, and M.J. Lazzara, *Diversity in Dimerization Topologies Enables Differential Control of Receptor Tyrosine Kinase Phosphorylation Dynamics.* Cellular and Molecular Bioengineering, 2014. **7**(1): p. 86-98.

114. Seshacharyulu, P., et al., *Targeting the EGFR signaling pathway in cancer therapy.* Expert opinion on therapeutic targets, 2012. **16**(1): p. 15-31.

115. Yarden, Y. and G. Pines, *The ERBB network: at last, cancer therapy meets systems biology.* Nat Rev Cancer, 2012. **12**(8): p. 553-63.

116. Ferguson, K.M., *A structure-based view of Epidermal Growth Factor Receptor regulation.* Annual review of biophysics, 2008. **37**: p. 353.

117. Yarden, Y. and M.X. Sliwkowski, *Untangling the ErbB signalling network.* Nature reviews Molecular cell biology, 2001. **2**(2): p. 127-137.

118. Furuuchi, K., et al., *Targeted antireceptor therapy with monoclonal antibodies leads to the formation of inactivated tetrameric forms of ErbB receptors.* The Journal of Immunology, 2007. **178**(2): p. 1021-1029.

119. Pinkas-Kramarski, R., et al., *ErbB tyrosine kinases and the two neuregulin families constitute a ligand-receptor network.* Molecular and cellular biology, 1998. **18**(10): p. 6090-6101.

120. Tzahar, E., et al., *Bivalence of EGF-like ligands drives the ErbB signaling network.* The EMBO Journal, 1997. **16**(16): p. 4938-4950.

121. Kozer, N., et al., *Recruitment of the adaptor protein Grb2 to EGFR tetramers.* Biochemistry, 2014. **53**(16): p. 2594-2604.

122. Graus-Porta, D., et al., *ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling.* The EMBO journal, 1997. **16**(7): p. 1647-1655.

123. Brown, P.M., et al., *THE EXTRACELLULAR DOMAIN OF THE EPIDERMAL GROWTH-FACTOR RECEPTOR - STUDIES ON THE AFFINITY AND STOICHIOMETRY OF BINDING, RECEPTOR DIMERIZATION AND A BINDING-DOMAIN MUTANT.* European Journal of Biochemistry, 1994. **225**(1): p. 223-233.

124. Hurwitz, D.R., et al., *EGF INDUCES INCREASED LIGAND-BINDING AFFINITY AND DIMERIZATION OF SOLUBLE EPIDERMAL GROWTH-FACTOR (EGF) RECEPTOR EXTRACELLULAR DOMAIN.* Journal of Biological Chemistry, 1991. **266**(32): p. 22035-22043.

125. Ali, M.H. and B. Imperiali, *Protein oligomerization: how and why.* Bioorganic & medicinal chemistry, 2005. **13**(17): p. 5013-5020.

126. Matthews, J.M. and M. Sunde, *Dimers, oligomers, everywhere.* Adv Exp Med Biol, 2012. **747**: p. 1-18.

127. Fricke, F., M.S. Dietz, and M. Heilemann, *Single-Molecule Methods to Study Membrane Receptor Oligomerization.* ChemPhysChem, 2014.

128. Newton, G.J., et al., *Regulated Shedding of EGFR Ligands is Controlled by Oligomerization*, in *45th Annual Meeting of the American Society for Cell Biology*. 2005.

129. Axelrod, D., et al., *Mobility measurement by analysis of fluorescence photobleaching recovery kinetics.* Biophysical journal, 1976. **16**(9): p. 1055.

130. Hofman, E.G., et al., *Ligand-induced EGF Receptor Oligomerization Is Kinase-dependent and Enhances Internalization.* Journal of Biological Chemistry, 2010. **285**(50): p. 39481-39489.

131. Főrster, T., *10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation.* Discussions of the Faraday Society, 1959. **27**: p. 7-17.

132. Defize, L., et al., *Signal transduction by epidermal growth factor occurs through the subclass of high affinity receptors.* The Journal of cell biology, 1989. **109**(5): p. 2495-2507.

133. Gamett, D.C., et al., *Secondary dimerization between members of the epidermal growth factor receptor family.* Journal of Biological Chemistry, 1997. **272**(18): p. 12052-12056.

134. Nagy, P., et al., *Activation-dependent clustering of the erbB2 receptor tyrosine kinase detected by scanning near-field optical microscopy.* Journal of cell science, 1999. **112**(11): p. 1733-1741.

135. Park, E., R. Baron, and R. Landgraf, *Higher-Order Association States of Cellular ERBB3 Probed with Photo-Cross-Linkable Aptamers†.* Biochemistry, 2008. **47**(46): p. 11992-12005.

136. Schlessinger, J., *Cell signaling by receptor tyrosine kinases.* Cell, 2000. **103**(2): p. 211-225.

137. Yi, J.Y., I. Han, and E.-S. Oh, *Transmembrane Domain Dependent Functional Oligomerization of Syndecans.* The Scientific World Journal, 2006. **6**: p. 457-459.

138. Bardin, C., et al., *Therapeutic drug monitoring in cancer – Are we missing a trick?* European Journal of Cancer, 2014. **50**(12): p. 2005-2009.

139. Bessman, N.J. and M.A. Lemmon, *Finding the missing links in EGFR.* Nat Struct Mol Biol, 2012. **19**(1): p. 1-3.

140. Clayton, A.H., et al., *Predominance of activated EGFR higher-order oligomers on the cell surface.* Growth Factors, 2008. **26**(6): p. 316-324.

141. Kozer, N., et al., *Evidence for extended YFP-EGFR dimers in the absence of ligand on the surface of living cells.* Physical biology, 2011. **8**(6): p. 066002.

142. Kozer, N., et al., *Exploring higher-order EGFR oligomerisation and phosphorylation—a combined experimental and theoretical approach.* Molecular BioSystems, 2013. **9**(7): p. 1849-1863.

143.	Yamashita, H., et al., *Oligomerization–function relationship of EGFR on living cells detected by the coiled-coil labeling and FRET microscopy.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2015. **1848**(6): p. 1359-1366.

144.	Zanetti-Domingues, L.C., et al., *Determining the geometry of oligomers of the human epidermal growth factor family on cells with 7 nm resolution.* Progress in biophysics and molecular biology, 2015.

145.	Becker, J., et al., *Role of receptor tyrosine kinases in gastric cancer: new targets for a selective therapy.* World journal of gastroenterology: WJG, 2006. **12**(21): p. 3297-3305.

146.	Ghobrial, I.M., T.E. Witzig, and A.A. Adjei, *Targeting apoptosis pathways in cancer therapy.* CA: A Cancer Journal for Clinicians, 2005. **55**(3): p. 178-194.

147.	Li, N., et al., *Inhibition of cell proliferation by an anti-EGFR aptamer.* PLoS One, 2011. **6**(6): p. e20299.

148.	GULLICK, W.J., et al., *Antibodies to the ATP-binding site of the human epidermal growth factor (EGF) receptor as specific inhibitors of EGF-stimulated protein-tyrosine kinase activity.* European Journal of Biochemistry, 1986. **158**(2): p. 245-253.

149.	Schrodinger, L., *Schrodinger Software Suite.* New York: Schrödinger, LLC, 2011.

150.	Yun, C.H., et al., *Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity.* Cancer Cell, 2007. **11**(3): p. 217-27.

151.	Kozer, N., et al., *Recruitment of the Adaptor Protein Grb2 to EGFR Tetramers.* Biochemistry, 2014. **53**(16): p. 2594-2604.

152.	Huang, H.-J., et al., *Current developments of computer-aided drug design.* Journal of the Taiwan Institute of Chemical Engineers, 2010. **41**(6): p. 623-635.

153.	Bacilieri, M. and S. Moro, *Ligand-based drug design methodologies in drug discovery process: an overview.* Current drug discovery technologies, 2006. **3**(3): p. 155-165.

154.	Blundell, T.L., *Structure-based drug design.* Nature, 1996. **384**(6604 Suppl): p. 23-26.

155.	Moro, S., et al., *Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: pyrazolo [4, 3-e] 1, 2, 4-triazolo [1, 5-c] pyrimidine derivatives as a key study.* Journal of medicinal chemistry, 2005. **48**(1): p. 152-162.

156.	Cohen, C., O. Fischel, and E. Cohen, *Molecular ConceptorTM for Training in Medicinal Chemistry, Drug Design, and Cheminformatics.* Chemical Biology & Drug Design, 2007. **69**(1): p. 75-82.

157.	Ferreira, L.G., et al., *Molecular docking and structure-based drug design strategies.* Molecules, 2015. **20**(7): p. 13384-13421.

158.	Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nature reviews Drug discovery, 2004. **3**(11): p. 935-949.

159.	Mohan, V., et al., *Docking: successes and challenges.* Current pharmaceutical design, 2005. **11**(3): p. 323-333.

160.	Prenzel, N., et al., *The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification.* Endocrine-related cancer, 2001. **8**(1): p. 11-31.

161.	Garrett, T.P., et al., *Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α.* Cell, 2002. **110**(6): p. 763-773.

162.	Friesner, R.A., et al., *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of medicinal chemistry, 2004. **47**(7): p. 1739-1749.

163. Eck, M.J. and C.-H. Yun, *Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2010. **1804**(3): p. 559-566.

164. Bologa, C.G., M.M. Olah, and T.I. Oprea, *Chemical database preparation for compound acquisition or virtual screening*, in *Bioinformatics and Drug Discovery*. 2006, Springer. p. 375-388.

165. Kalyaanamoorthy, S. and Y.-P.P. Chen, *Structure-based drug design to augment hit discovery.* Drug Discovery Today, 2011. **16**(17–18): p. 831-839.

166. Baurin, N., et al., *Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds.* Journal of chemical information and computer sciences, 2004. **44**(2): p. 643-651.

167. Ghose, A.K., V.N. Viswanadhan, and J.J. Wendoloski, *A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases.* Journal of combinatorial chemistry, 1999. **1**(1): p. 55-68.

168. Zuegg, J. and M. A Cooper, *Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening.* Current topics in medicinal chemistry, 2012. **12**(14): p. 1500-1513.

169. Muegge, I., S.L. Heald, and D. Brittelli, *Simple selection criteria for drug-like chemical matter.* Journal of medicinal chemistry, 2001. **44**(12): p. 1841-1846.

170. Veber, D.F., et al., *Molecular properties that influence the oral bioavailability of drug candidates.* Journal of medicinal chemistry, 2002. **45**(12): p. 2615-2623.

171. Friesner, R.A., et al., *Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes.* Journal of medicinal chemistry, 2006. **49**(21): p. 6177-6196.

172. Adeniyi, A.A. and P.A. Ajibade, *Comparing the suitability of autodock, gold and glide for the docking and predicting the possible targets of ru (ii)-based complexes as anticancer agents.* Molecules, 2013. **18**(4): p. 3760-3778.

173. Kellenberger, E., et al., *Comparative evaluation of eight docking tools for docking and virtual screening accuracy.* Proteins, 2004. **57**(2): p. 225-42.

174. Schneider, G., *Virtual screening: an endless staircase?* Nature Reviews Drug Discovery, 2010. **9**(4): p. 273-276.

175. Weber, L., *JChem Base-ChemAxon*. 2008, ROYAL SOC CHEMISTRY THOMAS GRAHAM HOUSE, SCIENCE PARK, MILTON RD, CAMBRIDGE CB4 0WF, CAMBS, ENGLAND. p. 65-66.

176. Singh, H., et al., *QSAR based model for discriminating EGFR inhibitors and non-inhibitors using Random forest.* Biology direct, 2015. **10**(1): p. 10.

177. Wang, D.D., et al., *Personalized prediction of EGFR mutation-induced drug resistance in lung cancer.* Scientific reports, 2013. **3**.

178. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5-32.

179. Qi, Y., et al., *Systematic prediction of human membrane receptor interactions.* Proteomics, 2009. **9**(23): p. 5243-5255.

180. Wiles, D. and T. Suprunchuk, *Carbodihydrazones and oxalyl dihydrazones.* Canadian Journal of Chemistry, 1968. **46**(5): p. 701-705.

181. Szmuszkovicz, J. and M.E. Greig, *A Study of the Inhibitory Effect of Various Hydrazides on Monoamine Oxidase in vitro and in vivo.* Journal of Medicinal Chemistry, 1961. **4**(2): p. 259-296.

182. Umesha, K., K. Rai, and M.H. Nayaka, *Antioxidant and Antimicrobial Activity of 5-methyl-2-(5-methyl-1, 3-diphenyl-1H-pyrazole-4-carbonyl)-2, 4-dihydro-pyrazol-3-one.* International journal of biomedical science: IJBS, 2009. **5**(4): p. 359.

183. Wu, D., et al., *Synthesis, Structure–Activity Relationship, and Pharmacophore Modeling Studies of Pyrazole-3-Carbohydrazone Derivatives as Dipeptidyl Peptidase IV Inhibitors.* Chemical biology & drug design, 2012. **79**(6): p. 897-906.

184. Katrizky, A.R., P. Barczynski, and D.L. Ostercamp, *Mechanisms of heterocycle ring formation. Part 5. A carbon-13 nuclear magnetic resonance study of pyrazolinone synthesis by the reaction of β-ketoesters with substituted hydrazines.* Journal of the Chemical Society, Perkin Transactions 2, 1987(8): p. 969-975.

185. Semple, G., et al., *3-(1 H-tetrazol-5-yl)-1, 4, 5, 6-tetrahydro-cyclopentapyrazole (MK-0354): a partial agonist of the nicotinic acid receptor, G-protein coupled receptor 109a, with antilipolytic but no vasodilatory activity in mice.* Journal of medicinal chemistry, 2008. **51**(16): p. 5101-5108.

186. Ley, J.P. and H.-J. Bertram, *Hydroxy-or methoxy-substituted benzaldoximes and benzaldehyde-O-alkyloximes as tyrosinase inhibitors.* Bioorganic & medicinal chemistry, 2001. **9**(7): p. 1879-1885.

187. Ley, J.P., et al., *New bitter-masking compounds: hydroxylated benzoic acid amides of aromatic amines as structural analogues of homoeriodictyol.* Journal of agricultural and food chemistry, 2006. **54**(22): p. 8574-8579.

188. Wilson, J., *Phenolic Analogs of Amino Carboxylic Acid Ligands for 99mTc. II. Synthesis and Characterization of N, N'-Ethylenebis [N-(o-Hydroxybenzyl)-glycines](ehbg).* Australian Journal of Chemistry, 1988. **41**(2): p. 173-182.

189. Meyers, H., et al., *Multiple simultaneous synthesis of phenolic libraries.* Molecular Diversity, 1995. **1**(1): p. 13-20.

190. Montalbetti, C.A.G.N. and V. Falque, *Amide bond formation and peptide coupling.* Tetrahedron, 2005. **61**(46): p. 10827-10852.

191. Bart, J., et al., *Room-temperature intermediate layer bonding for microfluidic devices.* Lab on a Chip, 2009. **9**(24): p. 3481-3488.

192. Barltrop, J.A., et al., *5-(3-carboxymethoxyphenyl)-2-(4, 5-dimethylthiazolyl)-3-(4-sulfophenyl) tetrazolium, inner salt (MTS) and related analogs of 3-(4, 5-dimethylthiazolyl)-2, 5-diphenyltetrazolium bromide (MTT) reducing to purple water-soluble formazans as cell-viability indicators.* Bioorganic & Medicinal Chemistry Letters, 1991. **1**(11): p. 611-614.

193. Vistica, D.T., et al., *Tetrazolium-based assays for cellular viability: a critical examination of selected parameters affecting formazan production.* Cancer Res, 1991. **51**(10): p. 2515-20.

194. Shoemaker, R.H., *The NCI60 human tumour cell line anticancer drug screen.* Nat Rev Cancer, 2006. **6**(10): p. 813-823.

195. Boyd, M., K. Paull, and L. Rubinstein, *Data display and analysis strategies for the NCI disease-oriented in vitro antitumor drug screen*, in *Cytotoxic Anticancer Drugs: Models and Concepts for Drug Discovery and Development*. 1992, Springer. p. 11-34.

196. Rubinstein, L.V., et al., *Comparison of in vitro anticancer-drug-screening data generated with a tetrazolium assay versus a protein assay against a diverse panel of human tumor cell lines.* J Natl Cancer Inst, 1990. **82**(13): p. 1113-8.

197. Monks, A., et al., *Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines.* J Natl Cancer Inst, 1991. **83**(11): p. 757-66.

198. Anderson, W.K., A. Gopalsamy, and P.S. Reddy, *Design, synthesis, and study of 9-Substituted ellipticine and 2-Methylellipticinium analogs as potential CNS-selective antitumor agents.* Journal of medicinal chemistry, 1994. **37**(13): p. 1955-1963.

199. Neubig, R.R., et al., *International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification. XXXVIII. Update on terms and symbols in quantitative pharmacology.* Pharmacological Reviews, 2003. **55**(4): p. 597-606.

200.   Pawson, C.T. and J.D. Scott, *Signal integration through blending, bolstering and bifurcating of intracellular information.* Nature structural & molecular biology, 2010. **17**(6): p. 653-658.

201.   Song, G., G. Ouyang, and S. Bao, *The activation of Akt/PKB signaling pathway and cell survival.* Journal of cellular and molecular medicine, 2005. **9**(1): p. 59-71.

202.   Mebratu, Y. and Y. Tesfaigzi, *How ERK1/2 activation controls cell proliferation and cell death: Is subcellular localization the answer?* Cell cycle, 2009. **8**(8): p. 1168-1175.

203.   Lowry, O.H., et al., *Protein measurement with the Folin phenol reagent.* J biol Chem, 1951. **193**(1): p. 265-275.

204.   Bramhall, S., et al., *A simple colorimetric method for determination of protein.* Analytical biochemistry, 1969. **31**: p. 146-148.

205.   Burnette, W.N., *"Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A.* Analytical biochemistry, 1981. **112**(2): p. 195-203.

206.   Derer, S., et al., *Impact of epidermal growth factor receptor (EGFR) cell surface expression levels on effector mechanisms of EGFR antibodies.* The Journal of Immunology, 2012. **189**(11): p. 5230-5239.

207.   Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Analytical biochemistry, 1976. **72**(1): p. 248-254.

208.   Recanatini, M., et al., *QT prolongation through hERG K+ channel blockade: Current knowledge and strategies for the early prediction during drug development.* Medicinal Research Reviews, 2005. **25**(2): p. 133-166.

209.   Ciaccio, M.F., et al., *Systems analysis of EGF receptor signaling dynamics with microwestern arrays.* Nat Meth, 2010. **7**(2): p. 148-155.

210.   Liu, Z.-Q., T. Mahmood, and P.-C. Yang, *Western blot: Technique, theory and trouble shooting.* North American journal of medical sciences, 2014. **6**(3): p. 160.

211.   Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nature methods, 2012. **9**(7): p. 671-675.

212.   Rakowicz-Szulczynska, E.M., et al., *Epidermal growth factor (EGF) and monoclonal antibody to cell surface EGF receptor bind to the same chromatin receptor.* Archives of biochemistry and biophysics, 1989. **268**(2): p. 456-464.

213.   Schaefer, G., et al., *A two-in-one antibody against HER3 and EGFR has superior inhibitory activity compared with monospecific antibodies.* Cancer cell, 2011. **20**(4): p. 472-486.

214.   Leach, A.R., B.K. Shoichet, and C.E. Peishoff, *Prediction of protein-ligand interactions. Docking and scoring: successes and gaps.* Journal of medicinal chemistry, 2006. **49**(20): p. 5851-5855.

215.   Lengauer, T. and M. Rarey, *Computational methods for biomolecular docking.* Current Opinion in Structural Biology, 1996. **6**(3): p. 402-406.

216.   Cohen, N.C., et al., *Molecular modeling software and methods for medicinal chemistry.* Journal of medicinal chemistry, 1990. **33**(3): p. 883-894.

217.   Blaney, J.M. and J.S. Dixon, *A good ligand is hard to find: Automated docking methods.* Perspectives in Drug Discovery and Design, 1993. **1**(2): p. 301-319.

218.   Webb, S.E.D., et al., *Nanometric molecular separation measurements by single molecule photobleaching.* Methods, 2015. **88**: p. 76-80.

219.   Ultra, C., *12.0 Suite, Perkin Elmer*. 2012.

220.   Sunderhaus, J.D. and S.F. Martin, *Applications of multicomponent reactions to the synthesis of diverse heterocyclic scaffolds.* Chemistry-A European Journal, 2009. **15**(6): p. 1300-1308.

221.   Brussel, W. and C. Van Sumere, *N-Acylamino acids and peptides VI. A Simple synthesis of N-acylglycines of the benzoyl-and cinnamyl type.* Bulletin des Sociétés Chimiques Belges, 1978. **87**(10): p. 791-797.

222.   Han, S.-Y. and Y.-A. Kim, *Recent development of peptide coupling reagents in organic synthesis.* Tetrahedron, 2004. **60**(11): p. 2447-2467.

223.   Wang, S., et al., *Enhancement of peptide coupling reactions by 4-dimethylaminopyridine.* International journal of peptide and protein research, 1981. **18**(5): p. 459-467.

224.   Huang, K.T., Y.H. Chen, and A.M. Walker, *Inaccuracies in MTS assays: major distorting effects of medium, serum albumin, and fatty acids.* Biotechniques, 2004. **37**(3): p. 406, 408, 410-2.

225.   Geenen, S., et al., *HPLC–MS/MS methods for the quantitative analysis of 5-oxoproline (pyroglutamate) in rat plasma and hepatic cell line culture medium.* Journal of pharmaceutical and biomedical analysis, 2011. **56**(3): p. 655-663.

226.   Wang, P., S.M. Henning, and D. Heber, *Limitations of MTT and MTS-based assays for measurement of antiproliferative activity of green tea polyphenols.* PloS one, 2010. **5**(4): p. e10202.

227.   Demetrashvili, N. and E.R. Van den Heuvel, *Confidence intervals for intraclass correlation coefficients in a nonlinear dose–response meta-analysis.* Biometrics, 2015.

228.   Gewirtz, D.A., M.L. Bristol, and J.C. Yalowich, *Toxicity issues in cancer drug development.* Curr Opin Investig Drugs, 2010. **11**(6): p. 612-4.

229.   Collins, J., *Prospective predictions and validations in anticancer therapy.* Pharmacokinetics in risk assessment: drinking water and health, 1987. **8**: p. 431-440.

230.   Benfield, P.A. and B.D. Car, *Pharmacokinetic and Toxicology Issues in Cancer Drug Discovery and Development*, in *Molecular Cancer Therapeutics*. 2005, John Wiley & Sons, Inc. p. 255-286.

231.   Kurien, B.T. and R.H. Scofield, *Western blotting.* Methods, 2006. **38**(4): p. 283-293.

232.   MacPhee, D.J., *Methodological considerations for improving Western blot analysis.* Journal of pharmacological and toxicological methods, 2010. **61**(2): p. 171-177.

233.   Alegria-Schaffer, A., A. Lodge, and K. Vattem, *Performing and optimizing Western blots with an emphasis on chemiluminescent detection.* Methods in enzymology, 2009. **463**: p. 573-599.

Appendix

Appendix A – Selected lead candidates in colours

Ambinter_										
Ligand	glide Gscore	glide energy	glide einternal	glide emodel	Prime Energy	MMGBSA dG Bind 0	Ligand Energy	Receptor Energy	Complex Energy	glide posenum
ZINC 72131821	-3.824	-24.203	1.266	-28.848	-23621.7	-50.486	-105.093	-23466.1	-23621.7	15
ZINC20 537922	-3.53	-27.684	3.837	-36.075	-23515.1	-48.796	-0.15	-23466.1	-23515.1	3
ZINC20 322301	-3.346	-27.477	5.886	-34.332	-23463.2	-47.496	50.474	-23466.1	-23463.2	12
ZINC72 131820	-3.301	-22.551	0.133	-30.292	-23613.6	-43.26	-104.215	-23466.1	-23613.6	17
ZINC67 847695	-3.099	-28.944	7.746	-33.793	-23481.9	-52.455	36.688	-23466.1	-23481.9	4
ZINC67 847694	-2.798	-21.11	0.464	-26.499	-23602.3	-45.114	-91.024	-23466.1	-23602.3	20
ZINC65 372018	-2.735	-22.328	1.216	-28.606	-23580.1	-37.862	-76.134	-23466.1	-23580.1	1
ZINC71 771450	-2.657	-22.547	1.221	-25.633	-23474.8	-38.787	30.099	-23466.1	-23474.8	2
ZINC67 741022	-2.608	-24.284	6.594	-28.105	-23503.4	-45.307	7.988	-23466.1	-23503.4	6
ZINC67 955954	-2.301	-25.522	3.836	-29.083	-23530.2	-42.784	-21.305	-23466.1	-23530.2	4
ZINC67 851668	-2.07	-16.432	6.243	-18.387	-23499.5	-26.138	-7.219	-23466.1	-23499.5	53
ZINC72 457948	-2.067	-16.656	3.45	-21.841	-23497.8	-43.59	11.898	-23466.1	-23497.8	26
ZINC67 774117	-2.027	-19.806	0.567	-20.478	-23445.1	-40.967	61.956	-23466.1	-23445.1	37

ZINC02887281	-1.962	-26.228	1.053	-33.118	-23501.3	-39.033	3.889	-23466.1	-23501.3	24
ZINC00052413	-1.946	-20.722	3.477	-23.432	-23557.8	-34.545	-57.171	-23466.1	-23557.8	5
ZINC67919965	-1.917	-19.821	2.928	-23.492	-23510.9	-28.437	-16.311	-23466.1	-23510.9	11
ZINC20907761	-1.815	-20.082	2.534	-13.178	-23460.8	-44.806	50.104	-23466.1	-23460.8	1
ZINC67955955	-1.712	-25.43	2.977	-28.91	-23533.6	-45.976	-21.467	-23466.1	-23533.6	1
ZINC39515547	-1.591	-19.8	2.194	-22.303	-23472.1	-36.521	30.517	-23466.1	-23472.1	20
ZINC22926613	-0.627	-18.233	6.405	-12.213	-23455.9	-38.403	48.653	-23466.1	-23455.9	11

Appendix B – Some Colorimetric (MTS) & Western blotting protocols

MTS Assay Protocol

The materials are [3-(4,5-dimethylthiazol-2-yl)-5-(carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium(MTS), Dulbecco's phosphate buffered saline(DPBS) and phenazine methosulphate(PMS)

- Already prepared MTS Solution (2 mL of DPBS + 100 μL of PMS), containers wrapped with foil to protect from light and refrigerated at -20 ^{o}C, is thawed and used. Or else
- 2.0 mL of MTS is removed from the amber reagent bottle aseptically into a test tube
- Add 100 μL PMS solution to the test tube of the 2.0 mL MTS immediately.
- A gentle swirl is required to effect a good mix of the solutions. Refrigerate at -20 ^{o}C until needed
- (a) 20 μL of the MTS/PMS solution is pipetted into each well of the 96 well assay plates (in this case 4 for each drug) containing 100 μL of cells in culture medium
- (b) Plates are incubated at 37°C with CO_2 for 1-2 hours
- (c) The absorbance is recorded at 490 nm using Microplate Reader, Bio-Rad Model 680, following the New Endpoint Protocol

SDS-PolyAcrylamide Gel Electrophoresis

<u>Gel Preparation</u>

Pour Resolving Gel into cast (~3 mL- up to top of green frame to leave room for Stacking gel)

Gently layer 20% isopropanol over top to flatten and remove air bubbles

Takes 45min to set

When completely set, remove isopropanol

Layer Stacking Gel over top and insert comb (10- or 15-well)

Takes 10min to set

12% Resolving Gel

(can make 8-12% depending on desired migration)

1.20 mL	40% Acrylamide
1.00 mL	1.5M Tris pH 8.8
1.76 mL	dH2O
40 μL	10%SDS
20 μL	APS
2.8 μL	Temed

<u>Sample Preparation</u>

Prepare 20 μg protein up to 15 μL total volume, including 3μL 5X Laemmli Sample Buffer

Puncture small hole in Eppendorf lid with syringe to prevent lid popping off

Boil sample, 95ºC 5min

Vortex then store on ice 5 min (careful, as tube is HOT!)

Vortex then centrifuge, 5 min 14000 rpm

4% Stacking Gel

200 μL	40% Acrylamide
500 μL	0.5M Tris pH 6.8
1.28 mL	dH2O
40 μL	10%SDS
40 μL	APS
2.8 μL	Temed

5X Laemmli Sample Buffer
12.5 mL Glycerol
12.5 mL 0.5M Tris-HCl pH 6.8
~1 mL 1% Bromophenol Blue (depends how blue you like it)
Add 75 mg/mL DTT before use or keep at 20ºC

10x Running Buffer 1L
30.29 g Tris-Base
144.13 g Glycine
10.00 g SDS
Make up to 1L with dH2O
Store at 4 ^{0}C and dilute 1:10 for use

10x Transfer Buffer 1L
Same as Running Buffer but without SDS
Dilute 10x Buffer:Methanol:Water 1:2:7

Transferring Gel (Semi-Dry) Transfers protein from gel to membrane
Moisten 3x blotting paper in 1x Transfer Buffer and place on cathode
Prepare PVDF Membrane by placing in Methanol 15sec, dH2O 1 min, 1x Transfer Buffer
Lay membrane over paper
Line up gel in correct orientation
Lay another 3x moistened blotting paper
Place cathode on top, and run at 15 V for 45 min-1hr

Wet Transfer:

Prepare and mark the membrane (right size, PVDF or Immobilon-P membrane) with a pencil; then soak it in Methanol for a few minutes, and rinse it with distilled water.

Equilibrate the membrane, pads, filter papers (4 pieces) and transfer foam in Transfer Buffer (store in 4°C) for 5 minutes.

Assemble transfer sandwich in the following order: black frame (negative electrode) >> foam >> 3 pieces of filter papers >> SDS-PAGE gel >> membrane >> 3 pieces of filter papers >> foam >> red frame (positive electrode). Make the marked side of membrane face the gel. Put it into transfer tank.

Transfer the gel at 80V for 90mins in cold room. Alternatively, use ice bag and magnet stirring bar to transfer the gel in room temperature.

Running Gel

Take glass plates out of green brackets
Secure in gel holder (front plates facing in) and place in tank
Fill inner chamber with fresh 1x Running Buffer
Fill outer chamber with sufficient recycled 1x Running Buffer
Load samples and marker(M) into wells (remember the order you load eg M,1,2,3,4,M,1,2,3,4)
Run gel at 200V for ~45min (until dye leaves gel)
Carefully take gel out of glass plates and place in 1x Transfer Buffer
(If gel comes off back plate, turn gel over to face the correct way- left to right)

Western Blotting

Once protein transferred, block membrane in 5% skim milk for 1hr
Wash 3x for 5min with PBS-Tween 0.01%
Probe with primary antibody for 1hr room temp, or overnight at 4 ºC
Wash 3x for 5 min with PBS-Tween 0.01%
Probe with secondary antibody (mouse or rabbit) for 1hr room temp
Wash 3x for 5 min with PBS-Tween 0.01%
Expose to x-ray film with ECL reagent and develop

Lowry Method for Protein Quantification

Prepare a Bovine Serum Albumin (BSA- stock solution 1mg/mL) Standard Curve up to a total volume of 400μL, in duplicate.

μL BSA	μL dH$_2$O
Blank	400 <mark>(make the blank with 5uL of lysis buffer and 395uL water)</mark>
2.5	397.5
5.0	395
7.5	392.5
10.0	390
15.0	385
25.0	375
50.0	350
100.0	300

Prepare samples up to 400μL (eg 5μL lysate in 395μL dH$_2$O) in duplicate

Add 1mL freshly made de Lowry Solution

Incubate for 10min (room temp) <mark>(start timer from when you first start adding solutions to the test tube)</mark>

Add 0.125mL Folin Solution (diluted 1:3)

Incubate 30min (room temp)

Read Absorbance at 750nm (plastic cuvettes) <mark>(make sure you vortex the solution so that colour is evenly distributed)</mark>

Use standard curve to calculate concentration of unknown samples.

De Lowry Solution (10mL) <mark>(make 1mL extra than what you need to make up for shortages)</mark>

       9.70mL Solution A
       0.15mL Solution B
       0.15mL Solution C

Solution A   3% Na$_2$CO$_3$
                   0.15M NaOH
Solution B   2% NaK Tartrate
Solution C   1% CuSO$_4$.5H$_2$O

Table showing the densitometry values obtained using ImageJ

DPT+EGF	EGF-induced activation	pY-EGFR	EGF-induced activation	pY-ERK1	EGF-induced activation	pY-ERK2
(-EGF)		0		0		0
(+EGF)		1		1		1
DPT1-1uM		2.502566641		85.17276679		104.2224412
DPT1-10uM		62.64462754		1.580823291		24.1959233
DPT1-100uM		4.342151116		176.5379204		216.0222535
DPT2-1uM		0.771180069		0.637417115		0.018254137
DPT2-10uM		67.81419457		0.386611231		0.387944671
DPT2-100uM		46.09391553		0.373124364		15.98103543
DPT3-1uM		1.067637307		0.296768789		15.98103543
DPT3-10uM		3.470947384		0.655056811		0.669164313
DPT3-100uM		0.19058069		2.852395573		2.909691707
DPT4-1uM		1.576258786		91.99206185		3.81722737
DPT4-10uM		1.071542476		2.70388399		91.94953502

Appendix C – Spectral data for the 'synthesized' hit leads (zm1 **7**, 'zm1a **7**' & zm2 **8**)

zm1 7 and zm1a 7

Acquisition Time (sec)	3.8339	Date		Apr 10 2015	Date Stamp		Apr 10 2015	File Name		G:\zm-1h_PROTON_dmso_20150410_01.fid\fid	
Frequency (MHz)	399.73	Nucleus		1H	Number of Transients		512	Original Points Count	24576	Points Count	32768
Pulse Sequence	s2pul	Receiver Gain		42.00	Solvent		DMSO-d6	Spectrum Offset (Hz)	2398.3701	Spectrum Type	STANDARD
Sweep Width (Hz)	6410.26	Temperature (degree C)	25.000								

236
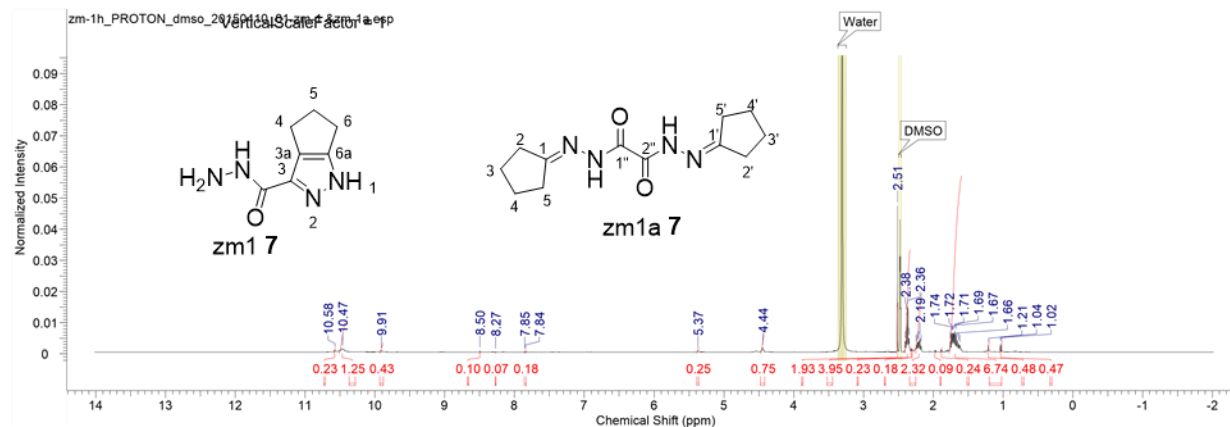
¹³CNMR: zm1 **7**, [1,4,5,6-tetrahydrocyclopenta[*c*]pyrazole-3-carbohydrazide]

zm1 7

Acquisition Time (sec)	1.9661	Date		18 Feb 2015 11:43:40				
File Name		G:\zm-1eP_Fr8-17\zm-1eP_Fr8-17_CARBON_dmso_20150209_01.fid\zm-1eP_Fr8-17_CARBON_dmso_20150209_01.fid.mrc						
Frequency (MHz)	100.52	Nucleus	13C	Origin	Varian	Original Points Count	49152	
Points Count	131072	Pulse Sequence	Unknown	Spectrum Offset (Hz)	11056.1631	Sweep Width (Hz)	25000.00	
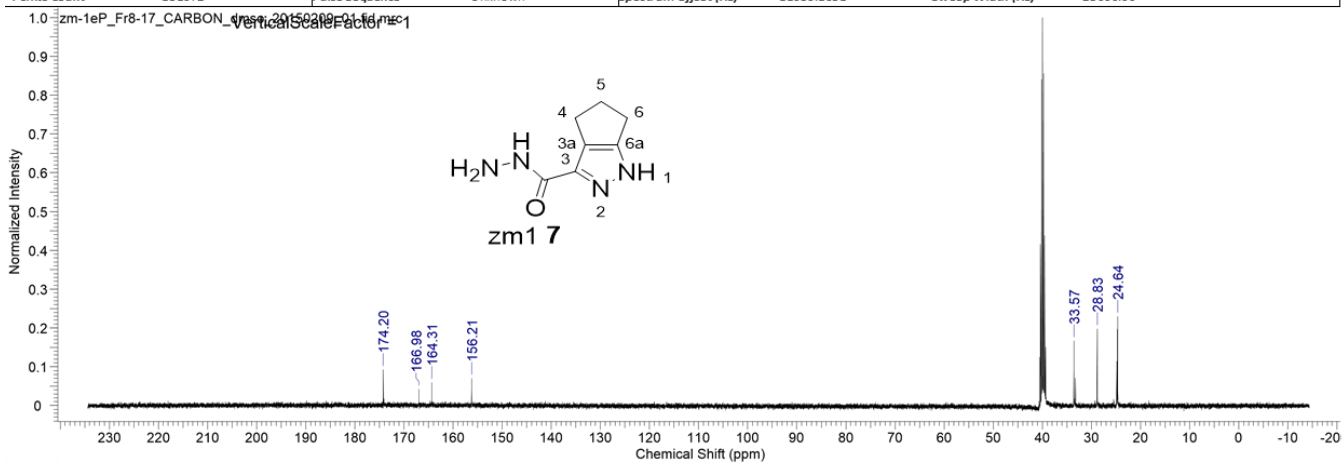
[M + H]$^+$: zm1 **7**, 1,4,5,6-tetrahydrocyclopenta[*c*]pyrazole-3-carbohydrazide
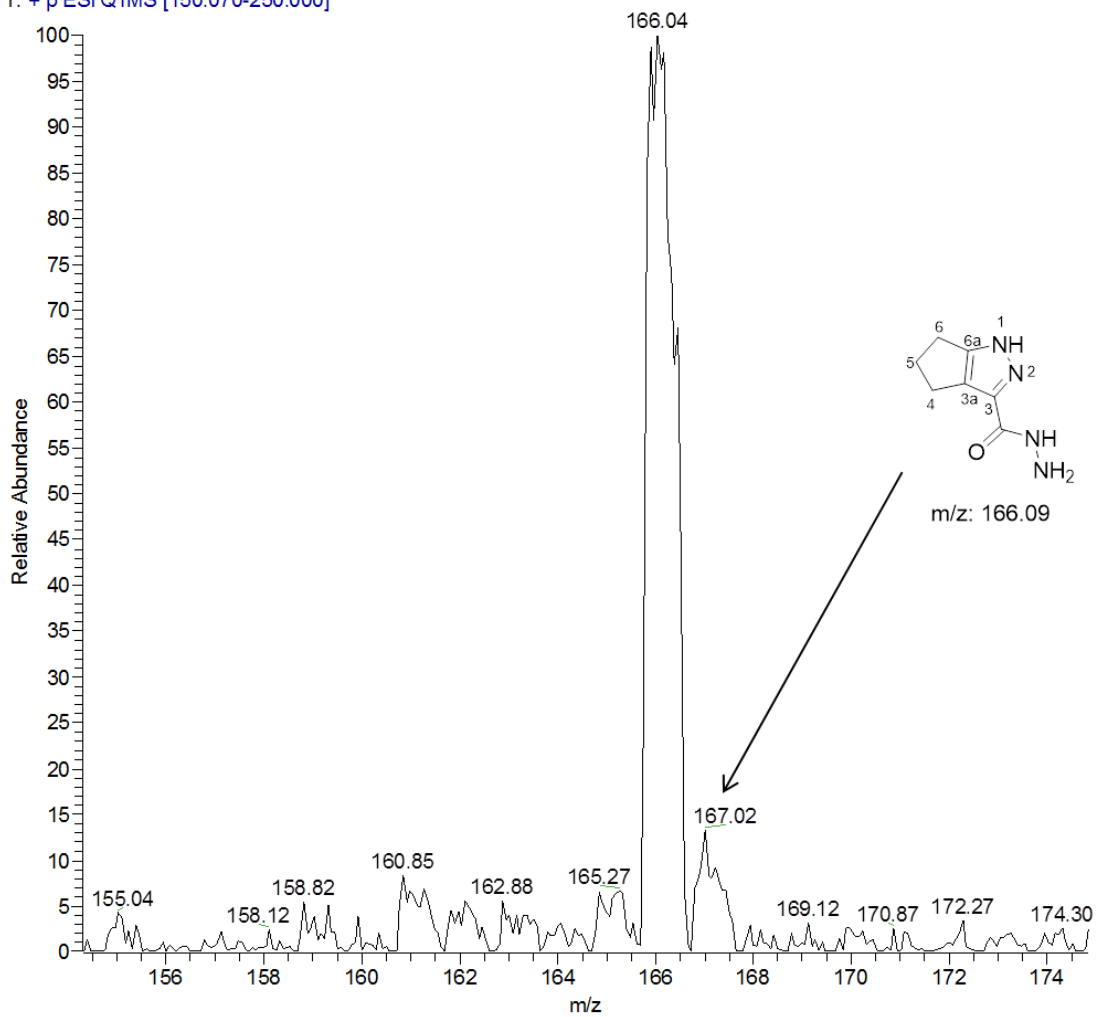
FTIR: zm1 **7**, 1,4,5,6-tetrahydrocyclopenta[*c*]pyrazole-3-carbohydrazide

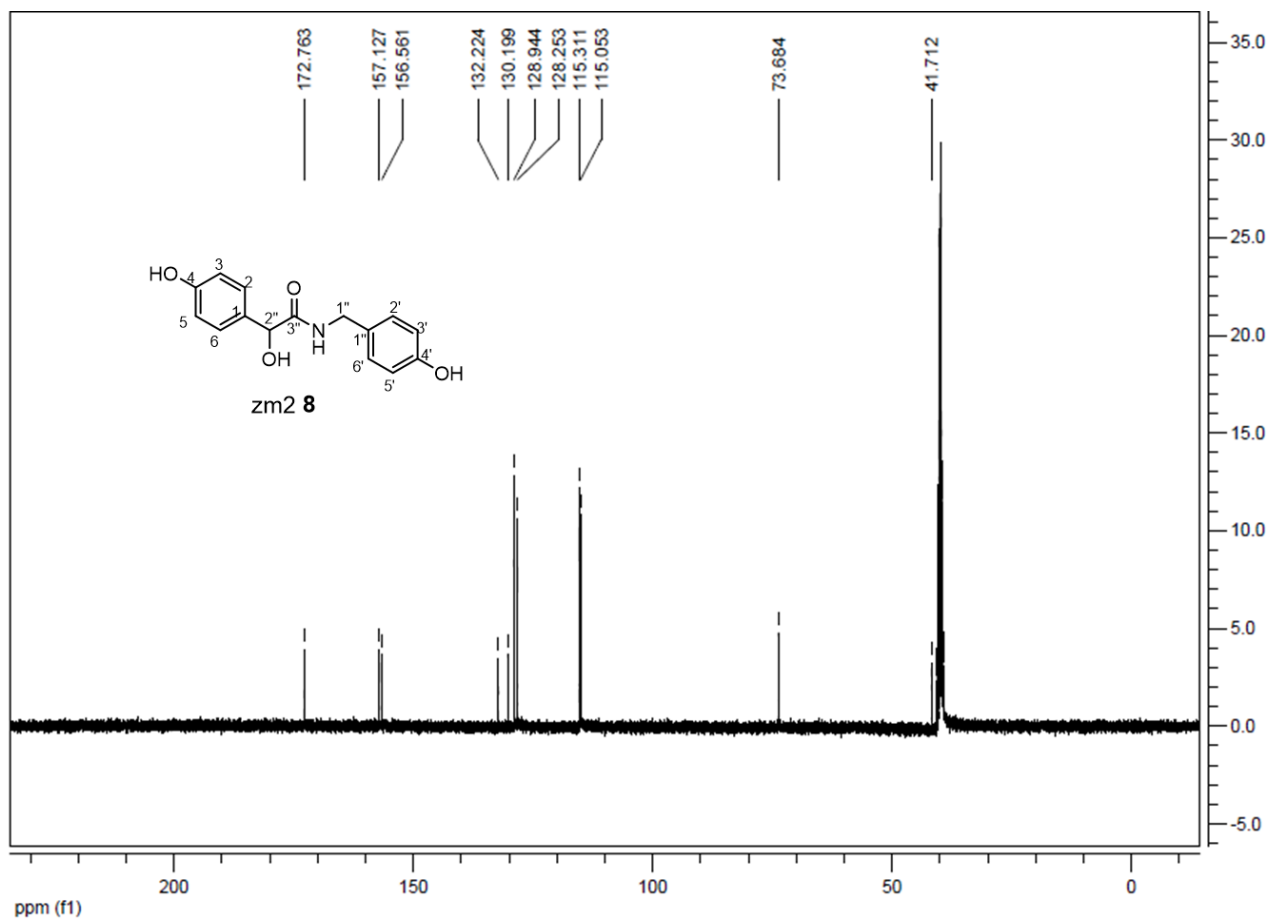# ¹HNMR: zm2 **8**, 2-hydroxy-N-(4-hydroxybenzyl)-2-(4-hydroxyphenyl) acetamide

zm2 8

¹³CNMR: zm2 **8**, 2-hydroxy-N-(4-hydroxybenzyl)-2-(4-hydroxyphenyl acetamide]

[M + H]⁺: zm2 **8**, 2-hydroxy-N-(4-hydroxybenzyl)-2-(4-hydroxyphenyl)acetamide

FTIR: zm2 **8**, 2-hydroxy-N-(4-hydroxybenzyl)-2-(4-hydroxyphenyl) acetamide

zm2 **8**