Global Scientific JOURNALS

Selection of Drug Candidates for Docking with the N-terminal Prion Protein

Pharmacokinetics and Pharmacodynamics.

Michael Kontaiteh Fonjang.

Key Words: Pharmacokinetics, pharmacodynamics, drug candidates, lipophilicity, molecular descriptors, N-term_PrP, Porphyrins, Zinc database, Alzheimer's disease, QSAR, QPAR, and Index of aromaticity.

1. INTRODUCTION.

Pharmacokinetics describes how the body affects specific drugs after administration through the processes of absorption, distribution, metabolism and elimination from the body[1, 2]. The primary goal of clinical pharmacokinetics includes enhancing efficiency and decreasing toxicity of drug therapy. The development of strong correlations between drug concentrations and their pharmacokinetic responses has enabled clinicians to apply pharmacokinetic principles to actual patient situations. Kinetic homogeneity describes the predictable relationship between plasma drug concentration and concentration at the receptor site where a drug produces its therapeutic effect. A change in the plasma drug concentration reflects the changes in drug concentrations at the receptor site, as well as in other tissues. When studying concentrations of a drug in plasma, we assume that these plasma concentrations directly relate to concentrations in tissues where the disease process is to be modified by the drug. Simplifications of body processes are necessary to predict a drug's behavior in the body. A possible way to make these simplifications is to apply mathematical principles, such as the compartmental model which is composed of a number of compartments needed to describe the drug's behavior in the body. Compartmental models are termed deterministic because the observed drug concentrations determine the type of compartmental model required to describe the pharmacokinetics of the drug. On the other hand, pharmacodynamics refers to the relationship between drug concentration at the site of action, and the resulting effect, including the time course of the intensity of therapeutic and adverse effects. The relationship between the concentration of a drug and its effect at the receptor site is a monotonic function; the concentration of a drug at the site of the receptor determines the intensity of its effect, even though, other factors such as density of receptors on the cell surface, the mechanism of transmission of signals by second messengers into the cell, or regulatory factors that control gene translation and protein synthesis also govern the effect of a drug.[3] This multilevel regulation results in variation of sensitivity to drug effect from one individual to another. Drug potency is the concentration at which 50% of the maximum effect is achieved and is referred to as 50% effective concentration or EC50.

2. THE BINDING of PrP WITH DRUG CANDIDATES.

The human prion protein (PrP) is attached to the selectively permeable plasma membrane of all cells at the C-terminal by a glycosylphosphatidylinositol (GP)[4] It comprises a folded C-terminal, also called the Globular Domain (GD), and a naturally disordered N-terminal with transient secondary structures.[5] The N-terminal binds the amyloid- **β** protein and this progresses to Alzheimer's disease.[6] The inhibition of the interaction between the prion and amyloid- **β** is a valuable strategy against Alzheimer's disease.[7] However, the compounds that bind the N-term_HuPrP such as the porphyrins (see figure 1) have poor blood-brain barrier permeability.[8] Porphyrins bind specific residues at the Octapeptide-Repeat (OR) region of the N-terminal. Compounds similar to the porphyrins, with good blood—brain barrier permeability might be ideal drug candidates against Alzheimer's disease.
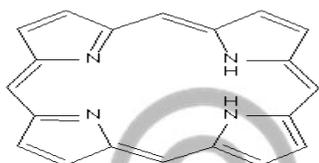
Figure 1. The simplest porphyrin or porphine. Substituted porphyrins might be good drug candidates against Alzheimer's disease. (Adapted from Zinc.docking.org)

3. THE PORPHYRIN STRUCTURE AND AROMATICITY.

The ancestor of porphyrins is the porphine macrocyclic ring; it is a heterocyclic organic compound with a certain degree of aromaticity, and its chemical structure consist of four pyrole rings connected by methine groups to form a larger macrocycle composed of Sp2 hybridised carbon atoms in a conjugated double bond system. The aromatic character in porphyrins might not be a directly measurable quantity and is universally understood by convention. Aromaticity was described in the following ways: (I) A system more stable than its acyclic or cyclic conjugated unsaturated analogues, quantified as resonance energy, or better still aromatic stabilization energy (ASE) [9].(II) The cyclic carbon-carbon bond lengths tend to be intermediate between those typical of single and double bonds, and the quantitative descriptors are the geometrical indices.(III) An external magnetic field induces a diatropic π-ring current, historically an exaltation of magnetic susceptibility(Λ), characteristic of proton NMR chemical shifts.

The recent analysis of porphyrin structure does not use the popular depiction of the porphyrin skeleton as a bridged 18 π-annulene derivative[10]. Instead, the

porphyrin moiety is subdivided into the macrocyclic "internal cross" (an 18 π-electron system of 16 atoms), and the four peripheral pyrole rings, see figure 2.

Figure 2. Two models of the electron system of porphyrin. Left: as bridged [18] annulene derivative: right: as macrocyclic internal cross (in bold), and four pyrole rings (1-4). Adapted from chemical Rev. 2001, 101, 1385-1419.

The C-C and C-N bond lengths in the porphyrin macrocycle can be influenced by exocyclic substitution and by coordination with metal cations. The analyses based on the harmonic oscillator model of aromaticity (HOMA) [11] consist of two components, EN and GEO.

$$HOMA = 1 - \frac{\alpha}{n}\sum(R_{opt} - R_i)^2 =$$
$$1 - \left[\alpha(R_{opt} - R_{av})^2 + \frac{\alpha}{n}\sum(R_{av} - R_i)^2\right] =$$
$$1 - EN - GEO$$

............14

The model makes it possible to separate the two terms describing different contributions to a decrease in aromaticity. The EN term describes changes in aromatic character due to deviation of the average bond length from the optimal value, while the GEO term reflects the consequences of bond length alternation[12]. Here, n is the number of bonds in the summation, α an empirical constant fixed to give HOMA = 0 for a hypothetical Kekule structure of aromatic systems, and the value 1 is a representative of a system with all bonds equal to the optimal value R_{opt}[13], R_{av} represents the average bond length, while the individual bond lengths R_i are obtained[12] from the Pauli definition of bond number[14]. In free base porphyrin, rings (1 & 3) are protonated, see figure 2. The positions of these protons were assigned on the pyrole rings by means of the harmonic stabilization energy (HOSE) model[15]. N-H tautomerism[16] tends to reduce the difference between the pyrole rings.

The pyrole rings with NH groups (1 & 3) are more aromatic[18], than the other five-membered rings (2 & 4). This can be ascribed to the HOMA values of 0.666 and 0.452 respectively; on the pyrole rings (2 & 4), the nitrogen is Sp2 hybridised with the lone pair of electrons in a π-orbital parallel to the plane of the ring, and consequently, the lone pairs of electrons are not available for delocalization. The NICS values computed in the ring centers show the difference

more clearly; the NICS (-15.2) for rings (1 & 3) is similar to the pyrole value (-15.1) [19], but rings (2 &4) have much lower values (-4.5), implying that the C_2H_2 groups function as exocyclic bridges.

Metal complexation reduces the difference between the two types of pyrole rings, but the porphyrin skeleton does not lead fully to D_{4h} symmetry. Computed NICS values at various points on the macrocyclic ring reveal that the porphyrin free base is more aromatic than the porphyrin metal complex, and in turn, the porphyrin metal complex is more aromatic than the porphyrin dianion. The results of this computation are summarised in figure 3.

Figure 3 .Computed NICS values for free porphyrin base, its dianion, and the Mg complex. The dashed lines indicate the delocalized π electrons deduced from the NICS and HOMA analysis. Adapted from chemical Rev. 2001, 101, 1385-1419.

4. BLOOD-BRAIN BARRIER PERMEABILITY.

One of the most important features of the brain and spinal cord is their complete separation from the blood by the blood-brain barrier (BBB), and the blood-spinal cord barrier respectively. A complex network of epithelial cells, astroglia, pericytes, perivascular macrophages, and basal lamina forms the BBB.[20, 21] Clear evidence for the existence of this permeability barrier emerged in 1909 with the demonstration by Erwin Goldman that a dye injected into the blood stream of a rat stained the whole body except the brain and spinal cord. On the contrary, the injection of the dye into the cerebral ventricles, stained the brain and spinal cord but not the rest of the body.[22] Consequently, the blood-brain barrier significantly impedes entry from blood to brain of virtually all molecules, except those that are small and lipophilic, or those that enter the brain through an active transport mechanism.[23] It is not sufficient for neurotherapeutic agents to move across the blood-brain barrier; they also have to stay in the brain long enough to exert their desired action. For CNS drug discovery, it is clearly essential to establish whether a compound will penetrate and distribute within the CNS because efficacy is largely dependent on sufficient exposure within the CNS.[24] The number and

strength of intermolecular forces between a drug and the surrounding water molecules determine the ease with which any particular drug diffuses across the blood-brain barrier.[25, 26]. By quantifying the molecular features that contribute to these forces, it should be possible to predict the in vivo blood-brain barrier permeability.[27]

Molecular lipophilicity, usually quantified as LogP, and aqueous solubility represented as LogS are important molecular descriptors in medicinal Chemistry. [28] Physiochemical features of CNS drugs are related to their ability to penetrate the blood-brain barrier and exhibit CNS activity. The bioavailability of a drug and its access to therapeutic targets are important considerations in rational drug design. Partition coefficients are useful in the estimation of drug distribution within an organism. The hydrophobic drugs with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells, while hydrophilic drugs with low partition coefficients are preferentially localised in hydrophilic compartments such as blood serum.[29] The LogP coefficient is well known as one of the principal parameters which estimates lipophilicity or solubility in lipids, and to a large degree, determine their pharmacokinetic properties. According to Lipinski rule of 5, a drug with a LogP coefficient greater than 5, would not have good absorption properties. Most CNS active drugs have LogP coefficient less than 5, for it to be easily absorbed into the blood-brain barrier, where active diffusion of drugs is likely to have most lipophilic areas such as lipid bilayers of membranes, and if the LogP value is less than 0, it will move towards hydrophilic compartments. The LogP value is a measure of the distribution of the ligand between an organic phase (octanol) and an aqueous phase (water). Very low LogP values result in poor lipid bilayer permeability, and very high values cause poor aqueous solubility, consequently, the compromising LogP value should be between 0 and 5.[30] The LogS coefficient is a measure of the concentration of a drug in aqueous solution. Drugs with LogS coefficients less than 0 and greater than -4 are easily absorbed in the blood and tissues.[31]

5. THE MOLECULAR DESCRIPTORS.

Molecular descriptors are numerical values that characterise the chemical and physical properties of a molecule. The descriptor might be structure-based or experimentally derived, and it varies in complexity of encoded information and in compute time. Quantitative Structure-activity Relationship (QSAR) or Quantitative Property-activity relationship (QPAR) studies require numerical molecular descriptors associated with the structural formulas, which are discrete entities.[32] Hydrophobic, steric or electronic descriptors associated with molecules or substituents are invaluable in QSAR studies. Topological indices alone or in combination with physicochemical descriptors have also proved their usefulness; they could be associated to molecular structure, and are simple and easily computed for large numbers of structures.[33] This is particularly valuable, especially with the fast development of combinatorial libraries, which require one to explore millions of structures for QSAR studies, prior to high-throughput synthesis and testing. Numerous activities and properties of molecules depend

on the presence of specific atoms or functional groups in their structures. These features can be used in deriving QSAR models via regression analysis or artificial neural networks. In QSAR or QPAR studies, an attempt is made to relate the structure of a molecule to a biological activity by means of a statistical tool. Such a relationship is codified in expression 1:[34]

QSAR or QPAR = F(molecular structure) = F(molecular descriptor)………………….1

Where F represents a mathematical function.

The Gaussian process (GP) is a newly introduced method in QSAR for the modeling of nonlinear relationships. It has a built in tool to prevent overtraining and does not require cross-validation, and was particularly used to construct solubility models[35]. In GP, the property Y to be modeled is described by a family of Gaussian processes priors F (X), where X is a vector of descriptors, Y a vector of the observations and F is a random functional value for any finite set of n points from a Gaussian distribution with zero mean value. The principle of statistical inference is used to identify the most likely posterior functions that combine the priors to model the observed data Y. The covariance function k (X, X') is used to relate one observation to another, and the covariance function is a squared exponential.

$$k(X,X') = \sigma_f^2 \exp\left[-\frac{1}{2}\sum_{i=1}^{n} w_i(x_i - x_i')^2\right] + \sigma_n^2 \delta(X,X')$$

………………2

Where δ (X, X') is the Kronecker delta function, σ_n is the variance of Gaussian noise with zero mean, n is the number of descriptors, w_i are the weights to be determined, K is the covariance matrix k(X,X'). For a given new vector X*, K* = [k (X*, X'), k(X*, X²)………k(X*, Xⁿ), the predicted value Y* and uncertainty are calculated using the formulas:

The summation is from descriptor n = 1, to n = n

Schwaighofer et al. used a different covariance function in the solubility model[35]

$$k(X,X') = \left(1 + \sum_{i=1}^{N} w_i(x_i - x_i')^2\right)^{-\nu}$$

…………………………… 3

The reliability of GP prediction depends on how well the parameters in the covariance function are optimised. Based on Bayer's theorem, the best parameter set $\theta = (w_i, \sigma_f, \sigma_n)$ is obtained by maximising LogP $(X/Y, \theta)$, described in equation 4.

$$\log p(Y|X,\theta) = -\frac{1}{2}Y^T K^{-1} Y - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi$$

...............4

Considering permeability, absorption, distribution of a drug in the system, plasma protein binding, metabolism, elimination, and toxicity, the following molecular descriptors may be used to select a ligand with drug-like properties and biological activity:

- Lipophilicity Coefficient (LogP)
- Aqueous Solubility (LogS)
- Apolar and Polar Desolvation (AD & PD)
- Hydrogen Bond Counts (HD & HA)
- Molecular Weight (MW)
- Molecular Flexibility (RB)
- Polar Surface Area (PSA)

5.1. LIPOPHILICITY

The lipohilicity coefficient represented as LogP describes a compounds ability to dissolve into lipophilic solutions. Experimentally, it is measured as a compounds distribution between a non-aqueous phase (1-Octanol), and aqueous phase (water), expressed as the logarithm to the base 10 of the concentration ratios:

$$\log P_{oct/wat} = \log\left(\frac{[\text{solute}]_{octanol}}{[\text{solute}]_{water}^{un-ionized}}\right)$$

.... 5

It can be observed from Table 2, the differences in lipophilicity coefficients for most drug candidates; these can be ascribed to the several approaches in the calculation of lipophilicity coefficients. The different approaches range from experimental, fragmental, atom-based, conformation-dependent, property-based, quantum chemical, molecular dynamics, lattice energy, property-based, continuum solvation, 3D structure representation, molecular lipophilicity potentials, and graph molecular connectivity approaches[36]. In one study, a total of 30 and 18 methods were tested for public and industrial datasets, respectively, and it was found that accuracy of models declined with the number of nonhydrogen-bonded atoms. The Arithmetic Average Model (AAM), was used

as the baseline model for comparism; methods with Root Mean Square Error (RMSE) greater than that of the baseline were unacceptable, while the majority of analysed methods produced reasonable results for the public dataset but only seven methods were successful on the in house datasets .[37] The predictive power of the theoretical approaches can be checked by comparing their results with those of reliable experimental methods. From literature[38], the predictive power of the calculation procedure is significantly better for simple organic molecules than chemically heterogeneous drug structures. Considering the structure-based approach, the calculation procedure can be arranged in three groups with significantly differing predictive power: Fragmental > atom-based > Conformation-dependent approaches.[38, 39]As an illustration of the differences in LogP coefficients for a single drug candidate, we will focus mainly on the fragmental and the atom based methods.

The fragmental methods, for instance, the Molinspiration (MiLogP) method, cut molecules into fragments and apply correction factors in other to compensate for intermolecular interactions. This is expressed in equation 6:

$$\log P = \sum_{i=1}^{n} a_i f_i + \sum_{j=1}^{m} b_j F_j$$

.... 6

Where, the first term considers the contribution of fragment content (f_i), and the incidence of the fragment (a_i) in the query structure. The second term considers the contribution of the correction factor (F_j), and its frequency (b_j). The summation is from fragment n = 1, to n = n, and the contribution of the correction factor ranges from m = 1 to m = m. A prime advantage of the fragmental method is the defining of fragments larger than single atoms which guarantee that significant electronic interactions are compromised within a fragment, while the main disadvantage is the arbitrary fragmentation and missing fragments.

The Atom based methods [40, 41] such as the Dragon ALOGP [42] cut molecules into single atoms and do not apply correction rules. This is described in equation 7:

Log P = $\sum n_i a_i$...7

 Where n_i represent the number of atoms of type i, and a_i the contribution of an atom of type i. Since the partition coefficient is not a simple additive property, the constitutive feature is covered by classifying huge numbers of atom types according to structural environment. An advantage of atom-based method is the avoidance of ambiguities, and its shortcoming is the failure to deal with long-range interactions.

The Molinspiration method of calculating lipophilicity coefficients can be reliable by virtue of its fragment-based approach, which applies a correction factor in order to compensate for intermolecular interactions. Molinspiration

Chemokinetics providers used datasets of more than 12 thousand molecules. It is based on group contributions and was developed using 35 small basic fragments as well as 185 larger fragments characterising intermolecular Hydrogen bond contributions and charge interactions. This procedure achieved $R^2=0.94$ and MAE = 0.33 for 12202 molecules[43]

Where R^2 represents squared Pearson correlation Coefficient and MAE is the Mean Absolute Error.

5.2. AQUEOUS SOLUBILITY.

Aqueous Solubility is one of the key factors that affect a drug's oral bioavailability[44, 45]. It governs both the rate of dissolution of the compound and the maximum concentration reached in the gastrointestinal fluid. Generally, a drug with high solubility and membrane permeability is considered free from bioavailability problems; otherwise, it is a problematic candidate or needs careful formulation work. Solubility is of various types, for instance, intrinsic, thermodynamic, apparent, and kinetic solubility. Most in silico models are developed to predict the intrinsic solubility S. In concept, the intrinsic solubility of a compound in its neutral form is the concentration in moles per dm^3 (Mol/L) of its saturated solution in equilibrium with the solid phase. Other units of measurement of solubility include solute per volume (mg/L), solute per weight (g/kg), molarity (mole/L), molality (mole/1000g), and percentage solution (g/100ml). For completely ionisable electrolytes, solubility without reference to pH and ionisation constant pKa is meaningless. The logarithm of solubility is often used for convenience. For a given solid state and solvent, the solubility S is almost exclusively dependent on intermolecular adhesive interactions between solute-solute, solute-solvent, and solvent-solvent molecules. In other words, the crystal packing cavitation and solvation energy determine the intrinsic solubility of a compound. The first type of mathematical expression exemplified to describe LogS was by Jain and Yalkowsky's in which LogS was correlated with experimentally determined temperature of melting point (TM) and the logarithm of octanol/water partition coefficient with n = 580, and AUE = 0.42[46]. Those models were referred to as the general solubility equation (GSE), but Wang et al noticed that there were several duplicated and erroneous entries in Jain and Yalkowsky's data set, and after eliminating those problematic entries and replacing the experimental LogP with the calculated (CLogP), a new general solubility equation was constructed.

LogS = 3.513 – 0.010 × TM – 1.112 × ClogP..8

Where TM represent melting point temperature of solid at -25°C, for a liquid the melting point temperature is 25°C, with r^2 = 0.937.

5.3. POLAR AND APOLAR DESOLVATION ENERGIES.

Apolar desolvation energy (Kcal/mol) at 298.15K is the reduction in Gibbs free energy when a compound is dissolved in solution and the solvent is assumed to

be apolar , while polar desolvation energy is the same but the solvent is polar . In terms of thermodynamics, it is expressed as:

$$\Delta G_{bind} = \Delta E_{elec} + \Delta G_{des} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 10$$

Where ΔE_{elec} represents total electrostatic energy and ΔG_{des} is the total desolvation energy. Based on equation 10, the desolvation energy of a protein can be approximated using expression 11

$$g(r) \sum\sum e_{ij} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots 11$$

This is the interaction between the ith atom of a ligand and the ith atom of a receptor, and e_{ij} is the atomic contact potential (ACP), $g(r)$ is a smooth function based on distance. We assume atoms to be within 7Å distance. Within this range, $g(r)$ varies from 0 to 1 and this information can be used in calculating desolvation energy[47]

5.4. HYDROGEN BOND COUNTS.

A hydrogen bond is an intermolecular or intramolecular force; it is a bond formed between a very electronegative atom such as Oxygen, Nitrogen, Sulphur with available lone pairs of electrons, and a hydrogen atom attached to a very electronegative atom. The atom types in a ligand may be hydrogen donor or hydrogen acceptor and based on Christopher A Lipinski rule of 5, the
Number of hydrogen bond donors should not be more than 5, and the number of hydrogen acceptors in a ligand should not exceed 10 [30]

5.5. MOLECULAR FLEXIBILITY.

The number of rotatable bonds (RBN) is the number of bonds, which allow free rotation. This is defined as any single bond, not in a ring, bound to a nonterminal heavy atom. Excluded from the count is the amide C–N bond, which is stabilized by resonance and have a very high rotational energy barrier[48].
Rotatable bonds describe the flexibility of a molecule and are also a determinant of drug likeliness as well as oral bioavailability. Lipinski rule of 5 suggest a drug ought to have fewer than 10 rotatable bonds.[30]

5.6. MOLECULAR POLAR SURFACE AREA

The polar surface area (PSA) of a molecule is the surface sum of all polar atoms such as oxygen and nitrogen. It measures the ability of a drug to penetrate the blood-brain barrier in order to bind on receptors. It is a very useful parameter for the prediction of drug transport properties and correlates very well with the human intestinal absorption, [49, 50] CaCo-2 monolayer penetrations, [51-54] and

blood-brain barrier penetration [55, 56]. Time requirements to calculate the PSA are generally high, up to tens of minutes per molecule when the geometry is optimised or conformational search is performed. Of recent, a fast single-conformer method for PSA calculation was reported with a throughput of several molecules per second. [57] A new methodology for the calculation of PSA termed topological polar surface area (TPSA) is based on the summation of tabulated surface contributions of polar fragments such as atoms and their bonding patterns. The contributions of polar fragments were determined by last-square fitting of the fragments-based TPSA to a single conformer 3D PSA for a large set of drug like structures. Molecules from the world drug index, [58] were used for this procedure. A better solution would be to scale contributions of polar fragments according to the strength of hydrogen bonds [59]. From the fitting procedures, the values for surface fragment contributions can be determined from equation 12.

$$3D\ PSA = \sum_{i}^{n_{types}} n_i \cdot c(fragment_i)$$

……………………………………………12

Where 3D PSA is the traditionally calculated PSA based on 3D molecular structure, n_{types} is the number of types of polar fragments, $c(fragment_i)$ is the optimised surface contribution of fragment i, and n_i is the frequency of fragment i in the molecule. The 3D PSA used as a target in the fitting was calculated from CORINA [60] geometries considering the Van der Waals surfaces belonging to Oxygen, Nitrogen, Sulphur, and Phosphorus atoms, including their attached hydrogen. All structure manipulation, processing of SMILES, identification of polar fragments, statistical analysis, was done using an in-house molecular development kit written Java. The statistical analysis provided good correlation between 3D PSA and TPSA with the following statistical parameters: $r^2 = 0.982$, $r = 0.991$, $\sigma = 7.83$ Å^2 and $_{average}$ absolute error = 5.62Å^2.

5.7. MOLECULAR WEIGHT.

Molecular weight or molar mass of a substance is the mass of one mole, or the mass of $6.0632*10^{23}$ particles of the substance. The SI units of molar mass are the kilogram per mole (kg/mol), but it is commonly expressed in grams per mole (g/mol). The molecular weight of a substance can be calculated from its molar refractivity from equation 13:

$$\frac{n^2 - 1}{n^2 + 2} \frac{MW}{d} = MR$$ …………………………………………13

Where MR is the molar refractivity, n is the index of refraction, d the density of substance, and MW is the Molecular weight.

Ligands with lower molecular mass will easily diffuse through the phospholipid bilayer of the plasma membrane than larger molecules, and based on these criteria, molecules with molecular masses between 150 and 500 are more suitable as drug candidates.[30]

6. METHODOLOGY.

Structure-based virtual screening has had several important successes in recent years[61], and is now a common technique in early stage drug discovery. The zinc database contains a library of 727842 molecoles[62], and catalogs of compounds from vendors. The molecules have been assigned biologically relevant protonation states and are annotated with properties such as logP coefficients, hydrogen-bond counts, number of rotatable bonds, polar surface area, desolvation energies, net charge, apolar and polar surface area, and molecular weights. Each molecule in the library contains vendor and purchasing information and is ready for docking simulations. The ZINC platform contains subsets of search methods such as, structure, catalog or vendor, property, target, ring, and cart. With the subset property, we used 'drug-like' to filter the drug-like properties such as molecular weight between 150 and 500, calculated logP in the range $-4 \le 5$, number of hydrogen bond donors < 6, number of hydrogen-bond acceptors < 11, and number of rotatable bonds < 14. A total of 2548 purchasable compounds, which satisfied the above constraints, were displayed.

To select the porphyrin-like compounds, we then imputed the porphyrin SMILE or its name, two compounds annotated with molecular descriptors where obtained. We then used the ZINC ID of the two compounds to search for more molecules similar to porphyrin. As the number of porphyrin-like molecules increased, we used their ZINC ID to obtain more, and finally, we arrived at 36 porphyrin-like molecules. From the 36, we selected 29 of them with lipophilicity coefficients in the range (0, 5), and molecular weight less than 500. However, we tolerated some lipophilicity coefficients than proscribed by the Lipinski's rule of 5 because of the uncertainties in the calculated values. The molecules can be downloaded in the following formats: MOL2, SDF, SMILES, or as flexibase.

Limits on molecular properties such as net charge and molecular weight are specified on the left-hand side of the search Web page. On the bottom left, individual ZINC database registration codes, the unique serial number assigned to each substance may be specified, either by typing or choosing a text file of codes to upload from the browsing computer, and molecules matching any of the

ZINC codes specified will be found. On the right, molecular substructures may be drawn using the Java Molecular Editor (JME) [63].

To calculate LogP or LogS coefficient using the Molinspiration algorithm or the ALOGP2.1, we copied the SMILE or draw the structure of the ligand from the ZINC database, imputed it on the platform, and then click on calculate LogP or LogS coefficient. The calculation is performed in a matter of seconds, and the results displayed.

7. RESULTS AND DISCUSSIONS.

7.1 DATA ANALYSIS.

Results of the search are reviewed using the Database Browser, and most of the web queries are answered in half a minute or less, complex queries or multiple simultaneous requests may take longer. The database Browser displays in a table containing ZINC registration codes, a 2D sketch, purchasing information, and molecular properties such as calculated LogP, number of rotatable bonds, hydrogen bond counts, topological polar surface area, net charge, molecular weight, and sometimes the properties are pH dependent for some molecules.

About 100000 molecules of the Zinc database violate Lipinski rule of 5, while more than 500000 are acceptable. These figures are a strong indication that the results from the ZINC database are reliable. Consequently, the probability is high that the 29 molecules selected are drug like. The differences in lipophlicity coefficients have been discussed in section 5.1; Different methods of calculation result in differing lipophilicity coefficients for a single drug candidate. The Molinspiration, ZINC, and the ALOGPS2.1 methods of lipophilicity calculations have been presented in table 1 & 2. In table 1, we have presented the molecular descriptors for each drug candidate based on the ZINC calculation. The minimum value of LogP being -0.4 for ZINC59380156 and the maximum is 4.55 for ZINC78206861. The negative logP values are a bit out of the Lipinski rule of 5 ranges, but other molecular descriptors such as molecular weight, molecular flexibility, and total polar surface area might compensate for the lower lipophilicity coefficients. In table 2, the lipophilicity and aqueous solubility coefficients of the 29 molecules based on the Molinspiration and ALOGPS2.1 calculations have been presented. See section 5.1 for the differences in the calculation of lipophilicity coefficients.

7.2 THE PORPHYRIN MOLECULAR STRUCTURE.

The porphyrins bind the NH_2- Term_PrP at the 5Trp residues, W57, W65, W73, W89 at the Octapeptide Repeat region (OR), and also a number of charged residues such as K23, K24, K100, K103 and K105[64]. The 29 molecules selected from the ZINC database are similar to porphyrin in chemical and physical properties. The chemical structures of the drug candidates are presented in table 3.

The molecules are heterocyclic macrocycles made from four pyrole subunits and linked by methine bridges. The porphyrin structure is characterised by high stability, which is as a consequence of its conjugated macrocyclic system. It can be viewed as four pyrole rings connected to each other by four meso- carbons. The pyroles are aromatic 5- membered rings with 6 π-electrons and high degree of aromaticity.

7.2.1 EFECT OF SUBSTITUENTS ON THE PORPHRIN MACROCYCLIC RING.

The substituents on the macrocyclic ring influence its chemical properties, as well as biological activity: (I) the electron donating groups such as –OH, -NH$_2$, -OR, -NHCOR, -R, where R represents an alkyl group increase the electron density and aromaticity of the ring. Electron donating groups contain an electronegative atom directly attached to the ring, and the electronegative atom has at least a lone pair of electrons for delocalisation by the positive mesomeric effect. The alkyl groups increase electron density by the positive inductive effect. (II) Electron withdrawing groups such as –COOH, -COOR, -CONH$_2$, -COR, -SO$_3$H, -CN, -NO$_2$ decrease the electron density and aromaticity of the ring by the negative mesomeric or resonance effect. The electron withdrawing groups all contain an electronegative atom, and at least a double or triple bond for delocalisation. It can be observed from table 3, that ZINC 60076174, 80774214, 60076186, 78069660, 7809660, 79132071, 60076177, 59380156, 78222613, 78139775 for instance, have strong electron withdrawing groups attached to the macrocyclic ring. This electron withdrawing groups pull electrons from the ring, thereby making it less aromatic and the entire molecule is more polarised, this result in lower lipophilicity coefficients and aromaticity indices. The polar drug candidates will easily bind polar residues at the NH$_2$-terminal PrP such as Lysine, Arginine and Histidine. On the other hand, ZINC 49783984, 78015143, 82153006, 78315702 contain electron donating groups which enhances the electron density and aromaticity of the macrocycle, and consequently, the porphyrin macrocycle has a high affinity for π-stacking interactions with aromatic residues on NH$_2$-terminal PrP such as tryptophan.

The value of the lipophilicity coefficient decreases on average with the strength of electron withdrawing groups on the porphyrin macrocyclic ring due to increasing polarity and aqueous solubility. Thus, ZINC 59380156 and 49783984 with very strong electron withdrawing groups are polar and have the lowest lipophilicity coefficients in table 1, while ZINC 78015143 and 78315702 with electron donating groups have very high lipophilicity coefficients due to an increase in the aromatic character of the porphyrin macrocycle. Although ZINC782068 has a lipophilicity coefficient of 4.55, its electron withdrawing group is not directly attached to a pyrole but to an exocyclic substituent (the methine group), and so its effect is minimal.

ZINC 78222613 is a dianion and has a lower logP value of -0.652 based on the Molinspiration calculation and 1.25 based on the ZINC calculation. Furthermore, it has an aromaticity index of -14.9 compared to that of the uncharged macrocyclic ring which is -16.5 based on the Nuclear Independent Chemical Shift (NICS). These lower values may be ascribed to the strong electron withdrawing groups which polarizes the molecule thereby making it less aromatic. The

dianion will obviously have a high affinity for the positively charged residues such as Arginine, lysine and Histidine at the N-terminal PrP.

Table 1. Classification of drug candidates based on the rule of 5 by Lipinski, namely; the LogP coefficient, number of rotatable bonds, Molecular weight (MW), Solubility (LogS), Apolar Desolvation (AD), Polar Desolvation (PD), Polar surface Area (PSA)

ZINC ID	LogP	PD	HD	HA	Charge	PSA	MW	RB
60076174	1.54	-40	7	8	1	145	427.45	5
80774214	2.64	-67	4	8	-1	141	428.47	4
60076186	2.35	-40	7	8	1	145	455.5	4
78069660	2.28	-59	7	10	1	160	486.55	6
79132071	0.96	-52	8	10	1	172	489.55	6
60076177	1.72	-41	9	12	1	202	587.66	8
49783638	2.61	-99	10	10	2	171	544.7	6
59380156	-0.4	-22	8	10	0	187	485.5	6
78222613	1.25	-85	6	12	-1	211	543.56	8
78139775	1.80	-99	6	12	-1	211	571.6	10
59380157	0.39	-99	11	10	3	192	530.65	9
60076177	1.72	-41	9	12	1	102	428.56	5
78222613	1.25	-99	5	12	-2	210	542.55	8
78315702	3.65	-14	6	8	0	138	378.3	0
60194322	3.69	-26	5	6	1	102	324.46	0
60194323	3.38	-82	7	8	2	132	451.5	4
82190886	3.40	-17	4	7	0	107	415.49	4
79824394	3.55	-67	3	8	-1	127	442.45	5
79231393	3.77	-16	4	7	0	107	429.52	4
82054668	3.82	-69	4	8	0	131	457.54	6
59651175	4.16	-16	2	5	0		328.37	0
82153006	4.23	-14	4	6	0	98	346.39	0
38792143	4.81	-13	2	4	0	57	314.39	0
60308698	3.15	-72	6	10	-1	178	519.57	3
60308695	3.47	-43	7	10	0	180	518.57	3
78015143	3.65	-15	6	8	0	138	378.38	
49783984	-0.3	-82	2	6	2	57	562.8	12
49784734	3.69	-87	8	10	2	147	606.8	12
78206861	4.55	-63	5	6	1	102	428.56	5

Table 2. Prediction of LogP coefficient based on the Molinspiration Method (MinLogP), the ALOGPS2.1 (ALogS), and the calculation of solubility based on the ALOGPS2.1 (LogS).

ZINC ID	MinLogP	LogS	ALogP
60076174	-1.48	-4.30	1.26
80774214	-0.07	-3.96	1.58
60076186	-0.676	-4.52	1.63
78069660	-0.351	-3.99	0.47
79132071	-1.248	-3.92	-0.31
60076177	-1.298	-4.64	1.5
49783638	1.206	-4.76	-0.35
59380156	-0.04	-3.24	-0.04

78222613	0.652	-3.93	0.01
78139775	-0.111	-4.11	0.07
59380157	-0.282	-5.38	-1.40
60076177	-1.298	-4.65	2.09
78222613	-0.652	-3.93	0.01
78315702	3.649	-1.84	-0.45
60194322	0.667	-5.38	2.97
60194323	0.396	-5.82	1.55
82190886	3.405	-3.77	0.96
79824394	0.839	-3.78	1.12
79231393	3.768	-3.85	1.35
82054668	0.959	-4.66	1.18
59651175	4.158	-3.84	1.48
82153006	4.232	-3.14	0.60
38792143	4.814	-4.46	2.77
60308698	0.084	-3.22	0.48
60308695	0.244	-3.88	1.60
78015143	3.649	-2.79	1.22
49783984	-0.881	-5.92	-0.25
49784734	0.639	-5.61	0.28
78206861	2.349	-5.20	0.92
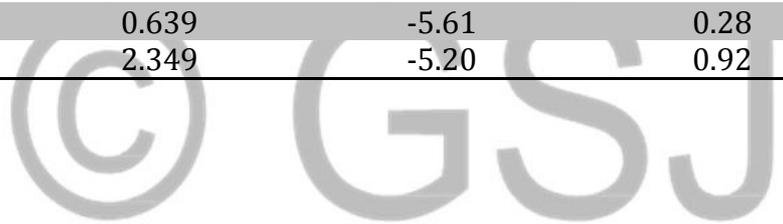
Table 3. Molecular structures of the selected drug candidates from the Zinc database.

ZINC ID	CHEMICAL STRUCTURE
60076174	
80774214	
60076186	

78069660
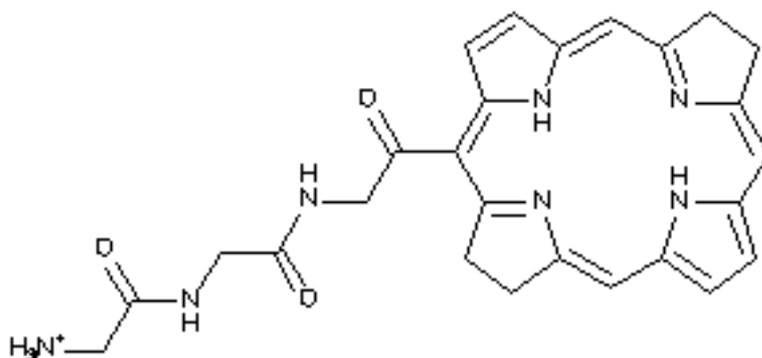
79132071
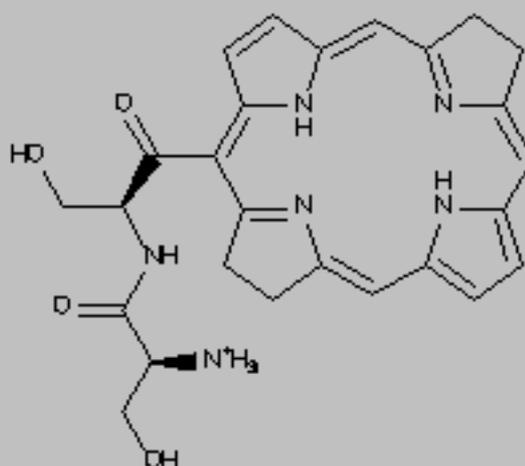
60076177
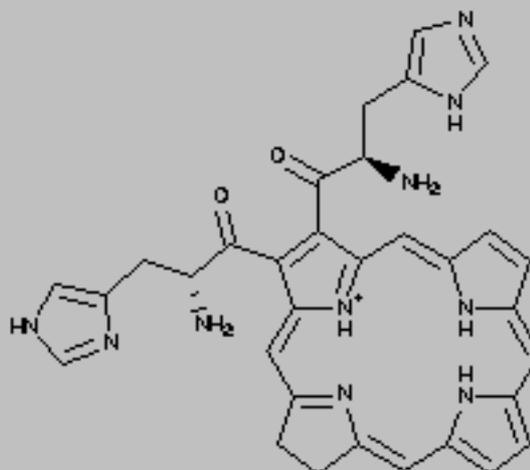
49783638
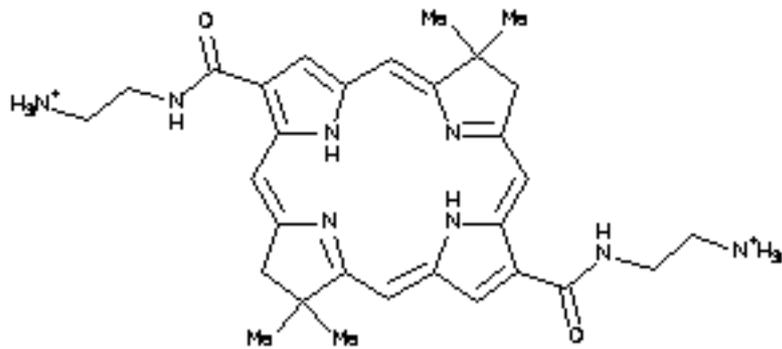
59380156
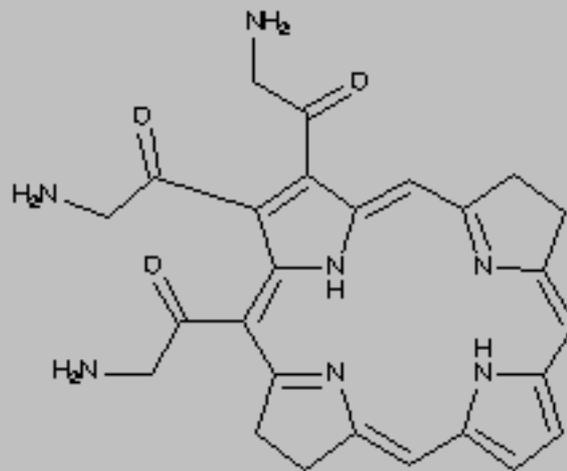
78222613
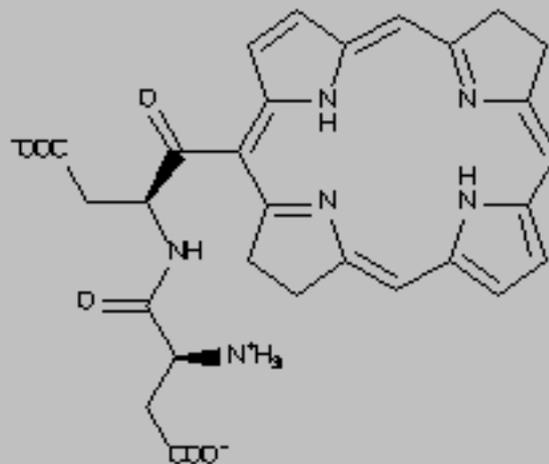
78139775
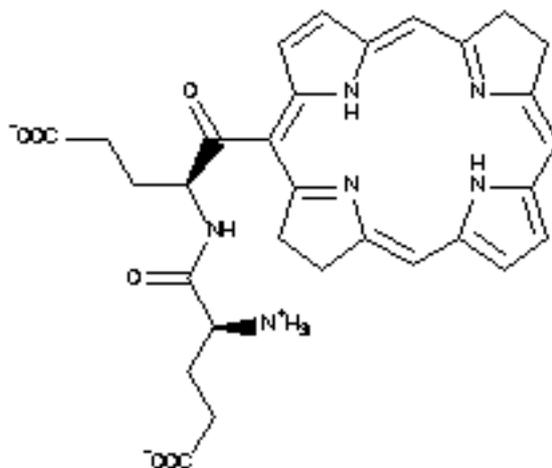
59380157
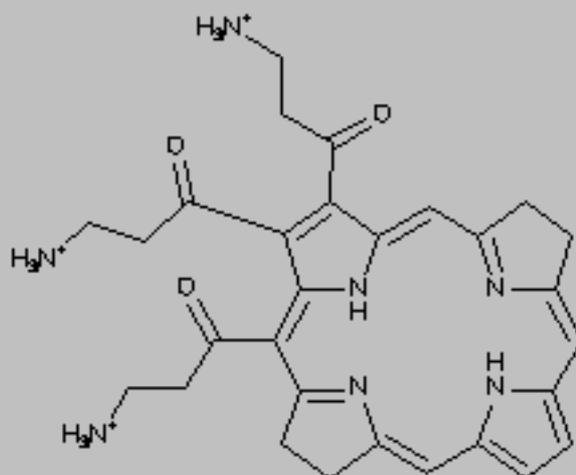
60076177
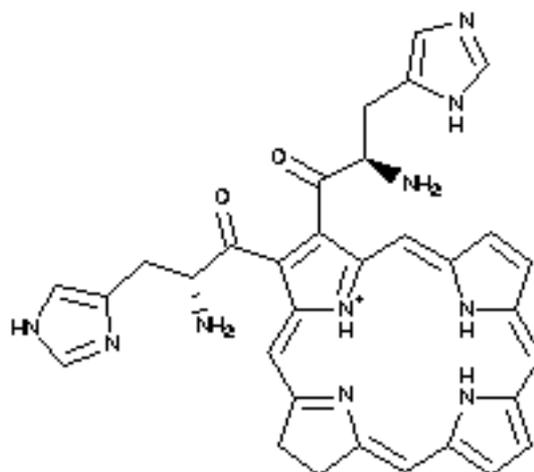
78222613
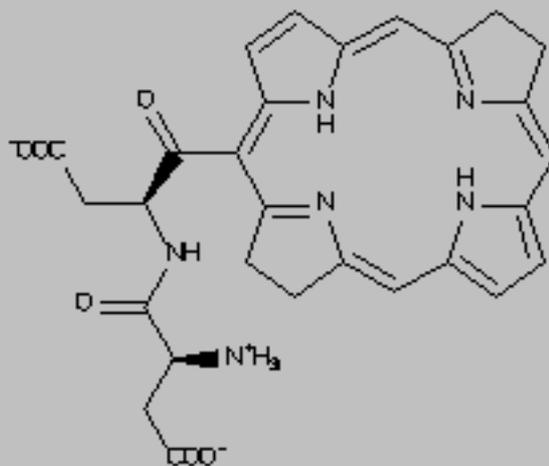
78315702
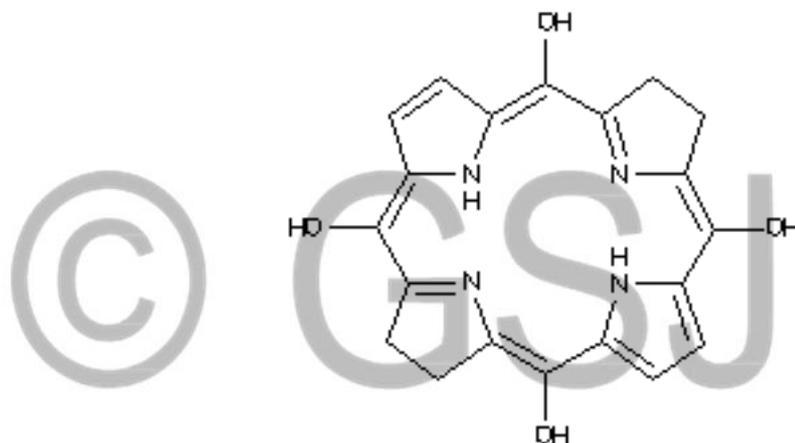
60194322
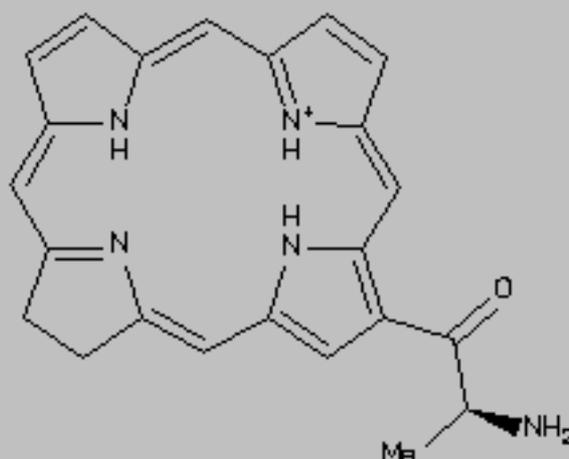
60194323
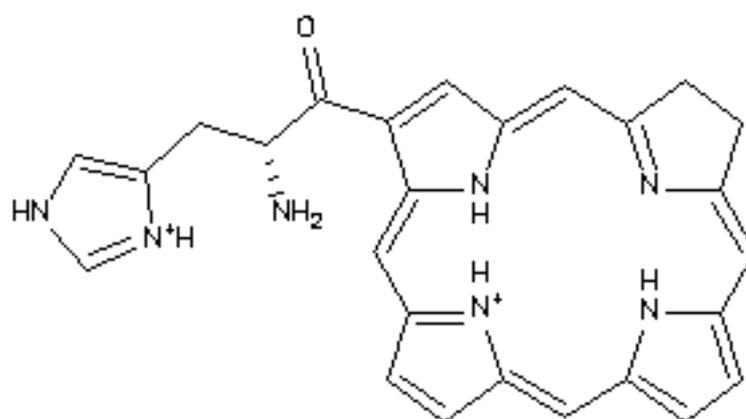
82190886
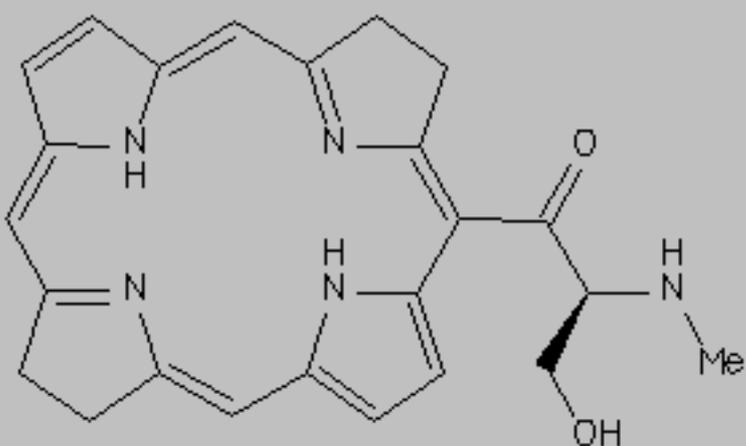
79824394
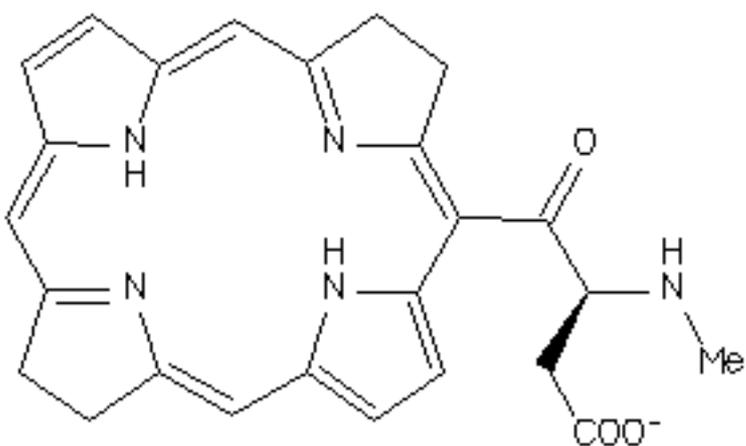
79231393
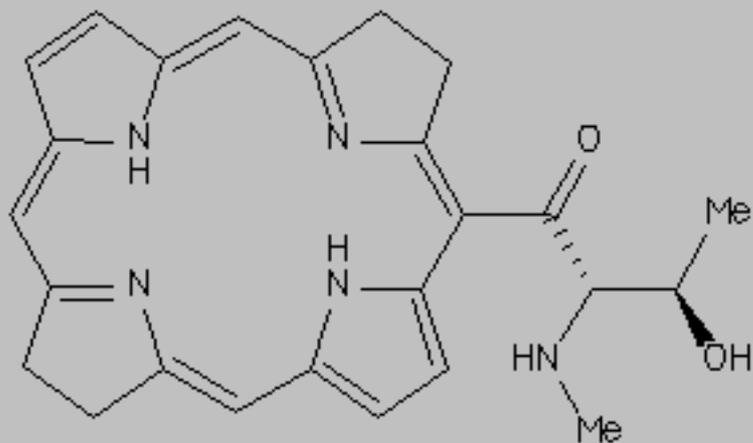
82054668
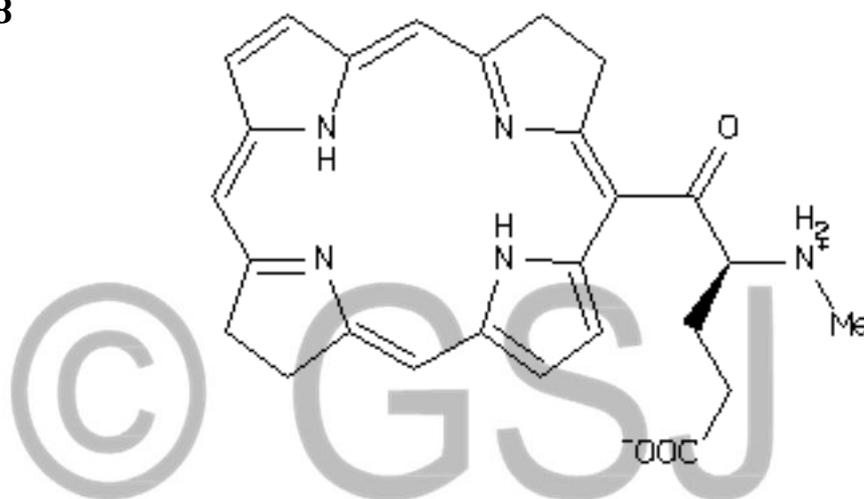
59651175
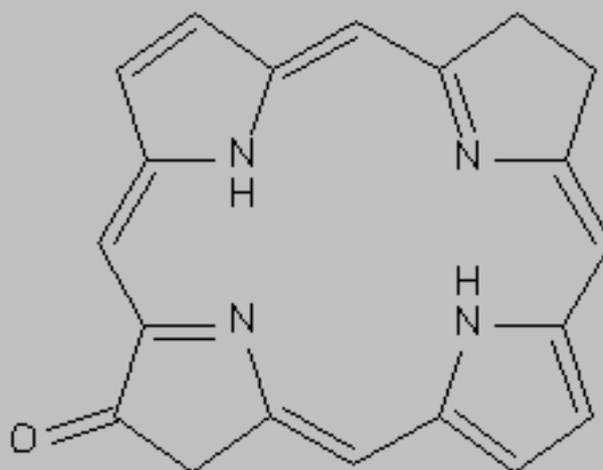
82153006
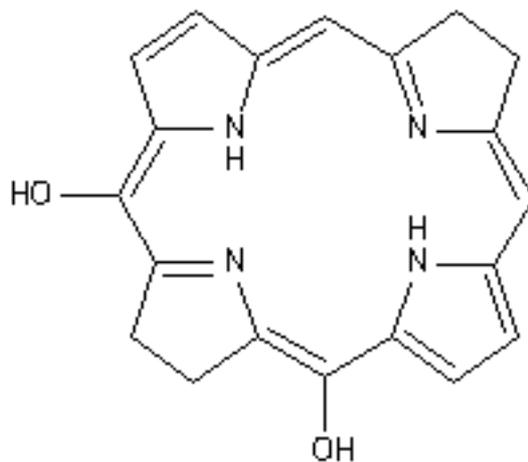
38792143
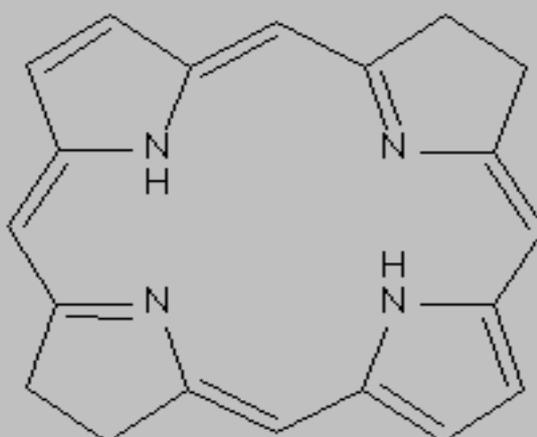
60308698
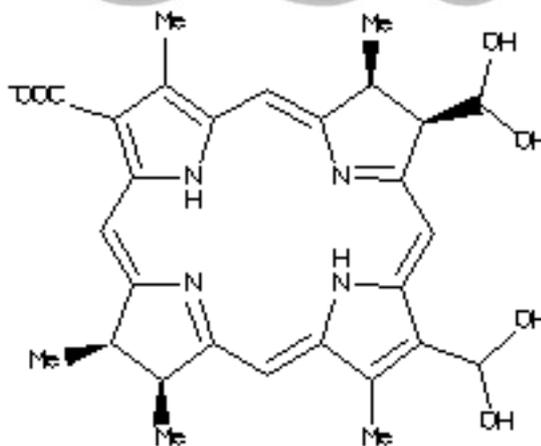
60308695
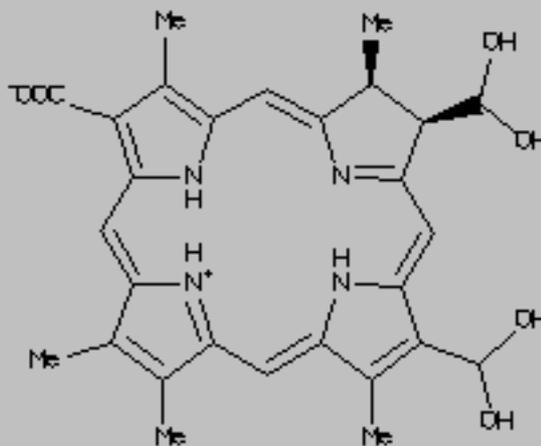
78015143
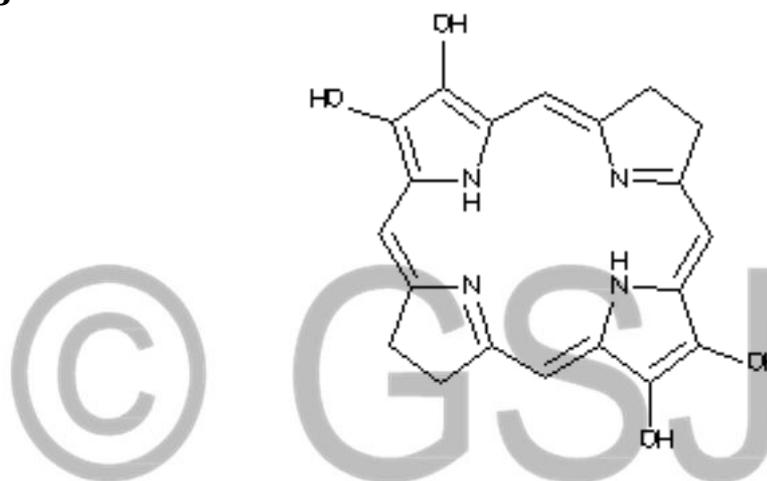
49783984
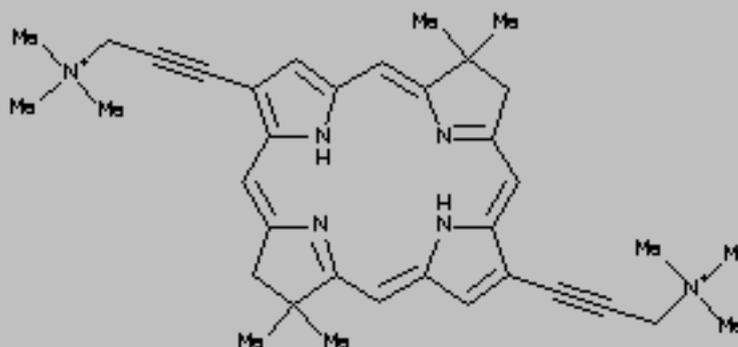
49784734
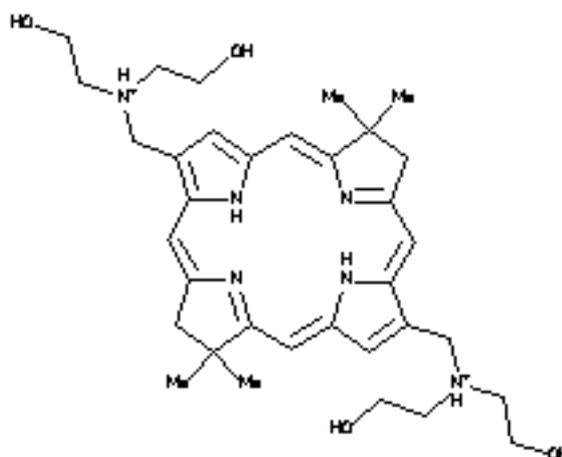
78206861
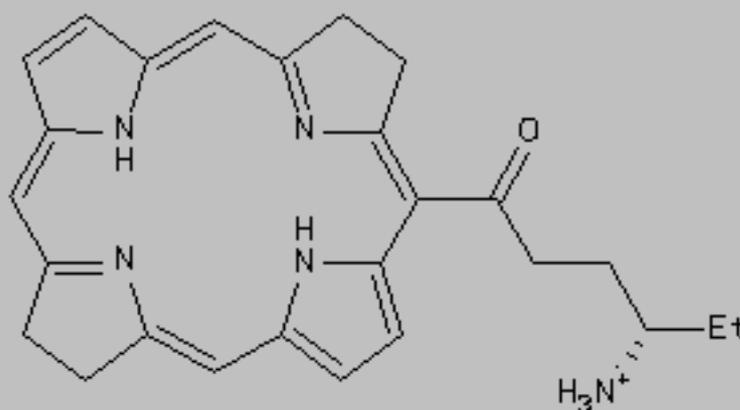
7. CONCLUSIONS.

The porphyrins bind specific residues at the octapeptide repeat region (OR) of the NH_2-term_PrP; this has been suggested to be a novel highly valuable strategy against Alzheimer's disease. The Lipinski's rule of 5 is a criterion to select the ligands with drug-like properties from the ZINC database, but lipohilicity and the aqueous solubility coefficients are the two cardinal molecular descriptors. The differences in lipophilicity coefficients for a single drug candidate may be ascribed to the numerous methods of calculation, but the fragmental method of Molinspiration can be trusted since it applies correction methods in order to compensate for intermolecular

interactions. The C-C and C-N bond lengths in porphyrins are influenced by exocyclic substitutions; this has been analysed with the harmonic oscillator method of aromaticity (HOMA), and the nucleus-independent chemical shift (NICS) models. Substituents on the porphyrin macrocyclic ring influence the values of molecular indices; electron withdrawing groups decrease lipophylicity coefficients and index of aromaticity, while electron donating groups increase lipophilicity coefficients and index of aromaticity. The ZINC database uses the fragmental method of Molinspiration in the calculation of the lipophilicity coefficients; consequently, we suggest the ZINC database classification of the 29 selected molecules will be used for molecular docking with NH_2-term_PrP, in order to select the best binding poses for molecular dynamics simulations.

8. REFERENCES.

(1)   Brenner GM, & Stevens CW. 2006, Pharmacology.2nd Edn. Philadelphia PA: Saunders Elsevier.

(2)   Barber & Wilson, Clinical Pharmacy Survival Guide, 2000.

(3)   www.ashp.org/doclibrary/p2418-resid 34-94

(4)   Prusiner SB (1998) P Natl Acad Sci USA 95 (7328): 47-52

(5)   Cong X, Casiraghi N, Rossetti G, Mohanty S, Giachin G, Legname G, & Caloni P (2013) Journal of Chemical Theory and Computation9(11): 5158-5167

(6)   Mangialasche F, Solomon A, Winblad B, Mecocci P, & Kivipelto M (2010) Lancet Neurol 9(7): 702-716.

(7)   Nicoll AJ, Panico S, Freir DB, Wright D, Terry C, Risse E, Herron CE, O'Malley T, Wadsworth JDF , Farrow MA, Walsh DM, Saibil HR, & Collinge J (2013) Nat Commun 4:2416

(8)   Caughey B, Caughey WS, Kocisko DA, Lee KS, Silveira JR, and Morrey JD (2006) Accounts Chem Res 39(9): 646-653.

(9)   Lamoen D, Parinello M, Chem. Phys. Lett. 1996, 248, 309-315 ; (b) Kozlowski PM, Jarzecki AA, Pulay P. J. Phys. Chem. 1996, 100, 7007 - 7013.

(10) Vogel E J. Heterocycle. Chem. 1996, 33, 1461 - 1487.

(11) Julg A. François P. Theoretical Chem. Acta 1967, 8, 249 - 259 ; (b) Bird C.W. Tetrahedron 1985, 41, 1409-1414; (c) Kruszewski J. T.M. Krygowski TM, Tetrahedron Lett. 1972, 3839 - 3842 ; (d) Krygowski J. Chem. Inf. Comput. Sci. 1993, 33, 70-78; (e) Krygowski TM, Ciesielski A, Bird WC, Kotschy A, ibid. 1995, 35, 203 - 210

(12) Krygowski TM, CyranÄski M. Tetrahedron1996, 52,1713-1722;(b) ibid. 1996, 52, 10 255 - 10 264.

(13) Julg A, François P. Theoretical Chem. Acta 1967, 8, 249 - 259 ; (b) Bird CW. Tetrahedron 1985, 41, 1409-1414; (c) Kruszewski J, Krygowski T.M, Tetrahedron Lett. 1972, 3839 - 3842; (d) Krygowski T.M, J. Chem. Inf. Comput. Sci. 1993, 33, 70-78;

(14) Pauling L . J. American. Chem. Soc. 1947, 69, 542 - 553

 (15) Krygowski TM, Anulewicz R, Kruszewski J, Acta Crystallogr. Sect. B 1983, 39, 732 - 739.

(16)   Boronat M.   Ortí E.   Viruela P.M.   TomaÂs F, J. Mol. Struct. Theoretical Chem 1997, 390, 149 - 156.

(17)   Frisch MJ, Trucks GW, Schlegel HB, Gill MW,   Johnson BG, Robb MA,   Cheeseman JR,   Keith T,   Petersson GA,   Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA, Gaussian 94, Pittsburgh, PA, 1995.

(18) Eldridge JL, Jackman LM, J. Chem. Soc. 1961, 859 - 866 ; (b) Dauben JH,   Wilson JD, J. L. Laity JL, Non-benzenoid Aromatics, Vol.2 (Ed: Snyder J), Academic Press, New York, 1971,

(19) Von P, Schleyer R, Maerker C, Dransfeld A, Jiao H, Eikema NJ , J. Am. Chem. Soc. 1996, 118, 6317-6318

(20)   Reese TS & Karnovsky MJ. J Cell Biol. #4, 207-217, 1967.

(21)   Nagy Peters H & Huethner I. Lab. Invest. 50, 313-322, 1984.

(22)   Clark E, O' Malley CD. The Human Brain and Spinal Cord. A historical study illustrated by writings from antiquity to the twentieth century, CA: University of California Press.1968.

(23) Begley DJ. ABC transporters and the blood-brain barrier. Curr Pharm Des 10; 1295-1312, 2004[PubMed]

(24) Ballabh P, Bruan A, Nedrgaad M. The blood-brain barrier; An overview: Structure, regulation. And clinical implications. Neourobiol Dis 16: 1-13, 2004.

(25)Van de Waterbeemd H. Camenisch G. Folkers G. Chretien JR. Raevsky OA. Estimation of blood-brain barrier crossing of drugs using molecular size, shape, and H-bonding descriptors. J Drug target 6: 151-165, 1998

(26) Feng MR. Assessment of blood-brain barrier penetration in silico, in vitro, and in vivo. Curr Drug Metab 3: 647-657, 2002.   .

(27)   Crone, C. Acta Physiol. Scand, 64, 407-417, 1965.

(28)   Betz. A.L, & Goldstein, G.W.A. Rev. Phsiol. 48, 241-250, 1986

(29)   Abolghasem J, Mohammed A.A , Ali S, Solubility Prediction Methods for Drug/Drug-like Molecules, 2008.

(30)   Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxico Methods 2000; 44:235-249,[PubMed:11274893]   .

(31)   Camer J A (2003) High throughput measurement of LogD and pka. In Artursson P, Lenernas H, and Van der waterbeemd H (Eds). Methods and Principles in Medicinal Chemistry 18(P.21-45) Veiheim Wiley-VCH.

(32)   Geladi, P, and Tasato ML (1990). Multivariate latent variable projection methods: SIMCA and Pls. In. Practical applications of Quantitative Structure-activity relationship (QSAR) in Environmental Chemistry and Toxicology (Karcher W, and Devilers J, Eds.) Kluwer Academic Publishers, Dordrecht. PP. 171-179

(33)   Devihers, J (1996), Neural Networks in QSAR and Drug Design. Academic Press, London, P, 284.

(34)   Kier, LB and Hall LH. (1986) Molecular connectivity in Structure-activity Analysis. Research Studies Press, UK, P. 257.

(35)   (a) Schwaighofer A, Schroeter T, Mika S. Laub J, ter Laak A, Sulzle D, Ganzer U, Heinrich N, Muller KR. Accurate Solubility prediction with error bars for electrolytes: A machine learning approach. J. Chem. Inf. Model. 2007, 47 (2), 407-424; (b) Obrezanova O, Csanyi G, Gola J M, Segall, MD. Gaussian Processes: A method for automatic QSAR Modeling of ADME Properties. J. Chem. Inf. Model. 2007, 47, 1847-1857

(36)   Blake JF. Chemoinformatics-Predicting the Physicochemical Properties of Drug-like Molecules. Curr. Opin. Biotechnol.2000, 11, 104-107

(37)   Wiley Liss, Inc. and the American Pharmacist Association J Pharm Sci 98:861-893, 2009.

(38)   Rekker, RF, and the Kort, HM, Eur. J Chem. 14, 479-488, 1979.

(39)   Nys, GG and Rekker, Eur, J Med Chem. 9, 361-375, 1974.

(40)   Viswananadhan VN, Ghose AK, Revanker GR, Robins RK, Atomic Physicochemical Parameters for Three-Dimensional Structures Directed Quantitative Structure Activity Relationship. J Chem Inf Comput Sci 9:163-172.

(41)   Wildman SA, Crippen GM. 1999. Prediction of Physicochemical Parameters by Atomic Contribution. J Chem Inf Comput Sci 39:868-873

(42)   Mauri A, Consonni V, Pavan M, Todeschini R. 2006. Dragon software: An easy approach to molecular descriptor calculation. Match-Commun Math Comput Chem 56: 237-248.

(43)   http://www.molinspiration.com/services/logp

(44)   Wang JM, Hou T J. Recent advances on in silico ADME modeling. Annual. Reports in Computational Chemistry 2009, 5, 101-127.

(45)   (a) Hou T, Wang J. Structure-ADME relationship: still a long way to go? Expert Opin. Drug Metab. Toxicol. 2008, 4 (6), 759- 770; (b) Tetko, I. V. Bruneau P, Mewes HW, Rohrer DC, Poda GI, Can we estimate the accuracy of ADME-Tox predictions? Drug Discov. Today 2006, 11 (15-16), 700-707

(46)   Jain N, Yalkowsky SH. Estimation of the aqueous solubility I: application to organic nonelectrolytes. J. Pharm. Sci., 2001, 90 (2), 234-252.

(47)
        http://sanjuan.cs.uwindsor.ca/ppipaw/index.php/help/document ation/59

(48)   F. Veber et al., J.Med.Chem. 2002, 45, 2615-2623.

(49) Palm K, Stenberg P, Luthman K, Artursson P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. Pharm. Res. 1997, 14, 568-571.

(50) Clark, D. E. Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption. J. Pharm. Sci. 1999, 88, 807-814.

(51) Stenberg P, Luthman K, Ellens H, Pin Lee Ch, Smith PL, Lago A, Elliot JD, Artursson P. Prediction of the Intestinal Absorption of Endothelin Receptor Antagonists Using Three Theoretical Methods of Increasing Complexity. Pharm. Res. 1999, 16, 1520-1526.

(52) Palm K, Luthman K, Ungell AL, Strandlund G, Artursson P. Correlation of Drug Absorption with Molecular Surface Properties. J. Pharm. Sci. 1996, 85, 32-39.

(53) Palm K, Luthman K, Ungell AL, Strandlund G, Beigi F, Lundahl P, Artursson P. Evaluation of Dynamic Polar Molecular Surface Area as Predictor of Drug Absorption: Comparison with Other Computational and Experimental Predictors. J. Med. Chem. 1998, 41, 5382-5392.

(54) Krarup, LH, Christensen IT, Hovgaard L, Frokjaer S. Predicting Drug Absorption from Molecular Surface Properties Based on Molecular Dynamics Simulations. Pharm. Res. 1998, 15, 972-978.

(55) Stenberg P, Luthman K, Artursson P. Prediction of Membrane Permeability to Peptides from Calculated Dynamic Molecular Surface Properties. Pharm. Res. 1999, 16, 205-212.

(56) Clark DE. Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 2. Prediction of Blood-Brain Barrier Penetration. J. Pharm. Sci. 1999, 88, 815-821.

(57) Kelder J, Grootenhuis PD, Bayada DM, Delbressine LP, Ploemen JP. Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. Pharm. Res. 1999, 16, 1514-15

(58) WorldDrugIndexDatabaseWDI97; DerwentPublicationsLtd.

(59) Clark DE. Rapid Calculation of Polar Molecular Surface Area and Its Application to the Prediction of Transport Phenomena. 1. Prediction of Intestinal Absorption. J. Pharm. Sci. 1999, 88, 807-814.

(60) Sadowski J, Gasteiger J. From Atoms and Bonds to Three-dimensional Atomic Coordinates: Automatic Model Builders. Chem. Rev. 1993, 93, 2567-2581.

(61).Molinspiration.com www.molinspiration.com 2004.

(62). Wang R, Fu Y, Lai L. A New Atom-Additive Method for Calculating Partition Coefficients. J Chem Inf Comput Sci 1997;37:615–621.

(63) Ihlenfeldt WD, Voigt JH, Bienfait B, Oellien F, Nicklaus MC. Enhanced CACTVS browser of the Open NCI Database. J Chem Inf Comput Sci 2002;42:46–57. [PubMed: 11855965]

(64)Caugheys B, Caugheys WS, Kocisko DA, Lee Ks, Siveiva JR, & Morrey JP (2006), Accounts Chem Res(9): 646-653.