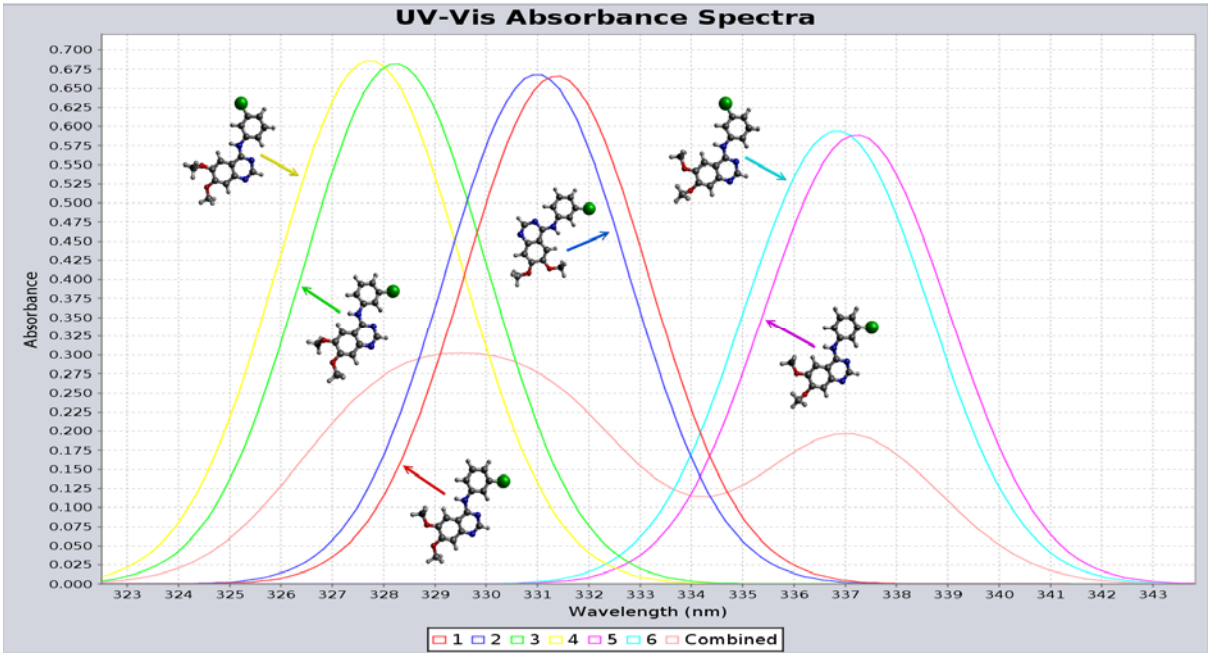
Accelerating optical reporting for conformation of tyrosine kinase inhibitors in solutions

Feng Wang[1]* and Vladislav Vasilyev[2]

[1]*Department of Chemistry and Biotechnology and Centre for Translational Atomaterials, School of Science, Swinburne University of Technology, Melbourne, Victoria 3122, Australia*

[2]*National Computational Infrastructure, Australian National University, Canberra, ACT 0200, Australia*

**E-mail addresses: fwang@swin.edu.au (F. Wang)*



Abstract

It has been a challenge in automated analysis of medical and chemical knowledge to extract represent quantitative structure–activity relationship (QSAR) using intelligent computing in drug discovery. One of many domain-specific bottlenecks in drug discovery is robust conformation search in three-dimensional (3D) space for flexible drug candidates. The process involves researchers and machines working together to achieve their own strengths for greater outcome. The present study has been developing a method for conformational sampling conformers in the class of 4-anilinoquinazoline derivatives for epidermal growth factor receptor (EGFR) tyrosine kinases inhibitors (TKIs). We use AG-1478 to demonstrate how the new intelligent computing method helps to quantum mechanically determine 22 target drug conformer clusters and their properties from conformational sampling, based on density functional theory (DFT) method, time-dependent (TD)-DFT in solvents and clustering analysis (CA). The UV-vis spectra of the preferred conformers agree well with earlier experimental measurements in which the conformer dependent UV-Vis spectral shift of AG-1478 can be as large as approximately 15 nm. We are further developing this method to study and design new 4-anilinoquinazoline derivatives of EGFR TKIs.

Keyword: TD-DFT calculations; UV-Vis spectroscopy; 4-anilinoquinazoline derivatives; EGFR-TKIs; Quantum mechanical conformational sampling; Intelligent computing; Quantum clustering (QC); AG-1478.

1. Introduction

A number of studies revealed that 4-anilinoquinazoline derivatives are potent and highly selective inhibitors of epidermal growth factor (EGFR) phosphorylation at the ATP binding site. Tyrosine kinases (TK), together with serine/threonine kinases and histidine kinases, are one of the three identified enzyme classes based on their catalytic specificity. (Ismail, 2016) Anilinoquinazolines are the most developed class of drugs that inhibit EGFR kinase intracellularly, (Bridges A. J., 1996) (Aparna, 2005) and drug candidates in this class have already reached various phases of clinical trials, such as Gefitinib, Erlotinib, Lapatinib, Afatinib, Vandetanib, Icotinib, Dacomitinib (PF-00299804), (Kabir, 2020) DP150335 and AG-1478. (Khattab M. A., 2016) The present study will focus on tyrphostin AG-1478 (*N*-(3-chlorophenyl)-6,7-dimethoxy-4-quinazolinamine, see Figure 1) as a model anilinoquinazoline to develop a robust quantum mechanically based conformational search for this class of TKIs. As AG-1478 has the important chemical features required for its binding affinity (Wilson, 2015) and potent activity toward EGFR, includes the topology of quinazoline nitrogens and -NH- linker. (Bridges A. J., 1999) The UV-Vis absorption spectroscopy of conjugated aromatic core (4-anilinoquinazoline moiety) of AG-1478 has been studied as optical reporting to probe its conformational changes quantum mechanically in solution (Khattab M. A., 2016) (Kabir, 2020) and experimental measurements. (Clayton, 2005) These results show that absorption in UV-vis spectra and particularly photoemission in fluorescence spectra are highly sensitive to solvent polarity, which is ideal for validation and training of the method to be developed.

Conformational sampling is still major hurdles to effective drug design methods in both the classical and QM approaches. The high demand of exploring and analyzing big data has encouraged the use of data-hungry machine learning algorithms such as intelligent computing. Over the past decade, there has been a remarkable increase in the amount of available compound activity and biomedical data. (Chen, 2018) Most of the algorithms in drug design are based on geometry, for the binding sites always locate on the concave surface of protein. (Wang, 2013) Estimating the range of three-dimensional (3D) structures (conformations) that are available to a molecule is a key component of computer-aided drug design. (Habgood, 2020) As drug design distinguishes itself from clear-cut engineering by the presence of error, nonlinearity and seemingly random events. (Schneider, 2016) Quantum mechanical simulation offers improved accuracy over forcefield methods at a high computational cost. It is a matter of balance between accuracy and costs. Three major challenges which were summarised by Habgood et al (Habgood, 2020) include (1) development a good method to generate ensembles

of a molecule's bioactive conformation; (2) rational analysis and modification of a pre-established bioactive conformation; and (3) approximation of real solution phase conformational ensembles in tandem with spectroscopic data such as IR, NMR and UV-vis. Many conformational search methods focused on the reproduction of crystal structures (i.e. bioactive conformations) (Cappel, 2015). While crystal structures usually provide a single 3D structures, they do not cover the full conformational space of a flexible molecule with multiple rotatable single bonds or unsaturated ring systems. (Habgood, 2020) The solution is to develop computational tools to generate a set of structures—a conformational ensemble—in silico. The present study targets (1) and (3) and is further developed into (2), i.e., design new bioactive anilinoquinazoline derivatives.

In order to achieve comprehensive conformational search with high quality results and computational efficiency, one needs intelligent computing (Foloppe, 2009) (Cappel, 2015). For example, random torsional angle changes and random coordinate changes using molecular mechanics (MM) (Mohamadi, 1990) (Chang, 1989) (Ferguson, 1989), distance geometry (Crippen, 1978) (Izrailev, 2006), rule-based methods (Rusinko, 1989) (Gasteiger, 1990), knowledge-based approaches from the Protein Databank (PDB) and Cambridge Structural Database (CSD) (Hawkins, 2020), low mode search for cyclic and acyclic alkanes and cyclic peptides (Kolossva´ry, 1996), or low mode MD for macrocycles and protein loops (Labute, 2010). Cappel et al (Cappel, 2015) used the common scaffold alignment (CSA) approach which focused on the coordinates of the atoms in the common scaffold for drug screening. Among the large number of different approaches of conformational ensemble generation of bioactive molecules, the simplest form of conformational 'ensemble' generation is a potential energy scan for a single dihedral angle, assessing the stability of a subset of conformations based around one structurally significant coordinate, such as AG-1478 (Khattab M. A., 2016) and Dacomitinib (PF-00299804) (Kabir, 2020). The planar (global minimum structure) and twisted non-planar (local minimum structure) conformers identified by quantum mechanical (QM) calculations through potential energy scans of selected flexible singles bonds. In this method, the researchers need to determine the rotatable single bonds and if the potential energy scan is not relaxed (in which each step is a single point calculation), it is unlikely to produce a comprehensive conformer search for molecules with multiple rotatable bonds as there can be hundreds of such conformers (Mackey, 2015).

The challenges in the case of the flexible ligands are that a large number of conformers needs to be considered. It is, however, almost impossible to include and consider all possible conformers. The critical issues in this regard are (Mackey, 2015): a) to include all possible structures at the conformational potential energy local minima; b) to deal with energy degenerated conformers, and c) to calculate too many possible conformers for flexible ligands with many rotatable bonds. In this study, we developed a robust quantum mechanical conformational search along with the rotatable flexible single bonds of a molecule combined with the optimisation calculations using intelligent computing.

2. Methods and computational details

2.1 Robust conformation search

To partially address the issues in conformational search (Mackey, 2015), we developed a simple but effective intelligent computing approach. It automatically determines all rotatable bonds of a given molecule or a drug candidate and generates a predetermined number of conformers using a random sampling in the conformational space of the molecule. It then produces input and batch submits for the related QM calculations including optimization and following property calculations; sorts out the large number of output files of the conformers, groups the conformers based on their total energy and geometry using a cluster analysis (Xu, 2015); and finally calculates target or measurable properties such as NMR and UV-vis spectra for a selected number of top ranked conformers for comparison.

To illustrate above mentioned method, the model TKI AG-1478 is employed. The structure of AG-1478 is given in Fig 1.

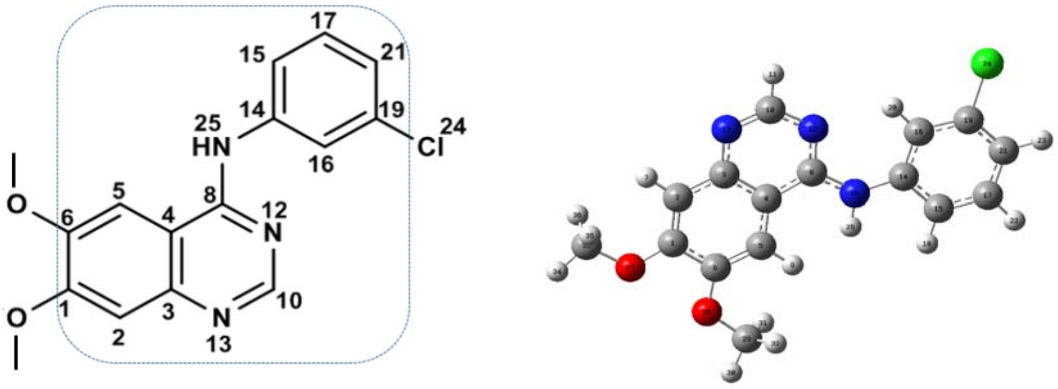


Figure 1. The chemical structure and nomenclature of AG-1478. (a) The structure in the dotted line box is the 4-anilinoquinazoline scaffold. (b) The global minimum structure of AG-1478 (Khattab M. A., 2016). Note that the atomic labels are not the IUPAC label.

In the approach, robust determination of the rotatable bonds (the dihedral angles) for a target ligand (AG-1478 in this case) uses two simple rules:

(a) all single bonds of a ligand which are not part of the ring(s) are rotatable;

(b) neither atoms which are joined by a single bond (A-B) can contain three identical substituents (e.g., three hydrogen atoms ($-CH_3$), three methyl groups ($-C(CH_3)_3$)) or else rotation about the single bond (A-B) will be irrelevant.

In the case of AG-1478 (see Fig 1), the rotatable single bonds are therefore restricted to $C_{(8)}$-$N_{(25)}$ and $C_{(14)}$-$N_{(25)}$, and the pair of O-C_{ring} bonds which are responsible for the conformational potential energy surface of AG-1478. The two $O-CH_3$ single bonds on the side of the quinazoline core are excluded for rotational single bonds applying rule (b). The method randomly assigned each of the four rotatable single bonds to a value between 0° and 360°. In the present study the number of randomly generated conformers of AG-1478 was set to 300, which is sufficiently enough for such a small flexible molecule as AG-1478.

To optimize the randomly generated conformers of AG-1478 at the quantum chemical level a program generated the input files for the Gaussian16 (Fisch, et al., 2016), which then were batch submitted for the optimization at the DFT level of theory to relax to the closest local energy minima. To sort out and rank the resulted collection of the optimized AG-1478 conformers a Cluster Analysis (CA) was used. As a criterion of proximity (similarity) between two conformers the Euclidian-type distance function was employed

$$D_{p,q} = \sqrt{\sum_{i=1}^{k} (p_i - q_i)^2} \qquad (1)$$

where the property variables $\boldsymbol{p_i}$ and $\boldsymbol{q_i}$ which represent each structure are: (a) the total energy; (b) the eigenvalues of the interatomic distance matrix, and optionally (c) the dipole moment.

The main advantage of using the eigenvalues of the interatomic distance matrix over the dihedral angle values is that the overall and the local symmetries of a molecule are automatically taken into account. For example, some rotations around the bonds generate the identical structures which are different in the dihedral angle space but have the same

eigenvalues of the interatomic distance matrix. As a result, a task of grouping performed by the Cluster Analysis offers new perspectives in machine learning technologies (Liu, 2016) by discovering patterns in the large data sets and by relevant information retrieval.

The 'strain energy' is a potential energy cut-off ΔU_{max} above the lowest the 'global minimum' energy of the ensemble. It is assumed that all generated conformations up to a given strain energy (i.e. the preferred conformers) are physically feasible and should be considered in further calculations. (Habgood, 2020) Any conformations higher in energy than the strain energy (ΔU_{max}) are not attainable, so that it is quite improbable that they can take part in any supermolecular complex. In the present study the strain energy, ΔU_{max}, was set up at 10.5 kcal/mol, as the maximum calculated strain energy of bioactive conformations needs to be between 5 and 25 kcal/mol. (Foloppe N, 2019) To rank the preferred conformers, their total energies were employed which were also used to estimate the probability distribution for each local minimum using a Boltzmann distribution (Landau, Lev Davidovich & Lifshitz, Evgeny Mikhailovich (1980) (Landau, 1976),

$$p_i = \frac{n_i}{Q} e^{-\varepsilon_i / kT} \tag{2}$$

where $\boldsymbol{p_i}$ is the probability of state $\boldsymbol{i}$, $\boldsymbol{\varepsilon_i}$ is the relative energy of state $\boldsymbol{i}$, $\boldsymbol{n_i}$ is the number of structures with the energy $\boldsymbol{\varepsilon_i}$ (i.e., cluster size), $\boldsymbol{k}$ is the Boltzmann constant, and $\boldsymbol{T}$ is the temperature of the system (in the present study, T=300 K). The normalization denominator $\boldsymbol{Q}$ is the canonical partition function,

$$Q = \sum_{i=1}^{M} e^{-\varepsilon_i / kT} \tag{3}$$

where $\boldsymbol{M}$ is the number of all states accessible to the system.

The above procedure implemented in the method presents an intelligent computing with minimal human supervision which partly resolves the issues concerned Mackey (Mackey, 2015), as the preferred conformers of a drug molecule are obtained using cluster analysis from a larger number of conformers through a systematic search for potential energy local minima, which are calculated accurately using DFT based quantum mechanical calculations.

2.2 Quantum mechanical computational details

In the present study, a predetermined 300 randomly generated conformers for AG-1478 are obtained from rotations of four rotational bonds of $C_{(8)}$-$N_{(25)}$ and $C_{(14)}$-$N_{(25)}$, O-$C_{(1)}$ and O-$C_{(6)}$ bonds. All generated geometries were reoptimized at the B3LYP/6-311+G(d) level of theory in dimethyl sulfoxide (DMSO), water and methanol solvents, respectively, using an implicit polarizable continuum model (PCM). (Mennucci, 2002) A total of 22 local minimum structures (including a global minimum one) for the AG-1478 were obtained from the 300 randomly generated geometries of AG-1478.

The UV-Vis spectra for each of the 22 clusters of local minimum structures were then calculated using time dependant density functional theory (TD-DFT) and the same B3LYP/6-311+G(d) model in the three solvents (DMSO, water and methanol). The absorption UV-Vis spectra of the 22 low lying AG-1478 conformers were each calculated for the lowest 60 singlet excited states while the absorption profiles were simulated using a half-width half-maximum (HWHM) of 4428 nm (0.28 eV). All calculations were performed using Gaussian 16 computational chemistry packages (Fisch, et al., 2016) at the National Computational Infrastructure (NCI). The conformers were compared using their alignment and their quantum mechanically calculated UV-vis spectra in the present study were compared with the measurements in the same solvent (Khattab M. A., 2016).

3. Results and discussion

3.1 Preferred conformers for AG-1478

It is well known that potency of drugs significantly depends on its 3D structure. The bioactive conformations are not necessarily at their energy minima in the gas or solution phases (the preferred conformers). The binding sites are very often located on the concave surface (Wang, 2013) so that binding to a protein can distort a ligand away from the local minima (Cappel, 2015). For example, it was discovered that the isolated planar conformer of AG-1478 is more stable than the twisted one (Khattab M. A., 2016). Also, our preliminary study discovered that a ligand can be a non-planar in a complex with protein. In addition, the anisotropic properties, such as the shape, dipole moment and orbital distributions, are very different between the isolated planar and non-planar conformers. So, it is a significant

challenge to systematically search all possible or at least meaningful conformers of AG-1478 using the conventional 1D potential energy scans. (Khattab M. A., 2016) (Kabir, 2020)
The present study develops a more versatile method because it uses a set of automated unsupervised operations (such as Cluster Analysis) with intelligent computing. Of the total 300 randomly generated conformers of AG-1478 which converged to the closest local minima on the potential energy surface after the DFT quantum mechanical optimization (except for the non-converged cases because of the interatomic clashes). The present study determined 185 local minimum conformers for clustering based on the 0.1 proximity (similarity) criterion (see Eq.1), which leads to 22 clusters (corresponding to 22 groups of conformers). All the conformers which were sorted out using the Cluster Analysis are presented in Table S1 of the Supplementary Materials. Table 1 presents the top 12 clusters of preferred conformers, which represent a total of 129 stable conformers of AG-1478 in the DMSO solvent.

Evidently, the drug candidates tend to a higher binding affinity toward the target protein if their binding conformation matches one of the low-energy solution structure, which is of practical interest to produce a candidate bioactive conformation. (Habgood, 2020) As seen in Table 1, the preferred AG-1478 3D conformers in DMSO solution can be very different, reflecting by the variation of dipole moments of the conformers from 3.31 Debye in Cluster 7 to 11.1 Debye in Cluster 5. The statistical weights calculated based on the Boltzmann distribution at temperature 300 K (last column of Table 1, see Eqs 2 and 3) indicate that the conformational space of AG-1478 is dominated by the first three preferred conformer clusters which account for 67.5%, 26.8% and 3.9%, respectively, with the other preferred conformer clusters having minor weights in the DMSO solvent.

Table 1: The twelve top AG-1478 conformers with the lowest energies in DMSO.

Cluster (conformer)	Total Energy (E_h)	Relative Energy (Kcal/mol)	Dipole Moment (Debye)	Cluster Size	Weight[§] (%)
1	-1392.755766	0.000	8.4504	10	67.646
2	-1392.754819	0.594	3.6481	11	26.819
3	-1392.752620	1.975	8.1783	17	3.871
4	-1392.751670	2.570	5.3047	11	0.900
5	-1392.751092	2.933	11.1002	9	0.395
6	-1392.750163	3.516	5.812	19	0.306
7	-1392.748488	4.567	3.3103	12	0.032
8	-1392.748387	4.631	4.6733	10	0.024
9	-1392.747203	5.373	9.2452	5	0.003
10	-1392.746252	5.970	4.3828	5	0.001
11	-1392.744797	6.884	4.1492	6	0.000

12	-1392.744774	6.898	4.4065	14	0.001

§ Estimation of the probability distribution of a given conformer is in %.

Figure 2 compares the top twelve conformer clusters of AG-1478 reported in Table 1 with respect to their relative strain energy (in kcal/mol), dipole moment (in D), cluster size and probability distribution (in %), together with the optimized structures in DMSO solution. The strain energy under cutoff (10.5 kcal/mol) is related to the statistical weight of the conformers - the larger a blue column in Figure 2, the less probability to find a given conformers (smaller the green columns). As a result, a global minimum structure of the isolated AG-1478 is a dominant one (67.65%) and the cumulative statistical weight of the top four preferred conformer clusters account for more than 99% (see Table 1). The dipole moment of the conformers varies largely, depending on the conformers in 3D structure. For the illustrative purpose, the dipole moments of conformers are plotted pointing downwards in Figure 2.

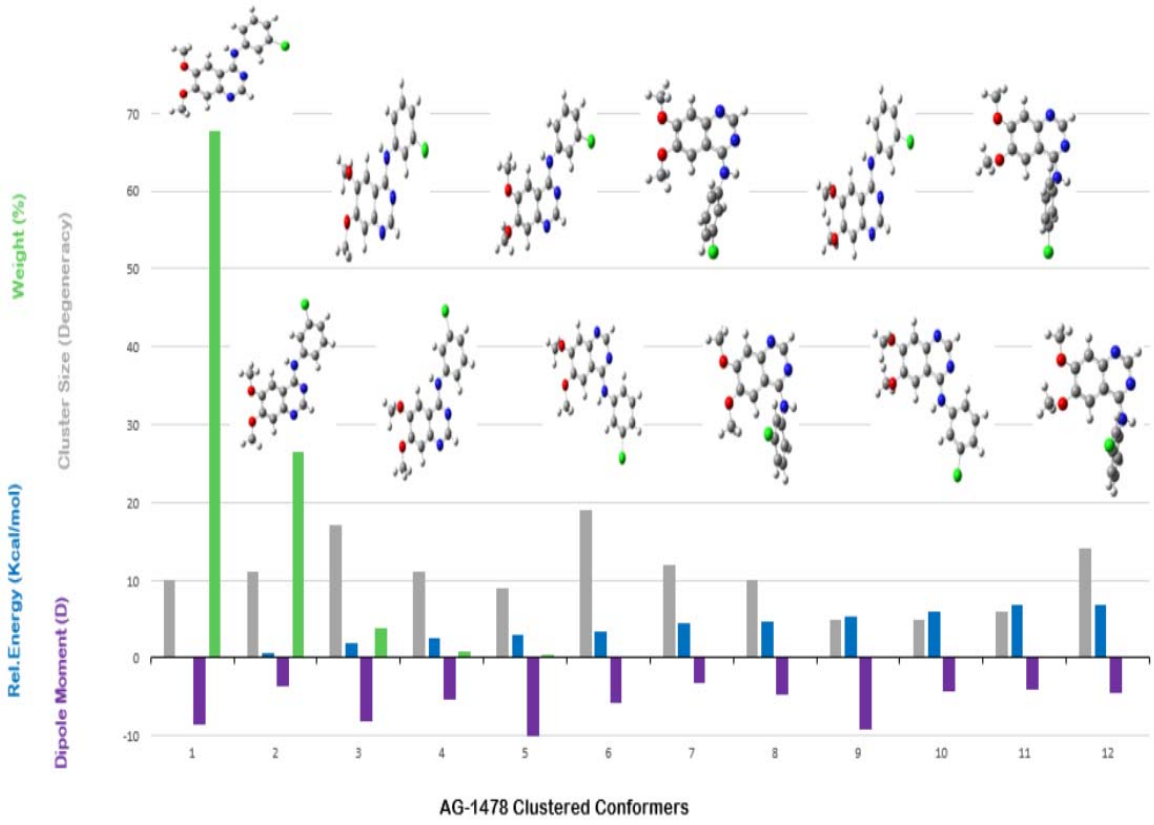


Figure 2 Comparison of the top twelve AG-1478 preferred conformers above the strain energy cut-off (10.5 kcal/mol) (blue), dipole moment (D, purple), cluster size (grey) and probability distribution (in %, green), together with the 3D optimized structures in DMSO solution.

The 3D structures of the twelve preferred AG-1478 conformer clusters are also shown in Figure 2. The top ranked energy favourite conformer clusters (up to Cluster 6) are presented by planar

structures, i.e., the quinazoline core and chlorophenyl lie approximately in the same plane, with varying positions of the chlorophenyl (flipping) and the methoxy chains ("tails") on the other side of the quinazoline core. The planes of the quinazoline core and chlorophenyl of the AG-1478 conformers start to twist by means of rotations around the -$N_{(25)}$H- linker and become non-planar conformer clusters at higher strain energy, such as Clusters 6 and 7 in Figure 2. In the present study, the whole molecule is allowed to relax to respond to the rotations around the all rotatable bonds rather than relaxes along a fixed single bond in the 1D rotational potential energy scan. (Khattab M. A., 2016) A number of low energy planar conformers of AG-1478 are obtained from flipping of the chlorophenyl ring, before the non-planar AG-1478 conformers in Clusters 7 and 8 are located. As shown in Figure 2 Cluster 8 is similar to the less stable twisted AG-1478A conformer identified earlier using 1D potential energy scan. (Khattab M. A., 2016) It is difficult to locate all possible conformers of AG-1478 without searching the entire potential energy surface of a flexible ligand. However, the important preferred conformers discovered in our previous study, AG-1478A (twisted) and AG-1478B (planar) conformers, are determined automatically in the present study as a result of the unsupervised conformational search using intelligent computing. One needs to note that the strain energy difference between the conformers was calculated to be 1.58 kcal/mol in gas phase using the 1D potential energy scan. (Khattab M. A., 2016)

3.2 Superposition of the preferred AG-1478 conformers in solvents

Of the twenty-two local minimum AG-1478 conformer clusters, the planar conformers, in which both rings are lying in the same plane, dominate the distribution on the conformational landscape in DMSO solution. The top six preferred conformer clusters, plus conformer Clusters 9 and 10 in Table 1 and Table S1 (DMSO), take the planar conformation, that is, the dihedral angle $\angle C_{(4)}C_{(8)}N_{(25)}C_{(14)}$ (where $C_{(14)}$ is connected to $C_{(l)}$ chlorophenyl ring, see Figure 1(a)) is either 0° or 180°. Figure 3 shows the structural superposition of all 22 local minimum conformer clusters of AG-1478, where all structures are aligned to the plane formed by three atoms $C_{(10)}$, $C_{(4)}$, and $C_{(1)}$ in the quinazoline core. It can be seen that the conformational changes occur at the two methoxy side chains on $C_{(1)}$ and $C_{(6)}$, which connect to the methoxy (-OCH_3) groups and flipping of the chlorophenyl ring in Figure 3 (also see Figure 1(a)).

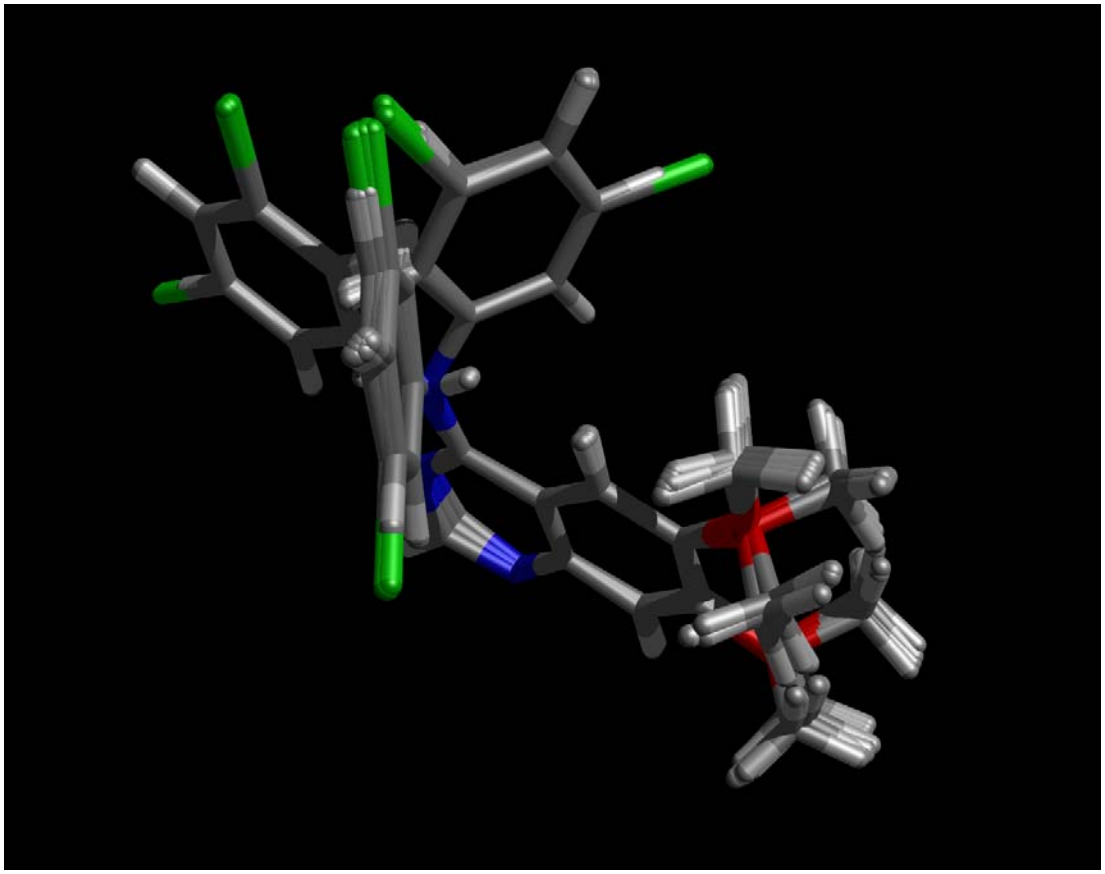
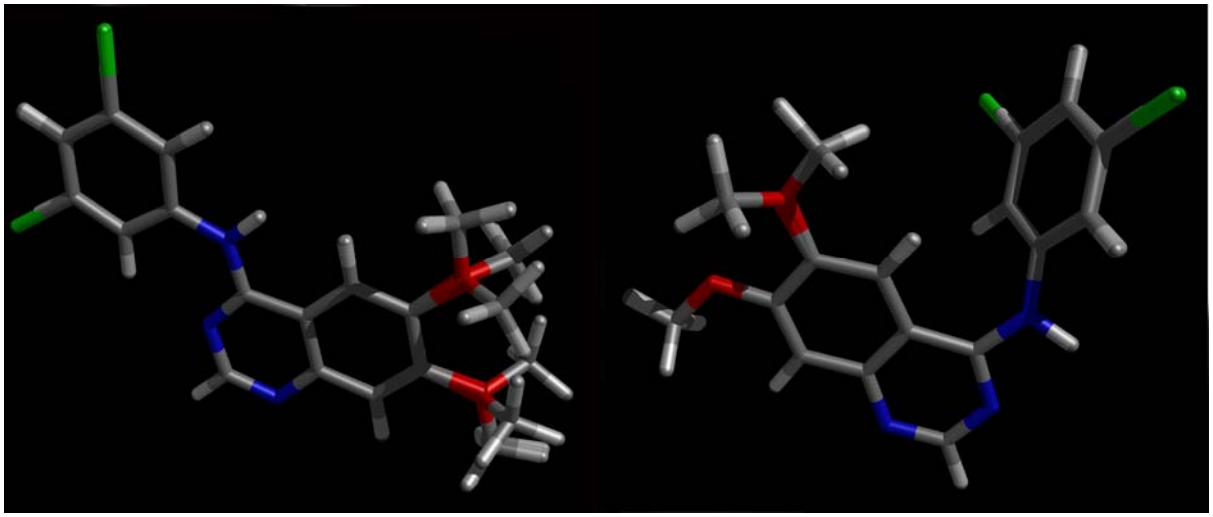


Figure .3 Superposition of all twenty-two local minimum (preferred) conformers of AG-1478. Three atoms of the quinazoline core, $C_{(10)}$, $C_{(4)}$, and $C_{(1)}$, were used for the superimposition (See Figure 1(a)).

To further explore the preferred conformers in the planar and non-planar conformations, Figures 4(a) and (b) align the planar and non-planar AG-1478 structures, respectively. The planar conformers including Clusters 1-6 and Clusters 9-10 with the strain energy less than 6.0 Kcal/mol vary either flipping positions of the quinazoline core and the chlorophenyl ring or the flexible methoxy side chains as shown in Figure 4(a). A fewer non-planar conformers, Clusters 7-8 and 11-12 appear in the range of strain energy cut-off, with the chlorophenyl ring being out of the quinazoline core plane are shown in Figure 4(b), including the twisted conformer found in our previous study. (Khattab M. A., 2016)

The non-planar AG-1478 conformers exhibit lager strain energy than the planar ones in DMSO, water and methanol solvents. The results from the conformation search in water and methanol are given in Tables S2 and S3, respectively in the Supplementary Materials. Different solvents do not substantially change the distribution of the local minima structures in 3D space when the implicit solvent model is applied. As seen from the conformer alignment in Figures 4(a)

and (b), rotating a central bond in the AG-1478 by a few degrees exhibits virtually small impact on the strain energy but the orientations of the chlorophenyl ring in Figure 4(a) and the methoxy tails in Figure 4(b) have much more marked effect. It also suggests that the long chain alkoxy groups at the $C_{(1)}$ and $C_{(6)}$ positions of quinazoline core favours the activity. (Noolvi, 2013) This finding supports the experimental observation that the presence of the bulky electronegative groups at the $C_{(1)}$ and $C_{(6)}$ positions leads to increase in activities of the compounds. It is evident from the molecular docking studies that the hydrophobic groups substituted at these positions of the quinazoline core result in the strong hydrophobic interactions with the nonpolar active site residues and likely may enhance the EGFR kinase inhibition.



(a) Planar conformers (b) Twisted conformers

Figure 4 Superposition of the conformers of AG-1478 for the planar (a) and twisted (b) configurations. Three atoms of the quinazoline core, $C_{(10)}$, $C_{(4)}$, and $C_{(1)}$, were used for the superimposition. (See Figure 1(a)).

Table 2 compares the properties of the preferred AG-1478 conformers in two protic (water and methanol) and aprotic (DMSO) solvents. As seen in the table, the top preferred AG-1478 conformer clusters exhibit very similar strain energies, dipole moments and the estimated statistical weight. As a result, in the next section we will concentrate on the properties of the AG-1478 conformers in the DMSO solvent.

Table 2: Comparison of the properties of top ranked AG-1478 conformers in DMSO, water and methanol.*

No	ΔE	μ	%	ΔE	μ	%	ΔE	μ	%
	DMSO			H_2O			MeOH		

1	0.000	8.450	67.65	0.000	8.487	68.30	0.000	8.411	66.98
2	0.594	3.648	26.82	0.607	3.657	26.48	0.580	3.639	27.19
3	1.975	8.178	3.87	2.014	8.205	3.65	1.932	8.150	4.12
4	2.570	5.305	0.90	2.622	5.307	0.83	2.516	5.302	0.98
5	2.933	11.100	0.40	2.940	11.152	0.39	2.925	11.046	0.35
6	3.516	5.812	0.31	3.538	5.842	0.28	3.493	5.781	0.32
7	4.567	3.310	0.03	4.612	3.320	0.03	4.520	3.300	0.03
8	4.631	4.673	0.02	4.672	4.703	0.02	4.586	4.643	0.03
9	5.373	9.245	0.003	5.421	9.290	0.003	5.323	9.199	0.004
10	5.970	4.383	0.001	6.029	4.401	0.001	5.908	4.362	0.001
11	6.884	8.450	0.000	6.948	4.146	0.000	6.816	4.152	0.000
12	6.898	3.648	0.001	6.951	4.429	0.001	6.842	4.383	0.001

*ΔE is in kcal/mol, dipole moment is in **D**, estimate of the statistical weight is in %.

3.3 UV-vis spectra of the preferred conformers in DMSO

A large amount of experimental data can be used to train, develop, and validate *in silico* models. Many quinazoline based drugs, such as AG-1478 can be chromophores, which are promising tools for optical reporting, as they may change colour or fluorescence when interacting with solvents and cell proteins. (Khattab M. A., 2016; Khattab M. W., 2016) (Van Dongen, 2017) (Khattab M. W., 2018) (Kabir, 2020) The UV-Vis absorption spectrum of a compound is the result of transitions among the occupied molecular orbitals (ground state) and the virtual orbitals (excited states). In a benchmarking study regarding the accuracy of the DFT functionals in the TD-DFT calculations over fifty molecules containing the fused-ring with experimental maximum absorption (λ_{max}), Ali et al (Ali, 2020) found that the B3LYP/631G(d,p) model is among the most accurate. In their study, the theoretically predicted vertical absorption wavelength ($\lambda_{ver\text{-}abso}$) of these compounds is in the order of accuracy as LC-ωPBE < CAM-B3LYP < PBE0 < B3LYP < PBE when combined with the 6-31G(d,p) basis set. (Ali, 2020) This is something akin to our previous studies of UV-Vis spectral simulations for AG-1478 which used B3LYP/6-311G* basis set. (Khattab M. A., 2016) (Kabir, 2020) Our earlier studies on AG-1478 (Khattab M. W., 2016) (Khattab M. A., 2016) revealed that the measured absorption spectrum of AG1478 exhibited two overlapping peaks at ca. 332 nm and 346 nm, respectively, in DMSO solution.

Figure 5 displays the calculated UV-Vis spectra concentrating in the region 320-345 nm of the top six preferred AG-1478 conformer clusters in DMSO solution, using time-dependent (TD)-DFT method. In this figure, the calculated AG-1478 UV-vis spectra in region of 320-345 nm is from the most significant HOMO-LUMO transitions in a quality measurable region. (Khattab

M. A., 2016; Khattab M. W., 2016) The major spectroscopic properties including the positions of the major transitions (λ), conformer structures as well as the calculated HOMO-LUMO energy gaps are summarized in Table 3. Figure 5 also depicts a synthesised UV-vis spectrum (combined, coloured in red) which is based on the 22 superimposed individual spectra of all local minimum AG-1478 conformers (the value of each spectrum was scaled). The lower wavenumber band at ca 330 nm which is more intensive receives contributions from the first four preferred AG-1478 conformers (Clusters 1-4) whereas the less intensive band at ca 337 nm results from the local minimum conformer Clusters 5 and 6 with higher energy. The largest band position (342.57 nm) of conformer Cluster 18 is indicated in Figure 5 (without spectrum). The calculated UV-vis spectra for all 22 local minimum preferred conformer clusters of AG-1478 in DMSO solution are presented in Figure S1 of the Supplementary Materials.

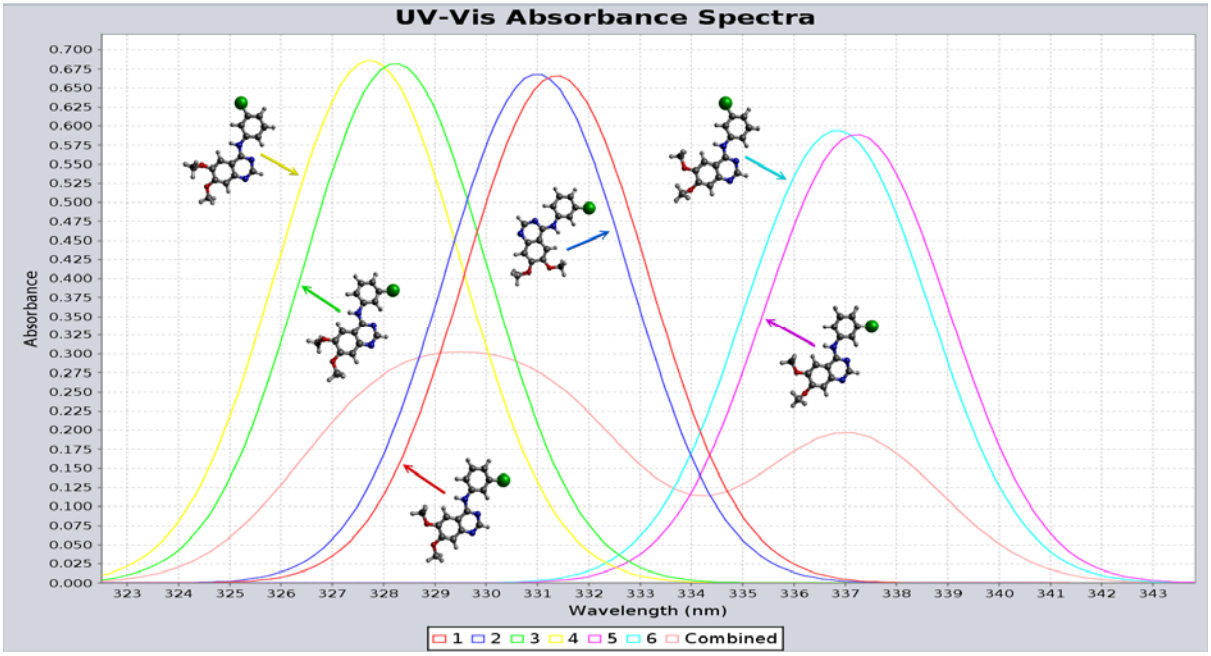


Figure 5. Calculated UV-Vis spectra of the first six preferred AG-1478 conformer clusters in DMSO solution. The combined spectrum (red) indeed shows the splitting bands. Cluster 4 and Cluster 18 AG-1478 conformers represent the smallest λ_{max} (328 nm) and largest λ_{max} (343 nm, position is pointed in the figure. The spectrum of Cluster 18 is not in the figure) of the spectrum.

The calculated UV-vis spectra of the preferred AG-1478 conformers in Figures 5 and S1 reveal that their UV-vis spectra are indeed sensitive to its conformation changes. The most intensive UV-vis band (λ_{max}) shift between the planar and non-planar conformers is as large as 14.80 nm from conformer (Cluster) 4 at 327.73 nm to conformer (Cluster) 18 at 342.53 nm, in excellent agreement with the UV-vis spectral measurement of 14 nm in the same solvent. (Khattab M. A., 2016; Khattab M. W., 2016)

Table 3: Comparison of the UV-Vis absorption spectral maxima (λ, nm) of AG-1478 conformers in DMSO using TD-DFT calculations.

#Cluster	Δε(HOMO-LUMO) eV	λ_1 (nm)	$\Delta\lambda_1$[a] (nm)	λ_2 (nm)	$\Delta\lambda_2$[a] (nm)
1	4.256	331.37	-0.63		
2	4.260	330.99	-1.01		
3	4.283	328.23	-3.77		
4	4.287	327.73	-4.27		
5	4.203			337.22	-8.78
6	4.206			336.84	-9.16
7	4.255			335.40	-10.6
8	4.241			336.96	-9.04
9	4.230	333.44	1.44		
10	4.234	332.87	0.87		
11	4.308	329.73	-2.27		
12	4.295	331.35	-0.65		
13	4.300	330.68	-1.32		
14	4.313	329.31	-2.69		
15	4.203			340.57	-5.43
16	4.189			342.06	-3.94
17	4.200			340.72	-5.28
18	4.185			342.53	-3.47
19	4.270	333.61	1.61		
20	4.256			335.23	-10.77
21	4.252	334.77	2.77		
22	4.269	333.33	1.33		
Combined		329		337	
Expt[a]		332		346	

[a]Experimental data in DMSO (Khattab M. A., 2016; Khattab M. W., 2016). Note that the experimental band positions were determined from simply the λ positions rather than being determined from fitting the spectrum.

Figure 5 and Table 3 indicate that conformation of AG-1478 contributes to the optical properties (spectra) through the changes in their electronic structures (excited states) of the conformers. It is further suggested that the non-planar conformers with higher strain energies are responsible to the optical spectral band shift to larger wavelength (nm) region. The largest UV-Vis spectral band shift (λ_{max}) at 343 nm results from a non-planar preferred conformer of AG-1478 (Cluster 18). It suggests that binding to the receptor such as EGFR (protein) pocket, a ligand more likely takes a non-planar conformation in the complex, as binding sites always locate on the concave surface. (Wang, 2013) It is quite different from when the ligand is in "isolation". The intelligent computing platform developed in this study is also able to batch calculate other measurable properties of the preferred conformers such as IR and NMR spectroscopy quantum mechanically.

4. Conclusion

The 4-anilinoquinazoline based chemical scaffold has led to development of the effective epidermal growth factor (EGFR) inhibitors, exemplified by the clinically approved drugs. As a result, one of the important drug development direction towards the novel small TKIs is to develop its derivatives from the 4-anilinoquinazoline template. (Aparna, 2005) (Bridges A. J., 1996) (Cappel, 2015) A significant challenge in anti-cancer drug development based on 4-anilinoquinazoline derivatives is the large number of conformers due to the ligand flexibility, (Mackey, 2015) in which full comprehensibility of deep learning models may be hard to achieve, (Schneider, 2016) (Jiménez-Luna, 2020). The present study developed an intelligent computing method in the direction of ligand conformation sampling. We demonstrated the method efficiency on the AG-1478 model ligand, a 4-anilinoquinazoline derivative to comprehensively search, batch calculate, sort out, rank and finally to simulate the measurable properties such as UV-vis spectra of the preferred conformers. The results which agree with available experimental measurements reveal that the AG-1478 conformers indeed shift the UV-Vis spectral band with approximately 14 nm, and non-planar conformers with higher energy exhibit different electronic behaviour with respect to their lower energy planar counterparts, in agreement with previous studies. (Khattab M. A., 2016; Khattab M. W., 2016) The same method can also be applied to study other 4-anilinoquinazoline derivatives with measurable properties, and to design new drugs with specific tyrosine kinase (EGFR) inhibitory activity. Using the intelligent computing method developed, we significantly reduce the time required in conformational search and sampling of 4-anilinoquinazoline derivatives and other prospective small molecule drugs from months to weeks. Finally, we noted that it is important to recognize that drug function is related with dynamism as well as with structures (Palopoli, 2016) and we are proceeding to combine this method of conformation search with molecular dynamics studies in near future.

Acknowledgment

FW acknowledges part of the funding from Excellerate Australia for “Spectroscopic and theoretical study of a potent anticancer drug.” The authors also acknowledge supercomputer support from National Computational Infrastructure (NCI) at Australian National University (ANU) and from Swinburne University of Technology Supercomputing Facilities (OzSTAR).

References

Ali, A. R. (2020). D-DFT benchmark for UV-visible spectra of fused-ring electron acceptors using global and range-separated hybrids. *Physical Chemistry and Chemical Physics, 22*, 7864.

Aparna, V. R. (2005). Virtual Screening of 4-Anilinoquinazoline Analogues as EGFR Kinase Inhibitors: Importance of Hydrogen Bonds in the Evaluation of Poses and Scoring Functions. *Journal of Chemical Information and Modeling, 45*, 725.

Bridges, A. J. (1996). Tyrosine Kinase Inhibitors. 8. An Unusually Steep Structure−Activity Relationship for Analogues of 4-(3-Bromoanilino)-6,7dimethoxyquinazoline (PD 153035), a Potent Inhibitor of the Epidermal Growth Factor Receptor. *Journal of Medicinal Chemistry, 39*, 267.

Bridges, A. J. (1999). The rationale and strategy used to develop a series of highly potent, irreversible, inhibitors of the epidermal growth factor receptor family of tyrosine kinases. *Current Medicinal Chemistry, 6*, 825.

Butler KT, L. F. (2009). Toward accurate relative energy predictiond of the bioactive conformation of drugs. *Journal of Computational Chemsitry, 30*, 601.

Cappel, D. D. (2015). Exploring conformational search protocols for ligand-based virtual screening and 3-D QSAR modeling. *Journal of Computer-Aided Molecular Design, 29*, 165.

Chang, G. G. (1989). An internal coordinate Monte Carlo method for searching conformational space. *Journal of the American Chemical Society, 111*, 4379.

Chen, H. E. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today, 23*, 1241.

Clayton, A. P. (2005). Fluorescence and analytical ultracentrifugation analyses of the interaction of the tyrosine kinase inhibitor, tyrphostin AG1478-mesylate, with albumin. *Analytical Biochemistry, 342*, 292.

Crippen, G. (1978). Note rapid calculation of coordinates from distance matrices. *Journal of Computational Chemistry, 26*, 449.

Ferguson, D. R. (1989). A new approach to probing conformational space with molecular mechanics: random incremental pulse search. *Journal of the American Chemical Society, 111*, 4371.

Fisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., . . . Gomper. (2016). *Gaussian 16, Revision C.01, .* Wallingford CT.: Gaussian, Inc., .

Foloppe N, C. I.-J. (2019). Energy windows for computed compound conformers: covering artefacts or truly large reorganization energies? *Future medicinal Chemistry, 11*, 97.

Foloppe, N. C. (2009). Conformational sampling and energetics of drug-like molecules. *Current Medicinal Chemistry, 16*, 3381.

Gasteiger, J. R. (1990). Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology, 3*, 537.

Habgood, M. J. (2020). Conformational Searching with Quantum Mechanics. In A. Heifetz, *Quantum Mechanics in Drug Discovery* (p. 207). New York: Springer Science+Business Media, LLC, part of Springer Nature.

Hao M-H, H. O. (2007). Torsion angle preference and energetics of small molecule ligands bound to proteins. *Journal of Chemical Information and Modelling, 47*, 2242.

Hawkins, P. S. (2020). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling, 50*, 572.

Horn, D. A. (2003). Novel clustering algorithm for microarray expression data in a truncated SVD space. *Bioinformatics, 19*, 1110.

Horn, D. G. (2001). Algorithm for data clustering in pattern recognition problems based on quantum mechanics. *Physical Review Letters, 88*, 018702.
Ismail, R. I. (2016). Recent advances in 4-aminoquinazoline based scaffold derivatives targeting EGFR kinases as anticancer agents. *Future Journal of Pharmaceutical Sciences, 2*, 9.
Izrailev, S. Z. (2006). A distance geometry. *Journal of Computational Chemistry, 27*, 1962.
Jiménez-Luna, J. G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence, 2*, 573.
Kabir, M. L. (2020). Deducing the Conformational Properties of a Tyrosine Kinase Inhibitor in Solution by Optical Spectroscopy and Computational Chemistry. *Frontiers in Chemistry, 8*, 596.
Kanal IY, K. J. (2018). A sobering assessment of small-molecule force field methods for low energy conformer predictions. *International Journal of Quantum Chemistry, 118*, e25512.
Khattab, M. A. (2016). Two conformers of a tyrosine kinase inhibitor (AG-1478) disclosed using simulated UV-Vis absorption spectroscopy. *New Journal of Chemistry, 40*, 8296.
Khattab, M. W. (2016). UV-Vis Spectroscopy and Solvatochromism of the Tyrosine Kinase Inhibitor AG-1478,. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 164*, 128.
Khattab, M. W. (2018). Conformational Plasticity in TKI-Kinase Interactions Revealed with Fluorescence Spectroscopy and Theoretical Calculations. *Journal of Physical Chemistry B, 122*, 4667.
Kolossva´ry, I. G. (1996). Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *Journal of the American Chemical Society, 118*, 5011.
Labute, P. (2010). LowModeMD-implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *Journal of Chemical Information and Modeling, 50*, 792.
Landau, L. L. (1976). *Statistical Physics. Course of Theoretical Physics. 5 (3 ed.).* Oxford: Pergamon Press.
Li, C. X. (2020). Platinum(II) Terpyridine Anticancer Complexes Possessing Multiple Mode of DNA Interaction and EGFR Inhibiting Activity. *Frontiers in Chemistry, 8*, 210.
Liu, D. J. (2016). Analyzing documents with Quantum Clustering: A novel pattern recognition algorithm based on quantum mechanics. *Pattern Recognition Letters, 77*, 8.
Mackey, M. (2015, December 15). *Virtual screening – how many conformations is enough?* Retrieved from Cresset: https://www.cresset-group.com/about/news/virtual-screening-how-many-conformations-is-enough/
Mennucci, B. T. (2002). Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotations of Chiral Molecules. *The Journal of Physical Chemistry A, 106*, 6102.
Mohamadi, F. R. (1990). MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry, 11*, 440.
Noolvi, M. P. (2013). A comparative QSAR analysis and molecular docking studies of quinazoline derivatives as tyrosine kinase (EGFR) inhibitors: A rational approach to anticancer drug design. *Journal of Saudi Chemical Society, 17*, 361.
Palopoli, N. M. (2016). Addressing the Role of Conformational Diversity in Protein Structure Prediction. *PLoS ONE, 11*, e0154923.

Rusinko, A. I. (1989). (1989) Using CONCORD to construct a large database of three-dimensional coordinates from connection tables. *Journal of Chemical Information and Modeling, 29*, 251.

Schneider, P. S. (2016). De novo design at the edge of chaos: iniperspective. *Journal of Medicinal Chemistry, 59*, 4077.

Van Dongen, M. K. (2017). Exploring the optical reporting characteristics of drugs: UV-Vis spectra and conformations of the tyrosine kinase inhibitor SKF86002. *New Journal of Chemistry, 41*, 14567.

Wang, K. G. (2013). An Accurate Method for Prediction of Protein-Ligand Binding Site on Protein Surface Using SVM and Statistical Depth Function. *BioMed Research International, 2013*, 409658.

Wilson, J. N. (2015). Binding-induced, turn-on fluorescence of the EGFR/ERBB kinase inhibitor, lapatinib. *Organic & Biomolecular Chemistry, 13*, 5006.

Xu, D. T. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science, 2*, 165.

Yang, H. S. (2018). ADMETopt: A Web Server for ADMET Optimization in Drug Design via Scaffold Hopping. *Journal of Chemical Information and Modeling, 58*, 2051.

Accelerating optical reporting for conformation of tyrosine kinase inhibitors in solutions

Feng Wang[1]* and Vladislav Vasilyev[2]

[1]*Department of Chemistry and Biotechnology and Centre for Translational Atomaterials, School of Science, Swinburne University of Technology, Melbourne, Victoria 3122, Australia*

[2]*National Computational Infrastructure, Australian National University, Canberra, ACT 0200, Australia*

**E-mail addresses: fwang@swin.edu.au (F. Wang)*

Supplementary Materials

Table S1 Information of 184 conformers collected from AG-1478 rotational potential energy landscape in DMSO solvent.

#	Relative Energy, kcal/mol	Dipole Moment	Cluster Size	Weight, %
1	0.000	8.450	10	67.646
2	0.594	3.648	11	26.819
3	1.974	8.178	17	3.871
4	2.570	5.305	11	0.900
5	2.933	11.100	9	0.395
6	3.516	5.812	19	0.306
7	4.567	3.310	12	0.032
8	4.630	4.673	10	0.024
9	5.373	9.245	5	0.003
10	5.970	4.383	5	0.001
11	6.883	4.149	6	0.0003
12	6.898	4.407	14	0.0007
13	6.975	2.214	7	0.0003
14	7.017	5.329	6	0.0002
15	7.580	6.298	12	0.0002
16	7.627	7.631	9	0.0001
17	7.638	5.298	9	0.0001
18	7.710	8.475	6	0.00007
19	10.334	4.232	1	0.000
20	10.336	5.484	2	0.000
21	10.494	5.476	2	0.000
22	10.524	4.243	1	0.000

Table S2 Information of 182 conformers collected from AG-1478 rotational potential energy landscape in water.

#	Relative Energy, Kcal/mol	Dipole Moment	Cluster Size	Weight, %
1	0.000	8.4873	10	68.302
2	0.607	3.6568	11	26.481
3	2.014	8.2046	17	3.649
4	2.622	5.3071	11	0.832
5	2.940	11.1519	9	0.394
6	3.538	5.8420	18	0.282
7	4.612	3.3203	12	0.030
8	4.672	4.7026	10	0.022
9	5.421	9.2901	5	0.003
10	6.029	4.4014	5	0.001
11	6.948	4.1458	6	0.0003
12	6.951	4.4294	14	0.0006
13	7.039	2.2212	6	0.0002
14	7.074	5.3444	6	0.0002
15	7.632	6.3314	12	0.00017
16	7.684	7.6692	9	0.00011
17	7.698	5.3235	9	0.00011
18	7.762	8.5223	6	0.00007
19	10.405	5.5106	2	0.000
20	10.409	4.2455	1	0.000
21	10.571	5.5039	2	0.000
22	10.595	4.2633	1	0.000

Table S3 Information of 183 conformers collected from AG-1478 rotational potential energy landscape in Methanol.

#	Relative Energy, Kcal/mol	Dipole Moment	Cluster Size	%
1	0.000	8.4115	10	66.975
2	0.580	3.6392	11	27.191
3	1.932	8.1500	17	4.121
4	2.516	5.3015	11	0.978
5	2.925	11.0456	8	0.352
6	3.493	5.7813	19	0.316
7	4.520	3.2999	12	0.0341
8	4.586	4.6426	10	0.0254
9	5.323	9.1986	5	0.0036
10	5.908	4.3618	5	0.0013
11	6.816	4.1520	6	0.0003
12	6.842	4.3832	14	0.0007
13	6.908	2.2066	7	0.0003
14	6.958	5.3140	6	0.0003
15	7.524	6.2664	12	0.0002
16	7.568	7.5876	9	0.0001
17	7.576	5.2722	9	0.0001
18	7.654	8.4249	6	0.00007
19	10.256	4.2175	1	0.0000
20	10.262	5.4541	2	0.0000
21	10.415	5.4465	2	0.0000
22	10.449	4.2218	1	0.0000

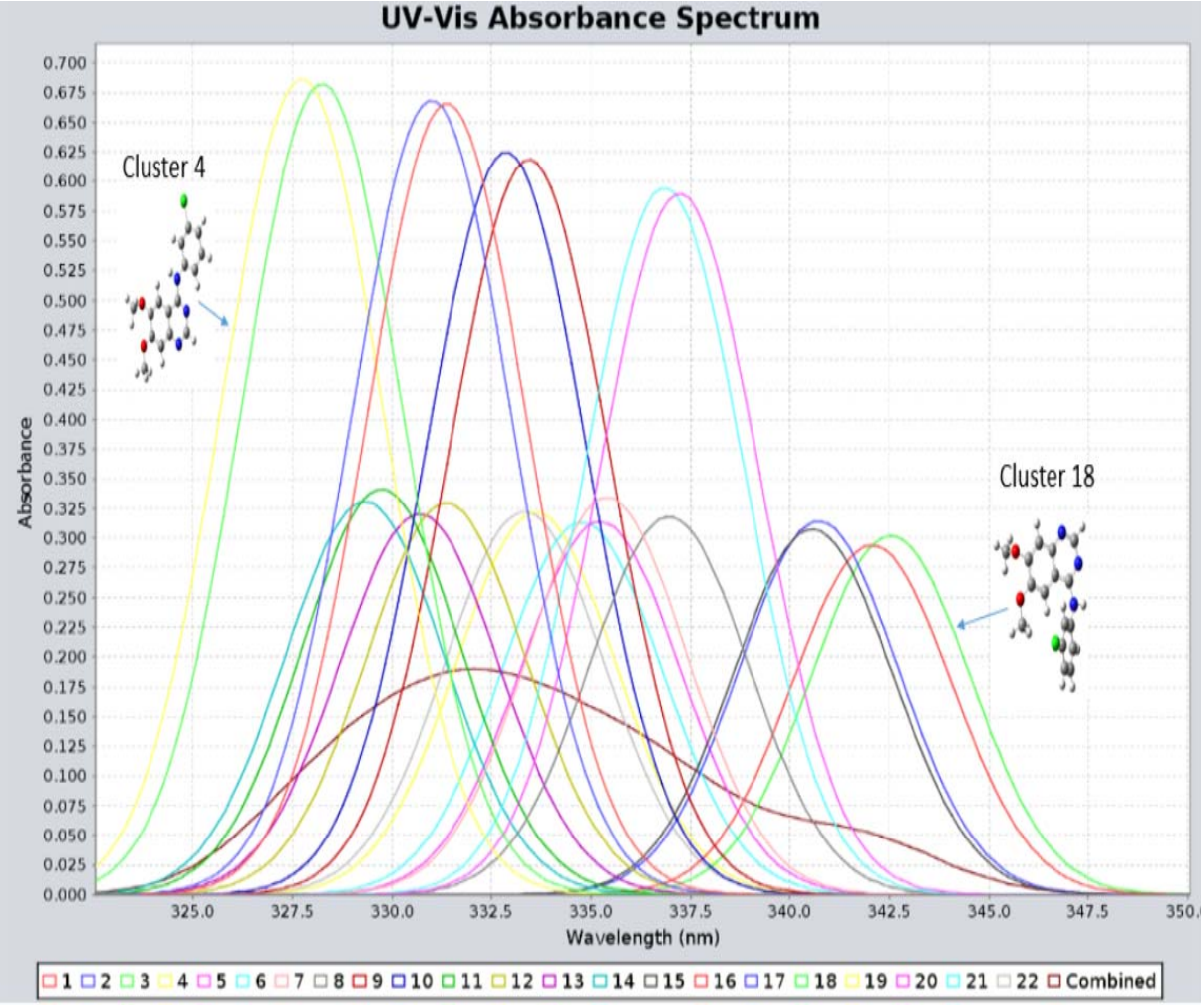


Figure S1 Conformation caused UV-Vis spectral band shift of all lowest-lying clustered AG-1478 conformers in DMSO solution up to 22 Clusters.