A Quantum-Inspired Approach to De-Novo Drug Design

David Snelling, Ganesh Shahane, William Shipman, Alexander Balaeff, Mark Pearce, Shahar Keinan

Design and optimization of targeted drug-like compounds is an important part of the early stage drug discovery process. In this paper, we describe the use of a novel technique for rapid design of lead-like compounds for the Dengue viral RNA-dependent-RNA polymerase (RdRp). Initially, a large (>billions) fragment-based chemical library is designed by mapping relevant pharmacophores to the target binding pocket. The de-novo synthesis of molecules from fragments is formulated as a quadratic unconstrained binary optimization problem that can be solved using the quantum-inspired Digital Annealer (DA), providing an opportunity to take advantage of this fledgling, groundbreaking technology. The DA constrains the search space of molecules with drug-like properties that match the binding pocket and then optimizes for synthetic feasibility and novelty, thus offering significant commercial advantages over existing techniques.

File list (1)

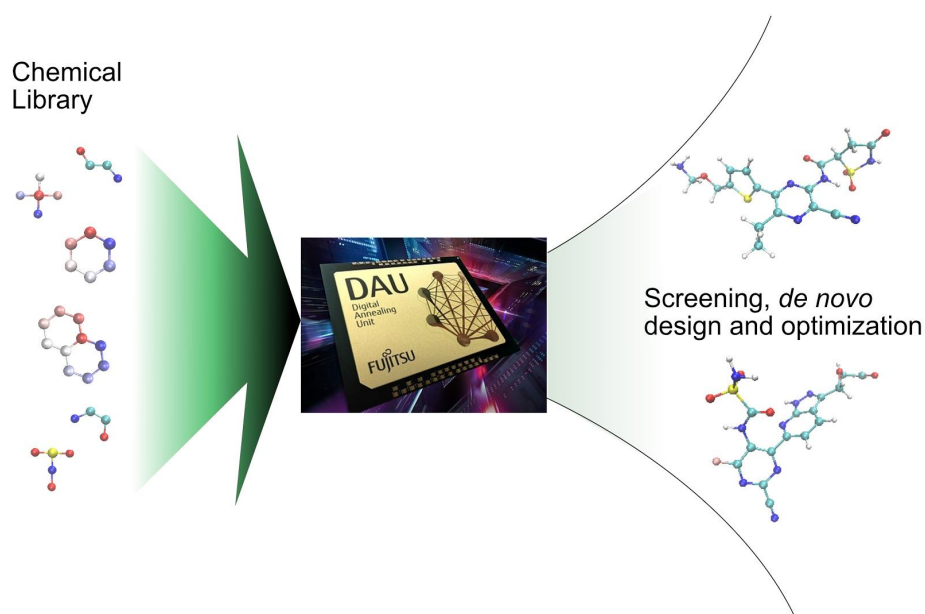|	
White Paper - Final Version.pdf (1.20 MiB)	view on ChemRxiv • download file

A Quantum-Inspired Approach to De-Novo Drug Design

David Snelling*[A], Ganesh Shahane[A], William J. Shipman[B], Alexander Balaeff[B], Mark Pearce[A] and Shahar Keinan*[B]

[A]Fujitsu UK, 22 Baker Street, Marylebone, London W1U 3BW
[B]Polaris[qb], 201 W Main St., Durham, NC USA 27701

Chemical Library

DAU
Digital Annealing Unit
FUJITSU

Screening, *de novo* design and optimization

Abstract: Design and optimization of targeted drug-like compounds is an important part of the early stage drug discovery process. In this paper, we describe the use of a novel technique for rapid design of lead-like compounds for the Dengue viral RNA-dependent-RNA polymerase (RdRp). Initially, a large (>billions) fragment-based chemical library is designed by mapping relevant pharmacophores to the target binding pocket. The de-novo synthesis of molecules from fragments is formulated as a quadratic unconstrained binary optimization problem that can be solved using the quantum-inspired Digital Annealer (DA), providing an opportunity to take advantage of this fledgling, groundbreaking technology. The DA constrains the search space of molecules with drug-like properties that match the binding pocket and then optimizes for synthetic feasibility and novelty, thus offering significant commercial advantages over existing techniques.

1. Introduction

1.1. Need for faster drug design:

The purpose of drug design is to identify novel molecules that bind to a specific protein (relevant to a specific disease) and block (or enhance) the protein activity, thus changing the course of the disease. Drug molecules also have other certain necessary properties such as selectivity and safety. Due to increasing costs of drug development and a high failure rate of potential drug candidates, there is a continuous need for development of new innovative medicines. For example, the FDA-approval rate for new drugs that enter clinical trials is only 19% [1]. In the last couple of years, there has been an interest in finding drug candidates from novel chemistry, leading designers to explore larger and larger chemical spaces [2]. However, for most computer systems it takes too long to explore a large chemical space, especially those including over 1 billion molecules. We present here the use of a quantum-inspired technology for searching large chemical spaces as the means to significantly accelerate the first step of any computational drug design campaign.

1.2. The Digital Annealer:

Computational drug design can be viewed as an optimization problem in which one is searching for molecules within a predefined chemical space that maximizes deriderable properties. The enumeration of billions of known molecules and then testing their properties using structure-based computational modeling methods is a cost-prohibitive task. Thus, to overcome this barrier we use the quantum-inspired Digital Annealer (DA), instead of brute-force enumeration. The DA uses a heuristic technique for solving hard optimization problems by mimicking quantum mechanical effects such as quantum tunneling, a promising approach with a potential to harness quantum mechanical concepts to solve hard optimization problems. The DA can thus be used for single or multi-objective optimization problems, allowing optimization over multiple dimensions and thus an efficient global search for promising solutions.

Fujitsu's Digital Annealer architecture uses a digital circuit design inspired by quantum phenomena with logical connections across all bits. It can solve large-scale combinatorial optimization problems very quickly and more accurately than possible before.

A convenient way to conceptualize the Digital Annealer is as a special accelerator to speed up combinatorial optimizations and most likely to be used with conventional hardware in a hybrid environment. It is important to emphasize that the DA is not based on an actual quantum computer and therefore does not suffer from that technology's engineering and practical constraints:

- Unlike a true quantum computer, it is commercially viable today (not at the prototype stage), and experimental access is possible.

- The Digital Annealer operates at room temperatures and does not need special cooling beyond what is usually used for computer systems.

- And yet, for combinatorial optimization calculations – one of the core advances promised by quantum computing – the Digital Annealer delivers results several orders of magnitude superior to those currently available using true quantum devices.

1.3. Dengue Fever:

As a case study for our approach, we chose the RNA-dependent RNA polymerase (RdRp) of the Dengue virus. Dengue fever is a mosquito-borne viral disease causing initial flu-like symptoms and, in a small number of cases, developing into life-threatening complications such as the Dengue Shock Syndrome (DSS) or the Dengue Hemorrhagic Fever (DHF). It is now widespread in over a 100 countries and 4 continents worldwide, thereby threatening up to 40% of the world's population [3]. According to the World Health Organization (WHO), there are as many as 50-100 million infections per year with 500,000 cases of severe dengue and 22,000 deaths. Though a number of attempts have been made to design antiviral drugs, there is still no cure available on the market. Sanofi's FDA-approved vaccine Dengvaxia can be

administered only to a limited number of people aged 9-45 with a history of sickness and recovery from the disease. For those not previously infected, Dengvaxia might in fact significantly increase the risk of a severe infection [4]. Hence, there is an urgent need to develop safe and effective drugs for the treatment of dengue infection.

The Dengue viruses (DENV) are divided into four different and closely related serotypes: DENV1, DENV2, DENV3 and DENV4. An ideal Dengue inhibitor should therefore exert a pan-serotype activity by targeting a conserved protein essential for viral replication. One such protein is the RNA-dependent RNA polymerase (RdRp) that performs RNA synthesis during DENV3 replication. The protein sequence for RdRp is conserved across all four DENV serotypes with more than 65% homology and hence serves an attractive target for drug design. We are targeting here an allosteric binding pocket (not the RNA binding active site) that was discovered in RdRp using fragments, and is located between the Fingers domain and the Priming loop. Several x-ray structures of the RdRp Dengue polymerase with ligands in this pocket are available, and these form the basis for our design work.

2. Workflow and Results

2.1. An Overview:

We present here the workflow for the De-Novo Digital Drug Design platform, and how we implement it for the RdRp domain of the DENV3 serotype NS5. For a specific disease and protein target, we start by building a chemical library: for the selected binding pocket we extract molecular characteristics, combine those with other drug-like properties to build the virtual chemical space [5]. We then assess the library and design molecular leads using the quantum-inspired Digital Annealer (DA). The leads are then evaluated using machine learning algorithms [6] for the prediction of ADMET properties such as solubility, toxicity and blood-brain barrier (BBB) penetration and further ranked according to the binding affinities calculated by QM/MM simulations [7]. In this white paper, we will use the DENV3 target to illustrate the process of building specific, very large virtual chemical libraries and searching those for novel drugs with the Digital Annealer.

2.2. Designing a Targeted Virtual Chemical Library:

A challenge in drug design is the need for novel (and patentable) chemistry that still obeys the rules for being "drug-like". We address this challenge here by designing a huge (over 1 billion molecules) virtual chemical library that is targeted to the Dengue Fever RdRp allosteric pocket and utilizing the DA optimization protocols to find the best "drug-like" molecules.

The library design starts from the targeted protein binding pocket. In this case, the 3-dimensional x-ray structure of the Dengue Fever RdRp allosteric pocket is available in the Protein Data Bank (Yokokawa et al. 2016 [8]; PDB ID 5HMZ). This pocket was discovered using Fragments campaign, with multiple structures that have bound ligands available, including the one shown here (figure 1). To build the library, we examined the binding pocket and identified the amino acids residues that are needed for ligand binding.
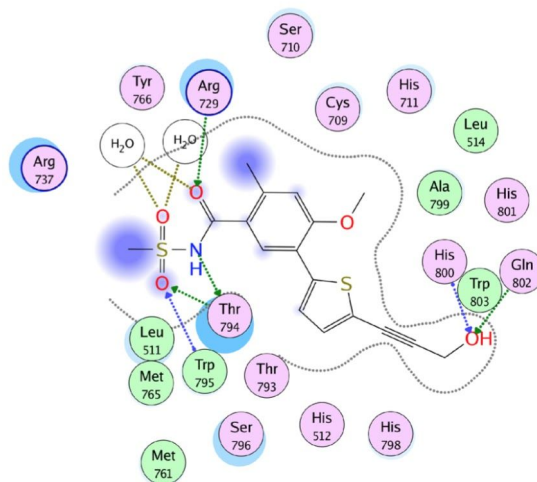

Figure 1: A two-dimensional ligand interaction map of compound-23 interacting with amino acid residues of the DENV3 RdRp (PDB: 5HMZ.pdb). The image is adapted from (Yokokawa et al. 2016) [8].

Several molecular scaffolds that will bind to these residues were identified, then the library was expanded by adding functional groups to the scaffolds. The functional groups are selected based on the characteristics of the 3-dimensional binding pocket i.e. the locations of polar/non-polar and aromatic amino acid side chains, the protein flexibility, excluded volume, etc. This process is similar in theory to scaffold hopping, but is done on a much larger scale,

with the resulting library including anywhere from hundreds of millions to over a billion molecules.

An example virtual chemical library is shown in figure 2. The library is based on a 6-atom ring scaffold (top) to which 6 functional groups are covalently attached at different positions. Each molecule in the library is thus described by a size-6 vector, whereby the i-th component of the vector represents the ID of the functional group chosen for the i-th position. The Maximal values of the vector component are {2,15,15,16,11,17} according to the number of possible functional groups at each site. All possible combinations of the R-groups yield the library of ~1.3 million molecules. An example molecule from the library and its corresponding molecular vector are shown in figure 2.
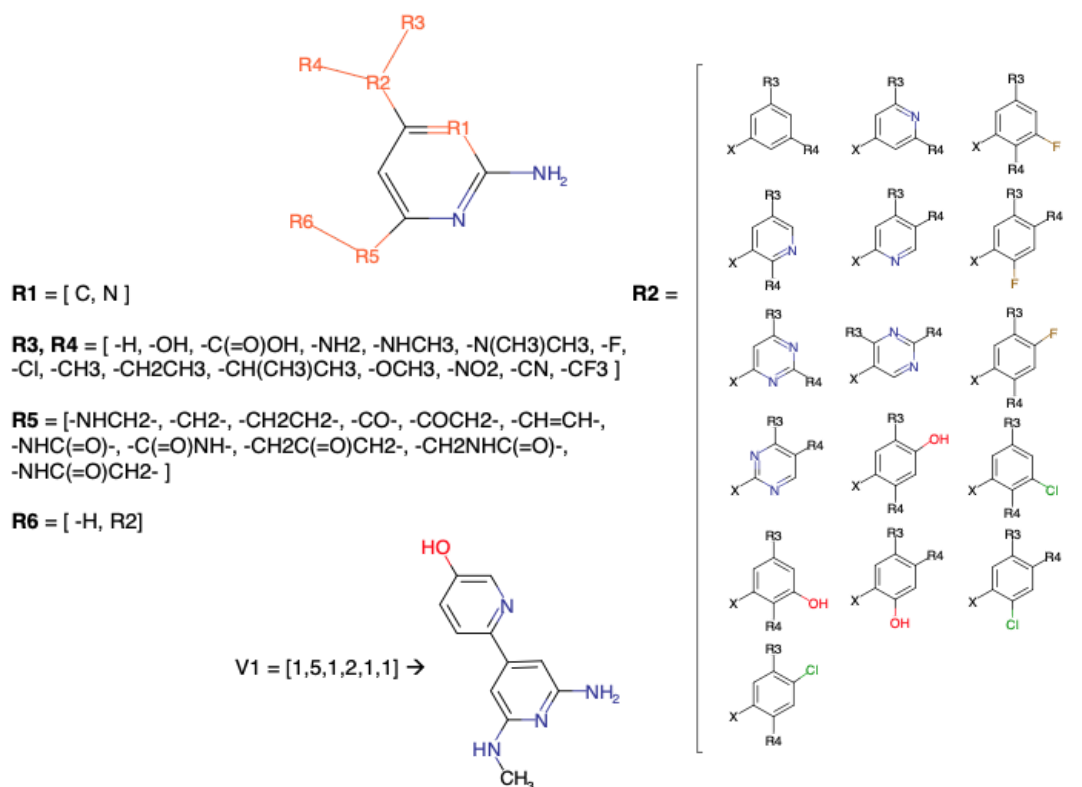

R1 = [C, N]

R3, R4 = [-H, -OH, -C(=O)OH, -NH2, -NHCH3, -N(CH3)CH3, -F, -Cl, -CH3, -CH2CH3, -CH(CH3)CH3, -OCH3, -NO2, -CN, -CF3]

R5 = [-NHCH2-, -CH2-, -CH2CH2-, -CO-, -COCH2-, -CH=CH-, -NHC(=O)-, -C(=O)NH-, -CH2C(=O)CH2-, -CH2NHC(=O)-, -NHC(=O)CH2-]

R6 = [-H, R2]

V1 = [1,5,1,2,1,1] →

Figure 2: A visual representation of an example of a targeted virtual chemical library. The molecular scaffold (top) has six attachment points for six different R-groups. Each of the R-groups from R1 to R6 can be represented by the atoms/fragments in the corresponding square brackets. For example, R1 can either be C or N. Every molecule designed is represented by a size-6 vector, where the numbers correspond to the atom/fragment that represents a particular R-group. The image is copied from (Keinan et al. 2018) [9].

The actual chemical library used in this study is similar to the one shown in figure 2, but had 8 different R groups and a larger choice of fragments for each than in figure 2. The resulting size of the library was approximately 1.3 billion molecules. The exact formulation of the library is omitted from this paper for proprietary reasons.

2.3. Defining the drug design criteria:

Next, we used the DA to search the library of 1.3 billion molecules and find the optimal ~1000 molecules for more accurate calculations. The following optimization constraints defined how we find the optimal molecules, and were divided into two categories:

1. **Drug-like properties** which are relevant to all drug design projects, and include limits on molecular weight (MW, 450-500), number of Hydrogen bond donors (HBD, 2-5), number of Hydrogen bond acceptors (HBA, 2-10), topological polar surface area (TPSA, 90-140), number of rotatable bonds (RB, 4-8) and the octanol-water partition coefficient (LogP, 0-5). Furthermore, we implemented the functional group filters as described in the GDB-17 paper [2], to avoid formation of possible unrealistic and toxic functional groups and checked for synthetic accessibility of each molecule.

2. **Structural properties** that are specific to the Dengue Fever RdRp allosteric pocket, and define critical distances between pharmacophores. In this case, we define the distances between 3 groups (figure 3) relevant to binding in the pocket at a bound geometry, as can be seen in the 5HMZ.pdb compound-23: a) a hydrogen bond donor (an -OH group in 5HMZ) and b) two hydrogen bond acceptors (in 5HMZ, -C=O and -S=O, respectively).

The goal of our optimization was to select optimal molecules from the whole chemical library. An optimal molecule is defined as the one for which the properties described above fall within the specified bounds.
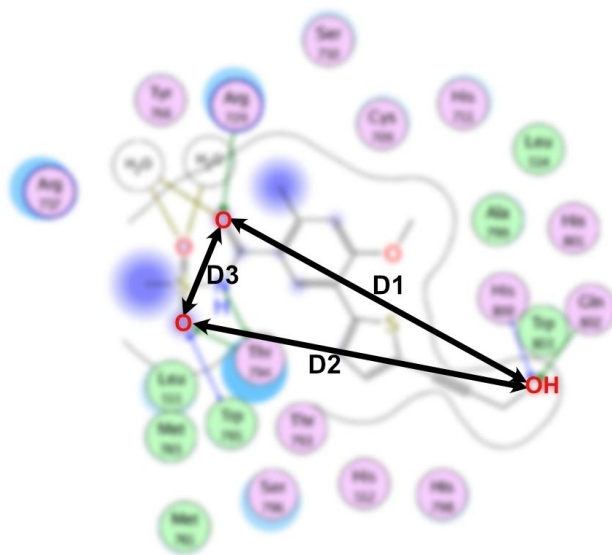
Figure 3: Defining three distances: D1, D2 and D3 between the key hydrogen bond donors and acceptors of compound-23, that make critical contacts with amino acids in the DENV3 RdRp.

2.4. Searching the library using the DA:

Once the library design is complete, and the optimization constraints are defined, the process of identifying the optimal molecules begins. In Methods section 3, we will discuss the Quadratic Unconstrained Binary Optimization (QUBO) algorithm developed to search the library. For this manuscript, the QUBO algorithm was used approximately 40 times to search the RdRp Dengue virus molecular library. On the DA, each search took 0.6 seconds to select 125 molecules and the top 25 were chosen from each set, resulting in 977 molecules.

2.5. Diversity of the Designed Molecules:

Figure 4 visualizes the chemical space of 977 optimal compounds detected by the DA. The chemical space is described as the principal components of the six drug-like physicochemical properties of pharmaceutical relevance: MW, HBA, HBD, RB, logP and TPSA. The first three

principal components captured ~68% of the variance and a lack of clustering indicates that the designed molecules cover the chemical space well.
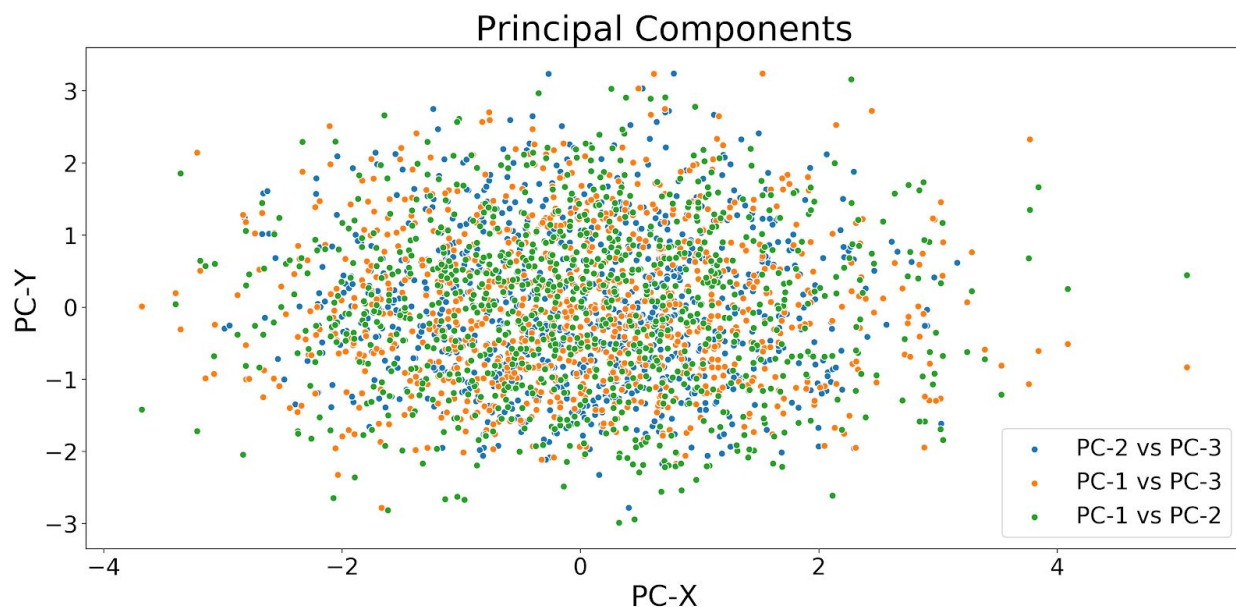

Figure 4: Visual representation of the chemical space of the 977 optimal molecules detected in the Dengue RdRp library by the DA. The visualization is based on the principal component analysis (PCA) of six drug-like properties as described in the text. The first three principal components (PC-1, PC-2 and PC-3) account for ~68% of the variance. Data points are colour-coded according to the principal components plotted against each other as described in the legend.

Another way to look at the diversity of these molecules, is to look at which R groups from the original library were selected by the DA. Figure 5 shows the frequency of each R group for each position on the scaffold in the constructed chemical library.
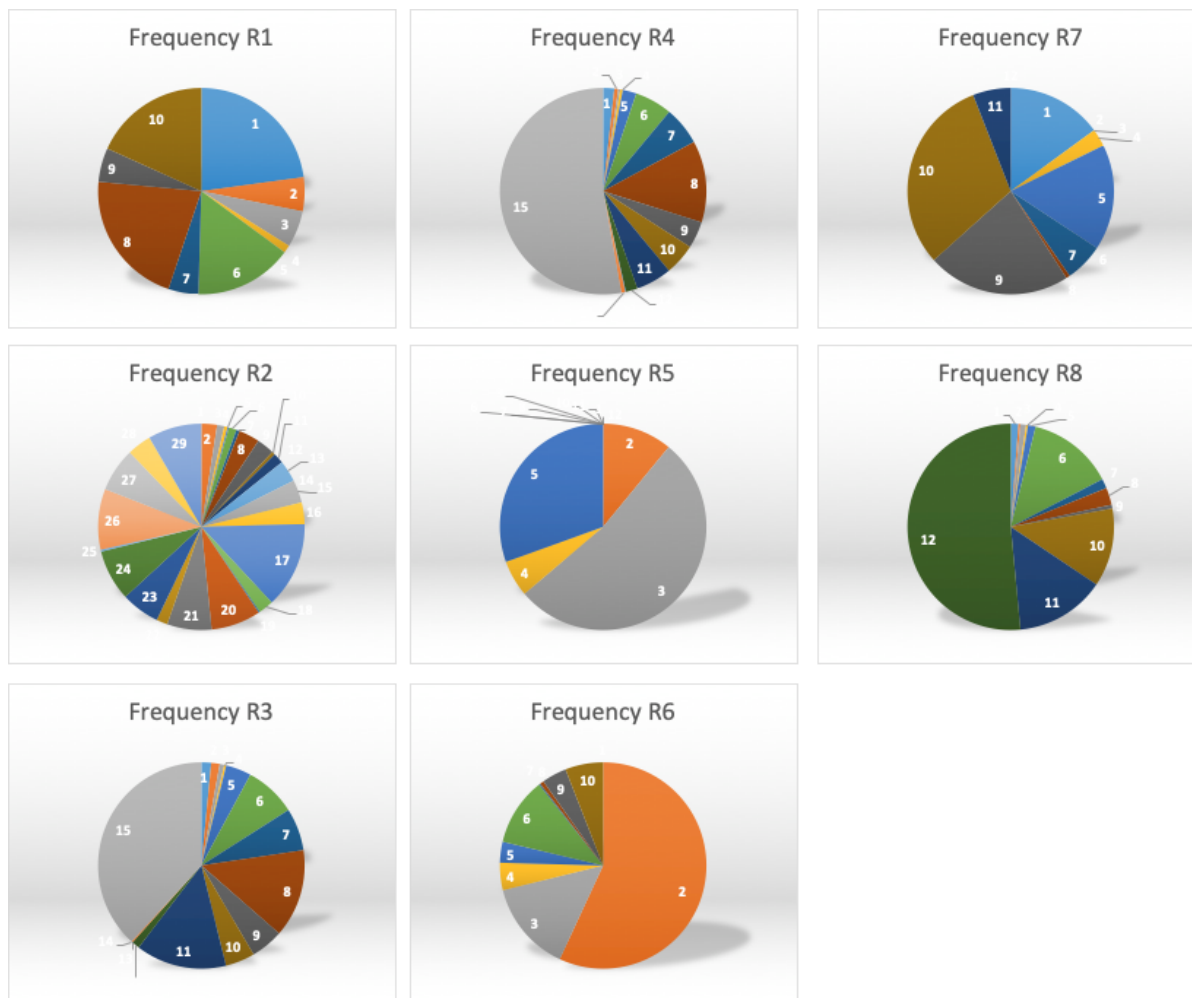
Figure 5: The frequencies of each of the fragments in the designed 977 molecules from all R-groups. R5, R6 and R7 groups show larger frequencies of fragments containing hydrogen bond donors and acceptors. R3 and R4 show a high frequency of -CF3 which predominantly contributes towards the molecular weight of the designed molecules.

Several interesting observations can be made from figure 5. One of the constraints we started from was the donor-acceptor distances. From the frequency of R5, R6 & R7 it can be seen that the DA identifies the right fragments from these R groups for donors and acceptors. Also, one of the target constraints used was MW. To reach MW of 450, the heavy group of CF3 is frequently selected by the DA in some R groups.

A question remains whether or not the molecules selected using rather simplistic geometric constraints will indeed fit well in the Dengue Fever RdRp allosteric pocket. We have answered this question by looking at two R groups. First, for the R1 group (defining the molecular scaffold) all fragments are present except for one (#5) where the angles between the donors and acceptors groups are significantly different than with other fragments in the R group. This indicates that the DA will identify the right geometry for the Dengue Fever RdRp allosteric pocket. Secondly, for the R group that binds Trp803 (cf. figure 3), one fragment (#12) occurs most frequently. This is the same group that was identified as the optimal group in the Novartis original paper (Yokokawa et al. 2016) [8], albeit after multiple experimental trials. Here, we identify it in 0.06 seconds, as it appears in many samples generated by the DA on each and every pass.

To better understand the nature of the chemical subspace we used here, we studied the distribution of all the six drug-like properties (figure 5). As can be seen in figure 5, none violate the given constraints, indicating that the DA algorithm was able to select an optimum in every run. The MW exhibits a positively skewed distribution (figure 5a) ranging from 449.99 to 496.13 g mol^{-1} and an average of 466.79 ± 10.67 g mol^{-1}. On the other hand, the TPSA exibits a negatively skewed distribution with a mean of 124.86 ± 9.31 Å^2. The RB and HBD (figure 5c and 5e) follow a slightly right-skewed distribution and range from 4 to 8 and 2 to 5 respectively. LogP and HBA on the other hand, follow a gaussian distribution, where a logP between 1 to 4 accounts for ~90% of the designed molecules.
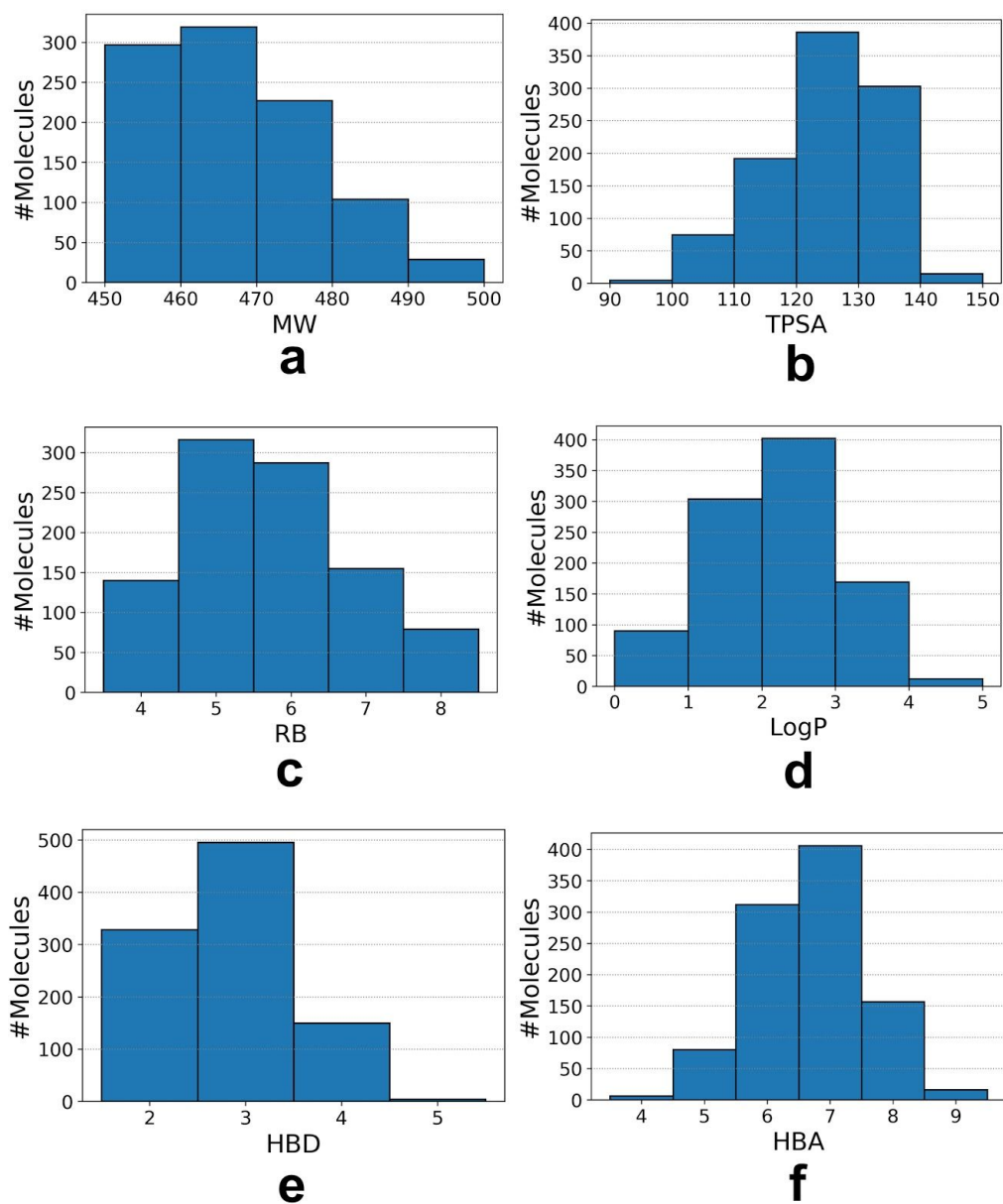
Figure 5: Distribution of molecular properties for 977 molecules designed by the DA **a:** MW- Molecular Weight, **b:** TPSA- Topological Polar Surface Area, **c:** RB-Number of Rotatable Bonds, **d:** LogP- octanol/water partition coefficient, **e:** HBD- Hydrogen Bond Donors and **f:** HBA- Hydrogen Bond Acceptors.

2.6. Benchmarking & Model Evaluation:

De novo drug design is a multiobjective optimisation problem [10]. Hence, we evaluate several properties of molecules produced by the DA; out of the 1.3 billion molecules in the library, 10,000 molecules constrained towards the Dengue RdRp pocket chemical space were generated and evaluated using the GuacaMol benchmark [11] for five properties:

1. **Validity:** If molecules with incorrect SMILES syntax are generated, then the model is penalized.
2. **Uniqueness:** If the model generates the same molecule more than once, it is penalized.
3. **Novelty:** This benchmark penalizes models when they generate molecules present in the training set.
4. **Frechet Chemnet Distance (FCD):** Introduced by [12], FCD is a measure of how close distribution of generated data is compared to the distribution of molecules (taken from ChemBL) in the training set.
5. **KL Divergence:** In this, the probability distributions of a variety of physicochemical descriptors for the training set and the DA-generated molecules are compared and corresponding KL divergences are calculated.

Table 1 lists the scores of all the above benchmarks for the DA and some other previously developed generative models.

Benchmark	DA	SMILES LSTM	ORGAN	VAE
Validity	1	0.959	0.379	0.870
Uniqueness	1	1	0.841	0.999
Novelty	1	0.912	0.687	0.974
KL divergence	0.043	0.991	0.267	0.982
FCD	0.0008 (35.51)	0.913	0	0.863

Table 1: Results of the DA and other generative models for GuacaMol benchmarks. Scores close to 0 indicate poor performance for the said metric.

It can be seen that the DA-generated molecules are characterized by a high degree of validity, uniqueness, and novelty. The FCD and KL divergence scores, however, can only be meaningfully used if models are pre-trained on the ChEMBL dataset and are used to generate ChEMBL-like molecules. Poor scores on these metrics indicate that the model either generates molecules highly specific for a target protein or the molecules belong to a non-drug-like chemical space. Thus, it is expected that the FCD and KL scores would be close to zero, as we are comparing the distribution of specialised molecules to a subset of ChEMBL data.

3. Methods

3.1. Quadratic Unconstrained Binary Optimization (QUBO):

In this section we will expand on how we programmed the Digital Annealer to perform the Monte Carlo optimization in the chemical space using the aforementioned optimization constraints (section 2.3). The Digital Annealer's power partially comes from its simplicity. The field of possible solutions is described by a bit vector x where each bit represents whether a single property is true or not. In the case of drug design, we are using each bit to represent whether a given molecular fragment is chosen for a particular R-group. For example, in our library design R group R5 has the possibility of being an -OH fragment. If this is the case, then the single bit corresponding to an -OH fragment in the R5 group position will be 1. Using this as a basis, we formulate the optimization problem as a Quadratic Unconstrained Binary Optimization (QUBO) polynomial [13]. The goal of the Digital Annealer is to perform a minimization of the QUBO, i.e. find values for the x such that the polynomial is minimized.
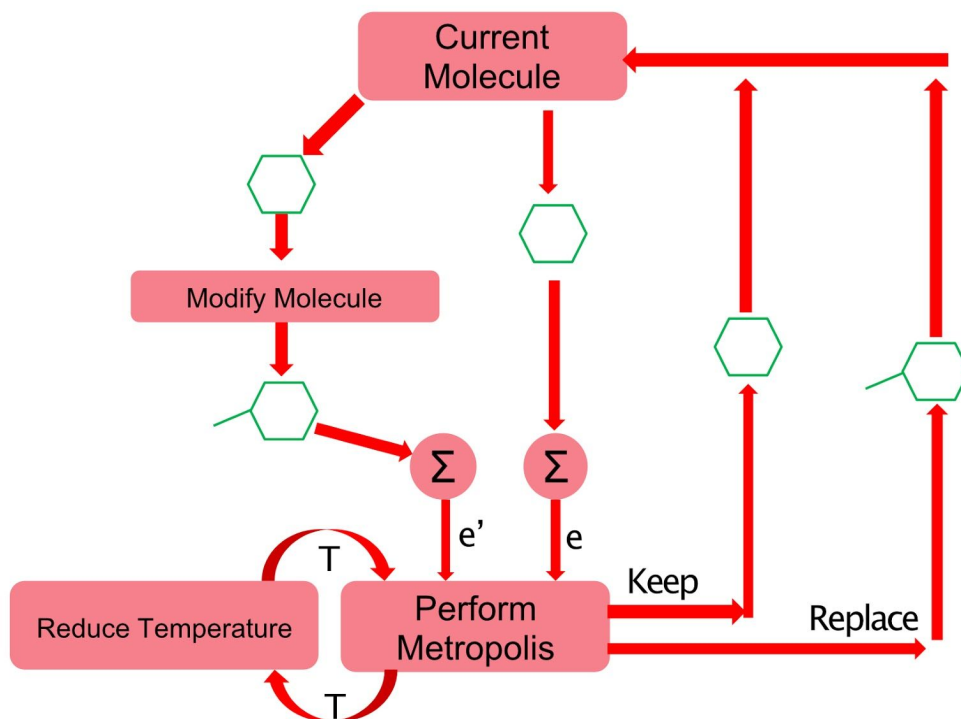
Figure 7: In this illustration of the DA algorithm workflow, random bit vectors x (each representing a possible molecule) are used to compute the value of the QUBO. Using a simulated annealing algorithm, executed millions of times in parallel, progressively better molecules are designed.

$$QUBO = \sum_{I,j} x_i x_j$$

The above polynomial is *binary,* because x can have either a 0 or 1 value (simple binary variables). The weights ($W_{i,j}$) are 64 bit coefficients. The polynomial is *quadratic*, since the x terms are multiplied by each other, but in no cases do we multiply three or more x's together. Notice that the value of a weight ($W_{i,j}$) is only added to the total when both x_i and x_j are one. The polynomial is *unconstrained*, because there are no other limits on what values the variables can take, i.e. there are no direct implementation of if-then-else-like conditions on the solution.

There are however, ways in which constraints can be encoded into a QUBO. Specifically for the work described in this manuscript, most of the DA search is done through constraint satisfaction rather than optimization. As we stated before, here we implement two types of constraints, structural and parametric. For structural constraints we require that only one fragment be assigned to an R group - we can't attach both an -OH fragment and a -C(=O)OH fragment to the same bonding point. The details on implementation are beyond the scope of this White Paper, but the reader is directed to (Glover et al. 2018) [13] for a very clear tutorial on QUBO design.

3.2. Searching the library:

Based on the constructed QUBO polynomial, we run the DA optimization process shown in figure 7. The process is repeated millions of times in parallel, producing hundreds of good quality molecules in seconds. The algorithm starts with a random molecule from the library and modifies it by changing a single fragment. Both the original and new molecules are evaluated using the QUBO formulation that models the desired constraints. The evaluation is done based on the Metropolis algorithm. It starts with a random configuration, and computes values for e and e' by evaluating the QUBO. It then uses the values e, e', and T to compute the probability that the new molecule will replace the current molecule. This probability reduces slowly as the temperature (T) decreases and the process is repeated until a minimum temperature is reached.

Figure 7 shows that annealing provides a much better sample than simple random selection. We show there that out of approximately 1000 annealed and random samples, both sorted by the QUBO energy (explained in the following section), the annealed molecules reach energy that is two orders of magnitude lower than random molecules. The horizontal axis is the sorted sample number and the vertical axis is the QUBO energy.
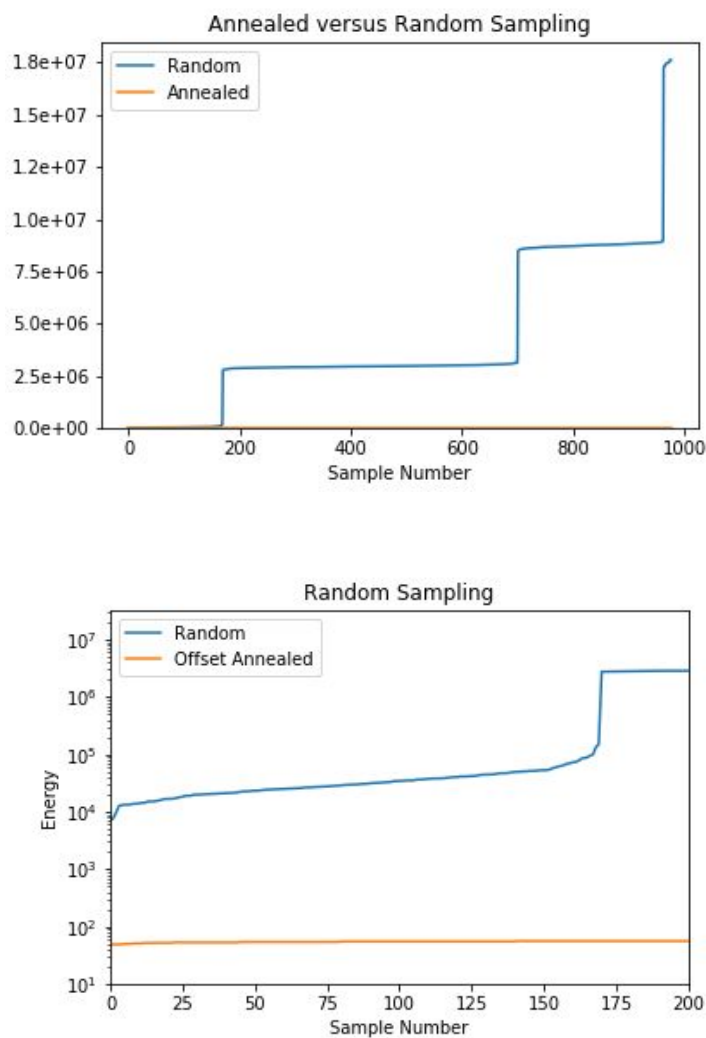
Figure 7: Graphs show the QUBO energies of 1000 random and annealed samples from the chemical space. The lower plot is a zoom in on the lower left corner, plotted using a log scale. [Note that the annealed energy is actually negative, so has been offset by 100 to facilitate display.] It is important to note that none of the random samples reach an energy value of zero, which indicates that the drug-like property constraints are satisfied.

3.3. Synthetic Accessibility (SA) score as the objective function:

Although there are many ways to construct a QUBO, in this case the QUBO has been designed such that any molecule that fails to meet all the constraints will have a positive energy and

those with a zero energy meet all the constraints. In addition, negative energy values can be incorporated into the model to reflect the expected simplicity of synthesizing the molecule. For example, using the Synthetic Accessibility (SA) score (Ertl and Schuffenhauer 2009) [14], the more negative the energy, the easier it is to synthesize the molecule.
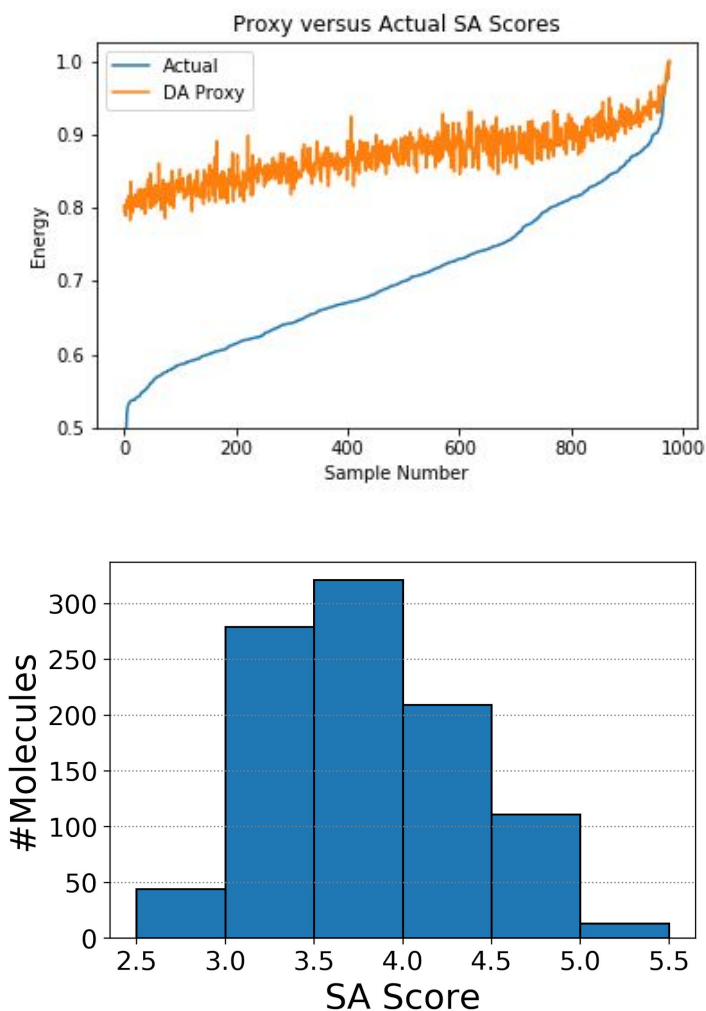

Figure 8: Correlation (top) between actual Synthetic Accessibility (SA) scores and the proxy scores used by the DA. The data are normalized in the range of 0-1, for clarity. Distribution (bottom) of actual SA scores of the designed 977 molecules as calculated from the scripts written by (Ertl and Schuffenhauer 2009) [14]. Scores below 6 indicate easy synthesizability.

Even though an exact implementation of the SA score is not possible in the DA, a proxy formulation was developed. The SA score of each molecule is approximated by the sum of all possible pairwise SA scores of the functional groups of which the molecule is composed (the SA scores for each pair of fragments is readily available). The sum is not equal to the accessibility score that would be obtained for the fully assembled molecule. However, the proxy score and the exact score correlate well. In this study, the correlation coefficient between the two was found to be 0.91 (figure 8, top). The correlation is clearly meaningful and sufficient for ranking several thousand molecules produced by the DA. The differences from the actual ranking order are unlikely to be significant. For reference, the average SA score was 3.8, while the average proxy value was 37.9.

Including the SA score in the QUBO formulation puts the energies of the DA-generated molecules in the range from -52.0 to -33.0. In contrast, the random sample energies range from 7,100 to 17,000,000. In a separate test of 100,000 random molecules, the lowest energy was equal to 5,200 indicating that constraints remained unsatisfied in all the 100,000 attempts. Brute force sampling is not a viable strategy for the libraries of this size and larger.

4. Conclusion

In this paper, we have presented a new technique for de-novo drug design using the quantum-inspired Digital Annealer. In summary, the DA improves the drug discovery process in two key ways. Firstly, it allows for rapid combinatorial enumeration of billions of possible molecules, larger than many alternatives in the market are able to review. Due to increased evaluated space, the DA is more likely to identify higher quality lead molecules, simply due to the fact that there are vastly more molecules to choose from. Secondly, despite scanning a broader space, the Digital Annealer can do so in a fraction of the time needed by alternative solutions, completing a single scan of the library, which generates 128 samples, in approximately 0.6 seconds, it is possible to generate thousands of candidate molecules in a matter of minutes. This elucidates the usefulness of our technology for successful and rapid

design of lead-like compounds that are synthetically feasible and potentially innovative with respect to existing intellectual properties.

*Corresponding authors:

Dr. David Snelling: dave.snelling@uk.fujitsu.com

Dr. Shahar Keinan: skeinan@polarisqb.com

REFERENCES:

1. Fisher, Jill A., Marci D. Cottingham, and Corey A. Kalbaugh. 2015. "Peering into the Pharmaceutical 'pipeline': Investigational Drugs, Clinical Trials, and Industry Priorities." *Social Science & Medicine* 131 (April): 322–30.

2. Ruddigkeit, Lars, Ruud van Deursen, Lorenz C. Blum, and Jean-Louis Reymond. 2012. "Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17." *Journal of Chemical Information and Modeling* 52 (11): 2864–75.

3. Whitehorn, Jamie, and Cameron P. Simmons. 2011. "The Pathogenesis of Dengue." Vaccine 29 (42): 7221–28.

4. WHO report:
   https://www.who.int/immunization/research/development/dengue_q_and_a/en/

5. Schneider, Petra, and Gisbert Schneider. 2016. "De Novo Design at the Edge of Chaos." *Journal of Medicinal Chemistry* 59 (9): 4077–86.

6. https://www.pharma-iq.com/pre-clinical-discovery-and-development/articles/an-effective-way-to-apply-ai-to-the-design-of-new-drug-lead-compounds

7. Frush, Elizabeth Hatcher, Sivakumar Sekharan, and Shahar Keinan. 2017. "In Silico Prediction of Ligand Binding Energies in Multiple Therapeutic Targets and Diverse Ligand Sets-A Case Study on BACE1, TYK2, HSP90, and PERK Proteins." *The Journal of Physical Chemistry. B* 121 (34): 8142–48.

8. Yokokawa, Fumiaki, Shahul Nilar, Christian G. Noble, Siew Pheng Lim, Ranga Rao, Stefani Tania, Gang Wang, et al. 2016. "Discovery of Potent Non-Nucleoside Inhibitors

of Dengue Viral RNA-Dependent RNA Polymerase from a Fragment Hit Using Structure-Based Drug Design." *Journal of Medicinal Chemistry* 59 (8): 3935–52.

9. Keinan, S., E. Hatcher Frush, and W. J. Shipman. 2018. "Leveraging Cloud Computing for In-Silico Drug Design Using the Quantum Molecular Design (QMD) Framework." *Computing in Science Engineering* 20 (4): 66–73.

10. Schneider, Gisbert. 2019. "Mind and Machine in Drug Design." *Nature Machine Intelligence* 1 (3): 128–30.

11. Brown, Nathan, Marco Fiscato, Marwin H. S. Segler, and Alain C. Vaucher. 2019. "GuacaMol: Benchmarking Models for de Novo Molecular Design." *Journal of Chemical Information and Modeling* 59 (3): 1096–1108.

12. Preuer, Kristina, Philipp Renz, Thomas Unterthiner, Sepp Hochreiter, and Günter Klambauer. 2018. "Fréchet ChemNet Distance: A Metric for Generative Models for Molecules in Drug Discovery." *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1803.09518.

13. Glover, Fred, Gary Kochenberger, and Yu Du. 2018. "A Tutorial on Formulating and Using QUBO Models." *arXiv [cs.DS]*. arXiv. http://arxiv.org/abs/1811.11538.

14. Ertl, Peter, and Ansgar Schuffenhauer. 2009. "Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions." *Journal of Cheminformatics* 1 (1): 8.