A Computational Intelligence Methodology for Unique, Low-Energy Glutamic Acid Isomers Discovery

Mostafa M. H. Ellabaan[1], Ong Yew Soon[1] and Hengye Man[2]

[1]School of Computer Engineering, Nanayang Technological University, Singapore.
[2]Department of Biology, Boston University, Boston, MA, USA.

***Abstract*—Computational intelligence has been largely inspired by the human brain, but has been seldom used for understanding the molecular mechanism of human brain functions of learning and memory. This study proposes, for the first time, a computational intelligence methodology inspiring memetic computing to discover, at the molecular level, a central component of the glutamatergic system of the human brain, the glutamic acid. Glutamic acid is the most abundant amino acids in the central nervous systems (CNS) in the human brain where serving as an excitatory neurotransmitter. Recent studies have shown its role in the functions of learning and memory as well as pathogenesis of many neurodegenerative disorders. Unlike most neurotransmitters, glutamic acid is characterized by a high degree of chemical flexibility. As such, it possesses a large number of low-energy conformations, which allow for selective interaction with various transporter and receptor proteins, comprising the glutamatergic systems in the CNS. Identifying these low-energy stable conformations of glutamic acid may open new doors for understanding the details of the interaction inside the glutamatergic system, allowing better understanding of the learning process as well as the design of new, yet effective drugs for neurodegenerative disorders, including Alzheimer disease. In this study, a novel molecular memetic algorithm (MOL-MA) has been proposed. MOL-MA is mainly composed of specially designed molecular evolutionary operators coupled with a life-time learning procedure and valley adaptive clearing scheme for the discovery of multiple precise glutamic acid isomers. The energies as well as first and second derivatives of glutamic acid structure, used in this algorithm during the evolution process, are calculated using an ab initio approach, namely, Hartree–Fock method, with the STO-3G basis set. The proposed algorithm has been compared against several of the state of the art algorithms. Experimental results have shown the advantage performance of the proposed algorithm in terms of number of uncovered isomers and computational cost incurred. This study has successfully led to a discovery of unprecedented database of 523 glutamic acid isomers.**

Index Terms— Glutamic acid, neurotransmitters, Isomers, Optimization, Memetic and Evolutionary Computation

I. INTRODUCTION

GLUTAMIC ACID exists in ample quantity in almost every cell type and more importantly in mammalian brain cells. Glutamic acid is known as a fast synaptic neurotransmitter in the central nervous system (CNS), which plays an essential role in several form of communication ranging from excitatory transmission to more complex signal processing necessary for learning and memory [1-4]. It has also been shown to be involved in several pathological conditions, including epilepsy, Alzheimer's disease, Parkinson's, and schizophrenia [5, 6]. The ability of glutamic acid to participate in such a wide range of activity is mostly due to the large number of transporters and receptors proteins that are all not only accessible to, but also activated by. Due its chemical flexibility, it can, in addition, accept many different low-energy conformations. As such, it has a greater ability in selective binding and varying receptor modulation effects when interacting with neuronal transporters and receptors. In particular, some conformations may work as activators, some other as inhibitors [7]. Inhibition of neuronal transporters and/or over-activation of receptors are believed to be involved in many destructive molecular cascades in neuronal cells[8].

As recent studies have shown the involvement of the glutamic acid in the functions of learning and memory and pathological disorders, chemists and neurologists have started to synthesize some Glutamate isomers for the purpose of studying their potency and inhibition effect on the glutamatergic system components. Nine isomers are thus successfully designed theoretically using the stereochemistry principles with the guide of Newman's diagram. Few isomers, however, have been experimentally identified to date. Moreover, experimental studies are most often challenging and more importantly expensive. Alternatively, glutamic acid isomers can be efficiently identified through optimization methods which, among the computational methods, is the most commonly used for such a purpose.

In the context of optimization, the task of isomers identification in molecular systems is equivalent to the undertaking of

multiple minima identification. However, finding multiple optima for molecular and chemical systems is far more challenging than that on mathematical test functions. The core challenges emerge from the presence of a large number of critical points including maxima and saddles - which increase exponentially in chemical and molecular systems - may hampered the progress of search for local optima. Finding local optima using high fidelity first-principles calculations also makes the search even more challenging due to the high computational cost of solving the Schrödinger equations. The challenge escalates exponentially with increasing molecular size. In particular, glutamic acid consists of 19 atoms where each atom in the molecule has three additional degrees of freedom. With each additional degree of freedom, there is an inseparably intertwined exponentially scaling positional optimization problem. Single first-principles calculation evaluation using HF/STO-3G for the glutamic acid, for instance, may take up to half-hour on a 2-Quad-core Lenovo machine. Moreover, molecular systems involve many rules and constraints that must be satisfied during the search process.

To identify isomers, optimization methods proposed in the last decade can be classified as conventional or stochastic approaches [9]. Generally, conventional approaches are established to have the benefit of fast convergence to high quality or precise solutions. However, useful a *priori* knowledge on the regions where glutamic acid isomers may lie within the vast landscape is mandatory for the approaches in this category to fare well. On the contrary, the stochastic approaches can identify the good regions of interest where isomers may lie reliably, but convergence to high quality or precise solutions are typically slow and inefficient.

Conventional approaches make use of domain knowledge on the initial guess where the isomers may lie. Due to the use of domain specific a *priori* information, conventional methods are well established to converge efficiently to a precise glutamic acid isomer. Pertaining to stochastic search approaches, population based schemes are typically considered. Nevertheless, in contrast to conventional approaches, significantly fewer stochastic population based approaches have been explored for locating isomers of molecular systems generally and of Glutamic acid, specifically, making it a fertile area for further research investigations to date. In light of this, this study explores for the first time, the use of memetic optimization method that attempts to fill this gap in order to identify glutamic acid isomers. A novel valley adaptive clearing molecular memetic algorithm (MOL-MA) is thus proposed.

The efficacy of MOL-MA against other recent algorithms proposed in the literature on first-principles calculations of glutamic acid has been investigated. In particular, the search performances of MOL-MA and other counterpart algorithms considered for glutamic acid isomers discovery have been studied in terms of computational effort and number of uncovered glutamic acid isomers. Experimental results have shown that the proposed algorithm has attained the largest set of glutamic acid isomers with minimal computation cost. This study has led to the discovery of unprecedented database of 523 glutamic acid isomers.

The report is organized in the following manner: Section II provides a brief mathematical definition of the glutamic acid isomers problem. The algorithmic details of the suggest MOL-MA are, then, presented in Section III, while Section IV present experimental results and Section V presents the uncovered set of glutamic acid isomers. A brief conclusion and future work is finally presented in Section VI.

II. Problem Statement

The fitness or energy landscape has proven to be a useful conceptual framework in various fields, from biological evolution and protein folding to combinatorial and molecular optimization [10-12]. A landscape can be formally defined as an ordered set of three components $\mathbf{L} = (\mathbf{X}; f; \varphi)$, where X is the set of possible solutions or structural configurations, $f(\mathbf{x})$ is the fitness function, and φ denotes the distance measure between two structural configurations in $\mathbf{X}$.

Mathematically, isomers can be defined as a stationary point ($\mathbf{x}_i$) on landscape $\mathbf{L}$, where the gradient vanishes and the Hessian matrix [13] is a positive definite, and can be expressed as:

$$\mathbf{X}_I = \left\{ \mathbf{x}_i \middle| \left(\left\| \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \right\| = 0 \right) \right\} \quad (1)$$

where $\mathbf{x}_i \in \mathrm{R}^d$, d is the dimensional size, $f(\mathbf{x}_i) \in \mathbf{R}, \mathbf{X}_I$ is the set of glutamic acid isomers. Hessian matrix H_i is positive definite.

More often, scientists are generally interested on glutamic acid isomers that are deemed as "good". To formalize the concept of "good" isomers, we define the ideal sample set $\mathrm{X_I}$ having the following properties:

- contains all structures with energies below a user-defined threshold, i.e. ($f(\mathbf{x}_i) < f_{max}$).
- contains no duplicates, i.e., there do not exist any $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{X}_I$ s.t. $\|f(\mathbf{x}_i) - f(\mathbf{x}_j)\| < j$) and ($\varphi(\mathbf{x}_i, \mathbf{x}_j) < \vartheta$)
- has a precision of $\left\| \frac{\partial f(\mathbf{x}_i)}{\partial \mathbf{x}_i} \right\| < \delta$.

where j and ϑ are the maximum acceptable similarities in the energy and configuration spaces, respectively, δ defines the precision of the gradient norm and $\varphi(\mathbf{x}_i, \mathbf{x}_j)$ denotes a structure similarity metric. In the present study, the Ultrafast Shape Recognition (USR) is considered. It follows naturally that the binding energies of structures in the ideal sample are bounded by $\mathrm{f_{max}}$ and j. A good sample set ($\mathbf{X}_k$) should thus be the one that approximates the ideal set $\mathbf{X}_I$ sufficiently well. Hence, one can test for closeness to $\mathbf{X}_I$ by measuring the cardinality of $\mathbf{X}_k \cap \mathbf{X}_I$.

III. The Proposed Methodology

Memetic computation represents an emerging field that has attracted increasing research attention in the recent few decades, with a growing number of success reported [14]. Memetic computation [15] in its earliest form was introduced as a memetic algorithm (MA), which is a marriage between

population-based global search and life-time learning, where the latter is often referred to as a meme, capable of individual refinement in converging to the precise local optima rapidly. They are inspired by Darwin's theory of natural evolution and Dawkins' notion of a meme. MA has manifested as a form of hybrid global-local approach that facilitates both exploration and exploitation in the search. Up to date, many MAs have been crafted for solving real-world problems more efficiently. Such hybrid algorithms have been applied successfully to solve various medical, scientific and engineering design problems [16], [17], where higher quality solutions are attained more efficiently than traditional evolutionary algorithms [18-20].

In this study, we propose a molecular-based memetic algorithm (MOL-MA), designed for the discovery of multiple low-lying glutamic acid isomers. In MOL-MA, a population of potential glutamic acid isomers is first randomly generated using a specialized molecular initialization procedure. The potential glutamic acid isomers are then assessed as candidate solutions for the problem at hand using a newly designed fitness function. Glutamic acid structures in the population survive to the next generation according to the stochastic universal sampling procedure (Section III.A). The surviving structures then undergo the specially designed molecular-based evolutionary operators of crossover and mutation (Section III.B). To facilitate diversity in the search, evolved glutamic acid isomers then undergo the valley-adaptive clearing scheme (Section III.C) where population individuals belonging to the same valley are segregated from others into niche or valley groups. The elite individual per valley group is then refined using the Berny algorithm and Lamarckian inheritance [21] (Section III.D). Valley elites that satisfy as glutamic acid isomers are archived and all other members of the valley groups are relocated randomly to other regions of the energy landscape to facilitate search exploration towards the discovery of other potential high quality glutamic acid isomers that may exist. In the search process, any newly uncovered glutamic acid isomers are archived and checked against possible duplicates (Section III.E). This process repeats for a maximum number of generations or until the convergence criteria are satisfied.

A. *Molecular Initialization Procedure*

Considering domain knowledge pertaining to glutamic acid molecule, a set of n initial glutamic acid structures is generated to form the initial population. Particularly, a basic glutamic acid molecule, available at [22] and depicted in Fig. 3, is used as a seed to produce the initial population. However, dealing with glutamic acid as set of atoms, ignoring the bonds and structure constraints, is neither conceivable nor rational. In order to grasp structure of glutamic acid seed structure and facilitate further structure processing, a tree representation is adopted as an acyclic graph representation of glutamic acid structure. Each node in the tree is a data-structure representing a specific atom with information about the location of the atom as well as the type and order of the bond that synergize it with its parent. An example for the node considered in this study presented in Fig. 2:

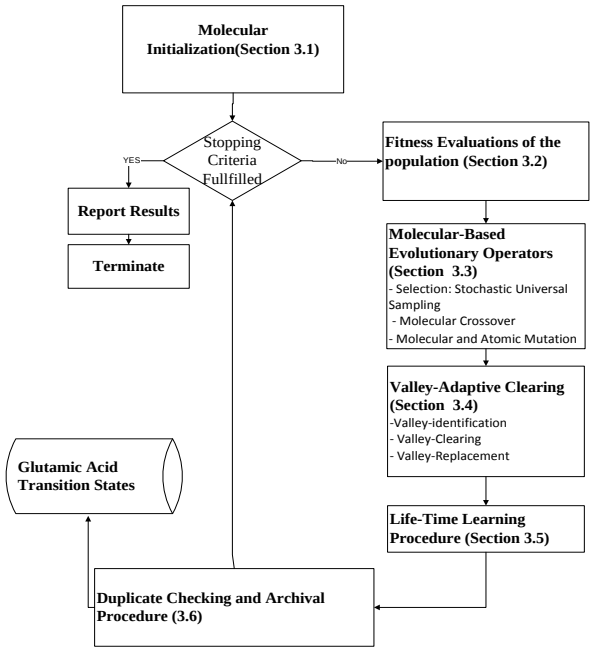


Fig. 1: The Proposed Molecular Memetic Algorithm (MOL-MA)

Atom ID: Atoms identifier in the molecule
Coords: Cartesian Coordinate of the atom
Pbonds: The Type and order of the bond (SC, DC) [1] with the parent

Fig. 2: Basic Tree Node

An example tree representation of glutamic acid seed structure is shown in Fig. 3, using C3 as a pivot atom or root node. Once the tree representation of the glutamic acid seed is available, a random set of mutations (See Section 3.2) is then employed on the seed structure. The randomly mutated variants of the glutamic acid seed structure, with structure similarity higher than 5%, are then considered for the initial population.

B. *Molecular-Based Evolutionary Operators*

The novel molecular reproduction operators, proposed in this study, are briefly described in what follows.

Molecular Crossover: Crossover operates by interchanging genetic information between individuals or chromosomes. The crossover process, considered in this study, between two glutamic structures is depicted in Fig. 5. The aim is to exchange some of the molecular substructures between glutamic acid isomers, thus allowing successful substructures to replicate across populations and generations. For that purpose, glutamic acid isomers in the population are first selected and then aligned. Both chromosomes are randomly. A random atom is then chosen and identified in the chromosomes. Its details as well as the subtree structure are thus exchanged between the parent structures.

[1] Sc: refers to a single covalent bond, while DC refers to a double covalent bond

Molecular and Atomic Mutations: Analogous to biological mutation, the mutation operators in EAs bring in subtle changes to a randomly chosen subset of chromosomes in the MA population or glutamic structures in the present context. We thus propose the following three molecular and atomic mutation strategies for glutamic structures evolution

ST-1. Sub-Trees/substructure Rotation: a set of parent nodes are randomly chosen ($\boldsymbol{\Omega_{S_R}}$). Their structures and sub trees are rotated randomly around the axis of the chosen node and its parent.

$$\forall \mathbf{s} \in \boldsymbol{\Omega_{S_R}}\ \forall \mathbf{x} \in \mathbf{a}(\mathbf{s}), \mathbf{x} = \mathbf{R} \times \mathbf{x} \qquad (2)$$

Where $\mathrm{a}(\mathbf{s})$ set of atoms that fall in the subtree precedent to $\mathbf{s}$, $\mathbf{x}$ is a vector containing the selected atom coordinates and $\mathbf{R}_{(\vartheta_1, \langle \mathrm{P}(\mathrm{P}(\mathbf{s})) - \mathrm{P}(\mathbf{s}) \rangle)}$ is the rotation matrix where rotation is performed around $\langle \boldsymbol{P}(\boldsymbol{P}(\mathbf{s})) - \boldsymbol{P}(\mathbf{s}) \rangle$ axis with a random angle ϑ_1 .

ST-2. Leave Nodes/Atoms Rotation: a set of randomly chosen leaves/atoms ($\boldsymbol{\Omega_R}$) are arbitrarily rotated. Each leave node is rotated around the bond between the grandparent and parent of the node) as follow

$$\forall \mathbf{x} \in \boldsymbol{\Omega_R}, \mathbf{x} = \mathbf{R} \times \mathbf{x} \qquad (3)$$

Where $\mathbf{x}$ is a selected atom and $\mathbf{R}_{(\vartheta_2, \langle \mathrm{P}(\mathrm{P}(\mathbf{x})) - \mathrm{P}(\mathbf{x}) \rangle)}$ is the rotation matrix where rotation is performed around $\langle \mathbf{P}(\mathbf{P}(\mathbf{x})) - \mathbf{P}(\mathbf{x}) \rangle$ axis with a random angle ϑ_2 .

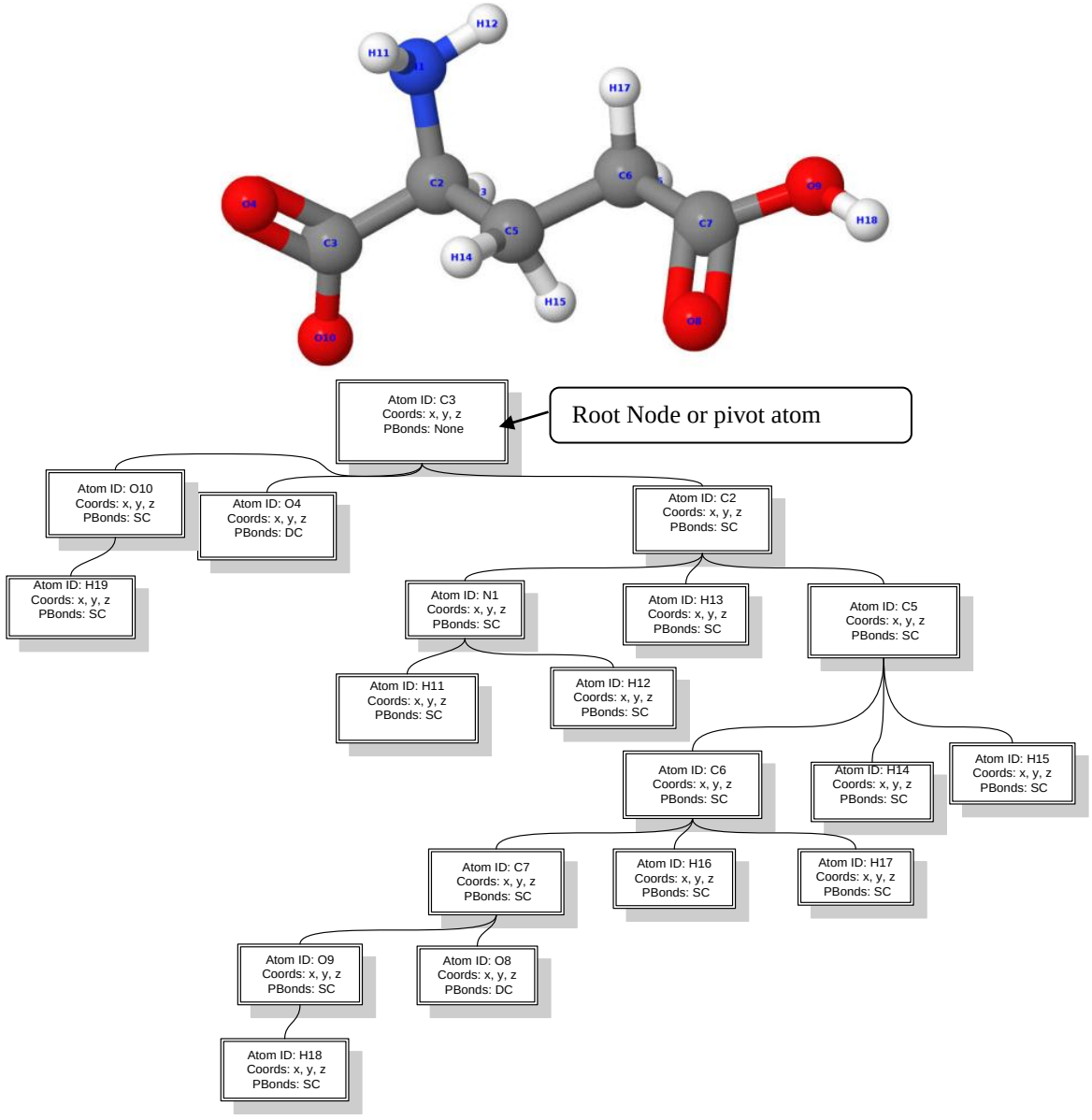


Fig. 3: Glutamic Acid and Its Tree Representation, Taking C3 as A Pivot Atom or Root Node.

ST-3. Nodes/Atoms Translation: a set of randomly chosen atoms/nodes ($\boldsymbol{\Omega_T}$) are arbitrarily translated such that their bonds with neighbours are maintained within length constraints detailed in Table I.

$$\forall \mathbf{x} \in \boldsymbol{\Omega}_{\mathbf{T}}, \mathbf{x} = \alpha_{b_T(\mathbf{P(x)},\mathbf{x})}(\mathbf{P(x)} - \mathbf{x})\ s.t\ \alpha_{b_T(\mathbf{P(x)}-\mathbf{x})} \in [\alpha_{b_T(\mathbf{P(x)}-\mathbf{x})_{min}}, \alpha_{b_T(\mathbf{P(x)}-\mathbf{x})_{max}}] \quad (4)$$

Where $\mathbf{x}$ is a selected atom and $\mathbf{P(x)}$ is parent atom of $\mathbf{x}$ and $b_T(\mathbf{P(x)}, \mathbf{x})$ is the bond type as shown in Table I, $\alpha_{b_T(\mathbf{P(x)},\mathbf{x})}$ is a random number between the minimum $(\alpha_{b_T(\mathbf{P(x)}-\mathbf{x})_{min}})$ and maximum $(\alpha_{b_T(\mathbf{P(x)}-\mathbf{x})_{min}})$ bond lengths.

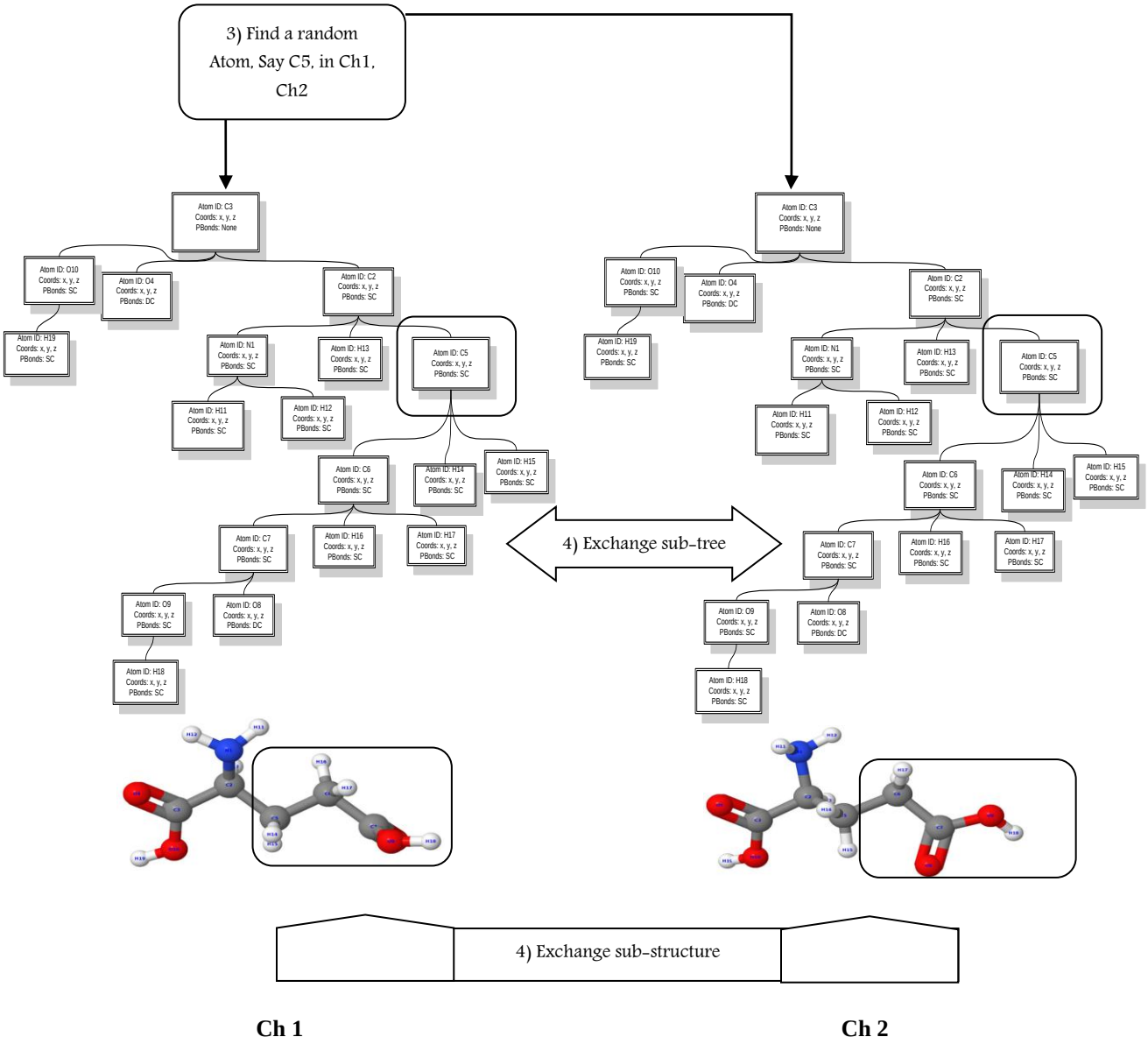


Fig. 4: Molecular Crossover Operator

TABLE I
BOND LENGTHS CONSTRAINTS, AS CONSIDERED BY DOMAIN EXPERTS.

Bond Type $(b_T(\mathbf{y}-\mathbf{x}))$	Min length ($\alpha_{b_T(\mathbf{y}-\mathbf{x})_{min}}$) A0	Max length ($\alpha_{b_T(\mathbf{y}-\mathbf{x})_{max}}$) A0
O-H	0.95	1.50
C-H	0.95	1.15
N-H	1.06	1.12
N-C	1.47	2.10
O-C	1.43	2.15
C-C	1.20	1.54

C. *Valley-Adaptive Clearing Scheme*

The valley-adaptive clearing scheme is designed to adapt to non-uniform width of valleys in the problem landscape and considered here to maintain the diversity of the evolving populations. The core idea of the valley-adaptive clearing scheme is to cluster the population individuals into groups that share common valleys and subsequently clearing the lowest fit individuals from each group by relocating them to random regions of the solution space. In particular, the valley-adaptive

clearing scheme is composed of three core phases. The valley identification phase categorizes the population of individuals into groups of individuals sharing the same valley, denoted as Vgroups. Subsequently, the dominant individual (i.e., in terms of fitness value) of a valley group or $\mathbf{V}_{\text{id}}$ is archived if it represents a unique glutamic acid isomer, while all other members of the same group undergo the valley replacement phase, where relocation of these individuals to new basin of attractions or valleys are made so that unique good quality glutamic acid isomer elsewhere may be uncovered. In the event that glutamic acid isomer exists in a valley group, all individuals of the group will undergo the valley clearing phase where elite individuals (x_{elite}) are ensured to survive across the search generation while all others are relocated to new valleys or basin of attractions, of the fitness landscape.

D. Life-Time Learning Procedure

The Berny algorithm is considered here as the life-time refinement procedure in MOL-MA due to its ability in converging to precise glutamic acid isomers efficiently. Considering the valley groups V_{groups} identified previously in section 4.4, the elite individual ($\mathbf{x}_{elite}$) of each valley group will undergo the Berny algorithm to arrive at a precise glutamic acid isomer found in the near vicinity of $\mathbf{x}_{elite}$. For each $\mathbf{x}_{elite}$ vector, the Berny algorithm proceeds based on the energy and its derivatives at $\mathbf{x}_{elite}$ and iterates, considering using L-BFGS updates until a convergence in term of maximum number of iterations or precision configured is fulfilled.

E. Duplicate Molecular Structure Identification of Configurations and Archival Procedure

In chemistry or physical sciences, both rotated and translated molecular configurations of a given configuration are considered identical, i.e. they are invariant to translation and rotation. In MOL-MA search, only unique representative configurations in the population are considered to facilitate discovery of multiple high quality glutamic acid isomers. To avoid potential wastage of resources due to redundant duplicate configurations, a procedure to identify such structures is designed based on USR, the Ultrafast Shape Recognition. USR measures the similarity between molecular structure shapes with a signature vector of twelve atomic distance statistics, **U**. This signature captures the mean, standard deviation and asymmetry of the distances from each atom in the structure to four anchor points: (a) the structure's centroids, b) the atom closest to a, c) the atom furthest from a and d) the atom furthest from c. This signature presents an effective measure, invariant to translational and rotational symmetries. The similarity $s(\mathbf{x}_i, \mathbf{x}_j)$ between two molecular structures $\mathbf{x}_i$ and $\mathbf{x}_j$, as inverses of distances between the signatures, can hence be easily defined by the inverse-scaled Manhattan distance as follows:

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{1+\frac{1}{12}\sum_{k=1}^{12}\left|\mathbf{U}_k^{\mathbf{x}_i} - \mathbf{U}_k^{\mathbf{x}_j}\right|} \quad (5)$$

where $\mathbf{U}_k^{\mathbf{x}_i}$ denotes the k^{th} component of $\mathbf{x}_i$'s USR signature. $s(x_i, x_j)$ maps [0; 1) with 1 indicating maximum similarity. $s(x_i, x_j)$ is also symmetric since $s(x_i, x_j) = s(x_i, x_j)$.

During the evolutionary process, molecular signatures are calculated for the glutamic acid structures and two glutamic acid structures (x_i, x_j) with $s(x_i, x_j) > \vartheta$ are considered as identical [6]. In such events, one of the glutamic acid structures in the population is replaced with a random configuration.

During duplicate identification and replacement process, all non-duplicated discovered structures are archived using a multi-map data structure. Multi-maps are associative data structures that store elements indexed by keys to ensure that comparisons and duplicate checking are performed efficiently. Multi-maps, in particular, allows for quick access and rapid retrieval of archival elements based on key values. From the computational cost prospective, the worst case access time is $O(m + 1)$ where m is the number of keys and worst case insertion time is $O(\log s_n)$ where s_n is the number of archived solutions. In our work, the archive is indexed based on the number of significant figures of binding energies as well as USR.

IV. EXPERIMENTAL STUDY

In this section, we study the efficacy of the proposed molecular memetic algorithm (MOL-MA) for locating glutamic acid isomers and pit it against several existing state-of-the-art optimization methods, including the stochastic multi-start local search (SMLS) [23], sequential niching memetic algorithm (SNMA) [24], as representatives of recent advances in niching algorithms and MAs. All methods work as baselines for comparison on realistic glutamic acid isomers optimization, using a comprehensive set of performance measures that are described in Table II. The energy, gradient and eigenvalues are calculated using first-principles approach at the Hartree–Fock level, with the STO-3G basis set [25], and all computations are executed on the PDCC cluster. The experimental settings and numerical results obtained are reported. In particular, the algorithmic parameter settings used in the present study are listed in Table III. All algorithms are initialized with a population size of 10 molecules and iterate until a maximum of 50 generations exceeded or the algorithm incurs a maximum number of 4 x 10^4 Gaussian calls. For the sake of statistical significance, the experimental results summarizing 10 independent experimental runs are reported. Note that more runs are deemed to be impractical due to the intractable computational resources that would be required.

In this section, we present next on the numerical results of MOL-MA pitted against other state-of-the-art evolutionary and memetic approaches for the discovery of isomers. For the sake of brevity, the unique properties of the algorithms considered in the present study, i.e., in terms of life time learning, specialized molecular evolutionary operators, niching operators and domain knowledge required are summarized in Table IV.

The performance efficacy of the algorithms assessed according to their success rate in Table V. According to the table, it is observed that only algorithms that consider local search operations have attained a non-zero success rate. Among those successfully attained a non-zero success rate, algorithms with molecular operators has observed to attain not only the highest

success rate of 100 but also the largest set of isomers uncovered, see Table VI. In particular, canonical SNMA method for example has attained a significant number of isomers. However, a better result has been attained for SNMA* which consider the molecular operators, where number of isomers arrives at 38 isomers, with improvement rate of 680% over canonical SNMA. Among all, MOL-MA has been observed to attain the largest number of isomers, with significant improvement over other algorithms ranging from 170-1600% improvement rate.

TABLE II
DESCRIPTIONS OF THE PERFORMANCE MEASURES CONSIDERED

Performance Measure	Description
Success Rate	The ratio of the successful runs to the total number of runs (10), where a run is called successful if the algorithm attains at least an isomer.
Number of Uncovered Glutamic Acid Isomers	Number of uncovered glutamic acid isomers (#IS) is used here to measure the efficacy of the algorithms.
Gaussian Calls	Number of Gaussian calls that the algorithm incurred during the execution

TABLE III
ALGORITHMIC PARAMETERS SETTINGS

Parameter	Value
ϑ	0.05
j	0.0001
δ	3.5 x 10-4
f_{max}	-100
Pmutation	0.5
Pcrossover	0.5
rnicheRadius	10 A0

TABLE IV
PROPERTIES OF THE ALGORITHMS CONSIDERED (Y DENOTES THE EXISTENCE OF A PROPERTY WHILE N REFERS TO ITS ABSENCE)

Properties	SMLS	CGA	SNMA	SNMA*	MOL-MA
Life-time learning	Y	Y	Y	Y	Y
Molecular Operators	N	N	N	Y	Y
Niching Operator	N	N	Y	Y	Y
Domain Knowledge Required	Initial guesses	N	N	N	N

TABLE V
SUCCESS RATE OF ALGORITHMS

Method	MOL-MA	SNMA*	SNMA	SMLS	CGA
Success rate	100	100	60	45	0

TABLE VI
NUMBER OF UNCOVERED ISOMERS BY DIFFERENT ALGORITHMS

Method	MOL-MA	SNMA*	SNMA	SMLS	CGA
Number of Uncovered Isomers	47	38.5	5.3	2.9	0
Improvement rate	100%	120%	878%	1603%	NA

The computational requirements of the algorithms considered in the present study are also summarized in Table VII. From the results, SNMA, CGA, and SMLS are observed to incur the most computational effort, exhausting the entire computational budget allowable. In spite of the largest percentage of precise isomers uncovered as discussed previously, MOL-MA remains to display superiority in terms of computational efficiency over the other state-of-the-art algorithms. In particular, the cost reductions of MOL-MA over the other counterpart algorithms are summarized in Table VII, indicating that MOL-MA maintains an average cost reduction of 345% over SNMA* and 490% over other algorithms.

TABLE VII
NUMBER OF GAUSSIAN CALLS INCURRED BY DIFFERENT ALGORITHMS

Method	MOL-MA	SNMA*	SMLS	CGA
Number of Gaussian Calls	6,771	30,185	40,000	40,000
Cost Reduction in percentage	100%	345%	490%	490%

V. Study on the Uncovered Set of Glutamic Isomers and Landscape Correlation Analysis

Since the proposed algorithm attained the largest set of glutamic acid isomers, the bond-aware molecular memetic algorithm has been allowed to run for longer with a 50-individual population and 10,000 generations with a maximum of a million gradient evaluations. A database of 523 isomers has been uncovered. The 10 glutamic isomers uncovered whose energies are the lowest have been reported in Table VIII. Among them, the folded and extended forms have been observed, while the first form has been observed to be with the minimal energy. Such a form is reported in literature to activate the receptor. Despite its importance for receptor activations and neuronal communications, the abundance of this form may cause a fatal molecular cascade in neuronal cells, leading to the neurotoxicity. Nonetheless, the energy difference between the folded and extended forms does not appear to be very high as indicated in Table VIII. In particular, the energy difference is below 1.5 KJ/MOL. This may suggest that the transformation from a folded to extended isomers can be manipulated quite easily. Uncovering such a transformation pathway may provide new therapeutic tools for neurotoxicity. However, the full details of such a transformation pathway would require knowledge on glutamic acid transition states lying between the extended and folded isomers.

The density distribution of the uncovered set of glutamic acid isomers in terms of relative energy and structure dissimilarity of glutamic acid isomers have been depicted in Fig. 5. It is clear from the Fig. that glutamic acid isomers are denser in the region between 25% and 45% USR structure dissimilarity and from 3 to 17 KJ/MOL as well as from 27 to 38 KJ/MOL. Moreover, more than 90% of the isomers uncovered were below 45 (KJ/MOL) and 45% USR Dissimilarity, indicating the most rugged regions.

In order to measure the difficulty of the glutamic acid isomers problem, we further study the landscape correlation of the glutamic acid isomers problem through the mean of fitness-distance correlation [26]. Fitness-Distance Correlation (FDC) (Eqn. 15) is the Pearson product moment correlation between the energy differences and the structural differences of the samples to the lowest energy isomer, shown below.

$$\text{FDC} = \frac{cov(\partial E;\partial D)}{\sigma(\partial E)\sigma(\partial D)} \qquad (15)$$

where $cov(\partial E;\partial D)$ is the covariance function, $(\partial E)$and (∂D) are the energy difference and USR dissimilarity between each solution and the lowest energy solution respectively. Likewise, $\sigma(\partial E)$ and $\sigma(\partial D)$ represent the standard deviations of the energy differences and the structural dissimilarity.

According to FDC, a landscape can be categorized into well-ordered, rough or deceptive landscape. In particular, high correlation (FDC >60) indicates that isomers are well-ordered and optimization methods can locate them quite easily, while small correlation indicates a rough landscape where an optimization algorithm may be mislead to sub-optimal region, and negative correlation indicates a deceptive landscape where the global minimum is located among high energy solutions. The landscape analysis on the uncovered set of glutamic acid isomers has revealed that the problem has a rough landscape with FDC of $4.6 \ll 60$.

VI. Conclusion

Due to the importance of glutamic acid isomers in both neuroscience and structural bioinformatics, this study will be utterly dedicated to the discovery of glutamic aci isomers. A mathematical formulation and description of this non-linear programming problem has been defined. A novel valley adaptive clearing molecular memetic algorithm (MOL-MA) that requires no a *priori* knowledge for the discovery of the glutamic acid isomers has been proposed. The proposed molecular memetic algorithm is composed of several core components, namely a specialized molecular structure initialization, an advance fitness function, molecular-based evolutionary operators, valley adaptive clearing scheme and the Berny-based life time learning procedure. Assessments made against several state-of-the-art approaches in the field, in terms of number of uncovered isomers and computational cost incurred, is conducted to demonstrate the efficacy of MOL-MA.

MOL-MA has been employed on glutamic acid isomers using high fidelity computational models based on first-principles calculations. The results of the MOL-MA application on glutamic acid have led not only to a discovery of previously known isomers, but also to locate newly established ones. The insights gained through isomers uncovered could provide better understanding of the thermodynamics properties and the isomerization processes in glutamic acid and also provided an organized manner for further studies to be held on glutamic acid isomers.

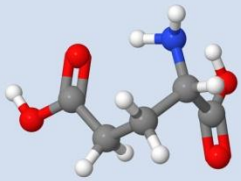
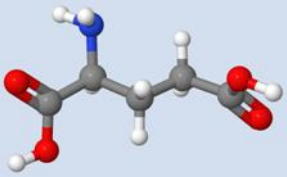
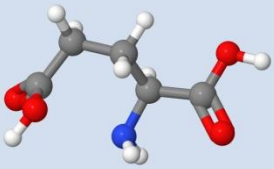
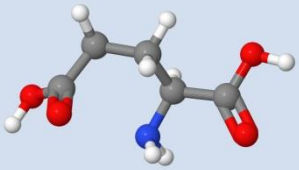
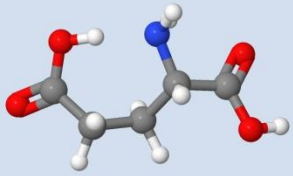
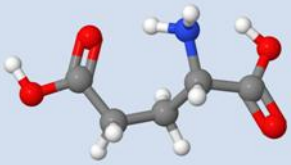
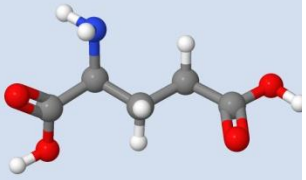
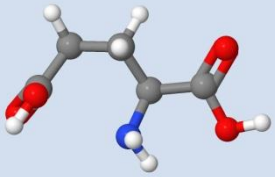
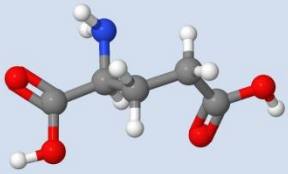
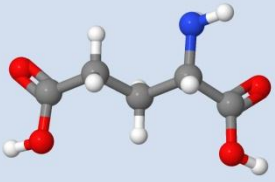
The success on glutamic acid encourages the efforts to further investigate the applicability of MOL-MA for other molecular systems such as drug molecules, other amino acids, other neurotransmitters, neuroreceptors and neurotransporters as well as protein folding. Although MOL-MA will be applied to glutamic acid in this study, it can be easily adapted for subsequent studies in computational chemistry and biology. Moreover, the analyses that will be conducted in this study can be replicated for other molecular systems to uncover landscape properties and provide insights into both physical chemists and evolutionary algorithmists.

The interaction of glutamic acid isomers that will be discovered with neurotransporters and neuroreceptors will be also a subject for further investigation. Such study on the interaction among glutamatergic system components is expected to reveal the cover over the magnificent complex neurosystems activities such as cognition, learning and memory. It may also explain the declination of learning with aging. It will also shed the light on the glutamic acid configurations that plays either positive or negative role in both physiological and pathological conditions.

Acknowledgement

This work was supported by IEEE Walter Karplus Summer Research Grant 2011.

TABLE VIII
THE 10 LOWEST ENERGY GLUTAMIC ACID ISOMERS

ID	Isomers	Energy (Hartree) Relative Energy (KJ/MOL)	ID	Isomers	Energy (Hartree) Relative Energy (KJ/MOL)
IS-1 (Folded)		-541.3665 (0)	IS-6 (Extended)		-541.3659 (1.575299772)
IS-2 (Folded)		-541.3663 (0.52509)	IS-7 (Folded)		-541.3659 (1.575299772)
IS-3 (Folded)		-541.3663 (0.52509)	IS-8 (Folded)		-541.3657 (2.100399696)
IS-4 (Extended)		-541.366 (1.31274981)	IS-9 (Folded)		-541.3655 (2.62549962)
IS-5 (Folded)		-541.3659 (1.575299772)	IS-10 (Extended)		-541.3654 (2.888049582)

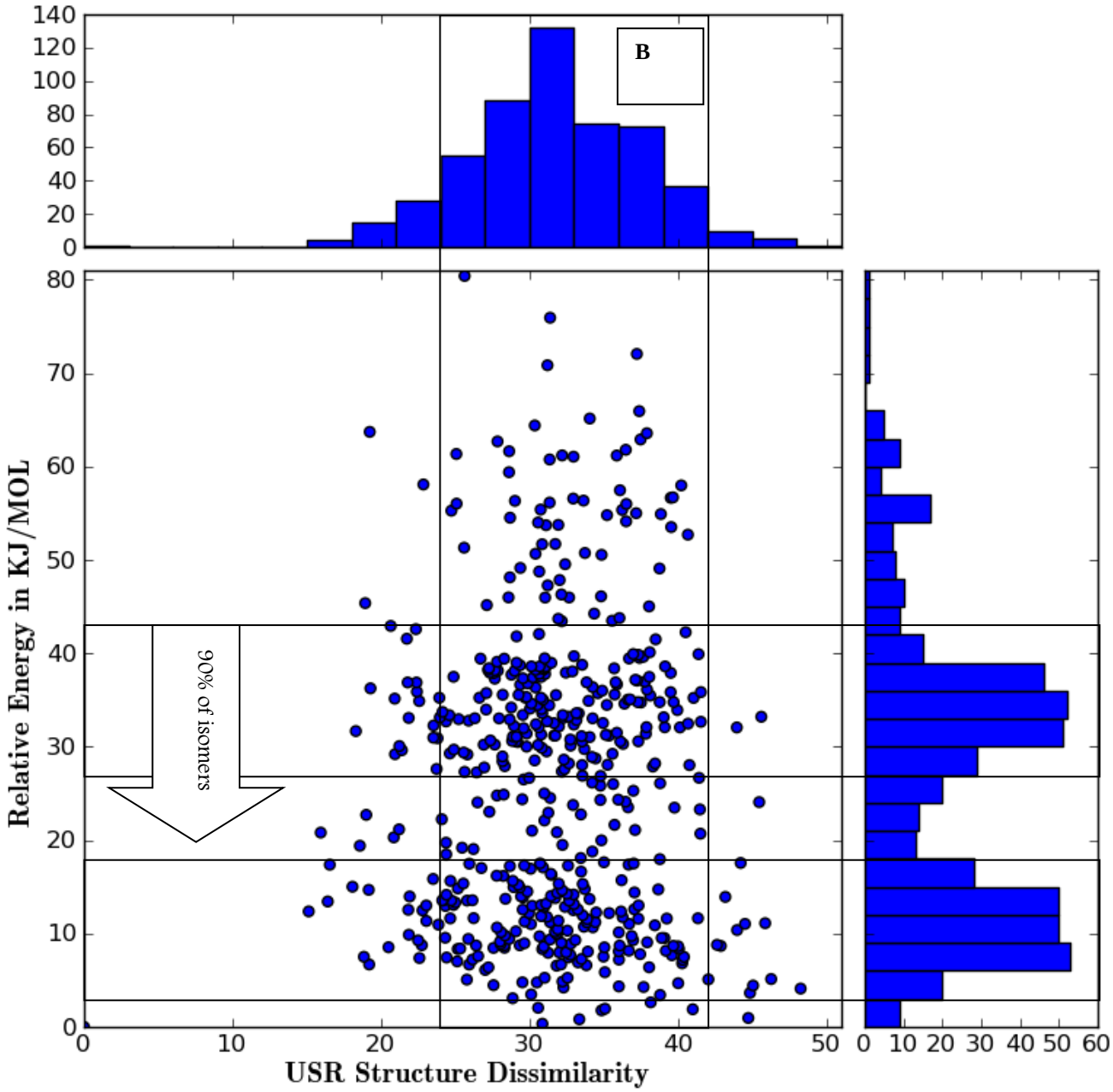


Fig 5: Scatter Plot With Marginal Histograms of USR Structure Dissimilarity Versus Relative Energy for the Glutamic Acid Isomers Uncovered.

References

[1] J. E. Mullins, Synthesis of Conformationally-Restricted Glutamate Analogs Via Novel Ring-Forming Strategies, University of Montana, 2007.

[2] A. Dinsmore, P. M. Doyle, and D. W. Young, Synthesis of Glutamate Agonists and Anatagonists by a Ring Switching Strategy, *Tetrahedron letters* 36 (1995) 7503-7506.

[3] S. Shuto, S. Ono, H. Imoto, K. Yoshii, and A. Matsuda, Synthesis and Biological Activity of Conformationally Restricted Analogues of Milnacipran:(1 S, 2 R)-1-Phenyl-2-[(R)-1-Amino-2-Propynyl]-N, N-Diethylcyclopropanecarboxamide Is a Novel Class of Nmda Receptor Channel Blocker, *Journal of medicinal chemistry* 41 (1998) 3507-3514.

[4] W. J. McEntee and T. H. Crook, Glutamate: Its Role in Learning, Memory, and the Aging Brain, *Psychopharmacology* 111 (1993) 391-401.

[5] J. T. Coyle and P. Puttfarcken, Oxidative Stress, Glutamate, and Neurodegenerative Disorders, *Science* 262 (1993) 689.

[6] L. Turski and W. Turski, Towards an Understanding of the Role of Glutamate in Neurodegenerative Disorders: Energy Metabolism and Neuropathology, *Cellular and Molecular Life Sciences* 49 (1993) 1064-1072.

[7] K. Shimamoto and Y. Ohfune, Syntheses and Conformational Analyses of Glutamate Analogs: 2-(2-Carboxy-3-Substituted-Cyclopropyl) Glycines as Useful Probes for Excitatory Amino Acid Receptors, *Journal of medicinal chemistry* 39 (1996) 407-423.

[8] L. P. Mark, R. W. Prost, J. L. Ulmer, M. M. Smith, D. L. Daniels, J. M. Strottmann, W. D. Brown, and L. Hacein-Bey, Pictorial Review of Glutamate Excitotoxicity: Fundamental Concepts for Neuroimaging, *American journal of neuroradiology* 22 (2001) 1813.

[9] M. Ellabaan, Y. Ong, M. Lim, and K. Jer-Lai, "Finding Multiple First Order Saddle Points Using a Valley Adaptive Clearing Genetic Algorithm," 2010, pp. 457-462.

[10] D. Wales, *Energy Landscapes:[with Applications to Clusters, Biomolecules and Glasses]*: Cambridge University Press Cambridge, 2003.

[11] H. Soh, Y. S. Ong, Q. C. Nguyen, Q. H. Nguyen, M. S. Habibullah, T. Hung, and J.-L. Kuo, Discovering Unique, Low-Energy Pure Water Isomers: Memetic Exploration, Optimization and Landscape Analysis, *IEEE Transactions on Evolutionary Computation* (2009)

[12] C. Brooks 3rd, J. Onuchic, and D. Wales, Statistical Thermodynamics. Taking a Walk on a Landscape, *Science (New York, NY)* 293 (2001) 612.

[13] S. Trygubenko and D. Wales, A Doubly Nudged Elastic Band Method for Finding Transition States, *Chem. Phy.* 120 (2004) 2082-2094.

[14] Y. S. Ong, M. H. Lim, and X. S. Chen, Research Frontier: Memetic Computation - Past, Present & Future, *IEEE Computational Intelligence Magazine* 5 (2010) 24-36.

[15] P.Moscato, "On Evolution, Search, Optimization ,Genetic Algorithms and Martial Arts: Towards Memetic Algorithms," 1989.

[16] F. Neri, J. Toivanen, G. L. Cascella, and Y. S. Ong, An Adaptive Multimeme Algorithm for Designing Hiv Multidrug Therapies, *IEEE/ACM Transactions on Computational Biology and Bioinformatics, Special Issue on Computational Intelligence Approaches in Computational Biology and Bioinformatics* 4 (2007) 264-278.

[17] S. D. Handoko, C. K. Kwoh, and Y. S. Ong, Feasibility Structure Modeling: An Effective Chaperon for Constrained Memetic Algorithms *IEEE Transactions on Evolutionary Computation* 14 (2010) 740-758.

[18] M. N. Le, Y. S. Ong, Y. Jin, and B. Sendhoff, Lamarckian Memetic Algorithms: Local Optimum and Connectivity Structure Analysis 1 (2009) 175-190.

[19] Q. C. Nguyen, Y. S. Ong, and M. H. Lim, A Probabilistic Memetic Framework, *IEEE Transactions on Evolutionary Computation* 13 (2009) 604-623.

[20] Z. Zongzhao, O. Yew Soon, B. N. Prasanth, J. K. Andy, and L. Kai Yew, Combining Global and Local Surrogate Models to Accelerate Evolutionary Optimization, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 37 (2007) 66-76.

[21] Y. S. Ong and A. J. Keane, Meta-Lamarckian Learning in Memetic Algorithm, *IEEE Transactions On Evolutionary Computation* 8 (2004)

[22] NCBI. (2010). *Glutamate Acid Structure*. Available: http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=611

[23] S. Kok and C. Sandrock, Locating and Characterizing the Stationary Points of the Extended Rosenbrock Function, *Evolutionary Computation* 17 (2009) 437-453.

[24] J. E. Vitela and O. Castaños, "A Real-Coded Niching Memetic Algorithm for Continuous Multimodal Function Optimization," 2008, pp. 2170-2177.

[25] F. Jensen, *Introduction to Computational Chemistry*: Wiley New York, 1999.

[26] H. Soh, Y. S. Ong, C. N. Quoc, H. N. Quang, M. S. Habibullah, T. Hung, and J. L. Kuo, Discovering Unique, Low-Energy Pure Water Isomers: Memetic Exploration, Optimization, and Landscape Analysis, *IEEE Transactions On Evolutionary Computation* 14 (2010) 419-437.