UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

LOWER-ENERGY CONFORMERS SEARCH OF TPP-1 POLYPEPTIDE VIA HYBRID PARTICLE SWARM OPTIMIZATION AND GENETIC ALGORITHM

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

GENWEI ZHANG
Norman, Oklahoma
2018

LOWER-ENERGY CONFORMERS SEARCH OF TPP-1 POLYPEPTIDE VIA HYBRID PARTICLE SWARM OPTIMIZATION AND GENETIC ALGORITHM

A THESIS APPROVED FOR THE
GALLOGLY COLLEGE OF ENGINEERING

BY

Dr. Charles D. Nicholson, Chair

Dr. Yihan Shao

Dr. Shima Mohebbi

To my parents, Jixing Zhang and Huanju Ma, neither of whom received higher education, but unconditionally directed and taught me how to behave towards the path to knowledge and truth.

To my wife, Nanxi Li, who continuously supports and encourages me over time. I will keep praying for our harmonious union.

I also dedicated this work to my grandmother, Xiangmei Wang, I hope there is no suffering for you any longer and wish peace with you forever in heaven!

Acknowledgements

Here and now, I want to express my sincere gratitude to several people who have been very helpful to me, especially with this thesis work.

First of all, I would like to give my sincere thanks to my thesis advisor, Dr. Charles Nicholson, for his continuing advice and encouragement over the period of my thesis study. Without his supervision and support, this thesis would not have been completed. Moreover, I want to deliver another special thanks to my advisory committee member, Dr. Yihan Shao, for his ceaseless advice, patience and support throughout the period after my Ph.D. I believe that his broad knowledge and expertise have benefited me tremendously towards my future career and this mentorship would impact me profoundly. I also want to thank Dr. Shima Mohebbi for her valuable advice on this thesis. In addition, I acknowledge the program assistant Nicola Manos, who assisted me a lot along the road of completing my master's degree. Another acknowledgement also needs to be given to the OU Supercomputing Center for Education and Research (OSCER).

Secondly, I want to thank my coworkers in the computational chemistry lab, especially the postdoc Dr. Xiaoliang Pan for his scientific and technical help throughout my work, and also several graduate students, including Vardhan Satalkar, Pei Zheng, and Renmeng Liu, for all the valuable times we have had together during the past two years.

Last but not least, I want to thank my parents for nurturing me for these many years unconditionally and my sister for her continuous financial support even after her marriage. More importantly, I thank my wife for showing up in my life as a godsend, and I will always cherish and protect you as we proceed to this long-life journey together. Please rest your hands on me, and I will bring you to a better life!

Table of Contents

List of Tables

List of Figures

Abstract

Low-energy conformation search on biological macromolecules remains a challenge in biochemical experiments and theoretical studies. Finding efficient approaches to minimize the energy of peptide structures is critically needed for researchers either studying peptide-protein interactions or designing peptide drugs. In this study, we aim to develop a heuristic-based algorithm to efficiently minimize a promising PD-L1 inhibiting polypeptide, TPP-1, and build its low-energy conformer pool to advance its subsequent structure optimization and molecular docking studies. Through our study, we find that, using backbone dihedral angles as the decision variables, both PSO and GA can outperform other existing heuristic approaches in optimizing the structure of Met-enkephalin, a benchmarking pentapeptide for evaluating the efficiency of conformation optimizers. Using the established algorithm pipeline, hybridizing PSO and GA minimized TPP-1 structure efficiently and a low-energy pool was built with an acceptable computational cost (a couple days using a single laptop). Remarkably, the efficiency of hybrid PSO-GA is hundreds-fold higher than the conventional Molecular Dynamic simulations running under the force filed. Meanwhile, the stereo-chemical quality of the minimized structures was validated using Ramachandran plot. In summary, hybrid PSO-GA minimizes TPP-1 structure efficiently and yields a low-energy conformer pool within a reasonably short time period. Overall, our approach can be extended to biochemical research to speed up the peptide conformation determinations and hence can facilitate peptide-involved drug development.

Chapter 1 : Introduction

1.1 The Rise of Immunotherapy

Cancer is one of the leading causes of human death on earth. According to the statistics from National Cancer institute (NCI), 8.2 million people died of cancer-related disease in 2012 with another 1.4 million new cases developed that year worldwide. To make it even worse, the number of newly developed cancer cases each year keeps increasing. For example, an estimated 1.7 million new cancer patients will be diagnosed in the United States in 2018 (https://www.cancer.gov/about-cancer).

Formed by irregularly shaped cells, tumor typically proliferates in an abnormally high growth rate, and it invades neighboring organs and causes damages until functional failure/loss of the hosting organ. Due to cell-to-cell heterogeneity and rapid gene mutations among tumor cells, malignant tumor is hard to predict at its early stages and pharmaceutically difficult to treat. In fact, the current prevailing cancer therapies, i.e. chemotherapy and radiotherapy, are still plagued by their accompanying severe adverse side effects and low efficacies [1-4].

Immunotherapy has become a promising alternative treatment for cancer patients in recent years after a striking success of several clinical trials targeting human immune checkpoints [5-9], mainly referring to the two proteins on T cell surface, i.e. CTLA-4 and PD-1 (Fig. 1-1). These two proteins were discovered by two cancer immunologists James P. Allison and Tasuku Honjo, respectively. Both of the two scientists were awarded the 2018 Nobel Prize in Physiology or Medicine for their contributions to unveil the complexity of molecular mechanisms regulating the immune responses based upon the

two aforementioned immune checkpoints [10], and their important findings laid a solid foundation for the blooming of immunotherapy. In plain words, cancer immunotherapy targets and destroys cancer cells through enhancing human body's immune activities by releasing some of the 'brakes' on the host immune system.

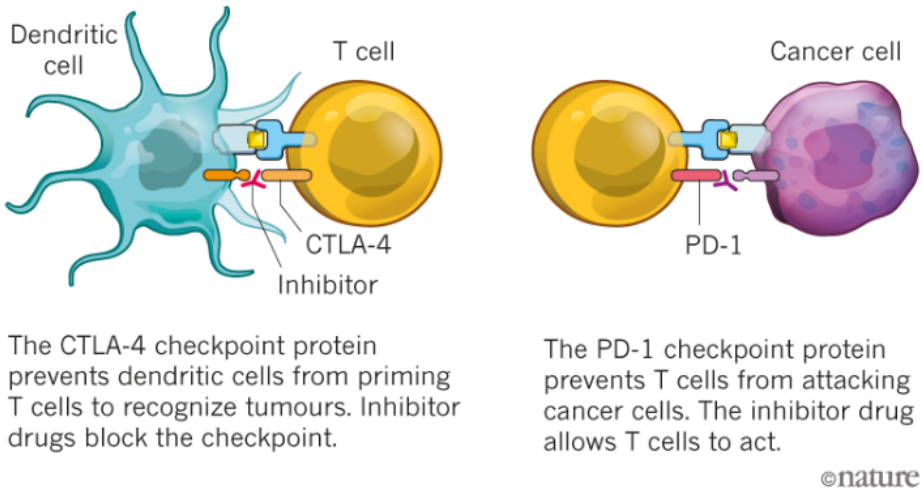


Figure 1-1. The immune checkpoints of T-cells.
Both PD-1 and CTLA-4 are the T-cell receptor responsible for the immune checkpoint. Inhibitor drugs exert their functions through blocking the interactions between PD-1 (or CTLA4) and its ligand.
[*Nature*] REF. 10 © (2018).

In the past several years, multiple novel cancer immunotherapy drugs have been developed as the immune checkpoint antagonists and these approaches have focused on anti-PD-1 and anti-PD-L1 checkpoint inhibitors, anti-CTLA-4 monoclonal antibodies and immune-modulatory drugs. Since 2015, the Food and Drug Administration (FDA) has approved eight antibody-related immunotherapy drugs targeting either one or multiple proteins [11]. In addition, multiple antibody immunotherapy candidates are being actively investigated in clinical trials, and some of them have already shown potentials to advance to real patient treatments [9, 12]. Chinese FDA also approved its first antibody drug used for immunotherapy in the summer of 2018. The pharmaceutical

value of immunotherapy has kept increasing and the immunotherapy drugs market is projected to reach $201.52 billion by 2021. This increases from $108.41 billion in 2016, with an annual growing rate at 13.5% during the forecasting period [13].

1.2 Low-energy Conformation Search of TPP-1

Organic molecules are still the dominating force on the pharmaceutical drug market. Approximately 70% of the top 200 pharmaceutical products (ranked by retail sales in 2016 by the Njardarson Group at The University of Arizona) are developed from small organic molecules. However, adverse side effect caused by the toxicity of organic compounds is a major challenge and concern for late stage clinical studies.

As an alternative strategy, peptide-based pharmaceutical drug development has gained increasing attention in recent years considering its natural biodegradability and lower toxicity. Whilst less toxic, studying and optimizing short peptide structures are more effort-intensive during pre-clinical investigations than designing/synthesizing organic compounds because the three dimensional conformations of short peptides tend to be more flexible while, in contrast, small organic compounds are usually rigid. The structural flexibility of short peptides is mainly attributed to a large number of feasible combinations of the dihedral angles within the peptide backbone. Therefore, this causes the peptide conformational searching space to be a vast number, rendering the computation very expensive. Since peptide low-energy conformers are needed in peptide conformation-related studies, such as peptide docking, binding free energy calculations and peptide structure optimization. Therefore, low-energy conformational search is unavoidable, and naturally, efficient approaches to expedite the peptide low-energy conformer searching is critically needed.

1.2.1 Discovery of Polypeptide TPP-1

TPP-1, with an amino acid sequence of SGQYASYHCWCWRNPGRSGGSK, has been actively studied in the field of cancer immunotherapy. It was shown to be effective in inhibiting cancer cell proliferation and could potentially improve current tumor therapy [14]. TPP-1, which works as a PD-L1 blockade agent, was discovered from biological high-throughput screening assay for its blocking of the signaling pathway to one major immune checkpoint PD-1 on the T-cells. PD-L1 is expressed regularly on the normal human cell surfaces and serves as a key cellular signal for T-cells to distinguish self-cells from non-self-cells. However, during the course of evolution, mutated tumor cells hijack this mechanism by expressing PD-L1 on its own cell surface and take advantage of this protein as an escaping tool of human immune surveillance. And discovery of TPP-1 has opened new ways to the improvement of immunotherapy tools.

1.2.2 Polypeptide TPP-1, as an Alternative for Tumor Immunotherapy

From the perspective of manufacturing pharmaceutical products, the cost of producing peptides is much lower than producing antibodies. Moreover, since antibody molecules are usually larger in size than peptides, their absorption efficiency is much lower in patients' body when compared with small organic molecules or short peptides. However, small organic molecules also suffer from high probability toxicity issues [15-18], especially in the late stage clinical studies. Considering all these above factors, developing short peptides as anti-cancer therapies is more promising. In fact, while peptides as a cancer treatment has attracted the therapeutic focus for many years [19, 20] only recently it gained a lot of momentum in the field of immunotherapy.

TPP-1, which has been shown to be effective in inhibiting cancer cell proliferation, is now regarded as, if not alone, at least a supplemental, promising checkpoint immunotherapy strategy through a proposed working mechanism as shown in Figure 1-2. Despite encouraging pre-clinical results of TPP-1 [14], structural optimization steps will be needed to advance TPP-1 to clinical studies for more comprehensive investigations. In fact, conventional computational methodologies to characterize peptide structures or to find low-energy conformations are still limited to molecular dynamic (MD) simulations using empirical force fields. While the result of MD simulations is usually reliable, the execution of this approach is computationally expensive. Thus, in order to meet the increasing demand of peptide sequence optimization and low-energy structure prediction, it is important to accelerate existing global optimizers through designing novel algorithms.

Finding the low-energy conformers of TPP-1 is beneficial to molecular docking studies to characterize the binding poses and binding free energy calculations to evaluate the binding strength. For instance, in the optimization of TPP-1 sequence to either strengthen or weaken the binding strength to its target, low-energy conformers can be utilized to guide the TPP-1 amino acid mutations. In this project, we aim to find the low-energy conformers of TPP-1 to provide a basis for the subsequent TPP-1: PD-L1 interaction studies both computationally and experimentally.

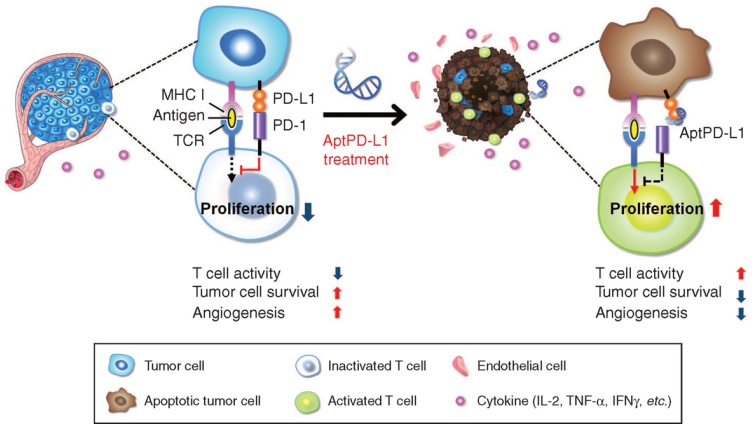


Figure 1-2. AptPD-L1, a DNA aptamer, blocks PD-1/PD-L1 interaction and attenuates T cell suppression.
The figure is adapted from a paper published in 2016 [21], and the inhibition of tumor cell proliferation by TPP-1 is proposed to perform in a similar manner to AptPD-L1.
[*Nucleic Acids*] REF. 21 © (2016).

1.3 Heuristic Approaches

Heuristic, meaning "to find" or "to discover", is a strategy that was derived from the previous experience accumulated from tackling similar problems. The study of heuristic in human decision-making can be dated back to 1970-1980s by psychologists Amos Tversky and Daniel Kahneman [22], however, the formal concept of heuristic approaches originated with Nobel laureate Herbert A. Simon.

Heuristic techniques are not guaranteed to find optimal solutions, but practically speaking, they do offer an approach to efficiently discover high-quality solutions which satisfy goals. Additionally, metaheuristics for optimization are general frameworks which provide rules for guiding heuristic search in optimization problems. Metaheuristic algorithms include well known techniques such as evolutionary algorithms and simulated annealing. Such techniques have been applied to many famous and typical NP-hard problems, such as scheduling problems [23-25], knapsack problems [26, 27] and travelling salesman problems [28, 29], *etc*.

Heuristic or metaheuristic approaches are usually chosen when finding an exact solution to the problem of interest is impractical as it cannot be obtained within an acceptable amount of time or with the available computing resources. There are four typical principles for applying heuristic methods. These principles can be ordered sequentially as: understanding the problem, making a plan, carrying out the plan, and evaluating/adapting. These principles were first proposed in a monograph titled "How to solve it" by a Hungarian mathematician György (George) Pólya in 1945. Nowadays, these principles have been widely applied to address difficult problems in computer science and many other scientific fields.

1.3.1 Heuristic Approaches in Addressing Scientific Problems

In theoretical chemistry, intensive calculations are needed to carry out simulations of biological molecule conformation changes in a dynamic process. Many heuristic methods are implemented in the field of computational chemistry. For example, before Molecular Dynamic (MD) simulations, typically steepest descent algorithm was applied to quickly orient the molecule to a local lower-energy conformation [30]; in computational protein design, multiple stochastic search algorithms were applied and compared [31]; in molecular docking software, simulated annealing approach was used to bring organic molecules to their low-energy conformations during the docking processes [32, 33].

The application of heuristic approaches can also be found in the field of bioinformatics. One example, sequence alignment problem can be very challenging when the size and number of sequences increases, researchers have taken the advantages of simulated annealing, ant colony optimization and particle swarm optimization approaches

to help finding the optimal or near-optimal solutions for efficient DNA sequence alignment [34].

1.3.2 Particle Swarm Optimization and Genetic Algorithm

Here, two metaheuristic approaches, Particle Swarm Optimization and Genetic Algorithm, are discussed.

Particle Swarm Optimization (PSO) is first introduced by James Kennedy and Russell C. Eberhart in 1995 and since then tens of thousands of papers have been published about particle swarms. PSO was inspired by the social behavior of birds and shoals of fish and it is a population-based optimization technique. The main advantage of PSO is that, with only a few tunable parameters, it is simple to implement. Another attractive feature of PSO is that, since it does not require the gradient of the problem being optimized, it is applicable to many not-differentiable problems. This makes it distinctive from classic optimization methods, such as gradient descent algorithm. Moreover, in many cases, PSO tends to converge to the 'best' solution quickly and hence it has the potential to solve many difficult problems with a high efficiency.

Genetic algorithm (GA), on the other hand, is inspired by the process of natural selection and thus Darwin's theory of evolution. It is first introduced by John Holland in 1960. In 1989, his student Goldberg further extended GA into a protocol that is in common use today. GA is a powerful meta-heuristic tool being frequently and widely used to find near-optimum solutions for many combinational problems in the field of business, scientific and engineering. The working principle of GA is rather straightforward: during each successive generation, individual solutions with desired

fitness, as measured by the evaluation function, are retained and advanced to generate offspring, and this step is repeated iteratively till optimized solutions are produced.

With GA being easily implementable, it is also suitable to hybridize with other optimization methods. Other advantages of GA are that it is normally guaranteed to improve the current solution and that it also works for problems with a 'noisy' environment. However, GA has its limitations as well. One major disadvantage is the curse of dimensionality. In other words, GA does not scale well with problem complexity. The problem search space increases exponentially with the dimension of the problem. In addition, for some dynamic problems, GA cannot always produce consistent solutions because early convergence issues can become prominent.

1.3.3 Pros and Cons of Heuristic Approaches

Overall, applying heuristic approaches for problem-solving can be very beneficial and efficient. It tends to provide a quick and relatively inexpensive feedback to us. Especially for the early-phase of a designing process, the intuitive feedback from heuristic technique would be helpful to problem-designers. However, one major shortcoming of using heuristic approaches is that the different heuristic methodologies may perform very differently, and no single approach can be used to solve all kinds of problem instances. Trained experts are usually required to implement heuristic approaches efficiently and effectively, especially in addressing practical scientific questions.

Chapter 2 : Both PSO and GA are Efficient in Minimizing Met-enkephalin

2.1 Introduction

Peptides are short biopolymers, whose components are the 20 unique amino acids in nature. The number of amino acid units within a peptide usually ranges from a few to a few tens. When the amino acid number grows even larger, it usually adopts a more rigid structure and becomes a protein. In biological systems, peptides perform important physiological roles. For example, peptides form one major class of hormones and are essential for human homeostasis regulations; some peptides can function as transporters mediating the trans-membrane processes of nutrients or small molecules; other peptides can function as enzymes to catalyze metabolism reactions. Overall, an appropriate concentration and activity levels of peptides are necessary to achieve body homeostasis and maintain health.

Studying peptides is important for not only understanding the fundamental molecular mechanism of the endogenous biological system, but also presenting tremendous values for pharmaceutical drug development. Especially in recent years, developing peptides as cancer therapy drugs have gained increasing attention, because of their lower production cost (than antibodies) and less toxicity (than small organic compounds). These two factors will be critically evaluated when a potential drug candidate advances into clinical trials. In pre-clinical research, however, emphasis is usually placed onto optimizing the binding affinity of the peptide of interest to its targeting proteins.

Based on the "thermodynamic hypothesis" by Anfinsen in 1973, the native state of protein is the structure that minimizes the free energy [35]. In other words, such a native state corresponds to the structure conformation with the global minimum energy. The immediate question to address is then how to quickly find such a structure with a global minimum energy. Performing the conformational search to find this global minimum energy is an obvious approach and it is called as a conformational optimization problem. Optimizing peptide structures has been a challenging task for many decades because of its large conformational searching space caused by a vast number of dihedral angle combinations for the peptide backbone. And notably, when the amino acid sequence becomes longer, the conformational search space grows exponentially.

Over the past years, many research have been carried out to apply heuristic approaches to facilitate the peptide structure minimization, such as Tabu Search (TS) [36], Simulated Annealing (SA) [37], Monte Carlo Minimization (MCM) [38] and Conformational Space Annealing (CSA) [39]. However, these optimizers can only minimize short (typically less than 10 amino acids) peptide structures because the computational cost of searching becomes too high to handle larger peptides. Meanwhile, restraints were applied to the dihedral angle combinations. Thus, it is important to fill the gap of technologies to perform long peptide conformation searches efficiently and effectively.

In this chapter, the adaptation of Genetic Algorithm, Memetic Algorithm, and Particle Swarm Optimization techniques to minimize a common testing peptide, Met-enkephalin, will be performed. Met-enkephalin, with a primary amino acid sequence YGGFM, is an endogenous pentapeptide that has an opioid effect. Met-enkephalin is

highly unstable and has a low bioavailability and short half-life (minutes). These properties cause difficulties to study this peptide using traditional biochemical techniques. That explains why it has attracted many chemists' attention to study its structure computationally. As a benchmarking peptide, it has been widely used to perform the evaluation of peptide conformation optimizers. For a direct comparison with other existing approaches, we will also use Met-enkephalin to test our algorithm.

2.2 Methodology

2.2.1 Energy Evaluation Function

Many different force fields for defining the potential energy of protein structures have been designed, and the most used ones are ECEPP, OPLS, GROMOS, CHARMM and AMBER. In this study, we choose the AMBER force field that is a preferable choice among computational chemists when performing the molecular mechanic studies [30].

Below is the functional form of the AMBER force field; this equation defines the potential energy of a macromolecule system (peptide in our case).

$$V(r^N) = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

The first two terms represent the bond and angle energies, respectively. The last term is composed of two parts: van der Waals (non-bonded energy between all atom pairs) and electrostatics. The third term, which depends on the torsion angles, will be the only energy to be minimized in this study. The principle/fact behind this is that, under biological conditions, the dihedral angle changes will primarily determine the overall peptide structures; whilst, the bond length and angle terms are relatively constant. Thus, to reduce the size of the problem, bond lengths and bond angles will be fixed at their

equilibrium values; and the changes to van der Waals and electrostatics during the conformation searches were also handled by using the default values in Amber.

2.2.2 Decision Variables

In Section 2.2.1, we mentioned that the dihedral angles are the only parameters we are trying to optimize. Hence, the decision variables are the dihedral angles for each amino acid that forms the peptide sequence. Specifically, they include the backbone dihedral angles: phi (ϕ), psi (φ) and omega (ω), and also the amino acid side chain dihedral angles, chi (χ). Table 2-1 shows the detailed numbers of the decision variables for peptide Met-enkephalin.

Table 2-1. Number of all the dihedral angles within peptide Met-enkephalin

Amino Acid	Psi angle (ϕ)	Psi angle (φ)	Omega angle (ω)	Chi angle (χ)
Tyr (Y)	1	1	1	3
Gly (G)	1	1	1	0
Gly (G)	1	1	1	0
Phe (F)	1	1	1	2
Met (M)	1	1	1	3
Sum of all dihedral angles**: 23**				

2.2.3 Particle Swarm Optimization

Particle Swarm Optimization has been applied to solve many NP-hard problems. We choose PSO in this study because it is simple to implement with only a few parameters to adjust. Furthermore, in many cases, PSO converges to the best solution quickly and hence it can solve some difficult problems efficiently.

As denoted in the name, there are two main components for PSO, the swarm and the particle. All particles together form a swarm and there is a social component to guide them move synergistically. For each individual particle, the movement is influenced by

three factors: inertia, cognitive influence and social influences where inertia velocity of a particle maintains its capability to explore the search space, cognitive influence comes from its personal best history and social influences consider the effect of found global best. Therefore, each individual particle tries to achieve self-improvement through both cognitive and social influences, and eventually the whole swarm moves towards the best area of solution quickly. The two variables being considered for each particle within the swarm are velocity and position, and they are updated at each iterative cycle based on the equations (1) and (2) below, respectively.

$$V_i^{t+1} = V_i^t + \varphi_1 * r_1(P_i - X_i^t) + \varphi_2 * r_2\left(P_g - X_i^t\right) \quad (1)$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad (2)$$

Both parameters, velocity and position, are initialized randomly from the solution space. During each cycle, both individual particle best solution (P_i) and global particle best solution (P_g) are retained, and velocity and position update their values in a sequential order. Briefly, the velocity updates itself by considering all three components (i.e. inertia, cognitive influence and social influences), and afterwards, the position simply updates itself by adding the updated velocity. The iterative cycle is repeated until termination criteria is satisfied.

2.2.4 Genetic Algorithm

Genetic Algorithm is a powerful meta-heuristic tool that has been frequently used to seek near-optimum solutions for many combinational problems. We also test GA as a Met-enkephalin conformation optimizer with the consideration that GA is suitable to hybridize with other existing methods, so that we can easily combine, if necessary, with other implemented heuristic methods in our study.

Overall, GA generates offspring through, firstly, a linear combination of two parents as denoted below in equations (3) and (4), where β is randomized between 0 and 1, and secondly, performing crossover with one parent and thirdly, randomly selecting one point for mutation to a random value within the solution space. The corresponding rate for crossover and mutation being optimized to be 0.8 and 0.1, respectively, in our study.

$$O1=P1 - \beta*(P1-P2) \quad (3)$$

$$O2=P2 + \beta*(P1-P2) \quad (4)$$

Importantly, it is usually not efficient to replace the entire parental generation with all new offspring in the children generation. Hence, we applied the elitism to maintain high fitness for the continuing population by retaining the top 50% individuals from the parental generation and another top 50% individuals from the children generation in terms of their fitness values determined by the evaluation function.

2.2.5 Memetic Algorithm

The Memetic Algorithm, as a hybrid Genetic Algorithm, was developed in the most recent decade and has been successfully applied to solve many real-world problems, such as maintenance scheduling [40, 41], gene expression clustering [42] and gene feature selection [43]. When combined with the local search, the MA performance proves to be an improvement to GA in some problem domains. In this study, we perform a very simple MA with the local neighborhood structure being defined as the value of one dihedral angle ± 0.5 degrees. Thus, the size of the neighborhood is two-fold of the problem dimension.

2.2.6 Amber and Pyrosetta Software

The evaluation function was implemented using the Amber fore field as stated in Section 2.2.1, thanks to the contribution from Amber developers who developed the Python Application Programmer Interface (API). The Python API makes it possible for us to use sander (an Amber module which carries out energy minimization) functionalities inside our own Python scripts without worrying about a) how strings map to the underlying Fortran code and additionally, or b) bugs arising from uninitialized variables (see Amber manual Section 17.13.4 [30]). Specifically, in this study, we employed the sander functionality of AMBER software to compute the total potential energy of our peptide systems.

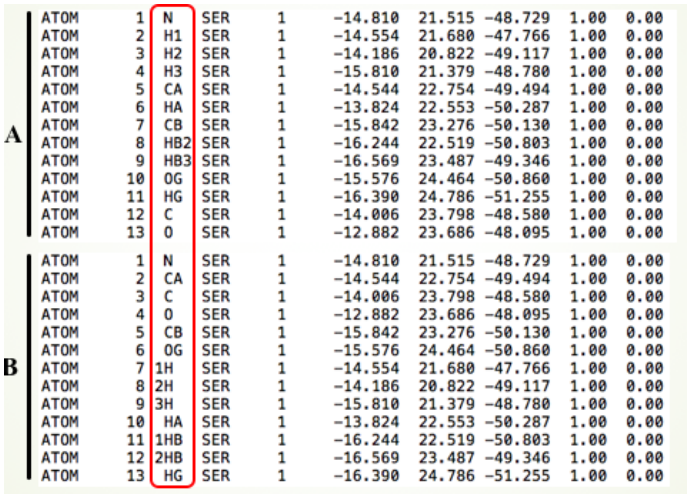
Another important question to tackle is how to generate the new atom coordinates after changing the dihedral angles during each iterative cycle. One useful software package we employed for this question is called Pyrosetta [44], which is an interactive Python-based interface modified based on Rosetta. Pyrosetta enables users to design their own molecular modeling algorithms, and we used the software to regenerate new atomic coordinates from a set of new dihedral angles.

2.3 Results

2.3.1 Conversion between Amber and Sybyl Atom Type

To compute the potential energy of the peptide system, all atomic coordinates need to be provided. In this study, we regenerate the coordinates of all peptide atoms using Pyrosetta software, which can directly take the dihedral angles of the peptide sequence as the input and quickly output the new coordinates. However, a major

challenge is that Pyrosetta code utilizes Sybyl atom type, whist Amber package only accept Amber atom type. The atom type incompatibility causes problematic cooperativity between the two software packages. After a detailed output file comparison, we find that the differences between using these two atom types appear as only different atom orders (Fig. 2-1). Specifically, the order of atom coordinates for each amino acid regenerated from Pyrosetta need to be reshuffled before passing to Amber software for energy calculations. As a solution, we manually find the matching orders for all 20 amino acids denoted using either Amber or Sybyl Atom Type, and then use this mapping to link the two software packages. For researchers who may face similar issues whenever need to connect the Amber and Pyrosetta, the matching orders from this study can be easily used.



A

ATOM	1	N	SER	1	-14.810	21.515	-48.729	1.00	0.00
ATOM	2	H1	SER	1	-14.554	21.680	-47.766	1.00	0.00
ATOM	3	H2	SER	1	-14.186	20.822	-49.117	1.00	0.00
ATOM	4	H3	SER	1	-15.810	21.379	-48.780	1.00	0.00
ATOM	5	CA	SER	1	-14.544	22.754	-49.494	1.00	0.00
ATOM	6	HA	SER	1	-13.824	22.553	-50.287	1.00	0.00
ATOM	7	CB	SER	1	-15.842	23.276	-50.130	1.00	0.00
ATOM	8	HB2	SER	1	-16.244	22.519	-50.803	1.00	0.00
ATOM	9	HB3	SER	1	-16.569	23.487	-49.346	1.00	0.00
ATOM	10	OG	SER	1	-15.576	24.464	-50.860	1.00	0.00
ATOM	11	HG	SER	1	-16.390	24.786	-51.255	1.00	0.00
ATOM	12	C	SER	1	-14.006	23.798	-48.580	1.00	0.00
ATOM	13	O	SER	1	-12.882	23.686	-48.095	1.00	0.00

B

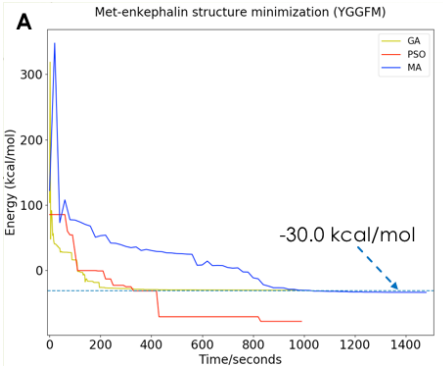
ATOM	1	N	SER	1	-14.810	21.515	-48.729	1.00	0.00
ATOM	2	CA	SER	1	-14.544	22.754	-49.494	1.00	0.00
ATOM	3	C	SER	1	-14.006	23.798	-48.580	1.00	0.00
ATOM	4	O	SER	1	-12.882	23.686	-48.095	1.00	0.00
ATOM	5	CB	SER	1	-15.842	23.276	-50.130	1.00	0.00
ATOM	6	OG	SER	1	-15.576	24.464	-50.860	1.00	0.00
ATOM	7	1H	SER	1	-14.554	21.680	-47.766	1.00	0.00
ATOM	8	2H	SER	1	-14.186	20.822	-49.117	1.00	0.00
ATOM	9	3H	SER	1	-15.810	21.379	-48.780	1.00	0.00
ATOM	10	HA	SER	1	-13.824	22.553	-50.287	1.00	0.00
ATOM	11	1HB	SER	1	-16.244	22.519	-50.803	1.00	0.00
ATOM	12	2HB	SER	1	-16.569	23.487	-49.346	1.00	0.00
ATOM	13	HG	SER	1	-16.390	24.786	-51.255	1.00	0.00

Figure 2-1. Example of atom order difference between Amber atom type and Sybyl atom type using amino acid Serine.
A) Amber atom type expression of amino acid Serine (SER); B) Sybyl atom type expression of amino acid Serine (SER). The atom symbol and order differences are highlighted with a red box.

2.3.2 PSO, GA and MA all Efficiently Minimize Met-enkephalin

Three heuristic approaches (i.e. PSO, GA and MA) are implemented in this study to test and compare the algorithm minimization efficiency on Met-enkephalin. All three algorithms are coded in Python and their corresponding experimental designs can be

found in detail in Section 2.2. As shown in the left panel of Figure 2-2, using the 23 dihedral angles within Met-enkephalin as the decision variables (refer to Table 2-1), all three heuristic approaches (i.e. PSO, GA and MA) can obtain minimized Met-enkephalin three-dimensional structures with the system total energy below -30.0 kcal/mol (Fig. 2-2A). Although both of GA and MA converge to minimized conformations with quite close total energies, GA performs 2-fold faster than MA. A possible explanation is that the MA method needs to spend a non-negligible amount of time to perform the local search for each individual before moving to the next iterative generation, which causes its slower convergence compared to GA without any local searches. Of note, the PSO algorithm stands out strikingly in regard to its efficiency in finding a further lower-energy conformation with the least computational effort among the three methods (Fig. 2-2B).



B

	Algorithm	Energy (kcal/mol)	Cost
This work	GA	-30.34	525s
	MA	-33.71	1087s
	PSO	-78.05	418s
Literature	MCM	-28.44	2-3 hr
	CSA	-31.48	0.75 hr
	αBB	-30.81	1.3 hr
	…		

Figure 2-2. Efficient structure minimization of Met-enkephalin using PSO, GA and MA heuristic methods.
A) Plot of the energy decrease during Met-encephalin structure minimization using PSO, GA and MA; B) Comparison of minimization efficiency between our algorithms and other existing global optimizers from literature, including MCM, CSA and **α**BB.

To further evaluate the exciting results using our methods, all three algorithms are compared with some existing approaches found in the literature, such as Monte Carlo Minimization (MCM) [38], Conformational Space Annealing (CSA) [39] and **α**-Branch

and Bound (αBB) [45] (Fig. 2-2B). The results demonstrate that all three approaches implemented herein are comparable, and in particular, our PSO method significantly outperforms all other methods. Overall, the study of these three algorithms on a testing peptide, Met-enkephalin, lays the foundation to study another novel and longer peptide, namely TPP-1. Given these results, we advance PSO and GA methods for the TPP-1 structure minimization considering their efficient optimization performance.

2.3.3 Minimized Met-enkephalin Structure by PSO

From the results in Section 2.3.2, PSO minimizes the Met-enkephalin structure to a much lower energy than all other applied techniques. Here, we use a molecule graphic software, Chimera [46], to represent the minimized lowest-energy Met-enkephalin structure (Fig. 2-3, colored in tan). And through comparison with the Met-enkephalin structure before running any minimization (Fig. 2-3, colored in cyan), we find that the two middle flexible residues, Glycine (G), exist as the overall structural turning points after minimization. In addition, the two peptide terminal ends, i.e. the amino and carboxylic acid groups, are found in close proximity, indicating electrostatics play a role in attracting these two groups after minimization. Furthermore, the optimized structure is considered feasible after being submitted to the online software, PROCHECK [47] for the stereo-chemical quality evaluation.

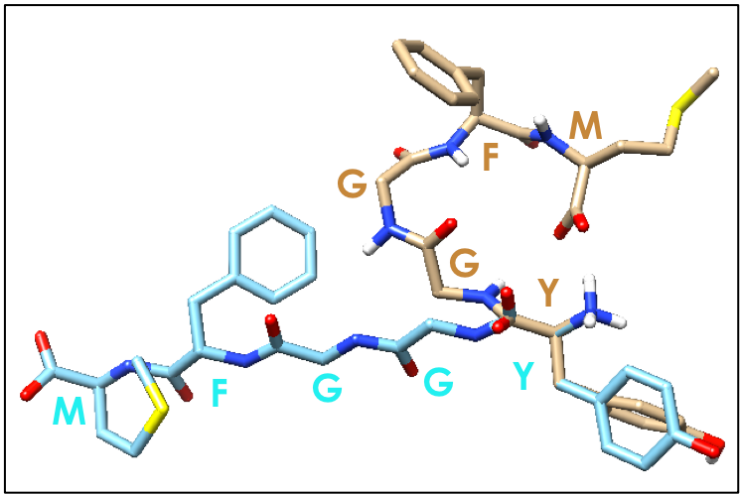


Figure 2-3. Ribbon representation of Met-enkephalin before and after minimization. The ribbon structure colored in Cyan is the linearized Met-enkephalin, i.e. before minimization; The ribbon structure colored in Tan is the Met-enkephalin after minimization. Corresponding amino acid residues are labeled adjacent to the ribbon structures, Y: Tyrosine, G: Glycine, F: Phenylalanine, M: Methionine.

2.4 Discussion

In the past decades, the Met-enkephalin has attracted considerable interest in the area of developing more efficient global optimizers for protein or peptide. In this chapter, we choose Met-enkephalin to evaluate our algorithms for the purpose of convenient comparison with other existing methods.

To exclude the discrepancies caused by the continuingly updated force fields, we re-calculated the results from literature using the same Amber force field that was used in our study. Another interesting point is the mismatching atom orders expressed by Amber and Pyrosetta. Although manually assigned atom order correction was used in our study to solve this problem, it would be more efficient to develop a script so that the atom type conversions can be done automatically within either Amber or Pyrosetta software.

Based on our experimental data, we found that all three algorithms, i.e. PSO, GA and MA, can efficiently minimize the structure of Met-enkephalin to a comparable low-

energy conformation with other existing methods. Of note, the PSO approach minimizes the Met-enkephalin structure very efficiently, outperforming all other methods as shown in Figure 2-2. Noticeably, when we compare the performance of MA and GA, MA, with the local search, found a slightly lower energy conformation than GA, however, the computational cost of MA becomes double of GA. Seemingly, this phenomenon is caused by an inefficient local structure refinements. In fact, we have tried multiple local neighborhood structures, and none of them can efficiently decrease the CPU cost in comparison with GA. In my opinion, a reasonable explanation is that the domain of dihedral angles is continuous, and local neighborhood structure works more efficient for discrete variables than continuous values. With the optimization results from Met-enkephalin, we decide to advance both PSO and GA for the TPP-1 structure minimization in the Chapter 3.

Chapter 3 : Building a Low-energy Pool of TPP-1 Conformers Using Hybrid PSO and GA

3.1 Introduction

Cancer immunotherapy was proposed as far as 100 years ago, but it is only revived until recent successes in the clinical studies that target the immune-checkpoint with antibodies [48]. The protein-protein interaction between PD-1, programmed death-1 protein, and its ligand PD-L1, programmed death-1 protein ligand, has become the focal target to develop new immune checkpoint antagonists in both clinical and pre-clinical trials. Although using antibodies as the immune-checkpoint antagonists have achieved effective results in clinical studies, before advancing further towards therapeutic drugs, some disadvantages of antibodies still give rise to concerns among researchers, such as low oral bioavailability, very long half-life time, low penetration rate and more importantly, difficult and expensive production. Taken all the above factors into consideration, developing small molecule or peptide antagonists for targeting immune checkpoints has become the alternative resort in immunotherapy.

Until recently, the first Å-resolution crystal structure of human PD-1/PD-L1 has been solved [49], which allows researchers to analyze the molecular interactions between these two proteins more in detail. Specifically, a large flat and hydrophobic interface was found between PD-1 and PD-L1. This makes it very difficult to develop small organic molecule ligands due to either low binding specificity (when enough hydrophilic portion was maintained to make them soluble) or low water solubility (when hydrophobic moieties were largely included to improve the binding affinity) [50]. Therefore, small

organic molecules do not appear to be very promising for this particular target, and developing suitable peptide or peptide derivatives to target the immune checkpoint PD-1/PD-L1 becomes more attractive.

TPP-1, a 22-mer peptide, was discovered from biological high-throughput screening assay and shown to be effective in inhibiting cancer cell proliferation and could potentially improve current tumor therapy [14]. TPP-1, which works as a PD-L1 blockade agent, binds PD-L1 and disrupts the PD-1/PD-L1 interaction. Since TPP-1 contains only naturally occurring amino acids, it is projected to have a higher bioavailability and stronger binding selectivity properties than organic molecules.

At current pre-clinical stage, TPP-1 optimization is critically needed in order to design more potent and effective peptide candidates before advancing to furthermore comprehensive clinical investigations. Serving as an important component towards this goal, building the low-energy pool of TPP-1 structures would be beneficial and expedite the process of TPP-1 drug development.

Most current peptide global optimizers, such as aforementioned MCM, CSA, TS or SA in Chapter 2, cannot handle large (>10aa) peptides structural minimization efficiently. Given that dihedral angles determine primarily the overall peptide structures under biological conditions, in this project, we aim to efficiently collect the low-energy conformers of TPP-1 by evolving the combination of amino acid dihedral angles in the backbone, i.e. phi (ϕ) and psi (φ), and subsequently, running a brief energy minimization to adjust any stereo-inappropriate local structures or remove side chain clashes.

With the advent of more and more complicated problems, single heuristic approach cannot solve problem as efficiently as required, alternatively, some researchers

have hybridized different heuristic approaches for solving complicated problems by taking each of their advantages [51-54]. Inspired by some of these work [52, 54], we are also interested in implementing the hybridized PSO and GA for TPP-1 conformational search.

3.2 Methodology

3.2.1 Energy Evaluation Function

The same evaluation function from Section 2.2.1 is used to implement energy calculations of TPP-1.

3.2.2 Decision Variables and Search Space

Here we define the decision variables for TPP-1 similar to the decision variables in Section 2.2.2 for Met-enkephalin. TPP-1 sequence includes 22 amino acids, i.e. SGQYASYHCWCWRNPGRSGGSK using one letter representation. However, due to the search space grows exponentially as the size of the decision variable linearly increases. Therefore, to simplify the problem, we only take into the consideration of the conformation-dominating dihedral angles, specifically, the backbone dihedral angles: phi (ϕ) and psi (φ). And the peptide plane dihedral angle omega (ω), and amino acid side chain dihedral angles chi (χ) were not included. For excluded dihedral angles, their values were taken by default from the software Pyrosetta. Table 3-1 below shows the detailed numbers of the decision variables for TPP-1 peptide.

Table 3-1. Number of dihedral angles within peptide TPP-1.

Amino Acid	Psi angle (ϕ)	Psi	Omega angle (ω)	Chi angle (χ)
Ser (S)	1	1	1	1
Gly (G)	1	1	1	0
Gln (Q)	1	1	1	3

Tyr (Y)	1	1	1	3
Ala (A)	1	1	1	0
Ser (S)	1	1	1	1
Tyr (Y)	1	1	1	3
His (H)	1	1	1	2
Cys (C)	1	1	1	1
Trp (W)	1	1	1	2
Cys (C)	1	1	1	1
Trp (W)	1	1	1	2
Arg (R)	1	1	1	5
Asp (D)	1	1	1	2
Pro (P)	1	1	1	2
Gly (G)	1	1	1	0
Arg (R)	1	1	1	5
Ser (S)	1	1	1	1
Gly (G)	1	1	1	0
Gly (G)	1	1	1	0
Ser (S)	1	1	1	1
Lys (K)	1	1	1	4
Sum of all backbone dihedral angles**: 44**				

Note: The values of dihedral angle ω and χ are taken from Pyrosetta by default.

3.2.3 Hybrid PSO-GA

Both Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) methods have been specified in detail in Section 2.2. In this chapter, we implement the algorithms similarly for each method *per se*.

Hybrid PSO and GA is introduced to this work by performing PSO first and then GA. Specifically, one-hundred particles were initialized randomly from the variable domain, i.e. from -180° to 180°, for PSO minimization. Taking the quick convergence of PSO into a consideration, we output and save the best population of particles after the last iterative cycle. Subsequently, fifty chromosomes of GA are randomly selected from the aforementioned one-hundred PSO particles instead of being randomly initialized again from the whole variable domain. This allows GA to start from lower-energy conformations, avoiding a large number of unnecessary searches within poor quality

solution space, and thus improving the GA search efficiency. The optimized parameters of PSO and GA from Section 2.2 are directly inherited in this chapter.

3.2.4 Molecular Dynamic Simulation

The linearized TPP-1 structure, generated from Pyrosetta, is used as the initial structure to run molecular dynamics (MD). Specifically, the peptide was solvated in 105 cubic Å water box, and a 50 nanoseconds (ns) MD simulation is implemented using software NAMD on OU Supercomputing Center for Education and Research (OSCER) with GPU acceleration. Amber force field is used to model peptide. The MD simulation system is equilibrated at 300K for 2 ns. Periodic boundary conditions are selected, and long-range electrostatic interactions are calculated with particle mesh Ewald method, with non-bonded cutoff set to 12.0 Å and SHAKE algorithm is used to constrain bonds involving hydrogen atoms. Time step is 2 femtoseconds (fs) and the trajectories are recorded every 10 picoseconds (ps).

3.3 Results

3.3.1 Hybrid PSO-GA Minimizes TPP-1 More Efficiently Than PSO or GA

In this experiment, we implement both PSO and GA in TPP-1 structure minimization with 44 peptide backbone dihedral angles, i.e. the phi (ϕ) and psi (φ) angles, being used as the decision variables. As the results show, neither PSO nor GA itself can minimize TPP-1 efficiently. Note that when the PSO implementation converges, no further improvement is obtained with more iteration cycles (Fig. 3-1A). In contrast, when the PSO and GA algorithms are hybridized following the procedures specified in Section 3.2.3, we find a significant increase of the structure minimization efficiency (Fig. 3-1B). Hybrid PSO-GA further minimizes the TPP-1 structure to a conformation at around -100

kcal/mol from the conformation minimized by PSO at around 300 kcal/mol. This hybridization also outperforms GA, which only minimizes TPP-1 to a conformation at about -9 kcal/mol. Overall, the performance of hybrid PSO-GA is significantly more efficient than using single heuristic method.

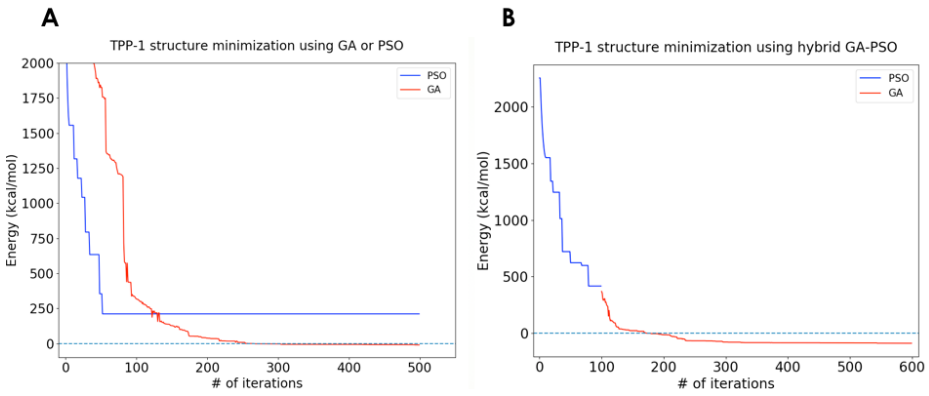


Figure 3-1.Comparison of the minimization efficiency between GA, PSO and hybrid PSO-GA.
A) The TPP-1 structure is minimized with either PSO algorithm (blue line) or GA (red line); B) The TPP-1 structure is minimized with a hybrid algorithm, i.e. GA (red line) is subsequently performed after PSO (blue line) minimization.

3.3.2 A Pool of Lower-energy TPP-1 Conformers

Following the same approach as used in Section 3.3.1, we collect 30 low-energy conformations of TPP-1 by randomly selecting 50 particles from the PSO minimization and advancing them to the GA step as the initial chromosomes. After computing the energies of all 30 TPP-1 conformers, we arbitrarily choose the top ten lowest-energy conformations to build the low-energy TPP-1 pool. Interestingly, we find that all ten low-energy conformers have unique structural conformations (Fig 3-2A), indicating the conformation flexibility property of peptide folding. Furthermore, Ramachandran plot ([55]) of the phi-psi (ϕ-φ) torsion angles are used to validate the stereo-chemical quality and feasibilities of all TPP-1 conformations minimized by hybrid PSO-GA. A Ramachandran plot built from 500 non-homologous proteins is used as a dihedral angle

feasibility reference [56] (shown in Figure 3-2B). And all our 30 minimized TPP-1 structures are used to generate a similar plot to compare (Figure 3-2C). Through this comparison, we find that a majority of the structures adapted β-strand as its main secondary structure with some of them also maintaining α-helix structures locally. See Figure 3-2A for a depiction of these results. Quantitatively, above 90% of the dihedral angles fell into the favorable or additional allowed regions in the Ramachandran plot, given by the online software PROCHECK [47].

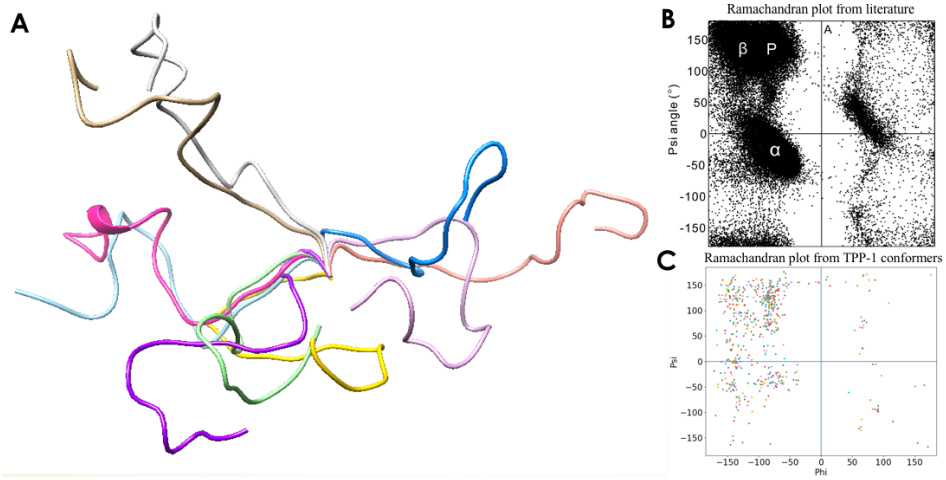


Figure 3-2. A Diversified low-energy conformation pool of TPP-1.

A) Ribbon representation of top 10 TPP-1 low-energy conformers, prepared using software UCSF Chimera; B) Ramachandran plot from published literature [56]; C) The Ramachandran plot for TPP-1 low-energy conformers from this study.

3.3.3 Comparable Results from Hybrid PSO-GA and Molecular Dynamic Simulation

Table 3-2. Comparison between hybrid PSO-GA and Molecular Dynamic simulation.

MD simulation length	Cost (GPU 20 cores)	Minimization energy*
10ns	~20h	-517.41
50ns	~75h	-563.67
PSO-GA (conformer index)	**Cost (CPU 1 core)**	**Minimization energy***
7	2-3h	-543.61
10	2-3h	-494.94
11	2-3h	-485.44
18	2-3h	-540.99
20	2-3h	-542.93
22	2-3h	-539.43
26	2-3h	-496.14
27	2-3h	-509.95
29	2-3h	-482.29
30	2-3h	-542.95

Note: * done by 10,000 steps steepest descent algorithm, unit: kcal/mol.

To rationally evaluate the efficiency of our PSO-GA algorithm, conventional theoretical calculation, i.e. Molecular Dynamics (MD) simulation under the force field, was performed to minimize TPP-1 structure from a linearized configuration. As the results in Table 3-2 show, after a 10 ns and 50 ns MD simulation, the TPP-1 structure was minimized to conformations at energy -517.41 and -563.67 kcal/mol, respectively. As a comparison, we find that the TPP-1 conformers minimized using our hybrid PSO-GA approach are comparable with the MD simulation results (Table 3-2). In terms of the

computational efficiency, a 10 ns MD simulation requires ~20 hours computational cost with GPU acceleration (20 cores) on the University of Oklahoma Supercomputing Center (OSCER). In contrast, our hybrid PSO-GA requires only 2-3 hours on only one CPU core and achieved even lower energy conformers. When considering the fact that normally GPU accelerates calculations 6~7 times (over a 20-core CPU node), our PSO-GA algorithm improves the conformational search efficiency by at least two orders of magnitude. Moreover, it is encouraging that the hybrid PSO-GA also obtained stereo-chemically reasonably structures. Therefore, using the established hybrid PSO-GA algorithm, we can quickly generate a pool of low-energy conformers for peptides, and this low-energy structure pool can be potentially used as valuable initial structures for subsequent peptide molecular docking and peptide-protein binding free energy calculations.

3.4 Discussion

Peptide conformational search has been a challenging scientific task for many decades and due to its conformation flexibility, finding only the global minimum structure may not suit some research needs.

In this study, in order to meet an increasing demand for peptide drug development in the field of cancer immunotherapy recently, we choose a promising PD-L1 inhibiting peptide, TPP-1, as our studying target. We apply a hybridized PSO and GA approach to minimize the TPP-1 structure. More importantly, our goal is to build a low-energy structure pool of TPP-1 other than finding only the global minimum structure.

As our results show in Figure 3-1, hybrid PSO-GA is more efficient in searching for minimized TPP-1 structures than either PSO or GA itself. And notably, multiple TPP-

1 low-energy structures can be collected within only couple days using one single laptop (Table 3-2). Overall, these minimized structures are also sterically feasible as shown on the Ramachandran plot (Fig. 3-2C). When compared with the conventional Molecular Dynamics simulation approach, our hybrid PSO-GA stands out significantly as it increases TPP-1 structure minimization efficiency by more than two orders of magnitude (Table 3-2).

In our study, no constraints were applied on the decision variables, meaning all the chosen dihedral angles were allowed to continuously change from -180° to 180°. However, it should be easy to adapt this algorithm for one who works with discrete variables. Previously, multiple studies applied heuristic methods to minimize peptide structures and they all applied restraints onto the dihedral angles in order to save computational cost [39, 45]. Therefore, our hybrid PSO-GA has higher chances of finding more low-energy conformations.

We conclude that the hybrid PSO-GA is well suited for conformational searches of peptides. Especially, it could be a very important study to conduct for peptide-involved structure optimization or binding affinity investigations. It is obvious that our hybrid PSO-GA can be further improved if one can spend more time on the parameter tuning and optimization. In addition, working on different peptides, some of the parameters may need re-adjustment, such as the population size of GA, the inertia constant and also the constants for cognitive or social components of PSO.

Chapter 4 : Overall Summary and Future Directions

How to efficiently search the low-energy conformation, especially for biological macromolecules remains a difficult question. This work aims to improve current peptide conformational search methods using meta-heuristic approaches. In particular, our study focused on improving peptide conformational searching in order to benefit researchers either studying peptide-protein interactions or developing peptide-involved drug design.

Instead of just searching the global minimum structure, we built a low-energy conformer pool for a promising PD-L1 inhibiting polypeptide, TPP-1. Using the peptide backbone dihedral angles as the decision variables, both PSO and GA can outperform other existing approaches in minimizing the Met-enkephalin, a benchmarking pentapeptide for judging the efficiency of conformation optimizers. However, neither PSO and GA performed well on minimizing TPP-1, instead, we found that hybridizing PSO and GA can minimize TPP-1 structure efficiently. Strikingly, the efficiency of hybrid PSO-GA is hundreds-fold faster than the conventional Molecular Dynamic simulations running under the force filed. Meanwhile, the stereo-chemical quality of the minimized structures was also validated using Ramachandran plot. Overall, our hybrid PSO-GA minimization approach can benefit biochemical and biomedical researches with the demand of determining polypeptide conformations and hence can advance the field of studying peptide-involved anti-cancer drugs.

For the future directions, another important question need to address is to obtain the trajectory cluster analysis result after peptide MD simulations. In theory, the overall structure moves towards low-energy conformations under the force field, but it may take

a long time for the simulation to escape certain local minima along the simulation path due to the reason that these local minima are already quite stable conformations with very low energies. Although the simulated structure at the simulation endpoint typically has a very low energy, it is still important to know all the low energy conformations within the entire MD simulation trajectory. Such a way, the comparison between minimized structures from hybrid PSO-GA and low-energy structures from the MD cluster analysis can give us more confidence on determining the robustness of the PSO-GA optimizer. Given more time on this thesis project, I would like to perform the trajectory cluster analysis using Amber Package. And furthermore, for structural similarity comparisons, the 'match maker' function in software UCSF Chimera can be utilized to align structure to structure.

In peptide structure optimization studies, there are typically two main cases: either seeking more potent peptide interactions to achieve effective results or reducing some peptide interactions to an appropriate range to maintain its interaction reversibility. In either scenario, besides natural amino acids, some unnatural amino acids, which only differ in the side chains from 20 natural amino acids, may be introduced into the peptide sequence to help tune the peptide-protein binding affinity. And here, we propose that our method can work efficiently in a similar manner to meet either goal because we only used the backbone dihedral angles as the decision variables in our hybrid PSO-GA. But the side chain dihedral angles or clashes will be taken care of by a short (~s) steepest minimization step at the end. However, we expect this method to work well for peptide derivatives, which involves small organic moieties, although extensive testing will be

needed. But this could be another interesting future project to expand the applicability of our methods onto different types of biomolecules optimizations.

Overall, as a proof of concept, this study demonstrates that hybridized meta-heuristic strategies can be more efficiently applied to expedite peptide conformational searching. It significantly accelerates short peptide conformation searching by at least two orders of magnitude (while reaching similar quantity of results) than conventional MD simulations.

References

1. Giannini, L., F. Incandela, M. Fiore, A. Gronchi, S. Stacchiotti, C. Sangalli, and C. Piazza, *Radiation-Induced Sarcoma of the Head and Neck: A Review of the Literature.* Front. Oncol., 2018. **8**: p. 449.

2. Shinoto, M., K. Terashima, H. Suefuji, A. Matsunobu, S. Toyama, K. Fukunishi, and Y. Shioyama, *A single institutional experience of combined carbon-ion radiotherapy and chemotherapy for unresectable locally advanced pancreatic cancer.* Radiother. Oncol., 2018.

3. de Gooijer, C.J., P. Baas, and J.A. Burgers, *Current chemotherapy strategies in malignant pleural mesothelioma.* Transl Lung Cancer Res, 2018. **7**(5): p. 574-583.

4. Chin, V., A. Nagrial, K. Sjoquist, C.A. O'Connor, L. Chantrill, A.V. Biankin, R.J. Scholten, and D. Yip, *Chemotherapy and radiotherapy for advanced pancreatic cancer.* Cochrane Database Syst. Rev., 2018. **3**: p. CD011044.

5. Yi, J.S., N. Ready, P. Healy, C. Dumbauld, R. Osborne, M. Berry, D. Shoemaker, J. Clarke, J. Crawford, B. Tong, D. Harpole, T.A. D'Amico, F. McSherry, F. Dunphy, S.J. McCall, J.D. Christensen, X. Wang, and K.J. Weinhold, *Immune Activation in Early-Stage Non-Small Cell Lung Cancer Patients Receiving Neoadjuvant Chemotherapy Plus Ipilimumab.* Clin. Cancer. Res., 2017. **23**(24): p. 7474-7482.

6. Schachter, J., A. Ribas, G.V. Long, A. Arance, J.J. Grob, L. Mortier, A. Daud, M.S. Carlino, C. McNeil, M. Lotem, J. Larkin, P. Lorigan, B. Neyns, C. Blank, T.M. Petrella, O. Hamid, H. Zhou, S. Ebbinghaus, N. Ibrahim, and C. Robert, *Pembrolizumab versus ipilimumab for advanced melanoma: final overall survival results of a multicentre, randomised, open-label phase 3 study (KEYNOTE-006).* Lancet, 2017. **390**(10105): p. 1853-1862.

7. Robert, C., J. Schachter, G.V. Long, A. Arance, J.J. Grob, L. Mortier, A. Daud, M.S. Carlino, C. McNeil, M. Lotem, J. Larkin, P. Lorigan, B. Neyns, C.U. Blank, O. Hamid, C. Mateus, R. Shapira-Frommer, M. Kosh, H. Zhou, N. Ibrahim, S. Ebbinghaus, A. Ribas, and K.-. investigators, *Pembrolizumab versus Ipilimumab in Advanced Melanoma.* N. Engl. J. Med., 2015. **372**(26): p. 2521-32.

8. Long, M., K. Beckwith, P. Do, B.L. Mundy, A. Gordon, A.M. Lehman, K.J. Maddocks, C. Cheney, J.A. Jones, J.M. Flynn, L.A. Andritsos, F. Awan, J.A.

Fraietta, C.H. June, M.V. Maus, J.A. Woyach, M.A. Caligiuri, A.J. Johnson, N. Muthusamy, and J.C. Byrd, *Ibrutinib treatment improves T cell number and function in CLL patients.* J. Clin. Invest., 2017. **127**(8): p. 3052-3064.

9. Lee, A., S. Sun, A. Sandler, and T. Hoang, *Recent progress in therapeutic antibodies for cancer immunotherapy.* Curr. Opin. Chem. Biol., 2018. **44**: p. 56-65.

10. Ledford, H., H. Else, and M. Warren, *Cancer immunologists scoop medicine Nobel prize.* Nature, 2018. **562**(7725): p. 20-21.

11. Webpage, a.; Available from: https://conquer-magazine.com/issues/special-issues/cancer-immunotherapy-june-2017/23-cancer-immunotherapy-drugs-approved-by-the-fda-since-2015.

12. Chang, H.N., B.Y. Liu, Y.K. Qi, Y. Zhou, Y.P. Chen, K.M. Pan, W.W. Li, X.M. Zhou, W.W. Ma, C.Y. Fu, Y.M. Qi, L. Liu, and Y.F. Gao, *Blocking of the PD-1/PD-L1 Interaction by a D-Peptide Antagonist for Cancer Immunotherapy.* Angew. Chem. Int. Ed. Engl., 2015. **54**(40): p. 11760-4.

13. Webpage, b.; Available from: https://www.marketsandmarkets.com/Market-Reports/immunotherapy-drug-market-137717755.html.

14. Li, C., N. Zhang, J. Zhou, C. Ding, Y. Jin, X. Cui, K. Pu, and Y. Zhu, *Peptide Blocking of PD-1/PD-L1 Interaction for Cancer Immunotherapy.* Cancer Immunol Res, 2018. **6**(2): p. 178-188.

15. Plonka, P.M., *Hair pigmentation disorders or 50 years of German-Polish alliance for study on a severe side effect of chemotherapy: Kostanecki's legacy.* Exp. Dermatol., 2015. **24**(1): p. 10-1.

16. Markman, M., *Chemotherapy-associated neurotoxicity: an important side effect-impacting on quality, rather than quantity, of life.* J. Cancer Res. Clin. Oncol., 1996. **122**(9): p. 511-2.

17. Cronin, P.A., E. Myers, H.P. Redmond, S. O'Reilly, and W.O. Kirwan, *Lipomatosis: an unusual side-effect of cytotoxic chemotherapy?* Acta Derm. Venereol., 2010. **90**(3): p. 303-4.

18. Breidenbach, M., D.T. Rein, T. Schondorf, T. Schmidt, E. Konig, M. Valter, and C.M. Kurbacher, *Hematological side-effect profiles of individualized chemotherapy regimen for recurrent ovarian cancer.* Anticancer Drugs, 2003. **14**(5): p. 341-6.

19. Zhang, X., Z. Zhang, X. Xu, Y. Li, Y. Li, Y. Jian, and Z. Gu, *Bioinspired therapeutic dendrimers as efficient peptide drugs based on supramolecular interactions for tumor inhibition.* Angew. Chem. Int. Ed. Engl., 2015. **54**(14): p. 4289-94.

20. Chang, D.K., C.T. Lin, C.H. Wu, and H.C. Wu, *A novel peptide enhances therapeutic efficacy of liposomal anti-cancer drugs in mice models of human lung cancer.* PloS one, 2009. **4**(1): p. e4171.

21. Lai, W.Y., B.T. Huang, J.W. Wang, P.Y. Lin, and P.C. Yang, *A Novel PD-L1-targeting Antagonistic DNA Aptamer With Antitumor Effects.* Mol Ther Nucleic Acids, 2016. **5**(12): p. e397.

22. Tversky, A. and D. Kahneman, *Citation Classic - Judgment under Uncertainty - Heuristics and Biases.* Cc/Soc Behav Sci, 1983(14): p. 22-22.

23. Xu, Z., Y. Zou, and X. Kong, *Meta-heuristic algorithms for parallel identical machines scheduling problem with weighted late work criterion and common due date.* Springerplus, 2015. **4**: p. 782.

24. Marques, I., M.E. Captivo, and M. Vaz Pato, *A bicriteria heuristic for an elective surgery scheduling problem.* Health Care Manag Sci, 2015. **18**(3): p. 251-66.

25. Lei, D., M. Li, and L. Wang, *A Two-Phase Meta-Heuristic for Multiobjective Flexible Job Shop Scheduling Problem With Total Energy Consumption Threshold.* IEEE Trans Cybern, 2018.

26. Raidl, G.R. and J. Gottlieb, *Empirical analysis of locality, heritability and heuristic bias in evolutionary algorithms: a case study for the multidimensional knapsack problem.* Evol. Comput., 2005. **13**(4): p. 441-75.

27. Drake, J.H., E. Ozcan, and E.K. Burke, *A Case Study of Controlling Crossover in a Selection Hyper-heuristic Framework Using the Multidimensional Knapsack Problem.* Evol. Comput., 2016. **24**(1): p. 113-41.

28. Ahmed, Z.H., *The ordered clustered travelling salesman problem: a hybrid genetic algorithm.* ScientificWorldJournal, 2014. **2014**: p. 258207.

29. Anaya Fuentes, G.E., E.S. Hernandez Gress, J.C. Seck Tuoh Mora, and J. Medina Marin, *Solution to travelling salesman problem by clusters and a modified multi-restart iterated local search metaheuristic.* PloS one, 2018. **13**(8): p. e0201868.

30. D.A. Case, I.Y.B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman., *Amber Manual 2018*. 2018.

31. Mignon, D. and T. Simonson, *Comparing three stochastic search algorithms for computational protein design: Monte Carlo, replica exchange Monte Carlo, and a multistart, steepest-descent heuristic.* J. Comput. Chem., 2016. **37**(19): p. 1781-1793.

32. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J. Comput. Chem., 2010. **31**(2): p. 455-61.

33. Handoko, S.D., X. Ouyang, C.T. Su, C.K. Kwoh, and Y.S. Ong, *QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization.* IEEE/ACM Trans. Comput. Biol. Bioinform., 2012. **9**(5): p. 1266-72.

34. Bucak, I.O. and V. Uslan, *An analysis of sequence alignment: heuristic algorithms.* Conf. Proc. IEEE Eng. Med. Biol. Soc., 2010. **2010**: p. 1824-7.

35. Anfinsen, C.B., *Principles that govern the folding of protein chains.* Science, 1973. **181**(4096): p. 223-30.

36. Morales, L.B., R. Garduno-Juarez, J.M. Aguilar-Alvarado, and F.J. Riveros-Castro, *A parallel tabu search for conformational energy optimization of oligopeptides.* J. Comput. Chem., 2000. **21**(2): p. 147-156.

37. Morales, L.B., R. Garduno-Juarez, and D. Romero, *Applications of simulated annealing to the multiple-minima problem in small peptides.* J. Biomol. Struct. Dyn., 1991. **8**(4): p. 721-35.

38. Nayeem, A., J. Vila, and H.A. Scheraga, *A Comparative-Study of the Simulated-Annealing and Monte Carlo-with-Minimization Approaches to the Minimum-Energy Structures of Polypeptides - [Met]-Enkephalin.* J. Comput. Chem., 1991. **12**(5): p. 594-605.

39. Lee, J., H.A. Scheraga, and S. Rackovsky, *New optimization method for conformational energy calculations on polypeptides: Conformational space annealing.* J. Comput. Chem., 1997. **18**(9): p. 1222-1232.

40. Xu, J.Y., Y.Q. Yin, T.C.E. Cheng, C.C. Wu, and S.S. Gu, *A memetic algorithm for the re-entrant permutation flowshop scheduling problem to minimize the makespan.* Appl Soft Comput, 2014. **24**: p. 277-283.

41. Jin, L.L., C.Y. Zhang, X.Y. Shao, and G.D. Tian, *Mathematical modeling and a memetic algorithm for the integration of process planning and scheduling considering uncertain processing times.* P I Mech Eng B-J Eng, 2016. **230**(7): p. 1272-1283.

42. Merz, P., *Analysis of gene expression profiles: an application of memetic algorithms to the minimum sum-of-squares clustering problem.* BioSyst., 2003. **72**(1-2): p. 99-109.

43. Zhu, Z.X., Y.S. Ong, and M. Dash, *Markov blanket-embedded genetic algorithm for gene selection.* Pattern Recogn, 2007. **40**(11): p. 3236-3248.

44. Chaudhury, S., S. Lyskov, and J.J. Gray, *PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta.* Bioinformatics, 2010. **26**(5): p. 689-691.

45. Androulakis, I.P., C.D. Maranas, and C.A. Floudas, *alpha BB: A global optimization method for general constrained nonconvex problems.* J Global Optim, 1995. **7**(4): p. 337-363.

46. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera--a visualization system for exploratory research and analysis.* J. Comput. Chem., 2004. **25**(13): p. 1605-12.

47. Laskowski, R.A., M.W. Macarthur, D.S. Moss, and J.M. Thornton, *Procheck - a Program to Check the Stereochemical Quality of Protein Structures.* J. Appl. Crystallogr., 1993. **26**: p. 283-291.

48. Domling, A. and T.A. Holak, *Programmed Death-1: Therapeutic Success after More than 100 Years of Cancer Immunotherapy.* Angew Chem Int Edit, 2014. **53**(9): p. 2286-2288.

49. Zak, K.M., R. Kitel, S. Przetocka, P. Golik, K. Guzik, B. Musielak, A. Domling, G. Dubin, and T.A. Holak, *Structure of the Complex of Human Programmed Death 1, PD-1, and Its Ligand PD-L1.* Structure, 2015. **23**(12): p. 2341-2348.

50. Zarganes-Tzitzikas, T., M. Konstantinidou, Y.Z. Gao, D. Krzemien, K. Zak, G. Dubin, T.A. Holak, and A. Domling, *Inhibitors of programmed cell death 1 (PD-1): a patent review (2010-2015).* Expert Opin. Ther. Pat., 2016. **26**(9): p. 973-977.

51. Li, S.T., X.X. Wu, and M.K. Tan, *Gene selection using hybrid particle swarm optimization and genetic algorithm.* Soft Comput, 2008. **12**(11): p. 1039-1048.

52. Wang, J., W.Y. Yuan, and D.J. Cheng, *Hybrid genetic-particle swarm algorithm: An efficient method for fast optimization of atomic clusters.* Comput Theor Chem, 2015. **1059**: p. 12-17.

53. Abualigah, L.M. and A.T. Khader, *Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering.* J Supercomput, 2017. **73**(11): p. 4773-4795.

54. Ali, A.F. and M.A. Tawhid, *A hybrid particle swarm optimization and genetic algorithm with population partitioning for large scale optimization problems.* Ain Shams Eng J, 2017. **8**(2): p. 191-206.

55. Ramachandran, G.N., C. Ramakrishnan, and V. Sasisekharan, *Stereochemistry of polypeptide chain configurations.* J. Mol. Biol., 1963. **7**: p. 95-9.

56. Wu, S.T. and Y. Zhang, *ANGLOR: A Composite Machine-Learning Algorithm for Protein Backbone Torsion Angle Prediction.* PloS one, 2008. **3**(10).

Appendix

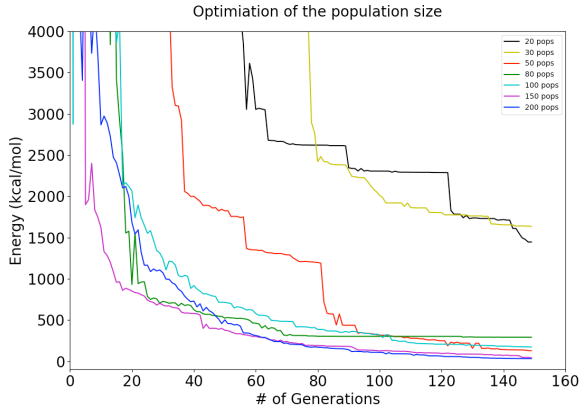


Figure 1. GA population size optimization results. TPP-1 energy minimization using seven different representative population sizes were shown above. Energy on the y-axis was cut at 4000kcal/mol for convenient illustration and comparison.

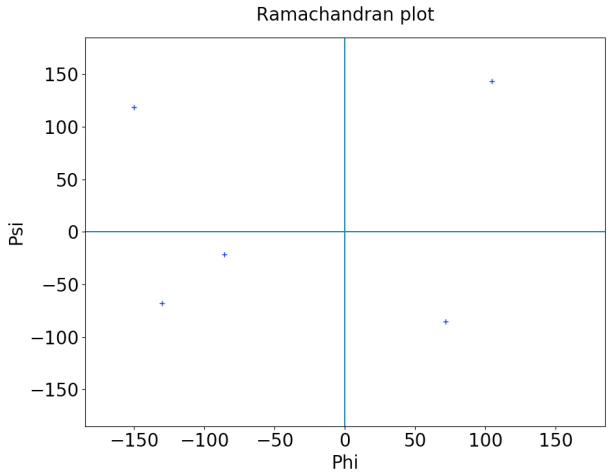


Figure 2. Ramachandran plot for optimized Met-enkephalin using PSO.

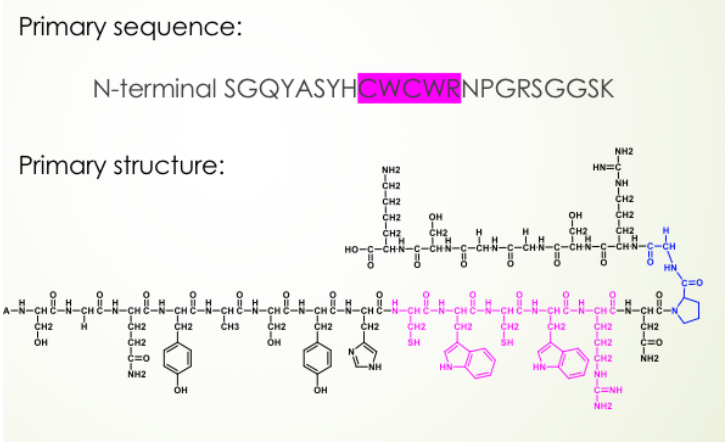


Figure 3. TPP-1 primary sequence and primary structure.

Table 1. All amino acids backbone dihedral angles for Met-enkephalin in the optimized structure by PSO.

#	Amino	ϕ (Phi)	φ (Psi)
1	Tyr	-150.3	118.8
2	Gly	104.4	143.3
3	Gly	71.7	-85.4
4	Phe	-130.2	-67.8
5	Met	-85.97	-21.6

Note: all the angles are expressed in degree (°).

Table 2. All amino acids backbone dihedral angles for TPP-1 in the lowest-energy structure optimized by Hybrid PSO-GA.

#	Amino acid	ϕ (Phi)	φ (Psi)	#	Amino acid	ϕ (Phi)	φ (Psi)
1	Ser	-47.6	156.1	**12**	Trp	-85.6	82.9
2	Gly	-122.0	-68.8	**13**	Arg	145.6	22.8
3	Gln	-153.5	55.1	**14**	Asp	-122.4	72.6
4	Tyr	154.9	122.0	**15**	Pro	-37.9	6.3
5	Ala	53.4	61.6	**16**	Gly	121.1	55.0
6	Ser	-61.9	-44.5	**17**	Arg	-80.0	90.8
7	Tyr	139.6	45.3	**18**	Ser	-140.4	-45.7
8	His	-67.7	120.8	**19**	Gly	121.1	22.8
9	Cys	-85.6	122.7	**20**	Gly	-139.5	88.8
10	Trp	84.4	93.3	**21**	Ser	-92.6	114.2
11	Cys	-73.6	116.3	**22**	Lys	135.4	48.8

Note: all the angles are expressed in degree (°).