Minimum Energy Conformations of Self-Interacting Polymer Chains via Multipopulation Genetic Algorithms (MpGA)

Luis Olivares-Quiroz[1,*] and Marcos A. González-Olvera[2]

[1] *Physics Department and Complex Sciences Graduate Program, Universidad Autónoma de la Ciudad de México. Prol San Isidro 151 CP 09760. Mexico City.*
[2] *Engineering Department. Colegio de Ciencia y Tecnología, Universidad Autónoma de la Ciudad de México. Prol San Isidro 151 CP 09760. Mexico City.*

Abstract. Identification of stable three-dimensional conformations in proteins and peptides that correspond to minima in potential and free energy hypersurfaces has been under intense scrutiny over the past decades since the paradigm structure-function was proposed [1]. This classical paradigm states that most of biologically active conformations in proteins and peptides can be associated with global minima energy states on the energy hypersurface of the polypeptide chain [2,3]. In this work we discuss the onset of macroscopic minimum-energy conformations on small interacting peptides composed by only two types of residues: hydrophobic (A) or polar (B). Based on a previous work in 2D [4], we consider here an interacting three dimensional potential $V = V_1 + V_2$ where V_1 corresponds to a intramolecular bending potential between adjacent residues whereas V_2 is a Lennard-Jones (LJ) type intermolecular potential with both an attractive and repulsive part. In addition, the V_2 term can switch to repulsive or attractive depending on the type of pair interaction AA, AB or BB considered. As a novel approach to the standard geometric-based minimization methods [5–7], we propose a Multipopulation Genetic Algorithm (MpGA) as a minimization algorithm [8]. The central advantage of this approach is a wider search on the energy hypersurface. In order to test the validity of our method, we reproduced in excellent agreement the results previously obtained in 2D by [4]. Our results show than in three dimensions our method enlarges the number of stable macroscopic conformations found for a given polypeptide. As an example of the role played by the interacting pairs AB, AA and BB we discuss as well the case of small diblock polymers and quantify the degree of compactness expected in the three dimensional structure as function of the composition of the chain.

AMS subject classifications: 92D20, 68T20, 92C05, 82C22, 82B30

Key words: Protein folding, genetic algorithms, energy minimization, interacting heteropolymers, coarse-grained models.

---

*Corresponding author. *Email addresses:* `luis.olivares@uacm.edu.mx` (L. Olivares-Quiroz), `marcos.angel.gonzalez@uacm.edu.mx` (M. A. González-Olvera)

1 Introduction

Since the early recognition of the structure-function paradigm in proteins and peptides [1–3], a lot of effort has been dedicated to elucidate the physical mechanisms underlying the passage from a uncoiled structure to a quasi-crystalline three dimensional structure known as the native state [9–11]. The classical view of protein folding in which a single three-dimensional, quasi-crystalline state, is uniquely associated with biological activity has been recently challenged due to the existence of *intrinsically disordered proteins* (IDP's) which do show biological activity even in the absence of a defined three-dimensional structure [12–15]. In spite of this, the large majority of biological macromolecules still obey the modified structure-function paradigm, in which the native state is usually associated with a small ensemble of given structures which do not deviate significantly among them [16]. Most proteins in their native state are actually highly efficient biological enzymes that perform a large variety of biochemical functions, thus, the search for effective computational algorithms that provide physical insight on mechanisms that drive protein folding and protein collapse has been boosted in recent years [17,18].

A usual approach to find conformations that correspond to stable and free energy-minimized conformations is to calculate the equilibrium partition function Z for the heteropolymer chain from where the free energy $F = -RT\ln Z$ can be evaluated and minimized [19,20]. However, due to the presence of complex intra and intermolecular interactions between residues themselves and solvent molecules [21], the exact evaluation of the partition function $Z = Tr(e^{\beta H})$ via the full interacting Hamiltonian H is beyond our current mathematical capabilities [22]. In order to overcome this difficulty, a wide and vast set of numerical methods have been launched. Configurational entropy S and internal energy E (also identified with potential energy V) are the two central contributions to free energy $F = E - TS$. As it has been previously discussed in the literature, both S and E play a central role on protein folding and collapse [23,24]. However, in this work, we shall focus our attention to potential energy V minimization processes in protein folding. In order to achieve this, a wide class of numerical methods have been focused on the search for algorithms to provide efficient potential energy minimization methods [25,26]. Such methods are usually geometric-based, like steepest-descent, Newton-Raphson of different orders and many others [5–7]. The main set back of this approach is that as the number N of residues increases, the number of metastable states in the multi-dimensional energy surface also increases exponentially with N [27]. The intrinsic roughness of the Energy Landscape prevents a full exploration of the energy hypersurface. Such issue becomes particularly relevant since most functional proteins have sizes between $N \sim 10^2 - 10^3$ aminoacid residues, then the potential energy surface $V = V(\{\gamma_i\})$ where γ_i are, for instance dihedral angles $\gamma_i = (\phi_j, \psi_j)$ for each residue, becomes rapidly a very rough surface as N increases with a number typically of the order e^N of metastable local minima states [28]. The probability that a geometric-based searching algorithm is trapped in one of these pseudo energy minimum also increases as the size of the polypeptide chain is increased [29].