AIMS Biophysics

AIMS Biophysics, 8(3): 291–306.
DOI: 10.3934/biophy.2021023
Received: 27 April 2021
Accepted: 12 July 2021
Published: 23 July 2021

http://www.aimspress.com/journal/biophysics

Research article

Contributions of topological polar-polar contacts to achieve better folding stability of 2D/3D HP lattice proteins: An *in silico* approach

Salomón J. Alas-Guardado[1], Pedro Pablo González-Pérez[2,*] and Hiram Isaac Beltrán[3,*]

[1] Departamento de Ciencias Naturales, Universidad Autónoma Metropolitana Unidad Cuajimalpa, CDMX 05300, México. orcid.org/0000-0001-8903-8766
[2] Departamento de Matemáticas Aplicadas y Sistemas, Universidad Autónoma Metropolitana, Unidad Cuajimalpa, CDMX 05300, México. orcid.org/0000-0001-7223-9035
[3] Departamento de Ciencias Básicas, Universidad Autónoma Metropolitana, Unidad Azcapotzalco, CDMX 02200, México. orcid.org/0000-0002-1097-455X

* **Correspondence:** Email: pgonzalez@cua.uam.mx, hibc@azc.uam.mx; Tel: +525558146500, +525558146500; Fax: +525558146500, +52555553189190.

Abstract: Many of the simplistic hydrophobic-polar lattice models, such as Dill's model (called **Model 1** herein), are aimed to fold structures through hydrophobic-hydrophobic interactions mimicking the well-known hydrophobic collapse present in protein structures. In this work, we studied 11 designed hydrophobic-polar sequences, S_1-S_8 folded in 2D-square lattice, and S_9-S_{11} folded in 3D-cubic lattice. And to better fold these structures we have developed **Model 2** as an approximation to convex function aimed to weight hydrophobic-hydrophobic but also polar-polar contacts as an augmented version of **Model 1**. In this partitioned approach hydrophobic-hydrophobic ponderation was tuned as α-1 and polar-polar ponderation as α. This model is centered in preserving required hydrophobic substructure, and at the same time including polar-polar interactions, otherwise absent, to reach a better folding score now also acquiring the polar-polar substructure. In all tested cases the folding trials were better achieved with **Model 2**, using α values of 0.05, 0.1, 0.2 and 0.3 depending of sequence size, even finding optimal scores not reached with **Model 1**. An important result is that the better folding score, required the lower α weighting. And when α values above 0.3 are employed, no matter the nature of the hydrophobic-polar sequence, banning of hydrophobic-hydrophobic contacts started, thus yielding misfolding of sequences. Therefore, the value of α to correctly fold structures is the result of a careful weighting among hydrophobic-hydrophobic and polar-polar contacts.

Keywords: HP model; protein folding/structure; polar contacts; genetic algorithm; convex function

1. Introduction

Nature teaches that there is a clever selection of copolymer sequences of biological molecular monomers aimed to overpass physicochemical challenges. These are well-known as ***foldamers*** in the biological sciences and could be generated from aminoacids, nucleic acids or sugars [1]. Such physicochemical challenges are those needed for achieving key functions as: structural shaping & templates, chaperones/transporters, molecular crowders & recognition motifs, phase transfer agents or surfactants, antioxidants, ionic/molecular reservoirs, detoxifying agents, catalysts or inhibitors, among other biophysicochemical functionalities. These tasks are molecularly fulfilled because folded structures were kinetic/thermodynamic balanced where particular structures corresponds to specific functions [2]. This is an intense research field in the proteins domain, being all-atom approaches the most complex case, and perhaps the hydrophobic-polar (HP) Dill's model represents the simplest case, that is why it's entitled "coarse grained model". All these deal with a classic biological problem called *protein folding paradigm* stating that: *the characteristic or native three-dimensional structure of a given proteic sequence is completely determined (encoded) by the aminoacid sequence itself* [3–5], very related to the "protein folding backbone approach" proposed by Rose *et al* [6]. Nevertheless, both the all atom and the coarse grained models [7], have such high complexity that represent NP-complete systems from the mathematical viewpoint, thus they are clustered into "computationally intractable" set of problems [8].

So no matter of which level of "coding" is the protein folding problem treated its complexity gave insights to many approaches to treat it, in parts, or as a whole, experimental of theoretically. In this line, there is the HP Dill's model, being a simple two letter coding, in which H is hydrophobic and P is polar, both aminoacid moieties, embedded in a more complex proteic sequence oversimplifying the natural 20-aminoacid coding [4,5,9,10]. More sophisticated models are three letter code (3LC), or four letter code (4LC), e.g., the 4LC named HPNX-model is constructed from the split of the polar charged monomers as positive (P), negative (N), and neutral (X) [11,12]. These diversity of models, ranged from simple lattice models [13–17], followed by a myriad of intermediate models, to finally reach the all atom classical molecular dynamics simulations, or even its quantum counterpart [18–22]. All these computational approaches have become cornerstone trials intended to understand such complexity, where authors have stated: computing has been able to provide *ab initio* abstract/conceptual hints or motifs leading to other derived theoretical or experimental developments. This latter aids even in cases where none evidence, nor experimental nor theoretical, is available [20]. But returning to basics the HP-model, no matter of its simplicity, cleverly states that hydrophobic interactions among aminoacids [23] represent one of the principal driving forces that yield native states in proteins. All the latter occurring in such a fast folding as required by natural or anthropogenic molecular processes. Hence, a lot of information and research is yet to be surveyed and discovered, even in the simplistic HP-folding approach, as will be seen forthcoming.

It is well known that the non-covalent interactions, both polar and nonpolar, play an important role during protein folding, mainly those formed by hydrophobic contacts, but as well as hydrogen bonds, aromatic interactions, and salt bridges, for example [24]. Therefore, only considering a hydrophobic core, there is not considered enough structural information nor the total interactions of

the real system. It has been reported in the literature that hydrophobic contacts contribute 60 ± 4% and hydrogen bonds 40 ± 4% to protein stability [25,26], hence both interactions are important factors that contribute in stabilizing the native folded state of proteins. Therefore, the polar counterpart of the protein system scored by polar-polar (P···P) interactions, should be taking into account.

Because of the nature of the side chains of aminoacids, physicochemical differences among them are present, developing important changes in polarity, size and conformation. These properties are responsible to tune the packing of aminoacids comprised in proteins. Nevertheless, itself hydrophobicity is not enough to achieve the accurate folding. Therefore, its polar counterpart accounts for the rest of molecular interactions between aminoacids, also molding the outer protein structure. Where, polar residues could strongly interact between them by P···P contacts developing mainly electrostatic interactions, e.g. of short range as salt bridges, zwitterionic contacts, etc. [27]. Indeed, salt bridges and zwitterionic contacts have resulted very important to enhance protein folding stability and they are generally formed in their outer/polar surface [28,29]. Moreover, polar aminoacids tend to accommodate towards this polar media attracting water molecules and forming hydration cores surrounding polar protein surface [28,29]. Some of these latter are the main contributions to justify taking into account the P···P interactions trying to enhance the simple HP model. Some authors argue that the inclusion of P···P interactions also favor structure compactness, as happens in real proteic systems [30]. Particularly, Kumar & Nussinov clearly state "*While the hydrophobic effect is the major driving force in protein folding, electrostatic interactions are important in protein folding, stability, flexibility, and function*" [29]. One major enhancement is that these augmented HP models are known to develop protein-like features undoubtedly indicating that lattice models include in these simple metaheuristics the fundamental physicochemical principles of proteins [30].

With all the latter in mind, in this current research the main goal is to test another very simple metaheuristic to include P···P interactions into the HP lattice model. In our work group we have explored the study and design of simple HP protein structures considering: 1) the hydrophobic-hydrophobic (H···H) interactions and 2) the restrictions imposed by the intracellular space as an osmolyte effect or molecular crowding [31–33]. In this work we are now proposing to survey the effect of also including P···P interactions to develop better folding strategies. In this new proposal of augmented HP model, the formation of the hydrophobic core is permanently prioritized, and later on the formation of the polar substructures, in contrast with other augmented HP models where this important requisite is not always maintained [30]. This by considering more contributions and physicochemical properties present in these simplistic HP protein systems. For this purpose, were tested 11 HP sequences, some that were difficult to fold in previous works and some other newly designed, in two types of lattices, 2D-square and 3D-cubic. These sequences were selected/designed taking into account including H···H as well as P···P contacts tracking the effect of these two contributions in the sequence folding procedure.

2. Materials and methods

2.1. A lattice model including H···H and P···P interactions

As can be seen below in expressions (1) to (2) describing **Model 1** (Dill's model) [34], the HP-model is abstracting the protein interactions by labeling the aminoacids as H or P. Where backbone structures will be energetically chosen by optimizing the amount of -H···H- contacts due to a

hydrophobic effect. Nevertheless, substructures based on -P···P- interactions, such as hydrogen bonds and salt bridges should also be considered for structure prediction. Hence, extending this traditional HP model to a variant where P···P interactions are also considered in the optimization process conducted to the design of a new approach named **Model 2**, were we use a variant of the convex function, as shown in the expressions (3) to (6) below. Here, this variant of convex function has been employed to tune the weight between H···H and P···P interactions. That is, each H···H interaction is assigned the value α–1 while each P···P interaction is assigned the value -α, with $0 \leq \alpha \leq 1$.

Model 1: Only the H···H interactions are optimized (Dill's model):

$$F = f(e) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} e_{ij} \tag{1}$$

where e_{ij} is defined as:

$$e_{ij} = \begin{cases} -1, if\ i\ and\ j\ are\ both\ H\ residues\ and\ topological\ neighbors \\ 0, otherwise \end{cases} \tag{2}$$

Model 2: Inspired by the convex function, where both H···H and P···P interactions are optimized:

$$F = (\alpha - 1)f(h) + (-\alpha)f(p) = (\alpha - 1)\sum_{i=1}^{n} \sum_{j=i+1}^{n} h_{ij} + (-\alpha)\sum_{i=1}^{n} \sum_{j=i+1}^{n} p_{ij} \tag{3}$$

$$where\ f(h) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} h_{ij}\ and\ f(p) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_{ij} \tag{4}$$

where:

i. h_{ij} counts the hydrophobic interactions defined as:

$$h_{ij} = \begin{cases} 1, if\ i\ and\ j\ are\ topological\ neighbors \\ 0, otherwise \end{cases} \tag{5}$$

ii. p_{ij} counts the polar interactions defined as:

$$p_{ij} = \begin{cases} 1, if\ i\ and\ j\ are\ topological\ neighbors \\ 0, otherwise \end{cases} \tag{6}$$

iii. α, $0 \leq \alpha \leq 1$, determines the weight given to each type of interaction. Alpha values very close to 0 favor the hydrophobic collapse observed in Dill's model (expression (1)).

iv. Expression (3) is subject to the following restriction: the best hydrophobic core obtained (optimum number of H···H interactions) must be conserved.

v. The term score is defined here as a double conditional in function F (expression (3)) that i) like Dill's model, generates the best score in the HH core, as the maximum amount of H···H interactions and ii) maximized the outer P···P contacts, without breaking/disassembling the HH core.

Note that in expression (3), when $\alpha = 0$, **Model 2** is simplified to **Model 1**. And the score term will be used to rank folding results.

One of the principal aims in this contribution is testing α values in **Model 2** that conduct to the formation of structures with better local minima, preserving hydrophobic core and now folding also PP substructures.

2.2. The assembly and optimization of 2D-HP structures & methodology

HP-model states that any aminoacid sequence (S_i) could be transcribed and defined as expression (7):

$$S_i \in \{H, P\}, i = 1,2, \dots, n \tag{7}$$

where "*n*" is the length of the chain.

As shown in Figure S1 (in Supplementary), the HP transcribed sequence could then be folded in a simplistic 2D or 3D-lattice, following discrete movements between the neighboring cells that conform the lattice. Particular details of this are given in Supplementary Material.

From the transcribed HP-sequence, a random population of HP 2D/3D-structures is generated in a 2D/3D-lattice, following the algorithm explained and illustrated in Figure S2 (in Supplementary). Subsequently, on this random population of structures, an optimization process begins executed by an evolutionary algorithm [31,32] whose fitness function is given precisely by **Model 2**. The main characteristics of the evolutionary algorithm used are listed in Table S1 (in Supplementary). Moreover, the folding methodology was carried out for the *in silico* experiments on our *Evolution* bioinformatics platform, see further details in Supplementary.

3. Results

3.1. Structural design: from HP-sequences to HP simplistic structures

In a previous work [33], we designed a set of simple HP-sequences and the corresponding optimal 2D-HP folding for each of these. The latter allowed us to know in advance the optimal score expressed only by the number of H···H interactions, that characterized each of the expected structures. On that research, the optimization of these 2D-HP structures was carried out employing an evolutionary algorithm whose fitness function was given by **Model 1**, only developing HH structures, as described above. In the present work, we return to some of these HP-sequences designed to fold in 2D-square lattice model, and plus we incorporated new HP-sequences as well as the corresponding expected 2D or 3D-structures. And moreover, we use **Model 2**, now including also P···P interactions to pursue a better folding score, but also to verify the contributions of H···H contacts in its own HH substructure, but now also exhibiting P···P contacts in order to form the missing PP substructure. Table 1 lists the characteristics of the HP-sequences in which we will test the contribution of P···P contacts in the formation of the corresponding expected structure. The 2D-sequences are S_1-S_8, meanwhile the 3D-structures are S_9-S_{11}.

Table 1. Designed target HP sequences and their best score achieved for folding when **Model 2** is used. That is, PP contacts are also rewarded. This variant of the convex function is used to tune the value of the reward between H···H and P···P contacts as $f(h)$ and $f(p)$ respectively.

S_{ID}	DTS	SL	FS	Model 1 (Dill's case)	Model 2 (from convex function)				
				$f(e)$ OF / BFF	$f(p)$	α_B	$f(h)\times(\alpha-1)$	$f(p)\times(-\alpha)$	OF / BFF
S_1	$P_{11}H_{16}P_{20}$	47	2D	-9 / -9	21	0.3	-6.3	-6.3	-12.6 / -12.6
S_2	$H_3P_4H_3P_4H_3$	17	2D	-6 / -6	2	0.3	-4.2	-0.6	-4.8 / -4.8
S_3	$P_4H_{12}P_4$	20	2D	-6 / -6	4	0.3	-4.2	-1.2	-5.4 / -5.4
S_4	$H_4P_4H_4P_4H_4P_4H_4$	28	2D	-12 / -12	5	0.3	-8.4	-1.5	-9.9 / -9.9
S_5	$P_4H_{25}P_4$	33	2D	-16 / -16	4	0.2	-12.8	-0.8	-13.6 / -13.6
S_6	$H_5P_4H_5P_4H_5P_4H_5P_4H_5$	41	2D	-20 / -20	**8**	**0.2**	**-16.0**	**-1.6**	**-17.6 / -17.2**
S_7	$P_3H_2P_2H_2P_5H_7P_2H_2P_4H_2P_2HP_2$	36	2D	-14 / -14	7	0.1	-12.6	-0.7	-13.3 /-13.3
S_8	$H_9P_2H_2P_4H_2P_2H_2P_4H_2P_2H_2P_4H_2P_2H_2P_4H_2$	49	2D	-24 / -22	8	0.05	-22.8	-0.4	-23.2 / -23.2
S_9	$P_8H_{27}P_8$	43	3D	-28 / -28	14	0.1	-25.2	-1.4	-26.6 / -26.6
S_{10}	$P_4H_9P_4H_9P_4H_9P_4$	43	3D	-30 / -30	4	0.1	-27	-0.4	-27.4 / -27.4
S_{11}	$P_4H_{16}P_4H_{16}P_4$	44	3D	-34 / -34	8	0.1	-30.6	-0.8	-31.4 / -31.4

S_{ID} = Sequence identification. DTS = Designed Target Sequence; SL = Sequence Length; SF = Folding Space; OF = Optimal Fitness; BFF = Best Fitness Found; $f(h)$ = Expected H···H contacts; $f(p)$ = Expected P···P contacts; α_B = Best α value. Bold means that $OF_{P\cdots P}$ was not found. Note: fitness is equivalent to score.

3.2. Role of the computational tool and the methodological approach in obtaining the results

The folding of the sequences proposed in Table 1 and, therefore, the generation of the expected 2D/3D-structures, was carried out as described in Supplementary (employing parameters stated in Table S2) using the *Evolution* computational tool (http://bioinformatics.cua.uam.mx/site/) [35,36] and guided by the methodological approach described in Supplementary (see Figures S2 & S3). *Evolution* is a bioinformatics platform developed by our work group [31–33], where its functionality has recently been enhanced from **Model 1**, to now implement optimization based on **Model 2**.

It is necessary to mention that the values of α (in **Model 2**) are the result of a preprocessing phase carried out by executing many batches of experiments *a priori*, finding that α values greater than 0.3 led rupture of hydrophobic core in the majority of structures, thus banning the hydrophobic collapse. When α is higher it moves away from the Dill's model, hence taking a bet for this concept of hydrophobic collapse as main driving force, and that it should be maintained but not as a unique contribution. The **Model 2** optimization normalizes between 0 and 1 the importance of ponderation or weighting given to both H···H and P···P interactions in the folding process. So that the sum of these weights always equals to 1. Moreover, α values equal or lower than 0.3 means that the formation of the hydrophobic core is only receiving 70% or less of the overall weight, in the expected 2D/3D-structures. And this value is just enough weighting to ensure that hydrophobic core is preserved, reason why this weighting scheme still provides the required hydrophobic collapse in structures.

3.3. *Analysis of the in silico experiments*

Also, in Table 1 are gathered the results of S_1-S_{11} folding, applying both **Model 1**, where $\alpha = 0$, and **Model 2**, where $\alpha \neq 0$. As an example, in Figure 1A is observed that topological H···H contacts are 9 for S_1, therefore applying **Model 1**, the optimal fitness (OF) is −9 and the value of best fitness found (BFF) in experiments resulted also −9, but all P···P contacts are not optimized (Figure 1B). When considering **Model 2**, we tested different α values and the best value of α (α_B), in this case 0.3, is the one that produced the expected optimal structure also folding PP substructure (Figure 1C). Therefore, applying **Model 2**, the HH fitness (F_{HH}) was obtained by multiplying the total expected hydrophobic contacts ($f(h)$) by (α_B-1), that is, $f(h)$ (α_B-1) = 9 (−0.7) = −6.3. And the PP fitness (F_{PP}) was obtained by multiplying the total expected polar contacts ($f(p)$) by ($-\alpha_B$), this is $f(p)$ ($-\alpha_B$) = 21(−0.3) = −6.3. Finally, the BFF is equal to the sum of F_{HH} and F_{PP}, that is (−6.3) + (−6.3) = −12.6, precisely matching this value with that of OF. And in this same way all the other experiments were carried out and BFF & OF values were obtained.

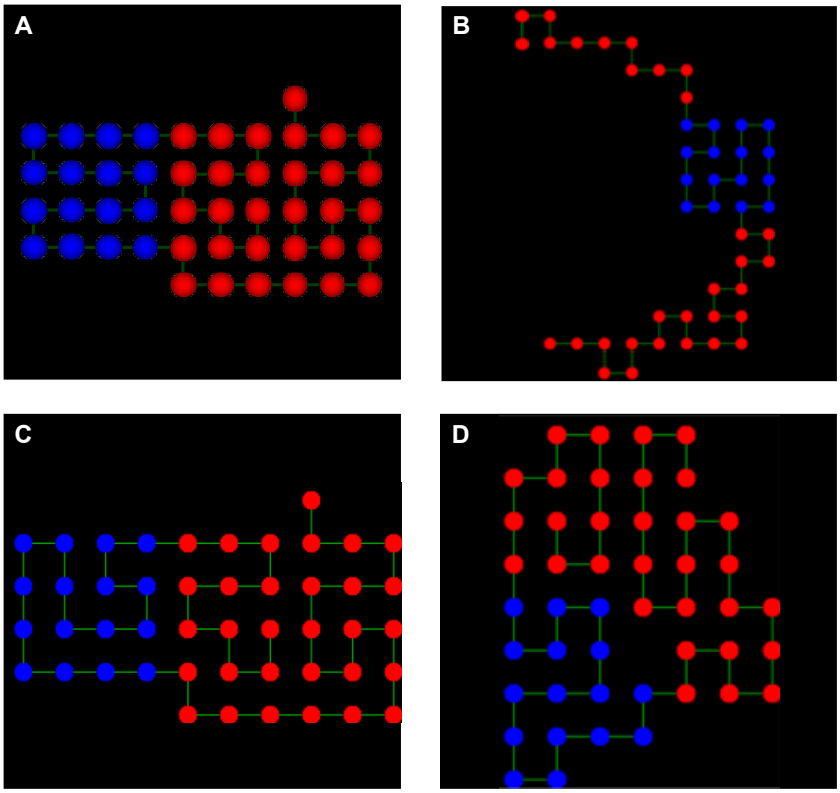

Figure 1. S_1 HP 2D structures: (A) Expected; (B) **Model 1**; (C) **Model 2**, α = 0.3 (degenerated optimum); (d) **Model 2**, α = 0.5 (overweighed). Hydrophobic core is able to form different degenerated structures, see Figure 1B vs. Figure 1C, evidencing that the correct weighting of P···P minimizes overall degeneracy of S_1, due to the formation of PP substructure. In Figure 1B the two P branches pointed out of the HH substructure in different sides. In Figure 1C, the two P branches pointed out of HH substructure in the same side to fold the PP substructure. In Figure 1D, with overweighed α values (higher than optimum), the expected conformation is not achieved but misfolded structures occurred. Note that H and P residues are labeled in blue and red colors, respectively.

4. Discussion

4.1. Structural analysis of folded S_1-S_{11}

Figure 1 shows the behavior of sequence S_1, taking account **Model 1** (Figure 1B) and its respectively modification using **Model 2** (Figure 1C). S_1 is a H_{16} hydrophobic core, nevertheless, the terminal P_{11} & P_{20} branches were stipulated to be one larger and one shorter, in order to survey their correct fold and its own PP substructure. This should cause a fine tune in the proposed α value to preserve the hydrophobic core but also to fold a new polar core in a neighboring fashion. Another possible effect with this is how the H···H core degeneracy behaves with this new weighting scheme. It was easy to observe that **Model 1** results only developed well folded hydrophobic core, obtaining the expected nine H···H contacts, but few P···P contacts were produced. **Model 2** usage with α value of 0.3, the hydrophobic core is maintained and a compact PP substructure with 21 topological contacts was developed, hence this value of $\alpha = 0.3$ was considered to be optimum (Figure 1C).

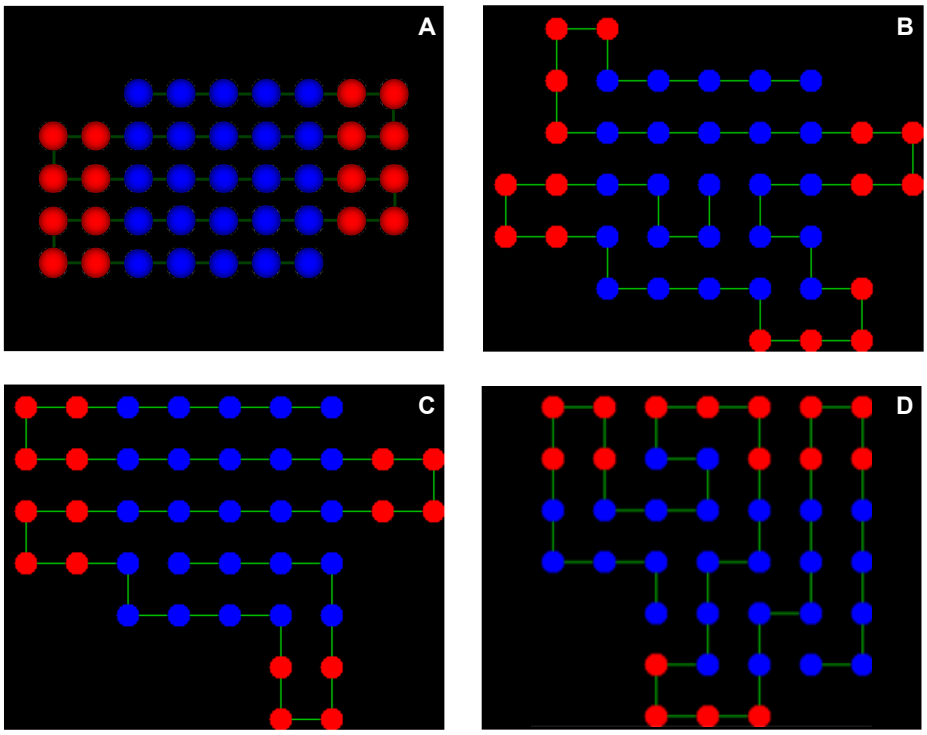

Figure 2. S_6 HP 2D structures: (A) Expected; (B) **Model 1**; (C) **Model 2**, $\alpha = 0.2$ (optimum); (D) Model 2, $\alpha = 0.4$ (overweighed). Hydrophobic core is able to form different degenerated structures, see Figure 2B vs. Figure 2C, evidencing again that the correct weighting of P···P minimizes overall degeneracy of S_6, due to the formation of the 4PP substructures. In Figure 2C the two left P branches interact to each other to score another P···P contact, but right branches this is not occurring, that is why suboptimal BFF = −17.2 was obtained. In Figure 2D, with overweighed α values, the expected conformation is not achieved, nor HH core, neither PP substructures. The colors labeled of the H and P residues are the same as Figure 1.

The 2D-HP structure expected as a result of the optimal folding of the S_6 sequence is illustrated in Figure 2A. There, the expected structure corresponds to a 5 × 5 hydrophobic core with short P_4 branches, positioned between each two rows of HH core. This 2D-HP structure was proposed by us [33], where its folding was studied using both **Model 1** and a simulated intracellular medium.

The resulting folding of S_6 sequence in 2D lattice is illustrated in Figure 2B-D. As can be seen in Figure 2B, when sequence folding is guided by **Model 1**, optimal conformation of hydrophobic core is reached ($OF_{H\cdots H} = -20$), however, only 2 of the 4 expected polar substructures achieve proper conformation. Note that the suitable conformation of a polar substructure is characterized by a single polar contact, which corresponds to a P···P score equal to −1. Adequate conformation observed of these two polar substructures is due to restrictions imposed by 5 × 5 hydrophobic core formation, there is no other way to achieve this. Figure 2C shows optimal conformation of hydrophobic core with $OF_{H\cdots H} = -20$, folded using $\alpha = 0.2$, also achieving conformation of the 4P substructures, although one of them does not show expected orientation. One of the goals of this experiment was to achieve $OF_{P\cdots P} = -8$ (see Figure 2C and Table 1), but only $OF_{P\cdots P} = -6$ was obtained at least in the batches of experiments carried out. But another goal was also to achieve the conformation of the 4P substructures and reach local $OF_{P\cdots P} = -4$, i.e., -1 for each one of the $4P_4$ substructures, which was well accomplished. It should be pointed that for $\alpha = 0.05$ and $\alpha = 0.1$ it was also possible to reach the best possible conformation shown in Figure 2C. However, for $\alpha > 0.2$, the expected 2D-HP optimal structure was not achieved (Figure 2D).

The S_8 sequence should produce the complex 2D-HP structure illustrated in Figure 3A as a result of its folding. There, the expected secondary structure corresponds to a 5 × 5 hydrophobic core included in a 7 × 7 polar-nonpolar structure. Figure 3B illustrates S_8 folding guided by **Model 1**, showing an imperfect hydrophobic core, since $OF_{H\cdots H} = -24$ and $BFF_{H\cdots H} = -22$ (Table 1). This result shows both complexity of expected 2D-HP structure and the need of restriction/boundary conditions provided by considering P···P contacts contribution. Figure 3C shows optimal conformation of HH square 5×5 with $\alpha = 0.05$, achieving $OF_{H\cdots H} = -24$, at the same time that surrounding polar perimeter reached expected $OF_{P\cdots P} = -8$. However, acquired optimal HP-structure was not exactly the expected in the surrounding polar substructure (see Figure 3A), being a degenerate optimal. For $\alpha > 0.05$ not optimal nor degenerated-optimal structures were found, see one example in Figure 3D.

For the 2D folded sequences (S_1-S_8), S_8 sequence was the one that produced the most complex structure. Hence, when **Model 2** was used, folding of S_8 sequence was precisely the one that required a lower α value to yield optimized structure. Seeming that at higher complexity of the hydrophobic core of the optimal structure, the required α value should be lower. Also, with α values higher than optimum, expected HH core is not achieved but misfolded structures occurred.

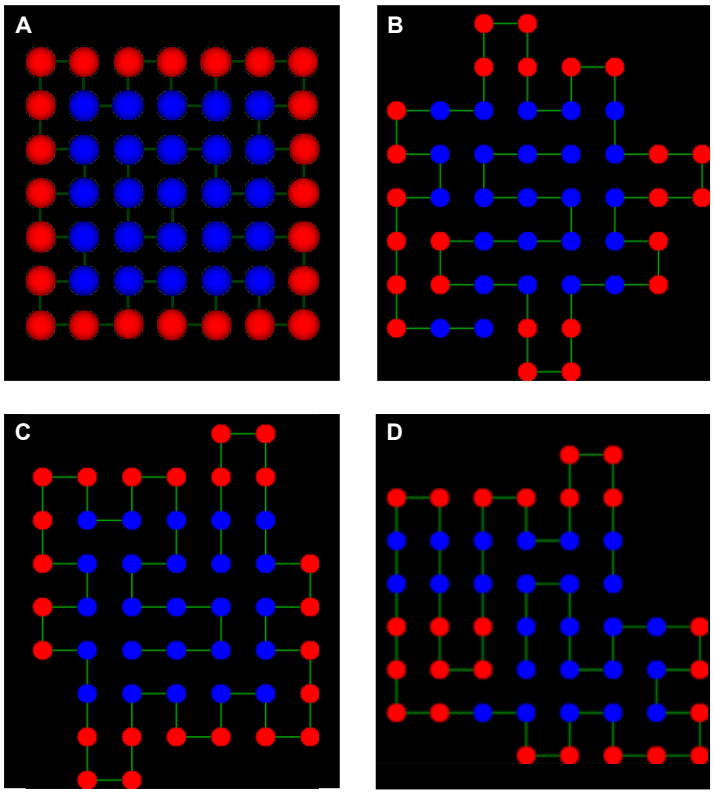

Figure 3. S_8 HP 2D structures: (A) Expected; (B) **Model 1**; (C) **Model 2**, α = 0.05 (optimum); (D) **Model 2**, $\alpha = 0.1$ (overweighed). Hydrophobic core in Figure 3B was not obtained, but in Figure 3C it was obtained in a degenerate form in comparison to that in Figure 3A. Here also degenerated PP substructure was achieved in which two corner P4 fragments fold outwards but giving OFFP···P = −8. For $\alpha > 0.05$ was not possible to accomplish optimal conformation nor its degenerated optimal as shown the example (Figure 3D), at least in batches of experiments carried out. In these cases, $0.05 < \alpha \leq 0.3$, it was perceived that as the α value increased, only suboptimal hydrophobic cores were generated. The colors labeled of the H and P residues are the same as Figure 1.

The folding of the 2D HP S_2, S_3, S_4, S_5, and S_7 sequences, using **Model 1** and **Model 2**, is provided in the Supplementary, Figures S4-S8, respectively.

The S_{11} sequence (see Table 1) is a very interesting case to fold in 3D-space. Looking at its primary structure ($P_4H_{16}P_4H_{16}P_4$) we could think that optimal should be a 3D-structure made up of two 4 × 4 hydrophobic layers, with 16H residues each; and 3PP substructures, each with P_4 residues, two located at the end of each 4 × 4 hydrophobic layer and one as a connecting unit between two 4 × 4 hydrophobic layers. However, finding the optimal folding does not only mean reaching the maximum number of H···H contacts, with **Model 1**, or H···H/P···P contacts, with **Model 2**. This process also involves maximizing the compactness of resulting 3D-structures, as we will see below.

The folding of S_{11} is shown Figure 4A using **Model 1**, and in Figure 4B-D using **Model 2**. In Figure 4A, optimal folding of hydrophobic core is achieved, with $OF_{H\cdots H} = -34$ (see Table 1). Note that P···P contacts were not obtained. Also optimal folding of hydrophobic core was obtained by compacting structure into 3HH and not 2HH layers, resulting in the most compact structure found by optimization algorithm. As shown in Figure 4B, with $\alpha = 0.1$, the optimal conformation of the

hydrophobic core is preserved while the polar substructures also reach their best folding, achieving expected 8 P···P contacts (Table 1). Note that each volume shares two rows between them as a boundary condition. With values higher than optimal $\alpha = 0.1$, conformation of hydrophobic core was not achieved nor PP substructure, e.g. Figure 4C-D show suboptimal structures with $\alpha = 0.2$ & 0.3, respectively.

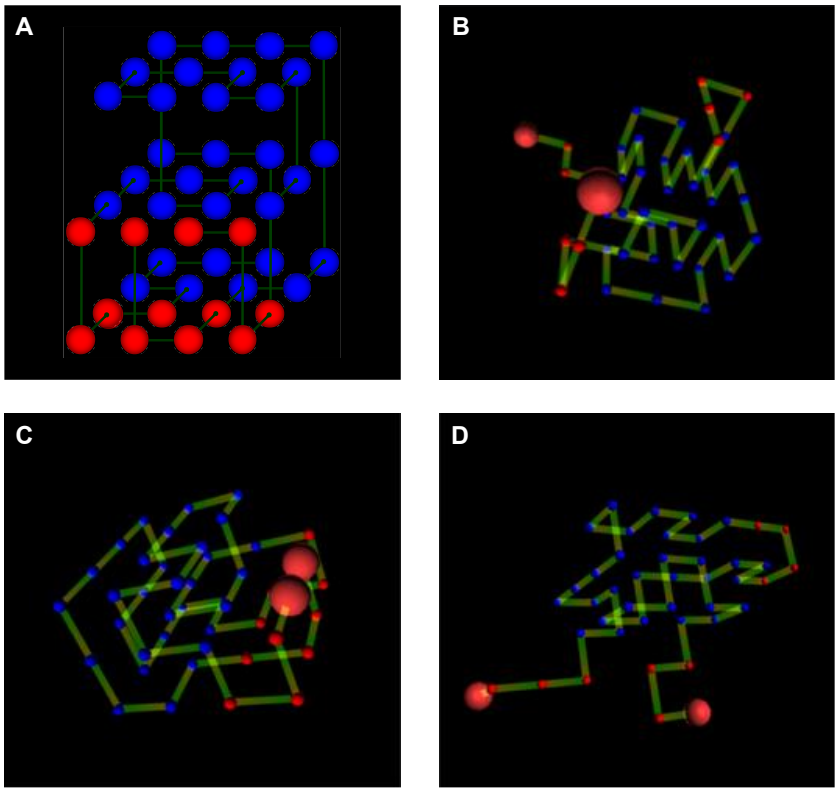

Figure 4. S_{11} HP 3D structures obtained by using: (A) Expected; (B) **Model 1**; (C) **Model 2**, $\alpha = 0.1$ (specular optimum); and (D) **Model 2**, $\alpha = 0.2$ (overweighed). Main core (largest volume) of the resulting 3D-HP structure (Figure 4A) is given by a HP rectangular prism, consisting of two 3 × 4 layers with 12H residues each and a hybrid 3 × 4 layer integrated by 12 residues, 8H + 4P. Alternatively, smallest volume of obtained 3D-HP structure corresponds to a rectangular prism consisting of two layers of 2 × 4 residues, one layer with 8P residues and other of 8 residues, 4P + 4H. Note that each volume shares two rows between them as a boundary condition. Note also that in Figure 4D both H···H & P···P interactions were diminished due to the effect of increasing α. The colors labeled of the H and P residues are the same as Figure 1.

The folding of the 3D HP S_9 and S_{10} sequences, using **Model 1** and **Model 2**, is provided in the Supplementary, Figures S9 and S10, respectively. On the other hand, degenerated structures are diminished using **Model 2**, due to conformational restraints imposed by PP substructure formation. Tested structures in 2D & 3D correctly folded with $\alpha = 0.05$–0.3 and $\alpha = 0.1$, respectively. Values of α beyond these yielded misfolded structures.

4.2. Summary discussion

The summarized results of the folding of S_1 to S_{11} sequences, using **Model 1** and **Model 2**, are shown in Table 2 and Figure 5. Here, it is necessary to mention that for each sequence, 10 batches of experiments were run per approach, each batch consisting of 10 trials. Table 2 provides for each sequence the following features: 1) best size of hydrophobic core, 2) best number of polar contacts, and 3) the best total number of contacts), using both **Model 1** and **Model 2**. The folding tendency for features 1), 2) and 3) is shown in Figure 5 captions (A), (B), and (C), respectively.

Table 2. Summarized results of the folding of S_1 to S_{11} sequences using **Model 1** and **Model 2**.

Folding features		**Sequences**										
		S_1	**S_2**	**S_3**	**S_4**	**S_5**	**S_6**	**S_7**	**S_8**	**S_9**	**S_{10}**	**S_{11}**
Number of H···H contacts	Expected	9	6	6	12	16	20	14	24	28	30	34
	Model 1	9	6	6	12	16	20	14	22	28	30	34
	Model 2	9	6	6	12	16	20	14	24	28	30	34
Number of P···P contacts	Expected	21	2	4	5	4	8	7	8	14	4	8
	Model 1	7	1	0	2	0	3	6	8	0	2	0
	Model 2	21	2	4	5	4	6	7	8	14	4	8
Total number of contacts (H···H + P···P)	Expected	30	8	10	17	20	28	21	32	42	34	42
	Model 1	16	7	6	14	16	23	20	30	28	32	34
	Model 2	30	8	10	17	20	26	21	32	42	34	42

As can be seen in Figure 5A, the expected hydrophobic core was reached for all eleven sequences when **Model 2** was used. Only one miss was found for **Model 1**, failing to reach the expected hydrophobic core of the S_8 sequence. This fact could be justified as a consequence of the role played by P···P contributions correcting in such a way the folding of the hydrophobic core. For instance, when analyzing some of the expected HP 2D structures (see Figures 1A, 2A & 3A), it could be seen that the expected 2D P substructures impose restrictions on the degrees of freedom of the hydrophobic elements, and as a result contributing to the formation of the expected hydrophobic core if the sequence allows the latter. Figure 5B shows the P···P folding tendency, therein it could be noted that when **Model 2** was used the expected polar substructure was achieved for ten of the eleven sequences, where the most contrasting examples for this polar effect are for S_1, S_9 & S_{11}. Only in the case of sequence S_6, **Model 2** failed to achieve the polar optimum substructure (Figure 2C). Lastly, in Figure 5C can be observed the global folding tendency, that is the joint contribution of H···H and P···P interactions, for sequences S_1 to S_{11} using **Model 1** and **Model 2** clearly showing that **Model 2** is in all cases better than **Model 1**, again where the most contrasting examples are for S_1, S_9 & S_{11}.

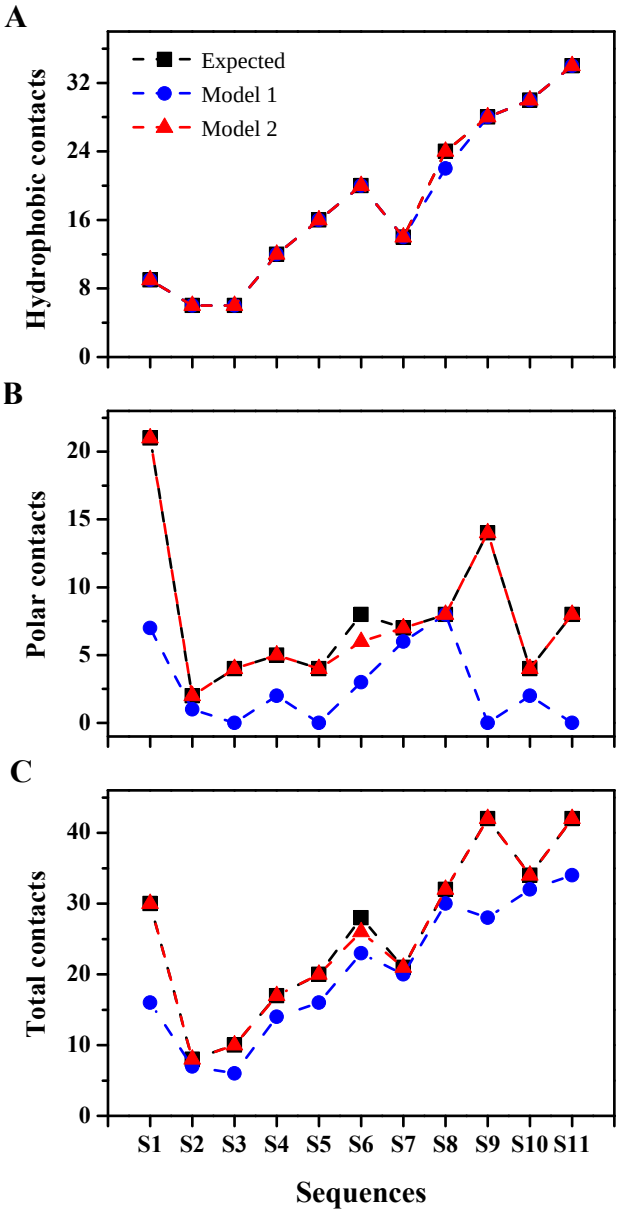

Figure 5. Folding tendency for sequences S_1 to S_{11}. The number of expected contacts is compared with those found using **Model 1** and **Model 2** for the cases: (A) H···H contacts, (B) P···P contacts, and (C) total number of contacts (H···H + P···P). Dashed lines are only guide for the eye.

5. Conclusions

Here were tested 11 HP-sequences, S_1-S_8 in 2D-square lattice and S_9-S_{11} in 3D-cubic lattice using two folding approaches, a) Dill's model, named **Model 1**, and b) a model inspired by convex function, named **Model 2**. Last model is heuristically aimed to weight H···H (Dill's model) and also P···P contacts, to gather more structural information in order to reach better folding solutions in any given HP-sequence. In **Model 2**, H···H interactions were tuned as α-1 and P···P interactions as -α, and HP folding in all cases was more successful than **Model 1**. When α values above 0.3 are employed to fold (high

P···P weighting) this started to ban H···H contacts, and misfolding occurred. There were needed α values very close to 0.3 to optimize longer or shorter sequences, with low difficulty in their folded structures. In comparison, more complex 2D-sequences, required much lower α values of 0.05-0.1 to achieve the accurate folding, being the length of the sequence not as important as its structural complexity. The 3D-sequences, of higher dimensionality and complexity, also required $\alpha = 0.1$ to improve folding results. Moreover, α parameter might be included into a multi-objective optimization scheme, in such a way that the same algorithm should be able to find both the best α value and the optimal folding at once.

Acknowledgments

Authors would like to thank the support provided by Oscar Sánchez Cortés, in the improvements of the *Evolution* bioinformatics platform. This research was supported by CONACyT (project 0222872 HIB and project A1-S-46202 SJAG) and by Universidad Autónoma Metropolitana.

Conflict of interest

The authors declare no conflict of interest.

Author contributions:

Salomon J. Alas-Guardado participated in the experiment design and drafted the manuscript; Pedro Pablo González-Pérez coordinated the software engineering work of the Evolution bioinformatics platform, participated in the experiment design, conducted the in silico experiments, and drafted the manuscript. Hiram Isaac Beltrán conceived the design of the target 2D/3D structures and foldamers, participated in the experiment design and drafted the manuscript. All authors gave final approval for publication.

References

1. Goodman CM, Choi S, Shandler S, et al. (2007) Foldamers as versatile frameworks for the design and evolution of function. *Nat Chem Biol* 3: 252–262.
2. Hill DJ, Mio MJ, Prince RB, et al. (2001) A field guide to foldamers. *Chem Rev* 101: 3893–4012.
3. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181: 223–230.
4. Dill KA, Ozkan SB, Shell MS, et al. (2008) The protein folding problem. *Annu Rev Biophys* 37: 289–316.
5. Dill KA, MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338: 1042–1046.
6. Rose GD, Fleming PJ, Banavar JR, et al. (2006) A backbone-based theory of protein folding. *Proc Natl Acad Sci U S A* 103: 16623–16633.
7. Hu J, Chen T, Wang M, et al. (2017) A critical comparison of coarse-grained structure-based approaches and atomic models of protein folding. *Phys Chem Chem Phys* 19: 13629–13639.
8. Berger B, Leighton TOM (1998) Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. *J Comput Biol* 5: 27–40.

9. Shatabda S, Newton MAH, Rashid MA, et al. (2014) How good are simplified models for protein structure prediction? *Adv Bioinf* 2014: 867179.
10. Madain A, Dalhoum ALA, Sleit A (2018) Computational modeling of proteins based on cellular automata: A method of HP folding approximation. *Protein J* 37: 248–260.
11. Backofen R, Will S, Bornberg-Bauer E (1999) Application of constraint programming techniques for structure prediction of lattice proteins with extended alphabets. *Bioinformatics* 15: 234–242.
12. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annu Rev Phys Chem* 48: 545–600.
13. Gupta A, Maňuch J, Stacho L (2005) Structure-approximating inverse protein folding problem in the 2D HP model. *J Comput Biol* 12: 1328–1345.
14. Hoque T, Chetty M, Sattar A (2009) Extended HP model for protein structure prediction. *J Comput Biol* 16: 85–103.
15. Shmygelska A, Hoos HH (2005) An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem. *BMC Bioinf* 6: 30.
16. Bechini A (2013) On the characterization and software implementation of general protein lattice models. *PLoS One* 8: e59504.
17. Abeln S, Vendruscolo M, Dobson CM, et al. (2014) A simple lattice model that captures protein folding, aggregation and amyloid formation. *PLoS One* 9: e85185.
18. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106: 1589–1615.
19. Ferina J, Daggett V (2019) Visualizing protein folding and unfolding. *J Mol Biol* 431: 1540–1564.
20. Compiani M, Capriotti E (2013) Computational and theoretical methods for protein folding. *Biochemistry* 52: 8601–8624.
21. Beck DAC, Daggett V (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* 34: 112–120.
22. Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr Opin Struct Biol* 24: 98–105.
23. Dill KA, Bromberg S, Yue K, et al. (1995) Principles of protein folding—a perspective from simple exact models. *Protein Sci* 4: 561–602.
24. Newberry RW, Raines RT (2019) Secondary forces in protein folding. *ACS Chem Biol* 14: 1677–1686.
25. Pace CN, Fu H, Fryar KL, et al. (2011) Contribution of hydrophobic interactions to protein stability. *J Mol Biol* 408: 514–528.
26. Pace CN, Scholtz JM, Grimsley GR (2014) Forces stabilizing proteins. *FEBS Lett* 588: 2177–2184.
27. Leonhard K, Prausnitz JM, Radke CJ (2003) Solvent–amino acid interaction energies in 3-D-lattice MC simulations of model proteins. Aggregation thermodynamics and kinetics. *Phys Chem Chem Phys* 5: 5291–5299.
28. Zhou HX, Pang X (2018) Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem Rev* 118: 1691–1741.
29. Kumar S, Nussinov R (2002) Close-range electrostatic interactions in proteins. *Chem Bio Chem* 3: 604–617.
30. Moreno-Hernández S, Levitt M (2012) Comparative modeling and protein-like features of hydrophobic-polar models on a two-dimensional lattice. *Proteins* 80: 1683–1693.

31. Alas SJ, González-Pérez PP (2016) Simulating the folding of HP-sequences with a minimalist model in an inhomogeneous medium. *Biosystems* 142: 52–67.
32. Gonzalez-Perez PP, Orta DJ, Peña I, et al. (2017) A computational approach to studying protein folding problems considering the crucial role of the intracellular environment. *J Comput Biol* 24: 995–1013.
33. de Jesús Alas S, González-Pérez PP, Beltrán HI (2019) In silico minimalist approach to study 2D HP protein folding into an inhomogeneous space mimicking osmolyte effect: First trial in the search of foldameric backbones. *Biosystems* 181: 31–43.
34. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29: 7133–7155.
35. Beltrán HI, Rojo-Domínguez A, Gutiérrez MES, et al. (2009) Exploring dimensionality, systematic mutations and number of contacts in simple HP ab-initio protein folding using a blackboard-based agent platform. *Int J Phys Math Sci* 3*:* 256–265.
36. Pérez PPG, Beltrán HI, Rojo-Domínguez A, et al. (2009) Multi-agent systems applied in the modeling and simulation of biological problems: A case study in protein folding. *World Acad Sci, Eng Technol* 58: 128.

AIMS Press