International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:3, No:4, 2009

Exploring Dimensionality, Systematic Mutations and Number of Contacts in Simple HP *ab-initio* Protein Folding Using a Blackboard-based Agent Platform

Hiram I. Beltrán, Arturo Rojo-Domínguez, Máximo Eduardo Sánchez Gutiérrez and Pedro Pablo González Pérez

Abstract**—**A computational platform is presented in this contribution. It has been designed as a virtual laboratory to be used for exploring optimization algorithms in biological problems. This platform is built on a blackboard-based agent architecture. As a test case, the version of the platform presented here is devoted to the study of protein folding, initially with a bead-like description of the chain and with the widely used model of hydrophobic and polar residues (HP model). Some details of the platform design are presented along with its capabilities and also are revised some explorations of the protein folding problems with different types of discrete space. It is also shown the capability of the platform to incorporate specific tools for the structural analysis of the runs in order to understand and improve the optimization process. Accordingly, the results obtained demonstrate that the ensemble of computational tools into a single platform is worthwhile by itself, since experiments developed on it can be designed to fulfill different levels of information in a self-consistent fashion. By now, it is being explored how an experiment design can be useful to create a computational agent to be included within the platform. These inclusions of designed agents –or software pieces– are useful for the better accomplishment of the tasks to be developed by the platform. Clearly, while the number of agents increases the new version of the virtual laboratory thus enhances in robustness and functionality.

Keywords**—**genetic algorithms, multi-agent systems, bioinformatics, optimization, protein folding, structural biology.

Dr. Hiram Isaac Beltrán* (corresponding author), Departamento de Ciencias Naturales, DCNI, Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Pedro Antonio de los Santos 84. Col. San Miguel Chapultepec. Miguel Hidalgo 11850. México, D.F. e-mail: hbeltran@correo.cua.uam.mx. Tel. +52 (55) 26363853.

Dr. Arturo Rojo-Domínguez, Departamento de Ciencias Naturales, DCNI, Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Pedro Antonio de los Santos 84. Col. San Miguel Chapultepec. Miguel Hidalgo 11850. México, D.F. e-mail: arojo@correo.cua.uam.mx.

Máximo Eduardo Sánchez, Departamento de Matemáticas Aplicadas y Sistemas, DCNI, Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Pedro Antonio de los Santos 84. Col. San Miguel Chapultepec. Miguel Hidalgo 11850. México, D.F.

Dr. Pedro Pablo González* (corresponding author), Departamento de Matemáticas Aplicadas y Sistemas, DCNI, Universidad Autónoma Metropolitana, Unidad Cuajimalpa. Pedro Antonio de los Santos 84. Col. San Miguel Chapultepec. Miguel Hidalgo 11850. México, D.F. Tel. +52 (55) 26363805. e-mail: pgonzalez@correo.cua.uam.mx.

I. Introduction

ABOUT 25 years ago, Robson stated, 'the mechanism by which a protein adopts its biologically active conformation after biosynthesis is becoming a little clearer' [1]. Nevertheless this type of challenge arises into a category of scientific problems where regardless of the increasing computational capacity, it is clear that a brute force solution is unattainable. The systematic solution involves visiting and evaluating all possible conformations, and this is an impossible task as Levinthal underlined in such a work published as early as 1968 [2]. The inherent variability and complexity of protein folding makes it a benchmark test case per excellence for high throughput optimization tools as we recently stated [3]. Even though tremendous effort has been done in this line, the needed clarity for understanding the protein folding problem and not to say the functionality of those biomolecules is not yet reached [4]. Indeed, these latter two points have been improved but a lot of work to do remains and it is to be studied by the next generations.

As stated, huge improvements related to theoretical [5], experimental [6] or combined approaches [7] have been aimed in order to understand and eventually handle the protein folding mechanisms [8, 9]. These approaches, not completely understood, have even been employed to attain more complex problems, e.g. structural design [10-13]. In a wider landscape, these efforts have been directed towards a further range of objectives: i) from the simple purpose of knowledge generation in basic sciences [14]; ii) to the successfully application in technological fields, such as medicine, food science, and green strategies regarding the obtaining of fuels; iii) also in the fields of chemicals and composite materials design, in bioremediation treatments, in water purification, among others.

As it is always observed during scientific and technological research, the systematic and also the serendipitous paths to obtain results have led to whole emerging fields. Some of them are *in vivo* folding [15], protein misfolding diseases [16], evolutionary analysis [17], epigenetic factors in protein function, biological structures and health [18, 19], enzymatic

biotechnological applications [14] and green industrial processes [20, 21], etc. Today we are situated in a post-genomic world where understanding protein folding has never been more important. This topic has a broad impact and interest in fields ranging from structural biology [22-25] to materials science [26-29] and beyond [17, 30, 31]. Almost one decade ago Radford underlined the importance of the protein folding field. This denoted with a dramatically increase in the number of scientific reports present in the literature and also due to the broad benefits obtained by its study and thorough application [32, 33].

At this level of understanding, the systematical mutation of residues in proteins has led to the design of biomolecules with enhanced properties useful in biotechnological applications [34, 35]. This importance has encouraged us to study, with basic *ab initio* tools, the intrinsic features of mutations in simplistic HP protein sequences (this means described only by ***h***ydrophobic and ***p***olar monomers). All this with a new bioinformatics platform recently developed and designed in order to trigger and permit an efficient feedback from computational experimentation as will be described below.

II. THE BIOINFORMATICS PLATFORM

A. Requirements

The bioinformatics platform must give, to any user e.g. particularly to biologists and bioinformatics, a set of tools for studying and exploring a number of problems belonging to structural biology. It must have some characteristics required by those scientists for supporting their activities. The system has to provide also a set of built-in functionalities for the elaboration and automatic analysis of these data. Taking into account the previous ideas, the bioinformatics platform must satisfy the requirements described in the next subsection.

1) Knowledge elaboration and cooperative work constitute the first requirement for the bioinformatics platform. It has to implement a set of methods to:
   - Elaborate the knowledge and heuristics on protein folding from biology, biochemistry, physics, etc.
   - Produce new information on the folding process and thus the protein structure.
   - Allow the integration, collaboration and coordination of some known models and algorithms, previously developed, to deal with the study of the biological problem already mentioned, and
   - Develop and test new models, not only based on the known ones, but also allowing the user to experiment on innovative forms of protein representation, fitting functions, structural analysis or combinations of these tools.
2) This requirement of bioinformatics platform deals with the organization of data and its representation. The data are distinguished between:
   - Biological knowledge and heuristics on the protein folding, necessary to the models and evolutionary algorithms that will be used, and
   - Starting data that has to be fed to the algorithms and models producing new data to be analyzed. This produced data has to be shared and accessible by all the users of the system, both “computational" and human ones.
3) In this requirement, the bioinformatics platform has to provide tools for allowing the interaction of human users, such as biologists and biochemists, with the system so that they can dynamically add knowledge on the properties of the protein or feedbacks on the data produced by the algorithms. In this way the experts can control the entire flow, which gives rise to the final solution of the protein folding prediction.
4) Finally, the bioinformatics platform has to be enriched with a method for controlling the overall flux of information, and for the coordination between the different participants mentioned above. This makes the platform a very flexible instrument, not only for developing models and algorithms, but also to produce relevant biologic data.

B. An effective computational approach for modeling biological systems

Multiagent systems (MAS) [36-39] have been considered a powerful paradigm for modeling and engineering complex systems [40, 41]. Biological systems are in fact complex systems, and might be effectively modeled in terms of MAS [3].

A MAS is an ensemble of interacting agents. Interaction among agents can be realized by direct communication, according to some agent communication language, or indirectly. The latter is achieved by exploiting some environmental resources acting as mediating communication/coordination abstractions [42, 43]. Such abstractions range from a simple communication channel to information spaces, such as blackboards [44], useful for agents to synchronize their tasks. The term blackboard-based agent architecture has been widely used in recent years for referring to MAS with agents interacting indirectly through a blackboard.

We claim here that an effective and appropriate approach that fulfills the bioinformatics platform requirements, described in the previous subsection, is gathered in the blackboard-based agent architecture. This approach has proven to be an effective and appropriate abstraction level to construct whole models of a diversity of biological systems [41, 45]. If an agent represents an individual active component of the system (i.e. a protein or a gene), the overall MAS, including the blackboard, captures the full set of the biological components including also the structures involved in their

interaction (i.e. cellular compartments, intracellular messengers, etc.) leading to a thorough mimic of the system under study. In particular, the blackboard can be adopted to model the various patterns of interaction that can be found in biological processes, such as chemical information exchanges. As we will see below, the bioinformatics platform presented in this work is an example of such architecture [3].

C. Architectural model and behavior

As can be seen in Fig. 1, the architecture of bioinformatics platform is defined in terms of the following main components:

- the blackboard
- conformation generator agents
- genetic search agents
- interface agents

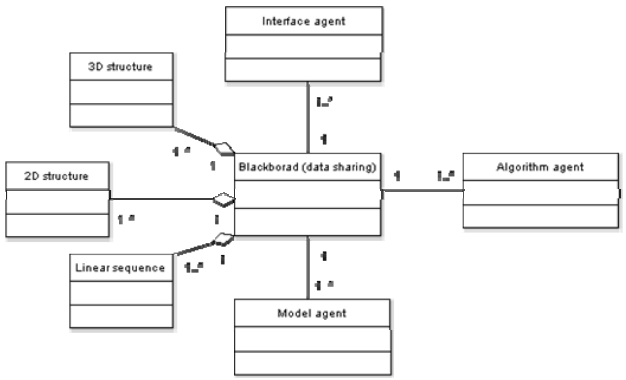

Fig. 1 Architectural model of the bioinformatics platform. The class diagram shows the main building blocks of the bioinformatics platform in its current version

In the architectural model, the blackboard is the data structure through which the agents communicate to each other, synchronize and also where the problem solution emerges. The data on the blackboard must be shared and accessible by all the users of the system, both "computational" and human ones. Such knowledge has to be organized in a set of hierarchical levels that collect data containing more and more information on the protein folding process (e. g. linear amino acid sequence, protein conformations in 2D, protein conformations in 3D, biological and physical knowledge about protein folding, etc.). This organization permits the solution of the problem through an opportunistic and incremental assembling process of solution elements, employed as building blocks, all this in order to obtain a satisfactory configuration supporting the main result. As can be seen in Fig. 2, the blackboard organization herein proposed defines five steps: i) the amino acid sequence level, ii) the HP sequence level, iii) the protein conformational space level, iv) the approximation algorithm workspace level, and v) the plausible protein conformation level.

Once the blackboard steps have been defined, the architectural model of bioinformatics platform defines three different types of agent: model agents (e.g. conformation generator), genetic search agents and interface agents.

"Conformation generator agents", have the ability to create a 2D/3D conformational space from a target linear sequence, implementing the HP model [46-48] on triangular or square (2D) lattices, or cubic (3D) lattices. Fig. 3 represents a *3D conformation generator agent*. As shown in this figure, *currentCoordinates* and *newCoordinates* are two kinds of attributes of this agent, whereas *moveStraightOn()* is a behavior. The condition part of this behavior is fulfilled when the agent can place the *(j)th* bead straight on relative to the bond between the previous two *(j-1)th* and *(j-2)th* beads.

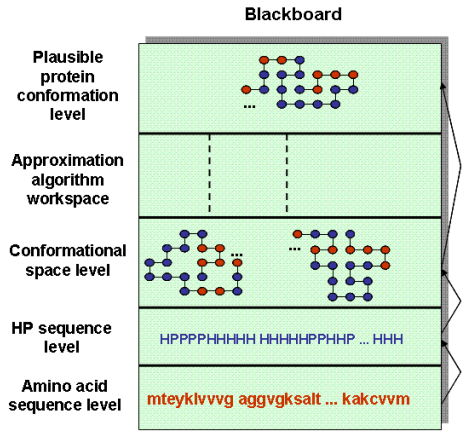

Fig. 2 Defined blackboard steps in the bioinformatics platform

"Genetic search agents", carry out the heuristic search and optimization processes on the conformational space. These agents encapsulate a wide variety of genetic operators and techniques necessary to find good-enough solutions to protein folding problem. Some of them are selection, reproduction, crossover, mutation and elitism. In Fig. 4 is depicted a genetic search agent of *selection operator* type. As can be seen in this figure, the behavior repertory of this agent involves a vast range of selection operators from genetic algorithm techniques [49-51].

"Interface agents", provide a set of functionality for allowing the activity of human users. They are the means for allowing human experts, such as biologists, to interact with the system so that they can dynamically add knowledge on the biological problem or direct the feedback from results provided by the algorithms. In this way, the experts can control the entire flow which gives rise to the final solution of the problem.

Fig. 5 shows the workflow of the major activities to be executed through the bioinformatics platform when studying and exploring protein folding problem. As depicted in this figure, the single bead HP chain, corresponding to a particular amino acid sequence, is submitted to the platform with particular parameters of lattice model, search and analysis,

predefined by the user. The platform runs its "conformation generator agents" and the initial conformational space (named generation 0) is constructed. Now, the user can visualize this initial conformational space and interact with it, exploring its characteristics through the wide range of 2D/3D charts. After that, the platform runs the genetic algorithm by means of the interaction among its "genetic search agents". Results yielded by this run or experiment are analyzed by the user which can experiment on different generations and conformations produced. This experimentation can also suggest changes in the system description that can be implemented in the platform for the next runs. During all this process, the invoking of "interface agents" is needed just for the communication of the user with the platform.

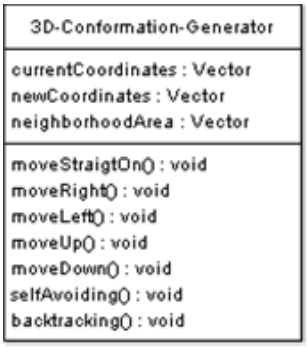

Fig. 3 The "3D-Conformation-Generator" agent

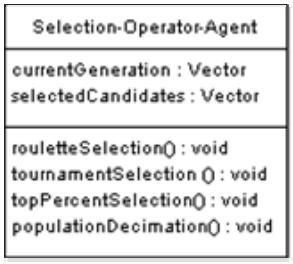

Fig. 4 The "Selection-Operator" agent

III. SOME INTERESTING PROBLEMS INVOLVING PROTEIN FOLDING

Design of the platform involved the study of any optimization problem in biological sciences, such as intermolecular association (docking), ligand and drug design, molecular clustering, etc. This first version of the platform deals with protein folding, being a NP problem, which will permit to test and show the ability of the software to deal with strongly difficult optimization tasks. In particular, such as enormous spaces are to be explored, complex fitting functions with several contributors or variables, and multiminima energetic surfaces are going to be needed and analyzed. In order to demonstrate its capacities we describe below a couple of computational experiments using a minimal description of the biological system; namely, using a bead-like model with HP representation of the polypeptide chain. All this using lattices or discrete space approaches for the movement of the chain during folding process. The simpler meaning is that every amino acid is symbolized by a hydrophobic (H) or by a polar (P) bead. Employed graphical representations use blue and red colors, respectively, for these types of residues. Also denoting the simplest stage of complexity the chosen fitting function assigns a (-1) score for H-H close non-covalent contacts and (0) elsewhere. As well, it is the purpose of both experiments, not only to explore relevant issues in protein folding and explore the ability of the platform in this minimal representation, but also to detect important tasks to be integrated in the next version of the platform. Additionally it is meant to demonstrate the flexibility and adaptability of this tool to be used as a real virtual laboratory, which architecture and design receives feedback, not only from the user, but also from the inherent requirements derived from its use in particular experiments.

IV. EXPERIMENTS

Experiment 1, *Search of local determinants of structure*. Protein folding is a complex process for several reasons; being one of them the enormous amount of possible conformations. Mother Nature, during evolution, has avoided the screening of an *in vivo* random search of the native conformation. The natural result is the intrinsically directed sampling to reach just to the relevant conformation to attain biological activity. This particular state is frequently hypothesized as being the lowest-energy conformation. The clue in this proposal, in order to avoid a random search being prohibitive both *in vivo* and *in silico*, is that short range local interactions can direct the first stages of the folding process. This merely depends in the existence of short segments of the polypeptide chain with the ability or potential of getting i) stable conformations, ii) fast folding sequences, or iii) both of the latter.

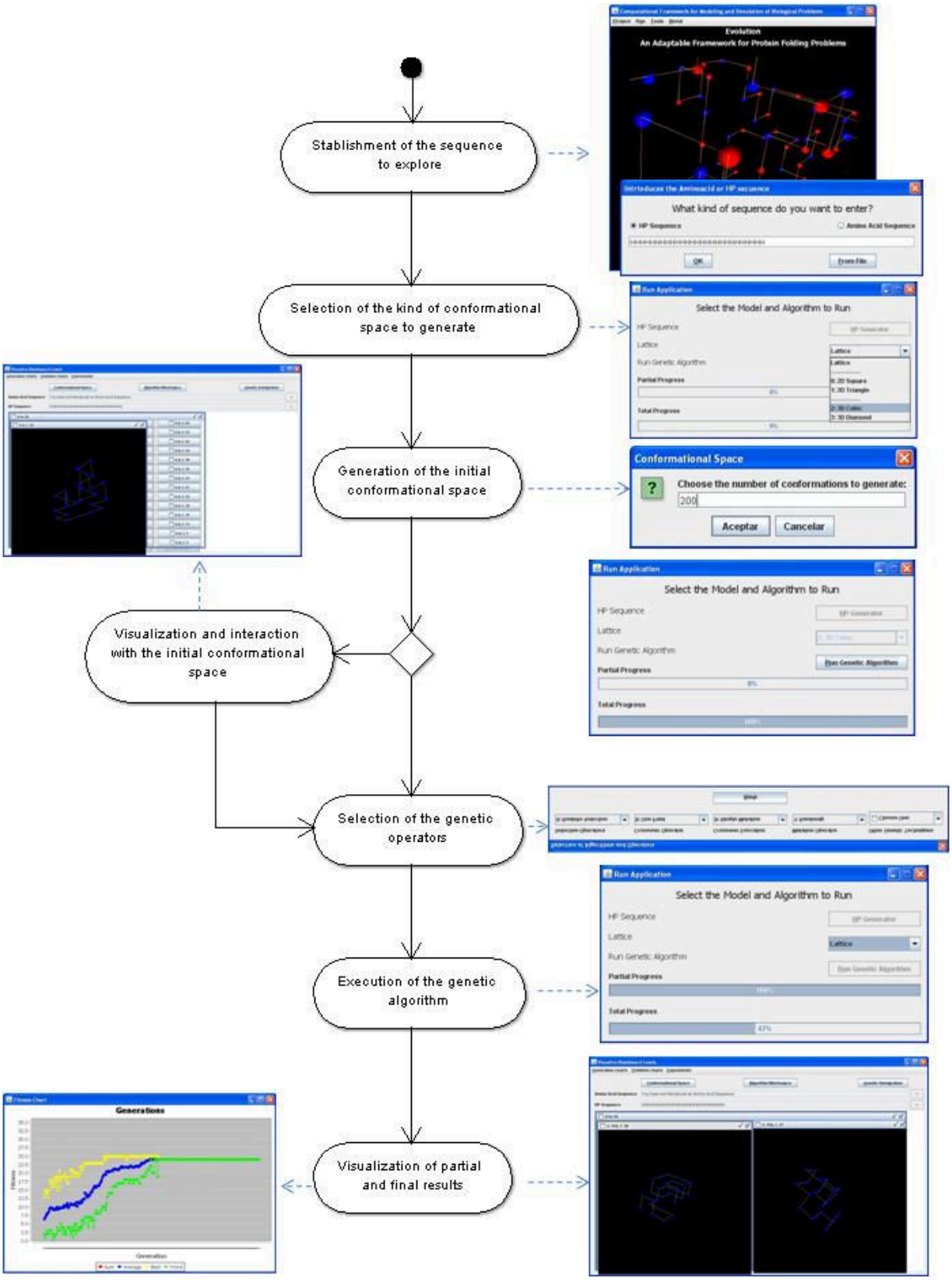

Fig. 5 Bioinformatics platform workflow. The activity diagram describes the behavior of the bioinformatics platform during a run

The existence of a folding unit was originally proposed in the 1970´s but coined as *foldons* until 1996 by Wolynes *et al.* [52], and references therein. The interest for them has recently reborn and it has been revisited by the same group [53], as well as utilized and reviewed by Englander [54, 55] and Oliveberg [56, 57]. A foldon is defined as a part of a protein which folds quasi-independently, guiding it to its folded structure without complex information needed. Alpha helices are one of the most stable local structures, forming the first hydrogen bond between the *(j)th* and the *(j+3)th* residues. With the previously defined fitting function and chain representation, the closest contacts can be formed just between these residues in the 2D-square and 3D-cubic lattices, and between the *(j)th* and the *(j+2)th* residues in a 2D-triangular lattice. In order to explore, if the simplest fitting function can detect particularly stable segments, able to direct the folding of the remaining segments of the chain, we performed a systematic study of all the possible HP sequences in the longest segment we can handle, namely an hexamer.

By symmetry, we discarded all palindromic sequences (those who read the same in the forward and backward directions) since they are indistinguishable by the current representation. Anyway, we are aware about the directionality of real polypeptide chains, but by now is not the case of study. In order to compare between different lattices, we also discarded all sequences without the possibility of forming a contact in any of the three types of lattice described above (square, cubic and triangular); being rejected, for example, hexamers such as PHHHPP, since this chain would not be able to establish H-H contacts in square or cubic lattices.

TABLE I.
SELECTED SEQUENCES OF HEXAMERIC HP SYSTEMS

HHHHHH	PPHPHH
PHHHHH	PPHHPH
HPHHHH	PHPPHH
HHPHHH	HPPPHH
PPHHHH	HPHPPH
PHPHHH	PHPHPH
PHHPHH	HPPHPH
PHHHPH	HHPPPH
PHHHHP	HPPPPH
HPPHHH	HPPHPP
HPHPHH	PHPPHP
HPHHPH	PHHPPH
HHPPHH	PPPPPP

Taking into account all the later details, for this experiment we finally selected the sequences shown in Table I. Results from these runs showed that the triangular lattice (Figs. 6 and 7) provides a higher number of contacts but also yields a higher number of equivalent structures with identical scores; situation known as a degenerated system. Also, we found that with the current simplest fitting function no particular sequence can be clearly detected as a folding unit. This result could have been found for different reasons as follows:

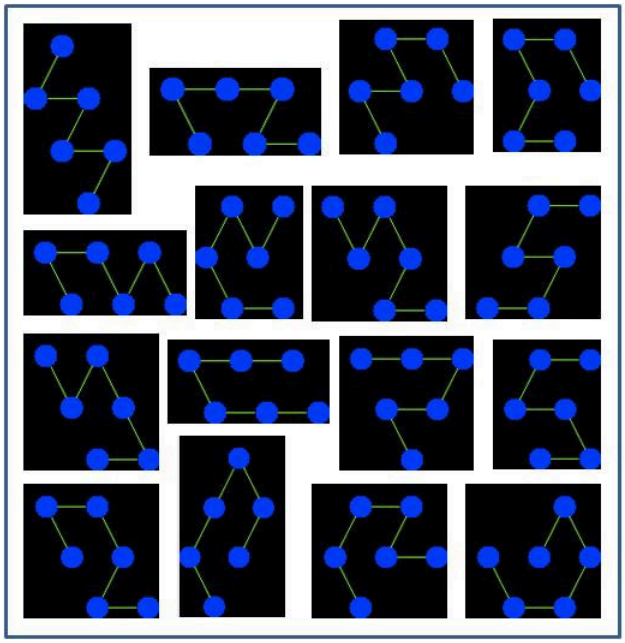

Fig. 6. Folding results of poly-H hexamer (**HHHHHH**) in a 2D-triangular lattice. Several degenerated structures, with a maximum score of (-4) can be detected with the use of the platform. A (-1) is added to the score for each **H** bead with another in a close neighbor point in the lattice with no bond between them. Meanwhile for the other sequences (see Fig. 7) varying degrees of degeneracy and maximum number of contacts can be tracked.

a) The hexamer might be smaller than needed to provide a unique conspicuous conformation in our HP representation in any of the assayed lattices. This can be now tested analyzing longer sequences, but due to the combinatorial explosion of different possible sequences for longer segments, a systematical search of each one is not longer attainable. So, a module to generate and test random HP sequences, of a particular length, and to test their folding uniqueness, stability or efficiency, must be added to the platform. This particular design is already under development. This will allow the user to compose and automatically launch experiments such as the one presented for the hexamer, in different lattices, but also to test the whole process with different fitting functions, and genetic algorithm operators, among other variables in the platform. This shows, on the one hand, the great variety of possible experiments that can be assayed by changing initial parameters and, on the other, the strong promissory potential of the platform.

b) Another reason to fail in the detection of folding units may be the simplicity of the fitting function. Some of the degenerated structures are not as globular as others; see for example the different conformers in Fig. 7 with the same score and different packing. A fitting function, which provides

favorable scores for "solvent exposed" polar residues could break the conformational degeneration found. Also, the calculation of some global indexes to describe geometry, such as the radius of gyration or the maximal diameter of the chain can be programmed in the platform and also to be included into the fitting function; both favoring more packed conformations from the more extended ones. These indexes are being integrated for the next version of the platform.

c) Finally, it must be considered that during local folding of short segments, the rest of the chain is also present; this means that both termini of the hexamer are connected to other beads thus restricting the conformational space of the potential foldons. Since our platform already includes a self-avoiding feature, which prevents two beads to occupy the same lattice site, we can add to the hexamers tails of poly-P beads in both termini to simulate the effect of the rest of the chain.

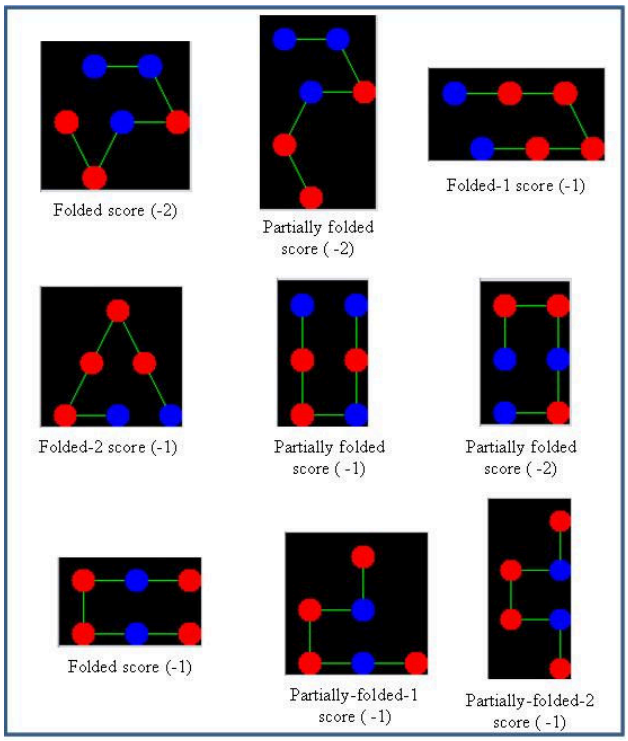

Fig. 7 Folding of poly-H hexamer in triangular and square lattices. Note that with the current fitting function the same optimal score can be obtained with close packed (globular) chains and also with low density (relatively extended) ones, even for short sequences. This suggests the inclusion of packing indexes in the fitting function, such as radius of gyration

We can conclude that a relatively modest experiment, such as the one just described, can yield a good amount of information to interact with the platform and to provide feedback for its adaptation to the particular issue to be explored. The bioinformatics platform is designed to incorporate any upgrades such as the ones suggested by the results discussed for hexamer folding experiments.

Experiment 2, *Standardization of optimal parameters for the genetic algorithm*. Before launching a folding process, two parameters, related with the genetic algorithm of search, are required from the user.

- The first is the initial number of conformations (C), which means the size of the population of each generation, constant during all the optimization process.
- The second is the number of such generations (G), which means the number of cycles, of offspring generation and new parents selection, which are considered necessary for obtaining an optimal conformation. No default can be suggested prior to this second experiment.

In this second experiment we explore the folding of an HP chain, only in the 2D-squared lattice, using different chain lengths with three different H/P ratios: unitary and around 1/3 and 2/3. This means that we will analyze the efficiency of folding regarding different number of (C) and (G) for chains with the same number of H and P residues and for chains with a proportion of around 33% of each of these types of amino acids. Since the number of possible sequences with these characteristics can be unachievable to be systematically studied, we only present in this first strategy uniformly composed sequences. This means those with H and P residues regularly placed along the chains. We also considered that P residues, in any of the ends of the chain, will not contribute to the folding since no scoring can be assigned with our fitting function (see fundamentals of hexamer folding in **Experiment 1**), which only considers H-H contacts (see Fig. 7), so again we avoided these sequences. Each folding experiment was run five times looking for robust results for each condition. The lengths (L) explored go from 6 to 20 residues and some of the assayed sequences are represented in Table II.

As expected for a system with more potential contacts, higher scores are found in chains with more H residues, but also in these sequences a higher number of optimal scored conformations can be found. This means that degeneration is strongly influenced by the number of favorable potential contacts. This at least with our current fitting function and with evenly dispersed HP residues, which avoids the local concentration of any type of residues in a particular segment of the chain. From these runs, it has been found that for the chain sizes assayed, populations of 20 or 40 conformers do not provide enough genetic variability, in order to permit reliable detection of any of the optimal conformations (Fig. 8). This means that it is necessary to start from a more abundant set of initial conformations to systematically reach good results at the end of the run, with the present genetic operators.

TABLE II. Some of the HP sequences assayed

L	HP ≈ 1/2	HP = 1	HP ≈ 2/1
10	HPPHPPHPPH	HPHPHPPHPH	HHPHHPHPHH
12	HPPPHPPHPPPH	HPHPHPPHPHPH	HPHHPHHPHHPH
14	HPPHPPHPPHPPPH	HPHPHPPHPHPHPH	HHPHHPHHPHHPHH
16	HPPHPPHPPHPPHPPH	HPHPHPPHPHPHPHPH	HHPHHPHHPHHPHHPH
18	HPPHPPPHPPHPPPHPPH	HPHPHPPHPPHPHPHPH	HPHHPHHPHHPHHPHHPH
20	HPPHPPHPPHHPPHPPHPPH	HPHPHPHPPHPHPHPHPHPH	HHPHHPHPHHHHPHHPHPHH

Concerning the parameters C and G, no default can be initially proposed since we have observed that the values of both parameters, which efficiently conduct the run to an optimal solution, strongly depend on chain length and complexity. Regarding the effect of the length of the chain, we can see that the minimum amount of conformations must be higher as the number of residues increases, being this correlation stronger with HP ratio > 1. Also with this type of hydrophobic chains, we frequently found that the genetic algorithm founds favorable conformations but gets lost them during the random search of new populations. For this reason we designed and implemented an elitism tool, where the best two members of any given population are guaranteed to *survive* to the next generation; thus assuring the permanence of the highest scores in each run. These optimal members are allowed to mix with the rest of the population and again the best two are promoted directly to the next generation. Another upgrade derived from the feedback of these folding assays concerns with the form of the crossover operator in the genetic algorithm.

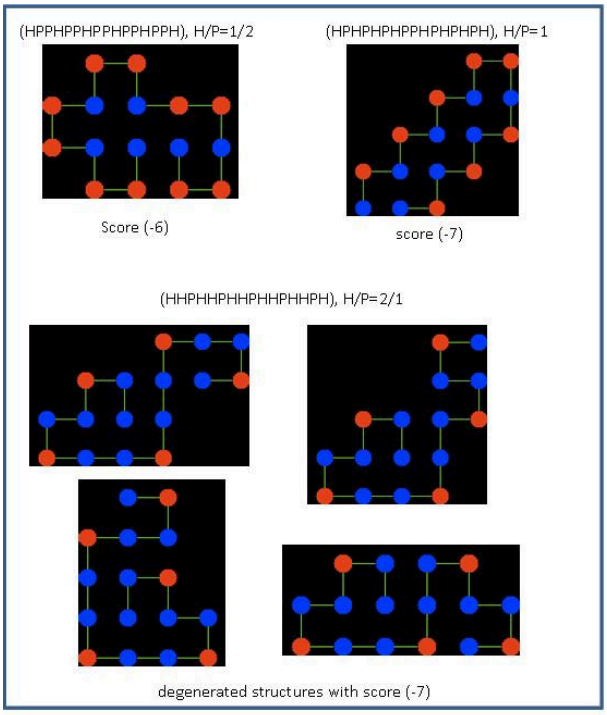

Fig. 8 Folding 16-mer sequences in a 2D-squared lattice

To generate an offspring from two members of the previous generation a crossing point is randomly selected, and then the first moiety of a member is combined with the second moiety of the other. But, since these moieties correspond to fragments of conformations, their association or fusion into a new complete chain is not an easy task if the self-avoiding principle is observed. Thus, in the original algorithm presented here the conformation, of the first beads in the second moiety, was randomly mutated in order to allow an association of the two segments without bead overlapping. But this procedure biased the search allowing conformational mutation only on the second moiety (thus always maintaining the first moiety in its original geometry). Hence we implemented a function to select by chance the first or the second moiety for this task.

Results obtained with the original algorithm or using elitism and random mutation showed no statistically significant difference when the number of initial conformations is small. But, using optimal values for the size of the generations, a notably increment in efficiency of the algorithm is found. This was observed not only in faster convergence, but also in a lower number of generations, and also in finding more favorable conformations. Nevertheless, in some runs, using elitism, we have detected an early convergence phenomenon. This is characteristic where the genetic variability of subsequent generations is gradually depleted and the whole population is converted to the same and suboptimal conformer, trapping the search in a local minimum. Currently, we are developing an index to detect this decrement of variability in the population, and the possibility of injection of

random conformers to allow further exploration of the conformational space.

V. CONCLUSIONS

In this work, we presented a bioinformatics platform obtained by integrating genetic algorithms and other computational techniques into a blackboard-based agent architecture. The version of the tool presented herein allows studying and exploring optimization algorithms in complex problems belonging to structural biology, such as folding. Our analysis considered the impact of several platform features for the protein folding problem: the conformational representation (2D triangular, 2D square and 3D cubic lattices), mutations in simplistic HP protein sequences and the activation of specific genetic techniques such as elitism and "non bias" crossover.

From the experiments carried out we can conclude that:

1. Relatively modest experiments can yield a good amount of information to interact with the platform and to provide feedback for its adaptation to the particular issue to be explored.
2. The bioinformatics platform is designed to incorporate any upgrades such as the ones suggested by the results discussed for Experiment 1.
3. The platform is able to provide and integrate a very useful set of tools to the bioinformatics or biologist user to explore a great amount of possibilities, incorporating and updating an endless amount of indexes, correlations and subroutines, depending on the problem under study.
4. The variety of tools will incrementally and opportunistically emerge depending on usage of the platform and knowledge of the biological problematic to be explored, these can be of physical, chemical, statistic and biological nature.
5. Finally, the fine tuning of the gathered tools enables a particular accuracy of the solution of each tested problem depending on its complexity. In the sense that simpler problems will require less tools and complex problems will require more of them.

ACKNOWLEDGMENT

We acknowledge for support from Acuerdos del Rector General UAM. A.R-D. thanks for financial support from CONACyT México.

REFERENCES

[1] Robson, B., *PROTEIN FOLDING.* Trends in Biochemical Sciences, 1976. **1**(3): p. 49-50.
[2] Levinthal C, *Are there pathways for protein folding?* Journal De Chimie Physique Et De Physico-Chimie Biologique, 1968. **65**(1): p. 44-&.
[3] Gonzalez, P.P., et al., *An adaptable multi-agent based virtual laboratory to explore complex problems in structural biology. Starting case study: simplistic protein folding.* Journal of Computational Biology, 2009: p. Submitted.
[4] Sohl, J.L., S.S. Jaswal, and D.A. Agard, *Unfolded conformations of alpha-lytic protease are more stable than its native state.* Nature, 1998. **395**(6704): p. 817-819.
[5] Sadqi, M., D. Fushman, and V. Munoz, *Atom-by-atom analysis of global downhill protein folding.* Nature, 2006. **442**(7100): p. 317-321.
[6] Creighton, T.E., *EXPERIMENTAL STUDIES OF PROTEIN FOLDING AND UNFOLDING.* Progress in Biophysics & Molecular Biology, 1978. **33**(3): p. 231-297.
[7] Snow, C.D., et al., *Absolute comparison of simulated and experimental protein-folding dynamics.* Nature, 2002. **420**(6911): p. 102-106.
[8] Oh, K., K.S. Jeong, and J.S. Moore, *Folding-driven synthesis of oligomers.* Nature, 2001. **414**(6866): p. 889-893.
[9] Shakhnovich, E., V. Abkevich, and O. Ptitsyn, *Conserved residues and the mechanism of protein folding.* Nature, 1996. **379**(6560): p. 96-98.
[10] Betz, S.F., et al., *Crystallization of a designed peptide from a molten globule ensemble.* Folding & Design, 1996. **1**(1): p. 57-64.
[11] Koide, S., et al., *Design of single-layer beta-sheets without a hydrophobic core.* Nature, 2000. **403**(6768): p. 456-460.
[12] Minor, D.L. and P.S. Kim, *Context-dependent secondary structure formation of a designed protein sequence.* Nature, 1996. **380**(6576): p. 730-734.
[13] Robertson, D.E., et al., *DESIGN AND SYNTHESIS OF MULTI-HEME PROTEINS.* Nature, 1994. **368**(6470): p. 425-431.
[14] Miller, W.T. and D.P. Raleigh, *Protein folding: From basic science to biotechnology.* Genetic Analysis-Biomolecular Engineering, 1996. **12**(5-6): p. 169-172.
[15] Gething, M.J. and J. Sambrook, *PROTEIN FOLDING IN THE CELL.* Nature, 1992. **355**(6355): p. 33-45.
[16] Carulla, N., et al., *Molecular recycling within amyloid fibrils.* Nature, 2005. **436**(7050): p. 554-558.
[17] Levy, E.D., et al., *Assembly reflects evolution of protein complexes.* Nature, 2008. **453**(7199): p. 1262-U66.
[18] Egger, G., et al., *Epigenetics in human disease and prospects for epigenetic therapy.* Nature, 2004. **429**(6990): p. 457-463.
[19] Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator.* Nature, 2008. **451**(7180): p. 846-850.
[20] Conway, G. and G. Toenniessen, *Feeding the world in the twenty-first century.* Nature, 1999. **402**(6761): p. C55-C58.
[21] Kato, M., et al., *Plant biotechnology - Caffeine synthase gene from tea leaves.* Nature, 2000. **406**(6799): p. 956-957.
[22] Schwartz, T.W. and W.L. Hubbell, *Structural biology - A moving story of receptors.* Nature, 2008. **455**(7212): p. 473-474.
[23] Ferguson, N., et al., *Structural biology - Analysis of 'downhill' protein folding.* Nature, 2007. **445**(7129): p. E14-E15.
[24] Zhou, Z. and Y.W. Bai, *Structural biology - Analysis of protein-folding cooperativity.* Nature, 2007. **445**(7129): p. E16-E17.
[25] Economou, A., *Structural biology - Clamour for a kiss.* Nature, 2008. **455**(7215): p. 879-880.
[26] Stefani, M., *Protein Folding and Misfolding on Surfaces.* International Journal of Molecular Sciences, 2008. **9**(12): p. 2515-2542.
[27] Cha, J.N., et al., *Biomimetic synthesis of ordered silica structures mediated by block copolypeptides.* Nature, 2000. **403**(6767): p. 289-292.
[28] Kloxin, A.M. and K.S. Anseth, *Materials science - Protein gels on the move.* Nature, 2008. **454**(7205): p. 705-706.
[29] Meldrum, F.C., et al., *SYNTHESIS OF INORGANIC NANOPHASE MATERIALS IN SUPRAMOLECULAR PROTEIN CAGES.* Nature, 1991. **349**(6311): p. 684-687.
[30] Howard, J., *Molecular motors: structural adaptations to cellular functions.* Nature, 1997. **389**(6651): p. 561-567.
[31] Schliwa, M. and G. Woehlke, *Molecular motors.* Nature, 2003. **422**(6933): p. 759-765.
[32] Brockwell, D.J., D.A. Smith, and S.E. Radford, *Protein folding mechanisms: new methods and emerging ideas.* Current Opinion in Structural Biology, 2000. **10**(1): p. 16-25.
[33] Radford, S.E., *Protein folding: progress made and promises ahead.* Trends in Biochemical Sciences, 2000. **25**(12): p. 611-618.
[34] Chiti, F., et al., *Rationalization of the effects of mutations on peptide and protein aggregation rates.* Nature, 2003. **424**(6950): p. 805-808.
[35] Serrano, L., et al., *EFFECT OF ALANINE VERSUS GLYCINE IN ALPHA-HELICES ON PROTEIN STABILITY.* Nature, 1992. **356**(6368): p. 453-455.
[36] Negrete, J. and P.P. Gonzalez, *A Net of Multi-Agent Expert Systems with Emergent Control.* Expert Systems With Applications, 1998. **14**(1-2).

[37] Wooldridge, M., *The Gaia methodology for agent-oriented analysis and design.* Autonomous Agents and Multi-Agent Systems, 2000. **3**(3): p. 285-312.
[38] Liu, H., M.X. Tang, and J.H. Frazer, *Supporting evolution in a multi-agent cooperative design environment.* Advances in Engineering Software, 2002. **33**(6): p. 319-328.
[39] Wooldridge, M.J., *Software engineering with agents: Pitfalls and pratfalls.* IEEE Internet Computing, 1999. **3**(3): p. 20-+.
[40] Gallimore, R.J., et al., *Cooperating agents for 3-D scientific data interpretation.* Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 1999. **29**(1): p. 110-126.
[41] Gonzalez, P.P., et al., *Cellulat: an agent-based intracellular signalling model.* Biosystems, 2003. **68**(2-3): p. 171-185.
[42] Ominici, A., A. Ricci, and M. Viroli, *Coordination artifacts: Environment-based coordination for intelligent agents.* AAMAS'04, 2006.
[43] Ominici, A., et al., *Coordination artifacts: Environment-based coordination for intelligen agents.* Proceedings of 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems, 2004: p. 286-293.
[44] Gonzalez, P.P. and J. Negrete, *REDSIEX: A Cooperative Network of Expert Systems with Blackboard Architectures.* Expert Systems, 1997. **14**(4): p. 180-189.
[45] Lagunez-Otero, J., et al., *Cellulat*, in *Artificial Life VIII: Proceedings of the Eight International Conference on Artificial Life*, R.K. Standish, M.A. Bedau, and H.A. Abbass, Editors. 2002: Sydney, Australia. p. 97-100.
[46] Dill, K.A., *THEORY FOR THE FOLDING AND STABILITY OF GLOBULAR-PROTEINS.* Biochemistry, 1985. **24**(6): p. 1501-1509.
[47] Dill, K.A., *Polymer principles and protein folding.* Protein Science, 1999. **8**(6): p. 1166-1180.
[48] Dill, K.A., et al., *PRINCIPLES OF PROTEIN-FOLDING - A PERSPECTIVE FROM SIMPLE EXACT MODELS.* Protein Science, 1995. **4**(4): p. 561-602.
[49] Goldberg, D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*. 1989: Addison-Wesley Professional.
[50] Lima, C., *Parameter Setting in Evolutionary Algorithms*. 2007, Berlin: Springer.
[51] Mitchell, M., *An Introduction to Genetic Algorithms (Complex Adaptive Systems)*. 1998: The MIT Press.
[52] Panchenko, A.R., Z. Luthey-Schulten, and P.G. Wolynes, *Foldons, protein structural modules, and exons.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(5): p. 2008-2013.
[53] Ferreiro, D.U., et al., *The energy landscapes of repeat-containing proteins: Topology, cooperativity, and the folding funnels of one-dimensional architectures.* Plos Computational Biology, 2008. **4**(5).
[54] Englander, S.W., L. Mayne, and M.M.G. Krishna, *Protein folding and misfolding: mechanism and principles.* Quarterly Reviews of Biophysics, 2007. **40**(4): p. 287-326.
[55] Bedard, S., et al., *Protein folding: Independent unrelated pathways or predetermined pathway with optional errors.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(20): p. 7182-7187.
[56] Lindberg, M.O. and M. Oliveberg, *Malleability of protein folding pathways: a simple reason for complex behaviour.* Current Opinion in Structural Biology, 2007. **17**(1): p. 21-29.
[57] Olofsson, M., et al., *Folding of S6 structures with divergent amino acid composition: Pathway flexibility within partly overlapping foldons.* Journal of Molecular Biology, 2007. **365**(1): p. 237-248.