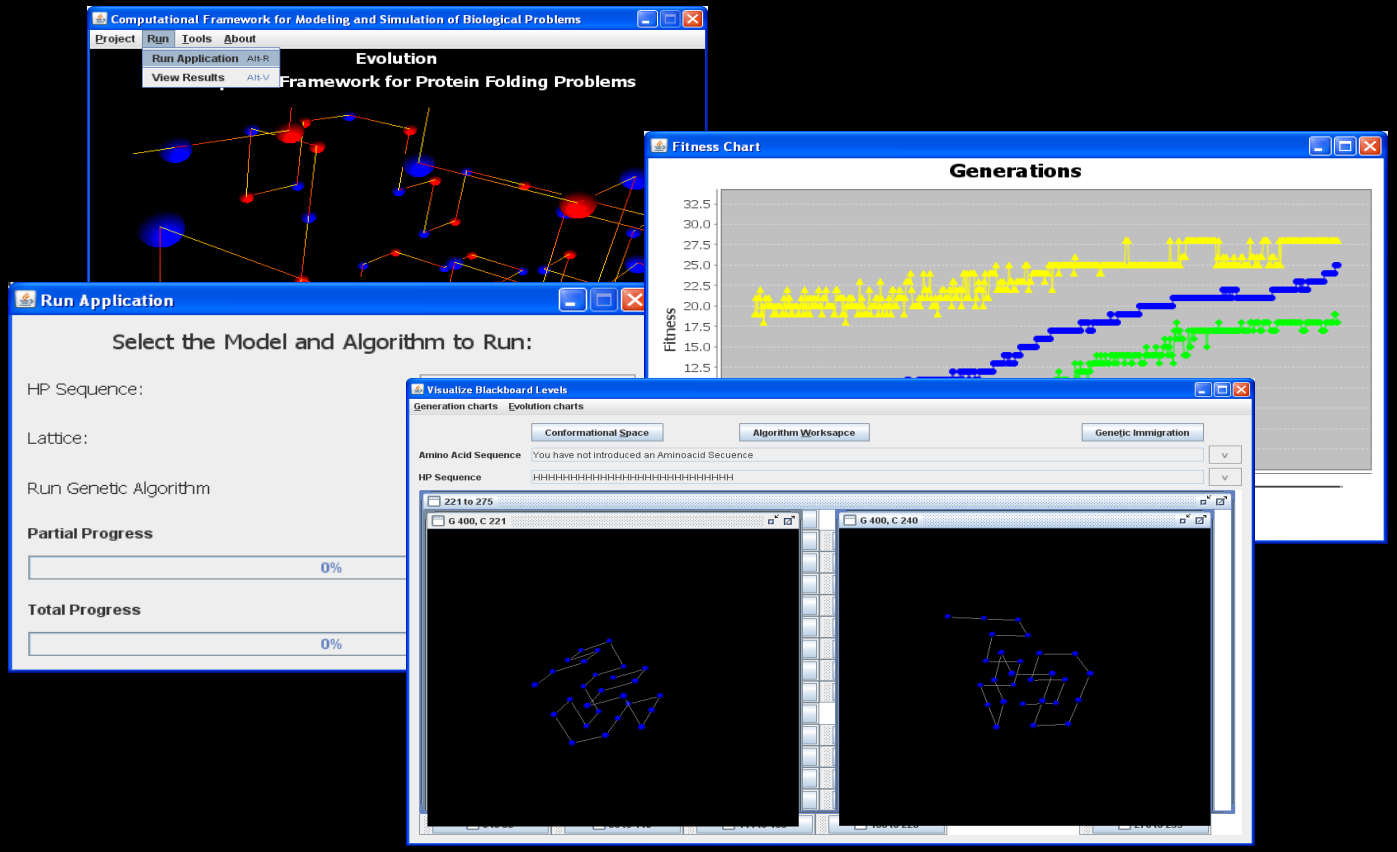
Un approccio multi-agente / sistemi blackboard per l'esplorazione di problemi complesi in biologia strutturale

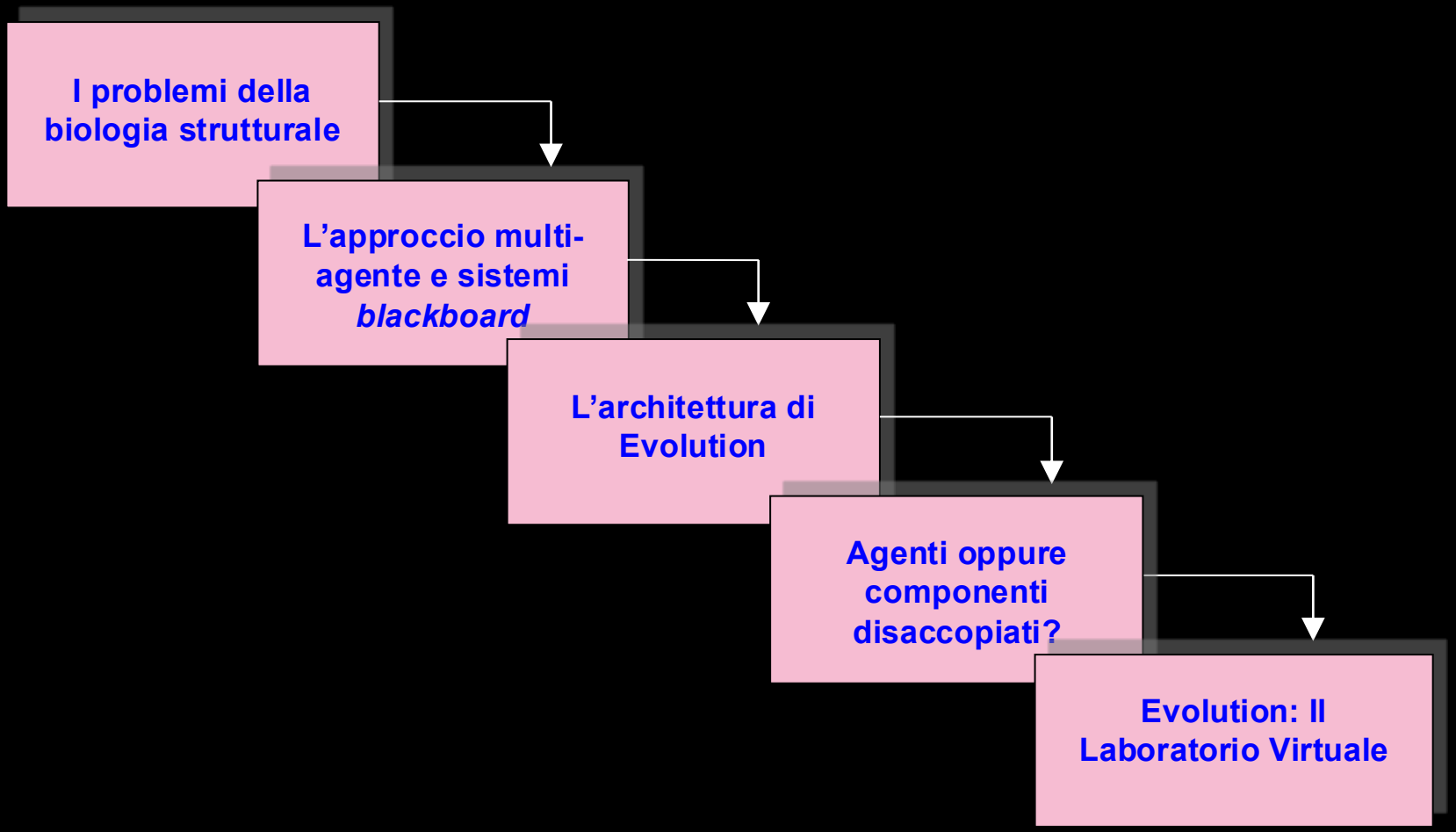
(An adaptable blackboard-based multiagent framework to explore complex problems in structural biology)

González-Pérez, P.P., Rojo-Domínguez, A., Beltrán, H., and Sánchez, M.

Universidad Autónoma Metropolitana, México



Struttura della presentazione



Biologia strutturale

La biologia strutturale è una branca della ***biologia*** molecolare, biochimica e biofisica concernente con la struttura molecolare di macromolecole biologiche – in particolare proteine e acidi nucleici.

Principali questioni in biologia strutturale

- Come queste macromolecole acquistano la struttura che hanno?
- Come alterazioni nella loro struttura influenzano la loro funzione?

Problemi complesi in biologia strutturale

Folding

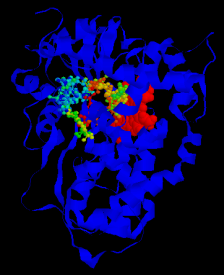
meaiakydfk ataddelsfk
... qtgmfprnyv tpvnrnv

Inverse folding

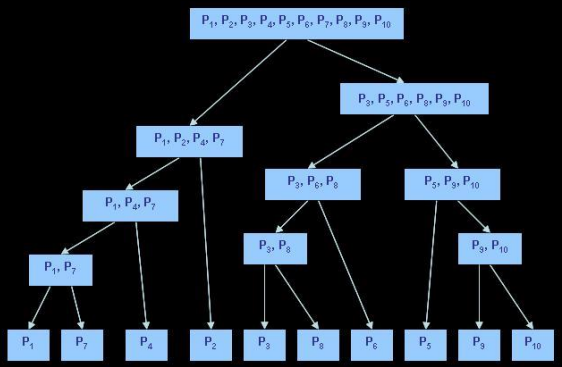
mteyk…vg … iqliqnhfvd
... kskakcvvm

mte… aggvgksalt
iqliqnhfvd ... ksk.. vvm

Docking



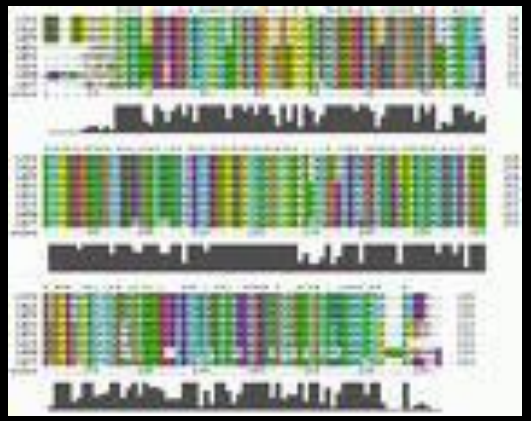
Clustering



Structure-activity and structure-property relationships

$$activity = f(structure)$$
$$activity = c_0 + c_1 MD_1 + c_2 MD_2 + \cdots + c_n MD_n$$
$$property = f(structure)$$
$$property = c_0 + c_1 MD_1 + c_2 MD_2 + \cdots + c_n MD_n$$

Sequence alignment



Il ripiegamento delle proteine (the protein folding problem)

Ripiegamento delle proteine:

- E' il meccanismo attraverso il quale le proteine raggiungono la loro conformazione biologicamente attiva dopo la biosintesi.
- Il ripiegamento delle proteine è essenzialmente un problema di ricerca dellla conformazione biologicamente attiva di una proteina (il cosiddetto stato nativo), per una determinata sequenza di residui di aminoacidi.

meaiakydfk ataddelsfk ... qtgmfprnyv tpvnrnv

Grb2

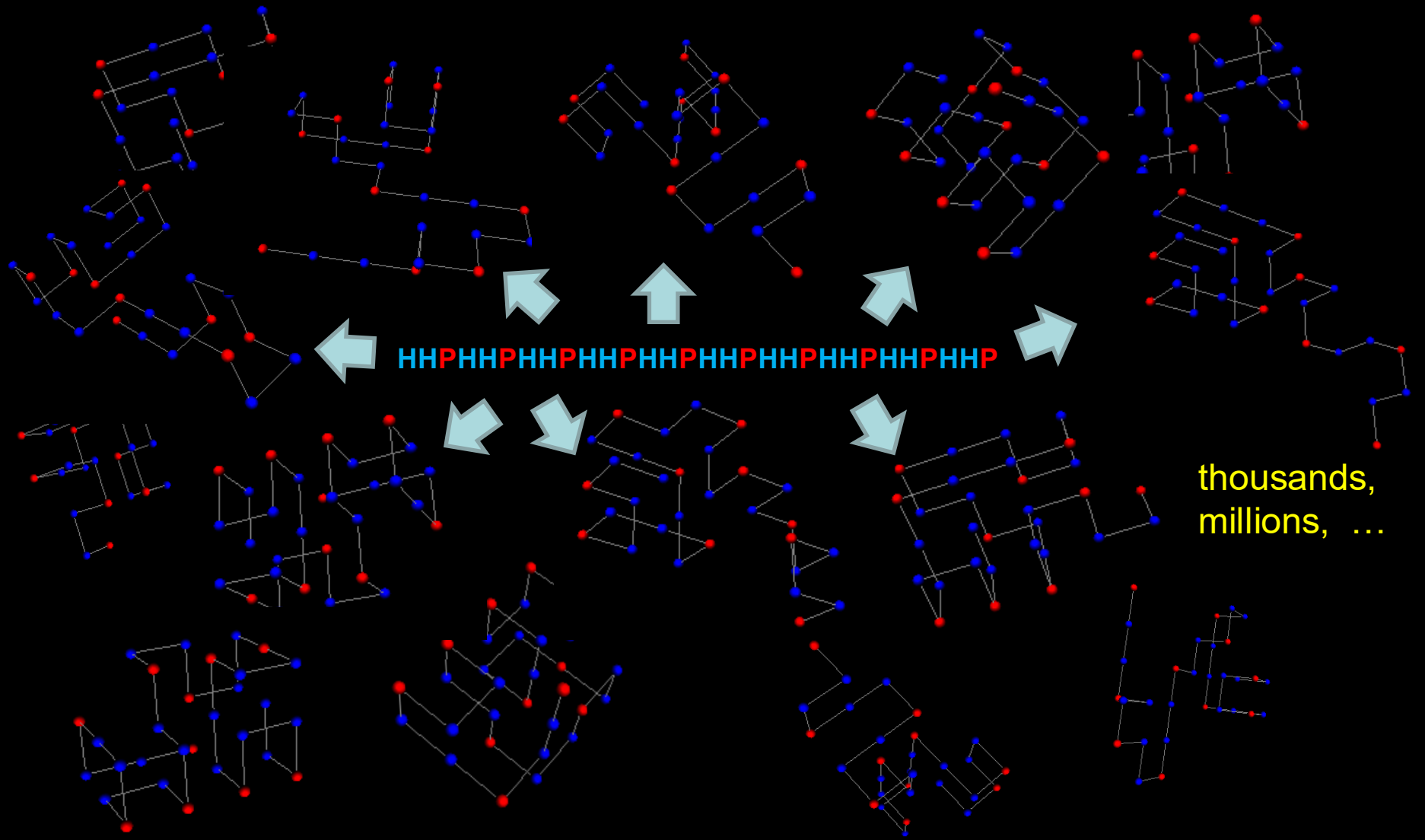
mddsevesta silasvkeqe drsgdlgdme ... plkgaplmqk i

p120

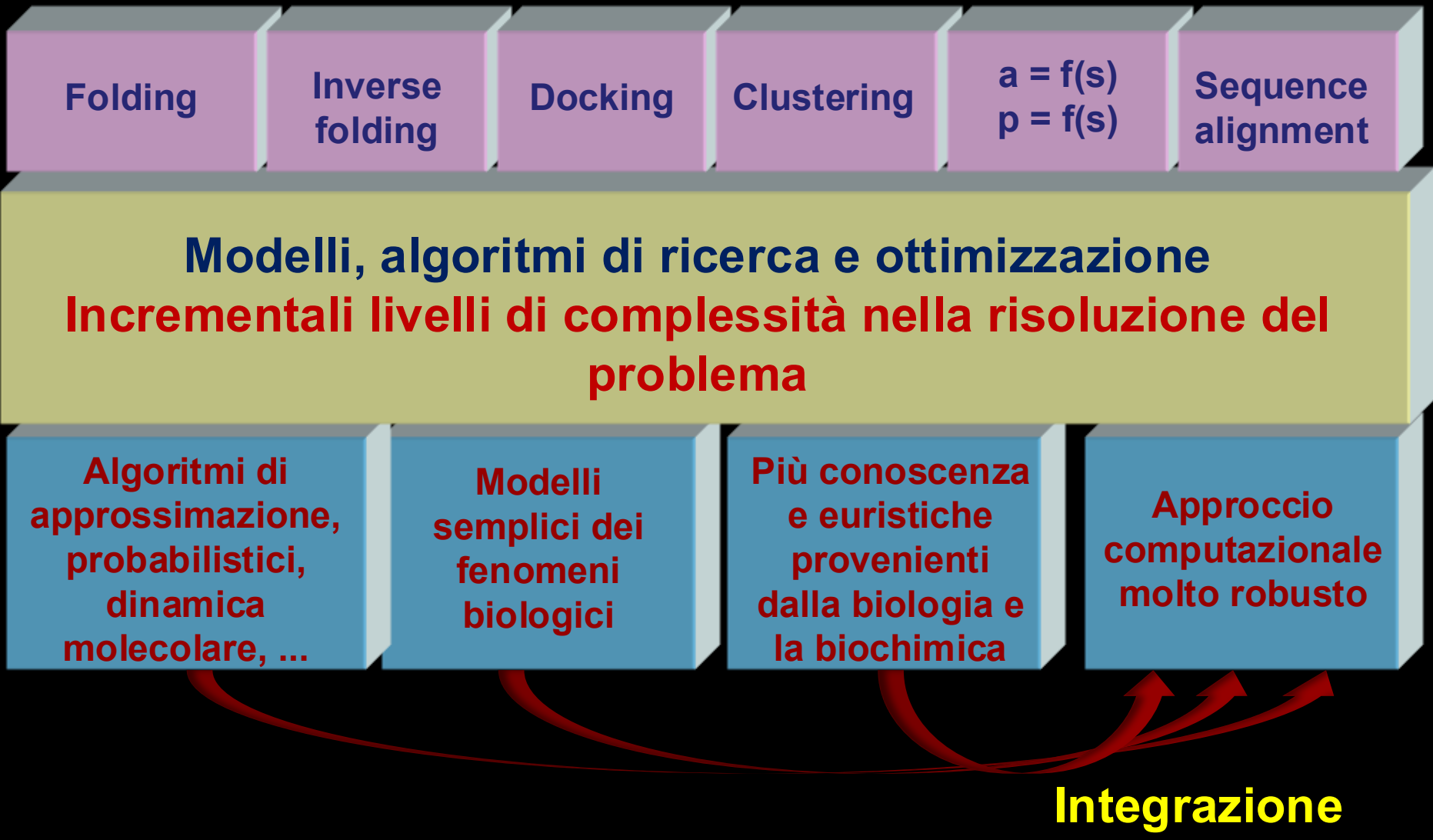
mteyklvvvg aggvgksalt iqliqnhfvd ... kskakcvvm

Ras

Il ripiegamento delle proteina



Cosa c'è di comune fra questi problemi biologici?



La proposta

- Gli approcci computazionali alla soluzione dei problemi di folding, docking e disegno molecolare, per diventare soluzioni robuste ed efficaci, devono integrare e far interagire diversi modelli, algoritmi e strategie che permettano una migliore fruizione dei diversi tipi di conoscenza (biologica, fisica, chimica, molecolare, ecc.) e di abilità (*lattice bead models*, *off-latice bead models*, *united atom models*, algoritmi di approssimazione, probabilistici, euristici, evolutivi, ecc.) che ocorrono per affrontare la gran complessità caratteristica di questi tipi di fenomeni biologici.
- In questo senso, noi postuliamo che un approccio di sistemi multiagenti / sistemi *blackboard* sarebbe in grado di fornire una infrastruttura idonea per raggiungere questo obiettivo.

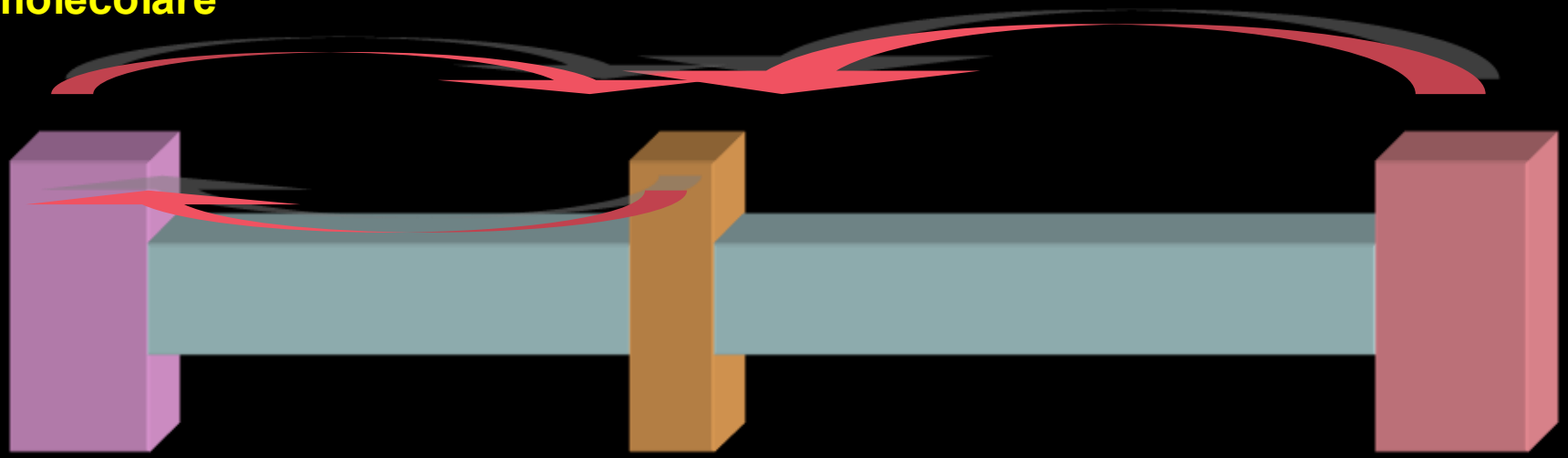
Fra le applicazioni bioinformatiche e i sistemi multi-agenti

Noi siamo qui

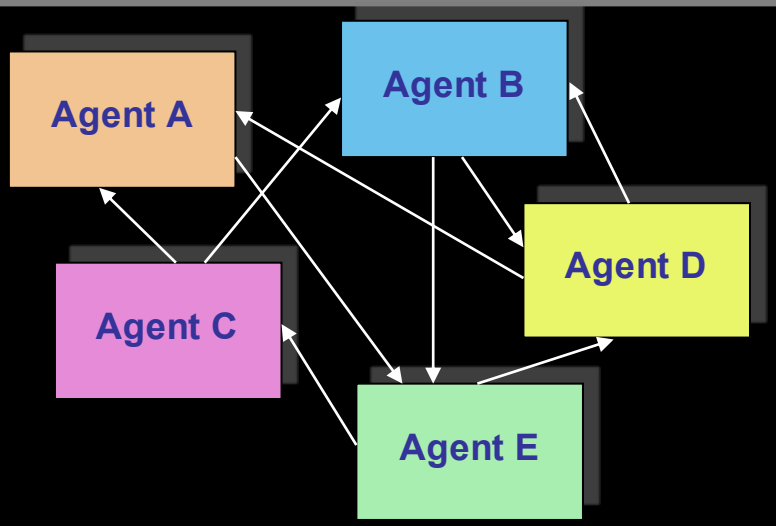
Biologia / Problemi non risolti in biologia molecolare

Sviluppo di applicazioni bioinformatiche

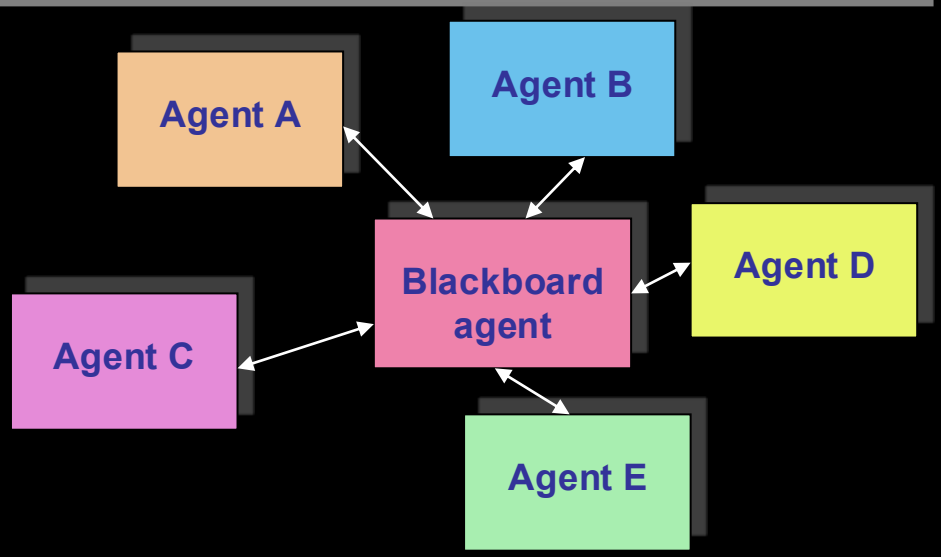
Sistemi multi-agenti e sistemi blackboard (principi e teorie)



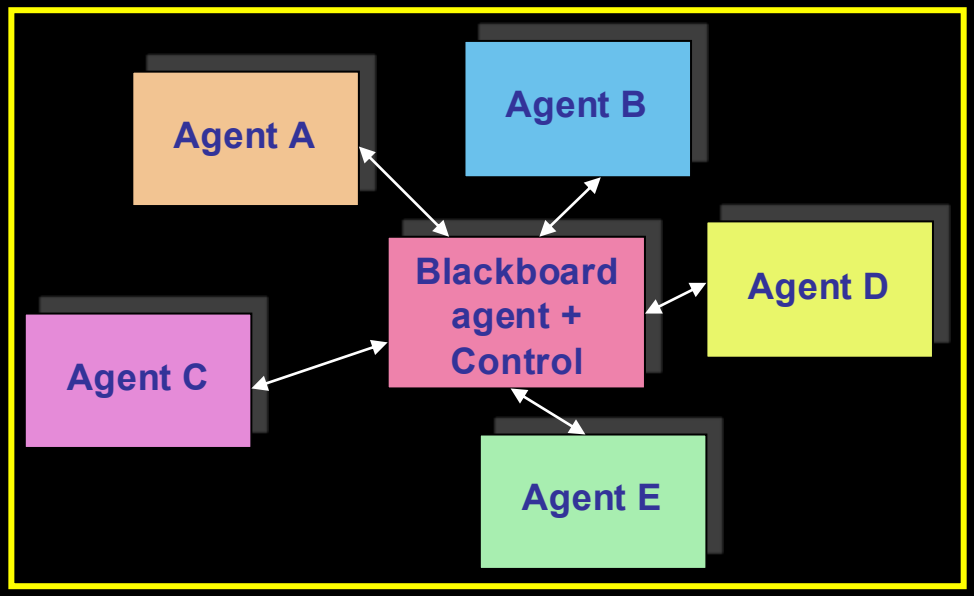
L'approccio multi-agente: Collaborating agents (Corkill, 2003)



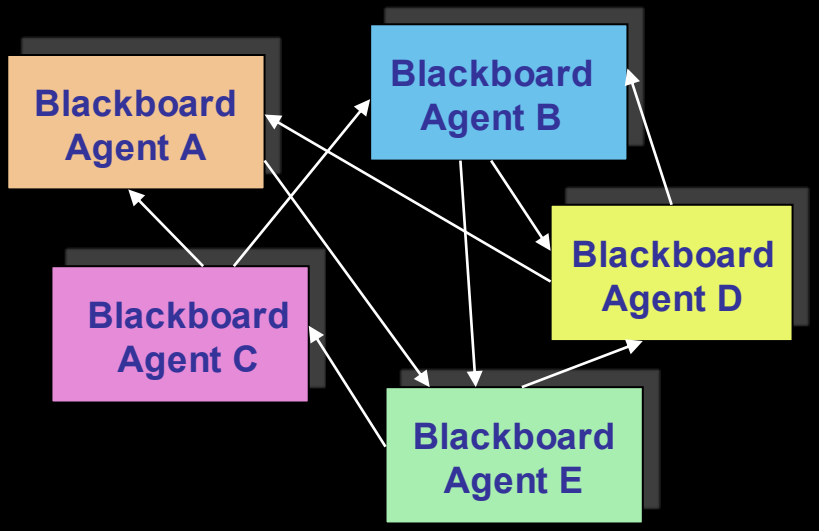
Directly interacting



With a Blackboard agent



With a Blackboard & control agent

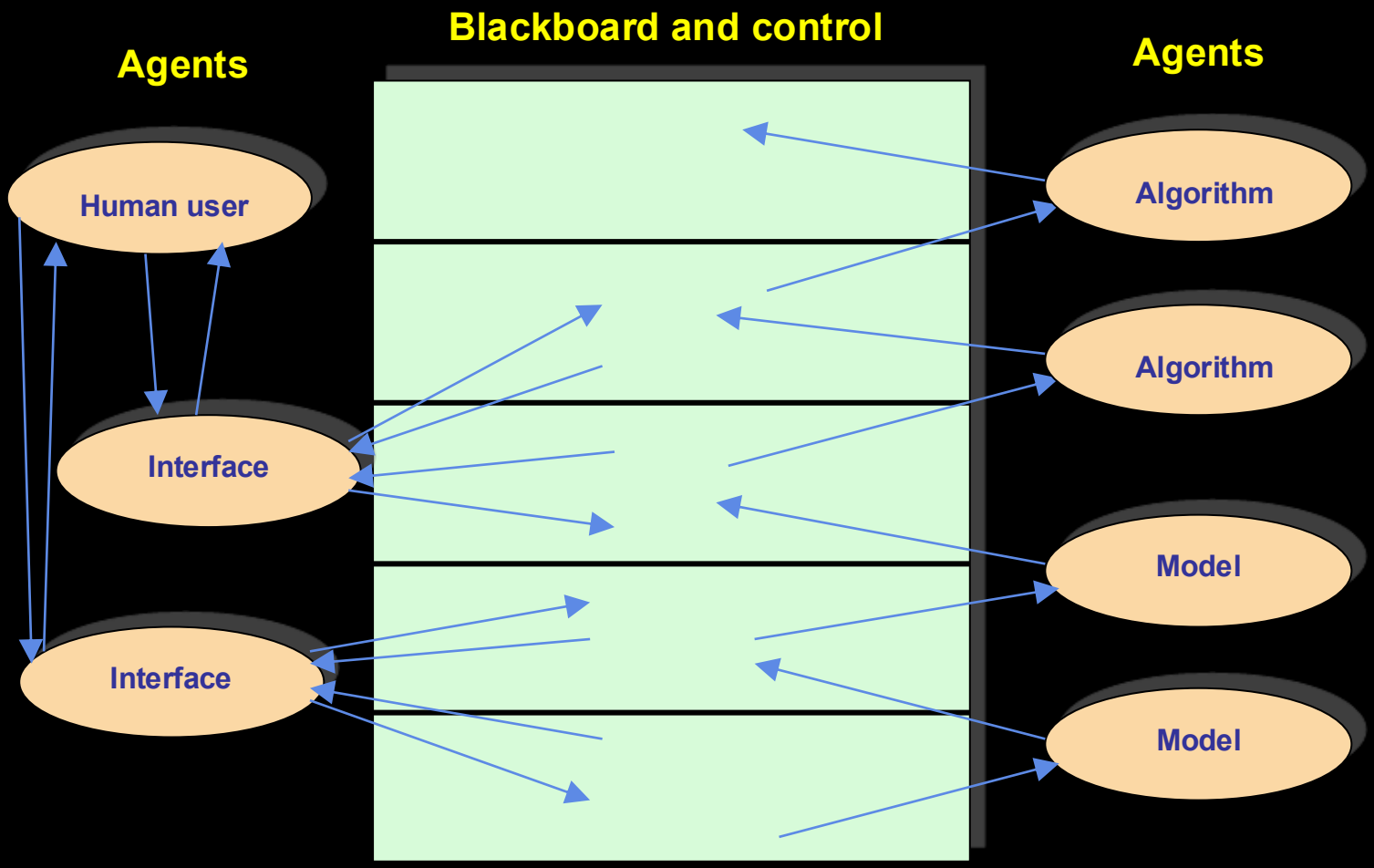


Full-fledged Blackboard agents

L’approccio multi-agente / sistemi blackboard

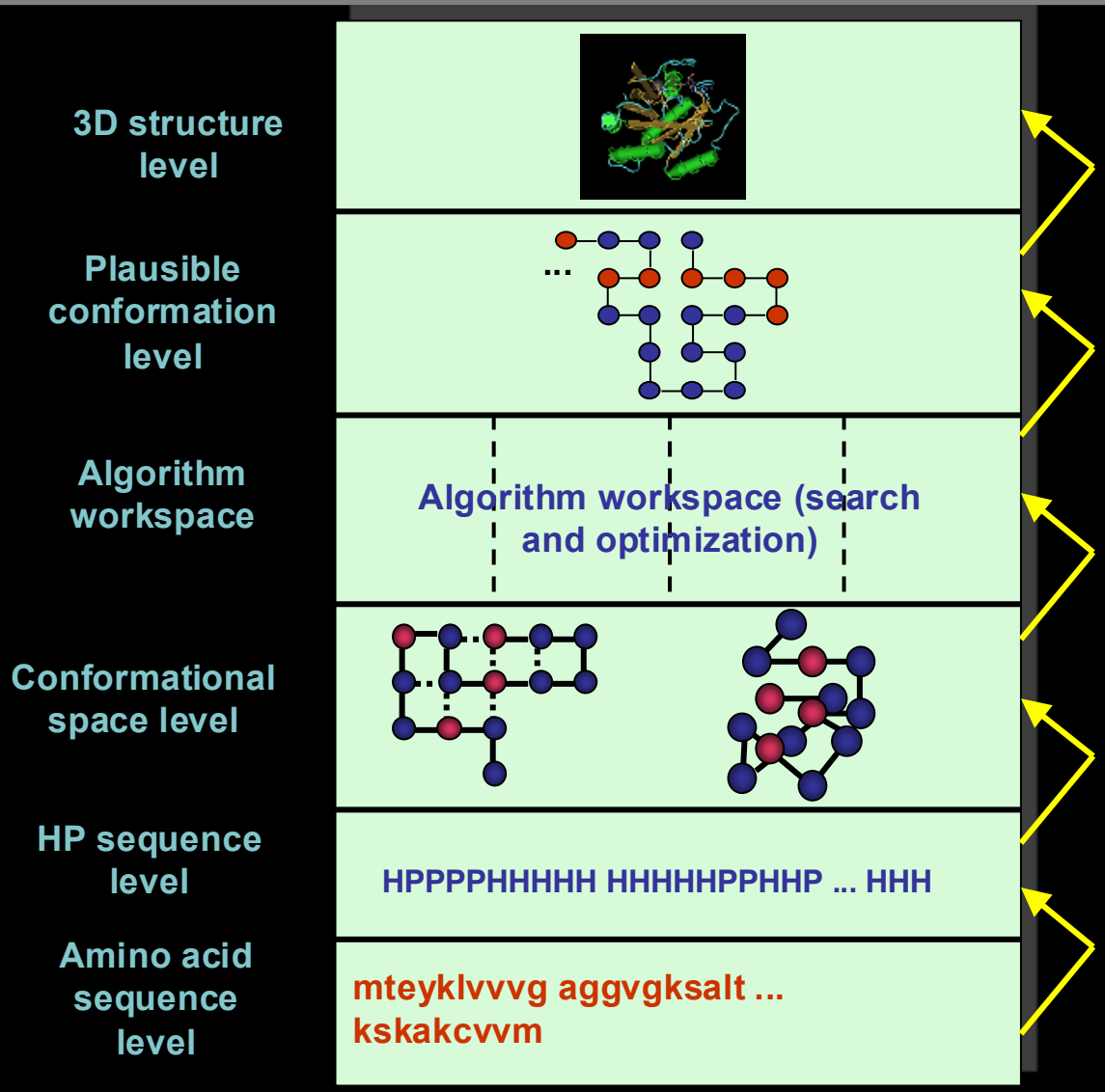
- Questo approccio ha dimostrato di essere un efficace e opportuno livello di astrazione per la costruzione di modelli di tutta una varietà di sistemi biologici (Cellulat, 2003).
- Se un agente rappresenta un singolo componente attivo del sistema (per esempio, proteine o geni), il sistema globale, compreso il *blackboard*, cattura l'insieme delle componenti biologiche comprese anche le strutture coinvolte nella loro interazione (ad esempio, i compartimenti cellulari, messaggeri intracellulari, ...) portando ad una accurata e completa simulazione del sistema sotto studio.
- In particolare, il *blackboard* può essere adottato per modellare i vari pattern di interazione che possono essere trovati nei processi biologici, come lo scambio di informazioni chimiche.

Evolution: architettura

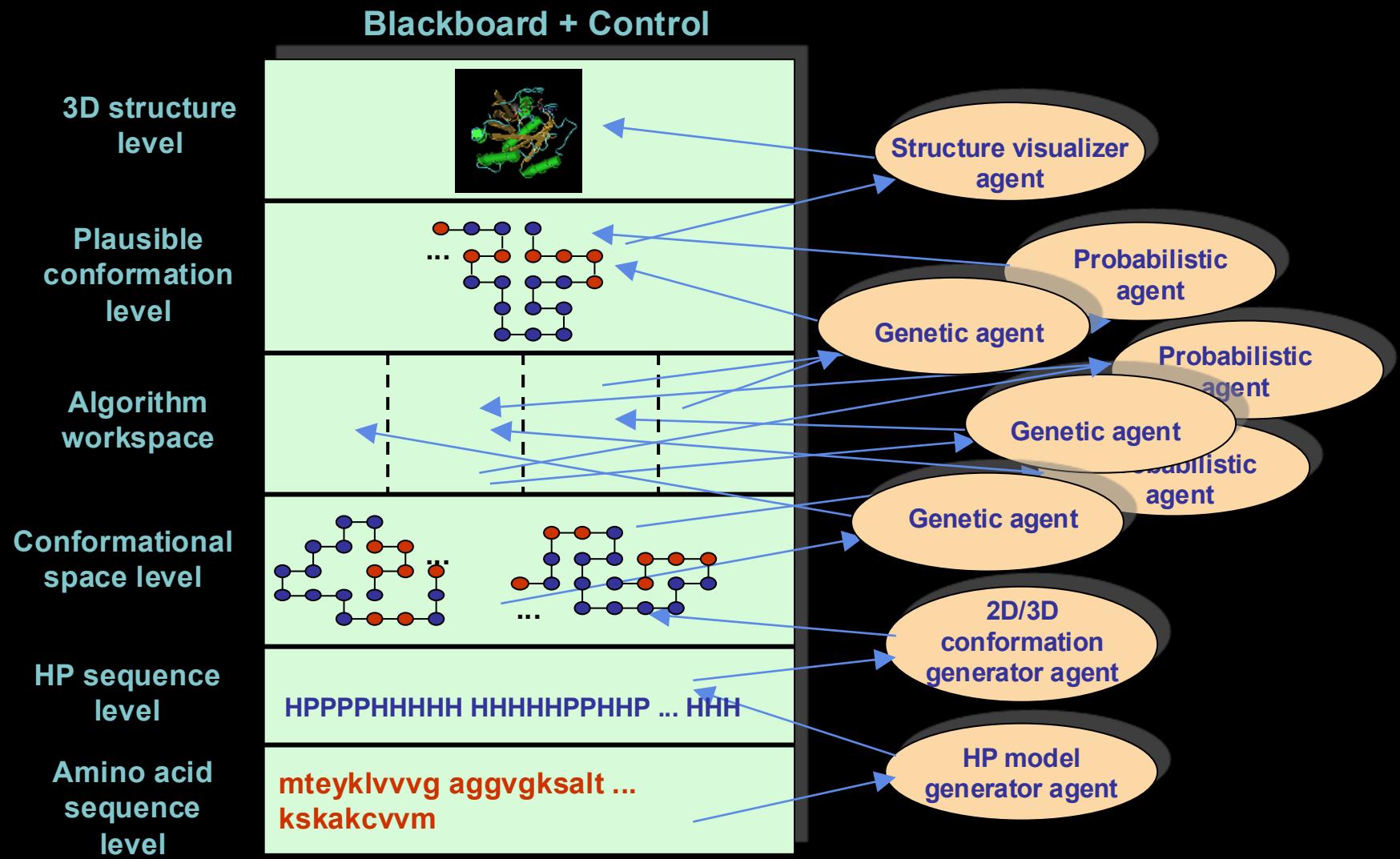


“Agenti” che incapsulano modelli e algoritmi e agenti che stabiliscono l’interfacia con l’utente.

The blackboard levels: the semantics for protein folding problem

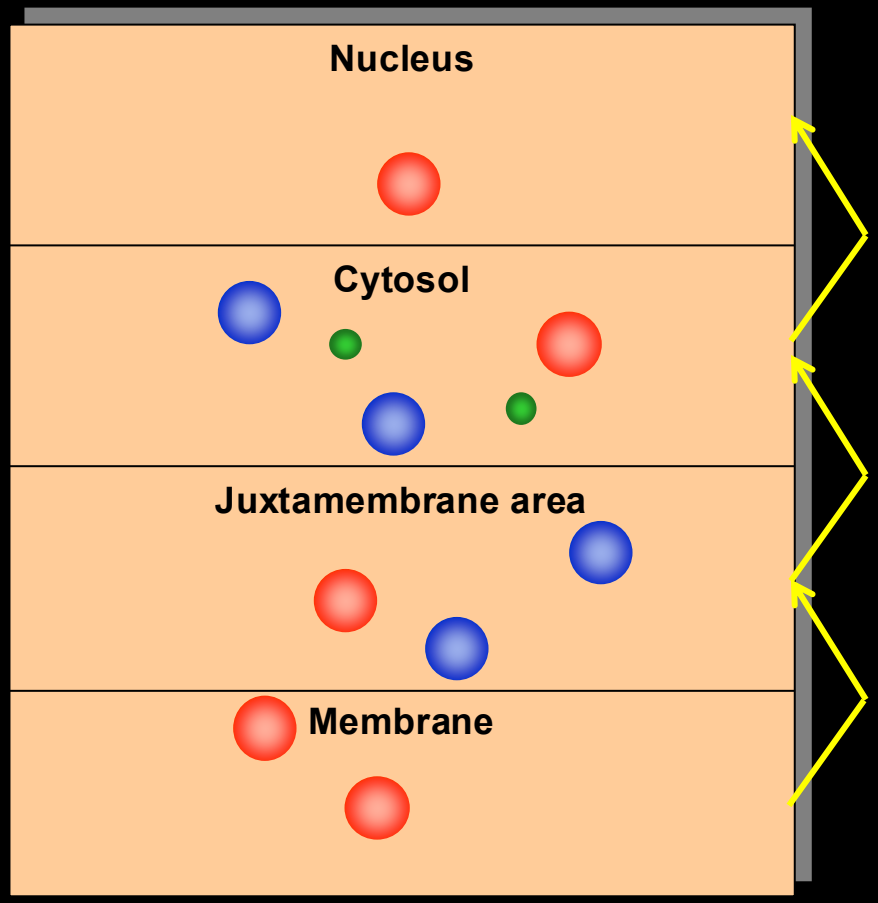


Blackboard-based multi-agent architecture: the semantics for protein folding problem

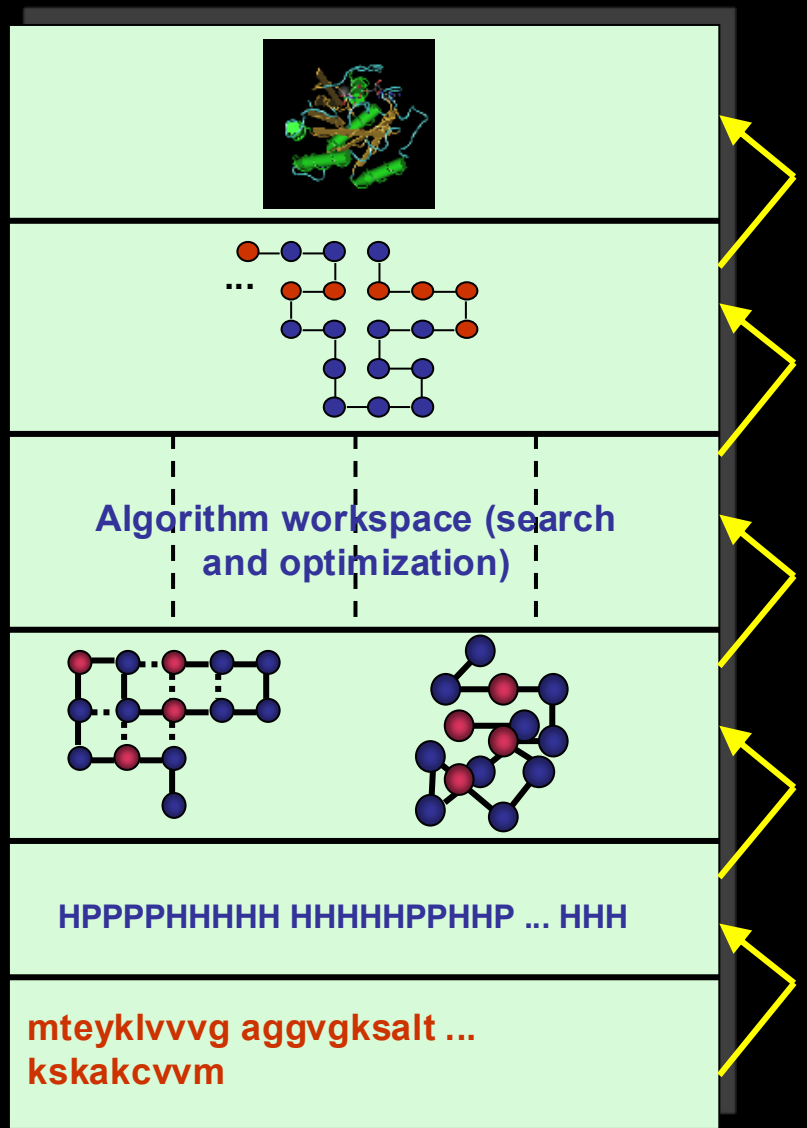


Dal blackboard come struttura biologica al blackboard come *workspace* per i meccanismi e modelli

Cellulat (2003)

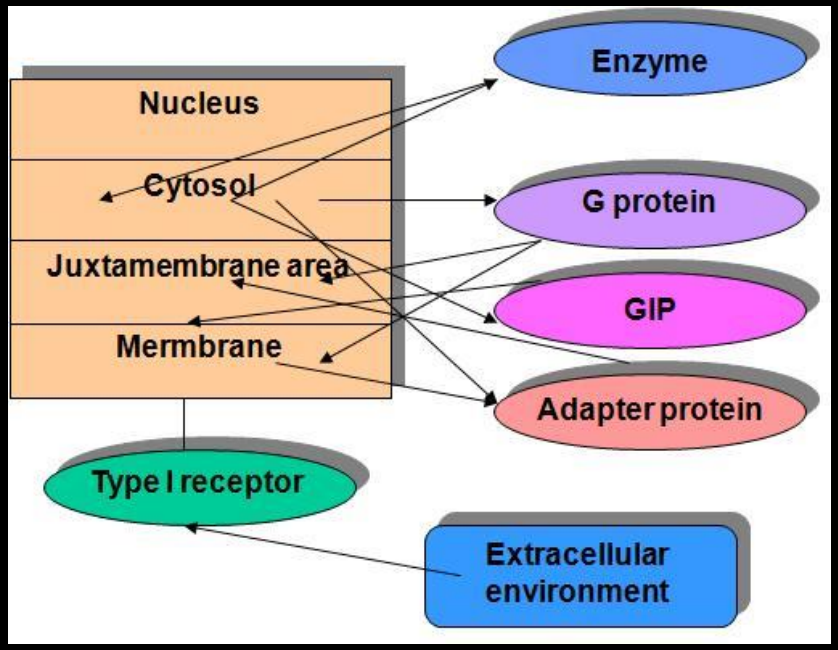


Evolution (2008)

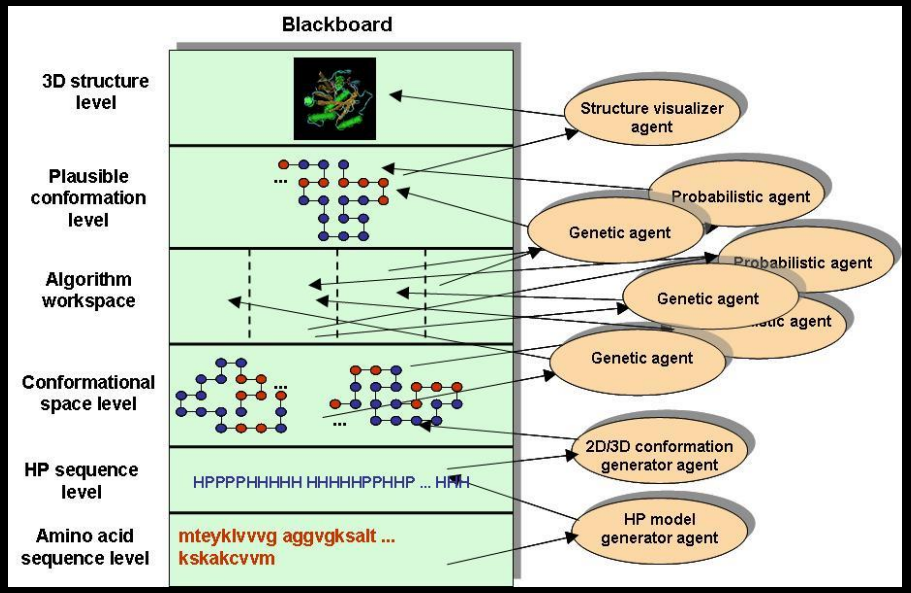


Dagli agenti come componenti biologici agli agenti come repertorio di meccanismi e modelli

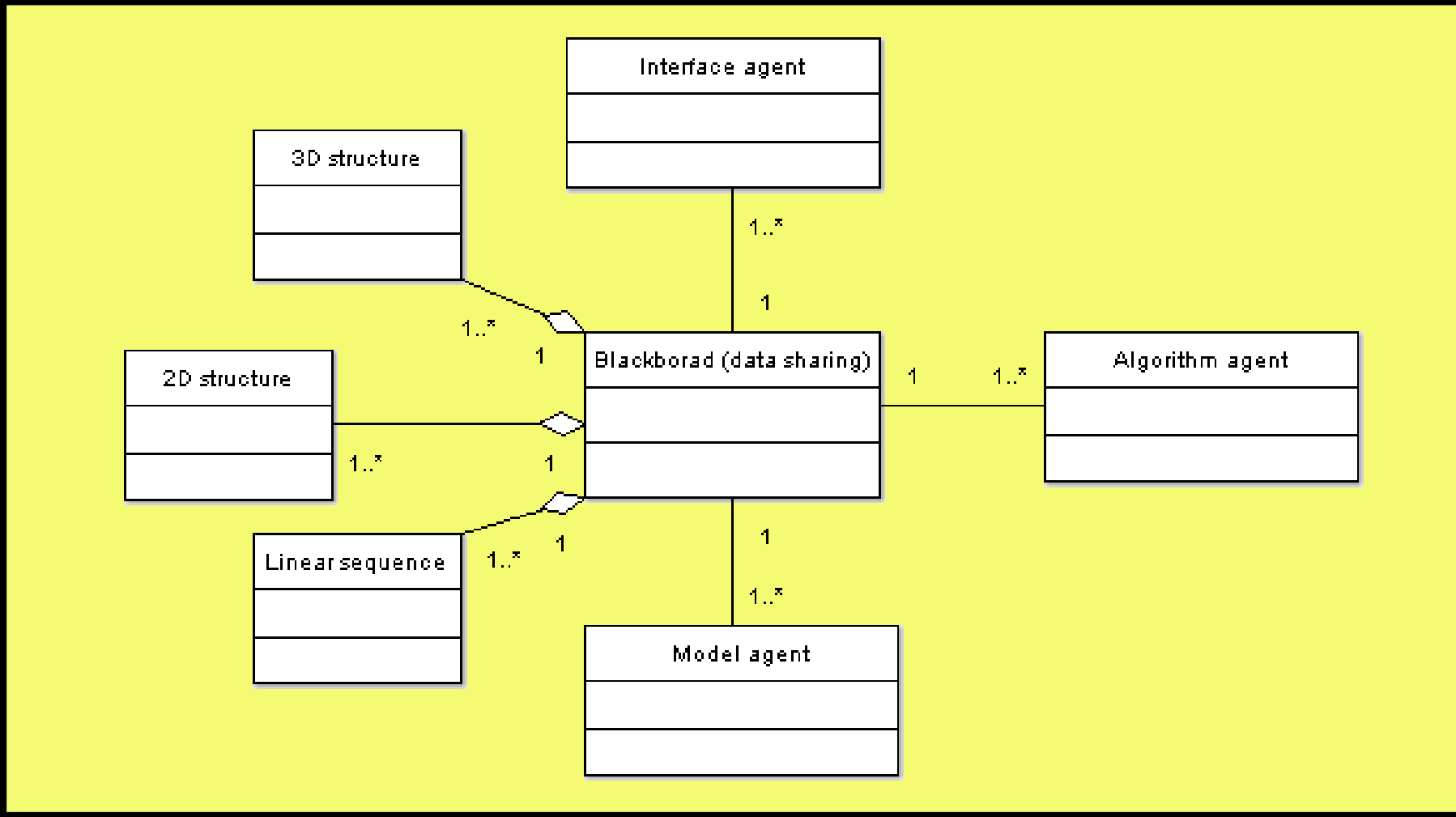
Cellulat (2003)



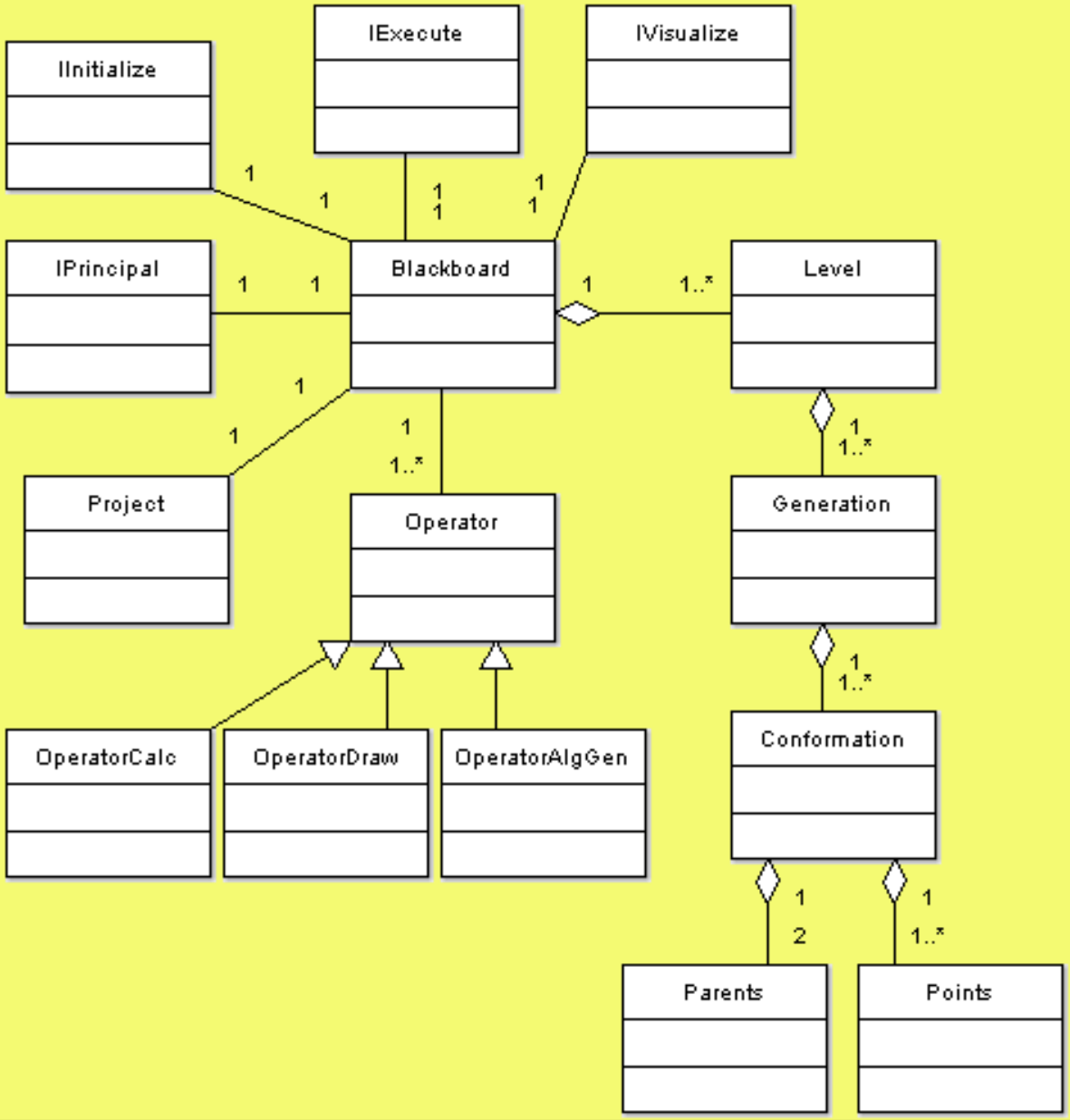
Evolution (2008)



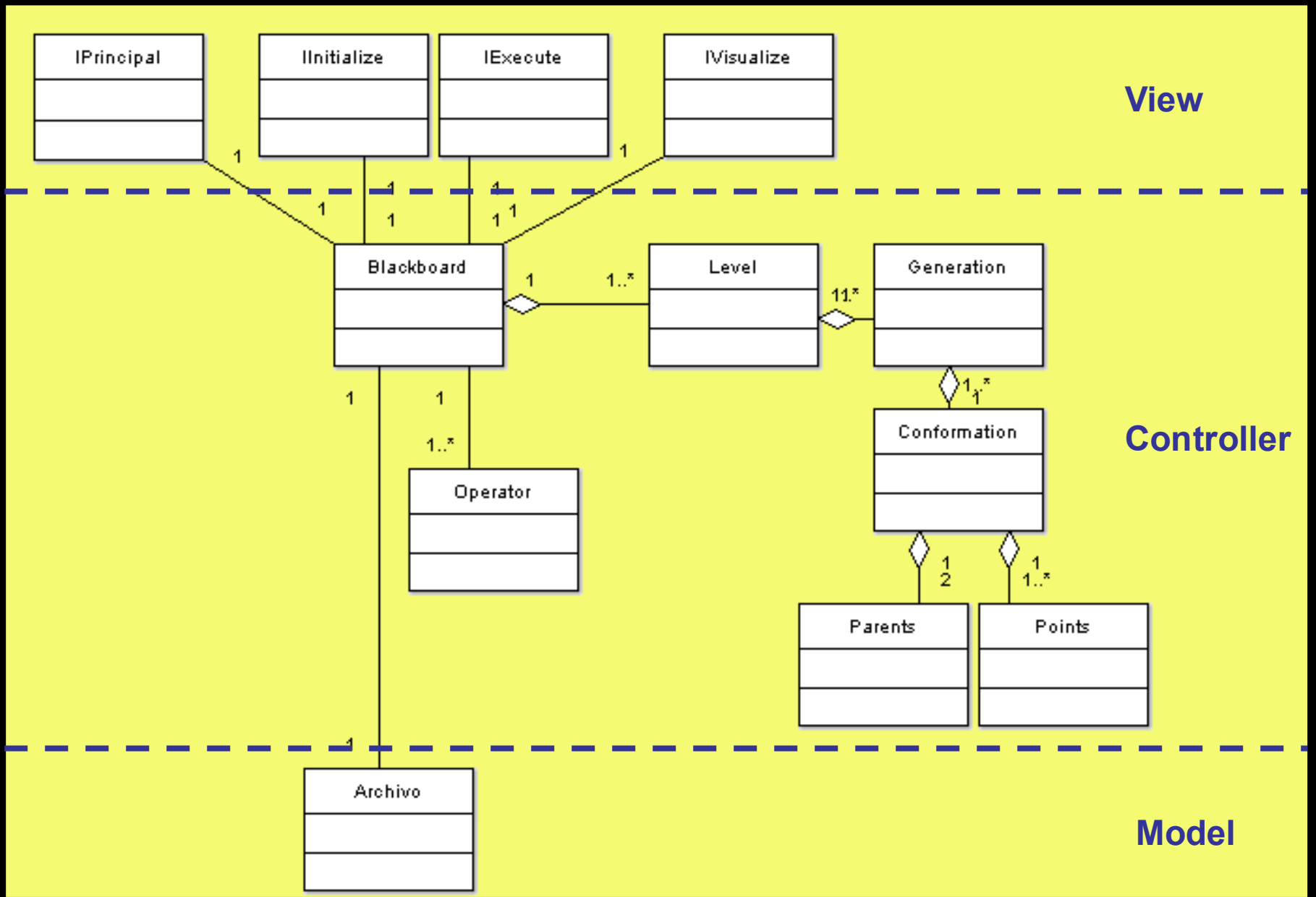
Blackboard-based multi-agent architecture: the architectural pattern



Architectural design

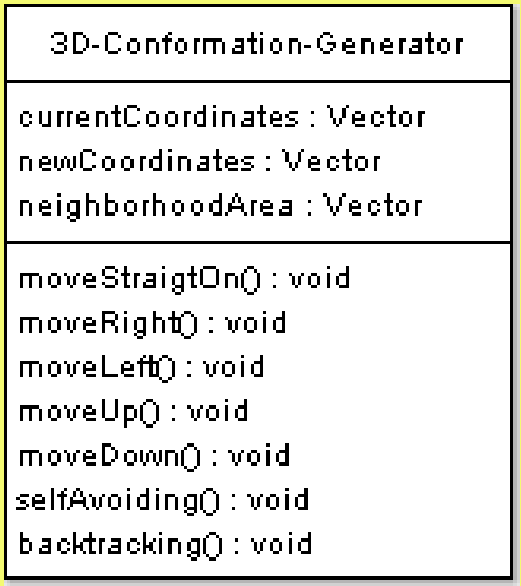


Architectural design

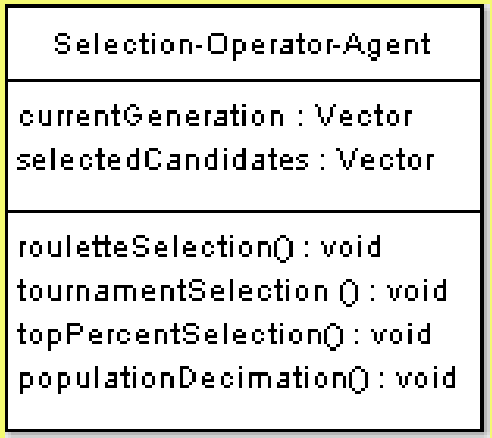


Agenti oppure componenti disaccopiati?

Model agent

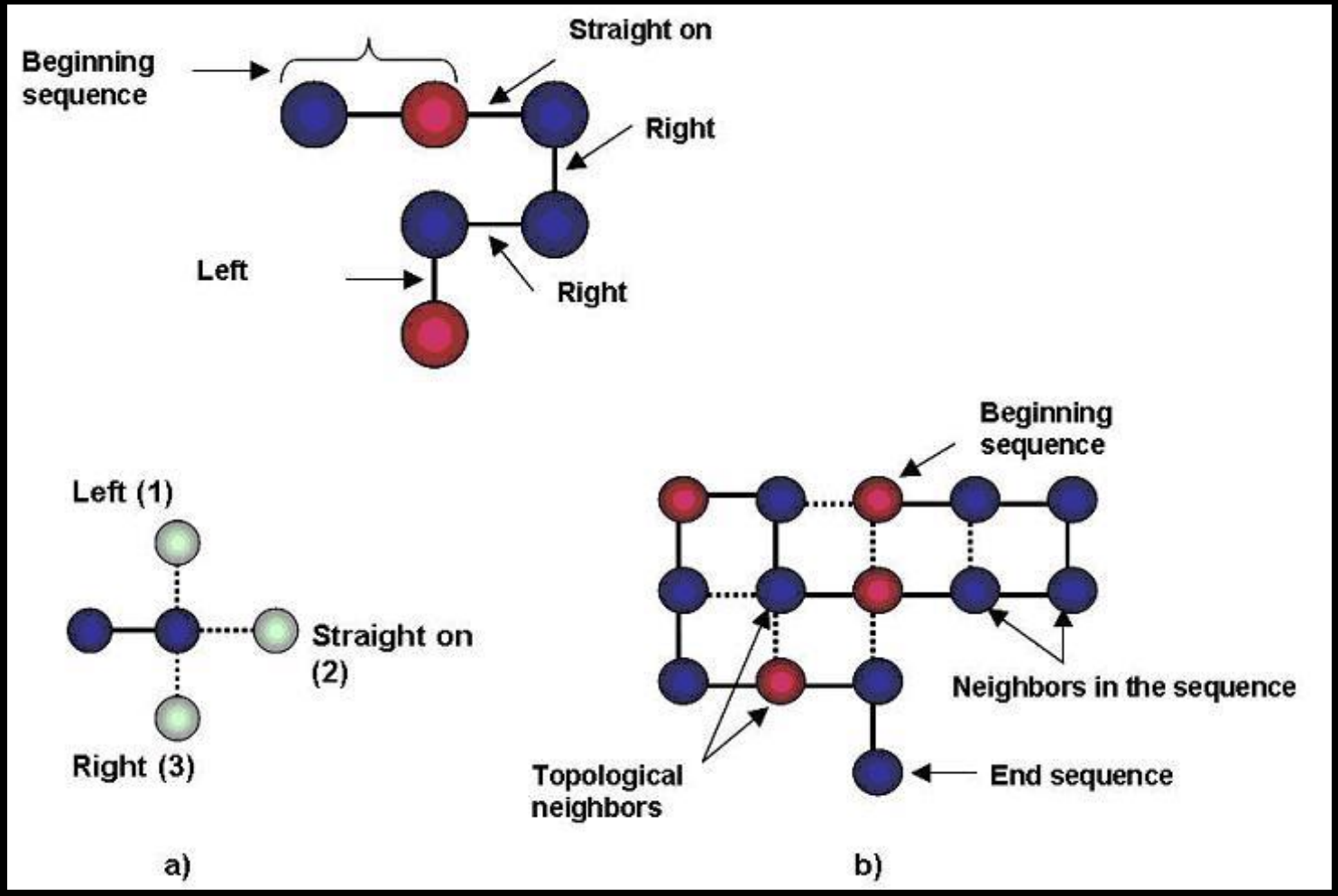


Algorithm agent



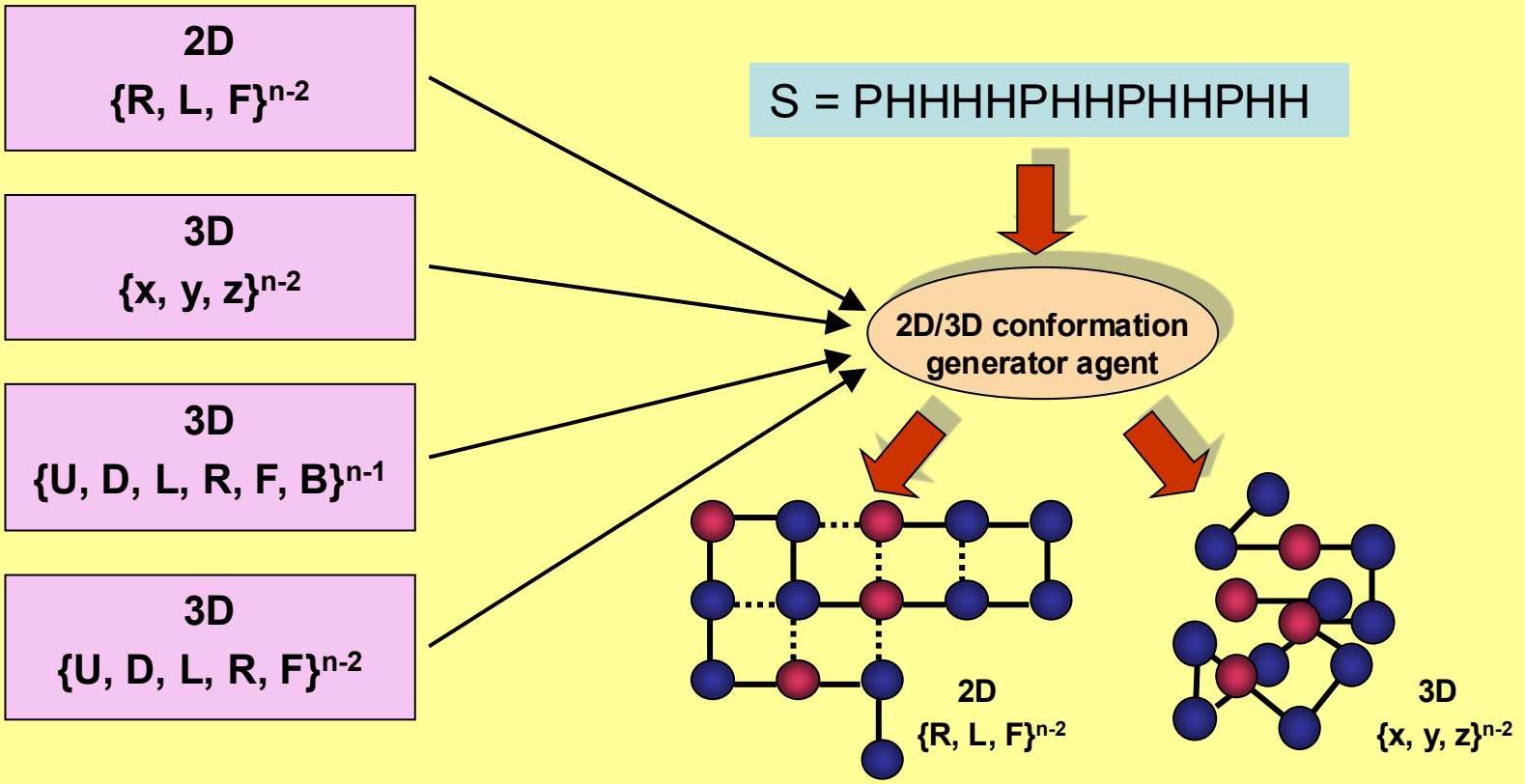
Modelli ed algoritmi incapsulati negli agenti

A 2D space, with a square lattice



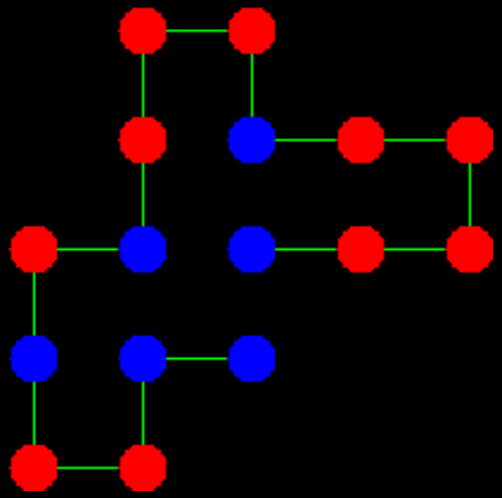
Modelli ed algoritmi incapsulati negli agenti

Local coordinate systems

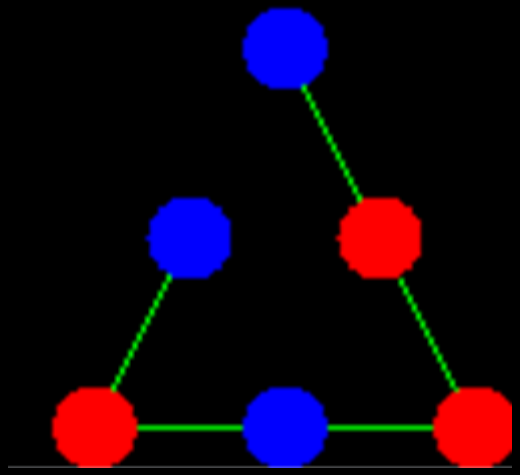


Modelli ed algoritmi incapsulati negli agenti

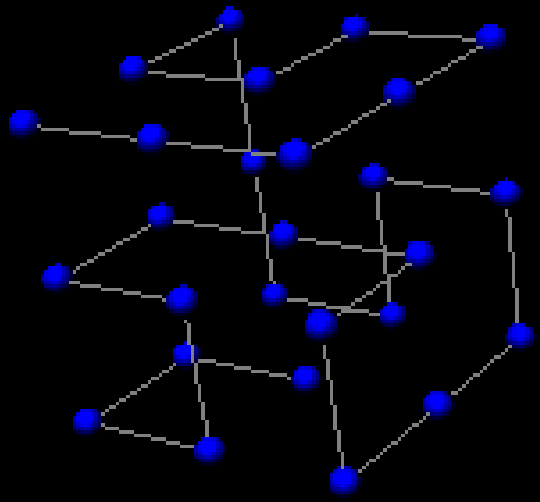
Different types of discrete spaces can be generated in Evolution



2D Squared

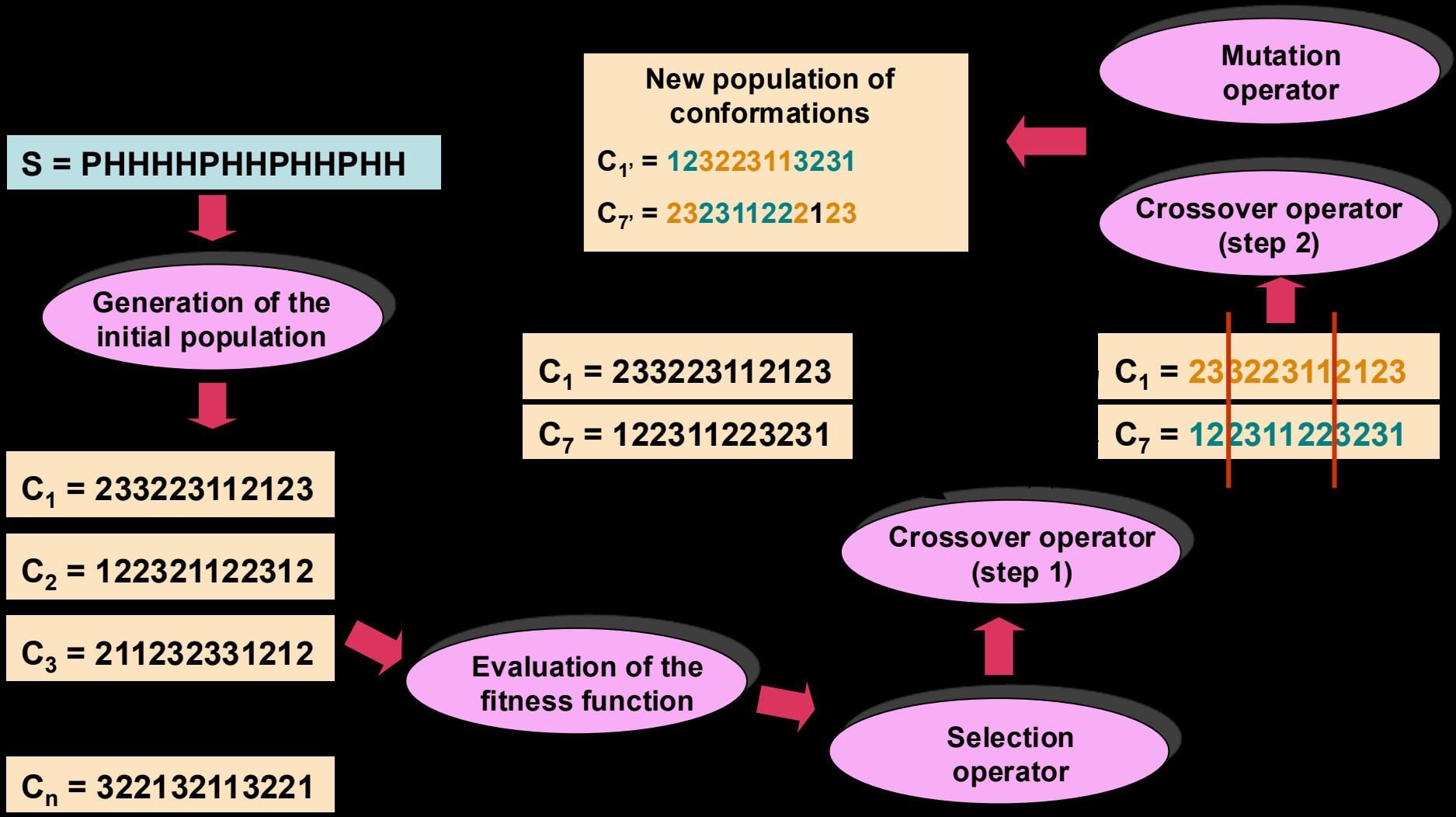


2D Triangular



3D Cubic

Modelli ed algoritmi incapsulati negli agenti



Modelli ed algoritmi incapsulati negli agenti

Model 1: the traditional model

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

$$e_{HH} = -1.0 \quad e_{HP} = 0.0 \; e_{PP} = 0.0$$

$$\Delta_{ij} = \begin{cases} 1 \; if \;\; i \; and \;\; j \; are \; topological, \; but \; not \; sequence \; neighbours \\ 0 \; otherwise \end{cases}$$

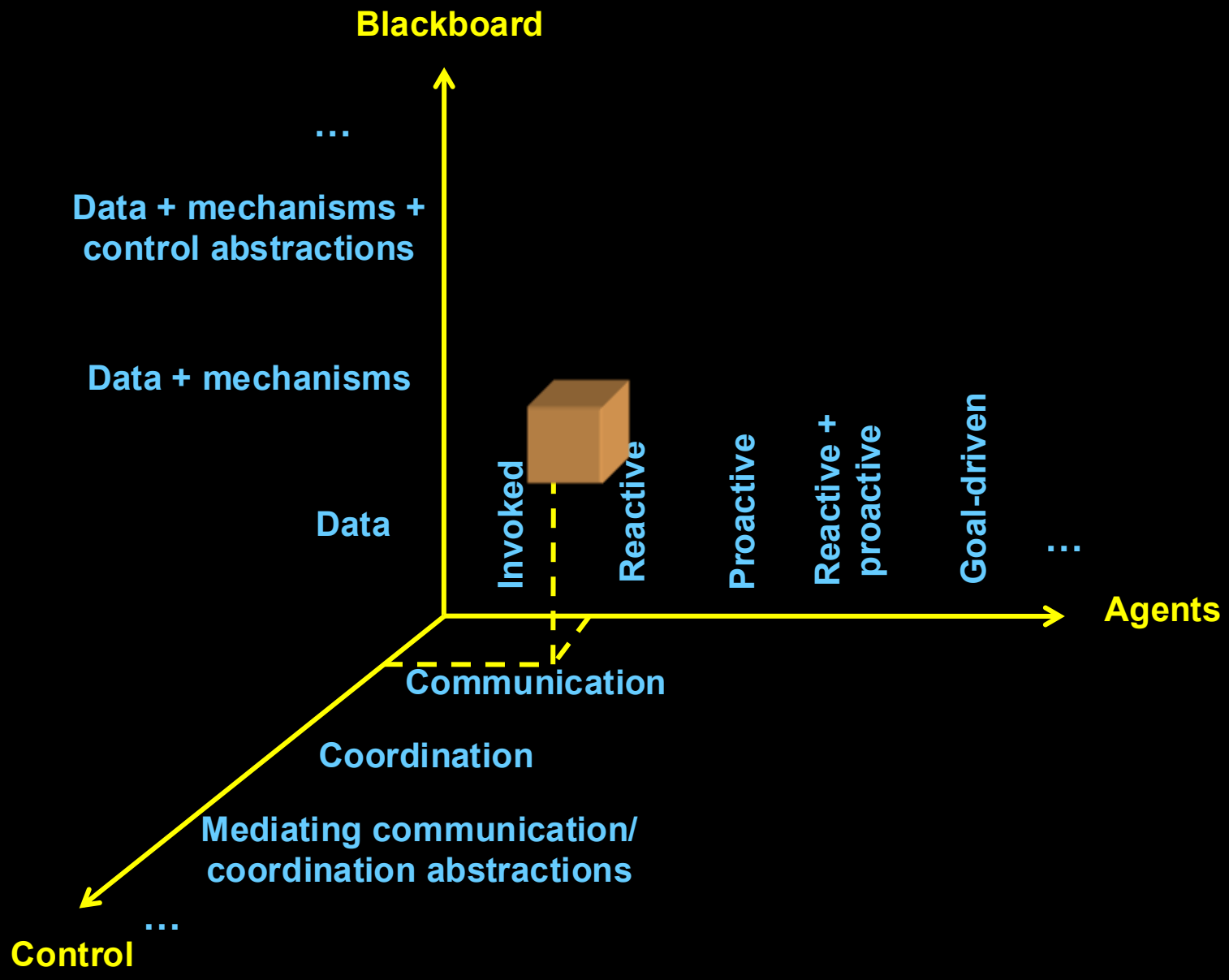
Model 2: there is a reward for PP bonds

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

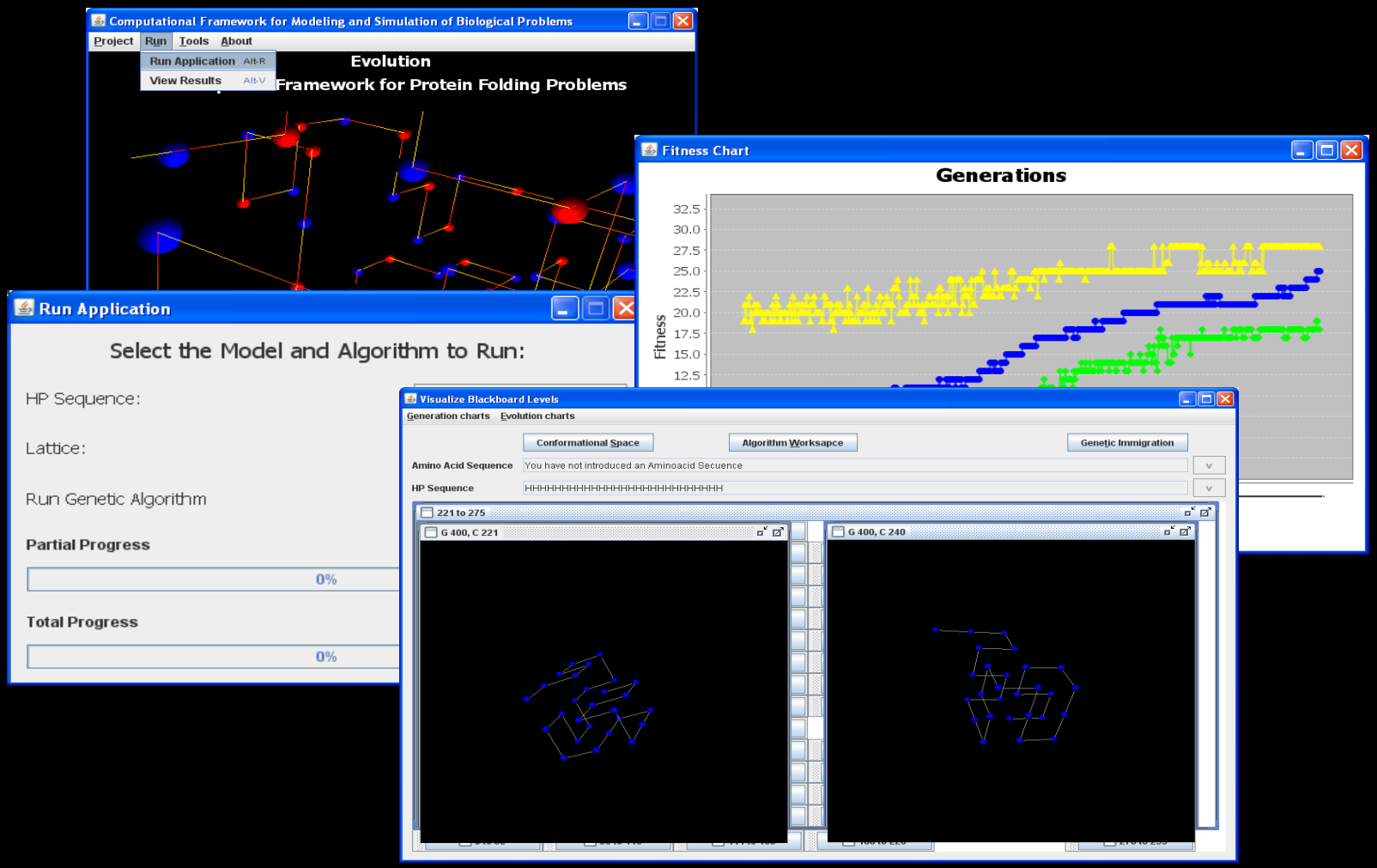
$$e_{HH} = -1.0 \quad e_{HP} = 0.0 \; e_{PP} = -0.5 \; or \; ???$$

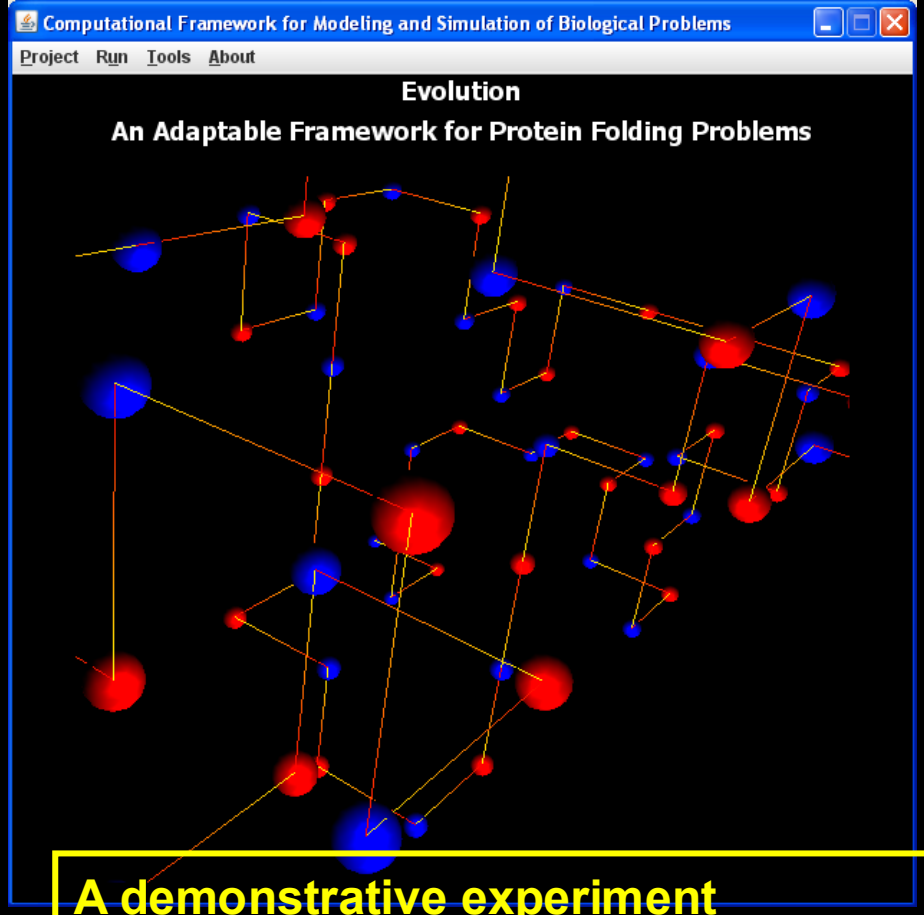
$$\Delta_{ij} = \begin{cases} 1 \; if \;\; i \; and \;\; j \; are \; topological, \; but \; not \; sequence \; neighbours \\ 0 \; otherwise \end{cases}$$

Verso un vero approccio multi-agente



Blackboard-based multi-agent architecture: the virtual laboratory



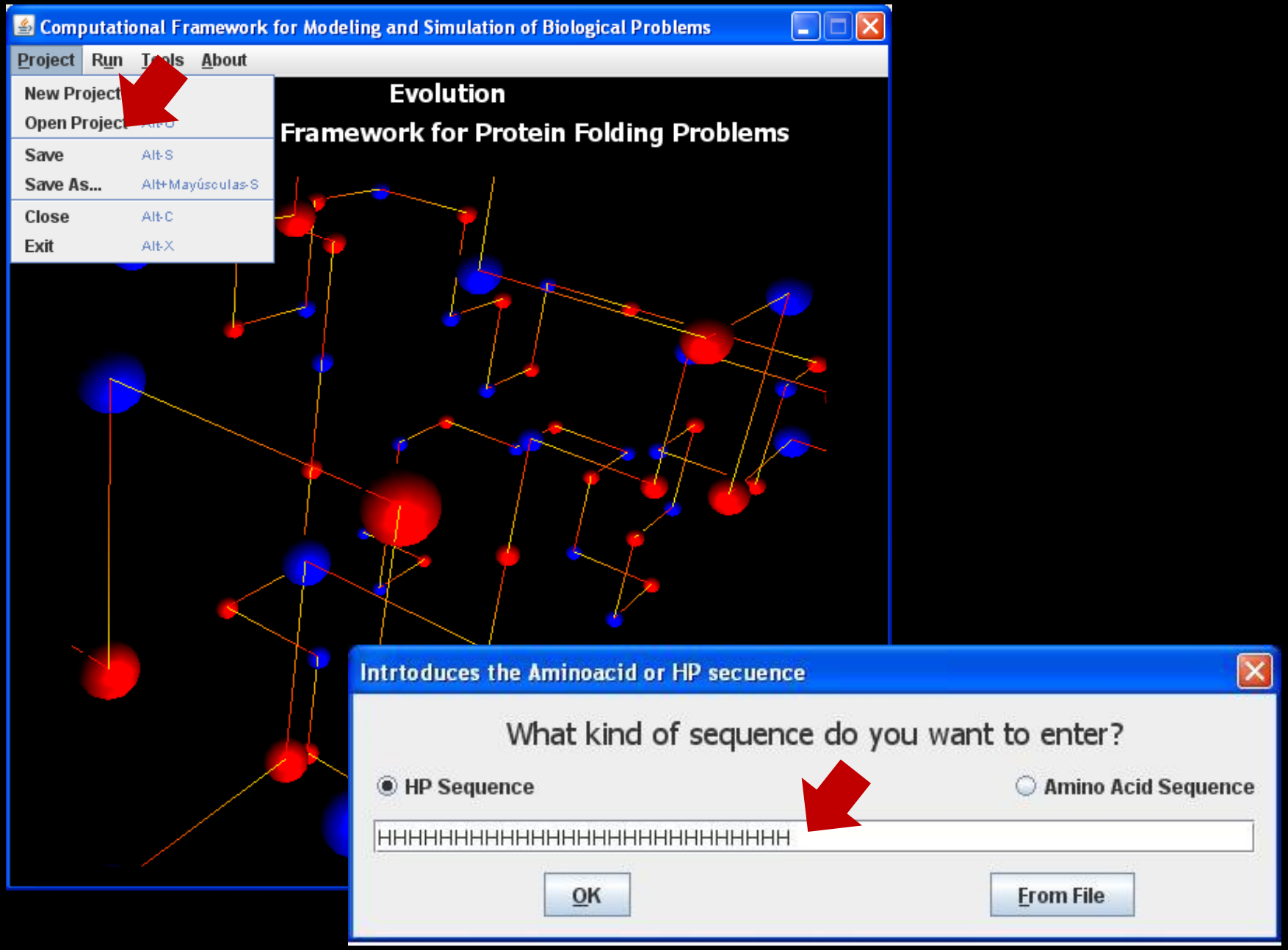


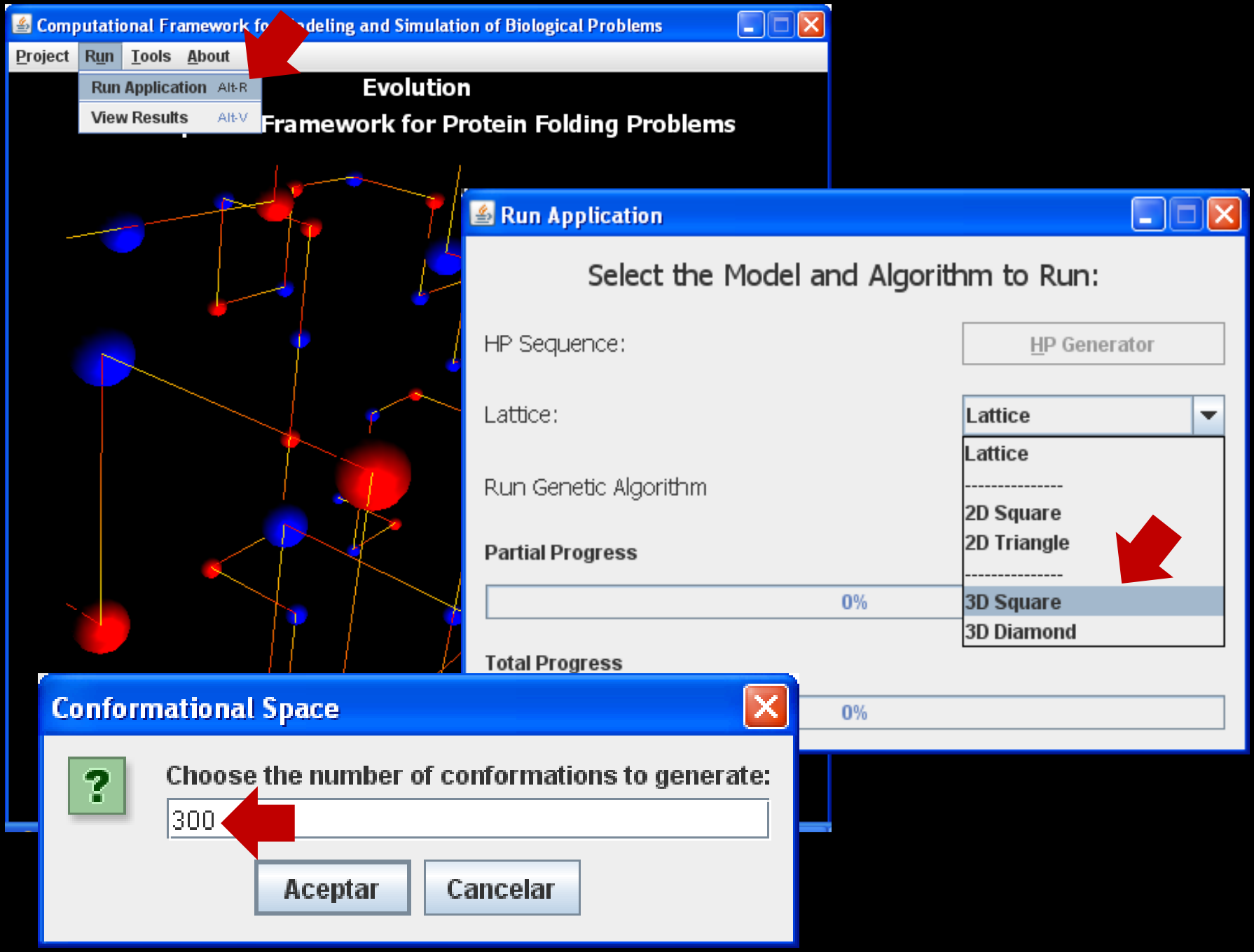
A demonstrative experiment

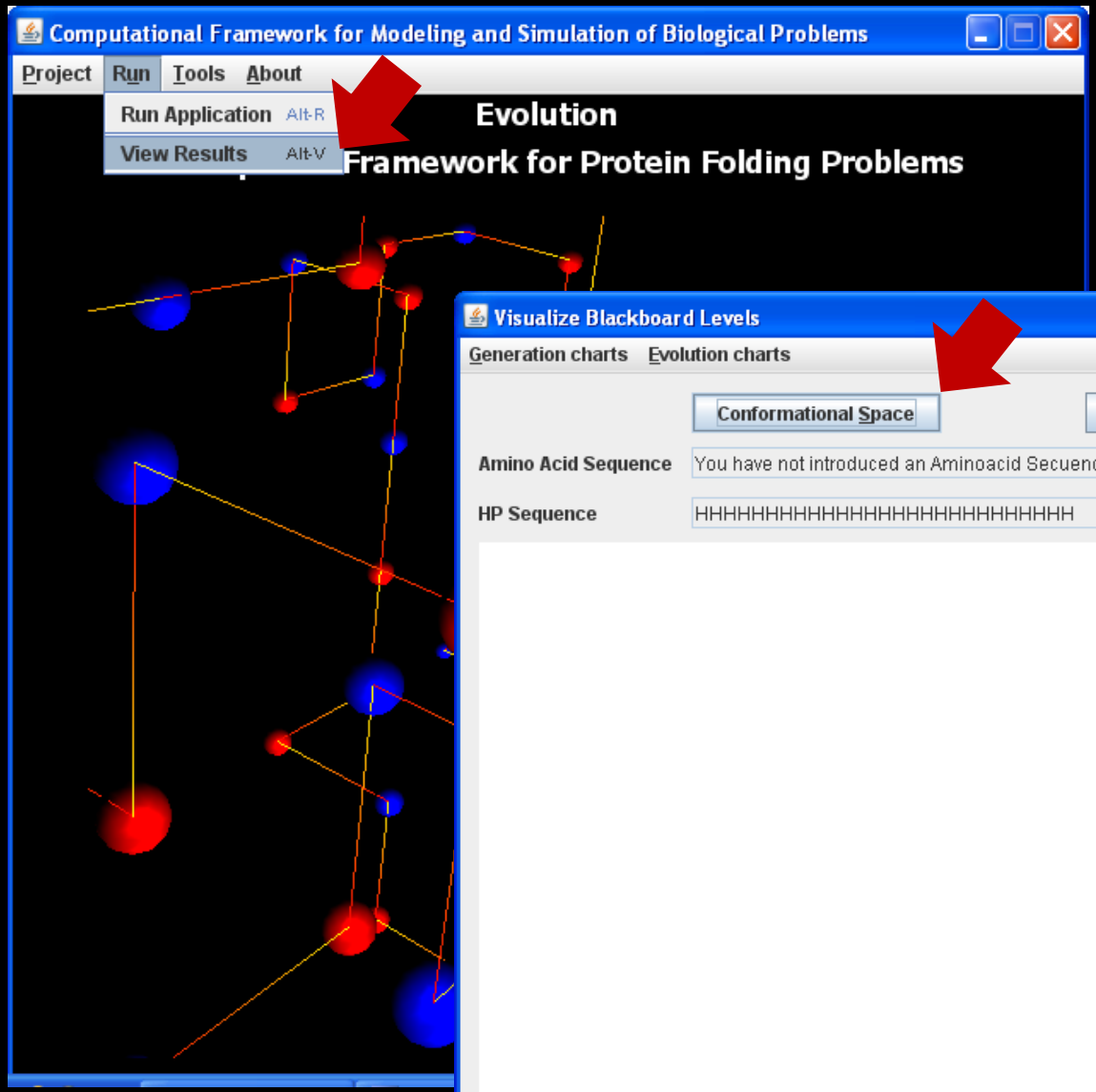
Sequence: HHHHHHHHHHHHHHHHHHHHHHHHHH

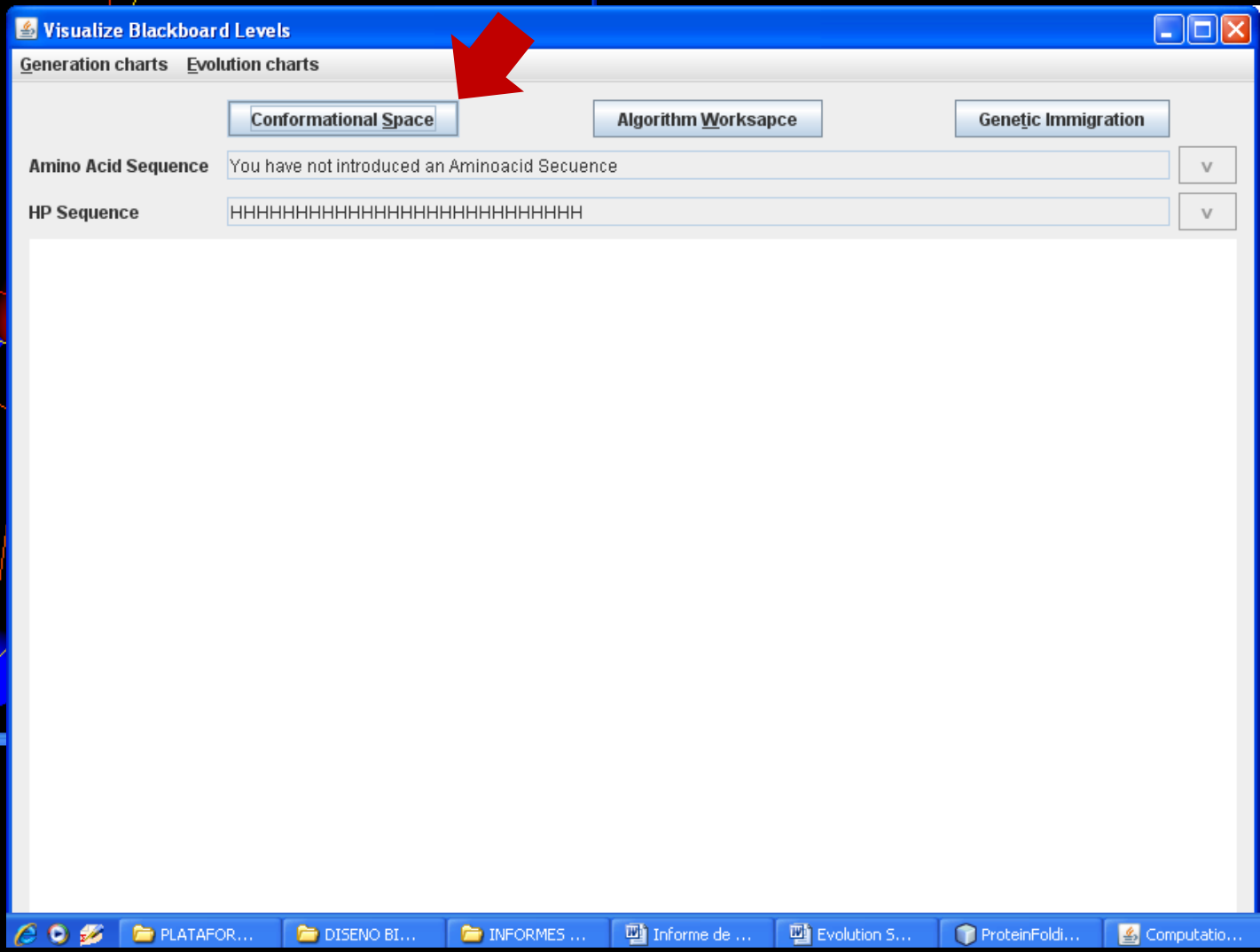
Parameters of the experiment:

- **Number of conformations per generation:** 300
- **Number of generations to be run by the evolutionary algorithm:** 500
- **Genetic operators:** *Roulette selection* method, *Single point crossover* and *Simple mutation.*
- **Other parameters:** *Self-avoiding* in the *crossover point* and *Elitism* not activated.

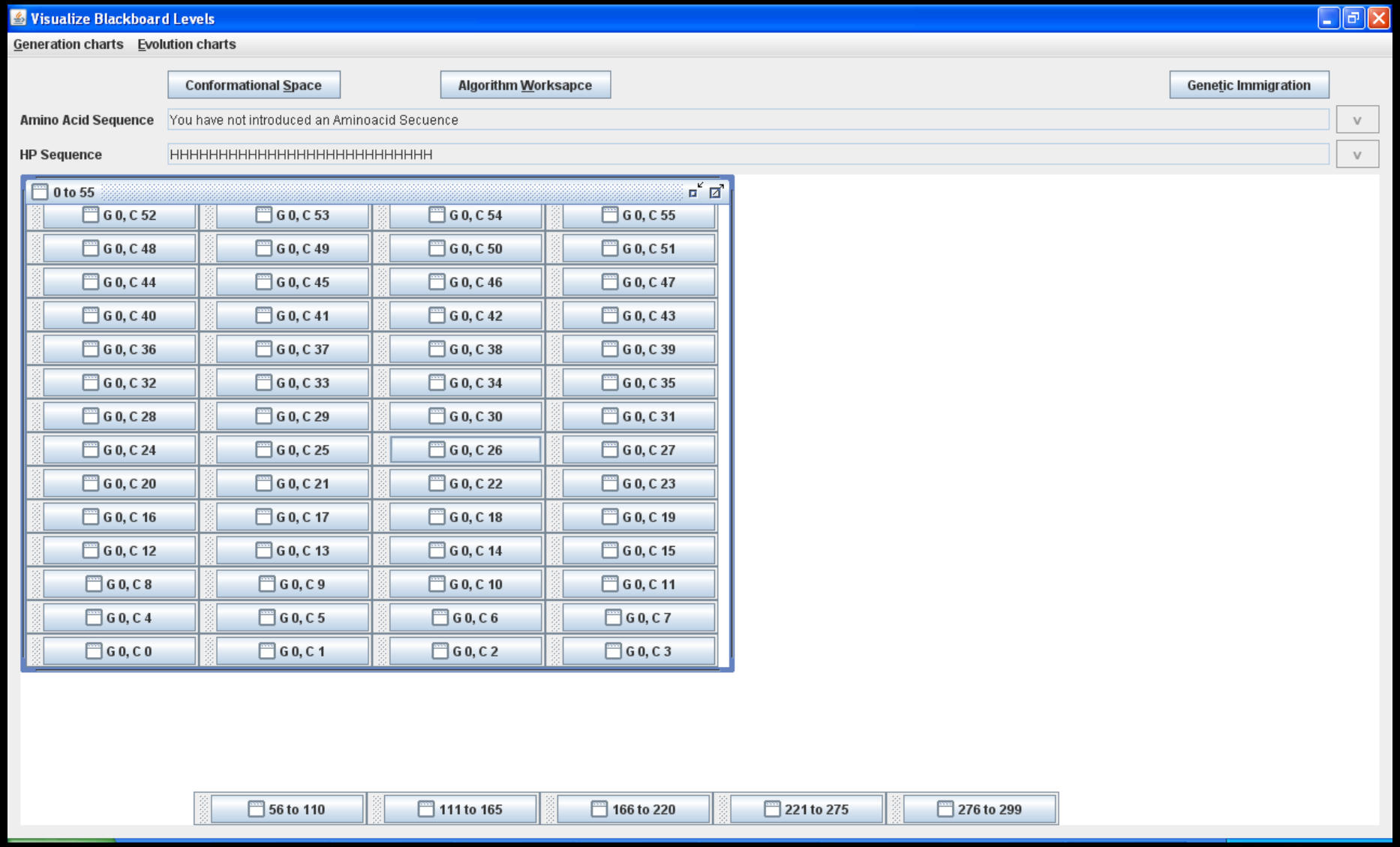

Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
New Project
Open Project
Save
Alt-S
Save As...
Alt+Mayúsculas-S
Close
Alt-C
Exit
Alt-X
Evolution
Framework for Protein Folding Problems
Intrtoduces the Aminoacid or HP secuence
What kind of sequence do you want to enter?
HP Sequence
Amino Acid Sequence
HHHHHHHHHHHHHHHHHHHHHHHHHH
OK
From File


Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
Run Application Alt-R
View Results Alt-V
Evolution
Framework for Protein Folding Problems
Run Application
Select the Model and Algorithm to Run:
HP Sequence:
HP Generator
Lattice:
Lattice
Lattice
2D Square
2D Triangle
3D Square
3D Diamond
Run Genetic Algorithm
Partial Progress
0%
Total Progress
0%
Conformational Space
Choose the number of conformations to generate:
300
Aceptar
Cancelar

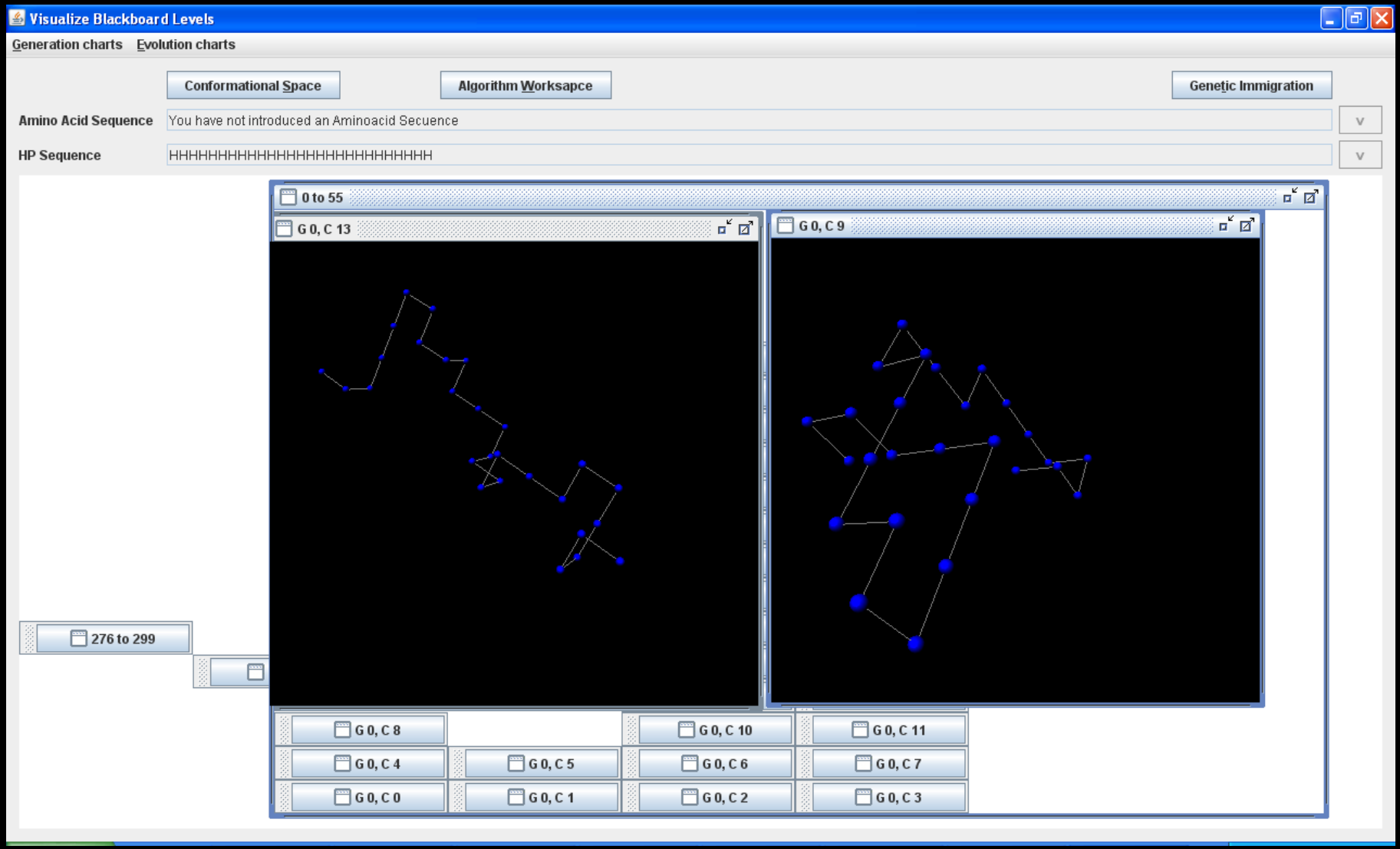

Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
Run Application Alt-R
View Results Alt-V
Evolution
Framework for Protein Folding Problems

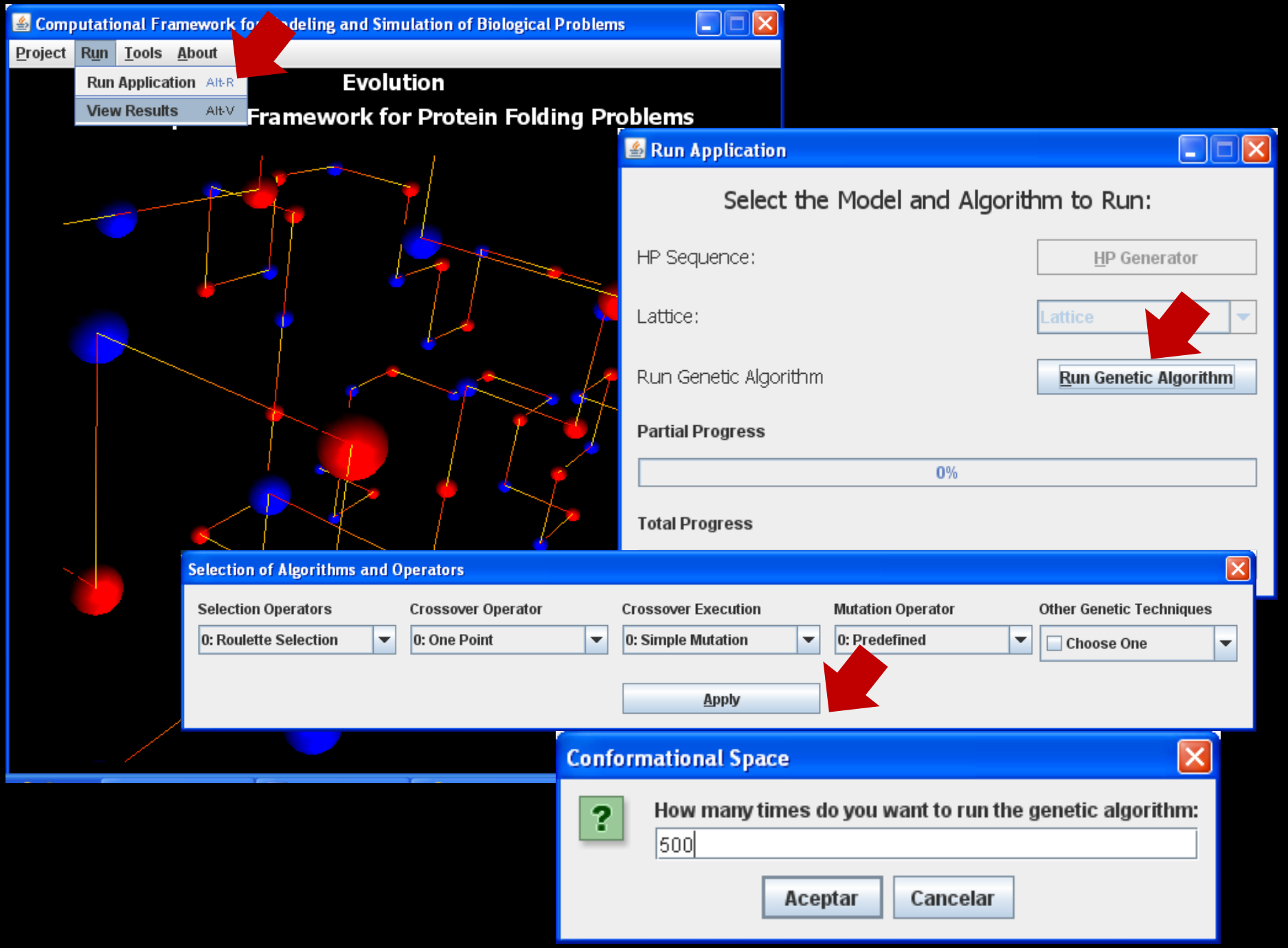

Visualize Blackboard Levels
Generation charts
Evolution charts
Conformational Space
Algorithm Worksapce
Genetic Immigration
Amino Acid Sequence
You have not introduced an Aminoacid Secuence
HP Sequence
HHHHHHHHHHHHHHHHHHHHHHHHHH
PLATAFOR...
DISENO BI...
INFORMES ...
Informe de ...
Evolution S...
ProteinFoldi...
Computatio...

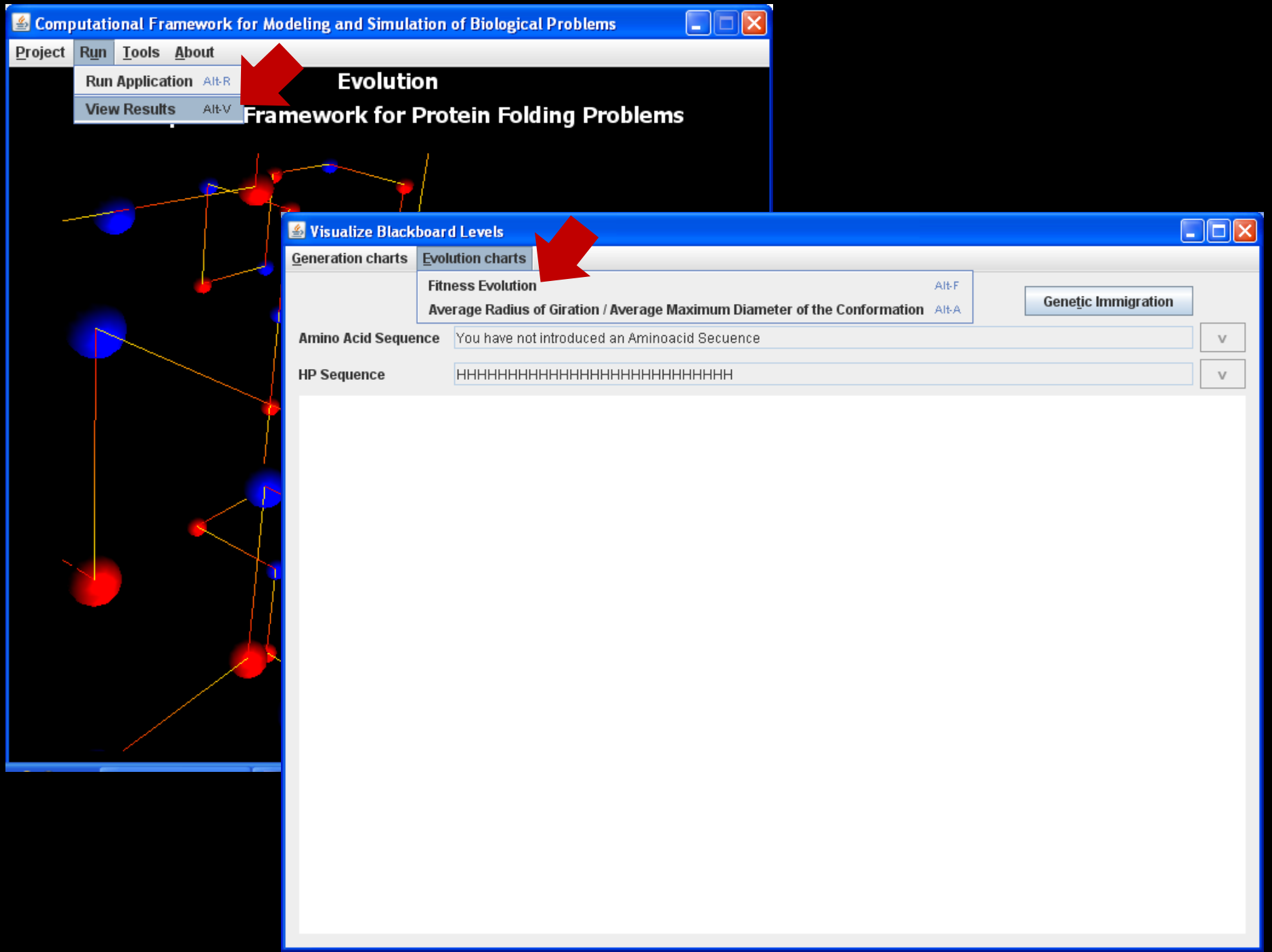
The initial generation

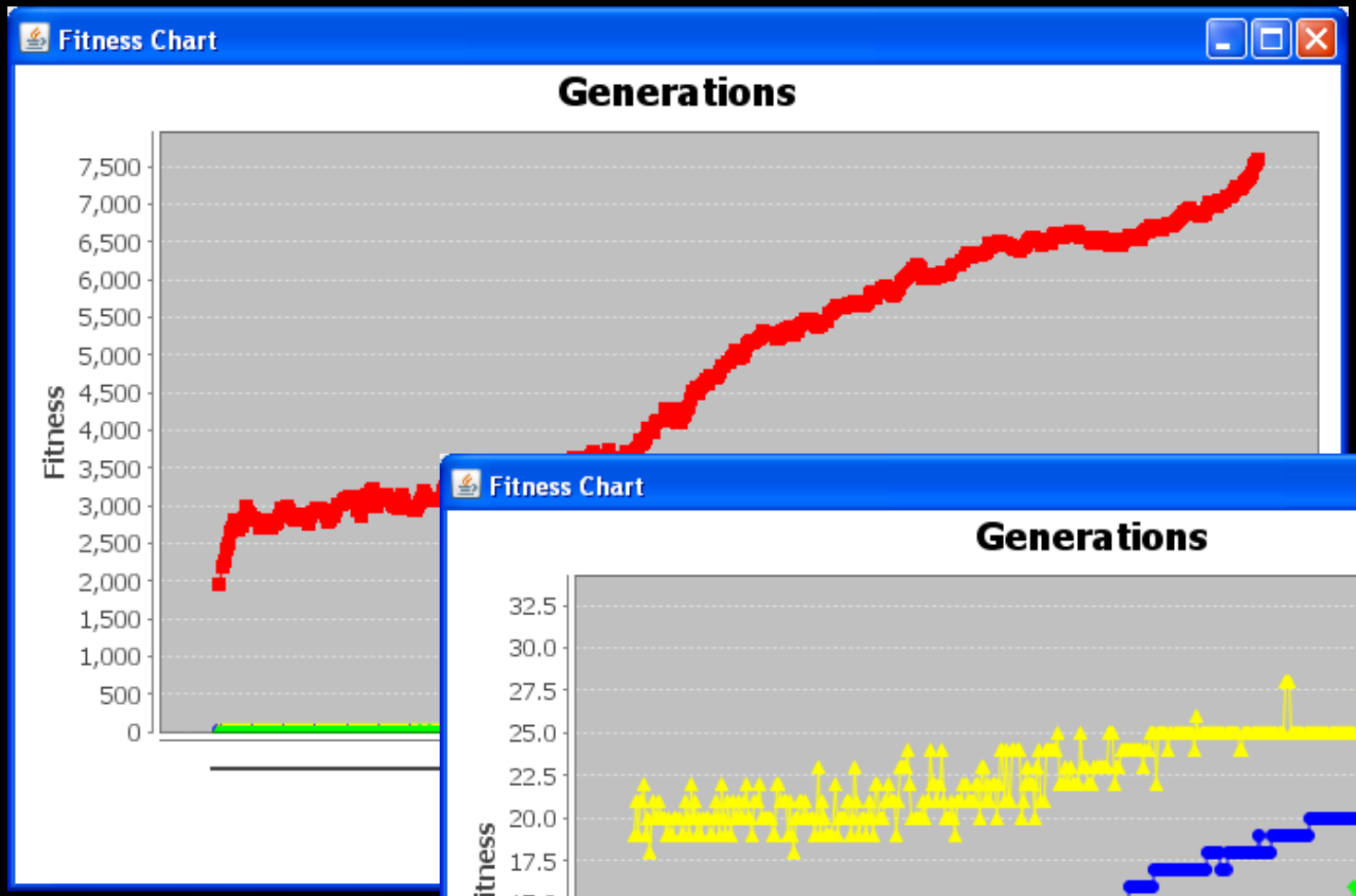


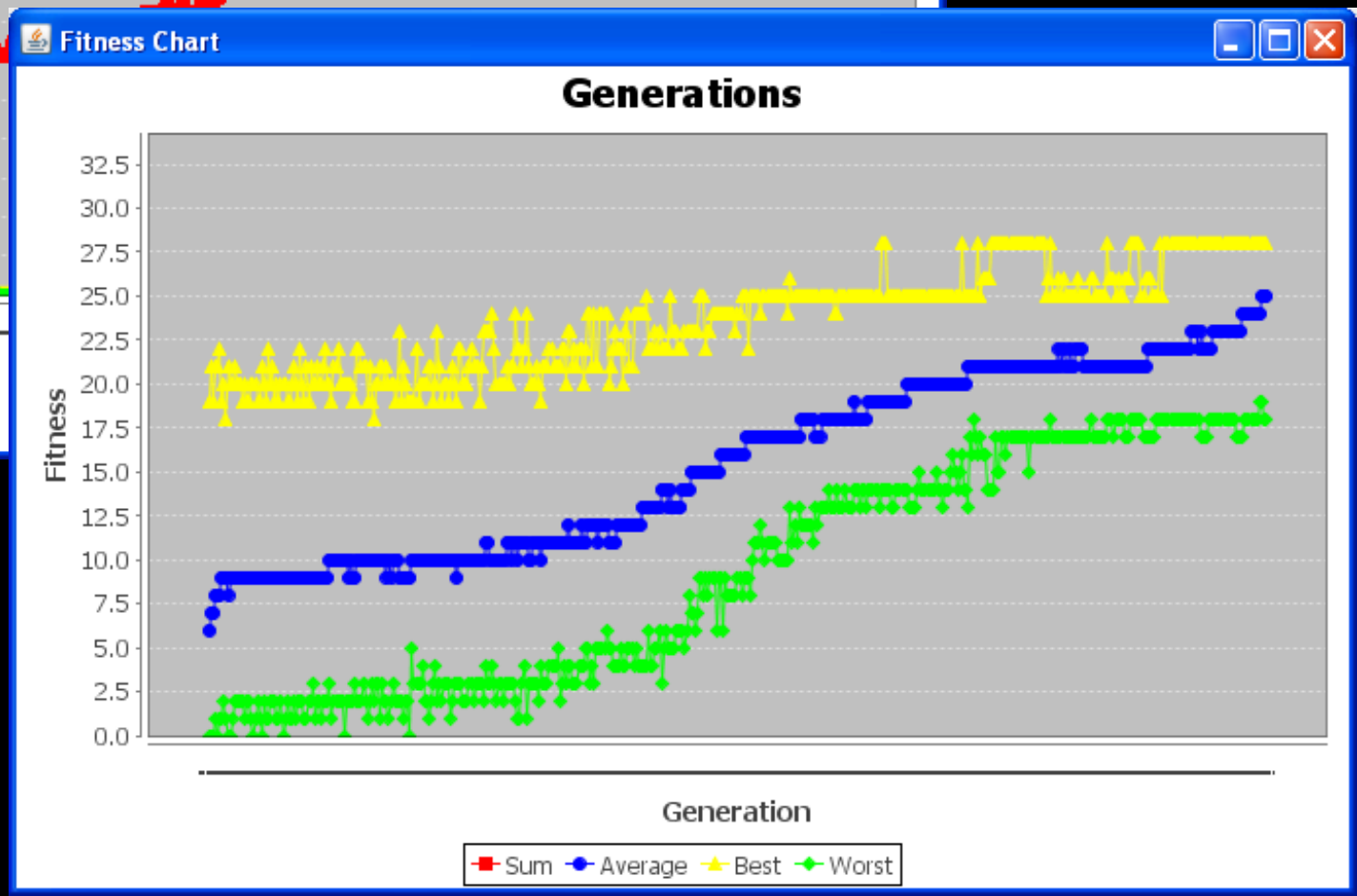
Two members from the initial generation



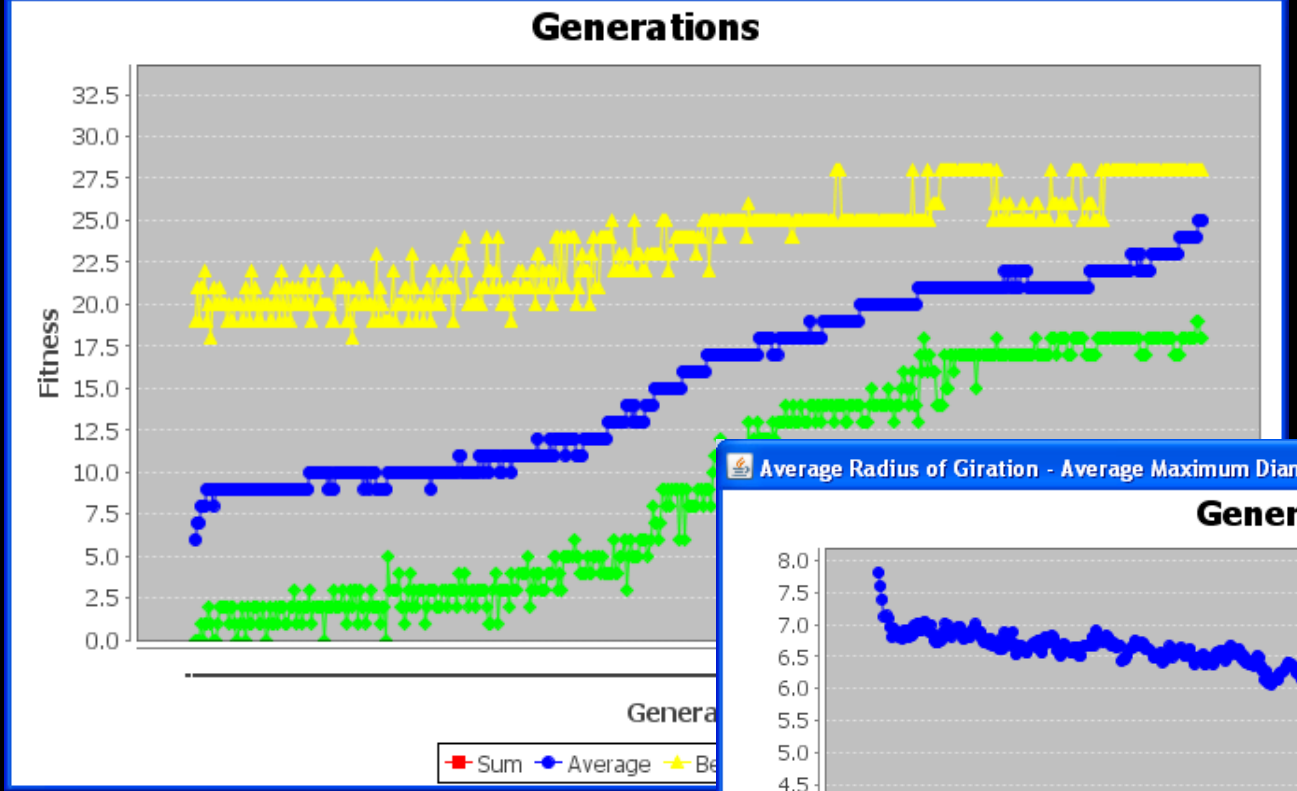

Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
Run Application Alt-R
View Results Alt-V
Evolution
Framework for Protein Folding Problems
Run Application
Select the Model and Algorithm to Run:
HP Sequence:
HP Generator
Lattice:
Lattice
Run Genetic Algorithm
Run Genetic Algorithm
Partial Progress
0%
Total Progress
Selection of Algorithms and Operators
Selection Operators
0: Roulette Selection
Crossover Operator
0: One Point
Crossover Execution
0: Simple Mutation
Mutation Operator
0: Predefined
Other Genetic Techniques
Choose One
Apply
Conformational Space
How many times do you want to run the genetic algorithm:
500
Aceptar
Cancelar


Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
Run Application Alt-R
View Results Alt-V
Evolution
Framework for Protein Folding Problems
Visualize Blackboard Levels
Generation charts
Evolution charts
Fitness Evolution Alt-F
Average Radius of Giration / Average Maximum Diameter of the Conformation Alt-A
Genetic Immigration
Amino Acid Sequence
You have not introduced an Aminoacid Secuence
HP Sequence
HHHHHHHHHHHHHHHHHHHHHHHHHH

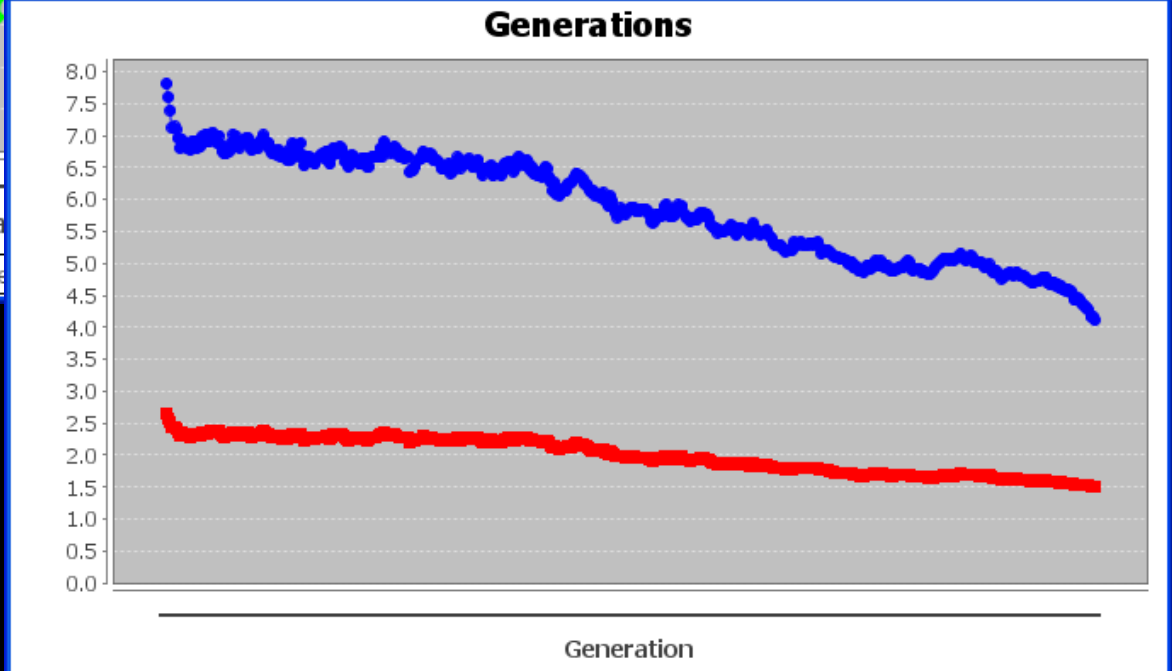

Fitness Chart
Generations
Fitness
7,500
7,000
6,500
6,000
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0


Fitness Chart
Generations
Fitness
32.5
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Generation
Sum
Average
Best
Worst

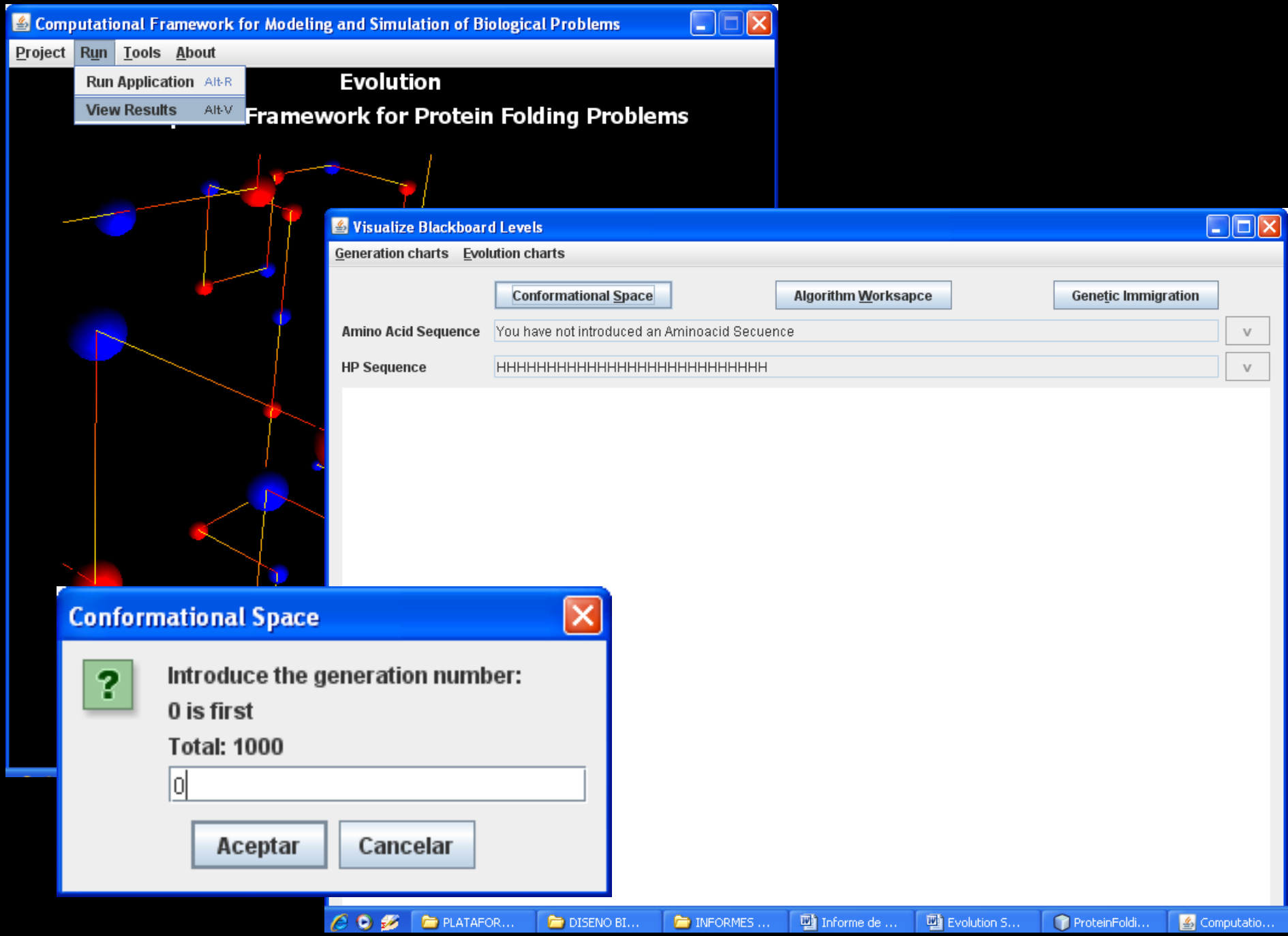
Fitness Chart

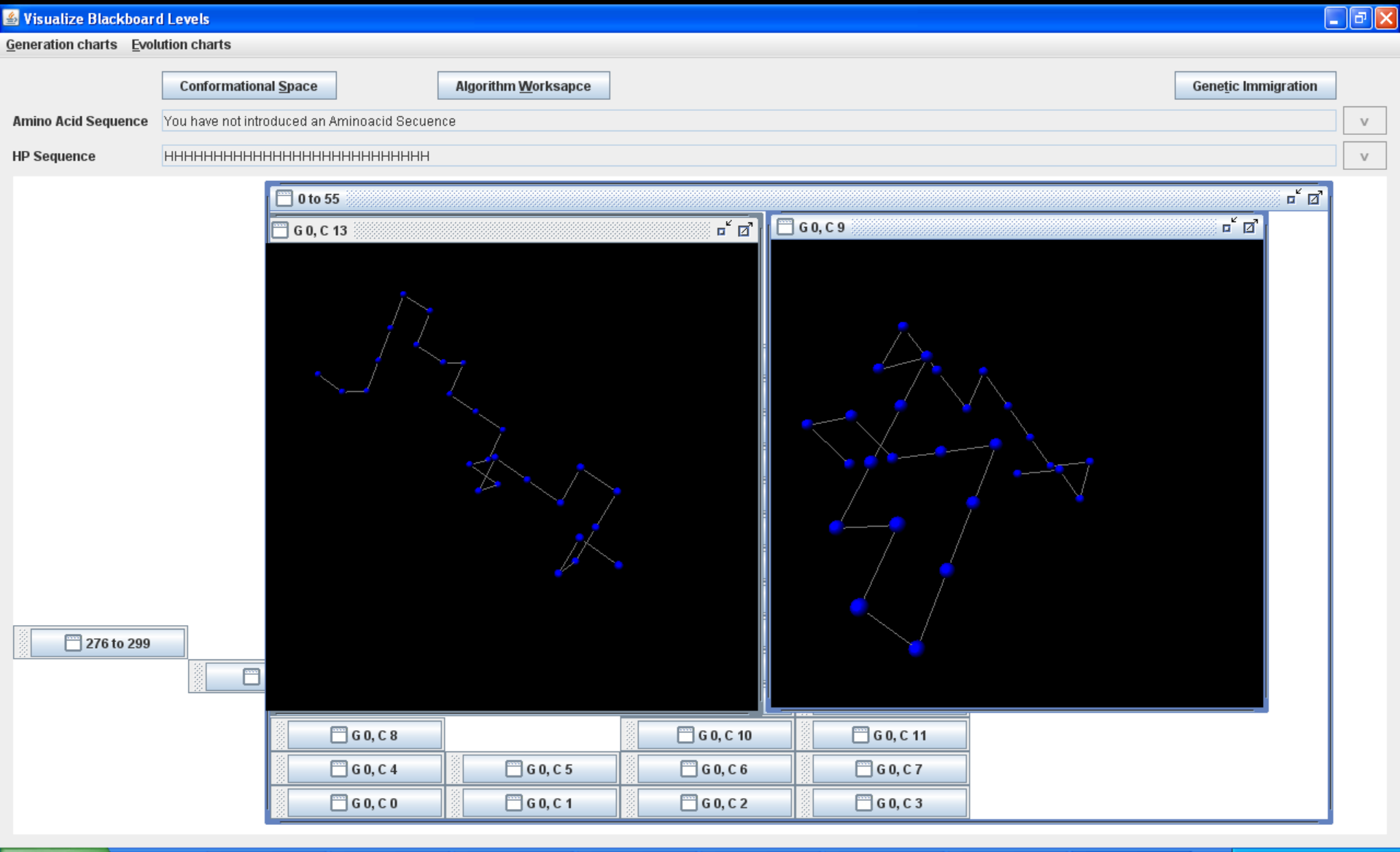


Average Radius of Giration - Average Maximum Diameter of the Conformation

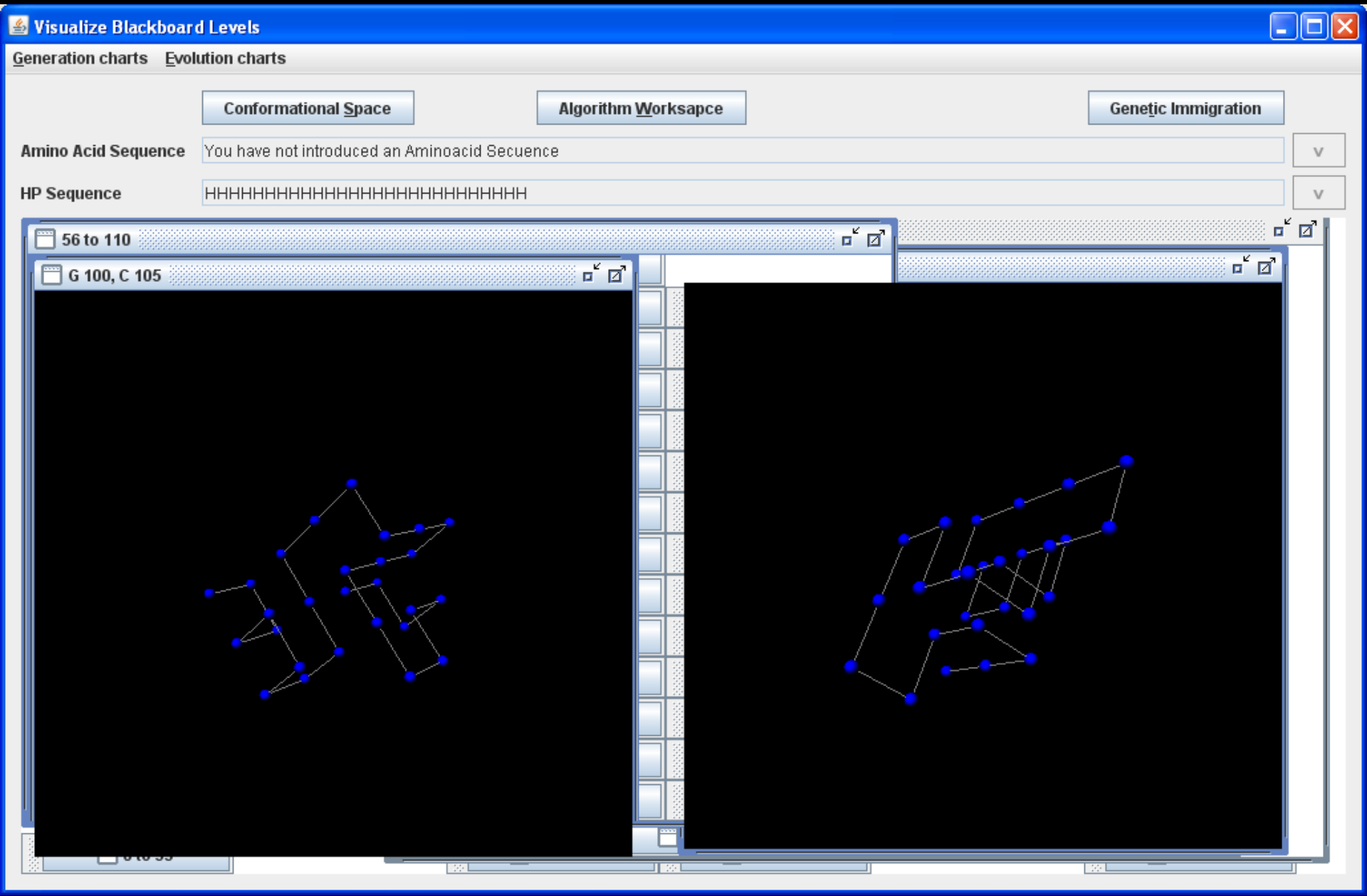


	0	1	2	3	4	5	6	7
Rgp	2.6573386...	2.5771096...	2.5165340...	2.4424663...	2.4304483...	2.3987645...	2.3470272...	2.3099633...
Dmaxp	7.8064430...	7.6010336...	7.3944657...	7.1237227...	7.1366601...	7.0970556...	6.9532254...	6.8098883...

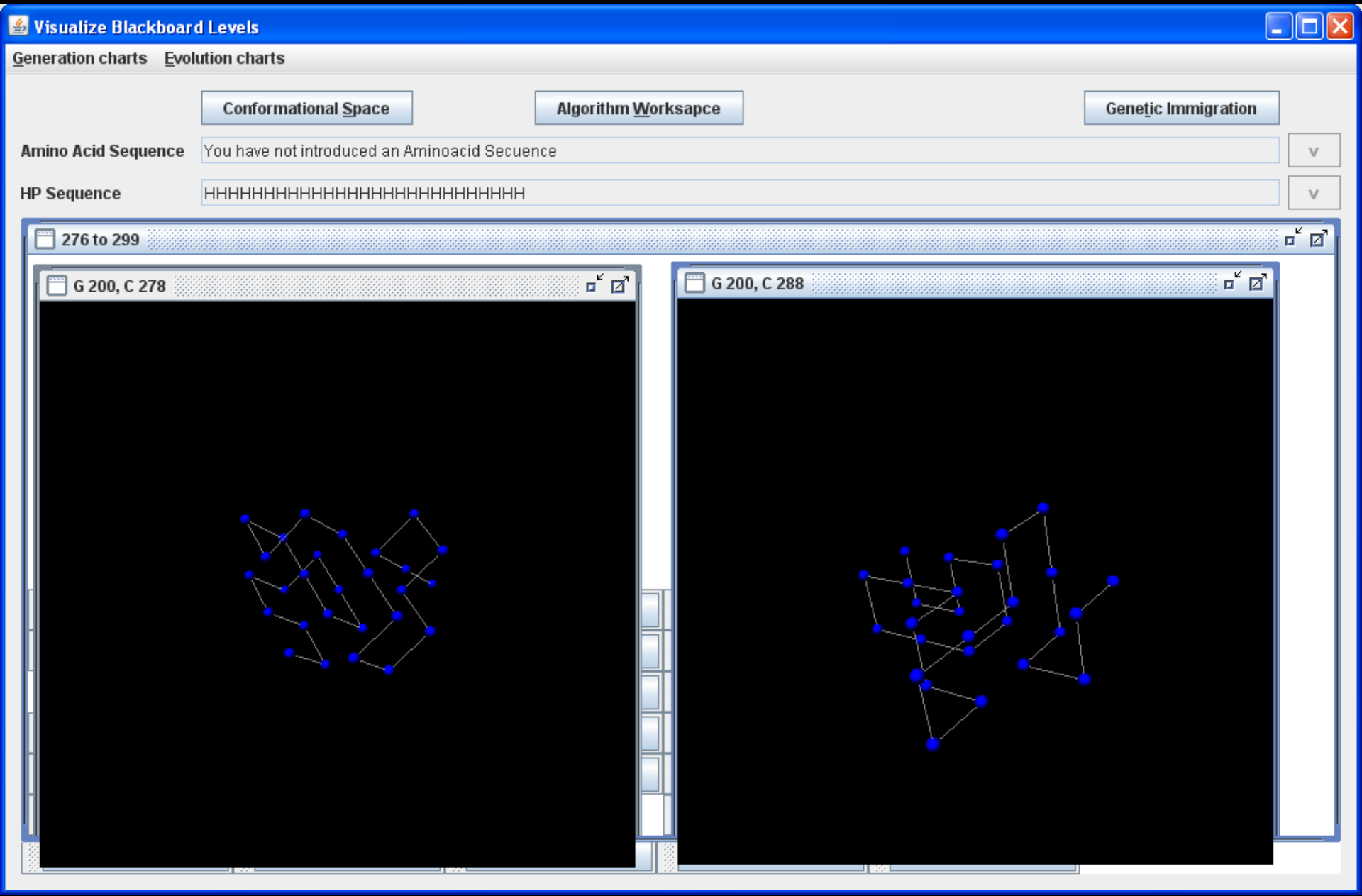

Computational Framework for Modeling and Simulation of Biological Problems
Project
Run
Tools
About
Run Application Alt-R
View Results Alt-V
Evolution
Framework for Protein Folding Problems
Visualize Blackboard Levels
Generation charts
Evolution charts
Conformational Space
Algorithm Worksapce
Genetic Immigration
Amino Acid Sequence
You have not introduced an Aminoacid Secuence
HP Sequence
HHHHHHHHHHHHHHHHHHHHHHHHHH
Conformational Space
Introduce the generation number:
0 is first
Total: 1000
0
Aceptar
Cancelar
PLATAFOR...
DISENO BI...
INFORMES ...
Informe de ...
Evolution S...
ProteinFoldi...
Computatio...



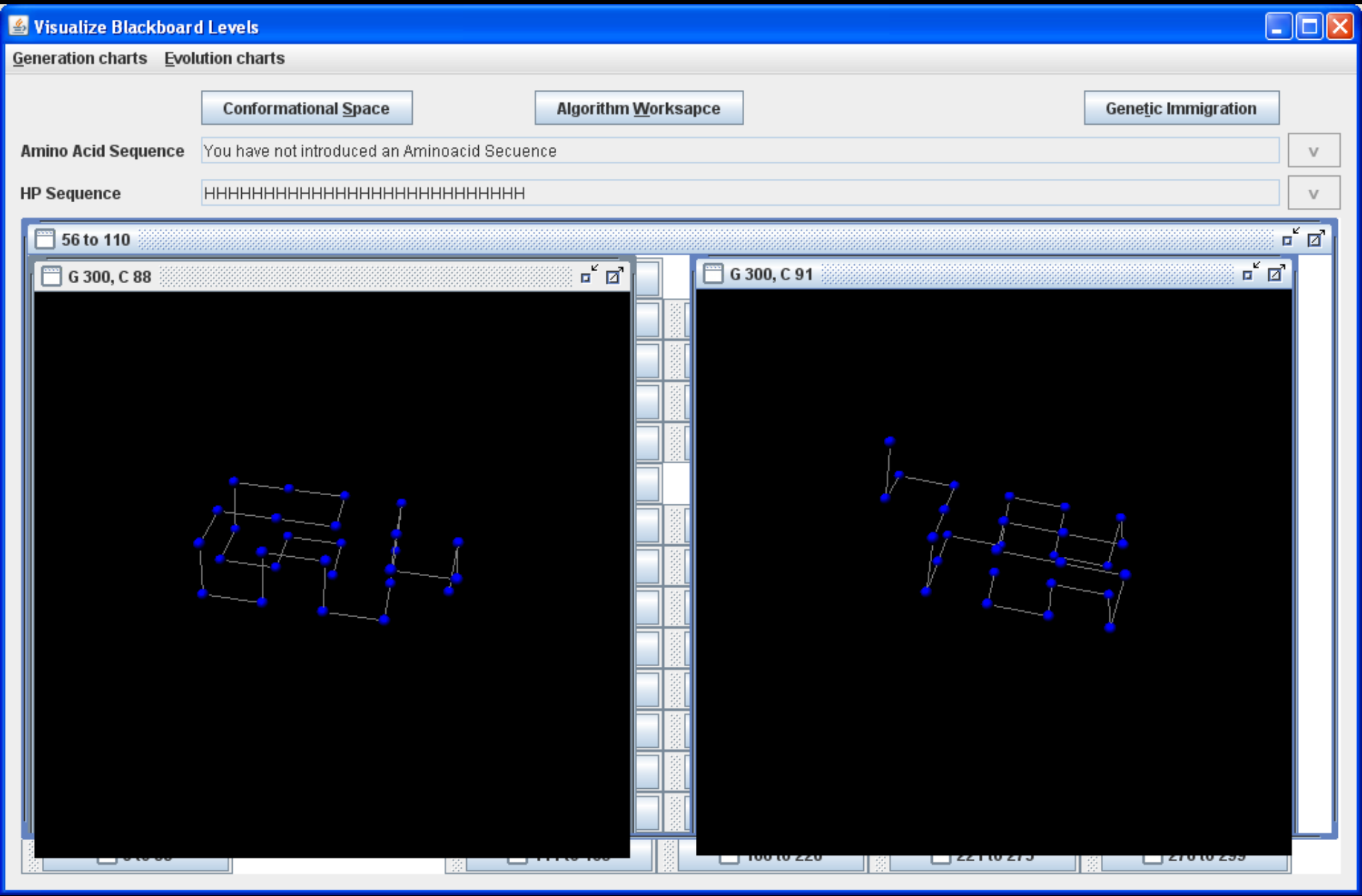
Two members from the initial generation



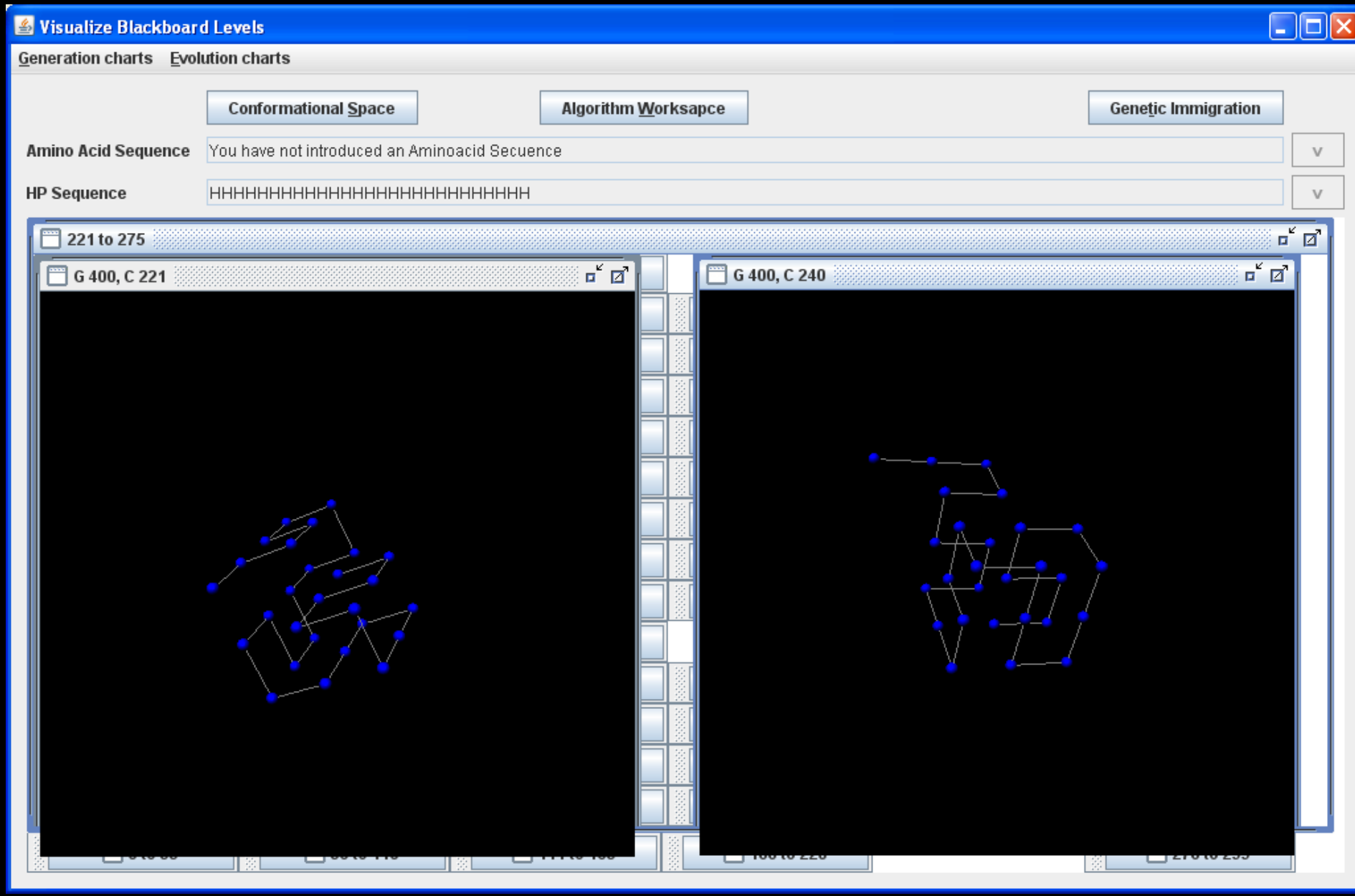
Two members from generation 100



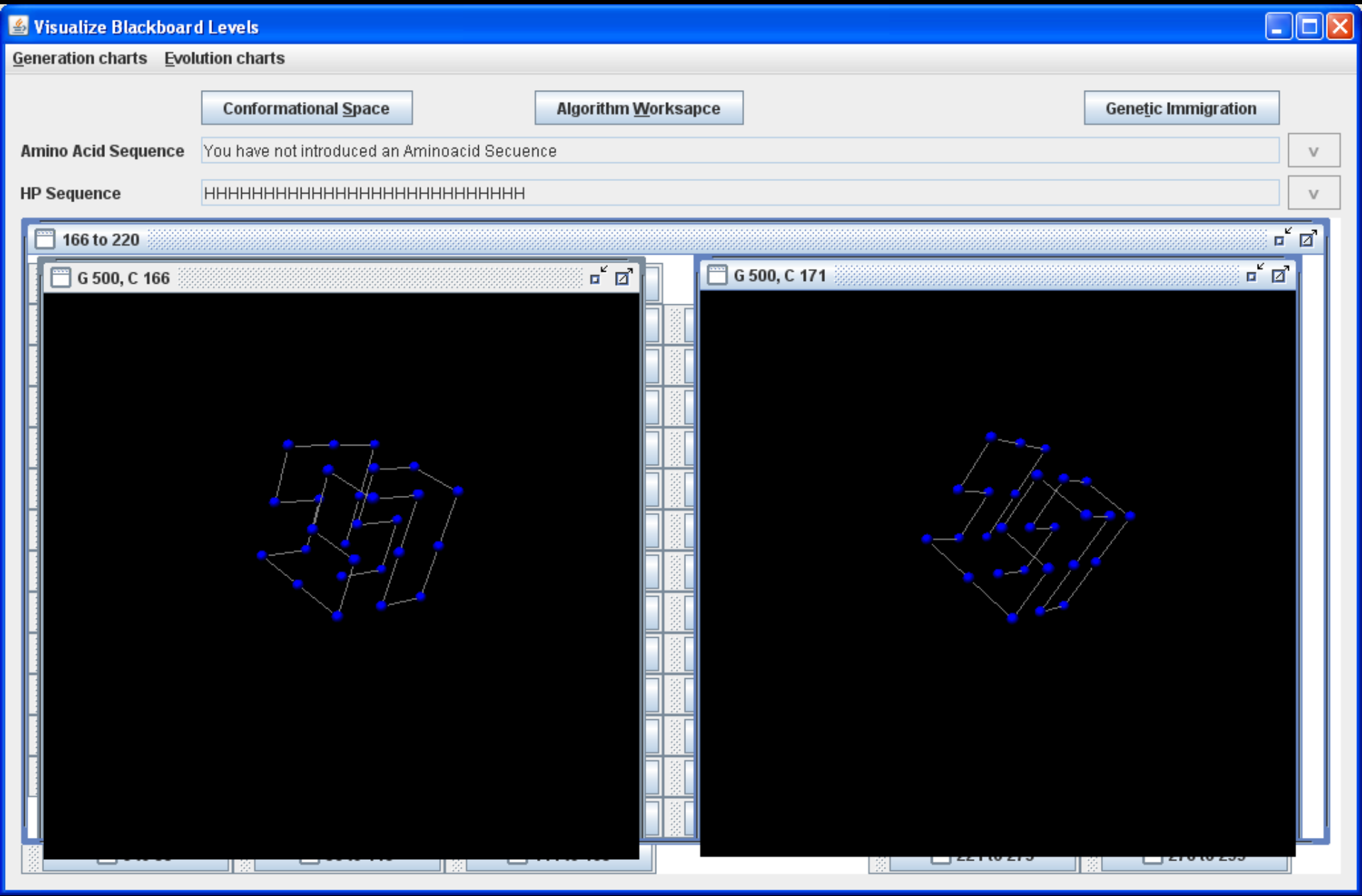
Two members from generation 200



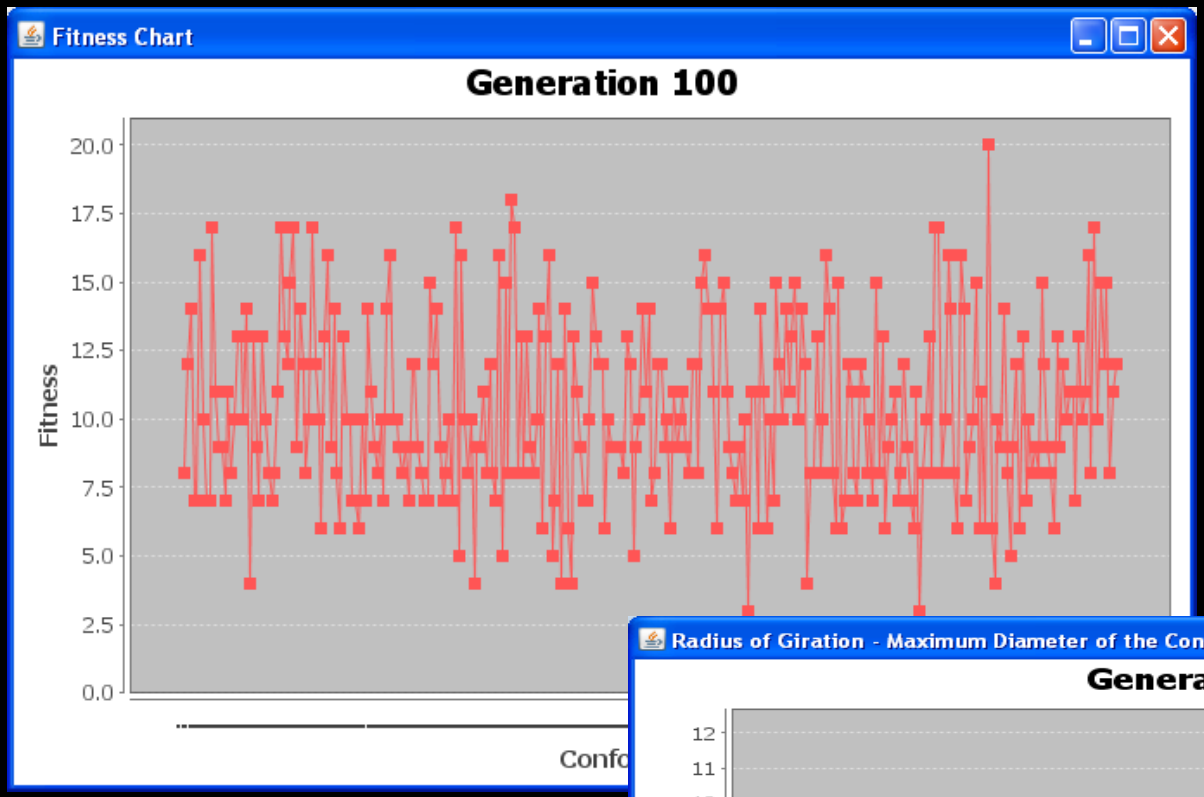
Two members from generation 300

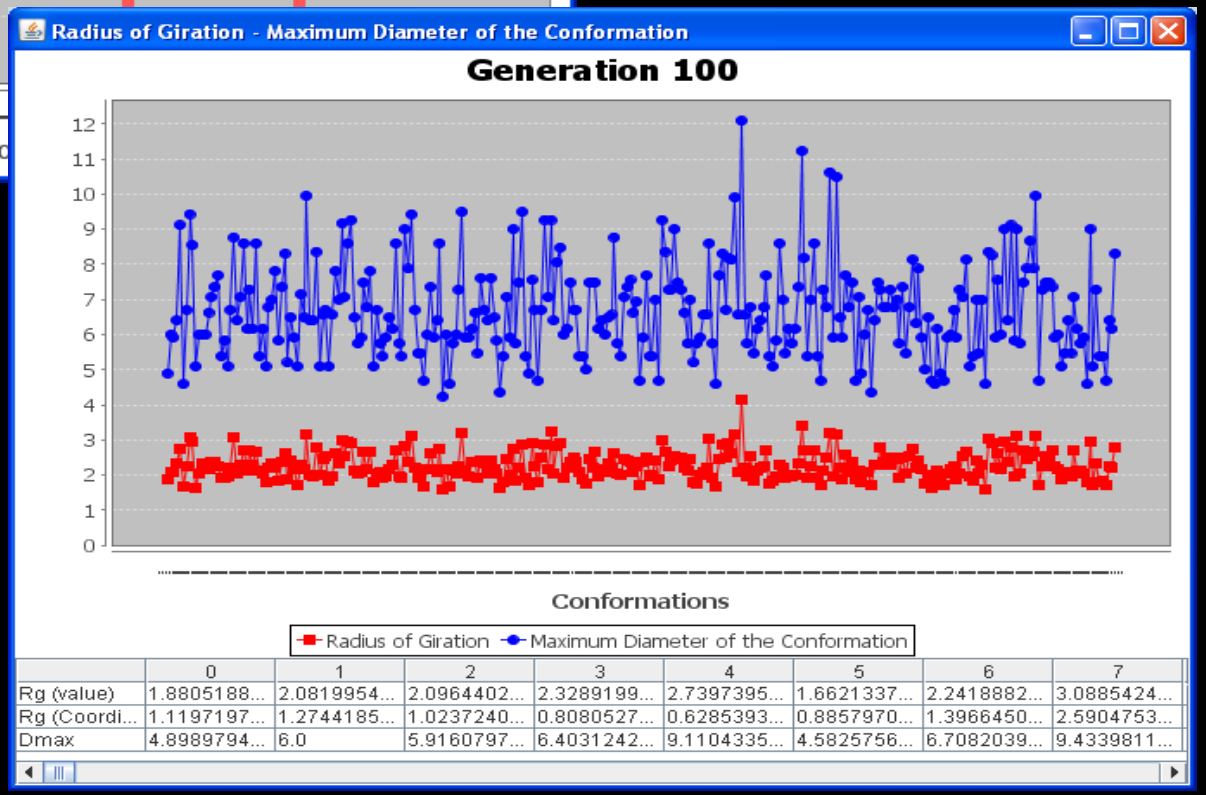


Two members from generation 400

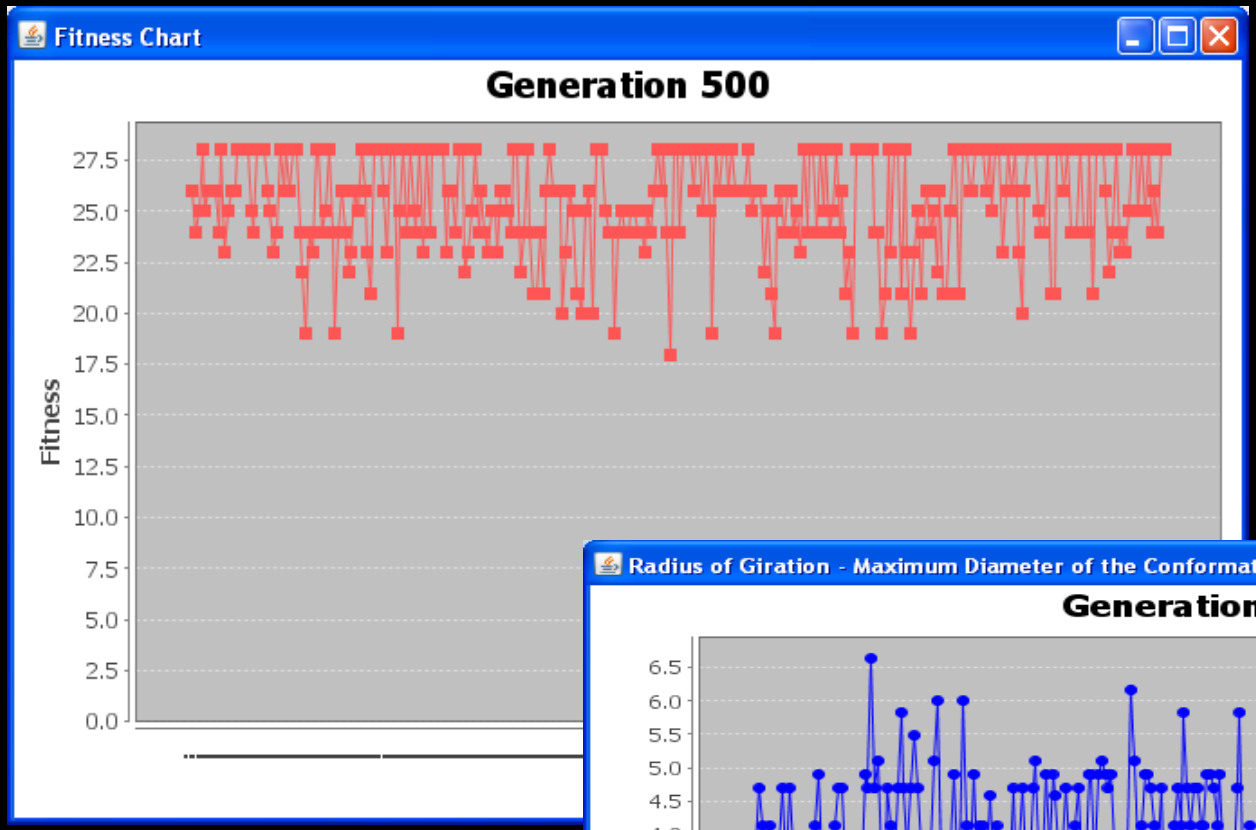


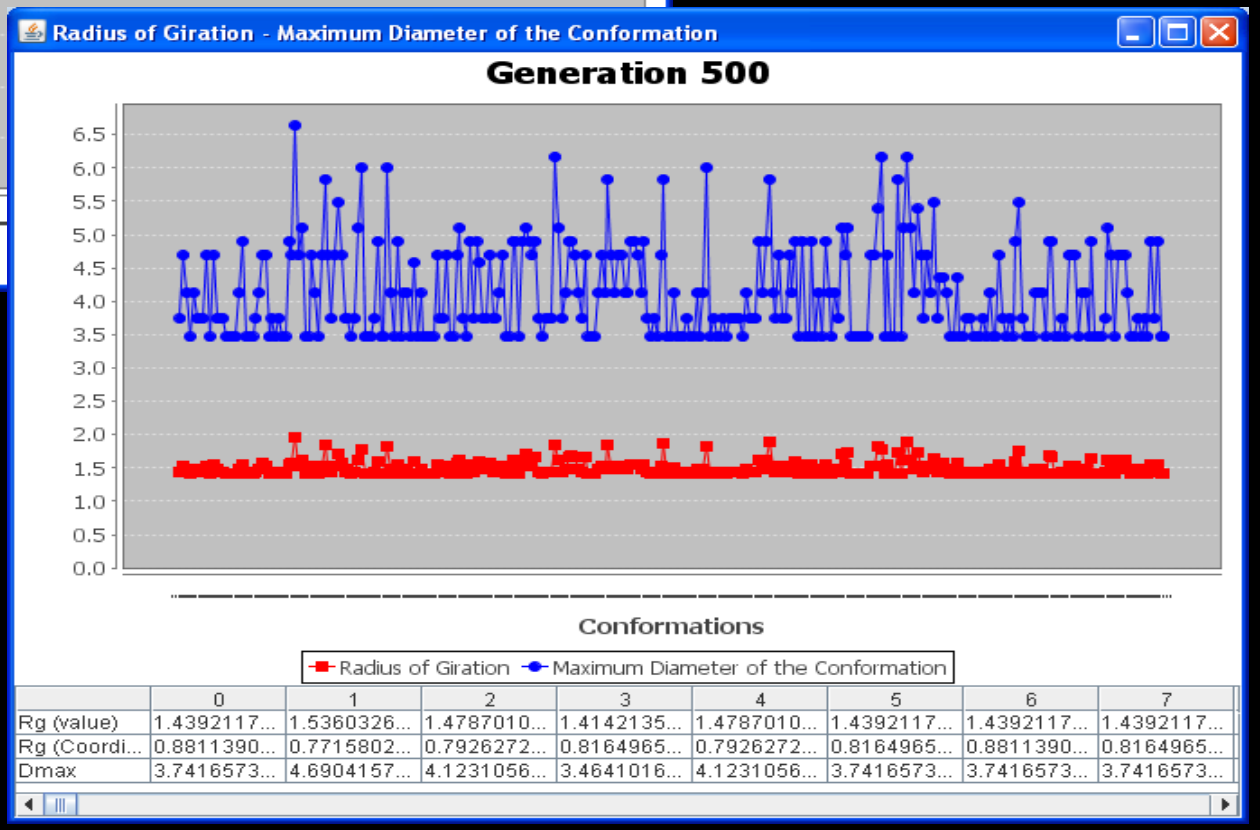
Two members from generation 500





	0	1	2	3	4	5	6	7
Rg (value)	1.8805188...	2.0819954...	2.0964402...	2.3289199...	2.7397395...	1.6621337...	2.2418882...	3.0885424...
Rg (Coordi...	1.1197197...	1.2744185...	1.0237240...	0.8080527...	0.6285393...	0.8857970...	1.3966450...	2.5904753...
Dmax	4.8989794...	6.0	5.9160797...	6.4031242...	9.1104335...	4.5825756...	6.7082039...	9.4339811...

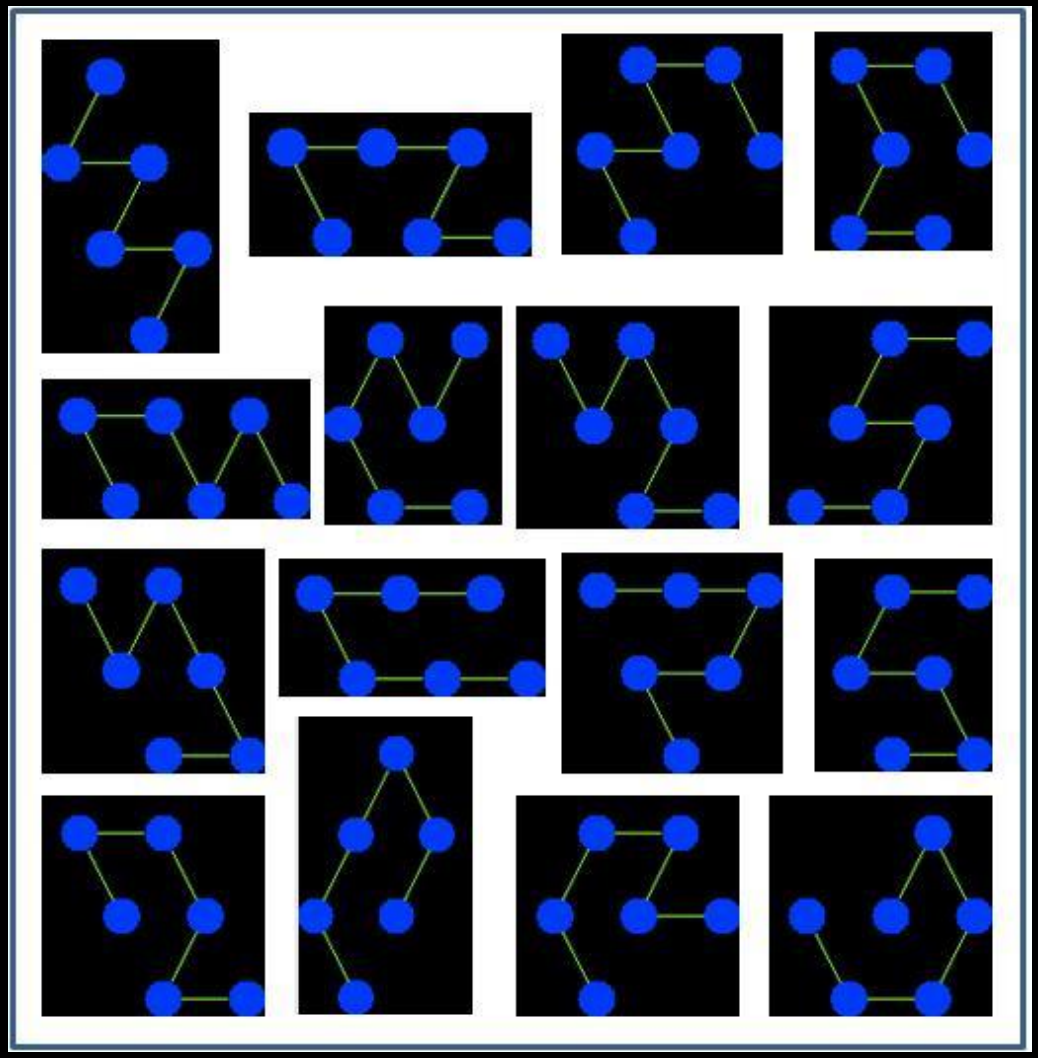




	0	1	2	3	4	5	6	7
Rg (value)	1.4392117...	1.5360326...	1.4787010...	1.4142135...	1.4787010...	1.4392117...	1.4392117...	1.4392117...
Rg (Coordi...	0.8811390...	0.7715802...	0.7926272...	0.8164965...	0.7926272...	0.8164965...	0.8811390...	0.8164965...
Dmax	3.7416573...	4.6904157...	4.1231056...	3.4641016...	4.1231056...	3.7416573...	3.7416573...	3.7416573...

Search of local determinants of structure

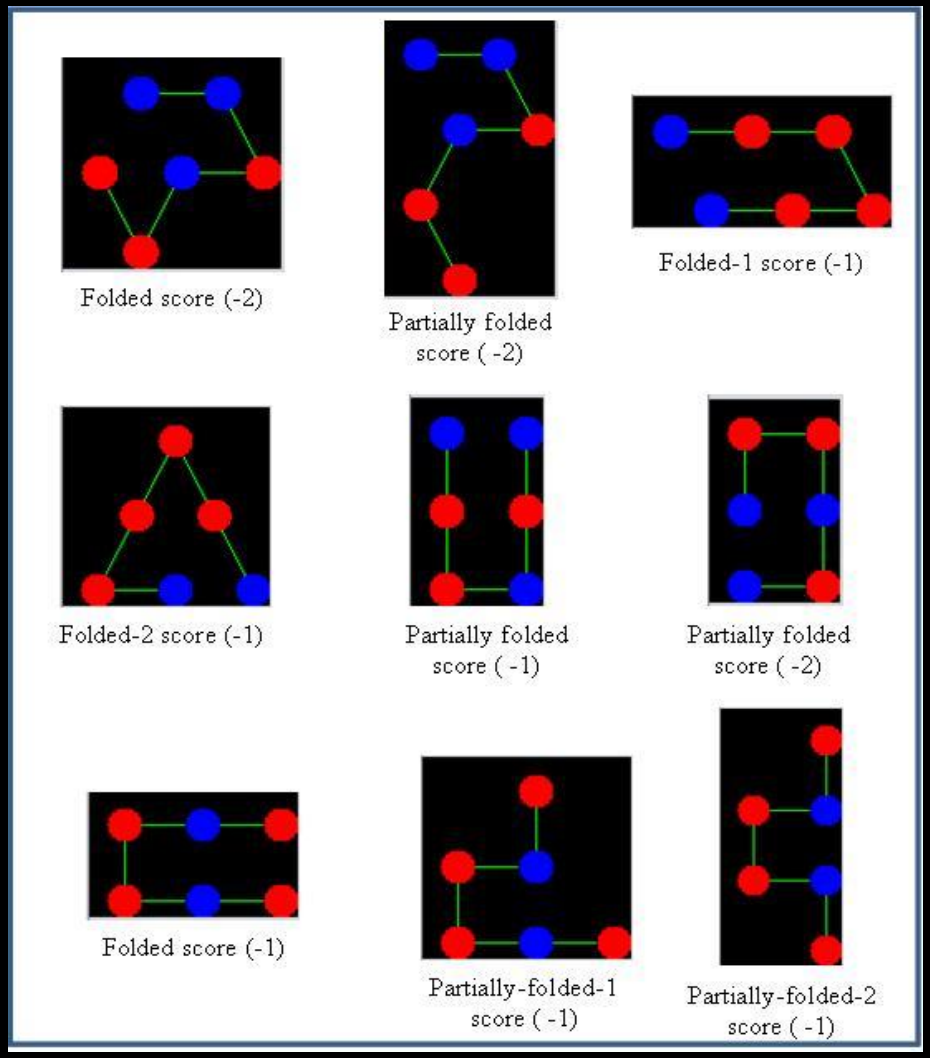
SELECTED SEQUENCES OF HEXAMERIC HP SYSTEMS	
HHHHHH	PPHPHH
PHHHHH	PPHHPH
HPHHHH	PHPPHH
HHPHHH	HPPPHH
PPHHHH	HPHPPH
PHPHHH	PHPHPH
PHHPHH	HPPHPH
PHHHPH	HHPPPH
PHHHHP	HPPPPH
HPPHHH	HPPHPP
HPHPHH	PHPPHP
HPHHPH	PHHPPH
HHPPHH	PPPPPP



Hexamer folding in 2D triangular lattice. For the HHHHHH several degenerated structures, with the maximum score of 4 can be detected with the use of the platform. A minus 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 varying degrees of degeneracy and maximum number of contacts can be tracked.

Search of local determinants of structure

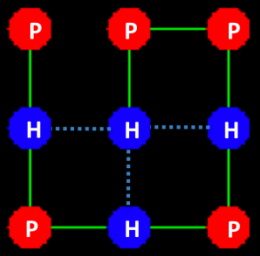
SELECTED SEQUENCES OF HEXAMERIC HP SYSTEMS	
HHHHHH	PPHPHH
PHHHHH	PPHHPH
HPHHHH	PHPPHH
HHPHHH	HPPPHH
PPHHHH	HPHPPH
PHPHHH	PHPHPH
PHHPHH	HPPHPH
PHHHPH	HHPPPH
PHHHHP	HPPPPH
HPPHHH	HPPHPP
HPHPHH	PHPPHP
HPHHPH	PHHPPH
HHPPHH	PPPPPP



Hexamer folding in 2D square and triangular lattice. Note that with the current fitting function the same optimal score can be obtained with close packed (globular) chains and with low density (relatively extended) ones, even for short sequences as the conformers shown in this figure. This suggest the inclusion of packing indexes in the fitting function.

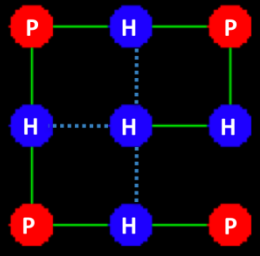
Algorithm performance

Sequence 1
HPPHPHPHP



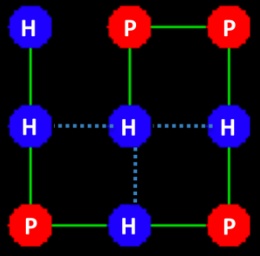
Score -3
4H/5P

Sequence 2
HHPHPHPHP



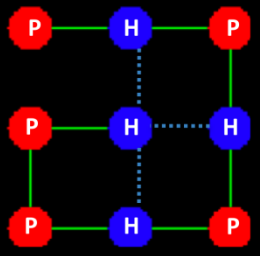
Score -3
5H/4P

Sequence 3
HPPHPHPHH



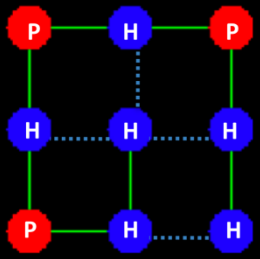
Score -3
5H/4P

Sequence 4
HPPHPHPHP



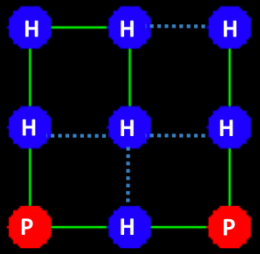
Score -3
4H/5P

Sequence 5
HHPHPHPHH



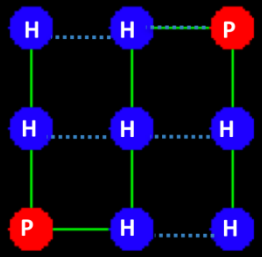
Score -4
6H/3P

Sequence 6
HHHHPHPHH



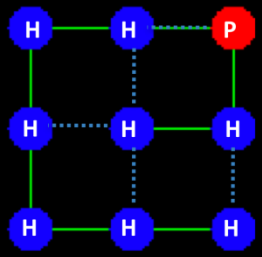
Score -4
7H/2P

Sequence 7
HHPHHHPHH



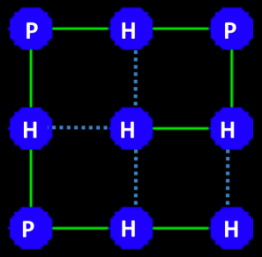
Score -4
7H/2P

Sequence 8
HHHHHHPHH



Score -4
8H/1P

Sequence 9
HHHHHHHHH

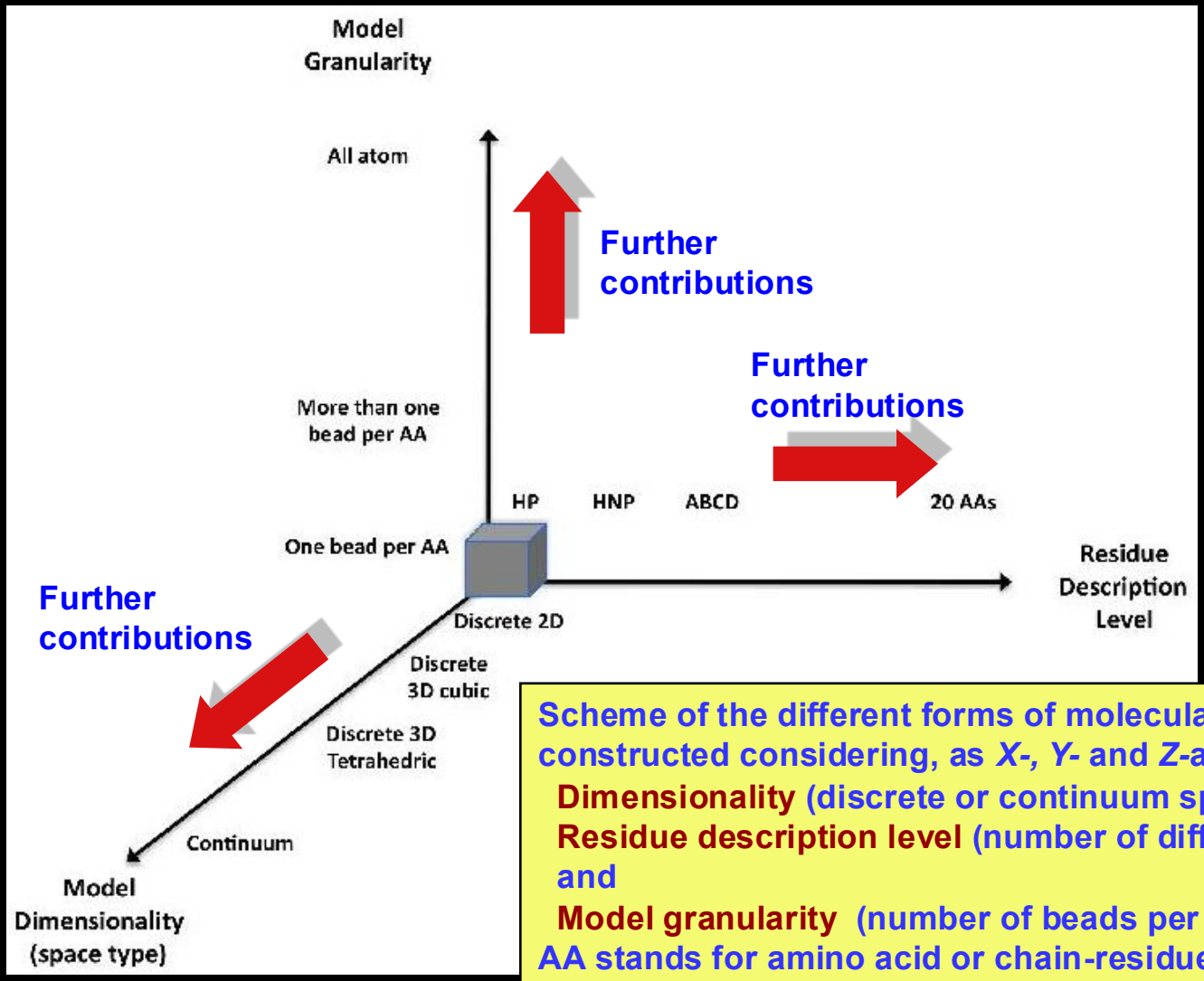


Score -4
9H/0P

Folded conformations of varying sequences differing not only in their H/P ratio but also in the positioning of the residues. The sequences were constructed trying to yield the highest scores possible upon folding into a 2D square, in order to test the performance of the platform.

Conclusioni

Description of the protein folding problem in the Evolution virtual laboratory



Scheme of the different forms of molecular folding representation constructed considering, as *X*-, *Y*- and *Z*-axes, the model:

Dimensionality (discrete or continuum space),

Residue description level (number of different amino-acid types), and

Model granularity (number of beads per monomer), respectively.

AA stands for amino acid or chain-residue.

H, N and P for hydrophobic, neutral and polar residues.

Conclusioni

Dagli esperimenti effettuati possiamo concludere che:

- Esperimenti relativamente semplici possono produrre una buona quantità di informazioni per interagire con la piattaforma bioinformatica e forniscono un feedback per il suo adeguamento al particolare problema ad essere esplorato.
- La piattaforma bioinformatica è stata progettata per integrare eventuali aggiornamenti, come quelli suggeriti dai risultati dei vari esperimenti.
- La piattaforma è in grado di fornire e integrare una serie di utili strumenti permettendo al utente (bioinformatico o biologo) di esplorare una grande quantità di possibilità nel problema sotto studio.

Produzione scientifica

- González, P.P., Beltrán, H.I., Rojo-Domínguez, A. and Sánchez, M. (2008) *An adaptable multi-agent based virtual laboratory to explore complex problems in structural biology. Starting case study: simplistic protein folding,* to be published at Journal of Computational Biology.
- Rojo-Domínguez, A., Beltrán, H.I. and González, P.P. *A Virtual Laboratory of Genetic Algorithms for Chain Folding*; Proceeding of USA-Mexico Workshop on Biological Chemistry: Multidisciplinary Approaches to Protein Folding, 2009.
- Beltrán, H.I., Rojo-Domínguez, A., Sánchez, M.E. and González, P.P.; *Exploring dimensionality, systematic mutations and number of contacts in simple HP ab-initio protein folding using a blackboard-based agent platform;* Proceedings of World Academy of Science, Engineering and Technology, Volume 40, April, 2009.
- Rojo-Domínguez, A., González, P.P. and Beltrán, H.I. (2009) *Feedback Experiments on Protein Folding using an Adaptable Virtual Laboratory. Folding 27-mer HP chains in a Cubic Lattice, it has been sent to Biosystems.*