



# Computational Nanoscience: *From Clusters to Proteins via Nature- Inspired Computation*

Roy L. Johnston  
School of Chemistry



UNIVERSITY OF  
BIRMINGHAM



# Overview

- Nature-inspired Computation
  - Genetic Algorithms
  - Ant Colony Optimization
  - Artificial Neural Networks
- Clusters
  - The Cluster GA
  - Bimetallic "Nanoalloys"
- Studies of Model Protein Folding
  - HP Lattice Bead Model
  - 3D Diamond Lattice
  - BLN Model
- Investigation of Energy Landscapes



# Nature-inspired Computation

## 1. Genetic Algorithms

- Examples of class of **Evolutionary Algorithms**.
- Procedures for optimising a function, structure or process which depends on a large number of variables.
- Based on principles of natural **evolution**.
- Consist of **mating (crossover)**, **mutation** and “**natural selection**” operators.

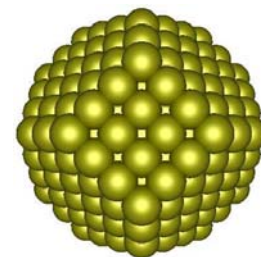
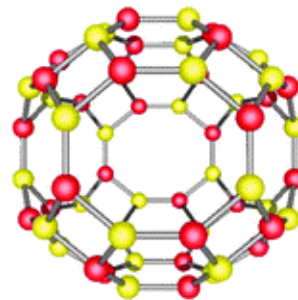


# Some Applications of Genetic Algorithms to Global Optimization Problems in Chemistry

## Geometry Optimization of Clusters

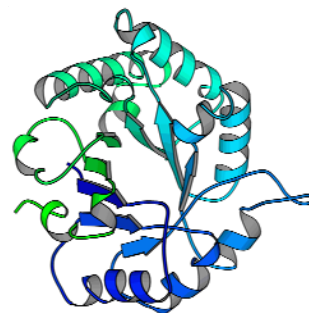
C. Roberts, RLJ, N.T. Wilson, *Theor. Chem. Acc.* **104**, 123 (2000).

RLJ, *Dalton Trans.* 4193 (2003).



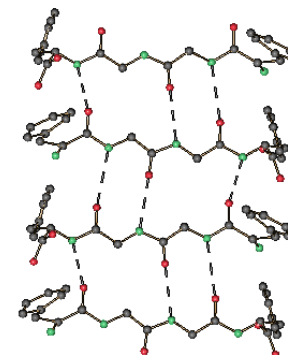
## Studies of Model Protein Folding

G.A. Cox, T. V. Mortimer-Jones, R. P. Taylor, RLJ, *Theor. Chem. Acc.* **112**, 163 (2004).



## Structure Solution from Powder Diffraction

K.D.M. Harris, S. Habershon, E.Y. Cheung, RLJ, *Z. Kristall.* **219**, 838 (2004).

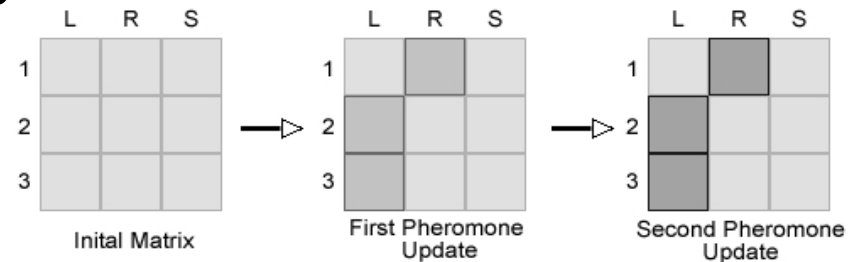
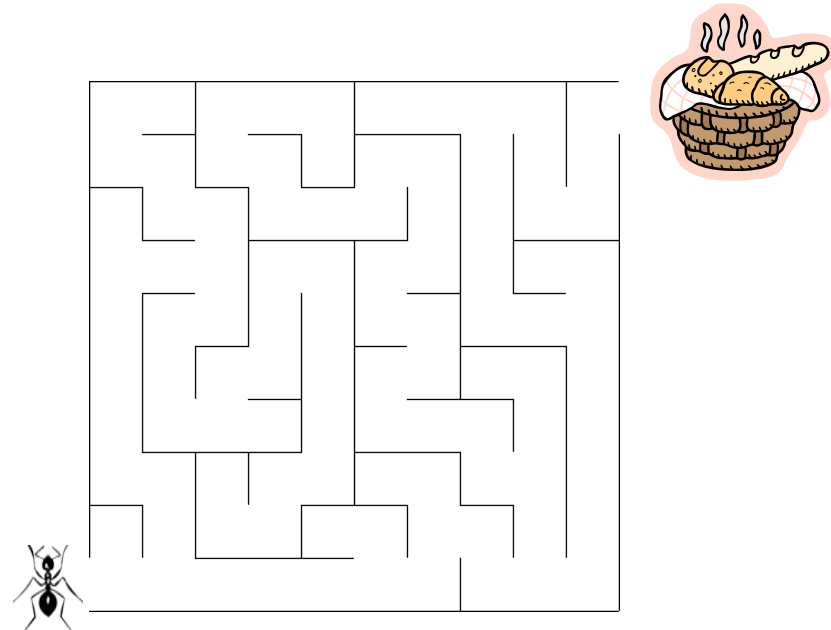




# Nature-inspired Computation

## 2. Ant Colony Optimization

- Based on foraging abilities of ants.
- Can be applied to problems which can be formulated as a walk on a lattice, decision tree *etc.*
- Ants lay down a **pheromone trail** (which decays with time). The more pheromone a path has, the more likely another ant will follow it.
- We have applied ACO to the protein folding problem.



Tom Mortimer-Jones, Ben Curley  
Graham Cox

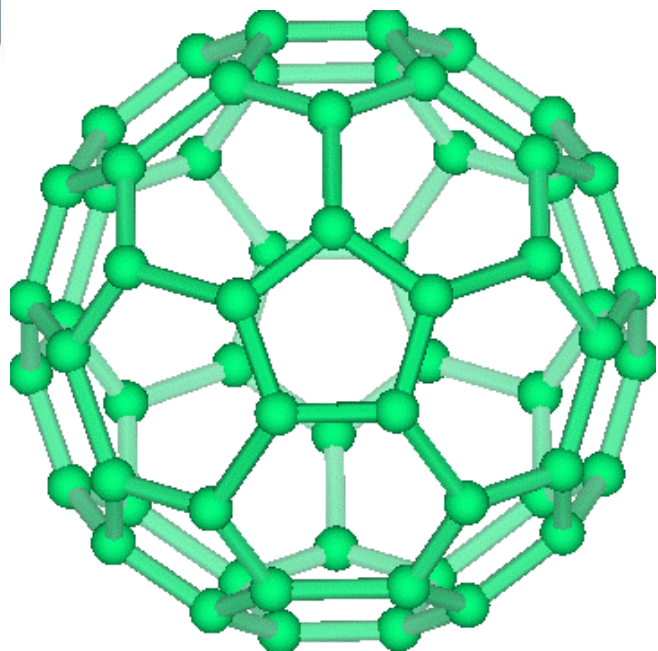


# Nature-inspired Computation

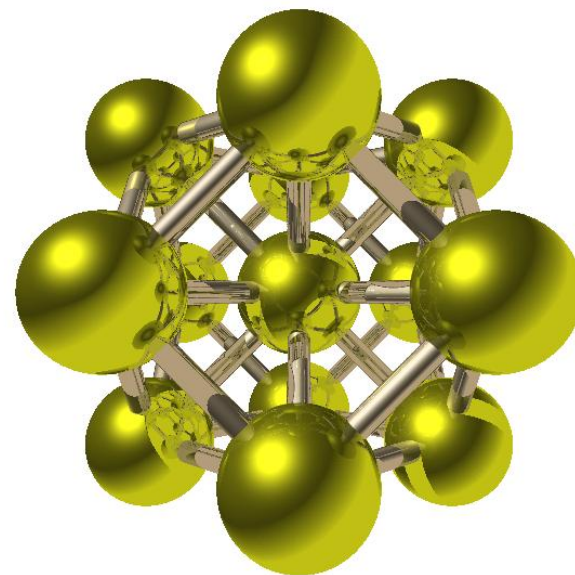
## 3. Artificial Neural Networks

- Based on model of neuron activity in the brain.
- We have applied an ANN to the problem of indexing unit cells from powder diffraction data.

*S. Habershon, E.Y. Cheung, K.D.M. Harris, RLJ, J. Phys. Chem. A 108, 711 (2004).*

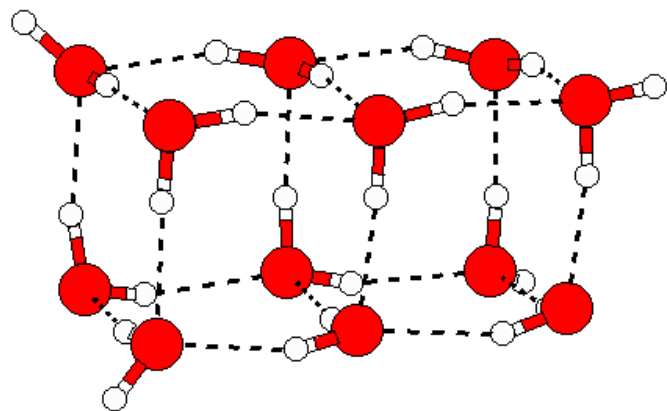


Fullerenes

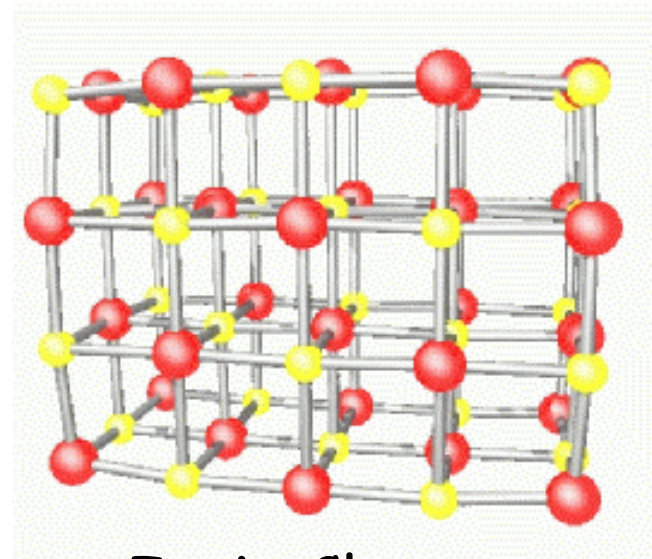


Metal Clusters

# Clusters



Molecular Clusters



Ionic Clusters



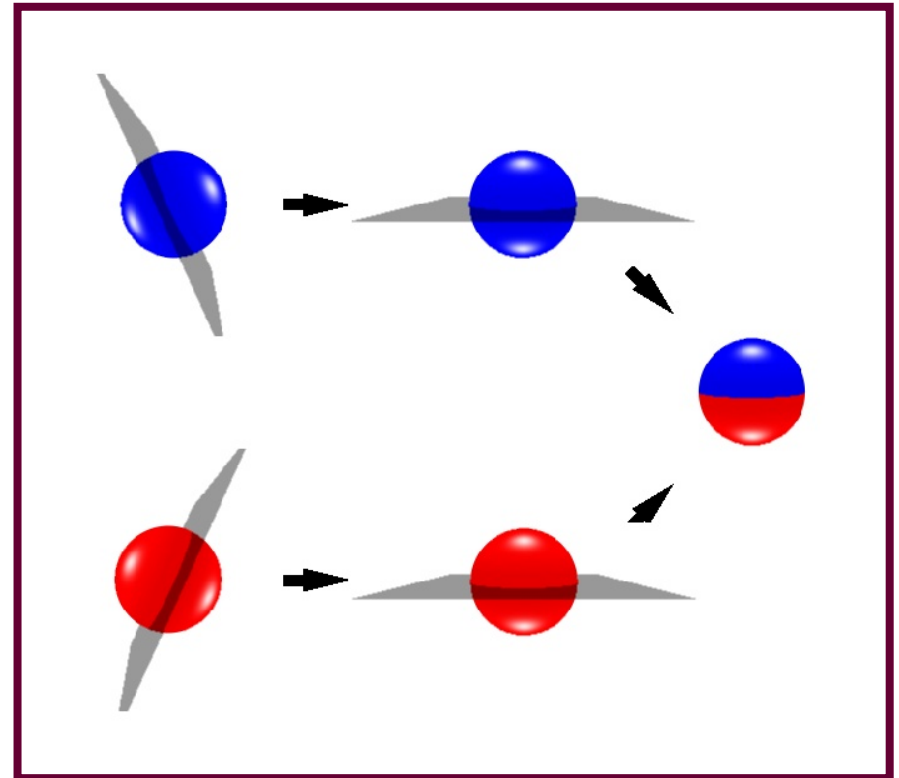
# Why search for the global minimum?

- To check trends in most stable structures – as a function of size, composition etc.
- To gain understanding of the nature of the potential energy surface.
- To determine stability of GM relative to other minima.
- To test that the interatomic potential, electronic structure calculation *etc.* is physically reasonable.



# The Cluster GA (1)

- Generate **initial population** of  $N_{\text{clust}} = 10-30$  clusters randomly:  $\{x,y,z\} = 0 \rightarrow N^{1/3}$
- Perform **local energy minimization** using L-BFGS quasi-Newton algorithm.
- Apply “**cut and paste**” phenotypic **crossover** operator:  
$$N_{\text{mate}} \approx 0.8 \times N_{\text{clust}}$$
- **Crossover Variations:**
  - ◆ take half of atoms from each parent
  - ◆ random cutting plane
  - ◆ fitness-weighted cutting plane
  - ◆ two cutting planes





# The Cluster GA (2)

- **Mutation** achieved by randomly perturbing a proportion of clusters.
- Mutation Types:
  - ◆ replace cluster
  - ◆ move a fraction of atoms
  - ◆ twist operation
  - ◆ permute unlike atoms
- **Relax** offspring and mutants by local minimization.



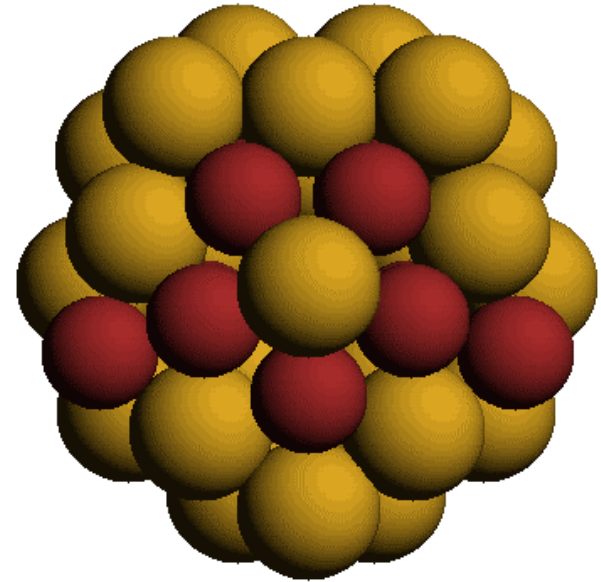
# Predator Operator

- Eliminates members of the population based on energy, shape, connectivity *etc.*
- The **energy predator** involves elimination of the previous lowest energy isomers from successive runs of the GA.
- This enables low-lying metastable isomers to be found.
- A predator can also be used to remove “**dead end**” or “**bottleneck**” structures or to remove identical or similar structures.



# Bimetallic "Nanoalloy" Clusters

- **Nanoalloys**: clusters of two or more metallic elements.
- A wide range of combinations and compositions are possible.
- Bimetallic nanoalloys ( $A_aB_b$ ) can be generated with controlled size ( $a+b$ ) and composition ( $a/b$ ).
- Structures and degree of A-B **segregation/mixing** can depend on method of generation.
- Nanoalloys can be generated in cluster beams or as colloids or by decomposing bimetallic organometallic complexes.



- Important in catalysis, and for opto-electronic & magnetic properties.



# Simulating Nanoalloys

- **Nanoalloys** are modelled using empirical many-body potentials – such as the **Gupta many-body potential** – and DFT calculations.

- Search for lowest energy isomers (**GM**) and low-lying metastable isomers using **Genetic Algorithms**.

*L.D. Lloyd, RLJ, S. Salhi, J. Comp. Chem. 26, 1069 (2005).*

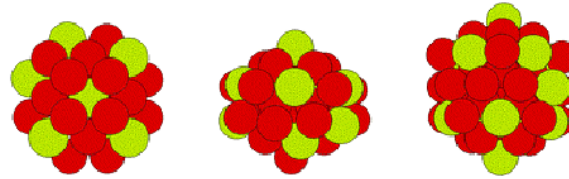
- For larger clusters ( $N \rightarrow 1000$ ), systematic studies are made for high-symmetry (*e.g.* icosahedral or cuboctahedral) structures.

*N.T. Wilson, RLJ, J. Mater. Chem. 12, 2913 (2002).*



# Example Systems

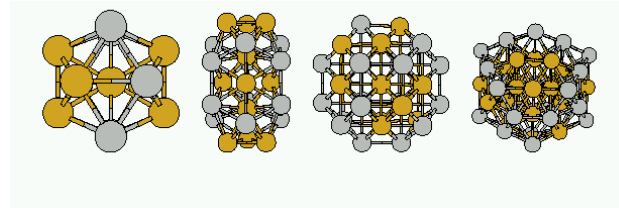
Ni-Al



Cu-Au



Ag-Au



Pd-Pt





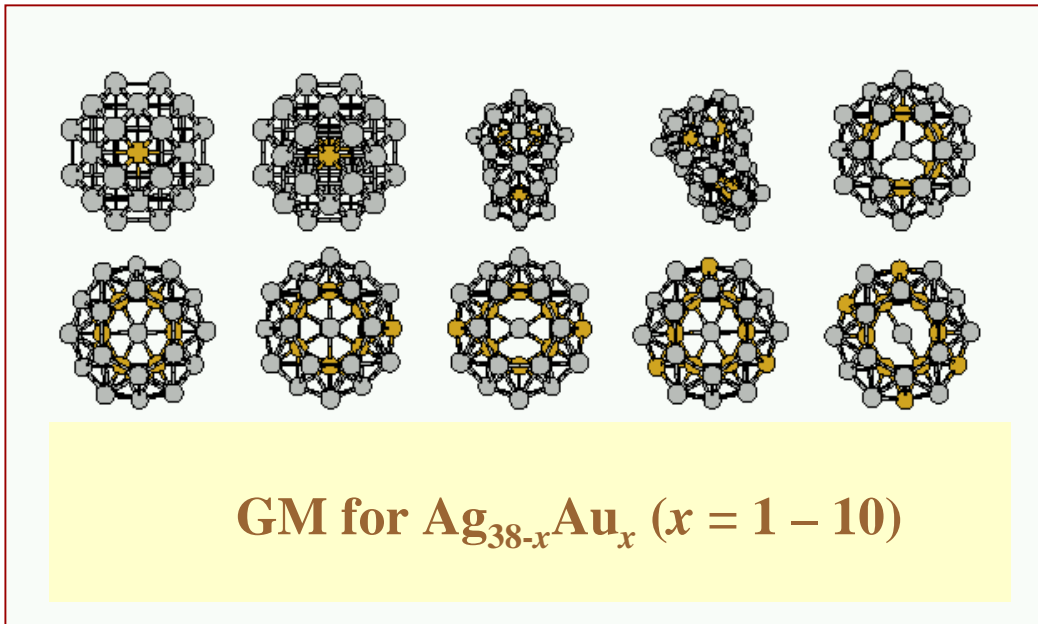
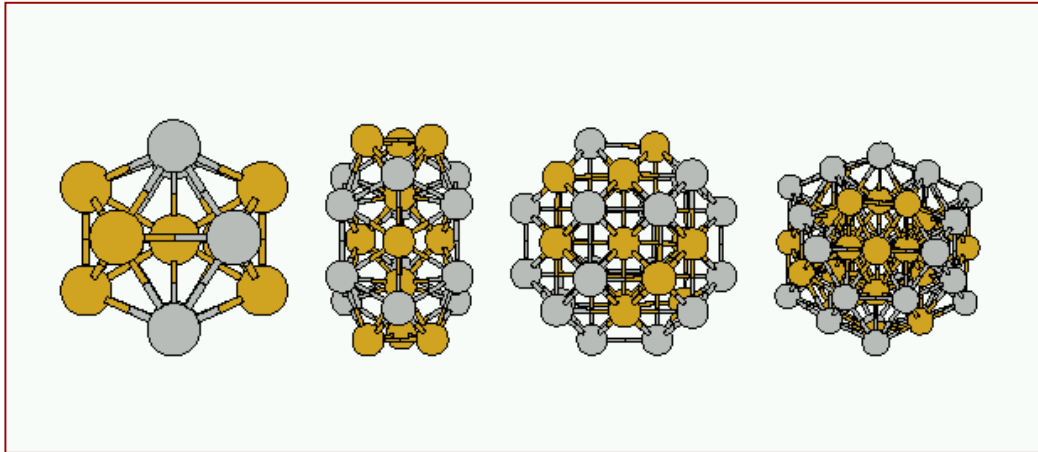
# Ag-Au nanoalloys

G. Rossi, R. Ferrando, A. Rapallo, A. Fortunelli, B.C. Curley, L.D. Lloyd, RLJ, *J. Chem. Phys.* **122**, 194309 (2005).

- Ag, Au and all Ag-Au bulk alloys exhibit fcc packing.
- In the bulk, Ag-Au forms solid solutions for all compositions (**no ordered phases!**).
- Mixing is weakly exothermic.
- There is experimental interest in how the shape and frequency of the plasmon resonance of Ag-Au clusters varies with composition and segregation/mixing.
- Recent STEM studies (Li, Wilcoxon) of core-shell Ag-Au clusters indicate a degree of inter-shell diffusion.
- This is consistent with XAFS measurements and MD simulations by Shibata *et al.* on core-shell (Au)Ag clusters.



# Structural motifs for Ag-Au clusters



- **Ag-Au:** segregates so the surface is richer in Ag (lower surface energy) and the core is richer in Au (strongest M-M bonds) – though the bulk alloy is a solid solution for all compositions.



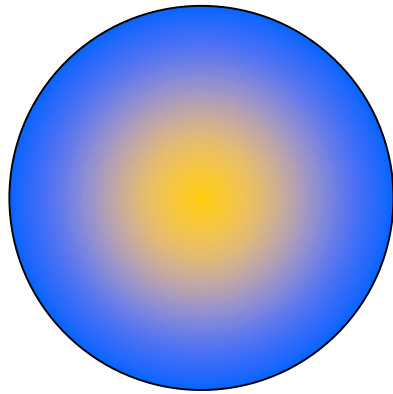
# Atomic ordering in $A_aB_b$ nanoalloys

Depends on:

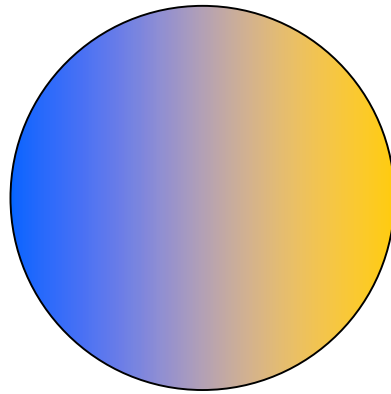
- Relative strengths of A-A, B-B and A-B bonds
- Surface energies of bulk elements A and B
- Relative atomic sizes
- Charge transfer
- Strength of binding to surface ligands (surfactants)
- Specific electronic/magnetic (quantum size) effects.



# Segregation patterns in nanoalloys

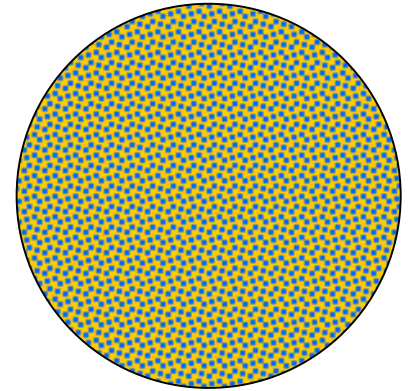


Core-Shell

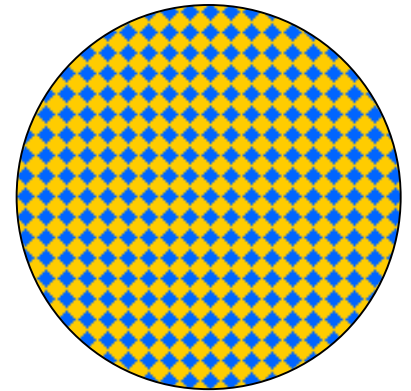


Layered

Random



Ordered



**Segregated**

**Mixed**



# Data Mining for Nanoalloys ?

- **Structure-Property Relationships**
  - **Structure**: size, geometry, composition, segregation
  - **Property**: energy, catalytic activity, physical and mechanical properties
- **Structure Prediction**
- **Potential-free Modelling**
  - **Neural Networks** – trained on experimental and *ab initio* calculated data

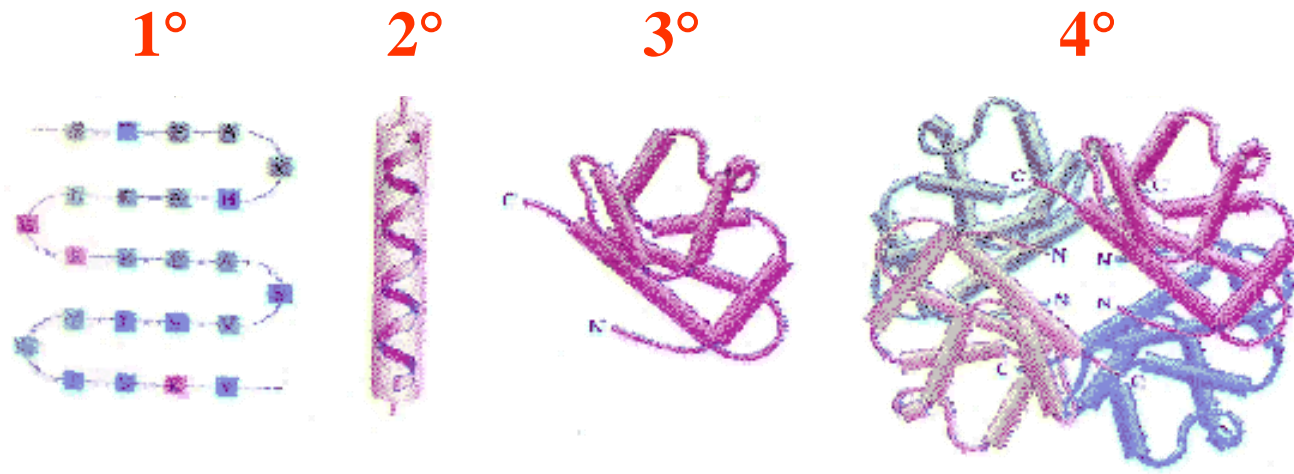
Josep Alsina (Sussex)



# Studies of Model Protein Folding

## The Protein Folding Problem

- To predict the 3D local spatial arrangement (**secondary structure**) and folded conformation (**tertiary structure**) of a protein from knowledge of its **primary structure** – the 1D **sequence** of amino acid residues.





# Protein Models

- **Bead Models**
  - **minimalist models** – each amino acid is represented by a bead, usually based on their **hydrophobic** or **hydrophilic** nature.
  - e.g. **HP** and **BNL** models.
  - beads may be constrained to a **lattice** or may be **off-lattice**.
- **United Atom Models**
  - with **backbone** and **side chain** beads.
- **All-Atom Models**
  - full **atomistic** treatment of protein (*e.g.* CHARMM).



# Features of the Protein Folding GA

- Initial valid conformations generated using a **Recoil Growth Algorithm**.
- **Brood** selection.
- 1-point crossover.
- Variety of mutation operators (local rotations, 1-point mutation, reptation).
- Monte Carlo **local search**.
- Diversity **predation**.

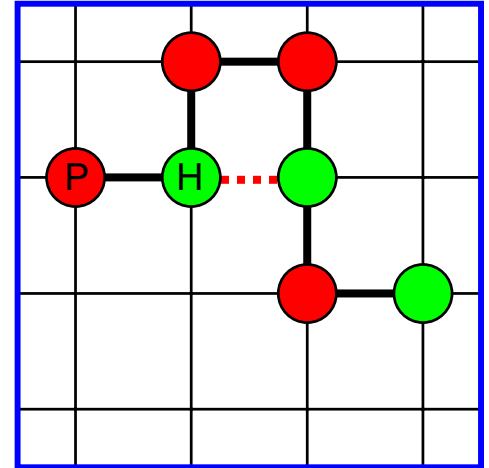


# The HP Lattice Bead Model

Amino acids are classed as either **Hydrophobic (H)** or **Polar (P)**.

Each amino acid is represented as a hard sphere (“bead”) on a lattice (e.g. 2-D square and 3-D diamond lattices).

Interactions occur between beads which are **adjacent on the lattice (topological neighbours)** but are **not directly bonded (sequence neighbours)**.



## Standard HP Model

$$\varepsilon_{HH} = -1 \quad \varepsilon_{HP} = \varepsilon_{PP} = 0$$

$$\Delta_{ij} = 1 \quad i \text{ and } j \text{ are topological neighbours} \\ \text{(but not sequence neighbours)}$$

$$\Delta_{ij} = 0 \quad \text{otherwise.}$$

$$E = \sum_{i < j} \varepsilon_{ij} \Delta_{ij}$$



# Comparison with Previous GA

Sequence	$E$ (min)	This Work		Unger & Moulton	
		%Success	$\langle N_{eval} \rangle$	$E$ (min)	$\langle N_{eval} \rangle$
HP-20	-9	100	18,338	-9	30,492
HP-24	-9	100	27,278	-9	30,491
HP-25	-8	100	35,128	-8	20,400
HP-36	-14	70	113,667	-14	301,339
HP-48	-23	13	261,311	-22	126,547
HP-50	-21	100	97,691	-21	592,887

•200 GA runs. Parameters: mating = 1.0, mutation = 0.1, elitism = 30%, DPL = 1, local search, brood size = 5.

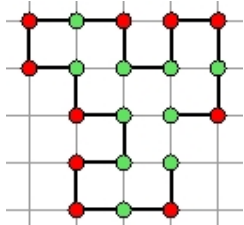
•Maximum generations = 100.

G.A. Cox, T. V. Mortimer-Jones, R. P. Taylor, RLJ, *Theor. Chem. Acc.* **112**, 163-178 (2004).

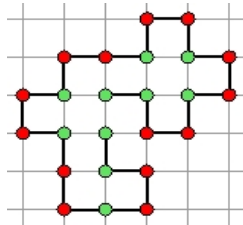


# Example Global Minima for Benchmark Sequences

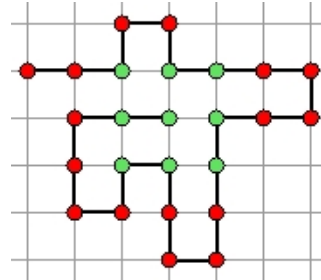
HP-20



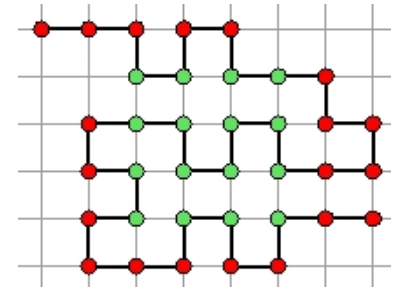
HP-24



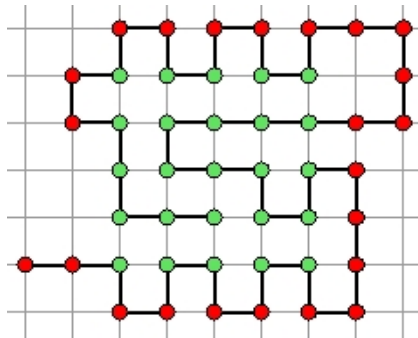
HP-25



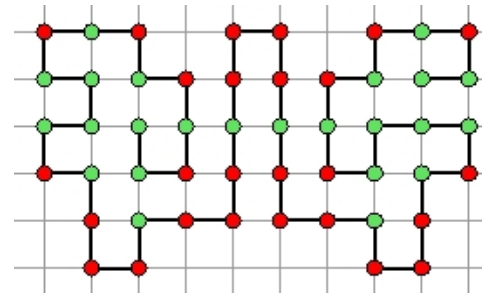
HP-36



HP-48



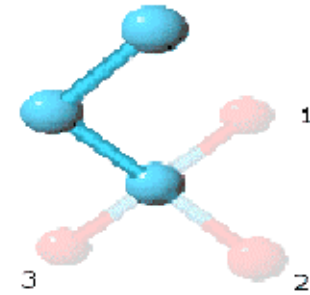
HP-50



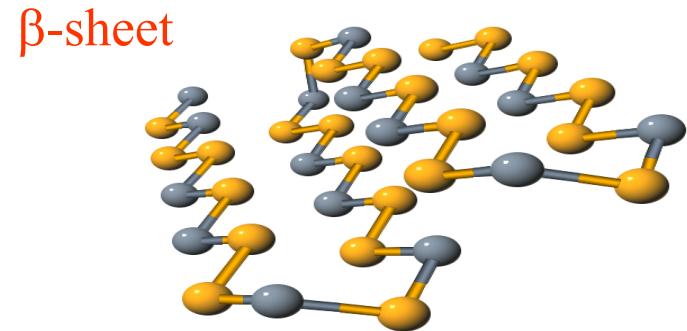
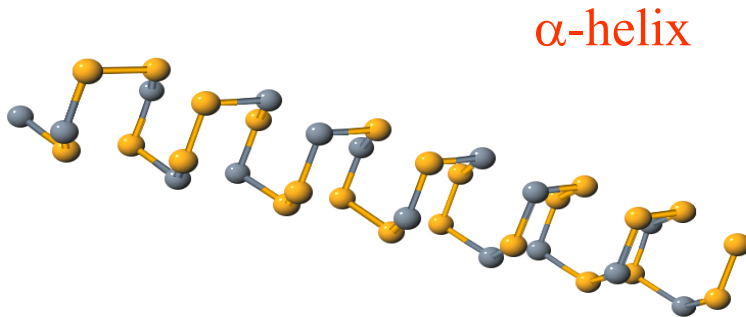


# HP Bead Model - 3D Diamond Lattice

- Same basic HP model but now beads lie on a 3-D diamond lattice.
- Again there are 3 choices for placing each new bead.
- Dimensionality of search space is the same – but for each  $N$  there are fewer invalid structures.



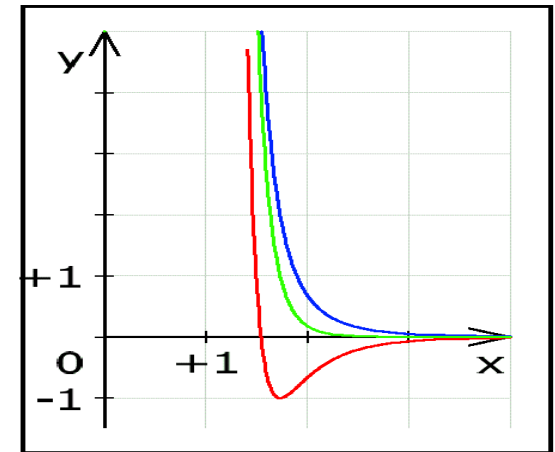
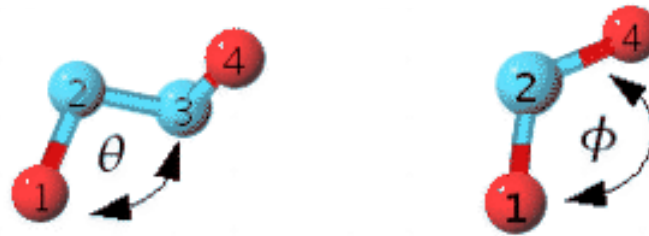
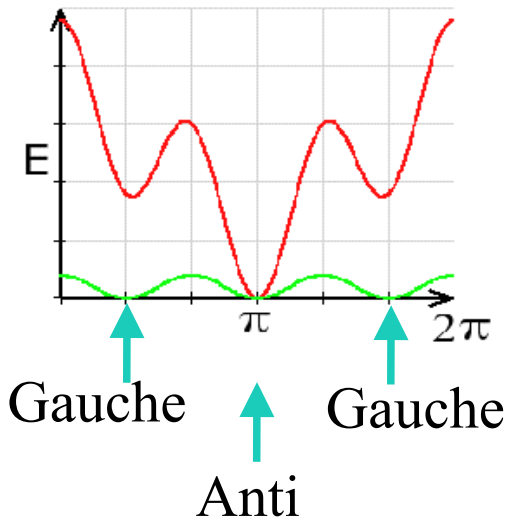
- More protein-like structures are possible.





# The BLN Model

- 3 Types of bead: Hydrophobic (**B**), Hydrophilic (**L**) and Neutral (**N**).
- Off-lattice model – has (bond  $r$ ) stretching, (angle  $\theta$ ) bending, torsional ( $\phi$ ) and through space (Lennard-Jones) components.

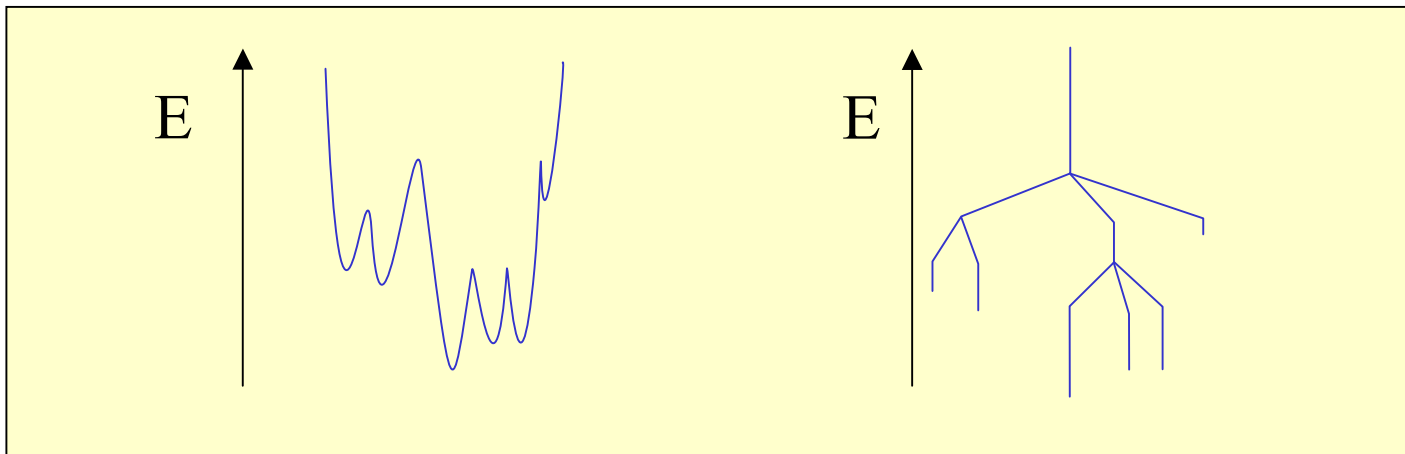


Gareth Rylance



# Investigation of Energy Landscapes for Protein Folding

- **Disconnectivity graph** approach (Becker & Karplus 1997) allows visualisation of the connectivity of high-dimensional PES (e.g. for protein folding).



- **BUT** – the  $x$ -coordinate has no meaning. Can more physically meaningful coordinates be obtained?



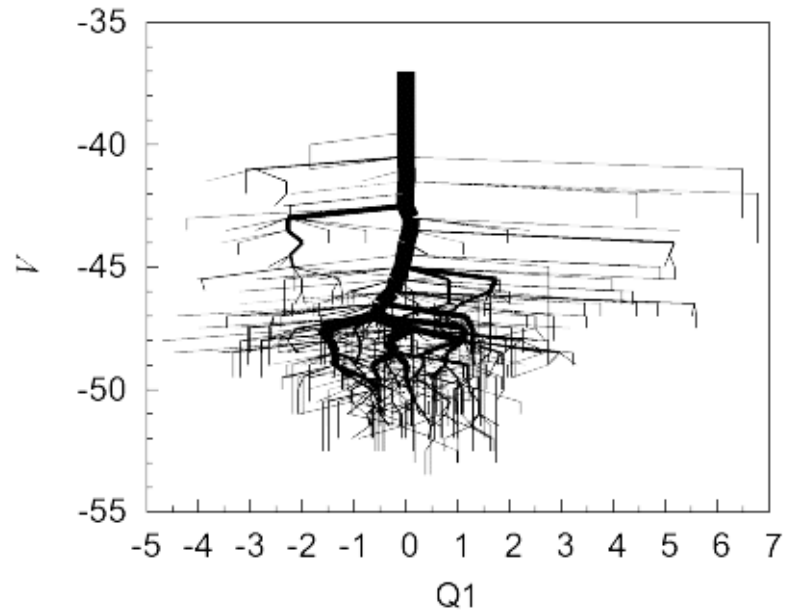
- **Principal Component (Coordinate) Analysis (PCA)**
  - Perform a linear transformation of coordinates of energy minima and transition states (1<sup>st</sup> rank saddles).
  - Identify the principal components – coordinates that maximise the variance of the system.
- **Example: 46-bead BLN model protein**
  - 1<sup>st</sup> PC contains approx. 30% of total variance.
  - 1<sup>st</sup> + 2<sup>nd</sup> PCs contain approx 45% of total variance.
- **1 and 2D disconnectivity graphs can be plotted against  $Q_1$  and  $Q_2$** 
  - Line thickness can be related to **number of structures** (local minima) or **structural diversity** within a superbasin.

T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G.J. Rylance, RLJ, D.J. Wales, *J. Chem. Phys.* **122**, 084714 (2005).



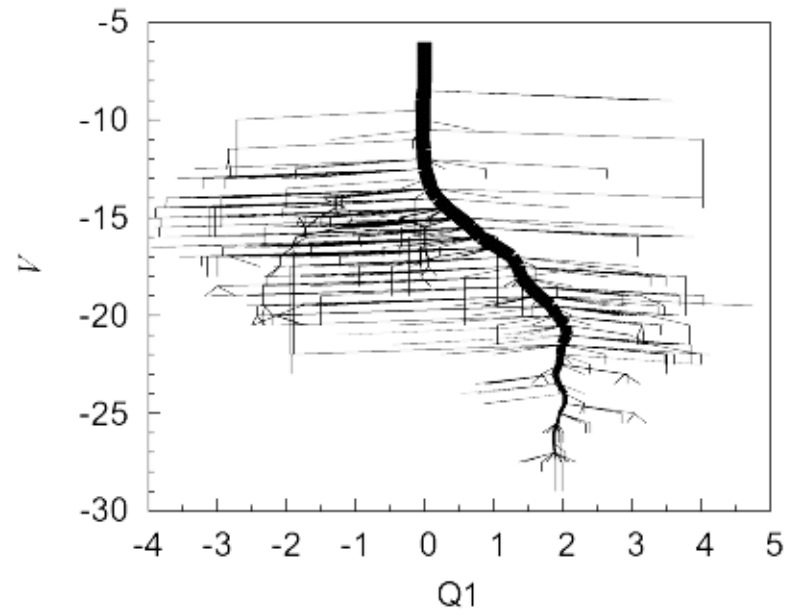
# BLN (46)

1D

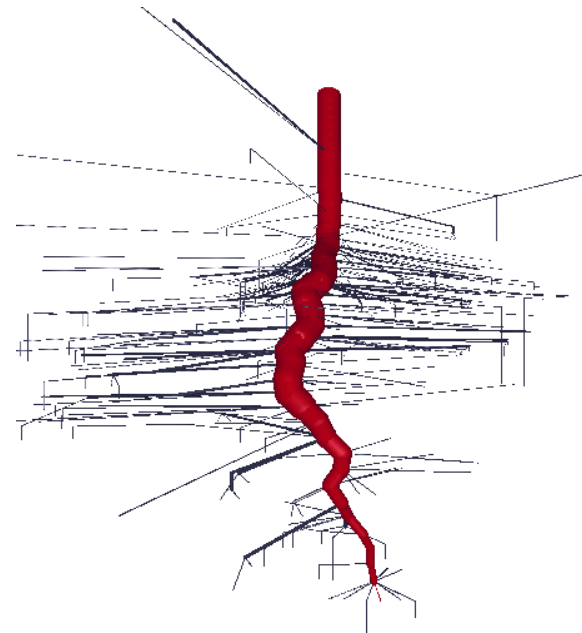
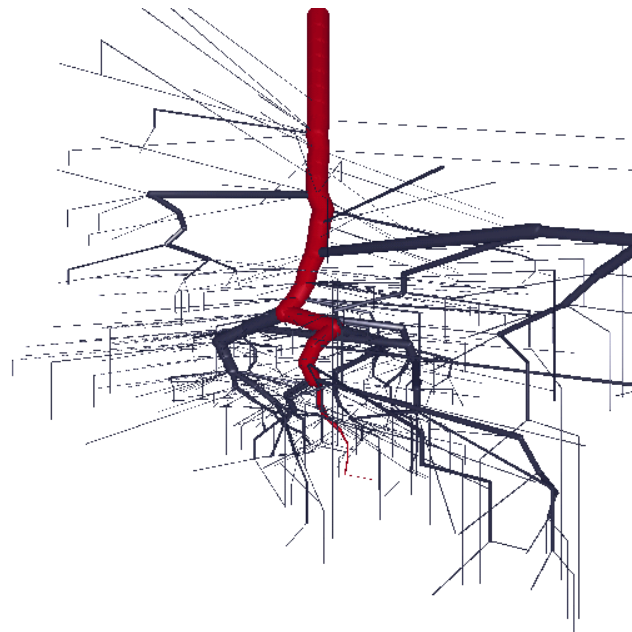


# Go

1D



2D





# Latest Work

- PCA-based disconnectivity graphs have been extended to:
  - **(ala)<sub>6</sub> peptide** – using CHARMM forcefield to compare PCA based on cartesian and internal coordinates ( $\omega$ ,  $\phi$ ,  $\psi$ ).
  - **Lennard-Jones LJ<sub>38</sub> clusters** – to explain the relative difficulty of finding the fcc-like GM compared to other low-lying icosahedral-like structures.



# Acknowledgements

- Dr Lesley Lloyd
- Dr Tom Mortimer-Jones
- Dr Chris Roberts
- Dr Nick Wilson
- Andy Bennett
- Graham Cox
- Ben Curley
- Gareth Rylance
- Dr Ziyou Li (Physics, B'ham)
- Prof. Said Salhi (Kent)
- Prof. Riccardo Ferrando (Genoa)
- Prof. Tamiki Komatsuzaki (Kobe)
- EPSRC
- The Leverhulme Trust
- The Royal Society
- Wellcome Trust
- University of Birmingham
- HP (Galway)