THE UNIVERSITY of NEW ORLEANS

Computational Research at UNO

10/31/2008

Presented by Zhiyu Zhao
on behalf of Scott Whittenburg
The University of New Orleans

UNO Computer Science

Stephen Winters-Hilt – CSCI/Children's Hospital/Bioinformatics

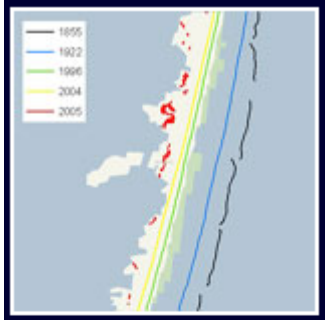
Vassil Roussev – CSCI/Information Assurance

Chris Summa – CSCI/Computational Biology & Bioinformatics

Dan Bilar – CSCI/Information Security (specifically network security)

Christopher Taylor – CSCI (presenting separately)

Maps and GIS



Modeling Software

Lab personnel and collaborators have experience with the following commercial and public domain computational fluid dynamics, coastal and oceanographic modes

Commercial Models

Surface Modeling System (SMS) 9.2: RMA2, RMA4, Full suite except CGWAVE

DELFT3D (hydrodynamics, Waves, Sediment Transport, Morphology) pending

FLOW3D - fully 3D CFD Model with Volume of Fluid Concept (VOF)

PCSWWM 2003 - Stormwater Management System

Public Domain and Research Models

Princeton Ocean Model (POM)

Estuarine Coastal Ocean Model (ECOMSED)

Finite Volume Coastal Ocean Model (FVCOM)

Advanced Circulation Model (ADCIRC)

EPA SWWM - US EPA Stormwater Management System

EPA NET - Steady State Flow In Pipe Networks

Visualization Software

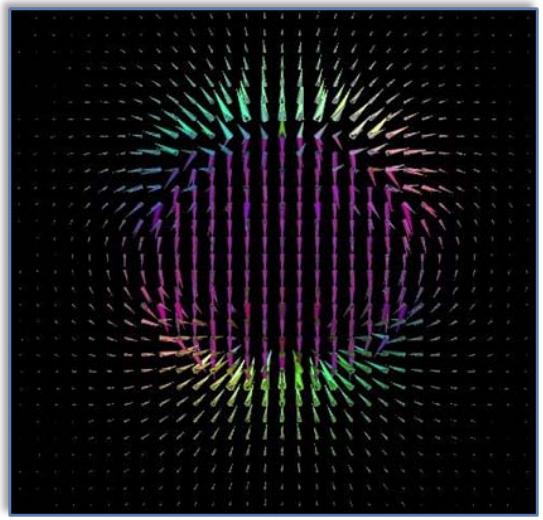
Multidimensional Visualization is Accomplished with TecplotC 360, MatlabC 6+, And SMS 9.2

Advanced Materials Research Institute

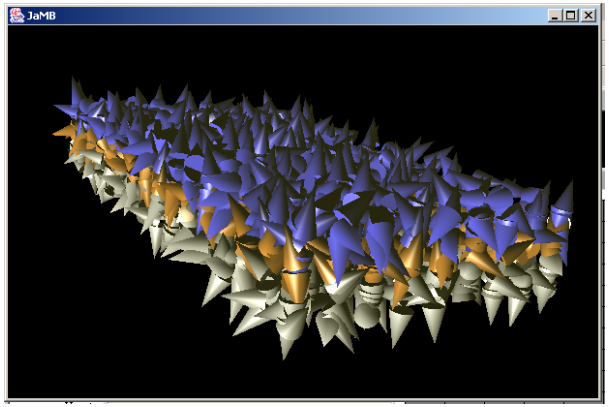
AMRI

Scott Whittenburg – Advanced Materials Research Institute (AMRI) - Micromagnetic computer simulations of nanometer-size magnetic materials used in magnetic recording media and sensors. The LONI supercomputer network will enable Dr. Whittenburg and his group to increase the detail and size of the magnetic regions leading to more realistic simulations and improved visualization.

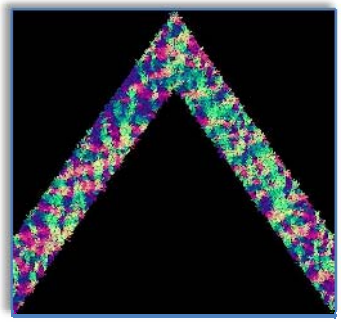
Spin Wave in Permalloy



Stray Magnetic Field Around Iron Nanosphere



Fe/Au/Py GMR Device



Remanent State of 4 Micron Cobalt ZigZag Wire

THE UNIVERSITY of NEW ORLEANS

Research on Protein 3-D Structure and Genome Sequence Related Problems*

10/31/2008

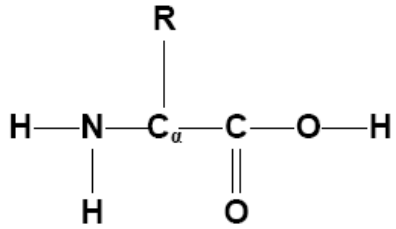
Presented by Zhiyu Zhao
The LONI Institute
&
The University of New Orleans

*Joint research with Dr. Bin Fu's group at the University of Texas – Pan American

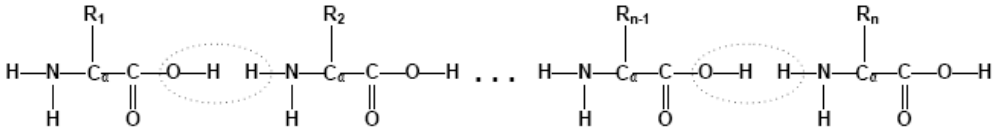
Outline

- **Protein structure related problems**
  - Protein 3-D structure alignment
  - Searching for similar protein structures in the Protein Data Bank
- **Genome sequence related problems**
  - Haplotype reconstruction
  - Genome rearrangement

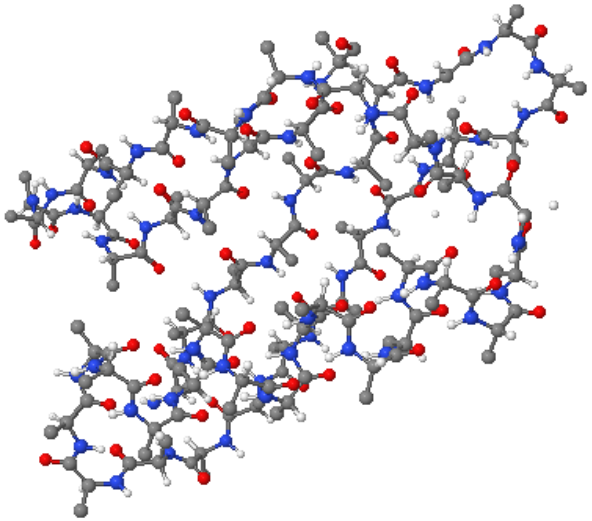
Protein 3-D structure

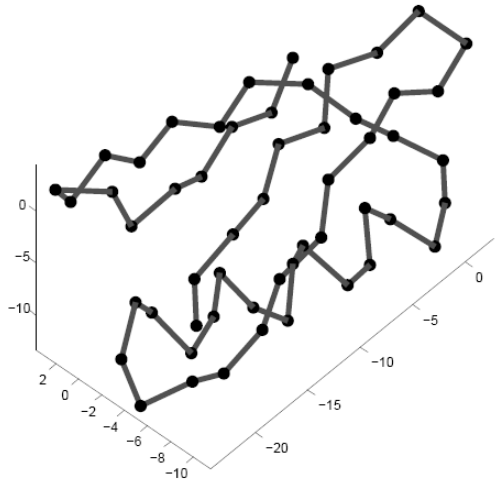


(a) The general structure of an α - amino acid



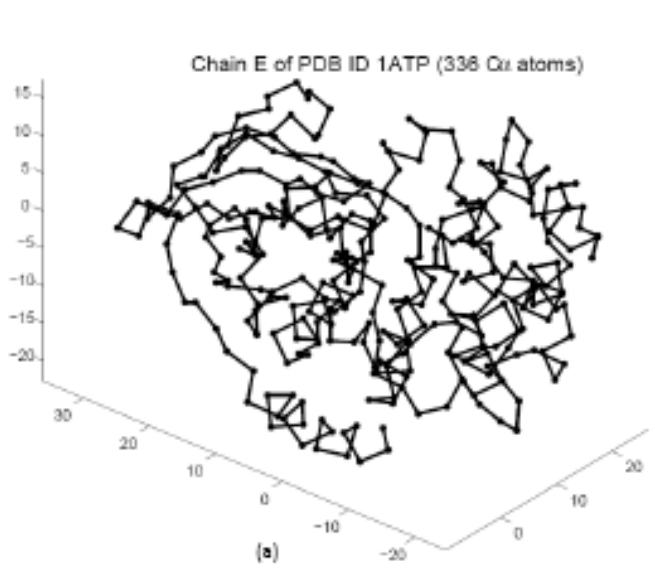
(b) The condensation of α - amino acids

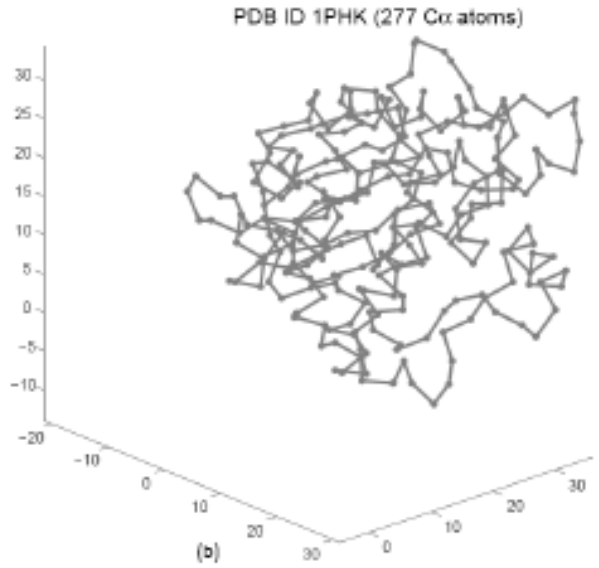


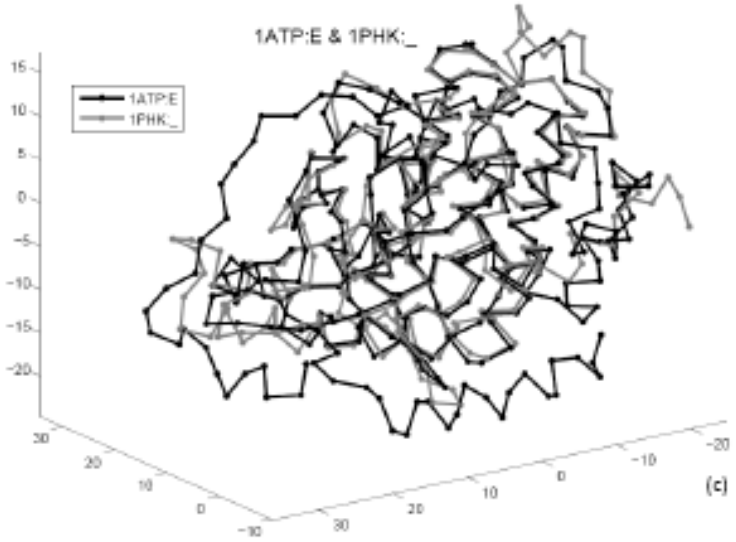


(c) A backbone representation of a protein chain with C_α atoms highlighted

Protein 3-D structure alignment







N_{mat} = 256

RMSD = 1.55Å

Translation vector :

T = [-6.8740 4.8834 18.6999]

Rotation matrix :

$$R = \begin{bmatrix} 0.7560 & 0.6393 & -0.1409 \\ 0.6061 & -0.6023 & 0.5194 \\ 0.2472 & -0.4781 & -0.8428 \end{bmatrix}$$

- Finding an optimal alignment is NP-complete
- Even approximate algorithms are very complex

Protein 3-D structure alignment

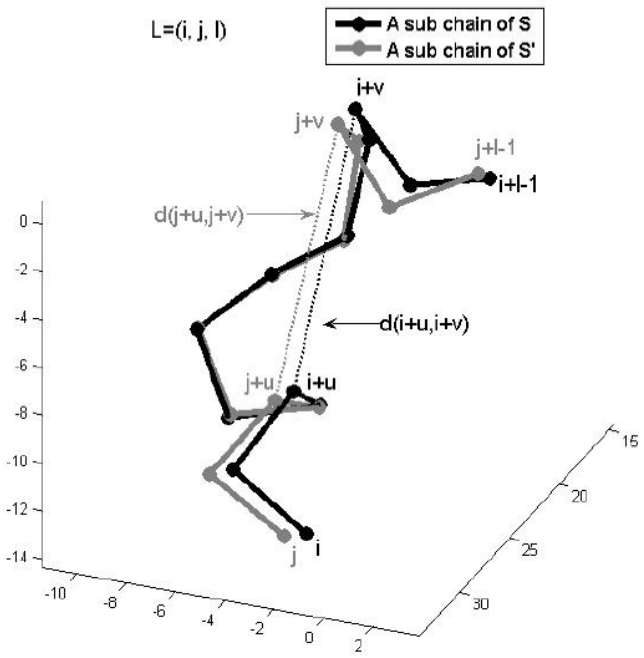
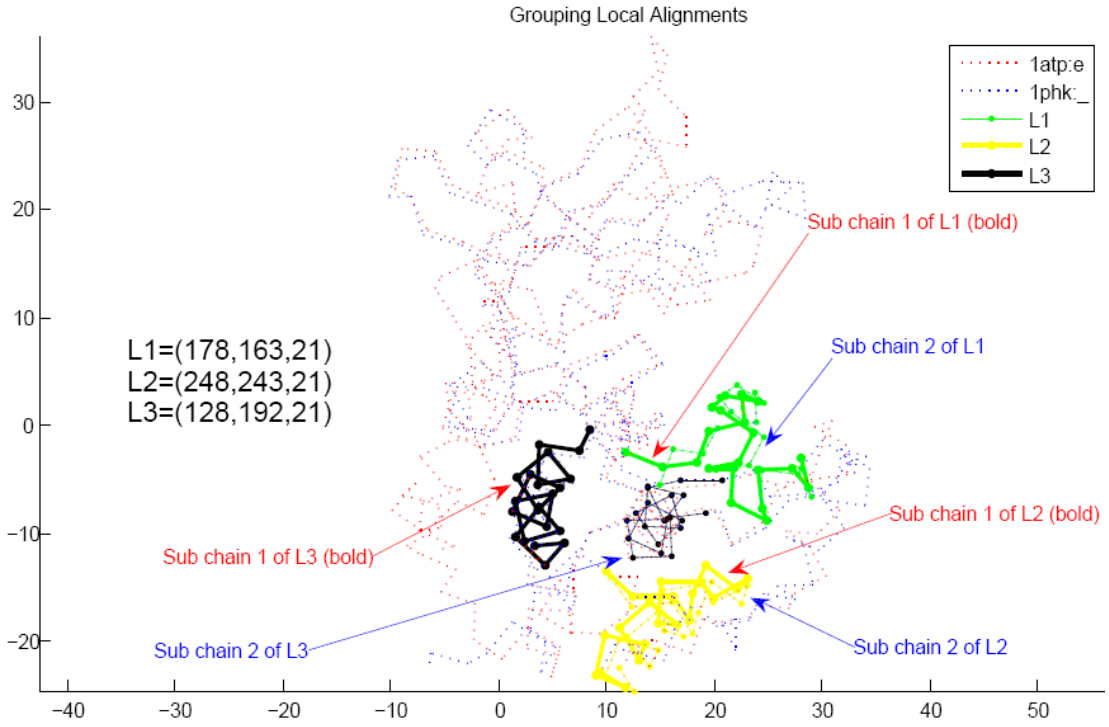
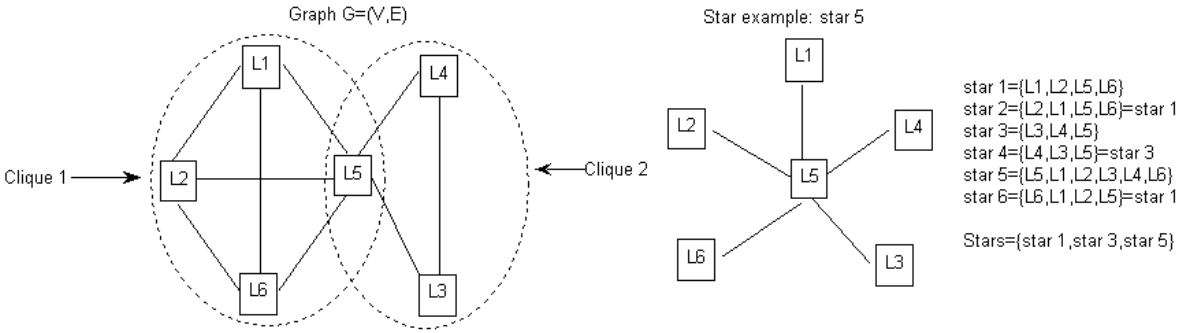


Figure 7: Local alignment L=(i, j, l)





Outline

- **Protein structure related problems**
  - Protein 3-D structure alignment
  - Searching for similar protein structures in the Protein Data Bank
- **Genome sequence related problems**
  - Haplotype reconstruction
  - Genome rearrangement

Finding Proteins with Similar 3-D Structures in the Protein Data Bank

- Given a protein structure S, find all the structures similar to S in the PDB .
- Application of protein structure alignment
- Difficulties
  - Over 100,000 chains in the PDB
  - alignment algorithms are not fast enough
- http://fpsa.cs.uno.edu/ http://fpsa.cs.panam.edu/
- Possible collaboration with LONI / TeraGrid

Outline

- **Protein structure related problems**
  - Protein 3-D structure alignment
  - Searching for similar protein structures in the Protein Data Bank
- **Genome sequence related problems**
  - Haplotype reconstruction
  - Genome rearrangement

Reconstructing haplotypes from SNP matrices with incomplete and inconsistent errors

- Genomes, SNPs and haplotypes
  - A genome: a complete sequence of DNAs; a string over the alphabet of {A, C, G, T}.
  - A SNP: a single nucleotide polymorphism between DNA sequences

Figure 3: A C/T Polymorphism between Two DNA Strands

DNA sequence 1: GCGCCAGTGGACTGCGTAGACCTATTTTCCAGCTGCGCCTGATGAAGGCG ...
DNA sequence 2: GCGCCAGCGGACTGCGTAGACCTATATTCCAGCTCCGCCTGATAAAGGCG ...
DNA sequence 3: GCGCCAGCGGACTGCGTAGACCTATTTTCCAGCTCCGCCTGATAAAGGCG ...
DNA sequence 4: GCGCCAGTGGACTGCGTAGACCTATTTTCCAGCTCCGCCTGATGAAGGCG ...
DNA sequence 5: GCGCCAGTGGACTGCGTAGACCTATATTCCAGCTCCGCCTGATAAAGGCG ...
...

Figure 4: SNPs in the single strand DNA sequences of individuals

Reconstructing haplotypes from SNP matrices with incomplete and inconsistent errors

- Genomes, SNPs and haplotypes
  - A haplotype: a set of SNPs on a particular chromosome copy

Figure 5: Two haplotypes of an individual

  - Importance of haplotyping individuals
    - A key step in the analysis of genetic variation
    - Help understand the complex causes of diseases
    - Individually tailored therapy design

Reconstructing haplotypes from a SNP matrix with incomplete and inconsistent errors

Haplotype 1: ABABABABBABBBABABBBABBBABBBABBAAABBBAAAABABAABABA

Haplotype 2: BBBBABABAABBBBAABBBABBABABAABBABBAAABBBAAABABAABABA

Dissimilarity between haplotypes: 20%

SNP matrix: incomplete error rate 10%; inconsistent error rate: 4%

```
-BBBABABAABBBBBABBBAABBABAABBA-AAAABBBAAABABAABABA      2
ABA---BBBABAAABABBBABABBBA-BBABB-AAB-B-AA-A-AABABA      1
BBBBABABAABBBB-ABBAA-BBABABABAB-AAA-BBAAA-ABAB-ABA      2
-B-BABABBABBBABABB-ABABBB-BB-ABBAAABBBAAAAABAABABA      1
ABABA--BBABBBABABBB-BABBBABABAB-A-ABBBAAAAABAABABA      1
ABA-ABABBAB-BABABBBAB-B-BBBBBABBAAABBBAA-AABBABABA      1
BBBBABABAABBBBAABBBBABBBBAABBABBAAABBBAAABAAAAAABA      2
A-AB-BABBABBBA-ABBBBBABBB-BBBABBAAABBBAAAAA-AA-A-A      1
BBBBABABAABBBBAABBBA-BBABAABBAB-AA-ABBAAABA-AABAAA      2
BBBBA-ABAABBBBA-BBB-A-BABAABB-BB-AABB-AAABAAAABABA      2
```

Figure 6: An SNP matrix with incomplete and inconsistent errors

Haplotype Reconstruction

- http://fpsa.cs.uno.edu/haprec/haprec.html
- Human genome size: 3,200,000,000
- Possible collaboration with LONI / TeraGrid

Outline

- **Protein structure related problems**
  - Protein 3-D structure alignment
  - Searching for similar protein structures in the Protein Data Bank
- **Genome sequence related problems**
  - Haplotype reconstruction
  - Genome rearrangement

Genome rearrangement

- Genome rearrangement
  - To compute the number of genetic operations to convert a source genome to a target genome
- Genome representation for rearrangement analysis
  - A genome
    - A sequence of n genes, each represented by a signed number between –n, …, -1 and 1, …, n.

    e.g. genome $\mathcal{G}$ = 1 -3 -2 1 4 8 5 2 -7 -6 5 9 10
  - An exemplar genome
    - A sequence of genes where duplicate genes are deleted.

    e.g. exemplar genome G = 1 -3 -2 4 8 5 -7 -6 9 10

Genome rearrangement based on Exemplar Non-breaking Similarity (ENbS)

- Exemplar non-breaking similarity
  - The number of non-breaking points between two exemplar genomes

e.g. exemplar genomes G = 1 2 3 4 5 6 7 8 9 10 and H = 1 -3 -2 4 8 5 -7 -6 9 10.
Non-breaking points in G are (2,3),(6,7) and (9,10).
Non-breaking points in H are (-3,-2),(-7,-6) and (9,10).
enbs(G,H) = 3.

- Problem definition
  - Given two genomes with a same alphabet, find two corresponding exemplar genomes to maximize the non-breaking similarity between them.
- NP-complete
- Possible collaboration with LONI / TeraGrid

Thank you!