A Fatgraph Model of Protein Structure

Carsten Wiuf
BiRC – Bioinformatics Research Center
University of Aarhus
DIMACS 2009, April 27-29

Bob Penner
Jorgen Ellegaard Andersen
Michael Knudsen
Short Intro and Aim

PROTEIN FATGRAPH MODEL
DESCRIPTIONS/ INVARIANTS CLASSIFICATION
Short Intro and Aim

PROTEIN
FATGRAPH MODEL
DESCRIPTORS/ INVARIANTS
CLASSIFICATION
Fatgraphs and Surfaces
(in math, originally due to Bob Penner)
Examples of Associated Surfaces

Euler characteristic
\[ \chi(F) = v(G) - e(G) \]
\[ \chi(F) = 2 - 2g - r \quad F \text{ orientable} \]
\[ \chi(F) = 2 - g - r \quad F \text{ non-orientable} \]

Moebius strip: Non-orientable
How to determine $g$ and $r$?
Permutations $\sigma$ and $\tau$ on stubs

\[
\sigma(G_1) = \sigma(G_2) = \sigma(G_3) = (1, 2, 3)(4, 5, 6)(7, 8, 9),
\]
\[
\tau_u(G_1) = (2, 8)(3, 6)(4, 7)(5, 9), \tau_t(G_1) = 1_S,
\]
\[
\tau_u(G_2) = (2, 8)(3, 6)(4, 9)(5, 7), \tau_t(G_2) = 1_S,
\]
\[
\tau_u(G_3) = (2, 8)(3, 6)(5, 9), \tau_t(G_3) = (4, 7).
\]
Protein to Fatgraph

Amino acid
Protein to Fatgraph

Peptide unit
Protein to Fatgraph
Building the Fatgraph

Twist vs Non-twist determined from the backbone
Protein Classification

• More than 50,000 known protein structures and 200,000 domains stored in PDB

• Protein Classification
  – CATH and SCOP; largely manual
  – Assisted by secondary structure knowledge

• Automated classification
  – Rogen and co-workers; geometric classification
The classification is hierarchical with the four main levels called **Class**, **Architecture**, **Topology**, and **Homology**, respectively.
CATH
Size of topology class in CATH

Alpha

Beta

Alpha-Beta

Few Secondary Structures
Mainly Alpha

Genus and Boundary
Wilcoxon
Significance p<0.005
Distorted Sandwich - 13 topologies (in "mainly beta")
Mainly Alpha – 24 largest topologies
(Nearest Neighbour with 25)
Mainly Alpha – 24 largest homologies
(Nearest Neighbour with 25)

76%
Classify “Unknown” Topology
(“Mainly beta”; 12 largest topologies)
Acknowledgement

• Joint work with
  – Bob Penner
  – Jorgen Ellegaard Andersen
  – Michael Knudsen
END
Ramachandran Plot