

Analysis of Protein Structure Using Euclidean Distance Matrix Analysis and DGEOM

Paul Richardson, Heather L. Gordon

Department of Chemistry, Brock University, St. Catharines, Ontario, L2S 3A1, Canada

Euclidean Distance Matrix Analysis (EDMA) is a method used to study the shape and form of structures.ⁱ EDMA has been used to compare statistically different structures in the bones of mammals. We want to know if EDMA can be used to differentiate between different tertiary structures of proteins.

The systems being used in our research are two different Protein Databank files. Both of the systems are insulin molecules, except one of the files, 1VKT, contains structures for wild type insulin, and the other, 1HIQ, contains structures for a mutant of insulin.

EDMA generates a mean inter-atomic distance matrix based on a sample of structures. Here we are using 10 conformations for each the wild type and the mutant, obtained from 2D NMR experiments. DGEOM is a distance geometry program. DGEOM is used to create (x, y, z) co-ordinates obtained from EDMA. EDMA is then used to generate a statistic to compare whether or not mean shapes obtained from two different samples are statistically different. This is done by using a statistical test. The computer programs that calculate these statistics are written in different programming languages, primarily FORTRAN 77 and AWK.

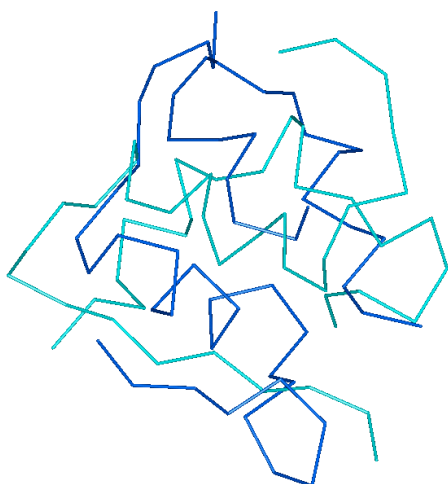


Figure 1: C α -trace of 2 Insulin Structures

Light Blue: 1VKT

Dark Blue: 1HIQ

Figure 2: Sequence Alignment of 1VKT and 1HIQ

1VKT: GIVEQ**S**CTSI**S**SLYQLENYCN

1HIQ: GIVEQ**C**CTSI**C**SLYQLENYCN

1VKT: FVNQHLCG**S**DLVEALYLVCGERG**F**FYTKPT

1HIQ: FVNQHLCG**S**HLVEALYLVCGERG**S**FYTPKT

ⁱ *An Invariant Approach to Statistical Analysis of Shapes*; Lele, S. R.; Richtsmeier, J. T.; Chapman & Hall; 2001