

**PS03.05.08 THE CHARACTERISTICS OF COMMON SEQUENCE PEPTIDES IN PROTEINS.** Luhua Lai, Dawei Lin, Youqi Tang, Institute of Physical Chemistry, Peking University, Beijing 100871, China

Characteristics of sequentially identical short peptides in a protein 3-dimensional representative data set and their relationship with protein folding were studied. Sequentially identical short peptides have the same intrinsic properties. They are ideal candidates for studying how short peptides conformation are influenced by their surrounding structural environment. We have carried out an exhaustive search on sequentially identical peptides in a protein structure representative set generated from the Brookhaven Protein Data Bank. The relationships among their sequence, secondary structure, protein fold class and short peptide structure adaptability in protein were analyzed. Six examples were found in the same protein fold class but having totally different structures. The proteins they are in have sequence identity less than 50%. This result implies that the conformation of some short peptides are influenced by more detailed structural environment than protein fold class, so that protein fold class dependent secondary structure prediction algorithm will still encounter the structural plasticity dilemma. From the sequence and structure analysis, we found that although some of the sequentially identical peptide pairs take different structures, most of the sequentially identical sequence peptide pairs have similar structure. A positive correlation was found between the accuracy of secondary structure prediction and the structure conservation of common sequence short peptides. The common sequence peptide pairs which preserve their structure in unrelated protein and different local structural environment have been proposed to be in the folding nucleation site or folding initiation site. One of them has found experimental evidence. The results of this study give helpful clues to protein secondary structure prediction and folding study.

**PS03.05.09 THE CUPREDOXIN FOLD: DEFINITION AND ANALYSIS OF THE COMMON CORE STRUCTURE.** Michael E.P. Murphy and Elinor T. Adman. Department of Biological Structure, University of Washington, Box 357742, Seattle, WA 98195-7742.

The cupredoxin fold is found in a family of copper containing proteins that include single domain electron transfer proteins and multi-domain multi-subunit enzymes that catalyze a variety of redox reactions including cytochrome c oxidase, nitrite reductase and the blue oxidase family. A method of multiple coordinate superposition was used to compare the three-dimensional structures of members of this family. Equivalent residues were identified by searching for a minimum of four consecutive residues with a r.m.s. deviation from the average coordinates under a defined cut off. The correctness of the alignment was validated by examining main chain hydrogen bonding patterns. A total of 53 alpha carbons (r.m.s. of 1.3 Å) were identified as being equivalent in 23 cupredoxin domains using a cut off of 2.5 Å. The core structure contains both beta-strands and loop regions. None of the helices present in these structures are conserved. A tree was constructed based on positional similarity of the core residue alpha carbons. This tree shows that domains of the multi-domain enzymes are more closely related between enzymes than domains within a single enzyme. Surprisingly, these relationships are independent of the metal content and function of these domains and suggests that domain duplication and differentiation occurred before functional specialization. The structure based tree agrees well with previously published phylogenetic trees. The sequence identity of the equivalent residues between all 23 domains is 19.5%. Despite this low sequence identity the resulting sequence alignment does contain recognizable patterns which may be of use in identifying other members of the structural family including proteins that do not contain copper.

**PS03.05.10 GAMMA: A NEW DOCKING PROGRAM UTILIZING AN ADVANCED EVOLUTION SYSTEM ALGORITHM AS AN ENGINE.** Steven Ness, Trevor Hart, Randy Read, Department of Medical Microbiology and Immunology, University of Alberta, Canada

Gamma (Genetic Algorithm for MacroMolecular-ligand Analysis) is a new program that uses an advanced Genetic Algorithm/Evolution System (GA/ES) engine for docking flexible ligands to macromolecular targets. Genetic Algorithms are powerful minimization/adaptation engines that have been used previously in many problem domains, and are beginning to be applied to the docking problem.

Previous approaches to the docking problem using GAs have mainly used the simple canonical GA. The canonical GA is defined as an algorithm that iteratively uses crossover and mutation operators on a population of bit-string organisms, selecting and reproducing organisms that have higher scores in terms of an objective fitness function. Good results have been obtained using this simplified evolutionary algorithm applied to the docking problem.

Theoretical research into GA/ES systems is progressing at a rapid pace, and new methods that eclipse the performance of the canonical GA are now appearing. Gamma utilizes some of these important new methods including: the use of integers and real numbers as genes, multiple gene types, prevention of lethal crossovers, hybrid minimization, multiple simultaneous runs with migration, and Demes (local neighbourhoods). These methods are combined together within an object-oriented framework that allows rapid prototyping and rapid design changes with little global program reorganization.

Preliminary results are very encouraging. Compared with a Monte Carlo Simulated Annealing engine, Gamma finds better (lower energy) solutions in considerably less time.

**PS03.05.11 MODES OF BINDING SYNTHETIC INHIBITORS TO FACTOR XA : AN AUTOMATED DOCKING APPROACH.** Mohan S. Rao and Arthur J. Olson, Department of Molecular Biology, The Scripps Research Institute, La Jolla CA

In order to understand the structural basis of Factor Xa specificity, complexes of synthetic inhibitors are generated using the AutoDock suite of programs (1). The inhibitors docked in the present study are 3-amidino benzyl phenyl ether, 4-amidino benzyl phenyl ether, 3-amidino benzyl pyruvic acid, 4-amidino benzyl pyruvic acid, dibasic amidino aryl propanoic acid, 3-amidinoaryl pyrrolidinyl phenyl propanoic acid. These inhibitors are different in size, nature of linkage and properties. Some of the lower energy conformers of these complexes together with their energy and possible hydrogen bond schemes are reported. Our results indicate the possibility of a hydrogen bond between amide hydrogen of amidino aryl moiety of these synthetic inhibitors and the side chain oxygen of Asp-189.

References:

- 1) Goodsell, D.S & Olson, A. J *Proteins: Str. Func. Genet.*, 8 (1990) 195