Rosetta is a de novo/comparative protein structure modeling algorithm. As one of the top-performing programs for protein structure prediction, it has predicted protein structures with high backbone accuracies. As a test case, we modeled the N-domain of Troponin C (NTnC) at both 4®"ƒ and 30®"ƒ, for which NMR structures are known. Rosetta-predicted structures of NTnC align with the NMR-determined structures at these two temperatures with ~ 3 Å backbone root mean square deviation. Our approach should be generally applicable to modeling temperature-dependent protein conformational rearrangements.

1178-Pos Board B70
Classification of Amyloidogenic Hexapeptides with Machine Learning Methods

Malgorzata Kotulska, Olgierd Unold, Jerzy Stanislawski.
Wroclaw University of Technology, Wroclaw, Poland.

Amyloids are proteins that form fibrils. Many of them underlie serious diseases, like Alzheimer disease. Recent studies indicate that amyloidogenic properties can be associated with short segments of aminoacids, which transform the structure when exposed. A few hundreds of such peptides have been experimentally found; experimental testing of all possible aminoacid combinations is currently not feasible. Instead, they can be predicted by computational methods. 3D Profile is a physicochemical method that has generated the most numerous dataset. However, it is computationally very demanding. Here, we show that dataset generation can be accelerated. Two methods to increase the classification efficiency of amyloidogenic candidates are presented and tested: simplified 3D profile generation and machine learning methods.

We generated a new dataset of hexapeptides, using modified 3D profile algorithm, which showed very good classification overlap with ZipperDB (93.5%). The new part of our dataset contains 1779 segments, with 204 classified as amyloidogenic. The dataset was applied for various machine learning methods. The most effective methods were Multilayer Perceptron and Alternating Decision Tree with areas under ROC curve of 0.96, accuracy of 91%, true positive rate of ca. 80%, and true negative rate 95%. A few other machine learning methods also achieved a good performance. We showed that the simplified profile generation method does not introduce an error with regard to the original method, while increasing the computational efficiency. Our new dataset proved representative enough to use simple statistical methods for testing the amylogenicity only based on six letter sequences. Statistical machine learning methods such as Alternating Decision Tree and Multilayer Perceptron can replace the energy based classifier, with advantage of very significantly reduced computational time and simplicity to perform the analysis. Additionally, a decision tree provides a set of very easily interpretable rules.

1179-Pos Board B71
N-Glycan Structure Modeling and in Silico Glycosylation: Template-Based Structure Prediction of Carbohydrate Structures of Glycoconjugates

Sunhwan Jo, Hui sun Lee, George Li, Jeffrey Skolnick, Wonpil Im.
The University of Kansas, Lawrence, KS, USA.

Obtaining the crystal structure of glycoconjugate is challenging due to the flexibility of the carbohydrate chains. Alternatively, computational modeling, which combines the primary sequence information of glycans determined by the mass spectrometry and known N-glycan structure, is an appealing approach. Here we present a survey of N-glycan structures of 35 different glycan sequences in the PDB, showing that N-glycan structures found on homologous glycoproteins are significantly conserved compared to the random background. This suggests that N-glycan chains can be confidently modeled to a glycoprotein if there exists a template N-glycan structure whose parent glycoprotein shares sequence similarity. On the other hand, N-glycan structures found on non-homologous glycoproteins have not shown significant structure similarity. However, despite that the global N-glycan structures are different, the internal substructure of those N-glycans found on the non-homologous glycoproteins, particularly, the substructure that are closer to the protein, showed significantly similar structure. Increased interaction with protein might be responsible to the restricted conformational space of N-glycan chains. Our results so far suggest that computational structure prediction of N-glycan portion of glycoconjugate using structure database would be effective, but different approaches must be needed depending on the availability of template structure. In addition, we also present a database for PDB glycan structural fragments (substructures) as well as PDB glycan-protein database, which are useful for glycan structure modeling.

1180-Pos Board B72
Adsorption of Bone Sialoprotein on Hydroxyapatite-A Combination Study with Bioinformatics and Molecular Dynamics Simulations

Yang Yang[1], Zhijun Xu[2], Qiang Cui[3], Nita Sahai[2].
[1]Rowan University, Glassboro, NJ, USA, [2]The University of Akron, Akron, OH, USA, [3]University of Wisconsin-Madison, Madison, WI, USA.

Bone sialoprotein (BSP), an acidic non-collagenous protein specific to bone, is proposed previously to modulate hydroxyapatite (HAP) nanocrystal growth. Two highly conserved phosphorylated acidic amino-acid sequences in BSP are hypothesized as the functional motifs. Specifically, we choose one of them, $(Sp)_2E_8$, where Sp represents a phosphoserine, as a model peptide to study the interactions between BSP and the HAP (001) face. A bioinformatics method helps predict the likely peptide conformations adsorbed on the HAP surface, which, subsequently, is subject to further examinations using molecular dynamics simulations with the explicit solvent model. The bioinformatics method predicts a Sp residue binds strongly to the surface, and the Glu residues show propensity to form a helical conformation. Long-time molecular dynamic simulations observe some variations of the sidechain orientations compared to the bioinformatics-predicted conformation, however, the backbone structure and the major binding features are largely preserved. In addition, no apparent geometrical templating between the peptide residues and the studied HAP surface sites is noticed, which implies that adsorption and subsequent crystal growth modulation by BSP may be structurally non-specific.

1181-Pos Board B73
Introducing Dinamo: A Package for Calculating Protein Circular Dichroism using Classical Electromagnetic Theory

Boris A. Sango[1], Neville Y. Forlemu[2], Sandeep Pothuganti[1], Rahul Nori[1], Yvonne E. Bongfen[3], Kathryn A. Thomasson[1].
[1]University of North Dakota, Grand Forks, ND, USA, [2]Georgia Gwinnett College, Lawrenceville, GA, USA, [3]Oklahoma Baptist University, Shawnee, OK, USA.

The dipole interaction model is a classical electromagnetic theory that has successfully been able to reproduce the experimental circular dichroism (CD) for the π-π* transitions for peptides and proteins. This theoretical model, pioneered by Jon B. Applequist, has been assembled into a package DInaMo that is written in C and Fortran allowing for treatment of whole proteins. The program reads Protein Data Bank formatted files of structures generated by molecular mechanics and molecular dynamics. Simple crystal structures need to at least be energy minimized for use in the model because they do not contain all the hydrogens. DInaMo reduces all the amide chromophores to points with anisotropic polarizability and all nonchromophoric aliphatic atoms to points with isotropic polarizability; all other atoms are ignored. By determining the interactions among the chromophoric and nonchromphoric parts of the molecule using empirically derived polarizabilities, the rotational and dipole strengths are determined leading to the calculation of the CD spectrum for each molecule. Theoretically predicted CD for a variety proteins (lysozyme, myoglobin, insulin, and collagen) are compared with synchrotron radiation CD data. Theory agrees with experiment showing bands with similar morphology and absorption maxima for the π-π* transitions.

1182-Pos Board B74
Bridging the Gap between Sequence and Function

Alexander Johs.
Oak Ridge National Laboratory, Oak Ridge, TN, USA.

Structural genomics initiatives have generated a massive quantity of high resolution structures and sequenced genomes from archaea, bacteria, viruses and eukaryotes. These numbers are expected to grow rapidly within the coming years. The available data constitute an invaluable resource for the prediction of protein structure and function and will allow us to obtain a more comprehensive structure-based understanding of biological function. The acquisition of new biochemical functionality in the course of evolution does not necessarily involve the transfer of whole genes, but may be limited to the transfer of functional domains. In fact, most of a protein's amino acids serve structural roles and may exhibit a low degree of sequence conservation even among closely related genomes. Therefore, it is critical to identify and compare structurally related domains across a wide spectrum of organisms to reveal unique metabolic functionalities.

This presentation will outline examples that have combined sequence alignments, homology modeling and biophysical approaches to predict function. Integration of models with existing knowledge about genomic context, biochemical pathways and sparse experimental data, such as small-angle X-ray scattering and spectroscopic data, enables us to accurately identify