27.1

Solution Structure of the Hypothetical Protein HP0495 from *Helicobacter pylori*

Chui-Lin Chiu, Ming-Tao Pai, C. C. Huang, Ya-Ping Tsao, and Jya-Wei Cheng

Institute of Biotechnology, Tsing Hua University, Taiwan, China

Structural genomics attempts to provide three-dimensional structural information about a significant fraction of the proteins encoded by the genes sequenced in various genome projects. In this study, we use NMR to determine the solution structures of HP0495, a hypothetical protein from the human pathogenic bacteria *Helicobacter pylori*. HP0495 has 4 similar sequences referred to as "hypothetical protein" by PSI-BLAST analysis against a non-redundant sequence database and is not associated with any known function based on its amino-acid sequence. Solution structure analysis revealed that the overall fold of this protein consists beta-strands and alpha-helices. We also investigate the conformational characteristics by hydrogen-deuterium exchange, and molecular dynamics. Furthermore, biochemical functions are analyzed to help the proteomic studies of the roles of *Helicobacter pylori* in human diseases.

27.2

A Special Genetic Algorithm for RNA Secondary Structure Prediction

Qinghua Cui, Bing Liu, Tianzi Jiang, and Songde Ma

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China

RNA's secondary structure plays an important role in knowing its function. With further investigation into the role and the structure of RNA, it has become quintessential to understand not only the primary structure of RNA but also the secondary and tertiary structures. Several methods have been tested in estimating these structures. Genetic Algorithms, a specialized algorithm of evolutionary computation, have proven a useful tool when simulating RNA secondary structure. The procedure is to simulate the RNA folding process by finding the most stable structure based upon the optimal minimum energy calculation. This paper proposes a special genetic algorithm explicitly designed to find RNA secondary structure. Differing from other approaches, this genetic algorithm combines methods not typically associated with genetic algorithms while still utilizing the typical genetic algorithm approach. The results provided a look into how different methods can improve the genetic algorithm used in secondary RNA structure prediction. And a comparison with traditional genetic algorithm is provided.

27.3

Protein Folding *In Silico*

Ulrich H. E. Hansmann

Department of Physics, Michigan Technological University, Houghton, MI, USA

The successful deciphering of the human genome has highlighted newly an old challenge in protein science: for most of the resolved protein sequences we do not know the corresponding structures and functions. Neither do we understand in detail the mechanism by which a protein folds into its biologically active form. Computer experiments offer one way to evaluate the sequence-structure relationship and the folding process but are extremely difficult for detailed protein models. This is because the energy landscape of all-atom protein models is characterized by a multitude of local minima separated by high energy barriers. Only over the last few years have been algorithms developed that allow one to overcome this multiple-minima problem in protein simulations. Prominent examples of these new and sophisticated techniques are *parallel tempering* and *generalized-ensemble* sampling. I will present applications of these methods to the simulation of small proteins such as the 34-residue PTH(1–34) and the 36-residue villin headpiece subdomain HP-36. We study for these molecules the folding mechanism and the relation between secondary structure formation and folding. Limitations of current energy functions are discussed.

[1] N.A. Alves and U.H.E. Hansmann (2000) *Phys. Rev. Lett.* **84,** 1836
[2] U.H.E. Hansmann and L. Wille (2002) *Phys. Rev. Lett.* **88,** 068105
[3] U.H.E. Hansmann (2003) *Comp. Sci. Eng.* **5,** 64
[4] C.-Y. Lin, C.-K. Hu, and U.H.E. Hansmann (2003) *Proteins: Structure, Function and Genetics*, **52,** 436
[5] U.H.E. Hansmann (2004) *J. Chem. Phys.* **120,** 417

Support by research grants of the National Science Foundation and the National Institutes of Health is acknowledged.

27.4

Unfolding and Aggregation Behavior of Amyloidogenic Proteins Unravelled by Combining Microelectroseparation, Spectroscopic, and Structural Proteomic Methods

N. H. H. Heegaard, T. J. D. Jørgensen, O. Trapp, N. Rozlosnik, J. S. Pedersen, D. B. Corlin, P. Roepstorff, and M. H. Nissen

Department of Autoimmunology, Statens Serum Institut, Copenhagen, Denmark

About 20 different proteins are known to cause disease in humans because they malfold and precipitate as insoluble fibrillar aggregates called amyloid. The consequences are serious since amyloid destroys the tissues that house the lesions as is the case with the brain in Alzheimer's disease and with articular structures in hemodialysis-related amyloidosis. Unraveling of pathological folding and aggregation is therefore of interest for disease proteomics. Here, we study the conformation of β_2-microglobulin (β_2m), a small protein involved in hemodialysis-associated amyloidosis. A specifically cleaved form of β_2m that may potentially be generated *in vivo* by cleavage at lysine-58 produced thioflavin T-stainable aggregates at 37°C in contrast to wild-type β_2m that remained soluble. Furthermore, the cleaved form was shown to be conformationally unstable using capillary electrophoresis and fluorescence spectroscopy. Aggregation was characterized by size-exclusion chromatography with light scattering detection and by atomic force microscopy to assess morphology and the effect of amyloid fibril seed addition. A marked increase in the unfolding rate constant of the cleaved β_2m as compared to unmodified β_2m was determined by capillary electrophoresis and the cooperative nature as well as the pronounced temperature dependency of the unfolding was revealed by hydrogen-deuterium exchange monitored by mass spectrometry. The protein and the methods used here constitute a paradigm for pathoproteomic investigations of protein conformational diseases. The detailed understanding of protein folding afforded by these studies will eventually pave the way for the development of novel methods to counteract malfolding processes and for methods to resolve amyloid lesions.

27.5

Crystal Structures of Membrane Proteins: Some Approaches and Some Results

N. Isaacs[1], I. Black[1], P. Blackburn[1], M. O'Hara[1], A. Roszak[1], C. Simpson[1], Y. Zhu[1], R. Cogdell[2], A. Gardiner[2], G. Graham[3], and R. Nibbs[3]

[1]Department of Chemistry, [2]IBLS, [3]Division of Immunity, Infection and Inflammation, University of Glasgow, Glasgow, United Kingdom

An area of structural genomics that is always recognised as being fundamentally important is membrane proteins. It is estimated that 30% of the genome codes for proteins inserted in the membrane. This figure reflects not only the importance of membrane proteins but also their wide range of biological activities as receptors, channels, transporters, etc. In 1985, when the first structure of an integral membrane protein was reported, there were only 268 protein structures available in the Protein Data Bank (PDB). Today, the PDB contains nearly 27,000 structures, but fewer than 200 of these are membrane proteins and only 80 are unique structures. Despite these low numbers, crystal structures of integral membrane proteins can be very informative as evidenced, for example, by our work on proteins from bacterial photosynthetic units where structures of light-harvesting complexes and reaction centres have revealed mechanisms of energy capture and transfer. Other studies on two different types of integral membrane proteins (a chemokine GPCR and a bacterial autotransporter) illustrate some of the difficulties of working with membrane proteins and indicate the need to develop high throughput methods using the availability of genomic data to provide a wider range of targets for study.

An exciting new development in the UK is the Membrane Protein Structure Initiative (Mpsi), which aims to use genomic data in developing high throughput methods for the expression, isolation, purification, crystallisation and structure determination of integral membrane proteins. Mpsi is a collaborative, formed between sixteen leading membrane protein structural biologists from eight Universities and Institutions. It is led from the University of Glasgow and has been awarded a five-year grant under the BBSRC's SPoRT initiative.

27.6
Structural Proteomics for Functional Discovery of Hypothetical Proteins

Byoung-Young Jeon[1], Bon-Kyoung Koo[1], Eunmi Hong[1], Jin-Won Jung[1], Adelinda Yee[2], Cheryl H. Arrowsmith[2], and Weontae Lee[1]

[1]Department of Biochemistry and HTSD-NMR & Application NRL, Yonsei University, Seoul, Korea; and [2]Ontario Cancer Institute, University of Toronto, Toronto, Canada

Recently, we initiated structural proteomics project targeting *Thermoplasma acidophilum,* which is one of the smallest organisms among archaeal genomes. In this report, the solution structures of two hypothetical proteins will be presented as an example of structural proteomics approach for functional discovery of the unknown gene products. For example, TA1092 protein belongs to the S24E family of 30S ribosomal proteins with 98 residues. However, the structure of ribosomal protein R32S in *achaeon* species was not determined yet. Therefore, the structural information will provide us an insight to understand its molecular functions. To characterize the biochemical properties of these proteins, biophysical techniques, such as NMR, CD and sedimentation equilibration experiments have also been applied. Multidimensional-heteronuclear NMR data together with automatic assignment tool have been used for resonance assignment. Solution structures were semi-automatically determined with CYANA. The secondary structure of TA1092 consists of two α-helices and four anti-parallel β-sheets. TA0956 protein is mainly composed of β-sheets and two small α-helices. The structure-function of these proteins will be also discussed related to their molecular functions.

27.7
Biomedical Structural Proteomics

Peter Kuhn

Scripps PARC Institute for Advanced Biomedical Sciences, The Scripps Research Institute, San Diego, CA, USA

Biomedical structural proteomics is aimed at identifying and experimentally characterizing novel targets for therapeutic intervention including the development of tool and small molecule inhibitor compounds. We have developed and are implementing technologies that enable biophysical characterization of individual proteins and protein protein interactions at high fidelity and minimal sample consumption. These biophysical tools determine both the 3-dimensional and thermodynamic interactions of the protein partners, and can determine modes of action and interaction. Technology advances made over the past few years have enabled a more exhaustive approach to the exploration of target selection, protein production, crystallization, and x-ray diffraction experiments. While these high-throughput automated systems are going into production in applications ranging from traditional structural biology, large multimeric complexes and structural genomics, there is a new need emerging concerning the required streamlining, automation, and miniaturization (reduction in sample consumption) of other biophysical methods to validate protein function and interaction with partners (signaling molecules, small molecule agonitsts/antagonists). This approach is currently being applied to a number of disease targets including SARS CoV related proteins.

27.8
Structural Alignment of Proteins by Curve Matching

Q. S. Li and C. B. Chen

College of Computer Science & Technology, Huazhong University of Science and Technology, Wuhan, P.R. China

The structural alignment between two protein structures has been studied using the technique of three-dimensional (3-D) curve matching. For the problem of structural alignment, proteins can be represented as 3-D curves defined by their backbones. The proposed B-Spline Structure Curve (BSSC), a new model of protein backbone, will exactly pass through the sampling atoms. Thus, the problem is changed to detect the most similar subcurves between two BSSCs. To solve the subcurve detection problem, the curvature at each sampling points on BSSC has been obtained by the theory of differential geometry. The curvature is a rotation and translation invariant signature of 3-D curve, and transforms the representation of protein structure from 3-D curve to 1-D sequence. The algorithm complexity of the similar subcurve detection procedure that offers the candidates of structural alignment is almost $O(n)$ by introducing the efficient hashing technique, where n is the number of sampling points on the curve that has more sampling points of the two matching curves. An additional subsection matching strategy has been adopted in the subcurve detection process. For each candidate alignment, the Root Mean Square Derivation (RMSD) is derived by the superimposition between the two atom sets that are corresponding to the detected sampling points. The ultimate alignment is the one which possesses the largest number of equivalent positions whose RMSD is less than the given threshold. This novel structural alignment approach opens new opportunities for the comparative analysis of protein structures and for more studies on protein folding principles and other issues about protein structure.

27.9

Proteomic Analysis of the Cellular Responses Induced by Hydroquinone in Human Embryonic Lung Fibroblasts

X. Y. Li[1,2], Z. X. Zhuang[1], J. J. Liu[2], H. Y. Huang[1,2], Q. Z. Wei[1,2], and X. H. Yang[1,2]

[1]Shenzhen Center for Disease Control and Prevention, Shenzhen, P.R. China; and [2]School of Public Health, Zhong Shan University, Guangzhou, P.R. China

Hydroquinone (HQ), a major metabolite of benzene, used widely as a reagent in photographic developers, as an antioxidant in manufacture of rubber, as a polymerization inhibitor for acrylic and vinyl acetate monomers, and in cosmetic products as a skin-lightening agent. But the mechanism of its effect on human cells is far from clear. In the present work, we studied on the cellular response induced by HQ using proteomic approaches. Human embryonic lung fibroblasts (HLF) were treated with 1×10^{-4} mol/L HQ for 24 h. This dose of HQ was assayed to significantly decrease cell viability. After treatment, cells were rinsed in magnesium acetate buffer then lysed in 500 μl of a buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, and 10 mM Tris. Supernatant were collected and precipitated with acetone prior to protein concentration determination. 2DE was performed using the Amersham Bioscience 2DE system following the manufacturer's instructions. For the first dimension, 1000 μg of total protein with a volume of 250 μl was subjected to isoelectric focusing (IEF) on IPG strip(pH 4~7L, 13 cm) with rehydration for 12 hr at 30 V, IEF for 1 hr at 500 V, 1 hr at 1000 V, 40000~45000 Vhr at 8000 V. Focused strips were equilibrated with reduction and alklyation, then subjected to SDS-PAGE for the second dimension in a 12.5% gel. Proteins were visualized by the staining with colloidal coomassie blue. Three separate experiments were repeated independently. Based on analysis with Imagemaster software, 490 $\pm$ 28 protein spots were detected in control HLF and 438 $\pm$ 23 in HQ-treated cells. Fifteen protein spots showed significant changes after HQ treatment. For protein identification, selected gel spots were excised and distained, then dehydrated with acetonitrile. After completely drying, trypsin was incorporated into the spots to swell on ice. Then digestion buffer without trypsin was added to submerge the gel piece. After overnight incubation at 37°C, tryptic peptides were extracted from the gel pieces then concentrated in a SpeedVac. Eleven protein spots were identified by peptide mass fingerprinting using MALDI-TOF or by peptide sequence tagging using MALDI-TOF-TOF. Among them including transaldolase, growth factor receptor-bound protein2, mutant β-actin,γ-actin, Lasp-1, TAR DNA-binding protein and similar to neural precursor cell expressed protein. These include proteins involved in oxidative stress, cellular signaling, RNA splicing and cytoskeleton reconstruction. Most of their involvements in the cellular responses to HQ have not been reported. Therefore, our findings may offer new insights into the mechanisms of HQ cytotoxicity and these proteins may serve as new biomarkers for detecting exposure of human populations to HQ. It is suggested that proteomic approaches may provide new strategies to evaluate the toxicity of xenobiotics.

27.10

A Two-dimensional Reference Map of Mouse Ovaries Proteins

Xiang Ma[1,2], Yun Qian[2], Yan Meng[2], Man-Xin Xie[2], Zhen Hou[2], Jin-Yong Liu[2], Ye-Fei Zhu[1], Yun-Dong Mao[2], Zuo-Min Zhou[1], Jia-Yin Liu[2], and Jia-Hao Sha[1]

[1]Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, P.R. China; and [2]Center of Repeoductive Medicine, First Affiliated Hospital of Nanjing Medical University, P.R. China

We used two-dimensional gel electrophoresis (2-DE) and matrix assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) to perform a preliminary proteomic analysis of mouse ovaries. A 12.5% sodium dodecyl sulfate (SDS) reference gel was generated by immobilized pH gradient isoelectric focusing of mouse ovaries proteins in a non-linear gradient (pH 3–10). Based on peptide mass fingerprinting, 172 proteins were identified and classified into seven functional groups: cell division proteins, cell signaling/communications proteins, cell structure/motility proteins, cell/organism defense proteins, metabolism proteins and unclassified proteins. This work provides a first step toward the establishment of a systematic ovaries protein database and stands as a valuable resource for molecular analyses of normal and pathologic conditions affecting mouse ovaries.

27.11

A Perspective on Structural Genomics Efforts in China—A Report, Review and Revision

Zihe Rao

Laboratory of Structural Biology, Tsinghua University & National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

The Structural Genomics initiative in China was proposed by the founder of Chinese Protein Crystallography, Professor Dong-cai Liang, immediately after the Yokohama Structural Genomics Conference in 2000.

The program in China was initially founded by the Chinese Academy of Sciences, China National Natural Science Foundation and the Ministry of Science and Technology of China. Structural biologists together with protein chemists in China formed two consortiums: one by the related institutions from Chinese Academy of Sciences and the other one by Universities. The main target genes of the Chinese group are primarily human with particular emphasis on disease and novel bacterial genes.

After two years efforts, including a break during the 2003 SARS outbreak, it is now the stage for us to report on the progress of Structural Genomics in China, and also to review and revise our program. According to the reports by both University and CAS groups, approximately 2,000 targets have been selected and cloned; 369 proteins have been purified; 157 proteins crystallized; and 118 structures have been solved (including NMR structures).

27.12

Structural Genomics Efforts for Proteins from Human and from Other Higher Eukaryotes

Y. Y. Shi

Hefei National Laboratory for Physical Sciences at Microscale, School of Life Science, University of Science and Technology of China, P.R. China

Structural genomics efforts at the Chinese Academy of Sciences are reported. The major targets for this structural genomics project include proteins expressed in human hematopoietic stem/progenitor cells, proteins related with blood diseases. Late on the targets are extended to other proteins from human and other higher eukaryotes that may related with medicine or may have important biological functions. We also interest small protein complexes. Up to now, 1306 targets genes have been constructed in expression vectors. 134 proteins have been purified for structure determination. 62 Crystals are obtained, 40 structures are determined, 31 by x-ray crystallography and 9 by NMR. The platforms for bioinformatics, gene cloning, protein production, isotope labeling, sample preparation and crystal growing and screening have been established. The major bottleneck still is to get enough proteins suitable for crystallization trials and suitable for NMR experiment, especially when these proteins from human and other higher eukaryotes. In this lecture I will give examples focusing protein structure determination and protein- protein interactions study by NMR.

27.13

Post-translational Modifications Characterisation Using MS^n on a MALDI-QIT-TOF MS

Chris Sutton[1], Rachel Martin[1], and Koichi Tanaka[2]

[1]Shimadzu Biotech, Wharfside, Trafford Wharf Road, Manchester, United Kingdom; and [2]Koichi Tanaka Mass Spectrometry Research Laboratory, Shimadzu Corporation, Kyoto, Japan

The MALDI QIT (quadrupole ion trap) TOF mass spectrometer allows detailed characterisation of post-translational modifications. The instrument combines the virtually infinite analysis time of MALDI, MS^n functionality of the ion trap, and resolution and accuracy of time-of-flight. Phosphopeptides were particularly labile during ionisation by mass spectrometric analysis with partial loss of 98 Da in MS mode from the parent, and virtually complete loss in MS^2 mode. However, MS^3 of the dephosphorylated peptide ion formed in MS^2 mode allowed the characterisation dephosphorylated serine, threonine and tyrosine using nested (comparison of the MS^2 and MS^3 spectra) MS^n analysis. Oligosaccharides represent a challenging class of biological structures for analysis by mass spectrometry due to the branched structures, anomeric linkages (α and β glycosidic bonds) and isobaric forms. The MALDI QIT TOF mass spectrometer was used to find novel diagnostic signals correlating to specific oligosaccharide fragments that could be confirmed through MS^n analysis. Analysis to MS^5 mode was critical to define unique mono- and di-saccharides building blocks that could be correlated with the original parent mass and therefore the intact structure of the oligosaccharide. This process was extrapolated further to the analysis of glycopeptides, which enabled site-specific glycosylation information to be obtained retained in the analyte. MS^2 analysis was used to sequence the peptide moiety and identify the product ion specific to the cleavage of the N-acetylglucosamine attachment site of the peptide. MS^3 analysis was then performed selectively to obtain the structure of the attached oligosaccharides.

27.14

Probing Protein Structure by Laser Induced Oxidation and Mass Spectrometry

T. T. Aye, T. Y. Low, A. Vellaichamy, X. W. Goh, and S. K. Sze

Genome Institute of Singapore, Singapore

Functions of proteins are largely determined by their 3-D structures. Currently, NMR and X-ray crystallography have provided considerable insights into the mechanisms of protein functions. However, these methods are low throughput, labour intensive, and require large amount of high purity sample.

Protein surface footprinting, coupled with mass spectrometry and protein folding calculation, is emerging as a powerful alternative to study protein structure and interaction. We have developed an ultra fast laser pulse method to produce extensive protein surface oxidation. A pulse UV laser is used to produce high concentration of free hydroxyl radicals in solution in nanosecond time scale. The hydroxyl radicals react with amino acid residues on the protein surface mainly by adding an oxygen atom to the reactive sites and increases the mass of each site by 16 Da. The modified protein is then analyzed by mass spectrometry to map the modified amino acids on the protein surface.

Ubiquitin, cytochrome C, myoglobin, carbonic anhydase and BSA are used as model proteins to develop this technique. The oxidized sites on the protein surface are compared to the X-ray and NMR protein structures in the PDB. The high concentration of hydroxyl radicals generated by nanosecond laser increases the specificity and extension of surface residue modification. This contributes to the high sensitivity, specificity and speed of this method in mapping the solvent accessible surface of any protein in its native state and environment. Currently, we are extending this technique to study protein-protein interaction.

27.15

Characterization of Human 20S Proteasome Isoforms by a Combination of Proteomic Approaches: Identification of Posttranslational Modifications and Protein Variants

S. Uttenweiler-Joseph, S. Claverol, M. P. Bousquet-Dubouch, L. Sylvius, A. Stella, O. Burlet-Schiltz, and B. Monsarrat

Institut de Pharmacologie et de Biologie Structurale, CNRS, Toulouse, France

The proteasome is an enzyme present in all cells which plays a critical role in the regulation of cell growth and survival. The 20S catalytic core of the proteasome is a cylindrical complex composed of 28 subunits assembled into two outer α rings and two inner β rings containing 7 subunits each. Three β subunits per ring possess a catalytic active site and are responsible for three major peptidase activities. The enzymatic activity of 20S proteasome can be modulated by the replacement of the catalytic subunits and by post-translational modifications, such as phosphorylations.

The aim of our work is to study the 20S proteasome structure/activity relationships. Here we present our results on the human 20S proteasome subunit characterization using proteomic approaches.

The 20S proteasome of human erythrocytes exhibits numerous isoforms associated to most subunits as revealed by previous studies performed in our laboratory with a classical proteomic approach (1). To characterize these protein isoforms (protein variants and/or posttranslational modifications), we report on the usefulness of a strategy combining "bottom up" and "top down" proteomic approaches. This strategy is based on the separation of protein isoforms by 2D PAGE, passive elution of the proteins out of the gel and MS analysis of the intact proteins (2). The determination of the molecular weight of the protein isoforms at a precision of 1 Da permits to draw hypotheses on the nature of their modification(s). In parallel, the protein isoforms are digested with different endoproteases and MS/MS analyses of the resulting peptides allow the modification(s) to be confirmed and localized. Using this strategy, phosphorylations, N-terminal acetylations and protein variants have been identified on various proteasome subunits.

1. S. Claverol *et al*. (2002) *Mol. Cell. Proteomics* **1,** 567–578
2. S. Claverol *et al*. (2003) *Mol. Cell. Proteomics* **2,** 483–493

27.16

Structural Basis of a Flavivirus Recognized by Its Neutralizing Antibody: Solution Structure of the Domain III of the Japanese Encephalitis Virus Envelope Protein

Chih-Wei Wu*, Ya-Ping Tsao*, Kuen-Phon Wu*, Suh-Chin Wu, and Jya-Wei Cheng

Institute of Biotechnology and Department of Life Science, Tsing Hua University, Taiwan, China; *these three authors contributed equally to this paper.

The flavivirus envelope protein is the dominant antigen in eliciting neutralizing antibodies and plays an important role in inducing immunologic responses in the infected host. We have determined the solution structure of the major antigenic domain (domain III) of the Japanese Encephalitis Virus (JEV) envelope protein. The JEV domain III forms a β-barrel type structure composed of six antiparallel β-strands resembling the immunoglobulin constant domain. We have also identified epitopes of the JEV domain III to its neutralizing antibody by chemical shift perturbation measurements. Site-directed mutagenesis experiments are performed to confirm the NMR results. Our study provides a structural basis for understanding the mechanism of immunologic protection and for rational design of vaccines effective against flaviviruses.

27.17

Structure-based Inhibitor Design for a Cancer Target in *De Novo* Purine Biosynthesis Pathway

Lan Xu[1], Chenglong Li[1], Arthur J. Olson[1], and Ian A. Wilson[1,2]

[1]Departments of Molecular Biology and [2]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA, USA

The bifunctional dimeric enzyme, Aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase (AICAR Tfase)/IMP cyclohydrolase (IMPCH) (ATIC) is a folate-dependent enzyme and catalyzes the penultimate and final reactions in the *de novo* purine biosynthesis pathway that produces inosine monophosphate (IMP)-the precursor of purine nucleotides. Recently, ATIC has emerged a promising target for the rational design of anti-neoplastic chemotherapeutic agent.

We have focused primarily on identification of novel non-folate based inhibitors that will selectively inhibit AICAR Tfase activity, which may cause less severe cytotoxicity than traditional anti-folate drugs, such as methotrexate. A systematic search for novel inhibitors of AICAR Tfase utilizing molecular docking coupled with *in vitro* kinetic assay has enabled the discovery of several novel non-folate based inhibitors. To understand the structural basis for inhibition of AICAR Tfase, the crystal structure of ATIC complexed with one of the novel inhibitors, 326203-A (K_i = 7.1 μM), 2-[5-hydroxy-3-methyl-1-(2-methyl-4-sulfo-phenyl)-1H-pyrazol-4-ylazo]-4-sulfo-benzoic acid was determined at 1.8 Å resolution. The complex structure of ATIC/Inhibitor 326203-A exhibited a novel binding mode compared to other folate- based inhibitors. Most of the favorable binding affinity and specificity appears to be attributed to the sulfonate group which forms the dominant interaction with Lys^{267} and the proposed oxyanion hole. The inhibitor represents the first non-folate based inhibitor scaffold recognized by AICAR Tfase. Furthermore, several more potent inhibitors were identified by a similarity search based on the structural information from this enzyme-inhibitor complex. The discovery of additional compounds with improved potency supports this combination approach of computational/structural rational design of novel non-folate inhibitors of AICAR Tfase.

27.18

Nuclear Localization of Spindlin 1 Is Led by a Nonlinear Nuclear Location Signal

P. Zhang, L. P. Qing, Y. H. Gao, H. F. Yuan, D. M. Wang, L. Chen, W. Yue, and X. T. Pei

Department of Stem Cell Biology, Beijing Institute of Transfusion Medicine, Beijing, China

Spindlin 1, one of the member of spin/ssty family which expressed during gametogenesis, plays important role in cell-cycle regulation and shares sequence homology in the C-terminal with the others members of this family. It encodes a protein of 237 amino acids and locates in the nucleus. To identify the sequence that responsible for the nuclear localization of spindlin 1, different truncated forms of spindlin 1 were examined for their sub-cellular localization indicated by marker protein -EGFP(enhanced green fluorescent protein). The C-terminal deleted spindlin 1 was inserted in frame into the C terminus of EGFP, while the N-terminal deletion into the N terminus. Such recombinants were used to tranfected cultured cells (COS-7, NIH 3T3, OVC Ar3) and localization observed by confocal analysis. Results showed that residues 28–105 from N-terminal and the C-terminal 9 residues were necessary for its nuclear localization, deletion either of them make the protein a diffused distribution in the cell. But neither of them was sufficient for leading the protein into nucleus. Take together, the above results demonstrated that spindlin 1 harbor a nonlinear nuclear location signal, which may have three-dimensional structure.

27.19

Protein Structure Prediction on a Genomic Scale

Yang Zhang and Jeffrey Skolnick

Center of Excellence in Bioinformatics, State University of New York, Buffalo, NY, USA

Despite considerable effort, the prediction of the native structure of a protein from its amino acid sequence remains an outstanding unsolved problem. In this postgenomic era, because protein structure can assist in functional annotation, the need for progress is even more crucial. In this talk, we present the recent structure prediction results by *TASSER* (*T*hreading/*ASSE*mbly/*R*efinement) method based on a large-scale benchmark test. *TASSER* is a new-developed hierarchical approach to the protein structure prediction, which consists of template identification by the threading program *PROSPECTOR_3*, followed by tertiary structure assembly via the rearrangement of continuous template fragments. The benchmark set includes 2,234 proteins up to 300 residues long that cover the whole Protein Data Bank (PDB) library at the level of 35% sequence identity. With homology proteins exclusively removed, *TASSER* can fold successfully 2/3 of single-domain proteins, 2/5 of multiple-domain proteins, and 2/5 of membrane proteins. For the first time, *TASSER* modeling shows significant improvements with respect to the initial templates in a systematic way. As an illustration, *TASSER* protocol is also used to all the medium-sized ORFs in *E. Coli* genome and the comparable result is attained based on the confidence criteria established in the PDB benchmark. When using as input the best possible templates identified by structurally aligning the native structure through the PDB, *TASSER* can fold almost all the proteins with the accuracy comparable to low-resolution experimental structures. This observation highlights the urgent need for progress in fold recognition alignment algorithms, which may lead to the eventual solution of the protein structure prediction problem.