

DIVISION OF PHYSICAL CHEMISTRY

Final Program, 226th ACS National Meeting, New York, NY, September 7-11, 2003

Combinatorial Biophysical Chemistry and Molecular Evolution

Abstracts

PHYS 1 [654598]: De novo design of peptides, proteins, and peptide mimetics

William F. DeGrado, Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, 1010 Stellar Chance laboratories, 36 th and Hamilton Walk, Philadelphia, PA 19104-6059, Fax: 215-573-7229, wdegrado@mail.med.upenn.edu

Our group has developed computational and synthetic methods for understanding the structural bases for protein function. In order to test some of the rules and concepts that are believed to be important for protein folding and stability we have designed several simple proteins that fold into predetermined three-dimensional structures. A number of helical bundle motifs have been designed, and structurally characterized to determine the features that are important for folding into a native-like structure. Further, the sequences of the designed helical bundles have been elaborated to introduce binding sites for small ligands and metal ions, including a diiron cluster similar to that observed in enzymes such as the R2 subunit of ribonucleotide reductase R2 E.coli and methane monooxygenase. We have also investigated the features required for folding and assembly of membrane proteins. In a similar vein, we have designed non-peptidic oligomers that mimic the structures and biological properties of helical peptides that interact with phospholipid bilayers (e.g., antimicrobial peptides such as magainin), as well as helices that engage in protein-protein interactions.

PHYS 2 [634970]: Combinatorial associations of the human bZIP transcription factor leucine zippers

Amy E. Keating, Department of Biology, MIT, 77 Massachusetts Ave., 68-622, Cambridge, MA 02139, Fax: 617-253-4043, keating@mit.edu

The sequencing of the human genome has revealed approximately 55 human bZIP transcription factors that can form homo- or heterodimers to regulate a wide variety of biological processes. The information necessary for dimerization specificity is encoded in the coiled-coil or "leucine-zipper" domains of these proteins. We have used protein microarrays to carry out a comprehensive analysis of the intrinsic interaction specificity of the bZIPs. By paying particular attention to issues such as purity, valency and oxidation state, we have obtained very high quality interaction data, as judged by reproducibility, symmetry and agreement with solution studies. Our measurements of over 1400 unique pairwise combinations show that bZIP interactions are sparse and highly-selective in vitro. The resulting data provide an excellent foundation for a computational study of sequence and structural features that are responsible for interaction specificity.

PHYS 3 [635265]: Structures and functions of de novo proteins from designed combinatorial libraries

Michael H. Hecht, Department of Chemistry, Princeton University, Frick Laboratory, Washington Road, Princeton, NJ 08544, Fax: 609-258-6746, hecht@princeton.edu

Combinatorial libraries of de novo amino acid sequences can provide a rich source of diversity for the discovery of novel proteins. Randomly generated sequences, however, rarely fold into well-ordered protein-like structures. To enhance the quality of a library, diversity must be focused into regions of sequence space consistent with well-folded structures. We have designed focused libraries of sequences by constraining the binary pattern of polar and nonpolar amino acids to favor structures that contain abundant secondary structure, while simultaneously burying hydrophobic side chains in the protein interior and exposing hydrophilic side chains to the surrounding solvent. The experimentally determined solution structure of a novel 102-residue protein from a binary patterned library is a four-helix bundle, as specified by the design. These results show that amino acid sequences that have neither been selected by evolution, nor isolated via high-throughput screening, nor designed by computer can form native-like protein structures. The implications and applications of these de novo proteins will be discussed as they pertain to

molecular evolution, protein folding, 3-dimensional structure (alpha-helical and beta-sheet), cofactor binding, enzymatic catalysis, biosensors, amyloid formation, self-assembly, and biomaterials.

PHYS 4 [654599]: A dynamic combinatorial library approach to protein design

George L McLendon, Department of Chemistry, Princeton University, Princeton, NJ 08544, Fax: 608-258-6746, glm@princeton.edu

For a given structural archetype, what amino-acid sequences define the most stable topologies?

Exchange-labile transition metal complexes of peptides augmented with appropriate metal-ion binding residues allow the thermodynamically controlled self-assembly of stable peptide oligomers which exhibit many of the biophysical properties characteristic of natural proteins.

Preliminary experiments clearly show that the folding free energy is coupled to metal-ion binding in the assembly of peptide helix dimeric and trimeric coiled coils. This coupling has been exploited to search for optimally stable sequence variants in virtual combinatorial libraries of over 200 million members.

Interrogation of the library using ion-cyclotron FT mass spectrometry permits stepwise dissociation of peptides from the oligomeric complexes, and subsequent sequencing. This allows us to definitively answer the question posed above.

PHYS 5 [662755]: Directed evolution of proteins

Andreas Plueckthun, Biochemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057, Zürich, Switzerland, plueckthun@bioc.unizh.ch

Abstract text not available.

PHYS 29 [644513]: Protein optimization with high throughput virtual and experimental screening

John R. Desjarlais, Department of Computational Biology, Xencor, 111 W Lemon Avenue, Monrovia, CA 91016, Fax: 626-305-0350, jrd@xencor.com

Many naturally occurring proteins have potentially powerful applications, but require modulation or optimization of their properties for use. Recently developed tools for structure-based protein design are being used to improve these proteins and increase their value. We use proprietary structure-based design tools such as Protein Design Automation® technology to pre-screen vast numbers of protein sequences in silico. By screening enormous diversity computationally, low fitness sequences are eliminated from consideration and highly focused sequence libraries are created for experimental screening. These libraries facilitate the rapid discovery of improved proteins with very efficient use of experimental resources. In silico library generation finds protein sequences that are completely novel by screening the full diversity of protein sequence space. This technology has been applied to numerous systems including important industrial and pharmaceutical proteins and has a demonstrated record of success.

PHYS 30 [654589]: Grid computing methods for enhancing sampling and accuracy in computational protein design

Vijay S. Pande, Department of Chemistry, Stanford University, MS 5080, Stanford, CA 94305-5080, Fax: 650-725-0259, pande@stanford.edu

One of the greatest challenges in computational protein design is the balancing of the use of accurate (but computationally expensive) models with sufficient sampling in sequence space. By using a grid computing resource of tens of thousands of CPUs, we have performed extensive sequence space sampling using state of the art models, leading to very large-scale design libraries. Our results indicate the degree of sampling (library size) needed and suggest how one can better balance the needs of sampling and accuracy. Finally, we apply our design libraries to other problems, such as the design of novel function and to the structure prediction problem.

PHYS 31 [641850]: Ab initio profile for a given backbone structure - application to protein design and the detection functionally important residues

Hidetoshi Kono, Research Group for Biomacromolecular Function, Japan Atomic Energy Research Institute, 8-1 Umemidai, Kizu-cho, Souraku-gun, Kyoto, Japan, Fax: +81-774-71-3460, kono@apr.jaeri.go.jp, and Jeffrey G. Saven, Department of Chemistry, University of Pennsylvania

We have developed a method based on a self-consistent statistical theory to calculate amino acid probabilities of sequences (i.e., an “ab initio” profile) consistent with a particular backbone structure. The method quantifies amino acid variability at each residue site from a physico-chemical point of view. These results may be compared with multiple sequence alignments of homologous sequences produced during the course of evolution. For a variety of protein structures, comparison of ab initio and evolutionary profiles shows that they are significantly consistent with each other at about half of the residue sites. Significant discrepancy of the profiles is often found at biologically functional sites, suggesting the potential use of this comparison for predicting functionally important residues. In protein design, the profile information can be used to direct the synthesis of particular sequences as well as to identify the range of amino acid variability at selected sites. Such information may be used to engineer functional sites into proteins.

PHYS 32 [642407]: Computational approaches for combinatorial protein library prescreening and optimization

Costas D. Maranas, Gregory L. Moore, Manish C. Saraf, and Sangbum Lee, Department of Chemical Engineering, The Pennsylvania State University, 112 Fenske Laboratory, University Park, PA 16802, Fax: 814-865-7846, costas@psu.edu

Combinatorial protein library generation and screening has emerged as a powerful strategy for protein engineering. We present methods for identifying and reducing the fraction of reassembled sequences with diminished functionality due to unfavorable "clashing" interactions between protein fragments. First, a rapid prescreen for identifying residue-residue clashes is presented that considers: introduction of repulsive pairs (+/+ or -/-), disruption of H-bonds, and generation of steric clashes or cavities. Second, mean-field theory calculations are performed on a structure-based computational model to further evaluate the plasticity of residue-residue substitution patterns. With these two methods, we found that crossovers are preferentially distributed to avoid the formation of clashes in the hybrids. Lastly, an optimization framework is introduced for redesigning clashes, either upstream in the parental sequences or downstream in promising hybrids. A deterministic algorithm for the rotamer optimization problem that combines dead-end elimination with an integer programming model is presented.

PHYS 55 [644953]: Computation driven directed evolution of a beta-lactamase

Virginia W. Cornish, Shalom Goldberg, Debleena Sengupta, Ben Gherman, and Richard A. Friesner, Department of Chemistry, Columbia University, 3000 Broadway, Havemeyer Hall, MC 3111, New York, NY 10027, Fax: 212-932-1289

The targets of beta-lactam antibiotics, penicillin-binding proteins (PBPs), are believed to have evolved into beta-lactamases, the bacterial resistance enzymes that hydrolyze and therefor inactivate these antibiotics. Both PBPs and beta-lactamases are serine-protease type enzymes that form an acyl-enzyme intermediate. But while beta-lactamases efficiently catalyze hydrolysis of the acyl-enzyme intermediate, PBPs are very poor catalysts of this hydrolysis reaction and so are trapped as the acyl-enzyme intermediate and inactivated. Here we present our results using mixed quantum mechanics/molecular mechanics (QM/MM) calculations to guide the directed evolution of a penicillin-binding protein into a beta-lactamase.

PHYS 56 [642682]: Dissecting molecular recognition with phage-displayed protein libraries

Gregory A. Weiss, Ken Sato, Phillip Tam, Katsuyuki Murase, and Sara K. Avrantinis, Department of Chemistry, U.C. Irvine, 516 Rowland Hall, University of California, Irvine, Irvine, CA 92697-2025, Fax: 949-824-8571, gweiss@uci.edu

Libraries of protein variants have been used to identify sidechain residues contributing to receptor function through favorable contacts. Shotgun scanning with phage-displayed libraries of wild-type and alanine residues substituted in specific positions can quantify sidechain contributions to protein function. This 'reverse protein engineering' approach has been used to identify forces responsible for the unusually strong streptavidin-biotin interaction, the stability of M13 bacteriophage coat proteins, and the model protein-DNA interaction engrailed homeodomain binding to a specific DNA sequence. Shotgun scanning with homologous substitutions offers a method to elucidate more subtle forces required for protein specificity and optimal binding. In addition to the structure/activity information obtained from homolog shotgun scanning, the method can be used to rapidly affinity mature polypeptides.

PHYS 57 [642354]: Novel diversity for the creation of protein function

Marc Ostermeier, Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, Fax: 410-516-5510, oster@jhu.edu

The directed evolution toolbox contains a number of well-established methodologies (e.g. random mutagenesis and DNA shuffling) for improving such protein properties as stability, catalytic activity and substrate specificity. What are lacking are methods for the creation of new function where no function exists. Traditional directed evolution methods generate diversity by mutating genes and recombining homologous genes but maintaining gene length and order of gene sequences. We are developing two strategies for the introduction of allosteric properties into proteins that utilize diversity created in novel ways: random domain insertion (which utilizes diversity in the order of sequences) and conditional heterodimerization (which utilizes diversity in the gene length).

PHYS 58 [644849]: Computational and experimental analysis of DNA shuffling

Narendra Maheshri, and David V. Schaffer, Department of Chemical Engineering and the Wills Neuroscience Institute, University of California at Berkeley, 201 Gilman Hall, Berkeley, CA 94720-1462, naren@uclink.berkeley.edu

We describe a computational model of DNA shuffling, SHUFFIT, based on the thermodynamics and kinetics of this process. The model independently tracks a representative ensemble of DNA molecules and records their states at every stage of a shuffling reaction. These data can subsequently be analyzed to yield information on any relevant metric, including reassembly efficiency, crossover number, type, and distribution, and DNA sequence length distributions. The predictive ability of the model was validated by comparison to three independent sets of experimental data, and analysis of the simulation results led to several unique insights into the DNA shuffling process. We examine a tradeoff between crossover frequency and reassembly efficiency and illustrate the effects of experimental parameters on this relationship. Furthermore, we discuss conditions that promote the formation of useless “junk” DNA sequences or multimeric sequences containing multiple copies of the reassembled product. We are making SHUFFIT accessible to researchers (<http://www.cchem.berkeley.edu/~schaffer/shuffling>) in order to aid in the design of optimal shuffling reaction conditions.

PHYS 82 [658814]: Evolutionary perspectives on protein folding, structure, and thermodynamics

Richard A. Goldstein, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, Fax: 734-647-4865, richardg@umich.edu

Proteins fold into their native-state conformations in milliseconds to seconds, ignoring theoretical estimates that this process should take many times the age of the universe. Much work is directed to understanding how proteins are so much smarter than theorists, who cannot even reliably predict what the final folded states will be. Proteins have one major advantage over theorists - proteins have been working on this problem for billions of years. We can consider different ways in which proteins may have evolved to solve the protein-folding problem. Using simple theoretical models, we can show how neutral evolution and population dynamics combined with the need to fold can explain many of the observed properties of proteins, including the way proteins fold, the distribution of observed protein structures, the marginal stability of proteins, and how the evolutionary robustness of protein structures co-exists with sequence plasticity.

PHYS 83 [644418]: Protein building blocks for recombination

Zhen-Gang Wang, Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd., Mail Code 210-41, Pasadena, CA 91125, Fax: 626-568-8743, zgw@cheme.caltech.edu

In vitro recombination is a powerful tool for tuning and optimizing protein functions. It promotes the rapid accumulation of beneficial traits from multiple parents onto a new single offspring. Recombination also plays a key role in natural evolution in the generation of diverse antibodies, synthases, and proteases through the swapping of well-defined structural domains. However, while there have been several approaches to identifying the building blocks of proteins, the principles that govern whether a polypeptide fragment can be exchanged among different proteins remain unclear. We have recently developed a computational algorithm called SCHEMA for identifying the structural units that can be swapped among homologous proteins without disrupting the three-dimensional structure

of the protein. This talk will discuss the SCHEMA method and its application in guiding recombination experiments.

PHYS 84 [644428]: Exploring sequence space by structure-guided recombination

Jonathan J. Silberg, Michelle M. Meyer, Christopher R. Otey, and Frances H. Arnold, Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail Code 210-41, Pasadena, CA 91125, Fax: 626 568 8743, joff@cheme.caltech.edu

Anecdotal evidence from DNA-shuffling experiments suggests that recombination can rapidly accumulate beneficial mutations onto a single offspring and possibly find novel and beneficial combinations of amino acids that are neutral in their original, parental contexts. Little information exists, however, on the relationship between mutation level and evolution of protein function, i.e. how evolution of function scales with the number of effective mutations incorporated by recombination. We are using structure-based tools to identify polypeptide elements that can be swapped among related proteins without disrupting their three-dimensional structure. By recombining these elements in a combinatorial fashion we are creating well-defined libraries of diverse sequences (using distantly-related β -lactamase or cytochrome P450 homologs as parents), a large fraction of which are predicted to retain parental fold. We are analyzing the structures and activities of proteins in these chimeric libraries, and investigating whether these proteins can display interesting behaviors, including ones not known in nature.

PHYS 85 [644365]: Focused saturation mutagenesis for antibody affinity maturation

K. Dane Wittrup¹, Katarina Midelfort², Shaun Lippow¹, Brenda A. Kellogg³, and Bruce Tidor⁴. (1) Chemical Engineering & Bioengineering, Massachusetts Institute of Technology, MIT 66-552, Cambridge, MA 02139, Fax: 617-258-5766, wittrup@mit.edu, (2) Biological Engineering, MIT, (3) Department of Chemistry, Massachusetts Institute of Technology, (4) Biological Engineering Division and Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology

There is substantial and expanding activity in the development of antibody drugs, and therapeutic efficacy is often directly related to antigen binding affinity. We have demonstrated previously that directed evolution by yeast surface display is a powerful method for identifying mutant antibodies with extremely high affinity, however such mutants often possess functionally consequential mutations in framework residues that raise the specter of human immunogenicity along the clinical development pathway. It is of interest to identify which sites in the complementarity determining regions (CDRs) can be mutated to obtain improved affinity. We have taken both computational and experimental approaches to finding those sites most amenable to mutagenic improvement, using as model systems the 4-4-20 antibody against fluorescein and the D1.3 antibody against lysozyme.

PHYS 86 [642659]: Evolution and mis-evolution of antibodies in the immune system

Michael W. Deem, and Jeong-Man Park, Departments of Bioengineering and Physics & Astronomy, Rice University, 6100 Main Street - MS 142, Houston, TX 77005-1892, Fax: 713-348-5811, mwdeem@rice.edu

We introduce a model of protein evolution to explain limitations in the immune system response to vaccination and disease. The phenomenon of original antigenic sin, wherein vaccination creates memory sequences that can *increase* susceptibility to future exposures to the same disease, is explained as stemming from localization of the immune system response in antibody sequence space. This localization is a result of the roughness in sequence space of the evolved antibody affinity constant for antigen and is observed for diseases with high year-to-year mutation rates, such as influenza.