tools for STRUCTURAL GENOMICS

accelerating the structure pipeline

The international structural genomics effort has resulted in a number of technological advancements that are accelerating the process of three-dimensional structure determination while continually decreasing the cost per structure. Significant strides have been made for all of the required experimental steps including protein expression, purification, crystallisation and structure solution. However, the overall success rates for producing structures are still quite low due to two main bottlenecks, protein production and crystallisation. The magnitude of the low success rate is accentuated by the fact that the first five years of the structural genomics efforts have focused on the easy proteins, (ie predominantly non-membranous proteins from prokaryotic organisms and small non-membranous proteins from eukaryotic organisms). However, a number of new technologies and experimental approaches provide a realistic optimism regarding the realisation of the ultimate goal to determine 10,000 new structures with at least one example of every possible protein family during the next five years. This accomplishment should provide the necessary protein structure scaffolds to allow scientists to predict the structures of all other proteins using only gene sequences, a result that will substantially benefit fundamental biology and medicine.

By Dr Lawrence J. DeLucas, Dr Christie G. Brouillette, Dr Kal Ramnarayan and Dr Shankari Mylvaganam

Table 1: High throughput structural genomics

Summary of crystallisation results reported February 1, 2004 for the NIH Structural Genomics Initiative. Statistics are based upon the total number of soluble proteins expressed by the original seven NIH Structural Genomics Consortia

	SOLUBLE PROTEINS EXPRESSED	CRYSTALLISED	DIFFRACTION QUALITY	STRUCTURES	DEPOSITED IN PDB
Total Number	7,652	1.793	766	555	409
Percentage	36.2%	23.4%	10.0%	7.3%	5.3%

The human genome or any other genome, for that matter, will not be truly understood until the functional roles of all the possible gene products are known. In pursuit of this quest, there has been an exponential demand for protein structural information, which is the focus of various structural genomics initiatives around the world (http://www.nigms.nih.gov/funding/psi/lay_summary.html). The hope is to determine the structure for one or more proteins from each family, reaching a total of 10,000 structures. These structures will provide valuable information regarding the relationship between a protein's sequence and its tertiary structure. It is believed that this knowledge will provide a foundation of data that will enable the prediction of all other structures from sequence information alone.

A 3D protein structure is critical to the advancement and efficiency of rationale drug design, as well as to protein structure-function studies, because the majority of drugs and natural effector molecules stereo-specifically interact with target proteins to affect the physiological and biological activity of a protein by blocking or altering its properties. According to some estimates, the number of disease targets could potentially increase from 500 or so today to more than 10,000. Within the next few years, however, the Human Genome Project and other academic and commercial initiatives are expected to identify the genes for more than 20,000 potential drug discovery targets from among the 30,000 or so genes believed to comprise the human genome[1,2]. The genes for hundreds of additional targets are also being identified from the genomes of pathogenic bacterial, viral and fungal organisms. Genetic sequences of thousands of protein targets from other pharmacologically relevant species, such as rat, mouse, dog and certain primates are being or will be determined for use as appropriate animal models. Finally, genetic sequences of selected targets from genomic research species such as *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), *Brachydanio rerio* (zebra fish) and *Saccharomyces cerevisiae* (yeast) will also be important in gene function identification studies and comparative genomics.

Last year, ***DDW*** published a review article describing high-throughput structure-based drug discovery technology[3], which provided a good foundation and historical perspective for the present article. Here we will focus on technologies not represented in last year's article as well as attempt to bring up to date topics covered in both.

Status of high throughput structural genomics

X-ray crystallography remains the predominant method contributing to the majority of new structures emerging from the structural genomics initiative. Unfortunately, in spite of the fact that most of the target organisms are prokaryotic, success rates for producing structures are extremely low. **Table 1** provides a summary of the status for the seven original US NIH-funded Research Centers, in operation for three years.

In spite of the overall dismal success rate, high throughput structural genomics programmes have resulted in a number of technological advancements for each of the critical steps necessary to determine three-dimensional structures. For instance, Nuclear Magnetic Resonance (NMR) techniques have already proven useful and contributions are expected to grow significantly in the future[60-63].

Protein production

The first step, protein production, has involved the use of multiple protein expression systems, adapted to high throughput by scaling the process

down to analytical volumes (<2ml) that can be accommodated in multi-well format. Each expression system has advantages and disadvantages regarding cost, yield, post-translational modifications, susceptibility to automation, process time, etc. A number of the structural genomics consortia (http://www.nigms.nih.gov/news/meetings/airlie.html#agree) have relied heavily on *E. coli* expression systems due to their simplicity, cost and amenability to parallel processing. Robotic systems originally designed for high throughput compound screening can be adapted to perform most or all aspects of *E. coli* protein expression[4]. Certainly, *E. coli* can be considered the expression system of choice for the 'low hanging fruit' (small proteins without post-translational modifications). However, as with any of the alternative systems, *E. coli* expression has limited the ability to express properly folded, large, eukaryotic proteins, post translationally modified proteins and biologically active membrane proteins. The combination of using different bacterial expression systems or cell-free[7,8], mammalian, yeast[6,9,10] and various viral expression systems expressed in insect cells[11] has also offered promise as alternative approaches for the high throughput structural genomics community[5,6]. Problems are being addressed using alternative approaches. Efforts have been established through consortiums such as MepNet (Membrane Protein Network) with the goal of concentrating research efforts on the expression and crystallisation of 101 G-coupled protein receptors (GCPRs) using three alternative expression vectors, *E. coli*, *P. pastoris*, and the Semliki Forest virus (SFV). To date, 60% of the targets in this consortium have been expressed at 1mg levels or higher and demonstrated as biologically relevant (personal communication).

It is clear that strategies must be developed and refined to adequately accommodate challenging proteins (ie membrane proteins and large protein complexes) that typically represent more than one third of the total number of proteins in prokaryotic and eukaryotic genomes.

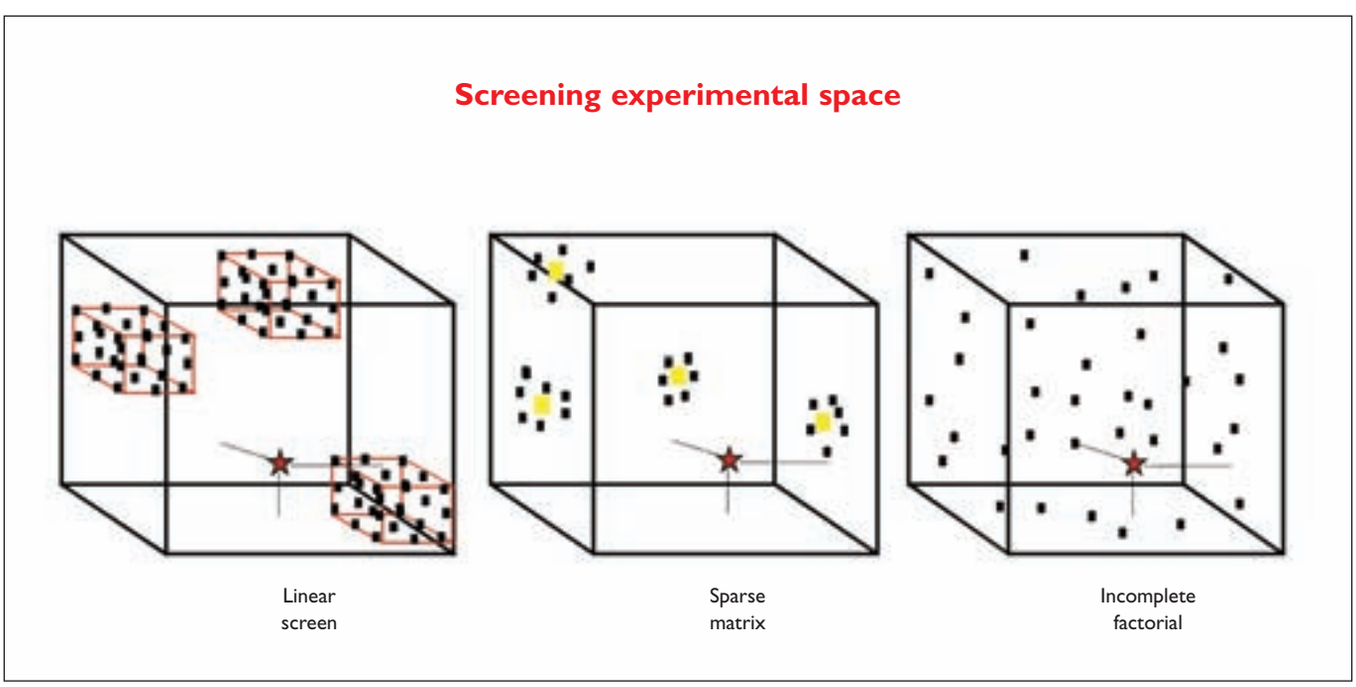
Protein purification

The process of purification may be more of a bottleneck than originally realised. Although automated high throughput purification systems and new chromatography media have all enhanced scientist's ability to produce 'purified protein'. Producing sufficient quantities with sufficient purity (>95%) and homogeneity to yield diffraction-quality crystals is still a problem. This is most likely one of the factors contributing to the low crystallisation success rate listed in **Table 1**. Application of high-throughput purification of more complicated aqueous and membrane proteins is expected to become more challenging.

Crystallisation

Crystallisation remains the most challenging and difficult problem as evidenced in **Table 1**. The low crystallisation success rate exists in spite of a num-

Figure 1
Comparison of crystallisation screens. Commercially available crystallisation screens generally use a sparse matrix and/or linear approach for sampling various combinations of the screening parameters. As seen in the figure, these methods provide a fine screening of particular regions of 'crystallisation space' but they fail to sample other regions. An incomplete factorial provides the minimum number of experiments required to evenly sample all possible permutations of the variable (ie the entire 'crystallisation space')



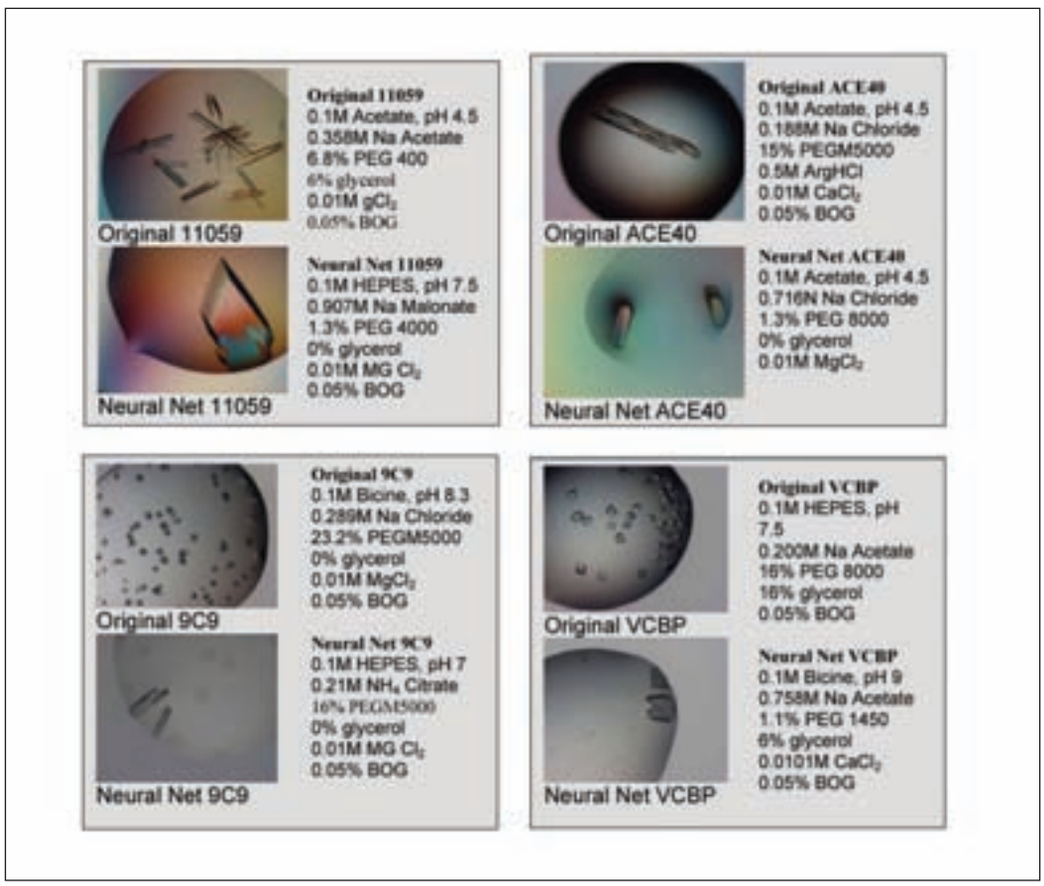


Figure 2
Neural net vs crystal optimisation. Images and chemical conditions for the best results obtained from an incomplete factorial screen from four (non-commercial) proteins are compared with experimental results obtained from new conditions predicted by a neural net analysis of results from the incomplete factorial screen

ber of advancements in the field including the availability of fully automated robotic vapour and liquid diffusion crystallisation systems, use of site-directed mutagenesis to engineer 'crystallisation constructs', development of new techniques such as microbatch (under-oil) crystallisation[12], and sophisticated systems that dynamically control the crystallisation kinetics[97].

Further reducing the scale of crystallisation experiments from micro to nanolitre volumes is a solution to the problem of producing sufficient quantities for crystallisation studies. Our research centre (Center for Biophysical Sciences and Engineering, CBSE) developed an automated in-house system that can prepare vapour diffusion nano-crystallisation experiments ranging from 15nl to 200nl drop volumes[13]. A variety of commercially available systems provide similar capabilities with experiment throughput ranging from hundreds to several thousand experiments/hr[14-19].

An alternative crystallisation approach to vapour diffusion is liquid diffusion. The Fluidigm Corporation developed an automated microfluidics system that essentially performs liquid diffusion experiments in nanolitre volumes. Liquid diffusion provides an additional capability that may prove useful, particularly for the crystallisation of membrane proteins because the detergent micelle concentration does not change appreciably during the crystallisation equilibration process. High-throughput systems have led to the development of several automated crystal observation and analysis systems (commercially available) that have significantly reduced the labour and time required to inspect individual experiments for crystal growth[3]. Most of the commercial systems automatically score the contents of each experiment into broad categories such as clear drop versus precipitate versus crystal. However, automatic image discrimination capable of assessing the quality of different crystals remains an improvement that the structural genomics community desires.

The crystallisation of a homologous protein from another organism[20] approaches crystallisation from a different angle. In many cases, the differences in sequence are minor and often found on the surface or in regions without direct biological activity (ie, not in the active site of an enzyme); therefore, co-ordinate information from the x-ray structure of the homologous protein can be used to accu-

rately model the structure of the original protein. Limited proteolysis can provide a protein form that is, by chance, more conducive to crystallisation[21]. Introduction of point mutations, truncations or deletions has also been demonstrated to help improve crystallisation success rates[22-30]. Alternatively, one can modify the target protein's surface by introducing co-factors, additives, antibodies, or through the removal of carbohydrates in an effort to produce more suitable crystalline lattice contacts[31-38]. This includes the use of site-directed changes to surface amino acids to create 'crystallisation constructs'[22,23,39]. A recent study[39] provides a strategy and theoretical rationalisation for making specific surface mutations, more likely to improve a protein's ability to crystallise. The strategy involves the replacement of large, and therefore flexible, side-chains (ie lysine, arginine, glutamine, etc) that might exist alone or in patches on the exposed surface of a protein, with alanine, a small, uncharged amino acid. This, it is hypothesised, allows the protein to interact more closely with itself and it reduces the unfavourable entropy hurdle that must be overcome to constrain, for example, a flexible lysine, as the protein tries to form a crystalline lattice. A new method, high-throughput deuterium exchange mass spectrometry (DXMS) can be used to rapidly identify unstructured regions on a protein's surface. Truncation of the disordered regions has been demonstrated to improve crystallisation[40]. Additional knowledge regarding a protein's surface characteristics (such as can be obtained via entropy considerations or empirical DXMS data), provide a rational and more time/cost-effective approach for engineering successful crystallisation constructs.

To improve our success rate with crystallisation trials, CBSE has optimised an incomplete factorial screen thereby allowing a small number of experiments to be performed, sampling all possible experiments in a statistically robust manner[41]. This approach provides for efficient determination of solution conditions suitable for crystallising proteins by performing experiments that take into account the independent and interdependent influences of each experimental parameter. Comparison

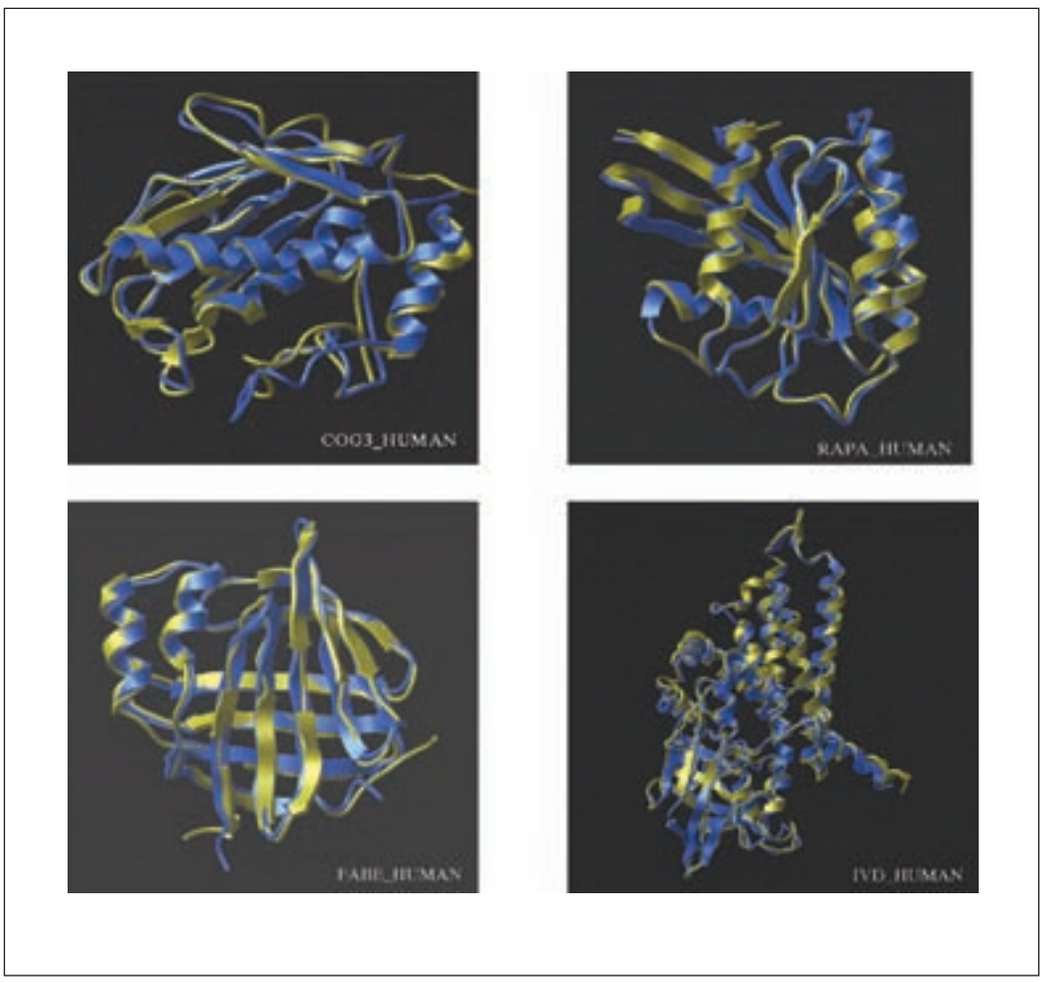


Figure 3
Ribbon representations of the superposition of augmented homology models (coloured yellow) with the corresponding PDB crystal structures (blue) for four proteins in the in-house protein set. The images from top left, clockwise correspond to cog3_human, rapa_human, ivd_human and fabe_human

between the incomplete versus sparse matrix screens suggests that the incomplete factorial method may find a larger number of conditions that are useful for crystal optimisation[13,42].

An extension of the incomplete factorial crystallisation screen involves its combination with automated predictive algorithm to evaluate all possible permutations of variables and their levels (ie, specific protein/crystallising component concentrations, pH, solution ionic strength, temperature, etc). If the correct variables and sample size are chosen to adequately represent the crystallisation nature of the protein, training the neural net with the incomplete factorial screen results in a stable set of hidden neurons and basis function weights. The 'trained' neural network can then be used to predict non-sampled complete factorial conditions that theoretically cover the entire 'crystallisation space' of possible experimental conditions. Our preliminary results indicate that this approach could increase the success rate for producing diffraction-quality macromolecular crystals has recently been published by our group[13,42]. The images in **Figure 2** demonstrate some of the dramatic improvements in crystal size and quality observed using this technique. Of particular significance is the fact that in a number of cases, the optimisation conditions predicted by the neural net are quite disparate from any of the screen conditions (including those of the initial crystals) used to 'train' the neural net[13,42].

Structure determination

Determination of the crystallographic structure of a protein (once sufficient crystals are obtained) has realised the most dramatic technological advancements, compared to other steps in the overall process. Synchrotron radiation facilities provide extremely brilliant sources, combined with automated crystal handling and preparation/alignment, have shortened the time needed to collect data from hours to minutes[16,17,43-53]. The ability to produce and crystallise selenomethionine-substituted proteins[54-56] eliminates the need to collect

additional data from crystals soaked with heavy metal complexes. This plus a number of alternative approaches for rapidly obtaining initial protein phases, the critical/enabling step for structure determination, have had a major impact in making high throughput structure analysis a reality. Throughput is enhanced even more by the availability of software packages that automate the search for initial phases through iterative improvements in electron density representations of the structure. Efforts to determine initial protein phases using the native sulfurs present in proteins also shows promise[57]. Finally, the tedious task of model building (fitting atomic models of proteins into electron densities) has, to a large extent, been automated[51,58,59].

It can be be logically assumed that large molecular complexes and membrane proteins will often yield crystal of marginal quality, a natural impediment to obtaining the x-ray phases from which the initial protein model is calculated. For weakly diffracting crystals or crystals of large molecular complexes low-dose electron tomography (ET) may prove useful in leapfrogging the x-ray phase acquisition process[64-66]. For example, a 20-angstrom structure obtained by ET can be used to determine the initial crystallographic protein phases by performing a molecular rotation search of the ET model within the crystallographic data[67]. One commercial service, provided by Sidec Technologies AB, exploits low dose cyroelectron tomography in combination with a proprietary algorithm (COMET, Constrained Maximum Entropy Tomography) that enhances signal to noise ratios using a small amount of non-crystallised sample. This technique enables three-dimensional molecular reconstruction within days to yield a 20-angstrom structure. ET can observe protein-protein interactions within a solution or cell membrane. This type of information can complement the corresponding functional studies that will be necessary to understand the involvement of these proteins in biological processes.

As noted previously, the ultimate goal of the international structural genomics effort is to determine from every possible protein family. This information is expected to improve the ability to accurately model or predict the three-dimensional structure of a new protein purely from its primary amino acid sequence. The goal appears to be a realistic one as evidenced by the significant improvements in structure prediction methods seen over the last five years[68-74]. Presently there are three ways to generate protein models from sequence information: comparative or homology modelling[75-77], fold recognition or threading[78-80] and *ab initio* methods[81-84]. In spite of known weaknesses in these methods, homology modelling is being widely used to derive 3D models of proteins with high sequence homology to known structures with an accuracy required for use in drug discovery[85-87]. The threading methods are being used for putative target function assignments[88-90] and the *ab initio* methods are being used to derive structures of small to medium-sized peptides[91-93] and in loop generation in proteins and antibodies[94,95].

Cengent Therapeutics developed a novel comparative modelling approach called Augmented Homology Modeling™ that relies on an iterative method of deriving the protein structure to extend the range of accessible structures available and useful for rational experimental design. The Augmented Homology Modeling™ method has, in a number of cases led to protein models of improved quality, more closely matching the actual protein structure. This method also allows good quality models to be generated for proteins that have lower homology than their template structures, ie belong to the so-called 'twilight region' of sequence identity. The ribbon superpositions of predicted models (yellow) with corresponding PDB structures (blue) for four Cengent modelled proteins are shown in **Figure 3**. The structural folds for the cores of the experimental and predicted coordinates were highly conserved. Deviations occur predominantly at the N- and C-termini and, to a lesser extent, for some of the loops.

Conclusions

As the structural genomics efforts continue to mature and structure determination for the easy proteins or 'low hanging fruit' is completed, attention will naturally turn to the more difficult problems. The future challenges include expression and purification of proteins from eukaryotic organisms containing post-translational modifications, membrane proteins and large multi-domain protein complexes. Novel high throughput, cost-effective expression protocols will need to be developed to address the protein production phase of the process. Equally challenging will be the development of crystallisation strategies for these complicated proteins, particularly membrane proteins. Although new experimental approaches for membrane protein crystallisation have emerged in recent years, adaptation to high-throughput, fully automated systems will require further work. A semi-automated system capable of accommodating the 'in cubo' crystallisation method for membrane

proteins was recently described[96]. Developments such as this are expected to lead to new high throughput protocols whereby a variety of lipids and detergent/lipid mixtures are rapidly screened for suitable crystallisation conditions.

If the international genomics technological achievements of the past five years are any indication of the future, there can only be optimism with regard to the community's goal of determining 10,000 new structures by 2010. Although more difficult protein targets await scientists, there is a substantial international commitment to this programme.

Acknowledgements

With thanks to Sharney Logan, Director of Business Development, Center for Biophysical Sciences and Engineering; Southeast Collaboratory for Structural Genomics, SECSG; University of Georgia (UGA), University of Alabama at Birmingham (UAB), University of Alabama at Huntsville (UAH), and Duke University; Charlie Carter, University of North Carolina Chapel Hill (UNC); and Funding by NIH grant P50-GM62407 and NASA Cooperative Agreement NCC8-246. DDW

Dr Larry DeLucas *is the Director of the Center for Biophysical Sciences and Engineering, Director of the Comprehensive Cancer Center X-ray Core Facility and Professor at the University of Alabama at Birmingham. His research includes structure-based drug design with x-ray crystallography in the fields of genomics and proteomics coupled with development of innovative technologies for drug discovery platforms. Dr DeLucas has published more than 100 research articles in various scientific journals, co-authored two books on protein crystal growth and is a co-inventor on 14 patents mainly involving protein crystal growth.*

Christie Brouillette, PhD *is Research Professor and Associate Director of Molecular Biophysics at the University of Alabama at Birmingham Center for Biophysical Sciences and Engineering. Her Biomolecular Analysis Group includes the Center's High Throughput Screening Facility and Biocalorimetry Laboratory. Her research includes studies of protein structural co-operativity and energetics using a variety of biophysical tools, including calorimetry and protein-protein, and protein-ligand interactions, especially relating to lead discovery/optimisation. She is co-inventor on three patents and has co-authored numerous review articles on the structure-function of apolipoproteins.*

Kal Ramnarayan, PhD*, is Vice-President and Chief Scientific Officer of Cengent Therapeutics Inc and co-founded Structural Bioinformatics, Inc (now Cengent Therapeutics Inc). He has co-invented US Patents 6,436,933, 5,571,821 and 6,541,498 and is on the Advisory Board of IBM's BlueGene programme, Strand Genomics and Keck Graduate Institute. Dr Ramnarayan has a PhD in Molecular Biophysics from IISc India.*

Shankari E. Mylvaganam, PhD *is Head of Crystallography and Associate Director at Cengent Therapeutics Inc. Dr Mylvaganam has established the protein production and Crystallography Department and manages contract crystallography. She has determined the structures of several phosphatase with inhibitors, kinases, E8 antibody/antibody-cytochrome-c and a novel haemoglobin. Her PhD is in Protein Crystallography from Birkbeck College, England.*

References

1 Venter, JC, et al (2001). The sequence of the human genome. Science 291,1304-1351.
2 O'Donovan, C, et al (2001). The human proteome initiative (HPI). Trends Biotechnol. 19, 178-181.
3 Stevens, RC (2003). The Cost and Value of Three-Dimensional Protein Structure. Drug Discovery World 4, 35-48.
4 Finley, JB, et al (2004). Expression, purification, and characterization of 3-deoxy—arabino-heptulosonate 7-phosphate synthase from Pyrococcus furiosus. Protein Expression and Purification 34, 49-55.
5 Lueking, A, et al (2000). A System for Dual Protein Expression in Pichia pastoris and Escherichia coli. Protein Expression and Purification 3, 372-378.
6 Boettner, M, et al (2002). High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoria. Journal of Biotechnol. 99, 51-62.
7 Endo, Y and Sawasaki, T (2003). High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. Biotechnology Advances 21, 695-713.
8 Sawasaki, T, et al (2002). A cell-free protein synthesis system for high-throughput proteomics. PNAS 12, 14652-14657.
9 Gilbert, M, and Albala, JS (2002). Accelerating code to function: sizing up the protein production line. Current Opinion in Chem. Biol. 6, 102-105.
10 Holz, C, et al (2002). A micro-scale process for high-throughput expression of cDNAs in the yeast Saccharomyces cerevisiae. Protein Expression and Purification 25, 372-378.
11 Coleman, TA, et al (1997). Production and purification of novel secreted human proteins. Gene 190, 163-171.
12 Cheyen, NE (1997). The role of oil in macromolecular crystallization. Structure 5, 1269-1274.
13 DeLucas, LJ, et al (2003). Efficient Protein Crystallization. J. Structural Biology 142, 188-206.
14 Chayen, NE, Stewart, PS and Baldock, P (1994). New developments of the IMPAX small volume automated cyrstallisation system. Acta Crystallogr D 50, 456-458.
15 deTitta, G (2000). Gearing up for ~40K crystallization experiments a day: meeting the needs of HT structural proteomics projects. 8th International Conference on the Crystallisation of Biological Macromolecules (ICCBM8).
16 Goodwill, KE, Tennant, MG and Stevens, RC (2001). High-throughput x-ray crystallography for structure-base drug design. Drug Discov. Today 6, S113-S188.
17 Kuhn, P, et al (2002). The genesis of high-throughput structure-based drug discovery using protein crystallography. Curr. Opin. Chem. Biol. 6, 704-710.
18 Stevens, RC (2000). High-throughput crystallization. Curr. Opin. Struct. Biol. 10, 558-563.
19 Rose, D (1999). Microdispensing technologies in drug discovery. Drug Discov. Today 5, 511-419.
20 Campbell, JW, et al (1972). X-ray diffraction studies on enzymes in the glocolytic pathway. Cold Spring Harbor Symp. Quant. Biol. 35, 165-170.
21 McPherson, A (1982). Preparation and Analysis of Protein Crystals. Wiley, New York.
22 Lawson, DM, et al (1991). Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. Nature 349, 541-544.
23 McElroy, HE, et al (1992). Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. J. Cryst. Growth 122, 265-272.
24 D'Arcy, A, et al (1999). Crystal engineering: a case study using the 24kDa fragment of the DNA gyrase B subunit from Escherichia coli. Acta Crystallogr. D 55, 1623-1625.
25 Ay, J, et al (1998). Structure and function of the Bacillus hybrid enzyme GluXyn-1: native-like jellyroll fold preserved after insertion of autonomous globular domain. Proc. Natl. Acad. Sci. USA 95, 6613-6618.
26 Betton, JM, et al (1996). Creating a bifunctional protein by insertion of beta-lactamase into the maltodextrin-binding protein. Nat. Biotechnol. 15, 1276-1279.
27 Nagi, AD and Regan, L (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. Fold Des. 2, 67-75.
28 Nugent, PG, et al (1996). Protein engineering loops in aspartic proteinases: site-directed mutagenesis, biochemical characterization and X-ray analysis of chymosin with a replaced loop for rhizopuspepsin. Protein Eng. 9, 884-893.
29 Thompson, MJ and Eisenberg, D (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. J. Mol. Biol. 290, 595-604.
30 Zhou, HX, Hoess, RH and DeGrado, WF (1996). In vitro evolution of thermodynamically stable turns. Nat. Struct. Bio. 3, 446-451.
31 Davis, SJ, et al (1990). Crystallization of a soluble form of the rat T-cell surface glycoprotein CD4 complexed with Fab from the W3/25 monoclonal antibody. J. Mol. Biol. 213, 7-10.
32 Ostermeier, C, et al (1997). Structure at 2.7Å resolution of the Paracoccus denitrificans two-subunit sytochrome c oxidase complexed with an antibody FV fragment. Proc. Natl. Acad. Sci. USA 94, 10547-10553.
33 Prongay, AJ, et al (1990). Preparation and crystallization of a human immunodeficiency virus p24-Fab complex. Protein Eng. 7, 933-939.
34 DeLucas, LJ, et al (1977). Preliminary X-Ray Study of Crystals of Human Transferrin. J. Mol. Biol. 123, 285-286.
35 Baker, HM, et al (1994). Enzymatic proteins. Acta Crystallogr. D 50, 380-384.
36 Grueninger-Leitch, F, et al (1996). Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. Protein Sci. 12, 2617-2622.
37 Kostrewa, D, et al 1997. Crystal structure of phytase from Aspergillus ficuum at 2.5Å resolution. Nat. Struct. Biol. 4, 185-190.
38 Oefner, C, et al (2000). Structure of human neutral endopeptidase (Neprilysis) complexed with phosphoramidon. J. Mol. Biol. 296, 341-349.
39 Derewenda, Z (2004). Rational Protein Crystallization by Mutational Surface Engineering. Structure 12, 529-535.
40 Pantazatos, D, et al (2004). Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. PNAS 101, 751-756.
41 Carter, Jr, CW and Carter, CW (1979). Protein crystallization using incomplete factorial experiments. J. Biol. Chem. 254, 12219-12223.
42 DeLucas, LJ, et al (accepted, in press). Protein Crystallization: Virtual Screening and Optimization, Progress in Biophys. and Mol. Biol.
43 Blundell, TL, Jhoti, H and Abell, C (2002). High-throughput crystallography for lead discovery in drug design. Nat. Rev. Drug Disc. 1, 45-54.
44 Bodenstaff, ER, et al (2002). The prospects of protein nanocrystallography. Acta Crystallogr. D58, 1901-1906.
45 Wilson, J (2002). Towards the automated evaluation of crystallization trials. Acta Crystallogr. D 58, 1907-1914.
46 Watanabe, N (2002). Semi-automatic protein crystallization system that allows in situ observation of x-ray diffraction from crystals in the drop. Acta Crystallogr. D 58, 1527-1530.
47 Muchmore, SW, et al (2000). Automated crystal mounting and data collection in protein crystallography. Structure 58, 243-246.
48 Sharff, AJ (2003). High-Throughput Crystallography on an in-house source, using ACTOR. Rigaku Journal.
49 Hope, H (1988). Cryocrystallography of biological macromolecules. Acta Crystallogr. B 44, 22-26.
50 Garman, E (1999). Cool data: quantity and quality. Acta Crystallogr. D 55, 1641-1653.
51 Terwilliger, TC and Berendzen, J (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D 55, 849-861.
52 Perrakis, A, Morris, R and Lamzin, VS (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. 6, 458-463.
53 Lamzin, VS and Perrakis, A (2000). Current State of automated crystallographic data analysis. Nat. Struct. Biol. 7, 978-981.

Continued on page 74

Continued from page 73

54 Hendrickson, WA, Horton, JR and LeMaster, DM (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. 9, 1665-1672.
55 Hendrickson, WA and Ogata, CM (1997). Phase determination from multiwavelength anomalous diffraction measurements. Methods Enzymol. 276, 494-523.
56 Rice, LM, Earnest, TN and Brunger, AT (2000). Single-wavelength anomalous diffraction phasing revisited. Acta Crystallogr. D 56, 1413-1420.
57 Adams, MWW, et al (2003). The Southeast Collaboratory for Structural Genomics: A High-Throughput Gene to Structure Factory. Accounts of Chemical Research 36 , 191-198.
58 Terwilliger, TC (2000). Maximum-likelihood density modification. Acta Crystallogr. D 56, 965-972.
59 Blundell, TL, et al (2002). High-throughput x-ray crystallography for drug discovery. Drug Design Special Publication 279, 53-59.
60 Leonor MP, et al (2003). High-throughput screening of structural proteomics targets using NMR. FEBS Letters 552, 207-213.
61 Valafar, H and Prestegard, JH (2003). Rapid classification of a protein fold family using a statistical analysis of dipolar couplings. Bioinformatics 19, 1549-1555.
62 Yakunin, AF, et al (2004). Structural proteomics: a tool for genome annotation. Current Opinion in Chemical Biology 8, 42-48.
63 Heinemann, U, et al (2000). An integrated approach to structural genomics. Progress in Biophysics and Molecular Biology 73, 347-362.
64 Engström, M (2002). Drug Discovery High-Resolution 3-D Protein Conformation. Genetic Engineering News 22.
65 Steven, AC and Aebi, U (2003). The next ice age: cryo-electron tomography of intact cells. Trends Cell Biol. 13, 107-110.
66 Sandin, S, et al (2004). Structure and Flexibility of Individual Immunoglobulin G Molecules in Solution. Structure 12, 409-415.
67 Moore, PB and Steitz, TA (2003). The structural basis of large ribosomal subunit function. Annu Rev Biochem. 72, 813-50.
68 Bujnicki, JM, et al (2001). Structure prediction meta server. Bioinformatics 17, 750-751.
69 Xiang, X and Honig, B (2001). Extending the accuracy limits of prediction for side chain conformations. J. Mol. Biol. 311, 421-430.
70 Kelley, LA, MacCallum, RM and Sternberg, MJE (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 299, 501-522.
71 Shi, J, Blundell, TL and Mizuguchi, K (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure dependent gap penalties. J. Mol. Biol. 310, 243-257.
72 McGuffin, LJ, Bryson, K and Jones, DT (2000). The PSIPRED protein structure prediction server. Bioinformatics 16, 404-405.
73 Venclovas, C (2001). Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. Proteins: Struct., Funct., and Genet. 45, 47-54.
74 Kihara, D, Lu, H, Kolinski, A and Skolnick, J (2001). TOUCHSTONE: An ab initio protein structure prediction method that uses threading-based tertiary restraints. Proc. Natl. Acad. Sci. USA 98, 10125-10130.
75 Sali, A and Blundell, T (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815.
76 Abagyan, R, et al (1997). Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. Proteins: Struct., Funct., and Genet. Suppl. 1, 29-37.
77 Lambert, C, et al (2002). ESyPred3D: prediction of protein 3D structures. Bioinformatics 18, 1250-1256.
78 Karplus, K, Barrett, C and Hughey, R (1998). Hidden Markov models for detecting remote protein homologies. Bioinformatics 14, 846-856.
79 Labesse, G and Mornon, J (1998). A tool for incrementing threading optimization (T.I.T.O.) to help alignment and modelling of remote homologues. Bioinformatics 14, 206-211.
80 Lundstrom, J, et al (2001). Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci. 10, 2354-2362.
81 Bonneau, R, et al (2001). Rosetta in CASP4: progress in ab initio protein structure prediction. Proteins: Struct., Funct., and Genet. 4, 119-126.
82 Skolnick, J, et al (2001). Ab initio protein structure prediction via a combination of threading, lattice folding, clustering, and structure refinement. Proteins: Struct., Funct., and Genet. 45, 149-156.
83 Hardin, C, Pogorelov, TV and Luthey-Schulten, Z (2002). Ab initio protein structure prediction. Curr. Opin. Struct. Biol. 12, 176-181.
84 Huang, ES, Samudrala, R and Ponder, JW (1999). Ab initio fold prediction of small helical proteins using distance geometry and knowledge-based scoring functions. J. Mol. Biol. 290, 267-281.
85 Zubrzycki, IZ (2002). Homology modeling and molecular dynamics study of NAD-dependent glycerol-3-phosphate dehydrogenase from Trypanosoma brucei rhodesiense, a potential target enzyme for anti-sleeping sickness drug development. Biophys. J. 82, 2906-2915.
86 Osawa, H and Toyoshima, C (2002). Homology modeling of the cation binding sites of Na+K+-ATPase. Proc. Natl. Acad. Sci. USA 99, 15977-15982.
87 Sabnis, YA, et al (2003). Probing the structure of falcipain-3, a cysteine protease from Plasmodium falciparum: comparative protein modeling and docking studies. Protein Sci. 12, 501-509.
88 Kitson, DH, et al (2002). Functional annotation of proteomic sequences based on consensus of sequence and structural analysis. Briefings in Bioinformatics 3, 32-44.
89 Koretke, KK, Russell, RB and Lupas, AN (2002). Fold recognition without folds. Protein Sci. 11, 1575-1579.
90 Fetrow, JS and Skolnick, J (1998). Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases. J. Mol. Biol. 281:949-968.
91 Gibbs, N, Clarke, AR and Sessions, RB (2001). Ab initio protein structure prediction using physicochemical potentials and a simplified off-lattice model. Proteins: Struct., Funct., and Genet. 43, 186-202
92 Liu, Y and Beveridge, DL (2002). Exploratory studies of ab initio protein structure prediction: multiple copy simulated annealing, AMBER energy fucntions, and a generalized born/solvent accessibility solvation model. Proteins: Struct., Funct., and Genet. 46, 128-146.
93 Srinivasan, R and Rose, GD (2002). Ab initio prediction of protein structure using LINUS. Proteins: Struct., Funct., and Genet. 47, 489-495.
94 De Bakker, PI, et al (2003). Ab initio construction of polypeptide fragments: accuracy of loop decoy discrimination by an all-atom statistical potential and the AMBER force field with the Generalized Born solvation model. Proteins: Struct., Funct., and Genet. 51, 21-40.
95 Tosatto, SC, et al (2002). A divide and conquer approach to fast loop modeling. Protein Eng. 15, 279-286.
96 Caffrey, M (2003). Membrane Protein Crystallization. Journal of Structural Biology 142, 108-132.
97 Collingsworth, PD, Bray, TL and Christopher GK (2000). Crystal Growth via Computer Controlled Vapor Diffusion. J. Crystal Growth 219, 283-289.