

The cost and value of three-dimensional protein structure

During the past few years we have seen the cost of protein three-dimensional structure determination begin to converge by protein family type, there has been a commoditisation of high-throughput structure based drug discovery technology components, and the reliability of timely structure determination of highly valued structures has increased. Although the cost of a structure determination is not insignificant, the value of that structure is much higher. For both basic and applied science, one has only to look at the impact of the Watson and Crick structure determination of DNA 50 years ago. Estimates are that for drug discovery, there is a 50% cost and time savings from target selection to IND filing when structure is centrally used as part of the drug discovery process. Given the large number of intriguing novel targets currently available and the low success rates still observed in going from target selection to lead compound to final marketed drug, tremendous potential remains in the area of structure based drug discovery.

One of the requirements for the rational design of novel drug compounds is structural information on the drug target. Over the past few years, advances in high-throughput (HT) technologies have enabled a pipeline approach for the traditionally long experimental process to proceed from the gene to a validated three-dimensional structure. This reduction in required effort, time and expertise has resulted in a

significant increase in applicability of structural techniques to a wide range of basic and applied sciences, from large macromolecular machines to structural genomics to drug discovery. Three-dimensional structural information of biological systems such as proteins, protein-nucleic acid complexes, macromolecular machines and protein-inhibitor complexes is of critical importance to the understanding of structure-function relationships

**By Professor
Raymond C. Stevens**

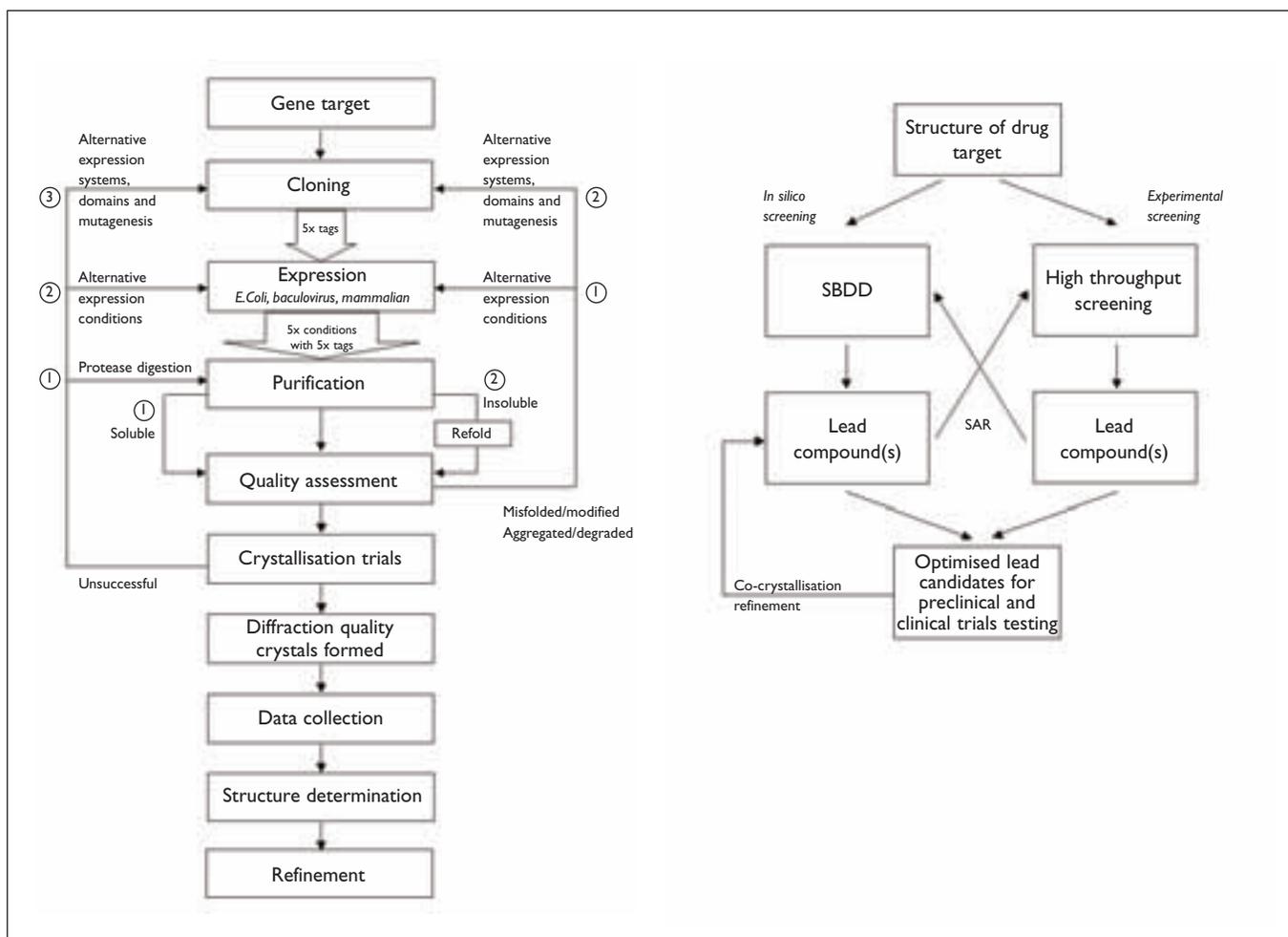


Figure 1

A Overview of the steps involved in HT protein production for structural studies³. With the correct positive and negative controls, one should be able to pass through such a flowchart only once in order to successfully obtain protein samples suitable for structural analysis. The numbers indicate preferences in the flowchart diagram.

B Outline of the steps involved in the rational drug development process. Both virtual (SBDD) and experimental (HT screening, HTS) development steps are required for optimal drug design⁵

and, hence, the functional interpretation of a structure and its interaction with binding partners.

The explosion in interest and investment in structural biology coincides with a number of very significant technological advances that have arisen over the past few years. Together, these advances have enabled the possibility of an overall industrialisation of the structure determination process, similar to the revolution that occurred in sequencing efforts less than a decade ago. However, the multi-dimensional complexity of the structure determination process significantly complicates the streamlining of the entire process and, at present, a significant investment is required to establish the systems for automated cloning, expression, purification, crystallisation, data collection, structure determination, ligand screening and medicinal chemistry optimisation of lead compounds which all need to be interconnected via an advanced information management system. High throughput environments are a prerequisite for

any large scale approach such as structure-based drug design (SBDD) or structural genomics, and is an interesting opportunity for corporate and academic institutions.

Gene to structure – high-throughput structural biology technologies

Historically, three-dimensional protein structure determination has required 1-20 years, depending on the difficulty of the structure determination process steps involved (Figure 1). The bottleneck of suitable protein sample generation was reduced with the molecular biology tools developed in the 1980s and 1990s. Methods currently exist for parallel expression and purification of large numbers of proteins¹⁻⁴, and these technologies enable the exploration of multiple constructs, homologs and variants for each specific protein target.

Crystallisation has remained one of the rate-limiting steps to determine three-dimensional macromolecular structures. Therefore, recent

efforts have focused on developing apparatus and robotics to process crystallisation trials in an accelerated, or HT fashion. Stand-alone workstations have been developed for specific tasks (eg protein drop dispensing) and fully-integrated HT systems have been developed with processing capacities of 2,500 to 140,000 experiments per day and these systems were recently summarised by Kuhn et al, 2002⁵ (Table 1). Automated crystallisation trials using smaller, higher-density drop plating configurations (96-well, 384-well and 1536-well formats) provide for a more condensed experimental scale, minimising crystallisation trial storage space requirements, maximising the number of experiments that can be pursued, significantly reducing overall material requirements, and reducing the cost per crystallisation trial plate. Reduction in drop size to the nanolitre scale has accelerated the formation of diffraction quality

crystals, helped many proteins form crystals which were not forming crystals with microlitre volumes, and further reduces the cost significantly⁶. During the past two years a few of these systems have created several million crystallisation trials and many more million crystallisation images, all with the positive and negative data stored for analysis and efficiency improvements.

Robotic crystallisation trials allow for less error, enable more systematic and routine experimentation, and provide for extensive data mining. Crystallisation trials using sub-microlitre drops produced three major benefits: firstly, 10-100 times reduced costs; secondly, smaller crystallisation drops enabled exploration of a larger crystallisation parameter space; thirdly, the shortened time of crystal formation using smaller drops enables the more rapid analysis of results. In addition, faster sample processing and

Table 1: Commercially available robotic crystallisation systems

EQUIPMENT	CRYSTALLISATION METHOD	REFERENCE
Douglas Instruments IMPAX 1-5/Oryx 6	Microbatch under oil	www.douglas.co.uk
Gilson/Cyberlab C-200/C-240	Sitting or hanging drop	www.gilson.com
Emerald BioStructure RoboHTC™	Sitting drop	www.emeraldbiostructures.com
Syrrx-RTS HTSB Factory	Sitting drop	www.syrrx.com
Matrix Technologies Hydra™	Sitting or hanging drop	www.matrixti.com
Cartesian Technologies Honeybee™, Hummingbird™	Sitting drop	www.cartesiantech.com
DataCentric Automation Rhombix Screen	Sitting drop	www.dcacorp.com
Diversified Scientific VaporPro™	Controlled vapour diffusion	www.dsitech.com
Fluidigm Topaz™	Microfluidic chip-based	www.fluidigm.com
Beckman	Sitting drop	www.beckman.com
Tecan	Sitting drop	www.tecan.com

Table 2: Commercially available crystallisation storage and image analysis systems

COMPANY	HARDWARE/SOFTWARE	REFERENCE
Emerald BioStructures	Crystal Monitor™	www.emeraldbiostructures.com
Syrrx/RTS HTSB Factory	Fully integrated system with CrystalBrain software	www.syrrx.com
RoboDesign/MS-C-Rigaku	RoboStorage™ RoboMicroScope™	www.robodesign.com
The Automation Partnership	HomeBase™	www.automationpartnership.com
Veeco	Optimag 1700, 1750, 1770	www.veeco.com
TriTek	Crystal Pro™	www.tritekcorp.com
Discovery Partners	Crystal Farm™	www.discoverypartners.com
Diversified Scientific	CrystalScore™	www.dsitech.com
DataCentric Automation	Rhombix Vision	www.dccorp.com
BioTom		www.biotom.net

faster crystallisation enables more preserved sample integrity and homogeneity, reducing decomposition, an important issue for selenomethionine-incorporated proteins that are used for MAD structure determinations⁷. Faster crystallisation rates generating smaller crystals can potentially produce more ordered crystals with less imperfections⁶, and these crystals can be used directly for structure determination with high-energy third-generation synchrotrons⁸. In addition, smaller crystals can be more rapidly flash-cooled because of their smaller mass and, hence, exhibit reduced cooling defects and lower crystal mosaicity⁹.

With the large number of crystallisation trial experiments generated using automation comes the concomitant need for automated downstream crystal image processing and analysis. Crystallisation trials require periodic examination of the individual wells over the equilibration cycle of each drop to monitor crystallisation and/or precipitation events. Manual inspection of crystallisa-

tion trials is both tedious and error prone, and is impractical for the number of plates generated robotically. Several robotic systems have been developed and are commercially available for automated plate storage, image acquisition and analysis (Table 2). Although we have seen strong growth in hardware development, the least developed 'protein to crystal' area to date is crystallisation and image system integration and data mining tools. The CrystalBrain system developed by Syrrx is perhaps the most advanced and powerful system currently available. System integration and data mining/efficiency improvements have the most potential for growth in the next phase of technology development.

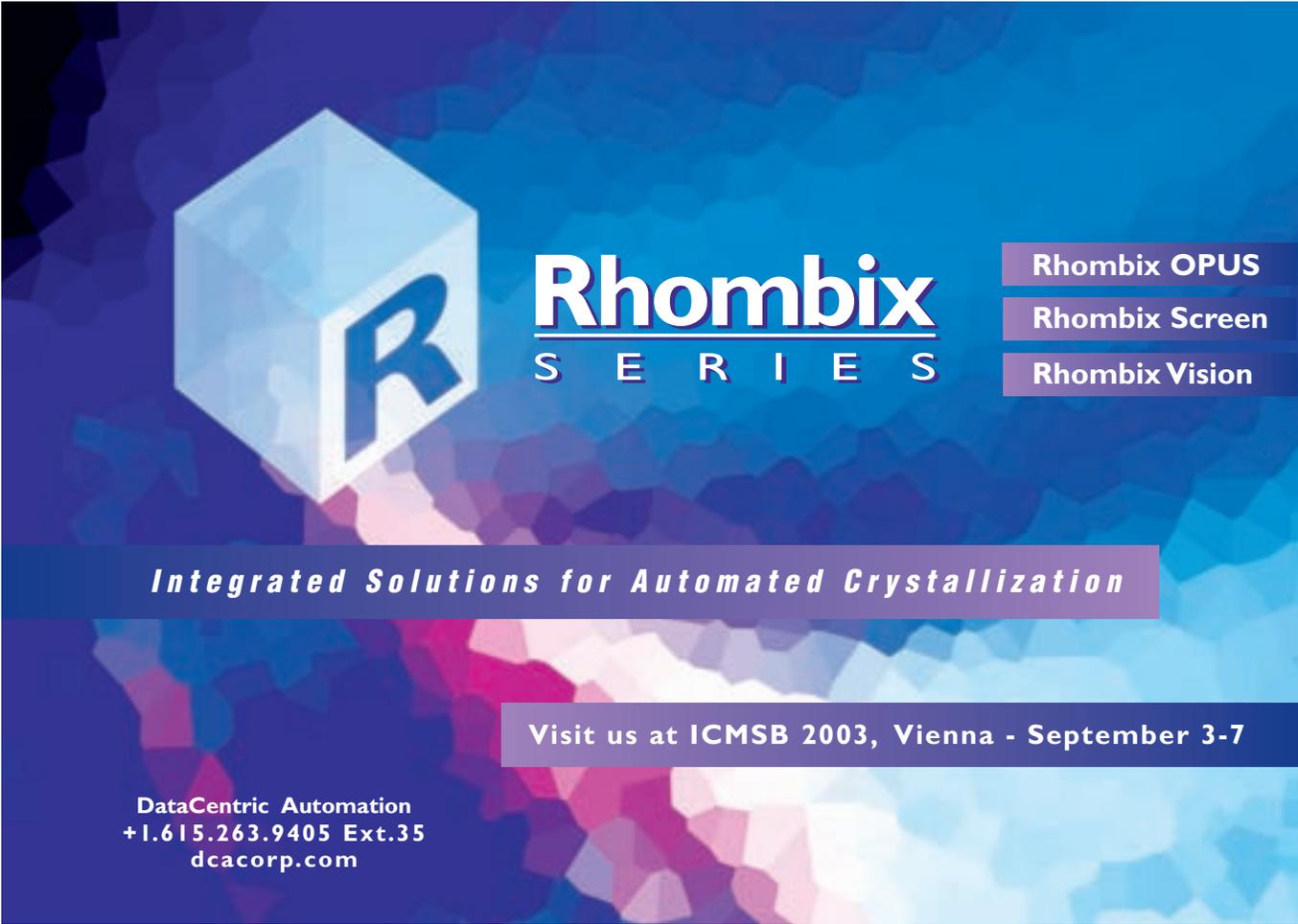
The last steps in the structure determination process are data collection and structure solution. Although data collection with x-rays can be conducted in-house with many samples, the majority of final data collection is conducted at synchrotron radiation facilities (less than 10 synchrotron facilities are available worldwide) that

produce highly parallel x-rays that can be finely collimated to produce a focused beam that is extremely intense (orders of magnitude more intense than in-house x-ray sources) and tunable. These two aspects make synchrotron-generated radiation an important tool for studying macromolecular structures. Some of the critical technology developments used at synchrotrons include the use of multiple-wavelength anomalous dispersion (MAD) using selenomethionine incorporation^{7,10} and diffraction data collection using flash-cooling¹¹. More recently, automated crystal mounting systems have been developed by Abbott Laboratories¹² (ACTOR system, now sold by Molecular Structure Corporation), RoboHutch developed at the Advanced Light Source¹³, and the 96-crystal cassette system developed at Stanford Synchrotron Radiation Laboratory¹⁴. Finally, several programmes have been developed that are able to take diffraction data and determine the three-dimensional struc-

ture with reduced user intervention such as PHENIX, CrystalNet, SHARP, Solve/Resolve and arp/Warp, although these systems are constantly being improved to handle the wide range of data quality cases observed in protein crystallography^{15,16}. All of these recent developments are powerful new tools, but there is still a need for significant improvements to increase reliability and speed, and the next few years will see continued major advancements.

Gene to structure – learning factory

Even though new technology has emerged in the field of structural biology, the current efficiency is not optimal, as the technology is only as good as what you place into the pipeline. One example of where this technology is allowing the collection of data in a more systematic manner for data mining is for the protein crystallisation step. The Joint Center for Structural Genomics published its efforts on cloning, expressing, purifying and



Rhombix
S E R I E S

- Rhombix OPUS
- Rhombix Screen
- Rhombix Vision

Integrated Solutions for Automated Crystallization

Visit us at ICMSB 2003, Vienna - September 3-7

DataCentric Automation
+1.615.263.9405 Ext.35
dcacorp.com

Table 3: Summary of results to date for high-throughput protein expression, purification and crystallisation efforts with the *T. maritima* proteome¹⁷

T. MARITIMA HT PIPELINE	# PROCESSED	% SUCCESS
Targets	1,877	100
Amplifiers	1,791	95
Expression clones	1,376	73
Proteins attempted expression	1,376	73
Proteins prepared for crystallography	542	29
Proteins crystallised	456	24

Table 4: Protein crystallisation condition success rate analysis of the *T. maritima* proteome screen¹⁸. Note: 542 proteins were passed through crystallisation robot with 480 conditions per screen

PERCENTILE SUCCESS RATE	# OF CONDITIONS	% OF CONDITIONS
100%	143	30%
95%	122	25%
90%	97	20%
85%	76	16%
80%	65	13%
50%	14	3%
43%	10	2%

References

- 1 Gilbert, M et al (2002). Accelerating code to function: sizing up the protein production line. *Curr Opin Chem Biol* 6:102–105.
- 2 Lesley, SA (2001). High-throughput proteomics: protein expression and purification in the postgenomic world. *Protein Expr Purif* 22:159–164.
- 3 Stevens, RC (2000). Design of high-throughput methods of protein production for structural biology. *Struct Fold Des* 8:R177–R185.
- 4 Waldo, GS et al (1999). Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17:691–695.
- 5 Kuhn, P et al (2002). The genesis of high-throughput structure based drug discovery using protein crystallography. *Curr Opin Chem Biol* 6:704–710.
- 6 Santarsiero, BD et al (2002). An approach to rapid protein crystallization using nanodroplets. *J Appl Crystallogr* 35:278–281
- 7 Hendrickson, WA et al (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.
- 8 Stevens, RC (2000). High-throughput protein crystallization. *Curr Opin Struct Biol* 10:558–563.
- 9 Goodwill, KE et al (2001). High-throughput X-ray crystallography for structure-based drug design. *Drug Discovery Today* 6 (Genomics Suppl):S113–118.
- 10 Guss, JM et al (1988). Phase determination by multiple-wavelength X-ray diffraction: crystal structure of a basic “blue” copper protein from cucumbers. *Science* 241:806–811.
- 11 Garman, E (1999). Cool data: quantity AND quality. *Acta Crystallogr D* 55:1641–1653.
- 12 Abola, E et al (2000). Automation of X-ray crystallography. *Nat Struct Biol* 7:973–977.

Continued on page 47

Table 5: The cost of novel drug target three-dimensional structure determination by protein crystallography and associated estimated success rates*

* The average cost is based on the listed success rate; if a higher or faster success rate is needed, the cost accordingly needs to be adjusted upwards. Bacterial and human soluble protein structure determination programmes have an estimated maximum success rate of 75%; membrane protein structure determination programmes do not have enough data to provide an estimated maximum success rate. The average cost estimates do not include the costs of technology development or large equipment amortisation. Estimates are based on both public and industry data

NOVEL DRUG TARGET	AVERAGE COST	SUCCESS RATE
Soluble bacterial targets	\$140,000	35%
Soluble human protein (eg kinases, proteases, NHRs)	\$450,000	35%
Bacterial membrane proteins	\$1.5 million	10%
Human membrane proteins	\$2.5 million	10%

crystallising as many of the proteins in the *Thermotoga maritima* proteome (1,877 genes) as possible (Table 3)¹⁷. With this data all compiled within six months, positive and negative data mining was possible in order to determine the most cost-efficient screening conditions (Table 4)¹⁸. For the crystallisation portion of its experimentation alone, it spent approximately \$140,000. If traditional technology (eg available four years prior to these efforts) had been used, the crystallisation trials would have cost more than \$1.4 million, with the largest cost reduction coming from process miniaturisation. Based on crystallisation trial data analysis, if only the most efficient crystallisation conditions were used, the entire proteome crystallisation screen analysis could have been performed for less than \$43,000. This represents a significant saving, and further savings such as this will continue to be required to reduce the cost and increase the efficiency of protein structure determinations. With these HT structure determination advances, not only are we seeing cost savings, but also increases in the reliability and speed to determine the necessary three-dimensional structures are being realised.

Structure to drug – modelling, virtual and physical ligand screening and medicinal chemistry

In addition to the described advances in ‘gene to structure’, a number of advances have also been

made in the ‘structure to drug’ pipeline and have been described elsewhere^{7,10–16}. Historically, experimental screening approaches (eg HTS) were used to find potential drug leads, however, virtual ligand screening (VLS, or *in silico* screening) methods have emerged to assist rational drug development¹⁹. VLS involves computer-based screening of large chemical libraries against structural and electrostatic information for target proteins and is a powerful and rapid tool to direct, or ‘focus’ the design of experimental libraries. A variety of *in silico* docking and scoring tools are used to computationally find favourable small molecule/protein interaction partners^{20–21}, incorporating more than one VLS method in parallel and comparing the results provides more favourable drug leads²². Unfortunately, the success rate for VLS is quite low, on the order of less than 10%. It is therefore critical to validate and/or supplement with experimental methods like NMR-based or x-ray crystallographic-based co-complex screening methods. The ‘SAR by NMR’ methodology pioneered by Fesik²³ (SAR, structure–activity relationship) yields important information for lead optimisation, and NMR-based binding assays are used in SBDD efforts (eg see triadtherapeutics.com). X-ray crystallographic screening is included in Abbott Laboratories (Abbott Park, IL, USA)²⁴, Vertex Pharmaceuticals and Astex Technology²¹ SBDD efforts. VLS and better predictive ADME/Tox filtering of compounds, in combination with the recent

Table 6: Drug targets with inhibitors designed using SBDD methods⁵

HIV protease	bcl-2
renin	neuraminidase
factor Xa	glyceraldehyde-3-phosphate dehydrogenase
thrombin	interleukin-1 beta-converting enzyme (ICE, caspase-1)
factor VIIa/TF	thymidylate synthase
urokinase	protein-tyrosine phosphatase
urokinase-type plasminogen activator	acetylcholinesterase
rhinovirus 3C protease	prostaglandin receptor
cathepsin B, L, and S	matrix metalloproteinase (MMP)
20S proteasome (chymotrypsin-activity)	trihydroxynaphthalene reductase (3HNR)
memapsin 2 (beta-secretase)	dihydrofolate reductase (MAC vs human)
Abl kinase	cyclooxygenase-2
CDKs	neuronal nicotinic receptor agonists
EGFR kinase	5 α -reductase
Lck kinase	monoamine oxidase-B
Src kinase	N-myristoyl transferase
aldose reductase	hypoxanthine-guanine-xanthine phosphoribosyltransferase
phospholipase-A2	DNA gyrase
STAT	human carbonic anhydrase II
bcl-2	peptide deformylase
DPP4	HDAC

advancements in structural biology, enable HT protein three-dimensional structure determination and rapid determination of small-molecule/protein co-complex structures, enabling SBDD lead identification and optimisation at a faster and more robust level. Lastly, the power to use structure-based methods ultimately lies in the hands of the chemists who synthesise the lead compounds, and until they more fully embrace the SBDD approach, which still appears that only a subset of chemist currently do, efficiencies of drug discovery will always be limited by the

amount of chemistry that one can apply to a given drug design programme.

Cost and value of protein structure determination

Important considerations in SBDD are the success rate for highly valued drug target structure determination, the time required to obtain the structure, and the overall cost of the structure determination process. The values listed in **Table 5** are cost estimates for four different protein classes, and provide a relative cost comparison for the major families of

drug targets. Although only very modest decreases in the cost of structure determination have been observed over the past few years, we have seen the success rate improve significantly and it is anticipated that these costs will be reduced in the coming years as the new HT technologies are embraced and attain improved efficiencies and economies of scale. The cost estimates listed in Table 5 do not apply to co-crystallisation experiments, which are 10-20% of the initial *de novo* structure cost and requires the crystal form used in the initial structure determination be amendable to co-crystallisation with small molecule compounds. Currently, the time average for soluble protein targets is one year for a novel structure determination, although this can be accomplished much faster at an increased cost. Finally, for small soluble proteins less than 25 kilodaltons in size, NMR appears to be faster and more economical.

Given an average cost from target identification

to investigational new drug (IND) filing estimated at approximately \$15-20 million for a single successful programme, SBDD methods can reduce this cost by more than 50%, with cost reductions largely attributed to the quality of lead candidates and number of different pharmacophore series that can be designed based on the structural information. The cost of the actual structure determination process is not a significant percentage of this overall amount, but the value of this structural information is obviously much higher due to the increase in lead compound quality. In addition to the reduced cost estimates for successful SBDD campaigns, the structural information also has tremendous value in terms of deciding which drug discovery projects to terminate, as the active site definition that is obtained with a protein three-dimensional structure allows one to rapidly ascertain whether or not an effective small molecule drug can be created against a particular target.



Main Conference: 24th – 25th September 2003 Focus Day: 23rd September 2003, Hilton Hotel, Mainz, Germany

Understanding, Analysing & Utilising Proteomic, Genomic & Pharmacogenomic Data

— 38 Essential Case Studies Including —

DATA INTEGRATION
HIGH THROUGHPUT SCREENING
DATABASES & DATAMINING
IN SILICO TECHNIQUES

PROTEOMICS
COMPUTER MODELING & VISUALISATION
MICROARRAYS
CHEMOINFORMATICS

“ This conference looks set to be the best commercially focused informatics conference happening this year ”

Dr. Simon Brocklehurst, Director of Informatics & Robotics,
CAMBRIDGE ANTIBODY TECHNOLOGY



For A full Conference Brochure Contact Colin Chapman & quote ref: DDW
 Tel: +44 (0)20 7368 9413 • Fax: +44 (0)20 7368 9401 • Email: colin12@wbr.co.uk • Web: www.wbr.co.uk

Table 7: Example of the diversity of experiments performed at Syrrx on several eukaryotic and prokaryotic drug discovery targets²⁸
Constructs made: total number of C-terminal- and N-terminal-tagged vectors constructed for microexpression studies. Purification diversity: total number of purification strategies employed. Coarse screen experiments: total number of sparse matrix and grid screen crystallisation experiments. Fine screen experiments: total number of fine screen crystallisation experiments for crystal optimisation. Total images: total number of images acquired and analysed. Resolution: average resolution of optimised crystals from nanolitre-volume experiments

TARGET	CONSTRUCTS MADE		PURIFICATION DIVERSITY		COARSE SCREEN EXPERIMENTS		FINE SCREEN EXPERIMENTS		TOTAL IMAGES		RESOLUTION
	N TAG	C TAG	N TAG	C TAG	N TAG	C TAG	N TAG	C TAG	N TAG	C TAG	
KINASES											
Aik	13	2	18	2	31,392	3,072	9,120	0	202,560	15,360	1.9
Fgfr2	3	1	29	0	48,192	0	10,752	0	294,720		2.8
EphA2	8	1	12	1	15,264	0	2,304	384	87,840	1,920	2.2
Akt3	17	2	22	0	15,360	0	7,104	0	112,320		2.1
Fak	6	1	26	0	34,656	0	5,376	0	220,160		1.4
Dap	15	5	10	2	22,272	4,416	6,912	0	145,920	22,080	1.9
Ick	6	1	30	2	11,424	0	3,744	0	75,840		2.3
PROTEASES											
DPP4	22	4	34	7	45,888	14,784	53,856	2,592	498,720	86,880	2.1
FAP	3	4	0	14	0	5,760	0	10,176	–		2.3
DPP8	6	2	10	4	2,400	2,688	1,824	384	21,120	15,360	2.7
PROKARYOTIC											
MurA-Hi	4	2	8	2	8,352	2,304	2,976	0	56,640	–	2.3
MurB-Hi	1	1	3	2	5,376	0	480	0	29,280	–	1.5
MurC-Hi	44	7	53	8	71,712	9,216	4,704	384	382,080	48,000	1.7
IspD-Hi	0	4	0	5	0	5,568	0	1,128	–	36,480	1.9
RfaD-Hi	0	4	0	3	0	2,016	0	2,304	–	21,600	2.3
AroE-Hi	0	2	0	5	0	8,448	0	4,128	–	62,880	2.0
Eno-Hi	0	2	0	5	0	5,760	0	96	–	29,280	2.3
Prb-Hi	0	2	0	4	0	2,304	0	960	–	16,320	2.2
Asd-Hi	0	1	0	2	0	2,304	0	96	–	12,000	1.7

Continued from page 42

Successes in structure-based drug discovery

The first companies to focus on SBDD included Vertex Pharmaceuticals in Cambridge, MA and Agouron in La Jolla, CA more than 15 years ago. Given the average time to develop a drug from the target selection stage to the market, it is not surprising that we are only now beginning to see the rewards from these early-stage innovators in the field of SBDD. Using SBDD, drugs have been designed for targets such as proteases, kinases and an expanding number of other biological macromolecules (Table 6). Notable structure-based marketed drugs include the HIV protease inhibitors Viracept^{TM25} (Agouron, USA and Eli Lilly, USA) and Agenerase^{TM26} (Vertex, USA; Kissei, Japan; and Glaxo Wellcome, UK) and the neuraminidase inhibitors Relenza^{TM27} (Biota, Australia and Glaxo Wellcome, UK) and Tamiflu^{TM28} (Gilead Sciences, USA and Roche, Switzerland). Based on the increasing number of small-molecule therapeutics derived using SBDD and the recent development of HT structural biology technology platforms, rational SBDD approaches will be increasingly important in future lead discovery and optimisation efforts.

As an example of the power of these next generation technologies, over the past 18 months Syrrx has crystallised and solved the structures of 73 unique structures representing drug discovery targets from both eukaryotic and prokaryotic sources. In addition, it has collected 3-Å or better diffraction data from 28 other proteins that have yet to be solved, and have crystallised 11 other proteins that did not initially diffract well enough to collect full datasets (Table 7)²⁹. The HT structure determination technology available at Syrrx enabled rapid success for each protein target by employing significant diversity at each step of the process. As expected, the number of experiments required to successfully determine a soluble prokaryotic protein structure was typically much less than that required to process human protein targets. In all cases, structure determination relied upon crystals grown only from nanolitre volume crystallisation experiments. Importantly, the fully integrated crystallisation platform available at Syrrx allowed each of the described targets to be performed in approximately three months. Since the impact of structural information on lead optimisation is highest when this information is available early, the parallel testing of multiple protein samples and crystallisation conditions realised with HT methods for rapid structure

determination accelerates the conversion of initial leads into clinical candidates.

Conclusions

Current HT systems have already shown a reduction in the time requirement from years to months or even weeks for novel structure determination, and even faster rates for lead discovery and optimisation. These advances have eliminated many of the previous bottlenecks and are presenting the most direct path to accelerate the development of novel therapeutics. Based upon the capabilities of HT structural biology, SBDD will increasingly be relied upon for therapeutic lead compound discovery and optimisation. What remains as a challenge to the field is the integration of the different technologies from gene to clinical candidate along with innovations for improved efficiencies. There is the need for better integration of the medicinal chemistry and structural biology efforts, as well as the complementation of compound screening (eg HTS) methods with SBDD methods. In addition, SBDD for the largest family of drug targets, namely membrane proteins has for the most part been ignored, except by the European consortium MepNet/BioXtal, a few academic labs, exploratory pilot projects within industry, and the dedicated SBDD GPCR and ion channel effort at Sagres Discovery. On a positive note, we have seen a tremendous increase in the number of membrane protein structures that have been solved at atomic resolution, many of which are highly valued drug targets such as COX-2 (Table 8). Unfortunately, these structures were not available during the drug discovery process and it remains a challenge to determine membrane protein structural information in a timely manner. Fortunately, this situation is reminiscent of the historical structure determination status for other protein families such as kinases and antibodies, where these first structures were solved in the early 1990s, and at that time they were considered to have large flexible loops and would be too challenging to solve on a regular basis, but are nowadays considered relatively routine structures to determine.

Acknowledgements

The author would like to thank Dr Peter Kuhn, Angela Walker and Dr Marianne G. Patch for assistance with the manuscript. **DDW**

Raymond C. Stevens is a professor of Molecular Biology and Chemistry at The Scripps Research Institute. He was one of the founders of Syrrx, a high-throughput structure-based company focused

- 13** Muchmore, SW et al (2000) Automated crystal mounting and data collection for protein crystallography. *Struct Fold Des* 8:R243–R246.
- 14** McPhillips, TM et al (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synchrotron Radiat.* 9:401–406.
- 15** Adams, PD et al (2000). Recent developments in software for the automation of crystallographic macromolecular structure determination. *Curr Opin Struct Biol* 10:564–568.
- 16** Perrakis, A, et al (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463
- 17** Lesley, SA et al (2002). Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci USA* 99:11664–11669.
- 18** Page, R et al (2003). Shotgun crystallization strategy for structural genomics: An optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome. *Acta Crystallogr D* 59:1028–1037.
- 19** Klebe, G (2000). Recent developments in structure-based drug design. *J Mol Med* 78:269–281.
- 20** Abagyan, R et al (2001). High-throughput docking for lead generation. *Curr Opin Chem Biol* 5:375–382.
- 21** Blundell, TL et al (2002). High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* 1:45–54.
- 22** Charifson, PS et al (1999). Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 42:5100–5109.

Continued on page 48

Continued from page 47

- 23** Shuker, SB et al (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534.
- 24** Nienaber, VL et al (2000). Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol* 18:1105–1108.
- 25** Kaldor, SV et al (1997). Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J Med Chem* 40:3979–3985.
- 26** Kim, EE et al (1995). Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J Am Chem Soc* 117:1181–1182.
- 27** von Itzstein, M et al (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363:418–423.
- 28** Kim, CU et al (1997). Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* 119:681–690.
- 29** Hosfield, D et al (2003). A fully integrated protein crystallization platform for small molecule drug discovery. *J Struct Biol* 142:207–217.

Table 8: Selected membrane protein structures organised by class with year of structure determination
 Information obtained from: www.mpibp.frankfurt.mpg.de/michell/public/memprotstruct.html
http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html

Monotopic membrane proteins	Beta-barrel
Prostaglandin H2 Synthase (1994)	TolC (2000)
Cyclooxygenase-2 (COX-2) (1996)	OmpA (1998)
Squalene Cyclase (1999)	OmpX (1999)
Monamine Oxidase (2002)	OmpLA (1999)
Fatty Acid Amide Hydroxylase (FAAH)	FhuA (1998)
(2002)	FecA (2002)
Rhodopsins	Prokaryotic Ion Channels
Bacterial Rhodopsins (1997)	KcsA K+ (1998)
Halo Rhodopsins (2000)	MthK K+ (2002)
Sensory Rhodopsins (2001)	MscL Mechano (1998)
GPCR Bovine Rhodopsin (2000)	Chloride (2002)
	Aquaporin (2001)
Photosynthetic Reaction Center (1985)	ATPase (2000)
Light Harvesting Complexes (1995)	P450s (2000)
Toxins	Respiratory
Alpha Hemolysin (1996)	Fumerate Reductase (1999)
LukF (1999)	ATP Synthase (1999)
Porins (6 different) (1992)	Cytochrome C oxidase (1995)
Photosystem I (1997)	Cytochrome bcl complex (1997)
Photosystem II (2001)	

on soluble drug targets; a founder of MemRx, a GPCR and ion channel structure-based drug discovery company (MemRx was recently acquired by Sagres Discovery); and a founding member of the Joint Center for Structural Genomics, one of the

NIH funded structural genomics centres. He has published more than 120 peer-reviewed articles, holds several patents in the area of structure-based drug discovery and has received several awards for his work in this area.