Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Three-dimensional structure determination of proteins related to human health in their functional context at The Israel Structural Proteomics Center (ISPC)

Shira Albeck,[a] Yigal Burstein,[b] Orly Dym,[a] Yossi Jacobovitch,[a] Nurit Levi,[a] Ran Meged,[a] Yigal Michael,[a] Yoav Peleg,[a] Jaime Prilusky,[c] Gideon Schreiber,[d] Israel Silman,[e] Tamar Unger[a] and Joel L. Sussman[a]*

[a]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel, [b]Department of Organic Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel, [c]Department of Biological Services, Weizmann Institute of Science, Rehovot 76100, USA, [d]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, USA, and [e]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel

Correspondence e-mail: joel.sussman@weizmann.ac.il

Received 7 September 2004
Accepted 22 July 2005

This paper was presented at ICCBM10.

The principal goal of the Israel Structural Proteomics Center (ISPC) is to determine the structures of proteins related to human health in their functional context. Emphasis is on the solution of structures of proteins complexed with their natural partner proteins and/or with DNA. To date, the ISPC has solved the structures of 14 proteins, including two protein complexes. It has adopted automated high-throughput (HTP) cloning and expression techniques and is now expressing in *Escherichia coli*, *Pichia pastoris* and baculovirus, and in a cell-free *E. coli* system. Protein expression in *E. coli* is the primary system of choice in which different parameters are tested in parallel. Much effort is being devoted to development of automated refolding of proteins expressed as inclusion bodies in *E. coli*. The current procedure utilizes tagged proteins from which the tag can subsequently be removed by TEV protease, thus permitting streamlined purification of a large number of samples. Robotic protein crystallization screens and optimization utilize both the batch method under oil and vapour diffusion. In order to record and organize the data accumulated by the ISPC, a laboratory information-management system (LIMS) has been developed which facilitates data monitoring and analysis. This permits optimization of conditions at all stages of protein production and structure determination. A set of bioinformatics tools, which are implemented in our LIMS, is utilized to analyze each target.

1. Introduction

Proteomics is a new field of research that has emerged in the past decade from spectacular advances in genomics, in particular the deciphering of the DNA sequences of the entire human genome and those of many other organisms. Genes provide cells with the 'dictionary' of the amino acids that determine a protein's primary sequence. It is the proteins that carry out the molecular functions of the human body: generation of energy, production of cellular components, degradation of waste products, regulation of cellular processes and fighting disease.

Advances in genomics provide valuable information about the composition of proteins, but little about their structure and, ultimately most crucially, little concerning their function. Indeed, the functions of most proteins are still unknown. In order to understand how and why proteins function as they do, it is essential to know their three-dimensional structures. Thus, in 2000, the US National Institute of General Medical Sciences (NIGMS) initially established seven Structural Genomics Centers and subsequently established two additional ones in order to develop and utilize efficient high-throughput (HTP) approaches and methodologies for achieving this difficult and

time-consuming task (Chance *et al.*, 2002; Service, 2000; see http://www.nigms.nih.gov/psi/). Subsequently, major initiatives were established elsewhere (Stevens *et al.*, 2001), including Canada (Yee *et al.*, 2003), Japan (Yokoyama, 2003) and Europe (Heinemann *et al.*, 2000; Leulliot *et al.*, 2005). These initiatives have already resulted in impressive achievements (Todd *et al.*, 2005) in helping biologists to study structure–function relationships and in the design of new drugs. In addition, they have spawned new developments in protein structure prediction (Shah *et al.*, 2005).

Following the large-scale NIH-funded US initiatives (see http://www.nigms.nih.gov/psi/), the European Commission funded the first pan-European project, Structural Proteomics in Europe (SPINE; see http://www.spineurope.org/), which is focused on target proteins related to human health and disease.

The seed money received from SPINE permitted the establishment of a structural proteomics initiative at the Weizmann Institute of Science (WIS). Partly as a consequence of WIS participation in SPINE, the Israel Ministry of Science and Technology, in the fall of 2002, selected it as the site of the 'Israel Structural Proteomics Center (ISPC)' (see http://www.weizmann.ac.il/ISPC), making possible the purchase of HTP robotic instruments and 'state-of-the-art' equipment. The goal of the ISPC is to determine the structures of proteins related to human health in their functional context. Behind each target lies a scientific question and, owing to the importance of each target, we apply several approaches in order to increase the chances of overcoming the numerous obstacles along the production pipeline. These include expression of each target in several expression systems, expression and/or purification of targets together with their natural binding partners, so as to increase their solubility or stabilization, and utilization of bioinformatics tools to assist manipulation and engineering of proteins so as to increase their solubility as well as their ability to crystallize. For this purpose, the center utilizes HTP technologies to facilitate handling of the large number of trial experiments generated for each target.

Much effort is being devoted to determination of the three-dimensional structures of protein complexes. In many biological processes proteins form complexes with other proteins. One may mention signal transduction, control of gene expression, enzyme inhibition, antibody–antigen interaction, hormone-receptor recognition and even the assembly of multi-domain proteins. Consequently, study of the structure of the complex of a protein with its binding partner provides a valuable approach to understanding how it functions in its cellular context. Moreover, solution of the three-dimensional structure of a protein complex provides important information for understanding the molecular basis of protein–protein interactions. In recent years, it has become apparent that some proteins are 'natively unfolded', *i.e.* intrinsically disordered in isolation; they may thus only adopt a folded conformation when complexed with a partner protein (Dyson & Wright, 2005). In such cases, we use either co-expression of the component proteins or co-refolding of the partners to achieve direct assembly of the functional complex.

As already mentioned, most of the targets studied at the ISPC are proteins related to human health or human diseases. This effort has already resulted in the determination of 14 protein structures, some of which are related to clinical conditions such as Gaucher's disease, atherosclerosis and Alzheimer's disease.

The strategy adapted for structure determination is outlined in Fig. 1. There are two entry points into the ISPC pipeline. One is at the cloning stage, where a gene of interest is cloned, expressed and purified at the ISPC, followed by crystallization and structure determination. Alternatively, scientists may submit purified proteins directly for crystallization and subsequent structure determination.

2. Methods

2.1. Cloning and expression in *Escherichia coli*

Small-scale cloning and expression employ HTP methodologies that utilize robotic equipment (Fig. 2). Both the ligation-independent cloning Gateway system (Invitrogen) and conventional methods are employed. We are using pET-based expression vectors (Novagen) containing various tags useful for subsequent purification. Each protein is expressed tagged and/or in its native form. The His tag (His_6), thioredoxin (Trx) and glutathione-*S*-transferase (GST) are all engineered with a protease-cleavage site (TEV) that permits subsequent removal of the tag. In cases where coexpression of two proteins is required, they are cloned under two separate promoters.

Cloned DNAs are introduced simultaneously into various *E. coli* strains, *e.g.* BL21(DE3)pLysS, Rosetta(DE3)pLysS, Rosetta–gamiB(DE3), thus increasing the probability that a given protein will be expressed in soluble form and in high yield. Expression is screened on a small scale in 4 ml cultures in 24-deep-well plates at two temperatures (288 and 303 K) using a Tecan robot (Fig. 2*a*). Soluble and insoluble cellular fractions are analyzed by SDS–PAGE for protein expression.

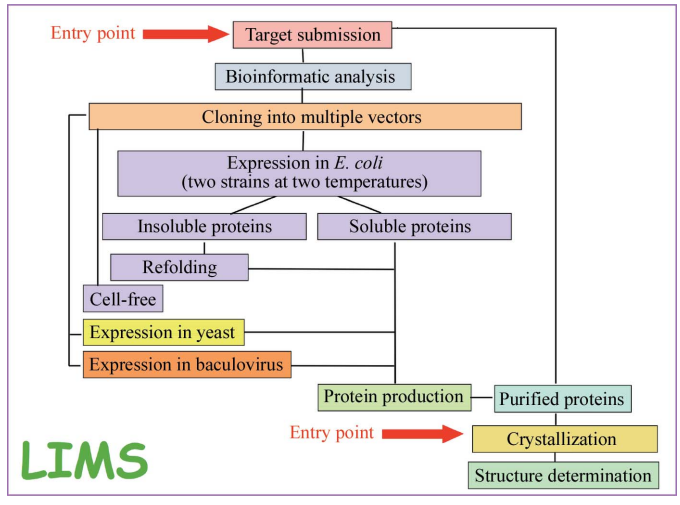

Figure 1
Flow diagram for handling of targets submitted to the ISPC.

Once optimal conditions have been determined, large-scale cultures (4.2 l) are used to obtain larger amounts of protein.

2.2. *In vitro* bacterial expression

For toxic proteins, which cannot be expressed in *E. coli*, cell-free expression is being applied for the assessment of expression and solubility. The cell-free system is based on an *in vitro* protein-synthesis system that couples transcription and translation from a recombinant DNA. We are using a bacterial extract prepared 'in-house' based on a procedure developed in the Genomic Science Center at RIKEN, Japan (Kigawa *et al.*, 1999). In cases in which the template DNA is linear, purified λ phage Gam protein is added to the lysate to inhibit exonuclease activity (ExoV). DNAs cloned using the Gateway system can be used for screening in the cell-free system in the 96-well format. While this procedure is fast and convenient, yields of protein are still low.

2.3. Expression in *Pichia pastoris*

Expression of proteins in the yeast *P. pastoris* (see Fig. 3) is directed to biosynthesis of either intracellular or secreted protein and in both cases the proteins bear a removable N-terminal His tag. Transformation of the linearized vector is performed by electroporation into a *P. pastoris his4* host, utilizing Invitrogen strains GS115, KM71 or SMD1168. Multiple integration events of the target gene are screened by selection for increasing resistance to the antibiotic G418. Colony PCR is performed on selected clones to verify the presence of an intact gene in the *Pichia* genome. Small-scale screening to identify the most effective clones is performed in 50 ml of BMGY medium in baffled flasks at 303 K. Induction with methanol is performed at different temperatures (293–303 K), using various media and additives. Methanol is added [1%(*v*/*v*)] every 24 h throughout the induction stage. Samples are taken periodically (up to 168 h induction). Large-scale production is performed on a 2 l scale in baffled flasks.

(*a*) (*b*) (*c*) (*d*)

Figure 2
Robots used at the ISPC. (*a*) Tecan Freedom Evo, (*b*) AKTA Explorer, (*c*) Douglas Oryx 6 crystallization robot, (*d*) TriTek CrystalPro visualization robot.

2.4. Refolding of inclusion bodies

About 75% of mammalian proteins express in *E. coli* as inclusion bodies. We are developing an automatized folding screen that utilizes a pipetting robot (Fig. 2*a*). The basic method involves solubilization of the inclusion bodies by a chaotropic agent such as urea or guanidinium chloride (with or without a reducing agent). The His-tagged protein is partially purified in the denatured state by capture on Ni–NTA. It is then diluted into various buffers containing additives such as salts, polar additives (*e.g.* arginine), osmolytes (*e.g.* PEG), detergents and chaotropes at three different pH values. Typically, up to 50 different combinations are screened. Folding of a protein is validated by subjecting its clear solution to analytical gel filtration or to gel electrophoresis under non-denaturing conditions.

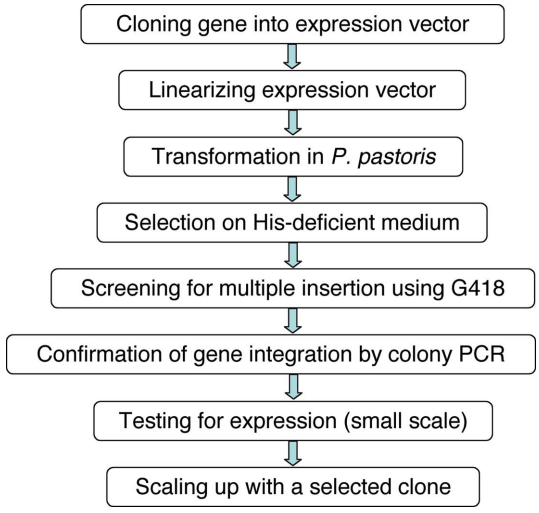

Figure 3
From gene to protein in *P. pastoris*. Sequential steps carried out in the yeast system for expression of heterologous proteins.

2.5. Protein purification

Our current strategy is to use tagged proteins, permitting streamlined purification and TEV cleavage for a large number of samples. Prior to crystallization, for which a protein must be >90% pure, three purification steps are conducted: (i) capture by affinity chromatography, (ii) an intermediate purification step involving either ion-exchange or hydrophobic chromatography and (iii) gel filtration. This last step is important since it removes aggregates, which can reduce the chances of success in crystallization screens.

For protein purification, the ISPC has purchased an AKTA 3D kit coupled to an AKTA Explorer system (Amersham) (Fig. 2*b*). This permits automated purification of multiple samples of soluble proteins fused to affinity tags.

Finally, the purity and homogeneity of a protein sample are established by analytical gel filtration and SDS–PAGE.

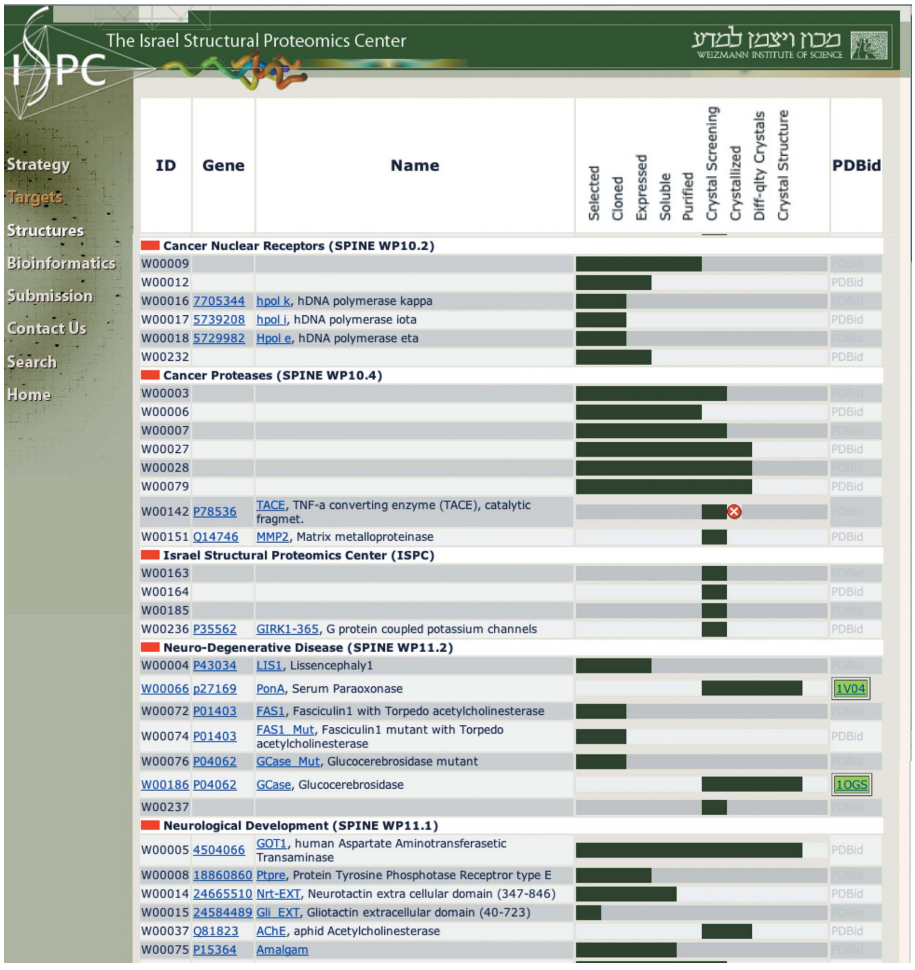

Figure 4
ISPC targets. A partial list of target proteins and protein complexes being handled by the ISPC, which also shows their status (see http://www.weizmann.ac.il/ISPC/status.html).

2.6. HTP protein crystallization

To obtain protein crystals suitable for three-dimensional structure determination we use the Douglas Instrument IMPAX 1–5 and Oryx 6 robots (see Fig. 2*c*). Both instruments employ the microbatch method under oil, which is very rapid and consumes only small amounts of protein and precipitation agents, making it suitable for HTP crystallization experiments. We use ~600 different conditions per target (~15 different commercial crystallization kits) varying in their precipitation agent, pH, salt, detergents and additives. In addition, we have prepared the PEG/Ion/pH screen (Newman *et al.*, submitted) and are in the process of preparing the random screen developed by Bernhard Rupp (Lawrence Livermore National Laboratory, UC Berkeley, USA). Once crystals have been obtained, optimization of their size

and diffracting power is performed. This is performed by slightly changing the composition of the precipitating solution, the pH, the temperature, the drop volume, the protein concentration and the type of oil used (paraffin oil, silicone oil or various ratios of the two). We are using a TriTek CrystalPro visualization robot to permit routine and rapid viewing and assessment of thousands of crystallization trials (Fig. 2*d*), which can easily be seen *via* a web-based browsing tool.

2.7. Laboratory information-management system (LIMS)

In order to record, organize and analyze the enormous amount of data that the ISPC is accumulating, we are collaborating with the data-management teams of the WIS Information Systems and Bioinformatics Centers in the development of a LIMS. This ORACLE-based system facilitates data analysis, thus permitting optimization of conditions at all stages of protein production and structure determination. In parallel, a number of the tools developed for this LIMS system have been ported to the HalX LIMS system (Prilusky, Oueillet *et al.*, 2005; see §3.2).

3. Results and discussion

3.1. Target proteins for structure determination

Genes for target proteins, as well as purified proteins, are being received from research groups throughout Israel, including local biotechnology companies interested in solving the structures of proteins for pharmaceutical purposes. The center has a particular interest in targets related to human health and disease in the following categories.

(i) Proteins and protein complexes linked to neurological development, such as nuclear receptors and cholinesterase-like adhesion proteins (CLAMs), and to neurodegenerative diseases, *e.g.* glucocerebrosidase, acetylcholinesterase and amyloid.

(ii) Proteins and protein complexes related to mechanisms of malignancy, such as cell-surface proteins, nuclear receptors, transcription factors, DNA-repair/replication enzymes and protein kinases and phosphatases.

(iii) Proteins associated with autoimmune diseases such as multiple sclerosis.

A sample of our target selection is shown in Fig. 4. A full description is not accessible for all of the targets, as some are confidential.

3.2. Bioinformatic tools

Each target is analyzed using a series of bioinformatics tools (see http://www.weizmann.ac.il/ISPC/biotools.html), which are implemented in our LIMS. These tools assist us in all the steps of the production and crystallization process, *e.g.* folding prediction (*FoldIndex*; Prilusky, Felder *et al.*, 2005; Fig. 5), domain analysis, physical characterization and data mining.

An online search is made *via* *SeqAlert* (see http://bioportal.weizmann.ac.il/salertb/main) to check whether the same or a similar protein exists in the PDB or is being studied in any other structural genomics center. The bioinformatic analysis, together with the literature search, helps us to design

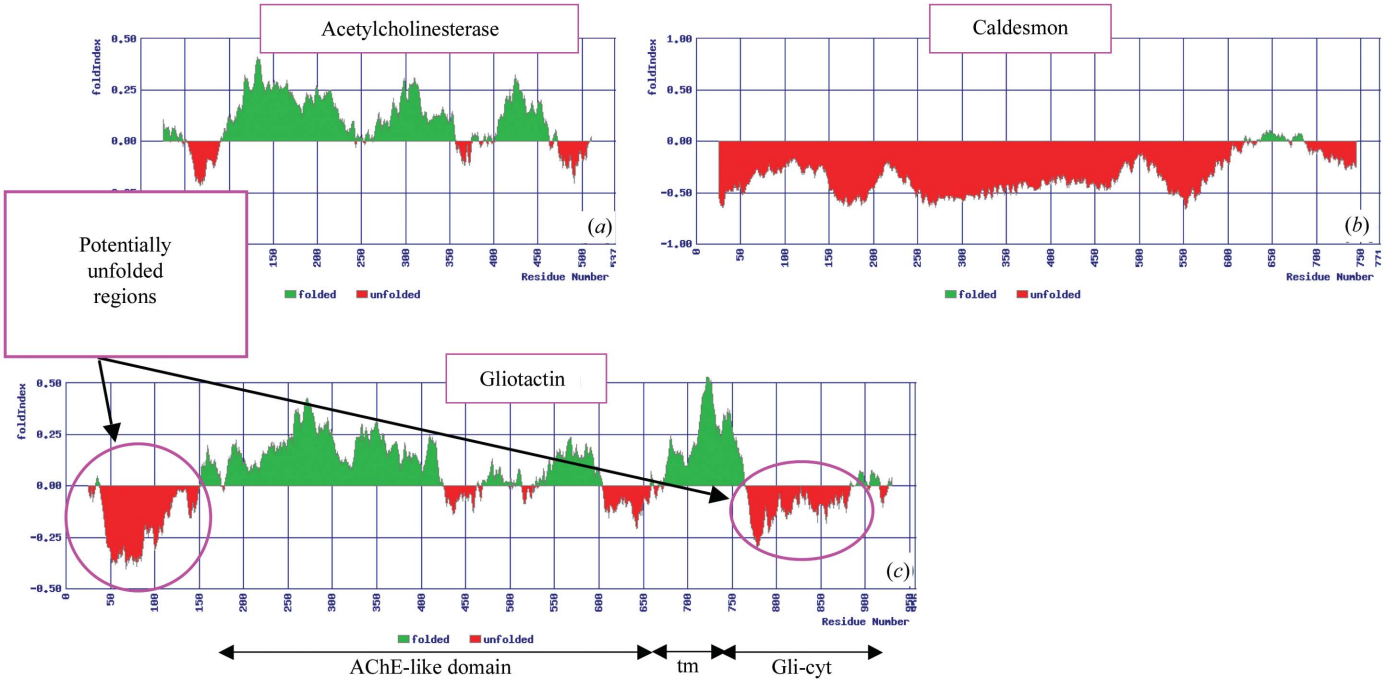

Figure 5
FoldIndex: 'will this protein fold?' (Prilusky, Felder *et al.*, 2005) (see http://bioportal.weizmann.ac.il/fldbin/findex). Green regions correspond to sequences predicted to be folded and those in red are predicted to be intrinsically unfolded. (*a*) A folded protein (acetylcholinesterase; Sussman *et al.*, 1991); (*b*) a protein experimentally shown to be intrinsically unfolded (caldesmon; Permyakov *et al.*, 2003); (*c*) an example of a protein containing both folded and unfolded domains (gliotactin; Zeev-Ben-Mordehai *et al.*, 2003).

our experimental protocol: expression system(s), whether or not mutations or deletions should be introduced and if the protein should be co-produced and/or co-crystallized with stabilizing binding partners. Owing to the growing interest in protein complexes, the ISPC has been helping to develop, together with Dr Anne Poupon (Gif-sur-Yvette), the HalX LIMS system to accommodate the various steps required for annotation of the stages involved in preparation and structure determination of proteins (Prilusky, Oueillet *et al.*, 2005).

3.3. Protein expression in *E. coli*

Following the submission of a target and its bioinformatic analysis, the ISPC utilizes HTP cloning and expression methodologies (see §2.1). A number of parameters are screened in parallel, including promoters, tags, inducers, temperature, strains and additives, in order to optimize production of soluble protein. Successful application of this approach is illustrated for two targets in Fig. 6. If soluble protein is obtained under a particular set of conditions, production is scaled up, usually to 4.2 l.

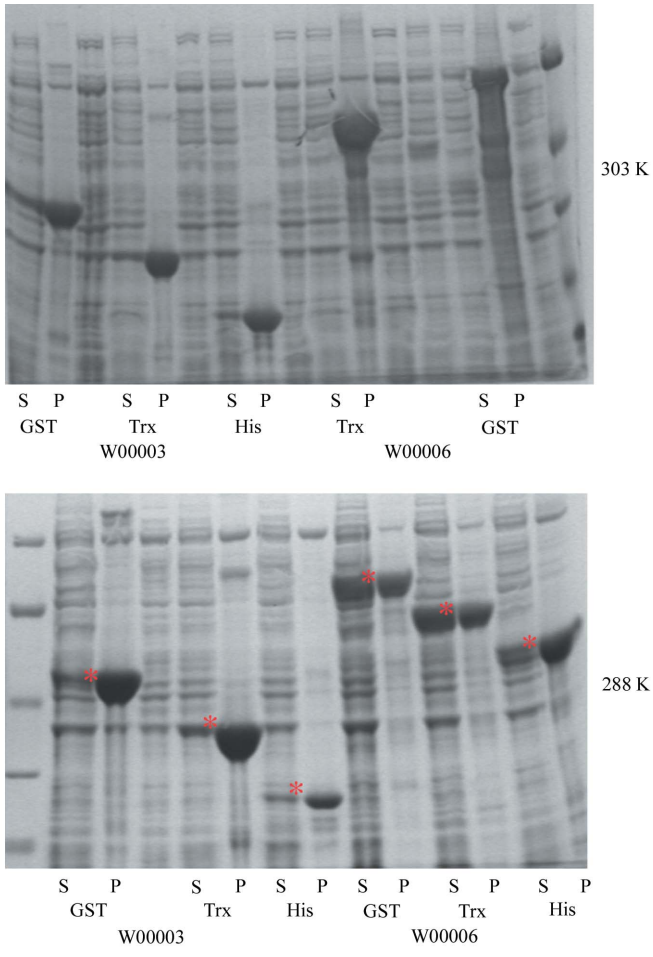

Figure 6
Temperature optimization in *E. coli* expression of two targets, W00003 and W00006, cloned in three different vectors, using His, Trx and GST tags, at two temperatures. Expression at 288 K resulted in a higher proportion of soluble protein (red asterisk). P, pellet; S, soluble.

In cases where only inclusion bodies form, refolding is employed. So far, seven proteins have been successfully refolded.

3.4. Protein expression in eukaryotic systems

In cases where no soluble or correctly folded protein can be obtained in *E. coli*, alternative expression systems are employed. *P. pastoris* or baculovirus are being used for expression of proteins for which post-translational modification is believed to be essential for obtaining a functional protein.

Expression in *P. pastoris* is directed towards production of either intracellular or secreted protein. Various parameters are being tested to optimize the yield of soluble protein, including the composition of the medium, promoters, tags, temperature and yeast strain (Fig. 7).

We have engineered Gateway expression vectors compatible with *P. pastoris* and have used them for expression of intracellular or secreted proteins. We have been able to express five eukaryotic proteins that we were unable to obtain in *E. coli*. All five were secreted into the culture medium in glycosylated form.

3.5. Protein purification and crystallization

Once a soluble protein has been obtained in one of the above systems, it is purified by affinity chromatography followed by at least two further purification steps. The tags are then removed proteolytically and the protein is analyzed to establish whether it is correctly folded. It is then prepared for the crystallization screen(s) (Fig. 8).

Screening and optimization of protein crystallization conditions are being carried out with a Douglas Instrument IMPAX 1-5 and an Oryx 6 robot, both of which employ the microbatch method under oil (see http://www.douglas.co.uk/impax.htm; Chayen *et al.*, 1990). The benefits of using this procedure include the requirement for very small volumes of both protein and reagent, the minimization of surface inter-

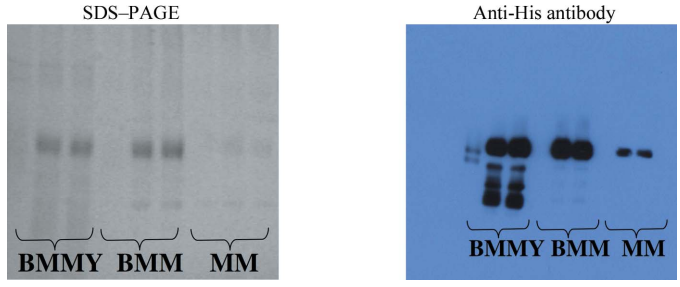

Figure 7
Optimization of the induction medium for protein expression in *P. pastoris*. Target W00075 was expressed in *P. pastoris* in three media, BMMY, BMM and MM, induction taking place with methanol at 293 K. Samples were taken at three different times points during induction, 72, 120 and 144 h (from left to right in each medium), and were analyzed by SDS–PAGE, stained with Gel Code (left) or subjected to Western analysis with anti-His_6 monoclonal antibodies (right). It can be seen that the choice of medium strongly influences both the amount and quality of the protein produced.

action with the protein and the ability to precisely control protein and reagent concentrations. Because the method is very rapid and consumes only small amounts of protein, it is suitable for HTP crystallization screening and optimization. Many target proteins have been successfully crystallized using the microbatch method and the robots employed are much less expensive than those that utilize hanging-drop or sitting-drop methods. Over 13 000 crystallization wells have been set up so far. Use of the microbatch method has resulted in a remarkably high success rate, with ∼75% of the experiments yielding crystals so far. The conventional hanging-drop and sitting-drop vapour-diffusion methods are still being used in cases where the crystallization conditions for a particular protein are known.

3.6. From gene to structure

In the eight months since production commenced (January–August 2004), 27 targets have already been handled. Each target has posed a unique challenge. It is commonly accepted that only ∼20% of the proteins expressed in *E. coli* are produced in soluble form. We have therefore applied HTP methodologies for optimization at all stages of production and crystallization. Our pilot study is summarized in Fig. 9. It is clear from Fig. 9(*a*) that our HTP screening procedure, which utilizes different expression systems and optimizes multiple parameters, has enabled us to increase the percentage of soluble proteins obtained from the commonly accepted figure of ∼20% to ∼50%. Nevertheless, from inspection of both Figs. 9(*a*) and 9(*b*), which includes the proteins entering the pipeline at the crystallization step, it is apparent that the major bottleneck in obtaining crystals is still production of soluble monodisperse protein.

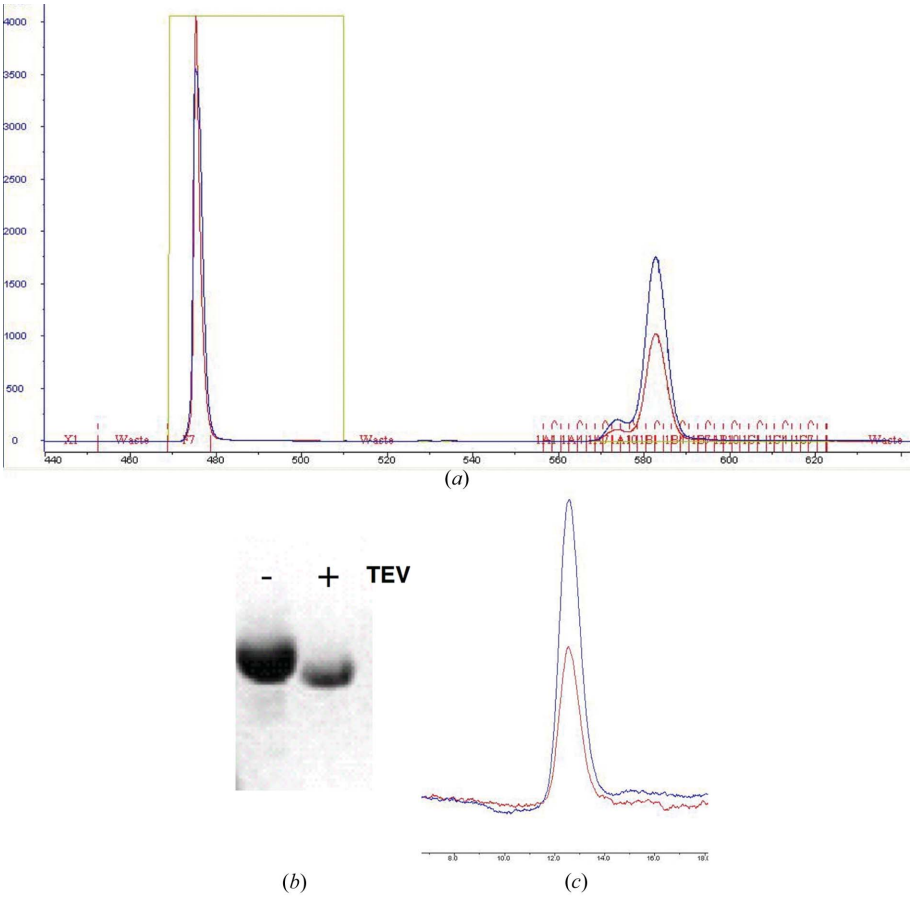

Figure 8
Protein purification. (*a*) Chromatogram of protein eluted in one step from a nickel column followed by size-exclusion chromatography; (*b*) SDS–PAGE of protein cleaved by TEV; (*c*) Analysis of final product by analytical size-exclusion chromatography.

3.7. Protein complexes

The ISPC aims to elucidate the structures of proteins related to human health in their functional context. Proteins can function either alone or complexed with one or more other proteins and/or nucleic acids. The structure of a protein in its complexed form is often different from that of the protein alone. It is therefore of interest to solve structures of protein complexes and to gain information about protein–protein interactions. This is usually achieved by expressing each soluble protein separately and then cocrystallizing the complex. However, often a protein that is active as a complex is unstable or unfolded in the absence of its partner(s). Consequently, we have adopted several additional strategies for obtaining soluble protein complexes. These include co-refolding of the denatured partners and co-refolding of a soluble protein with its denatured partner. Alternatively, we are using coexpression and purification of the protein complex. To date, we have screened crystallization conditions for four different protein complexes and solved two complex structures.

3.8. Initial fruits of the ISPC

The ISPC has stimulated the interest in structural biology of biochemists and biologists within the WIS. They now realise that it is possible to determine three-dimensional structures of proteins much more rapidly, with 14 structures being solved in the past eight months. Three important three-dimensional structures (Fig. 10) that were determined with the help of the ISPC are the following.

(i) Paraoxonase (PDB code 1v04): a multi-purpose enzyme that has been shown to perform a variety of jobs in the body, including ridding the arteries of plaque-forming clumps of LDL ('bad' cholesterol) that lead to arteriosclerosis and degrading toxic chemicals such as pesticides and nerve gases (Aharoni *et al.*, 2004; Harel *et al.*, 2004). For a recent popular report, see http://news.bbc.co.uk/2/hi/science/nature/3671827.stm.

(ii) TbADH: a mutant of alcohol dehydrogenase from the thermophilic bacterium, *Thermoanaerobium brockii*, involved in reduction of ketones and secondary aldehydes. It is being studied to examine structure, function and thermal stability of enzymes from thermophilic and mesophilic microorganisms (Levin *et al.*, 2004).

(iii) Glucocerebrosidase (PDB code 1ogs): mutations occurring in this enzyme cause Gaucher disease, an often severe lipid-storage disease that mainly affects Ashkenazi Jews. The solution of the structure of this enzyme may result in new therapies for the disease (Dvir *et al.*, 2003).

Furthermore, the ISPC has already had an impact on the Israeli biotechnology industry. Two small/medium-size enterprises are currently working with the ISPC in the development of new drugs *via* X-ray crystallographic determination of the structures of complexes of putative lead molecules with their protein targets. The ISPC has also benefitted enormously from being part of SPINE, since a substantial number of our scientists and students have been able to participate in workshops and to work for short periods in other SPINE laboratories, *e.g.* at Oxford, Hinxton, Berlin, Marseille, Munich, Gif-sur-Yvette, Strasbourg, Grenoble, Hamburg, Uppsala, Barcelona, Amsterdam and York. Being part of SPINE has allowed us to make contact with people at the bench, to have informal discussions by e-mail and telephone about technical problems and to share experiences and protocols. We have also been able to obtain modified expression vectors from our SPINE colleagues. Furthermore, we have been able to take advantage of these interactions to make informed decisions as to, for example, which expression systems to develop and which robots to purchase.

Some of the ideas developed at the ISPC, in particular in the area of bioinformatics, are now implemented on a web-based server (Fig. 11). In addition, in close collaboration with

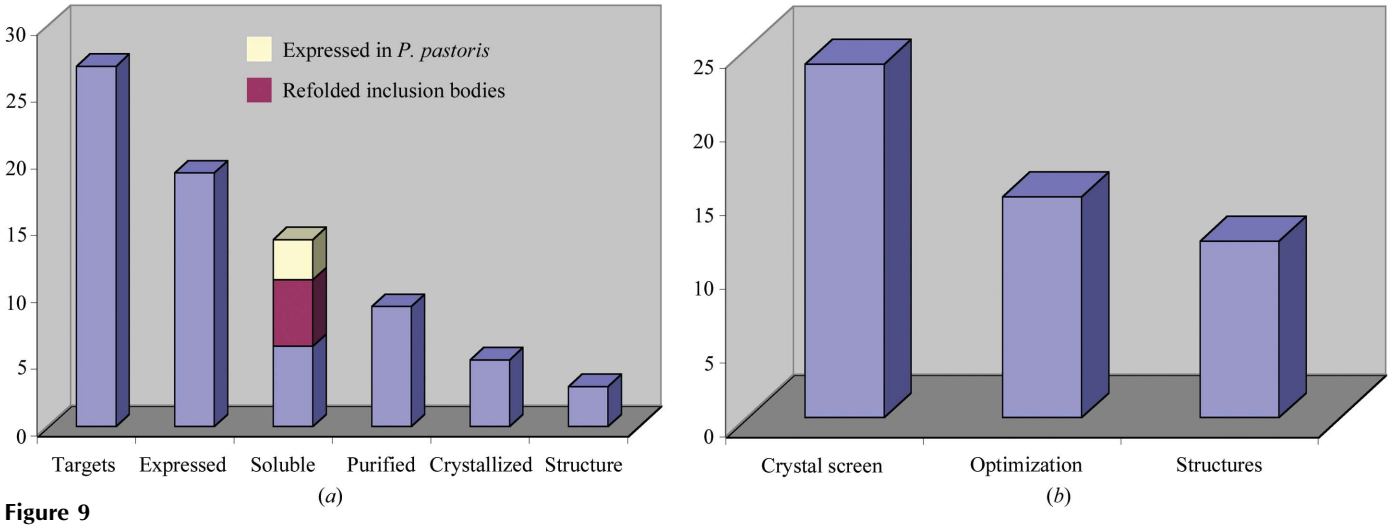

Figure 9
Fall-off in success through various stages of protein production, crystallization and structure determination. (*a*) Entry point at the cloning stage; (*b*) entry point at the crystallization stage.

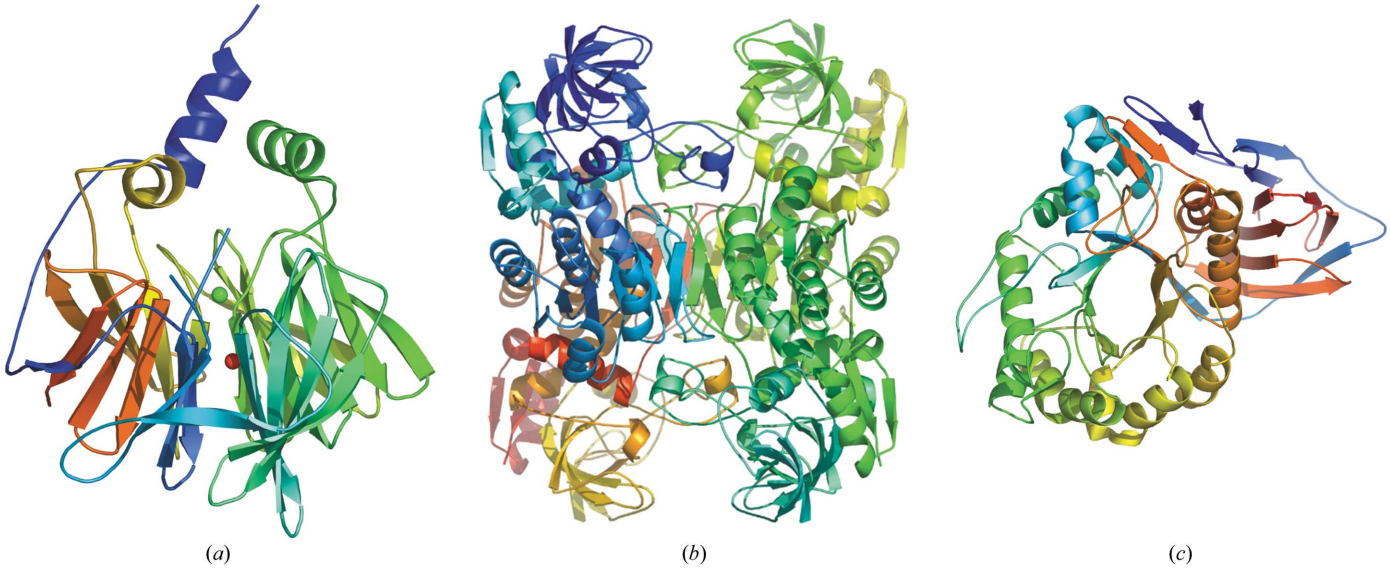

Figure 10
Three major structures determined with the aid of the ISPC. (*a*) Mammalian paraoxonase; (*b*) *T. brockii* alcohol dehydrogenase; (*c*) human glucocerebrosidase.

Bioinformatics Tools Developed at the Weizmann that have been adopted by ISPC

FoldIndex© – http://bip.weizmann.ac.il/fldbin/findex – tries to answer the question 'will this protein fold'?
SeqAlert© – http://bip.weizmann.ac.il/salertb/main – is a sequence-alerting service that will periodically compare your sequence(s) to sequences that may have had their three-dimensional structures determined, or those that have been submitted to the PDB and not yet been released, as well to those listed in TargetDB, the database of target sequences from worldwide structural genomics projects.
OCA© – http://bip.weizmann.ac.il/oca-bin/ocamain – is a structure–function browser which, in addition to searching and browsing PDB data, integrates data from other resources including GPCRDB, KEGG, OMIM, UNIPROT *etc*., so as to facilitate the understanding of the genomics/proteomics of biological data.
SeqFacts© – http://bip.weizmann.ac.il/sqfbin/seqfacts – is a tool for sequence identification, analysis, characterization and annotation. This server will try to find relevant information related to your sequence.
RecentReferences© – http://bip.weizmann.ac.il/sqfbin/recentReferences – will try to find recent publications related to your sequence.
BestPrimers© – http://bip.weizmann.ac.il/sqfbin/bestPrimers – design of primers.
SuggestES - http://bioportal.weizmann.ac.il/expsysb/suggestES - is a tool that takes a protein sequence you provide, and scans a large database of protein sequences with known results for different expression systems, in order to predict the best expression system to use for the particular sequence input.

Figure 11
Bioinformatics tools developed by the ISPC.

Anne Poupon (Gif-sur-Yvette) and Jaime Prilusky (WIS) has extended the HalX Protein Production LIMS that she developed to cover the specific requirements of the ISPC (Prilusky, Oueilliet *et al.*, 2005). All such modifications and improvements are now being included as part of the official HalX release, so that other European centers will be able to benefit from them. For example, HalX is now capable of querying and retrieving information from remote servers. The first implementation of this Web Services feature was for primer design, which is now being performed over the Internet by the 'BestPrimers' server at WIS.

The research described is being supported by the European Commission Fifth Framework 'Quality of Life and Management of Living Resources' 'SPINE' Project grant No. QLG2-CT-2002-00988, the Israel Ministry of Science and Technology Grant for the ISPC, the Divadol Foundation and a Minerva Grant. JLS is the Morton and Gladys Pickman Professor of Structural Biology.

References

Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I. & Tawfik, D. S. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 482–487.
Chance, M. R., Bresnick, A. R., Burley, S. K., Jiang, J.-S., Lima, C. D., Sali, A., Almo, S. C., Bonanno, J. B., Buglino, J. A., Boulton, S., Chen, H., Eswar, N., He, G., Huang, R., Ilyin, V., McMahan, L., Pieper, U., Ray, S., Vidal, M. & Wang, L. K. (2002). *Protein Sci.* **11**, 723–738.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Dvir, H., Harel, M., McCarthy, A. H., Toker, L., Silman, I., Futerman, A. H. & Sussman, J. L. (2003). *EMBO Rep.* **4**, 704–709.
Dyson, H. J. & Wright, P. E. (2005). *Nature Rev. Mol. Cell. Biol.* **6**, 197–208.
Harel, M., Aharoni, A., Gaidukov, L., Brumshtein, B., Khersonsky, O., Meged, R., Dvir, H., Ravelli, R. B., McCarthy, A., Toker, L., Silman, I., Sussman, J. L. & Tawfik, D. S. (2004). *Nature Struct. Mol. Biol.* **11**, 412–419.
Heinemann, U., Frevert, J., Hofmann, K.-P., Illing, G., Maurer, C., Oschkinat, H. & Saenger, W. (2000). *Prog. Biophys. Mol. Biol.* **73**, 347–362.
Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. & Yokoyama, S. (1999). *FEBS Lett.* **442**, 15–19.
Leulliot, N., Tresaugues, L., Bremang, M., Sorel, I., Ulryck, N., Graille, M., Aboulfath, I., Poupon, A., Liger, D., Quevillon-Cheruel, S., Janin, J. & van Tilbeurgh, H. (2005). *Acta Cryst.* D**61**, 664–670.
Levin, I., Meiri, G., Peretz, M., Burstein, Y. & Frolow, F. (2004). *Protein Sci.* **13**, 1547–1556.
Permyakov, S. E., Millett, I. S., Doniach, S., Permyakov, E. A. & Uversky, V. N. (2003). *Proteins*, **53**, 855–862.
Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). *Bioinformatics*, **21**, 3435–3438.
Prilusky, J., Oueillet, E., Ulryck, N., Pajon, A., Bernauer, J., Krimm, I., Quevillon-Cheruel, S., Leulliot, N., Graille, M., Liger, D., Trésaugues, L., Sussman, J. L., Janin, J., van Tilbeurgh, H. & Poupon, A. (2005). *Acta Cryst.* D**61**, 671–678.
Service, R. F. (2000). *Science*, **289**, 2254–2255.
Shah, A. K., Liu, Z. J., Stewart, P. D., Schubot, F. D., Rose, J. P., Newton, M. G. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 123–129.
Stevens, R. C., Yokoyama, S. & Wilson, I. A. (2001). *Science*, **294**, 89–92.
Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L. & Silman, I. (1991). *Science*, **253**, 872–879.
Todd, A. E., Marsden, R. L., Thornton, J. M. & Orengo, C. A. (2005). *J. Mol. Biol.* **348**, 1235–1260.
Yee, A., Pardee, K., Christendat, D., Savchenko, A., Edwards, A. M. & Arrowsmith, C. H. (2003). *Acc. Chem. Res.* **36,** 183–189.
Yokoyama, S. (2003). *Curr. Opin. Chem. Biol.* **7**, 39–43.
Zeev-Ben-Mordehai, T., Rydberg, E. H., Solomon, A., Toker, L., Botti, S., Auld, V. J., Silman, I. & Sussman, J. L. (2003). *Proteins*, **53**, 758–767.