The Status of Analytical Protein Chemistry at the Australian National University

June 2003

A report commissioned by the Management Committee of the Australian National University Biomolecular Resource Facility.

Prepared by

Dr P.J. Milburn
Technical Specialist
ANU Biomolecular Resource Facility
peter.milburn@anu.edu.au

Contents

Introduction

Integrative Biology

Global Trends in Analytical Protein Chemistry

The Techniques of Analytical Protein Chemistry

Amino acid composition analysis

Atomic absorbance spectroscopy

Chemical polypeptide determination

N-terminal sequence determination.

C-terminal sequence determination.

Protein charcterisation by mass spectrometry

Peptide mass fingerprinting

Peptide sequence determination by tandem mass spectrometry

Comparison of the utility of the different Techniques of Analytical Protein Chemistry

Amino acid composition analysis

Atomic absorbance spectroscopy

Chemical polypeptide determination

Protein characterisation by mass spectrometry

Analytical Protein Chemistry at the Australian National University

History

Current status of Analytical Protein Chemistry at the ANU

Comparison of resources supporting integrative biology at the ANU

Recommendations for the future of Analytical Protein Chemistry at the ANU

References cited

Introduction

In the context of this report, Analytical Protein Chemistry encompasses those methodologies that enable the chemical nature of discrete gene products to be determined. The type of information that is gathered for this purpose includes amino acid composition, primary sequence, peptide-mass fingerprints, accurate native molecular mass and the site and nature of post-translational modifications and bound metals.

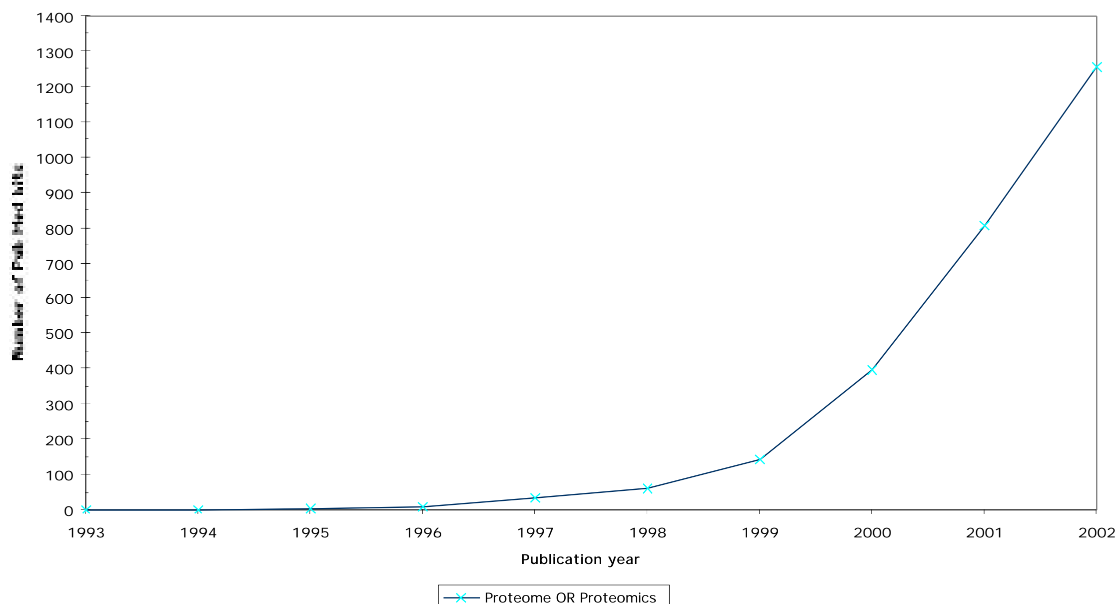
Integrative Biology

The integrative approach to understanding living systems seeks to elucidate every aspect of a given biological state at the most fundamental level. This holistic treatment has the capacity to deconvolute the highly complex relationships between organisms and their environment and provide descriptions of biological processes at a molecular level.

The first Human Gene Mapping Workshop in 1973 heralded the dawn of Integrative Biology. Nucleic acid molecular biology and genetics were married when genetic linkage between thalassaemia and the loss of globin genes was established by the use of sequence specific molecular probes (Kan *et al.*, 1975). The utility of polymorphic sequence markers in linkage analysis was firmly established by 1979 and in 1980 Botstein *et al.* published "Construction of a genetic linkage map in man using restriction fragment polymorphism". This was a milestone in the development of both genetics and human biology. By 1985, the outrageous prospect of sequencing the entire human genome (estimated at some 3×10^8 bases) had been mooted. This task seemed impossible at the time since it took a skilled scientist a whole day to sequence 50-100 bases at a cost of some 10$ *per* base! It is worth noting how rapidly technology was both developed and deployed such that by 1989 "Mapping and Sequencing the Human Genome" (McKusick 1989) was a fully developed concept. This era spawned the discipline of genomics, which may be described as the quantitative study of genes, regulatory and non-coding DNA sequences.

During the previous decade it has become apparent that the availability of complete consensus genome sequences for a large set of diverse organisms would transform biological investigations irrevocably. It has been recognised that a major challenge of this post-genomic era is the elucidation of the expression, structure and function of all the proteins encoded in the genome and this is being addressed by a novel discipline termed proteomics. In comparison to genomics, proteomics represents a much greater challenge because the former is a relatively static entity, while the latter is dynamic in character and changes both with the development of an individual and in response to environmental factors. A crude bibliometric survey of scientific papers pertinent to the study of the proteome was conducted over the last decade by searching the PubMed database for proteome OR proteomics and the results are presented in Figure 1.

Figure 1. The number of hits obtained *per annum* searching PubMed for proteome OR proteomics.



It is clear that investigations aimed at defining systems at the level of the proteome are now an important part of biology and one that is undergoing rapid expansion globally.

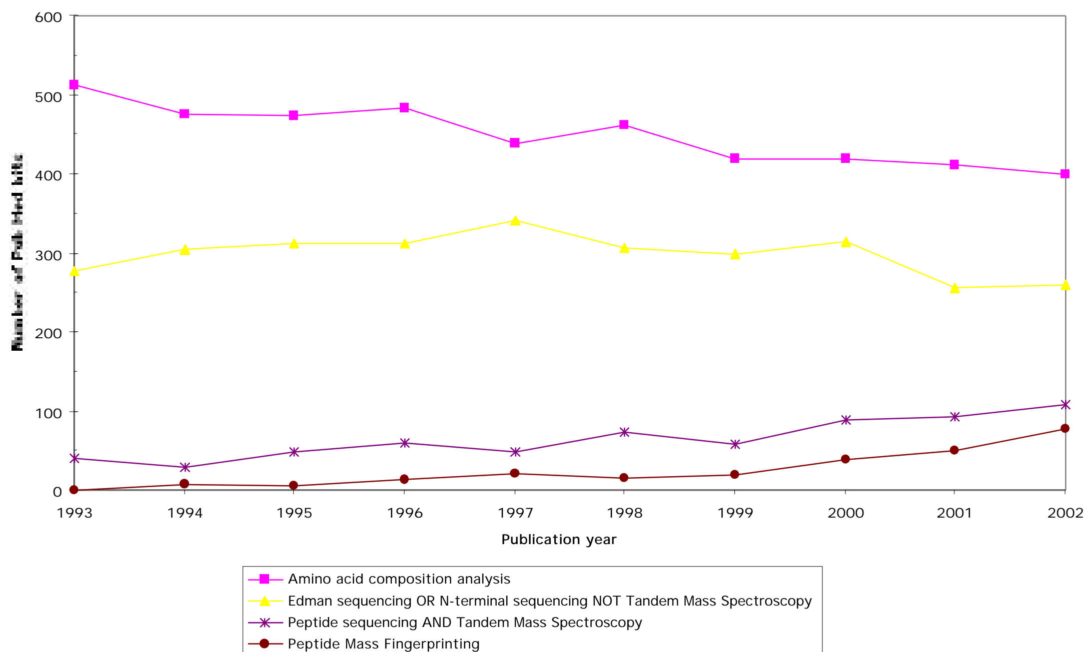
The integrative approach to the study of biology has continued to develop and non-parametric investigations of phenotype, gene expression and metabolic state are now important disciplines.   Just as the product of genotype and circumstance is phenotype, the product of the genome and circumstance is the phenome. Similarly, the quantitative study of gene expression as mRNA is termed transcriptomics and metabolites and metabolic networks is termed metabolomics.

The integrative approach to protein chemistry includes some important sub-disciplines. The function and properties of proteins may be modulated significantly by post-translational chemical modifications. Consequently, phosphoproteomics and glycomics (the quantitative study of the expression, structure and function of all the phosphoproteins and glycoproteins, respectively, encoded in the genome) are emerging fields.

<u>Global trends in Analytical Protein Chemistry</u>

Analytical protein chemistry has been an important resource to biochemists and biologists for a great many years but never more so than in the era of proteomics. While the more traditional applications remain important, the constraints of studying a large number of gene products simultaneously have forced the development of high throughput strategies. A crude bibliometric survey of scientific papers relating to analytical protein chemistry was conducted over the last decade by searching the PubMed database for the appropriate terms and the results are presented in Figure 2.

Figure 2. The number of hits obtained *per annum* searching PubMed for terms relating to methods of analytical protein chemistry.



It is clear that both amino acid composition analysis and chemical N-terminal protein sequence determination are mature techniques that support a large number of scientific publications each year. It is possible that there is a slight trend toward decline in the data but it must be remembered that, because these are established techniques, reference to their use is often buried in these reports. Random checking has shown that the PubMed search failed to find numerous papers that reported data obtained by these means but the trends are probably reliable. Industry sources confirm that these are indeed mature markets; while instrument purchases are infrequent reagent sales are steady.

The number of scientific papers relating to the use of MS in the analysis of proteins is growing on a yearly basis. Interestingly, fewer papers employing MS were found in the search than those employing either of the more established techniques. However, the trend in the number of papers employing MS in the analysis of proteins is clearly upward.

The Techniques of Analytical Protein Chemistry

Amino acid composition analysis

Amino acid composition analysis is a classical method of protein analysis that finds continued and extensive application in medical and food science. It is indispensable for both protein and peptide quantification. It is a complex technique that combines protein hydrolysis with subsequent chromatographic separation and detection of the mixture of constituent amino acids. The process is both time-consuming and technically demanding if rigorous precision and accuracy are to be attained with very high sensitivity. A number of significant developments have occurred during the last decade (for review see Fountlakis & Lahm, 1998).

Microwave energy sources have been adopted in place of traditional hot ovens, which has reduced hydrolysis times from hours to minutes.

Gas-phase hydrolysis methods have been developed that preserve chemically labile amino acids, which enables their quantitative determination.

Reversed-phase HPLC has been adopted over ion-exchange methods, which has led to more robust and rapid separation of amino acid derivatives.

A variety of chemical procedures have been developed that yield an amino acid derivative that is fluorescent, thereby enabling detection to the sub-femtomole range.

Amino acid composition data has been included in algorithms for large-scale protein identification by database searches.

This last development has been a major factor in the continuing utility of this technique as a means of identifying proteins (Wilkins *et al.*, 1999). Accurate amino acid composition, molecular weight and isoelectric point data combine to form a powerful means of identifying a protein by database searches. This information may be generated readily by means of 2D-gel analysis with subsequent hydrolysis of the PVDF blots (Wilkins *et al.*, 1996).

As a technique for high-throughput protein identification, amino acid composition analysis is facing enormous competition from mass spectrometry techniques because it is perceived as being relatively slow and difficult to perform accurately.

Potential developments include the use of capillary zone electrophoresis (CZE) to accelerate separation of amino acid derivatives and tandem mass spectrometry (MS-MS) both to enhance sensitivity and to accelerate separation.

<u>Atomic absorbance spectroscopy</u>

It has been known since classical times that the introduction of certain metals or their salts, *e.g. cuprum*, to a heat source causes them to "burn" with a characteristic pure coloured flame. Atomic absorbance spectroscopy is a classical technique of analytical chemistry with application to any field where the quantitation of elements is required, *e.g.* astronomy, soil science, physiology *etc*. Protein chemists are interested in the specific binding of metals to proteins because there are many established instances where this is essential for maintenance of structural motifs, biological activity or, indeed, both.

Inductively-coupled plasma furnaces have greatly increased the range of elements detected by raising the temperature of the source, thereby generating emissions at shorter wavelengths.

For metals such as copper nanomole amounts are detected accurately.

Chemical polypeptide sequence determination

Chemical polypeptide sequence determination is a classical method of protein chemistry that relies on chemical cycles to remove amino acids singly and successively from the terminus of a polypeptide chain. Like amino acid composition analysis it is a complex technique, which, in this case, combines terminal amino acid cleavage with subsequent chromatographic detection and identification relative to the mixture of amino acid standards.

N-terminal sequence determination.

The chemistry introduced by Edman (1950) for N-terminal sequence determination remains the standard for today, its virtues including exceptionally high repetitive yields and both excellent resolvability (Zimmerman *et al.*, 1977) and stability (Tarr 1977) of the amino acid derivatives produced. The introduction of automated protein sequenators (Edman & Begg, 1967) powered a revolution in Molecular Biology, in which the complete sequences of a range of proteins were determined. For the first time, protein chemists were able to compare a library of primary structure with a library of three-dimensional structure.

A number of significant developments have occurred during the last twenty years or so.

The Edman degradation is performed with delivery of the acid and base in the gas phase and the sample adsorbed either on a glass-fibre filter or poly(vinylidene difluoride) membrane, which has been pre-treated with a cationic polymer. These procedural modifications have greatly improved the retention of sample in the solid phase generating greater read lengths and, more importantly, higher sensitivity.

Introduction of low dead-volume sequentators and micro HPLC systems has provided sensitivity enhancements such that less than 1 picomole of sample is required to sequence up to 20 or more residues.

Use of improved solvents and organic bases have seen a reduction in chromatographic interference and an increase in yields of sensitive amino acid residues.

Potential developments include the use of pro-fluorescent Edman reagents (Okiyama *et al.* 2002) and laser-induced fluorescence (LIF) detection to provide order-of-magnitude sensitivity enhancements, CZE to accelerate separation of cleaved amino acid derivatives and MS-MS both to enhance sensitivity and to accelerate separation.

C-terminal sequence determination.

The chemistry employed for C-terminal sequence determination was introduced originally in the 1920s (Schlack & Kumpf, 1926). The *thiocyanate method* for C-terminal sequencing, thus, predated the *isothiocyanate method* for N-terminal sequencing by some 30 years. The *thiocyanate method* is prone to many side reactions and the sequencing yields are low compared to Edman's method. For these reasons C-terminal sequencing has been largely ignored until the 1990s when it became clear that C-terminal analysis of recombinant proteins was important for quality control. Post-translational truncation at the C-terminus is commonly encountered and is difficult to quantify by any alternative technique.

In a recent publication, Bergman *et al.* (2001) described the use of the commercially available system optimised to obtain 4-5 residues of sequence from as little as 10 picomole of protein **after** the sample had been subjected previously to N-terminal sequencing.

As a technique for high-throughput protein identification, chemical sequence determination suffers because it is relatively slow, identifying only one or two amino acids *per* hour using conventional configurations. Its principal utility resides in the unambiguous sequence analysis of protein termini. Being an equilibrium technique it provides reliable information from heterogeneous samples and is thus invaluable as a quality control analysis of protein products (Bergman *et al.*, 2001).

<u>Protein characterisation by mass spectrometry</u>

Peptide-mass fingerprinting.

The digestion of a given protein by a hydrolytic enzyme such as trypsin, which is specific for cleavage of the peptide bond C-terminal to a positively-charged amino acid side chain, yields a mixture of products of molecular mass that may be predicted from the primary sequence. If the masses of a sufficient number of these peptides are determined with appropriate accuracy, then the originating protein may be identified unequivocally by searching a sequence database. This type of data set is usually obtained with a Matrix-Assisted LASER Desorption-Ionisation (MALDI) Time-of-Flight (TOF) Mass Spectrometer (MS) because of the inherent high sensitivity and ability to tolerate mixtures that are characteristic of this instrument. Atmospheric Pressure Ionspray (API) Ion-Trap (IT) Mass Spectrometers are also widely employed to obtain peptide-mass fingerprints. This latter type of instrument is somewhat less sensitive and requires higher maintenance than the MALDI-TOF MS but is better suited to obtaining sequence information from difficult samples.

Peptide sequencing determination by tandem mass spectrometry.

The peptide-mass fingerprint fails to identify a protein unequivocally if few there is a paucity of peptide masses that are obtained with good mass accuracy, the protein has extensive post-translational modification or if it belongs to a family of closely related proteins that generate very similar fingerprints. Under these circumstances obtaining Peptide Sequence Tags (PSTs), elements of a peptide's sequence that are not necessarily continuous, may clinch identification. By exploiting the spontaneous process of Post-Source Decay (PSD) in the MALDI-TOF MS a rudimentary MS-MS experiment allows one to determine PSTs. However, one has little control over the events occurring in PSD and perhaps only 30% of peptides yield useful PSD PSTs. However, in most cases there are several peptides from which one may attempt to observe the formation of ion progeny.

Assuming sufficient sample availability, proteins that resist a first round of analysis by MALDI-TOF may be re-analysed using a platform better suited to MS-MS techniques such as an API-IT MS. Additionally, the peptide-mass fingerprint technique fails by definition if the protein does not exist in a searchable database. Obtaining useful lengths of continuous *de novo* peptide sequence by mass spectrometry that are suitable for genetic cloning requires a triple sector instrument that enables one to perform high-energy Collisionally-Induced-Dissociation (CID) MS-MS.

As a technique for high-throughput protein identification, MS has great potential because of the relatively high speed of data acquisition and the possibilities for the intelligent automation of data collection and analysis.

Comparison of the utility of the different Techniques of Analytical Protein Chemistry

Amino acid composition analysis

Quantitative amino acid composition analysis remains the method of choice for the determination of the absolute concentration of pure peptides and proteins. However, in most cases, once an accurate extinction coefficient has been determined using this technique concentration may be estimated reliably by measuring optical density relative to a standard solution.

The amino acid composition of a discrete gene product is a primary descriptor of its chemical nature but is of little use in a biological context. The primary biological descriptor of a protein is the amino acid sequence and, of course, the amino acid composition may be calculated from knowledge of this. For this reason, while amino acid composition analysis remains a useful technique it is no longer a principal tool for the identification of proteins or their derived fragments.

Atomic absorbance spectroscopy

Atomic absorbance spectroscopy is the method of choice for the determination of both the identity of protein-bound metal atoms and the associated stoichiometry of binding.

Chemical polypeptide sequence determination

In the pre-genomic era, chemical polypeptide sequence determination was a major force in molecular biology and the complete sequence of hundreds of proteins was determined in this way. The development of DNA sequencing and its ancillary technologies proceeded in such a way that it became far more efficient to clone and sequence a cDNA copy of the RNA message that encoded a protein rather than to determine complete sequences at the protein level. Partial sequence of an unknown protein was obtained by chemical polypeptide sequencing and this information was used to design a probe that could be used to identify components of an expression library that might encode the protein. Progress in the genome projects rapidly destroyed the chances of discovering a novel gene product and by the mid 1990s a homologue for virtually every protein sequence obtained was to be found in a database. "The Sequence of the Human Genome" was published in 2001 (Venter *et al.*, 2001) marking a watershed in human biology. At the end of June 2003 there were 236 bacterial, 18 archaeal and 44 eukaryotic genomes available for bioinformatic searches, although many of these are not yet completed. Apart from human, animal genome sequence data from mouse, rat, zebra fish and fruit fly are now available, which are important experimental species. The available plant genome sequence data are biased heavily towards commercial species and include rice, soybean wheat, corn and tomato.

High-throughput protein identification is now achieved by utilisation of MS methods that allow protein characterisation by searching databases with unique numerical parameters, which has largely replaced techniques that determine the actual chemical code of protein sequence. The large range of genomic databases currently available allows researchers to study many different proteomes using MS techniques.

Chemical polypeptide sequence determination remains an important tool for proteomic analysis of organisms for which genome data is not available. It also retains utility in proteomics by virtue of its ability to delimit the open reading frame (Wasinger & Humphery-Smith 1998).

In more specific applications, chemical microsequencing remains a principal tool for the characterisation of both N- and C-terminal post translational modifications and quality assessment of recombinant expression products (Bergman *et al.*, 2001). Chemical microsequencing is also a powerful tool for mapping sites of post-translational modification unambiguously (Wang *et al.*, 1999). Chemical methods are equilibrium techniques so are useful for investigating genetic variation *e.g.* in the determination of relative allelic expression, especially in polyploid organisms.

Apart from inadequate amount or preparation of sample, the primary cause of failure in chemical microsequencing is chemical modification that renders the polypeptide terminus refectory to the degradation chemistry.

Protein characterisation by mass spectrometry

The various techniques of protein MS are rapidly gaining credibility as benchmark techniques in analytical protein chemistry. Much of the developmental progress has been driven by the need for high-throughput protein identification required by proteomic analysis.

Determination of an accurate native molecular mass for a specific protein by MS allows one to screen for splice variants, alternate initiations and terminations and, also, post-translational modifications. However, characterisation of post-translational modifications requires that the protein be fragmented into peptides of a size suitable for further analysis.

Peptide-mass fingerprinting and both the determination of sequence tags or *de novo* protein sequence all require suitably sized protein fragments. Generating these fragments from very small amounts of protein becomes the bottleneck of analytical protein chemistry by MS. Dedicated equipment and skilled personnel become a requirement for efficient sample analysis.

Apart from inadequate amount or preparation of sample, the primary cause of failure in MS analysis of protein fragments is failure to generate soluble peptides in sufficient yields. Identification of gene products by a combination of peptide-mass fingerprinting and PSTs fails if the number of fragments covered is small, if the data quality is low due to limiting amounts of peptide and, of course, if the genome database is sparse. The characterisation of genetic variation may also be problematic using MS techniques because mutations cause mass shifts relative to reference sequences and heterozygosity may be either difficult to characterise or completely masked.

Analytical Protein Chemistry at the Australian National University

History

The ANU Biomolecular Resource Facility (BRF) was instigated in 1987 and subsumed the Amino Acid Composition Analysis and N-terminal (Edman) Protein Sequence Analysis services offered informally by the Physical Biochemistry Group at the John Curtin School of Medical Research (JCSMR). By the mid

1990s, the BRF operated two N-terminal and one C-terminal chemical protein sequencer, the purchase of which had been funded in part by the Resource & Infrastructure, Equipment and Facilities (RIEF) program of the Australian Research Council (ARC). The sequencing service was supported by High-sensitivity Amino Acid Composition Analysis. The analytical capability of these services were rated very highly, being rated amongst the top several laboratories in the world by annual surveys of the Association of Biomolecular Resource Facilities (ABRF). At the time the ANU Facility was superior to the Australian Proteome Analysis Facility (APAF) and many samples from around Australia were processed at the ANU. Subsequently, other similar facilities in Australia were successively upgraded to the point where researchers were able to utilise their local resources.

A formal service in amino acid composition analysis was withdrawn from the BRF at the end of 1999 due to a lack of staff resources and the very low demand from within the ANU. The equipment was maintained for use by researchers until mid 2002, when it was decommissioned.

Contrary to the global trend (Figure 2), the demand for chemical protein sequencing by researchers at the ANU has collapsed in recent years. This has resulted in the withdrawal of both the N-terminal and C-terminal Protein Sequencing Services offered by the ANU Biomolecular Resource Facility (BRF) at the end of 2002.

In spite of the global upsurge in proteomics research (figure 1) there has been limited activity in this area at the ANU to date. The Genomic Interactions Group (GIG) at the Research School of Biological Sciences (RSBS) have been a notable exception to this generalisation. In the past, this group has utilised the N-terminal protein sequencing facility at the ANU BRF extensively but, more recently, have been using principally MS approaches at APAF.

The emerging techniques in mass spectrometry (Figure 2) have not been embraced widely by biological researchers at the ANU. This has resulted from the limited biological applicability of the instrumentation previously available at the ANU and, to a great extent, the lack of accessibility that formal service support would provide. The Research School of Chemistry has provided an excellent Mass Spectrometry service for chemists but has never promoted itself for biological applications.

Current Status of Analytical Protein Chemistry at the Australian National University

In 2003, Analytical Protein Chemistry is at low ebb in the ANU. The only formal microanalytical service relevant to protein chemists at the ANU is that of metal analysis of biological samples provided through the atomic absorption spectroscopy facility in the Department of Forestry.

The ABI PROCISE C Protein Sequencer is a state-of-the-art chemical C-terminal sequencer but has been decommissioned due to a lack of demand for this technology at the ANU. Similarly, the ABI PROCISE HT Protein Sequencer and the ABI PROCISE cLC Protein Sequencer, state-of-the-art high-throughput and high-sensitivity N-terminal chemical sequencers, respectively, are due to be decommissioned. Researchers requiring these analyses may avail themselves of the services offered by APAF, for example.

Recently, initiatives from the GIG at RSBS have led to the acquisition of state-of-the-art MS equipment that will enable the pursuit of high quality biological MS on campus.

The Bruker Omniflex is a MALDI-TOF instrument that supports automated high-throughput sample analysis suitable for identifying proteins by peptide-mass fingerprinting. Additionally, PSD MS-MS experiments are supported allowing PSTs to be obtained when peptide-mass fingerprint searches provide ambiguous results. However, this instrument has yet to be commissioned fully and no formal analytical service is provided. From previous experience, the latter is likely to form a considerable barrier to researchers that are not familiar with MS of proteins and peptides.

The Thermo Finnigan ProteomeX LCQ Deca XP LC-API-IT MS represents a new generation ion of liquid-chromatograph coupled ion trap mass spectrometers. The development of an orthogonal gas swept API source and improved transfer ion optics have combined in the Deca XP plus to produce sensitivity to rival that of a MALDI-TOF system. The ProteomeX system includes a Surveyor Capillary HPLC system that allows 2-dimensional HPLC resolution of mixtures upstream to the Deca XP MS. The system is an ideal platform for 'Shotgun Proteomics' utilising Multidimensional Protein Identification Technology (MudPIT) (Wu & MacCoss 2002). 'Shotgun Proteomics' refers to the direct digestion and analysis of complex protein mixtures to generate a profile of the constituent protein complement. This instrument has yet to be commissioned.

Comparison of resources supporting Integrative Biology at the ANU

It is clear that the ANU must maintain resources to support research in the various realms of Integrative Biology if it is to remain competitive in biological sciences at a national level, let alone in an international context. This requires investment not only in capital equipment but also in personnel. Experienced personnel are required not only to operate equipment but also to provide an environment that supports experimental design, preparation of grant applications and training for students.

Traditionally, biology has been a discipline that has not required high technology on a grand scale but these days are clearly gone. The existing resources in integrative biology at the ANU have been created from individual initiatives rather than from a coordinated program of resource allocation. These resources are detailed in Table 1.

Genomic biology at the ANU has been provided with a high level of resources by the ANU BRF. This was defined as a core activity at its inception and has always been supported by at least one full-time technical staff position. The BRF has been able to lobby effectively to obtain funding to maintain a state-of-the-art equipment inventory.

Metabolomic biology has yet to be embraced at the ANU. The newly installed Thermo Finnigan ProteomeX LCQ Deca XP LC-API-IT MS is an appropriate MS platform with which to investigate complex mixtures of metabolites. One anticipates that research in this area will develop rapidly once the power of the technology is recognised within the University.

Proteomic biology has received surprisingly little support at the ANU given its prominence as an emerging discipline (Figure 1). It seems likely that this has been the result of scepticism on the part of the ANU research community. On the whole, the general importance of proteomic biology was not communicated well by its early proponents. Too much emphasis was placed upon trying to display vast numbers of proteins in 2D gels with hugely expensive robotic systems cutting out and digesting gel pieces, with proteins being identified at a mind boggling rate by a squadron of mass spectrometers. The BRF provided an excellent service in chemical polypeptide sequencing in the 1990s (discussed above) but the demand for its traditional applications has fallen and the relevance

of this technology to proteomic investigations has decreased. At this juncture, it is no longer viable to provide this service at the ANU. Given the current prominence of analytical protein chemistry internationally, it remains a puzzle as to why the demand for chemical sequencing has not been replaced by a demand for biological mass spectrometry. It is clear that this lack of general support for protein analytical chemistry has translated into the current lack of infrastructure. Researchers at the ANU are increasingly aware of the power of proteomic approaches to specific hypothesis-driven investigations in addition to the large-scale non-parametric applications. There is growing frustration at the lack of infrastructure support in this very important area.

Transcriptomic biology at the ANU receives support from both the ANU BRF and the RSBS SPMF. As a result of a strategic initiative of the JCSMR, the BRF provides a comprehensive service based around the Affymetrix GeneChip Technology (www.affymetrix.com). This system utilises pre-manufactured high-density arrays. Complementing this resource, Affymetrix spotted array technology, which allows custom preparation of low-density arrays, is available to researchers at the RSBS SPMF.

Recommendations for the future of Analytical Protein Chemistry at the Australian National University

The great biological research institutions maintain service support for all of the techniques of analytical protein chemistry discussed herein. Clearly, where resources are realistically limited it is necessary to respond strategically to both current and projected circumstances. The ANU is in a position to provide a very effective research environment for protein chemists over the next few years if its resources are deployed appropriately.

It is recommended that the development of analytical protein chemistry at the ANU be an integral component of a strategic plan for the development of integrative biology at the ANU.

Amino acid composition analysis

While amino acid composition analysis continues to have extensive application in many areas of biology (Figure 2), the technique is clearly not one of great relevance currently to researchers at the ANU. Occasional demand for quantitative analysis in order to assay standard solutions of peptides and proteins does not warrant provision of an in-house service. Excellent extramural services are provided by both Auspep Pty and APAF that are appropriate for the current needs of the ANU. This situation is unlikely to change unless there are significant strategic changes in ANU research programmes.

It is recommended that the limited demand for amino acid composition analysis arising at the ANU be met by external facilities.

Atomic absorbance spectroscopy

The interaction of proteins with metals has important implications for both their structure and biochemistry. Given the recent discovery of a huge number of genes for which the gene product is essentially undescribed one anticipates that atomic absorbance spectroscopy will play an important role in the verification of metal binding properties predicted by bioinformatic methods. An excellent facility exists within the ANU but, currently, has a low profile among protein chemists.

It is recommended that the relationship between the wider community of biological researchers at the ANU and the atomic absorbance spectroscopy facility in the department of Forestry be formalised. Ideally, this facility should engage with the network of biological resource facilities at the ANU.

Chemical polypeptide sequence determination

The techniques of chemical polypeptide sequence determination continue to have extensive application in biological sciences (Figure 2). Indeed, it is anticipated that the demand may even increase, driven by the need for product verification in the biotechnology sector. In particular, the technical advances in C-terminal sequencing that have occurred in the last several years should lead to a significant surge of interest in this technique. In the academic research sector chemical polypeptide sequence determination retains importance for the precise characterisation of recombinant products destined for use in applications such as crystallography, spectroscopy, enzymology and pharmacology. Given that the immediate future of protein chemistry is likely to be a period of great expansion globally it is quite possible that the needs of the ANU research community will follow suit. Any large-scale activity requiring high-throughput protein expression demands in-house product verification and allocation of funds for this research will depend upon its presence.

It is recommended that the current demand for chemical polypeptide sequence determination within the ANU, which is at an all time low, be met by external service facilities. However, it is recommended that the capability to provide a service in both N-terminal and C-terminal chemical sequencing be maintained at the ANU until a clear strategic plan for the future of protein chemistry and proteomics at the ANU is established.

Protein characterisation by mass spectrometry

The importance of the various techniques of mass spectrometry in analytical protein chemistry on a global scale increases annually (Figure 2). The application of this technology by biologists at the ANU has been tardy by most standards. The most obvious reason for this has been a lack of suitable equipment in an appropriate biologically oriented environment. Recent acquisitions at RSBS have introduced suitable instrumentation but in an undersupported environment. As the field of proteomics matures, methods for visualising gene-product display are diverging from simply 2D electrophoresis to include shotgun proteomics and protein microarrays. A state-of-the –art biological mass spectrometry facility is vital to the pursuit of research of this type. In addition to powerful MS resources, protein microarray technology requires high-throughput monoclonal antibody and recombinant protein expression facilities but this is beyond the scope of the present discussion.

It is recommended that the immediate and short-term future interests of analytical protein chemistry at the ANU be best served by supporting the development of biological mass spectrometry. An application currently under consideration by the ARC Linkage Infrastructure and Facilities program for a triple-sector mass spectrometer (a Micromass Q-TOF) should be given full institutional support. Funding should be provided not only for technical support to researchers but also intellectual support to train students, assist with data interpretation, project design and funding applications. Skilled personnel and dedicated equipment should be provided to generate fragments suitable for MS-analysis from very small amounts of protein.

Table 1. Infrastructure supporting Integrative Biology at the ANU

Principal Equipment	Location	Service	Staff
Genomics			
ABI PRISM 3730 Genetic Analyser	ANU BRF	Full or partial	1.5
ABI PRISM 7700 Sequence Detection System	ANU BRF	Self service	
Metabolomics			
Thermo Finnigan ProteomeX LCQ Deca XP LC-API-IT MS	RSBS SPMF	Self service	
Proteomics			
Amino acid composition analysis			
PE AminoQuant Workstation[a]	ANU BRF (JCSMR)	Not available	
Chemical polypeptide sequence determination			
ABI PROCISE HT Protein Sequencer[a]	ANU BRF (JCSMR)	Not available	
ABI PROCISE cLC Protein Sequencer	ANU BRF (JCSMR)	None	
ABI PROCISE C Protein Sequencer[a]	ANU BRF (JCSMR)	Not available	
Biological Mass Spectrometry			
Bruker Omniflex MALDI TOF MS	RSBS SPMF	Self service	
Thermo Finnigan ProteomeX LCQ Deca XP LC-API-IT MS	RSBS SPMF	Self service	
Transcriptomics			
Affymetrix 417 Microarrayer	RSBS SPMF	Self service	
Affymetrix 428 Array Scanner	RSBS SPMF	Self service	
Affymetrix (Agilent) Gene Array Scanner	ANU BRF	Full or partial	2.0
ABI PRISM 7700 Sequence Detection System	ANU BRF	Self service	
[a]decommissioned			

References cited

Bergman T, Cederlund E & Jornvall H, 2001, *Anal. Chem.*, **290**, 74-82.

Botstein D, White RL, Skolnick M & Davis, RW, 1980, *Am. J. Human Genetics,* **32,** 314-331.

Edman P, 1950, *Acta Chem. Scand.*, **4**, 277.

Fountlakis M & Lahm HW, 1998, *J. Chromatogr. A*, **826**, 109-134.

Kan YW, Dozy AM, Varmus HE, Taylor HM, Holland JP, Lie-Injo LE, Ganesan J & Todd D, 1975, *Nature*, **15**, 255-256

McKusick VA, 1989, *New Engl. J. Med.*, **230**, 910-915.

Okiyama N, Santa T, Toriba A, Nakagomi K, Imai K, Hiranuma H & Tanaka H, 2002, *Biomed. Chromatogr.*, **16**, 183-186.

Schlack P & Kumpf W, 1926, *Hoppe-Seyler's Z. Physiol. Chem.*, **154**, 125-170.

Tarr GE, 1977, *Meth. Enzymol.*, **47**, 355.

Venter JC *et al.,* 2001, *Science*, **291**, 1304-1351.

Wang X, Bruderer S, Rafi Z, Xue J, Milburn PJ, Kramer A, & Robinson PJ, 1999, *EMBO J.*, **18**, 4549-59.

Wasinger VC & Humphery-Smith I, 1998, *FEMS Microbiol. Lett.*, **169**, 375-382.

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, & Hochstrasser DE, 1996, *Biotechnology*, **14**, 61-65.

Wilkins MR, Yan JX & Gooley AA, 1999, *Methods Mol. Biol.*, **112**, 445-460.

Wu CC & MacCoss MJ, 2002, *Curr. Opin. Mol. Ther.*, **4**, 242-250.

Zimmerman CL, Appella E & Pisano JJ, 1977, *Anal. Biochem.*, **77**, 569.