Annu. Rev. Biophys. Biomol. Struct. 2001. 30:67–85

MASS SPECTROMETRY AS A TOOL FOR PROTEIN CRYSTALLOGRAPHY

Steven L. Cohen and Brian T. Chait
Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, New York, NY 10021; email: chait@rockvax.rockefeller.edu, steven_cohen@merck.com

Key Words X-ray crystallography, domain elucidation, protein crystals, heavy atom, phasing, modeling

■ **Abstract** Atomic resolution structure determinations of proteins by X-ray crystallography are formidable multidisciplinary undertakings, requiring protein-construct design, expression and purification, crystallization trials, phase determination, and model building. Modern mass spectrometric methods can greatly facilitate these obligate tasks. Thus, mass spectrometry can be used to verify that the desired protein construct has been correctly expressed, to define compact domains in the target protein, to assess the components contained within the protein crystals, and to screen for successful incorporation of seleno-methionine and other heavy metal reagents used for phasing. In addition, mass spectrometry can be used to address issues of modeling, topology, and side-chain proximity. Here, we demonstrate how rational use of mass spectrometry assists and expedites high resolution X-ray structure determination through each stage of the process of protein crystallography.

CONTENTS

1056-8700/01/0610-0067$14.00

PERSPECTIVES AND OVERVIEW

Accurate measurement of the molecular mass of proteins and peptides has opened up attractive new avenues for studying proteins. In this chapter we review the utility of modern mass spectrometric methods for investigating protein structure. We discuss how mass spectrometry (MS) has become an enabling tool for the researcher intent on defining protein structures at atomic resolution.

The growing utility of modern mass spectrometry in biological research derives from the development of two extraordinary methods for ionizing biomolecules, matrix-assisted laser desorption/ionization (MALDI) (3, 22, 41) and electrospray ionization (ESI) (31, 44). These techniques enable the determination of the molecular mass of peptides and proteins with unprecedented accuracy and speed. Thus, protein masses can be routinely measured with accuracies as high as 1 part in 10,000, and peptides can be measured with accuracies higher than 1 part in 100,000. Complete analyses can often be made in a matter of minutes. The total amount of sample required for an analysis is usually in the range 0.1–10 pmol. For the applications discussed here, two types of mass spectrometric measurements are needed. The first is single-stage mass measurement of proteins or fragments of proteins produced by enzymatic/chemical treatment in solution. The second is double stage mass measurement (i.e. tandem MS) in which a chosen peptide is isolated within the mass spectrometer, fragmented in the gas phase, and the masses of the resulting fragments are determined. Single-stage MS measurements provide accurate molecular masses of proteins and proteolytic fragments of proteins, whereas tandem MS measurements provide additional information concerning the amino acid sequence of the chosen peptides.

Atomic resolution structure determinations of proteins are formidable multidisciplinary undertakings encompassing tasks such as protein cloning, expression and purification procedures, crystallization trials, phase determination, and model building. We and others have shown that mass spectrometry can greatly facilitate these obligate tasks (9, 12). Thus, mass spectrometry can be used to verify that the desired construct has been correctly expressed, to define compact domains in the target protein, to assess the components contained within the protein crystals, and to screen for successful incorporation of seleno-methionine and heavy metal reagents used for phasing. In addition, mass spectrometry can be used to address issues of modeling, topology, and side-chain proximity. We demonstrate how rational use of mass spectrometry can greatly facilitate high resolution X-ray structure determination through each stage of the process (Figure 1).

PROTEIN CONSTRUCT VERIFICATION

Isolating milligram quantities of target protein is a prerequisite for protein structure determination. Of equal importance is having available, rapid, accurate means for validating the integrity of the purified product. Techniques for assaying the

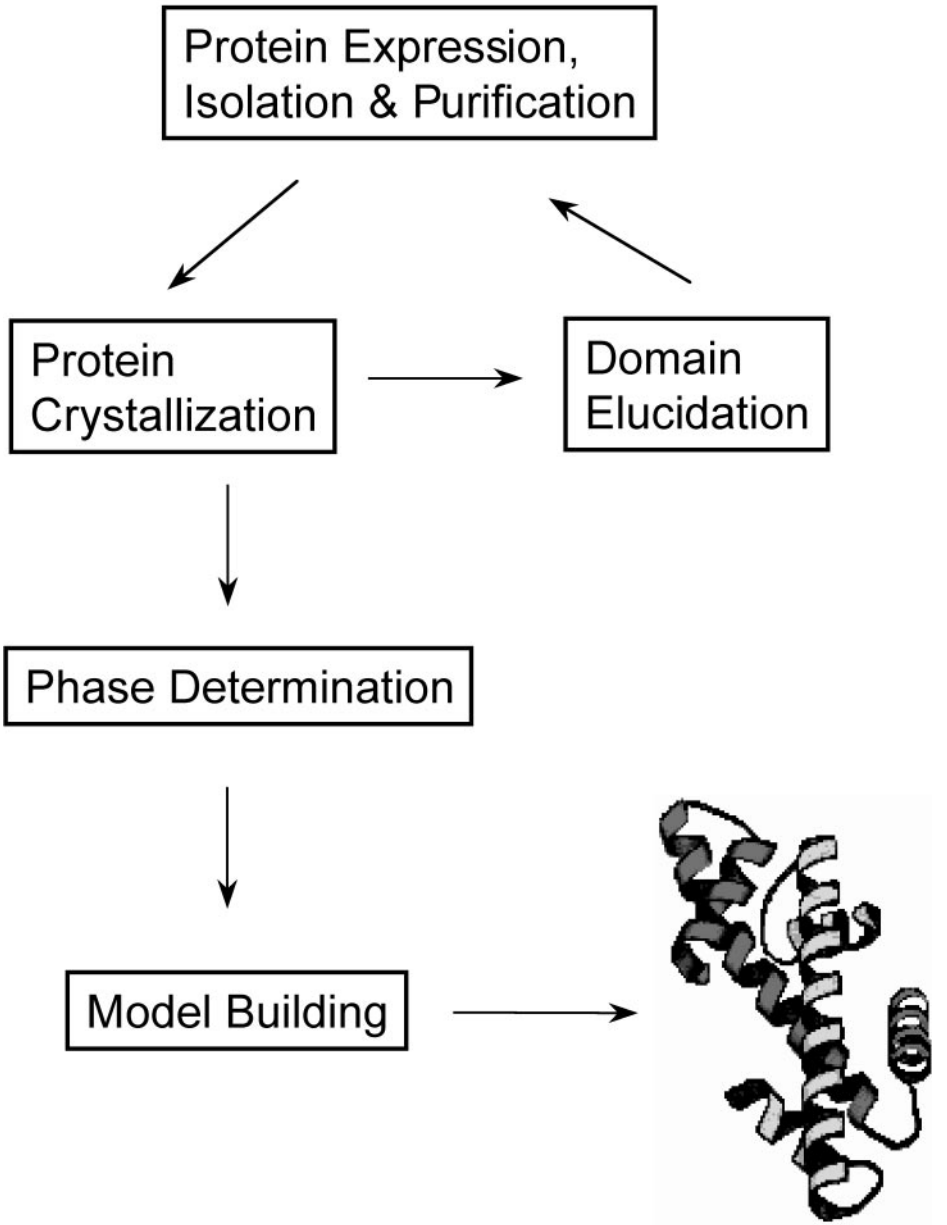

Figure 1 Stages of the X-ray crystal structure determination process where mass spectrometry proves useful.

correctness and purity of target protein constructs include chromatographic methods, gel electrophoresis, dynamic light scattering, ultracentrifugation, and mass spectrometry (9). Mass spectrometry's unsurpassed accuracy and speed for measuring mass allows one to quickly test whether a given protein has been faithfully expressed—i.e., that the primary structure of the protein construct is correct. The measurement allows for detection of PCR errors, mistranslation errors (17), unwanted modifications (e.g., from attachment of β-mercaptoethanol to thiols or oxidation of methionyl residues), and major protein impurities and byproducts (e.g., from degradation). Additional information concerning the nature of the errors/modifications can be obtained by MS peptide mapping (29). To pinpoint the precise site of error/modification, individual proteolytic peptides can be subjected to MS/MS analysis (29).

Mass spectrometry is also particularly useful for detecting desired modifications such as phosphorylation. To illustrate this utility, we consider STAT-1, a transcription factor containing 750 amino acid residues. In the cytoplasm, activation of STAT-1 requires a single phosphorylation at Tyr-701, whereupon the protein dimerizes and is translocated to the nucleus. To obtain a crystal structure of the STAT-1 dimer bound to DNA (11), it was necessary to ensure correct in vitro phosphorylation by the EGF receptor kinase. MALDI-MS peptide mapping demonstrated that the recombinant protein was fully and exclusively phosphorylated on Tyr-701, as desired (43). In another case, mass spectrometry was useful for defining phosphorylation inhomogeneity in a structural study of the *Schizosaccharomyces pombe* casein kinase-1(CK-1). CK-1 undergoes autophosphorylation, leading to multiple forms of the protein that differ in the degree of phosphorylation. Different forms of this protein were resolved by cation exchange chromatography and characterized by electrospray ionization mass spectrometry (8). Using this information, the unphosphorylated form of a truncated variant of CK-1 was isolated and its crystal structure determined in complex with MgATP at 2.0 Å resolution (46).

DOMAIN ELUCIDATION

In order to obtain high-resolution structures, it appears important to work with compact, well-defined proteins. This generally accepted notion presumes that restricting the degrees of conformational motion in a protein increases the chance that it will crystallize into well-ordered lattices. A precisely defined folding domain also results in a protein with reduced tendencies for aggregation.

Numerous computer-aided approaches to help define folding domains are available. These include the use of homologous sequence alignment tools, secondary structure prediction software, and various "threading" algorithms. Progress in genomic sequencing has greatly facilitated homology searching, and there are increasingly available more reliable secondary structure prediction programs. Although these methods may be sufficient to permit an initial design of protein constructs for crystallization, ambiguities concerning the precise boundaries of

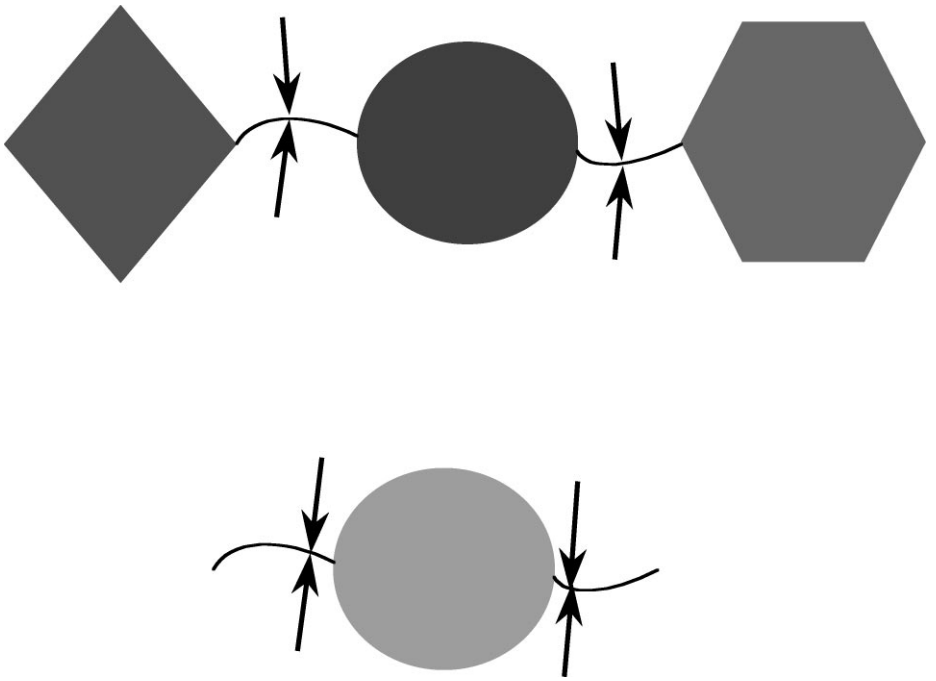

Figure 2 Domain elucidation by mass spectrometry. *Top*: Hypothetical three-domain protein. Folding domains are represented as a diamond, circle, and hexagon. Interdomain linkages are denoted as thin black lines. Limited proteolysis in solution cleaves within the polypeptide linkages (arrows) releasing the individual folding domains. *Bottom*: Single protein domain that has unstructured polypeptide segments extending beyond the compact folding domain. Limited proteolysis trims off these flexible chains (arrows) leaving a more compact folding domain. In both cases, mass spectrometry allows rapid, accurate definition of the domain boundaries.

folding domains often remain. This is particularly true in cases of sequences that exhibit low homology alignments or poorly predicted secondary structural elements. Great efforts are often undertaken to express, purify, and crystallize proteins, only to be thwarted by poorly diffracting crystals. Limited proteolysis combined with mass spectrometry (2, 13) can provide solid biochemical data, usually in a matter of hours, which complements and often resolves ambiguities raised by insufficiencies in the alignment and prediction procedures (Figure 2; 12). We provide examples of several challenging crystallization problems that were resolved with the assistance of mass spectrometric domain elucidation.

The Cap-Binding Protein

The cap-binding protein (known as eukaryotic translation initiation factor 4E, eIF4E) recognizes the 5′ mRNA cap structure during the initiation step of

translation. For many years, crystallization trials of the cap-binding protein were attempted by several structural groups, but without success. Typically the trials would yield thin, long needle-shaped protein crystals (Figure 3a) that were too fragile to handle and diffracted X-rays poorly. The breakthrough in this project was provided by accurately defining the compact domain of the cap-binding protein (32). Full-length cap-binding protein (residues 1-217) complexed with 7-methyl-GDP was subjected to limited proteolysis; the products were analyzed by MALDI-MS. The sites of maximal proteolytic susceptibility were mapped primarily to the N-terminal of the protein (Figure 3b). In addition, sequence alignment of the cap-binding protein from various organisms showed a phylogenetically conserved C-terminus and a divergent N-terminus. Thus, the mass spectrometric results suggested that the conserved C-terminal region corresponded to a proteolytically resistant domain between residues 28-217 or 33-217. These rationally designed constructs were expressed and found to be functionally active. In contrast to the full-length protein, the 28-217 cap construct yielded large cocrystals with 7-methyl-GDP that diffracted to 2.2 Å resolution, providing the first glimpse of this key component of the translation machinery (Figure 3c; 32).

dTAFS 42/62

The TATA box–binding protein associated factors (TAFs) appear to be required for some forms of activated transcription. Two TAF proteins (from *Drosophila*), dTAF42 and dTAF62, show regions of sequence similarity to the histone proteins H3 and H4, respectively. Like H3 and H4, dTAF42 and dTAF62 strongly associate in solution to form heterodimers and tetramers (45). Sequences for the initial constructs for the crystallographic study (residues 1-100 for dTAF42 and 1-91 for dTAF62) were based on the sequence homologies with the histone proteins. Crystals were obtained from these constructs, but they diffracted to no better than 4 Å. It was suspected that the poor diffraction was a result of disorder in the terminal portions of the two proteins. Adding to this suspicion were MS data that revealed extensive protein degradation occurring in a sample of the proteins in solution that was stored at 4°C for 2 months. These MS findings also hinted at potential proteolytically-resistant core domains for both proteins. Limited proteolysis experiments were conducted on fresh sample of the TAF proteins. MALDI-MS mapping of the resulting digests revealed proteolytic susceptibility at the N- and C-terminals of dTAF42 and at the C-terminal of dTAF62 (12, 45). Based on these results, new constructs (11-95 for dTAF42 and 1-82 for dTAF62) were expressed and crystallized. The crystals were large, diffracting to better than 1.4 Å, and yielded the first high-resolution view of the histone fold (7, 45).

Nova RNA-Binding Proteins

Nova-1 and Nova-2 are multidomain neuronal proteins that have been implicated in autoimmune disorders (23). Nova proteins bind RNA through three K-homology (KH) RNA-binding motifs. Initial attempts to design constructs of

the Nova proteins for crystallization were based on sequence alignments of the known KH-motifs. Despite the availability of a large number of such homologous sequences, the precise size of each of the three KH-motifs in the Nova proteins was unclear. To better define the boundaries of these motifs, domain elucidation by MALDI-MS was performed on full-length Nova-1 (511 residues). The ease and speed with which MALDI-MS can provide precise mapping information proves particularly useful for large multidomain proteins such as Nova-1. Rapidly obtained mapping results derived from an initial trial digest guided selection of proteases for subsequent digests. Ultimately, four time-course proteolysis experiments using Asp-N, Glu-C, trypsin, and chymotrypsin were required to fully map the domain organization of the protein (27). The mapping experiments on Nova-1 revealed three proteolytically resistant KH-motifs, with the majority of the proteolytically susceptible sites mapping to the inter-domain regions. The digests provided boundary information for all three KH domains. Cleavage within the KH-motifs mapped to variable loop regions. A construct of the third KH domain (KH3) from Nova-1 and -2 proteins yielded suitable crystals for obtaining the first high X-ray resolution structures of the KH-motif (27) and subsequently for the Nova-2 KH3/RNA complex (28).

Other Examples of Domain Elucidation by Mass Spectrometry

HIV-1 Nef-SH3 Complex ESI and MALDI-MS were used to map limited proteolysis digests of a complex between full-length HIV-1 Nef and the SH3 domain of Hck. The data helped define a crystallizable domain core of the Nef protein. The information obtained from ESI-MS was essentially equivalent to that obtained by MALDI-MS. The crystallographic study resulted in the first high resolution structure of the Nef-SH3 complex (25).

Glutamate Receptor Limited proteolysis analyzed by MALDI-MS combined with N-terminal sequencing, deletion mutagenesis, and sequence alignment helped elucidate the minimal functional ligand-binding domain of the Glu2R glutamate receptor and provided crystallizable constructs that yielded a 1.9 Å resolution structure (1, 10).

Rieske Iron-Sulfur Protein A soluble fragment of the Rieske iron-sulfur protein was isolated by liquid chromatographic methods from limited thermolysin proteolysis of a chloroplast cytochrome b_6f complex. ESI-MS and N-terminal sequencing were used to characterize the protein fragments. Crystals of the 139 residue Rieske fragment that were obtained diffracted X-rays to a resolution better than 2.5 Å (50).

Adenylyl Cyclase Adenylyl cyclase is a large multidomain protein that plays an important role in signal transduction. Initial constructs of the catalytically active cytoplasmic domain of mammalian type II adenylyl cyclase yielded large crystals

that diffracted poorly (48). Limited proteolysis and MALDI-MS mapping revealed a proteolytic-resistant core domain that retained catalytic activity. Based on this MS data, a new slightly trimmed construct was engineered that produced crystals that diffracted to 2.2 Å (48) and provided the first high resolution view of the adenylyl cyclase catalytic domain (49).

RNA Polymerase σ^{70} Subunit Fragment The *Escherichia coli* RNA polymerase σ^{70} subunit is a large (>600 residues) multidomain protein. Trypsin digestion, MALDI-MS, and N-terminal sequencing were used to probe the domain architecture (40). One of the trypsin fragments was found to retain core RNA polymerase binding activity and formed crystals that ultimately yielded a 2.6 Å structure (30).

PROTEIN CRYSTAL ANALYSIS

The process of obtaining protein crystals of sufficient quality for X-ray diffraction analysis often requires the screening of hundreds or even thousands of different protein mother liquor solutions. Such crystallization trials can benefit from mass spectrometric analysis of both the mother liquor and the resulting protein crystals. In the latter case, mass spectrometry ensures the presence of the desired protein or proteins in a given crystal. It further allows for the assessment of protein integrity, permitting the sensitive detection of degradation, oxidation, etc. Finally, MS allows the detection of heavy metal incorporation into individual protein crystals. Examples are provided below.

Determining the Presence of the Expected Protein Components

The TFIIB-TBP-DNA ternary complex is composed of the transcription factor IIB (TFIIB), the TATA box-binding protein (TBP), and DNA containing the TATA-element. Previous work had elucidated the 2.1 Å structure of TBP in complex with its cognate DNA (36). During crystallization trials of the ternary complex, it was predicted that many of the crystals would contain the binary rather than the ternary complex. To optimize time spent at the synchrotron light source, unencumbered by the uncertainty as to which crystals contained the ternary complex of interest, we analyzed several different protein crystals by MALDI-MS. Figure 4 shows the results obtained from two different crystallization trials. The panels labeled "crystal A" show mass spectra obtained from the mother liquor (top) and crystal (bottom) of the human ternary complex. Although both proteins are present in the mother liquor, crystal A was found to contain human TBP but not human TFIIB. In contrast, crystal B, produced from a combination of *Arabidopsis thaliana* TBP and human TFIIB proteins, was seen to contain both proteins. The crystal B form yielded a 2.7 Å structure of the ternary complex (37).

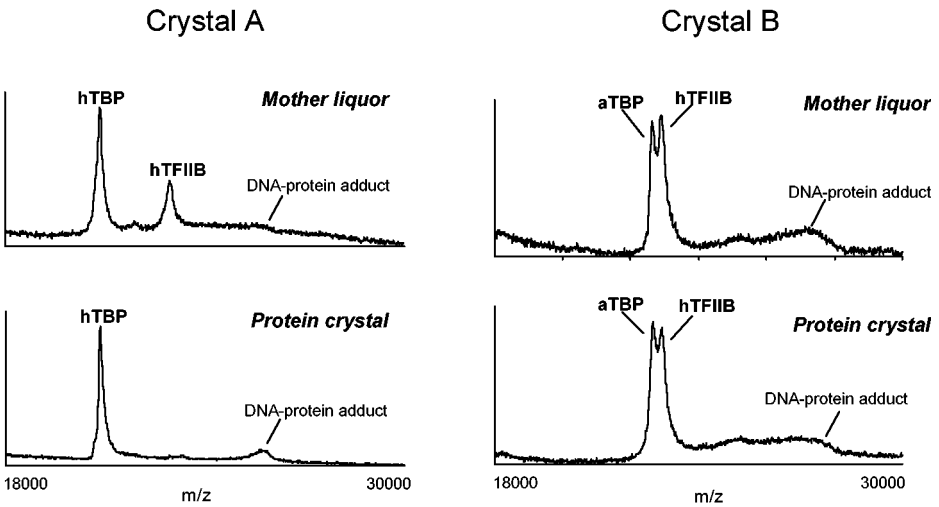

Figure 4 Direct MALDI mass spectrometric analysis of protein crystals of the TFIIB-TBP-DNA ternary complex (37) allowed determination of their protein composition. Crystal A contained TBP but no TFIIB, whereas crystal B contained both TBP and TFIIB.

Examining Alteration of Proteins During Crystallization

Sometimes proteins only crystallize over long periods of time (weeks to months). During such extended crystallization experiments, the proteins may degrade or become modified (e.g., by oxidation). Alterations of this type may prove highly detrimental as in the case of the crystallization of the heterodimer of the TBP-associated factors, dTAF42/dTAF62 (45), where heterogeneity arising from proteolysis degraded the quality of the crystals (see Domain Elucidation section). Although degradation is considered undesirable, it has occasionally been found to be a precondition for crystallization. For example, *Streptococcus pyogenes* β recombinase required $\sim$6 months to produce useful crystals (38). MALDI-MS analysis of these crystals demonstrated that the starting protein (205 amino acid residues in length) had extensively degraded, yielding crystallizable protein consisting of polypeptide fragments, ranging from 126 to 136 amino acids in length (Figure 5). Apparently, this adventitious proteolytic cleavage was key to successful production of diffraction grade crystals.

Assessing Heavy Metal Incorporation

Phase determination is often accomplished by soaking crystals in solutions containing heavy metal reagents (below). Because successful incorporation of these reagents is subject to much trial and error, a rapid method for assessing the

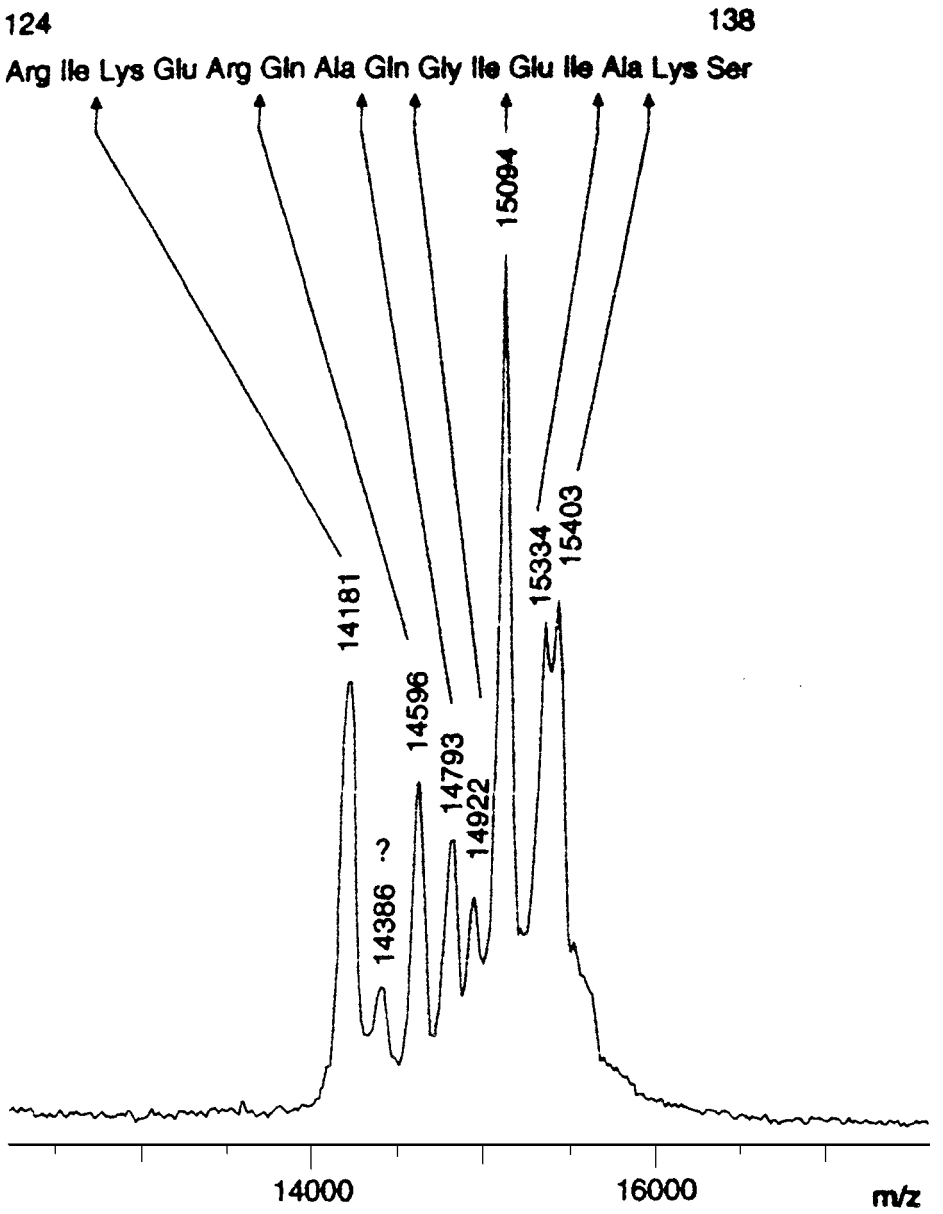

Figure 5 MALDI mass spectrum of dissolved crystal of β-recombinase demonstrated that the starting protein (205 amino acid residues in length) had extensively degraded during the 6-month crystallization period. The data indicate removal of the C-terminal portion of the protein to yield a crystallized product with a frayed end that spanned residues 126-136. Figure reproduced from (38).

efficiency of incorporation is valuable for choosing the optimum reagents. Direct MS analysis of heavy metal-doped crystals serves this purpose well, and we describe an example of this application in the case study of the potassium channel provided at the end of the chapter.

PHASING

Extraction of electron density and structural information from X-ray diffraction data of protein crystals requires knowledge of the magnitudes and phase angles of the diffracted X-rays. The diffraction data itself contains only the magnitudes; the associated phases must be obtained by other means. Available "phasing" methods are isomorphous replacement and molecular replacement techniques. In isomorphic replacement, a heavy atom is introduced into the protein with minimal perturbation of the native protein fold. Phasing information is obtained from X-ray diffraction of a crystal of heavy-atom derivatized protein.

Heavy atoms used for phasing range in size from selenium to uranium. Isomorphic replacement with an anomalous scatterer such as selenium (as selenomethionine, SeMet) is increasingly being used in multiple wavelength anomalous dispersion (MAD) phasing (21). Typically, use of SeMet requires that the protein be recombinantly expressed in an *Escherichia coli* strain that is auxotrophic for methionine and grown on a medium containing selenomethionine (15). Mass spectrometry has been successfully used to ensure high stoichiometric incorporation of SeMet. The mass difference between Met and SeMet is $\sim$47 Da, a difference readily distinguished by mass spectrometry. SeMet incorporation into recombinant protein has been measured by MALDI methods (6, 20) and by ESI methods (19, 24, 39). Bioincorporation of telluromethionine (TeMet), another anomalous scatterer, has also been studied by ESI-MS (5).

More classical isomorphic replacement approaches (e.g., multiple isomorphous replacement, MIR) commonly involve soaking protein crystals with selected heavy atom derivatization reagents. Solvent accessible groups, located in the protein [preferably free sulfhydryls (from cysteine) or histidines or methionines] can aid in heavy atom binding. Sometimes these preferred amino acids have to be introduced into the protein through simple mutation of the original protein sequence. Here, mass spectrometry can confirm that these mutations have been correctly made by measuring the mass of the intact protein and if needed by analyzing a peptide digest of the protein. Successful incorporation of heavy atoms, however, remains a hit-or-miss procedure, requiring the screening of multiple combinations of reagents and derivatization conditions. Knowing the extent to which derivatization has occurred would certainly hasten the screening process. Except for the X-ray analysis itself, there are only two available ways to assess whether crystal soakings are successful at incorporating heavy atoms. The first involves the use of native polyacrylamide electrophoretic gels (4). The second approach uses mass spectrometry (14). Although both methods can be very useful in screening

suitable heavy atom reagents, MS methods can provide additional information such as stoichiometry of incorporation, and for crystal soaking experiments, the ability to analyze single protein crystals to determine the extent of heavy atom incorporation (see next section).

To better understand the potential applications of MS methods to analyze heavy atom derivatives, we undertook a comprehensive study of the mass spectrometry of mercury-derivatized protein (14). Mercurating reagents are commonly used to derivatize proteins because of the highly specific interaction between mercury compounds and the sulfhydryl group. A concern in using MS methods (i.e., ESI and MALDI) to measure the incorporation of mercury or any other heavy atom [except selenomethionine (see above)] is whether the heavy-atom group in the derivatized protein remains bound during the MS measurement process. Because some heavy atom groups are only weakly bound to the protein, we suspected that they may be labile during the MS process. We showed that ESI-MS provides more accurate information concerning the levels of mercuration than does MALDI-MS (ESI is an energetically gentler method for forming intact, labile gas phase ions compared to MALDI). Indeed, our study showed that ESI-MS provides a quantitative measurement of the extent of mercury incorporation into the protein of two mercurating reagents (ethylmercury and p-(chloromercuri)phenylsulfonate) (14). The utility of ESI in a screening of a wide variety of derivatizing reagents was also demonstrated in a structural study of the human immunoglobulin Fc receptor (42). In this study, ESI mass spectra showed successive addition of up to five gold cyanide [$Au(CN)_2^-$] groups attaching to the protein, with the number of additions correlating with increasing concentrations of the derivatizing reagent. Similar binding data were shown for other reagents, such as mercuric chloride, platinum tetrachloride, and lead acetate. Two other reagents (ethylmercuric thiosalicylate and gold chloride) failed to react with the protein. This data guided the authors in their selection of suitable heavy atom reagents for subsequent crystal soaking experiments (from which complete X-ray data sets were obtained). The number of bound heavy metal atoms measured by the ESI-MS correlated well with the heavy-atom refinement data (42).

The ESI-MS method, however, has a practical drawback—the need for thorough sample cleanup of salts, buffers, and other additives prior to the MS analysis. Protein samples that can withstand the clean-up process (by, e.g., dialysis, solid-phase extraction techniques, or HPLC) are excellent candidates for ESI-MS analysis. Otherwise, MALDI-MS can be used to analyze the derivatized protein because MALDI is more tolerant of the salts and additives commonly found in recombinant protein preparations (3). However, our study of mercurated protein showed that mercury compounds bound to protein are partially labile to the MALDI process, leading to underestimation of the levels of metal incorporation (14). The lability is dependent on the choice of MALDI matrix, the matrix solution acidity, and the nature of the mercury substituent. Although not as accurate as ESI in determining the extent of derivatization, suitable MALDI conditions can be found to provide a useful lower limit to the level of metal incorporation. This lower

limit proves sufficient to assist in choosing appropriate derivatization reagents and in selecting protein crystals that have been successfully spiked with metal. The high throughput capabilities and quick turnaround of information provided by the MALDI method positions it as a more practical choice over ESI-MS for screening purposes. This is particularly relevant to a crystallographer whose pending brief visit to a synchrotron facility does not afford the opportunity of a lengthy sample cleanup for ESI analysis.

MODEL BUILDING AND REFINEMENT

After the electron density has been determined from an X-ray diffraction study of a protein, a model of the protein structure is built. One question that frequently arises during model building is whether a stretch of sequence that is not observed in the electron density map is either disordered or altogether missing in the protein crystal (perhaps as a result of degradation over time; see examples above). This question is readily resolved by direct analysis of the crystal of interest. Thus, for example, during the structure determination of the the HERG K+ channel N-terminal domain, we demonstrated using MALDI-MS that residues 1–25 were present in the crystal, although they were absent from the electron density map (35). In a similar manner, it was shown that the entire 172-residue N-terminal FMN-binding domain of *Escherichia coli* NADPH-sulfite reductase α-subunit, although completely absent in the electron density map, was present in the crystallized 60 kDa protein as was determined by MALDI-MS (18). It was concluded that the FMN-binding domain was highly flexible because the remaining portions of the protein (containing the FAD and NADPH binding domains) yielded a 1.9 Å structure.

Cross-linking can be used to provide restraints for model building. For example, during the structure determination of the KcsA potassium channel (16), questions arose as to the connectivity and spatial positioning of the transmembrane helices of the protein. Two possible models were consistent with the diffraction data (Figure 6). To distinguish between these models, strategic sites on the protein were mutated to cysteine and subjected to cross linking. A combination of mass spectrometry and gel electrophoresis readouts of the cross-linking experiments resolved the ambiguity (Figure 6).

Recently, intramolecular protein cross-linking in combination with mass spectrometry provided restraint information that is potentially useful for protein model building (47). Here, lysine residues that are sufficiently close to one another in the protein are cross-linked, and the sites of cross-linking were determined by using proteolysis and mass spectrometry. The resulting data provided limits on the inter-residue distances.

Limited proteolysis, in conjunction with mass spectrometry, can also be used to provide information concerning the topology of integral membrane proteins (33, 34). This information can be used in lieu of high resolution data. The recent

study of the human homopentameric glycine receptor is one example (26). Limited proteolysis of a recombinantly expressed form of the receptor was undertaken and analyzed by HPLC and ESI-MS. The results from this study placed upper limits on the lengths of transmembrane segments and proposed a re-evaluation of previous topology models for the receptor (26).

THE POTASSIUM CHANNEL: A Case Study

Mass spectrometry played an active role throughout the first structural determination of the potassium channel (16). The KcsA potassium channel, from *Streptomyces lividans*, consists of four identical integral membrane protein subunits. As with all membrane proteins, the crystallization experiments were quite formidable. Initial protein constructs were obtained by purifying a full-length C-terminal His-Tag protein (170 residues; 18.8 kDa), followed by trypsin treatment to remove the His-Tag. Although the trypsin treatment yielded sufficient quantities of protein, this material did not produce diffraction grade crystals.

Domain elucidation by mass spectrometry was undertaken to better define the KcsA domain to improve the chances of crystallization and obtain higher quality crystals. Proteolysis experiments were carried out on the full-length C-terminal His-Tag construct of KcsA, and the peptides were mapped by MALDI-MS. An important consideration in analyzing membrane proteins by MALDI-MS is ensuring that the detergent used to solubilize the protein is also well tolerated by the MALDI technique. Detergents (e.g., Mega-9, LDAO) were found to be suitable for protein purification and crystallization and did not interfere with MALDI-MS analysis (7a). Within an hour of starting three different time-course digests (i.e., trypsin, chymotrypsin, and subtilisin), the precise boundaries of a proteolytically resistant "core" of the KcsA channel were revealed. (The choice of proteases used here ensured broad specificity of cleavage.) The MALDI mapping of the digests was straightforward: Each protease quickly removed about 45 residues from the C-terminal of the His-Tag KcsA, leaving a 122-residue (by trypsin) or 125-residue (by chymotrypsin or subtilisin) core domain. There was little indication that the 45-residue portion that was proteolytically trimmed from the C-terminal of full-length KcsA had any inherent structure. Beyond one day of digestion, the C-terminally-truncated protein began to show an additional cleavage site 19 residues (by trypsin) or 25 residues (by subtilisin) from the N-terminal. Chymotrypsin did not cleave anywhere in the N-terminal.

The limited proteolysis MS results guided in designing a new construct for crystallization. The full-length, 170-residue C-terminal His-Tag protein was expressed, and chymotrypsin was used to cleave off the C-terminal to yield the new construct (residues 1-125). We showed that chymotrypsin was a better choice than trypsin because it did not cleave KcsA at the N-terminal. Interestingly, during the initial production of the new KcsA construct with chymotrypsin, the utility of mass spectrometry again came into play. MS revealed that in addition to the expected

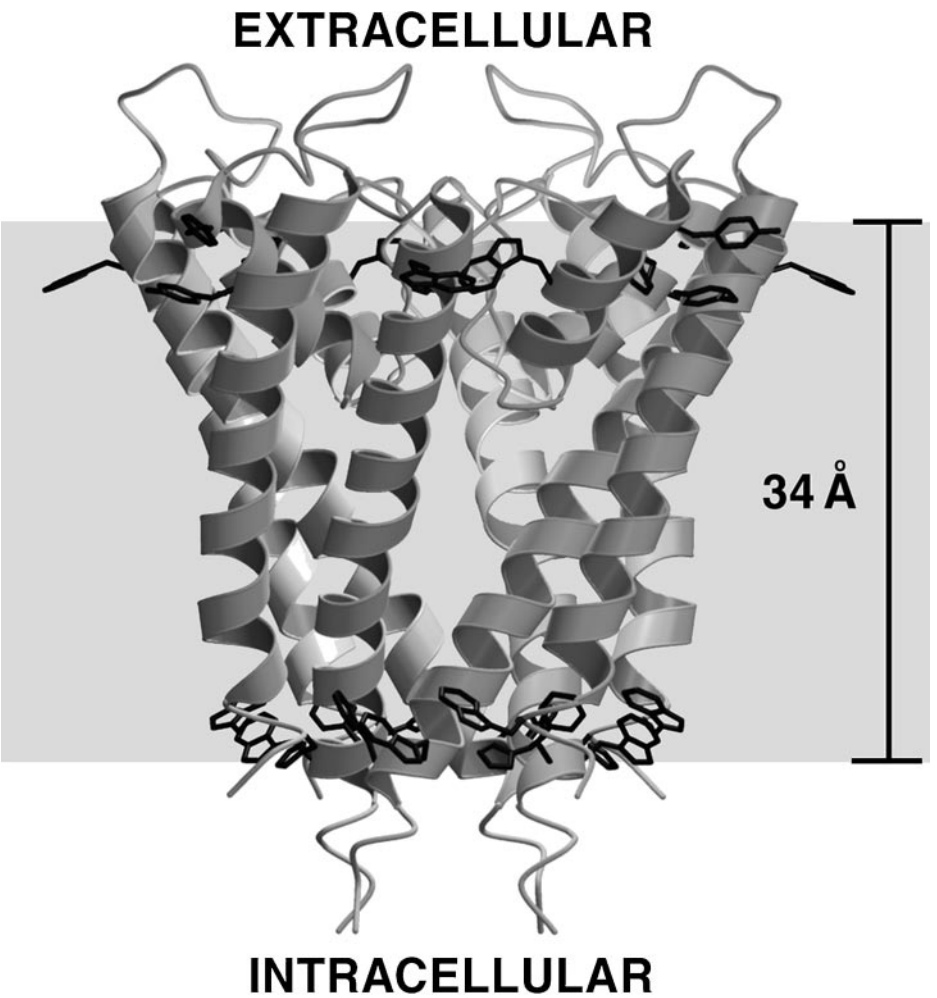

Figure 8 Ribbon representation of the KcsA potassium channel (16).

chymotryptic cleavage product (KcsA 1-125), an unanticipated second population of the protein appeared (KcsA 20-125). This second cleavage product was not easily distinguished from the expected product by SDS-PAGE, and its presence likely would have adversely affected the crystallization trials. MS precisely identified the second product as having originated from a trypsin-like cleavage in the N-terminal, and its formation was thus attributed to the presence of trace levels of trypsin, a common contaminant found in chymotrypsin preparations. To avoid the trypsin activity, TLCK-treated chymotrypsin was used to ensure a homogenous population of proteolytically resistant KcsA. This chymotrypsin construct provided high quality diffracting crystals of the potassium channel (16).

Phasing of the KcsA structure was also greatly facilitated by mass spectrometry. Wild-type KcsA does not contain cysteinyl residues. Thus, it was decided to introduce a cysteine via mutation to permit site-specific attachment of a heavy atom. Because it was not clear where to place the cysteine within the protein, a large number of KcsA cysteine mutants were prepared. In addition, it was uncertain which heavy metal derivatizing reagent would prove optimal. MALDI-MS was used to screen several of the mutants after reaction in solution with a variety of different mercurating reagents. In this way, we were able to identify the most effective mercurating reagents and mutants that labeled with high efficiency. Based partially on this data, a series of KcsA cysteine mutants were subjected to crystallization trials (the crystals were grown in the absence of the mercurating reagents). Once large crystals were obtained, they were soaked in the optimal mercurating reagents, and the individual crystals were analyzed by MALDI-MS to determine the extent of mercury incorporation (see Phasing Section for details; 14). MS identified crystals that were efficiently labeled and thus likely to be useful for phasing. Figure 7 shows a superposition of the sites of mercury atom attachment determined from the large collection of single cysteine mutants that successfully incorporated mercury.

Finally, mass spectrometry was used to assist in the model building process (see above) that yielded the beautiful and informative structure shown in Figure 8 (16).

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Center for Research Resources, NIH (Grant RR00862-27) to BTC. It is a genuine pleasure to acknowledge our many fine collaborators at Rockefeller University including Rod MacKinnon, Stephen Burley, Seth Darst, John Kuriyan, Joseph Marcotrigiano, Hal Lewis, João Morais Cabral, Declan Doyle, Adrian Ferré-D'Amaré, Xiaoling Xie, and Dimitar Nikolav. We also wish to thank Júlio Cesar Padovan, Urooj Mirza, Martine Cadene, Wenzhu Zhang, David Fenyö, and Ron Beavis for their great intellectual input and friendship.

LITERATURE CITED

1. Armstrong N, Sun Y, Chen GQ, Gouaux E. 1998. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* 395:913–17
2. Bantscheff M, Weiss V, Glocker MO. 1999. Identification of linker regions and domain borders of the transcription activator protein NtrC from *Escherichia coli* by limited proteolysis, in-gel digestion, and mass spectrometry. *Biochemistry* 38:11012–20
3. Beavis RC, Chait BT. 1996. Matrix-assisted laser desorption ionization mass-spectrometry of proteins. *Methods Enzymol.* 270:519–51
4. Boggon TJ, Shapiro L. 2000. Screening for phasing atoms in protein crystallography. *Struct. Fold Des.* 8:R143–49
5. Budisa N, Karnbrock W, Steinbacher S,

Humm A, Prade L, et al. 1997. Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. *J. Mol. Biol.* 270:616–23

6. Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, et al. 1995. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* 230:788–96
7. Burley SK, Xie X, Clark KL, Shu F. 1997. Histone-like transcription factors in eukaryotes. *Curr. Opin. Struct. Biol.* 7:94–102

7a. Cadene M, Chait BT. 2000. A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* 72:5655–58

8. Carmel G, Leichus B, Cheng X, Patterson SD, Mirza U, et al. 1994. Expression, purification, crystallization, and preliminary x-ray analysis of casein kinase-1 from Schizosaccharomyces pombe. *J. Biol. Chem.* 269:7304–09
9. Chait BT. 1994. Mass spectrometry—a useful tool for the protein X-ray crystallographer and NMR spectroscopist. *Structure* 2:465–67
10. Chen GQ, Sun Y, Jin R, Gouaux E. 1998. Probing the ligand binding domain of the GluR2 receptor by proteolysis and deletion mutagenesis defines domain boundaries and yields a crystallizable construct. *Protein Sci.* 7:2623–30
11. Chen X, Vinkemeier U, Zhao Y, Jeruzalmi D, Darnell JE Jr, et al. 1998. Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* 93:827–39
12. Cohen SL. 1996. Domain elucidation by mass spectrometry. *Structure* 4:1013–16
13. Cohen SL, Ferré-D'Amaré AR, Burley SK, Chait BT. 1995. Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci.* 4:1088–99
14. Cohen SL, Padovan JC, Chait BT. 2000. Mass spectrometric analysis of mercury incorporation into proteins for X-ray diffraction phase determination. *Anal. Chem.* 72:574–79
15. Doublie S. 1997. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* 276:523–30
16. Doyle DA, Cabral JM, Pfuetzner RA, Kuo A, Gulbis JM, et al. 1998. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* 280:69–77
17. Forman MD, Stack RF, Masters PS, Hauer CR, Baxter SM. 1998. High level, context dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*. *Protein Sci.* 7:500–03
18. Gruez A, Pignol D, Zeghouf M, Coves J, Fontecave M, et al. 2000. Four crystal structures of the 60 kDa flavoprotein monomer of the sulfite reductase indicate a disordered flavodoxin-like module. *J. Mol. Biol.* 299:199–212
19. Hädener A, Matzinger PK, Malashkevich VN, Louie GV, Wood SP, et al. 1993. Purification, characterization, crystallisation and X-ray analysis of selenomethionine-labelled hydroxymethylbilane synthase from *Escherichia coli*. *Eur. J. Biochem.* 211:615–24
20. Hage T, Reinemer P, Sebald W. 1998. Crystals of a 1:1 complex between human interleukin-4 and the extracellular domain of its receptor alpha chain. *Eur. J. Biochem.* 258:831–36
21. Hendrickson WA. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254:51–58
22. Hillenkamp F, Karas M, Beavis RC, Chait BT. 1991. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 63:1193A–1203A

23. Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, et al. 2000. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* 25:359–71
24. Kozak M, Jankowska E, Janowski R, Grzonka Z, Grubb A, et al. 1999. Expression of a selenomethionyl derivative and preliminary crystallographic studies of human cystatin C. *Acta Crystallogr. D Biol. Crystallogr.* 55:1939–42
25. Lee CH, Saksela K, Mirza UA, Chait BT, Kuriyan J. 1996. Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. *Cell* 85: 931–42
26. Leite JF, Amoscato AA, Cascio M. 2000. Coupled proteolytic and mass spectrometry studies indicate a novel topology for the glycine receptor. *J. Biol. Chem.* 275:13683–89
27. Lewis HA, Chen H, Edo C, Buckanovich RJ, Yang YY, et al. 1999. Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure* 7:191–203
28. Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, et al. 2000. Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* 100:323–32
29. Li X, Himanen JP, Martin de Llano JJ, Padovan JC, Chait BT, et al. 1999. Mutational analysis of sickle haemoglobin (Hb) gelation. *Biotechnol. Appl. Biochem.* 29:165–84
30. Malhotra A, Severinova E, Darst SA. 1996. Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127–36
31. Mann M, Wilm M. 1995. Electrospray mass spectrometry for protein characterization. *Trends Biochem. Sci.* 20:219–24
32. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. 1997. Cocrystal structure of the messenger RNA 5′ cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* 89:951–61
33. Mel SF, Falick AM, Burlingame AL, Stroud RM. 1993. Mapping a membrane-associated conformation of colicin Ia. *Biochemistry* 32:9473–79
34. Moore CR, Yates JR III, Griffin PR, Shabanowitz J, Martino PA, et al. 1989. Proteolytic fragments of the nicotinic acetylcholine receptor identified by mass spectrometry: implications for receptor topography. *Biochemistry* 28:9184–91
35. Cabral JHM, Lee A, Cohen SL, Chait BT, Li M, et al. 1998. Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain. *Cell* 95:649–55
36. Nikolov DB, Burley SK. 1994. 2.1-Å resolution refined structure of a TATA box-binding protein (TBP). *Nat. Struct. Biol.* 1:621–37
37. Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, et al. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377:119–28
38. Orth P, Jekow P, Alonso JC, Hinrichs W. 1999. Proteolytic cleavage of gram-positive beta recombinase is required for crystallization. *Protein Eng.* 12:371–73
39. Pieper M, Betz M, Budisa N, Gomis Rüth FX, Bode W, Tschesche H. 1997. Expression, purification, characterization, and X-ray analysis of selenomethionine 215 variant of leukocyte collagenase. *J. Protein Chem.* 16:637–50
40. Severinova E, Severinov K, Fenyö D, Marr M, Brody EN, et al. 1996. Domain organization of the *Escherichia coli* RNA polymerase sigma 70 subunit. *J. Mol. Biol.* 263:637–47
41. Stults JT. 1995. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). *Curr. Opin. Struct. Biol.* 5: 691–98
42. Sun PD, Hammer CH. 2000. Mass-spectrometry assisted heavy-atom derivative screening of human Fc gamma RIII crystals. *Acta Crystallogr. D Biol. Crystallogr.* 56:161–68

43. Vinkemeier U, Cohen SL, Moarefi I, Chait BT, Kuriyan J, Darnell JE Jr. 1996. DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. *EMBO J.* 15:5616–26
44. Wilm M. 2000. Mass spectrometric analysis of proteins. *Adv. Protein. Chem.* 54:1–30
45. Xie X, Kokubo T, Cohen SL, Mirza UA, Hoffmann A, et al. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* 380:316–22
46. Xu RM, Carmel G, Sweet RM, Kuret J, Cheng X. 1995. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J.* 14:1015–23
47. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, et al. 2000. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. USA* 97:5802–06
48. Zhang G, Liu Y, Qin J, Vo B, Tang WJ, et al. 1997. Characterization and crystallization of a minimal catalytic core domain from mammalian type II adenylyl cyclase. *Protein Sci.* 6:903–08
49. Zhang G, Liu Y, Ruoho AE, Hurley JH. 1997. Structure of the adenylyl cyclase catalytic core. *Nature* 386:247–52 Erratum. 1997. *Nature* 388 (6638):204
50. Zhang H, Carrell CJ, Huang D, Sled V, Ohnishi T, et al. 1996. Characterization and crystallization of the lumen side domain of the chloroplast Rieske iron-sulfur protein. *J. Biol. Chem.* 271:31360–66

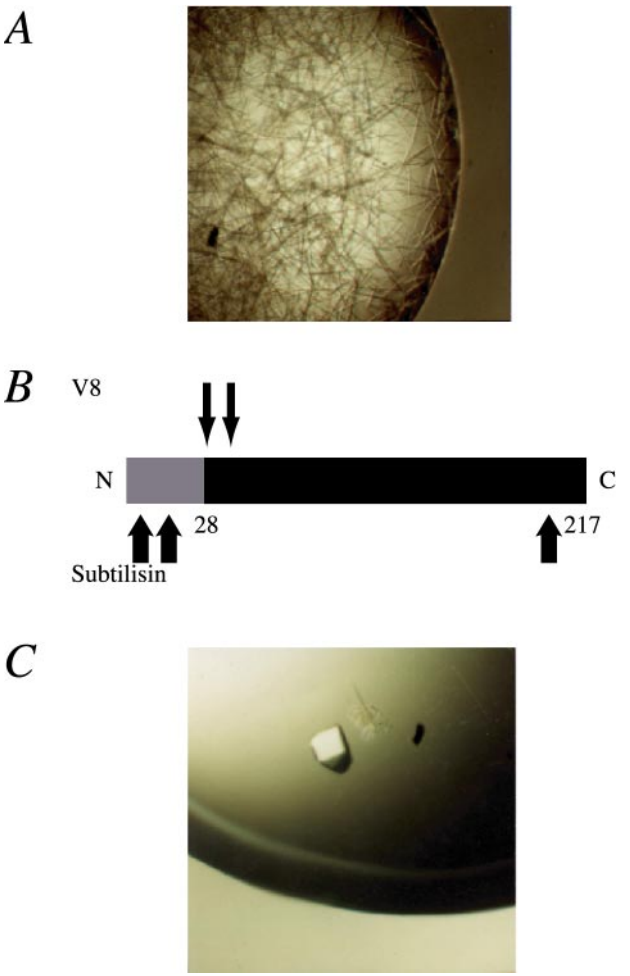

Figure 3 (*A*) Full-length cap-binding protein (residues 1–217) in complex with 7-methyl-GDP crystallizes as fine needles that diffract poorly. (*B*) The complex was subjected to limited proteolysis, and the resulting products were analyzed by MALDI-MS. The arrows indicate the sites of proteolytic susceptibility. These data were used to design a new compact construct (residues 28–217). (*C*) The new compact construct yields large crystals that diffract to 2.2 Å resolution. Data are from reference 32.

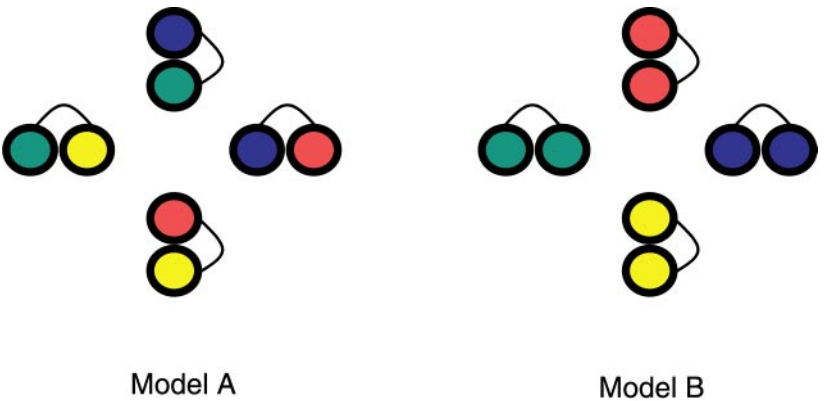

Figure 6 Chemical cross-linking provided restraints for model building in the solution of the structure of the KcsA potassium channel (16). The figure depicts a view looking down the pore of the channel. The discs represent the transmembrane helices (2 per monomer unit), and the four colors denote the four monomer units of the homotetrameric channel. Indicated are the predicted outcomes of cross-linking experiments (denoted by the arcs between the discs) assuming two possible arrangements of the transmembrane helices (i.e., Model A and Model B). A combination of mass spectrometry and SDS-PAGE was used to differentiate between these two possible cross-linked forms of the channel. The data clearly demonstrated cross-linking between the two transmembrane helices within each monomer unit consistent with Model B, rather than intermonomer cross-links that would be predicted for the transmembrane helices arrangement depicted in Model A.

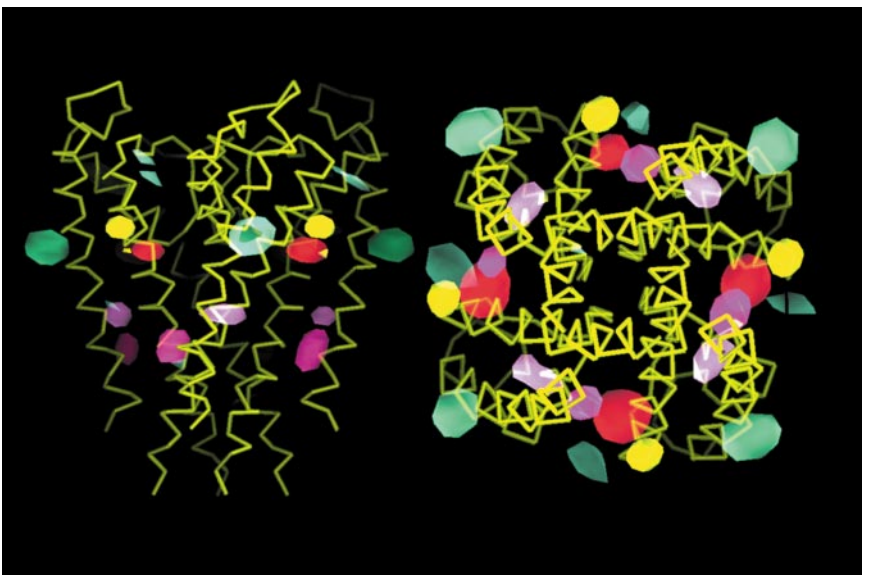

Figure 7 Sites of mercury atom attachment (colored blobs) superimposed on the backbone structure of the KcsA potassium channel. Side-view and top-view are shown. The different colors represent the sites of attachment to different single cysteine mutants. Figure courtesy of R MacKinnon.