X-ray Diffraction at Synchrotron Light Sources

John R Helliwell, *University of Manchester, Manchester, UK*

The elucidation of biological structure and function for macromolecules and their complexes is undertaken extensively using X-ray diffraction. Synchrotron X-radiation is widely used, due to its exceptional X-ray brilliance and tunability properties, in protein crystallography, fibre diffraction and solution scattering.

Introductory article

Article Contents

Historical Context and Introduction

Synchrotron radiation (SR) is emitted when charged particles, usually electrons but also positrons, travel at speeds close to the speed of light, and are held in circular orbits by a magnetic field. These machines, accelerators, were originally developed for particle physics experiments where the charged particles struck a target or collided with counter-rotating oppositely charged particles. The emission of synchrotron light, in a continuum from X-rays to infrared wavelengths, was a nuisance by-product, an energy loss that consumed electricity (as radio frequency power), as far as particle physics was concerned. However, it was clear from calculations by J. Schwinger in the late 1940s and later by L. Parratt, as well as experimentally, that this was a potentially very bright, very versatile X-ray source for the study of the structure of matter by diffraction, as well as by spectroscopic and imaging methods. By comparison, the 'standard' X-ray tube, invented by W. Roentgen in 1895, emits rather weak specific wavelengths (Kα, Kβ lines), albeit sitting on a white background of X-rays, but which is exceedingly weak compared with white SR.

There is a long history to X-ray crystallography, which commenced in 1912, before SR was first observed in 1948 and the first SR diffraction experiment was performed by Y. Cauchois in Frascati, Italy in 1963 and the first biological diffraction from muscle fibres by G. Rosenbaum, K. C. Holmes and J. Witz (published in 1971 in *Nature*) recorded in Hamburg, Germany. The first SR protein crystal diffraction patterns were recorded at the Stanford Linear Accelerator Center's (SLAC) electron storage ring 'SPEAR' by J. C. Phillips and K. O. Hodgson in 1976. The first dedicated SR X-ray source was the Synchrotron Radiation Source ('SRS') Daresbury, UK (1981), followed by various others around the world, and from these 'second generation' of SR sources came a steady flow of innovations in methods. Gradually, large numbers of protein structure determinations occurred. The study of noncrystalline samples such as muscle fibres as well as solutions also progressed well on these sources.

A third generation of SR sources has come on-line in the 1990s. These are much larger than most of the second-generation sources and thus allow 'insertion device' periodic magnet sources (wigglers and undulators) to be incorporated in the ring (**Figure 1**), and are of exceptional brilliance. Since biological molecular samples are usually small (< 0.5 mm, often 50 μm or even smaller) the brilliance property is very important. Thus micro-samples of all types, crystals, fibres and solutions, have all become amenable to study. In another application the intensities are now such that the electron bunch structure of the circulating particle beam can be exploited. Even single bunch sub-nanosecond time-resolved structural studies are possible on protein crystals. These experiments involve some perturbation or stimulation of the sample followed, after a preset time gap, by the recording of the X-ray diffraction patterns. These polychromatic patterns (**Figure 2**) are known as Laue patterns as they involve a stationary crystal, as in the original 1912 X-ray tube-based experiment. The largest single category of SR instruments in the X-ray region comprises those used for data collection on crystals of biological macromolecules or larger biological macromolecular complexes (**Figure 3**).

A general consideration for X-ray diffraction structural studies is that of radiation damage to the sample, whether it is a crystal, fibre or a solution of biological macromolecules. The basic effect, and problem, is that X-rays are absorbed and the energy deposited is converted into reactive free radicals which can diffuse around the sample and disrupt the weak molecular bonds that exist between the molecules and thereby cause a deterioration of the state of sample order. For a protein crystal, for example, the resolution deteriorates. In the most brilliant SR X-ray beams the energy deposited is also enough to heat the sample by several degrees per second (although the exact temperature is difficult to measure or calculate exactly). Several approaches have helped to get around these

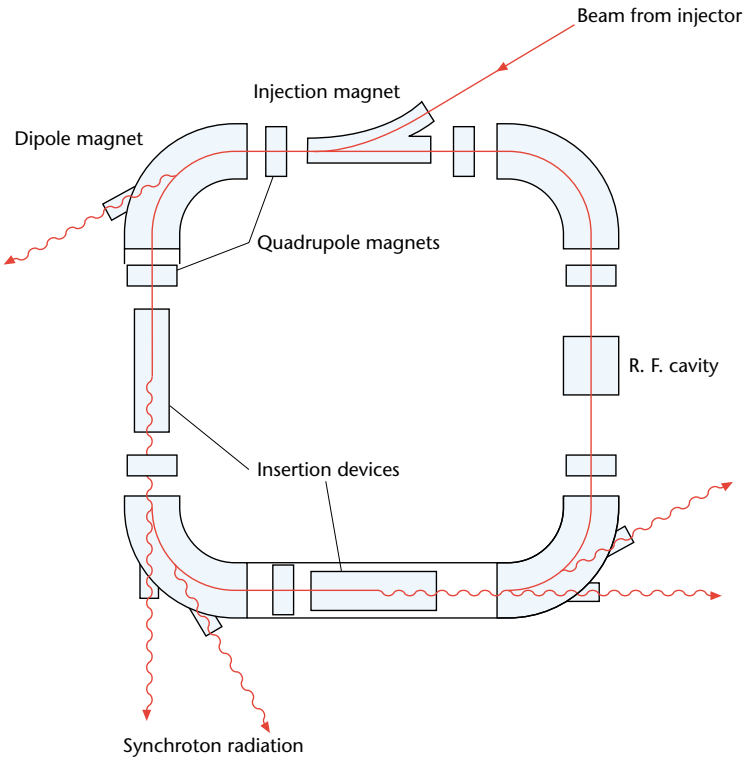


Figure 1 A schematic diagram of the layout of a synchrotron radiation (SR) source. In practice, SR storage rings are much larger, having many more machine 'cells' than the four shown here. At the European SR Facility (ESRF) in Grenoble, for example, the third-generation high-brilliance source is 840 m in circumference and has 32 cells.

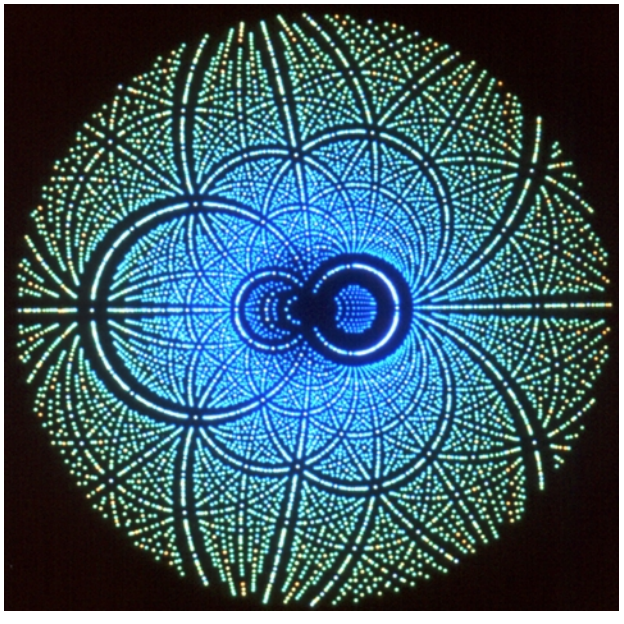
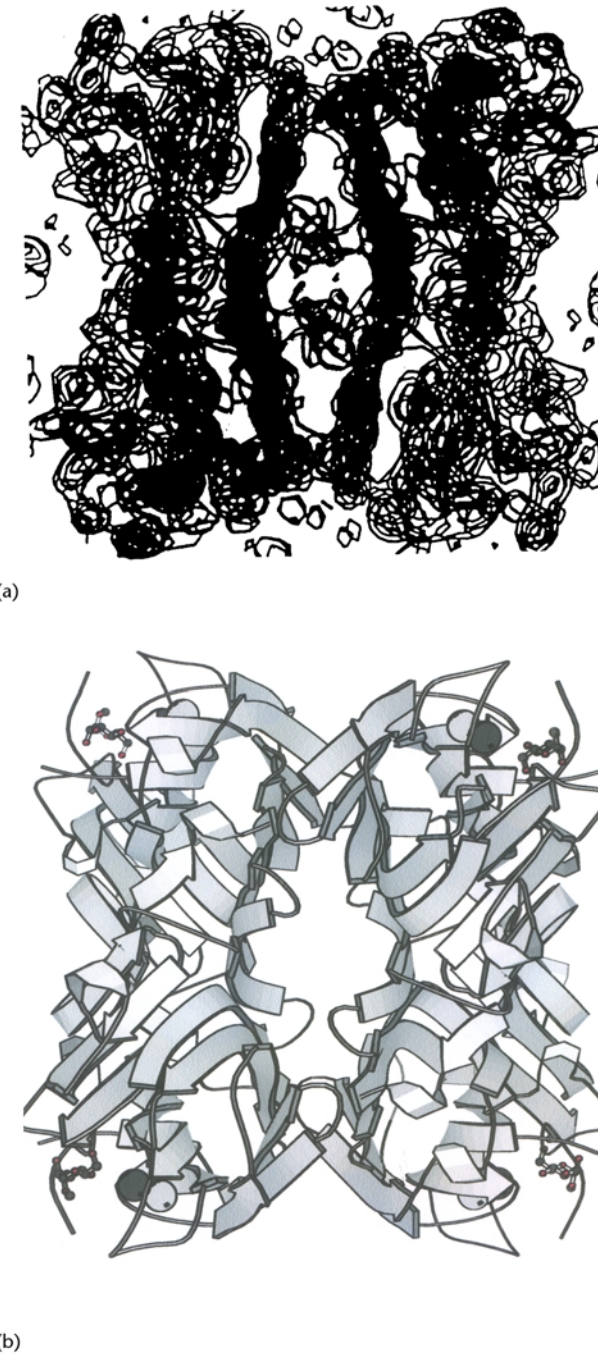


Figure 2 A polychromatic Laue diffraction pattern from a protein crystal with spots colour-coded for wavelength. Short wavelengths in blue and then through the rainbow colours to red at the longest wavelengths. Actual wavelength bandpass used 0.5 Å to 2.5 Å. Reproduced with the permission of Acta Crystallographica.

chemically or thermally damaging effects. The use of short wavelengths reduces the absorption and since it varies as wavelength cubed versus the diffraction signal varying as wavelength squared there is a benefit of the short wavelength. The faster exposure time at SR sources is also valuable since the diffusion time of the free radicals is significant. The most effective approach of all is to cool the sample to cryo temperature (100 K), thus preventing the diffusion of free radicals altogether. However, the sample lifetime is not infinite. The deposited energy will break bonds and eventually cause physical break up of the sample order. While a big benefit, there are subtle structural changes upon cryo-cooling which increase the number of multiple conformations of amino acid side-chains and affect the vibrational signatures of atoms. Neutron spectroscopy techniques show that there is a glass-like transition at around 180 K and a loss of protein function occurs below such a temperature. Naturally, there is a continued interest in room temperature measurements.



(a)

(b)

Figure 3 (a) Electron density map image of a protein crystal. (b) Ribbon diagram of the same protein crystal to the same scale and the same view. The protein is the tetramer of concanavalin A from jack beans with glucoside in atom-by-atom format and metal atoms also shown as individual spheres. Concanavalin A–sugar complexes are typical of the proteins studied to high resolution at a second-generation SR source (in this case Daresbury SRS), i.e. up to 100 kDa in the asymmetric unit of the crystal.

Structure Elucidation of Large Biomolecules at or Near to Atomic Resolution

Use is made of the high intensity and collimation of SR for recording the weak, high-angle outer parts of the diffraction pattern which allows a high-resolution protein crystal structure refined model to be achieved.

The resolution achievable at the synchrotron is now often reaching 'atomic resolution' (i.e. where the electron density images of the protein crystal unit cell resolve individual atoms). Indeed the data available now can reach a sufficient number to allow molecular model refinement where the individual motions of atoms are described by a full anisotropic model (ellipsoids) and single versus double bonds are clearly resolved as is done routinely in chemical (smaller molecule) structure refinement.

The technical advance of cryo-cooling the protein crystal sample (typically at 100 K) to prevent X-ray irradiation damage has also widened the number of such studies. This method was introduced in the study of ribosome crystals, pioneered by A. Yonath and H. Hope and co-workers, and is now used for almost all protein crystal samples especially at high-brilliance third-generation SR sources. At such sources beam damage even of cryo-cooled crystals is observed on the insertion device beamlines.

Studies of Crystals with Large Unit Cell Dimensions

Large macromolecular complexes, i.e. of proteins with proteins or proteins with nucleic acids, in cases such as spherical or cylindrical viruses or the ribosome, have been crystallized and their structures determined. These are at resolutions typically of 3 Å, occasionally 2.5 Å. Nevertheless, for the size of these structures these are major achievements of structural biological crystallography. The first animal virus determined was the common cold virus (rhinovirus) by M. G. Rossmann using the Cornell 'CHESS' Synchrotron, USA in 1985, with preliminary work undertaken at Daresbury, UK and Hamburg, Germany. **Figure 4** shows one of the diffraction patterns measured from a rhinovirus crystal at CHESS. Approximately 160 different crystals, each yielding one such pattern, were used in the analysis. The crystal structure of the large subunit of the ribosome has been determined by P. D. Moore and T. Steitz and colleagues at Yale University in 2000 using SR data from the National Synchrotron Light Source at Brookhaven National Laboratory, USA. A variety of ribosomal subunit structural studies are now being reported in the literature. The ribosome is now known to be a ribozyme.

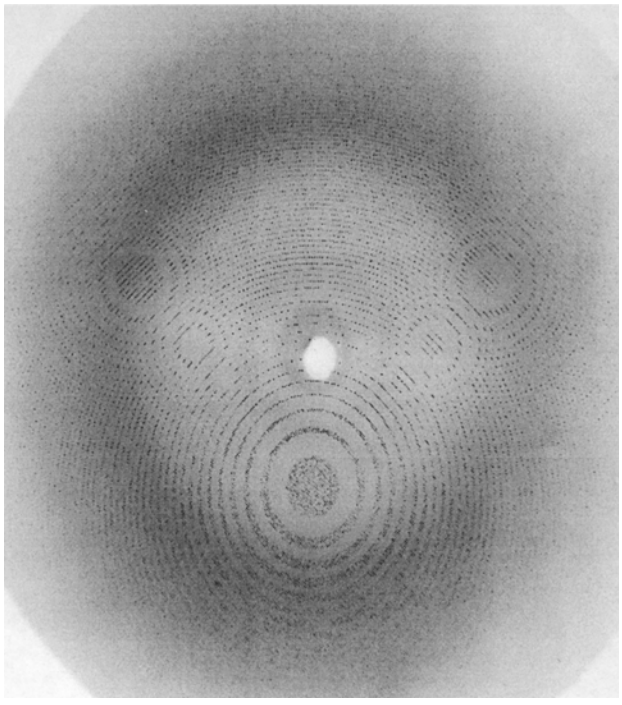


Figure 4 A monochromatic rotating crystal diffraction pattern recorded from a human rhinovirus crystal at the Cornell High Energy SR Source (CHESS), USA. Rhinovirus was the first mammalian virus structure to be solved (in 1985). Reproduced with the permission of Professor M. G. Rossmann, Purdue University, USA and *Acta Crystallographica*.

These crystals make optimal use of the high intensity of the SR beam as they are weakly diffracting, as one can readily imagine because there are rather few unit cells in the crystal (e.g. 10^8 rather than 10^{14}). Moreover the diffraction spots are very close together in angle and so the fine collimation (small divergence) of the SR beam is needed. SR diffraction pattern recording for crystal unit cells in excess of 1000 Å edge length is now quite feasible.

Multiple Wavelength Anomalous Dispersion (MAD) Technique

The tunability of synchrotron radiation has made this method of structure elucidation feasible, and it is growing in importance. The core problem in solving a protein crystal structure, which this method addresses in a direct way, is the determination of the phase angle of each and every X-ray reflected beam from a protein crystal. From the 1950s through to the 1980s the method used involved heavy atom soaking of protein crystals with measurements being made from each crystal type, using a method called multiple isomorphous replacement (MIR). However, instead, it is possible to utilize the wavelength-dependent changes ('anomalous dispersion') to the X-ray scattering of metal atoms that are in a protein naturally or are introduced via derivatization or, more frequently today, by deliberately producing selenomethionine-substituted proteins. Exploitation of these effects (at maximum approximately only 10–15% of the wavelength-independent scattering of an atom) required the improvements available from SR X-ray intensity data collection to reliably measure the effects, which are typically at the 1–2% change in intensity level. The technical requirements of the SR source and beamline necessitated rather exacting developments of source and beamline optics stability, i.e. exact wavelength and wavelength-band setting, as well as improved diffraction detectors (i.e. image plates and charge coupled device detectors). Today the MAD method is allowing a much higher throughput of protein crystal structure determinations since the recording of the data, the determination of the positions of the metal sites and the calculation of an interpretable electron density image is achievable on a third-generation source undulator beamline in hours. The bottlenecks that remain are high-throughput expression and purification of pure protein as well as high-throughput crystallization. A new field of 'structural genomics' has thus opened up.

Time-resolved Structural Analysis and Polychromatic X-ray Diffraction

The purpose of structure elucidation is to understand the biological function. A direct way to probe function is to stimulate the functional cycle of a molecule or molecular biological system and then to probe the structural intermediate state in real time using X-ray diffraction measurements at given time points after the stimulation. This is a relatively new area of SR diffraction research and development, and to date only 10 successful biochemical crystallographic real-time studies have been published. The methods of stimulation include diffusion of substrate (for enzymes), pH jump, light flash or temperature jump. The most prompt method of stimulation is light flash and is applicable to naturally light-sensitive proteins (e.g. bacteriorhodopsin, phospho yellow protein, carbon monoxide myoglobin to name three cases that have been studied this way). Also amenable to this approach are cases where a 'caged' substrate is pre-equilibrated in the crystal sample (e.g. by diffusion or co-crystallization) and the natural substrate is released from the cage at the protein active site upon light flash. Protein crystals might be thought to be too rigid an environment to allow protein structure conformational change. Of course this can indeed be the case but the fact that there is a large fraction of the crystal that is open solvent, even up to 80% in some cases, allows some conformational change to occur, unless the active site is truly blocked by molecule-to-molecule crystal packing. Obviously large multidomain rearrangements are not

feasible in the crystalline state without break up of the crystal, and these cases can only be investigated in solution.

A different but related approach, at least for slower structural processes, is to freeze the protein crystal to trap structural intermediates. Since the intrinsic time to freeze the crystal (of around < 0.5 mm size) is in the sub-second range this sets a limit to the time resolution achievable. A faster approach involving freeze trapping is to use very thin samples and to use electron diffraction techniques; then the freezing time is a matter of milliseconds. Electrons scatter more strongly than X-rays but are more damaging. They are used as preference, however, where three-dimensional crystals cannot be obtained. In all freeze-trapping methods the precise moment to trap the intermediate structure has to be determined and this decision can be guided by spectral measurements on the protein crystal or if there is not a spectral signal associated with the structural change then it can be done via repeated recording of a small subset of the diffraction spots to look for intensity changes.

For faster structural changes only the real time approach is viable. Using polychromatic SR diffraction, whereby a large bandpass of wavelengths is used, parallel acquisition of the diffraction X-ray reflections is possible (**Figure 2**). This is known as the Laue method as it is the same diffraction geometry used in the original 1912 experiment, but of course using a hugely powerful X-ray source by comparison! Exposure times in the sub-nanosecond range have been realized on third-generation SR sources. Clearly on such time scales the rotation of the crystal, as employed with the standard monochromatic diffraction geometry, is not viable. The Laue method uses a stationary crystal for each exposure and to record a full angular range of the data requires a number of separate exposures with the crystal orientation reset between each. Substantial methods development has been required to harness this technique for accurate structural analysis, in particular for wavelength normalization, i.e. to bring all the spot intensities onto a common scale recorded from the various wavelengths, and also involving a thorough re-examination of the Laue diffraction geometry. Instrumentation and software engineering development has also been a major effort. The main SR centres for these developments and initiatives have been Cornell (CHESS synchrotron), Daresbury (SRS), University of Chicago and Grenoble (European SR Facility, ESRF). An interesting spin-off has been the application to neutron protein crystallography where neutron reactor 'steady state' sources produce a white spectrum of neutrons.

The longer term promises femtosecond X-ray time resolution capability via the X-ray free electron laser installation proposed for the 'DESY' accelerator physics laboratory in Hamburg (scheduled for 2010 for first beam, if approved for funding). A similar project is proposed at SLAC in Stanford.

Noncrystalline Diffraction

This embraces all other states of order, namely solutions, fibres or amorphous solids. The case of polycrystalline powders is not included in this category. Powders are finding increased importance in smaller molecule 'chemical crystallography' but their use is not widespread in the life sciences where the powder lines would be exceedingly close together, i.e. overlapping. (One test case, for example, is that at ESRF, bacteriorhodopsin.)

Fibres are very important in the life sciences since many naturally occurring fibres exist, such as DNA, muscle, collagen, silk and filamentous viruses for example. Helical structures produce a very distinctive signature of a criss-cross diffraction pattern. The high SR intensity is needed for single fibres and/or fast time resolution. In such cases conventional X-ray sources are too weak. Of special significance are muscle fibres and especially time-resolved muscle diffraction under induced electrical twitch. Indeed, the first example of biological diffraction with SR in 1971 involved muscle fibre diffraction as mentioned in the introduction.

Amorphous samples in the life sciences occur naturally, such as connective tissue, bone and tendon. Because these tend to form large samples, high intensity is generally unnecessary, nor is fast recording needed for time-resolved diffraction studies. However, recording of diffraction from tiny areas of a large sample in a scanned matrix of points allows local variations in order and structure to be probed as well as normal and pathological tissues to be compared (e.g. young versus old or normal versus cancerous tissue). Microfocus SR beams on third generation sources have had a big impact here. In one instance a sample of hair from Napoleon (the diameter of a hair is around only 20 μm) was examined to determine the arsenic content (it is thought that he was systematically poisoned by his British captors on St Helena).

Solutions of dissolved proteins are used to determine the radius of gyration and molecular weight of the macromolecule in question from the 'small angle X-ray scattering pattern' produced. Synchrotron radiation is used where time-induced conformational re-arrangements are made. Examples include virus assembly or disassembly, enzyme domain re-arrangements, etc. A closely related technique is 'small angle neutron scattering' (SANS). With the SANS technique, by altering the H_2O/D_2O mix or the H/D balance, the contrast of macromolecules against solvent or protein versus protein or protein versus nucleic acid in complexes can be varied. The SANS technique allows then the dissection by structure analysis of complex molecular systems including molecular machines, e.g. the ribosome, chaperonin complex, viruses, etc.

Other Probes of Structure; Neutron and Electron Diffraction and NMR Spectroscopy

It is worth emphasizing that some 40% of a given genome involves membrane-bound proteins and many of these (most?) may not be amenable to crystallization. Electron diffraction of two-dimensional arrays allows for a different approach. Also it is possible to record electron microscope images of large numbers of individual molecules in many different orientations on an em sample grid to reconstruct a three-dimensional image of the whole molecule in the computer. This has been particularly successful for obtaining medium resolution (10 Å) of ribosomal particles already; and improvements towards 5-Å resolution are taking place. Nuclear magnetic resonance (NMR) is a completely different approach which uses the radio frequency resonance signals of different isotopes to elucidate structure. The use of NMR for structure determination for proteins with molecular weights up to 20 kDa is routine, and 30 kDa is feasible. The big advantage is that the protein does not need to be crystallized as a solution is used. The precision of the atomic positions is not as good as from a protein crystal structure analysis but if a crystal cannot be obtained the NMR structure is invaluable. In another category of study by NMR of protein ligand binding, weakly interacting complexes can be studied whereas protein crystallography requires at least approximately 30% occupancy of binding to pick up the structural details in the difference electron density image that is determined. In rational drug design programmes, i.e. involving computer design, weak ligand binding may result and so NMR studies can be used to guide the design process through this stage.

An Overview of Achievements of SR Diffraction in the Life Sciences

In the 20 years of wide SR usage in structural studies involving diffraction techniques there have been spectacular successes, especially for huge multi-macromolecular complexes such as viruses and ribosomes at second-generation and now also third-generation SR sources (i.e. for yet larger complexes). Sub-nanosecond time-resolved protein crystallography has been realized at third-generation SR sources using the Laue method. Tiny crystals and single fibres can now yield usable diffraction data at third-generation SR sources. High-throughput protein crystal structure determination, 'structural genomics', is becoming a reality where the step of optimizing the crystal size can be relaxed and MAD phasing is applied. As more and more dedicated SR facilities become available and more easily accessible geographically, i.e. on a regional as well as a national or an international basis, the number, range and type of application will be much more diverse and the numbers of structures available will increase dramatically.